Eberhard Zeidler

Quantum Field Theory

I Basics in Mathematics and Physics

A Bridge Between Mathematicians and Physicists

Springer

Quantum Field Theory I:
Basics in Mathematics and Physics

Eberhard Zeidler

Quantum Field Theory I: Basics in Mathematics and Physics

A Bridge between Mathematicians and Physicists

With 94 Figures and 19 Tables

Springer

Eberhard Zeidler
Max Planck Institute
for Mathematics in the Sciences
Inselstrasse 22
04103 Leipzig
Germany
e-mail: ezeidler@mis.mpg.de

Library of Congress Control Number: 2006929535

Mathematics Subject Classification (2000): 35Qxx, 58-xx, 81Txx, 82-xx, 83Cxx

ISBN-10 3-540-34762-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-34762-0 Springer Berlin Heidelberg New York

Typesetting: by the author using a Springer LaTeX macro package
Production: LE-TeX Jelonek, Schmidt & Vöckler GbR, Leipzig
Cover design: Erich Kirchner, Heidelberg

Printed on acid-free paper 46/3100/YL 5 4 3 2 1 0

TO THE MEMORY OF JÜRGEN MOSER
(1928–1999)

Preface

Daß ich erkenne, was die Welt im Innersten zusammenhält.[1]
Faust

Concepts without intuition are empty, intuition without concepts is blind.
Immanuel Kant (1724–1804)

The greatest mathematicians like Archimedes, Newton, and Gauss have always been able to combine theory and applications into one.
Felix Klein (1849–1925)

The present comprehensive introduction to the mathematical and physical aspects of quantum field theory consists of the following six volumes:

Volume I: Basics in Mathematics and Physics
Volume II: Quantum Electrodynamics
Volume III: Gauge Theory
Volume IV: Quantum Mathematics
Volume V: The Physics of the Standard Model
Volume VI: Quantum Gravity and String Theory.

Since ancient times, both physicists and mathematicians have tried to understand the forces acting in nature. Nowadays we know that there exist four fundamental forces in nature:

- Newton's gravitational force,
- Maxwell's electromagnetic force,
- the strong force between elementary particles, and
- the weak force between elementary particles (e.g., the force responsible for the radioactive decay of atoms).

In the 20th century, physicists established two basic models, namely,

- the Standard Model in cosmology based on Einstein's theory of general relativity, and
- the Standard Model in elementary particle physics based on gauge theory.

[1] So that I may perceive whatever holds the world together in its inmost folds. The alchemist Georg Faust (1480–1540) is the protagonist of Goethe's drama *Faust* written in 1808.

One of the greatest challenges of the human intellect is the discovery of a unified theory for the four fundamental forces in nature based on first principles in physics and rigorous mathematics. For many years, I have been fascinated by this challenge. When talking about this challenge to colleagues, I have noticed that many of my colleagues in mathematics complain about the fact that it is difficult to understand the thinking of physicists and to follow the pragmatic, but frequently non-rigorous arguments used by physicists. On the other hand, my colleagues in physics complain about the abstract level of the modern mathematical literature and the lack of explicitly formulated connections to physics. This has motivated me to write the present book and the volumes to follow.

It is my intention to build a bridge between mathematicians and physicists.

The main ideas of this treatise are described in the Prologue to this book. The six volumes address a broad audience of readers, including both undergraduate students and graduate students as well as experienced scientists who want to become familiar with the mathematical and physical aspects of the fascinating field of quantum field theory. In some sense, we will start from scratch:

- For students of mathematics, I would like to show that detailed knowledge of the physical background helps to motivate the mathematical subjects and to discover interesting interrelationships between quite different mathematical questions.
- For students of physics, I would like to introduce fairly advanced mathematics which is beyond the usual curriculum in physics.

For historical reasons, there exists a gap between the language of mathematicians and the language of physicists. I want to bridge this gap.[2] I will try to minimize the preliminaries such that undergraduate students after two years of studies should be able to understand the main body of the text. In writing this monograph, it was my goal to follow the advise given by the poet Johann Wolfgang von Goethe (1749–1832):

Textbooks should be attractive by showing the beauty of the subject.

Ariadne's thread. In the author's opinion, the most important prelude to learning a new subject is strong motivation. Experience shows that highly motivated students are willing to take great effort to learn sophisticated subjects.

I would like to put the beginning of Ariadne's thread into the hands of the reader.

[2] On November 7th 1940, there was a famous accident in the U.S.A. which was recorded on film. The Tacoma Narrows Bridge broke down because of unexpected nonlinear resonance effects. I hope that my bridge between mathematicians and physicists is not of Tacoma type.

Remember the following myth. On the Greek island of Crete in ancient times, there lived the monster Minotaur, half human and half bull, in a labyrinth. Every nine years, seven virgins and seven young men had to be sacrificed to the Minotaur. Ariadne, the daughter of King Minos of Crete and Pasiphaë fell in love with one of the seven young men – the Athenian Prince Theseus. To save his life, Ariadne gave Theseus a thread of yarn, and he fixed the beginning of the thread at the entrance of the labyrinth. After a hard fight, Theseus killed the Minotaur, and he escaped from the labyrinth by the help of Ariadne's thread.[3] For hard scientific work, it is nice to have a kind of Ariadne's thread at hand. The six volumes cover a fairly broad spectrum of mathematics and physics. In particular, in the present first volume the reader gets information about

- the physics of the Standard Model of particle physics and
- the magic formulas in quantum field theory,

and we touch the following mathematical subjects:

- finite-dimensional Hilbert spaces and a rigorous approach to the basic ideas of quantum field theory,
- elements of functional differentiation and functional integration,
- elements of probability theory,
- calculus of variations and the principle of critical action,
- harmonic analysis and the Fourier transform, the Laplace transform, and the Mellin transform,
- Green's functions, partial differential equations, and distributions (generalized functions),
- Green's functions, the Fourier method, and functional integrals (path integrals),
- the Lebesgue integral, general measure integrals, and Hilbert spaces,
- elements of functional analysis and perturbation theory,
- the Dirichlet principle as a paradigm for the modern Hilbert space approach to partial differential equations,
- spectral theory and rigorous Dirac calculus,
- analyticity,
- calculus for Grassmann variables,
- many-particle systems and number theory,
- Lie groups and Lie algebras,
- basic ideas of differential and algebraic topology (homology, cohomology, and homotopy; topological quantum numbers and quantum states).

We want to show the reader that many mathematical methods used in quantum field theory can be traced back to classical mathematical problems. In

[3] Unfortunately, Theseus was not grateful to Ariadne. He deserted her on the Island of Naxos, and she became the bride of Dionysus. Richard Strauss composed the opera *Ariadne on Naxos* in 1912.

particular, we will thoroughly study the relation of the procedure of renormalization in physics to the following classical mathematical topics:

- singular perturbations, resonances, and bifurcation in oscillating systems (renormalization in a nutshell on page 625),
- the regularization of divergent infinite series, divergent infinite products, and divergent integrals,
- divergent integrals and distributions (Hadamard's finite part of divergent integrals),
- the passage from a finite number of degrees of freedom to an infinite number of degrees of freedom and the method of counterterms in complex analysis (the Weierstrass theorem and the Mittag–Leffler theorem),
- analytic continuation and the zeta function in number theory,
- Poincaré's asymptotic series and the Ritt theorem in complex analysis,
- the renormalization group and Lie's theory of dynamical systems (one-parameter Lie groups),
- rigorous theory of finite-dimensional functional integrals (path integrals).

The following volumes will provide the reader with important additional material. A summary can be found in the Prologue on pages 11 through 15.

Additional material on the Internet. The interested reader may find additional material on my homepage:

Internet: `www.mis.mpg.de/ezeidler/`

This concerns a carefully structured panorama of important literature in mathematics, physics, history of the sciences and philosophy, along with a comprehensive bibliography. One may also find a comprehensive list of mathematicians, physicists, and philosophers (from ancient until present time) mentioned in the six volumes. My homepage also allows links to the leading centers in elementary particle physics: CERN (Geneva, Switzerland), DESY (Hamburg, Germany), FERMILAB (Batavia, Illinois, U.S.A.), KEK (Tsukuba, Japan), and SLAC (Stanford University, California, U.S.A.). One may also find links to the following Max Planck Institutes in Germany: Astronomy (Heidelberg), Astrophysics (Garching), Complex Systems in Physics (Dresden), Albert Einstein Institute for Gravitational Physics (Golm), Mathematics (Bonn), Nuclear Physics (Heidelberg), Werner Heisenberg Institute for Physics (Munich), and Plasmaphysics (Garching).

Apology. The author apologizes for his imperfect English style. In the preface to his monograph *The Classical Groups*, Princeton University Press, 1946, Hermann Weyl writes the following:

> The gods have imposed upon my writing the yoke of a foreign tongue that was not sung at my cradle.
>
> "Was das heissen will, weiss jeder,
> Der im Traum pferdlos geritten ist,"[4]

[4] Everyone who has dreamt of riding free, without the need of a horse, will know what I mean.

I am tempted to say with the Swiss poet Gottfried Keller (1819–1890). Nobody is more aware than myself of the attendant loss in vigor, ease and lucidity of expression.

Acknowledgements. First of all I would like to thank the Max Planck Society in Germany for founding the Max Planck Institute for Mathematics in the Sciences (MIS) in Leipzig in 1996 and for creating a superb scientific environment here. This treatise would have been impossible without the extensive contacts of the institute to mathematicians and physicists all over the world and without the excellent library of the institute. My special thanks go to the intellectual fathers of the institute, Friedrich Hirzebruch (chairman of the Founder's Committee) and Stefan Hildebrandt in Bonn, Karl-Heinz Hoffmann and Julius Wess in Munich, and the late Jürgen Moser in Zurich who was an external scientific member of the institute. I would like to dedicate this volume to the memory of Jürgen Moser who was a great mathematician and an amiable man. Moreover, I would like to thank Don Zagier (Max Planck Institute for Mathematics in Bonn and Collège de France in Paris), one of the greatest experts in number theory, for the kindness of writing a beautiful section on useful techniques of number theory in physics.

I am very grateful to numerous colleagues in mathematics and physics from all over the world for illuminating discussions. It is not possible to mention the names of all of them, since the list is very long. In particular, I would like to thank the professors from the Institute of Theoretical Physics at Leipzig University, Bodo Geyer, Wolfhard Janke, Gerd Rudolph, Manfred Salmhofer, Klaus Sibold, Armin Uhlmann, and Rainer Verch for nice cooperation. For many stimulating discussions on a broad spectrum of mathematical problems, I would like to thank the co-directors of the MIS, Wolfgang Hackbusch, Jürgen Jost, and Stefan Müller.

For getting information about new research topics, I am very grateful to my former and present collaborators: Günther Berger, Ludmilla Bordag, Friedemann Brandt, Friedemann Brock, Chand Devchand, Bertfried Fauser, Felix Finster, Christian Fleischhack, Jörg Frauendiener, Hans-Peter Gittel, Matthias Günther, Bruce Hunt, Konrad Kaltenbach, Satyanad Kichenassamy, Klaus Kirsten, Christian Klein, Andreas Knauf, Alexander Lange, Roland Matthes, Johannes Maul†, Erich Miersemann, Mario Paschke, Hoang Xuan Phu, Karin Quasthoff, Olaf Richter†, Alexander Schmidt, Rainer Schumann, Friedemann Schuricht, Peter Senf†, Martin Speight, Jürgen Tolksdorf, Dimitri Vassilevich, Hartmut Wachter, and Raimar Wulkenhaar.

For experienced assistance in preparing this book, I would like to thank Kerstin Fölting (graphics, tables, and a meticulous proof-reading of my entire latex-file together with Rainer Munck), Micaela Krieger–Hauwede (graphics, tables, and layout), and Jeffrey Ovall (checking and improving my English style). For supporting me kindly in various aspects and for helping me to save time, I am also very grateful to my secretary, Regine Lübke, and to the staff of the institute including the librarians directed by Ingo Brüggemann,

the computer group directed by Rainer Kleinrensing, and the administration directed by Dietmar Rudzik. Finally, I would like to thank the Springer-Verlag for a harmonious collaboration.

I hope that the reader of this book enjoys getting a feel for the unity of mathematics and physics by discovering interrelations between apparently completely different subjects.

Leipzig, Fall 2005 *Eberhard Zeidler*

Contents

Prologue

We begin with some quotations which exemplify the philosophical underpinnings of this work.

Theoria cum praxi.
Gottfried Wilhelm Leibniz (1646–1716)

> It is very difficult to write mathematics books today. If one does not take pains with the fine points of theorems, explanations, proofs and corollaries, then it won't be a mathematics book; but if one does these things, then the reading of it will be extremely boring.
>
> Johannes Kepler (1571–1630)
> *Astronomia Nova*

> The interaction between physics and mathematics has always played an important role. The physicist who does not have the latest mathematical knowledge available to him is at a distinct disadvantage. The mathematician who shies away from physical applications will most likely miss important insights and motivations.
>
> Marvin Schechter
> *Operator Methods in Quantum Mechanics*[5]

> In 1967 Lenard and I found a proof of the stability of matter. Our proof was so complicated and so unilluminating that it stimulated Lieb and Thirring to find the first decent proof. Why was our proof so bad and why was theirs so good? The reason is simple. Lenard and I began with mathematical tricks and hacked our way through a forest of inequalities without any physical understanding. Lieb and Thirring began with physical understanding and went on to find the appropriate mathematical language to make their understanding rigorous. Our proof was a dead end. Theirs was a gateway to the new world of ideas collected in this book.
>
> Freeman Dyson
> From the Preface to *Elliott Lieb's Selecta*[6]

The state of the art in quantum field theory. One of the intellectual fathers of quantum electrodynamics is Freeman Dyson (born in 1923) who

[5] North-Holland, Amsterdam, 1982.
[6] Stability of Matter: From Atoms to Stars, Springer, New York, 2002.

works at the Institute for Advanced Study in Princeton.[7] He characterizes the state of the art in quantum field theory in the following way:

> All through its history, quantum field theory has had two faces, one looking outward, the other looking inward. The outward face looks at nature and gives us numbers that we can calculate and compare with experiments. The inward face looks at mathematical concepts and searches for a consistent foundation on which to build the theory. The outward face shows us brilliantly successful theory, bringing order to the chaos of particle interactions, predicting experimental results with astonishing precision. The inward face shows us a deep mystery. After seventy years of searching, we have found no consistent mathematical basis for the theory. When we try to impose the rigorous standards of pure mathematics, the theory becomes undefined or inconsistent. From the point of view of a pure mathematician, the theory does not exist. This is the *great unsolved paradox of quantum field theory.*
>
> To resolve the paradox, during the last twenty years, quantum field theorists have become string-theorists. String theory is a new version of quantum field theory, exploring the mathematical foundations more deeply and entering a new world of multidimensional geometry. String theory also brings gravitation into the picture, and thereby unifies quantum field theory with general relativity. String theory has already led to important advances in pure mathematics. It has not led to any physical predictions that can be tested by experiment. We do not know whether string theory is a true description of nature. All we know is that it is a rich treasure of new mathematics, with an enticing promise of new physics. During the coming century, string theory will be intensively developed, and, if we are lucky, tested by experiment.[8]

Five golden rules. When writing the latex file of this book on my computer, I had in mind the following five quotations. Let me start with the mathematician Hermann Weyl (1885–1930) who became a follower of Hilbert in Göttingen in 1930 and who left Germany in 1933 when the Nazi regime came to power. Together with Albert Einstein (1879–1955) and John von Neumann (1903–1957), Weyl became a member of the newly founded Institute for Advanced Study in Princeton, New Jersey, U.S.A. in 1933. Hermann Weyl wrote in 1938:[9]

> The stringent precision attainable for mathematical thought has led many authors to a mode of writing which must give the reader an impression of being shut up in a brightly illuminated cell where every detail sticks out with the same dazzling clarity, but without relief. I prefer the open landscape under a clear sky with its depth of perspective, where the wealth of sharply defined nearby details gradually fades away towards the horizon.

[7] F. Dyson, Selected Papers of Freeman Dyson with Commentaries, Amer. Math. Soc., Providence, Rhode Island, 1996. We recommend reading this fascinating volume.

[8] In: Quantum Field Theory, A 20th Century Profile. Edited by A. Mitra, Indian National Science Academy and Hindustan Book Agency, 2000 (reprinted with permission).

[9] H. Weyl, The Classical Groups, Princeton University Press, 1938 (reprinted with permission).

For his fundamental contributions to electroweak interaction inside the Standard Model in particle physics, the physicist Steven Weinberg (born 1933) was awarded the Nobel prize in physics in 1979 together with Sheldon Glashow (born 1932) and Abdus Salam (1926–1996). On the occasion of a conference on the interrelations between mathematics and physics in 1986, Weinberg pointed out the following: [10]

> I am not able to learn any mathematics unless I can see some problem I am going to solve with mathematics, and I don't understand how anyone can teach mathematics without having a battery of problems that the student is going to be inspired to want to solve and then see that he or she can use the tools for solving them.

For his theoretical investigations on parity violation under weak interaction, the physicist Cheng Ning Yang (born 1922) was awarded the Nobel prize in physics in 1957 together with Tsung Dao Lee (born 1926). In an interview, Yang remarked:[11]

> In 1983 I gave a talk on physics in Seoul, South Korea. I joked "There exist only two kinds of modern mathematics books: one which you cannot read beyond the first page and one which you cannot read beyond the first sentence. The *Mathematical Intelligencer* later reprinted this joke of mine. But I suspect many mathematicians themselves agree with me.

The interrelations between mathematics and modern physics have been promoted by Sir Michael Atiyah (born 1929) on a very deep level. In 1966, the young Atiyah was awarded the Fields medal. In an interview, Atiyah emphasized the following: [12]

> The more I have learned about physics, the more convinced I am that physics provides, in a sense, the deepest applications of mathematics. The mathematical problems that have been solved, or techniques that have arisen out of physics in the past, have been the lifeblood of mathematics... The really deep questions are still in the physical sciences. For the health of mathematics at its research level, I think it is very important to maintain that link as much as possible.

The development of modern quantum field theory has been strongly influenced by the pioneering ideas of the physicist Richard Feynman (1918–1988). In 1965, for his contributions to the foundation of quantum electrodynamics, Feynman was awarded the Nobel prize in physics together with Julian Schwinger (1918–1994) and Sin-Itiro Tomonaga (1906–1979). In the beginning of the 1960s, Feynman held his famous *Feynman lectures* at the California Institute of Technology in Pasadena. In the preface to the printed version of the lectures, Feynman told his students the following:

> Finally, may I add that the main purpose of my teaching has not been to prepare you for some examination – it was not even to prepare you to

[10] Notices Amer. Math. Soc. **33** (1986), 716–733 (reprinted with permission).
[11] Mathematical Intelligencer **15** (1993), 13–21 (reprinted with permission).
[12] Mathematical Intelligencer **6** (1984), 9–19 (reprinted with permission).

> serve industry or military. I wanted most to give you some appreciation of the wonderful world and the physicist's way of looking at it, which, I believe, is a major part of the true culture of modern times.[13]

The fascination of quantum field theory. As a typical example, let us consider the anomalous magnetic moment of the electron. This is given by the following formula

$$\boxed{\mathbf{M}_e = -\frac{e}{2m_e}\, g_e \mathbf{S}}$$

with the so-called gyromagnetic factor

$$g_e = 2(1 + a)$$

of the electron. Here, m_e is the mass of the electron, $-e$ is the negative electric charge of the electron. The spin vector $\mathbf{S}$ has the length $\hbar/2$, where h denotes Planck's quantum of action, and $\hbar := h/2\pi$. High-precision experiments yield the value

$$a_{\text{exp}} = 0.001\,159\,652\,\mathit{188}\;\mathit{4} \pm 0.000\,000\,000\,\mathit{004}\;\mathit{3}.$$

Quantum electrodynamics is able to predict this result with high accuracy. The theory yields the following value

$$a = \frac{\alpha}{2\pi} - 0.328\,478\,965\left(\frac{\alpha}{\pi}\right)^2 + (1.175\,62 \pm 0.000\,56)\left(\frac{\alpha}{\pi}\right)^3 - (1.472 \pm 0.152)\left(\frac{\alpha}{\pi}\right)^4 \tag{0.1}$$

with the electromagnetic fine structure constant

$$\alpha = \frac{1}{137.035\,989\,\mathit{500} \pm 0.000\,000\,\mathit{061}}.$$

Explicitly,

$$a = 0.001\,159\,652\,\mathit{164} \pm 0.000\,000\,000\,\mathit{108}.$$

The error is due to the uncertainty of the electromagnetic fine structure constant α. Observe that 9 digits coincide between the experimental value a_{exp} and the theoretical value a.

The theoretical result (0.1) represents a highlight in modern theoretical physics. The single terms with respect to powers of the fine structure constant α have been obtained by using the method of perturbation theory. In order to represent graphically the single terms appearing in perturbation theory, Richard Feynman (1918–1988) invented the language of Feynman diagrams in about 1945.[14] For example, Fig. 0.1 shows some simple Feynman diagrams

[13] R. Feynman, R. Leighton, and M. Sands, The Feynman Lectures in Physics, Addison-Wesley, Reading, Massachusetts, 1963.

[14] For the history of this approach, see the quotation on page 27.

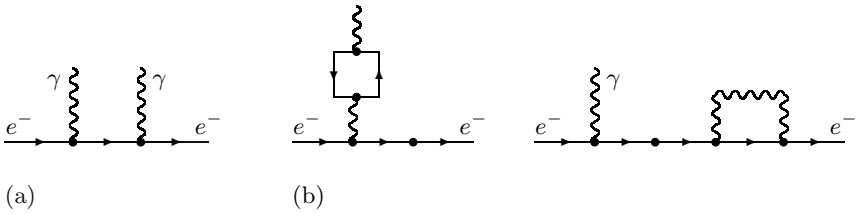


Fig. 0.1. Feynman diagrams

for the Compton scattering between electrons and photons. In higher order of perturbation theory, the Feynman diagrams become more and more complex. In particular, in order to get the α^3-term of (0.1), one has to use 72 Feynman diagrams. The computation of the α^3-term has taken 20 years. The α^4-term from (0.1) is based on 891 Feynman diagrams. The computation has been done mainly by numerical approximation methods. This needed years of supercomputer time.[15] The mathematical situation becomes horrible because of the following fact.

Many of the Feynman diagrams correspond to divergent higher-dimensional integrals called algebraic Feynman integrals.

Physicists invented the ingenious method of renormalization in order to give the apparently meaningless integrals a precise interpretation. Renormalization plays a fundamental role in quantum field theory. Physicists do not expect that the perturbation series (0.1) is part of a convergent power series expansion with respect to the variable α at the origin. Suppose that there would exist such a convergent power series expansion

$$a = \sum_{n=1}^{\infty} a_n \alpha^n, \qquad |\alpha| \leq \alpha_0$$

near the origin $\alpha = 0$. This series would then converge for small negative values of α. However, such a negative coupling constant would correspond to a repelling force which destroys the system. This argument is due to Dyson.[16]

Therefore, we do not expect that the series (0.1) *is convergent.*

In Sect. 15.5.2, we will show that each formal power series expansion is indeed the asymptotic expansion of some analytic function in an angular domain, by the famous 1916 Ritt theorem in mathematics.

[15] See M. Veltman, Facts and Mysteries in Elementary Particle Physics, World Scientific, Singapore, 2003; this is a beautiful history of modern elementary particle physics.

[16] F. Dyson, Divergence of perturbation theory in quantum electrodynamics, Phys. Rev. **85** (1952), 631–632.

From the mathematical point of view, the best approach to renormalization was created by Epstein and Glaser in 1973. The Epstein–Glaser theory avoids the use of divergent integrals and their regularization, but relies on the power of the modern theory of distributions (generalized functions).

Physicists have also computed the magnetic moment of the myon. As for the electron, the coincidence between theory and experiment is of fantastic accuracy. Here, the theory takes all of the contributions coming from electromagnetic, weak, strong, and gravitative interaction into account.[17]

> *It is a challenge for the mathematics of the future to completely understand formula* (0.1).

Let us now briefly discuss the content of Volumes I through VI of this monograph.

Volume I. The first volume entitled *Basics in Mathematics and Physics* is structured in the following way.

Part I: Introduction
- Chapter 1: Historical Introduction
- Chapter 2: Phenomenology of the Standard Model in Particle Physics
- Chapter 3: The Challenge of Different Scales in Nature.

Part II: Basic Techniques in Mathematics
- Chapter 4: Analyticity
- Chapter 5: A Glance at Topology
- Chapter 6: Many-Particle Systems
- Chapter 7: Rigorous Finite-Dimensional Magic Formulas of Quantum Field Theory
- Chapter 8: Rigorous Finite-Dimensional Perturbation Theory
- Chapter 9: Calculus for Grassmann Variables
- Chapter 10: Infinite-Dimensional Hilbert Spaces
- Chapter 11: Distributions and Green's Functions
- Chapter 12: Distributions and Quantum Physics.

Part III: Heuristic Magic Formulas of Quantum Field Theory
- Chapter 13: Basic Strategies in Quantum Field Theory
- Chapter 14: The Response Approach
- Chapter 15: The Operator Approach
- Chapter 16: Peculiarities of Gauge Theories
- Chapter 17: A Panorama of the Literature.

Describing the content of Volume I by a parable, we will first enter a mountain railway in order to reach easily and quickly the top of the desired mountain and to admire the beautiful mountain ranges. Later on we will try to climb to the top along the rocks.

[17] See M. Böhm, A. Denner, and H. Joos, Gauge Theories of the Strong and Electroweak Interaction, Teubner, Stuttgart, 2001, p. 80.

In particular, the heuristic magic formulas from Part III should help the reader to understand quickly the language of physicists in quantum field theory. These magic formulas are non-rigorous from the mathematical point of view, but they are extremely useful for computing physical effects.

Modern elementary particle physics is based on the Standard Model in particle physics introduced in the late 1960s and the early 1970s. Before studying thoroughly the Standard Model in the next volumes, we will discuss the phenomenology of this model in the present volume. It is the goal of quantum field theory to compute

- the cross sections of scattering processes in particle accelerators which characterize the behavior of the scattered particles,
- the masses of stable elementary particles (e.g., the proton mass as a bound state of three quarks), and
- the lifetime of unstable elementary particles in particle accelerators.

To this end, physicists use the methods of perturbation theory. Fortunately enough, the computations can be based on only a few basic formulas which we call magic formulas. The magic formulas of quantum theory are extremely useful for describing the experimental data observed in particle accelerators, but they are only valid on a quite formal level.

This difficulty is typical for present quantum field theory.

To help the reader in understanding the formal approach used in physics, we consider the finite-dimensional situation in Chapter 6.

In the finite-dimensional case, we will rigorously prove all of the magic formulas used by physicists in quantum field theory.

Furthermore, we relate physics to the following fields of mathematics:

- causality and the analyticity of complex-valued functions,
- many-particle systems, the Casimir effect in quantum field theory, and number theory,
- propagation of physical effects, distributions (generalized functions), and the Green's function,
- rigorous justification of the elegant Dirac calculus,
- duality in physics (time and energy, time and frequency, position and momentum) and harmonic analysis (Fourier series, Fourier transformation, Laplace transformation, Mellin transformation, von Neumann's general operator calculus for self-adjoint operators, Gelfand triplets and generalized eigenfunctions),
- the relation between renormalization, resonances, and bifurcation,
- dynamical systems, Lie groups, and the renormalization group,
- fundamental limits in physics,
- topology in physics (Chern numbers and topological quantum numbers),
- probability, Brownian motion, and the Wiener integral,

- the Feynman path integral,
- Hadamard's integrals and algebraic Feynman integrals.

In fact, this covers a broad range of physical and mathematical subjects.

Volume II. The second volume entitled *Quantum Electrodynamics* consists of the following parts.

Part I : Introduction
- Chapter 1: Mathematical Principles of Natural Philosophy
- Chapter 2: The Basic Strategy of Extracting Finite Information from Infinities
- Chapter 3: A Glance at the Mathematical Structure Behind Renormalization in Physics.

Part II : Basic Ideas in Quantum Mechanics
- Chapter 4: The Principle of Critical Action and the Harmonic Oscillator as a Paradigm
- Chapter 5: Quantization of the Harmonic Oscillator.

Part III : Scattering Processes in Quantum Mechanics
- Chapter 6: Quantum Particles on the Real Line – Ariadne's Thread in Scattering Theory
- Chapter 7: Three-Dimensional Motion of a Quantum Particle
- Chapter 8: Observables and Operator Theory – the Trouble with Infinite Dimensions
- Chapter 9: The Hydrogen Atom as a Paradigm in Functional Analysis
- Chapter 10: Large Atoms and Molecules.

Part IV : Relativistic Fields
- Chapter 11: Einstein's Theory of Special Relativity
- Chapter 12: The Electromagnetic Field
- Chapter 13: Dirac's Relativistic Electron.

Part V : Basic Ideas of Quantum Field Theory
- Chapter 14: Chain of Quantized Harmonic Oscillators
- Chapter 15: Quantum Electrodynamics.

Part VI : Renormalization
- Chapter 16: Radiative Corrections
- Chapter 17: A Glance at the Bogoliubov–Parasiuk–Hepp–Zimmermann Renormalization
- Chapter 18: The Beauty of the Epstein–Glaser Approach
- Chapter 19: A Glance at Algebraic Renormalization
- Chapter 20: The Renormalization Flow Method.

The final goal of quantum field theory is the foundation of a rigorous mathematical theory which contains the Standard Model as a special low-energy approximation. At present we are far away from reaching this final goal. From the physical point of view, the most successful quantum field theory is quantum electrodynamics. This will be studied in Volume II along with some

applications to important physical processes like Compton scattering between electrons and photons, the spontaneous emission of light by molecules, Cherenkov radiation of fast electrons, the Lamb shift in the hydrogen spectrum, the anomalous magnetic moment of the electron, and the Hawking radiation of black holes. We also study the physics and mathematics behind the crucial phenomenon of renormalization and the change of scales in physics culminating in the modern theory of the renormalization group. Generally, we try to include both interesting mathematics and interesting physics. In particular, we will discuss the relation of renormalization in physics to the following mathematical subjects: Euler's gamma function, the Riemann–Liouville integral, and dimensional regularization; Borel summation of divergent series; pseudo-convergence of iterative methods for ill-posed problems, Hopf algebras and Rota–Baxter algebras; theory of categories; wave front sets and the theory of distributions, Euler's and Feynman's mathemagics.

Volume III. The fundamental forces in the universe are described by gauge field theories which generalize both Gauss' surface theory and Maxwell's theory of electromagnetism. The third volume entitled *Gauge Theories* is divided into the following parts.

Part I : The Euclidean Space as a Paradigm
- Chapter 1: The Algebraic Structure of the Euclidean Space
- Chapter 2: The Differential Structure of the Euclidean Space
- Chapter 3: Changing Observers and Tensor Analysis.

Part II: Interactions and Gauge Theory
- Chapter 4: Basic Principles in Physics
- Chapter 5: Observers, Physical Fields, and Bundles
- Chapter 6: Symmetry Breaking in Physics
- Chapter 7: Gauss' Surface Theory.

Part III: Fundamental Gauge Theories in Physics
- Chapter 8: Einstein's Theory of Special Relativity
- Chapter 9: Maxwell's Theory of Electromagnetism
- Chapter 10: Dirac's Relativistic Electron
- Chapter 11: The Standard Model in Particle Physics
- Chapter 12: Einstein's Theory of General Relativity and Cosmology
- Chapter 13: A Glance at String Theory and the Graviton
- Chapter 14: The Sigma Model.

Interestingly enough, it turns out that the Standard Model in particle physics is related to many deep questions in both mathematics and physics. We will see that the question about the structure of the fundamental forces in nature has influenced implicitly or explicitly the development of a large part of mathematics. One of our heros will be Carl Friedrich Gauss (1777–1855), one of the greatest mathematicians of all time. We will encounter his highly influential work again and again. In the German Museum in Munich, one can read the following inscription under Gauss' impressive portrait:

> His spirit lifted the deepest secrets of numbers, space, and nature; he measured the orbits of the planets, the form and the forces of the earth; in his mind he carried the mathematical science of a coming century.

On the occasion of Gauss' death, Sartorius von Waltershausen wrote the following in 1855:

> From time to time in the past, certain brilliant, unusually gifted personalities have arisen from their environment, who by virtue of the creative power of their thoughts and the energy of their actions have had such an overall positive influence on the intellectual development of mankind, that they at the same time stand tall as markers between the centuries... Such epoch-making mental giants in the history of mathematics and the natural sciences are Archimedes of Syracuse in ancient times, Newton toward the end of the dark ages and Gauss in our present day, whose shining, glorious career has come to an end after the cold hand of death touched his at one time deeply-thinking head on February 23 of this year.

Another hero will be Bernhard Riemann (1826–1866) – a pupil of Gauss. Riemann's legacy influenced strongly mathematics and physics of the 20th century, as we will show in this treatise.[18]

The two Standard Models in modern physics concerning cosmology and elementary particles are closely related to modern differential geometry. This will be thoroughly studied in Volume III. We will show that both Einstein's general theory of relativity and the Standard Model in particle physics are gauge theories. From the mathematical point of view, the fundamental forces in nature are curvatures of appropriate fiber bundles. Historically, mathematicians have tried to understand the curvature of geometric objects. At the very beginning, there was Gauss' *theorema egregium*[19] telling us that curvature is an intrinsic property of a surface. On the other side, in the history of physics, physicists have tried to understand the forces in nature. Nowadays we know that both mathematicians and physicists have approached the same goal coming from different sides. We can summarize this by saying briefly that

$$\textit{force} = \textit{curvature}.$$

For the convenience of the reader, we will also discuss in Volume II that many of the mathematical concepts arising in quantum field theory are rooted in the geometry of the Euclidean space (e.g., Lie groups and Lie algebras, operator algebras, Grassmann algebras, Clifford algebras, differential forms and cohomology, Hodge duality, projective structures, symplectic structures, contact structures, conformal structures, Riemann surfaces, and supersymmetry).

Volume IV. Quantum physics differs from classical relativistic field theories by adding the process of quantization. From the physical point of view,

[18] We also recommend the beautiful monograph written by Krzysztof Maurin, Riemann's Legacy, Kluwer, Dordrecht, 1997.

[19] The Latin expression *theorema egregium* means *the beautiful theorem.*

there appear additional quantum effects based on random quantum fluctuations. From the mathematical point of view, one has to deform classical theories in an appropriate way. Volume IV is devoted to the mathematical and physical methods of quantization. For this, we coin the term *Quantum Mathematics*. Volume IV represents the first systematic textbook on *Quantum Mathematics*. This volume will be divided into the following parts.

Part I: Finite Quantum Mathematics
- Chapter 1: Many-Particle Systems, Probability, and Information
- Chapter 2: Quantum Systems and Hilbert Spaces
- Chapter 3: Quantum Information.

Part II: Symmetry and Quantization
- Chapter 4: Finite Groups and the Paradigm of Symmetric Functions
- Chapter 5: Compact Lie Groups
- Chapter 6: The Poincaré Group
- Chapter 7: Applications to Analytic S-Matrix Theory
- Chapter 8: The Yang–Baxter Equation, Hopf Algebras, and Quantum Groups.

Part III: Operators Algebras and Quantization
- Chapter 9: States and Observables
- Chapter 10: Local Operator Algebras and Causality.

Part IV: Topology and Quantization
- Chapter 11: Basic Ideas
- Chapter 12: Cohomology and Homology in Physics
- Chapter 13: The Atiyah–Singer Index Theorem and Spectral Geometry.

Part V : Interactions between Mathematics and Physics
- Chapter 14: Geometric Quantization
- Chapter 15: Stochastic Quantization
- Chapter 16: Progress in Mathematics by Using Ideas Originated in Quantum Physics
- Chapter 17: Mathematics – a Cosmic Eye of Humanity.

Typically, quantum fields are interacting physical systems with an infinite number of degrees of freedom and very strong singularities. In mathematics,

- interactions lead to nonlinear terms, and
- infinite-dimensional systems are described in terms of functional analysis.

Therefore, the right mathematical setting for quantum field theory is nonlinear functional analysis. This branch of mathematics has been very successful in the rigorous treatment of nonlinear partial differential equations concerning elasticity and plasticity theory, hydrodynamics, and the theory of general relativity. But the actual state of the art of nonlinear functional analysis does not yet allow for the rigorous investigation of realistic models in quantum field theory, like the Standard Model in particle physics. Physicists say, we cannot wait until mathematics is ready. Therefore, we have to develop our

own non-rigorous methods, and we have to check the success of our methods by comparing them with experimental data. In order to help mathematicians to enter the world of physicists, we will proceed as follows.

(i) Rigorous methods: We first develop quantum mathematics in finite-dimensional spaces. In this case, we can use rigorous methods based on the theory of Hilbert spaces, operator algebras, and discrete functional integrals.
(ii) Formal methods. The formulas from (i) can be generalized in a straightforward, but formal way to infinite-dimensional systems.

This way, the mathematician should learn where the formulas of the physicists come from and how to handle these formulas in order to compute physical effects. What remains is to solve the open problem of rigorous justification.

> *The point will be the investigation of limits and pseudo-limits if the number of particles goes to infinity.*

By a pseudo-limit, we understand the extraction of maximal information from an ill-defined object, as in the method of renormalization. The experience of physicists and mathematicians shows that we cannot expect the limits or pseudo-limits to exist for all possible quantities. The rule of thumb is as follows: concentrate on quantities which can be measured in physical experiments. This seriously complicates the subject. We will frequently encounter the Feynman functional integral. From the mnemonic point of view, this is a marvellous tool. But it lacks mathematical rigor. We will follow the advise given by Evariste Galois (1811–1832):

> Unfortunately what is little recognized is that the most worthwhile scientific books are those in which the author clearly indicates what he does not know; for an author most hurts his readers by concealing difficulties.

Volume V. The mathematician should notice that it is the ultimate goal of a physicist to compute real numbers which can be measured in physical experiments. For reaching this goal, the physicist mixes rigorous arguments with heuristic ones in an ingenious way. In order to make mathematicians familiar with this method of doing science, in Volume V we will study the physics of the Standard Model in particle physics. In particular, we will show how to compute a number of physical effects. In this respect, symmetries will play an important role. For example, this will concern the representation theory of compact Lie groups (e.g., gauge groups in gauge theory), noncompact Lie groups (the Poincaré group and its universal covering group in relativistic physics), infinite-dimensional Lie algebras (e.g., the Virasoro algebra in string theory), and supersymmetric generalizations.

Volume VI. The last volume will be devoted to combining the Standard Model in particle physics with gravitation. We will study several possible approaches to this fascinating, but still completely open problem. The leading candidate is string theory. In connection with the string theory of physicists,

a completely new way of thinking has emerged which we will call *physical mathematics*, a term already used in Kishore Marathe's nice survey article on the role of knot theory in modern mathematics, physics, and biology.[20] Distinguish the following:

- By mathematical physics, we traditionally understand a branch of mathematics which answers questions coming from physics by applying rigorous mathematical methods. The heart of mathematical physics are mathematical proofs (e.g., existence proofs for solutions of partial differential equations or operator equations).
- By physical mathematics, we understand a branch of physics which is motivated by the question about the fundamental forces in nature. Using physical pictures, physicists are able to conjecture deep mathematical results (e.g., the existence and the properties of new topological invariants for manifolds and knots). The heart of physical mathematics is physical intuition, but not the mathematical proof.

The hero of physical mathematics is the physicist Edward Witten (born 1951) from the Institute for Advanced Study in Princeton. At the International Congress of Mathematicians in Kyoto (Japan) in 1990, Witten was awarded the Fields medal. In the last 15 years, physical mathematics was very successful in feeding fascinating new ideas into mathematics. The main method of physical mathematics goes like this:

- start with a model in quantum field theory based on an appropriate Lagrangian;
- quantize this model by means of the corresponding Feynman functional integral;
- extract essential information from the functional integral by using the method of stationary phase.

The point is that this method yields beautiful mathematical conjectures, but it is not able to give rigorous proofs. Unfortunately, for getting proofs, mathematicians have to follow quite different sophisticated routes. It is a challenge to mathematicians to understand better the magic weapon of physical mathematics.

> *The magic weapon of physical mathematics will be called the Witten functor. This functor translates physical structures into mathematical structures.*

With respect to the Witten functor, one observes the following general evolution principle in mathematics.

[20] K. Marathe, A chapter in physical mathematics: theory of knots in the sciences, pp. 873–888. In: Mathematics Unlimited – 2001 and Beyond edited by B. Engquist and W. Schmid, Springer, Berlin, 2001.

(i) From quantity to quality: In the 1920s, modern algebra was founded by passing from concrete mathematical objects like numbers to abstract mathematical structures like groups, rings, fields, and algebras. Here, one only considers the relations between the objects, but not the individual structure of the objects. For example, Emmy Noether emphasized in the 1920s that, in the setting of algebraic topology created by Poincaré at the end of the 19th century, it is very useful to pass from Betti numbers to homology groups. In turn, it was discovered in the 1930s that cohomology groups are in fact richer in structure than homology groups. The point is that cohomology groups possess a natural multiplicative structure which generates the cohomology ring of topological spaces. For example, the product $\mathbb{S}^2 \times \mathbb{S}^4$ of a 2-dimensional sphere with a 4-dimensional sphere has the same homology and cohomology groups as the 3-dimensional complex projective space $\mathbb{P}^3_{\mathbb{C}}$. However, these two manifolds are not topologically equivalent, since their cohomology rings are different.

(ii) Combining abstract structures with each other: For example, Lie groups are obtained by combining the notion of manifold with the notion of group. In turn, fiber bundles occur by combining manifolds with Lie groups.

(iii) Functors between abstract structures: In the late 1940s, the theory of categories emerged in the context of algebraic topology. For example, the Galois functor simplifies the study of field extensions by mapping fields to groups. The Lie functor simplifies the investigation of Lie groups by mapping Lie groups to Lie algebras. Moreover, the homology functor simplifies the structural analysis of topological spaces (geometric objects) by mapping topological spaces to groups called homology groups. Combining the homology functor with the general concept of duality, we arrive at the cohomology functor which maps topological spaces to cohomology groups. Cohomology plays a fundamental role in modern physics.

(iv) Statistics of abstract structures: In physical mathematics, one considers the statistics of physical states in terms of functional integrals. The point is that the states are equivalence classes of mathematical structures. In the language of mathematics, the physical state spaces are moduli spaces. For example, in string theory the states of strings are Riemann surfaces modulo conformal equivalence. Thus, the state space of all those strings which possess a fixed genus g is nothing other than Riemann's famous moduli space $\mathcal{M}_g$ which can be described by a universal covering space of $\mathcal{M}_g$ called the Teichmüller space $\mathcal{T}_g$. Mathematicians know that the theory of moduli spaces is a challenge in algebraic geometry, since such objects carry singularities, as a rule. Physicists expect that those singularities are responsible for essential physical effects.

Another typical feature of physical mathematics is the description of many-particle systems by *partition functions* which encode essential information. As we will show, the Feynman functional integral is nothing other than a

partition function which encodes the essential properties of quantum fields. From the physical point of view, the Riemann zeta function is a partition function for the infinite system of prime numbers. The notion of partition function unifies

- statistical physics,
- quantum mechanics,
- quantum field theory, and
- number theory.

Summarizing, I dare say that

> *The most important notion of modern physics is the Feynman functional integral as a partition function for the states of many-particle systems.*

It is a challenge of mathematics to understand this notion in a better way than known today.

A panorama of mathematics. For the investigation of problems in quantum field theory, we need a broad spectrum of mathematical branches. This concerns

(a) algebra, algebraic geometry, and number theory,
(b) analysis and functional analysis,
(c) geometry and topology,
(d) information theory, theory of probability, and stochastic processes,
(e) scientific computing.

In particular, we will deal with the following subjects:

- Lie groups and symmetry, Lie algebras, Kac–Moody algebras (gauge groups, permutation groups, the Poincaré group in relativistic physics, conformal symmetry),
- graded Lie algebras (supersymmetry between bosons and fermions),
- calculus of variations and partial differential equations (the principle of critical action),
- distributions (also called generalized functions) and partial differential equations (Green's functions, correlation functions, propagator kernels, or resolvent kernels),
- distributions and renormalization (the Epstein–Glaser approach to quantum field theory via the S-matrix),
- geometric optics and Huygens' principle (symplectic geometry, contact transformations, Poisson structures, Finsler geometry),
- Einstein's Brownian motion, diffusion, stochastic processes and the Wiener integral, Feynman's functional integrals, Gaussian integrals in the theory of probability, Fresnel integrals in geometric optics, the method of stationary phase,

- non-Euclidean geometry, covariant derivatives and connections on fiber bundles (Einstein's theory of general relativity for the universe, and the Standard Model in elementary particle physics),
- the geometrization of physics (Minkowski space geometry and Einstein's theory of special relativity, pseudo-Riemannian geometry and Einstein's theory of general relativity, Hilbert space geometry and quantum states, projective geometry and quantum states, Kähler geometry and strings, conformal geometry and strings),
- spectral theory for operators in Hilbert spaces and quantum systems,
- operator algebras and many-particle systems (states and observables),
- quantization of classical systems (method of operator algebras, Feynman's functional integrals, Weyl quantization, geometric quantization, deformation quantization, stochastic quantization, the Riemann–Hilbert problem, Hopf algebras and renormalization),
- combinatorics (Feynman diagrams, Hopf algebras),
- quantum information, quantum computers, and operator algebras,
- conformal quantum field theory and operator algebras,
- noncommutative geometry and operator algebras,
- vertex algebras (sporadic groups, monster and moonshine),
- Grassmann algebras and differential forms (de Rham cohomology),
- cohomology, Hilbert's theory of syzygies, and BRST quantization of gauge field theories,
- number theory and statistical physics,
- topology (mapping degree, Hopf bundle, Morse theory, Lyusternik–Schnirelman theory, homology, cohomology, homotopy, characteristic classes, homological algebra, K-theory),
- topological quantum numbers (e.g., the Gauss–Bonnet theorem, Chern classes, and Chern numbers, Morse numbers, Floer homology),
- the Riemann–Roch–Hirzebruch theorem and the Atiyah–Singer index theorem,
- analytic continuation, functions of several complex variables (sheaf theory),
- string theory, conformal symmetry, moduli spaces of Riemann surfaces, and Kähler manifolds.

The role of proofs. Mathematics relies on proofs based on perfect logic. The reader should note that, in this treatise, the terms

- proposition,
- theorem (important proposition), and
- proof

are used in the rigorous sense of mathematics. In addition, for helping the reader in understanding the basic ideas, we also use 'motivations', 'formal proofs', 'heuristic arguments' and so on, which emphasize intuition, but lack rigor. Because of the rich material to be studied, it is impossible to provide the reader with full proofs for all the different subjects. However, for missing

proofs we add references to carefully selected sources. Many of the missing proofs can be found in the following monographs:

- E. Zeidler, Applied Functional Analysis, Vols. 1, 2, Springer, New York. 1995.
- E. Zeidler, Nonlinear Functional Analysis and its Applications, Vols. 1–4, Springer, New York, 1985–88.

For getting an overview, the reader should also consult the following book:[21]

- E. Zeidler (Ed.), Oxford Users' Guide to Mathematics, Oxford University Press, 2004 (1300 pages).

At the end of the *Oxford Users' Guide to Mathematics*, the interested reader may find a chronology of mathematics and physics from ancient to present times embedded in the cultural history of mankind.

Perspectives. At the International Congress of Mathematicians in Paris in 1900, Hilbert formulated 23 open problems for the mathematics of the 20th century. Many of them have been solved.[22] Hilbert said the following to the audience in 1900:

> Each progress in mathematics is based on the discovery of stronger tools and easier methods, which at the same time makes it easier to understand earlier methods. By making these stronger tools and easier methods his own, it is possible for the individual researcher to orientate himself in the different branches of mathematics...
> When the answer to a mathematical problem cannot be found, then the reason is frequently that we have not recognized the general idea from which the given problem only appears as a link in a chain of related problems...
> The organic unity of mathematics is inherent in the nature of this science, for mathematics is the foundation of all exact knowledge of natural phenomena.

For the 21th century, the open problem of quantum field theory represents a great challenge. It is completely unclear how long the solution of this problem will take. In fact, there are long-term problems in mathematics. As an example, let us consider *Fermat's Last Theorem* where the solution needed more than 350 years. In ancient times, Pythagoras (508–500 B.C.) knew that the equation

$$x^2 + y^2 = z^2$$

has an infinite number of integer solutions (e.g., $x = 3, y = 4, z = 5$). In 1637, Pierre de Fermat (1601–1665), claimed that the equation

$$x^n + y^n = z^n, \qquad n = 3, 4, \dots$$

[21] The German version reads as E. Zeidler, Teubner-Taschenbuch der Mathematik, Vols. 1, 2, Teubner, Wiesbaden, 2003. The English translation of the second volume is in preparation.

[22] See D. Hilbert, Mathematical Problems, Bull. Amer. Math. Soc. **8** (1902), 437–479, and B. Yandell, The Honors Class: Hilbert's Problems and Their Solvers, Natick, Massachusetts, 2001.

has no nontrivial integer solution. In his copy of the *Arithmetica* by Diophantus (250 A.C.), Fermat wrote the following:

> It is impossible to separate a cube into two cubes, or a biquadrate into two biquadrates, or generally any power except a square into two powers with the same exponent. I have discovered a truly marvellous proof of this, which however the margin is not large enough to contain.

The history of this problem can be found in the bestseller by Simon Singh, Fermat's Last Theorem: The Story of a Riddle that Confounded the World's Greatest Minds for 358 Years, Fourth Estate, London, 1997. The final proof was given by Andrew Wiles (born 1953) in Princeton in 1994.[23] The proof, based on the Galois functor, is of extraordinary complexity, and it uses many sophisticated tools from number theory and algebraic geometry developed in the 19th and 20th century. However, in the sense of Hilbert's philosophy for hard problems quoted above, let us describe the basic idea behind the solution. In this connection, it turns out that there is a beautiful geometric result of general interest behind Fermat's Last Theorem.[24] The fundamental geometric result tells us that[25]

(M) *Each elliptic curve is modular.*

Roughly speaking, the proof of Fermat's last theorem proceeds now like this:

(i) Suppose that Fermat's claim is wrong. Then, there exists a nontrivial triplet x, y, z of integers such that $x^n + y^n = z^n$ for some fixed natural number $n \geq 3$.
(ii) The triplet x, y, z can be used in order to construct a specific elliptic curve (the Frey curve), which is not modular, a contradiction to (M).

It remains to sketch the meaning of the geometric principle (M). To begin with, consider the equation of the complex unit circle

$$x^2 + y^2 = 1$$

where x and y are complex parameters. The unit circle allows a parametrization either by periodic functions,

$$x = \cos\varphi, \qquad y = \sin\varphi, \qquad \varphi \in \mathbb{C},$$

or by rational functions,

[23] A. Wiles, Modular elliptic curves and Fermat's Last Theorem, Ann. Math. **142** (1994), 443–551.

[24] We refer to the beautiful lecture given by Don Zagier, Leçon inaugurale, Jeudi 17 Mai 2001, Collège de France, Paris. See also H. Darmon, A proof of the full Shimura–Taniyama–Weil conjecture is announced, Notices Amer. Math. Soc. **46** (1999), 1397–1401. Much background material can be found in the fascinating textbook by Y. Hellagouarch, Invitation to the Mathematics of Fermat–Wiles, Academic Press, New York.

[25] A comprehensive survey article on modular forms can be found in Zagier (1995).

$$x = \frac{2}{1+t^2} - 1, \qquad y = \frac{2t}{1+t^2}, \qquad t \in \mathbb{C},$$

provided we set $t := \tan\frac{\varphi}{2}$. Recall that each compact Riemann surface of genus zero is conformally and topologically equivalent to the real two-dimensional sphere called the Riemann sphere. In particular, the complex unit circle considered above is such a Riemann surface of genus zero. Moreover, compact Riemann surfaces of genus one are conformally and topologically equivalent to some real two-dimensional torus. Such Riemann surfaces are also called elliptic curves. For example, given three pairwise different complex numbers e_1, e_2, e_3, the equation

$$y^2 = 4(x - e_1)(x - e_2)(x - e_3)$$

with complex parameters x and y represents an elliptic curve which allows the global parametrization

$$x = \wp(t), \qquad y = \wp'(t), \qquad t \in \mathbb{C}$$

by the Weierstrass $\wp$-function. This is an elliptic (i.e., double-periodic) function whose two complex periods depend on e_1, e_2, e_3. The fundamental geometric result reads now as follows:

(i) Each compact Riemann surface of genus zero (i.e., each complex curve of circle type) allows two global parametrizations by either periodic functions or rational functions.
(ii) Each compact Riemann surface of genus one (i.e., each elliptic curve) allows two global parametrizations by either double-periodic functions or modular functions.
(iii) Each compact Riemann surface of genus $g \geq 2$ can be globally parametrized by automorphic functions.[26]

The global parametrization (i) of elliptic curves by elliptic functions is one of the most famous results of 19th century mathematics due to Jacobi, Riemann, and Weierstrass. The general result (ii) on the global parametrization of elliptic curves by modular functions was only proved in 1999, i.e., it was shown that the full Shimura–Taniyama–Weil conjecture is true. Statement (iii) represents the famous uniformization theorem for compact Riemann surfaces which was proved independently by Koebe and Poincaré in 1907 after strong efforts made by Poincaré and Klein. The existence of double-periodic functions was discovered by Gauss in 1797 while studying the geometric properties of the lemniscate introduced by Jakob Bernoulli (1654–1705). Therefore, the innocent looking three statements (i), (ii), (iii) above are the result of 200 years of intense mathematical research. Summarizing, in the sense of Hilbert,

[26] Much material on Riemann surfaces, elliptic curves, zeta functions, Galois theory, and so on, can be found in the volume edited by M. Waldschmidt et al., From Number Theory to Physics, Springer, New York, 1995.

the famous Fermat conjecture could finally be solved because it could be reduced to the general idea of modular curves. In a fascinating essay on the future of mathematics, Arthur Jaffe (born 1937) from Harvard University wrote the following:[27]

> Mathematical research should be as broad and as original as possible, with very long range-goals. We expect history to repeat itself: we expect that the most profound and useful future applications of mathematics cannot be predicted today, since they will arise from mathematics yet to be discovered.

Studying the physics and mathematics of the fundamental forces in nature, there arises the question about the philosophical background. Concerning this, let me finish with two quotations. Erich Worbs writes in his Gauss biography:

> Sartorius von Waltershausen reports that Gauss once said there were questions of infinitely higher value than the mathematical ones, namely, those about our relation to God, our determination, and our future. Only, he concluded, their solutions lie far beyond our comprehension, and completely outside the field of science.

In the Harnack Building of the Max-Planck Society in Berlin, one can read the following words by Johann Wolfgang von Goethe:

> The greatest joy of a thinking man is to have explored the explorable and just to admire the unexplorable.

[27] Ordering the universe: the role of mathematics, Notices Amer. Math. Soc. **236** (1984), 589–608.

1. Historical Introduction

If we wish to foresee the future of mathematics our proper course is to study the history and present condition of the science.

Henri Poincaré (1854–1912)

It is not the knowledge but the learning, not the possessing, but the earning, not the being there but the getting there, which gives us the greatest pleasure.

Carl Friedrich Gauss (1777–1855)
to his Hungarian friend Janos Bólyai

For me, as a young man, Hilbert (1858–1943) became the kind of mathematician which I admired, a man with an enormous power of abstract thought, combined with a fully developed sense for the physical reality.

Norbert Wiener (1894–1964)

In the fall 1926, the young John von Neumann (1903–1957) arrived in Göttingen to take up his duties as Hilbert's assistant. These were the hectic years during which quantum mechanics was developing with breakneck speed, with a new idea popping up every few weeks from all over the horizon. The theoretical physicists Born, Dirac, Heisenberg, Jordan, Pauli, and Schrödinger who were developing the new theory were groping for adequate mathematical tools. It finally dawned upon them that their 'observables' had properties which made them look like Hermitean operators in Hilbert space, and that by an extraordinary coincidence, the 'spectrum' of Hilbert (which he had chosen around 1900 from a superficial analogy) was to be the central conception in the explanation of the 'spectra' of atoms. It was therefore natural that they should enlist Hilbert's help to put some mathematical sense in their formal computations. With the assistance of Nordheim and von Neumann, Hilbert first tried integral operators in the space L_2, but that needed the use of the Dirac delta function δ, a concept which was for the mathematicians of that time self-contradictory. John von Neumann therefore resolved to try another approach.

Jean Dieudonné (1906–1992)
History of Functional Analysis[1]

Stimulated by an interest in quantum mechanics, John von Neumann began the work in operator theory which he was to continue as long as he lived. Most of the ideas essential for an abstract theory had already been

[1] North–Holland, Amsterdam, 1981 (reprinted with permission).

> developed by the Hungarian mathematician Fryges Riesz, who had established the spectral theory for bounded Hermitean operators in a form very much like as regarded now standard. Von Neumann saw the need to extend Riesz's treatment to *unbounded* operators and found a clue to doing this in Carleman's highly original work on integral operators with singular kernels...
> The result was a paper von Neumann submitted for publication to the *Mathematische Zeitschrift* but later withdrew. The reason for this withdrawal was that in 1928 Erhard Schmidt and myself, independently, saw the role which could be played in the theory by the concept of the adjoint operator, and the importance which should be attached to *self-adjoint operators.* When von Neumann learned from Professor Schmidt of this observation, he was able to rewrite his paper in a much more satisfactory and complete form... Incidentally, for permission to withdraw the paper, the publisher exacted from Professor von Neumann a promise to write a book on quantum mechanics. The book soon appeared and has become one of the classics of modern physics.[2]
>
> Marshall Harvey Stone (1903–1989)

1.1 The Revolution of Physics

At the beginning of the 20th century, Max Planck (1858–1947) and Albert Einstein (1879–1955) completely revolutionized physics. In 1900, Max Planck derived the universal radiation law for stars by postulating that

The action in our world is quantized.

Let us discuss this fundamental physical principle. The action is the most important physical quantity in nature. For any process, the action is the product of energy $\times$ time for a small time interval. The total action during a fixed time interval is then given by an integral summing over small time intervals. The fundamental principle of least (or more precisely, critical) action tells us the following:

A process in nature proceeds in such a way that the action becomes minimal under appropriate boundary conditions.

More precisely, the action is critical. This means that the first variation of the action S vanishes, $\delta S = 0$. In 1918 Emmy Noether (1882–1935) proved a fundamental mathematical theorem. The famous Noether theorem tells us that

Conservation laws in physics are caused by symmetries of physical systems.

[2] J. von Neumann, Mathematical Foundations of Quantum Mechanics, Princeton University Press, 1955 (first German edition: Springer, Berlin, 1932). This quotation is taken from F. Browder (Ed.), Functional Analysis and Related Fields, Springer, Berlin, 1970 (reprinted with permission).

To explain this basic principle for describing nature in terms of mathematics, consider our solar system. The motion of the sun and the planets only depends on the initial positions and initial velocities. Obviously, the motion of the solar system is invariant under time translations, spatial translations, and rotations. This is responsible for conservation of energy, momentum, and angular momentum, respectively. For example, invariance under time translations means the following. If a process of the physical system is possible, $\mathbf{x} = \mathbf{x}(t)$, then each process is also possible which is obtained by time translation, $\mathbf{x} = \mathbf{x}(t + \text{const})$. According to Planck, the smallest amount of action in nature is equal to

$$\boxed{h = 6.260\,0755 \cdot 10^{-34}\text{Js}} \tag{1.1}$$

where 1 Joule = 1 kg $\cdot$ m^2/s^2. We also introduce $\hbar := h/2\pi$. The universal constant h is the famous Planck quantum of action (or the Planck constant). Observe that the action of typical processes in daily life has the magnitude of 1 Js. Therefore, the Planck constant is tiny. Nevertheless, the quantization of action has enormous consequences. For example, consider a mass point on the real line which moves periodically,

$$q(t) = \text{const} \cdot \sin(\omega t)$$

where t denotes time, and ω is called the angular frequency of the harmonic oscillator. Since the sine function has period 2π, the harmonic oscillator has the time period $T = 2\pi/\omega$. By definition, the frequency ν is the number of oscillations per second. Hence $T = 1/\nu$, and $\omega = 2\pi\nu$. If E denotes the energy of the harmonic oscillator, then the product ET is a typical action value for the oscillations of the harmonic oscillator. Therefore, according to Planck's quantization of action, it seems to be quite natural to postulate that $ET = nh$ for $n = 0, 1, 2, \ldots$ This yields Planck's quantization rule for the energy of the harmonic oscillator,

$$E = n\hbar\omega, \qquad n = 0, 1, 2, \ldots$$

from the year 1900. About 25 years later, the young physicist Werner Heisenberg (1901–1976) invented the full quantization procedure of classical mechanics. Using implicitly the commutation rule

$$\boxed{qp - pq = \mathrm{i}\hbar} \tag{1.2}$$

for the position q and the momentum p of a quantum particle, Heisenberg obtained the precise formula

$$\boxed{E = \Big(n + \frac{1}{2}\Big)\hbar\omega, \qquad n = 0, 1, 2, \ldots} \tag{1.3}$$

for the quantized energy levels of a harmonic oscillator. Heisenberg's quantum mechanics changed completely the paradigm of physics. In classical physics,

observables are real numbers. In Heisenberg's approach, observables are abstract quantities which obey certain commutation rules. More than fifty years before Heisenberg's discovery, the great Norwegian mathematician Sophus Lie (1842–1899) found out that commutation rules of type (1.2) play a fundamental role when trying to study continuous symmetry groups by means of linearization. In 1934, for this kind of algebraic structure, the term "Lie algebra" was coined by Hermann Weyl (1885–1955). Lie algebras and their generalizations to infinite dimensions, like the Virasoro algebra and supersymmetric algebras in string theory and conformal quantum field theory, are crucial for modern quantum physics. The Heisenberg formula (1.3) tells us that the ground state of each harmonic oscillator has the non-vanishing energy

$$E = \frac{\hbar\omega}{2}. \tag{1.4}$$

This fact causes tremendous difficulties in quantum field theories. Since a quantum field has an infinite number of degrees of freedom, the ground state has an infinite energy. There are tricks to cure the situation a little bit, but the infinite ground state energy is the deeper reason for the appearance of nasty divergent quantities in quantum field theory. Physicists have developed the ingenious method of renormalization in order to extract finite quantities that can be measured in physical experiments. Surprisingly enough, in quantum electrodynamics there is an extremely precise coincidence with the renormalized theoretical values and the values measured in particle accelerator experiments. No one understands this. Nowadays many physicists are convinced that this approach is not the final word. There must be a deeper theory behind. One promising candidate is string theory.

At the end of his life, Albert Einstein wrote the following about his first years.

> Between the ages of 12–16, I familiarized myself with the elements of mathematics. In doing so I had the good fortune of discovering books which were not too particular in their logical rigor.
> In 1896, at the age of 17, I entered the Swiss Institute of Technology (ETH) in Zurich. There I had excellent teachers, for example, Hurwitz (1859–1919) and Minkowski (1864–1909), so that I really could get a sound mathematical education. However, most of the time, I worked in the physical laboratory, fascinated by the direct contact with experience. The rest of the time I used, in the main, to study at home the works of Kirchhoff (1824–1887), Helmholtz (1821–1894), Hertz (1857–1894), and so on. The fact that I neglected mathematics to a certain extent had its cause not merely in my stronger interest in the natural sciences than in mathematics, but also in the following strange experience. I saw that mathematics was split up into numerous specialities, each of which could easily absorb the short life granted to us. Consequently, I saw myself in the position of Buridan's ass which was unable to decide upon any specific bundle of hay. This was obviously due to the fact that my intuition was not strong enough in the field of mathematics in order to differentiate clearly that

> which was fundamentally important, and that which is really basic, from the rest of the more or less dispensable erudition, and it was not clear to me as a student that the approach to a more profound knowledge of the basic principles of physics is tied up with the most intricate mathematical methods. This only dawned upon me gradually after years of independent scientific work. True enough, physics was also divided into separate fields. In this field, however, I soon learned to scent out that which was able to lead to fundamentals.[3]

After his studies, Einstein got a position at the Swiss patent office in Bern. In 1905 Einstein published four fundamental papers on the theory of special relativity, the equivalence between mass and energy, the Brownian motion, and the light particle hypothesis in Volume 17 of the journal *Annalen der Physik.*

> *The theory of special relativity completely changed our philosophy about space and time.*

According to Einstein, there is no absolute time, but time changes from observer to observer. This follows from the surprising fact that the velocity of light has the same value in each inertial system, which was established experimentally by Albert Michelson (1852–1931) in 1887. From his principle of relativity, Einstein deduced that a point particle of rest mass m_0 and momentum vector $\mathbf{p}$ has a positive energy E given by

$$\boxed{E^2 = m_0^2 c^4 + c^2 \mathbf{p}^2} \tag{1.5}$$

where c denotes the velocity of light in a vacuum. If the particle moves with sub-velocity of light, $\mathbf{x} = \mathbf{x}(t)$, than it has the mass

$$m = \frac{m_0}{\sqrt{1 - \dot{\mathbf{x}}(t)^2/c^2}}. \tag{1.6}$$

If the particle rests, then we get

$$E = m_0 c^2. \tag{1.7}$$

This magic energy formula governs the energy production in our sun by helium synthesis. Thus, our life depends crucially on this formula. Unfortunately, the atomic bomb is based on this formula, too.

Let us now discuss the historical background of Einstein's light particle hypothesis. Maxwell (1831–1879) conjectured in 1862 that light is an electromagnetic wave. In 1886 Heinrich Hertz established the existence of electromagnetic waves by a famous experiment carried out at Kiel University (Germany). When electromagnetic radiation is incident on the surface

[3] This is the English translation of Einstein's handwritten letter copied in the following book: Albert Einstein als Philosoph und Naturforscher (Albert Einstein as philosopher and scientist). Edited by P. Schilpp, Kohlhammer Verlag, Stuttgart (printed with permission).

of a metal, it is observed that electrons may be ejected. This phenomenon is called the photoelectric effect. This effect was first observed by Heinrich Hertz in 1887. Fifteen years later, Philipp Lenard (1862–1947) observed that the maximum kinetic energy of the electrons does not depend on the intensity of light. In order to explain the photoelectric effect, Einstein postulated in 1905 that electromagnetic waves are quantized. That is, light consists of light particles (or light quanta) which were coined photons in 1926 by the physical chemist Gilbert Lewis. According to Einstein, a light particle (photon) has the energy E given by Planck's quantum hypothesis,

$$E = h\nu. \tag{1.8}$$

Here, ν is the frequency of light, which is related to the wave length λ by the dispersion relation $\lambda\nu = c$. Hence $E = hc/\lambda$. This means that a blue photon has more energy than a red one. Since a photon moves with light speed, its rest mass must be zero. Thus, from (1.5) we obtain $|\mathbf{p}| = E/c$. If we introduce the angular frequency $\omega = 2\pi\nu$, then we obtain the final expression

$$\boxed{E = \hbar\omega, \qquad \mathbf{p} = \hbar\mathbf{k}, \qquad |\mathbf{k}| = \frac{\omega}{c}} \tag{1.9}$$

for the energy E and the momentum vector $\mathbf{p}$ of a photon. Here, the wave vector $\mathbf{k}$ of length $k = \omega/c$ is parallel to the vector $\mathbf{p}$. Nowadays we know that light particles are quanta, and that quantum particles are physical objects which possess a strange structure. Quanta combine features of both waves and particles. In the photoelectric effect, a photon hits an electron such that the electron leaves the metal. The energy of the electron is given by

$$E = \hbar\omega - W$$

where the so-called work function W depends on the binding energy of the electrons in the atoms of the metal. This energy formula suggests that for small angular frequencies ω no electrons can leave the metal, since there would be $E < 0$, a contradiction. In fact, this has been observed in experiments. Careful experiments were performed by Millikan (1868–1953) in 1916. He found out that a typical constant in his experiments coincided with the Planck constant, as predicted by Einstein. In 1921 Einstein was awarded the Nobel prize in physics for his services to theoretical physics, and especially for his discovery of the law of the photoelectric effect. As a curiosity let us mention, that Max Planck, while recommending Einstein enthusiastically for a membership in the Prussian Academy in Berlin, wrote the following:

> That sometimes, in his speculations, he went too far, such as, for example, in his hypothesis of the light quanta, should not be held too much against him.

1.2 Quantization in a Nutshell

In 1926 Born discovered the fundamental fact that quantum physics is intrinsically connected with random processes. Hence the mathematical theory of probability plays a crucial role in quantum physics. Already Maxwell (1831–1879) had emphasized:

The true logic of this world lies in probability theory.

Before discussing the randomness of quantum processes and the challenge of quantization, let us mention that Maxwell strongly influenced the physics of the 20th century. As we will show later on, Einstein's theory of special relativity follows from the invariance of the Maxwell equations in electromagnetism under Lorentz transformations. Moreover, the generalization of the Maxwell equations from the commutative gauge group $U(1)$ to the noncommutative gauge groups $SU(2)$ and $SU(3)$ leads to the Standard Model in particle physics. Finally, statistical physics can be traced back to Maxwell's statistical velocity distribution of molecules.

From the physical point of view, quantum mechanics and quantum field theory are described best by the Feynman approach via Feynman diagrams, transition amplitudes, Feynman propagators (Green's functions), and functional integrals. In order to make the reader familiar with the fascinating story of this approach, let us start with a quotation taken from Freeman Dyson's book *Disturbing the Universe*, Harper & Row, New York, 1979:[4]

> Dick Feynman (1918–1988) was a profoundly original scientist. He refused to take anybody's word for anything. This meant that he was forced to rediscover or reinvent for himself almost the whole physics. It took him five years of concentrated work to reinvent quantum mechanics. He said that he couldn't understand the official version of quantum mechanics that was taught in the textbooks and so he had to begin afresh from the beginning. This was a heroic enterprise. He worked harder during those years than anybody else I ever knew. At the end he had his version of quantum mechanics that he could understand...
>
> The calculations that I did for Hans Bethe,[5] using the orthodox method, took me several months of work and several hundred sheets of paper.
>
> > *Dick could get the same answer, calculating on a blackboard, in half an hour...*
>
> In orthodox physics, it can be said: Suppose an electron is in this state at a certain time, then you calculate what it will do next by solving the Schrödinger equation introduced by Schrödinger in 1926. Instead of this, Dick simply said:

[4] Reprinted by permission of Basic Books, a member of Perseus Books, L.L.C.

[5] Hans Bethe (1906–2005) was awarded the 1967 Nobel prize in physics for his contributions to nuclear reactions, especially his discoveries concerning the energy production in stars. See H. Bethe, R. Bacher, and M. Livingstone, Basic Bethe: Seminal Articles on Nuclear Physics 1936–37, American Institute of Physics, 1986.

The electron does whatever it likes.

A history of the electron is any possible path in space and time. The behavior of the electron is just the result of adding together all the histories according to some simple rules that Dick worked out. I had the enormous luck to be at Cornell in 1948 when the idea was newborn, and to be for a short time Dick's sounding board...

Dick distrusted my mathematics and I distrusted his intuition.

Dick fought against my scepticism, arguing that Einstein had failed because he stopped thinking in concrete physical images and became a manipulator of equations. I had to admit that was true. The discoveries of Einstein's earlier years were all based on direct physical intuition. Einstein's later unified theories failed because they were only sets of equations without physical meaning...

Nobody but Dick could use his theory. Without success I tried to understand him... At the beginning of September after vacations it was time to go back East. I got onto a Greyhound bus and travelled nonstop for three days and nights as far as Chicago. This time I had nobody to talk to. The roads were too bumpy for me to read, and so I sat and looked out of the window and gradually fell into a comfortable stupor. As we were droning across Nebraska on the third day, something suddenly happened. For two weeks I had not thought about physics, and now it came bursting into my consciousness like an explosion. Feynman's pictures and Schwinger's equations began sorting themselves out in my head with a clarity they had never had before. I had no pencil or paper, but everything was so clear I did not need to write it down.

Feynman and Schwinger were just looking at the same set of ideas from two different sides.

Putting their methods together, you would have a theory of quantum electrodynamics that combined the mathematical precision of Schwinger with the practical flexibility of Feynman...

During the rest of the day as we watched the sun go down over the prairie, I was mapping out in my head the shape of the paper I would write when I got to Princeton. The title of the paper would be *The radiation theories of Tomonaga, Schwinger, and Feynman.*[6]

For the convenience of the reader, in what follows let us summarize the prototypes of basic formulas for the passage from classical physics to quantum physics. These formulas are special cases of more general approaches due to

- Newton in 1666 (equation of motion),
- Euler in 1744 and Lagrange in 1762 (calculus of variations),
- Fourier in 1807 (Fourier method in the theory of partial differential equations, Fourier series, and Fourier integral),
- Poisson in 1811 (Poisson brackets and conservation laws),

[6] F. Dyson, Phys. Rev. **75** (1949), 486–502. Freeman Dyson (born 1923) is a member of the Institute for Advanced Study in Princeton (New Jersey, U.S.A.). He made fundamental contributions to quantum field theory, statistical physics, stability of matter, and number theory. This can be found in F. Dyson, Selected Papers, Amer. Math. Soc., Providence, Rhode Island, 1996.

- Hamilton in 1827 (Hamiltonian and canonical equations),
- Green in 1828 (the method of Green's function in electromagnetism),
- Lie in 1870 (continuous transformation groups (Lie groups) and infinitesimal transformation groups (Lie algebras)),
- Poincaré in 1892 (small divisors in celestial mechanics, and the renormalization of the quasi-periodic motion of planets by adding regularizing terms (also called counterterms) to the Poincaré–Lindsted series),
- Fredholm in 1900 (integral equations),
- Hilbert in 1904 (integral equations, and spectral theory for infinite-dimensional symmetric matrices),
- Emmy Noether in 1918 (symmetry, Lie groups, and conservation laws),
- Wiener in 1923 (Wiener integral for stochastic processes including Brownian motion for diffusion processes),
- von Neumann in 1928 (spectral theory for unbounded self-adjoint operators in Hilbert spaces, and calculus for operators),
- Stone in 1930 (unitary one-parameter groups and the general dynamics of quantum systems).

From the physical point of view, the following formulas are special cases of more general formulas due to Heisenberg in 1925, Born and Jordan in 1926, Schrödinger in 1926, Dirac in 1927, Feynman in 1942, Heisenberg in 1943, Dyson in 1949, Lippmann and Schwinger in 1950. In fact, we will study the following approaches to quantum mechanics:

- the 1925 Heisenberg *particle picture* via time-dependent operators as observables, and the Poisson–Lie operator equation of motion,
- the 1926 Schrödinger *wave picture* via time-dependent quantum states, and the Schrödinger partial differential equation of motion,
- the 1927 Dirac *interaction picture* which describes the motion under an interacting force as a perturbation of the interaction-free dynamics, and
- the 1942 Feynman picture based on a *statistics for possible classical motions* via the *Feynman path integral*, which generalizes the 1923 Wiener integral for the mathematical description of Einstein's Brownian motion in diffusion processes from the year 1905.

The fact that it is possible to describe quantum particles in an equivalent way by either Heisenberg's particle picture or Schrödinger's wave picture reflects a general duality principle in quantum physics:

> *Quantum particles are more general objects than classical particles and classical waves.*

This has been discovered in the history of physics step by step. Note that, for didactic reasons, in this section we will not follow the historical route, but we will present the material in a manner which is most convenient from the modern point of view.[7] Nowadays most physicists prefer the Feynman

[7] Remarks on the historical route of quantum mechanics can be found in Sect. 1.3 on page 60.

approach to quantum physics. In what follows we restrict ourselves to formal considerations.

Hints for quick reading. After reading Sects. 1.2.1 through 1.2.4, the reader may pass to Sects. 15.1 through 15.5 on the operator approach to quantum field theory. A rigorous approach to the basic ideas in quantum field theory in terms of a finite-dimensional Hilbert space setting can be found in Chap. 7. The true mathematical difficulties in quantum field theory are related to the infinite-dimensional setting. However, rigorous finite-dimensional models help to understand the mathematical substance of the magic formulas used by physicists in quantum field theory in a formal way. These magic formulas are due to Dyson, Feynman, Schwinger, Gell-Mann and Low, Faddeev and Popov. They can be found in Chaps. 14 through 16. The reader who wants to become familiar, as quickly as possible, with applications of quantum field theory to concrete physical processes in quantum electrodynamics should pass to Volume II.

1.2.1 Basic Formulas

The classical principle of critical action. For the mathematical description of physics, it is crucial that the fundamental processes in nature are governed by an optimality principle called the principle of least action. In fact, the action is not always minimal in nature, but sometimes the action is only critical (also called stationary). Therefore, we have to speak about the principle of critical action. In the history of physics, the role of variational principles was underlined by Fermat (1601–1665), Maupertius (1698–1759), Euler (1707–1783), Lagrange (1736–1813), Gauss (1777–1855), Hamilton (1788–1856), and Jacobi (1804–1851). As the simplest example for the principle of critical action, let us start with the following variational problem

$$\boxed{\int_{t_0}^{t_1} L(q(t), \dot{q}(t))\, dt = \text{critical!}} \tag{1.10}$$

for the motion, $q = q(t)$, of a classical particle with mass m on the real line (Fig. 1.1). This is called the principle of critical action. Here, $q(t)$ denotes the position of the classical particle at time t. Following Newton, the dot, $\dot{q}$, denotes the derivative with respect to time. We have to add the boundary condition

$$q(t_0) = x_0, \qquad q(t_1) = x_1 \tag{1.11}$$

for given initial point x_0 at the initial time t_0, and given final point x_1 at the final time t_1. The function $L = L(q, \dot{q})$ is called Lagrangian; it has the physical dimension of energy. More important than energy is the action S of the classical motion, $q = q(t)$, during the time interval $[t_0, t_1]$. Explicitly,

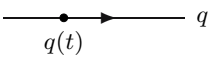


Fig. 1.1. Motion on the real line

$$S[q] := \int_{t_0}^{t_1} L(q(t), \dot{q}(t))\, dt.$$

Here, the action S has the physical dimension of *energy* $\times$ *time*. Each smooth solution of the variational problem (1.10) satisfies the following Euler–Lagrange equation[8]

$$\boxed{\frac{d}{dt} L_{\dot{q}}(q(t), \dot{q}(t)) = L_q(q(t), \dot{q}(t))} \tag{1.12}$$

for all times $t \in [t_0, t_1]$. Since the Lagrangian $L = L(q, \dot{q})$ does not depend explicitly on time t, it is invariant under time translations $t \mapsto t + \text{const}$. By the Noether theorem, each solution, $q = q(t)$, of the Euler–Lagrange equation satisfies the conservation law

$$\frac{dE(t)}{dt} = 0 \qquad \text{for all} \quad t \in [t_0, t_1] \tag{1.13}$$

with the momentum function $p(t) := L_{\dot{q}}(q(t), \dot{q}(t))$, and the energy function

$$E(t) := p(t)\dot{q}(t) - L(q(t), \dot{q}(t)).$$

The relation (1.13) tells us that the energy $E(t)$ does not depend on time along the classical trajectory. This is called conservation of energy.[9] Generally, from the mathematical point of view, the fundamental notion of energy is intimately related to symmetry properties of physical systems, namely, the invariance under time translations.

The Feynman picture of quantum mechanics and the Feynman path integral. We want to consider the motion of a quantum particle on the real line. This motion is described by a complex-valued function $\psi = \psi(x, t)$ whose physical meaning will be explained below. From the physical point of view, the best interpretation of the passage from the classical motion of the particle to the corresponding quantum motion can be obtained by the famous Feynman formula

[8] The symbol $L_{\dot{q}}$ denotes the partial derivative of the function $L = L(q, \dot{q})$ with respect to the variable $\dot{q}$.

[9] Without referring to the general Noether theorem, energy conservation follows directly from the Euler–Lagrange equation, $\dot{p} = L_q$. In fact, differentiation with respect to time t yields

$$\dot{E} = \dot{p}\dot{q} + p\ddot{q} - L_{\dot{q}}\ddot{q} - L_q\dot{q} = (\dot{p} - L_q)\dot{q} = 0.$$

$$\boxed{\mathcal{P}(x_1, t_1; x_0, t_0) = \int \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}q.} \tag{1.14}$$

Formally, we sum over all possible classical trajectories $q = q(t)$ which satisfy the boundary condition (1.11). The integral from (1.14) is called a Feynman path integral (or functional integral). It represents a statistic over the possible trajectories of the classical particle. The statistical weight $\mathrm{e}^{\mathrm{i}S[q]/\hbar}$ for each trajectory depends on the classical action $S[q]$ of the trajectory. Since the quantity $S[q]$ has the physical dimension of *energy* $\times$ *time*, we have to divide the action $S[q]$ by a constant $\hbar$ of the same dimension in order to get a dimensionless argument of the exponential function.

This way, Planck's constant of action, $\hbar$, appears in a natural way.

It was noticed by Feynman that, because of the smallness of the Planck constant $\hbar$, a formal application of Kelvin's method of stationary phase in optics tells us that, naturally enough, the main contribution to the integral (1.14) comes from that classical trajectory which corresponds to the solution of the principle of critical action (1.10).

Quantum motion is obtained from classical motion by adding random quantum fluctuations.

In particular, quantum mechanics passes over to classical mechanics if

$$\hbar \to 0.$$

This limit corresponds to the passage from wave optics to geometric optics if the wave length of light goes to zero, $\lambda \to 0$.

In terms of mathematics, quantization of a classical theory corresponds to a deformation of the classical theory which depends on the Planck parameter $\hbar$.

It remains to discuss the physical meaning of the Feynman propagator kernel $\mathcal{P}$ and of the function ψ. According to Feynman's 1942 dissertation at Princeton University, the dynamics of the function ψ is governed by the crucial formula

$$\boxed{\psi(x_1, t_1) = \int_{-\infty}^{\infty} \mathcal{P}(x_1, t_1; x_0, t_0)\psi(x_0, t_0)\, dx_0} \tag{1.15}$$

for all positions x_1 and all times $t_1 \geq t_0$. We write this briefly as

$$\psi(t_1) = P(t_1, t_0)\psi(t_0). \tag{1.16}$$

The operator $P(t_1, t_0)$ is called the Feynman propagator.

The simplest example corresponds to the motion of a free particle on the real line. In this case, the Lagrangian reads as

$$L(q,\dot{q}) := \frac{m\dot{q}^2}{2}.$$

The corresponding Euler–Lagrange equation, $m\ddot{q} = 0$, has the general solution

$$q(t) = x_0 + v(t - t_0), \qquad t \in \mathbb{R}.$$

This describes the motion of a point particle on the real line with the constant velocity v. The corresponding Feynman propagator kernel is given by

$$\mathcal{P}(x_1,t_1;x_0,t_0) = \sqrt{\frac{m}{2\pi \mathrm{i}\hbar(t_1-t_0)}} \cdot \mathrm{e}^{-\frac{m(x_1-x_0)^2}{2\mathrm{i}\hbar(t_1-t_0)}}.$$

Here, we choose the value $\sqrt{\mathrm{i}} := \mathrm{e}^{\mathrm{i}\pi/4}$ for the square root of the imaginary unit. Using the replacement $\mathrm{i}t \mapsto t$, the Feynman kernel, $\mathcal{P}$, passes over to Fourier's heat kernel for both the propagation of heat and the diffusion of particles (Brownian motion) on the real line. This will be studied in Sect. 11.1.3 on page 586.

Born's interpretation of the Schrödinger ψ-function. In 1926 Schrödinger formulated his famous partial differential equation for some wave function ψ which will be considered below. Surprisingly enough, Schrödinger was very successful in computing the quantized energy levels of the hydrogen atom, but he did not know the physical meaning of the function ψ. This was discovered by Born a few months later by studying scattering processes for electrons.

> *According to Born, the value $|\psi(x,t)|^2$ plays the role of a probability density.*

This changes physics dramatically. In contrast to classical physics, quantum processes are random processes. More precisely, we have to distinguish between the following two cases.

(i) Case 1: One single particle. Suppose that $\int_{-\infty}^{\infty} |\psi(x,t)|^2 dx < \infty$. Then the function ψ describes a single particle on the real line. The quotient

$$\frac{\int_a^b |\psi(x,t)|^2 \, dx}{\int_{-\infty}^{\infty} |\psi(x,t)|^2 \, dx}$$

is equal to the probability of finding the particle in the interval $[a,b]$ of the real line at time t. In addition, let us introduce the momentum operator P and the position operator Q,

$$(P\psi)(x) := -\mathrm{i}\hbar \frac{d}{dx}\psi(x), \qquad (Q\psi)(x) := x\psi(x), \qquad x \in \mathbb{R},$$

along with the inner product

$$\langle \chi | \psi \rangle := \int_{-\infty}^{\infty} \chi(x)^\dagger \psi(x) \, dx.$$

For a classical particle on the real line, position q and momentum $p = mv$ (mass $\times$ velocity) can be measured with arbitrarily high precision at the same time. This is not true anymore for a quantum particle on the real line. In fact, in the quantum state $\psi = \psi(x)$ it is only possible to measure the mean position $\overline{q}$ and the fluctuation Δq of the position. Explicitly,

$$\overline{q} = \frac{\langle \psi | Q\psi \rangle}{\langle \psi | \psi \rangle} = \frac{\int_{-\infty}^{\infty} \psi(x)^\dagger x \psi(x)\, dx}{\int_{-\infty}^{\infty} \psi(x)^\dagger \psi(x)\, dx}$$

and

$$(\Delta q)^2 = \frac{\langle \psi | (Q - \overline{q}\,)^2 \psi \rangle}{\langle \psi | \psi \rangle} = \frac{\int_{-\infty}^{\infty} \psi(x)^\dagger (x - \overline{q})^2 \psi(x)\, dx}{\int_{-\infty}^{\infty} \psi(x)^\dagger \psi(x)\, dx}.$$

Similarly, for the measured mean momentum $\overline{p}$ and its fluctuation Δp, we get

$$\overline{p} = \frac{\langle \psi | P\psi \rangle}{\langle \psi | \psi \rangle}, \qquad (\Delta p)^2 = \frac{\langle \psi | (P - \overline{p}\,)^2 \psi \rangle}{\langle \psi | \psi \rangle}.$$

If the function $\psi = \psi(x,t)$ also depends on time t, then so do the measured values. In 1927 Heisenberg showed that there holds the following fundamental inequality for the fluctuations:

$$\boxed{\Delta p \Delta q \geq \frac{\hbar}{2}.} \tag{1.17}$$

This famous uncertainty inequality tells us that, in contrast to classical particles, it is not possible to measure precisely position and momentum of a quantum particle at the same time. If the measurement of the position of a quantum particle is fairly precise (i.e., Δq is small), then the velocity of the particle is quite uncertain (i.e., Δp is large). Conversely, if the velocity is known fairly precisely (i.e., Δp is small), then the position of the quantum particle is quite uncertain (i.e., Δq is large).

(ii) Case 2: Homogeneous stream of particles. Suppose that

$$\int_{-\infty}^{\infty} |\psi(x,t)|^2 dx = \infty.$$

Then the function ψ describes a flow of identical particles on the real line which has the current density vector

$$\mathbf{J}(x,t) := \varrho(x,t)\mathbf{v}(x,t)$$

with the particle density $\varrho(x,t) := |\psi(x,t)|^2$ and the velocity vector $\mathbf{v}(x,t)$ at the point x at time t. Explicitly,

$$\mathbf{J}(x,t) = \Re\left(\psi(x,t)^\dagger \cdot \frac{P}{m}\,\psi(x,t)\right)\mathbf{e}.$$

The unit vector $\mathbf{e}$ points from left to right on the real line. For example, choose the function

$$\psi(x,t) := \sqrt{\varrho_0} \cdot \mathrm{e}^{\mathrm{i}(px - E_p t)/\hbar}$$

with fixed real number p and $E_p := p^2/2m$. This function describes a homogeneous stream of particles with momentum p, energy E_p, and particle density $\varrho(x,t) = |\psi(x,t)|^2 = \varrho_0$. The particle momentum p is an eigenvalue of the momentum operator P,

$$P\psi = p\psi,$$

and the particle energy E_p is an eigenvalue of the energy operator,

$$\mathrm{i}\hbar \frac{\partial}{\partial t}\psi = E_p \psi.$$

Moreover, we have $\mathbf{J}(x,t) = \varrho_0 \mathbf{v}$ along with the velocity vector $\mathbf{v} = p\mathbf{e}/m$.

Observe that the measured values of a single particle are based on the inner product $\langle \psi | \varphi \rangle$. This is the key to John von Neumann's Hilbert space approach to quantum mechanics from the late 1920s. However, the Hilbert space setting is not sufficient, since states ψ with $\langle \psi | \psi \rangle = \infty$ appear which do not lie in the Hilbert space under consideration. To include such states, one has to use Gelfand's theory of rigged Hilbert spaces from the 1950s which is based on the notion of generalized functions (distributions). This will be studied in Sect. 12.2 on page 675.

The Schrödinger wave picture for quantum mechanics on the real line. Consider the special case where the Lagrangian is given by

$$L(q,\dot{q}) := \frac{m\dot{q}^2}{2} - \kappa U(q).$$

Here, the Euler–Lagrange equation of motion reads as

$$m\ddot{q} = -\kappa U'(q).$$

This is the Newtonian equation of motion with the force $F(q) := -\kappa U'(q)$. For the momentum, $p(t) = m\dot{q}(t)$. The function U is called the potential, and the real number κ is called coupling constant.[10] For the motion $q = q(t)$ of the classical particle on the real line, we have conservation of energy, i.e., there exists a constant E such that

$$\boxed{E = \frac{p(t)^2}{2m} + \kappa U(q(t))} \tag{1.18}$$

[10] In many cases, the coupling constant κ is small. Then, it is possible to apply the methods of perturbation theory. This is of fundamental importance for quantum field theory.

for all times t. Here, E is called the energy of the motion. Using now the elegant Schrödinger quantization rule

$$E \;\Rightarrow\; \mathrm{i}\hbar\frac{\partial}{\partial t}, \qquad p \;\Rightarrow\; P,$$

the classical energy equation (1.18) passes over to the following famous Schrödinger equation

$$\boxed{\mathrm{i}\hbar\frac{\partial}{\partial t}\psi = \left(\frac{P^2}{2m} + \kappa U\right)\psi.} \tag{1.19}$$

Explicitly,

$$\mathrm{i}\hbar\dot{\psi}(x,t) = -\frac{\hbar^2}{2m}\psi_{xx}(x,t) + \kappa U(x)\psi(x,t). \tag{1.20}$$

Recall that the dot, $\dot{\psi}$, denotes the partial derivative with respect to time. In the history of mathematics and physics, it was gradually discovered that the solutions of partial differential equations can be represented by integral formulas with appropriate kernels which are called Green's functions. In particular, the corresponding solution formula for the initial-value problem to the Schrödinger equation (1.20) reads as

$$\boxed{\psi(x,t) = \int_{-\infty}^{\infty} \mathcal{G}(x,t;y,t_0)\psi(y,t_0)dy} \tag{1.21}$$

for all positions $x \in \mathbb{R}$ and all times $t \geq t_0$. Comparing this with Feynman's formula (1.15), we see that

$$\mathcal{G}(x,t;y,t_0) \equiv \mathcal{P}(x,t;y,t_0).$$

Consequently, the Feynman propagator kernel is nothing than the Green's function to the initial-value problem for the Schrödinger equation (1.20). In fact, if we know the initial state $\psi(x,t_0)$ of the quantum particle at the initial time t_0, then formula (1.21) determines the state $\psi = \psi(x,t)$ for all times $t \geq t_0$ in the future. This tells us that

> *The Feynman propagator kernel knows all about the motion of the quantum particle on the real line.*

By physical considerations, Feynman discovered that the Feynman kernel can be represented by a path integral of the form (1.14). This discovery was crucial for the further development of quantum field theory.

Schrödinger's method for computing quantized energies by solving eigenvalue problems. Motivated by the classical Fourier method, Schrödinger made the separation ansatz

$$\psi(x,t) = \varphi(x)\mathrm{e}^{-\mathrm{i}Et/\hbar}.$$

Then, the time-dependent Schrödinger equation (1.19) passes over to the stationary Schrödinger equation

$$E\varphi(x) = -\frac{\hbar^2}{2m}\,\varphi''(x) + \kappa\, U(x)\varphi(x), \qquad x \in \mathbb{R}.$$

This is an eigenvalue problem for the unknown energy E. Using this method, Schrödinger computed the quantized energies of quantum particles. Interestingly enough, Schrödinger did not know the precise physical meaning of the complex wave function ψ. This problem was solved by Born in 1926; Born discovered the statistical interpretation of $|\psi(x,t)|^2$ discussed above. Schrödinger and Born were awarded the Nobel prize in physics in 1933 and 1954, respectively.

Von Neumann's solution of the Schrödinger equation. Let us introduce the free Hamiltonian

$$H_0 := \frac{P^2}{2m} = -\frac{\hbar^2}{2m}\,\frac{\partial^2}{\partial x^2}$$

along with the Hamiltonian

$$H := H_0 + \kappa U.$$

Then, the Schrödinger equation (1.19) can be written as an operator equation of the form

$$\boxed{\mathrm{i}\hbar\dot{\psi}(t) = H\psi(t)} \tag{1.22}$$

for all times $t \in \mathbb{R}$. If we consider equation (1.22) as a classical ordinary differential equation, then the solution reads as

$$\boxed{\psi(t) = \mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}\,\psi(t_0) \qquad \text{for all} \quad t \in \mathbb{R}.} \tag{1.23}$$

Let us now discuss the mathematical meaning of the operator $\mathrm{e}^{-\mathrm{i}tH/\hbar}$.

Von Neumann's operator calculus. In the late 1920s, von Neumann developed a calculus for unbounded self-adjoint operators on Hilbert spaces which gives the exponential function for operators a precise meaning under appropriate assumptions on the Hamiltonian H.[11] In the sense of von Neumann's operator calculus, formula (1.23) solves the Schrödinger equation (1.19), and it describes completely the dynamics of the quantum particle on the real line. Comparing this with equation (1.16), $\psi(t) = P(t,t_0)\psi(t_0)$, we see that the Feynman propagator is given by

[11] The operator H has to be a self-adjoint operator on the Hilbert space $L_2(\mathbb{R})$. See Zeidler (1995), Vol. 1.

$$P(t,t_0) = \mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar} \qquad \text{for all} \quad t, t_0 \in \mathbb{R}.$$

Generalizing Hilbert's spectral theory from 1904 for bounded symmetric operators to unbounded self-adjoint operators in 1928, von Neumann justified the formula

$$\boxed{H = \int_{\lambda=-\infty}^{\infty} \lambda dE_\lambda}$$

where the operators $E_\lambda : X \to X$ are orthogonal projection operators on the Hilbert space $X = L_2(\mathbb{R})$. Moreover, $f(H) = \int_{\lambda=-\infty}^{\infty} f(\lambda) dE_\lambda$. In particular,

$$\boxed{\mathrm{e}^{-\mathrm{i}tH/\hbar} = \int_{\lambda=-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}t\lambda/\hbar} dE_\lambda.}$$

The family $\{E_\lambda\}$ of operators E_λ with $\lambda \in \mathbb{R}$ is called the spectral family of the Hamiltonian H. In order to discuss the physical meaning of the spectral family, choose a fixed particle state on the real line, $\psi \in X$, normalized by

$$\langle \psi | \psi \rangle = \int_{-\infty}^{\infty} |\psi(x,t)|^2 dx = 1.$$

Suppose we are measuring the energy E of the particle in the state ψ. Then, the probability P of finding the measured value E in the interval J is given by the integral

$$\boxed{\mathsf{P} := \int_J dF(\lambda).}$$

Here, the function $F(\lambda_0) := \int_{]-\infty,\lambda_0[} d\sigma(\lambda)$ with $\sigma(\lambda) := \langle \psi | E_\lambda \psi \rangle$ represents the probability distribution of the energy of the quantum particle on the real line in the given state ψ. By the theory of probability, for the mean value $\overline{E}$ and the fluctuation ΔE of the measured energy, we get

$$\overline{E} = \int_{-\infty}^{\infty} \lambda dF(\lambda), \qquad (\Delta E)^2 = \int_{-\infty}^{\infty} (\lambda - \overline{E})^2 dF(\lambda).$$

In particular, if the function F is smooth, then $dF(\lambda)$ can be replaced by $F'(\lambda)d\lambda$ where $F'(\lambda) = \sigma(\lambda)$.

According to von Neumann, the spectral family can be constructed for all self-adjoint operators on Hilbert spaces. Such operators represent observables in quantum mechanics. Therefore, the theory of spectral families allows us to describe the random measurements of observables in quantum mechanics.[12]

Heisenberg's S-matrix for scattering processes. The most important processes for elementary particles are scattering processes in particle

[12] As an introduction to functional analysis and its applications to mathematical physics, we recommend Zeidler (1995), Vols. 1, 2.

accelerators. Therefore, physicists are mainly interested in computing scattering processes. For this, the main tool is the S-matrix which was introduced by Wheeler in 1937 and Heisenberg in 1943. The further development was strongly influenced by Heisenberg's paper from 1943.

> *The S-matrix is closely related to physical quantities which can be measured in experiments.*

This underlines the importance of the S-matrix. Let us discuss this in terms of scattering processes on the real line. To this end, set

$$\psi_{\mathrm{in}}(x,t) := \sqrt{\varrho_{\mathrm{in}}} \cdot \mathrm{e}^{\mathrm{i}(p_{\mathrm{in}}x - E_{\mathrm{in}}t)/\hbar}$$

and

$$\psi_{\mathrm{out}}(x,t) := \sqrt{\varrho_{\mathrm{out}}} \cdot \mathrm{e}^{\mathrm{i}(p_{\mathrm{out}}x - E_{\mathrm{out}}t)/\hbar}.$$

We regard ψ_{in} as a stream of incoming free particles with

- momentum vector $\mathbf{p}_{\mathrm{in}} = p_{\mathrm{in}}\mathbf{e}$ of the incoming particles,
- velocity vector $\mathbf{v}_{\mathrm{in}} = \mathbf{p}_{\mathrm{in}}/m$ of the incoming particles,
- energy $E_{\mathrm{in}} = p_{\mathrm{in}}^2/2m$ of the incoming particles, and
- particle density ϱ_{in} of the incoming particles.

The corresponding quantities of the outgoing particles are defined similarly. We assume that the potential U is concentrated in a neighborhood of the origin. The incoming free particles are then scattered at the potential κU. This means that some particles are reflected at the potential wall and some of them penetrate the potential wall. We are interested in those particles which penetrate the potential wall. This yields a stream of outgoing particles described by the function ψ_{out}. Let us introduce the Feynman transition amplitude

$$\boxed{\mathsf{S}(q,t_1;p,t_0) := \langle \psi_{\mathrm{out}} | P(t_1,t_0)\psi_{\mathrm{in}} \rangle.}$$

In this connection, observe that the Feynman propagator $P(t_1,t_0)$ sends the incoming state ψ_{in} at time t_0 to the state

$$\psi(t_1) = P(t_1,t_0)\psi(t_0)$$

at time t_1, and we compare the actual state $\psi(t_1)$ with the possible outgoing state $\psi_{\mathrm{out}}(t_1)$ at time t_1 by computing the inner product $\langle \psi_{\mathrm{out}}(t_1) | \psi(t_1) \rangle$. In terms of the Feynman propagator kernel $\mathcal{P}$, the transition amplitude $\mathsf{S}(q,t_1;p,t_0)$ is equal to

$$\int_{-\infty}^{\infty} \psi_{\mathrm{out}}(x_1,t_1)^{\dagger} \left(\int_{-\infty}^{\infty} \mathcal{P}(x_1,t_1;x_0,t_0)\psi_{\mathrm{in}}(x_0,t_0)\, dx_0 \right) dx_1.$$

Carrying out a physical experiment, we are interested in the transition probability from the incoming particle stream at time t_0 to the outgoing stream

$\psi_{\text{out}}(t_1)$ at time t_1. It turns out that this transition probability[13] is given by the expression

$$|\mathsf{S}(q,t_1;p,t_0)|^2.$$

To free ourselves from the arbitrary choice of the initial time t_0 and the final time t_1, we pass over to the formal limit $t_0 \to -\infty$ and $t_1 \to +\infty$. This way, we get

$$\boxed{\mathsf{S}(q,+\infty;p,-\infty) := \lim_{t_1\to+\infty}\lim_{t_0\to-\infty}\mathsf{S}(q,t_1;p,t_0).}$$

Physicists call this the S-matrix element for particles with incoming momentum p and outgoing momentum q. The set of all these S-matrix elements forms the S-matrix S. The corresponding transition probability for the particle stream is then given by the key formula

$$\boxed{|\mathsf{S}(q,+\infty;p,-\infty)|^2.}$$

It turns out that these S-matrix elements vanish if $q \neq p$, by energy conservation.

The Lippmann–Schwinger integral equation. It is possible to compute the S-matrix elements by solving the following Lippmann–Schwinger integral equation

$$\boxed{\varphi(x) = \varphi_0(x) - \kappa\int_{-\infty}^{\infty}\mathcal{G}(x,y)U(y)\varphi(y)dy} \tag{1.24}$$

with $\varphi_0(x) := \mathrm{e}^{\mathrm{i}px/\hbar}$ and the kernel

$$\mathcal{G}(x,y) := \mathrm{i}m\,\frac{\mathrm{e}^{\mathrm{i}p|x-y|/\hbar}}{\hbar p}.$$

The function

$$\psi_{\text{in}}(x,t) := \mathrm{e}^{-\mathrm{i}E_p(t-t_0)/\hbar}\varphi_0(x)$$

describes the incoming particle stream with momentum p. If the function $\varphi = \varphi(x)$ is a solution of (1.24), then

$$\psi(x,t) := \mathrm{e}^{-\mathrm{i}E_p(t-t_0)/\hbar}\varphi(x)$$

with $E_p := p^2/2m$ is a solution of the Schrödinger equation (1.19) which describes the scattering of the incoming particle stream ψ_{in}. Finally, we compare the scattered particle stream with the outgoing particle stream

$$\psi_{\text{out}}(x,t) := \mathrm{e}^{-\mathrm{i}E_p(t-t_1)/\hbar}\varphi_0(x)$$

of momentum p. This implies

[13] In mathematics, transition probability is called conditional probability.

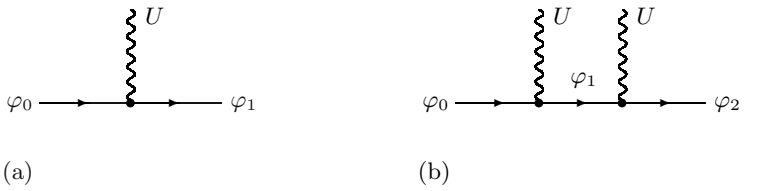


Fig. 1.2. Feynman diagrams

$$\langle\psi_{\text{out}}|\psi\rangle = \int_{-\infty}^{\infty} \psi_{\text{out}}(x,t)^\dagger \psi(x,t)dx$$
$$= \mathrm{e}^{-\mathrm{i}E_p(t_1-t_0)} \int_{-\infty}^{\infty} \varphi_0(x)^\dagger \varphi(x)dx = \mathrm{e}^{-\mathrm{i}E_p(t_1-t_0)}\langle\varphi_0|\varphi\rangle.$$

Letting $t_0 \to -\infty$ and $t_1 \to +\infty$, we obtain the formal limit

$$\boxed{|\mathsf{S}(p,+\infty;p,-\infty)|^2 = |\langle\varphi_0|\varphi\rangle|^2}$$

which depends on the solution φ of the time-independent Lippmann–Schwinger integral equation.[14] The point is that the solution φ can be computed approximately by using the following iterative method

$$\varphi_{n+1}(x) = \varphi_0(x) - \kappa \int_{-\infty}^{\infty} \mathcal{G}(x,y)U(y)\varphi_n(y)dy, \qquad n = 0,1,2,\ldots$$

The method of Feynman diagrams. The basic idea of Feynman diagrams is to represent graphically the approximations $\varphi_1, \varphi_2, \ldots$. This technique is widely used in elementary particle physics; it helps to simplify the computation of scattering processes, and gives physical insight. In the present case, for small coupling constant κ, the first approximation

$$\boxed{\varphi_1(x) = \varphi_0(x) + \kappa \int_{-\infty}^{\infty} \mathcal{G}(x,y)U(y)\varphi_0(y)dy}$$

is called the Born approximation; it was used by Born in 1926 in order to compute scattering processes for electrons. The Feynman diagram for φ_1 is pictured in Fig. 1.2(a). Intuitively, the interaction between the "particle" φ_0 and the potential U yields the "particle" φ_1. The second approximation φ_2 is given by

$$\varphi_2(x) = \varphi_0(x) + \kappa \int_{-\infty}^{\infty} \mathcal{G}(x,y)U(y)\varphi_1(y)dy.$$

The corresponding Feynman diagram is pictured in Fig. 1.2(b). Intuitively, the interaction between the "particle" φ_1 and the potential U yields the "particle" φ_2. Observe that

[14] This approach is called stationary scattering theory. See Zeidler (1995), Vol. 1, Sect. 5.24.5.

As a rule, each iterative method in mathematics and physics can be represented graphically by Feynman diagrams.

The Heisenberg–Born–Jordan commutation relation. For the momentum operator P and the position operator Q,

$$QP\psi - PQ\psi = -\mathrm{i}\hbar\big(x\psi'(x) - (x\psi(x))'\big) = \mathrm{i}\hbar\psi(x).$$

Letting $[Q,P]_- := QP - PQ$, we obtain

$$[Q,P]_- = \mathrm{i}\hbar I$$

where I denotes the identity operator. This is the commutation rule (1.2) which appeared at the birth of modern quantum mechanics in 1925. The interesting history of the commutation relation will be discussed in Sect. 1.3 on page 60.

The Heisenberg particle picture for quantum mechanics on the real line. It was discovered in the 1920s that completely different approaches to quantum mechanics are in fact equivalent. Let us discuss the equivalence between the Schrödinger picture and the Heisenberg picture which was invented before the Schrödinger equation. In the Schrödinger picture, the dynamics of the time-dependent wave function $\psi = \psi(t)$ is governed by the equation

$$\psi(t) = \mathrm{e}^{-\mathrm{i}tH/\hbar}\psi(0).$$

In the Heisenberg picture, we introduce the time-dependent momentum operator

$$P(t) := \mathrm{e}^{\mathrm{i}tH/\hbar}P\mathrm{e}^{-\mathrm{i}tH/\hbar},$$

the time-dependent position operator

$$Q(t) := \mathrm{e}^{\mathrm{i}tH/\hbar}Q\mathrm{e}^{-\mathrm{i}tH/\hbar},$$

and the time-independent state $\psi(0)$. For the measured mean-value of momentum in the Schrödinger picture, we get

$$\overline{p} = \frac{\langle\psi(t)|P\psi(t)\rangle}{\langle\psi(t)|\psi(t)\rangle}.$$

Moreover, we get

$$\overline{p} = \frac{\langle\psi(0)|P(t)\psi(0)\rangle}{\langle\psi(0)|\psi(0)\rangle}.$$

in the Heisenberg picture.[15] Differentiation with respect to time t yields

[15] Note that $\langle A\chi|\psi\rangle = \langle\chi|A^\dagger\psi\rangle$. Since $H^\dagger = H$ and $(\mathrm{e}^{-\mathrm{i}tH/\hbar})^\dagger = \mathrm{e}^{\mathrm{i}tH/\hbar}$, we obtain that the inner product $\langle\psi(t)|P\psi(t)\rangle$ is equal to

$$\langle\mathrm{e}^{-\mathrm{i}tH/\hbar}\psi(0)|P\mathrm{e}^{-\mathrm{i}tH/\hbar}\psi(0)\rangle = \langle\psi(0)|\mathrm{e}^{\mathrm{i}tH/\hbar}P\mathrm{e}^{-\mathrm{i}tH/\hbar}\psi(0)\rangle = \langle\psi(0)|P(t)\psi(0)\rangle.$$

As an introduction to the theory of Hilbert spaces, we recommend Zeidler (1995), Vol. 1.

$$\mathrm{i}\hbar\dot{P}(t) = P(t)H - HP(t).$$

This way, we obtain the equations of motion in the Heisenberg picture,

$$\boxed{\mathrm{i}\hbar\dot{P}(t) = [P(t), H]_-, \qquad \mathrm{i}\hbar\dot{Q}(t) = [Q(t), H]_-,}$$

along with the commutation relations $[Q(t), P(t)]_- = \mathrm{i}\hbar I$ for all times $t \in \mathbb{R}$.

The Dirac interaction picture for quantum mechanics on the real line. In order to study the dynamics of perturbed systems, Dirac introduced the so-called interaction picture which is crucial for quantum field theory. Let us discuss the basic idea of this approach. We start with the Schrödinger equation

$$\boxed{\mathrm{i}\hbar\dot{\psi}(t) = (H_0 + \kappa U(t))\psi(t).} \tag{1.25}$$

Here, the potential $U = U(t)$ is allowed to depend on time t. The following arguments are well-known for classical ordinary differential equations. The point is that we will apply formally these arguments to operator differential equations, too. The first trick is to introduce the new function

$$\Psi(t) := \mathrm{e}^{\mathrm{i}(t-t_0)H_0/\hbar}\,\psi(t).$$

Then $\Psi(t_0) = \psi(t_0)$. From the Schrödinger equation (1.25), we get the new differential equation

$$\boxed{\mathrm{i}\hbar\dot{\Psi}(t) = \kappa \mathsf{U}(t)\Psi(t), \qquad t \in \mathbb{R}} \tag{1.26}$$

with the transformed potential[16]

$$\mathsf{U}(t) := \mathrm{e}^{\mathrm{i}(t-t_0)H_0/\hbar} U(t)\, \mathrm{e}^{-\mathrm{i}(t-t_0)H_0/\hbar}.$$

Differentiation of the function

$$F(t) = F(t_0) + \int_{t_0}^{t} f(\tau)d\tau$$

with respect to time t yields $\dot{F}(t) = f(t)$. Therefore, the differential equation (1.26) is equivalent to the Volterra integral equation

$$\Psi(t) = \Psi_0 + \frac{\kappa}{\mathrm{i}\hbar}\int_{t_0}^{t} \mathsf{U}(\tau)\Psi(\tau)d\tau \tag{1.27}$$

[16] In fact, for simplifying notation, we set $\hbar := 1$ and $t_0 := 0$. Then

$$\mathrm{i}\dot{\Psi}(t) = -H_0\mathrm{e}^{\mathrm{i}tH_0}\psi(t) + \mathrm{e}^{\mathrm{i}tH_0}\mathrm{i}\dot{\psi}(t).$$

Using $\mathrm{i}\dot{\psi} = (H_0 + \kappa U)\psi$, we obtain the claim (1.26).

with $\Psi_0 := \Psi(t_0) = \psi(t_0)$. This integral equation can be solved by means of the following iterative method

$$\Psi_{n+1}(t) = \Psi_0 + \frac{\kappa}{\mathrm{i}\hbar}\int_{t_0}^{t} \mathsf{U}(\tau)\Psi_n(\tau)d\tau, \qquad n = 0, 1, 2, \ldots$$

This yields the solution

$$\Psi(t) = \Psi_0 + \sum_{n=1}^{\infty} \frac{\kappa^n}{(\mathrm{i}\hbar)^n} \int \mathsf{U}(t_1)\cdots\mathsf{U}(t_n)\Psi_0 \tag{1.28}$$

where we use the convention $\int := \int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2 \cdots \int_{t_0}^{t_{n-1}} dt_n$.

The magic Dyson series. It is our goal to simplify the solution formula (1.28) by introducing the chronological operator $\mathcal{T}$. We will obtain the elegant formula

$$\boxed{\Psi(t) = \mathcal{T}\exp\left(\frac{\kappa}{\mathrm{i}\hbar}\int_{t_0}^{t}\mathsf{U}(\tau)d\tau\right)\Psi_0.} \tag{1.29}$$

Explicitly, we get the following Dyson series

$$\Psi(t) = \Psi_0 + \sum_{n=1}^{\infty} \frac{\kappa^n}{n!(\mathrm{i}\hbar)^n} \int_{t_0}^{t}\cdots\int_{t_0}^{t} \mathcal{T}(\mathsf{U}(t_1)\cdots\mathsf{U}(t_n))\Psi_0\, dt_1\cdots dt_n.$$

Here, the chronological operator $\mathcal{T}$ organizes the factors in such a way that time is increasing from right to left.[17] For example,

$$\mathcal{T}(\mathsf{U}(t_1)\mathsf{U}(t_2)) := \begin{cases} \mathsf{U}(t_1)\mathsf{U}(t_2) & \text{if} \quad t_1 \geq t_2, \\ \mathsf{U}(t_2)\mathsf{U}(t_1) & \text{if} \quad t_2 > t_1. \end{cases}$$

More generally,

$$\mathcal{T}(\mathsf{U}(t_1)\mathsf{U}(t_2)\cdots\mathsf{U}(t_n)) := \mathsf{U}(t_{1'})\mathsf{U}(t_{2'})\cdots\mathsf{U}(t_{n'})$$

where $t_{1'}, \ldots, t_{n'}$ is a permutation of $t_1, \ldots, t_n$ such that $t_{1'} \geq t_{2'} \geq \ldots \geq t_{n'}$. Consider now the integral

$$J := \int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2 \mathsf{U}(t_1)\mathsf{U}(t_2) = \int_{t_0}^{t}\int_{t_0}^{t} \mathsf{U}(t_1)\mathsf{U}(t_2)\theta(t_1 - t_2)dt_1 dt_2$$

where $\theta(t) := 1$ if $t \geq 0$, and $\theta(t) := 0$ if $t < 0$. Using a permutation of indices,

[17] In the present case, the use of the chronological operator $\mathcal{T}$ is trivial, since $\mathsf{U}(t_1)$ commutes with $\mathsf{U}(t_2)$. However, the chronological operator is crucial if one wants a straightforward generalization of the argument above to finite-dimensional systems of classical ordinary differential equations and infinite-dimensional operator equations appearing in quantum field theory.

$$J = \int_{t_0}^{t}\int_{t_0}^{t} \frac{1}{2}\Big(\mathsf{U}(t_1)\mathsf{U}(t_2)\theta(t_1-t_2) + \mathsf{U}(t_2)\mathsf{U}(t_1)\theta(t_2-t_1)\Big)dt_1dt_2.$$

Hence $J = \int_{t_0}^{t}\int_{t_0}^{t} \frac{1}{2}\mathcal{T}(\mathsf{U}(t_1)\mathsf{U}(t_2))dt_1dt_2$. Similarly, we proceed for $n = 3, 4, \ldots$ in order to get the desired relation (1.29). □

Summarizing, in the Dirac interaction picture we pass from the Schrödinger state function $\psi = \psi(t)$ and the potential $U = U(t)$ to the Dirac state function Ψ and the Dirac potential U, respectively. Explicitly,

$$\Psi(t) := \mathrm{e}^{\mathrm{i}(t-t_0)H_0/\hbar}\psi(t), \qquad \mathsf{U}(t) := \mathrm{e}^{\mathrm{i}(t-t_0)H_0/\hbar}U(t)\,\mathrm{e}^{-\mathrm{i}(t-t_0)H_0/\hbar}$$

for all times $t \in \mathbb{R}$. Here, the Hamiltonian H_0 describes the free dynamics in the Schrödinger picture. The Dirac state function Ψ satisfies the modified Schrödinger equation

$$\boxed{\mathrm{i}\hbar\dot{\Psi}(t) = \kappa\mathsf{U}(t)\Psi(t), \qquad t \in \mathbb{R}}$$

and the Dirac potential U satisfies the equation of motion

$$\mathrm{i}\hbar\dot{\mathsf{U}}(t) = [\mathsf{U}(t), H_0]_- + V(t), \qquad t \in \mathbb{R}$$

with $V(t) := \mathrm{e}^{\mathrm{i}(t-t_0)H_0/\hbar}\dot{U}(t)\mathrm{e}^{-\mathrm{i}(t-t_0)H_0/\hbar}$. If the potential $U(t)$ is time-independent, then

$$\boxed{\mathrm{i}\hbar\dot{\mathsf{U}}(t) = [\mathsf{U}(t), H_0]_-, \qquad t \in \mathbb{R}.}$$

This is the Heisenberg equation of motion with respect to the unperturbed Hamiltonian H_0.

In Chap. 15 we will use this type of argument in order to reduce the investigation of interacting quantum fields to free quantum fields.

Perturbation theory. Suppose that the coupling constant κ is small. We then obtain the first-order (or Born) approximation

$$\Psi(t) = \Psi_0 + \frac{\kappa}{\mathrm{i}\hbar}\int_{t_0}^{t}\mathsf{U}(\tau)\Psi_0 d\tau.$$

Thus, for the solution $\psi(t) = \mathrm{e}^{-\mathrm{i}(t-t_0)H_0/\hbar}\Psi(t)$ of the Schrödinger equation (1.25), we obtain the first-order approximation

$$\boxed{\psi(t) = \psi_0(t) + \frac{\kappa}{\mathrm{i}\hbar}\int_{t_0}^{t}\mathrm{e}^{-\mathrm{i}(t-\tau)H_0/\hbar}U(\tau)\psi_0(\tau)d\tau, \qquad t \in \mathbb{R}}$$

which represents the first-order perturbation of the free dynamics

$$\psi_0(t) := \mathrm{e}^{-\mathrm{i}(t-t_0)H_0/\hbar}\psi_0(t_0).$$

1.2.2 The Fundamental Role of the Harmonic Oscillator in Quantum Field Theory

> The present paper seeks to establish a basis for theoretical quantum mechanics founded exclusively upon relationships between quantities which in principle are observable... We shall restrict ourselves to problems involving one degree of freedom.[18]
>
> Werner Heisenberg, 1925

Since the 1920s, the experience of physicists has shown that

> *Quantum fields can be treated as nonlinear perturbations of an infinite number of uncoupled quantized harmonic oscillators.*

All the computations of physical effects in quantum field theory done by physicists have been based on this general principle. It is the long-term desire of physicists to replace this local approach by a more powerful global approach. The harmonic oscillator and its relations to quantum field theory will be thoroughly studied in Volume II. At this point, let us only sketch the basic ideas.

The classical harmonic oscillator and Poisson brackets. The Newtonian equation of motion for a harmonic oscillator of mass m and coupling constant $\kappa > 0$ on the real line reads as

$$m\ddot{q}(t) = -\kappa q(t), \qquad t \in \mathbb{R}. \tag{1.30}$$

This is the simplest oscillating system in physics. The equation of motion (1.30) possesses the following general solution

$$\boxed{q(t) = \sqrt{\frac{\hbar}{2m\omega}}\,(a\mathrm{e}^{-\mathrm{i}\omega t} + a^\dagger \mathrm{e}^{\mathrm{i}\omega t}), \qquad t \in \mathbb{R}} \tag{1.31}$$

along with the angular frequency $\omega := \sqrt{\kappa/m}$. The Fourier coefficient a is a complex number. Introducing the momentum $p(t) := m\dot{q}(t)$ at time t, the relations between the Fourier coefficient a, the conjugate complex value $a^\dagger$, and the initial values of the harmonic oscillator are given by

$$q(0) = \sqrt{\frac{\hbar}{2m\omega}}\;\,(a + a^\dagger), \qquad p(0) = \mathrm{i}\sqrt{\frac{\hbar m\omega}{2}}\;\,(a^\dagger - a). \tag{1.32}$$

The following three equivalent formulations were studied in the history of classical mechanics.[19]

[18] W. Heisenberg, Quantum-theoretical re-interpretation of kinematic and mechanical relations, Zeitschrift für Physik **33** (1925), 879–893. This paper founded quantum mechanics. Heisenberg was awarded the Nobel prize in physics in 1932.

[19] A detailed investigation of the harmonic oscillator can be found in Volume II.

(i) The Lagrangian approach: The function

$$L(q,\dot{q}) := \frac{1}{2}m\dot{q}^2 - \frac{1}{2}\kappa q^2$$

represents the Lagrangian of the harmonic oscillator. The Euler–Lagrange equation $\dot{p} = L_q$ is equivalent to the Newtonian equation (1.30). Along each trajectory $q = q(t)$ of the harmonic oscillator, we have energy conservation

$$E = \frac{p(t)^2}{2m} + \frac{\kappa q(t)^2}{2} \qquad \text{for all} \quad t \in \mathbb{R}.$$

(ii) The Hamiltonian approach: Introducing the Hamiltonian

$$H(q,p) := \frac{p^2}{2m} + \frac{\kappa q^2}{2},$$

the Newtonian equation (1.30) is equivalent to the following Hamiltonian equations of motion

$$\dot{p} = -H_q, \qquad \dot{q} = H_p. \tag{1.33}$$

This is also called canonical equation. Explicitly, $\dot{p}(t) = -\kappa q(t)$ and $\dot{q}(t) = p(t)/m$. We will show in Volume II that there is a symplectic structure behind the Hamiltonian approach.

(iii) The Poissonian approach: Let us introduce the Poisson bracket

$$\{A(q,p), B(q,p)\} := A_q B_p - B_q A_p$$

where A_q denotes the partial derivative with respect to the variable q. The Hamiltonian equations (1.33) of motion can then be written as

$$\boxed{\dot{p} = \{p, H\}, \qquad \dot{q} = \{q, H\}.} \tag{1.34}$$

This reveals the Poissonian structure behind classical mechanics. Moreover, we have

$$\boxed{\{q,p\} = 1.} \tag{1.35}$$

We will show below that the equations (1.34) and (1.35) are the key to the quantization of the harmonic oscillator. This was discovered gradually by Heisenberg, Born, Jordan, and Dirac in 1925/1926. Let us discuss this.

Heisenberg's philosophical principle. In 1925 Heisenberg wanted to understand atomic spectra. As a mathematical model, he considered the infinite scheme

$$q(t) = (q_{nm} \mathrm{e}^{\mathrm{i}\omega_{nm}t}), \qquad n,m = 1,2,\ldots$$

of angular frequencies ω_{nm} and complex-valued amplitudes q_{nm}. Following Einstein and Bohr, Heisenberg postulated that the angular frequencies are related to the possible energies of the system by the equation

$$\omega_{nm} = \frac{E_n - E_m}{\hbar}.$$

It was his goal to compute the energy levels $E_1, E_2, \ldots$ and the intensities of the spectral lines which are proportional to the squares $|q_{nm}|^2$. To this end, Heisenberg developed some simple rules for the scheme. Finally, he got the crucial energy relation

$$\boxed{E_n = \left(n + \frac{1}{2}\right)\hbar\omega, \qquad n = 0, 1, 2, \ldots}$$

This was the birth of modern quantum mechanics. From the philosophical point of view, Heisenberg did only use quantities which are directly related to physical experiments in the spectroscopy of atoms and molecules. In particular, he did not use the notion of trajectory or velocity of a quantum particle. In the same philosophical spirit, Heisenberg introduced the S-matrix in 1943; this approach has been very successful in elementary particle physics.

Heisenberg did not know the mathematical notion of matrix. In fact, in his 1925 paper he invented matrix multiplication by using physical arguments. When reading Heisenberg's manuscript, Born remembered some course in matrix calculus from the time of his studies; he conjectured that there should hold the following commutation relation

$$\boxed{q(t)p(t) - p(t)q(t) = \mathrm{i}\hbar I \qquad \text{for all} \quad t \in \mathbb{R}} \tag{1.36}$$

with $p(t) := m\dot{q}(t) = (\mathrm{i}m\omega_{nm}q_{nm}\mathrm{e}^{\mathrm{i}\omega_{nm}t})$. Recall that the symbol I denotes the identity operator. Born, himself, could prove (1.36) only for the diagonal elements. The general proof was then obtained with the help of his young assistant Pascual Jordan in Göttingen. For this historical reason, the commutation relation (1.36) will be called the Heisenberg–Born–Jordan commutation relation in what follows.

The Heisenberg picture of the quantum harmonic oscillator and Lie brackets. Let us now formulate Heisenberg's approach to quantum mechanics in the manner polished by Born, Jordan, and Dirac. For the quantum harmonic oscillator, the classical motion

$$q = q(t), \qquad t \in \mathbb{R}$$

is replaced by the operator-valued function $q = q(t)$. Moreover, in order to obtain the equation of motion, we use the Poissonian approach, and we replace the classical Poisson bracket by the Lie bracket. Explicitly,

$$\boxed{\{A, B\} \;\Rightarrow\; \frac{1}{\mathrm{i}\hbar}\,[A, B]_-}$$

where $[A, B]_- := AB - BA$.[20] From (1.34) and (1.35), we get the equations of motion

$$\boxed{\mathrm{i}\hbar\dot{p}(t) = [p(t), H(t)]_-, \qquad \mathrm{i}\hbar\dot{q}(t) = [q(t), H(t)]_-,} \tag{1.37}$$

and the commutation relation (1.36) along with the Hamiltonian

$$H(t) = \frac{p(t)^2}{2m} + \frac{\kappa q(t)^2}{2}.$$

It turns out that

> *This problem can be solved easily by using the classical solution* (1.31) *and by replacing the Fourier coefficient a by an operator. Here, we have to assume that the operator a and its adjoint $a^\dagger$ satisfy the following commutation relation*

$$\boxed{[a, a^\dagger]_- = I.} \tag{1.38}$$

This method is called Fourier quantization. In Volume II, we will use this method in order to obtain quantum electrodynamics as a quantum field theory which generalizes classical Maxwell's theory of electromagnetism.

The Schrödinger picture. The Schrödinger equation for the harmonic oscillator reads as

$$\boxed{\mathrm{i}\hbar\dot{\psi}(x,t) = H\psi(x,t), \qquad x, t \in \mathbb{R}}$$

with the momentum operator $(P\psi)(x,t) := -\mathrm{i}\hbar\psi_x(x,t)$, the position operator $(Q\psi)(x,t) := x\psi(x,t)$, and the Hamiltonian

$$H := \frac{P^2}{2m} + \frac{\kappa Q^2}{2}.$$

Explicitly, the Schrödinger equation reads as

$$\mathrm{i}\hbar\dot{\psi}(x,t) = -\frac{\hbar^2}{2m}\,\psi_{xx}(x,t) + \frac{\kappa}{2}\,x^2\psi(x,t).$$

The ansatz $\psi(x,t) = \varphi(x)\mathrm{e}^{-\mathrm{i}Et/\hbar}$ yields the stationary Schrödinger equation

$$\boxed{E\varphi = H\varphi.} \tag{1.39}$$

[20] This general rule is due to Dirac. In 1928, Jordan and Wigner discovered that one has to replace the commutator $[A, B]_-$ for bosons (e.g., photons) by the anticommutator $[A, B]_+ := AB + BA$ in the case of fermions (e.g., electrons).

Explicitly, $E\varphi(x) = -\frac{\hbar^2}{2m}\varphi''(x) + \frac{\kappa}{2}x^2\varphi(x)$. This is an eigenvalue problem for computing the unknown energy E. From classical analysis it is known that the Hermite functions $\varphi_0, \varphi_1, \dots$ are eigenfunctions of (1.39). Let us use the language of physicists in order to obtain these eigenfunctions in a very elegant manner. Motivated by (1.32), we set $p(0) := P, q(0) := Q$, and hence

$$Q = \sqrt{\frac{\hbar}{2m\omega}}\ (a + a^\dagger), \qquad P = \mathrm{i}\sqrt{\frac{\hbar m\omega}{2}}\ (a^\dagger - a).$$

To simplify notation, let $m = \omega = \hbar = 1$. This implies $\kappa = 1$. Then,

$$a = \frac{1}{\sqrt{2}}\,(Q + \mathrm{i}P), \qquad a^\dagger = \frac{1}{\sqrt{2}}\,(Q - \mathrm{i}P).$$

It follows from the commutation relation $[Q, P]_- = \mathrm{i}I$ that this choice of the operator a satisfies the commutation relation (1.38). We will show in Volume II, using only the commutation relations, that the functions

$$\varphi_n := \frac{1}{\sqrt{n!}}\,(a^\dagger)^n\varphi_0, \qquad n = 0, 1, 2, \dots$$

with $\varphi_0(x) := c_0\mathrm{e}^{-x^2/2}$ are eigensolutions of the equation

$$E_n\varphi_n = H\varphi_n, \qquad n = 0, 1, 2, \dots$$

with the eigenvalues $E_n := (n + \frac{1}{2})$.[21] If we choose the constant $c_0 := \pi^{-1/4}$, then

$$\langle\varphi_n|\varphi_m\rangle := \int_{-\infty}^{\infty} \varphi_n(x)^\dagger\varphi_m(x)\,dx = \delta_{nm}, \qquad n, m = 0, 1, 2, \dots$$

In other words, the eigenfunctions $\varphi_0, \varphi_1, \dots$ form an orthonormal system in the Hilbert space $L_2(\mathbb{R})$.[22] For the original Schrödinger equation, we get the solutions

$$\psi_n(x, t) = \varphi_n(x)\mathrm{e}^{-\mathrm{i}E_nt/\hbar}, \qquad n = 0, 1, 2, \dots$$

which describe quantum oscillations of the quantum particle on the real line with energy $E_n = (n + \frac{1}{2})\hbar\omega$.

The Feynman picture. Using the eigenfunctions $\psi_0, \psi_1, \dots$, we can construct the Feynman propagator kernel

$$\mathcal{P}(x, t; y, t_0) = \sum_{n=0}^{\infty} \psi_n(x, t)\psi_n(y, t_0)^\dagger.$$

[21] This corresponds to $E_n = \left(n + \frac{1}{2}\right)\hbar\omega$ when our simplification $\hbar = \omega = m = 1$ drops out.

[22] In fact, it is shown in Zeidler (1995), Vol. 1, Sect. 3.4 that this orthonormal system is complete.

This kernel knows all about the dynamics of the quantum harmonic oscillator. In fact, suppose that we are given the wave function $\psi(x, t_0) := \varphi(x)$ at the initial time t_0. For arbitrary points x on the real line and arbitrary real time t, the wave function is then given by the formula

$$\boxed{\psi(x,t) = \int_{-\infty}^{\infty} \mathcal{P}(x,t;y,t_0)\varphi(y)dy.} \tag{1.40}$$

The fundamental role of Green's functions in mathematics and physics. In terms of physics, the Feynman propagator kernel $\mathcal{P}$ allows the following intuitive interpretation. Choose the initial state

$$\varphi(x) := \varphi_0 \delta(x - x_0)$$

where φ_0 is a fixed complex number.[23] Formally, this corresponds to an initial state which is sharply concentrated at the point x_0 at the initial time t_0. By (1.40), we get the solution

$$\psi(x,t) = \varphi_0 \mathcal{P}(x,t;x_0,t_0)$$

for all positions $x \in \mathbb{R}$ and all times $t \geq t_0$. Thus, the Feynman propagator describes the propagation of a sharply concentrated initial state. Formula (1.40) tells us then that the general dynamics is the superposition of sharply concentrated initial states $\varphi(x_0)\delta(x - x_0)$. This is the special case of a general strategy in mathematics and physics called the strategy of the Green's function:

- Study first the response of a given physical system under the action of a sharply concentrated external force. This response corresponds to the Green's function of the system.
- The total response of the system under the action of a general external force can then be described by the superposition of sharply concentrated forces.

The response approach to quantum field theory will be studied in Chap. 14.

The importance of Fock states in quantum field theory. In the example above, the states

$$\varphi_n := \frac{1}{\sqrt{n!}} \ (a^\dagger)^n \varphi_0, \qquad n = 0, 1, 2, \ldots$$

span the Hilbert space $L_2(\mathbb{R})$. These states are called Fock states, and $L_2(\mathbb{R})$ is called the corresponding Fock space. The state φ_0 represents the ground state, and we have

$$a\varphi_0 = 0.$$

[23] The meaning of the Dirac delta function δ can be found on page 589.

Furthermore, for the operator $N := a^\dagger a$, we get

$$N\varphi_n = n\varphi_n, \qquad n = 0, 1, 2, \ldots$$

In Chap. 15 we will generalize this model to quantum field theory. Then, the following will happen:

- The state φ_0 passes over to the vacuum state $|0\rangle$ of a free quantum field.
- The operator $a^\dagger$ is called creation operator.
- The Fock state φ_n corresponds to a state which consists of n particles.
- Because $a\varphi_0 = 0$, the operator a is called annihilation operator.
- The Fock state φ_n is a common eigenstate of the energy operator H and the particle number operator N with the eigenvalue n which counts the number of particles of φ_n.

1.2.3 Quantum Fields and Second Quantization

Quantum field theory was founded by Heisenberg and Pauli in 1929.[24] From the physical point of view the following is crucial:

> *A quantum field can be treated as a system of an infinite number of quantum particles where creation and annihilation of particles are possible.*

In particular, for studying the radiation of atoms and molecules, one has to consider the quantum field of photons. In quantum electrodynamics, one investigates the quantum field of electrons, positrons, and photons.

The second quantization of the Schrödinger equation. As a prototype, let us consider the quantum field corresponding to the Schrödinger equation. We will proceed in several steps.

- Step 1: Classical mechanics. We start with a classical particle on the real line. The principle of critical action reads as

$$\int_{t_0}^{t_1} L(q(t), \dot{q}(t))dt = \text{critical!}$$

along with the boundary condition "$q(t) = \text{given}$" for $t = t_0, t_1$. This leads to the Euler–Lagrange equation

$$\frac{d}{dt} L_{\dot{q}}(q(t), \dot{q}(t)) = L_q(q(t), \dot{q}(t))$$

which describes the motion, $q = q(t)$, of the classical particle on the real line.[25] Let us consider the special case where

[24] W. Heisenberg and W. Pauli, On quantum field theory (in German), Zeitschrift für Physik **56** (1929), 1–61; **59** (1930), 108–190.

[25] The derivation of the Euler–Lagrange equation in classical mechanics along with symplectic and Poissonian geometry will be studied in Chaps. 4 and 5 of Vol. II.

$$L(q,\dot{q}) := \frac{m\dot{q}^2}{2} - \kappa U(q).$$

We define the momentum $p := L_{\dot{q}}$ and the Hamiltonian $H := p\dot{q} - L$. Hence $p = m\dot{q}$, and

$$H = \frac{p^2}{2m} + \kappa U(q).$$

Set $H(t) := \frac{p(t)^2}{2m} + \kappa U(q(t))$. By energy conservation, we have

$$H(t) = H(0) \qquad \text{for all} \quad t \in \mathbb{R},$$

for each smooth solution $q = q(t)$ of the Euler–Lagrange equation.

- Step 2: First quantization by using Heisenberg's particle picture. We want to describe a quantum particle on the real line.

  To this end, we replace the classical trajectory $q = q(t)$ by an operator-valued function.

  This implies the operators $p(t) := m\dot{q}(t)$ and $H(t)$ as given above. More precisely, for each time t, we have the commutation relation

$$[q(t), p(t)]_- = \mathrm{i}\hbar I$$

  and the following equations of motion[26]

$$\boxed{\mathrm{i}\hbar\dot{q}(t) = [q(t), H(t)]_-, \qquad \mathrm{i}\hbar\dot{p}(t) = [p(t), H(t)]_-.}$$

  We will show in Volume II that this implies the Newtonian equation of motion $m\ddot{q}(t) = -\kappa U'(q(t))$. Furthermore, the energy operator $H(t)$ does not depend on time t. To simplify notation, this operator is denoted by the symbol H.

- Step 3: First quantization by using Schrödinger's wave picture. Here, the quantum particle on the real line is described by the complex-valued wave function $\psi = \psi(x,t)$ which satisfies the Schrödinger equation

$$\mathrm{i}\hbar\dot{\psi}(x,t) = -\frac{\hbar^2}{2m}\,\psi_{xx}(x,t) + \kappa U(x)\psi(x,t). \tag{1.41}$$

  First of all we want to derive the Schrödinger equation by the principle of critical action of the form

$$\int_{t_0}^{t_1} \left(\int_{x_0}^{x_1} \mathcal{L}\, dx \right) dt = \text{critical!} \tag{1.42}$$

  along with the boundary condition "$\psi(x,t) =$ given" on the boundary $\partial\Omega$ of the rectangle $\Omega := \{(x,t) \in \mathbb{R}^2 : x_0 \le x \le x_1,\ t_0 \le t \le t_1\}$. Explicitly, for the Lagrangian density,

[26] Recall that $[A,B]_- := AB - BA$.

$$\mathcal{L}(\psi, \dot{\psi}, \psi_x; \psi^\dagger, \dot{\psi}^\dagger, \psi_x^\dagger) := \mathrm{i}\hbar\psi^\dagger\dot{\psi} - \frac{\hbar^2}{2m}\psi_x^\dagger\psi_x - \kappa U\psi^\dagger\psi$$

with the real potential $U = U(x)$. Recall that $\dot{\psi}$ denotes the partial derivative with respect to time t. Each classical solution $\psi = \psi(x,t)$ of (1.42) satisfies the two Euler–Lagrange equations

$$\frac{\partial}{\partial x}\mathcal{L}_{\psi_x} + \frac{\partial}{\partial t}\mathcal{L}_{\dot{\psi}} = \mathcal{L}_\psi \tag{1.43}$$

and

$$\frac{\partial}{\partial x}\mathcal{L}_{\psi_x^\dagger} + \frac{\partial}{\partial t}\mathcal{L}_{\dot{\psi}^\dagger} = \mathcal{L}_{\psi^\dagger}. \tag{1.44}$$

Equation (1.43) is precisely the Schrödinger equation (1.41), whereas equation (1.44) is obtained from the Schrödinger equation by applying the operation of complex conjugation, that is,

$$-\mathrm{i}\hbar\dot{\psi}^\dagger = -\frac{\hbar^2}{2m}\psi_{xx}^\dagger + \kappa U\psi^\dagger.$$

Thus, equation (1.44) does not provide us any new information. Introduce the momentum

$$\pi := \mathcal{L}_{\dot{\psi}}.$$

Explicitly, $\pi(x,t) = \mathrm{i}\hbar\psi^\dagger(x,t)$. Moreover, we introduce the Hamiltonian density

$$\mathcal{H} := \pi\dot{\psi} - \mathcal{L}$$

and the Hamiltonian $H := \int_{-\infty}^{\infty} \mathcal{H}dx$. Explicitly,

$$\mathcal{H} = \frac{\hbar^2}{2m}\,\psi_x^\dagger\psi_x + U\psi^\dagger\psi.$$

Here, H represents the energy of the classical field ψ.

- Step 4: Second quantization of the Schrödinger equation and the quantum field. We now want to describe an infinite number of quantum particles on the real line including the creation and annihilation of particles.

  To this end, we replace the classical wave function $\psi = \psi(x,t)$ by an operator-valued function.

  More precisely, $\psi(x,t)$ is an operator which, for all positions $x, y \in \mathbb{R}$ and all times $t \in \mathbb{R}$, satisfies the so-called canonical commutation relations

$$\begin{aligned}[\psi(x,t), \pi(y,t)]_- &= \mathrm{i}\hbar\delta(x-y),\\ [\psi(x,t), \psi(y,t)]_- &= [\pi(x,t), \pi(y,t)]_- = 0\end{aligned}$$

  along with the equations of motion

$$\boxed{\mathrm{i}\hbar\dot{\psi} = [\psi, H]_-, \qquad \mathrm{i}\hbar\dot{\pi} = [\pi, H]_-.}$$

It turns out that this implies the Schrödinger equation for the quantum field $\psi = \psi(x,t)$.[27]

The prototype of a quantum field and the method of Fourier quantization. Suppose that we know a system $\varphi_0, \varphi_1, \ldots$ of eigensolutions of the stationary Schrödinger equation,

$$E\varphi_n = -\frac{\hbar^2}{2m}(\varphi_n)_{xx} + \kappa U\varphi_n, \qquad n = 0, 1, 2, \ldots$$

where $\varphi_0, \varphi_1, \ldots$ represents a complete orthonormal system in the Hilbert space $L_2(\mathbb{R})$. The Fourier series

$$\psi(x,t) = \sum_{n=0}^{\infty} \varphi_n(x)\mathrm{e}^{-\mathrm{i}E_n t/\hbar} a_n \tag{1.45}$$

with complex numbers $a_0, a_1, \ldots$ is then a solution of the Schrödinger equation. Replace now the classical Fourier coefficients by operators $a_0, a_1, \ldots$ which, for all $n, m = 0, 1, \ldots$ satisfy the commutation relations

$$\left[a_n, a_m^\dagger\right]_- = \delta_{nm} I, \qquad [a_n, a_m]_- = \left[a_n^\dagger, a_m^\dagger\right]_- = 0.$$

The classical field ψ from (1.45) passes then over to a quantum field which consists of an infinite number of particles having the energies $E_0, E_1, \ldots$ We assume that there exists a state $|0\rangle$ which is free of particles. This state of lowest energy E_0 is called ground state (or vacuum).The symbol

$$a_{i_1}^\dagger a_{i_2}^\dagger \cdots a_{i_N}^\dagger |0\rangle$$

represents then a state of the quantum field which consists of precisely N free particles possessing the energies $E_{i_1}, \ldots, E_{i_N}$. Moreover, the symbol

$$\psi_{\mathrm{free}}^\dagger(x_1, t) \cdots \psi_{\mathrm{free}}^\dagger(x_N, t)|0\rangle$$

represents a state at time t which is related to N free particles. Here, it is important to distinguish between

- the ground state $|0\rangle$ of the free quantum field ψ_{free} without any interactions,
- and the ground state $|0_{\mathrm{int}}\rangle$ of the interacting quantum field ψ.

The main trouble of quantum field theories concerns the investigation of interacting quantum fields in rigorous mathematical terms.

[27] The Dirac delta function δ represents a generalization of the Kronecker symbol δ_{ij} to infinite degrees of freedom. In particular, $\delta(x-y) = 0$ if $x \neq y$. For the heuristic and rigorous definition of δ, see pages 590 and 609, respectively.

Commutation relations for creation and annihilation operators. In elementary particle physics, we have to distinguish between bosons (particles of integer spin, e.g., photons) and fermions (particles of half-integer spin, e.g., electrons). The prototype of commutation relations for annihilation operators $a(\mathbf{p})$ and creation operators $a^\dagger(\mathbf{p})$ of bosonic particles of a momentum vector $\mathbf{p}$ reads as [28]

$$\boxed{\big[a(\mathbf{p}), a^\dagger(\mathbf{q})\big]_- = \delta_{\mathbf{pq}} I, \qquad [a(\mathbf{p}), a(\mathbf{q})]_- = \big[a^\dagger(\mathbf{p}), a^\dagger(\mathbf{q})\big]_- = 0}$$

for all momentum vectors $\mathbf{p}, \mathbf{q}$ which lie on a fixed lattice of width Δp in 3-dimensional momentum space. Here, we use the 3-dimensional Kronecker symbol defined by $\delta_{\mathbf{pp}} := 1$ and $\delta_{\mathbf{pq}} = 0$ if $\mathbf{p} \neq \mathbf{q}$. Physicists pass to the formal continuum limit. To consider this, let us rescale the annihilation and creation operators by setting

$$\mathbf{a}(\mathbf{p}) := \frac{a(\mathbf{p})}{\sqrt{(\Delta p)^3}}, \qquad \mathbf{a}^\dagger(\mathbf{p}) := \frac{a^\dagger(\mathbf{p})}{\sqrt{(\Delta p)^3}}.$$

Hence

$$\big[\mathbf{a}(\mathbf{p}), \mathbf{a}^\dagger(\mathbf{q})\big]_- = \frac{\delta_{\mathbf{pq}}}{(\Delta p)^3}\, I, \qquad [\mathbf{a}(\mathbf{p}), \mathbf{a}(\mathbf{q})]_- = \big[\mathbf{a}^\dagger(\mathbf{p}), \mathbf{a}^\dagger(\mathbf{q})\big]_- = 0.$$

The formal continuum limit $\Delta p \to 0$ yields then

$$\big[\mathbf{a}(\mathbf{p}), \mathbf{a}^\dagger(\mathbf{q})\big]_- = \delta^3(\mathbf{p} - \mathbf{q}) I, \qquad [\mathbf{a}(\mathbf{p}), \mathbf{a}(\mathbf{q})]_- = \big[\mathbf{a}^\dagger(\mathbf{p}), \mathbf{a}^\dagger(\mathbf{q})\big]_- = 0$$

for all 3-dimensional momentum vectors $\mathbf{p}$ and $\mathbf{q}$. The relation between the discrete Dirac delta function and its continuum limit is studied on page 673. The rigorous mathematical approach to creation and annihilation operators for free quantum particles in terms of the so-called Fock space can be found in Volume II.

The fundamental role of correlations of a quantum field. The experience of physicists in quantum physics shows that one should prefer the study of quantities which are related to measurements in physical experiments. From the physical point of view, we can measure

- cross sections of scattering processes for elementary particles, and
- masses of bound particles (like the proton as a bound state of three quarks).

It turns out that these quantities are related to correlations between different space-time points of the quantum field. According to Feynman, the basic quantity is the correlation function

$$\boxed{G_2(x_1, t_1; x_2, t_2) := \langle 0_{\text{int}} | \mathcal{T} \psi(x_1, t_1) \psi^\dagger(x_2, t_2) | 0_{\text{int}} \rangle}$$

[28] Recall that $[A, B]_- := AB - BA$ and $[A, B]_+ := AB + BA$. For fermions, one has to replace the Lie bracket $[.,.]_-$ by the Jordan–Wigner bracket $[.,.]_+$.

which is also called the 2-point Green's function of the interacting quantum field ψ. This function describes the correlation between the quantum field at position x_1 at time t_1 and the quantum field at position x_2 and time t_2. Here, the symbol $\mathcal{T}$ denotes the chronological operator. Explicitly,

$$\mathcal{T}(\psi(x_1,t_1)\psi^\dagger(x_2,t_2)) := \begin{cases} \psi(x_1,t_1)\psi^\dagger(x_2,t_2) & \text{if} \quad t_1 \geq t_2, \\ \psi^\dagger(x_2,t_2)\psi(x_1,t_1) & \text{if} \quad t_2 > t_1. \end{cases}$$

It turns out that

The 2-point Green's function G_2 of a quantum field is a highly singular mathematical object.

This fact causes serious mathematical difficulties. Similarly, the $2n$-point Green's function is obtained by replacing the product $\psi(x_1,t_1)\psi^\dagger(x_2,t_2)$ by a product of $2n$ field operators. For example, the 4-point Green's function G_4 is given by

$$\langle 0_{\text{int}}|\mathcal{T}\psi(x_1,t_1)\psi(x_2,t_2)\psi^\dagger(x_3,t_3)\psi^\dagger(x_4,t_4)|0_{\text{int}}\rangle.$$

The Green's functions $G_2, G_4, G_6, \ldots$ of a quantum field are closely related to the moments of the quantum field which contain the information on the probability structure of the quantum field.

1.2.4 The Importance of Functional Integrals

The Feynman picture of quantum field theory and the method of moments. For quantum field theory, it is crucial that the $2n$-point Green's functions can be expressed by Feynman functional integrals (also called path integrals). For example,

$$\boxed{G_2(x,t;y,s) = \frac{\int \psi(x,t)\psi^\dagger(y,s)\mathrm{e}^{\mathrm{i}S[\psi,\psi^\dagger]/\hbar}\,\mathcal{D}\psi\,\mathcal{D}\psi^\dagger}{\int \mathrm{e}^{\mathrm{i}S[\psi,\psi^\dagger]/\hbar}\,\mathcal{D}\psi\,\mathcal{D}\psi^\dagger}}$$

where we integrate over all possible classical fields $\psi, \psi^\dagger$. In this connection, we use the classical action

$$S[\psi,\psi^\dagger] := \int_{\mathbb{R}^2} \mathcal{L}\, dxdt$$

where the Lagrangian density $\mathcal{L}$ depends on $\psi, \psi^\dagger$, and their first-order partial derivatives. The crucial point is that the formula for G_2 makes also sense if the Lagrangian $\mathcal{L}$ contains nonlinear terms in ψ and $\psi^\dagger$ which describe self-interactions of the quantum field. For example, we may replace the potential U by the field product $\psi^\dagger\psi$. Then

$$\mathcal{L} := \mathrm{i}\hbar\psi^\dagger\dot{\psi} - \frac{\hbar^2}{2m}\psi_x^\dagger\psi_x - \kappa(\psi^\dagger)^2\psi^2.$$

The relation to the Gaussian distribution in the theory of probability. For $k = 0, 1, 2, \ldots$, the quantity

$$M_k := \frac{\int_{\mathbb{R}} x^k \mathrm{e}^{-x^2/2\sigma^2} dx}{\int_{\mathbb{R}} \mathrm{e}^{-x^2/2\sigma^2} dx}$$

is called the kth moment of the Gaussian distribution in the theory of probability. By the classical moment theorem, a probability distribution is uniquely determined by its infinite series $M_0, M_1, M_2, \ldots$ of moments.[29] There exists the following simple trick for computing the moments. We introduce the function

$$Z(J) := \mathcal{C} \int_{\mathbb{R}} \mathrm{e}^{-x^2/2\sigma^2} \ \mathrm{e}^{Jx} \ dx$$

of the real variable J where the normalization constant $\mathcal{C}$ is chosen in such a way that $Z(0) = 1$. Then, for $k = 0, 1, 2, \ldots$,

$$\boxed{M_k = \frac{d^k Z(0)}{dJ^k}.}$$

Naturally enough, the function $Z = Z(J)$ is called the generating function of the moments. Physicists call this the Wick moment trick. It is quite remarkable that the investigation of general interacting quantum fields can be based on an infinite-dimensional version of the Wick trick. The point is that classical integrals have to be replaced by Feynman functional integrals. Here, physicists start with the so-called generating functional integral

$$\boxed{Z(J, J^\dagger) = \mathcal{C} \int \mathrm{e}^{\mathrm{i}S[\psi,\psi^\dagger]/\hbar} \ \mathrm{e}^{\int_{\mathbb{R}^2}(\psi J + \psi^\dagger J^\dagger)dxdt} \ \mathcal{D}\psi \ \mathcal{D}\psi^\dagger}$$

where the normalization $\mathcal{C}$ is chosen in such a way that $Z(0, 0) = 1$. Then

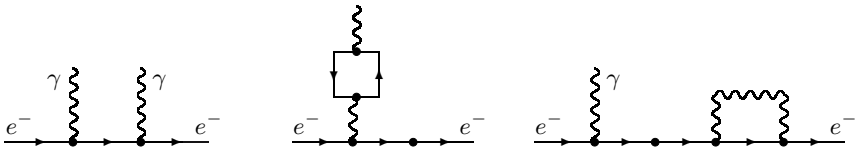
$$G_2(x, t; y, s) = \left(\frac{\hbar}{\mathrm{i}}\right)^2 \frac{\delta^2 Z(0, 0)}{\delta J(x, t) \delta J^\dagger(y, s)}.$$

Analogously, one obtains the higher-order Green's functions by applying functional derivatives of higher order. The precise definition of functional derivatives can be found in Sect. 7.20.1 on page 396. Note that

> *Functional derivatives and functional integrals are natural generalizations of classical partial derivatives and classical multidimensional integrals to infinite dimensions, respectively.*

They have been used systematically by physicists in the 20th century in order to generalize the classical calculus due to Newton and Leibniz to an infinite number of degrees of freedom which appear typically in quantum field theory.

[29] See Zeidler (1995), Vol. 2, Sect. 1.4.



(a) $n = 2$ (b) $n = 4$

Fig. 1.3. Virtual particles

The Feynman approach to interacting quantum fields can be based on the moments of "infinite-dimensional Gaussian distributions".

This will be thoroughly studied in the volumes of this treatise.

Virtual particles. It turns out that the Feynman diagrams from Fig. 1.2 on page 41 can be generalized to quantum fields. Fig. 1.3 displays some Feynman diagrams which describe the scattering of one electron with one photon (Compton effect). The diagrams are nothing more than a graphical representation of the analytic expressions of perturbation theory. If one gives the diagrams a physical interpretation, then there occur, for example, photons for which the physical relation

$$E^2 = c^2\mathbf{p}^2$$

between energy E and momentum vector $\mathbf{p}$ is violated. Such particles are called virtual photons by physicists.[30] Without using virtual particles, the computation of scattering processes would give wrong results. This shows that

The interactions between quantum fields are based on both real and virtual particles.

The Bethe–Salpeter equation for bound states. The state

$$\psi^\dagger(x,t)\psi^\dagger(y,s)|0_{\text{int}}\rangle$$

of the quantum field ψ describes the physics of systems consisting of two quantum particles. In particular, we expect that this state also contains bound states between two particles. Let $|p\rangle$ denote a one-particle state of momentum p. In 1951 Bethe and Salpeter used the so-called Bethe amplitudes

$$\chi_p(x,t;y,s) := \langle 0_{\text{int}}|\mathcal{T}\psi(x,t)\psi^\dagger(y,s)|p\rangle$$

in order to derive an integral equation for χ_p which contains the unknown energy E of the bound state as an eigenvalue parameter. This is the famous

[30] The Compton effect will be thoroughly studied in Volume II in the context of quantum electrodynamics.

Bethe–Salpeter equation which will be considered in Volume V on the physics of the Standard Model.[31]

1.3 The Role of Göttingen

> One cannot comprehend what it is one possesses if one has not understood what one's predecessors possessed.
>
> Johann Wolfgang von Goethe (1749–1832)

> I am ill-mannered, for I take a lively interest in a mathematical object only where I see a prospect of a clever connection of ideas or of results recommended by elegance or generality.
>
> Carl Friedrich Gauss (1777–1855)

In 1807, Carl Friedrich Gauss got the position as professor for astronomy and director of the observatory in Göttingen. His successors were Lejeune Dirichlet (1805–1859) and Bernhard Riemann (1826–1866) in Göttingen. In 1871, Felix Klein (1849–1925) finished his habilitation in Göttingen. After that he received professorships in Erlangen, Munich, and Leipzig. In 1881 he founded the Mathematical Institute of Leipzig University. After his move to Göttingen in 1886, Sophus Lie (1842–1899) became Klein's successor in Leipzig. Initiated by Klein, Hilbert received a professorship at Göttingen University in 1895. Under Hilbert, Göttingen became an extremely active place in mathematics. In the 1920s, Emmy Noether (1882–1935) revolutionized algebra in Göttingen. The young mathematician Bartel Leendert van der Waerden (1903–1996) attended the lectures given by Emmy Noether and Emil Artin. The result was his highly influential monograph *Modern Algebra*, Springer, Berlin 1930.

The emergence of quantum mechanics. In the late 1920s, Göttingen was the intellectual center for the development of the new quantum physics by Heisenberg (1901–1976), Born (1882–1970), Jordan (1902–1980), Pauli (1900–1958), and von Neumann (1903–1957). Students and scientists from all over the world came to Göttingen in order to take part in this scientific revolution. Among the visitors were the young physicists Vladimir Fock (1890–1974), Lev Landau (1908–1968), and Robert Oppenheimer (1904–1967). In order to get an impression of the flair of the early days of quantum mechanics, let us first quote Werner Heisenberg (1901–1976) and Paul Dirac (1902–1984) who gave *Evening Lectures* at the International Center for Theoretical Physics in Trieste (Italy) in 1968. They were invited by Abdus Salam (1926–1996).[32] Heisenberg pointed out the following:

[31] See Bethe and Salpeter (1957), Itzykson and Zuber (1981), Sect. 10.2, and Gross (1993), Sect. 12.5.

[32] A. Salam (Ed.) (1968). The author would like to thank Professor Armin Uhlmann (Leipzig) for drawing his attention to these beautiful "Evening Lectures". The history of quantum physics can be found in J. Mehra and H. Rechenberg (2002),

I had the impression from my conversation with Bohr (1885–1962) that one should go away from all these classical concepts, one should not speak of the orbit of an electron...

When I came back from Copenhagen to Göttingen I decided that I should again try to do some kind of guess work there, namely, to guess the intensities in the hydrogen spectrum... That was early in the summer 1925 and I failed completely. The formulae got too complicated... At the same time I also felt, if the mechanical system would be simpler, then it might be possible just to do the same thing as Kramers (1894–1952) and I had done in Copenhagen and to guess the amplitudes. Therefore I turned from the hydrogen atom to the anharmonic oscillator, which was a very simple model. Just then I became ill and went to the island of Heligoland to recover. There I had plenty of time to do my calculations. It turned out that it really was quite simple to translate classical mechanics into quantum mechanics. But I should mention one important point. It was not sufficient simply to say "let us take some frequencies and amplitudes to replace orbit quantities" and use a kind of multiplication which we had already used in Copenhagen and which later turned out to be equivalent to matrix multiplication...

It turned out that one could replace the quantum conditions of Bohr's theory by a formula which was essentially equivalent to the sum-rule by Thomas and Kuhn... I was however not able to get a neat mathematical scheme out of it. Very soon afterwards both Born and Jordan in Göttingen and Dirac in Cambridge were able to invent a perfectly closed mathematical scheme; Dirac with very ingenious new methods on q-numbers and Born and Jordan with more conventional methods of matrices[33]...

When you try too much for rigorous mathematical methods you fix your attention on those points which are not important from the physics point and thereby you get away from the experimental situation. If you try to solve a problem by rather dirty mathematics, as I have mostly done, then you are forced always to think of the experimental situation; and whatever formulae you write down, you try to compare the formulae with reality and thereby, somehow, you get closer to reality than by looking for the rigorous methods. But this may, of course, be different for different people...

In 1926 Niels Bohr and I discussed the question on the physical interpretation of quantum mechanics many, many nights and we were frequently in a state of despair. Bohr tried more in the direction of duality between waves and particles; I preferred to start from the mathematical formalism and to look for a consistent interpretation. Finally Bohr went to Norway to think alone about the problem and I remained in Copenhagen. Then I remembered Einstein's remark in our discussion. I remembered that Einstein had said that "It is the theory which decides what can be observed." From there it was easy to turn around our question and not to ask "How can I represent in quantum mechanics this orbit of an electron in a cloud

Vols. 1–6. As an introduction to the development of quantum mechanics in the 1920s, we recommend van der Waerden (1968). For the history of quantum field theory, we refer to Schweber (1994) and Weinberg (1995), Vol. 1, Chap. 1.

[33] Dirac's q-numbers (quantum numbers) are abstract operators in the sense of modern functional analysis, whereas Born and Jordan used concrete realizations of the operators in the form of infinite-dimensional complex matrices.

> chamber?", but rather to ask "Is it not true that always only such situations occur in nature, even in a cloud chamber, which can be described by the mathematical formalism of quantum mechanics?" By turning around I had to investigate what can be described in this formalism; and then it was very easily seen, especially when one used the new mathematical discoveries of Dirac and Jordan about transformation theory, that one could not describe at the same time the exact position and the exact velocity of an electron; one had these *uncertainty relations.* In this way things became clear. When Bohr returned to Copenhagen, he had found an equivalent interpretation with his concept of complementarity, so finally we all agreed that now we had understood quantum theory...
>
> Again we met a difficult situation in 1927 when Einstein and Bohr discussed these matters at the Solvay Conference. Almost every day the sequence of events was the following. We all lived in the same hotel. In the morning for breakfast Einstein would appear and tell Bohr a new fictitious experiment in which he could disprove the uncertainty relations and thereby our interpretation of quantum theory. Then Bohr, Pauli and I would be very worried, we would follow Bohr and Einstein to the meeting and would discuss this problem all day. But at night for dinner usually Bohr had solved the problem and he gave the answer to Einstein, so then we felt that everything was alright and Einstein was a bit sorry about that and said he would think about it. Next morning he would bring a new fictitious experiment, again we had to discuss, and so on. This went on for quite a number of days and at the end of the conference the Copenhagen physicists had the feeling that they had won the battle and that actually Einstein could not make any real objection... Einstein never accepted the probabilistic interpretation of quantum mechanics. He said :"God does not play at dice."

Dirac emphasized the following in his *Evening Lecture* at Trieste:

> I have the best of reasons for being an admirer of Werner Heisenberg. He and I were young research students at the same time, about the same age, working on the same problem. Heisenberg succeeded where I failed. There was a large mass of spectroscopic data accumulated at that time and Heisenberg found out the proper way of handling it. In doing so, he started the golden age of theoretical physics...
>
> One can distinguish between two main procedures for a theoretical physicist. One of them is to work from the *experimental basis.* For this, one must keep in close touch with the experimental physicists. One reads about all the results they obtain and tries to fit them into a comprehensive and satisfying scheme.
>
> The other procedure is to work from the *mathematical basis.* One examines and criticizes the existing theory. One tries to pin-point the faults in it and then tries to remove them. The difficulty here is to remove the faults without destroying the very great success of the existing theory...
>
> This is illustrated by the discovery of quantum mechanics. Two men are involved, Heisenberg and Schrödinger. Heisenberg was working from the experimental basis, using the results of spectroscopy, which by 1925 had accumulated an enormous amount of data... It was Heisenberg's genius that he was able to pick out the important things from the great wealth

> of information and arrange them in a natural scheme. He was thus led to matrices...
>
> Schrödinger's approach was quite different. He worked from the mathematical basis. He was not well informed about the latest spectroscopic results, like Heisenberg was, but had the idea at the back of his mind that spectral frequencies should be fixed by eigenvalue equations, something like those that fix the frequencies of systems of vibrating strings. He had this idea for a long time, and was eventually able to find the right equation, in an indirect way...
>
> Heisenberg and Schrödinger gave us two forms of quantum mechanics, which were soon found to be equivalent. They provided two pictures, with a certain mathematical transformation connecting them. I joined in the early work on quantum mechanics, following the procedure based on mathematics, with a very abstract point of view. I took the noncommutative algebra which was suggested by Heisenberg's matrices as the main feature for a new dynamics...

The following quotation is taken from Max Born's fascinating book *Physics in my Generation*, Springer, New York, 1969:

> In Göttingen we also took part in the attempts to distill the unknown mechanics of the atom out of the experimental results. The logical difficulty became ever more acute. Investigations on scattering and dispersions of light showed that Einstein's conception of transition probability as a measure of the strength of an oscillation was not adequate... The art of guessing correct formulas, which depart from the classical formulas but pass over into them in the sense of Bohr's correspondence principle, was brought to considerable perfection...
>
> This period was brought to a sudden end by Heisenberg, who was my assistant at that time. He cut the Gordian knot by a philosophical principle and replaced guesswork by a mathematical rule. The principle asserts that concepts and pictures that do not correspond to physically observable facts should not be used in theoretical description. When Einstein, in setting up his theory of relativity, eliminated his concepts of the absolute velocity of a body and of the absolute simultaneity of two events at different places, he was making use of the same principle. Heisenberg banished the picture of electron orbits with definite radii and periods of rotation, because these quantities are not observable; he demanded that the theory should be built up by means of quadratic arrays. Instead of describing the motion by giving a coordinate as a function of time $x = x(t)$, one ought to determine an array of transition probabilities (x_{ij}). To me the decisive part in his work is the requirement that one must find a rule whereby from a given array
>
> $$\begin{pmatrix} x_{11} & x_{12} & \dots \\ x_{21} & x_{22} & \dots \\ \dots & & \end{pmatrix}$$
>
> the array for the square $(x^2)_{ij}$ may be found (or, in general, the multiplication law of such arrays).
>
> By consideration of known examples discovered by guesswork, Heisenberg found this rule and applied it with success to simple examples such as

the harmonic and anharmonic oscillator. This was in the summer 1925. Heisenberg, suffering from a severe attack of hay fever, took leave of absence for a course of treatment at the seaside and handed over his paper to me for publication, if I thought I could do anything about it.

The significance of the idea was immediately clear to me, and I sent the manuscript to the publisher.[34] Heisenberg's rule of multiplication left me no peace, and after a week of intensive thought and trial, I suddenly remembered an algebraic theory that I had learned from my teacher, Rosanes, in Breslau. Such quadratic arrays are quite familiar to mathematicians and are called matrices, in association with a definite rule of multiplication. I applied this rule to Heisenberg's quantum condition and found that it agreed for the diagonal elements. It was easy to guess what the remaining elements must be, namely, null; and immediately there stood before me the strange formula

$$qp - pq = \mathrm{i}\hbar. \tag{1.46}$$

This meant that the coordinates q and momenta p are not to be represented by the values of numbers but by symbols whose product depends on the order of multiplication – which do not "commute", as we say.

My excitement over this result was like that of the mariner who, after long voyaging, sees the desired land from afar, and my only regret was that Heisenberg was not with me. I was convinced from the first that we had stumbled on the truth. Yet again a large part was only guesswork, in particular the vanishing of the non-diagonal elements in the foregoing expression. For this problem, I secured the collaboration of my pupil Pascual Jordan, and in a few days we succeeded in showing that I had guessed correctly. The joint paper written by Jordan and myself[35] contains the most important principles of quantum mechanics, including its extension to electrodynamics...

There followed a hectic period of collaboration among the three of us, rendered difficult by Heisenberg's absence. There was a lively interchange of letters... The result was a three-man paper,[36] which brought the formal side of the investigation to a certain degree of completeness. Before this paper appeared, the first dramatic surprise occurred: Paul Dirac's paper on the same subject.[37] The stimulus received through a lecture by Heisenberg in Cambridge led him to results similar to ours in Göttingen, with the difference that he did not have recourse to the known matrix theory of the mathematicians but discovered for himself and elaborated the doctrine of such non-commuting symbols.

[34] W. Heisenberg, Quantum-theoretical re-interpretation of kinematic and mechanical relations, Zeitschrift für Physik **33** (1925), 879–893.

[35] M. Born and P. Jordan, On quantum mechanics, Zeitschrift für Physik **34** (1925), 883–888.

[36] M. Born, W. Heisenberg, and P. Jordan, On quantum mechanics II, Zeitschrift für Physik **35** (1926), 557–615.

[37] P. Dirac, Quantum mechanics and a preliminary investigation of the hydrogen atom, Proc. Roy. Soc. A **110** (1926), 561–569.

The first nontrivial and physically important application of quantum mechanics was made soon afterwards by Wolfgang Pauli,[38] who calculated the stationary energy values of the hydrogen atom by the matrix method and found complete agreement with Bohr's 1913 formulas. From this moment there was no longer any doubt about the correctness of the theory among physicists...

What the real significance of the formalism might be was, however, by no means clear. Mathematics, as often happens, was wiser than interpretative thought. While we were still discussing the point, there occurred the second dramatic surprise: the appearance of Schrödinger's celebrated paper.[39] He followed quite a different line of thought, which derived from Prince Louis de Broglie (1892–1987). The latter had a few years previously made the bold assertion, supported by brilliant theoretical considerations, that wave-corpuscle dualism, familiar to physicists in the case of light, must also be exhibited by electrons; to each freely movable electron there belongs, according to these ideas, a plane wave of perfectly definite wave length, determined by Planck's constant and mass... Schrödinger extended de Broglie's wave equation, which applied to free motion, to the case in which forces act... and he succeeded in deriving the stationary states of the hydrogen atom as monochromatic solutions of his wave equation not extending to infinity. For a short while, at the beginning of 1926, it looked as if suddenly there were two self-contained but entirely distinct systems of explanation in the field – matrix mechanics and wave mechanics. But Schrödinger himself soon demonstrated their complete equivalence.

Wave mechanics enjoyed much greater popularity than the Göttingen or Cambridge version of quantum mechanics. Wave mechanics operates with a wave function ψ, which – at least in the case of one particle – can be pictured in space, and it employs the mathematical methods of partial differential equations familiar to every physicist.

It appeared to me that it was not possible to arrive at a clear interpretation of the Schrödinger ψ-function by considering bound electrons. I had therefore been at pains, as early as the end of 1925... I was at that time the guest of the Massachusetts Institute of Technology in the U.S.A., and there I found in Norbert Wiener (1894–1964) a distinguished collaborator. In a joint 1926 paper we replaced the matrix by the general concept of an operator and, in this way, made possible the description of aperiodic processes... Once more an idea of Einstein's gave the lead. He had thought to make the duality of particles (light quanta or photons) and waves comprehensible by interpreting the square of the optical wave amplitudes as probability density for the occurrence of photons. This idea could at once be extended to Schrödinger's ψ-function:

> *The square of the amplitude, $|\psi|^2$, must represent the probability density for electrons (or other particles).*

To assert this was easy; but how could I prove this? For this purpose atomic scattering processes suggested themselves. A shower of electrons coming from an infinite distance, represented by a wave of known intensity (that

[38] W. Pauli, On the hydrogen spectrum from the standpoint of the new quantum mechanics, Zeitschrift für Physik **36** (1926), 336–363.

[39] E. Schrödinger, Quantization as an eigenvalue problem, Ann. Phys. **9** (1926), 361–376.

> is, $|\psi|^2$) impinge on an obstacle say a heavy atom... In the same way that the water wave caused by a steamer excites secondary circular waves in striking a pile, the incident electron wave is partly transformed by the atom into a secondary spherical wave, whose amplitude of oscillations ψ is different in different directions. The square of the amplitude $|\psi|^2$ of this wave at a great distance from the scattering center then determines the relative probability of scattering in its dependence of direction... Soon Wentzel succeeded in deriving Rutherford's celebrated 1911 formula for the scattering of α-particles from my theory.
>
> But the factor that contributed more than these successes to the speed of acceptance of my statistical interpretation of the ψ-function was a 1927 paper by Heisenberg that contained his celebrated uncertainty relationship, through which the revolutionary character of the new conception was first made clear.

In 1927 Heisenberg left Göttingen in order to get a professorship at Leipzig University. Four years later, van der Waerden came to Leipzig as a professor of mathematics. According to his own words, he liked very much to attend Heisenberg's seminars. In 1932 van der Waerden published his nicely written book *Group Theory and Quantum Mechanics* about applications of group theory to the spectra of non-relativistic molecules and Dirac's relativistic electron.

The challenge of quantum electrodynamics. Already in the early days of quantum mechanics, physicists tried to understand the quantization of the electromagnetic field. Heisenberg and Pauli published two fundamental papers in 1929 and 1930.[40] Quantum electrodynamics was fully developed in the late 1940s. In order to handle meaningless infinite expressions, physicists developed the method of renormalization in the 1930s and 1940s. This ingenious method allows physicists to extract the relevant physical information from mathematically meaningless expressions. We will thoroughly study this important point later on. From the physical point of view, the following is crucial:

- The singularities of the Green's functions reflect both the complicated structure of the ground state of a quantum field and the complex interactions between the unobservable ground state and the real world, by means of quantum fluctuations.
- The renormalization procedure indicates that the present quantum field theory is not a basic theory, but only an effective theory which averages a deeper physical structure, at a fairly low energy scale.

In 1965, Julian Schwinger said in his Nobel prize speech:

> The relativistic quantum theory of fields was born some thirty-five years ago through the paternal efforts of Dirac, Heisenberg, Pauli and others. It was a somewhat retarded youngster, however, and first reached adolescence seventeen years later, an event we are gathered here to celebrate.

[40] W. Heisenberg and W. Pauli, On quantum field theory (in German) Zeitschrift für Physik **56** (1929), 1–61; **59** (1930), 168–190.

1.4 The Göttingen Tragedy

> Wistfully I recall how, during the Nazi occupation of Poland (1939–1945), Edward Marczewski introduced the arcana of analysis to me. In those dark days these were for me bright moments, for which I am infinitely grateful to him.
>
> Krysztof Maurin (born 1923)

In 1933 the Nazi regime reached the political power in Germany. Because of racist repression, the best scientists left Germany, among them Emmy Noether (1882–1935), Emil Artin (1898–1972), Paul Bernays (1888–1977), Max Born (1882–1970), Richard Courant (1882–1972), Albert Einstein (1879–1955), Kurt-Otto Friedrichs (1901–1982), Leon Lichtenstein (1878–1933), John von Neumann (1903–1957), and Hermann Weyl (1885–1955). Furthermore, Edmund Landau (1877–1938) and the Nobel laureate James Franck (1884–1964) lost their positions at Göttingen University. Hilbert's best friend, Otto Blumenthal (1876–1944), was murdered in the Nazi concentration camp Terežin (Theresienstadt). In 1934 Hilbert was asked by the Nazi minister of education about the flourishing scientific life in Göttingen. Hilbert answered:

> There is no mathematics anymore in Göttingen.

The Göttingen tradition moved to the United States of America. Richard Courant founded the famous Courant Institute at New York University (NYU). In this context, we recommend reading the two beautiful biographies about David Hilbert and Richard Courant written by Constance Reid.[41]

After working as professor in Zurich and Prague, Einstein was appointed as director of the Kaiser-Wilhelm Institute for Physics in Berlin in 1914. In 1933 Einstein emigrated to the United States of America where he got a position at the Institute for Advanced Study in Princeton, New Jersey.[42] Einstein lived there until his death in 1955. As Einstein did, John von Neumann left Germany in 1933 and moved to the Institute for Advanced Study in Princeton, too. Von Neumann was one of the greatest mathematicians of the 20th century. His fundamental contributions concern game theory, mathematical economics, mathematical logic, lattice theory, operator theory in Hilbert spaces, operator algebras, mathematical foundations of quantum mechanics, theory of Lie groups, measure theory, statistical physics, ergodic theory, construction of the first computers ENIAC and MANIAC, and foundations of computer science in the 1940s, shock waves, turbulence, meteorology, and

[41] C. Reid, Hilbert, Springer 1970, New York, and C. Reid, Courant in Göttingen and New York: the Life of an Improbable Mathematician, Springer, New York, 1976.

[42] This institute was founded in 1930. The history of this famous research institute can be found in E. Regis, Who Got Einstein's Office? Eccentricity and Genius at the Institute for Advanced Study in Princeton, Addison-Wesley, Reading, Massachusetts.

numerical mathematics. In the context with John von Neumann, let us quote Peter Hilton.[43]

> For a mentor of Ph.D. candidates it would be most easy to educate a poor applied mathematician. The next simplest thing would be to educate a poor pure mathematician. Then an entire quantum gap lies between the education of a good pure mathematician, and finally, an enormous quantum gap, the education of a good applied mathematician. For the latter task, especially after the death of John von Neumann, I would consider no one sufficiently qualified.
>
> The knowledge and abilities which are nowadays required of a really successful applied mathematician, presume an extraordinary high intellectual standard, and, even for the career of our present-day students, it is almost impossible to predict which parts of mathematics will prove most suited for applications.

Besides John von Neumann, another hero of the 20th century mathematics is Hermann Weyl. He studied mathematics in Göttingen from 1903 until 1908. He attended lectures given by Carathéodory (1873–1950), Hilbert (1858–1943), Klein (1849–1925), Koebe (1882–1945), and Zermelo (1871–1953). He was Hilbert's most gifted pupil. From 1913 until 1930, Weyl worked at the Swiss Institute of Technology (ETH) in Zurich. In 1930 he became Hilbert's successor at Göttingen University. Three years later, Hermann Weyl left Germany and joined Einstein and von Neumann at the Institute for Advanced Study in Princeton. Hermann Weyl influenced very strongly the relations between mathematics and physics. He wrote a number of classical monographs about Riemann surfaces, theory of general relativity, group theory and quantum mechanics, representation theory of the classical Lie groups, symmetry, and philosophical questions. His books became bibles for both physicists and mathematicians. In particular, the idea of gauge field theory can be traced back to Weyl's monograph *Space, Time, Matter* from 1923 where he presented Einstein's theory of general relativity along with his own ideas about a general theory of matter, based on scaling invariance and the conformal group. In 1944 Hermann Weyl wrote the following in the Bulletin of the American Mathematical Society:

> A great master of mathematics passed away when Hilbert died in Göttingen on February 14, 1943, at the age of eighty-one. In retrospect, it seems to us that the era of mathematics upon which he impressed the seal of his spirit, and which is now sinking below the horizon, achieved a more perfect balance than has prevailed before or since, between the mastering of single concrete problems and the formation of general abstract concepts. Hilbert's own work contributed not a little to bringing about this happy equilibrium, and the direction we have since proceeded can in many instances be traced back to this impulse. No mathematician of equal stature has arisen from our generation...
>
> Hilbert was singularly free from national and racial prejudice; in all public questions, be they political, social, or spiritual, he stood forever on the side

[43] In: M. Otte (Ed.), Mathematiker über Mathematik, Springer, Berlin 1973.

of freedom, frequently in isolated opposition against the compact majority of his environment... It was not mere chance, when the Nazis "purged" the universities in 1933 and their hand fell most heavily on the Hilbert school that Hilbert's most intimate collaborators left Germany voluntarily or under the pressure of Nazi persecution. He himself was too old, and stayed behind; but the years after 1933 became years of ever-deepening tragic loneliness.

1.5 Highlights in the Sciences

1.5.1 The Nobel Prize

The history of quantum physics in the 20th century is reflected best by the Nobel laureates in physics, chemistry, and medicine. For the convenience of the reader, here is a selection of topics.[44]

(i) **Radioactivity**: Bequerel, and Marie and Pierre Curie 1903 (natural radioactivity), Rutherford 1908 (chemistry of radioactive substances), Marie Curie 1911 (radium), Irène Joliot-Curie and Frédérik Joliot 1935 (artificial radioactivity), Hahn 1944 (uranium fission).

(ii) **Rays**: Röntgen 1901 (X-rays or Röntgen rays), Lorentz and Zeeman 1902 (influence of magnetism on radiation phenomena), Lenard 1905 (cathode rays), Michelson 1907 (spectroscopic experiments), Wien 1911 (radiation of heat), Sir William Henry Bragg and Sir William Lawrence Bragg 1913 (analysis of crystal structure by means of X-rays), Laue 1914 (diffraction of X-rays by crystals), Stark 1919 (splitting of spectral lines in electric fields), Franck and Hertz 1925 (observation of energy quantization in mercury atoms), Perrin 1926 (measurement of the size of atoms), Compton 1927 (Compton effect), Wilson 1927 (electron tracks in a cloud chamber), Sir Raman 1930 (scattering of light by atoms), Debye 1936 (investigation of molecular structure by diffraction of X-rays and electrons in gases), Hess 1936 (discovery of cosmic rays), Cherenkov 1958 (Cherenkov radiation), Bethe 1967 (energy production in stars), Gabor 1971 (holograph method), Townes, Basov, and Prochorov 1973 (laser), Ryle and Hewish 1974 (radio astronomy and pulsars), Penzias and Wilson 1978 (discovery of cosmic microwave radiation coming from the early universe), Chandrasekhar 1983 (theory of the structure and formation of stars), Fowler 1983 (theory of the formation of chemical elements in the universe), Hulse and Taylor 1993 (detection of gravitational waves coming from binary neutron stars).

(iii) **Structure of matter**: Sir Joseph John Thomson 1906 (conduction of electricity by gases), Einstein 1921 (photoelectric effect), Millikan 1923 (measurement of the charge of the electron), Chadwick 1935 (discovery of the neutron), Anderson 1936 (discovery of the positron), Fermi 1938 (experimental production of new radioactive elements and the discovery of nuclear reactions brought about by slow neutrons), Lawrence 1939 (construction of the cyclotron as particle accelerator), Stern 1943 (measurement of the magnetic moment of the proton), Yukawa 1949 (theoretical prediction of the existence of π-mesons in 1935), Powell 1950 (experimental discovery of π-mesons), Lamb 1955 (precision test of quantum electrodynamics by measurement of the $2s$ and $2p$ energy difference

[44] For more details, we refer to the literature about Nobel laureates summarized on page 929.

in the hydrogen atom), Kusch 1955 (measurement of the magnetic moment of the electron), Chamberlain and Segré 1959 (discovery of the antiproton), Hofstadter 1961 (experimental discovery of the internal structure of the proton), Richter and Ting 1976 (experimental discovery of the charm quark), Fitch and Cronin 1980 (experimental discovery of CP violation in weak interaction), Rubbia and van der Meer 1984 (experimental discovery of the $W^{\pm}$-bosons and the Z^0-boson in electroweak interaction), Lederman, Schwartz, and Steinberger 1988 (experimental discovery of the muon neutrino), Friedman, Kendall, and Taylor 1990 (experimental discovery of the quark structure of the proton by investigating deep inelastic scattering of electrons on protons), Perl 1995 (experimental discovery of the tau lepton), Reines 1995 (detection of the neutrino), Curl, Kroto, and Smalley 1996 (fullerenes).

(iv) **Quantum mechanics**: Planck 1918 (existence of energy quanta and foundation of quantum physics), Bohr 1922 (semiclassical model of the atom), de Broglie 1929 (wave nature of electrons), Heisenberg 1932 (foundation of quantum mechanics), Dirac and Schrödinger 1933 (new productive forms of atomic theory), Pauli 1945 (exclusion principle), Born 1954 (statistical interpretation of Schrödinger's wave function), Pauling 1954 (chemical bond), Wigner 1963 (symmetry principles), Goeppert-Mayer and Jensen 1963 (nuclear shell).

(v) **Quantum chemistry**: Fukui and Hoffmann 1981) (course of chemical reactions), Kohn 1998 (density functional theory), Pople 1998 (computational methods in quantum chemistry).

(vi) **The emergence of the Standard Model of elementary particle physics**: Yang and Lee 1957 (theory of parity violation in weak interaction), Feynman, Schwinger, and Tomonaga 1965 (theory of quantum electrodynamics), Gell-Mann 1969 (classification of elementary particles and interactions), Glashow, Salam, and Weinberg 1979 (theory of electroweak interaction as a gauge theory), Rubbia and van der Meer 1984 (experimental discovery of the intermediate vector bosons), 't Hooft and Veltman 1999 (renormalization of the gauge field theory for electroweak interaction), Gross, Politzer, and Wilczek 2004 (asymptotic freedom of quarks).

(vii) **Solid states and condensed matter**: Kammerlingh-Onnes 1913 (superconductivity), Bardeen, Brattain, and Shocklee 1956 (transistor), Mößbauer 1961 (recoilless gamma emission from nuclei), Landau 1962 (theory of phase transitions in condensed matter and liquid helium), Alfvén 1970 (magnetohydrodynamics and plasma physics), Néel 1970 (antiferromagnetism), Bardeen, Cooper, and Schrieffer 1972 (theory of Cooper pairs and superconductivity), Esaki and Giaever 1973 (tunnelling effects in semiconductors and superconductors), Josephson 1973 (super-current through a tunnel barrier and the Josephson effect), Wilson 1980 (critical phenomena, phase transitions, and the renormalization group technique for combining different scales in nature with each other), von Klitzing 1985 (experimental discovery of the quantum Hall effect), Müller and Bednorz 1987 (experimental discovery of high-temperature superconductivity in ceramic materials), Ramsey 1989 (hydrogen maser), Dehmelt and Paul 1989 (trapping of ions), de Gennes 1991 (liquid crystals and polymers), Chu, Tannoudji, and Phillips 1998 (cooling and trapping of atoms with laser light), Laughlin, Störmer, and Tsu 1999 (experimental discovery of the fractional quantum Hall effect), Cornell, Ketterle, and Wieman 2001 (experimental realization of Bose–Einstein condensation in dilute gases), Abrikosov, Ginzburg, and Leggett 2003 (superconductivity and superfluidity).

(viii) **Non-equilibrium thermodynamics**: Onsager 1968 (Onsager's law), Prigogine 1977 (dissipative structures).

(ix) **Computer technology**: Alferov and Kroemer 2000 (semiconductor heterostructures used in high-speed electronics and opto-electronics), Kirby 2000 (invention of integrated circuits).
(x) **Molecular biology**: Crick and Watson 1962 (double helix of DNA), Holley, Har Gobind Khorana, and Nirenberg 1968 (genetic code and protein synthesis), Delbrück, Hershey, and Luria 1969 (replication mechanism and the genetic structure of viruses), Diesenhofer, Huber, and Michel 1988 (determination of a 3-dimensional structure of a photosynthetic reaction), Fenn and Tanaka 2002 (development of soft desorption ionization methods for mass spectrometric analysis of biological macromolecules), Wüthrich 2002 (development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution).

1.5.2 The Fields Medal in Mathematics

The International Congress of Mathematicians (ICM) takes place every four years. In 1924, a resolution was adopted that at each ICM, two gold medals should be awarded to recognize outstanding mathematical achievement. Professor J. C. Fields, a Canadian mathematician who was secretary of the 1924 Congress, later donated funds establishing the medals which were named in his honor. Consistent with Field's wish that the awards recognize both existing work and the promise of future achievement, it was agreed to restrict the medal to mathematicians not over forty in the year of the congress. In 1966 it was agreed, because of great extension of mathematical research, up to four medals could be awarded at each Congress. The following list of Fields medallists reflects important progress in mathematics. The Fields medal has a very high reputation.[45]

- 1936 Oslo: Ahlfors (quasi-conformal maps), Douglas (minimal surfaces).
- 1950 Cambridge, Massachusetts: Laurent Schwartz (generalized functions), Selberg (elementary proof of the prime number theorem).
- 1954 Amsterdam: Kodaira (harmonic integrals in algebraic geometry), Serre (homotopy groups of spheres).
- 1958 Edinburgh: Roth (rational approximations to algebraic numbers), Thom (cobordism theory for manifolds).
- 1962 Stockholm: Hörmander (general theory of linear partial differential equations), Milnor (exotic spheres).
- 1966 Moscow: Atiyah (K-theory for vector bundles), Cohen (continuum hypothesis), Smale (proof of the Poincaré conjecture for n-dimensional spheres with $n \geq 5$, general structure of dynamical systems), Grothendieck (nuclear spaces, schemes in algebraic geometry).
- 1970 Nice: Baker (theory of transcendental numbers), Hironaka (blowing-up of singularities of algebraic varieties), Novikov (homology and homotopy theory), Thompson (group theory).
- 1974 Vancouver: Bombieri (analytic number theory and geometry of numbers), Mumford (Abelian varieties).

[45] We refer to M. Atiyah and D. Iagolnitzer (Eds.), Fields Medallists' Lectures, World Scientific, Singapore, 2003.

- 1978 Helsinki: Deligne (proof of the modified Riemann conjecture due to Weil for algebraic varieties over finite fields), Fefferman (singular integral operators, analytic functions of several variables), Margulis (structure of discrete Lie subgroups with fixed volume), Quillen (proof of the Serre conjecture on projective modules, cohomology of groups).
- 1982 Warsaw: Connes (structure of von Neumann algebras of type III), Thurston (hyperbolic structure of 3-dimensional manifolds), Shing-Tung Yau (positive mass theorem in general relativity, proof of the Calabi conjecture for Kähler manifolds).
- 1986 Berkeley: Donaldson (Yang–Mills equations and the differential topoloy of 4-dimensional manifolds), Faltings (proof of the Mordell conjecture for Diophantine equations), Freedman (proof of the Poincaré conjecture for 4-dimensional spheres).
- 1990 Kyoto: Drinfeld (quantum groups and Galois groups), Jones (von Neumann algebras and Jones polynomials in knot theory), Mori (classification of 3-dimensional algebraic varieties), Witten (supersymmetry and Morse theory, global anomalies, supersymmetric index theory, rigidity theorems for representations of Lie groups in string theory, spin structure, and a new approach to the positive mass theorem).
- 1994 Zurich: Bourgain (nonlinear Schrödinger equation, geometry of Banach spaces, ergodic theory, and analytic number theory), Piere-Louis Lions (nonlinear partial differential equations, Boltzmann equation, viscosity method for the Hamilton-Jacobi equation, compressible fluids, Hartee-Fock equation, anisotropic diffusion and image processing), Yoccoz (stability of dynamical systems), Zelmanov (Lie algebras and Jordan algebras).
- 1998 Berlin: Borcherds (representation of the monster group, modular forms), Gowers (geometry of Banach spaces, combinatorics), Kontsevich (Poisson structures and quantum deformations, equivalence of two models in quantum gravitation, effective knot invariants in topology), McMullen (complex dynamics and hyperbolic geometry), Wiles (special tribute for proving Fermat's last theorem).
- 2002 Beijing: Lafforgue (Langlands program for function fields, deep connections between number theory, analysis, and group representation theory), Voevodsky (proof of the Milnor conjecture in algebraic K-theory, motivic cohomology theory).

This list shows convincingly that the great achievements of mathematics in the 20th century are related to the efforts made by the great masters of mathematics and physics in the 18th and 19th century.

1.5.3 The Nevanlinna Prize in Computer Sciences

Since 1982, parallel to the Fields medal, the Nevanlinna prize has bee awarded for outstanding contributions to computer sciences.

- 1983 Warsaw: Tarjan (construction of highly effective algorithms).
- 1986 Berkeley: Valiant (complexity theory, random algorithms).
- 1990 Kyoto: Razborov (complexity of networks).
- 1994 Avi Widgerson (verification of proofs).
- 1998 Berlin: Shore (theory of quantum computation).
- 2002 Beijing: Sudan (probabilistic algorithms for checking the correctness of proofs).

1.5.4 The Wolf Prize in Physics

Every year the Israelitic Parliament (the Knesseth) confers the Wolf prize to outstanding scientists for their life-work in the fields of agriculture, arts, chemistry, mathematics, medicine, and physics. Here is the list of physicists who were awarded the Wolf prize.

- 1978 Wu (experimental discovery of parity violation in weak interaction in 1957).
- 1979 Uhlenbeck (experimental discovery of the electron spin in 1922, together with the late Goudsmith), Occhialini (experimental discovery of electron pair production and the charged pion).
- 1980 Fisher, Kadanoff, and Wilson (critical phenomena and phase transitions).
- 1981 Dyson, 't Hooft, and Weisskopf (quantum theory of fields).
- 1982 Lederman and Perl (discovery of the bottom quark).
- 1983/84 Hahn (spin echo), Hirsh (transmission electron microscope), Maiman (first operating laser).
- 1984/85 Herring and Nozieres (electrons in metals).
- 1986 Feigenbaum and Libchhaber (universal laws in turbulence; theory and experiments).
- 1987 Friedman, Rossi, and Giacconi (solar X-rays).
- 1988 Penrose and Hawking (necessity of cosmic singularities).
- 1989 not awarded.
- 1990 de Gennes and Thouless (complex condensed matter, liquid crystals, and disordered 2-dimensional systems).
- 1991 Goldhaber and Telegdi (weak interaction).
- 1992 Taylor (radio pulsar).
- 1993 Mandelbrot (fractals).
- 1994/95 Ginzburg (superconductivity and Ginzburg–Landau equation), Nambu (superconductivity, colored quarks).
- 1996 not awarded.
- 1996/97 Wheeler (black holes, quantum gravity, and nuclear fission).
- 1998 Aharonov and Berry (global quantum effects, Aharonov–Bohm effect, and Berry phase).
- 1999 Shechtman (experimental discovery of quasi-crystals).
- 2000 Davis and Koshiba (neutrino astronomy).
- 2001 not awarded.
- 2002/03 Leggett (superfluidity of the light helium isotope and macroscopic quantum phenomena), Halperin (two-dimensional melting, disordered systems, and strongly interacting electrons).
- 2004 Brout and Higgs (mass generation by local gauge symmetry in the world of subatomic particles).
- 2005 Kleppner (hydrogen maser, Rydberg atoms, and Bose–Einstein condensation).

1.5.5 The Wolf Prize in Mathematics

The following mathematicians were awarded the Wolf prize.[46]

[46] We refer to S. Chern and F. Hirzebruch (Eds.), Wolf Prize in Mathematics, Vols. 1, 2, World Scientific, Singapore, 2001.

- 1978 Gelfand (functional analysis, group representations, and seminal contributions to many parts of mathematics), Siegel (number theory, analytic functions of several variables, celestial mechanics).
- 1979 Leray (application of topological methods to differential equations), Weil (algebraic-geometric methods in number theory).
- 1980 Henri Cartan (algebraic topology, homological algebra, sheaf theory, and analytic functions of several variables), Kolmogorov (foundation of probability theory, stochastic processes, ergodic theory, Fourier analysis, and celestial mechanics).
- 1981 Ahlfors (geometric function theory), Zariski (commutative algebra and algebraic geometry).
- 1982 Whitney (algebraic topology and differential topology), Krein (functional analysis and its applications).
- 1983/84 Chern (global differential geometry), Erdös (discrete mathematics: number theory, combinatorics, graph theory, probability).
- 1984/85 Kodaira (complex manifolds and algebraic varieties), Hans Lewy (partial differential equations).
- 1986 Eilenberg (algebraic topology and homological algebra), Selberg (number theory, discrete groups, and automorphic functions).
- 1987 Ito (stochastic differential equations), Lax (linear and nonlinear partial differential equations, direct and inverse scattering theory, and shock waves).
- 1988 Hirzebruch (topology, algebraic geometry, and index theory), Hörmander (partial differential equations, pseudo-differential operators, Fourier integral operators).
- 1989 Calderon (singular integral operators and partial differential equations), Milnor (geometry, algebraic and differential topology).
- 1990 de Georgi (calculus of variations and partial differential equations).
- 1991 not awarded.
- 1992 Carleson (Fourier analysis, complex analysis, and quasi-conformal mappings), Thompson (finite groups).
- 1993 Gromov (global Riemannian geometry and symplectic geometry), Tits (algebraic groups, Tits buildings).
- 1994 Moser (stabilitity in Hamiltonian mechanics, celestial mechanics, and nonlinear partial differential equations).
- 1995/96 Langlands (number theory, automorphic forms, group representations, and the Langlands program on noncommutative class field theory), Wiles (proof of Fermat's last theorem).
- 1997 Joseph Keller (electromagnetic, acoustic, and optical wave propagation; fluid, solid, and quantum mechanics, statistical physics), Sinai (ergodic theory and statistical mechanics, dynamical systems).
- 1998 not awarded.
- 1999 Lovasz (combinatorics, theoretical computer sciences, combinatorial optimization), Stein (Fourier analysis and harmonic analysis).
- 2000 Bott (topology, differential geometry, Lie groups), Serre (topology, algebra, and algebraic geometry).
- 2001 Arnold (dynamical systems and singularity theory).
- 2002/03 Sato (hyperfunctions and microfunction theory, holonomic quantum field theory), Tate (algebraic number theory).
- 2004 not awarded.
- 2005 Margulis (theory of lattices in semi-simple Lie groups and striking applications of this to ergodic theory, representation theory, number theory, combinatorics, and measure theory), Novikov (algebraic and differential topology, algebraic-geometric methods in mathematical physics).

1.5.6 The Abel Prize in Mathematics

The Abel prize was founded in 2003 by the Norwegian government. This new prize is intended to play the role of the Nobel prize in mathematics.

- 2003 Serre (algebra, number theory, and topology).
- 2004 Atiyah and Singer (analysis, differential geometry, topology, and the Atiyah–Singer index theorem).
- 2005 Lax (linear and nonlinear partial differential equations, solitons and Lax pairs, scattering theory, shock waves).

1.6 The Emergence of Physical Mathematics – a New Dimension of Mathematics

At the International Congress of Mathematicians in Kyoto in 1990, the young physicist Edward Witten (Institute for Advanced Study in Princeton) was awarded the Fields medal in mathematics. In his laudation for Edward Witten, Sir Michael Atiyah emphasized the following:[47]

> The past decade has seen a remarkable renaissance in the interaction between mathematics and physics. This has been mainly due to the increasingly sophisticated mathematical models employed by elementary particle physicists, and the consequent need to use the appropriate mathematical machinery. In particular, because of the strongly non-linear nature of the theories involved, topological ideas and methods have played a prominent part.
>
> The mathematical community has benefited from this interaction in two ways. First, and more conventionally, mathematicians have been spurred into learning some of the relevant physics and collaborating with colleagues in theoretical physics. Second, and more surprisingly, many of the ideas emanating from physics have led to significant new insights in purely mathematical problems, and remarkable discoveries have been made in consequence. The main input from physics has come from quantum field theory. While the analytic foundations of quantum field theory have been intensively studied by mathematicians for many years, the new stimulus has involved the more formal (algebraic, geometric, topological) aspects.
>
> In all this large and exciting field, which involves many of the leading physicists and mathematicians in the world, Edward Witten stands clearly as the most influential and dominating figure. Although he is definitely a physicist his command of mathematics is rivalled by few mathematicians, and his ability to interpret physical ideas in mathematical form is quite unique. Time and again he has surprised the mathematical community by a brilliant application of physical insight leading to new and deep mathematical theorems.

[47] M. Atiyah and D. Iagolnitzer (Eds.), Fields Medallists' Lectures, 1997. Reprinted by permission of World Scientific Publishing Co. Pte. Ltd, Singapore.

In 1986, the American Mathematical Society invited mathematicians and physicists to a joined symposium devoted to *Mathematics: the Unifying Thread in Science.* In his quite remarkable speech, the physicist Steven Weinberg pointed out the following.[48]

> String theory is right now the hot topic of theoretical physics. According to this picture, the fundamental constituents of nature are not, in fact, particles, or even fields, but are instead little strings, little elementary rubber bands that go zipping around, each in its own state of vibration. In these theories what we call a particle is just a string in a particular state of vibration, and what we call a reaction among particles, is just the collision of two or more strings, each in its own state of vibration, forming a single joined string which then later breaks up, forming several independent strings, each again in its own mode of vibration.
>
> It seems like a strange notion for physicists to have come to after all these years of talking about particles and fields, and it would take too long to explain why we think this is not an unreasonable picture of nature, but perhaps I can summarize it in one sentence:
>
> *String theories incorporate gravitation.*
>
> In fact, not only do they incorporate it, you cannot have a string theory without gravitation. The graviton, the quantum of gravitational radiation, the particle which is transmitted when a gravitational force is exerted between two masses, is just the lowest mode of vibration of a fundamental closed string (closed meaning that it is a loop). Not only do they include and necessitate gravitation, but these string theories for the first time allow a description of gravitation on a microscopic quantum level which is free of mathematical inconsistencies.
>
> All other descriptions of gravity broke down mathematically, gave nonsensical results when carried to very small distances or very high energies. String theory is our first chance at a reasonable theory of gravity which extends from the very large down to very small and as such, it is natural that we are all agog over it. String theory itself has focused the attention of physicists on branches of mathematics that most of us weren't fortunate enough to have learned when we were students. You can easily see that a string (just think of a little bit of cord) travelling through space, sweeps out a two-dimensional surface.
>
> *A very convenient description of string theory is to say that it is the theory of these two-dimensional Riemann surfaces.*
>
> The theory of two-dimensional surfaces is remarkably beautiful. There are ways of classifying all possible two-dimensional surfaces according to their handles on them and the number of boundaries, which simply don't exist in any higher dimension. The theory of two-dimensional surfaces is a branch of mathematics that when you get into it is one of the loveliest things you can learn. It was developed in the 19th century, starting with Riemann,

[48] Notices Amer. Math. Soc. **33** (1986), 716–733 (reprinted with permission). For his fundamental contributions to the theory of the unified weak and electromagnetic interaction between elementary particles, Steven Weinberg was awarded the Nobel prize in physics in 1979. He wrote the standard textbook of modern quantum field theory, S. Weinberg, Quantum Field Theory, Vols. 1–3, Cambridge University Press.

and further developed by mathematicians working in the late 19th century motivated by problems in complex analysis, and then continuing into the 20th century.

There are mathematicians who have spent their whole lives working on this theory of two-dimensional surfaces, who have never heard of string theory (or at least not until very recently). Yet when the physicists started to figure out how to solve the dynamical problems of strings, and they realized what they had to do was to perform sums[49] over all possible two-dimensional surfaces in order to add up all the ways that reactions could occur, they found the mathematics just ready for their use, developed over the past 100 years.

String theory involves another branch of mathematics which goes back to group theory.

> *The equations which govern these surfaces have a very large group of symmetries, known as the conformal group.*

One description of these symmetries is in terms of an algebraic structure representing all the possible group transformations, which is actually infinite dimensional. Mathematicians have been doing a lot of work developing the theory of these *infinite dimensional Lie algebras* which underlie symmetry groups, again without a clear motivation in terms of physics, and certainly without knowing anything about string theory. Yet when the physicists started to work on it, there it was.

Speaking quite personally, I have found it exhilarating at my stage of life to have to go back to school and learn all this wonderful mathematics. Some of us physicists have enjoyed our conversations with mathematicians, in which

> *We beg them to explain things to us in terms we can understand.*

At the same time the mathematicians are pleased and somewhat bemused that we are paying attention to them after all these years. The mathematics department of the University of Texas at Austin now allows the physicists to use one of their lounges – which would have been unlikely in previous years.

Unfortunately, I must admit that there is no experimental evidence yet for string theory, and so, if theoretical physicists are spending more time talking to the mathematicians, they are spending less time talking to the experimentalists, which is not good.

1.7 The Seven Millennium Prize Problems of the Clay Mathematics Institute

At the Second World Congress of Mathematicians in Paris in 1900, in a seminal lecture, Hilbert formulated his famous 23 open problems.[50] The hundredth anniversary of Hilbert's lecture was celebrated in Paris, in the

[49] These sums correspond to Feynman functional integrals.

[50] See F. E. Browder (Ed.) Mathematical Developments Arising from Hilbert's Problems, Amer. Math. Soc., New York, 1976, and B. Yandell, The Honors Class: Hilbert's Problems and Their Solvers, Peters, Natick, Massachusetts, 2001.

"Amphithéatre" of the Collège de France, on May 24, 2000. The Scientific Advisory Board of the newly founded Clay Mathematics Institute (CMI) in Cambridge, Massachusetts, U.S.A., selected seven Millennium prize problems. The Scientific Advisory Board consists of Arthur Jaffe (director of the CMI, Harvard University, U.S.A), Alain Connes (Institut des Hautes Études Scientifiques (IHÉS) and Collège de France), Andrew Wiles (Princeton University, U.S.A.), and Edward Witten (Institute for Advanced Study, Princeton, U.S.A.). The CMI explains its intention as follows:

> Mathematics occupies a privileged place among the sciences. It embodies the quintessence of human knowledge, reaching into every field of human endeavor. The frontiers of mathematical understanding evolve today in deep and unfathomable ways. Fundamental advances go hand in hand with discoveries in all fields of science. Technological applications of mathematics underpin our daily life, including our ability to communicate thanks to cryptology and coding theory, our ability to navigate and to travel, our health and well-being, our security, and they also play a central role in our economy. The evolution of mathematics will remain a central tool to shaping civilization. To appreciate the scope of mathematical truth challenges the capabilities of the human mind.
>
> In order to celebrate mathematics in the new millennium, the CMI has named seven "Millennium prize problems". The Scientific Advisory Board of the CMI selected these problems, focusing on important classic questions that have resisted solution over the years. The Board of Directors of CMI designated a $ 7 million prize fund to these problems, with $ 1 million allocated to each.

The seven Millennium prize problems read as follows:

(i) The Riemann conjecture in number theory on the zeros of the Riemann zeta function and the asymptotics of prime numbers.
(ii) The Birch and Swinnerton–Dyer conjecture in number theory on the relation between the size of the solution set of a Diophantine equation and the behavior of an associated zeta function near the critical point $s = 1$.
(iii) The Poincaré conjecture in topology on the exceptional topological structure of the 3-dimensional sphere.
(iv) The Hodge conjecture in algebraic geometry on the nice structure of projective algebraic varieties.
(v) The Cook problem in computer sciences of deciding whether an answer that can be quickly checked with inside knowledge, may without such help require much longer to solve, no matter how clever a program we write.
(vi) The solution of the turbulence problem for viscous fluids modelled by the Navier–Stokes partial differential equations.
(vii) The rigorous mathematical foundation of a unified quantum field theory for elementary particles.

A detailed description of the problems can be found on the following Internet address:

`http://www.claymath.org/prize_problems/`

For a detailed discussion of the seven prize problems, we refer to K. Devlin, The Millennium Problems: the Seven Greatest Unsolved Mathematical Puzzles of Our Time, Basic Books, New York, 2002.

2. Phenomenology of the Standard Model for Elementary Particles

> *First Law of Progress in Theoretical Physics:* You will get nowhere by crunching equations.
> *Second Law:* Do not trust arguments based on the lowest order of perturbation theory.
> *Third Law:* You may use any degrees of freedom you like to describe a physical system, but if you use the wrong ones, you'll be sorry.[1]
>
> Steven Weinberg, 1983

For the motivation and convenience of the reader, let us sketch important basic ideas of elementary particle physics in this chapter. It is our philosophy that before studying a complex subject, one should know the main goals in nontechnical terms. As an introduction to particle physics, we recommend the textbooks by Nachtmann (1990), Coughlan and Dood (1991), Sibold (2001), and Seiden (2005). For the history of elementary particle physics, see the beautiful books by the two Nobel laureates Steven Weinberg (1983), (1995) and Martinus Veltman (2003).

In the 1960s and early 1970s, Gell-Mann and Fritzsch, Glashow, Salam, and Weinberg founded the so-called Standard Model of particle physics which is of fundamental importance for modern physics.[2] The Standard Model of particle physics is based on

- the principle of critical action, and
- the principle of local symmetry (gauge theory).

The Lagrangian density of the Standard Model will be thoroughly studied in Volume III. It turns out that the Standard Model of particle physics is governed by the same mathematical approach as Einstein's theory of general relativity on gravitation from 1915. The common mathematical tool is the theory of curvature in modern differential geometry (called gauge theory in physics). A survey on the Standard Model in particle physics and its possible generalizations along with a summary of the most important literature can be found in

[1] S. Weinberg, Why the renormalization group is a good thing. In: A. Guth, K. Huang, and R. Jaffe (Eds.), Asymptotic Realms of Physics: Essays in Honor of Francis Low, MIT Press, Cambridge, Massachusetts, 1983, pp. 1–19 (reprinted with permission).

[2] The Nobel prize in physics was awarded to Gell-Mann in 1969 and to Glashow, Salam, and Weinberg in 1979.

J.Rosner, Resource letter SM-1: The standard model and beyond, 2003. Internet: http://arXiv:hep-ph/0206176

This serves as a survey on modern physics including the following topics: quarks and leptons, the Higgs mass, CP violation, strong CP problem and massless axions, dynamics of heavy quarks, precision electroweak measurements, neutrino oscillations and neutrino masses, grand unification of interactions and extended gauge groups, proton decay, baryon asymmetry of the universe, supersymmetry, the riddle of dark matter and dark energy in the present universe, and string theory. Up-dated particle data are summarized by

Particle Data Group. Internet: http://pdg.lbl.gov

For the cosmic microwave background radiation and its information on the early universe see

NASA home page, WMAP, Internet: http://www.nasa.gov/home/

The WMAP (Wilkinson Microwave Anisotropy Probe) satellite experiment of NASA allows us to see the state of the universe at the age of 400 000 years after the Big Bang. In particular, the WMAP experiment shows that our universe is $13.7 \cdot 10^9$ years old. The five ages of our expanding universe starting from the Big Bang are studied in Adams and Laughlin (1997), (1999) (inside the physics of eternity). The five ages of our universe read as follows:

- the primordial era (from the Big Bang until the age of 10^5 years),
- the stelliferous era of the present universe (10^6–10^{14} years),
- the degenerate era (10^{15}–10^{39} years) (brown and white dwarfs, neutron stars and black holes dominate the universe),
- the black hole era (10^{40}–10^{100} years) (black holes dominate the universe), and
- the dark era ($> 10^{101}$ years).

It turns out that the expansion of our universe is accelerated. In the final dark era, the energy density of the universe goes to zero after the vaporization of the last black holes by Hawking radiation. However, it is possible that a new Big Bang is generated by large quantum fluctuations of the ground state (vacuum) of the universe. For modern astrophysics and cosmology, we refer to Schutz (2003) (phenomenology), Shore (2003) (tapestry of astrophysics), Börner (2003) (early universe), and Straumann (2004) (general relativity and astrophysics).

2.1 The System of Units

In this monograph, if not stated explicitly the contrary, we will use the international system of physical units, SI. This system is based on meter (length), second (time), kilogram (mass), Kelvin (temperature), and Coulomb (electric charge). Systems of units in physics are thoroughly discussed in the Appendix

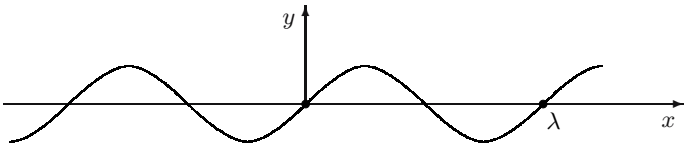


Fig. 2.1. Harmonic wave

(e.g., the SI system, the dimensionless Planck system, and the energetic system of units in particle physics based on powers of energy). In the Appendix, the interested reader also finds the very useful method of dimensional analysis along with tables about the values of the universal constants and the magnitudes of important physical quantities in nature. The knowledge of this is important for a deeper understanding of physical phenomena appearing in the real world.

2.2 Waves in Physics

Waves play a crucial role in physics, since they allow us to describe the transport of energy and information in nature.

2.2.1 Harmonic Waves

To begin with, let us explain the terminology used by physicists for describing waves. We are given the positive numbers ν, λ, a and the real number α. The real function

$$y = a \sin 2\pi \left(\frac{x}{\lambda} - \nu t + \frac{\alpha}{2\pi} \right) \tag{2.1}$$

is called a sinusoidal or harmonic wave. Let us discuss this. For fixed time t, the function $y = y(x)$ from (2.1) has the period λ with respect to the position variable x (Fig. 2.1). For fixed x, the function $y = y(t)$ has the time period $T = 1/\nu$. In order to avoid the factor 2π, physicists introduce both the angular frequency $\omega := 2\pi\nu$ and the wave number $\kappa := 2\pi/\lambda$. Then

$$\boxed{y = a \sin(\kappa x - \omega t + \alpha).} \tag{2.2}$$

If position x and time t are related by

$$\kappa x - \omega t = \text{const},$$

then $y = a \sin(\kappa x - \omega t + \alpha)$ remains constant during all times. This means that the wave propagates with the speed $v = \omega/\kappa = \nu\lambda$ from left to right. Physicists use the following terminology:

a	amplitude,
α	phase shift,
λ	wave length,
T	time period of oscillations,
$\nu = 1/T$	frequency (number of oscillations per second),
$\kappa = 2\pi/\lambda$	wave number,
$\omega = 2\pi\nu$	angular frequency,
$v = \omega/\kappa = \nu\lambda$	phase velocity of the wave.

The function[3]

$$\boxed{y = a\sin(\mathbf{kx} - \omega t + \alpha)} \tag{2.3}$$

describes a so-called sinusoidal plane wave with the wave vector $\mathbf{k}$ and the wave number $\kappa := |\mathbf{k}|$. Let $\mathbf{n}$ be a unit vector. If we set $\mathbf{x} = x\mathbf{n}$ and $\mathbf{k} = \kappa\mathbf{n}$, we obtain the function (2.2). Thus, the expression (2.3) corresponds to a wave propagating with speed v in direction of the vector $\mathbf{k}$ and angular frequency

$$\omega = v|\mathbf{k}|.$$

Using the famous Euler exponential formula

$$\mathrm{e}^{\mathrm{i}z} = \cos z + \mathrm{i}\sin z \qquad \text{for all} \quad z \in \mathbb{C}, \tag{2.4}$$

and letting $A := a\mathrm{e}^{\mathrm{i}\alpha}$, the wave (2.3) can be written as

$$\boxed{y = \Im(A\mathrm{e}^{\mathrm{i}(\mathbf{kx}-\omega t)}).}$$

2.2.2 Wave Packets

Group velocity. Typically, there exists a so-called dispersion relation

$$\omega = \omega(\kappa)$$

between the wave number κ and the angular frequency ω of a wave. In contrast to the phase velocity, $v := \omega(\kappa_0)/\kappa_0$, the number

$$\boxed{v_g := \omega'(\kappa_0)}$$

is called the group velocity corresponding to the wave number κ_0. We want to show that the group velocity can be regarded as the propagation speed of small wave packets (Fig. 2.2). In terms of physics, the group velocity is

[3] The symbol $\mathbf{kx}$ denotes the inner product between the vectors $\mathbf{k}$ and $\mathbf{x}$. Moreover, we set $\mathbf{x}^2 := \mathbf{xx}$.

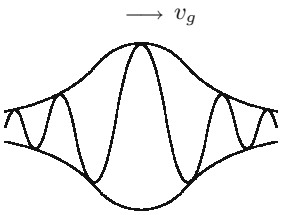


Fig. 2.2. Group velocity

decisive, since energy propagates with group velocity. By definition, a wave packet is obtained by the superposition of harmonic waves,

$$y(x,t) := \int_{-\infty}^{\infty} a(\kappa) \mathrm{e}^{\mathrm{i}(\kappa x - \omega(\kappa)t)} d\kappa, \qquad x,t \in \mathbb{R}.$$

We assume that the amplitude function $a = a(\kappa)$ is mainly concentrated on a small interval centered at the wave number κ_0. In order to understand the typical behavior of a wave packet, let us choose the amplitude function

$$a(\kappa) := C\mathrm{e}^{-\frac{1}{2}\left(\frac{\kappa-\kappa_0}{\Delta\kappa}\right)^2}$$

where the number $\Delta\kappa > 0$ is fixed, and C denotes a positive constant. Furthermore, we assume that the function $\omega = \omega(\kappa)$ is a quadratic polynomial,

$$\omega(\kappa) = \omega_0 + v_g(\kappa - \kappa_0) + \frac{b}{2}(\kappa - \kappa_0)^2.$$

Then

$$\omega(\kappa)t - \kappa x = (\omega_0 t - \kappa_0 x) + (\kappa - \kappa_0)(v_g t - x) + \frac{b}{2}(\kappa - \kappa_0)^2 t.$$

An explicit computation yields

$$\boxed{y(x,t) = A(x - v_g t) \cdot \mathrm{e}^{\mathrm{i}(\omega_0 t - \kappa_0 x)}, \qquad x,t \in \mathbb{R}.} \tag{2.5}$$

Explicitly, we have

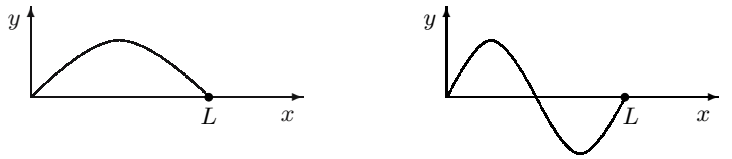
$$A(x) := \frac{C}{\sqrt{2\pi\Delta(t)}} \cdot \mathrm{e}^{-\frac{1}{2}\left(\frac{x}{\Delta(t)}\right)^2}$$

where

$$\Delta(t) := \sqrt{(\Delta\kappa)^2 - \mathrm{i}bt}.$$

Here, we take the principal part of the square root.[4] The wave packet $y = y(x,t)$ from (2.5) can be regarded as a modified harmonic wave which propagates with the phase velocity ω_0/κ_0, and the amplitude A propagates with the group velocity v_g.

[4] Let $z = |z|\,\mathrm{e}^{\mathrm{i}\varphi}$ where $-\pi < \varphi \le \pi$. Then, the principal part of the square root is given by $\sqrt{z} := \sqrt{|z|} \cdot \mathrm{e}^{\mathrm{i}\varphi/2}$.



(a) $n = 1$ (b) $n = 2$

Fig. 2.3. Standing waves

- If $b = 0$, then $\Delta(t) = \Delta\kappa$ for all times t. In this case, the amplitude is mainly concentrated on the interval $[\kappa_0 - \Delta\kappa, \kappa_0 + \Delta\kappa]$.
- If $b \neq 0$, then

$$|\Delta(t)|^4 = (\Delta\kappa)^4 + b^2t^2 \qquad \text{for all} \quad t \in \mathbb{R}.$$

Hence $|\Delta(t)| \to +\infty$ as $t \to +\infty$. This means that the wave packet is diffusing as times goes to infinity.

2.2.3 Standing Waves

The following investigations on the number of standing waves in a box were critically used by Planck in 1900 in order to derive his famous radiation law from Boltzmann statistics (see Sect. 2.3.2). Our goal is formula (2.6) below. Consider first the function

$$y = a \sin \kappa x$$

where $\kappa := n\pi/L$, $n = 1, 2, \ldots$ Since $y(0) = y(L) = 0$, this is called a standing wave on the interval $[0, L]$ (Fig. 2.3). Let $N(R)$ denote the number of standing waves on $[0, L]$ with the restriction $\kappa \leq R$. Approximately,

$$N(R) = \frac{R}{\pi/L} = \frac{RL}{\pi}.$$

Let us now pass to the 3-dimensional case. By definition, the function

$$y = A \sin \kappa_1 x \cdot \sin \kappa_2 y \cdot \sin \kappa_3 z$$

with $\kappa_j = n_j\pi/L$ and $n_j = 1, 2, ..., j = 1, 2, 3$ represents a 3-dimensional standing wave in the box $B := \{(x, y, z) \in \mathbb{R}^3 : 0 \leq x, y, z \leq L\}$ of volume $V = L^3$. Set $\mathbf{k} := (\kappa_1, \kappa_2, \kappa_3)$. Let $N(R)$ denote the number of standing waves in the box B with $|\mathbf{k}| \leq R$, and let $\mathcal{V}$ denote the volume of the positive quadrant $\{\mathbf{k} : |\mathbf{k}| \leq R,\ \kappa_1, \kappa_2, \kappa_3 \geq 0\}$ of a ball of radius R (Fig. 2.4). Approximately,

$$N(R) = \frac{\mathcal{V}}{(\pi/L)^3}.$$

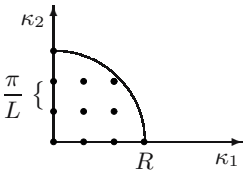


Fig. 2.4. Lattice points in a disc

Since $\mathcal{V} = \frac{1}{8} \cdot \frac{4}{3}\pi R^3$, we have $N(R) = R^3 V/6\pi^2$. From $N'(R) = R^2 V/2\pi^2$ we get

$$\Delta N = R^2 V \Delta R/2\pi^2$$

for the number of standing waves in the box with $R \leq |\mathbf{k}| \leq R + \Delta R$. Passing to the angular frequency $\omega = c|\mathbf{k}|$, we finally obtain that

$$\boxed{\Delta N = \frac{V\omega^2}{2\pi^2 c^3}\Delta\omega} \tag{2.6}$$

is the number of standing waves in the box B of volume V corresponding to the interval $[\omega, \omega + \Delta\omega]$ of angular frequency.

2.2.4 Electromagnetic Waves

The Maxwell equations for the electric field vector $\mathbf{E} = \mathbf{E}(\mathbf{x}, t)$ and the magnetic field vector $\mathbf{B} = \mathbf{B}(\mathbf{x}, t)$ in vacuum read as

$$\begin{aligned} \operatorname{div} \mathbf{E} &= 0, & \operatorname{div} \mathbf{B} &= 0, \\ \mathbf{curl}\, \mathbf{E} &= -\dot{\mathbf{B}}, & c^2\, \mathbf{curl}\, \mathbf{B} &= \dot{\mathbf{E}}, \end{aligned} \tag{2.7}$$

where c is the velocity of light in a vacuum. The dot denotes the partial time derivative. Let $\mathbf{i}, \mathbf{j}, \mathbf{n}$ be a right-handed orthonormal vector basis. An explicit computation shows that the harmonic waves

$$\boxed{\mathbf{E} = a\cos(\mathbf{kx} - \omega t) \cdot \mathbf{i} + b\sin(\mathbf{kx} - \omega t + \alpha) \cdot \mathbf{j}, \qquad \mathbf{B} = \frac{\mathbf{n} \times \mathbf{E}}{c}}$$

along with $\mathbf{k} := \kappa\mathbf{n}$ and the dispersion relation $\omega = c\kappa$ represent a solution of the Maxwell equations. This electromagnetic wave (light wave) propagates in direction of the unit vector $\mathbf{n}$. From $c_g = \omega'(\kappa) = c$ we obtain that the group velocity of light is equal to the phase velocity. Thus, the light is polarized. This means that the amplitude of the harmonic wave $\mathbf{E} = \mathbf{E}(\mathbf{x}, t)$ is a vector. Explicitly, we have the following possibilities:

- If $a^2 + b^2 \neq 0$, the polarization is called elliptic.

- If $a = b$ and $a \neq 0$, the polarization is called circular.
- If $a \neq 0,\ b = 0$ or $a = 0,\ b \neq 0$, the polarization is called linear.

In a bounded region Ω of volume V, the energy E and the momentum vector $\mathbf{p}$ of the electromagnetic field are given by

$$E = \int_\Omega \eta \, d^3x, \qquad \mathbf{p} = \int_\Omega (\mathbf{D} \times \mathbf{B}) d^3x, \qquad \eta := \frac{1}{2}(\mathbf{ED} + \mathbf{HB}),$$

where $\mathbf{D} = \varepsilon_0 \mathbf{E}$ and $\mathbf{B} = \mu_0 \mathbf{H}$. Here, ε_0 (resp. μ_0) is the electric (resp. magnetic) field constant of a vacuum. Moreover, $c^2 = 1/\varepsilon_0\mu_0$. In addition, the energy current density vector is equal to $\mathbf{S} = \mathbf{E} \times \mathbf{H}$. For our light wave, we get the energy density $\eta = a^2 + b^2$, and hence $E = \eta V$ as well as $\mathbf{S} = c\eta\mathbf{n}$. This yields the crucial relation $E^2 = c^2\mathbf{p}^2$. Hence

$$\boxed{E = c|\mathbf{p}|}$$

which corresponds to the Einstein relation $E^2 = m_0^2c^4 + c^2\mathbf{p}^2$ with vanishing rest mass, $m_0 = 0$.

2.2.5 Superposition of Waves and the Fourier Transform

> The profound study of nature is the most fertile source of mathematics.
> Joseph Fourier (1768–1830)

The Fourier transform represents a fundamental duality between time and frequency (resp. time and energy).

The frequency space. Let $a : \mathbb{R} \to \mathbb{C}$ be a measurable (e.g., continuous) function with $\int_{\mathbb{R}} |a(\omega)| d\omega < \infty$. Then, the integral

$$\boxed{f(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} a(\omega) \, \mathrm{e}^{-\mathrm{i}\omega t} d\omega}$$

exists for all times $t \in \mathbb{R}$. The function f represents the superposition of harmonic waves. The amplitude function $a = a(\omega)$ is called the Fourier transform of the time-dependent function f with respect to angular frequency ω. Properties of the Fourier transform will be considered in Sect. 10.3.3 on page 535 by using the space $\mathcal{S}(\mathbb{R})$ of smooth complex-valued functions on the real line which are rapidly decreasing at infinity. For example, if $f \in \mathcal{S}(\mathbb{R})$, then we get $a \in \mathcal{S}(\mathbb{R})$ along with the inverse transformation formula

$$a(\omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(t) \, \mathrm{e}^{\mathrm{i}\omega t} dt \qquad \text{for all} \quad \omega \in \mathbb{R}.$$

In particular, if we choose the Gaussian function $f(t) := \mathrm{e}^{-\frac{t^2}{2}}$ for all $t \in \mathbb{R}$, then

$$a(\omega) = f(\omega) \qquad \text{for all} \quad \omega \in \mathbb{R}.$$

This means that the Gaussian function is distinguished by the fact that it is invariant under the Fourier transform. This transform was extensively used by Fourier in his treatise *Théorie analytique de la chaleur* (analytic heat theory) from 1822.

The Fourier transform plays a fundamental role in quantum field theory.

In particular, Feynman's elegant approach to quantum physics via path integrals is closely related to Fourier's heat kernel method for solving the heat equation (see page 588). Furthermore,

- the Dirac calculus, and
- von Neumann's spectral theory for self-adjoint operators in Hilbert space

represent far-reaching generalizations of the Fourier method to quantum field theory.

The energy space. Let us use the transformation

$$E = \hbar\omega$$

in order to pass from the frequency space to the energy space. From the physical point of view, E is the energy of a photon which has the frequency $\nu = \omega/2\pi$. Setting $b(E) := a(E/\hbar)/\sqrt{\hbar}$, we get

$$\boxed{f(t) = \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-\mathrm{i}Et/\hbar}}{\sqrt{2\pi\hbar}}\, b(E)dE \qquad \text{for all} \quad t \in \mathbb{R}} \tag{2.8}$$

along with the inverse transformation formula

$$\boxed{b(E) = \int_{-\infty}^{\infty} \frac{\mathrm{e}^{\mathrm{i}Et/\hbar}}{\sqrt{2\pi\hbar}}\, f(t)dt \qquad \text{for all} \quad E \in \mathbb{R}.} \tag{2.9}$$

In terms of quantum physics, the time-dependent function $t \mapsto f(t)$ is represented as a superposition of the functions

$$t \mapsto \frac{\mathrm{e}^{-\mathrm{i}Et/\hbar}}{\sqrt{2\pi\hbar}}$$

which correspond to quantum states of energy E.

For historical reasons, engineers, mathematicians, and physicists use different definitions of the Fourier transform.

This will be discussed on page 538. Our definition (2.9) fits the needs of quantum physics. In particular, the normalization factor $\frac{1}{\sqrt{2\pi\hbar}}$ guarantees that the map $f \mapsto b$ is a unitary operator on the Hilbert space $L_2(\mathbb{R})$. Furthermore, in

terms of the elegant Dirac calculus, formulas (2.8) and (2.9) can be written symmetrically as

$$\langle t|f\rangle = \int_{\mathbb{R}} \langle t|E\rangle\langle E|f\rangle dE$$

and

$$\langle E|f\rangle = \int_{\mathbb{R}} \langle E|t\rangle\langle t|f\rangle dt,$$

respectively, where we set $\langle t|f\rangle := f(t)$ and $\langle E|f\rangle := b(E)$, as well as

$$\langle t|E\rangle = \langle E|t\rangle^{\dagger} := \frac{\mathrm{e}^{-\mathrm{i}Et/\hbar}}{\sqrt{2\pi\hbar}}.$$

The choice of the normalization factor $\frac{1}{\sqrt{2\pi\hbar}}$ from (2.9) also yields the appropriate continuous orthonormality relation

$$\langle E|E_0\rangle = \int_{\mathbb{R}} \langle E|t\rangle\langle t|E_0\rangle dt = \frac{1}{2\pi\hbar}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}(E-E_0)t/\hbar}\, dt = \delta(E-E_0)$$

for all $E, E_0 \in \mathbb{R}$. Here, the symbol δ denotes the Dirac delta function. This will be discussed in Sect. 12.1ff.

The local inversion theorem. As a preparation for the Laplace transform below, let us recall the classical main theorem on the inversion of the Fourier transform. The key formulas are given by (2.9) and the following modification of the superposition formula (2.8):

$$f(t_0) = \quad PV\int_{-\infty}^{\infty} \frac{\mathrm{e}^{-\mathrm{i}Et_0/\hbar}}{\sqrt{2\pi\hbar}}\, b(E)dE. \tag{2.10}$$

Here, the symbol $PV\int_{-\infty}^{\infty}\ldots$ stands for

$$PV\int_{-\infty}^{\infty} g(E)dE := \lim_{N\to\infty}\int_{-N}^{N} g(E)dE.$$

We assume that this finite limit exists which is called Cauchy's 'principal value' of the integral.

Theorem 2.1 *Suppose that the function $f : \mathbb{R} \to \mathbb{C}$ is continuous except for a finite number of points and $\int_{-\infty}^{\infty} |f(t)|dt < \infty$. Then, the integral* (2.9) *exists.*

If, in addition, the function f is differentiable in an open neighborhood of the point t_0, then the inverse formula (2.10) *holds true.*

The classical proof can be found in Doetsch (1956), Vol. 1, Sect. 4.2.

2.2.6 Damped Waves, the Laplace Transform, and Dispersion Relations

> In his 1812 treatise *Théorie analytique des probabilités*, Pierre-Simon Laplace (1749–1824) extensively used special integrals of the type
>
> $$F(s) := \int_0^\infty f(t)\mathrm{e}^{-st}dt$$
>
> in order to solve both differential equations and difference equations and to obtain asymptotic expansions. In modern terminology, these integrals correspond to the Laplace transform. Historically, in 1737 Euler was the first who applied such integrals for solving differential equations. However, it seems that Laplace did not know Euler's results.
>
> Gustav Doetsch
> *Handbook of the Laplace Transform*, Birkhäuser, Basel, 1956.

Damped wave. In particle accelerators, physicists observe unstable particles of finite lifetime. Such particles are also called resonances. Let us model this by using the so-called damped wave

$$\boxed{y(t) = A\mathrm{e}^{-\gamma_0 t}\cdot \mathrm{e}^{-\mathrm{i}\omega_0 t} = A\mathrm{e}^{-\mathrm{i}(\omega_0 - \mathrm{i}\gamma_0)t}, \qquad t\in\mathbb{R}} \tag{2.11}$$

where the real number ω_0 and the positive number γ_0 are called the angular frequency and the damping constant, respectively. The nonzero complex number A is called the amplitude. Obviously,

$$\lim_{t\to+\infty} y(t) = 0,$$

that is, the wave vanishes at large time. The number

$$\boxed{t_{\mathrm{mean}} := \frac{1}{\gamma_0}}$$

is called the mean lifetime of the damped wave. Observe that during the time interval $[0, t_{\mathrm{mean}}]$ the damping function $t\mapsto \mathrm{e}^{-\gamma_0 t}$ decreases from $\mathrm{e}^0 = 1$ to $\mathrm{e}^{-1}\sim\frac{1}{2}$. Explicitly, if $A = a\mathrm{e}^{\mathrm{i}\alpha}$ with $a>0$ and $\alpha\in\mathbb{R}$, then

$$y(t) = a\mathrm{e}^{-\gamma_0 t}\cdot(\cos(\omega_0-\alpha) - \mathrm{i}\sin(\omega_0-\alpha)), \qquad t\in\mathbb{R}.$$

Summarizing, we obtain that

> *Damped waves can be described by complex frequencies* $\omega_0 - \mathrm{i}\gamma_0$. *Here, the imaginary part is related to the mean lifetime of the wave by* $t_{\mathrm{mean}} = 1/\gamma_0$.

Truncated damped wave. By definition, the Heaviside function θ switches on physical processes at time $t=0$. Explicitly, we set $\theta(t) := 1$ if $t\geq 0$ and $\theta(t) := 0$ if $t<0$. The function

$$y(t) = A\theta(t)\mathrm{e}^{-\gamma_0 t} \cdot \mathrm{e}^{-\mathrm{i}\omega_0 t}, \qquad t \in \mathbb{R} \tag{2.12}$$

describes a damped wave which is switched on at time $t = 0$. Roughly speaking, the notion of truncated damped wave is related to causality. We will show below that this corresponds to analyticity properties of the Fourier transform.

Generally, there exists a deep relation between causality and analyticity in physics.

The dispersion relation. The Fourier transform of the damped wave (2.11) reads as[5]

$$a(\omega) = \int_{-\infty}^{\infty} \left(A\mathrm{e}^{-\gamma_0 t}\ \mathrm{e}^{-\mathrm{i}\omega_0 t}\right) \mathrm{e}^{\mathrm{i}\omega t}\ dt.$$

This integral does not exist in the classical sense because of $\mathrm{e}^{-\gamma_0 t} \to +\infty$ as $t \to -\infty$. The situation changes completely if we replace the damped wave by a truncated damped wave. Then

$$a(\omega) = \int_{-\infty}^{\infty} (A\theta(t)\mathrm{e}^{-\gamma_0 t}\mathrm{e}^{-\mathrm{i}\omega_0 t}\ \mathrm{e}^{\mathrm{i}\omega t})\ dt = \int_0^{\infty} A\mathrm{e}^{-\gamma_0 t + \mathrm{i}(\omega-\omega_0)t} dt.$$

Explicitly, for all $\omega \in \mathbb{R}$,

$$a(\omega) = \lim_{T\to+\infty} A \cdot \frac{\mathrm{e}^{-\gamma_0 T}\mathrm{e}^{\mathrm{i}(\omega-\omega_0)T} - 1}{\mathrm{i}(\omega-\omega_0) - \gamma_0} = \frac{\mathrm{i}A}{\omega - (\omega_0 - \mathrm{i}\gamma_0)}.$$

Analytic continuation yields the meromorphic function

$$\boxed{a(\omega) = \frac{\mathrm{i}A}{\omega - (\omega_0 - \mathrm{i}\gamma_0)}, \qquad \omega \in \mathbb{C} \setminus \{\omega_0 - \mathrm{i}\gamma_0\}}$$

on the complex ω-plane which has precisely one singularity at the point $\omega_0 - \mathrm{i}\gamma_0$ in the open lower half-plane. This singularity encodes both the angular frequency of the damped wave, ω_0, and the lifetime, $t_{\mathrm{mean}} = 1/\gamma_0$. By Theorem 2.1 on page 88, the inverse Fourier transform yields

$$A\theta(t)\mathrm{e}^{-\gamma_0 t}\mathrm{e}^{-\mathrm{i}\omega_0 t} = \frac{1}{2\pi}\ PV \int_{-\infty}^{\infty} a(\omega)\mathrm{e}^{-\mathrm{i}\omega t}\ d\omega, \qquad t \in \mathbb{R} \setminus \{0\}.$$

In particular, the Laplace transform $\omega \mapsto a(\omega)$ is holomorphic in the closed upper half-plane. Moreover, the real and imaginary part of $a(\omega)$ satisfy the relations

$$\boxed{\Im(a(\omega)) = \frac{1}{\pi}\ PV \int_{\mathbb{R}} \frac{\Re(a(\xi))}{\omega - \xi}\ d\xi \qquad \text{for all} \quad \omega \in \mathbb{R}} \tag{2.13}$$

[5] To simplify notation, we rescale the Fourier transform by using the replacement $a(\omega) \Rightarrow a(\omega)/\sqrt{2\pi}$.

and

$$\Re(a(\omega)) = -\frac{1}{\pi}\, PV \int_{\mathbb{R}} \frac{\Im(a(\xi))}{\omega - \xi}\, d\xi \qquad \text{for all} \quad \omega \in \mathbb{R}. \tag{2.14}$$

The integral is to be understood in the sense of Cauchy's principal value

$$PV \int_{\mathbb{R}} \frac{g(\xi)d\xi}{\omega - \xi} := \lim_{\varepsilon \to +0} \left(\int_{-\infty}^{\omega - \varepsilon} \frac{g(\xi)d\xi}{\omega - \xi} + \int_{\omega + \varepsilon}^{\infty} \frac{g(\xi)d\xi}{\omega - \xi} \right)$$

with respect to the singularity of the integrand at the point $\omega \in \mathbb{R}$. This has the following important consequence. If we know, say, the real part $\Re(a(\omega))$ on the real line, then we get the imaginary part $\Im(a(\omega))$ on the real line, by (2.13). Furthermore, the classical Cauchy residue formula

$$a(\omega) = \frac{1}{2\pi \mathrm{i}} \lim_{R \to +\infty} \int_{-R}^{R} \frac{a(\xi)d\xi}{\xi - \omega}, \qquad \Im(\omega) > 0$$

yields the values of $a(\omega)$ on the open upper-half plane.[6] Using analytic continuation, we then get the function $a = a(\omega)$ on the complex plane.

> *Summarizing, for a truncated damped wave, the imaginary part of the Laplace transform $a = a(\omega)$ on the real line knows all about the meromorphic function a on the complex plane.*

In mathematics, the map $\Im(a) = H\Re(a)$ given by (2.13) is called the Hilbert transform (see Sect. 11.9.3 on page 663). In physics, the relations (2.13) and (2.14) are called dispersion relations (see Sect. 12.4.5 on causality and analyticity).

The basic trick of the Laplace transformation. Let us now consider the critical case where $\gamma_0 \leq 0$. The trick is to introduce a damping factor $\mathrm{e}^{-\gamma t}$. In fact, in order to get a convergent integral, we change the Laplace transform by setting

$$a(\omega + \mathrm{i}\gamma) := \int_0^{\infty} \left(A\mathrm{e}^{-(\gamma_0 + \gamma)t}\, \mathrm{e}^{-\mathrm{i}\omega_0} \right) \mathrm{e}^{\mathrm{i}\omega t}\, dt.$$

To get a damping factor, we choose the real number γ in such a way that $\gamma_0 + \gamma > 0$. This can be written as

$$a(\omega + \mathrm{i}\gamma) = \int_0^{\infty} \left(A\mathrm{e}^{-\gamma_0 t} \mathrm{e}^{-\mathrm{i}\omega_0 t} \right) \mathrm{e}^{\mathrm{i}(\omega + \mathrm{i}\gamma)t}\, dt.$$

Explicitly,

$$a(\omega + \mathrm{i}\gamma) = \frac{\mathrm{i}A}{\omega + \mathrm{i}\gamma - (\omega_0 - \mathrm{i}\gamma_0)}$$

for all real numbers ω, γ with $\gamma > -\gamma_0$. By the inverse Fourier transform,

[6] This follows as in Problem 12.1 on page 732.

$$A\theta(t)\mathrm{e}^{-(\gamma_0+\gamma)t}\mathrm{e}^{-\mathrm{i}\omega_0} = \frac{1}{2\pi}\, PV \int_{-\infty}^{\infty} a(\omega+\mathrm{i}\gamma)\mathrm{e}^{-\mathrm{i}\omega t}\, d\omega, \qquad t \in \mathbb{R}\setminus\{0\}.$$

Hence

$$A\theta(t)\mathrm{e}^{-\gamma_0 t}\mathrm{e}^{-\mathrm{i}\omega_0 t} = \frac{1}{2\pi}\, PV \int_{-\infty}^{\infty} a(\omega+\mathrm{i}\gamma)\mathrm{e}^{-\mathrm{i}(\omega+\mathrm{i}\gamma)t}\, d\omega, \qquad t \in \mathbb{R}\setminus\{0\}.$$

The energy space. Setting $E = \hbar\omega$, let us pass over to the space of complex energy $\mathcal{E} = E - \mathrm{i}\Gamma$. The damped wave (2.12) can be equivalently written as

$$y(t) = A\mathrm{e}^{-\mathrm{i}\mathcal{E}_0 t/\hbar} \qquad t \in \mathbb{R}$$

where $\mathcal{E}_0 = E_0 - \mathrm{i}\Gamma$ with $E_0 := \hbar\omega_0$ and $\Gamma_0 := \hbar\gamma_0$. The Laplace transform (2.16) reads as

$$(\mathcal{L}y)(\mathcal{E}) = \frac{\mathrm{i}\hbar A}{\mathcal{E}-\mathcal{E}_0} \tag{2.15}$$

for all $\mathcal{E} \in \mathbb{C}$ with $\Im(\mathcal{E}) > \Im(\mathcal{E}_0)$.

The Laplace transform in the energetic space. The two key formulas read as

$$\boxed{(\mathcal{L}f)(\mathcal{E}) := \int_0^{\infty} \mathrm{e}^{\mathrm{i}\mathcal{E}t/\hbar} f(t)\, dt \qquad \text{for all} \quad \mathcal{E} \in \mathcal{H}_{\gamma_0}} \tag{2.16}$$

and

$$f(t) = \frac{1}{2\pi\hbar}\, PV \int_L \mathrm{e}^{-\mathrm{i}\mathcal{E}t/\hbar}(\mathcal{L}f)(\mathcal{E})\, d\mathcal{E} \qquad \text{for all} \quad t > 0. \tag{2.17}$$

The function $\mathcal{L}$ is called the Laplace transform of the function f in the energy space. Furthermore, for given real number γ_0 we write $f \in \mathcal{K}_{\gamma_0}$ iff[7] the continuous function $f : [0,\infty[\to \mathbb{C}$ satisfies the growth condition

$$|f(t)| \le \text{const} \cdot \mathrm{e}^{-\gamma_0 t} \qquad \text{for all} \quad t \ge 0.$$

Moreover, we introduce the open half-plane $\mathcal{H}_{\gamma_0} := \{\mathcal{E} \in \mathbb{C} : \Im(\mathcal{E}) > -\hbar\gamma_0\}$.

Theorem 2.2 *Let $f \in \mathcal{K}_{\gamma_0}$. Then, the Laplace transform $\mathcal{L}f$ given by* (2.16) *is a holomorphic function on the open half-plane $\mathcal{H}_{\gamma_0}$.*

The inverse transform $f = \mathcal{L}^{-1}F$ is given by (2.17)*. Here, we integrate over a fixed, but otherwise arbitrary line L in $\mathcal{H}_{\gamma_0}$ which is parallel to the real axis.*

[7] In mathematics, 'iff' stands for 'if and only if'.

Using the same argument as for damped waves above, the inversion formula (2.17) follows easily from the inversion formula for the Fourier transform (2.10).[8] It is crucial for physics that the Laplace transform $\mathcal{L}f$ is holomorphic in an open half-plane of the complex energy space. As we will show later on, after carrying out an analytic continuation, the singularities of the Laplace transform $\mathcal{L}$ are of great physical importance. They correspond to both the energies of bound states and their lifetimes.

The Laplace transform in the frequency space. This is obtained from (2.16) by using the replacement $E \Rightarrow \omega$ and setting $\hbar := 1$. In terms of physics, this corresponds to using the energetic system of units.

Three basic rules. The following rules allow us to solve differential equations and to get information about the solutions in terms of superpositions of harmonic waves or damped waves. Moreover, the resonance effects become transparent. To simplify notation, let us set $\hbar := 1$ in (i)-(iii) below (i.e., we work in the energetic system).

(i) *Wave rule.* Let $k = 0, 1, \ldots$. For complex energy $\mathcal{E}_0 = E_0 - \mathrm{i}\Gamma_0$, the Laplace transform of the function

$$f(t) := t^k \mathrm{e}^{-\mathrm{i}\mathcal{E}_0 t}, \quad t \geq 0$$

is equal to

$$(\mathcal{L}f)(\mathcal{E}) = \frac{\mathrm{i}^{k+1} k!}{(\mathcal{E} - \mathcal{E}_0)^{k+1}}$$

for all complex energies $\mathcal{E}$ with $\Im(\mathcal{E}) > \Im(\mathcal{E}_0)$.

(ii) *Derivative rule.* We are given the function $f \in \mathcal{K}_{\gamma_0}$. Moreover, we assume that the time derivatives $\dot{f}, \ddot{f}$ also lie in $\mathcal{K}_{\gamma_0}$. Then

$$\boxed{\big(\mathcal{L}\dot{f}\big)(\mathcal{E}) = -\mathrm{i}\mathcal{E} \cdot \mathcal{L}(f)(\mathcal{E}) - f(+0)}$$

and

$$\big(\mathcal{L}\ddot{f}\big)(\mathcal{E}) = (-\mathrm{i}\mathcal{E})^2 \mathcal{L}(f)(\mathcal{E}) + \mathrm{i}\mathcal{E} \cdot f(+0) - \dot{f}(+0)$$

for all $\mathcal{E} \in \mathbb{C}$ with $\Im(\mathcal{E}) > -\gamma_0$. This way, the Laplace transform converts differentiation into multiplication.[9]

(iii) *Convolution rule.* Let $f, g \in \mathcal{K}_{\gamma_0}$. Set

$$\big(f * g\big)(t) := \int_0^t f(t-\tau) g(\tau)\, d\tau \qquad \text{for all} \quad t \geq 0.$$

The function $f * g$ is called the causal convolution between the functions f and g. Then

[8] See the standard textbooks by Widder (1944) and Doetsch (1956). As an introduction to the Laplace transform, we recommend Zeidler (2004).

[9] As usual, the symbol $g(+0)$ stands for the limit $\lim_{t \to +0} g(t)$.

$$\mathcal{L}(f * g) = \mathcal{L}f \cdot \mathcal{L}g \qquad \text{on} \quad \mathcal{H}_{\gamma_0}.$$

This tells us that the product in the energy space corresponds to the causal convolution in the original time space.

Proof. Ad (i). If $k = 0$, the claim follows from (2.15). For $k = 1, 2, \ldots$, differentiate successively equation (2.16) with respect to the parameter $\mathcal{E}_0$.

Ad (ii). Integration by parts yields

$$\int_0^T \mathrm{e}^{\mathrm{i}\mathcal{E}t} \dot{f}(t)\, dt = \mathrm{e}^{\mathrm{i}\mathcal{E}t} f(t)\Big|_0^T - \mathrm{i}\mathcal{E} \int_0^T \mathrm{e}^{\mathrm{i}\mathcal{E}t} f(t) dt.$$

Finally, apply the limit $T \to +\infty$. Observe that

$$|\mathrm{e}^{\mathrm{i}\mathcal{E}T} f(T)| \leq \text{const} \cdot \mathrm{e}^{-(\Im(\mathcal{E})+\gamma_0)T} \to 0 \qquad \text{as} \quad T \to +\infty$$

if $\Im(\mathcal{E}) > -\gamma_0$. The same argument applied to $\dot{f}$ yields

$$\mathcal{L}(\ddot{f}) = -\mathrm{i}\mathcal{E} \cdot \mathcal{L}(\dot{f}) - \dot{f}(+0).$$

Ad (iii). Set $h := f * g$. It follows from $\int_0^t \mathrm{e}^{\tau} \mathrm{e}^{t-\tau}\, d\tau = t\mathrm{e}^t$ that $h \in \mathcal{K}_{\gamma_0 - \varepsilon}$ for all $\varepsilon > 0$. Define $f(t) = g(t) = h(t) := 0$ if $t < 0$. Then

$$h(t) := \int_{-\infty}^{\infty} f(t-\tau) g(\tau)\, d\tau, \quad t \in \mathbb{R}.$$

Using the substitution $s := t - \tau$,

$$\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\mathcal{E}t} h(t)\, dt = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\mathcal{E}s} f(s)\, ds \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\mathcal{E}\tau} g(\tau)\, d\tau.$$

□

2.2.7 The Response Function, the Feynman Propagator, and Causality

We want to use the Laplace transform in order to construct the prototype of the Feynman propagator. The idea is to describe mathematically the response of a physical system to an external force which is switched on at the initial time $t = 0$.

The causal initial-value problem for the harmonic oscillator. As a simple example, let us consider the following initial-value problem

$$\begin{aligned} &m\ddot{x}(t) = -\kappa_0 x(t) + F(t), \qquad t > 0, \\ &x(+0) = x_0, \quad \dot{x}(+0) = v_0 \end{aligned} \tag{2.18}$$

along with $x(t) = 0$ for all $t < 0$. Here, the function $x = x(t)$ describes the motion of a classical particle of mass $m > 0$ on the real line under the influence of

- the repulsive force $-\kappa_0 x$ with the coupling constant $\kappa_0 > 0$, and
- the external force F which is switched on at the initial time $t = 0$, that is, $F(t) = 0$ for all $t < 0$. We assume that the given function $F : \mathbb{R} \to \mathbb{R}$ is continuous on the time interval $[0, \infty[$.

We are given the initial position x_0 and the initial velocity v_0 of the particle at time $t = 0$. We are looking for a solution $x : \mathbb{R} \to \mathbb{R}$ which is differentiable twice on the open time interval $]0, \infty[$. In addition, we postulate that the functions x and $\dot{x}$ can be continuously extended to the closed time interval $[0, \infty[$. We set $\omega_0 := \sqrt{\kappa_0/m}$.

Theorem 2.3 *The initial-value problem* (2.18) *has the unique solution*

$$x(t) = \theta(t) \left(x_0 \cos \omega_0 t + \frac{v_0}{\omega_0} \cdot \sin \omega_0 t + \frac{1}{m\omega_0} \int_0^t \sin \omega_0 (t - \tau) \cdot F(\tau)\, d\tau \right)$$

for all times $t \in \mathbb{R}$.

Proof. To simplify notation, set $m := 1$.

(I) Existence. Let $t > 0$. Differentiation with respect to time t yields

$$\dot{x}(t) = -\omega_0 x_0 \sin \omega_0 t + v_0 \cos \omega_0 t + \int_0^t \cos \omega_0 (t - \tau) \cdot F(\tau)\, d\tau.$$

Note that the differentiation of the integral yields the additional term $\sin \omega_0 (t - t) \cdot F(t)$ which vanishes. Furthermore,

$$\ddot{x}(t) = -\omega_0^2 x_0 \cos \omega_0 t - \omega_0 v_0 \sin \omega_0 t - \omega_0 \int_0^t \sin \omega_0 (t - \tau) \cdot F(\tau) d\tau + F(t).$$

Hence $\ddot{x}(t) = -\omega_0^2 x(t) + F(t)$ if $t > 0$.

(II) Uniqueness. Suppose that the two functions $x_1, x_2 : \mathbb{R} \to \mathbb{R}$ are solutions of (2.18). Define $x(t) := x_1(t) - x_2(t)$. The function $x : \mathbb{R} \to \mathbb{R}$ satisfies then problem (2.18) in the special case where $x_0 = v_0 = 0$ and $F(t) \equiv 0$. By a standard result, this implies $x(t) \equiv 0$. □

The retarded Green's function. Set

$$\mathcal{G}(t) := \begin{cases} \frac{\sin \omega_0 t}{m\omega_0} & \text{if} \quad t \geq 0, \\ 0 & \text{if} \quad t < 0. \end{cases}$$

If $x_0 = v_0 = 0$, that is, the particle rests at time $t = 0$, then the solution from Theorem 2.3 can be written as

$$\boxed{x(t) = \int_{-\infty}^{\infty} \mathcal{G}(t - \tau) F(\tau)\, d\tau \qquad \text{for all} \quad t \in \mathbb{R}.} \tag{2.19}$$

Synonymously, the function $\mathcal{G}$ is called

- the response function,
- the Feynman propagator, and
- the retarded Green's function

of the causal initial-value problem (2.18) for the harmonic oscillator on the real line. In terms of physics, the function $\mathcal{G}$ describes the response of the harmonic oscillator to the action of the external force F which is switched on at time $t = 0$. Here, the harmonic oscillator rests on the time interval $]-\infty, 0]$. Observe that the construction of the function $\mathcal{G}$ critically depends on causality, that is, this function refers to the behavior of the harmonic oscillator in the future.

Response functions play a fundamental role in quantum field theory.

We will study this in Chap. 14.

The Dirac delta function. Physicists like to use the Dirac delta function $\delta = \delta(t)$ in order to introduce the Green's function $\mathcal{G}$ in a formal, but elegant way. To this end, they choose the external force

$$F(t) := \delta(t) \qquad \text{for all} \quad t \in \mathbb{R}$$

which only acts at the initial time $t = 0$. Using the characteristic property (11.19) of the Dirac delta function on page 590, it follows from (2.19) that

$$\boxed{x(t) = \int_{-\infty}^{\infty} \mathcal{G}(t-\tau)\delta(\tau)d\tau = \mathcal{G}(t) \qquad \text{for all} \quad t \in \mathbb{R}.}$$

Intuitively, the retarded Green's function $t \mapsto \mathcal{G}(t)$ describes the motion of a harmonic oscillator on the real line

- which rests on the time interval $]-\infty, 0]$, and
- which starts to move under the influence of a large external force that acts precisely at the time point $t = 0$.

Observe that the Dirac delta function is not a classical function, but a generalized function also called distribution. The rigorous theory of distributions will be studied in Chap. 11.

Application of the Laplace transform to the harmonic oscillator. We want to show how the solution from Theorem 2.3 can be obtained by using the Laplace transform. To display the elegant basic idea as clearly as possible, we restrict ourselves to a formal argument by ignoring the range of validity of the following formulas. Moreover, in order to simplify notation we set $m := 1$ and $\hbar := 1$. Then $\mathcal{E} = \omega$. We will use the basic rules (i), (ii), (iii) for the Laplace transform considered on page 93.

To begin with, suppose that the function $x = x(t)$ is a solution of the causal initial-value problem (2.18) above. By the derivative rule (ii), the Laplace transform $\mathcal{L}x$ satisfies the equation

$$-\omega^2 \mathcal{L}x + \mathrm{i}\omega x_0 - v_0 + \omega_0^2 \mathcal{L}x = \mathcal{L}F.$$

Hence

$$\boxed{\mathcal{L}x = -\frac{\mathrm{i}\omega x_0}{\omega_0^2 - \omega^2} + \frac{v_0}{\omega_0^2 - \omega^2} + \frac{\mathcal{L}F}{\omega_0^2 - \omega^2}.}$$

This means that the Laplace transform $\mathcal{L}x$ of the motion possesses a very simple structure in the frequency space. This solution depends critically on the function

$$R(\omega) := \frac{1}{\omega_0^2 - \omega^2}$$

which is called the response function of the harmonic oscillator in the frequency space. The singularities of the function R at the points $\omega = \pm\omega_0$ correspond to the eigenoscillations $t \mapsto \mathrm{e}^{\pm \mathrm{i}\omega_0 t}$ of the harmonic oscillator. We will see at the end of the proof that the function R is the Laplace transform of the retarded Green's function $\mathcal{G}$. It remains to reconstruct the function $x = x(t)$ from its Laplace transform. Let us discuss this.

Using partial fractions and the wave rule (i), we get

$$\frac{\mathrm{i}\omega}{\omega^2 - \omega_0^2} = \frac{1}{2}\left(\frac{\mathrm{i}}{\omega - \omega_0} + \frac{\mathrm{i}}{\omega + \omega_0}\right) = \mathcal{L}\left(\frac{\mathrm{e}^{-\mathrm{i}\omega_0 t} + \mathrm{e}^{\mathrm{i}\omega_0 t}}{2}\right) = \mathcal{L}(\cos \omega_0 t).$$

Similarly,

$$\frac{1}{\omega_0^2 - \omega^2} = \frac{1}{2\omega_0}\left(\frac{1}{\omega + \omega_0} - \frac{1}{\omega - \omega_0}\right) = \mathcal{L}\left(\frac{\mathrm{e}^{\mathrm{i}\omega_0 t} - \mathrm{e}^{-\mathrm{i}\omega_0 t}}{2\mathrm{i}\omega_0}\right) = \mathcal{L}\left(\frac{\sin \omega_0 t}{\omega_0}\right).$$

Finally, the convolution rule (iii) tells us that

$$\frac{1}{\omega_0^2 - \omega^2} \cdot \mathcal{L}F = \mathcal{L}\left(\frac{\sin \omega_0 t}{\omega_0} * F\right).$$

Summarizing, for all times $t \geq 0$,

$$x(t) = x_0 \cos \omega_0 t + \frac{v_0}{\omega_0} \cdot \sin \omega_0 t + \int_0^t \frac{\sin \omega_0 (t - \tau)}{\omega_0} \cdot F(\tau)\, d\tau.$$

This is precisely the solution from Theorem 2.3 on page 95.

2.3 Historical Background

The explicit statement that matter is composed of indivisible particles called atoms (from the Greek $\alpha\tau o\mu o\sigma$, atomos, "uncuttable") we trace to the ancient town of Abdera, on the seacoast of Thrace. There, in the latter part of the fifth century B.C., the Greek philosophers Leucippus

and Democritus[10] taught that all matter is made up of atoms and empty space... After the birth of modern science, the idea of atoms came to be used as a basis of quantitative theories of matter. In the 17th century Newton attempted to account for the expansion of gases in terms of the outrush of their atoms into empty space. More influential, in the early 19th century Dalton explained the fixed ratios of the weights of chemical elements that make up common compounds in terms of the relative weights of the atoms of these elements...

By the end of the 19th century the idea of the atom had become familiar to most scientists – familiar but not yet universally accepted... All this changed in the first decades of the 20th century...

Just as ancient Abdera symbolizes for us the birth of atomism, there is one place with which the discovery of constituents of the atom is especially associated: It is the Cavendish Laboratory of the University of Cambridge (England) founded in 1874.[11] There, in 1895, Thomson performed the experiments on cathode rays that led him to conclude that there is a particle – the electron – that is both the carrier of electricity and a basic constituent of all atoms. It was at the Cavendish in 1895–1898 that Rutherford began his work on radioactivity, and to the Cavendish in 1919 that Rutherford returned, after his discovery of the atomic nucleus in 1911, to succeed Thomson as Cavendish Professor of Experimental Physics and to found what was long the preeminent center for nuclear physics. The list of constituents of the atom was completed at the Cavendish Laboratory in 1932, when Chadwick discovered the neutron ... All ordinary matter is composed of atoms which in turn consist of protons, neutrons, and electrons...

The particle menu changed drastically in the 1950s when large accelerators like the Betatron at Berkeley and new devices for detecting particles like the bubble chamber began to become available. In the debris of collisions of the high-energy protons from these accelerators were found a great variety of new hadrons, labelled $\rho, \omega, \eta, \phi, \Delta, \Xi, \Omega$, and so on – so many that the Greek alphabet was in danger of exhaustion. They were all unstable with extremely short lifetimes, which is why they are absent in ordinary matter... Soon an attempt was made to restore some economy to the multitude of hadrons. In the early 1960s Murray Gell-Mann and George Zweig of the California Institute of Technology, building on earlier work of Gell-Mann and Yuval Ne'eman of Tel-Aviv, proposed independently that the hadrons are all composites of a few types of really elementary building blocks, called quarks by Gell-Mann.[12]

Steven Weinberg
The Discovery of Subatomic Particles[13]

[10] Democritus (460–371 B.C.), Sir Isaac Newton (1643–1727), John Dalton (1766–1844).

[11] The first Cavendish Professor was James Clerk Maxwell (1831–1879) from 1874 until his death. The next Cavendish Professors were Lord Rayleigh (1842–1919), Nobel prize in physics in 1904, and after him Sir Joseph John Thomson (1856–1940), Nobel prize in physics in 1906, Lord Ernest Rutherford (1871–1937), Nobel prize in chemistry in 1908, for the chemistry of radioactive substances, James Chadwick (1891–1974), Nobel prize in physics in 1935.

[12] Quarks – a whimsical name was taken from a passage in James Joyce's *Finnegans Wake* by Gell-Mann.

[13] Scientific American Library, New York, 1983 (reprinted with permission).

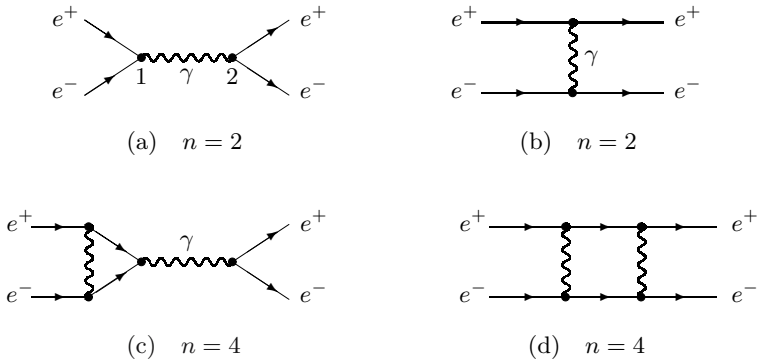


Fig. 2.5. Feynman diagrams for electron-positron scattering

Quantum field theory is a set of ideas and tools that combine three of the major themes of modern physics: the quantum theory, the field concept, and the principle of relativity. Today, most working physicists need to know some quantum field theory, and many others are curious about it. The theory underlies modern elementary particle physics, and supplies essential tools to nuclear physics, atomic physics, condensed matter physics, and astrophysics. In addition, quantum field theory has led to new bridges between physics and mathematics.
One might think that a subject of such power and widespread application would be complex and difficult. In fact, the central concepts and techniques of quantum field theory are quite simple and intuitive. This is especially true of the many pictorial tools (Feynman diagrams, renormalization group flows, and spaces of symmetry transformations) that are routinely used by quantum field theorists. Admittedly, these tools take time to learn, and tying the subject together with rigorous proofs can become extremely technical. Nevertheless, we feel that the basic concepts and tools of quantum field theory can be made accessible to all physicists, not just an elite group of experts.

Michael Peskin and Daniel Schroeder[14]

The gauge theories for the strong and electroweak interaction have become the Standard Model of particle physics. They realize in a consistent way the requirements of quantum theory, special relativity and symmetry principles. For the first time, we have a consistent theory of the fundamental interactions that allows for precision calculations for many experiments. The Standard Model has, up to now, successfully passed all experimental tests. This success establishes the importance of gauge theories, despite the fact that gravity is not included and that the Standard Model is most likely an effective theory resulting from the low-energy limit of a more fundamental theory.

Manfred Böhm, Ansgar Denner, and Hans Joos[15]

[14] From the preface to *An Introduction to Quantum Field Theory*, Addison–Wesley, Reading, Massachusetts, 1995 (reprinted with permission).
[15] From the preface to *Gauge Theories of the Strong and Electroweak Interaction*, Teubner, Stuttgart, 2001 (reprinted with permission).

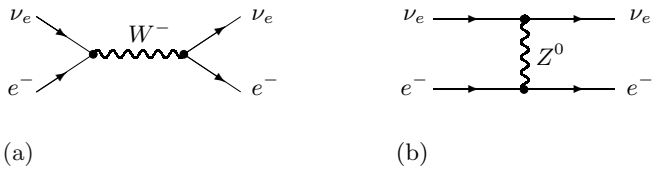


Fig. 2.6. Neutrino-electron scattering

> If supersymmetry plays the role in physics that we suspect it does, then it is very likely to be discovered by the next generation of particle accelerators, either at Fermilab in Batavia, Illinois, or at CERN in Geneva, Switzerland.
> Edward Witten, 2000[16]

If we are looking around, we see matter and light in the physical world. There arises the question about the deeper structure of matter and light and the final causes for these structures. The causes are called forces (or interactions) by physicists.

Classical gravitation. In the 17th century, Newton introduced the notion of gravitational force. This way, Newton could justify Kepler's empirical laws on the motion of planets. The precision of Newton's mechanics was so high that, by using the methods of perturbation theory, astronomers could successfully predict the existence of new planets, namely, Neptune and Pluto in 1846 and 1930, respectively.

The electromagnetic field. In the 19th century, Faraday studied experimentally the electromagnetic force. In particular, he discovered the law of induction on the relation between electricity and magnetism in 1821. Maxwell formulated mathematically his celebrated theory of electromagnetism in 1864. The point is that the Maxwell equations unify two apparently different forces in nature, namely, the electric and the magnetic force. To this end, Maxwell used Faraday's fundamental idea of the electromagnetic field. The existence of electromagnetic waves was demonstrated experimentally by Hertz in 1886. In the 20th century, classical physics was revolutionized by

- quantum physics based on Planck's quantum hypothesis from the year 1900,
- Einstein's 1905 theory of special relativity (e.g., electrodynamics of moving bodies),
- Einstein's 1905 theory of light quanta (photons), and
- Einstein's 1915 theory of general relativity for gravitation.

Let us discuss some crucial points.

[16] From the Foreword to G. Kane, Supersymmetry: Squarks, Photinos, and the Unveiling of the Ultimate Laws of Nature, Perseus Publishing, Cambridge, Massachusetts, 2000.

2.3.1 Planck's Radiation Law

> My new universal radiation formula was submitted to the Berlin Physical Society, at the meeting on October 19, 1900. On the very day when I formulated this law by using purely formal arguments, I began to devote myself to the task of investing it with a true physical meaning. This quest let me to study the interrelation between entropy and probability. Since the entropy S is an additive magnitude, but the probability W is a multiplicative one, I simply postulated that
>
> $$S = k \ln W,$$
>
> where k is a universal constant.[17] I found that k represents the so-called absolute gas constant. It is, understandably, often called Boltzmann's constant. However, this calls for the comment that Boltzmann (1844–1906) never introduced this constant.
>
> Now, as for the quantity W with respect to radiation in a cavity, I found that in order to interpret it as a probability, it was necessary to introduce a universal constant which I called h. Since it had the physical dimension of action (energy times time), I gave it the name *elementary quantum of action*. Thus the nature of entropy as a measure of probability, in the sense indicated by Boltzmann, was established in the domain of radiation, too.
>
> While the significance of the quantum of action for the interrelation between entropy and probability was established, the part played by the new constant in general physical processes still remained an open question. I therefore tried immediately to weld the elementary quantum of action into the framework of classical theory. But in the face of all such attempts, which continued for a number of years, this constant showed itself to be obdurate. For it heralded the advent of something entirely unprecedented (namely, quantum physics), and was destined to remodel basically the physical outlook and thinking of man which, ever since Newton (1643–1727) and Leibniz (1646–1716) laid the groundwork for infinitesimal calculus, were founded on the assumption that all causal interactions in nature are continuous.[18]
>
> Max Planck (1858–1947)

The birth of quantum physics in 1900. Consider a star with surface temperature T. For example, the surface temperature of the sun is equal to 6000K. The famous Planck radiation law from 1900 describes the amount E_A of electromagnetic energy which is emitted from the surface area A of the star during the time interval $[0, t]$ with respect to the interval $[\omega_1, \omega_2]$ of angular frequency. Explicitly,

$$\boxed{E_A = \frac{ctA}{4} \int_{\omega_1}^{\omega_2} \mathcal{P}(\omega, T) d\omega} \tag{2.20}$$

with the Planck function

[17] Note that in modern terminology, W is not a probability, but a statistical weight.
[18] M. Planck, Wissenschaftliche Selbstbiographie, Barth, Leipzig 1948.

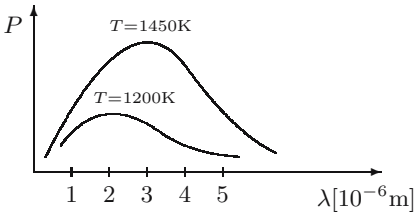


Fig. 2.7. Planck's radiation law

$$\boxed{\mathcal{P}(\omega, T) := \frac{\hbar\omega^3}{\pi^2 c^3 (e^{\hbar\omega/kT} - 1)}.} \tag{2.21}$$

Here, we use the following notation: c velocity of light in vacuum, k Boltzmann constant, h Planck constant, $\hbar := h/2\pi$, λ wave length of the radiation, ν frequency, ω angular frequency. Recall that $\nu = c/\lambda = \omega/2\pi$. More precisely, the Planck radiation law refers to the radiation of a black body. By definition, a body is called 'black' by physicists iff it absorbs all the incoming electromagnetic radiation. In other words, there is no reflection of incoming radiation. The radiation coming from stars can be described approximately by black-body radiation. If we pass to the wave length λ, then we obtain

$$E_A = \frac{ctA}{4} \int_{\lambda_1}^{\lambda_2} P(\lambda, T) d\lambda$$

with

$$P(\lambda, T) := \frac{8\pi hc}{\lambda^5 (e^{hc/kT\lambda} - 1)}$$

(Fig. 2.7). The values of wave length λ observed today by physicists range from 1 femtometer $= 10^{-15}$m (high energy cosmic rays) to 10^7m (low energy radio waves in the universe). Let us first discuss some important special cases of the Planck radiation law.

The Stefan–Boltzmann law. From (2.20) we obtain the total amount of emitted energy

$$E_A = \frac{ctA}{4} \int_0^{\infty} \mathcal{P}(\omega, T) d\omega.$$

Hence

$$E_A = \sigma t A T^4$$

where $\sigma = 2\pi^5/15c^2h^3 = 5.7 \cdot 10^{-8}\mathrm{J/sm^2K^4}$.

The Wien displacement law. For fixed temperature T, the function $P = P(\lambda, T)$ has a maximum at the wave length

$$\lambda_{\max} = \frac{a}{T},$$

where $a := 2.898 \cdot 10^{-3}$mK. In fact, from $P_\lambda(\lambda, T) = 0$ we get the fixed point equation

$$x = f(x) \quad \text{where} \quad f(x) := 5e^{-x}(e^x - 1),$$

and $x := hc/(kT\lambda)$. The corresponding iteration scheme

$$x_{n+1} = f(x_n), \qquad n = 0, 1, 2, \ldots$$

converges rapidly to the value $x = 4.965$.

The Wien approximation for high frequencies. If $\hbar\omega/kT \gg 1$, then $e^{\hbar\omega/kT} - 1$ can be approximated by $e^{\hbar\omega/kT}$. Hence

$$\mathcal{P}(\omega, T) = \frac{\hbar\omega^3}{\pi^2 c^3} e^{-\hbar\omega/kT}, \qquad \omega \to \infty. \tag{2.22}$$

Rayleigh–Jeans approximation for low frequencies. If $\hbar\omega/kT \ll 1$, then $e^{\hbar\omega/kT} - 1$ can be approximated by $\hbar\omega/kT$. Hence

$$\mathcal{P}(\omega, T) = \frac{kT\omega^2}{\pi^2 c^3}, \qquad \omega \to 0. \tag{2.23}$$

Cavity radiation. Consider a box of volume V filled with electromagnetic radiation in thermodynamic equilibrium. Let T be the temperature of the walls. According to Planck, the energy E of the radiation with respect to the interval $[\omega_1, \omega_2]$ of angular frequency is given by

$$E = \eta(T)V$$

with the energy density

$$\boxed{\eta(T) = \int_{\omega_1}^{\omega_2} \mathcal{P}(\omega, T)d\omega} \tag{2.24}$$

where

$$\mathcal{P}(\omega, T) = \frac{\hbar\omega^3}{\pi^2 c^3 (e^{\hbar\omega/kT} - 1)}. \tag{2.25}$$

The classical radiation argument. It was discovered by Kirchhoff (1824–1884) that there exists a close connection between the energy density $\eta(T)$ of the electromagnetic radiation in a box and the energy E_A radiated by a black body. Explicitly,

$$E_A = \frac{ctA\eta(T)}{4}. \tag{2.26}$$

The basic idea is to equip the box with a small slot of surface area A. The radiation coming from the slot then represents the black body radiation (2.26). The complete argument can be found in Zeidler (1986), Vol. IV, p. 410.

Consequently, in order to derive the Planck radiation law (2.20), it is sufficient to justify (2.24), (2.25). This will be done in the following. We will use two different arguments based on phenomenological thermodynamics and the Boltzmann statistics.

Planck's first justification of the radiation law. Let us consider the argument from phenomenological thermodynamics that was used first by Planck in 1899 to obtain the energy density (2.24). Later on Planck applied Boltzmann's method of statistical mechanics to the radiation law. In 1893 Wien published the radiation law

$$\mathcal{P}(\omega, T) = A\omega^3 e^{-B\omega/T}. \tag{2.27}$$

Tests showed, however, that this law failed for low frequencies. In 1899, Lord Rayleigh formulated the law (2.23) for small frequencies. The idea of Planck was to get the universal law for all frequencies by interpolating the Wien law and the Rayleigh–Jeans law. To this end, he used the notion of entropy. To explain this, let us start with a general thermodynamic system which has one degree of freedom. Such a system is described by the fundamental equation

$$\boxed{TdS = dE + pdV} \tag{2.28}$$

between volume V, pressure p, temperature T, inner energy E, and entropy S. Choose E and V as basic variables.[19] Then

$$S = S(E, V), \qquad T = T(E, V), \qquad p = p(E, V).$$

From (2.28) we get $TS_E dE + TS_V dV = dE + pdV$. Hence

$$S_E = \frac{1}{T}, \quad S_V = \frac{p}{T}.$$

Consider now the cavity radiation in a box of volume V with respect to the interval $[\omega, \omega + \Delta\omega]$ of angular frequency, where $\Delta\omega$ is sufficiently small. Then

$$E = V\mathcal{P}(\omega, T)\Delta\omega$$

with the unknown function $\mathcal{P}$. To simplify notation, in what follows we will denote different constants by the same symbols C and D.

(i) The Rayleigh–Jeans law. From (2.23) we obtain

$$E = \frac{kTV\omega^2\Delta\omega}{\pi^2 c^3} = CT.$$

Hence $E'(T) = C$. Since $S_E = T^{-1}$, it follows from the chain rule that

[19] A detailed study of both phenomenological and statistical thermodynamics along with substantial applications can be found in Zeidler (1986), Vol. IV, Chaps. 67 and 68.

$$S_{EE} = \frac{dT^{-1}}{dE} = \frac{dT^{-1}}{dT}\frac{dT}{dE} = -\frac{1}{T^2E'(T)}. \tag{2.29}$$

This implies

$$S_{EE} = -\frac{C}{E^2}.$$

(ii) The Wien law. It follows from (2.27) that $E = Ce^{D/T}$. Hence

$$E'(T) = -\frac{CD}{T^2}e^{D/T} = -\frac{DE}{T^2}.$$

By (2.29),

$$S_{EE} = -\frac{C}{E}.$$

(iii) Planck's interpolation. Planck started from the ansatz

$$S_{EE} = -\frac{D}{E(E+C)}. \tag{2.30}$$

This equation can be satisfied by the choice

$$E = \frac{C}{e^{D/T}-1}.$$

In fact,

$$E'(T) = \frac{De^{D/T}}{T^2(e^{D/T}-1)^2} = \frac{DE(E+C)}{T^2}.$$

From (2.29) we get relation (2.30). This way, Planck obtained the desired function

$$\mathcal{P} = \frac{C}{e^{D/T}-1}$$

of the radiation law. This method, however, does not determine the constants C and D. As we will show below, the stronger methods of the Boltzmann statistics along with the quantum hypothesis allow us the determination of the unknown constants.

The Planck function and number theory. Letting $x := \hbar\omega/kT$, we get

$$\mathcal{P}(\omega, T) = \frac{aT^3x^3}{e^x-1},$$

where $a := k^3/\pi^2\hbar^2c^3$. For all complex numbers x with $0 < |x| < 2\pi$, one has the classical relation

$$\frac{x^3}{e^x-1} = \sum_{k=0}^{\infty}\frac{B_k}{k!}x^{k+2}.$$

The so-called Bernoulli numbers B_k were introduced by Jakob Bernoulli (1675–1705). To explain the meaning of B_k, define the sum

$$S_n^p := 1^p + 2^p + \ldots + n^p.$$

Then

$$S_n^p = \frac{n^{p+1}}{p+1} + \frac{n^p}{2} + \frac{B_2}{2}\binom{p}{1}n^{p-1} + \frac{B_3}{3}\binom{p}{2}n^{p-2} + \ldots + \frac{B_p}{p}\binom{p}{p-1}n$$

for $n = 1, 2, \ldots$ and $p = 1, 2, \ldots$ The Bernoulli numbers can be computed by means of the recursion formula

$$\binom{p+1}{n}B_n = -\sum_{k=0}^{n-1}\binom{p+1}{k}B_k.$$

Explicitly,

$$B_0 = 1,\ B_1 = -\frac{1}{2},\ B_2 = \frac{1}{6},\ B_3 = 0,\ B_4 = -\frac{1}{30}.$$

For odd natural numbers $n \geq 3$, $B_n = 0$. The relation between the Bernoulli numbers and the Riemann zeta function will be discussed on page 278.

2.3.2 The Boltzmann Statistics and Planck's Quantum Hypothesis

Entropy and Boltzmann's partition function. Consider a system which can attain the possible energy states $E_0, E_1, \ldots, E_m$. According to Boltzmann statistics, the mean value $\overline{E}$ of energy and the energy fluctuation ΔE at temperature T are given by

$$\boxed{\overline{E} = \textstyle\sum_{n=0}^{m} p_n E_n} \tag{2.31}$$

and

$$(\Delta E)^2 = \sum_{n=0}^{m} p_n (E_n - \overline{E})^2$$

along with the probabilities

$$p_n = \frac{e^{-E_n/kT}}{\sum_{n=0}^{m} e^{-E_n/kT}}. \tag{2.32}$$

Let us motivate this by means of the principle of maximal entropy. To this end, define the entropy

$$\boxed{S := -k\sum_{n=0}^{m} p_n \ln p_n}$$

and consider the problem

$$S = \max!$$

along with the side conditions

$$\sum_n p_n E_n - \overline{E} = 0 \quad \text{and} \quad \sum_n p_n - 1 = 0$$

where $0 \le p_n \le 1$ for all n. According to the Lagrange multiplier rule,[20] we have to study the critical points of the Lagrange function

$$L := S + \lambda \left(\sum_n p_n E_n - \overline{E} \right) + \mu \left(\sum_n p_n - 1 \right).$$

The condition $L_{p_n} = 0$ for all n yields

$$-k \ln p_n - k + \lambda E_n + \mu = 0, \qquad n = 0, 1, \ldots, m.$$

Hence $p_n = \text{const} \cdot e^{\lambda E_n / k}$. Set $\lambda = -1/T$. Using $\sum_n p_n = 1$, we get the desired result (2.32). From a formal point of view, this argument shows that temperature is nothing else than a Lagrange multiplier.

Boltzmann's law for the energy distribution. Let us now apply this to the special situation where

$$E_n = n\hbar\omega, \qquad n = 0, 1, \ldots$$

Introducing the so-called *partition function*

$$Z := \sum_{n=0}^{\infty} e^{-E_n/kT},$$

we get

$$Z = \sum_{n=0}^{\infty} e^{-n\gamma} = \frac{1}{1 - e^{-\gamma}}$$

where $\gamma := \hbar\omega/kT$. By (2.31),

$$\overline{E} = -\frac{\hbar\omega Z'(\gamma)}{Z(\gamma)} = -\hbar\omega \, \frac{d \ln Z(\gamma)}{d\gamma} = \frac{\hbar\omega}{e^{\gamma} - 1}.$$

Hence

$$\overline{E} = \frac{\hbar\omega}{e^{\hbar\omega/kT} - 1}. \tag{2.33}$$

In particular, if $\hbar\omega/kT \ll 1$, we get the approximation

$$\overline{E} = kT$$

which corresponds to the famous Boltzmann law. This law tells us that for sufficiently high temperature T, each degree of freedom in a many-particle

[20] A detailed investigation of the Lagrange multiplier rule with respect to statistical physics can be found in Zeidler (1986), Vol. III, p. 296.

system contributes the energy kT to the total mean energy of the system. Relation (2.33) will be used now in order to obtain the Planck radiation law.

Planck's quantum hypothesis. In 1900, Max Planck postulated that the energy of the electromagnetic radiation emitted from the walls of a cavity is quantized. The energy quanta are given by the formula

$$\boxed{\Delta E = \hbar\omega.} \tag{2.34}$$

Here, ω denotes the angular frequency of the electromagnetic radiation. Recall that $\hbar := h/2\pi$ where

$$h = 6.626 \cdot 10^{-34}\text{Js}$$

is a fundamental constant in nature called the Planck constant (or Planck's quantum of action). Note that actions have the physical dimension of "energy times time". This is the most important physical quantity in nature. In daily life, the actions have themagnitude of 1 Js. Planck's quantum hypothesis marked a revolutionin physics. It paved the way to modern quantum physics.

Let us discuss how Planck obtained his radiation law from the quantum hypothesis. Consider a cubic cavity of length L and volume $V = L^3$ filled with electromagnetic radiation in thermodynamic equilibrium. By (2.6) on page 85, the number of standing electromagnetic waves with respect to the interval $[\omega, \omega + \Delta\omega]$ of angular frequencies is equal to

$$2\Delta N = \frac{V\omega^2\Delta\omega}{\pi^2 c^3}.$$

The factor 2 can be motivated by the fact that plane electromagnetic waves possess 2 linearly independent directions of polarization. Thus, the radiation energy of the cavity with respect to $[\omega, \omega + \Delta\omega]$ is

$$2\Delta N \cdot \overline{E} = \frac{V\hbar\omega^3\Delta\omega}{\pi^2 c^3(e^{\hbar\omega/kT} - 1)}$$

by (2.33). According to (2.24) on page 103, this yields the Planck function

$$\mathcal{P}(\omega, T) = \frac{\hbar\omega^3}{\pi^2 c^3(e^{\hbar\omega/kT} - 1)}.$$

Heisenberg's ground state energy of the harmonic oscillator. Quantum mechanics was founded by Heisenberg in 1925. Born noticed that Heisenberg's ideas could be converted into a theory for infinite-dimensional matrices. Heisenberg showed that the quantized energy of a harmonic oscillator of angular frequency ω is given by the key formula

$$\boxed{E = \hbar\omega\left(n + \frac{1}{2}\right), \qquad n = 0, 1, 2, \ldots} \tag{2.35}$$

This result justifies Planck's quantum hypothesis for the harmonic oscillator. In particular, Heisenberg obtained the value $E = \frac{1}{2}\hbar\omega$ for the energy of the ground state corresponding to $n = 0$. The same argument as in (2.33) shows that the mean energy of a large system of harmonic oscillators at temperature T is given by

$$\boxed{\overline{E} = \frac{\hbar\omega}{e^{\hbar\omega/kT} - 1} + \frac{\hbar\omega}{2}.}$$

In 1926 Schrödinger formulated his wave quantum mechanics based on a partial differential equation – the famous Schrödinger equation. We will show in Volume II how the key formula (2.35) follows from the Schrödinger equation for the harmonic oscillator.

2.3.3 Einstein's Theory of Special Relativity

In the year 1905, in the same volume of the journal *Annalen der Physik*, Einstein published four seminal papers on

(i) special relativity (electrodynamics of moving bodies),
(ii) the relation $E = m_0c^2$ between the rest mass m_0 of a particle and its rest energy E,
(iii) the photoelectric effect (light particle hypothesis), and
(iv) the Brownian motion.

Let us discuss a few basic ideas.

The fundamental role played by inertial systems. Each physical theory has to be described by both

- fundamental equations of motion and
- the systems of reference in which the equations of motion are valid.

In addition, one has to know how to transform the equations of motion between different systems of reference.

For example, the Newtonian equations of mechanics are valid in each inertial system. By definition, a Cartesian coordinate system is an inertial system iff there exists a system time t for it such that each mass point, which is far enough from other masses and shielded against fields (e.g., light pressure), remains at rest or moves rectilinearly with constant velocity. For example, a spaceship represents an inertial system if it is located at far distance from stars and planets and it flies without rocket propulsion. The experience of astronomers shows that each Cartesian system Σ_{sun} is a good approximation of an inertial system. Such a system is defined as follows.

- The system has its origin at the center of gravity of our solar system, which lies within the sun.
- The axes of Σ_{sun} point towards fixed stars which can be chosen arbitrarily.

In addition, note that a Cartesian coordinate system represents an inertial system if it is obtained from Σ_{sun} by a constant translatory motion. The rotating earth is not an inertial system.

In 1905 Einstein derived his theory of special relativity from the postulate that

Physical processes proceed the same way in all inertial systems.

This is called the principle of special relativity. We will show later on that this principle implies that relativistic physics has to be described mathematically in terms of invariants of the Poincaré group. Around 1900 it was a famous open problem in physics to determine the systems of reference in which the Maxwell equations are valid and to understand how the electromagnetic field has to be transformed under a general change of the observer. Einstein solved this problem by postulating that

The Maxwell equations are valid in each inertial system.

In particular, this implies the surprising fact that the velocity of light in a vacuum, c, is the same in each inertial system. This forced Einstein to change Newton's classical mechanics and the classical thinking about space and time. In Newton's classical mechanics, the formula

$$x' = x - vt, \quad y' = y, \quad z' = z \quad t' = t \tag{2.36}$$

describes the transformation from a right-handed Cartesian (x, y, z)-system to a right-handed Cartesian (x', y', z')-system. Here, we assume that the origin of the (x', y', z')-system moves with the constant positive velocity v along the x-axis. At time $t = 0$, the two systems coincide. Einstein showed in 1905 that one has to replace the Galilei transformation (2.36) by the following Lorentz transformation

$$\boxed{x' = \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad y' = y, \quad z' = z, \quad t' = \frac{t - \frac{vx}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}}.} \tag{2.37}$$

The point is that in contrast to classical mechanics, time is not an absolute quantity in Einstein's theory of special relativity, but it depends on the choice of the inertial system used by the observer. If the velocity v is small compared with the velocity c of light in vacuum, that is, $v/c \ll 1$, then the classical Galilei transformation (2.36) represents a good approximation of the relativistic Lorentz transformation (2.37). Motivated by the singularity of the Lorentz transformation (2.37) for $v = c$, Einstein postulated that

The maximal speed for the propagation of physical effects is the speed of light.

As a consequence of his new relativistic mechanics, Einstein obtained the fundamental relation

$$\boxed{E = m_0 c^2}$$

between the energy E and the rest mass m_0 of a resting particle. This is the most important law in physics, since it governs the energy production in stars, for example, in the sun.

2.3.4 Einstein's Theory of General Relativity

Newton's theory of gravitation implies a propagation of gravitational effects with infinite speed. This fact violates Einstein's postulate on the existence of a maximal signal velocity. To explain this, suppose that there occurs a huge explosion at the center C of our Milky Way which changes the mass of the center drastically. According to Newton, this explosion would change immediately the orbit of the sun around C. Following Einstein, the light needs 30 000 years from C to the sun, and hence the sun will notice the explosion 30 000 years later. In the years from 1905 until 1915, Einstein made strong efforts to modify Newton's theory of gravitation in such a way that gravitational effects propagate with the velocity of light in vacuum. In 1915 Einstein published his theory of general relativity. In this theory, Newton's gravitational force is replaced by the

Riemann curvature of the 4-dimensional space-time manifold.

This curvature is caused by the existing masses in the universe.

In Newton's mechanics, the equations of motion for the planets are invariant under 3-dimensional rotations (the group $SO(3)$). This leads to conservation of angular momentum which implies that the orbits of planets lie in a plane. In addition, the equations of motion possess an additional hidden symmetry which is related to 4-dimensional rotations (the group $SO(4)$). This additional symmetry fixes the axes of the elliptic orbits of the planets (Lenz vector). In general relativity, the latter symmetry does not exist anymore. Because of this symmetry breaking, the great semi-axis (perihelion) of Mercury is not fixed, but it rotates 43 arc seconds per century. This rotation was discovered by Le Verrier (1811–1877). In contrast to Newtonian mechanics, Einstein's theory of general relativity yields precisely the observed motion of the perihelion of Mercury. Moreover, Einstein's theory of general relativity implies

- the existence of black holes,
- the Big Bang,
- and the expansion of our universe.

As an introduction to Einstein's theory of special and general relativity, we recommend the author's monograph Zeidler (1986), Vol. IV.

2.3.5 Einstein's Light Particle Hypothesis

In his 1905 paper on the photoelectric effect, Einstein postulated that

Electromagnetic waves possess particle properties, too.

Nowadays these light particles are called photons.[21] According to Einstein, a photon of frequency ν has the energy

$$\boxed{E = h\nu}$$

where h is Planck's quantum of action. The perfect dualism between particles and waves was postulated by de Broglie in 1924. In 1932 Tamm and Ivanenko formulated the hypothesis that all fundamental forces in nature are related to the exchange of field quanta called messenger particles. The photon is the messenger particle for the electromagnetic force. Einstein's light particle hypothesis was experimentally established by

- the photoelectric effect (Millikan 1916), and
- the Compton effect (Compton 1922).

For their contributions to the quantum physics of light particles, Einstein, Millikan, and Compton were awarded the Nobel prize in physics in 1921, 1923, and 1927, respectively. Let us briefly discuss the photoelectric effect first investigated by Thomson and Lenard at the end of the 19th century. If light hits a metal, then electrons are ejected.

(i) *The classical picture.* Each electron is bound in the metal by the average binding energy U. In order to eject an electron, we need the light energy E where

$$E > U.$$

The classical electromagnetic wave is able to eject electrons if the intensity of the light is large enough. The frequency ν of light does not play any role.

(ii) *Einstein's quantum picture.* According to Einstein, light consists of light particles (photons) of energy $E = h\nu$. Such a photon can only eject an electron if its energy $h\nu$ is larger than the binding energy U of the electron. This means that

$$\nu > \nu_{\text{crit}}$$

where $\nu_{\text{crit}} := U/h$. Consequently, the photoelectric effect only occurs above a threshold frequency, in contrast to (i).

The experiment verifies case (ii).

Photons and gravitation. Photons do not have any rest mass. Nevertheless, they are affected by gravitation. For example, light is deflected by the

[21] The term 'photon' was introduced as the name for the particle of light, by the physical chemist Gilbert Lewis, at Berkeley University (California) in 1926.

sun. In 1959, using the Mößbauer effect (recoilless absorption of γ-quanta), the red shift of light in the gravitational field of earth was established in a tower of height 22.5m at Harvard University in Boston, by Pound and Repka. In 1928 Hubble discovered a red shift in the spectra of distant galaxies. This red shift is caused by the expansion of our universe after the Big Bang. s

Photons and the early universe. In 1965 Penzias and Wilson discovered a weak cosmic background microwave radiation which is a relict of the Big Bang (Nobel prize in physics in 1978). In 1992, the 'Cosmic Background Explorer (COBE)' satellite sent data to earth which showed that the background radiation is not perfectly anisotropic. This slightly anisotropic structure of the background radiation is very important for understanding the formation of galaxies in the early universe caused by energy fluctuations. More recent very precise measurements of the background radiation by the WMAP (Wilkinson Microwave Anisotropic Probe) experiment of NASA contribute to a better understanding of the evolution of our universe after the Big Bang (see page 80).

2.3.6 Rutherford's Particle Scattering

Coulomb force. Consider a particle with positive electric charge Q, mass m, and initial velocity vector $\mathbf{v} = v\mathbf{e}_1$ which is scattered at a fixed particle with positive charge Q_0, as shown in Fig. 2.8. Here, $\mathbf{e}_1$ is a fixed unit vector. The motion $\mathbf{x} = \mathbf{x}(t)$ of the incoming particle is governed by the Newtonian equation of motion

$$m\ddot{\mathbf{x}} = \frac{QQ_0}{4\pi\varepsilon_0||\mathbf{x}||^2} \cdot \frac{\mathbf{x}}{||\mathbf{x}||}$$

with respect to the Coulomb force between the charges Q and Q_0. The orbit of the scattered particle is a hyperbola. Explicitly, if the distance between the incoming particle and the x-axis is equal to d, then the scattering angle ϑ is given by

$$d = \frac{QQ_0}{4\pi\varepsilon_0 mv^2} \cot\frac{\vartheta}{2} .$$

The proof of this classical result will be given in the chapter on the hydrogen atom to be found in Volume II. The mathematics is the same as for the motion of a comet scattered by the sun in celestial mechanics.[22] The proof is essentially based on conservation of energy and angular momentum. In particular, if the scattering angle is equal to $\vartheta - \Delta\vartheta$, then the distance $d + \Delta d$ is given by

$$d + \Delta d = \frac{QQ_0}{4\pi\varepsilon_0 mv^2} \cot\frac{\vartheta - \Delta\vartheta}{2} .$$

[22] Replace Q_0, Q, and $1/4\pi\varepsilon_0$ by the mass of the sun, the mass of the planet, and the gravitational constant, respectively.

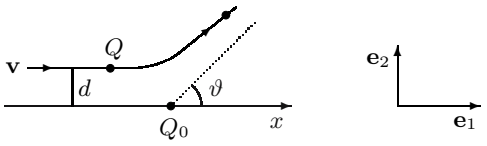


Fig. 2.8. Scattering of a charged particle

Differential cross section. Consider now a homogeneous current with the velocity vector $\mathbf{v} = v\mathbf{e}_1$ and particle number density ϱ. This corresponds to the electric current density vector

$$\mathbf{J} = Q\varrho\mathbf{v}$$

of incoming particles. Let ΔN denote the number of incoming particles which, during the time interval $[0, t]$, are scattered about angles which lie in the interval $[\vartheta, \vartheta + \Delta\vartheta]$. We then get the key formula for Coulomb scattering

$$\boxed{\Delta N = \varrho v t \Delta\sigma} \tag{2.38}$$

along with the so-called differential cross section

$$\boxed{\Delta\sigma = \left(\frac{QQ_0}{8\pi\varepsilon_0 m v^2}\right)^2 \frac{2\pi \sin\vartheta}{\sin^4 \frac{\vartheta}{2}} \cdot \Delta\vartheta + o(\Delta\vartheta), \quad \Delta\vartheta \to 0.}$$

To prove this, consider an annulus A perpendicular to the x-axis of radii d and $d+\Delta d$. Suppose that the incoming particles passing through the annulus A possess a scattering angle that lies in the interval $[\vartheta, \vartheta + \Delta\vartheta]$. The annulus A has the surface area

$$\Delta\sigma = \pi\big((d + \Delta d)^2 - d^2\big).$$

Hence

$$\Delta\sigma = \pi \left(\frac{QQ_0}{4\pi\varepsilon_0 m v^2}\right)^2 \left(\cot^2 \frac{\vartheta - \Delta\vartheta}{2} - \cot^2 \frac{\vartheta}{2}\right).$$

For small angles $\Delta\vartheta$, use the Taylor expansion

$$f(\vartheta - \Delta\vartheta) - f(\vartheta) = -f'(\vartheta)\Delta\vartheta + o(\Delta\vartheta), \quad \Delta\vartheta \to 0.$$

This yields

$$\Delta\sigma = \pi \left(\frac{QQ_0}{4\pi\varepsilon_0 m v^2}\right)^2 \frac{\cos \frac{\vartheta}{2}}{\sin^3 \frac{\vartheta}{2}} \cdot \Delta\vartheta + o(\Delta\vartheta), \quad \Delta\vartheta \to 0.$$

In addition, note that $\sin\vartheta = 2\sin\frac{\vartheta}{2}\cos\frac{\vartheta}{2}$. Physicists frequently write

$$\Delta\sigma = \left(\frac{QQ_0}{8\pi\varepsilon_0 m v^2}\right)^2 \frac{\Delta\Omega}{\sin^4\frac{\vartheta}{2}} + o(\Delta\Omega), \quad \Delta\Omega \to 0$$

where $\Delta\Omega := 2\pi \sin\vartheta\, \Delta\vartheta$. Intuitively, $\Delta\Omega$ is the surface area of a region on the unit sphere whose points possess a geographic latitude that lies in the interval $[\vartheta, \vartheta + \Delta\vartheta]$.

The structure of atoms. Around 1910, Ernest Rutherford (1871–1937) performed scattering experiments in order to discover the structure of atoms. He directed a beam of α-particles (generated by radioactive decay) to a thin metal foil (e.g., gold), and he measured the distribution of the angle ϑ of the scattered particles. This way, Rutherford discovered experimentally that an atom consists of a nucleus and electrons. For his contributions to radioactive chemistry, Rutherford was awarded the Nobel prize in chemistry in 1908. In 1932, Chadwick discovered the neutron which has no electric charge. This discovery clarified the structure of the nucleus of atoms. The final result reads as follows.

- The nucleus of an atom consists of Z protons p and N neutrons n.
- The nucleus is surrounded by Z electrons e^-.

Each electron e^- has the negative electric charge $-e$, whereas each proton p has the positive charge e. Here,

$$e = 1.602 \cdot 10^{-19}\text{As}.$$

The number Z of protons is called the atomic number. For example, the hydrogen atom ($Z = 1$) consists of one proton and one electron. The helium atom ($Z = 2$) consists of two protons, two neutrons, and two electrons. As Rutherford discovered, α-particles are nuclei of the helium atom. For the rest masses m_e, m_p, m_n of electron, proton, and neutron, respectively, we get

$$m_e = 9.108 \cdot 10^{-31}\text{kg}, \quad m_p = 1\,836\, m_e, \quad m_n = 1.009\, m_p\,.$$

Rutherford's scattering experiments also proved that the radius of an atom is about

$$r = 10^{-10}\text{m}.$$

Nowadays we know the following:

- radius of the proton: 10^{-15}m,
- radius of molecules: $10^{-9}\text{m} - 10^{-10}\text{m}$,
- radius of the nucleus: $10^{-14}\text{m} - 10^{-15}\text{m}$,
- radius of quarks and electrons: less than 10^{-18}m.

2.3.7 The Cross Section for Compton Scattering

In order to investigate the properties of elementary particles, physicists perform scattering experiments. The most important quantity of a scattering process is the cross section. There exist two important tasks for quantum field theory:

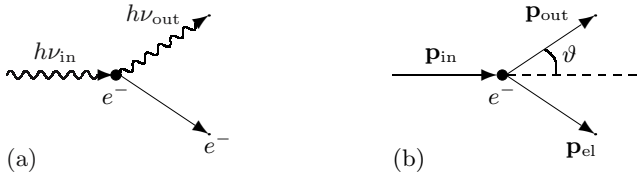


Fig. 2.9. Compton effect

(i) the computation of cross sections σ (scattering states), and
(ii) the computation of the rest energies of elementary particles (bound states).

It turns out that the task (i) is much easier to handle than (ii), since we can use the methods of perturbation theory pictured by Feynman diagrams. The ultimate, extremely ambitious goal is the creation of a theory which predicts the existence and properties (e.g., the masses and the magnetic moments) of all fundamental particles and forces in nature.

Compton scattering. In each scattering process, physicists measure the crucial cross section σ. In 1929, for the cross section of the Compton scattering of light at crystals, Klein and Nishina computed the formula

$$\boxed{\sigma = \int_{\mathbb{S}^2} f(\vartheta) d\Omega}$$

with

$$\boxed{f(\vartheta) := \Big(\beta^2(1+\cos^2\vartheta) + \beta(1-\beta)^2\Big) \cdot \frac{\lambda_e^2 \alpha^2}{8\pi^2}} \tag{2.39}$$

where $\beta := \lambda_{\text{in}}/\lambda_{\text{out}}$. Here, we use the following notation:

- m_e mass of the electron, $-e$ charge of the electron,
- λ_{in} (resp. λ_{out}) wave length of the incoming (resp. outgoing) photon,
- h Planck's quantum of action,
- Compton wave length of the electron

$$\lambda_e := \frac{h}{m_e c} = 10^{-12}\text{m},$$

- dimensionless fine structure constant in quantum electrodynamics

$$\boxed{\alpha := \frac{e^2}{4\pi\varepsilon_0 c\hbar} = \frac{1}{137.04}}$$

where ε_0 is the electric field constant of a vacuum.

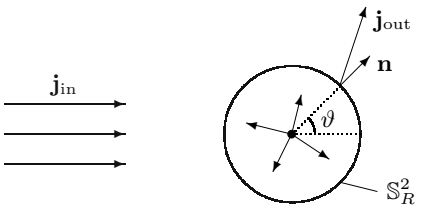


Fig. 2.10. Cross section for the Compton scattering of light

We integrate over the unit sphere $\mathbb{S}^2$ with the scattering angle ϑ and the surface element

$$d\Omega = \cos\vartheta \, d\vartheta d\varphi$$

where the geographic latitude ϑ and the geographic longitude φ vary in the interval $[-\pi/2, \pi/2]$ and $[-\pi, \pi]$, respectively. The differential

$$d\sigma := f(\vartheta) d\Omega$$

and the integral $\sigma = \int_{\mathbb{S}^2} d\sigma$ are called the differential cross section and the total cross section, respectively.

The famous Klein–Nishina formula (2.39) shows that Compton scattering is a second-order effect with respect to the fine structure constant α. Explicitly,

$$\sigma = \frac{\lambda_e^2 \alpha^2}{2\pi} \Big(\frac{4}{3} - \frac{8\gamma}{3} + \frac{104\gamma^2}{15} + \dots \Big), \qquad \gamma := \frac{\lambda_{\mathrm{e}}}{\lambda_{\mathrm{in}}}.$$

If the energy of the incoming photons is sufficiently low, $\gamma \ll 1$, then we get the classical formula

$$\sigma = \frac{2\lambda_e^2 \alpha^2}{3\pi} = 0.665 \cdot 10^{-28} \mathrm{m}^2$$

which was obtained by Joseph John Thomson at the end of the 19th century.[23] Observe that this classical approximation formula does not depend on the scattering angle ϑ. Physicists measure cross sections in barns. By definition, 1 barn $= 10^{-28} \mathrm{m}^2$.

Physical interpretation of the cross section. Let us now discuss the physical meaning of the cross section σ. Consider the situation pictured in Figure 2.10. We choose a sufficiently large sphere $\mathbb{S}^2_R$ of radius R about the scattering center. Let $\mathbf{n}$ and ΔS be the outer normal unit vector and the surface element of the sphere $\mathbb{S}^2_R$, respectively. The incoming photon stream can be described by the energy current density vector

[23] This can be found in the standard textbook on electrodynamics by Jackson (1975).

$$\mathbf{j}_{\text{in}} = \varrho_{\text{in}} \mathbf{v}_{\text{in}}$$

where ϱ_{in} and $\mathbf{v}_{\text{in}}$ denote the energy density and the velocity vector, respectively. In a typical experiment, the incoming photon stream is homogeneous. Therefore, we assume that the vector $\mathbf{j}_{\text{in}}$ is constant. By the scattering process, we obtain the outgoing energy current density vector field

$$\mathbf{j}_{\text{out}} = \varrho_{\text{out}} \mathbf{v}_{\text{out}}$$

which depends on the position vector $\mathbf{x}$, but not on time. Now to the point. The decisive quantity

$$E = (t_1 - t_0) \int_{\mathbb{S}^2_R} \mathbf{j}_{\text{out}} \mathbf{n} \, dS \tag{2.40}$$

is equal to the amount of outgoing energy that flows through the sphere $\mathbb{S}^2_R$ during the time interval $[t_0, t_1]$. This amount of energy can be measured by experiment. Naturally enough, E is proportional to $||\mathbf{j}_{\text{in}}||$ (incoming energy flow). The coefficient of proportionality σ defined by

$$\boxed{E = \sigma(t_1 - t_0) \, ||\mathbf{j}_{\text{in}}||} \tag{2.41}$$

has the physical dimension of area (m^2). Therefore, σ is called the total cross section of the scattering process. We want to show that there exists a function f such that

$$\boxed{\sigma = \int_{\mathbb{S}^2_R} f(\varphi, \vartheta) d\Omega.} \tag{2.42}$$

In fact, from (2.40) and (2.41) we get

$$\sigma = \int_{\mathbb{S}^2_R} \frac{\mathbf{j}_{\text{out}} \mathbf{n}}{||\mathbf{j}_{\text{in}}||} \, dS.$$

Naturally enough, the outgoing energy E does not depend on the choice of the radius R if the radius is sufficiently large. Because of the equality $dS = R^2 d\Omega$, we assume that the product $R^2 \mathbf{j}_{\text{out}}$ does not depend on R, and hence

$$\frac{(\mathbf{j}_{\text{out}} \mathbf{n})(P)}{||\mathbf{j}_{\text{in}}||} = f(\varphi, \vartheta), \qquad P \in \mathbb{S}^2_R$$

where φ and ϑ are the geographic longitude and the geographic latitude of the point P, respectively. This implies the desired formula (2.42).

Concerning Rutherford's experiment on the scattering of α-particles at protons (Fig. 2.8 on page 114), observe that it does not make any sense to consider the total cross section in this case, since the integral $\int_{\mathbb{S}^2_R} f d\Omega$

is divergent. Therefore, we need a localized version $\Delta\sigma$ of the cross section called the differential cross section. The idea is to consider a regular subset $\mathcal{S}$ of the sphere $\mathbb{S}^2_R$ that surrounds the point $P \in \mathbb{S}^2_R$. We now measure the scattered energy flow that passes through the part $\mathcal{S}$ of the sphere $\mathbb{S}^2_R$. The quantity

$$E(\mathcal{S}) := (t_1 - t_0) \int_{\mathcal{S}} \mathbf{j}_{\text{out}} \mathbf{n} \, dS$$

is equal to the amount of outgoing energy that flows through the part $\mathcal{S}$ of the sphere $\mathbb{S}^2_R$ during the time interval $[t_0, t_1]$. Similarly as above, we define the cross section $\sigma(\mathcal{S})$ with respect to $\mathcal{S}$ by the relation

$$\boxed{E(\mathcal{S}) = \sigma(\mathcal{S})(t_1 - t_0) \, ||\mathbf{j}_{\text{in}}||.}$$

Hence $\sigma(\mathcal{S}) = \int_{\mathcal{S}} f(\varphi, \vartheta) d\Omega$. Contracting the set $\mathcal{S}$ to the point $P \in \mathbb{S}^2_R$, we define

$$\frac{d\sigma}{d\Omega}(P) := \lim_{\mathcal{S} \to P} \frac{\sigma(\mathcal{S})}{\text{area } (\mathcal{S})} = f(\varphi, \vartheta).$$

Mnemonically, it is convenient to write

$$\boxed{\sigma(\mathcal{S}) = \int_{\mathcal{S}} d\sigma \quad \text{where} \quad d\sigma := f(\varphi, \vartheta) d\Omega.}$$

Compton effect and quantum electrodynamics. The Klein–Nishina formula (2.39) for the cross section of the Compton effect can be obtained from quantum electrodynamics. This highlight of quantum electrodynamics will be thoroughly studied in Volume II. Quantum electrodynamics represents a quantum field theory which describes the interactions between electrons, positrons, and photons. This is a perturbation theory with respect to the fine structure constant

$$\alpha = \frac{1}{137.04}.$$

The smallness of the fine structure constant is responsible for the great success of perturbation theory in quantum electrodynamics. We will show in Volume II that the Klein–Nishina formula follows from using second-order Feynman diagrams along with time-consuming computations based on Dirac matrices. Observe that the Klein–Nishina formula does not depend on the polarization of the photons. In fact, this formula averages over the polarizations of the incoming photons and sums over the polarizations of the outgoing photons.

General cross sections in elementary particle physics. The definition of total cross section σ and differential cross section $d\sigma$ introduced in (2.41ff) above applies to all types of scattering processes in physics. In particle accelerators, one defines cross sections with respect to the particle number. In this case, the incoming particle stream is described by the particle number current density vector,

$$\mathbf{j}_{\text{in}} = \varrho_{\text{in}} \mathbf{v}_{\text{in}}$$

where ϱ_{in} and $\mathbf{v}_{\text{in}}$ denote the particle density and the velocity vector, respectively. In definition (2.41) we then have to replace the energy E by the particle number N.

2.3.8 Bohr's Model of the Hydrogen Atom

> In 1922 Sommerfeld asked me whether I would be willing to follow him to a meeting in Göttingen at which Bohr would present his theory. These days in Göttingen we now always refer to the "Bohr festival". There for the first time I learned how a man like Bohr worked on problems of atomic physics.[24]
>
> Werner Heisenberg, 1968

Many properties of atoms, molecules, stars, and galaxies can be detected by measuring the energy spectrum of the electromagnetic radiation emitted from such objects. In 1884, the physicist Balmer tried to find a mathematical relation between the measured wave lengths λ in the spectrum of the hydrogen atom. By trial and guesswork, he found out that

$$\lambda = \text{const} \cdot \frac{m^2}{m^2 - 4} \quad \text{for} \quad m = 3, 4, 5.$$

This is a discrete sequence for the wave length. A few years later, Rydberg and Ritz postulated that all the possible values of the wave length observed in the spectrum of the hydrogen atom are given by the following empirical formula

$$\boxed{\lambda_{nm} = \frac{2\pi c}{\omega_{nm}}, \qquad n < m, \quad n, m = 1, 2, \ldots}$$

where

$$\omega_{nm} := R \left(\frac{1}{n^2} - \frac{1}{m^2} \right). \tag{2.43}$$

The experiment yields the value $R = 2.07 \cdot 10^{16} \text{s}^{-1}$; this is called the Rydberg frequency. In 1911, based on scattering experiments, Rutherford formulated the hypothesis that each atom consists of electrons which surround the nucleus like the planets surround the sun. In 1913 Bohr postulated the following for the hydrogen atom.

(i) The hydrogen atom consists of one electron which travels around the nucleus (one proton) in a circular orbit (Fig. 2.11(a)).

[24] Heisenberg's lecture about the history of quantum mechanics can be found in the collection of lectures edited by A. Salam (1968).

(a) bound state (b) scattering state

Fig. 2.11. Hydrogen atom

(ii) The angular momentum a of the electron is quantized, that is,

$$\boxed{a = n\hbar, \qquad n = 1, 2, \ldots}$$

where h is Planck's quantum of action and $\hbar := h/2\pi$.

(iii) If the electron jumps from an orbit with energy E_n to an orbit with lower energy E_m, the energy difference

$$\Delta E = E_n - E_m$$

is emitted by electromagnetic radiation. According to Einstein's light quanta hypothesis, the angular frequency ω of the emitted light quantum (photon) is given by

$$\Delta E = \hbar\omega.$$

This implies the fundamental energy-frequency relation

$$\boxed{\omega_{nm} = \frac{E_n - E_m}{\hbar}.} \tag{2.44}$$

From this we get the cocycle relation

$$\omega_{nm} + \omega_{mk} + \omega_{kn} = 0 \qquad \text{for all} \quad n, m, k = 0, 1, 2, \ldots.$$

From the mathematical point of view, the cocycle relation shows that the emitted frequencies of the hydrogen atom have a cohomological structure. This observation was the crucial starting point of Heisenberg's ingenious approach to quantum mechanics in 1924. In fact, Heisenberg spoke about the Ritz combination principle, but not on cohomology which was unknown at that time. As a crucial topological tool, cohomology only emerged in the 1930s.

In order to discuss Bohr's model, let E_n, r_n, and v_n denote energy, radius, and velocity of the nth orbit of the electron. We will show below that postulates (i) through (iii) imply

$$\boxed{E_n = -\frac{e^2}{8\pi\varepsilon_0 r_1} \cdot \frac{1}{n^2}, \qquad n = 1, 2, \ldots} \tag{2.45}$$

as well as

$$r_n = \frac{\lambda_e}{2\pi\alpha} \cdot n^2 \quad \text{and} \quad v_n = \frac{\alpha c}{n}, \qquad n = 1, 2, \dots$$

Recall the definition of both the fine structure constant

$$\alpha := \frac{e^2}{4\pi\varepsilon_0 \hbar c} = \frac{1}{137.04}$$

and the Compton wave length of the electron, $\lambda_e := \frac{h}{m_e c}$. Here, we use the following notations: e charge of the proton, $-e$ charge of the electron, m_p mass of the proton, m_e mass of the electron, ε_0 electric field constant of vacuum.

From (2.45) along with (iii) we obtain immediately the Rydberg–Ritz formula (2.43). For the smallest orbit we get the energy

$$E_1 = -13.6\text{eV}.$$

The electron is bound to the nucleus with this energy. Note that this energy is needed in order to ionize the hydrogen atom, that is, to eject the electron. Moreover, this gives the order of magnitude for the energies which occur in chemical reactions per atom. The radius of the smallest orbit

$$r_1 = 5 \cdot 10^{-11}\text{m}$$

is called the Bohr radius of the hydrogen atom. Recall that 1 nm $= 10^{-9}$m (nanometer) and 1 fm $= 10^{-15}$m (femtometer). The radius of the proton

$$r_{\text{proton}} = 1.4 \cdot 10^{-15}\text{m}$$

is much less than the Bohr radius. If we represent the proton by a pea of radius 5mm, then the electron would surround the pea at a distance of 30m. In addition, the mass quotient

$$\frac{m_p}{m_e} = 1836$$

shows that the proton is much heavier than the electron. The speed

$$v_1 = \alpha c = \frac{c}{137.04}$$

of the electron on the smallest orbit is much less than the velocity of light c. Therefore, relativistic effects do not play any role in Bohr's atomic model. Equivalently, formula (2.45) can be written as

$$\boxed{E_n = -\frac{m_e c^2 \alpha^2}{2n^2}, \qquad n = 1, 2, \dots}$$

Here, $m_e c^2$ is the rest energy of the electron.

Let us now prove the claim (2.45). The motion of the electron around the proton on a circle of radius r is described by the equation

$$\mathbf{x} = r(\cos\omega t\,\mathbf{i} + \sin\omega t\,\mathbf{j})$$

where **i, j** are fixed right-handed orthonormal vectors. Hence

$$\dot{\mathbf{x}}(t) = \omega r(-\sin\omega t\,\mathbf{i} + \cos\omega t\,\mathbf{j})$$

and

$$\ddot{\mathbf{x}}(t) = -\omega^2 r(\cos\omega t\,\mathbf{i} + \sin\omega t\,\mathbf{j}).$$

As usual, the dot denotes the time derivative. For the angular momentum of the electron, we obtain

$$\mathbf{L} = m_e(\mathbf{x}\times\dot{\mathbf{x}}) = m_e\omega r^2\mathbf{k}$$

where $\mathbf{k} := \mathbf{i}\times\mathbf{j}$. There act two attractive forces between the proton and the electron, namely, Newton's gravitational force

$$\mathbf{F}_N = -\frac{Gm_pm_e}{r^2}\cdot\frac{\mathbf{x}}{r}$$

and the electrostatic Coulomb force

$$\mathbf{F}_C = -\frac{e^2}{4\pi\varepsilon_0 r^2}\cdot\frac{\mathbf{x}}{r}.$$

Here $r := ||\mathbf{x}||$. Since $||\mathbf{F}_N||/||\mathbf{F}_C|| = 5\cdot 10^{-42}$, the gravitational force can be neglected. Thus, the motion of the electron is governed by the Newton equation

$$m_e\ddot{\mathbf{x}} = \mathbf{F}_C.$$

Hence

$$m_e\omega^2 r = \frac{e^2}{4\pi\varepsilon_0 r^2}. \tag{2.46}$$

The energy is equal to

$$E = \frac{m_e}{2}\dot{\mathbf{x}}^2 + U = \frac{m_e\omega^2 r^2}{2} - \frac{e^2}{4\pi\varepsilon_0 r} = -\frac{e^2}{8\pi\varepsilon_0 r}. \tag{2.47}$$

Bohr postulated that the angular momentum only attains the following quantized values

$$||\mathbf{L}|| = n\hbar, \qquad n = 1, 2, \ldots$$

This gives $m_e\omega r^2 = n\hbar$ for $n = 1, 2, \ldots$ From (2.46) we obtain the orbit radii

$$r_n = r_1 n^2, \qquad n = 1, 2, \ldots$$

with $r_1 := 4\pi\varepsilon_0\hbar^2/m_e e^2 = \lambda_e/2\pi\alpha$. The energy is given by

$$E = -\frac{e^2}{8\pi\varepsilon_0 r_n} = -\frac{e^2}{8\pi\varepsilon_0 r_1 n^2},$$

and the orbital velocity is equal to $v_n = ||\dot{\mathbf{x}}|| = \omega r_n = \alpha c/n$.

The Bohr model described the observed spectrum of the hydrogen atom in a perfect manner. However, physicists did not understand this model for the following reason. In terms of classical electrodynamics, the rotating electron represents an accelerated electric charge. Such a charge steadily loses energy by emitting electromagnetic radiation. Therefore, the electron cannot move on stable orbits. In addition, classical physics cannot explain why the electron jumps spontaneously from one orbit to another one. This fundamental problem was solved by Heisenberg in 1925 and Schrödinger in 1926, by introducing quantum mechanics, namely, Heisenberg's matrix mechanics and Schrödinger's wave mechanics, respectively. In terms of Schrödinger's wave mechanics, the wave function of the electron of the hydrogen atom performs eigenoscillations which correspond to Bohr's semiclassical circular orbits. In 1927 Heisenberg discovered the crucial quantum-mechanical uncertainty relation which tells us that

The electron is a quantum particle for which the classical notions of position and velocity do not make sense anymore.

2.3.9 Einstein's Radiation Law and Laser Beams

In 1917 Einstein wrote a fundamental paper on the quantum theory of radiation.[25] Let us explain the basic ideas. Consider N atoms where each of them can be in a state of energy $E_0, E_1, E_2, \ldots$ There exist the following three fundamental processes.

(i) Spontaneous emission: The atom passes spontaneously from the energy E_n to the lower energy E_m by emitting a photon of angular frequency

$$\omega = \frac{E_n - E_m}{\hbar}\,. \tag{2.48}$$

Here, we assume that $E_n > E_m$.

(ii) Stimulated absorption: Suppose that the atom is embedded into an external electromagnetic field of energy density $\eta(T)$. This stimulates the atom to jump from the energy E_m to the higher energy E_n by absorbing a photon of angular frequency (2.48).

(iii) Stimulated emission: The energy density $\eta(T)$ also stimulates jumps of the atom from the energy E_n to the lower energy E_m. In this case the atom emits a photon of angular frequency (2.48).

[25] A. Einstein, On the quantum theory of radiation (in German), Phys. Z. **18** (1917), 121–135. The English translation of this paper can be found in van der Waerden (1968), pp. 63–68.

We assume that the energy density $\eta(T)$ refers to the absolute temperature T and to electromagnetic waves whose angular frequencies lie in the small interval $[\omega, \omega + \Delta\omega]$. In order to describe the situation quantitatively, consider a large system of N atoms at absolute temperature T. According to Boltzmann statistics, Einstein assumed that at each fixed time t, the number of atoms of energy E_n is equal to

$$N_n(t) = N\mathrm{e}^{-E_n/kT}.$$

At the later time $t + \Delta t$, we get

$$\begin{aligned} N_n(t+\Delta t) = N_n(t) - N_n(t)(\omega_{nm}\Delta\omega)\Delta t - \\ -N_n(t)(\omega_{nm}^{\mathrm{stim}}\eta(T))\Delta t + N_m(t)(\mathcal{A}_{nm}^{\mathrm{stim}}\eta(T))\Delta t, \end{aligned}$$

up to higher-order terms of the form $o(\Delta t)$ as $\Delta t \to 0$. The coefficient $\mathcal{E}_{nm}$ describes the spontaneous emission of photons whereas $\mathcal{E}_{nm}^{\mathrm{stim}}$ and $\mathcal{A}_{nm}^{\mathrm{stim}}$ describe the stimulated emission and absorption of photons, respectively. In thermodynamic equilibrium, Einstein assumed that the energy jumps of the atom from E_n to E_m and from E_m to E_n do not change the number of atoms at the energy level E_n. Hence

$$N_n(t + \Delta t) = N_n(t).$$

This implies the Einstein relation

$$N_n(t)\big(\mathcal{E}_{nm}\Delta\omega + \eta(T)\,\mathcal{E}_{nm}^{\mathrm{stim}}\big) = N_m(t)\eta(T)\,\mathcal{A}_{nm}^{\mathrm{stim}}.$$

Hence

$$\mathrm{e}^{-(E_n-E_m)/kT}\left(\frac{\mathcal{E}_{nm}\Delta\omega}{\eta(T)} + \mathcal{E}_{nm}^{\mathrm{stim}}\right) = \mathcal{A}_{nm}^{\mathrm{stim}}.$$

Introducing the angular frequency ω of the emitted photon, $E_n - E_m = \hbar\omega$, we get

$$\mathrm{e}^{-\hbar\omega/kT}\left(\frac{\mathcal{E}_{nm}\Delta\omega}{\eta(T)} + \mathcal{E}_{nm}^{\mathrm{stim}}\right) = \mathcal{A}_{nm}^{\mathrm{stim}}. \tag{2.49}$$

Assuming that $\eta(T) \to +\infty$ as $T \to +\infty$, we obtain the Einstein relation

$$\boxed{\mathcal{E}_{nm}^{\mathrm{stim}} = \mathcal{A}_{mn}^{\mathrm{stim}}.}$$

This tells us that the coefficients of stimulated emission and stimulated absorption coincide. Moreover, for the energy density $\eta(T)$ of the external electromagnetic radiation, it follows from (2.49) that

$$\eta(T) = \frac{\mathcal{E}_{nm}\Delta\omega}{\mathcal{E}_{nm}^{\mathrm{stim}}\left(\mathrm{e}^{\hbar\omega/kT} - 1\right)}. \tag{2.50}$$

Let us compare this with the energy density of photons following from Planck's radiation law (2.24) on page 103,

$$\eta(T) = \mathcal{P}(\omega, T)\Delta\omega = \frac{\hbar\omega^3 \Delta\omega}{\pi^2 c^3 \left(\mathrm{e}^{\hbar\omega/kT} - 1\right)}. \tag{2.51}$$

From (2.50) and (2.51) we get

$$\boxed{\mathcal{E}_{nm} = \frac{\hbar\omega^3}{\pi^2 c^3} \cdot \mathcal{E}_{nm}^{\mathrm{stim}}.} \tag{2.52}$$

This equation shows the relation between spontaneous and stimulated emission of photons of angular frequency ω.

The term "laser" stands for "light amplification by stimulated emission of radiation." The first lasers were realized experimentally around 1960. In 1964 Basov, Prochorov, and Townes were awarded the Nobel prize in physics for "fundamental work in the field of quantum electronics, which has led to the construction of oscillators and amplifiers based on the maser-laser principle." Theoretically, the laser is based on Einstein's 1917 paper discussed above.

In 1927 Dirac wrote a fundamental paper on emission and absorption of radiation where he showed that Einstein's radiation law follows from the new quantum theory due to Heisenberg and Schrödinger from 1925 and 1926, respectively. Dirac's paper is part of a collection edited by Schwinger (1958) which contains the most important papers in quantum electrodynamics. See also the classic textbooks by Dirac (1930) and Heitler (1936). The modern quantum radiation theory can be found in Pike and Sarkar (1995) and Mandel and Wolf (1995) (laser theory).

2.3.10 Quantum Computers

In order to speed up computers, physicists are thinking about the realization of quantum computers based on the laws of quantum physics. The basic idea is to use the fact that there exist superpositions of quantum states φ and ψ given by

$$\alpha\varphi + \beta\psi$$

where α and β are complex numbers. This leads to 'quantum bits' called

$$qbits$$

which transport an essentially higher amount of information than can be done by using traditional 'bits'. We will study this in Volume IV on quantum mathematics. Many strange properties of quantum information are based on the existence of composed entangled states

$$\varphi \otimes \psi \pm \psi \otimes \varphi$$

where the single states φ and ψ lose their individuality. As an introduction to quantum information, we recommend Boumeester, Ekert, and Zeilinger (2000), Nielsen and Chuang (2001), and Heiss (2002).

Table 2.1. Fundamental forces in nature

Interactions				
force	strong	electroweak		gravitation
		electromagnetic	weak	
acting on	quarks and gluons	electric charges	leptons, quarks, and $W^{\pm}$, Z^0 bosons	masses, photons, and gravitons
relative strength	1	10^{-2}	10^{-5}	10^{-38}
range	10^{-15}m	∞	10^{-18}m	∞

2.4 The Standard Model in Particle Physics

> I do not know what I may appear to the world, but to myself I seem to have been only like a boy playing on the seashore, and diverting myself now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me.
>
> Isaac Newton (1643–1727)

It is our goal to study the basic ideas of the Standard Model in terms of physics.

2.4.1 The Four Fundamental Forces in Nature

In the 20th century, physicists used radioactive decay, cosmic rays, and more and more powerful particle accelerators in order to get a fairly complete picture of the structure of matter. Nowadays physicists assume that there exist the following four fundamental forces in nature.

(i) *Strong force* (e.g., the proton as a bound state of three quarks).
(ii) *Electromagnetic force* (e.g., the chemical binding of molecules).
(iii) *Weak force* (e.g., the β-decay of the neutron).
(iv) *Gravitational force* (e.g., the motion of planets around the sun, the expansion of the universe, and black holes).

Some important properties of these forces are summarized in Table 2.1. As we will see below, the electromagnetic force and the weak force are part of a unified force called the electroweak force.

Strong force. The nuclear force is responsible for the stability of the proton and for the relative stability of the neutron and the atomic nucleus, which consists of protons and neutrons (nucleons). The range of the strong force is equal to the diameter of the proton, namely, 1 fermi = 10^{-15}m.

Table 2.2. Typical experimental data

interaction	typical cross section in millibarn ($10^{-31}\mathrm{m}^2$)	typical mean energy of resonances in MeV	typical mean lifetime of resonances in seconds
strong	1	10^2	10^{-23}
electromagnetic	10^{-3}	10^{-3}	10^{-18}
weak	10^{-9}	10^{-14}	10^{-7}

The proton consists of two u quarks and one d quark, whereas the neutron consists of two d quarks and one u quark. The proton p is the nucleus of the hydrogen atom. The neutron was discovered experimentally by Chadwick in 1932 (Nobel prize in physics in 1935). In 1932 Ivanenko predicted that the nucleus consists of protons and neutrons.

Electromagnetic force. The electromagnetic force is responsible for the stability of atoms and molecules by acting on the protons and electrons of the atoms. All physical and chemical properties of solid states, liquids, and gases are based on the electromagnetic force. The range of the electromagnetic force is infinite. The electron was discovered by Thomson in 1895 who investigated cathode rays (Nobel prize in physics in 1906). If electrons hit a metal, electromagnetic rays of high energy are generated. These so-called X-rays were discovered by Röntgen in 1895 (first Nobel prize in physics in 1901). A semiclassical model of the atom was formulated by Bohr in 1913 (Nobel prize in physics in 1922). Bohr was strongly influenced by Rutherford's scattering experiments performed in the years 1909–1911. The final model of the atom was based on quantum mechanics invented by Heisenberg in 1925 (Nobel prize in physics in 1932) and Schrödinger in 1926 (Nobel prize in physics together with Dirac in 1933). The shell structure of the atom can only be understood by using the electron spin and Pauli's exclusion principle for fermions (e.g., electrons) from 1926. Pauli was awarded the Nobel prize in physics in 1945.

Weak force. This force is responsible for the radioactive decay of atoms discovered by Bequerel in 1892 (Nobel prize in physics together with Marie and Pierre Curie in 1903). The basic reaction is the β-decay of the neutron,

$$n \to p + e^- + \overline{\nu}_e.$$

That is, the neutron n decays into one proton p, one electron e^-, and one anti-electron neutrino $\overline{\nu}_e$. The mean lifetime of the neutron is 15 minutes. Experiments show that the mean lifetime of the proton is larger than 10^{32} years. This is an incredibly huge number.[26] The existence of the neutrino was

[26] Note that the age of our universe is $13.7 \cdot 10^9$ years.

predicted by Pauli in 1933 in order to guarantee momentum conservation in the neutron decay. Radioactive decay and the chemical properties of radioactive substances were studied around 1900 by Marie Curie (Nobel prizes in physics and chemistry in 1903 and 1911, respectively), Pierre Curie (Nobel prize in physics in 1903), and Rutherford (Nobel prize in physics in 1908). These three scientists found out that radioactive decay generates three types of rays:

- α-rays (helium nuclei ${}_2^4\mathrm{He}$),
- β-rays (fast electrons),
- γ-rays (high-energy photons).

Visible light has a wave length between $4 \cdot 10^{-7}$m and $8 \cdot 10^{-7}$m. In contrast to this, X-rays and γ-rays have a short wave length of 10^{-10}m and 10^{-13}m, respectively. The energy of X-rays and γ-rays is much stronger than the energy of light rays. Note that the energy of photons increases if the wave length decreases. Cosmic rays were discovered by Hess in 1911 (Nobel prize in physics in 1936). In 1928 Gamow explained the production of α-rays. If α-particles would be classical particles, they could not leave the nucleus because of the existence of a strong potential barrier. Therefore, it is crucial that α-particles are quantum particles. They possess stochastic properties. In particular, they can leave the nucleus by "tunnelling" the potential barrier. In 1934 Fermi used slow neutrons in order to produce new radioactive elements containing a large number of nucleons. Fermi was awarded the Nobel prize in physics in 1938.

The neutrino was experimentally discovered outside a nuclear reactor in 1956. At the Savannah River reactor (Georgia, U.S.A.), the number of neutrinos emerging per second was extremely high, and physicists waited with their detector until they eventually detected some. Other neutrino sources are our sun and outbursts of supernovae. Note that the energy production of the sun is based on a series of nuclear reactions which converts hydrogen into helium. These reactions start by the process

$$p + p \to {}_1^2\mathrm{D} + e^+ + \nu_e.$$

Here, the fusion of two protons yields one deuteron ${}_1^2\mathrm{D}$ plus one positron e^+ and one electron neutrino ν_e. This process is caused by the weak force.[27] For his theory of the energy production in stars, Bethe was awarded the Nobel prize in physics in 1967. Neutrinos coming from the sun were detected by Davies in 1970. In 1987, a supernova explosion took place about 160 000

[27] The nucleus of an atom consists of protons and neutrons called nucleons. The symbol ${}_Z^N\mathrm{A}$ stands for the nucleus of an atom that consists of N nucleons and Z protons. Hence the number of neutrons is equal to $N - Z$.

The symbol e^- (resp. e^+) tells us that the electron (resp. positron) has the negative (resp. positive) elementary electric charge $-e$ (resp. e). Similarly, Z^0 tells us that the Z-boson has no electric charge.

light years away in the Magellanic Cloud next to our Milky Way galaxy. This star had 8 sun masses. The released energy was enormous,

$$E = 10^{46}\text{J}.$$

More then 99% of energy came out in invisible form – as neutrinos – based on the reaction

$$e^- + p \to n + \nu_e.$$

Two experimental groups in the United States and Japan reported detecting neutrinos at the time of the supernova. The experimental detection of neutrinos is a highly nontrivial task, since their interaction with matter is extremely weak.

Phase transitions of the fundamental forces in the early universe. Many physicists assume that there was only one fundamental force at the time of the Big Bang. The cooling of the universe was responsible for phase transitions of this fundamental force which led to a splitting into the four fundamental forces observed in nature today.

2.4.2 The Fundamental Particles in Nature

Basic ideas of the Standard Model. The Standard Model in elementary particle physics concerns the strong, weak, and electromagnetic force. The main idea of the Standard Model is the following.

(a) *Fundamental particles.* There exist precisely 12 fundamental particles in nature, namely, the 6 quarks u, d, c, s, t, b and the 6 leptons

$$e^- \text{ (electron)},\ \mu^- \text{ (muon)},\ \tau^- \text{ (tauon)},\ \nu_e,\ \nu_\mu,\ \nu_\tau \text{ (3 neutrinos)}.$$

These 12 fundamental particles are fermions (i.e., they have half-integer spin, $\frac{1}{2}\hbar$.) For the six quarks u, d, c, s, t, b, physicists invented fancy names. They call them up, down, charm, strange, top, and bottom quarks, respectively. The 12 fundamental particles are divided into three generations:
First generation: u, d, e^-, ν_e.
Second generation: c, s, μ^-, ν_μ.
Third generation: t, b, τ^-, ν_τ.

(b) *Messenger particles.* The 12 fundamental particles experience the three fundamental forces (strong, electromagnetic, and weak) by the exchange of 12 messenger particles, namely,

$$\gamma \text{ (the photon)},\ W^+, W^-, Z^0 \text{ (3 weak gauge bosons)},\ \text{ and 8 gluons.}$$

These 12 messenger particles are bosons (integer spin, 0, $\hbar$). The 8 gluons provide the "glue" for keeping the quarks together.

Table 2.3. Fundamental particles in the Standard Model

leptons			
particle	mass MeV/c^2	electric charge	spin
electron-neutrino ν_e	$<15{\cdot}10^{-6}$	0	$\frac{1}{2}\hbar$
electron e^-	0.511	$-e$	$\frac{1}{2}\hbar$
myon-neutrino ν_μ	<0.17	0	$\frac{1}{2}\hbar$
myon μ^-	105.7	$-e$	$\frac{1}{2}\hbar$
tauon-neutrino ν_τ	<24	0	$\frac{1}{2}\hbar$
tauon τ^-	1 777	$-e$	$\frac{1}{2}\hbar$

quarks			
particle	mass MeV/c^2	electric charge	spin
up u	$2-8$	$\frac{2}{3}e$	$\frac{1}{2}\hbar$
down d	$5-15$	$-\frac{1}{3}e$	$\frac{1}{2}\hbar$
charm c	$1000-1600$	$\frac{2}{3}e$	$\frac{1}{2}\hbar$
strange s	$100-300$	$-\frac{1}{3}e$	$\frac{1}{2}\hbar$
top t	$180\,000$ $\pm 12\,000$	$\frac{2}{3}e$	$\frac{1}{2}\hbar$
bottom b	$4100-4500$	$-\frac{1}{3}e$	$\frac{1}{2}\hbar$

(c) *Antiparticles.* To each fundamental particle, there exists an antiparticle of same mass and opposite electric charge. The 12 antiparticles are denoted in the following way.
First generation: $\overline{u}, \overline{d}$ (antiquarks), e^+ (positron), $\overline{\nu}_e$ (anti-electron neutrino).
Second generation: $\overline{c}, \overline{s}$ (antiquarks), μ^+ (antimuon), $\overline{\nu}_\mu$ (anti-muon neutrino).
Third generation: $\overline{t}, \overline{b}$ (antiquarks), τ^+ (antitauon), $\overline{\nu}_\tau$ (anti-tau neutrino).

It is a typical property of our universe that matter highly dominates antimatter. In 1928 Dirac used his relativistic equation for the electron in order to predict the existence of the antiparticle e^+ to the electron e^-. This particle (called positron) was experimentally discovered in a Wilson cloud chamber in 1932 by Anderson (Nobel prize in physics in 1936). The antiproton $\overline{p}$ was discovered by Chamberlain and Segré in 1955 (Nobel prize in physics in 1959), and the antineutron was found by Cook in 1956.

Let us now discuss some more details of the Standard Model in particle physics. A thorough investigation of the physics of the Standard Model can be found in Volume V.

Strong force and the eight gluons. As a typical example, the proton p consists of two u quarks and one d quark. Symbolically, $p = uud$. The rest mass m_0 of a proton is equal to

Table 2.4. Messenger particles in the Standard Model

electroweak force			
particle	mass GeV/c^2	electric charge	spin
photon	0	0	$\hbar$
W^-	80.3	$-e$	$\hbar$
W^+	80.3	e	$\hbar$
Z^0	91.2	0	$\hbar$

strong force			
particle	mass	electric charge	spin
8 gluons	0	0	0

$$m_0 = 1.67 \cdot 10^{-27} \text{ kg}.$$

This corresponds to the rest energy $E = m_0 c^2 = 1.5 \cdot 10^{-18}$J. Physicists like to measure energies in eV (electron volt). In this energy scale, the rest energy of the proton is equal to

$$E = 0.938 \cdot 10^9 \text{ eV}.$$

As a rule of thumb, the rest energy of a proton is equal to 1 GeV (giga electron volt), and this equals the rest energy of 1836 electrons. Physicists use the prefixes giga, mega, nano, femto for the corresponding factors $10^9, 10^6, 10^{-9}, 10^{-15}$. For more information, see Table A.1 on page 935.

In 2008, the particle accelerator at CERN (Geneva, Switzerland)[28] will reach particle energies which equal the rest energy of 7 000 protons (7TeV). The radius of the proton is equal to 10^{-15}m = 1 fermi. Experiments show that electrons and quarks have a radius which is less than 0.001 fermi. In fact, nowadays the electron and the six quarks are considered to be point-like particles. Observe that most space of the proton is filled with massless gluons. Each quark has three charges called red, green, and blue by physicists. The gluons see the color charge. The action of gluons onto the color charge causes the strong force. There are 8 gluons. As we will see later on, this depends on the fact that the dimension of the Lie algebra su(3) is equal to 8. Since the strong force is based on the color charge, the theory of the strong force is called quantum chromodynamics. An atom has a radius of 10^{-10}m = 100 000 fermi. Therefore, the strong force does not play any role for the interaction between the electrons and the nucleus of an atom.

[28] CERN stands for *Conseil Européen pour la Recherche Nucléaire* (European Organization for Nuclear Research at Geneva); this was founded in 1953.

The total rest energy of the three quarks of the proton is equal to 30MeV. Consider the rough form of Heisenberg's uncertainty relation

$$\Delta x \cdot \Delta p \sim \hbar.$$

The quark moves in a proton of radius $\Delta x = 10^{-15}$m. Therefore, the quark has the momentum $\Delta p = \hbar/\Delta x$, and the total energy

$$E = \sqrt{m_0^2 c^4 + c^2(\Delta p)^2} = 200 \text{ MeV}.$$

Hence the total energy of the three quarks of the proton is equal to 0.6 GeV. The remaining energy 0.4 GeV of the proton corresponds to the motion of gluons. This means that the proton has an extremely large binding energy of quarks which cannot be computed with the usual methods of perturbation theory. Physicists use highly specialized supercomputers in order to compute the binding energy of the proton, the neutron, and mesons. Mathematically, the computations are based on grid models in the framework of gauge lattice theory. The results are in good agreement with experiments.

Baryons and mesons as composite particles. An elementary particle is called a hadron iff it experiences the strong force. Quarks are elementary hadrons. Concerning composite hadrons, one has to distinguish between baryons and mesons.

- Each baryon consists of three quarks (e.g., the proton, the neutron, the lambda, and the sigma). Baryons have half-integer spin.
- Each meson consists of quark-antiquark pairs (e.g., the three mesons π^+, π^0, π^- or the kaon K^0). Mesons have integer spin.

The existence of mesons, which have about 250 electron masses (or 1/7 of the proton mass), was predicted by Yukawa in 1935 (Nobel prize in physics in 1949). Experimentally, mesons were discovered in cosmic rays by Powell in 1947 (Nobel prize in physics in 1950). Baryons and mesons are white, that is, the color charges red, green, and blue neutralize each other so that we cannot see the color charges. Only the gluons can see the color charges. The existence of gluons was experimentally verified at the DESY accelerator PETRA (Hamburg, Germany) in 1979.

Electroweak force, the photon, and the three weak gauge bosons. Maxwell's classical theory of electromagnetism from 1864 unified the electric force with the magnetic force into the electromagnetic force. In the Standard Model of particle physics, the electromagnetic force is unified with the weak force. This yields the so-called electroweak force. The corresponding messenger particles are the photon γ and the three weak gauge bosons W^-, W^+, and Z^0. As we will show later on, the appearance of 4 bosons depends on the fact that the dimension of the Lie algebra u(1)×su(2) is equal to 4. The process

$$d \to u + e^- + \overline{\nu}_e \tag{2.53}$$

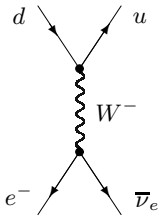


Fig. 2.12. Beta decay of the d quark

describes the decay of a d quark into a u quark, an electron, and an anti-electron neutrino. This decay is caused by the exchange of a W^--boson (Fig. 2.12). Since we have $n = ddu$ and $p = uud$ for the neutron and proton, respectively, the process (2.53) is responsible for the crucial beta decay of the neutron $n \to p + e^- + \overline{\nu}_e$.

Lifetime of elementary particles. As a rule of thumb, physicists use the formula

$$\boxed{\Delta t = \frac{\hbar}{m_0 c^2}}$$

for the lifetime, Δt, of an elementary particle. This is a consequence of the energy-time uncertainty mentioned on page 142. Here, h, c, and m_0 denote Planck's constant, the velocity of light in a vacuum, and the rest mass of the particle, respectively. Recall that $\hbar := h/2\pi$. In particular, massless particles like the photon, the gluon, and the graviton have an infinite lifetime. If a messenger particle has the lifetime Δt, then it can move the distance r during its lifetime. This tells us that the range of the corresponding force is equal to

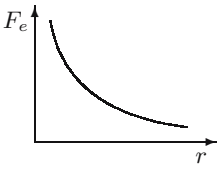
$$r = c \cdot \Delta t = \frac{\hbar}{m_0 c}.$$

This is equal to the so-called reduced Compton wave length $\lambda\!\!\!^{-}_{\mathrm{C}}$ of the particle. Since the weak gauge bosons $W^\pm$ and Z^0 have a rest mass of about 100 GeV/c^2, their lifetime is 10^{-23}s, and the range of the weak force is less than 0.01 fermi = 10^{-17}m.

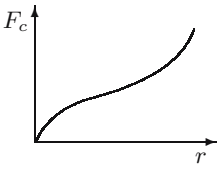
The quark confinement. Note that there exists a fundamental difference between the electromagnetic force and the strong force (Fig. 2.13).

- The electromagnetic force vanishes for large distances and it goes to infinity if the distance goes to zero.
- In contrast to this, the strong force vanishes if the distance between the quarks goes to zero and it becomes infinite if the distance between the quarks goes to infinity.

This implies the crucial fact that the quarks behave like free particles for small distances less than 0.2 fermi = $0.2 \cdot 10^{-15}$m (asymptotic freedom). For



(a) strength of electric force F_e (r distance)



(b) strength of color force F_c

Fig. 2.13. The quark confinement

distances more than one fermi between the quarks, the strong force is very large. This property of the strong force is responsible for the fact that *free* quarks have never been observed. This is the so-called quark confinement. A complete theoretical understanding of the quark confinement is still missing.

History of the Standard Model of particle physics. The quark hypothesis was formulated by Gell-Mann in 1964 (Nobel prize in physics in 1969). A similar theory was independently proposed by Zweig in 1964. In Zweig's approach the quarks were called aces. In the very beginning of his theory, Gell-Mann was not sure whether the quarks are merely mathematical constructions (based on the representation theory of the group $SU(3)$) or real physical objects. The breakthrough came from physical experiments. In 1968, deeply inelastic electron-proton scattering experiments were performed at SLAC of Stanford University (California, U.S.A.). These experiments established that the proton possesses an internal structure which corresponds to a decomposition of the proton into three quarks.

The first theory of the weak force (β-decay) dates back to Fermi in 1933. This model worked successfully for fairly low energies. In particular, Fermi was able to compute the cross sections for neutrino reactions. Since the infinities of Fermi's quantum field theory could not be renormalized, physicists were looking for an improved theory. In 1967 and 1968, Weinberg and Salam, respectively, formulated independently a model which unified the weak and electromagnetic force. They based their models in part on work developed by Glashow in 1961. Therefore, this model is called the Glashow–Salam–Weinberg model (in 1979 Glashow, Salam, and Weinberg were awarded the Nobel prize in physics). The sophisticated renormalization of this model was shown by 't Hooft in 1971 (Nobel prize in physics together with Veltman in 1999). He used mathematical tools developed for Feynman integrals by Faddeev and Popov in the 1960s (cancellation of ghosts by factorizing with respect to gauge orbits). In 1974, the reaction

$$p + \nu_\mu \to n + \pi^+ + \nu_\mu$$

was observed at Argonne National Laboratory. This process is based on the exchange of an electrically neutral Z^0 boson, predicted by the Glashow–

Salam–Weinberg model. This model was finally established experimentally by the discovery of the three weak gauge bosons $W^{\pm}$ and Z^0 at the CERN particle accelerator (Geneva, Switzerland) in 1983. This experiment needs very high energies, since the W and Z bosons have a rest mass of approximately 100 proton masses. Rubbia and van der Meer were awarded the 1984 Nobel prize in physics for performing this fundamental experiment, together with a large group of experimentalists at the CERN proton-antiproton collider. The particles of the Standard Model were discovered in the following years:

- Free leptons: electron (1895), muon (1937), electron neutrino (1956), muon neutrino (1961), tauon (1975), tauon neutrino (1975).
- Bound quarks: u, d, s (1970), c (1974), b (1977), t (1994).
- Messenger particles: photon (1922), gluons (1979), three weak gauge bosons (1983).
- Composite particles: proton (1914), neutron (1932), π-meson (1947), and J/φ-meson (1974).
- Antiparticles: anti-electron (positron) (1932), antiproton (1955), antineutron (1956).

These particles have the following lifetimes: photon and gluon (∞), u quark and proton ($> 10^{32}$ years), electron ($> 10^{23}$ years), d quark and neutron (887s), muon (10^{-6}s), s quark (10^{-8}s), c and b quark (10^{-12}s), Z and $W^{\pm}$ bosons (10^{-25}s), t quark (10^{-25}s). Nowadays physicists know about 80 composite particles.

Gravitational waves and the graviton. If we assume that all of the fundamental forces in nature are based on the exchange of messenger particles, then we have to postulate the existence of an additional messenger particle called graviton which is responsible for the gravitational force. This is not a pure speculation. Let us discuss this. In 1974 Hulse and Taylor observed the pulsar PSR 1913+16 which has a distance of 20 000 light years from earth (Hulse and Taylor were awarded the Nobel prize in physics in 1993). This pulsar consists of two neutron stars. Each of them has 1.4 sun masses and a diameter of approximately 20 km. This means that the mass density is very large. The pulsation period of 0.0590299952709 seconds undergoes a periodic change because of the companion star. This is one of the stablest clocks in the universe. The two stars slowly approach each other because of a loss of gravitational energy due to gravitational radiation. On the basis of a post-Newtonian approximation to general relativity, computations verify a number of predictions, including the formula for the energy loss from a binary system due to gravitational radiation.[29] Observe that Hulse and Taylor established the existence of gravitational waves only in an indirect manner. In the near future, physicists will perform highly sensitive laser experiments in order to prove directly the existence of gravitational waves.

[29] This can be found in Straumann (2004).

(a) point particle (b) string

Fig. 2.14. Motion of particles and strings

The idea is the following. If a gravitational wave hits a mirror, then the small change of the mirror position can be observed by the small deflection of a laser beam of length between 500 meters (in Germany) and some kilometers (the LIGO (Laser Interferometer Gravitational-Wave Observatory) project in the United States of America). In the future, it is planned to use spacelabs where the effective distance is about five million kilometers (the LISA (Laser Interferometer Space Antenna) project to be launched by NASA and ESA in about 2011).[30] Right now extensive computer simulations are being performed in order to understand the pattern of the gravitational waves caused by supernova explosions or the collision of two black holes (resp. two collapsing binary neutron stars). Einstein's theory of general relativity predicts that gravitational waves propagate with the speed of light and they have two different directions of polarization. Therefore, the hypothetical massless graviton of spin 2 should propagate with the speed of light.

String theory and the graviton. In the 1970s a true revolution took place in the thinking of theoretical physicists. Up to this time, it was assumed that the fundamental constituents of matter are particles. In contrast to this, modern string theory is based on the following fascinating hypothesis:

> *Elementary particles are not point-particles, but they are tiny strings living below the Planck length* $l = 10^{-35}$m.

The motion of a point-particle (resp. string) corresponds to a 1-dimensional world-line (resp. 2-dimensional world-sheet) (Fig. 2.14). There exists a very rich mathematical theory of 2-dimensional surfaces called Riemann surfaces. It is typical for Riemann surfaces that they possess a conformal structure. This explains why conformal field theory is closely related to string theory. Note the following important fact:

> *The larger the symmetry of a physical system is, the more information about the structure of the system can be obtained from the mathematics of the relevant symmetry group.*

In contrast to other dimensions, the two-dimensional (local) conformal group is huge. This is reflected by the richness of the classical theory of analytic functions on the complex plane. For example, in conformal field theory, the

[30] Details can be found in the monograph by Schutz (2003), Chap. 22.

structure of the fundamental Green's function is mainly determined by the conformal symmetry. Surprisingly enough, each string theory contains a particle of spin 2 which can be identified with the graviton. String theory is the most promising candidate for a unified theory of all four interactions in nature. However, one should also note that there is no experimental evidence for strings so far. It is expected that typical string effects can be only observed at extremely high energies. Therefore, one is looking for indirect effects which can be observed at much lower energy ranges as virtual particles or as a relic of the Big Bang (cosmic strings, magnetic monopoles, dark matter, dark energy, and so on). As an introduction to string theory, we recommend Lüst and Theissen (1989) and Zwiebach (2004). Moreover, we refer to the standard textbooks by Green, Schwarz and Witten (1987), Vols. 1, 2, and Polchinski (1998), Vols. 1, 2. The history of string theory can be found in Greene (1999).

Supersymmetry. Physicists assume that there exists a perfect symmetry between fermions and bosons at extremely high energies. This means that, for each fermion there exists precisely one boson called the supersymmetric partner of the fermion. For example, the supersymmetric partners of electrons, quarks, photons, and gravitons are called electrinos, quarkinos, photinos, and gravitinos, respectively. Note that this so-called supersymmetry is not observed in our real world today. Physicists assume that perfect supersymmetry did exist only shortly after the Big Bang at extremely high energies. However, physicists expect that the particle accelerators of the next generation will be able to prove the existence of supersymmetric particles. The relevant calculations have been already performed in the framework of the so-called minimal supersymmetric Standard Model. For the renormalization of the minimal supersymmetric Standard Model see Hollik et al. (2002). As an introduction to supersymmetry, we recommend Martin (1997) (a supersymmetry primer), Bailin and Love (1997), Kalka (1997), and Kane (2000). We also refer to Wess and Bagger (1991) and Weinberg (1995), Vol. 3.

The Higgs particles. In gauge theories, the messenger particles are massless for mathematical reasons. In sharp contrast to this, the gauge bosons $W^{\pm}$ and Z^0 possess a large mass of about 100 proton masses which corresponds to a rest energy of 100GeV. In order to explain theoretically the particle masses of $W^{\pm}$ and Z^0, physicists invented a mathematical trick called the Higgs mechanism, by using gauge invariance and adding appropriate mass terms to the Lagrangian. In terms of physics, this means that the Standard Model has to be supplemented by a number of hypothetical particles called Higgs particles. Computations show that the mass of the lightest Higgs particle should be between 114 and 193 proton masses. In 2008, the energy of the new CERN collider LHC[31] will be large enough in order to establish the

[31] The letters LHC stand for *Large Hadron Collider*. A detailed discussion of the LHC can be found in the article by B. Mansoulié, Physics at the large hadron collider. In: Duplantier and Rivasseau (Eds.) (2003), pp. 311–331.

existence of Higgs particles on a sound experimental basis. Note that the Standard Model of particle physics would break down if the Higgs particle did not exist.

Noncommutative geometry and the Standard Model in particle physics. It was discovered by Connes and Lott in 1990 that there is a new kind of geometry behind the Standard Model of particle physics called noncommutative geometry.[32] As an introduction to noncommutative geometry, we recommend the two monographs by Connes (1994) and Gracia-Bondia, Vàrilly, and Figueroa (2001).

Originally, the Higgs particle was inserted into the Standard Model by hand. Noncommutative geometry implies the appearance of the Higgs particle in a natural way. This will be studied in Volume V on the physics of the Standard Model. Noncommutative geometry is a new branch of mathematics which studies the generalization of geometric properties in terms of operator algebras.

Quantum gravity. Most physicists assume that below the Planck length $l_P = 10^{-35}$m and the Planck time $t_P = 10^{-44}$s, space and time lose their classical geometric properties, and there appear new physical effects combining gravitation and quantum physics in a strange manner. This has been coined as quantum gravity. Moreover, it is thinkable that space and time did not exist at the very beginning of the universe. They were created later on. In contrast to space and time, physical states always exist. They can be described mathematically by operator algebras.

> *In the setting of noncommutative geometry, physical states are primary and space-time is secondary.*

As an introduction to different approaches to quantum gravity, we recommend the collection of articles by Giulini, Kiefer, and Lämmerzahl (2003) (from theory to experimental search), the survey article by Ashtekhar and Lewandowski (2004) (loop quantum gravity), and the monograph by Kiefer (2004).

> *Most physicists expect that the creation of the final theory of quantum gravity will dramatically change our knowledge about space and time.*

The main tasks of quantum field theory. There exist two fundamental kinds of quantum states, namely, scattering states and bound states. In terms of classical celestial mechanics, scattering states correspond to comets and bound states correspond to closed orbits of planets (Fig. 2.15). Physicists use quantum field theory in order to compute

- the cross section of scattering processes,

[32] A. Connes and J. Lott, Particle models and noncommutative geometry, Nucl. Phys. B (Proc. Suppl.) **18** (1990), 29–47. See also the collection of survey articles edited by F. Scheck, W. Wend, and H. Upmeier, Noncommutative Geometry and the Standard Model of Elementary Particle Physics, Springer, Berlin, 2002.

(a) bound state (b) scattering state

Fig. 2.15. Classification of states

- the lifetime of decaying particles,
- the energy of composite particles (bound states), and
- the magnetic moment of particles.

A final theory should also allow us to compute the electric charge and further properties of elementary particles. Basic literature on quantum field theory can be found on page 907.

2.5 Magic Formulas

Let us summarize the most important formulas which govern the world of elementary particles.

(i) Planck's quantization of the energy of the harmonic oscillator from 1900:

$$\boxed{E = \hbar\omega\left(n + \frac{1}{2}\right), \qquad n = 0, 1, 2, \ldots}$$

Here, we use the following notation: E energy of the harmonic oscillator, ω angular frequency, h Planck's quantum of action. Furthermore, the symbol $\hbar := h/2\pi$ is called the reduced quantum of action. For$n = 0$, we get the ground state energy of the harmonic oscillator, $E_0 = \frac{1}{2}\hbar\omega$, which was obtained by Heisenberg in 1925 in the framework of his new quantum mechanics.

(ii) Planck's mean energy of the harmonic oscillator from 1900:

$$E = \hbar\omega\left(\frac{1}{2} + \frac{1}{\mathrm{e}^{\hbar\omega/kT} - 1}\right)$$

along with the following notation: E mean energy, ω angular frequency, T absolute temperature, k Boltzmann constant. For high temperatures, $kT \gg \hbar\omega$, we get approximately

$$E = kT.$$

According to Boltzmann, this is the mean energy of an oscillating degree of freedom in a many-particle system at the sufficiently high temperature T. For low temperatures, $kT \ll \hbar\omega$, we get

$$E = E_0 + \hbar\omega \mathrm{e}^{-\hbar\omega/kT}$$

where $E_0 = \frac{1}{2}\hbar\omega$ is the ground state energy of the harmonic oscillator. In a hot universe of absolute temperature T, an elementary particle with nonzero rest mass m_0 can only exist if

$$T \leq T_{\mathrm{crit}}$$

where $T_{\mathrm{crit}} := m_0 c^2/k$.

(iii) Einstein's energy-mass relation for relativistic particles from 1905:

$$\boxed{E^2 = m_0^2 c^4 + c^2 \mathbf{p}^2.}$$

Here, we use the following notation: E energy of the particle, m_0 rest mass, $\mathbf{p}$ momentum vector, c velocity of light in vacuum. For a photon, the rest mass vanishes, $m_0 = 0$.

(iv) Einstein's energy-frequency relation for photons from 1905:

$$\boxed{E = \hbar\omega.}$$

Equivalently, $E = h\nu$. Here, we use the following notation: E energy of the photon, ν frequency of the photon, $\omega = 2\pi\nu$ angular frequency.

(v) Bohr's quantized energy of the hydrogen atom from 1913:

$$E_n = -\frac{m_e e^4}{8\varepsilon_0^2 h^2} \cdot \frac{1}{n^2}, \qquad n = 1, 2, \ldots$$

Here, we use the following notation: E_n energy of the electron on the nth orbit, $-e$ electric charge of the electron, m_e mass of the electron, ε_0 electric field constant of the vacuum. The hydrogen atom consists of one proton and one electron.

Equivalently, the Bohr energy can be written as

$$E_n = -\frac{e^2}{8\pi\varepsilon_0 r_1} \cdot \frac{1}{n^2}, \qquad n = 1, 2, \ldots$$

along with

$$r_1 := \frac{\lambda_e}{2\pi\alpha}, \quad \lambda_e := \frac{h}{m_e c}, \qquad \alpha := \frac{e^2}{4\pi\varepsilon_0\hbar c} = \frac{1}{137.04}$$

where r_1, λ_e, α are called the Bohr radius of the hydrogen atom, the Compton wave length of the electron, and the fine structure constant in quantum electrodynamics, respectively.

(vi) The Schwarzschild radius of a star or a black hole from 1916:

$$r = \frac{GM}{c^2}$$

where M denotes the mass of the star or the black hole, and G is Newton's gravitational constant.

(vii) De Broglie's wave length of the matter wave corresponding to a relativistic particle from 1924 (duality between particles and waves in quantum physics):

$$\boxed{\lambda_C = \frac{h}{m_0 c}}$$

along with the following notation: m_0 rest mass of the particle, λ_C wave length of the matter wave (Compton wave length). Note that the symbol $\lambda\!\!\!^{-}_C := \lambda/2\pi$ is called the reduced Compton wave length.

(viii) Heisenberg's uncertainty relation for position and momentum from 1927:

$$\boxed{\Delta q \Delta p \geq \frac{\hbar}{2}.}$$

Here, Δq and Δp denote the mean error of the position coordinate q and the momentum coordinate p of the quantum particle, respectively.

(ix) The energy-time uncertainty relation for unstable particles:

$$\boxed{\Delta E \Delta t \geq \frac{\hbar}{2}}$$

along with the following notation: ΔE mean energy and Δt mean lifetime of the particle.

(x) The golden Breit–Wigner lifetime rule from 1930: Consider the scattering amplitude $f = f(E)$ as an analytic function of the energy E. If this function f has a pole at the complex energy value $E = E_0 + \mathrm{i}\Gamma$,

$$f(E) = \frac{C}{E - (E_0 + \mathrm{i}\Gamma)} + \text{regular terms},$$

then this corresponds to an unstable particle (resonance) of mean energy E_0 and mean lifetime,

$$\Delta t_{\text{mean}} := \frac{\hbar}{\Gamma}.$$

The interval $[E_0 - \frac{1}{2}\Delta E, E_0 + \frac{1}{2}\Delta E]$ with $\Delta E := \Gamma$ contains the fluctuating energy values E of the particle. For the modulus of f, we have

$$|f(E)|^2 = \frac{|C|^2}{(E - E_0)^2 + \Gamma^2} + \text{regular terms}.$$

(xi) Yukawa's meson mass formula from 1935:

$$m_0 = \frac{\hbar}{rc}.$$

This equation relates the rest mass, m_0, of the messenger particle (meson) to the range, r, of the corresponding force (radius of the proton). The typical length, r, is also called the reduced Compton wave length of the messenger particle.

(xii) The Hawking temperature of a black hole from 1975:

$$T = \frac{hc^3}{10kGM}.$$

Here, T and M denote the temperature and the mass of the black hole, respectively. Moreover, G denotes Newton's gravitational constant.

(xiii) Hawking's finite lifetime of a black hole caused by evaporation from 1975:

$$t = \frac{5M^3G^2}{hc^4}.$$

Here, t denotes the lifetime of the black hole.

(xiv) Friedman's mean energy density of the early universe from 1922:

$$\eta = \frac{3c^2}{32\pi t^2}.$$

Here, η and t denote the energy density and the age of the universe, respectively.

(xv) Hubble's law for the red shift in the spectrum of distant galaxies from 1928:

$$\frac{\Delta\lambda}{\lambda} = Ht.$$

Here, we use the following notation: H Hubble constant, λ wave length of light, $\Delta\lambda$ red shift of wave length λ, t time needed by light for travelling from the observed galaxy to earth. The red shift proves experimentally the expansion of the universe. The Hubble law is only an approximation. In fact, recent measurements of astronomers show that the expansion of the universe is accelerated (see Börner (2003)).

2.6 Quantum Numbers of Elementary Particles

It is a typical feature of elementary particles that in contrast to classical particles, important properties can be described by discrete numbers called quantum numbers. The spin of an elementary particle serves as a typical example.

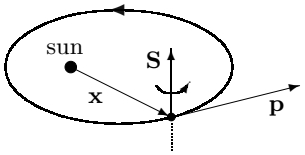


Fig. 2.16. Spin of a planet

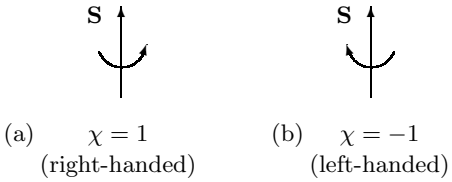


Fig. 2.17. Chirality of a planet

2.6.1 The Spin

Orbital angular momentum, spin, and chirality of a planet. Let us first consider the spin in classical physics. Each planet of our solar system travels around the sun in an elliptic orbit with position vector $\mathbf{x} = \mathbf{x}(t)$ (Fig. 2.16). The vector

$$\mathbf{p}(t) := m\dot{\mathbf{x}}(t)$$

is called the momentum vector of the planet at time t, where m and the time derivative $\dot{\mathbf{x}}(t)$ are the mass and the velocity vector of the planet, respectively. The vector

$$\boxed{\mathbf{a}(t) := \mathbf{x}(t) \times \mathbf{p}(t)}$$

is called the orbital angular momentum vector of the planet at time t. The length $||\mathbf{a}||$ of the vector $\mathbf{a}$ is called the orbital angular momentum of the planet at time t. In the special case where the planet rotates counterclockwise on a circle of radius r about the unit vector $\mathbf{n}$ with constant angular velocity ω, the position vector $\mathbf{x}(t)$ of length r is orthogonal to the velocity vector $\dot{\mathbf{x}}(t)$ of length $r\omega$. Hence

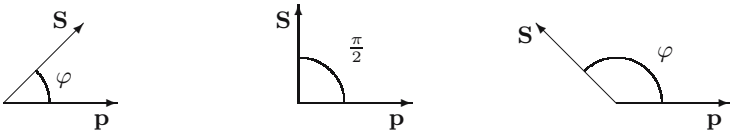
$$\mathbf{a} = (\omega r^2 m)\mathbf{n}.$$

In addition, the planet rotates counterclockwise about its own axis, given by the unit vector $\mathbf{n}_s$, with constant angular velocity ω_s. The corresponding intrinsic angular momentum vector

$$\boxed{\mathbf{S} := \chi\omega_s\Big(\int r(P)^2\varrho(P)dV\Big)\mathbf{n}_s}$$

is called the spin vector of the planet. Here, $\varrho(P)$ is the mass density of the planet at the point P, $r(P)$ is the distance between the point P and the rotation axis, and we take the volume integral over the points of the planet. The length of the spin vector $\mathbf{S}$ is called the spin of the planet. The definition of the spin vector $\mathbf{S}$ depends on the number $\chi = \pm 1$ called chirality.

- If $\chi = 1$, then the planet rotates counterclockwise about the vector $\mathbf{S}$. The planet is called right-handed (Fig. 2.17(a)).
- If $\chi = -1$, then the planet rotates clockwise about the vector $\mathbf{S}$. The planet is called left-handed (Fig. 2.17(b)).



(a) $\zeta > 0$ (b) $\zeta = 0$ (c) $\zeta < 0$

Fig. 2.18. Helicity ζ of a planet

For a planet, this is only a matter of convention. We agree to choose $\chi := 1$. Then the spin vector $\mathbf{S}$ points in the direction of the axis vector $\mathbf{n}_s$. The situation changes completely in elementary particle physics. For an elementary particle, we can measure the spin vector $\mathbf{S}$. The chirality $\chi = \pm 1$ is then an additional degree of freedom of an elementary particle which plays a crucial role in the Standard Model of particle physics. Observe the following peculiarity.

In the classical form of the Standard Model of particle physics, it is assumed that neutrinos are massless and always left-handed.

However, this is only an approximation of reality. On the basis of recent experiments, physicists assume that neutrinos possess a small mass and there exist also right-handed neutrinos in nature. This is based on the following experimental observation. In the burning sun, only electron neutrinos are produced. The measurements of astrophysicists show that there appears a shortage of sun neutrinos by a factor two. This neutrino defect problem in the sun can be solved in the following way: If we assume that neutrinos have a small mass, then this small neutrino mass makes it possible that neutrino oscillations occur which convert the electron neutrino into other types of neutrinos on its way from sun to earth; this changes the number of observed neutrinos.[33] As we will show later on, the Dirac equation allows us to define the chirality of a fermion in an elegant way.

Helicity of a planet. The real number

$$\zeta := \frac{\mathbf{S}(t)\mathbf{p}(t)}{||\mathbf{p}(t)||} = ||\mathbf{S}(t)|| \cos\varphi(t)$$

is called the helicity of a planet at time t, where $\varphi(t)$ is the angle between the spin vector $\mathbf{S}(t)$ and the direction of motion $\mathbf{p}(t)$ at time t (Fig. 2.18). If $\zeta > 0$ (resp. $\zeta < 0$), then the spin vector $\mathbf{S}$ has the same (resp. opposite) direction as the momentum vector $\mathbf{p}$.

Fermions and bosons. Each elementary particle possesses an intrinsic angular momentum called spin. The spin is characterized by a spin quantum number

[33] For more details, we refer to M. Fukugita and T. Yanagita, Physics of Neutrinos and Applications to Astrophysics, Springer, Berlin, 2003.

$$j = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, 2, \ldots$$

The particle is called a fermion iff $j = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$, and the particle is called a boson iff $j = 0, 1, 2, \ldots$ For example, the electron has the spin number $j = \frac{1}{2}$ (fermion), and the photon has the spin number $j = 1$ (boson). For the quantity $\mathbf{S}^2$, we measure the mean value

$$\hbar^2 j(j+1)$$

in each quantum state of the elementary particle. More precisely, we have the following situation. Consider first an electron which has spin quantum number $j = \frac{1}{2}$. Let the unit vector $\mathbf{n}$ be given. Choose a right-handed Cartesian coordinate system such that the z-axis points in direction of $\mathbf{n}$. Then there exist two quantum states denoted by

$$|\tfrac{1}{2}, \mathbf{n}\rangle \qquad \text{and} \qquad |-\tfrac{1}{2}, \mathbf{n}\rangle$$

such that we measure the spin vector

$$\mathbf{S} = j_z \hbar \mathbf{n} \qquad \text{in the electron state} \quad |j_z, \mathbf{n}\rangle \qquad \text{where} \quad j_z = \pm\tfrac{1}{2}.$$

The number j_z is called the spin projection number with respect to the unit vector $\mathbf{n}$.[34] Consider now the general case where the elementary particle has spin number j. Then there exist $2j+1$ quantum states of the particle denoted by $|j_z, j, \mathbf{n}\rangle$ where $j_z = j, j-1, j-2, \ldots, -j$. We measure the spin vector

$$\mathbf{S} = j_z \hbar \mathbf{n} \qquad \text{in the particle state} \quad |j_z, j, \mathbf{n}\rangle.$$

It is possible that not all of these quantum states can be realized in nature. In particular, for photons only the quantum states $|\pm 1, 1, \mathbf{n}\rangle$ are realized. This corresponds to the classical fact that the polarization of electromagnetic waves is always transversal. This means that both the electric and the magnetic field are transversal to the direction of propagation of the wave. Since the methods of statistical physics are based on the counting of quantum states, the following holds true.

> *If certain quantum states are forbidden by a general principle, then this has crucial consequences for the physical properties of quantum systems.*

For example, this concerns the Pauli exclusion principle below. The spin of an elementary particle is related to the 3-dimensional rotation group $SO(3)$. This is only part of the truth. Since the group $SO(3)$ is not simply connected, it has a simply connected universal covering group given by the group $SU(2)$ which is also called $Spin(3)$ (3-dimensional spin group).

[34] Instead of j_z, we also write j^3.

The spin of elementary particles results from the fact that the irreducible representations of the group $SU(2)$ can be characterized by a number j which coincides with the spin quantum number.

Since the irreducible representations of the group $SO(3)$ have integer spin quantum numbers, $j = 0, 1, \ldots$, one can say that

The existence of fermions is related to the nontrivial topological structure of the 3-dimensional rotation group.

For more details, we refer to Sect. 5.7.1.

The sum rule for spin quantum numbers. Consider two elementary particles which have the spin numbers j_1 and j_2. Then the composite particle has a spin number j which may attain precisely one of the following values:

$$j = j_1 + j_2, \;\; j_1 + j_2 - 1, \;\; j_1 + j_2 - 2, \;\; \ldots, \;\; |j_1 - j_2|.$$

For example, a system of two protons with $j_1 = j_2 = \frac{1}{2}$ may have the spin number $j = 1$ or $j = 0$. A system of three protons may have the spin numbers $j = 1 \pm \frac{1}{2}$ or $j = 0 + \frac{1}{2}$, hence $j = \frac{3}{2}, \frac{1}{2}$. As we will show later on, this sum rule can be based on the representations of the spin group $SU(2)$ on tensor products of linear spaces along with the decomposition into irreducible representations.

Three general principles for elementary particles. The following three principles play a fundamental role in quantum physics.

(i) *The principle of indistinguishability.* Quantum particles are not individuals. They cannot be distinguished individually.
(ii) *Pauli's exclusion principle.* In contrast to bosons, two fermions can never be in the same quantum state of a given quantum system.
(iii) *Pauli's spin-statistics principle.* Bosons (resp. fermions) obey Bose (resp. Fermi) statistics.[35]

We will show later on that principle (i) is responsible for the fact that the quantum states of bosons (resp. fermions) are symmetric (resp. antisymmetric) under permutations of elementary particles. Many strange properties of quantum systems are consequences of (ii) and (iii) (e.g., Bose–Einstein condensation – a group of bosons which are all in the same quantum state at extremely low temperature, and which behave like a single entity[36]).

The color charge of quarks is a consequence of the Pauli exclusion principle.

[35] Bose (resp. Fermi) statistics is also called Bose–Einstein (resp. Fermi–Dirac) statistics. These kinds of quantum statistical physics along with important applications can be found in Zeidler (1986), Vol. IV (e.g., Planck's radiation law, the critical Chandrasekhar mass of a white dwarf star (Nobel prize in physics in 1983), neutron stars, black holes, the death of a star, and the early universe).

[36] For the experimental realization, Cornell, Ketterle, and Wiemann were awarded the Nobel prize in physics in 2001.

In fact, we will show later on that quarks violate the Pauli exclusion principle if we do not add additional degrees of freedom (called color). In the framework of axiomatic quantum field theory created by Gårding and Wightman in the 1960s, the spin-statistics principle is a rigorous mathematical theorem. This can be found in Streater and Wightman (1968).

Historical remark. In 1922 Stern and Gerlach observed an unexpected splitting of a ray of silver atoms in a magnetic field (Stern–Gerlach effect). This splitting was explained by Goudsmit and Uhlenbeck in 1925. They postulated that the electron has a spin, and hence a magnetic moment. More precisely, if the electron is in the quantum state $|\pm\frac{1}{2},\mathbf{n}\rangle$, then it has the magnetic moment vector

$$\mathbf{m} = -\frac{e}{m_e}\mathbf{S}, \qquad \mathbf{S} = \pm\frac{\hbar}{2}\mathbf{n}.$$

The vector $\mathbf{m}$ possesses the following physical meaning.

(i) An external magnetic field $\mathbf{B}$ acts onto the electron by the force

$$\mathbf{F}(\mathbf{x}) = \mathbf{B}'(\mathbf{x})\mathbf{m} \tag{2.54}$$

and the torque $\mathbf{T}(\mathbf{x}) = \mathbf{m} \times \mathbf{B}(\mathbf{x})$.[37]

(ii) The constant magnetic field $\mathbf{B}$ adds the energy $E = -\mathbf{mB}$ to the electron.

(iii) The electron induces a magnetic field $\mathbf{B}_0$ given by

$$\mathbf{B}_0(\mathbf{x}) = \frac{\mu_0}{4\pi r^3}\Big(\frac{3(\mathbf{xm})\mathbf{x}}{r^2} - \mathbf{m}\Big)$$

where we set $r := ||\mathbf{x}||$. Here, μ_0 is the magnetic field constant of a vacuum.

The strength of the magnetic field $\mathbf{B}$ is measured in Tesla, $1\mathrm{T} = \mathrm{Vs/m^2}$. The magnetic field of the earth has the strength of $0.5 \cdot 10^{-4}\mathrm{T}$. The force (2.54) is responsible for the Stern–Gerlach effect. Similarly, we expect that the proton has the magnetic moment

$$\mathbf{m} = \frac{e}{m_p}\mathbf{S}, \qquad \mathbf{S} = \pm\frac{\hbar}{2}\mathbf{n}$$

where e and m_p denote the electric charge and mass of the proton, respectively. The point is that these values are only approximations of the measured values. Therefore, physicists say that both the electron and the proton have an anomalous magnetic. For the magnetic moment $\mu := |\mathbf{m}|$ of the electron and the proton, experiments yield the values

$$\mu_e = \frac{e\hbar}{2m_e} \cdot 1.001\,159\,652\,193 = 9.284\,770\,1(31) \cdot 10^{-24}\mathrm{J/T}$$

[37] As usual, the directional derivative $\mathbf{B}'(\mathbf{x})\mathbf{m}$ is defined to be the derivative of $\mathbf{B}(\mathbf{x}+\varepsilon\mathbf{m})$ with respect to the real variable ε at the point $\varepsilon = 0$.

and

$$\mu_p = \frac{e\hbar}{2m_p} \cdot 2.792\,847\,386 = 1.410\,607\,61(47) \cdot 10^{-26}\mathrm{J/T},$$

respectively. Thus, the magnetic moment of the electron is much stronger than that of the proton. The relative error of the magnetic moment of the electron and the proton is $3.4 \cdot 10^{-7}$. The reason for these anomalous magnetic moments are typical effects of quantum field theory which are based on vacuum fluctuations (interactions with virtual electrons, virtual positrons and virtual photons). The sophisticated renormalization methods of quantum electrodynamics based on large computer programs yield a theoretical value of the magnetic moment of the electron which is in fantastic coincidence with the experiments. In 1949, Schwinger used his new theory of quantum electrodynamics in order to compute the first-order correction of the magnetic moment of the electron. He obtained the formula

$$\mu_e = \frac{e\hbar}{2m_e}\Big(1 + \frac{\alpha}{\pi}\Big)$$

where α is the fine structure constant. Precision experiments yield the value

$$\alpha = \frac{e^2}{4\pi\varepsilon_0\hbar c} = \frac{1}{137.035\,989\,5(61)}.$$

Based on huge computer programs and the methods of perturbation theory in quantum electrodynamics (Feynman diagrams and renormalization methods), physicists computed higher order corrections of the magnetic moment of the electron which are in fantastic coincidence with experimental data.

The perturbation methods of quantum electrodynamics were invented in the late 1940s by Feynman, Schwinger, and Tomonaga.

These methods along with Dyson's renormalization approach allow the computation of electromagnetic effects in terms of powers of the fine structure constant α. Quantum electrodynamics investigates the interactions between electrons, positrons, and photons. The extremely accurate coincidence between theory and experiment can still be improved by including physical effects that are caused by additional particles in the setting of the Standard Model of particle physics. For example, the fourth-order approximation of the magnetic moment of the electron also includes effects which come from the muon. For the ratio between the masses and the magnetic moments of the muon and the electron, we get approximately

$$\frac{m_\mu}{m_e} = 2, \qquad \frac{\mu_\mu}{\mu_e} = 5 \cdot 10^{-3},$$

respectively. Similarly, for the ratio between the masses and magnetic moments of the neutron and proton, we obtain approximately

$$\frac{m_n}{m_p} = 1, \qquad \frac{\mu_n}{\mu_p} = \frac{2}{3}.$$

The exact values for the muon and the neutron are known with the same high precision as for the electron and the proton. The magnetic moment of the proton was measured by Stern in 1933 (Nobel prize in physics in 1943). Generally, both the experimental and the theoretical determination of magnetic moments of elementary particles is a highly nontrivial task. The magnetism of materials in nature is caused by both the intrinsic magnetic moments of elementary particles (based on spin) and the motion of electrically charged elementary particles.

In 1925 Heisenberg created quantum mechanics. Born and Jordan noticed quickly that Heisenberg's approach based on Fourier series could be reformulated in the mathematical language of infinite-dimensional matrices. Therefore, Heisenberg's quantum mechanics is also called matrix mechanics. In 1926 Schrödinger formulated an alternative approach to quantum mechanics based on the non-relativistic Schrödinger equation for a wave function. In the same year, Pauli generalized the Schrödinger equation to the non-relativistic Pauli–Schrödinger equation by inserting the electron spin. Pauli noticed that this approach, along with the exclusion principle, explains the structure of atoms encoded in the periodic table of chemical elements.

The final explanation for the existence of the electron spin was obtained by Dirac in 1928. Dirac formulated a relativistic equation for the electron by combining quantum mechanics with Einstein's theory of special relativity (Nobel prize in physics together with Schrödinger in 1933). The Dirac equation implies the existence of the electron spin, by using a simple group-theoretical argument.

The electron spin represents a typical relativistic quantum effect.

The same is true for the spin of all elementary particles. As we will see later on, the Dirac equation is of fundamental importance for describing the 12 fundamental particles of the Standard Model of particle physics. From the mathematical point of view, the Dirac equation is based on a Clifford algebra. Fundamentally, the spin of elementary particles is related to the transformation of quantum fields under the rotation of inertial systems. The general approach was developed by Wigner in 1939 by classifying the irreducible unitary representations of the Poincaré group in infinite-dimensional Hilbert spaces (Nobel prize in physics in 1963).

The periodic table of chemical elements, Pauli's exclusion principle, and the shell structure of atoms. In 1869, great progress was made in chemistry when Mendeleev (1834-1907) and Meyer (1830-1895) were able to systematically order the chemical elements according to phenomenological criteria. Table 2.5 shows the beginning of the periodic table. In the horizontal direction the atomic number Z increases, while in the vertical direction elements behave similarly. In the late 1920s, Pauli discovered how the periodic

Table 2.5. The periodic table of chemical elements

1 H $1s$							2 He $1s^2 = \mathrm{K}$
3 Li K $2s$	4 Be K $2s^2$	5 B K $2s^2, 2p$	6 C K $2s^2, 2p^2$	7 N K $2s^2, 2p^3$	8 O K $2s^2, 2p^4$	9 F K $2s^2, 2p^5$	10 Ne K $2s^2, 2p^6 = \mathrm{L}$
11 Na K L $3s$	12 Mg K L $3s^2$	13 Al K L $3s^2, 3p$	14 Si K L $3s^2, 3p^2$	15 P K L $3s^2, 3p^3$	16 S K L $3s^2, 3p^4$	17 Cl K L $3s^2, 3p^5$	18 Ar K L $3s^2, 3p^6 = \mathrm{M}$

table of chemical elements can be understood in terms of quantum physics. Let us discuss the main ideas.

(i) The atomic number Z is equal to the number of electrons and equal to the number of protons in the nucleus.

(ii) The number of neutrons in the nucleus may vary for fixed Z. Thereby isotopes occur.

(iii) An electron state

$$|n, l, m, j_z\rangle$$

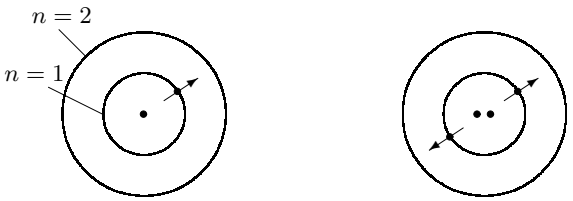
is characterized by the quantum numbers

- $n = 1, 2, 3 \ldots$ (orbit),
- $l = 0, \ldots, n-1$ (orbital angular momentum),
- $m = l, l-1, \ldots, -l$ (orbital angular momentum projection),
- $j_z = \pm\frac{1}{2}$ (spin projection).

In particular, the ground state is described by the quantum numbers $n = 1, l = m = 0,\ j_z = \pm\frac{1}{2}$. By Pauli's exclusion principle, two electrons cannot coincide in all four quantum numbers n, l, m, j_z.

(iv) The energy of the electron depends on the quantum number n with $E_1 < E_2 < E_3 < \ldots$ For energetic reasons, the orbits with $n = 1, 2, 3 \ldots$ are filled successively.

(v) The inner orbits are maximally filled. Such orbits are called closed shells and denoted by K, L, M, ... for $n = 1, 2, 3 \ldots$, respectively. The electrons of the incomplete outer orbit are responsible for the chemical behavior of the element. The similarity of chemical elements is a consequence of the same number of outer electrons. Table 2.5 shows the number of electrons in different orbits. In horizontal direction new electrons are added continuously. Here, the symbols s and p stand for electron states with the quantum number $l = 0$ and $l = 1$, respectively. Furthermore, the symbol ns^k means that the number of s-electrons is equal to k in the nth orbit. The maximal number of s-electrons in the nth orbit ($n = 1, 2, 3 \ldots$) is equal to 2 because of



(a) hydrogen ${}^{1}_{1}\mathrm{H}$ (b) helium ${}^{4}_{2}\mathrm{He}$

Fig. 2.19. Structure of atoms

$$n = 1, \quad l = 0, \quad m = 0, \quad j_z = \pm\tfrac{1}{2}.$$

Similarly, the maximal number of p-electrons in the nth orbit is equal to six since

$$n = 2,\ l = 1,\ m = 1, 0, -1,\ j_z = \pm\tfrac{1}{2}.$$

In the vertical lines of Table 2.5 we have precisely the same number of outer s- and p-electrons. This results in a similar behavior of the corresponding elements. The inert gases 2 He (helium), 10 Ne (neon), and 18 Ar (argon) have only closed shells K, L, M, respectively. This is the reason for their chemical inactivity.

Spin-orbit coupling. The states from (iii) above neglect the interactions between orbit and spin (called spin-orbit coupling) and the nonlinear interaction between different electrons. If we take the spin-orbit coupling into account, then we cannot anymore distinguish between orbital angular momentum and intrinsic spin of the electron. Only the total angular momentum makes sense. By the sum rule for angular momenta, the electron states $|n, j, j_z\rangle$ are now described by the following quantum numbers:

- $n = 1, 2, \ldots$ (orbit),
- $j = l + \frac{1}{2}, l - \frac{1}{2}$ (total angular momentum), $l = 1, \ldots, n-1$,
- $j = \frac{1}{2}$ if $l = 0$, and
- $j_z = j, j-1, \ldots, -j$ (projection of total angular momentum).

The ground state corresponds to $n = 1, j = \frac{1}{2}, j_z = \pm\frac{1}{2}$. Starting with calcium $Z = 19$, the irregularities of the filling of the orbits begin. This follows from the fact that for a large number Z of electrons the interaction between the electrons becomes stronger and the energetic situation becomes more complicated.[38]

Symmetry breaking in atomic spectra. If an electron jumps from an energy level E_n to a lower energy level E_m, then a photon of energy

$$h\nu = E_n - E_m$$

[38] As an introduction to the application of group-theoretic methods to atomic spectra, we recommend van der Waerden (1930) and Mizushima (1970).

and frequency ν is emitted. In first approximation, the electron energy E_n only depends on the orbit quantum number n. This means that different electron states may have the same energy. This degeneracy of energy levels is a consequence of the rotational symmetry of the electrostatic Coulomb force. This symmetry can be broken by external magnetic (resp. electric) fields or by internal forces between the electrons. As a rule, symmetry breaking reduces the degeneracy of energy levels. In experiments, this leads to a splitting of spectral lines. For example, in a constant magnetic field, electron states with different magnetic quantum numbers m possess different energy levels. This splitting of the spectrum is called the Zeeman effect (in 1902 Zeeman was awarded the Nobel prize in physics).

Quantum chemistry. Pauli's approach to the periodic table of chemical elements marked the emerging of quantum chemistry. Nowadays huge computer programs (e.g., the "Gaussian" soft ware package) are used in order to compute the properties of atoms and molecules. In 1998 Pople and Kohn were awarded the Nobel prize in chemistry for the development of computational methods in quantum chemistry and the development of the density-functional theory, respectively. The density-functional method allows us the approximate computation of ground state energies of large molecules.

- The first task of quantum chemistry is to compute the energy levels of an atom or molecule. This is done by the Ritz method for eigenvalue problems.[39] The main point is to find appropriate linear combinations of given functions with unknown coefficients for the computation of approximate eigenfunctions.
- Furthermore, chemists want to get information on the geometry of the molecules (distances and angles between atoms, symmetries).
- Finally, chemists want to compute the dynamics of a chemical reaction, that is, the change of the structure of the molecules during a reaction process (reaction path). In this connection, saddle points on the energy surface of molecules are crucial. The points of the energy surface depend on the geometric molecule parameters.

Heisenberg's isospin. In 1932, motivated by the theory of the electron spin and the fact that protons and neutrons have approximately the same rest energy, Heisenberg regarded the proton p and the neutron n as two different states $|\frac{1}{2}\rangle$ and $|-\frac{1}{2}\rangle$ of a quantum object (nucleon), and he called $\frac{1}{2}$ and $\pm\frac{1}{2}$ the isospin number and the isospin projection of the nucleon, respectively. The isospin and the so-called weak isospin play an important role in the Standard Model.

The particle zoo. Beginning in the 1950's, physicists used particle accelerators of increasing power in order to discover a large zoo of unstable "elementary particles" of extremely short lifetimes. These particles were called resonances. The quite natural goal of physicists was to reduce this zoo to a

[39] The Ritz method is thoroughly studied in Zeidler (1986), Vol. IIA.

small number of fundamental particles. They tried to understand the zoo of elementary particles in the same way as the chemical elements can be understood by means of the periodic table. To this end, physicists used quantum numbers based on symmetries. Nowadays this goal is realized by the Standard Model of elementary particle physics.

The mathematical formulation of the Standard Model in particle physics. In the next section, we will discuss the basic ideas of the Standard Model in terms of quantum numbers. The detailed formulation of the Standard Model in particle physics based on the full Lagrangian density will be postponed to Volume III after a detailed study of quantum electrodynamics in Volume II. Quantum electrodynamics represents the Standard Model in a nutshell. The Lagrangian density of the Standard Model contains fields which describe

- the six quarks,
- the six leptons,
- the 12 messenger particles, and
- the Higgs particle.

The point is that the Lagrangian density of the Standard Model possesses crucial symmetry properties, namely, the gauge symmetry

$$U(1) \times SU(2) \times SU(3)$$

and the relativistic symmetry under Poincaré transformations. Furthermore, the Clifford algebra of Dirac–Pauli matrices plays a crucial role. This Clifford algebra is closely related to the Minkowski metric of the 4-dimensional space-time manifold. The interactions between elementary particles mediated by the messenger particles are described by using the modern language of differential geometry (curvature of fiber bundles) The prototype of this geometric concept in physics will be discussed in Sect. 2.9.1 on page 183.

2.6.2 Conservation of Quantum Numbers

> The Mendeleev of elementary particle physics is Murray Gell-Mann.
> Folklore

The most important quantum numbers of elementary particles read as follows.

- Q/e electric charge number (global gauge group $U(1)$),[40]
- j spin number (j^3 third component of spin) (symmetry group $SU(2)$ as subgroup of the universal covering group $SL(2,\mathbb{C})$ of the Poincaré group in special relativity),[41]

[40] The group $U(1)$ consists of all complex numbers z with $|z| = 1$.

[41] The group $SU(n)$ consists of all complex $(n \times n)$-matrices A with $A^{-1} = A^\dagger$ and $\det A = 1$.
The group $SL(2,\mathbb{C})$ consists of all complex (2×2)-matrices A with $\det A = 1$.

Table 2.6. Quantum numbers of leptons

particle	e^-	ν_e	μ^-	ν_μ	τ^-	ν_τ
Q/e	-1	0	-1	0	-1	0
$\mathcal{P}$	1	1	1	1	1	1
$\mathcal{L}$	1	1	1	1	1	1
$\mathcal{L}_e$	1	1	0	0	0	0
$\mathcal{L}_\mu$	0	0	1	1	0	0
$\mathcal{L}_\tau$	0	0	0	0	1	1

- rest mass m_0 (infinite-dimensional representations of the universal covering group of the Poincaré group due to Wigner),
- energy-momentum 4-vector p (group of space-time translations),
- $\mathcal{P}, \mathcal{P}_{\text{total}}$ internal parity, total parity, respectively (group of space reflections),
- χ chirality (universal covering group $SL(2,\mathbb{C})$ of the Poincaré group),
- $\mathcal{L}$ lepton number,
- $\mathcal{L}_e, \mathcal{L}_\mu, \mathcal{L}_\tau$ electron lepton number, muon lepton number, tau lepton number, respectively,
- $\mathcal{B}$ baryon number,
- $\mathcal{I}$ strong isospin ($\mathcal{I}^3$ third component of the strong isospin),
- S, C, B, T strangeness, charm, bottomness, topness, respectively,
- $\mathcal{Y}$ strong hypercharge,[42]
- r, g, b color charge of quarks: red, green, blue, respectively (local gauge group $SU(3)$ of strong interaction),
- $\mathcal{I}_w$ weak isospin ($\mathcal{I}_w^3$ third component of weak isospin) (local gauge group $SU(2)$ of electroweak interaction),
- $\mathcal{Y}_w$ weak hypercharge (local gauge group $U(1)$ of electroweak interaction).

Important examples can be found in Tables 2.6–2.9.

Antiparticles. If we pass from an elementary particle to its antiparticle, then rest mass m_0, spin j, strong isospin $\mathcal{I}$, and weak isospin $\mathcal{I}_w$ remain unchanged, whereas the quantum numbers

$$Q/e, \mathcal{B}, \mathcal{L};\ \ \mathcal{L}_e, \mathcal{L}_\mu, \mathcal{L}_\tau;\ \ \mathcal{I}^3, \mathcal{I}_w^3, \mathcal{Y}, \mathcal{Y}_w, S, B, C, T;\ \ \chi, \mathcal{P}$$

change their sign. The color charges r, g, b have to be replaced by anticolor charges $\overline{r}, \overline{g}, \overline{b}$, respectively. The values of the color charges change their sign.

> *Particles and antiparticles are described by mathematical objects which are dual to each other (elements of a Hilbert space and of its dual space).*

[42] $\mathcal{I}, \mathcal{I}^3, S, B, C, T, \mathcal{Y}$ are based on the global symmetry group $SU(6)$ of the six quarks called the flavor symmetry group of strong interaction.

Table 2.7. Quantum numbers of quarks

particle	d	u	s	c	b	t
Q/e	$-\frac{1}{3}$	$\frac{2}{3}$	$-\frac{1}{3}$	$\frac{2}{3}$	$-\frac{1}{3}$	$\frac{2}{3}$
$\mathcal{P}$	1	1	1	1	1	1
$\mathcal{I}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0
$\mathcal{I}^3$	$\frac{1}{2}$	$-\frac{1}{2}$	0	0	0	0
$\mathcal{Y}$	$\frac{1}{3}$	$\frac{1}{3}$	$-\frac{2}{3}$	$\frac{4}{3}$	$-\frac{2}{3}$	$\frac{4}{3}$
$\mathcal{B}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
S	0	0	1	0	0	0
C	0	0	0	1	0	0
B	0	0	0	0	-1	0
T	0	0	0	0	0	1

Table 2.8. Quark content

baryon	p	n	Σ^-	Σ^0	Σ^+	Ξ^-	Ξ^0	Λ
	uud	ddu	dds	uds	uus	dss	uss	uds
meson	K^0	K^+	π^-	π^0	π^+	K^-	$\overline{K}^0$	η
	$d\bar{s}$	$u\bar{s}$	$d\bar{u}$	$u\bar{u},\ d\bar{d}$	$u\bar{d}$	$s\bar{u}$	$\bar{d}s$	$u\bar{u},\ d\bar{d},\ s\bar{s}$

This will be discussed in Volume II on quantum electrodynamics in the framework of the Fock space.

Electric charge. The electric charge of an elementary particle is conserved for all processes. For composed particles, the electric charge is an additive quantity. Recall that baryons consist of three quarks, whereas mesons consist of quark-antiquark pairs. The quantum number Q/e is an integer for leptons, baryons, and mesons.

Lepton number. The lepton numbers are additive quantities. Using Table 2.6 on page 155, we set

$$\mathcal{L} := \mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\tau.$$

In particular, for leptons, $\mathcal{L} = 1$. For quarks, baryons, and mesons, all of the lepton numbers are defined to be zero.

The lepton numbers $\mathcal{L}$ and $\mathcal{L}_e, \mathcal{L}_\mu, \mathcal{L}_\tau$ are conserved for all processes.

For example, the beta decay of the neutron

$$n \to p + e^- + \overline{\nu}_e \tag{2.55}$$

Table 2.9. Electroweak interaction

right-handed lepton singlets (gauge group $U(1)$)	$\mathcal{I}_w$	$\mathcal{I}_w^3$	$\mathcal{Y}_w$	Q/e	χ
$e_R^-\quad \mu_R^-\quad \tau_R^-$	0	0	-1	-1	1
left-handed lepton doublets (gauge group $SU(2)$)					
$\begin{pmatrix}\nu_{eL}\\ e_L^-\end{pmatrix}\quad \begin{pmatrix}\nu_{\mu L}\\ \mu_L^-\end{pmatrix}\quad \begin{pmatrix}\nu_{\tau L}\\ \tau_L^-\end{pmatrix}$	$\frac{1}{2}$ $\frac{1}{2}$	$\frac{1}{2}$ $-\frac{1}{2}$	$-\frac{1}{2}$ $-\frac{1}{2}$	0 -1	-1 -1
right-handed quark singlets (gauge group $U(1)$)					
$u_R\quad c_R\quad t_R$	0	0	$\frac{2}{3}$	$\frac{2}{3}$	1
$d'_R\quad s'_R\quad b'_R$	0	0	$-\frac{1}{3}$	$-\frac{1}{3}$	1
left-handed quark doublets (gauge group $SU(2)$)					
$\begin{pmatrix}u_L\\ d'_L\end{pmatrix}\quad \begin{pmatrix}c_L\\ s'_L\end{pmatrix}\quad \begin{pmatrix}t_L\\ b'_L\end{pmatrix}$	$\frac{1}{2}$ $\frac{1}{2}$	$\frac{1}{2}$ $-\frac{1}{2}$	$\frac{1}{6}$ $\frac{1}{6}$	$\frac{2}{3}$ $-\frac{1}{3}$	-1 -1
Higgs boson H	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	0	

conserves the lepton numbers $\mathcal{L}$ and $\mathcal{L}_e$. In this case, the right-hand side has the lepton number $\mathcal{L} = \mathcal{L}_e = 1-1 = 0$, and the left-hand side has $\mathcal{L} = \mathcal{L}_e = 0$ (see Table 2.6). Since the lepton number $\mathcal{L}$ of the electron neutrino, ν_e, is equal to one, the process

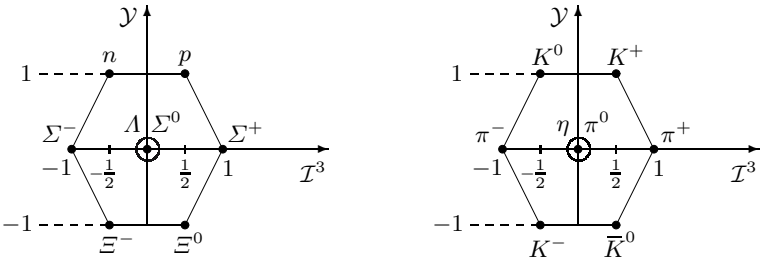
$$n \to p + e^- + \nu_e \tag{2.56}$$

is impossible because it violates the conservation of $\mathcal{L}$. In fact, in contrast to (2.55), the right-hand (resp. left-hand) side of (2.56) has $\mathcal{L} = 0+1+1$ (resp. $\mathcal{L} = 0$). This argument shows that the production of an electron must always be accompanied by the production of an anti-electron neutrino. Similarly, the process

$$n \to p + \mu^- + \overline{\nu}_e$$

is impossible because it violates the conservation of the electron lepton number $\mathcal{L}_e$. In fact, the right-hand (resp. left-hand) side has $\mathcal{L}_e = 0+0-1$ (resp. $\mathcal{L}_e = 0$).

Baryon number. The baryon number is an additive quantity. For each quark, the baryon number is equal to $\frac{1}{3}$. Since each baryon consists of three quarks, the baryon number of baryons is equal to 1. For mesons and leptons, the baryon number is equal to zero.



(a) baryon octet (b) meson octet

Fig. 2.20. Composed particles

The baryon number $\mathcal{B}$ is conserved for all processes.

For example, one baryon can never decay into two leptons. Note that physicists assume that the conservation of lepton number and baryon number was violated shortly after the Big Bang. This way, the creation of leptons and baryons was possible in the early universe. A similar asymmetry of the laws of physics in the early universe is responsible for the dominance of particles over antiparticles in the present universe.

Strong isospin, strangeness, and strong hypercharge. The relation between electric charge Q and strong isospin $\mathcal{I}^3$ is given by

$$\boxed{\frac{Q}{e} = \mathcal{I}^3 + \frac{\mathcal{Y}}{2}}$$

where the so-called strong hypercharge is defined by $\mathcal{Y} := \mathcal{B} + S + C + B + T$. The eight baryons (resp. eight mesons) pictured in Fig. 2.20 possess approximately the same mass.[43] They are distinguished by the "quantum number hypercharge $\mathcal{Y}$" and the "quantum number third component of isospin $\mathcal{I}^3$." The quark content can be found in Table 2.8 on page 156. For example, a proton consists of two up quarks u and one down quark d. For the proton, we get $S = C = B = T = 0$, and hence $\mathcal{Y} = 1$. Moreover, the proton has the isospin $\mathcal{I} = \frac{1}{2}$ and the third component of isospin $\mathcal{I}^3 = \frac{1}{2}$. Thus, the electric charge number of the proton is given by

$$\frac{Q}{e} = \frac{1}{2} + \frac{1}{2},$$

[43] In contrast to strong hypercharge $\mathcal{Y}$, weak hypercharge $\mathcal{Y}_w$ is defined by the slightly modified relation

$$\frac{Q}{e} = \mathcal{I}_w^3 + \mathcal{Y}_w$$

where Q and $\mathcal{I}_w$ denote electric charge and weak isospin, respectively. This convention is used in order to simplify the notation in the Standard Model of particle physics (see Table 2.9 on page 157).

implying $Q = e$, as desired. For historical reasons, a baryon or meson is called strange iff it contains a strange quark s or a strange antiquark $\overline{s}$. Such particles were discovered in the 1950s. All of the strange particles are produced by strong interaction and often decay via weak interaction.

The quantum numbers $\mathcal{I}, \mathcal{I}_3, \mathcal{Y}, S, C, B, T$ are conserved for processes via strong interaction.

Chirality, weak isospin, and weak hypercharge. Table 2.9 on page 157 shows the classical Standard Model of electroweak interaction. This is an approximation based on the assumption that neutrinos are massless and there only exist left-handed neutrinos. Explicitly, the three left-handed pairs are

$$\boxed{(\nu_e)_L, e_L^-; \quad (\nu_\mu)_L, \mu_L^-; \quad (\nu_\tau)_L, \tau_L^-.}$$

Since these pairs possess specific transformation laws with respect to the gauge group $SU(2)$, they are called doublets. In addition, there exist right-handed leptons

$$e_R^-, \mu_R^-, \tau_R^-$$

called singlets. Similarly, one has to use the left-handed quark doublets

$$u_L, d'_L; \quad c_L, s'_L; \quad t_L, b'_L$$

and the right-handed quark singlets

$$u_R, c_R, t_R; \quad d'_R, s'_R, b'_R.$$

Here, d'_L replaces the quark d by a linear combination of quarks. The coefficients have to be determined by experiment. They are called CKM (Cabibbo–Kobayashi–Maskawa) coefficients. Moreover, we assign

- the weak isospin $\mathcal{I}_w = 0$ (singlets), $\mathcal{I}_w = \frac{1}{2}$ (doublets), and
- the weak hypercharge $\mathcal{Y}_w = 2Q/e$ (singlets), $\mathcal{Y}_w = \mathcal{B} - \mathcal{L}$ (doublets).

The explicit values can be found in Table 2.9 on page 157.

Parity. To each lepton and quark, we assign the internal parity number $\mathcal{P} = 1$. For antileptons and antiquarks, $\mathcal{P} = -1$. In addition, for the photon, $\mathcal{P} = -1$. For composed particles, the internal parity is a multiplicative quantity. Since mesons consist of quark-antiquark pairs, the internal parity of mesons is equal to

$$\mathcal{P} = 1(-1) = -1.$$

Baryons consist of three quarks. Thus, for baryons, $\mathcal{P} = 1^3 = 1$. In addition, let us introduce the orbital parity by letting

$$\mathcal{P}_{\text{orbital}} := (-1)^l.$$

Here, l is the quantum number of angular momentum of the particle. Finally, the total parity is defined to be

Table 2.10. Color charge of quarks

quarks	red r	green g	blue b
	$Q_r = \frac{1}{3}$	$Q_g = \frac{1}{3}$	$Q_b = -\frac{2}{3}$
antiquarks	antired $\bar{r}$	antigreen $\bar{g}$	antiblue $\bar{b}$
	$Q_{\bar{r}} = -\frac{1}{3}$	$Q_{\bar{g}} = -\frac{1}{3}$	$Q_{\bar{b}} = \frac{2}{3}$

$$\mathcal{P}_{\text{total}} := \mathcal{P}(-1)^l.$$

Color of quarks. The electromagnetic interaction depends on the electric charge of particles. Similarly, the strong interaction depends on a specific charge of quarks called color charge. Each quark possesses three color charges called red r, green g, and blue b. Let us assign the color charge quantum numbers

$$Q_r = \frac{1}{3}, \quad Q_g = \frac{1}{3}, \quad Q_b = -\frac{2}{3}.$$

Antiparticles carry the color charges antired $\bar{r}$ (or cyan), antigreen $\bar{g}$ (or magenta), antiblue $\bar{b}$ (or yellow). Here, the color charge numbers change sign. Explicitly,

$$Q_{\bar{r}} = -\frac{1}{3}, \quad Q_{\bar{g}} = -\frac{1}{3}, \quad Q_{\bar{b}} = \frac{2}{3}.$$

The color charge is an additive quantity. We postulate that

The total color charge of baryons and mesons is equal to zero.

Since the superposition of red, green, and blue yields white, physicists also say that baryons and mesons are white.

States of elementary particles and group theory. Mathematically, states of elementary particles are described by unit vectors in complex Hilbert spaces. The vectors depend on space and time. Let us consider the following three examples. To simplify notation, we do not take the dependence on space and time coordinates into account.

(i) Two electrons with spin. We are given a right-handed (x, y, z)-Cartesian coordinate system. The symbol

$$|e^-, j, j^3\rangle$$

denotes the state of an electron e^- with spin numbers $j = \frac{1}{2}$ and $j^3 = \pm\frac{1}{2}$. This means that, in this state, we measure the spin $j^3\hbar$ in direction of the z-axis. Assume that the two unit vectors

$$|e^-, \tfrac{1}{2}, \tfrac{1}{2}\rangle, \qquad |e^-, \tfrac{1}{2}, -\tfrac{1}{2}\rangle$$

form the basis of a 2-dimensional complex Hilbert space X with the inner product $\langle\psi|\varphi\rangle$. The linear operator $S^3 : X \to X$ given by the following eigenvectors and eigenvalues

$$S^3|e^-, \tfrac{1}{2}, \pm\tfrac{1}{2}\rangle = \pm\tfrac{1}{2}\hbar\, |e^-, \tfrac{1}{2}, \pm\tfrac{1}{2}\rangle$$

is called the spin operator in direction of the z-axis. For complex numbers α, β with $|\alpha|^2 + |\beta|^2 = 1$, the vector

$$\psi = \alpha\, |e^-, \tfrac{1}{2}, \tfrac{1}{2}\rangle + \beta\, |e^-, \tfrac{1}{2}, -\tfrac{1}{2}\rangle$$

corresponds to a state where the electron is in the state $|e^-, \frac{1}{2}, \frac{1}{2}\rangle$ and $|e^-, \frac{1}{2}, -\frac{1}{2}\rangle$ with probability $|\alpha|^2$ and $|\beta|^2$, respectively. In the state ψ, we measure the value

$$\langle\psi|S^3\psi\rangle = \tfrac{1}{2}\hbar|\alpha|^2 - \tfrac{1}{2}\hbar|\beta|^2$$

of the electron spin in z-direction. It is assumed that two vectors ψ_1 and ψ_2 represent the same physical state of the electron iff

$$\psi_1 = \sigma\psi_2$$

where σ is a complex number with $|\sigma| = 1$. In this case, the measured value of the electron spin does not change. The *antisymmetric* tensor product[44]

$$\varphi := \frac{|e^-, \tfrac{1}{2}, j_1^3\rangle \otimes |e^-, \tfrac{1}{2}, j_2^3\rangle \;\; - \;\; |e^-, \tfrac{1}{2}, j_2^3\rangle \otimes |e^-, \tfrac{1}{2}, j_1^3\rangle}{\sqrt{2}} \tag{2.57}$$

represents the state of two electrons with spin numbers j_1^3 and j_2^3. Observe the following two crucial facts.

- A permutation of the two electrons changes the sign of φ, but not the physical state. Thus, the two electrons are not distinguishable. They lose their individuality (principle of indistinguishable elementary particles).
- If $j_1^3 = j_2^3$, then $\varphi = 0$. Consequently, the two electrons cannot be in the same spin state (Pauli's exclusion principle for fermions).

(ii) Two photons with spin. Electrons are *fermions*. In contrast to this, photons γ are *bosons* and hence Pauli's exclusion principle is not valid for photons. This motivates us to use the following *symmetric* tensor product

$$\varphi := \frac{|\gamma, 1, j_1^3\rangle \otimes |\gamma, 1, j_2^3\rangle \;\; + \;\; |\gamma, 1, j_2^3\rangle \otimes |\gamma, 1, j_1^3\rangle}{\sqrt{2}}$$

[44] Following Dirac's notation, physicists write briefly the products

$$\varphi := \frac{|e^-, \tfrac{1}{2}, j_1^3\rangle\, |e^-, \tfrac{1}{2}, j_2^3\rangle \;\; - \;\; |e^-, \tfrac{1}{2}, j_2^3\rangle\, |e^-, \tfrac{1}{2}, j_1^3\rangle}{\sqrt{2}}.$$

with $j_k^3 = \pm 1$ for $k = 1, 2$. This describes the composed state of two photons of spin number $j = 1$ and polarizations j_1^3 and j_2^3. Here, $j_1^3 = j_2^3$ is possible.

(iii) The proton consisting of three colored quarks u, u, d and spin. The state

$$\begin{aligned}\psi := \frac{1}{\sqrt{108}} \sum_{\alpha,\beta,\gamma=1}^{3} \delta_{\alpha\beta\gamma}^{123} \Big(& 2|u_+^\alpha d_-^\beta u_+^\gamma\rangle + 2|u_+^\alpha u_+^\beta d_-^\gamma\rangle + 2|d_-^\alpha u_+^\beta u_+^\gamma\rangle \\ & - |u_+^\alpha u_-^\beta d_+^\gamma\rangle - |u_+^\alpha d_+^\beta u_-^\gamma - |u_-^\alpha d_+^\beta u_+^\gamma\rangle \\ & - |d_+^\alpha u_-^\beta u_+^\gamma\rangle - |d_+^\alpha u_+^\beta u_-^\gamma\rangle - |u_-^\alpha u_+^\beta d_+^\gamma\rangle \Big)\end{aligned}$$

describes a proton state with spin $\hbar/2$ in z-direction. Here, the symbol $\delta_{\alpha\beta\gamma}^{123}$ denotes the sign of the permutation $(123) \mapsto (\alpha\beta\gamma)$. The indices $\alpha, \beta, \gamma = 1, 2, 3$ describe the color charges red, green, blue, respectively. For example, the symbol $|u_+^\alpha u_+^\beta d_-^\gamma\rangle$ stands for the tensor product

$$\Big(|u\rangle \otimes |\tfrac{1}{2}\rangle \otimes |\alpha\rangle\Big) \otimes \Big(|u\rangle \otimes |\tfrac{1}{2}\rangle \otimes |\beta\rangle\Big) \otimes \Big(|d\rangle \otimes |-\tfrac{1}{2}\rangle \otimes |\gamma\rangle\Big).$$

This represents the composition of

- one u quark with spin $\hbar/2$ in the z-direction and color α,
- one u quark with spin $\hbar/2$ in the z-direction and color β, and
- one d quark with spin $-\hbar/2$ in the z-direction and color γ.

The state ψ follows in a quite natural way by using the representation theory of the compact Lie group $SU(3)$ due to Hermann Weyl. This will be studied in Volume V on the physics of the Standard Model.

2.7 The Fundamental Role of Symmetry in Physics

Symmetries essentially simplify computations in mathematics and physics.
Folklore

Symmetry, as wide or as narrow you may define its meaning, is one idea by which man through the ages has tried to comprehend and create order, beauty, and perfection.

Hermann Weyl, 1952
Symmetry[45]

Symmetry and invariance considerations have long played important roles in physics. The 32 crystal classes – that is, groups of rotations in three-dimensional space all the elements of which are of the order 2, 3, 4 or 6 – were determined 1830 by Hessel, in the same year as group theory was born by Galois.
The determination of the 230 space groups in 1891, independently by Schoenflies and Fedorov (these are the discrete subgroups of the Euclidean

[45] Princeton University Press, 1952

group which contain three non-coplanar translations) was a masterpiece of analysis and so was the determination by Groth of the possible properties of crystals with the symmetries of these space groups in 1926...
Symmetry transformations in pre-quantum theory were rather obvious transformations of 3-dimensional space; in quantum theory they became unitary transformations of Hilbert space. These form subgroups of all unitary transformations which are *essentially* homomorphic to the symmetry group in question, essentially homomorphic only because a unitary transformation in quantum mechanics is equivalent to any of its multiples by a numerical factor (of modulus 1). However, this essential homomorphy could be reduced, particularly, as a result of Bargmann's investigations in most cases to a true homomorphy to an extended group which is called, then, the quantum mechanical symmetry group.[46] The quantum mechanical operations of the symmetry group break up the Hilbert space of all states into subspaces each of which is invariant under the operations in question...

The unitary representations of the Poincaré group were determined in the late 30's; except for the trivial one, they were all shown to be infinite-dimensional.[47] This is equivalent with the statement that no system can be relativistically invariant unless it can be in an infinity of orthogonal states. By calling attention to the properties of the unitary representations of noncompact Lie groups, the physicists have stimulated the mathematicians' interest in this field. The mathematicians are now very much ahead of us in this field, and it is not easy to catch up with the results of Gelfand, Naimark, Harish-Chandra, and so many others.

Eugene Wigner, Gibbs Lecture 1968[48]
Symmetry Principles in Old and New Physics

The most important lesson that we have learned in this century is that the secret of nature is *symmetry.* Starting with relativity, proceeding through the development of quantum mechanics, and culminating in the Standard Model of particle physics, symmetry principles have assumed a central position in the fundamental theories of nature. Local gauge theories provide the basis of the Standard Model and of Einstein's theory of gravitation...
In recent years we have discovered a new and extremely powerful symmetry – *supersymmetry* – which might explain many mysteries of the Standard Model.
Another part of the lesson of symmetry is that much of the texture of the world is due to mechanisms of symmetry breaking. In quantum mechanical systems with a finite number of degrees of freedom global symmetries are realized in only one way. The laws of physics are invariant and the ground state of the theory is unique and symmetric... However, in systems with an infinite number of degrees of freedom a second realization of symmetry is possible, in which the ground state is asymmetric. This spontaneous

[46] V. Bargmann, On unitary ray representations of continuous groups, Ann. of Math. **59** (1954), 1–46.

[47] E. Wigner, On unitary representations of the inhomogeneous Lorentz group, Ann. of Math. **40** (1939), 149–204.

[48] Bull. Amer. Math. Soc. **74** (1968), 793–815. Wigner (1902–1995) was awarded the 1963 Nobel prize in physics for his contributions to symmetry principles in quantum physics.

symmetry breaking is responsible for magnetism, superconductivity, and the structure of the unified electroweak theory in the Standard Model.

The second important lesson we have learned is the idea of *renormalization group* and effective dynamics. The *decoupling* of physical phenomena at different scales of energy is an essential characteristic of nature. It is this feature of nature that makes it possible to understand the limited range of physical phenomena without having to understand everything at once. The characteristic behavior of the solutions of the renormalization equations is that they approach a finite dimensional submanifold in the infinite dimensional space of all theories. This defines an effective low energy theory... Thus, for example quantum chromodynamics is the theory of quarks whose interactions are mediated by gluons. This is the appropriate description at energies of billions of electron volts. However, if we want to describe the properties of ordinary nuclei at energies of millions of electron volts, we employ instead an effective theory of nucleons, composites of the quarks, whose interactions are mediated by other quark composites – mesons... There may be more than one, equally fundamental, formulation of a particular quantum field theory, each appropriate at a different scale of energy.

David Gross
The triumph and limitations of quantum field theory.[49]

Symmetries have always played an important role in physics. With quantum mechanics, however, the interplay between physics and symmetries has reached a new dimension. The very structure of quantum mechanics invites the application of group theoretical methods...

Symmetries are also a direct mediator between experimental facts and the theoretical structure of a theory. This is the case because there is a direct connection between *symmetries and conservation laws.* Space-time symmetries are an obvious example. Conservation of energy, momentum and angular momentum are linked to invariance under time translation, space translation and rotation in space.

It was in atomic physics that space-time symmetries became significant, but they are also important in nuclear physics and in all the physics discovered after that. In nuclear physics, however, a new concept of symmetries, *symmetries in an internal space*, was discovered with isospin. All our so-called fundamental models describing what we know about strong-weak and electromagnetic interactions are built on symmetries in space-time and internal spaces. These symmetries are not only used to extract information from a theory, they are also used to construct these theories, and this for good reasons. It turned out that only theories possessing such symmetries make sense as quantum field theories. Thus symmetries are not only a good tool to deal with quantum field theoretical models, they are necessary to define such models.

Following this line of thought, a new type symmetry, the so-called *supersymmetry*, proved to be extremely successful. Quantum theory seems to have a very deep relation to supersymmetry.[50] Thus it is not surprising

[49] In: Tian Yu Cao (Ed.), Conceptual Foundations of Quantum Field Theory (with contributions made by leading physicists), Cambridge University Press 1999, pp. 56–67 (reprinted with permission).

[50] Supersymmetry is already present in the 3-dimensional Euclidean space. This will be studied in Volume III.

> that the most promising models of physics are based on supersymmetry, even when they go beyond a local quantum field theory, as string theory does.[51]
>
> Julius Wess
> *Quantum Theory Centenary, Berlin, 2000*

For a deeper understanding of physical processes in nature, it is of fundamental importance to answer the following question.

> *Suppose that we observe a process in a physical system. Which transformed versions of this process can also happen?*

As we will show, this question is closely connected with the concept of symmetry which plays a fundamental role in modern physics. From the mathematical point of view, symmetry is described by group theory which we will encounter very frequently in this monograph. At this point, we only want to discuss some basic ideas. First let us consider three important examples: energy conservation, irreversible processes, and parity violation in weak interaction.

(i) *Time translation and energy conservation.* Consider the motion $q = q(t)$ of a classical particle of mass $m > 0$ on the real line. The Newtonian equation reads as

$$m\ddot{q}(t) = F(q(t)), \qquad t \in \mathbb{R}. \tag{2.58}$$

Since the given smooth force $F = F(q)$ only depends on position q, but not on time t, the following hold true for each fixed time t_0:

> *If the smooth function $q = q(t)$ is a solution of* (2.58)*, then so is the function $q = q(t + t_0)$.*

We say that equation (2.58) is invariant under time translations. This implies conservation of energy. To show this, choose a function U such that $F(q) = -U'(q)$ for all $q \in \mathbb{R}$. Introduce the energy function

$$E(t) := \frac{m\dot{q}(t)^2}{2} + U(q(t)).$$

For each solution $q = q(t)$ of (2.58), we then get

$$E(t) = \text{const} \qquad \text{for all} \quad t \in \mathbb{R}.$$

In fact, $\dot{E}(t) = m\ddot{q}(t)\dot{q}(t) + U'(q(t))\dot{q}(t) = 0$. Note that this argument fails if the force $F = F(q, t)$ depends on time t.

[51] Max Planck presented his radiation law to the Deutsche Physikalische Gesellschaft on December 14, 1900. The Proceedings of the *Symposia Quantum Theory Cetenary* December 2000 in Berlin appeared in Annalen der Physik (Leipzig), Vol. 9, 11/12 (2000) and Vol. 10, 1/2 (2001) (reprinted with permission). These two volumes contain survey articles on the modern development of quantum physics.

(ii) *Irreversible processes.* A process is called reversible iff it is invariant under time-reversal. Otherwise, the process is called irreversible. In classical mechanics, the motion of a mass point is reversible if the force is time-independent. For example, if $q = q(t)$ is a solution of (2.58), then so is $q = q(-t)$. The situation changes completely if there acts a friction force. For example, consider the equation

$$m\ddot{q}(t) = -\kappa\dot{q}(t), \qquad t \in \mathbb{R} \tag{2.59}$$

with the friction force $F = -\kappa\dot{q}$ depending on velocity. Here, the real number κ is positive. This problem has the solution

$$q(t) = \mathrm{e}^{-\kappa t/m}, \qquad t \in \mathbb{R}.$$

The motion is irreversible. To see this, observe that the time-reversed process $q = q(-t) = e^{\kappa t/m}$ is not a solution of (2.59). Since the equation (2.59) is invariant under time translations, we expect energy conservation. In fact, introducing the kinetic energy function

$$E(t) := \frac{m\dot{q}(t)^2}{2}$$

along with the heat function $Q(t) := \kappa \int_0^t \dot{q}(\tau)^2 d\tau$, we get

$$\boxed{E(t) - E(0) + Q(t) = \text{const} \qquad \text{for all} \quad t \in \mathbb{R}.}$$

This is the first law of thermodynamics, telling us that the loss of kinetic energy $E(t) - E(0)$ is compensated by the production of heat energy $Q(t)$ during the time interval $[0, t]$. In fact,

$$\dot{E}(t) = m\ddot{q}(t)\dot{q}(t) = -\kappa\dot{q}(t)^2 = -\dot{Q}(t).$$

By friction, the real line is heated up. Let $T(t)$ denote the temperature of the real line at time t. Define the external entropy function S_e by the differential equation

$$\boxed{\dot{S}_e(t) := \frac{\dot{Q}(t)}{T(t)}.}$$

Explicitly,

$$S_e(t) = S_e(0) + \int_0^t \frac{\kappa\dot{q}(\tau)^2}{T(\tau)} d\tau \qquad \text{for all} \quad t \in \mathbb{R}.$$

The second law of thermodynamics tells us that there exists a function $S = S(t)$ called total entropy such that

$$\dot{S}(t) \geq \frac{\dot{Q}(t)}{T(t)} \qquad \text{for all} \quad t \in \mathbb{R}.$$

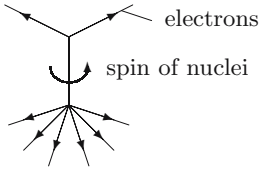


Fig. 2.21. Parity violation in β-decay of $^{60}_{27}$Co

Introducing the internal entropy S_i by setting

$$S(t) = S_e(t) + S_i(t),$$

we get the inequality $\dot{S}_i(t) \geq 0$ or all times $t \in \mathbb{R}$. Different choices of the function S_i lead to different thermodynamic models. The model is called reversible iff

$$\dot{S}(t) = \frac{\dot{Q}(t)}{T(t)} \qquad \text{for all} \quad t \in \mathbb{R}.$$

This means that $S_i(t) = \text{const}$ for all $t \in \mathbb{R}$. For example, a simple reversible model is obtained by setting $\kappa = 0$ along with $S_i(t) :\equiv 0$, and

$$S(t) = S_e(t) := \text{const} \qquad \text{for all} \quad t \in \mathbb{R}.$$

A detailed study of both phenomenological and statistical thermodynamics can be found in Zeidler (1986), Vol. IV. We will show in Volume IV of the present monograph concerning quantum mathematics that entropy is equivalent to information. This explains why entropy plays such a crucial role in nature.

The time-evolution of a living being is always irreversible. Indeed, one never observes that an old man develops back in time into a baby. Since past and future can be distinguished by observing living beings, physicists say that time is directed, or, time is equipped with an arrow.

(iii) *Parity violation in weak interaction.* The Andromeda galaxy is 2 million light years far away from the earth. Suppose that there exist intelligent creatures on a star of the Andromeda galaxy. We want to tell them the definition of right-handed orientation. On earth we only need the first three fingers of our right hand. But how can we communicate this positive orientation to distant creatures which might not have a right hand? The answer to this is an experiment that was performed by Mrs. Chien-Shiung Wu and coworkers at Columbia University, New York in 1957. The effect is that the β-decay of $^{60}_{17}$Co-kernels results in the emission of 30 per cent more electrons antiparallel to the spin direction than parallel (Fig. 2.21 on page 167). This shows that processes of weak interaction are not always invariant under space reflections. Physicists call this the violation of parity in weak interaction. Mrs. Wu starts her fundamental paper (Phys. Rev. **105** (1957), 1431–1432) in the following way:

> In a recent paper on the question of parity in weak interaction (Phys. Rev. **104** (1956)), Lee and Yang critically surveyed the experimental information concerning this question and reached the conclusion that there is no existing evidence either to support or to refuse parity conservation in weak interactions. They proposed a number of experiments on beta decays and hyperon and meson decays which would provide the necessary evidence for parity conservation or non-conservation...

In 1957 Chen Ning Yang and Tsung Dao Lee were awarded the Nobel prize in physics for their theoretical contributions to parity laws in weak interaction. Mrs. Wu was awarded the first Wolf prize in physics in 1978.

Asymmetry under space reflections plays an important role in chemistry, too. For example, amino acids existing on earth are left-handed.

2.7.1 Classical Symmetries

Consider the motion $\mathbf{x} = \mathbf{x}(t),\ t \in \mathbb{R}$, of a mass point in the 3-dimensional Euclidean space. The following transformations play a fundamental role.

(i) Translation: $\mathbf{x}(t) + \mathbf{a}$.
(ii) Rotation: $R\mathbf{x}(t)$.
(iii) Space reflection: $-\mathbf{x}(t)$.
(iv) Time translation: $\mathbf{x}(t + t_0)$.
(v) Time reversal: $\mathbf{x}(-t)$.
(vi) Similarity transformation: $\lambda\mathbf{x}(t)$.
(vii) Combined rescaling: $\lambda^k\mathbf{x}(\lambda^n t)$.
(viii) Rescaling of all physical units (choice of a new system of physical units).
(ix) Charge conjugation: Passage from a particle to its antiparticle. Replace the charge Q by $-Q$.

If, in a physical system, the realization of the process $\mathbf{x} = \mathbf{x}(t)$ implies that the process $\mathbf{x} = \mathbf{x}(t) + \mathbf{a}$ is also possible, then we say that the physical system is invariant under space translations. Similarly, we say that the system is invariant under rotations, reflections, and so on. Such symmetries have important consequences. For example,

> *Invariance under time translations, space translations, and rotations imply conservation of energy, momentum, and angular momentum, respectively.*

This is a special case of the famous Noether theorem telling us that

> *Symmetries imply conservation laws.*

This will be studied in Volume II.

Dimensional analysis. The very useful method (viii) above of dimensional analysis in physics is based on the fact that physical quantities carry a physical dimension. In this context, dimensionless physical quantities, like

Table 2.11. Symmetry

invariance under transformations	physical effect
time translation	conservation of energy
space translation	conservation of momentum
rotation	conservation of angular momentum (spin)
Lorentz transformation of space-time coordinates	theory of special relativity
Poincaré transformation of space-time coordinates	energy-momentum tensor of relativistic fields; CPT invariance of elementary particle processes; quantum numbers
time reversal	reversible processes; antiparticles
space reflection	orientation; parity of elementary particles
diffeomorphism (change of general space-time coordinates)	general theory of relativity
permutation	bosons and fermions; Pauli's exclusion principle; supersymmetry
unitary gauge transformation	gauge theory and interaction; quantum numbers (e.g., electric charge); Standard Model of particle physics
similarity transformation	rescaling of processes; dimensional analysis; renormalization group (semigroup)
conformal transformation	rescaling of processes; critical phenomena and phase transitions

the fine structure constant α, play an outstanding role in physics. This will be studied in the Appendix A.5 on page 944. There we will show that, surprisingly enough,

- Newton's gravitational law, and
- Kolmogorov's law for turbulent flow

can be obtained from general rescaling arguments without using the specifics of the physical situation.

Einstein's theory of special relativity. Newton's mechanics is invariant under the Galileian transformation[52]

[52] This is thoroughly studied in Zeidler (1986), Vol. IV, p. 28.

$$\mathbf{x}' = \mathbf{x} - \mathbf{v}t, \qquad t' = t.$$

In contrast to this, the Maxwell equations are not invariant under Galileian transformations, but under Lorentz transformations (2.37) on page 110. Einstein postulated that one has to use Lorentz transformations instead of Galileian transformations. This led him to his revolutionary theory of special relativity in 1905 which did not change the Maxwell equations, but Newton's classical equation of motion.

Prediction of new elementary particles via symmetry. Important symmetries appearing in modern physics are listed in Table 2.11 on page 169. The development of modern elementary particle physics has strongly been influenced by the use of symmetries. The existence of new particles has always been predicted in order to avoid lack of symmetry. For example, Gell-Mann predicted the existence of quarks by using symmetry arguments based on the representation theory of the Lie group $SU(3)$.

Conformal field theory. There holds the following rule of thumb:

> *The larger a symmetry group is, the stronger are its implications for the structure of the physical system.*

For example, in two dimensions, analytic functions generate conformal mappings. Therefore, the family of conformal mappings is huge in two dimensions. Conformal field theory studies physical systems that are invariant under conformal mappings (general rescaling). Typical examples are

- string theory,
- 2-dimensional quantum field theories, and
- 2-dimensional lattice models near phase transitions in statistical physics.

In these cases, conformal symmetry determines the Green's functions in an almost unique manner. Let us motivate why conformal symmetry might be important for phase transitions. Typically, the physical system (e.g., the lattice model) possesses a characteristic length scale given by the finite correlation length. However, near a phase transition, the fluctuations become large and hence the correlation length becomes infinite in the limit. In other words, the system loses its finite characteristic length scale in a phase transition, and it becomes invariant under rescaling (conformal transformations).

Observe that the Maxwell equations in electromagnetism do not distinguish a characteristic length scale. Therefore, we expect that the Maxwell equations are invariant under 4-dimensional conformal transformations of the Minkowski space. This is indeed true. We will study this in Volume III.

2.7.2 The CPT Symmetry Principle for Elementary Particles

> Elementary particle processes are invariant under the CPT symmetry. This is one of the most fundamental symmetries in physics.
>
> Folklore

The Maxwell equations in electrodynamics. Suppose that we are given an electric field $\mathbf{E} = \mathbf{E}(\mathbf{x}, t)$ and a magnetic field $\mathbf{B} = \mathbf{B}(\mathbf{x}, t)$ in vacuum. According to Maxwell, these fields always satisfy the following four equations

$$\begin{aligned} \varepsilon_0 \operatorname{div} \mathbf{E} &= \varrho, & \operatorname{div} \mathbf{B} &= 0, \\ \mathbf{curl}\, \mathbf{E} &= -\dot{\mathbf{B}}, & \mathbf{curl}\, \mathbf{B} &= \frac{\dot{\mathbf{E}}}{c^2} + \mu_0 \mathbf{J}. \end{aligned} \tag{2.60}$$

Here, we use the following notation: ϱ electric charge density, $\mathbf{J}$ electric current density vector, ε_0 electric field constant of a vacuum, c velocity of light in a vacuum, $\mu_0 = 1/\varepsilon_0 c^2$ magnetic field constant of a vacuum. The Maxwell equations (2.60) are valid in each inertial system. If we pass from one inertial system to another one, then we have to use the Lorentz transformation for the space-time coordinates and the tensor transformation law for the electromagnetic field tensor. The explicit transformation formulas for the electromagnetic field will be considered in Volume III. Note that it may happen that the electric field is transformed into a magnetic field and vice versa. Choosing a right-handed Cartesian (x, y, z)-coordinate system with orthonormal basis vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$, we have

$$\mathbf{x} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}, \qquad \boldsymbol{\partial} := \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}.$$

Let $\mathbf{E} = E^1\mathbf{i} + E^2\mathbf{j} + E^3\mathbf{k}$. By classical vector calculus, recall that

$$\mathbf{grad}\, U := \boldsymbol{\partial} U = U_x\mathbf{i} + U_y\mathbf{j} + U_z\mathbf{k}, \qquad \operatorname{div} \mathbf{E} := \boldsymbol{\partial}\mathbf{E} = E^1_x + E^2_y + E^3_z,$$

and

$$\mathbf{curl}\, \mathbf{E} := \boldsymbol{\partial} \times \mathbf{E} = (E^3_y - E^2_z)\mathbf{i} + (E^1_z - E^3_x)\mathbf{j} + (E^2_x - E^1_y)\mathbf{k}.$$

Let us introduce the following three operators C, P, T.

(i) Charge conjugation $Q \mapsto -Q$:

$$\begin{aligned} (C\varrho)(\mathbf{x}, t) &:= -\varrho(\mathbf{x}, t), & (C\mathbf{J})(\mathbf{x}, t) &:= -\mathbf{J}(\mathbf{x}, t), \\ (C\mathbf{E})(\mathbf{x}, t) &:= -\mathbf{E}(\mathbf{x}, t), & (C\mathbf{B})(\mathbf{x}, t) &:= -\mathbf{B}(\mathbf{x}, t). \end{aligned}$$

(ii) Parity transformation $\mathbf{x} \mapsto -\mathbf{x}$:

$$\begin{aligned} (P\varrho)(\mathbf{x}, t) &:= \varrho(-\mathbf{x}, t), & (P\mathbf{J})(\mathbf{x}, t) &:= -\mathbf{J}(-\mathbf{x}, t), \\ (P\mathbf{E})(\mathbf{x}, t) &:= -\mathbf{E}(-\mathbf{x}, t), & (P\mathbf{B})(\mathbf{x}, t) &:= \mathbf{B}(-\mathbf{x}, t). \end{aligned}$$

(iii) Time reversal $t \mapsto -t$:

$$\begin{aligned} (T\varrho)(\mathbf{x}, t) &:= \varrho(\mathbf{x}, -t), & (T\mathbf{J})(\mathbf{x}, -t) &:= -\mathbf{J}(\mathbf{x}, -t), \\ (T\mathbf{E})(\mathbf{x}, t) &:= \mathbf{E}(\mathbf{x}, -t), & (T\mathbf{B})(\mathbf{x}, t) &:= -\mathbf{B}(\mathbf{x}, -t). \end{aligned}$$

A simple computation shows that

The Maxwell equations (2.60) *are invariant under the transformations* $C, P,$ *and* T.

For the transformation T, this means the following: If $\mathbf{E}, \mathbf{B}, \varrho, \mathbf{J}$ are solutions of the Maxwell equations (2.60), then so are

$$T\mathbf{E},\ T\mathbf{B},\ T\varrho,\ T\mathbf{J}.$$

The same is true, if we replace T by either C or P.

It was discovered by Maxwell that the electromagnetic field can be represented in the form

$$\boxed{\mathbf{E} = -\,\mathbf{grad}\, U - \dot{\mathbf{A}}, \qquad \mathbf{B} = \mathbf{curl}\, \mathbf{A}.}$$

Here, U and $\mathbf{A}$ are called the scalar and the vector potential of the electromagnetic field, respectively. As we will show on page 174, the potential functions U and $\mathbf{A}$ are not uniquely determined by the electromagnetic field $\mathbf{E}, \mathbf{B}$. This is intimately related to the idea of gauge symmetry in elementary particle physics. In order to obtain crucial transformation laws for the electromagnetic field, we introduce the operators C, P, T in the following way.

(i) Charge conjugation $Q \mapsto -Q$:

$$(CU)(\mathbf{x}, t) := -U(\mathbf{x}, t), \qquad (C\mathbf{A})(\mathbf{x}, t) := -\mathbf{A}(\mathbf{x}, t).$$

(ii) Parity transformation $\mathbf{x} \mapsto -\mathbf{x}$:

$$(PU)(\mathbf{x}, t) := U(-\mathbf{x}, t), \qquad (P\mathbf{A})(\mathbf{x}, t) := -\mathbf{A}(-\mathbf{x}, t).$$

(iii) Time reversal $t \mapsto -t$:

$$(TU)(\mathbf{x}, t) := U(\mathbf{x}, -t), \qquad (T\mathbf{A})(\mathbf{x}, t) := -\mathbf{A}(\mathbf{x}, -t).$$

Note that $(c\varrho, \mathbf{J})$ and $(U/c, \mathbf{A})$ transform like $(ct, \mathbf{x})$ under time reversal and space reflections.[53]

The Schrödinger–Maxwell equation. The equation for the wave function $\psi = \psi(\mathbf{x}, t)$ of a quantum particle of mass m and electric charge Q in an electromagnetic field $\mathbf{E}, \mathbf{B}$ reads as

$$\boxed{\mathrm{i}\hbar\dot{\psi} = \frac{(\mathbf{P} - Q\mathbf{A})^2}{2m}\psi + QU\psi} \tag{2.61}$$

[53] As we will show in Volume III, this remains true for arbitrary Lorentz transformations.

along with the generalized momentum operator $\mathbf{P} = -\mathrm{i}\hbar\,\mathbf{grad}$ and

$$\mathbf{E} = -\,\mathbf{grad}\,U - \dot{\mathbf{A}}, \qquad \mathbf{B} = \mathbf{curl}\,\mathbf{A}.$$

If we use the normalization condition $\int_{\mathbb{R}^3} |\psi(\mathbf{x},t)|^2 d^3x = 1$, then the integral

$$\int_{\Omega} |\psi(\mathbf{x},t)|^2 d^3x$$

tells us the probability for finding the particle in the region Ω of the Euclidean space at time t. The operators C, P, and T for the wave function are defined as follows.

(i) Charge conjugation $Q \mapsto -Q$: $(C\psi)(\mathbf{x},t) := \psi(\mathbf{x},t)$.
(ii) Parity operation $\mathbf{x} \mapsto -\mathbf{x}$: $(P\psi)(\mathbf{x},t) := \psi(-\mathbf{x},t)$.
(iii) Time reversal $t \mapsto -t$: $(T\psi)(\mathbf{x},t) := \psi(\mathbf{x},-t)^\dagger$.

Suppose that the potential U only depends on the modulus $||\mathbf{x}||$ of the position vector $\mathbf{x}$. By a simple computation, we obtain that

> *The Schrödinger–Maxwell equation* (2.61) *is invariant under the combined CPT transformation.*

In addition, the Schrödinger–Maxwell equation is also invariant under gauge symmetry. This will be considered in Sect. 2.7.3.

The basic symmetry of elementary particle processes. In contrast to the Schrödinger–Maxwell equation, general processes for elementary particles are not always invariant under the single operations C, P, T. However, the following holds true.

> *Elementary particle processes are invariant under the combined operation CPT.*

Intuitively, suppose that we observe a certain process for elementary particles. Perform the following operations:

- reverse the time direction,
- replace particles by antiparticles and vice versa, and
- consider the process in a mirror.

This way, we get a new process which is also possible in nature.

In the 1950s, the fundamental CPT invariance principle was discovered by Schwinger and Lüders and perfected by Pauli. A rigorous proof was given by Res Jost (1965) on the basis of the Gårding–Whightman axiomatic quantum field theory introduced in the 1950s. Jost used sophisticated arguments based on the analytic continuation of functions of several complex variables.[54]

[54] See also Streater and Whightman (1968), Reed and Simon (1972), Vol. II, and Bogoliubov et al. (1990).

2.7.3 Local Gauge Symmetry

Local gauge symmetry plays a crucial role in elementary particle physics. In fact, the Standard Model of particle physics is based on local gauge symmetry. At this point, let us only discuss the basic idea of gauge invariance by considering the Schrödinger–Maxwell equation as a prototype.

The Maxwell equations. Let $f = f(\mathbf{x}, t)$ be a real-valued smooth function on $\mathbb{R}^4$. The transformation

$$U^g := U - \dot{f}, \qquad \mathbf{A}^g := \mathbf{A} + \mathbf{grad}\, f$$

is called a gauge transformation of the 4-potential $U, \mathbf{A}$. Recall that the dot denotes the time derivative. The point is that such a gauge transformation leaves the electromagnetic field invariant, that is,

$$\boxed{\mathbf{E}^g = \mathbf{E}, \qquad \mathbf{B}^g = \mathbf{B}.}$$

This follows from

$$\begin{aligned}\mathbf{E}^g &= -\,\mathbf{grad}\; U^g - (\dot{\mathbf{A}})^g = -\,\mathbf{grad}\; U + \mathbf{grad}\, \dot{f} - \dot{\mathbf{A}} - \mathbf{grad}\, \dot{f} \\ &= -\,\mathbf{grad}\; U - \dot{\mathbf{A}} = \mathbf{E},\end{aligned}$$

and $\mathbf{B}^g = \mathbf{curl}\, \mathbf{A}^g = \mathbf{curl}\, \mathbf{A} = \mathbf{B}$. In particular, the Maxwell equations (2.60) are invariant under gauge transformations.

The Schrödinger–Maxwell equation. Let $\alpha = \alpha(\mathbf{x}, t)$ be a smooth real-valued function on $\mathbb{R}^4$. Consider the phase transformation

$$\boxed{\psi^g(\mathbf{x}, t) := \mathrm{e}^{\mathrm{i}\alpha(\mathbf{x},t)}\, \psi(\mathbf{x}, t)}$$

along with the transformation

$$\boxed{U^g := U - \dot{f}, \qquad \mathbf{A}^g := \mathbf{A} + \mathbf{grad}\, f.}$$

In addition, we assume that the phase function α is related to the transformation function f by

$$f(\mathbf{x}, t) = \frac{\hbar}{Q} \cdot \alpha(\mathbf{x}, t).$$

The transformation $(\psi, U, \mathbf{A}) \mapsto (\psi^g, U^g, \mathbf{A}^g)$ is called a gauge transformation of the Schrödinger–Maxwell equation (2.61). Since the wave function ψ is transformed by a phase factor $\mathrm{e}^{\mathrm{i}\alpha(\mathbf{x},t)}$ which depends on space and time, the gauge transformation is also called a local symmetry transformation.

Theorem 2.4 *The Schrödinger–Maxwell equation* (2.61) *is invariant under gauge transformations.*

Explicitly, if $\psi, U, \mathbf{A}$ satisfy the equation (2.61) on page 172, then so do $\psi^g, U^g, \mathbf{A}^g$. This tells us that the Schrödinger–Maxwell equation describes the prototype of a gauge field theory. In fact, we will show in Volume III that the Standard Model of particle physics is of this type.

Proof. To simplify notation, choose the energetic system with $\hbar = 1$.

(I) Covariant derivative. In order to get insight, let us introduce the covariant derivatives

$$\boxed{\nabla_t \psi := \frac{\partial \psi}{\partial t} + \mathrm{i}QU\psi} \tag{2.62}$$

and

$$\boxed{\nabla \psi := \mathbf{grad}\, \psi - \mathrm{i}Q\mathbf{A}\psi} \tag{2.63}$$

which generalize the classical partial derivatives of the wave function ψ. For the components in a right-handed Cartesian coordinate systems, we also write

$$\nabla_j \psi := \frac{\partial \psi}{\partial x^j} - \mathrm{i}QA_j\psi, \qquad j = 1, 2, 3$$

where A_1, A_2, A_3 denote the components of the vector $\mathbf{A}$. The point is that

> *In contrast to the classical partial derivatives, the covariant derivatives transform like the wave function ψ under gauge transformations.*

Explicitly, we will prove below that

(i) $\nabla_t^g \psi^g = \mathrm{e}^{\mathrm{i}\alpha} \nabla_t \psi$ and
(ii) $\nabla^g \psi^g = \mathrm{e}^{\mathrm{i}\alpha} \nabla \psi$.

Generally, transformation laws of this type are crucial for gauge field theory, as we will show in Volume III.

(II) Gauge invariance of the Schrödinger–Maxwell equation. Observe that

$$\mathbf{P} - Q\mathbf{A} = -\mathrm{i}(\mathbf{grad} - \mathrm{i}Q\mathbf{A}) = -\mathrm{i}\nabla.$$

Therefore, using covariant derivatives, the Schrödinger–Maxwell equation (2.61) on page 172 reads elegantly as

$$\mathrm{i}\nabla_t \psi = -\frac{\nabla^2 \psi}{2m}.$$

By (i) and (ii), this implies

$$\mathrm{i}\nabla_t^g \psi^g = -\frac{(\nabla^g)^2 \psi^g}{2m}$$

which is the desired invariance of the Schrödinger–Maxwell equation under gauge transformations. In fact,

$$(\nabla^g)^2\psi^g = \nabla^g(\nabla^g\psi^g) = \nabla^g(\mathrm{e}^{\mathrm{i}\alpha}\nabla\psi) = \mathrm{e}^{\mathrm{i}\alpha}\nabla(\nabla\psi).$$

In addition, $\nabla_t^g\psi^g = \mathrm{e}^{\mathrm{i}\alpha}\nabla_t\psi$.

(III) Proof of (i). Observe that

$$\begin{aligned}\nabla_t^g\psi^g &= \frac{\partial}{\partial t}(\mathrm{e}^{\mathrm{i}\alpha}\psi) + \mathrm{i}QU^g\psi^g \\ &= (\mathrm{i}\dot{\alpha}\psi + \dot{\psi})\mathrm{e}^{\mathrm{i}\alpha} + \mathrm{i}QU\psi\mathrm{e}^{\mathrm{i}\alpha} - \mathrm{i}\dot{\alpha}\psi\mathrm{e}^{\mathrm{i}\alpha}.\end{aligned}$$

Hence $\nabla_t^g\psi^g = \mathrm{e}^{\mathrm{i}\alpha}(\dot{\psi} + \mathrm{i}QU\psi) = \mathrm{e}^{\mathrm{i}\alpha}\nabla_t\psi$.

(IV) Proof of (ii). By the product rule,

$$\begin{aligned}\nabla^g\psi^g &= \mathbf{grad}\,(\mathrm{e}^{\mathrm{i}\alpha}\psi) - \mathrm{i}Q\left(\mathbf{A} + \mathbf{grad}\,\frac{\alpha}{Q}\right)\mathrm{e}^{\mathrm{i}\alpha}\psi \\ &= \psi\,\mathbf{grad}\,\mathrm{e}^{\mathrm{i}\alpha} + \mathrm{e}^{\mathrm{i}\alpha}\,\mathbf{grad}\,\psi - \mathrm{i}Q\mathbf{A}\mathrm{e}^{\mathrm{i}\alpha}\psi - \mathrm{i}\mathrm{e}^{\mathrm{i}\alpha}\psi\,\mathbf{grad}\,\alpha.\end{aligned}$$

Since $\mathbf{grad}\,\mathrm{e}^{\mathrm{i}\alpha} = \mathrm{i}\mathrm{e}^{\mathrm{i}\alpha}\,\mathbf{grad}\,\alpha$, we get

$$\nabla^g\psi^g = \mathrm{e}^{\mathrm{i}\alpha}(\mathbf{grad}\,\psi - \mathrm{i}Q\mathbf{A}\psi) = \mathrm{e}^{\mathrm{i}\alpha}\nabla\psi.$$

□

2.7.4 Permutations and Pauli's Exclusion Principle

The states of an elementary particle system are invariant under permutations of the particles (principle of indistinguishable particles). More precisely, we use

- even permutations for bosons, and
- odd permutations for fermions.

This guarantees that, as in (2.57), two identical fermions can never stay together in the same state (Pauli's exclusion principle). According to Hermann Weyl, permutation groups also play a crucial role in order to determine the irreducible representations of the Lie group $SU(3)$ which are fundamental for elementary particle physics.[55]

2.7.5 Crossing Symmetry

Suppose that a reaction of the form

$$A + B \to C + D$$

[55] This will be studied in Volume V on the physics of the Standard Model. See also the classic monograph by Weyl (1938) on the representations of the classical Lie groups.

for elementary particles A, B, C, D is observed in an experiment. Replacing the particle B by its antiparticle $\overline{B}$ and crossing it over to the other side of the equation, we get

$$A \to \overline{B} + C + D.$$

Similarly,

$$A + \overline{C} \to \overline{B} + D, \qquad \overline{C} + \overline{D} \to \overline{A} + \overline{B}.$$

The crossing symmetry principle tells us that all of these processes are possible if conservation of energy is not violated. The same is true for the reversed process $C + D \to A + B$.

2.7.6 Forbidden Spectral Lines in Molecules

> The representation theory for groups is an extremely useful tool in order to understand the structure of both molecule spectra and scattering processes for elementary particles. The same concerns the structure of processes for nuclei of molecules.
>
> Folklore

Suppose that an electron of a molecule jumps from the nth orbit with energy E_n to the mth orbit with lower energy E_m. According to Heisenberg and Schrödinger, the transition probability for this jump is equal to

$$\gamma_{nm} := \frac{e^2 \omega_{nm}^3}{3\pi c^3 \hbar} \, |\langle \psi_m | \mathbf{x} \psi_n \rangle|^2$$

with the dipol transition amplitude

$$\langle \psi_m | \mathbf{x} \psi_n \rangle := \int_{\mathbb{R}^3} \psi_m(\mathbf{x}, t)^\dagger \mathbf{x} \psi_n(\mathbf{x}, t) d^3x. \tag{2.64}$$

Here, e is the electric charge of the electron, and $\omega_{nm} = (E_n - E_m)/\hbar$ is the angular frequency of the emitted photon. Moreover,

$$\psi_n(\mathbf{x}, t) = \mathrm{e}^{-\mathrm{i}E_n t/\hbar} \varphi(\mathbf{x})$$

is the Schrödinger wave function of the electron on the nth orbit. The jump of the electron is impossible as dipole radiation iff

$$\boxed{\langle \psi_m | \mathbf{x} \psi_n \rangle = 0.} \tag{2.65}$$

For example, condition (2.65) is satisfied if ψ_n and ψ_m have the same parity. In fact, in this case, ψ_n and ψ_m are both even (resp. odd) with respect to $\mathbf{x}$. Thus, the integrand from (2.64) is an odd function with respect to $\mathbf{x}$, and hence the integral vanishes. More generally, one can use the representation theory of the rotation group $SO(3)$ in order to get (2.65) for special values of angular momentum. In addition, the Wigner–Eckardt theorem tells

us the structure of the transition amplitude in terms of general values of angular momentum. Similarly, symmetry properties of the S-matrix determine the structure of transition probabilities for scattering processes of elementary particles, including cross sections. This will be studied in later volumes. As an introduction to the use of group theory in physics, we recommend the monographs by van der Waerden (1932), Novoshilov (1973), Fuchs and Schweigert (1997).

2.8 Symmetry Breaking

Imperfection of matter sows the seed of death.
Thomas Mann (1875–1955)

Many phenomena in nature can be understood by using the fact that symmetries are disturbed under an external influence. We speak of symmetry breaking which is frequently related to phase transition. As a typical process, consider the cooling of water. At a critical temperature, water is transformed into bizarre flowers of ice by a phase transition. Obviously, the ice flowers possess a lower degree of symmetry than the homogeneous water. Mathematically, it is much easier to describe water than ice flowers. Similarly, physicists expect that shortly after the Big Bang, there existed only one fundamental force. In the process of cooling the universe, the gravitational force, the strong force, the weak force, and the electromagnetic force crystallized out step by step.

2.8.1 Parity Violation and CP Violation

Recall that the violation of mirror symmetry (i.e., parity P) in weak interaction was experimentally discovered by Mrs. Wu in 1955. After that discovery, physicists hoped for some time that at least the combination of mirror symmetry and passage to antiparticles would be a universal symmetry in the world of elementary particles (i.e., CP symmetry). For example, consider the kaon decay

$$K^0 \to \pi^+ + e^- + \overline{\nu}_e \tag{2.66}$$

caused by weak interaction. If CP symmetry would hold true, then the corresponding antiparticle reaction

$$K^0 \to \pi^- + e^+ + \nu_e$$

considered in a mirror would have the same decay rate as the original process (2.66). But in 1964 it was established experimentally by Cronin and Fitch that this is not true. In 1980, Cronin and Fitch were awarded the Noble prize in physics for the discovery of CP violation in nature. However, note that the effect of CP violation is small.

2.8.2 Irreversibility

The time-evolution of living beings is not reversible. There arises the important question whether processes for elementary particles are always reversible. The answer is 'no'. In fact, each process for elementary particles is invariant under the combined CPT symmetry transformation. If time reversal T would be a universal symmetry, then the CP symmetry would be universally realized. However, this contradicts the CP violation observed in experiments.

2.8.3 Splitting of Spectral Lines in Molecules

Many effects in quantum physics are based on the perturbation of degenerate quantum states. For example, we speak of a degenerate energy E iff there exist n linearly independent states

$$|1\rangle, |2\rangle, \ldots, |n\rangle$$

in a Hilbert space which possess the same energy E. Under a small perturbation, we obtain n slightly perturbed states

$$|1'\rangle, |2'\rangle, \ldots, |n'\rangle$$

with the n energies $E'_1, E'_2, \ldots, E'_n$. This splitting of energy values leads to a splitting of spectral lines which can be observed by physical experiment. As a simple mathematical model, consider the eigenvalue problem

$$H(\varepsilon)|k\rangle = E_k|k\rangle, \qquad k = 1, 2$$

with the matrices

$$H(\varepsilon) := \begin{pmatrix} 1+\varepsilon & 0 \\ 0 & 1-\varepsilon \end{pmatrix}, \quad |1\rangle := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |2\rangle := \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

For $\varepsilon = 0$, we get the degenerate eigenvalue $E_1 = E_2 = 1$, which splits into the two eigenvalues $E_1 = 1 + \varepsilon$ and $E_2 = 1 - \varepsilon$ for $\varepsilon \neq 0$.

The Zeeman effect. For example, consider the electron of the hydrogen atom on the nth orbit with angular momentum l. This electron can be in the $2l + 1$ states

$$|l, m\rangle, \qquad m = l, l-1, \ldots, -l$$

with energy E_n where $l = 0, \ldots, n-1$, and $n = 1, 2, \ldots$. The hydrogen atom is invariant under rotations. If we put the hydrogen atom into a constant magnetic field $\mathbf{B}$, then the full rotational symmetry is broken; it has to be replaced by the rotational symmetry about the axis $\mathbf{B}$. In 1926 Schrödinger introduced his wave equation for the hydrogen atom, he computed the energies and wave functions for the hydrogen atom, and he developed stationary

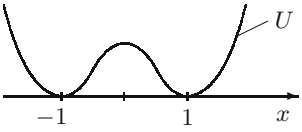


Fig. 2.22. Spontaneous symmetry breaking

perturbation theory.[56] Using this, Schrödinger showed that, under the influence of the magnetic field $\mathbf{B}$, the energy E_n of the unperturbed electron splits into the $2l+1$ values

$$E_n + m\Delta E, \qquad m = l, l-1, \ldots, -l.$$

This energy splitting is observed in the spectrum of the hydrogen atom as Zeeman effect. For this experimental discovery, Pieter Zeeman was awarded the Nobel prize in physics in 1902. The numbers l and m of the electron state $|l, m\rangle$ are called angular quantum number and magnetic quantum number, respectively.

The quark model. Similar group-theoretic arguments were used by Gell-Mann in 1964. Organizing the mass spectrum of baryons and mesons as pictured in Fig. 2.20 on page 158 and using the symmetry group $SU(3)$, Gell-Mann concluded that these particles (e.g., proton and neutron) are composed particles. This led him to formulate the quark hypothesis which was established experimentally a few years later by proton scattering experiments at the Stanford linear accelerator SLAC.

2.8.4 Spontaneous Symmetry Breaking and Particles

If a physical process does not possess maximal symmetry, then we speak of spontaneous symmetry breaking. For example, the gravitational field of the sun is invariant under rotations. Orbits of maximal symmetry are circles. In order to compute the orbit of the planet Mars on the basis of Tycho Brahe's experimental data, Kepler started with the ansatz of a circle. After time-consuming computations,[57] Kepler found out that the orbit of Mars is an ellipse which was published in his *Astronomia Nova* from 1609. It was discovered in the 1960s by Goldstone and Higgs that spontaneous symmetry breaking of the ground state of a quantum field enables us to model mathematically the emergence of new particles called Goldstone and Higgs particles. To explain the basic idea, consider the motion $q = q(t)$ of a particle of mass m in the complex plane

[56] Perturbation theory for time-dependent processes was invented by Dirac in 1927. As an introduction to perturbation theory, we recommend the classic monographs by Kato (1966), Reed and Simon (1972), Vol. 4.

[57] Note that Kepler could not use logarithms at this time, since logarithms were introduced by John Napier only in 1614.

$$\boxed{m\ddot{q}(t) = -U'(|q(t)|), \qquad t \in \mathbb{R}} \tag{2.67}$$

under the influence of the Landau–Ginzburg potential

$$U(x) = (x^2 - 1)^2, \qquad x \in \mathbb{R}$$

as pictured in Fig. 2.22. Here, $q(t)$ is a complex number.

(i) Symmetry of the vacuum. Observe that the potential $U(|q|)$ is invariant under the group $U(1)$ of rotations

$$q \mapsto \mathrm{e}^{\mathrm{i}\varphi} q$$

about the origin in the complex plane $\mathbb{C}$. Here, the rotation angle φ is a real number. The particle states q of lowest potential energy are given by the equation

$$U(|q|) = 0, \qquad q \in \mathbb{C}.$$

Hence $|q| = 1$, i.e., the set of ground states coincides with the unit circle. This set is called the vacuum by physicists. Observe that the vacuum is invariant under the rotation group $U(1)$.

(ii) Spontaneous symmetry breaking. Fix some ground state, say, $q_0 = 1$, and set

$$q = q_0 + h$$

where h is a small real quantity. In other words, we consider a perturbation of the vacuum state q_0. Since $U'(x) = 2x(x^2 - 1)$, the original equation of motion (2.67) passes over to

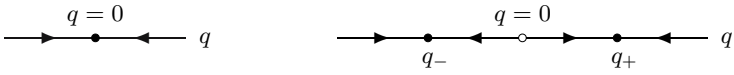
$$m\ddot{h} = -2(1 + h)(2h + h^2).$$

Setting $\kappa_0 := 4$ and neglecting terms of order $O(h^2)$ as $h \to 0$, we get the modified equation of motion

$$\boxed{m\ddot{h}(t) + \kappa_0 h(t) = 0.} \tag{2.68}$$

This describes the motion of a harmonic oscillator with the angular frequency $\omega_0 := \sqrt{\kappa_0/m}$. Physicists say that problem (2.68) is obtained from the original problem (2.67) by spontaneous symmetry breaking.

The Higgs particle. Note the crucial fact that the Standard Model of particle physics represents a gauge field theory. Such theories are generalizations of the Maxwell equations; they were considered first by Yang and Mills in 1954. Originally, the messenger particles of each gauge theory are massless. However, if one introduces an additional massive particle called Higgs particle, then it is possible to equip the massless messenger particle with a mass. The procedure (ii) above represents an oversimplified version of the so-called Higgs mechanism which will be studied in Volume III. This way, in



(a) $\mu < 0$ (b) $\mu > 0$

Fig. 2.23. Bifurcation of a flow

the Standard Model of particle physics, the messenger bosons $W^{\pm}, Z^0$ obtain a mass of approximately 90 proton masses. These three particles were discovered experimentally at CERN in 1983 by a large group of experimentalists. Rubbia and van der Meer were awarded the Nobel prize in physics in 1984 for their decisive contributions to this challenging project.

2.8.5 Bifurcation and Phase Transitions

> Bifurcation is caused by a loss of stability; it is accompanied by an increase of complexity, and hence a decrease of entropy.
>
> Folklore

Bifurcation theory studies the change of the qualitative behavior of systems under an external influence. For example, this is related to pattern formation. As a typical example, consider a viscous fluid between two parallel plates heated from below. Let T_u and T_l denote the constant temperature of the upper plate and the variable temperature of the lower plate, respectively. If the temperature difference $T_l - T_u$ reaches a critical value, then bifurcation occurs. We observe the formation of hexagonal cells called Bénard cells first observed by Bénard in 1901. As a simple example, consider the differential equation

$$\dot{q}(t) = \mu q(t) - q(t)^3, \qquad t \in \mathbb{R}$$

for the motion of fluid particles on the real line. The real parameter μ describes an external influence. The behavior of the trajectories is pictured in Fig. 2.23.

- For $\mu < 0$, the equation $\mu q - q^3 = 0$ has the only solution $q = 0$. This corresponds to an attractor at the point $q = 0$.
- For $\mu > 0$, the equation $\mu q - q^3 = 0$ has three solutions $q = 0,\ q_{\pm} := \pm\mu$. This corresponds to an repeller at the point $q = 0$ and two attractors at the points q_- and q_+.

The bifurcation of this fluid appears at the critical parameter value $\mu = 0$. There exists a rich arsenal of bifurcations for dynamical systems describing, for example, ecological catastrophes in nature (e.g., the extinction of animals). Many results on bifurcation theory along with substantial applications in the sciences can be found in Zeidler (1986), Vols. I–IV.

2.9 The Structure of Interactions in Nature

Force equals curvature in modern physics.
Folklore

2.9.1 The Electromagnetic Field as Generalized Curvature

Concerning the Schrödinger–Maxwell equation, it is a crucial observation that the electromagnetic field $\mathbf{E}, \mathbf{B}$ can be obtained by the commutation relation for covariant derivatives of the wave function ψ. Explicitly, with respect to a right-handed Cartesian coordinate system,[58] we have

$$\boxed{\nabla_t \nabla - \nabla \nabla_t = \mathrm{i}Q\mathbf{E}} \tag{2.69}$$

and

$$\boxed{\nabla_1 \nabla_2 - \nabla_2 \nabla_1 = -\mathrm{i}QB_3.} \tag{2.70}$$

Using cyclic permutation, we also get

$$\nabla_2 \nabla_3 - \nabla_3 \nabla_2 = -\mathrm{i}QB_1, \qquad \nabla_3 \nabla_1 - \nabla_1 \nabla_3 = -\mathrm{i}QB_2.$$

In modern differential geometry, curvature is defined by the commutation relations for covariant derivatives. According to (2.69) and (2.70), the electromagnetic field $\mathbf{E}, \mathbf{B}$ can be regarded as the curvature to the covariant derivatives generated by the Schrödinger–Maxwell equation. In order to prove (2.69) and (2.70), observe that

$$\nabla_t = \frac{\partial}{\partial t} + \mathrm{i}QU, \qquad \nabla = \mathbf{grad} - \mathrm{i}Q\mathbf{A},$$

by (2.62) on page 175. Hence

$$\nabla_t \nabla = \frac{\partial}{\partial t}\,\mathbf{grad}\, - \mathrm{i}Q\mathbf{A}_t$$

and

$$\nabla \nabla_t = \mathbf{grad}\,\frac{\partial}{\partial t} + \mathrm{i}Q\,\mathbf{grad}\,U.$$

Since **grad** commutes with $\frac{\partial}{\partial t}$, we get

$$\nabla_t \nabla - \nabla \nabla_t = -\mathrm{i}Q\mathbf{A}_t - \mathrm{i}Q\,\mathbf{grad}\,U = \mathrm{i}Q\mathbf{E}.$$

This is (2.69). Similarly, it follows from

$$\nabla_k = \partial_k - \mathrm{i}QA_k, \qquad k = 1, 2, 3$$

[58] Recall that $\mathbf{x} = x_1\mathbf{i} + x_2\mathbf{j} + x_3\mathbf{k}$, and $\partial_k := \partial/\partial x_k$. Similarly, the components of $\mathbf{B}$ are denoted by B_1, B_2, B_3, and so on.

that

$$\nabla_j \nabla_k = \partial_j \partial_k - \mathrm{i}Q\partial_j A_k.$$

Since $\partial_j \partial_k = \partial_k \partial_j$,

$$\nabla_1 \nabla_2 - \nabla_2 \nabla_1 = -\mathrm{i}Q(\partial_1 A_2 - \partial_2 A_1) = -\mathrm{i}QB_3.$$

This proves (2.70).

2.9.2 Physics and Modern Differential Geometry

In modern differential geometry, one starts with the notion of parallel transport, which corresponds to the transport of information in physics. Parallel transport allows us the construction of covariant derivatives.[59] Finally, commutation relations between covariant derivatives lead us to the crucial notion of curvature. This will be studied in Volume III by using the modern language of fiber bundles, which fits best the idea of parallel transport of mathematical objects. This approach called gauge field theory applies to

- the curvature of curves,
- the classical Gaussian curvature of 2-dimensional surfaces,
- the Riemann curvature of n-dimensional Riemannian manifolds,
- Einstein's theory of general relativity and the Standard Model of modern cosmology,
- the Cartan–Ehresmann curvature of fiber bundles,
- the Maxwell theory of electromagnetism with respect to the gauge group $U(1)$,
- the Yang–Mills gauge field theory with respect to the gauge group $SU(n)$ where $n = 2, 3, \ldots$,
- the Standard Model of elementary particle physics with respect to the gauge group $U(1) \times SU(2) \times SU(3)$,
- supergravity, and
- string theory.

Historical remarks. In the 20th century, physicists discovered step by step that the fundamental interactions in nature can be described mathematically by so-called gauge field theories. At this point let us only mention the following. For example, in nature we observe electrons as basic particles. Mathematically, electrons are governed by the 1928 Dirac equation which combines Einstein's 1905 theory of special relativity with Schrödinger's 1926 quantum mechanics. Dirac noticed immediately that his equation predicts the existence of an antiparticle to the electron which has the positive electric charge e. In 1932 Anderson discovered the positron experimentally in cosmic rays. Now to the point of gauge field theory.

[59] Parallel transport and covariant derivatives are closely related to the notion of 'connection' which is basic for modern differential geometry.

If we postulate that the Dirac equation is invariant under local symmetries (i.e., suitable phase transformations), then we have to introduce mathematically an additional field.

It turns out that this additional field corresponds to the electromagnetic field. According to Einstein, the electromagnetic field consists of light quanta (photons). Roughly speaking, we can say that

The existence of the electron implies the existence of its antiparticle and of the photon which mediates the interaction between electrons and positrons.

The same remains true for the other interactions described by the Standard Model of particle physics. The existence of the 12 basic particles (quarks and leptons) of the Standard Model implies the existence of their antiparticles and of the 12 interacting particles (8 gluons, the photon, and the three vector bosons W^+, W^-, Z^0.) The number of the interacting particles is closely related to the fact that the gauge group $U(1)\times SU(2)\times SU(3)$ of the Standard Model of particle physics has $1 + 3 + 8 = 12$ dimensions.

In his 1915 theory of general relativity, Einstein described Newton's gravitational force by the curvature of the 4-dimensional pseudo-Riemannian space-time manifold. Élie Cartan discovered in the 1920's that one can assign the notion of curvature to fairly general mathematical objects. This generalizes Gauss' famous theorema egregium. In the 1950s, Ehresmann formulated the final abstract mathematical theory of the curvature of fiber bundles. Yang and Mills discovered in 1954 that it is possible to generalize Maxwell's theory of electromagnetism to more general noncommutative symmetry groups. Nowadays we know that the curvature of fiber bundles is behind

- Einstein's theory of general relativity,
- Maxwell's theory of electromagnetism,
- quantum electrodynamics,
- the Standard Model of particle physics as a generalization of quantum electrodynamics,
- string theory,
- supergravity theory, and so on.

In the terminology of physicists, all of these theories are gauge theories. Mnemonically,

$$\boxed{\text{force} = \text{curvature.}}$$

This is the most important principle of modern physics. Since ancient times, physicists have made strong efforts to understand the forces in nature. Mathematicians studied geometric objects and wanted to understand their curvature. It turns out that physicists and mathematicians studied in fact the same problem. This beautiful interaction between mathematics and physics will be thoroughly studied in Volume II on quantum electrodynamics and in Volume III on gauge theory.

3. The Challenge of Different Scales in Nature

> Between quantum length scales (atomic diameters of about 10^{-10}m) and the earth's diameter (10^{6}m) there are about 16 length scales. Most of technology and much of science occurs in this range. Between the Planck length (10^{-35}m) and the diameter of the visible universe there are 70 length scales; 70, 16, or even 2 is a very large number. Most theories become intractable when they require coupling between even two adjacent length scales. Computational resources are generally not sufficient to resolve multiple length scales in three-dimensional problems and even in many two-dimensional problems. The problem is not merely one of presently available computational resources, which are growing at a rapid rate. To obtain an extra factor of 10 in computational resolution requires in the most favorable case a factor 10^{4} in computational resources for time-dependent three-dimensional problems. When multiple length scales are in question, the under-resolution of computations performed with today's algorithms will be with us for some time to come, and the essential role which must be assigned to theory, and to the design of algorithms of a new nature, becomes evident. It is for this reason that nonlinear and stochastic phenomena, often described by the theory of coherent and chaotic structures, coupling adjacent and multiple length scales, is a vital topic.[1]
>
> James Glimm, 1991

3.1 The Trouble with Scale Changes

In physics we frequently have to perform singular limits when passing from one essential scale to another one. There are the following typical examples:

(i) the singular limit from Einstein's theory of general relativity to Newtonian mechanics,
(ii) the singular limit from the mesoscopic Boltzmann equation to the macroscopic Navier–Stokes equations in continuum mechanics,
(iii) phase transitions as a singular limit related to the Ginzburg–Landau equation,
(iv) thin films as singular limits,
(v) thin plates as singular limits of 3-dimensional elasticity theory,

[1] J. Glimm, Nonlinear and stochastic phenomena: The grand challenge for partial differential equations, SIAM Review **33** (1991), 626–643 (reprinted with permission).

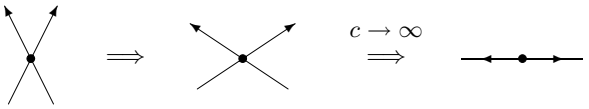


Fig. 3.1. Degeneration of the light cone

(vi) the emergence of microstructures in nature and high-technology.

In what follows, let us discuss some basic ideas.

Ad (i). Letting the velocity of light go to infinity, $c \to \infty$, the wave equation

$$\frac{U_{tt}}{c^2} - U_{xx} - U_{yy} - U_{zz} = 0$$

passes over to the Laplace equation

$$-U_{xx} - U_{yy} - U_{zz} = 0.$$

Roughly speaking, this is the limit from Einstein's theory of relativity to Newtonian mechanics. The causal structure of the theory of relativity is described by the light cone

$$\frac{x^2 + y^2 + z^2}{c^2} - t^2 = 0.$$

This cone degenerates into the plane $t = 0$ as $c \to \infty$. This is represented schematically in Fig. 3.1. In terms of mathematics, this limit was studied first by Friedrichs (1927).

Note that this limit is highly nontrivial, since the type of the partial differential equation changes completely from the hyperbolic wave equation to the elliptic Laplace equation. Important recent results were obtained by Rendall (1994).

Ad (ii). In 1913 Hilbert investigated the Boltzmann equation. He posed the problem of justifying the limit from the Boltzmann equation for a viscous fluid to the Navier–Stokes equations in continuum mechanics. This corresponds to a passage

- from the mesoscopic length scale $l = 10^{-5}$cm (the free path length of a molecule)
- to the macroscopic length scale $l = 1$cm.

This famous, highly nontrivial problem was finally solved by Golse and Saint-Raymond (2001).

Ad (iii). Phase transitions in superconductivity and superfluidity can be modelled mathematically by the stationary Ginzburg–Landau equation. To this end, one studies the singular limit $\varepsilon \to 0$ for the following variational problem

$$\int_\Omega |\,\mathbf{grad}\,\psi_\varepsilon|^2 d^3x + \frac{1}{\varepsilon^2}\int_\Omega (1-|\psi_\varepsilon|^2)^2 d^3x = \min!,$$
$$\psi_\varepsilon = g \qquad \text{on} \quad \partial\Omega.$$

where the boundary function g is given. This was done by Bethuel, Brézis, and Hélein (1994) in a sophisticated manner. We refer to this monograph. The basic ideas can be found in Zeidler (1995), Vol. 2, p. 152.

Ad (iv). Thin films play an important role in modern high technology. From the mathematical point of view, the main task is to start with the equations of continuum mechanics and to rigorously study the limit $d \to 0$, that is, the thickness goes to zero (Fig. 3.2). Recent sophisticated results can be found in DeSimone, Kohn, Müller, and Otto (2002).

Ad (v). To study the limit $d \to 0$ for thin plates of thickness d, it is possible to apply de Giorgi's powerful technique of Γ-convergence for variational integrals. For this, we refer to Friesecke, James, and Müller (2002). A brief introduction to Γ-convergence can be found in Jost and Li-Jost (1998).

Ad (vi). The mathematical theory of microstructures tries to understand the mathematical mechanisms which are responsible for the formation of microstructures. In this connection, singular scaling limits play a fundamental role. As an introduction to this quite interesting, modern field of mathematical analysis, we recommend the lecture notes by Müller (1998) and Dolzmann (2003).

Summarizing, the transition from one essential scale to another one is a highly nontrivial task for mathematics. The dream of physicists is to create a final theory for the four fundamental forces in the universe. However, the experience of mathematicians in the past tells us that it will be extremely difficult to pass from the Planck length $l = 10^{-33}$cm to the macroscopic scale $l = 1$ cm, in a mathematically rigorous way.

3.2 Wilson's Renormalization Group Theory in Physics

Each Lagrangian density represents a physical theory. The idea of flows for ordinary differential equations can be generalized to flows (or semi-flows) in the space of physical theories. It turns out that appropriate fixed points of the semi-flow correspond to phase transitions. The physical idea behind this fixed point is the observation that

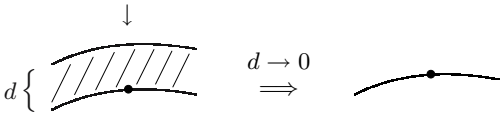


Fig. 3.2. Thin film

Physical systems are invariant under rescaling at phase transitions.

Intuitively, this is based on the following fact: Since the correlation length becomes infinite at a phase transition (large fluctuations), the system loses its typical length scale. We will study this in great detail in the later volumes. In the collection of seminal papers that appeared in the journal *Physical Review*, Joel Lebowitz writes the following:[2]

> The Wilson renormalization group in statistical physics had antecedents in quantum field theory by Gell-Mann and Low.[3] Following Wilson's work, the renormalization group method has spread and had enormous influence on almost all fields of science. It provides a method for quantitative analysis of the "essential" features of a large class of nonlinear phenomena exhibiting *self-similar structures*. This includes not only scale invariant critical systems (phase transitions) where fluctuations are "infinite" on the microscopic spatial and temporal scale, but also fractals, dynamical systems exhibiting Feigenbaum period doubling, Kolmogorov–Arnold–Moser theory (KAM theory) on critical resonances in celestial mechanics, singular behavior in nonlinear partial differential equations, and "chaos". Even where not directly applicable, the renormalization group often provides a paradigm for the analysis of complex phenomenas... A lot of mathematical work remains to be done to make it into a well-defined theory of phase transitions.

For his theory of critical phenomena in terms of the renormalization group, Kenneth Wilson (born 1935) was awarded the Nobel prize in physics in 1982. Wilson's ideas changed the paradigm of theoretical physics.

(i) In the past, physicists studied specific theories like the motion of planets around the sun or the motion of the electron around the nucleus of the hydrogen atom.
(ii) Nowadays physicists want to study the behavior of physical phenomena at quite different scales. The idea of the renormalization group helps to bridge the different scales.

As a typical example for (ii), consider the cooling of the universe after the Big Bang. To understand this, we have to study the behavior of elementary particles at completely different energy scales. Let us mention two fundamental phase transitions in the early universe:

- First phase transition: Using the method of running coupling constants in renormalized quantum field theory, physicists discovered that, 10^{-35} seconds after the Big Bang at a temperature of 10^{28}K, strong interaction

[2] J. Lebowitz, Statistical mechanics in the 20th century. In: H. Stroke (Ed.), The Physical Review: The First Hundred Years, American Institute of Physics, New York 1995, pp. 363–389 (reprinted with permission).

[3] M. Gell-Mann and F. Low, Quantum electrodynamics at small distances, Rev. Phys. **95**(5) (1954), 1300–1317.
K. Wilson and J. Kogut, The renormalization group and the ε-expansion, Physics Reports **12C** (1974), 75–199.

and electroweak interaction decoupled. This phase transition corresponds to a particle energy of 10^{15} GeV.
- Second phase transition: 10^{-12} seconds after the Big Bang at a temperature of 10^{16}K, weak and electromagnetic interaction decoupled. This corresponds to a particle energy of 10^3 GeV.

Note that the most powerful particle accelerator in the world will begin to work at CERN (Geneva, Switzerland) in the year 2008. There, the particle energy will be about 10^3 GeV. This means that we will be able in near future to reach the energy scale of the second phase transition in a huge laboratory on earth.

If we want to create a unified theory for all fundamental interactions in the universe including gravitation, then we have to bridge 60 scales,

- from the radius of the visible universe, $r = 10^{28}$cm,
- down to the Planck length, $l = 10^{-33}$cm (see Table 3.1).

3.2.1 A New Paradigm in Physics

To emphasize the role of Wilson's new paradigm in physics, let us quote from the introduction to a series of lectures on the renormalization group given by David Gross:[4]

> Physics is scale dependent. For example, consider a fluid. At each scale of distances, we need a different theory to describe its behavior:
> - at ~ 1 cm – classical continuum mechanics (Navier–Stokes equations),
> - at $\sim 10^{-5}$cm – theory of granular structure,
> - at 10^{-8} cm – theory of atom (nucleus plus electronic cloud),
> - at 10^{-13}cm – nuclear physics (nucleons),
> - at $\sim 10^{-13}$cm – 10^{-18}cm – quantum chromodynamics (quarks)
> - at $\sim 10^{-33}$cm – string theory.
>
> At each scale, we have different degrees of freedom and different dynamics. Physics at a larger scale (largely) decouples from the physics at a smaller scale. For example, to describe the behavior of a fluid at the scale $\sim$ 1cm, we do not need to know about the granular structure, nor about atoms or nucleons. The only things we need to know are the viscosity and the density of the fluid. Of course, these values can be computed from the physics of a smaller scale, but if we found them out in some way (for example, measurement), we can do without smaller scale theories at all. Similarly, if we want to describe atoms, we do not need to know anything about the nucleus except its mass and electric charge.
>
> Thus, a theory at a larger scale remembers only finitely many parameters from the theories at smaller scales, and throws the rest of the details away.

[4] These lectures were part of the Special Year 1996-97 at the Institute for Advanced Study in Princeton, New Jersey, devoted to the physics and mathematics of quantum field theories. See P. Deligne, E. Witten et al., Lectures on Quantum Field Theory, Vol. 1, pp. 551–593, Amer. Math. Soc., Providence, Rhode Island 2000 (reprinted with permission).

Table 3.1. From macrocosmos to microcosmos

length scale in centimeter	physical phenomenon
$\mathbf{10^{28}}$cm	radius of the visible universe
10^{23}	radius of the Milky Way
10^{18}	light year
10^{15}	radius of the solar system
10^{11}	radius of the sun
10^{9}	radius of the earth
10^{5}	wave length of radio waves
$10^{3} - 10^{-1}$	wave length of sound
3.75 cm	wave length of the cosmic background radiation
10^{-3}	wave length of heat radiation
$4 \cdot 10^{-7} - 8 \cdot 10^{-7}$	wave length of visible light
10^{-7}	radius of molecules
10^{-8}	radius of atoms
10^{-8}	wave length of X rays
10^{-13}	radius of protons
10^{-13}	wave length of cosmic rays
10^{-14}	Compton wave length of the electron
10^{-17}	Compton wave length of the proton
$< 10^{-18}$	radius of quarks and electrons
$\mathbf{10^{-33}}$ cm	Planck scale, strings, quantum gravitation

More precisely, when we pass from a smaller scale to a larger scale, we average over irrelevant degrees of freedom. Mathematically, this means that they become integration variables and thus disappear in the answer.

This decoupling is the reason why we are able to do physics. If there was no decoupling, it would be necessary for Newton to know string theory to describe the motion of a viscous fluid...

The general aim of the renormalization group method is to explain how this decoupling takes place and why exactly information is transmitted from scale to scale through finitely many parameters. In quantum theory, decoupling of scales is not at all obvious. Indeed, because of the uncertainty principle, we have to work at all scales at once. The renormalization group describes why decoupling survives in quantum theory.

Table 3.2. Typical energy scales

energy in Giga electron volt	physical phenomenon
10^{-8}	binding energy of the electron in the hydrogen atom; energy of chemical processes
10^{-3}	rest energy of the electron
1 GeV	rest energy of the proton
10^{2}	rest energy of the vector bosons W^+, W^-, Z^0
10^{3}	weak and electromagnetic interaction decouple at this particle energy
10^{3}	particle energy at the new CERN accelerator in 2008
10^{14}	energy consumption of a human being per day
10^{15}	strong and electroweak energy decouple at this particle energy
10^{19}	Planck energy
10^{27}	rest energy of a stone (mass of 1kg)
10^{29}	energy production of a hydrogen bomb
10^{36}	energy production of the sun per day
10^{52}	rest energy of the earth
10^{57}	rest energy of the sun
10^{68}	rest energy of the Milky Way
10^{79}	rest energy of the visible universe

3.2.2 Screening of the Coulomb Field and the Renormalization Group of Lie Type

> The infinitesimal transformations of a Lie group know all about the local structure of the Lie group itself.
>
> Folklore

My text for today is a paper by Francis Low and Murray Gell-Mann.[5] It is "Quantum Electrodynamics at Small Distances" published in the *Physical Review* in 1954. This paper is one of the most important ever published in quantum field theory... Gell-Mann and Low started by considering an ancient problem, the Coulomb force between two charges and how this force behaves at very short distances... The important thing about the Gell-Mann–Low paper was the fact that they realized that quantum field

[5] S. Weinberg, Why the renormalization group is a good thing. In: A. Guth, K. Huang, and A. Jaffe (Eds.) (1983), Asymptotic Realms of Physics, Essays in Honor of Francis Low, MIT Press, Cambridge, Massachusetts, pp. 1–19 (reprinted with permission).

> theory has a scale invariance, that the scale invariance is broken by particle masses but these are negligible at very high energy or very short distances if you renormalize in an appropriate way, and that then the only things that's breaking scale invariance is the renormalization procedure, and that one can take that into account by keeping track of the running coupling constant $a(r)$.
>
> Steven Weinberg, 1983

It is typical for the coupling constants of the Standard Model of particle physics that they depend on the physical situation, for example, the energy-momentum transfer of a scattered particle. We speak of running coupling constants. This has been established experimentally. In 1994, the Particle Data Group reported that, for scattering processes of particles at an energy of 91 GeV,[6] one does not have to use the classical electromagnetic fine structure constant $\alpha = 1/137.04$, but rather the effective fine structure constant

$$a = \frac{1}{128.87 \pm 0.12}.$$

To begin with, we want to study the running coupling constant (i.e., the running fine structure constant) in the case of quantum electrodynamics. We will work in the energetic system, that is, we set $c = \hbar = \varepsilon_0 := 1$.

The renormalized electric charge. The classical electrostatic Coulomb field of an electron with electric charge $-e$ and mass m_e is given by

$$\mathbf{E} = -\,\mathbf{grad}\, U$$

along with the Coulomb potential

$$U(\mathbf{x}) := -\frac{e}{4\pi r} \tag{3.1}$$

where $r := ||\mathbf{x}||$. Recall that the electric charge $-e$ is related to the electromagnetic fine structure constant α by

$$\alpha := \frac{e^2}{4\pi} = \frac{1}{137.04}.$$

We will also use the reduced Compton wave length [7] $\lambda\!\!\!^{-}_e = 1/m_e$. The force $\mathbf{F}$ between two electrons is equal to

$$\mathbf{F} = -\,\mathbf{grad}\, V$$

with $V = -eU$. Hence

$$V(\mathbf{x}) = \frac{\alpha}{r}.$$

[6] This is the rest energy of the Z^0 boson.
[7] In the SI system, $\lambda\!\!\!^{-}_e = 3.86 \cdot 10^{-11}$m.

In 1935 Sperber and Uehling showed that quantum effects change the electric field $\mathbf{E} = -\,\mathbf{grad}\, U$ of a charged particle. For an electron, we have to replace (3.1) by the expression[8]

$$U(\mathbf{x}) = -\frac{e}{4\pi r}\left(1 + \frac{2\alpha}{3\pi}\ln\left(\frac{\lambda\!\!\!^{-}_e}{r}\right) - \frac{2\alpha}{3\pi}\Big(C + \frac{5}{6}\Big)\right) + O(\alpha^2),\ \ \alpha \to 0 \quad (3.2)$$

if $r/\lambda\!\!\!^{-}_e \ll 1$, and

$$U(\mathbf{x}) = -\frac{e}{4\pi r}\left(1 + \frac{\alpha}{4\sqrt{\pi}}\left(\frac{\lambda\!\!\!^{-}_e}{r}\right)^{3/2} \mathrm{e}^{-2r/\lambda\!\!\!^{-}_e}\right) + O(\alpha^2),\ \ \alpha \to 0$$

if $r/\lambda\!\!\!^{-}_e \gg 1$. Therefore, quantum effects influence only the short-distance behavior of the electric field $\mathbf{E}$. These effects are called vacuum polarization by physicists. Here, $C = 0.577\ldots$ is the famous Euler constant. For the force between two electrons, $\mathbf{F} = -e\mathbf{E} = -\,\mathbf{grad}\, V$, we get

$$V(\mathbf{x}) = \frac{\alpha}{r}\left(1 + \frac{\alpha}{4\sqrt{\pi}}\left(\frac{\lambda\!\!\!^{-}_e}{r}\right)^{3/2} \mathrm{e}^{-2r/\lambda\!\!\!^{-}_e}\right) + O(\alpha^2),\ \ \alpha \to 0$$

if $r/\lambda\!\!\!^{-}_e \gg 1$. In a classical experiment, we measure the force between two electrons at a large distance r. It turns out that this yields the fine structure constant $\alpha = 1/137.04$, and hence the electron charge $e = \sqrt{4\pi\alpha}$. For short distances, $r/\lambda\!\!\!^{-}_e \ll 1$, we obtain

$$V(\mathbf{x}) = \frac{a(r)}{r}$$

with[9]

$$\boxed{a(r) := \alpha + \frac{2\alpha^2}{3\pi}\left(\ln\left(\frac{\lambda\!\!\!^{-}_e}{r}\right) - C - \frac{5}{6}\right) + O(\alpha^3),\ \ \alpha \to 0.} \quad (3.3)$$

We call $a(r)$ the running fine structure constant, since it depends on the distance, r. The effective electron charge $-e(r)$ is defined by

$$\boxed{e(r)^2 := 4\pi a(r).}$$

In quantum electrodynamics, the coupling constant between the Dirac field of the electron and the electromagnetic field is given by $\kappa = e$. Therefore, we call

[8] In terms of modern quantum electrodynamics, the potential U is obtained by using renormalized second-order perturbation theory. This will be studied in Volume II. See also Landau and Lifshitz (1982), Vol. IV, Sect. 111, and Greiner, Reinhardt (1996a), p. 252.

[9] Note that $\ln(\lambda\!\!\!^{-}_e/r) > 0$ if $0 < r < \lambda\!\!\!^{-}_e$.

$$\kappa(r) = e(r) = \sqrt{4\pi a(r)}$$

the running coupling constant of quantum electrodynamics. In contrast to the classical Coulomb law (3.1), the modified Coulomb law (3.2) contains the distinguished length scale $\lambda\!\!\!^{-}_e$. This is an important consequence of the renormalization procedure.

The basic ideas of the renormalization group in quantum electrodynamics. In 1954 Gell-Mann and Low studied the Coulomb potential for short distances in all orders of perturbation theory. They found out that the approach can be substantially simplified by using general symmetry arguments which correspond to the renormalization group. Let us discuss the basic ideas by using some heuristic arguments. Choose the distance $R > 0$ as a fixed, sufficiently small parameter. Suppose that the coupling constant $a(R)$ is known at the given distance R. It is our goal to determine the coupling constant $a(r)$ for distances r different from the initial distance R. Below we will obtain the following approximative formula

$$\boxed{a(r) = \frac{a(R)}{1 - \frac{2a(R)}{3\pi} \ln\left(\frac{R}{r}\right)}} \tag{3.4}$$

for all distances $r > r_{\text{crit}}(R)$. Naturally enough, the critical distance $r_{\text{crit}}(R)$ is given by the zero of the denominator, that is,

$$1 - \frac{2a(R)}{3\pi} \ln\left(\frac{R}{r_{\text{crit}}(R)}\right) = 0.$$

In particular, it follows from (3.3) that $a(R) = \alpha$, up to terms of order $O(\alpha^2)$. Hence

$$r_{\text{crit}}(R) = 10^{-293}\text{m}.$$

This critical distance is extremely small. Finally, we obtain the crucial limit

$$\boxed{\lim_{r \to r_{\text{crit}} + 0} a(r) = +\infty.}$$

Summarizing, we get the following in quantum electrodynamics.

> *There exists a tiny critical distance r_{crit} such that the effective electric charge of two electrons at distance $r > r_{\text{crit}}$ becomes infinite if the distance r goes to r_{crit}.*

We will not get any information about the behavior for distances $r < r_{\text{crit}}$. The situation changes dramatically if we replace electrons by quarks. The strong force between two quarks vanishes if the distance between two quarks goes to zero. This will be discussed in Sect. 3.2.3 on page 201.

In order to get the crucial relation (3.4), we will use the elegant technique of the renormalization group. The essential ingredients of this technique read as follows:

- dimensional analysis,
- the algebraic renormalization-group equation (self-similarity),
- the equivalent renormalization-group differential equation (also called the infinitesimal renormalization-group equation), and
- the solution of the renormalization-group differential equation.

We will also show that the renormalization group represents a local 1-dimensional group (also called local flow), in the sense of the classical theory due to Sophus Lie (1842–1899).

Dimensional analysis and the renormalization-group equation. The quantities $a(r), r/R$, and $r/\lambda\!\!\!^{-}_e = rm_e$ are dimensionless in the energetic system. By dimensional analysis, we start with the ansatz

$$a(r) = G\left(a(R), \frac{r}{R}, rm_e\right)$$

for the running coupling constant $a(r)$. We now make the crucial assumption that, for small distances r, the function G does not depend on the electron mass m_e. Thus, we get the key equation

$$\boxed{a(r) = G\left(a(R), \frac{r}{R}\right).} \tag{3.5}$$

We call this the renormalization-group equation. This equation tells us that the running coupling constant $a(r)$ at the distance r depends on both the running coupling constant $a(R)$ at the distance R and the ratio r/R. Generally, in physics, such a situation is characterized by the key word *self-similarity*. The classical Lie theory can be used in order to analyze the general structure of the solutions of equation (3.5). The idea is to pass over from (3.5) to an equivalent differential equation which is called the renormalization-group differential equation (or the infinitesimal renormalization-group equation). As a preparation for this, let us write $G = G(a, x)$, and let us introduce the crucial Gell–Mann–Low beta function

$$\beta(a) := -G_x(a, 1).$$

It follows from (3.5) that $a = G(a, 1)$. By Taylor expansion,

$$\begin{aligned} G(a, x) &= G(a, 1) + G_x(a, 1)(x - 1) + o(x - 1) \\ &= a - \beta(a)(x - 1) + o(x - 1), \qquad x \to 1. \end{aligned}$$

Roughly speaking, the beta function $a \mapsto \beta(a)$ knows all about the local behavior of the function G.

The renormalization-group differential equation. Using the replacement $R \Rightarrow r$ and $r \Rightarrow r + \Delta r$, it follows from (3.5) that

$$a(r + \Delta r) = G\left(a(r), \frac{r + \Delta r}{r}\right), \qquad a(r) = G(a(r), 1).$$

This implies

$$a'(r) = \lim_{\Delta r \to +0} \frac{a(r+\Delta r) - a(r)}{\Delta r} = \frac{G_x(a(r),1)}{r}.$$

This yields the crucial renormalization-group differential equation

$$\boxed{a'(r) = -\frac{\beta(a(r))}{r}.} \tag{3.6}$$

Determination of the beta function. It remains to compute the beta function. To this end, we will use the additional information given by equation (3.3) which follows from renormalized perturbation theory. By (3.3), we get

$$a(r) = \alpha,$$

and

$$a'(r) = -\frac{2\alpha^2}{3\pi r},$$

up to terms of order $O(\alpha^3)$. This motivates the differential equation

$$\boxed{a'(r) = -\frac{2a(r)^2}{3\pi r}.} \tag{3.7}$$

This is the desired renormalization-group differential equation (3.6) with the beta function $\beta(a) := 2a^2/3\pi$. Summarizing,

- the general structure of the renormalization-group differential equation (3.6) follows from dimensional analysis (Lie group theory);
- the specific form of this equation depends on the beta function which follows from renormalized perturbation theory.

The same method can be applied to the Standard Model of particle physics.

Rescaling. To simplify the approach, it is convenient to rescale the distance r by introducing the new variable

$$t := \ln\left(\frac{r}{R}\right), \qquad r > 0. \tag{3.8}$$

Setting $b(t) := a(r)$, the differential equation (3.7) passes over to the rescaled renormalization-group differential equation:

$$\boxed{b'(t) = -\frac{2b(t)^2}{3\pi}} \tag{3.9}$$

along with the initial condition $b(0) = a(R)$. This equation has the unique solution

$$b(t) = \frac{a(R)}{1 + \frac{2}{3\pi}\, a(R)t}$$

for all $t > -3\pi/2a(R)$, which proves the claim (3.4).

Summarizing, the renormalization group approach adds a global aspect to local renormalized perturbation theory (3.3).

This is one of the crucial advantages of renormalization group theory. It remains to show that there is indeed a symmetry and hence a group behind (3.5). To this end, we will use the classical Lie theory for differential equations.

The relation to Lie's theory of additive one-parameter groups. The renormalization-group differential equation (3.9) can be written as

$$b'(t) = -\beta(b(t)),$$

along with the initial condition $b(0) = a(R)$. Since the beta function does not explicitly depend on time t, this differential equation describes a dynamical system. Let $b = b(t)$ be a solution. In terms of physics, we regard this as the trajectory of a fluid particle on the real line. Define

$$\boxed{\mathcal{F}_t b(0) := b(t).}$$

The flow operator $\mathcal{F}_t$ sends the initial position $b(0)$ of the fluid particle at time $t = 0$ to the position $b(t)$ of the particle at time t. Obviously, $\mathcal{F}_0 b(0) = b(0)$, and

$$\mathcal{F}_s(\mathcal{F}_t b(0)) = \mathcal{F}_s b(t) = b(t+s) = \mathcal{F}_{t+s} b(0).$$

This implies that

(i) $\mathcal{F}_0 = \mathrm{id}$, and
(ii) $\mathcal{F}_t \mathcal{F}_s = \mathcal{F}_{t+s}$ for all real numbers t, s in some open neighborhood of $t_0 = 0$.

Following Sophus Lie, we call the family $\{\mathcal{F}_t\}$ of operators $\mathcal{F}_t$ a local additive one-parameter group (or a local flow).

The multiplicative one-parameter group. Let us now reformulate this in terms of the original parameter r. Consider the renormalization-group differential equation

$$\boxed{a'(r) = -\frac{\beta(a(r))}{r}.} \tag{3.10}$$

Let $a = a(r)$ be a solution. Define

$$\boxed{\mathcal{G}_r a(1) := a(r).}$$

With respect to the variable $t = \ln r$, the function $a(r) := \mathcal{F}_{\ln r}\, a_0$ is a solution of (3.10) with $a(1) = a_0$. Hence

$$\mathcal{G}_r a(1) = \mathcal{F}_{\ln r}\, a(1).$$

This implies that

(i) $\mathcal{G}_1 = \mathrm{id}$, and
(ii) $\mathcal{G}_r \mathcal{G}_R = \mathcal{G}_{rR}$ for all real numbers r and R in some open neighborhood of $r_0 = 1$.

Define $G(a, r) := \mathcal{G}_r a$. We want to show that

$$\boxed{G\left(a(R), \frac{r}{R}\right) = a(r).}$$

This is the renormalization-group equation. In fact, $\mathcal{G}_R \mathcal{G}_{1/R} = \mathrm{id}$ implies $\mathcal{G}_{1/R} = \mathcal{G}_R^{-1}$. Therefore,

$$\mathcal{G}_{r/R}\, a(R) = \mathcal{G}_r \mathcal{G}_{1/R}\, a(R) = \mathcal{G}_r \mathcal{G}_R^{-1} a(R) = \mathcal{G}_r\, a(1) = a(r).$$

The family $\{\mathcal{G}_r\}$ is called a local multiplicative one-parameter group.

Historical remarks. The idea of the renormalization group in quantum electrodynamics can be traced back to a paper by Stückelberg, Petermann (1953), and by Gell-Mann, Low (1954). Based on Lie's theory of infinitesimal transformations and the Dyson transformation in renormalization theory, the mathematical theory of the renormalization group in quantum field theory was developed by Bogoliubov and Shirkov (1956). Nine years later, Kadanoff (1965) applied the idea of renormalization group to studying the Ising model for ferromagnets near the phase transition. As a powerful general tool for modern physics, the theory of renormalization group was established by Wilson around 1970 (see Wilson and Kogut (1974)). For this, he was awarded the Nobel prize in physics in 1982. In particular, Wilson used the idea of renormalization group in order to compute

- critical exponents for the physical laws which govern phase transitions of many-particle systems (e.g., fluids, gases, materials), and
- the high-energy asymptotics of quantum fields via the method of the singularities of operator products on the light cone, and
- the structure functions for the cross sections of scattering processes via the operator-product method.

In the 1980s, Polchinski (1984) developed a new approach to the renormalization of quantum field theories by using a semi-flow which approaches the renormalized Lagrangian in a large number of steps. This revolutionized renormalization theory in both quantum field theory and statistical physics. This can be found in the monograph by M. Salmhofer (1999). For a detailed history of the renormalization group, we refer to L. Brown, Renormalization: From Lorentz to Landau and Beyond, Springer, New York, 1993. See also Rivasseau's remark on the modern strategy in renormalization theory quoted on page 849. This strategy has been strongly influenced by Wilson's work on the renormalization group. For recent progress in the theory of the renormalization group due to Connes, Kreimer, and Marcolli, we refer to Sect. 15.4.6 on page 859.

3.2.3 The Running Coupling Constant and the Asymptotic Freedom of Quarks

> In 1973 Politzer, Gross, and Wilczek realized that the plus sign in the logarithmic term (3.2) which prevented the use of perturbation theory in quantum electrodynamics at short distances, for non-Abelian gauge theories is a minus sign[10]... Politzer, Gross, and Wilczek instantly realized that this explains an experimental fact which had been observed in a famous experiment on deep inelastic electron-proton scattering done by an MIT-SLAC collaboration in 1968. This was that at very high momentum transfer, in other words, at very short distances, the strong interactions seem to turn off and the formulas for the form factors in deep inelastic electron scattering seem to obey a kind of naive scaling,"Bjorken scaling".[11]
>
> Steven Weinberg

We say that some interaction between particles is asymptotically free iff the particles move freely for short distances (or high energies). In terms of the running coupling constant $\kappa(r)$, by asymptotic freedom we understand that the function $r \mapsto \kappa(r)$ is continuous on a reasonably large interval $]0, r_0[$, and

$$\lim_{r \to +0} \kappa(r) = 0.$$

This means that the interaction vanishes for short distances. Typically, quarks are free particles at short distances.

Lack of asymptotic freedom in quantum electrodynamics. By (3.4),

$$a(r) = \frac{a(R)}{1 - \frac{2a(R)}{3\pi} \ln(\frac{R}{r})} \tag{3.11}$$

for all distances $r > r_{\text{crit}}(R)$. The running coupling constant $a = a(r)$ is defined on the interval $]r_{\text{crit}}(R), \infty[$, and

$$\lim_{r \to r_{\text{crit}} + 0} a(r) = +\infty.$$

Note that $r_{\text{crit}} = 10^{-293}$m. This is an extremely small distance compared with the radius of the proton, $r = 10^{-15}$m. At the critical distance r_{crit}, the electromagnetic interaction becomes infinite, i.e., quantum electrodynamics breaks down. Observe that physicists expect that already below the Planck length, $r = 10^{-35}$m, quantum field theory breaks down. For such short distances, physicists believe that the structure of space and time change dramatically, and quantum field theory has to be replaced by quantum gravitation. Nowadays, there exist several proposals for quantum gravitation. However,

[10] See Politzer (1973), Gross and Wilzek (1973). The three physicists were awarded the Nobel prize in physics in 2004.

[11] In: A. Guth, K. Huang, and A. Jaffe (Eds.) (1983), pp. 1–19 (reprinted with permission).

a convincing and generally accepted theory of quantum gravitation is still missing.

A model for asymptotic freedom. For quantum electrodynamics, by (3.3) we have

$$a(r) = \alpha + \sigma \cdot \frac{2\alpha^2}{3\pi}\left(\ln\left(\frac{\lambda\!\!\!^{-}_e}{r}\right) - C - \frac{5}{6}\right) \tag{3.12}$$

with $\sigma = 1$ if $r/\lambda\!\!\!^{-}_e \ll 1$. According to (3.7), the renormalization-group differential equation reads as

$$a'(r) = -\sigma \cdot \frac{2a(r)^2}{3\pi r},$$

with the solution

$$\boxed{a(r) = \frac{a(R)}{1 - \sigma \cdot \frac{2a(R)}{3\pi}\ln(\frac{R}{r})}.}$$

Choose now the converse sign, $\sigma := -1$. Then, the situation changes completely compared with quantum electrodynamics. We now have $a(r) > 0$ for all $r \in]0, R[$, and

$$\lim_{r \to +0} a(r) = 0.$$

This is a model for asymptotic freedom which resembles the situation in strong interaction to be considered below.

Running coupling constants in gauge theories. The method described above for quantum electrodynamics can be generalized to gauge field theories with the gauge groups $U(1)$ and $SU(n)$ where $n = 2, 3, ...$, along with the coupling constant κ. Let us introduce the energy-momentum 4-vector

$$P := (E, \mathbf{p})$$

with the Minkowski square $P^2 = E^2 - \mathbf{p}^2$.Recall that we work in the energetic system with $c = \hbar = \varepsilon_0 = 1$. The coupling constant

$$\kappa(P)$$

depends on the energy-momentum 4-vector P of the physical experiment under consideration. We speak of running coupling constant. Let us also introduce the running fine structure constant of the theory by letting

$$\alpha(P) := \frac{\kappa(P)^2}{4\pi}.$$

In the special case of quantum electrodynamics, we have $\kappa(P) = e(P)$, where $-e(P)$ is the running effective electron charge which depends on P. Computing the renormalized radiative corrections in lowest order, one obtains the following renormalization group differential equation

$$\boxed{\frac{d\alpha(P)}{d\ln P^2} = -\frac{\beta_n \alpha(P)^2}{4\pi}} \tag{3.13}$$

along with the initial condition $\alpha(P_0) = \alpha_0$. The constant β_n depends on the choice of the gauge group and the number of fermions, N_{fermion}. Explicitly,

$$\beta_n := \begin{cases} -\frac{2}{3}N_{\text{fermion}} & : \text{ gauge group } U(1), \quad n = 1 \\ \frac{11}{3}n - \frac{2}{3}N_{\text{fermion}} & : \text{ gauge group } SU(n), \quad n = 2, 3, \ldots \end{cases}$$

The solution of the differential equation (3.13) reads as

$$\boxed{\alpha(P) = \frac{\alpha(P_0)}{1 + \frac{\alpha(P_0)}{4\pi} \cdot \beta_n \ln\left(\frac{P^2}{P_0^2}\right)}.} \tag{3.14}$$

For fixing the energy-momentum scale P_0, the value $\alpha(P_0)$ of the fine structure constant has to be determined by the experiment. As typical examples, let us consider both quantum electrodynamics and quantum chromodynamics.

(i) Quantum electrodynamics (electromagnetic interaction): In this case, we have the gauge group $U(1)$ and one fermion, namely, the electron. Hence $\beta_1 = -\frac{2}{3}$. For the running fine structure constant,

$$\alpha(P) = \frac{\alpha(P_0)}{1 - \frac{\alpha(P_0)}{6\pi} \ln\left(\frac{P^2}{P_0^2}\right)}.$$

Let us pass from energy-momentum to distance r. High energies P^2 correspond to small distances r. Therefore, it is reasonable to replace the dimensionless quantity P^2/P_0^2 by $(r_0/r)^\gamma$ for some $\gamma > 0$. Hence

$$\alpha(r) = \frac{\alpha(r_0)}{1 - \frac{\alpha(r_0)\gamma}{6\pi} \cdot \ln\left(\frac{r_0}{r}\right)}$$

for small distances, r. In first-order approximation, $\alpha(r_0) = \alpha = 1/137.04$. This implies

$$\boxed{\alpha(r) = \frac{\alpha}{1 - \frac{\alpha\gamma}{6\pi} \cdot \ln\left(\frac{r_0}{r}\right)}.}$$

Writing $a(r)$ instead of $\alpha(r)$ and choosing $\gamma = 4$, this is precisely formula (3.11) which has been motivated in the preceding section.

(ii) Quantum chromodynamics (strong interaction): Here, we have six quarks along with the gauge group $SU(3)$. Hence $\beta_{\text{QCD}} = \beta_3 = 11 - 4 = 7$. This implies

$$\boxed{\alpha_{\text{QCD}}(P) = \frac{\alpha_{\text{QCD}}(P_0)}{1 + \frac{7\alpha_{\text{QCD}}(P_0)}{4\pi} \cdot \ln(\frac{P^2}{P_0^2})}.} \tag{3.15}$$

Observe that the renormalization-group differential equation above is only an approximation. If we pass to higher-order perturbation theory, then this differential equation has to be refined which leads to corrections of the basic formula (3.14). Higher-order corrections can be found in the monograph by Böhm, Denner, and Jost (2001), Sect. 2.6.

High-energy asymptotic freedom in strong interaction. For strong interaction, equation (3.15) tells us that

$$\lim_{P^2 \to +\infty} \alpha_{\mathrm{QCD}}(P) = 0.$$

This is the famous asymptotic freedom of strong interaction discovered in 1973 by Politzer, Gross, and Wilczek. From the physical point of view, this means that quarks are free particles at very high energies or very small distances (e.g., in the early universe or inside the proton).

3.2.4 The Quark Confinement

In 1964 Gell-Mann introduced quarks as a mathematical construction for simplifying the classification of elementary particles. But he hesitated to consider them as real particles, since the experiments in particle accelerator did not prove the existence of such particles. The situation changed in 1967 when electron-proton scattering experiments at SLAC (Stanford University) revealed an internal structure of the proton. Nowadays, we assume that there exists a quark confinement, which prevents quarks from living outside of hadrons in our present world. A deep theoretical understanding of quark confinement is a true challenge for modern theoretical physics. In 1974 Wilson initiated a discrete version of the Standard Model called lattice gauge field theory. On the basis of this lattice approach, special supercomputers have been constructed which are able to compute, for example, the proton mass as a bound state of three quarks.

In addition, there exist phenomenological models for explaining quark confinement. For example, bag models are based on the construction of effective potentials which do not allow the quarks to leave the bag. Other so-called chromoelectric models use similar methods as in superconductivity by replacing magnetic fields with electric fields. In this direction, a fundamental approach was developed by Seiberg and Witten (1994). This led to the formulation of the famous Seiberg–Witten equation which plays a fundamental role in modern differential geometry and topology.[12]

[12] As an introduction to the Seiberg–Witten equation, we recommend Moore (1996) and Jost (2000a). See also the survey article by Donaldson (1996). The Seiberg–Witten equation simplifies dramatically the sophisticated Donaldson theory for constructing topological invariants of 4-dimensional manifolds via the Yang–Mills equation.

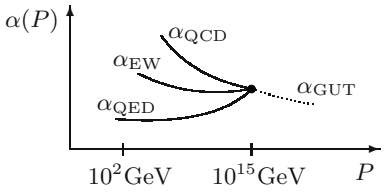


Fig. 3.3. Unification of fundamental interactions

3.2.5 Proton Decay and Supersymmetric Grand Unification

> There may be all sorts of new physical effects that come into play at particle energies above 10^{15} GeV. For example, there's no real reason to believe that baryon number would be conserved at such energies... We might expect a proton mean lifetime of the order of magnitude of $(10^{15}\mathrm{GeV})^4/(\alpha^2 m_p^5)$, essentially as estimated in a paper by Georgi, Quinn, and me (Phys. Rev. Lett. **33**, 451). This comes out to be about 10^{32} years.
>
> Steven Weinberg[13]

Figure 3.3 shows the running coupling constants $\alpha_{QED}, \alpha_{EW}, \alpha_{QCD}$ of quantum electrodynamics, electroweak interaction and strong interaction (quantum chromodynamics), respectively. Surprisingly enough, the curves of the three coupling constants intersect each other at a typical particle energy of 10^{15} GeV. Therefore, physicists conjecture that above this critical particle energy, there exists only one fundamental interaction which can be described by a theory called Grand Unified Theory (GUT). In the framework of GUT, the proton is not stable anymore. Experiments show that the mean lifetime of a proton is more than 10^{32} years. This disqualifies the original GUT based on the gauge group $SU(5)$. However, supersymmetric variants of GUT are still possible, since they possess additional degrees of freedom. Interestingly enough, the intersection of the running coupling constant curves in Fig. 3.3 is especially sharply concentrated at one point in the case of supersymmetric GUT. This seems to underline the importance of superymmetric theories.

3.2.6 The Adler–Bell–Jackiw Anomaly

In 1969 Adler, Bell, and Jackiw pointed out that there exist special Feynman diagrams in the theory of electroweak interaction which cause nasty divergent expressions called Adler–Bell–Jackiw anomalies. Fortunately enough, these anomalies disappear if one postulates the following lepton–quark symmetry:

The number of leptons is equal to the number of quarks.

This condition is fulfilled in the Standard Model of particle physics. Here, we have six leptons and six quarks. For the theory of anomalies, we refer to the monograph by Fujikawa and Suzuki (2004).

[13] In: A. Guth, K. Huang, and A. Jaffe (Eds.) (1983), pp. 1–19 (reprinted with permission).

(a) attracting (b) repelling

Fig. 3.4. Discrete dynamical system

3.3 Stable and Unstable Manifolds

The method of Wilson's renormalization group reduces the computation of critical phenomena in physics to the study of dynamical systems near a stationary point. To explain the basic ideas concerning stability and instability, let us consider discrete dynamical systems.

Iterative methods, stable manifolds, and unstable manifolds. For given real parameters a and x_0, the iterative method

$$x_{n+1} = ax_n, \qquad n = 0, 1, \ldots$$

is convergent (resp. divergent) for $|a| < 1$ (resp. $|a| > 1$). In the case of convergence, we have

$$\lim_{n\to\infty} x_n = 0.$$

The limit $x = 0$ is the unique solution of the fixed point equation

$$x = ax$$

if $a \neq 0$. We say that this fixed point is attracting (resp. repelling) if $|a| < 1$ (resp. $|a| > 1$) (Fig. 3.4). In terms of physics, we set $x(n\Delta t) := x_n$. This can be regarded as the position of a particle at the discrete time $n\Delta t$.

Let us now study the more general 2-dimensional iterative method

$$\boxed{x_{n+1} = \tfrac{1}{2}x_n, \qquad y_{n+1} = -2y_n, \qquad n = 0, 1, 2, \ldots} \tag{3.16}$$

The corresponding equation

$$x = \tfrac{1}{2}x, \quad y = -2y$$

has the unique fixed point $(x, y) = (0, 0)$. The behavior of the iterative method (3.16) critically depends on the choice of the starting point.

- For the starting point $(x_0, 0)$, the iterative method (3.16) converges to the fixed point $(0, 0)$ along the x-axis which is called a stable manifold.
- In contrast to this, for the starting point $(0, y_0)$ with $y_0 \neq 0$, the iterative method (3.16) runs away along the y-axis which is called an unstable manifold.

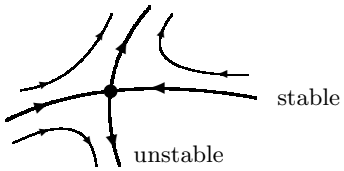


Fig. 3.5. Stable and unstable manifolds

Fig. 3.5 shows the flow of a fluid which is a continuous version of the iterative method (3.16). As an introduction to the theory of stable and unstable manifolds for dynamical systems, we recommend the monograph by Amann (1990).

> *In statistical mechanics, it turns out that the renormalization-group flow (or semi-flow) corresponds to iterative methods which possess a fixed point along with the typical property that there exist stable and unstable manifolds which are passing through the fixed point. In terms of physics, the fixed point describes a phase transition.*

We refer to the collection of survey articles by Duplantier and Rivasseau (2003).

3.4 A Glance at Conformal Field Theories

In terms of mathematics, rescaling of a physical parameter corresponds to a special conformal transformation. Therefore, it is quite natural that the theory of conformal transformations plays an important role in statistical mechanics in connection with phase transitions and in 2-dimensional quantum field theory. The point is that the conformal invariance in two dimensions determines almost uniquely the structure of the Green's functions (correlation functions). Conformal field theory was founded by Polyakov in 1970.[14] For the last 30 years, physicists have developed the so-called conformal field theory. We refer to Fuchs (1992) (introduction), Fradkin and Palchik (1996), De Francesco et al. (1997) (general theory), and Polchinski (1998), Vols. 1, 2 (string theory).

For example, the conformal symmetry plays a fundamental role in string theory, since string theories are conformally invariant. From the mathematical point of view, the theory of conformal transformations was created by Gauss in the 1820s and further developed by Riemann in the 1850s. The notion of Riemann surfaces and their uniformization by Poincaré and Koebe in 1907 is the mathematical background of conformal field theory and string theory. This will be studied in Volume VI.

[14] Conformal symmetry of critical fluctuations, JETP Lett. **12** (1970), 381.

4. Analyticity

> Mathematicians and physicists like holomorphic and meromorphic functions, since the local behavior of such functions determines completely their global behavior.
>
> Folklore

A complex number has the form

$$\boxed{z = x + y\mathrm{i}}$$

where x and y are real numbers and $\mathrm{i}^2 = -1$. We write $\Re(z) := x$ and $\Im(z) := y$ and call this the real and imaginary part of z, respectively. The set of all complex numbers is denoted by $\mathbb{C}$; it is also called the complex plane. Using polar coordinates, each complex number $z = x + y\mathrm{i}$ can be written uniquely as

$$x = r\cos\varphi, \qquad y = r\sin\varphi, \qquad -\pi < \varphi \le \pi$$

where $r := \sqrt{x^2 + y^2}$. The real numbers

$$|z| := r, \qquad \arg z := \varphi$$

are called the modulus and the principal argument of z, respectively (Fig. 4.1). Using the Euler formula $\mathrm{e}^{\mathrm{i}\varphi} = \cos\varphi + \mathrm{i}\sin\varphi$, each complex number z can also be uniquely represented as

$$z = |z| \cdot \mathrm{e}^{\mathrm{i}\varphi}, \qquad -\pi < \varphi \le \pi.$$

Sometimes it is useful to use the representation

$$z = |z|\mathrm{e}^{\mathrm{i}\varphi_*}$$

where $\varphi_* = \arg(z) + 2\pi k$ with $k = 0, \pm 1, \pm 2, \ldots$ We call φ_* an argument of the complex number, and we write $\arg_*(z) = \varphi_+$.

Convention. In what follows, the symbol U always denotes an open subset of the complex plane.

The theory of complex-valued holomorphic functions is one of the most beautiful parts of mathematics; it has played a key role in the development of modern mathematics (analysis, topology, algebraic geometry, and number theory). As an introduction to complex function theory, we recommend Hurwitz and Courant (1964) (classic), Remmert (1991), (1998), and Zeidler (2004).

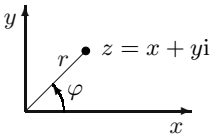


Fig. 4.1. Complex number

4.1 Power Series Expansion

Holomorphic functions. A function $f : U \to \mathbb{C}$ is called holomorphic on the open set U iff it is differentiable at each point z of U, that is, there exists the derivative

$$f'(z) = \lim_{h \to 0} \frac{f(z+h) - f(z)}{h}.$$

Theorem 4.1 *A function $f : U \to \mathbb{C}$ is holomorphic on the open set U iff, for each point $z_0 \in U$, there exists a power series such that*

$$f(z) = a_0 + a_1(z - z_0) + a_2(z - z_0)^2 + \dots$$

for all points z in some open neighborhood of z_0.

For example, polynomials are holomorphic. The power series expansion

$$\mathrm{e}^z := \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2} + \dots$$

is convergent for all $z \in \mathbb{C}$. Thus, the exponential function $z \mapsto \mathrm{e}^z$ is holomorphic on the complex plane. The same is true for the functions $z \mapsto \sin z, \cos z$ given by

$$\sin z := \frac{\mathrm{e}^{\mathrm{i}z} - \mathrm{e}^{-\mathrm{i}z}}{2\mathrm{i}}, \qquad \cos z := \frac{\mathrm{e}^{\mathrm{i}z} + \mathrm{e}^{-\mathrm{i}z}}{2} \qquad \text{for all} \quad z \in \mathbb{C}.$$

For all $z, w \in \mathbb{C}$, there holds the crucial addition theorem

$$\boxed{\mathrm{e}^{z+w} = \mathrm{e}^z \mathrm{e}^w.}$$

The exponential function has the period $2\pi\mathrm{i}$. For all $z \in \mathbb{C}$, we have the Euler relation

$$\boxed{\mathrm{e}^{\mathrm{i}z} = \cos z + \mathrm{i} \sin z.}$$

Entire functions. By definition, a function $f : \mathbb{C} \to \mathbb{C}$ is called entire iff it is holomorphic on the complex plane. For example, polynomials, the exponential function $z \mapsto \mathrm{e}^z$, and the trigonometric functions $z \mapsto \sin z$ and $z \mapsto \cos z$ are entire functions.

Locally holomorphic functions. Let z_0 be a point of the complex plane. A complex-valued function f is called locally holomorphic at the point z_0 iff there exists an open ball $B(z_0)$ centered at z_0 such that the function $f : B(z_0) \to \mathbb{C}$ is holomorphic.

Biholomorphic functions. Let U and V be open subsets of the complex plane $\mathbb{C}$. A function

$$f : U \to V$$

is called biholomorphic iff it is bijective and both f and f^{-1} are holomorphic. Biholomorphic maps are always angle-preserving, that is, the oriented angles between intersecting curves are preserved.

Conformal maps. Fix the point z_0 of the complex plane and the complex numbers a, b with $b \neq 0$. The function

$$f(z) := a + b(z - z_0), \qquad z \in \mathbb{C}$$

is the superposition of a translation, a rotation around the center z_0, and a similarity transformation with respect to z_0. Obviously, this map is angle-preserving and a biholomorphic map $f : \mathbb{C} \to \mathbb{C}$ from the complex plane onto itself. Such a map is called a conformal map of the complex plane onto itself. We want to generalize this concept. To this end, let

$$f : U \to \mathbb{C} \tag{4.1}$$

be a holomorphic function on the nonempty open subset U of the complex plane. This map is called a conformal map from U onto $f(U)$ iff it is an angle-preserving diffeomorphism[1] from the set U onto the set $f(U)$. For a function (4.1) on the nonempty open subset U of the complex plane $\mathbb{C}$, the following three properties are equivalent.

(i) The map f is conformal from U onto $f(U)$.
(ii) The function $f : U \to \mathbb{C}$ is holomorphic, injective, and $f'(z_0) \neq 0$ for all points z_0 in U.[2]
(ii) The set $f(U)$ is open and the function $f : U \to f(U)$ is biholomorphic.

In the case (ii), the function f looks locally like

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + \dots .$$

in a sufficiently small open neighborhood of each point $z_0 \in U$. Because of the condition $f'(z_0) \neq 0$, the map f is not locally degenerate at z_0.

Integrals. It turns out that integrals of holomorphic functions reflect topological properties of the domain of definition. Let the function $f : U \to \mathbb{C}$ be holomorphic, and let $C : z = z(t), t_0 \leq t \leq t_1$ be an oriented smooth curve in the set U. Define

$$\int_C f(z)dz := \int_{t_0}^{t_1} f(z(t)) \, \frac{dz(t)}{dt} \cdot dt.$$

This definition of the curve integral does not depend on the parametrization of the oriented smooth curve C. The integral changes sign if the orientation of the curve changes.

The fundamental theorem of calculus. Let $f : U \to \mathbb{C}$ be holomorphic. Then for each smooth curve C in U with initial point z_0 and end point z_1,

$$\boxed{\int_C f'(z)dz = f(z_1) - f(z_0).}$$

[1] By definition, the map $f : U \to f(U)$ is a diffeomorphism iff it is bijective and both the functions f and f^{-1} are smooth with respect to the two real variables on U and $f(U)$, respectively.

[2] This implies that the set $f(U)$ is open in the complex plane.

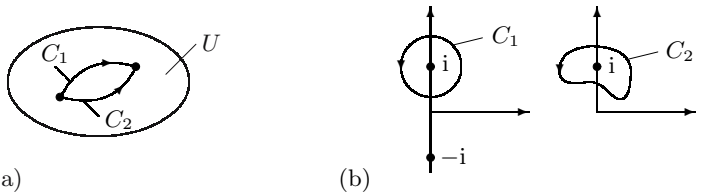


Fig. 4.2. Continuous deformation of the integration path

Proof. By the chain rule,

$$\int_{t_0}^{t_1} f'(z(t))z'(t)dt = \int_{t_0}^{t_1} \frac{df(z(t))}{dt}\, dt = f(z(t_1)) - f(z(t_0)).$$

□

4.2 Deformation Invariance of Integrals

Let the function $f : U \to \mathbb{C}$ be holomorphic. Then

$$\boxed{\int_{C_1} f(z)dz = \int_{C_2} f(z)dz} \tag{4.2}$$

if the following holds true: The two smooth curves C_1 and C_2 have the same initial and end point, and they can be continuously deformed into each other without leaving the set U and without changing the initial and end point (Fig. 4.2(a)). This result remains true if C_1 and C_2 are reasonable piecewise smooth curves (e.g., polygons).

4.3 Cauchy's Integral Formula

In recent years, progress in the theory of partial differential equations has been tremendous, often in unexpected directions, while also solving classical problems in more general settings. New fields have been added, like the study of variational inequalities, of solitons, of wave front sets, of pseudo-differential operators, of differential forms on manifolds, and so on. Much of the progress has been made by the use of functional analysis. However, in the process much of the original simplicity of the theory has been lost. This is perhaps connected with the emphasis of solving problems, which often requires piling up mountains of *a priori* estimates and the skillful juggling of function spaces to make ends meet. It is good to remember that mathematics is *not* only concerned with solving problems, but also with studying the structure and behavior of objects that it creates. One of the best examples is classical theory of functions of a complex variable. It, incidentally, does solve problems as in the Riemann mapping theorem. But much of its beauty lies in statements that can hardly be considered as "solving" anything, like the calculus of residues, or Cauchy's integral

> formula (4.3). The only "problem" solved by (4.3) is the improper one of determining the function f from its boundary values, which generally has no solution. Formula (4.3) is not only strikingly beautiful but also extremely useful. It shows easily that each holomorphic function can be differentiated infinitely often and can be represented by convergent power series.[3]
>
> Fritz John, 1984

Let C be a counterclockwise oriented circle living in the open set U. If the function f is holomorphic on U, then

$$\boxed{f(z) = \int_C \frac{f(\zeta)d\zeta}{\zeta - z}} \tag{4.3}$$

for all points z inside the circle C. This fundamental formula was discovered by Cauchy (1789–1857) in 1831 during his exile in Turin (Italy).

4.4 Cauchy's Residue Formula and Topological Charges

> The word residue was first used by Cauchy in 1826, but to be sure the definition there is quite complicated.[4]
>
> Reinhold Remmert

Meromorphic functions. The function $f : U \to \mathbb{C}$ has an isolated singularity at the point z_0 iff it is holomorphic in a punctured open neighborhood of z_0.[5] Then, there exist complex numbers $\ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots$ such that

$$f(z) = \sum_{k=-\infty}^{\infty} a_k (z - z_0)^k$$

for all z in some punctured open neighborhood of z_0.[6] The number

$$\mathrm{res}_{z_0}(f) := a_{-1}$$

is called the residue of the function f at the point z_0. If there exists a positive integer m such that $a_{-m} \neq 0$ and $a_k = 0$ for all $k < -m$, then we say that the function f has an isolated pole of order m at the point z_0. Then,

$$f(z) = \frac{a_{-m}}{(z - z_0)^m} + \ldots + \frac{a_{-1}}{z - z_0} + a_0 + a_1(z - z_0) + \ldots$$

for all points z in some punctured open neighborhood of z_0. The sum

$$\frac{a_{-m}}{(z - z_0)^m} + \ldots + \frac{a_{-1}}{z - z_0}$$

[3] F. John, A walk through partial differential equations. In: S. Chern (Ed.), Seminar on Nonlinear Partial Differential Equations, Springer, New York, 1984, pp. 73–84 (reprinted with permission).

[4] R. Remmert, Theory of Complex Functions, Springer, New York, 1991.

[5] This means that there exists an open neighborhood V of z_0 such that f is holomorphic on $V \setminus \{z_0\}$.

[6] The symbol $\sum_{k=-\infty}^{\infty} b_k$ stands for $\sum_{k=-1}^{-\infty} b_k + \sum_{k=0}^{\infty} b_k$.

is called the principal part of the function f at the point z_0.

The function $f : U \to \mathbb{C}$ is called meromorphic on the open set U iff it is holomorphic up to isolated poles. Rational functions (i.e., quotients of polynomials) and the functions

$$\tan z := \frac{\sin z}{\cos z}, \qquad \cot z := \frac{\cos z}{\sin z}$$

are meromorphic on the complex plane. The poles correspond to the zeros of the denominator. Meromorphic functions play a fundamental role in physics.

The poles of meromorphic functions describe essential physical properties (e.g., the masses of elementary particles).

The following theorem is called Cauchy's residue theorem.

Theorem 4.2 *For a meromorphic function $f : U \to \mathbb{C}$, there holds*

$$\boxed{\frac{1}{2\pi i}\int_C f(z)dz = \sum_{k=1}^{m} \mathrm{res}_{z_k}(f)} \tag{4.4}$$

if the following conditions are met: the circle C lies in the open set U, the function f has precisely the poles $z_1, \ldots, z_m$ inside the circle C and no poles on C.

The integral is equal to zero if there are no poles on the closed disc bounded by the circle C.

This is one of the most useful theorems in mathematics. For example, the function

$$f(z) := \frac{2\mathrm{i}Q}{1+z^2} = \frac{Q}{z-\mathrm{i}} - \frac{Q}{z+\mathrm{i}}$$

has the residue $\mathrm{res}_{z=\mathrm{i}}(f) = 1$ at the point $z = \mathrm{i}$. If C_1 is a counterclockwise oriented circle of radius $R < 2$ centered at the point i, then the residue theorem tells us that

$$\frac{1}{2\pi\mathrm{i}}\int_{C_1} f(z)dz = Q.$$

If we continuously deform the circle C_1 into the closed curve C_2 such that the singularities $z = \pm\mathrm{i}$ are not touched during the deformation (Fig. 4.2(b) on page 212), then

$$\frac{1}{2\pi\mathrm{i}}\int_{C_1} f(z)dz = \frac{1}{2\pi\mathrm{i}}\int_{C_2} f(z)dz = Q.$$

We say that the function f has the topological charge Q at the point $z = \mathrm{i}$.

The attribute 'topological' refers to the fact that this charge is invariant under continuous deformations of the integration path.

Topological charges are also called topological quantum numbers.

4.5 The Winding Number

Let the functions $f, g : U \to \mathbb{R}$ be meromorphic on the open set U, and let C be a counterclockwise oriented circle lying in U. Suppose that no zeros or singularities of f lie on C. Define

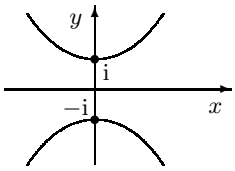


Fig. 4.3. The system $y^2 - x^2 = 1$, $xy = 0$

$$\boxed{w(f) := \frac{1}{2\pi \mathrm{i}} \int_C \frac{f'(z)}{f(z)}\, dz.}$$

This is an integer called the winding number of the function f on the circle C. For example, if $f(z) := z^n$ for a fixed integer n and the origin lies inside the circle C, then

$$w(f) = \frac{1}{2\pi \mathrm{i}} \int_C \frac{ndz}{z} = n,$$

by Cauchy's residue theorem. Generally, the winding number tells us how often the image curve $f(C)$ winds counterclockwise around the origin (see Fig. 5.6(a) on page 231). There hold the following two crucial properties.

(E) Existence principle: If $w(f) \neq 0$, then the equation $f(z) = 0$ has a solution inside the circle C.
(P) Invariance under perturbations: We have $w(f + g) = w(g)$ if

$$\sup_{z \in C} |g(z)| < \sup_{z \in C} |f(z)|.$$

Let us apply the winding number to the fundamental theorem of algebra.

4.6 Gauss' Fundamental Theorem of Algebra

> Finally, I note that it is not at all impossible that the proof, which I have based on geometric principles here, be given in a purely analytic form; but I believed the presentation which I developed here to be less abstract and to expose better the essence of the proof than one could expect from an analytic proof.
>
> Carl Friedrich Gauss (1777–1855)
> *Thesis*, 1799

Complex numbers were introduced by Bombelli in 1550; he used complex numbers systematically in order to solve algebraic equations of third degree. The imaginary unit, i, solves the equation $z^2 + 1 = 0$, and we have the factorization

$$z^2 + 1 = (z - \mathrm{i})(z + \mathrm{i}).$$

The fundamental theorem of algebra tells us the following much more general result.

Theorem 4.3 *Let $n = 1, 2, \ldots$ For each polynomial*

$$p(z) := z^n + a_{n-1} z^{n-1} + \ldots + a_1 z + a_0$$

with complex coefficients $a_0, a_1, \ldots$, *there exist complex numbers* $z_1, \ldots, z_n$ *such that*

$$p(z) = (z - z_1)(z - z_2) \cdots (z - z_n) \qquad \textit{for all} \quad z \in \mathbb{C}.$$

Gauss proved this theorem for the first time in his Ph.D. thesis[7] in 1799. He reduced this problem to the intersection problem for two real algebraic curves. To explain the basic idea, consider the equation

$$z^2 + 1 = 0, \qquad z \in \mathbb{C}.$$

Letting $z = x + y\mathrm{i}$, we get $x^2 - y^2 + (2xy)\mathrm{i} + 1 = 0$. This problem is equivalent to the system of equations

$$y^2 - x^2 = 1, \qquad 2xy = 0.$$

In terms of geometry, this system describes the intersection of a hyperbola with the y-axis (Fig. 4.3). In the general case, Gauss had to study the intersection problem for two algebraic curves of nth degree. The second and third proofs of this theorem, given by Gauss in the years 1815 and 1816, were purely analytic, while the fourth proof, on the 50th anniversary of his doctorate in 1849, was closely related to the first. In fact, the first and the fourth proof contain a gap. Gauss assumed certain properties of algebraic curves to be self-evident, which is not at all the case. This gap was filled by Alexander Ostrowski in 1920 (see Ostrowski's article in the Collected Works of Gauss, Vol. 10). Let us give an elegant proof of the fundamental theorem of algebra by using the winding number.

Proof. We proceed in two steps.

- Step 1: Existence of at least one zero. Set $f(z) := z^n$, and $p(z) = f(z) + g(z)$. Choose a counterclockwise oriented circle C of radius R centered at the origin. Then $|f(z)| = R^n$ on C. Furthermore,

$$|g(z)| \le \sum_{k=0}^{n-1} |a_k| \cdot |z^k| \le \text{const} \cdot R^{n-1} \qquad \text{on } C.$$

  If we choose the radius R sufficiently large, then the perturbation property (P) of the winding number above tells us that

$$w(p) = w(f + g) = w(f) = n.$$

  By the existence principle (E), $p(z_1) = 0$ for some z_1.
- Step 2: Factorization. By the Taylor theorem,

$$p(z) = p(z_1) + (z - z_1)\big(p'(z_1) + \frac{1}{2}p''(z_1)(z - z_1)^2 + \ldots\big).$$

  Hence $p(z) = (z - z_1)q(z)$ where q is a polynomial of degree $n - 1$. By induction, we obtain the claim. □

Intuition. The geometric idea behind the argument from step 1 is quite simple. The map $z \mapsto z^n$ sends the circle C to a circle which winds n times around the origin. If the radius R is sufficiently large, then the map $z \mapsto p(z)$ sends the circle C to a curve which winds n times around the origin as well, by a perturbation argument. By continuity, there must exist some point z_1 inside the circle C which

[7] C. F. Gauß, The four proofs of the fundamental theorem of algebra (in German), Ostwald's Klassiker, No. 14, Engelmann, Leipzig, 1890. The book series *Ostwald's classic library* (Ostwald's Klassiker) is the best source for classic papers in mathematics and physics.

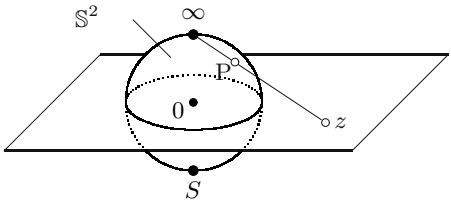


Fig. 4.4. Riemann sphere

is mapped to the origin, that is, $p(z_1) = 0$. The point z_1 is the desired zero of the given polynomial p.

Gauss implicitly used this intuitive argument, but he had to argue in a more sophisticated way, since he did not have the rigorous theory of the winding number at hand.

4.7 Compactification of the Complex Plane

Each bounded sequence (z_n) in the complex plane contains a convergent subsequence. This important property is not always true for unbounded sequences. In order to cure this defect, we add the symbol ∞ to the complex plane $\mathbb{C}$. The set

$$\boxed{\overline{\mathbb{C}} := \mathbb{C} \cup \{\infty\}}$$

is called the closed complex plane.

The Riemann sphere. The unit sphere

$$\mathbb{S}^2 := \{(x, y, \zeta) \in \mathbb{R}^3 : \ x^2 + y^2 + \zeta^2 = 1\}$$

in the 3-dimensional Euclidean space is called the Riemann sphere. Naturally enough, the point $N := (0, 0, 1)$ is called the North Pole of $\mathbb{S}^2$. Using Fig. 4.4, we set $\chi(z) := P$. The map

$$\chi : \mathbb{C} \to \mathbb{S}^2 \setminus \{N\}$$

can be extended to a map $\chi : \overline{\mathbb{C}} \to \mathbb{S}^2$ by setting $\chi(\infty) := N$. A sequence (z_n) in the closed complex plane is said to converge to the point z iff this is true for the corresponding points on the Riemann sphere. We now get the desired convergence theorem:

Each sequence in the closed complex plane has a convergent subsequence.

We call the closed complex plane a compactification of the complex plane. In modern mathematics, the compactification of mathematical objects is used very frequently. Using the topological terminology introduced in Sect. 5.5 on page 239, the Riemann sphere $\mathbb{S}^2$ is a compact topological space, and the closed complex plane $\overline{\mathbb{C}}$ is homeomorphic to the Riemann sphere. The inverse map $\chi^{-1} : \mathbb{S}^2 \setminus N \to \mathbb{C}$ is called stereographic projection.

By definition, a function f is called locally holomorphic at the point $z = \infty$ iff the function

$$g(z) := f\left(\frac{1}{z}\right)$$

is locally holomorphic at the point $z = 0$. For example, the functions

$$f(z) := \frac{1}{z^n}, \qquad n = 1, 2, \ldots$$

are locally holomorphic at the point $z = \infty$.

Theorem 4.4 *Each holomorphic function from the closed complex plane into itself is constant.*

This theorem is named after Liouville, but it was first proved by Cauchy in 1844. We want to use this theorem in order to give an alternative proof of the fundamental theorem of algebra. By our argument above, it is sufficient to prove that each polynomial of order $n \geq 1$ has a zero.

Application to the fundamental theorem of algebra. Suppose that the polynomial p has no zeros on the complex plane. Define

$$f(z) := \frac{1}{p(z)}, \qquad z \in \mathbb{C},$$

and set $f(\infty) := 0$. Since the function f behaves like $1/z^n$ near the point ∞, it is locally holomorphic at ∞. Thus, the function f is holomorphic on the closed complex plane, and hence it is constant, a contradiction. □

The Liouville Theorem 4.4 is the prototype of a mathematical theorem which relates the structure of analytic objects to the topology of the space on which the objects live. In terms of physics, this corresponds to the situation that the structure of physical fields depends on the topological structure of the underlying space-time manifold.

4.8 Analytic Continuation and the Local-Global Principle

Let $f, g : U \to \mathbb{R}$ be holomorphic functions on the open, arcwise connected set U such that

$$f(z_n) = g(z_n) \qquad \text{for all} \quad n = 1, 2, \ldots,$$

and the sequence (z_n) with $z_n \neq a$ for all n converges to some point a in U as $n \to +\infty$. Then $f = g$ on U.

Let U and V be open, arcwise connected sets in the complex plane such that $U \subset V$. If $f : U \to \mathbb{C}$ and $F : V \to \mathbb{C}$ are holomorphic and $f = F$ on U, then F is called the holomorphic (or analytic) extension of f. This holomorphic (or analytic) continuation is uniquely determined. For example, the power series

$$f(z) = 1 + z + z^2 + \ldots$$

is convergent for all $z \in \mathbb{C}$ with $|z| < 1$. Since $f(z) = \frac{1}{1-z}$, the function f can be uniquely extended to the holomorphic function $F : \mathbb{C} \setminus \{1\} \to \mathbb{C}$ given by

$$F(z) := \frac{1}{1-z} \qquad \text{for all} \quad z \in \mathbb{C}, z \neq 1.$$

As a second example, consider the energy function

$$\boxed{f(E) := \sqrt{E} \qquad \text{for all} \quad E > 0.}$$

This function allows the power series expansion

$$f(E) = \{1 + (E-1)\}^{1/2} = 1 + \tfrac{1}{2}(E-1) + \binom{\frac{1}{2}}{2}(E-1)^2 + \dots$$

for all complex energy values E with $|E-1| < 1$. The unique global analytic continuation is given by the double-valued expression

$$\boxed{F(E) := \pm\sqrt{|E|} \cdot \mathrm{e}^{\frac{1}{2}\mathrm{i}\arg E} \qquad \text{for all} \quad E \in \mathbb{C}.}$$

We also write $F(E) = \sqrt{E}$.

Counterexample. The fact that a function allows a holomorphic extension represents important and highly nontrivial information, as the following simple counterexample (ii) shows.

(i) Consider the real function $f : [\frac{1}{2}, 1] \to \mathbb{R}$ defined by $f(x) := x$. There exist infinitely many smooth functions $g : \mathbb{R} \to \mathbb{R}$ which are extensions of f, but there is only one holomorphic extension, namely, $f(z) := z$ for all $z \in \mathbb{C}$.

(ii) Modifying the preceding example, set

$$f(x) := |x|$$

if either $x \in [\frac{1}{2}, 1]$ or $x \in [-1, -\frac{1}{2}]$. There exist infinitely many smooth extensions $g : \mathbb{R} \to \mathbb{R}$ of the function f, but *no* holomorphic extension. To see this, suppose that there exists such an extension. This extension would be uniquely determined by the values of f on the interval $[\frac{1}{2}, 1]$. By (i), $f(x) = x$ on $[-1, -\frac{1}{2}]$, a contradiction.

Generalization of terminology. Let D be an arbitrary subset of the complex plane (e.g., the closed unit disc). The function

$$f : D \to \mathbb{C}$$

is called holomorphic iff there exists a holomorphic function $g : U \to \mathbb{C}$ such that D is a subset of the open set U and $f = g$ on D. A function is called locally holomorphic at a point z_0 iff it is defined and holomorphic on some open neighborhood of the point z_0.

4.9 Integrals and Riemann Surfaces

Algebraic integrals. Let $R = R(z, w)$ be a rational function with respect to the complex variables z and w (i.e., R is the quotient of two polynomials with respect to z and w). By an algebraic integral or Abelian integral, we mean an integral of the form

$$\int_C R(z, w(z))dz$$

where C is a curve, and the function $w = w(z)$ is given by the equation

$$\boxed{w^n + p_{n-1}(z)w^{n-1} + \dots + p_1(z)w + p_0(z) = 0, \qquad z, w \in \mathbb{C}.} \tag{4.5}$$

Here, $p_0, p_1, \ldots$ are polynomials. Such integrals were first studied by Abel (1802–1829). In an ingenious way, Riemann (1826–1866) discovered that algebraic integrals can be understood best by using the topological concept of the Riemann surface. Let us explain this by considering a simple example.

The Riemann surface of the logarithmic function. For each given complex number $z \neq 0$, the equation

$$\boxed{z = \mathrm{e}^w, \qquad w \in \mathbb{C}}$$

has a set of solutions which we denote by $w = \ln z$. Explicitly,

$$\ln z = \ln|z| + \mathrm{i}\arg z + 2\pi k, \qquad k = 0, \pm 1, \pm 2, \ldots \tag{4.6}$$

This function is many-valued on the punctured complex plane $\mathbb{C} \setminus \{0\}$. It was Riemann's idea to construct a set $\mathcal{R}$ such that the function

$$\boxed{\ln : \mathcal{R} \to \mathbb{C}}$$

is single-valued on $\mathcal{R}$. This can be done easily. To this end, set

$$S_k := \mathbb{C} \setminus \{0\}, \qquad k = 0, \pm 1, \pm 2, \ldots$$

We cut each sheet S_k along the negative real axis, and we glue the sheets along the cuts together in the following way.

- If we start from the point $z = 1$ on the sheet S_0, and we move counterclockwise along the unit circle, then we change from the sheet S_k to the sheet S_{k+1} at the point $z = -1$ where $k = 0, 1, 2, \ldots$.
- Similarly, if we start from the point $z = 1$ at the sheet S_0 and we move clockwise along the unit circle, then we change from S_k to S_{k-1} at the point $z = -1$ where $k = 0, -1, -2, \ldots$

Furthermore, if $z \in S_k$, then we define $\ln z := \ln|z| + \mathrm{i}\arg z + 2\pi k$. The set $\mathcal{R}$ is called the Riemann surface of the function $\ln$. This is an 'infinite round staircase'. If we cut the sheet S_0 along the negative real axis, then the restriction

$$\ln : S_0 \setminus\ ]-\infty, 0] \to \mathbb{C}$$

is called the principal branch of the logarithmic function; this function is single-valued and holomorphic. Explicitly, for the principal branch,

$$\ln z = \ln|z| + \mathrm{i}\arg z, \qquad -\pi < \arg z < \pi.$$

For all complex numbers z with $|z| < 1$, the principal branch allows the power series expansion

$$\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \ldots$$

Application to integrals. The idea of the Riemann surface is extremely useful for computing curve integrals.

The point is that we have to choose curves on the Riemann surface of the primitive function to the integrand.

For example, consider the smooth curve $C : z = z(t), t_0 \le t \le t_1$. By the fundamental theorem of calculus,

$$\int_C \frac{dz}{z} = \ln z(t_1) - \ln z(t_0).$$

However, this value is not well-defined, since the logarithmic function is many-valued. This defect can be cured completely if we regard C as a curve on the Riemann surface $\mathcal{R}$ of the logarithmic function. For example, let C denote the counterclockwise oriented unit circle on the complex plane. Now regard C as a curve on the Riemann surface $\mathcal{R}$ of the logarithmic function. For example, let us start at the point $z = 1$ on the sheet S_0. Moving counterclockwise, we end up at the point $z = 1$ on the sheet S_1. Hence

$$\int_C \frac{dz}{z} = \ln_{S_1} z(t_1) - \ln_{S_0} z(t_0) = \ln_{S_1} 1 - \ln_{S_0} 1 = 2\pi\mathrm{i}.$$

The same result is obtained by using Cauchy's residue theorem. The Riemann surface $\mathcal{R}$ reveals the natural background of the residue theorem.

Algebraic curves and Riemann surfaces. The real equation

$$w - z = 0, \qquad z, w \in \mathbb{R}$$

represents a real curve, namely, a straight line through the origin. The complex equation

$$w - z = 0, \qquad z, w \in \mathbb{C}$$

represents a complex curve (also called 1-dimensional complex manifold). This complex curve corresponds to the complex plane, $\mathbb{C}$, which is equivalent to the two-dimensional real plane,[8] $\mathbb{R}^2$. If we compactify the complex plane, $\mathbb{C}$, then we obtain the closed complex plane, $\overline{\mathbb{C}}$, which is in one-to-one correspondence to the Riemann sphere, $\mathbb{S}^2$. In terms of the theory of manifolds to be introduced in Sect. 5.4 on page 234, Riemann surfaces are defined to be 1-dimensional, complex, arcwise connected manifolds. Using this terminology, the following hold true:

The closed complex plane and the Riemann sphere are conformally equivalent compact Riemann surfaces of genus zero.

Furthermore, the real equation

$$w^2 - z = 0, \qquad z, w \in \mathbb{R}$$

describes a parabola. The complex extension

$$w^2 - z = 0, \qquad z, w \in \mathbb{C}$$

describes the Riemann surface to the function $w = \sqrt{z}$. After compactification, this is a compact Riemann surface which is conformally equivalent to the Riemann sphere. More generally, each of the algebraic equations (4.5) describes, after passing to connected components and using compactification, a compact Riemann surface of genus g where $g = 0, 1, 2, \ldots$ In 1907, Poincaré and Koebe proved independently the famous uniformization theorem telling us that

Each algebraic curve (and hence each compact Riemann surface) possesses a global smooth parametrization $z = z(t)$, $w = w(t)$ for all parameters $t \in \mathcal{T}$

[8] Traditionally, the notion complex curve (resp. plane or surface) corresponds to a space of complex dimension one (resp. real dimension two).

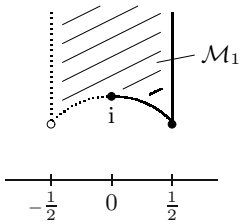


Fig. 4.5. The Riemann moduli space $\mathcal{M}_1$

where the parameter space $\mathcal{T}$ has to be chosen in an appropriate manner. This way, many special functions of mathematical physics appear as parametrizations of algebraic curves (trigonometric functions, elliptic functions, modular functions, and automorphic functions). For example, the complex unit circle, that is, the complex curve

$$z^2 + w^2 - 1 = 0, \qquad z, w \in \mathbb{C}$$

possesses the global parametrization $z = \cos t,\ w = \sin t,\ t \in \mathbb{C}$. Let e_1, e_2, e_3 be three pairwise different complex numbers. Then, the algebraic curve

$$w^2 - 4(z - e_1)(z - e_2)(z - e_3) = 0, \qquad z, w \in \mathbb{C} \tag{4.7}$$

possesses the global parametrization $z = \wp(t),\ w = \wp'(t),\ t \in \mathbb{C}$ where $\wp$ denotes the Weierstrass elliptic function (i.e., $\wp$ is double-periodic). The Riemann surface of the algebraic curve (4.7) is conformally equivalent to a torus which has always the genus $g = 1$. The Riemann surface to the complex curve

$$z - \mathrm{e}^{w} = 0, \qquad z, w \in \mathbb{C}$$

is nothing other than the Riemann surface to the function $w = \ln z$.

The space of string states. Riemann surfaces and their applications in string theory will be studied in Volume VI. At this point let us only mention that

Compact Riemann surfaces correspond to string states; conformally equivalent Riemann surfaces represent the same string state.

In this modern terminology, the space of string states corresponds to Riemann's classical moduli space. If g denotes the genus of the string state, then the following are met:

(i) If $g \geq 2$, then the space of string states can be parametrized by $3g - 3$ complex parameters.

(ii) If $g = 0$, then there exists a unique string state which corresponds to the Riemann sphere.

(ii) If $g = 1$, then the string states correspond to tori. The space of string states is in one-to-one correspondence to the subset $\mathcal{M}_1$ of the complex upper half-plane pictured in Fig. 4.5. Therefore, the string states can be parameterized by one complex parameter. Explicitly, we start with the strip

$$\{-\tfrac{1}{2} < \Re(z) \leq \tfrac{1}{2},\ \Im(z) > 0\}.$$

We take away the points of the strip that lie in the closed unit disc, and we add the arc $\{z \in \mathbb{C} :\ |z| = 1,\ \Re(z) \geq 0\}$ of the unit circle.

For genus $g \geq 1$, the string state space is *not* a manifold, but has singularities. The theory of Riemann surfaces and its generalizations has played a key role in the development of modern algebraic geometry.

Suggested reading. As an introduction to Riemann surfaces, we recommend the following two books:

- A. Hurwitz and R. Courant, Lectures on Complex Function Theory and Elliptic Functions (in German), Springer, Berlin, 1964.
- J. Jost, Compact Riemann Surfaces: An Introduction to Contemporary Mathematics, Springer, Berlin, 1997.

Furthermore, we refer to the following books:

- M. Farkas and I. Kra, Riemann Surfaces, Springer, New York, 1992.
- M. Farkas and I. Kra, Theta Constants, Riemann Surfaces and the Modular Group: An Introduction with Applications to Uniformization Theorems, Partition Identities and Combinatorial Number Theory, American Mathematical Society, Providence, Rhode Island, 2001.
- L. Ford, Automorphic Functions, Chelsea, New York, 1972.
- O. Forster, Lectures on Riemann Surfaces, Springer, Berlin, 1981.
- K. Maurin, Riemann's Legacy: Riemann's Ideas in Mathematics and Physics of the 20th Century, Kluwer, Dordrecht, 1997.
- R. Narasimhan, Compact Riemann Surfaces, Birkhäuser, Basel, 1996.

A collection of beautiful survey articles can be found in

- R. Waldschmidt et al. (Eds.), From Number Theory to Physics, Springer, New York, 1995.

4.10 Domains of Holomorphy

> Physically interpretable functions obtained by analytic continuation from functions describing physical phenomena also describe physical phenomena: they are not mere mathematical chimeras... It turns out that the Wightman distributions[9] are boundary values of holomorphic functions of N variables, the Wightman functions, and these have proved to be a very useful tool for investigating axiomatic quantum field theory.
>
> Res Jost, 1965
> *The General Theory of Quantized Fields*[10]

It turns out that there exists a crucial difference between holomorphic functions on the complex plane and holomorphic functions on the higher-dimensional complex spaces $\mathbb{C}^2, \mathbb{C}^3, \ldots$ This concerns analytic continuation. Let us discuss this. For an open set $\mathcal{U}$ in $\mathbb{C}^N$, the function $f : \mathcal{U} \to \mathbb{C}^N$ is called holomorphic on $\mathcal{U}$ iff the partial derivatives

$$\frac{\partial f}{\partial z_j}, \qquad j = 1, \ldots, N$$

exist on $\mathcal{U}$. This is equivalent to the fact that the function f can be locally represented by a power series expansion which converges absolutely in some open neighborhood of each point in $\mathcal{U}$.

[9] The Wightman distributions are the averaged vacuum expectation values of products of quantum fields.

[10] American Mathematical Society, Providence, Rhode Island, 1965 (reprinted with permission).

An open, arcwise connected subset $\mathcal{U}$ of $\mathbb{C}^N$ is called a domain of holomorphy iff there exists a holomorphic function $f : \mathcal{U} \to \mathbb{C}$ which cannot be extended to a holomorphic function on a larger open, arcwise connected set.

On the complex plane $\mathbb{C}$, each open, arcwise connected set U is a domain of holomorphy. This theorem does not remain true for $\mathbb{C}^2, \mathbb{C}^3, \ldots$ For example, consider the set

$$\mathcal{W} := \{(z_1, z_2) \in \mathbb{C}^2 : \tfrac{1}{2} < |z| < 1\}$$

with $|z| := \sqrt{|z_1|^2 + |z_2|^2}$. Then, each holomorphic function $f : \mathcal{W} \to \mathbb{C}$ can be extended to a holomorphic function $f : \mathcal{U} \to \mathbb{C}$ on the open unit ball

$$\mathcal{U} := \{(z_1, z_2) \in \mathbb{C}^2 : |z| < 1\},$$

but a further extension to a larger open set is not always possible. This tells us that the open set $\mathcal{W}$ is *not* a domain of holomorphy. In contrast to this, the open ball $\mathcal{U}$ is a domain of holomorphy. This is a special case of the following more general result:

Each convex open subset of $\mathbb{C}^N$ is a domain of holomorphy.

The famous 1938 Bochner theorem on the analytic continuation of functions of several variables tells us the following. Consider a so-called tube

$$\mathbb{R}^N + \mathrm{i}\Omega := \{(x + y\mathrm{i}) : \ x \in \mathbb{R}^N, \ \ y \in \Omega\}$$

where Ω is a convex open subset of $\mathbb{R}^N$, $N = 1, 2, \ldots$.

Theorem 4.5 *If a function f is locally holomorphic at some point of the tube $\mathbb{R} + \mathrm{i}\omega$, then it can be holomorphically extended to the tube. But a further holomorphic extension to a larger open set in $\mathbb{C}^N$ is not always possible.*

This theorem tells us that the tube $\mathbb{R}^N + \mathrm{i}\Omega$ is a domain of holomorphy in $\mathbb{C}^N$. The proofs can be found in Vladimirov (1966).

4.11 A Glance at Analytic *S*-Matrix Theory

In the 1950s and 1960s, physicists thoroughly studied scattering processes for elementary particles by using analytic continuation for the S-matrix and the Green's functions. If these functions have a singularity at the complex energy

$$E = E_0 + \mathrm{i}\Delta E,$$

then there exists an elementary particle with rest mass $m_0 = E_0/c^2$ and mean lifetime $\Delta t = \hbar/\Delta E$. As an introduction to analytic S-matrix theory, we recommend the monographs by

- A. Barut, The Theory of the Scattering Matrix, MacMillan, New York, 1967.
- G. Chew, The Analytic S-Matrix: A Basis for Nuclear Democracy, Benjamin, New York, 1966.
- R. Eden, High Energy Collisions of Elementary Particles, Cambridge University Press, 1967.
- I. Todorov, Analytic Properties of Feynman Diagrams in Quantum Field Theory, Pergamon Press, London, 1971.

We will study this in Volume V on the physics of the Standard Model.

4.12 Important Applications

Analyticity plays a fundamental role in physics. In this treatise, we will encounter the following topics:

- residue theorem, Fresnel integrals, and the diffraction of light (Vol. I);
- causality and dispersion relations (Vol. I);
- analytic continuation and path integrals (Vol. I);
- analytic continuation of the zeta function, the prime number theorem, and the Casimir effect in quantum field physics (Vol. I);
- asymptotic formulas for the Laplace transform (Vol. I);
- the Paley–Wiener theorem, wave front sets, and the propagation of singularities (Vol. I);
- distributions as boundary values of holomorphic functions (Vol. I);
- the Riemann–Hilbert problem and renormalization (Vols. I, IV);
- singularities of analytic functions and phase transitions (Vols. I, IV);
- scattering processes, Riemann surfaces, and energy resonances (Vol. II);
- Borel summation (Vol. II);
- analytic S-matrix theory for studying scattering processes for elementary particles (Vol. V);
- analytic continuation of Wightman functions (the edge-of-the-wedge theorem, the Bargmann–Hall–Wightman theorem), the CPT symmetry principle, and the spin-statistics theorem in axiomatic quantum field theory (Vol. V);
- analytic continuation and the Euclidean approach to axiomatic quantum field theory (the Osterwalder–Schrader axioms) (Vol. V);
- Riemann surfaces and string theory (Vol. VI);
- Kähler manifolds and string theory (Vol. VI);
- holomorphic functions, strings, conformal field theory, the structure of Green's functions, and phase transitions in statistical physics (Vol. VI).

5. A Glance at Topology

> Topology is precisely that mathematical discipline which allows a passage from the local to the global.
>
> René Thom (1923–2002)

Topology studies the qualitative behavior of mathematical and physical objects. The following results discussed in the preceding chapter are related to topology:

- deformation invariance of the integral of holomorphic functions,
- Cauchy's residue theorem,
- properties of the winding number,
- Liouville's theorem,
- analytic continuation of holomorphic functions,
- Abelian integrals and Riemann surfaces.

Topology was created by Poincaré (1854–1912) at the end of the 19th century and was motivated by the investigation of Riemann surfaces and the qualitative behavior of the orbits of planets, asteroids, and comets in celestial mechanics.[1] Topology studies far-reaching generalizations of the results summarized above.

5.1 Local and Global Properties of the Universe

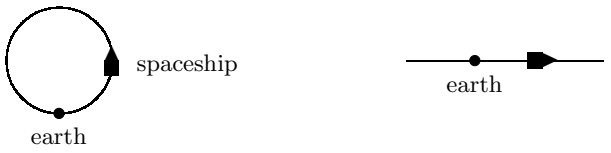
Since ancient times, scientists have made enormous efforts to understand

- the macrocosmos – our universe – and
- the microcosmos – the world of elementary particles.

A unified theory for the four fundamental forces in nature (i.e., strong, weak, electromagnetic, and gravity) has the task to combine the macrocosmos with the microcosmos. In Fig. 5.1, two 1-dimensional models of the universe are pictured. These two models possess the same local structure near the earth, but the global structures are completely different. To illustrate this, consider a spaceship starting on earth and moving the same direction all the time. In a closed universe, the spaceship may return to earth, whereas this is impossible in an open universe. Note that in reality, the universe is expanding. This means that the radius of the closed universe in Fig. 5.1(a) is expanding in time. At the time of the Big Bang, the universe was concentrated at one point.

The study of the global behavior of geometric objects is the subject of topology. We expect that the topology of the global universe influences the physics of elementary particles. As an example, consider a photon of wave length λ. If the

[1] H. Poincaré, Analysis situs, J. Math. École Polytechnique **1** (1895), 1–121.



(a) closed universe (b) open universe

Fig. 5.1. Local and global structure of the universe

radius of our universe r is finite, then we get the inequality $\lambda \leq r$. This implies that there exists a lower bound (or cut-off) for the energy of photons,

$$E = \frac{hc}{\lambda} \geq \frac{hc}{r}.$$

In fact, the so-called infrared (or low-energy) catastrophe of quantum electrodynamics concerns mathematical problems which arise at low energies ($\lambda \to +\infty$).

It is thinkable that our universe possesses a very strange global structure which results from the cooling process after the Big Bang. For example, we know that the freezing of water can produce bizarre ice flowers. In Fig. 5.2(a), the earth E and a distant galaxy G are connected by a wormhole W which allows fast travelling from E to G by leaving the regular universe and entering the hyperspace.[2]

In Fig 5.2(b), we picture a singularity S which is observed in a 1-dimensional universe. This singularity can be obtained by projecting a regular space curve without singularities (e.g., without self-intersections) onto a plane. In algebraic geometry, this method is widely used for the so-called blowing-up of singularities.

> *It is thinkable that the complex structure of the observed 4-dimensional space-time physics results from a much simpler physics in higher dimensions.*

This is one of the hopes of all the physicists who are working in the theory of strings and d-dimensional branes.[3] This explains why topology enters more and more into modern physics. Of course, there is much room for fancy ideas and wild speculations.

5.2 Bolzano's Existence Principle

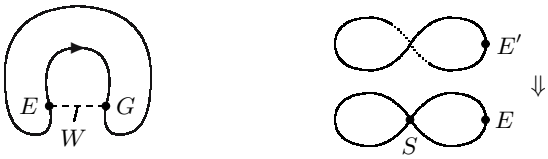
We are going to consider the prototype of a topological theorem on the intersection of continuous curves which dates back to Bolzano (1781-1848).

The mapping degree. For a real continuous function $f : [0,1] \to \mathbb{R}$ with $f(0) \neq 0$ and $f(1) \neq 0$, we define the mapping degree $\deg(f)$ by setting

$$\deg(f) := \begin{cases} 1 & \text{if} \quad f(0) < 0, \quad f(1) > 0, \\ -1 & \text{if} \quad f(0) > 0, \quad f(1) < 0, \\ 0 & \text{if} \quad f(0)f(1) > 0. \end{cases}$$

[2] The theory of wormholes was developed by Wheeler in the 1950s (see Thorne (1993)).

[3] 'Branes' generalize membranes to higher dimensions.



(a) wormhole (b) projection

Fig. 5.2. Strange one-dimensional universes

If the function is smooth and has only a finite number of zeros $x_1, ..., x_N$ along with $f'(x_j) \neq 0$ for all j, then

$$\deg(f) = \operatorname{sgn} f'(x_1) + ... + \operatorname{sgn} f'(x_N).$$

In geometric terms, the mapping degree coincides with the intersection number between the x-axis and the graph of f (Fig. 5.3).

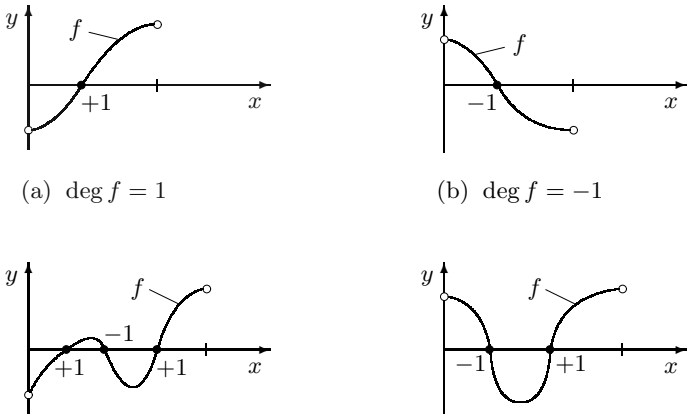
In 1817 Bolzano published a fundamental paper. In the introduction, he wrote the following.

> There are two theorems of which it could be said that until recently a completely correct proof of the same was unknown. One is the fundamental theorem of algebra proved by Gauss in his 1799 dissertation; the other one is the following theorem: That between two quantities of unknown size which evince a result of opposite sign, there must lie at least one root of the equation.

In modern language, Bolzano's theorem reads as follows.

The continuous function $f : [0, 1] \to \mathbb{R}$ has a zero if $f(0)f(1) < 0$.

Intuitively, the condition $f(0)f(1) < 0$ tells us that the function f has different signs at the two end points $x = 0$ and $x = 1$, and hence the graph of f has to



(a) $\deg f = 1$ (b) $\deg f = -1$

(c) $\deg f = 1 - 1 + 1 = 1$ (d) $\deg f = -1 + 1 = 0$

Fig. 5.3. Existence of at least one zero

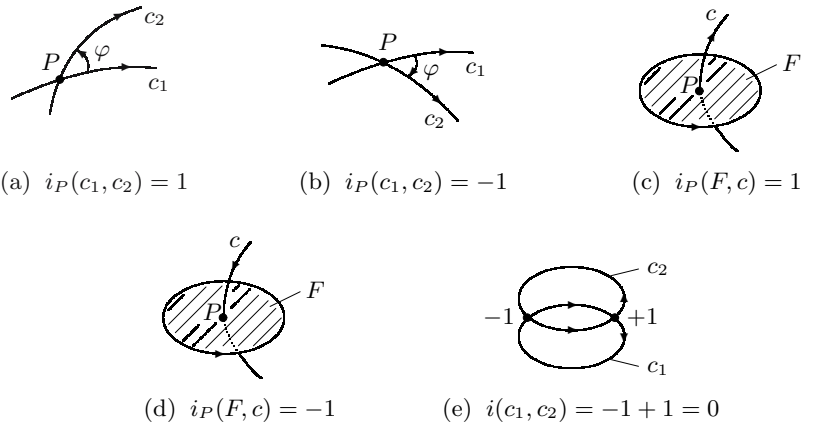


Fig. 5.4. Intersection number

intersect the x-axis. In terms of the mapping degree, the Bolzano theorem reads as follows. We are given a continuous function $f : [0, 1] \to \mathbb{R}$ which does not vanish on the boundary of the interval.

If $\deg(f) \neq 0$*, then the function* f *has a zero.*

Invariance of the mapping degree under deformations of the maps. Consider a finite time interval $[t_0, t_1]$. Suppose that we are given two continuous functions $f, g : [0, 1] \to \mathbb{R}$. Then, we get

$$\deg(f) = \deg(g)$$

if there exists a continuous map

$$H : [0, 1] \times [t_0, t_1] \to \mathbb{R}$$

such that $H(x, t_0) = f(x)$ and $H(x, t_1) = g(x)$ for all $x \in [0, 1]$. Furthermore, we have to assume that $H(x, t) \neq 0$ for the boundary points $x = 0, 1$ and all times $t \in [t_0, t_1]$.

Perspectives. The Bolzano theorem represents one of the most fundamental topological existence principles in mathematics. It allows far-reaching generalizations

- to the Euclidean spaces $\mathbb{R}^n, n = 1, 2, \ldots$ (the Brouwer mapping degree introduced in 1910)
- and to infinite-dimensional Banach spaces (the Leray–Schauder mapping degree introduced in 1936)
- along with numerous applications to differential and integral equations.

This can be found in the author's monographs Zeidler (1986), Vols. I–IV.

5.3 Elementary Geometric Notions

The most important elementary geometric notion is the concept of intersection number (Fig. 5.4).

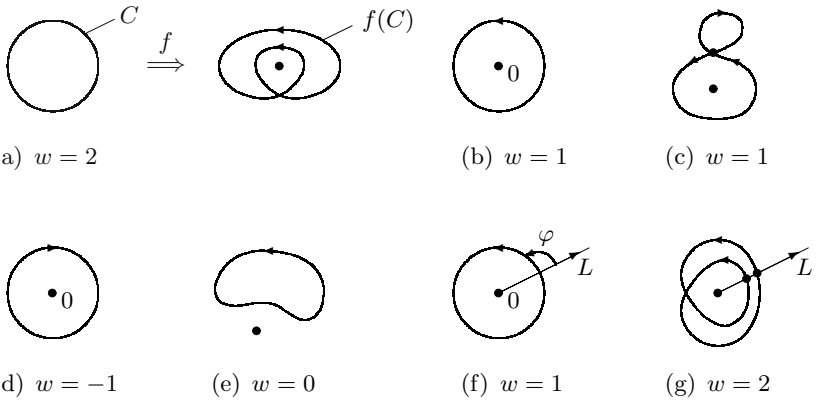
(a) $i = 1$ (b) $i = -1$

Fig. 5.5. Linking number

- Local intersection number for two oriented plane curves c_1 and c_2: At the intersection point P, define

$$i_P(c_1, c_2) := 1 \quad (\text{resp.} \quad i_P(c_1, c_2) := -1)$$

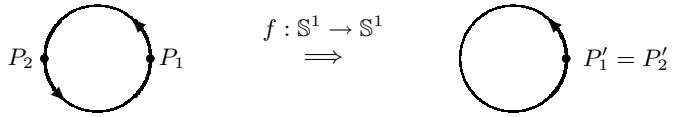
  iff the curve c_2 can be obtained from the curve c_1 by a counterclockwise (resp. clockwise) rotation about the point P with positive rotation angle $0 < \varphi < \pi$ (Figs. 5.4(a), (b)).
- Global intersection number $i(c_1, c_2)$ for two oriented plane curves: This is equal to the sum of local intersection numbers provided this number is finite (Fig. 5.4(e)).
- Local intersection number of an oriented curve and an oriented 2-dimensional surface: This is defined as pictured in Figs. 5.4(c), (d).
- Global intersection number of an oriented curve and an oriented 2-dimensional surface: This is defined to be the sum of the local intersection numbers.
- Linking number for two closed curves in the 3-dimensional space: This is reduced to the intersection number between an oriented curve and an oriented 2-dimensional surface as pictured in Fig. 5.5.
- Winding number: For a given plane curve, the winding number w tells us how many times the curve surrounds the origin, taking orientation into account, as pictured in Figs. 5.6(b)-(g). Note that the winding number is equal to the intersection number between a ray L through the origin and the curve (Figs. 5.6(f), (g)).



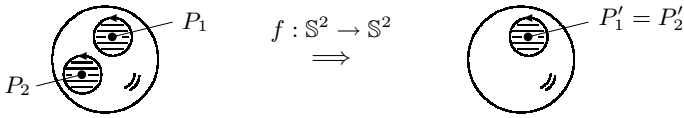
(a) $w = 2$ (b) $w = 1$ (c) $w = 1$

(d) $w = -1$ (e) $w = 0$ (f) $w = 1$ (g) $w = 2$

Fig. 5.6. Winding number



(a) $\deg f = \deg_{P_1} f + \deg_{P_2} f = 1 + 1 = 2$



(b) $\deg f = \deg_{P_1} f + \deg_{P_2} f = 1 - 1 = 0$

Fig. 5.7. Mapping degree

- Local mapping degree: Consider maps as pictured in Fig. 5.7. If $f(P) = P'$ and f preserves (resp. reverses) orientation in a sufficiently small neighborhood of the point P, then we set

$$\deg_P f := 1 \quad (\text{resp. } \deg_P f = -1).$$

- Global mapping degree: If a map $f : A \to B$ globally preserves (resp. reverses) orientation, then we set

$$\deg f := 1 \quad (\text{resp. } \deg f = -1).$$

  If this is not the case, then we choose an image point P' and define

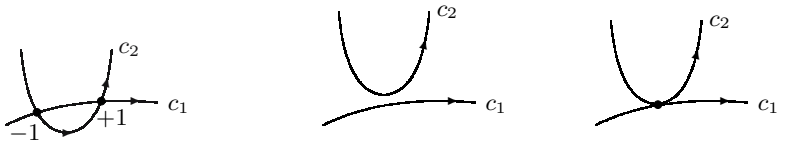
$$\boxed{\deg f := \deg_{P_1} f + ... + \deg_{P_n} f} \tag{5.1}$$

  where the points $P_1, ..., P_n$ are precisely the preimage points of P'. Mnemonically, the global mapping degree is the sum of local mapping degrees with respect to a fixed image point P'. It turns out that, roughly speaking, this definition does not depend on the choice of the image point P'. In the plane, winding number and mapping degree coincide.

The strategy of putting in general position. The definition given above for the intersection number of two curves refers tacitly to transversal intersection. This excludes the degenerate case where the two curves touch each other (Fig. 5.8(c)). In the degenerate case, we proceed in the following quite natural manner: We slightly perturb the curves such that

(a) either they intersect transversally each other (Fig. 5.8(a)) or
(b) they do not intersect at all (Fig. 5.8(b)).

It is now crucial that in both perturbed regular cases, we obtain the same well-defined intersection number, namely, zero. Therefore, it is reasonable to assign the intersection number zero to the degenerate situation of Fig. 5.8(c). This is part of a general strategy used in modern differential topology. This strategy reads as follows:



(a) $i(c_1, c_2) = -1 + 1 = 0$ (b) $i(c_1, c_2) = 0$ (c) $i(c_1, c_2) = 0$

Fig. 5.8. General position

Put geometric objects in general position by sufficiently small perturbations, and assign integers to the general position configuration.

It turns out that general position can always be formulated in terms of transversality. The point is that

Transversality is generic.

This means that the transversal situation can always be obtained by sufficiently small perturbations. Finally, one has to show that the integer assigned to the transversal situation is stable, that is, it does not depend on the specific form of the small perturbation, as in Fig. 5.8. The prototype for the success of this strategy is the following simple variant of the Sard theorem. Consider the equation

$$\boxed{f(x) = y_0, \qquad x \in \mathbb{R}.} \tag{5.2}$$

We are given the real number y_0. Let $f : \mathbb{R} \to \mathbb{R}$ be a smooth function such that

$$\lim_{|x| \to \infty} |f(x)| = \infty.$$

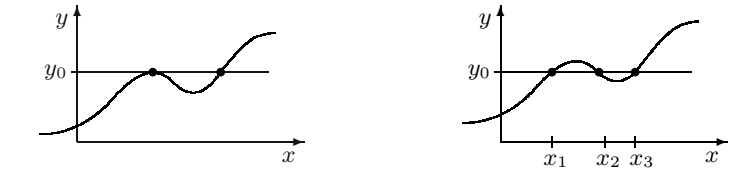
Then, there exist a sequence (ε_n) with $\varepsilon_n \to 0$ such that, for each index n, the perturbed equation

$$f(x) = y_0 + \varepsilon_n, \qquad x \in \mathbb{R}$$

has at most a finite number of solutions $x_1, ..., x_N$, and each solution is nondegenerate, that is,

$$f'(x_j) \neq 0 \qquad \text{for all} \quad j.$$

In terms of geometry, this means that the graph of f transversally intersects the



(a) (b)

Fig. 5.9. Sard's theorem

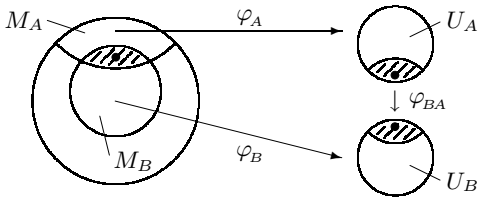


Fig. 5.10. Surface of earth

horizontal line $y = y_0$ (Fig. 5.9).[4] The importance of transversality and genericity for differential topology was emphasized by René Thom in the 1950s.[5]

5.4 Manifolds and Diffeomorphisms

The theory of manifolds studies smooth maps between geometric objects. Roughly speaking, manifolds are smooth curves and smooth surfaces along with higher-dimensional generalizations (e.g., m-dimensional spheres or Lie groups).

Intuitive motivation. The prototype of a real 2-dimensional manifold is the surface of earth, M. In geography, we describe parts of the earth locally by charts (subsets of $\mathbb{R}^2$) of an atlas (Fig. 5.10). Here, the same city can be located on different charts. Therefore, we need transition maps between charts. The following two properties are typical:

- the surface M is locally described by local coordinates living in open subsets of the 2-dimensional Euclidean space $\mathbb{R}^2$, and
- the change of local coordinates is carried out by smooth functions.

Generally, a real m-dimensional manifold M looks locally like an open subset of $\mathbb{R}^m$, and the change of local coordinates is done by diffeomorphisms. To explain this, let U and V be two open subsets of $\mathbb{R}^m$. The map

$$f : U \to V \tag{5.3}$$

is called smooth iff it is continuous and its components have continuous partial derivatives of each order. The smooth map (5.3) is called a diffeomorphism iff it is bijective and smooth, and the inverse map $f^{-1} : V \to U$ is smooth as well. For example, set $f(x) := \mathrm{e}^x$. Then, the map

$$f : \mathbb{R} \to \ ]0, \infty[$$

is a diffeomorphism. The inverse map reads as $f^{-1}(x) := \ln x$ for all $x \in]0, \infty[$.

The definitions can be immediately generalized to the situation where U and V are open subsets of the complex plane, $\mathbb{C}$, or more generally, open subsets of the m-dimensional complex space, $\mathbb{C}^m$. In particular, the map $f : U \to V$ is a

[4] The proof along with far-reaching generalizations to infinite-dimensional spaces due to Smale can be found in Zeidler (1986), Vol. I, Sect. 4.18. As an introduction to differential topology based on the transversality strategy, we recommend Guillemin and Pollack (1974).

[5] In 1958 Thom was awarded the Fields medal for his pioneering contributions to the global theory of manifolds (cobordism theory).

diffeomorphism iff it is bijective and the maps f and f^{-1} are smooth. In the special case where $m = 1$, diffeomorphisms coincide with conformal maps.

Definition of real manifolds. The precise definition of a manifold reads like this (Fig. 5.10).

(M1) A real m-dimensional manifold is a set M along with a (finite or countable) family of bijective maps[6]

$$\varphi_A : M_A \to U_A. \tag{5.4}$$

Here, M_A is a subset of M, and U_A is a nonempty open subset of $\mathbb{R}^m$. We call $x_A = \varphi_A(x)$ the local coordinate of the point $x \in M$, and φ_A is called a chart map.

(M2) If the point x lies both in M_A and M_B, then there are assigned the two local coordinates

$$x_A = \varphi_A(x) \qquad \text{and} \qquad x_B = \varphi_B(x)$$

to the point x. In this case, we assume that the transition map

$$\boxed{x_B = \varphi_{BA}(x_A)}$$

is a diffeomorphism on its natural domain of definition. Explicitly, we have $\varphi_{BA} = \varphi_B \circ \varphi_A^{-1}$ on the set $\varphi_A(M_A \cap M_B)$.[7]

The manifold is called oriented iff all of the transition maps preserve orientation.

Morphisms. Let M and N be manifolds. By a manifold morphism, we understand a smooth map

$$f : M \to N,$$

that is, this map is smooth with respect to local coordinates. Furthermore, the smooth map $f : M \to N$ is called a manifold isomorphism iff it is bijective and the inverse map $f^{-1} : N \to M$ is also smooth. Manifold isomorphisms are also called diffeomorphisms.

Complex manifolds. Similarly, the notion of complex m-dimensional manifold is obtained by replacing open subsets of the real m-dimensional space $\mathbb{R}^m$ by open subsets of the complex m-dimensional space $\mathbb{C}^m$. Complex 1-dimensional manifolds are also called regular complex curves. By a Riemann surface, we understand an arcwise connected, regular complex curve. Diffeomorphic Riemann surfaces are also called conformally equivalent.

Convention. If we do not explicitly state otherwise, by a manifold we always understand a real finite-dimensional manifold.

5.5 Topological Spaces, Homeomorphisms, and Deformations

Topology studies continuous maps between geometric objects. The notion of topological space was introduced by Hausdorff in 1914.[8] Topological spaces generalize manifolds. For example, the graph of the curve

[6] The index A lies in the set $\mathcal{A}$. We assume that the index set $\mathcal{A}$ is either finite or equal to the set of natural numbers. Naturally enough, suppose that each point of M lies in some set M_A.

[7] Naturally enough, we assume that the set $\varphi_A(M_A \cap M_B)$ is an open subset of the chart space $\mathbb{R}^m$.

[8] F. Hausdorff, Foundation of Set Theory (in German), Teubner, Leipzig, 1914.

$$y = x^{2/3}, \qquad x \in \mathbb{R}$$

is a topological space, but not a manifold because of the singularity (cusp) at the point $(0, 0)$. For fixed genus $g \geq 2$, spaces of string states (i.e., the Riemann moduli spaces) are topological spaces, but not manifolds because of the appearance of singularities. The prototype of a topological space is a subset of $\mathbb{R}^n$ with $n = 1, 2, \ldots$ or a point. Roughly speaking, we pass from manifolds to topological spaces by replacing 'smooth' by 'continuous.'

Definition. A set X is called a topological space iff certain subsets S of X are distinguished as *open sets* such that the following hold:

(T1) Both the set X and the empty set $\emptyset$ are open.
(T2) The union of each arbitrary family of open sets is open.
(T3) The intersection of each finite family of open sets is open.

The family of all open sets is called a topology. By an open open neighborhood of the point x in X, we understand an open subset of X which contains the point x.[9] The topological space X is called separated iff for any two different points $x, y \in X$, there exist disjoint open neighborhoods of x and y. A subset C of the topological space X is called closed iff the complement $X \setminus \mathbb{C}$ is open.

Each subset M of a topological space X is also a topological space.

By definition, a subset V of M is called open iff there exists an open subset U of X such that

$$V = U \cap M.$$

The topology of M is called the topology induced by X.

Morphisms. Let X and Y be topological spaces. The map

$$f : X \to Y$$

is called continuous (or a topological morphism) iff the preimage of open sets is again open. The continuous map $f : X \to Y$ is called a topological isomorphism iff it is bijective and the inverse map $f^{-1} : Y \to X$ is continuous, too. Topological isomorphisms are also called homeomorphisms.

Standard examples. In order to give the reader a feeling for the universality of the notion of topological space, let us consider the following examples.

- For $n = 0, 1, \ldots$, the sets $\mathbb{R}^n$ and $\mathbb{C}^n$ and their subsets are topological spaces.
- Every Hilbert space and Banach space and their subsets are topological spaces.
- Every real (resp. complex) finite-dimensional manifold is a topological space.
- For $n = 1, 2, \ldots$, the n-dimensional unit sphere

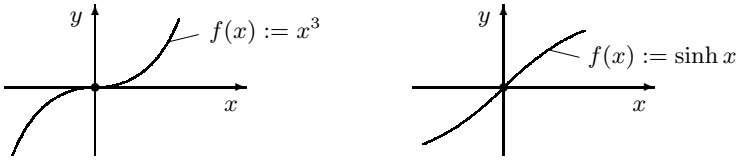
$$\mathbb{S}^n := \{x \in \mathbb{R}^{n+1} : x_1^2 + x_2^2 + \ldots + x_{n+1}^2 = 1\}$$

  is an n-dimensional real manifold.
- Define $F(\varphi) := \mathrm{e}^{\mathrm{i}\varphi}$ for all $\varphi \in \mathbb{R}$. The continuous map

$$F : \mathbb{R} \to \mathbb{S}^1$$

  sends the real line onto the unit circle $\mathbb{S}^1$. This is not a homeomorphism. We say that the real line $\mathbb{R}$ is a covering space of the unit circle.

[9] More generally, by a neighborhood of the point x, we understand an arbitrary set which contains an open neighborhood of the point x.



(a) homeomorphism (b) diffeomorphism

Fig. 5.11. Global behavior of a function

- Consider the interval $[-\pi, \pi]$ and identify the point $-\pi$ with π. This way, we obtain a topological space denoted by $[-\pi, \pi]/\{-\pi, \pi\}$ which is homeomorphic to the unit circle $\mathbb{S}^1$. Explicitly, the homeomorphism is given by the map

$$h : [-\pi, \pi]/\{-\pi, \pi\} \to \mathbb{S}^1$$

  where $h(\varphi) := \mathrm{e}^{\mathrm{i}\varphi}$ for all $\varphi \in [-\pi, \pi]$.
- The n-dimensional real projective space $\mathbb{P}^n$ is an n-dimensional real manifold obtained by considering antipodal pairs as one point. That is, the elements of $\mathbb{P}^n$ are antipodal pairs $\{x, -x\}$.
- For $n = 1, 2, \ldots$. the groups $SU(n), U(n), SO(n)$ are manifolds. More general, each Lie group is a real manifold.
- For the rotation group of the 3-dimensional Euclidean space,

$$SO(3) = \{(\mathbf{n}, \varphi) : \quad \mathbf{n} \in \mathbb{P}^2, \quad \varphi \in \mathbb{S}^1\}.$$

  Here, the unit vector $\mathbf{n}$ represents the rotation axis, and φ represents the rotation angle. Note that each rotation about the axis $\mathbf{n}$ can also be described by a rotation about the axis $-\mathbf{n}$. We briefly write

$$SO(3) = \mathbb{P}^2 \times \mathbb{S}^1.$$

  More precisely, the manifold $SO(3)$ is diffeomorphic to the manifold $\mathbb{P}^2 \times \mathbb{S}^1$. This shows that the Lie group $SO(3)$ possesses a nontrivial topological structure. We will show in Sect. 5.7.1 on page 265 that the topology of $SO(3)$ is responsible for the electron spin.
- Riemann surfaces are 2-dimensional real manifolds and 1-dimensional complex manifolds.

Each diffeomorphism is a homeomorphism, but the converse is not true. To explain this, consider the two maps

$$f(x) := x^3, \qquad g(x) := \sinh x \qquad \text{for all} \quad x \in \mathbb{R}$$

pictured in Fig. 5.11. Then, the map $g : \mathbb{R} \to \mathbb{R}$ is a diffeomorphism, whereas $f : \mathbb{R} \to \mathbb{R}$ is a homeomorphism, but not a diffeomorphism. In fact, the inverse map $f^{-1}(x) := x^{1/3}$ is not smooth at the point $x = 0$.

Deformations. We want to study the deformation of both geometric objects (topological spaces) and continuous maps. We expect that

> *Crucial phenomena in nature should be invariant under continuous deformations.*

The prototypes of deformations are given by the following two maps

(i) $H(x,t) := (1-t)x$ for all $x \in \mathbb{R},\ t \in [0,1]$, and
(ii) $H(x,t) := (1-t)f(x) + tg(t)$ for all $x \in \mathbb{R},\ t \in [0,1]$.

We regard the variable t as time. The map (i) contracts the real line into the origin during the time interval $[0,1]$. The trajectory

$$x(t) := (1-t)x, \qquad t \in [0,1]$$

sends the point x at the initial time $t = 0$ to the origin $x = 0$ at the final time $t = 1$. If the functions $f, g : \mathbb{R} \to \mathbb{R}$ are continuous, then the map (ii) continuously deforms the function f into g. This motivates the following definitions.

- Homotopic continuous maps: Let X and Y be topologically spaces. Two continuous maps $f, g : X \to Y$ are called homotopic iff there exists a continuous map
$$H : X \times [0,1] \to Y$$
such that $H(x,0) = f(x)$ and $H(x,1) = g(x)$ for all $x \in X$. Intuitively, the map f is continuously deformed into the map g during the time interval $[0,1]$. The map $f : X \to Y$ is called homotopically trivial iff it is homotopic to a constant map $g : X \to Y$.
- Homotopically equivalent topological spaces: The two topological spaces X and Y are called homotopically equivalent iff there exists a map
$$f : X \to Y$$
which is invertible up to deformations. We briefly write $X \sim Y$. Explicitly, this means the following. There exists a map
$$g : Y \to X$$
such that $f \circ g : Y \to Y$ is homotopic to the identity map on Y, and $g \circ f : X \to X$ is homotopic to the identity map on X. We briefly write
$$f \circ g \sim \mathrm{id}_Y \qquad \text{and} \qquad g \circ f \sim \mathrm{id}_X.$$
If $f : X \to Y$ is a homeomorphism, then the topological space X is called topologically equivalent to the topological space Y. This implies that X is homotopically equivalent to Y. To see this, choose $g := f^{-1}$.

Standard examples. We want to consider contractible topological spaces and deformation retracts.

- A topological space is called contractible iff it is homotopically equivalent to a one-point space.[10] Explicitly, this means the following. Let x_0 be a point of the topological space X. Suppose that there exists a continuous map
$$H : X \times [0,1] \to X$$
such that $H(x,0) = x$ and $H(x,1) = x_0$ for all $x \in X$. Then, the space X is contractible to the point x_0.
- Retract: Let Y be a subset of a the topological space X. The continuous map
$$r : X \to Y$$
is called a retract iff it is continuous and it fixes the points of Y, that is, $r(y) = y$ for all $y \in Y$.

[10] Contractible topological spaces are also called homotopically trivial.

(a) simply connected disc (b) not simply connected annulus

Fig. 5.12. Connectivity

- Deformation retract: If, in addition, the retraction r is homotopic to the identity map on X, then Y is called a deformation retract of X. Explicitly, this means that there exists a continuous map

$$H : X \times [0,1] \to X$$

such that $H(x,0) = x$ for all $x \in X$ and $H(y,1) = y$ for all $y \in Y$.

For example, the real line and each interval is contractible, but the unit sphere $\mathbb{S}^1$ is not contractible. Furthermore, for $n = 2, 3, \ldots$, each n-dimensional ball is contractible. Let $x, y \in \mathbb{R}^2$. Then, the map

$$(x, y) \mapsto (x, 0)$$

is a retraction, and the x-axis is a deformation retract of the (x, y)-plane.

Topological invariants. Suppose that we assign a mathematical object to a class of topological spaces.

- This object is called a topological invariant iff it is invariant under homeomorphisms.
- This object is called a homotopy invariant iff it is the same for homotopically equivalent topological spaces.

The most important topological invariant is the Euler characteristic to be introduced below. Topological invariants play a fundamental role in mathematics and physics (e.g., Betti numbers, Gauss' linking number, genus of a Riemannian surface, winding number, mapping degree, Morse index of a dynamical system, characteristic classes, Stiefel–Whitney classes, Chern classes and Chern numbers, Atiyah–Singer index, Gromov–Witten invariants, Donaldson invariants of 4-dimensional manifolds, Seiberg–Witten invariants, Jones polynomials of knots, homology groups, cohomology groups, homotopy groups, K-groups as generalized cohomology groups). We will study this thoroughly in Volume IV on quantum mathematics. At this point, we will only sketch same basic ideas.

As a rule, topological invariants are also homotopy invariants.

Compactness. A topological space X is called compact iff each covering of X by open sets contains a finite family of open subsets which already covers the space X. For example, a subset of $\mathbb{R}^n, n = 1, 2, \ldots$ is a compact topological space iff it is closed and bounded (e.g., closed balls or spheres).

Connectedness. The topological space X is called arcwise connected iff for any two points x_0, x_1 in X there exists a continuous map

$$x : [0,1] \to X$$

such that $x(0) = x_0$ and $x(1) = x_1$. Intuitively, the continuous curve $x = x(t)$ with $0 \leq t \leq 1$ connects the point x_0 with the point x_1. To each point x_0 in X, there

exists a maximal arcwise connected subset of X which contains the point x_0. This set is called the component of the point x_0 in the topological space X. For example, the component of the point $x = 1$ on the punctured real line $\mathbb{R} \setminus \{0\}$ is equal to the interval $]0, \infty[$.

By a loop in X, we understand a continuous map $x : [0,1] \to X$ with the property $x(0) = x(1)$. The topological space X is called simply connected iff each loop in X can be continuously contracted into a point of X. Explicitly, there exists a continuous map $(t,s) \mapsto H(t,s)$,

$$H : [0,1] \times [0,1] \to X,$$

and a point x_0 such that $H(t,0) = x(t)$ and $H(t,1) = x_0$ for all $t \in [0,1]$. The unit circle $\mathbb{S}^1$ is not simply connected. Any disc in $\mathbb{R}^2$ is simply connected, but an annulus is not simply connected (Fig. 5.12). The 2-dimensional sphere $\mathbb{S}^2$ is simply connected, but a torus is not simply connected.

The Jordan curve theorem. A topological space is called a closed Jordan curve iff it is homeomorphic to the unit circle $\mathbb{S}^1$. Obviously, the topological space

$$\mathbb{R}^2 \setminus \mathbb{S}^1$$

consists of two components, namely, the interior and the exterior of the unit circle. In 1887 Camille Jordan (1838-1922) proved the much deeper result that for any closed Jordan curve C living in $\mathbb{R}^2$, the topological space

$$\mathbb{R}^2 \setminus C$$

consists of two components; one component is bounded and the other component is unbounded. These two components are called the interior and the exterior of the curve C, respectively.[11]

Suggested reading. For first reading, we recommmend the elegant book by Guillemin and Pollack (1974) on elementary differential topology, which emphasizes geometric intuition. We also refer to the beautiful modern textbook by Jost (2002a) (Riemannian geometry, spin structures and the Dirac operator, geometric analysis, de Rham cohomology, geodesics, Morse theory, Floer homology, harmonic maps, Kähler manifolds, Chern classes, variational problems from quantum field theory, the Yang–Mills functional, the Chern–Simons functional, the Ginzburg–Landau functional, and the Seiberg–Witten functional).

As an introduction to topology and its applications to physics, we recommend the following books: Milnor (1963) (Morse theory), (1965) (differential topology), Pontryagin (1965) and Spanier (1989) (classic textbooks), Hirzebruch (1966) (classic monograph on topology and algebraic geometry), Milnor and Stasheff (1974) (characteristic classes), Naber (1980), (1982), (1997) (applications to physics, e.g., singularities of our universe and gauge theory), Bott and Tu (1982) (differential forms and characteristic classes), Dubrovin, Fomenko, and Novikov (1992) (geometry, topology, and physics), Marathe and Martucci (1992) (gauge field theory), Bredon (1993) (topology and geometry), Schwarz (1993), (1994) (topology and quantum field theory), Gilkey (1995) (Atiyah–Singer index theorem and spectral geometry), Knörrer (1995) (the beauty of geometry), Jost (1997) (Riemann surfaces), Friedrich (2000) (spin geometry and Dirac operators), Hatcher (2002), (2005) (modern textbook).

[11] The proof based on the mapping degree and the generalization to higher dimensions (the Jordan–Brouwer separation theorem) can be found in Zeidler (1986), Vol. I, Sect. 13.8. The Appendix of Zeidler (1986), Vol. I, contains a summary of important notions and theorems from topology.

General surveys on topology can be found in Novikov (1996), Dodson and Parker (1997), and Zeidler (2002a), Vol. II. For the history of algebraic topology and algebraic geometry, we recommend Dieudonné (1978), (1985), (1989).

5.6 Topological Quantum Numbers

In the 20th century, physicists learned that quantum phenomena in nature can be classified by quantum numbers. There arises the question how to describe quantum numbers in terms of mathematics. It turns out that there are two important possibilities to obtain quantum numbers , namely,

(S) symmetry (the representation theory of compact Lie groups), and
(T) topology (topological invariants as topological charges or topological quantum numbers).

In what follows, we will sketch some basic ideas by emphasizing the relations to physical applications.

5.6.1 The Genus of a Surface

Classification of real 1-dimensional manifolds. Let M be a real, 1-dimensional, arcwise connected manifold.

- If M is compact, then it is diffeomorphic to the unit circle $\mathbb{S}^1$.
- If M is not compact, then it is diffeomorphic to the real line $\mathbb{R}$.

The proof can be found in Milnor (1965) (appendix).

Classification of real 2-dimensional manifolds. The following theorem was one of the highlights of topology at the end of the 19th century:[12]

> *Each real, 2-dimensional, compact, oriented, arcwise connected manifold is homeomorphic to a surface that is obtained by attaching a finite number of handles to the 2-dimensional sphere* $\mathbb{S}^2$.

The number of handles, $g = 0, 1, 2, \ldots$, is called the genus of the manifold. For example, the 2-dimensional sphere $\mathbb{S}^2$ has the genus $g = 0$. The torus is homeomorphic to a sphere with one handle (Fig. 5.13(a)). Therefore, the genus of a torus is equal to 1. Equivalently, the genus is the number of 'holes' (Fig. 5.13(b)). The genus is a topological invariant. More precisely, two real, 2-dimensional, compact, oriented, arcwise connected manifolds are homeomorphic iff they have the same genus. This means that the crucial qualitative structure of such manifolds only depends on the genus of the manifold. Below we will illuminate this by considering

[12] The proof of the more general result on the classification of real, 2-dimensional, oriented and non-orientable, compact, arcwise connected topological manifolds can be found in W. Rinow, Topologie (in German), Sect. 50, Verlag der Wissenschaften, Berlin, 1975.
The complete proof was given first by Dehn and Heegard in 1907 based on earlier work by Möbius, Riemann, Betti, Felix Klein, Poincaré, and van Dyck in the 19th century. We refer to E. Scholz, History of the Theory of Manifolds from Riemann to Poincaré (in German), Birkhäuser, Basel, 1980.
The classification of non-compact, real, 2-dimensional topological manifolds can be found in B. von Kerékjártó, Vorlesungen über Topologie (Lectures on topology) (in German), Springer, Berlin, 1923.

(a) $g = 1$ (b) $g = 2$

Fig. 5.13. The genus g of an oriented surface

- the Poincaré–Hopf index theorem on stationary points of dynamical systems,
- the Gauss–Bonnet theorem on the total curvature of surfaces, and
- the Morse theorem on critical points of energy functions on surfaces.

In addition, the crucial properties of Abelian integrals only depend on the genus g of the Riemann surface of the corresponding algebraic function. In particular, the number of additive periods of an Abelian integral is equal to $2g$. In particular, the Riemann surface of an elliptic integral is a torus with genus $g = 1$. This is responsible for the fact that elliptic integrals possess two additive periods and the inverse functions to elliptic integrals are double-periodic (i.e., they are elliptic functions).

5.6.2 The Euler Characteristic

We will decompose geometric objects into cells. This allows us to compute the Euler characteristic as an alternating sum of the numbers of cells of increasing dimension. Let us start with the following definitions.

- A 0-cell is a point.
- A 1-cell is a topological space which is homeomorphic to the open interval $]0, 1[$.
- A 2-cell is a topological space which is homeomorphic to the open unit square $\{(x, y) \in \mathbb{R}^2 : 0 < x, y < 1\}$.
- For $n = 1, 2, \ldots$, an n-cell is a topological space which is homeomorphic to the open n-dimensional cube $\{x \in \mathbb{R}^n : 0 < x_1, \ldots, x_n < 1\}$.

We define the Euler characteristic of the topological space X by

$$\boxed{\chi(X) := c_0 - c_1 + c_2 - \ldots} \tag{5.5}$$

Here, $c_0, c_1, \ldots$ is the number of 0-cells, 1-cells, ..., respectively. This definition seems to depend on the decomposition of the space X into cells. The point is that the Euler characteristic of X does not depend on the choice of the decomposition.

> *It turns out that the Euler characteristic of homotopically equivalent topological spaces is the same.*

This is a deep result of modern topology.

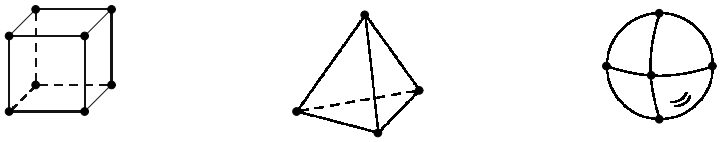
Prototypes of the Euler characteristic χ. To get some feeling for the Euler characteristic, let us consider a few simple examples.

(i) For a point, $\chi = c_0 = 1$.
(ii) Compact interval $[a, b]$. The interval $[a, b]$ consists of the two boundary points a, b and the open interval $]a, b[$. Hence

$$\chi = c_0 - c_1 = 2 - 1 = 1.$$

(a) $\chi = 2 - 2 = 0$ (b) $\chi = 3 - 3 = 0$

Fig. 5.14. Euler characteristic of a circle



(a) $\chi = 8 - 12 + 6 = 2$ (b) $\chi = 4 - 6 + 4 = 2$ (c) $\chi = 6 - 12 + 8 = 2$

Fig. 5.15. Euler characteristic of a sphere

The point is that the Euler characteristic of $[a, b]$ does not depend on the choice of any decomposition of the interval. In fact, if we consider the decomposition

$$a < x_1 < \ldots < x_n < b,$$

we have $n + 2$ points $a, x_1, \ldots, b$ and $n + 1$ open intervals $]a, x_1[, \ldots,]x_n, b[$. Hence

$$\chi = c_0 - c_1 = n + 2 - (n + 1) = 1.$$

(iii) Unit circle. For $n = 2, 3, \ldots$, every decomposition of the unit sphere $\mathbb{S}^1$ consists of n points and n 1-cells (Fig. 5.14). Hence

$$\chi(\mathbb{S}^1) = c_0 - c_1 = n - n = 0.$$

(iv) 2-dimensional unit sphere. The sphere $\mathbb{S}^2$ can be decomposed into one point (e.g., the North Pole) and the complement $\mathbb{S}^2 \setminus \{N\}$ which is a 2-cell. Hence

$$\chi(\mathbb{S}^2) = c_0 - c_1 + c_2 = 1 - 0 + 1 = 2.$$

Consider now a triangulation of $\mathbb{S}^2$ as pictured in Fig. 5.15(c). This triangulation consists of six vertices, twelve 1-cells, and eight 2-cells. Hence

$$\chi(\mathbb{S}^2) = c_0 - c_1 + c_2 = 6 - 12 + 8 = 2.$$

(v) Surface of a cube. By 5.15(b), there are eight vertices, twelve 1-cells, and six 2-cells. Hence

$$\chi = c_0 - c_1 + c_2 = 8 - 12 + 6 = 2.$$

We expect this, since the surface of the cube is homeomorphic to the 2-dimensional sphere. For the solid cube, we get

$$\chi = c_0 - c_1 + c_2 - c_3 = 8 - 12 + 6 - 1 = 1.$$

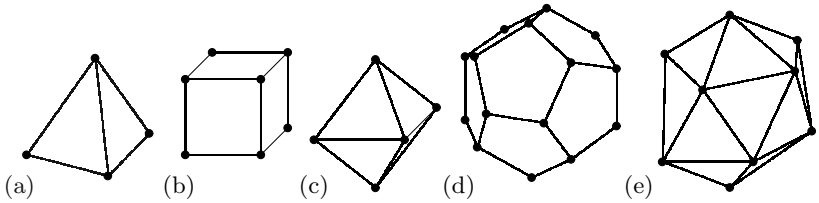


Fig. 5.16. Tetrahedron, cube, octahedron, dodecahedron, icosahedron

The following are true.

- The Euler characteristic of a contractible space is equal to one.
- The Euler characteristic of a real, 2-dimensional, compact, oriented, arcwise connected manifold of genus $g = 0, 1, 2, \ldots$ is equal to $\chi = 2 - 2g$. This generalizes (iv). In particular, for the torus we have $g = 1$, and hence we get $\chi = 2 - 2 = 0$.

5.6.3 Platonic Solids and Fullerenes

> In 387 B.C. Plato founded in Athens his Academy, which in most respects was like a modern university... Though the main center shifted to Alexandria around 300 B.C., the Academy remained preeminent in philosophy throughout the Alexandrian period. It lasted nine hundred years until it was closed by the Christian emperor Justinian in A.D. 529 because it taught "pagan and perverse learning"... Plato, one of the most informed men of his day, was not a mathematician; but his enthusiasm for the subject and his belief in its importance for philosophy and for the understanding of the universe encouraged mathematicians to pursue it. It is noteworthy that almost all of the important mathematical work of the fourth century was done by Plato's friends and pupils. Plato himself seems to have been more concerned to improve and perfect what was known.
>
> Morris Kline, 1972
> *Mathematical Thought from Ancient to Modern Times*[13]

The Greeks were charmed by the fact that there exist precisely five regular polyhedra called the Platonic solids: tetrahedron, cube, octahedron, dodecahedron, and icosahedron (Fig. 5.16).

- The number of faces is equal to 4 (triangles), 6 (squares), 8 (triangles), 12 (pentagons), 20 (triangles), respectively. These numbers are responsible for the designation of the Platonic solids.
- The number of vertices is equal to 4, 8, 6, 20, 12, respectively.
- The number of edges is equal to 6, 12, 12, 30, 30, respectively.

By definition, a regular polyhedron is a polyhedron whose faces all have the same number of sides, and which also has the same number of faces meeting at each vertex. It can be proven that the Platonic solids are the only regular polyhedra homeomorphic to the 2-sphere. In 1750 Euler discovered that there holds

$$\boxed{c_0 - c_1 + c_2 = 2}$$

[13] Reprinted by permission of Oxford University Press.

for the Platonic polyhedra. Here, c_0 is the number of vertices, c_1 is the number of edges, and c_2 is the number of faces. This formula follows from the fact that the Euler characteristic of the 2-dimensional sphere is equal to 2, and the surfaces of all the Platonic solids are homeomorphic to the 2-dimensional sphere. Felix Klein wrote a beautiful book on the icosahedron[14] and its symmetry group of order 60. He used the symmetry group of the icosahedron in order to solve fifth-order algebraic equations by means of automorphic functions. If the polyhedron has g holes, then

$$\boxed{\chi = c_0 - c_1 + c_2 = 2 - 2g.}$$

This was discovered by Simon l'Huilier in 1815. The number g is called the genus of the polyhedron. In particular, for a torus-like polyhedron, we get $g = 1$.

In 1985, a new class of molecules was experimentally discovered. These molecules of carbon have a roughly spherical shape. They are called fullerenes. For example, the sixty carbon atoms of the C_{60}-molecule form a polyhedron which has 60 vertices and 32 faces (20 six-angles and 12 five-angles). By Euler's polyhedron formula, the number of edges is equal to

$$c_1 = c_0 + c_2 - 2 = 60 + 32 - 2 = 90.$$

A soccer ball has the same shape. For the discovery of fullerenes, Sir Harold Kroto, Robert Curl, and Richard Smalley were awarded the 1996 Nobel Prize in chemistry. These molecules are named after the American architect Robert Buckminster Fuller.

5.6.4 The Poincaré–Hopf Theorem for Velocity Fields

We want to show that the topology of a manifold restricts the structure of velocity vector fields on the manifold. Consider a velocity vector field

$$\mathbf{v}(\mathbf{x}) = v(x, y)\mathbf{i} + w(x, y)\mathbf{j}$$

on the Euclidean plane. Suppose that $\mathbf{v}(\mathbf{x}_0) = 0$ and that the stationary point $\mathbf{x}_0$ is regular, that is, we define

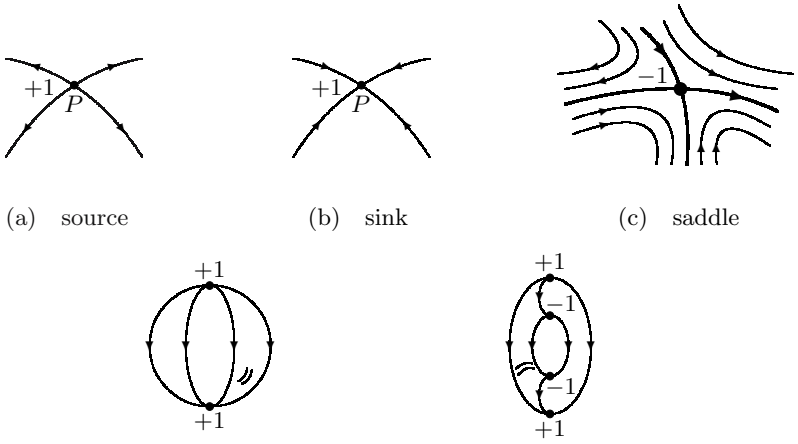
$$\det \mathbf{v}'(\mathbf{x}_0) := \begin{vmatrix} v_x(x_0, y_0) & v_y(x_0, y_0) \\ w_x(x_0, y_0) & w_y(x_0, y_0) \end{vmatrix},$$

and we assume that $\det \mathbf{v}'(\mathbf{x}_0) \neq 0$. The number $\operatorname{ind} \mathbf{v}(\mathbf{x}_0) := \operatorname{sgn} \det \mathbf{v}'(\mathbf{x}_0)$ is called the index of the stationary point $\mathbf{x}_0$. Some typical situations are pictured in Fig. 5.17. For example, if

$$v(x, y) := \lambda x, \qquad w(x, y) := \mu y$$

with real nonzero numbers λ and μ, then $\operatorname{ind} \mathbf{v}'(0) = \operatorname{sgn}(\lambda\mu)$. In particular, the index of a sink (or source) is equal to $+1$, whereas the index of a saddle is equal to -1. Consider a smooth velocity vector field on a real, 2-dimensional, compact, arcwise connected manifold (e.g., a sphere) which has only a finite number of stationary points $\mathbf{x}_1, ..., \mathbf{x}_N$. In addition, suppose that all of the stationary points are regular. Then, the Poincaré–Hopf theorem tells us that

[14] Vorlesungen über das Ikosaeder, Teubner, Leipzig, 1884. (English translation: Lectures on the Icosahedron, Dover, 1956.)



(a) source (b) sink (c) saddle

(d) $\chi = 1 + 1 = 2$ (e) $\chi = 1 - 1 - 1 + 1 = 0$

Fig. 5.17. Velocity vector field of a fluid

$$\boxed{\sum_{j=1}^{N} \operatorname{ind} \mathbf{v}'(\mathbf{x}_j) = \chi} \tag{5.6}$$

where χ is the Euler characteristic of the manifold.[15] Let us consider some examples.

- The velocity field pictured in Fig. 5.17(d) has one sink and one source, and the Euler characteristic of a sphere is equal to 2.
- Since the Euler characteristic of a sphere is different from zero, each smooth velocity vector field on a sphere has at least one stationary point (Poincaré's hairy ball theorem).
- The velocity field pictured in Fig. 5.17(e) has one source, one sink, and two saddles, and the Euler characteristic of a torus is equal to zero.

Conversely, the Poincaré–Hopf theorem can be used in order to compute the Euler characteristic of a manifold easily. To this end, one has to construct a simple flow of fluid particles on the manifold.

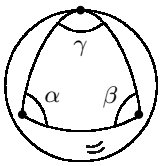
5.6.5 The Gauss–Bonnet Theorem

Consider the triangle pictured in Fig. 5.18(a) on a sphere $\mathbb{S}^2_R$ of radius R where the sides are geodesics. For the sum of angles,

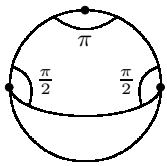
$$\alpha + \beta + \gamma - F = \pi.$$

Here, F is the surface area of the triangle. Introducing the Gaussian curvature of the sphere, $K := 1/R^2$, we get the formula

[15] The index refers to local coordinates. However, it turns out that this definition is independent of the choice of local coordinates.



(a) $\alpha + \beta + \gamma - F = \pi$



(b) $\frac{\pi}{2} + \frac{\pi}{2} + \pi - \pi = \pi$

Fig. 5.18. The Gauss–Bonnet theorem

$$\boxed{\alpha + \beta + \gamma - \int_{\mathcal{T}} K dS = \pi} \tag{5.7}$$

where we integrate over the triangle. The sphere $\mathbb{S}^2_R$ can be decomposed into four triangles as pictured in Fig. 5.18(b). Summing over the four triangles, we get

$$\int_{\mathbb{S}^2_R} K dS = 4 \cdot 2\pi - 4\pi = 4\pi.$$

Letting $M = \mathbb{S}^2_R$, this can be written as

$$\boxed{\frac{1}{2\pi} \int_M K dS = \chi(M)} \tag{5.8}$$

where $\chi(M) = 2$ is the Euler characteristic (or the first Chern number) of the sphere. The integral is to be understood as an integral with respect to the surface measure of the sphere. This Gauss–Bonnet formula is the prototype for a topological quantum number. It tells us that a topological quantity – the Euler characteristic – can be expressed by an analytical quantity – the Gaussian curvature. For a torus, equation (5.18) is valid with $K \equiv 0$ and $\chi = 0$.

The Gauss–Bonnet–Chern theorem. The Gauss–Bonnet theorem is one of the most beautiful theorems in mathematics. In Volume IV on quantum mathematics, we will consider far-reaching generalizations in the framework of the theory of characteristic classes: the Gauss–Bonnet–Chern theorem for n-dimensional Riemannian manifolds. In order to obtain such a generalization, one has to reformulate the classical Gauss–Bonnet theorem in the modern language of vector bundles. At this point, let us only sketch a few basic ideas.

Riemannian geometry of the 2-dimensional sphere. To begin with, let us introduce spherical coordinates

$$-\pi < \varphi \leq \pi, \qquad -\frac{\pi}{2} \leq \vartheta \leq \frac{\pi}{2}.$$

Here, φ and ϑ denote geographic longitude and geographic latitude, respectively. Moreover, we get the following:

- equator: $\vartheta = 0$;
- North Pole: $\vartheta = \frac{\pi}{2}$;
- South Pole: $\vartheta = -\frac{\pi}{2}$;
- meridian: $\varphi = \text{const}$;
- parallel of latitude: $\vartheta = \text{const}$.

In terms of Cartesian coordinates x, y, z, the sphere $\mathbb{S}^2_R$ can be parametrized in the following way:

$$x = R\cos\varphi\cos\vartheta, \qquad y = R\sin\varphi\cos\vartheta, \qquad z = R\sin\vartheta.$$

In fact, it follows from $\cos^2\alpha + \sin^2\alpha = 1$ that $x^2 + y^2 + z^2 = R^2$. Now consider a smooth curve

$$C : \varphi = \varphi(t), \qquad \vartheta = \vartheta(t), \qquad t_0 \le t \le t_1$$

on the sphere. In Cartesian coordinates,

$$x(t) = R\cos\varphi(t)\cos\vartheta(t), \quad y(t) = R\sin\varphi(t)\cos\vartheta(t), \quad z(t) = R\sin\vartheta(t).$$

Differentiation with respect to time yields

$$\dot{x}(t) = -R\dot{\varphi}(t)\sin\varphi(t)\cos\vartheta(t) - R\dot{\vartheta}(t)\cos\varphi(t)\sin\vartheta(t).$$

Similarly, we get $\dot{y}$ and $\dot{z}$. Using again $\cos^2\alpha + \sin^2\alpha = 1$,

$$\dot{s}(t) := \sqrt{\dot{x}^2(t) + \dot{y}^2(t) + \dot{z}^2(t)} = R\sqrt{\dot{\vartheta}^2(t) + \dot{\varphi}(t)^2\cos^2\vartheta(t)}.$$

This yields the arc length of the curve C,

$$s := \int_{t_0}^{t_1} \dot{s}(t)dt = \int_{t_0}^{t_1} R\sqrt{\dot{\vartheta}(t)^2 + \dot{\varphi}(t)^2\cos^2\vartheta(t)}\; dt.$$

Setting $u^1 := \varphi, u^2 := \vartheta$, we get

$$\dot{s}(t)^2 = g_{ij}(u^1(t), u^2(t))\dot{u}^i(t)\dot{u}^j(t)$$

where we sum over $i, j = 1, 2$. The functions

$$g_{11}(\varphi, \vartheta) := R^2\cos^2\vartheta(t), \qquad g_{22}(\varphi, \vartheta) := R^2, \qquad g_{12} = g_{21} = 0$$

are called the components of the metric tensor of the sphere. Set $g := \det(g_{ij})$. For the sphere, $g = g_{11}g_{22} = R^4\cos^2\vartheta$. The differential form

$$\upsilon := \sqrt{g}\; d\varphi \wedge d\vartheta = R^2\cos\vartheta\; d\varphi \wedge d\vartheta$$

is called the volume form of the sphere $\mathbb{S}^2_R$. The integral

$$\int_{\mathbb{S}^2_R} \upsilon = \int_{-\pi}^{\pi}\int_{-\pi/2}^{\pi/2} R^2\cos\vartheta\; d\varphi d\vartheta = 4\pi R^2$$

is equal to the surface area of the sphere.

In the 1820s, the importance of the metric tensor was first noticed by Gauss in the context of his surface theory. The theorema egregium (the beautiful theorem) of Gauss tells us that the curvature of a surface can be computed by means of the second partial derivatives of the functions g_{ij} (see Vol. III). In the 1850s, Riemann introduced the components R_{ijkl} of the Riemann curvature tensor for n-dimensional Riemannian manifolds. In the special case of the sphere, there is only one essential component of the Riemann curvature tensor, namely,

$$R_{1212} = Kg$$

where $K = 1/R^2$ is the Gaussian curvature of the sphere $\mathbb{S}^2_R$. In 1915 Einstein critically used the Riemann curvature tensor of the 4-dimensional space-time manifold

in order to describe the gravitational force in our universe in the framework of the theory of general relativity.

Velocity fields and the tangent bundle of the 2-dimensional sphere. For each point x of the sphere $\mathbb{S}^2_R$, let $T_x\mathbb{S}^2_R$ denote the tangent space at x. In terms of physics, this tangent space consists of all possible velocity vectors $\mathbf{v}$ at the point x. In order to get a global object, let us introduce the tangent bundle $T\mathbb{S}^2_R$ of the sphere $\mathbb{S}^2_R$ which consists of all possible pairs $(x, \mathbf{v})$ where $\mathbf{v}$ is a velocity vector at the point x. In other words,

$$T\mathbb{S}^2_R := \{(x, \mathbf{v}) : \ x \in \mathbb{S}^2_R, \ \mathbf{v} \in T_x\mathbb{S}^2_R\}.$$

One can show that this is a real 4-dimensional manifold. It turns out that the important geometric properties of the sphere like

- the parallel transport of velocity vectors and
- the curvature properties

can be formulated in terms of a connection on the tangent bundle. The basic idea is to compute the curvature of the sphere at a point by measuring the parallel transport of a velocity vector along a sufficiently small loop around the point. This concept is intrinsic, that is, it is independent of the surrounding space of the surface.

> *In order to best understand the geometry of the sphere $\mathbb{S}^2_R$, do not study the sphere itself, but its tangent bundle $T\mathbb{S}^2_R$.*

This strategy is crucial for modern differential geometry. Replacing the tangent bundle by more general bundles, we get a mathematical approach to differential geometry which is identical with gauge theory in physics. The two approaches were created independently by mathematicians (Gauss, Riemann, Lie, Levi-Civita, Élie Cartan, Ehresmann) and physicists (Yang, Mills) and completed in the 1950s.[16] In particular, the Standard Model in particle physics can be described this way. This will be studied in great detail in Volume III on gauge theory. We will see that

- connection (resp. curvature) in mathematics
- corresponds to potential (resp. interacting force) in physics.

The first Chern class of the tangent bundle of the sphere. By definition, the differential form

$$\omega = \frac{K}{2\pi} \cdot \upsilon$$

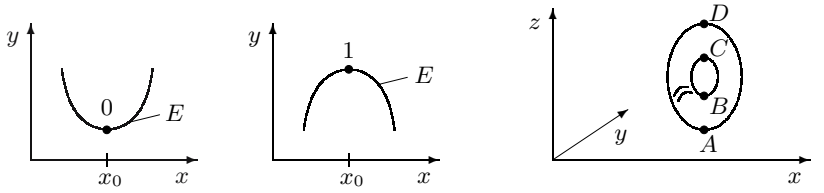
is a representative of the first Chern class $c_1(\mathbb{S}^2_R)$ of the tangent bundle $T\mathbb{S}^2_R$. The Gauss–Bonnet theorem (5.8) can equivalently be written as

$$\boxed{\int_{\mathbb{S}^2_R} \omega = \chi(\mathbb{S}^2_R).}$$

As we will show in Volume IV, the first Chern class is an element of the second de Rham cohomology group $H^2(\mathbb{S}^2_R)$ of the sphere $\mathbb{S}^2_R$. In 1945 Chern discovered how to generalize this to higher dimensions.[17]

[16] See S. Kobayashi, K. Nomizu, Foundations of Differential Geometry, Vols. 1, 2, Wiley, New York, 1963.

[17] S. Chern, A simple intrinsic proof of the Gauss–Bonnet formula for closed Riemannian manifolds, Ann. of Math. **45** (1945), 747–752.



(a) $E''(x_0) > 0$ (b) $E''(x_0) < 0$ (c) $i_A = 0,\ i_B = i_C = 1,\ i_D = 2$

Fig. 5.19. Morse index

5.6.6 The Morse Theorem on Critical Points of Energy Functions

The Morse index of an energy function. Suppose that the smooth real function $E : \mathbb{R} \to \mathbb{R}$ has a regular critical point at x_0, that is, $E'(x_0) = 0$ and $E''(x_0) \neq 0$. Then, the Morse index of x_0 is given by

$$i(x_0) := \begin{cases} 0 & \text{if} \quad E''(x_0) > 0, \\ 1 & \text{if} \quad E''(x_0) < 0. \end{cases}$$

This corresponds to a minimum (resp. maximum) of E at the point x_0 (Fig. 5.19). More generally, consider a smooth real function $E : \mathbb{R}^2 \to \mathbb{R}$. Then

$$E'(x,y) := (E_x(x,y), E_y(x,y)), \qquad E''(x,y) := \begin{pmatrix} E_{xx}(x,y) & E_{xy}(x,y) \\ E_{xy}(x,y) & E_{yy}(x,y) \end{pmatrix}.$$

Let (x_0, y_0) be a regular critical point of E, that is, $E'(x_0, y_0) = 0$ and

$$\det E''(x_0, y_0) \neq 0.$$

Then the Morse index $i(x_0, y_0)$ of (x_0, y_0) is defined to be the number of negative eigenvalues of the matrix $E''(x_0, y_0)$. For example, the function

$$f(x,y) := \lambda x^2 + \mu y^2$$

with the real nonzero numbers λ and μ has the origin $(0,0)$ as regular critical point with the Morse index

$$i(0,0) := \begin{cases} 0 & \text{if} \quad \lambda > 0, \mu > 0, \\ 1 & \text{if} \quad \lambda\mu < 0, \\ 2 & \text{if} \quad \lambda < 0, \mu < 0. \end{cases}$$

This corresponds to a minimum (resp. saddle point, maximum) of the function E at the point $(0,0)$.

The Morse theorem. Let $E : M \to \mathbb{R}$ be a smooth function on the real, 2-dimensional, compact, oriented, arcwise connected manifold M of genus g (e.g., a sphere or a torus). Suppose that the function E has only a finite number of critical

Fig. 5.20. Gauss' linking number

points, and all of them are regular. Let m_i be the number of critical points of E with Morse index i.[18] Then

$$\boxed{m_0 - m_1 + m_2 = \chi(M).}$$

In addition, we have the famous Morse inequalities[19]

$$m_0 \geq \beta_0, \qquad m_1 \geq \beta_1, \qquad m_2 \geq \beta_2.$$

The lower bounds $\beta_0 := 1$, $\beta_1 := 2g$, $\beta_2 := 1$ are the so-called Betti numbers of the manifold M. For the Euler characteristic of the manifold M, we get

$$\chi(M) = \beta_0 - \beta_1 + \beta_2 = 2 - 2g.$$

In particular, for a torus M, $g = 1$.[20] Thus, on a torus, the function E has at least one minimum, one maximum, and two saddle points. In addition, from $\chi = 0$ we get $m_0 + m_2 = m_1$, that is, the number of minima plus the number of maxima of E is equal to the number of saddle points.

For example, consider the function $z = h(x, y)$ on the torus where $h(x, y)$ denotes the height of the point (x, y) (Fig. 5.19(c)). The function h has a minimum (resp. maximum) at the point A (resp. D) with Morse index $i_A = 0$ (resp. $i_D = 2$). Furthermore, the function h has saddles at the points B, C with Morse index $i_B = i_C = 1$. For the Euler characteristic of the torus, we get

$$\chi = m_0 - m_1 + m_2 = 1 - 2 + 1 = 0.$$

5.6.7 Magnetic Fields, the Gauss Integral, and the Linking Number

> Topology is rooted in Maxwell's theory on the electromagnetic field.
> Folklore

[18] The Morse index refers to local coordinates. However, it is independent of the choice of local coordinates. The proof of the Morse theorem can be found in Jost (2002a), 2nd edn., Sect. 5.3.

[19] Betti (1823–1892), Morse (1892–1977).

[20] The first Betti number $\beta_1 = 2$ tells us that on the torus, there are two essentially different loops (also called fundamental 1-cycles) which cannot be continuously deformed into each other. For example, think of the equator and some meridian.

> Of the geometria situs, which Leibniz (1646–1716) sensed, and of which only a few geometers, Euler (1707–1783) and Vandermonde (1735–1796), were granted an obscured view, we know and have, after a hundred and fifty years, still little more than nothing.
>
> Carl Friedrich Gauss, 1833

> It was the discovery by Gauss of this very integral expressing the work done on a magnetic pole while describing a closed curve in presence of a closed electric current and indicating the geometric connection between the two closed curves, that led him to lament the small progress made in the *Geometry of Position* since the time of Leibniz, Euler and Vandermonde. We now have some progress to report, chiefly due to Riemann, Helmholtz and Listing.
>
> James Clerk Maxwell, 1873
> *A Treatise on Electricity and Magnetism*

> In obtaining a topological invariant by using a physical field theory, Gauss had anticipated *Topological Field Theory* by almost 150 years. Even the term topology was not used then. It was introduced by Johann Listing (1806-1882), a student and protegé of Gauss, in his 1847 essay *Preliminary Studies on Topology*. Gauss' linking number formula can also be interpreted as the equality of topological and analytical degree of a suitable function. Starting with this a far-reaching generalization of the Gauss integral to higher linking self-linking integrals can be obtained. This forms a small part of a program initiated by Maxim Kontsevich to relate topology of low-dimensional manifolds, homotopical algebras, and non-commutative geometry with topological field theories and Feynman diagrams in physics.
>
> Kishore Marathe, 2001
> *A chapter in physical mathematics: theory of knots in the sciences*[21]

Consider the situation pictured in Fig. 5.20. The electric current of strength $\mathcal{J}'$ flowing in the wire C', generates the magnetic field

$$\mathbf{B}(\mathbf{x}) = \frac{\mu_0 \mathcal{J}'}{4\pi} \int_{C'} \frac{(\mathbf{x}' - \mathbf{x}) \times d\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|^3},$$

by the Biot–Savart law. From the global Maxwell equation $\int_C \mathbf{B} d\mathbf{x} = \mu_0 \mathcal{J}$ it follows that

$$\mu_0 \mathcal{J} = \frac{\mu_0 \mathcal{J}'}{4\pi} \int_C \int_{C'} \frac{((\mathbf{x}' - \mathbf{x}) \times d\mathbf{x}')d\mathbf{x}}{|\mathbf{x} - \mathbf{x}'|^3}.$$

Hence

$$\mu_0 \mathcal{J} = \frac{\mu_0 \mathcal{J}'}{4\pi} \int_C \int_{C'} \frac{(\mathbf{x} - \mathbf{x}')(d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}.$$

In terms of physics, the linking number is defined to be the quotient

$$l(C, C') := \frac{\mathcal{J}}{\mathcal{J}'}.$$

For example, in Fig. 5.20(b) the linking number is equal to 2. In this case, the current strength $\mathcal{J}'$ in the wire C' flows twice through the surface S spanned by

[21] This nice survey article is contained in the monumental survey on modern mathematics edited by Engquist and Schmid (Eds.) (2001), pp. 873–888 (reprinted with permission). References to modern topological quantum field theory can be found on page 265.

the wire C. For the electric current strength $\mathcal{J}$ flowing through S, we then get $\mathcal{J} = 2\mathcal{J}'$. Generally, this leads us to the definition of the Gaussian integral for the linking number between the two curves C and C',

$$\boxed{l(C, C') := \frac{1}{4\pi} \int_C \int_{C'} \frac{(\mathbf{x} - \mathbf{x}')(d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}.}$$

Note that $l(C', C) = l(C, C')$.

Historically, the Biot–Savart law was discovered in 1820. In the 1830s in Göttingen, Gauss (1777–1855) experimentally studied the magnetic field of the earth with Wilhelm Weber (1804–1891). After Gauss' death, the fragment of a 1833 note was found. In this note entitled "On Electrodynamics", Gauss introduced his linking integral. [22]

5.6.8 Electric Fields, the Kronecker Integral, and the Mapping Degree

Electric Fields. Parallel to the Biot–Savart law for magnetic fields and Gauss' linking integral, let us now use the electric Coulomb field in order to give the Kronecker degree integral a physical motivation. In what follows, we will use the elegant language of differential forms which generalizes the classical Newton–Leibniz calculus to functions of several variables in a natural way. This calculus is basic for modern physics and mathematics. We will thoroughly study this in Volume III. As an introduction to differential forms and their applications to physics, we recommend the author's handbook Zeidler (2002a), Vols. 1, 2.

The localization principle. Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ be points in the interior of a closed 3-dimensional ball $\mathbb{B}$. For $j = 1, \ldots, N$, choose a small closed ball $\mathbb{B}_j$ centered at the point $\mathbf{x}_j$ such that $\mathbb{B}_j$ lies in the interior of $\mathbb{B}$, and $\mathbb{B}_1, \ldots, \mathbb{B}_N$ are pairwise disjoint (Fig. 5.21). Then

$$\boxed{\int_{\partial \mathbb{B}} \omega = \sum_{j=1}^{N} \int_{\partial \mathbb{B}_j} \omega} \tag{5.9}$$

if the differential 2-form ω is smooth and a cocycle. i.e.,

$$d\omega = 0 \quad \text{on} \quad \mathbb{B} \setminus \bigcup_{j=1}^{N} \operatorname{int} \mathbb{B}_j.$$

In addition, we assume that the spheres $\partial\mathbb{B}$ and $\partial\mathbb{B}_1, \ldots, \partial\mathbb{B}_N$ are oriented in such a way that the unit normal vector points outwards on the corresponding ball. In fact, let $N = 2$. By the Poincaré–Stokes theorem,

$$0 = \int_{\mathbb{B} \setminus (\mathbb{B}_1 \cup \mathbb{B}_2)} d\omega = \int_{\partial \mathbb{B}} \omega - \int_{\partial \mathbb{B}_1} \omega - \int_{\partial \mathbb{B}_2} \omega.$$

The same argument applies to $N \geq 2$.

[22] As an introduction to the relations between Gauss' linking number and modern knot theory, we recommend the article by A. Hirshfeld, Knots and physics: old wine in new bottles, Am. J. Phys. **66**(12) (1998), 1060–1066.

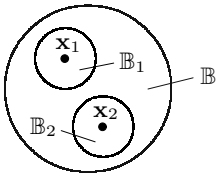


Fig. 5.21. Singularities

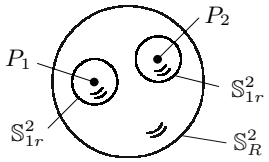


Fig. 5.22. Electric field

Generalization of Cauchy's residue theorem to electric fields. Suppose that $\mathbf{E} = \mathbf{E}(\mathbf{x})$ is an electric field which is smooth on the closed ball $\mathbb{B}$ up to the interior points $\mathbf{x}_1, \ldots, \mathbf{x}_N$ along with

$$\operatorname{div} \mathbf{E} = 0$$

for all $\mathbf{x} \in \mathbb{B}$ with $\mathbf{x} \neq \mathbf{x}_1, \ldots, \mathbf{x}_N$. In addition, suppose that the electric field behaves like the Coulomb field near the point $\mathbf{x}_j$, that is,

$$\mathbf{E}(\mathbf{x}) = \frac{Q_j}{4\pi\varepsilon_0 \|\mathbf{x} - \mathbf{x}_j\|^2} \cdot \frac{\mathbf{x} - \mathbf{x}_j}{\|\mathbf{x} - \mathbf{x}_j\|} + \mathbf{E}_j(\mathbf{x})$$

for all $\mathbf{x} \neq \mathbf{x}_1, \ldots, \mathbf{x}_N$ where the field $\mathbf{E}_j$ is smooth on $\mathbb{B}_j$. Then

$$\boxed{\varepsilon_0 \int_{\partial\mathbb{B}} (\mathbf{E}\mathbf{n}) dS = \sum_{j=1}^{N} Q_j} \tag{5.10}$$

where $\mathbf{n}$ denotes the outer unit normal vector of $\mathbb{B}$. This is a special case of the localization principle (5.9). In fact, let $N = 2$. By the Gauss integral theorem

$$0 = \int_{\mathbb{B}\setminus(\mathbb{B}_1 \cup \mathbb{B}_2)} \operatorname{div} \mathbf{E}\, d^3x = \int_{\partial\mathbb{B}} (\mathbf{E}\mathbf{n})\, dS - \int_{\partial\mathbb{B}_1} (\mathbf{E}\mathbf{n})\, dS - \int_{\partial\mathbb{B}_2} (\mathbf{E}\mathbf{n})\, dS.$$

Moreover,

$$\int_{\partial\mathbb{B}_j} \frac{Q_j}{4\pi\varepsilon_0 \|\mathbf{x} - \mathbf{x}_j\|^2} \cdot \frac{(\mathbf{x} - \mathbf{x}_j)\mathbf{n}}{\|\mathbf{x} - \mathbf{x}_j\|}\, dS = \frac{Q_j}{4\pi\varepsilon_0 r^2} \cdot 4\pi r^2 = \frac{Q_j}{\varepsilon_0}\,.$$

The claim follows now by contracting the ball $\mathbb{B}_j$ to the center $\mathbf{x}_j$.

The pull-back of a differential form. Let $f : \mathbb{R} \to \mathbb{R}$ be a smooth function. Then $dy = f'(x)\, dx$. For the differential form

$$\omega = a(y)\, dy,$$

we define the pull-back with respect to f by

$$f^*\omega = a\big(f(x)\big) f'(x)\, dx.$$

If $f : \mathbb{R} \to \mathbb{R}$ is a diffeomorphism then the classical substitution rule tells us that

$$\int_\alpha^\beta f^*\omega = \int_{f(\alpha)}^{f(\beta)} \omega$$

for all real numbers $\alpha \leq \beta$. This technique can be generalized to higher dimensions in a quite natural way. Explicitly, for a smooth map $f : \mathbb{R}^n \to \mathbb{R}^m$, the pull-back of the differential form

$$\omega = \sum_{j_1,\ldots,j_p=1}^{m} a_{j_1\ldots j_p}(y)\, dy^{j_1} \wedge \cdots \wedge dy^{j_p}$$

is obtained by the substitution

$$f^*\omega := \sum_{j_1,\ldots,j_p=1}^{m} a_{j_1\ldots j_p}\big(f(x)\big)\, dy^{j_1} \wedge \cdots \wedge dy^{j_p}$$

with

$$dy^{j_k} = \sum_{l=1}^{n} \frac{\partial f_{j_k}(x)}{\partial x^l}\, dx^l$$

for all indices $j_1, \ldots, j_p$. The pull-back has the crucial property

$$\boxed{d(f^*\omega) = f^*(d\omega).}$$

If $f : \mathbb{R}^n \to \mathbb{R}^n$ is a diffeomorphism, then

$$\boxed{\int_\Omega f^*\omega = \sigma \int_{f(\Omega)} \omega}$$

for all open (or closed) bounded subsets Ω of $\mathbb{R}^n$ and all smooth differential n-forms ω. Here, $\sigma = 1$ (resp. $\sigma = -1$) if f preserves (resp. reverses) the orientation. If we contract the set Ω into the point $\mathbf{x}$, then

$$\lim_{\Omega \to \mathbf{x}} \frac{\int_\Omega f^*\omega}{\int_{f(\Omega)} \omega} = \operatorname{sgn} \det f'(\mathbf{x}). \tag{5.11}$$

The Kronecker integral. Let $f : \mathbb{B} \to \mathbb{R}^3$ be a smooth map on the closed ball $\mathbb{B}$ such that the restriction

$$f : \partial\mathbb{B} \to \mathbb{S}^2$$

maps the boundary of $\mathbb{B}$ to the unit sphere $\mathbb{S}^2$. In addition, assume that the map f has precisely the zeros $\mathbf{x}_1, \ldots, \mathbf{x}_N$. In addition, let $\det f'(\mathbf{x}_j) \neq 0$ for all j.[23] Define the mapping degree

$$\deg f := \sum_{j=1}^{N} \operatorname{sgn} \det f'(\mathbf{x}_j).$$

The famous Kronecker integral formula from 1869 tells us that[24]

$$\boxed{\int_{\partial\mathbb{B}} f^*\upsilon = \deg f \int_{\mathbb{S}^2} \upsilon} \tag{5.12}$$

[23] The derivative $f'(\mathbf{x})$ is the matrix of the first-order partial derivatives of the components of f at the point $\mathbf{x}$.

[24] Kronecker (1823–1891)

where υ is the volume form on the unit sphere $\mathbb{S}^2$. Explicitly, in spherical coordinates,

$$\upsilon = \cos\vartheta \, d\varphi \wedge d\vartheta.$$

In Cartesian (X, Y, Z)-coordinates,

$$\upsilon = X \, dY \wedge dZ + Y \, dZ \wedge dX + Z \, dX \wedge dY.$$

Let us prove (5.12). The trick is to use the Coulomb field

$$\mathbf{E}(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|^3}$$

along with the corresponding differential form

$$\mathcal{E} = E^1 dY \wedge dZ + E^2 dZ \wedge dX + E^3 dX \wedge dY.$$

Then

$$d\mathcal{E} = \left(\frac{\partial E^1}{\partial X} + \frac{\partial E^2}{\partial Y} + \frac{\partial E^3}{\partial Z} \right) dX \wedge dY \wedge dZ = \operatorname{div} \mathbf{E} \, dX \wedge dY \wedge dZ.$$

Hence $d\mathcal{E} = 0$ if $\mathbf{x} \neq 0$. By the Poincaré–Stokes theorem,

$$\int_{\partial \mathbb{B}} \mathcal{E} = \int_{\mathbb{B}} d\mathcal{E} = \int_{U(0)} d\mathcal{E}.$$

Here, $\mathbb{B}$ is the closed unit ball, and $U(0)$ is an arbitrary open neighborhood of the origin. Since

$$d(f^* \mathcal{E}) = f^*(d\mathcal{E}) = 0$$

for all $\mathbf{x} \neq \mathbf{x}_1, \ldots, \mathbf{x}_N$, it follows from the Poincaré–Stokes theorem that

$$\int_{\partial \mathbb{B}} f^* \mathcal{E} = \int_{\mathbb{B}} d(f^* \mathcal{E}) = \sum_{j=1}^{N} \int_{\mathbb{B}_j} f^*(d\mathcal{E})$$

where $\mathbb{B}_j$ is a small ball centered at $\mathbf{x}_j$. Contracting the ball,

$$\lim_{\mathbb{B}_j \to \mathbf{x}_j} \frac{\int_{\mathbb{B}_j} f^* \mathcal{E}}{\int_{f(\mathcal{B}_j)} \mathcal{E}} = \operatorname{sgn} \det f'(\mathbf{x}_j),$$

by (5.11). This implies

$$\int_{\partial \mathbb{B}} f^* \mathcal{E} = \sum_{j=1}^{N} \operatorname{sgn} \det f'(\mathbf{x}_j) \int_{\mathbb{S}^2} \mathcal{E}.$$

Finally, observe that $\mathcal{E} = \omega$ on $\mathbb{S}^2$.

5.6.9 The Heat Kernel and the Atiyah–Singer Index Theorem

> Like others, we came to the heat kernel via one direction of mathematics. However, as we progressed in that direction, we realized that the heat kernel plays a central role in almost all directions we can think of. That it bears a name related to physics only indicates that it was originally discovered in direction with heat.[25] But even in physics, its significance goes way beyond the giving a mathematical model for heat distribution. The name cannot be changed – it's too late for that – but the impression some people may have that an occurrence of a certain kernel called the heat kernel means one is necessarily doing physics is a false impression. Maybe one is and maybe one isn't... There is a universal gadget which is a dominant factor practically everywhere in mathematics, also in physics, and has very simple and powerful properties. We have no a priori explanation (psychological, philosophical, mathematical) for the phenomenon of the existence of such a universal gadget.
>
> Jay Jorgenson and Serge Lang, 2001
> *The Ubiquitous Heat Kernel*[26]

To explain the basic ideas of the far-reaching heat kernel approach to mathematics, let us study the most simple example, namely, heat conduction on the unit circle. To this end, let $C^\infty(\mathbb{S}^1)$ denote the space of smooth complex-valued functions

$$f : \mathbb{S}^1 \to \mathbb{C}$$

on the unit circle $\mathbb{S}^1$. Alternatively, these functions can be regarded as smooth functions $f = f(\varphi)$ on the real line of period 2π, with respect to the angle φ. In this case, we have the convergent Fourier series expansion

$$f(\varphi) = \sum_{n=-\infty}^{\infty} a_n(f)\, \mathrm{e}^{\mathrm{i}n\varphi} \tag{5.13}$$

for all $\varphi \in \mathbb{R}$ where $a_n(f) := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\varphi)\mathrm{e}^{-\mathrm{i}n\varphi}\, d\varphi$. We are given the function $g \in C^\infty(\mathbb{S}^1)$. Let us first study the following differential equation on the unit circle,

$$\boxed{Df = g, \qquad f \in C^\infty(\mathbb{S}^1),} \tag{5.14}$$

with the differential operator $D := -\mathrm{i}\frac{d}{d\varphi}$. We want to show that

$$\boxed{\operatorname{ind} D = \chi(\mathbb{S}^1) = 0.} \tag{5.15}$$

This tells us that an analytic quantity – the index of the differential operator D on the unit circle $\mathbb{S}^1$ – equals a topological quantity – the Euler characteristic $\chi(\mathbb{S}^1)$ of the unit circle.[27]

[25] Joseph Fourier, *La théorie de la chaleur* (heat theory), Paris, 1822.

[26] This survey article covers rich material from number theory, algebraic geometry, differential geometry, analysis, approximation theory, and topology. See Engquist and Schmid (2001), pp. 655–683 (reprinted with permission).

[27] Euler (1707–1783), Laplace (1749–1827), Fourier (1768–1830), Gauss (1777–1855), Jacobi (1804–1851), Dirichlet (1805–1859), Riemann (1826–1866), Weyl (1885–1955), Feynman (1918–1988). For their index theory, Atiyah (born 1929) and Singer (born 1924) were awarded the Abel prize in 2004.

The classical Fourier method. Let us prove (5.15). Using Fourier series expansion, it follows from

$$g(\varphi) = a_0(g) + a_1(g)\mathrm{e}^{\mathrm{i}\varphi} + a_{-1}(g)\mathrm{e}^{-\mathrm{i}\varphi} + ...$$

and $f'(\varphi) = \mathrm{i}a_1(f)\mathrm{e}^{\mathrm{i}\varphi} - \mathrm{i}a_{-1}(f)\mathrm{e}^{-\mathrm{i}\varphi} + ...$ that the original problem (5.14) has a solution iff $a_0(g) = 0$. In other words, we have the solvability condition

$$\int_{-\pi}^{\pi} g(\varphi)d\varphi = 0.$$

The solution f is then given by (5.13) where

$$a_0(f) = \text{const}, \qquad a_n(f) = -\frac{a_n(g)}{n}, \qquad n = \pm 1, \pm 2, ...$$

By definition, the index of the equation (5.14) reads as

$$\operatorname{ind} D := d - s$$

where d is the dimension of the solution space, and s is the number of linearly independent solvability conditions. Since $d = 1$ and $s = 1$, $\operatorname{ind} D = 0$. This proves the claim.

The heat equation on the unit circle. Let us pass from the operator D to the Laplacian $\Delta := D^2$ on the unit circle. Explicitly,

$$\Delta = -\frac{d^2}{d\varphi^2}.$$

Following Fourier, we want to study the initial-value problem for the unit circle,

$$\begin{aligned} \frac{\partial T(\varphi,t)}{\partial \varphi} &= -D^2 T(\varphi,t), \qquad \varphi \in [-\pi,\pi], \;\; t > 0, \\ T(\varphi,0) &= T_0(\varphi). \end{aligned} \tag{5.16}$$

This equation describes the propagation of heat on the unit circle (e.g., heat conduction within a thin wire). Here, T denotes temperature. Alternatively, this can be viewed as the propagation of heat on the real line where the temperature T has the period 2π with respect to the position coordinate $\varphi \in \mathbb{R}$. We are given the initial temperature $T_0 \in C^\infty(\mathbb{S}^1)$ at the initial time $t = 0$. By using the classical Fourier method, our goal is to describe the unique solution of (5.16) by the integral formula

$$\boxed{T(\varphi,t) = \int_{-\pi}^{\pi} \mathcal{P}_+(\varphi,\psi,t)T_0(\psi)d\psi} \tag{5.17}$$

for all $\varphi \in [-\pi,\pi]$, $t > 0$. Here, the heat kernel $\mathcal{P}_+$ is given by the infinite series

$$\boxed{\mathcal{P}_+(\varphi,\psi,t) := \sum_{k=-\infty}^{\infty} \mathrm{e}^{-t\lambda_k} f_k(\varphi) f_k(\psi)^\dagger}$$

along with the functions

$$f_k(\varphi) := \frac{\mathrm{e}^{\mathrm{i}k\varphi}}{\sqrt{2\pi}}, \qquad \varphi \in [-\pi, \pi], \qquad k = 0, \pm 1, \pm 2, \ldots$$

which represent eigenfunctions of the operator D. Explicitly,

$$Df_k = kf_k \qquad \text{on} \quad \mathbb{S}^1, \qquad k = 0, \pm 1, \pm 2, ...$$

This implies

$$D^2 f_k = \lambda_k f_k \qquad \text{on} \quad \mathbb{S}^1, \qquad k = 0, \pm 1, \pm 2, ...$$

with the eigenvalue $\lambda_k = k^2$.

In terms of Feynman's approach to physics, the heat kernel is identical to the Feynman propagator kernel for the heat equation.

Let us introduce the complex Hilbert space $L_2(\mathbb{S}^1)$ of all measurable functions $f : \mathbb{R} \to \mathbb{C}$ of period 2π with the inner product

$$\langle f|h\rangle := \int_{-\pi}^{\pi} f(\varphi)^\dagger h(\varphi) d\varphi.$$

The functions $f_0, f_1, ...$ form a complete orthonormal system in the Hilbert space $L_2(\mathbb{S}^1)$. Explicitly, the heat kernel of (5.16) reads as

$$\mathcal{P}_+(\varphi, \psi, t) := \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \mathrm{e}^{-k^2 t} \mathrm{e}^{\mathrm{i}k(\varphi-\psi)}, \qquad \varphi, \psi \in \mathbb{R}, \quad t > 0.$$

Note that the series for $\mathcal{P}_+$ is convergent for times $t > 0$, but divergent for $t < 0$. This reflects the fact that the temperature distribution T_0 at time $t = 0$ does not determine the temperature distribution in the past.

Theorem 5.1 *The initial-value problem* (5.16) *for heat conduction on the unit circle has the unique solution* (5.17).

Proof. It follows from (5.17) that

$$T(\varphi, t) = \sum_{k=-\infty}^{\infty} a_k(T_0) \mathrm{e}^{-k^2 t} \mathrm{e}^{\mathrm{i}k\varphi}.$$

Differentiation shows that this is a solution of the initial-value problem (5.16). Uniqueness of the solution follows from the classical maximum principle for the heat equation.[28] □

The language of the Dirac calculus. For the convenience of the reader, we want to reformulate the classical Fourier method above in terms of the elegant Dirac calculus. We restrict ourselves to formal arguments. To begin with, let us denote the function f_k by $|k\rangle$. Moreover, we set $f_k(\varphi) := \langle\varphi|k\rangle$. We have the orthogonality conditions

$$\langle k|l\rangle = \delta_{kl}, \qquad k, l = 0, \pm 1, \pm 2, \ldots$$

and the two completeness relations

$$\sum_{k=-\infty}^{\infty} |k\rangle\langle k| = I, \qquad \int_{-\pi}^{\pi} d\varphi \, |\varphi\rangle\langle\varphi| = I.$$

[28] The Fourier method is studied in greater detail in Zeidler (1995), Vol. 1, Chap. 5. Many applications of the maximum principle to linear and nonlinear differential equations can be found in Zeidler (1986), Vol. I, Chap. 7.

For example, this implies the inner product

$$\langle k|l\rangle = \int_{-\pi}^{\pi} \langle k|\varphi\rangle\langle\varphi|l\rangle \, d\varphi = \int_{-\pi}^{\pi} f_k(\varphi)^\dagger f_l(\varphi) d\varphi.$$

Now to the point. Introducing the Feynman propagator $P_+(t) := \theta(t)\mathrm{e}^{-D^2 t}$ for all $t \in \mathbb{R}$, the solution of the initial-value problem (5.16) reads as

$$\boxed{T(t) = P_+(t)T_0, \qquad t \geq 0.}$$

Let us motivate this in a formal way. By the completeness relation,

$$\langle\varphi|T(t)\rangle = \int_{-\pi}^{\pi} d\psi \langle\varphi|P_+(t)|\psi\rangle\langle\psi|T_0\rangle.$$

Since $D|k\rangle = k|k\rangle$, we get the crucial eigensolutions

$$P_+(t)|k\rangle = \mathrm{e}^{-k^2 t}|k\rangle, \qquad t \geq 0$$

of the Feynman propagator. Introducing the Feynman propagator kernel

$$\mathcal{P}_+(\varphi, \psi, t) := \langle\varphi|P_+(t)|\psi\rangle, \qquad t \geq 0,$$

the completeness relation yields

$$\begin{aligned}\mathcal{P}_+(\varphi, \psi, t) &= \sum_{k=-\infty}^{\infty} \langle\varphi|P_+(t)|k\rangle\langle k|\psi\rangle \\ &= \sum_{k=-\infty}^{\infty} \mathrm{e}^{-k^2 t}\langle\varphi|k\rangle\langle k|\psi\rangle = \sum_{k=-\infty}^{\infty} \mathrm{e}^{-k^2 t} f_k(\varphi) f_k(\psi)^\dagger.\end{aligned}$$

This is precisely the solution formula (5.17).

The zeta function of a compact manifold. Before continuing the study of the heat equation on the unit circle, let us embed this problem into the general context of real d-dimensional compact manifolds M.[29] On such a manifold, the Laplacian Δ has a countable set of eigenvalues $0 \leq \lambda_0 < \lambda_1 < \ldots$ which possesses the following famous Weyl asymptotics

$$\lambda_n \simeq \left(\frac{2\pi n}{V}\right)^{\frac{2}{d}} \qquad \text{as} \quad n \to \infty, \tag{5.18}$$

where d and V denote the dimension and the volume of the manifold M, respectively.[30] By definition, the zeta function of the manifold M reads as

$$\zeta_M(s) := \sum_{n=1}^{\infty} \frac{1}{\lambda_n^s}, \qquad s \in \mathbb{C}, \;\; \Re(s) > \frac{d}{2}.$$

[29] The proofs of the following statements can be found in the standard textbook on spectral geometry by Gilkey (1995). Many applications to physical problems are investigated in Kirsten (2002).

[30] Hermann Weyl's first investigations on (5.18) from 1911 were motivated by Planck's radiation law. In modern language, Weyl wanted to prove that the properties of a photon gas in a box do not depend on the shape of the box, but only on the volume.

This series converges for all complex numbers s with $\Re(s) > \frac{d}{2}$. Introducing the energy $E_n := -\ln \lambda_n$, we get the partition function

$$\zeta_M(s) := \sum_{n=1}^{\infty} \mathrm{e}^{-sE_n} \qquad \Re(s) > \frac{d}{2}.$$

This zeta function is closely related to the Dirichlet series

$$K_M(t) := \sum_{n=1}^{\infty} \mathrm{e}^{-t\lambda_n}, \qquad t > 0$$

which is called the global heat kernel of the manifold M. Letting $\tau := \sqrt{t}$, the global heat kernel possesses the following asymptotic expansion for small times,

$$\boxed{K_M(t) = \frac{c_{-d}}{\tau^d} + \frac{c_{-d+1}}{\tau^{d-1}} + ... + \frac{c_{-1}}{\tau} + c_0 + c_1\tau + ..., \qquad \tau \to +0.}$$

The coefficients $c_{-d}, c_{-d+1}, ...$ are called heat kernel coefficients. The normalized Mellin transform of the heat kernel yields the zeta function,

$$\zeta_M(s) = \frac{1}{\Gamma(s)} \int_0^{\infty} K(t) t^{s-1} dt, \qquad s \in \mathbb{C},\ \Re(s) > \frac{d}{2}$$

along with the relations

$$c_{-d+k} = \Gamma(s) \mathrm{Res}\, \zeta_M(s)_{|s=(d-k)/2}, \qquad k = 0, 1, 2, ...$$

In particular, the zeta function ζ_M is holomorphic in a neighborhood of the origin $s = 0$. In order to compute Feynman functional integrals, physicists need the product

$$\det \Delta := \prod_{n=0}^{\infty} \lambda_n$$

called the determinant of Δ (see page 658). Obviously, the derivative $\zeta'_M(0)$ is equal to the sum $-\lambda_0 - \lambda_1 - ...$ of eigenvalues. Hence we get the basic formula

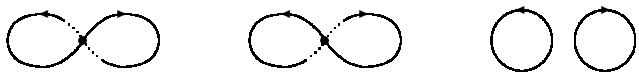
$$\boxed{\det \Delta = \mathrm{e}^{-\zeta'_M(0)}} \tag{5.19}$$

frequently used by physicists.

The heat kernel method for computing the index. One of the deepest results of modern mathematics is the 1963 Atiyah–Singer index theorem. This theorem generalizes the Riemann–Roch theorem for Riemann surfaces; it tells us that the index of each elliptic operator (e.g., the Laplacian) on a compact manifold M is a topological invariant of the manifold and can be computed by means of the heat kernel coefficients. In particular, the index does not change if both the manifold and the differential operator are perturbed in a reasonable manner. Thus, the Atiyah–Singer index theorem tells us that

> *There exists a deep relation between the topology (i.e., the qualitative behavior) of a manifold and the structure of analytic objects on the manifold.*

This index theorem also applies to pseudo-differential operators including large classes of differential and integral equations. The Atiyah–Singer theorem will be studied in Volume IV; it is a crucial ingredient of quantum mathematics. Interestingly enough, there exists a rigorous supersymmetric proof of the Atiyah–Singer



(a) L_+ (b) L_- (c) L_0

Fig. 5.23. Knots and links

index theorem which was motivated by physical mathematics. This proof can be found in Cycon et al. (1986). See also Berline, Getzler, and Vergne (1991).

At this point, let us consider the special case of the Atiyah–Singer theorem for the operator D on the 1-dimensional unit circle $\mathbb{S}^1$ with $d = 1$. The eigenvalues of the Laplacian Δ are given by

$$\lambda_n = n^2, \qquad n = 0, 1, 2, ...$$

This corresponds to the Weyl asymptotics (5.18) where $V = 2\pi$ denotes the volume of the unit circle. This yields the zeta function

$$\zeta_{\mathbb{S}^1}(s) = \sum_{n=1}^{\infty} \frac{1}{n^{2s}} = \zeta(2s), \qquad s \in \mathbb{C},\ \Re(s) > \frac{1}{2},$$

and the heat kernel

$$K_{\mathbb{S}^1}(t) = \sum_{n=0}^{\infty} \mathrm{e}^{-tn^2} = \tfrac{1}{2}(\vartheta_3(0,t) + 1), \qquad t > 0.$$

The function ϑ_3 is one of the four Jacobi theta functions. The heat kernel allows the following asymptotic expansion,

$$\mathcal{P}_+(\varphi, \varphi, t) \sim \frac{1}{\sqrt{4\pi t}} + \mathcal{P}_0(\varphi) + t\mathcal{P}_1(\varphi) + t^2\mathcal{P}_2(\varphi)...,$$

as time goes to zero, $t \to +0$. Here, $\mathcal{P}_0 = 0$. The Atiyah–Singer index theorem tell us the index formula

$$\boxed{\operatorname{ind} D = \int_{\mathbb{S}^1} \mathcal{P}_0 d\varphi.}$$

Hence $\operatorname{ind} D = 0$, as obtained above in an explicit way.

there exists an orientation preserving diffeomorphism

$$F : f(\mathbb{S}^1) \to \mathbb{R}^3$$

which maps the set $f(\mathbb{S}^1)$ onto $g(\mathbb{S}^1)$. The knot is called an unknot iff it is equivalent to the unit circle.

By definition, a link is a finite collection of pairwise disjoint knots. Note that knots are special links. Graphically, links are represented by projections onto a fixed plane with crossings marked as over and under (Fig. 5.23). There exists a complete classification of all knots having at most 16 crossing points. This list consists of 1, 701, 936 knots.[31] Two links are called equivalent iff they consist of equivalent knots.

[31] See the introduction to knot theory by Adams (1994) and the article by Hoste et al. (1998).

The main task of knot theory is to decide whether two knots or links are equivalent.

For example, one wants to know whether a given knot is trivial, that is, it is equivalent to the unknot. To this end, one introduces link

5.6.10 Knots and Topological Quantum Field Theory

> In the last twenty years a body of mathematics has evolved with strong direct input from theoretical physics, for example from classical and quantum field theories, statistical mechanics and string theory. In particular, in the geometry and topology of low dimensional manifolds (i.e., manifolds of dimensions 2, 3 and 4) we have seen new results, some of them quite surprising, as well as new ways of looking at old results. Donaldson's work based on his study of the solution space of the Yang–Mills equations, monopole equations of Seiberg–Witten, Floer homology, quantum groups and topological quantum field theoretical interpretation of the Jones polynomial and other knot invariants are some of the examples of this development. Donaldson, Jones and Witten have received the Fields medal for their work. We think the name "Physical Mathematics" is appropriate to describe this new, exciting and fast growing area of mathematics. Recent developments in knot theory make it an important chapter in "Physical Mathematics." Until the early 1980s it was an area in the backwaters of topology. Now it is a very active area of research with its own journal.[32]
>
> Kishore Marathe, 2001
>
> *A chapter in physical mathematics: theory of knots in the sciences*

Intuitively, a knot is a closed curve in the 3-dimensional Euclidean space without self-intersections; two knots are called equivalent iff they can be deformed into each other in the 3-dimensional Euclidean space, by avoiding self-intersections during the deformation process. More precisely, by definition, a knot is a smooth embedding

$$\boxed{f : \mathbb{S}^1 \to \mathbb{R}^3}$$

of the unit circle $\mathbb{S}^1$ into $\mathbb{R}^3$, that is, the image $f(\mathbb{S}^1)$ of the smooth map f is a submanifold of $\mathbb{R}^3$, and the induced map

$$f : \mathbb{S}^1 \to f(\mathbb{S}^1)$$

is a diffeomorphism.[33] Two knots $f, g : \mathbb{S}^1 \to \mathbb{R}^3$ are called equivalent (or ambient isotopic) iff

invariants which possess the characteristic property that equivalent links have the same link invariant. A crucial link invariant is the Jones polynomial

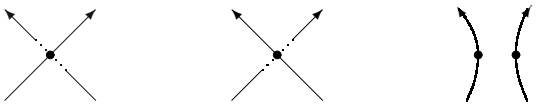
$$J(t) := \sum_{k=-n}^{n} a_k t^{k/2}$$

generated by $t^{1/2}$ and $t^{-1/2}$ with integer coefficients a_k, $n = 0, 1, 2, ...$ The Jones polynomials have the following properties:

(J1) Two oriented equivalent links have the same Jones polynomial.

[32] See the footnote on page 252.

[33] The precise definition of submanifolds will be given in Volume III.



(a) L_+ (b) L_- (c) L_0

Fig. 5.24. Crossing points

(J2) For the unknot, $J(t) :\equiv 1$.

(J3) If the projective link diagrams of three links differ exactly at one distinguished crossing point, then the corresponding Jones polynomials satisfy the following skein relation

$$\boxed{t^{-1}J_+(t) - tJ_-(t) = (t^{1/2} - t^{-1/2})J_0(t)}$$

where the Jones polynomials J_+, J_-, J_0 correspond to the behavior of the link diagrams L_+, L_-, L_0, respectively, at the distinguished crossing point (Fig. 5.24).

(J4) For the mirror image of a link, the Jones polynomial is transformed by $t \mapsto t^{-1}$.

For a given link, the idea of computing the Jones polynomial is to successively change the link diagram at crossing points according to Fig. 5.24 in order to finally obtain the unknot. The corresponding skein relations then yield recursive formulas for the desired Jones polynomial. In particular, for the links L_+, L_-, L_0 pictured in Fig. 5.23, the skein relation with $J_\pm = 1$ by (J2) yields

$$t^{-1} - t = (t^{1/2} - t^{-1/2})J_0(t).$$

Hence the Jones polynomial J_0 of the link L_0 in Fig. 5.23(c) is equal to

$$J_0(t) = -t^{1/2} - t^{-1/2}.$$

Considering Fig. 5.23, the link L_+ (resp. L_-) is not equivalent to the link L_0, since the corresponding Jones polynomial $J_\pm = 1$ is different from J_0.

The Jones polynomials were discovered by Jones in 1985 when studying the structure of von Neumann operator algebras.[34] A few years later, Witten (1989) published a fundamental paper on a beautiful physical interpretation of the Jones polynomials via a special model in quantum field theory. The idea is to use

- the principle of critical action for the Chern–Simons Lagrangian on the 3-dimensional sphere $\mathbb{S}^3$ with respect to the gauge group $SU(2)$;
- quantization of this classical field theory yields the corresponding Feynman functional integral, as a formal partition function;
- to each knot on $\mathbb{S}^3$ one can assign a physical quantity called the Wilson loop;
- finally, the Jones polynomial of a knot or link is the vacuum expectation value of the corresponding Wilson loop.

Roughly speaking, the Chern–Simons Lagrangians represent quite natural gauge field theories on 3-dimensional manifolds.

Topological quantum field theory. This new branch of topology was founded by E. Witten, Topological quantum field theory, Commun. Math. Phys. **117** (1988), 353–386. The basic idea of topological quantum field theory is to use the Lagrangians of special gauge field theories and the corresponding Feynman

[34] Vaughan Jones was awarded the Fields medal in 1990.

functional integrals in order to construct sophisticated topological invariants for low-dimensional manifolds. As an introduction to modern knot theory and its relations to physics, chemistry, and biology (DNA), we recommend the survey article by Marathe (2001) and the monographs by Kaufman (2001) and Flappan (2000) (molecular chirality in chemistry). For topological quantum field theory, see the lectures given by Atiyah (1990b) and Witten (1999c). We also refer to the monograph by Jost (2002a) (geometric analysis and classical models in quantum field theory).

For a wealth of material on the relation between topological charges and nonlinear partial differential equations arising in modern physics, we recommend the monographs by Felsager (1997), Naber (1997), and Yang (2001).

5.7 Quantum States

In the next sections we want to show that the study of quantum states is closely related to crucial topological concepts. We will consider

- the covering group of the rotation group and the electron spin,
- the space of spin states of an electron with the Hopf fibration of the 3-dimensional sphere (the $U(1)$-Hopf bundle), and
- the relation between the space of spin states, Grassmann manifolds, and projective geometry.

5.7.1 The Topological Character of the Electron Spin

> Our 3-dimensional intuition sees rotations, but not the corresponding universal covering group $Spin(3)$ which lives on the 4-dimensional unit sphere. Nature sees this in terms of the electron spin.
>
> Folklore

Let us sketch why the electron spin is a topological effect.

> *The point is that the rotation group $SO(3)$ of the 3-dimensional Euclidean space of our intuition is not simply connected, whereas the multiplicative group $Spin(3)$ of unit quaternions is simply connected.*

Let us start with the famous Cayley rotation formula[35]

$$\boxed{\mathbf{x}' = Q\mathbf{x}Q^{\dagger}} \tag{5.20}$$

where we use the following quaternions

$$Q := \cos\frac{\theta}{2} + \mathbf{n}\sin\frac{\theta}{2}, \qquad Q^{\dagger} := \cos\frac{\theta}{2} - \mathbf{n}\sin\frac{\theta}{2}.$$

Formula (5.20) describes the rotation of the position vector $\mathbf{x}$ about an axis given by the unit vector $\mathbf{n}$ and the rotation angle $\theta \in [0, 2\pi[$. Here, we use Hamilton's quaternions. A quaternion $(\alpha, \mathbf{a})$ is an ordered pair of a real number α and an vector $\mathbf{a}$ in the 3-dimensional Euclidean space. We briefly write $\alpha + \mathbf{a}$. The sum and the product of two quaternions are defined by

$$(\alpha + \mathbf{a}) + (\beta + \mathbf{b}) = (\alpha + \beta) + (\mathbf{a} + \mathbf{b})$$

[35] Hamilton (1805–1865), Cayley (1821–1895).

and

$$(\alpha + \mathbf{a})(\beta + \mathbf{b}) = \alpha\beta + \alpha\mathbf{b} + \beta\mathbf{a} + \mathbf{a} \times \mathbf{b} - \mathbf{ab}.$$

If we replace Q by $-Q$, then we get the same rotated vector $\mathbf{x}'$. This corresponds to a passage from $\mathbf{n}$ to $-\mathbf{n}$ and from θ to $2\pi - \theta$. Consequently, there are two additional degrees of freedom hidden in the Cayley formula (5.20).

Nature sees these hidden degrees of freedom in terms of the electron spin.

Let us discuss this. The length of a quaternion is defined by

$$||\alpha + \mathbf{a}|| := \sqrt{\alpha^2 + \mathbf{a}^2}.$$

In particular, $||Q|| = \sqrt{\cos^2 \frac{\theta}{2} + \sin^2 \frac{\theta}{2}\ \mathbf{n}^2} = 1$. The set of all unit quaternions Q forms a multiplicative group called the spin group $Spin(3)$. The Cayley formula (5.20) describes a group epimorphism

$$\boxed{\chi : Spin(3) \to SO(3)} \tag{5.21}$$

from the spin group $Spin(3)$ onto the group $SO(3)$ of rotations $SO(3)$ in three-dimensional space, that is, the surjective map χ respects products.

The group $SO(3)$ is not simply connected, but the group $Spin(3)$, which corresponds to the 4-dimensional unit sphere, is simply connected. We call $Spin(3)$ the universal covering group of the group $SO(3)$. Let us now study this in terms of Lie groups and Lie algebras.

(i) The Lie group $U(1)$ of rotations in the plane consists of all complex numbers A with $|A| = 1$ Hence $A = \mathrm{e}^{\mathrm{i}\theta}$ with real parameter θ. For all the complex numbers $z = x + y\mathrm{i}$, the transformation

$$z' = \mathrm{e}^{\mathrm{i}\theta} z$$

describes a rotation in the plane about the origin with rotation angle θ. Set $\varrho(\theta) := \mathrm{e}^{\mathrm{i}\theta}$. The map

$$\varrho : \mathbb{R} \to U(1)$$

is a group epimorphism from the additive group $\mathbb{R}$ onto the multiplicative group $U(1)$. Explicitly,

$$\varrho(\theta + \varphi) = \varrho(\theta)\varrho(\varphi) \qquad \text{for all} \quad \theta, \varphi \in \mathbb{R}.$$

The unit circle $U(1)$ is not simply connected, whereas the real line $\mathbb{R}$ is simply connected. The additive group $\mathbb{R}$ is called the universal covering group of $U(1)$.

(ii) The Lie algebra $u(1)$. By Taylor expansion,

$$\mathrm{e}^{\mathrm{i}\theta} = 1 + \mathrm{i}\theta + o(\theta), \qquad \theta \to 0.$$

The set $u(1) := \{\mathrm{i}\theta : \ \theta \in \mathbb{R}\}$ is called the Lie algebra to $U(1)$.

(iii) The Lie group $SO(3)$. Choose a right-handed Cartesian (x, y, z)-system in the Euclidean space of our intuition. All possible rotations about the origin can be described by the matrix transformations

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = A \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

where the real (3×3)-matrix has the properties $\det A = 1$ and $A^{-1} = A^d$. All of these matrices form the group $SO(3)$.[36]

(iv) The Lie algebra $so(3)$ of the group $SO(3)$ consists of all real (3×3)-matrices B with $B = -B^d$. The set $so(3)$ is a real 3-dimensional linear space equipped with the Lie product

$$[B, C]_- := BC - CB.$$

That is, $B, C \in so(3)$ implies $[B, C]_- \in so(3)$. The matrices in $so(3)$ are called infinitesimal rotations. Set

$$\varrho(B) := \mathrm{e}^B.$$

The map $\varrho : so(3) \to SO(3)$ is surjective. In other words, each given matrix $A \in SO(3)$ can be represented in the form $A = \mathrm{e}^B$ with $B \in so(3)$. The relation between the commutator of $SO(3)$ and the Lie product of the corresponding Lie algebra $so(3)$ is given by

$$[A, B]_- = \frac{\partial^2}{\partial t \partial s} \left(\mathrm{e}^{tA} \mathrm{e}^{sB} \mathrm{e}^{-tA} \mathrm{e}^{-sB} \right)_{|t,s=0}$$

where t and s are real parameters.

(v) The Lie group $SU(n), n \geq 2$ consists of all the complex $(n \times n)$-matrices $\mathcal{A}$ with $\det \mathcal{A} = 1$ and $\mathcal{A}^{-1} = \mathcal{A}^\dagger$.

(vi) The Lie algebra $su(n)$ of the Lie group $SU(n)$ consists of all the complex $(n \times n)$-matrices $\mathcal{B}$ with $\mathcal{B}^\dagger = -\mathcal{B}$ and $\operatorname{tr} \mathcal{B} = 0$. If $\mathcal{B}, \mathcal{C} \in su(n)$, then we have $[\mathcal{B}, \mathcal{C}]_- \in su(n)$. Set

$$\varrho(\mathcal{B}) := \mathrm{e}^{\mathcal{B}}.$$

The map $\varrho : su(n) \to SU(n)$ is surjective. The group $SU(n)$ is simply connected for $n = 2, 3, \ldots$.

The group $SU(2)$ is isomorphic to the group $Spin(3)$, that is, there exists a bijective map from $SU(2)$ onto $Spin(3)$ which respects products.

The map (5.21) induces the group epimorphism

$$\chi : SU(2) \to SO(3).$$

The linearization of this map at the unit element,

$$\chi'(I) : su(2) \to so(3),$$

is a Lie algebra isomorphism, that is, $\chi'(I)$ is a linear isomorphism which respects Lie products.[37]

(vii) Representations of $SO(3)$ and integer angular momentum. For each natural number $l = 0, 1, 2, \ldots$, there exist both a complex Hilbert space X_l of dimension $2l + 1$ and a group morphism

$$r : SO(3) \to GL(X_l),$$

that is, the map r respects products. Here, $GL(X_l)$ denotes the group of all linear isomorphisms from X_l onto itself. The linearization at the unit element

$$r'(I) : so(3) \to gl(X_l)$$

[36] The symbol $A^d = (b_{jk})$ denotes the dual (or transposed) matrix to the given matrix $A = (a_{jk})$. Explicitly, $b_{jk} := a_{kj}$ for $j, k = 1, 2, 3$. Instead of A^d, one also uses the symbols A^T, A^t, or ${}^tA$ in the literature.

[37] The definition of the derivative $\chi'(I)$ can be found in Sect. 7.20.1 on page 396.

is a Lie algebra morphism, that is, $r'(I)$ is a linear operator which respects Lie products. Here, $gl(X_l)$ consists of all linear operators on X_l equipped with the Lie product $[A, B]_- := AB - BA$.

(ix) Representations of $SU(2)$ and half-integer spin. For each given rational number $s = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \ldots$, there exist both a complex Hilbert space X_s of dimension $2s + 1$ and a group morphism

$$r : SU(2) \to GL(X_s)$$

such that the linearization at the unit element,

$$r'(I) : su(2) \to gl(X_s),$$

is a Lie algebra morphism. The elements of X_l (resp. X_s) represent quantum states with the angular momentum quantum number l (resp. the spin quantum number s.)

Summarizing, the Lie group $SO(3)$ and its universal covering group $SU(2)$ are locally isomorphic near the unit element, but they are not globally isomorphic.

This leads to different representations of the groups $SO(3)$ and $SU(2)$ which is responsible for the existence of the electron spin. The Lie algebras $so(3)$ and $su(2)$ are isomorphic. Therefore, the operators for both angular momentum and spin satisfy the same commutation rules in quantum physics. This will be studied in Volume III on quantum mathematics.

5.7.2 The Hopf Fibration of the 3-Dimensional Sphere

The space of quantum states of a quantum system possesses a nontrivial topology.

Folklore

In 1931, Heinz Hopf (1894–1971) proved the existence of a continuous map

$$\boxed{h : \mathbb{S}^3 \to \mathbb{S}^2} \tag{5.22}$$

from the 3-dimensional unit sphere onto the 2-dimensional unit sphere which is not homotopic to a constant map, that is, this map cannot be continuously deformed into a constant map without leaving the spheres during the deformation process.[38] This was a sensation in topology. The map h is called the Hopf map or the Hopf $U(1)$-principal fiber bundle. Hopf did not refer to physics. We want to show that the Hopf map possesses a quite natural physical interpretation in terms of the classification of quantum states of a non-relativistic electron. There arises the following fundamental question:

What is a quantum state?

[38] For $n = 1, 2, ...$, we define

$$\mathbb{S}^n := \{(x_1, ..., x_{n+1}) \in \mathbb{R}^{n+1} : x_1^2 + ... + x_{n+1}^2 = 1\}.$$

This is called the n-dimensional unit sphere in $\mathbb{R}^{n+1}$. Moreover, the set

$$\mathbb{B}^{n+1} := \{(x_1, ..., x_{n+1}) \in \mathbb{R}^{n+1} : x_1^2 + ... + x_{n+1}^2 \leq 1\}$$

is called the $(n + 1)$-dimensional closed unit ball in $\mathbb{R}^{n+1}$.

To answer this question, we will consider non-relativistic electrons. We will see that quantum states are highly nonlinear mathematical objects.

Quantum states. Let $\mathbb{C}^2$ denote the linear space of all the complex 1-column matrices

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$$

where ψ_1, ψ_2 are complex numbers. We also introduce the inner product

$$\langle \psi | \varphi \rangle := \sum_{j=1}^{2} \psi_j^\dagger \varphi_j \qquad \text{for all} \quad \varphi, \psi \in \mathbb{C}^2$$

along with the norm $||\psi|| := \sqrt{\langle \psi | \psi \rangle}$. Consider the unit sphere

$$\mathbb{S}(\mathbb{C}^2) := \{\psi \in \mathbb{C}^2 : ||\psi|| = 1\}$$

of the complex 2-dimensional Hilbert space $\mathbb{C}^2$. By definition, each element ψ of $\mathbb{S}(\mathbb{C}^2)$ is called the state of a non-relativistic electron. Writing $\psi_j = \alpha_j + \mathrm{i}\beta_j$ with real numbers α_j, β_j, we get

$$\langle \psi | \psi \rangle = |\psi_1|^2 + |\psi_2|^2 = \alpha_1^2 + \beta_1^2 + \alpha_2^2 + \beta_2^2 = 1$$

for all $\psi \in \mathbb{S}(\mathbb{C}^2)$. Hence $\mathbb{S}(\mathbb{C}^2)$ can be identified with the unit sphere,

$$\mathbb{S}(\mathbb{C}^2) = \mathbb{S}^3.$$

Two states $\varphi, \psi \in \mathbb{S}(\mathbb{C}^2)$ are called equivalent iff there exists a complex number $\lambda \in U(1)$ (i.e., $|\lambda| = 1$) such that $\varphi = \lambda\psi$. By a *quantum state* of the electron, we understand the equivalence class

$$[\psi] := \{\lambda\psi : \lambda \in U(1)\} \tag{5.23}$$

where $\psi \in \mathbb{S}(\mathbb{C}^2)$ and $|\lambda| = 1$. We have the fibration

$$\boxed{\mathbb{S}(\mathbb{C}^2) = \bigcup_{\psi \in \mathbb{S}(\mathbb{C}^2)} [\psi],}$$

that is, the sphere is decomposed into the sets $[\psi]$ which are called $U(1)$-fibers in mathematics. Consequently, the set of all quantum states coincides with the set of all $U(1)$-fibers of the unit sphere $\mathbb{S}(\mathbb{C}^2)$. This set is denoted by

$$\mathbb{S}(\mathbb{C}^2)/U(1).$$

In terms of orbits, this definition reads like this. Choose a point $\psi \in \mathbb{S}(\mathbb{C}^2)$. For each complex number $\lambda \in U(1)$, define a map

$$\varrho_\lambda : \mathbb{S}(\mathbb{C}^2) \to \mathbb{S}(\mathbb{C}^2)$$

given by $\varrho_\lambda(\psi) := \lambda\psi$. We say that the group $U(1)$ acts on the unit sphere $\mathbb{S}(\mathbb{C}^2)$. Naturally enough, the set $[\psi]$ from (5.23) is called an orbit through the point ψ. In this setting, the space of quantum states coincides with the space of $U(1)$-orbits on the unit sphere of the Hilbert space $\mathbb{C}^2$.

The Hopf map. Our next goal is to understand the topological structure of the space $\mathbb{S}(\mathbb{C}^2)/U(1)$ of quantum states. To this end, we choose the Pauli matrices

$$\sigma^1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^2 := \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma^3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and we define the spin operators $S^j := \frac{\hbar}{2}\sigma^j, j = 1, 2, 3$. For each state $\psi \in \mathbb{S}(\mathbb{C}^2)$, we define

$$\bar{S}^j := \langle \psi | S^j \psi \rangle, \qquad j = 1, 2, 3.$$

Choose a Cartesian coordinate system with the right-handed orthonormal basis vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$. From the physical point of view, if the electron is in the state ψ, then we measure the mean spin vector $\bar{\mathbf{S}} := \bar{S}^1\mathbf{i} + \bar{S}^2\mathbf{j} + \bar{S}^3\mathbf{k}$. We now construct a map

$$h : \mathbb{S}(\mathbb{C}^2) \to \mathbb{S}^2 \tag{5.24}$$

by setting

$$\boxed{h(\psi) := \frac{\bar{\mathbf{S}}(\psi)}{||\bar{\mathbf{S}}(\psi)||}.}$$

That is, we assign the measured normalized mean spin vector to each given state $\psi \in \mathbb{S}^2(\mathbb{C}^2)$. This measured value does not change if we pass to an equivalent state φ. Thus, the function h only depends on the equivalence class $[\psi]$. Setting

$$h([\psi]) := \frac{\bar{\mathbf{S}}(\psi)}{||\bar{\mathbf{S}}(\psi)||},$$

we obtain a well-defined induced map

$$h : \mathbb{S}(\mathbb{C}^2)/U(1) \to \mathbb{S}^2. \tag{5.25}$$

Now to the point. We claim that the following hold true.

Theorem 5.2 (i) *The map h introduced in* (5.24) *is a continuous map from the unit sphere $\mathbb{S}(\mathbb{C}^2)$ of the 2-dimensional complex Hilbert space $\mathbb{C}^2$ onto the 2-dimensional unit sphere $\mathbb{S}^2$ of the 3-dimensional Euclidean space.*

(ii) *The map h is essential, that is, it is not homotopic to a constant map.*

(iii) *There exists a one-to-one correspondence between the space of quantum states $\mathbb{S}(\mathbb{C}^2)/U(1)$ and the 2-dimensional unit sphere which is given by the induced map h from* (5.25).

The proof will be given in Volume IV on quantum mathematics (see also Knörrer (1996)). In terms of physics, this result tells us that

> *Two normalized vectors ψ and χ represent the same quantum state iff the measured normalized spin vectors coincide.*

This is a quite natural result. Using (5.24) and $\mathbb{S}^2(\mathbb{C}^2) = \mathbb{S}^3$, we get the desired Hopf map (5.22).

5.7.3 The Homotopy Functor

> Functors reduce the solution of topological problems to simpler algebraic problems.
>
> Folklore

The homotopy groups of a topological space. We are going to sketch how the Hopf map is related to the classification of continuous maps up to deformations. Let X be a topological space. Consider the continuous map

$$f : \mathbb{S}^k \to X, \qquad k = 1, 2, \ldots . \tag{5.26}$$

By definition, the homotopy class $[f]$ of the map f consists of all continuous maps $g : \mathbb{S}^k \to X$ that are homotopic to f. The symbol $\pi_k(X)$ denotes the set of all homotopy classes $[f]$ of maps of the form (5.26).

The set $\pi_k(X)$ can be equipped with the structure of a group.

This yields the kth homotopy group $\pi_k(X)$ of the topological space X. In particular, the group $\pi_1(X)$ is Poincaré's fundamental group. Note the following.

(i) The topological space X is simply connected iff $\pi_1(X) = 0$.

(ii) If $\pi_k(X) = 0$, then each continuous map $f : \mathbb{S}^k \to X$ is homotopic to a constant map. In this case, we say that the k-connectivity of the space X is trivial.

(iii) If $\pi_k(X) = \mathbb{Z}$, then it is possible to assign an integer to each continuous map $f : \mathbb{S}^k \to X$ which is called the mapping degree $\deg(f)$. Two continuous maps

$$f, g : \mathbb{S}^k \to X$$

have the same mapping degree iff they are homotopic. In other words, the mapping degree is a homotopy invariant.

(iv) Homotopically equivalent topological spaces have the same homotopy groups.

Let us consider the following examples.

- For the unit circle, $\pi_1(\mathbb{S}^1) = \mathbb{Z}$ and $\pi_k(\mathbb{S}^1) = 0,\ k = 2, 3, \ldots$ The mapping degree of $f : \mathbb{S}^1 \to \mathbb{S}^1$ is called the winding number.
- For the fundamental group of the 2-dimensional torus, $\pi_1(X) = \mathbb{Z} \oplus \mathbb{Z}$.
- For the 2-dimensional unit sphere,

$$\boxed{\pi_1(\mathbb{S}^2) = 0, \qquad \pi_2(\mathbb{S}^2) = \pi_3(\mathbb{S}^2) = \mathbb{Z}.}$$

- For the k-dimensional unit sphere with $k = 1, 2, 3, \ldots$, we get $\pi_k(\mathbb{S}^k) = \mathbb{Z}$.

The complete computation of all the homotopy groups of the spheres of arbitrary dimension is an open problem in topology.[39]

The mapping degree for smooth mappings. Let $k = 1, 2, \ldots$ For a smooth map $f : \mathbb{S}^k \to \mathbb{S}^k$, the mapping degree $\deg(f)$ is given by the Kronecker integral

$$\deg(f) = \int_{\mathbb{S}^k} f^* \upsilon$$

where υ denotes the volume form on $\mathbb{S}^k$ normalized by $\int_{\mathbb{S}^k} \upsilon = 1$. Let us consider some examples.

[39] A comprehensive list of known homotopy groups can be found in Dodson and Parker (1997) (appendix).

(a) Suppose that the smooth map $f : \mathbb{S}^1 \to \mathbb{S}^1$ is given by the function $\Phi = \Phi(\varphi)$ where φ and Φ are angle coordinates. Then $\upsilon = \frac{1}{2\pi} \cdot d\Phi$. Hence

$$\deg(f) = \int_{\mathbb{S}^1} f^*\upsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi'(\varphi) d\varphi = \frac{\Phi(\pi) - \Phi(-\pi)}{2\pi}.$$

Since Φ is an angle variable on the unit circle, the mapping degree is an integer. For example, if $\Phi(\varphi) = n\varphi$ for some integer n, then $\deg(f) = n$.

(b) Assume that the smooth map $f : \mathbb{S}^2 \to \mathbb{S}^2$ is given by the functions

$$\phi = \Phi(\varphi, \vartheta), \quad \Theta = \Theta(\varphi, \vartheta)$$

with respect to spherical coordinates. Then $\upsilon = \frac{1}{4\pi} \cos\Theta \, d\Phi \wedge d\Theta$. Hence

$$\deg(f) = \int_{\mathbb{S}^2} \upsilon = \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_{-\pi/2}^{\pi/2} \cos\Theta(\varphi, \vartheta) \, \frac{\partial(\Phi, \Theta)}{\partial(\varphi, \vartheta)} \, d\varphi d\vartheta.$$

Here, noting that $d\varphi \wedge d\vartheta = -d\vartheta \wedge d\varphi$, we use

$$\begin{aligned} d\Phi \wedge d\Theta &= (\Phi_\varphi d\varphi + \Phi_\vartheta d\vartheta) \wedge (\Theta_\varphi d\varphi + \Theta_\vartheta d\vartheta) \\ &= (\Phi_\varphi \Theta_\vartheta - \Phi_\vartheta \Theta_\varphi) \, d\varphi \wedge d\vartheta = \frac{\partial(\Phi, \Theta)}{\partial(\varphi, \vartheta)} \, d\varphi \wedge d\vartheta. \end{aligned}$$

(c) Suppose that the smooth map $f : \mathbb{S}^2 \to \mathbb{S}^2$ is described by the functions

$$X = X(\varphi, \vartheta), \quad Y = Y(\varphi, \vartheta), \quad Z = Z(\varphi, \vartheta)$$

where X, Y, Z are Cartesian coordinates. Then

$$\upsilon = \frac{X dY \wedge dZ + Y dZ \wedge dX + Z dX \wedge dY}{4\pi}.$$

For the mapping degree, this implies

$$\deg(f) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_{-\pi/2}^{\pi/2} \begin{vmatrix} X & Y & Z \\ X_\varphi & Y_\varphi & Z_\varphi \\ X_\vartheta & Y_\vartheta & Z_\vartheta \end{vmatrix} \, d\varphi d\vartheta.$$

This follows from

$$\begin{aligned} Z dX \wedge dY &= Z(X_\varphi d\varphi + X_\vartheta d\vartheta) \wedge (Y_\varphi d\varphi + Y_\vartheta d\vartheta) \\ &= Z(X_\varphi Y_\vartheta - Y_\varphi X_\vartheta) \, d\varphi \wedge d\vartheta \end{aligned}$$

along with the cyclic permutation $X \mapsto Y \mapsto Z \mapsto X$.

(d) For a smooth map $f : \mathbb{S}^3 \to \mathbb{S}^2$, the mapping degree can be computed by using the formula

$$\deg(f) = \int_{\mathbb{S}^3} f^*\upsilon \wedge \omega$$

along with the normalized volume form υ on $\mathbb{S}^2$ and the 1-form

$$\omega := \frac{8\psi^2 \sin\vartheta \, d\psi}{\pi(1+\psi^2)^3} + \frac{2\psi^2 \cos^2\vartheta \, d\varphi}{\pi(1+\psi^2)^2}.$$

Here, φ, ϑ, ψ denote spherical coordinates on $\mathbb{S}^3$ where $-\pi < \varphi \leq \pi$ along with $-\frac{\pi}{2} \leq \vartheta \leq \frac{\pi}{2}$ and $0 \leq \psi \leq \pi$. The Hopf map has the mapping degree one.[40]

[40] The proof can be found in Schwarz (1994), p. 164 and Spanier (1989), p. 489.

The homotopy functor and the mapping degree for continuous maps. Let

$$f : X \to Y$$

be a continuous map from the topological space X into the topological space Y. It can be shown that for each $k = 1, 2, \ldots$, the map f induces a group epimorphism

$$f^* : \pi_k(X) \to \pi_k(Y)$$

in a quite natural manner. The following crucial properties hold.

(C) Composition of maps: For continuous maps $f : X \to Y$ and $g : Y \to Z$, the composition $g \circ f$ corresponds to the composition $f^* \circ g^*$.
(I) Identity map: The identity map on X corresponds to the identity map on $\pi_k(X)$ for all k.
(H) Homeomorphism: If $f : X \to Y$ is a homeomorphism, then the induced map $f^* : \pi_k(X) \to \pi_k(Y)$ is a group isomorphism for all k.

This homotopy functor allows us to define the mapping degree for continuous maps

$$f : \mathbb{S}^k \to \mathbb{S}^k, \qquad k = 1, 2, \ldots$$

In fact, the homotopy functor generates the group epimorphism

$$f^* : \pi_k(\mathbb{S}^k) \to \pi_k(\mathbb{S}^k).$$

where $\pi_k(\mathbb{S}^k) = \mathbb{Z}$. We now define $\deg(f) := f^*(1)$.

Generalization of the mapping degree. Consider a continuous map

$$F : \mathbb{S}^k \to \mathbb{R}^{k+1} \setminus \{0\}, \qquad k = 1, 2, \ldots.$$

This yields the normalized map

$$f : \mathbb{S}^k \to \mathbb{S}^k, \qquad k = 1, 2, \ldots$$

by setting $f(x) := F(x)/||F(x)||$. We now define the mapping degree of the mapping F by

$$\deg(F) := \deg(f).$$

Suppose that we are given the smooth map

$$F : \mathbb{B}^{k+1} \to \mathbb{R}^{k+1}, \qquad k = 1, 2, \ldots$$

on the closed $(k+1)$-dimensional unit ball such that the following are met:

- $F(x) \neq 0$ on the boundary $\mathbb{S}^k$.
- The map F has precisely the zeros $x_1, \ldots, x_N$ on $\mathbb{B}^{k+1}$.
- The zeros are regular, that is, $\det F'(x_j) \neq 0$ for all j.[41]

Then, the mapping degree can be computed by the simple formula

$$\deg(F) = \operatorname{sgn} \det F'(x_1) + \ldots + \operatorname{sgn} \det F'(x_N).$$

This shows that the mapping degree generalizes the 1-dimensional concept introduced on page 228

The homotopy functor is the prototype of functors in algebraic topology. The crucial homology and cohomology functors will be studied in Volume IV on quantum mathematics.

[41] The symbol $\det F'(x_j)$ denotes the Jacobian of the map F at the point x_j; this is the determinant of the first partial derivatives of the components of F at the point x_j.

These functors allow us to reduce the study of topological spaces and continuous maps to the study of groups and group morphisms.

To illustrate this, let us prove that there is no homeomorphism

$$f : \mathbb{B}^2 \to \mathbb{S}^1$$

In fact, suppose that there exists such a homeomorphism. Then, the map

$$f^* : \pi_1(\mathbb{B}^2) \to \pi_1(\mathbb{S}^1) \tag{5.27}$$

is a group isomorphism. Since the unit ball is simply connected, $\pi_1(\mathbb{B}^2) = 0$. However, $\pi_1(\mathbb{S}^1) = \mathbb{Z}$. Thus, there is no group isomorphism of the form (5.27). This is the desired contradiction.

5.7.4 Grassmann Manifolds and Projective Geometry

Grassmann manifold. By a 1-dimensional linear subspace of the complex Hilbert space $\mathbb{C}^2$, we understand a set of the form

$$L(\psi) := \{\lambda\psi : \lambda \in \mathbb{C}\}$$

where ψ is a fixed nonzero element of $\mathbb{C}^2$. Without loss of generality, we can choose ψ in such a way that $||\psi|| = 1$. Then

$$\boxed{\mathbb{C}^2 = \bigcup_{\psi \in \mathbb{S}(\mathbb{C}^2)} L(\psi).}$$

Obviously, the map $L(\psi) \mapsto [\psi]$ is a one-to-one correspondence between the set of linear 1-dimensional subspaces $L(\psi)$ and the set of quantum states $[\psi]$. Therefore, a quantum state can also be defined as a linear 1-dimensional subspace of $\mathbb{C}^2$. In mathematics, the space of all one-dimensional subspaces of $\mathbb{C}^2$ is called the 1-Grassmann manifold $\mathbb{G}_1(\mathbb{C}^2)$ of $\mathbb{C}^2$.

Rays and projective space. By a ray, we understand a set of the form

$$R(\psi) := \{\lambda\psi : \lambda \in \mathbb{C},\ \lambda \neq 0\}$$

where $\psi \in \mathbb{S}(\mathbb{C}^2)$. The map

$$R(\psi) \mapsto [\psi]$$

yields a one-to-one correspondence between the set of rays $\mathbb{P}(\mathbb{C}^2)$ and the set of quantum states. In mathematics, the set of rays $L(\psi)$ is also called the complex 1-dimensional projective space $\mathbb{P}^1_{\mathbb{C}}$.

Summarizing, there exist the following three equivalent possibilities for defining the space of quantum states of the Hilbert space $\mathbb{C}^2$, namely,

(i) the orbit space $S(\mathbb{C}^2)/U(1)$ of the unit sphere $\mathbb{S}(\mathbb{C}^2)$ under the action of the group $U(1)$ on $\mathbb{S}(\mathbb{C}^2)$,
(ii) the 1-Grassmann manifold $\mathbb{G}_1(\mathbb{C}^2)$ of $\mathbb{C}^2$ which consists of all complex one-dimensional subspaces of $\mathbb{C}^2$,
(iii) the complex one-dimensional projective space $\mathbb{P}^1_{\mathbb{C}}$ of rays in $\mathbb{C}^2$.

These objects are well-known in algebraic geometry. In some sense, projective spaces are the simplest geometric objects which possess a nontrivial topology. Explicitly, we have the following homeomorphisms:

$$\boxed{\mathbb{S}(\mathbb{C}^2)/U(1) \cong \mathbb{S}^2 \cong \mathbb{G}_1(\mathbb{C}^2) \cong \mathbb{P}^1_{\mathbb{C}}.}$$

The complex projective space $\mathbb{P}^1_{\mathbb{C}}$ is a Kähler manifold. This is proved in Jost (2002a), Chap. 5. The Hopf bundle is an example for the following fascinating development in sciences:

> *Physicists and mathematicians have studied completely different deep questions posed by nature and by intrinsic mathematics, respectively. Finally, they have arrived at the same highly nontrivial mathematical tools.*

5.8 Perspectives

For further prototypes of important topological phenomena in physics, we refer to the following sections of the present volume:

- cohomology and potentials of physical fields (Sect. 16.8.2);
- the cohomology of the unite circle, the unit sphere, and the torus (Sect. 16.8.3);
- cohomology and atomic spectra (Sect. 16.8.4);
- cohomology and BRST symmetry (the classification of physical states and the elimination of ghosts, and the cohomology of Lie groups) (Sect. 16.8.5).

Much material on topology can be found in Volume IV on quantum mathematics. In particular, we will investigate there homology groups, cohomology groups, homotopy groups, and characteristic classes. In particular, we will show that the fundamental concepts of homology and cohomology are rooted in the properties of electric circuits.

6. Many-Particle Systems in Mathematics and Physics

> Partition functions are the main tool for studying many-particle systems.
>
> Folklore

Many-particle systems play a fundamental role in both mathematics and physics.

- In physics, we encounter systems of molecules (e.g., gases or liquids) or systems of elementary particles in quantum field theory.
- In mathematics, for example, we want to study the system of prime numbers.

In the 19th century, physicists developed the methods of statistical mechanics for studying many-particle systems, whereas mathematicians proved the distribution law for prime numbers. It turns out that the two apparently different approaches can be traced back to the same mathematical root, namely, the notion of partition function. In modern quantum field theory, the Feynman functional integral can be viewed as a partition function, as we will discuss later on. The typical procedure proceeds in the following two steps.

(i) Coding: The many-particle system is encoded into one single function called a partition function (e.g., the Boltzmann partition function in statistical physics, Riemann's zeta function or Dirichlet's L-function for describing prime numbers, the Feynman functional integral in quantum field theory). The idea goes back to Euler; in 1737 he proved the identity

$$\prod_p \left(1 - \frac{1}{p^s}\right)^{-1} = \sum_{n=1}^{\infty} \frac{1}{n^s} = \zeta(s)$$

for all real numbers $s > 1$. The product refers to all prime numbers p.

(ii) Decoding: The task is to get crucial information about the many-particle system by studying the properties of the partition function. The idea goes back to Riemann. He recognized that the zeta function ζ extends holomorphically to the punctured complex plane $\mathbb{C} \setminus \{0\}$, and that the detailed knowledge on the distribution of the zeros of the zeta function allows far-reaching statements about the asymptotic distribution of the prime numbers. This is related to the famous Riemann hypothesis to be considered on page 296.

The complexity of justifying the Riemann hypothesis and the mathematical trouble with the Feynman path integral prove that step (ii) is much more complex than step (i). By analytic continuation, the series

$$\sum_{n=1}^{\infty} \frac{1}{n^s}$$

can be assigned the value $\zeta(s)$ for all complex numbers s with $s \neq 1$. In particular, for the uncritical value $s = 0$, we can assign the value

$$\boxed{1+1+1+\ldots=\zeta(0)=-\frac{1}{2}.} \tag{6.1}$$

Note that the series $1+1+1\ldots$ is divergent in the classical sense. Furthermore,

$$\sum_{n=1}^{\infty} n=\zeta(-1)=-\frac{1}{12}.$$

Interestingly enough, this formula is closely related to a physical experiment called the Casimir effect. This crucial effect in quantum field theory will be studied in Sect. 6.6. More generally, one has the identity[1]

$$\zeta(s)=\frac{1}{s-1}+\frac{1}{2}+\frac{s}{12}-\frac{s(s+1)(s+2)}{720}+\ldots+$$
$$+\frac{B_n}{n!}\cdot s(s+1)\cdots(s+n-2)$$

for $s=0,-1,-2\ldots,-n+1$ and $n=2,3,\ldots$ Here, the Bernoulli numbers, B_n, are defined by

$$\boxed{\frac{z}{\mathrm{e}^z-1}=\sum_{k=0}^{\infty}\frac{B_k}{k!}\,z^k, \qquad 0<|z|<2\pi.}$$

This function is called the generating function of the Bernoulli numbers.[2] Recall that this function enters the 1900 Planck radiation law on page 101. Explicitly,

$$B_0=1,\ \ B_1=-\frac{1}{2},\ \ B_2=\frac{1}{6},\ \ B_4=-\frac{1}{30},\ \ B_6=\frac{1}{42},\ldots$$

and $B_n=0$ for $n=3,5,7,\ldots$ This yields

$$\zeta(0)=-\frac{1}{2},\quad \zeta(-1)=-\frac{1}{12},\quad \zeta(-2)=0,\quad \zeta(-3)=\frac{1}{120}.$$

More generally,

$$\zeta(1-2m)=-\frac{B_{2m}}{2m}, \qquad \zeta(-2m)=0, \qquad m=1,2,3,\ldots$$

In 1734 Euler proved the famous formulas

$$\zeta(2)=\frac{\pi^2}{6}, \qquad \zeta(4)=\frac{\pi^4}{90}.$$

More generally,

$$\zeta(2m)=\frac{(-1)^{m-1}2^{2m-1}B_{2m}}{(2m)!}\cdot\pi^{2m}, \qquad m=1,2,\ldots$$

The properties of $\zeta(2m+1)$ for $m=1,2,3,..$ represent a famous open problem of mathematics. In 1986 Roger Apéry created a sensation by proving the irrationality of the number $\zeta(3)=1.2020569\ldots$[3]

[1] The proof can be found in Zagier (1981), p. 26.

[2] These numbers were introduced by Jakob Bernoulli (1654–1705).

[3] See the survey article by M. Kontsevich and D. Zagier, Periods. In: B. Enquist and W. Schmid (Eds.) (2001), pp. 771–808. This article contains a wealth of sophisticated material on the values of integrals over rational integrands and their properties. Such numbers are generally called periods. Prototypes are the number π and the periods of elliptic integrals.

As an introduction to the Riemann zeta function and its numerous relations to other mathematical subjects, we recommend the book by J. Havil, Gamma: Exploring Euler's Constant, Princeton University Press, 2003. Recommendations for further reading can be found on page 298.

6.1 Partition Function in Statistical Physics

Statistical physics is based on the partition function

$$\boxed{Z(T,\mu,V) := \sum_{m=1}^{M} \mathrm{e}^{(\mu N_m - E_m)/kT}\,.} \tag{6.2}$$

Here, we assume that the physical system under consideration can be located in precisely M different states $\mathcal{S}_1, \ldots, \mathcal{S}_M$. Each state $\mathcal{S}_m$ is characterized by

- the energy E_m, and
- the particle number N_m.

Energy and particle number depend on the volume V of the system. The positive real parameter T is called absolute temperature, and the real parameter μ is called chemical potential. Finally, k denotes the Boltzmann constant. In SI units, $k = 1.380 \cdot 10^{-23}$ Joule/Kelvin. The point is that the knowledge of the partition function allows us to compute all of the important thermodynamic quantities of the physical system. Let us summarize this. The probability of finding the system in the state $\mathcal{S}_m$ is given by

$$\boxed{p_m := \frac{\mathrm{e}^{(\mu N_m - E_m)/kT}}{Z(T,\mu,V)}\,, \qquad m = 1, \ldots, M.}$$

This implies the mean energy

$$\boxed{\overline{E} := \sum_{m=1}^{M} p_m E_m,}$$

the square of the energy fluctuation

$$(\Delta E)^2 := \sum_{m=1}^{M} p_m (E_m - \overline{E})^2,$$

the mean particle number

$$\boxed{\overline{N} := \sum_{m=1}^{M} p_m N_m,}$$

and the square of the particle number fluctuation

$$(\Delta N)^2 := \sum_{m=1}^{M} p_m (N_m - \overline{N})^2.$$

The fundamental quantity

$$\boxed{S := -k \sum_{m=1}^{M} p_m \ln p_m}$$

is called the entropy of the physical system. We will show in Volume IV on quantum mathematics that the entropy measures the disorder of the system.

- The entropy S is minimal if, say, $p_1 = 1$, $p_2 = p_3 = \ldots = p_M = 0$ (maximal order). Then $S = 0$.
- The entropy S is maximal if $p_1 = p_2 = \ldots = p_M = \frac{1}{M}$ (maximal disorder). Then $S = kM \ln M$.

In addition, we will show in Volume IV that the entropy S is equal to the mean number of questions with yes-no answers which are necessary in order to determine the state of the physical system. Thus, the entropy is equal to the mean information gained after measuring the state of the physical system. In terms of biology, the complexity of a biological system increases if its order increases, and hence its entropy decreases.

The statistical potential. The function

$$\boxed{\Omega(T, \mu, V) := -kT \ln Z(T, \mu, V)}$$

is called the statistical potential (or the Gibbs potential) of the physical system. We get the partial derivatives

$$S = -\Omega_T, \qquad \overline{N} = -\Omega_\mu, \qquad P = -\Omega_V$$

where P denotes the pressure of the system. Moreover, we obtain the mean free energy

$$\boxed{\overline{F} := \Omega + \mu \overline{N},}$$

and the mean inner energy $\overline{E} = \overline{F} + ST$. Physicists introduce the fugacity

$$z := \mathrm{e}^{\mu/kT}.$$

Using this, the partition function can be written as

$$Z(T, \mu, V) = \sum_{N=0}^{N_{\max}} z^N Z_N(T, \mu, V)$$

where we set

$$Z_N(T, \mu, V) := \sum_{N_m = N} \mathrm{e}^{-E_m/kT}.$$

Here, we sum over the energies E_m of all the possible states which correspond to N particles.

The extended statistical potential. Replacing the energy E_m by the sum $E_m + J_m$, we introduce

$$\Omega_+(\mu, T, V; J) := -kT \ln \sum_{m=1}^{M} \mathrm{e}^{(\mu N_m - E_m - J_m)/kT}.$$

The additional real parameters $J_1, \ldots, J_M$ are called source parameters by physicists. For the probability, we then get

$$\boxed{p_m = \frac{\partial \Omega_+}{\partial J_m}\Big|_{J_1=\ldots=J_M=0}, \qquad m = 1, \ldots, M.}$$

The ground state energy formula. Consider the case where the particle number is fixed, that is, $\mu = 0$. Then

$$Z = \sum_{m=1}^{M} \mathrm{e}^{-E_m/kT} = \mathrm{e}^{-E_1/kT}\left(1 + \sum_{m=2}^{M} \mathrm{e}^{-(E_m-E_1)/kT}\right).$$

Suppose that the ground state energy E_1 is not degenerate. This means that we have $E_1 < E_2 \leq \ldots \leq E_m$. Then, for fixed volume V,

$$\boxed{E_1 = -\lim_{T\to+\infty} kT \ln Z(T, V).} \tag{6.3}$$

This formula is frequently used by physicists in order to compute the ground state energy of non-degenerate ground states.

The Riemann zeta function represents a special partition function. To see this, choose the chemical potential $\mu = 0$ and the special energy levels

$$E_n := skT \ln n, \qquad n = 1, 2, \ldots$$

for fixed real number $s > 1$. Then

$$Z = \sum_{n=1}^{M} \mathrm{e}^{-E_n/kT} = \sum_{n=1}^{M} \frac{1}{n^s}.$$

This special partition function is a truncation of the Riemann zeta function

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}, \qquad \Re(s) > 1.$$

Von Neumann's density matrix. Let us formulate the partition function in the language of the density matrix which was introduced by John von Neumann (1903–1957) around 1930. Define the Hamiltonian H and the particle number operator N by letting

$$H := \begin{pmatrix} E_1 & & 0 \\ & E_2 & \\ & & \ddots \\ 0 & & E_M \end{pmatrix}, \qquad N := \begin{pmatrix} N_1 & & 0 \\ & N_2 & \\ & & \ddots \\ 0 & & N_M \end{pmatrix}.$$

Then, for the partition function,

$$\boxed{Z(T, \mu, V) = \operatorname{tr} \mathrm{e}^{(\mu N - H)/kT}.} \tag{6.4}$$

The matrix

$$\varrho := \begin{pmatrix} p_1 & & 0 \\ & p_2 & \\ & & \ddots \\ 0 & & p_M \end{pmatrix}$$

is called the density matrix. Explicitly,

$$\boxed{\varrho = \frac{\mathrm{e}^{(\mu N - H)/kT}}{Z(T, \mu, V)} \, .}$$

For the mean energy and the mean particle number,

$$\boxed{\overline{E} = \mathrm{tr}(\varrho H), \qquad \overline{N} = \mathrm{tr}(\varrho N).}$$

In the language of the Dirac calculus,

$$H = \sum_{m=1}^{M} E_M |m\rangle\langle m|, \qquad N = \sum_{m=1}^{M} N_m |m\rangle\langle m|$$

where $|1\rangle, |2\rangle, \ldots, |M\rangle$ form a complete orthonormal system of joint eigenvectors of the operators H and N. Explicitly,

$$|1\rangle := \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad , \ldots , \qquad |M\rangle := \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} .$$

Hence $H|m\rangle = E_m|m\rangle$ and $N|m\rangle = N_m|m\rangle$. Furthermore, note that matrix multiplication yields

$$|1\rangle\langle 1| = \begin{pmatrix} 1\ 0 \ldots 0 \\ 0\ 0 \ldots 0 \\ \vdots \\ 0\ 0 \ldots 0 \end{pmatrix}$$

along with similar matrices for $|m\rangle\langle m|$. The density matrix ϱ can be written as

$$\boxed{\varrho = \sum_{m=1}^{M} p_m |m\rangle\langle m|}$$

where

$$p_m = \frac{\langle m|\mathrm{e}^{(\mu N - H)/kT}|m\rangle}{\mathrm{tr}\, \mathrm{e}^{(\mu N - H)/kT}} \tag{6.5}$$

is the probability of finding the physical system in the state $|m\rangle$.

Historical remarks. Classical statistical physics was created by Maxwell (1831–1879) and Boltzmann (1844–1906); it was completed by Gibbs (1839–1903) who was professor for mathematical physics at Yale University (Connecticut, U.S.A.). Gibbs used the following classical partition function

$$Z_{\mathrm{class}} = \int_{\mathbb{R}^6} \mathrm{e}^{-E(\mathbf{x},\mathbf{p})/kT} \, \frac{d^3x \, d^3p}{h^3}$$

for sufficiently high temperature T. Here, $\mathbf{x}$ and $\mathbf{p}$ are the position vector and the momentum vector of a particle in 3-dimensional Euclidean space, respectively. Since Z_{class} has to be a dimensionless physical quantity, we introduce an arbitrary

constant h^3 where h has the physical dimension of action. In classical physics, there is no such distinguished physical constant. However, the situation changes completely in quantum physics. In 1900 Planck (1858–1947) introduced his quantum of action, h. We now have the following heuristic principle which is successfully used by physicists:

> *The high-temperature approximation of quantum statistics is obtained from classical statistical physics by subdividing the phase space into cells of volume h^3.*

For example, we will use this principle of phase space quantization for computing the cross section of scattering processes on page 838. Important contributions to the development of quantum statistics were made by Einstein (1879–1955), Pauli (1900–1958) (exclusion principle), Bose (1894–1974) (Bose–Einstein statistics for bosons), Dirac (1902–1984), Fermi (1901–1954) (Dirac–Fermi statistics for fermions), von Neumann (1903–1957) (density matrix), Lev Landau (1908–1968) (low temperatures and liquid helium), and Schwinger (1918–1994) (Green's function and KMS states of thermodynamic equilibrium). In fact, high-technology is based on solid state physics and quantum statistics (e.g., semiconductors, superconductors, and lasers). Moreover, the structure of stars (e.g., white dwarfs, neutron stars, supernovae, black holes), and the evolution of the universe after the Big Bang can only be understood by using the methods of quantum statistics. This can be found in Zeidler (1986), Vol. IV. We will come back to this fascinating topic in Volume III of this treatise.

6.2 Euler's Partition Function

> Read Euler, he is the master of us all.
> Marquis de Pierre Simon Laplace[4]

We want to study the decomposition of natural numbers into the sum of natural numbers. This is a basic problem in additive number theory. To this end, let n be a nonzero natural number. We define

$$p(n) := \begin{array}{l}\text{number of decompositions of } n\\ \text{into a sum of nonzero natural numbers.}\end{array}$$

For example, one has $p(3) = 3$, since

$$3 = 1 + 1 + 1, \quad 3 = 2 + 1, \quad 3 = 3.$$

We also set $p(0) := 1$. Following Euler, we define the partition function

$$P(q) := \sum_{n=0}^{\infty} p(n) q^n.$$

This represents the generating function for $p(0), p(1), p(2), \ldots$

Euler's theorem. For all complex numbers q with $|q| < 1$, one has the convergent product representation

[4] Euler (1707–1783), Laplace (1749–1827), Fourier (1768–1830), Hardy (1877–1947), Ramanujan (1887–1920), Hecke (1887–1947), Rademacher (1892–1969).

$$P(q) = \frac{1}{\prod_{n=1}^{\infty}(1-q^n)}$$

together with

$$\prod_{n=1}^{\infty}(1-q^n) = \sum_{n=-\infty}^{\infty}(-1)^n q^{\frac{3n^2+n}{2}} = 1 - q - q^2 + q^5 + q^7 + \dots$$

After finding the formulas numerically, Euler had to work a long time to get a rigorous proof.

The Hardy–Ramanujan theorem. For large numbers n, we have the asymptotic equation

$$p(n) \simeq \frac{\mathrm{e}^{\pi\sqrt{2n/3}}}{4n\sqrt{3}}, \qquad n \to \infty.$$

This theorem was obtained by Hardy and Ramanujan in 1918. Rademacher discovered in 1937 that $p(n)$ can be expanded in a convergent series in n. His proof used the Dedekind eta function

$$\eta(\tau) := \mathrm{e}^{\pi \mathrm{i}\tau/12}\prod_{n=1}^{\infty}(1-q^n) \qquad \text{with} \quad q := \mathrm{e}^{2\pi\tau\mathrm{i}}.$$

The proof can be found in T. Apostol, Modular Functions and Dirichlet Series in Number Theory, Springer, New York, 1990. Let $\mathbb{C}_> := \{\tau \in \mathbb{C} : \ \Im(\tau) > 0\}$ be the open upper half-plane. The Dedekind eta function is holomorphic on $\mathbb{C}_>$, and it has the symmetry property

$$\eta\left(\frac{a\tau+b}{c\tau+d}\right) = \varepsilon(c\tau+d)^{\frac{1}{2}}\eta(\tau) \qquad \text{for all} \quad \tau \in \mathbb{C}_> \tag{6.6}$$

and all modular transformations (i.e., for all integers a, b, c, d with $ad - bc = 1$). Here, ε is a 24th root of unity (i.e., $\varepsilon^{24} = 1$).

Modular forms of weight k. Let k be an integer. By definition, a modular form f of weight k has the following properties.

- The function $f : \mathbb{C}_> \to \mathbb{C}$ is holomorphic.
- The function f satisfies the symmetry condition

$$f\left(\frac{a\tau+b}{c\tau+d}\right) = (c\tau+d)^k f(\tau) \qquad \text{for all} \quad \tau \in \mathbb{C}_>$$

  and all modular transformations.
- The Fourier expansion $f(\tau) = \sum_n a_n \mathrm{e}^{2\pi\mathrm{i}\tau}$ has the property that $a_n = 0$ for all $n = -1, -2, \dots$

Typically, modular forms satisfy sophisticated relations. Don Zagier writes in his beautiful survey article on modular forms:[5]

> The importance of modular forms stems from the conjunction of the following two facts.
> (i) They arise naturally in a wide variety of contexts in mathematics and physics and often encode the arithmetically interesting information about a problem.

[5] Introduction to modular forms, pp. 238–291. In: M. Waldschmidt et al. (Eds.), From Number Theory to Physics, Springer, Berlin, 1995 (reprinted with permission).

(ii) The space M_k of modular forms of weight k is finite-dimensional for each integer k.

The point is that if $\dim M_k = d$ and we have more than d situations giving rise to modular forms in M_k, then we automatically have a linear relation among these functions and hence get 'for free' information – often highly nontrivial – relating these different situations... What we meant by (i) above is that nature – both physically and mathematically – often produces situations described by numbers which turn out to be the Fourier coefficients of a modular form. These can be as disparate as multiplicities of energy levels, numbers of vectors in a lattice of given length, sums over the divisors of integers, special values of zeta functions, or numbers of solutions of Diophantine equations. But the fact that all of these different objects land in the little space M_k forces the existence of relations among them...

The key to the rich internal structure of the theory of modular forms is the existence of a commutative algebra (Hecke algebra) of operators T_n ($n \in \mathbb{N}$) acting on the space M_k. The space M_k has a canonical basis of simultaneous eigenvectors of all the Hecke operators T_n; these special modular forms have the property that their Fourier coefficients $a(n)$ are algebraic integers and satisfy the multiplicative property $a(nm) = a(n)a(m)$ whenever n and m are relative prime.

6.3 Discrete Laplace Transformation

The two key formulas of the discrete Laplace transformation are given by

$$\boxed{f(z) = \sum_{n=0}^{\infty} a_n z^n} \tag{6.7}$$

along with the inverse formula

$$\boxed{a_n = \frac{1}{2\pi \mathrm{i}} \int_{|z|=r} \frac{f(z)}{z^{n+1}}\, dz, \qquad n = 0, 1, 2, \ldots} \tag{6.8}$$

for fixed radius $r > 0$. The transformation

$$f = \mathcal{L}(a_0, a_1, a_2, \ldots)$$

sends each sequence (a_0, a_1, a_2) of complex numbers to a formal power series expansion f. We want to study the properties of the map $\mathcal{L}$.

The Tauberian theorem. If the function f is holomorphic on the open unit disc $\{z \in \mathbb{C} : |z| < 1\}$ and it has the power series expansion (6.7), then the function f is called the generating function of the sequence $(a_0, a_1, \ldots)$. In this case, we have the inverse transformation formula (6.8) for each real number $r \in]0, 1[$. We now want to investigate the formula

$$\boxed{\lim_{\varepsilon \to +0} f(1 - \varepsilon) = \sum_{n=0}^{\infty} a_n} \tag{6.9}$$

where ε is a positive real parameter. Mnemonically, this can be written as

$$\lim_{z\to 1-0} f(z) = f(1).$$

(i) The 1826 Abelian theorem: Relation (6.9) is valid if the right-hand series is convergent.
(ii) The 1897 Tauberian theorem: Relation (6.9) is valid if the left-hand limit exists and

$$a_n = o\left(\frac{1}{n}\right), \qquad n \to \infty. \tag{6.10}$$

Theorem (i) describes a specific property of the map $\mathcal{L}$, whereas theorem (ii) refers to the inverse map $\mathcal{L}^{-1}$. These two theorems are due to Abel (1802–1829) and Tauber (1866–1942).[6] In 1911, Hardy (1877–1947) and Littlewood (1885–1977) proved a sophisticated result saying that the condition (6.10) can be replaced by the weaker condition

$$a_n = O\left(\frac{1}{n}\right), \qquad n \to \infty.$$

The theorems (i) and (ii) are prototypes of numerous sophisticated mathematical theorems about sums and integrals. This is summarized in the modern standard text by Korevaar (2004). The transformation

$$z = \mathrm{e}^{\mathrm{i}E}, \qquad \Im(E) > 0$$

maps the open upper half-plane conformally onto the open unit disc. In terms of the variable E, the discrete Laplace transform looks like

$$F(E) = \sum_{n=0}^{\infty} a_n \mathrm{e}^{\mathrm{i}En}, \qquad \Im(E) > 0.$$

The corresponding integral transform reads as

$$\boxed{G(E) = \int_0^{\infty} a(t)\mathrm{e}^{\mathrm{i}Et/\hbar}dt, \qquad \Im(E) > 0.}$$

This is called the Laplace transform. In physics, the variable t (resp. E) has the meaning of time (resp. energy). Engineers use the Laplace transform for computing electric circuits. Then, we set $E = \omega\hbar$ where the variable ω corresponds to frequency. The Laplace transform describes a crucial duality between time and energy (resp. time and frequency).

The rabbit problem and Fibonacci numbers. The discrete Laplace transform can be used in order to solve difference equations. Let us consider a classic problem. The most accomplished mathematician of the Middle Ages was Leonardo de Pisa (1175–1250) called Fibonacci who lived in Italy. Fibonacci is a shortened form of Filius Bonaccio (son of Bonaccio). In his famous 1202 *Liber Abaci* (Book of the Abacus), Fibonacci introduced the Arabic system of numbers to Europe.[7] Fibonacci formulated and solved the rabbit problem which leads to the discrete dynamical system

[6] Alfred Tauber was murdered in the German concentration camp Theresienstadt (Terežin).
[7] The decimal system was introduced much later on the European mainland by the Dutch engineer Simon Stevin in 1585.

$$\boxed{N((n+2)\Delta t) = N((n+1)\Delta t) + N(n\Delta t), \qquad n = 0, 1, 2, \ldots} \tag{6.11}$$

Explicitly, the rabbit problem reads as follows:

- A pair of adult rabbits produces a pair of baby rabbits once each month.
- Each pair of baby rabbits requires one month to grow to be adults and subsequently produces one pair of baby rabbits each month thereafter.
- Determine the number $N(n\Delta t)$ of adult rabbit pairs after n months. In this model, Fibonacci assumed that rabbits are immortal.

Noting that the number of pairs of baby rabbits is equal to the number of adult rabbit pairs in the previous month, we get the difference equation (6.11). For given initial values $N(0) = 0$ and $N(\Delta t) = 1$, one checks explicitly that the solution is given by the formula

$$\boxed{N(n\Delta t) = \frac{1}{\sqrt{5}} \left(r^n - \frac{(-1)^n}{r^n} \right), \qquad n = 0, 1, \ldots} \tag{6.12}$$

due to Binnet (1786-1856). The irrational number

$$\boxed{r = \frac{1}{2}(1 + \sqrt{5}) = 0.681\ldots}$$

is the positive solution of the equation

$$\lambda^2 - \lambda - 1 = 0. \tag{6.13}$$

which is called the characteristic equation to the original difference equation (6.11). The nonnegative integers $N(n\Delta t)$ are called the Fibonacci numbers. The first Fibonacci numbers read as

$$0,\ 1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ 34,\ 55,\ 89,\ 144,\ 233,\ 377,\ 610.$$

For example, after 15 months the number of adult rabbit pairs equals 610. For the mathematicians in the 13th century, it was a great surprise that the integer Fibonacci numbers are generated by the irrational numbers r and $-\frac{1}{r}$. Obviously, we have the asymptotic law

$$\lim_{n\to\infty} \frac{N(n\Delta t + \Delta t)}{N(n\Delta t)} = r.$$

For large times, we get the asymptotic equation[8]

$$N(n\Delta t) \simeq \frac{r^n}{\sqrt{5}} \qquad \text{as} \quad n \to \infty.$$

Let us motivate the solution by using two different methods.

(i) *The ansatz method*: Letting $N(n\Delta T) := \lambda^n$, we get $\lambda^{2+n} = \lambda^{1+n} + \lambda^n$, implying the characteristic equation (6.13) which has the two solutions $\lambda_\pm := \frac{1}{2}(1 \pm \sqrt{5})$. Hence $\lambda_+ = r$ and $\lambda_- = -\frac{1}{r}$. By superposition,

$$N(n\Delta t) = a\lambda_+^n + b\lambda_-^n.$$

[8] Recall that $f(n) \simeq g(n)$ as $n \to \infty$ means that $\frac{f(n)}{g(n)} \to 1$ as $n \to \infty$.

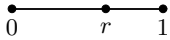


Fig. 6.1. The golden ratio

From the initial conditions $N(0) = 0$ and $N(\Delta t) = 1$, we get the constants a and b, giving the solution (6.12). This solution method resembles the ansatz $N(t) := \mathrm{e}^{\lambda t}$ for solving the ordinary differential equation

$$N'' = N' + N$$

which leads to the characteristic equation (6.13) as well.

(ii) *Laplace's method of generating function* F: Set $a_n := N(n\Delta t)$ and

$$F(z) := a_0 + a_1 z + a_2 z^2 + \dots$$

The original difference equation reads as $a_{n+2} = a_{n+1} + a_n$. Hence

$$\frac{F - a_0 - a_1 z}{z^2} = \frac{F - a_0}{z} + F.$$

Since $a_0 = 0$ and $a_1 = 1$,

$$F = \frac{z}{1 - z - z^2}.$$

It remains to compute the coefficients a_n of F. This can be done easily by using the method of partial fractions. To begin with, note that the equation $1 - z - z^2 = 0$ has the zeros $z = \frac{1}{\lambda_\pm}$, since the substitution $z = \frac{1}{\lambda}$ yields the characteristic equation $\lambda^2 - \lambda - 1 = 0$. Explicitly, we get the following decomposition

$$F = \frac{1}{\lambda_+ - \lambda_-} \left(\frac{1}{1 - \lambda_+ z} - \frac{1}{1 - \lambda_- z} \right).$$

By applying the geometric series,

$$\frac{1}{1 - \lambda_\pm z} = 1 + \lambda_\pm + \lambda_\pm^2 z^2 + \dots$$

This way we obtain the desired solution $a_n = \frac{1}{\sqrt{5}}(\lambda_+^n - \lambda_-^n)$ from (6.12).

Note that the proof (ii) has a formal character, since we did not check the convergence of the power series expansion. It turns out that it is not necessary to check the convergence; in Volume II we will establish the Mikusiński calculus for hyperfunctions which leads to expressions being always convergent, in a generalized sense.

The golden ratio and the chaotic motion of asteroids. The number r is one of the most famous irrational numbers. It is called the golden ratio for the following reason. Subdivide the unit interval by the point r into two subintervals such that

$$\frac{r}{1 - r} = \frac{1}{r}.$$

Then $r^2 - r - 1 = 0$ (Fig. 6.1) In ancient times, the golden ratio played an important role in architecture for aesthetical reasons.[9] The iterative method

[9] Applications of the golden ratio in art and sciences including quasi-crystals and biological objects can be found in Dunlap (1997) and Koshy (2001).

$$r_{n+1} = \frac{1}{1 + r_n}, \qquad n = 0, 1, \ldots, r_0 = 1 \tag{6.14}$$

converges to the golden ratio r as $n \to \infty$. This is the most regular continued fraction,

$$r = \frac{1}{1 + \frac{1}{1+\ldots}}$$

In this sense, the golden ratio is the "most irrational" real number. If one wants to simulate chaotic motion of asteroids on computers in the framework of the Kolmogorov–Arnold–Moser theory (KAM theory), one uses the iterative method (6.14) (see Scheck (2000), Vol. 1). Surprisingly enough, this method works perfectly despite the rational character of the approximations.

Applications of discrete dynamical systems. In mathematics, the equations of motion of a discrete dynamical system are called difference equations. Nowadays such equations are used

- for solving partial differential equations on computers (see Knabner and Angermann (2003)),
- for modelling deterministic chaos in physics (see Schuster (1994)),
- for studying mathematical models on both the origin of life, taking metabolism into account (see Dyson (1999a)), and
- virus dynamics in immunology (see Noack and May (2000)). This concerns, for example, the spread of aids.

We also refer to W. de Melo and S. van Strien, One-Dimensional Dynamics, Springer, Berlin, 1993 and to J. Jost, Dynamical Systems: Examples of Complex Behavior, Springer, Berlin, 2005.

6.4 Integral Transformations

Integral transformations are extremely useful in mathematics and physics.
Folklore

The discrete two-sided Laplace transform. Replace the discrete Laplace transform (6.7) by the more general formula

$$\boxed{f(z) = \sum_{n=-\infty}^{\infty} a_n z^n} \tag{6.15}$$

along with the inverse formula

$$\boxed{a_n = \frac{1}{2\pi \mathrm{i}} \int_{|z|=r} \frac{f(z)}{z^{n+1}}\, dz, \qquad n = 0, 1, 2, \ldots} \tag{6.16}$$

for fixed radius $r > 0$. The transformation

$$f = \mathcal{M}(\ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots)$$

sends each sequence $(\ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots)$ of complex numbers to a formal Laurent series expansion, f. Let $r_0 \in]0, 1[$ be a fixed number. Suppose that the function f is holomorphic in the open annulus $\{z \in \mathbb{C} : r_0 < |z| < 1\}$ where

$r_0 > 0$. Then, we have the inverse transformation formula (6.16) for each real number $r \in]r_0, 1[$.

The two-sided Laplace transform. The two-sided Laplace transform

$$\boxed{A(s) := \int_{-\infty}^{\infty} a(t)\mathrm{e}^{-st}dt} \tag{6.17}$$

represents the continuous analogue of the two-sided discrete Laplace transform by setting $z = \mathrm{e}^{-s}$.

The Mellin transform. Substituting $\tau := \mathrm{e}^{-t}$ and $b(\tau) := a(t)$, the two-sided Laplace transform (6.17) passes over to the integral transformation

$$\boxed{A(s) := \int_0^{\infty} b(\tau)\tau^{s-1}\, d\tau} \tag{6.18}$$

which is called the Mellin transform. The Mellin transform (6.18) of the function $\mathrm{e}^{-\tau}$ is precisely Euler's gamma function,

$$\boxed{\Gamma(s) = \int_0^{\infty} \mathrm{e}^{-\tau}\tau^{s-1}\, d\tau \qquad \text{for all} \quad s \in \mathbb{C},\ \Re(s) > 0.} \tag{6.19}$$

The gamma function can be analytically continued to a meromorphic function in the complex plane which has poles only at the points $s = 0, -1, -2, \ldots$, and these poles are simple. The Laurent series at the pole $s = -n$ with $n = 0, 1, \ldots$ reads as

$$\Gamma(s) = \frac{(-1)^n}{n!(s+n)} + a_0 + a_1(s+n) + \ldots$$

The gamma function has the following characteristic properties:

(i) $\Gamma(1) = 1$.
(ii) $\Gamma(n) = (n-1)!$ for $n = 2, 3, \ldots$
(iii) $\Gamma(z+1) = z\Gamma(z)$ for all $z \in \mathbb{C}$ with $z \neq 0, -1, -2, \ldots$

The integral transformation

$$\boxed{F(s) = \frac{1}{\Gamma(s)} \int_0^{\infty} f(\tau)\tau^{s-1}\, d\tau} \tag{6.20}$$

is called the normalized Mellin transform f to F. This transformation is named after Robert Mellin (1854–1933) who wrote a fundamental paper about this transformation in 1895. In particular, the Riemann zeta function can be represented by the following normalized Mellin transform

$$\zeta(s) = \frac{1}{\Gamma(s)} \int_0^{\infty} \frac{\tau^{s-1}}{\mathrm{e}^{\tau} - 1}\, d\tau \qquad \text{for all} \quad s \in \mathbb{C},\ \Re(s) > 1.$$

We will show on page 665 that the Mellin transformation is closely related to the Haar measure on the multiplicative group of positive real numbers. This will be discussed in Sect. 11.9.4 on page 665.

The Euler beta function and string theory. For all complex numbers p and q with $\Re(p) > 0$ and $\Re(q) > 0$, the integral

$$B(p, q) := \int_0^1 \tau^{p-1}(1-\tau)^{q-1}\, d\tau$$

exists. This function is called Euler's beta function. The relation between the beta function and the gamma function reads as

$$B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

The beta function played an important role in the history of string theory. In the 1970s, the physicist Veneziano looked for a function that fulfills a number of natural postulates for the scattering amplitudes of elementary particles in strong interaction. He found out that Euler's beta function did the job. This observation initiated string theory as a theory for strong interaction. Later on, the paradigm changed completely. Nowadays string theory is a candidate for describing all of the fundamental forces in the universe.

6.5 The Riemann Zeta Function

> I feel that I am expressing my gratitude for the honor that the Berlin Academy has endowed upon me by appointing me as one of its correspondents, by making immediate use of my privileges thereof and presenting the results of an investigation on the frequency of prime numbers; a subject which, by the interest which Gauss (1777–1855) and Dirichlet (1805–1859) have given it over a long period of time, seems to be worthy of renewed mention.[10]
>
> Berhard Riemann, 1859
> *On the distribution of prime numbers*

By definition, the Riemann zeta function looks like

$$\boxed{\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}.}$$

This series converges in the classical sense if s is a complex number with real part larger than 1. This function plays a fundamental role in both number theory and modern physics. Let us first investigate applications to number theory. Applications to the famous Casimir effect in quantum field physics will be studied on page 299.

6.5.1 The Prime Number Theorem – a Pearl of Mathematics

A natural number p is said to be a prime iff $p \geq 2$ and the only divisors of p are 1 and p itself. The first prime numbers are $2, 3, 5, 7, 11, 13, 17, 19$. Euclid (300 B.C.) knew that

[10] This is the beginning of one of the most famous works in all of mathematics. In this short paper Riemann develops his new ideas and presents the famous "Riemann hypothesis". The German title of this paper reads as "Über die Anzahl von Primzahlen unter einer gegebenen Grösse", Berliner Monatsberichte 1859, pp. 671–680. The collected works of Riemann (1990) are all in all just one volume. However, every single one of these papers is a jewel of mathematics. Riemann (1826–1866) has profoundly influenced the mathematics and physics of the twentieth century with his treasure of ideas. We refer to K. Maurin, Riemann's Legacy, Kluwer, Dordrecht, 1997.

- there exists an infinite number of primes, and
- each natural number $n \geq 2$ allows the unique factorization

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_m^{\alpha_m}$$

where $p_1, p_2, \ldots$ is an increasing sequence of primes, and the exponents $\alpha_1, \alpha_2, \ldots$ are positive integers.

In 1737 Euler proved that for each real number $s > 1$,

$$\boxed{\prod_p \left(1 - \frac{1}{p^s}\right)^{-1} = \sum_{n=1}^{\infty} \frac{1}{n^s}.} \tag{6.21}$$

Here, the product refers to all primes p. Euler's formula combines number theory with analysis. In fact, from unique factorization we have

$$\zeta(s) = \sum_{\mu_2, \mu_3, \ldots \geq 0} (2^{\mu_2} 3^{\mu_3} \cdots)^{-s} = \prod_p \left(\sum_{\mu \geq 0} p^{-\mu s} \right) = \prod_p \frac{1}{1 - p^{-s}}$$

by using the geometric series.

Legendre (1752–1832) introduced the prime number distribution function $\pi(x)$. For an arbitrary real number $x \geq 2$,

$$\pi(x) := \{\text{number of primes} \leq x\}.$$

Counting the primes below 40 000, Legendre conjectured in 1785 that[11]

$$\pi(x) \simeq \frac{x}{\ln x - 1.083\,66}, \qquad x \to +\infty.$$

The young Gauss did not know Legendre's work. In 1792, the fifteen-years old Gauss conjectured by counting primes that

$$\pi(x) \simeq \mathrm{Li}(x), \qquad x \to +\infty$$

where

$$\mathrm{Li}(x) := PV \int_2^x \frac{dt}{\ln t}, \qquad x \geq 2.$$

The principal value stands for the limit $\lim_{\varepsilon \to +0} \int_2^{1-\varepsilon} + \int_{1+\varepsilon}^x$. The prime number theorem reads as follows.

Theorem 6.1 *The asymptotic distribution of primes is given by*

$$\boxed{\pi(x) \simeq \frac{x}{\ln x} \qquad \text{as } x \to +\infty.}$$

[11] The symbol $f(x) \simeq g(x), x \to +\infty$ means that the quotient $\frac{f(x)}{g(x)}$ goes to 1 as $x \to +\infty$.

For more than hundred years, the proof of the prime number conjecture was a famous open problem in mathematics. In 1896, Jacques Hadamard (1865–1963) and Charles de la Valeé-Poussin (1866–1962) proved independently the prime number theorem. Their proofs were long and intricate. A simple analytic proof was given by Newman (1980). Our proof follows Zagier (1996).

Proof. For complex numbers s with $\Re(s) > 1$ and real numbers x, let us introduce the two auxiliary functions

$$\Phi(s) := \sum_p \frac{\ln p}{p^s}, \qquad \vartheta(x) := \sum_{p \le x} \ln p.$$

Here, we sum over all the primes p. The following proof uses analytic continuation of the Riemann ζ function along with Newman's adiabatic theorem based on Cauchy's integral formula. We will proceed in the following steps.

(i) For the Riemann zeta function, the series

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

converges absolutely for all complex numbers s with $\Re(s) > 1$.

(ii) Extension of Euler's formula: For all complex numbers s with $\Re(s) > 1$,

$$\zeta(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}.$$

(iii) $\zeta(s) - \frac{1}{s-1}$ extends holomorphically to the set $\{s \in \mathbb{C} : \Re(s) > 0\}$.

(iv) $\vartheta(x) = O(x)$ as $x \to +\infty$.

(v) $\zeta(s) \neq 0$ for all $s \in \mathbb{C}$ with $\Re(s) \ge 1$, and $\Phi(s) - \frac{1}{s-1}$ extends holomorphically to the set $\{s \in \mathbb{C} : \Re(s) \ge 1\}$.

(vi) Newman's adiabatic theorem: Let $a : [0, \infty[\to \mathbb{R}$ be a bounded and locally integrable function and suppose that the function

$$A(z) := \int_0^{\infty} a(t) \mathrm{e}^{-zt}\, dt, \qquad \Re(z) > 0$$

extends holomorphically to the set $\{z \in \mathbb{C} : \Re(z) \ge 0\}$. Then, the integral $\int_0^{\infty} a(t)dt$ exists and equals $A(0)$.

(vii) The integral $\int_1^{\infty} \frac{\vartheta(x)-x}{x^2}\, dx$ is convergent.

(viii) $\vartheta(x) \simeq x$ as $x \to \infty$.

Let us first show that claim (viii) implies easily the prime number theorem. In fact, for any $\varepsilon > 0$, we get the following two inequalities

$$\vartheta(x) = \sum_{p \le x} \ln p \le \sum_{p \le x} \ln x = \pi(x) \ln x$$

and

$$\begin{aligned} \vartheta(x) &\ge \sum_{x^{1-\varepsilon} \le p \le x} \ln p \ge \sum_{x^{1-\varepsilon} \le p \le x} (1-\varepsilon) \ln x \\ &\ge (1-\varepsilon) \ln x \cdot \{\pi(x) - O(x^{1-\varepsilon})\} \qquad x \to +\infty. \end{aligned}$$

To simplify notation, set $f(x) := \frac{\pi(x)\ \ln x}{x}$. Then

$$(1-\varepsilon)f(x) - O\left(\frac{1}{x^\varepsilon}\right)\ \ln x \le \frac{\vartheta(x)}{x} \le f(x), \qquad x \to +\infty.$$

Letting $x \to +\infty$, it follows from (viii) above that

$$(1-\varepsilon)\limsup f(x) \le 1 \le \liminf f(x)$$

for all $\varepsilon > 0$. Hence

$$\limsup f(x) \le 1 \le \liminf f(x) \le \limsup f(x), \qquad x \to +\infty.$$

Therefore, $\liminf f(x) = \limsup f(x) = 1$ as $x \to +\infty$. This implies

$$\lim_{x\to+\infty} f(x) = 1$$

which is the claim of the prime number theorem.

It remains to prove (i) through (viii). The symbol p denotes primes in what follows.

Ad (i). Note that $|n^s| = n^{\Re(s)}$ for $n = 1, 2, \ldots$

Ad (ii). Use Euler's argument from (6.21) by taking the absolute convergence from (i) into account.

Ad (iii). For $\Re(s) > 1$ we have

$$\zeta(s) - \frac{1}{s-1} = \sum_{n=1}^{\infty}\frac{1}{n^s} - \int_1^\infty \frac{dx}{x^s} = \sum_{n=1}^{\infty}\int_n^{n+1}\left(\frac{1}{n^s} - \frac{1}{x^s}\right)dx.$$

The series on the right converges absolutely for $\Re(s) > 0$ because

$$\begin{aligned}\left|\int_n^{n+1}\left(\frac{1}{n^s} - \frac{1}{x^s}\right)dx\right| &= \left|s\int_n^{n+1}dx\int_n^x \frac{du}{u^{s+1}}\right| \\ &\le \max_{n\le u\le n+1}\left|\frac{s}{u^{s+1}}\right| = \frac{|s|}{n^{\Re(s)+1}}\end{aligned}$$

by the mean theorem for integrals.

Ad (iv). For $n = 1, 2, \ldots$ we have $\binom{2n}{n} = (2n-1)\binom{2n-2}{n-1}$. By induction,

$$\binom{2n}{n} \ge \prod_{n<p\le 2n} p.$$

This implies

$$2^{2n} = (1+1)^{2n} = \binom{2n}{0} + \ldots + \binom{2n}{2n} \ge \binom{2n}{n} \ge \prod_{n<p\le 2n} p = \mathrm{e}^{2\vartheta(2n)-\vartheta(n)}.$$

Hence, since $\vartheta(x)$ changes by $O(\ln x)$ if x changes by $O(1)$ for large x,

$$\vartheta(x) - \vartheta\left(\frac{x}{2}\right) \le Cx$$

for any $C > \ln 2$ and all $x \ge x_0$ where x_0 depends on C. Summing this over

$$x, \frac{x}{2}, \ldots, \frac{x}{2^r},$$

where $x/2^r \geq x_0 > x/2^{r+1}$, we obtain $\vartheta(x) \leq 2Cx + O(1)$ as $x \to +\infty$.

Ad (v). For $\Re(s) > 1$, the convergent product in (ii) implies that $\zeta(s) \neq 0$ and that

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_p \frac{\ln p}{p^s - 1} = \Phi(s) + \sum_p \frac{\ln p}{p^s(p^s - 1)}.$$

The final sum converges for $\Re(s) > \frac{1}{2}$, so this and (iii) imply that $\Phi(s)$ extends meromorphically to $\Re(s) > \frac{1}{2}$, with poles only at $s = 1$ and at the zeros of $\zeta(s)$, and that, if $\zeta(s)$ has a zero of order μ at $s = 1 + \mathrm{i}\alpha$ ($\alpha \in \mathbb{R},\ \alpha \neq 0$) and a zero of order ν at $1 + 2\mathrm{i}\alpha$ (so $\mu, \nu \geq 0$ by (iii)), then

- $\lim_{\varepsilon \to +0}\ \varepsilon\Phi(1 + \varepsilon) = 1$,
- $\lim_{\varepsilon \to +0}\ \varepsilon\Phi(1 + \varepsilon \pm \mathrm{i}\alpha) = -\mu$,
- $\lim_{\varepsilon \to +0}\ \varepsilon\Phi(1 + \varepsilon \pm 2\mathrm{i}\alpha) = -\nu$.

The inequality

$$\sum_{r=-2}^{2} \binom{4}{2+r} \Phi(1 + \varepsilon + \mathrm{i}r\alpha) = \sum_p \frac{\ln p}{p^{1+\varepsilon}} \, (p^{\mathrm{i}\alpha/2} + p^{-\mathrm{i}\alpha/2})^4 \geq 0$$

then implies that $6 - 8\mu - 2\nu \geq 0$, so $\mu = 0$.

Ad (vi). The elementary proof will be given on page 686 in the context of adiabatic limits in physics.

Ad (vii). The function ϑ is piecewise constant and it jumps at the prime numbers. This implies that for $\Re(s) > 1$, we have

$$\Phi(s) = \sum_p \frac{\ln p}{p^s} = s \int_1^\infty \frac{\vartheta(x)}{x^{s+1}} \, dx = s \int_0^\infty \mathrm{e}^{-st} \vartheta(\mathrm{e}^t) dt.$$

Therefore, the claim (vii) is obtained by applying Newman's adiabtatic theorem (vi) to the functions $a(t) := \vartheta(\mathrm{e}^t)\mathrm{e}^{-t} - 1$ and

$$A(z) = \frac{\Phi(z+1)}{z+1} - \frac{1}{z},$$

which satisfy its hypotheses by (iv) and (v).

Ad (viii). Assume that for some $\lambda > 0$ there are arbitrarily large x which satisfy the inequality $\vartheta(x) \geq \lambda x$. Since ϑ is non-decreasing, we have

$$\int_x^{\lambda x} \frac{\vartheta(t) - t}{t^2} \, dt \geq \int_x^{\lambda x} \frac{\lambda x - t}{t^2} \, dt = \int_1^\lambda \frac{\lambda - t}{t^2} \, dt > 0$$

for such x, contradicting (vi). Similarly, the inequality $\vartheta(x) \leq \lambda x$ with $\lambda < 1$ would imply

$$\int_{\lambda x}^x \frac{\vartheta(t) - t}{t^2} \, dt \leq \int_{\lambda x}^x \frac{\lambda x - t}{t^2} \, dt = \int_\lambda^1 \frac{\lambda - t}{t^2} \, dt < 0,$$

again a contradiction for λ fixed and x big enough. □

6.5.2 The Riemann Hypothesis

> Billions of zeros cannot be wrong. Recent work by van de Lune has shown that the first 10 billion zeros of the zeta function are on the critical line... The Riemann hypothesis tells us that the primes are distributed in as nice a way as possible. If the Riemann hypothesis would be false, there would be some strange irregularities in the distribution of primes. The first zero off the line would be a very important mathematical constant. It seems unlikely that nature is that perverse!
>
> Brian Conrey, 2003
> *The Riemann hypothesis*[12]

In 1859 Riemann wrote a fundamental paper about the zeta function. For the first time, he studied this function for complex arguments. He proved that the zeta function is well-defined for all complex numbers $s \neq 1$, in the sense of an analytic continuation. To this end, Riemann used the following functional equation for the zeta function,

$$\frac{2^{s-1}\pi^s}{\Gamma(s)}\,\zeta(1-s) = \zeta(s)\,\cos\frac{\pi s}{2} \qquad \text{for all} \quad s \in \mathbb{C}.$$

The zeta function has the trivial zeros $s = -2, -4, -6, \ldots$ The Riemann hypothesis claims that all of the other zeros s satisfy the condition

$$\Re(s) = \frac{1}{2}.$$

This is the most famous open problem in mathematics. If the Riemann hypothesis is right, then one has the following sharp inequality

$$\boxed{|\pi(x) - \mathrm{Li}(x)| \leq \mathrm{const} \cdot \sqrt{x}\,\ln x \qquad \text{for all} \quad x \geq 2} \tag{6.22}$$

where $\pi(.)$ denotes the prime number distribution function.

6.5.3 Dirichlet's *L*-Function

> Over the years striking analogies have been observed between the Riemann ζ-function and other L-functions. While these functions are seemingly independent of each other, there is growing evidence that they are all somehow connected in a way that we do not fully understand... There is a growing body of evidence that there is a conspiracy between L-functions – a conspiracy which is preventing us from solving the Riemann hypothesis.
>
> Brian Conrey, 2003
> *The Riemann hypothesis*

In his famous 1801 treatise on number theory *Disquisitiones arithmeticae*, the young Gauss studied the theory of quadratic number fields. Such fields are generated by adding the solution of a quadratic algebraic equation to the field of rational numbers $\mathbb{Q}$. For example, adding the solution of the quadratic equation

$$x^2 - 5 = 0$$

[12] Notices of the American Mathematical Society **50** (3) (2003), pp. 341–353 (reprinted with permission).

to the field $\mathbb{Q}$ we get the quadratic number field $\mathbb{Q}(\sqrt{5})$ consisting of all numbers $a+b\sqrt{5}$ where a and b are rational numbers. In 1837 Dirichlet (1805–1859) founded analytic number theory. He used L-functions in order to get his famous explicit formulas for the class numbers of quadratic number fields. These formulas describe deep structural properties of quadratic number fields. In 1855 Dirichlet became the follower of Gauss in Göttingen. Dirichlet also used L-series in order to prove the following crucial theorem:

If the integers n and d are positive, then the sequence $n, n+d, n+2d, \ldots$ contains infinitely many primes.

For the proofs, we refer to Zagier (1981), and Stein, Shakarchi (2003), Vol. 1. In order to explain the basic idea of Dirichlet's L-function, let us consider a simple example.

The ring $\mathbb{Z}/\mathrm{mod}\ 4$. Let $\mathbb{Z}$ denote the ring of integers. For $n, m \in \mathbb{Z}$, we write

$$n \equiv m \mod 4$$

iff the difference $n-m$ is divisible by 4. The set

$$[n] := \{m \in \mathbb{Z} : m \equiv n \mod 4\}$$

is called the rest class modulo 4, and the elements of $[n]$ are called the representatives of the rest class. For example,

$$[3] := \{3, 3 \pm 4, 3 \pm 8, 3 \pm 12, \ldots\}.$$

Addition and multiplication of rest classes are defined by the corresponding operations for the representatives. Explicitly,

$$[n] + [m] := [n+m], \qquad [n][m] := [nm].$$

These definitions do not depend on the choice of representatives. The rest classes $[0], [1], [2], [3]$ form the ring $\mathbb{Z}/\mathrm{mod}\ 4$ with the zero element $[0]$ and the unit element $[1]$. Since

$$[2][2] = [4] = [0], \qquad [3][3] = [9] = [1],$$

precisely the elements $[1], [3]$ are invertible.

Characters of the ring $\mathbb{Z}/\mathrm{mod}\ 4$. By a character of the ring $\mathbb{Z}/\mathrm{mod}\ 4$, we understand a multiplicative map

$$\chi : \mathbb{Z}/\mathrm{mod}\ 4 \to \mathbb{C}.$$

More precisely, for all elements a and b of $\mathbb{Z}/\mathrm{mod}\ 4$, we postulate that

$$\chi(ab) = \chi(a)\chi(b).$$

Furthermore, $|\chi(a)| = 1$ for all invertible elements a, and $\chi(b) = 0$ for all non-invertible elements b. Explicitly, such a character is given by

$$\chi([0]) := \chi([2]) = 0, \qquad \chi([1]) := 1, \qquad \chi([3]) := -1.$$

The L-function of the ring $\mathbb{Z}/\mathrm{mod}\ 4$. Set $\chi(n) := \chi([n])$ for all integers n, and define

$$L_\chi(s) := \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}$$

for all complex numbers s with $\Re(s) > 1$. Then

$$\boxed{L_\chi(s) = \prod_p \left(1 - \frac{\chi(p)}{p^s}\right)^{-1}.}$$

Here, the product runs over all prime numbers. This generalizes Euler's product formula.

Suggested reading. For number theory including applications to physics, we refer to the following collection of beautiful survey articles

- M. Waldschmidt, P. Moussa, J. Luck, and C. Itzykson (Eds.), From Number Theory to Physics, Springer, New York, 1995

and to the following books and papers:

- T. Apostol, Introduction to Analytic Number Theory, Springer, New York, 1986.
- P. Bateman and H. Diamond, A hundred years of prime numbers, American Math. Monthly **103** (1996), 729–741.
- A. Bytsenko, G. Cognola, E. Elizalde, V. Moretti, and S. Zerbini, Analytic Aspects of Quantum Fields, World Scientific, Singapore, 2003.
- P. Cartier, An introduction to zeta functions. In: M. Waldschmidt et al. (1995) (see above), pp. 1–63.
- J. Conrey, The Riemann hypothesis, Notices Amer. Math. Soc. **50**(3) (2003), 341–353.
- A. Connes, Noncommutative geometry and the Riemann zeta function, Frontieres and perspectives, International Mathematical Union, 2000.
- J. Havil, Gamma: Exploring Euler's Constant, Princeton University Press, Princeton, 2003.
- H. Edwards, Riemann's Zeta Function, Academic Press, New York, 1974.
- E. Elizalde, Ten Physical Applications of Spectral Zeta Functions, Springer, Berlin, 1995.
- P. Gilkey, Invariance Theory, the Heat Equation, and the Atiyah–Singer Index Theorem, CRC Press, Boca Raton, Florida, 1995.
- P. Gilkey, Asymptotic Formulae in Spectral Geometry, Chapman, CRC Press, Boca Raton, Florida, 2003.
- K. Jorgenson and S. Lang, The ubiquitous heat kernel. In: B. Enquist and W. Schmid (Eds.), Mathematics Unlimited – 2001 and Beyond, pp. 655–684. Springer, New York, 2001.
- K. Kirsten, Spectral Functions in Mathematics and Physics, Chapman, Boca Raton, Florida, 2002.
- M. Kontsevich and D. Zagier, Periods. In: B. Enquist and W. Schmid (Eds.), pp. 771–808 (see above).
- S. Lang, Introduction to Modular Forms, Springer, Berlin, 1976.
- D. Newman, Simple analytic proof of the prime number theorem, American Math. Monthly, **87** (1980), 693–696.
- S. Patterson, An Introduction to the Theory of the Riemann Zeta Function, Cambridge University Press, 1995.
- P. Ribenboim, The New Book of Prime Number Records, Springer, New York, 1996.
- B. Riemann, Collected Mathematical Works with Commentaries. Edited by R. Narasimhan, Springer, New York, and Teubner, Leipzig, 1990.
- E. Stein and R. Shakarchi, Princeton Lectures in Analysis, I: Fourier Analysis, II: Complex Analysis, III: Measure Theory, IV: Selected Topics, Princeton University Press, 2003.
- E. Titchmarsh, The Theory of the Riemann Zeta Function. Edited by D. Heath-Brown, Cambridge University Press, 1986.

- D. Vassilievich, Heat Kernel Expansion: User's Manual, Physics Reports **388** (2003), 279–360.
- D. Zagier, Zeta Functions and Quadratic Number Fields: An Introduction to Advanced Number Theory, Springer, Berlin, 1981 (in German).
- D. Zagier, Introduction to Modular Forms, pp. 238–291. In: M. Waldschmidt et al. (Eds.) (1995) (see above).
- D. Zagier, Newman's short proof of the prime number theorem, Amer. Math. Monthly **104**, 705–708.

6.6 The Casimir Effect in Quantum Field Theory and the Epstein Zeta Function

> How can it be that mathematics, being after all a product of human thought independent of experience, is so admirably adapted to the objects of reality.
>
> Albert Einstein (1879–1955)

Let us study some physical effect which allows us to prove experimentally the existence of the ground state energy of a quantum field. This is the so-called Casimir effect. Consider two parallel, uncharged, perfectly conducting metallic plates (see Fig. 6.2). From the classical point of view, there should not be any electromagnetic force between the two plates. However, Casimir predicted in 1948 that the fluctuations of the ground state energy of the electromagnetic quantum field cause an attractive electromagnetic force between the two plates given by

$$\boxed{\mathbf{F}(d) = -\frac{\pi^2 \hbar c L^2}{240 d^4}\, \mathbf{e}_x}$$

where d is the distance between the two plates, and L^2 is the surface area of each of the two plates. Here, the unit vector $\mathbf{e}_x$ points in direction of the x-axis. We write $\mathbf{F}(d) = F(d)\mathbf{e}_x$. Explicitly, in the SI system,

$$F(d) = -1.30 \cdot 10^{-27}\, \frac{[L]^2}{[d]^4} \cdot \mathrm{N}$$

where $[L]$ and $[d]$ are the numerical values of L and d, respectively, measured in m (meter). Note that 1 N (Newton) = 1 kg m/s^2. The tiny Casimir force $F(d)$ can be established experimentally. This effect was first studied by H. Casimir, On the

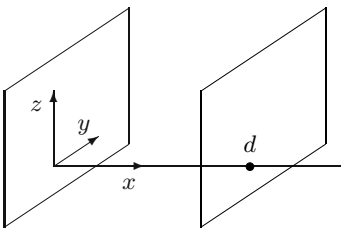


Fig. 6.2. The Casimir effect

attraction between two perfectly conducting plates, Kon. Ned. Akad. Wetensch Proc. **51** (1948), 793–795.

Let us show how this Casimir force can be obtained by using both physical and mathematical arguments. From the physical point of view, the two plates change the situation of the ground state of the electromagnetic quantum field. The ground state is confined by the two plates. We now postulate that the confinement of the ground state of a quantum fields causes a force. In the present case, we get

$$F(d) = -E'(d)$$

where $E(d)$ is the ground state energy of the quantum field between the two plates.[13]

The one-dimensional model. We first consider a one-dimensional model. Let $A : [0, d] \to \mathbb{R}$ be a classical physical quantity which describes a standing wave with $A(0) = A(d) = 0$. That is,

$$A(x) := A_0 \sin \frac{n\pi x}{d}, \qquad x \in [0, d]$$

where $n = 0, 1, 2, \ldots$. Here, we have the wave number

$$k(n) := \frac{n\pi}{d}.$$

The corresponding travelling wave reads as

$$A(x, t) = A_0 \sin(k(n)x - \omega(n)t), \qquad x, t \in \mathbb{R} \tag{6.23}$$

with the propagation speed $c = \frac{\omega(n)}{k(n)}$. In 1925 Heisenberg showed that the energy levels of a harmonic quantum oscillator with angular frequency ω are given by

$$E = \hbar\omega \left(N + \frac{1}{2} \right), \qquad N = 0, 1, 2, \ldots$$

The quantum number $N = 0$ describes the zero-point energy

$$E = \frac{\hbar\omega}{2}.$$

Furthermore, the quantum numbers $N = 1, 2, \ldots$ correspond to excited energy states.

Let us now consider a quantum field which is obtained by quantizing the classical field (6.23). Heuristically, the quantum field is given by an infinite number of harmonic quantum oscillators of frequencies $\omega(n) = ck(n)$ with $n = 0, 1, \ldots$ Naturally enough, we assume that the ground state energy $E(d)$ of the quantum field is given by the sum of the zero-point energies of the single harmonic oscillators. Hence

$$E(d) = \sum_{n=0}^{\infty} \frac{\hbar\omega(n)}{2} = \sum_{n=0}^{\infty} \frac{\hbar ck(n)}{2}.$$

This implies

$$E(d) = \frac{\pi c\hbar}{2d} \sum_{n=1}^{\infty} n.$$

[13] Note that the derivative $E'(d)$ of the energy $E(d)$ with respect to the length parameter d has the physical dimension of $energy/length = force$.

Obviously, this series is divergent. This difficulty is typical for quantum field theories. In order to get a finite value as observed in the physical experiment, we will argue as follows. To begin with, we choose a large cut-off constant $R > 0$, and we introduce the truncated ground state energy

$$E_R(d) := \frac{\pi c\hbar}{2d} \sum_{n=1}^{\infty} n\theta(R-n)$$

where θ denotes the Heaviside function. Explicitly, $\theta(t) := 1$ if $t \geq 0$ and $\theta(t) := 0$ if $t < 0$. Naturally enough, we postulate that

> *The true ground state energy does not depend on the arbitrarily chosen cut-off.*

In order to get such an energy expression, we introduce the truncated Riemann zeta function

$$\zeta_R(s) := \sum_{n=1}^{\infty} \frac{1}{n^s} \cdot \theta(R-n)$$

for all complex numbers s. Obviously,

$$E_R(d) = \frac{\pi c\hbar}{2d}\, \zeta_R(-1).$$

Let us now study the limit $R \to +\infty$.

- If $\Re(s) > 1$, then the classical limit

$$\lim_{R\to+\infty} \zeta_R(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \zeta(s)$$

  exists. The Riemann zeta function ζ is a meromorphic function on the complex plane $\mathbb{C}$ with one simple pole at the point $s = 1$ and no other poles. Thus, $\zeta(s)$ is well-defined for all $s \in \mathbb{C} \setminus \{1\}$. In particular, $\zeta(-1) = -\frac{1}{12}$.
- We now define the generalized limit

$$\lim_{R\to+\infty} \zeta_R(s) := \zeta(s) \qquad \text{for all} \quad s \in \mathbb{C} \setminus \{1\}$$

  which is finite. We briefly write this as

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \zeta(s) \qquad \text{for all} \quad s \in \mathbb{C} \setminus \{1\}. \tag{6.24}$$

Now we define the ground state energy $E(d)$ by setting

$$E(d) := \frac{\pi c\hbar}{2d}\zeta(-1) = -\frac{\pi c\hbar}{24d}.$$

This method is called the zeta function regularization.

In what follows, we will show that a similar argument can be used in order to get the energy expression (6.30) below for the Casimir effect. The idea is to replace the Riemann zeta function by the Epstein zeta function[14] and to study the

[14] P. Epstein, On the theory of general zeta functions I, II, Math. Ann. **56** (1903), 615–644; **63** (1907), 205–216 (in German).

asymptotics of the Epstein zeta function by using sophisticated arguments from analytic number theory.

The ground state energy of the electromagnetic quantum field. The truncated ground state energy $E_R(d)$ is equal to

$$\frac{1}{8}\sum_{l,m,n\in\mathbb{Z}} \pi\hbar c\left(\frac{l^2}{L^2}+\frac{m^2}{L^2}+\frac{n^2}{d^2}\right)^{1/2}\cdot\theta(R^2-l^2-m^2-n^2). \tag{6.25}$$

Here, we fix the large cut-off parameter $R>0$. We sum over all integers l,m,n. Note that $\sum_{l\in\mathbb{Z}} l^2\theta(R^2-l^2)=2\sum_{l\in\mathbb{N}} l^2\theta(R^2-l^2)$. Similarly, we obtain that $E_R(d)$ is equal to

$$\sum_{l,m,n\in\mathbb{N}} w_{lmn}\pi\hbar c\left(\frac{l^2}{L^2}+\frac{m^2}{L^2}+\frac{n^2}{d^2}\right)^{1/2}\cdot\theta(R^2-l^2-m^2-n^2). \tag{6.26}$$

Here, we only sum over lattice points (l,m,n) whose components are natural numbers, and we use the following weights:

- $w_{lmn}=1$ if the tupel (l,m,n) is non-degenerate, that is, none of the indices l,m,n is equal to zero;
- $w_{lmn}:=\frac{1}{2}$ if precisely one of the indices l,m,n is equal to zero;
- $w_{lmn}:=\frac{1}{4}$ if precisely two of the indices l,m,n are equal to zero.

The expression (6.25) for the ground state energy of the electromagnetic quantum field will be thoroughly motivated in Volume II on quantum electrodynamics by using the free photon field. At this point let us only discuss the basic ideas. Suppose that the functions $\mathbf{A}, U$ satisfy the wave equations

$$\Box\mathbf{A}=0, \qquad \Box U=0$$

and the Lorentz gauge condition $\operatorname{div}\mathbf{A}+\frac{\dot{U}}{c^2}=0$. Then the electromagnetic field

$$\mathbf{E}=-\dot{\mathbf{A}}-\mathbf{grad}\,U, \qquad \mathbf{B}=\mathbf{curl}\,\mathbf{A}$$

satisfies the Maxwell equations without electric charges and electric currents. According to the experimental situation pictured in Fig. 6.2 on page 299, we want to construct standing electromagnetic waves on the parallelepiped

$$\mathcal{P}:=\left\{(x,y,z)\in\mathbb{R}^3:\ \ 0\le x\le d,\ \ -\frac{L}{2}\le y,z\le\frac{L}{2}\right\}.$$

To this end, we set $U:=0$ and

$$\mathbf{A}(\mathbf{x},t):=\mathbf{A}_0\sin(\mathbf{k}_{lmn}\mathbf{x}-\omega_{lmn}t) \tag{6.27}$$

with $\mathbf{x}=x\mathbf{e}_x+y\mathbf{e}_y+z\mathbf{e}_z$ and $\mathbf{k}_{lmn}=k_x\mathbf{e}_x+k_y\mathbf{e}_y+k_z\mathbf{e}_z$. The unit vectors $\mathbf{e}_x,\mathbf{e}_y,\mathbf{e}_z$ point into the direction of the x-axis, y-axis, z-axis, respectively. Here, we choose the wave numbers

$$k_x:=\frac{n\pi}{d}, \quad k_y:=\frac{l\pi}{L}, \quad k_z:=\frac{m\pi}{L}, \qquad l,m,n=0,1,2,\ldots$$

and the angular frequencies $\omega_{lmn}:=c||\mathbf{k}_{lmn}||$ along with

$$\mathbf{A_0}\mathbf{k}_{lmn}=0. \tag{6.28}$$

This transversality condition guarantees that the gauge condition $\operatorname{div} \mathbf{A} = 0$ is satisfied. If $\mathbf{k}_{lmn} \neq 0$, then the equation (6.28) has two linearly independent vector solutions $\mathbf{A}_0$ which correspond to two possible polarization directions of electromagnetic waves.

We now consider the corresponding quantum field. Let us motivate that the truncated ground state energy of the quantum field is given by

$$E_R(d) = \sum_{l,m,n \in \mathbb{N}} w_{lmn} \cdot \hbar\omega_{lmn} \cdot \theta(R^2 - l^2 - m^2 - n^2).$$

Case 1: Non-degenerate oscillations. If the wave vector $\mathbf{k}_{lmn}$ is non-degenerate, that is, all of the natural numbers l, m, n are different from zero, then the oscillations with angular frequency ω_{lmn} contribute the amount of zero-point energy

$$2 \cdot \frac{\hbar\omega_{lmn}}{2} \tag{6.29}$$

to the ground state energy of the quantum field. The factor 2 corresponds to the two polarization degrees of freedom of electromagnetic waves. The energy contribution (6.29) is equal to

$$\hbar\omega_{lmn} = w_{lmn} \cdot \hbar c ||k_{lmn}|| = w_{lmn} \cdot \pi\hbar c \left(\frac{l^2}{L^2} + \frac{m^2}{L^2} + \frac{n^2}{d^2} \right)^{1/2}$$

with the weight $w_{lmn} := 1$.

Case 2: Degenerate oscillations. If the wave vector $\mathbf{k}_{lmn}$ is degenerate (i.e., at least one of the natural numbers l, m, n is equal to zero), then the oscillation (6.27) degenerates. This means that the oscillation is not 3-dimensional, but only r-dimensional with $r = 2, 1, 0$. In this case, we use the weights $w_{lmn} := \frac{1}{2}, \frac{1}{4}, \frac{1}{8}$, respectively.[15] This way we get the truncated ground state energy $E_R(d)$ from (6.25) above. Setting $\lambda := \frac{L}{d}$, we obtain

$$E_R(d) = \frac{\pi c\hbar}{8L} \sum_{l,m,n \in \mathbb{Z}} (l^2 + m^2 + \lambda^2 n^2)^{1/2} \cdot \theta(R^2 - l^2 - m^2 - n^2).$$

Now define the Epstein zeta function

$$Z(\lambda, s) := \sum\nolimits'_{l,m,n \in \mathbb{Z}} \frac{1}{(l^2 + m^2 + \lambda^2 n^2)^s}$$

for all $\lambda > 0$ and all complex numbers s with $\Re(s) > \frac{3}{2}$. The prime on the summation sign means that the term $(l, m, n) = (0, 0, 0)$ is to be omitted. This series converges. For understanding the Casimir effect, the following theorem is crucial.

Theorem 6.2 (i) *For each $\lambda > 0$, the Epstein zeta function $s \mapsto Z(\lambda, s)$ is a meromorphic function on the complex plane with a simple pole at the point $s = \frac{3}{2}$ and no other poles.*

(ii) *We have the following asymptotics*

$$Z\left(\lambda, -\frac{1}{2}\right) = -\frac{\pi}{90}\lambda^3 - \frac{1}{\pi^2}\zeta\left(\frac{3}{2}\right) L_4\left(\frac{3}{2}\right) + O(\sqrt{\lambda} e^{-\pi\lambda}), \qquad \lambda \to +\infty$$

[15] From the mathematical point of view, this choice of weights will allow us to apply the Epstein zeta function.

and

$$Z\left(\lambda, -\frac{1}{2}\right) = -\frac{L_4(2)}{3\pi\lambda} - \frac{\lambda}{6} + O(\lambda e^{-\pi/\lambda}), \qquad \lambda \to +0.$$

Here, $L_4(s) := 1 - \frac{1}{3^s} + \frac{1}{5^s} - \ldots$ *for all complex numbers* s *with* $\Re(s) > 1$.

The proof due to Don Zagier will be given in the next section. As for the one-dimensional model above, we define the generalized limit $E(d) = \lim_{R\to+\infty} E_R(d)$ by setting

$$E(d) := \frac{\pi c\hbar}{8L} \cdot Z\left(\frac{L}{d}, -\frac{1}{2}\right).$$

If the distance d between the two plates is small with respect to the length L of the plates, then the quotient $\lambda = \frac{L}{d}$ is large and we get the following asymptotics

$$\boxed{E(d) \simeq -\frac{\pi^2\hbar c}{720L}\left(\frac{L}{d}\right)^3, \qquad \frac{L}{d} \to +\infty.} \tag{6.30}$$

This yields the desired Casimir force

$$F(d) = -E'(d) = -\frac{\pi^2\hbar c L^2}{240 d^4}.$$

In 1826, the ingenious young mathematician Niels Hendrik Abel (1802–1829) wrote the following:

> The divergent series are the invention of the devil, and it is a shame to base on them any demonstration whatsoever. By using them, one may draw any conclusion he pleases and that is why these series have produced so many fallacies and so many paradoxes.

The experience of physicists shows that the definition of a convergent series is too restrictive; sometimes nature also sees the information which is encoded in a divergent series. In particular, one can say that

Nature sees analytic continuation.

Suggested reading. As an introduction to the Casimir effect, we recommend Gottfried and Tung-Mow Yan (2003), Sect. 10.3. A detailed study of the Casimir effect can be found in the following monographs:

K. Milton, The Casimir Effect: Physical Manifestations of Zero-Point Energy, World Scientific, Singapore, 2001.
P. Milonni, The Quantum Vacuum: An Introduction to Quantum Electrodynamics, Academic Press, Boston, 1994.
E. Elizalde, Ten Physical Applications of Spectral Zeta Functions, Springer, Berlin, 1995.

P. van Baal, A Course in Field Theory, University of Leiden, 1998.
Internet: `http://rulgm4.leidenuniv.nl/van-baal/FTcourse.html`
K. Kirsten, Spectral Functions in Mathematics and Physics, Chapman, Boca Raton, 2002.
A. Bytsenko, G. Cognola, E. Elizalde, V. Moretti, and S. Zerbini, Analytic Aspects of Quantum Fields, World Scientific, Singapore, 2003.

We also refer to

S. Blau, M. Visser, and A. Wipf, Zeta functions and the Casimir energy. Nucl. Phys. B **310** (1988), 163–180

and to the two survey articles on the Casimir effect written by B. Duplantier and R. Balian. These articles are contained in the following collection:

> B. Duplantier and V. Rivasseau (Eds.), Vacuum Energy – Renormalization. Poincaré Seminar 2002, Birkhäuser, Basel, 2003.

Experimental results about the Casimir effect can be found in

> T. Ederth, Template-stripped gold surfaces with 0.4-nm rms roughness suitable for force measurements: Applications to the Casimir force in the 20-100-nm range, Physical Reviews **A62** (6) (2000), 062104.

6.7 Appendix: The Mellin Transformation and Other Useful Analytic Techniques by Don Zagier

> The Mellin transformation is a magic wand.
> Folklore

The following material is not sufficiently well known to a broad audience. The reader should note that the tools to be described are extremely useful. These tools enlarge the arsenal of weapons used in mathematical physics. They allow interesting applications concerning the asymptotic behavior of functions occurring in mathematics and physics.

6.7.1 The Generalized Mellin Transformation

The Mellin transformation is a basic tool for analyzing the behavior of many important functions in mathematics and mathematical physics, such as the zeta functions occurring in number theory and in connection with various spectral problems including the Casimir effect. We describe it first in its simplest form and then explain how this basic definition can be extended to a much wider class of functions, important for many applications.

Let $\varphi :]0,\infty[\to \mathbb{C}$ be a function on the positive real axis which is reasonably smooth (actually, continuous or even piecewise continuous would be enough) and decays rapidly at both 0 and ∞, i.e., the function $t^A\varphi(t)$ is bounded on $]0,\infty[$ for any $A \in \mathbb{R}$. Then the integral

$$\tilde{\varphi}(s) = \int_0^\infty \varphi(t) t^{s-1} dt \tag{6.31}$$

converges for any complex value s and defines a holomorphic function of s called the *Mellin transform* of $\varphi(t)$. The following small table, in which α denotes a complex number and λ a positive real number shows how $\tilde{\varphi}(s)$ changes when $\varphi(t)$ is modified in various simple ways:

$\varphi(\lambda t)$	$t^\alpha \varphi(t)$	$\varphi(t^\lambda)$	$\varphi(t^{-1})$	$\varphi'(t)$
$\lambda^{-s}\tilde{\varphi}(s)$	$\tilde{\varphi}(s+\alpha)$	$\lambda^{-s}\tilde{\varphi}(\lambda^{-1}s)$	$\tilde{\varphi}(-s)$	$(1-s)\tilde{\varphi}(s-1)$

(6.32)

We also mention, although we will not use it in the sequel, that the function $\varphi(t)$ can be recovered from its Mellin transform by the *inverse Mellin transformation* formula

$$\varphi(t) = \frac{1}{2\pi \mathrm{i}} \int_{C-\mathrm{i}\infty}^{C+\mathrm{i}\infty} \tilde{\varphi}(s) t^{-s} ds \qquad (0 < t < \infty),$$

where C is any real number. (That this is independent of C follows from Cauchy's formula (4.2) on page 212.)

The need for generalizing the Mellin transformation. In contrast to the assumption formulated for the classical Mellin transformation above, most functions which we encounter in practise are not very small at both zero and infinity.

- If we assume that $\varphi(t)$ is of rapid decay at infinity but grows like t^{-A} for some real number A as $t \to +0$, then the integral (6.31) converges and defines a holomorphic function only in the right half-plane $\Re(s) > A$.
- Similarly, if $\varphi(t)$ is of rapid decay at zero but grows like t^{-B} at infinity for some real number B, then $\tilde{\varphi}(s)$ makes sense and is holomorphic only in the left half-plane $\Re(s) < B$,
- while if $\varphi(t)$ has polynomial growth at both ends, say like t^{-A} at 0 and like t^{-B} at ∞ with $A < B$, then $\tilde{\varphi}(s)$ is holomorphic only in the strip $A < \Re(s) < B$.

But it turns out that in many cases the function $\tilde{\varphi}(s)$ has a meromorphic extension to a larger half-plane or strip than the one in which the original integral (6.31) converges, or even to the whole complex plane. Moreover, this extended Mellin transform can sometimes be defined even in cases where $A > B$, in which case the integral (6.31) does not converge for any value of s at all.

Investigation of the generalized Mellin transform. We are going to consider three different cases.

Case 1: Let us start with the frequently occurring case where $\varphi(t)$ is of rapid decay at infinity and is smooth in a neighborhood of zero, i.e., it has an asymptotic expansion[16]

$$\varphi(t) \sim \sum_{n=0}^{\infty} a_n t^n \qquad (t \to +0).$$

Then for s with $\Re(s) > 0$ and any positive integer N, the integral (6.31) converges and can be decomposed as follows:

$$\begin{aligned} \tilde{\varphi}(s) &= \int_0^1 \varphi(t) t^{s-1} dt + \int_1^\infty \varphi(t) t^{s-1} dt \\ &= \int_0^1 \left(\varphi(t) - \sum_{n=0}^{N-1} a_n t^n \right) t^{s-1} dt + \sum_{n=0}^{N-1} \frac{a_n}{n+s} + \int_1^\infty \varphi(t) t^{s-1} dt. \end{aligned}$$

The first integral on the right converges in the larger half-plane $\Re(s) > -N$ and the second for all $s \in \mathbb{C}$, so we deduce that $\tilde{\varphi}(s)$ has a meromorphic continuation to $\Re(s) > -N$ with simple poles of residue a_n at $s = -n$ $(n = 0, \dots, N-1)$ and no other singularities. Since this holds for every n, it follows that the Mellin transform $\tilde{\varphi}(s)$ in fact has a meromorphic continuation to all of $\mathbb{C}$ with simple poles of residue a_n at $s = -n$ $(n = 0, 1, 2, \dots)$ and no other poles. The same argument shows that, more generally, if $\varphi(t)$ is of rapid decay at infinity and has an asymptotic expansion

[16] Recall that this means that the difference $\varphi(t) - \sum_{n=0}^{N} a_n t^n$ is $o(t^N)$ as $t \to +0$ for any integer $N \geq 1$; it is not required that the series $\sum_{n=0}^{\infty} a_n t^n$ be convergent for any positive t.

$$\varphi(t) \sim \sum_{j=1}^{\infty} a_j\, t^{\alpha_j} \qquad (t \to 0) \tag{6.33}$$

as t tends to zero, where the α_j are real numbers tending to $+\infty$ as $j \to \infty$ or complex numbers with real parts tending to infinity, then the function $\tilde\varphi(s)$ defined by the integral (6.31) for $\Re(s) > -\min_j \Re(\alpha_j)$ has a meromorphic extension to all of $\mathbb{C}$ with simple poles of residue a_j at $s = -\alpha_j$ ($j = 1, 2, \dots$) and no other poles. Yet more generally, we can allow terms of the form $t^\alpha (\ln t)^m$ with $\lambda \in \mathbb{C}$ and $m \in \mathbb{Z}_{\ge 0}$ in the asymptotic expansion of $\varphi(t)$ at $t = 0$ and each such term contributes a pole with principal part $(-1)^m m!/(s+\alpha)^{m+1}$ at $s = -\alpha$, because $\int_0^1 t^{\alpha+s-1}(\ln t)^m\, dt = \frac{\partial^m}{\partial \alpha^m} \int_0^1 t^{\alpha+s-1}\, dt = (-1)^m m!/(\alpha+s)^{m+1}$ for $\Re(s+\alpha) > 0$.

Case 2: By exactly the same considerations, or by replacing $\varphi(t)$ by $\varphi(t^{-1})$, we find that if $\varphi(t)$ is of rapid decay (faster than any power of t) as $t \to +0$ but has an asymptotic expansion of the form

$$\varphi(t) \sim \sum_{k=1}^{\infty} b_k\, t^{\beta_k} \qquad (t \to \infty) \tag{6.34}$$

at infinity, where now the exponents β_k are complex numbers whose real parts tend to $-\infty$, then the function $\tilde\varphi(s)$, originally defined by (6.31) in a left half-plane $\Re(s) < -\max_k \Re(\beta_k)$, extends meromorphically to the whole complex s-plane with simple poles of residue $-b_k$ at $s = -\beta_k$ and no other poles. (More generally, again as before, we can allow terms $b_k t^{\beta_k} (\ln t)^{n_k}$ in (6.33) which then produce poles with principal parts $(-1)^{n_k+1} n_k!\, b_k/(s+\beta_k)^{n_k+1}$ at $s = -\beta_k$.)

Case 3: Now we can use these ideas to define $\tilde\varphi(s)$ for functions which are not small either at 0 or at ∞, even when the integral (6.31) does not converge for any value of s. We simply assume that $\varphi(t)$ is a smooth (or continuous) function on $]0, \infty[$ which has asymptotic expansions of the forms (6.33) and (6.34) at zero and infinity, respectively. (Again, we could allow more general terms with powers of $\ln t$ in the expansions, as already explained, but the corresponding modifications are easy and for simplicity of expression we will assume expansions purely in powers of t.) For convenience we assume that the numbering is such that $\Re(\alpha_1) \le \Re(\alpha_2) \le \cdots$ and $\Re(\beta_1) \ge \Re(\beta_2) \ge \cdots$. Then, for any $T > 0$ (formerly we took $T = 1$, but the extra freedom of being able to choose any value of T will be very useful later) we define two "half-Mellin transforms" $\tilde\varphi_{\le T}(s)$ and $\tilde\varphi_{\ge T}(s)$ by

$$\tilde\varphi_{\le T}(s) = \int_0^T \varphi(t)\, t^{s-1}\, dt \qquad \big(\Re(s) > -\Re(\alpha_1)\big)\,,$$
$$\tilde\varphi_{\ge T}(s) = \int_T^\infty \varphi(t)\, t^{s-1}\, dt \qquad \big(\Re(s) < -\Re(\beta_1)\big)\,.$$

Just as before, we see that for each integer $J \ge 1$ the function $\tilde\varphi_{\le}(s)$ extends by the formula

$$\tilde\varphi_{\le T}(s) = \int_0^T \left(\varphi(t) - \sum_{j=1}^{J} a_j t^{\alpha_j}\right) t^{s-1}\, dt + \sum_{j=1}^{J} \frac{a_j}{s+\alpha_j}\, T^{s+\alpha_j}$$

to the half-plane $\Re(s) > -\Re(\alpha_{J+1})$ and hence, letting $J \to \infty$, that $\tilde\varphi_{\le}(s)$ is a meromorphic function of s with simple poles of residue a_j at $s = -\alpha_j$ ($j = 1, 2, \dots$) and no other poles. Similarly, $\tilde\varphi_{\ge t}(s)$ extends to a meromorphic function whose only poles are simple ones of residue $-b_k$ at $s = -\beta_k$. We now define

$$\boxed{\tilde{\varphi}(s) := \tilde{\varphi}_{\le T}(s) + \tilde{\varphi}_{\ge T}(s).} \tag{6.35}$$

This is a meromorphic function of s and is independent of the choice of T, since the effect of changing T to T' is simply to add the everywhere holomorphic function $\int_T^{T'} \varphi(t)\, t^{s-1}\, dt$ to $\tilde{\varphi}_{\le T}(s)$ and subtract the same function from $\tilde{\varphi}_{\ge}(s)$, not affecting the sum of their analytic continuations.

In summary, if $\varphi(t)$ is a function of t with asymptotic expansions as a sum of powers of t (or of powers of t multiplied by integral powers of $\ln t$) at both zero and infinity, then we can define in a canonical way a Mellin transform $\tilde{\varphi}(s)$ which is meromorphic in the entire s-plane and whose poles reflect directly the coefficients in the asymptotic expansions of $\varphi(t)$. This definition is consistent with and has the same properties (6.32) as the original definition (6.31).

We end this section by giving two simple examples, while Sections 6.7.2 and 6.7.3 will give further applications of the method.

Example 1. Let $\varphi(t) := t^\alpha$, where α is a complex number. Then φ has an asymptotic expansion (6.33) at 0 with a single term $\alpha_1 = \alpha$, $a_1 = 1$, and an asymptotic expansion (6.34) at ∞ with a single term $\beta_1 = \alpha$, $b_1 = 1$. We immediately find that $\tilde{\varphi}_{\le T}(s) = T^{s+\alpha}/(s+\alpha)$ for $\Re(s+\alpha) > 0$ and

$$\tilde{\varphi}_{\ge T}(s) = -T^{s+\alpha}/(s+\alpha)$$

for $\Re(s+\alpha) < 0$, so that, although the original Mellin transform integral (6.31) does not converge for any value of s, the function $\tilde{\varphi}(s)$ defined as the sum of the meromorphic continuations of $\tilde{\varphi}_{\le T}(s)$ and $\tilde{\varphi}_{\ge T}(s)$ makes sense, is independent of T, and in fact is identically zero. More generally, we find that

$$\tilde{\varphi}(s) \equiv 0$$

whenever $\varphi(t)$ is a finite linear combination of functions of the form $t^\alpha \ln^m t$ with the exponents $\alpha \in \mathbb{C}$ and $m \in \mathbb{Z}_{\ge 0}$. (These are exactly the functions whose images $\varphi_\lambda(t) := \varphi(\lambda t)$ under the action of the multiplicative group $\mathbb{R}_+^\times$ of positive real numbers span a finite-dimensional space.) In particular, we see that the generalized Mellin transformation is no longer injective.

Example 2. Let $\varphi(t) := \mathrm{e}^{-t}$. Here the integral (6.31) converges for $\Re(s) > 0$ and defines Euler's gamma-function

$$\Gamma(s) = \int_0^\infty \mathrm{e}^{-t} t^{s-1} dt \qquad (\Re(s) > 0). \tag{6.36}$$

From the fact that $\varphi(t)$ is of rapid decay at infinity and has the asymptotic (here even convergent) expansion $\sum_{n=0}^\infty (-t)^n/n!$ at zero, we deduce that $\Gamma(s) = \tilde{\varphi}(s)$ has a meromorphic continuation to all s with a simple pole of residue $(-1)^n/n!$ at $s = -n$ $(n = 0, 1, \dots)$ and no other poles.[17] From the first of the properties listed in (6.32), we find the following formula, which we will use many times:

$$\boxed{\varphi(t) = \mathrm{e}^{-\lambda t} \quad \Rightarrow \quad \tilde{\varphi}(s) = \Gamma(s)\, \lambda^{-s} \qquad (\lambda > 0).} \tag{6.37}$$

[17] Of course, in this special case these well-known properties can also be deduced from the functional equation $\Gamma(s+1) = s\Gamma(s)$ (proved for $\Re(s) > 0$ by integration by parts in the integral (6.36) defining $\Gamma(s)$), N applications of which gives the meromorphic extension $\Gamma(s) = s^{-1}(s+1)^{-1} \cdots (s+N-1)^{-1}\Gamma(s+N)$ of $\Gamma(s)$ to the half-plane $\Re(s) > -N$.

6.7.2 Dirichlet Series and their Special Values

In this section we look at functions $\varphi(t)$ for which the Mellin transform defined in Sect. 6.7.1 is related to a Dirichlet series. The key formula is (6.37), because it allows us to convert Dirichlet series into exponential series, which are much simpler.
Example 3. Define $\varphi(t)$ for $t > 0$ by $\varphi(t) = 1/(\mathrm{e}^t - 1)$. This function is of rapid decay at infinity and has an asymptotic expansion (actually convergent for $t < 2\pi$)

$$\frac{1}{\mathrm{e}^t - 1} = \frac{1}{t + \frac{t^2}{2} + \frac{t^3}{6} + \cdots} = \sum_{r=0}^{\infty} \frac{B_r}{r!}\, t^{r-1} \tag{6.38}$$

with certain rational coefficients $B_0 = 1$, $B_1 = -\frac{1}{2}$, $B_2 = \frac{1}{6}, \ldots$ called *Bernoulli numbers*. From the results of Sect. 6.7.1 we know that the Mellin transform $\tilde{\varphi}(s)$, originally defined for $\Re(s) > 1$ by the integral (6.31) has a meromorphic continuation to all s with simple poles of residue $B_r/r!$ at $s = 1 - r$ $(r = 0, 1, 2, \ldots)$. On the other hand, since $\mathrm{e}^t > 1$ for $t > 0$, we can expand $\varphi(t)$ as a geometric series $\mathrm{e}^{-t} + \mathrm{e}^{-2t} + \mathrm{e}^{-3t} + \cdots$, so (6.37) gives (first in the region of convergence) $\tilde{\varphi}(s) = \Gamma(s)\zeta(s)$, where

$$\zeta(s) = \sum_{m=1}^{\infty} \frac{1}{m^s} \qquad (\Re(s) > 1) \tag{6.39}$$

is the Riemann zeta function. Since $\Gamma(s)$, as we have seen is also meromorphic, with simple poles of residue $(-1)^n/n!$ at non-positive integral arguments $s = -n$ and no other poles, and since $\Gamma(s)$ (as is well-known and easily proved) never vanishes, we deduce that $\zeta(s)$ has a meromorphic continuation to all s with a unique simple pole of residue $1/\Gamma(1) = 1$ at $s = 1$ and that its values at non-positive integral arguments are rational numbers expressible in terms of the Bernoulli numbers:[18]

$$\zeta(-n) = (-1)^n \frac{B_{n+1}}{n+1} \qquad (n = 0,\ 1,\ 2,\ \ldots). \tag{6.40}$$

Example 4. To approach $\zeta(s)$ in another way, we choose for $\varphi(t)$ the *theta function*

$$\vartheta(t) := \sum_{n=-\infty}^{\infty} \mathrm{e}^{-\pi n^2 t} \qquad (t > 0). \tag{6.41}$$

(The factor π in the exponent has been included for later convenience.) We can write this out as

$$\vartheta(t) = 1 + 2\,\mathrm{e}^{-\pi t} + 2\,\mathrm{e}^{-4\pi t} + \cdots, \tag{6.42}$$

and since the generalized Mellin transform of the function 1 is identically 0 by Example 1, we deduce from (6.37) that $\tilde{\varphi}(s) = 2\,\zeta^*(2s)$, where

$$\zeta^*(s) := \pi^{-s/2}\,\Gamma\left(\frac{s}{2}\right)\,\zeta(s). \tag{6.43}$$

To obtain the analytic properties of $\zeta(s)$ from the results of Sect. 6.7.1, we need the asymptotics of $\vartheta(t)$ at zero and infinity. They follow immediately from the following famous result, due to Jacobi:

[18] Jakob Bernoulli (1654–1705), Maclaurin (1698–1746), Euler (1707–1783), Fourier (1768–1830), Gauss (1777-1855), Poisson (1781–1840), Jacobi (1804–1851), Dirichlet (1805–1859), Catalan (1814–1894), Riemann (1826–1866), Dedekind (1831–1916), Mellin (1854–1933), Hurwitz (1859–1919), Epstein (1871–1939).

Proposition 6.3 *The function $\vartheta(t)$ satisfies the functional equation*

$$\vartheta(t) = \frac{1}{\sqrt{t}}\,\vartheta\left(\frac{1}{t}\right) \qquad (t > 0). \tag{6.44}$$

Proof. Formula (6.44) is a special case of the *Poisson summation formula*, which says that

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{n\in\mathbb{Z}} \hat{f}(n) \tag{6.45}$$

for any sufficiently well-behaved (i.e., smooth and small at infinity) complex-valued function f on the real line, where

$$\hat{f}(y) := \int_{-\infty}^{\infty} f(x)\,\mathrm{e}^{2\pi\mathrm{i}xy}\,dx$$

is the Fourier transform of f, up to rescaling.[19] Now let us consider the function $f_t(x) := \mathrm{e}^{-\pi t x^2}$. Its Fourier transform is given by

$$\hat{f}_t(y) = \int_{-\infty}^{\infty} \mathrm{e}^{-\pi t x^2 + 2\pi\mathrm{i}xy}\,dx = \mathrm{e}^{-\pi y^2/t}\int_{-\infty}^{\infty} \mathrm{e}^{-\pi t(x+\mathrm{i}y/t)^2}\,dx = \frac{c}{\sqrt{t}}\,f_{1/t}(y),$$

where $c > 0$ is the constant $c = \int_{-\infty}^{\infty} \mathrm{e}^{-\pi x^2}dx$. Applying (6.45) with $f = f_t$ therefore gives $\vartheta(t) = ct^{-1/2}\vartheta(1/t)$, and taking $t = 1$ in this formula gives $c = 1$ and proves equation (6.44). □

Now we find from (6.42) that $\vartheta(t)$ has the asymptotic expansions

$$\vartheta(t) = 1 + O(t^{-N}) \qquad \text{as } t \to \infty$$

and $\vartheta(t) = t^{-1/2} + O(t^N)$ as $t \to 0$, where $N > 0$ is arbitrary. It follows from the results of Sect. 6.7.1 that its Mellin transform $\tilde{\vartheta}(s)$ has a meromorphic extension to all s with simple poles of residue 1 and -1 at $s = 1/2$ and $s = 0$, respectively, and no other poles. From the formula $\zeta^*(s) = \frac{1}{2}\tilde{\vartheta}(s/2)$ we deduce that the function $\zeta^*(s)$ defined in (6.43) is meromorphic having simple poles of residue 1 and 0 at $s = 1$ and $s = 0$ and no other poles and hence (using once again that $\Gamma(s)$ has simple poles at non-positive integers and never vanishes) that $\zeta(s)$ itself is holomorphic except for a single pole of residue 1 at $s = 1$ and vanishes at negative even arguments $s = -2, -4, \ldots$. This is weaker than (6.40), which gives a formula for $\zeta(s)$ at all non-positive arguments (and also shows the vanishing at negative even integers because it is an exercise to deduce from the definition (6.38) that B_r vanishes for odd $r > 1$). The advantage of the second approach to $\zeta(s)$ is that from equation (6.44) and the properties of Mellin transforms listed in (6.32) we immediately deduce the famous functional equation

$$\zeta^*(s) = \zeta^*(1 - s) \tag{6.46}$$

of the Riemann zeta-function which was discovered (for integer values $\neq 0, 1$ of s) by Euler in 1749 and proved (for all complex values $\neq 0, 1$ of s) by Riemann in 1859 by just this argument.

[19] To prove this, note that the function $F(x) = \sum_{n\in\mathbb{Z}} f(n + x)$ is periodic with period 1, so has a Fourier expansion $F(x) = \sum_{m\in\mathbb{Z}} c_m \mathrm{e}^{2\pi\mathrm{i}mx}$ with the Fourier coefficients $c_m = \int_0^1 F(x)\mathrm{e}^{-2\pi\mathrm{i}mx}dx = \hat{f}(-m)$. Now set $x = 0$.

We next generalize the method of Example 3. Consider a generalized Dirichlet series

$$L(s) = \sum_{m=1}^{\infty} c_m \, \lambda_m^{-s} \tag{6.47}$$

where the λ_m are real numbers satisfying $0 < \lambda_1 < \lambda_2 < \cdots$ and growing at least as fast as some positive power of m. (This is an ordinary Dirichlet series if $\lambda_m = m$ for all m.) Assume that the series converges for at least one value s_0 of s. Then it automatically converges in a half-plane (for instance, if $\lambda_m = m$ then the fact that $c_m = O(m^{s_0})$ implies convergence in the half-plane $\Re(s) > \Re(s_0) + 1$) and the associated exponential series

$$\varphi(t) = \sum_{m=1}^{\infty} c_m \, \mathrm{e}^{-\lambda_m t} \qquad (t > 0) \tag{6.48}$$

converges for all positive values of t. We then have:

Proposition 6.4 *Let $L(s)$ be a generalized Dirichlet series as in* (6.47)*, convergent somewhere, and assume that the function $\varphi(t)$ defined by* (6.48) *has an asymptotic expansion of the form*

$$\varphi(t) \;\sim\; \sum_{n=-1}^{\infty} a_n \, t^n \qquad (t \to +0). \tag{6.49}$$

Then $L(s)$ has a meromorphic continuation to all s, with a simple pole of residue a_{-1} at $s = 1$ and no other singularities, and its values at non-positive integers are given by

$$L(-n) = (-1)^n \, n! \, a_n \qquad (n \,=\, 0,\, 1,\, 2,\, \ldots)\,. \tag{6.50}$$

Proof. The function $\varphi(t)$ is of rapid decay at infinity and has the asymptotic expansion (6.49) at zero, so by the results of Sect. 6.7.1 we know that its Mellin transform $\tilde{\varphi}(s)$ extends meromorphically to all s, with simple poles of residue a_n at $s = -n$ ($n = -1,\, 0,\, 1, \ldots$). On the other hand, $\tilde{\varphi}(s)$ is equal to $\Gamma(s)L(s)$ by formula (6.37), and we know that $\Gamma(s)$ has simple poles of residue $(-1)^n n!$ at $s = -n$ ($n = 0,\, 1, \ldots$), has no other zeros or poles, and equals 1 at $s = 1$. The result follows. □

Example 5. Consider the Dirichlet series defined by

$$L_4(s) := \frac{1}{1^s} \,-\, \frac{1}{3^s} + \frac{1}{5^s} \,-\, \cdots \qquad (\Re(s) > 1). \tag{6.51}$$

Here the function defined by (6.48) is given by

$$\varphi(t) = \mathrm{e}^{-t} \,-\, \mathrm{e}^{-3t} \,+\, \mathrm{e}^{-5t} \,-\, \cdots = \frac{1}{\mathrm{e}^t + \mathrm{e}^{-t}} = \frac{1}{2\,\cosh t}$$

(geometric series). The asymptotic expansion of this function at $t = 0$ has the form

$$\varphi(t) = \frac{1/2}{1 + t^2/2! + t^4/4! + \cdots} = \frac{1}{2} \sum_{n=0}^{\infty} E_n \, \frac{t^n}{n!}$$

where the coefficients $E_0 = 1,\ E_1 = 0,\ E_2 = -1, \ldots$ are certain integers called the *Euler numbers*. It follows from the proposition that the function $L_4(s)$ has a

holomorphic continuation to all s and that $L_4(-n) = E_n/2$ for all $n \geq 0$. (We can omit the factor $(-1)^n$ because $E_n = 0$ for n odd.) The same method works for any Dirichlet series of the form $\sum_{m=1}^{\infty} \chi(m) m^{-s}$ with coefficients $\chi(m)$ which are periodic of some period (here 4), the most important case being that of Dirichlet L-series, where the coefficients $\chi(m)$ also satisfy $\chi(m_1 m_2) = \chi(m_1)\chi(m_2)$ for all m_1 and m_2.

Example 6. As a final example, consider the *Hurwitz zeta function*, defined by

$$\zeta(s,a) := \sum_{n=0}^{\infty} \frac{1}{(n+a)^s} \qquad (a > 0,\ \Re(s) > 0). \tag{6.52}$$

Here $\varphi(t) = \sum_{n=0}^{\infty} \mathrm{e}^{-(n+a)t} = \frac{\mathrm{e}^{-at}}{1-\mathrm{e}^{-t}}$. But for any x we have the expansion

$$\frac{\mathrm{e}^{xt}}{\mathrm{e}^t - 1} \sim \sum_{n=0}^{\infty} \frac{B_n(x)}{n!}\, t^n \qquad (t \to 0) \tag{6.53}$$

where the $B_n(x)$ are the *Bernoulli polynomials*

$$B_n(x) := \sum_{r=0}^{n} \binom{n}{r} B_r\, x^{n-r} \tag{6.54}$$

($B_0(x) = 1$, $B_1(x) = x - \frac{1}{2}$, $B_2(x) = x^2 - x + \frac{1}{6}, \dots$). We deduce that $\zeta(s,a)$ has a meromorphic continuation in s with a simple pole of residue 1 (independent of a) at $s = 1$, and, generalizing (6.40), that its values at non-positive integers are given by

$$\zeta(s,-n) = -\frac{B_{n+1}(a)}{n+1} \qquad (n = 0,\ 1,\ 2, \dots). \tag{6.55}$$

6.7.3 Application: the Casimir Effect

In the study of the Casimir effect, described in Sect. 6.6, one encounters the "function" defined by the series

$$F(\lambda) = -2\pi \sum_{l,\,m,\,n \in \mathbb{Z}} \sqrt{l^2 + m^2 + \lambda^2 n^2} \tag{6.56}$$

where λ is a positive real variable. (Here the factor -2π has been included for later convenience.) Of course this series is divergent. One would like to answer the following questions:

A. How can $F(\lambda)$ be defined rigorously?

B. How can $F(\lambda)$ be computed effectively for a given $\lambda > 0$?

C. How does $F(\lambda)$ behave asymptotically as $\lambda \to 0$ and as $\lambda \to \infty$?

(For the analysis of the Casimir effect, it is the asymptotics at $\lambda \to \infty$ which are important.) Using the ideas explained in the previous two sections, we will show that the answers are as follows:

A. Let $\lambda > 0$. For complex s with $\Re(s) > \frac{3}{2}$, define

$$Z(\lambda, s) := {\sum_{l,\,m,\,n \in \mathbb{Z}}}' \quad \frac{1}{(l^2 + m^2 + \lambda^2 n^2)^s} \tag{6.57}$$

(the prime on the summation sign means that the term $(l,m,n) = (0,0,0)$ is to be omitted), a so-called Epstein zeta function. Then $Z(\lambda, s)$ has a meromorphic

continuation to all s, with a simple pole of residue $2\pi/\lambda$ at $s = 3/2$ and no other poles, and satisfies the functional equation

$$Z^*(\lambda, s) = \frac{1}{\lambda}\, Z^*\left(\frac{1}{\lambda}, \frac{3}{2} - s\right), \tag{6.58}$$

where

$$Z^*(\lambda, s) \;:=\; \pi^{-s}\, \Gamma(s)\, Z(\lambda, s)\,. \tag{6.59}$$

Then one makes sense of (6.56) by setting $F(\lambda) = Z^*(\lambda, -\frac{1}{2})$.[20]

B. The value of the function $F(\lambda)$ is given for any positive real number T by

$$\begin{aligned} F(\lambda) = \frac{1}{\sqrt{T}} \sum_{l,\,m,\,n} \gamma_{-\frac{1}{2}}\big(\pi T(l^2 + m^2 + \lambda^2 n^2)\big) \\ + \frac{1}{\lambda T^2} \sum_{l,m,n} \gamma_2\left(\frac{\pi}{T}(l^2 + m^2 + \lambda^{-2} n^2)\right), \end{aligned} \tag{6.60}$$

where the sums are taken over all triples $(l, m, n) \in \mathbb{Z}^3$ and the functions $\gamma_{-\frac{1}{2}}(x)$ and $\gamma_2(x)$ are defined by the formulas

$$\gamma_{-\frac{1}{2}}(x) := \int_1^\infty \mathrm{e}^{-xt}\, \frac{dt}{t^{3/2}} \qquad (x \geq 0) \tag{6.61}$$

(a variant of the error function) and

$$\gamma_2(x) := \begin{cases} \left(\frac{1}{x} + \frac{1}{x^2}\right) \mathrm{e}^{-x} & \text{if} \quad x > 0 \\ -\frac{1}{2} & \text{if} \quad x = 0, \end{cases} \tag{6.62}$$

respectively. Since both $\gamma_{-\frac{1}{2}}(x)$ and $\gamma_2(x)$ are $O(\mathrm{e}^{-x})$ as $x \to \infty$, formula (6.60) makes $F(\lambda)$ rapidly computable. More precisely, if we choose $T = \lambda^{-2/3}$ then there are (uniformly in λ) only $O(M^{3/2})$ terms in the two sums in (6.60) for which the arguments of $\gamma_{-\frac{1}{2}}$ or γ_2 are $\leq M$, so that a relatively small number of terms suffices to compute $F(\lambda)$ to high precision. Here are some sample values:

t	$F(t)$
0.1	6.211157049634453208312778213637817391711766753 71 . . .
0.5	1.744906668422350545220029681769409791799015923 36 . . .
1	1.675073821392163756773789658549957277747090020 78 . . .
2	3.192402282746911828637014055947382864058909476 11 . . .
10	220.762287791317835036587359031113282945900247337 . . .

(6.63)

Each of these numbers was computed independently using (6.60) for several different values of T. The fact that the answers agreed to the precision given, even though the individual terms of the sums are completely different, gives a high degree of confidence in the correctness of the theoretical and numerical calculations.

[20] Note that $\pi^{1/2}\, \Gamma(-\frac{1}{2}) = -2\pi$.

Of course we could also use the functional equation (6.58) to give a much simpler convergent series expansion for $F(\lambda)$ than (6.60), namely

$$F(\lambda) = \frac{1}{\lambda}\, Z^*\left(\frac{1}{\lambda}, 2\right) = \frac{\lambda^3}{\pi^2} \sum_{l,\,m,\,n}{}' \frac{1}{(\lambda^2 l^2 + \lambda^2 m^2 + n^2)^2}\,, \tag{6.64}$$

but this formula would be useless for practical purposes because of the slow convergence: summing over $|l|$, $|m|$, $|n| \le N$ involves $O(N^3)$ terms and gives an error of the order of $1/N$, so that we would need some 10^{10} terms to achieve even three digits of precision.

C. The value of $F(\lambda)$ is given for small λ by

$$F(\lambda) = \frac{2C}{3\lambda} + \frac{\pi}{3}\,\lambda + O(\lambda \mathrm{e}^{-\pi/\lambda}) \qquad (\lambda \to 0) \tag{6.65}$$

and for large λ by

$$F(\lambda) = \frac{\pi^2}{45}\,\lambda^3 + C' + O(\sqrt{\lambda}\,\mathrm{e}^{-\pi\lambda}) \qquad (\lambda \to \infty) \tag{6.66}$$

where $C = 1 - \frac{1}{9} + \frac{1}{25} - \cdots = L_4(2)$ (with L_4 as in (6.51) is Catalan's constant and $C' = \frac{2}{\pi}\zeta(\frac{3}{2})L_4(\frac{3}{2})$. The values obtained for $\lambda = 0.1$ and $\lambda = 10$ by retaining only the first two terms in equations (6.65) or (6.66), respectively, are

$$F(0.1) = 6.211157049634453208312778212325210834634234808 52\ldots,$$
$$F(10) = 220.762287791317835036587358989193358212711493834\ldots$$

extremely close to the exact values for these two numbers given above. Let us now prove each of these assertions.

Proof of A. From (6.37) and the fact that the generalized Mellin transform of the constant function 1 vanishes, we deduce that the function $Z^*(\lambda, s)$ defined by (6.59) equals the Mellin transform $\tilde{\varphi}_\lambda(s)$ of the function

$$\varphi_\lambda(t) := \sum_{l,\,m,\,n\in\mathbb{Z}} \mathrm{e}^{-\pi(l^2+m^2+\lambda^2 n^2)t} = \vartheta(t)^2\,\vartheta(\lambda^2 t) \qquad (\lambda,\ t > 0)\,,$$

where $\vartheta(t)$ is the theta series defined in (6.41). The functional equation (6.44) of $\vartheta(t)$ implies the functional equation $\varphi_\lambda(t) = \lambda^{-1}t^{-3/2}\varphi_{1/\lambda}(t^{-1})$ of $\varphi_\lambda(t)$, and the meromorphic continuation, description of poles and functional equation (6.58) of $Z^*(s)$ then follow as in Example 4 above.

Proof of B. Since the function $\varphi(t) = \varphi_\lambda(t)$ equals 1 to all orders in t as $t \to \infty$ and (by virtue of its functional equation) equals $\lambda^{-1}t^{-3/2}$ to all orders in t as $t \to 0$, the two pieces of the decomposition (6.35) of its Mellin transform are given by

$$\tilde{\varphi}_{\lambda,\ge T}(s) = \int_T^\infty \bigl(\varphi_\lambda(t) \,-\, 1\bigr)\, t^{s-1}\, dt \,-\, \frac{T^s}{s} \tag{6.67}$$

and

$$\tilde{\varphi}_{\lambda,\le T}(s) = \int_0^T \bigl(\varphi_\lambda(t) \,-\, \lambda^{-1}\, t^{-3/2}\bigr)\, t^{s-1}\, dt + \frac{\lambda^{-1}\, T^{s-\frac{3}{2}}}{s-\frac{3}{2}},$$

respectively. Using the functional equation of $\varphi_\lambda(t)$, we see that these are exchanged when we replace λ, s and T by λ^{-1}, $\frac{3}{2} - s$ and T^{-1}, respectively (again making

the functional equation of $Z^*(\lambda, s) = \tilde{\varphi}_\lambda(s)$ evident), so we only have to study $\tilde{\varphi}_{\lambda,\geq T}(s)$. Substituting the definition of $\varphi_\lambda(t)$ into (6.67), we find

$$\tilde{\varphi}_{\lambda,\geq T}(s) = T^s \sum_{l,m,n}{}' \gamma_s\big(\pi T(l^2+m^2+\lambda^2 n^2)\big) \;-\; \frac{T^s}{s}\,,$$

where $\gamma_s(x)$, essentially the incomplete gamma function, is defined for $x > 0$ by

$$\gamma_s(x) := \int_1^\infty t^{s-1}\,\mathrm{e}^{-xt}\,dt \qquad (x>0)\,.$$

The extra term $-T^s/s$ in this formula can be omitted if we drop the prime from the summation signs and define the value of $\gamma_s(0)$ as $-1/s$ (which is indeed the limiting value of $\gamma_s(x)$ as $x \to 0$ if $\Re(s) < 1$). The final result, with this convention for $\gamma_s(0)$, is therefore

$$\begin{aligned} Z^*(\lambda,\, s) = T^s \sum_{l,m,n} \gamma_s\big(\pi T(l^2+m^2+\lambda^2 n^2)\big) \\ + \left[s \mapsto \frac{3}{2} - s,\ \lambda \mapsto \lambda^{-1},\ T \mapsto T^{-1}\right]. \end{aligned}$$

The special case $s = -\frac{1}{2}$ gives the expansion (6.60).

Proof of C. The leading term in the expansion of $F(\lambda)$ when λ is very big or very small can be obtained from equation (6.64): if λ is small then the dominating terms in (6.64) are those with $n = 0$, so $F(\lambda)$ is asymptotically equal to

$$\pi^{-2}\lambda^{-1}\sum_{l,m}{}'(l^2+m^2)^{-2} = 4\pi^{-2}\zeta(2)L_4(2)\lambda^{-1}$$

(here we have used that $\frac{1}{4}\sum'_{l,m}(l^2+m^2)^{-s}$ is the Dedekind zeta function of the field $\mathbb{Q}(\mathrm{i})$, which factors as $\zeta(s)$ times $L_4(s)$), while for large λ the dominating terms are those with $l = m = 0$ and we get $F(\lambda) \sim 2\pi^{-2}\zeta(4)\lambda^3$. To get a more precise estimate, we use equation (6.60).

Consider first the case $\lambda \to 0$, and choose T in (6.60) so that $T \to \infty$, $\lambda^2 T \to 0$. (The best choice will turn out to be $T = \lambda^{-1}$.) Then in view of the exponential decay $\gamma_s(x) = O_s(x^{-1}\mathrm{e}^{-\pi x})$ of $\gamma_s(x)$ when $x \to \infty$, we find that the only terms in (6.60) which are not exponentially small are those with $(l,m) = (0,0)$ in the first sum and those with $n = 0$ in the second one. Hence

$$\begin{aligned} F(\lambda) &= \frac{1}{\sqrt{T}}\left[f_1(\lambda^2 T) + O\left(\frac{1}{T}\sum_{l,m}{}'\mathrm{e}^{-\pi T(l^2+m^2)}\cdot\sum_n \mathrm{e}^{-\pi\lambda^2 T n^2}\right)\right] \\ &\quad + \frac{1}{\lambda T^2}\left[f_2\left(\frac{1}{T}\right) + O\left(\sum_{l,m}\mathrm{e}^{-\pi(l^2+m^2)/T}\cdot\lambda^2 T\sum_n{}'\mathrm{e}^{-\pi n^2/\lambda^2 T}\right)\right] \\ &= \frac{1}{\sqrt{T}}\,f_1(\lambda^2 T) + \frac{1}{\lambda T^2}\,f_2\left(\frac{1}{T}\right) + O\left(\frac{1}{\lambda T^2}\,\mathrm{e}^{-\pi T} + \lambda\,\mathrm{e}^{-\pi/\lambda^2 T}\right), \end{aligned} \tag{6.68}$$

where

$$f_1(\varepsilon) := \sum_{n\in\mathbb{Z}} \gamma_{-1/2}(\pi\varepsilon n^2)\,, \qquad f_2(\varepsilon) := \sum_{l,\,m\in\mathbb{Z}} \gamma_2(\pi\varepsilon(l^2+m^2))\,.$$

From the definition (6.61) and the functional equation of $\vartheta(x)$ we find (for ε small)

$$\begin{aligned} f_1(\varepsilon) &= \int_1^\infty t^{-3/2}\,\vartheta(\varepsilon t)\,dt = \sqrt{\varepsilon}\int_\varepsilon^\infty \vartheta(t)\,\frac{dt}{t^{3/2}} = \sqrt{\varepsilon}\int_0^{1/\varepsilon}\vartheta(x)\,dx \\ &= \frac{1}{\sqrt{\varepsilon}} + \sqrt{\varepsilon}\int_0^{1/\varepsilon}\big(\vartheta(x)-1\big)\,dx = \frac{1}{\sqrt{\varepsilon}} + \sqrt{\varepsilon}\,\Big(\frac{\pi}{3} + O\big(\mathrm{e}^{-\pi/\varepsilon}\big)\Big)\,, \end{aligned}$$

since

$$\int_0^\infty \big(\vartheta(x)-1\big)\,dx = 2\sum_{n=1}^\infty \int_0^\infty \mathrm{e}^{-\pi n^2 x}\,dx = \frac{2}{\pi}\sum_{n=1}^\infty \frac{1}{n^2} = \frac{\pi}{3}\,.$$

Similarly, noting the exceptional case $x = 0$ in the definition (6.62) of $\gamma_2(x)$, we find

$$\begin{aligned} f_2(\varepsilon) + \frac{1}{2} &= \int_1^\infty t\,\big(\vartheta(\varepsilon t)^2 - 1\big)\,dt = \frac{1}{\varepsilon^2}\int_\varepsilon^\infty x\,\big(\vartheta(x)^2-1\big)\,dx \\ &= \frac{1}{\varepsilon^2}\,\left(\int_0^\infty x\,\big(\vartheta(x)^2-1\big)\,dx \;-\; \int_0^\varepsilon (1 - x + O(\mathrm{e}^{-\pi/x}))\,dx\right) \\ &= \frac{2C}{3\varepsilon^2} \;-\; \frac{1}{\varepsilon} + \frac{1}{2} + O(\mathrm{e}^{-\pi/\varepsilon})\,, \end{aligned}$$

where this time we have used

$$\begin{aligned} \int_0^\infty x\,\big(\vartheta(x)^2-1\big)\,dx &= \sum_{l,\,m}{}' \int_0^\infty x\mathrm{e}^{-\pi(l^2+m^2)x}\,dx \\ &= \frac{1}{\pi^2}\sum_{l,\,m}{}' \frac{1}{(l^2+m^2)^2} = \frac{4}{\pi^2}\,\zeta(2)L_4(2)\,. \end{aligned}$$

Inserting these formulas into (6.68), we find

$$\begin{aligned} F(\lambda) &= \frac{1}{\sqrt{T}}\,\left[\frac{1}{\lambda\sqrt{T}} + \frac{\pi}{3}\,\lambda\sqrt{T} + O(\lambda\sqrt{T})\,\mathrm{e}^{-\pi/\lambda^2 T})\right] \\ &\qquad + \frac{1}{\lambda T^2}\,\big[\frac{2C}{3}\,T^2 \;-\; T + O(\mathrm{e}^{-\pi T})\big] + O\left(\frac{1}{\lambda T^2}\,\mathrm{e}^{-\pi T} + \lambda\,\mathrm{e}^{-\pi/\lambda^2 T}\right) \\ &= \frac{2C}{3\lambda^2} + \frac{\pi}{3}\,\lambda + O\left(\frac{1}{\lambda T^2}\,\mathrm{e}^{-\pi T} + \lambda\,\mathrm{e}^{-\pi/\lambda^2 T}\right)\,. \end{aligned}$$

(Note how the two non-exponentially small terms which depend on T cancel, as they have to.) Taking $T = 1/\lambda$ gives (6.65).

The proof of (6.66) is completely analogous. From (6.60) and the exponential decay of $\gamma_s(x)$ we get

$$F(\lambda) = \frac{1}{\sqrt{T}}\,f_3(T) + \frac{1}{\lambda T^2}\,f_4\left(\frac{1}{\lambda^2 T}\right) + O\left(\frac{\mathrm{e}^{-\pi\lambda^2 T}}{\lambda^2 T^{5/2}} + \frac{\mathrm{e}^{-\pi/T}}{T^{1/2}}\right)$$

as $\lambda \to \infty$, where $T \to 0$ with $\lambda^2 T \to \infty$ and where $f_3(\varepsilon)$ and $f_4(\varepsilon)$ are defined by

$$f_3(\varepsilon) := \sum_{l,\,m\in\mathbb{Z}} \gamma_{-1/2}(\pi\varepsilon(l^2+m^2))\,, \qquad f_4(\varepsilon) := \sum_{n\in\mathbb{Z}} \gamma_2(\pi\varepsilon n^2)\,.$$

This time we find the expansions

$$f_3(\varepsilon) = \frac{2}{3\varepsilon} + C'\sqrt{\varepsilon} + O(\mathrm{e}^{-\pi/\varepsilon})$$

and

$$f_4(\varepsilon) = \frac{\pi^2}{45\varepsilon^2} - \frac{2}{3\sqrt{\varepsilon}} + O(\sqrt{\varepsilon}\,\mathrm{e}^{-\pi/\varepsilon})$$

for ε small, the non-exponentially small terms which depend on T again cancel, and taking $T = 1/\lambda$ to make the error as small as possible we obtain (6.66). The details are left to the reader.

As a final comment, we mention that the constants C and C' occurring in (6.65) and (6.66) can be evaluated rapidly by the same method as in **B** above, with $\vartheta(t)^2$ instead of $\vartheta(t)^2\vartheta(\lambda t)$, using the fact that the Dirichlet series $4\zeta(s)L_4(s)$ is the Mellin transform of $\vartheta(t)^2$. Their numerical values are

$$C = 0.91596559417721901505460351493238411077414937428167213 4\ldots\,,$$
$$C' = 1.43774554488764350693200343638999965018484034037993399 7\ldots$$

6.7.4 Asymptotics of Series of the Form $\sum f(nt)$

In this section we describe an extremely useful, and not sufficiently well known, asymptotic formula for functions given by expansions of the form

$$g(x) = f(x) + f(2x) + f(3x) + \cdots \tag{6.69}$$

where $f :]0,\infty[\to \mathbb{C}$ is a smooth function of sufficiently rapid decay to ensure the convergence of the series (say $f(x) = O(x^{-1-\varepsilon})$ as $x \to \infty$) and having a known asymptotic expansion at $x = 0$. In the simplest situation, we assume that f has a power series expansion (which may be only asymptotic rather than convergent, i.e. f need only be differentiable rather than analytic at the origin)

$$f(x) \;\sim\; \sum_{n=0}^{\infty} b_n\, x^n \qquad\qquad (x \to 0)\,. \tag{6.70}$$

First, let us try to guess what the answer should be. On the one hand we can argue à la Euler, simply substituting the expansion (6.70) into (6.69) and interchanging the order of summation, without worrying about convergence problems. This gives

$$g(x) \;\sim\; \sum_{m=1}^{\infty}\sum_{n=0}^{\infty} b_n\,(mx)^n = \sum_{n=0}^{\infty} b_n \left(\sum_{m=1}^{\infty} m^n\right) x^n = \sum_{n=0}^{\infty} b_n\,\zeta(-n)\,x^n\,. \tag{6.71}$$

Of course this calculation is meaningless, since not only is the interchange of summation not permitted, but each of the interior sums $\sum_m m^n$ is divergent. Nevertheless, we know from Sect. 6.7.2, and Euler knew non-rigorously in 1749, that the numbers $\zeta(-n)$ do make sense and are certain rational numbers given by equation (6.40), so that at least the final expression in (6.71) makes sense as a formal power series. Alternatively, we can proceed à la Riemann and consider (6.69) for x small as $1/x$ times an approximation to the integral

$$I_f = \int_0^{\infty} f(t)\,dt\,. \tag{6.72}$$

This suggests instead that the correct asymptotic expansion of $g(x)$ near 0 should be given by

$$g(x) \sim \frac{I_f}{x} \qquad (x \to 0)\,, \tag{6.73}$$

and indeed this formula is true, by the very definition of Riemann integrals as limits of sums, if we interpret the symbol of asymptotic equality "$\sim$" in its weak sense, as saying simply that the ratio of the expressions on its left and right tends to 1 as $x \to 0$. However, we are using "$\sim$" in its strong sense, where $g(x) \sim \sum_\lambda a_\lambda x^\lambda$ means that the difference between $g(x)$ and the finite sum $\sum_{\lambda<C} a_\lambda x^\lambda$ is $O(x^C)$ as $x \to 0$ for any value of C, no matter how large. In this stronger sense, neither (6.71) nor (6.73) gives the correct asymptotic development of g. Remarkably enough, however, their sum does give the right answer:

Proposition 6.5 *Let f be a smooth function on the positive real line which has the asymptotic development* (6.70) *at the origin and, together with all its derivatives, is of rapid decay at infinity. Then the function $g(x)$ defined by* (6.69) *has the asymptotic development*

$$g(x) \sim \frac{I_f}{x} + \sum_{n=0}^{\infty} b_n \frac{B_{n+1}}{n+1} (-x)^n \tag{6.74}$$

as $x \to 0$, where I_f is defined by equation (6.72).

Proof. We begin by describing the Euler–Maclaurin summation formula. To state it, we need the Bernoulli polynomials $B_n(x)$, which can be described by the generating function (6.54), by the explicit formula (6.55) in terms of Bernoulli numbers, or, most beautiful, by the property that

$$\int_a^{a+1} B_n(x)dx = a^n \qquad \text{for every } a.$$

(It is easy to see that there is only one polynomial $B_n(x)$ with this property for each n.) From any of these definitions we can deduce without difficulty that $B_n'(x) = nB_{n-1}(x)$ and that $B_n(x+1) - B_n(x) = nx^{n-1}$.

Now let f be a smooth function on the positive reals. Integration by parts and the fact that $B_1(0) = \frac{1}{2} = -B_1(0)$ but $B_{n+1}(1) = B_{n+1}(0) = B_{n+1}$ for $n \geq 2$ gives

$$\int_0^1 f^{(n)}(t) \frac{B_n(t)}{n!}\, dt = -\int_0^1 f^{(n+1)}(t) \frac{B_{n+1}(t)}{(n+1)!}\, dt + r_n$$

where $r_0 := \frac{1}{2}(f(0) + f(1))$ and

$$r_n := \frac{B_{n+1}}{(n+1)!}[f^{(n)}(1) - f^{(n)}(0)] \qquad (n \geq 1).$$

Hence, by induction on N,

$$\int_0^1 f(t)\, dt = \frac{1}{2}[f(0) + f(1)] + \sum_{n=1}^{N-1} \frac{(-1)^n B_{n+1}}{(n+1)!}[f^{(n)}(1) - f^{(n)}(0)] + (-1)^N \int_0^1 f^{(N)}(t) \frac{B_N(t)}{N!}\, dt$$

for every integer $N \geq 1$. Replacing $f(t)$ by $f(t+m-1)$ and summing over the natural numbers $m = 1, 2, \ldots, M$ gives

$$\int_0^M f(t)\,dt = \frac{f(0)}{2} + \sum_{m=1}^{M-1} f(m) + \frac{f(M)}{2} + \sum_{n=1}^{N-1} \frac{(-1)^n B_{n+1}}{(n+1)!}[f^{(n)}(M) - f^{(n)}(0)]$$
$$+(-1)^N \int_0^M f^{(N)}(t)\,\frac{\bar{B}_N(t)}{N!}\,dt \tag{6.75}$$

where $\bar{B}_N(t) := B_N(t - [t])$. This is the Euler-Maclaurin summation formula.[21]

Now assume that f and each of its derivatives is of rapid decay at infinity, so that $\int_0^\infty |f^{(N)}(t)|dt$ converges. Since $\bar{B}_N(t)$ is periodic and hence bounded, we can let $M \to \infty$ in (6.75) to get

$$\sum_{m=1}^{\infty} f(m) = \int_0^\infty f(t)\,dt + \sum_{n=0}^{N-1} \frac{(-1)^n B_{n+1}}{(n+1)!}\,f^{(n)}(0)$$
$$-(-1)^N \int_0^\infty f^{(N)}(t)\,\frac{\bar{B}_N(t)}{N!}\,dt\,.$$

Replacing $f(t)$ by $f(xt)$ and then t by t/x with $x > 0$ changes this formula to

$$\sum_{m=1}^{\infty} f(mx) = \frac{1}{x}\int_0^\infty f(t)\,dt + \sum_{n=0}^{N-1} \frac{(-1)^n B_{n+1}}{(n+1)!}\,f^{(n)}(0)\,x^n$$
$$+(-x)^{N-1} \int_0^\infty f^{(N)}(t)\,\frac{\bar{B}_N(t/x)}{N!}\,dt\,.$$

The last integral is bounded as $x \to 0$ with N fixed for the same reason as before, so the final term is $O_N(x^{N-1})$. Substituting $f^{(n)}(0) = n!\,b_n$ from (6.70), we obtain the desired asymptotic formula (6.74). □

Before giving examples of Prop. 6.5, we mention three extensions to more general sums.

1. First, instead of (6.69) we can look at shifted sums of the form

$$g(x) = \sum_{m=0}^{\infty} f((m+a)x)$$

where $a > 0$. Here the "Riemann Ansatz" and the "Euler Ansatz" predict $I_f x^{-1}$ and $\sum_{n=0}^\infty b_n \zeta(-n, a)x^n$ for the asymptotic expansion of g, and again the correct answer is the sum of these two:

$$\sum_{m=0}^{\infty} f((m+a)x) \;\sim\; \frac{I_f}{x} + \sum_{n=0}^{\infty} b_n\,\frac{B_{n+1}(a)}{n+1}\,x^n \tag{6.76}$$

(cf. equation (6.55)). The proof is similar to that of Prop. 6.5 and will be omitted. By taking rational values of a in (6.76) and forming suitable linear combinations, we can use this formula to give the asymptotic development of $\sum_{m\ge1} \chi(m)f(mx)$ as $x \to 0$ for any periodic function $\chi(m)$, such as a Dirichlet character.

2. Next, we can allow non-integral exponents of x in (6.70). If the expansion of $f(x)$ at the origin contains terms $b_\lambda x^\lambda$ with arbitrary real numbers $\lambda > -1$ (or complex numbers with real part greater than -1), then the formula for g need only be modified by adding the corresponding terms $b_\lambda \zeta(-\lambda)x^\lambda$. Terms with $\lambda < -1$

[21] The symbol $[t]$ denotes the largest integer m with $m \le t$. For example, $[2.5] = 2$.

are not interesting since they can simply be subtracted from $f(x)$, since the sum $\sum_m b_m(mx)\lambda$ then converges absolutely. The limiting case $\lambda = -1$ is of interest because it occurs in various applications. Here the answer (proved most easily by taking one function, like $x^{-1}\mathrm{e}^{-x}$, which has a $1/x$ singularity at the origin and for which $\sum f(mx)$ can be computed exactly) is

$$\begin{aligned} f(x) \;&\sim\; \sum_{\lambda\geq -1} b_\lambda\, x^\lambda \\ \Rightarrow \quad \sum_{m=1}^{\infty} f(mx) \;&\sim\; \frac{1}{x}\left(b_{-1}\,\ln\frac{1}{x} + I_f^*\right) + \sum_{\lambda>-1}^{\infty} b_\lambda\,\zeta(-\lambda)\,(-x)^\lambda\,. \end{aligned} \tag{6.77}$$

Here, we introduce the following integral $I_f^* := \int_0^\infty (f(t) - b_{-1}\mathrm{e}^{-t}/t)dt$.

3. Finally, we can also allow terms of the form $x^\lambda(\ln x)^n$ in the expansion of $f(x)$, the corresponding contribution to $g(x)$ being simply the nth derivative with respect to λ of $\zeta(-\lambda)x^\lambda$, e.g. a term $x^\lambda \ln\frac{1}{x}$ in the expansion of $f(x)$ at 0 leads to a term $x^\lambda\big(\zeta(-\lambda)\ln\frac{1}{x} + \zeta'(-\lambda)\big)$ in the expansion at 0 of $g(x)$. In particular, using the known value $\zeta'(0) = -\frac{1}{2}\ln(2\pi)$ we find

$$\begin{aligned} f(x) \;&\sim\; b\,\ln\frac{1}{x} + \sum_{n=0}^{\infty} b_n x^n \\ \Rightarrow \quad \sum_{m=1}^{\infty} f(mx) \;&\sim\; \frac{I_f}{x} \;-\; \frac{b}{2}\,\ln\frac{2\pi}{x} + \sum_{n=0}^{\infty} b_n\,\zeta(-n)\,x^n\,. \end{aligned} \tag{6.78}$$

We end by giving two easy and two harder examples to illustrate how these asymptotic formulas work.

Example 1. Take $f(x) := \mathrm{e}^{-\lambda x}$ with $\lambda > 0$. This function is smooth, small at infinity, and has an expansion (6.70) at $x = 0$ with $b_n = (-\lambda)^n/n!$. The integral I_f equals $1/\lambda$. Hence (6.74) gives

$$g(x) \;\sim\; \frac{1}{\lambda x} + \sum_{n=0}^{\infty} \frac{B_{n+1}}{(n+1)!}\,(\lambda x)^n$$

as the asymptotic expansion of $g(x) = \sum\limits_{m=1}^{\infty} f(mx) = \frac{1}{\mathrm{e}^{\lambda x}-1}$, in accordance with the definition (6.38) of the Bernoulli numbers.

Example 2. Now take $f(x) := \mathrm{e}^{-\lambda x^2}$ with $\lambda > 0$. This function is again smooth and small at infinity, and has an expansion (6.70) at $x = 0$ with $b_{2n} = (-\lambda)^n/n!$ and $b_n = 0$ for n odd. Since all Bernoulli numbers with odd indices > 1 vanish, the asymptotic expansion in (6.74) breaks off after two terms, and we find

$$g(x) = \frac{I_f}{x} - \frac{1}{2} + O(x^N)$$

for all N, with $I_f = \sqrt{\pi/4\lambda}$. In this case, of course, we know much more, because $g(x)$ is simply $\frac{1}{2}\big(\vartheta(\pi x^2) - 1\big)$ and therefore equation (6.44) gives the much more precise statement

$$g(x) = \frac{I_f}{x} - \frac{1}{2} + O\left(x^{-1}\mathrm{e}^{-\pi^2/\lambda x^2}\right).$$

The same applies to any function $f(x)$ whose expansion at $x = 0$ has only even powers of x. For such a function, the expansion (6.74) collapses to

$$g(x) \sim \frac{I_f}{x} - \frac{b_0}{2}$$

as $x \to 0$, but this is always just a weakening of the Poisson summation formula (6.45), because we can extend $f(x)$ by $f(-x) = f(x)$ to a smooth even function on the real line and use (6.46) to get the exact formula

$$2g(x) + f(0) = \sum_{n\in\mathbb{Z}} f(nx) = \frac{1}{x} \sum_{n\in\mathbb{Z}} \hat{f}(nx) = \frac{2I_f}{x} + \frac{2}{x} \sum_{n=1}^{\infty} \hat{f}(nx)\,,$$

and the smoothness of f implies that the function $\hat{f}(y)$ decays at infinity more rapidly than any negative power of y. The right way to think of Prop. 6.5 is therefore as a replacement for the Poisson summation formula when one is confronted with a sum over only positive integers rather than a sum over all of $\mathbb{Z}$. Such sums are very much harder to study than sums over all integers – just think of the special values of the Riemann zeta function, where the numbers $\zeta(2k)$ can be obtained in closed form because they can be written as

$$\zeta(2k) = \frac{1}{2} \sum_{n\in\mathbb{Z}} \frac{1}{n^k},$$

while the numbers $\zeta(2k+1)$, which cannot be reduced to sums over all of $\mathbb{Z}$ in this way, are *not* known exactly.

Example 3. As our next example, we consider the function

$$g_k(q) := \sum_{n=1}^{\infty} \sigma_{k-1}(n)\, q^n \qquad (0 < q < 1)\,,$$

where k is an integer greater than 1 and $\sigma_{k-1}(n)$ denotes the sum of the $(k-1)$st powers of the divisors of a natural number n. In the theory of modular forms, it is shown that if k is even and larger than 2, then g_k satisfies the functional equation[22]

$$-\frac{B_k}{2k} + g_k\big(\mathrm{e}^{-2\pi x}\big) = \frac{(-1)^{k/2}}{x^k} \left(-\frac{B_k}{2k} + g_k\big(\mathrm{e}^{-2\pi/x}\big) \right) \qquad (k = 4,\, 6,\, 8, \ldots;\; x > 0).$$

In particular, $g_k(\mathrm{e}^{-2\pi x})$ has the terminating asymptotic expansion

$$g_k(\mathrm{e}^{-2\pi x}) = -(-1)^{k/2} \frac{B_k}{2k}\, x^{-k} + \frac{B_k}{2k} + O(x^N) \qquad (N = 1, 2, \ldots) \qquad (6.79)$$

as $x \to 0$ in these cases. Let us see how Prop. 6.5 permits us to recover this asymptotic formula without knowing the modularity, and at the same time tells us why (6.79) fails for $k = 2$ or k odd and what replaces it in those cases.

We first note that

$$g_k(q) = \sum_{n=1}^{\infty} \Big(\sum_{m|n} m^{k-1} \Big) q^n = \sum_{m=1}^{\infty} m^{k-1} \big(q^m + q^{2m} + \cdots \big) = \sum_{m=1}^{\infty} m^{k-1}\, \frac{q^m}{1-q^m}$$

and hence that g_k can be written after a change of variables in the form

$$g_k\big(\mathrm{e}^{-x}\big) = \frac{1}{x^{k-1}} \sum_{m=1}^{\infty} f_k(mx)\,, \qquad f_k(x) = \frac{x^{k-1}}{\mathrm{e}^x - 1}\,.$$

[22] For the theory of modular forms, we refer to Sect. 6.2.

The function $f = f_k$ satisfies the hypotheses of Prop. 6.5 with integral

$$I_f = \int_0^\infty \frac{x^{k-1}}{\mathrm{e}^x - 1}\, dx = \int_0^\infty x^{k-1}\big(\mathrm{e}^{-x} + \mathrm{e}^{-2x} + \cdots\big)\, dx = (k-1)!\,\zeta(k)$$

and with Taylor expansion $f(x) = \sum\limits_{r=0}^{\infty} \frac{B_r}{r!} x^{r+k-2}$ at zero. Prop. 6.5 therefore gives

$$x^{k-1} g_k\big(\mathrm{e}^{-x}\big) \;\sim\; \frac{(k-1)!\,\zeta(k)}{x^k} + \sum_{r=0}^{\infty} (-1)^{r+k}\, \frac{B_r}{r!}\, \frac{B_{r+k-1}}{r+k-1}\, x^{r-1} \tag{6.80}$$

as $x \to 0$. If k is even and ≥ 4, then all products $B_r B_{r+k-1}$ with $r \neq 1$ vanish, since r and $r+k-1$ have opposite parity and all odd-index Bernoulli numbers except B_1 are zero. Therefore only two terms of (6.80) survive, and replacing x by $2\pi x$ and using the well-known formula for $\zeta(k)$ in terms of B_k, we recover (6.79). If $k = 2$, then the argument is the same except that now the $r = 0$ term also gives a non-zero contribution, so that we find

$$g_2\big(\mathrm{e}^{-2\pi x}\big) = \frac{1}{24x^2} \;-\; \frac{1}{4\pi x} + \frac{1}{24} + O(x^N) \qquad\qquad (N = 1, 2, \ldots),$$

instead of (6.79), in accordance with the known near-modularity property of g_2. Finally, if k is odd then we still get an explicit asymptotic formula with rational coefficients, but it now no longer terminates. Thus for $k = 3$ we find the expansion

$$g_3\big(\mathrm{e}^{-x}\big) \;\sim\; \frac{2\,\zeta(3)}{x^3} \;-\; \frac{1}{12x} + \frac{x}{1440} + \frac{x^3}{181440} + \frac{x^5}{7257600} + \frac{x^7}{159667200} + \cdots$$

as $x \to 0$, even though g_k has no modularity properties in this case. We leave it as an exercise to the reader to calculate the corresponding expansion when $k = 1$, where one has to use equation (6.77) instead of equation (6.74).

Example 4. As our final example, consider the function

$$P(q) = \prod_{m=1}^{\infty} \frac{1}{1-q^m} \qquad\qquad (|q| < 1)\,,$$

the generating power series of the partition function. To study the behavior of the partition function, we need to know how $P(q)$ blows up as q approaches 1 from below (or, more generally, any root of unity from within the unit circle). Here again the known modularity properties of the function $P(q)$ imply the non-trivial functional equation

$$\mathrm{e}^{\pi x/12}\, P\big(\mathrm{e}^{-2\pi x}\big) = \sqrt{x}\, \mathrm{e}^{\pi/12x}\, P\big(\mathrm{e}^{-2\pi/x}\big) \qquad\qquad (x > 0),$$

from which one immediately obtains the asymptotic expansion

$$\ln P\big(\mathrm{e}^{-2\pi x}\big) = \frac{\pi}{12x} + \frac{1}{2}\ln x \;-\; \frac{\pi}{12}\, x + O(x^N) \qquad\qquad (N = 1, 2, \ldots)\,. \tag{6.81}$$

To obtain this formula without knowing anything about the modularity of P, we observe that $\ln P\big(\mathrm{e}^{-x}\big)$ has the form $\sum_{m=1}^{\infty} f(mx)$ with $f(x) = -\ln(1-\mathrm{e}^{-x})$. This function is small at infinity, has integral $I_f = \zeta(2)$ (as one sees by integrating by parts once and then calculating as in Example 2), and has an asymptotic development

$$f(x) \;\sim\; \ln\frac{1}{x} \;-\; \sum_{n=1}^{\infty} \frac{B_n}{n\cdot n!}\, x^n \,,$$

as one sees easily by differentiating once. Hence equation (6.79) applies and gives

$$\ln P\bigl(\mathrm{e}^{-x}\bigr) \;\sim\; \frac{\zeta(2)}{x} + \frac{1}{2}\ln\frac{x}{2\pi} \;-\; \sum_{n=1}^{\infty}(-1)^n\, \frac{B_n}{n\cdot n!}\,\frac{B_{n+1}}{n+1}\, x^n \,.$$

Again all terms except for the first in the sum on the right vanish because n and $n+1$ have opposite parity, and replacing x by $2\pi x$ and using $\zeta(2) = \pi^2/6$ we recover (6.81). The same method using (6.76) with rational values of a lets us compute the exact asymptotics of $P(q)$ as q approaches any root of unity, recovering precisely the same result as that given by the modularity. Moreover, just as in the case of Example 2 for k odd, the method applies even when modularity fails. For example, if we define

$$P_2(q) = \prod_{m=1}^{\infty} \frac{1}{(1-q^m)^m} \qquad (|q|<1)\,,$$

a generating function that occurs in connection with the theory of plane partitions, then an analysis like the one just given for $P(q)$, but now with $f(x) = -x\ln(1-\mathrm{e}^{-x})$, produces the complete asymptotic expansion

$$P_2(\mathrm{e}^{-x}) = c\, x^{1/12}\, \mathrm{e}^{\zeta(3)/x^2} \left(1 \;-\; \frac{x^2}{2880} \;-\; \frac{17x^4}{12902400} \;-\; \cdots\right) \qquad (x\to 0)$$

with $c = \mathrm{e}^{\zeta'(-1)} = 0.847536694\cdots$, and using (6.78) one can get the expansion of $P_2(\alpha\mathrm{e}^{-x})$ for any root of unity α. Furthermore, for reasons similar to those which applied to g_k with k even, one finds that the corresponding expansions for the logarithm of $P_3(q) = \prod(1-q^n)^{-n^2}$ when q tends to a root of unity are terminating, even though there are no modularity properties in this case.

7. Rigorous Finite-Dimensional Magic Formulas of Quantum Field Theory

> Everything should be made as simple as possible, but not simpler.
> Albert Einstein (1879–1955)

> The important things in the world appear as invariants... The things we are immediately aware of are the relations of these invariants to a certain frame of reference... The growth of the use of transformation theory, as applied first to relativity and later to the quantum theory, is the essence of the new method in theoretical physics.
> Paul Dirac, 1930[1]
> St. John's College, Cambridge, England

This chapter is completely elementary, but it is very important for understanding both the basic ideas behind quantum field theory and the language used by physicists. Mathematicians should note that we introduce two crucial tools which are not mentioned in the standard literature on finite-dimensional linear algebra, namely,

- the Dirac calculus and
- discrete path integrals (functional integrals).

These tools are also very useful for mathematics itself.

7.1 Geometrization of Physics

In his 1915 theory of general relativity, Einstein described observers by local coordinate systems. However, since physics has to be independent of the choice of observers, physical quantities like the gravitational force have to be described by geometric objects which do *not* depend on the choice of the observer. In the late 1920s, Dirac tried to translate this general philosophy to quantum mechanics. To this end, he invented his transformation theory in the setting of Hilbert spaces. More precisely, as we will show in Sect. 12.2, one needs the concept of a rigged Hilbert space or Gelfand triplet. This is intimately related to Laurent Schwartz's theory of distributions which generalizes the Newton–Leibniz calculus. Summarizing, modern theoretical physics is based on the following concept:

[1] P. Dirac, The Principles of Quantum Mechanics, Clarendon Press, Oxford, 1930.

physical quantities (observables) and states
$\Rightarrow$ invariant geometric objects;
observers $\Rightarrow$ coordinate systems.

This underlines the importance of geometry for modern physics and the usefulness of abstract mathematical notions for describing physical quantities. In quantum physics, one has to use the geometry of Hilbert spaces.

7.2 Ariadne's Thread in Quantum Field Theory

Quantum field theory is based on the following tools:

- Fourier series (7.17) and Dirac calculus (7.20);
- the Fourier representation of the Green's operator (7.28);
- the Laplace transform of the Green's operator (7.30);
- the Dyson series for the Feynman propagator (7.51).

The magic formulas of quantum field theory. The following three magic formulas lie at the heart of quantum field theory:

(i) the magic Dyson perturbation formula for the S-matrix (Sect. 7.18);
(ii) the magic Feynman propagator formula and the Feynman kernel in terms of discrete path integrals (Sect. 7.21.1);
(iii) the magic Gell-Mann–Low formula for perturbed causal correlation functions (Sect. 7.22.2).

Furthermore, the magic Gell-Mann–Low formula is closely related to the magic Feynman formula for time-ordered products (Sect. 7.21.2) and the Wick rotation trick for vacuum expectation values (Sect. 7.22.1).

Basic strategy. These formulas are used by physicists in a formal manner for infinite-dimensional systems. There is a lack of rigorous justification. Our strategy is the following one:

(F) Rigorous finite-dimensional approach: We prove the magic formulas rigorously in finite-dimensional Hilbert spaces. In particular, we introduce a rigorous discrete Feynman path integral. This will be done in the present chapter.
(I) Formal infinite-dimensional approach: We translate straightforwardly the rigorous finite-dimensional formulas to infinite dimensions in a formal manner. This will be done in
  - Chapter 14 (response approach) and
  - Chapter 15 (operator approach).

This way, a mathematician should learn quickly the background of the formulas used by physicists in quantum field theory. The most important quantities read as follows:

- transition amplitude $\langle\varphi|A\psi\rangle$ for the observable A;
- transition probability $|\langle\varphi|A\psi\rangle|^2$;
- transition amplitude for the propagator,

$$\langle\varphi|\mathrm{e}^{-\mathrm{i}Ht/\hbar}\varphi_0\rangle,$$

  describing the dynamics of the quantum system in terms of the crucial Hamiltonian H (energy operator);
- ground state expectation value $\langle 0|A|0\rangle$ (also called vacuum expectation value for the observable A);[2]
- causal correlation coefficient

$$\langle 0|A(t)B(s)|0\rangle$$

  with the time ordering $t > s$ (also called 2-point Green's function) describing the correlations between two observables A and B at different time points t and s;
- higher causal correlation coefficient

$$\langle 0|A_1(t_1)A_2(t_2)\cdots A_n(t_n)|0\rangle$$

  with the time ordering $t_1 > t_2 > \cdots > t_n$ (also called n-point Green's function);
- the singularities of the Green's operator which correspond to the rest energies (and hence the rest masses) of the bound states.

The advantage of Feynman's approach to quantum field theory is the fact that both the transition amplitudes and the causal correlation functions can be represented by functional integrals which only depend on the classical action that appears in the principle of critical action.

Resonances and renormalization. In quantum field theory, a crucial role is played by the methods of renormalization theory. In Sect. 8.1, we will discuss the basic ideas of renormalization.

> *Using a simple rigorous finite-dimensional model, we will show that the phenomenon of renormalization is related to resonances which can be treated rigorously in terms of bifurcation theory.*

The point is that in the resonance case, naive perturbation theory fails completely; it has to be replaced by a more sophisticated approach.

The challenge for mathematics. The reader should note that physicists successfully use the formal infinite-dimensional methods in order to predict experimental data with extremely high precision.

[2] Note that the ground state $|0\rangle$ refers either to the state $|E_1\rangle$ of least energy of the full system described by the perturbed Hamiltonian $H = H_0 + V$ or to the ground state $|E_1^0\rangle$ of the unperturbed (or free) Hamiltonian H_0. In the following, we will clearly distinguish between the two cases.

From the mathematical point of view, the main problem is to justify the passage from finite to infinite dimensions for quantities which can be measured in physical experiments.

The study of the appropriate limits represents an important mathematical task for the future.

7.3 Linear Spaces

Functional analysis uses linear spaces and equips them with additional structures. This way, algebra, analysis, and geometry are combined with each other. Physical systems with a finite (resp. infinite) number of degrees of freedom are described by finite-dimensional (resp. infinite-dimensional) spaces.

The real, N-dimensional, linear space $\mathbb{R}^N$. By definition, the space $\mathbb{R}^N$ consists of all the column matrices

$$x = \begin{pmatrix} x^1 \\ \vdots \\ x^N \end{pmatrix} \tag{7.1}$$

where $x^1, \ldots, x^N$ are real numbers.

The complex, N-dimensional, linear space $\mathbb{C}^N$. By definition, the space $\mathbb{C}^N$ consists of all the column matrices

$$\psi = \begin{pmatrix} \psi^1 \\ \vdots \\ \psi^N \end{pmatrix} \tag{7.2}$$

where $\psi^1, \ldots, \psi^N$ are complex numbers. In particular, the zero vector $\mathbf{0}$ is obtained by setting $\psi^j = 0$ for $j = 1, \ldots, N$. The space $\mathbb{C}^N$ is the prototype of a complex linear space. Let us discuss this. To begin with, for all $\varphi, \psi \in \mathbb{C}^N$ and all complex numbers α, β, we introduce the linear combination $\alpha\varphi + \beta\psi$ by setting

$$\alpha \begin{pmatrix} \varphi^1 \\ \vdots \\ \varphi^N \end{pmatrix} + \beta \begin{pmatrix} \psi^1 \\ \vdots \\ \psi^N \end{pmatrix} := \begin{pmatrix} \alpha\varphi^1 + \beta\psi^1 \\ \vdots \\ \alpha\varphi^N + \beta\psi^N \end{pmatrix}.$$

These linear combinations satisfy the following quite natural rules:

(L1) $(\varphi + \psi) + \chi = \varphi + (\psi + \chi)$ (associative law);
(L2) $\varphi + \psi = \psi + \varphi$ (commutative law);
(L3) $\varphi + \mathbf{0} = \varphi$ (zero element);

(L4) $(\alpha+\beta)\varphi = \alpha\varphi + \beta\varphi$ and $\alpha(\varphi+\psi) = \alpha\varphi + \alpha\psi$ (distributive laws);
(L5) $\alpha(\beta\varphi) = (\alpha\beta)\varphi$ (associative law);
(L6) $\alpha\varphi = \varphi$ if $\alpha = 1$ (normalization);
(L7) $\alpha\varphi = \mathbf{0}$ if $\alpha = 0$.

For $N = 2$, the two elements

$$e_1 := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad e_2 := \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

form a basis of the space $\mathbb{C}^2$. That is, for all $\psi \in \mathbb{C}^2$,

$$\psi = \psi^1 e_1 + \psi^2 e_2.$$

A linear operator $A : \mathbb{C}^2 \to \mathbb{C}^2$ is given by the matrix equation $A\psi = \varphi$ which reads explicitly as

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \psi^1 \\ \psi^2 \end{pmatrix} = \begin{pmatrix} \varphi^1 \\ \varphi^2 \end{pmatrix}$$

with complex entries. Equivalently, this stands for the system

$$\begin{aligned} a_{11}\psi^1 + a_{12}\psi^2 &= \varphi^1, \\ a_{21}\psi^1 + a_{22}\psi^2 &= \varphi^2. \end{aligned}$$

This can be generalized immediately to the spaces $\mathbb{C}^N$ for $N = 1, 2, 3, \ldots$

Complex linear space. A set X is called a complex linear space iff it contains a distinguished element $\mathbf{0}$ and for all $\varphi, \psi \in X$ and all complex numbers α, β, the linear combination

$$\alpha\varphi + \beta\psi$$

is defined in such a way that the rules (L1) through (L7) above are satisfied.[3]

A subset L of X is called a linear subspace of X iff $\varphi, \psi \in L$ implies $\alpha\varphi + \beta\psi \in L$ for all complex numbers α, β.

Linear hull. Let S be a subset of the complex linear space X. By definition, the linear hull, $\operatorname{span} S$, is the smallest linear subspace of X which contains the set S. Explicitly, the set $\operatorname{span} S$ consists of all possible finite linear combinations

$$\alpha_1\psi_1 + \ldots + \alpha_n\psi_n, \qquad n = 1, 2, \ldots$$

where $\psi_j \in S$ and $\alpha_j \in \mathbb{C}$ for all $j = 1, \ldots, n$. Analogously, we define the linear hull of a subset of a real linear space.

[3] To simplify notation, in what follows we will write 0 instead of $\mathbf{0}$. This cannot lead to any mistakes.

Morphisms and isomorphisms in modern mathematics. In this monograph, we will encounter many mathematical structures, for example, linear spaces, Hilbert spaces, groups, algebras, topological spaces, manifolds, and so on. Typical maps are called morphisms; they preserve the structure under consideration. Isomorphic structures are always characterized by isomorphisms. Let us explain this for linear spaces. Note that

> *The definitions of the standard notions "surjective, injective, and bijective" can be found in the Appendix on page 931.*

Linear morphisms. Linear spaces are characterized by the linear structure based on linear combinations. Linear morphisms preserve linear structure. Explicitly, let X and Y be complex linear spaces. By a linear morphism (or a linear operator) $A : X \to Y$ we understand a map which preserves linear combinations, that is,

$$\boxed{A(\alpha\varphi + \beta\psi) = \alpha A\varphi + \beta A\psi}$$

for all $\varphi, \psi \in X$ and all complex numbers α, β.

Isomorphic linear spaces. By a linear isomorphism, we understand a linear morphism $A : X \to Y$ which is bijective. Two complex linear spaces X and Y are said to be isomorphic iff there exists a linear isomorphism $A : X \to Y$.

Dimension. Let $N = 1, 2, \ldots$ By definition, the elements $\varphi_1, \ldots, \varphi_N$ form a basis of the complex linear space X iff each element ψ in X can be represented by the linear combination

$$\psi = c_1\varphi_1 + \ldots + c_N\varphi_N$$

in such a way that the complex numbers $c_1, \ldots, c_N$ are uniquely determined by ψ. The complex numbers $c_1, \ldots, c_N$ are called the coordinates of ψ with respect to the basis $\varphi_1, \ldots, \varphi_N$. The main theorem on linear spaces reads as follows.

Theorem 7.1 *If a complex linear space has a finite basis, then every basis has the same number of elements.*

The number of basis elements is called the dimension of the linear space X; the dimension is denoted by $\dim X$. For the trivial linear space $\{0\}$ which only contains the zero element, we define $\dim\{0\} = 0$. By an infinite-dimensional linear space X, we understand a nontrivial linear space which does not have any finite basis. We write $\dim X = \infty$.

The proof of Theorem 7.1 can be found in all textbooks on linear algebra. We recommend A. Kostrikin and Yu. Manin, Linear Algebra and Geometry, Springer, New York, 1989. This book emphasizes the relations to modern physics.

Each complex linear space X of finite dimension N is linearly isomorphic to the space $\mathbb{C}^N$.

To prove this, choose a fixed basis $\varphi_1, \ldots, \varphi_N$ in X. For each element φ of X, we have

$$\varphi = c_1\varphi_1 + \ldots + c_N\varphi_N$$

where the complex numbers $c_1, \ldots, c_N$ are uniquely determined by φ. The map

$$\varphi \mapsto (c_1, \ldots, c_N)$$

is the desired linear isomorphism from the space X onto the space $\mathbb{C}^N$. This shows that finite-dimensional complex linear spaces possess only one essential invariant, namely, the dimension.

Two finite-dimensional complex linear spaces are isomorphic iff they have the same dimension.

Examples. The space $C^\infty(\mathbb{R})$ of smooth functions $\psi : \mathbb{R} \to \mathbb{C}$ forms a complex linear space. For fixed $n = 0, 1, 2, \ldots$, the space P_n of all complex polynomials

$$p(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n, \qquad x \in \mathbb{R}$$

with complex coefficients $a_0, a_1, \ldots, a_n$ represents an n-dimensional linear subspace of $C^\infty(\mathbb{R})$. Set $p_j(x) := x^j$. Then, the polynomials $p_0, p_1, \ldots, p_n$ form a basis of P_n. Let $-\infty < a < b < \infty$. Define

$$(A\psi)(x) := \frac{d\psi(x)}{dx}, \qquad B\psi := \int_a^b \psi(x)dx, \qquad (S\psi)(x) := \psi(x)^2$$

for all $x \in \mathbb{R}$ and all $\psi \in C^\infty(\mathbb{R})$. Then, the operators

$$A : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R}), \qquad B : C^\infty(\mathbb{R}) \to \mathbb{C}$$

are linear operators, whereas the operator $S : C^\infty(\mathbb{R}) \to \mathbb{C}^\infty(\mathbb{R})$ is *not* linear.

Nonlinear operators describe interactions in physics.

Therefore, nonlinear operators play a fundamental role in physics.

Real linear space. If we only consider real linear combinations $\alpha\varphi + \beta\psi$ where α and β are real numbers, then we get the notion of a linear real space. All of the notions and results discussed for complex linear spaces above remain valid for real linear spaces if we restrict ourselves to real coefficients. For example, the space $\mathbb{R}^N$ is a real, N-dimensional space. Moreover, every real N- dimensional linear space is linear isomorphic to $\mathbb{R}^N$.

Terminology. In what follows, the symbol $\mathbb{K}$ denotes either $\mathbb{R}$ or $\mathbb{C}$. By a linear space over $\mathbb{K}$ we understand a real (resp. complex) linear space iff $\mathbb{K} = \mathbb{R}$ (resp. $\mathbb{K} = \mathbb{C}$).

The linear space $L(X,Y)$. The symbol $L(X,Y)$ denotes the set of all linear operators

$$A : X \to Y$$

where X and Y are linear spaces over $\mathbb{K}$. Let $A, B \in L(X,Y)$ and $\alpha, \beta \in \mathbb{K}$. The linear combination $\alpha A + \beta B$ is defined by setting

$$(\alpha A + \beta B)\psi := \alpha A\psi + \beta B\psi \qquad \text{for all} \quad \varphi, \psi \in X.$$

This way, the set $L(X,Y)$ becomes a linear space over $\mathbb{K}$. We now choose $Y = X$.

The algebra $L(X,X)$. For two linear operators $A, B : X \to X$ define the product AB by setting

$$(AB)\psi := A(B\psi) \qquad \text{for all} \quad \psi \in X.$$

The set $L(X,X)$ becomes an algebra. For all $A, B, C \in L(X,X)$ and all $\alpha, \beta \in \mathbb{K}$, this means the following.

- $L(X,Y)$ is a linear space over $\mathbb{K}$.
- $(\alpha A + \beta B)C = \alpha AC + \beta BC$ and $C(\alpha A + \beta B) = \alpha CA + \beta CB$ (distributivity).
- $(AB)C = A(BC)$ (associativity).

The following simple criterion is useful.

Proposition 7.2 *The linear operator $A : X \to X$ is bijective iff there exist operators $B, C \in L(X,X)$ such that $AB = CA = I$. Then, $A^{-1} = B = C$.*

Proof. It follows from $AB\varphi = \varphi$ that each element φ of X lies in the image of A. Hence A is surjective (see page 931).

If $A\varphi = A\psi$, then $\varphi = CA\varphi = CA\psi = \psi$. Therefore, A is injective. Consequently, the operator A is bijective, and hence the inverse operator $A^{-1} : X \to X$ exists with $AA^{-1} = A^{-1}A = I$. Hence $A^{-1} = A^{-1}AB = B$ and $A^{-1} = CAA^{-1} = C$. □

As we will show in Volume III on quantum mathematics, operator algebras play a fundamental role in the algebraic approach to quantum physics (e.g., algebraic quantum field theory). As an introduction to this, we recommend R. Haag, Local Quantum Physics: Fields, Particles, Algebras, Springer, 1996, and H. Araki, Mathematical Theory of Quantum Fields, Oxford University Press, 1999.

Linear functionals and the dual space X^d. Let X be a linear space over $\mathbb{K}$. By a linear functional on X, we understand a linear map

$$F : X \to \mathbb{K}.$$

The set $L(X,\mathbb{K})$ of all linear functionals on X forms a linear space over X which is called the dual space to the linear X. This space is also denoted by the symbol X^d.

For example, the most general linear functional $F : \mathbb{C}^2 \to \mathbb{C}$ has the form

$$F(\psi) = c_1\psi^1 + c_2\psi^2 \qquad \text{for all} \quad \psi \in \mathbb{C}^2$$

where c_1, c_2 are fixed complex numbers. Setting

$$F(\psi) := \int_0^1 x\psi(x)dx \qquad \text{for all} \quad \psi \in C^\infty(\mathbb{R}),$$

we get a linear functional $F : \mathbb{C}^\infty(\mathbb{R}) \to \mathbb{C}$ on the function space $C^\infty(\mathbb{R})$.

Multilinear functionals. Again let X be a linear spaces over $\mathbb{K}$. We speak of a bilinear map

$$B : X \times X \to \mathbb{K}$$

iff this map is linear in each argument. Explicitly, for all $\varphi, \psi, \chi \in X$ and all $\alpha, \beta \in \mathbb{K}$, we have

$$B(\alpha\varphi + \beta\psi, \chi) = \alpha B(\varphi, \chi) + \beta B(\psi, \chi),$$
$$B(\chi, \alpha\varphi + \beta\psi) = \alpha B(\chi, \varphi) + \beta B(\chi, \psi).$$

Furthermore, B is called symmetric (resp. antisymmetric) iff for all $\varphi, \psi \in X$,

$$B(\varphi, \psi) = B(\psi, \varphi)$$

(resp. $B(\varphi, \psi) = -B(\psi, \varphi)$). If B is antisymmetric, then $B(\varphi, \varphi) = 0$. Similarly, the map

$$(\varphi_1, \ldots, \varphi_m) \mapsto M(\varphi_1, \ldots, \varphi_m)$$

from $X \times \cdots \times X$ (m factors) to $\mathbb{K}$ is called m-linear iff it is linear with respect to each argument. Moreover, M is called symmetric (resp. antisymmetric) iff it remains unchanged (resp. changes sign) under a transposition of two arguments.

The determinant of a linear operator. Let X be an m-dimensional linear space over $\mathbb{K}$ with $m = 1, 2, \ldots$. Let $A, B : X \to X$ be linear operators. One can show that there exists a uniquely determined number $\det(A)$ in $\mathbb{K}$ such that

$$\boxed{M(A\varphi_1, \ldots, A\varphi_m) = \det(A) \cdot M(\varphi_1, \ldots, \varphi_m)}$$

for all m-linear antisymmetric functionals M on X and all $\varphi_1, \ldots, \varphi_m \in X$. The so-called determinant has the following properties.

(i) $\det(AB) = \det(A)\det(B)$,
(ii) $\det(I) = 1$ (identity operator),
(iii) $\det(A) \neq 0$ iff A is a linear isomorphism.
(iv) If $\det(A) \neq 0$, then $\det A^{-1} = (\det A)^{-1}$.

For example, if the linear operator $A : \mathbb{C}^2 \to \mathbb{C}^2$ is given by the matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \text{then} \quad \det(A) = ad - bc.$$

In fact, if we choose the basis e_1, e_2 of $\mathbb{C}^2$ introduced above, then

$$Ae_1 = ae_1 + ce_2, \quad Ae_2 = be_1 + de_2.$$

For an antisymmetric bilinear functional B, we get $B(e_1, e_1) = B(e_2, e_2) = 0$ and $B(e_1, e_2) = -B(e_2, e_1)$. Hence

$$B(Ae_1, Ae_2) = (ad - bc)B(e_1, e_2) = \det(A) \cdot B(e_1, e_2).$$

The determinant was introduced by Leibniz (1646–1716). The general formula can be found in (7.3) below.

Matrix elements as coordinates of a linear operator. Consider an arbitrary linear operator $A : X \to X$ on the N-dimensional linear space X over $\mathbb{K}$ where $\mathbb{K} = \mathbb{R}, \mathbb{C}$. Let $\varphi, \psi \in X$. Choose any basis $e_1, \dots, e_N$ of X. The element Ae_k is then a linear combination of the basis vectors $e_1, \dots, e_N$. Explicitly,

$$\boxed{Ae_k = \sum_{k=1}^{N} a_{jk} e_j, \qquad k = 1, \dots, N.}$$

Consequently, it follows from $\psi = \sum_{k=1}^{N} \psi^k e_k$ that

$$A\psi = \sum_{j,k=1}^{N} (a_{jk}\psi^k) e_j.$$

This way, we assign the matrix $\mathcal{A} := (a_{jk})$ to the operator A. The matrix elements a_{kj} are also called the coordinates of the operator A with respect to the basis $e_1, \dots, e_N$. If we set $\varphi = A\psi$, then this corresponds to the equation

$$\boxed{\varphi^j = \sum_{k=1}^{N} a_{jk}\psi^k, \qquad j = 1, \dots, N}$$

for the corresponding coordinates of φ, ψ, and A.

In physics, the choice of a basis corresponds to the choice of an observer who measures coordinates.

Matrix calculus. There exists a perfect correspondence between operator algebra and matrix algebra. To explain this, let $A, B : X \to X$ be two linear operators. Moreover, let $\mathcal{A} := (a_{ij})$ and $\mathcal{B} := (b_{ij})$ denote the matrices corresponding to A and B with respect to a fixed basis $e_1, \dots, e_N$ of X. Then the following are met.

(i) Linear combination: The matrix $\mathcal{C} := (c_{ij})$ corresponding to the linear combination $C := \alpha A + \beta B$ is given by

$$c_{ij} = \alpha a_{ij} + \beta b_{ij}, \qquad i, j = 1, \dots N$$

for all numbers $\alpha, \beta \in \mathbb{K}$. The sum of matrices is defined in such a way that the operator equation $C = \alpha A + \beta B$ passes over to the matrix equation $\mathcal{C} = \alpha\mathcal{A} + \beta\mathcal{B}$.

(ii) Operator product: The matrix $\mathcal{P} := (p_{ij})$ corresponding to $P := AB$ reads as

$$\boxed{p_{ij} = \sum_{s=1}^{N} a_{is} b_{sj}, \qquad i, j = N.}$$

The product of matrices is defined in such a way that the operator equation $P = AB$ passes over to the matrix equation $\mathcal{P} = \mathcal{A}\mathcal{B}$.

(iii) Determinant: We have $\det A = \det \mathcal{A}$, that is, the determinant $\det A$ of the operator A is equal to the determinant of the matrix $\mathcal{A} := (a_{ij})$ given by

$$\boxed{\det \mathcal{A} := \sum_{j_1, \dots, j_N = 1}^{N} a_{1j_1} a_{2j_2} \cdots a_{Nj_N} \varepsilon_{j_1 \dots j_N}} \tag{7.3}$$

where $\varepsilon_{12\dots N} := 1$, and this symbol is antisymmetric with respect to the indices. For example, $\varepsilon_{12} = -\varepsilon_{21} = -1$ and $\varepsilon_{11} = \varepsilon_{22} = 0$.

The advantage of the definition of the determinant $\det(A)$ of the linear operator A, in terms of multi-linear functionals above, is that it does not depend on the choice of any basis of the space X.

The determinant is an invariant of linear operators on finite-dimensional linear spaces.

7.4 Finite-Dimensional Hilbert Spaces

For quantum physics, it is crucial to use linear spaces which are equipped additionally with an inner product. This leads us to the notion of Hilbert space. In 1906 Hilbert (1862–1943) introduced the space l_2 in order to generalize the classical principal axis transformation for finite-dimensional quadratic forms to infinite dimensions. By definition, the elements of l_2 are infinite sequences $(\psi^1, \psi^2, \dots)$ of complex numbers $\psi^1, \psi^2, \dots$ such that $\sum_{j=1}^{\infty} |\psi^j|^2 < \infty$. The space l_2 becomes a complex infinite-dimensional Hilbert space equipped with the inner product

$$\langle \varphi | \psi \rangle = \sum_{j=1}^{\infty} (\varphi^j)^\dagger \psi^j.$$

This inner product converges for all $\varphi, \psi \in l_2$. Motivated by the needs of quantum mechanics, the young John von Neumann (1903–1957) introduced the notion of an abstract Hilbert space in 1929.[4] At this time, von Neumann was Hilbert's assistant in Göttingen. Hilbert's 1906 paper founded the spectral theory for bounded self-adjoint operators. Von Neumann generalized this to unbounded self-adjoint operators. Such operators are typical for quantum mechanics. In particular, von Neumann found out that in infinite-dimensional Hilbert spaces, the domain of definition of an operator plays a crucial role.

> *Since two different operators may possess the same matrix elements, infinite-dimensional matrices are not sufficient for the mathematics of quantum physics.*

One has to use John von Neumann's abstract language of operators.

The N-dimensional real linear Hilbert space $\mathbb{R}^N$. Using the Euclidean inner product

$$\langle x|y\rangle := \sum_{j=1}^{N} x^j y^j, \qquad x, y \in \mathbb{R}^N,$$

the space $\mathbb{R}^N$ becomes an N-dimensional real Hilbert space.

The N-dimensional complex Hilbert space $\mathbb{C}^N$. The space $\mathbb{C}^N$ becomes a complex N-dimensional Hilbert space if we introduce the following inner product

$$\boxed{\langle \varphi|\psi\rangle := \sum_{j=1}^{N} (\varphi^j)^\dagger \psi^j.} \tag{7.4}$$

Here, the symbol $z^\dagger$ denotes the conjugate complex number to z.[5] We also introduce the norm of the element ψ given by

$$||\psi|| := \sqrt{\langle \psi|\psi\rangle} = \left(\sum_{j=1}^{N} |\psi^j|^2 \right)^{1/2}. \tag{7.5}$$

[4] J. von Neumann, General spectral theory for Hermitean operators, Math. Annalen **102** (1929), 49–131 (in German).

[5] Recall that $(\alpha+\beta\mathrm{i})^\dagger = \alpha-\beta\mathrm{i}$ for real numbers α and β. As a rule, mathematicians use the modified inner product

$$\langle \varphi|\psi\rangle = \sum_{j=1}^{N} \varphi^j (\psi^j)^\dagger.$$

However, as we will see below, this convention does *not* fit the elegant Dirac calculus which is very useful for mathematics as well. Therefore, we adopt the convention (7.4) used by physicists.

The elements

$$e_1 := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad e_2 := \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

form an orthonormal basis of the Hilbert space $\mathbb{C}^2$, that is,

$$\langle e_j | e_k \rangle = \delta_{jk}, \qquad j, k = 1, 2.$$

Recall that $\delta_{jk} := 1$ if $j = k$ and $\delta_{jk} = 0$ if $j \neq k$. Traditionally, δ_{jk} is called theKronecker symbol. For

$$\psi = \begin{pmatrix} \psi^1 \\ \psi^2 \end{pmatrix},$$

we get $\langle e_1 | \psi \rangle = \psi^1$ and $\langle e_2 | \psi \rangle = \psi^2$. Hence

$$\boxed{\psi = \langle e_1 | \psi \rangle e_1 + \langle e_2 | \psi \rangle e_2.}$$

This is the prototype of a Fourier series. For all $\varphi, \psi, \chi \in \mathbb{C}^N$ and all complex numbers α, β, the inner product satisfies the following conditions:

(P1) $\langle \varphi | \psi \rangle$ is a complex number,
(P2) $\langle \varphi | \psi \rangle^\dagger = \langle \psi | \varphi \rangle$ (antiduality),
(P3) $\langle \psi | \psi \rangle > 0$ if $\psi \neq 0$, and $\langle \psi | \psi \rangle = 0$ if $\psi = 0$ (definiteness),
(P4) $\langle \chi | \alpha\varphi + \beta\psi \rangle = \alpha \langle \chi | \varphi \rangle + \beta \langle \chi | \psi \rangle$ (linearity),
(P5) $\langle \alpha\varphi + \beta\psi | \chi \rangle = \alpha^\dagger \langle \varphi | \chi \rangle + \beta^\dagger \langle \psi | \chi \rangle$ (antilinearity).

Pre-Hilbert space. By a complex pre-Hilbert X, we understand a complex linear space equipped with an inner product which satisfies the conditions (P1) through (P5) for all $\varphi, \psi, \chi \in X$ and all complex numbers α, β. The definition of the norm

$$||\psi|| := \sqrt{\langle \psi | \psi \rangle} \tag{7.6}$$

allows us to introduce the notion of convergence. Let $\psi_n, \psi \in X$ for the indices $n = 1, 2, \ldots$ We write

$$\boxed{\lim_{n \to \infty} \psi_n = \psi} \tag{7.7}$$

iff the condition $\lim_{n \to \infty} ||\psi_n - \psi|| = 0$ is satisfied.

Hilbert space. A pre-Hilbert space is called a Hilbert space iff it is complete. This means the following. A sequence (ψ_n) in X is called a Cauchy sequence iff for each number $\varepsilon > 0$, there exists an index $n_0(\varepsilon)$ such that

$$||\psi_m - \psi_n|| < \varepsilon \qquad \text{for all} \quad m, n \geq n_0(\varepsilon).$$

By definition, the pre-Hilbert space X is called complete iff each Cauchy sequence converges to an element of X. Summarizing, a complex Hilbert space X is characterized by

- complex linear combinations $\alpha\varphi + \beta\psi$,
- an inner product $\langle\varphi|\psi\rangle$, and
- the validity of the completeness condition.

Hilbert space morphism. Let X and Y be complex Hilbert spaces. By a Hilbert space morphism (or an isometric operator) $U : X \to Y$, we understand a linear operator which preserves the inner product, that is,

$$\langle U\varphi|U\psi\rangle = \langle\varphi|\psi\rangle \qquad \text{for all} \quad \varphi, \psi \in X. \tag{7.8}$$

Isomorphic Hilbert spaces. Bijective Hilbert space morphisms are called Hilbert space isomorphisms. Explicitly, the operator $U : X \to Y$ is called a Hilbert space isomorphism (or a unitary operator) iff it is linear, bijective, and there holds relation (7.8). By definition, the Hilbert X is isomorphic (or unitarily equivalent) to the Hilbert space Y iff a Hilbert space isomorphism $U : X \to Y$ exists. Intuitively, isomorphic Hilbert spaces describe the same physics.

Finite-dimensional Hilbert spaces. It can be shown that every finite-dimensional pre-Hilbert space is complete, and hence it is a Hilbert space. This is not true anymore for infinite-dimensional pre-Hilbert spaces, as we will discuss in Sect. 10.2.

An N-dimensional complex (resp. real) Hilbert space is isomorphic to the Hilbert space $\mathbb{C}^N$ (resp. $\mathbb{R}^N$).

The Schwarz inequality. The most important inequality which is valid in a Hilbert space is the Schwarz inequality.[6] For example, we will show in Sect. 10.1 that this inequality implies Heisenberg's famous uncertainty principle in quantum physics.

Theorem 7.3 *For all elements φ and ψ of a pre-Hilbert space, we have the Schwarz inequality*

$$\boxed{|\langle\varphi|\psi\rangle| \leq ||\varphi|| \cdot ||\psi||.}$$

Proof. Let $\psi \neq 0$. For all complex numbers α, $0 \leq \langle\varphi - \alpha\psi|\varphi - \alpha\psi\rangle$. Hence

$$0 \leq \langle\varphi|\varphi\rangle - \alpha\langle\varphi|\psi\rangle - \alpha^\dagger\{\langle\psi|\varphi\rangle - \alpha\langle\psi|\psi\rangle\}.$$

Finally, choose $\alpha := \frac{\langle\psi|\varphi\rangle}{\langle\psi|\psi\rangle}$. □

The triangle inequality. For all elements φ, ψ, χ of a pre-Hilbert space,

$$\boxed{|\ ||\varphi|| - ||\psi||\ | \leq ||\varphi - \psi|| \leq ||\varphi|| + ||\psi||.} \tag{7.9}$$

[6] In the 19th century, classical variants of this inequality were discovered by several mathematicians including Victor Bunyakovski (1804–1889), Augustin Cauchy (1775–1857), and Amandus Schwarz (1843–1921). In the standard literature on functional analysis, one speaks usually of the Schwarz inequality.

Proof. Since $\langle\psi|\varphi\rangle = \langle\varphi|\psi\rangle^\dagger$, it follows from the Schwarz inequality that

$$\begin{aligned} ||\varphi+\psi||^2 &= \langle\varphi+\psi|\varphi+\psi\rangle = \langle\varphi|\varphi\rangle + \langle\varphi|\psi\rangle + \langle\varphi|\psi\rangle^\dagger + \langle\psi|\psi\rangle \\ &= ||\varphi||^2 + 2\Re(\langle\varphi|\psi\rangle) + ||\psi||^2 \\ &\le ||\varphi||^2 + 2||\varphi||\cdot||\psi|| + ||\psi||^2 = (||\varphi|| + ||\psi||)^2. \end{aligned}$$

Hence $||\varphi+\psi|| \le ||\varphi|| + ||\psi||$. This implies

$$||\varphi|| = ||\psi + (\varphi-\psi)|| \le ||\psi|| + ||\varphi-\psi||.$$

Thus, $||\varphi|| - ||\psi|| \le ||\varphi-\psi||$. Finally, interchanging φ with ψ, we obtain the inequality $\pm(||\varphi|| - ||\psi||) \le ||\varphi-\psi||$. □

Continuity of the inner product. Suppose that $\varphi_n \to \varphi$ and $\psi_n \to \psi$ as $n \to \infty$ in a pre-Hilbert space. Then

$$\boxed{\lim_{n\to\infty} \langle\varphi_n|\psi_n\rangle \to \langle\varphi|\psi\rangle.} \tag{7.10}$$

Proof. It follows from $|\,||\varphi_n|| - ||\varphi||\,| \le ||\varphi-\varphi_n|| \to 0$ as $n \to \infty$ that the sequence $(||\varphi_n||)$ converges to $||\varphi||$, and hence it is bounded. Similarly, the sequence $(||\psi_n||)$ is bounded. By the Schwarz inequality,

$$\begin{aligned} |\langle\varphi_n|\psi_n\rangle - \langle\varphi|\psi\rangle| &= |\langle\varphi_n-\varphi|\psi_n\rangle + \langle\varphi|\psi_n-\psi\rangle| \\ &\le |\langle\varphi_n-\varphi|\psi_n\rangle| + |\langle\varphi|\psi_n-\psi\rangle| \\ &\le ||\varphi_n-\varphi||\cdot||\psi_n|| + ||\varphi||\cdot||\psi_n-\psi|| \to 0 \ \text{ as } n\to\infty. \end{aligned}$$

This finishes the proof. □

The continuity of the inner product implies the continuity of the norm. Explicitly, it follows from $\varphi_n \to \varphi$ as $n \to \infty$ in a pre-Hilbert space that

$$\lim_{n\to\infty} ||\varphi_n|| = ||\varphi||.$$

The topology of a Hilbert space. A subset $\mathcal{O}$ of a Hilbert space X is called open iff for each point $\psi_0 \in X$, there exists a number $\varepsilon > 0$ such that the open ball

$$\{\psi \in X : ||\psi-\psi_0|| < \varepsilon\}$$

is contained in the set $\mathcal{O}$. It turns out that

Each Hilbert space is a topological space.

For a subset $\mathcal{C}$ of a Hilbert space X, the following two conditions are equivalent.

(i) The set $\mathcal{C}$ is closed, that is, the complement $X \setminus \mathcal{C}$ is open.
(ii) The set $\mathcal{C}$ is sequentially closed, that is, if $\psi_n \in \mathcal{C}$ for all n and $\psi_n \to \psi$ in X as $n \to \infty$, then $\psi \in \mathcal{C}$.

The same is true for pre-Hilbert spaces.

7.5 Groups

> Groups describe symmetries in physics, whereas the linearization of symmetries leads to Lie algebras.
>
> Folklore

Let X be an n-dimensional real or complex linear space with $n = 1, 2, \ldots$. The set $GL(X)$ of all linear isomorphisms $A : X \to X$ is the prototype of a group; it is called the general linear group of the space X. Set $\mathcal{G} := GL(X)$. Then, for all $A, B, C \in \mathcal{G}$, the following are met.

(G1) Consistency: $AB \in \mathcal{G}$.
(G2) Associativity: $(AB)C = A(BC)$.
(G3) Unit element: There exists a uniquely determined element I in $\mathcal{G}$ such that $IA = AI = A$ for all $A \in \mathcal{G}$.
(G4) For each $A \in \mathcal{G}$, there exists a uniquely determined element B in $\mathcal{G}$ such that $AB = BA = I$. We write A^{-1} instead of B.

A set $\mathcal{G}$ equipped with a uniquely defined product AB for all $A, B \in \mathcal{G}$ is called a group iff the conditions (G1) through (G4) are satisfied.

Group morphism. A map $\chi : \mathcal{G} \to \mathcal{H}$ between two groups $\mathcal{G}$ and $\mathcal{H}$ is called a group morphism iff it respects products, that is,

$$\chi(AB) = \chi(A)\chi(B) \qquad \text{for all} \quad A, B \in \mathcal{G}.$$

A bijective (resp. surjective) group morphism $\chi : \mathcal{G} \to \mathcal{H}$ is called group isomorphism (resp. group epimorphism).

Subgroup. A subset $\mathcal{S}$ of a group $\mathcal{G}$ is called a subgroup iff it is a group with respect to the product on $\mathcal{G}$. Again let X be a complex n-dimensional Hilbert space with $n = 1, 2, \ldots$ Then the following are true.

- Unitary group $U(X)$: The set $U(X)$ of all unitary operators $A : X \to X$ forms a subgroup of $GL(X)$.
- Special unitary group $SU(X)$: The set of all $A \in U(X)$ with $\det(A) = 1$ forms a subgroup of $U(X)$ denoted by $SU(X)$.

Matrix algebra. Let $A = (a_{ij})$ be a real or complex $(n \times n)$-matrix with $n = 1, 2, \ldots$.

- The dual matrix $A^d = (b_{ij})$ is defined by $b_{ij} := a_{ji}$ for $i, j = 1, \ldots n$.
- The adjoint matrix $A^\dagger = (c_{ij})$ is defined by $c_{ij} := a^\dagger_{ji}$ for $i, j = 1, \ldots n$.
- The conjugate complex matrix $A^c = (d_{ij})$ is defined by $d_{ij} := a^\dagger_{ij}$ for $i, j = 1, \ldots, n$.
- The number $\mathrm{tr}(A) := \sum_{j=1}^n a_{jj}$ is called the trace of the matrix A. The expression for the determinant $\det(A)$ can be found in (7.3) on page 335.

For all complex linear $(n \times n)$-matrices A, B and all complex numbers α, β, the following ar true:

- $(\alpha A + \beta B)^\dagger = \alpha^\dagger A^\dagger + \beta^\dagger B^\dagger$,
- $(AB)^\dagger = B^\dagger A^\dagger$ and $(AB)^d = B^d A^d$,
- $\det I = 1$ and $\det A^d = \det A$,
- $\det(AB) = \det A \cdot \det B$,
- $\det A^{-1} = (\det A)^{-1}$ if $\det A \neq 0$.
- $\operatorname{tr}(\alpha A + \beta B) = \alpha \operatorname{tr} A + \beta \operatorname{tr} B$,
- $\operatorname{tr} A^d = \operatorname{tr} A$ and $\operatorname{tr} A^\dagger = (\operatorname{tr} A)^\dagger$,
- $\operatorname{tr}(AB) = \operatorname{tr}(BA)$,
- $\operatorname{tr} I = n$.

The proofs of all the matrix rules can be found in Hein (1990).

Multiplicative matrix groups. The following groups with respect to multiplication AB of numbers A, B or matrices will be used quite often.

(i) The group $GL(1, \mathbb{R})$ consists of all nonzero real numbers. This group is also denoted by $\mathbb{R}^\times$.
(ii) The group $\mathbb{R}_+^\times$ consists of all positive numbers. This is a subgroup of $\mathbb{R}^\times$.
(iii) The group $U(1)$ consists of all complex numbers z with $|z| = 1$.
(iv) The real, general linear group $GL(n, \mathbb{R})$ consists of all invertible real $(n \times n)$-matrices.
(v) The real, special linear group $SL(n, \mathbb{R})$ consists of all matrices A in $GL(n, \mathbb{R})$ with $\det(A) = 1$.
(vi) The complex, general linear group $GL(n, \mathbb{C})$ consists of all invertible complex $(n \times n)$-matrices.
(vii) The complex, special linear group $SL(n, \mathbb{C})$ consists of all matrices A in $GL(n, \mathbb{C})$ with $\det(A) = 1$.
(viii) The unitary group $U(n)$ consists of all complex $(n \times n)$-matrices with $AA^\dagger = A^\dagger A = I$. Such matrices are called unitary matrices. This is a subgroup of $GL(n, \mathbb{C})$.
(ix) The special unitary group $SU(n)$ consists of all matrices $A \in U(n)$ with $\det A = 1$.
(x) The orthogonal group $O(n)$ consists of all real $(n \times n)$-matrices A with $AA^d = A^d A = I$. Such matrices are called orthogonal matrices. This is a subgroup of $GL(n, \mathbb{R})$.
(xi) The special orthogonal group $SO(n)$ consists of all matrices $A \in O(n)$ with $\det(A) = 1$. In particular, $O(1) = \{1, -1\}$ and $SO(1) = \{1\}$.

If X is a complex n-dimensional Hilbert space with $n = 1, 2, \ldots$, then we have the following group isomorphisms:

- $U(n) = U(X), SU(n) = SU(X), GL(n, \mathbb{C}) = GL(X), SL(n, \mathbb{C}) = SL(X)$.

Similarly, if X is a real n-dimensional Hilbert space, then we have the following group isomorphisms:

- $O(n) = U(X), SO(n) = SU(X), GL(n, \mathbb{R}) = GL(X), SL(n, \mathbb{R}) = SL(X)$.

A group is called commutative (or Abelian) iff $AB = BA$ for all $A, B \in \mathcal{G}$.

- The groups (i)–(iii) above are commutative.
- If $n = 1$, then the groups (iv)–(xi) are commutative.
- If $n \geq 2$, then the groups (iv)–(xi) above are not commutative.

Additive groups. The prototype of an additive group is the real line $\mathbb{R}$ equipped with addition. Set $\mathcal{A} := \mathbb{R}$. Then, for all $a, b, c \in \mathcal{A}$, the following are met.

(A1) Consistency: $a + b \in \mathcal{A}$.
(A2) Associativity: $(a + b) + c = a + (b + c)$.
(A3) Commutativity: $a + b = b + a$.
(A4) Zero element: There exists precisely one element 0 in the set $\mathcal{A}$ such that $a + 0 = a$ for all $a \in \mathcal{A}$.
(A5) Inverse element: For each $a \in \mathcal{A}$, there exists precisely one element b such that $a + b = 0$. We write $(-a)$ instead of b.

A set $\mathcal{A}$ is called an additive group iff it is equipped with a uniquely determined addition $a + b$ for all $a, b \in \mathcal{A}$ such that the conditions (A1) through (A5) are met. A map

$$\chi : \mathcal{A} \to \mathcal{B}$$

between two additive groups $\mathcal{A}$ and $\mathcal{B}$ is called an additive group morphism iff it respects addition, that is, $\chi(a + b) = \chi(a) + \chi(b)$ for all $a, b \in \mathcal{A}$.

7.6 Lie Algebras

Let X be a real linear space. The prototype of a Lie algebra is the set $gl(X)$ of all linear operators $A : X \to X$. Define the Lie product

$$[A, B] := AB - BA.$$

Set $\mathcal{L} := gl(X)$. Then, for all $A, B, C \in \mathcal{L}$ and all real numbers α, β the following are met.

(L1) Linearity: $\mathcal{L}$ is a real linear space.
(L2) Consistency: $[A, B] \in \mathcal{L}$.
(L3) Anticommutativity: $[B, A] = -[A, B]$.
(L4) Bilinearity: $[\alpha A + \beta B, C] = \alpha[A, C] + \beta[B, C]$.
(L5) Jacobi identity: $[[A, B], C] + [[B, C], A] + [[C, A], B] = 0$.

A set $\mathcal{L}$ is called a real Lie algebra iff it is equipped with a uniquely determined Lie product $[A, B]$ for all elements A, B of $\mathcal{L}$ such that conditions (L1) through (L5) are met.[7] The Lie product $[A, B]$ is also called the Lie bracket.

[7] Similarly, we speak of a complex Lie algebra $\mathcal{L}$ iff a natural modification of (L1) through (L5) holds true. Explicitly, we postulate that $\mathcal{L}$ is a complex linear space, and condition (L4) is valid for all complex numbers α, β.

Lie algebra morphism. A map $\chi : \mathcal{L} \to \mathcal{M}$ between two real (resp. complex) Lie algebras is called a Lie algebra morphism iff it is a linear morphism and it respects Lie products, that is,

$$\chi([A, B]) = [\chi(A), \chi(B)] \qquad \text{for all} \quad A, B \in \mathcal{L}.$$

Lie subalgebra. A subset $\mathcal{S}$ of a Lie algebra $\mathcal{L}$ is called a Lie subalgebra iff it is a Lie algebra with respect to the Lie product on $\mathcal{L}$. For example, let X be a complex n-dimensional Hilbert space with $n = 1, 2, \ldots$ Then the following are true.[8]

- Skew-adjoint operators: The set $u(X)$ of all linear operators $A : X \to X$ with $A^\dagger = -A$ forms a Lie subalgebra of $gl(X)$.
- Traceless skew-adjoint operators: The set of all $A \in u(X)$ with $\mathrm{tr}(A) = 0$ forms a Lie subalgebra of $u(X)$ denoted by $su(n)$.

Classification of morphisms. The following terminology is used in mathematics for all kind of morphisms:

- bijective morphisms are called isomorphisms if the inverse map is also a morphism;[9]

Furthermore, for linear morphisms (or group morphisms, Lie algebra morphisms), we use the following convention:[10]

- surjective morphisms are called epimorphisms;
- injective morphisms are called monomorphisms.

For example, two groups $\mathcal{G}$ and $\mathcal{H}$ are called isomorphic iff there exists a group isomorphism $\chi : \mathcal{G} \to \mathcal{H}$.

> *Isomorphic groups and isomorphic Lie algebras describe the same mathematics and physics.*

Matrix Lie algebras. The following real Lie algebras with respect to the Lie product $[A, B] := AB - BA$ are frequently used.[11]

(i) The Lie algebra $gl(n, \mathbb{R})$ consists of all real $(n \times n)$-matrices.
(ii) The Lie algebra $sl(n, \mathbb{R})$ consists of all $A \in gl(n, \mathbb{R})$ with $\mathrm{tr}(A) = 0$.
(iii) The Lie algebra $gl(n, \mathbb{C})$ consists of all complex $(n \times n)$-matrices.
(iv) The Lie algebra $sl(n, \mathbb{C})$ consists of all matrices $A \in gl(n, \mathbb{C})$ with vanishing trace, $\mathrm{tr}(A) = 0$.

[8] The definition of the adjoint operator $A^\dagger$ and the trace $\mathrm{tr}(A)$ of the linear operator A can be found on page 357 and 363, respectively.
[9] For linear isomorphisms, group isomorphisms, and Lie algebra isomorphisms, the latter condition is satisfied automatically.
[10] For general morphisms, the definition of epimorphisms and monomorphisms will be given in Volume II, Chap. 3, in the setting of category theory.
[11] The Lie algebras $gl(n, \mathbb{C}), sl(n, \mathbb{C})$ are also complex Lie algebras, but $u(n)$ and $su(n)$ are not complex Lie algebras.

(v) The Lie algebra $u(n)$ consists of all $A \in gl(n, \mathbb{C})$ with $A^\dagger = -A$. These matrices are called skew-adjoint.
(vi) The Lie algebra $su(n)$ consists of all complex matrices $A \in u(n)$ with $\mathrm{tr}(A) = 0$.
(vii) The Lie algebra $o(n)$ consists of all $A \in gl(n, \mathbb{R})$ with $A^d = -A$. These matrices are called skew-symmetric.
(viii) The Lie algebra $so(n)$ coincides with $o(n)$.

For example, in order to prove that $su(n)$ is a real Lie algebra, we have to show that the following are true.

(a) If $A, B \in su(n)$ and $\alpha, \beta \in \mathbb{R}$, then $\alpha A + \beta B \in su(n)$.
(b) If $A, B \in su(n)$, then $[A, B] \in su(n)$.

in fact, the claim (a) follows from

$$(\alpha A + \beta B)^\dagger = \alpha A^\dagger + \beta B^\dagger = -\alpha A - \beta B.$$

In order to prove (b), note that

$$[A, B]^\dagger = (AB - BA)^\dagger = B^\dagger A^\dagger - A^\dagger B^\dagger = BA - AB = -[A, B].$$

In addition, $\mathrm{tr}([A, B]) = \mathrm{tr}(AB - BA) = \mathrm{tr}(AB) - \mathrm{tr}(BA) = 0$.

If X is a complex n-dimensional Hilbert space with $n = 1, 2, \ldots$, then we have the following real Lie algebra isomorphisms:

$$u(n) = u(X), \quad su(n) = su(X), \quad gl(n, \mathbb{C}) = gl(X), \quad sl(n, \mathbb{C}) = sl(X).$$

Similarly, if X is a real n-dimensional Hilbert space, then we have the following real Lie algebra isomorphisms:

$$o(n) = so(n) = u(X), \quad gl(n, \mathbb{R}) = gl(X), \quad sl(n, \mathbb{R}) = sl(X).$$

The Hilbert space $gl(n, \mathbb{C})$. This space is a complex Hilbert space of dimension n^2 with respect to the inner product

$$\boxed{\langle A|B\rangle := \mathrm{tr}(A^\dagger B).}$$

For the norm,

$$||A|| := \langle A|A\rangle^{1/2} = (\mathrm{tr}(A^\dagger A))^{1/2} = \left(\sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}.$$

In particular, the convergence $A_k \to A$ as $k \to \infty$ in $gl(n, \mathbb{C})$ means that

$$\lim_{k \to \infty} ||A_k - A||^2 = \sum_{i,j=1}^{n} ||(a_{ij})_k - a_{ij}||^2 = 0.$$

This is equivalent to the convergence $\lim_{k \to \infty} (a_{ij})_k = a_{ij}$ of the corresponding matrix elements,

7.7 Lie's Logarithmic Trick for Matrix Groups

In 1614, John Napier (1550–1617) published his book "Mirifici logarithmorum canonis descriptio" where he popularized the use of logarithms. His idea was to simplify the computation of products by reducing this to the computation of sums. The basic rule reads as

$$\ln(ab) = \ln a + \ln b \qquad \text{for all} \quad a, b \in \mathbb{R}.$$

Johannes Kepler (1571–1630) used this new mathematical tool for simplifying substantially his huge computations in celestial mechanics.

Our goal is to simplify the products of matrices in a similar manner.

This approach was invented by Sophus Lie (1842–1899) in the 1870s and lies at the heart of the theory of Lie groups and Lie algebras.

The exponential function. For each matrix $A \in GL(n, \mathbb{C})$, we define

$$\boxed{\mathrm{e}^A := I + A + \frac{A^2}{2!} + \ldots = \sum_{k=0}^{\infty} \frac{A^k}{k!}.}$$

This infinite series converges in the Hilbert space $gl(n, \mathbb{C})$. Observe that for all matrices $A, B \in gl(n, \mathbb{C})$, the following useful formulas are valid.

- If $AB = BA$, then $\mathrm{e}^{A+B} = \mathrm{e}^A \mathrm{e}^B$.[12]
- $\det(\mathrm{e}^A) = \mathrm{e}^{\mathrm{tr}(A)}$.
- $\left(\mathrm{e}^A\right)^\dagger = \mathrm{e}^{A^\dagger}$ and $\left(\mathrm{e}^A\right)^d = \mathrm{e}^{A^d}$.
- $\left(\mathrm{e}^A\right)^{-1} = \mathrm{e}^{-A}$.
- $\mathrm{e}^{tA}\mathrm{e}^{sA} = \mathrm{e}^{(t+s)A}$ (causality relation). In other words, the set $\{\mathrm{e}^{tA} : t \in \mathbb{R}\}$ is a subgroup of $GL(n, \mathbb{C})$, and the map $t \mapsto \mathrm{e}^{tA}$ is a group morphism from the additive group $\mathbb{R}$ into the group $GL(n, \mathbb{C})$.
- Let $\psi_0 \in \mathbb{C}^n$. The function $\psi(t) := \mathrm{e}^{tA}\psi_0$ for all times $t \in \mathbb{R}$ is the unique solution of the differential equation

$$\frac{d\psi(t)}{dt} = A\psi(t), \quad t \in \mathbb{R}, \quad \psi(0) = \psi_0.$$

  This equation and its infinite-dimensional generalizations govern the dynamics of many physical systems (e.g., quantum systems). If we set $A := -\mathrm{i}H/\hbar$ where $H^\dagger = H$, then we get the Schrödinger equation

$$\boxed{\mathrm{i}\hbar \frac{d\psi(t)}{dt} = H\psi(t), \quad t \in \mathbb{R}, \quad \psi(0) = \psi_0}$$

  for the Hamiltonian H (energy operator).

[12] It is crucial for quantum physics that this formula is not valid, as a rule, if $AB \neq BA$. Then, one has to use the magic Baker–Campbell–Hausdorff formula (see page 508).

Finally, let us comment on the causality relation. To this end, introduce the time-dependent operator $\mathcal{T}_t\psi_0 := \psi(t) = \mathrm{e}^{-\mathrm{i}tH/\hbar}\psi_0$. Then the causality relation tells us that

$$\boxed{\mathcal{T}_{s+t}\psi_0 = \mathcal{T}_s(\mathcal{T}_t\psi_0)} \qquad \text{for all} \quad t, s \geq 0.$$

In terms of physics, this means that the following two procedures yield the same result.

(I) Start from the initial state ψ_0 and compute the state $\psi(s+t) = \mathcal{T}_{s+t}$ of the system at the later time $s+t$.
(II) Compute first the state at time $\psi(t) = \mathcal{T}_t\psi_0$. Use this as new initial state for computing state $\mathcal{T}_s\psi(t)$ of the system at the later time s.

Finally, it follows from

$$\psi_0 = \mathcal{T}_{-t}\psi(t), t \geq 0$$

that the state of the system at the later time t uniquely determines the initial state of the system at time $t = 0$.

Summarizing, the exponential function describes the causality of physical processes.

The logarithmic function. For all matrices $A \in gl(n, C)$ with $||A|| < 1$, we define

$$\boxed{\ln(I + A) := A - \frac{A^2}{2} + \ldots = \sum_{k=1}^{\infty}(-1)^{k+1}\frac{A^k}{k}.}$$

This series converges in the Hilbert space $gl(n, \mathbb{C})$. There exists a small number $\varepsilon > 0$ such that for all matrices $B, C \in gl(n, \mathbb{C})$ with $||B - I|| < \varepsilon$ and $||C - I|| < \varepsilon$, the following are true.

- $\ln(BC) = \ln B + \ln C$,
- $\ln I = 0$,
- $(\ln B)^\dagger = \ln B^\dagger$ and $(\ln B)^d = \ln B^d$,
- $\mathrm{tr}(B) = \ln(\det \mathrm{e}^B)$,
- $\mathrm{e}^{\ln A} = A$ if $||A - I|| < 1$,
- $\ln \mathrm{e}^A = A$ if $||\mathrm{e}^A - I|| < 1$.

The local parametrization of the group $GL(n, \mathbb{C})$. We are given the matrix $A_0 \in GL(n, \mathbb{C})$ and the number $\varepsilon > 0$. By an ε-neighborhood of A_0 in the group $GL(n, \mathbb{C})$, we understand the set

$$\mathcal{U}_\varepsilon(A_0) := \{A \in GL(n, \mathbb{C}) : ||A - A_0|| < \varepsilon\}$$

where the norm refers to the Hilbert space $gl(n, \mathbb{C})$. In addition, the subset $\mathcal{U}(A_0)$ of the group $GL(n, \mathbb{C})$ is called a neighborhood of A_0 iff it contains any ε-neighborhood $\mathcal{U}_\varepsilon(A_0)$. If we replace the group $GL(n, \mathbb{C})$ by the Lie algebra $gl(n, \mathbb{C})$, then we obtain the corresponding notions for $gl(n, \mathbb{C})$. The following

theorem tells us that the group $GL(n,\mathbb{C})$ can be locally parametrized by the elements of the Lie algebra $gl(n,\mathbb{C})$. This main trick of the theory of Lie groups simplifies the computations with respect to the group $GL(n,\mathbb{C})$.

Theorem 7.4 *The map $A \mapsto \ln A$ sends the neighborhood*

$$\{A \in GL(n,\mathbb{C}) : ||A - I|| < I\}$$

of the unit element I of the group $GL(n,\mathbb{C})$ bijectively onto some neighborhood of the origin in the Lie algebra $gl(n,\mathbb{C})$. The inverse map is given by $B \mapsto e^A$.

The matrix $B := \ln A$ is called the local coordinate of the group element A near the unit element. For an arbitrary group element A_0 in $GL(n,\mathbb{C})$, we write

$$A = A_0 \mathrm{e}^B \qquad \text{for all} \quad B \in gl(n,\mathbb{C}).$$

If the matrix A lies in a sufficiently small neighborhood of A_0, then the local coordinate of A is given by $B := \ln(A_0^{-1}A)$. Analogously, the groups

$$SL(n,\mathbb{C}), GL(n,\mathbb{R}), SL(n,\mathbb{R}), U(n), SU(n), O(n), SO(n)$$

can be locally parametrized by the matrices of the Lie algebras

$$sl(n,\mathbb{C}), gl(n,\mathbb{R}), sl(n,\mathbb{R}), u(n), su(n), o(n), so(n),$$

respectively. For example, consider the case of the group $SU(n)$. Choose a matrix $A \in SU(n)$ near the unit element. It follows from $AA^\dagger = I$ that

$$0 = \ln I = \ln A + \ln A^\dagger = \ln A + (\ln A)^\dagger. \tag{7.11}$$

In addition, $\det A = 1$ implies $\mathrm{tr}(\ln A) = \ln(\det A) = 0$. Hence the local coordinate $\ln A$ is a skew-adjoint and traceless matrix which lies in the Lie algebra $su(n)$.

7.8 Lie Groups

Lie groups. The notion of Lie group combines the notion of group with the notion of manifold. By definition, a finite-dimensional real manifold is called a Lie group iff it is equipped with a group structure such that both the multiplication map

$$(A, B) \mapsto AB$$

and the inversion map $A \mapsto A^{-1}$ are smooth. The following theorem is merely a reformulation of Theorem 7.4.

Theorem 7.5 *For $n = 1, 2, \ldots$, the classical groups*

$$GL(n,\mathbb{C}), SL(n,\mathbb{C}), GL(n,\mathbb{R}), SL(n,\mathbb{R}), U(n), SU(n), O(n), SO(n)$$

are Lie groups. The tangent space of these Lie groups at the unit element are the Lie algebras

$$gl(n,\mathbb{C}), sl(n,\mathbb{C}), gl(n,\mathbb{R}), sl(n,\mathbb{R}), u(n), su(n), o(n), so(n),$$

respectively. Here, $so(n) = o(n)$. The Lie groups $U(n), SU(n), O(n), SO(n)$ are compact subsets of $\mathbb{C}^{n^2}$.

Observe that the 2-dimensional sphere (e.g., the surface of earth) cannot be equipped with the structure of a Lie group. However, if we identify the points of the 3-dimensional unit sphere $\mathbb{S}^3$ with the unit quaternions, then the sphere $\mathbb{S}^3$ becomes a Lie group with respect to the multiplication of quaternions.

Lie's linearization principle. The tangent space of a Lie group $\mathcal{G}$ at the unit element can always be equipped with the structure of a Lie algebra which is denoted by $\mathcal{LG}$.

> *It is crucial for the theory of Lie groups that the Lie algebra $\mathcal{LG}$ knows all about the local structure of the Lie group $\mathcal{G}$.*

This is the famous Lie linearization principle for Lie groups which we will encounter quite often.

Lie group morphism. Let $\mathcal{G}$ and $\mathcal{H}$ be Lie groups. By a Lie group morphism

$$f : \mathcal{G} \to \mathcal{H}$$

we understand a map which is both a group morphism and a manifold morphism. In other words, this is a smooth group morphism. The map $f : \mathcal{G} \to \mathcal{H}$ is called a Lie group isomorphism iff it is both a group isomorphism and a manifold isomorphism. In other words, this is a group isomorphism which is a diffeomorphism, too.

The theory of manifolds is fundamental for modern physics, as we will see at many different places of the volumes of this treatise. In particular, the theory of Lie groups and Lie algebras (and their representations) is basic for the study of symmetry phenomena in physics and mathematics.

The physical meaning of the tangent space. If we consider the motion $x = x(t)$ of a particle on a manifold M such that the particle is at the point x_0 at time $t = 0$, then the velocity vector $\dot{x}(0)$ lies in the tangent space $T_{x_0}M$ of the manifold M at the point x_0. Generally, the space $T_{x_0}M$ consists of all possible velocity vectors at the point x_0. As an example, consider the group $SU(n)$ at the point I (unit element).

(a) We show that each velocity vector at the unit element lies in the Lie algebra $su(n)$. To this end, consider an arbitrary smooth curve

$$A = A(t), \qquad t \in \mathcal{U}(0)$$

on the manifold $SU(n)$ with $A(0) = I$. Here, $\mathcal{U}(0)$ is a neighborhood of the origin on the real line. Differentiation with respect to time t at the point $t = 0$ yields the velocity vector $\dot{A}(0)$. Set

$$B(t) := \ln A(t), \qquad t \in \mathcal{U}(0).$$

Hence $\dot{B}(0) = A(0)^{-1}\dot{A}(0) = \dot{A}(0)$. By (7.11), $B(t) \in su(n)$ for all times $t \in \mathcal{U}(0)$. This implies $\dot{B}(0) \in su(n)$, noting that $su(n)$ is a linear space.

(b) We show that each vector B in $su(n)$ is the velocity vector of a smooth curve on $SU(n)$ which passes through the point I. Explicitly, we choose the curve

$$A(t) := \mathrm{e}^{tB}, \qquad t \in \mathbb{R}.$$

Then $\dot{A}(0) = B$. It remains to show that $A(t) \in SU(n)$ for all $t \in \mathbb{R}$. In fact, it follows from $B + B^\dagger = 0$ and $\operatorname{tr} B = 0$ that

$$A(t)A(t)^\dagger = \mathrm{e}^{t(B+B^\dagger)} = I, \qquad \det A(t) = \mathrm{e}^{t \operatorname{tr} B} = 1.$$

Infinitesimal transformations. The elements B of the Lie algebra $su(n)$ are called the infinitesimal transformations of the Lie group $SU(n)$. This is motivated by the approximation

$$\mathrm{e}^{tB} = I + tB + O(t^2), \qquad t \to 0. \tag{7.12}$$

The use of infinitesimal transformations substantially simplifies the investigation of Lie groups. Finally, let us motivate the Lie bracket. Suppose that $B, C \in su(n)$. Then $\mathrm{e}^{\pm tB}, \mathrm{e}^{sC} \in SU(n)$ for all $t, s \in \mathbb{R}$. For fixed $s \in \mathbb{R}$, set

$$A(t) := \mathrm{e}^{sB}\mathrm{e}^{tC}\mathrm{e}^{-sB}, \qquad t \in \mathbb{R}.$$

Since $SU(n)$ is a group, the product $A(t)$ lies in $SU(n)$ for all $t \in \mathbb{R}$, and $A(0) = I$. By (a) above, the velocity vector $\dot{A}(0)$ lies in $su(n)$. Explicitly, $\dot{A}(0) = \mathrm{e}^{sB}C\mathrm{e}^{-sB}$. Therefore,

$$\mathrm{e}^{sB}C\mathrm{e}^{-sB} \in su(n) \qquad \text{for all} \quad s \in \mathbb{R}.$$

Differentiating this with respect to s at the point $s = 0$, we obtain that $BC - CB \in su(n)$. This argument shows that the group property of $SU(n)$ implies that the tangent space $su(n)$ of $SU(n)$ at the point I is invariant under the Lie bracket $[B, C] := BC - CB$, that is,

$$B, C \in su(n) \quad \Rightarrow \quad [B, C] \in su(n).$$

7.9 Basic Notions in Quantum Physics

> Quantum physics is based on the study of states, costates, and observables.
> Folklore

Let X and Y be complex finite-dimensional Hilbert spaces of dimension N and M, respectively, $N, M = 1, 2, \ldots$

7.9.1 States, Costates, and Observables

States. The nonzero elements ψ of the Hilbert space X are called states in quantum physics. Two states ψ and φ are said to be equivalent iff there exists a nonzero complex number μ with

$$\psi = \mu\varphi.$$

Intuitively, equivalent states represent the same physics. The notions like mean value, mean fluctuation, and transition probability to be introduced below are the same for equivalent states. The state ψ is called normalized iff $||\psi|| = 1$.

Duality and costates. By a linear functional F on the space X, we understand a linear map $F : X \to \mathbb{C}$, that is,

$$F(\alpha\psi + \beta\varphi) = \alpha F(\psi) + \beta F(\varphi)$$

for all $\psi, \varphi \in X$ and all complex numbers α, β. By definition, the set of all linear functionals on X forms the dual space of X denoted by X^d. Linear functionals are called costates in quantum physics.

Observables. The operator $A : X \to Y$ is called linear iff

$$A(\alpha\psi + \beta\varphi) = \alpha A\psi + \beta A\varphi$$

for all $\psi, \varphi \in X$ and all complex numbers α and β. By an observable on the Hilbert space X, we understand a linear self-adjoint operator $A : X \to X$, that is, we have the symmetry property

$$\langle\psi|A\varphi\rangle = \langle A\psi|\varphi\rangle \qquad \text{for all} \quad \varphi, \psi \in X.$$

Intuitively, observables are physical quantities like energy, position, momentum, angular momentum (spin). It follows from

$$\langle\psi|A\psi\rangle^\dagger = \langle A\psi|\psi\rangle = \langle\psi|A\psi\rangle$$

that $\langle\psi|A\psi\rangle$ is a real number for all states ψ. The real number

$$\boxed{\bar{A} := \frac{\langle\psi|A\psi\rangle}{||\psi||^2}} \tag{7.13}$$

is called the mean value of the observable A in the state ψ. Intuitively, this represents the mean value of the physical quantity A (for example, energy) measured in the state ψ. Moreover, the nonnegative mean fluctuation, ΔA, of the observable A in the state ψ is defined by[13]

[13] To simplify notation, we write $\bar{A}$ instead of $\bar{A} \cdot I$ where I denotes the identity operator.

$$(\Delta A)^2 = \overline{(A - \bar{A})^2}.$$

Explicitly,

$$(\Delta A)^2 = \frac{\langle \psi | (A - \bar{A})^2 \psi \rangle}{||\psi||^2}. \tag{7.14}$$

As in the theory of probability,

$$\boxed{(\Delta A)^2 = \overline{A^2} - \bar{A}^2.}$$

In fact, this follows from

$$\overline{A^2 - 2\bar{A}A + \bar{A}^2} = \overline{A^2} - 2\bar{A}\bar{A} + \bar{A}^2.$$

Furthermore, from $\langle \psi | (A - \bar{A})^2 \psi \rangle = \langle (A - \bar{A})\psi | (A - \bar{A})\psi \rangle$ we get

$$\Delta A = \frac{||(A - \bar{A})\psi||}{||\psi||}.$$

Eigenstates as sharp states. A state ψ is called an eigenstate of the observable A iff there exists a complex number λ such that

$$A\psi = \lambda\psi.$$

In an eigenstate ψ, the mean value is equal to the eigenvalue λ, and the mean fluctuation is equal to zero,

$$\bar{A} = \lambda, \qquad \Delta A = 0.$$

In fact, $\langle \psi | A\psi \rangle = \langle \psi | \lambda\psi \rangle = \lambda \langle \psi | \psi \rangle$, and $||(A - \lambda I)\psi|| = 0$. In particular, the eigenvalues of an observable are always real. Because of $\Delta A = 0$, the eigenstate ψ of the observable A is called a sharp state with the sharp value λ. Let us consider two examples:

- Molecule states are eigenstates of the energy operator (Hamiltonian). The sharp energy values are responsible for the spectrum of molecules.
- Elementary particle are eigenstates of the mass operator and the spin operator. The corresponding eigenvalues represent sharp values of mass and spin.

Transition amplitude. The complex number

$$\langle \psi | A\varphi \rangle$$

is called the transition amplitude from the state φ to the state ψ with respect to the observable A. The real number

$$\tau := \frac{|\langle\psi|A\varphi\rangle|^2}{||\psi||^2 \cdot ||\varphi||^2}$$

is called the transition probability from the state φ to the state ψ with respect to the observable A. The great importance of transition amplitudes for quantum physics was emphasized by Feynman in the 1940s. He based his approach to quantum mechanics, quantum statistics, and quantum field theory on the notion of transition amplitude. In a natural way, this leads to path integrals for the propagators, as we will discuss in Sect. 7.21.

Correlation coefficient. Let $A, B : X \to X$ be two observables. The complex number

$$\gamma := \frac{\overline{(A-\bar{A})(B-\bar{B})}}{\Delta A \cdot \Delta B}$$

is called the (generalized) correlation coefficient for the ordered pair of observables A, B in the state ψ. There holds

$$\boxed{\gamma = \frac{\overline{AB} - \bar{A}\bar{B}}{\Delta A \cdot \Delta B}.}$$

This follows from $\overline{(A-\bar{A})(B-\bar{B})} = \overline{AB} - 2\bar{A}\bar{B} + \bar{A}\bar{B}$. Explicitly,

$$\gamma = \frac{\langle\psi|(A-\bar{A})(B-\bar{B})\psi\rangle}{\Delta A \cdot \Delta B \cdot ||\psi||^2} = \frac{\langle(A-\bar{A})\psi|(B-\bar{B})\psi\rangle}{\Delta A \cdot \Delta B \cdot ||\psi||^2}.$$

By the Schwarz inequality,

$$|\gamma| \leq \frac{||(A-\bar{A})\psi|| \cdot ||(B-\bar{B})\psi||}{\Delta A \cdot \Delta B \cdot ||\psi||^2} = 1.$$

Hence $|\gamma| \leq 1$.

Commuting observables and the Gauss method of least squares. Let us now consider the special case of two observables $A, B : X \to X$ which commute, that is, $AB = BA$. It follows then from

$$\langle\psi|AB\varphi\rangle = \langle A\psi|B\varphi\rangle = \langle BA\psi|\varphi\rangle = \langle AB\psi|\varphi\rangle$$

for all $\psi, \varphi \in X$ that AB is again an observable. Thus, $\overline{AB}$ is a real number, and hence the correlation coefficient γ for A and B (in each state ψ) is also a real number. In addition, $-1 \leq \gamma \leq 1$. In order to get an intuitive interpretation of the correlation coefficient, consider the following minimum problem

$$\overline{(B-\alpha A-\beta)^2} = \min!, \qquad \alpha, \beta \in \mathbb{R}. \tag{7.15}$$

That is, by Gauss' method of least squares, we want to optimally fit the observable B by the real linear combination $\alpha A + \beta$. Assume that ΔA and ΔB do not vanish.

Theorem 7.6 *For given observables A and B with $AB = BA$, the minimum problem* (7.15) *has the solution*

$$\alpha := \gamma \cdot \frac{\Delta B}{\Delta A}, \qquad \beta := \bar{B} - \alpha \bar{A}.$$

The minimal value is equal to $(\Delta B)^2(1-\gamma^2)$.

Intuitively, this theorem tells us that the correlation coefficient γ measures the dependence between the physical quantities A and B.

- This dependence is optimal iff

$$(\Delta B)^2 \cdot (1-\gamma^2) = 0,$$

  that is, $|\gamma| = 1$. We call this strong correlation.
- The coefficient $\gamma = 0$ corresponds to weak correlation.

Proof. (I) Assume first that $\bar{A} = \bar{B} = 0$ and $\Delta A = \Delta B = 1$. Then $\gamma = \overline{AB}$ and $\overline{A^2} = (\Delta A)^2 + \bar{A}^2 = 1$. Similarly, $\overline{B^2} = 1$. Therefore,

$$\overline{(B-\gamma A)^2} = \overline{B^2} - 2\gamma\overline{AB} + \gamma^2\overline{A^2} = 1 - \gamma^2.$$

Noting that $\overline{(B-\gamma A)A} = \overline{BA} - \gamma = 0$, the trick is to set

$$\overline{(B-\alpha A-\beta)^2} = \overline{((B-\gamma A)-(\alpha-\gamma)A-\beta)^2}.$$

This is equal to

$$\overline{(B-\gamma A)^2} + (\alpha-\gamma)^2\overline{A^2} + \beta^2 = 1 - \gamma^2 + (\alpha-\gamma)^2 + \beta^2.$$

This quadratic expression is minimal if $\alpha = \gamma$ and $\beta = 0$. The minimal value is equal to $1-\gamma^2$. This proves the theorem in the present special case.

(II) For general mean values $\bar{A}, \bar{B}$, assume that $\Delta A = \Delta B = 1$. Setting $\mathcal{A} := A - \bar{A}$ and $\mathcal{B} := B - \bar{B}$, we get

$$\overline{(B-\alpha A-\beta)^2} = \overline{(\mathcal{B}-\alpha\mathcal{A}-\beta')^2} = \min!$$

with $\beta' := \beta - \bar{B} + \alpha\bar{A}$. By (I), this problem has the solution $\alpha = \gamma$ and $\beta' = 0$.

(III) For general nonzero mean fluctuations $\Delta A, \Delta B$, the proof can be easily reduced to (II) by using the rescaling $A \mapsto A/\Delta A$ and $B \mapsto B/\Delta B$. □

Causal correlation functions. Let $A = A(t)$ and $B = B(t)$ be functions depending on time t such that $A(t), B(t) : X \to X$ are observables for each $t \in \mathbb{R}$. Fix the state $\psi \in X$. For $t > s$, we set

$$\boxed{\mathcal{C}(t,s) := \langle\psi|A(t)B(s)\psi\rangle.}$$

The function $\mathcal{C}$ is called the causal correlation function. Motivated by the correlation coefficients introduced above, we expect that the value $\mathcal{C}(t,s)$ of

the causal correlation function contains information about the correlation between the observable B at time s and the observable A at the later time t, in the state ψ. The experience of physicists is that causal correlation functions represent an important tool in quantum field theory.

Quantum field theory is mainly based on the computation of causal correlations functions which describe the correlations of the quantum field between different space points at different time points.

The causal correlation functions of a quantum field are also called 2-point Green's functions and higher-order Green's functions.

7.9.2 Observers and Coordinates

Orthogonality. Two states ψ and φ of the Hilbert space X are called orthogonal iff

$$\langle\psi|\varphi\rangle = 0.$$

This is the most important notion in Hilbert spaces.

The geometry of Hilbert spaces and hence quantum physics are governed by orthogonality.

For example, this concerns Fourier series and Dirac calculus, as we will show below.

Proposition 7.7 *Two eigenstates of an observable A with different eigenvalues are orthogonal.*

To prove this, let $A\psi = \lambda\psi$ and $A\varphi = \mu\varphi$ wit $\lambda \neq \mu$. Since the eigenvalues λ and μ are real, $(\lambda - \mu)\langle\psi|\varphi\rangle = \langle A\psi|\varphi\rangle - \langle\psi|A\varphi\rangle = 0$. Hence $\langle\psi|\varphi\rangle = 0$.

Fourier coefficients of states. If an observer measures the state ψ by a measurement device, then this corresponds to the decomposition

$$\boxed{\psi = c_1\varphi_1 + \ldots + c_N\varphi_N} \tag{7.16}$$

where $\varphi_1, \ldots, \varphi_N$ is a basis of the Hilbert space X, and $c_1, \ldots, c_N$ are complex numbers. In order to get a physical interpretation of this decomposition, assume additionally that the basis $\varphi_1, \ldots, \varphi_N$ forms an orthonormal system, that is, $\langle\varphi_j|\varphi_k\rangle = \delta_{jk}$ for all j, k.[14] Then

$$\langle\psi|\psi\rangle = |c_1|^2 + \ldots + |c_N|^2.$$

Finally, assume that the state ψ is normalized, that is, $\langle\psi|\psi\rangle = 1$. Then

$$\boxed{|c_1|^2 + \ldots + |c_N|^2 = 1.}$$

This allows us to give the following interpretation:

[14] For the Kronecker symbol, recall that $\delta_{jk} := 1$ if $j = k$, and $\delta_{jk} := 0$ if $j \neq k$. Here, $j, k = 1, \ldots N$.

- Our measurement device is able to detect the states $\varphi_1, \ldots, \varphi_N$.
- If the particle is in the state ψ, then $|c_j|^2$ is the probability of measuring the state φ_j.

We call $c_1, \ldots, c_N$ the coordinates of the state ψ with respect to the basis $\varphi_1, \ldots, \varphi_N$. Explicitly,

$$c_j = \langle \varphi_j | \psi \rangle, \qquad j = 1, \ldots N.$$

This follows from (7.16) by using $\langle \varphi_k | \psi \rangle = \langle \varphi_k | c_k \varphi_k \rangle = c_k$. The complex number c_k is called the Fourier coefficient of the state ψ.

Matrix elements of observables. The matrix elements

$$\boxed{a_{jk} := \langle \varphi_j | A \varphi_k \rangle, \qquad j, k = 1, \ldots, N}$$

are called the coordinates of the observable A with respect to the basis $\varphi_1, \ldots, \varphi_N$. The Dirac calculus was invented by Dirac in the late 1920s; it allows us to pass elegantly from one coordinate system to another one. Dirac called this transformation theory. The Dirac calculus translates well-known notions from linear algebra into a convenient mnemonical language. In this monograph, we will always use Dirac's language. To begin with, let us formulate the relevant notions in the traditional mathematical language.

7.10 Fourier Series

Theorem 7.8 *Each nontrivial finite-dimensional Hilbert space possesses an orthonormal basis.*

The standard proof based on the Schmidt orthogonalization method can be found in Zeidler (1995), Vol. 1, Sect. 3.3.

Key relations. For fixed $N = 1, 2, \ldots$, let X be an N-dimensional complex Hilbert space.

(i) Orthonormal system: By definition, the states $\varphi_1, \ldots, \varphi_K$ form an orthonormal system of the Hilbert space X iff

$$\langle \varphi_j | \varphi_k \rangle = \delta_{jk}, \qquad j, k = 1, \ldots, K.$$

This orthonormal system is called complete iff the states form a basis of X, that is, $K = N$. Then, each state ψ of X can be represented by the finite Fourier series

$$\boxed{\psi = \sum_{j=1}^{N} \langle \varphi_j | \psi \rangle \varphi_j.} \tag{7.17}$$

The complex numbers $\langle \varphi_j | \psi \rangle$, $j = 1, \ldots, N$, are called the Fourier coefficients of the state ψ.

(iii) Parseval equation: If $\varphi_1, \ldots, \varphi_N$ is a complete orthonormal system in the Hilbert space X, then[15]

$$\boxed{\langle\varphi|\psi\rangle = \sum_{j=1}^{N} \langle\varphi|\varphi_j\rangle\langle\varphi_j|\psi\rangle.}$$

This so-called Parseval equation represents the inner product in terms of Fourier coefficients.[16]

(iv) Duality: For each state φ in X, set

$$F_\varphi(\chi) := \langle\varphi|\chi\rangle \qquad \text{for all} \quad \chi \in X.$$

This is a linear functional on the Hilbert space X called the costate, F_φ, corresponding to the state φ.

The Riesz theorem tells us that each linear functional on the Hilbert space X is the costate to some state of X.[17]

(v) Two different products: For the costate $F \in X^d$ and the state $\psi \in X$, we introduce the inner product

$$F \cdot \psi := F(\psi),$$

and the tensor product $\psi \otimes F$. By definition, this is a linear operator $\psi \otimes F : X \to X$ given by

$$(\psi \otimes F)(\chi) := F(\chi)\psi \qquad \text{for all} \quad \chi \in X.$$

(vi) Dirac's completeness relation: An orthonormal system $\varphi_1, \ldots, \varphi_N$ is complete iff

$$\boxed{\sum_{j=1}^{N} \varphi_j \otimes F_{\varphi_j} = I}$$

where I denotes the identity operator on X. In fact,

$$\sum_{j=1}^{N} (\varphi_j \otimes F_{\varphi_j})\psi = \sum_{j=1}^{N} F_{\varphi_j}(\psi)\varphi_j = \sum_{j=1}^{N} \langle\varphi_j|\psi\rangle\varphi_j = \psi.$$

[15] To prove this, note that the inner product $\langle\varphi|\psi\rangle$ is equal to

$$\langle \sum_{j=1}^{N} c_j\varphi_j | \sum_{k=1}^{N} d_k\varphi_k\rangle = \sum_{j,k=1}^{N} c_j^\dagger d_k \langle\varphi_j|\varphi_k\rangle = \sum_{j,k=1}^{N} c_j^\dagger d_k \delta_{jk} = \sum_{j=1}^{N} c_j^\dagger d_j$$

along with $c_j = \langle\varphi_j|\varphi\rangle$ and $d_k := \langle\varphi_k|\psi\rangle$, as well as $c_j^\dagger = \langle\varphi|\varphi_j\rangle$.

[16] Parseval des Chénes (1755–1836)

[17] For the proof, see Zeidler (1995), Vol. 1, Sect. 2.10.

(vii) Adjoint operator: For each linear operator $A : X \to X$, there exists a unique linear operator $A^\dagger : X \to X$ such that

$$\langle A\varphi|\psi\rangle = \langle\varphi|A^\dagger\psi\rangle \qquad \text{for all} \quad \varphi, \psi \in X. \tag{7.18}$$

The operator $A^\dagger$ is called the adjoint operator to A. An observable A is characterized by the condition $A = A^\dagger$. Such operators are called self-adjoint.

(viii) Dual operator: For each linear operator $A : X \to X$, the dual operator $A^d : X^d \to X^d$ is defined as follows. To each costate $F \in X^d$, we assign the costate $A^d F$ given by

$$(A^d F)(\varphi) := F(A\varphi) \qquad \text{for all} \quad \varphi \in X.$$

Observe that the adjoint operator $A^\dagger$ and the dual operator A^d act on different spaces. Let $\varphi_1, \ldots, \varphi_N$ be an orthonormal basis of the Hilbert space X. Then, the matrix elements of the adjoint operator $A^\dagger$ are given by

$$b_{jk} := \langle\varphi_j|A^\dagger\varphi_k\rangle, \qquad j, k = 1, \ldots, N.$$

Hence

$$b_{jk} = \langle A\varphi_j|\varphi_k\rangle = \langle\varphi_k|A\varphi_j\rangle^\dagger = a_{kj}^\dagger, \qquad j, k = 1, \ldots, N.$$

The $(N \times N)$-matrix $\mathcal{A}^\dagger := (b_{jk})$ is called the adjoint matrix to the matrix $\mathcal{A} := (a_{jk})$.

The main theorem on observables. The following theorem is called the principal axis theorem in mathematics, since it guarantees the existence of a principal axis for quadratic curves and surfaces (for example, ellipses and ellipsoids). This theorem can be traced back to Euler (1707–1783) and Cauchy (1789–1857).[18]

Theorem 7.9 *Each observable $A : X \to X$ possesses a complete orthonormal system of eigenstates $\varphi_1, \ldots, \varphi_N$.*

Let $A\varphi_j = \lambda_j\varphi_j$ for $j = 1, \ldots, N$ where $\varphi_1, \ldots, \varphi_N$ is a complete orthonormal system in X. Then

$$A\psi = \sum_{j=1}^{N} \lambda_j \langle\varphi_j|\psi\rangle\varphi_j \qquad \text{for all} \quad \psi \in X.$$

This follows from $A\psi = \sum_{j=1}^N \langle\varphi_j|\psi\rangle A\varphi_j$.

Functions of observables. Let $F : \mathbb{R} \to \mathbb{C}$ be an arbitrary function. For given observable $A : X \to X$, we define the operator $F(A) : X \to X$ by setting

[18] The proof can be found in Zeidler (1995), Vol. 1, Sect. 4.2.

$$\boxed{F(A)\varphi := \sum_{j=1}^{N} F(\lambda_j)\langle\varphi_j|\varphi\rangle\varphi_j}$$

for all $\varphi \in X$. This way, we get the operator function $A \mapsto F(A)$. More generally, the definition of the operator $F(A)$ makes sense if the function $F : \sigma(A) \to \mathbb{C}$ is only defined on the spectrum $\sigma(A)$ of the observable A, that is, on the set of eigenvalues $\lambda_1, \ldots, \lambda_N$ of A. For example, if the complex number λ is not in the spectrum $\sigma(A)$, then it makes sense to define

$$R_\lambda(A)\varphi := \sum_{j=1}^{N} \frac{\langle\varphi_j|\varphi\rangle}{\lambda - \lambda_j}\,\varphi_j \qquad \text{for all} \quad \varphi \in X.$$

Obviously, $R_\lambda(A)(\lambda I - A) = (\lambda A - I)R_\lambda(A) = I$. Hence $R_\lambda(A) = (\lambda I - A)^{-1}$. This operator is called the resolvent of the observable A at the point $\lambda \in \mathbb{C}$. In the language of the Dirac calculus to be considered in the next section, we get the elegant formula

$$\boxed{F(A) = \sum_{j=1}^{N} F(\lambda_j)\ |\varphi_j\rangle\langle\varphi_j|.} \tag{7.19}$$

Fourier transform. Let X be a finite-dimensional complex Hilbert space of dimension $N = 1, 2, \ldots$. Recall that the Hilbert space $\mathbb{C}^N$ is equipped with the inner product

$$\langle x|y\rangle := \sum_{j=1}^{N} x_j^\dagger y_j.$$

The operator $\mathcal{F} : X \to \mathbb{C}^N$ is defined by

$$(\mathcal{F}\psi)(j) := \langle\varphi_j|\psi\rangle, \qquad j = 1, \ldots, N.$$

That is, the operator $\mathcal{F}$ maps each state ψ to the sequence of its Fourier coefficients,

$$\mathcal{F}\psi = (\langle\varphi_1|\psi\rangle, \ldots, \langle\varphi_N|\psi\rangle).$$

The operator $\mathcal{F}$ is called the discrete Fourier transformation with respect to the orthonormal basis $\varphi_1, \ldots, \varphi_N$. The Parseval equation tells us that

$$\boxed{\langle\varphi|\psi\rangle = \langle\mathcal{F}\varphi|\mathcal{F}\psi\rangle \qquad \text{for all} \quad \varphi, \psi \in X.}$$

Consequently, the operator $\mathcal{F} : X \to \mathbb{C}^N$ is unitary, and the Hilbert space X is isomorphic (or unitarily equivalent) to the Hilbert space $\mathbb{C}^N$.

Let $A : X \to X$ be an observable with the eigenvalues $\lambda_1, \ldots, \lambda_N$ and the eigenvectors $\varphi_1, \ldots, \varphi_N$, respectively. Then

$$\mathcal{F}(A\psi) = (\lambda_1\langle\varphi_j|\psi\rangle, \ldots, \lambda_N\langle\varphi_N|\psi\rangle).$$

This means that the observable A corresponds to the multiplication operator

$$\langle\varphi_j|\psi\rangle \mapsto \lambda_j\langle\varphi_j|\psi\rangle, \qquad j = 1, \ldots, N$$

in the Fourier space $\mathbb{C}^N$. We will show later on that there exists a far-reaching generalization to self-adjoint operators in infinite-dimensional Hilbert spaces due to John von Neumann.

Reformulation in terms of some measure integral. When generalizing the notion of finite Fourier series to infinite dimensions, we will encounter infinite series and integrals with appropriate weights which can be regarded as continuous sums (Fourier–Stieltjes integral transformations). All of these notions can be described in a unique manner by the basic notion of measure integral to be considered later on. In particular, for an arbitrary function $f : \mathbb{R} \to \mathbb{C}$, we get

$$\boxed{\int_{\mathbb{R}} f(x)d\mu(x) = \sum_{j=1}^{N} f(j).}$$

Here, the measure μ corresponds to a mass distribution on the real line where precisely the points $x_1 = 1, \ldots, x_N = N$ possess the mass one. Furthermore, for an arbitrary subset Ω of the real line,

$$\int_{\Omega} f(x)d\mu(x) = \sum_{j\in\Omega} f(j).$$

This reformulation of finite sums in terms of a measure integral serves as preparation for generalizing the Dirac calculus to infinite dimensions.

In what follows we will translate the relations above into Dirac's language which is used in physics textbooks.

7.11 Dirac Calculus in Finite-Dimensional Hilbert Spaces

> The Dirac calculus works on its own.
> Folklore

Dirac's bra and ket symbols. The inner product $\langle\varphi|\psi\rangle$ represents a bracket. Mnemonically, it was Dirac's idea to introduce the bra symbol $\langle\varphi|$ and the ket symbol $|\psi\rangle$ which possess the following meaning.

- States and costates: Let $\varphi, \psi \in X$ and $F \in X^d$. We write

$$|\psi\rangle := \psi, \qquad \langle\varphi| := F_\varphi, \qquad |\psi\rangle\langle\varphi| := \psi \otimes F_\varphi.$$

This yields the inner product $\langle\varphi| \cdot |\psi\rangle = \langle\varphi|\psi\rangle$, and the Dirac product

$$\boxed{|\psi\rangle\langle\varphi| \cdot |\chi\rangle = |\psi\rangle\langle\varphi|\chi\rangle.}$$

- Adjointness: We set $|\psi\rangle^\dagger = \langle\psi|$ and $\langle\varphi|^\dagger = |\varphi\rangle$, as well as

$$(ab)^\dagger = b^\dagger a^\dagger$$

for operators and states. It turns out that this convention implies the right rules for both the inner product and the adjoint operator. In fact, we obtain

$$\langle\varphi|\psi\rangle^\dagger = (\langle\varphi| \cdot |\psi\rangle)^\dagger = \langle\psi| \cdot |\varphi\rangle = \langle\psi|\varphi\rangle$$

where the dot denotes the inner product. Furthermore,

$$\langle A\varphi|\psi\rangle = \langle\psi|A\varphi\rangle^\dagger = (\langle\psi|A \cdot |\varphi\rangle)^\dagger = \langle\varphi| \cdot A^\dagger|\psi\rangle = \langle\varphi|A^\dagger\psi\rangle.$$

- Dirac's completeness relation: An orthonormal system $\varphi_1, \ldots, \varphi_N$ is complete iff

$$\boxed{\sum_{j=1}^{N} |\varphi_j\rangle\langle\varphi_j| = I.} \tag{7.20}$$

Dirac's substitution trick and coordinates. Dirac's completeness relation allows us to elegantly reformulate equations for states, costates, and operators in terms of coordinates (that is, Fourier coefficients and matrix elements). As a prototype for Dirac's substitution trick, use the completeness relation (7.20) in order to get

$$\boxed{\langle\varphi|\psi\rangle = \langle\varphi|I|\psi\rangle = \sum_{j=1}^{N} \langle\varphi|\varphi_j\rangle\langle\varphi_j|\psi\rangle.}$$

This is the Parseval equation.

(i) Transition amplitude: Let $A : X \to X$ be a linear operator. Then[19]

$$\langle\varphi|A|\psi\rangle = \sum_{j,k=1}^{N} \langle\varphi|\varphi_j\rangle\langle\varphi_j|A|\varphi_k\rangle\langle\varphi_k|\psi\rangle.$$

If we set $b_j := \langle\varphi_j|\varphi\rangle$ and $c_k := \langle\varphi_k|\psi\rangle$, as well as $a_{jk} := \langle\varphi_j|A|\varphi_k\rangle$, then we get the matrix equation

$$\langle\varphi|A|\psi\rangle = \sum_{j,k=1}^{N} b_j a_{jk} c_k = b^d \mathcal{A} c$$

with the matrix $\mathcal{A} := (a_{jk})$ corresponding to the operator A.

[19] We use synonymously the two symbols $\langle\varphi|A\psi\rangle$ and $\langle\varphi|A|\psi\rangle$.

(ii) Operator equation: Our next goal is to reformulate the equation

$$\boxed{\varphi = A\psi}$$

in terms of coordinates. By Dirac's substitution trick,

$$\langle \varphi_j|\varphi\rangle = \sum_{k=1}^{N} \langle \varphi_j|A|\varphi_k\rangle\langle \varphi_k|\psi\rangle, \qquad j = 1, \dots, N.$$

This can be written as $b_j = \sum_{k=1}^{N} a_{jk}c_k$. In the language of matrices, $b = \mathcal{A}c$. Explicitly,

$$\begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ \vdots & & & \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_N \end{pmatrix}.$$

(iii) Product of operators: Finally, let us reformulate the product

$$C = AB$$

of the two linear operators $A, B : X \to X$ in terms of coordinates. Dirac's substitution trick yields

$$\langle \varphi_j|AB|\varphi_k\rangle = \sum_{s=1}^{N} \langle \varphi_j|A|\varphi_s\rangle\langle \varphi_s|B|\varphi_k\rangle.$$

For the matrix elements,

$$c_{jk} = \sum_{s=1}^{N} a_{js}b_{sk}, \qquad j, k = 1, \dots, N.$$

This corresponds to the matrix equation $\mathcal{C} = \mathcal{AB}$.

In Sect. 7.21 we will use Dirac's substitution trick in order to obtain the Feynman path integral.

Costates. Let F and ψ be a costate and a state, respectively, that is, $F \in X^d$ and $\psi \in X$. According to Dirac, we write

$$\langle F| \cdot |\psi\rangle = \langle F|\psi\rangle := F(\psi).$$

In particular, this yields

$$\langle F_\varphi|\psi\rangle = \langle \varphi|\psi\rangle \qquad \text{for all} \quad \varphi, \psi \in X.$$

By Dirac's completeness relation (7.20),

$$\boxed{\langle F|\psi\rangle = \sum_{j=1}^{N} \langle F|\varphi_j\rangle\langle\varphi_j|\psi\rangle.}$$

Therefore, we write $\langle F| = \sum_{j=1}^{N} \langle F|\varphi_j\rangle\langle\varphi_j|$.[20] The complex numbers

$$\langle F|\varphi_1\rangle, \ldots, \langle F|\varphi_N\rangle$$

are called the coordinates (or the Fourier coefficients) of the costate F.

Dual operator. We are given the linear operator $A : X \to X$. According to Sect. 7.10, the dual operator $A^d : X^d \to X^d$ is defined by the relation $(A^d F)(\varphi) = F(A\varphi)$. In the Dirac calculus, this can be written as

$$\boxed{\langle A^d F|\varphi\rangle = \langle F|A\varphi\rangle \qquad \text{for all} \quad F \in X^d, \varphi \in X.}$$

For each $F \in X^d$, let us formulate the equation

$$G = A^d F$$

in terms of coordinates. By Dirac's completeness relation (7.20),

$$\langle G|\varphi_j\rangle = \langle F|A\varphi\rangle = \sum_{k=1}^{N} \langle F|\varphi_k\rangle\langle\varphi_k|A\varphi_j\rangle.$$

This yields the matrix equation

$$\langle G|\varphi_j\rangle = \sum_{k=1}^{N} b_{jk}\langle F|\varphi_k\rangle, \qquad j = 1, \ldots, N$$

where we set

$$b_{jk} := a_{kj} = \langle\varphi_k|A\varphi_j\rangle, \qquad j, k = 1, \ldots, N.$$

The $(N \times N)$-matrix $\mathcal{A}^d := (b_{jk})$ is called the dual (or transposed) matrix to the matrix $\mathcal{A} = (a_{jk})$. In terms of the dual operator,

[20] In the traditional language of mathematics,

$$F = \sum_{j=1}^{N} F(\varphi_j) F_{\varphi_j}.$$

This follows from $\psi = c_1\varphi_1 + \ldots + c_N\varphi_N$ along with $c_j = \langle\varphi_j|\psi\rangle = F_{\varphi_j}(\psi)$. Hence

$$F(\psi) = F\left(\sum_{j=1}^{N} c_j\varphi_j\right) = \sum_{j=1}^{N} c_j F(\varphi_j) = \sum_{j=1}^{N} F(\varphi_j) F_{\varphi_j}(\psi).$$

$$b_{jk} = \langle A^d F_{\varphi_k} | \varphi_j \rangle, \qquad j,k = 1,\dots,N.$$

In fact, $\langle A^d F_{\varphi_k} | \varphi_j \rangle = \langle F_{\varphi_k} | A\varphi_j \rangle = \langle \varphi_k | A\varphi_j \rangle = a_{kj} = b_{jk}$.

Finally, let us mention that for an observable $A : X \to X$, we get

$$\boxed{\langle A^d F_\varphi | \psi \rangle = \langle A\varphi | \psi \rangle \qquad \text{for all} \quad \varphi, \psi \in X.}$$

This follows from $\langle A^d F_\varphi | \psi \rangle = \langle F_\varphi | A\psi \rangle = \langle \varphi | A\psi \rangle = \langle A\varphi | \psi \rangle$.

7.12 The Trace of a Linear Operator

> As all roads lead to Rome so I find in my own case at least that all algebraic inquiries, sooner or later, end at the Capitol of Modern Algebra over whose shining portal is inscribed the Theory of Invariants.[21]
>
> James Sylvester (1814–1897)

> The theory of invariants came into existence about the middle of the nineteenth century somewhat like Minerva: a grown-up virgin, mailed in the shining armor of algebra, she sprang forth from Cayley's Jovian head:[22] Her Athens over which she ruled and which she served as a tutelary and beneficent goddess was projective geometry.[23]
>
> Hermann Weyl (1885–1955)

> Geometry is the invariant theory of groups of transformations.
>
> Felix Klein (1849–1925)
> *Erlanger Program,* 1872

Let $A : X \to X$ be a linear operator on the complex N-dimensional Hilbert space X with $N = 1, 2, \dots$ Choose a fixed complete orthonormal system $\varphi_1, \dots, \varphi_N$ and define the trace of A by setting

$$\boxed{\operatorname{tr}(A) := \sum_{j=1}^{N} \langle \varphi_j | A\varphi_j \rangle.} \tag{7.21}$$

Proposition 7.10 *The value* $\operatorname{tr}(A)$ *does not depend on the choice of the complete orthonormal system.*

In terms of physics this means that different observers use different matrices (a_{jk}) for describing the abstract observable A (for example, energy), but all of them compute the same matrix trace, $\operatorname{tr}(A) = \sum_{j=1}^{N} a_{jj}$.

[21] As an introduction to invariant theory, we recommend the textbook by P. Olver, Classical Invariant Theory, Cambridge University Press, 1999.

[22] Cayley (1821–1895)

[23] H. Weyl, The Classical Groups, Princeton University Press, 1938 (8th edition, 1973). H. Weyl, Invariants, Duke Math. J. **5** (1939), 489–502.

Proof. Let $\psi_1, \ldots, \psi_N$ be an orthonormal basis. By Dirac's completeness relation (7.20), we obtain

$$\operatorname{tr}(A) = \sum_{j=1}^{N} \langle \varphi_j | A \varphi_j \rangle = \sum_{j,k,r=1}^{N} \langle \varphi_j | \psi_k \rangle \langle \psi_k | A \psi_r \rangle \langle \psi_r | \varphi_j \rangle$$

along with $\sum_{j=1}^{N} \langle \psi_r | \varphi_j \rangle \langle \varphi_j | \psi_k \rangle = \langle \psi_r | \psi_k \rangle = \delta_{rk}$. This yields the claim $\operatorname{tr}(A) = \sum_{r=1}^{N} \langle \psi_r | A \psi_r \rangle$. □

Proposition 7.11 *For linear operators $A, B : X \to X$ and complex numbers α, β the following are met:*

(i) $\operatorname{tr}(\alpha A + \beta B) = \alpha \operatorname{tr} A + \beta \operatorname{tr} B$,
(ii) $\operatorname{tr}(AB) = \operatorname{tr}(BA)$,
(iii) $\operatorname{tr} A^\dagger = (\operatorname{tr} A)^\dagger$.

Proof. Ad (i). $\sum_j \langle \varphi_j | \alpha A + \beta B | \varphi_j \rangle = \alpha \sum_j \langle \varphi_j | A | \varphi_j \rangle + \beta \sum_j \langle \varphi_j | B | \varphi_j \rangle$.

Ad (ii). By Dirac's completeness relation (7.20),

$$\operatorname{tr}(AB) = \sum_{j=1}^{N} \langle \varphi_j | AB \varphi_j \rangle = \sum_{j,k=1}^{N} \langle \varphi_j | A \varphi_k \rangle \langle \varphi_k | B \varphi_j \rangle.$$

This is equal to

$$\sum_{j,k=1}^{N} \langle \varphi_k | B \varphi_j \rangle \langle \varphi_j | A \varphi_k \rangle = \sum_{k=1}^{N} \langle \varphi_k | BA \varphi_k \rangle = \operatorname{tr}(BA).$$

Ad (iii). $\operatorname{tr} A^\dagger = \sum_j \langle \varphi_j | A^\dagger \varphi_j \rangle = \sum_j \langle A \varphi_j | \varphi_j \rangle = \sum_j \langle \varphi_j | A \varphi_j \rangle^\dagger$. □

For three linear operators $A, B, C : X \to X$, it follows from the associative law, $ABC = A(BC) = (AB)C$, that

$$\operatorname{tr}(ABC) = \operatorname{tr}(BCA) = \operatorname{tr}(CAB).$$

This means that the trace is invariant under cyclic permutations. The same argument shows that the trace of a finite product of operators is invariant under cyclic permutations of the factors. For example,

$$\operatorname{tr}(ABCD) = \operatorname{tr}(BCDA).$$

Linear invariance. Let $T : X \to X$ be a linear invertible operator. Then, $TT^{-1} = I$ implies that

$$\operatorname{tr}(TAT^{-1}) = \operatorname{tr}(AT^{-1}T) = \operatorname{tr}(A).$$

This tells us that the trace of a linear operator on a finite-dimensional Hilbert space is invariant under linear isomorphisms.

The theory of invariants plays a fundamental role in physics.

We will encounter this quite often in this treatise. For example, the trace of observables is fundamental for statistical physics and quantum physics. This importance of the trace stems from the fact that $\mathrm{tr}(A)$ assigns a real number to the observable A. This real number can be measured in a physical experiment.

The characteristic equation. Let $A : X \to X$ be a linear operator on the complex N-dimensional linear space X with $N = 1, 2, \ldots$. The equation

$$\boxed{\det(\lambda I - A) = 0, \qquad \lambda \in \mathbb{C}} \tag{7.22}$$

is called the characteristic equation (or the secular equation) of the linear operator A. For example, this equation was used by Lagrange (1736–1813) and Laplace (1749–1827) for computing the long-term (secular) perturbations of the orbits of planets. The solutions of (7.22) are precisely the eigenvalues $\lambda_1, \ldots, \lambda_N$ of the linear operator A. The set

$$\sigma(A) := \{\lambda_1, \ldots, \lambda_N\}$$

is called the spectrum of A, whereas the complement $\varrho(A) := \mathbb{C} \setminus \sigma(A)$ is called the resolvent set of A.

The inverse operator $(\lambda I - A)^{-1}$ *exists iff* $\lambda \in \varrho(A)$.

This operator is called the resolvent of the operator A at the point $\lambda \in \mathbb{C}$. Equation (7.22) can be written as

$$\lambda^N + p_{N-1}\lambda^{N-1} + \ldots + p_1\lambda + p_0 = 0.$$

Since the determinant of a linear operator is an invariant, the coefficients $p_0, p_1, \ldots$ are invariants of the linear operator A. They are called the principal invariants of A. In particular,

$$p_0 = (-1)^N \det A, \qquad p_{N-1} = -\,\mathrm{tr}(A).$$

In terms of the eigenvalues of A,

$$\det(\lambda I - A) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_N) = 0.$$

This implies

$$\boxed{\det(A) = \lambda_1\lambda_2\cdots\lambda_N, \qquad \mathrm{tr}(A) = \lambda_1 + \lambda_2 + \ldots + \lambda_N.}$$

For example, the characteristic equation $\det(\lambda I - A) = 0$ of the (2×2)-matrix $A = (a_{ij})$ reads as

$$(\lambda - a_{11})(\lambda - a_{22}) - a_{12}a_{21} = \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21} = 0.$$

This is identical to $\lambda^2 - \mathrm{tr}(A)\,\lambda + \det(A) = 0$.

7.13 Banach Spaces

> Banach spaces are used in order to study the convergence of iterative methods which play a fundamental role in perturbation theory.
>
> Folklore

Let $-\infty < a < b < \infty$. The prototype of a Banach space is the space $C[a,b]$ of all continuous functions $f : [a,b] \to \mathbb{C}$ equipped with the norm

$$||f|| := \max_{a \le t \le b} |f(t)|.$$

Normed space. Set $X := C[a,b]$. For all $f, g \in X$ and all complex numbers α, the following are met:

(B0) X is a complex linear space.
(B1) To each element f of X, there is assigned a nonnegative real number $||f||$. Moreover, $||f|| = 0$ iff $f = 0$.
(B2) Triangle inequality: $||f + g|| \le ||f|| + ||g||$.
(B3) $||\alpha f|| = |\alpha| \cdot ||f||$.

Generally, a space X is called a complex normed space iff the conditions (B0) through (B3) are satisfied. By definition, a subset M of a normed space is bounded iff there exists a positive number r such that $||f|| \le r$ for all $f \in M$. Furthermore, let X and Y be normed spaces, and let M be a subset of X. The operator $A : M \to Y$ is called bounded iff it transforms bounded sets into bounded sets.

Completeness. The convergence $\lim_{n\to\infty} f_n = f$ in a complex normed space X is defined by

$$\lim_{n\to\infty} ||f - f_n|| = 0.$$

A sequence (f_n) in X is called a Cauchy sequence iff for each given $\varepsilon > 0$, there exists an index $n_0(\varepsilon)$ such that

$$||f_n - f_m|| < \varepsilon \qquad \text{for all} \quad n, m \ge n_0(\varepsilon).$$

A complex normed space is called a Banach space iff it is complete, that is, each Cauchy sequence is convergent. It can be shown that each finite-dimensional complex normed space is a Banach space. For the proof, see Zeidler (1995), Vol. 1, Sect. 1.12. Moreover, with respect to the norm

$$||\varphi|| := \sqrt{\langle \varphi | \varphi \rangle},$$

each pre-Hilbert space (resp. Hilbert space) is a normed space (resp. Banach space).

The Banach fixed-point theorem. Consider the operator equation

$$\boxed{\varphi = \kappa A \varphi, \qquad \varphi \in \mathcal{B}.} \tag{7.23}$$

We are given the complex number κ. We are looking for an element φ of the complex Banach space X which lies in the ball $\mathcal{B} := \{\varphi \in X : ||\varphi|| \leq r\}$ of fixed radius $r > 0$. In addition, we set $\varphi_0 := 0$, and we consider the iterative method

$$\varphi_{n+1} = \kappa A\varphi_n, \qquad n = 0, 1, \ldots. \tag{7.24}$$

Theorem 7.12 *Suppose that the operator $A : \mathcal{B} \to X$ satisfies the Lipschitz condition*

$$||A\varphi - A\chi|| \leq \text{const}\, ||\varphi - \chi|| \qquad \textit{for all} \quad \varphi, \chi \in \mathcal{B}.$$

Then, there exists a number $\kappa_0 > 0$ such that, for each given complex number κ with $|\kappa| \leq \kappa_0$, equation (7.23) *has a unique solution φ, and the iterative method* (7.24) *converges to φ as $n \to \infty$.*

The proof of this theorem along with applications to numerous problems can be found in Zeidler (1986), Vol. I.

The Banach space $L(X, Y)$. Let X and Y be finite-dimensional complex Banach spaces. By $L(X, Y)$, we denote the set of all linear operators

$$A : X \to Y.$$

For all $A, B \in L(X, Y)$ and all complex numbers α, β, the linear combination $\alpha A + \beta B$ is defined by

$$(\alpha A + \beta B)\varphi := \alpha A\varphi + \beta B\varphi \qquad \text{for all} \quad \varphi \in X.$$

This way, $L(X, Y)$ becomes a complex linear space. For each $A \in L(X, Y)$, the norm

$$||A|| := \max_{||\varphi|| \leq 1} ||A\varphi||$$

is a well-defined real number, and it can be shown that $L(X, Y)$ becomes a finite-dimensional complex Banach space. For the proof, see Zeidler (1995), Vol. 1, Sect. 1.20. In particular, if we choose $Y := \mathbb{C}$ then the dual space X^d of all linear functionals $F : X \to \mathbb{C}$ becomes a complex Banach space with respect to the norm

$$||F|| := \max_{||\varphi|| \leq 1} |F(\varphi)|.$$

Analytic operator functions. Let $a_0, a_1, a_2, \ldots$ be complex numbers. Suppose that the power series expansion

$$f(z) = a_0 + a_1 z + a_2 z^2 + \ldots$$

is convergent for all complex numbers z with $|z| < r$. Define

$$\boxed{f(A) := a_0 I + a_1 A + a_2 A^2 + \ldots = \sum_{k=0}^{\infty} a_k A^k.} \tag{7.25}$$

This series converges for all operators $A \in L(X, X)$ with $||A|| < r$, and we have $f(A) \in L(X, X)$. Explicitly,

$$\lim_{n \to \infty} ||f(A) - \sum_{k=0}^{n} a_k A^k|| = 0.$$

This convergence is equivalent to the convergence of the corresponding matrix elements with respect to any fixed basis of the linear space X. As an example, let us consider the geometric series

$$f(z) := 1 + z^2 + z^3 + \ldots$$

which converges for all $z \in \mathbb{C}$ with $|z| < 1$. Explicitly, $f(z) = \frac{1}{1-z}$. For all $A \in L(X, X)$ with $||A|| < 1$, we set

$$f(A) := I + A + A^2 + \ldots$$

It turns out that $f(A) = (I - A)^{-1}$. Analogously, we get the exponential function

$$\mathrm{e}^A = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \ldots$$

for all operators $A \in L(X, X)$. The proofs can be found in Zeidler (1995), Vol. 1, Sect. 1.22.

7.14 Probability and Hilbert's Spectral Family of an Observable

We want to formulate basic notions in finite-dimensional Hilbert spaces in such a language that it can be easily generalized to infinite dimensions later on, by using John von Neumann's operator calculus. In this section, we use the measure integral to be introduced in Sect. 10.2.1. The reader who is not familiar with this notion of integral should only look at the finite sums below.

Hilbert's spectral family. Let $A : X \to X$ be an observable on the complex N-dimensional Hilbert space X. Furthermore, let $\varphi_1, \ldots, \varphi_N$ be a complete orthonormal system of eigenstates of the operator A with the eigenvalues $\lambda_1, \ldots, \lambda_N$, respectively. That is, $A\varphi_j = \lambda_j \varphi_j$ for $j = 1, \ldots, N$. In 1904 Hilbert introduced the spectral operator $E_\lambda : X \to X$ given by

$$\boxed{E_\lambda := \sum_{\lambda_j < \lambda} |\varphi_j\rangle\langle\varphi_j|, \qquad \lambda \in \mathbb{R}.}$$

The symbol $\sum_{\lambda_j<\lambda}$ means that we sum over all indices j with the property that $\lambda_j < \lambda$. We call $\{E_\lambda\}_{\lambda\in\mathbb{R}}$ the spectral family of the observable A. Spectral families play a crucial role in quantum physics for describing the energy spectrum of atoms, molecules, and elementary particles.

Hilbert's spectral integral. For each given observable A,

$$A = \int_{\mathbb{R}} \lambda dE_\lambda = \sum_{j=1}^{N} \lambda_j |\varphi_j\rangle\langle\varphi_j|.$$

Let $\varphi, \psi \in X$. For the transition amplitude,

$$\langle\varphi|A\psi\rangle = \int_{\mathbb{R}} \lambda d\langle\varphi|E_\lambda\psi\rangle = \sum_{j=1}^{N} \lambda_j \langle\varphi|\varphi_j\rangle\langle\varphi_j|\psi\rangle.$$

Randomness of quantum processes. Suppose that we measure the observable A (e.g., energy) in the normalized state ψ, and we obtain the value a. The probability, $P(\Omega)$, of finding the measured value a in the set Ω is defined to be

$$P(\Omega) := \sum_{\lambda_j\in\Omega} |\langle\varphi_j|\psi\rangle|^2.$$

Observe that if Ω is the real line, then $P(\mathbb{R}) = \sum_{j=1}^{N} |\langle\varphi_j|\psi\rangle|^2 = \langle\psi|\psi\rangle = 1$, by the Parseval equation, as expected. In the language of measure integrals,

$$P(\Omega) = \int_\Omega d\langle E_\lambda\psi|E_\lambda\psi\rangle.$$

This integral refers to the variable λ. Since the operator E_λ is self-adjoint and $E_\lambda^2 = E_\lambda$, we get $\langle E_\lambda\psi|E_\lambda\psi\rangle = \langle\psi|E_\lambda^2\psi\rangle = \langle\psi|E_\lambda\psi\rangle$. Hence

$$P(\Omega) = \int_\Omega d\langle\psi|E_\lambda\psi\rangle.$$

If we choose the open interval $\Omega :=]-\infty, \lambda[$, then we obtain the so-called probability distribution function

$$F(\lambda) := P(]-\infty, \lambda[) = \sum_{\lambda_j<\lambda} |\langle\varphi_j|\psi\rangle|^2.$$

The function F is a step function which jumps at the points $\lambda_1, \ldots, \lambda_N$. Moreover, the function F is continuous from the left. According to the rules of the theory of probability, the mean value of the observable A is given by

$$\bar{A} = \int_{\mathbb{R}} \lambda dF(\lambda).$$

Hence

$$\bar{A} = \sum_{j=1}^{N} \lambda_j |\langle \varphi_j | \psi \rangle|^2 = \sum_{j=1}^{N} \lambda_j \langle \psi | \varphi_j \rangle \langle \varphi_j | \psi \rangle = \langle \psi | A\psi \rangle$$

which coincides with the definition of the mean value given by formula (7.13) on page 350.

7.15 Transition Probabilities, *S*-Matrix, and Unitary Operators

In elementary particle physics, the S-matrix plays a fundamental role for describing scattering processes. In what follows, we want to describe the basic ideas of this fundamental concept in terms of a complex finite-dimensional Hilbert space X which has the dimension $N = 1, 2, \ldots$ Suppose that the two systems

$$\varphi_1, \ldots \varphi_N \quad \text{and} \quad \psi_1, \ldots, \psi_N$$

form an orthonormal basis of the Hilbert space X. By the Dirac calculus, we get the following key relation:

$$\sum_{r=1}^{N} \langle \psi_j | \varphi_r \rangle \langle \varphi_r | \psi_k \rangle = \langle \psi_j | \psi_k \rangle = \delta_{jk}, \qquad j, k = 1, \ldots, N. \tag{7.26}$$

S-matrix. By definition, the transition amplitude from the state ψ_k to the state φ_j reads as

$$s_{jk} := \langle \varphi_j | \psi_k \rangle, \qquad j, k = 1, \ldots, N. \tag{7.27}$$

The idea is to regard this system (s_{jk}) of transition amplitudes as the elements of a complex $(N \times N)$-matrix called the S-matrix. From (7.26) we obtain immediately the following crucial property.

Proposition 7.13 *The S-matrix (s_{jk}) is unitary. Explicitly,*

$$\sum_{r=1}^{N} s_{rj}^{\dagger} s_{rk} = \delta_{jk}, \qquad j, k = 1, \ldots, N.$$

Transition probabilities. Letting $j = k$, we get

$$\sum_{r=1}^{N} |s_{rk}|^2 = 1, \qquad k = 1, \ldots, N.$$

This allows us the following interpretation. For given fixed state ψ_k with $k = 1, \ldots, N$, the number

$$\boxed{|s_{rk}|^2 = |\langle \varphi_r | \psi_k \rangle|^2, \qquad r = 1, \ldots, N}$$

is the transition probability from the state ψ_k to the state φ_r. We now want to show that precisely unitary operators are related to systems of transition probabilities.

From transition probabilities to unitary operators. Let $\varphi_1, \ldots, \varphi_N$ and $\psi_1, \ldots, \psi_N$ be chosen as above. Define a linear operator $S : X \to X$ by setting

$$S\varphi_k := \psi_k, \qquad k = 1, \ldots, N.$$

Then, the operator S is unitary, that is,

$$\boxed{\langle S\varphi | S\psi \rangle = \langle \varphi | \psi \rangle \qquad \text{for all} \quad \varphi, \psi \in X.}$$

Let us prove this. Note first that

$$\langle S\varphi_j | S\varphi_l \rangle = \langle \psi_j | \psi_l \rangle = \delta_{jl} \qquad j, l = 1, \ldots, N.$$

Thus, it follows from $\varphi = \sum_{k=1}^N \langle \varphi_k | \varphi \rangle \varphi_k$ and $\psi = \sum_{l=1}^N \langle \varphi_l | \psi \rangle \varphi_l$ that $\langle S\varphi | S\psi \rangle$ is equal to

$$\sum_{k,l=1}^N \langle \varphi_k | \varphi \rangle^\dagger \langle \varphi_l | \psi \rangle \langle S\varphi_k | S\varphi_l \rangle = \sum_{k=1}^N \langle \varphi | \varphi_k \rangle \langle \varphi_k | \psi \rangle = \langle \varphi | \psi \rangle.$$

From unitary operators to transition probabilities. Conversely, suppose that we are given a linear unitary operator $S : X \to X$. Let $\varphi_1, \ldots, \varphi_N$ be an orthonormal basis of X. Finally, set

$$\psi_k := S\varphi_k, \qquad k = 1, \ldots, N.$$

Then, the system $\psi_1, \ldots, \psi_N$ is also an orthonormal basis. In fact,

$$\langle \psi_k | \psi_l \rangle = \langle S\varphi_k | S\varphi_l \rangle = \langle \varphi_k | \varphi_l \rangle = \delta_{jk}.$$

The matrix elements of the operator S are given by

$$s_{jk} := \langle \varphi_j | S\varphi_k \rangle, \qquad j, k = 1, \ldots, N.$$

Hence $s_{jk} = \langle \varphi_j | \psi_k \rangle$, showing us that s_{jk} represents a transition probability, as in (7.27) above.

Physical interpretation. Suppose that the following situation is met.

- The given quantum system is in one of the states $\varphi_1, \ldots, \varphi_N$ at the initial time t_{in}.
- The initial state φ_k passes over to the state $S\varphi_k$ at the final time t_{out}.

Then, the real number

$$|s_{jk}|^2 = |\langle \varphi_j | S \varphi_k \rangle|^2$$

represents the transition probability of the quantum system from the initial state φ_k to the final state φ_j during the time interval $[t_{\rm in}, t_{\rm out}]$. It is crucial that the operator S is unitary. Otherwise, an interpretation in terms of transition probabilities is not possible.

Typically, the original S-matrix constructed in the Standard Model in particle physics is not unitary. However, this shortcoming can be fixed by introducing ghosts and antighosts (see Chap. 16).

7.16 The Magic Formulas for the Green's Operator

Let $H : X \to X$ be an observable on the complex Hilbert space X of finite dimension $N = 1, 2, , \ldots$ The three magic formulas read as follows: [24]

(i) Fourier representation formula in terms of energy eigenstates:

$$\boxed{G(E) = \sum_{j=1}^{N} \frac{|E_j\rangle\langle E_j|}{E_j - E}, \qquad E \in \mathbb{C} \setminus \sigma(H).} \tag{7.28}$$

(ii) The Laplace transform of the retarded propagator in the energy space:

$$G(E) = \frac{\mathrm{i}}{\hbar} \int_0^\infty P^+(t, t_0)\, \mathrm{e}^{\mathrm{i}(t-t_0)E/\hbar}\, dt, \qquad \Im(E) > 0. \tag{7.29}$$

The retarded propagator P^+ is also called the Feynman propagator.

(iii) The inverse Laplace transform:

$$\boxed{P^+(t, t_0) = \frac{1}{2\pi\mathrm{i}} \lim_{R \to +\infty} \int_{-R+\mathrm{i}\gamma}^{R+\mathrm{i}\gamma} \mathrm{e}^{-\mathrm{i}(t-t_0)E/\hbar}\, G(E)\, dE.} \tag{7.30}$$

This is true for all real times $t \neq t_0$. For fixed parameter $\gamma > 0$, we integrate over the line $\{E \in \mathbb{C} : \ \Im(E) = \gamma\}$ parallel to the real axis.

These formulas reflect the duality between time t and energy E which is crucial for quantum field theory. We will use the following terminology:

- energy operator H (Hamiltonian);
- energy eigenvalues $E_1, \ldots E_N$ with $H|E_j\rangle = E_j|E_j\rangle$ for $j = 1, \ldots, N$;
- energy eigenstates $|E_1\rangle, \ldots, |E_N\rangle$;

[24] Fourier (1768–1830) and Laplace (1749–1827); John von Neumann (1903–1957) founded far-reaching generalizations of these formulas in the framework of his operator calculus for self-adjoint operators in infinite-dimensional Hilbert spaces.

- energy spectrum $\sigma(H) := \{E_1, \ldots, E_N\}$;
- Green's operator $G(E) := (H - EI)^{-1}$ defined for complex energies E with $E \notin \sigma(H)$;
- Green's matrix $G_{jk}(E) := \langle j|G(E)|k\rangle$ for $j, k = 1, \ldots, N$ with respect to the complete orthonormal system $|1\rangle, \ldots, |N\rangle$;
- propagator $P(t, t_0) := \mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}$ for all real times, $t \in \mathbb{R}$, and fixed real initial time t_0.
- Retarded propagator (or Feynman propagator)

$$P^+(t, t_0) := \begin{cases} P(t, t_0), & \text{if } t \geq t_0 \\ 0 & \text{if } t < t_0. \end{cases}$$

In terms of the Heaviside function, $P^+(t, t_0) = \theta(t - t_0)P(t, t_0)$.[25] The retarded propagator is switched on at time t_0. The eigenstates form a complete orthonormal system in the Hilbert space X. In the physics textbooks, the retarded propagator (resp. the Green's operator) is closely related to instationary (resp. stationary) scattering theory. For physics it is crucial that

> *The singularities of the Green's operator in* (7.28) *are precisely the points in the energy spectrum.*

Physicists developed methods for computing the Green's operator and the retarded propagator. For example,

> *Feynman's approach to quantum physics is based on representing the retarded propagator by a path integral.*

In what follows, we will discuss this. We start with the instationary Schrödinger equation

$$\boxed{\mathrm{i}\hbar\dot{\psi}(t) = H\psi(t), \qquad t \in \mathbb{R}.} \tag{7.31}$$

Here, the dot denotes the time derivative.

7.16.1 Non-Resonance and Resonance

Motivated by Fourier's approach to eigenoscillations of the vibrating string, we are looking for special solutions of the instationary Schrödinger equation (7.31) which have the following form

$$\psi(t) = \mathrm{e}^{-\mathrm{i}Et/\hbar}\varphi, \qquad t \in \mathbb{R}.$$

This leads to the stationary Schrödinger equation

$$\boxed{H\varphi = E\varphi, \qquad \varphi \in X.} \tag{7.32}$$

[25] By definition, $\theta(t) := 1$ if $t \geq 0$, and $\theta(t) := 0$ if $t < 0$.

In mathematics, this is called an eigenvalue problem. The corresponding inhomogeneous problem looks like

$$H\varphi - E\varphi = \chi, \qquad \varphi \in X. \tag{7.33}$$

In terms of physics, the right-hand side χ describes an external force. In Schwinger's terminology, χ is called a source term.

Basic ideas. There is a duality between the two problems (7.33) and (7.32). To explain this, fix the complex number E. We are interested in the operator

$$G(E) := (H - EI)^{-1}$$

called the Green's operator in physics.

(N) Non-resonance case: Suppose $E \notin \sigma(H)$. Then, the Green's operator

$$G(E) : X \to X$$

exists.[26] For each given $\chi \in X$, the inhomogeneous equation (7.33) has the unique solution $\varphi = G(E)\chi$. Explicitly, the Green's operator looks like

$$\boxed{G(E) = \sum_{j=1}^{N} \frac{|E_j\rangle\langle E_j|}{E_j - E}.}$$

In terms of physics, the energy parameter E is different from the energies $|E_j\rangle, \ldots, |E_N\rangle$ of the eigenoscillations of the system.

(R) Resonance case: Suppose $E \in \sigma(H)$. Without any loss of generality, we assume that $E = E_1$. Moreover, assume that the energy eigenvalue E_1 has the multiplicity m, that is, $E_1 = E_2 = \ldots = E_m$ and $E_j \neq E_1$ if $j > m$.[27] Then, the inhomogeneous problem (7.33) has a solution iff the resonance condition

$$\langle E_j|\chi\rangle = 0, \qquad j = m+1, \ldots, N \tag{7.34}$$

is satisfied. In this case, the general solution of (7.33) looks like

$$\varphi = G_{\mathrm{reg}}(E_1)\chi + c_1|E_1\rangle + \ldots + c_m|E_m\rangle$$

where $c_1, \ldots, c_m$ are arbitrary complex numbers, and the regularized Green's operator is given by

$$\boxed{G_{\mathrm{reg}}(E_1) := \sum_{j=1}^{m} |E_j\rangle\langle E_j| + \sum_{j=m+1}^{N} \frac{|E_j\rangle\langle E_j|}{E_j - E_1}.}$$

[26] In mathematics, the operator $-G(E) = (EI - H)^{-1}$ is called the resolvent of the Hamiltonian H at the point E. The functional analytic theory of resolvents can be found in Lax (2002).

[27] Suppose that $1 \leq m < N$ in order to exclude the trivial case $H = E_1 I$.

In terms of physics, the resonance condition (7.34) tells us that the external force χ has to be compatible with the eigenoscillations of the system. If this compatibility is violated, then the physical process is impossible. Furthermore, the external force χ never uniquely determines the state of the system. This complicates the physics. In order to obtain a unique state, one has to add a side condition (see Theorem 7.15 below).

In Sect. 8.1 we will show how the difficulties of the resonance case are related to renormalization.

In quantum field theory, the Epstein–Glaser approach to renormalization displays most clearly that there are additional degrees of freedom which have to be determined by side conditions coming from physics. The Epstein–Glaser approach is based on the theory of generalized functions. This will be investigated in Volume II.

Proof. We will use the Fourier method, that is, we will determine the solution by computing its Fourier coefficents.

Ad (N). By (7.33), $\langle E_j|(H-E)\varphi\rangle = \langle E_j|\chi\rangle$ for all j. If $H\varphi_j = E_j\varphi_j$, then $\langle\varphi_j|H\varphi\rangle = \langle H\varphi_j|\varphi\rangle = E_j\langle\varphi_j|\varphi\rangle$. Hence

$$\langle E_j|H\varphi\rangle = E_j\langle E_j|\varphi\rangle.$$

Thus, we obtain

$$(E_j - E)\langle E_j|\varphi\rangle = \langle E_j|\chi\rangle, \qquad j = 1,\dots,N. \tag{7.35}$$

If $E \neq E_j$ for all j, then

$$|\varphi\rangle = \sum_{j=1}^{N} |E_j\rangle\langle E_j|\varphi\rangle = \sum_{j=1}^{N} \frac{|E_j\rangle\langle E_j|\chi\rangle}{E_j - E} = G(E)|\chi\rangle.$$

Ad (R). If $E = E_1$, then $\langle E_j|\chi\rangle = 0$ for $j = m+1,\dots,N$, by (7.35). This yields the special solution

$$|\varphi\rangle = \sum_{j=m+1}^{N} \frac{|E_j\rangle\langle E_j|\chi\rangle}{E_j - E}.$$

Note that the resonance condition (7.35) does not restrict the Fourier coefficients $\langle E_j|\varphi\rangle$ for $j = 1,\dots,m$. Therefore, the general solution is obtained by adding the linear combination $\sum_{j=1}^{m}(c_j+1)|E_j\rangle$. □

The orthogonal projector P onto the eigenspace of the eigenvalue E_1. Let us consider the resonance case (R) above. Define the operator

$$\boxed{P := \sum_{j=1}^{m} |E_j\rangle\langle E_j|.}$$

It is our goal to describe the regularized Green's operator $G_{\rm reg}(E_1)$ in terms of the operator P. We will show that $G_{\rm reg}(E_1) = (H + P - E_1 I)^{-1}$.

Let us first study typical properties of the operator P. For all $\varphi \in X$, we have $\varphi = \sum_{j=1}^{N} |E_j\rangle\langle E_j|\varphi\rangle$, and

$$P\varphi = \sum_{j=1}^{m} |E_j\rangle\langle E_j|\varphi\rangle.$$

This implies the decomposition

$$\varphi = P\varphi + (I - P)\varphi$$

where $P\varphi$ is orthogonal to $(I - P)\varphi$. We write this as $X = PX \oplus (I - P)X$. The space PX is identical with the eigenspace of the Hamiltonian H with respect to the eigenvalue E_1. We call P the orthogonal projector onto the eigenspace PX.

Proposition 7.14 (i) *$P\varphi = \varphi$ iff $H\varphi = E_1\varphi$.*
(ii) *$P^2 = P$ and $P^\dagger = P$.*

Proof. Ad (i). Note that $P\varphi = \varphi$ iff $\langle E_j|\varphi\rangle = 0$ for $j = m+1, \ldots, N$.

Ad (ii). Since $P|E_j\rangle = |E_j\rangle$ for $j = 1, \ldots, m$, we get $P^2 = P$. Moreover, for all $\varphi, \psi \in X$,

$$\langle\varphi|P\psi\rangle = \sum_{j=1}^{N} \langle\varphi|E_j\rangle\langle E_j|\psi\rangle.$$

Similarly, $\langle\psi|P\varphi\rangle = \sum_{j=1}^{N} \langle\psi|E_j\rangle\langle E_j|\varphi\rangle$. Hence

$$\langle P\varphi|\psi\rangle = \langle\psi|P\varphi\rangle^\dagger = \sum_{j=1}^{N} \langle E_j|\psi\rangle\langle\varphi|E_j\rangle = \langle\varphi|P\psi\rangle.$$

Therefore, $P^\dagger = P$. □

Let us now use the operator P in order to study the operator equation

$$\boxed{H\varphi - E_1\varphi = \chi, \qquad \varphi \in X} \tag{7.36}$$

in the resonance case (R) above. We are given $\chi \in X$. Reformulating (R), we obtain the following result.

Theorem 7.15 *Equation* (7.36) *has a solution iff $P\chi = 0$. Assume that $P\chi = 0$. Then, $\varphi = G_{\rm reg}\chi$ is the unique solution of equation* (7.36) *along with the side condition $P\varphi = 0$.*

Consequently, the linear operator $G_{\rm reg} : (I-P)X \to (I-P)X$ is invertible.

The pseudo-resolvent in the resonance case. Consider now the modified operator equation

$$\boxed{H\varphi + P\varphi - E_1\varphi = \chi, \qquad \varphi \in X} \tag{7.37}$$

in the resonance case (R) above.

Theorem 7.16 *For each $\chi \in X$, equation* (7.37) *has the unique solution*

$$\varphi = (H + P - E_1 I)^{-1}\chi = G_{\rm reg}(E_1)\chi.$$

Therefore, the operator $-G_{\rm reg}(E_1)$ is called the pseudo-resolvent of the Hamiltonian H at the eigenvalue E_1.

Proof. (I) Uniqueness. Let φ_1 and φ_2 be solutions of (7.37). Set $\varphi := \varphi_1 - \varphi_2$. Then

$$H\varphi - E_1\varphi = -P\varphi.$$

By Theorem 7.15, $P^2\varphi = 0$. Hence $P\varphi = 0$. Finally, Theorem 7.15 tells us that $\varphi = 0$.

(II) Existence. Set

$$\varphi := G_{\rm reg}(E_1)\chi = \sum_{j=1}^{m} |E_j\rangle\langle E_j|\chi\rangle + \sum_{j=m+1}^{N} \frac{|E_j\rangle\langle E_j|\chi\rangle}{E_j - E_1}.$$

Then $(H + P - E_1 I)\varphi = P\varphi + (H - E_1 I)\varphi$. This is equal to

$$\sum_{j=1}^{m} |E_j\rangle\langle E_j|\chi\rangle + \sum_{j=m+1}^{N} |E_j\rangle\langle E_j|\chi\rangle = \chi.$$

Therefore, $(H + P - E_1 I)\varphi = \chi$. □

In Sect. 8.1, we will show how to use Theorem 7.16 in the nonlinear resonance case which is related to the concept of renormalization.

7.16.2 Causality and the Laplace Transform

Damped oscillations and complex energy. Let us introduce the complex energy

$$\boxed{E = \hbar(\omega - \Gamma \mathrm{i})}$$

where ω and $\Gamma > 0$ are real numbers. For a fixed nonzero complex number x_0, the function

$$x(t) = x_0 \mathrm{e}^{-\mathrm{i}Et/\hbar} = x_0 \mathrm{e}^{-\mathrm{i}\omega t}\mathrm{e}^{-\Gamma t}, \qquad t \in \mathbb{R}$$

describes a damped wave of angular frequency ω. It is reasonable to call the parameter

$$\boxed{\Delta t := \frac{1}{\Gamma}}$$

the mean lifetime of the damped oscillation. Since the function $x = x(t)$ increases as $t \to -\infty$, the integral

$$\int_{-\infty}^{\infty} x(t)dt = x_0 \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\omega t}\mathrm{e}^{-\gamma t}dt \tag{7.38}$$

does not exist. The situation changes completely if we truncate the function $x = x(t)$ by introducing the Heaviside function θ. This yields the function

$$y(t) := \theta(t)x(t) = x_0\theta(t)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{e}^{-\gamma t}, \qquad t \in \mathbb{R}.$$

Explicitly, $y(t) = x(t)$ if $t \geq 0$, and $y(t) = 0$ if $t < 0$. Set $x_0 := 1$. For the integral,

$$\int_{-\infty}^{\infty} y(t)dt = \int_0^{\infty} \mathrm{e}^{(-\mathrm{i}\omega-\gamma)t}\, dt = -\lim_{T\to+\infty} \frac{\mathrm{e}^{(-\mathrm{i}\omega-\gamma)t}}{\mathrm{i}\omega+\gamma}\Big|_0^T = \frac{1}{\mathrm{i}\omega+\gamma}.$$

Note that $\mathrm{e}^{-\gamma T} \to 0$ as $T \to +\infty$.

The role of negative energies. In classical mechanics, the energy E is always nonnegative. However, in quantum physics, negative energies are allowed. In 1928 Dirac predicted that negative energies correspond to positive energies of antiparticles.

The importance of the retarded propagator. We now want to investigate the Laplace transform of the retarded propagator. By definition, the retarded operator reads as

$$P^+(t,t_0) := \theta(t-t_0)\mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}, \qquad t,t_0 \in \mathbb{R}.$$

By the operator calculus formula (7.19) on page 358,

$$\boxed{P^+(t,t_0) = \theta(t-t_0)\sum_{j=1}^{N} \mathrm{e}^{-\mathrm{i}(t-t_0)E_j/\hbar}|E_j\rangle\langle E_j|.} \tag{7.39}$$

Here, $|E_1\rangle, \ldots, |E_N\rangle$ represents a complete orthonormal system of eigenvectors of the Hamiltonian H corresponding to the eigenvalues $E_1, \ldots, E_N$, respectively. Observe that it is crucial to use the retarded propagator $P^+(t,t_0)$ instead of the propagator $P(t,t_0) = \mathrm{e}^{-\mathrm{i}(t-t_0)E/\hbar}$. Indeed, the Fourier transform

$$\int_{-\infty}^{\infty} P(t,t_0)\mathrm{e}^{\mathrm{i}E(t-t_0)/\hbar}dE = \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}\, \mathrm{e}^{\mathrm{i}E(t-t_0)/\hbar}dE$$

does not exist. To check this, it is sufficient to consider the corresponding integrals for the eigenvalues, by (7.39). However, if we replace the operator

H by the eigenvalue E_j, then we get a divergent integral. As the following theorem shows, the situation changes completely if we consider the integral

$$\int_{-\infty}^{\infty} P^+(t,t_0)\mathrm{e}^{\mathrm{i}E(t-t_0)/\hbar}dE = \int_0^{\infty} \mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}\ \mathrm{e}^{\mathrm{i}E(t-t_0)/\hbar}dE,$$

and if we use the complex energy $E = \hbar(\omega + \gamma\mathrm{i})$ with $\gamma > 0$. The passage from the propagator $P(t,t_0)$ to the retarded propagator $P^+(t,t_0)$ means that the process is switched on at time t_0, and we study the process in the future, $t \geq t_0$. Therefore, the retarded propagator reflects causality.

Theorem 7.17 *The retarded propagator P^+ is related to the Green's operator G by the magic Laplace transform* (7.29) *and the inverse Laplace transform* (7.30).

Proof. To simplify notation, we set $\hbar := 1$ and $t_0 := 0$.

Ad (7.29). Let $\Im(E) > 0$. Then

$$\frac{1}{E_j - E} = \mathrm{i}\int_0^{\infty} \mathrm{e}^{-\mathrm{i}tE_j}\ \mathrm{e}^{\mathrm{i}Et}\ dt.$$

By (7.39), we get the claim (7.29).

Ad (7.30). We will use Cauchy's residue method. Using a translation, we have to show that

$$\theta(t)\mathrm{e}^{-\mathrm{i}tH} = \frac{1}{2\pi\mathrm{i}} \lim_{R\to+\infty} \int_{-R}^{R} \mathrm{e}^{-\mathrm{i}t(E+\mathrm{i}\gamma)}\ G(E+\mathrm{i}\gamma)dE.$$

By (7.39), it is enough to prove this relation for the eigenvalues of H. Thus, we have to show that

$$\theta(t)\mathrm{e}^{-\mathrm{i}tE_j} = \frac{1}{2\pi\mathrm{i}} \lim_{R\to+\infty} \int_{-R}^{R} \frac{\mathrm{e}^{-\mathrm{i}t(E+\mathrm{i}\gamma)}}{E_j - (E+\mathrm{i}\gamma)}\ dE \tag{7.40}$$

for all real times $t \neq 0$ and fixed parameter $\gamma > 0$.

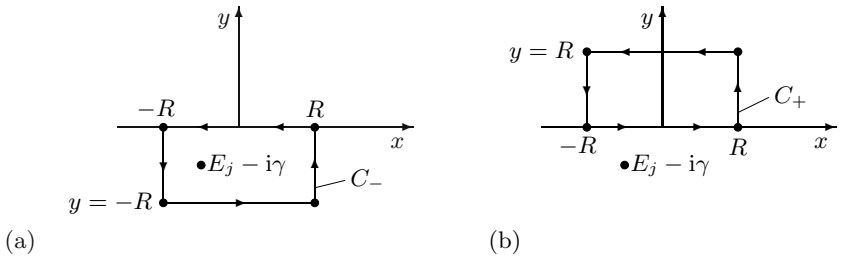


Fig. 7.1. Cauchy's residue method

Case 1: $t > 0$. Choose the curve C_- as pictured in Fig. 7.1(a). By Cauchy's residue method, we get

$$\frac{1}{2\pi\mathrm{i}} \int_{C_-} \frac{\mathrm{e}^{-\mathrm{i}Et}}{E - (E_j - \mathrm{i}\gamma)}\, dE = \mathrm{e}^{-\mathrm{i}t(E_j - \mathrm{i}\gamma)},$$

noting that the integrand has a first-order pole at the point $E = E_j - \mathrm{i}\gamma$. Letting $E = x + \mathrm{i}y$,

$$\frac{\mathrm{e}^{-\mathrm{i}Et}}{E - (E_j - \mathrm{i}\gamma)} = \frac{\mathrm{e}^{-\mathrm{i}xt}\, \mathrm{e}^{yt}}{x - E_j + \mathrm{i}(y + \gamma)}.$$

For fixed time $t > 0$, the function $y \mapsto \mathrm{e}^{yt}$ decays fast as $y \to -\infty$. Hence

$$\lim_{R \to +\infty} \int_{C_- \setminus [-R,R]} \frac{\mathrm{e}^{-\mathrm{i}Et}}{E - (E_j - \mathrm{i}\gamma)}\, dE = 0.$$

Observing the orientation of the curve C_-,

$$\lim_{R \to +\infty} \frac{1}{2\pi\mathrm{i}} \int_{-R}^{R} \frac{\mathrm{e}^{-\mathrm{i}Et}}{E - (E_j - \mathrm{i}\gamma)}\, dE = -\mathrm{e}^{-\mathrm{i}(E_j - \mathrm{i}\gamma)t}.$$

This yields the claim (7.40).

Case 2: $t < 0$. Now use Fig. 7.1(b). Noting that the curve C_+ does not surround the singularity $E_j - \mathrm{i}\gamma$ of the integrand,

$$\int_{C_+} \frac{\mathrm{e}^{-\mathrm{i}Et}}{E - (E_j - \mathrm{i}\gamma)}\, dE = 0.$$

For fixed time $t < 0$, the function $y \mapsto \mathrm{e}^{ty}$ decays rapidly as $y \mapsto +\infty$. Hence

$$\lim_{R \to +\infty} \int_{-R}^{R} \frac{\mathrm{e}^{-\mathrm{i}Et}}{E - (E_j - \mathrm{i}\gamma)}\, dE = 0.$$

□

The lifetime of particles. In particle accelerators, physicists frequently observe particles which have only a finite lifetime. To mathematically model this phenomenon, we have to modify the instationary Schrödinger equation,

$$\boxed{\mathrm{i}\hbar\dot{\psi}(t) = (H - \mathrm{i}V)\psi(t), \qquad t \in \mathbb{R}}$$

by adding the complex perturbation $-\mathrm{i}V$ to the self-adjoint Hamiltonian H. Making the Fourier ansatz $\psi(t) = \mathrm{e}^{-\mathrm{i}Et/\hbar}\varphi$, we obtain the stationary Schrödinger equation

$$(H - \mathrm{i}V)\varphi = E\varphi.$$

Assume that this equation has the complex eigenvalues

$$E_j = \hbar(\omega_j - \Gamma_j \mathrm{i}), \qquad j = 1, \ldots, N$$

with the corresponding eigenvectors $|E_1\rangle, \ldots, |E_N\rangle$. Suppose that $\Gamma_j > 0$ for all j. Then, the state

$$\psi(t) = \mathrm{e}^{-\mathrm{i}E_j t/\hbar}|E_j\rangle = \mathrm{e}^{-\mathrm{i}\omega_j t}\mathrm{e}^{-\Gamma_j t}|E_j\rangle, \qquad t \in \mathbb{R}$$

can be regarded as a state which has the energy $\hbar\omega_j$ and the mean lifetime $\Delta t = 1/\Gamma_j$. From the functional analytic point of view, this important phenomenon is studied in the monograph by Hislop and Sigal (1996).

7.17 The Magic Dyson Formula for the Retarded Propagator

We want to study the initial value problem

$$\boxed{\mathrm{i}\hbar\dot{\psi}(t) = H(t)\psi(t), \qquad t \geq t_0, \quad \psi(t_0) = \psi_0.} \tag{7.41}$$

We make the following assumptions:

(A1) Let X be a finite-dimensional complex Hilbert space. We are given the initial state $\psi_0 \in X$ at the initial time t_0.
(A2) The operator $H(t) : X \to X$ is linear for each time $t \in \mathbb{R}$.
(A3) The map $t \mapsto H(t)$ is smooth from $\mathbb{R}$ to the space $L(X, X)$.[28]
(A4) The operator $H_0 : X \to X$ is linear, and it does not depend on time t.
(A5) The function $f : \mathbb{R} \to X$ is smooth.

The assumptions (A4), (A5) refer to problems considered below. In the special case where $X = \mathbb{C}^N$, all of the operators and states are matrices. By choosing an orthonormal basis of the Hilbert space X and passing to matrices, our problem can be always reduced to the space $\mathbb{C}^N$. If we pass to matrix elements, then problem (7.41) represents a linear system of ordinary differential equations. The unique solution of (7.41) can be written as

$$\psi(t) = P^+(t, t_0)\psi_0, \qquad t \geq t_0.$$

It is our goal to prove the magic Dyson formula

$$\boxed{P^+(t, t_0) = \mathcal{T}\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}\int_{t_0}^{t} H(\tau)d\tau}, \qquad t \geq t_0} \tag{7.42}$$

for the retarded propagator (Feynman propagator) P^+ and to derive the Dyson series. Here, we use the so-called time-ordering operator $\mathcal{T}$ which

[28] This means that the matrix elements of $H(t)$ are smooth functions of time t. Similarly, we assume that the Fourier coefficients of $f(t)$ are smooth functions with respect to time t. Naturally enough, these definitions of smoothness do not depend on the choice of the basis vectors.

describes the causal structure of time-dependent processes. Explicitly, the chronological operator $\mathcal{T}$ organizes the factors in such a way that time is increasing from right to left:

$$\mathcal{T}\{H(\tau_1)H(\tau_2)\} := \begin{cases} H(\tau_1)H(\tau_2) & \text{if} \quad \tau_1 \geq \tau_2, \\ H(\tau_2)H(\tau_1) & \text{if} \quad \tau_2 \geq \tau_1. \end{cases}$$

More generally,

$$\mathcal{T}\{H(\tau_1)H(\tau_2)\cdots H(\tau_n)\} := H(\tau_{1'})H(\tau_{2'})\cdots H(\tau_{n'}). \tag{7.43}$$

Here, $\tau_{1'}, \ldots, \tau_{n'}$ is a permutation of $\tau_1, \ldots, \tau_n$ such that $\tau_{1'} \geq \tau_{2'} \geq \ldots \geq \tau_{n'}$. The Dyson formula (7.42) is to be understood mnemonically. Using the power series expansion of the exponential function, for all $t \geq t_0$, we get

$$P^+(t,t_0) = \sum_{k=0}^{\infty} \frac{(-\mathrm{i})^k}{\hbar^k k!} \int_{t_0}^{t} \cdots \int_{t_0}^{t} \mathcal{T}(H(\tau_1)\cdots H(\tau_k))d\tau_1 \cdots d\tau_k.$$

Thus, for all $t \geq t_0$,

$$P^+(t,t_0) = I - \frac{\mathrm{i}}{\hbar}\int_{t_0}^{t} H(\tau)\, d\tau - \frac{1}{2!}\int_{t_0}^{t}\int_{t_0}^{t} \mathcal{T}(H(\tau_1)H(\tau_2))d\tau_1 d\tau_2 + \ldots$$

This is the famous Dyson series. The convergence of the Dyson series will be proved on page 388. At this point, let us only mention the following two special cases.

- Let $X = \mathbb{C}$. Then, $H(t)$ is a complex number for each time t, and we are looking for a complex-valued function $\psi : \mathbb{R} \to \mathbb{C}$. The Dyson formula now corresponds to the classical solution

  $$\psi(t) = \psi_0 \mathrm{e}^{-\frac{\mathrm{i}}{\hbar}\int_{t_0}^{t} H(\tau)d\tau}$$

  of equation (7.41). This formula was known to Euler (1707–1783) and Lagrange (1736–1813).
- Suppose that $H(t) = H_0$ for all times t. For (7.41), we then get the solution formula

  $$\psi(t) = \mathrm{e}^{-\mathrm{i}tH_0/\hbar}\psi_0.$$

For understanding quantum field theory, observe that

> *The infinite-dimensional version of the Dyson series lies at the heart of the operator approach to quantum field theory.*

This will be considered in Chap. 15. Applications to quantum electrodynamics will be studied in Volume II. Alternatively, the retarded propagator $P^+(t,t_0)$ can be represented by a path integral. This will be investigated in Sect. 7.21.1 on page 417. It is now our goal to prove the convergence of the Dyson series. To begin with, we need some preparatory results.

7.17.1 Lagrange's Variation of the Parameter

The inhomogeneous linear problem. Let us start our approach to the Dyson series by considering the initial-value problem

$$\boxed{\mathrm{i}\hbar\dot{\psi}(t) = H_0\psi(t) + f(t), \qquad t \in \mathbb{R}, \qquad \psi(t_0) = \psi_0.} \tag{7.44}$$

We are looking for a solution $\psi : \mathbb{R} \to X$. By the classical theory of linear differential equations, the initial-value problem (7.44) has a unique solution which exists for all times $t \in \mathbb{R}$. Explicitly, this solution is given by the formula

$$\boxed{\psi(t) = P(t, t_0)\psi_0 + \frac{1}{\mathrm{i}\hbar}\int_{t_0}^{t} P(t, s)f(s)ds} \tag{7.45}$$

for all times $t \in \mathbb{R}$ with the propagator

$$P(t, s) := \mathrm{e}^{-\mathrm{i}H_0(t-s)/\hbar}.$$

Consider the special case where $X = \mathbb{C}$. Then, differentiation with respect to time t yields

$$\mathrm{i}\hbar\dot{\psi}(t) = \mathrm{i}\hbar P_t(t, t_0)\psi_0 + P(t, t)f(t) + \int_{t_0}^{t} P_t(t, s)f(s)ds.$$

The propagator satisfies the initial-value problem

$$\mathrm{i}\hbar P_t(t, t_0) = H_0 P(t, t_0), \qquad t \in \mathbb{R}, \qquad P(t_0, t_0) = 1.$$

In addition, $P(t, t) = 1$ for all $t \in \mathbb{R}$. Hence

$$\mathrm{i}\hbar\dot{\psi}(t) = H_0 P(t, t_0)\psi_0 + f(t) + H_0\int_{t_0}^{t} P(t, s)f(s)ds = H_0\psi(t) + f(t).$$

The same argument applies to all finite-dimensional Hilbert spaces X.

Variation of the parameter. Let us explain how the basic formula (7.45) is related to the method of the variation of parameters. Again set $X = \mathbb{C}$. Then H_0 is a complex number. The differential equation

$$\mathrm{i}\hbar\dot{\psi}(t) = H_0\psi(t)$$

with vanishing external influence, $f(t) \equiv 0$, has the general solution

$$\psi(t) = P(t, t_0)C, \qquad t \in \mathbb{R}$$

where C is an arbitrary constant. Following Lagrange, we make the ansatz

$$\psi(t) = P(t, t_0)C(t), \qquad t \in \mathbb{R}$$

for getting the solution of the initial-value problem

$$\mathrm{i}\hbar\dot{\psi}(t) = H_0\psi(t) + f(t), \qquad t \in \mathbb{R}, \qquad \psi(t_0) = C(t_0).$$

From the physical point of view, Lagrange's idea was to describe the influence of the external force f on the physical state ψ by a variation of the constant C in time. It remains to determine the function $C = C(t)$. In fact, differentiation with respect to time yields

$$\dot{\psi}(t) = P_t(t, t_0)C(t) + P(t, t_0)\dot{C}(t).$$

Hence

$$\mathrm{i}\hbar\dot{\psi}(t) = H_0P(t, t_0)C(t) + \mathrm{i}\hbar P(t, t_0)\dot{C}(t) = H_0\psi(t) + \mathrm{i}\hbar P(t, t_0)\dot{C}(t).$$

Therefore,

$$\mathrm{i}\hbar P(t, t_0)\dot{C}(t) = f(t).$$

Observe that we have the crucial propagator equation

$$\boxed{P(t, s)P(s, \tau) = P(t, \tau), \qquad P(t, t) = 1, \qquad t, s, \tau \in \mathbb{R}.}$$

Hence $P(t, s) = P(s, t)^{-1}$. Thus, $\mathrm{i}\hbar\dot{C}(t) = P(t_0, t)f(t)$. This differential equation has the solution

$$\mathrm{i}\hbar C(t) = \mathrm{i}\hbar C(t_0) + \int_{t_0}^{t} P(t_0, s)f(s)ds.$$

Multiplying this by $P(t, t_0)$ and using the propagator equation, we get the desired result

$$\mathrm{i}\hbar\psi(t) = \mathrm{i}\hbar P(t, t_0)C(t_0) + \int_{t_0}^{t} P(t, s)f(s)ds.$$

The propagator. We introduce the retarded propagator

$$\boxed{P^+(t, t_0) := \theta(t - t_0)P(t, t_0), \qquad t \in \mathbb{R}} \tag{7.46}$$

and the advanced propagator

$$P^-(t, t_0) := -\theta(t_0 - t)P(t, t_0), \qquad t \in \mathbb{R}.$$

For the propagator,

$$P(t, t_0) = P^+(t, t_0) - P^-(t, t_0), \qquad t \in \mathbb{R}.$$

7.17.2 Duhamel's Principle

The solution formula from (7.45) tells us that

> *The solution of the special problem with vanishing external influence,* $f(t) \equiv 0$, *knows the propagator* $P(t,s)$, *and hence it knows all about the general solution by superposition.*

This is called Duhamel's principle.[29] More precisely, the general solution ψ of the initial-value problem (7.44) can be obtained as follows.

(H1) General homogeneous problem: Consider first the case where the external influence vanishes, $f(t) \equiv 0$. Then, the initial-value problem

$$\mathrm{i}\hbar\dot{\psi}(t) = H_0\psi(t), \qquad t \in \mathbb{R}, \qquad \psi(t_0) = \psi_0$$

has the unique solution

$$\boxed{\psi(t) = P(t,t_0)\psi_0, \qquad t \in \mathbb{R}.}$$

This determines uniquely the propagator $P(t,t_0)$.

(H2) Special homogeneous problem: Consider the case where the external influence vanishes, $f(t) \equiv 0$, but the initial condition depends on the external influence $f(s)$ at time s. Then, for each given initial time $s \in \mathbb{R}$, the initial-value problem

$$\mathrm{i}\hbar\dot{\psi}(t) = H_0\psi(t), \qquad t \in \mathbb{R}, \qquad \psi(s) = f(s)$$

has the unique solution $\psi(t) = P(t,s)f(s), \ \ t \in \mathbb{R}$.

(I1) Inhomogeneous problem with special initial condition: Consider the case where the external influence is acting, but the system rests at the initial time t_0. Then, the initial-value problem

$$\mathrm{i}\hbar\psi(t) = H_0\psi(t) + f(t), \qquad t \in \mathbb{R}, \qquad \psi(t_0) = 0$$

has the unique solution

$$\boxed{\psi(t) = \int_{t_0}^{t} P(t,s)f(s)ds, \qquad t \in \mathbb{R}}$$

which is obtained as a superposition of the solutions from (H2).

(I2) General inhomogeneous problem: The solution (7.45) of the general initial-value problem (7.44) is obtained by a superposition of the special solutions from (H1) and (I1).

[29] Duhamel (1792–1872)

7.17.3 The Volterra Integral Equation

Let us replace equation (7.44) on page 383 by the more general equation

$$\boxed{\mathrm{i}\hbar\dot{\psi}(t) = H(t)\psi(t) + f(t), \qquad t \in \mathbb{R}, \qquad \psi(t_0) = \psi_0} \tag{7.47}$$

where the Hamiltonian, $H = H(t)$, depends on time t. It turns out that the initial-value problem (7.47) is equivalent to the following Volterra integral equation[30]

$$\boxed{\psi(t) = \psi_0 + \frac{1}{\mathrm{i}\hbar}\int_{t_0}^{t} H(s)\psi(s)ds + F(t), \qquad t \in \mathbb{R}} \tag{7.48}$$

where $F(t) := \frac{1}{\mathrm{i}\hbar}\int_{t_0}^{t} f(s)ds$. In fact, integrating equation (7.47), we get (7.48). Conversely, suppose that the integral equation (7.48) has a continuous solution $\psi : \mathbb{R} \to X$. Then, the function ψ is differentiable, and it is the solution of the differential equation (7.47). The standard method for solving the Volterra integral equation (7.48) is the following iterative method

$$\psi_{n+1}(t) = \psi_0 + \frac{1}{\mathrm{i}\hbar}\int_{t_0}^{t} H(s)\psi_n(s)ds + F(t), \qquad t \in \mathbb{R} \tag{7.49}$$

for all indices $n = 0, 1, 2, \ldots$

Theorem 7.18 *The Volterra integral equation* (7.48) *has a unique continuous solution $\psi : \mathbb{R} \to X$. As $n \to \infty$, the sequence (ψ_n) converges uniformly on each compact interval to the solution ψ.*

Proof. To simplify notation, set $t_0 := 0$ and $\hbar := 1$. Fix $T > 0$. Let $C[-T, T]$ denote the space of all continuous functions $\psi : [-T, T] \to X$. Equipped with the norm

$$||\psi|| := \max_{-T \le t \le T} ||\psi(t)||,$$

the space $C[-T, T]$ becomes a complex Banach space.[31]

(I) Existence on the time interval $[-T, T]$. Set $\mathcal{H} := \max_{-T \le t \le T} ||H(t)||$. Let $t > 0$. Then

$$||\psi_2(t) - \psi_1(t)|| = \Big|\Big| \int_0^t H(s)(\psi_1(s) - \psi_0)ds \Big|\Big| \le t\mathcal{H}\, ||\psi_1 - \psi_0||.$$

Hence

[30] Volterra (1860–1950)

[31] In the special case where $X = \mathbb{C}$, the norm $||\psi(t)||$ coincides with the modulus of the complex number $\psi(t)$.

$$||\psi_3(t) - \psi_2(t)|| = \Big|\Big| \int_0^t H(s)(\psi_2(s) - \psi_1(s))ds \Big|\Big|$$
$$\leq \int_0^t s\mathcal{H}^2 ||\psi_1 - \psi_0|| ds = \frac{t^2\mathcal{H}^2}{2!}\, ||\psi_1 - \psi_0||.$$

Similarly, for $n = 1, 2, \ldots$ we get

$$||\psi_{n+1} - \psi_n|| \leq \frac{(T\mathcal{H})^n}{n!}\, ||\psi_1 - \psi_0||.$$

Since the series $\mathrm{e}^{T\mathcal{H}} = 1 + T\mathcal{H} + \ldots$ is convergent, the triangle inequality tells us that

$$||\psi_{n+m+1} - \psi_n|| \leq \sum_{k=n}^{n+m} ||\psi_{k+1} - \psi_k|| \leq \sum_{k=n}^{n+m} \frac{(T\mathcal{H})^k}{k!} < \varepsilon$$

for all $n \geq n_0(\varepsilon)$ and all $m = 1, 2, \ldots$ Thus, the sequence (ψ_n) is Cauchy, and hence it converges to an element ψ of the Banach space $C[-T, T]$, that is,

$$\lim_{n\to\infty} ||\psi_n - \psi|| = 0.$$

Therefore, the sequence (ψ_n) converges uniformly on the interval $[-T, T]$ to the function ψ. Letting $n \to \infty$, it follows from (7.49) that the function $\psi : [-T, T] \to X$ is a solution of (7.48) on the time interval $[-T, T]$.

(II) Uniqueness. Let ψ and φ be two solutions of (7.48). As in (I), it follows from

$$\psi(t) - \varphi(t) = \frac{1}{\mathrm{i}} \int_0^t H(s)(\psi(s) - \varphi(s))ds$$

that $||\psi - \varphi|| \leq \frac{(T\mathcal{H})^n}{n!} ||\psi - \varphi||$. Letting $n \to \infty$, we see that $\psi = \varphi$ on $[-T, T]$. Since the choice of $T > 0$ is arbitrary, we get uniqueness for all times $t \in \mathbb{R}$.

(III) Continuation. By (I) and (II), there exists a unique solution of (7.48) on each time interval $[-T, T]$. Choosing $T = 1, 2, \ldots$, the corresponding solutions can be continued to a unique solution on the real line. □

The propagator equation. Fix the initial time $t_0 \in \mathbb{R}$. The unique solution $P = P(t, t_0)$ of the initial-value problem

$$\boxed{\mathrm{i}\hbar P_t(t, t_0) = H(t)P(t, t_0), \qquad t \in \mathbb{R}, \qquad P(t_0, t_0) = I} \tag{7.50}$$

is called the propagator of equation (7.47). This designation is motivated by the following fact.

Theorem 7.19 *The function*

$$\psi(t) := P(t, t_0)\psi_0 + \frac{1}{i\hbar} \int_{t_0}^t P(t, s)f(s)ds, \qquad t \in \mathbb{R}$$

is the unique solution of the initial-value problem (7.47).

Proof. (I) Existence. Differentiation with respect to time t yields

$$\mathrm{i}\hbar\dot{\psi}(t) = \mathrm{i}\hbar P_t(t,t_0)\psi_0 + P(t,t)f(t) + \int_{t_0}^{t} P_t(t,s)f(s)ds.$$

Hence

$$\mathrm{i}\hbar\dot{\psi}(t) = H(t)P(t,t_0)\psi_0 + f(t) + \frac{1}{\mathrm{i}\hbar}\int_{t_0}^{t} H(t)P(t,s)f(s)ds.$$

This implies $\mathrm{i}\hbar\dot{\psi}(t) = H(t)\psi(t) + f(t)$.

(II) Uniqueness. This follows from Theorem 7.18. □

7.17.4 The Dyson Series

Let us now study the following Dyson series

$$P^{+}(t,t_0) = I - \frac{\mathrm{i}}{\hbar}\int_{t_0}^{t} H(\tau)d\tau + \\ + \frac{1}{2!}\left(-\frac{\mathrm{i}}{\hbar}\right)^2 \int_{t_0}^{t}\int_{t_0}^{t} \mathcal{T}\{H(\tau_1)H(\tau_2)\}d\tau_1 d\tau_2 + \dots.$$

Mnemonically, we write

$$\boxed{P^{+}(t,t_0) = \mathcal{T}\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}\int_{t_0}^{t} H(\tau)d\tau}, \qquad t \geq t_0.} \tag{7.51}$$

The definition of the chronological operator $\mathcal{T}$ can be found on page 382.

Theorem 7.20 *The Dyson series converges for all times $t \geq t_0$. The solution of the initial-value problem*

$$i\hbar\dot{\psi}(t) = H\psi(t), \quad t \geq t_0, \quad \psi(t_0) = \psi_0$$

is given by $\psi(t) = P^{+}(t,t_0)\psi_0$ for all $t \geq t_0$.

Proof. Let $t \geq t_0$. The propagator equation (7.50) is equivalent to the Volterra integral equation

$$P(t,t_0) = I - \frac{\mathrm{i}}{\hbar}\int_{t_0}^{t} H(\tau)P(\tau,t_0)ds.$$

The iterative method

$$P_{n+1}(t,t_0) = I - \frac{\mathrm{i}}{\hbar}\int_{t_0}^{t} H(\tau)P_n(\tau,t_0)d\tau, \qquad n = 0,1,\dots$$

with the initial approximation $P_0(t,t_0) :\equiv I$ yields

$$P(t,t_0) = I + \sum_{n=1}^{\infty} \left(-\frac{\mathrm{i}}{\hbar}\right)^n \int H(\tau_1)\cdots H(\tau_n)$$

where $\int := \int_{t_0}^{t} d\tau_1 \int_{t_0}^{\tau_1} d\tau_2 \cdots \int_{t_0}^{\tau_{n-1}} d\tau_n$. From this we get the claim by using the chronological operator $\mathcal{T}$. In fact, for example, consider the integral

$$J := \int_{t_0}^{t} d\tau_1 \int_{t_0}^{\tau_1} d\tau_2 \; H(\tau_1)H(\tau_2)$$

over a triangle. We want to transform this into an integral over a square. To this end, let us equivalently write

$$J = \int_{t_0}^{t}\int_{t_0}^{t} H(\tau_1)H(\tau_2)\theta(\tau_1-\tau_2)\, d\tau_1 d\tau_2,$$

noting that $\theta(\tau_1-\tau_2) = 0$ if $\tau_2 > \tau_1$. Permuting the indices,

$$\begin{aligned} J &= \int_{t_0}^{t}\int_{t_0}^{t} \tfrac{1}{2}\{H(\tau_1)H(\tau_2)\theta(\tau_1-\tau_2) + H(\tau_2)H(\tau_1)\theta(\tau_2-\tau_1)\} d\tau_1 d\tau_2 \\ &= \int_{t_0}^{t}\int_{t_0}^{t} \tfrac{1}{2}\mathcal{T}\{H(\tau_1)H(\tau_2)\}\, d\tau_1 d\tau_2. \end{aligned}$$

Similarly, we transform the other integral terms. □

Physical motivation of the magic Dyson formula. Using a formal argument, we want to show that

> *The magic Dyson formula* (7.51) *is quite natural from the physical point of view; it follows from the superposition principle.*

The point is that we have to apply the superposition principle to time-dependent interactions. To this end, we decompose the time interval $[t_0, t]$ into small pieces of equal length, i.e., we set $t_k := t_0 + k\Delta t$, $k = 0, 1, \ldots, n$. Then

$$t_0 < t_1 < \ldots < t_n = t.$$

The following formulas are valid up to terms of higher order with respect to the time difference Δt. From the key differential equation $\mathrm{i}\hbar\dot{\psi}(t) = H(t)\psi(t)$ we get

$$\mathrm{i}\hbar\, \frac{\psi(t+\Delta t) - \psi(t)}{\Delta t} = H(t)\psi(t).$$

This implies

$$\psi(t+\Delta t) = \left(1 - \frac{\mathrm{i}}{\hbar} H(t)\Delta t\right)\psi(t).$$

It is convenient to write this as

$$\psi(t+\Delta t) = \mathrm{e}^{-\frac{\mathrm{i}}{\hbar}H(t)\Delta t}\psi(t),$$

up to terms of order $(\Delta t)^2$. Consider first the special case where $n = 2$. Then $t = t_2$, and

$$\psi(t) = \mathrm{e}^{-\frac{\mathrm{i}}{\hbar}H(t_1)\Delta t}\psi(t_1) = \mathrm{e}^{-\frac{\mathrm{i}}{\hbar}H(t_1)\Delta t}\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}H(t_0)\Delta t}\psi(t_0).$$

Since $t_1 > t_0$, we may add the chronological operator $\mathcal{T}$. Hence

$$\psi(t) = \mathcal{T}\{\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}H(t_1)\Delta t}\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}H(t_0)\Delta t}\}\psi(t_0).$$

By the addition theorem for the exponential function,

$$\psi(t) = \mathcal{T}\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}(H(t_1)+H(t_0))\Delta t}\psi(t_0).$$

Similarly, for $n = 2, 3, \ldots,$

$$\boxed{\psi(t) = \mathcal{T}\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}\sum_{k=0}^{n-1}H(t_k)\Delta t}\psi(t_0).}$$

Letting $n \to \infty$, we formally get

$$\psi(t) = \mathcal{T}\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}\int_{t_0}^{t}H(\tau)d\tau}\psi(t_0).$$

Using the propagator, $\psi(t) = P(t, t_0)\psi(t_0)$. This yields the Dyson propagator formula (7.51).

7.18 The Magic Dyson Formula for the *S*-Matrix

It is our goal to reformulate the Dyson propagator formula (7.51) in terms of perturbation theory. This way, we will obtain the magic Dyson formula for the S-matrix. This formula is the key to computing all kind of scattering processes for elementary particles in quantum field theory. The following perturbed Hamiltonian

$$\boxed{H = H_0 + \kappa V}$$

is typical for perturbation theory. Here, the operators $H_0, V : X \to X$ are linear and self-adjoint on the complex N-dimensional Hilbert space X where $1 \leq N < \infty$. The real constant $\kappa \geq 0$ is called the coupling constant. Recall that the retarded propagator with respect to the perturbed Hamiltonian H is given by

$$P^+(t, s) := \mathrm{e}^{-\mathrm{i}(t-s)H/\hbar}, \qquad t \geq s.$$

The key formula reads as

$$\boxed{P^+(t, s) = \mathrm{e}^{-\mathrm{i}tH_0/\hbar}\mathsf{S}(t, s)\mathrm{e}^{\mathrm{i}sH_0/\hbar}, \qquad t \geq s} \tag{7.52}$$

along with the magic Dyson formula

$$\boxed{\mathsf{S}(t, s) = \mathcal{T}\mathrm{e}^{-\frac{\mathrm{i}\kappa}{\hbar}\int_s^t V_I(\eta)d\eta}, \qquad t \geq s.} \tag{7.53}$$

Here, we set $V_I := \mathrm{e}^{\mathrm{i}tH_0/\hbar}V\mathrm{e}^{-\mathrm{i}tH_0\hbar}$. In physics, the operator $\mathsf{S}(t, s)$ is called the S-matrix operator with respect to the finite time-interval $[s, t]$.

Theorem 7.21 *For all real times t and s with $t \geq s$, the formulas* (7.52) *and* (7.53) *are valid.*

Let us first discuss the physical interpretation of the S-matrix. The crucial relation reads as

$$\boxed{\langle \mathrm{e}^{-\mathrm{i}tH_0/\hbar}\varphi_{\mathrm{out}}|P^+(t,s)\mathrm{e}^{-\mathrm{i}sH_0/\hbar}\varphi_{\mathrm{in}}\rangle = \langle \varphi_{\mathrm{out}}|\mathsf{S}(t,s)\varphi_{\mathrm{in}}\rangle.} \tag{7.54}$$

In terms of physics, this means the following:

- The function $\psi_{\mathrm{in}}(s) = \mathrm{e}^{-\mathrm{i}sH_0/\hbar}\varphi_{\mathrm{in}}$ describes the dynamics of a free particle at time s. We regard this particle as an incoming free particle running from time $s = -\infty$ until time $s = +\infty$. At time $s = 0$, the particle is in the state φ_{in}.
- The function $\psi_{\mathrm{out}}(t) = \mathrm{e}^{-\mathrm{i}tH_0/\hbar}\varphi_{\mathrm{out}}$ describes the dynamics of a free particle at time t. We regard this particle as an outgoing free particle running from time $t = -\infty$ until time $t = +\infty$. At time $t = 0$, the particle is in the state φ_{out}.
- The transition amplitude

$$a := \langle \psi_{\mathrm{out}}(t)|P^+(t,s)\psi_{\mathrm{in}}(s)\rangle$$

  corresponds to the transition probability

$$|a|^2 = |\langle \psi_{\mathrm{out}}(t)|P^+(t,s)\psi_{\mathrm{in}}(s)\rangle|^2$$

  from the free incoming particle at time s to the free outgoing particle at time t with $t > s$. By (7.54),

$$|a|^2 = |\langle \varphi_{\mathrm{out}}|\mathsf{S}(t,s)\varphi_{\mathrm{in}}\rangle|^2.$$

This way, the transition probability $|a|^2$ is related to the S-matrix operator $\mathsf{S}(t,s)$. Using the transition probability $|a|^2$, one can compute the corresponding cross sections which can be measured in particle accelerators. This justifies the terminology S-matrix (or scattering matrix). More precisely, using a complete orthonormal system $|1\rangle, \ldots, |N\rangle$ of the Hilbert space X, the matrix to the operator $\mathsf{S}(t,s)$ reads as $(\mathsf{S}_{jk}(t,s))$ where we set

$$\mathsf{S}_{jk}(t,s) := \langle j|\mathsf{S}(t,s)|k\rangle, \qquad j,k = 1,\ldots,N.$$

In order to simplify terminology, we frequently use the two terms 'S-matrix' and 'scattering operator' synonymously.

Proof of Theorem 7.21. Set $\hbar := 1$. Fix the initial time $s \in \mathbb{R}$. By (7.50), the Feynman propagator satisfies the differential equation

$$\mathrm{i}P_t(t,s) = HP(t,s), \quad t \geq s, \quad P(s,s) = I. \tag{7.55}$$

Note that $P^+(t,t_0) = P(t,t_0)$ for all $t \geq t_0$. The trick is to introduce the operator

$$\mathsf{S}(t,s) := \mathrm{e}^{\mathrm{i}tH_0} P^+(t,s) \mathrm{e}^{-\mathrm{i}sH_0}.$$

By (7.55), it follows from $P^+(t,s) = \mathrm{e}^{-\mathrm{i}tH_0}\mathsf{S}(t,s)\mathrm{e}^{\mathrm{i}sH_0}$ that

$$H_0\mathrm{e}^{-\mathrm{i}tH_0}\mathsf{S}(t,s)\mathrm{e}^{\mathrm{i}sH_0} + \mathrm{i}\mathrm{e}^{-\mathrm{i}tH_0}\mathsf{S}_t(t,s)\mathrm{e}^{\mathrm{i}sH_0} = (H_0+\kappa V)\mathrm{e}^{-\mathrm{i}tH_0}\mathsf{S}(t,s)\mathrm{e}^{\mathrm{i}sH_0}.$$

Hence

$$\mathrm{i}\mathrm{e}^{-\mathrm{i}tH_0}\mathsf{S}_t(t,s)\mathrm{e}^{\mathrm{i}sH_0} = \kappa V \mathrm{e}^{-\mathrm{i}tH_0}\mathsf{S}(t,s)\mathrm{e}^{\mathrm{i}sH_0}.$$

Setting $V_I(t) := \mathrm{e}^{\mathrm{i}tH_0} V \mathrm{e}^{-\mathrm{i}tH_0}$, we get the modified differential equation

$$\mathrm{i}\mathsf{S}_t(t,s) = \kappa V_I(t)\mathsf{S}(t,s), \quad t \geq s, \quad \mathsf{S}(s,s) = I.$$

Theorem 7.20 tells us that $\mathsf{S}(t,s) = \mathcal{T}\mathrm{e}^{-\frac{\kappa \mathrm{i}}{\hbar}\int_s^t V_I(\eta)d\eta}$. □

7.19 Canonical Transformations

Canonical transformations were introduced by Jacobi (1804–1851) in order to simplify the solution of difficult problems in celestial mechanics. It is a typical property of canonical transformations that

- they do not change the physics and
- they preserve the mathematical structure.

In quantum physics, the typical mathematical structure is given by the Hilbert space structure. This motivates the following definition:

Canonical transformations are Hilbert space isomorphisms.

Recall that by a Hilbert space isomorphism (or a unitary operator), we understand a linear bijective map $U : X \to Y$ from the complex Hilbert space X onto the complex Hilbert space Y which preserves the inner product,

$$\langle U\varphi | U\psi\rangle = \langle \varphi | \psi\rangle \qquad \text{for all} \quad \varphi, \psi \in X.$$

We want to study two important canonical transformations in quantum physics, namely,

- the passage from the Schrödinger picture to the Heisenberg picture, and
- the passage from the Schrödinger picture to Dirac's interaction picture.

7.19.1 The Schrödinger Picture

Let $H : X \to X$ be a linear self-adjoint operator on the complex N-dimensional Hilbert space X where $1 \leq N < \infty$. For given $\psi_0 \in X$, the Schrödinger equation reads as

$$\boxed{\mathrm{i}\hbar\dot{\psi}(t) = H\psi(t), \qquad t \geq 0, \quad \psi(0) = \psi_0.} \tag{7.56}$$

We are looking for the function $\psi : [0, \infty[\to X$. The unique solution of (7.56) is given by

$$\psi(t) = \mathrm{e}^{-\mathrm{i}tH/\hbar}\psi_0, \qquad t \geq 0.$$

Here, $\psi(t)$ describes the state of the quantum system at time t. By an observable, we understand an arbitrary linear self-adjoint operator

$$A : X \to X.$$

Suppose that the initial state $\psi(0)$ is normalized. Then, the state $\psi(t)$ is normalized for all times $t \in \mathbb{R}$. The real number

$$\bar{A}(t) = \langle \psi(t) | A\psi(t) \rangle$$

is the mean value of the observable A measured in the state $\psi(t)$ at time t. Obviously, we get

$$\bar{A}(t) = \langle \psi(0) | \mathrm{e}^{\mathrm{i}tH/\hbar} A \mathrm{e}^{-\mathrm{i}tH/\hbar} \psi(0) \rangle.$$

This identity is the basis for the passage to the Heisenberg picture to be considered now.

7.19.2 The Heisenberg Picture

Fix time $t \in \mathbb{R}$. To each observable $A : X \to X$, we assign the operator

$$\boxed{A_H(t) := \mathrm{e}^{\mathrm{i}tH/\hbar} A \mathrm{e}^{-\mathrm{i}tH/\hbar}.}$$

In addition, to each state $\varphi \in X$, we assign the element $\varphi_H(t)$ of X where

$$\boxed{\varphi_H(t) := \mathrm{e}^{\mathrm{i}tH/\hbar}\varphi.}$$

We call $A_H(t)$ and $\varphi_H(t)$ the Heisenberg picture of the observable A and the state φ at time t, respectively. Set

$$U(t) := \mathrm{e}^{\mathrm{i}tH/\hbar}.$$

The operator $U(t) : X \to X$ is unitary, that is,

$$\langle \varphi | \chi \rangle = \langle U(t)\varphi | U(t)\chi \rangle \tag{7.57}$$

for all $\varphi, \chi \in X$ and all $t \in \mathbb{R}$. Therefore, the transformation $\varphi_H(t) = U(t)\varphi$ represents a canonical transformation at time t. Observe the following:

(i) Dynamics: For the solution of the Schrödinger equation,

$$\psi(t) = \mathrm{e}^{-\mathrm{i}tH/\hbar}\psi(0), \qquad t \in \mathbb{R},$$

the Heisenberg picture is given by

$$\psi(t)_H = \mathrm{e}^{\mathrm{i}tH/\hbar}\psi(t) = \psi(0), \qquad t \in \mathbb{R}.$$

This state does not depend on time t.

(ii) Observable: The Heisenberg picture $A_H(t)$ of the observable A satisfies the following equation of motion[32]

$$\boxed{\mathrm{i}\hbar\dot{A}_H(t) = [A_H(t), H]_-, \qquad t \in \mathbb{R}}$$

which replaces the Schrödinger equation in the Heisenberg picture.

(iii) Transition amplitude: It follows from (7.57) and $A_H(t) = U(t)AU(t)^{-1}$ that

$$\langle\varphi|A\chi\rangle = \langle\varphi_H(t)|A_H(t)\chi_H(t)\rangle$$

for all $\varphi, \chi \in X$ and all $t \in \mathbb{R}$.

7.19.3 The Dirac Interaction Picture

Dirac modified the Heisenberg picture in terms of perturbation theory. To explain this, consider the situation

$$\boxed{H = H_0 + \kappa V}$$

which is typical for perturbation theory. Let $H_0, V : X \to X$ be linear self-adjoint operators on the complex N-dimensional Hilbert space X where $1 \leq N < \infty$. The real constant $\kappa \geq 0$ is called coupling constant. In physics, the Hamiltonian H is called a perturbation of the free Hamiltonian H_0. The idea of Dirac's interaction picture is to modify the passage to the Heisenberg picture by replacing the Hamiltonian H by the free Hamiltonian H_0. Explicitly, fix time $t \in \mathbb{R}$. To each observable $A : X \to X$, we assign the operator

$$\boxed{A_I(t) := \mathrm{e}^{\mathrm{i}tH_0/\hbar}A\mathrm{e}^{-\mathrm{i}tH_0/\hbar}.}$$

Moreover, to each state $\varphi \in X$, we assign the element $\varphi_I(t)$ of X where

$$\boxed{\varphi_I(t) := \mathrm{e}^{\mathrm{i}tH_0/\hbar}\varphi.}$$

We call $A_I(t)$ and $\varphi_I(t)$ the Dirac interaction picture of the observable A and the state φ at time t, respectively. Set

$$U_0(t) := \mathrm{e}^{\mathrm{i}tH_0/\hbar}.$$

The operator $U_0(t) : X \to X$ is unitary, that is,

$$\langle\varphi|\chi\rangle = \langle U_0(t)\varphi|U_0(t)\chi\rangle$$

for all $\varphi, \chi \in X$ and all $t \in \mathbb{R}$. Thus, the transformation $\varphi_I(t) = U_0(t)\varphi$ represents a canonical transformation at time t. Observe the following.

[32] In fact, differentiation with respect to time t yields

$$\mathrm{i}\hbar\dot{A}_H(t) = -H\mathrm{e}^{\mathrm{i}tH/\hbar}A\mathrm{e}^{-\mathrm{i}tH/\hbar} + \mathrm{e}^{\mathrm{i}tH/\hbar}A\mathrm{e}^{-\mathrm{i}tH/\hbar}H.$$

Moreover, recall that $[A, B]_- := AB - BA$.

(i) Dynamics: Let

$$\psi(t) = \mathrm{e}^{-\mathrm{i}tH/\hbar}\psi(0), \qquad t \in \mathbb{R}$$

be a solution of the Schrödinger equation $\mathrm{i}\hbar\dot{\psi}(t) = H\psi(t), t \in \mathbb{R}$ with respect to the perturbed Hamiltonian H. Then, the interaction picture $\psi_I = \psi_I(t)$ satisfies the differential equation[33]

$$\boxed{\mathrm{i}\hbar\dot{\psi}_I(t) = \kappa V_I(t)\psi_I(t), \qquad t \in \mathbb{R}.}$$

Here, $V_I(t) := \mathrm{e}^{\mathrm{i}tH_0/\hbar}V\mathrm{e}^{-\mathrm{i}tH_0/\hbar}$.

(ii) Observable: The Dirac interaction picture $A_I(t)$ of the observable A satisfies the following equation of motion

$$\boxed{\mathrm{i}\hbar\dot{A}_I(t) = [A_I(t), H_0]_-, \qquad t \in \mathbb{R}}$$

which replaces the Schrödinger equation in the Dirac interaction picture. This equation only depends on the free Hamiltonian H_0.

(iii) Transition amplitude: For all $\varphi, \chi \in X$ and all $t \in \mathbb{R}$,

$$\langle\varphi|A\chi\rangle = \langle\varphi_I(t)|A_I(t)\chi_I(t)\rangle.$$

7.20 Functional Calculus

> Functionals generalize classical functions to systems with a finite or infinite number of degrees of freedom. Folklore

It was emphasized by Feynman (1918–1988) and Schwinger (1918–1994) in the 1940s that it is very useful for quantum field theory to extend the classical calculus due to Newton (1643–1727) and Leibniz (1646–1716) to functionals. In mathematics, the differentiation of functionals and operators was introduced by Fréchet (1878–1973), Gâteaux (1889–1914), and Volterra (1860–1950) in about 1900. The goal was to give the calculus of variations a rigorous basis.[34] In the 1920s, functional integrals (Euclidean path integrals) were introduced by Wiener (1894–1964) in order to mathematically describe Einstein's 1905 approach to Brownian motion (theory of stochastic processes). We are going to discuss the basic ideas.

[33] In fact, $\psi_I(t) = \mathrm{e}^{\mathrm{i}tH_0/\hbar}\psi(t)$. Differentiation with respect to time yields

$$\mathrm{i}\hbar\dot{\psi}_I(t) = -H_0\mathrm{e}^{\mathrm{i}tH_0/\hbar}\psi(t) + \mathrm{e}^{\mathrm{i}tH_0/\hbar}\mathrm{i}\hbar\dot{\psi}(t).$$

This is equal to $-H_0\mathrm{e}^{\mathrm{i}tH_0/\hbar}\psi(t) + \mathrm{e}^{\mathrm{i}tH_0/\hbar}(H_0 + \kappa V)\psi(t) = \kappa V_I(t)\psi_I(t)$.

[34] This modern version of calculus is thoroughly studied in Zeidler (1986), Vol. I, Chap. 4.

7.20.1 Functional Derivatives

> Classical derivatives are generalized to functional derivatives; differentials are linear functionals in modern mathematics.
>
> Folklore

Let $Z : X \to \mathbb{C}$ be a functional on the complex Hilbert space X. We write

$$Z = Z(J).$$

That is, to each element J of X we assign the complex number $Z(J)$. In quantum field theory, such functionals arise in a natural way. Prototypes are the action, $S(\psi)$, of a quantum field ψ and the generating functional Z for the correlation functions (see Chap. 13). Then, $Z(J)$ is the value of the generating functional at the point J. Intuitively, the source function J describes an external force acting on the physical system. The functional derivative $Z'(J)$ tells us then the response of the physical system under a small change of the source. It is our goal to investigate the following generalizations:

- derivative $\Rightarrow$ functional derivative;
- partial derivative $\Rightarrow$ partial functional derivative;
- integral $\Rightarrow$ functional integral.

Notation. In mathematics, the following notions possess a precise meaning:

- functional derivative and partial functional derivative,
- directional derivative,
- variation,
- differential,
- infinitesimal transformation.

The confusion caused by infinitesimals. The idea of infinitesimals was introduced by Newton and Leibniz in the 17th century. They used the relation

$$(\delta x)^2 = 0 \tag{7.58}$$

for the 'infinitesimally small quantity' δx. Obviously, the only real number δx which satisfies the magic relation (7.58) is given by $\delta x = 0$, which does not fit the intention of Newton and Leibniz. Thus, there is a lot of confusion concerning (7.58), which has survived in parts of the physical literature. Nowadays, the notions are completely clarified in mathematics. There exist two approaches, namely,

- the standard approach, and
- the non-standard approach.

In the standard approach, one completely avoids the relation (7.58). This approach will be applied in the present first volume and in most parts of the further volumes. In non-standard analysis, one also introduces the following notions:

- infinitesimally small number, and
- infinitely large number.

In the non-standard setting, the relation $(\delta x)^2 = 0$ takes on a precise meaning for an infinitesimally small non-standard number, δx.[35] In the volumes of this treatise, we will show that relations of the form

$$\boxed{\partial^2 = 0, \qquad d^2 = 0, \qquad \delta^2 = 0 \qquad Q^2 = 0}$$

play a fundamental role in modern mathematics and physics. In particular,

- the relation $\partial^2 = 0$ for the boundary operator ∂ is responsible for Poincaré's homology theory in algebraic topology,
- the Poincaré lemma $d^2 = 0$ for differential forms (that is, $d(d\omega) = 0$) is the basis for de Rham's cohomology theory in differential topology,
- the relation $\delta^2 = 0$ allows us to introduce the Hodge homology on Riemannian manifolds which is dual to the de Rham cohomology, and
- the operator relation $Q^2 = 0$ is crucial for the BRST quantization (or cohomological quantization) of gauge theories (e.g., the Standard Model in particle physics and string theories).

There exists a branch of mathematics called homological algebra which studies the far-reaching consequences of the relation $Q^2 = 0$ in terms of exact sequences and homology groups (resp. the dual cohomology groups).[36] In the

[35] Non-standard analysis was rigorously founded in 1960 by the logician Abraham Robinson (1918–1974). This will be considered in Volume II, by using ultrafilters. The elegant basic idea of non-standard analysis is to construct a field $*\mathbb{R}$ of mathematical objects called non-standard numbers such that $*\mathbb{R}$ represents an extension of the field $\mathbb{R}$ of real numbers,

$$\mathbb{R} \subset *\mathbb{R}.$$

Besides the real numbers, the field $*\mathbb{R}$ contains infinitesimally small numbers and infinitely large numbers. For all the elements of $*\mathbb{R}$, the operations of addition, multiplication, and division (by nonzero elements) are well defined. In terms of algebra, the set $*\mathbb{R}$ is a field which extends the field $\mathbb{R}$ of classical real numbers. For two positive elements x and y of $\mathbb{R}$, there exists always a nonzero natural number n such that $x < ny$. This so-called Archimedian property of the field $\mathbb{R}$ is not anymore valid for $*\mathbb{R}$.

[36] The classical book is H. Cartan and S. Eilenberg, Homological Algebra, Princeton University Press. In terms of physics, we recommend M. Henneaux and C. Teitelboim, Quantization of Gauge Systems, Princeton University Press, 1993. We will show in Volume IV on quantum mathematics that the physical origin of homology and cohomology is rooted in electric circuits and the Maxwell equations in electrodynamics.

early 1950s, Jean Leray and Henri Cartan showed that the theory of holomorphic functions of several variables can be reformulated elegantly in the language of sheaf cohomology[37]

The one-dimensional case. As starting point, consider the Taylor expansion of the smooth function $f : \mathbb{R} \to \mathbb{R}$,[38]

$$f(x + \Delta x) = f(x) + f'(x)\Delta x + o(\Delta x), \qquad \Delta x \to 0.$$

We define

$$\boxed{\delta x := \Delta x, \qquad \delta f := f'(x)\delta x,}$$

and

$$\boxed{df(x)(h) := f'(x)h, \qquad dx(h) := h \qquad \text{for all} \quad h \in \mathbb{R}.}$$

With a view to generalizations to be considered below, note that

- the variations δx and δf are real numbers,
- whereas the differentials $dx, df(x) : \mathbb{R} \to \mathbb{R}$ are linear mappings (functionals) on the real line $\mathbb{R}$.

Obviously, we have $df(x)(h) = f'(x)dx(h)$ for all $h \in \mathbb{R}$. This is equivalent to the rigorous formula

$$\boxed{df(x) = f'(x)dx,}$$

in the sense of mappings.

The language of physicists. In the physics literature, one proceeds frequently as follows. By Taylor expansion,

$$\delta f = f(x + \delta x) - f(x) = f'(x)\delta x + \frac{f''(x)\delta x^2}{2} + \dots \tag{7.59}$$

Using $(\delta x)^2 = 0$, we get

$$\delta f = f'(x)\delta x. \tag{7.60}$$

This is formally the same result as above. In terms of mathematics, let us write

$$f(x) = g(x) \mod o(x), \qquad x \to 0$$

iff $f(x) - g(x) = o(x)$ as $x \to 0$. In particular, $(\delta x)^2 = o(x)$ as $x \to 0$ means that

$$(\delta x)^2 = 0 \mod o(x), \qquad x \to 0.$$

In standard mathematics, this replaces the magic relation (7.58) due to Newton and Leibniz. The passage to the language of physicists consists then in

[37] See K. Maurin, Methods of Hilbert Spaces, PWN, Warsaw, 1972.

[38] The definition of the classical Landau symbols $o(\Delta x)$ and $O(\Delta x)$ can be found on page 932. In particular, we write $r(\chi) = o(\chi)$ as $\chi \to 0$ iff $\lim_{\chi \to 0} \frac{r(\chi)}{\chi} = 0$.

dropping out the symbol mod $o(x)$ for simplifying notation. In this sense, equation (7.59) implies (7.60).

The general case. Consider now the functional $Z : X \to \mathbb{C}$ on the complex Hilbert space X. Our starting point is the definition of the functional derivatives

$$\boxed{\frac{\delta Z(J)}{\delta J}, \qquad \frac{\delta^2 Z(J)}{\delta J^2}.}$$

Synonymously, we will write

$$Z'(J) = \frac{\delta Z(J)}{\delta J}, \qquad Z''(J) = \frac{\delta^2 Z(J)}{\delta J^2}.$$

Naturally enough, we will formulate the corresponding definitions in such a way that, in the special case where $Z = Z(J)$ is a complex-valued function of the real variable J (i.e., $X = \mathbb{R}$), the functional derivative coincides with the classical derivative. That is,

$$\boxed{Z'(J) = \frac{\delta Z(J)}{\delta J} = \frac{dZ(J)}{dJ}, \qquad Z''(J) = \frac{\delta^2 Z(J)}{\delta J^2} = \frac{d^2 Z(J)}{dJ^2}.}$$

Moreover, our notation will be chosen in such a way that in the classical case, $X = \mathbb{R}$, we get

$$\frac{\delta Z(J)}{\delta J}(h) = Z'(h)h, \qquad \frac{\delta^2 Z(J)}{\delta J^2}(h,k) = Z''(J)hk$$

for all real numbers h, k. In the volumes of this treatise, we will extensively use the calculus of differential forms. In modern mathematics, differentials are not infinitesimally small quantities, but functionals. Explicitly,

$$dZ(J)(h) = Z'(J)(h), \qquad d^2 Z(J)(h,k) = Z''(J)(h,k).$$

In this sense, $dZ(J) = Z'(J)$ and $d^2 Z(J) = Z''(J)$.

In the calculus of variations, one writes

$$\delta Z := \frac{\delta Z(J)}{\delta J}(h).$$

This is called the first variation of the functional Z at the point J in direction of the vector h. More precisely, one has to write

$$\delta Z(J;h) := \frac{\delta Z(J)}{\delta J}(h).$$

Similarly,

$$\delta^2 Z := \frac{\delta^2 Z(J)}{\delta J^2}(h,k).$$

More precisely,

$$\delta^2 Z(J;h,k) = \frac{\delta^2 Z(J)}{\delta J^2}(h,k).$$

Basic definition of the functional derivative. Fix the point $J \in X$. For given $h \in X$, define

$$\frac{\delta Z(J)}{\delta J}(h) := \lim_{t\to 0} \frac{Z(J+th) - Z(J)}{t}.$$

Here, t is a real parameter. If this limit exists, then it is called the directional derivative of the functional Z at the point J in direction of the vector h. Equivalently,

$$\boxed{\frac{\delta Z(J)}{\delta J}(h) := \frac{d}{dt}\left(Z(J+th)\right)_{|t=0}.}$$

This allows us the following physical interpretation. Think of Z as temperature and regard J as a point in 3-dimensional Euclidean space. Starting at the point J, we move along a straight line in direction of the vector h. At time t we reach the point $J+th$ and we observe the temperature

$$Z(J+th).$$

Differentiating this with respect to time t at the initial time, $t=0$, we get the directional derivative of temperature Z at the point J in direction of the vector h. This quantity is also called the temperature gradient at the point J in direction of h. For the change of temperature, we get

$$Z(J+th) = Z(J) + t\frac{\delta Z(J)}{\delta J}(h) + o(t), \quad t \to 0.$$

In the general case, the map $h \mapsto \frac{\delta Z(J)}{\delta J}(h)$ represents an operator of the form

$$\frac{\delta Z(J)}{\delta J} : X \to \mathbb{C}.$$

Parallel to classical calculus, this operator is also denoted by the symbol

$$\boxed{Z'(J) := \frac{\delta Z(J)}{\delta J}.}$$

We call $Z'(J)$ the functional derivative[39] of the functional Z at the point J.

Higher-order functional derivatives. Fix $h,k \in X$. Naturally enough, we define

$$\frac{\delta^2 Z(J)}{\delta J}(h,k) := \frac{d}{dt}\left(\frac{\delta Z(J+tk)}{\delta J}(h)\right)_{|t=0}.$$

The map $(h,k) \mapsto \frac{\delta^2 Z(J)}{\delta J^2}(h,k)$ represents an operator of the form

[39] As a rule, the map $Z'(J) : X \to \mathbb{C}$ is linear, but this is not always the case.

$$\frac{\delta^2 Z(J)}{\delta J^2} : X \times X \to \mathbb{C}.$$

We also introduce the notation

$$\boxed{Z''(J) := \frac{\delta^2 Z(J)}{\delta J^2}.}$$

This operator is called the second functional derivative[40] of the functional Z at the point J. Summarizing, the first and second functional derivatives of the functional Z at the point J are operators of the form

$$\boxed{Z'(J) : X \to \mathbb{C}, \qquad Z''(J) : X \times X \to \mathbb{C}.}$$

Higher-order functional derivatives are defined analogously. For example, fix $h, k, l \in X$. We then define

$$\frac{\delta^3 Z(J)}{\delta J^3}(h,k,l) := \frac{d}{dt}\left(\frac{\delta^2 Z(J+tl)}{\delta J^2}(h,k)\right)_{|t=0}.$$

Example. Define $F(\psi) := \langle\psi|\psi\rangle$ for all $\psi \in X$ where X is a complex Hilbert space.[41] Then, for all $h, k \in X$,

$$F'(\psi)(h) = \frac{\delta F(\psi)}{\delta \psi}(h) = \langle\psi|h\rangle + \langle\psi|h\rangle^\dagger$$

and $F''(\psi)(h,k) = \langle h|k\rangle + \langle h|k\rangle^\dagger$.

Proof. Set $\chi(t) := F(\psi + th)$ for all $t \in \mathbb{R}$. Explicitly,

$$\chi(t) = \langle\psi|\psi\rangle + t(\langle\psi|h\rangle + \langle h|\psi\rangle) + t^2\langle h|h\rangle.$$

This implies $\chi'(0) = F'(\psi)(h) = \langle\psi|h\rangle + \langle h|\psi\rangle$. Moreover, let $k \in X$. Set

$$\varrho(t) := F'(\psi + kt)(h) = \langle\psi + tk|h\rangle + \langle h|\psi + tk\rangle, \qquad t \in \mathbb{R}.$$

Hence $\varrho'(0) = F''(\psi)(h,k) = \langle k|h\rangle + \langle h|k\rangle$. □

7.20.2 Partial Functional Derivatives

> Here, one finds a method which requires only a simple use of the principles of differential and integral calculus; above all I must call attention to the fact that I have introduced in my calculations a new characteristic δ since this method requires that the same quantities vary in two different ways.
>
> Comte de Joseph Louis Lagrange, 1762

[40] As a rule, $Z''(J)(h,k)$ is linear with respect to h and k, and we have the symmetry property $Z''(J)(h,k) = Z''(J)(k,h)$, but this is not always the case.

[41] In a real Hilbert space X, we have $F'(\psi)(h) = 2\langle\psi|h\rangle$ and

$$F''(\psi)(h,k) = 2\langle h|k\rangle \qquad \text{for all} \quad h, k \in X.$$

> By generalizing Euler's 1744 method, Lagrange (1736–1813) got the idea for his remarkable formulas, where in a single line there is contained the solution of all problems of analytic mechanics.
>
> Carl Gustav Jacob Jacobi (1804–1851)

It is our goal to generalize the classical partial derivatives

$$\frac{\partial f(x,y)}{\partial x}, \qquad \frac{\partial^2 f(x,y)}{\partial x \partial y}$$

to the partial functional derivatives

$$\frac{\delta Z(J)}{\delta J(x)}, \qquad \frac{\delta^2 Z(J)}{\delta J(x)\delta J(y)},$$

respectively. In classical calculus, the problem

$$f(x^1, x^2) = \text{critical!}$$

is equivalent to

$$\frac{\partial f(x^1, x^2)}{\partial x^j} = 0 \qquad \text{for all indices } j.$$

In the calculus of variations, the solutions of the principle of critical action

$$\boxed{S(\psi) = \text{critical!}, \qquad \psi \in X}$$

satisfy the so-called variational equation

$$\frac{\delta S(\psi)}{\delta \psi} = 0.$$

This implies

$$\boxed{\frac{\delta S(\psi)}{\delta \psi(x)} = 0 \qquad \text{for all indices } x} \tag{7.61}$$

which represents the desired equation of motion for the field ψ. This equation is also called the Euler–Lagrange equation.

> *We will show in this treatise that all of the fundamental field equations in physics are of the type* (7.61).

For example, this concerns the electromagnetic field, non-relativistic and relativistic quantum mechanics, the Standard Model in particle physics, and the theory of general relativity. The basic tool for introducing partial functional derivatives is the notion of the density of a given functional; this generalizes the classical mass density

Density of a functional. Consider a set M equipped with a measure μ. Let X be an appropriate subspace of the space of all the functions $h : M \to \mathbb{C}$. For given function $\varrho \in X$, define

$$F(h) := \int_C \varrho(x)h(x)d\mu(x) \qquad \text{for all} \quad h \in X.$$

The function ϱ is called the density function of the functional F. This definition is well formulated if the function ϱ is uniquely determined by the functional F. This uniqueness has to be checked in each case. Examples will be considered below. Suppose now that the functional derivative $\frac{\delta Z(J)}{\delta J}$ has a density ϱ, that is,

$$\frac{\delta Z(J)}{\delta J}(h) = \int_M \varrho(x)h(x)d\mu(x) \qquad \text{for all} \quad h \in X.$$

We then define

$$\frac{\delta Z(J)}{\delta J(x)} := \varrho(x).$$

This is called the partial functional derivative of the functional Z at the point J with respect to the 'index' x. Summarizing, for all $h \in X$, we get the suggestive formula

$$\boxed{\frac{\delta Z(J)}{\delta J}(h) = \int_M \frac{\delta Z(J)}{\delta J(x)}\, h(x)d\mu(x).}$$

The variational lemma. Let us now study a few examples which are prototypes for general situations arising in quantum field theory. For checking the density property, we will use the following result. Let $-\infty < a < b < \infty$, and let $C[a,b]$ denote the space of all continuous functions $f : [a,b] \to \mathbb{R}$.

Proposition 7.22 *Suppose that for given two functions $f, g \in C[a,b]$, we have*

$$\int_a^b f(x)h(x)dx = \int_a^b g(x)h(x)dx \qquad \textit{for all} \quad h \in C[a,b]. \tag{7.62}$$

Then $f(x) = g(x)$ for all $x \in [a,b]$.

Proof. Set $F(x) := f(x) - g(x)$. Then $\int_a^b F(x)h(x)dx = 0$ for all functions $h \in C[a,b]$. Choosing $h = F$,

$$\int_a^b F(x)^2 dx = 0.$$

Hence $F(x) = 0$ for all $x \in [a,b]$. □

It is important for the calculus of variations that there exists a stronger variant of the preceding proposition. To this end, let $\mathcal{D}(a,b)$ denote the set of all smooth functions $h : [a,b] \to \mathbb{R}$ which have compact support in the open interval $]a,b[$, that is, they vanish in some open neighborhoods of the boundary points a and b.

Proposition 7.23 *Suppose that for given two functions $f, g \in C[a,b]$, we have*

$$\int_a^b f(x)h(x)dx = \int_a^b g(x)h(x)dx \qquad \text{for all} \quad h \in \mathcal{D}(a,b). \tag{7.63}$$

Then $f(x) = g(x)$ for all $x \in [a,b]$.

This is a special case of the variational lemma to be considered in Prop. 10.15 on page 543. The idea of proof is to use a limiting process in order to get the identity (7.62) from (7.63).

Example. Set $X := C[a,b]$. Fix the function $\varrho \in X$, and define

$$Z(J) := \int_a^b \varrho(x)J(x)dx \qquad \text{for all} \quad J \in X.$$

This is a functional $F : X \to \mathbb{R}$ on the real, linear function space X.

Proposition 7.24 *For fixed $J \in X$, the functional derivative is given by*

$$\frac{\delta Z(J)}{\delta J}(h) = \int_a^b \varrho(x)h(x)dx \qquad \text{for all} \quad h \in X.$$

This functional has the function ϱ as density.

According to this fact, for each given point $x \in [a,b]$ we define the partial functional derivative as

$$\frac{\delta Z(J)}{\delta J(x)} := \varrho(x).$$

Proof. Fix $h \in X$. Define

$$\chi(t) := Z(J + th) = \int_a^b \varrho(x)\{J(x) + th(x)\}dx, \qquad t \in \mathbb{R}.$$

Hence $\chi'(0) = Z'(J)(h) = \int_a^b \varrho(x)h(x)dx$. The uniqueness of the density function follows from Prop. 7.22.

Example. Choose again $X := C[a,b]$. We now set

$$Z(J) := \frac{1}{2}\int_C \varrho(x,y)J(x)J(y)dxdy \qquad \text{for all} \quad J \in X.$$

Here, we are given the continuous function $\varrho : M \to \mathbb{R}$ on the closed square $M := [a,b] \times [a,b]$. In addition, we assume that ϱ is symmetric, that is, we have $\varrho(x,y) = \varrho(y,x)$ for all $(x,y) \in M$.

Proposition 7.25 *Fix $J \in X$. For all $h, k \in X$, the first functional derivative and the second functional derivative are given by*

$$\frac{\delta Z(J)}{\delta J}(h) = \int_M \varrho(x,y)h(x)J(y)dxdy$$

and

$$\frac{\delta^2 Z(J)}{\delta J^2}(h,k) = \int_M \varrho(x,y)h(x)k(y)dxdy,$$

respectively. The second functional derivative has the function density ϱ.

Therefore, for all $x, y \in [a,b]$, we define

$$\frac{\delta^2 Z(J)}{\delta J(x)\delta J(y)} := \varrho(x,y).$$

This is the second partial functional derivative of the functional Z.

Proof. (I) First functional derivative. Let $t \in \mathbb{R}$. From

$$\chi(t) := Z(J+th) = \frac{1}{2}\int_M \varrho(x,y)\{J(x)+th(x)\}\{J(y)+th(y)\}dxdy$$

we get

$$\chi'(0) = Z'(J)(h) = \frac{1}{2}\int_M \{\varrho(x,y)h(x)J(y) + \varrho(x,y)J(x)h(y)\}dxdy.$$

By symmetry of ϱ, $Z'(J)(h) = \int_M \varrho(x,y)h(x)J(y)dxdy$.

(II) Second partial derivative. Differentiating the function

$$\sigma(t) := Z'(J+tk)(h)$$

with respect to t at the point $t = 0$, we obtain

$$Z''(J)(h,k) = \int_M \varrho(x,y)h(x)k(y)dxdy.$$

(III) Uniqueness of the density. Let $\varrho_* : M \to \mathbb{R}$ be continuous, and suppose that

$$\int_M \varrho(x,y)h(x)k(y)dxdy = \int_M \varrho_*(x,y)h(x)k(y)dxdy \tag{7.64}$$

for all $h, k \in X$. Let $F : M \to \mathbb{R}$ be an arbitrary continuous function. By the classical Weierstrass theorem, there exists a sequence $p_n : M \to \mathbb{R}$ of polynomials in two variables such that $\max_{a\le x,y\le b} |f(x,y) - p_n(x,y)| \to 0$ as $n \to \infty$. By (7.64), $\int_M \varrho(x,y)p_n(x,y)dxdy = \int_M \varrho_*(x,y)p_n(x,y)dxdy$ for all indices n. Letting $n \to \infty$, we get

$$\int_M \varrho(x,y)F(x,y)dxdy = \int_M \varrho_*(x,y)F(x,y)dxdy$$

for all continuous functions $F : M \to \mathbb{R}$. The same argument as in the proof of Prop. 7.24 tells us now that $\varrho = \varrho_*$ on M. □

Example. In the calculus of variations, the density functions of functionals are obtained by using integration by parts. As a prototype, consider the functional

$$S[q] := \frac{1}{2}\int_{t_0}^{t_1} \dot{q}(t)^2 dt \qquad \text{for all} \quad q \in X.$$

The dot denotes the time derivative. By definition, the symbol X represents the space of all smooth functions $q : [t_0, t_1] \to \mathbb{R}$ on the compact time interval $[t_0, t_1]$ which vanish on the boundary, that is, $q(t_0) = q(t_1) = 0$.

Proposition 7.26 *Let $q \in X$. The functional derivative is given by*

$$\frac{\delta S[q]}{\delta q}(h) = \int_{t_0}^{t_1} -\ddot{q}(t)h(t)dt \qquad \textit{for all} \quad h \in X.$$

This functional has the density $-\ddot{q}$.

Therefore, for each point $t \in [t_0, t_1]$, we define the partial functional derivative

$$\frac{\delta S[q]}{\delta q(t)} := -\ddot{q}(t).$$

Proof. (I) Functional derivative. Fix $q, h \in X$. For each parameter $\tau \in \mathbb{R}$, define

$$\chi(\tau) := S[q + \tau h] = \frac{1}{2}\int_{t_0}^{t_1} (\dot{q}(t) + \tau\dot{h}(t))^2 dt, \qquad \tau \in \mathbb{R}.$$

Hence $\chi'(0) = S'[q](h) = \int_{t_0}^{t_1} \dot{q}(t)\dot{h}(t)dt$. Observing the boundary condition $h(t_0) = h(t_1) = 0$, integration by parts yields

$$S'[q](h) = -\int_{t_0}^{t_1} \ddot{q}(t)h(t)dt \qquad \text{for all} \quad h \in X.$$

This proves the claim for the functional derivative.

(II) Uniqueness of the density. Let $g : [t_0, t_1] \to \mathbb{R}$ be a continuous function. Suppose that

$$\int_{t_0}^{t_1} -\ddot{q}(t)h(t)dt = \int_{t_0}^{t_1} g(t)h(t)dt \qquad \text{for all} \quad h \in X.$$

In particular, this identity is true for all functions $h \in \mathcal{D}(t_0, t_1)$. By the variational lemma (Prop. 7.23), $g = -\ddot{q}$ on $[t_0, t_1]$. □

The principle of least action for the classical harmonic oscillator. Let us study the motion $q = q(t)$ of a particle of mass m on the real line. Fix the compact time interval $[t_0, t_1]$. The functional

$$S[q] := \int_{t_0}^{t_1} \left(\frac{m\dot{q}(t)^2}{2} - \frac{\kappa q(t)^2}{2} + F(t)q(t) \right) dt, \qquad q \in X$$

is called the action of the particle. Here, κ is a positive number called coupling constant, and the given function $F : [t_0, t_1] \to \mathbb{R}$ is smooth. By definition, the space X consists of all smooth functions $q : [t_0, t_1] \to \mathbb{R}$ which satisfy the following boundary condition

$$q(t_0) = 0, \quad q(t_1) = 0.$$

The principle of least action for the motion of the particle reads as

$$\boxed{S[q] = \min!, \qquad q \in X.} \tag{7.65}$$

Let us first compute the functional derivative. To this end, fix $q, h \in X$. Set

$$\chi(\tau) := S[q + \tau h], \qquad \tau \in \mathbb{R}.$$

Differentiating with respect to the parameter τ at the point $\tau = 0$,

$$\chi'(0) = S'[q](h) = \int_{t_0}^{t_1} \Big(m\dot{q}(t)\dot{h}(t) - \kappa q(t)h(t) + F(t)h(t) \Big) dt.$$

Observing the boundary condition $h(t_0) = h(t_1) = 0$, integration by parts yields

$$S'[q](h) = \int_{t_0}^{t_1} \left(-m\ddot{q}(t) - \kappa q(t) + F(t) \right) h(t) dt.$$

Hence, for all $t \in [t_0, t_1]$, we obtain the partial functional derivative

$$\frac{\delta S[q]}{\delta q(t)} = -m\ddot{q}(t) - \kappa q(t) + F(t).$$

Theorem 7.27 *Each solution $q = q(t)$ of the principle of least action* (7.65) *satisfies the Euler–Lagrange equation*

$$\frac{\delta S[q]}{\delta q(t)} = 0 \qquad \textit{for all} \quad t \in [t_0, t_1].$$

Explicitly, $m\ddot{q}(t) = -\kappa q(t) + F(t)$.

This is the equation of a so-called harmonic oscillator with the restoring force $-\kappa q$ and the external force F. If the external force vanishes, $F = 0$, then we get the special solution

$$q(t) = \text{const} \cdot \sin(\omega t), \qquad t \in \mathbb{R}.$$

This motion represents an oscillation on the real line with the positive angular frequency ω given by the relation $\omega^2 = \kappa/m$.

Proof. Let $q \in X$ be a solution of the minimum problem (7.65). Fix the function $h \in \mathcal{D}(t_0, t_1)$. Introduce the function

$$\chi(\tau) := S[q + \tau h] \qquad \text{for all} \quad \tau \in \mathbb{R}.$$

Since $h(t_0) = h(t_1) = 0$, we get $h \in X$. Consequently, the simplified problem

$$\chi(\tau) = \min!, \qquad \tau \in \mathbb{R}$$

has the solution $\tau = 0$. By classical calculus,

$$\chi'(0) = 0.$$

This yields $S'[q](h) = 0$ for all $h \in \mathcal{D}(t_0, t_1)$. Hence

$$\int_{t_0}^{t_1} \frac{\delta S[q]}{\delta q(t)}\, h(t)dt = \frac{\delta S[q]}{\delta q}(h) = 0$$

for all $h \in \mathcal{D}(t_0, t_1)$. This implies

$$\frac{\delta S[q]}{\delta q(t)} = 0 \qquad \text{for all} \quad t \in [t_0, t_1],$$

by the variational lemma (Prop. 7.23). This finishes the classical proof invented by the young Lagrange in 1762. □

The same argument applies to all kinds of variational problems in mathematics and physics.

In the volumes of this treatise, we will encounter plenty of such variational problems.

The principle of critical action. Consider first the smooth real function $f : \mathbb{R} \to \mathbb{R}$. By definition, the problem

$$f(x) = \text{critical!}, \qquad x \in \mathbb{R}$$

is equivalent to $f'(x) = 0$. The solutions are called the critical points of the function f. This includes minimal points, maximal points, and horizontal inflection points. Similarly, by definition, the following critical point problem

$$\boxed{S[q] = \text{critical!}, \qquad q \in X} \tag{7.66}$$

is equivalent to $S'[q] = 0$. The same argument as above shows that each solution q of (7.66) satisfies the same Euler–Lagrange equation as obtained in Theorem 7.27.

The principle of least action versus the principle of critical action. Let us finish with the following remark. Consider first the real function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) := x^3$. The minimum problem

$$f(x) = \min!, \qquad x \in \mathbb{R}$$

has no solution, but the critical point problem

$$f(x) = \text{critical!}, \qquad x \in \mathbb{R}$$

has the solution $x = 0$. In fact, $f'(x) = 0$ implies $3x^2 = 0$, and hence $x = 0$. The same happens to more general variational problems. Therefore, we will not use the principle of least action, but the more general principle of critical action.

The language of physicists. In Sect. 11.2.3 on page 591, we will consider a formal definition of partial functional derivatives based on the Dirac delta function. This formal definition is used in most physics textbooks. The experience shows that both our rigorous approach introduced above and the formal approach based on the Dirac delta function lead to the same results.

7.20.3 Infinitesimal Transformations

> Infinitesimal symmetry transformations know much, but not all about global symmetry transformations.
>
> Folklore

In order to investigate the invariance of physical processes under symmetries, physicists simplify the considerations by using infinitesimal transformations. This theory was created by Sophus Lie (1849–1899) in about 1870. Let us discuss some basic ideas in rigorous terms.

> *Roughly speaking, infinitesimal transformations are obtained by neglecting terms of higher order than one.*

The prototype of infinitesimal transformations are infinitesimal rotations. Let us study this first.

Infinitesimal rotations. The transformation

$$\boxed{x' = x\cos\theta - y\sin\theta, \qquad y' = x\sin\theta + y\cos\theta} \tag{7.67}$$

represents a counterclockwise rotation about the origin in the plane, with rotation angle θ (Fig. 7.2). For small rotation angle θ, we get

$$x' = x - y\theta + o(\theta), \qquad y' = y + x\theta + o(\theta), \qquad \theta \to 0.$$

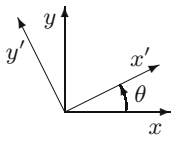


Fig. 7.2. Rotation

Thus, the linearization of the rotation (7.67) reads as

$$\boxed{x' = x - y\theta, \qquad y' = y + x\theta} \tag{7.68}$$

where θ is a fixed real number. This linear transformation is called an infinitesimal rotation with rotation angle θ. Physicists set

$$\delta\theta := \theta, \qquad \delta x := x' - x, \qquad \delta y := y' - y.$$

Therefore, the infinitesimal rotation (7.68) reads as

$$\boxed{\delta x = -y\delta\theta, \qquad \delta y = x\delta\theta.} \tag{7.69}$$

Invariant functions. The smooth function $f : \mathbb{R}^2 \to \mathbb{R}$ is called invariant under rotations iff for all rotation angles θ and all $x, y \in \mathbb{R}$, we have

$$f(x', y') = f(x, y).$$

Moreover, the function f is called invariant under infinitesimal rotations iff for all rotation angles $\delta\theta$ and all $x, y \in \mathbb{R}$, we get

$$f(x + \delta x, y + \delta y) = f(x, y) + o(\delta\theta), \qquad \delta\theta \to 0$$

where δx and δy are given by (7.69). As the prototype of the classical Lie theory on invariant functions, let us prove the following result. The point is that global symmetry properties can be described by a local equation, namely, a partial differential equation.

Proposition 7.28 *For each smooth function $f : \mathbb{R}^2 \to \mathbb{R}$, the following three conditions are equivalent.*

(i) *The function f is invariant under infinitesimal rotations.*

(ii) *The function f satisfies the Lie partial differential equation*

$$xf_y(x, y) - yf_x(x, y) = 0 \qquad \text{for all} \quad (x, y) \in \mathbb{R}^2. \tag{7.70}$$

(iii) *The function f is invariant under rotations.*

(iv) *In polar coordinates φ, r, the function f only depends on r.*

Proof. (i) $\Leftrightarrow$ (ii). By Taylor expansion, as $\delta\theta \to 0$,

$$\begin{aligned} f(x+\delta x, y+\delta y) - f(x,y) &= f_x(x,y)\delta x + f_y(x,y)\delta y + o(\theta) \\ &= (-yf_x(x,y) + xf_y(x,y))\delta\theta + o(\delta\theta). \end{aligned}$$

If (i) holds true, then as $\delta\theta \to 0$,

$$o(\delta\theta) = (-yf_x(x,y) + xf_y(x,y))\delta\theta + o(\delta\theta).$$

Dividing this by the real number $\delta\theta$ and letting $\delta\theta \to 0$, we get (ii).

Conversely, if (ii) holds true, then $f(x+\delta x, y+\delta y) - f(x,y) = o(\delta\theta)$ as $\delta\theta \to 0$. This implies (i).

(ii) $\Leftrightarrow$ (iii). Introduce polar coordinates,

$$x = r\cos\varphi, \qquad y = r\sin\varphi,$$

and set $F(\varphi, r) := f(x,y)$. By the chain rule,

$$F_\varphi(\varphi, r) = -f_x(x,y)r\sin\varphi + f_y(x,y)r\cos\varphi = -yf_x(x,y) + xf_x(x,y).$$

The function f is invariant under rotations iff the function F does not depend on the variable φ, that is, $F_\varphi(\varphi, r) = 0$ for all φ, r. This is equivalent to (ii).

(iii) $\Leftrightarrow$ (iv). This is obvious. □

Conservation laws. From the infinitesimal transformation (7.69) we get

$$\frac{\delta x}{\delta\theta} = -y, \quad \frac{\delta y}{\delta\theta} = x.$$

Letting $\theta \to 0$, we obtain the dynamical system

$$\boxed{\dot{x}(\theta) = -y(\theta), \quad \dot{y}(\theta) = x(\theta)} \tag{7.71}$$

which is called the characteristic system to the Lie equation (7.70). The solutions of the characteristic system (7.71) read as

$$x(\theta) = x_0\cos\theta - y_0\sin\theta, \qquad y(\theta) = x_0\sin\theta + y_0\cos\theta.$$

Regarding θ as time, $\theta := t$, we get rotations around the origin with angular velocity $\omega = 1$. A smooth function $f : \mathbb{R}^2 \to \mathbb{R}$ is a conservation law for this dynamical system (i.e., the function is constant along the trajectories) iff f is invariant under rotations. By Prop. 7.28, the function f satisfies the Lie partial differential equation (7.70).

The language of complex numbers. Set

$$z := x + y\mathrm{i}, \qquad z' := x' + y'\mathrm{i}.$$

Using the Euler equation $\mathrm{e}^{\mathrm{i}\theta} = \cos\theta + \mathrm{i}\sin\theta$, the rotation formula (7.67) can be elegantly written as

$$\boxed{z' = \mathrm{e}^{\mathrm{i}\theta} z.}$$

From the Taylor expansion $\mathrm{e}^{\mathrm{i}\theta} = 1 + \mathrm{i}\theta + o(\theta)$, we get

$$z' = z + \mathrm{i}z\theta + zo(\theta), \qquad \theta \to 0.$$

This yields the linearization

$$z' = z + \mathrm{i}z\theta.$$

Setting $\delta\theta := \theta$ and $\delta z := z' - z$, we get $z' = z + \delta z$ along with

$$\boxed{\delta z = \mathrm{i}z\delta\theta.}$$

This represents an infinitesimal rotation. The addition theorem for the exponential function yields the so-called group equation

$$\boxed{\mathrm{e}^{\mathrm{i}(\varphi+\theta)} = \mathrm{e}^{\mathrm{i}\varphi}\mathrm{e}^{\mathrm{i}\theta} \qquad \text{for all} \quad \varphi, \theta \in \mathbb{R}} \tag{7.72}$$

which tells us that the composition of rotations corresponds to the addition of the rotation angles φ and θ.

The language of matrices. Introducing the matrix

$$R(\theta) := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \qquad \text{for all} \quad \theta \in \mathbb{R}$$

the rotation formula (7.67) can be written as

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = R(\theta) \begin{pmatrix} x \\ y \end{pmatrix}.$$

The group equation (7.72) corresponds to

$$R(\varphi + \theta) = R(\varphi)R(\theta) \qquad \text{for all} \quad \varphi, \theta \in \mathbb{R}.$$

Explicitly,

$$R(\varphi + \theta) = \begin{pmatrix} \cos(\varphi+\theta) & -\sin(\varphi+\theta) \\ \sin(\varphi+\theta) & \cos(\varphi+\theta) \end{pmatrix}$$

and

$$R(\varphi)R(\theta) = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Hence

$$R(\varphi)R(\theta) = \begin{pmatrix} \cos\varphi\cos\theta - \sin\varphi\sin\theta & -\cos\varphi\sin\theta - \sin\varphi\cos\theta \\ \cos\varphi\sin\theta + \sin\varphi\cos\theta & \cos\varphi\cos\theta - \sin\varphi\sin\theta \end{pmatrix}.$$

From $R(\varphi+\theta) = R(\varphi)R(\theta)$ we obtain the addition theorem for trigonometric functions. For example,

$$\sin(\varphi+\theta) = \cos\varphi\sin\theta + \sin\varphi\cos\theta \qquad \text{for all} \quad \varphi,\theta \in \mathbb{R}.$$

Differentiation of $R(\theta)$ with respect of θ at the point $\theta = 0$ yields

$$R'(0) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Moreover, differentiating the group equation $R(\varphi+\theta) = R(\varphi)R(\theta)$ with respect to the angle φ at $\varphi = 0$, we obtain the differential equation

$$\boxed{R'(\theta) = AR(\theta), \qquad \theta \in \mathbb{R}, \qquad R(0) = I}$$

by setting $A := R'(0)$. This initial-value problem has the unique solution[42]

$$R(\theta) = \mathrm{e}^{\theta A} \qquad \text{for all} \quad \theta \in \mathbb{R}.$$

By Taylor expansion,

$$R(\theta) = I + \theta A + o(\theta), \qquad \theta \to 0.$$

Infinitesimal transformations correspond to the matrix θA. Explicitly, writing $\delta\theta$ instead of θ, we get

$$\begin{pmatrix} \delta x \\ \delta y \end{pmatrix} = \delta\theta \cdot A \begin{pmatrix} x \\ y \end{pmatrix} \qquad \text{for all} \quad \delta\theta \in \mathbb{R}.$$

This is equivalent to the infinitesimal rotation $\delta x = -y\delta\theta, \delta y = x\delta\theta$.

The infinitesimal strategy for one-parameter Lie groups and conservation laws. We now want to generalize the results for rotations in the plane to one-parameter Lie matrix groups. To this end, choose either $\mathbb{K} := \mathbb{R}$ (set of real numbers) or $\mathbb{K} := \mathbb{C}$ (set of complex numbers). Set

$$x := \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix}$$

where $x^1, \ldots, x^n \in \mathbb{K}$. Furthermore, fix an $(n \times n)$-matrix $A = (a_{ij})$ with entries $a_{11}, a_{12}, \ldots$ in $\mathbb{K}$. Now to the point. The family $\mathcal{G}$ of matrices,

[42] See Zeidler (1995), Vol. 1, Sect. 1.24.

$$\{\mathrm{e}^{tA}\}_{t\in\mathbb{R}}$$

is called a one-parameter Lie matrix group with the generator A. Obviously,

$$A = \frac{d\mathrm{e}^{tA}}{dt}\Big|_{t=0}.$$

For all parameters $t, s \in \mathbb{R}$, we have the so-called group equation (or causality equation)

$$\boxed{\mathrm{e}^{(t+s)A} = \mathrm{e}^{tA}\mathrm{e}^{sA}.}$$

By Taylor expansion, $\mathrm{e}^{tA} = I + tA + o(t)$ as $t \to 0$. Thus, the linearization of the transformation

$$\boxed{x' = \mathrm{e}^{tA}x, \qquad t \in \mathbb{R}}$$

looks like

$$x' = x + tAx, \qquad t \in \mathbb{R}.$$

Writing δt instead of t and setting $\delta := x' - x$, we get $x' = x + \delta x$ along with the so-called infinitesimal transformation

$$\boxed{\delta x = \delta t \cdot Ax, \qquad \delta t \in \mathbb{R}.} \tag{7.73}$$

By definition, the function $f : \mathbb{K}^n \to \mathbb{K}$ is called invariant under the Lie group $\mathcal{G}$ iff

$$f(\mathrm{e}^{tA}x) = f(x)$$

for all $x \in \mathbb{K}^n$ and all $t \in \mathbb{R}$. Moreover, the function f is called invariant under the infinitesimal transformations (7.73) iff

$$f(x + \delta x) - f(x) = o(\delta t), \qquad \delta t \to 0$$

for all $x \in \mathbb{K}^n$ and all real parameters δt.

Theorem 7.29 *For a smooth function $f : \mathbb{K}^n \to \mathbb{K}$, the following three statements are equivalent.*

(i) *The function f is invariant under the infinitesimal transformations of the one-parameter Lie group $\mathcal{G}$.*

(ii) *The function f satisfies the Lie equation*

$$f'(x)(Ax) = 0 \qquad \textit{for all} \quad x \in \mathbb{K}^n. \tag{7.74}$$

(iii) *The function f is invariant under the Lie group $\mathcal{G}$.*

Explicitly, the Lie equation (7.74) reads as

$$\left(\frac{\partial f(x)}{\partial x^1}, \ldots, \frac{\partial f(x)}{\partial x^n}\right) Ax = 0 \qquad \text{for all} \quad x \in \mathbb{K}^n. \tag{7.75}$$

From the physical point of view, the solutions of the Lie system (7.75) of first-order partial differential equations are the conservation laws to the dynamical system $x(t) = \mathrm{e}^{tA}x(0), t \in \mathbb{R}$ described by the system

$$\dot{x}(t) = Ax(t), \qquad t \in \mathbb{R} \tag{7.76}$$

of ordinary differential equations. Equation (7.76) is called the characteristic equation to (7.75); it follows from the infinitesimal transformation (7.73) by passing to the quotient $\frac{\delta x}{\delta t} = Ax$ and letting $\delta t \to 0$.

Proof. By Taylor expansion,

$$f(x + th) = f(x) + f'(x)(h) + o(t), \qquad t \to 0$$

with

$$f'(x)(h) = \left(\frac{\partial f(x)}{\partial x^1}, \ldots, \frac{\partial f(x)}{\partial x^n}\right) h = \sum_{k=1}^{n} \frac{\partial f(x)}{\partial x^k}\, h^k.$$

(i) $\Leftrightarrow$ (ii). Fix the point $x \in \mathbb{K}^n$. Introduce the function

$$\gamma(t) := f(x + tAx) \qquad \text{for all} \quad t \in \mathbb{R}.$$

By the chain rule, $\gamma'(0) = f'(x)(Ax)$. Condition (i) is equivalent to

$$\gamma(t) - \gamma(0) = o(t), \qquad t \to 0.$$

In turn, this is equivalent to $\gamma'(0) = 0$.

(ii) $\Leftrightarrow$ (ii). Now set $g(t) := f(\mathrm{e}^{tA}x)$ for all $t \in \mathbb{R}$. By the chain rule,

$$g'(t) = f'(\mathrm{e}^{tA}x)(A\mathrm{e}^{tA}x) \qquad \text{for all} \quad t \in \mathbb{R}.$$

Replacing x with $\mathrm{e}^{tA}x$, condition (ii) is equivalent to

$$g'(t) = 0 \qquad \text{for all} \quad t \in \mathbb{R}.$$

In turn, this is equivalent to $g(t) = \text{const}$ for all $t \in \mathbb{R}$, which corresponds to condition (iii). □

Lack of global information. The theory of Lie matrix groups is fundamental for elementary particle physics. This will be studied in the following volumes. In Sect. 5.7.1 we have discussed the crucial fact that the two matrix Lie groups $SO(3)$ and $SU(2)$ have isomorphic Lie algebras, and hence they are locally isomorphic, but they are not globally isomorphic. In this important case, the infinitesimal transformations do not know all about the global transformations. It is quite remarkable that nature sees this difference in terms of the electron spin.

7.20.4 Functional Integration

We want to generalize the notion of classical integral to *integrals over functionals*.

Finite measure integral. Let $\mathcal{Q}$ be a finite nonempty set. Suppose that a nonnegative real number $\mu(q)$ is assigned to each point q of $\mathcal{Q}$. We regard this number as the mass of the point q. For each function $F : \mathcal{Q} \to \mathbb{C}$, we define the integral with respect to the measure μ by

$$\boxed{\int_{\mathcal{Q}} F(q)d\mu(q) := \sum_{q\in\mathcal{Q}} F(q)\mu(q).}$$

Path integral. Let us now consider the special case where the set $\mathcal{Q}$ consists of a finite number of curves. To this end, we divide the time interval $[s,t]$ into small equidistant time intervals

$$s = t_0 < t_1 < \ldots < t_m = t$$

where $t_j := t_0 + j\Delta t$ for all $j = 0, 1, \ldots, m$. Consider continuous, piecewise linear curves

$$q : [s,t] \to \mathbb{C}$$

that attain some of the given discrete values $q_1, \ldots, q_N$ at the discrete time points $t_0, \ldots, t_m$ (Fig. 7.3). Let $\mathcal{Q}_{\Delta t}$ denote the set of all such curves. Suppose that we are given the functional $F : \mathcal{Q}_{\Delta t} \to \mathbb{C}$ which assigns a complex number $F(q)$ to each curve $q \in \mathcal{Q}_{\Delta t}$ (e.g., the curve length). Then, the measure integral

$$\boxed{\int_{\mathcal{Q}_{\Delta t}} F(q)d\mu(q) := \sum_{q\in\mathcal{Q}_{\Delta t}} F(q)\mu(q)} \tag{7.77}$$

is called a path integral. In what follows we will use the terms 'functional integral' and 'path integral' synonymously.

7.21 The Discrete Feynman Path Integral

> I found myself thinking of a large number of integrals, one after the other in sequence. In the integrand was the product of the exponentials, which, of course, was the exponential of the sum of the terms like $\epsilon\mathcal{L}$. Now $\mathcal{L}$ is the Lagrangian and ϵ is like the time interval dt, so that if you took a sum of such terms, that's exactly like an integral. That's like Riemann's formula for the integral $\int \mathcal{L}dt$; you just take the value at each point and add them together. We are to take the limit as $\epsilon \to 0$, of course. Therefore, the connection between the wave function of one instant and and the wave

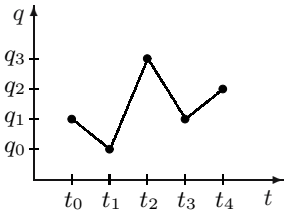


Fig. 7.3. Piecewise linear curve

> function of another instant a finite time later could be obtained by an infinite number of integrals of $e^{iS/\hbar}$, where S is the action expression... This led later on to the idea of the amplitude of the path; that for each possible way that the particle can go from one point to another in space-time, there's an amplitude $e^{iS/\hbar}$, where S is the action along the path. Amplitudes from various paths superpose by addition. This then is another, a third way, of describing quantum mechanic, which looks quite different than that of Heisenberg or Schrödinger, but is equivalent to them.[43]
>
> Richard Feynman (1918–1988)
> *Nobel Lecture in* 1965

In modern textbooks in physics, the approach to quantum field theory is mainly based on functional integrals which can be viewed as generalized path integrals. In the following we want to discuss the basic ideas.

7.21.1 The Magic Feynman Propagator Formula

The magic Feynman propagator formula is given by the following discrete path integral

$$\boxed{\langle Q_{\text{out}}|P(t_{\text{out}},t_{\text{in}})Q_{\text{in}}\rangle = \int_{q\in\Pi_{\Delta t}(q_{\text{in}},q_{\text{out}})} e^{iS[q]/\hbar}\, d\mu[q].} \tag{7.78}$$

In terms of physics, this is the transition amplitude of the propagator with respect to the finite time interval $[t_{\text{in}}, t_{\text{out}}]$. This transition amplitudes describes the dynamics of the quantum system. The real number

$$|\langle Q_{\text{out}}|P(t_{\text{out}},t_{\text{in}})Q_{\text{in}}\rangle|^2$$

is the transition probability from the initial state Q_{in} at the initial time t_{in} to the final state Q_{out} at the final time t_{out}. The explicit form of

- the action $S[q]$ along the curve $q : [t_{\text{in}}, t_{\text{out}}] \to \mathbb{R}$,
- and the measure $\mu[q]$ of the curve $q = q(t)$

[43] Nobel Prize Lectures. Reprinted by permission of the Nobel Foundation, Stockholm.

are given by[44]

$$\langle Q(t_{r+1})|\mathrm{e}^{-\mathrm{i}\Delta t\cdot H/\hbar}Q(t_r)\rangle = \varrho_r \mathrm{e}^{\mathrm{i}\Delta S_r[q]/\hbar}, \tag{7.79}$$

and

$$\boxed{S[q] := \sum_{r=1}^{m} \Delta S_r[q],}$$

as well as

$$\boxed{\mu[q] := \prod_{r=1}^{m} \varrho_r.}$$

The path integral (7.78) represents a finite sum of the form

$$\sum_{q\in\Pi_{\Delta t}(q_{\mathrm{in}},q_{\mathrm{out}})} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\,\mu[q].$$

The real number $\Delta S_r[q]$ is called the local action of the curve q on the time interval $[t_r, t_{r+1}]$. By definition, the total action of the curve q is then obtained by summing up the local actions. The precise formulation will be given in (A1) through (A5) below.

> *The importance of the magic formula* (7.78) *relies on the fact that it represents the crucial transition amplitude as the superposition of simpler physical effects given by* $e^{\mathrm{i}\Delta S_r[q]/\hbar}$ *which are acting on small time intervals* $[t_r, t_{r+1}]$ *of length* Δt.

This can be regarded as a generalization of Newton's classical strategy of describing the dynamics of mechanical systems by their behavior during infinitesimally small time intervals. We make the following assumptions.

(A1) Energy operator: We are given a linear self-adjoint operator

$$H : X \to X$$

on the N-dimensional complex Hilbert space X, $1 \le N < \infty$. This operator is called the energy operator (or the Hamiltonian). There exists a complete orthonormal basis $|E_1\rangle, \ldots, |E_N\rangle$ of the Hilbert space X such that

$$H|E_j\rangle = E_j|E_j\rangle, \qquad j = 1, \ldots, N.$$

The real eigenvalues $E_1, \ldots, E_N$ of the operator H are the possible energy values of the system with the corresponding energy eigenstates $|E_1\rangle, \ldots, |E_N\rangle$.

[44] Here, $\Delta S_r[q]/\hbar$ and ϱ_r are the argument and the modulus of the complex number $\langle Q(t_{r+1})|\mathrm{e}^{-\mathrm{i}\Delta t\cdot H/\hbar}Q(t_r)\rangle$, respectively. Equation (7.79) determines the real argument $\Delta S_r[q]/\hbar$ only up to a multiple of the number 2π. The same is true for $S[q]/\hbar$. However, the point is that the factor $\mathrm{e}^{\mathrm{i}S[q]/\hbar}$ arising in the magic Feynman propagator formula (7.78) is uniquely determined.

(A2) Position operator: Let $Q_1, \ldots, Q_N$ denote an arbitrary orthonormal basis of the Hilbert space X. In order to get some intuitive interpretation, suppose that there exists a linear self-adjoint operator $Q : X \to X$ such that

$$QQ_j = q_j Q_j, \qquad j = 1, \ldots N.$$

We regard the operator Q as position operator, and the eigenvalues $q_1, \ldots, q_N$ may be regarded as the positions of a quantum particle on the real line. The position q_j corresponds to the normalized state Q_j. As in the Dirac calculus, we will use the notation $|Q_j\rangle$ instead of Q_j.

(A3) Feynman propagator: Let us also introduce the propagator

$$\boxed{P(t,s) := \mathrm{e}^{-\mathrm{i}(t-s)H/\hbar}, \qquad s,t \in \mathbb{R}.} \tag{7.80}$$

The truncated operator

$$P^+(t,s) := \theta(t-s)P(t,s), \qquad t,s \in \mathbb{R}$$

is called the retarded propagator (or the Feynman propagator).

(A4) Piecewise linear paths: Decompose the given time interval $[t_{\mathrm{in}}, t_{\mathrm{out}}]$ into m pieces of equal length,

$$t_j = t_{\mathrm{in}} + j\Delta t, \qquad j = 0, 1, \ldots, m$$

where $\Delta t := (t_{\mathrm{out}} - t_{\mathrm{in}})/m$. As pictured in Fig. 7.3 on page 417, we consider all possible curves $q : [t_{\mathrm{in}}, t_{\mathrm{out}}] \to \mathbb{R}$ which have the following properties:

- The curves are piecewise linear and continuous. In other words, they are polygonal curves in the plane.
- At the node points $t_0 = t_{\mathrm{in}}, t_1, \ldots, t_{m-1}, t_m = t_{\mathrm{out}}$, the curves attain some of the real values $q_1, \ldots, q_N$.
- Both the initial point and the final point of the curves are fixed, i.e.,

$$q(t_{\mathrm{in}}) := q_{\mathrm{in}}, \quad q(t_{\mathrm{out}}) := q_{\mathrm{out}}.$$

The space of these curves is denoted by $\Pi_{\Delta t}(q_{\mathrm{in}}, q_{\mathrm{out}})$.

(A5) Transition amplitudes: Let $q \in \Pi_{\Delta t}(q_{\mathrm{in}}, q_{\mathrm{out}})$. Suppose that the curve has the node points

$$q(t_r) = q_j, \qquad q(t_{r+1}) = q_k$$

at time t_r and t_{r+1}, respectively. We set $Q(t_r) := Q_j$ and $Q(t_{r+1}) := Q_k$. Then

$$\langle Q(t_{r+1})|\mathrm{e}^{-\mathrm{i}\Delta t H/\hbar} Q(t_r)\rangle = \langle Q_k|\mathrm{e}^{-\mathrm{i}\Delta t \cdot H/\hbar} Q_j\rangle.$$

This complex number can be written as $\varrho_r \mathrm{e}^{\mathrm{i}\Delta S_r[q]/\hbar}$ with the real argument $\Delta S_r[q]/\hbar$ and the modulus ϱ_r.

By the Dirac calculus, $\sum_{k=1} |E_k\rangle\langle E_k| = I$. Hence $\langle Q(t_{r+1})|\mathrm{e}^{-\mathrm{i}\Delta t\cdot H/\hbar}Q(t_r)\rangle$ is equal to

$$\sum_{j,k=1}^{N} \langle Q(t_{r+1})|E_j\rangle\langle E_j|\mathrm{e}^{-\mathrm{i}\Delta t\cdot H/\hbar}|E_k\rangle\langle E_k|Q(t_r)\rangle.$$

Since $\langle E_j|\mathrm{e}^{-\mathrm{i}\Delta t\cdot H/\hbar}|E_k\rangle = \mathrm{e}^{-\mathrm{i}E_k\Delta t/\hbar}\langle E_j|E_k\rangle = \mathrm{e}^{-\mathrm{i}E_k\Delta t/\hbar}\,\delta_{jk}$, we get

$$\varrho_r\mathrm{e}^{\mathrm{i}\Delta S_r[q]/\hbar} = \sum_{j=1}^{N} \mathrm{e}^{-\mathrm{i}E_j\Delta t/\hbar}\langle Q(t_{r+1})|E_j\rangle\langle E_j|Q(t_r)\rangle.$$

Theorem 7.30 *The magic Feynman formula* (7.78) *holds true.*

Proof. To simplify notation, set $\hbar := 1$. To begin with, let $m = 2$, that is, $t_0 < t_1 < t_2$, and $\Delta t = t_1 - t_0 = t_2 - t_1$. According to the Dirac calculus,

$$\sum_{j=1}^{N} |Q_j\rangle\langle Q_j| = I.$$

By (7.80), the propagator satisfies the relation

$$P(t_2, t_0) = P(t_2, t_1)P(t_1, t_0).$$

Thus, Dirac's substitution trick (7.20) yields

$$\langle Q_{\mathrm{out}}|P(t_2, t_0)|Q_{\mathrm{in}}\rangle = \sum_{j=1}^{N}\langle Q_{\mathrm{out}}|P(t_2, t_1)Q_j\rangle\langle Q_j|P(t_1, t_0)Q_{\mathrm{in}}\rangle.$$

An arbitrary curve $q \in \mathit{\Pi}_{\Delta\mathrm{t}}(q_{\mathrm{in}}, q_{\mathrm{out}})$ has the form

$$q(t_0) = q_{\mathrm{in}}, \quad q(t_1) = q_j, \quad q(t_2) = q_{\mathrm{out}}.$$

Therefore, the transition amplitude $\langle Q_{\mathrm{out}}|P(t_2, t_0)|Q_{\mathrm{in}}\rangle$ is equal to the sum

$$\sum_{q\in\mathit{\Pi}_{\Delta\mathrm{t}}(q_{\mathrm{in}},q_{\mathrm{out}})} \langle Q(t_2)|P(t_2, t_1)|Q(t_1)\rangle\langle Q(t_1)|P(t_1, t_0)|Q(t_0)\rangle$$

over all possible curves. This is the basic trick. This tells us that the transition amplitude $\langle Q_{\mathrm{out}}|P(t_2, t_0)|Q_{\mathrm{in}}\rangle$ is equal to

$$\sum_{q\in\mathit{\Pi}_{\Delta\mathrm{t}}(q_{\mathrm{in}},q_{\mathrm{out}})} \langle Q(t_2)|\mathrm{e}^{-\mathrm{i}\Delta tH}|Q(t_1)\rangle\langle Q(t_1)|\mathrm{e}^{-\mathrm{i}\Delta tH}|Q(t_0)\rangle$$
$$= \sum_{q\in\mathit{\Pi}_{\Delta\mathrm{t}}(q_{\mathrm{in}},q_{\mathrm{out}})} \mathrm{e}^{\mathrm{i}(\Delta S_1[q]+\Delta S_2[q])}\varrho_1\varrho_2.$$

This finishes the proof for $m = 2$. Now let $m = 3$, that is, $t_0 < t_1 < t_2 < t_3$. By Dirac's substitution trick, the transition amplitude $\langle Q_{\text{out}}|P(t_3, t_0)|Q_{\text{in}}\rangle$ is equal to

$$\sum_{j,k=1}^{N} \langle Q_{\text{out}}|P(t_3, t_2)Q_j\rangle\langle Q_j|P(t_2, t_1)Q_k\rangle\langle Q_k|P(t_1, t_0)|Q_{\text{in}}\rangle.$$

The proof proceeds now as above. A similar argument can be applied in the case where $m = 4, 5, \ldots$ □

The Feynman propagator kernel. Let us now explain the basic trick of Feynman's approach to quantum physics. For a given Hamiltonian H, the dynamics of the quantum system is described by the equation

$$\boxed{\psi(t_{\text{out}}) = P(t_{\text{out}}, t_{\text{in}})\psi(t_{\text{in}}), \qquad t_{\text{out}} > t_{\text{in}}} \tag{7.81}$$

where $\psi(t)$ denotes the state of the system at time t. Introducing the Fourier coefficients

$$\psi_j(t) := \langle Q_j|\psi(t)\rangle, \qquad j = 1, \ldots, N$$

and the so-called Feynman propagator kernel

$$\mathcal{K}(Q_j, t_{\text{out}}; Q_k, t_{\text{in}}) := \langle Q_j|P(t_{\text{out}}, t_{\text{in}})Q_k\rangle,$$

we get the Feynman propagator kernel equation

$$\boxed{\psi_j(t_{\text{out}}) = \sum_{k=1}^{N} \mathcal{K}(Q_j, t_{\text{out}}; Q_k, t_{\text{in}})\psi_k(t_{\text{in}}), \qquad t_{\text{out}} > t_{\text{in}}} \tag{7.82}$$

for $j = 1, \ldots, N$. This equation is equivalent to (7.81). In fact, by the Dirac calculus,

$$\langle Q_j|\psi(t_{\text{out}})\rangle = \sum_{k=1}^{N} \langle Q_j|P(t_{\text{out}}, t_{\text{in}})Q_k\rangle\langle Q_k|\psi(t_{\text{in}})\rangle.$$

Using the magic Feynman propagator formula (7.78), the Feynman propagator kernel can be expressed by a path integral. Explicitly,

$$\mathcal{K}(Q_j, t_{\text{out}}; Q_k, t_{\text{in}}) = \int_{q \in \Pi_{\Delta t}(q_j, q_k)} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, d\mu[q]. \tag{7.83}$$

Rescaling. For a given positive number Δq, let us consider the special case where

$$q_j := q_1 + (j-1)\Delta q, \qquad j = 1, \ldots, N.$$

Define the rescaled states

$$|q_j\rangle := \frac{|Q_j\rangle}{\sqrt{\Delta q}}, \qquad j = 1, \ldots N.$$

Introducing the discrete Dirac delta function,

$$\delta_{\Delta q}(q_j - q_k) := \frac{\delta_{jk}}{\Delta q}, \qquad j,k = 1,\ldots,N, \tag{7.84}$$

we get the rescaled orthogonality condition

$$\langle q_j | q_k \rangle = \delta_{\Delta q}(q_j - q_k), \qquad j,k = 1,\ldots,N$$

along with the rescaled completeness relation

$$\sum_{j=1}^{N} |q_j\rangle\langle q_j| \; \Delta q = I.$$

The magic Feynman propagator formula (7.78) now reads as

$$\langle q_{\rm out}|P(t_{\rm out},t_{\rm in})|q_{\rm in}\rangle = \sum_{q\in \Pi_{\Delta t}(q_{\rm in},q_{\rm out})} {\rm e}^{{\rm i}\mathcal{S}[q]/\hbar}\; \mu(q)(\Delta q)^{m-1}.$$

Perspective. Roughly speaking, physicists use the formal limits

$$N,m \to +\infty, \quad \Delta t \to 0, \quad \Delta q \to 0, \quad q_1 \to -\infty, \quad q_N \to +\infty$$

in order to investigate quantum particles on the real line in terms of path integrals. In particular, the discrete Dirac delta function passes over to the Dirac delta function in a formal manner,

$$\delta_{\Delta q} \to \delta \qquad \text{as} \quad \Delta q \to 0.$$

7.21.2 The Magic Formula for Time-Ordered Products

We now modify the magic formula (7.78) by considering the following transition amplitude:

$$\langle Q_{\rm out}|P(t_{\rm out},t_r)QP(t_r,t_{\rm in})Q_{\rm in}\rangle = \int_{q\in \Pi_{\Delta t}(q_{\rm in},q_{\rm out})} q(t_r){\rm e}^{{\rm i}S[q]/\hbar}\; d\mu[q]. \tag{7.85}$$

This means that the physical influence of the operator Q acting at time t_r can be described simply by inserting the value of the curve $q(t_r)$ at time t_r into the path integral. Here, we use again the decomposition $t_0 < t_1 < \ldots < t_m$ with $t_{\rm in} := t_0$ and $t_{\rm out} := t_m$. Moreover, we choose t_r for fixed $r = 1,\ldots,m-1$.

Theorem 7.31 *Formula* (7.85) *holds true.*

Proof. To explain the basic idea, choose $t_0 < t_1 < t_2$, and set $t_r := t_1$. By the Dirac calculus, $\sum_{j=1}^{N} |Q_j\rangle\langle Q_j| = I$. Hence the transition amplitude

$$\langle Q_{\text{out}}|P(t_{\text{out}}, t_1)QP(t_1, t_{\text{in}})Q_{\text{in}}\rangle$$

is equal to

$$\sum_{j,k=1}^{N} \langle Q_{\text{out}}|P(t_{\text{out}}, t_1)Q_j\rangle\langle Q_j|QQ_k\rangle\langle Q_k|P(t_1, t_{\text{in}})Q_{\text{in}}\rangle.$$

Since $\langle Q_j|QQ_k\rangle = q_k\langle Q_j|Q_k\rangle = q_k\delta_{jk}$, this is equal to

$$\sum_{j=1}^{N} q_j\langle Q_{\text{out}}|P(t_{\text{out}}, t_1)Q_j\rangle\langle Q_j|P(t_1, t_{\text{in}})Q_{\text{in}}\rangle,$$

which is the expression for the path integral from the proof of Theorem 7.30, up to the additional factor q_j which corresponds to the value $q(t_1)$ of the curve q at time t_1. □

The same argument applies to transition amplitudes where the operator Q acts at different times $t_{r_1} < t_{r_2} < \ldots < t_{\text{r}_1}$. For example, let $t_{\text{in}} < t_1 < \ldots < t_m = t_{\text{out}}$, and choose $t_{r_1} < t_{r_2}$ with $1 < r_1 < r_2 < m$. Then, the time-ordered transition amplitude

$$\boxed{\langle Q_{\text{out}}|P(t_{\text{out}}, t_{r_2})QP(t_{r_2}, t_{r_1})QP(t_{r_1}, t_{\text{in}})Q_{\text{in}}\rangle}$$

is equal to the path integral

$$\boxed{\int_{q\in\Pi_{\Delta t}(q_{\text{in}},q_{\text{out}})} q(t_{r_2})q(t_{r_1})\mathrm{e}^{\mathrm{i}S[q]/\hbar}\, d\mu[q].} \tag{7.86}$$

The proof proceeds similarly to the proof of Theorem 7.31 by inserting the identity $\sum_{j=1}^{N} |Q_j\rangle\langle Q_j| = I$ at time t_{r_2} and time t_{r_1}.

7.21.3 The Trace Formula

The following formula plays an important role in quantum statistics. For the trace of the propagator $P(t_{\text{out}}, t_{\text{in}}) = \mathrm{e}^{-\mathrm{i}(t_{\text{out}}-t_{\text{in}})H/\hbar}$, we get

$$\boxed{\operatorname{tr} P(t_{\text{out}}, t_{\text{in}}) = \int_{q\in\Pi_{\Delta t,\text{periodic}}} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, d\mu[q].} \tag{7.87}$$

Here, we sum over all of the curves $q \in \Pi_{\Delta t}(q_{\text{in}}, q_{\text{out}})$ with $q_{\text{in}} = q_{\text{out}}$.
Proof. Set $Q_{\text{in}} := Q_j$ and $Q_{\text{out}} := Q_j$ for fixed $j = 1, \ldots, N$. By (7.78),

$$\langle Q_j|P(t_{\text{out}}, t_{\text{in}})Q_j\rangle = \int_{\Pi_{\Delta t}(q_j,q_j)} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, d\mu[q].$$

Finally, note that $\operatorname{tr} P(t_{\text{out}}, t_{\text{in}}) = \sum_{j=1}^{N}\langle Q_j|P(t_{\text{out}}, t_{\text{in}})Q_j\rangle$. □

7.22 Causal Correlation Functions

It is our goal to compute causal correlation functions which lie at the heart of quantum field theory. Assume that the energies are ordered by

$$E_1 \leq E_2 \leq \ldots \leq E_N.$$

The least energy value E_1 is also called the ground state energy (or the vacuum energy) of the system. For the following, we make the crucial assumption that the ground state energy is not degenerate. By definition, this means that the eigenvalue E_1 of the energy operator (Hamiltonian) H is simple, that is,

$$E_1 < E_2 \leq \ldots \leq E_N.$$

Thus, the normalized states of least energy are of the form $\varrho|E_1\rangle$ where ϱ is a complex number with $|\varrho| = 1$. For arbitrary real times $s < t$, the causal correlation function is defined by

$$\boxed{C(t,s) := \langle E_1|\mathrm{e}^{\mathrm{i}tH/\hbar}Q\mathrm{e}^{-\mathrm{i}tH/\hbar} \cdot \mathrm{e}^{\mathrm{i}sH/\hbar}Q\mathrm{e}^{-\mathrm{i}sH/\hbar}|E_1\rangle.}$$

This function is also called the 2-point Green's function. Using the propagator, this can be written as

$$C(t,s) := \langle E_1|P(0,t)QP(t,s)QP(s,0)|E_1\rangle.$$

In terms of physics, $C(t,s)$ describes the correlation between quantum fluctuations of the ground state $|E_1\rangle$ at time s and the later time t. Because of this time-ordering, we speak of causal correlation functions. The quantum fluctuations refer to the operator Q acting at time s and the later time t.

It is our goal to reduce the computation of causal correlation functions to simpler expressions.

This will correspond to both the magic Wick rotation formula (7.89) below and the magic Gell-Mann–Low perturbation formula from Theorem 7.33 on page 427. To this end, we will use the trick of Wick rotation. This trick is based on the time transformation

$$\boxed{\tau = t(1 - \varepsilon\mathrm{i}), \qquad t \in \mathbb{R}} \tag{7.88}$$

for fixed real parameter $\varepsilon > 0$. The basic idea is to simplify computations by passing from real time t to complex time τ. In addition, we will apply the limits

$$t_{\mathrm{out}} \to +\infty, \qquad t_{\mathrm{in}} \to -\infty, \qquad \varepsilon \to +0.$$

Introducing the modified correlation function

$$C_\varepsilon(t,s) := C(t(1 - \varepsilon\mathrm{i}), s(1 - \varepsilon\mathrm{i})), \qquad s < t,$$

we obtain the following limit relation

$$\boxed{C(t,s) = \lim_{\varepsilon \to +0} C_\varepsilon(t,s), \qquad s < t.}$$

We will now show how to compute the modified correlation function $C_\varepsilon(t,s)$ by applying the limits $t_{\rm out} \to +\infty$ and $t_{\rm in} \to -\infty$ to appropriate transition amplitudes.

7.22.1 The Wick Rotation Trick for Vacuum Expectation Values

We are given the states $Q_{\rm out}$ and $Q_{\rm in}$ as eigenstates of the linear self-adjoint operator $Q : X \to X$ on the N-dimensional complex Hilbert space X where $1 \le N < \infty$. Assume that $\langle Q_{\rm out}|E_1\rangle \neq 0$ and $\langle E_1|Q_{\rm in}\rangle \neq 0$. Fix the parameter $\varepsilon > 0$, and choose the real times s and t with $s < t$. The following formula (7.89) is called the magic Wick rotation formula.

Theorem 7.32 *The modified causal correlation function is equal to the limit*

$$C_\varepsilon(t,s) = \lim_{t_{\rm out} \to +\infty} \lim_{t_{\rm in} \to -\infty} \frac{\mathcal{A}}{\mathcal{B}} \tag{7.89}$$

where

$$\mathcal{A} := \langle Q_{\rm out}|P(\tau_{\rm out},\tau)QP(\tau,\sigma)QP(\sigma,\tau_{\rm in})|Q_{\rm in}\rangle$$

along with $\tau := t(1-\varepsilon{\rm i})$ *and* $\sigma := s(1-\varepsilon{\rm i})$, *and*

$$\mathcal{B} := \langle Q_{\rm out}|P(\tau_{\rm out},\tau_{\rm in})|Q_{\rm in}\rangle.$$

Observe that the key formula (7.89) represents the vacuum expectation value $C_\varepsilon(t,s) = \langle E_1|\ldots|E_N\rangle$ by the transition amplitudes $\mathcal{A}$ and $\mathcal{B}$ extended over an infinite time interval. Before proving this, let us reformulate this in terms of path integrals. In fact, we obtain

$$\mathcal{A} = \int_{q \in \Pi_{\Delta\tau}(q_{\rm in},q_{\rm out})} q(\tau)q(\sigma){\rm e}^{{\rm i}S[q]/\hbar}d\mu[q]$$

and

$$\mathcal{B} = \int_{q \in \Pi_{\Delta\tau}(q_{\rm in},q_{\rm out})} {\rm e}^{{\rm i}S[q]/\hbar}d\mu[q].$$

These discrete path integrals are obtained from the corresponding discrete path integrals introduced in Sect. 7.21 by replacing real time t by the complex time $t(1-\varepsilon{\rm i})$ for fixed parameter $\varepsilon > 0$. In particular, $t = t_{r_1}$ and $s = t_{r_2}$ fit the decomposition of the time interval, that is, $r_1 < r_2$, and

$$t_0 := t_{\rm in} < t_1 < \ldots < t_m := t_{\rm out}.$$

Proof of Theorem 7.32. (I) Basic trick. Let $a_1, \dots, a_N$ be complex numbers. Fix $\varepsilon > 0$. Then

$$\lim_{t\to+\infty} \frac{\sum_{j=1}^N \mathrm{e}^{-\varepsilon t E_j} a_j}{\sum_{j=1}^N \mathrm{e}^{-\varepsilon t E_j}} = a_1.$$

In fact, this follows from the decomposition

$$\mathrm{e}^{-\varepsilon t E_j} = \mathrm{e}^{-\varepsilon t E_1}\ \mathrm{e}^{-\varepsilon t (E_j - E_1)}$$

along with $E_1 < E_2 \le \ldots \le E_N$. Hence

$$\lim_{t\to+\infty} \frac{\mathrm{e}^{-\varepsilon t E_1}(a_1 + \sum_{j=2}^N \mathrm{e}^{-\varepsilon t (E_j - E_1)} a_j)}{\mathrm{e}^{-\varepsilon t E_1}(1 + \sum_{j=2}^N \mathrm{e}^{-\varepsilon t (E_j - E_1)})} = a_1.$$

(II) Similarly, for given complex numbers a_{jk}, b_{jk}, $j,k = 1, \dots, N$ with $a_{11} \neq 0$, we obtain

$$\lim_{t\to+\infty} \lim_{s\to-\infty} \frac{\sum_{j,k=1}^N \mathrm{e}^{-\mathrm{i}t(1-\varepsilon\mathrm{i})E_j} \mathrm{e}^{\mathrm{i}s(1-\varepsilon\mathrm{i})E_k} a_{jk} b_{jk}}{\sum_{j,k=1}^N \mathrm{e}^{-\mathrm{i}t(1-\varepsilon\mathrm{i})E_j} \mathrm{e}^{\mathrm{i}s(1-\varepsilon\mathrm{i})E_k} a_{jk}} = b_{11}.$$

(III) To simplify notation, set $\hbar := 1$. Observe that

$$\mathcal{A} := \langle \mathrm{e}^{\mathrm{i}\tau_{\mathrm{out}} H} Q_{\mathrm{out}} | \mathrm{e}^{\mathrm{i}\tau H} Q \mathrm{e}^{-\mathrm{i}\tau H} \cdot \mathrm{e}^{\mathrm{i}\sigma H} Q \mathrm{e}^{-\mathrm{i}\sigma H} \cdot \mathrm{e}^{\mathrm{i}\tau_{\mathrm{in}} H} Q_{\mathrm{in}} \rangle.$$

By the Dirac calculus,

$$\mathcal{A} = \sum_{j,k=1}^N \langle \mathrm{e}^{\mathrm{i}\tau_{\mathrm{out}} H} Q_{\mathrm{out}} | E_j \rangle \langle E_j | \mathrm{e}^{\mathrm{i}\tau H} Q \mathrm{e}^{-\mathrm{i}\tau H} \cdot \mathrm{e}^{\mathrm{i}\sigma H} Q \mathrm{e}^{-\mathrm{i}\sigma H} | E_k \rangle \langle E_k | \mathrm{e}^{\mathrm{i}\tau_{\mathrm{in}} H} Q_{\mathrm{in}} \rangle.$$

Since $\langle \mathrm{e}^{\mathrm{i}\tau_{\mathrm{out}} H} Q_{\mathrm{out}} | E_j \rangle = \langle Q_{\mathrm{out}} | \mathrm{e}^{-\mathrm{i}\tau_{\mathrm{out}} H} | E_j \rangle = \mathrm{e}^{-\mathrm{i}E_j \tau_{\mathrm{out}}} \langle Q_{\mathrm{out}} | E_j \rangle$, we get

$$\mathcal{A} = \sum_{j,k=1}^N \mathrm{e}^{-\mathrm{i}\tau_{\mathrm{out}} E_j} \langle Q_{\mathrm{out}} | E_j \rangle \mathrm{e}^{\mathrm{i}\tau_{\mathrm{in}} E_k} \langle E_k | Q_{\mathrm{in}} \rangle c_{jk}$$

with

$$c_{jk} := \langle E_j | \mathrm{e}^{\mathrm{i}\tau H} Q \mathrm{e}^{-\mathrm{i}\tau H} \cdot \mathrm{e}^{\mathrm{i}\sigma H} Q \mathrm{e}^{-\mathrm{i}\sigma H} | E_k \rangle.$$

Similarly,

$$\mathcal{B} = \sum_{j,k=1}^N \mathrm{e}^{-\mathrm{i}\tau_{\mathrm{out}} E_j} \langle Q_{\mathrm{out}} | E_j \rangle \mathrm{e}^{\mathrm{i}\tau_{\mathrm{in}} E_k} \langle E_k | Q_{\mathrm{in}} \rangle.$$

By (II), the quotient $\mathcal{A}/\mathcal{B}$ goes to c_{11} as $t_{\mathrm{out}} \to +\infty$ and $t_{\mathrm{in}} \to -\infty$. □

7.22.2 The Magic Gell-Mann–Low Reduction Formula

> The magic Dyson formula for the S-matrix and the magic Gell-Mann–Low formula for the causal correlation functions are the key to applying perturbation theory in quantum field theory.
>
> Folklore

Consider the typical perturbation of the Hamiltonian,

$$\boxed{H = H_0 + \kappa V,}$$

where $H_0, V : X \to X$ are linear self-adjoint operators on the given complex N-dimensional Hilbert space X, $N = 1, 2, \ldots$ The real constant $\kappa \geq 0$ is called the coupling constant. It is our goal to reformulate the Wick rotation trick from the preceding section in terms of the Dyson S-matrix operator,

$$\boxed{\mathsf{S}(t,s) = \mathcal{T}\mathrm{e}^{-\frac{\mathrm{i}\kappa}{\hbar}\int_s^t V_I(\eta)d\eta},} \tag{7.90}$$

introduced in Sect. 7.18. Here, $V_I(t) := \mathrm{e}^{\mathrm{i}tH_0/\hbar}V\mathrm{e}^{-\mathrm{i}tH_0/\hbar}$. Let

$$H_0|E_j^0\rangle = E_j^0|E_j^0\rangle, \qquad j = 1, \ldots N,$$

that is, $|E_1^0\rangle, \ldots |E_N^0\rangle$ forms a complete orthonormal system of eigenvectors of the unperturbed Hamiltonian H_0. In addition, assume that the ground state of the unperturbed Hamiltonian is not degenerate, that is, the lowest eigenvalue E_1^0 is simple,

$$E_1^0 < E_2^0 \leq \ldots E_N^0.$$

Similarly, let

$$H|E_j\rangle = E_j|E_j\rangle, \qquad j = 1, \ldots, N,$$

that is, $|E_1\rangle, \ldots, |E_N\rangle$ forms a complete orthonormal system of eigenvectors of the perturbed Hamiltonian H. Furthermore, assume that E_1 is simple and that

$$\langle E_1|E_1^0\rangle \neq 0.$$

By classical perturbation theory, these two conditions are always satisfied if the coupling constant κ is sufficiently small.[45]

Theorem 7.33 *We are given the real times t and s with $s < t$. For fixed parameter $\varepsilon > 0$, set $\tau := t(1 - \varepsilon i)$ and $\sigma := s(1 - \varepsilon i)$. Then, the modified correlation function is given by the following limit*

$$C_\varepsilon(t,s) = \lim_{t_{\mathrm{out}} \to +\infty} \lim_{t_{\mathrm{in}} \to -\infty} \frac{\langle E_1^0|\mathsf{S}(\tau_{\mathrm{out}}, \tau)Q_0(\tau)\mathsf{S}(\tau,\sigma)Q_0(\sigma)\mathsf{S}(\sigma,\tau_{\mathrm{in}})|E_1^0\rangle}{\langle E_1^0|\mathsf{S}(\tau_{\mathrm{out}},\tau_{\mathrm{in}})|E_1^0\rangle}.$$

Here, we set $Q_0(\tau) := e^{i\tau H_0/\hbar}Qe^{-i\tau H_0/\hbar}$.

[45] As an introduction to perturbation theory, we recommend the standard textbooks by Kato (1966) and Reed, Simon (1972), Vol. 4.

Mnemonically, using the time-ordering operator $\mathcal{T}$ from Sect. 7.17.4, physicists write this briefly as

$$\boxed{C_\varepsilon(t,s) = \lim_{t_{\rm out}\to+\infty} \lim_{t_{\rm in}\to-\infty} \frac{\langle E_1^0|\mathcal{T}\{\mathsf{S}(\tau_{\rm out},\tau_{\rm in})Q_0(\tau)Q_0(\sigma)\}|E_1^0\rangle}{\langle E_1^0|\mathsf{S}(\tau_{\rm out},\tau_{\rm in})|E_1^0\rangle}.}$$

This is called the magic Gell-Mann–Low perturbation formula. Observe the following. The causal correlation function is given by the limit

$$C(t,s) = \lim_{\varepsilon\to+0} C_\varepsilon(t,s).$$

By Sect. 7.22, $C(t,s) = \langle E_1|\dots|E_1\rangle$, that is, the causal correlation function refers to the ground state $|E_1\rangle$ of the perturbed quantum system. In contrast to this, the Gell-Mann–Low formula refers to the ground state $|E_1^0\rangle$ of the unperturbed (free) quantum system. Moreover, the operator $Q_0 = Q_0(\tau)$ only depends on the unperturbed Hamiltonian H_0. The interaction is described by the S-operator switched on at times $\tau = t(1-\varepsilon{\rm i})$ and $\sigma = s(1-\varepsilon{\rm i})$.

Proof of Theorem 7.33. This follows from Theorem 7.32 by replacing the states $|Q_{\rm in}\rangle$ and $|Q_{\rm out}\rangle$ by $|E_1^0\rangle$. In addition, for the propagator use the following Dyson formula (7.52), namely, $P(t,s) = {\rm e}^{-{\rm i}tH_0/\hbar}\mathsf{S}(t,s){\rm e}^{{\rm i}sH_0/\hbar}$ if $t \geq s$. □

In Chap. 15, we will consider the operator approach to quantum field theory. We will start there with the magic Dyson formula (7.90) in order to elegantly compute the S-matrix. The causal correlation functions then follow from the magic Gell-Mann–Low formula. In the following two sections, as an alternative, we will consider the rigorous response approach based on Gaussian integrals and discrete functional integrals. The corresponding applications to quantum field theory will be studied in Chap. 14.

7.23 The Magic Gaussian Integral

The most important probability distribution is the Gaussian distribution.
Folklore

Quantum theory is a stochastic theory. Therefore, it is not a big surprise that the Gaussian distribution plays a fundamental role in quantum physics. The following formulas are frequently used by physicists when computing Feynman functional integrals.

7.23.1 The One-Dimensional Prototype

The Fourier–Gauss formula. Let us start with the classical key formula

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} {\rm e}^{-x^2/2}{\rm e}^{-{\rm i}px}\,dx = {\rm e}^{-p^2/2} \tag{7.91}$$

which is valid for all complex numbers p. This magic formula tells us that the Gaussian function $\mathrm{e}^{-x^2/2}$ is invariant under Fourier transformation. The integral (7.91) exists in the classical sense because of the damping factor $\mathrm{e}^{-x^2/2}$. The proof will be given in Problem 7.1.

Use now the rescaling $x \mapsto \sqrt{a}\, x$ and $p \mapsto p/\sqrt{a}$. Then, for all $a > 0$ and $p \in \mathbb{C}$, we obtain

$$\boxed{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-ax^2/2} \mathrm{e}^{-\mathrm{i}px}\, dx = \frac{\mathrm{e}^{-p^2/2a}}{\sqrt{a}}.} \tag{7.92}$$

The Gaussian probability distribution. The function

$$\varrho(x) := \frac{\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}, \qquad x \in \mathbb{R}$$

is called the density function of the Gaussian probability distribution. Here, $\mu \in \mathbb{R}$ and $\sigma > 0$ (Fig. 7.4). We have the following normalization condition

$$\int_{-\infty}^{\infty} \varrho(x) dx = 1. \tag{7.93}$$

The number

$$\int_c^d \varrho(x) dx$$

is equal to the probability of measuring the quantity x in the interval $[c, d]$. In addition, for the mean value $\bar{x}$ and the mean fluctuation Δx of the random variable x, we get

$$\bar{x} = \int_{-\infty}^{\infty} x \varrho(x) dx = \mu,$$

$$(\Delta x)^2 = \int_{-\infty}^{\infty} (x - \bar{x})^2 \varrho(x) dx = \sigma^2.$$

By the central limit theorem of probability theory, roughly speaking, a random variable possesses a Gaussian distribution if it results from the superposition of a large number of independent random variables.

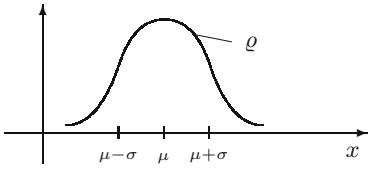


Fig. 7.4. Gaussian probability distribution

In quantum field theory, we will use perturbed Gaussian distributions.

The perturbations are caused by the interactions between quantum fields.

The Dirac delta distribution. In what follows we will use the space $\mathcal{S}(\mathbb{R})$ (resp. $\mathcal{D}(\mathbb{R})$) of smooth test functions $\chi : \mathbb{R} \to \mathbb{C}$ which go rapidly to zero at infinity (resp. vanish outside a compact interval). Furthermore, we will use elementary facts about generalized functions (see Sect. 11.3). By (7.93), we get the classical limiting relation

$$\boxed{\lim_{\sigma \to +0} \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-\frac{x^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}\, \chi(x)\, dx = \chi(0)} \tag{7.94}$$

for all test functions $\chi \in \mathcal{S}(\mathbb{R})$. The proof will be given in Problem 7.2. In the language of generalized functions,

$$\lim_{\sigma \to +0} \frac{\mathrm{e}^{-\frac{x^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} = \delta(x), \qquad x \in \mathbb{R}.$$

It follows from (7.92) and (7.94) that

$$\boxed{\lim_{a \to +0} \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}ax^2 - \mathrm{i}px}\, dx = \delta(p), \qquad p \in \mathbb{R},} \tag{7.95}$$

in the sense of generalized functions. This means explicitly that for all test functions $\chi \in \mathcal{S}(\mathbb{R})$,

$$\lim_{a \to +0} \int_{-\infty}^{\infty} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}ax^2 - \mathrm{i}px}\, dx \right) \chi(p) dp = \chi(0).$$

The method of stationary phase. The following observation is crucial. Formula (7.92) can be written as

$$\boxed{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{\Phi(x)} dx = \frac{\mathrm{e}^{\Phi(x_0)}}{\sqrt{a}}.} \tag{7.96}$$

Here, we introduce the phase function

$$\Phi(x) := -\tfrac{1}{2}ax^2 - \mathrm{i}px.$$

The critical value x_0 is determined by the stationary phase condition

$$\boxed{\Phi'(x_0) = 0.} \tag{7.97}$$

Explicitly, $-ax_0 - \mathrm{i}p = 0$ implies $x_0 = -\frac{\mathrm{i}p}{a}$. Hence $\Phi(x_0) = -\frac{p^2}{2a}$. The classical identity (7.96) is called the stationary phase identity.

Let h be a small positive real parameter. Set $\hbar := h/2\pi$. For each test function $\chi \in \mathcal{D}(\mathbb{R})$, we have the asymptotic expansion of the Gaussian integral,

$$\int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2\hbar}\, \chi(x)\, dx = h^{\frac{1}{2}} \chi(0) + O\big(h^{\frac{3}{2}}\big), \qquad h \to +0, \tag{7.98}$$

which follows from (7.94). This is the key to the method of stationary phase to be studied in Sect. 12.5.3 in connection with geometric optics, by generalizing Fresnel integrals. In particular, it turns out that formula (7.98) remains valid if we replace h by $\pm \mathrm{i}h$. We then get

$$\int_{-\infty}^{\infty} \mathrm{e}^{\pm \mathrm{i}x^2/2\hbar}\, \chi(x)\, dx = (\pm \mathrm{i}h)^{\frac{1}{2}} \chi(0) + O\big(h^{\frac{3}{2}}\big), \qquad h \to +0 \tag{7.99}$$

where we set $(\pm \mathrm{i})^{\frac{1}{2}} := \mathrm{e}^{\pm \frac{\mathrm{i}\pi}{4}}$. Intuitively, for small values of h, the integral (7.99) is rapidly oscillating. Therefore, the contributions of the integrand to the integral cancel each other, except at the position $x = 0$ where the rapid oscillations do not appear.

Physicists use the method of stationary phase in order to compute semiclassical approximations of quantum effects where the Planck constant $\hbar$ is assumed to be small. More precisely, we assume that $S(x)/\hbar$ is small, where $S(x)$ denotes the classical action. Applications of this so-called WKB method for computing semiclassical approximations of the Feynman propagator kernel can be found in both Kleinert (1996) and Grosche, Steiner (1996). It turns out that the topology of the corresponding classical dynamical system (focal points, closed orbits) plays a fundamental role in terms of the Maslov index. Summarizing, for an important class of interactions, the method of stationary phase allows us to compute the most important contributions to the corresponding physical processes.

The moment trick. For fixed $a > 0$ and all $J \in \mathbb{C}$, define

$$Z(J) := \sqrt{\frac{a}{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-ax^2/2}\, \mathrm{e}^{\mathrm{i}Jx}\, dx.$$

By (7.92),

$$\boxed{Z(J) = \mathrm{e}^{-J^2/2a} \qquad \text{for all} \quad J \in \mathbb{C}.}$$

In particular, $Z(0) = 1$. Using the nth derivative of the function $Z = Z(J)$, the integral

$$\boxed{\sqrt{\frac{a}{2\pi}} \int_{-\infty}^{\infty} x^n \mathrm{e}^{-\frac{1}{2}ax^2}\, dx = \frac{1}{\mathrm{i}^n} \frac{d^n Z(0)}{dJ^n} \qquad n = 1, 2, \ldots}$$

can be computed easily. In the theory of probability, this integral is called the nth moment of the Gauss distribution if we choose $a = 1/\sigma^2$. In particular,

the nth moment vanishes if n is odd. Physicists call this the moment trick or the Wick trick. The function $Z = Z(J)$ is called the generating function for the moments. For example,

$$\sqrt{\frac{a}{2\pi}} \int_{-\infty}^{\infty} x^2 \mathrm{e}^{-\frac{1}{2}ax^2} dx = -Z''(0) = \frac{1}{a} .$$

Perturbation theory. As a typical example, consider the integral

$$Z(J,\kappa) := \int_{-\infty}^{\infty} dx \ \mathrm{e}^{-\frac{1}{2}ax^2} \mathrm{e}^{\mathrm{i}\kappa x^4} \mathrm{e}^{\mathrm{i}Jx}$$

where κ is a positive constant called the coupling constant. By power series expansion,

$$Z(J,\kappa) = \int_{-\infty}^{\infty} dx \ \mathrm{e}^{-\frac{1}{2}ax^2} (1 + \mathrm{i}\kappa x^4 - \tfrac{1}{2}\kappa^2 x^8 + \ldots) \mathrm{e}^{\mathrm{i}Jx}.$$

By the Wick trick,

$$\begin{aligned} Z(J,\kappa) &= \int_{-\infty}^{\infty} dx \ \mathrm{e}^{-\frac{1}{2}ax^2} \left\{ 1 + \mathrm{i}\kappa \left(\frac{1}{\mathrm{i}} \frac{d}{dJ} \right)^4 - \tfrac{1}{2}\kappa^2 \left(\frac{1}{\mathrm{i}} \frac{d}{dJ} \right)^8 + \ldots \right\} \mathrm{e}^{\mathrm{i}Jx} \\ &= \left\{ 1 + \mathrm{i}\kappa \left(\frac{1}{\mathrm{i}} \frac{d}{dJ} \right)^4 - \frac{1}{2}\kappa^2 \left(\frac{1}{\mathrm{i}} \frac{d}{dJ} \right)^8 + \ldots \right\} \int_{-\infty}^{\infty} dx \ \mathrm{e}^{-\frac{1}{2}ax^2} \ \mathrm{e}^{\mathrm{i}Jx}. \end{aligned}$$

Mnemonically, this can be written as

$$\boxed{Z(J,\kappa) = \exp\left\{ \mathrm{i}\kappa \left(\frac{1}{\mathrm{i}} \frac{d}{dJ} \right)^4 \right\} Z(J,0).} \tag{7.100}$$

Explicitly, we get

$$Z(J,0) = \sqrt{\frac{2\pi}{a}} \ \mathrm{e}^{-J^2/2a}.$$

Recall that this can be obtained by using the method of stationary phase. Therefore, the first and second approximation of $Z(J,\kappa)$ reads as

$$Z(J,\kappa) = \sqrt{\frac{2\pi}{a}} \left(1 + \mathrm{i}\kappa \frac{d^4}{dJ^4} \right) \mathrm{e}^{-J^2/2a}$$

and

$$Z(J,\kappa) = \sqrt{\frac{2\pi}{a}} \left(1 + \mathrm{i}\kappa \frac{d^4}{dJ^4} - \frac{\kappa^2}{2} \frac{d^8}{dJ^8} \right) \mathrm{e}^{-J^2/2a}.$$

Formula (7.100) is the prototype of the fundamental quantum action principle in quantum field theory. More generally, let $U = U(x)$ be a polynomial with real coefficients. Then the integral

$$Z(J,\kappa) = \int_{-\infty}^{\infty} dx\, \mathrm{e}^{-\frac{1}{2}ax^2 + \mathrm{i}\kappa U(x)} \mathrm{e}^{\mathrm{i}Jx}$$

is equal to

$$\boxed{Z(J,\kappa) = \exp\left\{\mathrm{i}\kappa U\left(\frac{1}{\mathrm{i}}\frac{d}{dJ}\right)\right\} Z(J,0).} \tag{7.101}$$

The heat kernel formula. Using $\mathrm{e}^{-x^2-y^2} = \mathrm{e}^{-x^2}\mathrm{e}^{-y^2}$, and so on, from the key formula (7.92) we obtain the 3-dimensional heat kernel relation

$$\frac{1}{(2\pi)^{\frac{3}{2}}} \int_{\mathbb{R}^3} d^3x\, \mathrm{e}^{-\frac{1}{2}a\mathbf{x}^2 - \mathrm{i}\mathbf{p}\mathbf{x}} = \frac{\mathrm{e}^{-\frac{\mathbf{p}^2}{2a}}}{a^{\frac{3}{2}}} \tag{7.102}$$

for all $a > 0$ and all vectors $\mathbf{p}$.

Analytic continuation. Let $z = a + \mathrm{i}b$ be a complex number with $a, b \in \mathbb{R}$ where

$$a + \mathrm{i}b = |a + \mathrm{i}b|\, \mathrm{e}^{\mathrm{i}\alpha}, \qquad -\pi < \alpha < \pi.$$

In other words, the number z lies in the complex plane outside the negative real axis. We define

$$\sqrt{a + \mathrm{i}b} := \sqrt{|a + \mathrm{i}b|}\, \mathrm{e}^{\mathrm{i}\alpha/2}.$$

This function is analytic on $\mathbb{C} \setminus \{a \in \mathbb{R} :\ a \le 0\}$. Let us study the integral

$$G(a + \mathrm{i}b) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}(a+\mathrm{i}b)x^2} \mathrm{e}^{-\mathrm{i}px} dx$$

where p is an arbitrary complex number. Our goal is the formula

$$\boxed{G(a + \mathrm{i}b) = \frac{\mathrm{e}^{-p^2/2(a+\mathrm{i}b)}}{\sqrt{a + \mathrm{i}b}}} \tag{7.103}$$

for all $p \in \mathbb{C}$. We have to discuss three cases.

Case 1: Let $a > 0$ and $b = 0$. By the key formula (7.91),

$$G(a) = \frac{\mathrm{e}^{-p^2/2a}}{\sqrt{a}}. \tag{7.104}$$

Case 2: Let $a > 0$ and $b \in \mathbb{R}$. The integral $G(a+\mathrm{i}b)$ exists, and the function $a+\mathrm{i}b \mapsto G(a+\mathrm{i}b)$ is analytic on the open half-plane $\{a+\mathrm{i}b : a > 0, b \in \mathbb{R}\}$. In fact, the derivative of G can be computed as in Problem 7.1. Formula (7.103) follows now from (7.104) by means of analytic continuation.

Case 3: Let $a \le 0$ and $b \neq 0$. Then, the integral $G(a+b\mathrm{i})$ does not exist in the classical sense. However, we can define it in a generalized sense by means of analytic continuation. This yields (7.103). For example, in the generalized sense,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{\mp\frac{1}{2}\mathrm{i}x^2} dx = \frac{1}{\sqrt{\pm\mathrm{i}}} = \mathrm{e}^{\mp\mathrm{i}\pi/4}.$$

7.23.2 The Determinant Trick

Let us now pass over to multidimensional generalizations of the formulas considered above. Let $N = 1, 2, \ldots$, and let $\lambda_1 > 0, \ldots, \lambda_N > 0$. Moreover, let $J_1, \ldots, J_N$ be complex numbers. Set

$$Z(J) := \frac{1}{(2\pi)^{N/2}} \int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\sum_{k=1}^N \lambda_k \varphi_k^2} \, \mathrm{e}^{\mathrm{i}\sum_{k=1}^N J_k \varphi_k} d\varphi_1 \cdots d\varphi_N.$$

By the key formula (7.92),

$$Z(J) := \frac{\mathrm{e}^{-\frac{1}{2}\sum_{k=1}^N \lambda_k^{-1} J_k^2}}{(\prod_{k=1}^N \lambda_k)^{1/2}}. \tag{7.105}$$

In what follows, we will use the matrices

$$\varphi := \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_N \end{pmatrix}, \qquad J := \begin{pmatrix} J_1 \\ \vdots \\ J_N \end{pmatrix} \tag{7.106}$$

where $\varphi_1, \ldots, \varphi_N$ are real numbers, and $J_1, \ldots, J_N$ are complex numbers.

Theorem 7.34 *Let A be a real symmetric $(N \times N)$-matrix with the eigenvalues $\lambda_1 > 0, \ldots, \lambda_N > 0$. Then*

$$\frac{1}{(2\pi)^{N/2}} \int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\varphi^d A \varphi} \mathrm{e}^{\mathrm{i}\varphi^d J} \, d\varphi_1 \cdots d\varphi_N = \frac{\mathrm{e}^{-\frac{1}{2}J^d A^{-1} J}}{(\det A)^{1/2}}. \tag{7.107}$$

The proof of this determinant formula will be given in Problem 7.3.

7.23.3 The Zeta Function Trick

Let the matrix A be given as in Theorem 7.34 above. The function

$$\zeta_A(s) := \sum_{k=1}^N \frac{1}{\lambda_k^s}, \qquad \text{for all} \quad s \in \mathbb{C}$$

is called the zeta function of the matrix A on the complex plane. Note that $\lambda_k^{-s} = \mathrm{e}^{-s \ln \lambda_k}$. Thus, for the derivative,

$$\zeta_A'(s) = -\sum_{k=1}^N \frac{\ln \lambda_k}{\lambda_k^s}, \qquad s \in \mathbb{C}.$$

Hence

$$\det A = \prod_{k=1}^{N} \lambda_k = \mathrm{e}^{-\zeta_A'(0)}.$$

Setting $\langle \varphi | A\varphi \rangle := \varphi^d A\varphi$ and $\langle \varphi | J \rangle := \varphi^d J$, formula (7.107) is equivalent to the magic zeta function formula

$$\boxed{\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\langle \varphi | A\varphi \rangle + \mathrm{i}\langle \varphi | J \rangle} \prod_{k=1}^{N} \frac{d\varphi_k}{\sqrt{2\pi}} = \mathrm{e}^{\frac{1}{2}\zeta_A'(0)}\, \mathrm{e}^{-\frac{1}{2}\langle J | A^{-1} J \rangle}.} \tag{7.108}$$

In the case of infinite-dimensional functional integrals, physicists use this formula; they then compute the derivative of the zeta function at the origin, $\zeta'(0)$, by using analytic continuation.

7.23.4 The Moment Trick

Let the matrix A be given as in Theorem 7.34 above. Define

$$Z(J) := \int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\varphi^d A\varphi}\, \mathrm{e}^{\mathrm{i}\varphi^d J} d\varphi_1 \cdots d\varphi_N.$$

Then

$$\frac{1}{\mathrm{i}} \frac{\partial Z(0)}{\partial J_k} = \int_{\mathbb{R}^N} \varphi_k \mathrm{e}^{-\frac{1}{2}\varphi^d A\varphi}\, d\varphi_1 \cdots d\varphi_N.$$

This implies

$$\boxed{\frac{1}{\mathrm{i}^n} \frac{\partial^n Z(0)}{\partial J_{k_1} \cdots \partial J_{k_n}} = \int_{\mathbb{R}^N} \varphi_{k_1} \cdots \varphi_{k_n} \mathrm{e}^{-\frac{1}{2}\varphi^d A\varphi}\, d\varphi_1 \cdots d\varphi_N.} \tag{7.109}$$

7.23.5 The Method of Stationary Phase

First prototype. For $N = 1, 2, \ldots$ and each fixed real number $\eta > 0$, let us consider the integral

$$W(J) := \int_{\mathbb{R}^N} \mathrm{e}^{\Phi(\varphi, J)}\, \mathrm{e}^{-\eta\varphi^d\varphi} d\varphi_1 \cdots d\varphi_N$$

with the phase function

$$\Phi(\varphi, J) := \tfrac{1}{2}\mathrm{i}\varphi^d A\varphi + \mathrm{i}\varphi^d J.$$

Here, φ and J are given by (7.106). We will show below that under suitable assumptions, the functional Φ has a unique critical point φ_0 depending on J, and

$$W(J) = \mathrm{e}^{\Phi(\varphi_0, J)} W(0).$$

Explicitly,

$$\boxed{W(J) = \mathrm{e}^{-\frac{1}{2}\mathrm{i} J^d A^{-1} J} W(0)} \tag{7.110}$$

for all real numbers $J_1, \ldots, J_N$.

Theorem 7.35 *Let A be a complex symmetric invertible $(N \times N)$-matrix such that*

$$\sup_{\varphi \in \mathbb{R}^N} |e^{\mathrm{i}\varphi^d A \varphi}| < \infty. \tag{7.111}$$

Then, equation (7.110) *holds true.*

In particular, this implies

$$\lim_{\eta \to +0} \frac{W(J)}{W(0)} = \mathrm{e}^{-\frac{1}{2}\mathrm{i} J^d A^{-1} J}. \tag{7.112}$$

For example, condition (7.111) is satisfied if $A := D + \mathrm{i}\varepsilon I$ where the matrix D is real and $\varepsilon \geq 0$. In fact, then

$$|\mathrm{e}^{\mathrm{i}\varphi^d A \varphi}| = |\mathrm{e}^{\mathrm{i}\varphi^d A \varphi}| \cdot \mathrm{e}^{-\varepsilon \varphi^d \varphi} = \mathrm{e}^{-\varepsilon \varphi^d \varphi}.$$

Proof. (I) The integral

$$W(J) = \int_{\mathbb{R}^N} \mathrm{e}^{\frac{1}{2}\mathrm{i}\varphi^d A \varphi}\, \mathrm{e}^{\mathrm{i}\varphi^d J}\, \mathrm{e}^{-\eta \varphi^d \varphi}\, d\varphi_1 \cdots d\varphi_N$$

exists because of the damping factor $\mathrm{e}^{-\eta\varphi^d\varphi}$.

(II) Critical point of the phase function Φ. For fixed φ_0, h and J, define

$$\chi(\tau) := \Phi(\varphi_0 + \tau h, J), \qquad \tau \in \mathbb{R}.$$

The function χ is quadratic. By Taylor expansion,

$$\chi(1) = \chi(0) + \chi'(0) + \tfrac{1}{2}\chi''(0). \tag{7.113}$$

Choose φ_0 in such a way that it is a critical point of Φ, that is, $\chi'(0) = 0$ for all h. Since A is symmetric, we get

$$\chi'(0) = \tfrac{1}{2}\mathrm{i} h^d A \varphi_0 + \tfrac{1}{2}\mathrm{i}\varphi_0^d A h + \mathrm{i} h^d J = \mathrm{i} h^d (A + J)\varphi_0 = 0$$

for all $h \in \mathbb{R}^N$. Hence $A\varphi_0 + J = 0$. This so-called response equation has the unique solution

$$\varphi_0 = -A^{-1} J. \tag{7.114}$$

Since the matrix A^{-1} is symmetric, $\varphi_0^d = -J^d A^{-1}$. Setting $h := \varphi - \varphi_0$, it follows from (7.113) that

$$\Phi(\varphi) = \Phi(\varphi_0 + h) = \Phi(\varphi_0) + \tfrac{1}{2}\chi''(0).$$

This implies

$$W(J) = \mathrm{e}^{\Phi(\varphi_0, J)} \int_{\mathbb{R}^N} \mathrm{e}^{\frac{1}{2}\chi''(0)}\ \mathrm{e}^{-\eta\varphi^d\varphi} d\varphi_1 \cdots d\varphi_N.$$

If $J = 0$, then $\Phi(\varphi_0) = \Phi(0) = 0$. Since the J-term of the function Φ is linear with respect to φ, the second derivative $\chi''(0)$ does not depend on J. Thus,

$$W(0) = \int_{\mathbb{R}^N} \mathrm{e}^{\frac{1}{2}\chi''(0)}\ \mathrm{e}^{-\eta\varphi^d\varphi} d\varphi_1 \cdots d\varphi_N.$$

By (7.114),

$$\Phi(\varphi_0, J) = \frac{1}{2}\mathrm{i}J^d A^{-1} J - \mathrm{i}J^d A^{-1} J = -\tfrac{1}{2}\mathrm{i}J^d A^{-1} J.$$

□

Analytic continuation. Motivated by (7.112), we define

$$\boxed{\frac{\int_{\mathbb{R}^N} \mathrm{e}^{\frac{1}{2}\mathrm{i}\varphi^d A\varphi + \mathrm{i}\varphi^d J} \prod_{k=1}^N d\varphi_k}{\int_{\mathbb{R}^N} \mathrm{e}^{\frac{1}{2}\mathrm{i}\varphi^d A\varphi} \prod_{k=1}^N d\varphi_k} := \mathrm{e}^{-\frac{1}{2}\mathrm{i}J^d A^{-1} J}} \tag{7.115}$$

for all complex symmetric invertible $(N \times N)$-matrices A and all complex numbers $J_1, \ldots, J_N$.

Second prototype. Fix $N = 1, 2, \ldots$ and $\eta > 0$. Let us now consider the integral

$$W(J, \overline{J}) := \int_{\mathbb{R}^N} \mathrm{e}^{\Phi(\psi, \overline{\psi}, J, \overline{J})}\ \mathrm{e}^{-\eta(\psi^d\psi + \overline{\psi}\ \overline{\psi}^d)}\ d\overline{\psi_1} \cdots d\overline{\psi}_N d\psi_1 \cdots d\psi_N$$

with the phase function

$$\Phi(\psi, \overline{\psi}, J, \overline{J}) := \mathrm{i}\overline{\psi}A\psi + \mathrm{i}\overline{\psi}J + \mathrm{i}\overline{J}\psi.$$

Here, we use the following real matrices

$$\psi = \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_N \end{pmatrix}, \qquad J = \begin{pmatrix} J_1 \\ \vdots \\ J_N \end{pmatrix}$$

along with $\overline{\psi} = (\overline{\psi}_1, \ldots, \overline{\psi}_N)$ and $\overline{J} = (\overline{J}_1, \ldots, \overline{J}_N)$.

Theorem 7.36 *Let A be a real invertible $(N \times N)$-matrix. Then,*

$$W(J, \overline{J}) = e^{-\mathrm{i}\overline{J}A^{-1}J}\, W(0,0)$$

for all real numbers $J_1, \ldots, J_N, \overline{J}_1, \ldots, \overline{J}_N$, and all $\eta > 0$.

In particular, this implies

$$\lim_{\eta \to +0} \frac{W(J, \overline{J})}{W(0,0)} = \mathrm{e}^{-\mathrm{i}\overline{J}A^{-1}J}. \tag{7.116}$$

Proof. For fixed $\psi_0, \overline{\psi}_0, h, \overline{h}, J, \overline{J}$, set

$$\chi(\tau) := \Phi(\psi_0 + \tau h, \overline{\psi}_0 + \tau \overline{h}, J, \overline{J}), \qquad \tau \in \mathbb{R}.$$

Note that

$$\chi'(0) = \mathrm{i}\overline{h}(A\psi_0 + J) + \mathrm{i}(\overline{\psi}_0 A + \overline{J})h.$$

If we choose

$$\psi_0 := -A^{-1}J, \qquad \overline{\psi}_0 := -\overline{J}A^{-1},$$

then $\chi'(0) = 0$. As in the proof of Theorem 7.35, we get

$$W(J, \overline{J}) = \mathrm{e}^{\Phi(\psi_0, \overline{\psi}_0, J, \overline{J})}\, W(0,0)$$

where $W(0,0)$ depends on η. Finally,

$$\Phi(\psi_0, \overline{\psi}_0, J, \overline{J}) = \mathrm{i}\overline{J}A^{-1}J - 2\mathrm{i}\overline{J}A^{-1}J = -\mathrm{i}\overline{J}A^{-1}J.$$

This finishes the proof. □

Analytic continuation. Motivated by (7.116), we define

$$\boxed{\frac{\int_{\mathbb{R}^N} \mathrm{e}^{\mathrm{i}\overline{\psi}A\psi + \mathrm{i}\overline{J}\psi + \mathrm{i}\overline{\psi}J} \prod_{k=1}^N d\overline{\psi}_k d\psi_k}{\int_{\mathbb{R}^N} \mathrm{e}^{\mathrm{i}\overline{\psi}A\psi} \prod_{k=1}^N d\overline{\psi}_k d\psi_k} := \mathrm{e}^{-\mathrm{i}\overline{J}A^{-1}J}} \tag{7.117}$$

for all complex invertible $(N \times N)$-matrices A and all complex numbers $J_1, \ldots, J_N$, $\overline{J}_1, \ldots, \overline{J}_N$.[46] Here, as a rule, the single integrals do not exist, only the quotient is well-defined. Expressions of the form (7.117) play a crucial role in the following response approach to quantum field theory.

7.24 The Rigorous Response Approach to Finite Quantum Fields

> Quantum field theory is based on only a few basic principles which we call magic formulas.
>
> Folklore

[46] Note that the symbol $\overline{J}_k$ does not denote the conjugate complex number to J_k, but an independent variable.

7.24.1 Basic Ideas

Classical fields are described by the principle of critical action. The universal response approach to quantum field theory combines the principle of critical action with the random aspects of infinite-dimensional Gaussian integrals. The idea is to study the response of the quantum field under the influence of an external source. In this section, we will rigorously study a finite-dimensional variant of the response approach. The notation will be chosen in such a way that the formal continuum limit can be carried out in a straightforward manner in Chap. 14. The basic idea is to introduce the so-called extended quantum action functional $Z(J, \varphi)$ which depends on both the quantum field φ and the external source J and to derive two magic formulas, namely,

(QA) the quantum action reduction formula, and
(LSZ) the Lehmann–Szymanzik–Zimmermann reduction formula on the relation between correlation functions

$$C_n(x_1, \ldots, x_n)$$

and scattering functions $\mathsf{S}_n(x_1, \ldots, x_n)$.

The main steps of our approach are the following ones.

(i) The principle of critical action. The point is that the classical action functional

$$S[\varphi, J]$$

depends on both the classical quantum field φ and the external source J. This yields the Euler–Lagrange equation of the motion of the classical quantum field under the influence of the external source. See (7.122) on page 446.

(ii) The response operator. The linearized (and regularized) Euler–Lagrange equation determines the response operator R_ε which describes the response

$$\varphi = R_\varepsilon J$$

of the interaction-free classical quantum field φ to the external source J. See (7.125) on page446. The small parameter $\varepsilon > 0$ regularizes the response. It turns out that the response operator R_ε knows all about the full quantum field. To this end, we will use the two magic formulas (QA) and (LSZ).

(iii) The quantum action reduction formula (QA). The response operator R_ε determines the free generating functional

$$Z_{\text{free}}(J, \varphi).$$

The formula (QA) tells us how the corresponding generating functional

$$Z(J,\varphi)$$

of the interacting quantum field can be obtained from the free functional $Z_{\text{free}}(J,\varphi)$ by functional differentiation. See (7.127) on page 448. Note that the key formula (QA) is valid in each order of perturbation theory with respect to the coupling constant κ.

The magic formula (QA) *describes the quantization of the classical quantum field* φ.

(iv) The n-point correlation function C_n. It is our philosophy that

The main properties of a quantum field are described by the sequence of correlation functions $C_2, C_4, \ldots$

These functions are obtained from $Z(J,0)$ by using functional differentiation at the point $J=0$. See (7.129) on page 448.

(v) The n-point scattering functions S_n. These functions are obtained by applying functional differentiation to $Z(0,\varphi)$ at the point $\varphi = 0$. See (7.130) on page 449. The scattering functions know all about scattering processes.

(vi) The LSZ reduction formula. This fundamental formula tells us how to compute the n-point scattering function S_n by means of the n-point correlation function C_n.

(vii) The local quantum action principle. The formula (QA) is the solution formula to the Dyson–Schwinger equation (7.133) on page 452 which is called the local quantum action principle.

We will use finite functional integrals in order to derive rigorously the magic formulas (QA) and (LSZ) by using the quite natural global quantum action principle.

The global quantum action principle is based on an averaging over the classical fields where the statistical weight $e^{\mathrm{i}S[\varphi,J]/\hbar}$ *depends on the classical action* $S[\varphi,J]$.

The explicit formulation can be found in Sect. 7.24.5 on page 447. However, the response approach can also be formulated in such a way that functional integrals do not appear explicitly. This is important for the infinite-dimensional approach, since functional integrals are not well-defined in infinite dimensions.

The basic idea is then to start with the definition of the extended quantum action functional $Z = Z(J,\varphi)$ *and to define the correlation functions* C_n *and the scattering functions* S_n *as functional derivatives of* $Z(\varphi,J)$.

In what follows, we will investigate the basic ideas sketched in (i) through (vii) above in a rigorous setting. The translation to quantum fields with an infinite number of degrees of freedom will be studied in Chap. 14 (response approach) and Chap. 15 (operator approach).

7.24.2 Discrete Space-Time Manifold

According to Einstein, the physics in an inertial system depends on space-time points

$$x = (\mathbf{x}, ct)$$

where $\mathbf{x}$ and t denote position vector and time, respectively. Moreover, c is the velocity of light in a vacuum. The set of all space-time points x forms the 4-dimensional space-time manifold $\mathbb{M}^4$. In this section, we use a finite-dimensional set

$$\mathcal{M} = (1, 2, \ldots, N)$$

as discrete space-time manifold where $N = 1, 2, \ldots$. Furthermore, we fix a real number $\Delta x > 0$, and we set $\Delta^4 x := (\Delta x)^4$. Each function

$$\varphi : \mathcal{M} \to \mathbb{R}$$

is called a discrete quantum field, and each function $J : \mathcal{M} \to \mathbb{R}$ is called an external source. In the language of matrices, we write

$$\varphi = \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_N \end{pmatrix}, \qquad J = \begin{pmatrix} J_1 \\ \vdots \\ J_N \end{pmatrix}.$$

Discrete integral. For each function $f : \mathcal{M} \to \mathbb{R}$, the sum

$$\sum_{x \in \mathcal{M}} f(x) \Delta^4 x$$

is called a finite integral on $\mathcal{M}$. The space of all functions $f : \mathcal{M} \to \mathbb{R}$ becomes a real N-dimensional Hilbert space with respect to the inner product

$$\boxed{\langle f | g \rangle := \sum_{x \in \mathcal{M}} f(x) g(x) \Delta^4 x.} \tag{7.118}$$

This Hilbert space is denoted by $L_2(\mathcal{M})$. Let $\mathcal{A}$ be a real $(N \times N)$-matrix with the entries $\mathcal{A}(x, y)$ where $x, y \in \mathcal{M}$. Let $\Delta^4 y := \Delta^4 x$. Define

$$(A\varphi)(x) := \sum_{x \in \mathcal{M}} \mathcal{A}(x, y) \varphi(y) \Delta^4 y \qquad \text{for all} \quad x \in \mathcal{M}.$$

The linear operator $A : L_2(\mathcal{M}) \to L_2(\mathcal{M})$ is called a discrete integral operator with the kernel $\mathcal{A}$. Obviously, each linear operator A on $L_2(\mathcal{M})$ is a discrete integral operator. This statement is not true anymore in infinite-dimensional Hilbert spaces. This fact causes trouble in infinite dimensions.

Discrete Dirac delta function. Let $\delta_{x,x} := 1$ and $\delta_{x,y} := 0$ if $x \neq y$. The 4-dimensional discrete Dirac delta function $\delta_{\Delta^4 x}$ is defined by

$$\boxed{\delta_{\Delta^4 x}(x,y) := \frac{\delta_{x,y}}{\Delta^4 x}, \qquad x, y \in \mathcal{M}.}$$

For each function $\varphi \in L_2(\mathcal{M})$, the key relation

$$\varphi(x) = \sum_{y \in \mathcal{M}} \delta_{\Delta^4 x}(x,y) \varphi(y) \Delta^4 y \qquad \text{for all} \quad x \in \mathcal{M}$$

tells us that the discrete Dirac delta function $\delta_{\Delta^4 x}$ is the kernel of the identity operator on the Hilbert space $L_2(\mathcal{M})$. The formal limit $N \to +\infty, \Delta x \to 0$ yields

$$\varphi(x) = \int_{\mathbb{M}^4} \delta(x,y) \varphi(y) d^4 y \qquad \text{for all} \quad x \in \mathbb{M}^4.$$

In terms of the formal Dirac delta function, $\delta(x,y) = \delta^4(x-y)$ (see Sect. 11.2.1 on page 590).

Discrete local functional derivative. For each function $F : \mathbb{R}^N \to \mathbb{R}$, define

$$\boxed{\frac{\delta F(\varphi)}{\delta \varphi(x)} := \frac{1}{\Delta^4 x} \frac{\partial F(\varphi(1), \ldots, \varphi(N))}{\partial \varphi(x)}.}$$

This functional derivative differs from the corresponding partial derivative with respect to the variable $\varphi(x)$ by multiplying with the factor $1/\Delta^4 x$. For example,

$$\frac{\delta}{\delta \varphi(y)} \sum_{x \in \mathcal{M}} \varphi(x) J(x) \Delta^4 x = J(y) \,.$$

Discrete functional integral. For each function $F : \mathbb{R}^N \to \mathbb{R}$, the key definition reads as

$$\boxed{\int_{L_2(\mathcal{M})} F(\varphi) \mathcal{D}\varphi := \int_{\mathbb{R}^N} F(\varphi(1), \ldots, \varphi(N)) \prod_{k=1}^{N} \sqrt{\frac{\Delta^4 x}{2\pi}}\, d\varphi(k).}$$

We briefly write

$$\int_{L_2(\mathcal{M})} F(\varphi) \mathcal{D}\varphi := \int_{\mathbb{R}^N} F(\varphi) \prod_{x \in \mathcal{M}} \sqrt{\frac{\Delta^4 x}{2\pi}}\, d\varphi(x).$$

Naturally enough, we assume that the classical integral exists on the right-hand side. Let us explain why this represents a functional integral over the Hilbert space $L_2(\mathcal{M})$. Obviously, the map

$$\varphi \mapsto (\varphi(1), \ldots, \varphi(N))$$

is bijective from the function space $L_2(\mathcal{M})$ onto the space $\mathbb{R}^N$. In this sense, we write $L_2(\mathcal{M}) = \mathbb{R}^N$. Therefore, we may regard F as a functional

$$F : \mathcal{L}_2(\mathcal{M}) \to \mathbb{R},$$

and $\int F(\varphi)\mathcal{D}\varphi$ is an integral over the function space $L_2(\mathcal{M})$.

Example. Let $\lambda(1) > 0, \ldots, \lambda(N) > 0$. Set

$$\zeta_N(s) := \sum_{k=1}^{N} \frac{1}{\lambda(k)^s}\,, \qquad s \in \mathbb{C}.$$

By the magic zeta function formula (7.108) on page 435,

$$\int_{L_2(\mathcal{M})} \mathrm{e}^{-\frac{1}{2}\sum_{x\in\mathcal{M}} \lambda(x)\varphi(x)^2 \Delta^4 x}\, \mathcal{D}\varphi = \prod_{x\in\mathcal{M}} \frac{1}{\sqrt{\lambda(x)}} = \mathrm{e}^{\frac{1}{2}\zeta_N'(0)}.$$

In the special case where $\lambda(x) = 1$ for all $x \in \mathcal{M}$,

$$\int_{L_2(\mathcal{M})} \mathrm{e}^{-\frac{1}{2}\sum_{x\in\mathcal{M}} \varphi(x)^2 \Delta^4 x}\, \mathcal{D}\varphi = 1.$$

The formal limit $N \to +\infty$ and $\Delta x \to 0$ motivates the definition

$$\boxed{\int_{L_2(\mathbb{M}^4)} \mathrm{e}^{-\frac{1}{2}\int_{\mathbb{M}^4} \varphi(x)^2 d^4x}\, \mathcal{D}\varphi := 1.}$$

Terminology. Let us summarize the basic notions to be introduced below. The starting point will be the classical action functional $S[\varphi, J]$. The principle of critical action yields the Euler–Lagrange equation of motion for the quantum field φ under the influence of the external source J.

The fundamental quantum action principal defines the quantum action functional

$$Z(J) := \frac{\int_{x\in\mathcal{M}} e^{\mathrm{i}S[\varphi,J]/\hbar}\; e^{-\eta\langle\varphi|\varphi\rangle}\; \mathcal{D}\varphi}{\int_{x\in\mathcal{M}} e^{\mathrm{i}S[\varphi,0]/\hbar}\; e^{-\eta\langle\varphi|\varphi\rangle}\; \mathcal{D}\varphi}$$

as a functional integral depending on the classical action functional $S[\varphi, J]$ and by averaging over all possible quantum fields φ.

Here, we introduce the fixed small parameter $\eta > 0$ in order to force the convergence of the integrals. The quantum action functional $Z = Z(J)$ is normalized by the condition $Z(0) = 1$. The quantum action functional $Z(J)$ is related to quantum fluctuations that are added to the classical quantum field. These quantum fluctuations are described by correlation functions. In addition, we will introduce the extended quantum action functional

$$Z = Z(J, \varphi)$$

in (7.127) below. The relation between the function $Z = Z(J)$ and the extended function $Z = Z(J, \varphi)$ will be given by

$$\boxed{Z(J) = \frac{Z(J,0)}{Z(0,0)}.}$$

To simplify notation, we will use the same symbol Z for both $Z(J)$ and $Z(J,\varphi)$. Using the extended function $Z = Z(J,\varphi)$, we also introduce the scattering functional

$$\boxed{\mathsf{S}(\varphi) := \frac{Z(0,\varphi)}{Z(0,0)}.}$$

This way, $Z(J)$ and $\mathsf{S}(\varphi)$ are dual quantities along with the normalization condition $Z(0) = \mathsf{S}(0) = 1$. For describing the crucial properties of quantum fields we will use the following functions:

(i) the n-point correlation function $C_n = C_n(x_1, \dots, x_n)$;
(ii) the n-point scattering function $\mathsf{S}_n = \mathsf{S}_n(x_1, \dots, x_n)$;
(iii) the causal n-point correlation function $\mathcal{C}_n = \mathcal{C}_n(x_1, \dots, x_n)$, and
(iv) the n-point vertex function $V_n = V_n(x_1, \dots, x_n)$.

By definition,

$$C_n(x_1, \dots, x_n) := \left(\frac{\hbar}{\mathrm{i}}\right)^n \frac{\delta^n Z(J)}{\delta J(x_1) \cdots \delta J(x_n)}|_{J=0}$$

and

$$\mathsf{S}_n(x_1, \dots, x_n) := \frac{\delta^n \mathsf{S}(\varphi)}{\delta\varphi(x_1) \cdots \delta\varphi(x_n)}|_{\,\varphi=0}\,.$$

Here, $Z = Z(J)$ and $\mathsf{S} = \mathsf{S}(\varphi)$ are called generating functionals. Switching off the interaction by putting the coupling constant to zero, $\kappa = 0$, the correlation function C_n passes over to the so-called free correlation function, $C_{n,\mathrm{free}}$. Quantum field theory can be based on two magic formulas (QA) and (LSZ). Let us briefly discuss this.

(QA) The quantum action formula reduces the correlation function C_n to the free correlation functions $C_{02}, C_{04}, \dots$ In addition, the free correlation functions are determined by the response function which describes the solution $\varphi = R_\varepsilon J$ of the linearized (and regularized) equation of motion with the regularization parameter $\varepsilon > 0$.
(LSZ) The Lehmann–Szymanzik–Zimmermann formula reduces the scattering function S_n to the correlation function C_n.

We will show in Sect. 14.2.5 that the scattering function $\mathsf{S}_{\mathsf{m+n}}$ allows us to compute the S-matrix elements. This yields the essential information about scattering processes of m incoming particles and n outgoing particles.

The causal correlation function $\mathcal{C}_n$ was introduced in Sect. 7.22. We will show in Chap. 15 on the operator approach to quantum field theory that $C_n = \mathcal{C}_n$ for all n.

The function $\mathcal{C}_n$ is also called the n-point Green's function.

In the literature, one uses the notations G_n or τ_n for C_n synonymously. The n-point vertex function V_n is defined in terms of the correlation function C_n. The vertex functions extract important information from the correlation functions. The local quantum action principle tells us that the functional derivatives

$$\frac{\delta Z(J)}{\delta J(x)}, \qquad x \in \mathcal{M}$$

satisfy an equation which is called the Dyson–Schwinger equation.

Warning. For historical reasons, the physical quantity 'action' is denoted by the letter S. In 1943 Heisenberg introduced the S-matrix. Here, the letter S refers to 'scattering'. We will use the italic symbol $S[\varphi]$ and the san-serif symbol $\mathsf{S}(\varphi)$ in order to distinguish between the action functional and the scattering functional, respectively.

7.24.3 The Principle of Critical Action

We are given the action

$$S[\varphi, J] := \sum_{x \in \mathcal{M}} \mathcal{L}(\varphi)(x) \Delta^4 x \tag{7.119}$$

along with the Lagrangian density

$$\mathcal{L}(\varphi)(x) := \tfrac{1}{2}\varphi(x)(D\varphi)(x) + \tfrac{1}{2}\mathrm{i}\varepsilon\varphi(x)^2 + \kappa\mathcal{L}_{\mathrm{int}}(\varphi)(x) + \varphi(x)J(x).$$

In terms of the inner product on $L_2(\mathcal{M})$, the action looks like

$$\boxed{S[\varphi, J] = \tfrac{1}{2}\langle\varphi|(D + \mathrm{i}\varepsilon I)\varphi\rangle + \kappa\langle 1|\mathcal{L}_{\mathrm{int}}(\varphi)\rangle + \langle\varphi|J\rangle.} \tag{7.120}$$

In the following sections, the response approach is essentially based on the action $S[\varphi, J]$ of the quantum field φ. This action depends on the fixed external source J. We make the following assumptions:

- The functions φ and J lie in the space $L_2(\mathcal{M})$.
- The operator $D : L_2(\mathcal{M}) \to L_2(\mathcal{M})$ is linear and self-adjoint.
- There exists a real number $\varepsilon_0 > 0$ such that the inverse operator

$$(D + \mathrm{i}\varepsilon I)^{-1} : L_2(\mathcal{M}) \to L_2(\mathcal{M})$$

  exists for all $\varepsilon \in]0, \varepsilon_0[$.
- We fix the coupling constant $\kappa \geq 0$, and we assume that the function $\mathcal{L}_{\mathrm{int}} : \mathbb{R} \to \mathbb{R}$ is a polynomial with real coefficients (e.g., $\mathcal{L}_{\mathrm{int}}(\chi) = -\chi^4$). The nonlinear term $\kappa\mathcal{L}_{\mathrm{int}}$ with $\kappa > 0$ describes interactions.

Theorem 7.37 *The problem of critical action,*

$$S[\varphi, J] = \text{critical!}, \qquad \varphi \in L_2(\mathcal{M}), \tag{7.121}$$

is equivalent to the Euler–Lagrange equation of motion,

$$(D\varphi)(x) + \mathrm{i}\varepsilon\varphi(x) + \kappa\mathcal{L}'_{\mathrm{int}}(\varphi)(x) + J(x) = 0 \tag{7.122}$$

for all $x \in \mathcal{M}$. *This equation of motion is equivalent to*

$$\frac{\delta S[\varphi, J]}{\delta\varphi(x)} = 0 \quad \textit{for all } x \in \mathcal{M}. \tag{7.123}$$

.

Proof. Fix $h \in L_2(\mathcal{M})$. Introduce the function

$$\chi(\tau) := S[\varphi + \tau h, J] \qquad \text{for all} \quad \tau \in \mathbb{R}.$$

By definition, problem (7.121) is equivalent to

$$\chi'(0) = 0 \qquad \text{for all} \quad h \in L_2(\mathcal{M}).$$

Recall that

$$S[\varphi, J] = \tfrac{1}{2}\langle\varphi|D\varphi\rangle + \tfrac{1}{2}\mathrm{i}\varepsilon\langle\varphi|\varphi\rangle + \langle\varphi|J\rangle + \kappa\sum_{x\in\mathcal{M}}\mathcal{L}_{\mathrm{int}}(\varphi)(x)\Delta^4 x.$$

Therefore, the derivative $\chi'(0)$ is equal to

$$\tfrac{1}{2}\langle h|D\varphi\rangle + \tfrac{1}{2}\langle\varphi|Dh\rangle + \mathrm{i}\varepsilon\langle h|\varphi\rangle + \langle h|J\rangle + \kappa\sum_{x\in\mathcal{M}}\mathcal{L}'_{\mathrm{int}}(\varphi)(x)h(x)\Delta^4 x.$$

Since $\langle\varphi|Dh\rangle = \langle D\varphi|h\rangle = \langle h|D\varphi\rangle$,

$$\langle h|D\varphi + \mathrm{i}\varepsilon\varphi + J + \kappa\mathcal{L}'_{\mathrm{int}}(\varphi)\rangle = 0 \qquad \text{for all} \quad h \in L_2(\mathcal{M}).$$

This implies (7.122). Furthermore,

$$\frac{\delta S[\varphi, J]}{\delta\varphi(x)}(h) = \chi'(0) = \sum_{x\in\mathcal{M}}\frac{\delta S[\varphi, J]}{\delta\varphi(x)}\,h(x)\Delta^4 x = 0$$

for all $h \in L_2(\mathcal{M})$. This yields (7.123). □

7.24.4 The Response Function

Set $\kappa = 0$. The linearized equation of motion

$$\boxed{D\varphi + \mathrm{i}\varepsilon\varphi = -J} \tag{7.124}$$

is called the response equation. For a sufficiently small parameter $\varepsilon > 0$, equation (7.124) has the unique solution

$$\varphi = -(D + \mathrm{i}\varepsilon I)^{-1}J.$$

We set $R_\varepsilon := -(D + \mathrm{i}\varepsilon I)^{-1}$. Then $\varphi = R_\varepsilon J$. We write

$$\boxed{\varphi(x) = \sum_{y\in\mathcal{M}}\mathcal{R}_\varepsilon(x, y)J(y)\Delta^4 y, \qquad x \in \mathcal{M}.} \tag{7.125}$$

The kernel $\mathcal{R}_\varepsilon$ of the response operator R_ε is called the response function.

7.24.5 The Global Quantum Action Principle

Fix the real number $\eta > 0$. We define the crucial quantum action functional

$$Z(J) := \mathcal{N} \int_{L_2(\mathcal{M})} \mathrm{e}^{\mathrm{i}S[\varphi,J]/\hbar}\, \mathrm{e}^{-\eta\langle\varphi|\varphi\rangle}\, \mathcal{D}\varphi.$$

The number $\mathcal{N}$ is chosen in such a way that $Z(0) = 1$. Explicitly,

$$\boxed{Z(J) := \frac{\int_{L_2(\mathcal{M})} \mathrm{e}^{\mathrm{i}S[\varphi,J]/\hbar}\, \mathrm{e}^{-\eta\langle\varphi|\varphi\rangle}\, \mathcal{D}\varphi}{\int_{L_2(\mathcal{M})} \mathrm{e}^{\mathrm{i}S[\varphi,0]/\hbar}\, \mathrm{e}^{-\eta\langle\varphi|\varphi\rangle}\, \mathcal{D}\varphi}.} \tag{7.126}$$

Correlation functions. Define

$$C_n(x_1, \ldots, x_n) := \left(\frac{\hbar}{\mathrm{i}}\right)^n \frac{\delta^n Z(J)}{\delta J(x_1) \cdots \delta J(x_n)}|_{J=0}.$$

Hence

$$C_n(x_1, \ldots, x_n) = \frac{\int_{L_2(\mathcal{M})}\ \varphi(x_1) \cdots \varphi(x_n)\, \mathrm{e}^{\mathrm{i}S[\varphi,0]/\hbar}\, \mathrm{e}^{-\eta\langle\varphi|\varphi\rangle}\, \mathcal{D}\varphi}{\int_{L_2(\mathcal{M})}\ \mathrm{e}^{\mathrm{i}S[\varphi,0]/\hbar}\, \mathrm{e}^{-\eta\langle\varphi|\varphi\rangle}\, \mathcal{D}\varphi}\,.$$

Explicitly, $C_n(x_1, \ldots, x_n)$ is equal to

$$\frac{\int_{\mathbb{R}^N} \varphi(x_1) \cdots \varphi(x_N)\, \mathrm{e}^{\sum_{k=1}^N F(\varphi(k)) + \mathrm{i}\varphi(k)J(k)/\hbar}\, \mathrm{e}^{-\eta\langle\varphi|\varphi\rangle}\, d\varphi(1) \cdots d\varphi(N)}{\int_{\mathbb{R}^N} \mathrm{e}^{\sum_{k=1}^N F(\varphi(k))}\, \mathrm{e}^{-\eta\langle\varphi|\varphi\rangle}\, d\varphi(1) \cdots d\varphi(N)}$$

for all $x_1, \ldots, x_n \in \mathcal{M}$. Here, we set

$$F(\varphi(k)) := \frac{\mathrm{i}}{2\hbar}\{\varphi(k)(D\varphi)(k) + 2\kappa\mathcal{L}_{\mathrm{int}}(\varphi)(k)\} - \frac{\varepsilon}{2\hbar}\, \varphi(k)^2.$$

These are classical integrals over the N real variables $\varphi(1), \ldots, \varphi(N)$. The damping factors $\mathrm{e}^{-\varepsilon\varphi(k)^2/2\hbar}$ and $\mathrm{e}^{-\eta\varphi(k)^2}$ with $\varepsilon, \eta > 0$ and $k = 1, \ldots, N$ guarantee the existence of these integrals.[47] Motivated by the theory of probability, we call C_n the n-moment function or the n-correlation function. In particular, we get the following.

(i) $C_n(x_1, \ldots, x_N) \equiv 0$ if $n = 1, 3, 5, \ldots$
(ii) $C_n(x_1, \ldots, x_n)$ is symmetric with respect to the variables $x_1, \ldots, x_n$ if $n = 2, 4, 6, \ldots$

[47] Since the matrix D is real, $|\mathrm{e}^{\mathrm{i}\varphi(k)(D\varphi)(k)}| = 1$.

7.24.6 The Magic Quantum Action Reduction Formula for Correlation Functions

The following two magic reduction formulas (QA) and (LSZ) will be based on the extended quantum action functional

$$\boxed{Z(J,\varphi) := \exp\left\{\frac{\mathrm{i}\kappa}{\hbar}\sum_{x\in\mathcal{M}}\Delta^4 x\;\mathcal{L}_{\mathrm{int}}\left(\frac{\hbar}{\mathrm{i}}\,\frac{\delta}{\delta J(x)}\right)\right\}\cdot Z_{\mathrm{free}}(J,\varphi)} \tag{7.127}$$

along with

$$Z_{\mathrm{free}}(J,\varphi) := \mathrm{e}^{\frac{1}{2}\mathrm{i}\langle J|R_\varepsilon J\rangle/\hbar}\;\mathrm{e}^{\mathrm{i}\langle\varphi|J\rangle/\hbar}.$$

Explicitly,

$$Z_{\mathrm{free}}(J,\varphi) = \mathrm{e}^{\frac{1}{2}\mathrm{i}\sum_{x,y\in\mathcal{M}}J(x)\mathcal{R}_\varepsilon(x,y)J(y)\Delta^4x\Delta^4y/\hbar}\;\mathrm{e}^{\mathrm{i}\sum_{x\in\mathcal{M}}\varphi(x)J(x)\Delta^4x/\hbar}.$$

Observe that, in contrast to the definition of $Z(J)$ above, the extended functional $Z(J,\varphi)$ is formulated without using functional integrals. This functional only depends on the response function $\mathcal{R}_\varepsilon$. In particular, for vanishing coupling constant, $\kappa = 0$, we get $Z(J,\varphi) = Z_{\mathrm{free}}(J,\varphi)$.

Theorem 7.38 *For each source $J \in L_2(\mathcal{M})$, the quantum action functional is given by*

$$Z(J) = \frac{Z(J,0)}{Z(0,0)}. \tag{7.128}$$

This so-called magic quantum action formula (QA) reduces the computation of the quantum action functional $Z(J)$ to the response function $\mathcal{R}_\varepsilon = \mathcal{R}_\varepsilon(x,y)$, by (7.127). Note that equation (7.127) is to be understood in the sense of a formal power series expansion with respect to the coupling constant κ. This means that the left-hand side and the right-hand side of (7.127) are formal power series expansions with identical complex coefficients. Consequently, equation (7.127) represents a rigorous relation in each order of perturbation theory with respect to κ. This follows from our assumption that $\mathcal{L}_{\mathrm{int}}(\zeta)$ is a polynomial with respect to the variable ζ. Recall that the correlation functions are given by the following functional derivatives

$$\boxed{C_n(x_1,\ldots,x_n) := \left(\frac{\hbar}{\mathrm{i}}\right)^n \frac{\delta^n Z(J)}{\delta J(x_1)\cdots\delta J(x_n)}\Big|_{\,J=0}} \tag{7.129}$$

if $n = 2,4,6,\ldots$. Furthermore, $C_n \equiv 0$ if $n = 1,3,5,\ldots$

Proof of Theorem 7.38. To simplify notation, set $\hbar := 1$. By (7.126), the functional $Z(J)$ is equal to

$$\mathcal{N}\int_{L_2(\mathcal{M})}\mathrm{e}^{\frac{1}{2}\mathrm{i}\langle\varphi|D\varphi\rangle-\frac{1}{2}\varepsilon\langle\varphi|\varphi\rangle+\mathrm{i}\langle\varphi|J\rangle}\;\mathrm{e}^{\mathrm{i}\kappa\sum_{x\in\mathcal{M}}\Delta^4x\;\mathcal{L}_{\mathrm{int}}(\varphi)(x)}\;\mathrm{e}^{-\eta\langle\varphi|\varphi\rangle}\;\mathcal{D}\varphi.$$

The number $\mathcal{N}$ has to be chosen in such a way that $Z(0) = 1$.

(I) The moment trick. We will use an argument which is similar to equation (7.100) on page 432. Define

$$Z_{\text{free}}(J) := \int_{L_2(\mathcal{M})} \mathrm{e}^{\frac{1}{2}\mathrm{i}\langle\varphi|D\varphi\rangle - \frac{1}{2}\varepsilon\langle\varphi|\varphi\rangle + \mathrm{i}\langle\varphi|J\rangle} \, \mathrm{e}^{-\eta\langle\varphi|\varphi\rangle} \, \mathcal{D}\varphi.$$

This corresponds to $Z(J)$ by switching off the interaction, $\kappa = 0$. For the functional derivative,

$$\frac{1}{\mathrm{i}} \frac{\delta Z_{\text{free}}(J)}{\delta J(x)} = \int_{L_2(\mathcal{M})} \varphi(x) \, \mathrm{e}^{\frac{1}{2}\mathrm{i}\langle\varphi|D\varphi\rangle - \frac{1}{2}\varepsilon\langle\varphi|\varphi\rangle + \mathrm{i}\langle\varphi|J\rangle} \, \mathrm{e}^{-\eta\langle\varphi|\varphi\rangle} \, \mathcal{D}\varphi.$$

Power series expansion yields $\mathrm{e}^z = 1 + z + \frac{1}{2}z^2 + \ldots$ Hence

$$\mathrm{e}^{\mathrm{i}\kappa \sum_{x\in\mathcal{M}} \Delta^4 x \, \mathcal{L}_{\text{int}}(\varphi)(x)} = 1 + \mathrm{i}\kappa \sum_{x\in\mathcal{M}} \Delta^4 x \, \mathcal{L}_{\text{int}}(\varphi)(x) + \ldots$$

This implies

$$\begin{aligned} Z(J) &= \mathcal{N} \left\{ 1 + \mathrm{i}\kappa \sum_{x\in\mathcal{M}} \mathcal{L}_{\text{int}} \left(\frac{1}{\mathrm{i}} \frac{\delta}{\delta J(x)} + O(\kappa^2) \right) \right\} Z_{\text{free}}(J) \\ &= \mathcal{N} \, \exp \left\{ \mathrm{i}\kappa \sum_{x\in\mathcal{M}} \mathcal{L}_{\text{int}} \left(\frac{1}{\mathrm{i}} \frac{\delta}{\delta J(x)} \right) \right\} \, Z_{\text{free}}(J). \end{aligned}$$

(II) The method of stationary phase. Recall that $R_\varepsilon := -(D + \mathrm{i}\varepsilon I)^{-1}$. By (7.110) on page 436,

$$Z_{\text{free}}(J) = \mathrm{e}^{\frac{1}{2}\mathrm{i}\langle J|R_\varepsilon J\rangle} \, Z_{\text{free}}(0).$$

This way, we get $Z(J) = \text{const} \cdot Z(J, 0)$. Finally, the constant is uniquely determined by the normalization condition $Z(0) = 1$. □

7.24.7 The Magic LSZ Reduction Formula for Scattering Functions

Use now the extended quantum action functional $Z(J, \varphi)$ in order to define the scattering functional

$$\boxed{\mathsf{S}(\varphi) := \frac{Z(0, \varphi)}{Z(0, 0)}.}$$

The functional derivative

$$\mathsf{S}_n(x_1, \ldots, x_N) := \frac{\delta^n \mathsf{S}(\varphi)}{\delta\varphi(x_1) \cdots \delta\varphi(x_n)} \Big|_{\varphi=0} \tag{7.130}$$

is called the n-point scattering function. The physical meaning of this function will be discussed in Sect. 14.2.5 on page 785.

Theorem 7.39 *For all $x_1, \ldots, x_n \in \mathcal{M}$ and $n = 1, 2, \ldots$, we have the following LSZ reduction formula:*

$$\mathsf{S}_n(x_1, \ldots, x_n) = \frac{1}{(\mathrm{i}\hbar)^n} \{(D_{x_1} + \mathrm{i}\varepsilon I) \cdots (D_{x_n} + \mathrm{i}\varepsilon I) C_n\} (x_1, \ldots, x_n).$$

This fundamental formula relates the scattering function S_n to the correlation function C_n. Here, the symbol D_{x_k} stands for the linear operator D from the equation of motion (7.122) on page 446 which acts on the kth variable of the correlation function C_n.

Proof. The proof is an elementary consequence of both the chain rule for partial derivatives and the relation

$$R_\varepsilon (D + \mathrm{i}\varepsilon)^{-1} = -I.$$

In order to simplify notation, we write D and R instead of $D + \mathrm{i}\varepsilon$ and R_ε, respectively.

Step 1: To understand the simple idea of the proof, let us start with the case where $N = 1$. Then

$$Z_{\mathrm{free}}(J, \varphi) := \mathrm{e}^{\mathrm{i}RJ^2/2\hbar}\, \mathrm{e}^{\mathrm{i}\varphi J/\hbar}, \qquad \varphi, J \in \mathbb{R}.$$

Moreover, D and R are nonzero real numbers. Let us also introduce the differential operators

$$\frac{\delta}{\delta J} := \frac{1}{\Delta^4 x} \frac{d}{dJ}, \qquad A := \frac{\delta^k}{\delta J^k}, \qquad k = 0, 1, 2, \ldots.$$

Let $n = 1, 2, \ldots$ We claim that

$$\boxed{\left(\frac{\hbar}{\mathrm{i}}\right)^n \frac{\delta^n A Z_{\mathrm{free}}(J, \varphi)}{\delta \varphi^n} = (-D)^n \left(\frac{\hbar}{\mathrm{i}}\right)^n \frac{\delta^n A Z_{\mathrm{free}}(J, \varphi)}{\delta J^n}} \tag{7.131}$$

at the point $J = 0, \varphi = 0$. This is the key relation of our proof.

(i) First choose $n = 1$. Obviously,

$$\frac{\hbar}{\mathrm{i}} \frac{\delta Z_{\mathrm{free}}(J, \varphi)}{\delta \varphi} = J Z_{\mathrm{free}}(J, \varphi).$$

Furthermore, noting that $(-D)R = 1$, relation (7.131) follows from

$$\frac{\hbar}{\mathrm{i}} \frac{\delta Z_{\mathrm{free}}(J, \varphi)}{\delta J} = (RJ + \varphi) Z_{\mathrm{free}}(J, \varphi).$$

(ii) Now choose $n = 2$. Then

$$\left(\frac{\hbar}{\mathrm{i}}\right)^2 \frac{\delta^2 Z_{\mathrm{free}}(J, \varphi)}{\delta \varphi^2} = J^2 Z_{\mathrm{free}}(J, \varphi). \tag{7.132}$$

Moreover,

$$\left(\frac{\hbar}{\mathrm{i}}\right)^2 \frac{\delta^2 Z_{\mathrm{free}}(J,\varphi)}{\delta J^2} = (RJ+\varphi)^2 Z_{\mathrm{free}}(J,\varphi) + \frac{\hbar R}{\mathrm{i}\Delta^4 x} Z_{\mathrm{free}}(J,\varphi).$$

This implies

$$(-D)^2 \left(\frac{\hbar}{\mathrm{i}}\right)^2 \frac{\delta^2 Z_{\mathrm{free}}(J,\varphi)}{\delta J^2} = J^2 Z_{\mathrm{free}}(J,\varphi) + W(J,\varphi).$$

Here, the symbol $W(J,\varphi)$ stands for the remaining terms. Below we will only need the fact that $W(0,0)=0$ and $AW(J,\varphi)=0$ if $J=0,\varphi=0$. Hence

$$\left(\frac{\hbar}{\mathrm{i}}\right)^2 \frac{\delta^2 AZ_{\mathrm{free}}(J,\varphi)}{\delta J^2} = A\{J^2 Z_{\mathrm{free}}(J,\varphi)\} + AW(J,\varphi).$$

Letting $J=\varphi=0$, it follows from (7.132) that equation (7.131) holds true for $n=2$.

(iii) For $n=3,4,\ldots$, the proof of (7.131) proceeds analogously, by induction.

Step 2: Let $N=2,3,\ldots$ In contrast to Step 1, we now have to use partial derivatives with respect to the real variables $J(1),\ldots,J(N),\varphi(1),\ldots,\varphi(N)$. Let $D(x,y)$ and $\mathcal{R}(x,y)\Delta^4 x$ denote the entries of the matrix D and R, respectively. Here, $R=-D^{-1}$. Then

$$\langle J|RJ\rangle = \sum_{x,y\in\mathcal{M}} J(x)\mathcal{R}(x,y)J(y)\Delta^4 x\Delta^4 y,$$

and $\sum_{y\in\mathcal{M}} D(x,y)\mathcal{R}(y,z)\Delta^4 y = \delta_{x,z}$. The same argument as in (7.131) tells us that the functional derivative

$$\left(\frac{\hbar}{\mathrm{i}}\right)^n \frac{\delta^n \mathsf{S}(\varphi)}{\delta\varphi(x_1)\cdots\delta\varphi(x_n)}$$

is equal to

$$(-1)^n \left(\frac{\hbar}{\mathrm{i}}\right)^n \sum_{y_1,\ldots,y_n\in\mathcal{M}} D(x_1,y_1)\cdots D(x_n,y_n)\frac{\delta^n Z(J)}{\delta J(y_1)\cdots\delta J(y_n)}$$

if we put $J=\varphi=0$. Replacing the functional derivatives by S_n and C_n, we obtain that $\left(\frac{\hbar}{\mathrm{i}}\right)^n \mathsf{S}_n(x_1,\ldots,x_n)$ is equal to

$$(-1)^n \sum_{y_1,\ldots,y_n\in\mathcal{M}} D(x_1,y_1)\cdots D(x_n,y_n)C_n(y_1,\ldots,y_n).$$

This is the desired LSZ reduction formula. □

7.24.8 The Local Quantum Action Principle

The following theorem is called the local quantum action principle.

Theorem 7.40 *The functional derivative of the quantum action functional $Z = Z(J)$ satisfies the following Dyson–Schwinger integro-differential equation*

$$\frac{\hbar}{\mathrm{i}} \frac{\delta Z(J)}{\delta J(x)} = \sum_{y \in \mathcal{M}} \mathcal{R}_\varepsilon(x,y) J(y) Z(J) \Delta^4 y + \\ + \kappa \sum_{y \in \mathcal{M}} \mathcal{R}_\varepsilon(x,y) \mathcal{L}'_{\mathrm{int}} \left(\frac{\hbar}{\mathrm{i}} \frac{\delta}{\delta J(y)} \right) Z(J) \, \Delta^4 y. \quad (7.133)$$

For each $x \in \mathcal{M}$, this equation is to be understood in the sense of formal power series expansions with respect to the coupling constant κ. Therefore, this equation holds true in each order of perturbation theory with respect to κ. Since $R_\varepsilon = -(D + \mathrm{i}\varepsilon I)^{-1}$, the Dyson–Schwinger equation (7.133) is equivalent to the following equation

$$\frac{\hbar}{\mathrm{i}} \left\{ (D + \mathrm{i}\varepsilon I) \frac{\delta Z(J)}{\delta J} \right\} (x) = -J(x) Z(J) - \\ - \kappa \, \mathcal{L}'_{\mathrm{int}} \left(\frac{\hbar}{\mathrm{i}} \frac{\delta}{\delta J(x)} \right) Z(J) \quad (7.134)$$

for all $x \in \mathcal{M}$. This equation resembles the equation of motion for the classical field φ. In the next section, we will show that the Dyson–Schwinger equation is the equation of motion for the mean field φ_{mean} which includes quantum fluctuations. The trick of the following proof is to use the identity (7.138) below which is a special case of the Baker–Campbell–Hausdorff formula in the theory of Lie algebras.

Proof of Theorem 7.40. Recall that

$$Z(J) = \mathrm{e}^B \, Z_{\mathrm{free}}(J)$$

where $Z_{\mathrm{free}}(J) := \mathrm{e}^{\mathrm{i}\langle J | R_\varepsilon J \rangle / 2\hbar}$ along with

$$\langle J | R_\varepsilon J \rangle = \sum_{x,y \in \mathcal{M}} J(x) \mathcal{R}_\varepsilon(x,y) J(y) \Delta^4 x \Delta^4 y.$$

Furthermore, we use the differential operator

$$B := \frac{\kappa \mathrm{i}}{\hbar} \sum_{y \in \mathcal{M}} \Delta^4 y \, \mathcal{L}'_{\mathrm{int}} \left(\frac{\hbar}{\mathrm{i}} \frac{\delta}{\delta J(y)} \right).$$

Observe that $\mathcal{L}_{\mathrm{int}}(\zeta) = a_0 + a_1 \zeta + \ldots + a_r \zeta^r$ is a real polynomial with respect to the variable ζ.

(I) Two key identities. We claim that

$$\frac{\hbar}{\mathrm{i}}\,\frac{\delta Z_{\mathrm{free}}(J)}{\delta J(x)} = \sum_{y\in\mathcal{M}} \mathcal{R}_\varepsilon(x,y)J(y)Z_{\mathrm{free}}(J)\Delta^4 y, \tag{7.135}$$

and

$$J(x)\mathrm{e}^B Z_{\mathrm{free}}(J) = \mathrm{e}^B J(x)Z_{\mathrm{free}}(J) - \kappa\mathcal{L}'_{\mathrm{int}}\left(\frac{\hbar}{\mathrm{i}}\frac{\delta}{\delta J(x)}\right) Z_{\mathrm{free}}(J). \tag{7.136}$$

(II) We first show that (I) implies the desired key relation (7.134). It follows from (7.135) that

$$\frac{\hbar}{\mathrm{i}}\left\{(D+\mathrm{i}\varepsilon I)\frac{\delta Z_{\mathrm{free}}(J)}{\delta J}\right\}(x) = -J(x)Z_{\mathrm{free}}(J).$$

Furthermore, the matrix operator D commutes with the differential operators $\frac{\delta}{\delta J(x)}$ and B. Therefore,

$$\begin{aligned}\frac{\hbar}{\mathrm{i}}\left\{(D+\mathrm{i}\varepsilon I)\frac{\delta Z(J)}{\delta J}\right\}(x) &= \mathrm{e}^B\cdot\frac{\hbar}{\mathrm{i}}\left\{(D+\mathrm{i}\varepsilon I)\frac{\delta Z_{\mathrm{free}}(J)}{\delta J}\right\}(x)\\ &= -\mathrm{e}^B J(x)Z_{\mathrm{free}}(J).\end{aligned}$$

By (7.136), this is equal to $-J(x)Z(J) - \kappa\mathcal{L}'_{\mathrm{int}}(\frac{\hbar}{\mathrm{i}}\,\frac{\delta}{\delta J(x)})Z(J)$.

(III) Proof of the first key relation (7.135). This follows immediately from the definition $Z_{\mathrm{free}}(J) := \mathrm{e}^{\mathrm{i}\langle J|R_\varepsilon J\rangle/2\hbar}$.

(IV) Proof of the second key relation (7.136). Recall the definition $[A,B]_- := AB - BA$.

Step 1: Let $n = 1, 2, \ldots$. We show that

$$\left[\frac{\delta^n}{\delta J(x)^n}, J(y)\right]_- F(y) = n\delta(x,y)\frac{\delta^{n-1}}{\delta J(y)^{n-1}}F(y). \tag{7.137}$$

First consider the case where $n = 1$. Then

$$\begin{aligned}\left[\frac{\delta}{\delta J(x)}, J(y)\right]_- F(y) &= \frac{\delta}{\delta J(x)}\{J(y)F(y)\} - J(y)\frac{\delta F(y)}{\delta J(x)}\\ &= \frac{\delta J(y)}{\delta J(x)}\,F(y) = \frac{\delta_{x,y}}{\Delta^4 x}\,F(y) = \delta(x,y)F(y).\end{aligned}$$

Now let $n = 2$. Then

$$\begin{aligned}\left[\frac{\delta^2}{\delta J(x)^2}, J(y)\right]_- F(y) &= \frac{\delta}{\delta J(x)}\left[\frac{\delta}{\delta J(x)}, J(y)\right]_- F(y) +\\ &+ \left[\frac{\delta}{\delta J(x)}, J(y)\right]_- \frac{\delta F(y)}{\delta J(y)} = 2\delta(x,y)\frac{\delta F(y)}{\delta J(y)}.\end{aligned}$$

The general induction proof proceeds analogously.

Step 2: We show that

$$[B, J(x)]_- F(z) = \kappa \mathcal{L}'_{\text{int}} \left(\frac{\hbar}{\mathrm{i}} \frac{\delta}{\delta J(x)} \right) F(z).$$

In fact, by (7.137),

$$\begin{aligned} [B, J(x)]_- F(z) &= \frac{\kappa \mathrm{i}}{\hbar} \sum_{y \in \mathcal{M}} \Delta^4 y \, \frac{\hbar}{\mathrm{i}} \, \delta(x, y) \, \mathcal{L}'_{\text{int}} \left(\frac{\hbar}{\mathrm{i}} \frac{\delta}{\delta y} \right) F(z) \\ &= \kappa \mathcal{L}'_{\text{int}} \left(\frac{\hbar}{\mathrm{i}} \frac{\delta}{\delta x} \right) F(z). \end{aligned}$$

Step 3: We show that

$$[\,[J(x), B]_-, B]_- = 0.$$

In fact, since $[J(x), 1]_- = 0$, it follows from (7.137) that

$$[J(x), B]_- = \sum_{y \in \mathcal{M}} b_0 \delta(x, y) + b_1 \frac{\delta}{\delta y} + \dots$$

This differential operator commutes with the differential operator B.

Step 4: We have the special Baker–Campbell–Hausdorff formula

$$J(x) \mathrm{e}^B = \mathrm{e}^B J(x) + [J(x), B]_-. \tag{7.138}$$

This is a consequence of Step 3. The proof for matrices can be found in Problem 7.4 on page 494. The proof of (7.138) proceeds analogously to the proof for matrices. Note that (7.138) is to be understood in the sense of formal power series expansions with respect to κ. Using Step 2 and Step 4, we get the key relation (7.136). □

7.24.9 Simplifying the Computation of Quantum Effects

We have seen that correlation functions know all about quantum fields including scattering processes via the LSZ reduction formula. However, as a rule, the computation of concrete physical effects is lengthy and time consuming. Therefore, physicists have invented tools in order to simplify computations, namely,

(i) the family of reduced correlation functions and
(ii) the mean field approach (averaged quantum fluctuations).

Reduced correlation functions. The basic idea is to start with so-called reduced correlation functions which allow us the computation of the correlation functions. Schematically,

response function ⇒ reduced correlation functions ⇒
⇒ correlation functions ⇒ scattering functions (S matrix).

The S matrix knows all about scattering processes.

Mean field approach. In order to get typical information about the influence of quantum fluctuations, we average the quantum fluctuations over all possible classical field configurations. Schematically,

- classical field $\varphi \Rightarrow$ mean field φ_{mean};
- classical action $S[\varphi, J] \Rightarrow$ effective quantum action S_{eff};
- response function ⇒ vertex functional $V \Rightarrow$ effective quantum action

$$S_{\text{eff}} = V(\varphi_{\text{mean}}).$$

The effective quantum action depends on the vertex functional V which can be described by vertex functions $V(x_1, \dots, x_n),\ n = 1, 2, \dots$

From the computational point of view, note the following.

The computations concerning the reduced correlation functions and the mean field approach depend on the coupling constant κ, and they can be carried out in each order of perturbation theory.

7.24.10 Reduced Correlation Functions

Since $Z(0) = 1$, it makes sense to define the reduced quantum action functional by

$$\boxed{Z_{\text{red}}(J) := \ln Z(J).}$$

Hence

$$Z(J) = \mathrm{e}^{Z_{\text{red}}(J)}.$$

Parallel to the n-correlation functions, we introduce the so-called n-correlation functions

$$C_{n,\text{reduced}}(x_1, \dots, x_n) = \left(\frac{\hbar}{\mathrm{i}}\right)^n \frac{\delta^n Z_{\text{red}}(J)}{\delta J(x_1) \cdots \delta J(x_n)}, \qquad n = 1, 2, \dots$$

The reduced correlation functions simplify the computation of correlation functions. Examples will be considered in Prop. 7.45 on page 472. Roughly speaking, the reduced correlation functions have the advantage that they do not contain contributions that arise trivially from the multiplication of correlation functions of lower order.

7.24.11 The Mean Field Approximation

In physics, effective quantities can be obtained by averaging.
Folklore

The classical field. Fix the external source J. By the equation of motion (7.122) on page 446, the classical field φ satisfies the following (discrete) integral equation:

$$\varphi(x) = \sum_{y\in\mathcal{M}} \mathcal{R}_\varepsilon(x,y)\,\{J(y) + \kappa\mathcal{L}'_{\text{int}}(\varphi)(y)\}\,\Delta^4 y. \tag{7.139}$$

The kernel of this integral equation is the response function $\mathcal{R}_\varepsilon$. Setting $\kappa = 0$, we obtain the classical free field

$$\varphi_{\text{free}}(x) = \sum_{y\in\mathcal{M}} \mathcal{R}_\varepsilon(x,y)J(y)\Delta^4 y.$$

We now compute the interacting field by the following iterative method

$$\varphi_{n+1}(x) = \varphi_{\text{free}}(x) + \kappa \sum_{y\in\mathcal{M}} \mathcal{R}_\varepsilon(x,y)\mathcal{L}'_{\text{int}}(\varphi_n)(y)\Delta^4 y, \qquad n = 0,1,2,\ldots$$

with the free field as initial approximation, that is, we set $\varphi_0 := \varphi_{\text{free}}$. This yields the first approximation

$$\varphi_1(x) = \varphi_{\text{free}}(x) + \kappa \sum_{y\in\mathcal{M}} \mathcal{R}_\varepsilon(x,y)\mathcal{L}'_{\text{int}}(\varphi_{\text{free}})(y)\Delta^4 y.$$

Quantum fields differ from classical fields by adding quantum effects. Intuitively, we have to add quantum fluctuations. The complete theory is based on the quantum action principle considered in Sect. 7.24.6 on page 448. At this point, we want to discuss a simplified version of the complete theory based on the averaging of quantum effects.

Averaging quantum fluctuations. Fix the external source J and define

$$\varphi_{\text{mean}}(x) := \frac{\hbar}{\mathrm{i}Z(J)}\frac{\delta Z(J)}{\delta J(x)}, \qquad x \in \mathcal{M}.$$

This is called the mean field approximation of the classical field (7.139) corresponding to J. To justify this terminology, note that the global quantum action principle (7.126) on page 447 implies that

$$\varphi_{\text{mean}}(x) = \frac{\int_{L_2(\mathcal{M})} \varphi(x)\mathrm{e}^{\mathrm{i}S[\varphi,J]/\hbar}\cdot \mathrm{e}^{-\eta\langle\varphi|\varphi\rangle}\,\mathcal{D}\varphi}{\int_{L_2(\mathcal{M})} \mathrm{e}^{\mathrm{i}S[\varphi,J]/\hbar}\cdot \mathrm{e}^{-\eta\langle\varphi|\varphi\rangle}\,\mathcal{D}\varphi}.$$

Thus, the mean field φ_{mean} is obtained by averaging over all possible classical fields φ. Note that the mean field φ_{mean} depends on the choice of the external source J and the small damping parameter $\eta > 0$.

The relation between the mean field and the reduced quantum action functional. Using the reduced quantum action functional, the mean field is given by

$$\boxed{\varphi_{\mathrm{mean}}(x) = \frac{\hbar}{\mathrm{i}} \frac{\delta Z_{\mathrm{red}}(J)}{\delta J(x)}.} \tag{7.140}$$

Switching off the coupling constant, $\kappa = 0$, we have

$$Z_{\mathrm{free}}(J) = \mathrm{e}^{Z_{\mathrm{free,red}}(J)}$$

along with

$$Z_{\mathrm{free,red}}(J) := \frac{\mathrm{i}}{2\hbar}\langle J|R_\varepsilon J\rangle = \frac{\mathrm{i}}{2\hbar}\sum_{x,y\in\mathcal{M}} J(x)\mathcal{R}_\varepsilon(x,y)J(y)\Delta^4 x\Delta^4 y.$$

It follows from the Dyson–Schwinger equation (7.133) on page 452 that

$$\varphi_{\mathrm{mean}}(x) = \varphi_{\mathrm{free}}(x) + \\ +\kappa \mathrm{e}^{-Z_{\mathrm{red}}(J)} \sum_{y\in\mathcal{M}} \mathcal{R}_\varepsilon(x,y)\mathcal{L}'_{\mathrm{int}}\left(\frac{\hbar}{\mathrm{i}}\frac{\delta}{\delta J(y)}\right)\mathrm{e}^{Z_{\mathrm{red}}(J)}\ \Delta^4 y.$$

This resembles the classical equation of motion

$$\varphi(x) = \varphi_{\mathrm{free}}(x) + \kappa \sum_{y\in\mathcal{M}} \mathcal{R}_\varepsilon(x,y)\mathcal{L}'_{\mathrm{int}}(\varphi)(y)\Delta^4 y.$$

From the physical point of view, the quantum field differs from the classical field by quantum corrections.

> *The mean field φ_{mean} can be regarded as a field which differs from the classical field by adding the average of quantum corrections.*

In what follows we will use the Legendre transformation in order to introduce vertex functions and the effective action of a quantum field.

7.24.12 Vertex Functions and the Effective Action

The Legendre transformation. Equation (7.140) defines a map

$$(J(1),\dots,J(N)) \mapsto (\varphi_{\mathrm{mean}}(1),\dots,\varphi_{\mathrm{mean}}(N)) \tag{7.141}$$

from the external sources to the mean fields. Let us assume that this map is a diffeomorphism from $\mathbb{R}^N$ onto $\mathbb{R}^N$. This means that there exists a one-to-one relation between external sources and mean fields. Let us write this map as

$$\varphi_{\mathrm{mean}} = \varphi_{\mathrm{mean}}(J).$$

Conversely, $J = J(\varphi_{\text{mean}})$. Our goal is to replace external sources, J, by mean fields, φ_{mean} according to (7.141). In classical mechanics, one uses the Legendre transformation in order to pass from velocity to momentum and from the Lagrangian density to the Hamiltonian function. Parallel to this classical tool, let us introduce the Legendre transformation

$$\boxed{J \mapsto \varphi_{\text{mean}}, \qquad Z_{\text{red}}(J) \mapsto V(\varphi_{\text{mean}})}$$

given by (7.140) and

$$V(\varphi_{\text{mean}}) := \frac{\hbar}{\mathrm{i}} Z_{\text{red}}(J) - \sum_{x \in \mathcal{M}} \varphi_{\text{mean}}(x) J(x) \Delta^4 x.$$

The functional V is called the vertex functional.

Proposition 7.41 *We have*

$$\frac{\delta V(\varphi_{\text{mean}})}{\delta \varphi_{\text{mean}}(x)} = -J(x)$$

where the source J is related to the mean field φ_{mean} by the Legendre transformation (7.141) *above.*

Proof. For the total differential,

$$dV = \sum_{x \in \mathcal{M}} \frac{\partial V(\varphi_{\text{mean}})}{\partial \varphi_{\text{mean}}(x)} \, d\varphi_{\text{mean}}(x).$$

Furthermore,

$$\begin{aligned} dV = \sum_{x \in \mathcal{M}} \frac{\hbar}{\mathrm{i}} \frac{\partial Z_{\text{red}}(J)}{\partial J(x)} \, dJ(x) - \\ - \sum_{x \in \mathcal{M}} \{ J(x) d\varphi_{\text{mean}}(x) + \varphi_{\text{mean}}(x) dJ(x) \} \Delta^4 x. \end{aligned}$$

By (7.140), the terms with $dJ(x)$ compensate each other. Hence

$$dV = - \sum_{x \in \mathcal{M}} J(x) d\varphi_{\text{mean}}(x) \; \Delta^4 x.$$

This implies

$$\frac{\delta V(\varphi_{\text{mean}})}{\delta \varphi_{\text{mean}}(x)} = \frac{1}{\Delta^4 x} \frac{\partial V(\varphi_{\text{mean}})}{\partial \varphi_{\text{mean}}(x)} = -J(x).$$

□

Vertex functions. Let us define

$$V_n(x_1, \ldots, x_n) := \frac{\delta^n V(\varphi)}{\delta \varphi(x_1) \cdots \delta \varphi(x_n)} \Big|_{\, \varphi = 0}.$$

These functions are called vertex functions. The quantum action functional $Z(J)$ can be expressed by the vertex functional. Explicitly,

$$\boxed{Z(J) = \mathrm{e}^{\mathrm{i}(V(\varphi_{\mathrm{mean}})+\langle\varphi_{\mathrm{mean}}|J\rangle)/\hbar}.} \tag{7.142}$$

Here, the source J is related to the mean field φ_{mean} by the Legendre transformation (7.141).

The effective quantum action. Switching off the interaction, $\kappa = 0$, we get $\varphi_{\mathrm{mean}} = \varphi_{\mathrm{free}}$, and

$$V(\varphi_{\mathrm{free}}) = \sum_{x\in\mathcal{M}} \frac{1}{2}\varphi_{\mathrm{free}}(x)(D\varphi_{\mathrm{free}} + \mathrm{i}\varepsilon\varphi_{\mathrm{free}})(x)\Delta^4 x. \tag{7.143}$$

This is the classical action functional with vanishing source term. To prove (7.143), note that

$$V(\varphi_{\mathrm{free}}) = \frac{\hbar}{\mathrm{i}} Z_{\mathrm{free,red}}(J) - \langle\varphi_{\mathrm{free}}|J\rangle = \frac{1}{2}\langle J|R_\varepsilon J\rangle - \langle\varphi_{\mathrm{free}}|J\rangle.$$

Since $\varphi_{\mathrm{free}} = R_\varepsilon J$, we get $V(\varphi_{\mathrm{free}}) = -\frac{1}{2}\langle J|R_\varepsilon J\rangle$. Finally, it follows from $J = R_\varepsilon^{-1}\varphi_{\mathrm{free}}$ that

$$V(\varphi_{\mathrm{free}}) = -\frac{1}{2}\langle R_\varepsilon^{-1}\varphi_{\mathrm{free}}|\varphi_{\mathrm{free}}\rangle = \frac{1}{2}\langle (D + \mathrm{i}\varepsilon I)\varphi_{\mathrm{free}}|\varphi_{\mathrm{free}}\rangle.$$

This proves (7.143). Now consider the case of general coupling constant, $\kappa \geq 0$. We define

$$\boxed{S_{\mathrm{eff}}[\varphi_{\mathrm{mean}}] := V(\varphi_{\mathrm{mean}})}$$

and call this the effective action of the quantum field. This definition is motivated by equation (7.143).

The effective quantum action adds the average of quantum corrections to the classical action.

7.25 The Discrete φ^4-Model and Feynman Diagrams

It is our goal to apply the preceding response approach to a special discrete model, which is obtained from the continuum φ^4-model by replacing the Fourier transform by the discrete Fourier transform.

The continuum φ^4-model. Set $x = (\mathbf{x}, ct)$. Let

$$\varphi = \varphi(x)$$

be a real-valued function $\varphi : \mathbb{R}^4 \to \mathbb{R}$ which depends on the position vector $\mathbf{x}$ and time t. The solution φ of the nonlinear Klein–Gordon equation

$$\boxed{-\hbar^2\Box\varphi - m_0^2c^2\varphi + \mathrm{i}\varepsilon\varphi - 4\kappa\varphi^3 + J = 0 \qquad \text{on } \mathbb{R}^4} \tag{7.144}$$

models the self-interaction of an uncharged meson of rest mass m_0 along with the coupling constant $\kappa \geq 0$ and the regularization parameter $\varepsilon > 0$. The given smooth function $J : \mathbb{R}^4 \to \mathbb{R}$ describes an external source. The operator

$$\Box := \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \sum_{k=1}^{3}\left(\frac{\partial}{\partial x^k}\right)^2$$

is called the wave operator. The action functional of the meson reads as

$$S[\varphi, J] := \int_\Omega \mathcal{L}(\varphi, \partial\varphi, J)\, d^4x$$

with the Lagrangian density

$$\mathcal{L}(\varphi, J) := \frac{1}{2}\varphi(D + \mathrm{i}\varepsilon)\varphi + \kappa\mathcal{L}_{\mathrm{int}}(\varphi) + \varphi J$$

where we set $D := -\hbar^2\Box - m_0^2c^2$. Because of

$$\boxed{\mathcal{L}_{\mathrm{int}}(\varphi) := -\varphi^4,}$$

we speak of the φ^4-model. Recall that the closure $\mathrm{cl}(\Omega)$ of the open set Ω is obtained from Ω by adding the boundary $\partial\Omega$. The quantity S has the physical dimension of action which equals the product of energy with time.

Theorem 7.42 *Let Ω be a nonempty bounded open subset of $\mathbb{R}^4$. Then, each smooth solution $\varphi : \mathrm{cl}(\Omega) \to \mathbb{R}$ of the variational problem*

$$S[\varphi, J] = \text{critical!} \tag{7.145}$$

with the boundary condition $\varphi = 0$ on $\partial\Omega$ satisfies equation (7.144) *on Ω.*

This is called the principle of critical action for the meson equation (7.144).
Proof. Let $\mathcal{D}(\Omega)$ denote the space of all smooth functions $h : \Omega \to \mathbb{R}$ which vanish outside a compact subset of Ω. In particular, if $h \in \mathcal{D}(\Omega)$, then $h = 0$ on $\partial\Omega$. Let φ be a solution of (7.145). Fix $h \in \mathcal{D}(\Omega)$, and set

$$\chi(\tau) := S[\varphi + \tau h, J] \qquad \text{for all} \quad \tau \in \mathbb{R}.$$

By definition, problem (7.145) is equivalent to

$$\chi'(0) = 0 \qquad \text{for all} \quad h \in \mathcal{D}(\Omega).$$

This means that

$$\int_\Omega \frac{1}{2}hD\varphi + \frac{1}{2}\varphi Dh + h(\mathrm{i}\varepsilon\varphi - 4\kappa\varphi^3 + J)\, d^4x = 0$$

for all $h \in \mathcal{D}(\Omega)$. Using integration by parts, we obtain

$$\int_\Omega h(D\varphi + \mathrm{i}\varepsilon\varphi - 4\kappa\varphi^3 + J)\, d^4x = 0 \qquad \text{for all} \quad h \in \mathcal{D}(\Omega).$$

This implies

$$D\varphi + \mathrm{i}\varepsilon\varphi - 4\kappa\varphi^3 + J = 0 \qquad \text{on } \Omega,$$

by the variational lemma on page 542. □

Let $\varphi \in \mathcal{S}(\mathbb{R}^4)$, that is, the smooth function $\varphi : \mathbb{R}^4 \to \mathbb{C}$ goes rapidly to zero at infinity.[48] Introducing the momentum vector $\mathbf{p}$, the energy E, and the 4-momentum vector $p := (\mathbf{p}, E/c)$ along with

$$px := Et - \mathbf{xp},$$

the Fourier–Minkowski transform of the function φ reads as

$$\hat{\varphi}(p) := \frac{1}{4\pi^2\hbar^2} \int_{\mathbb{R}^4} \varphi(x) \mathrm{e}^{\mathrm{i}px/\hbar}\, d^4x.$$

The inverse transformation is given by

$$\varphi(x) := \frac{1}{4\pi^2\hbar^2} \int_{\mathbb{R}^4} \varphi(x) \mathrm{e}^{-\mathrm{i}px/\hbar}\, d^4p.$$

Using the Fourier–Minkowski transform, the equation

$$(-\hbar^2\Box - m_0^2c^2 + \mathrm{i}\varepsilon)\varphi(x) = f(x), \qquad x \in \mathbb{R}^4$$

passes over to

$$(p^2 - m_0^2c^2 + \mathrm{i}\varepsilon)\hat{\varphi}(p) = \hat{f}(p), \qquad p \in \mathbb{R}^4 \tag{7.146}$$

with $c^2p^2 = E^2 - c^2\mathbf{p}^2$. Hence

$$\hat{\varphi}(p) = \frac{\hat{f}(p)}{p^2 - m_0^2c^2 + \mathrm{i}\varepsilon}, \qquad p \in \mathbb{R}^4. \tag{7.147}$$

Set $D_\varepsilon := -\hbar^2\Box - m_0^2c^2 + \mathrm{i}\varepsilon$. By (7.146),

- the operator $D_\varepsilon := D + \mathrm{i}\varepsilon I$ corresponds to multiplication by the function $p^2 - m_0^2c^2 + \mathrm{i}\varepsilon$ in the Fourier space,

$$(\hat{D}_\varepsilon\hat{\varphi})(p) = (p^2 - m_0^2c^2 + \mathrm{i}\varepsilon)\hat{\varphi}(p), \qquad p \in \mathbb{R}^4;$$

[48] The precise definition can be found in Sect. 11.3.3 on page 614.

- the response operator $R_\varepsilon := -D_\varepsilon^{-1}$ corresponds to division by the function $-(p^2 - m_0^2c^2 + \mathrm{i}\varepsilon)$ in the Fourier space. Hence

$$(\hat{R}_\varepsilon \hat{f})(p) = -\frac{\hat{f}(p)}{p^2 - m_0^2c^2 + \mathrm{i}\varepsilon}, \qquad p \in \mathbb{R}^4. \tag{7.148}$$

The set of all 4-momentum vectors $p \in \mathbb{R}^4$ with

$$p^2 - c^2m_0^2 = 0$$

is called the mass hyperboloid or the mass shell of the meson. Explicitly, this means that

$$E^2 = c^2\mathbf{p}^2 + m_0^2c^4,$$

which is the relation between energy E and momentum vector $\mathbf{p}$ of a free meson. Note that equation (7.148) makes sense, since $\varepsilon > 0$. However, in the critical case where $\varepsilon = 0$, the operator $\hat{R}_0$ is singular if p lives on the mass shell. This is the reason for introducing the regularization term $\mathrm{i}\varepsilon\varphi$ into the meson equation (7.144). The relation of the response operator to the Feynman propagator for the Klein–Gordon equation will be considered in Sect. 14.2.2 on page 775.

Discretization. It is our goal to introduce a lattice approximation of the meson model above by discretizing space, time, momentum, and energy. To this end, we choose real numbers $\Delta x > 0$, $\Delta p > 0$ and a natural number $n_{\max}$. We will introduce the notation in such a way that the limit

$$\Delta x \to 0, \quad \Delta p \to 0, \quad n_{\max} \to +\infty$$

yields the continuum model. To this end, we choose a right-handed orthonormal system $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. For integers $n^\mu = 0, \pm 1, \ldots, \pm n_{\max}$ with $\mu = 0, 1, 2, 3$, we set

$$\mathbf{x} = (n^1\mathbf{e}_1 + n^2\mathbf{e}_2 + n^3\mathbf{e}_3)\Delta x, \qquad ct = n^0\Delta x.$$

This way, we obtain a finite number $x(1), \ldots, x(N)$ of space-time points living in $\mathbb{R}^4$. To simplify notation, we write

$$\sum_{x \in \mathcal{M}} f(x)\Delta^4 x := \sum_{k=1}^{N} f(k)\Delta^4 x,$$

that is, we identify $x(1), \ldots, x(N)$ with $1, \ldots, N$, respectively. Recall that $\Delta^4 x := (\Delta x)^4$. Similarly, we set

$$\mathbf{p} = (n^1\mathbf{e}_1 + n^2\mathbf{e}_2 + n^3\mathbf{e}_3)\Delta p, \qquad E = n^0c\Delta p.$$

This yields the finite number $p(1), \ldots, p(N)$ of 4-momentum vectors. We write $\Delta^4 p := (\Delta p)^4$, and

$$\sum_{p\in\mathcal{M}} \hat{f}(p)\Delta^4 p = \sum_{k=1}^{N} \hat{f}(k)\Delta^4 p.$$

In what follows, we will use the matrices

$$\varphi = \begin{pmatrix} \varphi(1) \\ \vdots \\ \varphi(N) \end{pmatrix}, \qquad \hat{\varphi} = \begin{pmatrix} \hat{\varphi}(1) \\ \vdots \\ \hat{\varphi}(N) \end{pmatrix}.$$

By a discrete Fourier transform, we understand a linear transformation

$$\boxed{\hat{\varphi} = U\varphi}$$

where U is a fixed complex unitary $(N\times N)$-matrix with the matrix elements $\mathcal{U}(p,x)\Delta^4 x$, that is, $U^\dagger = U^{-1}$. Explicitly,

$$\hat{\varphi}(p) = \sum_{x\in\mathcal{M}} \mathcal{U}(p,x)\varphi(x)\Delta^4 x, \qquad p\in\mathcal{M}.$$

The inverse transformation $\varphi = U^{-1}\hat{\varphi}$ will be written as

$$\varphi(x) = \sum_{p\in\mathcal{M}} \hat{\mathcal{U}}(x,p)\hat{\varphi}(p)\Delta^4 p, \qquad x\in\mathcal{M}.$$

Let $D : L_2(\mathcal{M}) \to L_2(\mathcal{M})$ be a linear operator. The equation

$$D\varphi = \psi$$

is transformed into $UDU^{-1}(U\varphi) = U\psi$. Hence

$$\hat{D}\hat{\varphi} = \hat{\psi}$$

where $\hat{D} := UDU^{-1}$ represents the Fourier transform of the operator D.

The discrete φ^4-model. As a discretization of the continuum φ^4-model above, we use the discrete model from Sect. 7.24.3 on page 445 along with the following special choices.

(i) Interaction: We set $\mathcal{L}_{\rm int}(\chi) := -\chi^4$. Then, $\mathcal{L}'_{\rm int}(\chi) = -4\chi^3$.

(ii) The linear operator D: Let $\varepsilon > 0$. We define $D + \mathrm{i}\varepsilon I := U^{-1}(\hat{D} + \mathrm{i}\varepsilon I)U$ along with the Fourier transform

$$\{(\hat{D} + \mathrm{i}\varepsilon I)\hat{\varphi}\}(k) := (p(k)^2 - m_0^2c^2 + \mathrm{i}\varepsilon)\hat{\varphi}(k), \qquad k = 1,\dots,N.$$

This definition is motivated by formula (7.146) on page 461.

For the response operator $R_\varepsilon := -(D + \mathrm{i}\varepsilon I)^{-1}$, we get $R_\varepsilon = U^{-1}\hat{R}_\varepsilon U$ along with

$$\hat{R}_\varepsilon\hat{\varphi}(k) = -\frac{\hat{\varphi}(k)}{p(k)^2 - m_0^2c^2 + \mathrm{i}\varepsilon}, \qquad k = 1, \ldots, N.$$

The linear response equation

$$(D + \mathrm{i}\varepsilon I)\varphi = -J$$

has the unique solution $\varphi = R_\varepsilon J$. Let $\mathcal{R}_\varepsilon(x, y)\Delta^4 y$ be the entries of the matrix R_ε where $\Delta^4 y = \Delta^4 x := (\Delta x)^4$. Then

$$\varphi(x) = \sum_{y\in\mathcal{M}} \mathcal{R}_\varepsilon(x, y)J(y)\Delta^4 y, \qquad x \in \mathcal{M}.$$

The function $(x, y) \mapsto \mathcal{R}_\varepsilon(x, y)$ is called the response function of the discrete φ^4-model.

For the discrete φ^4-model let us now compute the free correlation functions, the full correlation functions, the mean field, and the vertex functions in low orders of perturbation theory with respect to the coupling constant, κ. By the magic formula (7.127) on page 448, the quantum action functional is given by

$$\boxed{Z(J) = \mathcal{N}\exp\left\{-\mathrm{i}\kappa\hbar^3 \sum_{z\in\mathcal{M}} \Delta^4 z\, \frac{\delta^4}{\delta J(z)^4}\right\} \mathrm{e}^{Z_{\mathrm{free,red}}(J)}} \tag{7.149}$$

with the reduced free quantum action functional

$$Z_{\mathrm{free,red}} := \frac{\mathrm{i}}{2\hbar}\,\langle J|R_\varepsilon J\rangle = \frac{\mathrm{i}}{2\hbar} \sum_{x,y\in\mathcal{M}} J(x)\mathcal{R}_\varepsilon(x, y)J(y)\ \Delta^4 x\Delta^4 y.$$

The normalization constant $\mathcal{N}$ has to be chosen in such a way that $Z(0) = 1$. Up to terms of order κ^2 with respect to the small coupling constant κ, we get

$$Z(J) = \mathcal{N}\left\{1 - \mathrm{i}\kappa\hbar^3 \sum_{z\in\mathcal{M}} \Delta^4 z\, \frac{\delta^4}{\delta J(z)^4}\right\} \mathrm{e}^{Z_{\mathrm{free,red}}(J)}.$$

Free correlation functions. First consider the case where the interaction is switched off by setting $\kappa = 0$.

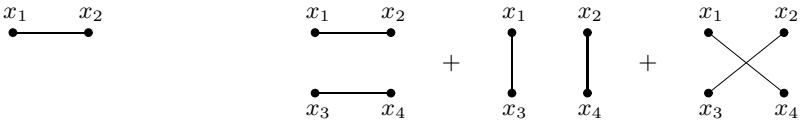
Proposition 7.43 *For the free correlation functions, we get the following expressions.*

(i) $C_{n,\mathrm{free}}(x_1, \ldots, x_n) \equiv 0$ *if* $n = 1, 3, 5, \ldots$.

(ii) *Free* 2*-point correlation function:* $C_{2,\mathrm{free}}(x_1, x_2) = -\mathrm{i}\hbar\mathcal{R}_\varepsilon(x_1, x_2)$.

(iii) *Free* 4*-point correlation function:*

$$C_{4,\mathrm{free}}(x_1, x_2, x_3, x_4) = C_{2,\mathrm{free}}(x_1, x_2)C_{2,\mathrm{free}}(x_3, x_4) +$$
$$+C_{2,\mathrm{free}}(x_1, x_3)C_{2,\mathrm{free}}(x_2, x_4) + C_{2,\mathrm{free}}(x_1, x_4)C_{2,\mathrm{free}}(x_2, x_3).$$



(a) $C_{2,\mathrm{free}}(x_1, x_2)$ (b) $C_{4,\mathrm{free}}(x_1, x_2, x_3, x_4)$

Fig. 7.5. Feynman diagrams for free correlation functions

Feynman diagrams. The functions $C_{2,\mathrm{free}}$ and $C_{4,\mathrm{free}}$ can be represented graphically by Feynman diagrams as pictured in Figure 7.5. The free 2-point correlation function $C_{2,\mathrm{free}}$ is also called the Feynman propagator of the discrete φ^4-model. Note that a complete list of the Feynman rules for the discrete φ^4-model can be found in Table 7.1 on page 469.

Proof of Proposition 7.43. Recall that

$$C_{n,\mathrm{free}}(x_1, \dots, x_n) := \left(\frac{\hbar}{\mathrm{i}}\right)^n \frac{\delta^n Z_{\mathrm{free}}(J)}{\delta J(x_1) \cdots \delta J(x_n)}\Big|_{J=0}. \tag{7.150}$$

Ad (i). Since $Z_{\mathrm{free,red}}$ is quadratic with respect to J, the functional

$$Z_{\mathrm{free}}(J) = \mathrm{e}^{Z_{\mathrm{free,red}}(J)}$$

contains only terms of even degree with respect to J.

Ad (ii). Power series expansion yields

$$\begin{aligned} Z_{\mathrm{free}}(J) &= 1 + \frac{\mathrm{i}}{2\hbar}\langle J | R_\varepsilon J\rangle + \dots \\ &= 1 + \frac{\mathrm{i}}{2\hbar} \sum_{x,y \in \mathcal{M}} J(x) \mathcal{R}_\varepsilon(x,y) J(y) \Delta^4 x \Delta^4 y + \dots \end{aligned}$$

The dots stand for terms of order $4, 6, 8, \dots$ with respect to J. Hence

$$\frac{\delta^2 Z_{\mathrm{free}}(J)}{\delta J(x_1) \delta J(x_2)} = \frac{\mathrm{i}}{\hbar}\, \mathcal{R}_\varepsilon(x,y).$$

By (7.150), $C_{2,\mathrm{free}}(x,y) = -\mathrm{i}\hbar \mathcal{R}_\varepsilon(x,y)$.

Ad (iii). Note that $Z_{\mathrm{free}}(J) = \mathrm{e}^{Z_{\mathrm{free,red}}}$. Hence

$$Z_{\mathrm{free}}(J) = 1 + Z_{\mathrm{free,red}}(J) + \frac{1}{2} Z_{\mathrm{free,red}}(J)^2 + \dots$$

The function $C_{4,\mathrm{free}}$ is determined by the term $\frac{1}{2} Z_{\mathrm{free,red}}(J)^2$ of order four with respect to J. Explicitly,

$$\frac{1}{4!}\left(\frac{\mathrm{i}}{\hbar}\right)^4 \sum_{x_1,x_2,x_3,x_4 \in \mathcal{M}} C_{\mathrm{free},4}(x_1,x_2,x_3,x_4) \prod_{k=1}^4 J(x_k) \Delta^4 x_k = \frac{1}{2} Z_{\mathrm{free,red}}(J)^2.$$

Here, $C_{4,\mathrm{free}}$ is symmetric with respect to x_1, x_2, x_3, x_4. By comparison, we get the claim. In fact, note that

$$Z_{\mathrm{free,red}}(J) = -\frac{1}{2\hbar^2}\langle J|C_{\mathrm{free},2}J\rangle.$$

This implies

$$\frac{1}{2}Z_{\mathrm{free,red}}(J)^2 = \frac{3}{4!\hbar^4}\,\langle J|C_{\mathrm{free},2}J\rangle^2.$$

Moreover,

$$\begin{aligned}\langle J|C_{\mathrm{free},2}J\rangle^2 = &\sum_{x_1,x_2\in\mathcal{M}} \Delta^4 x_1 \Delta^4 x_2 J(x_1) C_{2,\mathrm{free}}(x_1,x_2) J(x_2)\\ &\times \sum_{x_3,x_4\in\mathcal{M}} \Delta^4 x_3 \Delta^4 x_4 J(x_3) C_{2,\mathrm{free}}(x_3,x_4) J(x_4).\end{aligned}$$

Hence

$$\langle J|C_{2,\mathrm{free}}J\rangle^2 = \sum_{x_1,x_2,x_3,x_4\in\mathcal{M}} f(x_1,x_2,x_3,x_4)\prod_{k=1}^{4} J(x_k)\Delta^4 x_k$$

where we set

$$f(x_1,x_2,x_3,x_4) := C_{2,\mathrm{free}}(x_1,x_2)C_{2,\mathrm{free}}(x_3,x_4).$$

The function f is not symmetric with respect to the variables x_1, x_2, x_3, x_4. Therefore, we have to symmetrize. Since $C_{2,\mathrm{free}}(x,y)$ is symmetric with respect to x, y, we get

$$3\langle J|C_{2,\mathrm{free}}J\rangle = \sum_{x_1,x_2,x_3,x_4\in\mathcal{M}} \mathcal{C}(x_1,x_2,x_3,x_4)\prod_{k=1}^{4} J(x_k)\Delta^4 x_k$$

where $\mathcal{C}(x_1,x_2,x_3,x_4)$ is equal to

$$f(x_1,x_2,x_3,x_4) + f(x_1,x_3,x_2,x_4) + f(x_1,x_4,x_2,x_3).$$

This finishes the proof of Proposition 7.43. □

Generally, for even n, the free correlation function $C_{n,\mathrm{free}}$ is given by

$$C_{n,\mathrm{free}}(x_1,\ldots,x_n) = \sum_{\pi} C_{2,\mathrm{free}}(x_{i_1},x_{i_2})\cdots C_{2,\mathrm{free}}(x_{i_{n-1}},x_{i_n}).$$

Here, we sum over all permutations of $x_1,\ldots,x_n$ in such a restricted way that the pairs are always ordered, that is $i_l < i_{l+1}$ for all l. The corresponding Feynman diagrams are obtained by drawing n nodes and by connecting all possible pairs. Since the node x_1 can be connected with the other $n-1$ nodes, it follows by induction that there are $1\times 3\times\cdots\times(n-1)$ Feynman diagrams

for $n = 2, 4, \ldots$. The Feynman graph to $C_{2,\text{free}}$ is connected, whereas the graphs to $C_{n,\text{free}}$ with $n = 4, 6, 8, \ldots$ are not always connected.

Summary of properties of the free field. For the convenience of the reader, let us summarize typical quantities of the free field. Below we will generalize this to the case of interacting fields with $\kappa > 0$.

- Free reduced 2-point correlation function:

$$\boxed{C_{2,\text{free,red}}(x_1, x_2) = C_{2,\text{free}}(x_1, x_2) = -\mathrm{i}\hbar\mathcal{R}_\varepsilon(x_1, x_2).}$$

- Free reduced 4-point correlation function:

$$C_{4,\text{free,red}}(x_1, x_2, x_3, x_4) = C_{2,\text{free}}(x_1, x_4)C_{2,\text{free}}(x_2, x_3).$$

  This corresponds to the connected Feynman graph pictured in Fig. 7.5(b) on page 465.[49]
- Free 2-point vertex function $V_{2,\text{free}}$:

$$\sum_{z\in\mathcal{M}} V_{2,\text{free}}(x, z)C_{2,\text{free}}(z, y)\Delta^4 z = \mathrm{i}\hbar\delta(x, y).$$

- Free 4-point vertex function:

$$V_{4,\text{free}}(x_1, x_2, x_3, x_4) = \frac{\mathrm{i}}{\hbar} \cdot V_{2,\text{free}}(x_1, x_4)V_{2,\text{free}}(x_2, x_3).$$

- Mean field transformation (Legendre transformation):

$$\boxed{\varphi_{\text{mean,free}}(x) = \frac{\mathrm{i}}{\hbar} \sum_{y\in\mathcal{M}} C_{2,\text{free}}(x, y)J(y)\Delta^4 y.}$$

  The free mean field $\varphi_{\text{mean,free}}$ coincides with the free field caused by the external source J, that is, $(D + \mathrm{i}\varepsilon I)\varphi_{\text{mean,free}} = -J$.
- Inverse mean field transformation:

$$\boxed{J(x) = -\sum_{y\in\mathcal{M}} V_{2,\text{free}}(x, y)\varphi_{\text{mean,free}}(y)\ \Delta^4 y.}$$

  Therefore, the inverse kernel to $\mathrm{i}C_{2,\text{free}}(x, y)/\hbar$ is equal to $-V_{2,\text{free}}(x, y)$.
- The effective free action $V(\varphi_{\text{mean,free}})$ is equal to the free action

$$S_{\text{free}}[\varphi] := \sum_{x\in\mathcal{M}} \frac{1}{2}\varphi^d(D + \mathrm{i}\varepsilon I)\varphi\ \Delta^4 x \tag{7.151}$$

  where we set $\varphi := \varphi_{\text{mean,free}}$.

[49] Reduced correlation functions are also called connected correlation functions. It can be shown that the corresponding Feynman graphs are always connected.

The formulas above show that the vertex functions are dual to the correlation functions. If we switch on the interaction such that the coupling constant $\kappa > 0$ is small, then the free functions summarized above are perturbed slightly. Let us compute this in first order with respect to κ, that is, we neglect terms of order κ^2.

Full correlation functions. The free correlation functions $C_{n,\mathrm{free}}$ refer to vanishing coupling constant, $\kappa = 0$. In contrast to this, the full correlation functions C_n describe the behavior of the meson under self-interactions corresponding to the coupling constant $\kappa > 0$. The full n-point correlation function C_n is defined by

$$C_n(x_1, \dots, x_n) := \left(\frac{\hbar}{\mathrm{i}}\right)^n \frac{\delta^n Z(J)}{\delta J(x_1) \cdots \delta J(x_n)} \Big|_{J=0}.$$

Theorem 7.44 *Up to terms of order κ^2, the full correlation functions of the discrete φ^4-model read as follows for all $x_1, x_2, \dots \in \mathcal{M}$:*

(i) $C_n(x_1, \dots, x_n) \equiv 0$ *if* $n = 1, 3, 5, \dots$

(ii) *Full 2-point correlation function:*

$$C_2(x_1, x_2) = C_{2,\mathrm{free}}(x_1, x_2) + 12\kappa D(x_1, x_2) \tag{7.152}$$

along with

$$D(x_1, x_2) := -\frac{\mathrm{i}}{\hbar} \sum_{z \in \mathcal{M}} C_{2,\mathrm{free}}(x_1, z) C_{2,\mathrm{free}}(z, z) C_{2,\mathrm{free}}(z, x_2)\, \Delta^4 z.$$

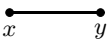
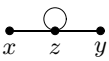
(iii) *Full 4-point correlation function:*

$$\begin{aligned} C_4(x_1, x_2, x_3, x_4) = {} & C_{4,\mathrm{free}}(x_1, x_2, x_3, x_4) \\ & + 12\kappa A(x_1, x_2, x_3, x_4) + 24\kappa B(x_1, x_2, x_3, x_4). \end{aligned}$$

Here, $A(x_1, x_2, x_3, x_4)$ is equal to

$$\begin{aligned} & -\frac{\mathrm{i}}{\hbar} \sum_{z \in \mathcal{M}} C_{2,\mathrm{free}}(x_1, z) C_{2,\mathrm{free}}(z, z) C_{2,\mathrm{free}}(z, x_2) C_{2,\mathrm{free}}(x_3, x_4) \Delta^4 z \\ & -\frac{\mathrm{i}}{\hbar} \sum_{z \in \mathcal{M}} C_{2,\mathrm{free}}(x_1, x_2) C_{2,\mathrm{free}}(x_3, z) C_{2,\mathrm{free}}(z, z) C_{2,\mathrm{free}}(z, x_4) \Delta^4 z \\ & -\frac{\mathrm{i}}{\hbar} \sum_{z \in \mathcal{M}} C_{2,\mathrm{free}}(x_1, z) C_{2,\mathrm{free}}(z, z) C_{2,\mathrm{free}}(z, x_3) C_{2,\mathrm{free}}(x_2, x_4) \Delta^4 z \\ & -\frac{\mathrm{i}}{\hbar} \sum_{z \in \mathcal{M}} C_{2,\mathrm{free}}(x_1, x_3) C_{2,\mathrm{free}}(x_2, z) C_{2,\mathrm{free}}(z, z) C_{2,\mathrm{free}}(z, x_4) \Delta^4 z \\ & -\frac{\mathrm{i}}{\hbar} \sum_{z \in \mathcal{M}} C_{2,\mathrm{free}}(x_1, z) C_{2,\mathrm{free}}(z, z) C_{2,\mathrm{free}}(z, x_4) C_{2,\mathrm{free}}(x_2, x_3) \Delta^4 z \\ & -\frac{\mathrm{i}}{\hbar} \sum_{z \in \mathcal{M}} C_{2,\mathrm{free}}(x_1, x_4) C_{2,\mathrm{free}}(x_2, z) C_{2,\mathrm{free}}(z, z) C_{2,\mathrm{free}}(z, x_3) \Delta^4 z. \end{aligned}$$

Table 7.1. Feynman diagrams for the discrete φ^4-model

propagator		$C_{2,\mathrm{free}}(x,y)$
loop		$C_{2,\mathrm{free}}(z,z)$
vertex of interaction	 	$-\frac{\mathrm{i}}{\hbar}\sum_{z\in\mathcal{M}}\Delta^4 z\ \ldots$
symmetry factor		s

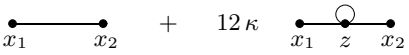


Fig. 7.6. First-order approximation of the full correlation function $C_2(x_1, x_2)$

Furthermore, $B(x_1, x_2, x_3, x_4)$ is equal to

$$-\frac{\mathrm{i}}{\hbar}\sum_{z\in\mathcal{M}}\{C_{2,\mathrm{free}}(x_1,z)C_{2,\mathrm{free}}(z,z)C_{2,\mathrm{free}}(z,x_4) + C_{2,\mathrm{free}}(x_2,z)C_{2,\mathrm{free}}(z,z)C_{2,\mathrm{free}}(z,x_3)\}\Delta^4 z.$$

Feynman diagrams. Using Table 7.1, the first-order approximation of the full 2-point correlation function C_2 from Theorem 7.44 is pictured in Fig. 7.6. Here, the symmetry factor is given by $s = 12$. Moreover, the self-energy term A and the vertex-interaction term B of the full 4-point correlation function C_4 from Theorem 7.44 are pictured in Fig. 7.7.

Proof of Theorem 7.44. To simplify notation, we set $\hbar := 1$.

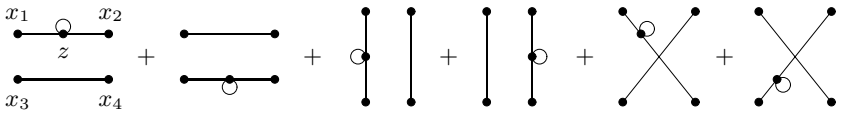
Ad (i). Observe that the functional $Z(J)$ only contains terms of even order with respect to the source J.

Ad (ii). We will proceed in several steps.

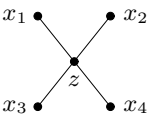
(I) By the magic quantum action formula,

$$Z(J) = \frac{E(J)}{E(0)}.$$

Here, we set



(a) self-energy graphs ($s = 4!/2$)



(b) Vertex interaction ($s = 4!$)

Fig. 7.7. First-order correction for the full correlation function $C_4(x_1, x_2, x_3, x_4)$

$$E(J) := \exp\left\{-\mathrm{i}\kappa \sum_{z\in\mathcal{M}} \Delta^4 z \, \frac{\delta^4}{\delta J(z)^4}\right\} \cdot \mathrm{e}^{B(J)}$$

along with

$$B(J) := -\frac{1}{2}\langle J|C_{2,\mathrm{free}}|J\rangle = -\frac{1}{2}\sum_{x,y\in\mathcal{M}} J(x)C_{2,\mathrm{free}}(x,y)J(y)\ \Delta^4 x \Delta^4 y.$$

(II) Up to terms of order κ^2,

$$E(J) = \left\{1 - \mathrm{i}\kappa \sum_{z\in\mathcal{M}} \Delta^4 z \, \frac{\delta^4}{\delta J(z)^4}\right\} \mathrm{e}^{B(J)}. \tag{7.153}$$

Introducing

$$A(J) := \mathrm{e}^{-B(J)} \left\{-\mathrm{i} \sum_{z\in\mathcal{M}} \Delta^4 z \, \frac{\delta^4}{\delta J(z)^4}\right\} \mathrm{e}^{B(J)},$$

we get $E(J) = \mathrm{e}^{B(J)}(1 + \kappa A(J))$. Up to terms of order κ^2, this implies

$$Z(J) = \frac{\mathrm{e}^{B(J)}(1 + \kappa A(J))}{1 + \kappa A(0)} = \mathrm{e}^{B(J)}\{1 + \kappa(A(J) - A(0))\}. \tag{7.154}$$

(III) Computation of the functional derivatives. To simplify notation, set

$$b := \frac{\delta B(J)}{\delta J(z)}, \qquad c := \frac{\delta^2 B(J)}{\delta J(z)^2}.$$

Because of the symmetry property $C_{2,\mathrm{free}}(x, y) = C_{2,\mathrm{free}}(y, x)$, we get

$$b = -\sum_{y\in\mathcal{M}} C_{2,\mathrm{free}}(z,y)J(y)\Delta^4 y, \qquad c = -C_{2,\mathrm{free}}(z,z).$$

(IV) Using the chain rule, we obtain

$$\begin{aligned}
\frac{\delta \mathrm{e}^{B(J)}}{\delta J(z)} &= b\mathrm{e}^{B(J)}, \qquad \frac{\delta^2 \mathrm{e}^{B(J)}}{\delta J(z)^2} = c\mathrm{e}^B + b^2\mathrm{e}^B,\\
\frac{\delta^3 \mathrm{e}^{B(J)}}{\delta J(z)^3} &= cb\mathrm{e}^B + 2bc\mathrm{e}^B + b^3\mathrm{e}^B = (3bc+b^3)\,\mathrm{e}^B,\\
\frac{\delta^4 \mathrm{e}^{B(J)}}{\delta J(z)^4} &= (3c^2 + 3b^2c + 3b^2c + b^4)\mathrm{e}^B.
\end{aligned}$$

This implies

$$A(J) = -\mathrm{i}\sum_{z\in\mathcal{M}} (3c^2 + 6b^2c + b^4)\ \Delta^4 z. \tag{7.155}$$

Since the function c does not depend on J,

$$A(J) - A(0) = -\mathrm{i}\sum_{z\in\mathcal{M}} (6b^2c + b^4)\ \Delta^4 z. \tag{7.156}$$

Here,

$$\sum_{z\in\mathcal{M}} b^2 c\ \Delta^4 z = \sum_{x,y\in\mathcal{M}} J(x)H(x,y)J(y)\Delta^4 x\Delta^4 y$$

along with $H(x,y) := \sum_{z\in\mathcal{M}} C_{2,\mathrm{free}}(x,z)C_{2,\mathrm{free}}(z,z)C_{2,\mathrm{free}}(z,y)\Delta^4 z$.

(V) For the correlation functions,

$$Z(J) = 1 - \frac{1}{2}\sum_{x,y\in\mathcal{M}} C_2(x,y)J(x)J(y)\Delta^4 x\Delta^4 y + \dots.$$

Since $\mathrm{e}^B = 1 + B + \dots$, it follows from (7.154) that the quadratic term of $Z(J)$ with respect to J is equal to

$$B(J) - \mathrm{i}\kappa\sum_{z\in\mathcal{M}} 6b^2c\ \Delta^4 z.$$

Multiplying this by -2, we get the desired result

$$C_2(x,y) = C_{2,\mathrm{free}}(x,y) - 12\mathrm{i}\kappa H(x,y).$$

Ad (iii). Use analogous arguments. See Problem 7.5 on page 495. □

Vacuum bubble. By (7.155), the functional A contains the expression

$$-3\mathrm{i}\sum_{z\in\mathcal{M}} c^2 = -3\mathrm{i}\sum_{z\in\mathcal{M}} C_{2,\mathrm{free}}(z,z)^2$$

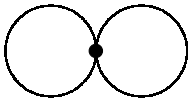


Fig. 7.8. Vacuum bubble

which is pictured graphically in Fig. 7.8, up to the symmetry factor 3. Because of the normalization condition $Z(0) = 1$, the difference $A(J) - A(0)$ appears instead of $A(J)$ in (7.156). This cancels the vacuum bubble. Similar cancellations of vacuum effects appear in higher order of perturbation theory.

The reduced correlation functions. The reduced quantum action functional Z_{red} is defined by $Z_{\mathrm{red}}(J) := \ln Z(J)$. Hence

$$\boxed{Z(J) = \mathrm{e}^{Z_{\mathrm{red}}(J)}.}$$

The reduced n-point correlation function is given by

$$C_{n,\mathrm{red}}(x_1, \ldots, x_n) := \left(\frac{\hbar}{\mathrm{i}}\right)^n \frac{\delta^n \ln Z(J)}{\delta J(x_1) \cdots \delta J(x_n)}.$$

Proposition 7.45 *There exist the following relations between full correlation functions and reduced correlation functions for all* $x_1, x_2, \ldots \in \mathcal{M}$:

(i) $C_{n,\mathrm{red}}(x_1, \ldots, x_n) \equiv 0$ *if* $n = 1, 3, 5, \ldots$

(ii) *Reduced 2-point correlation function:* $C_{2,\mathrm{red}}(x_1, x_2) = C_2(x_1, x_2)$.

(iii) *Reduced 4-point correlation function:*

$$\begin{aligned} C_{4,\mathrm{red}}(x_1, x_2, x_3, x_4) = {} & C_4(x_1, x_2, x_3, x_4) - C_2(x_1, x_2) C_2(x_3, x_4) \\ & - C_2(x_1, x_3) C_2(x_2, x_4) - C_2(x_1, x_4) C_2(x_2, x_3). \end{aligned}$$

Up to terms of order κ^2*, the reduced 4-point function is given by the Feynman diagram from Fig.* 7.9*. Explicitly,*

$$C_{4,\mathrm{red}}(x_1, x_2, x_3, x_4) = C_{2,\mathrm{free}}(x_1, x_4) C_{2,\mathrm{free}}(x_2, x_4) + 24\kappa B(x_1, x_2, x_3, x_4).$$

The definition of $B(x_1, x_2, x_3, x_4)$ can be found in Theorem 7.44 on page 468. The Feynman graphs from Fig. 7.9 are precisely the connected graphs from Fig. 7.5(b) and Fig. 7.7 on pages 465 and 470, respectively.

Proof. Ad (i). Since $Z(J)$ does not contain terms of odd order with respect to J, the function $\ln Z(J)$ has the same property.

Ad (ii). Note that

$$\begin{aligned} \frac{\delta^2 \ln Z(J)}{\delta J(x) \delta J(y)} &= \frac{\delta}{\delta J(x)} \left(\frac{1}{Z(J)} \frac{\delta Z(J)}{\delta J(y)} \right) \\ &= -\frac{1}{Z(J)^2} \frac{\delta Z(J)}{\delta J(x)} \frac{\delta Z(J)}{\delta J(y)} + \frac{1}{Z(J)} \frac{\delta^2 Z(J)}{\delta J(x) \delta J(y)}. \end{aligned}$$

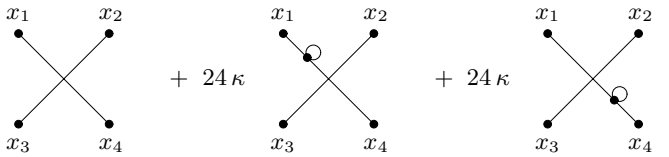


Fig. 7.9. First-order approximation of $C_{4,\mathrm{red}}$

Setting $J = 0$ and observing that $Z(0) = 1$, we get

$$C_{2,\mathrm{red}}(x_1, x_2) = C_1(x)C_1(y) + C_2(x, y) = C_2(x, y).$$

Ad (iii). Compute higher derivatives as in the proof to (ii).

Ad (iv). Use the correlation function C_2 from (7.152) on page 468. By (ii), the terms corresponding to disconnected graphs are cancelled. □

Vertex functions. The vertex functions are dual objects to the reduced correlation functions. By the chain rule, there follows from (7.140) and (7.142) on pages 457 and 459, respectively, that each correlation function can be represented by vertex functions and 2-point correlation functions. In this sense, vertex functions and 2-point correlation functions are the building blocks of general correlation functions (see (7.158) on page 474). In fact, vertex functions were introduced by physicists in order to simplify lengthy computations in higher orders of perturbation theory.

Proposition 7.46 *For all $x_1, x_2, x_3, x_4 \in \mathcal{M}$, the following hold true.*

(i) $V_n(x_1, \ldots, n) \equiv 0$ *if* $n = 1, 3, 5, \ldots$.

(ii) *The 2-point vertex function V_2 satisfies the equation*

$$\sum_{z \in \mathcal{M}} C_{2,\mathrm{red}}(x_1, z)V_2(z, x_2)\Delta^4 z = \mathrm{i}\hbar\delta_{\Delta^4 x}(x_1, x_2).$$

(iii) *The 4-point vertex function $V_4(x_1, x_2, x_3, x_4)$ is equal to*

$$\frac{1}{\mathrm{i}\hbar^3} \sum_{z_1, z_2, z_3, z_4 \in \mathcal{M}} C_{4,\mathrm{red}}(z_1, z_2, z_3, z_4) \prod_{k=1}^{4} V_2(x_k, z_k)\Delta^4 z_k.$$

Before giving the proof, let us discuss some special cases in order to get some feeling for vertex functions. In what follows we refer to the statements (i) through (iii) from Proposition 7.46 above.

- Inverse response kernel: First switch off the interaction by setting $\kappa = 0$. By statement (i) above, it follows from the response equation

$$\varphi_{\mathrm{free}}(x) = \frac{\mathrm{i}}{\hbar} \sum_{y \in \mathcal{M}} C_{2,\mathrm{free}}(x, y)J(y)\ \Delta^4 y, \qquad x \in \mathcal{M}$$

with the response kernel $\mathcal{R}_\varepsilon = \mathrm{i}C_{2,\mathrm{free}}/\hbar$ that

$$J(x) = -\sum_{y\in\mathcal{M}} V_{2,\mathrm{free}}(x,y)\varphi(y)\ \Delta^4 y.$$

- Free 4-point vertex function: By (iii) above, it follows from the reduced free 4-point function

$$C_{4,\mathrm{red,free}}(x_1,x_2,x_3,x_4) = C_{2,\mathrm{free}}(x_1,x_4)C_{2,\mathrm{free}}(x_2,x_3)$$

that

$$V_{4,\mathrm{free}}(x_1,x_2,x_3,x_4) = \frac{\mathrm{i}}{\hbar}\cdot V_{2,\mathrm{free}}(x_1,x_4)V_{2,\mathrm{free}}(x_2,x_3).$$

- Full 2-point vertex function: By (i) above, the equation

$$\varphi(x) = \frac{\mathrm{i}}{\hbar}\sum_{y\in\mathcal{M}} C_2(x,y)J(y)\ \Delta^4 y$$

implies that

$$J(x) = -\sum_{y\in\mathcal{M}} V_2(x,y)\varphi(y)\ \Delta^4 y. \tag{7.157}$$

- Representation of the full reduced 4-point correlation function by vertex functions: It follows from (i) above that equation (iii) above is equivalent to

$$C_{4,\mathrm{red}}(x_1,x_2,x_3,x_4) = \frac{\mathrm{i}}{\hbar}\sum_{z_1,z_2,z_3,z_4\in\mathcal{M}} V_4(z_1,z_2,z_3,z_4) \times\prod_{k=1}^{4} C_2(x_k,z_k)\Delta^4 z_k. \tag{7.158}$$

- First order approximation of the full 4-point vertex function: Up to terms of order κ^2,

$$C_{4,\mathrm{red}}(x_1,x_2,x_3,x_4) = C_{4,\mathrm{red,free}}(x_1,x_2,x_3,x_4) + 24\kappa B(x_1,x_2,x_3,x_4)$$

where $B(x_1,x_2,x_3,x_4)$ is equal to

$$-\frac{\mathrm{i}}{\hbar}\sum_{z\in\mathcal{M}}\{C_{2,\mathrm{free}}(x_1,z)C_{2,\mathrm{free}}(z,z)C_{2,\mathrm{free}}(z,x_4) +C_{2,\mathrm{free}}(x_2,z)C_{2,\mathrm{free}}(z,z)C_{2,\mathrm{free}}(z,x_3)\}\Delta^4 z.$$

By (7.158) along with (i) above,

$$V_4(x_1,x_2,x_3,x_4) = V_{4,\mathrm{free}}(x_1,x_2,x_3,x_4) + \kappa\mathcal{B}(x_1,x_2,x_3,x_4)$$

where $\mathcal{B}(x_1,x_2,x_3,x_4)$ is equal to

$$\sum_{z\in\mathcal{M}} \{V_{2,\mathrm{free}}(x_1,z)C_{2,\mathrm{free}}(z,z)V_{2,\mathrm{free}}(z,x_4) + V_{2,\mathrm{free}}(x_2,z)C_{2,\mathrm{free}}(z,z)V_{2,\mathrm{free}}(z,x_3)\}\Delta^4 z.$$

Feynman diagrams. The formula for $V_4(x_1,x_2,x_3,x_4)$ above displays a nice duality between the reduced correlation function $C_{4,\mathrm{red}}$ and the vertex function V_4. In the language of Feynman graphs for correlation functions (see Fig. 7.9 on page 473), one has only to replace

- the correlation function $C_{2,\mathrm{free}}$ corresponding to the external lines
- by the vertex function $V_{2,\mathrm{free}}$, up to an additional factor.

Proof of Proposition 7.46. If $\kappa = 0$, then the map $J \mapsto \varphi_{\mathrm{mean}}$ is invertible. By perturbation theory, this invertibility property remains true if κ is sufficiently small. This tells us that the Legendre transformation

$$\varphi_{\mathrm{mean}}(x) = \frac{\hbar}{\mathrm{i}}\frac{\delta Z_{\mathrm{red}}(J)}{\delta J(x)}, \qquad J(x) = -\frac{\delta V(\varphi_{\mathrm{mean}})}{\delta\varphi_{\mathrm{mean}}(x)} \tag{7.159}$$

is well-defined for each small coupling constant κ. Recall that

$$V_n(x_1,\ldots,x_n) := \frac{\delta^n V(\varphi_{\mathrm{mean}})}{\delta\varphi_{\mathrm{mean}}(x_1)\cdots\delta\varphi_{\mathrm{mean}}(x_n)}.$$

To simplify notation, replace φ_{mean} by φ, and set $\hbar := 1$.

Ad (i). Note that the map $\varphi \mapsto V(\varphi)$ contains only terms which are even with respect to φ.

Ad (ii). We first prove the following key identity:

$$\sum_{z\in\mathcal{M}} \Delta^4 z\, \frac{\delta^2 Z_{\mathrm{red}}(J)}{\delta J(x)\delta J(z)}\, \frac{\delta^2 V(\varphi)}{\delta\varphi(z)\delta\varphi(y)} = -\mathrm{i}\delta_{\Delta^4 x}(x,y). \tag{7.160}$$

In fact, it follows from the Legendre transformation (7.159) that the left-hand side of (7.160) is equal to

$$-\mathrm{i}\sum_{z\in\mathcal{M}} \Delta^4 z\, \frac{\delta\varphi(x)}{\delta J(z)}\frac{\delta J(z)}{\delta\varphi(y)} = -\mathrm{i}\frac{\delta\varphi(x)}{\delta\varphi(y)} = -\mathrm{i}\delta_{\Delta^4 x}(x,y).$$

Setting $J = 0$, equation (7.160) implies the claim (ii).

Ad (iii). Changing variables and differentiating (7.160) with respect to $J(x_3)$, we get

$$\sum_{z_1\in\mathcal{M}} \Delta^4 z_1\, \frac{\delta^3 Z_{\mathrm{red}}(J)}{\delta J(x_1)\delta J(z_1)\delta J(x_3)} \cdot \frac{\delta^2 V(\varphi)}{\delta\varphi(z_1)\delta\varphi(z_2)}$$
$$= -\sum_{z_1,z_3\in\mathcal{M}} \Delta^4 z_1 \Delta^4 z_3\, \frac{\delta^2 Z_{\mathrm{red}}(J)}{\delta J(x_1)\delta J(z_1)} \cdot \frac{\delta^3 V(\varphi)}{\delta\varphi(z_1)\delta\varphi(z_2)\delta\varphi(z_3)} \cdot \frac{\delta\varphi(z_3)}{\delta J(x_3)}.$$

By the Legendre transformation (7.159), this is equal to

$$\mathrm{i}\sum_{z_1,z_3\in\mathcal{M}} \Delta^4 z_1 \Delta^4 z_3\, \frac{\delta^2 Z_{\mathrm{red}}(J)}{\delta J(x_1)\delta J(z_1)} \cdot \frac{\delta^3 V(\varphi)}{\delta\varphi(z_1)\delta\varphi(z_2)\delta\varphi(z_3)} \cdot \frac{\delta^2 Z_{\mathrm{red}}(J)}{\delta J(z_3)\delta J(x_3)}.$$

Multiplying this by $\delta^2 Z_{\mathrm{red}}/\delta J(x_2)\delta J(z_2)$ and summing over z_2, we get

$$\begin{aligned}
&\sum_{z_1,z_2\in\mathcal{M}} \Delta^4 z_1 \Delta^4 z_2\, \frac{\delta^3 Z_{\mathrm{red}}}{\delta J(x_1)\delta J(z_1)\delta J(x_3)} \cdot \frac{\delta^2 V}{\delta\varphi(z_1)\delta\varphi(z_2)} \cdot \frac{\delta^2 Z_{\mathrm{red}}}{\delta J(x_2)\delta J(z_2)} \\
&\quad = \mathrm{i}\sum_{z_1,z_2,z_3\in\mathcal{M}} \Delta^4 z_1 \Delta^4 z_2 \Delta^4 z_3\, \frac{\delta^2 Z_{\mathrm{red}}}{\delta J(x_1)\delta J(z_1)} \cdot \frac{\delta^3 V}{\delta\varphi(z_1)\delta\varphi(z_2)\delta\varphi(z_3)} \\
&\qquad \times \frac{\delta^2 Z_{\mathrm{red}}}{\delta J(z_3)\delta J(x_3)} \cdot \frac{\delta^2 Z_{\mathrm{red}}}{\delta J(x_2)\delta J(z_2)}.
\end{aligned}$$

Using (7.160), there appears the factor $-\mathrm{i}\delta_{\Delta^4 x}(z_1, x_2)$ after summing over z_2. Then, summation over z_1 yields the key relation

$$\begin{aligned}
\frac{\delta^3 Z_{\mathrm{red}}(J)}{\delta J(x_1)\delta J(x_2)\delta J(x_3)} &= -\sum_{z_1,z_2,z_3\in\mathcal{M}} \Delta^4 z_1 \Delta^4 z_2 \Delta^4 z_3 \frac{\delta^3 V(\varphi)}{\delta\varphi(z_1)\delta\varphi(z_2)\delta\varphi(z_3)} \\
&\quad \times \frac{\delta^2 Z_{\mathrm{red}}(J)}{\delta J(x_1)\delta J(z_1)} \cdot \frac{\delta^2 Z_{\mathrm{red}}(J)}{\delta J(x_2)\delta J(z_2)} \cdot \frac{\delta^2 Z_{\mathrm{red}}(J)}{\delta J(x_3)\delta J(z_3)}.
\end{aligned}$$

Differentiating this with respect to $J(x_4)$ and setting $J = 0$, we obtain the desired relation (7.158). □

The mean field. By definition, the mean field is given by

$$\varphi_{\mathrm{mean}}(x) := \frac{\hbar}{\mathrm{i}} \frac{\delta Z_{\mathrm{red}}(J)}{\delta J(x)}.$$

Suppose that the components of the external source J are sufficiently small. Then, we may use the following approximation

$$\varphi_{\mathrm{mean}}(x) = \frac{\hbar}{\mathrm{i}} \frac{\delta Z_{2,\mathrm{red}}(J)}{\delta J(x)}.$$

By Proposition 7.45 on page 472,

$$Z_{2,\mathrm{red}}(J) = Z_2(J) = \frac{1}{2}\left(\frac{\mathrm{i}}{\hbar}\right)^2 \sum_{x,y\in\mathcal{M}} J(x) C_2(x,y) J(y)\ \Delta^4 x \Delta^4 y.$$

This yields

$$\boxed{\varphi_{\mathrm{mean}}(x) = \frac{\mathrm{i}}{\hbar}\sum_{y\in\mathcal{M}} C_2(x,y) J(y)\ \Delta^4 y.}$$

According to Theorem 7.44 on page 468,

$$C_2(x,y) = C_{2,\mathrm{free}}(x,y) + 12\kappa D(x,y),$$

in first-order approximation of perturbation theory with respect to the small coupling constant κ.

The effective action. By definition, the effective action is given by

$$S_{\mathrm{eff}}[\varphi_{\mathrm{mean}}] := V(\varphi_{\mathrm{mean}}).$$

If the components of the mean field, φ_{mean}, are sufficiently small, we may use the approximation

$$\boxed{S_{\mathrm{eff}}[\varphi] := V_2(\varphi) = \frac{1}{2} \sum_{x,y \in \mathcal{M}} \varphi(x) V_2(x,y) \varphi(y) \Delta^4 x \Delta^4 y.}$$

According to (7.143) on page 459, this differs from the classical free action (7.151) on page 467 by a term of order κ.

7.26 The Extended Response Approach

The following modification of the response model from Sect. 7.24 is the prototype of models used in quantum electrodynamics and in the Standard Model of particle physics (gauge field theory). The basic idea is to replace the action functional from (7.120) on page 445 by the following action functional

$$\boxed{\begin{aligned} S[\psi, \overline{\psi}, J, \overline{J}] := \langle \overline{\psi} | (D + \mathrm{i}\varepsilon I)\psi \rangle + \kappa \langle 1 | \mathcal{L}_{\mathrm{int}}(\psi, \overline{\psi}) \rangle \\ + \langle \overline{J} | \psi \rangle + \langle \overline{\psi} | J \rangle. \end{aligned}} \tag{7.161}$$

In contrast to (7.120), two independent fields $\psi, \overline{\psi}$ and two independent source functions $J, \overline{J}$ appear. Explicitly, the functional $S[\psi, \overline{\psi}, J, \overline{J}]$ is equal to

$$\begin{aligned} &\sum_{x,y \in \mathcal{M}} \overline{\psi}(x) D(x,y) \psi(y) \Delta^4 x \Delta^4 y + \kappa \sum_{x \in \mathcal{M}} \mathcal{L}_{\mathrm{int}}(\psi(x), \overline{\psi}(x)) \Delta^4 x \\ &\quad + \sum_{x \in \mathcal{M}} \{ \overline{J}(x) \psi(x) + \overline{\psi}(x) J(x) + \mathrm{i}\varepsilon \overline{\psi}(x) \psi(x) \} \Delta^4 x. \end{aligned}$$

Here, we introduce the coupling constant $\kappa \geq 0$ and the regularization parameter $\varepsilon > 0$. With a view to later applications, it is convenient to use the column matrices

$$\psi := \begin{pmatrix} \psi(1) \\ \vdots \\ \psi(N) \end{pmatrix}, \qquad J := \begin{pmatrix} J(1) \\ \vdots \\ J(N) \end{pmatrix}$$

along with the row matrices

$$\overline{\psi} := (\overline{\psi}(1), \dots, \overline{\psi}(N)), \qquad \overline{J} := (\overline{J}(1), \dots, \overline{J}(N)).$$

We make the following assumptions:

- $\psi(1), \dots, \psi(N), \overline{\psi}(1), \dots, \overline{\psi}(N)$ denote $2N$ independent real field variables.[50]
- $J(1), \dots J(N), \overline{J}(1), \dots, \overline{J}(N)$ denote $2N$ independent real source variables.
- The $(N \times N)$-matrix D is real.
- There is a number $\varepsilon_0 > 0$ such that the inverse matrix $(D + \mathrm{i}\varepsilon I)^{-1}$ exists for all $\varepsilon \in]0, \varepsilon_0[$.
- The function $\mathcal{L}_{\mathrm{int}} : \mathbb{R}^2 \to \mathbb{R}$ is a real polynomial of the two real variables ζ and $\overline{\zeta}$. For example, $\mathcal{L}_{\mathrm{int}}(\zeta, \overline{\zeta}) := (\zeta\overline{\zeta})^2$.

The proofs of the following statements proceed as in Sect. 7.24 on page 438.

The principle of critical action. We are given the source functions $J, \overline{J} \in L_2(\mathcal{M})$.

Theorem 7.47 *Each solution $\psi, \overline{\psi} \in L_2(\mathcal{M})$ of the variational problem*

$$S[\psi, \overline{\psi}, J, \overline{J}] = \text{critical!} \tag{7.162}$$

satisfies the following Euler-Lagrange equations

$$\frac{\delta S}{\delta \overline{\psi}(x)} = 0, \qquad \frac{\delta S}{\delta \psi(x)} = 0 \qquad \textit{for all} \quad x \in \mathcal{M}. \tag{7.163}$$

Explicitly, the Euler–Lagrange equations read as

$$(D\psi)(x) + \mathrm{i}\varepsilon\psi(x) + \kappa \frac{\partial \mathcal{L}_{\mathrm{int}}}{\partial \overline{\zeta}}(\psi(x), \overline{\psi}(x)) + J(x) = 0$$

and

$$(\overline{\psi}D)(x) + \mathrm{i}\varepsilon\overline{\psi}(x) + \kappa \frac{\partial \mathcal{L}_{\mathrm{int}}}{\partial \zeta}(\psi(x), \overline{\psi}(x)) + \overline{J}(x) = 0.$$

Proof. (I) For fixed $\overline{h} \in L_2(\mathcal{M})$, set

$$\chi(\tau) := S[\psi, \overline{\psi} + \tau\overline{h}, J, \overline{J}], \qquad \tau \in \mathbb{R}.$$

From the variational problem (7.162), we get $\chi'(0) = 0$. This yields

$$\langle \overline{h} | (D + \mathrm{i}\varepsilon I)\psi \rangle + \langle \overline{h} | J \rangle + \kappa \sum_{x \in \mathcal{M}} \frac{\partial \mathcal{L}_{\mathrm{int}}}{\partial \overline{\zeta}}(\psi(x), \overline{\psi}(x))\overline{h}(x) = 0.$$

[50] The case of complex variables will be considered in Sect. 7.27 on page 483.

This can be written as

$$\sum_{x\in\mathcal{M}} \frac{\delta S}{\delta\overline{\psi}(x)}\overline{h}(x) = 0 \qquad \text{for all} \quad \overline{h} \in L_2(\mathcal{M}) \tag{7.164}$$

along with

$$\frac{\delta S}{\delta\overline{\psi}(x)} = \{(D + \mathrm{i}\varepsilon I)\psi\}(x) + J(x) + \kappa\frac{\partial\mathcal{L}_{\mathrm{int}}}{\partial\overline{\zeta}}(\psi(x),\overline{\psi}(x)).$$

From equation (7.164) we obtain $\frac{\delta S}{\delta\overline{\psi}(x)} = 0$.

(II) Similarly, for fixed $h \in L_2(\mathcal{M})$, set

$$\chi(\tau) := S[\psi + \tau h, \overline{\psi}, J, \overline{J}], \qquad \tau \in \mathbb{R}.$$

From $\chi'(0) = 0$ we get

$$\langle\overline{\psi}|(D + \mathrm{i}\varepsilon I)h\rangle + \langle\overline{J}|h\rangle + \kappa\sum_{x\in\mathcal{M}}\frac{\partial\mathcal{L}_{\mathrm{int}}}{\partial\zeta}(\psi(x),\overline{\psi}(x))h(x) = 0.$$

This implies

$$\sum_{x\in\mathcal{M}} \frac{\delta S}{\delta\psi(x)}h(x) = 0 \qquad \text{for all} \quad h \in L_2(\mathcal{M})$$

along with

$$\frac{\delta S}{\delta\psi(x)} = \{\overline{\psi}(D + \mathrm{i}\varepsilon I)\}(x) + \overline{J}(x) + \kappa\frac{\partial\mathcal{L}_{\mathrm{int}}}{\partial\zeta}(\psi(x),\overline{\psi}(x)).$$

Hence $\frac{\delta S}{\delta\psi(x)} = 0$. □

The response function. The inverse operator $R_\varepsilon := -(D + \mathrm{i}\varepsilon I)^{-1}$ is called response operator. Switching off the interaction by setting $\kappa = 0$, the so-called response equations

$$(D + \mathrm{i}\varepsilon I)\psi = -J, \qquad \overline{\psi}(D + \mathrm{i}\varepsilon I) = -\overline{J}$$

have the unique solution $\psi = R_\varepsilon J$ and $\overline{\psi} = \overline{J}R_\varepsilon$, respectively. Explicitly,

$$\psi(x) = \sum_{y\in\mathcal{M}} \mathcal{R}_\varepsilon(x,y)J(y)\Delta^4 y, \qquad \overline{\psi}(y) = \sum_{x\in\mathcal{M}} \overline{J}(x)\mathcal{R}_\varepsilon(x,y)\Delta^4 x. \tag{7.165}$$

The kernel $(x,y) \mapsto \mathcal{R}_\varepsilon(x,y)$ is called response function. Equation (7.165) tells us how the source function J influences the free field ψ, and how the second source function $\overline{J}$ influences the free field $\overline{\psi}$.

Double functional integrals. For each function $F : \mathbb{R}^{2N} \to \mathbb{R}$, the functional integral

$$\int_{L_2(\mathcal{M})\times L_2(\mathcal{M})} F(\psi,\overline{\psi})\ \mathcal{D}\overline{\psi}\mathcal{D}\psi$$

is defined by the following classical $2N$-dimensional integral

$$\int_{\mathbb{R}^{2N}} F(\psi(1),\dots,\psi(N),\overline{\psi}(1),\dots,\overline{\psi}(N))\prod_{k=1}^{N}\left(\frac{\Delta^4 x}{2\pi}\right)^N d\overline{\psi}(k)d\psi(k).$$

We also briefly write

$$\int_{\mathbb{R}^{2N}} F(\psi,\overline{\psi})\prod_{x\in\mathcal{M}}\left(\frac{\Delta^4 x}{2\pi}\right)^N d\overline{\psi}(x)d\psi(x).$$

The global quantum action principle. By definition, the fundamental quantum action functional reads as

$$\boxed{Z(J,\overline{J}) := \mathcal{N}\int_{\mathcal{F}} \mathrm{e}^{\mathrm{i}S[\psi,\overline{\psi},J,\overline{J}]/\hbar} f_\eta(\psi,\overline{\psi})\ \mathcal{D}\overline{\psi}\mathcal{D}\psi.}$$

Here, we use the function space $\mathcal{F} := L_2(\mathcal{M}) \times L_2(\mathcal{M})$ and the regularizing factor

$$f_\eta(\psi,\overline{\psi}) := \mathrm{e}^{-\eta\langle\psi|\psi\rangle-\eta\langle\overline{\psi}|\overline{\psi}\rangle}$$

for fixed $\eta > 0$. The number $\mathcal{N}$ has to be chosen in such a way that $Z(0,0) = 1$. The local functional derivatives of the quantum action functional Z represent moments of the fields ψ and $\overline{\psi}$. For example, since $\mathrm{e}^{\mathrm{i}S/\hbar} = \mathrm{e}^{\mathrm{i}\langle\overline{\psi}|J\rangle/\hbar}\cdot\mathrm{e}^{\mathrm{i}\langle\overline{J}|\psi\rangle/\hbar}\cdots$, we get

$$\frac{\hbar}{\mathrm{i}}\frac{\delta Z}{\delta J(y)} = \mathcal{N}\int_{\mathcal{F}}\overline{\psi}(y)\mathrm{e}^{\mathrm{i}S[\psi,\overline{\psi},J,\overline{J}]/\hbar}\ f_\eta(\psi,\overline{\psi})\ \mathcal{D}\overline{\psi}\mathcal{D}\psi.$$

This is the first moment with respect to the field variable $\overline{\psi}$ at the space-time point y. Similarly,

$$\left(\frac{\hbar}{\mathrm{i}}\right)^2\frac{\delta^2 Z}{\delta\overline{J}(x)\delta J(y)} = \mathcal{N}\int_{\mathcal{F}}\overline{\psi}(y)\psi(x)\mathrm{e}^{\mathrm{i}S[\psi,\overline{\psi},J,\overline{J}]/\hbar}\ f_\eta(\psi,\overline{\psi})\ \mathcal{D}\overline{\psi}\mathcal{D}\psi.$$

This is the second moment with respect to the field variables $\overline{\psi}$ and ψ at the space-time points y and x, respectively.

The extended quantum action functional. We define the functional

$$Z(J,\overline{J},\psi,\overline{\psi})$$

by the following expression

$$\boxed{\exp\left\{\frac{\mathrm{i}\kappa}{\hbar}\sum_{x\in\mathcal{M}}\Delta^4 x\;\mathcal{L}_{\mathrm{int}}\left(\frac{\hbar}{\mathrm{i}}\,\frac{\delta}{\delta\overline{J}(x)},\frac{\hbar}{\mathrm{i}}\frac{\delta}{\delta J(x)}\right)\right\}Z_{\mathrm{free}}(J,\overline{J},\psi,\overline{\psi})}$$

along with

$$Z_{\mathrm{free}}(J,\overline{J},\psi,\overline{\psi}) := Z_{\mathrm{free}}(J,\overline{J})\;\mathrm{e}^{\mathrm{i}\langle\overline{J}|\psi\rangle/\hbar}\;\mathrm{e}^{\mathrm{i}\langle\overline{\psi}|J\rangle/\hbar}.$$

The magic quantum action reduction formula for correlation functions. This formula tells us that

$$\boxed{Z(J,\overline{J}) = \frac{Z(J,\overline{J},0,0)}{Z(0,0,0,0)}.}$$

Explicitly,

$$Z(J,\overline{J}) = \mathcal{M}\exp\left\{\frac{\mathrm{i}\kappa}{\hbar}\sum_{x\in\mathcal{M}}\Delta^4 x\;\mathcal{L}_{\mathrm{int}}\left(\frac{\hbar}{\mathrm{i}}\,\frac{\delta}{\delta\overline{J}(x)},\frac{\hbar}{\mathrm{i}}\frac{\delta}{\delta J(x)}\right)\right\}Z_{\mathrm{free}}(J,\overline{J})$$

along with

$$Z_{\mathrm{free}}(J,\overline{J}) := \mathrm{e}^{\mathrm{i}\langle\overline{J}|R_\varepsilon J\rangle/\hbar}. \tag{7.166}$$

The number $\mathcal{M}$ is uniquely determined by the normalization condition $Z(0,0)=1$. The proof proceeds analogously to the proof of Theorem 7.38 on page 448 by using the principle of stationary phase. For example, switching off the interaction by setting $\kappa = 0$, we get

$$\boxed{\left(\frac{\hbar}{\mathrm{i}}\right)^2\frac{\delta^2 Z_{\mathrm{free}}}{\delta\overline{J}(x)\delta J(y)}|_{\,J=\overline{J}=0} = -\mathrm{i}\hbar\mathcal{R}_\varepsilon(x,y).}$$

This tells us that the response kernel function $\mathcal{R}_\varepsilon = \mathcal{R}_\varepsilon(x,y)$ represents a special 2-point correlation function. This corresponds to Prop. 7.43 on page 464. For $n,m = 1,2,\ldots$, the correlation functions

$$C_{n,m}(x_1,\ldots,x_n,y_1,\ldots,y_m)$$

are defined by

$$\left(\frac{\hbar}{\mathrm{i}}\right)^{n+m}\frac{\delta Z^{n+m}(J,\overline{J})}{\delta\overline{J}(x_1)\cdots\delta\overline{J}(x_n)\delta J(y_1)\cdots\delta J(y_m)}|_{\,J=\overline{J}=0}.$$

Furthermore, switching off the interaction by setting $\kappa = 0$, the free correlation functions

$$C_{n,m,\mathrm{free}}(x_1,\ldots,x_n,y_1,\ldots,y_m)$$

are given by

$$\left(\frac{\hbar}{\mathrm{i}}\right)^{n+m} \frac{\delta Z^{n+m}_{\mathrm{free}}(J,\overline{J})}{\delta\overline{J}(x_1)\cdots\delta\overline{J}(x_n)\delta J(y_1)\cdots\delta J(y_m)}\Big|_{J=\overline{J}=0}.$$

For example,

$$C_{1,1,\mathrm{free}}(x,y) = -\mathrm{i}\hbar\mathcal{R}_\varepsilon(x,y).$$

The magic LSZ reduction formula. Let us define the scattering functional S by setting

$$\boxed{\mathsf{S}(\psi,\overline{\psi}) := \frac{Z(0,0,\psi,\overline{\psi})}{Z(0,0,0,0)}.}$$

For $n,m = 1,2,\ldots$, the so-called scattering function

$$\mathsf{S}_{n,m}(x_1,\ldots,x_n,y_1,\ldots,y_m)$$

is defined by the functional derivative

$$\frac{\delta^{n+m}\mathsf{S}(\psi,\overline{\psi})}{\delta\psi(x_1)\cdots\delta\psi(x_n)\delta\overline{\psi}(y_1)\cdots\delta\overline{\psi}(y_m)}\Big|_{\psi=\overline{\psi}=0}.$$

Theorem 7.48 *The scattering function* $\mathsf{S}_{n,m}(x_1,\ldots,x_n,y_1,\ldots,y_m)$ *is equal to*

$$\frac{1}{(\mathrm{i}\hbar)^{n+m}}\left\{\prod_{l=1}^{n}(D_{x_l}+\mathrm{i}\varepsilon I)\cdot C_{n,m}\prod_{k=1}^{m}(D_{y_k}+\mathrm{i}\varepsilon I)\right\}(x_1,\ldots,x_n,y_1,\ldots,y_m).$$

The symbol $D_{x_l}+\mathrm{i}\varepsilon I$ denotes the operator $D+\mathrm{i}\varepsilon I$ acting on the variable x_l of the function $C_{n,m}$. Similarly, $D_{y_k}+\mathrm{i}\varepsilon I$ denotes the operator $D_{y_k}+\mathrm{i}\varepsilon I$ acting on the variable y_k of $C_{n,m}$. Explicitly, $\mathsf{S}_{n,m}(x_1,\ldots,x_n,y_1,\ldots,y_m)$ is equal to

$$\frac{1}{(\mathrm{i}\hbar)^{n+m}}\prod_{l=1}^{n}D(x_l,\xi_l)C_{n,m}(\xi_1,\ldots,\xi_n,\eta_1,\ldots,\eta_m)\prod_{k=1}^{m}D(\eta_k,y_k)$$

where the symbol $D(x,\xi)$ with $x,\xi\in\mathcal{M}$ denotes the matrix elements of the operator $D+\mathrm{i}\varepsilon I$. The proof of Theorem 7.48 proceeds similarly to the proof of Theorem 7.39 on page 450.

The mean field approximation. The reduced quantum action functional is defined by

$$Z_{\mathrm{red}}(J,\overline{J}) := \ln Z(J,\overline{J}).$$

Hence

$$Z(J,\overline{J}) = \mathrm{e}^{Z_{\mathrm{red}}(J,\overline{J})}.$$

By definition, the mean field reads as

$$\boxed{\psi_{\mathrm{mean}}(x) := \frac{\hbar}{\mathrm{i}} \frac{\delta Z_{\mathrm{red}}(J,\overline{J})}{\delta \overline{J}(x)}, \qquad \overline{\psi}_{\mathrm{mean}}(x) := \frac{\hbar}{\mathrm{i}} \frac{\delta Z_{\mathrm{red}}(J,\overline{J})}{\delta J(x)}.}$$

We assume that the map $(J,\overline{J}) \mapsto (\psi_{\mathrm{mean}}, \overline{\psi}_{\mathrm{mean}})$ is a diffeomorphism from $\mathbb{R}^{2N}$ onto $\mathbb{R}^{2N}$.

Vertex functions and effective action. Let us define the vertex functional

$$\boxed{V(\psi_{\mathrm{mean}}, \overline{\psi}_{\mathrm{mean}}) := Z_{\mathrm{red}}(J,\overline{J}) - \langle \overline{J}|\psi\rangle - \langle \overline{\psi}|J\rangle.}$$

It follows as in Sect. 7.24.12 on page 457 that

$$J(x) = -\frac{\delta V(\psi_{\mathrm{mean}}, \overline{\psi}_{\mathrm{mean}})}{\delta \overline{\psi}_{\mathrm{mean}}(x)}, \qquad \overline{J}(x) = -\frac{\delta V(\psi_{\mathrm{mean}}, \overline{\psi}_{\mathrm{mean}})}{\delta \psi_{\mathrm{mean}}(x)}.$$

The map

$$(J,\overline{J},Z_{\mathrm{red}}) \mapsto (\psi_{\mathrm{mean}}, \overline{\psi}_{\mathrm{mean}}, V)$$

is called Legendre transformation. Let us write ψ and $\overline{\psi}$ instead of ψ_{mean} and $\overline{\psi}_{\mathrm{mean}}$, respectively. For $n,m = 1,2,\ldots$, the functions

$$V_{n,m}(x_1,\ldots,x_n,y_1,\ldots,y_m) := \frac{\delta^{n+m} V(\psi,\overline{\psi})}{\delta\psi(x_1)\cdots\delta\psi(x_n)\delta\overline{\psi}(y_1)\cdots\delta\overline{\psi}(y_m)}\Big|_{\psi=\overline{\psi}=0}$$

are called vertex functions. Finally, as in Sect. 7.24.12 on page 457, the value

$$S_{\mathrm{eff}}[\psi_{\mathrm{mean}}, \overline{\psi}_{\mathrm{mean}}] := V(\psi_{\mathrm{mean}}, \overline{\psi}_{\mathrm{mean}})$$

is called the effective action of the mean field.

7.27 Complex-Valued Fields

Let us now modify the model from Sect. 7.26 on page 477 by setting

$$\overline{\psi} := \psi^\dagger, \qquad \overline{J} := J^\dagger$$

and by assuming that the components of ψ and J are complex numbers. In particular, the action functional reads as

$$\boxed{\begin{aligned} S[\psi,\psi^\dagger,J,J^\dagger] &:= \langle \psi^\dagger | D\psi\rangle + \kappa\langle 1|\mathcal{L}_{\mathrm{int}}(\psi,\psi^\dagger)\rangle \\ &\quad + \langle J^\dagger|\psi\rangle + \langle \psi^\dagger|J\rangle. \end{aligned}} \tag{7.167}$$

Explicitly, the functional $S[\psi,\psi^\dagger,J,J^\dagger]$ is equal to

$$\sum_{x,y\in\mathcal{M}} \psi(x)^\dagger D(x,y)\psi(y)\Delta^4 x\Delta^4 y + \kappa \sum_{x\in\mathcal{M}} \mathcal{L}_{\rm int}(\psi(x),\psi(x)^\dagger)\Delta^4 x$$
$$+ \sum_{x\in\mathcal{M}} \{J(x)^\dagger\psi(x) + \psi(x)^\dagger J(x)\}\Delta^4 x.$$

Recall that $\psi(x)^\dagger$ denotes the complex-conjugate of $\psi(x)$. We are given the coupling constant $\kappa \geq 0$, and we make the following assumptions:

- $\psi(1),\ldots,\psi(N)$ are complex field variables.
- $J(1),\ldots J(N)$ denote complex source variables.
- The complex $(N\times N)$-matrix D is self-adjoint.
- There is a number $\varepsilon_0 > 0$ such that the inverse matrix $(D+\mathrm{i}\varepsilon I)^{-1}$ exists for all $\varepsilon \in]0,\varepsilon_0[$.
- The function $\mathcal{L}_{\rm int} : \mathbb{C}^2 \to \mathbb{R}$ is a complex polynomial of the two variables ζ and $\zeta^\dagger$. For example, $\mathcal{L}_{\rm int}(\zeta,\zeta^\dagger) := (\zeta^\dagger\zeta)^2$.

The special case where $\mathcal{L}_{\rm int} := -(\psi\psi^\dagger)^2 = -|\psi|^4$ is called the discrete complex φ^4-model. Let $L_2(\mathcal{M})$ denote the space of all functions $\psi : \mathcal{M} \to \mathbb{C}$. As in the real case, we set

$$\langle f|g\rangle := \sum_{x\in\mathcal{M}} f(x)g(x)\Delta^4 x.$$

Note that in contrast to the dual pairing $\langle f|g\rangle$, the inner product on the complex Hilbert space $L_2(\mathcal{M})$ is denoted by

$$\langle f|g\rangle_{L_2(\mathcal{M})} := \sum_{x\in\mathcal{M}} f(x)^\dagger g(x)\ \Delta^4 x.$$

The principle of critical action. We are given the source function $J \in L_2(\mathcal{M})$.

Theorem 7.49 *Each solution $\psi \in L_2(\mathcal{M})$ of the variational problem*

$$S[\psi,\psi^\dagger,J,J^\dagger] = \text{critical!} \tag{7.168}$$

satisfies the following Euler-Lagrange equations

$$\frac{\delta S}{\delta\psi^\dagger(x)} = 0, \qquad \frac{\delta S}{\delta\psi(x)} = 0 \qquad \textit{for all} \quad x\in\mathcal{M}.$$

Explicitly, the Euler–Lagrange equations read as

$$(D\psi)(x) + \kappa\frac{\partial\mathcal{L}_{\rm int}}{\partial\zeta^\dagger}(\psi(x),\psi(x)^\dagger) + J(x) = 0 \tag{7.169}$$

and

$$(\psi^\dagger D)(x) + \kappa \frac{\partial \mathcal{L}_{\text{int}}}{\partial \zeta}(\psi(x), \psi(x)^\dagger) + J(x)^\dagger = 0. \tag{7.170}$$

Proof. (I) To begin with, let us prove the following complex variational lemma. We are given $f, g \in L_2(\mathcal{M})$. Suppose that

$$\langle h^\dagger | f \rangle + \langle g | h \rangle = 0 \qquad \text{for all} \quad h \in L_2(\mathcal{M}). \tag{7.171}$$

Then $f \equiv 0$ and $g \equiv 0$. In fact, equation (7.171) reads explicitly as

$$\sum_{x \in \mathcal{M}} (h(x)^\dagger f(x) + g(x) h(x)) \Delta^4 x = 0.$$

First use real values $h(x)$. Then, $f(x) + g(x) = 0$ for all $x \in \mathcal{M}$. Secondly, use purely imaginary values $h(x)$. Then, $-f(x) + g(x) = 0$ for all $x \in \mathcal{M}$. Hence $f(x) = g(x) = 0$ for all $x \in \mathcal{M}$.

(II) For fixed $h \in L_2(\mathcal{M})$, set

$$\chi(\tau) := S[\psi + \tau h, \psi^\dagger + \tau h^\dagger, J, J^\dagger], \qquad \tau \in \mathbb{R}.$$

From the variational problem (7.168), we get $\chi'(0) = 0$. This yields

$$\begin{aligned} &\langle h^\dagger | D\psi \rangle + \langle h^\dagger | J \rangle + \kappa \sum_{x \in \mathcal{M}} \frac{\partial \mathcal{L}_{\text{int}}}{\partial \zeta^\dagger}(\psi(x), \psi(x)^\dagger) h(x)^\dagger \\ &+ \langle \psi^\dagger | Dh \rangle + \langle J^\dagger | h \rangle + \kappa \sum_{x \in \mathcal{M}} \frac{\partial \mathcal{L}_{\text{int}}}{\partial \zeta}(\psi(x), \psi(x)^\dagger) h(x) = 0. \end{aligned}$$

The claim follows now from (I). □

The response function. Switching off the interaction by setting $\kappa = 0$, the regularized Euler–Lagrange equation (7.169) reads as

$$\boxed{D\psi + \mathrm{i}\varepsilon\psi = -J} \tag{7.172}$$

where the regularization parameter $\varepsilon > 0$ is small enough that the inverse operator $R_\varepsilon := -(D + \mathrm{i}\varepsilon I)^{-1}$ exists. Using matrix elements, the unique solution $\psi = R_\varepsilon J$ of equation (7.172) is given by

$$\psi(x) = \sum_{y \in \mathcal{M}} \mathcal{R}_\varepsilon(x, y) J(y) \Delta^4 y, \qquad x \in \mathcal{M}.$$

The function $\mathcal{R}_\varepsilon$ is called the response kernel. Using $D^\dagger = D$, it follows from (7.172) that

$$\boxed{\psi^\dagger D - \mathrm{i}\varepsilon\psi^\dagger = -J^\dagger.}$$

This is the regularized version of the linearized Euler–Lagrange equation (7.170) with the unique solution

$$\psi(x)^\dagger = \sum_{y\in\mathcal{M}} J(y)^\dagger \mathcal{R}_\varepsilon(x,y)^\dagger \Delta^4 y, \qquad x \in \mathcal{M}.$$

We will show in Problem 7.6 on page 495 that the regularized equation (7.172) does not follow from a variational problem.

The extended quantum action functional. The functional

$$Z(J, J^\dagger, \psi, \psi^\dagger)$$

is defined by

$$\boxed{\exp\left\{\frac{\mathrm{i}\kappa}{\hbar}\sum_{x\in\mathcal{M}} \Delta^4 x\, \mathcal{L}_{\mathrm{int}}\left(\frac{\hbar}{\mathrm{i}}\frac{\delta}{\delta J^\dagger(x)}, \frac{\hbar}{\mathrm{i}}\frac{\delta}{\delta J(x)}\right)\right\} Z_{\mathrm{free}}(J, J^\dagger, \psi, \psi^\dagger)}$$

along with

$$Z_{\mathrm{free}}(J, J^\dagger, \psi, \psi^\dagger) := Z_{\mathrm{free}}(J, J^\dagger)\, \mathrm{e}^{\mathrm{i}\langle J^\dagger|\psi\rangle/\hbar}\, \mathrm{e}^{\mathrm{i}\langle\psi^\dagger|J\rangle/\hbar}.$$

The magic quantum action reduction formula for correlation functions. We define the quantum action functional of our complex model by

$$\boxed{Z(J, J^\dagger) := \frac{Z(J, J^\dagger, 0, 0)}{Z(0,0,0,0)}.}$$

Explicitly,

$$Z(J, J^\dagger) := \mathcal{M}\exp\left\{\frac{\mathrm{i}\kappa}{\hbar}\sum_{x\in\mathcal{M}} \Delta^4 x\, \mathcal{L}_{\mathrm{int}}\left(\frac{\hbar}{\mathrm{i}}\frac{\delta}{\delta J^\dagger(x)}, \frac{\hbar}{\mathrm{i}}\frac{\delta}{\delta J(x)}\right)\right\} Z_{\mathrm{free}}(J, J^\dagger)$$

along with

$$Z_{\mathrm{free}}(J, J^\dagger) := \mathrm{e}^{\mathrm{i}\langle J^\dagger|R_\varepsilon J\rangle/\hbar}.$$

The number $\mathcal{M}$ is uniquely determined by the normalization condition $Z(0,0) = 1$. This yields the correlation function

$$C_{n,m}(x_1, \dots, x_n, y_1, \dots, y_m)$$

defined by

$$\left(\frac{\hbar}{\mathrm{i}}\right)^{n+m} \frac{\delta^{n+m} Z(J, J^\dagger)}{\delta J^\dagger(x_1)\cdots\delta J^\dagger(x_n)\delta J(y_1)\cdots\delta J(y_m)}\Big|_{J=0}.$$

For example,

$$C_{1,1,\mathrm{free}}(x,y) = \left(\frac{\hbar}{\mathrm{i}}\right)^2 \frac{\delta Z_{\mathrm{free}}(J, J^\dagger)}{\delta J^\dagger(x)\delta J(y)} = -\mathrm{i}\hbar\mathcal{R}_\varepsilon(x,y).$$

The magic LSZ reduction formula. Let us define the scattering functional S by setting

$$\mathsf{S}(\psi, \psi^\dagger) := \frac{Z(0,0,\psi,\psi^\dagger)}{Z(0,0,0,0)}$$

For $n, m = 1, 2, \ldots$, the scattering function

$$\mathsf{S}_{n,m}(x_1, \ldots, x_n, y_1, \ldots, y_m)$$

is defined by the functional derivative

$$\frac{\delta^{n+m}\mathsf{S}(\psi,\psi^\dagger)}{\delta\psi(x_1)\cdots\delta\psi(x_n)\delta\psi^\dagger(y_1)\cdots\delta\psi^\dagger(y_m)}\Big|_{\psi=0}.$$

Theorem 7.50 *The scattering function* $\mathsf{S}_{n,m}(x_1, \ldots, x_n, y_1, \ldots, y_m)$ *is equal to*

$$\frac{1}{(\mathrm{i}\hbar)^{n+m}}\left\{\prod_{l=1}^{n}(D_{x_l}+\mathrm{i}\varepsilon I)\cdot C_{n,m}\prod_{k=1}^{m}(D_{y_k}+\mathrm{i}\varepsilon I)\right\}(x_1,\ldots,x_n,y_1,\ldots,y_m).$$

The proof of this theorem proceeds as for the $(\psi, \overline{\psi})$-model in Sect. 7.26. Parallel to the $(\psi, \overline{\psi})$-model, it is also possible to obtain the quantum action functional for the present $(\psi, \psi^\dagger)$-model from a global quantum action principle based on a finite functional integral.

7.28 The Method of Lagrange Multipliers

> Reduce constrained extremal problems to free extremal problems by changing the Lagrangian.
>
> Folklore

The Standard Model of particle physics is based on gauge field theory. The fixing of a gauge corresponds to some constraint. Therefore, constrained variational problems will play a fundamental role in the Standard Model later on. At this point, let us only summarize the basic ideas of the rigorous theory for finite-dimensional constrained problems. The point is that the Lagrange multiplier rule only applies to regular solutions of constrained problems. From the physical point of view, Lagrange multipliers correspond to

- constraining forces in classical mechanics,
- ghosts and antighosts (BRST symmetry) in non-commutative gauge field theories (e.g., electroweak and strong interaction in the Standard Model).[51]

[51] In the author's monograph, Zeidler (1986), Vol. III, the interested reader may find the general functional analytic theory of Lagrange multipliers along with many applications to extremal problems in the calculus of variations, linear and nonlinear optimization (e.g., the Kuhn–Tucker theory for convex problems) and optimal control (e.g., the Pontryagin maximum principle applied to optimal moon landing and the optimal return of a spaceship to earth).

Simple constrained problem. Consider the minimum problem

$$y = \min! \tag{7.173}$$

under the constraint

$$x^2 + y^2 = 1. \tag{7.174}$$

Obviously, the solution is $x_0 = 0, y_0 = -1$. The problem

$$y = \text{critical!} \tag{7.175}$$

under the constraint (7.174) means the following. Set $F(x, y) := y$. Consider a smooth curve $x = x(\tau), y = y(\tau)$ which satisfies the constraint (7.174) along with $x(0) = x_0, y(0) = y_0$. Set

$$\chi(\tau) := F(x(\tau), y(\tau)), \qquad \tau \in \mathbb{R}.$$

By definition, the point (x_0, y_0) is a solution of problem (7.175) along with the constraint (7.174) iff $\chi'(0) = 0$. Explicitly, the solutions are $x_0 = 0$ and $y_0 = \pm 1$ which correspond to a minimum and a maximum, respectively.

Lagrange's idea. Consider the problem

$$F(x, y) = \text{critical!} \tag{7.176}$$

under the constraint

$$G(x, y) = 0. \tag{7.177}$$

Suppose that the functions $F, G : \mathbb{R}^2 \to \mathbb{R}$ are smooth. The solution (x_0, y_0) of (7.176), (7.177) is called regular iff for each real number c, the linearized constraint

$$G_x(x_0, y_0)a + G_y(x_0, y_0)b = c$$

has a solution $(a, b) \in \mathbb{R}^2$. In other words,

$$(G_x(x_0, y_0), G_y(x_0, y_0)) \neq (0, 0).$$

This means that the rank of this matrix is maximal (i.e., equal to 1).[52] It was the idea of Lagrange to replace the original constrained problem (7.176), (7.177) by the free problem

$$F(x, y) + \lambda G(x, y) = \text{critical!} \tag{7.178}$$

where λ is a fixed, but unknown real parameter. This parameter is called Lagrange multiplier.

[52] In terms of geometry, we postulate that the constraint $G(x, y) = 0,\ (x, y) \in \mathbb{R}^2$ represents a manifold (i.e., a curve which has a tangent line at each point).

Proposition 7.51 *If (x_0, y_0) is a regular solution of the constrained problem* (7.176), (7.177)*, then there exists a real number λ such that (x_0, y_0) is a solution of the free problem* (7.178)*, that is,*

$$F_x(x_0, y_0) + \lambda G_x(x_0, y_0) = 0, \quad F_y(x_0, y_0) + \lambda G_y(x_0, y_0) = 0. \quad (7.179)$$

As a rule, this equation along with the constraint $G(x_0, y_0) = 0$ determines uniquely the Lagrange multiplier λ. In classical mechanics, the value of λ describes the strength of the constraining force (see the spherical pendulum below). The proof of the classical Proposition 7.51 can be found in Zeidler (1986), Vol. III, Sect. 43.10.

In the case of problem (7.173), (7.174), we have $F(x, y) := y$ and

$$G(x, y) := x^2 + y^2 - 1.$$

The constraint $G(x, y) = 0$ describes the unit circle. Obviously, this is a manifold. Equation (7.179) yields

$$2\lambda x_0 = 0, \quad 1 + 2\lambda y_0 = 0.$$

Hence $\lambda \neq 0$. This implies $x_0 = 0$. From $x^2 + y^2 = 1$ we get $y_0 = \pm 1$. Finally, $\lambda = \mp 1/2$.

The principle of critical action under constraints. Consider the variational problem

$$S[\varphi] = \text{critical!} \quad (7.180)$$

under the constraints

$$G_k[\varphi] = 0, \qquad k = 1, \dots, K. \quad (7.181)$$

Suppose that $S, G_1, \dots, G_K : \mathbb{R}^N \to \mathbb{R}$ are smooth functions where $K < N$. Let us also consider the system of linearized constraints:[53]

$$\sum_{x \in \mathcal{M}} \frac{\delta G_k[\varphi_0]}{\delta \varphi(x)} \, h_k(x) = c_k, \qquad k = 1, \dots, K. \quad (7.182)$$

A solution φ_0 of (7.180), (7.181) is called regular iff for arbitrary real numbers $c_1, \dots, c_K$, equation (7.182) has always a real solution $h_1, \dots h_k$.

Theorem 7.52 *Let φ_0 be a regular solution of the constrained problem* (7.180), (7.181)*. Then, there exist real numbers λ_k indexed by $k = 1, \dots, K$ such that*

$$\frac{\delta S[\varphi_0]}{\delta \varphi(x)} + \sum_{k=1}^{K} \lambda_k \frac{\delta G_k[\varphi_0]}{\delta \varphi(x)} = 0 \qquad \textit{for all} \quad x \in \mathcal{M}.$$

[53] Recall that $\mathcal{M} = \{1, \dots, N\}$. Moreover, the components of φ are denoted by $\varphi(1), \dots, \varphi(N)$.

The proof of this finite-dimensional theorem along with infinite-dimensional generalizations and numerous applications can be found in Zeidler (1986), Vol. III, Sect. 43.10.

The universal Gaussian principle of critical constraint. We want to study the motion $\mathbf{x} = \mathbf{x}(t)$ of a classical particle of mass $m > 0$ under the constraints

$$f(\mathbf{x}(t), t) = 0, \qquad g(\mathbf{x}(t), \dot{\mathbf{x}}(t), t) \ = 0. \tag{7.183}$$

From this, we get the linearized constraints[54]

$$\begin{aligned} f_{\mathbf{x}}(X)\ddot{\mathbf{x}} + \dot{\mathbf{x}} f_{\mathbf{xx}}(X)\dot{\mathbf{x}} + 2f_{\mathbf{x}t}(X)\dot{\mathbf{x}} + f_{tt}(X) = 0,\\ g_{\dot{\mathbf{x}}}(X)\ddot{\mathbf{x}} + g_{\mathbf{x}}(X)\dot{\mathbf{x}} + g_t(X) = 0 \end{aligned} \tag{7.184}$$

where $X := (\mathbf{x}, \dot{\mathbf{x}}, t)$. We assume that the functions $f : \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}$ are smooth, and the regularity condition

$$f_{\mathbf{x}}(X)^2 g_{\dot{\mathbf{x}}}(X)^2 - f_{\mathbf{x}}(X) g_{\dot{\mathbf{x}}}(X) \neq 0 \tag{7.185}$$

is satisfied for all points X with $f(X) = g(X) = 0$. According to Gauss, the equation of motion for the particle reads as

$$\boxed{m\ddot{\mathbf{x}} = \mathbf{F} + \lambda f_{\mathbf{x}} + \mu g_{\dot{\mathbf{x}}}.} \tag{7.186}$$

The additional force $\mathbf{F}_c := \lambda f_{\mathbf{x}} + \mu g_{\dot{\mathbf{x}}}$ is called constraining force. Explicitly,

$$m\ddot{\mathbf{x}} = \mathbf{F}(X) + \lambda(X) f_{\mathbf{x}}(X) + \mu(X) g_{\dot{\mathbf{x}}}(X).$$

Here, $\lambda(X)$ and $\mu(X)$ are uniquely determined by the linearized constraint (7.184). In fact, it follows from (7.184) and (7.186) that

$$\begin{aligned} \lambda f_{\mathbf{x}}^2 + \mu g_{\dot{\mathbf{x}}} f_{\mathbf{x}} &= -\mathbf{F} f_{\mathbf{x}} - \dot{\mathbf{x}} f_{\mathbf{xx}}\dot{\mathbf{x}} - 2f_{\mathbf{x}t}\dot{\mathbf{x}} - f_{tt},\\ \lambda g_{\dot{\mathbf{x}}} f_{\mathbf{x}} + \mu g_{\dot{\mathbf{x}}}^2 &= -\mathbf{F} g_{\dot{\mathbf{x}}} - g_{\mathbf{x}}\dot{\mathbf{x}} - g_t \end{aligned}$$

at the point $X := (\mathbf{x}, \dot{\mathbf{x}}, t)$. By the regularity condition (7.185), this linear system has a unique solution (λ, μ). Let us now show that the equation of motion follows from the Gaussian principle of critical constraint:

[54] Equation (7.184) is obtained from (7.183) by differentiating with respect to time t. In fact, it follows from $f(\mathbf{x}(t), t) = 0$ that

$$f_{\mathbf{x}}(\mathbf{x}(t), t)\dot{\mathbf{x}}(t) + f_t(\mathbf{x}(t), t) = 0.$$

Again differentiating this with respect to t, we obtain

$$f_{\mathbf{x}}(\mathbf{x}(t), t)\ddot{\mathbf{x}}(t) + \dot{\mathbf{x}}(t) f_{\mathbf{xx}}(\mathbf{x}(t), t)\dot{\mathbf{x}}(t) + 2f_{\mathbf{x}t}(\mathbf{x}(t), t)\dot{\mathbf{x}}(t) + f_{tt}(\mathbf{x}(t), t) = 0.$$

$$\frac{m}{2}\left(\ddot{\mathbf{x}} - \frac{\mathbf{F}}{m}\right)^2 = \text{critical!}, \qquad (7.187)$$
$$f_{\mathbf{x}}(X)\ddot{\mathbf{x}} + \dot{\mathbf{x}} f_{\mathbf{xx}}(X)\dot{\mathbf{x}} + 2f_{\mathbf{x}t}(X)\dot{\mathbf{x}} + f_{tt}(X) = 0,$$
$$g_{\dot{\mathbf{x}}}(X)\ddot{\mathbf{x}} + g_{\mathbf{x}}(X)\dot{\mathbf{x}} + g_t(X) = 0.$$

This represents a variant of the Gaussian least square method. We are looking for a solution $\ddot{\mathbf{x}}$ which depends on the fixed parameter $X := (\mathbf{x}, \dot{\mathbf{x}}, t)$. A solution of (7.187) is called regular iff condition (7.185) is satisfied.

Theorem 7.53 *Every regular solution of the Gaussian principle of critical constraint* (7.187) *satisfies the equation of motion* (7.186).

Proof. (I) Regularity. By (7.185), the rank of the matrix

$$(f_{\mathbf{x}}(X), g_{\dot{\mathbf{x}}}(X))$$

is maximal (i.e., equal to 2). To show this suppose that $\lambda f_{\mathbf{x}}(X) + \mu g_{\dot{\mathbf{x}}}(X) = 0$. Hence

$$\lambda f_{\mathbf{x}}(X)^2 + \mu f_{\mathbf{x}}(X) g_{\dot{\mathbf{x}}}(X) = 0,$$
$$\lambda g_{\dot{\mathbf{x}}}(X) f_{\mathbf{x}}(X) + \mu g_{\dot{\mathbf{x}}}(X)^2 = 0.$$

By (7.185), $\lambda = \mu = 0$.

(II) We are now able to use the general Lagrange multiplier rule on $\mathbb{R}^N$ whose proof can be found in Zeidler (1986), Sect. 43.10. By this rule, the original constrained problem (7.187) is equivalent to the free problem

$$\frac{m}{2}\left(\ddot{\mathbf{x}} - \frac{\mathbf{F}}{m}\right)^2 + \lambda(f_{\mathbf{x}}\ddot{\mathbf{x}} + \dot{\mathbf{x}} f_{\mathbf{xx}}\dot{\mathbf{x}} + 2f_{\mathbf{x}t}\dot{\mathbf{x}} + f_{tt}) +$$
$$+ \mu(g_{\dot{\mathbf{x}}}\ddot{\mathbf{x}} + g_{\mathbf{x}}\dot{\mathbf{x}} + g_t) = \text{critical!}$$

Differentiating this with respect to $\ddot{\mathbf{x}}$, we get the equation of motion (7.186). □

Special case. Suppose that $g \equiv 0$. In this case, we have to replace the regularity condition (7.185) by $f_{\mathbf{x}}(\mathbf{x}, t) \neq 0$ for all $(\mathbf{x}, t)$ with $f(\mathbf{x}, t) = 0$. The equation of motion reads now as

$$\boxed{m\ddot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, \dot{\mathbf{x}}, t) + \lambda(\mathbf{x}, \dot{\mathbf{x}}, t) f_{\mathbf{x}}(\mathbf{x}, t).}$$

The parameter λ follows uniquely from

$$f_{\mathbf{x}}(X)^2\lambda = -\mathbf{F}(X) f_{\mathbf{x}}(X) - \dot{\mathbf{x}} f_{\mathbf{xx}}(X)\dot{\mathbf{x}} - 2f_{\mathbf{x}t}(X)\dot{\mathbf{x}} - f_{tt}(X).$$

Application to the spherical pendulum. Let us investigate the motion of a point of mass $m > 0$ on a sphere of radius R under the action of the gravitational force of earth. This corresponds to the constraint

$$\mathbf{x}^2 - R^2 = 0$$

and the force $\mathbf{F} := -mg\mathbf{k}$. Here, we use the outer normal unit vector $\mathbf{k}$ of earth and the acceleration constant $g = 9.81\text{m/s}^2$. Setting $f(\mathbf{x}) := \frac{1}{2}(\mathbf{x}^2 - R^2)$, we get the equation of motion

$$m\ddot{\mathbf{x}} = -mg\mathbf{k} + \mathbf{F}_c.$$

The constraining force is given by $\mathbf{F}_c := \lambda f_{\mathbf{x}}(\mathbf{x}) = \lambda\mathbf{x}$ (normal force) along with

$$f_{\mathbf{x}}^2\lambda = mg \cdot \mathbf{k}f_{\mathbf{x}} - \dot{\mathbf{x}}f_{\mathbf{xx}}\dot{\mathbf{x}} = mg \cdot \mathbf{kx} - \dot{\mathbf{x}}^2.$$

Hence

$$\lambda = \frac{mg \cdot \mathbf{kx} - \dot{\mathbf{x}}^2}{R^2}.$$

Here, $z = \mathbf{kx}$ is the height of the point $\mathbf{x}$. For example, if the mass point rests at the South Pole of the sphere, then $\ddot{\mathbf{x}} = 0$. As expected, this implies

$$\mathbf{F}_c = -\mathbf{F}_{\text{gravitation}} = mg\mathbf{k}.$$

The Gaussian principle of critical constraint is the most general principle for the motion of n particles under constraints in classical mechanics. The general formulation of this principle can be found in Zeidler (1986), Vol. IV, Sect. 58.10.

7.29 The Formal Continuum Limit

The passage from the discrete φ^4-model to the corresponding continuum model is based on the formal limit

$$N \to +\infty, \qquad \Delta x \to +0 \tag{7.188}$$

Explicitly, this means the following:

- Space-time manifold: $\mathcal{M} \to \mathbb{M}^4$.
- Integral:

$$\sum_{x \in \mathcal{M}} f(x)\Delta^4 x \to \int_{\mathbb{M}^4} f(x)d^4x.$$

- Dirac's delta function: $\delta_{\Delta^4 x}(x, y) \to \delta^4(x - y)$.
- Functional derivative:[55]

$$\frac{1}{\Delta^4 x}\,\frac{\partial Z(J)}{\partial J(x)} \to \frac{\delta Z(J)}{\delta J(x)}.$$

[55] To simplify notation, we use the same symbol $\frac{\delta Z(J)}{\delta J(x)}$ in both the discrete case and the continuum case.

- Functional integral:

$$\int F(\varphi) \prod_{x\in\mathcal{M}} d\varphi(x) \to \int F(\varphi) \prod_{x\in\mathbb{M}^4} d\varphi(x).$$

The point is that, as a rule, the formal limits above produce divergent expressions. The idea of the method of counterterms is to change the Lagrangian of the theory in such a way that

(i) new terms appear which regularize the divergent integrals, and
(ii) the original physical parameters are replaced by new finite parameters which have to be determined by the physical experiment.

This is the procedure of renormalization to be discussed in Sect. 15.4. The counterterms for the continuum φ^4-model can be found in Sect. 15.4.2.

Problems

7.1 *The Fourier–Gauss integral.* Prove (7.91). Solution: Set

$$h(p) := \int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2}\mathrm{e}^{-\mathrm{i}px}\, dx$$

where $p \in \mathbb{C}$. Differentiating this,

$$h'(p) = -\mathrm{i}\int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2} x\mathrm{e}^{-\mathrm{i}px} dx = \mathrm{i}\int_{-\infty}^{\infty} \frac{d}{dx}\left(\mathrm{e}^{-x^2/2}\right)\mathrm{e}^{-\mathrm{i}px} dx.$$

This derivative exists, by the majorant criterion for integrals (see Zeidler (1995), Vol. 1, p. 440). In fact, let $p = \alpha + \mathrm{i}\beta$ where $\alpha, \beta \in \mathbb{R}$. Choose a real number $C > 0$ and a natural number $n = 0, 1, 2, \ldots$. Then

$$|x^n \mathrm{e}^{-x^2/2}\mathrm{e}^{-\mathrm{i}px}| = |x^n|\, \mathrm{e}^{-x^2/2}\, \mathrm{e}^{\beta x} \le \mathrm{const}\cdot \mathrm{e}^{-x^2/4} \tag{7.189}$$

for all $x, \alpha \in \mathbb{R}$ and all $\beta \in [-C, C]$. The constant depends on C and n. The right-hand side of (7.189) represents a majorant for the integrands of the integrals corresponding to h and h'. The point is that this majorant is integrable over $\mathbb{R}$.
Using integration by parts, $h'(p) = -ph(p)$. Hence $h(p) = \mathrm{e}^{-p^2/2}h(0)$. Furthermore,

$$h(0)^2 = \left(\int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2} dx\right)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-(x^2+y^2)/2} dxdy.$$

Using polar coordinates,

$$h(0)^2 = \int_0^{\infty}\int_{-\pi}^{\pi} \mathrm{e}^{-r^2/2} r dr d\varphi = -2\pi\int_0^{\infty} \frac{d}{dr}\left(\mathrm{e}^{-r^2/2}\right) dr = 2\pi.$$

7.2 *The limiting relation.* Prove (7.94). Solution: Set $f_\sigma(x) := \frac{\mathrm{e}^{-x^2/2\sigma^2}}{\sigma\sqrt{2\pi}}$ if $\sigma > 0$. Using the rescaling $x = \sigma y$,

$$\lim_{\sigma\to+0}\int_{-\infty}^{\infty} f_\sigma(x)\chi(x)dx = \lim_{\sigma\to+0}\int_{-\infty}^{\infty} f_1(y)\chi(\sigma y)dy$$
$$= \int_{-\infty}^{\infty} f_1(y)\chi(0)dy = \chi(0).$$

This limit exists because of the majorant condition. In fact, for all $y \in \mathbb{R}$,

$$|f_1(y)\chi(\sigma y)| \le \mathrm{const}\cdot f_1(y),$$

and the right-hand side is integrable over the real line.

7.3 *The determinant trick.* Prove Theorem 7.34. Solution: Let T be a real orthogonal $(N\times N)$-matrix, that is $T^{-1} = T^d$. Define

$$Z(J) := \frac{1}{(2\pi)^{N/2}}\int_{\mathbb{R}^N} \exp^{-\frac{1}{2}\varphi^d A\varphi}\ \mathrm{e}^{\mathrm{i}J^d\varphi}\ d\varphi_1\cdots d\varphi_N.$$

Using the transformation $\varphi = T\psi, p^d = J^dT$, we get

$$Z(J) = \int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\psi^d(T^dAT)\psi}\ \mathrm{e}^{\mathrm{i}p^d\psi}\ |\det T|\ d\psi_1\cdots d\psi_N.$$

It follows from $(\det T)^{-1} = \det T^{-1} = \det T^d = \det T$ that $\det T = \pm 1$. By the principal axis theorem, the orthogonal matrix T can be chosen in such a way that

$$\varphi^d A\varphi = \psi^d(T^dAT)\psi = \lambda_1\psi_1^2 + \ldots + \lambda_N\psi_N^2.$$

Hence T^dAT is equal to the diagonal matrix $\mathrm{diag}(\lambda_1,\ldots,\lambda_N)$. By (7.105),

$$Z(J) = \frac{\mathrm{e}^{-\frac{1}{2}\sum_{k=1}^N \lambda_k p_k^2}}{\sqrt{\lambda_1\cdots\lambda_N}}.$$

Furthermore, $\prod_{k=1}^N \lambda_k = \det(T^dAT) = \det T^d \det A \det T = \det A$. Finally,

$$\mathrm{diag}(\lambda_1^{-1},\ldots,\lambda_N^{-1}) = (T^dAT)^{-1} = T^{-1}A^{-1}(T^d)^{-1}.$$

Hence $J^dA^{-1}J = p^dT^{-1}A^{-1}(T^d)^{-1}p = \sum_{k=1}^N \lambda_k^{-1}p_k^2$.

7.4 *Commutation rule for matrices.* Let A, B be complex $(N\times N)$-matrices for $N = 2, 3, \ldots$ Set $[A,B] := AB - BA$. Show that

$$[[A,B],B] = 0 \tag{7.190}$$

implies

$$[A, \mathrm{e}^B] = \mathrm{e}^B[A,B]. \tag{7.191}$$

This is a simple special case of the famous Baker–Campbell–Hausdorff formula from Sect. 8.4 on page 508.

Solution: It is sufficient to show that

$$AB^{n+1} + nB^{n+1}A = (n+1)B^nAB, \qquad n = 1,2,\ldots \tag{7.192}$$

In fact, this implies

$$\frac{AB^{n+1}}{(n+1)!} - \frac{B^{n+1}A}{(n+1)!} = \frac{B^n}{n!}\,(AB - BA). \tag{7.193}$$

Since $\mathrm{e}^B = I + B + \frac{1}{2}B^2 + \ldots$, it follows from (7.193) that

$$A\mathrm{e}^B - \mathrm{e}^B A = \mathrm{e}^B(AB - BA).$$

This is the claim. It remains to prove (7.192) by induction.
(I) $n = 1$. By (7.190), $(AB - BA)B = B(AB - BA)$. Hence

$$AB^2 + B^2A = 2BAB. \tag{7.194}$$

(II) $n = 2$. Multiply equation (7.194) by B from the right and from the left. Then

$$AB^3 + B^2AB = 2BAB^2,$$
$$2BAB^2 + 2B^3A = 4B^2AB.$$

Adding this, we get the claim (7.192) for $n = 2$. The general induction proof proceeds analogously.

7.5 *Proof of Theorem* 7.44(iii). Proceed as in the proof of (ii). Hint: The explicit computations can be found in Greiner and Reinhardt (1996b), p. 397.

7.6 *Corollary to Theorem* 7.49. Let $\varepsilon > 0$. Parallel to Theorem 7.49 on page 484, consider the modified variational problem

$$\langle\psi^\dagger|(D + \mathrm{i}\varepsilon I)\psi\rangle + \langle J^\dagger|\psi\rangle + \langle\psi^\dagger|J\rangle = \text{critical!} \tag{7.195}$$

Recall that $D^\dagger = D$, by assumption. Show the following:
(i) For given $J \in L_2(\mathcal{M})$, the variational problem (7.195) is equivalent to the Euler–Lagrange equations

$$(D + \mathrm{i}\varepsilon)\psi = -J, \qquad \psi^\dagger(D + \mathrm{i}\varepsilon I) = -J^\dagger. \tag{7.196}$$

(ii) If $J \neq 0$, the variational problem (7.195) has no solution.
(iii) If $\varepsilon = 0$ and D is invertible, then the variational problem (7.195) has the unique solution $\psi = -D^{-1}J$.

Solution: Ad (i). Argue as in the proof of Theorem 7.49.
Ad (ii). Let ψ be a solution of (7.195). This implies (7.196). It follows from $(D + \mathrm{i}\varepsilon I)\psi = -J$ along with $D^\dagger = D$ that

$$\psi^\dagger(D - \mathrm{i}\varepsilon I) = -J^\dagger.$$

By (7.196), $2\mathrm{i}\varepsilon\psi^\dagger = 0$. Hence $\psi = 0$. Again by (7.196), we get $J = 0$, a contradiction.
Ad (iii). Observe that $\psi^\dagger D = -J^\dagger$ is a consequence of $D\psi = -J$.

8. Rigorous Finite-Dimensional Perturbation Theory

Perturbation theory is the most important method in modern physics.
Folklore

8.1 Renormalization

In quantum field theory, a crucial role is played by renormalization. Let us now study this phenomenon in a very simplified manner.

- We want to show how mathematical difficulties arise if nonlinear equations are linearized in the incorrect place.
- Furthermore, we will discuss how to overcome these difficulties by using the methods of bifurcation theory.

The main trick is to replace the original problem by an equivalent one by introducing so-called regularizing terms. We have to distinguish between

- the non-resonance case (N) (or regular case), and
- the resonance case (R) (or singular case).

In celestial mechanics, it is well-known that resonance may cause highly complicated motions of asteroids.[1]

> *In rough terms, the complexity of phenomena in quantum field theory is caused by resonances.*

In Sect. 7.16, the non-resonance case and the resonance case were studied for linear operator equations. We now want to generalize this to nonlinear problems.

8.1.1 Non-Resonance

Consider the nonlinear operator equation

$$\boxed{H_0\varphi + \kappa(v_0 + V(\varphi)) = E\varphi, \qquad \varphi \in X.} \tag{8.1}$$

We make the following assumptions.

(A1) The complex Hilbert space X has the finite dimension $N = 1, 2, \ldots$

[1] This is described mathematically by KAM theory (Kolmogorov–Arnold–Moser theory). As an introduction, we recommend Scheck (2000), Vol. 1, and Thirring (1997).

(A2) The operator $H_0 : X \to X$ is linear and self-adjoint. Furthermore,

$$H_0|E_j^0\rangle = E_j^0|E_j^0\rangle, \qquad j = 1, \ldots, N.$$

Here, the energy eigenstates $|E_1^0\rangle, \ldots, |E_N^0\rangle$ form a complete orthonormal system of X.

(A3) We set $V(\varphi) := W(\varphi, \varphi, \varphi)$ for all $\varphi \in X$, where the given operator $W : X \times X \times X \to X$ is linear in each argument. For example, we may choose $V(\varphi) := \langle\varphi|\varphi\rangle\varphi$.

(A4) We are given the complex constant κ called the coupling constant, and we are given the fixed element v_0 of the space X.

We are looking for an element φ of X.

Theorem 8.1 *Suppose that we are given the complex number E different from the energy values $E_1^0, \ldots, E_N^0$. Then, there exist positive numbers κ_0 and r_0 such that, for each given coupling constant κ with $|\kappa| \le \kappa_0$, equation* (8.1) *has precisely one solution $\varphi \in X$ with $||\varphi|| \le r_0$.*

Proof. Equation (8.1) is equivalent to

$$\varphi = -\kappa(H_0 - EI)^{-1}(v_0 + V(\varphi)), \qquad \varphi \in X.$$

The statement follows now from the Banach fixed-point theorem in Sect. 7.13. □

In particular, the solution φ can be computed by using the following iterative method

$$\varphi_{n+1} = -\kappa(H_0 - EI)^{-1}(v_0 + V(\varphi_n)), \qquad n = 0, 1, \ldots \tag{8.2}$$

with $\varphi_0 := 0$. This method converges to φ as $n \to \infty$ in the Hilbert space X. For the first approximation, we get

$$\varphi_1 = -\kappa(H_0 - EI)^{-1}v_0 = \kappa \sum_{j=1}^{N} \frac{|E_j^0\rangle\langle E_j^0|v_0\rangle}{E - E_j^0}. \tag{8.3}$$

Let us discuss this.

(N) The non-resonance case (regular case). The expression (8.3) makes sense, since we assume that the parameter E is different from the eigenvalues $E_1^0, \ldots, E_N^0$. We say that the value E is not in resonance with the eigenvalues $E_1^0, \ldots, E_N^0$. Then, the Green's operator $(H_0 - EI)^{-1}$ is well-defined. Explicitly,

$$(H_0 - EI)^{-1} = \sum_{j=1}^{N} \frac{|E_j^0\rangle\langle E_j^0|}{E_j^0 - E}.$$

(R) The resonance case (singular case). The situation changes completely if we choose

$$E := E_1^0.$$

Here, we say that the value E is in resonance with the eigenvalue E_1^0. Then, the Green's operator $(H_0 - E_1^0 I)^{-1}$ does not exist, and the iterative method (8.2) above fails completely. As a rule, φ_1 is an infinite quantity. Furthermore, if we set

$$E := E_1^0 + \varepsilon, \qquad \varepsilon \neq 0,$$

then we obtain

$$(H_0 - EI)^{-1} = \sum_{j=1}^{N} \frac{|E_j^0\rangle\langle E_j^0|}{E_j^0 - E_1^0 - \varepsilon}.$$

Since

$$\lim_{\varepsilon \to +0} \frac{1}{E_1^0 - E} = -\lim_{\varepsilon \to +0} \frac{1}{\varepsilon} = -\infty,$$

some of the expressions arising from perturbation theory become very large if the perturbation ε is very small.

Summarizing, it turns out that

Naive perturbation theory fails completely in the resonance case.

This situation is typical for the naive use of perturbation theory in quantum field theory. In what follows, we will show how to obtain a rigorous result. To this end, we will replace the naive iterative method (8.2) above by the rigorous, more sophisticated iterative method (8.12) below.

8.1.2 Resonance, Regularizing Term, and Bifurcation

Set $E := E_1^0 + \varepsilon$. Consider the nonlinear operator equation

$$\boxed{H_0\varphi + \kappa V(\varphi) = E\varphi, \qquad \varphi \in X.} \tag{8.4}$$

In addition to (A1) through (A4) above, we assume that the energy eigenvalue E_1^0 is simple, that is, the eigenvectors to E_1^0 have the form $\varrho|E_1^0\rangle$ where ϱ is an arbitrary nonzero complex number. We are looking for a solution (φ, E) of (8.4) with $\varphi \in X$ and $E \in \mathbb{C}$. The proof of the following theorem will be based on the use of regularizing terms.

Theorem 8.2 *There exist positive constants κ_0, s_0, η_0 and r_0 such that for given complex parameters κ and s with*

$$|\kappa| \le \kappa_0, \quad 0 < |s| \le s_0,$$

equation (8.4) *has precisely one solution φ, E which satisfies the normalization condition*

$$\langle E_1^0|\varphi\rangle = s$$

and the smallness conditions $|E - E_1^0| \le \eta_0$ and $||\varphi|| \le r_0$.

Before proving this, let us discuss the physical meaning of this result. We will show below that the zeroth approximation of the solution looks like

$$\varphi = s|E_1^0\rangle, \qquad E = E_1^0.$$

The first approximation of the energy is given by

$$E = E_1^0 + \kappa s^2 \langle G_{\mathrm{reg}} V(\psi_1)|\psi_1\rangle$$

where we set $\psi_1 := |E_1^0\rangle$. Observe the following point which is crucial for understanding the phenomenon of renormalization in physics.

From the mathematical point of view, we obtain a branch of solutions which depends on the parameter s.
The free parameter s has to be determined by physical experiments.

Let us discuss this. Suppose that we measure the

- energy E and
- the running coupling constant κ.

We then obtain the approximation

$$\kappa = \frac{E - E_1^0}{s^2 \langle G_{\rm reg} V(\psi_1) | \psi_1 \rangle}.$$

This tells us the value of the parameter s. This phenomenon is typical for renormalization in quantum field theory. The energy E_1^0 is called the bare energy. However, this bare energy is not a relevant physical quantity. In a physical experiment we do not measure the bare energy E_1^0, but the energy E and the running coupling constant κ. In elementary particle physics, this corresponds to the fact that the rest energy of an elementary particle (e.g., an electron) results from complex interaction processes. Therefore, the rest energy E differs from the bare energy E_1^0. In the present simple example, interactions are modelled by the nonlinear term $\kappa V(\varphi)$.

Proof of Theorem 8.2. (I) The resonance condition. To simplify notation, set $\psi_j := |E_j^0\rangle, j = 1, \ldots, N$. For given $\chi \in X$, consider the linear operator equation

$$H_0\varphi - E_1^0\varphi = \chi, \qquad \varphi \in X. \tag{8.5}$$

By Theorem 7.15 on page 376, this problem has a solution iff the so-called resonance condition

$$\langle \psi_1 | \chi \rangle = 0$$

is satisfied. The general solution is then given by

$$\varphi = s\psi_1 + \sum_{j=2}^{N} \frac{\langle \psi_j | \chi \rangle}{E_j^0 - E_1^0}\, \psi_j \tag{8.6}$$

where s is an arbitrary complex parameter.

(II) The regularized Green's operator $G_{\rm reg}$. Set $P\varphi := \langle \psi_1 | \varphi \rangle \psi_1$. The operator $P : X \to \mathrm{span}(\psi_1)$ projects the Hilbert space X orthogonally onto the 1-dimensional eigenvector space to the energy eigenvalue E_1^0. We now consider the modified equation

$$H_0\varphi + P\varphi - E_1^0\varphi = \chi, \qquad \varphi \in X. \tag{8.7}$$

Theorem 7.16 on page 377 tells us that, for each given $\chi \in X$, equation (8.7) has the unique solution

$$\varphi = (H_0 + P - E_1^0 I)^{-1}\chi.$$

We define $G_{\rm reg} := (H_0 + P - E_1^0 I)^{-1}$. Explicitly,[2]

$$\varphi = G_{\rm reg}\chi = \langle \psi_1 | \chi \rangle \psi_1 + \sum_{j=2}^{N} \frac{\langle \psi_j | \chi \rangle}{E_j^0 - E_1^0}\, \psi_j.$$

In particular, $G_{\rm reg}\psi_1 = \psi_1$. The term $P\varphi = \langle \psi_1 | \varphi \rangle \psi_1$ in (8.7) is called regularizing term.

(III) The trick of regularizing term. The original equation (8.4) on page 499 can be written equivalently as

[2] In fact, $(H_0 + P - E_1^0 I)\varphi$ is equal to $\langle \psi_1 | \chi \rangle \psi_1 + \sum_{j=2}^{N} \langle \psi_j | \chi \rangle \psi_j = \chi$.

$$\boxed{H_0\varphi - E_1^0\varphi + \kappa V(\varphi) + \langle\psi_1|\varphi\rangle\psi_1 = s\psi_1 + \varepsilon\varphi, \qquad \varphi \in X} \tag{8.8}$$

along with the normalization condition

$$\boxed{\langle\psi_1|\varphi\rangle = s.} \tag{8.9}$$

By (II), this is equivalent to the operator equation

$$\varphi = G_{\rm reg}(s\psi_1 - \kappa V(\varphi) + \varepsilon\varphi)$$

along with (8.9). Finally, since $G_{\rm reg}\psi_1 = \psi_1$, we obtain the equivalent operator equation

$$\varphi = s\psi_1 - \kappa G_{\rm reg}V(\varphi) + \varepsilon G_{\rm reg}\varphi \tag{8.10}$$

along with (8.9). We have to solve the system (8.9), (8.10). To this end, we will use both a rescaling and the Banach fixed-point theorem.

(IV) Rescaling. Set $\varphi := s(1+\varepsilon)\psi_1 + s\chi$. Equation (8.9) yields

$$s\langle\psi_1|\psi_1 + \varepsilon\psi_1 + \chi\rangle = s.$$

Since $s \neq 0$ and $\langle\psi_1|\psi_1\rangle = 1$, we get $\varepsilon = -\langle\psi_1|\chi\rangle$. Furthermore, it follows from (8.10) that

$$\begin{aligned} s(1+\varepsilon)\psi_1 + s\chi = s\psi_1 - \kappa s^3 G_{\rm reg}V((1+\varepsilon)\psi_1 + \chi) \\ + s\varepsilon(1+\varepsilon)\psi_1 + s\varepsilon G_{\rm reg}\chi. \end{aligned}$$

Consequently, the system (8.9), (8.10) corresponds to the following equivalent system

$$\begin{aligned} &\chi = A(\chi, \varepsilon, \kappa, s), \\ &\varepsilon = -\langle\psi_1|A(\chi, \varepsilon, \kappa, s)\rangle, \qquad \chi \in X, \;\; \varepsilon \in \mathbb{C} \end{aligned} \tag{8.11}$$

along with

$$A(\chi, \varepsilon, \kappa, s) := -\kappa s^2 G_{\rm reg}V((1+\varepsilon)\psi_1 + \chi) + \varepsilon G_{\rm reg}\chi + \varepsilon^2\psi_1.$$

(V) The Banach fixed-point theorem. The system (8.11) represents an operator equation on the Banach space $X \times \mathbb{C}$ with the norm

$$||(\chi, \varepsilon)|| := ||\chi|| + |\varepsilon|.$$

We are given the complex parameters s and κ with $0 < |s| \leq s_0$ and $|\kappa| \leq \kappa_0$ where $s_0 > 0$ and $\kappa_0 > 0$ are sufficiently small numbers. By the Banach fixed-point theorem in Sect. 7.13 on page 366, there exists a small closed ball $\mathcal{B}$ about the origin in the Banach space $X \times \mathbb{C}$ such that the operator equation (8.11) has a unique solution in the closed ball $\mathcal{B}$.

(V) Iterative method. By the Banach fixed-point theorem, the solution (χ, ε) of (8.11) can be computed by using the following iterative method

$$\begin{aligned} &\chi_{n+1} = A(\chi_n, \varepsilon_n, \kappa, s), \\ &\varepsilon_{n+1} = -\langle\psi_1|A(\chi_n, \varepsilon_n, \kappa, s)\rangle, \qquad n = 0, 1, 2, \ldots \end{aligned} \tag{8.12}$$

with $\chi_0 := 0$ and $\varepsilon_0 := 0$. This method converges in the Banach space $X \times \mathbb{C}$. In particular, we get

$$\chi_1 = -\kappa s^2 G_{\rm reg}V(\psi_1), \qquad \varepsilon_1 = \kappa s^2\langle\psi_1|G_{\rm reg}V(\psi_1)\rangle.$$

□

Bifurcation. On the product space $X \times \mathbb{C}$, the original nonlinear problem (8.4) on page 499 has two different solution curves, namely,

- the trivial solution curve $\varphi = 0, E =$ arbitrary complex number,
- and the nontrivial solution curve $(\varphi = \varphi(s,\kappa), E = E(s,\kappa))$ given by Theorem 8.2 on page 499.

The two curves intersect each other at the point $\varphi = 0, E = E_1^0$. Therefore, we speak of bifurcation. The nontrivial solution branch of equation (8.4) represents a perturbation of the curve

$$\varphi = s\psi_1, \quad E = E_1^0, \qquad s \in \mathbb{C}$$

which corresponds to the linearized problem $H_0\varphi = E_1^0\varphi$. Bifurcation theory is part of nonlinear functional analysis. A detailed study of the methods of bifurcation theory along with many applications in mathematical physics and mathematical biology can be found in Zeidler (1986).

8.1.3 The Renormalization Group

The method of renormalization group plays a crucial role in modern physics.

> *Roughly speaking, this method studies the behavior of physical effects under the rescaling of typical parameters.*

We are going to study a very simplified model for this. Let $(\varphi(s,\kappa), E(s,\kappa))$ be the solution of the original equation (8.4) on page 499, that is,

$$H_0\varphi(s,\kappa) + \kappa V[\varphi(s,\kappa)] = E(s,\kappa)\varphi(s,\kappa)$$

along with $\langle\psi_1|\varphi(s,\kappa)\rangle = s$. Choose the fixed real number $\lambda > 0$. Replacing $s \mapsto \lambda s$ and $\kappa \mapsto \frac{\kappa}{\lambda^2}$, we get

$$H_0\varphi\left(\lambda s, \frac{\kappa}{\lambda^2}\right) + \frac{\kappa}{\lambda^2} V\left[\varphi\left(\lambda s, \frac{\kappa}{\lambda^2}\right)\right] = E\left(\lambda s, \frac{\kappa}{\lambda^2}\right)\varphi\left(\lambda s, \frac{\kappa}{\lambda^2}\right)$$

along with $\langle\psi_1|\varphi\left(\lambda s, \frac{\kappa}{\lambda^2}\right)\rangle = \lambda s$. Define

$$\psi(s,\kappa) := \frac{1}{\lambda}\cdot\varphi\left(\lambda s, \frac{\kappa}{\lambda^2}\right).$$

Noting that $V(\lambda\psi) = \lambda^3 V(\psi)$, we obtain

$$H_0\psi(s,\kappa) + \kappa V[\psi(s,\kappa)] = E\left(\lambda s, \frac{\kappa}{\lambda^2}\right)\psi(s,\kappa)$$

along with $\langle\psi_1|\psi(s,\kappa)\rangle = s$. By the uniqueness statement from Theorem 8.2 on page 499, we get

$$\boxed{\frac{1}{\lambda}\cdot\varphi\left(\lambda s, \frac{\kappa}{\lambda^2}\right) = \varphi(s,\kappa)} \tag{8.13}$$

along with

$$\boxed{E\left(\lambda s, \frac{\kappa}{\lambda^2}\right) = E(s,\kappa)} \tag{8.14}$$

for all nonzero complex parameters s and κ in a sufficiently small neighborhood of the origin.

Summarizing, the homogeneity of the potential, $V(\lambda\varphi) = \lambda^3 V(\varphi)$, *implies the symmetries* (8.13), (8.14) *of the solution branch.*

Differentiating equation (8.13) with respect to the parameter λ, and setting $\lambda = 1$, we obtain

$$\boxed{\varphi(s,\kappa) - s\varphi_s(s,\kappa) + 2\kappa\varphi_\kappa(s,\kappa) = 0.} \tag{8.15}$$

In our model, the differential equation (8.15) can be regarded as a simplified version of the Callan–Szymanzik equation in quantum field theory.

Let $\mathbb{R}_+^\times$ denote the set of all positive real numbers; that is, $x \in \mathbb{R}_+^\times$ iff $x > 0$. For each parameter $\lambda \in \mathbb{R}_+^\times$, define the map $T_\lambda : \mathbb{C}^2 \to \mathbb{C}^2$ given by

$$T_\lambda(s,\kappa) := \left(\lambda s, \frac{\kappa}{\lambda^2}\right).$$

For all parameters $\lambda, \mu \in \mathbb{R}_+^\times$,

$$T_{\lambda\mu} = T_\lambda T_\mu.$$

Therefore, the family $\{T_\lambda\}_{\lambda\in\mathbb{R}_+^\times}$ of all operators T_λ forms a group. This group is called the renormalization group of the original operator equation (8.4) on page 499.

8.1.4 The Main Bifurcation Theorem

Let us now study the general case of the nonlinear equation

$$\boxed{H_0\varphi + \kappa V(\varphi) = E\varphi, \qquad \varphi \in X} \tag{8.16}$$

where the eigenvalue E_1^0 of the linearized problem

$$H_0\varphi = E_1^0\varphi$$

is not simple as in Sect. 8.1.2, but it has general multiplicity. To this end, we will reduce the problem to the nonlinear system (8.17) below. We make the following assumptions.

(A1) The complex Hilbert space X has the finite dimension $N = 1, 2, \ldots$

(A2) Linear operator: The operator $H_0 : X \to X$ is linear and self-adjoint. Furthermore,

$$H_0|E_j^0\rangle = E_j^0|E_j^0\rangle, \qquad j = 1, \ldots, N.$$

Here, the energy eigenstates $|E_1^0\rangle, \ldots, |E_N^0\rangle$ form a complete orthonormal system of X.

(A3) Multiplicity: The eigenvalue E_1^0 has the multiplicity m, that is, the eigenvectors $|E_1^0\rangle, \ldots, |E_m^0\rangle$ form a basis of the eigenspace of H_0 to the eigenvalue E_1^0. Let $1 \le m < N$. To simplify notation, set $\psi_j := |E_j^0\rangle$. Define the orthogonal projection operator $P : X \to X$ by setting

$$P\varphi := \sum_{j=1}^{m} \langle\psi_j|\varphi\rangle\psi_j \qquad \text{for all} \quad \varphi \in X.$$

(A4) Nonlinearity: We set $V(\varphi) := W(\varphi,\varphi,\varphi)$ for all $\varphi \in X$, where the given operator $W : X \times X \times X \to X$ is linear in each argument. For example, $V(\varphi) := \langle\varphi|\varphi\rangle\varphi$.

(A5) Resonance condition: The nonlinear equation[3]

$$\boxed{\sigma = \kappa PV(\sigma), \qquad \sigma \in PX,\ \kappa \in \mathbb{C}} \tag{8.17}$$

has a solution (σ_0, κ_0) where $\sigma_0 \neq 0$ and $\kappa_0 \neq 0$. This solution is regular, that is, the linearized equation

$$h = \kappa_0 P \cdot V'(\sigma_0)h, \qquad h \in X \tag{8.18}$$

has only the trivial solution $h = 0$.

Theorem 8.3 *There exists a number $\alpha_0 > 0$ such that for each given complex number α with $|\alpha| \leq \alpha_0$, the nonlinear problem* (8.16) *with the coupling constant κ_0 has a solution*

$$\varphi = \alpha\sigma_0 + O(\alpha^2), \qquad E = E_1^0 + \alpha^2, \qquad \alpha \to 0.$$

Proof. (I) The regularized Green's operator. For all $\chi \in X$, define

$$G_{\mathrm{reg}}\chi := \sum_{j=1}^{m} \langle\psi_j|\chi\rangle\psi_j + \sum_{j=m+1}^{N} \frac{\langle\psi_j|\chi\rangle}{E_j^0 - E_1^0}\,\psi_j.$$

Suppose that we are given $\chi \in X$ with $P\chi = 0$. By Theorem 7.16 on page 377, the equation

$$H_0\varrho - E_1^0\varrho = \chi$$

has precisely one solution $\varrho \in X$ with $P\varrho = 0$. This solution is given by

$$\varrho = G_{\mathrm{reg}}\chi = \sum_{j=m+1}^{N} \frac{\langle\psi_j|\chi\rangle}{E_j^0 - E_1^0}\,\psi_j.$$

(II) Equivalent system. Set $E := E_1^0 + \varepsilon$, and introduce the orthogonal projection operator $Q := I - P$. Explicitly,

$$Q\varphi = \sum_{j=m+1}^{N} \langle\psi_j|\varphi\rangle\psi_j \qquad \text{for all} \quad \varphi \in X.$$

Then, the original nonlinear problem (8.16) is equivalent to

$$\begin{aligned} Q(H_0\varphi - (E_1^0 + \varepsilon)\varphi + \kappa V(\varphi)) &= 0, \\ P(H_0\varphi - (E_1^0 + \varepsilon)\varphi + \kappa V(\varphi)) &= 0. \end{aligned} \tag{8.19}$$

The idea of the following proof is

[3] Set $\sigma := s_1\psi_1 + \ldots + s_m\psi_m$. Equation (8.17) is then equivalent to the system

$$g_j(s_1, \ldots, s_m; \kappa) = 0 \qquad j = 1, \ldots, m, \qquad s_1, \ldots, s_m \in \mathbb{C}.$$

Here, $g_j(s_1, \ldots, s_m; \kappa) := s_j - \kappa \cdot \langle\psi_j|V(s_1\psi_1 + \ldots + s_m\psi_m)\rangle$. Condition (8.18) means that

$$\det\left(\frac{\partial g_j(s_{01}, \ldots, s_{0m}; \kappa_0)}{\partial s_k}\right)|_{j,k=1,\ldots,m} \neq 0.$$

(i) to solve the first equation from (8.19) by the Banach fixed-point theorem,
(ii) to insert the solution from (i) into the second equation from (8.19), and
(iii) to solve the resulting equation by using the implicit function theorem near the solution (σ_0, κ_0) of equation (8.17).

For each $\varphi \in X$, define $\chi := P\varphi$ and $\varrho := Q\varphi$. Then

$$\varphi = \chi + \varrho, \qquad \chi \in PX,\ \varrho \in QX.$$

Therefore, system (8.19) is equivalent to

$$\begin{aligned} Q\{H_0(\chi+\varrho) - (E_1^0+\varepsilon)(\chi+\varrho) + \kappa V(\chi+\varrho)\} &= 0, \\ P\{H_0(\chi+\varrho) - (E_1^0+\varepsilon)(\chi+\varrho) + \kappa V(\chi+\varrho)\} &= 0. \end{aligned} \tag{8.20}$$

Observe that $\langle \psi_j | H_0\varphi - E_1^0\varphi\rangle = \langle H_0\psi_j - E_1^0\psi_j | \varphi\rangle = 0$ for $j = 1, \ldots, m$. Hence

$$P(H_0 - E_1^0 I) = 0.$$

Furthermore, $P\chi = \chi, Q\chi = 0$ and $Q\varrho = \varrho, P\varrho = 0$. Thus, choosing the coupling constant $\kappa := \kappa_0$ and recalling that $Q := I - P$, the system (8.20) is equivalent to

$$\begin{aligned} H_0\varrho - E_1^0\varrho &= \varepsilon\varrho - \kappa_0 QV(\chi+\varrho), \\ \varepsilon\chi &= \kappa_0 PV(\chi+\varrho). \end{aligned} \tag{8.21}$$

Finally, using the regularized Green's operator, this system is equivalent to the equation

$$\boxed{\varrho = \varepsilon G_{\mathrm{reg}}\varrho - \kappa_0 G_{\mathrm{reg}} QV(\chi+\varrho)} \tag{8.22}$$

along with

$$\boxed{\varepsilon\chi = \kappa_0 PV(\chi+\varrho).} \tag{8.23}$$

(III) Rescaling. We set $\chi := \alpha\sigma$ and $\varepsilon := \alpha^2$. Equation (8.22) passes then over to

$$\varrho = \alpha^2 G_{\mathrm{reg}}\varrho - \kappa_0 G_{\mathrm{reg}} QV(\alpha\sigma + \varrho). \tag{8.24}$$

(IV) The Banach fixed-point theorem. By Theorem 7.12 on page 367, there exist positive parameters α_0, β_0 and r_0 such that for given $\alpha \in \mathbb{C}$ and $\sigma \in PX$ with

$$|\alpha| \le \alpha_0, \quad ||\sigma|| \le \beta_0,$$

equation (8.24) has precisely one solution $\varrho \in QX$ with $||\varrho|| \le r_0$. This solution will be denoted by

$$\varrho = \varrho(\alpha, \sigma).$$

By the analytic form of the implicit function theorem,[4] the components of $\varrho(\alpha, \sigma)$ depend holomorphically on the complex parameter α. The iterative method

$$\varrho_{n+1} = \alpha^2 G_{\mathrm{reg}}\varrho_n - \kappa_0 G_{\mathrm{reg}} QV(\alpha\sigma + \varrho_n), \qquad n = 0, 1, \ldots$$

with $\varrho_0 := 0$ (or comparison of coefficients) shows that

[4] This can be found in Zeidler (1986), Vol. I, Sect. 8.3.

$$\varrho(\alpha,\sigma) = -\alpha^3\kappa_0 G_{\mathrm{reg}} QV(\sigma) + O(\alpha^4), \qquad \alpha \to 0.$$

(V) The bifurcation equation. Inserting $\varrho(\alpha,\sigma)$ into equation (8.23), we get

$$\alpha^3\sigma = \kappa_0 PV(\alpha\sigma + \varrho(\alpha,\sigma)).$$

Dividing this by α^3, we obtain the so-called bifurcation equation

$$\sigma = \kappa_0 PV(\sigma) + O(\alpha), \qquad \alpha \to 0. \tag{8.25}$$

For $\alpha = 0$, this equation has the solution $\sigma = \sigma_0$, by assumption (A5). Choose $h \in X$. Differentiating the equation

$$\sigma_0 + th = \kappa_0 PV(\sigma_0 + th)$$

with respect to the real parameter t at $t = 0$, we get

$$h = \kappa_0 P \cdot V'(\sigma_0)h. \tag{8.26}$$

This is the linearization of (8.25) at the point $\sigma = \sigma_0, \alpha = 0$. By assumption (A5), equation (8.26) has only the trivial solution $h = 0$. By the implicit function theorem, the bifurcation equation (8.25) has a solution of the form

$$\sigma = \sigma_0 + O(\alpha), \qquad \alpha \to 0.$$

□

Modification. If the resonance condition (A5) above is satisfied for the modified equation

$$\sigma = -\kappa PV(\sigma), \qquad \sigma \in PX,\ \kappa \in \mathbb{C},$$

then Theorem 8.3 remains true if we replace $E = E_1^0 + \alpha^2$ by $E = E_1^0 - \alpha^2$.

8.2 The Rellich Theorem

Let X be a complex Hilbert space of finite dimension $N = 1, 2, \ldots$. Consider the eigenvalue equation

$$A\varphi = \lambda\varphi, \qquad \lambda \in \mathbb{R},\ \ \varphi \in X \setminus \{0\}$$

along with the perturbed problem

$$\boxed{A(\varepsilon)\varphi(\varepsilon) = \lambda(\varepsilon)\varphi(\varepsilon), \qquad \lambda(\varepsilon) \in \mathbb{R},\ \ \varphi(\varepsilon) \in X \setminus \{0\}}$$

where ε is a small real perturbation parameter, and $A(0) = A$. We assume the following.

(H1) The linear operator $A : X \to X$ is self-adjoint.

(H2) There exists an open neighborhood $U(0)$ of the origin of the real line such that for each $\varepsilon \in U(0)$, the operator $A(\varepsilon) : X \to X$ is linear and self-adjoint, and it depends holomorphically on the parameter ε. Explicitly,

$$A(\varepsilon) = A + \varepsilon A_1 + \varepsilon^2 A_2 + \ldots$$

This means that for each arbitrary, but fixed basis $|1\rangle, \ldots, |N\rangle$ of the space X, all of the matrix elements $\langle m|A(\varepsilon)|n\rangle$ are power series expansions which are convergent for all real parameters $\varepsilon \in U(0)$.

By the principal axis theorem, each operator $A(\varepsilon)$ with $\varepsilon \in U(0)$ possesses a complete orthonormal system of eigenvectors with real eigenvalues.

Theorem 8.4 *There exists a small neighborhood of the origin $V(0)$ of the real line such that the eigenvalues and eigenvectors of the operator $A(\varepsilon)$ depend holomorphically on the real parameter $\varepsilon \in V(0)$.*

Explicitly, this means the following. Let φ be an eigenvector of multiplicity m of the operator A with eigenvalue $\lambda \in \mathbb{R}$. Then, there exist power series expansions

$$\begin{aligned} \lambda_j(\varepsilon) &= \lambda + \varepsilon\lambda_{j1} + \varepsilon^2\lambda_{j2} + \dots, \\ \varphi_j(\varepsilon) &= \varphi + \varepsilon\varphi_{j1} + \varepsilon^2\varphi_{j2} + \dots, \qquad j = 1, 2, \dots, m \end{aligned}$$

which converge for all $\varepsilon \in V(0)$, and which are eigensolutions of $A(\varepsilon)$.[5] In addition, there exists a number $\delta > 0$ such that $\lambda_1(\varepsilon), \dots, \lambda_m(\varepsilon)$ are the only eigenvalues of $A(\varepsilon)$ which lie in the interval $]\lambda - \delta, \lambda + \delta[$.

Theorem 8.4 is the special case of a general theorem due to Rellich which is valid for a broad class of self-adjoint operators in Hilbert spaces. The proof can be found in Riesz and Nagy (1978), Sect. 136.

8.3 The Trotter Product Formula

Theorem 8.5 *Let $A, B : X \to X$ be linear operators on the finite-dimensional Hilbert space X. Then*

$$e^{A+B} = \lim_{N\to\infty} (e^{A/N} e^{B/N})^N.$$

Proof. Set $C := \mathrm{e}^{(A+B)/N}$ and $D := \mathrm{e}^{A/N}\mathrm{e}^{B/N}$. Then

$$C - D = \frac{1}{N^2} \sum_{m=2}^{\infty} \frac{1}{N^{m-2}} \left(\frac{(A+B)^m}{m!} - \sum_{k+l=m} \frac{A^k B^l}{k!l!} \right),$$

since the terms for $m = 1$ cancel each other. Using

$$||(A+B)^m|| \le (||A|| + ||B||)^m, \qquad ||A^k B^l|| \le ||A||^k ||B||^l,$$

we get

$$||C - D|| = O\left(\frac{1}{N^2}\right) \qquad N \to \infty.$$

Moreover, $\max\{||C||, ||D||\} \le \mathrm{e}^{(||A||+||B||)/N}$. It follows from the identity

$$C^N - D^N = \sum_{k=0}^{N-1} C^k (C - D) D^{N-1-k}$$

that $||C^N - D^N|| \le N||C - D|| \cdot \mathrm{e}^{(||A||+||B||)(N-1)/N}$. Hence

$$||C^N - D^N|| = O\left(\frac{1}{N}\right), \qquad N \to \infty.$$

This implies $\lim_{N\to+\infty} C^N - D^N = 0$. □

[5] The convergence of $\varphi_j(\varepsilon)$ refers to the components.

8.4 The Magic Baker–Campbell–Hausdorff Formula

Let $A, B : X \to X$ be linear operators on the complex finite-dimensional Hilbert space X with $AB = BA$. Then

$$\mathrm{e}^A \mathrm{e}^B = \mathrm{e}^{A+B}.$$

However, the commutation relation $AB = BA$ is frequently violated in mathematics and physics. We then have to use the following Baker–Campbell–Hausdorff formula

$$\boxed{\mathrm{e}^A \mathrm{e}^B = \mathrm{e}^{A+B+\frac{1}{2}[A,B]_- + r(A,B)}} \tag{8.27}$$

if the operator norms $||A||$ and $||B||$ are sufficiently small. Again, we see the Lie product $[A, B]_- := AB - BA$. The remainder $r(A, B)$ has the form

$$r(A, B) = \sum_{k=3}^{\infty} p_k(A, B) \tag{8.28}$$

where $p_k(A, B)$ is a polynomial of order k of the variables A, B with respect to the Lie product $[., .]_-$. The coefficients of $p_k(A, B)$ are rational numbers. For example,

$$\begin{aligned} 12p_3(A, B) &= A \cdot (A \cdot B) + B \cdot (B \cdot A), \\ 24p_4(A, B) &= B \cdot (A \cdot (B \cdot A)) \end{aligned}$$

where we write $A \cdot B$ instead of $[A, B]_-$. The point is that the exponent

$$A + B + \tfrac{1}{2}[A, B]_- + r(A, B)$$

lies in the Lie algebra generated by the operators A and B. Thus, the generalized addition theorem (8.27) for the exponential function leads us in a natural way to the concept of Lie algebra.

Theorem 8.6 *Let $A, B : X \to X$ be linear operators on the finite-dimensional Hilbert space X. Then there exists a number $r > 0$ such that* (8.27), (8.28) *hold true if $||A|| \le r$ and $||B|| \le r$.*

Formula (8.27) is named after contributions made independently by Campbell, Baker, and Hausdorff around 1900. In 1950 Dynkin discovered the following explicit formula:

$$\begin{aligned} r(A, B) = \sum_{k=2}^{\infty} \frac{(-1)^k}{k+1} \sum{}' \frac{1}{l_1 + \ldots + l_k + 1} \\ \times \left(\frac{(\mathrm{ad}\, A)^{l_1}}{l_1!} \frac{(\mathrm{ad}\, B)^{m_1}}{m_1!} \cdots \frac{(\mathrm{ad}\, A)^{l_k}}{l_k!} \frac{(\mathrm{ad}\, B)^{m_k}}{m_k!} \right) (A). \end{aligned}$$

Here, we use the operator $\mathrm{ad} : L(X) \to L(X)$ given on the space $L(X)$ of linear operators on X. Explicitly, for each $C \in L(X)$, the linear operator $\mathrm{ad}\, C : X \to X$ is given by

$$(\mathrm{ad}\, C)D := [C, D]_- \qquad \text{for all} \quad D \in L(X).$$

The sum $\sum'$ refers to all integers $l_1, \ldots, l_k \ge 0$ and $m_1, \ldots, m_k \ge 0$ with $m_j + l_j > 0$ for all j. The proof can be found in Duistermaat and Kolk (2000), p. 30.

8.5 Regularizing Terms

The naive use of perturbation theory in quantum field theory leads to divergent mathematical expressions. In order to extract finite physical information from this, physicists use the method of renormalization. In Volume II we will study quantum electrodynamics. In this setting, renormalization can be understood best by proceeding as follows.

(i) Put the quantum system in a box of finite volume V.
(ii) Consider a finite lattice in momentum space of grid length Δp and maximal momentum $P_{\max}$.

The maximal momentum corresponds to the choice of a maximal energy, $E_{\max}$. We then have to carry out the limits

$$V \to +\infty, \quad E_{\max} \to +\infty, \quad \Delta p \to 0.$$

Unfortunately, it turns out that the naive limits do not always exist. Sometimes divergent expressions arise.

The idea of the method of regularizing terms is to force convergence of divergent expressions by introducing additional terms. This technique is well-known in mathematics. In what follows we will study three prototypes, namely,

- the construction of entire functions via regularizing factors (the Weierstrass product theorem),
- the construction of meromorphic functions via regularizing summands (the Mittag–Leffler theorem), and
- the regularization of divergent integrals by adding terms to the integrand via Taylor expansion.

In this monograph, we distinguish between

- regularizing terms and
- counterterms.

By convention, regularizing terms are mathematical objects which give divergent expressions a well-defined rigorous meaning. Counterterms are added to Lagrangian densities in order to construct regularizing terms. Roughly speaking, this allows us a physical interpretation of the regularizing terms. In quantum field theory, renormalization theory is based on counterterms.

8.5.1 The Weierstrass Product Theorem

Recall that by an entire function, we mean a function $f : \mathbb{C} \to \mathbb{C}$ which is holomorphic on the complex plane. The entire function f has no zeros iff there exists an entire function $g : \mathbb{C} \to \mathbb{C}$ such that

$$f(z) = \mathrm{e}^{g(z)} \qquad \text{for all} \quad z \in \mathbb{C}.$$

Suppose that the function f is a polynomial which has the zeros $z_0, z_1, \dots z_m$ with the multiplicities $n_0, \dots, n_m$, respectively, where $z_0 := 0$ and $z_j \neq 0$ if $j = 1, \dots, m$. Then

$$f(z) = a z^{n_0} \prod_{k=1}^{m} \left(1 - \frac{z}{z_k}\right)^{n_k} \qquad \text{for all} \quad z \in \mathbb{C}. \tag{8.29}$$

Here, a is a complex number. If $z = 0$ is not a zero of f, then the factor z^{n_0} drops out. Now consider the case where the function f has an infinite number of zeros. The key formula reads as

$$\boxed{f(z) = \mathrm{e}^{g(z)} z^{n_0} \prod_{k=1}^{\infty} \left(1 - \frac{z}{z_k}\right)^{n_k} \mathrm{e}^{p_k(z)} \qquad \text{for all} \quad z \in \mathbb{C}.} \tag{8.30}$$

The point is that the naive generalization of (8.29) fails, but we have to add the regularizing factors $\mathrm{e}^{p_k(z)}$ which force the convergence of the product.

Theorem 8.7 *Let $f : \mathbb{C} \to \mathbb{C}$ be an entire function which has an infinite number of zeros $z_0, z_1, \ldots$ ordered by modulus, $|z_0| < |z_1| < \ldots$ with $z_0 := 0$. Let n_k be the multiplicity of the zero z_k. Then, there exist polynomials $p_1, p_2, ..$ and an entire function g such that the product formula* (8.30) *holds true.*

This classical theorem is due to Weierstrass (1815–1897). The proof can be found in Remmert (1998), Sect. 3.1.

8.5.2 The Mittag–Leffler Theorem

We want to generalize the decomposition into partial fractions from rational functions to meromorphic functions. As prototypes, let us consider the two functions

$$f(z) := \frac{2z}{(z - \mathrm{i})(z + \mathrm{i})} = \frac{A_-}{z - i} + \frac{A_+}{z + \mathrm{i}}$$

with $A_\pm = \lim_{z \to \pm \mathrm{i}} f(z)(z \pm \mathrm{i}) = 1$, and

$$\pi \cot \pi z = \frac{\pi \cos \pi z}{\sin \pi z}.$$

The function $z \mapsto \sin \pi z$ has precisely the zeros $z_k := k$ with $k = 0, \pm 1, \pm 2, \ldots$ Since $\lim_{z \to z_k} \pi (z - z_k) \cot \pi z = 1$, we get the representation

$$\pi \cot \pi z = \frac{1}{z - z_k} + g_k(z), \qquad k = 0, \pm 1, \pm 2, \ldots$$

for all z different from z_k in a sufficiently small neighborhood of the point z_k. The function g_k is locally holomorphic at the point z_k. Thus, the given function $z \mapsto \cot \pi z$ has a pole of first order at each point z_k with the principal part $1/(z - z_k)$. Motivated by the decomposition into partial fractions of the function f, we make the ansatz[6]

$$\pi \cot \pi z = \sum_{k=-\infty}^{\infty} \frac{1}{z - z_k}.$$

However, this ansatz does not work, since the series is not convergent. We have to pass to the modified sum

$$\boxed{\pi \cot \pi z = \sum_{k=-\infty}^{\infty} \frac{1}{z - z_k} + C_k} \tag{8.31}$$

[6] The sum $\sum_{k=-\infty}^{\infty} \ldots$ stands for $\sum_{k=0}^{\infty} \ldots + \sum_{k=-1}^{-\infty} \ldots$

with the so-called regularizing terms $C_k := 1/k$ for $k = \pm 1, \pm 2, \ldots$ and $C_0 := 0$. These regularizing terms force the convergence of the series from (8.31) for all complex points z different from the critical points z_k with $k = 0, \pm 1, \pm 2, \ldots$ In 1748 Euler incorporated this formula in his *Introductio*.[7] Interestingly enough, the regularizing terms cancel if we combine the right terms with each other. Explicitly,

$$\pi \cot \pi z = z \sum_{k=-\infty}^{\infty} \frac{1}{z^2 - z_k^2} = \frac{1}{z} + \sum_{k=1}^{\infty} \frac{2z}{z^2 - z_k^2}.$$

A similar cancellation was observed by Brown and Feynman in 1952 when computing radiative corrections to Compton scattering.[8] Generally, such cancellations occur in renormalization theory for low energies if one takes the full set of possible Feynman diagrams into account.

Theorem 8.8 *Let $f : \mathbb{C} \to \mathbb{C}$ be a meromorphic function on the complex plane which has an infinite number of poles $z_0, z_1, \ldots$ ordered by modulus, $|z_0| < |z_1| < \ldots$ Let f_k denote the principal part of the function f at the pole z_k. Then, there exist polynomials $p_0, p_1, ..$ and an entire function g such that*

$$f(z) = g(z) + \sum_{k=0}^{\infty} f_k(z) - p_k(z)$$

for all complex numbers z different from $z_0, z_1, \ldots$

The polynomials p_k are called regularizing terms. This classical theorem is due to Mittag–Leffler (1846–1927). The proof can be found in Remmert (1998), Sect. 6.1.

8.5.3 Regularization of Divergent Integrals

Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function, and let ϱ be a real number. Consider the integral

$$E(R) := \int_{\varrho}^{R} f(x)dx.$$

Suppose that

$$\boxed{E(R) = a \ln R + g(R)} \tag{8.32}$$

for a fixed nonzero real number number a and all sufficiently large real real numbers R. In addition, suppose that the finite limit $\lim_{R\to+\infty} g(R)$ exists. In the classical sense,

$$\int_{\varrho}^{\infty} f(x)dx = \lim_{R\to+\infty} \int_{\varrho}^{R} f(x)dx = (\operatorname{sgn} a)\, \infty.$$

The regularized integral is defined by

$$\boxed{\operatorname{reg} \int_{\varrho}^{\infty} f(x)dx := \lim_{R\to+\infty} g(R).} \tag{8.33}$$

[7] A proof of this formula can be found in Remmert (1991), Sect. 11.2.

[8] L. Brown and R. Feynman, Radiative corrections to Compton scattering, Phys. Rev. **85**(2) (1952), 231–244.

This value is well-defined. In fact, suppose that there exists a second decomposition

$$\int_\varrho^R f(x)dx = a_1 \ln R + g_1(R)$$

where $a \neq a_1$. Then, $(a - a_1) \ln R + g(R) - g_1(R) = 0$. Letting $R \to +\infty$, we get a contradiction. Therefore, $a = a_1$ and $g = g_1$.

Regularizing terms. Suppose that the function f behaves asymptotically like

$$f(x) = \frac{a}{x} + O\left(\frac{1}{x^2}\right), \qquad x \to +\infty \tag{8.34}$$

where a is a nonzero real number. Let $\varrho > 0$. Then

$$\int_\varrho^R f(x)dx = a \ln R - a \ln \varrho + \int_\varrho^R \left(f(x) - \frac{a}{x}\right) dx.$$

This implies

$$\boxed{\operatorname{reg} \int_\varrho^\infty f(x)dx = -a \ln \varrho + \int_\varrho^\infty \left(f(x) - \frac{a}{x}\right) dx.} \tag{8.35}$$

The second integral is finite. The term $-\frac{a}{x}$ is called regularizing term.

Example. It follows from

$$\int_2^R \frac{dx}{x+1} = \ln(R+1) - \ln 3 = \ln R + \ln \frac{R+1}{R} - \ln 3$$

that

$$\operatorname{reg} \int_2^\infty \frac{dx}{x+1} = -\ln 3.$$

Physical interpretation. Regard $E(R)$ above as the energy of a quantum system on the interval $[\varrho, R]$. This energy is very large if the size of the system is very large. Such extremely large energies are not observed in physical experiments. Physicists assume that we only measure relative energies with respect to the ground state. In our model above, we measure $E(R) - a \ln R$. In the limit $R \to +\infty$, we get the regularized value $\operatorname{reg} E(\infty)$.

The method of subtracting infinities. Suppose that the integral $\int_\varrho^\infty f(x)dx$ is finite. For given nonzero real number b,

$$\int_\varrho^\infty (b + f(x))dx = (\operatorname{sgn} b)\, \infty.$$

Since

$$\lim_{R \to +\infty} \left(\int_\varrho^R (b + f(x))dx - \int_\varrho^R b dx \right) = \int_\varrho^R f(x)dx,$$

we define

$$\int_\varrho^\infty (b + f(x))dx - \int_\varrho^\infty b dx := \int_\varrho^\infty f(x)dx.$$

This procedure helps to cancel infinities in renormalization theory.

8.5.4 The Polchinski Equation

Suppose again that the function f has the asymptotic behavior given in (8.34). Then

$$\lim_{R\to+\infty} Rf(R) = a.$$

Consequently, the coefficient a of the regularizing term can be uniquely determined by using the equation

$$\boxed{\lim_{R\to+\infty} R\frac{d}{dR}\int_{\varrho}^{R}\left(f(x)-\frac{a}{x}\right)\,dx = 0.} \tag{8.36}$$

This is the prototype of the so-called Polchinski equation which plays an important role in modern renormalization theory based on the renormalization group. We will study this in a later volume. We also refer to J. Polchinski, Renormalization and effective Lagrangians, Nucl. Phys. B **231** (1984), 269–295.

9. Fermions and the Calculus for Grassmann Variables

> In 1844, Hermann Grassmann (1809–1877) emphasized the importance of the wedge product (Grassmann product) for geometry in higher dimensions. But his contemporaries did not understand him. Nowadays the wedge product is fundamental for modern mathematics (cohomology) and physics (fermions and supersymmetry).
>
> Folklore

Recall that we distinguish between bosons (elementary particles with integer spin like photons or mesons) and fermions (elementary particles with half-integer spin like electrons and quarks). The rigorous finite-dimensional approach from the preceding Chap. 7 refers to bosons. However, it is possible to extend this approach to fermions by replacing complex numbers by Grassmann variables. In this chapter, we are going to discuss this.

9.1 The Grassmann Product

Vectors. Let X be a complex linear space. For two elements φ and ψ of X, we define the Grassmann product $\varphi \wedge \psi$ by setting

$$(\varphi \wedge \psi)(f, g) := f(\varphi)g(\psi) - f(\psi)g(\varphi) \qquad \text{for all} \quad f, g \in X^d.$$

Recall that the dual space X^d consists of all linear functionals $f : X \to \mathbb{C}$. The map

$$\varphi \wedge \psi : X^d \times X^d \to \mathbb{C}$$

is bilinear and antisymmetric. Explicitly, for all $f, g, h \in X^d$ and all complex numbers α, β, we have

- $(\varphi \wedge \psi)(f, g) = -(\varphi \wedge \psi)(g, f)$;
- $(\varphi \wedge \psi)(f, \alpha g + \beta h) = \alpha(\varphi \wedge \psi)(f, g) + \beta(\varphi \wedge \psi)(f, h)$.

The two crucial properties of the Grassmann product are

- $\varphi \wedge \psi = -\psi \wedge \varphi$ (anticommutativity), and
- $(\alpha\varphi + \beta\chi) \wedge \psi = \alpha\varphi \wedge \psi + \beta\chi \wedge \psi$ (distributivity)

for all $\varphi, \psi, \chi \in X$ and all complex numbers α, β. If we write briefly $\varphi\psi$ instead of the wedge product $\varphi \wedge \psi$, then

- $\varphi\psi = -\psi\varphi$, and
- $(\alpha\varphi + \beta\chi)\psi = \alpha\varphi\psi + \beta\chi\psi$.

This implies the key relation

$$\boxed{\varphi^2 = 0 \qquad \text{for all} \quad \varphi \in X.}$$

Functionals. Dually, for $f, g \in X^d$, we define

$$(f \wedge g)(\varphi, \psi) := f(\varphi)g(\psi) - f(\psi)g(\varphi) \qquad \text{for all} \quad \varphi, \psi \in X.$$

The map $f \wedge g : X \times X \to \mathbb{C}$ is bilinear and antisymmetric.

9.2 Differential Forms

Dual basis. Let $b_1, \ldots, b_n$ be a basis of the complex linear space X. We define the linear functional $b^i : X \to \mathbb{C}$ by setting

$$b^i(\beta^1 b_1 + \ldots + \beta^n b_n) := \beta^i, \qquad i = 1, \ldots, n$$

for all complex numbers $\beta^1, \ldots, \beta^n$. We call $b^1, \ldots, b^n$ the dual basis to $b_1, \ldots, b_n$. The expressions

$$\sum_{i=1}^{n} \alpha_i b^i \qquad \text{and} \qquad \frac{1}{2} \sum_{i,j=1}^{n} \alpha_{ij} b^i \wedge b^j \tag{9.1}$$

are called 1-forms and 2-forms on X, respectively. Here, the coefficients $\alpha_1, \ldots, \alpha_n$ and $\alpha_{12}, \ldots$ are complex numbers with $\alpha_{ij} = -\alpha_{ji}$ for all i, j.

Terminology. In modern mathematics, one writes

$$dx^i \quad \text{instead of} \quad b^i.$$

Using this convention, the differential forms (9.1) are written as

$$\sum_{i=1}^{n} \alpha_i dx^i \qquad \text{and} \qquad \frac{1}{2} \sum_{i,j=1}^{n} \alpha_{ij} dx^i \wedge dx^j. \tag{9.2}$$

9.3 Calculus for One Grassmann Variable

Consider the set of all formal power series expansions

$$\alpha + \beta\eta + \gamma\eta^2 + \ldots \tag{9.3}$$

with respect to the variable η and complex coefficients $\alpha, \beta, \ldots$. Add the relations

$$\boxed{\eta^2 = 0}$$

and $\alpha\eta = \eta\alpha$ for all complex numbers α. This way, the expansion (9.3) reduces to $\alpha + \beta\eta$. This procedure allows us to define functions of the Grassmann variable η. For example, for each complex number α, we define

$$\mathrm{e}^{\alpha\eta} := 1 + \alpha\eta.$$

This is motivated by the formal power series expansion

$$\mathrm{e}^{\alpha\eta} = 1 + \alpha\eta + \tfrac{1}{2}(\alpha\eta)^2 + \ldots = 1 + \alpha\eta.$$

For all complex numbers α, β, we define

- derivative: $\frac{d}{d\eta}(\alpha + \beta\eta) := \beta$;
- integral: $\int(\alpha + \beta\eta)d\eta := \beta$.

The reader should note that in the case of Grassmann variables, derivative and integral coincide.

9.4 Calculus for Several Grassmann Variables

We now consider the set of all formal power series expansions

$$\alpha_0 + \alpha_1\eta_1 + \ldots + \alpha_n\eta_n + \alpha_{12}\eta_1\eta_2 + \ldots$$

with respect to the variables $\eta_1, \ldots, \eta_n$ and complex coefficients $\alpha_1, \alpha_2, \ldots$. We add the relations

$$\eta_i\eta_j = -\eta_j\eta_i, \qquad \alpha\eta_i = \eta_i\alpha, \qquad i, j = 1, \ldots, n, \quad \alpha \in \mathbb{C}.$$

This implies $\eta_i^2 = 0$ for all i.

- The left partial derivative $\frac{\partial_l}{\partial\eta_k} f(\eta_1, \ldots, \eta_n)$ is performed after moving the variable η_k to the left. For example,

$$\frac{\partial_l}{\partial\eta_2}(\eta_2\eta_1) = \eta_1$$

and

$$\frac{\partial_l}{\partial\eta_2}(\eta_1\eta_2) = \frac{\partial_l}{\partial\eta_2}(-\eta_2\eta_1) = -\eta_1.$$

In general we have the anticommutativity property

$$\boxed{\frac{\partial_l^2}{\partial\eta_1\partial\eta_2} = -\frac{\partial_l^2}{\partial\eta_2\partial\eta_1}.} \tag{9.4}$$

- Similarly, the right partial derivative $\frac{\partial_r}{\partial\eta_k} f(\eta_1, \ldots, \eta_n)$ is performed after moving the variable η_k to the right. For example,

$$\frac{\partial_r}{\partial\eta_2}(\eta_2\eta_1) = \frac{\partial_r}{\partial\eta_2}(-\eta_1\eta_2) = -\eta_1.$$

As for the left partial derivative, we have the anticommutativity property

$$\frac{\partial_r^2}{\partial\eta_1\partial\eta_2} = -\frac{\partial_r^2}{\partial\eta_2\partial\eta_1}. \tag{9.5}$$

- By definition, the integral

$$\boxed{\int f(\eta_1, \ldots, \eta_n)d\eta_1 \cdots d\eta_n := c}$$

is equal to the coefficient c of $\eta_n\eta_{n-1}\cdots\eta_1$ in the expansion of f,

$$f(\eta_1, \ldots, \eta_n) = c(\eta_n\eta_{n-1}\cdots\eta_1) + \ldots.$$

For example, $\eta_1\eta_2 = -\eta_2\eta_1$ implies

$$\int \eta_1\eta_2 d\eta_1 d\eta_2 = -1.$$

9.5 The Determinant Trick

Gaussian integrals play a fundamental role in the functional integral approach to the Standard Model in particle physics. We want to generalize the formula

$$\int \mathrm{e}^{\zeta\alpha\eta} d\eta d\zeta = \alpha \qquad \text{for all} \quad \alpha \in \mathbb{C}$$

with respect to the Grassmann variables η, ζ to $2n$ variables. To this end, let $\eta_1, \ldots, \eta_n, \zeta_1, \ldots, \zeta_n$ variables which satisfy the following relations

$$\eta_i \eta_j = -\eta_j \eta_i, \qquad \zeta_i \zeta_j = -\zeta_j \zeta_i, \qquad \eta_i \zeta_j = -\zeta_j \eta_i + \gamma \delta_{ij}$$

for all $i, j = 1, \ldots, n$ and fixed complex number γ.

Theorem 9.1 *For each complex $(n \times n)$-matrix $A = (a_{ij})$,*

$$\int \exp\left(\sum_{i,j=1}^{n} \zeta_i a_{ij} \eta_j\right) \prod_{i=1}^{n} d\eta_i d\zeta_i = \det A. \tag{9.6}$$

Proof. (I) Let $n = 2$. We have to compute the coefficient c of $\eta_2\zeta_2\eta_1\zeta_1$ in the expansion of the integrand,

$$1 + \sum_{i,j=1}^{2} \zeta_i a_{ij} \eta_j + \frac{1}{2}\left(\sum_{i,j=1}^{2} \zeta_i a_{ij} \eta_j\right)^2 = c(\eta_2\zeta_2\eta_1\zeta_1) + \ldots$$

The dots denote the remaining terms. It turns out that

$$c = a_{11}a_{22} - a_{12}a_{21} = \det(A).$$

Let us show this. Since $\zeta_2\eta_2 = -\eta_2\zeta_2 + \gamma$, we get

$$\zeta_1\eta_1\zeta_2\eta_2 = -\zeta_1\eta_1\eta_2\zeta_2 + \gamma\zeta_1\eta_1.$$

Furthermore, it follows from $\eta_1\eta_2 = -\eta_2\eta_1$ and $\zeta_1\eta_2 = -\eta_2\zeta_1$ that

$$\zeta_1\eta_1\zeta_2\eta_2 = -\eta_2\zeta_1\eta_1\zeta_2 + \ldots = -\eta_2\zeta_2\zeta_1\eta_1 + \ldots = \eta_2\zeta_2\eta_1\zeta_1 + \ldots.$$

The dots denote terms that contain less than four factors. This implies

$$\frac{1}{2}(\zeta_1 a_{11}\eta_1)(\zeta_2 a_{22}\eta_2) = \frac{1}{2}\, a_{11}a_{22}(\eta_2\zeta_2\eta_1\zeta_1) + \ldots$$

The same expression is obtained for $\frac{1}{2}(\zeta_2 a_{22}\eta_2)(\zeta_1 a_{11}\eta_1)$. Similarly, we get

$$\frac{1}{2}(\zeta_1 a_{12}\eta_2)(\zeta_2 a_{21}\eta_1) = -\frac{1}{2}\, a_{12}a_{21}(\eta_2\zeta_2\eta_1\zeta_1) + \ldots,$$

and the same expression is obtained for $\frac{1}{2}(\zeta_2 a_{21}\eta_1)(\zeta_1 a_{12}\eta_2)$.

(II) For $n = 3, 4, \ldots$, the proof proceeds analogously. □

Remark. Observe that Theorem 9.1 is related to the classical Gaussian integral

$$\int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\sum_{i,j=1}^{n} x^i a_{ij} x^j\right) \frac{dx^1}{\sqrt{2\pi}} \cdots \frac{dx^n}{\sqrt{2\pi}} = \frac{1}{\sqrt{\det A}} \tag{9.7}$$

for all real symmetric $(n \times n)$-matrices $A = (a_{ij})$ whose eigenvalues are positive.

The Grassmannian-Gaussian integral (9.6) *has the advantage over the classical Gaussian integral* (9.7) *that the determinant* $\det A$ *appears in the numerator.*

This simplifies computations in physics. As a typical application, we will introduce the Faddeev–Popov trick in Sect. 16.6 on page 889. This trick introduces ghosts into gauge field theories in order to guarantee the crucial unitarity of the S-matrix.

9.6 The Method of Stationary Phase

We want to compute the following Grassmann integral

$$W(J,\overline{J}) := \int \mathrm{e}^{\Phi(\psi,\overline{\psi},J,\overline{J})} \prod_{k=1}^{N} d\overline{\psi}(k) d\psi(k) \tag{9.8}$$

with the phase function

$$\Phi(\psi,\overline{\psi},J,\overline{J}) := \mathrm{i}\overline{\psi}A\psi + \mathrm{i}\overline{\psi}J + \mathrm{i}\overline{J}\psi.$$

Let us first explain the notation.

- For fixed $N = 1, 2, \ldots$, the complex $(N \times N)$-matrix A is invertible.
- The quantities $\psi(k), \overline{\psi}(k), J(k), \overline{J}(k)$ with $k = 1, \ldots N$ form a sequence $\chi_1, \ldots \chi_{4N}$ of Grassmann variables, that is,

  $$\chi_i \chi_j = -\chi_j \chi_i, \qquad i,j = 1, \ldots, 4N.$$

  In particular, the symbol $\prod_{k=1}^{N} d\overline{\psi}(k) d\psi(k)$ stands for the ordered product $d\overline{\psi}(1) d\psi(1) \cdots d\overline{\psi}(N) d\psi(N)$.
- We use the following matrices

  $$\psi = \begin{pmatrix} \psi(1) \\ \vdots \\ \psi(N) \end{pmatrix}, \qquad J = \begin{pmatrix} J(1) \\ \vdots \\ J(N) \end{pmatrix}$$

  along with $\overline{\psi} = (\overline{\psi}(1), \ldots, \overline{\psi}(N))$ and $\overline{J} = (\overline{J}(1), \ldots, \overline{J}(N))$.

The proof of the following theorem proceeds analogously to the proof of Theorem 7.36 on page 438.

Theorem 9.2 $W(J,\overline{J}) = e^{-\mathrm{i}\overline{J}A^{-1}J}\, W(0,0)$.

9.7 The Fermionic Response Model

The global quantum action principle. Parallel to Sect. 7.26 on page 477, we study the generating functional

$$\boxed{Z(J,\overline{J}) = \mathcal{N} \int \mathrm{e}^{\mathrm{i}S[\psi,\overline{\psi},J,\overline{J}]/\hbar}\, \mathcal{D}\overline{\psi}\mathcal{D}\psi}$$

along with the action functional

$$S[\psi,\overline{\psi},J,\overline{J}] := \langle\overline{\psi}|(D+\mathrm{i}\varepsilon I)\psi\rangle + \kappa\langle 1|\mathcal{L}_{\mathrm{int}}(\psi,\overline{\psi})\rangle + \\ +\langle\overline{J}|\psi\rangle + \langle\overline{\psi}|J\rangle.$$

The detailed notation can be found in (7.161) on page 477. In contrast to Sect. 7.26, we now assume that the quantities

$$\psi(x),\ \ \overline{\psi}(x),\ \ \ J(x),\ \ \overline{J}(x)$$

are not complex numbers, but independent Grassmann variables. Here, the index x denotes an arbitrary discrete space-time point, that is, $x \in \mathcal{M}$. Explicitly, the functional integral is to be understood as the following integral

$$Z(J,\overline{J}) = \mathcal{N}\int \mathrm{e}^{\mathrm{i}S[\psi,\overline{\psi},J,\overline{J}]/\hbar} \prod_{x\in\mathcal{M}} d\overline{\psi}(x)d\psi(x)$$

with respect to Grassmann variables. The normalization factor $\mathcal{N}$ has to be chosen in such a way that $Z(0,0) = 1$. The symbol $\prod_{x\in\mathcal{M}} d\overline{\psi}(x)d\psi(x)$ stands for the product

$$d\overline{\psi}(1)d\psi(1)\cdots d\overline{\psi}(N)d\psi(N)$$

where the discrete space-time points $x \in \mathcal{M}$ are numbered in a fixed order.

The magic quantum action reduction formula. The point is that this formula reads as in the case of the extended response model considered in Sect. 7.26. Explicitly,

$$\boxed{Z(J,\overline{J}) = \exp\left\{\frac{\mathrm{i}\kappa}{\hbar}\sum_{x\in\mathcal{M}} \Delta^4 x\ \mathcal{L}_{\mathrm{int}}\left(\frac{\hbar}{\mathrm{i}}\frac{\delta}{\delta\overline{J}(x)}, \frac{\hbar}{\mathrm{i}}\frac{\delta}{\delta J(x)}\right)\right\} Z_{\mathrm{free}}(J,\overline{J})}$$

along with

$$Z_{\mathrm{free}}(J,\overline{J}) := \mathrm{e}^{\mathrm{i}\langle\overline{J}|R_\varepsilon J\rangle/\hbar}$$

where $R_\varepsilon := -(D+\varepsilon I)^{-1}$. Here, we use the partial functional derivatives

$$\frac{\delta}{\delta J(x)} := \frac{1}{\Delta^4 x}\frac{\partial_l}{\partial J(x)}, \qquad \frac{\delta}{\delta\overline{J}(x)} := \frac{1}{\Delta^4 x}\frac{\partial_l}{\partial\overline{J}(x)}.$$

The proof proceeds analogously to the proof given in Sect. 7.24 on page 438. Observe that, in the present case, we have to use the principle of stationary phase for integrals with Grassmann variables from Sect. 9.6. The different determinant tricks for Gaussian and Grassmannian–Gaussian integrals do not matter, since the choice of the normalization factor $\mathcal{N}$ always cancels the different determinants.

The magic LSZ reduction formula. From the quantum action reduction formula above, we get the LSZ reduction formula as in Sect. 7.26. We will come back to this in connection with the Standard Model in particle physics. We also refer to Faddeev and Slavnov (1980).

10. Infinite-Dimensional Hilbert Spaces

> Quantum fields possess an infinite number of degrees of freedom. This causes a lot of mathematical trouble.
>
> Folklore

Smooth functions. Let Ω be an open subset of $\mathbb{R}^N$, $N = 1, 2, \ldots$ The function $f : \Omega \to \mathbb{C}$ is called smooth iff it is continuous and the partial derivatives of f of arbitrary order are also continuous on Ω. For the theory of infinite-dimensional Hilbert spaces and its applications in physics, it is important to use not only smooth functions, but also reasonable discontinuous functions which are limits of smooth functions. Here, we use pointwise limits, $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in \Omega$, and more general limits in the sense of the averaging over integrals, for example,

$$\lim_{n\to\infty} \int_\Omega |f_n(x) - f(x)|^2 d^N x = 0.$$

Furthermore, it is important to replace the classical Riemann integral by the modern Lebesgue integral.

10.1 The Importance of Infinite Dimensions in Quantum Physics

We want to discuss why an infinite number degrees of freedom for quantum physics is inevitable. To this end, we will show that the Heisenberg uncertainty relation cannot be realized in a finite-dimensional Hilbert space.

10.1.1 The Uncertainty Relation

> Before you start to axiomatize things, be sure that you first have something of mathematical substance.
>
> Hermann Weyl (1885–1955)

In 1927 Heisenberg (1901–1976) discovered that in contrast to Newton's classical mechanics, it is impossible to measure precisely position and momentum of a quantum particle at the same time. Heisenberg based his mathematical argument on the commutation relation

$$\boxed{QP - PQ = \mathrm{i}\hbar I} \tag{10.1}$$

for the position operator Q and the momentum operator P, along with the Schwarz inequality.

Finite-dimensional Hilbert spaces fail. Observe first that the fundamental commutation relation (10.1) cannot be realized for observables Q and P living in a nontrivial finite-dimensional Hilbert space X if the Planck constant $\hbar$ is different from zero. [1] Indeed, suppose that there exist two self-adjoint linear operators

$$Q, P : X \to X$$

such that (10.1) holds true. By Proposition 7.11 on page 364, $\operatorname{tr}(QP) = \operatorname{tr}(PQ)$. This implies

$$0 = \operatorname{tr}(QP - PQ) = \mathrm{i}\hbar \cdot \operatorname{tr} I = \mathrm{i}\hbar \dim X.$$

Thus, relation (10.1) forces the vanishing of the Planck constant $\hbar$ in the setting of a nontrivial finite-dimensional Hilbert space.

A nontrivial mathematical model. Our goal is to construct a nontrivial model which realizes the commutation relation (10.1). To this end, we choose the space $C_2(\mathbb{R})$ which consists of all continuous functions $\psi : \mathbb{R} \to \mathbb{C}$ with the property $\int_{-\infty}^{\infty} |\psi(x)|^2 dx < \infty$. Define $\langle \varphi | \psi \rangle := \int_{-\infty}^{\infty} \varphi(x)^\dagger \psi(x) dx$.

Proposition 10.1 *For all functions $\varphi, \psi \in C_2(\mathbb{R})$, the integral $\langle \varphi | \psi \rangle$ is finite.*

Proof. Set

$$a_n := \int_{-n}^{n} \varphi(x)^\dagger \psi(x) dx, \qquad n = 1, 2, \ldots$$

By the classical Schwarz inequality for integrals,

$$|a_n|^2 \le \int_{-n}^{n} |\varphi(x)|^2 dx \int_{-n}^{n} |\psi(x)|^2 dx.$$

This implies

$$|a_n|^2 \le \int_{-\infty}^{\infty} |\varphi(x)|^2 \, dx \int_{-\infty}^{\infty} |\psi(x)|^2 \, dx, \qquad n = 1, 2, \ldots$$

Thus, the sequence $a_1, a_2, \ldots$ is increasing and bounded. Consequently, the finite limit $\lim_{n\to\infty} a_n$ exists. □

Proposition 10.2 *The space $C_2(\mathbb{R})$ is a complex pre-Hilbert space.*

Proof. It can be checked easily that $\langle \varphi | \psi \rangle$ possesses the properties (P1) through (P5) from page 337. In particular, for a given continuous function $\psi : \mathbb{R} \to \mathbb{C}$, we have $\langle \psi | \psi \rangle = 0$, that is,

$$\int_{\mathbb{R}} |\psi(x)|^2 dx = 0$$

iff the function ψ vanishes identically. □

Consider now the Schwartz space $\mathcal{D}(\mathbb{R})$ which consists of all smooth functions $\psi : \mathbb{R} \to \mathbb{C}$ that vanish outside some finite interval. For all functions $\psi \in \mathcal{D}(\mathbb{R})$, we define the so-called position operator Q,

$$\boxed{(Q\psi)(x) := x\psi(x) \qquad \text{for all} \quad x \in \mathbb{R},}$$

and the so-called momentum operator P,

[1] In the trivial Hilbert space $\{0\}$, relation (10.1) is obviously true.

$$\boxed{(P\psi)(x) := -\mathrm{i}\hbar\frac{d\psi(x)}{dx} \qquad \text{for all} \quad x \in \mathbb{R}.}$$

For every function $\varphi \in \mathcal{D}(\mathbb{R})$, we have the commutation relation

$$QP\varphi - PQ\varphi = \mathrm{i}\hbar\varphi \qquad \text{on } \mathbb{R}. \tag{10.2}$$

In fact, $PQ\varphi = -\mathrm{i}\hbar\varphi + QP\varphi$ follows from the product rule,

$$\frac{d}{dx}(x\varphi(x)) = \varphi(x) + x\varphi'(x).$$

Finally, for all $\varphi \in \mathcal{D}(\mathbb{R})$ with $||\varphi|| = 1$, set

$$\overline{Q} := \langle\varphi|Q\varphi\rangle, \qquad \overline{P} := \langle\varphi|P\varphi\rangle,$$

as well as $\Delta Q := ||(Q - \overline{Q}I)\varphi||$ and $\Delta P := ||(P - \overline{P}I)\varphi||$. In terms of physics, if the quantum particle is in the state φ on the real line, then $\overline{Q}$ is the mean position and ΔQ is the mean fluctuation of the particle position. Similarly, $\overline{P}$ is the mean momentum and ΔP is the mean fluctuation of the particle momentum.

Proposition 10.3 *There holds the Heisenberg uncertainty inequality*

$$\Delta Q \cdot \Delta P \geq \frac{\hbar}{2}.$$

Proof. This is a special case of Theorem 10.4 below with the pre-Hilbert space $X := C(\mathbb{R})_2$, the linear subspace $\mathcal{D} := \mathcal{D}(\mathbb{R})$ of X, and the identity operator $C := I$. Note that the operators $Q, P : \mathcal{D}(\mathbb{R}) \to \mathcal{D}(\mathbb{R})$ are formally self-adjoint. In fact, for all $\varphi, \psi \in \mathcal{D}(\mathbb{R})$, integration by parts yields

$$\langle\varphi|P\psi\rangle = \int_{\mathbb{R}} \varphi(x)^\dagger(-\mathrm{i}\hbar)\psi'(x)dx = \int_{\mathbb{R}} (-\mathrm{i}\hbar\varphi'(x))^\dagger\psi(x)dx = \langle P\varphi|\psi\rangle.$$

Moreover,

$$\langle\varphi|Q\psi\rangle = \int_{\mathbb{R}} \varphi(x)^\dagger x\psi(x)dx = \int_{\mathbb{R}} (x\varphi(x))^\dagger\psi(x)dx = \langle Q\varphi|\psi\rangle.$$

□

The abstract uncertainty theorem. We make the following assumptions.

(H1) We are given a linear subspace $\mathcal{D}$ of the complex infinite-dimensional pre-Hilbert space X.

(H2) The linear operators $Q, P : \mathcal{D} \to X$ are formally self-adjoint, that is, we have

$$\langle Q\varphi|\psi\rangle = \langle\varphi|Q\psi\rangle \qquad \text{for all} \quad \varphi, \psi \in \mathcal{D}.$$

The analogous formula is true if we replace Q by P.

(H3) If $\varphi \in \mathcal{D}$, then $Q\varphi \in \mathcal{D}$ and $P\varphi \in \mathcal{D}$.

(H4) For all $\varphi \in \mathcal{D}$, there holds the commutation relation

$$QP\varphi - PQ\varphi = \mathrm{i}\hbar C\varphi.$$

Furthermore, for each $\varphi \in \mathcal{D}$ with $||\varphi|| = 1$, set $\overline{C} := \langle\varphi|C\varphi\rangle$. Similarly, we define $\overline{Q} := \langle\varphi|Q\varphi\rangle$ and $\overline{P} := \langle\varphi|P\varphi\rangle$. Furthermore, define

$$\Delta Q := ||(Q - \overline{Q}I)\varphi||, \qquad \Delta P := ||(P - \overline{P}I)\varphi||.$$

The following theorem is called the abstract uncertainty theorem.

Theorem 10.4 *There holds the inequality* $\Delta Q \cdot \Delta P \geq \frac{1}{2}\hbar|\overline{C}|$.

Proof. Set $z := \langle (Q - \overline{Q}I)\varphi | (P - \overline{P}I)\varphi \rangle$. By the Schwarz inequality,

$$|z| \leq ||(Q - \overline{Q}I)\varphi|| \cdot ||(P - \overline{P}I)\varphi|| = \Delta Q \Delta P.$$

Since $\overline{Q}$ and $\overline{P}$ are real numbers,

$$\begin{aligned} z - z^\dagger &= \langle (Q - \overline{Q}I)\varphi | (P - \overline{P}I)\varphi \rangle - \langle (P - \overline{P}I)\varphi | (Q - \overline{Q}I)\varphi \rangle \\ &= \langle \varphi | (Q - \overline{Q}I)(P - \overline{P}I)\varphi \rangle - \langle \varphi | (P - \overline{P}I)(Q - \overline{Q}I)\varphi \rangle \\ &= \langle \varphi | (QP - PQ)\varphi \rangle = \mathrm{i}\hbar \langle \varphi | C\varphi \rangle. \end{aligned}$$

Finally,

$$\hbar|\overline{C}| = |\hbar\langle \varphi | C\varphi \rangle| = |z - z^\dagger| = |2\mathrm{i}\Im(z)| \leq 2|z| \leq 2\Delta Q \Delta P.$$

□

10.1.2 The Trouble with the Continuous Spectrum

Let $H : X \to X$ be a linear self-adjoint operator on the complex finite-dimensional Hilbert space X. Then, the inverse operator

$$(H - EI)^{-1} : X \to X$$

exists for all complex energy parameters $E \in \mathbb{C} \setminus \sigma(H)$ up to a finite set of real energy values, $\sigma(H) := \{E_1, \ldots, E_N\}$. The set $\sigma(H)$ is called the energy spectrum of the operator H. In the finite-dimensional case, there exists a complete orthonormal system $|E_1\rangle, \ldots |E_N\rangle$ in the Hilbert space X such that

$$H|E_j\rangle = E_j|E_j\rangle, \qquad j = 1, \ldots, N.$$

The situation may change dramatically in an infinite-dimensional Hilbert space X. In such a space it is possible that the energy spectrum $\sigma(H)$ contains a continuum of energy values. For example, consider the electron of the hydrogen atom. Then, in non-relativistic quantum mechanics, the spectrum of the corresponding Hamiltonian H has the form

$$\sigma(H) = \{E_1, E_2, \ldots\} \cup [0, \infty[$$

where

$$E_n := -\frac{m_e c^2 \alpha^2}{2n^2}, \qquad n = 1, 2, \ldots.$$

Here, we use the following notation: m_e rest mass of the electron, c velocity of light in a vacuum, $\alpha = 137.04$ fine structure constant. Note the following two crucial facts:

- the energy values $E_1, E_2, \ldots$ correspond to bound states of the electron in the hydrogen atom, and
- the energy values $E \in [0, \infty[$ correspond to scattering states (scattering of an electron at the nucleus (proton) of the hydrogen atom).

In terms of our solar system, the bound states correspond to planets moving on ellipses, and the scattering states correspond to comets moving on hyperbolas and leaving the solar system for ever (Fig. 2.11 on page 121). For the hydrogen atom, the point is that the bound states correspond to eigenvectors $|E_1\rangle, |E_2\rangle, \ldots$ of the Hamiltonian H in the Hilbert space $X = L_2(\mathbb{R}^3)$, that is,

$$H|E_n\rangle = E_n|E_n\rangle, \qquad n = 1, 2, \ldots.$$

In contrast to this, the energy values E which lie in the continuous spectrum $\sigma_{\text{cont}} := [0, \infty[$ do not correspond to eigenvectors in the Hilbert space X, but to more general objects $\langle E|$,

$$H\langle E| = E\langle E|, \qquad \langle E| \in Y.$$

Here, the costate $\langle E|$ does not lie in the original Hilbert space X, but in some larger space Y. For the spectral theory in terms of costates, see Sect. 12.2 on page 675.

Historical remarks. In Sect. 7.4, we introduced the notion of a finite-dimensional Hilbert space. For quantum physics, it is crucial that the situation of finite-dimensional Hilbert spaces can be generalized to infinite dimensions. This represents a far-reaching generalization of the classical Fourier method for solving partial differential equations. The main contributions are due to the following mathematicians and physicists:

- Fourier (1786–1830) in 1822 (the Fourier method for solving the heat equation),
- Cauchy (1789–1857) in 1826 (principal axis transformation for finite-dimensional quadratic forms),
- Hilbert (1862–1943) in 1904 (principal axis transformation for infinite-dimensional symmetric quadratic forms (symmetric matrices)),
- Schrödinger (1887–1961) in 1926 (application of the Fourier method to the spectrum of the hydrogen atom),
- von Neumann (1903–1957) in 1928 (spectral theory for unbounded self-adjoint operators and the mathematical foundation of quantum mechanics),
- Dirac (1902–1984) in 1930 (transformation theory and Dirac calculus),
- Laurent Schwartz in 1945 (theory of distributions),
- Gelfand and Kostyuchenko in 1955 (generalized eigenfunctions and rigorous justification of the Dirac calculus).

Nowadays this is part of functional analysis and harmonic analysis. The point is that the classical Fourier transform is closely related to the translation group on the real line. If we replace the translation group by more general groups, then we get more general eigenfunction expansions and integral transformations for broad classes of special functions in mathematical physics. This goes back to

- Sturm (1803–1855) and Liouville (1818–1882) in 1836 and 1837, respectively.

In the 20th century, important contributions were made by

- Hermann Weyl (1885–1955) (representation theory for compact Lie groups),
- and Eugene Wigner (1902–1995) (representation theory of the noncompact Poincaré group).

10.2 The Hilbert Space $L_2(\Omega)$

Counterexample. In a finite-dimensional Hilbert space as defined in Sect. 7.4 on page 335, the completeness condition is automatically satisfied. The situation changes essentially in the infinite-dimensional case. To illustrate this, consider the

space $C_2(\mathbb{R})$ of continuous functions $\psi : \mathbb{R} \to \mathbb{C}$ with $\int_{\mathbb{R}} |\psi(x)|^2 dx < \infty$. It follows from the classical Schwarz inequality that for all $\varphi, \psi \in C_2(\mathbb{R})$, the integral

$$\langle \psi | \varphi \rangle := \int_{\mathbb{R}} \psi(x)^\dagger \varphi(x) dx$$

is finite. Equipped with this inner product, the space $C_2(\mathbb{R})$ becomes a complex pre-Hilbert space.

Unfortunately, the space $C_2(\mathbb{R})$ is not a Hilbert space, since the completeness condition fails.

In fact, it can be shown that there exists a sequence (ψ_n) of continuous functions ψ_n in the space $C_2(\mathbb{R})$ such that for each number $\varepsilon > 0$, there exists an index $n_0(\varepsilon)$ with

$$||\psi_m - \psi_n|| = \left(\int_{\mathbb{R}} |\psi_m(x) - \psi_n(x)|^2 dx \right)^{1/2} < \varepsilon$$

for all $m, n \geq n_0(\varepsilon)$. But there does *not* exist any continuous function ψ in the space $C_2(\mathbb{R})$ such that

$$\lim_{n \to \infty} ||\psi_n - \psi|| = \lim_{n \to \infty} \left(\int_{\mathbb{R}} |\psi_n(x) - \psi(x)|^2 dx \right)^{1/2} = 0.$$

The point is that such a function ψ only exists if

- we allow the use of discontinuous functions, and
- we replace the classical Riemann integral by the more general Lebesgue integral.

This integral was introduced by Henri Lebesgue in his 1902 Paris dissertation. Using the Lebesgue integral, the space $C_2(\mathbb{R})$ can be extended to some space $L_2(\mathbb{R})$, that is,

$$\boxed{C_2(\mathbb{R}) \subset L_2(\mathbb{R})}$$

such that $L_2(\mathbb{R})$ is a complex Hilbert space. In particular, the space $L_2(\mathbb{R})$ contains a class of reasonable discontinuous functions $\psi : \mathbb{R} \to \mathbb{C}$. Moreover, the properties of the Lebesgue integral force the validity of the completeness relation with respect to the inner product

$$\langle \psi | \varphi \rangle := \int_{\mathbb{R}} \psi(x)^\dagger \varphi(x) dx.$$

Explicitly, the completeness condition means the following. Suppose that we are given a sequence (ψ_n) of functions ψ_n in the space $L_2(\mathbb{R})$ such that for each number $\varepsilon > 0$, there exists an index $n_0(\varepsilon)$ with

$$||\psi_m - \psi_n|| = \left(\int_{\mathbb{R}} |\psi_m(x) - \psi_n(x)|^2 dx \right)^{1/2} < \varepsilon$$

for all $m, n \geq n_0(\varepsilon)$. There exists then a function $\psi \in L_2(\mathbb{R})$ such that $\psi_n \to \psi$ in $L_2(\mathbb{R})$ as $n \to \infty$. This means that

$$\lim_{n \to \infty} ||\psi_n - \psi|| = \lim_{n \to \infty} \left(\int_{\mathbb{R}} |\psi_n(x) - \psi(x)|^2 dx \right)^{1/2} = 0.$$

Let us discuss the main points. In what follows, we formulate the statements for the N-dimensional space $\mathbb{R}^N$ with $N = 1, 2, \ldots$ Note the following special cases.

- For $N = 1$, the space $\mathbb{R}^1$ coincides with the real line, and 1-dimensional cubes are intervals.
- For $N = 2$, the space $\mathbb{R}^2$ coincides with the plane, and 2-dimensional cubes are squares.

For $x \in \mathbb{R}^N$, recall that $||x|| := \sqrt{x_1^2 + \dots x_N^2}$.

Using an intuitive picture, the incomplete space $C_2(\mathbb{R})$ and the complete space $L_2(\mathbb{R})$ correspond to the incomplete space of rational numbers $\mathbb{Q}$ and the complete space of real numbers $\mathbb{R}$, respectively.

A Cauchy sequence (x_n) of rational numbers $x_1, x_2, \dots$ is does not always converge to a rational number. However, if we complete the space of rational numbers to the space of real numbers by introducing irrational numbers, then each Cauchy sequence of real numbers converges to a real number. The term 'irrational number' indicates that in the history of mathematics, mathematicians had philosophical trouble with understanding this notion.

10.2.1 Measure and Integral

> Almost all concepts which relate to the modern measure and integration theory, go back to the works of Lebesgue (1875–1941). The introduction of these concepts was the turning point in the transition from mathematics of the 19th century to mathematics of the 20th century.
>
> Naum Vilenkin

Our goal is to introduce the integral

$$\int_{\mathbb{R}^N} f(x) d\mu(x) = \sum_{j=1}^{J} f(x_j) m_j$$

and to generalize it to the prototype

$$\int_{\mathbb{R}^N} f(x) d\mu(x) = \int_{\mathbb{R}^N} f(x) \varrho(x) d^N x + \sum_{j=1}^{\infty} f(x_j) m_j.$$

Intuitively, the function ϱ represents a mass density on $\mathbb{R}^N$, and m_j is the mass at the point x_j in $\mathbb{R}^N$. On the real line, the value

$$\overline{x} = \frac{\int_{\mathbb{R}} x d\mu(x)}{\int_{\mathbb{R}} d\mu(x)}$$

represents the center of gravity of the mass distribution. This measure integral includes finite sums, infinite series, and traditional integrals, as special cases. This generality is needed for obtaining expansion formulas of the following form,

$$\boxed{F(y) = \int_{\mathbb{R}^N} K(y,x) d\mu(x), \qquad y \in \mathbb{R}^N.}$$

Such general expansions appear in von Neumann's and Dirac's operator calculus. This plays a fundamental role in quantum physics. These expansion formulas generalize the Fourier series of periodic functions and the Fourier transform. Furthermore, the measure integral is the natural setting for the theory of probability. Mass has then to be replaced by probability.

Measure theory begins with Archimedes of Syracus (287–212 B.C.) who computed the measure of the unit circle, $\mathbb{S}^1$. Using a polygon with 96 nodes, he obtained the approximation $\mu(\mathbb{S}^1) = 6.28$ which corresponds to $\pi = 3.14$. Around 1900, modern measure theory was founded by Borel (1871–1956) and Lebesgue (1875–1941). In 1932 Kolmogorov (1903–1987) used general measure theory in order to found the modern theory of probability. For example, the crucial mean value of the random function $f : \mathbb{R}^N \to \mathbb{R}$ is defined by

$$\bar{f} := \frac{\int_{\mathbb{R}^N} f(x) d\mu(x)}{\int_{\mathbb{R}^N} d\mu(x)}.$$

Definition of measure. The notion of measure generalizes the intuitive notion of volume, mass, positive electric charge, and probability. Suppose we are given an arbitrary set S. To certain subsets A of the set S we want to assign a number, $\mu(A)$, with

$$0 \le \mu(A) \le \infty.$$

The number $\mu(A)$ is called the measure of the set A.[2] More precisely, by a measure we understand a map $\mu : \mathcal{A} \mapsto [0, \infty]$ which has the following properties:

(P1) σ-algebra: The members of $\mathcal{A}$ are subsets of S including the set S and the empty set, $\emptyset$. If $A, B, A_1, A_2, ..$ are members of $\mathcal{A}$, then so are

$$A \setminus B, \quad A \cup B, \quad A \cap B, \quad \bigcup_{n=1}^{\infty} A_n, \quad \bigcap_{n=1}^{\infty} A_n.$$

(P2) Additivity: If A and B are disjoint members of $\mathcal{A}$, then

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

(P3) σ-additivity: If (A_n) is a pairwise disjoint family of members from $\mathcal{A}$, then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

The members of $\mathcal{A}$ are called measurable sets (with respect to μ). The measure is called finite iff $\mu(S) < \infty$. By a zero set, we understand a set whose measure is zero. The measure is called complete iff subsets of zero sets are always zero sets.

Measurable functions. The function $f : S \to \mathbb{C}$ is called measurable iff the preimage of open sets is measurable.

Step functions. Let us start with the formula

$$\int_{\mathbb{R}^N} f(x) d\mu(x) = \sum_{j=1}^{J} f_j \mu(A_j).$$

for step functions. By definition, the function $f : \mathbb{R}^N \to \mathbb{C}$ is called a step function iff there exists a finite number $A_1, \ldots, A_m$ of subsets of $\mathbb{R}^N$ with finite measure such that f is constant on A_j with value f_j for all j, and f vanishes outside the sets $A_1, \ldots, A_J$.

Measure integral. For a function $f : \mathbb{R}^N \to \mathbb{C}$, we define the integral by the key formula

[2] In the theory of probability, $\mu(S) := 1$.

$$\boxed{\int_{\mathbb{R}^N} f(x)d\mu(x) := \lim_{n\to\infty} \int_{\mathbb{R}^N} f_n(x)d\mu(x)} \tag{10.3}$$

where $f_n : \mathbb{R}^N \to \mathbb{C}$ are step functions, $n = 1, 2, \ldots$ More precisely, assume that the following hold true.

(H1) Convergence almost everywhere: There exists a sequence (f_n) of step functions such that

$$\lim_{n\to\infty} f_n(x) = f(x)$$

for all $x \in \mathbb{R}^N$ with the possible exception of a zero set.

(H2) Mean square approximation: The sequence (f_n) is Cauchy with respect to the mean square norm, that is, for each number $\varepsilon > 0$, there exists an index $n_0(\varepsilon)$ with

$$\int_{\mathbb{R}^N} |f_n(x) - f_m(x)|^2 d\mu(x) < \varepsilon \qquad \text{for all} \quad n, m \geq n_0(\varepsilon).$$

This integral is to be understood in the sense of step functions.

Theorem 10.5 *The finite limit* (10.3) *exists and is independent of the choice of the step functions.*

Precisely in this case, we say that the integral $\int_{\mathbb{R}^N} f(x)d\mu(x)$ exists. This is equivalent to the existence of the integral $\int_{\mathbb{R}^N} |f(x)|d\mu(x)$.[3]

The integral $\int_S f(x)d\mu(x)$ is defined analogously by replacing the set $\mathbb{R}^N$ by S. The value of the integral $\int_S f(x)d\mu(x)$ remains invariant if we change the function f on a zero set.

Majorant criterion. If $f, g : S \to \mathbb{C}$ are measurable and

$$|f(x)| \leq g(x) \qquad \text{for all} \quad x \in S,$$

then the existence of the integral $\int_S g(x)d\mu(x)$ implies the existence of the integral $\int_S f(x)d\mu(x)$, and we have the inequalities

$$\left|\int_S f(x)d\mu(x)\right| \leq \int_S |f(x)|d\mu(x) \leq \int_S g(x)d\mu(x).$$

10.2.2 Dirac Measure and Dirac Integral

The Dirac measure. Fix a point x_0 in $\mathbb{R}^N$. For an arbitrary subset A of $\mathbb{R}^N$, define

$$\mu(A) := \begin{cases} 1 & \text{if } x_0 \in A, \\ 0 & \text{if } x_0 \notin A. \end{cases}$$

Each subset A of $\mathbb{R}^N$ and each function $f : \mathbb{R}^N \to \mathbb{C}$ is measurable. For the integral,[4]

$$\int_{\mathbb{R}^N} f(x)d\mu(x) = f(x_0).$$

[3] For the function $f(x) \equiv 1$, the integral $\int_{\mathbb{R}} f(x)dx$ does not exist. Since $\lim_{R\to+\infty} \int_{-R}^{R} dx = \infty$, we say that the integral is divergent.

[4] Physicists write formally $\int_{\mathbb{R}^N} f(x)\delta(x - x_0)d^N x = f(x_0)$.

Finite number of mass points. Assign the positive mass $m_1, \ldots, m_J$ to the points $x_1, \ldots, x_J$ in $\mathbb{R}^N$, respectively. By definition, the measure $\mu(A)$ of an arbitrary subset A of $\mathbb{R}^N$ is equal to the total mass of this set. For example, if $x_1, x_2 \in A$ and $x_2, \ldots, x_J$ lie outside A, then $\mu(A) = m_1 + m_2$. For an arbitrary function $f : \mathbb{R}^N \to \mathbb{C}$,

$$\int_{\mathbb{R}^N} f(x) d\mu(x) = \sum_{j=1}^{J} f(x_j) m_j.$$

10.2.3 Lebesgue Measure and Lebesgue Integral

Characterization of the Lebesgue measure by translation invariance. The Lebesgue measure generalizes the classical volume of a set in $\mathbb{R}^N$.

Theorem 10.6 *There exists precisely one complete measure on $\mathbb{R}^N$ which generalizes the elementary measure of open cubes and which is invariant under translations.*

This measure is called the N-dimensional Lebesgue measure on $\mathbb{R}^N$. In particular, open and closed subsets of $\mathbb{R}^N$ are measurable. As a rule of thumb, non-measurable sets and functions with respect to the Lebesgue measure are highly pathological.

Zero sets. A subset of $\mathbb{R}^N$ is called a zero set iff it has the N-dimensional Lebesgue measure zero. The set A is a zero set iff for each $\varepsilon > 0$, there exists a system $J_1, J_2, \ldots$ of open N-dimensional cubes which cover the set A and whose total volume is less than ε. For example, the sets

$$\{x_1, \ldots, x_n\}, \qquad n = 1, 2, \ldots \qquad \text{or} \qquad \{x_1, x_2, \ldots\}$$

of a finite or countable number of points in $\mathbb{R}^N$ are zero sets. Roughly speaking, a subset of $\mathbb{R}^N$ is a zero set if its dimension is less than N. For example, the boundary of a ball or a cube in $\mathbb{R}^N$ is a zero set.

Almost everywhere continuous functions. By definition, the given function $f : \mathbb{R}^N \to \mathbb{C}$ is almost everywhere continuous iff it is continuous for all points of $\mathbb{R}^N$ with possible exception of a zero set.[5]

The Lebesgue integral. The Lebesgue integral refers to the Lebesgue measure. Note the following:

> *If the classical Riemann integral $\int_{\mathbb{R}^N} |f(x)| d^N x$ is finite, then the Lebesgue integral $\int_{\mathbb{R}^N} f(x) d^N x$ exists and is equal to the Riemann integral.*

In this monograph, all the integrals $\int_{\mathbb{R}^N} f(x) d^N x$ are to be understood in the sense of Lebesgue.

Example. If the function $f : \mathbb{R}^N \to \mathbb{C}$ is almost every continuous and satisfies the following growth condition

$$|f(x)| \leq \frac{\text{const}}{1 + ||x||^{N+1}} \qquad \text{for all} \quad x \in \mathbb{R}^N,$$

then the Lebesgue integral $\int_{\mathbb{R}^N} f(x) d^N x$ exists. In contrast to the Riemann integral, the Lebesgue integral possesses the following nice property concerning limits. We have

[5] Generally, a property is true almost everywhere on $\mathbb{R}^N$ iff it is true for all points of $\mathbb{R}^N$ with possible exception of a zero set.

$$\boxed{\lim_{n\to\infty}\int_{\mathbb{R}^N} f_n(x)dx = \int_{\mathbb{R}^N}\lim_{n\to\infty} f_n(x)d^N x}$$

if the following are met:

(H1) The integral $\int_{\mathbb{R}^N} f_n(x)d^N x$ exists for each n.

(H2) The limit $f(x) := \lim_{n\to\infty} f_n(x)$ exists almost everywhere on $\mathbb{R}^N$.

(H3) Majorant condition: There exists an integrable function $g : \mathbb{R}^N \to \mathbb{C}$ such that $|f_n(x)| \le g(x)$ almost everywhere on $\mathbb{R}^N$ for all n.

10.2.4 The Fischer–Riesz Theorem

By definition, the Lebesgue space $L_2(\mathbb{R}^N)$ consists of all complex-valued measurable functions $\psi : \mathbb{R}^N \to \mathbb{C}$ with $\int_{\mathbb{R}^N} |\psi(x)|^2 d^N x < \infty$.

Theorem 10.7 *The space $L_2(\mathbb{R}^N)$ is a complex infinite-dimensional Hilbert space with respect to the inner product*

$$\langle\psi|\varphi\rangle := \int_{\mathbb{R}^N} \psi(x)^\dagger \varphi(x) d^N x.$$

This fundamental theorem of modern analysis was proven independently by Ernst Fischer (1875–1954) and Fryges Riesz (1880–1956) in 1907. Note that, by definition, two functions ψ and φ represent the same element of the Hilbert space $L_2(\mathbb{R}^N)$ iff they differ on a zero set. In the Hilbert space $L_2(\mathbb{R}^N)$, the convergence

$$\lim_{n\to\infty} \psi_n = \psi$$

means explicitly that

$$\lim_{n\to\infty} ||\psi_n - \psi|| = \left(\int_{\mathbb{R}^N} |\psi_n(x) - \psi(x)|^2 d^N x\right)^{1/2} = 0.$$

This is called mean-square convergence.

If we replace $\mathbb{R}^N$ by a measurable subset Ω of $\mathbb{R}^N$ (e.g., Ω is open or closed), then we obtain the Hilbert space $L_2(\Omega)$ with the inner product

$$\langle\psi|\varphi\rangle := \int_{\Omega} \psi(x)^\dagger \varphi(x) d^N x.$$

Suggested reading. As an introduction to the theory of infinite-dimensional Hilbert spaces, we recommend the following two textbooks:

- P. Lax, Functional Analysis, Wiley, New York, 2002.
- E. Zeidler, Applied Functional Analysis: Applications to Mathematical Physics, Springer, New York, 1995.

In the Appendix to the latter book, the interested reader finds a summary of the basic properties of the Lebesgue integral. A detailed summary on general modern measure and integration theory can be found in the Appendix to the author's monograph,

- E. Zeidler, Nonlinear Functional Analysis and its Applications, Vol. IIB, Springer, New York, 1990.

As an introduction to the Lebesgue integral, we recommend the following textbooks:

- S. Lang, Real Analysis and Functional Analysis, Springer, New York, 1993.
- E. Lieb and M. Loss, Analysis, American Mathematical Society, Providence, Rhode Island, 1997.
- E. Stein and R. Shakarchi, Measure Theory, Princeton University Press, 2003.

10.3 Harmonic Analysis

The proofs to the statements of the following summary can be found in the author's textbook, Zeidler (1995), Vol. 1, Chap. 3.

10.3.1 Gauss' Method of Least Squares

Orthonormal system. Let X be a complex infinite-dimensional Hilbert space. By definition, the elements $\varphi_1, \ldots, \varphi_n$ of X form an orthonormal system iff

$$\langle \varphi_j | \varphi_k \rangle = \delta_{jk}, \qquad j,k = 1, \ldots, n.$$

For each given $\psi \in X$, we define the Fourier coefficients by setting

$$a_k(\psi) := \langle \varphi_k | \psi \rangle, \qquad k = 1, \ldots, n.$$

As we will show below, this definition generalizes the classical Fourier coefficients.

Theorem 10.8 *The Fourier coefficients $a_1(\psi), \ldots, a_n(\psi)$ are the unique solution of the minimum problem*

$$||\psi - a_1\varphi_1 - \ldots - a_n\varphi_n||^2 = \min!, \qquad a_1, \ldots, a_n \in \mathbb{C}$$

which is the abstract form of Gauss' least square method.

Completeness. The orthonormal system $\varphi_1, \varphi_2, \ldots$ in the Hilbert space X is called complete iff for all $\psi \in X$, the Fourier series is convergent,

$$\psi = \sum_{k=1}^{\infty} \langle \varphi_k | \psi \rangle \varphi_k. \tag{10.4}$$

This means that $\lim_{n\to\infty} || \sum_{k=1}^{n} \langle \varphi_k | \psi \rangle \varphi_k - \psi || = 0$. Then, for all $\varphi, \psi \in X$, we have the Parseval equation

$$\langle \varphi | \psi \rangle = \sum_{k=1}^{\infty} \langle \varphi | \varphi_k \rangle \langle \varphi_k | \psi \rangle. \tag{10.5}$$

The Dirac calculus. From the mnemonic formula

$$\boxed{\sum_{k=1}^{\infty} |\varphi_k\rangle\langle\varphi_k| = I}$$

called the completeness relation by physicists, we immediately get

$$|\psi\rangle = \sum_{k=1}^{\infty} |\varphi_k\rangle\langle\varphi_k|\psi\rangle$$

and

$$\langle \varphi | \psi \rangle = \sum_{k=1}^{\infty} \langle \varphi | \varphi_k \rangle \langle \varphi_k | \psi \rangle$$

which is identical with (10.4) and (10.5), respectively.

10.3.2 Discrete Fourier Transform

The Hilbert space $L_2(-\pi,\pi)$. Fourier (1768–1830) used the functions

$$\boxed{\varphi_p(x) := \frac{\mathrm{e}^{\mathrm{i}px}}{\sqrt{2\pi}}, \qquad p = 0, \pm 1, \pm 2, \dots}$$

The key observation is the following integral relation

$$\int_{-\pi}^{\pi} \varphi_p(x)^\dagger \varphi_q(x) dx = \delta_{pq}, \qquad p, q = 0, \pm 1, \pm 2, \dots$$

To translate this classical identity into the language of Hilbert spaces, let $L_2(-\pi,\pi)$ denote the space of all measurable[6] functions $\psi :]-\pi,\pi[\to \mathbb{C}$ with

$$\int_{-\pi}^{\pi} |\psi(x)|^2 dx < \infty.$$

The space $L_2(-\pi,\pi)$ becomes a complex Hilbert space equipped with the inner product

$$\langle \varphi | \psi \rangle := \int_{-\pi}^{\pi} \varphi(x)^\dagger \psi(x) dx.$$

Then $\langle \varphi_p | \varphi_q \rangle = \delta_{pq}$ for all $p, q = 0, \pm 1, \pm 2, \dots$

Theorem 10.9 *The system $\varphi_0, \varphi_1, \varphi_{-1}, \dots$ forms a complete orthonormal system in the Hilbert space $L_2(-\pi,\pi)$.*

The corresponding Fourier series in the Hilbert space $L_2(-\pi,\pi)$ reads as

$$\psi = \sum_{p=-\infty}^{\infty} a_p(\psi) \varphi_p \tag{10.6}$$

with $a_p(\psi) := \langle \psi | \varphi_p \rangle$.

Convergence of the Fourier series. The series (10.6) corresponds to the classical Fourier series

$$\boxed{\psi(x) = \frac{1}{\sqrt{2\pi}} \sum_{p=-\infty}^{\infty} a_p(\psi) \mathrm{e}^{\mathrm{i}px}} \tag{10.7}$$

along with the Fourier coefficients

$$a_p(\psi) := \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \psi(x) \mathrm{e}^{-\mathrm{i}px} dx, \qquad p = 0, \pm 1, \pm 2, \dots$$

The formulas for the derivatives are given by

$$D^\alpha \psi(x) = \frac{1}{\sqrt{2\pi}} \sum_{p=-\infty}^{\infty} (\mathrm{i}p)^\alpha a_p(\psi) \mathrm{e}^{\mathrm{i}px} \tag{10.8}$$

where we set $D := \frac{d}{d\varphi}$, and we choose $\alpha = 1, 2, \dots$ The following hold:

[6] We use the Lebesgue measure on the interval $]-\pi,\pi[$.

(i) If $\psi \in L_2(-\pi, \pi)$, then the Fourier series (10.7) converges in the Hilbert space $L_2(-\pi, \pi)$. Explicitly,

$$\lim_{n\to\infty} \int_{-\pi}^{\pi} \left| \psi(x) - \frac{1}{\sqrt{2\pi}} \sum_{p=-n}^{n} a_p(\psi) \mathrm{e}^{\mathrm{i}px} \right|^2 dx = 0.$$

This is called mean-square convergence.

(ii) Let the function $\psi : \mathbb{R} \to \mathbb{C}$ be 2π-periodic. The function ψ is smooth iff for each positive integer M, we have the decay condition

$$a_p(\psi) = O\left(\frac{1}{p^M}\right), \qquad p \to +\infty. \tag{10.9}$$

In this case, the Fourier series (10.7) and (10.8) converge uniformly and absolutely on the real line for all derivatives, $\alpha = 1, 2, \ldots$

The Hilbert space l_2. By definition, the space l_2 consists of all sequences $(a_0, a_1, a_{-1}, a_2, a_{-2}, \ldots)$ of complex numbers with

$$|a_0|^2 + |a_1|^2 + |a_{-1}|^2 + \ldots < \infty.$$

This is a complex infinite-dimensional Hilbert space equipped with the inner product

$$\langle a|b\rangle := \sum_{p=-\infty}^{\infty} a_p^\dagger b_p.$$

This series is absolutely convergent. Define the discrete Fourier transform by setting $\mathcal{F}\psi := (a_0(\psi), a_1(\psi), a_{-1}(\psi), \ldots)$.

Theorem 10.10 *The operator $\mathcal{F} : L_2(-\pi, \pi) \mapsto l_2$ is unitary.*

We call l_2 the momentum space. The derivative D^α, $\alpha = 1, 2, \ldots$, corresponds to the multiplication operator $a_p \mapsto (\mathrm{i}p)^\alpha a_p$ in the momentum space.

The convolution theorem. Let $f, g, h : \mathbb{R} \to \mathbb{C}$ be 2π-periodic functions that are integrable over the interval $[-\pi, \pi]$. We define the convolution

$$(f * g)(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x - y) g(y) dy, \qquad x \in \mathbb{R}.$$

The following properties are met:

- Consistency: $f * g : \mathbb{R} \to \mathbb{C}$ is 2π-periodic and continuous.
- Commutativity: $f * g = g * f$.
- Associativity: $(f * g) * h = f * (g * h)$.
- Linearity: $(\alpha f + \beta g) * h = \alpha f * h + \beta g * h$ for all complex numbers α, β.

Theorem 10.11 *For all $n = 0, \pm 1, \pm 2, \ldots$,*

$$\mathcal{F}(f * g)(n) = (\mathcal{F}f)(n)(\mathcal{F}g)(n).$$

The proof can be found in Stein and Shakarchi (2003), Vol. 1, Sect. 2.3.

10.3.3 Continuous Fourier Transform

Let us first summarize the key formulas which will be used frequently in this monograph. The validity of these formulas will be discussed below.

Key formulas on the real line. For the Fourier transform $\mathcal{F}\psi : \mathbb{R} \to \mathbb{C}$ of a function $\psi : \mathbb{R} \to \mathbb{C}$ on the real line, the following are met.

(i) Fourier transform:

$$\boxed{(\mathcal{F}\psi)(p) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) \mathrm{e}^{-\mathrm{i}px}\, dx, \qquad p \in \mathbb{R}.} \tag{10.10}$$

(ii) Inverse Fourier transform:

$$\psi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mathcal{F}\psi)(p) \mathrm{e}^{\mathrm{i}px}\, dx, \qquad x \in \mathbb{R}. \tag{10.11}$$

(iii) Transformation of derivatives: For $\alpha = 1, 2, \ldots$,

$$\frac{d^\alpha \psi(x)}{dx^\alpha} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mathrm{i}p)^\alpha (\mathcal{F}\psi)(p) \mathrm{e}^{\mathrm{i}px}\, dx, \qquad x \in \mathbb{R}. \tag{10.12}$$

(iv) The Parseval equation:

$$\langle \psi | \chi \rangle = \int_{-\infty}^{\infty} \psi(x)^\dagger \chi(x) dx = \langle \mathcal{F}\psi | \mathcal{F}\chi \rangle. \tag{10.13}$$

Physical interpretation in quantum mechanics. In terms of physics, the original function ψ acts in position space, whereas the Fourier transform $\mathcal{F}\psi$ acts in momentum space. In quantum mechanics, the function $\psi = \psi(x)$ is the wave function of a quantum particle on the real line. The operator

$$\boxed{(P\psi)(x) := -\mathrm{i}\hbar \frac{d}{dx}\, \psi(x)}$$

represents the momentum operator of the particle. Setting

$$\psi_p(x) := \frac{\mathrm{e}^{\mathrm{i}px/\hbar}}{\sqrt{2\pi\hbar}},$$

we get $P\psi_p = p\psi_p$ for each real value p of momentum. To simplify notation, we set $\hbar = 1$.

The Fourier transform represents an expansion with respect to the eigenfunctions of the momentum operator.

The normalization factor $1/\sqrt{2\pi}$ of the eigenfunction ψ_p is dictated by the Dirac calculus (see Sect. 11.2 on page 589) and the fact that the operator $\mathcal{F}$ can be uniquely extended to a unitary operator on the Hilbert space $L_2(\mathbb{R})$.

By (iii) above, the momentum operator P corresponds to the multiplication operator

$$(\mathcal{F}\psi)(p) \mapsto p(\mathcal{F}\psi)(p)$$

in momentum space. This tells us that the Fourier transform is related to a diagonalization of the momentum operator, $P = -\mathrm{i}\hbar \frac{d}{dx}$. John von Neumann (1903–1957)

proved that each self-adjoint operator on a Hilbert space can be realized as multiplication operator on a suitable function space (see the von Neumann spectral theorem in Sect. 11.2.3 on page 678).

Key formulas on $\mathbb{R}^N$. For the Fourier transform $\mathcal{F}\psi : \mathbb{R}^N \to \mathbb{C}$ of a function $\psi : \mathbb{R}^N \to \mathbb{C}$, the following hold true.

(i) Fourier transform:

$$\boxed{(\mathcal{F}\psi)(p) := \frac{1}{(2\pi)^{N/2}} \int_{\mathbb{R}^N} f(x)\, \mathrm{e}^{-\mathrm{i}\langle p|x\rangle}\, d^N x, \qquad p \in \mathbb{R}^N.}$$

(ii) Inverse Fourier transform:

$$\psi(x) = \frac{1}{(2\pi)^{N/2}} \int_{\mathbb{R}^N} (\mathcal{F}\psi)(p)\, \mathrm{e}^{\mathrm{i}\langle p|x\rangle}\, d^N x, \qquad x \in \mathbb{R}^N. \tag{10.14}$$

(iii) Transformation of derivatives: For partial derivatives of arbitrary order,

$$\partial^\alpha \psi(x) = \frac{1}{(2\pi)^{N/2}} \int_{\mathbb{R}^N} \mathrm{i}^{|\alpha|} p^\alpha (\mathcal{F}\psi)(p)\, \mathrm{e}^{\mathrm{i}\langle p|x\rangle}\, d^N x, \qquad x \in \mathbb{R}^N.$$

(iv) The Parseval equation:

$$\langle\psi|\chi\rangle = \int_{\mathbb{R}^N} \psi(x)^\dagger \chi(x) d^N x = \int_{\mathbb{R}^N} (\mathcal{F}\psi)(x)^\dagger (\mathcal{F}\chi)(x) d^N x = \langle \mathcal{F}\psi|\mathcal{F}\chi\rangle.$$

This implies the modified Parseval equation

$$\int_{\mathbb{R}^N} \psi(x)(\mathcal{F}\chi)(x) d^N x = \int_{\mathbb{R}^N} (\mathcal{F}\psi)(x)\chi(x) d^N x.$$

Here, we use the following notation

$$x = (x^1, \ldots, x^N), \;\; p = (p_1, \ldots, p_N), \;\; \langle p|x\rangle := \sum_{j=1}^N p_j x^j,$$

along with the norm $||x|| := \sqrt{(x^1)^2 + \ldots + (x^N)^2}$. Furthermore, we set

$$\partial_j := \frac{\partial}{\partial x^j}, \qquad j = 1, \ldots, N.$$

For partial derivatives of arbitrary order, we will use the symbol

$$\partial^\alpha \psi := \partial_1^{\alpha_1} \partial_2^{\alpha_2} \cdots \partial_N^{\alpha_N} = \frac{\partial^{|\alpha|}\psi}{(\partial x^1)^{\alpha_1} (\partial x^2)^{\alpha_2} \cdots (\partial x_N)^{\alpha_N}}. \tag{10.15}$$

Here, the multi-index $\alpha := (\alpha_1, \ldots, \alpha_N)$ has nonnegative integers $\alpha_1, \ldots, \alpha_N$ as components. Moreover, the order of the partial derivative ∂^α is given by the integer $|\alpha| := \alpha_1 + \ldots + \alpha_N$. Similarly,

$$p^\alpha := p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_N^{\alpha_N}.$$

To simplify notation, we include the trivial multi-index $\alpha = (0, \ldots, 0)$ which corresponds to the function itself, $\partial^\alpha \psi = \psi$.

The Schwartz space $\mathcal{S}(\mathbb{R}^N)$ of rapidly decreasing functions. In order to guarantee the existence of the Fourier integral, we have to assume that the functions ψ decrease rapidly at infinity. For example, the Gaussian function, given by $\psi(x) := \mathrm{e}^{-||x||^2/2}$, has the Fourier transform

$$(\mathcal{F}\psi)(p) = \frac{1}{(2\pi)^{N/2}} \int_{\mathbb{R}^N} \mathrm{e}^{-\mathrm{i}\langle p|x\rangle} \mathrm{e}^{-||x||^2/2}\, d^N x = \mathrm{e}^{-||p||^2/2}$$

for all $p \in \mathbb{R}^N$. This is a typical element of the space $\mathcal{S}(\mathbb{R}^N)$. By definition, the space $\mathcal{S}(\mathbb{R}^N)$ consists of all the smooth functions $\psi : \mathbb{R}^N \to \mathbb{C}$ which satisfy the decay condition

$$|\partial^\alpha \psi(x)| = O\left(\frac{1}{||x||^M}\right), \qquad ||x|| \to \infty$$

for all positive integers M and all multi-indices α.

Theorem 10.12 *The Fourier transform is a bijective map*

$$\mathcal{F} : \mathcal{S}(\mathbb{R}^N) \to \mathcal{S}(\mathbb{R}^N)$$

from the space $\mathcal{S}(\mathbb{R}^N)$ onto itself. In addition, the inverse Fourier transform is given by the classical formula above. For all $\psi, \chi \in \mathcal{S}(\mathbb{R}^N)$, both the Parseval equation and the modified Parseval equation are valid.

In order to formulate continuity properties of the Fourier transform, we introduce the semi-norms

$$|\psi|_{M,\alpha} := \sup_{x \in \mathbb{R}^N} (1 + ||x||^M)|D^\alpha \psi(x)|$$

for all integers M and all multi-indices α. We write

$$\lim_{n\to\infty} \psi_n = \psi \qquad \text{in} \quad \mathcal{S}(\mathbb{R}^N) \tag{10.16}$$

iff $\lim_{n\to+\infty} |\psi_n - \psi|_{M,\alpha} = 0$ for all possible M, α. The convergence (10.16) implies the convergence

$$\lim_{n\to\infty} \mathcal{F}(\psi_n) = \mathcal{F}(\psi) \qquad \text{in} \quad \mathcal{S}(\mathbb{R}^N).$$

The same is true for the inverse operator $\mathcal{F}^{-1}$. We say briefly that the operators $\mathcal{F}, \mathcal{F}^{-1} : \mathcal{S}(\mathbb{R}^N) \to \mathcal{S}(\mathbb{R}^N)$ are sequentially continuous.

The convolution theorem. Let $N = 1, 2, \ldots$. We are given the functions $\varphi, \psi, \chi \in \mathcal{S}(\mathbb{R}^N)$. The function $\varphi * \psi$ defined by

$$(\varphi * \psi)(x) := \int_{\mathbb{R}^N} \varphi(x - y)\psi(y) d^N y \qquad \text{for all} \quad x \in \mathbb{R}^N$$

is called the convolution of φ with ψ. We have the following properties:

- Consistency: $\varphi * \psi \in \mathcal{S}(\mathbb{R}^N)$.
- Commutativity: $\varphi * \psi = \psi * \varphi$.
- Associativity: $(\varphi * \psi) * \chi = \varphi * (\psi * \chi)$.
- Linearity: $(\alpha\varphi + \beta\psi) * \chi = \alpha\varphi * \chi + \beta\psi * \chi$ for all complex numbers α, β.

The proof of the following theorem can be found in the monograph by Hörmander (1983), Vol. 1, Sect. 7.1.

Theorem 10.13 *For the Fourier transform,*

$$\mathcal{F}(\varphi * \psi) = (2\pi)^{N/2}(\mathcal{F}\varphi)(\mathcal{F}\psi).$$

Conversely, $\mathcal{F}(\varphi\psi) = (2\pi)^{-N/2}(\mathcal{F}\varphi) * (\mathcal{F}\psi)$.

This theorem tells us that up to a real factor, the Fourier transform sends convolutions to products and, conversely, products to convolutions.

Discussion of terminology.. In order to avoid confusion, the reader should note that there exist different variants of the Fourier transform in the literature. To discuss this, let us start with the rescaled Fourier transformation formulas

$$\boxed{\begin{aligned}\hat{\varphi}(\xi) &= \frac{\alpha}{(2\pi)^{N/2}} \int_{\mathbb{R}^N} \varphi(x) \mathrm{e}^{-\mathrm{i}\langle x|\xi\rangle} d^N x, \\ \varphi(x) &= \frac{1}{\alpha(2\pi)^{N/2}} \int_{\mathbb{R}^N} \hat{\varphi}(\xi) \mathrm{e}^{\mathrm{i}\langle x|\xi\rangle} d^N \xi.\end{aligned}} \tag{10.17}$$

Here, $x = (x^1, \ldots, x^N)$ and $\xi = (\xi_1, \ldots, \xi_N)$ lie in $\mathbb{R}^N$, and

$$\langle x|\xi\rangle := \sum_{j=1}^{N} x^j \xi_j. \tag{10.18}$$

The normalization factor α is positive. Setting $\mathcal{F}_\alpha \varphi := \hat{\varphi}$, we obtain the following convolution formulas

$$\mathcal{F}_\alpha(\varphi * \psi) = \frac{(2\pi)^{N/2}}{\alpha} \, (\mathcal{F}_\alpha \varphi)(\mathcal{F}_\alpha \psi) \tag{10.19}$$

and

$$\mathcal{F}_\alpha(\varphi\psi) = \frac{1}{\alpha(2\pi)^{N/2}} \, (\mathcal{F}_\alpha \varphi) * (\mathcal{F}_\alpha \psi). \tag{10.20}$$

Note the following.

- The choice $\alpha = 1$ yields symmetric transformation formulas (10.17) which we have used in (10.14) above. This is called the Euclidean Fourier transform, since $\langle x|\xi\rangle$ represents the inner product on the N-dimensional Euclidean space $\mathbb{R}^N$. The choice $\alpha = 1$ has the advantage that the Fourier transform $\mathcal{F}\varphi := \hat{\varphi}$ is a unitary operator $\mathcal{F} : L_2(\mathbb{R}^N) \to L_2(\mathbb{R}^N)$ on the complex Hilbert space $L_2(\mathbb{R}^N)$. This is important, since unitary operators between Hilbert spaces preserve the structure of the corresponding quantum physics.
- However, the choice $\alpha = 1$ has the disadvantage that there appears an additional factor in the convolution rule (10.19). To avoid this factor, one has to choose $\alpha = 1/(2\pi)^{N/2}$.

From the physical point of view, the variables x and ξ possess different physical meaning in different situations. Let us discuss some typical examples.

(i) The Fourier transform from 1-dimensional position space to 1-dimensional momentum space:

$$N = 1, \quad x \Rightarrow x, \quad \xi \Rightarrow \frac{p}{\hbar}, \quad \alpha := \frac{1}{\hbar^{N/2}} = \frac{1}{\sqrt{\hbar}}.$$

From (10.17) we get

$$\hat{\varphi}\left(\frac{p}{\hbar}\right) = \frac{\alpha}{\sqrt{2\pi}} \int_{\mathbb{R}} \varphi(x) \mathrm{e}^{-\mathrm{i}xp/\hbar} dx,$$
$$\varphi(x) = \frac{1}{\alpha\sqrt{2\pi}} \int_{\mathbb{R}} \hat{\varphi}\left(\frac{p}{\hbar}\right) \mathrm{e}^{\mathrm{i}xp/\hbar} \frac{dp}{\hbar}.$$

Replacing $\hat{\varphi}\left(\frac{p}{\hbar}\right)$ by $\hat{\varphi}(p)$, we obtain

$$\boxed{\begin{aligned} \hat{\varphi}(p) &= \frac{1}{\sqrt{2\pi\hbar}} \int_{\mathbb{R}} \varphi(x) \mathrm{e}^{-\mathrm{i}xp/\hbar} dx, \\ \varphi(x) &= \frac{1}{\sqrt{2\pi\hbar}} \int_{\mathbb{R}} \hat{\varphi}(p) \mathrm{e}^{\mathrm{i}xp/\hbar} dp. \end{aligned}} \tag{10.21}$$

(ii) The Fourier transform from 3-dimensional position space to 3-dimensional momentum space:

$$N = 3, \quad x \Rightarrow \mathbf{x}, \quad \xi \Rightarrow \frac{\mathbf{p}}{\hbar}, \quad \alpha := \frac{1}{\hbar^{N/2}} = \frac{1}{\hbar^{3/2}}.$$

Hence

$$\boxed{\begin{aligned} \hat{\varphi}(\mathbf{p}) &= \frac{1}{(2\pi\hbar)^{3/2}} \int_{\mathbb{R}^3} \varphi(\mathbf{x}) \mathrm{e}^{-\mathrm{i}\mathbf{xp}/\hbar} d^3x, \\ \varphi(\mathbf{x}) &= \frac{1}{(2\pi\hbar)^{3/2}} \int_{\mathbb{R}^3} \hat{\varphi}(\mathbf{p}) \mathrm{e}^{\mathrm{i}\mathbf{xp}/\hbar} d^3p. \end{aligned}} \tag{10.22}$$

(iii) The Fourier transform from time space to energy space:

$$N = 1, \qquad x \Rightarrow t, \quad \xi \Rightarrow -\frac{E}{\hbar}, \quad \alpha := \frac{1}{\hbar^{N/2}} = \frac{1}{\sqrt{\hbar}}.$$

This substitution implies

$$\boxed{\begin{aligned} \hat{\varphi}(E) &= \frac{1}{\sqrt{2\pi\hbar}} \int_{\mathbb{R}} \varphi(t) \mathrm{e}^{\mathrm{i}tE/\hbar} dt, \\ \varphi(t) &= \frac{1}{\sqrt{2\pi\hbar}} \int_{\mathbb{R}} \hat{\varphi}(E) \mathrm{e}^{-\mathrm{i}tE/\hbar} dE. \end{aligned}} \tag{10.23}$$

(iv) The Fourier transform from time space to angular frequency space: Use (10.23) and set $\hbar := 1$ along with $E := \omega$.

(v) The Fourier transform from the 4-dimensional Minkowski space for space and time to the 4-dimensional momentum space:

$$N = 4, \quad x \Rightarrow (ct, \mathbf{x}), \quad \xi \Rightarrow \left(-\frac{E}{c\hbar}, \frac{\mathbf{p}}{\hbar}\right), \quad \alpha := \frac{1}{\hbar^{N/2}} = \frac{1}{\hbar^2}.$$

Set $x := (x^0, x^1, x^2, x^3)$ with $x^0 := ct$, as well as $p = (p^0, p^1, p^2, p^3)$ with $p^0 = p_0 := E/c$ and $p^j = -p_j$ for $j = 1, 2, 3$. Furthermore, define

$$xp := \sum_{j=0}^{3} x^j p_j = Et - \mathbf{xp}.$$

Then $\mathrm{e}^{\mathrm{i}px/\hbar} = \mathrm{e}^{-\mathrm{i}Et/\hbar}\,\mathrm{e}^{\mathrm{i}\mathbf{p}\mathbf{x}/\hbar}$. Hence

$$\boxed{\begin{aligned}\hat{\varphi}\,(p) &= \frac{1}{(2\pi\hbar)^2}\int_{\mathbb{R}^4}\varphi(x)\mathrm{e}^{\mathrm{i}xp/\hbar}d^4x,\\ \varphi(x) &= \frac{1}{(2\pi\hbar)^2}\int_{\mathbb{R}^4}\hat{\varphi}(p)\mathrm{e}^{-\mathrm{i}xp/\hbar}d^4p.\end{aligned}} \tag{10.24}$$

This so-called Fourier–Minkowski transform generalizes the Fourier transform from (ii) and (iii).

(vi) The Fourier–Laplace transform:

$$N = 1, 2, \ldots, \quad x \in \mathbb{R}^N, \quad \xi \in \mathbb{C}^N, \quad \alpha = 1.$$

Hence

$$\boxed{\begin{aligned}\hat{\varphi}(\xi) &= \frac{1}{(2\pi)^{N/2}}\int_{\mathbb{R}^N}\varphi(x)\mathrm{e}^{-\mathrm{i}\langle x|\xi\rangle}d^Nx,\\ \varphi(x) &= \frac{1}{(2\pi)^{N/2}}\int_{\mathbb{R}^N}\hat{\varphi}(\xi)\mathrm{e}^{\mathrm{i}\langle x|\xi\rangle}d^N\xi.\end{aligned}} \tag{10.25}$$

Here, the inner product $\langle x|\xi\rangle$ is given by (10.18). The point is that in contrast to the Fourier transform, the variable ξ does not live in the real space $\mathbb{R}^N$, but in the complex space $\mathbb{C}^N$. The Fourier–Laplace transform can be regarded as an analytic extension of the Fourier transform.

10.4 The Dirichlet Problem in Electrostatics as a Paradigm

> The Dirichlet principle is an exciting example for a problem that came from physics and could be solved by mathematics; this famous problem strongly influenced the development of far-reaching mathematical theories in the 20th century.
>
> Folklore

> By the Dirichlet principle we understand a method for solving boundary value problems via minimum problems for variational integrals. This principle goes back to Green (1793–1841), Gauss (1777–1855), Lord Kelvin (1824–1907), and Dirichlet (1805–1859). In 1870 Weierstrass (1815–1897) was the first to underline the shortcomings of this principle. He showed that there are variational problems which do not have any solution. In 1900 I showed that it is possible to rigorously justify the Dirichlet principle.[7]
>
> David Hilbert (1862–1943)

In order to understand the great achievement of Hilbert in the field of analysis, it is necessary to first comment on the state of analysis at the end of the nineteenth century. After Weierstrass had made sure of the foundations of complex function theory, and it has reached an impressive level, research switched to boundary-value problems, which first arose in

[7] D. Hilbert, On the Dirichlet principle (in German), Math. Ann. **59** (1904), 161–186.

physics (e.g., electrostatics). The work of Riemann (1826–1866) on complex function theory and conformal maps, however, had shown that boundary-value problems have great importance for pure mathematics as well. Two problems had to be solved:

(i) the problem of the existence of an electrostatic potential function for given boundary values, and

(ii) the problem of eigenoscillations of elastic bodies, for example, string and membrane.

The state of the theory was bad at the end of the nineteenth century. Riemann had believed that, by using the Dirichlet principle, one could deal with these problems in a simple and uniform way. After Weierstrass' substantial criticism of the Dirichlet principle in 1870, special methods had to be developed for these problems. These methods, by Carl Neumann, Amandus Schwarz, and Henri Poincaré, were very elaborate and still have great aesthetic appeal today; but because of their variety they were confusing, although at the end of the nineteenth century, Poincaré (1854–1912), in particular, endeavored with great astuteness to standardize the theory. There was, however, a lack of "simple basic facts" from which one could easily get complete results without sophisticated investigations of limiting processes.

Hilbert first looked for these "simple basic facts" in the calculus of variations. In 1900 he had an immediate and great success; he succeeded in justifying the Dirichlet principle.

While Hilbert used variational methods, the Swedish mathematician Fredholm (1866–1927) approached the same goal by developing Poincaré's work by using linear integral equations. In the winter semester 1900/01 Holmgren, who had come from Uppsala (Sweden) to study under Hilbert in Göttingen (Germany), held a lecture in Hilbert's seminar on Fredholm's work on linear integral equations which had been published the previous year. This was a decisive day in Hilbert's life. He took up Fredholm's discovery with great zeal, and combined it with his variational method. In this way he succeeded in creating a uniform theory which solved problems (i) and (ii) above.[8]

Hilbert believed that with this theory he had provided analysis with a great general basis which corresponds to an axiomatics of limiting processes. The further development of mathematics has proved him to be right.[9]

Otto Blumenthal, 1932

The creation of a rigorous mathematical quantum field theory is a challenge for modern mathematics. There is hope for solving this hard problem in the future. The optimism is motivated by the success of mathematics in the past. As an example, let us consider the Dirichlet principle in this section.

Physical motivation. Let Ω be a nonempty bounded open set in $\mathbb{R}^3$. We want to study electric fields $\mathbf{E} = -\,\mathbf{grad}\, U$ which are generated by a charge density ϱ. The corresponding Maxwell equation in a vacuum, $\varepsilon_0 \operatorname{div} \mathbf{E} = \varrho$, reads as

$$\boxed{\varepsilon_0 \Delta U = \varrho \text{ on } \Omega, \qquad U = U_0 \text{ on } \partial\Omega} \tag{10.26}$$

[8] See D. Hilbert, Foundations of the Theory of Integral Equations, Teubner, Leipzig, 1912 (in German).

[9] O. Blumenthal, Hilbert's biography (in German). In: D. Hilbert, Collected Works, Springer, Berlin, 1932, Vol. 3, pp. 388–429.
Hilbert's friend Otto Blumenthal (1876–1944) was murdered in the Nazi concentration camp Theresienstadt (Terežin).

with the Laplacian $\Delta U := -U_{xx} - U_{yy} - U_{zz}$. The boundary-value problem (10.26) is called the Dirichlet problem for the Poisson equation. The electric energy in the region Ω is given by the following integral

$$E(U) := \int_\Omega \frac{\varepsilon_0}{2} \, (U_x^2 + U_y^2 + U_z^2) - \varrho U \;\; dxdydz.$$

In particular, if $\varrho = 0$, then $E(U) = \frac{\varepsilon_0}{2} \int_\Omega \mathbf{E}^2 \; dxdydz$. The principle of minimal energy reads as

$$\boxed{E(U) = \min!, \qquad U = U_0 \;\text{ on }\; \partial\Omega.} \tag{10.27}$$

We are given the charge density ϱ and the boundary values U_0 of the potential U. We are looking for a potential U such that the electric energy is minimal.

From the physical point of view, it seems to be obvious that there exists a solution for problem (10.27).

However, from the mathematical point of view, the existence proof is a highly nontrivial task. In the first half of the 19th century, mathematicians and physicists studied gravitational, electric and magnetic forces in the setting of potential theory.[10] Green and Gauss independently published two fundamental papers on the properties of gravitational and electrostatic forces in 1828 and 1840, respectively. However, the full understanding of such problems was only completed in the 1960s on the basis of functional analytic interpolation theory.[11] For the history of the Dirichlet principle, we refer to the following two beautiful articles:

S. Hildebrandt, Remarks on the Dirichlet Principle (in German). In: H. Weyl, Die Idee der Riemannschen Fläche (The concept of a Riemann surface). Edited by R. Remmert, Teubner, Leipzig, 1997, pp. 197–215 (in German).

H. Brézis and F. Browder, Partial differential equations in the 20th century, Advances in Math. **135** (1998), 76–144.

System of units. To simplify notation, in this section we choose a system of units where the electric field constant of the vacuum is equal to one, $\varepsilon_0 = 1$.

10.4.1 The Variational Lemma

The space $\mathcal{D}(\Omega)$ of test functions. For the calculus of variations and the theory of generalized functions (distributions), it is typical to use test functions. In order to fix the notation, let Ω be a nonempty open subset of the N-dimensional space $\mathbb{R}^N$ where $N = 1, 2, \ldots$

(i) The space $C^\infty(\Omega)$ of smooth functions: This space consists of all smooth functions $f : \Omega \to \mathbb{C}$. This means that f is continuous on Ω and has partial derivatives of arbitrary order which are continuous on Ω, too.

[10] Laplace (1749–1827), Gauss (1777–1855), Poisson (1781–1859), Green (1793–1841), Dirichlet (1805–1859), Weierstrass (1815–1897), Riemann (1826–1866), Hilbert (1862–1943), Tonelli (1885–1946), Weyl (1885–1955), Sobolev (1908–1989).

[11] J. Lions and E. Magenes, Nonhomogeneous Boundary-Value Problems and Applications, Vols. 1–3, Springer, New York, 1972.

(ii) The space $\mathcal{D}(\Omega)$ of smooth functions which have compact support: By definition, the smooth function $f : \Omega \to \mathbb{C}$ has compact support iff it vanishes outside a compact subset of Ω. The set of these functions is denoted by the symbol $\mathcal{D}(\Omega)$. In other words, we have

$$f \in \mathcal{D}(\Omega)$$

iff $f \in C^\infty(\Omega)$ and the function f vanishes in some open neighborhood of the boundary $\partial\Omega$. Intuitively, these test functions do not see the boundary $\partial\Omega$.

(iii) The space $C^\infty(\overline{\Omega})$ of functions which are smooth up to the boundary: Recall that the closure $\overline{\Omega}$ of the set Ω is obtained from Ω by adding the boundary,

$$\overline{\Omega} = \Omega \cup \partial\Omega.$$

By definition, the function $f : \Omega \to \mathbb{C}$ lives in the space $C^\infty(\overline{\Omega})$ iff $f \in C^\infty(\Omega)$ and the function f, together with all of its partial derivatives of arbitrary order, can be continuously extended to the closure $\overline{\Omega}$.

Obviously, we have the inclusions

$$\mathcal{D}(\Omega) \subseteq C^\infty(\overline{\Omega}) \subseteq C^\infty(\Omega).$$

The space $\mathcal{D}(\Omega)$ was used by Laurent Schwartz in the 1940s in order to introduce the space $\mathcal{D}'(\Omega)$ of generalized functions (see Chap. 11). In the literature, one also uses the symbol $C_0^\infty(\Omega)$ instead of $\mathcal{D}(\Omega)$. The following theorem describes the typical property of test functions.

Theorem 10.14 *The set of test functions $\mathcal{D}(\Omega)$ is dense in the Hilbert space $L_2(\Omega)$.*

Explicitly, this means the following. For each given function $\varphi \in L_2(\Omega)$, there exists a sequence (φ_n) of test functions, $\varphi_n \in \mathcal{D}(\Omega)$ for all n, such that

$$\lim_{n\to\infty} ||\varphi - \varphi_n||^2 = \lim_{n\to\infty} \int_\Omega |\varphi(x) - \varphi_n(x)|^2 d^N x = 0.$$

The proof can be found in Zeidler (1995), Vol. 1, Sect. 2.2.3.

The real variational lemma. The following result is crucial for the calculus of variations (i.e., the principle of critical action in physics).

Proposition 10.15 *We are given the continuous function $f : \Omega \to \mathbb{C}$. Suppose that*

$$\int_\Omega f(x)h(x)d^N x = 0$$

for all smooth functions $h : \Omega \to \mathbb{R}$ which have compact support. Then, the function f vanishes on Ω.

Proof. Suppose first that the set Ω is bounded. Choosing $h := \Re(\varphi)$ and $h := \Im(\varphi)$, we get

$$\langle\varphi|f\rangle = \int_\Omega \varphi(x)^\dagger f(x) d^N x = 0 \qquad \text{for all} \quad \varphi \in \mathcal{D}(\Omega).$$

Let $g \in L_2(\Omega)$. By Theorem 10.14, there exists a sequence (φ_n) of test functions $\varphi_n \in \mathcal{D}(\Omega)$ for all n such that $\varphi_n \to g$ in $L_2(\Omega)$ as $n \to \infty$. It follows from

$$\langle\varphi_n|f\rangle = 0, \qquad n = 1, 2, \ldots$$

that $\langle g|f\rangle = \lim_{n\to\infty}\langle\varphi_n|f\rangle = 0$. Choosing $g = f$, we get $f = 0$ in $L_2(\Omega)$. This tells us that $f(x) = 0$ for almost all $x \in \Omega$. Since the function f is continuous, this implies $f(x) = 0$ for all $x \in \Omega$.

If the set Ω is unbounded (e.g. $\Omega = \mathbb{R}^N$), then we choose an open ball B_R of radius R, and we redefine $f(x) := 0$ outside B_R. Then, the argument above shows that $f = 0$ on $\Omega \cap B_R$. Choosing $R = 1, 2, \ldots$, we obtain that $f = 0$ on Ω. □

The complex variational lemma. Since the fields arising in quantum physics are complex-valued, the following variant of the variational lemma is useful for the principle of critical action.

Proposition 10.16 *We are given the continuous functions $f, g : \Omega \to \mathbb{C}$. Suppose that*

$$\int_\Omega (f(x)h(x) + g(x)h(x)^\dagger)d^N x = 0$$

for all smooth functions $h : \Omega \to \mathbb{C}$ which have compact support. Then the functions f and g vanish on Ω.

Proof. First let h be real. By the real variational lemma above,

$$f(x) + g(x) = 0 \qquad \text{on } \Omega.$$

Secondly, if $h(x) := \mathrm{i}k(x)$ where k is real, then $f(x) - g(x) = 0$ on Ω. This implies $f(x) = 0$ and $g(x) = 0$ on Ω. □

10.4.2 Integration by Parts

> Integration by parts is the key to the calculus of variations and to the modern theory of linear and nonlinear partial differential equations.
>
> Folklore

> Gauss (1777–1855) had very early been interested in the Laplace equation, both in two variables in connection with his work on complex numbers, and in three variables in relation with his astronomical studies. In his 1813 paper on the attraction of spheroids, he had proved particular cases of the Green formula. After 1830, he devoted much of his time to the study of magnetism, both experimentally and theoretically, and thus was led to new research on potential theory which he published in 1839. In that paper, he quotes no other work on the subject, and it is very unlikely that he ever heard of Green (1793–1841) whose work was not widely known, even in England.
>
> Jean Dieudonné (1906–1992)
> *History of Functional Analysis*[12]

> Green applied his concepts based on the Green's integral formula and the Green's function to electrical and magnetic problems.[13] He also took up

[12] North Holland, Amsterdam, 1981 (reprinted with permission).

[13] G. Green, Essay on the application of mathematical analysis to the theory of electricity and magnetism, 1828. Reprinted in Ostwald's Klassiker No. 61, Geest & Portig, Leipzig, 1895. This famous essay founded potential theory as a branch of mathematics and physics.
C. F. Gauss, General theorems on attractive and repulsive forces which act according to the inverse square of the distance (in German), 1839. See Gauss (1863/1933), Collected Works, Vol. V, pp. 197–242.

> in 1833 the problem of the gravitational potential of ellipsoids of variable density. In this work Green showed that when the potential U is given on the boundary of a body, there is just one function that satisfies $\Delta U = 0$ throughout the body, has no singularities, and has the given boundary values. To make his proof, Green assumed the existence of a function that minimizes the integral
>
> $$\int U_x^2 + U_y^2 + U_z^2 \, dxdydz.$$
>
> This is the first use of the Dirichlet principle. Green's work strongly inspired the great Cambridge school of mathematical physicists which included Sir Gabriel Stokes (1819–1903), Lord Kelvin (1824–1907), Lord Rayleigh (1842–1919), and Clerk Maxwell (1831–1879).
>
> Morris Kline, 1972
> *Mathematical Thought from Ancient to Modern Times*[14]

The 1-dimensional case. Let $-\infty < a < b < \infty$. Newton (1643–1727) and Leibniz (1646–1716) discovered the formula

$$\boxed{\int_a^b f'(x)dx = f|_a^b} \tag{10.28}$$

which is called the fundamental theorem of calculus. Here, we set

$$f|_a^b := f(b) - f(a).$$

The idea is to reduce an integral over the interval $\Omega :=]a, b[$ to an "integral over the boundary" $\partial\Omega$. For example, formula (10.28) is valid if the function $f : [a, b] \to \mathbb{R}$ is smooth up to the boundary. If the functions $f, g : [a, b] \to \mathbb{R}$ are smooth up to the boundary, then $(fg)' = f'g + fg'$. By (10.28),

$$\int_a^b f'g + fg' \, dx = fg|_a^b.$$

This implies the following integration-by-parts formula

$$\boxed{\int_a^b f'gdx = -\int_a^b fg'dx + fg|_a^b.} \tag{10.29}$$

In particular, the boundary term drops out if the function f has compact support in the open interval Ω, that is, $f \in \mathcal{D}(\Omega)$. Conversely, the integration-by-parts formula (10.29) implies (10.28) by setting $g = 1$.

The N-dimensional case. It is our goal to generalize the fundamental theorem of calculus and hence the integration-by-parts formula to higher dimensions. To this end, let Ω be a nonempty bounded open subset of $\mathbb{R}^N$ with $N = 2, 3 \ldots$ Set $x := (x^1, \ldots, x^N)$ and $\partial_j := \frac{\partial}{\partial x^j}$ if $x \in \mathbb{R}^N$.

(i) Suppose that $f, g : \Omega \to \mathbb{R}$ are smooth functions where f has compact support. Then, for $j = 1, \ldots, N$,

$$\boxed{\int_\Omega \partial_j f \cdot g \, d^N x = -\int_\Omega f\partial_j g \, d^N x.} \tag{10.30}$$

[14] Reprinted by permission of Oxford University Press.

(ii) Suppose that the functions $f, g : \mathrm{cl}(\Omega) \to \mathbb{R}$ are continuous. In addition, assume that f and g are smooth on Ω, and their first partial derivatives can be extended to continuous functions on $\mathrm{cl}(\Omega)$. Furthermore, suppose that the boundary of Ω is sufficiently regular (e.g., Ω is a ball or a cuboid). Then, for $j = 1, \ldots, N$,

$$\boxed{\int_\Omega \partial_j f \cdot g \; d^N x = -\int_\Omega f \partial_j g \; d^N x + \int_{\partial\Omega} f g n_j \; dS.} \tag{10.31}$$

Here, $n(x) = (n_1(x), \ldots, n_N(x))$ is the outer unit normal vector at the boundary point $x \in \partial\Omega$, and dS refers to the surface measure.

For the precise formulation one needs the precise definition of the admissible boundaries. We distinguish two cases.

(a) Smooth boundaries: Here, we assume that the closure $\mathrm{cl}(\Omega) = \Omega \cup \partial\Omega$ is a manifold with boundary. Roughly speaking, each boundary point x has a neighborhood in $\mathrm{cl}\,(\Omega)$ which, in suitable local coordinates, looks like the sufficiently small neighborhood of a boundary point y of the half-space

$$\{(y^1, \ldots, y^N) \in \mathbb{R}^N : y^N \geq 0\},$$

and the change of local coordinates is carried out by diffeomorphisms which are smooth up to the boundary of the half-space. The precise definition of the standard term "manifold with boundary" can be found in Zeidler (1986), Vol. 4, p. 584. The prototype of a manifold with boundary is a closed ball in $\mathbb{R}^N$.

(b) Piecewise smooth boundaries: Note that a cuboid in $\mathbb{R}^3$ is not a "manifold with boundary" because of the corners and edges. However, the integration-by-parts formula is also valid for cuboids and a general class of bounded open sets in $\mathbb{R}^N$ with "piecewise smooth" boundaries (i.e., Lipschitz-continuous boundaries). The precise formulation of the integration-by-parts formula in this case along with the proof can be found in J. Nečas, Les méthodes directes en théorie des equations elliptiques, Academia, Prague, p. 121.

The general Stokes theorem. The integration-by-parts formula dates back to Lagrange (1736–1813) and Gauss (1777–1855). The formula

$$\boxed{\int_\Omega \mathrm{div}\, \mathbf{v} \; d^3 x = \int_{\partial\Omega} \mathbf{v}\mathbf{n} \; dS} \tag{10.32}$$

for velocity fields $\mathbf{v}$ is called the Gauss theorem. Here, $\mathbf{n}$ denotes the outer unit normal vector at boundary points. From (10.32) we immediately get the integration-by-parts formula by choosing $\mathbf{v} := (fg, 0, 0)$. Then we obtain $\mathrm{div}\, \mathbf{v} = f_x g + g_x f$. Green (1793–1848) based his fundamental 1828 paper, concerning the Green's function in potential theory, on the Green's integral formula

$$\int_\Omega (U \Delta V - V \Delta U) \; d^3 x = \int_{\partial\Omega} \left(V \frac{\partial U}{\partial n} - U \frac{\partial V}{\partial n} \right) \; dS \tag{10.33}$$

which follows from repeated integration by parts.[15] Stokes (1819–1903) announced the integral relation

[15] Note that $\Delta U := -\sum_j \partial_j^2 U$ and $\frac{\partial U}{\partial n} := \sum_j n_j \partial_j U$.
In 1828 the Green's integral formula (10.33) was obtained independently by Ostrogradski (1801–1862).

$$\int_S \mathbf{N}\,\mathbf{curl}\,\mathbf{v}\; dS = \int_{\partial S} \mathbf{v} d\mathbf{x} \tag{10.34}$$

in 1854. Here, $\mathbf{N}$ denotes the unit normal vector of the surface S in $\mathbb{R}^3$, and the boundary curve ∂S is counterclockwise oriented with respect to $\mathbf{N}$. For a region in the Euclidean plane, the Stokes formula (10.34) coincides with the 2-dimensional version of the Gauss formula (10.32). The modern variant of the Gauss formula (10.32) and the Stokes formula (10.34), basically due to Poincaré (1854–1912), reads elegantly as

$$\boxed{\int_M d\omega = \int_{\partial M} \omega.} \tag{10.35}$$

Here, ω is a smooth $(N-1)$-dimensional differential form on the N-dimensional manifold M with boundary. For the proof of the general Stokes theorem (10.34) and its numerous applications, we refer to the following monographs:

H. Amann and J. Escher, Analysis III, Sec. 12.3, Birkhäuser, Basel, 2004.

V. Zorich, Analysis II, Sec. 13.3, Springer, Berlin, 2003.

V. Guillemin and A. Pollack, Differential Topology, Sec. 4.7, Prentice Hall, Englewood Cliffs, New Jersey, 1974.

R. Bott and L. Tu, Differential Forms in Algebraic Topology, Sec. 1.3, Springer, New York, 1982.

In Volume III on gauge field theory, we will show that the formula (10.35) is crucial for understanding the topological structure of general potentials in physics, and it lies at the heart of cohomology theory which is crucial for modern topology and quantum field theory.

The general Stokes theorem is one of the most beautiful and most useful theorems in mathematics and physics.

10.4.3 The Variational Problem

One needs to have delved but little into the principles of differential calculus to know the method of how to determine the greatest and least ordinates of curves. But there are maxima or minima problems of a higher order, which in fact depend on the same method, which however can not be subjected to this method. These are the problems where it is a matter of finding the curves themselves.
The first problem of this type, which the geometers solved, is that of the brachistochrone or the curve of fastest fall, which Johann Bernoulli (1667–1748) proposed toward the end of the preceding century. One attained this only in special ways, and it was only some time later and on the occasion of the investigations concerning isoperimetric problems that the geometer of whom we just spoke and his excellent brother Jakob Bernoulli (1654–1705) gave some rules in order to solve several other problems of this type.
But since these rules were not of sufficient generality, the famous Euler (1707–1783) undertook in 1744 to refer all investigations of this type to a general method. But even as sophisticated and fruitful as his method is, one must nevertheless confess that it is not sufficiently simple... Now, here one finds a method which requires only a simple use of the principles of differential and integral calculus.

Comte de Joseph Louis Lagrange, 1762

> I am very happy that this theory which I have treated since the first attempts almost alone, has been brought precisely by you to the highest perfection.
>
> Euler in a letter to Lagrange

The Euler–Lagrange equation. Let Ω be a nonempty bounded open set in $\mathbb{R}^2$. We are given the continuous function $\varrho : \overline{\Omega} \to \mathbb{R}$ and the continuous boundary function $U_0 : \partial\Omega \to \mathbb{R}$.[16]

Theorem 10.17 *Each smooth solution $U : \overline{\Omega} \to \mathbb{R}$ of the variational problem*

$$\int_\Omega \tfrac{1}{2}(U_x^2 + U_y^2) - \varrho U \;\; dxdy = \min! \tag{10.36}$$

with the boundary condition $U = U_0$ on $\partial\Omega$ satisfies the differential equation

$$-U_{xx} - U_{yy} = \varrho \qquad \text{on} \quad \Omega$$

which is called the Euler–Lagrange equation to (10.36).

The following elegant proof is identical with Lagrange's classical argument.
Proof. Let $U \in C^\infty(\overline{\Omega})$ be a solution of the minimum problem. Choose a smooth test function $h : \Omega \to \mathbb{R}$ which has compact support. Replacing U by $U + \tau h$, we get

$$\mathcal{J}(\tau) := \int_\Omega \tfrac{1}{2}(U_x + \tau h_x)^2 + \tfrac{1}{2}(U_y + \tau h_y)^2 - \varrho(U + \tau h)\, dxdy, \qquad \tau \in \mathbb{R}.$$

The test function h vanishes on the boundary $\partial\Omega$. Therefore, the function $U + \tau h$ has the same boundary values on $\partial\Omega$ as the function U. Since U is a solution of the minimum problem (10.36), the function $\tau \mapsto \mathcal{J}(\tau)$ has a minimum at the point $\tau = 0$. This implies that the derivative of the real function $\mathcal{J}$ vanishes at the point $\tau = 0$, that is, $\mathcal{J}'(0) = 0$. Hence

$$\int_\Omega U_x h_x + U_y h_y - \varrho h \; dxdy = 0.$$

Integration by parts yields

$$\int_\Omega (-U_{xx} - U_{yy} - \varrho) h \; dxdy = 0$$

for all test functions h. By the variational lemma, $-U_{xx} - U_{yy} - \varrho = 0$ on Ω. □

Many important variational problems related to the principle of critical action in physics can be found in the problem section to Chap. 14 on page 805.

[16] Recall that the symbol $\overline{\Omega}$ denotes the closure of the set Ω. The closure is also denoted by $\mathrm{cl}(\Omega)$.

10.4.4 Weierstrass' Counterexample

Karl Weierstrass (1815–1897) studied law in Bonn. In 1839 he decided to study mathematics and therefore went to Münster, where he attended a mathematical lecture by Gudermann (1798–1852) on elliptic functions. This was the only mathematical lecture which he ever heard in his life. After that Weierstrass worked from 1840 until 1855 as schoolteacher, totally isolated, in small towns. During this time he published a series of fundamental papers on the most difficult problems of complex function theory.[17]

In his academic inaugural speech in 1857 as a newly elected member of the Berlin Academy, he described how the theory of elliptic functions had tremendously attracted him, even before his studies, and that he regarded it as one of the main tasks of mathematics to create a general theory on periodic functions of several complex variables (Abelian functions).

To prepare himself for this difficult task, he first studied the available tools and occupied himself with less difficult problems. As a result of these studies he published, during the years 1841–1843, papers on the theory of power series and on the definition of analytic functions by means of algebraic differential equations.

In 1854 Weierstrass first succeeded in finding the solution to the famous "inverse problem" posed by Jacobi (1804–1851) in 1832.[18] The generalization of this result to general Abelian integrals, which was later obtained by Riemann and then by Weierstrass, is one of the greatest achievements in analysis. Whereas Riemann based his theory on the imperfect Dirichlet principle, Weierstrass' theory is based on purely algebraic arguments.

From Hilbert's obituary for Weierstrass

Riemann (1826–1866) is the man with glowing intuition. Through his overall genius he stands far above all his peers. Everywhere that his interest was aroused, he started the theory from scratch without worrying about tradition or constraints of existing systems.

Weierstrass (1815–1897) was above all a logician; he works slowly, systematically, step by step. Where he works, he heads for a finished form of the theory.

Without new ideas and without the formulation of new objectives, mathematics would soon amount to nothing more than the rigidity of its proofs and would begin to stagnate, its fuel running out. So, in a certain sense,

[17] In 1864 Weierstrass became professor in mathematics at the Berlin University.

[18] Jacobi formulated the following hypothesis: Let $p = p(z)$ be a polynomial of degree six. Consider the two functions $u = u(a, b)$ and $v = v(a, b)$ of the two complex arguments a, b given by the following system of hyperelliptic integrals

$$\int_{u_0}^{u} \frac{dz}{\sqrt{p(z)}} + \int_{v_0}^{v} \frac{dz}{\sqrt{p(z)}} = a,$$
$$\int_{u_1}^{u} \frac{dz}{\sqrt{p(z)}} + \int_{v_1}^{v} \frac{dz}{\sqrt{p(z)}} = b.$$

Then the two functions $u+v$ and uv are univalent and they possess four different periods.

> mathematics has been furthered most by those who distinguished themselves more through their intuition than through strong deduction.[19]
>
> Felix Klein (1849–1925)

Riemann's argument. Riemann based his seminal approach to complex function theory (e.g., the Riemann mapping theorem) and to Abelian integrals on both the notion of Riemann surface and on the following boundary-value problem for the Laplace equation[20]

$$\Delta U = 0 \ \text{ on } \ \Omega, \qquad U = U_0 \ \text{ on } \ \partial\Omega \tag{10.37}$$

where Ω is a bounded region in $\mathbb{R}^2$. According to the Dirichlet principle, Riemann reduced the boundary-value problem (10.37) to the following minimum problem

$$\boxed{\int_\Omega U_x^2 + U_y^2 \, dxdy = \min!, \qquad U = U_0 \ \text{ on } \ \partial\Omega.} \tag{10.38}$$

He followed the opinion of his contemporaries that the existence of a solution for (10.37) is evident, by physics. In fact, if we set $\mathbf{E} = -\,\mathbf{grad}\, U$, then problem (10.38) concerns the determination of an electrostatic field $\mathbf{E}$ of minimal energy under given boundary values of the potential U. By physical intuition, such a field always exists.

The counterexample. Let $C^1[-1,1]$ be the space of all continuous functions $f : [-1,1] \to \mathbb{R}$ which have a continuous first derivative f' on the open interval $]-1,1[$, such that f' can be extended to a continuous function on the closed interval $[-1,1]$. In 1870 Weierstrass published the following variational problem

$$\boxed{\begin{gathered} E(f) := \int_{-1}^{1} (xf'(x))^2 dx = \min!, \quad f \in C^1[-1,1], \\ f(-1) = 1, \quad f(1) = 0. \end{gathered}} \tag{10.39}$$

Obviously, $E(f) \geq 0$ for all $f \in C^1[-1,1]$. Choose

$$f_n(x) := \frac{1}{2} - \frac{1}{2} \cdot \frac{\arctan nx}{\arctan n}, \qquad n = 1, 2, \ldots$$

Then, $f_n(-1) = 1$ and $f_n(1) = 0$ for all n. Moreover, $\lim_{n\to\infty} E(f_n) = 0$. Consequently, we have

$$\inf E(f) = 0$$

where the infimum is taken over all admissible functions $f \in C^1[-1,1]$ which satisfy the boundary conditions $f(-1) = 1$ and $f(1) = 0$.

Proposition 10.18 *Problem* (10.39) *has no solution.*

---

[19] F. Klein, Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert (Development of mathematics in the 19th century), Vol. 1, Springer, Berlin, 1926.

[20] B. Riemann, Foundations of a general theory of complex-valued functions (in German), Dissertation, Göttingen, 1851.
B. Riemann, Theory of Abelian functions (in German), J. reine und angew. Math. **54** (1857), 115–155.
These two fundamental papers can be found in the collected works of Riemann (1990).

Proof. Suppose that f is a solution. Then $E(f) = 0$. This implies that $xf'(x) = 0$ for all $x \in [-1, 1]$. Hence $f'(x) = 0$ for all $x \in [-1, 1]$ with $x \neq 0$. Since f is continuous on $[-1, 1]$, we get $f(x) = \text{const}$ for all $x \in [-1, 1]$. This implies the relation $f(-1) = f(1)$, which contradicts the boundary condition $f(-1) = 1$ and $f(1) = 0$. □

If we replace the boundary condition from (10.39) by $f(-1) = f(1) = 0$, then the modified problem has a solution, namely, $f = 0$.

The idea of completion. Let X be the set of all rational numbers living in the interval $[0, 2]$. For the two minimum problems

$$f(x) = \min!, \qquad x \in [0, 2], \tag{10.40}$$

and

$$f(x) = \min!, \qquad x \in X, \tag{10.41}$$

the following two statements can be proved.

(i) The Weierstrass theorem: If $f : [0, 2] \to \mathbb{R}$ is continuous, then problem (10.40) has a solution.
(ii) If $f : X \to \mathbb{R}$ is continuous, then problem (10.41) does not always have a solution.

To prove (ii), choose $f(x) := (x - \sqrt{2})^2$. Problem (10.40) then has the unique solution $x = \sqrt{2}$. Since this is an irrational number, $\sqrt{2}$ is not contained in the set X. Therefore, problem (10.41) has no solution. If a mathematician knew only rational numbers, but not irrational numbers, then he could not prove a general existence theorem about minimal problems of the type (10.41). But if, one day, our mathematician invented irrational numbers, thereby completing the set of rational numbers $\mathbb{Q}$ to the set of real numbers $\mathbb{R}$, then he could prove a general existence theorem for problem (10.40). For the Dirichlet principle, we will encounter the same situation in Sect. 10.4.6 on page 555 by using the following replacements:

- rational numbers ⇒ smooth functions;
- irrational numbers ⇒ distributions;
- completion of the set of rational numbers to the set of real numbers ⇒ completion of the space of smooth functions to Sobolev spaces.

It is a general strategy of mathematics to simplify the solution of important problems by introducing ideal objects.

Weyl's elegant presentation of Riemann's theory. As a highlight in complex analysis, the young Hermann Weyl (1885–1955) wrote a famous monograph on Riemann's theory:

> H. Weyl, Die Idee der Riemannschen Fläche (The concept of a Riemann surface), Teubner, Leipzig, 1913 (in German).[21]

Recall that, by definition, a Riemann surface is a connected 2-dimensional complex manifold.[22] This means that the change of local coordinates is given by diffeomorphisms χ, that is, the bijective map χ and its inverse map χ^{-1} are holomorphic functions with respect to local coordinates. We call such functions biholomorphic. Two Riemann surfaces X and Y are called conformally equivalent iff there exists a biholomorphic map

$$f : X \to Y,$$

[21] A new edition of this classic with commentaries has been published in 1997.
[22] Naturally enough, we exclude the trivial case where the manifold is empty.

that is, this map is bijective and both f and f^{-1} are holomorphic functions with respect to local coordinates. The following theorems are basic for the theory of Riemann surfaces:

(i) The Riemann mapping theorem:

Each nonempty, simply connected, open subset of the complex plane which differs from the plane is conformally equivalent to the open unit disc.

An elegant proof of this theorem can be found in Remmert (1998), Sect. 8.2.

(ii) The uniformization theorem for simply connected Riemann surfaces:

Each simply connected Riemann surface is conformally equivalent to one and only one of the following three normal forms: the open unit disc, the complex plane, the Riemann sphere.

The proof can be found in Farkas and Kra (1992), Sect. 4.4. .

(iii) The uniformization theorem for compact Riemann surfaces: The classification of compact Riemann surfaces via conformal equivalence is crucial for string theory.

The equivalence classes of compact Riemann surfaces with fixed genus $g = 0, 1, 2, \ldots$ *form Riemann's moduli space* $\mathcal{M}_g$ *of complex dimension* $0,\ 1,\ 3g-3$ *if* $g = 0,\ 1,\ g \geq 2$, *respectively.*

The moduli spaces $\mathcal{M}_g$ are closely related to Teichmüller spaces. We will study this in Volume VI. As an introduction to the theory of compact Riemann surfaces, we recommend the Lecture Notes by Jost (1997). This monograph is based on harmonic maps which are related to a generalization of the Dirichlet principle to mappings between manifolds.

10.4.5 Typical Difficulties

We want to explain why the notion of classical solution fails for the Dirichlet principle in reasonable situations.

Euler's paradox. The variational problem

$$\boxed{\begin{aligned} &\int_{t_0}^{t_1} \sqrt{x(t)}\sqrt{\dot{x}(t)^2 + \dot{y}(t)^2}\, dt = \min!, \\ &x(t_0) = x_0, y(t_0) = y_0, \qquad x(t_1) = x_1, y(t_1) = y_1 \end{aligned}} \tag{10.42}$$

refers to a curve of minimal length between the points (x_0, y_0) and (x_1, y_1). The calculus of variations was founded by Euler (1707–1783) in his famous 1744 paper. In 1779 Euler discovered that for appropriate boundary points, the solution of (10.42) is only a polygon, but not a smooth curve. Euler called this the paradox of the calculus of variations. It seems that Euler's paradox was not known to the great masters of mathematics in the first half of the 19th century. In 1871 there appeared the paper by L. Todhunter, Researches in the calculus of variations, principally of discontinuous solutions, MacMillan, London.

The microstructure of materials and random ground states. Consider the variational problem

$$E(u) := \int_0^1 (u'(x)^2 - 1)dx = \min!,\ \ u \in C^1[0,1] \tag{10.43}$$

along with the boundary condition $u(0) = u(1) = 0$.

Proposition 10.19 *This minimum problem has no solution.*

Proof. The idea is to use an equipartition of the interval $[0,1]$ of mesh size Δx and piecewise linear continuous functions $p:[0,1]\to\mathbb{R}$ which attain the derivatives $p'(x)=\pm 1$ outside the grid points. Smoothing these zig-zag functions a little bit and letting $\Delta x\to 0$, we get a sequence (u_n) of admissible functions such that $E(u_n)\to 0$ as $n\to\infty$. Hence $\inf E(u)=0$ where the infimum is taken over all admissible functions.

Suppose now that u is a solution of the minimum problem. Then $E(u)=0$. Hence $u'(x)=\pm 1$ for all $x\in[0,1]$. By continuity, either $u'(x)=1$ on $[0,1]$ or $u'(x)=-1$ on $[0,1]$. But this contradicts $u(0)=u(1)=0$. □

In terms of physics, problem (10.43) represents an oversimplified model of some material which has the energy $E(u)$ in the state u. Proposition 10.19 tells us that there is no state of minimal energy, but only a sequence of states (i.e., zig-zag functions with $\Delta x\to 0$) which approach the infimum of possible energies.

> *Therefore, it is meaningful to say that the generalized solution of problem* (10.43) *is a random distribution of states where the derivatives* $+1$ *and* -1 *are attained with probability* $\frac{1}{2}$.

This situation is well-known in the modern mathematical theory for materials equipped with microstructure. As an introduction into this new branch of analysis, we recommend the Lecture Notes by Müller (1999) (variational models for microstructure and phase transitions) and Dolzmann (2003) (variational methods for crystalline microstructure: analysis and computation).

Lack of smoothness of the ground state. The minimum problem

$$E(u):=\int_{-1}^{1} u(x)^2(2x-u'(x))^2dx=\min!,\quad u\in C^1[-1,1]$$

with the boundary condition $u(-1)=0,\ u(1)=1$ has the unique solution

$$u_0(x):=0\ \text{ on }\ [-1,0],\qquad u_0(x):=x^2\ \text{ on }\ ]0,1].$$

The second derivative of this solution jumps at the point $x=0$.

Proof. (I) Existence. Obviously, $E(u_0)=0$. Since $E(u)\geq 0$ for all admissible functions u, the function u_0 is a solution.

(II) Uniqueness. Let v be a solution. Then $E(v)=0$. Consequently, we obtain $v(x)^2(2x-v'(x))^2=0$ for all $x\in[-1,1]$. Hence

$$v'(x)=2x\quad\text{or}\quad v(x)=0\qquad\text{for all}\quad x\in[-1,1]$$

along with $v(-1)=0$ and $v(1)=1$. At the point $x=1$, the boundary condition $v(1)=1$ implies $v'(x)=2x$ near $x=1$. Let $]a,1]$ be the largest subinterval of $[0,1]$ such that $v(x)\neq 0$. Then $v'(x)=2x$ on $]a,1]$. Hence $v(x)=x^2$ on $]a,1]$. This implies $a=0$. Similarly, we start at the boundary point $x=-1$ in order to conclude that $v(x)=0$ on $[-1,0]$. Thus, $v=u_0$. □

States of infinite energy. Let Ω be the open unit disc in the plane. We are given the continuous boundary function $g_0:\partial\Omega\to\mathbb{R}$ by the Fourier series

$$g_0(\varphi):=\sum_{n=1}^{\infty}\frac{\sin n!\varphi}{n^2}.$$

We are looking for a continuous function $U:\mathrm{cl}(\Omega)\to\mathbb{R}$ which is smooth on Ω such that

$$\Delta U = 0 \text{ on } \Omega, \qquad U = g_0 \text{ on } \partial\Omega. \tag{10.44}$$

This problem has the unique solution

$$U(\varphi, r) = \sum_{n=1}^{\infty} \frac{\sin n!\varphi}{n^2}\, r^{n!}$$

where we use polar coordinates. For the energy of the electrostatic potential U, we get

$$\boxed{E(U) = \frac{1}{2}\int_{\Omega} U_x^2 + U_y^2 \, dxdy = \frac{\pi}{2}\sum_{n=1}^{\infty} \frac{n^4}{n!} = \infty.}$$

This counterexample due to Hadamard (1865–1963) tells us that the solution of the boundary-value problem (10.44) cannot be obtained by solving the variational problem of minimal energy.

In other words, the Dirichlet principle fails completely for the special boundary values g_0 on the unit circle.

Hölder spaces. Let Ω be a nonempty bounded open set in $\mathbb{R}^N$ with dimension $N = 1, 2, \ldots$. Fix the number $\alpha \in]0, 1[$. We say that the function $f : \Omega \to \mathbb{R}$ is Hölder continuous with exponent α iff the following supremum

$$H_\alpha(f) := \sup_{x,y\in\Omega, x\neq y} \frac{|f(x) - f(y)|}{||x - y||^\alpha}$$

is finite. We write $f \in C^\alpha(\Omega)$. This implies that

$$|f(x) - f(y)| \leq H_\alpha(f)||x - y||^\alpha \qquad \text{for all} \quad x, y \in \Omega. \tag{10.45}$$

Note that the given function f can be uniquely extended to a continuous function $f : \mathrm{cl}(\Omega) \to \mathbb{R}$ on the closure of Ω, and the inequality (10.45) remains valid for all $x, y \in \mathrm{cl}(\Omega)$.

- The Hölder space $C^\alpha(\Omega)$ becomes a real Banach space equipped with the norm

$$||f||_\alpha := \max_{x\in\mathrm{cl}(\Omega)} |f(x)| + H_\alpha(f).$$

- Let $k = 1, 2, \ldots$ The Hölder space $C^{k,\alpha}(\Omega)$ consists of all the functions $f \in C^\alpha(\Omega)$ whose partial derivatives up to order k lie in $C^\alpha(\Omega)$. These partial derivatives on Ω can then be continuously extended to the closure $\mathrm{cl}(\Omega)$. Equipped with the norm

$$||f||_{k,\alpha} := \sum_{|\alpha|\leq k} \max_{x\in\mathrm{cl}(\Omega)} |\partial^\alpha f(x)| + \sum_{|\alpha|=k} H_\alpha(\partial^\alpha f),$$

the space $C^{k,\alpha}(\Omega)$ becomes a real Banach space.

Lipschitz spaces. The function $F : \Omega \to \mathbb{R}$ is called Lipschitz continuous iff condition (10.45) is satisfied with $\alpha = 1$. The set of all these functions is denoted by $C^{0,1}(\Omega)$. This is a real Banach space equipped with the norm

$$||f||_{0,1} := \max_{x\in\mathrm{cl}(\Omega)} |f(x)| + H_1(f).$$

Setting $k = 1, 2, \ldots$ and $\alpha = 1$, we get the real Banach space $C^{k,1}(\Omega)$ as above. The Rademacher theorem tells us that[23]

[23] Lipschitz (1832–1903), Rademacher (1892–1969).

If $f \in C^{0,1}(\Omega)$, then the function f is differentiable almost everywhere on the set Ω.

Lack of smoothness of the Newtonian potential. Let Ω be an open ball in $\mathbb{R}^3$. The Newtonian potential

$$U(\mathbf{x}) := \int_\Omega \frac{\varrho(\mathbf{y}) d^3y}{4\pi||\mathbf{x}-\mathbf{y}||} \tag{10.46}$$

describes the electrostatic potential generated by the charge density ϱ. The corresponding electric field is given by $\mathbf{E} := -\,\mathbf{grad}\, U$. The Newtonian potential is also called volume potential.

(i) If the charge density $\varrho : \mathrm{cl}(\Omega) \to \mathbb{R}$ is continuous, then the Newtonian potential U is continuous and has continuous first partial derivatives on the total space $\mathbb{R}^3$. Hence the electric field $\mathbf{E}$ is continuous on $\mathbb{R}^3$.

(ii) If $\varrho : \mathrm{cl}(\Omega) \to \mathbb{R}$ is continuous, then it is not always true that the Newtonian potential U has partial derivatives of second order on Ω. This means that the electric field $\mathbf{E} = -\,\mathbf{grad}\, U$ is not always differentiable on the ball Ω.

(iii) However, if ϱ is Hölder continuous, that is, $\varrho \in C^\alpha(\Omega)$ with $0 < \alpha < 1$, and if ϱ vanishes in an open neighborhood of the boundary of the ball Ω, then we have $U \in C^{2,\alpha}(\Omega)$ along with $\Delta U = \varrho$ on Ω.

For the proof of (i) and (iii), see Jost (2000b), p. 256. The classical counterexample to (ii) is given by

$$\varrho(\mathbf{x}) := \left(\frac{3z^2}{r^2} - 1\right) \frac{1}{\ln r}$$

with $r := ||\mathbf{x}||$. This is studied in the monograph by Günter (1957), Sect. 2.14. The classical result (iii) was first obtained by Otto Hölder (1859–1937) in his dissertation at the University of Tübingen (Germany) in 1882. For further material, we refer to the treatise on the calculus of variations by Giaquinta and Hildebrandt (1995).

Summarizing, the solvability theory for variational problems is complicated, since there exist strange states in nature (e.g., discontinuous microstructures or concentrations of energy described by infinities of energy.)

Similarly, the mathematical difficulties arising in quantum field theory are caused by the complexity of the fundamental interactions in nature. To overcome the mathematical difficulties, it is necessary to develop the right mathematical tools. Concerning the Dirichlet principle, mathematics was successful in the 20th century in the framework of functional analysis (Hilbert's axiomatics of limiting processes). In what follows we shall discuss the basic ideas.

Historical remark. In 1782 Laplace (1749–1827) remarked in a study on the shape of the planets that the Newtonian potential (10.46) satisfies the equation $\Delta U = 0$ outside the region Ω of the planet. Interestingly enough, Laplace found this equation in the complex form of spherical coordinates. Five years later, Laplace discovered the simpler expression in Cartesian coordinates, $\Delta U = U_{xx} + U_{yy} + U_{zz}$. In 1813 Poisson (1781–1840) found that for a ball Ω with constant mass density, there holds $-\Delta U = \varrho$ on Ω which is now called the Poisson equation. Observe that in modern differential geometry, the sign of Δ has been changed in order to obtain an operator with positive eigenvalues of Δ. We follow this sign change.

10.4.6 The Functional Analytic Existence Theorem

The role of functional analysis has been decisive exactly in connection with classical problems. Almost all problems are on the applications, where

functional analysis enables one to focus on a specific set of concrete analytical tasks and organize material in a clear and transparent form so that you know what the difficulties are.

Felix Browder, 1975[24]

The generalized Dirichlet principle. Let Ω be a nonempty open bounded subset of $\mathbb{R}^N$ for fixed $N = 2, 3, \dots$ Set $x = (x^1, \dots, x^N)$ and $\partial_j := \partial/\partial x^j$. Furthermore, set $\Delta U := -\sum_{j=1}^N \partial_j^2 U$. We want to study the minimum problem

$$\boxed{\int_\Omega \Big(\sum_{j=1}^N \tfrac{1}{2}(\partial_j U)^2 - \varrho U\Big)\, d^N x = \min!, \qquad U \in W_2^1(\Omega), \quad U - U_0 \in \overset{\circ}{W}{}_2^1} \tag{10.47}$$

along with the corresponding Euler–Lagrange equation

$$\Delta U = \varrho \ \text{ on } \ \Omega, \qquad U - U_0 \in \overset{\circ}{W}{}_2^1. \tag{10.48}$$

We are given the functions $\varrho, U_0 : \Omega \to \mathbb{R}$ such that we have $\varrho \in L_2(\Omega)$ and $U_0 \in W_2^1(\Omega)$.

Theorem 10.20 *The minimum problem* (10.47) *has a unique solution U. This function satisfies the partial differential equation* (10.48) *in the sense of distributions.*

If the boundary is smooth, that is, the closure $\mathrm{cl}(\Omega)$ *is a manifold with boundary, then the solution U lies in the Sobolev space $W_2^2(\Omega)$. The solution only depends on the generalized boundary values of the given function U_0*

Roughly speaking, the solutions gain regularity if the boundary of the set Ω gains regularity (e.g., Ω is a ball).

Basic notions. Let us explain the terminology used above. We will apply the language of distributions to be introduced in Sect. 11.3 on page 607. As above, let Ω be a nonempty bounded open subset of $\mathbb{R}^N$.

(i) The Schwartz space $\mathcal{D}(\Omega)$ of test functions: Recall that this space consists of all smooth functions $\varphi : \Omega \to \mathbb{R}$ which vanish outside a compact subset of Ω. In particular, we have

$$\varphi = 0 \ \text{ on } \ \partial\Omega. \tag{10.49}$$

This boundary condition will be generalized below.

(ii) Distributions: Each function $U : \Omega \to \mathbb{R}$ with $U \in L_2(\Omega)$ represents a distribution, that is, $U \in \mathcal{D}'(\Omega)$.

Therefore, the function U has partial derivatives of all orders.

This is the decisive advantage of the theory of distributions. In particular, the symbols $\partial_j U$, $j = 1, 2, \dots$, and ΔU are well-defined. If $U, f_j \in L_2(\Omega)$, then the equation $\partial_j U = f_j$ is equivalent to

$$\int_\Omega f_j \varphi \, d^N x = -\int_\Omega U \partial_j \varphi \, d^N x \qquad \text{for all} \quad \varphi \in \mathcal{D}(\Omega). \tag{10.50}$$

[24] F. Browder, The relation of functional analysis to concrete analysis in 20th century mathematics, Historia Math. **2** (1975), 577–590.

Moreover, for given $\varrho \in L_2(\Omega)$, the equation $\Delta U = \varrho$ is equivalent to

$$\int_\Omega U \Delta\varphi \, d^N x = \int_\Omega \varrho\varphi \, d^N x \qquad \text{for all} \quad \varphi \in \mathcal{D}(\Omega). \tag{10.51}$$

In the special case where f, f_j, U, ϱ are smooth functions, formulas (10.50) and (10.51) coincide with the classical integration-by-parts formulas.

(iii) The Sobolev space $W_2^1(\Omega)$: By definition, this space consists of all functions $U : \Omega \to \mathbb{R}$ with $U, \partial_1 U, \ldots, \partial_N U \in L_2(\Omega)$. The space $W_2^1(\Omega)$ becomes a real Hilbert space equipped with the inner product

$$\langle U|V\rangle_{1,2} := \int_\Omega (UV + \sum_{j=1}^N \partial_j U \partial_j V) \, d^N x \qquad \text{for all} \quad U, V \in W_2^1(\Omega).$$

(iv) The Sobolev space $\overset{\circ}{W}{}_2^1(\Omega)$: This space is the smallest closed linear subspace of the Hilbert space $W_2^1(\Omega)$ which contains the set $\mathcal{D}(\Omega)$ of test functions.

(v) The Sobolev space $W_2^k(\Omega)$ with $k = 1, 2, \ldots$ This space consists of all functions $U : \Omega \to \mathbb{R}$ which lie in $L_2(\mathbb{R}^N)$ and whose partial derivatives up to order k lie in $L_2(\Omega)$, too. Equipped with the inner product

$$\langle U|V\rangle_{k,2} := \int_\Omega \sum_{|\alpha|\le k} \partial^\alpha U \cdot \partial^\alpha V \, d^N x,$$

the space $W_2^k(\Omega)$ becomes a real Hilbert space.[25] We also introduce the norm $||U||_{k,2} := \sqrt{\langle U|U\rangle_{k,2}}$.[26]

(vi) The fractional Sobolev space $W_2^\alpha(\Omega)$ with $0 < \alpha < 1$. This space consists of all functions $U \in L_2(\Omega)$ for which the following integral is finite:

$$||U||_{\alpha,2}^2 := \int_\Omega |U(x)|^2 d^N x + \int_{\Omega\times\Omega} \frac{|U(x) - U(y)|^2}{||x - y||^{N+2\alpha}} \, d^N x d^N y.$$

(vii) The fractional Sobolev space $W_2^{1+\alpha}(\Omega)$ with $0 < \alpha < 1$. This space consists of all functions $U \in W_2^1(\Omega)$ for which the following integral is finite:

$$||U||_{1+\alpha,2}^2 := ||U||_{1,2}^2 + \int_{\Omega\times\Omega} \sum_{j=1}^N \frac{|\partial_j U(x) - \partial_j U(y)|^2}{||x - y||^{N+2\alpha}} \, d^N x d^N y.$$

(viii) The Sobolev space $W_2^{-1}(\Omega)$ of negative order: This space is the dual space to the Hilbert space $\overset{\circ}{W}{}_2^1(\Omega)$, that is, this space consists of all linear continuous functionals

$$F : \overset{\circ}{W}{}_2^1(\Omega) \to \mathbb{R}$$

equipped with the norm $||F|| := \sup_{||U||_{1,2}\le 1} |F(U)|$.

[25] Recall that $\partial^\alpha U \partial^\alpha V = UV$ if $|\alpha| = 0$.

[26] For an introduction to the theory of Sobolev spaces, see Evans (1998), Gilbarg and Trudinger (1983), and Kufner, John, and Fučik (1977).

(ix) Generalized boundary values: Suppose that the behavior of the closure $\mathrm{cl}(\Omega)$ is sufficiently regular near the boundary. More precisely, we assume that $\mathrm{cl}(\Omega)$ is a manifold with boundary (e.g., a ball). The classical boundary operator

$$B : C(\overline{\Omega}) \to C(\partial\Omega)$$

maps the continuous function $U : \mathrm{cl}(\Omega) \to \mathbb{R}$ to its boundary values $BU : \partial\Omega \to \mathbb{R}$. This map can be uniquely extended to a linear continuous operator

$$B : W_2^1(\Omega) \to L_2(\partial\Omega). \tag{10.52}$$

We call the function BU in $L_2(\partial\Omega)$ the generalized boundary values of the function $U \in W_2^1(\Omega)$. Naturally enough, if $U \in \overset{\circ}{W}{}_2^1(\Omega)$, then $BU = 0$, that is, $U = 0$ on $\partial\Omega$ in the generalized sense. Observe that the boundary operator from (10.52) is not surjective. This means that there are boundary functions $g : \partial\Omega \to \mathbb{R}$ with $\int_{\partial\Omega} g^2 dS < \infty$ such that g is not the generalized boundary value of a function $U : \Omega \to \mathbb{R}$ which lies in $W_2^1(\Omega)$. However, the boundary operator

$$B : W_2^1(\Omega) \to W_2^{1/2}(\partial\Omega)$$

is surjective, linear, and continuous. Observe that the fractional Sobolev space $W_2^{1/2}(\partial\Omega)$ is defined similarly to $W_2^{1/2}(\Omega)$ above, by using a decomposition of the boundary $\partial\Omega$ along with local coordinates.

10.4.7 Regularity of the Solution

We want to show that the generalized solution U from Theorem 10.20 is indeed a classical smooth solution if the given data (i.e., density ϱ, boundary $\partial\Omega$, and boundary values U_0) are smooth.

The Sobolev embedding theorem. As above, let Ω be a nonempty bounded open subset of $\mathbb{R}^N$ with $N = 1, 2, \ldots$ such that the closure $\mathrm{cl}(\Omega)$ is a manifold with boundary. If k is an integer with $k > N/2$, then the embedding

$$\boxed{W_2^k(\Omega) \subseteq C(\overline{\Omega})}$$

is continuous. Explicitly this means that if $U \in W_2^k(\Omega)$, then there exists a uniquely determined continuous function $V : \mathrm{cl}(\Omega) \to \mathbb{R}$ such that

$$U(x) = V(x) \qquad \text{for almost all } x \in \Omega.$$

Moreover, if we have the convergence $U_n \to U$ in $W_2^k(\Omega)$ as $n \to \infty$, then we get the convergence $V_n \to V$ in $C(\overline{\Omega})$ for the corresponding continuous functions. In the special case where $N = 1$ (resp. $N = 2, 3$), we need $k \geq 1$ (resp. $k \geq 2$). This implies the following crucial theorem:

$$\boxed{\textit{If } U \in W_2^k(\Omega), k = 1, 2, \ldots, \textit{ then } U \in C^\infty(\overline{\Omega}),}$$

after changing the values of U on a set of N-dimensional Lebesgue measure zero, if necessary.[27] In particular, in order to get smooth solutions of the Dirichlet problem,

[27] The proof of the Sobolev embedding theorem can be found in Evans (1998), p. 270. There, one also finds stronger results. For example, if $N = 3$, then the embedding $W_2^2(\Omega) \subseteq C^{1/2}(\Omega)$ is continuous. The simple proof for a prototype of the Sobolev embedding theorem will be given in Problem 10.3 based on the fundamental theorem of calculus.

one has to show that the solution U from Theorem 10.20 lies in all the Sobolev spaces $W_2^k(\Omega)$ with $k = 1, 2, \ldots$ This is the basic idea of regularity proofs in the modern theory of partial differential equations.

Smooth solutions of the Dirichlet problem. Consider the classical boundary value problem

$$\boxed{\Delta U = \varrho \text{ on } \Omega, \qquad U = U_0 \text{ on } \partial\Omega.} \tag{10.53}$$

The corresponding electrostatic energy reads as

$$E(U) = \int_\Omega (\sum_{j=1}^N \tfrac{1}{2}(\partial_j U)^2 - \varrho U)\, d^N x.$$

Let us make the following assumptions.

(A1) Smooth boundary: Let Ω be a nonempty bounded open subset of $\mathbb{R}^N$ with $N = 2, 3, \ldots$ such that the closure $\mathrm{cl}(\Omega)$ is a manifold with boundary (e.g., a ball).

(A2) Smooth charge density function: We are given the function $\varrho \in C^\infty(\overline{\Omega})$.

(A3) Smooth boundary values of the electrostatic potential: The given boundary function $U_0 : \partial\Omega \to \mathbb{R}$ is smooth (with respect to local coordinates of the boundary manifold $\partial\Omega$).

Theorem 10.21 *The boundary-value problem* (10.53) *has a unique real solution* $U \in C^\infty(\overline{\Omega})$. *The function* U *is also the unique solution of the variational problem*

$$E(U) = \min!, \qquad U \in C^\infty(\overline{\Omega}), \qquad U = U_0 \text{ on } \partial\Omega.$$

More generally, the classical solution U from Theorem 10.21 coincides with the generalized solution from Theorem 10.20. The proof can be found in Jost (2000b), p. 226.

The extension of the classical Laplacian for zero boundary values. It is convenient to reformulate the concept of generalized solution in the language of operator theory in Hilbert spaces. Let us first consider the homogeneous boundary value problem

$$\Delta U = \varrho \text{ on } \Omega, \qquad U = 0 \text{ on } \partial\Omega. \tag{10.54}$$

We start with the classical Laplacian

$$\Delta : \mathcal{D}(\Omega) \to \mathcal{D}(\Omega) \tag{10.55}$$

given by $\Delta U = \varrho$. Our goal is to construct function spaces X and Y such that there exists an extension

$$\boxed{\Delta : X \to Y}$$

of the Laplacian from (10.55) where the extended operator is a linear homeomorphism from X onto Y. Then there exists the linear continuous operator

$$\Delta^{-1} : Y \to X.$$

This means that equation (10.54) has the unique solution $U = \Delta^{-1}\varrho$, and this solution depends continuously on the charge density ϱ. The proof of Theorem 10.20 shows that one can choose the Sobolev spaces

$$X = \overset{\circ}{W}{}_2^1(\Omega) \quad \text{and} \quad Y = W_2^{-1}(\Omega).$$

The extension of the classical Laplacian for nonzero boundary values. We now consider the more general inhomogeneous boundary-value problem

$$\Delta U = \varrho \ \text{ on } \ \Omega, \qquad U = U_0 \ \text{ on } \ \partial\Omega. \tag{10.56}$$

We assume that Ω is a nonempty bounded open set in $\mathbb{R}^N$ such that $\mathrm{cl}(\Omega)$ is a manifold with boundary. Let us start with the classical operator

$$\Delta : C^\infty(\overline{\Omega}) \to C^\infty(\overline{\Omega}) \times C^\infty(\partial\Omega)$$

given by $\Delta U := (\varrho, U_0)$. This operator can be uniquely extended to a linear homeomorphism

$$\boxed{\Delta : X \to Y \times Z}$$

if we choose the following spaces.

(i) Hölder spaces: $X = C^{2,\alpha}(\Omega),\ Y = C^\alpha(\Omega),\ Z = C^{2,\alpha}(\partial\Omega)$.[28]

(ii) Sobolev spaces: $X = W_2^2(\Omega),\ Y = L_2(\Omega),\ Z = W_2^{3/2}(\partial\Omega)$.

The extension of linear differential operators in Hilbert spaces was first studied by Friedrichs (1901–1982) in the 1930s, by applying von Neumann's spectral theory for unbounded self-adjoint operators. The theory of the Friedrichs extension and important applications to the partial differential equations of mathematical physics can be found in Zeidler (1995), Vol. 1, Chap. 5.

Interpolation theory. In order to get operators of the form X, Y, Z above in a systematic way, interpolation theory was developed in the 1960s. The most important function spaces are the Besov spaces $B^s_{pq}(\Omega)$ and the Triebel–Lizorkin spaces $F^s_{pq}(\Omega)$. As special cases, we mention

$$B^{k+\alpha}_{\infty,\infty}(\Omega) = C^{k,\alpha}(\Omega), \quad F^k_{22}(\Omega) = W_2^k(\Omega), \quad k = 0, 1, \ldots,\ 0 < \alpha < 1$$

and $F^{1/2}_{22}(\Omega) = W_2^{1/2}(\Omega)$. Here, $C^{k,\alpha}(\Omega) := C^\alpha(\Omega)$ and $W_2^k(\Omega) := L_2(\Omega)$ if $k = 0$. As an elementary introduction to this subject, we recommend Zeidler (1995), Vol. 2, p. 360ff and Reed and Simon (1972), Vol. 2, Sect. IX.4 (interpolation theory in Banach spaces). Much material can be found in Triebel, Theory of Function Spaces, Birkhäuser, Basel, 1992.

10.4.8 The Beauty of the Green's Function

> The Green's function allows us to reduce the solution of general boundary values to the solution of a special boundary value problem.
>
> Folklore

The fundamental solution of the Poisson equation. Let us fix the point $x_0 \in \mathbb{R}^N$. For all $x \in \mathbb{R}^N$ with $x \neq x_0$, define

$$U_{x_0}(x) := \begin{cases} \frac{Q}{(N-2)\ \mathrm{meas}(\mathbb{S}^{N-1})} \cdot ||x - x_0||^{-N+2} & \text{if} \quad N = 3, 4, \ldots, \\ -\frac{Q}{2\pi}\ \ln ||x - x_0|| & \text{if} \quad N = 2. \end{cases}$$

[28] This approach was developed by Schauder (1899–1943) in the 1930s. A modern version of the Schauder theory based on sophisticated a priori estimates can be found in Jost (2002b).

Here, $\mathrm{meas}(\mathbb{S}^{N-1})$ is the surface measure of the unit sphere in $\mathbb{R}^N$. Explicitly,

$$\mathrm{meas}(\mathbb{S}^{2M-1}) = \frac{2\pi^M}{(M-1)!}, \qquad \mathrm{meas}(\mathbb{S}^{2M}) = \frac{2^{2M+1}M!\,\pi^M}{(2M)!}, \quad M = 1, 2, \ldots$$

In terms of physics, for $N = 3$, the function

$$U_{x_0}(x) = \frac{Q}{4\pi\,||x - x_0||}$$

is the electrostatic potential of the charge Q located at the point x_0, along with the electric field $\mathbf{E} = -\,\mathbf{grad}\,U_{x_0}$.

Proposition 10.22 *For all test functions $\varphi \in \mathcal{D}(\mathbb{R}^N)$ with $N = 2, 3, \ldots$,*

$$\int_\Omega U_{x_0}(x)\Delta\varphi(x)\,d^N x = Q\varphi(x_0).$$

The proof will be given in Problem 10.4. Let us translate this into the modern language of distributions to be introduced in Sect. 12.6.1 on page 730. To this end, for all $\varphi \in \mathcal{D}(\mathbb{R}^N)$, we define the following notions.

- Dirac's delta distribution: $\delta_{x_0}(\varphi) := \varphi(x_0)$;
- electrostatic potential as generalized function (distribution):

$$\mathcal{U}_{x_0}(\varphi) := \int_{\mathbb{R}^N} U_{x_0}(x)\varphi(x)\,d^N x;$$

- derivative $\partial_j \mathcal{U}_{x_0}(\varphi) = -\mathcal{U}_{x_0}(\partial_j\varphi),\ j = 1, \ldots, N$;
- Laplacian: $\Delta\mathcal{U}_{x_0}(\varphi) = \mathcal{U}_{x_0}(\Delta\varphi)$.

Therefore, it follows from Prop. 10.22 that

$$\boxed{\Delta\mathcal{U}_{x_0} = Q\delta_{x_0}.} \tag{10.57}$$

The existence-and uniqueness theorem for the Green's function. Let Ω be a nonempty bounded open subset of $\mathbb{R}^N$ with $N = 2, 3, \ldots$ such that the closure $\mathrm{cl}(\Omega)$ is a manifold with boundary. Fix the point $x_0 \in \Omega$, and consider the following boundary-value problem:

$$\Delta W_{x_0} = 0 \ \text{ on } \ \Omega, \qquad W_{x_0} = -U_{x_0} \ \text{ on } \ \partial\Omega. \tag{10.58}$$

The following theorem is a special case of Theorem 10.21.

Theorem 10.23 *The boundary-value problem* (10.58) *has precisely one classical solution $W_{x_0} \in C^\infty(\overline{\Omega})$.*

Using the function W_{x_0}, we define the function

$$G(x, x_0) := U_{x_0}(x) + W_{x_0}(x) \qquad \text{for all} \quad x \in \Omega \setminus \{x_0\}$$

which is called the Green's function of the Poisson equation on the set Ω. We set $G_{x_0}(x) := G(x, x_0)$. By (10.57),

$$\boxed{\Delta G_{x_0} = Q\delta_{x_0} \ \text{ on } \ \Omega, \qquad G_{x_0} = 0 \ \text{ on } \ \partial\Omega.} \tag{10.59}$$

It turns out that G is symmetric, that is, $G(x, x_0) = G(x_0, x)$ for all $x, x_0 \in \Omega$ with $x \neq x_0$.

The magic Green's solution formula. In his 1828 paper, Green underlined the importance of the following formula.

Theorem 10.24 *The solution U from Theorem 10.21 can be represented by the magic Green formula*

$$U(x_0) = \int_\Omega G(x, x_0)\varrho(x)\, d^N x - \int_{\partial\Omega} \frac{\partial G(x, x_0)}{\partial n_x} U_0(x)\, dS$$

where $\frac{\partial}{\partial n_x}$ denotes the directional derivative with respect to the outer normal vector at the boundary point x.

This is the prototype of representation formulas for the solutions of partial differential equations by using integral kernels. This formula shows how the charge density and the boundary values influence the electrostatic potential U. For example, if the charge density ϱ is only concentrated near the point x_1 and the boundary values of the potential are only concentrated near the point x_2, then the solution looks approximately like

$$U(x_0) = G(x_1, x_0)\varrho(x_1)(\Delta x)^3 - \frac{\partial G(x_2, x_0)}{\partial n_{x_2}} U_0(x_2)\Delta S.$$

Therefore, the Green's function localizes physical effects. This fact is crucial for quantum field theory.

Proof. Let B_R be an open ball of sufficiently small radius R centered at the point x_0. By Green's formula (10.33) on page 546,

$$\int_{\Omega\setminus B_R} U\Delta G_{x_0} - G_{x_0}\Delta U\, d^N x = \int_{\partial\Omega} G_{x_0}\frac{\partial U}{\partial n} - U\frac{\partial G_{x_0}}{\partial n}\, dS.$$

This formula can be simplified by observing the following.

- $\Delta G_{x_0} = 0$ on $\Omega \setminus B_R$.
- $G_{x_0} = 0$ on $\partial\Omega$.
- $\Delta U = \varrho$ on Ω.

Finally, let $R \to 0$ and observe the Green lemma proved in Problem 10.5. □

The Green's function for the unit ball. Let $\mathcal{B}$ be the open unit ball in $\mathbb{R}^3$. In this case, the Green's function for the Poisson equation is explicitly known, namely, if $x_0 \in \mathcal{B} \setminus \{0\}$, then

$$G(x, x_0) := \frac{Q}{4\pi||x - x_0||} - \frac{Q^*}{4\pi||x - x_0^*||} \qquad \text{for all} \quad x \in \mathcal{B} \setminus \{x_0\}.$$

Here, the point x_0^* is obtained from x_0 by inversion with respect to the unit sphere. Explicitly,

$$x_0^* := \frac{1}{||x_0||} \cdot \frac{x_0}{||x_0||} \qquad \text{for all} \quad x_0 \in \mathcal{B} \setminus \{0\}.$$

Moreover, we choose the charges $Q := 1$ and $Q^* := Q||x_0^*||/||x_0||$. The conformal map $x_0 \mapsto x_0^*$ is called Kelvin transformation.. If $x_0 = 0$, then we set

$$G(x, 0) := \frac{1}{4\pi||x||} - \frac{1}{4\pi} \qquad \text{for all} \quad x \in \mathcal{B} \setminus \{0\}.$$

In terms of physics, the Green's function allows the following interpretation. Note that in a metallic electric conductor, the electrostatic potential vanishes. Now consider a metallic unit sphere which bounds the open unit ball $\mathcal{B}$. Put an electric charge at the point x_0. The Green's function $x \mapsto G(x, x_0)$ is then the corresponding electrostatic potential. In fact, as the superposition of two potentials, the

function $x \mapsto G(x, x_0)$ satisfies the Laplace equation outside both the point x_0 and the mirror point x_0^*. Furthermore, we have chosen the point x_0^* in such a way that $G(x, x_0) = 0$ on the unit sphere $\partial\mathcal{B}$. In fact, if $||x|| = 1$, then

$$\frac{||x - x_0^*||^2}{||x - x_0||^2} = \frac{1 - xx_0^* + ||x_0^*||^2}{1 - xx_0 + ||x_0||^2}.$$

Set $r := ||x_0||$ and $r_* := ||x_0^*||$. Noting that $rr_* = 1$,

$$\frac{r^2||x - x_0^*||^2}{||x - x_0||^2} = \frac{r^2(1 - 2r_* \cos\vartheta + r_*^2)}{1 - 2r\cos\vartheta + r^2} = 1.$$

From Theorem 10.24 we obtain the so-called Poisson formula

$$\boxed{U(x_0) := \frac{1 - ||x_0||^2}{4\pi} \int_{\partial\mathcal{B}} \frac{U_0(x)dS}{||x - x_0||^3} \qquad \text{for all} \quad x_0 \in \mathcal{B}.} \tag{10.60}$$

By Theorem 10.24, this is the unique solution of the boundary-value problem

$$\Delta U = 0 \text{ on } \mathcal{B}, \qquad U = U_0 \text{ on } \partial\mathcal{B} \tag{10.61}$$

if the given boundary function U_0 is smooth. The following classical proposition shows that the continuity of the boundary function is sufficient.

Proposition 10.25 *If the given boundary function $U_0 : \partial\mathcal{B} \to \mathbb{R}$ is continuous, then the Poisson formula* (10.60) *yields the unique solution of the Dirichlet problem* (10.61) *which is smooth on the open ball $\mathcal{B}$ and can be continuously extended to the closed unit ball* $\mathrm{cl}(\mathcal{B})$.

For the proof, see Problem 10.6.

The elegant formal language of physicists. In the late 1920s, Dirac reformulated the classical 1828 approach due to Green by introducing the Dirac delta function δ. Formally, we have $\delta(x - x_0) = 0$ if $x \in \mathbb{R}^N \setminus \{x_0\}$ and

$$\int_{\mathbb{R}^N} \delta(x - x_0) f(x)\, d^N x = f(x_0).$$

By (10.22), for all test functions $\varphi \in \mathcal{D}(\Omega)$,

$$\int_\Omega U_{x_0}(x) \Delta\varphi(x)\, d^N x = \int_\Omega Q\delta(x - x_0)(x)\varphi(x)\, d^N x.$$

Formal integration by parts tells us that

$$\int_\Omega (\Delta U_{x_0}(x) - Q\delta(x - x_0))\varphi(x)\, d^N x = 0 \qquad \text{for all} \quad \varphi \in \mathcal{D}(\Omega).$$

Formal application of the variational lemma yields $\Delta U_{x_0}(x) = Q\delta(x - x_0)$ on the set Ω. Thus, for each fixed $x_0 \in \Omega$,

$$\boxed{\Delta G(x, x_0) = Q\delta(x - x_0) \text{ on } \Omega, \qquad G(x, x_0) = 0 \text{ on } \partial\Omega.} \tag{10.62}$$

This equation is used by physicists in order to characterize the Green's function of the Poisson equation on Ω. Equations (10.62) and (10.59) reflect the difference between the language of physicists and mathematicians, respectively.

10.4.9 The Method of Orthogonal Projection

> I am convinced that it will be possible to get these existence proofs by a general basic idea towards which the Dirichlet principle points. Perhaps it will then also be possible to answer the question of whether or not every regular variational problems possesses a solution if, with regard to boundary conditions, certain assumptions are fulfilled and if, when necessary, one sensibly generalizes the concept of solution.
>
> David Hilbert
> *Paris lecture* 1900[29]

The method of orthogonal projection. We consider the following minimum problem

$$\boxed{||U_0 - U||^2 = \min!, \qquad U \in L.} \tag{10.63}$$

We are given the element U_0 of the real Hilbert space X and the closed linear subspace L of X. Explicitly, $||U_0 - U||^2 = \langle U_0 - U | U_0 - U\rangle$. Therefore, problem (10.63) is called a quadratic variational problem. Intuitively, if X is the 3-dimensional Euclidean space, then we are looking for the minimal distance between the given point U_0 and the plane L. We expect that there exists a unique solution U which is the orthogonal projection of the position vector U_0 onto the plane L, and the Pythagorean theorem from elementary geometry motivates the relation

$$||U_0||^2 = ||U||^2 + ||U - U_0||^2. \tag{10.64}$$

Intuitively, the following theorem tells us that in a real Hilbert space, there exists a perpendicular from a given point to the closed linear subspace L.

Theorem 10.26 *Problem* (10.63) *has precisely one solution* U.
The vector $U_0 - U$ *is orthogonal to the plane* L*, and there holds the Pythagorean relation* (10.64).

Proof. (I) The parallelogram identity. For all $U, V \in X$, we have

$$2||U||^2 + 2||V||^2 = ||U - V||^2 + ||U + V||^2$$

which generalizes the Pythagorean theorem. To prove this, note that

$$\langle U \pm V | U \pm V\rangle = \langle U|U\rangle \pm \langle U|V\rangle \pm \langle V|U\rangle + \langle V|V\rangle$$

implies $||U \pm V||^2 = ||U||^2 \pm 2\langle U|V\rangle + ||V||^2$.

(II) Existence. Set $\kappa := \inf_{U\in L} ||U_0 - U||^2$. Obviously, $\kappa \geq 0$. There exists a sequence (V_n) in the linear subspace L such that

$$\lim_{n\to\infty} ||U_0 - V_n||^2 = \kappa. \tag{10.65}$$

Applying the parallelogram identity,

$$2||U_0 - V_n||^2 + 2||U_0 - V_m||^2 = ||V_n - V_m||^2 + 4||U_0 - \tfrac{1}{2}(V_n + V_m)||^2.$$

Since $V_n, V_m \in L$, we get $\frac{1}{2}(V_n + V_m) \in L$. Hence

[29] D. Hilbert, Mathematical problems, Bull. Amer. Math. Soc. **8** (1902), 437–479.

$$2||U_0 - V_n||^2 + 2||U_0 - V_m||^2 \geq ||V_n - V_m||^2 + 4\kappa.$$

Letting $n \to \infty$ and $m \to \infty$, the left-hand side goes to 4κ. Consequently, (V_n) is a Cauchy sequence, and hence it is convergent, say, $V_n \to U$ as $n \to \infty$. By (10.65), $||U_0 - U||^2 = \kappa$. Thus, U is a solution of (10.63).

(III) The abstract Euler–Lagrange equation. If U is a solution of (10.63), then

$$\langle U_0 - U|V\rangle = 0 \qquad \text{for all} \quad V \in L. \tag{10.66}$$

To show this, fix the point $V \in L$ and set

$$\mathcal{J}(t) := ||U_0 - (U + tV)||^2 \qquad \text{for all} \quad t \in \mathbb{R}.$$

Then the function $\mathcal{J}$ has a minimum at $t = 0$. Hence $\mathcal{J}'(0) = 0$. Since $\mathcal{J}(t)$ is equal to

$$\langle U_0 - (U + tV)|U_0 - (U + tV)\rangle = \langle U_0|U_0\rangle + 2t\langle U_0 - U|V\rangle + t^2\langle V|V\rangle,$$

we get $\mathcal{J}'(0) = 2\langle U_0 - U|V\rangle = 0$.

(IV) The Pythagorean relation. It follows from $\langle U - U_0|U\rangle = 0$ that

$$\langle U_0|U_0\rangle = \langle U + (U_0 - U)|U + (U_0 - U)\rangle = \langle U|U\rangle + \langle U - U_0|U - U_0\rangle.$$

(V) Uniqueness. If U and U_1 are solutions of (10.63), then U and U_1 satisfy the Euler–Lagrange equation (10.66). Hence $\langle U - U_1|V\rangle = 0$ for all $V \in L$. Letting $V := U - U_1 = 0$, we get $U - U_1 = 0$. □

The main theorem on quadratic minimum problems. For the minimum problem

$$\boxed{\tfrac{1}{2}a(U,U) - b(U) = \min!, \qquad U \in L,} \tag{10.67}$$

let us make the following assumptions.

(A1) The map $b : L \to \mathbb{R}$ is linear on the real Hilbert space L. There exists a constant $\beta > 0$ such that $|b(U)| \leq \beta||U||$ for all $U \in L$.

(A2) The map $a : L \times L \to \mathbb{R}$ is linear in each argument and symmetric, that is, $a(U,V) = a(V,U)$ for all $U, V \in L$. There exist constants $\alpha > 0$ and $\gamma > 0$ such that

$$\gamma||U||^2 \leq a(U,U) \leq \alpha||U||^2 \qquad \text{for all} \quad U \in L.$$

Theorem 10.27 *Problem* (10.67) *has a unique solution.*

The proof proceeds analogously to the proof of Theorem 10.26 on orthogonal projection. This proof can be found in Problem 10.1. In the special case where $L = \mathbb{R}$, our problem reads as $\frac{1}{2}U^2 - bU = \min!$, $U \in \mathbb{R}$ with the given real number b. The unique solution reads as $U = b$. This follows from $\frac{1}{2}U^2 - bU = \frac{1}{2}(U - b)^2 - b^2$.

Application to the Dirichlet principle. Let Ω be a nonempty bounded open set in $\mathbb{R}^2$. We consider the variational problem

$$\boxed{\int_\Omega \tfrac{1}{2}(U_x^2 + U_y^2) - \varrho U \, dxdy = \min!, \qquad U \in \overset{\circ}{W}{}_2^1(\Omega).} \tag{10.68}$$

This corresponds to the homogenous boundary condition $U = 0$ on $\partial\Omega$.

Theorem 10.28 *For given function $\varrho \in L_2(\Omega)$, the problem* (10.68) *of minimal energy has a unique solution.*

Proof. We will use Theorem 10.27. To this end, we set $L := \overset{\circ}{W}{}_2^1(\Omega)$ with the norm $||U||^2 := \int_\Omega U^2 + U_x^2 + U_y^2 \, dxdy$, as well as

$$a(U,V) := \int_\Omega U_x V_x + U_y V_y \, dxdy, \qquad b(U) := \int_\Omega \varrho U \, dxdy.$$

By the Schwarz inequality on the Hilbert space $L_2(\Omega)$, for all $U \in L$,

$$|b(U)|^2 \le \int_\Omega \varrho^2 \, dxdy \int_\Omega U^2 \, dxdy \le \left(\int_\Omega \varrho^2 \, dxdy \right) ||U||^2.$$

Analogously, $|a(U,V)|^2 \le ||U||^2 \cdot ||V||^2$ for all $U \in L$. Finally, it follows from the Poincaré–Friedrichs inequality (see Problem 10.2) that

$$||U||^2 \le (1+C) a(U,U) \qquad \text{for all} \quad U \in L.$$

□

Suggested reading. For first reading on the modern theory of partial differential equations, we recommend:

C. Evans, Partial Differential Equations, Amer. Math. Soc., Providence, Rhode Island, 1998.

G. Evans, J. Blackledge, and P. Yardley, Analytic Methods for Partial Differential Equations, Springer, London, 2000.

Yu. Egorov, A. Komech, and M. Shubin, Elements of the Modern Theory of Partial Differential Equations, Springer, New York, 1999.

The Hilbert space approach to the Dirichlet problem can be extended to the linear partial differential equations of elliptic, parabolic, and hyperbolic type. In the 1960s and 1970s, this approach was generalized to nonlinear partial differential equations in the setting of the theory of monotone operators. This can be found in Zeidler (1986), Vol. IIA (linear theory), Vol. IIB (nonlinear theory), Vol. III (variational problems and optimization), Vol. IV (applications to nonlinear problems in mathematical physics).

For the regularity theory of generalized solutions to linear elliptic partial differential equations and variational problems, see the following textbooks:

D. Gilbarg and N. Trudinger, Elliptic Partial Differential Equations of Second Order, Springer, New York, 1983.

J. Jost and Xianqing Li-Jost, Calculus of Variations, Cambridge University Press, 1998.

J. Jost, Partial Differential Equations, Springer, New York, 2002.

J. Jost, Postmodern Analysis, Springer, Berlin, 2005.

For a comprehensive presentation of the modern theory of partial differential equations, we refer to the following two treatises:

R. Dautray and J. Lions, Mathematical Analysis and Numerical Methods for Science and Technology, Vols. 1–6, Springer, New York, 1988.

M. Taylor, Partial Differential Equations, Vols. 1–3, Springer, New York.

As impressive monographs on the scope of Hilbert space theory, we recommend:

K. Maurin, Methods of Hilbert Spaces, Polish Scientific Publisher, Warsaw, 1972.

K. Maurin, Generalized Eigenfunction Expansions and Unitary Representations of Topological Groups, Polish Scientific Publisher, Warsaw, 1968.

K. Maurin, Riemann's Legacy: Riemann's Ideas in Mathematics and Physics of the 20th Century, Kluwer, Dordrecht, 1997.

10.4.10 The Power of Ideas in Mathematics

> A mathematician like a painter or poet, is a maker of patterns. If his patterns are more permanent as theirs, it is because they are made with ideas.
>
> Godfrey Harold Hardy (1877–1947)

In a right triangle the side opposite to the right angle is called the hypotenuse. The theorem of Pythagoras says that

$$c^2 = a^2 + b^2.$$

In words: the square of the length of the hypotenuse is equal to the sum of the squares of the lengths of the other two legs. Mathematicians of the Pythagorean school in ancient Greece attributed the Pythagorean theorem to the master of their school, Pythagoras of Samos (circa 560 B.C.–480 B.C.). It is said that Pythagoras sacrificed one hundred oxen to the gods in gratitude. In fact, this theorem was already known in Babylon at the time of King Hammurabi (circa 1728 B.C.–1686 B.C.). Presumably, however, it was a mathematician of the Pythagorean school who first proved the Pythagorean theorem. This famous theorem appears as Proposition 47 in Book I of Euclid's *Elements* (300 B.C.).

In 1940 Hermann Weyl wrote a fundamental paper where he emphasized that the justification of the Dirichlet principle can be based on the method of orthogonal projection. In 1943 he applied this method to the Dirichlet prinicple for harmonic differential forms. This way, Hodge theory obtained a sound analytic foundation.[30] The notion of Hilbert space is the abstract realization of the idea of orthogonality. To this end, one introduces the inner product

$$\langle \varphi | \psi \rangle$$

between the two elements φ and ψ of the Hilbert space. Recall that we say that φ is orthogonal to ψ iff $\langle \varphi | \psi \rangle = 0$. The proofs of the essential theorems about Hilbert spaces (e.g., spectral theory) are based on the notion of orthogonality. In the late 1920s, John von Neumann (1903–1955) discovered that the mathematical foundation of quantum mechanics can be based on the theory of Hilbert spaces. Therefore, the concept of orthogonality is crucial for quantum mechanics. In the early 1940s, Feynman (1918–1988) developed a new approach to quantum mechanics. He called the inner product $\langle \varphi | \psi \rangle$ the transition amplitude between the two quantum states φ and ψ, and he used this in order to construct his path integral.

> *There are ideas in mathematics, like the idea of orthogonality, which remain eternally young and which lose nothing of their intellectual freshness after thousands of years.*

[30] H. Weyl, The method of orthogonal projection in potential theory, Duke Math. J. **7**, (1940), 411–440.
H. Weyl, On Hodge's theory of harmonic integrals, Ann. of Math. **44** (1943), 1–6.

10.4.11 The Ritz Method

In order to solve the Dirichlet problem on computers, one uses the method of finite elements in modern scientific computing. This is a special case of the Ritz method.[31]

The basic idea. To the original minimum problem

$$||U_0 - U||^2 = \min!, \qquad U \in L, \tag{10.69}$$

we add the family of finite-dimensional problems

$$\boxed{||U_0 - V_n||^2 = \min!, \qquad V_n \in L_n, \quad n = 1, 2, \dots} \tag{10.70}$$

and we make the following assumptions.

(A1) Let L be a closed linear infinite-dimensional subspace of the real infinite-dimensional Hilbert space X. We are given $U_0 \in X$.

(A2) We choose a sequence $L_1, L_2, \dots$ of finite-dimensional linear subspaces of L such that the union of these subspaces is dense in L.

Theorem 10.29 *Each finite-dimensional problem from* (10.70) *has a unique solution V_n, and the sequence (V_n) converges in the Hilbert space L to the unique solution U of* (10.69) *as $n \to \infty$.*

The method of finite elements corresponds to the special case where L_n consists of piecewise linear (or piecewise polynomial) functions. Theorem 10.29 is the special case of a general theorem on quadratic minimum problems in Hilbert spaces. In order to get two-sided estimates for the minimal value of the original problem one uses dual maximum problems. This can be found in Zeidler (1995), Vol. 1, Sect. 2.6ff. In order to obtain optimal speed of convergence, one uses different scales. This so-called multi-grid method was developed in the 1980s. We recommend the following monographs:

W. Hackbusch, Multigrid Methods and Applications, Springer, Berlin, 1985.

W. Hackbusch, Elliptic Differential Equations: Theory and Numerical Treatment, Springer, New York, 1992.

P. Knabner and L. Angermann, Numerical Methods for Elliptic and Parabolic Partial Differential Equations, Springer, New York, 2003.

Application to the ground state energy of the Helium atom. In 1928 Hylleraas used the Ritz method in order to approximately compute the ground state energy of the Helium atom which has two electrons.[32] We refer to the following monographs:

H. Bethe and E. Salpeter, Quantum Mechanics of One- and Two-Electron Atoms, Springer, Berlin, 1957.

M. Reed and B. Simon, Methods of Modern Physics, Vol. 4, Chap. XIII, Academic Press, New York, 1978.

[31] W. Ritz, On a new method for solving a class of variational problems (in German), J. Reine Angew. Math. **135** (1909), 1–61.

[32] E. Hylleraas, On the ground state of the Helium atom (in German), Z. Phys. **48** (1928), 469–494; **54** (1929), 347–366; **65** (1930), 209–225.
E. Hylleraas, Reminiscences from early quantum mechanics of two-electron atoms, Rev. Mod. Phys. **35** (1963), 421–436.

In quantum chemistry, the huge software system "Gaussian" is used in order to compute the energy levels of fairly large molecules by means of the Ritz method. In this connection, the method of the density functional plays an important role. This method was developed by Walter Kohn in 1965 who was awarded the Nobel prize in chemistry in 1998. As an introduction to this topic, we refer to

> H. Eschrig, The Fundamentals of Density Functional Theory, Teubner, Leipzig, 2003.

A complete mathematical justification of the density functional theory is still missing.

10.4.12 The Main Existence Principle

> Tonelli (1885–1946) very successfully introduced lower-semicontinuity arguments into existence proofs by direct methods. He collected and presented his ideas, methods, and results in his treatise *Fondamenti di Calcolo delle Variazioni* the two volumes of which appeared in 1921 and 1923, respectively.
>
> Mariano Giaquinta and Stefan Hildebrandt[33]

We want to generalize the method of orthogonal projection for quadratic variational problems to more general minimum problems. This way, we will obtain a universal abstract existence principle in the calculus of variations.

The classical existence theorem. Let us make the following assumptions.

(A1) The real Hilbert space X is finite-dimensional.

(A2) Semicontinuity: The function $F : X \to \mathbb{R}$ is lower semicontinuous. By definition, this means that if $U_n \to U$ in X as $n \to \infty$, then[34]

$$F(U) \le \liminf_{n\to\infty} F(U_n).$$

(A3) Coercivity: $\lim_{||U||\to\infty} ||F(U)|| = \infty$ (growth condition at infinity).

Theorem 10.30 *Each lower semicontinuous, coercive functional on a finite-dimensional real Hilbert space has a minimum.*

For example, the real function $F : \mathbb{R} \to \mathbb{R}$ given by

$$F(U)) := |U| \qquad \text{for all} \quad U \in \mathbb{R} \setminus \{0\}, \qquad F(0) \le 0$$

is lower semicontinuous and coercive. This function attains the minimum at the point $U = 0$.

Proof. By (A3), there exists a number $R > 0$ such that

$$||F(U)|| \ge ||F(0)|| \qquad \text{for all} \quad U \in X \text{ with } ||U|| \ge R.$$

Thus, in order to compute the minimum of F on X, we can restrict ourselves to the minimum problem

$$F(U) = \min!, \qquad U \in X, \quad ||U|| \le R.$$

[33] Calculus of Variations, Vols. 1, 2, Springer, Berlin, 1995.

[34] Explicitly, if $F(U_{n'}) \to \kappa$ as $n' \to \infty$ for any subsequence of (U_n), then $F(U) \le \kappa$.

Set $\kappa := \inf_{||U|| \le R} F(U)$. Then there exists a sequence (U_n) with $||U_n|| \le R$ for all n such that

$$\kappa = \lim_{n \to \infty} F(U_n).$$

Since the sequence (U_n) is bounded, there exists a convergent subsequence, $U_{n'} \to U$ as $n' \to \infty$. Hence $||U|| \le R$. By (A2),

$$F(U) \le \lim_{n' \to \infty} F(U_{n'}) = \kappa.$$

This implies $F(U) = \kappa$. □

Lack of compactness in infinite-dimensional Hilbert spaces. We now want to generalize Theorem 10.30 to infinite dimensions. There arises the following difficulty.[35]

Proposition 10.31 *For a Hilbert space X , the following three properties are equivalent.*

(i) *The dimension of X is finite.*
(ii) *Each bounded sequence in X has a convergent subsequence.*
(iii) *The closed unit ball of X is compact.*

Consequently, the proof of Theorem 10.30 fails in infinite-dimensional Hilbert spaces. However, it is possible to modify this proof by passing from convergence to weak convergence.

The idea of weak convergence. By definition, a sequence (U_n) in the Hilbert space X is weakly convergent, denoted by

$$U_n \rightharpoonup U \quad \text{as} \quad n \to \infty,$$

iff $\langle V|U_n\rangle \to \langle V|U\rangle$ as $n \to \infty$ for all $V \in X$.

Proposition 10.32 *Each bounded sequence in a Hilbert space has a weakly convergent subsequence.*

If a closed ball contains a weakly convergent sequence, then it also contains the weak limit.

On a finite-dimensional Hilbert space, convergence and weak convergence coincide.

The main theorem. Let us modify the assumptions (A1)–(A3) of Theorem 10.30 above in the following way.

(B1) The dimension of the real Hilbert space X is finite or infinite.
(B2) Weak semicontinuity: The function $F : X \to \mathbb{R}$ is weakly lower semicontinuous, that is, it follows from $U_n \rightharpoonup U$ in X as $n \to \infty$ that

$$F(U) \le \liminf_{n \to \infty} F(U_n).$$

(B3) Coercivity: $\lim_{||U|| \to \infty} ||F(U)|| = \infty$.

Theorem 10.33 *Each weakly lower semicontinuous, coercive functional on a real Hilbert space has a minimum.*

[35] For the proofs of Propositions 10.31, 10.32, see Zeidler (1995), Vol. 2, Sect. 2.4.

Corollary 10.34 *Each continuous convex coercive functional on a real Hilbert space has a minimum.*

For the proof of Theorem 10.33, replace convergence by weak convergence and argue as in the proof of Theorem 10.30. For the proof of Corollary 10.34, we refer to Zeidler (1995), Vol. 2, Sect. 2.5. Generalizations of this theorem and many applications to variational problems and optimization problems can be found in the author's monograph Zeidler (1986), Vol. III. In particular, the violation of convexity is responsible for the appearance of microstructures. This is studied in Müller (1999) and Dolzmann (2003).

Problems

10.1 *The main theorem on quadratic variational problems.* Prove Theorem 10.27 analogously to Theorem 10.26 on orthogonal projection.
Solution: (I) Existence. Set $F(U) := \frac{1}{2}a(U,U) - b(U)$. Let $\kappa := \inf_{u\in L} F(U)$. By (A3),

$$F(U) \geq \frac{\gamma}{2}\,||U||^2 - \beta||U||.$$

Hence $\lim F_{||U||\to\infty}F(U) = \infty$. Furthermore, $F(u) \geq -R$ if $||U|| \leq R$. This implies $\kappa > -\infty$. Choose a sequence (U_n) with $\lim_{n\to\infty} F(U_n) = \kappa$, and $||U_n|| \leq R$ for all n. From the identity

$$2a(U_n,U_n) + 2a(U_mU_m) = a(U_n - U_m, U_n - U_m) + a(U_n + U_m, U_n + U_m)$$

we get

$$\begin{aligned} F(U_n) + F(U_m) &= \frac{1}{4}a(U_n - U_m, U_n - U_m) + 2F\left(\frac{U_n + U_m}{2}\right) \\ &\geq \frac{\gamma}{4}||U_n - U_m||^2 + 2\kappa. \end{aligned}$$

Since $F(U_n) + F(U_m) \to 2\kappa$ as $n, m \to \infty$, the sequence (U_n) is Cauchy. Hence $U_n \to U$ as $n \to \infty$. As on page 339 for the inner product, it follows that F is continuous. Therefore, $F(U_n) \to F(U)$ as $n \to \infty$. Hence $F(U) = \kappa$.
(II) The Euler–Lagrange equation. If U is a solution of the minimum problem (10.67), then the same argument as in the proof of Theorem 10.26 shows that

$$a(U,V) = b(V) \qquad \text{for all} \quad V \in L.$$

(III) Uniqueness. If U and U_1 are solutions of (10.67), then $a(U - U_1, V) = 0$ for all $V \in L$, by (II). Choosing $V := U - U_1$, we get

$$\gamma||U - U_1||^2 \leq a(U - U_1, U - U_1) = 0,$$

and hence $U - U_1 = 0$.

10.2 *The Poincaré–Friedrichs inequality.* Let Ω be a nonempty bounded open set in $\mathbb{R}^N$ with $N = 1, 2, \ldots$. Show that there exists a constant $C > 0$ such that for all $U \in \overset{\circ}{W}{}_2^1(\Omega)$,

$$\int_\Omega U^2 \, d^N x \le C \int_\Omega \sum_{j=1}^N (\partial_j U)^2 \, d^N x.$$

Hint: Let $N = 1$ and $\Omega :=]a, b[$ Set $X := \overset{\circ}{W}{}_2^1(\Omega)$. For all $U \in \mathcal{D}(\Omega)$, it follows from the fundamental theorem of calculus that

$$U(x) = \int_a^x f(\xi) \cdot U'(\xi) \, d\xi, \qquad x \in [a, b]$$

with $f(x) \equiv 1$. Hence

$$U(x)^2 \le \int_a^x f(\xi)^2 d\xi \int_a^x U'(\xi)^2 d\xi \le (b-a) \int_a^b U'(\xi)^2 d\xi, \qquad (10.71)$$

by the Schwarz inequality. This implies

$$\int_a^b U(x)^2 dx \le (b-a)^2 \int_a^b U'(x)^2 dx.$$

If $U \in X$, then there exists a sequence (U_n) in $\mathcal{D}(\Omega)$ such that $U_n \to U$ in X as $n \to \infty$. Applying the limit $n \to \infty$ to

$$\int_a^b U_n(x)^2 \, dx \le (b-a)^2 \int_a^b |U_n'(x)|^2 dx,$$

we obtain the same inequality by replacing U_n by U. In the general case, use an N-dimensional interval (see Zeidler (1995), Vol. 1, Sect. 2.5.6).

10.3 *Prototype of the Sobolev embedding theorem.* Consider the bounded open interval $\Omega :=]a, b[$. Show that the embedding $\overset{\circ}{W}{}_2^1(\Omega) \subseteq C[a, b]$ is continuous.
Solution: Let $U \in \mathcal{D}(\Omega)$. By (10.71),

$$\max_{a \le x \le b} |U(x)| \le \left((b-a) \int_a^b U'(x)^2 dx \right)^{1/2}.$$

Noting that $||U||_{1,2} := (\int_a^b U(x)^2 + U'(x)^2 \, dx)^{1/2}$, we get

$$||U||_{C[a,b]} \le \text{const} \; ||U||_{1,2}.$$

This remains true for all functions $U \in W_2^1(\Omega)$, by using convergent sequences, as in the proof of Problem 10.2. The proofs of further prototypes for the crucial Sobolev embedding theorems can be found in Zeidler (1995), Vol. 1, Sect. 2.5.

10.4 *The fundamental solution of the Poisson equation.* Use the Green's formula (10.33) in order to prove Proposition 10.22.
Solution: Let B_R be an open ball of radius R centered at the point x_0. The trick invented by Green is to replace the set Ω by $\Omega \setminus B_R$ and letting $R \to 0$. By the Green's formula (10.33) on page 546,

$$\int_{\Omega \setminus B_R} U_{x_0} \Delta\varphi - \varphi \Delta U_{x_0} \, d^N x = \int_{\partial\Omega} \varphi \frac{\partial U_{x_0}}{\partial n} - U_{x_0} \frac{\partial \varphi}{\partial n} \, dS + \int_{\partial B_R} \varphi \frac{\partial U_{x_0}}{\partial n} - U_{x_0} \frac{\partial \varphi}{\partial n} \, dS$$

where $\frac{\partial}{\partial n}$ denotes the derivative in direction of the outer normal vector. Note that $\frac{\partial}{\partial n} = -\frac{\partial}{\partial r}$, since the outer normal vector of $\Omega \setminus B_R$ at the sphere ∂B_R points inside the ball B_R. Observe that $\Delta U_{x_0} = 0$ on $\Omega \setminus B_R$ and that the integral over $\partial\Omega$ vanishes, since the test function $\varphi \in \mathcal{D}(\Omega)$ vanishes in a boundary strip of $\partial\Omega$. Finally, letting $R \to 0$, we get

$$\lim_{R\to 0} \int_\Omega U_{x_0} \Delta\varphi \, d^N x = \varphi(x_0),$$

by using the following Green lemma.

10.5 *The Green lemma.* Let f be a real valued function which is continuous on some open neighborhood of the point x_0 in $\mathbb{R}^N$ with $N = 2, 3, , \ldots$. Show that

$$\lim_{R\to 0} \int_{||x-x_0||=R} f(x) \frac{\partial U_{x_0}(x)}{\partial r} \, dS = -f(x_0).$$

Solution: To simplify notation, we set $x_0 = 0$ and $N = 3$ along with $r := ||x||$. Then

$$U_{x_0}(x) = \frac{1}{4\pi r}, \qquad \frac{\partial U_{x_0}(x)}{\partial r} = -\frac{1}{4\pi r^2}.$$

By the mean value theorem for integrals, there exists a point y with $||y|| = R$ such that

$$\int_{||x||=R} f(x) \frac{\partial U_{x_0}(x)}{\partial r} \, dS = -\frac{f(y)}{4\pi R^2} \int_{||x||=R} dS = -f(y).$$

Finally, $\lim_{R\to 0} f(y) = f(0)$.

10.6 *The Poisson formula for the ball.* Prove Prop. 10.25. Hint: Study the proof in Jost (2000b), p. 14.

10.7 *The fundamental solution of the Helmholtz equation.* Fix $k \in \mathbb{R}$ Set $r := ||\mathbf{x}||$. Show that the function

$$\mathcal{G}(\mathbf{x}) := \frac{\mathrm{e}^{-ikr}}{r} \qquad \text{for all} \quad \mathbf{x} \in \mathbb{R}^3$$

satisfies the identity

$$\int_{R^3} \mathcal{G}(\mathbf{x})(\Delta\varphi(\mathbf{x}) - k^2\varphi(\mathbf{x})) \, d^3x = \varphi(\mathbf{0}) \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}^3).$$

In the language of distributions, this means that

$$(\Delta - k^2)\mathcal{G} = \delta \quad \text{on} \quad \mathbb{R}^3.$$

Hint: Argue as in the proof of Prop. 10.22 on page 561.

10.8 *Green's classical argument for the Helmholtz equation.* Let Ω be a nonempty bounded open subset in $\mathbb{R}^3$. Suppose that the function $U \in C^\infty(\Omega) \cap C(\overline{\Omega})$ is a solution of the Helmholtz equation $\Delta U - k^2 U = 0$ on Ω. Show that for all points $\mathbf{x}_1 \in \Omega$,

$$U(\mathbf{x}_1) = \int_{\partial\Omega} \mathcal{G}(\mathbf{x}_1 - \mathbf{x}) \, \frac{\partial U(\mathbf{x})}{\partial n} - \frac{\partial \mathcal{G}(\mathbf{x}_1 - \mathbf{x})}{\partial n} U(\mathbf{x}) \, dS.$$

Hint: Argue as in the proof of Theorem 10.24.

Further explicit material on the Green's function for important equations in mathematical physics can be found in Evans, Blackledge, and Yardley (2000).

11. Distributions and Green's Functions

Whoever understands Green's functions can understand forces in nature.
Folklore

The invention of the Green's function brought about a tool-driven revolution in mathematical physics, similar in character to the more famous tool-driven revolution caused by the invention of electronic computers a century and a half later... The Green's function and the computer are prime examples of intellectual tools. They are tools for clear thinking...

Invented in 1828 by George Green (1793–1841) and successfully applied to classical electromagnetism, acoustics, and hydrodynamics, Green's functions were the essential link between the theories of quantum electrodynamics by Schwinger, Feynman, and Tomonaga in 1948 and are still alive and well today...
I began the application of the Green's function to condensed matter physics in 1956 with a study of spin-waves in ferromagnets. I found that all the Green's function tricks that had worked so well in quantum electrodynamics worked even better in the theory of spin waves...
Meanwhile, the Green's functions method was applied systematically by Bogoliubov and other people to a whole range of problems in condensed matter physics. The main novelty in condensed matter physics was the appearance of temperature as an additional variable... A beautiful thing happens when you make the transition from ordinary Green's functions to thermal Green's functions. To make the transition, all you have to do is to replace the real frequency of any oscillation by a complex number whose real part is frequency and whose imaginary part is temperature[1]...
Soon after thermal Green's functions were invented, they were applied to solve the outstanding unsolved problem of condensed matter physics, the problem of superconductivity. They allowed Cooper, Bardeen, and Schrieffer to understand superconductivity as an effect of a particular thermal Green's functions expressing long-range phase-coherence between pairs of electrons (called Cooper pairs)...
In the 1960s, after Green's functions had become established as the standard working tools of theoretical analysis in condensed matter physics, the wheel of fashion in particle physics continued to turn. For a decade, quantum field theory and Green's functions were unfashionable in particle

[1] See G. Mahan, Many-Particle Physics, Plenum Press, New York, 1990.

physics. The prevailing view was that quantum field theory had failed in the domain of strong interactions[2]...

Then in the 1970s, the wheel of fashion turned once more. Quantum field theory was back in the limelight with two enormous successes, the Weinberg–Salam unified theory of electromagnetic and weak interactions, and the gauge theory of strong interactions now known as quantum chromodynamics. Green's functions were once again the working tools of calculation, both in particle physics and in condensed matter physics. And so they have remained up to the present day.

In the 1980's, quantum field theory moved off in a new direction, to lattice gauge theories in one direction and to superstring theory in another... The Wilson loop is the reincarnation of a Green's function in a lattice gauge theory[3] and there is a reincarnation of Green's functions in superstring theory.

Freeman Dyson
George Green and physics[4]

Between 1930 and 1940, several mathematicians began to investigate systematically the concept of a "weak" solution of a linear partial differential equation, which appeared episodically (and without a name) in Poincaré's work.

It was one of the main contributions of Laurent Schwartz when he saw, in 1945, that the concept of distribution introduced by Sobolev in 1936 (which he had rediscovered independently) could give a satisfactory generalization of the Fourier transform including all the preceding ones... By his own research and those of his numerous students, Laurent Schwartz began to explore the potentialities of distributions (generalized functions) and gradually succeeded in convincing the world of mathematicians that this new concept should become central in all problems of mathematical analysis, due to the greater freedom and generality it allowed in the fundamental operations of calculus, doing away with a great many unnecessary restrictions and pathology.

The role of Laurent Schwartz (born 1915) in the theory of distributions is very similar to the one played by Newton (1643–1727) and Leibniz (1646–1716) in the history of Calculus. Contrary to popular belief, they of course did not invent it, for derivation and integration were practiced by men such as Cavalieri (1598–1647), Fermat (1601–1665) and Roberval (1602–1675) when Newton and Leibniz were merely schoolboys. But they were able to systematize the algorithms and notations of Calculus in such a way that it became a versatile and powerful tool which we know, whereas before them it could only be handled via complicated arguments and diagrams.

Jean Dieudonné, 1981
History of Functional Analysis[5]

[2] See G. Chew (1966), The Analytic S-Matrix: A Basis for Nuclear Democracy, Benjamin, New York, and A. Barut, The Theory of the Scattering Matrix, MacMillan, New York, 1967.

[3] See I. Montvay and G. Münster, Quantum Fields on a Lattice, Cambridge University Press, 1994.

[4] Physics World, August 1993, pp. 33–38 (reprinted with permission).

[5] North Holland, Amsterdam, 1981 (reprinted with permission).

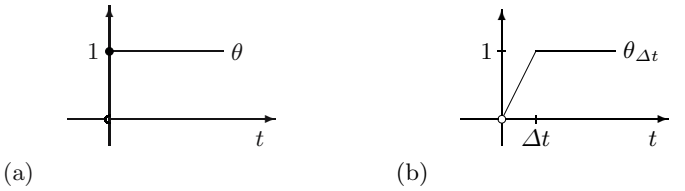


Fig. 11.1. Heaviside function

The local propagation of physical effects is described mathematically by Green's functions. In quantum field theory, special Green's functions are

- n-point correlation functions of quantum fields (n-point Green's functions).[6]

They are closely related to

- propagators,
- retarded propagators,
- advanced propagators.

The prototypes can be found in Sect. 11.1.2. In terms of physics, Green's functions describe physical processes under the influence of sharply concentrated external forces described by the Dirac delta function. General external forces are then obtained by the superposition principle. The Dirac delta function is not a classical object, but a generalized function (also called distribution). Therefore, distributions play a crucial role in physics, in particular, in quantum field theory. We do not suppose that the reader is familiar with this fundamental mathematical tool. Therefore, in this chapter, we will give an introduction to the mathematical theory of distributions and its physical interpretation.

The theory of distributions was created by Laurent Schwartz in 1945; it was motivated by Dirac's approach to quantum mechanics. This was represented in Dirac's famous 1930 monograph *Foundations of Quantum Mechanics.*[7] Distributions generalize a broad class of continuous and discontinuous functions. For a classical function, one has always to worry about the existence of derivatives. The situation changes completely for distributions.

Distributions possess derivatives of all orders.

Therefore, distributions are the right tool for the investigation of linear partial differential equations. For example, let us introduce the discontinuous Heaviside function

$$\theta(t) := \begin{cases} 1 & \text{if } t \geq 0, \\ 0 & \text{if } t < 0 \end{cases} \tag{11.1}$$

which jumps at the initial time $t = 0$ from zero to one (Fig. 11.1(a)). By convention, the Heaviside function is continuous from the right. In the language of distributions introduced in this chapter, the Heaviside function has the derivatives

$$\theta' = \delta, \quad \theta'' = \delta', \quad \theta''' = \delta'', \ldots$$

where δ denotes the Dirac delta distribution, and δ' denotes the first derivative of δ, and so on.[8] We will introduce the following spaces of test functions:

[6] The 2-point Green's function is also called the Feynman propagator.
[7] Laurent Schwartz was awarded the Fields medal in 1950.
[8] Heaviside (1850–1925) founded the theory of telephoning in his 1892 treatise *Electrical Papers.*

- the Schwartz space $\mathcal{D}(\mathbb{R}^N)$ of smooth functions $\varphi : \mathbb{R}^N \to \mathbb{C}$ which have compact support, and
- the Schwartz space $\mathcal{S}(\mathbb{R}^N)$ of smooth functions $\varphi : \mathbb{R} \to \mathbb{C}$ which are rapidly decreasing at infinity.

In addition, we will use

- the Hilbert space $L_2(\mathbb{R}^N)$ of square-integrable functions $\varphi : \mathbb{R}^N \to \mathbb{C}$,

$$\int_{\mathbb{R}^N} |\varphi(x)|^2 \, d^N x < \infty,$$

- the Schwartz space $\mathcal{D}'(\mathbb{R}^N)$ of distributions, and
- the Schwartz space $\mathcal{S}'(\mathbb{R}^N)$ of tempered distributions.

We have the following inclusions[9]

$$\boxed{\mathcal{D}(\mathbb{R}^N) \subset \mathcal{S}(\mathbb{R}^N) \subset L_2(\mathbb{R}^N) \subset \mathcal{S}'(\mathbb{R}^N) \subset \mathcal{D}'(\mathbb{R}^N).}$$

In particular, each tempered distribution is also a distribution, but the converse is not always true. Furthermore, the functions in the Hilbert space $L_2(\mathbb{R}^N)$ are tempered distributions, too. Tempered distributions are crucial for the Fourier transform. The N-dimensional Dirac delta distribution δ is characterized by the equation

$$\boxed{\delta(\varphi) = \varphi(0)}$$

for all test functions $\varphi \in \mathcal{S}(\mathbb{R}^N)$ (resp. all test functions $\varphi \in \mathcal{D}(\mathbb{R}^N)$). Therefore, δ is both a tempered distribution and a distribution, i.e., $\delta \in \mathcal{S}'(\mathbb{R}^N)$ and $\delta \in \mathcal{D}'(\mathbb{R}^N)$. If L a linear differential operator with constant coefficients, then the solutions F of the differential equation

$$\boxed{LF = \delta}$$

are called fundamental solutions of the differential operator L. One of the deepest theorems in the theory of distribution tells us that

> *Each nonzero linear differential operator L has a fundamental solution which is a tempered distribution.*

Green's functions are special fundamental functions which satisfy additional boundary conditions and initial conditions characterizing the specific physical situation. The Hilbert space $L_2(\mathbb{R}^N)$ is equipped with the inner product

$$\langle \psi | \varphi \rangle := \int_{\mathbb{R}^N} \psi(x)^\dagger \varphi(x) \, d^N x.$$

This space plays a fundamental role in quantum physics. However, in order to rigorously justify the elegant Dirac calculus used by physicists, one has to pass from the Hilbert space $L_2(\mathbb{R}^N)$ to the larger space $\mathcal{S}'(\mathbb{R}^N)$ of tempered distributions. More precisely, one has to use the triplet of spaces

$$\boxed{\mathcal{S}(\mathbb{R}^N) \subset L_2(\mathbb{R}^N) \subset \mathcal{S}'(\mathbb{R}^N)}$$

called the Gelfand triplet (or rigged Hilbert space $L_2(\mathbb{R}^N)$).

[9] We write $A \subseteq B$ (resp. $A \subset B$) iff A is a subset of B (resp. A is a proper subset of B).

Suggested reading. As an introduction to Green's functions, we recommend Economu (1988), Barton (1989) (the language of physicists), and Evans, Blackledge, and Yardley (2000) (the language of mathematicians). The following monographs serve as introductions to the theory of distributions and its applications to the partial differential equations of mathematical physics:

- Zeidler (1995), Vol. 1,
- Gelfand and Shilov (1964), Vol. 1,
- Stein and Shakarchi (2003)(harmonic analysis),
- Vladimirov (1966) (modern reformulation of the classic theory for partial differential equations in mathematical physics),
- Egorov, Komech, and Shubin (1999) (pseudo-differential operators and Fourier integral operators).

The relations between the theory of distributions and functional analysis can be found in

- Schwartz (1965), (1978), Vols. 1, 2,
- Reed and Simon (1972), Vols. 1–4,
- Triebel (1989), and
- Yosida (1995).

Comprehensive standard texts on the theory of distributions and its applications are

- Gelfand and Shilov (1964), Vols. 1–5,
- Hörmander (1983), Vols. 1–4,
- Egorov and Shubin (1991), Vols. 1–4.

Applications of the theory of distributions to quantum field theory are studied in

- Bogoliubov and Shirkov (1959),
- Bogoliubov et al. (1975), (1990),
- Jost (1965),
- Streater and Wightman (1968) (classic monograph on axiomatic quantum field theory),
- Reed and Simon (1972), Vol. 2 (axiomatic quantum field theory), Vol. 3 (Haag–Ruelle theorem in scattering theory),
- Simon (1974) (the $P(\varphi)_2$-model),
- Scharf (1995) (the Epstein–Glaser approach to quantum electrodynamics and renormalization), (2001) (gauge field theories),
- Manoukian (1983) (BPHZ renormalization).

A comprehensive list of fundamental solutions for partial differential equations can be found in Ortner and Wagner (1997).

System of units. In this chapter, we use the international SI system.

11.1 Rigorous Basic Ideas

The main goal of this chapter is to show how rigorous mathematics and the formal, but very useful language of physicists fit together. To begin with, we consider the discrete Dirac delta function. This will serve as a preparation for introducing the Dirac delta distribution below.

11.1.1 The Discrete Dirac Delta Function

The discrete Dirac delta function is the key to the Dirac delta distribution.
Folklore

For a fixed number $\Delta t > 0$, the discrete delta function is defined by

$$\delta_{\Delta t}(t) := \begin{cases} \frac{1}{\Delta t} & \text{if } 0 \le t \le \Delta t, \\ 0 & \text{otherwise.} \end{cases}$$

Intuitively, this is a force which acts on the small time interval $[0, \Delta t]$. This force goes to infinity as $\Delta t \to 0$. The Dirac delta function to be considered in Sect. 11.2 on page 589 refers to the limit of $\delta_{\Delta t}$ as $\Delta t \to +0$.

Proposition 11.1 *Fix time t_0. For each continuous function $\varphi : \mathbb{R} \to \mathbb{C}$,*

$$\lim_{\Delta t \to +0} \int_{\mathbb{R}} \delta_{\Delta t}(t - t_0)\varphi(t)dt = \varphi(t_0).$$

Proof. By the mean value theorem for integrals, there exists a real number $\tau \in [t_0, t_0 + \Delta t]$ such that

$$\int_{\mathbb{R}} \delta_{\Delta t}(t - t_0)\varphi(t)dt = \frac{1}{\Delta t} \int_{t_0}^{t_0+\Delta t} \varphi(t)dt = \varphi(\tau).$$

By continuity, $\varphi(\tau)$ goes to $\varphi(t_0)$ as $\Delta t \to +0$. □

The Fourier transform. Let us now compute the Fourier transform of the discrete Dirac delta function,

$$(\mathcal{F}\delta_{\Delta t})(\omega) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \delta_{\Delta t}(t)\mathrm{e}^{-\mathrm{i}\omega t}dt = \frac{1}{\Delta t\sqrt{2\pi}} \int_0^{\Delta t} \mathrm{e}^{-\mathrm{i}\omega t}dt.$$

For all nonzero real frequencies ω,

$$(\mathcal{F}\delta_{\Delta t})(\omega) = \frac{1 - \mathrm{e}^{-\mathrm{i}\omega\Delta t}}{\mathrm{i}\omega\Delta t\sqrt{2\pi}}.$$

If $\omega = 0$, then $(\mathcal{F}\delta_{\Delta t})(0) = 1/\sqrt{2\pi}$. This yields the following limit.

Proposition 11.2 *For all real frequencies ω,*

$$\lim_{\Delta t \to +0} (\mathcal{F}\delta_{\Delta t})(\omega) = \frac{1}{\sqrt{2\pi}}.$$

Fix real time t_0. For all real times t, we obtain the formula

$$\delta_{\Delta t}(t - t_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mathcal{F}\delta_{\Delta t})(\omega)\mathrm{e}^{\mathrm{i}\omega(t-t_0)}d\omega.$$

Furthermore, for all real frequencies ω,

$$\mathrm{e}^{-\mathrm{i}\omega t_0}(\mathcal{F}\delta_{\Delta t})(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \delta_{\Delta t}(t - t_0)\mathrm{e}^{-\mathrm{i}\omega t}dt.$$

Approximation of white noise. In terms of physics, the discrete Dirac delta function can be represented as a superposition of plane waves,

$$\delta_{\Delta t}(t) = \int_{-\infty}^{\infty} a_{\Delta t}(\omega) \mathrm{e}^{\mathrm{i}\omega t} d\omega, \qquad t \in \mathbb{R}$$

with the amplitude

$$a_{\Delta t}(\omega) := \frac{(\mathcal{F}\delta_{\Delta t})(\omega)}{\sqrt{2\pi}}, \qquad \omega \in \mathbb{R}.$$

Since $\lim_{\Delta t \to +0} a_{\Delta t}(\omega) = 1/2\pi$, the amplitudes do not depend on the frequency ω, as $\Delta t \to +0$. Physicists say that

> *The discrete Dirac delta function $\delta_{\Delta t}$ approximates white noise as $\Delta t \to +0$.*

The term 'white' comes from the fact that white light is a superposition of electromagnetic waves of all frequencies, and each frequency contributes approximately the same amplitude.

11.1.2 Prototypes of Green's Functions

The notion of Green's function was introduced by George Green (1793–1841) in the year 1828.[10]

> *The Green's function describes the behavior of a physical system by kicking it with a force which acts only during a very small time interval and which is concentrated on a very small neighborhood of some point in the position space.*

This generalizes Newton's infinitesimal strategy from mechanics to field theories.

Motion of a classical particle on the real line. Consider the following initial-value problem

$$\boxed{m\ddot{x}(t) = F(t), \quad t \in \mathbb{R}, \quad x(t_0) = x_0,\ \dot{x}(t_0) = v_0.} \tag{11.2}$$

We are given the initial time t_0, the initial position x_0, the initial velocity v_0, and the smooth force function $F : \mathbb{R} \to \mathbb{R}$. We are looking for the smooth position function $x : \mathbb{R} \to \mathbb{R}$. In terms of physics, problem (11.2) describes the motion $x = x(t)$ of a classical particle of mass $m > 0$ on the real line under the influence of the force F. This force acts from left to right if it is positive. To simplify notation, set $m := 1$.

Proposition 11.3 *The initial-value problem* (11.2) *has the unique solution*

$$x(t) = x_0 + P(t, t_0)v_0 + \int_{t_0}^{t} P(t, \tau)F(\tau)d\tau, \qquad t \in \mathbb{R}. \tag{11.3}$$

The function $P(t, \tau) := t - \tau$ is called the propagator defined for all $t, \tau \in \mathbb{R}$.

This proposition tells us that the propagator knows all about the motion of the mass point.

Proof. (I) Uniqueness. If there are two solutions x and x_*, then set $y := x - x_*$. Thus, $\ddot{y} = 0$ on $\mathbb{R}$ with $y(0) = \dot{y}(0) = 0$. Hence $y(t) = a + bt$ with $a = b = 0$.

(II) Existence. Differentiation with respect to time t yields the equation

[10] See the footnote on page 544.

$$\dot{x}(t) = v_0 + P(t,t)F(t) + \int_{t_0}^{t} P_t(t,\tau)F(\tau)d\tau = v_0 + \int_{t_0}^{t} F(\tau)d\tau.$$

This implies $\ddot{x}(t) = F(t)$. □

For $t, t_0 \in \mathbb{R}$, the following two functions

$$P_{\text{ret}}(t,t_0) := \theta(t-t_0)P(t,t_0)$$

and

$$P_{\text{adv}}(t,t_0) := -\theta(t_0-t)P(t,t_0)$$

are called the retarded and advanced propagator, respectively. Obviously,

$$P(t,t_0) = P_{\text{ret}}(t,t_0) - P_{\text{adv}}(t,t_0) \qquad \text{for all} \quad t,t_0 \in \mathbb{R}.$$

The retarded propagator describes the motion of the mass point in the case where the force F is switched on at time t_0. That is, we replace $F = F(t)$ by the function $F_{\text{ret}}(t) := \theta(t-t_0)F(t)$. Explicitly, the function

$$x(t) := x_0 + P_{\text{ret}}(t,t_0)v_0 + \int_{t_0}^{t} P_{\text{ret}}(t,\tau)F(\tau)d\tau, \qquad t \in \mathbb{R}$$

is a solution of problem (11.2) for $t > t_0$. Furthermore, $x(t) = x_0$ if $t \le t_0$ (i.e., the particle rests until time t_0).

The Green's function. To simplify notation, set $t_0 = 0$. The solution from Prop. 11.3 can be written as

$$\boxed{x(t) = x_0 + P(t,0)v_0 + \int_{-\infty}^{\infty} G(t,\tau)F(\tau)d\tau, \qquad t \in \mathbb{R}.} \tag{11.4}$$

The function G is called the Green's function of the initial-value problem (11.2). Explicitly, we set

$$G(t,\tau) := G_{\text{ret}}(t,\tau) - G_{\text{adv}}(t,\tau), \qquad t,\tau \in \mathbb{R}$$

along with the retarded Green's function

$$G_{\text{ret}}(t,\tau) := \begin{cases} P(t,\tau) & \text{if} \quad 0 \le \tau \le t, \\ 0 & \text{otherwise} \end{cases}$$

and the advanced Green's function

$$G_{\text{adv}}(t,\tau) := \begin{cases} P(t,\tau) & \text{if} \quad t \le \tau < 0, \\ 0 & \text{otherwise.} \end{cases}$$

The equations (11.2) and (11.4) reflect the duality between differential- and integral relations. In order to understand the physical meaning of the Green's function, choose the initial condition $x(0) = \dot{x}(0) = 0$, that is, the particle rests at time $t = 0$. Furthermore, for small $\Delta t > 0$ and fixed time t_1, we choose the force

$$F(t) := \delta_{\Delta t}(t - t_1) \qquad t \in \mathbb{R}.$$

The initial-value problem (11.2) has then the classical solution

$$x(t) = \int_{-\infty}^{\infty} G(t,\tau)\delta_{\Delta t}(\tau - t_1)d\tau$$

for all times $t \in \mathbb{R}$ different from the points t_1 and $t_1 + \Delta t$ where the force jumps. The position function $x = x(t)$ is continuous for all times $t \in \mathbb{R}$.

Proposition 11.4 *For all times $t, t_1 \in \mathbb{R}$ with $t \neq t_1$,*

$$\lim_{\Delta t \to +0} \int_{-\infty}^{\infty} G(t,\tau)\delta_{\Delta t}(\tau - t_1)d\tau = G(t,t_1).$$

This proposition tells us that the Green's function describes the response of the particle to a constant force F acting on the small time interval $[t_1, t_1 + \Delta t]$ normalized by $\int_{-\infty}^{\infty} F(t)dt = 1$. In addition, we assume that the particle rests at the initial time $t = 0$.

Proof. The Green's function $G = G(t,\tau)$ is continuous for all $(t,\tau) \in \mathbb{R}^2$ with $t \neq \tau$. Thus, for $t \neq t_1$,

$$\lim_{\Delta t \to +0} \int_{-\infty}^{\infty} G(t,\tau)\delta_{\Delta t}(\tau - t_1)\, d\tau = \lim_{\Delta t \to +0} \frac{1}{\Delta t} \int_{t_1}^{t_1+\Delta t} G(t,\tau)d\tau = G(t,t_1).$$

□

One checks easily that

$$G(t,t_1) = G(t - t_1, 0) \qquad \text{for all} \quad t, t_1 \in \mathbb{R}.$$

This relation reflects causality.

The prototype of the Schrödinger equation. Let A be a fixed complex number. We want to study the initial-value problem

$$\dot{x}(t) = Ax(t) + F(t), \quad t \in \mathbb{R}, \quad x(t_0) = x_0. \tag{11.5}$$

We are given the initial time t_0, the initial position $x_0 \in \mathbb{C}$, and the smooth external source $F : \mathbb{R} \to \mathbb{C}$. We are looking for a smooth function $x : \mathbb{R} \to \mathbb{C}$.

Proposition 11.5 *Problem* (11.5) *has the unique solution*

$$x(t) = P(t,t_0)x_0 + \int_{t_0}^{t} P(t,\tau)F(\tau)d\tau, \qquad t \in \mathbb{R}$$

with the propagator $P(t,\tau) := e^{A(t-\tau)}$ for all $t, \tau \in \mathbb{R}$.

The proof parallels the proof of Proposition 11.3. Replacing the propagator $P(t,\tau) := t - \tau$ by the new propagator $P(t,\tau) := \mathrm{e}^{A(t-\tau)}$, we introduce the functions $P_{\mathrm{ret}}, P_{\mathrm{adv}}, G, G_{\mathrm{ret}}, G_{\mathrm{adv}}$ as described above. In particular, Prop. 11.4 remains valid.

Oscillating classical mass point with non-critical angular frequency. We are given the time interval $[-\frac{T}{2}, \frac{T}{2}]$ of length $T > 0$. Introduce the following critical angular frequencies

$$\omega_k := \frac{2\pi k}{T}, \qquad k = 0, 1, 2, \ldots.$$

Consider the problem

$$\boxed{m\ddot{x}(t) + \omega^2 x(t) = F(t), \qquad t \in \mathbb{R}.} \tag{11.6}$$

To simplify notation, choose the unit mass, $m := 1$. For fixed $K = 1, 2, \ldots$, we are given the smooth T-periodic force function

$$F(t) := \sum_{k=-K}^{K} b_k \mathrm{e}^{2\pi \mathrm{i} k t/T}, \qquad t \in \mathbb{R}$$

where the Fourier coefficients b_k are complex numbers. Furthermore, we are given the angular frequency $\omega > 0$ such that $\omega \neq \omega_k$ for all $k = 1, 2, \ldots$ We are looking for the smooth T-periodic position function $x : \mathbb{R} \to \mathbb{C}$. The real and imaginary parts of the function $x = x(t)$ describe the oscillation of a mass point on the real line of time period T.

Proposition 11.6 *Problem* (11.6) *has the unique solution*

$$x(t) = \int_{-T/2}^{T/2} G(t,\tau)F(\tau)d\tau, \qquad t \in \mathbb{R}$$

with the continuous Green's function

$$G(t,\tau) := \sum_{k=-\infty}^{\infty} \frac{e^{2\pi \mathrm{i}k(t-\tau)/T}}{T(\omega^2 - \omega_k^2)}, \qquad t, \tau \in \mathbb{R}.$$

Proof. To simplify notation, set $T := 2\pi$. The general case follows then by rescaling of time.

(I) Existence. Choose $F(t) := b_k \mathrm{e}^{\mathrm{i}kt}$ for fixed $k = 0, \pm 1, \pm 2, \ldots$ Then the function

$$x(t) = \frac{b_k \mathrm{e}^{\mathrm{i}kt}}{\omega^2 - k^2}, \qquad t \in \mathbb{R}$$

is a solution of (11.6). By the orthogonality relation $\int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}nt} dt = 2\pi\delta_{n,0}$ for all integers n, we get

$$b_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}k\tau} \, F(\tau) d\tau.$$

Hence

$$x(t) = \int_{-\pi}^{\pi} \frac{\mathrm{e}^{\mathrm{i}k(t-\tau)}}{2\pi(\omega^2 - k^2)} \, F(\tau) \, d\tau.$$

Again by the orthogonality relation,

$$x(t) = \sum_{n=-K}^{K} \int_{-\pi}^{\pi} \frac{\mathrm{e}^{\mathrm{i}n(t-\tau)}}{2\pi(\omega^2 - n^2)} \, F(\tau) \, d\tau$$

for all K with $K > |k|$. Letting $K \to +\infty$, we get

$$x(t) = \int_{-\pi}^{\pi} G(t,\tau)F(\tau) \, d\tau, \qquad t \in \mathbb{R}.$$

Note that integration and the limiting process $K \to \infty$ can be interchanged by means of the majorant condition

$$\left| \frac{\mathrm{e}^{\mathrm{i}k(t-\tau)}}{2\pi(\omega^2 - n^2)} \right| \leq \frac{\text{const}}{1 + n^2}, \qquad n = 0, \pm 1, \pm 2, \ldots$$

along with $\sum_{n=0}^{\infty} \frac{1}{1+n^2} < \infty$. For the finite sum $F(t) = \sum_k b_k \mathrm{e}^{\mathrm{i}kt}$, we use the superposition principle.

(II) Uniqueness. Let $t \mapsto x_1(t)$ and $t \mapsto x_2(t)$ be two solutions of (11.6). For the difference $x(t) := x_1(t) - x_2(t)$, we get $\ddot{x} + \omega^2 x = 0$. Since $x : \mathbb{R} \to \mathbb{C}$ is 2π-periodic and smooth, we have the Fourier series

$$x(t) = \sum_{k=-\infty}^{\infty} a_k \mathrm{e}^{\mathrm{i}kt}, \qquad t \in \mathbb{R}.$$

Moreover, for each $r = 1, 2, \ldots$, we have the estimate

$$|a_k| \le \frac{\mathrm{const}(r)}{1 + k^r} \qquad \text{for all} \quad k = 0, \pm 1, \pm 2, \ldots, \tag{11.7}$$

by (10.9) on page 534. Hence the Fourier series can be differentiated term by term. This yields

$$\ddot{x}(t) + \omega^2 x(t) = \sum_{k=-\infty}^{\infty} (\omega^2 - k^2) a_k \mathrm{e}^{\mathrm{i}kt} = 0.$$

This implies $(\omega^2 - k^2)a_k = 0$ for all integers k. Hence $a_k = 0$ for all k. Therefore, $x(t) \equiv 0$. □

The critical resonance case and the method of regularization. For fixed $k_0 = 1, 2, \ldots$, choose the critical angular frequency

$$\omega = \omega_{k_0}.$$

In this case, the homogeneous problem (11.6) with $F(t) \equiv 0$ has the general solution

$$x(t) = a_{k_0} \mathrm{e}^{2\pi \mathrm{i} k_0 t/T} + a_{-k_0} \mathrm{e}^{-2\pi \mathrm{i} k_0 t/T}, \qquad t \in \mathbb{R}$$

with arbitrary complex numbers $a_{\pm k_0}$. These solutions correspond to eigenoscillations with the angular frequency $\omega_{k_0} = 2\pi k_0/T$.

Proposition 11.7 *The inhomogeneous problem* (11.6) *has a solution iff the external force F satisfies the non-resonance condition $b_{k_0} = b_{-k_0} = 0$. The general solution of* (11.6) *is then given by*

$$x(t) = a_{k_0} e^{2\pi \mathrm{i} k_0 t/T} + a_{-k_0} e^{-2\pi \mathrm{i} k_0 t/T} + \lim_{\varepsilon \to +0} \int_{-T/2}^{T/2} G_\varepsilon(t, \tau) F(\tau) d\tau, \qquad t \in \mathbb{R}$$

where $a_{\pm k_0}$ are arbitrary complex numbers. For each $\varepsilon > 0$, the function

$$G_\varepsilon(t, \tau) := \sum_{k=-\infty}^{\infty} \frac{e^{2\pi \mathrm{i} k(t-\tau)/T}}{T(\omega^2 - \omega_k^2 + \mathrm{i}\varepsilon)}, \qquad t, \tau \in \mathbb{R}$$

is called the regularized Green's function.

Proof. Use the same argument as in the proof of Prop. 11.6 above. Note the following peculiarity. Let $T = 2\pi$. If $x = x(t)$ is a solution of (11.6), then

$$(-k^2 + \omega_{k_0}^2) a_k = b_k, \qquad k = 0, \pm 1, \pm 2, \ldots$$

Since $\omega_{k_0} = k_0^2$, we get $b_{k_0} = b_{-k_0} = 0$. Conversely, let $b_{k_0} = b_{-k_0} = 0$. Then, the dangerous terms

$$\frac{b_{\pm k_0} \mathrm{e}^{\pm \mathrm{i} k_0 t}}{\omega_{k_0}^2 - k_0^2 + \mathrm{i}\varepsilon}$$

drop out. Therefore, the limit

$$\lim_{\varepsilon \to +0} \int_{-T/2}^{T/2} G_\varepsilon(t, \tau) F(\tau) d\tau, \qquad t \in \mathbb{R}$$

is well-defined, and it represents a special solution of (11.6). □

11.1.3 The Heat Equation and the Heat Kernel

We now want to study the prototype of the Green's function for a classical field theory. This is Fourier's famous heat kernel. Let $\psi(\mathbf{x}, t)$ be the temperature of a homogeneous body at position $\mathbf{x}$ and time t. By Fourier's law, variations of the temperature generate a heat current density vector given by

$$\mathbf{J} = -\kappa\, \mathbf{grad}\, \psi$$

where the material constant $\kappa > 0$ is called the heat conductivity of the body. The conservation law for heat,

$$\frac{\partial \psi}{\partial t} + \operatorname{div} \mathbf{J} = 0,$$

yields the heat conduction equation

$$\boxed{\frac{\partial \psi}{\partial t} = -\kappa \Delta \psi.} \tag{11.8}$$

In his 1822 treatise *On the Theory of Heat*, Fourier used the Fourier transform in order to obtain the famous solution formula

$$\boxed{\psi(\mathbf{x}, t) = \int_{\mathbb{R}^3} \mathcal{K}(\mathbf{x}, t; \mathbf{y}, s) \psi(\mathbf{y}, s) d^3\mathbf{y}, \qquad \mathbf{x} \in \mathbb{R}^3, \;\; t > s} \tag{11.9}$$

with the so-called heat kernel

$$\mathcal{K}(\mathbf{x}, t; \mathbf{y}, s) := \left(\frac{1}{4\pi\kappa(t-s)} \right)^{\frac{3}{2}} \mathrm{e}^{-\frac{(\mathbf{x}-\mathbf{y})^2}{4\kappa(t-s)}}$$

for all position vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$, and all times $t, s \in \mathbb{R}$ with $t > s$. If we know the temperature $\psi(\mathbf{y}, s)$ at the initial time s, then formula (11.9) yields the temperature $\psi(\mathbf{x}, t)$ at each later time $t > s$. In order to get the physical interpretation of the heat kernel, choose an initial temperature $\psi(\mathbf{y}, s) := \delta_\varepsilon(\mathbf{y})$ at time s which is localized in a small neighborhood of the point $\mathbf{y} = 0$. Explicitly,

$$\delta_\varepsilon(\mathbf{y}) := \begin{cases} \frac{1}{V(\varepsilon)} & \text{if} \quad ||\mathbf{y}|| \leq \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

where $V(\varepsilon) = \frac{4}{3}\pi\varepsilon^3$ is the volume of a sphere of radius ε. Letting $\varepsilon \to 0$, we get the temperature

$$\mathcal{K}(\mathbf{x}, t; 0, s) = \lim_{\varepsilon \to 0} \int_{\mathbb{R}^3} \mathcal{K}(\mathbf{x}, t; \mathbf{y}, s) \delta_\varepsilon(\mathbf{y}) d^3\mathbf{y}$$

at the point $\mathbf{x}$ at time $t > 0$. Physicists say briefly that the initial temperature $\delta(\mathbf{y})$, which is sharply concentrated at the origin at time s, induces the temperature $\mathcal{K}(\mathbf{x}, t; 0, s)$ at each point $\mathbf{x}$ and each time $t > s$.

The classical approach. For the temperature function $\psi = \psi(\mathbf{x}, t)$, consider the initial-value problem

$$\begin{aligned} \frac{\partial \psi(\mathbf{x}, t)}{\partial t} &= -\kappa \Delta \psi(\mathbf{x}, t) + f(\mathbf{x}, t), \qquad && \mathbf{x} \in \mathbb{R}^3, \; t \geq 0, \\ \psi(\mathbf{x}, 0) &= \psi_0(\mathbf{x}), && \mathbf{x} \in \mathbb{R}^3. \end{aligned} \tag{11.10}$$

Here, $\kappa > 0$. We are given both the smooth initial temperature $\psi_0 \in \mathcal{S}(\mathbb{R}^3)$, and the external heat source $f = f(\mathbf{x}, t)$ which is smooth on $\mathbb{R}^3 \times [0, \infty[$. In addition, we assume that the function $\mathbf{x} \mapsto f(\mathbf{x}, t)$ lies in the space $\mathcal{S}(\mathbb{R}^3)$ for each time $t \geq 0$. We are looking for a solution $\psi = \psi(\mathbf{x}, t)$ which is smooth on $\mathbb{R}^3 \times]0, \infty[$, continuous on $\mathbb{R}^3 \times [0, \infty[$, and bounded on $\mathbb{R}^3 \times [0, t_1]$ for each $t_1 > 0$.

Theorem 11.8 *The initial-value problem* (11.10) *has a unique solution given by*

$$\psi(\mathbf{x}, t) = \int_{\mathbb{R}^3} \mathcal{K}(\mathbf{x}, t; \mathbf{y}, 0)\, \psi_0(\mathbf{y})\, d^3y + \int_0^t d\tau \int_{\mathbb{R}^3} \mathcal{K}(\mathbf{x}, t; \mathbf{y}, \tau)\, f(\mathbf{y}, \tau)\, d^3y \tag{11.11}$$

for all position vectors $\mathbf{x} \in \mathbb{R}$ *and all times* $t > 0$.

The heat kernel is also called the Feynman propagator kernel of the heat equation. The proof of this classical result can be found in Triebel (1989), Sect. 41. In fact, careful differentiation of (11.11) shows that ψ is a solution of (11.10). Uniqueness follows from the maximum principle.

Perturbation theory. Instead of (11.10), let us consider the modified initial-value problem

$$\begin{aligned} \frac{\partial \psi(\mathbf{x}, t)}{\partial t} &= -\kappa \Delta \psi(\mathbf{x}, t) - U(\mathbf{x}) \psi(\mathbf{x}, t), && \mathbf{x} \in \mathbb{R}^3,\ t \geq 0, \\ \psi(\mathbf{x}, 0) &= \psi_0(\mathbf{x}), && \mathbf{x} \in \mathbb{R}^3. \end{aligned} \tag{11.12}$$

By Theorem 11.8, we obtain the Volterra integral equation

$$\psi(\mathbf{x}, t) = \int_{\mathbb{R}^3} \mathcal{K}(\mathbf{x}, t; \mathbf{y}, 0)\, \psi_0(\mathbf{y})\, d^3y - \int_0^t d\tau \int_{\mathbb{R}^3} \mathcal{K}(\mathbf{x}, t; \mathbf{y}, \tau) U(\mathbf{y})\, \psi(\mathbf{y}, \tau)\, d^3y$$

which can be solved by an iterative method. As first approximation, we get

$$\psi(\mathbf{x}, t) = \int_{\mathbb{R}^3} \mathcal{K}(\mathbf{x}, t; \mathbf{y}, 0)\, \psi_0(\mathbf{y})\, d^3y - \int_0^t d\tau \int_{\mathbb{R}^3} \mathcal{K}(\mathbf{x}, t; \mathbf{y}, \tau) U(\mathbf{y})\, \psi_0(\mathbf{y})\, d^3y.$$

11.1.4 The Diffusion Equation

The heat equation (11.8) also describes a diffusion process in 3-dimensional space. In this case, $\psi(\mathbf{x}, t)$ describes the density of a fluid at the point $\mathbf{x}$ at time t. The number $\kappa > 0$ is called the diffusion coefficient.

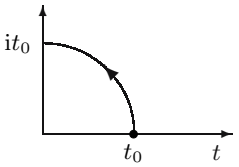


Fig. 11.2. Wick rotation

11.1.5 The Schrödinger Equation and the Euclidean Approach

We now consider the Schrödinger equation for a free particle,

$$\boxed{\mathrm{i}\frac{\partial\psi}{\partial t} = \frac{\hbar}{2m}\Delta\psi.} \tag{11.13}$$

This equation is obtained from the heat conduction equation (11.8) with the parameter $\kappa := \hbar/2m$ by replacing real time t by imaginary time $\mathrm{i}t$. This way, we formally obtain the Feynman propagator kernel for a free quantum particle

$$\boxed{\mathcal{P}_+(\mathbf{x},t;\mathbf{y},s) = \mathcal{K}(\mathbf{x},\mathrm{i}t;\mathbf{y},\mathrm{i}s) = \left(\frac{m}{\mathrm{i}h(t-s)}\right)^{\frac{3}{2}} \mathrm{e}^{\frac{\mathrm{i}m(\mathbf{x}-\mathbf{y})^2}{2\hbar(t-s)}}} \tag{11.14}$$

for $t > s$. This corresponds to an analytic continuation from the real t axis to the imaginary axis by a counterclockwise rotation of angle $\frac{\pi}{2}$. This rotation is called a Wick rotation (Fig. 11.2). Thus, $\mathrm{i}^{\frac{3}{2}}$ is to be understood as $\mathrm{e}^{\frac{3\pi\mathrm{i}}{4}}$.

The free quantum particle. Consider the initial-value problem

$$\begin{aligned} \mathrm{i}\hbar\frac{\partial\psi}{\partial t}(\mathbf{x},t) &= \frac{\mathbf{P}^2}{2m}\psi(\mathbf{x},t), \qquad && \mathbf{x}\in\mathbb{R}^3,\ t>0, \\ \psi(\mathbf{x},0) &= \psi_0(\mathbf{x}), && \mathbf{x}\in\mathbb{R}^3. \end{aligned} \tag{11.15}$$

Here, $\mathbf{P} = -\mathrm{i}\hbar\boldsymbol{\partial}$. This Schrödinger equation describes the motion of a free quantum particle of mass m (e.g., an electron) on the 3-dimensional Euclidean space. Replacing time t by $\mathrm{i}t$ and applying formal analytic continuation to the solution formula (11.11), we get

$$\boxed{\psi(\mathbf{x},t) = \int_{\mathbb{R}^3} \mathcal{P}_+(\mathbf{x},t;\mathbf{y},0)\psi_0(\mathbf{y})d^3y, \qquad \mathbf{x}\in\mathbb{R}^3,\ t>0.}$$

Explicitly,

$$\psi(\mathbf{x},t) = \Big(\frac{m}{\mathrm{i}ht}\Big)^{\frac{3}{2}} \int_{\mathbb{R}^3} \mathrm{e}^{-\frac{m(\mathbf{x}-\mathbf{y})^2}{2\mathrm{i}\hbar t}}\, \psi_0(\mathbf{y})\, d^3y, \quad \mathbf{x}\in\mathbb{R}^3,\ t>0. \tag{11.16}$$

In order to justify this formula rigorously, let us introduce the Hilbert space $L_2(\mathbb{R}^3)$. Define

$$H_0 := \frac{\mathbf{P}^2}{2m}\,.$$

More precisely, as the domain of definition of the operator

$$H_0 : D(H_0) \subseteq L_2(\mathbb{R}^3) \to L_2(\mathbb{R}^3),$$

we choose the Sobolev space $D(H_0) := W_2^2(\mathbb{R}^3)$. By definition, this space consists of all functions $\varphi \in L_2(\mathbb{R}^3)$ whose first and second partial derivatives, in the sense of generalized functions, lie in $L_2(\mathbb{R}^3)$, too (see page 611). Since the operator H_0 is self-adjoint, the operator $\mathrm{e}^{-\mathrm{i}H_0 t/\hbar} : L_2(\mathbb{R}^3) \to L_2(\mathbb{R}^3)$ is unitary, by the classical Stone theorem. The Hilbert space function

$$\boxed{\psi(t) := \mathrm{e}^{-\mathrm{i}H_0 t/\hbar}\,\psi_0, \qquad t \in \mathbb{R}} \tag{11.17}$$

represents the generalized solution of (11.15) for each given $\psi_0 \in L_2(\mathbb{R}^3)$.

Theorem 11.9 *For each given function $\psi_0 \in \mathcal{S}(\mathbb{R}^3)$, the solution* (11.17) *coincides with* (11.16).

The proof can be found in Zeidler (1995), Vol. 1, Sect. 5.22.2.

11.2 Dirac's Formal Approach

> Our work led us to consider quantities involving a certain kind of infinity. To get a precise notation for dealing with these infinities, we introduce a quantity $\delta(t)$ depending on a parameter t satisfying a condition
>
> $$\int_{-\infty}^{\infty} \delta(t)dt = 1 \qquad \text{and} \qquad \delta(t) = 0 \quad \text{for} \quad t \neq 0. \tag{11.18}$$
>
> This is not a function according to the usual mathematical definition of a function, but is something more general... Thus $\delta(t)$ is not a quantity which can be generally used in mathematical analysis like an ordinary function, but its use must be confined to certain simple kinds of expressions for which it is obvious that no inconsistency can arise.
>
> Paul Dirac, 1930
> *Principles of Quantum Mechanics*[11]

> In order to understand the Dirac delta function δ, use the discrete Dirac delta function $\delta_{\Delta t}$ and carry out the limit $\Delta t \to 0$ at the right place.
>
> Folklore

In this section, let us discuss the Dirac delta function in the language used by physicists. The following formal considerations will be given a rigorous meaning later on. However, it is useful to know the language of physicists as a mnemonic tool for the rigorous approach.

[11] Clarendon Press, Oxford, 1930. Fourth edition 1981. Reprinted by permission of Oxford University Press.

11.2.1 Dirac's Delta Function

The use of the Dirac delta function in physics is based on the following formal formulas:

(i) Limit: $\lim_{\Delta t\to 0} \delta_{\Delta t}(t) = \delta(t)$ for all $t \in \mathbb{R}$.
(ii) Support: $\delta(t) = 0$ for all $t \in \mathbb{R}$ with $t \neq 0$.
(iii) Characteristic property: Fix the real time t_0. For each continuous function $\varphi : \mathbb{R} \to \mathbb{C}$,

$$\boxed{\int_{\mathbb{R}} \delta(t - t_0)\varphi(t)dt = \varphi(t_0).} \tag{11.19}$$

Moreover, $\delta(t - t_0) = \delta(t_0 - t)$ for all $t, t_0 \in \mathbb{R}$.
(iv) Fourier transform: Fix the real time t_0. For all real frequencies ω,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \delta(t - t_0)\mathrm{e}^{-\mathrm{i}\omega t}dt = \frac{\mathrm{e}^{-\mathrm{i}\omega t_0}}{\sqrt{2\pi}}.$$

In particular, for $t_0 = 0$ we get

$$(\mathcal{F}\delta)(\omega) = \frac{1}{\sqrt{2\pi}}, \qquad \omega \in \mathbb{R}.$$

(v) Inverse Fourier transform: For all real times t,

$$\delta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\omega t}d\omega.$$

This corresponds to $\delta(t) = \mathcal{F}^{-1}\left(\frac{1}{\sqrt{2\pi}}\right)(t)$.

Motivation. The relations (ii) through (v) can be obtained from Sect. 11.1.1 by using the corresponding relations for the discrete Dirac delta function $\delta_{\Delta t}$ and formally carrying out the limit $\Delta t \to 0$. Mnemonically, the Fourier transformation formula (iv) follows immediately from the characteristic property (iii). This implies (v) by using the inverse Fourier transform from Sect. 10.3.3 on page 535 in a formal way.

The need for introducing generalized functions. There does not exist any classical function δ which satisfies relations (i) through (v) above. In fact, if a classical solution $f : \mathbb{R} \to \mathbb{C}$ has the property that $f(t) = 0$ for all $t \in \mathbb{R}$ with $t \neq 0$, then

$$\int_{\mathbb{R}} f(t)\varphi(t)dt = 0$$

for all smooth functions $\varphi : \mathbb{R} \to \mathbb{C}$. Therefore, relations (ii) and (iii) are contradictory for a classical function. However, we will show below that there exists a generalized function δ in the sense of Laurent Schwartz which has rigorous properties that resemble the formal properties (i) through (v) summarized above.

11.2.2 Density of a Mass Distribution

> The Dirac delta function allows us to handle continuous and discrete mass distributions on equal footing by using mass density functions.
>
> Folklore

Continuous mass distribution. Let $\varrho : \mathbb{R} \to \mathbb{R}$ be a continuous nonnegative function. Physically, this function can be regarded as the mass density of a mass distribution on the real line. The number

$$m = \int_a^b \varrho(x)dx$$

then represents the mass on the finite interval $[a, b]$.

Discrete mass distribution. Consider now a point of mass $m > 0$ at the coordinate x_0 on the real line. Formally, we assign the mass density function

$$\varrho(x) := m \cdot \delta(x - x_0) \qquad \text{for all} \quad x \in \mathbb{R}$$

to the mass point. To motivate this formal convention, note that $\varrho(x) = 0$ for all points x on the real line different from x_0. This means that there is no mass at the points different from x_0. Moreover,

$$\int_{\mathbb{R}} \varrho(x)dx = m,$$

that is, the total mass equals m.

Approximation of a mass point. For fixed $\Delta x > 0$, choose the mass density

$$\varrho_{\Delta x}(x) := m \cdot \delta_{\Delta x}(x - x_0) \qquad \text{for all} \quad x \in \mathbb{R}.$$

Then, we have $\varrho_{\Delta x}(x) = 0$ for all points x on the real line outside the small interval $[x_0, x_0 + \Delta x]$. Furthermore,

$$\int_{\mathbb{R}} \varrho_{\Delta x}(x)dx = \frac{1}{\Delta x} \int_{x_0}^{x_0+\Delta x} m dx = m.$$

Letting $\Delta x \to 0$, the interval $[x_0, x_0 + \Delta x]$ shrinks to the point x_0. Formally, we write

$$\lim_{\Delta x \to +0} \varrho_{\Delta x}(x) = m \cdot \delta(x - x_0) \qquad \text{for all} \quad x \in \mathbb{R}.$$

11.2.3 Local Functional Derivative

> Local functional derivatives are frequently used in quantum field theory. They generalize classical partial derivatives to an infinite number of variables.
>
> Folklore

The Dirac delta function as a generalized Kronecker symbol. Recall that $\delta_{kk} := 1$ and $\delta_{kl} = 0$ if $k \neq l$. For a given finite interval $[a, b]$ on the real line, consider the partition points

$$x_k := a + k\Delta x, \qquad k = 0, 1, \ldots N$$

with $\Delta x := (b - a)/N$. Then

$$\sum_{j=1}^{N} f(x_j)\frac{\delta_{jk}}{\Delta x}\cdot \Delta x = f(x_k), \qquad k = 1, \dots, N.$$

The formal limit yields

$$\int_a^b f(x)\delta(x-y)\cdot dx = f(y), \qquad a < y < b.$$

Classical partial derivatives. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function of n real variables $x_1, \dots, x_n$. The partial derivative is defined by

$$\frac{\partial f(x^1, \dots, x^n)}{\partial x^k} = \frac{d}{dt} f(x^1 + t\delta_{1k}, \dots, x^n + t\delta_{kn})_{|t=0}.$$

Therefore

$$\frac{\partial x^j}{\partial x^k} = \delta_{jk}, \qquad j, k = 1, \dots, n. \tag{11.20}$$

This corresponds to $f(x^1, \dots, x^n) := x^j$ for all $x^1, \dots, x^n \in \mathbb{R}$.

Local functional derivative. By a real functional on the real line, we mean a map $J \mapsto Z(J)$ which assigns a real number $Z(J)$ to each function $J : \mathbb{R} \to \mathbb{R}$. If necessary, we restrict ourselves to continuous functions J, smooth functions, integrable functions, and so on. Replacing the Kronecker symbol by the Dirac delta function, we formally define the functional derivative by setting

$$\boxed{\frac{\delta Z(J)}{\delta J(y)} := \frac{d}{dt} Z(J + t\delta_y)_{|t=0}} \tag{11.21}$$

where $\delta_y(x) := \delta(x - y)$. Parallel to (11.20), we get

$$\frac{\delta J(x)}{\delta J(y)} = \delta(x - y), \qquad x, y \in \mathbb{R}.$$

This corresponds to (11.21) by choosing the functional $Z(J) := J(x)$ for all functions $J : \mathbb{R} \to \mathbb{R}$ and fixed $x \in \mathbb{R}$. Let us consider some examples.

(i) Consider the functional

$$Z(J) := \int_{\mathbb{R}} \varrho(x) J(x) dx$$

for all continuous functions $J : \mathbb{R} \to \mathbb{R}$ and fixed continuous function $\varrho : \mathbb{R} \to \mathbb{R}$. Formally differentiating the expression

$$Z(J + t\delta_y) = \int_{\mathbb{R}} \varrho(x)(J(x) + t\delta(x - y))\, dx$$

with respect to the variable t at the point $t = 0$, we get

$$\frac{\delta Z(J)}{\delta J(y)} = \int_{\mathbb{R}} \varrho(x)\delta(x - y)\, dx = \varrho(y).$$

(ii) Let $\varrho(x,y) = \varrho(y,x)$ for all $x, y \in \mathbb{R}$. For the functional

$$Z(J) := \tfrac{1}{2} \int_{\mathbb{R}^2} \varrho(x,y) J(x) J(y)\, dxdy$$

we obtain

$$\frac{\delta^2 Z(J)}{\delta J(y_0) \delta J(x_0)} = \varrho(x_0, y_0) \qquad \text{for all} \quad x_0, y_0 \in \mathbb{R}. \tag{11.22}$$

In fact, formally differentiating the function

$$Z(J + t\delta_{x_0}) = \tfrac{1}{2} \int_{\mathbb{R}^2} \varrho(x,y)(J(x) + t\delta(x - x_0))(J(y) + t\delta(y - x_0))\, dxdy$$

with respect to t at the point $t = 0$, we get that the local functional derivative

$$\frac{\delta Z(J)}{\delta J(x_0)}$$

is equal to

$$\begin{aligned} &\tfrac{1}{2} \int_{\mathbb{R}^2} \varrho(x,y)\delta(x - x_0) J(y)\, dxdy + \tfrac{1}{2} \int_{\mathbb{R}^2} \varrho(x,y) J(x)\delta(y - x_0)\, dxdy \\ &= \tfrac{1}{2} \int_{\mathbb{R}} \varrho(x_0, y) J(y) dy + \tfrac{1}{2} \int_{\mathbb{R}} \varrho(x, x_0) J(x) dx = \int_{\mathbb{R}} \varrho(x_0, y) J(y) dy. \end{aligned}$$

Furthermore, formally differentiating the function

$$\frac{\delta Z(J + t\delta_{y_0})}{\delta J(x_0)} = \int_{\mathbb{R}} \varrho(x_0, y)(J(y) + t\delta(y - y_0))\, dy$$

with respect to t at $t = 0$, we get $\int_{\mathbb{R}} \varrho(x_0, y)\delta(y - y_0) dy = \varrho(x_0, y_0)$. This is the claim (11.22).

(iii) Let $-\infty < t_1 < \tau < t_2 < \infty$. For the functional

$$S[q] := \int_{t_1}^{t_2} \{\tfrac{1}{2}\dot{q}(t)^2 - \tfrac{1}{2} q(t)^2 + q(t) F(t)\}\, dt,$$

we obtain

$$\frac{\delta S[q]}{\delta q(\tau)} = -\ddot{q}(\tau) - q(\tau) + F(\tau).$$

In fact, formally differentiating the function

$$\begin{aligned} S[q + s\delta_\tau)] = \int_{t_1}^{t_2} &\{\tfrac{1}{2}(\dot{q}(t) + s\dot{\delta}(t - \tau))^2 \\ &-\tfrac{1}{2}(q(t) + s\delta(t - \tau))^2 + (q(t) + \delta(t - \tau)) F(t)\}\, dt \end{aligned}$$

with respect to the variable s at the point $s = 0$, we get

$$\frac{\delta S[q]}{\delta q(\tau)} = \int_{t_1}^{t_2} \{\dot{q}(t)\dot{\delta}(t - \tau) - q(t)\delta(t - \tau) + F(t)\delta(t - \tau)\}\, dt.$$

Finally, integration by parts yields

$$\frac{\delta S[q]}{\delta q(\tau)} = \int_{t_1}^{t_2} (-\ddot{q}(t) - q(t) + F(t))\delta(t - \tau)\, dt = -\ddot{q}(\tau) - q(\tau) + F(\tau).$$

The relation to functional derivatives used in mathematics. Let us consider a typical example which arises in the calculus of variations. To this end, introduce the space $C_b^\infty[t_1,t_2]$ of smooth functions $q:[t_1,t_2]\to\mathbb{R}$ which satisfy the boundary condition $q(t_1)=q(t_2)=0$. Define the functional $S:C_b^\infty[t_1,t_2]\to\mathbb{R}$ by setting

$$S[q]:=\int_{t_1}^{t_2}\{\tfrac{1}{2}\dot{q}(t)^2-\tfrac{1}{2}q(t)^2+q(t)F(t)\}\,dt.$$

Here, the continuous function $F:[t_1,t_2]\to\mathbb{R}$ is fixed. For given functions $q,h\in C_b^\infty[t_1,t_2]$, the functional derivative of the functional S at the point q in direction of h is defined by

$$\boxed{\frac{\delta S[q]}{\delta q}(h)=\frac{d}{ds}S[q+sh]_{|s=0}.}$$

Explicitly, differentiating the function

$$S[q+sh]=\int_{t_1}^{t_2}\{(\dot{q}(t)+s\dot{h}(t))^2-(q(t)+sh(t))^2+(q(t)+h(t))F(t)\}\,dt$$

of the real variable s at the point $s=0$, we obtain

$$\frac{\delta S[q]}{\delta q}(h)=\int_{t_1}^{t_2}(\dot{q}(t)\dot{h}(t)-q(t)h(t)+h(t)F(t))\,dt.$$

Integration by parts yields

$$\boxed{\frac{\delta S[q]}{\delta q}(h)=\int_{t_1}^{t_2}(-\ddot{q}(t)-q(t)+F(t))h(t)\,dt.}$$

Defining

$$\frac{\delta S[q]}{\delta q(t)}:=-\ddot{q}(t)-q(t)+F(t),$$

we obtain

$$\frac{\delta S[q]}{\delta q}(h)=\int_{t_1}^{t_2}\frac{\delta S[q]}{\delta q(t)}\cdot h(t)\,dt. \tag{11.23}$$

Naturally enough, the function

$$t\mapsto\frac{\delta S[q]}{\delta q(t)}$$

is called the density function of the functional derivative

$$\frac{\delta S[q]}{\delta q}:C_b^\infty[t_1,t_2]\to\mathbb{R}.$$

Note that the density function coincides with the local functional derivative introduced above in a formal way. Setting $\delta q:=h$, let us write the key formula (11.23) as

$$\boxed{\delta S=\int_{t_1}^{t_2}\frac{\delta S[q]}{\delta q(t)}\cdot\delta q(t)\,dt.}$$

This formula resembles the classical formula

$$df = \sum_{k=1}^{n} \frac{\partial f(x)}{\partial x^k} \cdot dx^k$$

for the function $f : \mathbb{R}^n \to \mathbb{R}$ with $x = (x^1, \ldots, x^n)$.

Application to the calculus of variations. Consider the problem

$$S[q] = \text{critical!}, \qquad q \in C_b^\infty[t_1, t_2]. \tag{11.24}$$

Here, the function q is a solution iff

$$\frac{\delta S[q]}{\delta q}(h) = 0 \qquad \text{for all} \quad h \in C_b^\infty[t_1, t_2].$$

By the variational lemma on page 542, this is equivalent to

$$\frac{\delta S[q]}{\delta q(t)} = 0 \qquad \text{for all} \quad t \in [t_1, t_2].$$

Explicitly, this means that $\ddot{q}(t) + q(t) = F(t)$ for all $t \in [t_1, t_2]$. This is the Euler–Lagrange equation to the variational problem (11.24).

11.2.4 The Substitution Rule

In quantum field theory, physicists use frequently the function $x \mapsto \delta(f(x))$. We define

$$\boxed{\delta(f(x)) := \sum_{j=1}^{N} \frac{\delta(x - x_j)}{|f'(x_j)|} \qquad \text{for all} \quad x \in \mathbb{R}.} \tag{11.25}$$

Here, we assume that the function $f : \mathbb{R} \to \mathbb{R}$ is smooth, and it has precisely the zeros $x_1, \ldots, x_N$. In addition, suppose that the zeros are non-degenerate, that is, $f'(x_j) \neq 0$ for all $j = 1, \ldots, N$. For example, if a is a nonzero real number and $x_0 \in \mathbb{R}$, then

$$\delta(a(x - x_0)) := \frac{\delta(x - x_0)}{|a|} \qquad \text{for all} \quad x \in \mathbb{R}.$$

Furthermore, if $a > 0$, then

$$\boxed{\delta(x^2 - a^2) = \frac{\delta(x - a)}{2a} + \frac{\delta(x + a)}{2a} \qquad \text{for all} \quad x \in \mathbb{R}.}$$

This implies

$$\int_{\mathbb{R}} \delta(x^2 - a^2)\varphi(x)dx = \frac{\varphi(a)}{2a} + \frac{\varphi(-a)}{2a} \tag{11.26}$$

for all continuous functions $\varphi : \mathbb{R} \to \mathbb{R}$. In turn, setting $x := (\mathbf{x}, t)$, formula (11.26) yields

$$\int_{\mathbb{R}^4} \delta(t^2 - \mathbf{x}^2)\varphi(x)d^4x = \sum_{\sigma=\pm 1} \int_{\mathbb{R}^3} \frac{\varphi(\mathbf{x}, \sigma||\mathbf{x}||)}{2||\mathbf{x}||}\, d^3x, \tag{11.27}$$

for all continuous functions $\varphi : \mathbb{R}^4 \to \mathbb{R}$ which vanish outside some ball. In order to get formula (11.27), one has to first integrate over time t.

Let us now motivate the definition (11.25). Suppose first that the smooth function $f : \mathbb{R} \to \mathbb{R}$ has the only zero x_1. In addition, let $f'(x_1) > 0$. Set

$$J := \int_{x_1-\Delta x}^{x_1+\Delta x} \delta(f(x))\varphi(x)dx.$$

For sufficiently small $\Delta x > 0$, the function f is monotone increasing on the interval $[x_1 - \Delta x, x_1 + \Delta x]$. Using the substitution $y = f(x)$, we get

$$J = \int_{f(x_1-\Delta x)}^{f(x_1+\Delta x)} \delta(y)\varphi(x(y))\frac{dx(y)}{dy}\, dy.$$

Since $f(x_1) = 0$,

$$J = \varphi(x(0))\frac{dx(0)}{dy} = \frac{\varphi(x_1)}{f'(x_1)}.$$

If $f'(x_1) < 0$, then we get an additional minus sign. Finally, the case of several zeros can be reduced to the case of one zero, by localization.

11.2.5 Formal Dirac Calculus and the Fourier Transform

> It is worth noting that notation facilitates discovery. This, in a most wonderful way, reduces the mind's labor.
>
> Gottfried Wilhelm Leibniz (1646–1716)

We want to show that the Dirac calculus represents the most elegant mnemonic method in order to memorize the Fourier transform and, more general, all kinds of eigenfunction expansions appearing in mathematics and physics. In this section, we are going to generalize the rigorous finite-dimensional Dirac calculus from Sect. 7.11 on page 359 to *infinite* dimensions in a formal, but very useful way. In 1955 Gelfand and Kostyuchenko published a rigorous justification of the general Dirac calculus in terms of an extension of von Neumann's spectral theory to rigged Hilbert spaces; they used distributions as generalized eigenfunctions (see Sect. 12.2 on page 675).

States, costates, and generalized states. Let $x, p \in \mathbb{R}$. It turns out that for the complete description of a quantum particle on the real line, the notion of states in a Hilbert space is not sufficient. One also needs costates and generalized states which are related to the duality theory of Hilbert spaces in mathematics. Following Dirac, we introduce the following symbols:

- state $|\psi\rangle$ of a quantum particle on the real line;
- costate $\langle F|$ of a quantum particle on the real line;
- generalized state $|F\rangle$ which is dual to the costate $\langle F|$;
- position costate $\langle x|$;[12]
- generalized position state $|x\rangle$;
- momentum costate $\langle p|$;
- generalized momentum state $|p\rangle$;

[12] Intuitively, $\langle x|\psi\rangle$ is the value $\psi(x)$ of the Schrödinger wave function ψ at the point x.
The state $|x\rangle$ describes a quantum particle on the real line which is localized at the point x. This is a highly idealized situation.
The state $|p\rangle$ describes a quantum particle on the real line which has the sharp momentum p.

- to each state $|\psi\rangle$ we assign a uniquely determined costate which is denoted by $\langle\psi|$.

Below we will use the following notation: $|\varphi\rangle, |\psi\rangle, |\chi\rangle$ are states, $\langle F|, \langle G|$ are costates, and α, β are complex numbers. We first postulate that the linear combinations

$$\alpha|\varphi\rangle + \beta|\psi\rangle \quad \text{and} \quad \alpha\langle F| + \beta\langle G|$$

and

$$\alpha|F\rangle + \beta|G\rangle$$

make sense, that is, states, costates, and generalized states form a complex linear space, respectively. Secondly, we postulate that the following products are defined:

- $\langle F| \cdot |\psi\rangle$ (dual pairing);
- $|\psi\rangle\langle F|$ (tensor product);
- $\langle\psi| \cdot |F\rangle$ (antidual pairing).

Here, $\langle F| \cdot |\psi\rangle$ is a complex number which relates the abstract quantities $\langle F|$ and $|\psi\rangle$ to physical measurement processes. We assume that the following properties hold true:

- Distributivity: $\langle F| \cdot (\alpha|\varphi\rangle + \beta|\psi\rangle) = \alpha\langle F| \cdot |\varphi\rangle + \beta\langle F| \cdot |\psi\rangle$, and

$$(\alpha\langle F| + \beta\langle G|) \cdot |\psi\rangle = \alpha\langle F| \cdot |\varphi\rangle + \beta\langle G| \cdot |\psi\rangle.$$

- Anticommutativity: $\langle\psi| \cdot |\varphi\rangle = (\langle\varphi| \cdot \psi\rangle)^\dagger$ and $\langle\psi| \cdot |F\rangle = (\langle F| \cdot |\psi\rangle)^\dagger$.
- Associativity: $(|\varphi\rangle \cdot \langle F|) \cdot |\psi\rangle := |\varphi\rangle \cdot (\langle F| \cdot |\psi\rangle)$.

To simplify notation, we write

$$\boxed{\langle F|\psi\rangle := \langle F| \cdot |\psi\rangle,}$$

and $\langle\psi|F\rangle := \langle\psi| \cdot |F\rangle$, as well as $|\alpha\varphi + \beta\psi\rangle := \alpha|\varphi\rangle + \beta|\psi\rangle$. This yields the following formulas:

(a) $\langle F|\alpha\varphi + \beta\psi\rangle = \alpha\langle F|\varphi\rangle + \beta\langle F|\psi\rangle$;
(b) $\langle\psi|\varphi\rangle = \langle\varphi|\psi\rangle^\dagger$;
(c) $\langle\psi|F\rangle = \langle F|\psi\rangle^\dagger$;
(d) $|\varphi\rangle\langle F| \cdot |\psi\rangle = |\varphi\rangle\langle F|\psi\rangle = \langle F|\psi\rangle\ |\varphi\rangle$.
(e) $\langle\alpha\varphi + \beta\psi|\chi\rangle = \alpha^\dagger\langle\varphi|\chi\rangle + \beta^\dagger\langle\psi|\chi\rangle$;
(f) $\langle\alpha\varphi + \beta\psi|F\rangle = \alpha^\dagger\langle\varphi|F\rangle + \beta^\dagger\langle\psi|F\rangle$.

In fact, $\langle\alpha\varphi + \beta\psi|\chi\rangle$ is equal to

$$\langle\chi|\alpha\varphi + \beta\psi\rangle^\dagger = (\alpha\langle\chi|\varphi\rangle + \beta\langle\chi|\psi\rangle)^\dagger = \alpha^\dagger\langle\varphi|\chi\rangle + \beta^\dagger\langle\psi|\chi\rangle.$$

Similarly, we argue for motivating (f). In terms of mathematics, the following hold true.

- By (a), costates $\langle F|$ are linear functionals, and $\langle F|\psi\rangle$ is the value $F(\psi)$ of the functional F at the point ψ.
- By (f), generalized states $|F\rangle$ are antilinear functionals where $\langle\psi|F\rangle$ is the value $F(\psi)$ of the functional at the point ψ.
- By (d), the product $|\varphi\rangle \cdot \langle F|$ corresponds to the tensor product $\varphi \otimes F$.

The key relations. In the late 1920's, Dirac noticed that the classical Fourier transform can be elegantly reformulated in terms of the following key relations for states and costates. Let $x, y, p, q \in \mathbb{R}$.

(K1) Dual pairing between states and position costates:

$$\langle x|\psi\rangle = \psi(x).$$

In terms of physics, the complex-valued function $x \mapsto \psi(x)$ is the famous Schrödinger wave function of the quantum particle on the real line. Furthermore,

$$\langle \psi|x\rangle = \langle x|\psi\rangle^\dagger = \psi(x)^\dagger.$$

(K2) Dual pairing between position costates and generalized momentum states:

$$\langle x|p\rangle = \frac{\mathrm{e}^{\mathrm{i}px/\hbar}}{\sqrt{2\pi\hbar}}.$$

The function $x \mapsto \langle x|p\rangle$ represents the Schrödinger wave function of a particle on the real line which has the momentum p.[13]

(K3) Orthogonality relation for position states: $\langle x|y\rangle = \delta(x-y)$.

(K4) Completeness relation for position states:

$$\boxed{\int_{\mathbb{R}} |x\rangle\langle x|\, dx = I}$$

where I denotes the identity operator.

(K5) Orthogonality relation for momentum states: $\langle p|q\rangle = \delta(p-q)$.

(K6) Completeness relation for momentum states:

$$\boxed{\int_{\mathbb{R}} |p\rangle\langle p|\, dp = I.}$$

(K7) Matrix elements of the position operator X: For all $x, y \in X$,

$$\langle x|X|y\rangle := x\delta(x-y).$$

This implies $\langle x|X|y\rangle^\dagger = \langle y|X|x\rangle$.

(K8) Matrix elements of the momentum operator P: For all $p, q \in \mathbb{R}$,

$$\langle p|P|q\rangle := p\delta(p-q).$$

Hence $\langle p|P|q\rangle^\dagger = \langle q|P|p\rangle$.

The Fourier transform. Let us now show that (K1) through (K8) summarize all of the properties of both the Fourier transform and the Dirac delta function.

(i) Inner product: For all states $|\psi\rangle$ and $|\varphi\rangle$,

$$\langle \varphi|\psi\rangle = \int_{\mathbb{R}} \varphi(x)^\dagger \psi(x) dx. \tag{11.28}$$

To get this in a formal way, we start with the trivial identity

$$\langle \varphi|\psi\rangle = \langle \varphi|I|\psi\rangle.$$

[13] Note that $\int_{\mathbb{R}} |\mathrm{e}^{\mathrm{i}px/\hbar}|^2 dx = \infty$. Therefore, the function $x \mapsto \langle x|p\rangle$ does not lie in the Hilbert space $L_2(\mathbb{R})$.

We now apply Dirac's substitution trick. Explicitly, we replace the identity operator I by the completeness relation (K4) for the position operator. Thus,

$$\boxed{\langle\varphi|\psi\rangle = \int_{\mathbb{R}} \langle\varphi|x\rangle\langle x|\psi\rangle \, dx.}$$

This is identical with (11.28).

(ii) Fourier transform: By Dirac's substitution trick,

$$\boxed{\langle p|\psi\rangle = \int_{\mathbb{R}} \langle p|x\rangle\langle x|\psi\rangle \, dx.}$$

Setting $\hat{\psi}(p) := \langle p|x\rangle$, we obtain

$$\hat{\psi}(p) = \frac{1}{\sqrt{2\pi\hbar}} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}px/\hbar}\psi(x) \, dx.$$

This shows that $\hat{\psi}$ is the Fourier transform of ψ.

(iii) Inverse Fourier transform: By Dirac's substitution trick and the completeness relation (K6) for the momentum states,

$$\boxed{\langle x|\psi\rangle = \int_{\mathbb{R}} \langle x|p\rangle\langle p|\psi\rangle \, dp.}$$

Explicitly, this is the inverse Fourier transform,

$$\psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}px/\hbar}\hat{\psi}(p) \, dp.$$

(iv) The Parseval equation: By Dirac's substitution trick,

$$\boxed{\langle\varphi|\psi\rangle = \int_{\mathbb{R}} \langle\varphi|x\rangle\langle x|\psi\rangle dx = \int_{\mathbb{R}} \langle\varphi|p\rangle\langle p|\psi\rangle dp.}$$

Hence

$$\int_{\mathbb{R}} \varphi(x)^\dagger\psi(x) \, dx = \int_{\mathbb{R}} \hat{\varphi}(p)^\dagger\hat{\psi}(p) \, dp.$$

This is the Parseval equation for the Fourier transform.

(v) The Dirac delta function in position space: By Dirac's substitution trick,

$$\boxed{\langle x|\psi\rangle = \int_{\mathbb{R}} \langle x|y\rangle\langle y|\psi\rangle \, dy.}$$

This means that

$$\psi(x) = \int_{\mathbb{R}} \delta(x-y)\psi(y)dy$$

which is the characteristic property of the Dirac delta function. Using the orthogonality relation for position states and Dirac's substitution trick with respect to the completeness relation for momentum states, we obtain

$$\boxed{\langle x|y\rangle = \int_{\mathbb{R}} \langle x|p\rangle\langle p|y\rangle \, dp.}$$

Hence

$$\delta(x-y) = \frac{1}{2\pi\hbar}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}p(x-y)/\hbar} \, dp. \tag{11.29}$$

This is the representation of the Dirac delta function by a Fourier integral.

(vi) The Dirac delta function in momentum space: Using the orthogonality relation for momentum states and Dirac's substitution trick with respect to the completeness relation for position states, we obtain

$$\boxed{\langle p|q\rangle = \int_{\mathbb{R}} \langle p|x\rangle\langle x|q\rangle \, dx.}$$

Hence

$$\delta(p-q) = \frac{1}{2\pi\hbar}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}x(q-p)/\hbar} \, dx.$$

Noting the invariance of the integral under the reflection $x \mapsto -x$, this formula is equivalent to (11.29).

(vii) The position operator X in position space: For each state $|\psi\rangle$ and all $x \in \mathbb{R}$,

$$\boxed{\langle x|X|\psi\rangle = x\langle x|\psi\rangle.} \tag{11.30}$$

We say that $\langle x|$ is an eigencostate of the position operator X. In fact, by Dirac's substitution trick,

$$\langle x|X|\psi\rangle = \int_{\mathbb{R}} \langle x|X|y\rangle\langle y|\psi\rangle \, dy. \tag{11.31}$$

Hence

$$\langle x|X|\psi\rangle = \int_{\mathbb{R}} x\delta(x-y)\psi(y)dy = x\psi(x) = x\langle x|\psi\rangle.$$

Equation (11.30) can be written as

$$\boxed{(X\psi)(x) = x\psi(x) \qquad \text{for all} \quad x \in \mathbb{R}.}$$

This means that the position operator X corresponds to the multiplication operator with respect to Schrödinger's wave function ψ of the quantum particle in the position space. Introducing the kernel

$$\mathcal{K}_X(x,y) := \langle x|X|y\rangle, \qquad x, y \in \mathbb{R}$$

of the position operator X, equation (11.31) corresponds to the so-called kernel equation

$$\boxed{(X\psi)(x) = \int_{\mathbb{R}} \mathcal{K}_X(x,y)\psi(y)dy \qquad \text{for all} \quad x \in \mathbb{R}}$$

with $\mathcal{K}_X(x,y) = x\delta(x-y)$. By definition, the real number

$$\overline{x} = \frac{\langle\psi|X|\psi\rangle}{\langle\psi|\psi\rangle} = \frac{\int_{\mathbb{R}} \psi(x)^\dagger x\psi(x)dx}{\int_{\mathbb{R}} \psi(x)^\dagger\varphi(x)dx}$$

is the mean position of the quantum particle in the nonzero state $|\psi\rangle$ on the real line. Finally, let us motivate the equality

$$\boxed{X|x\rangle = x|x\rangle \qquad \text{for all} \quad x \in \mathbb{R}.}$$

This means that $|x\rangle$ is a generalized eigenstate of the position operator X with the corresponding real eigenvalue x. In fact, by Dirac's substitution trick,

$$X|x\rangle = \int_{\mathbb{R}} |y\rangle\langle y|X|x\rangle \, dy = \int_{\mathbb{R}} |y\rangle \, y\delta(y-x) \, dy = x|x\rangle.$$

(viii) The momentum operator P in momentum space: The operator P plays the same role in the momentum space as the position operator in the position space. Replacing $X, |x\rangle, \langle x|$ by $P, |p\rangle, \langle p|$, respectively, we obtain that for all states ψ and all $p \in \mathbb{R}$,

$$\boxed{\langle p|P\psi\rangle = p \, \langle p|\psi\rangle} \tag{11.32}$$

and

$$\boxed{P|p\rangle = p|p\rangle.} \tag{11.33}$$

Equation (11.32) tells us that

$$\mathcal{F}(P\psi)(p) = p\mathcal{F}(\psi)(p) \qquad \text{for all} \quad p \in \mathbb{R}.$$

This means that the momentum operator P represents the multiplication operator after carrying out the Fourier transform. Instead of (11.33) and (11.32), we say that $\langle p|$ and $|p\rangle$ are an eigencostate and a generalized eigenstate, respectively, of the momentum operator P with the real eigenvalue p.

(ix) The momentum operator P in position space: It follows from

$$\psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}px/\hbar} \hat{\psi}(p)dp$$

that

$$-\mathrm{i}\hbar\frac{d}{dx}\psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}px/\hbar} p\hat{\psi}(p)dp. \tag{11.34}$$

Consequently, the multiplication operator $\hat{\psi}(p) \mapsto p\hat{\psi}(p)$ in the Fourier space corresponds to the operator $\psi(x) \mapsto -\mathrm{i}\hbar\psi'(x)$ in the position space. Explicitly,

$$(P\psi)(x) = \left(-\mathrm{i}\hbar\frac{d}{dx}\psi\right)(x) \qquad \text{for all} \quad x \in \mathbb{R}.$$

In the language of the Dirac calculus, we have

$$\boxed{\langle x|P\psi\rangle = -\mathrm{i}\hbar\frac{d}{dx}\langle x|\psi\rangle.}$$

This follows from (11.32) and (11.34) along with

$$\begin{aligned}\langle x|P\psi\rangle &= \int_{\mathbb{R}} \langle x|p\rangle\langle p|P|\psi\rangle\, dp = \int_{\mathbb{R}} \langle x|p\rangle p\langle p|\psi\rangle\, dp \\ &= -\mathrm{i}\hbar\frac{d}{dx}\psi(x) = -\mathrm{i}\hbar\frac{d}{dx}\langle x|\psi\rangle.\end{aligned}$$

The energy operator. In terms of the formal Dirac calculus, we now want to study the energy of a quantum particle on the real line. To this end, we postulate that there exists a measure μ on the real line along with the measure integral

$$\int_{\mathbb{R}} f(E)d\mu(E).$$

By definition, the zero set, $\mathrm{zero}(\mu)$, of the measure μ is the largest open subset of the real line $\mathbb{R}$ such that the measure vanishes on this set. By definition, the complement

$$\mathrm{supp}(\mu) := \mathbb{R} \setminus \mathrm{zero}(\mu)$$

is called the support of the measure μ. For a reasonable subset S of the real line, the number

$$\mu(S) = \int_S d\mu(E)$$

represents the measure of the set S. Moreover, we introduce the formal Dirac delta function δ_μ with respect to the measure μ by postulating the characteristic property

$$\int_{\mathbb{R}} \delta_\mu(E, E_0)f(E)d\mu(E) = f(E_0) \qquad \text{for all} \quad E_0 \in \mathbb{R}.$$

Furthermore, we postulate that

- $\delta_\mu(E, E_0) = 0$ if $E \neq E_0$.
- $\delta_\mu(E, E) = 0$ if $E \notin \mathrm{supp}(\mu)$.
- $\delta_\mu(E, E)^\dagger = \delta_\mu(E, E)$.

To illustrate the notion of measure, let us consider two typical examples.

- Discrete mass distribution: Consider n points $E_1, \ldots, E_n$ on the real line which have the positive masses $m_1, \ldots, m_n$, respectively. Then

$$\int_{\mathbb{R}} f(E)d\mu(E) = \sum_{k=1}^{n} f(E_k)m_k.$$

  The support of the measure μ is equal to the finite set $\{E_1, \ldots, E_n\}$.
- Continuous mass distribution: Let $\varrho : \mathbb{R} \to \mathbb{R}$ be a nonnegative continuous function. In terms of physics, the function ϱ represents a mass density on the real line. Then

$$\int_{\mathbb{R}} f(E)d\mu(E) = \int_{\mathbb{R}} f(E)\varrho(E)dE$$

  if the right-hand integral exists. The set $\mathrm{zero}(\varrho)$ is the largest open subset of $\mathbb{R}$ on which the density function ϱ vanishes.

Following Dirac, we introduce the following symbols for all $E \in \mathbb{R}$:

- the energetic costate $\langle E|$;
- the generalized energetic state $|E\rangle$.[14]

[14] Intuitively, $|E\rangle$ corresponds to a quantum particle on the real line which has the energy E. If the function $x \mapsto \langle x|E\rangle$ lies in the Hilbert space $L_2(\mathbb{R})$, then $|E\rangle$ is called a classical quantum state. This is not always the case. Therefore, we speak of the generalized energy state $|E\rangle$.

We also set $|E\rangle = 0$ and $\langle E| = 0$ if $E \notin \mathrm{supp}(\mu)$. Furthermore, we postulate that the following key relations hold true for all $E, E' \in \mathbb{R}$:

(E1) Orthogonality: $\langle E'|E\rangle = \delta_\mu(E', E)$.
(E2) Completeness:

$$\boxed{\int_{\mathbb{R}} |E\rangle\langle E|\, d\mu(E) = I.}$$

(E3) Matrix elements of the energy operator H:

$$\langle E'|H|E\rangle = E\delta_\mu(E', E).$$

Furthermore, $\langle E'|H|E\rangle^\dagger = \langle E|H|E'\rangle$.

Instead of the symbol $\mathrm{supp}(\mu)$ we synonymously write $\sigma(H)$, and we call this the spectrum of the energy operator (or Hamiltonian) H. Intuitively, the energy spectrum $\sigma(H)$ is the set of all possible energy values E of quantum particles on the real line which correspond to the physical situation described by the operator H (e.g., a fixed potential). Let us discuss some formal consequences of the conditions (E1) through (E3).

(i) Energetic Fourier transform: By the completeness condition for position states,

$$\boxed{\langle E|\psi\rangle = \int_{\mathbb{R}} \langle E|x\rangle\langle x|\psi\rangle\, dx.}$$

Setting $\hat{\psi}(E) := \langle E|\psi\rangle$ and $\chi_E(x) := \langle x|E\rangle$, this reads as

$$\hat{\psi}(E) = \int_{\mathbb{R}} \chi_E(x)^\dagger \psi(x) dx \qquad \text{for all} \quad E \in \mathbb{R}.$$

The function $E \mapsto \hat{\psi}(E)$ is called the energetic Fourier transform of the Schrödinger wave function $x \mapsto \psi(x)$ in the position space.

(ii) The inverse energetic Fourier transform: By the energetic completeness relation (E2),

$$\langle x|\psi\rangle = \int_{\mathbb{R}} \langle x|E\rangle\langle E|\psi\rangle\, d\mu(E).$$

This means that

$$\psi(x) = \int_{\mathbb{R}} \chi_E(x)\hat{\psi}(E) d\mu(E) \qquad \text{for all} \quad x \in \mathbb{R}.$$

(iii) The Parseval equation for the inner product: By the energetic completeness relation (E2),

$$\boxed{\langle \varphi|\psi\rangle = \int_{\mathbb{R}} \langle \varphi|E\rangle\langle E|\psi\rangle\, d\mu(E).}$$

This tells us that for all states $|\varphi\rangle$ and $|\psi\rangle$,

$$\langle \varphi|\psi\rangle = \int_{\mathbb{R}} \varphi(x)^\dagger \psi(x) dx = \int_{\mathbb{R}} \hat{\varphi}(E)^\dagger \hat{\psi}(E) d\mu(E).$$

(iv) Energetic eigencostates: For all states ψ and all $E \in \mathbb{R}$,

$$\boxed{\langle E|H|\psi\rangle = E\langle E|\psi\rangle.}$$

We say that $\langle E|$ is an eigencostate of the energy operator H to the energy value E. In fact, by the completeness relation (E2),

$$\begin{aligned}\langle E|H|\psi\rangle &= \int_{\mathbb{R}} \langle E|H|E'\rangle\langle E'|\psi\rangle \, d\mu(E') \\ &= \int_{\mathbb{R}} E\delta_\mu(E,E')\langle E'|\psi\rangle \, d\mu(E') = E\langle E|\psi\rangle.\end{aligned}$$

(v) Generalized energetic eigenstates: For all $E \in \mathbb{R}$, we get

$$\boxed{H|E\rangle = E|E\rangle.}$$

We say that $|E\rangle$ is a generalized energetic eigenstate with the eigenvalue E. To motivate this, note that by the completeness relation (E2),

$$H|E\rangle = \int_{\mathbb{R}} |E'\rangle\langle E'|H|E\rangle d\mu(E') = \int_{\mathbb{R}} E'\delta_\mu(E',E)|E'\rangle \, d\mu(E') = E|E\rangle.$$

(vi) Functions of the energy operator: Let $k = 1, 2, \ldots$. For all $E \in \mathbb{R}$, we have

$$H^k|E\rangle = E^k|E\rangle.$$

In fact, $H^2|E\rangle = H(H|E\rangle) = H(E|E\rangle) = E(H|E\rangle) = E^2|E\rangle$, and so on. This motivates the following definition. Let $f : \mathbb{R} \to \mathbb{C}$ be a polynomial or an analytic function. For all $E \in \mathbb{R}$, we set

$$\boxed{f(H)|E\rangle := f(E)|E\rangle.}$$

(iv) Transition amplitudes: For all states $|\varphi\rangle$ and $|\psi\rangle$,

$$\boxed{\langle\varphi|f(H)|\psi\rangle = \int_{\mathbb{R}} \langle\varphi|E\rangle f(E)\langle E|\psi\rangle d\mu(E).}$$

Equivalently,

$$\langle\varphi|f(H)|\psi\rangle = \int_R \hat{\varphi}(E)^\dagger f(E)\hat{\psi}(E)d\mu(E).$$

In fact, by the completeness relation (E2),

$$\begin{aligned}\langle\varphi|f(H)|\psi\rangle &= \int_{\mathbb{R}} d\mu(E') \int_{\mathbb{R}} d\mu(E)\langle\varphi|E'\rangle\langle E'|f(H)E\rangle\langle E|\psi\rangle \\ &= \int_{\mathbb{R}} d\mu(E') \int_{\mathbb{R}} d\mu(E) f(E)\langle\varphi|E'\rangle\langle E'|E\rangle\langle E|\psi\rangle \\ &= \int_{\mathbb{R}} d\mu(E) f(E)\langle\varphi|E\rangle\langle E|\psi\rangle.\end{aligned}$$

(v) Dynamics of a quantum particle: Let $|\psi(t)\rangle$ be the state of the quantum particle on the real line at the time t. We postulate that

$$\boxed{|\psi(t)\rangle = \mathrm{e}^{-\mathrm{i}tH/\hbar}|\psi(0)\rangle \qquad \text{for all} \quad t \in \mathbb{R}.} \tag{11.35}$$

Differentiation with respect to time t yields the Schrödinger equation

$$\mathrm{i}\hbar\frac{d}{dt}|\psi(t)\rangle = H|\psi(t)\rangle \qquad \text{for all} \quad t \in \mathbb{R}.$$

By the completeness relation in the position state, it follows from (11.35) that

$$\boxed{\langle x|\psi(t)\rangle = \int_{\mathbb{R}} \langle x|\mathrm{e}^{-\mathrm{i}tH/\hbar}|y\rangle\langle y|\psi(0)\rangle \, dy} \tag{11.36}$$

for all positions $x \in \mathbb{R}$ and all times $t \in \mathbb{R}$. Setting $\psi(x,t) := \langle x|\psi(t)\rangle$ and introducing the so-called propagator kernel,

$$\mathcal{P}(x,y;t) := \langle x|\mathrm{e}^{-\mathrm{i}tH/\hbar}|y\rangle,$$

equation (11.36) reads as

$$\psi(x,t) = \int_{\mathbb{R}} \mathcal{P}(x,y;t)\psi(y,0)dy \qquad \text{for all} \quad x,t \in \mathbb{R}.$$

Introducing the so-called retarded propagator kernel (or Feynman propagator kernel)

$$\mathcal{P}_+(x,y;t) := \theta(t)\mathcal{P}(x,y;t), \qquad x,y,t \in \mathbb{R},$$

then

$$\psi(x,t) = \int_{\mathbb{R}} \mathcal{P}_+(x,y;t)\psi(y,0)dy \qquad \text{for all} \quad x \in \mathbb{R}, \;\; t > 0.$$

Observables. Let $A \in \mathbb{R}$. If we replace $H, |E\rangle, \langle E|$ by $A, |A\rangle, \langle A|$, respectively, then we get the Dirac calculus with respect to the observable A. Special observables are energy H, position X, and momentum P.

In terms of this general approach, the spectral measure of both the position operator X and the momentum operator P is the Lebesgue measure on the real line. This measure is characterized by the fact that it is invariant under translations. In fact, the relation

$$\left(\frac{d}{dx}\psi\right)(x) = \lim_{\varepsilon \to +0} \frac{\psi(x+\varepsilon) - \psi(x)}{\varepsilon}$$

shows that the differential operator $\frac{d}{dx}$ represents the infinitesimal operator of the group of all translations, $x \mapsto x + \varepsilon$, on the real line. For the momentum operator, we obtain

$$P := -\mathrm{i}\hbar\frac{d}{dx}$$

on the position state. Therefore, it is quite natural that the translation-invariant Lebesgue measure represents the spectral measure of the momentum operator P.

11.2.6 Formal Construction of the Heat Kernel

We want to show how the Dirac calculus can be used in order to motivate the heat kernel formulas discussed on page 586. Set $\mathbf{P} := -\mathrm{i}\hbar\boldsymbol{\partial}$ with $\hbar = 1$. Hence

$$\mathbf{P}^2 = -\boldsymbol{\partial}^2 = \Delta.$$

Generalizing the one-dimensional Dirac calculus above, we will use the following key formulas:

- $\langle \mathbf{x}|\mathbf{y}\rangle = \delta^3(\mathbf{x}-\mathbf{y})$;
- $\langle \mathbf{x}|\mathbf{p}\rangle = \mathrm{e}^{\mathrm{i}\mathbf{p}\mathbf{x}}/(2\pi)^{\frac{3}{2}}$;
- $\int_{\mathbb{R}^3} |\mathbf{x}\rangle\langle\mathbf{x}|\, d^3x = I$;
- $\int_{\mathbb{R}^3} |\mathbf{p}\rangle\langle\mathbf{p}|\, d^3p = I$;
- $\mathbf{P}|\mathbf{p}\rangle = \mathbf{p}\,|\mathbf{p}\rangle$.

The homogeneous heat equation. Let us prove the solution formula (11.9) for the homogeneous heat equation (11.8) on page 586. In fact, the initial-value problem (11.8),

$$\dot{\psi}(t) = -\kappa\mathbf{P}^2\psi(t), \quad t > 0, \qquad \psi(0) = \psi_0$$

has the solution

$$\psi(t) = \mathrm{e}^{-\kappa t\mathbf{P}^2}\psi_0, \quad t > 0.$$

By the completeness relation $\int_{\mathbb{R}^3} |\mathbf{y}\rangle\langle\mathbf{y}|\, d^3y = I$,

$$\boxed{\langle \mathbf{x}|\psi(t)\rangle = \int_{\mathbb{R}^3} \langle \mathbf{x}|\mathrm{e}^{-\kappa t\mathbf{P}^2}|\mathbf{y}\rangle\langle\mathbf{y}|\psi_0\rangle\, d^3y.} \tag{11.37}$$

Since $\mathbf{P}|\mathbf{p}\rangle = \mathbf{p}|\mathbf{p}\rangle$, we get the crucial eigensolutions

$$\mathrm{e}^{-\kappa t\mathbf{P}^2}|\mathbf{p}\rangle = \mathrm{e}^{-\kappa t\mathbf{p}^2}|\mathbf{p}\rangle.$$

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ and $t > 0$. Introducing the heat kernel

$$\mathcal{K}(\mathbf{x}, t; \mathbf{y}, 0) := \langle \mathbf{x}|\mathrm{e}^{-\kappa t\mathbf{P}^2}|\mathbf{y}\rangle,$$

the completeness relation $\int_{\mathbb{R}^3} |\mathbf{p}\rangle\langle\mathbf{p}|\, d^3p = I$ yields

$$\begin{aligned}\mathcal{K}(\mathbf{x}, t; \mathbf{y}, 0) &= \int_{\mathbb{R}^3} \langle \mathbf{x}|\mathrm{e}^{-\kappa t\mathbf{P}^2}|\mathbf{p}\rangle\langle\mathbf{p}|\mathbf{y}\rangle\, d^3p \\ &= \int_{\mathbb{R}^3} \mathrm{e}^{-\kappa t\mathbf{p}^2}\langle \mathbf{x}|\mathbf{p}\rangle\langle\mathbf{p}|\mathbf{y}\rangle\, d^3p = \frac{1}{(2\pi)^3}\int_{\mathbb{R}^3} \mathrm{e}^{-\kappa t\mathbf{p}^2}\,\mathrm{e}^{\mathrm{i}(\mathbf{x}-\mathbf{y})\mathbf{p}}\, d^3p.\end{aligned}$$

Using the Gaussian integral (7.102) on page 433,

$$\mathcal{K}(\mathbf{x}, t; \mathbf{y}, 0) = \frac{\mathrm{e}^{-\frac{(\mathbf{x}-\mathbf{y})^2}{4\kappa t}}}{\left(4\pi\kappa t\right)^{\frac{3}{2}}}.$$

This is the heat kernel. From (11.37) we get the desired formula (11.9) on page 586.

The inhomogeneous heat equation and Duhamel's principle. The initial-value problem

$$\dot{\psi}(t) = -\kappa\mathbf{P}^2\psi(t) + f(t), \quad t > 0, \qquad \psi(0) = \psi_0$$

has the solution

$$\psi(t) = \mathrm{e}^{-\kappa t\mathbf{P}^2}\psi_0 + \int_0^t \mathrm{e}^{-\kappa(t-\tau)\mathbf{P}^2} f(\tau)d\tau, \qquad t > 0. \tag{11.38}$$

This formula is called Duhamel's principle.[15] The proof of (11.38) follows by formal differentiation parallel to (11.3). By the completeness relation,

$$\langle \mathbf{x}|\psi(t)\rangle = \int_{\mathbb{R}^3} \langle \mathbf{x}|\mathrm{e}^{-\kappa t\mathbf{P}^2}|\mathbf{y}\rangle\langle \mathbf{y}|\psi_0\rangle \, d^3y + \\ + \int_0^t \int_{\mathbb{R}^3} \langle \mathbf{x}|\mathrm{e}^{-\kappa(t-\tau)\mathbf{P}^2}|\mathbf{y}\rangle\langle \mathbf{y}|f(\tau)\rangle \, d^3y d\tau.$$

This is the desired solution formula (11.11) on page 587.

11.3 Laurent Schwartz's Rigorous Approach

11.3.1 Physical Measurements and the Idea of Averaging

Our measurement instruments measure only averages. The theory of distributions is based on different methods of averaging by using different types of test functions. For example, we can replace the mass density function $\varrho = \varrho(\mathbf{x}, t)$ by the average

$$\boxed{\varrho[\varphi] = \int_{\mathbb{R}^3} \varrho(\mathbf{x}, t)\varphi(\mathbf{x})d^3x} \tag{11.39}$$

where the test function $\varphi : \mathbb{R}^3 \to \mathbb{C}$ is smooth and vanishes outside some ball. The average $\varrho[\varphi]$ depends on the choice of the function φ. Dirac's relation

$$\int_{\mathbb{R}^3} \delta^3(\mathbf{x} - \mathbf{x}_0)\varphi(\mathbf{x})d^3x = \varphi(\mathbf{x}_0) \tag{11.40}$$

does not make any sense in classical analysis. However, we can define

$$\delta_{\mathbf{x_0}}(\varphi) := \varphi(\mathbf{x}_0)$$

for all test functions φ. The map $\varphi \mapsto \delta_{\mathbf{x}_0}(\varphi)$ describes a mass $m = 1$ at the point $\mathbf{x}_0$. Formula (11.40) serves as a very useful mnemonic tool.

> *The strategy of the theory of distributions is to reformulate classical properties in terms of averages of type* (11.39) *and then to generalize this to functionals* $\varphi \mapsto F(\varphi)$.

In quantum field theory, the method of averaging allows us to handle strong singularities in a reasonable way.

[15] Duhamel (1797–1872)

11.3.2 Distributions

Let $N = 1, 2, \ldots$. We will study complex-valued functions

$$f : \mathbb{R}^N \to \mathbb{C}$$

and generalizations of such functions called generalized functions or distributions. The real components of the point x in $\mathbb{R}^N$ are denoted by $x^1, \ldots, x^N$. We will use the notation introduced on page 536. In particular we write

$$\partial^\alpha := \partial_1^{\alpha_1} \partial_2^{\alpha_2} \cdots \partial_N^{\alpha_N}$$

along with the partial derivative $\partial_j := \partial/\partial x^j$ and $\alpha = (\alpha_1, \ldots, \alpha_N)$.

The support of a function. Let $f : \mathbb{R}^N \to \mathbb{C}$ be an arbitrary function. By definition, the zero set, $\text{zero}(f)$, is the maximal open subset of $\mathbb{R}^N$ on which the function f vanishes. The complement is called the support of f,

$$\boxed{\text{supp}(f) := \mathbb{R}^N \setminus \text{zero}(f).}$$

The support of a function is always a closed set. In particular, for the Heaviside function $\theta : \mathbb{R} \to \mathbb{R}$,

$$\text{supp}(\theta) = [0, \infty[.$$

The function, $f(x) := \sin x$ vanishes precisely at the points $x = 0, \pm\pi, \pm 2\pi, \ldots$. The support of this function is a closed set, and hence it equals the real line.

The space $\mathcal{D}(\mathbb{R}^N)$ of test functions. Let $N = 1, 2, \ldots$ By definition, the space $\mathcal{D}(\mathbb{R}^N)$ consists of all smooth functions

$$\varphi : \mathbb{R}^N \to \mathbb{C}$$

which have compact support.[16] Let $\varphi_n, \varphi \in \mathcal{D}(\mathbb{R}^N)$ for all n. We write

$$\boxed{\lim_{n\to\infty} \varphi_n = \varphi \quad \text{in } \mathcal{D}(\mathbb{R}^N)}$$

iff there exists a compact subset B of $\mathbb{R}^N$ such that

- $\text{supp}(\varphi_n) \subseteq B$ for all n;
- $\lim_{n\to\infty} \sup_{x\in B} |\varphi_n(x) - \varphi(x)| = 0$;
- $\lim_{n\to\infty} \sup_{x\in B} |\partial^\alpha \varphi_n(x) - \partial^\alpha \varphi(x)| = 0$ for all possible derivatives ∂^α.

The space $\mathcal{D}'(\mathbb{R}^N)$ of distributions. By definition, a distribution is a linear, sequentially continuous map

$$F : \mathcal{D}(\mathbb{R}^N) \to \mathbb{C}.$$

Explicitly, we assign a complex number $F(\varphi)$ to each test function φ in the space $\mathcal{D}(\mathbb{R}^N)$. Moreover, for all $\varphi, \psi \in \mathcal{D}(\mathbb{R}^N)$ and all complex numbers α, β,

$$F(\alpha\varphi + \beta\psi) = \alpha F(\varphi) + \beta F(\psi).$$

Finally, $\lim_{n\to\infty} \varphi_n = \varphi$ in $\mathcal{D}(\mathbb{R}^N)$ always implies

[16] Recall that a subset of $\mathbb{R}^N$ is compact iff it is bounded and closed (e.g., a closed ball). For example, the smooth function $\varphi : \mathbb{R} \to \mathbb{R}$ lies in $\mathcal{D}(\mathbb{R})$ iff it vanishes outside some finite interval.

$$\lim_{n\to\infty} F(\varphi_n) = F(\varphi).$$

The space of all such distributions is denoted by $\mathcal{D}'(\mathbb{R}^N)$.

The Dirac delta distribution. Define

$$\boxed{\delta[\varphi] := \varphi(0) \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}^N).}$$

Then, δ is a distribution living in $\mathcal{D}'(\mathbb{R}^N)$. In fact, for all $\varphi, \psi \in \mathcal{D}(\mathbb{R}^N)$ and all complex numbers α and β,

$$\delta[\alpha\varphi + \beta\psi] = \alpha\varphi(0) + \beta\varphi(0) = \alpha\delta[\varphi] + \beta\delta[\psi].$$

Moreover, $\lim_{n\to\infty} \varphi_n = \varphi$ in $\mathcal{D}(\mathbb{R}^N)$ implies $\lim_{n\to\infty} \varphi_n(0) = \varphi(0)$, and hence

$$\lim_{n\to\infty} \delta[\varphi_n] = \delta[\varphi].$$

Furthermore, for each fixed point $x_0 \in \mathbb{R}^N$, we define

$$\boxed{\delta_{x_0}[\varphi] := \varphi(x_0) \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}^N).}$$

This distribution, $\delta_{x_0} \in \mathcal{D}'(\mathbb{R}^N)$, is called the Dirac delta distribution at the point x_0.

Classical functions as distributions. The key formula reads as

$$\boxed{f[\varphi] := \int_{\mathbb{R}^N} f(x)\varphi(x)d^N x \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}^N).}$$

Let $L_{\text{loc}}(\mathbb{R}^N)$ be the space of all locally integrable functions $f : \mathbb{R}^N \to \mathbb{C}$, that is, the Lebesgue integral $\int_B f(x)d^N x$ exists for all compact subsets B of $\mathbb{R}^N$.[17] Then, the map $\varphi \mapsto f[\varphi]$ is a distribution which lies in the space $\mathcal{D}'(\mathbb{R}^N)$. We briefly write

$$\boxed{f \in \mathcal{D}'(\mathbb{R}^N).}$$

In this sense, we also write $L_{\text{loc}}(\mathbb{R}^N) \subset \mathcal{D}'(\mathbb{R}^N)$. Suppose that we are given functions $f, g \in L_{\text{loc}}(\mathbb{R}^N)$ such that

$$f[\varphi] = g[\varphi] \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}^N).$$

Then, $f(x) = g(x)$ for almost all $x \in \mathbb{R}^N$, in the sense of the N-dimensional Lebesgue measure. In particular, if $f, g : \mathbb{R}^N \to \mathbb{C}$ are continuous, then $f = g$ on $\mathbb{R}^N$. For example, the Heaviside function $\theta : \mathbb{R} \to \mathbb{R}$ corresponds to the distribution

$$\theta[\varphi] = \int_{\mathbb{R}} \theta(x)\varphi(x)dx = \int_0^\infty \varphi(x)dx \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}).$$

Convention. In what follows, we say that the distribution $F \in \mathcal{D}'(\mathbb{R}^N)$ is a classical function iff there exists a function $f : \mathbb{R}^N \to \mathbb{C}$ living in the space $L_{\text{loc}}(\mathbb{R}^N)$ such that

[17] For example, the function f is continuous (or almost everywhere continuous and bounded on bounded sets).

$$F(\varphi) = \int_{\mathbb{R}^N} f(x)\varphi(x)d^N x \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}^N).$$

To simplify notation, we also write f instead of F.

The product between classical functions and distributions. Let the function $\chi : \mathbb{R}^N \to \mathbb{C}$ be smooth. Then, for all test functions $\varphi \in \mathcal{D}(\mathbb{R}^N)$, the product function $\chi\varphi$ also lies in the space $\mathcal{D}(\mathbb{R}^N)$ of test functions. For given distribution $F \in \mathcal{D}'(\mathbb{R}^N)$, we define the product χF by setting

$$\boxed{(\chi F)(\varphi) := F(\chi\varphi) \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}^N).}$$

Then, χF is a distribution living in $\mathcal{D}'(\mathbb{R}^N)$. For example, $\chi\delta$ is given by

$$(\chi\delta)(\varphi) = \chi(0)\varphi(0) \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}^N).$$

The support of a distribution. Let Ω be an open subset of $\mathbb{R}^N$. For a distribution $F \in \mathcal{D}'(\mathbb{R}^N)$, we write

$$F = 0 \qquad \text{on} \ \ \Omega$$

iff $F(\varphi) = 0$ for all test functions $\varphi \in \mathcal{D}(\mathbb{R}^N)$ with $\text{supp}(\varphi) \subseteq \Omega$. By definition, the set $\text{zero}(F)$ is the maximal open subset of $\mathbb{R}^N$ such that $F = 0$ on this set. The complement

$$\text{supp}(F) := \mathbb{R}^N \setminus \text{zero}(F)$$

is called the support of the distribution F. For example, the Dirac delta distribution δ has the origin as support, and

$$\text{supp}(\delta_{x_0}) = \{x_0\}.$$

Derivative of a distribution. Let $f : \mathbb{R}^N \to \mathbb{C}$ be a smooth function. Integration by parts yields

$$\int_{\mathbb{R}^N} \partial^\alpha f(x)\varphi(x)d^N x = (-1)^{|\alpha|} \int_{\mathbb{R}^N} f(x)\partial^\alpha\varphi(x)d^N x$$

for all $\varphi \in \mathcal{D}(\mathbb{R}^N)$. This motivates the following definition. We are given the distribution $F \in \mathcal{D}'(\mathbb{R}^N)$. The derivative $\partial^\alpha F$ is defined by

$$\boxed{(\partial^\alpha F)(\varphi) := (-1)^{|\alpha|} F(\partial^\alpha\varphi) \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}^N).}$$

This is a distribution living in $\mathcal{D}'(\mathbb{R}^N)$. Note that

Each distribution has derivatives of all orders.

In particular, each function $f \in L_{\text{loc}}(\mathbb{R}^N)$ has derivatives of each order in the sense of distributions. Explicitly,

$$\partial^\alpha f[\varphi] = (-1)^{|\alpha|} \int_{\mathbb{R}^N} f(x)\partial^\alpha\varphi(x)d^N x \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}^N).$$

For the Dirac delta distribution, the derivative $\partial^\alpha\delta_{x_0}$ is given by

$$(\partial^\alpha\delta_{x_0})(\varphi) = (-1)^{|\alpha|}\partial^\alpha\varphi(x_0) \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}^N).$$

Typical examples. The following derivatives of distributions are used frequently.

(i) For the Heaviside function $\theta : \mathbb{R} \to \mathbb{R}$,

$$\boxed{\theta' = \delta \quad \text{on} \quad \mathbb{R}.}$$

In fact, for all test functions $\varphi \in \mathcal{D}(\mathbb{R})$, integration by parts yields

$$\int_{\mathbb{R}} -\theta(t)\varphi'(t)dt = \int_0^\infty -\varphi'(t)dt = \varphi(0).$$

Hence $\theta[-\varphi'] = \varphi(0) = \delta[\varphi]$ for all $\varphi \in \mathcal{D}(\mathbb{R})$.

(ii) Fix the complex number a and set $f(t) := \theta(t)\mathrm{e}^{at}$ for all $t \in \mathbb{R}$. Then

$$f' = af + \delta \quad \text{on} \quad \mathbb{R}.$$

To prove this, note that integration by parts yields

$$\int_{\mathbb{R}} \theta(t)\mathrm{e}^{at}(-\varphi'(t))dt = -\int_0^\infty \mathrm{e}^{at}\varphi'(t)dt = a\int_0^\infty \mathrm{e}^{at}\varphi(t)dt + \varphi(0).$$

Hence $f[-\varphi'] = af[\varphi] + \delta[\varphi]$ for all $\varphi \in \mathcal{D}(\mathbb{R})$.[18]

(iii) Set $g(t) := \theta(t)t$ for all $t \in \mathbb{R}$. Then

$$g'' = \delta \quad \text{on} \quad \mathbb{R}. \tag{11.41}$$

In fact, integration by parts tells us that

$$g[-\varphi'] = \int_0^\infty -t\varphi'(t)dt = \int_0^\infty \varphi(t)dt = \theta[\varphi].$$

Hence $g' = \theta$ on $\mathbb{R}$. By (i), $g'' = \theta' = \delta$, as claimed above.[19]

(iv) Suppose that the distribution $F \in \mathcal{D}'(\mathbb{R}^N)$ satisfies the equation[20]

$$\partial_j F = 0 \quad \text{on} \quad \mathbb{R}^N, \quad j = 1, \ldots, N.$$

Then, F is a constant classical function. Explicitly,

$$F(\varphi) = \text{const} \int_{\mathbb{R}^N} \varphi(x)d^N x \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}^N). \tag{11.42}$$

(v) The Weyl lemma: Fix the complex number a. If the distribution $F \in \mathcal{D}'(\mathbb{R}^N)$ satisfies the equation

$$\Delta F + aF = 0 \quad \text{on} \quad \mathbb{R}^N,$$

then F is a classical smooth function on $\mathbb{R}^N$.

The Sobolev space $W_2^k(\mathbb{R}^N)$. Let $k = 1, 2, \ldots$ By definition, the space $W_2^k(\mathbb{R}^N)$ consists of all functions $f \in L_2(\mathbb{R}^N)$ such that

$$\partial^\alpha f \in L_2(\mathbb{R}^N) \quad \text{for all} \ \ \alpha \ \ \text{with} \ |\alpha| \le k.$$

Explicitly, there exist functions g_α living in $L_2(\mathbb{R}^N)$ such that

[18] Formally, $f'(t) = \theta'(t)\mathrm{e}^{at} + \theta(t)a\mathrm{e}^{at} = \delta(t)\mathrm{e}^{at} + af(t) = \delta(t)\mathrm{e}^0 + af(t)$.

[19] Formally, $g'(t) = \delta(t)t + \theta(t) = \theta(t)$. Hence $g''(t) = \theta'(t) = \delta(t)$.

[20] The proof of (iv) and (v) can be found in Hörmander (1983), Vol. 1, Sect. 3.1 and Zeidler (1986), Vol. IIA, Sect. 18.15, respectively.

$$\int_{\mathbb{R}^N} f(x)(-1)^{|\alpha|}\partial^\alpha\varphi(x)d^Nx = \int_{\mathbb{R}^N} g_\alpha(x)\varphi(x)d^Nx$$

for all test functions $\varphi \in \mathcal{D}(\mathbb{R}^N)$ and all indices α with $|\alpha| \le k$. Then, $\partial^\alpha f = g_\alpha$. The space $W_2^k(\mathbb{R}^N)$ is a complex Hilbert space equipped with the inner product

$$\langle f|g\rangle := \int_{\mathbb{R}^N} \{f(x)^\dagger g(x) + \sum_{0<|\alpha|\le k} \partial^\alpha f(x)^\dagger \partial^\alpha g(x)\} d^Nx.$$

Limit of distributions. Let $F_n, F \in \mathcal{D}'(\mathbb{R}^N)$ for all n. We write

$$\lim_{n\to\infty} F_n = F \quad \text{in} \quad \mathcal{D}'(\mathbb{R}^N) \tag{11.43}$$

iff $\lim_{n\to\infty} F_n(\varphi) = F(\varphi)$ for all $\varphi \in \mathcal{D}(\mathbb{R}^N)$.

Proposition 11.10 *For all derivatives ∂^α, it follows from* (11.43) *that*

$$\lim_{n\to\infty} \partial^\alpha F_n = \partial^\alpha F.$$

For each continuous function $\psi : \mathbb{R}^N \to \mathbb{C}$, there exists a sequence (p_n) of polynomials $p_n : \mathbb{R}^N \to \mathbb{C}$ such that

$$\lim_{n\to\infty} \sup_{x\in B} |\psi(x) - p_n(x)| = 0$$

for all N-dimensional balls, B. This is the Weierstrass approximation theorem. Hence

$$\lim_{n\to\infty} \int_{\mathbb{R}^N} p_n(x)\varphi(x)d^Nx = \int_{\mathbb{R}^N} \psi(x)\varphi(x)d^Nx \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}^N).$$

Consequently, $\lim_{n\to\infty} p_n = \psi$ in $\mathcal{D}'(\mathbb{R}^N)$.

Application to the vibrating string. Fix $c > 0$. Let $f, g : \mathbb{R} \to \mathbb{R}$ be smooth functions. Then, the function

$$\psi(x,t) := f(x-ct) + g(x+ct), \qquad x,t \in \mathbb{R}$$

is a classical solution of the wave equation

$$\frac{\psi_{tt}(x,t)}{c^2} - \psi_{xx}(x,t) = 0, \qquad (x,t) \in \mathbb{R}^2. \tag{11.44}$$

If $f, g : \mathbb{R} \to \mathbb{R}$ are continuous, then the function ψ describes a vibrating string (superposition of two waves propagating with the speed c to right and to left, respectively). Thus, the function ψ possesses a well-defined physical meaning. However, as a rule, ψ is not a classical solution of the wave equation because of a lack of smoothness. However, the following hold true.

Proposition 11.11 *If $f, g : \mathbb{R} \to \mathbb{R}$ are continuous, then ψ solves the wave equation in the sense of distributions in $\mathcal{D}'(\mathbb{R}^2)$.*

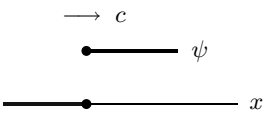


Fig. 11.3. Shock wave

Proof. To simplify notation, set $g = 0$. Approximate the function f by a sequence (p_n) of polynomials. For the function ψ_n corresponding to p_n,

$$\frac{(\psi_n)_{tt}}{c^2} - (\psi_n)_{xx} = 0.$$

Letting $n \to \infty$, $\lim_{n\to\infty} \psi_n = \psi$ in $\mathcal{D}'(\mathbb{R}^2)$. It follows from Prop. 11.10 that

$$\frac{\psi_{tt}}{c^2} - \psi_{xx} = 0 \quad \text{in} \quad \mathcal{D}'(\mathbb{R}^2).$$

Explicitly, this means that

$$\int_{\mathbb{R}^2} \psi(x,t) \left(\frac{\varphi_{tt}(x,t)}{c^2} - \varphi_{xx}(x,t) \right) dxdt = 0$$

for all test functions $\varphi \in \mathcal{D}(\mathbb{R}^2)$. □

This proof shows that the theory of distributions can be used in order to construct solutions of differential equations by using smooth approximations.

Propagation of singularities. The function

$$\psi(x,t) := \theta(x - ct), \qquad (x,t) \in \mathbb{R}^2 \tag{11.45}$$

is a solution of the wave equation (11.44) in the sense of distributions in the space $\mathcal{D}'(\mathbb{R}^2)$.

Proof. For fixed $\Delta t > 0$, choose the continuous function $\theta_{\Delta t} : \mathbb{R} \to \mathbb{R}$ as pictured in Fig. 11.1 on page 577. By Prop. 11.11, the function $\psi_{\Delta t}(x,t) := \theta_{\Delta t}(x - ct)$ satisfies the wave equation, that is,

$$\int_{\mathbb{R}^2} \theta_{\Delta t}(x - ct) \left(\frac{\varphi_{tt}(x,t)}{c^2} - \varphi_{xx}(x,t) \right) dxdt = 0$$

for all test functions $\varphi \in \mathbb{R}^2$. Letting $\Delta t \to 0$, we get

$$\int_{\mathbb{R}^2} \theta(x - ct) \left(\frac{\varphi_{tt}(x,t)}{c^2} - \varphi_{xx}(x,t) \right) dxdt = 0.$$

□

The solution ψ from (11.45) can be regarded as the propagation of a mass density ψ on the real line with velocity c. The jump of mass density at the point $x = ct$ at time t represents a simplified model for a shock wave in gas dynamics (Fig. 11.3). For example, supersonic aircrafts generate strongly audible shock waves in air. Shock waves are studied in the the standard textbook by Smoller (1994).

The propagation of singularities at the speed of light is typical for quantum field theory.

Nonlinear transformation of the Dirac delta distribution. We are given the smooth function $f : \mathbb{R}^N \to \mathbb{R}^N$ which has precisely the zeros $x_1, \ldots, x_n$. In addition, we assume that the zeros are non-degenerate, that is, $\det f'(x_j) \neq 0$ for $j = 1, \ldots, n$. We define[21]

$$\boxed{\delta \circ f := \sum_{j=1}^{n} \frac{\delta_{x_j}}{|\det f'(x_j)|}.}$$

Then, $\delta \circ f$ is a distribution living in $\mathcal{D}'(\mathbb{R}^N)$.

The space $\mathcal{D}'(\Omega)$. Let Ω be a nonempty open subset of $\mathbb{R}^N$. Replacing $\mathbb{R}^N$ by Ω, all of the considerations above can be generalized immediately to the set Ω. In particular, the space $\mathcal{D}(\Omega)$ of test functions consists of all smooth functions

$$\varphi : \Omega \to \mathbb{R}^N$$

which have compact support.[22] Furthermore, the space of distributions $\mathcal{D}'(\Omega)$ consists of all linear, sequentially continuous maps

$$F : \mathcal{D}(\Omega) \to \mathbb{C}.$$

Such distributions generalize functions of the form $f : \Omega \to \mathbb{C}$.

The Sobolev space $W_2^k(\mathbb{R}^N)$ is replaced by $W_2^k(\Omega)$. Here, we have to replace $\int_{\mathbb{R}^N}$ by $\int_\Omega$ above.

The space $\mathcal{E}'(\mathbb{R}^N)$. By definition, the space $\mathcal{E}(\mathbb{R}^N)$ consists of all smooth functions $\varphi : \mathbb{R}^N \to \mathbb{C}$. Let $\varphi_n, \varphi \in \mathcal{E}(\mathbb{R}^N)$ for all n. As $n \to \infty$, we write

$$\varphi_n \to \varphi \qquad \text{in } \mathcal{E}(\mathbb{R}^N)$$

iff φ_n (resp. every derivative $\partial^\alpha \varphi_n$) converges uniformly to φ (resp. to $\partial^\alpha \varphi$) on each compact subset of $\mathbb{R}^N$. By definition, the space $\mathcal{E}'(\mathbb{R}^N)$ consists of all linear, sequentially continuous functionals

$$F : \mathcal{E}(\mathbb{R}^{\mathbb{N}}) \to \mathbb{C}.$$

One can show that the space $\mathcal{E}'(\mathbb{R}^N)$ coincides with a linear subspace of $\mathcal{D}(\mathbb{R}^N)$, namely, the space of distributions with compact support. We have the inclusions

$$\mathcal{D}(\mathbb{R}^N) \subset \mathcal{S}(\mathbb{R}^N) \subset \mathcal{E}(\mathbb{R}^N) \subset \mathcal{E}'(\mathbb{R}^N) \subset \mathcal{S}'(\mathbb{R}^N) \subset \mathcal{D}'(\mathbb{R}^N).$$

11.3.3 Tempered Distributions

Let us now introduce a subclass of distributions which are called tempered distributions and which play a crucial role for the Fourier transform. The definition of the space $\mathcal{S}(\mathbb{R}^N)$ of rapidly decreasing functions

$$\varphi : \mathbb{R}^N \to \mathbb{C}$$

[21] This definition is motivated by the heuristic formula (11.25) for $\delta(f(x))$.

[22] The symbol $C^\infty(\Omega)$ denotes the space of smooth functions $\varphi : \Omega \to \mathbb{R}$. The space $\mathcal{D}(\Omega)$ is also denoted by $C_0^\infty(\Omega)$ in the literature.

along with the appropriate notion of convergence can be found in Sect. 10.3.3 on page 535. By a tempered distribution on $\mathbb{R}^N$, we mean a linear, sequentially continuous map

$$F : \mathcal{S}(\mathbb{R}^N) \to \mathbb{C}.$$

Explicitly, we assign a complex number $F(\varphi)$ to each test function φ in $\mathcal{S}(\mathbb{R}^N)$. Moreover, for all $\varphi, \psi \in \mathcal{S}(\mathbb{R}^N)$ and all complex numbers α, β,

$$F(\alpha\varphi + \beta\psi) = \alpha F(\varphi) + \beta F(\psi).$$

Finally, $\lim_{n\to\infty} \varphi_n = \varphi$ in $\mathcal{S}(\mathbb{R}^N)$ always implies

$$\lim_{n\to\infty} F(\varphi_n) = F(\varphi).$$

The space of all such distributions is denoted by $\mathcal{S}'(\mathbb{R}^N)$. In particular, fix the point x_0 in $\mathbb{R}^N$, and set

$$\delta_{x_0}(\varphi) := \varphi(x_0) \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}^N).$$

This tempered distribution is called the tempered Dirac distribution. Note that $\mathcal{D}(\mathbb{R}^N) \subset \mathcal{S}(\mathbb{R}^N)$ and that the convergence in $\mathcal{D}(\mathbb{R}^N)$ implies the convergence in $\mathcal{S}(\mathbb{R}^N)$. Consequently, each tempered distribution is also a distribution, that is,

$$\boxed{\mathcal{S}'(\mathbb{R}^N) \subset \mathcal{D}'(\mathbb{R}^N).}$$

Let us now summarize some classes of functions which are tempered distributions. To this end, set

$$f[\varphi] := \int_{\mathbb{R}^N} f(x)\varphi(x)d^N x \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}^N)$$

where the function $f : \mathbb{R}^N \to \mathbb{C}$ is given.

Proposition 11.12 *The functional $\varphi \mapsto f[\varphi]$ is a tempered distribution on $\mathbb{R}^N$ if one of the following two conditions is satisfied.*

(i) *The function f is either a polynomial or $f \in L_2(\mathbb{R}^N)$.*

(ii) *The function $f \in L_{\mathrm{loc}}(\mathbb{R}^N)$ is of moderate growth at infinity. Explicitly, there exists a nonnegative integer m such that*

$$|f(x)| = O(||x||^m), \qquad ||x|| \to \infty.$$

The proof can be found in Problem 12.7. Finally, we define the product $x^\alpha F$ by setting

$$(x^\alpha F)(\varphi) := F(\chi\varphi) \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}^N)$$

where we set $\chi(x) := x^\alpha = (x^1)^{\alpha_1}(x^2)^{\alpha_2}\cdots(x^N)^{\alpha_N}$ for all $x \in \mathbb{R}^N$. If F is a tempered distribution on $\mathbb{R}^N$, then so is $x^\alpha F$.

The tensor product of distributions. Let $N, M = 1, 2, \ldots$. For given functions $\varphi \in \mathcal{D}(\mathbb{R}^N)$ and $\psi \in \mathcal{D}(\mathbb{R}^M)$, we define the tensor product $\varphi \otimes \psi$ by setting

$$(\varphi \otimes \psi)(x, y) := \varphi(x)\psi(y) \qquad \text{for all} \quad x, y \in \mathbb{R}^{N+M}.$$

If $f \in L_{\mathrm{loc}}(\mathbb{R}^N)$ and $g \in L_{\mathrm{loc}}(\mathbb{R}^M)$, then

$$\int_{\mathbb{R}^{M+N}} f(x)g(y)\varphi(x)\psi(y)d^N x d^M y = \int_{\mathbb{R}^N} f(x)\varphi(x)d^N x \int_{\mathbb{R}^M} g(y)\psi(y)d^M y.$$

Furthermore, for all test functions $\chi \in \mathcal{D}(\mathbb{R}^{N+M})$,

$$\int_{\mathbb{R}^{M+N}} f(x)g(y)\chi(x,y)d^N x d^M y = \int_{\mathbb{R}^M} \left(\int_{\mathbb{R}^N} f(x)\chi(x,y)d^N x \right) g(y)d^M y.$$

This motivates the following definition of the tensor product of two distributions.

Proposition 11.13 *For given distributions $F \in \mathcal{D}'(\mathbb{R}^N)$ and $G \in \mathcal{D}'(\mathbb{R}^M)$, there exists precisely one distribution $T \in \mathcal{D}'(\mathbb{R}^{N+M})$ such that*

$$T(\varphi \otimes \psi) = F(\varphi)G(\psi) \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}^N), \psi \in \mathcal{D}(\mathbb{R}^M).$$

If F and G are tempered distributions, then so is T.

We write $T = F \otimes G$ and call this the tensor product between F and G. The proof can be found in Triebel (1989), Sect. 13. Furthermore, for all test functions $\chi \in \mathcal{D}(\mathbb{R}^{N+M})$,

$$(F \otimes G)(\chi) = G(F(\chi)).$$

This means that, we first apply F to the test function $x \mapsto \chi(x,y)$ for fixed $y \in \mathbb{R}^M$, and we then apply G to the test function $y \mapsto F(\chi(.,y))$.

The convolution of distributions. Recall that the function $f * g$ given by

$$(f * g)(x) := \int_{\mathbb{R}^N} f(x-y)g(y)d^N y \qquad \text{for all} \quad x \in \mathbb{R}^N$$

is called the convolution between f and g. The convolution makes sense if one of the following two conditions is satisfied:[23]

(a) The functions $f, g : \mathbb{R}^N \to \mathbb{C}$ are continuous and g has compact support.
(b) $f, g \in \mathcal{S}(\mathbb{R}^N)$.

Using the substitution $z = x - y$, for each test function $\chi \in \mathcal{D}(\mathbb{R}^N)$ we get

$$\int_{\mathbb{R}^N} (f * g)(x)\chi(x)d^N x = \int_{\mathbb{R}^N} f(x) \left(\int_{\mathbb{R}^N} g(y)\chi(x+y)d^N y \right) d^N x.$$

This motivates the definition

$$\boxed{(F * G)(\chi) := F\{G(\chi(x+...))\} \qquad \text{for all} \quad \chi \in \mathcal{D}(\mathbb{R}^N).}$$

This means that we first apply G to the test function $y \mapsto \chi(x+y)$ for fixed $x \in \mathbb{R}$, and we then apply F to the test function[24] $x \mapsto G(\chi(x+...))$. The proof of the following theorem can be found in Triebel (1989), Sect. 13.

Theorem 11.14 *We are given the distributions $F, G \in \mathcal{D}'(\mathbb{R}^N)$ where G has compact support. Then:*
(i) *$F * G \in \mathcal{D}'(\mathbb{R}^N)$.*
(ii) *$\partial^\alpha(F * G) = \partial^\alpha F * G = F * \partial^\alpha G$ for all derivatives.*
(iii) *If F has compact support, then $F * G = G * F$. In particular, we obtain $F * \delta = \delta * F$.*
(iv) *If $F \in \mathcal{S}'(\mathbb{R}^N)$, then $F * G \in \mathcal{S}'(\mathbb{R}^N)$.*

[23] Properties of the classical convolution can be found on page 537.
[24] The symbol $\chi(x+...)$ stands for the function $y \mapsto \chi(x+y)$.

11.3.4 The Fourier Transform

Recall from Sect. 10.3.3 on page 535 that the Fourier transform

$$\mathcal{F} : \mathcal{S}(\mathbb{R}^N) \to \mathcal{S}(\mathbb{R}^N)$$

is linear, bijective, and sequentially continuous. For all functions g, φ in $\mathcal{S}(R^N)$, we have the identity

$$\int_{\mathbb{R}^N} (\mathcal{F}g)(x)\varphi(x)d^N x = \int_{\mathbb{R}^N} g(x)(\mathcal{F}\varphi)(x)d^N x.$$

This motivates the following definition. For each tempered distribution $G \in \mathcal{S}'(\mathbb{R}^N)$, we define the Fourier transform $\mathcal{F}G$ by setting

$$\boxed{(\mathcal{F}G)(\varphi) := G(\mathcal{F}\varphi) \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}^N).} \tag{11.46}$$

It can be shown that $\mathcal{F}G$ is a tempered distribution. The map

$$\mathcal{F} : \mathcal{S}'(\mathbb{R}^N) \to \mathcal{S}'(\mathbb{R}^N)$$

is linear, bijective, and sequentially continuous.[25] In addition, for the inverse map $\mathcal{F}(G) \mapsto G$, we get

$$G(\varphi) = (\mathcal{F}G)(\mathcal{F}^{-1}\varphi) \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}^N).$$

The Fourier transform has the following properties:

(P1) $\mathcal{F}\delta = (2\pi)^{-N/2}$.
(P2) $\mathcal{F}\delta_{x_0} = \chi_{x_0}$ where $\chi_{x_0}(x) := (2\pi)^{-N/2}\mathrm{e}^{-\mathrm{i}\langle x_0|x\rangle}$ for all $x \in \mathbb{R}^N$.
(P3) $\mathcal{F}(\partial^\alpha G) = \mathrm{i}^{|\alpha|}x^\alpha \mathcal{F}G$ for all derivatives.

To prove this, choose an arbitrary test function $\varphi \in \mathcal{S}(\mathbb{R}^N)$. Obviously, (P1) is a special case of (P2). In order to prove (P2), note that

$$(\mathcal{F}\delta_{x_0})(\varphi) = \delta_{x_0}(\mathcal{F}\varphi) = (\mathcal{F}\varphi)(x_0) = \int_{\mathbb{R}^N} (2\pi)^{-N/2}\mathrm{e}^{-\mathrm{i}\langle x_0|x\rangle}\ \varphi(x)d^N x.$$

The derivative rule (P3) follows from

$$\partial^\alpha \int_{\mathbb{R}^N} \mathrm{e}^{-\mathrm{i}\langle p|x\rangle}\ \varphi(p)d^N p = \int_{\mathbb{R}^N} \mathrm{e}^{-\mathrm{i}\langle p|x\rangle}\ (-\mathrm{i})^{|\alpha|}p^\alpha \varphi(p)\ d^N p$$

along with

$$\begin{aligned}\mathcal{F}(\partial^\alpha G)(\varphi) &= \partial^\alpha G(\mathcal{F}\varphi) = (-1)^{|\alpha|}G(\partial^\alpha \mathcal{F}\varphi) \\ &= G(\mathcal{F}(\mathrm{i}^{|\alpha|}x^\alpha\varphi)) = (\mathcal{F}G)(\mathrm{i}^{|\alpha|}x^\alpha\varphi) = \mathrm{i}^{|\alpha|}(x^\alpha \mathcal{F}G)(\varphi).\end{aligned}$$

Let $f \in L_2(\mathbb{R}^N)$. If we regard the function f as tempered distribution, then the Fourier transform $\mathcal{F}f$ is a well-defined tempered distribution. It turns out that $\mathcal{F}f$ represents a function which lies in the space $L_2(\mathbb{R}^N)$. More generally, the following hold true.

Theorem 11.15 *The Fourier transform $\mathcal{F} : L_2(\mathbb{R}^N) \to L_2(\mathbb{R}^N)$ is a unitary transformation on the Hilbert space $L_2(\mathbb{R}^N)$.*

For the proof, see Triebel (1989), p. 109.

[25] The proof can be found in Zeidler (1995), Vol. 1, Sect. 3.8.

11.4 Hadamard's Regularization of Integrals

As a rule, the Green's functions for hyperbolic differential equations possess strong singularities. In order to represent such Green's functions by integral formulas, Hadamard (1865–1963) introduced the notion of the finite part of a divergent integral.[26]

11.4.1 Regularization of Divergent Integrals

In order to discuss the basic idea, choose the function $\varphi \in \mathcal{D}(\mathbb{R})$. The integral

$$J(\varphi) := \int_{\mathbb{R}} \frac{\varphi(x)dx}{x}$$

is convergent if $\varphi(0) = 0$ and divergent if $\varphi(0) \neq 0$. By definition, the regularization of this integral is given by

$$\boxed{J_{\text{reg}}(\varphi) = \int_{\mathbb{R}} \frac{\varphi(x) - \varphi(0)}{x}\, dx}$$

for all functions $\varphi \in \mathcal{D}(\mathbb{R})$. The additional term $-\frac{\varphi(0)}{x}$ is called regularizing term. We will also use the notation

$$J_{\text{reg}} = \mathcal{P}\left(\frac{1}{x}\right).$$

The following are met:

(i) The regularized functional, J_{reg}, is a distribution on $\mathbb{R}$.
(ii) For all test functions $\varphi \in \mathcal{D}(\mathbb{R})$,

$$J_{\text{reg}}(\varphi) = PV \int_{\mathbb{R}} \frac{\varphi(x)}{x}\, dx.$$

As usual, the symbol PV stands for Cauchy's principal value of the integral. This is defined by the following limit

$$PV \int_{\mathbb{R}} \frac{\varphi(x)}{x}\, dx = \lim_{\varepsilon \to +0} \left(\int_{-\infty}^{-\varepsilon} \frac{\varphi(x)}{x}\, dx + \int_{\varepsilon}^{\infty} \frac{\varphi(x)}{x}\, dx \right).$$

Proof. Ad (i). If $\varphi_n \to \varphi$ in $\mathcal{S}(\mathbb{R})$ as $n \to \infty$, then $J_{\text{reg}}(\varphi_n) \to J_{\text{reg}}(\varphi)$.

Ad (ii). Choose $\varepsilon > 0$. Since the following integrand is odd,

$$\int_{-\infty}^{-\varepsilon} \frac{\varphi(0)}{x}\, dx + \int_{\varepsilon}^{\infty} \frac{\varphi(0)}{x}\, dx = 0.$$

Finally, choose the following decomposition

$$\frac{\varphi(x)}{x} = \frac{\varphi(x) - \varphi(0)}{x} + \frac{\varphi(0)}{x}.$$

□

Note that if the integral $J(\varphi)$ is convergent, then $J_{\text{reg}}(\varphi) = J(\varphi)$. However, if the integral $J(\varphi)$ is divergent, then the well-defined integral $J_{\text{reg}}(\varphi)$ is called the finite part of $J(\varphi)$, in the sense of Hadamard.

[26] J. Hadamard, Le problème de Cauchy et les équations aux dérivées partielles linéaire hyperboliques, Paris, 1932. For a modern version of the theory, see the monograph by P. Günther (1988).

11.4.2 The Sokhotski Formula

The classical function

$$f(x) := \frac{1}{x} \qquad \text{for all} \quad x \in \mathbb{R} \setminus \{0\}$$

has a singularity at the point $x = 0$. There exist three different distributions which can be regarded as regularizations of this singular function, namely,

$$\mathcal{P}\left(\frac{1}{x}\right), \qquad \frac{1}{x + 0_+\mathrm{i}}, \qquad \frac{1}{x - 0_+\mathrm{i}}.$$

To discuss this, let z be a fixed, but otherwise arbitrary complex number. For each test function $\varphi \in \mathcal{D}(\mathbb{R})$, define

$$\mathcal{P}\left(\frac{1}{x-z}\right)(\varphi) := PV \int_{\mathbb{R}} \frac{\varphi(x)dx}{x-z}.$$

This is a distribution living in $\mathcal{D}'(\mathbb{R})$. If the parameter z is not real, then the integral exists in the classical sense. If z is a real number, then it follows as in the preceding section that

$$\mathcal{P}\left(\frac{1}{x-z}\right)(\varphi) := \int_{\mathbb{R}} \frac{\varphi(y+z) - \varphi(z)}{y}\, dy.$$

Choose $z := \mp\varepsilon\mathrm{i}$ with $\varepsilon > 0$.

Theorem 11.16 *For every test function $\varphi \in \mathcal{D}(\mathbb{R})$, we have the classical Sokhotski formula*

$$\lim_{\varepsilon \to +0} \mathcal{P}\left(\frac{1}{x \pm \varepsilon\mathrm{i}}\right)(\varphi) = \mathcal{P}\left(\frac{1}{x}\right) \mp \mathrm{i}\pi\varphi(0). \tag{11.47}$$

Proof. We have to show that for all test functions $\varphi \in \mathcal{D}(\mathbb{R})$,

$$\lim_{\varepsilon \to +0} \int_{-\infty}^{\infty} \frac{\varphi(x)dx}{x + \varepsilon\mathrm{i}} = -\mathrm{i}\pi\varphi(0) + \int_{-\infty}^{\infty} \frac{\varphi(x) - \varphi(0)}{x}\, dx.$$

The key to this limit is the identity

$$\begin{aligned} \int_{-R}^{R} \frac{dx}{x + \varepsilon\mathrm{i}} &= \int_{-R}^{R} \frac{x - \varepsilon\mathrm{i}}{x^2 + \varepsilon^2}\, dx \\ &= -\int_{-R}^{R} \frac{\varepsilon\mathrm{i}}{x^2 + \varepsilon^2}\, dx = -2\mathrm{i} \arctan \frac{R}{\varepsilon} \end{aligned}$$

which converges to $-\mathrm{i}\pi$ as $\varepsilon \to +0$. Since the test function φ has compact support, there exists an interval $[-R, R]$ such that φ vanishes outside this interval. Therefore, it remains to use the decomposition

$$\int_{-R}^{R} \frac{\varphi(x)dx}{x + \varepsilon\mathrm{i}} = \int_{-R}^{R} \frac{\varphi(x) - \varphi(0)}{x + \varepsilon\mathrm{i}}\, dx + \varphi(0) \int_{-R}^{R} \frac{dx}{x + \varepsilon\mathrm{i}},$$

and to perform the limit $\varepsilon \to +0$. □

Let us translate this into the language of distributions. The Sokhotski formula tells us that

$$\lim_{\varepsilon\to+0}\mathcal{P}\left(\frac{1}{x\pm\varepsilon\mathrm{i}}\right)=\mathcal{P}\left(\frac{1}{x}\right)\mp\mathrm{i}\pi\delta,$$

in the sense of distributions. This is called an adiabatic limit. We now define the distribution

$$\boxed{\frac{1}{x+0_+\mathrm{i}}:=\mathcal{P}\left(\frac{1}{x}\right)-\mathrm{i}\pi\delta.} \tag{11.48}$$

Similarly, we define the distribution

$$\frac{1}{x-0_+\mathrm{i}}:=\mathcal{P}\left(\frac{1}{x}\right)+\mathrm{i}\pi\delta.$$

If $\Im(z)\neq 0$, then

$$\mathcal{P}\left(\frac{1}{x-z}\right)(\varphi):=\int_{\mathbb{R}}\frac{\varphi(x)dx}{x-z}$$

where the integral exists in the classical sense. According to the usual identification between functions and the corresponding distributions, we write

$$\mathcal{P}\left(\frac{1}{x-z}\right)=\frac{1}{x-z}\quad\text{if }\ \Im(z)\neq 0.$$

Thus, in the sense of distributions,

$$\frac{1}{x\pm 0_+\mathrm{i}}=\lim_{\varepsilon\to+0}\frac{1}{x\pm\varepsilon\mathrm{i}}.$$

The considerations above remain valid if the test functions φ lie in the space $\mathcal{S}(\mathbb{R})$. Thus, $\frac{1}{x+0_+\mathrm{i}}$ and $\frac{1}{x-0_+\mathrm{i}}$ are also tempered distributions.

11.4.3 Steinmann's Renormalization Theorem

> The quest for the existence of a non-trivial quantum field in four space-time dimensions is still without any conclusive result. Nonetheless, physicists are working daily, with success, on concrete models which describe very efficiently physics at wide energy scales. This description is based on expansion of physical quantities like amplitudes of scattering processes of power series of "physical" parameters, as coupling constants, masses, charges. The higher order terms of these power series are usually ill-defined, in a naive approach, but physicists have soon learned how to make sense of them through the procedure now known as renormalization... On Minkowski space-time, Steinmann's concept of the scaling degree of a generalized function at a point leads to a rather smooth and economic method of renormalization...
>
> Romeo Brunetti and Klaus Fredenhagen, 2000[27]

[27] Microlocal analysis and interacting quantum fields: renormalization on physical backgrounds, Commun. Math. Phys. **208** (2000), 623–661 (reprinted with permission).

The support theorem. Let $c_0, \ldots, c_m$ be complex numbers. Set

$$F = c_0\delta + c_1\delta' + c_2\delta'' + \ldots + c_m\delta^{(m)}. \tag{11.49}$$

Then, the distribution F has the origin as support (or it vanishes identically). Interestingly enough, the converse is also true. The proof of the following theorem can be found in Triebel (1989), Sect. 4.

Theorem 11.17 *If the distribution $F \in \mathcal{D}'(\mathbb{R}^N)$ has the origin as support, then it has the form* (11.49).

The local degree of homogeneity of a distribution. Let σ_0 be a real number. The function $f : \mathbb{R}^N \to \mathbb{C}$ is said to be homogeneous of order σ_0 iff

$$\boxed{f(\lambda x) = \lambda^{\sigma_0} f(x) \qquad \text{for all} \quad x \in \mathbb{R}^N,\ x \neq 0,\ \ \lambda > 0.} \tag{11.50}$$

For example, if $N = 1$, then the functions $f(x) := x, x^2, x^{-1}, x^{-2}$ are homogeneous of order (or degree) $\sigma_0 = 1,\ 2,\ -1\ -2$, respectively. We write

$$\deg_h f := \sigma_0.$$

We want to generalize this notion to distributions. Let $\lambda > 0$. Suppose first that the function f is locally integrable along with (11.50). For all test functions $\varphi \in \mathcal{D}(\mathbb{R}^N)$, we then have the two identities

$$\lambda^{-\sigma} \int_{\mathbb{R}^N} f(\lambda y)\varphi(y)d^N y = \lambda^{\sigma_0 - \sigma} \int_{\mathbb{R}^N} f(y)\varphi(y)d^N y,$$

and

$$\int_{\mathbb{R}^N} f(\lambda y)\varphi(y)d^N y = \int_{\mathbb{R}^N} f(x)\lambda^{-N}\varphi(\lambda^{-1}x)d^N x,$$

by the substitution $x := \lambda y$. Now consider a general distribution $F \in \mathcal{D}'(\mathbb{R}^N)$. Set

$$(S_\lambda \varphi)(x) := \lambda^{-N}\varphi(\lambda^{-1}x) \qquad \text{for all} \quad x \in \mathbb{R}^N,\ \lambda > 0.$$

Suppose that there exists a real number σ such that

$$\boxed{\lim_{\lambda \to +0} \lambda^{-\sigma} F(S_\lambda \varphi) = 0 \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}^N).}$$

Let σ_0 be the supremum of all such numbers σ. We then write $\sigma_0 = \deg_h F$, and we call this the local degree of homogeneity of the distribution F. If there is no such real number σ, we set $\deg_h F := -\infty$. Therefore, each distribution $F \in \mathcal{D}'(\mathbb{R}^N)$ has a local degree of homogeneity, and

$$-\infty \leq \deg_h F \leq +\infty.$$

Let us consider some typical examples.

(i) Homogeneous distribution: Suppose that there exists a real number σ_0 such that

$$F(S_\lambda \varphi) = \lambda^{\sigma_0} F(\varphi) \qquad \text{for all} \quad \lambda > 0$$

and all test functions $\varphi \in \mathcal{D}(\mathbb{R}^N)$. Then $\sigma_0 = \deg F$.

(ii) Dirac's delta distribution δ on $\mathbb{R}^N$: For all $\lambda > 0$ and $\varphi \in \mathcal{D}(\mathbb{R}^N)$,

$$\delta(S_\lambda \varphi) = (S_\lambda \varphi)(0) = \lambda^{-N} \varphi(0) = \lambda^{-N} \delta(\varphi).$$

Hence $\deg_h \delta = -N$.

Steinmann's renormalization theorem for distributions. We are given a distribution

$$F_0 \in \mathcal{D}'(\mathbb{R}^N \setminus \{0\})$$

with finite local degree of homogeneity, $\deg_h F_0$. [28] By an admissible extension of F_0, we mean a distribution

$$F \in \mathcal{D}'(\mathbb{R}^N)$$

with $F(\varphi) = F_0(\varphi)$ for all $\varphi \in \mathcal{D}(\mathbb{R}^N \setminus \{0\})$, and $\deg_h F = \deg_h F_0$. The following theorem is basic for renormalization theory.

Theorem 11.18 *Each distribution F_0 above has an admissible extension F. If $\deg_h F_0 > -N$, then the admissible extension of F_0 is unique.*

If $\deg_h F_0 \leq -N$, then the family of all admissible extensions of F_0 depends on a finite number of complex parameters.

More precisely, in the critical case where $\deg_h F_0 \leq -N$, suppose that the degree $\deg_h F_0$ is an integer. Then the set of all admissible extensions of F_0 is given by

$$F + \sum_{|\alpha| \leq m} c_\alpha \partial^\alpha \delta$$

where F is a fixed admissible extension, and $m = -N - \deg_h F_0$. In addition, c_α are arbitrary complex numbers. The admissible extensions are also called Steinmann extensions. In quantum field theory, the free parameters c_α have to be determined by additional symmetry conditions (e.g., Ward–Takehashi identities). This is the method of renormalization. The proof of Theorem 11.18 will be given in Volume II in connection with the Epstein–Glaser approach to quantum field theory.[29] Let us consider a typical application.

11.4.4 Regularization Terms

> Modern functional analysis solves classical problems in a generalized setting by extending function spaces, operators, and functionals to more general objects. For example, this concerns variational problems, the partial differential equations of mathematical physics, and optimization problems. The key words for typical extensions are Sobolev spaces, the Friedrichs extension of a formally self-adjoint operator, the Hahn–Banach extension of functionals, and the Steinmann extension of distributions.[30]
>
> Folklore

[28] This degree is defined as above by restricting the test functions φ to the space $\mathcal{D}(\mathbb{R}^N \setminus \{0\})$.

[29] See also Steinmann (1971), and Brunetti, Fredenhagen (2000).

[30] This is thoroughly studied in the author's textbook, E. Zeidler, Applied Functional Analysis, Vols. 1, 2, Springer, New York, 1995.

Consider the integral

$$\boxed{F_0(\varphi) := \int_{\mathbb{R}} \frac{\varphi(x)}{x} dx.}$$

If we restrict ourselves to smooth test functions $\varphi : \mathbb{R} \setminus \{0\} \to \mathbb{C}$ having compact support, that is, these functions vanish outside a compact subset of the open set $\mathbb{R} \setminus \{0\}$, then φ vanishes in some open neighborhood of the origin and outside some sufficiently large interval. Consequently, the integral $F_0(\varphi)$ is well-defined; it represents a distribution in the space $\mathcal{D}'(\mathbb{R} \setminus \{0\})$ with local degree of homogeneity, $\deg_h F_0 = -1$. By the Steinmann Renormalization Theorem 11.18, we obtain the following.

Theorem 11.19 *The most general extension of the distribution F_0 to a distribution $F \in \mathcal{D}'(\mathbb{R})$ with the same degree of homogeneity, $\deg_h F = -1$, is given by*

$$F = J_{\rm reg} + c_0 \delta$$

where c is an arbitrary complex number.

Explicitly, we obtain the following expression

$$F(\varphi) = \int_{\mathbb{R}} \left(\frac{\varphi(x)}{x} - \frac{\varphi(0)}{x} \right) dx + c\varphi(0)$$

for all test functions $\varphi \in \mathcal{D}(\mathbb{R})$.

As next example, consider the divergent integral

$$\boxed{\int_{\mathbb{R}} \frac{dx}{x^2}.} \tag{11.51}$$

In order to renormalize this integral, we proceed as follows.

(i) We introduce the functional

$$F_0(\varphi) := \int_{\mathbb{R}} \frac{\varphi(x)}{x}\, dx$$

for all test functions $\varphi \in \mathcal{D}(\mathbb{R} \setminus \{0\})$. This is a distribution living in the space $\mathcal{D}'(\mathbb{R} \setminus \{0\})$. By Taylor expansion,

$$\varphi(x) = \varphi(0) + \varphi'(0)x + O(x^2), \qquad x \to 0.$$

Therefore, we construct the regularized integral[31]

$$F_{\rm reg}(\varphi) := \int_{\mathbb{R}} \frac{\varphi(x) - \varphi(0) - \varphi'(0)x}{x^2}\, dx$$

which exists for all test functions $\varphi \in \mathcal{D}(\mathbb{R})$. The distribution $F_{\rm reg}$ living in $\mathcal{D}'(\mathbb{R})$ represents an extension of F_0.

[31] This integral represents the finite part of the integral $F_0(\varphi)$, in the sense of Hadamard.

(iii) The distribution F_0 has the local degree of homogeneity, $\deg_h F_0 = -2$. The most general distribution $F \in \mathcal{D}'(\mathbb{R})$ of local degree of homogeneity -2, which is an extension of F_0, reads as

$$F = F_{\text{reg}} + c_0\delta + c_1\delta'$$

where c_0 and c_1 are arbitrary complex numbers. Explicitly,

$$F(\varphi) = \int_{\mathbb{R}} \frac{\varphi(x) - \varphi(0) - \varphi'(0)x}{x^2}\, dx + c_0\varphi(0) + c_1\varphi'(0)$$

for all test functions $\varphi \in \mathcal{D}(\mathbb{R})$. Note that also this makes sense for all test functions $\varphi \in \mathcal{S}(\mathbb{R})$. Consequently, F is not only a distribution on the real line, but also a tempered distribution.

By definition, the tempered distribution F is called the regularization of the divergent integral (11.51).

The philosophy of renormalization. Typically, the process of renormalization in physics introduces additional parameters which have to be fixed by additional considerations, frequently based on physical arguments.

The philosophy is that we extract relevant physical information from given, possibly ill-defined mathematical objects.

As a typical example, consider the Fourier series

$$\boxed{\sum_{p=-\infty}^{\infty} a_p(\psi)\varphi_p(x)}$$

with the basis functions $\varphi_p(x) := \mathrm{e}^{\mathrm{i}px}/\sqrt{2\pi}$ and the Fourier coefficients

$$a_p(\psi) := \int_{-\pi}^{\pi} \varphi_p(x)^\dagger \psi(x) dx, \qquad p = 0, \pm 1, \pm 2, \ldots$$

If the 2π-periodic function $\psi : \mathbb{R} \to \mathbb{C}$ is measurable, and $\int_{-\pi}^{\pi} |\psi(x)| dx < \infty$, then the Fourier coefficients are well-defined. In contrast to this, the Fourier series is not always well-defined as a classical convergent series.

The task is to reconstruct the function ψ from its Fourier coefficients.

A nice answer to a special case of this problem is given by the following Fejér theorem from 1904.[32]

Proposition 11.20 *Let $\psi : \mathbb{R} \to \mathbb{C}$ be a continuous 2π-periodic function. Introduce the partial sums*

$$s_n(x) := \sum_{p=-n}^{n} a_p(\psi)\varphi_p(x), \qquad x \in \mathbb{R}, \;\; n = 0, 1, \ldots$$

Then, the classical limit

$$\psi(x) = \lim_{n\to\infty} s_n(x)$$

does not always exist. However, for all $x \in \mathbb{R}$, we have the convergence

[32] The proof can be found in Zorich (2003), Vol. I, p. 538.

$$\psi(x) = \lim_{N\to\infty} \frac{s_0(x) + s_1(x) + \ldots + s_N(x)}{N+1}$$

for the mean averages of the partial sums. In addition, the convergence is uniform on the real line.

Similarly, it is possible to extract important information from divergent integrals. This fact is used critically in renormalization theory. For the regularization of divergent integrals, we refer to

- Sect. 8.5.3 on page 511 (counterterms), and
- Sect. 11.6.2 on page 636 (dimensional regularization).

11.5 Renormalization of the Anharmonic Oscillator

Resonances are dangerous for the mathematics of physical systems.
Folklore

11.5.1 Renormalization in a Nutshell

This section should help the reader to understand the basic ideas of renormalization theory from both the mathematical and physical point of view. We will use a simple example in order to demonstrate the relation between the critical phenomenon of resonance of an oscillating system and the renormalization of physical parameters. Our rigorous approach will be based on the classical methods of bifurcation theory in nonlinear functional analysis.[33] This way, we will also clarify the role of renormalized Green's functions which represent an important tool used by physicists in renormalization theory.

11.5.2 The Linearized Problem

In nature, oscillating systems play a crucial role. As examples, consider electromagnetic waves or the periodic motion of planets around the sun. As a rule, the motion of coupled oscillating systems becomes complicated if resonance occurs, that is, the difference between two frequencies becomes small. As a simple model, consider the motion $q = q(t)$,

$$\ddot{q}(t) + \omega^2 q(t) = F(t), \quad q(0) = q_0, \;\; \dot{q}(0) = q_1, \tag{11.52}$$

of a particle on the real line under the influence of the smooth external force F. Here, the angular frequency ω is a fixed positive number. The unique solution is given by

$$q(t) = q_0 \cos\omega t + \frac{q_1}{\omega} \sin\omega t + \int_0^t G(t,\tau) F(\tau)\, d\tau, \qquad t \in \mathbb{R}$$

with the Green's function

[33] A detailed study of bifurcation theory can be found in Zeidler (1986), Vol. 1, Chap. 8.

$$G(t,\tau) := \frac{1}{\omega}\sin\omega(t-\tau), \qquad t,\tau \in \mathbb{R}.$$

Now choose the periodic external force

$$F(t) := \sin\alpha t, \qquad t \in \mathbb{R}$$

with the fixed angular frequency $\alpha > 0$. We have to distinguish between the following two cases.

(i) Non-resonance. Let $\alpha \neq \omega$. Then

$$q(t) = q_0\cos\omega t + \frac{q_1}{\omega}\sin\omega t + \frac{\sin\alpha t + \sin\omega t}{2(\alpha+\omega)\omega} - \frac{\sin\alpha t - \sin\omega t}{2(\alpha-\omega)\omega}.$$

If α is near ω, then we speak of the small divisor $\alpha - \omega$.

(ii) Resonance. Let $\alpha = \omega$. The limit $\alpha \to \omega$ yields

$$q(t) = q_0\cos\omega t + \frac{q_1}{\omega}\sin\omega t + \frac{\sin\omega t}{2\omega^2} - \frac{t}{2\omega}\cos\omega t.$$

This is the unique solution of (11.52).

The appearance of the term $t\cos\omega t$ corresponds to a motion with increasing amplitude. In practice, such resonant motions may destroy bridges. From the physical point of view, for vanishing external force, the system (11.52) possesses the eigenoscillations

$$q(t) = q_0\cos\omega t + \frac{q_1}{\omega}\sin\omega t.$$

Since the external force $F(t) := \sin\omega t$ has the same angular frequency ω as the eigenoscillations, the external force amplifies the eigenoscillations. From the mathematical point of view, small divisors are responsible for resonance effects.

In the 1950s and 1960s, Kolmogorov, Arnold, and Moser showed that resonance effects can cause chaotic motions of bodies in celestial mechanics (KAM theory). Chaotic motions of asteroids have been observed by astronomers.[34]

The following material serves as preparation for renormalization. Let us consider the boundary-value problem

$$\boxed{\ddot{q}(t) + \omega^2 q(t) = F(t), \quad q(0) = q(\pi) = 0} \tag{11.53}$$

where $\omega > 0$. Let C_π^- denote the space of continuous π-periodic odd functions $f : \mathbb{R} \to \mathbb{R}$ with the norm

$$\|f\|_\pi := \max_{0\le t\le\pi} |f(t)|.$$

We are given the external force $F \in \mathbb{C}_\pi^-$. We are looking for a solution $q \in C_\pi^-$ which is twice continuously differentiable on the interval $[0,\pi]$. Such a solution is called a classical solution of (11.53).

Eigenoscillations. Consider first the case where the external force vanishes, $F \equiv 0$. Problem (11.53) possesses then a nontrivial solution iff we choose $\omega = k$ for some $k = 1, 2, \ldots$. The solution reads as

$$q(t) = \text{const} \cdot \sin kt.$$

To each function $q \in C_\pi^-$, we assign the Fourier series

[34] As an introduction to KAM theory, we recommend Scheck (2000), Vol. 1 and Thirring (1997).

$$\boxed{b_1(q)\sin t + b_2(q)\sin 2t + \dots} \tag{11.54}$$

with the Fourier coefficients

$$b_k(q) := \sqrt{\frac{2}{\pi}} \int_0^\pi q(t)\sin kt\,dt, \qquad k = 1, 2, \dots$$

If the function q is smooth, then the Fourier series (11.54) converges to $q(t)$ uniformly on the real line. Moreover, all the derivatives $q', q'', \dots$ can be computed by differentiating the Fourier series (11.54) term by term, and these series converge uniformly on the real line, by (10.8) on page 533.

The non-resonance case. Let $\omega \neq k$ for all $k = 1, 2, \dots$. If q is classical solution of problem (11.53), then the Fourier coefficients of q read as

$$\boxed{b_k(q) = \frac{b_k(F)}{\omega^2 - k^2}, \quad k = 1, 2, \dots} \tag{11.55}$$

This follows from

$$b_k(F) = \sqrt{\frac{2}{\pi}} \int_0^\pi F(t)\sin kt\,dt = \sqrt{\frac{2}{\pi}} \int_0^\pi \big(\ddot{q}(t) + \omega^2 q(t)\big)\sin kt\,dt$$

along with integration by parts

$$b_k(F) = \sqrt{\frac{2}{\pi}} \int_0^\pi (\omega^2 - k^2) q(t)\sin kt\,dt = (\omega^2 - k^2) b_k(q). \tag{11.56}$$

Proposition 11.21 *Suppose we are given the function $F \in C_\pi^-$ and the positive parameter ω with $\omega \neq 1, 2, \dots$ Then, the original boundary-value problem* (11.53) *has the unique classical solution*

$$q(t) = \int_0^\pi G(t, \tau; \omega) F(\tau)\,d\tau, \qquad t \in [0, \pi] \tag{11.57}$$

with the kernel

$$G(t, \tau; \omega) := \sqrt{\frac{2}{\pi}} \cdot \sum_{k=1}^\infty \frac{\sin kt\, \sin k\tau}{\omega^2 - k^2}$$

which is called the Green's function.

Proof. For all indices $k = 1, 2, \dots$,

$$\left| \frac{\sin kt\, \sin k\tau}{\omega^2 - k^2} \right| \leq \frac{\text{const}}{k^2} \qquad \text{for all} \quad t, \tau \in \mathbb{R}.$$

Therefore, the series for G converges uniformly for all $t, \tau \in \mathbb{R}$. Hence the Green's function G is continuous. We now restrict ourselves to the special case where the function F is a finite Fourier series. By (11.55),

$$q(t) = \sum_{k=1}^\infty b_k(q)\, \sin kt = \sum_{k=1}^\infty \frac{b_k(F)\sin kt}{\omega^2 - k^2}\,.$$

Inserting the integral $b_k(F) = \sqrt{\frac{2}{\pi}} \int_0^\pi F(\tau)\sin k\tau\,d\tau$, we get the claim. Note that summation and integration can be interchanged because of uniform convergence. □

Let us now discuss the general case where the given function F lies in the Banach space C_π^-.

- Step 1: Classical solution: Our proof above shows that the function $q = q(t)$ from (11.57) is a classical solution of the original problem (11.53) if the given function F is a finite Fourier series.
- Step 2: Generalized solution: If $F \in C_\pi^-$, then the integral (11.57) is still well-defined. Since the Green's function G is continuous, the function q is at least continuous. Therefore, the function $q = q(t)$ from (11.57) can be regarded as a generalized solution of the original problem (11.53). It can be shown that this generalized solution $q \in C_\pi^-$ satisfies the original problem (11.53) in the sense of distribution theory, that is, the derivatives of q are to be understood as distributional derivatives.
- Step 3: Regularity of the generalized solution: Standard arguments from the regularity theory for boundary-value problems of second-order ordinary differential equations show that q is indeed a classical solution of (11.53) which has continuous derivatives of second order.

For the functional analytic proof of the statement from step 3, see Problem 12.10. The proof will be based on the theory of Fredholm operators of index zero in Banach spaces and the approximation theorem of periodic continuous functions by finite Fourier series (the Fejér theorem).

The procedure from step 1 through step 3 is typical for the modern, functional analytic theory of differential equations.

The resonance case. Let $\omega = 1$. If problem (11.53) has a solution q, then it follows from (11.56) that

$$b_1(F) = (\omega^2 - 1)b_1(q) = 0. \tag{11.58}$$

In terms of physics, the external force $F(t) = \sin t$ is in resonance with the eigenoscillation $q(t) = \sin t$. In this case, the original boundary-value problem (11.53) has no solution. Observe that the Green's function G from Prop. 11.21 has a singularity at $\omega = 1$. Suppose that the external force F satisfies the resonance condition $b_1(F) = 0$ and that q is a solution of (11.53). Then, by (11.56), the Fourier coefficients of q satisfy the equation

$$b_k(q) = \frac{b_k(F)}{1 - k^2}, \qquad k = 2, 3, \ldots \tag{11.59}$$

By (11.58), the Fourier coefficient $b_1(q)$ is undetermined.

Proposition 11.22 *If $\omega = 1$ and $b_1(F) = 0$, then problem* (11.53) *has the general solution*

$$q(t) = \text{const} \cdot \sin t + \int_0^\pi G_{\text{reg}}(t, \tau; 1) F(\tau)\, d\tau$$

with the regularized Green's function

$$G_{\text{reg}}(t, \tau; 1) := \lim_{\omega \to 1} \left(G(t, \tau; \omega) - \sqrt{\frac{2}{\pi}} \cdot \frac{\sin t \, \sin \tau}{\omega^2 - 1} \right).$$

The proof proceeds similarly to the proof of Prop. 11.21. The expression

$$-\sqrt{\frac{2}{\pi}} \cdot \frac{\sin t \, \sin \tau}{\omega^2 - 1}$$

is called regularizing counterterm. Explicitly,

$$G_{\text{reg}}(t,\tau;1) = \sqrt{\frac{2}{\pi}} \cdot \sum_{k=2}^{\infty} \frac{\sin kt \, \sin k\tau}{1-k^2} \, .$$

Note that the singular part of the Green's function G disappears.

The method of pseudo-resolvents in the resonance case. Let $\omega = 1$. Consider the modified problem

$$\boxed{\ddot{q}(t) + q(t) + b_1(q) \sin t = F(t), \quad q(0) = q(\pi) = 0.} \tag{11.60}$$

If q is a solution, then

$$b_1(q) = b_1(F), \quad b_k(q) = \frac{b_k(F)}{1-k^2}, \qquad k = 2, 3, \ldots, \tag{11.61}$$

by (11.56).

Proposition 11.23 *For given $F \in C_\pi^-$, problem* (11.60) *has the unique solution $q \in C_\pi^-$ given by*

$$q(t) = \int_0^\pi G_{\text{pseudo}}(t,\tau;1) F(\tau) \, d\tau, \qquad t \in [0,\pi]$$

along with the pseudo-Green's function

$$G_{\text{pseudo}}(t,\tau;1) := \sqrt{\frac{2}{\pi}} \left(\sin t \, \sin \tau + \sum_{k=2}^{\infty} \frac{\sin kt \, \sin k\tau}{1-k^2} \right).$$

If $b_1(F) = 0$, then q is a special solution of the original resonance problem

$$\ddot{q}(t) + q(t) = F(t), \quad q(0) = q(\pi) = 0.$$

The proof proceeds as for the non-resonance case above, by using Fourier series. Note that if $b_1(F) = 0$, then $\int_0^\pi F(\tau) \sin \tau \, d\tau = 0$. Hence

$$\int_0^\pi G_{\text{pseudo}}(t,\tau;1) F(\tau) \, d\tau = \int_0^\pi G_{\text{reg}}(t,\tau;1) F(\tau) \, d\tau.$$

The following investigations should help the reader to understand the phenomenon of renormalization for nonlinear oscillations in the resonance case. We will use the methods of bifurcation theory. Let us study the nonlinear oscillation problem

$$\boxed{\ddot{q}(t) + \omega^2 q(t) = \kappa \, q(t)^3 + \mu \sin t, \quad q(0) = q(\pi) = 0.} \tag{11.62}$$

The term $\mu \sin t$ describes an external periodic force, whereas the nonlinear term $\kappa q(t)^3$ models self-interaction. We call κ the coupling constant.

11.5.3 The Nonlinear Problem and Non-Resonance

Non-resonance case. Let $\omega > 0$ where $\omega \neq k$ for all $k = 1, 2, \ldots$. There exist positive numbers κ_0, μ_0 and r_0 such that the following hold true.

Proposition 11.24 *For given real parameters κ and μ with $|\kappa| \leq \kappa_0$ and $|\mu| \leq \mu_0$, problem* (11.62) *has a unique solution $q \in C_\pi^-$ with $\|q\|_\pi \leq r_0$.*

Explicitly,

$$q(t) = \frac{\mu \sin t}{\omega^2 - 1} + \frac{\mu^3 \kappa}{4(\omega^2 - 1)^3} \left(\frac{3 \sin t}{\omega^2 - 1} - \frac{\sin 3t}{\omega^2 - 9} \right) + \dots$$

The dots stand for higher-order terms with respect to κ and μ.

Proof. By Prop. 11.21, our problem is equivalent to the integral equation

$$q(t) = \int_0^\pi G(t, \tau; \omega) \big(\kappa\, q(\tau)^3 + \mu \sin \tau \big) d\tau.$$

This equation can be solved by means of the iterative method

$$q_{n+1}(t) = \int_0^\pi G(t, \tau; \omega) \big(\kappa\, q_n(\tau)^3 + \mu \sin \tau \big)\, d\tau, \qquad n = 0, 1, 2, \dots$$

with $q_0(\tau) \equiv 0$. This corresponds to the following boundary-value problem

$$\ddot{q}_{n+1} + \omega^2 q_{n+1} = F_n, \quad q_{n+1}(0) = q_{n+1}(\pi) = 0$$

with $F_n(t) := \kappa q_n(t)^3 + \mu \sin t$. To begin with, let $F_0(t) = \mu \sin t$. By (11.55),

$$q_1(t) = \frac{\mu \sin t}{\omega^2 - 1}\,.$$

Furthermore, $F_1(t) = \kappa q_1(t)^3 + \mu \sin t$. Hence

$$F_1(t) = \frac{\kappa \mu^3}{4(\omega^2 - 1)^3} (3 \sin t - \sin 3t) + \mu \sin t.$$

By (11.55),

$$q_2(t) = q_1(t) + \frac{\kappa \mu^3}{4(\omega^2 - 1)^3} \left(\frac{3 \sin t}{\omega^2 - 1} - \frac{\sin 3t}{\omega^2 - 9} \right).$$

As $n \to \infty$, this iterative method converges in the Banach space C^-_π, according to the Banach fixed-point theorem (see Zeidler (1995), Vol. 1, Sect. 1.7). □

11.5.4 The Nonlinear Problem, Resonance, and Bifurcation

The resonance case, bifurcation, and renormalization. The situation changes substantially if we consider the resonance case $\omega = 1$. To simplify notation, let us set $\mu = 0$, that is, we only study the crucial nonlinear self-interaction. The point is that in contrast to the non-resonance case, the linearized problem

$$\boxed{\ddot{q}(t) + q(t) = 0, \quad q(0) = q(\pi) = 0}$$

has the nontrivial solution

$$q(t) = s \cdot \sin t$$

which we call ground state. Here, s is a real parameter. If we switch on the self-interaction, then we have to study the perturbed solutions

$$q(t) = s \cdot \sin t + h(t) \tag{11.63}$$

of the boundary-value problem

$$\boxed{\ddot{q} + \omega^2 q = \kappa q^3, \quad q(0) = q(\pi) = 0} \tag{11.64}$$

along with

$$\omega = 1 + \delta\omega.$$

Here, both the function h and the real parameter $\delta\omega$ depend on the small parameters s and κ. In terms of physics, the observed angular frequency ω depends not only on the coupling constant κ, but also on the parameter s of the actual ground state $q = s \cdot \sin t + \ldots$ of the system. This serves as a very simple model for the situation encountered in quantum field theory where physicists argue as follows.

- In an experiment, we measure both the coupling constant κ and the characteristic quantity $\omega = 1 + \delta\omega$ (e.g., the effective mass and the effective electric charge of an electron).
- The knowledge of κ and ω determines the parameter s. This way, we get the ground state (11.63).
- Physicists call $\omega = 1 + \delta\omega$ the renormalization of the bare quantity $\omega = 1$.

In terms of mathematics, we get a family of solutions q, $1 + \delta\omega$ depending on the small coupling constant κ and an additional small parameter s. In fact, there exist positive numbers $\kappa_0, s_0, \varepsilon_0$ and r_0 such that the following hold true.

Proposition 11.25 *For given real parameters κ and s with $|\kappa| \le \kappa_0$ and $|s| \le s_0$, there exists a unique solution $q \in C^-_\pi$, $\omega \in \mathbb{R}$ of* (11.64) *with*

$$\|q\|_\pi \le r_0, \quad b_1(q) = s,$$

and $|\omega^2 - 1| \le \varepsilon_0$.

Explicitly, the solution reads as

$$q(t) = s \cdot \sin t + \ldots, \quad \omega^2 = (1 + \delta\omega)^2 = 1 + \frac{3}{4}\kappa s^2 + \ldots$$

The dots stand for higher-order terms with respect to κ and s. If we know κ and s, then we approximately get the value $s = \pm\sqrt{\frac{4}{3\kappa}(\omega^2 - 1)}$ and the ground state

$$q(t) = \pm\sqrt{\frac{4(\omega^2 - 1)}{3\kappa}} \cdot \sin t, \qquad t \in \mathbb{R}.$$

The following proof of Prop. 11.25 is based on a special variant of the Lyapunov–Schmidt method in bifurcation theory which was invented around 1900 (see Zeidler (1986), Vol. 1).

Proof. Let $\omega^2 = 1 + \varepsilon$. Motivated by the resonance case (11.60), we pass to the equivalent problem

$$\boxed{\ddot{q} + q + b_1(q)\sin t = \kappa q^3 - \varepsilon q + s \cdot \sin t, \quad q(0) = q(\pi) = 0} \tag{11.65}$$

along with the additional equation

$$\boxed{b_1(q) = s}$$

which is called the bifurcation equation in mathematics. First regard ε as a free parameter. To solve (11.65), we will use the iterative method

$$\ddot{q}_{n+1} + q_{n+1} + b_1(q_{n+1}) \sin t = F_n, \quad q_{n+1}(0) = q_{n+1}(\pi) = 0$$

for $n = 0, 1, \ldots$ with

$$F_n(t) := \kappa q_n(t)^3 - \varepsilon q_n(t) + s \cdot \sin t,$$

and $q_0 \equiv 0$. By $F_0(t) = s \cdot \sin t$ and (11.61),

$$q_1(t) = s \cdot \sin t.$$

Furthermore, since $F_1(t) = \frac{1}{4}\kappa s^3(3 \sin t - \sin 3t) - \varepsilon s \cdot \sin t + s \cdot \sin t$,

$$q_2(t) = \frac{1}{4}\kappa s^3 \left(3 \sin t + \frac{1}{8} \sin 3t\right) - \varepsilon s \cdot \sin t + s \cdot \sin t.$$

By the Banach fixed-point theorem, the iterative method converges to the function $q = q_1 + \ldots$. Observe the crucial fact that each term of q contains the parameter s as a factor. Therefore, the bifurcation equation $b_1(q) = s$ yields

$$\frac{3}{4}\kappa s^3 - \varepsilon s + s + \ldots = s.$$

This fixes the parameter ε, namely,

$$\varepsilon = \frac{3}{4}\kappa s^2 + O(s^3), \quad s \to 0.$$

Hence $q(t) = s \cdot \sin t + O(s^2), s \to 0$. □

11.5.5 The Importance of the Renormalized Green's Function

Let $\kappa \neq 0$ and $s \neq 0$, and suppose that κ and s are sufficiently small. The solution $q = q(t; \kappa, s)$ from Prop. 11.25 satisfies the boundary-value problem

$$\ddot{q} + \omega(\kappa, s)^2 q = \kappa q^3, \quad q(0) = q(\pi) = 0.$$

By Prop. 11.21, this is equivalent to the integral equation

$$\boxed{q(t; \kappa, s) = \int_0^\pi G_{\text{ren}}(t, \tau; 1)\, \kappa\, q(\tau; \kappa, s)^3 d\tau}$$

where $G_{\text{ren}}(t, \tau; 1) := G\big(t, \tau; \omega(\kappa, s)\big)$ is called the renormalized Green's function. Explicitly,

$$G_{\text{ren}}(t, \tau; 1) = \sqrt{\frac{2}{\pi}} \cdot \sum_{k=1}^{\infty} \frac{\sin kt \sin k\tau}{\omega(\kappa, s)^2 - k^2}, \qquad s \neq 0, \kappa \neq 0$$

along with

$$\omega(\kappa, s) = 1 + \tfrac{3}{4}\kappa s^2 + \ldots$$

The function $\omega = \omega(\kappa, s)$ is an absolutely convergent power series with respect to the small parameters κ and s. The well-defined renormalized Green's function G_{ren} differs from the ill-defined naive Green's function

$$G(t, \tau; \omega = 1) = \sqrt{\frac{2}{\pi}} \cdot \sum_{k=1}^{\infty} \frac{\sin kt \sin k\tau}{\omega^2 - k^2} \tag{11.66}$$

with the singularity at $k = 1$ because of $\omega = 1$. The passage from G to $G_{\rm ren}$ corresponds to the passage

$$1^2 - k^2 \Rightarrow \omega(\kappa, s)^2 - k^2.$$

We say that we add counterterms in order to force well-defined expressions.

Roughly speaking, the ill-defined Green's function G from (11.66) *is obtained by linearizing at the wrong angular frequency $\omega = 1$.*

Our rigorous approach linearizes at the perturbed angular frequency $1 + \delta\omega$ and we replace the ill-defined Green's function G at $\omega = 1$ by the pseudo-Green's function.

11.5.6 The Renormalization Group

In quantum field theory, renormalized quantities depend on the energy E. The method of the renormalization group studies the dependence of physical quantities on the positive parameter E (or on other characteristic parameters). Let us explain the basic idea by considering the following simple model. Consider the boundary-value problem

$$\frac{d^2Q(\tau)}{d\tau^2} + \omega_T^2 Q(\tau) = \kappa_T Q(\tau)^3, \quad Q(0) = Q(T) = 0. \tag{11.67}$$

Using the rescalings $\tau := \pi t/T$ and $Q(\tau) := q(t)$, we get

$$\frac{d^2q(t)}{dt^2} + \omega^2 q(t) = \kappa\, q(t)^3, \quad q(0) = q(\pi) = 0$$

where

$$\boxed{\omega_T = \frac{\pi\omega}{T}, \quad \kappa_T = \frac{\pi^2\kappa}{T^2}\,.}$$

Introducing $w(T) := \frac{\omega_T}{\pi\omega} = \frac{1}{T}$ and $k(T) := \frac{\kappa_T}{\pi^2\kappa} = \frac{1}{T^2}$, we get the renormalization group equation

$$\boxed{w(T_1T_2) = w(T_1)w(T_2), \quad k(T_1T_2) = k(T_1)k(T_2)}$$

for all positive numbers T_1 and T_2. This is a representation of the multiplicative group $\mathbb{R}_+^\times$ of positive real numbers. Differentiation yields the renormalization-group differential equation

$$\frac{dw(T)}{dT} = -\frac{w(T)}{T}, \qquad \frac{dk(T)}{dT} = -\frac{2k(T)}{T}$$

along with the initial condition $w(1) = 1$ and $k(1) = 1$. Setting $\xi := \ln T$, we obtain the dynamical system

$$\frac{dw}{d\xi} = -w, \qquad \frac{dk}{d\xi} = -2k$$

with the solution $w = \mathrm{e}^{-\xi}$, $k = \mathrm{e}^{-2\xi}$.

11.6 The Importance of Algebraic Feynman Integrals

At the beginning of the 19th century, Abel (1802–1829) studied algebraic integrals of the form

$$\mathcal{J} = \int R(z, w(z))\, dz$$

where R is a rational function of the two complex variables z and w, and the function $w = w(z)$ satisfies an algebraic equation

$$P(z, w(z)) = 0, \qquad z \in \mathbb{C},$$

that is, $P(z, w)$ denotes a polynomial with respect to z and w with complex coefficients. If $P(w, z) := w^2 - (z - e_1)(z - e_2)(z - e_3)$ with pairwise different complex numbers e_1, e_2, e_3, then $\mathcal{J}$ is an elliptic integral. For example,

$$\mathcal{J} = \int \frac{dz}{\sqrt{(z - e_1)(z - e_2)(z - e_3)}}\,.$$

Riemann (1826–1866) showed that algebraic integrals can be understood best by studying the topology of the corresponding Riemann surface. In the 1940s, Feynman systematically used higher-dimensional algebraic integrals.

11.6.1 Wick Rotation and Cut-Off

As a prototype, let us consider the algebraic Feynman integral

$$\boxed{J(m) := \int_{-\infty}^{\infty} dE \int_{\mathbb{R}^3} \frac{d^3\mathbf{p}}{(E^2 - c^2\mathbf{p}^2 - m^2c^4)^2}.}$$

When computing cross sections for scattering processes of elementary particles via the Feynman diagram technique, one encounters integrals of such type. These integrals correspond to the internal lines of the Feynman diagrams. Here, we use the following notation: E energy, $\mathbf{p}$ momentum vector, and $m > 0$ bare mass of the elementary particle. Note that the integral $J(m)$ is divergent for the following two reasons.

- First, the integrand has a singularity if

  $$E^2 = c^2\mathbf{p}^2 + m^2c^4.$$

  This is a hyperboloid in $\mathbb{R}^4$ called the mass shell with respect to the mass parameter m. From the physical point of view, the energy E and momentum vector $\mathbf{p}$ of an elementary particle of rest mass m live on the mass shell, by Einstein's energy-momentum relation for a free particle.
- Second, the integrand decreases too slowly as $|\mathbf{p}| \to \infty$ and $|E| \to \infty$.

The following definition is basic. To simplify notation, we replace the SI system by the energetic system, that is, we set $c = 1$. For given large cut-off mass $M_{\max} > 0$, we define the regularized form of the integral $J(m)$ by

$$\boxed{\operatorname{reg}_{M_{\max}} J(m) := 2\pi^2 \mathrm{i} \cdot \ln\left(\frac{M_{\max}}{m}\right).} \tag{11.68}$$

In terms of physics, this means that we only consider such physical situations, say, in a particle accelerator where energy E and momentum vector $\mathbf{p}$ satisfy the relation

$$E^2 = \mathbf{p}^2 + M^2$$

along with $0 \leq M \leq M_{\max}$. Under this mass restriction, we assign the value (11.68) to the integral $J(m)$ if $M_{\max}/m \gg 1$. Let us motivate this definition.

- Step 1: Wick rotation. To eliminate the indefinite metric of the Minkowski space, we first perform a Wick rotation. This means that we pass over to imaginary energies by replacing the real axis by the imaginary axis.To this end, we define

$$J_{\mathrm{Wick}}(m) := \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} dE \int_{\mathbb{R}^3} \frac{d^3\mathbf{p}}{(E^2 - \mathbf{p}^2 - m^2)^2}.$$

  Use now the substitution

$$E = \mathrm{i}q_0, \qquad \mathbf{p} := \mathbf{q}.$$

  Here, q_0 is a real parameter, and we set $q^2 = q_0^2 + \mathbf{q}^2$. We then get

$$J_{\mathrm{Wick}}(m) := \mathrm{i}\int_{-\infty}^{\infty} dq^0 \int_{\mathbb{R}^3} \frac{d^3\mathbf{q}}{(q_0^2 + \mathbf{q}^2 + m^2)^2} = \mathrm{i}\int_{\mathbb{R}^4} \frac{d^4q}{(q^2 + m^2)^2}.$$

  This integral is infinite.
- Step 2: Mass cut-off. Fix $M_{\max} > 0$. We define the cut-off by setting

$$J_{\mathrm{Wick}}(m, M_{\max}) := \mathrm{i}\int_{q^2 \leq M_{\max}^2} \frac{d^4q}{(q^2 + m^2)^2}\ .$$

  This is a well-defined classical integral which can be computed easily. In fact, introducing spherical coordinates,

$$J_{\mathrm{Wick}}(m, M_{\max}) = \mathrm{i}\int_{r=0}^{M_{\max}} \int_{\mathbb{S}^3} \frac{r^3 dr d\Omega}{(r^2 + m^2)^2}.$$

  Using the measure $\mathrm{meas}(\mathbb{S}^3) = 2\pi^2$ of the 3-dimensional unit sphere $\mathbb{S}^3$ and the rescaling $r = m\varrho$,

$$\begin{aligned} J_{\mathrm{Wick}}(m, M_{\max}) &= 2\pi^2\mathrm{i}\int_0^{\frac{M_{\max}}{m}} \frac{\varrho(\varrho^2+1) - \varrho}{(\varrho^2+1)^2}\, d\varrho \\ &= \pi^2\mathrm{i}\left(\ln(\tfrac{M_{\max}^2}{m^2} + 1) + \frac{1}{\frac{M_{\max}^2}{m^2} + 1} - 1\right). \end{aligned}$$

  Hence

$$J_{\mathrm{Wick}}(m, M_{\max}) = \pi^2\mathrm{i}(1 + o(1)) \ln \tfrac{M_{\max}^2}{m^2}, \qquad \tfrac{M_{\max}}{m} \to \infty.$$

Finally, we obtain the desired expression (11.68) by neglecting the term $o(1)$ if $M_{\max}/m \gg 1$.

11.6.2 Dimensional Regularization

For given mass parameters $m > 0$ and $M_{\max} > 0$, we want to show that the expression (11.68) can also be obtained by a completely different approach called dimensional regularization. Nowadays physicists prefer to use the method of dimensional regularization in renormalization theory, since this method does not destroy the gauge invariance.[35]

- Step 1: The classical integral. Let $q := (q_0, \ldots, q_{D-1})$. Moreover, introduce the sum of squares $q^2 := q_0^2 + \ldots + q_{D-1}^2$. We start with the classical key formula

$$\boxed{\int_{\mathbb{R}^D} \frac{d^D q}{(q^2 + m^2)^N} = \frac{\pi^{\frac{D}{2}} \, \Gamma\left(N - \frac{D}{2}\right)}{m^{2N-D} \, \Gamma(N)}} \tag{11.69}$$

  where $D = 1, 2, \ldots$ and $N > D/2$. Recall that Euler's gamma function $\Gamma = \Gamma(z)$ is holomorphic on the complex plane up to first-order poles at the singular points $z = 0, -1, -2, \ldots$ At the origin,

$$\boxed{\Gamma(z) = \frac{1}{z} - \mathcal{C} + O(z), \qquad z \to 0}$$

  where $\mathcal{C} := \lim_{n\to\infty}(\sum_{k=1}^{n} \frac{1}{k} - \ln(n+1)) = 0.5722\ldots$ is Euler's constant.[36] Generally, for $n = 1, 2, \ldots$ we have[37]

$$\Gamma(z - n) = \frac{(-1)^n}{n!} \left(\frac{1}{z} - \mathcal{C} + \sum_{k=1}^{n} \frac{1}{k} \right) + O(z), \quad z \to 0. \tag{11.70}$$

  Furthermore, $\Gamma(1) = \Gamma(2) = 1$. In general

$$\Gamma(n) := (n-1)!, \qquad n = 1, 2, 3, \ldots$$

  In formula (11.69), the pole of the function $N \mapsto \Gamma(N - D/2)$ at the point $N = D/2$ indicates that the integral is divergent in this situation. For example, this is the case if $D = 4$ and $N = 2$.
- Step 2: Analytic continuation: Note that

  The basic trick of dimensional regularization is to apply analytic continuation to formula (11.69) *with respect to the dimension* D.

  This way, we define the left-hand integral from (11.69) by the right-hand side of (11.69) for all complex numbers D with

$$N - \frac{D}{2} \neq 0, -1, -2, \ldots$$

  For example, choose $D := 4 - \delta$ and $N = 2$ where $0 < \delta < 1$. This yields

[35] Much material can be found in 't Hooft and Veltman (1973), Veltman (1995). See also Peskin and Schroeder (1995), Kugo (1997), Ryder (1999), and Zinn-Justin (2004).

[36] The experience of physicists shows that this constant is cancelled when computing quantities that can be observed in physical experiments. A detailed investigation of Euler's constant can be found in Havil (2003).

[37] The proof can be found in Ryder (1999), p. 385.

$$\int_{\mathbb{R}^{4-\delta}} \frac{d^{4-\delta}q}{(q^2+m^2)^2} := \frac{\pi^{2-\delta/2}\Gamma\left(\frac{\delta}{2}\right)}{m^\delta \Gamma(2)}.$$

Using this, we define

$$\mathcal{J}_\delta(m) := \mathrm{i}\kappa^\delta \int_{\mathbb{R}^{4-\delta}} \frac{d^{4-\delta}q}{(q^2+m^2)^2} = \mathrm{i}\pi^2 \Gamma\left(\frac{\delta}{2}\right)\left(\frac{m\pi^{1/2}}{\kappa}\right)^{-\delta}.$$

Here, the positive parameter κ has the physical dimension of energy; κ has been introduced in order to get the dimensionless quotient m/κ. Let $x > 0$. As $\delta \to +0$, it follows from

$$\Gamma\left(\frac{\delta}{2}\right) = \frac{2}{\delta} - \mathcal{C} + o(1)$$

and

$$x^{-\delta} = \mathrm{e}^{-\delta \ln x} = 1 - \delta \ln x + o(\delta)$$

that

$$\mathcal{J}_\delta(m) = \mathrm{i}\pi^2 \left(\frac{2}{\delta} - \mathcal{C} - 2\ln\left(\frac{m\pi^{1/2}}{\kappa}\right)\right) + o(1), \qquad \delta \to +0.$$

- Step 3: The Pauli–Villars regularization method: In order to cancel the singular term $\frac{1}{\delta}$, we consider the difference

$$\mathcal{J}_\delta(m) - \mathcal{J}_\delta(M_{\max}) = 2\pi^2 \mathrm{i} \cdot \ln\left(\frac{M_{\max}}{m}\right) + o(1), \qquad \delta \to +0$$

by introducing the large fictitious mass $M_{\max}$. Generally, the classic Pauli–Villars method introduces additional physical fields which contain large fictitious masses. In contrast to the cut-off method, Pauli–Villars regularization does not destroy the relativistic invariance.[38]
- Step 4: The limit: Letting $\delta \to +0$, we get

$$\lim_{\delta \to +0} (\mathcal{J}_\delta(m) - \mathcal{J}_\delta(M_{\max})) = 2\pi^2 \mathrm{i} \cdot \ln\left(\frac{M_{\max}}{m}\right).$$

This coincides with (11.68).

The same method can be applied to broad classes of algebraic Feynman integrals. This will be studied in Volume V on the physics of the Standard Model in particle physics.

The Euler beta function. Let $\alpha > 0$ and $\beta > 0$. In dimensional regularization, one frequently uses the following two classic integral formulas

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\mathrm{e}^{-x}\, dx \tag{11.71}$$

and

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{11.72}$$

where

$$B(\alpha,\beta) := \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx$$

is the Euler beta function. The proofs along with further important material can be found in Zorich (2003), p. 439ff.

[38] W. Pauli and F. Villars, On the invariant regularization in relativistic quantum theory, Rev. Mod. Phys. **21** (1949), 434–444.

11.6.3 Weinberg's Power-Counting Theorem

> An attack is made on the problem of determing the asymptotic behavior at high energies and momenta of the Green's function, using new mathematical methods from the theory of real variables. We define a class A_n of functions of n real variables whose asymptotic behavior may be specified in a certain manner by means of certain "asymptotic coefficients". The Feynman integrands of perturbation theory (with energies taken imaginary) belong to such classes. We then prove that if certain conditions on the asymptotic coefficients are satisfied then an integral over k of the variables converges, and belongs to the class A_{n-k} with new asymptotic coefficients simply related to the old ones. When applied to perturbation theory this theorem validates the renormalization procedure of Dyson and Salam, proving that the renormalized integrals actually do always converge, and provides a simple rule for calculating the asymptotic behavior of any Green's function to any order of perturbation theory.
>
> I particularly wish to thank Professor Wightman for his many valuable suggestions, and for enabling this paper to satisfy Salam's criterion.
>
> Steven Weinberg, 1960
> *High energy behavior in quantum field theory*[39]

The prototypes of convergent and divergent integrals are given by the following one-dimensional examples:

(i) Convergence for $N > 1$:

$$\int_1^{+\infty} \frac{dp}{p^N} = \lim_{P\to+\infty} \frac{1}{N-1}\left(1 - \frac{1}{P^{N-1}}\right) = \frac{1}{N-1}.$$

(ii) Logarithmic divergence:

$$\int_1^{+\infty} \frac{dp}{p} = \lim_{P\to+\infty} \ln P = +\infty.$$

(iii) Divergence for $0 < N < 1$:

$$\int_1^{+\infty} \frac{dp}{p^N} = \lim_{P\to+\infty} \frac{P^{1-N} - 1}{1-N} = +\infty.$$

Using spherical coordinates, this can be easily translated to higher dimensions.

Simple model. Consider the integral

$$\boxed{\Sigma(q) := \int_{-\infty}^{\infty} \frac{p\, dp}{(m^2 + p^2)(\mu^2 + (p-q)^2)}.} \tag{11.73}$$

This is a simplified one-dimensional version of a Feynman integral corresponding to the self-energy of a fermion (e.g., an electron) of mass $m > 0$ and a boson (e.g., a meson) of mass $\mu > 0$. The real variables p and q correspond to the momentum of the particles. Explicitly,

[39] Reprinted excerpt with permission from S. Weinberg, Phys. Rev. **118**(3) (1960), 838–849. Copyright 1960 by the American Physical Society. The reader should note that the meaning of Salam's criterion is explained on page 848.

$$\Sigma(q) = \frac{\pi q(q^2 + (m-\mu)^2)}{\mu(q^4 + 2q^2(m^2+\mu^2) + (m^2-\mu^2)^2)}.$$

Thus, the integral possesses the following asymptotics for large momenta:

$$\boxed{\Sigma(q) = O\left(\frac{1}{q}\right), \qquad q \to +\infty}$$

for all $q \in \mathbb{R}$. Weinberg's fundamental power-counting theorem generalizes this simple example to very general situations. This theorem tells us that under quite general assumptions, the asymptotics of the integrand implies both the existence and the asymptotics of the integral for large momenta. The full formulation of this theorem along with the sophisticated proof can be found in Volume V. The proof starts with the one-dimensional case by using a compactness argument. One then proceeds by induction. At this point, we will only consider a special case concerning the existence of multidimensional integrals with rational integrands.[40]

Critical subdivergences. To begin with, let us discuss some typical pitfalls. Let $N \geq 0$.

- Inspect the integral
$$J(a) := \int_{\mathbb{R}^2} \frac{dxdy}{(1 + x^2 + ay^2)^N}.$$
Suppose first that $a > 0$. In this favorite case, the integral is convergent if $N > 2$. For $N \leq 2$, the integral is divergent.
- The situation changes completely if $a \leq 0$. The integral is then divergent for all exponents $N \geq 0$. In fact, for $a = 0$, the divergence of the integral is simply a consequence of
$$\int_{-\infty}^{\infty} dy = \infty.$$
If $a < 0$, then the denominator has zeros along the hyperbola $|a|y^2 - x^2 = 1$ which is responsible for the divergence of the integral.
- The integral
$$\int_{\mathbb{R}^2} \frac{dxdy}{(1+x^2)(1+|y|^N)}$$
is convergent for $N > 1$, and it is divergent for $N < 1$. This follows from the fact that the integral
$$\int_{-\infty}^{\infty} \frac{dy}{1+|y|^N}$$
is convergent (resp. divergent) if $N > 1$ (resp. $N < 1$).

Observe that in these examples, the danger comes from divergent subintegrals extended over lower-dimensional domains of integration.

The power-counting method. Let us consider the following integral

$$\boxed{\Sigma(q) := \int_{\mathbb{R}^{4n}} \frac{\mathcal{P}(q,p)}{\Pi_{i=1}^N (\mathbf{s}_i^2 + m_i^2)}\, d^{4n}p} \tag{11.74}$$

for all $q \in \mathbb{R}^{4n}$ along with the subintegrals

[40] We recommend reading Weinberg (1960). Generalizations can be found in Hahn and Zimmermann (1968), Manoukian (1983), and Etingof (1999).

$$\int_{\mathcal{H}} \frac{\mathcal{P}(q,p)}{\Pi_{i=1}^{N}(\mathbf{s}_i^2 + m_i^2)}\, dV_{\mathcal{H}}. \tag{11.75}$$

Here, $\mathcal{P}$ is a polynomial with respect to the real components of the 4-vectors $\mathbf{p}_1, \ldots, \mathbf{p}_n \in \mathbb{R}^4$ and $\mathbf{q}_1, \ldots, \mathbf{q}_n \in \mathbb{R}^4$. Moreover, we set

$$\mathbf{p}_j^2 := p_{j0}^2 + p_{j1}^2 + p_{j2}^2 + p_{j3}^2.$$

Furthermore, we have

$$\mathbf{s}_i := \mathbf{q}_i + \sum_{j=1}^{n} a_{ij}\mathbf{p}_j, \qquad i = 1, .., N$$

where a_{ij} are real numbers. The mass coefficients $m_1, \ldots, m_N$ are assumed to be positive. Finally, $\mathcal{H}$ is an arbitrary plane in $\mathbb{R}^{4n}$ described by a linear system of the form

$$\sum_{j=1}^{n} b_{ij}\mathbf{p}_j = c_i, \qquad i = 1, \ldots, n$$

where b_{ij}, c_i are real numbers for all $i, j = 1, \ldots, n$. In addition, $dV_{\mathcal{H}}$ denotes the volume element of $\mathcal{H}$. Now to the point. The superficial degree of the integral $\Sigma(q)$ is defined to be

$$\begin{aligned} \deg_{\sup} \Sigma(q) : = & \text{ degree of the integrand} \\ & \text{plus dimension of the domain of integration.} \end{aligned}$$

By definition, the degree of the integrand is equal to "degree of the numerator minus degree of the denominator". For example, the integrand from (11.73) has the degree $1 - 4 = -3$. Therefore, the integral $\Sigma(q)$ from (11.73) has the superficial degree

$$\deg_{\sup} = -3 + 1 = -2.$$

A very useful special case of the Weinberg theorem reads as follows.

Theorem 11.26 *The integral* (11.74) *is absolutely convergent if its superficial degree is negative and the superficial degrees of all of the subintegrals* (11.75) *are also negative.*

11.6.4 Integration Tricks

We want to summarize prototypes of integration tricks which are frequently used by physicists in quantum field theory.

Cauchy's residue trick. Consider the rational function

$$f(x) := \frac{a(x)}{b(x)}, \qquad x \in \mathbb{R}$$

where a and b are polynomials of degree α and β, respectively, with complex coefficients. We assume that $\beta - \alpha \geq 2$, $\alpha \geq 0$, and $b(x) \neq 0$ for all $x \in \mathbb{R}$. Let $z_1, \ldots, z_K$ be precisely the poles of the function f on the upper half-plane. Then

$$\boxed{\int_{\mathbb{R}} f(x)dx = 2\pi \mathrm{i} \sum_{k=1}^{K} \mathrm{res}_{z_k}(f).} \tag{11.76}$$

This integral exists in the classical sense.[41] In addition, we have $\int_{\mathbb{R}} f(x)dx = 0$ if the function f has no poles on the upper-half plane. The proof proceeds as in Problem 12.1 on page 732.

Example. Let $m > 0$. Consider the integral

$$J(m) := \int_{\mathbb{R}} \frac{dE}{m^2 - E^2}.$$

This integral does not exist in the classical sense because of the singularity of the integrand at the points $E = \pm m$. It is our goal to describe two different methods which motivate the definition

$$\boxed{\int_{\mathbb{R}} \frac{dE}{m^2 - E^2} := \frac{\mathrm{i}\pi}{m}} \tag{11.77}$$

of the regularized value of the divergent integral.

(i) Mass perturbation. We start with the replacement

$$\boxed{m \Rightarrow m - \mu\mathrm{i},} \qquad \mu > 0 \tag{11.78}$$

where μ is small. In terms of physics, the choice of the sign of μ is motivated by the argument (14.14) discussed on page 773 (damped oscillation of a quantum particle of mass m). By definition, the regularized integral reads as

$$J(m - \mu\mathrm{i}) = \int_{\mathbb{R}} \frac{dE}{(m - \mu\mathrm{i})^2 - E^2}, \qquad \mu > 0.$$

The decomposition

$$\frac{1}{(m - \mu\mathrm{i})^2 - E^2} = \frac{1}{2(m - \mu\mathrm{i})} \left(\frac{1}{m - \mu\mathrm{i} - E} + \frac{1}{m - \mu\mathrm{i} + E} \right)$$

tells us that the function$E \mapsto \frac{1}{(m-\mu\mathrm{i})^2 - E^2}$ has a pole at the point $-m + \mu\mathrm{i}$ in the upper half-plane. By (11.76),

$$J(m - \mu\mathrm{i}) = \frac{\mathrm{i}\pi}{m - \mu\mathrm{i}}.$$

The limit

$$\lim_{\mu \to +0} J(m - \mu\mathrm{i}) = \frac{\mathrm{i}\pi}{m}$$

coincides with (11.77). The same argument shows that the modified regularization

$$J_\varepsilon(m) := \int_{\mathbb{R}} \frac{dE}{m^2 - \varepsilon\mathrm{i} - E^2}, \qquad \varepsilon > 0$$

yields the same value, namely,

$$\lim_{\varepsilon \to +0} J_\varepsilon(m) = \lim_{\varepsilon \to +0} \frac{\mathrm{i}\pi}{\sqrt{m^2 - \varepsilon\mathrm{i}}} = \frac{\mathrm{i}\pi}{m}.$$

[41] The definition of the residue $\mathrm{res}_{z_k}(f)$ of the function f at the point z_k can be found on page 213.

(ii) Wick rotation. Replacing the real energy axis by the imaginary energy axis, we get

$$J_{\text{Wick}}(m) := \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{dE}{m^2 - E^2}.$$

Using the substitution $E = \mathrm{i}q$,

$$J_{\text{Wick}}(m) = \mathrm{i}\int_{-\infty}^{\infty} \frac{dq}{m^2 + q^2}.$$

Since $\frac{1}{q^2+m^2} = \frac{1}{2m\mathrm{i}}(\frac{1}{q-m\mathrm{i}} - \frac{1}{q+m\mathrm{i}})$, the residue method (11.76) yields

$$J_{\text{Wick}}(m) = \frac{\mathrm{i}\pi}{m}$$

which coincides with (11.77).

The differentiation trick. Let $m > 0$. We want to discuss an iterative method in order to regularize divergent integrals of the type

$$J_N(m) := \int_{\mathbb{R}} \frac{dE}{(m^2 - E^2)^N}, \qquad N = 1, 2, \ldots.$$

We already know the definition $J_1(m) := \mathrm{i}\pi/m$. Formal differentiation with respect to the mass parameter m yields

$$J_N'(m) = -2mN \int_{\mathbb{R}} \frac{dE}{(m^2 - E^2)^{N+1}} = -2mN J_{N+1}(m).$$

This motivates the recursive definition

$$\boxed{J_{N+1}(m) :== -\frac{1}{2Nm}\,\frac{dJ_N(m)}{dm}, \qquad N = 1, 2, \ldots}$$

with $J_1(m) := \mathrm{i}\pi/m$. For example, $J_2(m) = \mathrm{i}\pi/2m^3$.

Feynman's product trick. Let $0 < a < b$. Then

$$\frac{1}{ab} = \frac{1}{b-a}\int_a^b \frac{dt}{t^2}.$$

Hence

$$\boxed{\frac{1}{ab} = \int_0^1 \frac{du}{[a + (b-a)u]^2}.} \tag{11.79}$$

For example,

$$J := \int_{\mathbb{R}^2} \frac{dxdy}{(1+x^2)(1+y^2)} = \left(\int_{\mathbb{R}} \frac{dx}{1+x^2}\right)^2 = \pi^2.$$

By (11.79),

$$J = \int_0^1 du \int_{\mathbb{R}^2} \frac{dxdy}{[1 + x^2 + (y^2 - x^2)u]^2.}.$$

Similarly, for $j, k = 1, 2, \ldots,$

$$\frac{1}{a^j b^k} = \frac{(j+k-1)!}{(j-1)!(k-1)!} \int_0^1 \frac{x^{j-1}(1-x)^{k-1}}{[ax+b(1-x)]^{j+k}}\, dx.$$

Schwinger's parametrization trick. Let $a > 0$. Then

$$\boxed{\frac{1}{a} = \int_0^\infty \mathrm{e}^{-at} dt.} \tag{11.80}$$

Here, t is called the Schwinger parameter by physicists. This identity allows us to reduce algebraic Feynman integrals to integrals of exponential type. For example,

$$\pi = \int_{\mathbb{R}} \frac{dx}{1+x^2} = \int_{\mathbb{R}} dx \int_0^\infty dt\, \mathrm{e}^{-(1+x^2)t}.$$

Example. Set $p := (p_0, \ldots, p_{D-1})$ and $p^2 = p_0^2 + \ldots + p_{D-1}^2$. For $p, q \in \mathbb{R}^D$ with $p \neq 0$, we want to compute the integral

$$J_D(p) := \int_{\mathbb{R}^D} \frac{d^D q}{q^2 (p+q)^2}.$$

To begin with, let $D = 3$. By Schwinger's trick,

$$\begin{aligned} J_D(p) &= \int_0^\infty dt \int_0^\infty d\tau \int_{\mathbb{R}^D} d^D q\, \mathrm{e}^{-tq^2 - \tau(p+q)^2} \\ &= \int_0^\infty dt \int_0^\infty d\tau \int_{\mathbb{R}^D} d^D q \exp\left[-(t+\tau)\left(q + \frac{\tau p}{t+\tau} \right)^2 - \frac{t\tau p^2}{t+\tau} \right]. \end{aligned}$$

Computing the Gaussian integral with respect to the variable q,

$$J_D(p) = \pi^{D/2} \int_0^\infty dt \int_0^\infty (t+\tau)^{-D/2} \cdot \exp\left(-\frac{t\tau p^2}{t+\tau} \right) d\tau.$$

The Jacobian of the transformation $t = sx,\ \tau = (1-x)s$ reads as

$$\frac{\partial(t,\tau)}{\partial(x,s)} = \frac{\partial t}{\partial x}\frac{\partial \tau}{\partial s} - \frac{\partial t}{\partial s}\frac{\partial \tau}{\partial x} = s(1-x) + xs = s.$$

Hence

$$J_D(p) = \pi^{D/2} \int_0^1 dx \int_0^\infty ds\, s^{1-D/2} \mathrm{e}^{-sx(1-x)p^2}.$$

For all numbers $\alpha > 0$, $\Gamma(\alpha) = \int_0^\infty s^{\alpha-1} \mathrm{e}^{-s}\, ds$. This implies

$$J_D(p) = \pi^{D/2} (p^2)^{-2+D/2}\, \Gamma\left(2 - \frac{D}{2}\right) \int_0^1 [x(1-x)]^{-2+D/2}\, dx.$$

Finally, using the Euler beta function and its relation to the gamma function (11.72) on page 637, we get

$$J_D(p) = \pi^{D/2} (p^2)^{-2+D/2} \cdot \frac{\Gamma\left(2 - \frac{D}{2}\right) \Gamma\left(\frac{D}{2} - 1\right)^2}{\Gamma(D-2)}.$$

Analytic continuation shows that the right-hand side is finite if

$$D > 0 \qquad \text{and} \qquad D \neq 2, 4, 6, \ldots .$$

This corresponds to the dimensional regularization of the integral $J_D(p)$. The asymptotic expansion of the function $J_D(p)$ at the critical values $D = 2, 4, \ldots$ follows from (11.70) on page 636.

The measure of the n-dimensional unit sphere. We have

$$\operatorname{meas}(\mathbb{S}^n) = \frac{2\pi^{(n+1)/2}}{\Gamma\left(\frac{n+1}{2}\right)}, \qquad n = 1, 2, \ldots .$$

For $n = 1, 2, 3$, we get $\operatorname{meas}(\mathbb{S}^n) = 2\pi, 4\pi, 2\pi^2$, respectively. This surface measure is useful when passing to spherical coordinates in $\mathbb{R}^{n+1}$. In fact,

$$\int_{\mathbb{R}^{n+1}} f(||x||)\, d^{n+1}x = \operatorname{meas}(\mathbb{S}^n) \int_0^\infty f(r) r^n dr.$$

Suggested reading. Lists of integrals useful for computations in quantum field theory can be found in Macke 1959, Vol. 6 (classical Feynman integration tricks), 't Hooft and Veltman (1973), Bogoliubov and Shirkov (1983), and Veltman (1995) (dimensional regularization). For computing algebraic Feynman integrals by using computers, we refer to the list of software systems on page 930.

We also recommend Itzykson and Zuber (1980), Collins (1984), Peskin and Schroeder (1995), Kugo (1997), Ryder (1999), and Zinn-Justin (2004). The relation between algebraic Feynman integrals and both the theory of distributions and Green's functions for hyperbolic differential equations (the classical theory of Hadamard and Marcel Riesz) are thoroughly studied in Gelfand and Shilov (1964), Vol. 1, Chap. III, and Egorov, Komech, and Shubin (1999).

The connection between algebraic Feynman integrals and topology in terms of the Picard–Lefschetz theorem can be found in Hwa and Teplitz (1966). See also the hints for further reading given in the footnote on page 639.

For the classical background, we refer to Zorich (2003), Vol. 2. An extensive collection of classical integrals and infinite series can be found in Prudnikov et al. (1986), Vols. 1–5.

11.7 Fundamental Solutions of Differential Equations

Let $N = 1, 2, \ldots$ We want to solve the differential equation

$$\boxed{L\psi = f \qquad \text{on } \mathbb{R}^N} \tag{11.81}$$

in the sense of distributions. Here, L is a differential operator with constant complex coefficients of order m,

$$L := \sum_{|\alpha| \leq m} a_\alpha \partial^\alpha$$

where each a_α is a complex number. Naturally enough, we assume that L is not trivial, that is, there exists at least one nonzero coefficient a_α. We are given the distribution $f \in \mathcal{D}'(\mathbb{R}^N)$ with compact support. We are looking for a distribution $\psi \in \mathcal{D}'(\mathbb{R}^N)$. Motivated by the idea of the Green's function, we want to reduce the solution of (11.81) to the equation

$$LG = \delta \qquad \text{on } \mathbb{R}^N. \tag{11.82}$$

Each solution $G \in \mathcal{D}'(\mathbb{R}^N)$ of (11.82) is called a fundamental solution of the differential operator L. Green's functions are special fundamental solutions which satisfy additional side conditions (e.g., initial conditions and boundary conditions).

Theorem 11.27 *If G is a fundamental solution of the differential operator L, then the convolution $G * f$ is a solution of* (11.81).

Proof. $L(G * f) = LG * f = \delta * f = f$. □

For many important differential equations in mathematical physics, fundamental solutions are explicitly known. A comprehensive summary can be found in Ortner and Wagner (1997).

(i) For given distribution $f \in \mathcal{D}'(\mathbb{R})$, consider the equation

$$\boxed{\psi' = f \qquad \text{on } \mathbb{R}.} \tag{11.83}$$

We are looking for a solution $\psi \in \mathcal{D}'(\mathbb{R})$. The fundamental solutions of (11.83) have the form

$$G(t) = \theta(t) + \text{const} \qquad \text{for all} \quad t \in \mathbb{R}.$$

This means that $G' = \delta$. The general solution of (11.83) reads as

$$\psi = \theta * f + \text{const}.$$

In particular, if $f : \mathbb{R} \to \mathbb{C}$ is continuous and has compact support, then

$$\psi(t) = \int_{-\infty}^{\infty} \theta(t-\tau) f(\tau) d\tau + \text{const} = \int_{-\infty}^{t} f(\tau) d\tau + \text{const}, \quad t \in \mathbb{R}.$$

(ii) Suppose we are given the distribution f as in (i) above. We are looking for a distribution $\psi \in \mathcal{D}'(\mathbb{R})$ such that

$$\boxed{\psi'' = f \qquad \text{on } \mathbb{R}.} \tag{11.84}$$

The fundamental solutions of (11.83) have the form

$$G(t) = t\theta(t) + \text{const} \cdot t + \text{const} \qquad \text{for all} \quad t \in \mathbb{R}.$$

This means that $G'' = \delta$. The general solution of (11.83) reads as

$$\psi = (g\theta) * f + \text{const} \cdot g + \text{const}$$

where $g(t) := t$. In particular, if $f : \mathbb{R} \to \mathbb{C}$ is continuous and has compact support, then

$$\psi(t) = \int_{-\infty}^{t} (t-\tau) f(\tau) d\tau + \text{const} \cdot t + \text{const}, \quad t \in \mathbb{R}.$$

Proof. Ad (i). By (11.42) on page 611, each solution $\psi \in \mathcal{D}'(\mathbb{R})$ of the homogeneous equation $\psi' = 0$ is constant. Thus, two solutions of (11.83) differ by a constant. Finally, $\theta' = \delta$.

Ad (ii). By the Weyl lemma on page 611, each solution $\psi \in \mathcal{D}'(\mathbb{R})$ of $\psi'' = 0$ is a smooth function. Hence $\psi(t) = \text{const} \cdot t + \text{const}$. Moreover, we have shown in (11.41) on page 611 that $(g\theta)'' = \delta$. □

11.7.1 The Newtonian Potential

Consider the Poisson equation

$$\boxed{\varepsilon_0 \Delta V = \varrho \qquad \text{on } \mathbb{R}^3.} \tag{11.85}$$

We are given the distribution $\varrho \in \mathbb{R}^3$ with compact support, and we are looking for the distribution $V \in \mathcal{D}'(\mathbb{R}^3)$. In terms of physics, the electric charge density ϱ generates the electric field $\mathbf{E} = -\,\mathbf{grad}\, V$. The equation

$$\varepsilon_0 \Delta G = \delta \qquad \text{on } \mathbb{R}^3 \tag{11.86}$$

corresponds to a charge $Q = 1$ at the origin.

Proposition 11.28 *The general solution of* (11.85) *is given by*

$$V = G * \varrho + W$$

where $W : \mathbb{R}^3 \to \mathbb{C}$ *is an arbitrary smooth function with* $\Delta W = 0$ *on* $\mathbb{R}^3$*, and*

$$G(\mathbf{x}) := \frac{1}{4\pi\varepsilon_0 ||x||}, \qquad \mathbf{x} \in \mathbb{R}^3 \setminus \{0\}$$

is a fundamental solution satisfying (11.86)*. If* $\varrho : \mathbb{R}^3 \to \mathbb{R}$ *is a continuous function with compact support, then*

$$(G * \varrho)(\mathbf{x}) = \int_{\mathbb{R}^3} \frac{\varrho(\mathbf{y})}{4\pi\varepsilon_0 ||\mathbf{x} - \mathbf{y}||}\, d^3\mathbf{y}, \qquad \mathbf{x} \in \mathbb{R}^3.$$

Proof. By the Weyl lemma on page 611, each distribution $V \in \mathcal{D}'(\mathbb{R}^3)$ with $\Delta V = 0$ is a smooth function. By Prop. 10.22,

$$\int_{\mathbb{R}^3} \varepsilon_0 G(\mathbf{x}) \Delta\varphi(\mathbf{x}) d^3\mathbf{x} = \varphi(\mathbf{0}) \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}^3).$$

This tells us that $\varepsilon_0 \Delta G = \delta$. □

11.7.2 The Existence Theorem

Theorem 11.29 *Each nontrivial linear differential operator with complex constant coefficients has a tempered distribution as fundamental solution.*

In 1958, this crucial theorem was proved by Hörmander in a sophisticated manner. An alternative proof was given by Łojasiewicz in 1959. Atiyah showed in 1970 that this theorem can be proved in an elegant way by using the deep Hironaka theorem. Let us discuss the main difficulty of the proof. By Fourier transform, the original equation $LG = \delta$ passes over to the equation

$$\boxed{P\hat{G} = 1, \qquad \hat{G} \in \mathcal{S}'(\mathbb{R}^n)} \tag{11.87}$$

where P is a polynomial with respect to the momentum variable $p \in \mathbb{R}^n$. Naively, we get

$$\hat{G} = \frac{1}{P}.$$

However, as a rule, this method does not work, since the zeros of P generate singularities. This complicates the proof substantially.

11.7.3 The Beauty of Hironaka's Theorem

> In this note I shall show how Hironaka's theorem[42] on the resolution of singularities leads quickly to a new proof of the Hörmander–Łojasiewicz theorem on the division of distributions (generalized functions) and hence to the existence of tempered fundamental solutions for constant-coefficient differential operators. Since most of the difficulties in the general theory of partial differential operators arise from the singularities of the characteristic variety, it is quite natural to expect Hironaka's theorem to be relevant. In fact, this note is primarily intended to draw the attention of analysts to the power of this theorem.
>
> Michael Atiyah, 1970[43]

In this section, we want to have a look at some deep mathematical theorems which are related to quantum field theory. At the top is the highly sophisticated Hironaka theorem. It turns out that

> *An appropriate mathematical tool for quantum field theory is the theory of tempered generalized functions based on the Fourier transform.*

In fact, the basic objects of Feynman's approach to quantum field theory, namely,

- the Dirac delta function,
- the Feynman propagators (e.g., the photon propagator and the electron propagator in quantum electrodynamics), and
- the algebraic Feynman integrals (corresponding to internal lines of Feynman diagrams)

are *not* classical mathematical functions, but tempered distributions. Let us mention the following fundamental results:

(i) the 1964 Hironaka theorem on the resolution of singularities in algebraic geometry;
(ii) the 1968 Hironaka–Atiyah–Bernstein–Gelfand (HABG) theorem on meromorphic families of tempered distributions;
(iii) propagators of quantum fields and the 1958 Hörmander–Łojasiewicz theorem on the existence of tempered fundamental solutions; this can be based on the HABG theorem;
(iv) dimensional regularization of Feynman integrals via the HABG theorem;
(v) the zero-mass limit as a tempered generalized function via the HABG theorem;
(vi) the multiplication of generalized functions by using Hörmander's wave front sets, and causal products of propagators (the 1973 Epstein–Glaser approach to quantum field theory via constructing the S-matrix in terms of tempered generalized functions);
(vii) Hörmander's 1971 theory of Fourier integral operators, microlocal analysis, and Radzikowski's 1996 theory for Hadamard states in quantum gravitation;
(viii) Analytic continuation of functions of many complex variables; analyticity properties of the Fourier transform or Laplace transform of tempered generalized functions (the Paley–Wiener theorem and Bogoliubov's "edge-of-the-wedge" theorem); proof of the fundamental CPT theorem in the setting of axiomatic quantum field theory by Res Jost in 1957.

[42] H. Hironaka, Resolution of singularities of an algebraic variety over a field of characteristic zero, Ann. Math. **79** (1964), 109–326. Heisuke Hironaka (born 1931) was awarded the Fields medal in 1970.

[43] Resolution of singularities and division of distributions, Commun. Pure Appl. Math. **23** (1970), 145–150 (reprinted with permission).

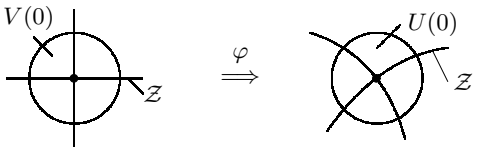


Fig. 11.4. Normal form of singularities

Mathematically, important contributions came from Laurent Schwartz (theory of generalized functions) in the 1940's and from Hörmander between 1955 and 1975 (general theory of linear partial differential operators). See Hörmander (1983), Vols. 1–4.[44]

In terms of quantum field theory, during the 1950s important impacts came from Arthur Wightman in Princeton and Nikolai Bogoliubov in Moscow and from their numerous collaborators. In what follows, let us sketch some of the main ideas.

The Hironaka theorem on the resolution of singularities. We want to study the behavior of the equation

$$\boxed{F(x) = 0, \qquad x \in \mathbb{R}^n} \tag{11.88}$$

near the origin, $x = 0$. The simplest case is the equation

$$y = 0, \qquad (x, y) \in \mathbb{R}^2$$

where the solutions are the points of the x-axis. However, we also allow singularities. As a prototype, consider the equation

$$xy = 0, \qquad (x, y) \in \mathbb{R}^2.$$

Here, the solutions consist of the points of both the x-axis, $y = 0$, and the y-axis, $x = 0$. We want to generalize this. To this end, we assume that the nonzero function

$$F : U(0) \to \mathbb{R}$$

is real analytic in a neighborhood $U(0)$ of the origin.[45] It is our goal to transform the function F into the local normal form

$$\boxed{F(\varphi(z)) = f(z) z_1^{\alpha_1} z_2^{\alpha_2} \cdots z_n^{\alpha_n}, \qquad z \in V(0)} \tag{11.89}$$

on a neighborhood $V(0)$ of the origin in $\mathbb{R}^n$ (Fig. 11.4). For the real exponents, $\alpha_1 \geq 0, \ldots, \alpha_n \geq 0$. Furthermore,

$$f(z) \neq 0 \qquad \text{on} \quad V(0),$$

[44] Laurent Schwartz, Lars Hörmander, Michael Atiyah, and Heisuke Hironaka were awarded the Fields medal in 1950, 1962, 1966, and 1970, respectively.

[45] This means that the function F can be represented as an absolutely convergent power series expansion on $U(0)$,

$$F(x) = a_0 + a_1 x_1 + \ldots + a_n x_n + a_{12} x_1 x_2 + \ldots$$

with real coefficients $a_0, a_1, \ldots$

and the function $f : V(0) \to \mathbb{R}$ is real analytic. The original equation (11.88) then passes over to the equivalent equation

$$z_1^{\alpha_1} z_2^{\alpha_2} \cdots z_n^{\alpha_n} = 0, \qquad z \in V(0)$$

which can be solved easily. Hironaka's theorem (1964) tells us that

The local normal form (11.89) *can be obtained by a regular local coordinate transformation,* $x = \varphi(z)$.

More precisely, there exists a surjective, real analytic map

$$\varphi : V(0) \to U(0)$$

on a neighborhood $V(0)$ of the origin such that the following hold true:

- The map φ is proper, that is, the preimages of compact sets are again compact.
- Let Z denote the set of zeros of the original equation (11.88), that is,

$$Z := \{x \in U(0) : \quad F(x) = 0\}.$$

  Let $\mathcal{Z}$ be the preimage of the zero set Z under the coordinate transformation $x = \varphi(z)$, i.e.,

$$\mathcal{Z} := \{z \in V(0) : \quad F(\varphi(z)) = 0\}.$$

- The complementary map $\varphi : V(0) \setminus \mathcal{Z} \to U(0) \setminus Z$ is a real analytic diffeomorphism.[46]

This is a special case of general results due to Hironaka (1964), pp. 109–326; the local normal form quoted above is a consequence of the theorem on page 170 of Hironaka's paper applied to analytic local rings.

Note that Hironaka's theorem represents a far-reaching generalization of the implicit function theorem. This theorem corresponds to the special case where $F'(0) \neq 0$. Then, the normal form looks like

$$F(\varphi(z)) = f(z) z_1, \qquad z \in V(0)$$

with $f(z) \neq 0$ on the neighborhood $V(0)$. This means that the solutions of the equation $F(x) = 0$ are given by the smooth hypersurface $x = \varphi(z)$ near the origin.

The genesis of the implicit function theorem and the Hironaka theorem goes back to Newton (1643-1727) who developed the method of Newton polygons in order to study the zeros of polynomial equations in two variables. For example, the two local solution branches of the equation

$$2x^3 + 2x^2 y + x y^2 + x^2 y^3 + y^4 = 0$$

near the origin, $x = 0, y = 0$, read as

$$y = (-1 \pm \mathrm{i})x + o(x) \qquad \text{and} \qquad y = \pm \mathrm{i} x^{1/2} + o(x^{1/2}), \quad x \to 0.$$

The appearance of fractional powers is typical for "this type of problem". This means that besides power series expansions we also encounter Puiseux expansions as solutions. See Zeidler (1986), Vol. 1, p. 431. In quantum field theory, such expansions are called singular Landau expansions.

[46] Explicitly, the map $\varphi : V(0) \setminus \mathcal{Z} \to U(0) \setminus Z$ is bijective and real analytic, and the inverse map is also real analytic.

Further important contributions to the implicit function theorem were made by Lagrange (1736–1813), Cauchy (1789–1857), and Weierstrass (1815–1897) (the Weierstrass preparation theorem is a forerunner of Hironaka's theorem). A fairly elementary discussion of the Hironaka approach to the blowing-up of singularities in algebraic geometry can be found in the survey article by Hauser (2003).

Besides the Hironaka theorem, there exists another deep generalization of the regular implicit function theorem called the hard implicit function theorem (or the Nash–Moser theorem). This theorem allows applications to chaotic motions in celestial mechanics (KAM theory) and to embedding theorems in the theory of manifolds.[47]

The Hironaka–Atiyah–Bernstein–Gelfand theorem on meromorphic families of generalized functions. Let us study the integral

$$\boxed{F_s(\varphi) := \int_{H_\geq} f(x)^s \varphi(x) d^n x} \tag{11.90}$$

for all test functions $\varphi \in \mathcal{D}(\mathbb{R}^n)$ and fixed complex exponent s with $\Re(s) > 0$. Our goal is to extend the function $s \mapsto F_s(\varphi)$ analytically to the complex plane. We cannot expect that this extended function is holomorphic on the complex plane. For physics, however, it is sufficient, that the extended function is meromorphic on the complex plane. In fact, the theorem below tells us that this property is met. To be precise, we assume that the nonzero function

$$f : \mathbb{R}^n \to \mathbb{R}$$

is real analytic and nonnegative. We set $x = (t, x_1, \ldots, x_{n-1})$, and define the closed half-space

$$H_\geq := \{x \in \mathbb{R}^n : \ t \geq 0\}.$$

In applications to physics, t plays the role of time. Then, for each complex number s with $\Re(s) > 0$, F_s is a generalized function which lies in the Schwartz space $\mathcal{D}'(\mathbb{R}^n)$.

Theorem 11.30 *For each test function $\varphi \in \mathcal{D}(\mathbb{R}^n)$, the function*

$$s \mapsto F_s(\varphi)$$

can be extended to a meromorphic function on the complex plane.

This theorem can be generalized in the following way.

- We replace $\mathbb{R}^n$ by an n-dimensional, real analytic, paracompact, and connected manifold X.
- We replace the closed half-space $H_\geq$ by the set

$$H_\geq := \{x \in X : \ g_1(x) \geq 0, \ldots, g_p(x) \geq 0\}$$

  where $g_1, \ldots, g_p : X \to \mathbb{R}$ are real analytic functions.
- The nonzero real analytic function $f : X \to \mathbb{R}$ is nonnegative.

For further material, we refer to the lectures by Etinghoff (1999).

[47] As an introduction to the implicit function theorem and the Nash–Moser theorem including detailed historical comments, we recommend the monograph by Krantz and Parks (2002). We also recommend Nirenberg (2001), Scheck (2000), Vol. 1 (applications to mechanics), and Zeidler (1986), Vol. 1 (see Chap. 8 on bifurcation theory).

11.8 Functional Integrals

> Feynman's path integrals are both infinite-dimensional Gaussian integrals and continuous partition functions.
>
> Folklore

By Theorem 11.8 on page 587, the initial-value problem for the heat equation is completely determined by the knowledge of the heat kernel.

> *In terms of the heat equation, the Feynman functional integral approach represents the heat kernel as an infinite-dimensional integral.*

This fundamental idea can be generalized to the time-evolution of general physical systems. The Feynman approach can be viewed as a generalization of the classical Fourier method via Fourier series or Fourier integral. Physicists like functional integrals very much, since they allow elegant explicit computations based on the following crucial methods:

- approximation by finite-dimensional Gaussian integrals,
- the method of stationary phase,
- infinite-dimensional Gaussian integrals and the method of zeta function regularization, and
- (formal) analytic continuation.

From the mathematical point of view, functional integrals represent an extraordinarily useful mnemonic tool for conjecturing rigorous results. Unfortunately, the rigorous proofs are frequently missing or they have to be based on sophisticated methods. Let us sketch the main ideas for the heat kernel. We will restrict ourselves to the elegant formal language used by physicists which can be immediately translated

- to the Schrödinger equation in quantum mechanics by replacing real time by imaginary time, $t \Rightarrow \mathrm{i}t/\hbar$,
- and to quantum field theory.

11.8.1 The Feynman Path Integral for the Heat Equation

The heat kernel. Let $\kappa > 0$. For a given initial time s, consider the initial-value problem

$$\begin{aligned} \frac{\partial T}{\partial t}(\mathbf{x},t) &= -\kappa \Delta T(\mathbf{x},t) - U(\mathbf{x})T(\mathbf{x},t), && \mathbf{x} \in \mathbb{R}^3,\ t \geq s, \\ T(\mathbf{x},s) &= T_0(\mathbf{x}), && \mathbf{x} \in \mathbb{R}^3. \end{aligned} \tag{11.91}$$

The solution formula reads as

$$T(\mathbf{x},t) = \int_{\mathbb{R}^3} \mathcal{K}(\mathbf{x},t;\mathbf{y},s)\, T_0(\mathbf{y})\, d^3y$$

for all $\mathbf{x} \in \mathbb{R}^3$ and $t \geq s$. For the heat kernel $\mathcal{K}$, we get Feynman's key formula

$$\boxed{\mathcal{K}(\mathbf{x},t;\mathbf{y},s) = \lim_{n\to\infty} \int_{\mathbb{R}^{3n}} \mathrm{e}^{-S}\, \frac{d^3x_1 d^3x_2 \cdots d^3x_n}{\left(4\pi\kappa\Delta t\right)^{\frac{3n}{2}}}} \tag{11.92}$$

along with the discrete action

$$S := \sum_{j=0}^{n} \left(\frac{1}{4\kappa} \left(\frac{\mathbf{x}_{j+1} - \mathbf{x}_j}{\Delta t} \right)^2 + U(\mathbf{x}_j) \right) \Delta t. \tag{11.93}$$

Here, we use the equidistant decomposition

$$t_0 = s < t_1 < t_2 < \ldots < t_{n+1} = t$$

of the time interval $[s, t]$ where $t_j := s + j\Delta t$. Equivalently, the key formula (11.92) can be written as

$$\mathcal{K}(\mathbf{x}, t; \mathbf{y}, s) = \lim_{n\to\infty} \int_{\mathbb{R}^{3n}} \prod_{j=0}^{n} \mathcal{K}_{\Delta t}(\mathbf{x}_{j+1}, t_{j+1}; \mathbf{x}_j, t_j)\; d^3x_1 d^3x_2 \cdots d^3x_n$$

with the so-called infinitesimal heat kernel

$$\mathcal{K}_{\Delta t}(\mathbf{x}_{j+1}, t_{j+1}; \mathbf{x}_j, t_j) :== \frac{\mathrm{e}^{-\frac{(\mathbf{x}_{j+1} - \mathbf{x}_j)^2}{4\kappa\,\Delta t}}}{\left(4\pi\kappa\Delta t\right)^{\frac{3}{2}}}\; \mathrm{e}^{-U(\mathbf{x}_j)\Delta t}\,.$$

Observe that the infinitesimal heat kernel coincides with the free heat kernel if the potential vanishes, $U \equiv 0$. Summarizing,

> *Feynman's key formula* (11.92) *reduces the full heat kernel under the influence of the potential U to an infinite-dimensional integral which depends on the simpler, explicitly known infinitesimal heat kernel.*

Let us give a formal proof via the elegant Dirac calculus.

Step 1: The heat kernel. Introduce the Hamiltonian

$$H := H_0 + U$$

with the unperturbed Hamiltonian $H_0 := \kappa\mathbf{P}^2 = \kappa\Delta$, and $\mathbf{P} := -\mathrm{i}\boldsymbol{\partial}$. The heat equation (11.91) reads then as

$$\dot{T}(t) = -HT(t), \qquad t \geq 0, \quad T(s) = T_0.$$

This equation has the solution

$$T(t) = \mathrm{e}^{-H(t-s)} T_0, \qquad t \geq s.$$

Using Dirac's completeness relation $\int_{\mathbb{R}^3} d^3y\, |\mathbf{y}\rangle\langle\mathbf{y}| = I$,

$$\langle\mathbf{x}|T(t)\rangle = \int_{\mathbb{R}^3} d^3y\, \langle\mathbf{x}|\mathrm{e}^{-H(t-s)}|\mathbf{y}\rangle\langle\mathbf{y}|T_0\rangle.$$

This implies

$$\boxed{T(\mathbf{x}, t) = \int_{\mathbb{R}^3} \mathcal{K}(\mathbf{x}, t; \mathbf{y}, s)\, T_0(\mathbf{y})\, d^3y}$$

for all $\mathbf{x} \in \mathbb{R}^3$ and all $t \geq s$ with the heat kernel

$$\mathcal{K}(\mathbf{x}, t; \mathbf{y}, s) := \langle\mathbf{x}|\mathrm{e}^{-H(t-s)}|\mathbf{y}\rangle.$$

Step 2: Causality. The heat kernel satisfies the product formula

$$\mathcal{K}(\mathbf{x}, t; \mathbf{y}, s) = \int_{\mathbb{R}^3} d^3z\, \mathcal{K}(\mathbf{x}, t; \mathbf{z}, \tau)\, \mathcal{K}(\mathbf{z}, \tau; \mathbf{y}, s)$$

if $t > \tau > s$. This follows from the causality relation

$$\mathrm{e}^{-H(t-s)} = \mathrm{e}^{-H(t-\tau)}\,\mathrm{e}^{-H(\tau-s)}$$

along with the completeness relation $\int d^3z\,|\mathbf{z}\rangle\langle\mathbf{z}| = I$. Hence

$$\langle\mathbf{x}|\mathrm{e}^{-H(t-s)}|\mathbf{y}\rangle = \int_{\mathbb{R}^3} d^3z\,\langle\mathbf{x}|\mathrm{e}^{-H(t-\tau)}|\mathbf{z}\rangle\langle\mathbf{z}|\mathrm{e}^{-H(\tau-s)}|\mathbf{y}\rangle.$$

Analogously, $\mathcal{K}(\mathbf{x},t;\mathbf{y},s)$ is equal to the following multiple integral

$$\boxed{\int_{\mathbb{R}^{3n}} d^3x_n\,d^3x_{n-1}\cdots d^3x_1\times \\ \times\mathcal{K}(\mathbf{x},t;\mathbf{x}_n,t_n)\,\mathcal{K}(\mathbf{x}_n,t_n;\mathbf{x}_{n-1},t_{n-1})\cdots\mathcal{K}(\mathbf{x}_1,t_1;\mathbf{y},s).}$$

For small time intervals Δt, the basic idea is now to replace the heat kernel

$$\mathcal{K}(\mathbf{x}_{j+1},t_{j+1}\,;\mathbf{x}_j,t_j) = \langle\mathbf{x}_{j+1}|\mathrm{e}^{-(H_0+U)\Delta t}|\mathbf{x}_j\rangle$$

by the infinitesimal heat kernel

$$\mathcal{K}_{\Delta t}(\mathbf{x}_{j+1},t_{j+1}\,;\mathbf{x}_j,t_j) := \langle\mathbf{x}_{j+1}|\mathrm{e}^{-H_0\Delta t}\,\mathrm{e}^{-U\Delta t}|\mathbf{x}_j\rangle$$

and to perform the limit $n\to\infty$, i.e., $\Delta t\to+0$. This will produce the claim (11.92). Let us motivate this.

Step 3: The Trotter product formula. By the addition theorem for the exponential function,

$$\mathrm{e}^{-H(t-s)} = \mathrm{e}^{-H(t-t_n)}\,\mathrm{e}^{-H(t_n-t_{n-1})}\cdots\mathrm{e}^{-H(t_1-s)}.$$

Noting that $H = H_0 + U$, the Trotter product formula tells us that

$$\mathrm{e}^{-H(t-s)} = \lim_{n\to\infty}\mathrm{e}^{-H_0(t-t_n)}\,\mathrm{e}^{-U(t-t_n)}\cdots\mathrm{e}^{-H_0(t_1-s)}\,\mathrm{e}^{-U(t_1-s)}. \tag{11.94}$$

This means that, for small time intervals Δt, we can replace approximately $\mathrm{e}^{-(H_0+U)\Delta t}$ by the product $\mathrm{e}^{-H_0\Delta t}\mathrm{e}^{-U\Delta t}$. This is motivated by

$$\mathrm{e}^{A\Delta t} = I + A\Delta t + o(\Delta t), \qquad \Delta t\to+0,$$

and hence

$$\mathrm{e}^{(A+B)\Delta t} = \mathrm{e}^{A\Delta t}\mathrm{e}^{B\Delta t} + o(\Delta t), \qquad \Delta t\to+0.$$

By the completeness relation $\int d^3x\,|\mathbf{x}\rangle\langle\mathbf{x}| = I$, it follows from (11.94) that the transition amplitude $\langle\mathbf{x}|\mathrm{e}^{-H(t-s)}|\mathbf{y}\rangle$ is equal to

$$\lim_{n\to\infty}\int_{\mathbb{R}^{3n}} d^3x_n\cdots d^3x_1\,\mathcal{K}_{\Delta t}(\mathbf{x},t;\mathbf{x}_n,t_n)\cdots\mathcal{K}_{\Delta t}(\mathbf{x}_1,t_1;\mathbf{y},s).$$

Step 4: Explicitly computing the infinitesimal heat kernel. Finally, we have to show that

$$\boxed{\mathcal{K}_{\Delta t}(\mathbf{x},t;\mathbf{y},s) = \frac{\mathrm{e}^{-\frac{(\mathbf{x}-\mathbf{y})^2}{4\kappa\Delta t}}}{\big(4\pi\kappa\Delta t\big)^{\frac{3}{2}}}\,\mathrm{e}^{-U(\mathbf{y})\Delta t}.} \tag{11.95}$$

In fact, by the completeness relation $\int_{\mathbb{R}^3} d^3p\,|\mathbf{p}\rangle\langle\mathbf{p}| = I$,

$$\langle\mathbf{x}|\mathrm{e}^{-H_0\Delta t}\,\mathrm{e}^{-U\Delta t}|\mathbf{y}\rangle = \int_{\mathbb{R}^3} d^3p\ \ \langle\mathbf{x}|\mathrm{e}^{-H_0\Delta t}|\mathbf{p}\rangle\langle\mathbf{p}|\mathrm{e}^{-U\Delta t}|\mathbf{y}\rangle.$$

Since $\mathbf{P}|\mathbf{p}\rangle = \mathbf{p}|\mathbf{p}\rangle$ and $H_0 = \kappa\mathbf{P}^2$,

$$\mathrm{e}^{-H_0\Delta t}|\mathbf{p}\rangle = \mathrm{e}^{-\kappa\mathbf{p}^2\Delta t}|\mathbf{p}\rangle.$$

Moreover, $\mathrm{e}^{-U\Delta t}|\mathbf{y}\rangle = \mathrm{e}^{-U(\mathbf{y})\Delta t}|\mathbf{y}\rangle$. Hence

$$\langle\mathbf{x}|\mathrm{e}^{-H_0\Delta t}\,\mathrm{e}^{-U\Delta t}|\mathbf{y}\rangle = \int_{\mathbb{R}^3} d^3p\,\mathrm{e}^{-\kappa\mathbf{p}^2\Delta t}\,\mathrm{e}^{-U(\mathbf{y})\Delta t}\langle\mathbf{x}|\mathbf{p}\rangle\langle\mathbf{p}|\mathbf{y}\rangle.$$

Since $\langle\mathbf{x}|\mathbf{p}\rangle = \frac{\mathrm{e}^{\mathrm{i}\mathbf{x}\mathbf{p}}}{(2\pi)^{3/2}}$, we get

$$\begin{aligned}\langle\mathbf{x}|\mathrm{e}^{-H_0\Delta t}\,\mathrm{e}^{-U\Delta t}|\mathbf{y}\rangle &= \left(\int_{\mathbb{R}^3} d^3p\,\mathrm{e}^{-\kappa\mathbf{p}^2\Delta t}\,\frac{\mathrm{e}^{\mathrm{i}(\mathbf{x}-\mathbf{y})\mathbf{p}}}{(2\pi)^3}\right)\mathrm{e}^{-U(\mathbf{y})\Delta t}\\ &= \frac{\mathrm{e}^{-\frac{(\mathbf{x}-\mathbf{y})^2}{4\kappa\Delta t}}}{\big(4\pi\kappa\Delta t\big)^{\frac{3}{2}}}\,\mathrm{e}^{-U(\mathbf{y})\Delta t},\end{aligned} \tag{11.96}$$

by using the classical Gaussian integral from (7.102) on page 433. This finishes the formal proof of claim (11.95).

11.8.2 Diffusion, Brownian Motion, and the Wiener Integral

The integral

$$\int_{\mathbb{R}^{3n}} \mathrm{e}^{-S}\,\frac{d^3x_1\,d^3x_2\cdots d^3x_n}{\big(4\pi\kappa\Delta t\big)^{\frac{3n}{2}}}$$

can be viewed approximately as a summation over all possible piecewise linear paths $\mathbf{q} = \mathbf{q}(t)$ from the point $\mathbf{y}$ at time s to the point $\mathbf{x}$ at time t along with the node points

$$\mathbf{q}(t_j) = \mathbf{x}_j, \qquad j = 0, \ldots, n+1.$$

Here, $\mathbf{x}_0 := \mathbf{y}$ and $\mathbf{x}_{n+1} := \mathbf{x}$. Then

$$S(\mathbf{q}) = \int_s^t \left(\frac{\dot{\mathbf{q}}(\tau)^2}{2\kappa} + U\big(\mathbf{q}(\tau)\big)\right) d\tau.$$

Therefore, it seems to be reasonable to write the key formula (11.92) as

$$\boxed{\mathcal{K}(\mathbf{x},t;\mathbf{y},s) = \int_{\mathbf{q}(s)=\mathbf{y}}^{\mathbf{q}(t)=\mathbf{x}} \mathrm{e}^{-S(\mathbf{q})}\ \mathcal{D}\mathbf{q}.} \tag{11.97}$$

Here, we sum over all paths $\mathbf{q} = \mathbf{q}(t)$ connecting the starting point $\mathbf{y}$ with the point final $\mathbf{x}$.

The point is that the famous Feynman–Kac formula (11.97) can be given a rigorous meaning, in the sense of a classical measure integral with respect to the Wiener measure on the space of continuous curves $\mathbf{q} : [s,t] \to \mathbb{R}^3$. This can be found in Reed and Simon (1972), Vol. II, Sect. X.11. In terms of physics, we regard the

differential equation (11.91) as diffusion equation for the mass density function T. Microscopically, diffusion corresponds to the stochastic motion of molecules (Brownian motion). This stochastic process can be described by some probability measure on the space of continuous paths. This measure was introduced by Norbert Wiener in 1923. The path integral (11.97) represents the corresponding Wiener integral.[48]

11.8.3 The Method of Quantum Fluctuations

We want to discuss a formal method which is frequently used by physicists in order to compute Feynman functional integrals in an elegant way. It is our goal

- to separate quantum fluctuations from the classical motion, and
- to compute the Feynman functional integral corresponding to quantum fluctuations by the method of zeta function regularization.

We will apply this method to the heat kernel. However, the same method also applies to the Feynman propagator kernel for the Schrödinger equation. Let us consider the heat equation (11.91) with vanishing potential, $U \equiv 0$. Below we want to show the following.

(i) Choose the action

$$S(\mathbf{q}) := \int_s^t \frac{\dot{\mathbf{q}}(\tau)^2}{4\kappa}\, d\tau.$$

Then

$$\int_{\mathbf{q}(s)=\mathbf{y}}^{\mathbf{q}(t)=\mathbf{x}} \mathrm{e}^{-S(\mathbf{q})} \mathcal{D}\mathbf{q} = \mathrm{e}^{-S(\mathbf{q}_{\mathrm{class}})} \int_{\mathbf{q}(s)=0}^{\mathbf{q}(t)=0} \mathrm{e}^{-S(\mathbf{q})}\, \mathcal{D}\mathbf{q}. \tag{11.98}$$

Here, $\mathbf{q}_{\mathrm{class}}$ is the solution to the action principle

$$S(\mathbf{q}) = \text{stationary}\,!\,, \qquad \mathbf{q}(s) = \mathbf{y},\ \mathbf{q}(t) = \mathbf{x}. \tag{11.99}$$

(ii) For the "classical contribution",

$$\mathrm{e}^{-S(\mathbf{q}_{\mathrm{class}})} = \mathrm{e}^{-\frac{(\mathbf{x}-\mathbf{y})^2}{4\kappa(t-s)}}\,. \tag{11.100}$$

(iii) For the "quantum fluctuations",

$$\int_{\mathbf{q}(s)=0}^{\mathbf{q}(t)=0} \mathrm{e}^{-S(\mathbf{q})}\, \mathcal{D}\mathbf{q} = \frac{1}{\big(4\pi\kappa(t-s)\big)^{\frac{3}{2}}}\,. \tag{11.101}$$

[48] N. Wiener, Differential space, J. Math. and Phys. of MIT **2** (1923), 131–174.
M. Kac, On distributions of certain Wiener functionals, Trans. Amer. Math. Soc. **65** (1949), 1–13.
R. Cameron, A family of integrals serving to connect the Wiener and Feynman integrals. J. of Math. and Phys. Sci. of MIT, **39** (1960), 126–140.
M. Kac, Wiener and integration in function spaces, Bull. Amer. Math. Soc. **72** (1966), 52–68.

This means that formula (11.98) yields the right expression for the heat kernel

$$\mathcal{K}(\mathbf{x},t;\mathbf{y},s)=\frac{\mathrm{e}^{-\frac{(\mathbf{x}-\mathbf{y})^2}{4\kappa(t-s)}}}{\big(4\pi\kappa(t-s)\big)^{\frac{3}{2}}}\,.$$

The crucial point is the verification of (11.101) below.

Step 1: Computation of $\mathbf{q}_{\mathrm{class}}$. The Euler–Lagrange equation to (11.99) reads as

$$\ddot{\mathbf{q}}(\tau)=0,\qquad \mathbf{q}(s)=\mathbf{y},\ \mathbf{q}(t)=\mathbf{x}$$

with the solution

$$\mathbf{q}_{\mathrm{class}}(\tau)=\left(\frac{\mathbf{x}-\mathbf{y}}{t-s}\right)\tau+\mathbf{y}-\left(\frac{\mathbf{x}-\mathbf{y}}{t-s}\right)s.$$

Hence

$$S(\mathbf{q}_{\mathrm{class}})=\int_s^t\frac{\dot{\mathbf{q}}_{\mathrm{class}}(\tau)^2}{4\kappa}\,d\tau=\frac{(\mathbf{x}-\mathbf{y})^2}{4\kappa(t-s)}\,.$$

Step 2: Decomposition. For an arbitrary motion, we set

$$\mathbf{q}(\tau):=\mathbf{q}_{\mathrm{class}}(\tau)+\mathbf{r}(\tau)$$

along with the boundary condition $\mathbf{r}(s)=\mathbf{r}(t)=0$ for the remainder $\mathbf{r}$. Hence

$$S(\mathbf{q})=S(\mathbf{q}_{\mathrm{class}})+S(\mathbf{r})+R$$

with $R:=\int_s^t\frac{1}{2\kappa}\dot{\mathbf{q}}_{\mathrm{class}}(\tau)\,\dot{\mathbf{r}}(\tau)\,d\tau$. Integration by parts yields

$$R=-\int_s^t\frac{1}{2\kappa}\,\ddot{\mathbf{q}}_{\mathrm{class}}(\tau)\mathbf{r}(\tau)\,d\tau=0.$$

This implies (11.98).

Step 3: Proof of (11.101). Let us first consider the special case of the motion $q=q(\tau)$ on the real line, that is, we want to show that

$$\boxed{\int_{q(s)=0}^{q(t)=0}\mathrm{e}^{-S(q)}\mathcal{D}q=\frac{1}{\big(4\pi\kappa(t-s)\big)^{\frac{1}{2}}}\,.}\tag{11.102}$$

Let $a>0$. Motivated by the formula

$$\int_{\mathbb{R}^3}dx\,dy\,dz\,\mathrm{e}^{-ax^2-ay^2-az^2}=\left(\int_{\mathbb{R}}dx\,\mathrm{e}^{-ax^2}\right)^3,$$

we then get the desired result

$$\int_{\mathbf{q}(s)=0}^{\mathbf{q}(t)=0}\mathrm{e}^{-S(\mathbf{q})}\mathcal{D}\mathbf{q}=\left(\int_{q(s)=0}^{q(t)=0}\mathrm{e}^{-S(q)}\mathcal{D}q\right)^3=\frac{1}{\big(4\pi\kappa(t-s)\big)^{\frac{3}{2}}}\,.$$

To simplify the argument, note that the rescaling $\tau\mapsto\alpha\tau$ of

$$S(q)=\int_s^t\frac{1}{4\kappa}\left(\frac{dq(\tau)}{d\tau}\right)^2d\tau\tag{11.103}$$

yields

$$\int_{\alpha s}^{\alpha t} \frac{1}{4\kappa\alpha} \left(\frac{dq(\tau)}{d\tau} \right)^2 d\tau.$$

For the integral from (11.102), this means that we have to replace κ by $\alpha\kappa$. Therefore, it is sufficient to motivate (11.102) in the special case where, say, $\kappa = \frac{1}{2}$. To this end, we will use the standard method of zeta function regularization.

11.8.4 Infinite-Dimensional Gaussian Integrals and Zeta Function Regularization

Choose $\kappa = \frac{1}{2}$. Let us motivate (11.102). The crucial trick is to write the action $S(q)$ as a quadratic form $\frac{1}{2}\langle q|\mathcal{A}q\rangle$ with a suitable differential operator $\mathcal{A}$, that is,

$$\int_{q(s)=0}^{q(t)=0} \mathrm{e}^{-S(q)}\, \mathcal{D}q = \int_{q(s)=0}^{q(t)=0} \mathrm{e}^{-\frac{1}{2}\langle q|\mathcal{A}q\rangle}\, \mathcal{D}q.$$

Then, motivated by the classical Gaussian integral

$$\int_{\mathbb{R}^n} \mathrm{e}^{-\frac{1}{2}x^{\mathrm{T}}Ax} \frac{d^n x}{(2\pi)^{\frac{n}{2}}} = \frac{1}{\sqrt{\det A}}$$

for all $n = 1, 2, \ldots$, let us try to make the formal ansatz

$$\boxed{\int_{q(s)=0}^{q(t)=0} \mathrm{e}^{-\frac{1}{2}\langle q|\mathcal{A}q\rangle}\, \mathcal{D}q := \frac{\gamma}{\sqrt{\det \mathcal{A}}}}$$

where γ is a normalization constant to be determined.[49] Finally, we have to compute the determinant $\det \mathcal{A}$ of the operator $\mathcal{A}$. To this end, we will use the method of zeta function regularization. Surprisingly enough, as we will show below, this formal approach provides us with the correct heat kernel formula if we choose $\gamma := \frac{1}{\sqrt{\pi}}$. To begin with, we start with (11.103). Setting $\kappa = \frac{1}{2}$, we get

$$S(q) = \frac{1}{2} \int_s^t \left(\frac{dq(\tau)}{d\tau} \right)^2 d\tau.$$

Integration by parts yields

$$S(q) = \frac{1}{2} \int_s^t q(\tau) \mathcal{A} q(\tau)\, d\tau$$

with the differential operator $\mathcal{A} := -\frac{d^2}{d\tau^2}$ (1-dimensional Laplacian). More precisely, the domain of definition, $D(\mathcal{A})$, of the operator

$$\mathcal{A} : D(\mathcal{A}) \subseteq L_2(s,t) \to L_2(s,t)$$

[49] Note the following nice fact. In applications of path integrals to quantum field theory, one wants to compute correlation functions. In this case, the explicit knowledge of the normalization factor is not necessary, since the correlation functions are given by quotients of functional integrals where the normalization factors are cancelled.

consists of all smooth functions $q : [s,t] \to \mathbb{C}$ with $q(s) = q(t) = 0$. This operator has the eigensolutions

$$\mathcal{A}q_n(\tau) = \lambda_n q_n(\tau), \qquad s \le \tau \le t$$

with $q_n(\tau) = \text{const} \cdot \sin \frac{\pi n \tau}{t-s}$ and

$$\lambda_n = \left(\frac{\pi n}{t-s}\right)^2, \qquad n = 1, 2, \dots .$$

Motivated by the classical zeta function determinant formula (5.19) on page 261, we introduce first the ζ-function

$$\zeta_{\mathcal{A}}(z) := \sum_{n=1}^{\infty} \frac{1}{\lambda_n^z}, \qquad \Re(z) > \frac{1}{2},$$

and we then define

$$\boxed{\det \mathcal{A} := \mathrm{e}^{-\zeta'_{\mathcal{A}}(0)}.} \tag{11.104}$$

Here, in some neighborhood of the origin, $z = 0$, the zeta function $\zeta_{\mathcal{A}}$ is to be understood in the sense of analytic continuation. Explicitly,

$$\zeta_{\mathcal{A}}(z) = \left(\frac{t-s}{\pi}\right)^{2z} \sum_{n=1}^{\infty} \frac{1}{n^{2z}} = \left(\frac{t-s}{\pi}\right)^{2z} \zeta(2z)$$

where ζ denotes the classical Riemann zeta function. By analytic continuation, Riemann showed that

$$\zeta(0) = -\frac{1}{2}, \qquad \zeta'(0) = -\frac{1}{2} \ln 2\pi.$$

Thus, it follows from $\zeta_{\mathcal{A}}(z) = \mathrm{e}^{2z \ln \frac{t-s}{\pi}} \zeta(2z)$ that

$$\zeta'_{\mathcal{A}}(0) = 2\zeta(0) \ln \frac{t-s}{\pi} + 2\zeta'(0) = -\ln 2(t-s).$$

Hence

$$\det \mathcal{A} = \mathrm{e}^{-\zeta'_{\mathcal{A}}(0)} = 2(t-s).$$

Choosing the normalization constant $\gamma := \frac{1}{\sqrt{\pi}}$, we get the desired formula

$$\int_{q(s)=0}^{q(t)=0} \mathrm{e}^{-\frac{1}{2}\langle q | \mathcal{A} q \rangle} \, \mathcal{D}q = \frac{\gamma}{\sqrt{\det \mathcal{A}}} = \frac{1}{\sqrt{2\pi(t-s)}} \, .$$

11.8.5 The Euclidean Trick and the Feynman Path Integral for the Schrödinger Equation

One of the main tricks used in physics is to reduce the Schrödinger equation to the Fourier heat equation by passing to imaginary time. This is called the Euclidean trick. Explicitly, replacing the time t by $\mathrm{i}t/\hbar$ and choosing the coupling constant $\kappa := \hbar^2/2m$, the initial problem (11.91) for the heat equation passes over to the initial-value problem for the Schrödinger equation

$$\begin{aligned} \mathrm{i}\hbar\frac{\partial T}{\partial t}(\mathbf{x},t) &= \tfrac{\hbar^2}{2m}\,\Delta T(\mathbf{x},t) + U(\mathbf{x})T(\mathbf{x},t), \qquad && \mathbf{x}\in\mathbb{R}^3,\ t\geq s, \\ T(\mathbf{x},s) &= T_0(\mathbf{x}), && \mathbf{x}\in\mathbb{R}^3. \end{aligned} \tag{11.105}$$

The solution formula for the heat equation (11.91) on page 651 passes over to

$$T(\mathbf{x},t) = \int_{\mathbb{R}^3} \mathcal{P}_+(\mathbf{x},t;\mathbf{y},s)\, T_0(\mathbf{y})\, d^3y$$

for all $\mathbf{x}\in\mathbb{R}^3$ and $t\geq s$. For the Feynman propagator kernel $\mathcal{P}_+$, we get Feynman's key formula[50]

$$\boxed{\mathcal{P}_+(\mathbf{x},t;\mathbf{y},s) = \lim_{n\to\infty}\int_{\mathbb{R}^{3n}} \mathrm{e}^{\mathrm{i}\mathcal{S}/\hbar}\ \frac{d^3x_1 d^3x_2\cdots d^3x_n}{\left(\frac{\mathrm{i}h\Delta t}{m}\right)^{\frac{3n}{2}}}} \tag{11.106}$$

along with the discrete action

$$\mathcal{S} := \sum_{j=0}^{n}\left(\frac{m}{2}\left(\frac{\mathbf{x}_{j+1}-\mathbf{x}_j}{\Delta t}\right)^2 - U(\mathbf{x}_j)\right)\Delta t,$$

and the equidistant decomposition $t_0 = s < t_1 < t_2 < \ldots < t_{n+1} = t$ of the time interval $[s,t]$ where $t_j := s + j\Delta t$. Explicitly, we start with the discrete action S for the heat equation introduced in (11.93). We obtain the action $\mathcal{S}$ for the Schrödinger equation by using the following replacements:

- $\Delta t \Rightarrow \mathrm{i}\Delta t/\hbar$,
- $-S \Rightarrow \mathrm{i}\mathcal{S}/\hbar$, and
- $\kappa \Rightarrow \frac{\hbar^2}{2m}$.

Furthermore, after carrying out the formal limit $\Delta t\to 0$, we get

$$\boxed{\mathcal{P}_+(\mathbf{x},t;\mathbf{y},s) = \int_{\mathbf{q}(s)=\mathbf{y}}^{\mathbf{q}(t)=\mathbf{x}} \mathrm{e}^{\mathrm{i}\mathcal{S}(\mathbf{q})/\hbar}\ \mathcal{D}\mathbf{q}.} \tag{11.107}$$

Here, we sum over all paths $\mathbf{q}=\mathbf{q}(t)$ connecting the point $\mathbf{y}$ with the point $\mathbf{x}$, and we use the action

$$\mathcal{S}(\mathbf{q}) = \int_s^t \left(\frac{m\dot{\mathbf{q}}(\tau)^2}{2} - U\big(\mathbf{q}(\tau)\big)\right) d\tau.$$

Note that this is the action of a classical particle of mass m which moves under the influence of the potential U along the trajectory $\mathbf{q}=\mathbf{q}(t)$.

Suggested reading. As an introduction to Feynman's path integral, we recommend Zeidler (1995), Vol. 1, Chap. 5. Much material on path integrals can be found in the following monographs: Schulmann (1981), Chaichian and Demichev (2001), Vols. 1, 2, Kleinert (1990), and Roepstorff (1996). In the handbook on path integrals by Grosche and Steiner (1998), one finds an extensive list of about 200 pages which summarizes all of the explicitly known path integrals in quantum mechanics. This handbook also contains quite interesting comments on the history of the Feynman path integral.

[50] We choose the convention $\mathrm{i}^{\frac{3n}{2}} := \mathrm{e}^{\frac{3n\mathrm{i}\pi}{4}}$. Furthermore, note that $4\pi\kappa\Delta t$ passes over to $4\pi\cdot\frac{\hbar^2\mathrm{i}\Delta t}{2m\hbar} = \frac{\mathrm{i}h\Delta t}{m}$.

For a rigorous approach of the path integral to quantum physics, we recommend Reed and Simon (1972), Vol. 2, Simon (1979), Albeverio and Høegh-Krohn (1975), Glimm and Jaffe (1981), Klauder and Daubchies (1982), Klauder (1989), (2000). We also refer to Freidlin (1985) and Johnson, Lapidus (2000).

For Brownian motion, stochastic processes and the Wiener integral, we refer to Hida (1970), Yeh (1973), and Chung, Zha (1995) (from Brownian motion to the Schrödinger equation).

The main trouble with Feynman integrals in quantum physics. The typical mathematical difficulty is based on the fact that, in contrast to the Wiener integral for the heat equation, the formal Feynman measures $\mathcal{D}\mathbf{q}$ and $\mathcal{D}\psi$ for the Schrödinger equation and the equations of quantum field theory, respectively, do not exist as classical measures. Therefore, the Feynman path integral and its generalizations to quantum field theory do not exist as standard measure integrals. We refer to Cameron (1960) and Johnson, Lapidus (2000), Sect. 4.6.

11.9 A Glance at Harmonic Analysis

11.9.1 The Fourier–Laplace Transform

For the Heaviside function θ, define

$$\boxed{a(\omega) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \theta(t) \mathrm{e}^{-\mathrm{i}\omega t} dt.}$$

If the frequency ω is real, then the integral does not exist. To overcome this difficulty, the idea is to pass to complex frequencies ω living in the open lower half-plane $\{\omega \in \mathbb{C} : \Im(\omega) < 0\}$, that is, we set $\omega := \eta - \varepsilon \mathrm{i}$ where η is a real number and $\varepsilon > 0$ This yields the damped oscillation

$$\mathrm{e}^{-\mathrm{i}\omega} = \mathrm{e}^{-\mathrm{i}\eta t} \mathrm{e}^{-\varepsilon t}, \qquad t \in \mathbb{R}.$$

Explicitly,

$$a(\omega) = \lim_{N \to \infty} \frac{\mathrm{i} \mathrm{e}^{-\eta \mathrm{i}\omega} \mathrm{e}^{-\varepsilon\omega}}{\omega\sqrt{2\pi}} \Bigg|_0^N = \frac{1}{\mathrm{i}\omega\sqrt{2\pi}}.$$

The function $a = a(\omega)$ is called the Fourier–Laplace transform of the Heaviside function θ; it is holomorphic in the open lower half-plane of the complex frequency plane. Let us consider the following two examples.

(a) Mnemonically, for the Dirac delta function δ, the Fourier–Laplace transform looks like

$$a(\omega) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \delta(t) \mathrm{e}^{-\mathrm{i}\omega t} dt = \frac{1}{\sqrt{2\pi}}, \qquad \omega \in \mathbb{C}.$$

Rigorously, we set $e_\omega(t) := \mathrm{e}^{-\mathrm{i}\omega t}$, and

$$\boxed{a(\omega) := \frac{\delta(e_\omega)}{\sqrt{2\pi}}.}$$

Hence $a(\omega) = \frac{e_\omega(0)}{\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}}$ for all $\omega \in \mathbb{C}$.

(b) For a smooth function $g : \mathbb{R} \to \mathbb{C}$ with compact support, the Fourier–Laplace transform

$$a(\omega) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(t) \mathrm{e}^{-\mathrm{i}\omega t} dt, \qquad \omega \in \mathbb{C}$$

is holomorphic on the complex frequency plane.

The Paley–Wiener–Schwartz theorem. This famous theorem of analysis shows that distributions with compact support and entire analytic functions with appropriate growth conditions are indeed the same via Fourier–Laplace transform. In what follows, let $n = 1, 2, \ldots$, and $R > 0$. To fix the terminology, recall that the symbol $\mathcal{D}(\mathbb{R}^n)$ denotes the set of all smooth functions $f : \mathbb{R}^n \to \mathbb{C}$ with compact support. The symbol $\mathbb{B}^n_R$ is used for the closed ball $\{x \in \mathbb{R}^n : \ ||x|| \leq R\}$. Finally, for each $p \in \mathbb{C}^n$, we set

$$e_p(x) := \mathrm{e}^{-\mathrm{i}\langle x|p\rangle}, \qquad x \in \mathbb{R}^n$$

where we use the convention $\langle x|p\rangle := \sum_{j=1}^n x^j p^j$.

(i) *The Paley–Wiener theorem for smooth functions with compact support*: For each function $f \in \mathcal{D}(\mathbb{R}^n)$ with support in the ball $\mathbb{B}^n_R$, the Fourier–Laplace transform

$$a(p) := \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(x) e_p(x) d^n x, \qquad p \in \mathbb{C}^n$$

is holomorphic on the complex momentum space $\mathbb{C}^n$, and it satisfies the following decay condition at infinity:

$$|a(p)| \leq \frac{\mathrm{const}(N) \cdot \mathrm{e}^{R|\Im(p)|}}{(1 + ||p||)^N} \tag{11.108}$$

for all $p \in \mathbb{C}^n$ and all exponents $N = 1, 2, \ldots$ The crucial point is that the converse is also true. Explicitly, if the function $a : \mathbb{C}^n \to \mathbb{C}$ is holomorphic and satisfies the decay condition (11.108), then it is the Fourier–Laplace transform of a function $f \in \mathcal{D}(\mathbb{R}^n)$ with support in the ball $\mathbb{B}^n_R$.

(ii) *The Schwartz theorem for distributions with compact support*: For each distribution $F \in \mathcal{D}'(\mathbb{R}^n)$ with support in the ball $\mathbb{B}^n_R$, the Fourier–Laplace transform

$$a(p) := F(e_p), \qquad p \in \mathbb{C}^n$$

is holomorphic on the complex momentum space $\mathbb{C}^n$, and it satisfies the following polynomial growth condition:

$$|a(p)| \leq \mathrm{const} \cdot (1 + ||p||)^N \ \mathrm{e}^{R|\Im(p)|} \tag{11.109}$$

for all $p \in \mathbb{C}^n$ and some fixed natural number N.[51] The converse is also true. Explicitly, if the function $a : \mathbb{C}^n \to \mathbb{C}$ is holomorphic and satisfies the growth condition (11.109), then it is the Fourier–Laplace transform of a distribution $F \in \mathcal{D}'(\mathbb{R}^n)$ with support in the ball $\mathbb{B}^n_R$.

(iii) *Polynomials*: Suppose that the support of the distribution $F \in \mathcal{D}'(\mathbb{R}^n)$ consists precisely of the origin. Then, there exists a natural number N such that

$$F = \sum_{|\alpha| \leq N} a_\alpha \partial^\alpha \delta,$$

[51] Naturally enough, the restriction of $a(p)$ to real momenta p coincides with the usual Fourier transform for tempered distributions.

where a_α denotes a complex number. This tells us that F is a finite sum of partial derivatives of Dirac's delta distribution. The Fourier–Laplace transform reads then as

$$a(p) = \sum_{|\alpha| \leq N} \frac{1}{(2\pi)^{n/2}} \cdot \mathrm{i}^{|\alpha|} a_\alpha p^\alpha, \qquad p \in \mathbb{C}^n.$$

Conversely, each polynomial $a = a(p)$ on the complex momentum space $\mathbb{C}^n$ is the Fourier–Laplace transform of a distribution which has precisely the origin as support. This shows that the Dirac delta distribution and its partial derivatives are quite natural objects, from the point of view of the Fourier–Laplace transform. For example, the distribution

$$F = a\delta + b\delta' + c\delta''$$

on the real line $\mathbb{R}$ with complex numbers a, b, c has the Fourier–Laplace transform

$$\frac{1}{\sqrt{2\pi}} \cdot (a + \mathrm{i}bp - cp^2), \qquad \text{for all} \quad p \in \mathbb{C}.$$

The proofs of statements (i)-(iii) above can be found in Hörmander (1983), Vol. 1, Sect. 7.3. As we will show in Volume II, statement (iii) plays a crucial role in the Epstein–Glaser approach to the renormalization of quantum field theories.

11.9.2 The Riemann–Hilbert Problem

Let us decompose the complex plane $\mathbb{C}$ into the real line, the open upper half-plane

$$\mathbb{C}_> := \{z \in \mathbb{C} : \ \Im(z) > 0\},$$

and the open lower half-plane $\mathbb{C}_< := \{z \in \mathbb{C} : \ \Im(z) < 0\}$. We also set $z = x + y\mathrm{i}$, that is, x and y are the real and imaginary part of the complex number z, respectively.[52] The prototype of the Riemann–Hilbert problem reads as follows. We are given the function $f \in \mathcal{D}(\mathbb{R})$. We are looking for a complex-valued function F which has the following properties:

(i) F is holomorphic on both the open half-planes $\mathbb{C}_>$ and $\mathbb{C}_<$.
(ii) For each point x on the real line, we have the following jump condition

$$\lim_{y \to +0} F(x + y\mathrm{i}) - \lim_{y \to -0} F(x + y\mathrm{i}) = f(x)$$

for the boundary values of F on the real line.[53]

Theorem 11.31 *The Riemann–Hilbert problem has the solution*

$$F = \frac{\mathrm{i}}{2} \cdot \mathcal{H}(f).$$

[52] The closed upper half-plane is defined by $\mathbb{C}_\geq := \{z \in \mathbb{C} : \ \Im(z) \geq 0\}$.
[53] We tacitly assume that the limits exist.

Explicitly, $\mathcal{H}(f)$ is given as follows. Fix the function $f \in \mathcal{D}(\mathbb{R})$. For arbitrary complex number z, define

$$\boxed{\mathcal{H}(f)(z) := \frac{1}{\pi} \cdot PV \int_{\mathbb{R}} \frac{f(\xi)d\xi}{z-\xi}.}$$

If z lies on the real line (resp. not on the real line), then the integral is to be understood in the sense of Cauchy's principal value (resp. in the classical sense). Theorem 11.31 follows now from the following two statements.

(P1) The function $\mathcal{H}(f)$ is holomorphic on $\mathbb{C}_>$ and $\mathbb{C}_<$.
(P2) For all points on the real line, $x \in \mathbb{R}$, we have the Sokhotski formula

$$\lim_{y \to \pm 0} \mathcal{H}(f)(x + \mathrm{i}y) = \mathcal{H}(f)(x) \mp \mathrm{i}f(x). \tag{11.110}$$

Proof. Ad (P1). This is a standard result for holomorphic functions. It follows from the Weierstrass theorem on sequences of holomorphic functions which converge on a compact set. See Remmert (1991), p. 249.

Ad (P2). This follows from Theorem 11.16 on page 619. □

11.9.3 The Hilbert Transform

Let us restrict the operator $\mathcal{H}$ to the real line. That is, we consider the function $z \mapsto \mathcal{H}(f)(z)$ on the real line. We then get the operator

$$\mathcal{H} : \mathcal{D}(\mathbb{R}) \to L_2(\mathbb{R}).$$

It turns out that this operator can be uniquely extended to a linear continuous operator

$$H : L_2(\mathbb{R}) \to L_2(\mathbb{R})$$

which is called the Hilbert transform. This operator has the following crucial property concerning the boundary values of holomorphic functions.

Theorem 11.32 *Let $F : \mathbb{C}_> \to \mathbb{C}$ be a holomorphic function on the open upper half-plane such that we have the following growth restriction*

$$\sup_{y>0} \int_{-\infty}^{\infty} |F(x + y\mathrm{i})|^2 dx < \infty.$$

Then the limit $u(x) + \mathrm{i}v(x) := \lim_{y \to +0} F(x + y\mathrm{i})$ exists for almost all $x \in \mathbb{R}$. Moreover, the real part u and the imaginary part v of the boundary values of the function F on the real line lie in the Hilbert space $L_2(\mathbb{R})$, and they are related to each other by the Hilbert transform. Explicitly,

$$v = Hu \quad \text{and} \quad u = -Hv.$$

The proof can be found in Titchmarsh (1967). For the Hilbert transform in distribution theory, we refer to Pandey (1996).

Example. The assumptions of Theorem 11.32 are satisfied if

$$F(z) := \frac{a}{(z-b)^n}, \qquad z \in \mathbb{C} \setminus \{b\}$$

where a and b are complex numbers with $\Im(b) < 0$, and $n = 1, 2, \ldots$

11.9.4 Symmetry and Special Functions

The experience of mathematicians and physicists shows that

Behind special functions, there lurk symmetry groups.

Roughly speaking, all of the important special functions that appear in mathematical physics are governed by symmetries. For example, we have shown in Sect. 7.20.3 on page 409 that the addition theorem for the exponential function

$$\mathrm{e}^{\mathrm{i}(\theta+\varphi)} = \mathrm{e}^{\mathrm{i}\theta}\mathrm{e}^{\mathrm{i}\varphi} \qquad \text{for all} \quad \theta, \varphi \in \mathbb{R} \tag{11.111}$$

is closely related to the group of rotations in the plane. By the Euler formula

$$\mathrm{e}^{\mathrm{i}\theta} = \cos\theta + \mathrm{i}\sin\theta \qquad \text{for all} \quad \theta \in \mathbb{R},$$

we get the addition theorem

$$\sin(\theta+\varphi) = \sin\theta\cos\varphi + \sin\varphi\cos\theta \qquad \text{for all} \quad \theta, \varphi \in \mathbb{R}$$

for the sine function. Symmetry groups will be carefully studied in the later volumes. At this point, let us only sketch the relation between the exponential function e^z (resp. the function z^α) and the corresponding Haar measures. The existence of an invariant measure for Lie groups was proven by Haar (1885–1933) in the year of his death. This result had far-reaching consequences for harmonic analysis.

(i) *The additive group* $\mathbb{R}$ *of real numbers*: This group is isomorphic to the group of translations on the real line. It follows from the addition theorem for the exponential function $\mathrm{e}^{\mathrm{i}(p+r)x} = \mathrm{e}^{\mathrm{i}px}\mathrm{e}^{\mathrm{i}rx}$ that for each fixed $x \in \mathbb{R}$, the map

$$\boxed{p \mapsto \mathrm{e}^{\mathrm{i}xp}} \tag{11.112}$$

is a group morphism from the additive group $\mathbb{R}$ onto the multiplicative group $U(1)$. Superposition of these exponential functions,

$$f(x) = \int_{\mathbb{R}} a(p)\mathrm{e}^{\mathrm{i}px}d\mu(p), \qquad x \in \mathbb{R},$$

generates the Fourier transform with the Haar measure

$$\int_a^b d\mu(p) = \int_a^b dp = b - a, \qquad a, b \in \mathbb{R}.$$

This measure has the characteristic property that it is invariant under the group action $a \mapsto a + c, b \mapsto b + c$ for each fixed $c \in \mathbb{R}$. Note that the Haar measure of the additive group $\mathbb{R}$ coincides with the classical Lebesgue measure on the real line.

(ii) *The additive group* $\mathbb{Z}$ *of integers*: The map (11.112) is a group morphism from the additive group $\mathbb{Z}$ into the multiplicative group $U(1)$. The superposition

$$f(x) = \int_{\mathbb{R}} a(p)\mathrm{e}^{2\pi\mathrm{i}px}d\mu(p) = \sum_{n=-\infty}^{\infty} a(n)\mathrm{e}^{2\pi\mathrm{i}nx}$$

generates the discrete Fourier transform. Here, in contrast to (i), the Haar measure μ of the group $\mathbb{Z}$ corresponds to a mass distribution on the real line which assigns the unit mass to each point of $\mathbb{Z}$.

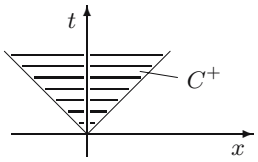


Fig. 11.5. The solid forward light cone

(iii) *The multiplicative group* $\mathbb{R}_+^\times$ *of positive real numbers*: Let $p, r, x > 0$. It follows from the product rule $(pr)^x = p^x r^x$ that, for each fixed $x \in \mathbb{R}_+^\times$, the map

$$\boxed{p \mapsto p^x}$$

is a group morphism from the multiplicative group $\mathbb{R}_+^\times$ into itself. The Haar measure μ of the group $\mathbb{R}_+^\times$ is given by

$$\int_a^b d\mu(p) = \int_a^b d\ln p = \ln\frac{b}{a}, \qquad b > a > 0.$$

This measure has the characteristic property that it is invariant under the group action $a \mapsto ac, b \mapsto bc$ for each $c \in \mathbb{R}_+^\times$. The superposition

$$f(x) = \int_0^\infty a(p)p^x d\ln p = \int_0^\infty a(p)p^{x-1}dp, \qquad x > 0$$

generates the Mellin transform (6.18) on page 290.

The Laplace transform is obtained from (i) and (ii) by restricting the measure μ to the interval $[0, \infty[$ and the set $\{0, 1, 2, \ldots\}$, respectively.

This summary shows that the simplest groups $\mathbb{R}, \mathbb{Z}$ and $\mathbb{R}_+^\times$ already provide us with a rich harmonic analysis. Observe that the group $\mathbb{Z} \otimes \mathbb{Z}$ yields the large class of double-periodic (i.e., elliptic) functions. Much material on the relations between symmetry groups and special functions can be found in the four-volume treatise by Vilenkin and Klimyk (1991). We also refer to Zagier (1995) (modular functions), Ford (1972) (automorphic functions), Knapp (1986) and Gelfand (1989) (non-compact groups).

11.9.5 Tempered Distributions as Boundary Values of Analytic Functions

Consider the equation

$$t^2 - x_1^2 - \ldots - x_n^2 = 0$$

of the light cone in the $(n+1)$-dimensional space-time $\mathbb{R}^{n+1}$ where $n = 1, 2, \ldots$ The set

$$C^+ := \{(x,t) \in \mathbb{R}^{n+1} : \ t \geq 0, \ t^2 - x_1^2 - \ldots - x_n^2 \geq 0\}$$

is called the solid forward light cone (Fig. 11.5).

Theorem 11.33 *Let T be a given tempered distribution on $\mathbb{R}^{n+1}$ with support in the solid forward light cone C^+. Then there exists always a holomorphic function $\mathcal{T} : \mathbb{R}^{n+1} - \mathrm{i}C^+ \to \mathbb{C}$ such that*

$$\lim_{\varepsilon \to +0} \mathcal{T}(x, t - \mathrm{i}\varepsilon) = T(x, t),$$

in the sense of tempered distributions on $\mathbb{R}^{n+1}$.

Explicitly, this means that

$$\lim_{\varepsilon \to +0} \int_{\mathbb{R}^{n+1}} \mathcal{T}(x, t - \mathrm{i}\varepsilon) \varphi(x, t) d^n x dt = T(\varphi)$$

for all test functions $\varphi \in \mathcal{S}(\mathbb{R}^{n+1})$. The proof can be found in Reed and Simon, Vol. 2, Sect. IX.3.

11.10 The Trouble with the Euclidean Trick

The passage $t \mapsto -\mathrm{i}t$ from real time t in quantum physics to imaginary time $-\mathrm{i}t$ is called the Euclidean trick. For example, this transformation sends the Schrödinger equation to the diffusion equation (see (11.13) on page 588). The experience of physicists shows that this trick is quite useful. From the mathematical point of view, note that

The Euclidean trick has to be handled very carefully.

Let us discuss two typical difficulties.

(i) Oscillating kernels. If we apply the Euclidean trick to the free quantum particle, then the heat kernel and the Feynman propagator kernel behave quite differently. Typically, this difference can be reduced to the fact that the Gaussian integral

$$\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{x^2}{2} - \mathrm{i}ax} dx = \sqrt{2\pi}\, \mathrm{e}^{-\frac{a^2}{2}}, \qquad a \in \mathbb{R}$$

is convergent, whereas the oscillating Fourier integral

$$\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}x} dx = \int_{-\infty}^{\infty} (\cos x + \mathrm{i} \sin x) dx$$

is divergent. The heat kernel is of the first type, whereas the Feynman propagator kernel corresponds to an oscillating integral. In the history of mathematics and physics, the following two oscillating integrals

$$\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}x^2}\, dx = (1 + \mathrm{i}) \sqrt{\frac{\pi}{2}}$$

and

$$\int_{-\infty}^{\infty} \frac{\sin x}{x}\, dx = \pi$$

played an important role in geometric optics and the theory of the Fourier integral, respectively. These two integrals are called the Fresnel integral and the sDirichlet integral, respectively.[54] The existence of these two integrals is closely related to the method of stationary phase (see Sect. 12.5.3 on page 714) which plays a crucial role for Feynman path integrals.

[54] Fresnel (1788–1827), Dirichlet (1805–1859).

(ii) Shock waves. Applying the Euclidean trick $t \mapsto -it$ to the wave equation

$$\frac{\partial^2 \psi}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} = 0 \tag{11.113}$$

and changing the notation, $y := ct$, we get the Laplace equation

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = 0. \tag{11.114}$$

Note that both equations describe completely different physical processes, namely,

- the wave propagation of a mass density ψ on the real line, and
- the stationary temperature distribution ψ on the Euclidean plane $\mathbb{R}^2$.

Whereas the wave equation (11.113) allows the propagation of discontinuities, the solutions of the stationary heat equation (Laplace equation) (11.114) are always smooth. More precisely, the following hold true.

- The function
$$\psi(x,t) := \theta(x - ct)$$
is a solution of the wave equation (11.113) on $\mathbb{R}^2$ in the sense of distributions, by the argument given on page 613. This solution describes the propagation of a singularity located at the point $x = ct$ at time t (see Fig. 11.3 on page 613).
- If the temperature function $\psi : \mathbb{R}^2 \to \mathbb{R}$ is absolutely integrable over every disk and satisfies the stationary heat equation (11.114) on $\mathbb{R}^2$ in the sense of distributions, then ψ is smooth on $\mathbb{R}^2$. More generally, the Weyl lemma on page 611 tells us that each distribution $\psi \in \mathcal{D}'(\mathbb{R}^2)$ with $\Delta\psi = 0$ is a smooth function on $\mathbb{R}^2$.

Example (ii) above tells us that

The Euclidean trick can never be used for describing shock waves.

This fact limits the use of the Euclidean trick in quantum field theory. In terms of mathematics, it is not possible to reduce the theory of hyperbolic partial differential equations to the much simpler theory of elliptic partial differential equations. Observe that quantum fields possess the typical character of hyperbolic partial differential equations.

A far-reaching mathematical analysis of propagating singularities can be based on the modern notion of the wave front set of a distribution (see Sect. 12.5.2 on page 706). In this setting, the wave front set of the solution of a homogeneous elliptic partial differential equation (e.g., the stationary heat equation) is empty, whereas the solutions of hyperbolic partial differential equations (e.g., the wave equation) possess nontrivial wave front sets which are related to the intuitive notion of propagating discontinuities.

12. Distributions and Physics

12.1 The Discrete Dirac Calculus

For understanding the Dirac calculus used by physicists, it is useful to start with a discrete variant of this calculus. This also helps to avoid divergent expressions, say, in quantum electrodynamics. In this chapter, we will use the international SI system of units.

12.1.1 Lattices

The basic idea is to use a truncated lattice in momentum space. To discuss this, let

$$\mathbf{x} = x^1\mathbf{i} + x^2\mathbf{j} + x^3\mathbf{k}$$

be the position vector with respect to a right-handed (x^1, x^2, x^3)-Cartesian coordinate system. Here, $\mathbf{i}, \mathbf{j}, \mathbf{k}$ is a right-handed orthonormal system of vectors. Consider the cube

$$\mathcal{C}(L) := \left\{(x^1, x^2, x^3) \in \mathbb{R}^3 : \ -\tfrac{L}{2} \le x^1, x^2, x^3 \le \tfrac{L}{2}\right\}$$

of side length $L > 0$. It is our goal to put the given physical field in the cube $\mathcal{C}(L)$ and to assume that the field has the period L with respect to the Cartesian coordinates x^1, x^2, x^3. The cube $\mathcal{C}(L)$ has the volume $\mathcal{V} = L^3$ which is called the normalization volume in physics. Furthermore, let $\mathcal{G}$ be a lattice in momentum space. Explicitly, let $\mathcal{G}$ be the set of all momentum vectors

$$\boxed{\mathbf{p} := \frac{2\pi\hbar}{L} \cdot \mathbf{m}} \tag{12.1}$$

with the lattice vectors $\mathbf{m} = m^1\mathbf{i} + m^2\mathbf{j} + m^3\mathbf{k}$ where m^1, m^2, m^3 are integers.

The following terminology is chosen in such a way that it fits well the Dirac calculus to be introduced below.

Orthogonality relation. For each momentum vector $\mathbf{p} \in \mathcal{G}$, the function

$$\boxed{f(\mathbf{x}) := \mathrm{e}^{\mathrm{i}\mathbf{p}\mathbf{x}/\hbar}}$$

has the period L with respect to the Cartesian coordinates x^1, x^2, x^3. Introducing the symbol $d^3x := dx^1dx^2dx^3$, we have the orthogonality relation

$$\frac{1}{L^3}\int_{\mathcal{C}(L)} \mathrm{e}^{\mathrm{i}(\mathbf{p}-\mathbf{q})\mathbf{x}/\hbar}\, d^3x = \delta_{\mathbf{pq}}, \qquad \mathbf{p}, \mathbf{q} \in \mathcal{G}. \tag{12.2}$$

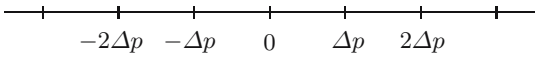


Fig. 12.1. Lattice in one-dimensional momentum space

Here, we set $\delta_{\mathbf{pq}} := 0$ (resp. $=1$) if $\mathbf{p} \neq \mathbf{q}$ (resp. $\mathbf{p} = \mathbf{q}$).

The truncated lattice in momentum space. For $N = 1, 2, 3, \ldots$, let $\mathcal{G}(N)$ denote the set of all momentum vectors (12.1) with

$$|m^1|, |m^2|, |m^3| \leq N.$$

For given complex numbers $a_{\mathbf{p}}$, the function

$$\boxed{f(\mathbf{x}) := \frac{1}{\sqrt{(2\pi\hbar)^3}} \sum_{\mathbf{p}\in\mathcal{G}(N)} a_{\mathbf{p}}\, \mathrm{e}^{\mathrm{i}\mathbf{p}\mathbf{x}/\hbar}\, \Delta^3 p} \tag{12.3}$$

has the period L with respect to the variables x^1, x^2, x^3. Here, we set

$$\Delta^3 p := \Delta p^1 \Delta p^2 \Delta p^3, \qquad \Delta p^j := \frac{2\pi\hbar}{L}, \qquad j = 1, 2, 3.$$

By the orthogonality relation (12.2),

$$\boxed{a_{\mathbf{p}} = \frac{1}{\sqrt{(2\pi\hbar)^3}} \int_{\mathcal{C}(L)} f(\mathbf{x}) \mathrm{e}^{-\mathrm{i}\mathbf{p}\mathbf{x}/\hbar}\, d^3x.} \tag{12.4}$$

We call (12.3) and (12.4) the discrete Fourier transform and the inverse discrete Fourier transform with respect to the lattice $\mathcal{G}(N)$, respectively. Note that

$$f(\mathbf{x})^\dagger := \frac{1}{\sqrt{(2\pi\hbar)^3}} \sum_{\mathbf{p}\in\mathcal{G}(N)} a_{\mathbf{p}}^\dagger \mathrm{e}^{-\mathrm{i}\mathbf{p}\mathbf{x}/\hbar}\, \Delta^3 p.$$

Therefore, if the function f is real-valued, then

$$f(\mathbf{x}) := \frac{1}{2\sqrt{(2\pi\hbar)^3}} \sum_{\mathbf{p}\in\mathcal{G}(N)} (a_{\mathbf{p}} \mathrm{e}^{\mathrm{i}\mathbf{p}\mathbf{x}/\hbar} + a_{\mathbf{p}}^\dagger \mathrm{e}^{-\mathrm{i}\mathbf{p}\mathbf{x}/\hbar})\, \Delta^3 p.$$

12.1.2 The Four-Dimensional Discrete Dirac Delta Function

The discrete delta function in position space. For arbitrary position vectors $\mathbf{x}$ and $\mathbf{y}$, define

$$\boxed{\delta_{\mathcal{C}(L)}(\mathbf{y} - \mathbf{x}) := \frac{1}{(2\pi\hbar)^3} \sum_{\mathbf{p}\in\mathcal{G}(N)} \mathrm{e}^{\mathrm{i}(\mathbf{y}-\mathbf{x})\mathbf{p}/\hbar} \Delta^3 p.} \tag{12.5}$$

The symmetry of the lattice $\mathcal{G}(N)$ under reflections at the origin allows us to replace $\mathbf{p}$ by $-\mathbf{p}$. Hence

$$\delta_{\mathcal{G}(N)}(\mathbf{y}-\mathbf{x}) = \delta_{\mathcal{G}(N)}(\mathbf{x}-\mathbf{y})$$

for all position vectors $\mathbf{x}$ and $\mathbf{y}$. For each function f of the form (12.3), we get the typical property

$$f(\mathbf{y}) = \int_{\mathcal{C}(L)} \delta_{\mathcal{C}(L)}(\mathbf{y}-\mathbf{x}) f(\mathbf{x}) d^3x \tag{12.6}$$

for all position vectors $\mathbf{y}$. This follows from the orthogonality relation (12.2).

The discrete delta function in momentum space. Let $\mathbf{p}, \mathbf{q} \in \mathcal{G}(N)$ be arbitrary lattice points. We define

$$\boxed{\delta_{\mathcal{G}(N)}(\mathbf{q}-\mathbf{p}) = \frac{1}{(2\pi\hbar)^3} \int_{\mathcal{C}(L)} \mathrm{e}^{\mathrm{i}(\mathbf{p}-\mathbf{q})\mathbf{x}/\hbar}\, d^3x.} \tag{12.7}$$

By the orthogonality relation (12.2),

$$\delta_{\mathcal{G}(N)}(\mathbf{p}-\mathbf{q}) := \frac{\delta_{\mathbf{pq}}}{\Delta^3 p}.$$

This is a rescaled Kronecker symbol. Obviously,

$$\delta_{\mathcal{G}(N)}(\mathbf{q}-\mathbf{p}) = \delta_{\mathcal{G}(N)}(\mathbf{p}-\mathbf{q}) \qquad \text{for all} \quad \mathbf{q}, \mathbf{p} \in \mathcal{G}(N).$$

For each function $g : \mathcal{G}(N) \to \mathbb{C}$, this implies

$$g(\mathbf{q}) = \sum_{\mathbf{p}\in\mathcal{G}(N)} \delta_{\mathcal{G}(N)}(\mathbf{q}-\mathbf{p}) g(\mathbf{p})\, \Delta^3 p \tag{12.8}$$

for all lattice points $\mathbf{q} \in \mathcal{G}(N)$.

The 4-dimensional discrete delta function. Consider the 4-dimensional momentum vector

$$p = (p^0, \mathbf{p})$$

where $p^0 := E/c$. Here, E denotes energy. Fix time $T > 0$. The discrete Dirac delta function δ_{dis} in the 4-dimensional momentum space is defined by the product

$$\boxed{\delta_{\mathrm{dis}}(p) := \delta_{cT}(p^0) \cdot \delta_{\mathcal{G}(N)}(\mathbf{p})}$$

with the discrete Dirac delta function $\delta_{\mathcal{G}(N)} : \mathcal{G}(N) \to \mathbb{R}$ in 3-dimensional momentum space,

$$\delta_{\mathcal{G}(N)}(\mathbf{p}) := \begin{cases} \frac{1}{\Delta^3 p} = \frac{\mathcal{V}}{(2\pi\hbar)^3} & \text{if} \quad \mathbf{p} = 0, \\ 0 & \text{if} \quad \mathbf{p} \neq 0, \end{cases} \tag{12.9}$$

and the discrete Dirac delta function $\delta_{cT} : \mathbb{R} \to \mathbb{R}$ in the energy space given by

$$\delta_{cT}(p^0) := \frac{1}{2\pi\hbar} \int_{-cT/2}^{cT/2} \mathrm{e}^{\mathrm{i}p^0 x^0/\hbar} dx^0.$$

Recall that $x^0 := ct$ and $p^0 := E/c$. Explicitly,

$$\delta_{cT}(p^0) = \begin{cases} \frac{cT}{2\pi\hbar} & \text{if} \quad p^0 = 0, \\ \frac{\sin\left(cTp^0/2\hbar\right)}{\pi p^0} & \text{if} \quad p^0 \neq 0. \end{cases}$$

Typical properties. For the computation of cross sections for scattering processes later on, we will use the following key relations.

Proposition 12.1 (i) *For all test functions* $\varphi \in \mathcal{S}(\mathbb{R})$,

$$\lim_{T\to+\infty} \int_{\mathbb{R}} \delta_{cT}(p^0)\varphi(p^0)dp^0 = \varphi(0).$$

This means that $\lim_{T\to+\infty} \delta_{cT} = \delta$, *in the sense of tempered distributions on the real line. Moreover,*

$$\lim_{T\to+\infty} \frac{\delta_{cT}(p^0)}{cT} = \begin{cases} \frac{1}{2\pi\hbar} & \text{if} \quad p^0 = 0, \\ 0 & \text{if} \quad p^0 \neq 0. \end{cases} \tag{12.10}$$

(ii) *For all* $p \in \mathbb{R}^4$, *as* $T \to +\infty$,

$$\delta_{\rm dis}(p)^2 = \frac{cT\mathcal{V}}{(2\pi\hbar)^4} \cdot \delta_{\rm dis}(p)\ (1 + o(1)). \tag{12.11}$$

(iii) *For all real numbers* p_1^0 *and* p_2^0, *as* $T \to +\infty$,

$$\int_{\mathbb{R}} \delta_{cT}(p^0 - p_1^0) \cdot \delta_{cT}(p^0 - p_2^0)dp^0 = \delta_{cT}(p_1^0 - p_2^0)\ (1 + o(1)).$$

For the proof, see Problem 12.6 on page 734. In applications, one frequently uses the energetic system. In this case, the formulas above are simplified by setting $c = \hbar = 1$ and $x^0 = t$ (time), as well as $p^0 = E$ (energy).

Fourier transform. Recall that $px = p^0x^0 - \mathbf{px}$. It follows from

$$\int_{-cT/2}^{cT/2} \mathrm{e}^{\mathrm{i}p^0x^0/\hbar}dx^0 \int_{\mathcal{C}(L)} \mathrm{e}^{-\mathrm{i}\mathbf{px}/\hbar}d^3x = (2\pi\hbar)^4\delta_{cT}(p^0) \cdot \delta_{\mathcal{G}(N)}(\mathbf{p})$$

that

$$\boxed{\frac{1}{(2\pi\hbar)^4} \int_{\Omega} \mathrm{e}^{\mathrm{i}px/\hbar}d^4x = \delta_{\rm dis}(p)} \tag{12.12}$$

where $\Omega := [-\frac{cT}{2}, \frac{cT}{2}] \times \mathcal{C}(L)$. This is a rigorous discrete variant of the mnemonic formula

$$\frac{1}{(2\pi\hbar)^4} \int_{\mathbb{R}^4} \mathrm{e}^{\mathrm{i}px/\hbar}d^4x = \delta^4(p)$$

used by physicists.

The magic square of the Dirac delta function in quantum field theory. In order to compute cross sections, the formula (12.11) is frequently applied by physicists in the following mnemonic form:

$$\boxed{\delta^4(p)^2 = \frac{cT\mathcal{V}}{(2\pi\hbar)^4} \cdot \delta^4(p).} \tag{12.13}$$

In fact, physicists always use the replacement

$$\delta^4(p) \Rightarrow \frac{cT\mathcal{V}}{(2\pi\hbar)^4}$$

where $\mathcal{V}$ and T are a typical volume and a typical time, respectively. We will show in Volume II on quantum electrodynamics that the final cross sections computed via Feynman diagrams do not depend on the choice of the normalization volume $\mathcal{V}$ and the normalization time T.

In this context, let us mention that we will show on page 730 that the square of the Dirac delta distribution does not exist as a reasonable mathematical object. Therefore, we will use the strategy of discrete approximations in this treatise.

12.1.3 Rigorous Discrete Dirac Calculus

Mnemonically, all the discrete formulas above can be elegantly obtained from the following two formal completeness relations in position space and momentum space:

(C1) $\int_{\mathcal{C}(L)} |\mathbf{x}\rangle\langle\mathbf{x}| \ d^3x = I$,
(C2) $\sum_{\mathbf{p}\in\mathcal{G}(N)} |\mathbf{p}\rangle\langle\mathbf{p}| \ \Delta^3 p = I$.

Formal relations. From these formal completeness relations, we immediately obtain the following relations.

(i) Discrete Fourier transform: $\langle\mathbf{x}|f\rangle = \sum_{\mathbf{p}\in\mathcal{G}(N)} \langle\mathbf{x}|\mathbf{p}\rangle\langle\mathbf{p}|f\rangle \ \Delta^3 p$.
(ii) Inverse discrete Fourier transform: $\langle\mathbf{p}|f\rangle = \int_{\mathcal{C}(L)} \langle\mathbf{p}|\mathbf{x}\rangle\langle\mathbf{x}|f\rangle \ d^3x$.
(iii) Dirac's discrete delta function in position space:

$$\langle\mathbf{y}|\mathbf{x}\rangle = \int_{\mathbf{p}\in\mathcal{G}(N)} \langle\mathbf{y}|\mathbf{p}\rangle\langle\mathbf{p}|\mathbf{x}\rangle \Delta^3 p$$

and $\langle\mathbf{y}|f\rangle = \int_{\mathcal{C}(L)} \langle\mathbf{y}|\mathbf{x}\rangle\langle\mathbf{x}|f\rangle \ d^3x$.
(iv) Dirac's discrete delta function in momentum space:

$$\langle\mathbf{q}|\mathbf{p}\rangle = \int_{\mathcal{C}(N)} \langle\mathbf{q}|\mathbf{x}\rangle\langle\mathbf{x}|\mathbf{p}\rangle \ d^3x$$

and $\langle\mathbf{q}|f\rangle = \int_{\mathbf{p}\in\mathcal{G}(N)} \langle\mathbf{q}|\mathbf{p}\rangle\langle\mathbf{p}|f\rangle \ \Delta^3 p$.

Rigorous formulas. The formal relations (i)-(iv) above correspond to rigorous formulas if we introduce the following notation:

- $\langle\mathbf{x}|f\rangle := f(\mathbf{x})$ and $\langle f|\mathbf{x}\rangle = \langle\mathbf{x}|f\rangle^\dagger$,
- $\langle\mathbf{p}|f\rangle := a_{\mathbf{p}}$ and $\langle f|\mathbf{p}\rangle = \langle\mathbf{p}|f\rangle^\dagger$,
- $\langle\mathbf{x}|\mathbf{p}\rangle := \mathrm{e}^{\mathrm{i}\mathbf{x}\mathbf{p}/\hbar}/(2\pi\hbar)^{3/2}$ and $\langle\mathbf{p}|\mathbf{x}\rangle = \langle\mathbf{x}|\mathbf{p}\rangle^\dagger$,
- $\langle\mathbf{y}|\mathbf{x}\rangle := \delta_{\mathcal{C}(L)}(\mathbf{y}-\mathbf{x})$,
- $\langle\mathbf{q}|\mathbf{p}\rangle := \delta_{\mathcal{G}(N)}(\mathbf{q}-\mathbf{p})$.

Furthermore, the fact that $\langle\mathbf{y}|\mathbf{x}\rangle \in \mathbb{R}$ implies the symmetry property

$$\langle\mathbf{y}|\mathbf{x}\rangle = \langle\mathbf{y}|\mathbf{x}\rangle^\dagger = \langle\mathbf{x}|\mathbf{y}\rangle.$$

Similarly, $\langle\mathbf{q}|\mathbf{p}\rangle = \langle\mathbf{q}|\mathbf{p}\rangle^\dagger = \langle\mathbf{p}|\mathbf{q}\rangle$.

12.1.4 The Formal Continuum Limit

> Mnemonically, Dirac's delta function is the formal lattice limit of the discrete Dirac delta function.
>
> Folklore

Later on we will carry out the following two limits.

(H) High-energy limit: $N \to +\infty$.
(L) Low-energy limit: $L \to +\infty$.

In terms of the momentum vector $\mathbf{p}$, observe that the limits (H) and (L) correspond to $|\mathbf{p}| \to \infty$ and $\Delta p^j \to 0$, respectively.

> *Our notation has to be chosen in such a way that these two limits lead to the classical formulas for the Fourier integral.*

In fact, from (12.3) and (12.4) we get the Fourier transform

$$f(\mathbf{x}) := \frac{1}{\sqrt{(2\pi\hbar)^3}} \int_{\mathbb{R}^3} a_{\mathbf{p}}\, \mathrm{e}^{\mathrm{i}\mathbf{p}\mathbf{x}/\hbar}\, d^3p$$

along with the inverse Fourier transform

$$a_{\mathbf{p}} = \frac{1}{\sqrt{(2\pi\hbar)^3}} \int_{\mathbb{R}^3} f(\mathbf{x})\mathrm{e}^{-\mathrm{i}\mathbf{p}\mathbf{x}/\hbar}\, d^3x.$$

Applying the limits $N \to +\infty$ and $L \to +\infty$ to (12.5) and (12.6), physicists write formally

$$\delta(\mathbf{y}-\mathbf{x}) := \frac{1}{(2\pi\hbar)^3} \int_{\mathbb{R}^3} \mathrm{e}^{\mathrm{i}(\mathbf{y}-\mathbf{x})\mathbf{p}/\hbar} d^3p \tag{12.14}$$

and

$$f(\mathbf{y}) = \int_{\mathbb{R}^3} \delta(\mathbf{y}-\mathbf{x}) f(\mathbf{x})\, d^3x. \tag{12.15}$$

Formula (12.7) in momentum space passes over to

$$\delta(\mathbf{q}-\mathbf{p}) = \frac{1}{(2\pi\hbar)^3} \int_{\mathbb{R}^3} \mathrm{e}^{\mathrm{i}(\mathbf{p}-\mathbf{q})\mathbf{x}/\hbar}\, d^3x.$$

Finally, from (12.8) we get

$$g(\mathbf{q}) = \int_{\mathbb{R}^3} \delta(\mathbf{q}-\mathbf{p})\, g(\mathbf{p}) d^3p.$$

This coincides with (12.14) and (12.15), respectively, by using the substitution $\mathbf{x} \Rightarrow -\mathbf{x}$.

In a straightforward manner, the discrete Dirac calculus can be generalized to the continuous case by carrying out the limits $N \to \infty$ and $L \to +\infty$ in a formal way. In particular, discrete delta functions in position and momentum space pass over to Dirac's delta function. Formally,

$$\boxed{\delta_{\mathcal{C}(L)}(\mathbf{x}) \to \delta(\mathbf{x}), \qquad \delta_{\mathcal{G}(N)}(\mathbf{p}) \to \delta(\mathbf{p}).}$$

The discrete completeness relations pass over to the following formal continuous completeness relations:

(C1) $\int_{\mathbb{R}^3} |\mathbf{x}\rangle\langle\mathbf{x}|\ d^3x = I$,
(C2) $\int_{\mathbb{R}^3} |\mathbf{p}\rangle\langle\mathbf{p}|\ d^3p = I$.

From these completeness relations, we formally obtain the following relations.

(i) Fourier transform: $\langle\mathbf{x}|f\rangle = \int_{\mathbb{R}^3} \langle\mathbf{x}|\mathbf{p}\rangle\langle\mathbf{p}|f\rangle\, d^3p$.
(ii) Inverse Fourier transform: $\langle\mathbf{p}|f\rangle = \int_{\mathbb{R}^3} \langle\mathbf{p}|\mathbf{x}\rangle\langle\mathbf{x}|f\rangle\, d^3x$.
(iii) Dirac's delta function in position space:

$$\langle\mathbf{y}|\mathbf{x}\rangle = \int_{\mathbb{R}^3} \langle\mathbf{y}|\mathbf{p}\rangle\langle\mathbf{p}|\mathbf{x}\rangle d^3p$$

and $\langle\mathbf{y}|f\rangle = \int_{\mathbb{R}^3} \langle\mathbf{y}|\mathbf{x}\rangle\langle\mathbf{x}|f\rangle\, d^3x$.

(iv) Dirac's delta function in momentum space:

$$\langle \mathbf{q}|\mathbf{p}\rangle = \int_{\mathbb{R}^3} \langle \mathbf{q}|\mathbf{x}\rangle\langle \mathbf{x}|\mathbf{p}\rangle \; d^3x$$

and $\langle \mathbf{q}|f\rangle = \int_{\mathbb{R}^3} \langle \mathbf{q}|\mathbf{p}\rangle\langle \mathbf{p}|f\rangle \; d^3p$.

This corresponds to the classical formulas for the Fourier transform if we introduce the following notation:

- $\langle \mathbf{x}|f\rangle := f(\mathbf{x})$ and $\langle f|\mathbf{x}\rangle = \langle \mathbf{x}|f\rangle^\dagger$,
- $\langle \mathbf{p}|f\rangle := a_{\mathbf{p}}$ and $\langle f|\mathbf{p}\rangle = \langle \mathbf{p}|f\rangle^\dagger$,
- $\langle \mathbf{x}|\mathbf{p}\rangle := \mathrm{e}^{\mathrm{i}\mathbf{x}\mathbf{p}/\hbar}/(2\pi\hbar)^{3/2}$ and $\langle \mathbf{p}|\mathbf{x}\rangle = \langle \mathbf{x}|\mathbf{p}\rangle^\dagger$,
- $\langle \mathbf{y}|\mathbf{x}\rangle := \delta(\mathbf{y}-\mathbf{x})$,
- $\langle \mathbf{q}|\mathbf{p}\rangle := \delta(\mathbf{q}-\mathbf{p})$.

Moreover, $\langle \mathbf{y}|\mathbf{x}\rangle = \langle \mathbf{y}|\mathbf{x}\rangle^\dagger = \langle \mathbf{x}|\mathbf{y}\rangle$ and $\langle \mathbf{q}|\mathbf{p}\rangle = \langle \mathbf{q}|\mathbf{p}\rangle^\dagger = \langle \mathbf{p}|\mathbf{q}\rangle$.

12.2 Rigorous General Dirac Calculus

We now want to show how the formal Dirac calculus introduced in Sect. 11.2 on page 589 can be given a rigorous meaning by using the notion of a Gelfand triplet (also called rigging of a Hilbert space).

12.2.1 Eigendistributions

The chain of inclusions

$$\mathcal{S}(\mathbb{R}) \subset L_2(\mathbb{R}) \subset \mathcal{S}'(\mathbb{R})$$

is called a Gelfand triplet with respect to the Hilbert space $L_2(\mathbb{R})$ (or a rigging of $L_2(\mathbb{R})$). Let us introduce the momentum operator

$$P : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$$

given by $(P\varphi)(x) := -\mathrm{i}\hbar\frac{d}{dx}\varphi(x)$ and the position operator

$$X : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$$

given by $(X\varphi)(x) := x\varphi(x)$ for all $x \in \mathbb{R}$ and all functions $\varphi \in \mathcal{S}(\mathbb{R})$. For a quantum particle on the real line, the position operator X and the momentum operator P do not possess eigenfunctions which lie in the Hilbert space $L_2(\mathbb{R})$. For example, setting $\varphi_p(x) := \mathrm{e}^{\mathrm{i}px/\hbar}/\sqrt{2\pi\hbar}$, we get

$$P\varphi_p = p\varphi_p \qquad \text{for all} \quad p \in \mathbb{R},$$

but $\int_{\mathbb{R}} |\varphi_p(x)|^2 dx = \infty$. Hence $\varphi_p \notin L_2(\mathbb{R})$. However, we will show that there exists a complete system of eigendistributions.

The momentum operator. Define $F_p(\varphi) := \int_{\mathbb{R}} \varphi_p(x)^\dagger \varphi(x) dx$ for all test functions $\varphi \in \mathcal{S}(\mathbb{R})$. Then, $F_p \in \mathcal{S}'(\mathbb{R})$.

The family $\{F_p\}_{p\in\mathbb{R}}$ of tempered distributions F_p forms a complete system of eigendistributions of the momentum operator P.

Explicitly, this means the following.

(i) Eigenvalue relation: For each real number p, we have

$$\boxed{F_p(P\varphi) = pF_p(\varphi) \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).} \tag{12.16}$$

(ii) Completeness relation: If

$$F_p(\varphi) = 0 \qquad \text{for all} \quad p \in \mathbb{R}$$

and fixed $\varphi \in \mathcal{S}(\mathbb{R})$, then $\varphi = 0$.

Let us prove this. Integration by parts yields

$$\int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}px/\hbar}(-\mathrm{i}\hbar\varphi'(x))dx = p\int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}px/\hbar}\varphi(x)dx.$$

This implies (i). Moreover, in order to prove (ii) let

$$F_p(\varphi) = \frac{1}{\sqrt{2\pi\hbar}}\int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}px/\hbar}\varphi(x)dx = 0 \qquad \text{for all} \quad p \in \mathbb{R}.$$

For the Fourier transform, $\hat{\varphi}(p) = F_p(\varphi) = 0$. Hence $\varphi = 0$. Using Dirac's notation to be thoroughly discussed on page 680, equation (12.16) is equivalent to

$$\boxed{P|F_p\rangle = p|F_p\rangle \qquad \text{for all} \quad p \in \mathbb{R}.}$$

Physicists briefly write $|p\rangle$ instead of $|F_p\rangle$.

The position operator. Similarly, we obtain that

The family $\{\delta_x\}_{x\in\mathbb{R}}$ of tempered distributions δ_x forms a complete system of eigendistributions of the position operator X.

Explicitly, this means the following.

(i) Eigenvalue relation: For each number $x \in \mathbb{R}$, we get

$$\boxed{\delta_x(X\varphi) = x\delta_x(\varphi) \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).} \tag{12.17}$$

(ii) Completeness relation: If $\delta_x(\varphi) = 0$ for all $x \in X$ and fixed $\varphi \in \mathcal{S}(\mathbb{R})$, then $\varphi = 0$.

This follows from $\delta_x(\varphi) = \varphi(x)$ and $\delta_x(X\varphi) = x\varphi(x)$. Using Dirac's notation to be introduced on page 680, equation (12.17) is equivalent to

$$\boxed{X|\delta_x\rangle = x|\delta_x\rangle \qquad \text{for all} \quad x \in \mathbb{R}.}$$

Physicists briefly write $|x\rangle$ instead of $|\delta_x\rangle$.

The energy operator. Define the Hamiltonian $H : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ by setting

$$H := \frac{P^2}{2m_e}.$$

This is the energy operator for the free motion of an electron of mass m_e on the real line.

The family $\{F_p\}_{p\in\mathbb{R}}$ represents a complete system of eigendistributions.

Explicitly, this means the following:

(i) Eigenvalue relation: For each real number p, we have

$$\boxed{F_p(H\varphi) = E(p)F_p(\varphi) \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).} \tag{12.18}$$

Here, the eigenvalue $E(p) := p^2/2m_e$ is the energy of a free particle of momentum p.

(ii) Completeness relation: If

$$F_p(\varphi) = 0 \qquad \text{for all} \quad p \in \mathbb{R}$$

and fixed $\varphi \in \mathcal{S}(\mathbb{R})$, then $\varphi = 0$.

Using Dirac's notation, equation (12.18) is equivalent to

$$\boxed{H|p\rangle = E(p)|p\rangle \qquad \text{for all} \quad p \in \mathbb{R}.}$$

12.2.2 Self-Adjoint Operators

It was discovered by John von Neumann around 1928 that the notion of self-adjoint operator on a Hilbert space plays a fundamental role in the mathematical approach to quantum physics. Let X be a complex Hilbert space. The operator $A : D(A) \to X$ is called self-adjoint iff

- the domain of definition $D(A)$ is a linear subspace of the Hilbert space X which is dense, that is, for each $\varphi \in X$, there exists a sequence (φ_n) in $D(A)$ such that $\lim_{n\to\infty} \varphi_n = \varphi$ in X;
- the operator A is linear and formally self-adjoint, that is,

$$\langle \varphi | A\psi \rangle = \langle A\varphi | \psi \rangle \qquad \text{for all} \quad \varphi, \psi \in D(A);$$

- the two operators $(A \pm \mathrm{i}I) : D(A) \to X$ are surjective.

The operator $B : D(B) \to X$ is called essentially self-adjoint iff there exists precisely one self-adjoint operator $A : D(A) \to X$ which is an extension of B, that is, $D(B) \subseteq D(A) \subseteq X$ and $A\varphi = B\varphi$ for all $\varphi \in D(B)$.

Proposition 12.2 *The three operators*

$$P, X, H : \mathcal{S}(\mathbb{R}) \to L_2(\mathbb{R}) \tag{12.19}$$

introduced in the preceding section are essentially self-adjoint on the Hilbert space $L_2(\mathbb{R})$.

The idea of proof is the following.[1] We first consider the natural extensions

$$\boxed{P, X, H : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})} \tag{12.20}$$

where derivative and multiplication are to be understood in the sense of tempered distributions. We then introduce the restricted domains of definition:

- $D(P) := \{\varphi \in L_2(\mathbb{R}) : \ P\varphi \in L_2(\mathbb{R})\}$;
- $D(X) := \{\varphi \in L_2(\mathbb{R}) : \ X\varphi \in L_2(\mathbb{R})\}$;
- $D(H) := \{\varphi \in L_2(\mathbb{R}) : \ H\varphi \in L_2(\mathbb{R})\}$.

The corresponding restrictions of the operators from (12.20), namely,

$$P : D(P) \to L_2(\mathbb{R}), \quad X : D(X) \to L_2(\mathbb{R}), \quad H : D(H) \to L_2(\mathbb{R})$$

are then the desired self-adjoint extensions of the operators from (12.19).

[1] The details can be found in Zeidler (1995), Vol. 1, p. 263ff and p. 414ff.

12.2.3 The von Neumann Spectral Theorem

The classical Fourier transform sends differentiation to multiplication. This nice property is crucial for the investigation of differential equations. John von Neumann showed around 1930 that the Fourier transform can be generalized to self-adjoint operators in Hilbert spaces. The two key relations read as follows:

(i) Parseval equation:

$$\langle \varphi | \psi \rangle = \int_{\mathcal{M}} \hat{\varphi}(m)^\dagger \hat{\psi}(m) d\mu(m) \qquad \text{for all} \quad \varphi, \psi \in X.$$

(ii) Multiplication operator:

$$\boxed{(\hat{A}\hat{\varphi})(m) = \lambda(m)\hat{\varphi}(m) \qquad \text{for all} \quad m \in \mathcal{M}}$$

where $\lambda : \mathcal{M} \to \mathbb{R}$ is a fixed function. We also set

$$\hat{\varphi}(m) := (\mathcal{F}\varphi)(m) \qquad \text{for all} \quad m \in \mathcal{M}.$$

The function $\hat{\varphi} : \mathcal{M} \to \mathbb{C}$ is called the generalized Fourier transform of the element φ in the Hilbert space X.

A Hilbert space X is called separable iff there exists a sequence $\varphi_1, \varphi_2, \ldots$ in X such that for each element φ of X, there exists a subsequence $\varphi_{1'}, \varphi_{2'}, \ldots$ which converges to φ in X. For example, the Hilbert space $L_2(\mathbb{R}^N)$ is separable for all $N = 1, 2, \ldots$.

Theorem 12.3 *Let $A : D(A) \to X$ be a linear self-adjoint operator defined on the dense subset $D(A)$ of the complex separable Hilbert space X. Then there exist both a Hilbert space $L_2(\mathcal{M}, \mu)$ and a unitary operator*

$$\mathcal{F} : X \to L_2(\mathcal{M}, \mu)$$

such that the operator A is transformed into the multiplication operator $\hat{A}$ given by (ii) *above.*

Let us explain the precise meaning of this theorem. We start with a measure μ on the nonempty set $\mathcal{M}$. By definition, the space $L_2(\mathcal{M}, \mu)$ consists of all μ-measurable functions

$$\hat{\varphi} : \mathcal{M} \to \mathbb{C}$$

such that $\int_{\mathcal{M}} |\hat{\varphi}(m)|^2 d\mu(m) < \infty$. The space $L_2(\mathcal{M}, \mu)$ becomes a complex Hilbert space with respect to the inner product

$$\langle \varphi | \psi \rangle := \int_{\mathcal{M}} \varphi(m)^\dagger \psi(m) d\mu(m).$$

Here, two functions φ and ψ represent the same element of the Hilbert space $L_2(\mathcal{M}, \mu)$ iff they differ on a set of μ-measure zero. Equation (ii) above is to be understood in the sense of the following commutative diagram:

$$\begin{array}{ccc} D(A) & \xrightarrow{A} & X \\ {\scriptstyle \mathcal{F}}\downarrow & & \downarrow{\scriptstyle \mathcal{F}} \\ D(\hat{A}) & \xrightarrow{\hat{A}} & L_2(\mathcal{M}, \mu). \end{array}$$

By definition, the function $\hat{\varphi}$ belongs to the set $D(\hat{A})$ iff

$$\hat{\varphi} \in L_2(\mathcal{M}, \mu) \quad \text{and} \quad \lambda\hat{\varphi} \in L_2(\mathcal{M}, \mu).$$

It turns out that $D(\hat{A}) = \mathcal{F}(D(A))$. Moreover, $\hat{A} = \mathcal{F}A\mathcal{F}^{-1}$.

For the proof of Theorem 12.3, we refer to Reed, Simon (1972), Vol. 1, Sect. VIII.3 and Berezin, Shubin (1991), supplement 1. The latter proof is elegantly based on the Riemann–Hilbert problem.

12.2.4 The Gelfand–Kostyuchenko Spectral Theorem

We now want to consider a refinement of the von Neumann spectral theorem which generalizes the results from Sect. 12.2.1 on eigendistributions. For the dimensions $N = 1, 2, \ldots$ we will use the Gelfand triplet

$$\mathcal{S}(\mathbb{R}^N) \subset L_2(\mathbb{R}^N) \subset \mathcal{S}'(\mathbb{R}^N).$$

The following theorem is a special case of a general result obtained by Gelfand and Kostyuchenko in 1955. The proof can be found in Gelfand and Shilov (1964), Vol. 4, Sect. I.4.

Theorem 12.4 *Let $A : \mathcal{S}(\mathbb{R}^N) \to \mathcal{S}(\mathbb{R}^N)$ be a linear, sequentially continuous operator which is essentially self-adjoint with respect to the Hilbert space $L_2(\mathbb{R}^N)$. Then, this operator has a complete system $\{F_m\}_{m\in\mathcal{M}}$ of eigendistributions.*

Explicitly, this means the following. There exists a nonempty set $\mathcal{M}$ such that $F_m \in \mathcal{S}'(\mathbb{R}^N)$ for all indices $m \in \mathcal{M}$.

(i) Eigendistributions: There exists a function $\lambda : \mathcal{M} \to \mathbb{R}$ such that

$$\boxed{F_m(A\varphi) = \lambda(m)F_m(\varphi)}$$

for all indices $m \in \mathcal{M}$ and all test functions $\varphi \in \mathcal{S}(\mathbb{R}^N)$.

(ii) Completeness: If $F_m(\varphi) = 0$ for all $m \in \mathcal{M}$ and fixed $\varphi \in \mathcal{S}(\mathbb{R}^N)$, then $\varphi = 0$.

The function $\hat{\varphi} : \mathcal{M} \to \mathbb{C}$ defined by

$$\hat{\varphi}(m) := F_m(\varphi) \qquad \text{for all} \quad m \in \mathcal{M}$$

is called the generalized Fourier transform of the function $\varphi \in \mathcal{S}(\mathbb{R}^N)$, with respect to the operator A.

12.2.5 The Duality Map

As a preparation for the formulation of the rigorous Dirac calculus in the next section, let us discuss the two key inclusion chains

$$\boxed{\mathcal{S}(\mathbb{R}^N) \subset L_2(\mathbb{R}^N) \subset \mathcal{S}'(\mathbb{R}^N)} \tag{12.21}$$

and

$$\mathcal{S}(\mathbb{R}^N) \subset L_2(\mathbb{R}^N) \subset \mathcal{S}'_a(\mathbb{R}^N) \tag{12.22}$$

along with the duality map

$$\boxed{J : L_2(\mathbb{R}^N) \to \mathcal{S}'(\mathbb{R}^N)} \tag{12.23}$$

and the antiduality map

$$J_a : \mathcal{S}'(\mathbb{R}^N) \to \mathcal{S}'_a(\mathbb{R}^N). \tag{12.24}$$

Here, $N = 1, 2, \ldots$. Let us summarize the relevant definitions.

(i) The Hilbert space $L_2(\mathbb{R}^N)$. Recall that the inner product on $L_2(\mathbb{R}^N)$ is given by

$$\langle \varphi | \psi \rangle := \int_{\mathbb{R}^N} \varphi(x)^\dagger \psi(x) d^N x \qquad \text{for all} \quad \varphi, \psi \in L_2(\mathbb{R}^N).$$

It is our goal below to extend this inner product to more general mathematical objects called costates and generalized states. The set $\mathcal{S}(\mathbb{R}^N)$ of rapidly increasing functions is dense in $L_2(\mathbb{R}^N)$.

(ii) Duality map: For each function $\psi \in L_2(\mathbb{R}^N)$, define

$$J(\psi)(\varphi) := \langle \psi | \varphi \rangle \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}^N).$$

For all $\psi, \chi \in L_2(\mathbb{R}^N)$ and all complex numbers α, β,

$$J(\alpha\psi + \beta\chi) = \alpha^\dagger J\psi + \beta^\dagger J\chi.$$

Moreover, if $J\psi = 0$, then $\psi = 0$. These relations tell us that the duality map $J : L_2(\mathbb{R}^N) \to \mathcal{S}'(\mathbb{R}^N)$ is antilinear and injective.

(iii) The antiduality map. By definition, the space of antidistributions $\mathcal{S}'_a(\mathbb{R}^N)$ consists of all antilinear, sequentially continuous maps[2] $G : \mathcal{S}(\mathbb{R}^N) \to \mathbb{C}$. For each tempered distribution $F \in \mathcal{S}'(\mathbb{R}^N)$, define

$$(J_a F)(\varphi) := F(\varphi)^\dagger \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}^N).$$

Then, $J_a F \in \mathcal{S}'_a(\mathbb{R}^N)$. For all $F, H \in \mathcal{S}'(\mathbb{R}^N)$ and all $\alpha, \beta \in \mathbb{C}$,

$$J_a(\alpha F + \beta H) = \alpha^\dagger J_a F + \beta^\dagger J_a H.$$

Thus, the antiduality map $J_a : \mathcal{S}'(\mathbb{R}^N) \to \mathcal{S}'_a(\mathbb{R}^N)$ is antilinear and bijective.

12.2.6 Dirac's Notation

Concerning the Gelfand triplet $\mathcal{S}(\mathbb{R}^N) \subset L_2(\mathbb{R}^N) \subset \mathcal{S}'(\mathbb{R}^N)$, let us introduce the following terminology.

- Each function $\psi \in L_2(\mathbb{R}^N)$ is called a state. We write $|\psi\rangle$ instead of ψ.
- Each linear functional $F \in \mathcal{S}'(\mathbb{R}^N)$ is called a costate. We write $\langle F|$ instead of F.
- Each antilinear functional $G \in \mathcal{S}'_a(\mathbb{R}^N)$ is called a generalized state. We write $|G\rangle$ instead of G.

[2] By definition, G is antilinear iff $G(\alpha\varphi + \beta\psi) = \alpha^\dagger G(\varphi) + \beta^\dagger G(\psi)$ for all complex numbers α, β and all $\varphi, \psi \in \mathcal{S}(\mathbb{R}^N)$.

To each state $\psi \in L_2(\mathbb{R}^N)$, we assign the costate $\langle\psi|$ given by the linear functional $\varphi \mapsto \langle\psi|\varphi\rangle$. Thus, the map

$$\boxed{|\psi\rangle \mapsto \langle\psi|}$$

is identical to the duality map $\psi \mapsto J\psi$. Furthermore, to each costate $\langle F|$ we assign the generalized state $|F\rangle$ given by the antilinear functional $\varphi \mapsto F(\varphi)^\dagger$. Thus, the map

$$\boxed{\langle F| \mapsto |F\rangle}$$

is equal to the antiduality map $F \mapsto J_a F$. Let the map $A : \mathcal{S}(\mathbb{R}^N) \to \mathcal{S}(\mathbb{R}^N)$ be formally self-adjoint with respect to the Hilbert space $L_2(\mathbb{R}^N)$. This means that

$$\langle\varphi|A\psi\rangle = \langle A\varphi|\psi\rangle \qquad \text{for all} \quad \varphi, \psi \in \mathcal{S}(\mathbb{R}^N).$$

We extend this by setting

$$\langle\varphi|A|G\rangle := \langle A\varphi|G\rangle \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}^N), \;\; G \in \mathcal{S}'_a(\mathbb{R}^N).$$

This defines the generalized state $A|G\rangle$ given by the antilinear functional

$$\varphi \mapsto \langle A\varphi|G\rangle.$$

In particular, fix $F \in \mathcal{S}'(\mathbb{R}^N)$ and the real number λ. Then, the equation

$$\boxed{F(A\varphi) = \lambda F(\varphi) \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}^N)} \tag{12.25}$$

is equivalent to the following three equations to be considered for all functions $\varphi \in \mathcal{S}(\mathbb{R}^N)$:

- $\langle F|A\varphi\rangle = \lambda\langle F|\varphi\rangle$;
- $\langle F|A\varphi\rangle^\dagger = \lambda\langle F|\varphi\rangle^\dagger$;
- $\langle A\varphi|F\rangle = \lambda\langle\varphi|F\rangle$.

Therefore, equation (12.25) is equivalent to $A|F\rangle = \lambda|F\rangle$.

12.2.7 The Schwartz Kernel Theorem

Consider the integral operator $\chi = K\varphi$ given by

$$\chi(x) = \int_{\mathbb{R}^2} \mathcal{K}(x,y)\varphi(y)dy, \qquad x \in \mathbb{R}.$$

The function $\mathcal{K}$ is called the kernel of the operator K. More precisely, if the function $\mathcal{K} : \mathbb{R}^2 \to \mathbb{C}$ is Lebesgue measurable and $\int_{\mathbb{R}^2} |\mathcal{K}(x,y)|^2 dxdy < \infty$, then the operator $K : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ is linear and continuous. Setting

$$B(\psi,\varphi) := \int_{\mathbb{R}^2} \mathcal{K}(x,y)\psi(x)\varphi(y)dxdy,$$

we obtain a bilinear functional $B : \mathcal{S}(\mathbb{R}) \times \mathcal{S}(\mathbb{R}) \to \mathbb{C}$. The following theorem is a far-reaching generalization of this result. It shows that a broad class of bilinear functionals have tempered distributions as their generalized kernels. The key formula reads as follows:

$$\boxed{B(\psi,\varphi) = F(\psi \otimes \varphi) \qquad \text{for all} \quad \psi \in \mathcal{S}(\mathbb{R}^N), \varphi \in \mathcal{S}(\mathbb{R}^M).} \tag{12.26}$$

Theorem 12.5 *Let $N, M = 1, 2, \ldots,$ and assume that the functional*

$$B : \mathcal{S}(\mathbb{R}^N) \times \mathcal{S}(\mathbb{R}^M) \to \mathbb{C}$$

is both linear and sequentially continuous with respect to each argument. Then, there exists a tempered distribution $F \in \mathcal{S}'(\mathbb{R}^{N+M})$ such that B can be represented by (12.26).

This theorem is the special case of a general kernel theorem valid in nuclear spaces. See Gelfand and Shilov (1964), Vol. 3, Chap. 1.

12.3 Fundamental Limits in Physics

Limits play a crucial role in physics. The idea is to approximate complicated phenomena by simpler ones. For example, we have the following limits:

- high-energy limit,
- low-energy-limit,
- thermodynamic limit and phase transitions,
- adiabatic limit and regularization,
- the limit from wave optics to geometric optics for short wavelengths of light,
- the limit from Einstein's relativistic physics to Newton's non-relativistic physics for sufficiently low velocities, and
- the limit from quantum physics to classical physics for sufficiently large action.

Let us discuss some of the basic ideas.

12.3.1 High-Energy Limit

Experiments in particle accelerators are carried out at a fixed energy scale E per particle. The high-energy limit corresponds to $E \to +\infty$. Such extremely high energies were present in the early universe shortly after the Big Bang. However, the high-energy limit and the low-energy limit $E \to +0$ are also crucial in quantum field theory in order to pass from lattices to the continuum limit. This is related to the procedure of renormalization. For example, quarks behave like free particles at very high particle energies.

12.3.2 Thermodynamic Limit and Phase Transitions

The most spectacular phenomena of thermodynamic systems are phase transitions. For example, the freezing of water to ice at the temperature 0° Celsius represents a phase transition. Mathematically, phase transitions can be studied by singularities (e.g. jumps) of thermodynamic quantities. As a rule, such singularities do not appear in thermodynamic systems of finite volume. One has to perform the limit

$$\boxed{V \to \infty,}$$

i.e., the volume V of the system has to go to infinity. This limit is called the thermodynamic limit. The importance of this limit in mathematical physics was emphasized by David Ruelle in his monograph *Statistical Mechanics: Rigorous Results,* New York, 1969. We also recommend the survey article by Griffith (1972) (rigorous results) and the monograph by Minlos (2000) (mathematical statistical

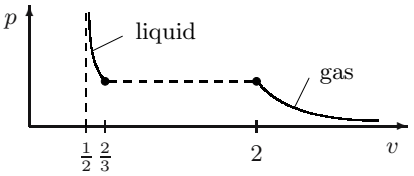


Fig. 12.2. Liquid-vapor isothermal line

physics). An extensive bibliography on statistical physics can be found in Emch and Liu (2002) (1500 references). A collection of seminal papers in 20th century statistical physics is contained in Stroke (1995).

The condensation of a gas. If one increases the pressure of a real gas at a fixed, sufficiently low temperature, then the gas condenses into a liquid. This is pictured schematically in Fig. 12.2. Here, p denotes the pressure, and v denotes the specific volume (i.e., volume per molecule).

The Yang–Lee condensation model. In order to understand Fig. 12.2 in terms of statistical physics, let us consider a purely mathematical model which is simple, but far from physical reality.[3] Let us start with the partition function

$$Z(T,V,\mu)=\sum_{N=0}^{2V} z^N \sum_{j=0}^{N} \binom{V}{N-j} = (1+z)^V \frac{1-z^{V+1}}{1-z}$$

where $z := \mathrm{e}^{\mu/kT}$. Here, we use the following notation: T temperature, μ chemical potential, V volume, k Boltzmann constant. Choose $V = 1, 2, \ldots$ By statistical mechanics, we get the pressure p and the mean particle number per volume, $\overline{N}/V$, by the formulas

$$p = kT \frac{\partial \ln Z}{\partial V} \tag{12.27}$$

and

$$\frac{\overline{N}}{V} = \frac{kT}{V} \frac{\partial \ln Z}{\partial \mu} = z \frac{\partial}{\partial z} \left(\frac{\ln Z}{V} \right). \tag{12.28}$$

Letting $V \to +\infty$, we will replace the pressure by the limit

$$p(z) := kT \lim_{V \to +\infty} \frac{\ln Z}{V}\,.$$

Moreover, motivated by (12.28) as $V \to +\infty$, we set

$$\frac{1}{v(z)} = \frac{\overline{N}}{V} := z \frac{\partial}{\partial z} \left(\frac{p(z)}{kT} \right).$$

An elementary computation yields

[3] More general results can be found in the following two papers:
C. Yang and T. Lee, Statistical theory of equations of state and phase transitions. I. Theory of condensation. II. Lattice gas and Ising model, Phys. Rev. **87** (1952), 404–409, 410–419.

$$\frac{p(z)}{kT} = \begin{cases} \ln(1+z) & \text{if} \quad |z| < 1, \\ \ln(1+z)z & \text{if} \quad |z| > 1, \end{cases}$$

and

$$v(z) = \begin{cases} \dfrac{1+z}{z} & \text{if} \quad |z| < 1, \\ \dfrac{1+z}{1+2z} & \text{if} \quad |z| > 1. \end{cases}$$

Let us discuss this. The physical region of the fugacity $z = \mathrm{e}^{\mu/kT}$ is given by $\mu \in \mathbb{R}$ and $T > 0$. Hence $0 < z < \infty$. Equation (12.27) shows that

The zeros of the partition function Z are critical.

Explicitly, the zeros of $Z = (1+z)^V \frac{1-z^{V+1}}{1-z}$ are given by the equation

$$1 - z^{V+1} = 0$$

except for $z = 1$. The zeros lie on the unit circle. As $V \to \infty$, the zeros fill a dense subset of the unit circle. This is the mechanism for getting a phase transition at $z = 1$ in the thermodynamic limit $V \to \infty$. In fact, the pressure remains continuous at $z = 1$, but the specific volume v jumps at $z = 1$, namely,

$$\lim_{z \to 1+0} v(z) = 2, \qquad \lim_{z \to 1-0} v(z) = \tfrac{2}{3}\,.$$

This allows us to regard the region $1 < z < \infty$ as *gas* (resp. $0 < z < 1$ as *liquid*). The jump of v corresponds to the condensation of the gas. This reflects the obvious fact that the specific volume of a gas is larger than that of a liquid. Moreover, the equation of state $p = f(v, T)$ reads as

$$p = \begin{cases} kT \ln \dfrac{v}{v-1} & \text{if} \quad |z| < 1, \\ kT \ln \dfrac{v(1-v)}{(2v-1)^2} & \text{if} \quad |z| > 1. \end{cases}$$

This is pictured in Fig. 12.2 on page 683 for fixed temperature T.

The importance of phase transitions in high technology processes. Phase transitions play a crucial role in understanding strange properties of matter. For example, at sufficiently low temperatures one observes

- superconductivity,
- superfluidity (e.g., liquid helium), and
- condensation of Bose–Einstein gases.

For example, Cornell, Ketterle, and Wieman were awarded the Nobel prize in physics in 2001 for the achievement of Bose–Einstein condensation in dilute gases of alkali atoms. As an introduction to the statistical physics of these phenomena, we recommend the classical lectures given by Feynman (1998) (14th edition).

Phase transitions in the early universe. Physicists also assume that the cooling of the universe after the Big Bang caused several phase transitions which were responsible for the splitting of the original unified force into gravitational, strong, weak, and electromagnetic interaction. In the setting of the inflationary theory, it is assumed that shortly after the Big Bang, a phase transition caused an enormous sudden expansion of the universe which is responsible for the almost flatness of the present universe. Moreover, the strange properties of the ground state of Fermi gases lead to the existence of neutron stars and white dwarfs which possess extreme mass densities. For this, we refer to Straumann (2004).

12.3.3 Adiabatic Limit

The basic idea is to compute integrals as limits of regularized integrals. The key formula reads as

$$\boxed{\int_0^\infty a(t)dt = \lim_{\varepsilon\to+0}\int_0^\infty a(t)\mathrm{e}^{-\varepsilon t}dt.} \tag{12.29}$$

This means the following. In order to compute the integral $\int_0^\infty a(t)dt$, we consider the regularized integral

$$A(\varepsilon) := \int_0^\infty a(t)\mathrm{e}^{-\varepsilon t}dt, \qquad \varepsilon > 0,$$

and we compute the limit (12.29). In terms of physics, we switch on the perturbation $a(t) \Rightarrow a(t)\mathrm{e}^{-\varepsilon t}$. For each time $t \geq 0$,

$$\lim_{\varepsilon\to+0} a(t)\mathrm{e}^{-\varepsilon t} = a(t).$$

In addition, for each parameter $\varepsilon > 0$, $\lim_{t\to+\infty} \mathrm{e}^{-\varepsilon t} = 1$. Physicists call this adiabatic perturbation. The point is that the adiabatic limit (12.29) does not always exist. In terms of mathematics, the function $A = A(\varepsilon)$ is called the Laplace transform of the given time-dependent function $a = a(t)$. The justification of the limit (12.29) corresponds to sophisticated Tauberian theorems for the Laplace transform.[4] As a typical example, consider the integral

$$J := \lim_{T\to+\infty}\int_0^T \frac{\sin t}{t}\,dt.$$

We will show below that

$$J = \lim_{\varepsilon\to+0}\int_0^\infty \frac{\sin t}{t}\cdot \mathrm{e}^{-\varepsilon t}\,dt = \frac{\pi}{2}.$$

This famous integral was used by Dirichlet (1809–1859) in order to investigate the convergence of Fourier integrals. Observe the following peculiarity. The integral

$$A(\varepsilon) := \int_0^\infty \frac{\sin t}{t}\,\mathrm{e}^{-\varepsilon t}dt$$

exists as Lebesgue integral for each complex number ε with $\Re(\varepsilon) > 0$. This follows from

$$\int_0^\infty \left|\frac{\sin t}{t}\,\mathrm{e}^{-\varepsilon t}\right| dt \leq \mathrm{const}\int_0^\infty \mathrm{e}^{-t\Re(\varepsilon)}\,dt < \infty,$$

by the majorant criterion on page 529. Moreover, the function $A = A(\varepsilon)$ is holomorphic on the open upper half-plane $\{\varepsilon \in \mathbb{C} : \Re(\varepsilon) > 0\}$. The situation changes critically at the value $\varepsilon = 0$. Then

$$\int_0^\infty \left|\frac{\sin t}{t}\right| dt = \infty.$$

[4] N. Wiener, Tauberian theorems, Ann. Math. **33** (2) (1933), 1–100.
J. Korevaar, Tauberian Theory: A Century of Developments, Springer, Berlin, 2004.

Therefore, the integral $A(0)$ is not absolutely convergent, and hence it does not exist as a Lebesgue integral; it is a so-called improper integral.

Newman's adiabatic limit theorem. The following theorem due to Newman (1980) is a crucial ingredient in the proof of the prime number theorem on page 291.

Theorem 12.6 *Let $a : [0,\infty[\to \mathbb{R}$ be a continuous (or, more generally, locally integrable) and bounded function. Suppose that the function*

$$A(\varepsilon) := \int_0^\infty a(t)e^{-\varepsilon t}\, dt, \qquad \Re(\varepsilon) > 0$$

extends holomorphically to an open neighborhood of the closed upper half-plane given by $\{\varepsilon \in \mathbb{C} : \Re(\varepsilon) \geq 0\}$. Then, the integral

$$\int_0^\infty a(t)dt$$

is convergent. Moreover, this integral is equal to $A(0)$.

Since the extended function A is continuous at the origin, we get

$$\int_0^\infty a(t)dt = \lim_{\varepsilon \to +0} A(\varepsilon).$$

Proof. Our proof follows Zagier (1996). For $T > 0$ set

$$A_T(z) := \int_0^T a(t)\mathrm{e}^{-zt}dt, \qquad z \in \mathbb{C}.$$

This function is clearly holomorphic on the complex plane. We must show that $\lim_{T\to+\infty} A_T(0) = A(0)$ where $A(0)$ denotes the extended value of the original function $\varepsilon \mapsto A(\varepsilon)$.

Let the radius R be a large real number and let C be the boundary of the compact set

$$D := \{z \in \mathbb{C} : \ |z| \leq R, \ \ \Re(z) \geq -\delta, \}$$

where $\delta > 0$ is small enough (depending on R) so that the extended function A is holomorphic on an open neighborhood of D. Then

$$A(0) - A_T(0) = \frac{1}{2\pi\mathrm{i}} \int_C (A(z) - A_T(z))\mathrm{e}^{zT} \left(1 + \frac{z^2}{R^2}\right) \frac{dz}{z}. \tag{12.30}$$

This key formula is a special case of Cauchy's integral formula for the representation of a holomorphic function by its boundary values (see (4.3) page 213). We now study the integral (12.30) on different parts of the curve C.

(I) On the semicircle $C_+ := C \cap \{z \in \mathbb{C} : \ \Re(z) > 0\}$, the integrand is bounded by $2B/R^2$, where $B := \sup_{t\geq 0} |a(t)|$, because

$$|A(z) - A_T(z)| = \left| \int_T^\infty a(t)\mathrm{e}^{-zt}dt \right| \leq B \int_T^\infty |\mathrm{e}^{-zt}|dt = \frac{B\mathrm{e}^{-\Re(z)T}}{\Re(z)}$$

if $\Re(z) > 0$. Furthermore,[5]

$$\left| \mathrm{e}^{zT} \left(1 + \frac{z^2}{R^2} \right) \frac{1}{z} \right| = \mathrm{e}^{\Re(z)T} \cdot \frac{2\Re(z)}{R^2}.$$

Hence the contribution to $A(0) - A_T(0)$ from the integral over C_+ is bounded in absolute value by B/R.

(II) For the integral over $C_- := C \cap \{z \in \mathbb{C} : \Re(z) < 0\}$, we look at $A(z)$ and $A_T(z)$ separately. Since A_T is holomorphic on the complex plane, the path of integration for the integral involving A_T can be replaced by the semicircle

$$C'_- := \{z \in \mathbb{C} : \ |z| = R, \ \Re(z) < 0\},$$

and the integral over C'_- is then bounded in absolute value by $2\pi B/R$ by exactly the same estimate as before since

$$|A_T(z)| = \left| \int_0^T a(t) \mathrm{e}^{-zt} dt \right| \leq B \int_{-\infty}^T |\mathrm{e}^{-zt}| dt = \frac{B \mathrm{e}^{-\Re(z)T}}{|\Re(z)|}, \qquad \Re(z) < 0.$$

(III) Finally, the remaining integral over C_- tends to zero as $T \to +\infty$ because the integrand is the product of the function $A(z)(1 + \frac{z^2}{R^2})\frac{1}{z}$, which is independent of T, and the function e^{zT}, which goes to zero rapidly and uniformly on compact sets as $T \to +\infty$ in the half-plane $\{z \in \mathbb{C} : \ \Re(z) < 0\}$. Hence

$$\limsup_{T \to +\infty} |A(0) - A_T(0)| \leq \frac{2B}{R}.$$

Since R is arbitrary, this proves the theorem. □

Adiabatic limit theorem of Hardy–Littlewood type. The following result is the special case of a fundamental Tauberian theorem for the Laplace transform which can be found in Korevaar (2004), p. 30.

Theorem 12.7 *The large-time limit*

$$\lim_{T \to +\infty} \int_0^T a(t) dt = A(+0)$$

exists if the following conditions are satisfied.

(H1) *The function* $a : [0, \infty[\to \mathbb{R}$ *is continuous, and the large-time limit*

$$A(\varepsilon) := \lim_{T \to +\infty} \int_0^T a(t) e^{-t\varepsilon} dt.$$

exists for each positive number ε.

(H2) *The finite limit* $A(+0) = \lim_{\varepsilon \to +0} A(\varepsilon)$ *exists.*

(H3) $\inf_{t>0} \ ta(t) > -\infty$.

[5] Note that $z = R\mathrm{e}^{\mathrm{i}\varphi}$, with the angle $-\pi/2 < \varphi < \pi/2$, implies

$$\left| \frac{1}{z} + \frac{z}{R} \right| = \frac{1}{R} \left| \mathrm{e}^{-\mathrm{i}\varphi} + \mathrm{e}^{\mathrm{i}\varphi} \right| = \frac{2\cos\varphi}{R} = \frac{2\Re(z)}{R^2}.$$

Application to the Dirichlet integral. For each $\varepsilon > 0$, consider the regularized Dirichlet integral

$$A(\varepsilon) := \int_0^\infty \frac{\sin t}{t}\, \mathrm{e}^{-\varepsilon t} dt.$$

Then

$$A'(\varepsilon) = -\int_0^\infty \mathrm{e}^{-\varepsilon t} \sin t \;\, dt = -\frac{1}{1+\varepsilon^2}, \qquad \varepsilon > 0.$$

This implies

$$A(\varepsilon) = -\int_\varepsilon^\infty A'(\eta) d\eta = \lim_{\eta \to +\infty} (\arctan \eta - \arctan \varepsilon) = \frac{\pi}{2} - \arctan \varepsilon.$$

Setting $a(t) := \frac{\sin t}{t}$, it follows from Theorem 12.7 that

$$\lim_{T \to +\infty} \int_0^T \frac{\sin t}{t}\, dt = \lim_{\varepsilon \to +0} A(\varepsilon) = \frac{\pi}{2}.$$

We will show in Problem 12.2 that the same result can be obtained for the Dirichlet integral by using Cauchy's residue method.

Adiabatic regularization of the Fourier transform of the Heaviside function. Typically, if ω is real parameter, then the oscillating integral

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \theta(t) \mathrm{e}^{\mathrm{i}\omega}\, dt = \frac{1}{\sqrt{2\pi}} \int_0^\infty \mathrm{e}^{\mathrm{i}\omega}\, dt$$

does not exist in the classical sense. Consequently, the classical Fourier transform of the Heaviside function does not exist. However, the Heaviside function θ represents a tempered distribution given by

$$\Theta(\varphi) := \int_{-\infty}^\infty \theta(t) \varphi(t) dt$$

for all test functions $\varphi \in \mathcal{S}(\mathbb{R})$. This implies the existence of the Fourier transform $\mathcal{F}(\Theta)$ as a tempered distribution. To simplify notation, we will write $\mathcal{F}(\theta)$ instead of $\mathcal{F}(\Theta)$. The basic trick of our approach is to replace the Heaviside function θ by the adiabatic approximation

$$\theta_\varepsilon(t) := \theta(t) \mathrm{e}^{-\varepsilon t}, \qquad t \in \mathbb{R},$$

and to study the limit $\varepsilon \to +0$. In mathematics, this is called Abelian regularization due to Abel (1802–1829). Physicists speak of adiabatic regularization introduced by Gell-Mann and Low (1951) for studying bound quantum states via S-matrix. The classical Fourier transform of θ_ε reads as

$$\mathcal{F}(\theta_\varepsilon)(\omega) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \mathrm{e}^{\mathrm{i}(\omega + \mathrm{i}\varepsilon)t} dt = \frac{\mathrm{i}}{\sqrt{2\pi}} \cdot \frac{1}{\omega + \varepsilon \mathrm{i}}$$

for all real numbers ω.

Proposition 12.8 *For the Fourier transform of the Heaviside function θ,*

$$\mathcal{F}(\theta) = \lim_{\varepsilon \to +0} \mathcal{F}(\theta_\varepsilon) = \frac{\mathrm{i}}{\sqrt{2\pi}} \cdot \frac{1}{\omega + 0_+ \mathrm{i}},$$

in the sense of tempered distributions.

Proof. For each test function $\varphi \in \mathcal{S}(\mathbb{R})$,

$$\lim_{\varepsilon \to +0} \int_{-\infty}^{\infty} \theta_\varepsilon(t)\varphi(t)dt = \int_{-\infty}^{\infty} \theta(t)\varphi(t)dt.$$

Thus, $\theta_\varepsilon \to \theta$ as $\varepsilon \to +0$, in the sense of tempered distributions. Since the Fourier transform $\mathcal{F} : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ is sequentially continuous,

$$\mathcal{F}(\theta) = \lim_{\varepsilon \to +0} \mathcal{F}(\theta_\varepsilon).$$

□

Recall that the Sokhotski formula (11.48) on page 620 tells us that

$$\frac{1}{\omega + 0_+\mathrm{i}} = \mathcal{P}\left(\frac{1}{\omega}\right) - \mathrm{i}\pi\delta.$$

12.3.4 Singular Limit

There exist the following crucial limiting processes:

(i) $\lambda \to 0$ (the wave length of light goes to zero): the passage from Maxwell's theory of electromagnetism to Fermat's geometric optics;
(ii) $\hbar \to 0$ (the Planck constant goes to zero): the passage from quantum mechanics to classical mechanics;
(iii) $c \to \infty$ (the velocity of light goes to infinity): the passage from the theory of relativity to classical mechanics.

Since the quantities $\lambda, \hbar, c$ carry physical dimensions, they depend on the choice of the unit system. Thus, more precisely, one has to use relative dimensionless quantities. For example, consider a physical experiment with visible light. Let L be a typical length scale of the experiment, and let λ be the wave length of light. The limit (i) corresponds then to

$$\frac{\lambda}{L} \to 0.$$

In (ii) and (iii), we need a typical velocity V and a typical action S (energy times time), respectively, and we have to study the limits

$$\frac{V}{c} \to 0, \qquad \frac{\hbar}{S} \to 0.$$

In daily life, we have the following typical ratios

$$\frac{\lambda}{L} \sim 10^{-6}, \qquad \frac{V}{c} \sim 10^{-8}, \qquad \frac{\hbar}{S} \sim 10^{-34}.$$

Here, we use the length $L = 1\mathrm{m}$, the velocity $V = 1\mathrm{m/s}$, and the action

$$S = 1\mathrm{kg} \cdot \mathrm{m} \cdot \mathrm{s}.$$

Consequently, relativistic effects and quantum effects can be neglected in daily life, and we can approximately apply the methods of geometric optics. For radio waves, X rays, γ-rays, and cosmic rays, we have

$$\frac{\lambda}{L} \sim l$$

with $l = 10^3,\ 10^{-10},\ 10^{-12},\ 10^{-15}$, respectively. Again, $L := 1\mathrm{m}$. The singular limits (i)-(iii) play an important role in physics. We will study them later on. For example, the short-wavelength limit for electromagnetic waves will be investigated in Sect. 12.5.4 on page 718. This is the prototype of a singular limit.

12.4 Duality in Physics

The concept of duality is crucial in both mathematics and physics.
Folklore

The goal is to relate apparently different problems to each other via duality in order to simplify the mathematical treatment.[6] There exist the following fundamental dualities in quantum physics related to the Fourier transform and the Laplace transform:

- particles and waves,
- time and frequency,
- time and energy,
- position and momentum,
- causality and analyticity,
- strong and weak interaction.

Let us sketch the basic ideas.

12.4.1 Particles and de Broglie's Matter Waves

De Broglies fundamental duality principle. Particles and waves are fundamental objects in physics for describing the transport of energy.

For quantum particles, there exists a duality between particle properties and wave properties.

In 1905 Einstein postulated that the frequency ν and the wave length λ of an electromagnetic light wave are related to the energy E and the momentum vector $\mathbf{p}$ of the corresponding light particle (photon) by

$$E = h\nu, \qquad \mathbf{p} = \frac{h}{\lambda}\,\mathbf{n}$$

with the unit vector $\mathbf{n}$. Equivalently,

$$\boxed{E = \hbar\omega, \qquad \mathbf{p} = \hbar\mathbf{k}.} \tag{12.31}$$

In his famous 1924 dissertation at the Sorbonne in Paris, de Broglie postulated that the duality principle (12.31) represents a universal law in quantum physics. In particular, he claimed that each massive or massless quantum particle possesses wave properties. For an electron of rest mass m_0 and velocity vector $\mathbf{v}$, Einstein's theory of special relativity tells us that

$$\mathbf{p} = \frac{m_0\mathbf{v}}{\sqrt{1 - \mathbf{v}^2/c^2}}.$$

Suppose that the electron moves slowly compared with the velocity of light, that is, $|\mathbf{v}|/c \ll 1$. Then we get the non-relativistic approximation $\mathbf{p} = m_0\mathbf{v}$. Introducing the Compton wave length of the electron, $\lambda_C := h/m_0c$, we obtain

$$\lambda = \frac{c}{|\mathbf{v}|}\,\lambda_C = \frac{c}{|\mathbf{v}|} \cdot 2.43 \cdot 10^{-12}\mathrm{m}$$

[6] A general duality theory with many application to problems in mathematics and physics can be found in Zeidler (1986), Vols. III and IV.

for the wave length λ of the electron. In 1927 Davisson, Germer, Thomson, and Reid demonstrated de Broglie's hypothesis on the wave character of the electron by studying the diffraction of electrons at crystals experimentally.[7] The powerful electron microscope is based on the wave character of the electron. The smallest wave length of visible light is about $400 \cdot 10^{-9}$ m $= 400$ nanometers. The wave length of the electron in an electron microscope is about $5 \cdot 10^{-9}$m $= 5$ nanometers. Therefore, an electron microscope is much more effective than a light microscope. The enlargement factor of an electron microscope is about 1 000 000.[8]

The Klein–Gordon equation for matter waves. De Broglie's matter wave can be formally described by a complex-valued wave function

$$\psi(\mathbf{x}, t) = A\mathrm{e}^{\mathrm{i}(\mathbf{px}-Et)/\hbar} \tag{12.32}$$

along with $E = \hbar\omega, \mathbf{p} = \hbar\mathbf{k}$ and the Einstein relation

$$E^2 = m_0^2c^4 + c^2\mathbf{p}^2. \tag{12.33}$$

The latter plays the role of a dispersion relation. Differentiation of ψ yields

$$\mathrm{i}\hbar\frac{\partial\psi}{\partial t} = E\psi, \qquad -\mathrm{i}\hbar\boldsymbol{\partial}\psi = \mathbf{p}\psi.$$

Consequently, if we use the substitution

$$\boxed{E \Rightarrow \mathrm{i}\hbar\frac{\partial}{\partial t}, \qquad \mathbf{p} \Rightarrow -\mathrm{i}\hbar\boldsymbol{\partial},} \tag{12.34}$$

then from the Einstein relation (12.33) we obtain the so-called Klein–Gordon equation

$$\Big(\frac{1}{c^2}\frac{\partial^2}{\partial^2 t} - \boldsymbol{\partial}^2 + \frac{m_0^2c^2}{\hbar^2}\Big)\psi = 0. \tag{12.35}$$

This construction guarantees that the wave function ψ from (12.32) is a solution of the Klein–Gordon equation. Explicitly, the Klein–Gordon equation can be written as

$$\frac{1}{c^2}\,\psi_{tt} - \psi_{xx} - \psi_{yy} - \psi_{zz} + \frac{m_0^2c^2}{\hbar^2}\,\psi = 0.$$

The substitution (12.34) is of great importance for quantum physics. As we will show later on, this elegant quantization rule also generates both the Schrödinger equation for non-relativistic electrons and the Dirac equation for relativistic electrons.

[7] Prince Louis-Victor de Broglie was awarded the 1929 Nobel prize in physics for his discovery of the wave nature of electrons; Clinton Davisson and Sir George Thomson were awarded the 1937 Nobel prize in physics for their electron diffraction experiments.

[8] Ernst Ruska was awarded the 1986 Nobel prize in physics for designing the first electron microscope in 1933, and Gerd Binnig and Heinrich Rohrer were awarded the 1986 Nobel prize in physics for their design of the scanning tunneling microscope in 1981, which is based on the quantum tunneling of particles.

12.4.2 Time and Frequency

> Many time-dependent processes in physics and engineering become simple in the frequency space via the Fourier transform.
>
> Folklore

Let $f \in \mathcal{S}(\mathbb{R})$. The function $a : \mathbb{R} \to \mathbb{C}$ given by

$$a(\omega) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)\, \mathrm{e}^{\mathrm{i}\omega t}\, dt \qquad \text{for all} \quad \omega \in \mathbb{R}$$

is called the Fourier transform of the function f from time space to angular frequency space. Recall that $a \in \mathcal{S}(\mathbb{R})$. The inverse transformation is given by

$$\boxed{f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} a(\omega) \mathrm{e}^{-\mathrm{i}\omega t}\, d\omega \qquad \text{for all} \quad t \in \mathbb{R}.}$$

This transformation represents a duality between time and frequency.

12.4.3 Time and Energy

Setting $E := \hbar\omega$ and $b(E) := a\left(\frac{E}{\hbar}\right)/\sqrt{\hbar}$, we get

$$b(E) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} f(t) \mathrm{e}^{\mathrm{i}Et/\hbar} dt, \qquad \text{for all} \quad E \in \mathbb{R}.$$

The map $f \mapsto b$ is called the Fourier transform from time space to energy space. The inverse transformation reads as

$$\boxed{f(t) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} b(E) \mathrm{e}^{-\mathrm{i}Et/\hbar} dE \qquad \text{for all} \quad t \in \mathbb{R}.} \tag{12.36}$$

For each real number E, we have

$$\mathrm{i}\hbar \frac{d}{dt}\left(\frac{\mathrm{e}^{-\mathrm{i}Et/\hbar}}{\sqrt{2\pi\hbar}}\right) = E \cdot \frac{\mathrm{e}^{-\mathrm{i}Et/\hbar}}{\sqrt{2\pi\hbar}} \qquad \text{for all} \quad t \in \mathbb{R}.$$

Formula (12.36) represents the time-dependent function $f \in \mathcal{S}(\mathbb{R})$ as a superposition of the eigenfunctions of Schrödinger's energy operator $\mathrm{i}\hbar\frac{d}{dt}$.

12.4.4 Position and Momentum

> In terms of quantum physics, the Fourier transform relates the position space to the momentum space.
>
> Folklore

We are given the function $\psi : \mathbb{R} \to \mathbb{C}$ which lies in the space $\mathcal{S}(\mathbb{R})$. Define

$$c(p) := \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \psi(x) \mathrm{e}^{-\mathrm{i}px/\hbar}\, dx \qquad \text{for all} \quad p \in \mathbb{R}. \tag{12.37}$$

Then, we have $c \in \mathcal{S}(\mathbb{R})$. The map $\psi \mapsto c$ is called the Fourier transform from the 1-dimensional position space to the 1-dimensional momentum space. The inverse transformation reads as

$$\boxed{\psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} c(p)\, \mathrm{e}^{\mathrm{i}px/\hbar} dp \qquad \text{for all} \quad x \in \mathbb{R}.} \tag{12.38}$$

For each real number p, we have

$$-\mathrm{i}\hbar \frac{d}{dx} \left(\frac{\mathrm{e}^{\mathrm{i}px/\hbar}}{\sqrt{2\pi\hbar}} \right) = p \cdot \frac{\mathrm{e}^{\mathrm{i}px/\hbar}}{\sqrt{2\pi\hbar}} \qquad \text{for all} \quad p \in \mathbb{R}. \tag{12.39}$$

Formula (12.38) represents the position function $\psi \in \mathcal{S}(\mathbb{R})$ as a superposition of the eigenfunctions of the momentum operator $-\mathrm{i}\hbar\frac{d}{dx}$. The normalization factor $\frac{1}{\sqrt{2\pi}}$ is chosen in such a way that we have the following formal continuous orthonormality condition

$$\int_{-\infty}^{\infty} \frac{\mathrm{e}^{-\mathrm{i}px/\hbar}}{\sqrt{2\pi\hbar}} \cdot \frac{\mathrm{e}^{\mathrm{i}p_0 x/\hbar}}{\sqrt{2\pi\hbar}}\, dx = \delta(p - p_0) \qquad \text{for all} \quad p, p_0 \in \mathbb{R}.$$

Quantum particle on the real line. We want to show that the Fourier transform represents a duality between position and momentum in quantum mechanics which culminates in the Heisenberg uncertainty relation. To begin with, consider the Hilbert space $L_2(\mathbb{R})$ with the inner product

$$\langle \chi | \psi \rangle := \int_{-\infty}^{\infty} \chi(x)^{\dagger} \psi(x)\, dx$$

and the norm

$$\|\psi\|^2 := \langle \psi | \psi \rangle = \int_{-\infty}^{\infty} |\psi(x)|^2\, dx.$$

Each function $\psi : \mathbb{R} \to \mathbb{C}$ with $\psi \in L_2(\mathbb{R})$ and the normalization condition $\|\psi\| = 1$ describes the state of a quantum particle on the real line where

$$\int_a^b |\psi(x)|^2\, dx$$

is the probability for finding the particle on the interval $[a, b]$. Let us introduce the position operator $X : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$,

$$\big(X\psi\big)(x) := x\psi(x) \qquad \text{for all} \quad x \in \mathbb{R},$$

and the momentum operator $P : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$,

$$\big(P\psi\big)(x) := -\mathrm{i}\hbar \frac{d\psi(x)}{dx} \qquad \text{for all} \quad x \in \mathbb{R}.$$

Then, for the particle state $\psi \in \mathcal{S}(\mathbb{R})$, the quantities

$$\bar{x} := \langle \psi | X\psi \rangle, \qquad (\Delta x)^2 := \langle \psi | (X - \bar{x})^2 \psi \rangle$$

and

$$\bar{p} := \langle \psi | P\psi \rangle, \qquad (\Delta p)^2 := \langle \psi | (P - \bar{p})^2 \psi \rangle$$

Fig. 12.3. Gaussian distribution

represent the mean value $\bar{x}$ and the mean fluctuation Δx of the particle position, as well as the mean value $\bar{p}$ and the mean fluctuation Δp of the particle momentum. Explicitly,

$$\bar{x} = \int_{-\infty}^{\infty} x\,|\psi(x)|^2\,dx, \qquad (\Delta x)^2 = \int_{-\infty}^{\infty} (x-\bar{x})^2 |\psi(x)|^2\,dx.$$

Using the Fourier transform,

$$\bar{p} = \int_{-\infty}^{\infty} p\,|c(p)|^2\,dp, \qquad (\Delta p)^2 = \int_{-\infty}^{\infty} (p-\bar{p})^2 |c(p)|^2\,dp.$$

By the Parseval equality, $\int_{-\infty}^{\infty} |c(p)|^2\,dp = \int_{-\infty}^{\infty} |\psi(x)|^2\,dx = 1$. In the terminology of probability theory, the two functions

$$x \mapsto |\psi(x)|^2 \quad \text{and} \quad p \mapsto |c(p)|^2$$

are the probability density functions of the random variables position x and momentum p of the quantum particle on the real line in the state ψ, respectively. The two functions

$$x \mapsto \int_{-\infty}^{x} |\psi(\xi)|^2\,d\xi \quad \text{and} \quad p \mapsto \int_{-\infty}^{p} |c(\eta)|^2\,d\eta$$

are the distribution functions of position x and momentum p of the quantum particle in the state ψ, respectively.

The Gaussian distribution. The most important distribution function in the the theory of probability is the Gaussian distribution. The central limit theorem due to Chebyshev (1821–1894) tells us that, roughly speaking, the Gaussian probability distribution always appears if the random variable under consideration is the superposition of many independent random effects. Explicitly, consider the quantum state

$$\psi(x) := \frac{\mathrm{e}^{-\frac{(x-\mu)^2}{4\sigma^2}}}{\sqrt{\sigma}\,\sqrt[4]{2\pi}}\,. \tag{12.40}$$

We then obtain the Gaussian distribution

$$|\psi(x)|^2 = \frac{\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\,\sqrt{2\pi}}\,, \qquad x \in \mathbb{R}$$

with the mean value $\bar{x} = \mu$ and the mean fluctuation $\Delta x = \sigma$ (Fig. 12.3). For the Fourier transform of the function ψ, we get

$$c(p) = \frac{\mathrm{e}^{-\frac{p^2}{4\sigma'^2}}}{\sqrt{\sigma'}\,\sqrt[4]{2\pi}}\,\mathrm{e}^{\mathrm{i}\bar{x}p/\hbar}$$

with $\sigma' = \hbar/2\sigma$. After rescaling, this follows from the well-known fact that the normalized Gaussian distribution $\frac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$ remains invariant under the Fourier transform. Hence

$$|c(p)|^2 = \frac{\mathrm{e}^{-\frac{p^2}{2\sigma'^2}}}{\sigma'\sqrt{2\pi}} \qquad \text{for all} \quad p \in \mathbb{R}.$$

Therefore, the Gauss state ψ from (12.40) possesses the mean momentum $\bar{p} = 0$ and the momentum fluctuation

$$\Delta p = \frac{\hbar}{2\Delta x}\ .$$

If Δx is small, then Δp is large. In other words, for a sharply located quantum particle, the momentum values p fluctuate strongly. This is a special case of the Heisenberg uncertainty relation

$$\boxed{\Delta x \Delta p \geq \frac{\hbar}{2}}$$

which is valid for each state $\varphi \in \mathcal{S}(\mathbb{R})$ (see Prop. 10.3 on page 523). The relation between harmonic analysis and Heisenberg's uncertainty inequality is thoroughly discussed in the article by Fefferman (1983).

12.4.5 Causality and Analyticity

> Linear response and causality force analyticity, and hence the Kramers–Kronig dispersion relations.
>
> Folklore

In the 1940s, physicists discovered that the polarization of the ground state of the quantum field in quantum electrodynamics is responsible for crucial physical effects (e.g., the Lamb shift of the hydrogen spectrum). In the late 1950s, physicists used dispersion relations for describing scattering processes for elementary particles via analyticity properties of the S-matrix. This led to the emergence of string theory in the 1970s. Let us discuss the basic ideas going back to classical electrodynamics. The key words are

- electric dipole and polarization,
- dielectricity of material media, and
- dispersion of light.

Our goal is to discuss the Kramers–Kronig dispersion relations which play a crucial role in all physical processes which are governed by linear response and causality. This concerns a broad class of phenomena in physics and engineering. Heisenberg's foundation of quantum mechanics in 1925 was strongly influenced by his joint 1924 paper with Kramers on the dispersion of light and earlier papers by Kramers.[9] Generally, dispersion means that processes depend on the energy spectrum. Singularities arise in the energy space which are generated by resonances. For scattering

[9] These papers along with historical comments can be found in the collection of articles edited by van der Waerden (1968).

processes of elementary particles, resonance appears for the energies of stable or unstable bound states of particles.

The Maxwell equations for material media. The basic equations read as

$$\boxed{\begin{aligned} &\operatorname{div}\mathbf{D} = \varrho, && \operatorname{div}\mathbf{B} = 0,\\ &\frac{\partial \mathbf{D}}{\partial t} = \mathbf{curl}\,\mathbf{H} - \mathbf{J}, && \frac{\partial \mathbf{B}}{\partial t} = -\,\mathbf{curl}\,\mathbf{E} \end{aligned}} \tag{12.41}$$

along with the constitutive laws

$$\mathbf{P} = \mathbf{P}(\mathbf{E}), \qquad \mathbf{M} = \mathbf{M}(\mathbf{B}) \tag{12.42}$$

and

$$\mathbf{D} = \varepsilon_0\mathbf{E} + \mathbf{P}(\mathbf{E}), \qquad \mathbf{H} = \frac{1}{\mu_0}\big(\mathbf{B} - \mathbf{M}(\mathbf{B})\big). \tag{12.43}$$

Here, we use the following notation:

- $\mathbf{E}$ electric field vector,,
- $\mathbf{B}$ magnetic field vector,
- ϱ electric charge density,
- $\mathbf{J}$ electric current density vector,
- $\mathbf{P}$ polarization (electric dipole moment density),
- $\mathbf{M}$ magnetization (magnetic dipole moment density),
- $\mathbf{D}$ electric intensity,
- $\mathbf{H}$ magnetic intensity,
- ε_0 electric field constant of a vacuum, and
- μ_0 magnetic field constant of a vacuum.

For the velocity of light in a vacuum,

$$\boxed{c = \frac{1}{\sqrt{\varepsilon_0\mu_0}}\,.}$$

In the SI system, $\mu_0 = 4\pi \cdot 10^{-7}\mathrm{Vs/Am}$. Hence $\varepsilon_0 = 1/\mu_0 c^2$. Naturally enough, in a vacuum we have $\mathbf{P} \equiv 0$ and $\mathbf{M} \equiv 0$. Therefore,

$$\mathbf{D} = \varepsilon_0\mathbf{E}, \qquad \mathbf{B} = \mu_0\mathbf{H}.$$

In material media, the molecules generate additional dipole moments described by $\mathbf{P}$ and $\mathbf{M}$ which depend on the electric field $\mathbf{E}$ and the magnetic field $\mathbf{B}$. Observe that the four Maxwell equations (12.41) are divided into two pairs of equations.

> *The two Maxwell equations from* (12.41) *which contain sources (i.e., electric charge density* ϱ *and electric current density vector* $\mathbf{J}$*) depend on the field intensities* $\mathbf{D}$ *and* $\mathbf{H}$*; the remaining two Maxwell equations depend on the fields* $\mathbf{E}$ *and* $\mathbf{B}$*.*

In Volume III on gauge field theory, we will study the relativistic formulation of the Maxwell equations (12.41). It turns out that $\mathbf{E}$ and $\mathbf{B}$ form the electric field tensor in Minkowski space, whereas $\mathbf{D}$ and $\mathbf{H}$ form the Weyl density tensor. The constitutive law (12.43) is a map of the form $(\mathbf{E}, \mathbf{B}) \mapsto (\mathbf{D}, \mathbf{H})$.

Energy. The conservation laws of the Maxwell equations will be thoroughly studied in Volume III, along with conservation laws for other important field theories in physics (like the Standard Model in particle physics and the theory of general relativity). At this point, let us only mention that

$$\eta = \frac{1}{2}\big(\mathbf{ED} + \mathbf{BH}\big)$$

represents the energy density of the electromagnetic field, and the Poynting vector $\mathbf{E} \times \mathbf{H}$ represents the energy current density vector.

Linear material media. If we have the constitutive laws

$$\mathbf{P} = \chi_e \mathbf{E}, \qquad \mathbf{M} = \chi_m \mathbf{B},$$

then we speak of linear material media. Here, the material constants χ_e and χ_m are called the electric and magnetic susceptibility. From the constitutive law (12.43) we obtain a relation between the electric field $\mathbf{E}$ and the electric intensity $\mathbf{D}$ (resp. the magnetic field $\mathbf{B}$ and the magnetic intensity $\mathbf{H}$). Explicitly,

$$\mathbf{D} = \varepsilon \mathbf{E}, \qquad \mathbf{H} = \frac{1}{\mu}\,\mathbf{B}.$$

Here, ε and μ are called the electric field constant and the magnetic field constant of the material medium. It follows as in Sect. 2.2.4 on page 85 that the velocity of light in the material medium is given by

$$\boxed{c = \frac{1}{\sqrt{\varepsilon\mu}}\,.}$$

It is crucial that ε and μ often depend on the angular frequency ω of the incoming electromagnetic wave. This way, the velocity of light in the material medium depends on the angular frequency, that is,

$$\boxed{c = c(\omega).}$$

This physical effect is called dispersion. As an example, consider water. For a constant field $\mathbf{E}$ (i.e., $\omega = 0$), we get $\chi_e = \varepsilon_0$, and hence

$$\varepsilon = \varepsilon_0 + \chi_e = 2\varepsilon_0.$$

For visible light, we obtain $\chi_e = 80\,\varepsilon_0$, and hence

$$\varepsilon = \varepsilon_0 + \chi_e = 81\varepsilon_0.$$

In what follows, let us study the phenomenon of dispersion.

Polarization. An electric charge Q at the origin generates the electric potential

$$U(\mathbf{x}) = \frac{Q}{4\pi\varepsilon_0 r}$$

and the electric Coulomb field

$$\mathbf{E}(\mathbf{x}) = -\,\mathbf{grad}\,U(\mathbf{x}) = \frac{Q}{4\pi\varepsilon_0 r^2} \cdot \frac{\mathbf{x}}{\|\mathbf{x}\|}$$

where $r := \|\mathbf{x}\|$. The Coulomb force

$$\boxed{\mathbf{F}(\mathbf{x}) = q\mathbf{E}(\mathbf{x})}$$

acts on an electric charge q at the point $\mathbf{x}$. Now consider the situation pictured in Fig. 12.4. There is a positive electric charge Q at the point $\mathbf{d}$ and a negative electric

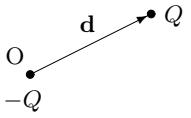


Fig. 12.4. Electric dipole

charge $-Q$ at the origin. If the distance $\|\mathbf{d}\|$ between the two charges is small, then we get the electric potential

$$U(\mathbf{x}) = \frac{Q}{4\pi\varepsilon_0\|\mathbf{x}-\mathbf{d}\|} - \frac{Q}{4\pi\varepsilon_0\|\mathbf{x}\|} = \frac{\mathbf{px}}{4\pi\varepsilon_0 r^3} + O(\|\mathbf{d}\|^2), \quad \mathbf{d}\to 0$$

where $\mathbf{p} := Q\mathbf{d}$ is called the dipole moment of the configuration.

More generally, consider a smooth electric charge density ϱ which vanishes outside some ball $\mathcal{B}$ about the origin. By superposition, the function ϱ generates the electric potential

$$U(\mathbf{x}) = \int_{\mathcal{B}} \frac{\varrho(\mathbf{y})}{4\pi\varepsilon_0\|\mathbf{x}-\mathbf{y}\|}\, d^3y\,.$$

For large distances $r = \|\mathbf{x}\|$,

$$U(\mathbf{x}) = \frac{Q}{4\pi\varepsilon_0 r} + \frac{\mathbf{px}}{4\pi\varepsilon_0 r^3} + \sum_{k,l=1}^{3} \frac{x_k Q_{kl} x_l}{8\pi\varepsilon_0 r^5} + O\left(\frac{1}{r^4}\right), \quad r\to\infty$$

with the dipole moment

$$\mathbf{p} = \int_{\mathcal{B}} \varrho(\mathbf{x})\mathbf{x}\, d^3x$$

and the quadrupole moments

$$Q_{kl} = \int_{\mathcal{B}} \big(3x_k x_l - \mathbf{x}^2\delta_{kl}\big)\varrho(\mathbf{x})\, d^3x, \qquad k,l = 1,2,3.$$

Here, x_1, x_2, x_3 are the Cartesian coordinates of the position vector $\mathbf{x}$.

The Maxwell equation for the electric intensity. We want to motivate the Maxwell equation (12.45) below. To this end, consider the electric potential

$$U(\mathbf{x}) = \int_{\mathbb{R}^3} \frac{\varrho(\mathbf{y})}{4\pi\varepsilon_0\|\mathbf{x}-\mathbf{y}\|}\, d^3y + \int_{\mathbb{R}^3} \frac{\mathbf{P}(\mathbf{y})(\mathbf{x}-\mathbf{y})}{4\pi\varepsilon_0\|\mathbf{x}-\mathbf{y}\|^3}\, d^3y. \tag{12.44}$$

By the superposition principle, this potential corresponds to an electric charge density $\varrho(\mathbf{y})$ and a dipole distribution with the polarization $\mathbf{P}(\mathbf{y})$ (electric dipole moment density) at the point $\mathbf{y}$. Suppose that the functions ϱ and $\mathbf{P}$ are smooth, and that they vanish outside some ball. By the methods of classical potential theory, the smooth function U satisfies the Poisson equation[10]

$$\boxed{\varepsilon_0 \Delta U = \varrho - \operatorname{div}\mathbf{P}.}$$

Noting that $\mathbf{E} = -\,\mathbf{grad}\, U$ and $\Delta U = -\operatorname{div}\mathbf{grad}\, U$,

[10] For example, the proof of this classical result can be found in Courant and Hilbert (1989), Vol. II.

$$\varepsilon_0 \operatorname{div} \mathbf{E} = \varrho - \operatorname{div} \mathbf{P}.$$

Finally, introducing $\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}$, we obtain the desired Maxwell equation

$$\boxed{\operatorname{div} \mathbf{D} = \varrho.} \tag{12.45}$$

Let us now translate this argument from electric fields to magnetic fields.

The Maxwell equation for the magnetic intensity. Traditionally, the potential

$$U(\mathbf{x}) = \frac{\mathbf{mx}}{4\pi\mu_0 r^3}$$

describes the magnetic field intensity $\mathbf{H} = -\operatorname{\mathbf{grad}} U$. Now equation (12.44) passes over to

$$U(\mathbf{x}) = \int_{\mathbb{R}^3} \frac{\varrho_{\text{magnetic}}(\mathbf{y})}{4\pi\varepsilon_0 \|\mathbf{x}-\mathbf{y}\|}\, d^3y + \int_{\mathbb{R}^3} \frac{\mathbf{M}(\mathbf{y})(\mathbf{x}-\mathbf{y})}{4\pi\mu_0 \|\mathbf{x}-\mathbf{y}\|^3}\, d^3y$$

where we use the following notation:

- $\varrho_{\text{magnetic}}(\mathbf{y})$ magnetic charge density at the point $\mathbf{y}$, and
- $\mathbf{M}(\mathbf{y})$ magnetization (magnetic dipole moment density) at the point $\mathbf{y}$.

As above for electric fields,

$$\mu_0 \Delta U = \varrho_{\text{magnetic}} - \operatorname{div} \mathbf{M}.$$

Electric and magnetic charges are called electric and magnetic monopoles, respectively. The experience of physicists show that

There are no magnetic monopoles in classical electrodynamics.

Hence $\varrho_{\text{magnetic}} \equiv 0$. Letting $\mathbf{H} = -\operatorname{\mathbf{grad}} U$, we get

$$\mu_0 \operatorname{div} \mathbf{H} = -\operatorname{div} \mathbf{M}.$$

Introducing the magnetic field $\mathbf{B} = \mu_0 \mathbf{H} + \mathbf{M}$, we obtain the Maxwell equation

$$\boxed{\operatorname{div} \mathbf{B} = 0.} \tag{12.46}$$

Magnetic monopoles. There exists the hypothesis that the high energy of the early universe caused the production of magnetic monopoles. Mathematically, the equations of gauge field theory possess solutions which can be regarded as generalized magnetic monopoles. The existence of such particles was predicted mathematically by Dirac in 1931.[11] We will study this in Volume III on gauge field theory. It turns out that the existence of monopoles is closely related to topology. We refer to the following monographs:

A. Jaffe and C. Taubes (1980). Vortices and Monopoles: The Structure of Static Gauge Fields, Birkhäuser, Boston.

W. Nahm, N. Craigie, and P. Goddard, Monopoles in Quantum Field Theory, World Scientific, Singapore, 1982.

M. Atiyah and M. Hitchin, The Geometry and Dynamics of Magnetic Monopoles, Princeton University Press, 1988.

G. Naber, Topology, Geometry, and Gauge Fields, Springer, New York, 1997.

[11] P. Dirac, Quantized singularities in the electromagnetic field, Proc. Royal Soc. A **133** (1931), 60–72.

B. Felsager, Geometry, Particles, and Fields, Springer, New York, 1998.

Y. Yang, Solitons in Field Theory and Nonlinear Analysis, Springer, New York, 2001.

Ya. Shuir, Magnetic Monopoles, Springer, Berlin, 2005.

C. Klein and O. Richter, Ernst Equation and Riemann Surfaces: Analytical and Numerical Methods, Springer, Berlin, 2005 (application to thin rotating discs of cosmic dust in the universe).

T. Dauxois and M. Peyrard, Physics of Solitons, Cambridge University Press, 2006 (the classical Fermi–Pasta–Ulam problem; solitons in hydrodynamics, solid state physics, atomic physics, magnetic systems, conducting polymers, Bose–Einstein condensates, biological molecules; energy localization and transfer in proteins, DNA).

Dispersion of light. The equation

$$\boxed{m_e\big(\ddot{\mathbf{x}} + 2\gamma\dot{\mathbf{x}} + \omega_0^2\mathbf{x}\big) = -e\mathbf{E}}$$

describes the classical motion $\mathbf{x} = \mathbf{x}(t)$ of an electron of mass m_e and electric charge $-e$ in a molecule under the influence of an oscillating external electric field

$$\mathbf{E}(t) = \mathbf{E}_0 \mathrm{e}^{-\mathrm{i}\omega t} \tag{12.47}$$

which corresponds to an electric wave of angular frequency ω. For the friction constant γ, we assume that $0 < \gamma < \omega_0$. The solution

$$\mathbf{x}(t) = \frac{e}{m_e} \cdot \frac{\mathbf{E}_0 \mathrm{e}^{-\mathrm{i}\omega t}}{\omega^2 + 2\gamma\omega\mathrm{i} - \omega_0^2}$$

is uniquely determined by the property that it vanishes if the external field $\mathbf{E}$ vanishes. The so-called characteristic equation

$$\omega^2 + 2\gamma\omega\mathrm{i} - \omega_0^2 = 0$$

has the zeros $\omega_\pm = -\gamma\mathrm{i} \pm \sqrt{\omega_0^2 - \gamma^2}$ which lie in the open lower half-plane. We assign the dipole moment

$$\mathbf{p}(t) = -e\Re(\mathbf{x}(t))$$

to the oscillating electron. If there are N electrons in the volume V, then we get the polarization $\mathbf{P}(t) = N\mathbf{p}(t)/V$. Using polar coordinates, we have

$$\frac{1}{\omega^2 + 2\mathrm{i}\gamma\omega - \omega_0^2} = -\frac{\mathrm{e}^{-\mathrm{i}\alpha(\omega)}}{|\omega^2 + 2\gamma\omega\mathrm{i} - \omega_0^2|}.$$

Thus, we obtain the key formula for the polarization

$$\mathbf{P}(t) = \frac{Ne^2\,\Re\big(\mathbf{E}_0 \mathrm{e}^{-\mathrm{i}(\omega t + \alpha(\omega))}\big)}{m_e V|\omega^2 + 2\gamma\mathrm{i} - \omega_0^2|}\,.$$

Note that the singularities of the function $\mathbf{P} = \mathbf{P}(t)$ lie in the open lower half-plane. If the friction constant γ is sufficiently small, then we can use the approximation $\gamma = 0$ along with $\alpha(\omega) \equiv 0$. This yields

$$\boxed{\mathbf{P}(t) = \chi_e(\omega)\Re(\mathbf{E}(t))}$$

with the electric susceptibility $\chi_e(\omega) = Ne^2/m_eV|\omega^2-\omega_0^2|$. For the electric intensity, we get

$$\boxed{\mathbf{D}(t) = \varepsilon_0\mathbf{E}(t) + \mathbf{P}(t) = \varepsilon(\omega)\mathbf{E}(t)}$$

with the electric field constant $\varepsilon(\omega) = \varepsilon_0 + \chi_e(\omega)$ of the material medium. Here, in order to simplify notation we replace $\Re(\mathbf{E}(t))$ by $\mathbf{E}(t)$.

Linear response and causality. Let us now describe a more general approach based on linear response theory. We assume that the relation between the electric field $\mathbf{E}$ and the induced polarization $\mathbf{P}$ is given by the equation

$$\mathbf{P}(t) = \int_{-\infty}^{\infty} \chi(t-\tau)\mathbf{E}(\tau)\, d\tau. \tag{12.48}$$

It is important that the linear response equation (12.48) between $\mathbf{E}$ and $\mathbf{P}$ has the following crucial additional property.

It follows from causality that $\chi(t) = 0$ *for all times* $t < 0$.

In fact, this implies

$$\mathbf{P}(t) = \int_{-\infty}^{t} \chi(t-\tau)\mathbf{E}(\tau)\, d\tau,$$

telling us that the polarization $\mathbf{P}(t)$ at time t only depends on the values of the electric field $\mathbf{E}(\tau)$ at the earlier time points $\tau \le t$.

The Fourier–Laplace transform. In order to study the function $\chi = \chi(t)$, we will use the rescaled Fourier–Laplace transform from the time space to the angular frequency space given by

$$\chi(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\chi}(\omega)\mathrm{e}^{-\mathrm{i}\omega t}d\omega \qquad \text{for all} \quad t \in \mathbb{R}$$

and

$$\hat{\chi}(\omega) := \int_{-\infty}^{\infty} \chi(t)\mathrm{e}^{\mathrm{i}\omega t}dt, \qquad \Im(\omega) > 0.$$

Suppose that $\chi \in \mathcal{S}(\mathbb{R})$. The causality condition $\chi(t) = 0$ if $t < 0$ implies that the function

$$\hat{\chi}(\omega) = \int_{0}^{\infty} \chi(t)\mathrm{e}^{\mathrm{i}\omega t}\, dt, \qquad \Im(\omega) > 0$$

is holomorphic on the open upper half-plane, and $\hat{\chi} \in \mathcal{S}(\mathbb{R})$ on the real line. Analogously, we introduce the rescaled Fourier–Laplace transforms

$$\hat{\mathbf{E}}(\omega) := \int_{-\infty}^{\infty} \mathbf{E}(t)\mathrm{e}^{\mathrm{i}\omega t}\, dt, \qquad \hat{\mathbf{P}}(\omega) := \int_{-\infty}^{\infty} \mathbf{P}(t)\mathrm{e}^{\mathrm{i}\omega t}\, dt \qquad \text{for all} \quad \omega \in \mathbb{R}.$$

The linear response equation (12.48) can be written as the convolution

$$\mathbf{P} = \chi * \mathbf{E}.$$

By the convolution rule for the Fourier transform, we get the product relation

$$\boxed{\hat{\mathbf{P}}(\omega) = \hat{\chi}(\omega)\hat{\mathbf{E}}(\omega) \qquad \text{for all} \quad \omega \in \mathbb{R}.} \tag{12.49}$$

This implies that $\hat{\mathbf{D}}(\omega) = \hat{\varepsilon}(\omega)\hat{\mathbf{E}}(\omega)$, where $\hat{\varepsilon}(\omega) := \varepsilon_0 + \hat{\chi}(\omega)$.

The Hilbert transform and the Kramers–Kronig dispersion relations. By Sect. 11.9.3 on page 663, we obtain

$$\hat{\chi}(\omega) = \frac{1}{2\pi \mathrm{i}} \int_{\mathbb{R}} \frac{\hat{\chi}(\xi)}{\xi - \omega}\, d\xi$$

for all complex numbers ω with $\Im(\omega) > 0$. Furthermore, for the boundary values of the function $\hat{\chi}$ on the real line, we obtain

$$\boxed{\Im(\hat{\chi}(\omega)) = \frac{1}{\pi}\, PV \int_{\mathbb{R}} \frac{\Re(\hat{\chi}(\xi))}{\omega - \xi}\, d\xi \qquad \text{for all} \quad \omega \in \mathbb{R}} \tag{12.50}$$

and

$$\boxed{\Re(\hat{\chi}(\omega)) = -\frac{1}{\pi}\, PV \int_{\mathbb{R}} \frac{\Im(\hat{\chi}(\xi))}{\omega - \xi}\, d\xi \qquad \text{for all} \quad \omega \in \mathbb{R}.} \tag{12.51}$$

Recall that Cauchy's principal value of the integral is to be understood in the sense of the following limit:

$$PV \int_{\mathbb{R}} \frac{g(\xi)}{\omega - \xi}\, d\xi := \lim_{\varepsilon \to +0} \left(\int_{-\infty}^{\omega - \varepsilon} \frac{g(\xi)}{\omega - \xi}\, d\xi + \int_{\omega + \varepsilon}^{\infty} \frac{g(\xi)}{\omega - \xi}\, d\xi \right).$$

The importance of the so-called dispersion relations (12.50) and (12.51) for physics was discovered independently by Kramers and Kronig in 1926. Let us mention the following two points.

(i) Energy dissipation: As a rule, the imaginary part $\Im(\hat{\chi})$ is responsible for energy dissipation. Therefore, the Kramers–Kronig dispersion relations (12.50) and (12.51) tell us that causality restricts possible energy dissipation.
(ii) Iterative method: Starting with a reasonable first approximation $\Re(\hat{\chi})_0$, we get $\Im(\hat{\chi})_0$ by (12.50). In turn, this yields $\Re(\hat{\chi})_1$ by (12.51), and so on.

Perspectives. Further information can be found in the following monographs:

- López and Zanette (1999) (dispersion in classical optics),
- Mandel and Wolf (1995) (dispersion in laser optics),
- Landau and Lifshitz (1982), Vol. 3 (dispersion and energy singularities in quantum-mechanical scattering processes),
- Bogoliubov et al. (1958), Bjorken and Drell (1965) (dispersion relations in high-energy physics),
- Chew (1962) (S-matrix theory, analyticity, and strong interaction),
- Bogoliubov et al. (1990) (dispersion relations in axiomatic quantum field theory),
- Todorov (1971) (analytic properties of Feynman diagrams),
- Nolting (2002), Vol. 6 (linear response theory),
- Schwabl (2003) and Landau, Lifshitz (1982), Vol. 9 (general fluctuation-dissipation theorem in statistical physics),
- van der Waerden (1968) (the importance of dispersion theory in the prehistory of Heisenberg's quantum mechanics).

12.4.6 Strong and Weak Interaction

We speak of strong or weak interaction if the coupling constant κ is large or small, respectively. In the case of weak interaction, physicists successfully use the method of perturbation theory. As a rule, this method fails for strong interaction.

In string theory, one studies models which are based on a duality between strong and weak interaction by using, roughly speaking, the replacement

$$\kappa \Rightarrow \frac{1}{\kappa}$$

in the framework of so-called T-duality and mirror symmetry. Physicists hope that, in the future, this will allow us to reduce difficult problems in strong interaction to the method of perturbation theory. We refer to Zwiebach (2004), Chap. 17, and Polchinski (1998), Vol. 2, Chap. 19.

12.5 Microlocal Analysis

Classical wave optics culminates in microlocal analysis.
Folklore

Generalizing the efforts made by physicists and mathematicians in the 19th and 20th century for understanding the propagation of physical effects in terms of the Fourier transform (e.g., in optics), microlocal analysis was founded by the following papers:

J. Kohn and L. Nirenberg, An algebra of pseudo-differential operators, Comm. Pure Appl. Math. **18** (1965), 269–305.

L. Hörmander, On the singularities of the solutions of partial differential equations, Comm. Pure Appl. Math. **23** (1970), 329–358 (wave front set).

V. Maslov, Théorie de perturbations et méthodes asymptotiques, Dunod, Paris, 1972 (translated from the 1965 Russian edition).

L. Hörmander, Fourier integral operators I, Acta Math. **127** (1971), 79–123.

J. Duistermaat and L. Hörmander, Fourier integral operators II, Acta Math. **128**, (1972), 183–269.

As an introduction to microlocal analysis and its applications to partial differential equations, we recommend the comprehensive survey:

Yu. Egorov, A. Komech, and M. Shubin, Elements of the Modern Theory of Partial Differential Equations, Springer, New York, 1999 (pseudo-differential operators and Fourier integral operators).

The relations of microlocal analysis to wave optics are studied in the monograph:

V. Guillemin and S. Sternberg, Geometric Asymptotics, Amer. Math. Soc. Providence, Rhode Island, 1989.

Much material can be found in the treatise:

L. Hörmander, The Analysis of Linear Partial Differential Operators, Vols. 1–4, Springer, New York, 1983.

The theory of Fourier integral operators is essentially based on symplectic differential geometry, which is fundamental for understanding the mathematical structure of geometric optics, classical mechanics, and quantum mechanics. This will be thoroughly studied in Volume II. We also refer to the following monographs:

V. Guillemin and S. Sternberg (1990), Symplectic Techniques in Physics, Cambridge University Press, 1990.

V. Arnold (1978), Mathematical Theory of Classical Mechanics, Springer, Berlin, 1978.

R. Abraham and J. Marsden, Foundations of Mechanics, Addison-Wesley, Reading, Massachusetts, 1978.

In terms of wave optics, singularities appear if light rays intersect each other (focal points) or they form an envelope line or envelope surface (caustic). The mathematical classification of singularities including caustics can be found in the monograph:

> V. Arnold, S. Gusein-Zade, and A. Varchenko, Singularities of Differentiable Maps, Vols. 1, 2, Birkhäuser, Basel, 1985.

Singularities can be handled by using the notion of a wave front set for distributions, which is of general importance for the theory of distributions. Two crucial results concern

- the propagation of singularities of the solutions of linear partial differential equations with smooth coefficients (Theorem 12.9 on page 711) and
- the multiplication of distributions (Theorem 12.11 on page 732).

These theorems play an important role in quantum field theory for

- analyzing the singularities of the Feynman propagator and
- renormalizing the singularities in the framework of the Epstein–Glaser approach.

This will be studied in Volume II in the physical context of quantum electrodynamics.

> *Hörmander's notion of wave front set introduced in 1970 fits the needs of physics in an optimal way. The wave front set describes a spectral analysis of singularities.*

In what follows, let us sketch the basic ideas. The proofs can be found in Hörmander (1983), Vol. 1. We speak of microlocal analysis, since we will study

- the local behavior of a distribution and
- the structure of its Fourier transform at infinity (the asymptotics for short wave lengths, high frequencies, and high energies).

In the sense of the terminology introduced below, microlocal analysis is local analysis in the cotangent bundle. The creation of microlocal analysis was one of the major achievements of the development of analysis in the 20th century. Microlocal analysis is also important for quantum field theory:

- It was discovered by Radzikowski (1996) in his Princeton thesis supervised by Arthur Wightman that the methods of microlocal analysis are very useful for quantum field theory for characterizing so-called Hadamard states which, roughly speaking, represent stable quantum states on curved 4-dimensional space-time manifolds (quantum gravity).
- This idea was further developed by Brunetti, Fredenhagen, and Köhler (1996), and Brunetti, Fredenhagen (2000).

12.5.1 Singular Support of a Distribution

We are given a distribution $G \in \mathcal{D}'(\mathbb{R}^N)$ where $N = 1, 2, \ldots$ Recall that G vanishes locally at the point $x_0 \in \mathbb{R}^N$ iff there exists some open neighborhood U of x_0 such that[12]

$$G(\varphi) = 0 \qquad \text{for all} \quad \varphi \in \mathcal{D}(U).$$

Similarly, the distribution G is called locally smooth at the point $x_0 \in \mathbb{R}^N$ iff there exists some smooth function $g : U \to \mathbb{C}$ on some open neighborhood U of the point x_0 such that

[12] Recall that the space $\mathcal{D}(U)$ consists of all smooth functions $\varphi : U \to \mathbb{C}$ which have compact support.

$$G(\varphi) = \int_{\mathbb{R}^N} g(x)\varphi(x)d^N x \qquad \text{for all} \quad \varphi \in \mathcal{D}(U).$$

Naturally enough, this leads us to the following definitions.

(a) Support $\operatorname{supp}(G)$: The point x_0 lies in the support of the distribution G iff the distribution is not locally zero at x_0.
(b) Singular support $\operatorname{sing\,supp}(G)$: The point x_0 lies in the singular support of the distribution G iff $x_0 \in \operatorname{supp}(G)$ and G is not locally smooth at x_0.

Examples. The following examples show the relation of classical wave propagation to the singular support of distributions.

(a) The Heaviside function $\theta : \mathbb{R} \to \mathbb{R}$ has a jump at the point $x = 0$ (see Fig. 11.1 on page 577). Hence

$$\operatorname{supp}(\theta) = [0, \infty[, \qquad \operatorname{sing\,supp}(\theta) = \{0\}.$$

For the Dirac delta distribution $\delta \in \mathcal{D}'(\mathbb{R}^N)$,

$$\operatorname{supp}(\delta) = \operatorname{sing\,supp}(\delta) = \{0\}.$$

This reflects the fact that the singularity of the Dirac delta function is located at the origin.

(b) Wave propagation on the real line: Define the function

$$f(x,t) := \theta(x - ct) \qquad \text{for all} \quad (x,t) \in \mathbb{R}^2$$

which describes the propagation of a discontinuity with the speed $c > 0$. Explicitly, the discontinuity is at the position $x = ct$ at time t. For the wave function f,

$$\operatorname{sing\,supp}(f) = \{(x,t) \in \mathbb{R}^2 : x - ct = 0\}.$$

This is a straight line in $\mathbb{R}^2$.

(c) Wave propagation in the 3-dimensional Euclidean space: Define

$$g(\mathbf{x},t) := \theta(\mathbf{n}\mathbf{x} - ct) \qquad \text{for all} \quad (\mathbf{x},t) \in \mathbb{R}^4.$$

This describes the propagation of a discontinuity in direction of the given unit vector $\mathbf{n}$ with the speed $c > 0$. The equation

$$\boxed{\mathbf{n}\mathbf{x} - ct = 0, \qquad (\mathbf{x},t) \in \mathbb{R}^4} \tag{12.52}$$

is called the classical wave front equation in $\mathbb{R}^4$. This describes a hyperplane in $\mathbb{R}^4$ which is orthogonal to the 4-dimensional vector $(\mathbf{n}, -c)$. Note that in the 3-dimensional space of our intuition, equation (12.52) describes a plane which moves with the velocity c in direction of its unit normal vector $\mathbf{n}$. If we regard g as a physical quantity (e.g., one component of the electric field vector), then this quantity jumps at the points of the moving plane. For the singular support of the function g, we have

$$\operatorname{sing\,supp}(g) = \{(\mathbf{x},t) \in \mathbb{R}^4 : \mathbf{n}\mathbf{x} - ct = 0\}.$$

12.5.2 Wave Front Set

The wave front set $WF(G)$ of a given distribution $G \in \mathcal{D}'(\mathbb{R}^N)$ represents a refinement of the notion of singular support. The definition below will be chosen in such a way that the following statements hold true.

(i) The wave front set $WF(G)$ is a subset of the product set

$$\{(x,k) : \ x \in \text{sing supp}(G), \ k \in \mathbb{R}^N \setminus \{0\}\}. \tag{12.53}$$

Intuitively, $x \in \mathbb{R}^N$ and $k \in \mathbb{R}^N$ can be regarded as N-dimensional position point and wave vector, respectively. Roughly speaking, large values of $||k||$ correspond to small wave lengths (high energies).

(ii) If $(x,k) \in WF(G)$, then $(x,\gamma k) \in WF(G)$ for all $\gamma > 0$ (homogeneity).

(iii) If $x \in \text{supp sing}(G)$, then there exists a nonzero wave vector $k \in \mathbb{R}^N$ such that $(x,k) \in WF(G)$.

The relation to the cotangent bundle. In modern mathematics, the set

$$T^*\mathbb{R}^N := \{(x,k) : x \in \mathbb{R}^N, k \in \mathbb{R}^N\}$$

is called the cotangent bundle of the position space $\{x : x \in \mathbb{R}^N\}$. We also introduce the dual pairing[13]

$$\langle k|x\rangle := k_j x^j, \qquad (x,k) \in T^*\mathbb{R}^N,$$

and the differential form

$$\Omega(k) := -k_j dx^j, \qquad k \in \mathbb{R}^N$$

with its derivative

$$d\Omega = dx^j \wedge dk_j.$$

The differential form $d\Omega$ equips the cotangent bundle $T^*\mathbb{R}^N$ with a natural symplectic structure. By a change of coordinates of the cotangent bundle, we mean the matrix transformation

$$x' = Ax, \quad k' = (A^{-1})^d k$$

where A is a real invertible $(N \times N)$-matrix, and $(A^{-1})^d$ denotes the dual (or transposed) matrix to the inverse matrix A^{-1}. The point is that $\langle k|x\rangle$ and Ω, $-d\Omega$ are invariant under coordinate changes of the cotangent bundle. This means that

$$\langle k'|x'\rangle = \langle k|x\rangle, \quad \Omega'(k') = \Omega(k), \quad d\Omega' = d\Omega$$

for all $x, k \in \mathbb{R}^N$. This invariance property follows immediately from a general mathematical principle in tensor calculus called the principle of the right index picture (see page 769). Alternatively, an explicit computation shows that

$$\langle k'|x'\rangle = (k')^d x' = k^d A^{-1} A x = k^d x = \langle k|x\rangle.$$

Similarly, we get $\Omega'(k') = \Omega(k)$. Finally, the Cartan derivative $d\Omega$ is invariant under coordinate transformations. Hence $d\Omega' = d\Omega$.

For fixed position x, the set $F_x := \{(x,k) : k \in \mathbb{R}^N\}$ is called the fiber of the cotangent bundle $T^*\mathbb{R}^N$ at the base point x. Intuitively, this fiber describes all possible wave vectors k at the position x.

[13] In what follows, we sum over equal upper and lower indices from 1 to N.

Summarizing, the wave front set of a distribution is a subset of the cotangent bundle of the position space.

In terms of fibers, we can write

$$\boxed{WF(G) \subseteq \cup_{x \in \text{supp sing}(G)} \ F_x \setminus \{0\}.}$$

The notion of cotangent bundle is important if one wants to generalize the notion of wave front set to functions and distributions that are defined on manifolds. Fiber bundles play a fundamental role in modern mathematics and physics. This will be thoroughly studied in Volume III on gauge field theory.

Prototypes. For the convenience of the reader, before giving the precise definition of the wave front set, let us present some examples.

(i) Smooth functions: The wave front set of a smooth function $g : \mathbb{R}^N \to \mathbb{C}$ is empty.

(ii) Heaviside function: The wave front set of the function $\theta = \theta(x)$ reads as

$$WF(\theta) = \{(x,k) \in \mathbb{R}^2 : x = 0, k \in \mathbb{R} \setminus \{0\}\}.$$

(iii) Dirac's delta distribution δ in $\mathcal{D}'(\mathbb{R}^N)$:

$$WF(\delta) = \{(x,k) \in \mathbb{R}^N \times \mathbb{R}^N : x = 0, k \in \mathbb{R}^N \setminus \{0\}\}.$$

(iv) Propagation of a discontinuity: We are given the unit vector $\mathbf{n}$ and the velocity $c > 0$. The function

$$\boxed{g(\mathbf{x},t) := \theta(\mathbf{nx} - ct), \qquad \mathbf{x} \in \mathbb{R}^3, \ t \in \mathbb{R}}$$

describes the propagation of a discontinuity in direction of the vector $\mathbf{n}$ with the velocity c. Set $x := (x^0, \mathbf{x})$ and $k := (k_0, \mathbf{k})$ with $x^0 := ct$. The hyperplane

$$\mathcal{W} := \{x \in \mathbb{R}^4 : \mathbf{nx} - ct = 0\}$$

in the 4-dimensional space-time $\mathbb{R}^4$ is called the classical wave front. The jumps of the function g correspond to the points of $\mathcal{W}$. The set

$$\mathcal{NW} := \{(x,k) \in \mathbb{R}^4 \times \mathbb{R}^4 : \ x \in \mathcal{W}, \ k \perp \mathcal{W}, \ k \neq 0\}$$

is called the normal bundle of the hyperplane $\mathcal{W}$. Geometrically, $\mathcal{NW}$ consists of all the pairs (x,k) where x lies on the hyperplane $\mathcal{W}$, and k is a nonzero normal vector of $\mathcal{W}$ at the point x. This means that

$$\langle k|x\rangle = k_0 x^0 + \mathbf{kx} = 0.$$

Explicitly, $k = \gamma(-1, \mathbf{n})$ for all nonzero real numbers γ. We now have

$$\boxed{\text{sing supp}(g) = \mathcal{W}, \qquad WF(g) = \mathcal{NW}.}$$

Thus, the singular support of the function g is equal to the hyperplane $\mathcal{W}$ (classical wave front plane), and the wave front set of g coincides with the normal bundle of $\mathcal{W}$.

(v) Let $\mathcal{P}$ be a plane through the origin in the 3-dimensional Euclidean space with the normal unit vector $\mathbf{n}$. Define the distribution

$$G(\varphi) := \int_{\mathcal{P}} g(x)\varphi(x)dS$$

for all test functions $\varphi \in \mathcal{D}(\mathbb{R}^3)$ where the function $g : \mathcal{P} \to \mathbb{C}$ is smooth, and dS refers to the surface area on the plane $\mathcal{P}$. The wave front set of G consists of all pairs

$$(\mathbf{x}, \gamma\mathbf{n})$$

where $\mathbf{x}$ lies in the support of the function g, and γ is an arbitrary nonzero real number.

(vi) Oscillating integral (superposition of plane waves): Define

$$g(x) := \int_{\mathbb{R}^N} \mathrm{e}^{\mathrm{i}\langle k|x\rangle}\, a(k)d^N k, \qquad x \in \mathbb{R}.$$

We assume that the amplitude function $a : \mathbb{R}^N \to \mathbb{C}$ is measurable (e.g., continuous) and $\int_{\mathbb{R}^N} |a(k)|\, d^N k < \infty$. Then

$$WF(g) = \{(0, k) : x \in \mathbb{R}^N,\ k \in \mathbb{R}^N \setminus \{0\}\}.$$

The function $\varphi(x, y) := \langle x|y\rangle$ is called phase function. In particular $\varphi_x(x, y) = y$ and $\varphi_y(x, y) = x$. The following example generalizes this situation.

(vii) Oscillating integral (caustic): Define

$$f(x) := \int_{\mathbb{R}^N} \mathrm{e}^{\mathrm{i}\varphi(x,y)}\, a(y)\, d^N y.$$

We make the following assumptions.

- The phase function $\varphi : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ is smooth. For all $x, y \in \mathbb{R}^N$ and $\gamma > 0$, we have $\varphi(x, \gamma y) = \gamma\varphi(x, y)$.
- $(\varphi_x(x, y), \varphi_y(x, y)) \neq 0$ for all $x, y \in \mathbb{R}^N$ with $y \neq 0$.
- The amplitude function $a : \mathbb{R}^N \to \mathbb{C}$ is measurable (e.g., continuous) and $\int_{\mathbb{R}^N} |a(y)|\, d^N y < \infty$.

For fixed $y \in \mathbb{R}^N$, the surface

$$\varphi(x, y) = 0, \qquad x \in \mathbb{R}^N \tag{12.54}$$

is called the phase surface. The critical points x of this surface are given by

$$\varphi_y(x, y) = 0. \tag{12.55}$$

The vector k is called a critical wave vector iff

$$k = \varphi_x(x, y), \qquad \varphi_y(x, y) = 0. \tag{12.56}$$

Then the wave front set $WF(f)$ of the function f is given by the set of all pairs

$$(x, k) \quad \text{with} \quad x, k \in \mathbb{R}^N, k \neq 0$$

where the wave vector k is critical, that is, condition (12.56) is satisfied for some $y \in \mathbb{R}^N$. Intuitively, this is related to the phenomenon of caustic surfaces. In fact, equation (12.54) describes a family of phase surfaces. If it is possible to solve the equation (12.55) by $y = y(x)$, then the equation $\varphi(x, y(x)) = 0$ represents the envelope surface of the surface family (12.54). The points of this envelope surface form the caustic surfaces in geometric optics.

(viii) If $G \in \mathcal{D}'(\mathbb{R})$, then

$$WF(G') = WF(G).$$

For example, $\theta' = \delta$ implies $WF(\delta) = WF(\theta)$.

Note that if $G \in \mathcal{D}'(\mathbb{R}^N)$ with $N > 1$ in (viii), then the situation changes essentially. A general result will be formulated in Theorem 12.9 on page 711.

Galaxies as gravitational lenses. Geometric optics was founded by Fermat (1601–1665). Recently, geometric optics has also played a crucial role in quasar astronomy. In fact, if light rays coming from a distant quasar pass through a galaxy, then the gravitational force of the galaxy acts like a lens. For example, astronomers observe several quasars instead of the original single quasar. This is the famous Einstein cross. Interestingly enough, caustic effects are also observed. We refer to the monograph:

J. Ehlers, E. Falco, and P. Schneider, Gravitational Lenses, Springer, New York, 1992.

Motivation of the wave front set. As a motivation for the general definition below, let us consider the Heaviside function θ. By Fourier transform,

$$\chi(x)\theta(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathcal{F}(\chi\theta)(k)\, \mathrm{e}^{\mathrm{i}kx} dk, \qquad x \in \mathbb{R} \tag{12.57}$$

and

$$\mathcal{F}(\chi\theta) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \chi(x)\theta(x)\, \mathrm{e}^{-\mathrm{i}kx} dx, \qquad k \in \mathbb{R}. \tag{12.58}$$

Choose the fixed point x_0 on the real line.

- Regular case: $x_0 \neq 0$. In order to localize the behavior of the function θ at the point x_0, we choose a sufficiently small neighborhood U of x_0 and a function $\chi \in \mathcal{D}(U)$ such that $\chi\theta \in \mathcal{D}(U)$. By the Paley–Wiener–Schwartz theorem on page 661,

$$|\mathcal{F}(\chi\theta)(k)| = O\left(\frac{1}{|k|^r}\right), \qquad |k| \to \infty, \quad r = 1, 2, \ldots. \tag{12.59}$$

That is, the Fourier transform $\mathcal{F}(\chi\theta)$ decreases rapidly as the wave number $|k|$ goes to infinity. In particular, there exists a sufficiently large number k_∞ such that the function

$$\theta_{k_\infty}(x) := \frac{1}{\sqrt{2\pi}} \int_{-k_\infty}^{k_\infty} \mathcal{F}(\chi\theta)(k) \mathrm{e}^{\mathrm{i}kx} dk \qquad \text{for all} \quad x \in U \tag{12.60}$$

represents a reasonable approximation of the localization $\chi\theta$ of the function θ near the point x_0. Since the wave number k is related to the wave length λ by the equation $k = 2\pi/\lambda$, formula (12.60) also represents a cutoff for the wave length near zero. That is, sufficiently small wavelengths can be neglected.
- Singular case: $x_0 = 0$. Now choose an open neighborhood $U =]-\varepsilon, \varepsilon[$ of the point x_0 and a nonnegative function $\chi \in \mathcal{D}(U)$. Then

$$\mathcal{F}(\chi\theta)(k) = O\left(\frac{1}{|k|}\right), \qquad |k| \to \infty.$$

This means that, as $|k| \to \infty$, the Fourier transform does not decrease as rapidly as in (12.59). To prove this, observe that, by the mean theorem, there exists a number $\eta \in U$ such that

$$\mathcal{F}(\chi\theta)(k) = \frac{\chi(\eta)}{\sqrt{2\pi}} \int_0^{\varepsilon} \mathrm{e}^{\mathrm{i}kx}\, dx = \frac{\chi(\eta)(\mathrm{e}^{\mathrm{i}\varepsilon} - 1)}{\mathrm{i}\sqrt{2\pi}} \cdot \frac{1}{k}.$$

Definition of the wave front set. By an open cone C in $\mathbb{R}^N$, we understand an open subset of $\mathbb{R}^N \setminus \{0\}$ which has the additional property that

$$k \in C \text{ implies } \gamma k \in C \text{ for all } \gamma > 0.$$

Motivated by the preceding example, let us base the definition of the wave front set $WF(G)$ of a distribution $G \in \mathcal{D}'(\mathbb{R}^N)$ on the violation of the following key condition:

$$\boxed{|\mathcal{F}(\chi G)(k)| \leq \frac{\text{const}(r)}{(1 + ||k||)^r}, \qquad r = 1, 2, \ldots} \tag{12.61}$$

The pair (x_0, k_0) with $x_0, k_0 \in \mathbb{R}^N$ and $k_0 \neq 0$ is called a regular microlocal pair of the distribution G iff

- there exists a test function $\chi \in \mathcal{D}(\mathbb{R}^N)$ with $\chi(x_0) \neq 0$ and
- an open cone C in $\mathbb{R}^N$ with $k_0 \in C$

such that the condition (12.61) is valid for all $k \in C$ and all indices r. This means that the localization χG of the distribution G near the point x_0 has a Fourier transform which decays rapidly with respect to wave numbers k that lie in the open cone C near k_0.

The wave front set $WF(G)$ consists of all pairs (x_0, k_0) with $x_0, k_0 \in \mathbb{R}^N$ and $k_0 \neq 0$ which are not regular microlocal pairs of the distribution G.

Propagation of singularities. Choose $N = 1, 2, , \ldots$ and $m = 1, 2, \ldots$. Let

$$\boxed{Lu(x) := \sum_{|\alpha| \leq m} a_\alpha(x) \partial^\alpha u(x)}$$

be a linear differential operator of order $m = 1, 2, \ldots$ with smooth coefficients $a_\alpha : \mathbb{R}^N \to \mathbb{C}$ for all multi-indices α. The replacement

$$\frac{\partial}{\partial x^j} \Rightarrow \mathrm{i}k_j, \qquad j = 1, \ldots, N$$

sends the differential operator L to its symbol

$$P(x, k) := \sum_{|\alpha| \leq m} a_\alpha \mathrm{i}^{|\alpha|} k^\alpha, \qquad x, k \in \mathbb{R}^N.$$

The principal symbol of L is defined by

$$P_m(x, k) := \sum_{|\alpha| = m} a_\alpha \mathrm{i}^{|\alpha|} k^\alpha, \qquad x, k \in \mathbb{R}^N.$$

If the coefficients a_α are constants, then $P(x, k)$ is independent of x, and Fourier transform tells us that

$$\mathcal{F}(Lu)(k) = P(k)\mathcal{F}(u)(k) \qquad \text{for all} \quad k \in \mathbb{R}^N$$

and all functions $u \in \mathcal{S}(\mathbb{R}^N)$. The set

$$\mathrm{Char}(L) := \{(x,k) \in \mathbb{R}^N \times \mathbb{R}^N : P_m(x,k) = 0, k \neq 0\}$$

is called the characteristic set of the differential operator L. In particular, the differential operator L is called elliptic iff $\mathrm{Char}(L)$ is empty. The following theorem is one of the most important theorems in the theory of linear partial differential equations.

Theorem 12.9 *We are given the distribution $f \in \mathcal{D}'(\mathbb{R}^N)$. Suppose that the distribution $u \in \mathcal{D}'(\mathbb{R}^N)$ is a solution of the linear differential equation*

$$Lu = f.$$

Then

$$\boxed{WF(u) \subseteq WF(f) \cup \mathrm{Char}(L).}$$

If L is elliptic, then $WF(u) = WF(f)$ and $\mathrm{sing\ supp}(u) = \mathrm{sing\ supp}(f)$.

If we choose $f = \delta$, then the solution G of $LG = \delta$ is a fundamental solution of L. In this special case,

$$WF(G) \subseteq WF(\delta) \cup \mathrm{Char}(L).$$

Here, $WF(\delta) = \{(0,k) : k \in \mathbb{R}^N \setminus \{0\}\}$. Observe that Green's functions are always fundamental solutions. As two typical examples let us consider the elliptic Poisson equation and the hyperbolic Klein–Gordon equation which contains the wave equation as a special case.

The elliptic Poisson equation. This equation reads as

$$\boxed{\varepsilon_0 \Delta U = \varrho \qquad \text{on} \quad \mathbb{R}^3.} \tag{12.62}$$

Here, we use a Cartesian (x,y,z)-system, and we introduce the Laplacian

$$\Delta := -\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2}.$$

The distribution $\varrho = \varrho(x,y,z)$ describes the electric charge density, and the distribution $U = U(x,y,z)$ represents the electrostatic potential. An electric point charge of strength Q at the origin corresponds to

$$\varrho = Q\delta$$

where δ is the Dirac delta distribution. The symbol of the Laplacian is given by

$$P(\mathbf{k}) = \mathbf{k}^2 \qquad \text{for all} \quad \mathbf{k} \in \mathbb{R}^3.$$

Consequently, the Laplacian is elliptic. For given distribution $\varrho \in \mathcal{D}'(\mathbb{R}^3)$, let the distribution $U \in \mathcal{D}'(\mathbb{R}^3)$ be a solution of the Poisson equation (12.62). Then, for the wave front sets,

$$\boxed{WF(U) = WF(\varrho).}$$

For the singular support,

$$\mathrm{sing\ supp}(U) = \mathrm{sing\ supp}(\varrho).$$

In particular, if the electric charge density $\varrho : \mathbb{R}^3 \to \mathbb{R}$ is smooth, then so is the electrostatic potential U, since the singular supports of ϱ and U are empty. In the

special case of a point charge at the origin, $\varrho = Q\delta$, the singular support of ϱ is equal to the origin , and so is the singular support of the corresponding electrostatic potential U. Explicitly,

$$U(\mathbf{x}) = \frac{Q}{4\pi\varepsilon_0||\mathbf{x}||} + V(\mathbf{x}) \qquad \text{for all} \quad \mathbf{x} \in \mathbb{R}^3, \mathbf{x} \neq 0.$$

Here, the function V is an arbitrary smooth solution of the homogeneous equation $\Delta V = 0$ on $\mathbb{R}^3$.

The hyperbolic Klein–Gordon equation. For fixed $m_0 > 0$, this equation reads as

$$\boxed{\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} + \Delta\psi + \frac{m_0^2c^2}{\hbar^2}\,\psi = \chi \qquad \text{on} \quad \mathbb{R}^4.} \tag{12.63}$$

Introducing the wave operator $\Box := \frac{1}{c^2}\,\frac{\partial^2}{\partial t^2} + \Delta$, the Klein–Gordon can be written as

$$\left(\Box + \frac{m_0^2c^2}{\hbar^2}\right)\psi = \chi.$$

Here, we use the following notation: $\psi = \psi(x, y, z, t)$ is the wave function of the Yukawa meson of rest mass $m_0 > 0$, c is the velocity of light in vacuum, h is the Planck constant, and $\hbar := h/2\pi$. In the special case where $m_0 = 0$, the Klein–Gordon equation passes over to the wave equation

$$\boxed{\frac{1}{c^2}\frac{\partial^2 U}{\partial t^2} + \Delta U = \frac{\varrho}{\varepsilon_0} \quad \text{on} \quad \mathbb{R}^4} \tag{12.64}$$

where the distributions $\varrho = \varrho(x, y, z, t)$ and $U = U(x, y, z, t)$ are the time-dependent electric charge density and electric potential, respectively.

Set $x := (x^0, \mathbf{x})$ and $k := (k_0, \mathbf{k})$ where $x^0 := ct$. Using the replacement

$$\frac{\partial}{\partial x^j} \Rightarrow \mathrm{i}k_j, \quad j = 0, 1, 2, 3,$$

we assign the symbol

$$P(x, k) := -k_0^2 + \mathbf{k}^2 + \frac{m_0^2c^2}{\hbar^2}$$

to the Klein–Gordon equation. By definition, the critical set $\mathrm{Char}(\Box + \frac{m_0^2c^2}{\hbar^2})$ of the Klein–Gordon operator is equal to the set of all the pairs

$$(x, k) \in \mathbb{R}^4 \times \mathbb{R}^4 \quad \text{with} \quad P(x, k) = -k_0^2 + \mathbf{k}^2 = 0, \;\; k \neq 0.$$

The solutions $x = x(\tau), k = k(\tau)$ of the system of ordinary differential equations

$$\dot{x} = P_k(x, k), \qquad \dot{k} = P_x(x, k)$$

are called the bicharacteristics of the Klein–Gordon equation. Explicitly,

$$\dot{x^0} = -2k_0, \quad \dot{x^j} = 2k_j, \quad \dot{k}_l = 0, \quad j = 1, 2, 3, \;\; l = 0, 1, 2, 3.$$

Therefore, the bicharacteristics are straight lines of the form

$$k_0, k_j = \mathrm{const}, \quad x^0(\tau) = -2k_0\tau + x^0(0), \quad x^j(\tau) = 2k_j(\tau) + x^j(0)$$

for all $\tau \in \mathbb{R}$ and $j = 1, 2, 3$. Using the energy E and the momentum vector $\mathbf{p} := p^1\mathbf{e}_1 + p^2\mathbf{e}_2 + p^3\mathbf{e}_3$ with respect to the right-handed orthonormal system $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$, physicists introduce the 4-momentum

$$p = \hbar k$$

along with $p_0 = p^0 := E/c$ and $p_j = -p^j$ if $j = 1, 2, 3$. Then, the point (x, k) lies in the characteristic set of the Klein–Gordon equation iff we satisfy the Einstein energy-momentum equation

$$\boxed{E^2 = c^2 m_0^2 + c^2 \mathbf{p}^2.}$$

This equation describes the so-called mass shell in the energy-momentum space corresponding to the mass parameter m_0. For $m_0 = 0$, the mass shell $E^2 = c^2\mathbf{p}^2$ is the counterpart to the light cone

$$c^2 t^2 = \mathbf{x}^2, \qquad \mathbf{x} \in \mathbb{R}^3,\ t \in \mathbb{R}.$$

For a given distribution $\chi \in \mathcal{D}'(\mathbb{R}^4)$, let $\psi \in \mathcal{D}'(\mathbb{R}^4)$ be a solution of the Klein–Gordon equation. Then, for the wave front set, we have the following two properties:

(i) $WF(\psi) \subseteq WF(\chi) \cup \operatorname{Char}\left(\Box + \frac{m_0^2 c^2}{\hbar^2}\right)$.

(ii) Let $f = 0$. If the point (x, k) lies in the wave front set $WF(u)$, then the same is true for all points of the bicharacteristics through the point (x, k).

In terms of physics, the appearance of the critical set $\operatorname{Char}(\Box + \frac{m_0^2 c^2}{\hbar^2})$ corresponds to the propagation of discontinuous wave fronts. This phenomenon does not appear for the Poisson equation, since this equation describes stationary processes.

Retarded and advanced fundamental solutions of the Klein–Gordon equation. Let $m_0 \geq 0$. We have

$$\left(\Box + \frac{m_0^2 c^2}{\hbar^2}\right) \mathcal{G}_\pm = \delta \qquad \text{on } \mathbb{R}^4$$

with the so-called retarded fundamental solution

$$\mathcal{G}_+(\mathbf{x}, t) := \frac{\theta(t)}{2\pi c}\, \delta(c^2 t^2 - r^2) - \frac{m_0 \theta(ct - r)}{2\pi\hbar} \cdot \frac{J_1\left(\frac{m_0 c^2}{\hbar}\sqrt{t^2 - \frac{r^2}{c^2}}\right)}{\sqrt{c^2 t^2 - r^2}}$$

and the so-called advanced fundamental solution

$$\mathcal{G}_-(\mathbf{x}, t) := \frac{\theta(-t)}{2\pi c}\, \delta(c^2 t^2 - r^2) - \frac{m_0 \theta(-ct - r)}{2\pi\hbar} \cdot \frac{J_1\left(\frac{m_0 c^2}{\hbar}\sqrt{t^2 - \frac{r^2}{c^2}}\right)}{\sqrt{c^2 t^2 - r^2}}.$$

Here, we use the radius $r := ||\mathbf{x}||$ and the Bessel function[14]

$$J_1(z) := \sum_{r=0}^{\infty} (-1)^r \frac{z^{2r+1}}{2^{2r+1} r!(r+1)!} \qquad \text{for all} \quad z \in \mathbb{C}.$$

The sophisticated proof can be found in Komech (1999), Chap. 4. The proof uses the classical method of Marcel Riesz (1886–1969) on the regularization of singular

[14] See G. Watson, A Treatise on the Theory of Bessel Functions, Cambridge University Press, 1944.

integrals by introducing a regularizing parameter and using analytic continuation with respect to this parameter.[15]

Remark on the notation used by physicists. The formal expression

$$f_+(\mathbf{x},t) := \theta(t)\delta(c^2t^2 - r^2)$$

stands for the rigorous distribution $F_+ \in \mathcal{D}'(\mathbb{R}^4)$ given by

$$\boxed{F_+(\varphi) := \frac{c^2}{2}\int_0^\infty dt \cdot t \int_{\mathbb{S}^2} dS_P\ \varphi(r,P,t)_{|r=ct}} \tag{12.65}$$

for all test functions $\varphi \in \mathcal{D}(\mathbb{R}^4)$. Similarly, for $f_-(\mathbf{x},t) := \theta(t)\delta(c^2t^2 - r^2)$ we get

$$\boxed{F_-(\varphi) := \frac{c^2}{2}\int_{-\infty}^0 dt \cdot t \int_{\mathbb{S}^2} dS_P\ \varphi(r,P,t)_{|r=-ct}} \tag{12.66}$$

for all $\varphi \in \mathcal{D}(\mathbb{R}^4)$. Define the forward light cone

$$C_+ := \{(\mathbf{x},t) \in \mathbb{R}^4 : t \geq 0,\ ||x||^2 = c^2t^2\}$$

and the backward light cone $C_- := \{(\mathbf{x},t) \in \mathbb{R}^4 : t \leq 0,\ ||x||^2 = c^2t^2\}$. Observe that the retarded (resp. advanced) distribution F_+ (resp. F_-) only depends on the values of the test function φ on the forward (resp. backward) light cone.

Mnemonically, this is obtained as follows. For fixed time t, the light-cone equation $c^2t^2 - r^2 = 0$ has the zeros $r = \pm ct$. By Sect. 11.2.4 on page 595,

$$\delta(c^2t^2 - r^2) = \frac{\delta(r-ct)}{2r} + \frac{\delta(r+ct)}{2r}. \tag{12.67}$$

Using spherical coordinates, the test function $\varphi = \varphi(r,P,t)$ depends on the radius r, the point P on the unit sphere $\mathbb{S}^2$, and time t. Now formal integration tells us that

$$\begin{aligned} F_+(\varphi) &= \int_{\mathbb{R}^4} f_+(\mathbf{x},t)\varphi(\mathbf{x},t)\ dxdydzd(ct) \\ &= \int_{-\infty}^\infty cdt \int_0^\infty r^2dr \int_{\mathbb{S}^2} dS_P\ \theta(t)\delta(c^2t^2 - r^2)\varphi(r,P,t). \end{aligned}$$

Applying the decomposition formula (12.67) and integrating over the radius r, we get the desired representation formula (12.65).

12.5.3 The Method of Stationary Phase

> The method of stationary phase has its roots in wave optics. Nowadays it is a magic tool in the hands of Edward Witten for studying quantum-field models in terms of the Feynman functional integral and for producing deep topological invariants. Here, the Feynman functional integral plays the role of a generating functional.
>
> Folklore

[15] M. Riesz, L'intégrale de Riemann–Liouville et le problème de Cauchy (The Riemann-Liouville integral and the initial-value problem for wave equations), Acta Math. **81** (1949), 1–223. See also Gelfand et al. (1964), Vol. 1, Chap. 3 (special types of distributions).

Rapidly oscillating integrals. In physics, one frequently encounters integrals which posses an integrand that oscillates with high frequency. The integral represents the mean value over a high-frequency process in nature or technology, and the task is to compute such mean values. Alternatively, one has to compute integrals which represent the mean value over short-wave processes. In quantum mechanics, one encounters integrals of the type

$$\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}S(x)/\hbar}\, a(x)dx$$

In the semiclassical case, the Planck constant h and hence $\hbar := h/2\pi$ can be assumed to be small. The Feynman functional integral

$$\int_{\psi\in X} \mathrm{e}^{\mathrm{i}S(\psi)/\hbar}\, \mathcal{D}(\psi)$$

is also of this type. Here X denotes the space of all quantum fields ψ, and the symbol $\mathcal{D}$ stands for a formal measure on the space X.

Two classical standard examples and Dirac's delta function. Fix $\varepsilon > 0$. In what follows, we take

$$\omega = \frac{1}{\varepsilon}, \qquad \lambda = \varepsilon, \qquad \hbar = \varepsilon.$$

Examples are given by

$$\mathrm{e}^{\mathrm{i}\omega t} = \mathrm{e}^{\mathrm{i}t/\varepsilon}, \qquad \mathrm{e}^{2\pi \mathrm{i}x/\lambda} \qquad \mathrm{e}^{\mathrm{i}Et/\hbar}, \qquad \mathrm{e}^{\mathrm{i}px/\hbar}.$$

Here, we use the following notation: t time, x space variable, E energy, and p momentum. Small values of ε correspond to situations where the angular frequency ω is high, the wave length λ is small, or the Planck's constant of action h is small. Two typical examples are the Dirichlet integral

$$\int_{-\infty}^{\infty} \frac{\sin\left(\frac{x}{\varepsilon}\right)}{x}\, dx = \pi, \qquad \varepsilon > 0,$$

and the Fresnel integrals

$$\int_{-\infty}^{\infty} \sin\left(\frac{x^2}{2\varepsilon}\right) dx = \int_{-\infty}^{\infty} \cos\left(\frac{x^2}{2\varepsilon}\right) dx = \sqrt{\pi\varepsilon}, \qquad \varepsilon > 0.$$

It is not obvious at first that these integrals are convergent. In fact, each of the integrands oscillates, and it turns out that many of the positive and negative contributions to the integral cancel each other. For the computation of these integrals via Cauchy's famous residue method, see Problems 12.2 and 12.3.

The Dirichlet integral played a key role in the classical convergence theory for both Fourier series and Fourier integrals due to Dirichlet (1805–1859).[16] Dirichlet used his integral for proving that

$$\lim_{\varepsilon\to+0} \int_0^a \frac{\sin\left(\frac{x}{\varepsilon}\right)}{x} f(x)dx = \frac{\pi}{2} \lim_{x\to+0} f(x) \tag{12.68}$$

if the function $f : [0, a] \to \mathbb{R}$ is monotone increasing and bounded on the compact interval $[0, a]$ with $a > 0$. This is called the Dirichlet lemma. Let $\varphi \in \mathcal{D}(\mathbb{R})$. This

[16] See Fikhtengol'ts (1965), Vol. 3, Chap. XIX.

function is smooth and hence of bounded variation. Consequently, the function φ is the difference of two continuous monotone increasing functions. Introducing the Dirichlet function

$$\delta_\varepsilon(x) := \frac{\sin\left(\frac{x}{\varepsilon}\right)}{\pi x}, \qquad x \in \mathbb{R} \setminus \{0\},$$

and $\delta_\varepsilon(0) := \frac{1}{\pi\varepsilon}$, it follows from (12.68) that

$$\boxed{\lim_{\varepsilon \to +0} \int_{-\infty}^{\infty} \delta_\varepsilon(x)\varphi(x)dx = \varphi(0)}$$

for all test functions $\varphi \in \mathcal{D}(\mathbb{R})$. Therefore,

$$\lim_{\varepsilon \to +0} \delta_\varepsilon(x) = \delta(x),$$

in the sense of distributions on $\mathcal{D}'(\mathbb{R})$. Consequently, the Dirichlet function family $\{\delta_\varepsilon\}_{\varepsilon>0}$ represents a forerunner of Dirac's delta function introduced by Dirac around 1930. For small parameter $\varepsilon > 0$, the function δ_ε is only concentrated near the origin. The method of stationary phase below will allow us to construct many such functions which possess the crucial property of concentration. Such function always appear for physical processes which are focused. A typical example of focusing is the patterns obtained by diffraction of light at a small slit.

The Fresnel integrals were used by Fresnel (1788–1827) in order to compute sophisticated second-order diffraction effects for light (see Sect. 12.5.5). Introducing the principal value integral

$$PV \int_{-\infty}^{\infty} f(x)dx := \lim_{\varepsilon \to +0} \left(\int_{-\infty}^{-\varepsilon} f(x)dx + \int_{\varepsilon}^{\infty} f(x)dx \right),$$

we get

$$PV \int_{-\infty}^{\infty} \frac{\cos\left(\frac{x}{\varepsilon}\right)}{x}\, dx = 0, \qquad \varepsilon > 0,$$

since the cosine function is even, and hence the integrand is odd. Using Euler's formula $\mathrm{e}^{\mathrm{i}\alpha} = \cos\alpha + \mathrm{i}\sin\alpha$, we obtain

$$PV \int_{-\infty}^{\infty} \frac{\mathrm{e}^{\mathrm{i}x/\varepsilon}dx}{x} = \mathrm{i}\pi,$$

and

$$\frac{1}{\sqrt{2\pi\varepsilon}} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}x^2/2\varepsilon}dx = \frac{1+\mathrm{i}}{\sqrt{2}}, \qquad \varepsilon > 0.$$

Kelvin's idea. Let $a \in \mathcal{D}(\mathbb{R})$ be a test function. That is, $a : \mathbb{R} \to \mathbb{C}$ is smooth and has compact support. The key formula of the method of stationary phase looks like

$$\boxed{\frac{1}{\sqrt{2\pi\varepsilon}} \int_{-\infty}^{\infty} a(x)\mathrm{e}^{\mathrm{i}x^2/2\varepsilon}\, dx = \frac{1+\mathrm{i}}{\sqrt{2}}\, a(0) + O(\varepsilon),} \tag{12.69}$$

as $\varepsilon \to +0$. This means that the main contributions to the integral come from the value of the test function a at the origin. This is a consequence of the fact that the Fresnel functions

$$\delta_\varepsilon(x) := \frac{\mathrm{e}^{\mathrm{i}x^2/2\varepsilon}}{(1+\mathrm{i})\sqrt{\pi\varepsilon}}, \qquad x \in \mathbb{R}$$

are concentrated near the origin if the parameter ε is small. Therefore, like the Dirichlet function family, the Fresnel function family $\{\delta_\varepsilon\}_{\varepsilon>0}$ also represents an approximation of the Dirac delta function, i.e., $\delta_\varepsilon \to \delta$ as $\varepsilon \to +0$, in the sense of distributions on $\mathcal{D}'(\mathbb{R})$. Generalizing this, consider the integral

$$J(\varepsilon) := \int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}\varphi(p)/\varepsilon} a(p) d^n p, \qquad \varepsilon > 0.$$

Our goal is to study the asymptotic behavior of this integral as $\varepsilon \to +0$. The method of stationary phase due to Lord Kelvin (1824–1907) tells us that the main contributions come only from momenta p for which the phase function φ is stationary, that is,

$$\varphi'(p) = 0.$$

We assume that both the amplitude function $a : \mathbb{R}^n \to \mathbb{C}$ and the phase function $\varphi : \mathbb{R}^n \to \mathbb{R}$ are smooth, and that the amplitude function a has compact support.

Prototypes. To begin with, let us study some typical examples.

(i) Choose $n = 1$. Fix the nonzero real number x. Then, for each $N = 1, 2, \ldots$,

$$\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}px/\varepsilon} a(p) dp = O(\varepsilon^N) \qquad \text{as} \quad \varepsilon \to +0.$$

This tells us that the integral decreases very rapidly as $\varepsilon \to +0$. This is an immediate consequence of the Paley–Wiener theorem. A completely elementary proof can be based on integration by parts. In fact, as $\varepsilon \to +0$, this integral is equal to

$$\int_{\mathbb{R}} \frac{\varepsilon}{\mathrm{i}x} \left(\frac{d}{dp} \mathrm{e}^{\mathrm{i}px/\varepsilon} \right) a(p) dp = -\frac{\varepsilon}{\mathrm{i}x} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}px/\varepsilon} a'(p) dp = O(\varepsilon).$$

(ii) Let $n = 1$.The phase function $\varphi(p) := p^2/2\varepsilon$ has a critical point at $p = 0$, since $\varphi'(0) = 0$. The modified Fresnel formula tells us that, as $\varepsilon \to +0$,

$$\frac{1}{\sqrt{2\pi\varepsilon}} \int_{\mathbb{R}} \mathrm{e}^{\pm \mathrm{i}p^2/2\varepsilon} a(p) dp = \frac{(1 \pm \mathrm{i})a(0)}{\sqrt{2}} + O(\varepsilon).$$

(iii) Let $n = 1, 2, \ldots$ Suppose that the real $(n \times n)$-matrix A is self-adjoint and $\det A \neq 0$. Then, as $\varepsilon \to +0$,

$$\frac{1}{\sqrt{(2\pi\varepsilon)^n}} \int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}\langle p|Ap\rangle/2\varepsilon} a(p) d^n p = \frac{1 + \mathrm{i}\,\mathrm{sgn}\, A}{\sqrt{2|\det A|}}\, a(0) + O(\varepsilon).$$

Here $\mathrm{sgn}\, A$, the signature of A, is equal to the number of positive eigenvalues of A minus the number of negative eigenvalues.

The main theorem. Suppose that the equation $\varphi'(p) = 0$ has precisely the zeros $p_1, \ldots, p_m$ on the support of the amplitude function a, and these zeros are regular, i.e., $\det \varphi''(p_j) \neq 0$ for $j = 1, \ldots, m$.[17] Then, as $\varepsilon \to +0$,

[17] Recall that $\varphi'(p) = 0$ means that all of the first-order partial derivatives of the phase function φ vanish at the point p,

$$\partial_k \varphi(p) = 0, \qquad k = 1, \ldots, n;$$

the Hessian $\det \varphi''(p)$ is equal to the determinant of the second partial derivatives of the function φ at the point p.

$$\boxed{\frac{1}{\sqrt{(2\pi\varepsilon)^n}}\int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}\varphi(p)/\varepsilon}a(p)d^n p = \sum_{j=1}^{m}\frac{(1+\mathrm{i}\,\mathrm{sgn}\,\varphi''(p_j))\cdot \mathrm{e}^{\mathrm{i}\varphi(p_j)/\varepsilon}a(p_j)}{\sqrt{2|\det\varphi''(p_j)|}},}$$

up to terms of order $O(\varepsilon)$. The proof can be found in Evans (1998) and Guillemin and Sternberg (1989). More general results on the higher asymptotics and degenerate critical phases are contained in the monograph by Hörmander (1983), Vol. 1, Sect. 7.7.

12.5.4 Short-Wave Asymptotics for Electromagnetic Waves

Let us study the singular limit, $\lambda \to 0$, where the wave length of light goes to zero. We want to explain the relations between the following differential equations:

- the Maxwell equations for the electromagnetic field,
- the wave front equation for the Maxwell system (characteristics),
- the characteristics of the wave front equation (bicharacteristics of the Maxwell system which correspond to light rays),
- the wave front equation (characteristics) for the wave equation,
- short waves, the eikonal equation, and the transport equations,
- Fermat's principle of shortest travelling time for light, the Euler–Lagrange equations for light rays, the Hamilton canonical equations along with the Hamilton–Jacobi partial differential equation.

We start with the Maxwell equations

$$\boxed{\begin{array}{ll}\varepsilon\,\mathrm{div}\,\mathbf{E} = \varrho, & \mathrm{div}\,\mathbf{B} = 0,\\ \mathbf{curl}\,\mathbf{E} = \dot{\mathbf{B}}, & \mathbf{curl}\,\mathbf{B} = \mu\mathbf{J} + \varepsilon\mu\dot{\mathbf{E}}.\end{array}} \tag{12.70}$$

These equations refer to homogeneous and isotropic material with electric field constant ε and magnetic field constant μ. Moreover, the quantities possess the following physical meaning: electric field $\mathbf{E}$, magnetic field $\mathbf{B}$, electric charge density ϱ, electric current density vector $\mathbf{J}$. Let $c = 1/\sqrt{\varepsilon_0\mu_0}$ denote the velocity of light in a vacuum. The velocity of light in the material is then given by

$$c_{\mathrm{mat}} = \frac{1}{\sqrt{\varepsilon\mu}} = \frac{n}{c}.$$

Here, n is called the refractive index of the material. We are first looking for wave fronts given by the equation

$$\chi(\mathbf{x}, t) = \mathrm{const}$$

such that discontinuities of the electric field $\mathbf{E}$ or the magnetic field $\mathbf{B}$ propagate along this wave front. It turns out that the function χ has to satisfy the wave front equation

$$\boxed{c_{\mathrm{mat}}^2(\mathbf{grad}\,\chi)^2 = \dot{\chi}^2} \tag{12.71}$$

or $\dot{\chi} = 0$. The dot denotes partial derivative with respect time t. This equation is also called the characteristic equation to the Maxwell equations.[18] Special solutions of the wave front equation are

[18] The characteristic equations for systems of partial differential equations and their applications to basic problems in mathematical physics are considered in Zeidler (2004), Sect. 1.13.3. See also Smirnov (1964), Vol. 4, Sect. III.3.

$$\chi(\mathbf{x},t) := \mathbf{e}\mathbf{x} - c_{\mathrm{mat}}t$$

where $\mathbf{e}$ is a fixed unit vector, and

$$\chi(\mathbf{x},t) := \sqrt{x^2+y^2+z^2} - c_{\mathrm{mat}}t.$$

The equation $\chi(\mathbf{x},t) = \text{const}$ describes then the propagation of both plane waves (in direction of the vector $\mathbf{e}$) and spherical waves with the velocity c_{mat}.

(a) Plane waves: We are given the unit vector $\mathbf{e}$. Suppose that $\varrho \equiv 0, \mathbf{J} \equiv 0$ (no external electric charges and no external electric currents). Then, the Maxwell equations possess the following solutions:[19]

$$\boxed{\mathbf{E}(\mathbf{x},t) := \mathbf{E}_0 \mathrm{e}^{\mathrm{i}k\mathbf{e}\mathbf{x}-\mathrm{i}\omega t}, \qquad \mathbf{B}(\mathbf{x},t) := \mathbf{B}_0 \mathrm{e}^{\mathrm{i}k\mathbf{e}\mathbf{x}-\mathrm{i}\omega t}.} \tag{12.72}$$

This electromagnetic wave propagates in direction of $\mathbf{e}$ with the velocity c_{mat}. The constant vector $\mathbf{E}_0$ cannot be prescribed completely arbitrarily; it has to be perpendicular to the direction $\mathbf{e}$ of propagation (transversal wave). Furthermore,

$$\mathbf{B}_0 = \frac{1}{c_{\mathrm{mat}}}\,(\mathbf{e} \times \mathbf{E}_0).$$

The plane wave has the wave length

$$\lambda = \frac{2\pi}{k}$$

where k is called the wave number. For the relation between frequency ν, angular frequency ω, and velocity of propagation c_{mat}, we get the so-called dispersion relation

$$\nu = \frac{\omega}{2\pi} = \frac{c_{\mathrm{mat}}}{\lambda}.$$

(b) Spherical waves: Let $r := ||\mathbf{x}||$ and $\mathbf{e}_r := \mathbf{x}/r$. The homogeneous wave equation

$$\boxed{\ddot{\psi} + c_{\mathrm{mat}}^2 \Delta\psi = 0} \tag{12.73}$$

has solutions of the form

$$\psi_+(\mathbf{x},t) := \frac{w(r - c_{\mathrm{mat}}t)}{r}, \qquad \mathbf{x} \in \mathbb{R}^3, t > 0,$$

and

$$\psi_-(\mathbf{x},t) := \frac{w(r + c_{\mathrm{mat}}t)}{r}, \qquad \mathbf{x} \in \mathbb{R}^3, t < 0$$

called outgoing and incoming spherical waves, respectively. We are given the smooth function $w :]0,\infty[\to \mathbb{R}$. In order to get such types of waves for the electromagnetic field, we use the scalar potential U and the vector potential potential $\mathbf{A}$ which satisfy the wave equations

$$\ddot{U} + c_{\mathrm{mat}}^2 \Delta U = \frac{c_{\mathrm{mat}}^2 \varrho}{\varepsilon}, \qquad \ddot{\mathbf{A}} + c_{\mathrm{mat}}^2 \Delta \mathbf{A} = \mu c_{\mathrm{mat}}^2 \mathbf{J}$$

along with the Lorentz gauge condition

[19] In order to get real electric and magnetic fields, we have to take the real part (resp. imaginary part) of $\mathbf{E}(\mathbf{x},t)$ and $\mathbf{B}(\mathbf{x},t)$.

$$\dot{U} + c_{\text{mat}}^2 \operatorname{div} \mathbf{A} = 0.$$

We are given smooth functions ϱ and $\mathbf{J}$ with compact support which satisfy the continuity equation $\varrho_t + \operatorname{div} \mathbf{J} = 0$. For $\mathbf{x} \in \mathbb{R}^3$ and $t \in \mathbb{R}^3$, define

$$\boxed{\begin{aligned} U(\mathbf{x},t) &:= \frac{1}{4\pi\varepsilon} \int_{\mathbb{R}^3} \frac{\varrho(\mathbf{y}, t_{\text{ret}})}{||\mathbf{x}-\mathbf{y}||}\, d^3y, \\ \mathbf{A}(\mathbf{x},t) &:= \frac{\mu}{4\pi} \int_{\mathbb{R}^3} \frac{\mathbf{J}(\mathbf{y}, t_{\text{ret}})}{||\mathbf{x}-\mathbf{y}||}\, d^3y \end{aligned}}$$

with the retarded time $t_{\text{ret}} := t - ||\mathbf{x}-\mathbf{y}||/c_{\text{mat}}$. Then, the electromagnetic field

$$\mathbf{E} := -\,\mathbf{grad}\, U - \dot{\mathbf{A}}, \qquad \mathbf{B} := \mathbf{curl}\, \mathbf{A}$$

is a solution of the Maxwell equations (12.70).

(c) The Hertz dipole (the electromagnetic field of an antenna): Consider a negative electric charge $-Q$ at the origin and a positive electric charge Q at the position $\mathbf{x}(t)$ near the origin at time t. Introduce the dipole moment

$$\mathbf{d}(t) := Q\mathbf{x}(t).$$

This situation generates an electromagnetic field. At large distances from the origin, this electromagnetic field looks approximately like

$$\begin{aligned} \mathbf{B}(\mathbf{x},t) &= \frac{\mu}{4\pi c_{\text{mat}} r} \left(\ddot{\mathbf{d}} \left(t - \frac{r}{c_{\text{mat}}} \right) \times \mathbf{e}_r \right), \\ \mathbf{E}(\mathbf{x},t) &= c_{\text{mat}} \mathbf{B}(\mathbf{x},t) \times \mathbf{e}_r. \end{aligned}$$

The eikonal equation and the transport equation. Suppose that we are given $\varrho \equiv 0, \mathbf{J} \equiv 0$. Each smooth solution $\mathbf{E}, \mathbf{B}$ of the Maxwell equations satisfies the wave equations

$$\ddot{\mathbf{E}} + c_{\text{mat}}^2 \Delta \mathbf{E} = 0, \qquad \ddot{\mathbf{B}} + c_{\text{mat}}^2 \Delta \mathbf{B} = 0.$$

However, the converse is not true. Making the ansatz

$$\mathbf{E}(\mathbf{x},t) = \mathrm{e}^{i\omega t}\, \mathbf{E}_0(\mathbf{x}), \qquad \mathbf{B}(\mathbf{x},t) = \mathrm{e}^{i\omega t}\, \mathbf{B}_0(\mathbf{x}),$$

and using the the wave equations, we obtain the Helmholtz equations

$$\Delta \mathbf{E}_0 = k^2 \mathbf{E}_0, \qquad \Delta \mathbf{B}_0 = k^2 \mathbf{B}_0$$

where $k = \omega/c_{\text{mat}}$. In order to study the short-wave asymptotics of the solutions of the Maxwell equations, we now make the crucial ansatz

$$\boxed{\mathbf{E}(\mathbf{x},t) = \mathbf{E_0}(\mathbf{x}) \mathrm{e}^{\mathrm{i}\omega S(\mathbf{x})} \mathrm{e}^{-\mathrm{i}\omega t} + \sum_{r=1}^{\infty} \left(\frac{\mathrm{i}}{\omega} \right)^r \mathbf{E}_r(\mathbf{x}) \mathrm{e}^{\mathrm{i}\omega S(\mathbf{x})} \mathrm{e}^{\mathrm{i}\omega t}.}$$

For the angular frequency, we get

$$\omega = \frac{2\pi c_{\text{mat}}}{\lambda}.$$

We now assume that the wave length λ is small. Substituting this into the wave equation and equating coefficients, we obtain the nonlinear first-order eikonal equation

$$\boxed{c_{\mathrm{mat}}^2 (\mathbf{grad}\, S)^2 = 1} \tag{12.74}$$

for eikonal S, and the linear first-order transport equation

$$\boxed{\mathbf{E}_0 \Delta S - 2\, \mathbf{grad}\, S\, \mathbf{grad}\, \mathbf{E}_0 = 0}$$

for the amplitude $\mathbf{E}_0$. The remaining amplitudes $\mathbf{E}_1, \mathbf{E}_2, \ldots$ follow recursively from the linear first-order transport equations

$$\mathbf{E}_r \Delta S - 2\, \mathbf{grad}\, S\, \mathbf{grad}\, \mathbf{E}_r = \Delta \mathbf{E}_{r-1}, \qquad r = 1, 2, \ldots$$

Replacing the electric field $\mathbf{E}$ by the magnetic field $\mathbf{B}$, we get analogous formulas. This way, we obtain approximate solutions of the Maxwell equations (12.70) which describe short electromagnetic waves. The limit case $\lambda \to 0$ corresponds to geometric optics founded by Fermat (1601–1665). We now want to discuss that

> *There exists a duality between wave fronts and light rays which can be traced back to Huygens* (1625–1695).

This duality was fully developed by the following authors:

- Cauchy (1789–1857) (solution theory for first-order partial differential equations via characteristic curves),
- Hamilton (1805–1865) (canonical equations),
- Jacobi (1804–1851) (Hamilton–Jacobi partial differential equation),
- Lie (1842–1899) (contact transformations),
- Hilbert (1862–1943) (invariant integral),
- Poincaré (1854–1912) and Élie Cartan (1859–1951) (integral invariants),
- Carathéodory (1873–1950) (field theory and the royal road to the calculus of variations), and
- Hölder (1901–1990) (Hölder's contact transformation and the Huygens principle).

We will study this in Volume II. It turns out that:

- Wave fronts are related to the Hamilton–Jacobi partial differential equation (eikonal equation), whereas
- light rays are described by the Euler–Lagrange ordinary differential equations to the Fermat variational principle (and, alternatively, by the Hamilton canonical equations.) The eikonal is the minimal time that is need by light in order to pass from a fixed point (e.g., the origin) to all the other points.
- In the 1950s, this duality played a fundamental role in the foundations of optimal control for dynamical systems. Bellman (1920–1984) based his dynamic programming on the notion of wave fronts (the Hamilton-Jacobi–Bellman equation), whereas Pontryagin (1908–1988) invented his maximum principle which is related to light rays (the canonical Hamilton–Pontryagin equations).[20]

In terms of mathematics,

- the wave fronts of the Maxwell equations correspond to solutions of the characteristic equation of the Maxwell system, and

[20] These interrelationships between the calculus of variations and optimization theory are thoroughly investigated in Zeidler (1986), Vol. 3.

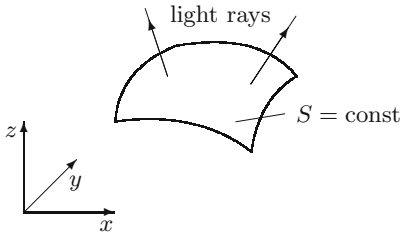


Fig. 12.5. Light rays and wave fronts

- the light rays correspond to the solutions of the bicharacteristic system, which is Cauchy's characteristic system to the characteristic equation of the Maxwell system.

Let us sketch some basic ideas. To simplify notation, we rescale time in such a way that

$$c_{\text{mat}} := 1.$$

The wave front equation (12.71) is a first-order partial differential equation. According to Cauchy, to each such equation there exists a characteristic system of ordinary differential equations.[21] Explicitly, if χ is a solution of the wave front equation (12.71), then the solutions of the following system

$$\dot{\mathbf{x}}(\tau) = -\,\mathbf{grad}\,\chi(\mathbf{x}(\tau), t(\tau)), \quad \dot{t}(\tau) = \chi_t(\mathbf{x}(\tau), t(\tau)) \tag{12.77}$$

for all $\tau \in \mathbb{R}$ are called characteristics of the wave front equation or bicharacteristics of the Maxwell system.[22] Now consider a solution $S = S(\mathbf{x})$ of the eikonal equation (12.74). Then

$$\chi(\mathbf{x}, t) := t - S(\mathbf{x})$$

is a solution of the wavefront equation (12.71). Then $\chi_t = 1$. Thus, $t = \tau$, and the corresponding bicharacteristics have the form

$$\dot{\mathbf{x}}(t) = -\,\mathbf{grad}\,S(\mathbf{x}(t)), \qquad t \in \mathbb{R}.$$

[21] Let $x \in \mathbb{R}^n$. Recall that for a solution $\psi = \psi(x)$ of the first-order partial differential equation

$$F(x, \psi(x), \psi'(x)) = 0, \tag{12.75}$$

Cauchy's characteristic curve system reads as

$$\dot{x} = F_{\psi'}, \quad \dot{p} = -F_x - pF_\psi, \quad \dot{\psi} = \psi' F_{\psi'}. \tag{12.76}$$

Here, the symbol ψ' stands for the tupel of first-order partial derivatives of ψ and the dot denotes the derivative with respect to the curve parameter. Explicitly,

$$\dot{x}(\tau) = F_{\psi'}(x(\tau), \psi(\mathbf{x}(\tau)), \psi'(\mathbf{x}(\tau))), \qquad \tau \in \mathbb{R},$$

and so on. The characteristics can be used in order to construct solutions $\psi(x)$ of (12.75). See Zeidler (2004), Sect. 1.13.3.

[22] Naturally enough, it turns out that the wave front equation (12.71) is also the characteristic equation to the wave equation (12.73).

At fixed time t, the wave front $t - S(\mathbf{x}) = \text{const}$ has the normal vector $\mathbf{grad}\, S(\mathbf{x})$ at the point $\mathbf{x}$. This tells us the crucial fact that

The bicharacteristics are orthogonal to the wave fronts (Fig. 12.5).

For a given wave front $S(\mathbf{x}) = 0$, we construct the family of wave fronts

$$t - S(x) = 0$$

which move with velocity one. The corresponding light rays are perpendicular to the wave fronts. They are straight lines. This simple construction is precisely the special case of Cauchy's general method of characteristics for solving the wave front equation (12.71).

We now want to discuss the relation to both Fermat's classical principle of shortest travelling time for light and the Hamilton–Jacobi theory.

(i) Fermat's principle: Consider the variational problem

$$\boxed{\int_0^{x_1} \sqrt{1 + y'(x)^2}\, dx = \min!} \tag{12.78}$$

with the boundary condition $y(0) = 0, y(x_1) = y_1$. We are looking for a curve $y = y(x)$ which connects the initial point $(0, 0)$ with the end point (x_1, y_1) such that the arc length is minimal. Since the velocity of light is assumed to be equal to 1, we are looking for a light ray which connects the two points in shortest time. This is Fermat's principle. Intuitively, the solution must be a straight line. Rigorously, let us apply the mathematical formalism of the calculus of variations.

(ii) The Euler–Lagrange equation: Introducing the Lagrangian,

$$L(x, y, y') := \sqrt{1 + y'^2},$$

each solution of (12.78) satisfies the Euler–Lagrange equation

$$\boxed{\frac{d}{dx} L_{y'} = L_y.} \tag{12.79}$$

This means that $y''/\sqrt{1 + y'^2} = 0$, and hence $y'' = 0$. Thus, the solutions are straight lines.

(iii) Hamilton's canonical equations: Introducing the momentum $p := L_{y'}$ and the Hamiltonian

$$H(x, y, p) := py' - L(x, y, y') = -\sqrt{1 - p^2},$$

the Euler–Lagrange equations are transformed into

$$\boxed{\dot{p} = -H_y, \qquad \dot{y} = -H_p.} \tag{12.80}$$

The dot denotes the derivative with respect to the curve parameter x.

(iv) The Hamilton–Jacobi differential equation: Let us define the eikonal $S(x_1, y_1)$. This is the time needed by a light ray which travels from the point $(0, 0)$ to the point (x_1, y_1). Explicitly,

$$S(x, y) = \sqrt{x^2 + y^2}.$$

Thus, S is a solution of the Hamilton–Jacobi differential equation

$$\boxed{S_x + H(x, S, S_y) = 0.} \tag{12.81}$$

In fact, $S_x^2 + S_y^2 = 1$.

Note that the canonical equations (12.80) correspond to Cauchy's characteristic curve equations (12.77) for the Hamilton–Jacobi partial differential equation.

The experience of physicists and mathematicians has been that this method can be applied to a broad class of problems in classical mechanics and geometry. We refer to Zeidler (1986), Vols. 3 and 4, Arnold (1978), Scheck (2000), Vol. 1, Boccaletti and Pucacco (1998), Vols. 1, 2 (theory of orbits in celestial mechanics).

12.5.5 Diffraction of Light

> Diffraction problems for light were studied by Fraunhofer (1787–1826), Fresnel (1788–1827), Helmholtz (1821–1894), Kelvin (1824–1907), Kirchhoff (1824–1887), Rayleigh (1842–1919), Poincaré (1854–1912), and Sommerfeld (1868–1951). In his famous lectures on light, Poincaré used Kelvin's method of stationary phase. In the 20th century, the rigorous mathematical treatment of diffraction problems was a challenge for the theory of integral equations. The Kirchhoff–Green representation formula is closely related to the Born approximation and the Lippmann–Schwinger integral equation for scattering processes in quantum physics. The Feynman path integral from the 1940s generalizes wave optics.
>
> Folklore

We want to sketch some basic ideas in diffraction theory. In particular, we want to show why rapidly oscillating integrals (e.g., Fresnel integrals) arise in connection with the diffraction of light.

The Kirchhoff–Green representation formula for the Helmholtz equation. Let Ω be a nonempty bounded open connected set in the 3-dimensional Euclidean space such that $\mathrm{cl}(\Omega)$ is a manifold with boundary. For a fixed real wave number k, let E be a solution of the Helmholtz equation

$$\Delta E - k^2 E = 0 \qquad \text{on} \quad \Omega. \tag{12.82}$$

which is smooth up to the boundary, $E \in C^\infty(\overline{\Omega})$. Then the function E satisfies the following Kirchhoff–Green representation formula

$$\boxed{E(\mathbf{x}_1) = \int_{\partial\Omega} \mathcal{G}(\mathbf{x}_1 - \mathbf{x}) \, \frac{\partial E(\mathbf{x})}{\partial n} - \frac{\partial \mathcal{G}(\mathbf{x}_1 - \mathbf{x})}{\partial n} E(\mathbf{x}) \, dS} \tag{12.83}$$

for all points $\mathbf{x}_1 \in \Omega$. We integrate over $\mathbf{x}$. This equation is fundamental for approximatively solving diffraction problems for light and sound. Here, we use the Helmholtz potential

$$\boxed{\mathcal{G}(\mathbf{x}) := \frac{\mathrm{e}^{-\mathrm{i}kr}}{4\pi r}}$$

with $r := ||\mathbf{x}||$, and $k = 2\pi/\lambda$. For $k = 0$, the Helmholtz potential passes over to the Coulomb potential. Note that

$$\Delta \mathcal{G} - k^2 \mathcal{G} = \delta,$$

in the sense of distributions (see Problem 10.8). In modern language,the Helmholtz potential is a Yukawa potential with imaginary mass $m := \mathrm{i}k$.The Kirchhoff–Green representation formula is a precise formulation of the Huygens principle: The value of the component $E(\mathbf{x}_1)$ of the electric field at the point $\mathbf{x}_1$ is obtained by a superposition of "waves" related to $\mathcal{G}$ which start at the boundary points $\mathbf{x} \in \partial\Omega$. For

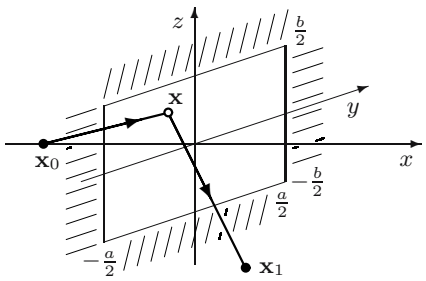


Fig. 12.6. Diffraction of light at a slit

a modern survey on the Hyugens principle, see the article by Belger, Günther, and Schimming (1997).

Sommerfeld's outgoing radiation condition. For fixed real wave number k, the asymptotic condition at infinity,

$$E(\mathbf{x}) = O\left(\frac{1}{r}\right), \qquad \left(\frac{\partial}{\partial r} + \mathrm{i}k\right) E(\mathbf{x}) = o\left(\frac{1}{r}\right), \qquad \text{as} \quad r \to \infty,$$

is called Sommerfeld's outgoing radiation condition. Mnemonically, the Helmholtz potential $\mathcal{G}$ has precisely this asymptotic behavior at infinity. This condition guarantees the uniqueness of the boundary-value problem for the Helmholtz equation. To explain this, suppose that, for fixed $k \in \mathbb{R}$, we are given a smooth function $E_0 : \partial\Omega \to \mathbb{C}$. Then the following are met:

(i) Internal boundary-value problem: The Helmholtz equation (12.82) has precisely one smooth solution $E : \Omega \cup \partial\Omega \to \mathbb{C}$ which attains the given values E_0 on the boundary $\partial\Omega$.

(ii) External boundary-value problem: The Helmholtz equation (12.82) has precisely one solution

$$E : \mathbb{R}^3 \setminus \Omega \to \mathbb{C}$$

which attains the given values E_0 on the boundary Ω and which satisfies Sommerfeld's outgoing radiation condition.

In the 20th century, mathematicians learned how to solve such elliptic problems which generalize the corresponding problems for the Laplace equation. There exist the following methods:

(a) Integral equations (see Kupradse (1956), Smirnov (1964), and Triebel (1989));
(b) Hilbert space methods, Sobolev spaces, and the regularity properties of generalized solutions (see Zeidler (1986), Vols. IIA, IIB, Zeidler (1995), Vol. I, Chap. 2, Evans (1998), and Jost (2002b));
(c) Schauder estimates and the continuation method (see Zeidler (1986), Vol. I, Chap. 6, and Jost (2002b)).

This program was initiated by Hilbert's problem number twenty that he formulated in his famous 1900 Paris lecture (see Hilbert (1900)).

The slit formula. We now want to apply the Kirchhoff–Green representation formula in order to approximately solve a diffraction problem for light as pictured in Fig. 12.6. We assume that the (y, z)-plane forms a screen which has a slit

$$\Sigma := \{(y, z) \in \mathbb{R}^2 : \ -\tfrac{1}{2}a \le y \le \tfrac{1}{2}a, \ -\tfrac{1}{2}b \le z \le \tfrac{1}{2}b\}.$$

Electromagnetic light waves start at the source point $\mathbf{x}_0$; they pass through an arbitrary point $\mathbf{x}$ of the slit, and they are observed at the point $\mathbf{x}_1$. We choose

$$\mathbf{x}_0 := -r_0\mathbf{i}, \qquad \mathbf{x}_1 := x_1\mathbf{i} + y_1\mathbf{x}_2 + z_1\mathbf{k}.$$

Furthermore, we introduce the unit vectors $\mathbf{e_0} := \mathbf{x}_0/r_0$ and $\mathbf{e}_1 := \mathbf{x}_1/r_1$. We want to compute the electromagnetic field $\mathbf{E}(\mathbf{x}_1), \mathbf{B}(\mathbf{x}_1)$ at the final point $\mathbf{x}_1$. More precisely, we are interested in the intensity of the light given by the energy density

$$I := \tfrac{1}{2}\varepsilon(|\mathbf{E}(\mathbf{x})|^2 + c_{\text{mat}}^2|\mathbf{B}(\mathbf{x})|^2)$$

at the point $\mathbf{x}$. To simplify the approach, we restrict ourselves to a fixed component of the electric harmonic waves

$$\mathbf{E}(\mathbf{x}) = \mathbf{E}_0(\mathbf{x})\mathrm{e}^{-\mathrm{i}\omega t}$$

with angular frequency ω. For the wave number and the wave length, we have

$$k = \frac{\omega}{c_{\text{mat}}}, \qquad \lambda = \frac{2\pi}{k} = \frac{c_{\text{mat}}}{\omega}.$$

Here, c_{mat} is the velocity of light in the medium (e.g., air). For example, we set

$$E(\mathbf{x}) := (E_0)_y(\mathbf{x}),$$

that is, we study the y-component of the electric field. Then, E is a solution of the Helmholtz equation (12.82). Since E will be complex-valued, we will measure the intensity of light I at the point $\mathbf{x}$ by the quantity

$$I = \frac{\varepsilon}{2}\,|E(\mathbf{x})|^2.$$

Physical experiments show that diffraction patterns may look rather complicated. Our approximate approach is based on the following assumptions.

(a) The source point $\mathbf{x}_0$ sends a wave $E(\mathbf{x}) = \mathcal{G}(\mathbf{x} - \mathbf{x}_0)$ to the point $\mathbf{x}$ of the slit.

(b) The wave observed at the final point $\mathbf{x}_1$ only depends on the light that comes from the points $\mathbf{x}$ of the slit Σ. By the Kirchhoff–Green representation formula,

$$E(\mathbf{x}_1) = \int_\Sigma \left(\mathcal{G}(\mathbf{x} - \mathbf{x}_1)\,\frac{\partial\mathcal{G}(\mathbf{x} - \mathbf{x}_0)}{\partial n} - \frac{\partial\mathcal{G}(\mathbf{x} - \mathbf{x}_1)}{\partial n}\mathcal{G}(\mathbf{x} - \mathbf{x}_0)\right) dS.$$

This is the desired formula for the electric field component $E(\mathbf{x}_1)$ at the point $\mathbf{x}_1$ of the observer.

(c) To simplify the approximate computation of the integral for $E(\mathbf{x}_1)$, we assume that

$$\frac{a}{\lambda} \sim \frac{b}{\lambda} \sim 1, \qquad \frac{r_0}{\lambda} \gg 1, \qquad \frac{r_1}{\lambda} \gg 1,$$

that is, the width of the slit a, b is of the magnitude of the wave length λ, and the distance between the source (resp. the observer) and the screen is large compared with the wave length λ. For visible light, $\lambda \sim 6 \cdot 10^{-7}$m. By Taylor expansion, we get the following approximation

$$\boxed{E(\mathbf{x}_1) = -\frac{\mathrm{e}^{-\mathrm{i}kr_0}}{4\pi r_0} \cdot \frac{\mathrm{e}^{-\mathrm{i}kr_1}}{4\pi r_1} \cdot \mathrm{i}kA\left(1 + \frac{y_1}{r_1}\right)} \tag{12.84}$$

along with

$$A := \int_{\Sigma} \mathrm{e}^{-kig(\mathbf{x})} dS_{\mathbf{x}}$$

and $g(\mathbf{x}) := ||\mathbf{x} - \mathbf{x_0}|| + ||\mathbf{x} - \mathbf{x}_1|| - r_0 - r_1$. Again by Taylor expansion, we obtain

$$\boxed{g(x) = -(\mathbf{e}_0 + \mathbf{e}_1)\mathbf{x} + \frac{(\mathbf{e}_0 \times \mathbf{x})^2}{2r_0} + \frac{(\mathbf{e}_1 \times \mathbf{x})^2}{2r_1},}$$

up to terms of order $O(\frac{1}{r_0^2}) + O(\frac{1}{r_1^2})$ as $r_0, r_1 \to \infty$.

Fraunhofer diffraction. For a point $\mathbf{x} = x\mathbf{i} + y\mathbf{j}$ of the slit, we use the first-order approximation

$$g(\mathbf{x}) = -(\mathbf{e}_0 + \mathbf{e}_1)\mathbf{x} = -y_1 y - z_1 z.$$

Hence

$$A = \int_{-\frac{1}{2}a}^{\frac{1}{2}a} \int_{-\frac{1}{2}b}^{\frac{1}{2}b} \mathrm{e}^{\mathrm{i}ky + \mathrm{i}kz} dy dz = \frac{\sin(\kappa_y a)}{\kappa_y} \cdot \frac{\sin(\kappa_z b)}{\kappa_z}$$

where

$$\kappa_y := \frac{\pi}{\lambda} \cdot \frac{y_1}{r_1}, \qquad \kappa_z := \frac{\pi}{\lambda} \cdot \frac{z_1}{r_1}.$$

For the intensity of the light $I = \frac{1}{2}\varepsilon |E(\mathbf{x})_1|^2$ at the point $\mathbf{x}_1$, we get

$$\boxed{I = I_0 \frac{\sin^2(\kappa_y a)}{\kappa_y^2} \cdot \frac{\sin^2(\kappa_z b)}{\kappa_z^2}}$$

along with

$$I_0 = \frac{\varepsilon}{132\pi^2\lambda^2 r_0^2 r_1^2} \left(1 + \frac{x_1}{r_1}\right)^2.$$

Fresnel diffraction. In the special case where we have

$$\mathbf{x}_0 = -\mathbf{x}_1$$

for the source point $\mathbf{x}_0$ and the observer point $\mathbf{x}_1$ in Fig. 12.6 on 725, we get $g(\mathbf{x}) = 0$ in first-order approximation. That is, the Fraunhofer approximation of diffraction vanishes. In order to describe the diffraction observed in physical experiments, we now have to take the second-order approximation

$$g(\mathbf{x}) = \frac{(\mathbf{i} \times \mathbf{x})^2}{r_0} = \frac{y^2 + z^2}{r_0}$$

into account, which corresponds to Fresnel diffraction. In this case, we get the electric field (12.84) with

$$A = \int_{-\frac{1}{2}a}^{\frac{1}{2}a} \mathrm{e}^{-\mathrm{i}ky^2/r_0} dy \int_{-\frac{1}{2}b}^{\frac{1}{2}b} \mathrm{e}^{-\mathrm{i}kz^2/r_0} dz,$$

and $k = 2\pi/\lambda$. This is a product of Fresnel integrals. For a more detailed study of diffraction problems, we refer to the following monographs:

W. Macke, Wellen (Waves) (in German), Geest & Portig, Leipzig, 1959.

G. Scharf, From Electrostatics to Optics, Springer, Berlin, 1994.

M. Born and E. Wolf, Principles of Optics, 4th edn., Pergamon Press, New York, 1970 (classic).

Our computations above on Fraunhofer diffraction and Fresnel diffraction follow the monograph by Macke.

12.5.6 Pseudo-Differential Operators

To explain the basic idea, consider the linear differential operator

$$(Lu)(t) := a(t)\frac{d^k}{dt^k}u(t), \qquad t \in \mathbb{R}$$

of order $k = 1, 2, \ldots$ with smooth coefficient function $a : \mathbb{R} \to \mathbb{R}$. For all functions $u \in \mathcal{S}(\mathbb{R})$, we have the Fourier transform

$$u(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\omega t} \mathcal{F}(u)(\omega) d\omega.$$

Hence

$$(Lu)(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\omega t} s(t, \omega) \mathcal{F}(u)(\omega) d\omega$$

where the smooth function

$$s(t, \omega) := a(t)(\mathrm{i}\omega)^k, \qquad t, \omega \in \mathbb{R}$$

is called the symbol of the differential operator L. Generally, a pseudo-differential operator P is given by an expression of the form

$$\boxed{(Pu)(x) := \frac{1}{\sqrt{(2\pi)^n}} \int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}\langle x|p\rangle}\ s(x, p)\mathcal{F}(u)(p)\ d^n p, \qquad x \in \mathbb{R}^n.}$$

The theory of pseudo-differential operators studies the properties of such operators in terms of the properties of the symbol $s(x, p)$. For example,

- linear partial differential operators with smooth coefficient functions and
- important classes of regular and singular integral operators

are pseudo-differential operators. In the history of mathematics and physics, efforts were made to reduce the solution of differential equations to the solution of simpler algebraic problems. Such efforts are due to Leibniz (1646–1716), Laplace (1749–1827), Fourier (1768–1830), and the British engineer Heaviside (1850–1925) to mention a few names. Nowadays this development has culminated in the theory of pseudo-differential operators.

12.5.7 Fourier Integral Operators

The replacement

$$\mathrm{e}^{\mathrm{i}\langle x|p\rangle} \Rightarrow \mathrm{e}^{\mathrm{i}\varphi(x,p)}$$

leads us immediately to the notion of Fourier integral operator,

$$\boxed{(Pu)(x) := \frac{1}{\sqrt{(2\pi)^n}} \int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}\varphi(x,p)}\ s(x, p)\mathcal{F}(u)(p)\ d^n p, \qquad x \in \mathbb{R}^n.}$$

Here, we assume that the phase function $\varphi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ has the following properties:

(i) Homogeneity: $\varphi(x, \lambda p) = \lambda\varphi(x, p)$ for all $\lambda > 0$ and all $x, p \in \mathbb{R}^n$.
(ii) Smoothness: $\varphi(x, p)$ is smooth for all $x, p \in \mathbb{R}^n$ with $p \neq 0$.
(iii) Nondegeneracy: $\varphi_x(x, p) \neq 0$ for all $x, p \in \mathbb{R}^n$ with $p \neq 0$.

Observe that the conditions (i)-(iii) are satisfied in the special case where the phase function is given by $\varphi(x,p) := \langle x|p\rangle$.

Typical application. Consider the initial- value problem for the wave equation:

$$U_{tt}(\mathbf{x},t) + \Delta U(\mathbf{x},t) = 0, \qquad \mathbf{x} \in \mathbb{R}^3, t \geq 0$$
$$U(\mathbf{x},0) = U_0(\mathbf{x}), \quad U_t(\mathbf{x},0) = U_1(\mathbf{x}), \qquad \mathbf{x} \in \mathbb{R}^3.$$

We are given the smooth functions $U_0, U_1 : \mathbb{R}^3 \to \mathbb{R}$ with compact support. Then there exists a unique smooth solution given by

$$\boxed{U(\mathbf{x},t) = \frac{1}{2\sqrt{(2\pi)^3}} \int_{\mathbb{R}^3} \sum_{s=0}^{1} \mathrm{e}^{\mathrm{i}\mathbf{x}\mathbf{p} + \mathrm{i}(-1)^s t|k|} \left(V_0(\mathbf{p}) - (-1)^s \mathrm{i} \frac{V_1(\mathbf{p})}{|\mathbf{p}|} \right) d^3p}$$

where V_0 and V_1 are the Fourier transform of U_0 and U_1, respectively. Note that the solution U is represented as the sum of two Fourier integral operators. To motivate this formula, let us use the classical Fourier method. To begin with, we introduce the Fourier transform of U with respect to the variable $\mathbf{x}$,

$$V(\mathbf{x},t) = \frac{1}{\sqrt{(2\pi)^3}} \int_{\mathbb{R}^3} \mathrm{e}^{-\mathrm{i}\mathbf{x}\mathbf{p}} \; U(\mathbf{x},t) d^3x.$$

The wave equation implies

$$V_{tt}(\mathbf{p},t) = -\mathbf{p}^2 V(\mathbf{p},t)$$

along with the initial conditions $V(\mathbf{p},0) = V_0(\mathbf{p})$ and $V_t(\mathbf{p},0) = V_1(\mathbf{p})$. This yields

$$V(\mathbf{p},t) = \frac{1}{2} \sum_{s=0}^{1} V_0(\mathbf{p}) - \mathrm{i}(-1)^s \frac{V_1(\mathbf{p})}{|\mathbf{p}|}.$$

Finally, by the inverse Fourier transform,

$$U(\mathbf{x},t) = \frac{1}{\sqrt{(2\pi)^3}} \int_{\mathbb{R}^3} \mathrm{e}^{\mathrm{i}\mathbf{x}\mathbf{p}} \; V(\mathbf{p},t) \; d^3p.$$

It turns out that

> *Fourier integral operators play a fundamental role in quantum field theory for describing the propagation of physical effects.*

Recommendations for further reading on both pseudo-differential operators and Fourier integral operators can be found on page 703.

12.6 Multiplication of Distributions

In a naive setting of quantum field theory, one encounters the square $\delta(x)^2$ of Dirac's delta function when computing cross sections of scattering processes. In 1954 Laurent Schwartz showed that it is not possible to construct a perfect theory of products FG for distributions $F, G \in \mathcal{D}'(\mathbb{R})$ without leaving the space $\mathcal{D}'(\mathbb{R})$ of distributions.

12.6.1 Laurent Schwartz's Counterexample

> Intelligence consists of this; that we recognize the similarity of different things and the difference between similar ones.
>
> Baron de la Brède et de Montesquieu (1689–1755)

Consider the space $\mathcal{D}(\mathbb{R})$ of smooth functions $\varphi : \mathbb{R} \to \mathbb{C}$ which have compact support, and let $\mathcal{D}'(\mathbb{R})$ be the corresponding space of distributions. To each function $f \in \mathcal{D}(\mathbb{R})$, there can be assigned a unique distribution F given by

$$F(\varphi) := \int_{\mathbb{R}} f(x)\varphi(x)dx \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}).$$

With respect to the usual product fg of functions, the space $\mathcal{D}(\mathbb{R})$ is a commutative and associative algebra with Leibniz rule. Explicitly, this means the following:

(i) Linearity: $\mathcal{D}(\mathbb{R})$ is a complex linear space.
(ii)) Product: If $f, g \in \mathcal{D}(\mathbb{R})$, then there exists a product $fg \in \mathcal{D}(\mathbb{R})$.
(iii)) Distributivity: $f(\alpha g + \beta h) = \alpha fg + \beta fh$ for all $f, g, h \in \mathcal{D}(\mathbb{R})$ and all complex numbers α, β.
(iv) Commutativity: $fg = gf$ for all $f, g \in \mathcal{D}(\mathbb{R})$.
(v) Associativity: $(fg)h = f(gh)$ for all $f, g, h \in \mathcal{D}(\mathbb{R})$.
(vi) Leibniz rule: The derivative D is a linear operator on $\mathcal{D}(\mathbb{R})$, and we have

$$D(fg) = (Df)g + f(Dg) \qquad \text{for all} \quad f, g \in \mathcal{D}(\mathbb{R}).$$

The following theorem was proved by Laurent Schwartz.[23]

Theorem 12.10 *It is not possible to equip the complex linear space $\mathcal{D}'(\mathbb{R})$ of distributions with a product such that it becomes a quite natural extension of the function algebra $\mathcal{D}(\mathbb{R})$.*

More precisely, the function algebra $\mathcal{D}(\mathbb{R})$ cannot be extended to a commutative and associative algebra $\mathcal{D}'(\mathbb{R})$ with Leibniz rule

$$D(FG) = (DF)G + F(DG) \qquad \text{for all} \quad F, G \in \mathcal{D}'(\mathbb{R}),$$

where D denotes the usual derivative for distributions.

Proof. We will argue by contradiction. Suppose that there exists a product on $\mathcal{D}'(\mathbb{R})$ which has the desired properties, that is, there hold true the distributive law, the commutative law, the associative law, and the Leibniz rule.

(I) The Dirac delta functional. The basic trick is to consider the following functions:

- $f(x) := x$ and $g(x) := |x|$ for all $x \in \mathbb{R}$.
- $h(x) := \ln|x| - 1$ for all $x \in \mathbb{R} \setminus \{0\}$.

Classically, the derivative $(x|x|)' = 2|x|$ exists for all $x \in \mathbb{R}$. Hence

$$Df = 1 \qquad \text{and} \qquad D(fg) = 2g.$$

Using the Heaviside function θ, we get

$$Dg = -1 + 2\theta \qquad \text{and hence} \quad D^2 g = 2\delta.$$

[23] L. Schwartz, Sur l'impossibilité de la multiplication des distributions. Comptes Rendus Acad. Sci. Paris **239** (1954), 847–848.

To obtain the desired contradiction, we will show that our assumption implies

$$\boxed{D^2 g = 0.} \tag{12.85}$$

(II) By the presupposed Leibniz rule on $\mathcal{D}'(\mathbb{R})$,

$$D(fg) = (Df)g + fDg = g + fDg.$$

Thus, again by the Leibniz rule,

$$D^2(fg) = 2Dg + fD^2 g. \tag{12.86}$$

On the other hand, from $D(fg) = 2g$ we get $D^2(fg) = 2Dg$. Combining this with (12.86), we obtain the key relation

$$fD^2 g = 0.$$

This yields the desired equation (12.85) if the following claim is true for all distributions $G \in \mathcal{D}'(\mathbb{R})$:

$$fG = 0 \qquad \text{implies} \quad G = 0. \tag{12.87}$$

It remains to prove (12.87).

(III) Observe that (12.87) follows from

$$fD^2(fh) = 1. \tag{12.88}$$

In fact, if $fG = 0$, then $Gf = 0$, by commutativity. Associativity tells us that

$$0 = GfD^2(fh) = G \cdot 1,$$

and hence $G = 0$. It remains to prove (12.88).

(IV) Proof of (12.88). Classically, $(x^2(\ln|x| - 1))' = 2x(\ln|x| - 1) + x$ for all $x \in \mathbb{R}$. Equivalently,

$$D(f^2 h) = 2fh + f. \tag{12.89}$$

By the presupposed Leibniz rule on $\mathcal{D}'(\mathbb{R})$ along with $Df = 1$, we obtain

$$D(f^2 h) = D(f \cdot fh) = fh + fD(fh).$$

Therefore, by the Leibniz rule, $D^2(f^2 h) = 2D(fh) + fD^2(fh)$, and hence

$$fD^2(fh) = D^2(f^2 h) - 2D(fh).$$

Finally, again by the Leibniz rule, it follows from (12.89) that

$$fD^2(fh) = 2D(fh) + 1 - 2D(fh) = 1.$$

□

12.6.2 Hörmander's Causal Product

Causality is crucial for quantum field theory.
Folklore

Based on his notion of the wave front set of a distribution, Hörmander proved the following theorem in 1970.

Theorem 12.11 *It is possible to define the product GH of two given distributions $G, H \in \mathcal{D}'(\mathbb{R}^N)$ if the wave front sets $WF(G)$ and $WF(H)$ possess the following crucial momentum property:*

$$(x,k) \in WF(G), \quad \textit{always implies} \quad (x,-k) \notin WF(H). \tag{12.90}$$

It turns out that this condition is satisfied in the Epstein–Glaser approach to quantum field theory, because of the validity of the causality principle for the S-matrix. This will be thoroughly studied in Volume II on quantum electrodynamics.

For two functions $g, h \in \mathcal{D}(\mathbb{R}^N)$, we have the convolution rule for the Fourier transform:

$$\mathcal{F}(gh) = \frac{1}{(2\pi)^{N/2}} \cdot \mathcal{F}(g) * \mathcal{F}(h).$$

Naturally enough, the product GH is given in a similar way after localization. Explicitly, for each point $x \in \mathbb{R}^N$ there exists a function $\chi \in \mathcal{D}(\mathbb{R}^N)$ such that $\chi = 1$ in some open neighborhood of x, and for the Fourier transform of the localizations we have the convolution rule

$$\boxed{\mathcal{F}(\chi^2 GH) = \frac{1}{(2\pi)^{N/2}} \cdot \mathcal{F}(\chi G) * \mathcal{F}(\chi H).} \tag{12.91}$$

As a counterexample, consider the Dirac delta distribution δ on $\mathbb{R}$. For the wave frontset

$$WF(\delta) = \{(0,k) : k \in \mathbb{R}, \ \ k \neq 0\}.$$

Therefore, condition (12.90) is violated for $G = H := \delta$. In this case, it is not possible to define the product δ^2.

The theory of products for distributions can be found in Reed and Simon (1972), Vol. II, Sect. IX.10, and in Hörmander (1983), Vol. I, Sect. 8.2. In order to get a general theory of products, one has to extend the class of Schwartz distributions to a broader class of generalized functions. In this connection, we refer to the monographs:

J. Colombeau, New Generalized Functions and Multiplication of Distributions, North-Holland, Amsterdam, 1984.

M. Oberguggenberger, Multiplication of Distributions and Applications to Partial Differential Equations, Longman, Harlow, 1992.

Problems

12.1 *Cauchy's residue method.* Prove that for each real number ω_0 and each $\varepsilon > 0$,

$$\int_{-\infty}^{\infty} \frac{d\omega}{\omega^2 - (\omega_0 + \varepsilon \mathrm{i})^2} = \frac{\pi \mathrm{i}}{\omega_0 + \varepsilon \mathrm{i}}.$$

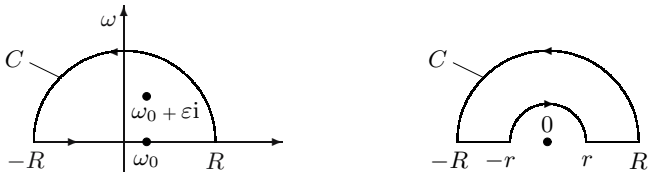


Fig. 12.7. Cauchy's residue method

Solution: Set $\omega_1 := \omega_0 + \varepsilon\mathrm{i}$. Then

$$\frac{1}{\omega^2 - \omega_1^2} = \frac{1}{2\omega_1}\left(\frac{1}{\omega - \omega_1} - \frac{1}{\omega + \omega_1}\right).$$

Using the semicircle C pictured in Fig. 12.7(a), Cauchy's residue theorem on page 214 tells us that

$$\int_C \frac{d\omega}{\omega^2 - \omega_1^2} = \frac{\pi\mathrm{i}}{\omega_1}.$$

Note that

$$\int_C f(\omega)\,d\omega = \int_{-R}^{R} f(x)\,dx + \int_0^{\pi} f(R\mathrm{e}^{\mathrm{i}\varphi})\mathrm{i}R\mathrm{e}^{\mathrm{i}\varphi}\,d\varphi.$$

Letting $R \to \infty$, the second integral goes to zero, since $(\omega^2 - \omega_1^2)^{-1} = O(R^{-2})$.

12.2 *The classical Dirichlet integral.* Use Cauchy's residue method in order to prove that

$$\int_{-\infty}^{\infty} \frac{\sin(\alpha x)}{x}\,dx = \pi \operatorname{sgn}\alpha \qquad \text{for all} \quad \alpha \in \mathbb{R}.$$

Solution: By rescaling $x \mapsto \alpha x$, it is sufficient to consider the case where $\alpha = 1$. Using the closed curve C from Fig. 12.7(b),

$$\int_C \frac{\mathrm{e}^{\mathrm{i}z}dz}{z} = 0,$$

since the integrand is analytic inside C. For the semicircle of radius R,

$$\lim_{R\to\infty} \int_0^{\pi} \frac{\mathrm{e}^{\mathrm{i}R\cos\varphi}\mathrm{e}^{-R\sin\varphi}\mathrm{i}R\mathrm{e}^{\mathrm{i}\varphi}d\varphi}{R\mathrm{e}^{\mathrm{i}\varphi}} = 0,$$

because of $\lim_{R\to\infty} \mathrm{e}^{-R\sin\varphi} = 0$. For the semicircle of radius r,

$$\lim_{r\to 0} \int_{\pi}^{0} \frac{\mathrm{e}^{\mathrm{i}r\cos\varphi}\mathrm{e}^{-r\sin\varphi}\mathrm{i}r\mathrm{e}^{\mathrm{i}\varphi}d\varphi}{r\mathrm{e}^{\mathrm{i}\varphi}} = \int_{\pi}^{0} \mathrm{i}\,d\varphi = -\mathrm{i}\pi.$$

Hence the sum

$$\int_{-R}^{-r} \frac{\cos x + \mathrm{i}\sin x}{x}\,dx + \int_{R}^{r} \frac{\cos x + \mathrm{i}\sin x}{x}\,dx$$

goes to $\mathrm{i}\pi$ as $R \to +\infty$ and $r \to +0$.

12.3 *The classical Fresnel integral.* Use Cauchy's residue method in order to prove that

$$\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}x^2}\, dx = \int_{-\infty}^{\infty} (\cos x^2 + \mathrm{i} \sin x^2) dx = (1+\mathrm{i})\sqrt{\frac{\pi}{2}}.$$

Hint: See Smirnov (1964), Vol. III/2, Sect. 56.

12.4 *The Dirichlet relation.* Prove that for each function $\varphi \in \mathcal{S}(\mathbb{R})$,

$$\lim_{\omega \to +\infty} \int_{-\infty}^{\infty} \frac{\sin \omega t}{\pi t}\, \varphi(t) dt = \varphi(0).$$

This means that

$$\lim_{\omega \to +\infty} \frac{\sin \omega t}{\pi t} = \delta,$$

in the sense of tempered distributions.
Hint: Justify the limit

$$\lim_{\omega \to +\infty} \int_{-\infty}^{\infty} \frac{\sin \omega t}{\pi t}\, \varphi(t) dt = \lim_{\omega \to +\infty} \int_{-\infty}^{\infty} \frac{\sin \tau}{\pi \tau}\, \varphi\left(\frac{\tau}{\omega}\right) d\tau = \int_{-\infty}^{\infty} \frac{\sin \tau}{\pi \tau}\, \varphi(0) d\tau.$$

The latter integral is equal to $\varphi(0)$. See also the proof given by Gelfand et al. (1964), Vol. 1, Sect. I.2.

12.5 *The Fourier relation.* Show that

$$\lim_{P \to +\infty} \frac{1}{2\pi\hbar} \int_{-P}^{P} \mathrm{e}^{\mathrm{i}px/\hbar}\, dp = \delta, \tag{12.92}$$

in the sense of tempered distributions. Mnemonically, we write

$$\frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}px/\hbar} dp = \delta(x).$$

Solution: Use Problem 11.4 and note that

$$\frac{1}{2\pi\hbar} \int_{-P}^{P} \mathrm{e}^{\mathrm{i}px/\hbar} dp = \frac{\sin(Px/\hbar)}{\pi x}.$$

12.6 *Proof of Proposition 12.1 on page 672 about the discrete Dirac function.*
Solution: To simplify notation, we work in the energetic system. Then, $p^0 = E$ and $c = \hbar = 1$.
Ad (i). Use Problem 11.4.
Ad (ii). Use (12.9) and (12.10) on page 672.
Ad (iii). We have to show that

$$\lim_{T \to +\infty} \frac{\int_{-\infty}^{\infty} \delta_T(E - E_1) \cdot \delta_T(E - E_2) dE}{\delta_T(E_1 - E_2)} = 1.$$

For $E_1 = E_2$, this is the classical formula

$$\int_{-\infty}^{\infty} \frac{2}{\pi T E^2} \sin^2\left(\frac{TE}{2}\right) dE = 1$$

which follows from

$$\int_{-\infty}^{\infty} \frac{\sin^2 x}{\pi x^2}\, dx = 1.$$

In the case where $E_1 \neq E_2$, let T be sufficiently large. Intuitively, the function $E \mapsto \delta_T(E - E_1)$ is sharply concentrated at the point E_1, and

$$\int_{-\infty}^{\infty} \delta_T(E - E_1) dE = 1.$$

Hence, up to some small error,

$$\int_{-\infty}^{\infty} \delta_T(E - E_1) \cdot \delta_T(E - E_2) dt = \delta_T(E_1 - E_2).$$

This intuitive argument can be made precise by using appropriate estimates.

12.7 *Proof of Proposition 11.12 on page 615.* Solution: Set

$$|\varphi|_k := \sup_{x \in \mathbb{R}^N} (1 + ||x||^k)|\varphi(x)|.$$

If we choose the integer k sufficiently large, then

$$|f(\varphi)\,| \leq \text{const} \cdot p_k(\varphi) \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}^N).$$

In fact,

$$\begin{aligned} |f(\varphi)| &= \left| \int_{\mathbb{R}^N} f(x)(1 + ||x||^k)^{-1} \varphi(x)(1 + ||x||^k) d^N x \right| \\ &\leq p_k(\varphi) \int_{\mathbb{R}^N} |f(x)|(1 + ||x||^k)^{-1} d^N x. \end{aligned}$$

If $\varphi_n \to \varphi$ as $n \to \infty$ in $\mathcal{S}(\mathbb{R}^N)$, then $p_k(\varphi_n - \varphi) \to 0$. Hence

$$|f(\varphi_n) - f(\varphi)\,| = |f(\varphi_n - \varphi)\,| \leq \text{const} \cdot p_k(\varphi_n - \varphi) \to 0.$$

12.8 *Special distribution.* We want to generalize the classical relation

$$\frac{d \ln |x|}{dx} = \frac{1}{x} \qquad \text{for all} \quad x \in \mathbb{R} \setminus \{0\}.$$

To this end, we set

$$\mathcal{P}(\ln |x|)(\varphi) := \int_{-\infty}^{\infty} (\varphi(x) - \varphi(0)) \ln |x| \, dx \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).$$

Show that this is a tempered distribution along with the derivative

$$\frac{d}{dx} \mathcal{P}(\ln |x|) = \mathcal{P}\left(\frac{1}{x}\right)$$

in the space $\mathcal{S}'(\mathbb{R})$.

Solution: Let $\varphi \in \mathcal{S}(\mathbb{R})$. Note that $\varphi(x) - \varphi(0) = O(x)$ as $x \to 0$ and $\lim_{x \to 0} \; x \ln |x| = 0$. By Problem 12.7, $\mathcal{P}(\ln |x|)$ is a tempered distribution. Furthermore, $\mathcal{P}(\ln |x|)(\varphi)$ is equal to

$$\lim_{\varepsilon \to +0} \left(\int_{-\infty}^{-\varepsilon} (\varphi(x) - \varphi(0)) \ln |x| \, dx + \int_{\varepsilon}^{\infty} (\varphi(x) - \varphi(0)) \ln |x| \, dx \right).$$

Integration by parts yields that $\mathcal{P}(\ln |x|)(-\varphi')$ is equal to

$$-\lim_{\varepsilon\to+0}\left(\int_{-\infty}^{-\varepsilon}(\varphi'(x)-\varphi'(0))\ln|x|\,dx+\int_{\varepsilon}^{\infty}(\varphi'(x)-\varphi'(0))\ln|x|\,dx\right)$$
$$=\lim_{\varepsilon\to+0}\left(\int_{-\infty}^{-\varepsilon}\frac{\varphi(x)-\varphi(0)}{x}\,dx+\int_{\varepsilon}^{\infty}\frac{\varphi(x)-\varphi(0)}{x}\,dx\right)=\mathcal{P}\left(\frac{1}{x}\right).$$

Note that the boundary terms vanish as $\varepsilon\to+0$ because of

$$\lim_{\varepsilon\to+0}(\varphi(\pm\varepsilon)-\varphi(0))\ln|\pm\varepsilon|=0.$$

12.9 *The jump trick for ordinary differential equations.* Let $a,b,c:\mathbb{R}\to\mathbb{C}$ be given smooth functions. Suppose that the smooth function $g:\mathbb{R}\to\mathbb{C}$ is a solution of the ordinary differential equation

$$ag''+bg'+cg=0\quad\text{on}\quad\mathbb{R}^2$$

with the initial conditions $g(0)=0$ and $g'(0)=1$. Show that the function

$$q(t):=\theta(t)g(t)\qquad\text{for all}\quad t\in\mathbb{R}$$

satisfies the differential equation

$$aq''+bq'+cq=a\delta\quad\text{on}\quad\mathbb{R}^2,$$

in the sense of distributions.
Solution: Let us start with a formal argument. Since $q(0)=0$, $q\delta=0$. Thus, by the product rule,
(i) $q'=\theta'g+\theta g'=g\delta+\theta g'=\theta g'$, and
(ii) $q''=\theta'g'+\theta g''=g'\delta+\theta g''=\delta+\theta g''$.
This yields the claim. To get a rigorous proof, we have to show that (i) and (ii) are valid in the sense of distributions.
Ad (i). Since $g(0)=0$ and $\theta(t)=0$ if $t<0$, integration by parts yields

$$-\int_{-\infty}^{\infty}\varphi'(t)\theta(t)g(t)\,dt=\int_{-\infty}^{\infty}\varphi(t)\theta(t)g'(t)\,dt$$

for all test functions $\varphi\in\mathcal{D}(\mathbb{R})$.
Ad(ii). We have to show that, for all $\varphi\in\mathcal{D}(\mathbb{R})$,

$$\int_{-\infty}^{\infty}\varphi''(t)\theta(t)g(t)\,dt=\varphi(0)+\int_{-\infty}^{\infty}\varphi(t)\theta(t)g''(t)\,dt.$$

This is equivalent to

$$\int_{0}^{\infty}\varphi''(t)g(t)\,dt=\varphi(0)+\int_{0}^{\infty}\varphi(t)g''(t)\,dt.$$

However, this follows from applying integration by parts twice. Note that $g(0)=0$ and $g'(0)=1$.

12.10 *Proof of Proposition 11.21 on page 627.* Complete the proof in the general case where the given function F lies in the Banach space C^-_π. Use the methods of functional analysis.
Hint: Let us sketch the main ideas. We will use results from the author's textbooks, Zeidler (1995), Vols. 1, 2. The notation is the same as introduced in Sect. 11.5 on page 625. To begin with, we define the function space $C^-_{2,\pi}$ which

contains precisely the π-periodic functions $f : \mathbb{R} \to \mathbb{C}$ which are odd. In addition, we assume that f is twice continuously differentiable on the interval $[0, \pi]$. Equipped with the norm

$$||f||_{2,\pi} := ||f||_\pi + ||f'||_\pi + ||f''||_\pi,$$

the space $C^-_{2,\pi}$ becomes a complex Banach space. For fixed $\omega \geq 0$, define

$$A_\omega q := q'' + \omega^2 q, \qquad q \in C^-_{2,\pi}.$$

This is a linear continuous operator

$$A_\omega : C^-_{2,\pi} \to C^-_\pi. \tag{12.93}$$

To begin with, let $\omega = 0$. For given $F \in C^-_\pi$, the equation

$$A_0 q = F, \qquad q \in C^-_{2,\pi} \tag{12.94}$$

is equivalent to the boundary-value problem

$$q'' = F, \qquad q(0) = q(\pi) = 0.$$

This problem has a unique solution. Consequently, the operator A_0 from (12.93) is a Fredholm operator of index zero. Since the embedding $C^-_{2,\pi} \subseteq C^-_\pi$ is compact, the operator A_ω from (12.93) represents a compact perturbation of the operator A_0, and hence it is also a Fredholm operator of index zero (see Zeidler (1995), Vol. 2, Sect. 5.8).
We are now given $\omega > 0$ with $\omega \neq k$ for all $k = 1, 2, \ldots$ Then

$$A_\omega q = 0, \qquad q \in C^-_{2,\pi}$$

implies $q = 0$. Since A_ω has index zero, the inverse operator

$$A_\omega^{-1} : C^-_\pi \to C^-_{2,\pi}$$

is linear and continuous. Let $F \in C^-_\pi$ be given. By the Fejér theorem on page 624, there exists a sequence (F_n) of finite Fourier series such that

$$\lim_{n\to\infty} F_n = F \qquad \text{in } C^-_\pi.$$

Set $q_n := A_\omega^{-1} F_n$ and $q := A_\omega^{-1} F$. Then

$$A_\omega q_n = F_n, \qquad q_n \in C^-_{2,\pi}.$$

This means that

$$q_n'' + \omega^2 q_n = F_n, \qquad q_n(0) = q_n(\pi) = 0$$

along with $q'' + \omega^2 q = F$, $q(0) = q(\pi) = 0$. Since q_n is a finite Fourier series, the argument used in the proof of Prop. 11.21 on page 627 shows that

$$q_n(t) = \int_0^\pi G(t, \tau; \omega) F_n(\tau) d\tau. \tag{12.95}$$

Letting $n \to \infty$, the continuity of the operator A_ω^{-1} yields the existence of the limit

$$q = \lim_{n\to\infty} q_n \qquad \text{in } C^-_{2,\pi}.$$

Since the Green's function G is continuous, it follows from (12.95) that

$$q(t) = \lim_{n\to\infty} q_n(t) = \lim_{n\to\infty} \int_0^\pi G(t, \tau; \omega) F_n(\tau) d\tau = \int_0^\pi G(t, \tau; \omega) F(\tau) d\tau.$$

13. Basic Strategies in Quantum Field Theory

> If one does not sometimes think the illogical, one will never discover new ideas in science.
>
> Max Planck (1858–1947)

> Mathematics is not a deductive science – that's a cliché. When you try to prove a theorem, you don't just list the hypotheses, and then start to reason. What you do is trial-and-error, experimentation, and guess work.
>
> Paul Halmos (born 1916)

> "I think, this is so", says Cicha, "in the fight for new insights, the breaking brigades are marching in the front row. The vanguard that does not look to left nor to right, but simply forges ahead – those are the physicists. And behind them there are following the various canteen men, all kinds of stretcher bearers, who clear the dead bodies away, or simply put, get things in order. Well, those are the mathematicians."
>
> From the criminal novel *Death Loves Poetry* of the Czech physicist Jan Klima (born in 1938)[1]

> Whatever the future may bring, it is safe to assert that the theoretical advances made in the unravelling of the constitution of matter since World War II (1939–1945) comprise one of the greatest intellectual achievements of mankind. They were based on the ground secured by Tomonaga, Bethe, Schwinger, Feynman, and Dyson to quantum field theory and renormalization theory in the period from 1946 to 1951.
>
> Silvan Schweber, 1994
> *QED and the Men Who Made it*[2]

The mathematical language of physicists is formal, like a short-hand writing, but this is extremely useful for getting very quickly the desired results that are related to the outcome of physical experiments. It is then the hard task of mathematicians to give rigorous proofs for the heuristic arguments of physicists. The flow of ideas from physics to mathematics is an indispensable source of inspiration for mathematicians.

Warning to the reader. In the following chapters, we will summarize the most important heuristic formulas used in quantum field theory.

[1] Reprinted with permission.
[2] Princeton University Press, 1994 (reprinted with permission).

These heuristic formulas are not to be understood in the sense of rigorous mathematics.

Nevertheless, this approach should help the reader to understand the language of physicists and to find his/her own way in the jungle of literature, which is full of inconsistencies and pseudo-proofs.

In Chaps. 14 and 15, the heuristic formulas are motivated by applying the formal continuum limit to the rigorous finite-dimensional approach from Chap. 7.

The elegance of the mnemonic language of physicists. Mathematicians should note that the language of physicists is optimal from the mnemonic point of view. Therefore, physicists are not willing to give up their language despite the mathematical shortcomings.

In each order of perturbation theory, the heuristic formulas used by physicists can be given a rigorous mathematical meaning.

This will be investigated in the following volumes of this monograph. Unfortunately, as a rule, the rigorous mathematical approach is equipped with technical details which obscure the basic ideas, in contrast to the language of physicists. Therefore, it is important for mathematicians to know the language of physicists as a guide to the rigorous approach. We will use the following two approaches in order to pass from heuristic quantum field theory to rigorous quantum field theory in each order of perturbation theory.

(D) Discretization: The idea is to pass from the heuristic continuum formulas to rigorous formulas by discretization of space, time, and momentum. We then supplement the finite formulas by additional terms called counterterms in order to guarantee the existence of the continuum limit. This second step is called renormalization.
(G) Generalized functions: We avoid discretization by using tempered distributions from the very beginning. This is the so-called Epstein–Glaser approach.

Survey on different approaches. Roughly speaking, there exist the following approaches to quantum field theory:

(i) the response approach based on the generating functional for correlation functions, the magic quantum action reduction formula, and the magic LSZ reduction formula for the S-matrix (Sects. 14.2.4 and 14.2.5);
(ii) the response approach based on the global quantum action principle via functional integrals, which implies (i) (Sect. 14.2.7);
(iii) the operator approach based on Dyson's S-matrix along with creation and annihilation operators for particles (Chap. 15);
(iv) gauge fields, functional integrals, Faddeev–Popov ghosts, and BRST symmetry (Chap. 16);

(v) functional integrals and quantum field theory at finite temperature (Sect. 13.8);
(vi) the rigorous Epstein–Glaser approach via the S-matrix as a tempered operator-valued distribution (Sect. 15.4.4);
(vii) the rigorous global approach of axiomatic and algebraic quantum field theory (Sect. 15.6);
(viii) the Ashtekar program based on the transport of quantum information along loops and holonomy (loop quantum gravity).[3]

In (i) through (vi), we get formal power series expansions which can be regarded as asymptotic series, by the classical Ritt theorem (Sect. 15.5.2). In higher-order perturbation theory, the expressions are rather involved from the analytical point of view. These expressions can be represented graphically by Feynman diagrams. This helps very much to get insight into the structure of perturbation theory and its renormalization. In this context,

- Zimmermann's forest formula and
- Kreimer's Hopf algebra play a crucial role (see Sect. 15.4.6).

Surprisingly enough, as a rule of thumb, the experience of physicists shows that the apparently different approaches (i) through (vi) above yield the same numerical results when applied to the measurements of concrete physical effects. It is the task for the future to understand this equivalence in the framework of a general mathematical theory. Hints for further reading on quantum field theory can be found on page 907ff.

Historical remarks. Originally, Heisenberg and Pauli started quantum field theory in 1929 by representing quantum fields as operator-valued functions

$$\varphi = \varphi(\mathbf{x}, t)$$

depending on the position vector $\mathbf{x}$ and time t. Here, the value $\varphi(\mathbf{x}, t)$ lies in a Hilbert space. However, it turned out that this approach is full of contradictions caused, for example, by not knowing the right commutation relations for interacting quantum fields from the very beginning.

> *Therefore, Feynman and Schwinger moved to a pragmatic point of view and tried to completely avoid the notion of operators in Hilbert space.*

The approaches (i) and (ii) above will be formulated independently of operator theory.

> *The idea is to relate the classical principle of critical action to correlation functions which describe the correlations of the quantum field at different space points and time points.*

[3] We will study this in Volume VI on quantum gravitation. We recommend the survey articles by Ashtekar, Lewandowski (2004) and Giulini, Kiefer, Lämmerzahl (2003). See also the 'Living reviews' quoted on page 929.

In Wightman's axiomatic approach to quantum field theory from 1956, quantum fields are tempered distributions with values in a Hilbert space. In algebraic quantum field theory dating back to Segal in 1947, the fundamental notions are

- observables (elements of an operator algebra), and
- states (positive functionals on the operator algebra).

This notion was generalized by Haag and Kastler in 1967, by passing to so-called nets of local operator algebras. As the standard textbook, we recommend R. Haag, Local Quantum Physics: Fields, Particles, Algebras, Springer, New York, 1996. Roughly speaking, a local operator net assigns an operator algebra to each open subset of the 4-dimensional space-time manifold. This operator algebra represents physical observables. It is important that this universal approach can be extended to many-particle systems in statistical physics and to curved space-time manifolds in order to include quantum gravitation. In this setting, it is possible to mathematically introduce two types of distinguished physical states, namely,

- Kubo–Martin–Schwinger (KMS) states which describe thermodynamic equilibrium states of many-particle systems, and
- Hadamard states which play a fundamental role in curved space–time manifolds.

For example, the theory of KMS states is equivalent to the Tomita–Takesaki theory for von Neumann algebras. It is also possible to formulate Einstein's principle of general relativity (or the covariance principle) in terms of both

- algebraic quantum field theory and
- the Ashtekar program.

In the framework of string theory, quantum field theory is a low-energy approximation of

- vibrating strings and
- higher-dimensional vibrating membranes called D-branes.

This is closely related to the theory of minimal surfaces and Riemannian geometry in mathematics. We will study this in Volume VI.

13.1 The Method of Moments and Correlation Functions

> The family of moments knows all about a given random phenomenon in nature.
>
> Folklore

We expect that a quantum field possesses a random structure depending on space and time with an infinite number of degrees of freedom. The idea of physicists is to describe the quantum field by its local moments which depend on position and time. For the local moments, physicists use the following notions synonymously:

- n-point correlation function,
- n-point Green's function, and
- n-point Feynman propagator.

To be honest, the formal definitions of these notions differ in the literature. But, roughly speaking, the physics behind these notions is the same in each order of perturbation theory. This follows from the nontrivial fact that the computations based on Feynman diagrams and the corresponding Feynman rules are the same. Observe the following. We have to carefully distinguish between

- free quantum fields, and
- full quantum fields.

Here, free quantum fields are free of interactions, whereas full quantum fields describe interactions. Naturally enough, the mathematics of free quantum fields is much simpler than the mathematics of full quantum fields. From the physical point of view, our main interest is to understand full quantum fields. As with quantum fields, we distinguish between

- free correlation functions, and
- full correlation functions.

Correlation functions of free quantum fields. In terms of functional integrals, the free n-point correlation function of a free quantum field φ_{free} is formally defined by

$$C_{n,\text{free}}(x_1, x_2, \ldots, x_n) := \frac{\int \varphi_{\text{free}}(x_1)\varphi_{\text{free}}(x_2)\cdots\varphi_{\text{free}}(x_n)\mathrm{e}^{\mathrm{i}S_{\text{free}}[\varphi_{\text{free}}]/\hbar}\mathcal{D}\varphi}{\int \mathrm{e}^{\mathrm{i}S_{\text{free}}[\varphi_{\text{free}}]/\hbar}\mathcal{D}\varphi}.$$

Here, $S_{\text{free}}[\varphi_{\text{free}}]$ is the classical action of the free quantum field φ_{free}. The symbol

$$x_j := (\mathbf{x}_j, t_j), \qquad j = 1, \ldots, n$$

denotes a space-time point with the position vector $\mathbf{x}_j$ and time t_j, in an inertial system. The free n-point Green's function of φ_{free} is defined by

$$G_{n,\text{free}}(x_1, x_2, \ldots, x_n) := \langle 0|\mathcal{T}\{\varphi_{\text{free}}(x_1)\varphi_{\text{free}}(x_2)\cdots\varphi_{\text{free}}(x_n)\}|0\rangle$$

where the symbol $|0\rangle$ denotes the ground state of the free quantum field φ_{free}. Therefore, the free n-point Green's function describes the vacuum expectation value of time-ordered products of free field operators. It turns out that on a formal level,

$$C_{n,\text{free}} = G_{n,\text{free}}, \qquad n = 1, 2, , \ldots .$$

This relates the functional-integral approach to the operator approach in quantum field theory.

Observe that, by definition, the chronological operator $\mathcal{T}$ orders time. Explicitly, for the quantum field φ, the definition of the operator $\mathcal{T}$ reads as

$$\mathcal{T}\{\varphi(x_1)\varphi(x_2)\} := \varphi(x_1)\varphi(x_2) \qquad \text{if} \quad t_1 \geq t_2.$$

This has to be replaced by

$$\mathcal{T}\{\varphi(x_1)\varphi(x_2)\} := \eta\varphi(x_2)\varphi(x_1) \qquad \text{if} \quad t_2 > t_1.$$

Here, we have to choose the sign $\eta := 1$ for bosonic fields (e.g., photons), and $\eta := -1$ for fermionic fields (e.g., electrons). For example, the free 2-point Green's function (resp. correlation function) reads as

$$G_{2,\text{free}}(\mathbf{x}, t; \mathbf{y}, s) := \langle 0|\mathcal{T}\{\varphi_{\text{free}}(\mathbf{x}, t)\varphi_{\text{free}}(\mathbf{y}, s)\}|0\rangle.$$

This function is also frequently called the Feynman propagator of the free quantum field φ. Similarly, we define the chronological operator for n-factors.

The chronological operator $\mathcal{T}$ takes causality into account.

Correlation functions of full quantum fields. Let us now consider quantum fields under interactions. The full n-point correlation function of the quantum field φ is formally defined by

$$\boxed{C_n(x_1, x_2, \ldots, x_n) := \frac{\int \varphi(x_1)\varphi(x_2)\cdots\varphi(x_n)\mathrm{e}^{\mathrm{i}S[\varphi]/\hbar}\mathcal{D}\varphi}{\int \mathrm{e}^{\mathrm{i}S[\varphi]/\hbar}\mathcal{D}\varphi}}$$

where $S[\varphi]$ denotes the classical action of the quantum field φ. In contrast to the free action $S_{\text{free}}[\varphi_{\text{free}}]$, the action $S[\varphi]$ includes nonlinear terms with respect to φ which describe the interactions. The full n-point Green's function of the quantum field φ is formally defined by

$$\boxed{G_n(x_1, x_2, \ldots, x_n) := \langle 0_{\text{int}}|\mathcal{T}\{\varphi(x_1)\varphi(x_2)\cdots\varphi(x_n)\}|0_{\text{int}}\rangle.}$$

Here, the symbol $|0_{\text{int}}\rangle$ denotes the ground state of the full quantum field which, as a rule, differs from the ground state $|0\rangle$ of the free quantum field. Note that the formulas for C_n and G_n are only mnemonic formulas. On a formal level, we have

$$\boxed{C_n = G_n, \qquad n = 1, 2, , \ldots .}$$

Intuitively, the n-point Green's functions (and hence the correlation functions) describe vacuum fluctuations of the quantum field.

The philosophy is that vacuum fluctuations know all about the quantum field.

Advantages and disadvantages of different approaches. The advantage of the correlation functions $C_1, C_2, \ldots$ is that they only depend on the classical action S, but not on operators in Hilbert space. The disadvantage is that functional integrals are beautiful mnemonic tools, but not well-defined mathematical objects, as a rule.

The advantage of the Green's functions $G_1, G_2, \ldots$ is that no ill-defined functional integrals appear, but operators in Hilbert space. The disadvantage is that the operators $\varphi(x)$ are highly singular objects. Furthermore, the commutation (resp. anticommutation) relations are not known *a priori* for interacting quantum fields, but they have to be determined. In particular, the naive assumption that the commutation (resp. anticommutation) relations for interacting quantum fields are the same as for free quantum fields does not hold.

The idea of local averaging over space and time. It turns out that, as a rule, the correlation functions C_n are not well-defined as classical local functions of the space-time variables. Intuitively, however, averages of the form

$$\int C_n(x_1, \ldots, ; x_n)\chi_1(x_1) \cdots \chi_n(x_n) d^4x_1 \cdots d^4x_n$$

make sense where $\chi_1, \ldots, \chi_n$ are suitable test functions. More precisely, the n-point correlation functions are distributions. Note that distributions are linear mathematical objects and the multiplicative structure of distributions is subtle, according to Laurent Schwartz. Therefore, the case of nonlinear interacting quantum fields has to be handled very carefully. In this connection, the Epstein–Glaser approach works successfully.

The philosophy behind the use of averages is that physical experiments are based on measurement devices which are only able to measure mean values.

The discussion of the method of moments in quantum field theory will be continued in Sect. 13.6 in terms of generating functions for moments, generating functionals, and external physical sources.

13.2 The Power of the *S*-Matrix

In particle accelerators, physicists measure the following quantities:

- cross sections of scattering processes;
- masses of stable particles;
- lifetimes of unstable particles;
- magnetic moments of elementary particles.

The task is to theoretically predict the measured values. To this end, physicists use two concepts, namely,

- the S-matrix and
- n-point correlation functions.

Roughly speaking, the two concepts are equivalent. The S-matrix encodes the transition probabilities of scattering processes. In fact, the S-matrix knows all about elementary particles.

- The S-matrix enables us to compute the cross sections of scattering processes.
- The singularities of the S-matrix refer to the masses of stable particles.
- Analytic continuation of the S-matrix in momentum space tells us the masses of unstable particles (also called resonances) and their lifetimes.
- The S-matrix allows us to compute the full correlation functions by the magic Gell-Mann–Low formula. This way, for example, we get so-called form factors which yield the magnetic moments of elementary particles (e.g., the anomalous magnetic moment of the electron and the muon).

13.3 The Relation Between the *S*-Matrix and the Correlation Functions

We expect that there exists a close connection between the S-matrix and the full correlation functions. This is indeed the case. In the 1950s, physicists developed magic formulas in order to pass from the S-matrix to the correlation functions and vice versa.

The response approach. This will be studied in Chap. 14. The two magic formulas read as follows.

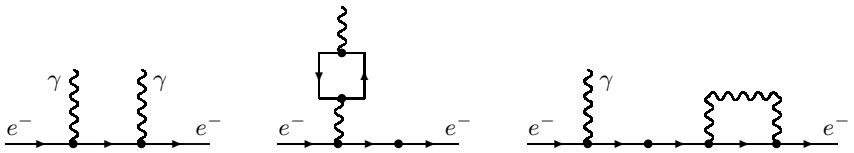
(QA) The quantum action reduction formula (14.38): computation of the full correlation functions by means of the free correlation functions.

(LSZ) The Lehmann–Szymanzik–Zimmermann reduction formula (14.41): computation of the S-matrix by using the full correlation functions obtained by (QA).

The basic formulas (QA) and (LSZ) can be obtained from the global quantum action principle which is based on Feynman's magic functional-integral formula (14.43). This functional integral describes quantum fluctuations by averaging over all possible classical fields. The statistical weight depends on the classical action.

The operator approach. This approach will be studied in Chap. 15; it will be based on the following two magic formulas.

(D) The Dyson formula (15.12): computation of the S-matrix by means of free quantum fields and the interaction terms of the classical action.



(a) S-matrix diagrams

(b) Amputated S-matrix diagrams

Fig. 13.1. Feynman diagrams for Compton scattering of electrons and photons

(GL) The Gell-Mann–Low formula (15.28): computation of the full correlation functions by using the S-matrix.

Roughly speaking, the response approach (QA), (LSZ) and the operator approach (D), (GL) are equivalent. This means that they lead to the same physical results.

Full quantum fields. The question arises of how to compute full quantum fields. In fact, this can be done by using Bogoliubov's formula (15.35) which allows us to compute the full quantum field as a functional derivative of the S-matrix formulated in the language of tempered distributions. In this setting, the S-matrix is a primary object, whereas the quantum field is a derived object.

13.4 Perturbation Theory and Feynman Diagrams

The magic formulas from the preceding section allow us to elegantly compute the S-matrix and the full correlation functions in terms of

- free correlation functions, and
- the wave functions of free particles.

To this end, we use the method of perturbation theory. The higher-order approximations yield rather complex analytic expressions. Fortunately enough, these expressions are structured, and they can be represented graphically by diagrams first introduced by Feynman. In elementary particle physics, physicists use

- Feynman diagrams and Feynman rules.

The Feynman rules yield analytic expressions for the S-matrix and the correlation functions in a very effective way. Figure 13.1 on page 747 shows some Feynman diagrams for the Compton scattering where a photon hits an electron. Observe the following:

- The number of nodes tells us the approximation order of perturbation theory. This causes the increasing complexity of Feynman diagrams when passing to higher-order approximations.
- The Feynman diagrams of the S-matrix contain external lines which correspond to incoming and outgoing particles of the scattering process. The internal lines can be regarded as virtual particles which are related to quantum fluctuations of the ground state.
- The amputated S-matrix diagrams are independent of the incoming and outgoing particles; these diagrams correspond to the magic LSZ formula (14.41) which relates the S-matrix to the correlation functions.

13.5 The Trouble with Interacting Quantum Fields

Let us now discuss a crucial difficulty which arises for full quantum fields. From the mathematical point of view, free quantum fields are well-understood mathematical objects; they are tempered distributions with values in a Hilbert space. In contrast to this, the experience of physicists and mathematicians shows that the notion of a full quantum field causes mathematical trouble. The partial differential equations for full quantum fields can be easily written down, as in the classical case. However, one has to add commutation (resp. anticommutation) relations for the field operators. The following question arises:

> *What are the right commutation (resp. anticommutation) relations for full quantum fields?*

If one uses the same commutation (resp. anticommutation) relations as for free fields, then one gets contradictions. In most physics textbooks, it is assumed that the full quantum field is unitarily equivalent to the free quantum field. This is used along with Dirac's interaction picture in order to get the S-matrix in the framework of perturbation theory. However, this assumption is completely wrong. Unitarily equivalent Hilbert space theories describe the same quantum physics. Intuitively, however, free quantum fields and full quantum fields correspond to different physical situations. In fact, Haag proved in 1955 that under appropriate assumptions, such a unitary equivalence along with Dirac's interaction picture does not exist. This is the famous Haag theorem.[4]

[4] See R. Haag, On quantum field theories, Dan. Mat. Medd. **29** (1955), No. 12, and the textbook by N. Bogoliubov et al., General Principles of Quantum Field Theory, Kluwer, Dordrecht, 1990, p. 388ff.

We have shown in Sect. 7.19.3 that the Dirac interaction picture exists in finite-dimensional Hilbert spaces. The point is that this does not remain true for infinite-dimensional Hilbert spaces. Therefore, we will never use the infinite-dimensional Dirac interaction picture in this monograph. To overcome this difficulty, one can apply the finite-dimensional Dirac interaction picture as an approximation. One then has to study the limit of passing to infinite dimensions. One expects that this limit does not exist for quantum fields, but the limit exists for those quantities that can be measured in physical experiments (e.g., cross sections). We will show later on that this approximation method works well.

Note that the right commutation (resp. anticommutation) relations for full quantum fields are unknown at the very beginning. They have to be determined at the end of the solution process. In the setting of the Epstein–Glaser approach, the basic idea is to determine the S-matrix as an operator-valued tempered distribution $S = S(g)$ by an iterative process, in each order of perturbation theory. The full quantum field $\Phi = \Phi(g)$ is then obtained from the S-matrix by using the Bogoliubov formula (15.35) on page 857 based on functional differentiation. This approach is rigorous on a perturbative level.

13.6 External Sources and the Generating Functional

> Study the response of physical systems under the influence of external sources.
>
> Folklore

In the classical theory of probability, one studies generating functions for the moments. These generating functions know all about the random situation under consideration (see page 58).

The generating functional. In quantum field theory, the classical generating function is replaced by the generating functional for the correlation functions given by

$$Z(J) := 1 + \frac{\mathrm{i}}{\hbar}\int_{\mathbb{R}^4} C_1(x)J(x)d^4x - \frac{1}{2\hbar^2}\int_{\mathbb{R}^8} C_2(x,y)J(x)J(y)d^4xd^4y + \sum_{n=3}^{\infty}\frac{\mathrm{i}^n}{n!\hbar^n}\int_{\mathbb{R}^{4n}} C_n(x_1,\cdots x_n)J(x_1)\cdots J(x_n)d^4x_1\cdots d^4x_n. \quad (13.1)$$

This functional depends on the function J. Following Schwinger, this function is called an external source.[5] It remains to compute the functional $Z(J)$. This will be done on page 753 by means of the magic Feynman formula (13.7). This formula is based on a functional integral which depends on the classical action. Similarly, we introduce the free generating functional,

[5] A general approach can be found in Schwinger (1970).

$$Z_{\text{free}}(J) := 1 + \sum_{n=1}^{\infty} \frac{\mathrm{i}^n}{n!\hbar^n} \int_{\mathbb{R}^{4n}} C_{n,\text{free}}(x_1, \cdots x_n) J(x_1) \cdots J(x_n) d^4x_1 \cdots d^4x_n.$$

Volterra's differential calculus. Generalizing the classical Taylor expansion, functionals of the type $Z(J)$ were investigated in mathematics by Volterra (1860-1950) around 1900. Motivated by classical partial derivatives, Volterra introduced partial functional derivatives in such a way that

$$\frac{\delta Z(J)}{\delta J(x)}|_{J=0} := \frac{\mathrm{i}}{\hbar}\, C_1(x), \tag{13.2}$$

and

$$\frac{\delta^n Z(J)}{\delta J(x_1) \cdots \delta J(x_n)}|_{J=0} := \frac{\mathrm{i}^n}{\hbar^n}\, C_n(x_1, \ldots, x_n), \qquad n = 2, 3, \ldots$$

Similarly,

$$\frac{\delta^n Z_{\text{free}}(J)}{\delta J(x_1) \cdots \delta J(x_n)}|_{J=0} := \frac{\mathrm{i}^n}{\hbar^n}\, C_{n,\text{free}}(x_1, \ldots, x_n), \qquad n = 2, 3, \ldots$$

The definition of partial functional derivatives can be found in Sect. 7.20.2 on page 401. Partial functional derivatives are also called local functional derivatives. The formulas above are well-defined, since we tacitly assume that the correlation functions are symmetric with respect to the variables $x_1, x_2, \ldots$

Cumulants and reduced correlation functions. In order to simplify computations, physicists introduce the so-called reduced generating functional by setting

$$Z_{\text{red}}(J) := \ln Z(J).$$

In particular, $Z(0) = 1$ implies $Z_{\text{red}}(0) = 0$. Conversely,

$$Z(J) = \mathrm{e}^{Z_{\text{red}}(J)}.$$

The so-called reduced (or connected) correlation functions, $C_{n,\text{red}}$ are given by the following expansion

$$Z_{\text{red}}(J) := \sum_{n=1}^{\infty} \frac{\mathrm{i}^n}{n!\hbar^n} \int_{\mathbb{R}^{4n}} C_{n,\text{red}}(x_1, \cdots x_n) J(x_1) \cdots J(x_n) d^4x_1 \cdots d^4x_n.$$

Passing from $Z(J)$ to $Z_{\text{red}}(J)$ eliminates redundant quantities and simplifies computations. In terms of Feynman graphs, connected correlation functions correspond to connected Feynman graphs. Thus, the passage to connected correlation functions avoids the use of redundant disconnected Feynman diagrams.

The method of connected correlation functions in physics corresponds to the method of cumulants in mathematics. Classical moments can be expressed by cumulants (or semi-invariants) and vice versa. This simplifies the

computation of moments. See Shiryaev (1996), Sect. II.12. We also recommend the classic monograph on the problem of moments by Shohat and Tamarkin (1950). The problem of moments concerns the reconstruction of a mass (or probability) distribution from its moments. This classical problem can be solved by using the Hahn–Banach theorem from functional analysis. This can be found in both Zeidler (1995), Vol. 2, Sect. 1.4 and Lax (2002).

The mean field approximation. The local functional derivative

$$\varphi_{\text{mean}}(x) := \frac{\hbar}{\mathrm{i}} \frac{\delta Z_{\text{red}}(J)}{\delta J(x)}|_{J=0}$$

is called the mean field of the quantum field. The experience of physicists shows that mean fields serve as a semi-classical approximation for describing quantum fields.

Effective quantum action. The expression

$$V(\varphi_{\text{mean}}) := \frac{\hbar}{\mathrm{i}} Z_{\text{red}}(J) - \int_{\mathbb{R}^4} \varphi_{\text{mean}}(x) J(x) d^4x$$

is called the effective quantum action. The expansion

$$\begin{aligned} V(\varphi_{\text{mean}}) &= 1 + \int_{\mathbb{R}^4} V_1(x) \varphi_{\text{mean}}(x) d^4x \\ &+ \sum_{n=2}^{\infty} \frac{1}{n!} \int_{\mathbb{R}^{4n}} V_n(x_1, \cdots x_n) \varphi_{\text{mean}}(x_1) \cdots \varphi_{\text{mean}}(x_n) d^4x_1 \cdots d^4x_n \end{aligned}$$

defines the so-called irreducible n-point vertex function V_n. It turns out that vertex functions correspond to Feynman diagrams which have a minimal redundance. We will study this in later volumes. We also refer to Greiner and Reinhardt (1996b), Sect. 12.7.

Correlation functions represent a basic tool for studying physical processes in both elementary particle physics and solid state physics. There exists a wealth of physical phenomena which have to be explained. Physicists use both

- the mean field φ_{mean} and
- the effective quantum action $V(\varphi_{\text{mean}})$

as approximations for describing quantum corrections.

13.7 The Beauty of Functional Integrals

> The action knows all about a quantum system via functional integrals.
> Folklore

In modern elementary particle physics, most physicists prefer the use of functional integrals because of their mnemonic elegance. Functional integrals allow us the economical formulation of basic principles in quantum field theory.

13.7.1 The Principle of Critical Action

It is quite interesting that many physical phenomena in nature are governed by the principle of critical action. As the prototype, consider a real field $\varphi : \mathbb{R}^4 \to \mathbb{R}$. We define the action S by setting

$$S[\varphi] := \int_\Omega \mathcal{L}(\varphi(x), \varphi_{\mathbf{x}}(x), \dot{\varphi}(x))\, d^4x.$$

Here, $x = (\mathbf{x}, t)$. The derivative $\varphi_{\mathbf{x}}$ stands for the tupel $(\varphi_{x^1}, \varphi_{x^2}, \varphi_{x^3})$ of partial derivatives, and $\dot{\varphi}$ denotes the partial time derivative φ_t. Moreover, the domain of integration is given by the product set

$$\Omega := \mathcal{C} \times [t_0, t_1]$$

where $\mathcal{C}$ is a nonempty bounded open subset of $\mathbb{R}^3$ (e.g., a cube), and $[t_0, t_1]$ is a finite time interval. The function $\mathcal{L}$ is called the Lagrangian density. The principle of critical action for the field φ reads as

$$\boxed{S[\varphi] = \text{critical!}} \tag{13.3}$$

We have to add the boundary condition $\varphi =$ fixed on $\partial\Omega$. Each smooth solution φ of the variational problem (13.3) satisfies the following Euler–Lagrange equation

$$\frac{\partial}{\partial t}\frac{\partial \mathcal{L}}{\partial \dot{\varphi}} + \sum_{j=1}^{3} \frac{\partial}{\partial x^j}\frac{\partial \mathcal{L}}{\partial \varphi_{x^j}} = \frac{\partial \mathcal{L}}{\partial \varphi}. \tag{13.4}$$

The proof will be given in Problem 14.2 on page 805. This equation generalizes Newton's classical equation of motion in mechanics: time derivative of momentum equals force. For example, if we choose the Lagrangian density

$$\mathcal{L}(\varphi, \varphi_{\mathbf{x}}, \dot{\varphi}) := \frac{\dot{\varphi}^2}{2c^2} - \frac{1}{2}\sum_{j=1}^{3} \varphi_{x_j}^2 - \frac{m_0^2 c^2 \varphi^2}{2\hbar^2} - \frac{\kappa\varphi^4}{4}, \tag{13.5}$$

then the Euler–Lagrange equation reads as

$$\frac{\varphi_{tt}}{c^2} - \sum_{j=1}^{3} \varphi_{x^j x^j} + \frac{m_0^2 c^2}{\hbar^2}\varphi + \kappa\varphi^3 = 0. \tag{13.6}$$

This is the so-called φ^4-model (or the nonlinear Klein–Gordon equation). The nonnegative real number κ is called the coupling constant. If $\kappa = 0$, then we get the linear Klein–Gordon equation. The nonlinear Klein-Gordon equation is a model for an uncharged spinless particle of mass m_0 (e.g., a meson). The κ-term models the self-interaction of the particle. Introducing

the wave operator $\Box := \frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \Delta$, the nonlinear Klein–Gordon equation can be written as

$$\boxed{\Box\varphi + \frac{m_0^2 c^2}{\hbar^2}\,\varphi + \kappa\varphi^3 = 0.}$$

This equation can be obtained in the following way. The energy E of a relativistic particle with rest mass m_0 and momentum vector $\mathbf{p}$ is given by the Einstein relation

$$E^2 = m_0^2 c^4 + c^2\mathbf{p}^2.$$

Motivated by Schrödinger quantization, let us use the replacement

$$E \Rightarrow \mathrm{i}\hbar\frac{\partial}{\partial t}, \qquad \mathbf{p} \Rightarrow -\mathrm{i}\hbar\boldsymbol{\partial}.$$

Hence $\Box\varphi + \frac{m_0^2 c^2}{\hbar^2}\varphi = 0$. This is the linear Klein–Gordon equation.

13.7.2 The Magic Feynman Representation Formula

The fundamental generating functional $Z(J)$ for the full correlation functions is given by the magic Feynman formula

$$\boxed{Z(J) = \mathcal{N}\int_{\mathcal{F}} \mathrm{e}^{\mathrm{i}S[\varphi]/\hbar}\,\mathrm{e}^{\mathrm{i}J[\varphi]/\hbar}\,\mathcal{D}\varphi.} \tag{13.7}$$

This is also called the global quantum action principle. Here, we set

$$J[\varphi] := \int_\Omega J(x)\varphi(x)d^4x.$$

The normalization factor $\mathcal{N}$ is uniquely determined by the normalization condition $Z(0) = 1$. Concerning the functional integral from (13.7), we integrate over all possible classical fields φ which satisfy the same boundary condition as in the principle of critical action (13.3). These classical fields φ form the space $\mathcal{F}$. Observe that the symbol $\mathcal{D}\varphi$ stands for a formal measure on the function space $\mathcal{F}$. Formula (13.7) is one of the most beautiful and most useful formulas in theoretical physics. It relates the classical action of a physical field φ to the full generating functional $Z(J)$ which describes the quantum fluctuations of the corresponding quantized field. Therefore, we call formula (13.7) the global quantum action principle (or the integral quantum action principle).[6]

[6] In mathematics, measure integrals are denoted by

$$Z(J) = \mathcal{N}\int_{\mathcal{F}} \mathrm{e}^{\mathrm{i}S[\varphi]/\hbar}\,\mathrm{e}^{\mathrm{i}J[\varphi]/\hbar}\,d\mu[\varphi].$$

13.7.3 Perturbation Theory

Let us assume that the Lagrangian density has the form

$$\boxed{\mathcal{L} = \mathcal{L}_{\text{free}} + \kappa \mathcal{L}_{\text{int}}}$$

where κ is a sufficiently small nonnegative number called the coupling constant. Here, $\mathcal{L}_{\text{free}}$ and $\kappa\mathcal{L}_{\text{int}}$ describe the free field and the interaction, respectively. Assume that the function $\mathcal{L}_{\text{int}}$ only depends on the field function φ, but not on its partial derivatives. In the special case (13.5), the interaction term $\kappa\mathcal{L}_{\text{int}}(\varphi)$ is equal to $-\kappa\varphi^4/4$. Setting

$$S_{\text{free}}[\varphi] := \int_\Omega \mathcal{L}_{\text{free}}(\varphi(x), \varphi_{\mathbf{x}}(x), \dot{\varphi}(x))\, d^4x$$

and $S_{\text{int}} := \int_\Omega \mathcal{L}_{\text{int}}(\varphi(x))\, d^4x$, we get

$$\boxed{S[\varphi] := S_{\text{free}}[\varphi] + \kappa S_{\text{int}}[\varphi].}$$

By (13.7), the free generating functional reads as

$$Z_{\text{free}}(J) = \mathcal{N}_{\text{free}} \int_{\mathcal{F}} \mathrm{e}^{\mathrm{i}S_{\text{free}}[\varphi]/\hbar}\, \mathrm{e}^{\mathrm{i}J[\varphi]/\hbar}\, \mathcal{D}\varphi.$$

The normalization factor $\mathcal{N}_{\text{free}}$ is uniquely determined by the normalization condition $Z_{\text{free}}(0) = 1$. The key formula reads as

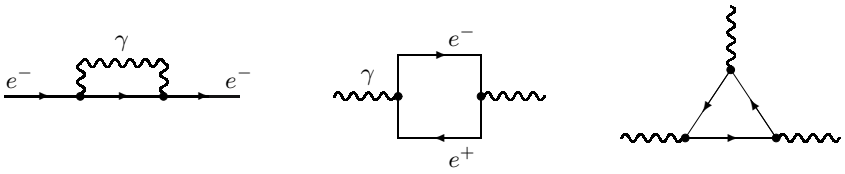
$$\boxed{Z(J) = \mathcal{N} \exp\Big\{-\frac{\mathrm{i}}{\hbar} \int_\Omega d^4x \cdot \kappa\mathcal{L}_{\text{int}} \left(\frac{\hbar}{\mathrm{i}} \frac{\delta}{\delta J(x)}\right)\Big\} Z_{\text{free}}(J).} \tag{13.8}$$

This formula shows us how to reduce the full generating functional to the free generating functional. This means that the full correlation functions can be computed by means of the free correlation functions. The normalization factor $\mathcal{N}$ is uniquely determined by the normalization condition $Z(0) = 1$. Observe that formula (13.8) is independent of the use of functional integrals.

13.7.4 Renormalization

As we will show in Volume II on quantum electrodynamics, there are approximations obtained by perturbation theory which are in good agreement

To simplify notation, physicists use the symbol $\mathcal{D}\varphi$ instead of $d\mu[\varphi]$. Mnemonically, the symbol $\mathcal{D}\varphi$ resembles the Leibniz notation, $\int f(x)dx$, for classical integrals. It was shown by Cameron (1960) that the Feynman functional integral does not exist as a measure integral, in the sense of mathematics. For the theory of functional integrals, we recommend Roepstorff (1996), Grosche and Steiner (1998) (handbook), Johnson and Lapidus (2000), Klauder (2000), Chaichian and Demichev (2001), Vols. 1, 2.



(a) self-interaction of an electron (b) self-interaction of a photon (c) vertex graph

Fig. 13.2. Critical Feynman diagrams in quantum electrodynamics

with physical experiments. For example, this concerns Compton scattering (see the Feynman diagram in Fig. 13.1(a) on page 747.) However, there are also processes which lead to divergent expressions in low-order perturbation theory. Typical processes of this kind in quantum electrodynamics are pictured in Fig. 13.2. In order to extract values from the divergent expression which coincide with those values obtained in physical experiments, one has to use the methods of renormalization theory (see Sect. 15.4).

13.7.5 Transition Amplitudes

Fix the finite time interval $[s,t]$. We are given the classical field $\varphi_0 = \varphi_0(\mathbf{x})$ at the initial time s and the classical field $\varphi_1 = \varphi_1(\mathbf{x})$ at the final time t. We assume that there exists a Hamiltonian operator (energy operator) H to the quantized version of the classical system with the action S. We are interested in the transition probability

$$|\langle\varphi_1|\mathrm{e}^{-\mathrm{i}H(t-s)/\hbar}|\varphi_0\rangle|^2$$

from the quantum state φ_0 at time s to the quantum state φ_1 at time t. Feynman's main hypothesis tells us that we have the key formula

$$\boxed{\langle\varphi_1|\mathrm{e}^{-\mathrm{i}H(t-s)/\hbar}|\varphi_0\rangle = \int_{\mathcal{F}} \mathrm{e}^{\mathrm{i}S[\varphi]/\hbar}\,\mathcal{D}\varphi.} \tag{13.9}$$

Here, the functional integral sums over all classical fields $\varphi = \varphi(\mathbf{x},\tau)$ which satisfy the boundary conditions

$$\varphi(\mathbf{x},s) = \varphi_0(\mathbf{x}), \qquad \varphi(\mathbf{x},t) = \varphi_1(\mathbf{x}) \qquad \text{for all} \quad \mathbf{x} \in \mathbb{R}^3$$

at time s and t, respectively. The inner product $\langle\varphi_1|\mathrm{e}^{-\mathrm{i}(t-s)H/\hbar}\varphi_0\rangle$ is called the transition amplitude. It turns out that transition amplitudes play a crucial role in quantum field theory. The importance of transition amplitudes for the computation of quantum processes was emphasized by Feynman in the 1940s.

13.7.6 The Magic Trace Formula

In statistical physics, the trace of operators plays a crucial role. Let us discuss the key formulas.

Propagator. For the trace of the operator $\mathrm{e}^{-\mathrm{i}(t-s)H/\hbar}$, we will use the formula

$$\boxed{\operatorname{tr} \mathrm{e}^{-\mathrm{i}H(t-s)/\hbar} = \int_{\mathcal{F}_{\text{periodic}}} \mathrm{e}^{\mathrm{i}S[\varphi]/\hbar}\, \mathcal{D}\varphi.} \tag{13.10}$$

Here, the functional integral sums over all classical fields φ which satisfy the following periodicity condition:

$$\varphi(\mathbf{x}, s) = \varphi(\mathbf{x}, t) \qquad \text{for all} \quad \mathbf{x} \in \mathbb{R}^3.$$

Formula (13.10) is motivated by applying a formal limit to the rigorous trace formula (7.87).

Statistical physics. Replacing $\mathrm{i}/\hbar$ by $1/\hbar$, we get

$$\boxed{\operatorname{tr} \mathrm{e}^{-H/kT} = \int_{\mathcal{F}_{\text{periodic}}} \mathrm{e}^{S[\varphi]/\hbar}\, \mathcal{D}\varphi.} \tag{13.11}$$

Here, the temperature T is equal to $\hbar/k(t-s)$ where k denotes the Boltzmann constant. In other words, the passage from the propagator to the partition function corresponds to the replacement

$$\frac{\mathrm{i}(t-s)}{\hbar} \Rightarrow \frac{1}{kT} \tag{13.12}$$

which sends imaginary time to inverse temperature. Planck's quantum of action, $\hbar$, and the Boltzmann constant, k, guarantee that the quantities have the correct physical dimensions.

> *The fundamental transformation* (13.12) *is responsible for the close connection between quantum field theory and statistical physics.*

Summarizing, the three key formulas (13.9), (13.10), and (13.11) above relate the action S of the classical field φ to the Hamiltonian operator H of the corresponding quantized field. Observe that the key formulas do not give us the Hamiltonian operator H itself, but only the transition probabilities as an averaging over classical fields. However, this is very useful for computing physical effects. A special model will be considered in Sect. 13.8.3.

The reader should observe that in quantum field theory, one has to distinguish between

(i) processes which are independent of temperature (e.g., scattering processes in particle accelerators) and
(ii) processes which critically depend on the temperature (e.g., processes in stars or in the early universe).

In case (ii), we speak of quantum fields at finite temperature.

13.8 Quantum Field Theory at Finite Temperature

> In mathematics and physics, many-particle systems are described by generating functions. Physicists call them partition functions.
>
> Folklore

> What happens when ordinary matter is so greatly compressed that the electrons form a relativistic degenerate gas, as in a white dwarf? What happens when the matter is compressed even further so that atomic nuclei overlap to form superdense nuclear matter, as in a neutron star? What happens when nuclear matter is heated to such great temperatures that the nucleons and pions melt into quarks and gluons, as in high-energy nucleus-nucleus collisions? What happened to the spontaneous symmetry breaking of the unified theory of the weak and electromagnetic interactions during the big bang? Questions such as these have been fascinated me for the past ten years. One reason is that a study of such systems involves statistical physics, elementary particle physics, nuclear physics, astrophysics, and cosmology, all of which I find interesting.
>
> Joseph Kapusta, 1993
> *Finite-Temperature Field Theory*[7]

Quantum statistics concerns many-particle systems (e.g., gases, systems of elementary particles, condensed matter). We want to show that the theory of such systems is governed by a single function, namely, the partition function Z. Both

- the generating functional $Z(J)$ (see Sect. 13.7.2) and
- the partition function Z (see Sect. 13.8.3)

can be represented by functional integrals. This underlines the close relation between quantum field theory and statistical physics. In order to emphasize this relationship, physicists use the same symbol Z.[8] Therefore, generating functionals $Z(J)$ are also called partition functionals. Interesting physical applications will be considered in the later volumes.

We have seen in Chap. 6 that the notion of partition function also plays a fundamental role in mathematics, for example, in number theory. Edward Witten uses partition functions in an ingenious manner in order to get deep insight into mathematics by using models motivated by physics.

13.8.1 The Partition Function

The general scheme of quantum statistics. Many-particle quantum systems are described by two operators in a Hilbert space, namely,

- the energy operator H (also called Hamiltonian), and
- the particle number operator N.

[7] Cambridge University Press, 1993 (reprinted with permission).

[8] Mnemonically, the symbol 'Z' refers to the German word 'Zustandssumme' which means 'partition function'.

In addition, we use the following three real parameters:

- the absolute temperature T,
- the chemical potential μ, and
- the size (volume) V of the system.

More generally, we want to describe s species of particles (e.g., electrons, photons, etc.) Therefore, we use s particle number operators

$$N_1, N_2, \ldots, N_s$$

along with the chemical potentials $\mu_1, \ldots, \mu_s$. Here, N_j refers to the jth species of particles. We start with the statistical operator

$$\boxed{\hat{\varrho} := \mathrm{e}^{\beta(\sum_{j=1}^{s} \mu_j N_j - H)}.} \tag{13.13}$$

Here, $\beta := 1/kT$, where k denotes the Boltzmann constant. This yields both the partition sum

$$\boxed{Z := \operatorname{tr} \hat{\varrho}} \tag{13.14}$$

and von Neumann's density operator $\varrho := \frac{\hat{\varrho}}{Z}$.

Mean values of observables. The density operator ϱ is related to the measurements of physical observables. Let A and B be observables (i.e., self-adjoint operators on the Hilbert space X). Then the mean value of the observable A measured in an experiment is given by

$$\boxed{\overline{A} := \operatorname{tr}(\varrho A).}$$

This allows us to introduce the nonnegative mean fluctuation ΔA of the observable A by setting

$$(\Delta A)^2 = \overline{(A - \overline{A})^2}.$$

In addition, we introduce the correlation coefficient

$$\gamma := \frac{\overline{(A - \overline{A})(B - \overline{B})}}{\Delta A \cdot \Delta B}$$

of the two observables A and B. For a many-particle system related to the operator-valued function $\varphi = \varphi(\mathbf{x}, t)$ of a quantum field, we set

$$C_n(x_1, \ldots, x_n) := \operatorname{tr}(\hat{\varrho} \mathcal{T} \{\varphi(x_1)\varphi_2(x) \cdots \varphi(x_n)\})$$

with $x_j := (\mathbf{x}_j, t_j)$. The chronological operator $\mathcal{T}$ refers to time $t_1, \ldots, t_n$. This is called the n-point correlation function. Note that physicists use the terms n-point correlation function, n-point Green's function and n-point Feynman propagator synonymously. We define

- the mean energy $\overline{E} := \operatorname{tr}(\varrho H)$,
- the mean particle number, $\overline{N}_j$ of the jth species,
- the entropy $S = -k\,\overline{\ln \varrho} = -k \operatorname{tr}(\varrho \ln \varrho)$, and
- the free energy $F := \overline{E} - TS$.

Physical quantities. We want to show that all of the interesting physical quantities of a many-particle system can be computed if we know its partition function Z. To this end, we introduce the statistical potential

$$\boxed{\Gamma(T, \mu_1, \ldots, \mu_s, V) := -kT \ln Z(T, \mu_1, \ldots, \mu_s, V).}$$

The partition function Z, and hence Γ, depends on the variables T (temperature), $\mu_1, \ldots, \mu_s$ (chemical potentials) and V (volume). The corresponding partial derivatives yield the following physical quantities:

- entropy: $S = -\Gamma_T$;
- mean particle numbers: $\overline{N}_j = -\Gamma_{\mu_j}$, $j = 1, \ldots, s$;
- pressure $P = -\Gamma_V$;
- free energy $F = \Gamma + \sum_{j=1}^{s} \mu_j \overline{N}_j$;
- mean (inner) energy $\overline{E} = F + TS$;
- free enthalpy $G = F + PV$;
- enthalpy: $\overline{E} + PV$.

In the 19th century, these quantities were introduced first in phenomenological thermodynamics. The relations between these quantities and their physical interpretation are thoroughly discussed in Zeidler (1986), Vol. IV, p. 387.

Finite-dimensional prototype. Let $|E_1\rangle, \ldots, |E_m\rangle$ be an orthonormal basis of the complex Hilbert space X. We set

- energy operator: $H := \sum_{k=1}^{m} E_k |E_k\rangle$;
- particle number operator $N_j := \sum_{k=1}^{m} n_{jk} |E_k\rangle$ of the jth species where $j = 1, \ldots, s$;
- density operator: $\varrho := \sum_{k=1}^{m} p_k |E_k\rangle$.

For the statistical operator, we get

$$\hat{\varrho} = \mathrm{e}^{\beta(\sum_{j=1}^{s} \mu_j N_j - H)} = \sum_{k=1}^{m} \mathrm{e}^{\beta(\sum_{j=1}^{s} \mu_j n_{jk} - E_k)} |E_k\rangle\langle E_k|.$$

This yields the partition sum

$$Z = \operatorname{tr} \hat{\varrho} = \sum_{k=1}^{m} \mathrm{e}^{\beta(\sum_{j=1}^{s} \mu_j n_{jk} - E_k)}$$

and the density operator $\varrho = \frac{\hat{\varrho}}{Z}$. Hence

$$p_k := \frac{\mathrm{e}^{\beta(\sum_{j=1}^{s} \mu_j n_{jk} - E_k)}}{Z}, \qquad k = 1, \ldots, m.$$

For the mean value of energy, we have

$$\overline{E} = \sum_{k=1}^{m} E_k p_k.$$

Similarly, the mean particle number of the jth species is given by

$$\overline{N}_j = \sum_{k=1}^{m} n_{jk} p_k, \qquad j = 1, \ldots, s.$$

For the state $|E_k\rangle$,

$$H|E_k\rangle = E_k|E_k\rangle, \qquad N_j|E_k\rangle = n_{jk}|E_k\rangle.$$

Here, $k = 1, \ldots, m$ and $j = 1, \ldots, s$. Intuitively, the state $|E_k\rangle$ describes one particle of energy E_k. In this state, there are n_{jk} particles of the jth species. The probability for realizing this situation is equal to p_k. This will be studied in greater detail along with physical applications in Volume IV on quantum mathematics.

13.8.2 The Classical Hamiltonian Approach

As preparation for the magic Feynman formula in Sect. 13.8.3, let us sketch the passage from the Lagrangian approach to the Hamiltonian approach.[9]

Special case. Choose the Lagrangian density $\mathcal{L}$ from the nonlinear Klein–Gordon model (13.5). Explicitly,

$$\mathcal{L}(\varphi, \varphi_{\mathbf{x}}, \dot{\varphi}) := \frac{\dot{\varphi}^2}{2c^2} - \frac{\varphi_{\mathbf{x}}^2}{2} - \frac{m_0^2 c^2 \varphi^2}{2\hbar^2} - \frac{\kappa \varphi^4}{4}.$$

We proceed in the following steps:

(I) Legendre transformation: Define the generalized momentum density

$$\pi := \frac{\partial \mathcal{L}}{\partial \dot{\varphi}}.$$

Explicitly, $\pi = \dot{\varphi}/c^2$. This allows us to eliminate the time derivative $\dot{\varphi}$.

(II) The Hamiltonian density $\mathcal{H}$: Define

$$\mathcal{H} := \pi \dot{\varphi} - \mathcal{L}(\varphi, \varphi_{\mathbf{x}}, \dot{\varphi})$$

[9] We will show in Volume II that symplectic geometry plays a crucial role in the Hamiltonian approach to physics.

where we eliminate $\dot{\varphi}$ with the help of π. Hence $\mathcal{H}$ only depends on $\varphi, \varphi_{\mathbf{x}}$, and π. Explicitly,

$$\mathcal{H}(\varphi, \pi) = \frac{c^2\pi^2}{2} + \frac{\varphi_{\mathbf{x}}^2}{2} + \frac{m_0^2 c^2 \varphi^2}{2\hbar^2} + \frac{\kappa\varphi^4}{4}.$$

We set

$$H(t; \varphi, \pi) := \int_{\mathcal{C}} \mathcal{H}(\varphi(\mathbf{x}, t), \varphi_{\mathbf{x}}(\mathbf{x}, t), \pi(\mathbf{x}, t))\, d^3\mathbf{x}.$$

This is the total energy of the field in the 3-dimensional open bounded set $\mathcal{C}$ at time t. Thus, $\mathcal{H}$ represents the energy density of the field. The energy density is positive, as expected.

(III) The Hamiltonian equations of motion reads as

$$\boxed{\dot{\pi}(\mathbf{x}, t) = -\frac{\delta\mathcal{H}(t; \varphi, \pi)}{\delta\varphi(\mathbf{x})}, \qquad \dot{\varphi}(\mathbf{x}, t) = \frac{\delta\mathcal{H}(t; \varphi, \pi)}{\delta\pi(\mathbf{x})}.}$$

Here, we use the local functional derivative at the point $\mathbf{x}$,

$$\frac{\delta\mathcal{H}(t; \varphi, \pi)}{\delta\varphi(\mathbf{x})} := \frac{\partial\mathcal{H}(P)}{\partial\varphi} - \frac{\partial}{\partial\mathbf{x}} \frac{\partial\mathcal{H}(P)}{\partial\varphi_{\mathbf{x}}}, \tag{13.15}$$

along with $P := (\varphi(x), \varphi_{\mathbf{x}}(x), \pi(x))$ and $x := (\mathbf{x}, t)$. Replacing φ by π, we get the corresponding local functional derivative with respect to π. Explicitly,

$$\dot{\pi} = -\frac{m_0^2 c^2}{\hbar^2} - \kappa\varphi^3 + \varphi_{\mathbf{xx}}, \qquad \dot{\varphi} = c^2\pi.$$

This is equivalent to the nonlinear Klein–Gordon equation (13.6).

(iv) The local functional derivative of the Hamiltonian density $\mathcal{H}$ is obtained as follows. Consider an arbitrary smooth test function $h = h(\mathbf{x})$ which has compact support on the 3-dimensional bounded open set $\mathcal{C}$. Differentiate the functional $H(t; \varphi + \varepsilon h, \pi)$ with respect to the real parameter ε at the point $\varepsilon = 0$. Using integration by parts, we get

$$\frac{\delta H(t; \varphi, \pi)}{\delta\varphi}(h) = \int_{\mathcal{C}} \frac{\delta\mathcal{H}(t; \varphi, \pi)}{\delta\varphi(\mathbf{x})} \cdot h(\mathbf{x})\, d^3\mathbf{x}$$

where $\frac{\delta\mathcal{H}(t;\varphi,\pi)}{\delta\varphi(\mathbf{x})}$ coincides with (13.15).

13.8.3 The Magic Feynman Functional Integral for the Partition Function

The partition function is given by the following magic formula

$$\boxed{Z(T, \mu) = \int_{\mathcal{P}_{\text{periodic}}} \mathrm{e}^{\mathcal{S}[\varphi, \pi; T, \mu]}\, \mathcal{D}\varphi\mathcal{D}\pi.} \tag{13.16}$$

Here, T is the absolute temperature, μ is the chemical potential, and k is the Boltzmann constant. Furthermore, we set

$$\mathcal{S}[\varphi,\pi;T,\mu] := \int_0^{1/kT} d\tau \left(\int_{\mathcal{C}} \pi \cdot \frac{\mathrm{i}}{\hbar} \frac{\partial \varphi}{\partial \tau} - \mathcal{H}(\varphi, \varphi_{\mathbf{x}}, \pi) + \mu \mathcal{N}(\varphi,\pi)\, d^3x \right).$$

Here, the arguments of the functions φ and π are $(\mathbf{x},\tau)$. The function $\mathcal{N}$ is the particle number density, that is,

$$N[\varphi,\pi] := \int_{\mathcal{C}} \mathcal{N}(\varphi(x),\pi(x)) d^3\mathbf{x}$$

is the number of particles in the 3-dimensional set $\mathcal{C}$. The functional integral from (13.16) refers to the function space $\mathcal{P}_{\text{periodic}}$ which consists of all the functions

$$\varphi, \pi : \mathcal{C} \times \mathbb{R} \to \mathbb{R}$$

of the argument $(\mathbf{x},\tau)$. In addition, we assume that φ is periodic with respect to the variable τ, that is,

$$\varphi(\mathbf{x},0) = \varphi\left(\mathbf{x}, \frac{1}{kT}\right) \qquad \text{for all} \quad \mathbf{x} \in \mathcal{C}.$$

Let us motivate the magic Feynman formula (13.16).

(a) The action functional for the classical field: In terms of classical field theory, the action functional reads as follows:

$$S[\varphi,\pi] = \int_{t_0}^{t_1} dt \left(\int_{\mathcal{C}} \pi \frac{\partial \varphi}{\partial t} - \mathcal{H}(\varphi,\pi)\, d^3x. \right)$$

To simplify notation, we set $t_0 := 0$. This yields the magic Feynman functional integral

$$Z = \int_{\mathcal{P}} \mathrm{e}^{\mathrm{i}S[\varphi,\pi]/\hbar}\, \mathcal{D}\varphi \mathcal{D}\pi \tag{13.17}$$

where we sum over all functions φ and π which satisfy appropriate side conditions to be specified in (d) below.

(b) The passage from time t to temperature T: Motivated by (13.12) on page 756, we use the replacement

$$\frac{\mathrm{i}t}{\hbar} \Rightarrow \frac{1}{kT}.$$

To this end, we set $t := -\mathrm{i}\hbar\tau$ with $\tau := 1/kT$. This yields

$$S[\varphi,\pi] = -\mathrm{i}\hbar \int_0^{1/kT} d\tau \left(\int_{\mathcal{C}} \pi \cdot \frac{\mathrm{i}}{\hbar} \frac{\partial \varphi}{\partial \tau} - \mathcal{H}(\varphi,\pi)\, d^3x \right).$$

(c) The passage from fixed particle number to variable particle number: Motivated by the basic formula (13.13) of statistical physics on page 758, we use the replacement

$$\text{energy} \Rightarrow \text{energy} - \mu \cdot \text{particle number}.$$

For the density functions, this means $\mathcal{H} \Rightarrow \mathcal{H} - \mu\mathcal{N}$.

(d) The trace formula and the periodicity condition for the field φ : According to (13.14) on page 758, the partition function corresponds to some trace. Motivated by the magic trace formula (13.10) on page 756, we use periodic fields φ. Summarizing this, the desired magic formula (13.16) is obtained from (13.17).

Applications of the magic partition function formula (13.16) will be considered in later volumes. We also refer to the textbooks by Feynman (1998) (14th edition) and Kapusta (1989).

13.8.4 The Thermodynamic Limit

In order to investigate phase transitions, one has to study the so-called thermodynamic limit where the volume of the system goes to infinity. Typical thermodynamic quantities of the system then become singular, which indicates the appearance of a phase transition.

14. The Response Approach

> All dynamical information about a quantum system may be extracted by studying the response of the ground state (vacuum state) of the field to an arbitrary external source J.
>
> Julian Schwinger, 1970[1]

The two magic formulas in quantum field theory. It is fascinating that the huge field of quantum field theory can be based on two magic formulas, namely,

(QA) the quantum action reduction formula for full correlation functions,
(LSZ) and the Lehmann–Szymanzik–Zimmermann reduction formula for the scattering matrix (also called S-matrix).

Since the 1950s, physicists have discovered different ways of formulating these two magic formulas.

> *In this chapter, our basic strategy is to generalize the rigorous finite-dimensional results from Sect. 7.24 to infinite dimensions by carrying out a formal limit.*

In Sect. 7.24 on page 438, we started with discrete functional integrals in order to derive the two magic formulas (QA) and (LSZ) by means of the principle of stationary phase.

> *In this chapter, we will start with the two magic formulas* (QA) *and* (LSZ) *as basic principles.*

The situation is similar to Newton's equation of motion in mechanics and to Maxwell's equations in electrodynamics. It is possible to motivate these equations by using physical and formal mathematical arguments; but one can also postulate the validity of these equations as the starting point of the theory. The prototypes of (QA) and (LSZ) can be found in the following two basic papers:

> M. Gell-Mann and F. Low, Bound states in quantum field theory. Phys. Rev. **84** (1951), 350–354.

[1] The importance of the source approach to quantum field theory is emphasized in J. Schwinger, Sources, Particles, and Fields, Vols. 1–3, Addison–Wesley, Reading, Massachusetts, 1970.

H. Lehmann, K. Szymanzik, and W. Zimmermann, The formulation of quantized field theories, Nuovo Cimento **1** (1955), 205–225 and **6** (1957), 319–333.

In the first paper, Gell-Mann and Low reduced the computation of the full correlation functions to the free correlation functions. The paper by Lehmann, Szymanzik, and Zimmermann showed how to reduce Heisenberg's S-matrix to the full correlation functions. In fact, this chapter combines several important contributions made to quantum field theory by Feynman, Schwinger, Dyson, Gell-Mann and Low. We will proceed as follows.

(i) We start with the classical principle of critical action

$$S[\varphi] = \text{critical!}$$

Here, the action depends on an additional source term which describes the influence of an external source on the quantum system. For our approach, it is crucial that the action of the free field is a quadratic form. This allows us then to apply the methods of Gaussian integrals.

(ii) From (i) we get the classical field equation (Euler–Lagrange equation)

$$D\varphi = -\kappa \mathcal{L}_{\text{int}}(\varphi) - J$$

by using the classical calculus of variations.

(iii) Switching off the interaction by setting $\kappa = 0$, we get the so-called response equation,

$$D\varphi = -J, \tag{14.1}$$

which tells us the response φ of the classical field φ in the presence of an arbitrary external source J. We replace this by the regularized equation

$$\boxed{(D + \mathrm{i}\varepsilon I)\varphi = -J} \tag{14.2}$$

with the small regularization parameter $\varepsilon > 0$, and we assume that the inverse operator $(D + \mathrm{i}\varepsilon I)^{-1}$ exists. Setting $R_\varepsilon := -(D + \mathrm{i}\varepsilon I)^{-1}$, the unique solution of the response equation reads as $\varphi = R_\varepsilon J$. In addition, we assume that the response operator can be represented by the following integral formula

$$\boxed{\varphi(x) = \int_{\mathbb{R}^4} \mathcal{R}_\varepsilon(x - y) J(y) d^4 y}$$

for the space-time points $x = (\mathbf{x}, t)$. The kernel $\mathcal{R}_\varepsilon$ is called the response function. Up to some normalization factor, the response function coincides with the Feynman propagator.[2]

[2] In gauge field theories, one has to modify the classical action by adding a Lagrange-multiplier term in order to guarantee the existence of the inverse operator $(D + \mathrm{i}\varepsilon I)^{-1}$.

(iv) *Interestingly enough, the response function $\mathcal{R}_\varepsilon$ knows all about the quantized field, which differs from the classical field by quantum fluctuations.* In fact, from the response function $\mathcal{R}_\varepsilon$ we get the free 2-point correlation function

$$\boxed{C_{2,\mathrm{free}}(x,y) := -\mathrm{i}\hbar\mathcal{R}_\varepsilon(x-y), \qquad x,y \in \mathbb{M}^4.}$$

The magic quantum action reduction formula (QA) then tells us how to obtain the full n-point correlation function, C_n, of the quantum field from the 2-point free correlation function, $C_{2,\mathrm{free}}$. Finally, the magic reduction formula (LSZ) tells us how to reduce the S-matrix to the full correlation functions, C_n, and the free particle functions which are solutions of the classical free Euler–Lagrange equations.[3]

(v) From the S-matrix we will obtain the cross sections of scattering processes.

(vi) In each order of perturbation theory with respect to powers of the coupling constant κ, the analytic expressions of the magic formulas (QA) and (LSZ) can be graphically represented by Feynman diagrams.

The physical idea behind this approach is the following one. The physics of a quantum field differs from the physics of the corresponding classical field by additional quantum fluctuations of the ground state of the quantum field.

> *Physicists use the following intuitive picture: quantum fluctuations are caused by virtual particles which jump from the ground state (also called vacuum state) of the quantum field to the real world and back to the ground state. These particles are called virtual, since they violate energy-momentum conservation. In the language of Feynman diagrams, the virtual particles correspond to internal lines.*

By the method of moments, we have to compute the moments of the stochastic quantum fluctuations of the ground state. To this end, we use perturbed infinite-dimensional Gaussian integrals where the Gaussian kernel depends on the classical action and the external source. The perturbation of the ideal Gaussian is governed by the interaction term of the Lagrangian density. This yields the global quantum action principle. The moments are given by the n-point correlation functions, C_n.

It is crucial that the magic formulas (QA) and (LSZ) have to be supplemented by the following procedures.

(a) Causality and response: The response function $\mathcal{R}_\varepsilon$ refers not to the classical Euler–Lagrange equation for the motion of the field, but to the *regularized* Euler–Lagrange equation (14.2). This guarantees the validity

[3] In Chap. 15, we will introduce the full n-point Green's function G_n. It turns out that G_n coincides with the full n-point correlation function C_n. Therefore, physicists use the terms 'n-point correlation function' and 'n-point Green's function' synonymously.

of the causality principle and uniquely determines the response function in Fourier space.

(b) Renormalization: As a rule, the magic formulas (QA) and (LSZ) include divergent integrals. The fundamental method of renormalization extracts physical information from these divergent integrals. The idea of the Bogoliubov–Parasiuk–Hepp–Zimmermann (BPHZ) renormalization method is to regularize the divergent integrals by adding counterterms to the Lagrangian and by changing the original parameters[4] into effective parameters which can be measured in physical experiments (e.g., the effective electron mass, the effective electron charge, and the effective coupling constant in quantum electrodynamics). Effective parameters are also called renormalized parameters.

(c) The Faddeev–Popov–De Witt technique in gauge theory: The Standard Model in particle physics represents a gauge theory. In the quantization of gauge theories, the presence of the local symmetry group (gauge group) causes difficulties. According to Faddeev and Popov, in order to overcome the difficulties one has to modify the generating functional by taking the side condition (gauge condition) into account. This leads to additional fields called ghosts which will be discussed in Sect. 16.6.

Notation. In the last chapters of this book, we will use the energetic system of units, that is, we simplify the formulas by setting $c = \hbar = \varepsilon_0 := 1$. Moreover, we will use the following terminology.

(i) Inertial system: We choose a fixed inertial system with right-handed Cartesian position coordinates x^1, x^2, x^3, and time t. The Cartesian coordinates refer to a fixed right-handed orthonormal system $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. For the position vector,

$$\mathbf{x} = x^1\mathbf{e}_1 + x^2\mathbf{e}_2 + x^3\mathbf{e}_3,$$

we briefly write $\mathbf{x} = x^k\mathbf{e}_k$. That is, we sum over the upper and lower Latin index k from 1 to 3. This is a special case of the Einstein summation convention described below.

(ii) Space-time point: Set $x^0 := t$, and

$$x = (x^0, x^1, x^2, x^3).$$

Furthermore, introduce the partial derivative $\partial_\mu f := \frac{\partial f}{\partial x^\mu}$.

(iii) The Minkowski metric: The final goal is to formulate physics in such a way that it respects Einstein's principle of special relativity, that is, the physical processes do not depend on the choice of the inertial system. To this end, in terms of mathematics, one has to use the methods of the theory of invariants applied to the Poincaré group. This will be studied in Volume III on general gauge theory. At this point, we only emphasize

[4] The original parameters are also called bare parameters.

that it is important to distinguish between upper and lower indices. For example, we define

$$\boxed{x_1 := -x^1, \quad x_2 := -x^2, \quad x_3 := -x^3, \quad x_0 := x^0.} \tag{14.3}$$

To systemize this, we introduce the Minkowski symbol $\eta_{\mu\nu}$ for the indices $\mu, \nu = 0, 1, 2, 3$. Explicitly,

$$\eta_{00} := 1, \quad \eta_{11} = \eta_{22} = \eta_{33} := -1,$$

along with $\eta_{\mu\nu} = 0$ if $\mu \neq \nu$. Moreover, we set $\eta^{\mu\nu} := \eta_{\mu\nu}$ for all μ, ν. The space of all space-time points x coincides with $\mathbb{R}^4$. However, in order to indicate that there is an additional structure induced by the Minkowski symbol $(\eta_{\mu\nu})$, we will denote the space of space-time points x by the symbol $\mathbb{M}^4$. This refers to Minkowski (1864–1909). In 1908 he emphasized the role of non-classical geometry for describing the structure of the four-dimensional space-time. Nowadays we use the terms Minkowski space and Minkowski geometry.

(iv) The Einstein summation convention: We sum over equal upper and lower Greek indices from 0 to 3. For example,

$$a_\mu b^\mu := \sum_{\mu=0}^{3} a_\mu b^\mu.$$

In particular, we will use the Minkowski symbol $\eta_{\mu\nu}$ in order to lift and to lower indices. For example, $x_\mu := \eta_{\mu\nu} x^\nu$, and

$$\partial^\mu := \eta^{\mu\nu}\partial_\nu, \qquad \partial_\mu = \eta_{\mu\nu}\partial^\nu, \qquad F^{\mu\nu} := \eta^{\mu\alpha}\eta^{\nu\beta}F_{\alpha\beta}.$$

Furthermore, if not stated otherwise, we sum over equal upper and lower Latin indices from 1 to 3. In particular, for the electric field $\mathbf{E}$ and the magnetic field $\mathbf{B}$, we write $\mathbf{E} = E^k\mathbf{e}_k = \sum_{k=1}^{3} E^k\mathbf{e}_k$ and $\mathbf{B} = B^k\mathbf{e}_k$.

(v) The principle of the right index picture: This very convenient principle tells us that an expression represents an invariant under the transformations of the Poincaré group if it has no free indices. For example, set

$$xy := x^\mu y_\mu.$$

Here, we sum over the upper and lower index μ. There is no free index anymore. Consequently, the inner product xy is an invariant under the Poincaré group. Such invariants are independent of the choice of the inertial system, and hence they possess a physical meaning. As another example, note that

$$F_{\mu\nu}F^{\mu\nu}$$

has no free index, and hence it is an invariant under the Poincaré group. We will show below that this is equal to $\mathbf{B}^2 - \mathbf{E}^2$. More generally, if an

equation has the right index picture, that is, the free indices on both sides of the equation are the same, then this equation is relativistically invariant, that is, it possesses the right transformation behavior under a change of inertial systems. For example, the equation

$$\partial_\mu \partial^\mu A^\nu = J^\nu_{\mathrm{QED}} \tag{14.4}$$

has the right index picture because we have the free index ν on both sides of the equation. Equation (14.4) is one of the equations of motion in quantum electrodynamics to be considered in Sect. 14.3.

The principle of the right index picture exists for general transformations including transformations of tensors, spinors and fiber bundles.

This principle is very useful for physics; it will be justified in Volume III on gauge theory.

14.1 The Fourier–Minkowski Transform

In quantum field theory, it is very useful to pass to momentum space via Fourier–Minkowski transform.

Folklore

For the convenience of the reader, let us summarize the most important properties of the Fourier–Minkowski transform in the energetic system of units.

Definition. Recall that $x = (\mathbf{x}, t)$. Similarly, $p = (\mathbf{p}, E)$, and

$$px := Et - \mathbf{px}.$$

For each function $g \in \mathcal{S}(\mathbb{M}^4)$, the Fourier–Minkowski transform is defined by

$$\boxed{(\mathcal{F}_M g)(p) := \frac{\alpha}{(2\pi)^2} \int_{\mathbb{M}^4} g(x) \mathrm{e}^{\mathrm{i}px}\, d^4x, \qquad p \in \mathbb{M}^4} \tag{14.5}$$

along with the inverse transformation[5]

$$g(x) = \frac{1}{(2\pi)^2 \alpha} \int_{\mathbb{M}^4} (\mathcal{F}_M g)(p)\, \mathrm{e}^{-\mathrm{i}px}\, d^4p, \qquad x \in \mathbb{M}^4.$$

The normalization factor α will be fixed in (14.12) below. We also write the symbol $\hat{f}$ for the Fourier–Minkowski transform $\mathcal{F}_M f$ of the function f. This transformation has the following properties:

[5] Recall that the spaces $\mathbb{M}^4$ and $\mathbb{R}^4$ coincide as linear spaces. Therefore, we have $\mathcal{S}(\mathbb{M}^4) = \mathcal{S}(\mathbb{R}^4)$ and $L_2(\mathbb{M}^4) = L_2(\mathbb{R}^4)$.

(i) The operator $\mathcal{F}_M : \mathcal{S}(\mathbb{M}^4) \to \mathcal{S}(\mathbb{M}^4)$ is linear, bijective, and sequentially continuous.

(ii) Parseval equation: For all $g, h \in \mathcal{S}(\mathbb{M}^4)$,

$$\langle \mathcal{F}_M g | \mathcal{F}_M h \rangle_{L_2(\mathbb{M}^4)} = \alpha^2 \langle g | h \rangle_{L_2(\mathbb{M}^4)}. \tag{14.6}$$

Here, we use the inner product $\langle g|h\rangle_{L_2(\mathbb{M}^4)} := \int_{\mathbb{M}^4} g(x)^\dagger h(x) d^4x$.

(iii) Duality: For all $g, h \in \mathcal{S}(\mathbb{M}^4)$,

$$\boxed{\langle \mathcal{F}_M g | h \rangle = \langle g | \mathcal{F}_M h \rangle.} \tag{14.7}$$

Here, we introduce the duality pairing $\langle g|h\rangle := \int_{\mathbb{M}^4} g(x)h(x)d^4x$. In other words, the operator $\mathcal{F}_M$ is self-dual with respect to the duality pairing $\langle .|. \rangle$. For all $f, \varphi \in \mathcal{S}(\mathbb{M}^4)$, this implies

$$\int_{\mathbb{M}^4} f(x)\varphi(x) d^4x = \int_{\mathbb{M}^4} (\mathcal{F}_M f)(p)(\mathcal{F}_M^{-1}\varphi)(p) d^4p.$$

Equivalently,

$$\int_{\mathbb{M}^4} f(x)\varphi(x) d^4x = \frac{1}{\alpha^2} \int_{\mathbb{M}^4} \hat{f}(p)\hat{g}(-p) d^4p. \tag{14.8}$$

(iv) Partial derivatives: For all $p \in \mathbb{M}^4$ and all $\mu = 0, 1, 2, 3$,

$$\boxed{\mathcal{F}_M(\mathrm{i}\partial_\mu g)(p) = p_\mu (\mathcal{F}_M g)(p).}$$

Similarly, $(\mathcal{F}_M^{-1})(\mathrm{i}\partial_\mu g)(p) = -p_\mu (\mathcal{F}_M^{-1} g)(p)$.

(v) Convolution: For all $g, h \in \mathcal{S}(\mathbb{R}^4)$,

$$\boxed{\mathcal{F}_M(g * h) = \frac{(2\pi)^2}{\alpha} (\mathcal{F}_M g)(\mathcal{F}_M h).} \tag{14.9}$$

(vi) Dirac's delta distribution:

$$\mathcal{F}_M \delta = \frac{\alpha}{(2\pi)^2}.$$

This means that for all test functions $\varphi \in \mathcal{S}(\mathbb{M}^4)$, we have $\delta(\varphi) = \varphi(0)$ along with

$$(\mathcal{F}_M \delta)(\varphi) = \int_{\mathbb{M}^4} \frac{\alpha}{(2\pi)^2} \varphi(x) \, d^4x$$

and

$$\delta(\varphi) = \int_{\mathbb{M}^4} \frac{\alpha}{(2\pi)^2} (\mathcal{F}_M^{-1}\varphi)(p) \, d^4p.$$

This is a special case of the general definition (14.10) below.

Let us study the simple relation between the Euclidean Fourier transform, $\mathcal{F}$, and the Fourier–Minkowski transform, $\mathcal{F}_M$. By Sect. 10.3.3, the Euclidean Fourier transform looks like

$$(\mathcal{F}g)(p) = \frac{1}{(2\pi)^2} \int_{\mathbb{M}^4} g(x)\, \mathrm{e}^{-\mathrm{i}Et} \mathrm{e}^{-\mathrm{i}\mathbf{px}} d^4x.$$

Define the time reflection operator operator $\mathcal{R}(\mathbf{x}, t) := (\mathbf{x}, -t)$. Since

$$\begin{aligned}(\mathcal{F}\mathcal{R}g)(p) &= \frac{1}{(2\pi)^2} \int_{\mathbb{M}^4} g(\mathbf{x}, -t) \mathrm{e}^{-\mathrm{i}Et} \mathrm{e}^{-\mathrm{i}\mathbf{px}} d^4x \\ &= \frac{1}{(2\pi)^2} \int_{\mathbb{M}^4} g(\mathbf{x}, t) \mathrm{e}^{\mathrm{i}Et} \mathrm{e}^{-\mathrm{i}\mathbf{px}} d^4x,\end{aligned}$$

we obtain

$$\boxed{\mathcal{F}_M g = \alpha \mathcal{F}\mathcal{R}g, \qquad g \in \mathcal{S}(\mathbb{R}^4).}$$

Therefore, the statements (i)-(vi) above are immediate consequences of the properties of the Euclidean Fourier transform, $\mathcal{F}$.

The Fourier–Minkowski transform, $\mathcal{F}_M T$, of a tempered distribution $T \in \mathcal{S}'(\mathbb{R}^4)$ is defined by

$$\boxed{(\mathcal{F}_M T)(\varphi) := T(\mathcal{F}_M \varphi) \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}^4).} \tag{14.10}$$

As an example, fix a continuous rational function $\hat{f} : \mathbb{M}^4 \to \mathbb{C}$. Define

$$T(\varphi) := \int_{\mathbb{M}^4} \hat{f}(p) (\mathcal{F}_M^{-1}\varphi)(p)\, d^4p$$

for all test functions $\varphi \in \mathcal{S}(\mathbb{R}^4)$. Equivalently,

$$T(\varphi) := \frac{1}{\alpha^2} \int_{\mathbb{M}^4} \hat{f}(p) \hat{\varphi}(-p)\, d^4p. \tag{14.11}$$

Then, T is a tempered distribution, $T \in \mathcal{S}'(\mathbb{R}^4)$, and the Fourier–Minkowski transform of T corresponds to the function $\hat{f}$. Explicitly,

$$(\mathcal{F}_M T)(\varphi) = T(\mathcal{F}_M \varphi) = \int_{\mathbb{R}^4} \hat{f}(p) \varphi(p) d^4p, \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}^4).$$

Convention. In this monograph, the normalization factor α of the Fourier–Minkowski transform (14.5) is chosen as

$$\boxed{\alpha := 1.} \tag{14.12}$$

By (i) above, this convention guarantees a simple form of the Parseval equation. This implies that the operator $\mathcal{F}_M$ can be uniquely continued to a unitary operator

$$\mathcal{F}_M : L_2(\mathbb{M}^4) \to L_2(\mathbb{M}^4)$$

on the Hilbert space $L_2(\mathbb{M}^4)$. Let us note that physicists frequently use the convention $\alpha = (2\pi)^2$. In this case, the convolution rule becomes simple, but the unitarity of the Fourier transform is destroyed. In quantum theory, unitary operators preserve the physics. In particular, probabilities, mean values, and mean fluctuations are preserved. This underlines the importance of unitary operators in quantum theory. The disadvantage of our convention is that the Feynman propagators get additional factors in momentum space.

14.2 The φ^4-Model

We are going to study the nonlinear Klein–Gordon equation

$$\boxed{\Box\varphi + m_0^2\varphi + 4\kappa\varphi^3 = J} \tag{14.13}$$

on an inertial system. Explicitly, $\Box\varphi := \varphi_{tt} + \Delta\varphi = \varphi_{tt} - \sum_{j=1}^3 \varphi_{x^j x^j}$. We are looking for a function $\varphi : \mathbb{M}^4 \to \mathbb{R}$. This is a model for an uncharged meson with spin zero and rest mass $m_0 > 0$. The constant $\kappa \geq 0$ is called the coupling constant of the self-interaction of the particle. The given smooth function $J : \mathbb{R}^4 \to \mathbb{R}$ describes an external source. Let $\kappa = 0$. For a given momentum vector $\mathbf{p}$, define the energy

$$E_{\mathbf{p}} := \sqrt{\mathbf{p}^2 + m_0^2}.$$

Then the function

$$u_{\mathbf{p}}(\mathbf{x}, t) := \text{const} \cdot \mathrm{e}^{\mathrm{i}(\mathbf{p}\mathbf{x} - E_{\mathbf{p}}t)} \tag{14.14}$$

is a solution of equation (14.13) with $\kappa = 0$. This function describes a free meson of rest mass m_0, momentum vector $\mathbf{p}$, and energy $E_{\mathbf{p}}$. In particular, if $\mathbf{p} = 0$, then $E_0 = m_0$.

Now choose $\kappa = 0$, and let $\mu > 0$. The damped oscillation

$$\varphi(t) = \text{const} \cdot \mathrm{e}^{-\mathrm{i}(E_0 - \mathrm{i}\mu)t} = \text{const} \cdot \mathrm{e}^{-\mathrm{i}E_0 t}\mathrm{e}^{-\mu t}, \qquad t \in \mathbb{R}$$

is a solution of the equation

$$\Box\varphi + (m_0 - \mathrm{i}\mu)^2\varphi = 0.$$

Setting $\varepsilon := 2\mu m_0$ and assuming that ε is small, then we get the approximative equation

$$\Box\varphi + m_0^2\varphi - \mathrm{i}\varepsilon\varphi = 0.$$

This argument motivates the negative sign of the regularizing term $-\mathrm{i}\varepsilon\varphi$.

14.2.1 The Classical Principle of Critical Action

Let Ω be a nonempty bounded open subset of $\mathbb{R}^4$. Choose $\varepsilon > 0$. The action functional of the meson reads as

$$S[\varphi, J] := \int_\Omega \mathcal{L}(\varphi, \partial\varphi, J)\, d^4x$$

with the Lagrangian density

$$\mathcal{L}(\varphi, \partial\varphi, J) := \frac{1}{2}\varphi(D + \mathrm{i}\varepsilon)\varphi + \kappa\mathcal{L}_{\mathrm{int}}(\varphi) + \varphi J.$$

Here, we set $D := -\Box - m_0^2$, and $\mathcal{L}_{\mathrm{int}}(\varphi) := -\varphi^4$. The variational problem

$$S[\varphi, J] = \text{critical!} \tag{14.15}$$

with the boundary condition $\varphi = 0$ on $\partial\Omega$ is called the principle of critical action for the meson. By Theorem 7.42 on page 460, each smooth solution of (14.15) satisfies the following regularized meson equation

$$\Box\varphi + m_0^2\varphi - \mathrm{i}\varepsilon\varphi + 4\kappa\varphi^3 = J. \tag{14.16}$$

Integration by parts yields

$$S[\varphi, J] = \int_\Omega \left(\frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi + \kappa\mathcal{L}_{\mathrm{int}}(\varphi) + \frac{1}{2}\mathrm{i}\varepsilon\varphi^2 + \varphi J\right) d^4x.$$

14.2.2 The Response Function and the Feynman Propagator

> In physics, use the right approximations and the right limits.
> The golden rule

Switching off the interaction by setting $\kappa = 0$, we get the so-called response equation

$$\Box\varphi + m_0^2\varphi - \mathrm{i}\varepsilon\varphi = J. \tag{14.17}$$

It is our goal to study this equation in rigorous terms by using the theory of tempered distributions and the Fourier–Minkowski transform. To motivate this, let us first apply formal arguments.

Formal response kernel function. We want to use the formal Fourier–Minkowski transform in order to solve equation (14.17). We will show in a formal manner that the solution of (14.17) can be written as

$$\varphi(x) = \int_{\mathbb{M}^4} \mathcal{R}_\varepsilon(x - y)J(y)d^4y, \qquad x \in \mathbb{M}^4 \tag{14.18}$$

with the response kernel function

$$\mathcal{R}_\varepsilon(x) := -\frac{1}{(2\pi)^4} \int_{\mathbb{M}^4} \frac{\mathrm{e}^{-\mathrm{i}px}}{p^2 - m_0^2 + \mathrm{i}\varepsilon}\, d^4p. \tag{14.19}$$

This function is also called the response function for mesons. The formal proof goes like this. We will use the Fourier–Minkowski transform from Sect. 14.1. Define $\hat{\varphi} := \mathcal{F}_M\varphi$, $\hat{J} := \mathcal{F}_M J$, and $\hat{\mathcal{R}}_\varepsilon := \mathcal{F}_M\mathcal{R}_\varepsilon$ Since

$$\mathcal{F}_M(-\partial_\mu\partial^\mu\varphi) = p_\mu p^\mu\hat{\varphi}(p) = p^2\hat{\varphi}(p),$$

it follows from the response equation (14.17) that

$$(-p^2 + m_0^2 - \mathrm{i}\varepsilon)\hat{\varphi}(p) = \hat{J}(p).$$

Hence

$$\hat{\varphi}(p) = -\frac{\hat{J}(p)}{p^2 - m_0^2 + \mathrm{i}\varepsilon}.$$

Applying the convolution formula (14.9) to (14.18), we obtain

$$\hat{\varphi}(p) = (2\pi)^2\hat{\mathcal{R}}_\varepsilon(p)\hat{J}(p).$$

Therefore,

$$\boxed{\hat{\mathcal{R}}_\varepsilon(p) = -\frac{1}{(2\pi)^2(p^2 - m_0^2 + \mathrm{i}\varepsilon)}.} \tag{14.20}$$

This implies (14.19). Choosing $J(x) := \delta^4(x)$, we get formally that

$$(\Box + m_0^2 - \mathrm{i}\varepsilon)\mathcal{R}_\varepsilon(x) = \delta^4(x). \tag{14.21}$$

The Feynman propagator and the free 2-point correlation function. Set

$$\mathcal{G}_{F,m_0,\varepsilon}(x) := -\mathrm{i}\mathcal{R}_\varepsilon(x)$$

along with

$$\boxed{C_{2,\mathrm{free},\varepsilon}(x,y) := \mathcal{G}_{F,m_0,\varepsilon}(x-y), \qquad x,y \in \mathbb{M}^4.}$$

The function $\mathcal{G}_{F,m_0,\varepsilon}$ is called the Feynman propagator of the regularized Klein–Gordon equation (or the meson propagator).[6] Note that the definition of the free 2-point correlation function $C_{2,\mathrm{free},\varepsilon}$ is motivated by the discrete φ^4-model on page 464. By (14.19),

$$\boxed{\mathcal{G}_{F,m_0,\varepsilon}(x) = \frac{\mathrm{i}}{(2\pi)^4} \int_{\mathbb{M}^4} \frac{\mathrm{e}^{-\mathrm{i}px}}{p^2 - m_0^2 + \mathrm{i}\varepsilon}\, d^4p.} \tag{14.22}$$

[6] Unfortunately, for historical reasons, there are different definitions of the Feynman propagator in the literature which differ by the factor i.

The Fourier–Minkowski transform of $\mathcal{G}_{F,m_0,\varepsilon}$ is given by

$$\hat{\mathcal{G}}_{F,m_0,\varepsilon}(p) = -\mathrm{i}\hat{\mathcal{R}}_\varepsilon(p) = \frac{\mathrm{i}}{(2\pi)^2(p^2 - m_0^2 + \mathrm{i}\varepsilon)}. \tag{14.23}$$

It follows formally from the duality relation (14.8) that

$$\int_{\mathbb{M}^4} \mathcal{G}_{F,m_0,\varepsilon}(x)\varphi(x)d^4x = \int_{\mathbb{M}^4} \frac{\mathrm{i}\hat{\varphi}(-p)}{(2\pi)^2(p^2 - m_0^2 + \mathrm{i}\varepsilon)}\, d^4p \tag{14.24}$$

for all test functions $\varphi \in \mathcal{S}(\mathbb{R}^4)$.

Warning to the reader. Observe that formulas (14.19), (14.22) look very elegant, but they are meaningless, since the integrals do not exist in the classical sense. In fact, the integrands decrease too slowly at infinity.

In contrast to this bad behavior in position space, both $\mathcal{R}_\varepsilon$ and $\mathcal{G}_{F,m_0,\varepsilon}$ are well-defined objects in momentum space, by (14.20).

This fact can be used in order to define R_ε and $\mathsf{G}_{F,m_0,\varepsilon}$ as well-defined tempered distributions on position space. In particular, we will show that formulas (14.19), (14.22) are valid in the theory of tempered distributions. In order to find the rigorous approach, we will use formulas (14.19), (14.22) as mnemonic tools.

The rigorous response operator. Setting $D := -\Box - m_0^2$, the response equation (14.17) reads as

$$\boxed{(D + \mathrm{i}\varepsilon)\varphi = -J.}$$

If $\varphi \in \mathcal{S}(\mathbb{R}^4)$, then $(D + \mathrm{i}\varepsilon I)\varphi \in \mathcal{S}(\mathbb{R}^4)$.Therefore the linear operator

$$D + \mathrm{i}\varepsilon I : \mathcal{S}(\mathbb{R}^4) \to \mathcal{S}(\mathbb{R}^4)$$

is well-defined.

Proposition 14.1 *The operator $D + \mathrm{i}\varepsilon I$ is invertible.*

Proof. We will use the Fourier–Minkowski transform along with the following diagram:

$$\begin{array}{ccc} \mathcal{S}(\mathbb{R}^4) & \underset{-R_\varepsilon}{\overset{D+\mathrm{i}\varepsilon I}{\rightleftarrows}} & \mathcal{S}(\mathbb{R}^4) \\ \Big\downarrow{\scriptstyle \mathcal{F}_M} & & \Big\uparrow{\scriptstyle \mathcal{F}_M^{-1}} \\ \mathcal{S}(\mathbb{R}^4) & \underset{-\widehat{R}_\varepsilon}{\overset{\widehat{(D+\mathrm{i}\varepsilon I)}}{\rightleftarrows}} & \mathcal{S}(\mathbb{R}^4) \end{array} \tag{14.25}$$

(I) Fourier space (momentum space). For each function $\hat{\varphi} \in \mathcal{S}(\mathbb{R}^4)$, define

$$(\hat{R}_\varepsilon\hat{\varphi})(p) := -\frac{\hat{\varphi}(p)}{p^2 - m_0^2 + \mathrm{i}\varepsilon}, \qquad p \in \mathbb{M}^4.$$

The linear operator $\hat{R}_\varepsilon : \mathcal{S}(\mathbb{R}^4) \to \mathcal{S}(\mathbb{R}^4)$ is well-defined and invertible. For the inverse operator, we have

$$(\hat{R}_\varepsilon^{-1}\hat{J})(p) = -(p^2 - m_0^2 + \mathrm{i}\varepsilon)\hat{J}(p), \qquad p \in \mathbb{M}^4.$$

This corresponds to the operator $-(D + \mathrm{i}\varepsilon I)$ in momentum space.

(II) Original position space. Using the inverse Fourier–Minkowski transform, we define

$$R_\varepsilon := \mathcal{F}_M^{-1}\hat{R}_\varepsilon\mathcal{F}_M.$$

Obviously, $(D + \mathrm{i}\varepsilon I) = -R_\varepsilon^{-1}$. Hence $R_\varepsilon = -(D + \mathrm{i}\varepsilon I)^{-1}$. □

The rigorous response distribution. We now want to show that the response kernel function $x \mapsto \mathcal{R}_\varepsilon(x)$ and hence the Feynman propagator

$$x \mapsto \mathcal{G}_{F,m_0,\varepsilon}(x)$$

exist as tempered distributions, and the response integral relation (14.18) is to be understood in the sense of the convolution of distributions. Motivated by (14.24), we define

$$\boxed{\mathsf{G}_{F,m_0,\varepsilon}(\varphi) := \int_{\mathbb{M}^4} \frac{\mathrm{i}\hat{\varphi}(-p)}{(2\pi)^2(p^2 - m_0^2 + \mathrm{i}\varepsilon)}\, d^4p} \tag{14.26}$$

for all test functions $\varphi \in \mathcal{S}(\mathbb{R}^4)$. This is a tempered distribution. In addition, we define the tempered response distribution by setting

$$\mathsf{R}_\varepsilon := \mathrm{i}\mathsf{G}_{F,m_0}.$$

According to (14.11), taking $\alpha = 1$, the Fourier-Minkowski transforms of the tempered distributions $\mathsf{G}_{F,m_0,\varepsilon}$ and R_ε are given by (14.23).

Theorem 14.2 (i) *The tempered distribution R_ε is a fundamental solution of the special response equation,*

$$(\Box + m_0^2 - \mathrm{i}\varepsilon)\mathsf{R}_\varepsilon = \delta.$$

(ii) *For each source function $J \in \mathcal{D}(\mathbb{R}^4)$, the distribution*

$$\varphi := \mathsf{R}_\varepsilon * J \tag{14.27}$$

is a solution of the response equation: $(\Box + m_0^2 - \mathrm{i}\varepsilon)\varphi = J$.

Proof. Ad (i). Obviously,

$$-\int_{\mathbb{M}^4} \frac{\{-p^2 + m_0^2 - \mathrm{i}\varepsilon\}(\mathcal{F}_M^{-1}\varphi)(p)}{(2\pi)^2(p^2 - m_0^2 + \mathrm{i}\varepsilon)}\, d^4p$$

is equal to

$$\frac{1}{(2\pi)^2}\int_{\mathbb{M}^4}(\mathcal{F}_M^{-1}\varphi)(p)\,d^4p=\varphi(0).$$

Note that the Fourier transform sends differential operators to multiplication operators. Explicitly,

$$\mathcal{F}_M^{-1}(\{\Box+m_0^2-\mathrm{i}\varepsilon\}\varphi)=\{-p^2+m_0^2-\mathrm{i}\varepsilon\}\mathcal{F}_M^{-1}\varphi$$

at the point p. Noting that $(\mathcal{F}_M^{-1}\varphi)(p)=\hat{\varphi}(-p)$ it follows from (14.26) that

$$\mathsf{R}_\varepsilon(\{\Box+m_0^2-\mathrm{i}\varepsilon\}\varphi)=\varphi(0)=\delta(\varphi)$$

for all test functions $\varphi\in\mathcal{S}(\mathbb{R}^4)$. This is the claim.

Ad (ii). This is a special case of Theorem 11.27 on page 645. □

The rigorous Feynman propagator distribution. The distribution $\mathsf{G}_{F,m_0,\varepsilon}$ depends on the regularizing parameter $\varepsilon>0$. In order to obtain a propagator distribution which is independent of ε, let us study the limit

$$\boxed{\mathsf{G}_{F,m_0}(\varphi):=\lim_{\varepsilon\to+0}\mathsf{G}_{F,m_0,\varepsilon}(\varphi)} \tag{14.28}$$

for all test functions $\varphi\in\mathcal{S}(\mathbb{R}^4)$. This is the limit in the space $\mathcal{S}'(\mathbb{R}^4)$ of tempered distributions. The idea is to use Schwinger's integration trick

$$\frac{\mathrm{i}}{(2\pi)^2(p^2-m_0^2+\mathrm{i}\varepsilon)}=\frac{1}{4(2\pi)^2}\int_0^\infty\exp\left\{\frac{\mathrm{i}(p^2-m_0^2+\mathrm{i}\varepsilon)}{4\xi}\right\}\frac{d\xi}{\xi^2}.$$

In fact, if $\varepsilon>0$ and $a\in\mathbb{R}$, then the substitution $\eta:=1/4\xi$ yields

$$\frac{1}{4}\int_0^\infty \mathrm{e}^{-\varepsilon/4\xi}\,\mathrm{e}^{\mathrm{i}a/4\xi}\,\frac{d\xi}{\xi^2}=-\int_0^\infty\mathrm{e}^{-\varepsilon\eta}\,\mathrm{e}^{\mathrm{i}a\eta}\,d\eta=\frac{4\mathrm{i}}{a+\mathrm{i}\varepsilon}.$$

Note the appearance of the damping factor $\mathrm{e}^{-\varepsilon\eta}$. Hence $\mathsf{G}_{F,m_0,\varepsilon}(\varphi)$ is equal to the following integral

$$\frac{1}{4(2\pi)^2}\int_{\mathbb{M}^4}d^4p\,\hat{\varphi}(p)\int_0^\infty\exp\left\{\frac{\mathrm{i}(p^2-m_0^2+\mathrm{i}\varepsilon)}{4\xi}\right\}\frac{d\xi}{\xi^2}$$

along with the Fourier–Minkowski transform $\hat{\varphi}(p)=(2\pi)^{-2}\int_{\mathbb{M}^4}\varphi(x)\mathrm{e}^{\mathrm{i}px}d^4x$. For the proof of the following theorem, see Problem 14.13 on page 811.

Theorem 14.3 *The limit* (14.28) *exists and defines the tempered distribution* G_{F,m_0}. *Explicitly,*

$$\mathsf{G}_{F,m_0}(\varphi)=\frac{1}{4\pi\mathrm{i}}\int_{\mathbb{M}^4}\delta(t^2-\mathbf{x}^2)\varphi(x)d^4x+\mathrm{i}\int_{\mathbb{M}^4}A(x)\varphi(x)d^4x$$

with

$$A(x) := \frac{m_0\theta(x^2)}{8\pi\sqrt{x^2}}\left(J_1(m_0\sqrt{x^2}) - \mathrm{i}N_1(m_0\sqrt{x^2})\right) + \\ + \frac{\mathrm{i}m_0\,\theta(-x^2)}{8\pi\sqrt{-x^2}}\,(J_1(\mathrm{i}m_0\sqrt{-x^2}) + \mathrm{i}N_1(\mathrm{i}m_0\sqrt{-x^2}))$$

where $x = (\mathbf{x}, t)$ and $x^2 := t^2 - \mathbf{x}^2$. Here, J_1 and N_1 are Bessel and Neumann functions, respectively, and θ denotes the Heaviside function.

The tempered distribution G_{F,m_0} is called the Feynman propagator of the Klein–Gordon equation (or the Feynman propagator for mesons). Note that we define the square root by $\sqrt{-l} := \mathrm{i}\sqrt{l}$ if $l > 0$. Moreover, according to Sect. 11.2.4 on page 595, we use the following convention

$$\int_{\mathbb{M}^4} \delta(t^2 - \mathbf{x}^2)\varphi(x)d^4x := \sum_{\sigma=\pm 1} \int_{\mathbb{R}^3} \frac{\varphi(\mathbf{x}, \sigma||\mathbf{x}||)}{2||\mathbf{x}||}\, d^3\mathbf{x}.$$

Mnemonically, we write

$$\boxed{\mathsf{G}_{F,m_0}(\varphi) = \int_{\mathbb{M}^4} \mathcal{G}_{F,m_0}(x)\varphi(x)d^4x}$$

along with

$$\mathcal{G}_{F,m_0}(x) := \frac{\delta(x^2)}{4\pi\mathrm{i}} + \mathrm{i}A(x).$$

This shows that in the position space, the main singularities of the distribution G_F are concentrated on the light cone given by the equation $t^2 - \mathbf{x}^2 = 0$. The Bessel and Neumann functions possess the following behavior near the origin $z = 0$ in the complex plane:

$$J_1(z) = \frac{z}{2} - \frac{z^3}{16} + O(z^5), \qquad z \to 0,$$
$$N_1(z) = -\frac{2}{\pi z} + \frac{z}{\pi}\left(\ln\frac{z}{2} + 1\right) + O(z^3).$$

Consequently, for small values $x^2 = t^2 - \mathbf{x}^2$, we obtain

$$\mathcal{G}_{F,m_0}(x) = \frac{\delta(x^2)}{4\pi\mathrm{i}} + \frac{\mathrm{i}m_0^2\theta(x^2)}{16\pi} - \\ - \frac{1}{4\pi^2x^2} + \frac{m_0^2}{8\pi^2}\,\ln\frac{m_0\sqrt{|x^2|}}{2} \tag{14.29}$$

up to terms of higher order with respect to x^2. This formula describes precisely the behavior of the Feynman propagator distribution near the light cone.

Finally, let us introduce some notation which is frequently used. For each tempered distribution, there exists the Fourier–Minkowski transform. Motivated by the investigations above, the Fourier–Minkowski transform of the tempered distribution G_{F,m_0} is denoted by

$$\frac{\mathrm{i}}{(2\pi)^2(p^2-m_0^2+0_+\mathrm{i})} := \mathcal{F}_M(\mathsf{G}_{F,m_0}).$$

Applying the inverse transformation, we get

$$\mathsf{G}_{F,m_0} = \mathcal{F}_M^{-1}\left(\frac{\mathrm{i}}{(2\pi)^2(p^2-m_0^2+0_+\mathrm{i})}\right).$$

Mnemonically, we write

$$\mathsf{G}_{F,m_0} = \frac{\mathrm{i}}{(2\pi)^4}\int_{\mathbb{M}^4}\frac{\mathrm{e}^{-\mathrm{i}px}}{p^2-m_0^2+0_+\mathrm{i}}\,d^4p. \tag{14.30}$$

Furthermore, using the equality $\mathsf{R}_\varepsilon = \mathrm{i}\mathsf{G}_{F,m_0}$ along with Theorem 14.2, we have

$$\boxed{(\Box+m_0^2)\mathsf{G}_{F,m_0} = -\mathrm{i}\delta.} \tag{14.31}$$

Singularities of the free 2-point correlation function. Motivated by the limit $\varepsilon \to +0$, we define the free 2-point correlation function of the φ^4-model by setting

$$\boxed{C_{2,\mathrm{free}}(x,y) := \mathcal{G}_{F,m_0}(x-y), \qquad x,y\in\mathbb{M}^4.}$$

If $(x-y)^2 = (t-\tau)^2 - (\mathbf{x}-\mathbf{y})^2$ is small, then there arise singularities given by (14.29). Explicitly,

$$C_{2,\mathrm{free}}(x,y) = \frac{\delta((x-y)^2)}{4\pi\mathrm{i}} + \frac{\mathrm{i}m_0^2\,\theta((x-y)^2)}{16\pi} - \\ -\frac{1}{4\pi^2(x-y)^2} + \frac{m_0^2}{8\pi^2}\,\ln\frac{m_0\sqrt{|(x-y)^2|}}{2}.$$

This is true for small values of $(x-y)^2$ up to terms of higher order with respect to $(x-y)^2$. In other words, consider the quantum field φ at the position $\mathbf{x}$ at time t and at the position $\mathbf{y}$ at time τ. Then the correlation of the quantum field between these two events is very large if $(t-\tau)^2-(\mathbf{x}-\mathbf{y})^2$ is very small. In the limit case where $(t-\tau)^2-(\mathbf{x}-\mathbf{y})^2 = 0$, the two events can be connected by a light signal.

The singularities of the correlation function $C_{2,\mathrm{free}}$ are related to causality.

The vertex distribution. We define the free 2-point vertex distribution by

$$\boxed{\mathsf{V}_{2,\mathrm{free}}(\varphi) := \int_{\mathbb{M}^4}(p^2-m_0^2)\hat{\varphi}(-p)\,d^4p}$$

for all test functions $\varphi \in \mathcal{S}(\mathbb{R}^4)$. Let us motivate this definition by using a formal argument. Recall first the response equation

$$(\Box + m_0^2 - \mathrm{i}\varepsilon)\varphi = J$$

along with its formal solution

$$\varphi(x) = \int_{\mathbb{M}^4} \mathcal{R}_\varepsilon(x-y)J(y)d^4y, \qquad x \in \mathbb{M}^4. \tag{14.32}$$

Using the Fourier–Minkowski transform,

$$\hat{\varphi}(p) = -\frac{\hat{J}(p)}{(2\pi)^2(p^2 - m_0^2 + \mathrm{i}\varepsilon)}.$$

Motivated by (7.157) on page 474, we introduce the vertex function $V_{2,\mathrm{free}}$ by writing

$$\boxed{J(x) = -\int_{\mathbb{M}^4} V_{2,\mathrm{free}}(x-y)\varphi(y)d^4y, \qquad x \in \mathbb{M}^4.} \tag{14.33}$$

Up to sign, this is the inverse transformation to (14.32). Using the Fourier–Minkowski transform,

$$\hat{J}(p) = -(2\pi)^2\hat{V}_{2,\mathrm{free}}(p)\hat{\varphi}(p).$$

Hence $\hat{V}_{2,\mathrm{free}}(p) = p^2 - m_0^2$, after letting $\varepsilon \to +0$. We now define the tempered distribution $\mathsf{V}_{2,\mathrm{free}}$ in such a way that its Fourier–Minkowski transform corresponds to $\hat{V}_{2,\mathrm{free}}$. This is the definition given above, by (14.11) with $\alpha = 1$.

The Feynman propagator distribution for massless particles. The definition of G_{F,m_0} also makes sense if we set $m_0 = 0$ (vanishing mass). Then, all the formulas above remain valid by setting

$$m_0 = 0, \qquad m_0^2 \ln m_0 := 0.$$

Instead of $\mathsf{G}_{F,m_0=0}$, we briefly write G_F. In particular,

$$\Box\mathsf{G}_F = -\mathrm{i}\delta, \tag{14.34}$$

and

$$\mathsf{G}_F = \frac{\mathrm{i}}{(2\pi)^4}\int_{\mathbb{M}^4} \frac{\mathrm{e}^{-\mathrm{i}px}}{p^2 + 0_+\mathrm{i}}\, d^4p. \tag{14.35}$$

In what follows, we will study the extended quantum action functional, the magic quantum action reduction formula for correlation functions, the magic LSZ reduction formula for the S-matrix, and the local quantum action principle.

14.2.3 The Extended Quantum Action Functional

> The extended quantum action functional knows all about the quantum field.
>
> Folklore

We define the extended quantum action functional $(J,\varphi) \mapsto Z(J,\varphi)$ of the φ^4-model by setting

$$\boxed{Z(J,\varphi) := \exp\Big\{\mathrm{i}\kappa \int_{\mathbb{M}^4} d^4x\ \mathcal{L}_{\mathrm{int}}\left(\frac{1}{\mathrm{i}}\ \frac{\delta}{\delta J(x)}\right)\Big\} \cdot Z_{\mathrm{free,source}}(J,\varphi)}$$

where

$$Z_{\mathrm{free,source}}(J,\varphi) := Z_{\mathrm{free}}(J)\ \mathrm{e}^{Z_{\mathrm{source}}(J)}. \tag{14.36}$$

Here, we set

$$Z_{\mathrm{free}}(J) := \mathrm{e}^{\frac{1}{2}\mathrm{i}\langle J|R_\varepsilon J\rangle}$$

along with the response functional

$$\langle J|R_\varepsilon J\rangle := \int_{\mathbb{M}^4\times\mathbb{M}^4} J(x)\mathcal{R}_\varepsilon(x-y)J(y)d^4xd^4y.$$

Moreover,

$$Z_{\mathrm{source}}(\varphi,J) := \int_{\mathbb{M}^4} \varphi(x)J(x)\ d^4x.$$

Finally, $\mathcal{L}_{\mathrm{int}}(\varphi) := -\varphi^4$. Hence

$$\mathcal{L}_{\mathrm{int}}\left(\frac{1}{\mathrm{i}}\ \frac{\delta}{\delta J(x)}\right) = -\frac{\delta^4}{\delta J(x)^4}.$$

14.2.4 The Magic Quantum Action Reduction Formula for Correlation Functions

The full correlation functions. Using the extended quantum action functional $Z = Z(J,\varphi)$, we define the quantum action functional $Z = Z(J)$ by setting

$$Z(J) := \frac{Z(J,0)}{Z(0,0)}. \tag{14.37}$$

Explicitly,

$$\boxed{Z(J) = \mathcal{N}\exp\left\{-\kappa\mathrm{i}\int_{\mathbb{M}^4} d^4x\ \frac{\delta^4}{\delta J(x)^4}\right\}\cdot Z_{\mathrm{free}}(J)} \tag{14.38}$$

along with

$$Z_{\text{free}}(J) = \mathrm{e}^{\frac{1}{2}\mathrm{i}\int_{\mathbb{M}^4\times\mathbb{M}^4} J(x)\mathcal{R}_\varepsilon(x-y)J(y)d^4xd^4y}.$$

The normalization constant $\mathcal{N}$ has to be chosen in such a way that we get $Z(0) = 1$. The full correlation functions C_n are defined by the following functional derivatives

$$\boxed{C_n(x_1,\ldots,x_n) := \frac{1}{\mathrm{i}^n}\,\frac{\delta^n Z(J)}{\delta J(x_1)\cdots\delta J(x_n)}|_{\,J=0}.} \tag{14.39}$$

This formula allows us to reduce the computation of the correlation functions $C_1, C_2, \ldots$ to the response function $\mathcal{R}_\varepsilon$. Equivalently, we get

$$Z(J) := 1 + \sum_{n=1}^{\infty} \frac{\mathrm{i}^n}{n!} \int_{\mathbb{M}^{4n}} C_n(x_1,\ldots,x_n)J(x_1)\cdots J(x_n)d^4x_1\cdots d^4x_n.$$

For each $n = 1, 2, \ldots$, the full correlation function C_n is symmetric with respect to the variables $x_1, \ldots, x_n$. This means that the quantum action functional $Z = Z(J)$ represents the generating functional for the family of full correlation functions $C_1, C_2, \ldots$. The operator

$$\exp\Big\{-\mathrm{i}\kappa \int_{\mathbb{M}^4} d^4x\, \frac{\delta^4}{\delta J(x)^4}\Big\}$$

is equal to

$$1 - \mathrm{i}\kappa \int_{\mathbb{M}^4} d^4x\, \frac{\delta^4}{\delta J(x)^4} - \frac{\kappa^2}{2!}\left(\int_{\mathbb{M}^4} d^4x \frac{\delta^4}{\delta J(x)^4}\right)^2 + \ldots$$

This can be written as

$$1 - \mathrm{i}\kappa \int_{\mathbb{M}^4} d^4x\, \frac{\delta^4}{\delta J(x)^4} - \frac{\kappa^2}{2!}\int_{\mathbb{M}^4} d^4x \frac{\delta^4}{\delta J(x)^4}\int_{\mathbb{M}^4} d^4y \frac{\delta^4}{\delta J(y)^4} + \ldots$$

Switching off the interaction by setting $\kappa = 0$, we get $Z(J) = Z_{\text{free}}(J)$. This way we obtain the free correlation functions by

$$C_{n,\text{free}}(x_1,\ldots,x_n) := \frac{1}{\mathrm{i}^n}\,\frac{\delta^n Z_{\text{free}}(J)}{\delta J(x_1)\cdots\delta J(x_n)}|_{\,J=0}.$$

Formula (14.38) allows us to reduce the computation of the full correlation functions $C_1, C_2, \ldots$ to the free correlation functions $C_{1,\text{free}}, C_{2,\text{free}}, \ldots$ In particular, we get $C_n \equiv 0$ for $n = 1, 3, 5 \ldots$

The quantum action axiom. Formula (14.39) is called the magic quantum action reduction formula for the correlation functions of the φ^4-model. In our approach above, we use this formula in order to define the correlation functions. From the physical point of view, we postulate the validity of this

magic formula as an axiom called the quantum action axiom (or the QA axiom) of the φ^4-model.

Motivation. In Sect. 7.24.6 on page 448, we studied the discrete φ^4-model in rigorous terms. The formal passage to infinite dimensions yields the formulas for $Z(\varphi, J)$ and C_n above.

Warning to the reader. Let $\varepsilon > 0$, and let Ω be a fixed nonempty bounded open subset of the 4-dimensional momentum space $\mathbb{M}^4$. The integral

$$\mathcal{R}_\varepsilon(x) = -\frac{1}{(2\pi)^4} \int_{\mathbb{M}^4} \frac{\mathrm{e}^{-\mathrm{i}px}}{p^2 - m_0^2 + \mathrm{i}\varepsilon}\, d^4p, \qquad x \in \mathbb{M}^4$$

for the response function $\mathcal{R}_\varepsilon$ is divergent at infinity. Therefore, let us replace $\mathcal{R}_\varepsilon$ by the cut-off function

$$\mathcal{R}_{\varepsilon,\Omega}(x) := -\frac{1}{(2\pi)^4} \int_{\Omega} \frac{\mathrm{e}^{-\mathrm{i}px}}{p^2 - m_0^2 + \mathrm{i}\varepsilon}\, d^4p, \qquad x \in \mathbb{M}^4.$$

This is a convergent integral. More general, we replace all of the integrals $\int_{\mathbb{M}^4} \dots$ by $\int_\Omega \dots$ This implies that the magic formula (14.39) is well-defined as a formal power series expansion with respect to the coupling constant κ.

This means that, in each order of perturbation theory with respect to the coupling κ, we get well-defined correlation functions.

However, these functions depend on the regularization parameter ε and the cut-off set Ω.

It remains to study the two limits $\varepsilon \to +0$ and $\Omega \to \mathbb{R}^4$.

It turns out that this limit leads to divergent expressions. Therefore, physicists invented the method of renormalization. The idea is

- to add counterterms to the Lagrangian density $\kappa\mathcal{L}_{\mathrm{int}}$ and
- to replace the bare mass m_0 by an renormalized (or effective) mass m_{ren}.

The goal is to get convergent expressions in each order of perturbation theory, as $\varepsilon \to +0$ and $\Omega \to \mathbb{R}^4$. We refer to Sects. 15.3ff on page 847ff.

The experience of physicists shows that one only needs the quantum action axiom and the LSZ axiom below, combined with the procedure of renormalization, in order to successfully compute a large number of physical processes for elementary particles such that the computed values coincide with the values measured in experiments.

Applications to the computation of concrete physical effects will be considered in Volume II and the later volumes.

14.2.5 The Magic LSZ Reduction Formula for the *S*-Matrix

Scattering functions. Parallel to the generating functional $Z(J)$ for the correlation functions, let us define the so-called scattering functional

$$\boxed{\mathsf{S}(\varphi) := \frac{Z(0,\varphi)}{Z(0,0)}} \tag{14.40}$$

by using the extended quantum action functional $Z = Z(J,\varphi)$ from Sect. 14.2.3. The scattering functional is normalized by $\mathsf{S}(0) = 1$. Furthermore, for $n = 1, 2, \ldots$, let us define the scattering functions by setting

$$\mathsf{S}_n(x_1, \ldots, x_n) := \frac{\delta^n \mathsf{S}(\varphi)}{\delta\varphi(x_1) \cdots \delta\varphi(x_n)} \Big|_{\varphi=0}.$$

The scattering functions $\mathsf{S}_n(x_1, \ldots, x_n)$ are symmetric with respect to the variables $x_1, \ldots, x_n$. Equivalently, we get

$$\mathsf{S}(\varphi) := 1 + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathbb{M}^{4n}} \mathsf{S}_n(x_1, \ldots, x_n)\varphi(x_1) \cdots \varphi(x_n) d^4x_1 \cdots d^4x_n.$$

This means that the scattering functional S is the generating functional for the family of scattering functions $\mathsf{S}_1, \mathsf{S}_2, \ldots$.

The LSZ axiom. The magic LSZ reduction formula for the φ^4-model allows us to compute the normalized scattering functions by means of the full correlation functions. Explicitly, for $n = 1, 2, \ldots$,

$$\boxed{S_n(x_1, \ldots, x_n) = \frac{1}{\mathrm{i}^n} \left\{ \prod_{k=1}^{n} (D_{x_k} + \mathrm{i}\varepsilon) \right\} C_n(x_1, \ldots, x_n).} \tag{14.41}$$

Here, we set

$$D_{x_k} := -\Box_{x_k} - m_0^2.$$

Note that the wave operator $\Box_{x_k}$ acts on the variable x_k of the full correlation function C_n.

Motivation. The formula (14.41) is motivated by generalizing Theorem 7.39 on page 450 in a formal manner. We postulate the validity of (14.41) as an axiom called the LSZ axiom.

***S*-matrix elements.** Let us now briefly discuss the physical meaning of the scattering functions. By (14.14), the function

$$u_{\mathbf{p}}(\mathbf{x}, t) := \frac{\mathrm{e}^{\mathrm{i}(\mathbf{p}\mathbf{x} - E_{\mathbf{p}}t)}}{\sqrt{\mathcal{V}}}$$

is a solution of the free Klein–Gordon equation $(\Box + m_0^2)\varphi = 0$. Here, we set

$$E_{\mathbf{p}} := \sqrt{m_0^2 + \mathbf{p}^2}.$$

The normalization factor $\mathcal{V} > 0$ is chosen in such a way that

$$\int_{\mathcal{C}} |u_{\mathbf{p}}(\mathbf{x}, t))|^2 \, d^3\mathbf{x} = \int_{\mathcal{C}} d^3\mathbf{x} = 1$$

where $\mathcal{C}$ is an arbitrary bounded open subset of $\mathbb{R}^3$. In terms of physics, the function $u_{\mathbf{p}}$ describes a homogenous stream of free mesons of particle number density $1/\mathcal{V}$. In other words, there is precisely one particle in the box $\mathcal{C}$ of volume $\mathcal{V}$. Each particle has the momentum vector $\mathbf{p}$ and the energy $E_{\mathbf{p}}$. Now consider a scattering process in the box $\mathcal{C}$ with

- n incoming particle streams of free mesons at remote past, $t = -\infty$, described by the functions

  $$u_{\mathbf{p}_1}, \ldots, u_{\mathbf{p}_n}$$

  of momentum vectors $\mathbf{p}_1, \ldots, \mathbf{p}_n$, respectively.
- After scattering, there appear n outgoing particle streams of free mesons at remote future, $t = +\infty$, described by the functions

  $$u_{\mathbf{p}_{n+1}}, \ldots, u_{\mathbf{p}_{2n}}$$

  of momentum vectors $\mathbf{p}_{n+1}, \ldots, \mathbf{p}_{2n}$, respectively.

We define

$$\boxed{\mathsf{S}_{\mathbf{p}_1,\ldots\mathbf{p}_{2n}} := \int_{\mathbb{R}^{2n}} S_{2n}(x_1, \ldots, x_{2n}) \prod_{k=1}^{2n} u_{\mathbf{p}_k}(x_k) \prod_{k=1}^{2n} d^4x_k.}$$

This is called the S-matrix element of the scattering process. The real number

$$|\mathsf{S}_{\mathbf{p}_1,\ldots\mathbf{p}_{2n}}|^2$$

is the transition probability of the scattering process. This transition probability can be used in order to compute the cross section of the scattering process. This is precisely the quantity which can be measured in scattering experiments. Many applications to concrete physical processes will be considered in the following volumes.

Renormalization. The LSZ quantum action axiom (14.41) tells us how to reduce the scattering functions to the correlation functions. By Sect. 14.2.4, we have to pass to the renormalized correlation functions. Therefore, physicists replace the correlation functions in the magic formula (14.41) by the renormalized correlation functions.

> *This way, physicists get the renormalized scattering functions, and hence the renormalized S-matrix elements.*

14.2.6 The Local Quantum Action Principle

The Dyson–Schwinger equation of the φ^4-model reads as follows:

$$\frac{\delta Z}{\delta J(x)} = \int_{\mathbb{M}^4} \mathcal{R}_\varepsilon(x,y) \left(\mathrm{i} J(x) Z(J) + 4\kappa \mathrm{i} \, \frac{\delta^3 Z(J)}{\delta J(x)^3} \right) d^4x. \tag{14.42}$$

Formally, this yields

$$\boxed{(\Box_x + m_0^2 - \mathrm{i}\varepsilon) \, \frac{\delta Z}{\delta J(x)} = \mathrm{i} J(x) Z(J) + 4\kappa \mathrm{i} \, \frac{\delta^3 Z(J)}{\delta J(x)^3} \, .}$$

These two equations are motivated by Theorem 7.40 on page 452.

14.2.7 The Mnemonic Functional Integral

The global quantum action principle. The formula

$$Z(J) = \mathcal{N} \int_{\mathcal{S}(\mathbb{R}^4)} \mathrm{e}^{\mathrm{i} S[\varphi, J]} \; \mathcal{D}\varphi \tag{14.43}$$

is called the global quantum action principle. The normalization factor $\mathcal{N}$ is uniquely determined by the normalization condition $Z(0) = 1$. Here, we integrate over all possible classical fields φ which lie in the function space $\mathcal{S}(\mathbb{R}^4)$ of rapidly decreasing smooth functions. The action functional is given by (14.15), that is,

$$S[\varphi, J] := \int_{\mathbb{M}^4} \tfrac{1}{2} \varphi(-\Box - m_0^2 + \mathrm{i}\varepsilon)\varphi - \kappa\varphi^4 + \varphi J \; d^4x.$$

Explicitly computing the normalization factor, the global quantum action principle reads as

$$\boxed{Z(J) = \frac{\int_{\mathcal{S}(\mathbb{R}^4)} \mathrm{e}^{\mathrm{i} S[\varphi, J]} \; \mathcal{D}\varphi}{\int_{\mathcal{S}(\mathbb{R}^4)} \mathrm{e}^{\mathrm{i} S[\varphi, 0]} \; \mathcal{D}\varphi}.} \tag{14.44}$$

Let $n = 1, 2, \ldots$. For the correlation functions,

$$C_n(x_1, \ldots, x_n) := \frac{\delta^n Z(J)}{\delta J(x_1) \cdots \delta J(x_n)} \Big|_{J=0},$$

we get

$$C_n(x_1, \ldots, x_n) = \frac{\int_{\mathcal{S}(\mathbb{R}^4)} \varphi(x_1) \cdots \varphi(x_n) \mathrm{e}^{\mathrm{i} S[\varphi, 0]} \; \mathcal{D}\varphi}{\int_{\mathcal{S}(\mathbb{R}^4)} \mathrm{e}^{\mathrm{i} S[\varphi, 0]} \; \mathcal{D}\varphi}.$$

These integrals are to be understood as formal limits of finite-dimensional integrals where the dimension of the integrals goes to infinity. Symbolically,

$$\mathcal{D}\varphi = \prod_{x \in \mathbb{M}^4} d\varphi(x).$$

Formal motivation of the magic formulas. In rigorous terms, we have shown in Sect. 7.25 on page 459 that finite-dimensional functional integrals can be used in order to justify both

- the magic quantum action reduction formula and
- the magic LSZ reduction formula.

In the case of the present continuum φ^4-model, the same arguments can be applied formally to (14.44) in order to get the magic formulas (14.38) and (14.41) from Sects. 14.2.4 and 14.2.5, respectively. Observe that this represents only a formal approach, since the infinite-dimensional functional integral (14.44) is not a well-defined mathematical object.

14.2.8 Bose–Einstein Condensation of Dilute Gases

The φ^4-model as a toy model in particle physics. From the modern point of view of elementary particle physics, the φ^4-model is only a toy model. It describes the interactions of elementary particles in the wrong way. It was discovered by physicists in the 1960s and in the early 1970s that

> *The Standard Model in particle physics has to be based on the idea of gauge field theory.*

Here, in contrast to the φ^4-model, the interactions between elementary particles are not described by self-interactions of the quantum field φ itself, but by additional gauge fields which correspond to particles called gauge bosons (photons, vector bosons W^+, W^-, Z^0, and eight gluons).

- The prototype of the Standard Model is quantum electrodynamics. In Sect. 14.3, we will sketch the basic ideas of quantum electrodynamics. In terms of gauge field theory, quantum electrodynamics refers to the commutative gauge group $U(1)$. A detailed study of quantum electrodynamics will be carried out in Volume II.
- The Standard Model in particle physics refers to the product gauge group $U(1) \times SU(2) \times SU(3)$. This has to be supplemented by the Higgs mechanism which equips the vector bosons W^+, W^-, Z^0 with large masses compared with the proton mass. The point is that the gauge groups $SU(2)$ and $SU(3)$ are not commutative, in contrast to $U(1)$. The prototype of the Standard Model in the form of a $SU(N)$-gauge field theory will be studied in Sect. 16.2 on page 878. In particular, in the case where $N = 3$, we get quantum chromodynamics which describes strong interactions based on quarks and gluons. A detailed study of the Standard Model in particle physics can be found in Volume III (gauge field theory) and Volume V (physics of the Standard Model).

Many physicists believe that the φ^4-model in 4-dimensional space-time $\mathbb{M}^4$ describes only a trivial free quantum field after renormalization. This means that the influence of the interaction term $\kappa\mathcal{L}_{\text{int}}$ vanishes after renormalization. In terms of physics, quantum fluctuations destroy the interaction. Recommendations for further reading can be found on page 871 (triviality of the φ^4-model). It is thinkable that a variant of the φ^4-model on a modified 4-dimensional space-time (in the setting of noncommutative geometry) corresponds to a nontrivial quantum field. This is the subject of recent research.

The φ^4-model as an effective theory for Bose–Einstein condensation. It was conjectured by Gross and Pitaevskii that the φ^4-model can be used in order to describe Bose–Einstein condensation of dilute gases, in the sense of an effective potential. This conjecture was rigorously proven in the fundamental paper by

> E. Lieb and R. Seiringer, Proof of Bose–Einstein condensation for dilute trapped gases, Phys. Rev. Lett. **88** (2002), No. 170409.
> Internet: `http://www.arXiv math-ph/0112032`

14.3 A Glance at Quantum Electrodynamics

Quantum electrodynamics studies the interactions between electrons, their antiparticles (positrons), and photons (gauge bosons). Quantum electrodynamics was created by Heisenberg and Pauli in 1929.[7] The final theory was established by Feynman, Schwinger, and Tomonaga in the 1940s and completed by Dyson in 1949.[8] At this point, let us only sketch a few basic ideas. For a detailed study of concrete physical processes between electrons, positrons and photons, we refer to Volume II.

Notation. In addition to the notations introduced at the beginning of this chapter, we will use the following symbols.

(i) The Dirac–Pauli matrices are given by

$$\gamma^0 := \begin{pmatrix} \sigma^0 & 0 \\ 0 & -\sigma^0 \end{pmatrix}, \quad \gamma^j := \begin{pmatrix} 0 & \sigma^j \\ -\sigma^j & 0 \end{pmatrix}, \qquad j = 1, 2, 3, \tag{14.45}$$

[7] P. Dirac, The quantum theory of the emission and absorption of radiation, Proc. Royal Soc. of London, Series A, Vol. 114 (1927), 243–265.
P. Dirac, The quantum theory of the electron, Proc. Royal Soc. London **A117** (1928), 610–624; **A118**, 351–361.
W. Heisenberg and W. Pauli, On the quantum electrodynamics of wave fields (in German), Z. Phys. **56** (1929), 1–61; **59** (1930), 108–190.

[8] F. Dyson, The radiation theories of Tomonaga, Schwinger, and Feynman, Phys. Rev. **75** (1949), 406–502; The S-matrix in quantum electrodynamics, Phys. Rev. **75** (1949), 1736–1755; The renormalization method in quantum electrodynamics, Proc Royal Soc. **207** (1951), 395–401.
A collection of important papers in quantum electrodynamics can be found in J. Schwinger (Ed.), Quantum Electrodynamics: 34 Selected Articles, Dover, New York, 1958.

along with the Pauli matrices

$$\sigma^0 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma^1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^2 := \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma^3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The Pauli matrices are self-adjoint, that is $(\sigma^\mu)^\dagger = \sigma^\mu$ for $\mu = 0, 1, 2, 3$. Moreover, we have the commutation relations

$$\sigma^1\sigma^2 - \sigma^2\sigma^1 = 2\mathrm{i}\sigma^3, \quad \sigma^2\sigma^3 - \sigma^3\sigma^2 = 2\mathrm{i}\sigma^1, \quad \sigma^3\sigma^1 - \sigma^1\sigma^3 = 2\mathrm{i}\sigma^2,$$

and the anticommutation relations

$$\sigma^r\sigma^s + \sigma^s\sigma^r = 2\delta^{rs} I, \qquad r, s = 1, 2, 3. \tag{14.46}$$

This implies $\mathrm{tr}(\sigma^r\sigma^s) = 2\delta^{rs}$ for $r, s = 1, 2, 3$. Furthermore, we introduce the chiral matrix

$$\gamma^5 := \mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} 0 & \sigma^0 \\ \sigma^0 & 0 \end{pmatrix}.$$

The choice of the Dirac–Pauli matrices is called the standard representation of the following Clifford anticommutation relations:

$$\boxed{\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2\eta^{\mu\nu} I, \qquad \mu, \nu = 0, 1, 2, 3.}$$

The symbol $\eta^{\mu\nu}$ is defined on page 769.

(ii) The four-potential of the electromagnetic field is given by

$$A(x) = (A_0(x), A_1(x), A_2(x), A_3(x)).$$

Each component is a smooth real-valued function $A_\mu : \mathbb{M}^4 \to \mathbb{R}$.

(iii) The electron field ψ has the form of a column matrix

$$\psi(x) := \begin{pmatrix} \psi^1(x) \\ \psi^2(x) \\ \psi^3(x) \\ \psi^4(x) \end{pmatrix}.$$

Here, each component is a smooth complex-valued function $\psi^k : \mathbb{M}^4 \to \mathbb{C}$. Furthermore, the adjoint matrix reads as

$$\psi(x)^\dagger = (\psi^1(x)^\dagger, \psi^2(x)^\dagger, \psi^3(x)^\dagger, \psi^4(x)^\dagger).$$

Following Dirac,[9] we define $\overline{\psi}(x) := \psi(x)^\dagger\gamma^0$.

[9] In mathematics, one frequently uses the symbol $\overline{z}$ for the conjugate complex number to the complex number z. In quantum field theory, the symbol $\overline{\psi}$ is fixed by the Dirac convention which is standard in the literature. Therefore, we use the symbol $z^\dagger$ for the conjugate complex of z.

14.3.1 The Equations of Motion

In quantum electrodynamics, the classical field equations with respect to a fixed inertial system read as follows.

(E1) The Maxwell equations for the electromagnetic field tensor $F_{\alpha\beta}$: For all $\alpha, \beta, \gamma = 0, 1, 2, 3$, we have the Bianchi identity

$$\boxed{\partial_\alpha F_{\beta\gamma} + \partial_\beta F_{\gamma\alpha} + \partial_\gamma F_{\alpha\beta} = 0.} \tag{14.47}$$

Moreover, for all $\beta = 0, 1, 2, 3$, the electromagnetic field tensor is related to the electric current generated by the electron field according to

$$\boxed{\partial_\alpha F^{\alpha\beta} = J^\beta_{\text{QED}}.} \tag{14.48}$$

Explicitly, the electric 4-current density vector is given by

$$J^\beta_{\text{QED}}(x) := -e\overline{\psi}(x)\gamma^\beta\psi(x).$$

Here, $-e$ is the negative electric charge of the electron, and m_e denotes the rest mass of the electron.

(E2) Dirac equation:

$$\boxed{\mathrm{i}\gamma^\alpha\partial_\alpha\psi - m_e\psi = -eA_\alpha\gamma^\alpha\psi.} \tag{14.49}$$

Here, the four-potential $A_\alpha, \alpha = 0, 1, 2, 3$, is related to the electromagnetic field tensor by the equation

$$F_{\alpha\beta} = \partial_\alpha A_\beta - \partial_\beta A_\alpha, \qquad \alpha, \beta = 0, 1, 2, 3.$$

We assume that the equations (i) and (ii) above are valid on $\mathbb{M}^4$. If we introduce the covariant derivative

$$\nabla_\alpha := \partial_\alpha - \mathrm{i}eA_\alpha,$$

and the Feynman dagger symbol $\not{\nabla} := \gamma^\alpha\nabla_\alpha$, then the Dirac equation (E2) can be elegantly written as

$$\boxed{(\mathrm{i}\not{\nabla} - m_e)\psi = 0.}$$

The language of vector calculus. For the electric field vector $\mathbf{E}$ and the magnetic field vector $\mathbf{B}$, set

$$\mathbf{E} := E^k\mathbf{e}_k, \quad \mathbf{B} := B^k\mathbf{e}_k.$$

In addition, let $U := A_0, \mathbf{A} := A^k\mathbf{e}_k$, as well as

$$\varrho := J^0_{\mathrm{QED}}, \qquad \mathbf{J}_{\mathrm{QED}} := J^k_{\mathrm{QED}} \mathbf{e}_k.$$

Here, ϱ is the electric charge density, and $\mathbf{J}_{\mathrm{QED}}$ is the electric current density vector. The equation $F_{\alpha\beta} = \partial_\alpha A_\beta - \partial_\beta A_\alpha$ corresponds to[10]

$$\mathbf{B} := \mathbf{curl}\, \mathbf{A}, \qquad \mathbf{E} := -\,\mathbf{grad}\, U - \dot{\mathbf{A}}$$

along with

$$F_{01} = -E_1 = E^1, \qquad F_{12} = B_3 = -B^3. \tag{14.50}$$

The remaining expressions are obtained by the cyclic permutation $1 \mapsto 2 \mapsto 3 \mapsto 1$. The Maxwell equations (14.47) and (14.48) read as

$$\begin{aligned} \operatorname{div} \mathbf{E} &= \varrho, & \operatorname{div} \mathbf{B} &= 0 \\ \mathbf{curl}\, \mathbf{E} &= -\dot{\mathbf{B}}, & \mathbf{curl}\, \mathbf{B} &= \mathbf{J}_{\mathrm{QED}} + \dot{\mathbf{E}}. \end{aligned}$$

In addition, this implies the continuity equation $\dot{\varrho} + \operatorname{div} \mathbf{J}_{\mathrm{QED}} = 0$.

The importance of quantum fluctuations. The reader should observe that

> *The classical equations of motion in quantum electrodynamics do not know all about the quantum system. It is crucial to include additional quantum fluctuations.*

This is the highly nontrivial process of quantizing the classical theory. This will be studied below. The mathematical difficulties of quantum field theory refer precisely to the quantization of the Standard Model in particle physics.

14.3.2 The Principle of Critical Action

Our goal is to obtain the equations of motion in quantum electrodynamics from a variational principle. Let Ω be a nonempty bounded open subset of the 4-dimensional space-time $\mathbb{M}^4$ (e.g., a 4-dimensional cube). Set

$$S[A, \psi] := \int_\Omega \mathcal{L}\, d^4x$$

along with the Lagrangian density

$$\mathcal{L} = -\tfrac{1}{4} F_{\alpha\beta} F^{\alpha\beta} + \overline{\psi}(\mathrm{i}\nabla_\alpha \gamma^\alpha - m_e)\psi.$$

This can be written as

$$\mathcal{L} = -\tfrac{1}{4} F_{\alpha\beta} F^{\alpha\beta} + \overline{\psi}(\mathrm{i}\partial_\alpha \gamma^\alpha - m_e)\psi + e\mathcal{L}_{\mathrm{int}}.$$

[10] The dot denotes the partial derivative with respect to time t.

The crucial term

$$\boxed{e\mathcal{L}_{\text{int}} := -J^{\alpha}_{\text{QED}} A_{\alpha}} \tag{14.51}$$

describes the interaction between the electromagnetic field A and the electron field ψ. The principle of critical action in quantum electrodynamics reads as follows:

$$\boxed{S[A, \psi] = \text{critical!}} \tag{14.52}$$

along with the boundary condition $A =$ fixed on $\partial\Omega$ and $\psi =$ fixed on $\partial\Omega$. The proof of the following theorem will be given in Problem 14.12.

Theorem 14.4 *Each smooth solution of* (14.52) *satisfies the equations of motion* (E1), (E2) *in quantum electrodynamics given above.*

Relativistic invariance. According to Einstein's principle of special relativity, quantum electrodynamics has to be independent of the choice of the inertial system. This is indeed the case for the equations of motion (E1), (E2) from Sect. 14.3.1. More precisely, for a change of inertial systems, note the following:

- in 1905 Einstein discovered the correct transformation law for space-time coordinates and the electromagnetic field tensor $F_{\alpha\beta}$;
- in 1928 Dirac discovered the correct transformation law for the electron field spinor ψ.

The transformation law for space-time coordinates is given by the Poincaré group, whereas the transformations laws for $F_{\alpha\beta}$ and ψ refer to representations of the Poincaré group. This will be studied in Volume III on gauge theory.

Discrete symmetries. The equations of motion (E1), (E2) are invariant under space reflection, time reflection, and charge conjugation. This will be studied in Volume III in connection with the Standard Model in particle physics.

Gauge symmetry. Finally, note that the equations of motion (E1), (E2) are invariant under the gauge transformation

$$\boxed{\psi^g(x) = \mathrm{e}^{\mathrm{i}\theta(x)}\psi(x), \qquad A^g_{\alpha}(x) := A_{\alpha}(x) + \frac{1}{e}\,\partial_{\alpha}\theta(x)}$$

for all space-time points $x \in \mathbb{M}^4$, and all $\alpha = 0, 1, 2, 3$. Here, the phase function

$$\theta : \mathbb{M}^4 \to \mathbb{R}$$

is assumed to be smooth. In fact, the covariant derivative $\nabla_{\alpha}\psi$ transforms in the same way as the electron field ψ. Explicitly,

$$\boxed{\nabla_\alpha^g \psi^g(x) = \mathrm{e}^{\mathrm{i}\theta(x)} \nabla_\alpha \psi(x).} \tag{14.53}$$

This follows from the product rule $\partial_\alpha \psi^g = \mathrm{i}e^{\mathrm{i}\theta} \partial_\alpha \theta \cdot \psi + \mathrm{e}^{\mathrm{i}\theta} \partial_\alpha \psi$, and hence

$$\nabla_\alpha^g \psi^g = (\partial_\alpha - \mathrm{i}eA^g)\psi^g = \mathrm{e}^{\mathrm{i}\theta}(\partial_\alpha - \mathrm{i}eA_\alpha)\psi.$$

In terms of physics, the gauge transformation (14.53) corresponds to a change of the phase of the electron field ψ.

It is typical for gauge field theory that the phase change of the electron field, $\psi(x) \mapsto e^{\mathrm{i}\theta(x)}\psi(x)$, given by the function $\theta = \theta(x)$ depends on space and time.

14.3.3 The Gauge Field Approach

In quantum field theory, try to fix gauge potentials by using convenient gauge conditions.

Folklore

It turns out that the computation of physical effects in quantum electrodynamics is easier to handle if we pass to a special variant of the equations of motion based on a special choice of the four-potential (also called gauge potential) A_0, A_1, A_2, A_3. This procedure will allow us to use the classical wave equation and to reduce the study of photons, electrons, and positrons to interacting harmonic oscillators (see Vol. II). The Lorentz condition (14.54) below is the prototype of a gauge condition.

Special variant of the equations of motion in quantum electrodynamics. Our goal is to modify the equations of motion (E1), (E2) in quantum electrodynamics on page 791. To this end, recall the 4-current density vector

$$J^\alpha_{\mathrm{QED}}(x) := -e\overline{\psi}(x)\gamma^\alpha \psi(x)$$

and the wave operator $\Box := \partial_\alpha \partial^\alpha = \partial_0^2 - \partial_1^2 - \partial_2^2 - \partial_3^2$. Now choose a fixed function $\lambda : \mathbb{R} \to \mathbb{R}$, and consider the following equations.

(M1) Wave equations for the gauge potential (photon field):

$$\Box A^\alpha(x) + (\lambda(x) - 1)\partial^\alpha \partial_\beta A^\beta(x) = J^\alpha_{\mathrm{QED}}(x), \qquad \alpha = 0, 1, 2, 3.$$

(M2) Dirac equation for the electron field:

$$\mathrm{i}\gamma^\alpha \partial_\alpha \psi(x) - m_e \psi(x) = -eA_\alpha(x)\gamma^\alpha \psi(x).$$

Furthermore, we define $F_{\alpha\beta} := \partial_\alpha A_\beta - \partial_\beta A_\alpha$. The equations above refer to all space-time points $x \in \mathbb{M}^4$.

Proposition 14.5 *Each smooth solution of the system* (M1), (M2) *which satisfies the Lorentz condition*

$$\partial_\alpha A^\alpha(x) = 0 \qquad \text{on} \quad \mathbb{M}^4 \tag{14.54}$$

is a solution of the equations of motion (E1), (E2) *in quantum electrodynamics.*

Proof. The Bianchi identity (14.47) follows immediately from the definition of $F_{\alpha\beta}$, by antisymmetry. Furthermore, by the Lorentz condition,

$$\partial_\alpha F^{\alpha\beta} = \partial_\alpha(\partial^\alpha A^\beta - \partial^\beta A^\alpha) = \Box A^\beta = J^\beta_{\text{QED}}.$$

This is the Maxwell system (14.48). □

In order to simplify the approach as much as possible, physicists prefer to use the gauge function $\lambda(x) := 1$. This is called the Feynman gauge. Generally, we expect that

> *Physical processes in quantum electrodynamics are invariant under gauge transformations.*

This fundamental principle is called the gauge invariance principle. In particular, we expect that the fixing of the gauge potential by some gauge condition does not influence the final physical results that can be measured in physical experiments (e.g., cross sections in scattering processes or the energies of bound states.)

Special variant of the principle of critical action. Again let Ω be a nonempty bounded open subset of the 4-dimensional space-time $\mathbb{M}^4$. Set

$$S_{\text{QED}}[A, \psi] := \int_\Omega \mathcal{L}_{\text{QED}}\, d^4x$$

along with the Lagrangian density

$$\mathcal{L}_{\text{QED}} := \mathcal{L}_{\text{free}} + e\mathcal{L}_{\text{int}}$$

where $e\mathcal{L}_{\text{int}} := -J^\alpha_{\text{QED}} A_\alpha$, and

$$\mathcal{L}_{\text{free}} := \tfrac{1}{2} A_\alpha \Box A^\alpha + \tfrac{1}{2}(1 - \lambda(x))(\partial_\beta A^\beta)^2 + \overline{\psi}(\mathrm{i}\gamma^\alpha \partial_\alpha - m_e)\psi.$$

Here, the smooth function $\lambda : \mathbb{R} \to \mathbb{R}$ is fixed. Now consider the variational problem

$$\boxed{S_{\text{QED}}[\overline{\psi}, A, \psi] = \text{critical!}} \tag{14.55}$$

along with fixed boundary values of A and ψ on $\partial\Omega$.

Theorem 14.6 (i) *Each smooth solution* A, ψ *of* (14.55) *satisfies both the wave equation* (M1) *and the Dirac equation* (M2) *on* Ω.

(ii) *If, in addition, the Lorentz condition* $\partial_\beta A^\beta = 0$ *on* Ω *is fulfilled, then* A, ψ *satisfies the equations of motion* (E1), (E2) *in quantum electrodynamics.*

Proof. Ad (i). Choose smooth test functions $h_\alpha : \Omega \to \mathbb{R}$, $\alpha = 0, 1, 2, 3$, which have compact support, that is, $h_\alpha \in \mathcal{D}(\Omega)$. Define the function

$$\mathcal{J}(\tau) := S_{\mathrm{QED}}[A + \tau h, \psi], \qquad \tau \in \mathbb{R}.$$

The variational problem (14.55) is equivalent to $\dot{\mathcal{J}}(0) = 0$ for all test functions $h_\alpha \in \mathcal{D}(\Omega)$, $\alpha = 0, 1, 2, 3$. Explicitly,

$$\int_\Omega \left(\tfrac{1}{2} h_\alpha \Box A^\alpha + \tfrac{1}{2} A_\alpha \Box h^\alpha + (1 - \lambda(x)) \partial_\beta A^\beta \cdot \partial_\alpha h^\alpha - J^\alpha_{\mathrm{QED}} h_\alpha \right) d^4x = 0.$$

Integration by parts yields

$$\int_\Omega A_\alpha \Box h^\alpha d^4x = \int_\Omega (\Box A_\alpha) h^\alpha \, d^4x.$$

Since $(\Box A_\alpha) h^\alpha = (\Box A^\alpha) h_\alpha$,

$$\int_\Omega \big(\Box A^\alpha - (1 - \lambda(x)) \partial^\alpha \partial_\beta A^\beta - J_{\mathrm{QED}} \big) h_\alpha \, d^4x = 0.$$

By the real variational lemma from Sect. 10.4.1 on page 542,

$$\Box A^\alpha - (1 - \lambda(x)) \partial^\alpha \partial_\beta A^\beta - J^\alpha_{\mathrm{QED}} = 0.$$

This is the wave equation (M1). In order to get the Dirac equation (M2), we set

$$\mathcal{J}(\tau) := S_{\mathrm{QED}}[A, \psi + \tau h], \qquad \tau \in \mathbb{R}$$

where $h^j \in \mathcal{D}(\Omega)$ for $j = 1, 2, 3, 4$. We now argue as in Problem 14.11.

Ad (ii). Use Proposition 14.5. □

Motivation of the action functional in quantum electrodynamics. Let Ω be a bounded open subset of $\mathbb{M}^4$. Choose the Lagrangian density

$$\mathcal{L} := -\tfrac{1}{4} F_{\alpha\beta} F^{\alpha\beta} + \overline{\psi} (\mathrm{i} \gamma^\alpha \nabla_\alpha - m_e) \psi,$$

and consider the principle of critical action

$$\int_\Omega \mathcal{L} \, d^4x = \text{critical!} \tag{14.56}$$

along with given boundary values of A and ψ on $\partial\Omega$ and the side condition

$$\partial_\alpha A^\alpha(x) = 0 \qquad \text{on} \quad \Omega. \tag{14.57}$$

This problem is obtained from the original problem (14.52) by adding the Lorentz condition (14.57). Motivated by the Lagrange multiplier rule, we introduce the modified Lagrangian density

$$\mathcal{L}_\lambda := \mathcal{L} - \tfrac{1}{2} \lambda(x) (\partial_\beta A^\beta(x))^2,$$

and we study the modified unconstrained variational problem

$$\int_\Omega \mathcal{L}_\lambda \, d^4x = \text{critical!} \tag{14.58}$$

along with given boundary values of A and ψ on $\partial\Omega$.

Proposition 14.7 *For given smooth function $\lambda : \mathbb{R} \to \mathbb{R}$, the variational problem* (14.58) *is equivalent to the principle of critical action* (14.55) *in quantum electrodynamics.*

Proof. We have to show that

$$\int_\Omega \mathcal{L}_\lambda d^4x = \int_\Omega \mathcal{L}_{\text{QED}} \, d^4x + \text{const}$$

for fixed boundary values of A and ψ on $\partial\Omega$. Here, the constant depends on the boundary values. In other words, we have to prove that the two Lagrangian densities $\mathcal{L}_\lambda$ and $\mathcal{L}_{\text{QED}}$ only differ by a null Lagrangian density (see Problem 14.5). In fact,

$$\begin{aligned} F_{\alpha\beta}F^{\alpha\beta} &= (\partial_\alpha A_\beta - \partial_\beta A_\alpha)(\partial^\alpha A^\beta - \partial^\beta A^\alpha) \\ &= 2\partial_\alpha A_\beta \cdot \partial^\alpha A^\beta - 2\partial_\alpha A_\beta \cdot \partial^\beta A^\alpha. \end{aligned}$$

Integration by parts yields

$$\int_\Omega F_{\alpha\beta}F^{\alpha\beta} \, d^4x = -2 \int_\Omega A_\beta \partial_\alpha \partial^\alpha A^\beta + \partial_\alpha A^\alpha \cdot \partial_\beta A^\beta \, d^4x + \text{const}$$

where the constant corresponds to the boundary terms. Hence

$$\int_\Omega -\tfrac{1}{4} F_{\alpha\beta}F^{\alpha\beta} \, d^4x = \int_\Omega \tfrac{1}{2} A_\beta \Box A^\beta + \tfrac{1}{2}(\partial_\beta A^\beta)^2 \, d^4x + \text{const}.$$

The other terms of $\mathcal{L}_\lambda$ and $\mathcal{L}_{\text{QED}}$ remain unchanged. □

14.3.4 The Extended Action Functional with Source Term

We now want to apply the response approach to quantum electrodynamics. In fact, we will use a combination of

- the bosonic model from Sect. 7.24 for photons and
- the fermionic model from Sect. 9.7 for electrons and positrons.

To this end, we replace the principle of critical action (14.52) by the extended principle of critical action

$$\boxed{S_{\text{QED,ext}}[A, \psi] = \text{critical!}} \tag{14.59}$$

by adding source terms. We are looking for smooth functions A and ψ whose components lie in the space $\mathcal{S}(\mathbb{R}^4)$. Here, we set

$$S_{\text{QED,ext}} := S_{\text{QED}} + S_{\text{source}} \tag{14.60}$$

with

$$S_{\text{QED}} := \int_{\mathbb{M}^4} (\mathcal{L}_{\text{free}} + e\mathcal{L}_{\text{int}}) d^4x, \qquad S_{\text{source}} := \int_{\mathbb{M}^4} \mathcal{L}_{\text{source}}\, d^4x.$$

The corresponding Lagrangian densities are given by

$$\begin{aligned} \mathcal{L}_{\text{free}} &:= \tfrac{1}{2} A_\alpha \eta^{\alpha\beta} \Box A_\beta + \overline{\psi}(\mathrm{i}\gamma^\alpha \partial_\alpha - m_e)\psi, \\ e\mathcal{L}_{\text{int}} &:= -J^\alpha_{\text{QED}} A_\alpha = e\overline{\psi} A_\alpha \gamma^\alpha \psi, \\ \mathcal{L}_{\text{source}} &:= A_\alpha(x) J^\alpha_A(x) + \overline{\mathcal{J}}(x)\psi(x) + \overline{\psi}(x) J(x). \end{aligned}$$

Observe that we use the Feynman gauge, $\lambda(x) \equiv 1$. The source terms are given by

$$J(x) := \begin{pmatrix} J^1(x) \\ J^2(x) \\ J^3(x) \\ J^4(x) \end{pmatrix}, \qquad \mathcal{J}(x) := \begin{pmatrix} \mathcal{J}^1(x) \\ \mathcal{J}^2(x) \\ \mathcal{J}^3(x) \\ \mathcal{J}^4(x) \end{pmatrix}.$$

Moreover, we set $\overline{\psi} := \psi^\dagger \gamma^0$ and $\overline{\mathcal{J}} := \mathcal{J}^\dagger \gamma^0$. The modulus e of the electric charge of the electron plays the role of the coupling constant. In our energetic system of units introduced on page 768, we have

$$e^2 = 4\pi\alpha = \frac{4\pi}{137.04}$$

where α is the fine structure constant. This means that the dimensionless coupling constant $\kappa = e$ is small in quantum electrodynamics.

Theorem 14.8 *Each smooth solution of the extended principle of critical action* (14.59) *satisfies the following equations.*

(i) *Photon field:*

$$\eta^{\alpha\beta} \Box A_\beta = -J^\alpha_A - e\overline{\psi}\gamma^\alpha \psi, \qquad \alpha = 0, 1, 2, 3.$$

(ii) *Electron field:*

$$(\mathrm{i}\gamma^\alpha \partial_\alpha - m_e)\psi = -J - eA_\alpha \gamma^\alpha \psi.$$

(iii) *Positron field:*

$$\overline{\psi}(\mathrm{i}\gamma^\alpha \partial_\alpha + m_e) = \overline{\mathcal{J}} + e\overline{\psi} A_\alpha \gamma^\alpha.$$

By convention, the differential operator ∂_α acts on the function $\overline{\psi}$ in (iii). The proof proceeds similarly to the proof of Theorem 14.6.

14.3.5 The Response Function for Photons

Let us switch off the interaction between the electromagnetic field and the electron field by setting the coupling constant equal to zero, that is, $e = 0$. From the photon field equation of Theorem 14.8(i) we obtain the following regularized response equation

$$\boxed{\eta^{\alpha\beta}(\Box - \mathrm{i}\varepsilon)A_\beta = -J_A^\alpha, \qquad \alpha = 0,1,2,3}$$

along with the regularizing parameter $\varepsilon > 0$. For the solution, we formally write

$$A_\alpha(x) = \int_{\mathbb{M}^4} \mathcal{R}_{\alpha\beta,\varepsilon}(x-y) J_A^\beta(y) d^4y, \qquad x \in \mathbb{M}^4, \;\; \alpha = 0,1,2,3.$$

The matrix function $(\mathcal{R}_{\alpha\beta,\varepsilon})$ is called the response function for photons. Explicitly,

$$\mathcal{R}_{\alpha\beta,\varepsilon}(x) := -\eta_{\alpha\beta}\mathcal{R}_\varepsilon(x). \tag{14.61}$$

Recall that

$$\mathcal{R}_\varepsilon(x) := -\frac{1}{(2\pi)^4} \int_{\mathbb{M}^4} \frac{\mathrm{e}^{-\mathrm{i}px}}{p^2 + \mathrm{i}\varepsilon}\, d^4p, \qquad x \in \mathbb{M}^4.$$

We also define $\mathcal{D}_{\alpha\beta,\varepsilon}(x) := -\mathrm{i}\mathcal{R}_{\alpha\beta,\varepsilon}(x)$. Hence

$$\mathcal{D}_{\alpha\beta,\varepsilon}(x) = -\frac{\mathrm{i}\eta_{\alpha\beta}}{(2\pi)^4} \int_{\mathbb{M}^4} \frac{\mathrm{e}^{-\mathrm{i}px}}{p^2 + \mathrm{i}\varepsilon}\, d^4p, \qquad x \in \mathbb{M}^4. \tag{14.62}$$

Formal motivation for the response function. It follows from

$$(\Box - \mathrm{i}\varepsilon)A^\alpha = -J_A^\alpha$$

along with (14.18) on page 774 that

$$A^\alpha(x) = -\int_{\mathbb{M}^4} \mathcal{R}_\varepsilon(x-y) J_A^\alpha(y) d^4p.$$

This yields (14.61).

Rigorous definition of the Feynman propagator for photons. Motivated by the formal considerations above, for $\alpha, \beta = 0,1,2,3$, we define the tempered distributions

$$\boxed{\mathsf{D}_F^{\alpha\beta} := -\eta^{\alpha\beta}\mathsf{G}_F.}$$

The matrix $\mathsf{D}_F := (\mathsf{D}_F^{\alpha\beta})$ is called the Feynman propagator for photons. By (14.34) and (14.35) on page 781, we get

$$\Box \mathsf{D}_F^{\alpha\beta} = \mathrm{i}\eta^{\alpha\beta}\delta \tag{14.63}$$

along with

$$\boxed{\mathsf{D}_F^{\alpha\beta} = -\frac{\mathrm{i}\eta^{\alpha\beta}}{(2\pi)^4} \int_{\mathbb{M}^4} \frac{\mathrm{e}^{-\mathrm{i}px}}{p^2 + 0_+\mathrm{i}}\, d^4p, \qquad \alpha, \beta = 0, 1, 2, 3.} \tag{14.64}$$

Finally, we define the tempered distributions $\mathsf{R}^{\alpha\beta} := \mathrm{i}\mathsf{D}_F^{\alpha\beta}$. As desired by the formal motivation (14.62), this definition yields

$$\mathsf{R}^{\alpha\beta} = \frac{\eta^{\alpha\beta}}{(2\pi)^4} \int_{\mathbb{M}^4} \frac{\mathrm{e}^{-\mathrm{i}px}}{p^2 + 0_+\mathrm{i}}\, d^4p.$$

14.3.6 The Response Function for Electrons

Set $e = 0$. From Theorem 14.8(ii) we get the following linearized response equation for the electron:

$$\boxed{(\mathrm{i}\gamma^\alpha \partial_\alpha - m_e)\psi = -J.}$$

We will show below that the formal solution looks like

$$\psi(x) = \int_{\mathbb{M}^4} \mathcal{R}_{\mathrm{el}}(x - y) J(y) d^4y, \qquad x \in \mathbb{M}^4$$

along with the definition of the response function for electrons:

$$\mathcal{R}_{\mathrm{el}}(x) := (\mathrm{i}\gamma^\alpha \partial_\alpha + m_e)\mathcal{R}_\varepsilon(x), \qquad x \in \mathbb{M}^4. \tag{14.65}$$

We also define $\mathcal{S}_F(x) := -\mathrm{i}\mathcal{R}_{\mathrm{el}}(x)$. Formally, this yields

$$\mathcal{S}_F(x) = \frac{\mathrm{i}}{(2\pi)^4} \cdot (\mathrm{i}\gamma^\alpha \partial_\alpha + m_e) \int_{\mathbb{M}^4} \frac{\mathrm{e}^{-\mathrm{i}px}}{p^2 - m_e^2 + \mathrm{i}\varepsilon}\, d^4p.$$

Here, we choose the regularization parameter $\varepsilon > 0$. Hence

$$\mathcal{S}_F(x) = \frac{\mathrm{i}}{(2\pi)^4} \int_{\mathbb{M}^4} \frac{\mathrm{i}\gamma^\alpha p_\alpha + m_e I}{p^2 - m_e^2 + \mathrm{i}\varepsilon} \cdot \mathrm{e}^{-\mathrm{i}px}\, d^4p.$$

Formal motivation of the response function for electrons. First set $\varepsilon = 0$. By (14.21) on page 775,

$$(\Box + m_e^2)\mathcal{R}(x) = \delta^{(4)}(x), \qquad x \in \mathbb{M}^4.$$

For the Dirac–Pauli matrices, $\gamma^\alpha\gamma^\beta + \gamma^\beta\gamma^\alpha = 2\eta^{\alpha\beta} I$. This implies Dirac's magic formula

$$\boxed{(\mathrm{i}\gamma^\alpha \partial_\alpha - m_e)(\mathrm{i}\gamma^\beta \partial_\beta + m_e) = -\Box - m_e^2.} \tag{14.66}$$

In fact, the product $(\mathrm{i}\gamma^\alpha\partial_\alpha - m_e)(\mathrm{i}\gamma^\beta\partial_\beta + m_e)$ is equal to

$$-\tfrac{1}{2}(\gamma^\alpha\gamma^\beta + \gamma^\beta\gamma^\lambda)\partial_\alpha\partial_\beta - m_e^2 = -\eta^{\alpha\beta}\partial_\alpha\partial_\beta - m_e^2 = -\Box - m_e^2.$$

Hence

$$(\mathrm{i}\gamma^\alpha\partial_\alpha - m_e)\mathcal{R}_{\mathrm{el}}(x) = -(\Box + m_e^2)\mathcal{R}(x) = -\delta^{(4)}(x).$$

Secondly, the formal definition (14.65) represents a regularization of the expression obtained by the formal proof.

Rigorous definition of the Feynman propagator for electrons. Motivated by the formal considerations above, we define the tempered distribution

$$\boxed{\mathsf{S}_F := (\mathrm{i}\gamma^\alpha\partial_\alpha + m_e)\mathsf{G}_{F,m_e}.} \tag{14.67}$$

By (14.30) and (14.31) on page 780, this implies

$$(\mathrm{i}\gamma^\alpha\partial_\alpha - m_e)\mathsf{S}_F = \mathrm{i}I\delta$$

along with the Fourier representation

$$\boxed{\mathsf{S}_F = \frac{\mathrm{i}}{(2\pi)^4}\int_{\mathbb{M}^4}\frac{\gamma^\alpha p_\alpha + m_e I}{p^2 - m_e^2 + 0_+\mathrm{i}}\cdot \mathrm{e}^{-\mathrm{i}px}d^4p.} \tag{14.68}$$

14.3.7 The Extended Quantum Action Functional

By definition, the extended quantum action functional of quantum electrodynamics,

$$Z(\overline{\mathcal{J}}, J_A, J; \overline{\psi}, A, \psi) \tag{14.69}$$

is given by

$$\boxed{\exp\left\{\mathrm{i}e\int_{\mathbb{M}^4} d^4x\ \mathcal{L}_{\mathrm{int}}\left(\frac{1}{\mathrm{i}}\frac{\delta}{\delta\overline{\mathcal{J}}(x)}, \frac{1}{\mathrm{i}}\ \frac{\delta}{\delta J_A(x)}, \frac{1}{\mathrm{i}}\frac{\delta}{\delta J(x)}\right)\right\} Z_{\mathrm{free,source}}}$$

along with

$$Z_{\mathrm{free,source}}(\overline{\mathcal{J}}, J_A, J; \overline{\psi}, A, \psi) := Z_{\mathrm{free}}(\overline{\mathcal{J}}, J_A, J)\ \mathrm{e}^{Z_{\mathrm{source}}(\overline{\mathcal{J}}, J_A, J;\ \overline{\psi}, A, \psi)}.$$

Here, we set

$$Z_{\mathrm{free}}(\overline{\mathcal{J}}, J_A, J) := \mathrm{e}^{Z_{\mathrm{free,red}}(\overline{\mathcal{J}}, J_A, J)}$$

where we introduce

$$Z_{\text{free,red}} := \int_{\mathbb{M}^4 \times \mathbb{M}^4} \tfrac{1}{2} J_A^\alpha(x) \mathrm{i}\mathcal{R}_{\alpha\beta,\varepsilon}(x-y) J_A^\beta(y)\, d^4x d^4y + \int_{\mathbb{M}^4 \times \mathbb{M}^4} \overline{\mathcal{J}}(x) \mathrm{i}\mathcal{R}_{\text{el}}(x-y) J(y)\, d^4x d^4y \qquad (14.70)$$

and

$$Z_{\text{source}} := \mathrm{i} \int_{\mathbb{M}^4} \overline{\mathcal{J}}(x)\psi(x) + \overline{\psi}(x) J(x) + J_A^\alpha(x) A_\alpha(x)\, d^4x.$$

This is related to the Feynman propagators for photons and electrons by

$$\mathrm{i}\mathcal{R}_{\alpha\beta,\varepsilon} = -\mathcal{D}_{\alpha\beta,\varepsilon}, \qquad \mathrm{i}\mathcal{R}_{\text{el}} = -\mathcal{S}_F, \qquad (14.71)$$

respectively. For the interaction term, we get

$$\mathcal{L}_{\text{int}}(\overline{\psi}, A, \psi) := \overline{\psi} \gamma^\alpha A_\alpha \psi.$$

This yields

$$\mathcal{L}_{\text{int}} \left(\frac{1}{\mathrm{i}} \frac{\delta}{\delta \overline{\mathcal{J}}(x)}, \frac{1}{\mathrm{i}} \frac{\delta}{\delta J_A(x)}, \frac{1}{\mathrm{i}} \frac{\delta}{\delta J(x)} \right) = \frac{1}{\mathrm{i}^3} \frac{\delta}{\delta \overline{\mathcal{J}}(x)} \cdot \gamma^\alpha \cdot \frac{\delta}{\delta J_A^\alpha(x)} \cdot \frac{\delta}{\delta J(x)}.$$

Bosonic variables. The variables A_0, A_1, A_2, A_3 of the photon field and the components $J_A^0, J_A^1, J_A^2, J_A^3$ of the source J_A are complex numbers.

Fermionic variables. Introduce the matrices

$$\psi = \begin{pmatrix} \psi^1 \\ \psi^2 \\ \psi^3 \\ \psi^4 \end{pmatrix}, \qquad J = \begin{pmatrix} J^1 \\ J^2 \\ J^3 \\ J^4 \end{pmatrix},$$

and

$$\overline{\psi} = (\bar{\psi}^1, \bar{\psi}^2, \bar{\psi}^3, \bar{\psi}^4), \qquad \overline{\mathcal{J}} = (\bar{\mathcal{J}}^1, \bar{\mathcal{J}}^2, \bar{\mathcal{J}}^3, \bar{\mathcal{J}}^4).$$

We assume that the components

$$\psi^1, \psi^2, \psi^3, \psi^4, \bar{\psi}^1, \bar{\psi}^2, \bar{\psi}^3, \bar{\psi}^4, J^1, J^2, J^3, J^4, \bar{\mathcal{J}}^1, \bar{\mathcal{J}}^2, \bar{\mathcal{J}}^3, \bar{\mathcal{J}}^4$$

of the fermionic fields ψ and $\overline{\psi}$ (electron and positron field) and the sources $\overline{\mathcal{J}}$ and J are independent Grassmann variables. This means that if we denote these variables by $\zeta^1, \ldots \zeta^{16}$, then

$$\zeta^k \zeta^m = -\zeta^m \zeta^k, \qquad k, m = 1, \ldots, 16.$$

Note that for Grassmann variables, we have to distinguish between left and right partial functional derivatives. We assume that

$$\frac{\delta}{\delta \overline{\mathcal{J}}(x)} \qquad \left(\text{resp. } \frac{\delta}{\delta J(x)} \right)$$

is a left (resp. right) partial functional derivative.

14.3.8 The Magic Quantum Action Reduction Formula

The quantum action axiom. Using the extended quantum action functional

$$Z(\overline{\mathcal{J}}, J_A, J; \overline{\psi}, A, \psi)$$

from Sect. 14.3.7, we define the quantum action functional in quantum electrodynamics by setting

$$Z(\overline{\mathcal{J}}, J_A, J) := \frac{Z(\overline{\mathcal{J}}, J_A, J; 0, 0, 0)}{Z(0, 0, 0; 0, 0, 0)}.$$

By definition, the correlation functions are the functional derivatives of the generating functional $(\overline{\mathcal{J}}, J_A, J) \mapsto Z(\overline{\mathcal{J}}, J_A, J)$ at vanishing source terms, $\overline{\mathcal{J}} = 0$ and $J_A = 0, J = 0$. In other words, the functional $Z = Z(\overline{\mathcal{J}}, J_A, J)$ is the generating functional for the correlation functions in quantum electrodynamics. For example, switching off the interaction by setting $e = 0$, we get the Feynman propagators for photons and electrons. Explicitly, by (14.70) and (14.71),

$$\mathcal{D}_{\alpha\beta,\varepsilon}(x - y) = \frac{1}{\mathrm{i}^2} \frac{\delta^2 Z_{\text{free}}}{\delta J_A^\alpha(x) \delta J_A^\beta(y)} | \, \overline{\mathcal{J}} = J_A = J = 0, \qquad \alpha, \beta = 0, 1, 2, 3,$$

and

$$\mathcal{S}_F(x - y) = \frac{1}{\mathrm{i}^2} \frac{\delta^2 Z_{\text{free}}}{\delta J(x) \delta \overline{\mathcal{J}}(y)} | \, \overline{\mathcal{J}} = J_A = J = 0.$$

14.3.9 The Magic LSZ Reduction Formula

Using the extended quantum action functional $Z(\overline{\mathcal{J}}; J_A, J; \overline{\psi}, A, \psi)$ from Sect. 14.3.7, the scattering functional $\mathsf{S}(\overline{\psi}, A, \psi)$ for electrons, positrons, and photons in quantum electrodynamics is defined by

$$\mathsf{S}(\overline{\psi}, A, \psi) := \frac{Z(0, 0, 0; \overline{\psi}, A, \psi)}{Z(0, 0, 0; 0, 0, 0)}.$$

The scattering function

$$\mathsf{S}_{n,m,l}(x_1, \ldots, x_n, y_1, \ldots, y_m, z_1, \ldots, z_l)$$

is defined by the functional derivative

$$\frac{\delta^{n+m+l} \mathsf{S}(\overline{\psi}, A, \psi)}{\delta \overline{\psi}(x_1) \cdots \delta \overline{\psi}(x_n) \delta A(y_1) \cdots \delta A(y_m) \delta \psi(z_1) \cdots \delta \psi(z_l)}$$

at the point $\psi = \overline{\psi} = 0, A = 0$. Parallel to Sect. 14.2.5, the S-matrix elements can be constructed by using the scattering functions and the free particle functions. Moreover, as in Theorem 7.48 on page 482, it is possible to relate the scattering functional $\mathsf{S}(\overline{\psi}, A, \psi)$ to the generating functional $Z(\overline{\mathcal{J}}, J_A, J)$ of the correlation functions by the magic LSZ reduction formula. A detailed study of the S-matrix in quantum electrodynamics along with interesting physical applications can be found in Volume II.

14.3.10 The Mnemonic Functional Integral

The basic functional integral of quantum electrodynamics reads as follows:

$$\boxed{Z(\overline{\mathcal{J}}, J_A, J) = \mathcal{N} \int \mathrm{e}^{\mathrm{i}S_{\mathrm{QED,ext}}} \, \mathcal{D}\overline{\psi} \, \mathcal{D}A \, \mathcal{D}\psi.} \tag{14.72}$$

The explicit form of the extended action functional

$$S_{\mathrm{QED,ext}}[\overline{\psi}, A, \psi; \overline{\mathcal{J}}, J_A, J]$$

can be found in (14.60) on page 797. The factor $\mathcal{N}$ has to be chosen in such a way that

$$Z(0, 0, 0) = 1.$$

Concerning the functional integral (14.72), the components of the photon field A have to be integrated over the space $\mathcal{S}(\mathbb{R}^4)$. Furthermore, the components of $\psi, \overline{\psi}, J, \overline{\mathcal{J}}$ represent independent Grassmann variables. Extending the rigorous arguments from Sect. 7.26 on page 477 and Sect. 9.7 on page 519 in a formal manner to infinite dimensions, we obtain

- the magic quantum action formula from Sect. 14.3.8 and
- the magic LSZ formula on the S-matrix in quantum electrodynamics from Sect. 14.3.9.

Therefore, the functional integral (14.72) is called the global quantum action principle of quantum electrodynamics.

> *Let us emphasize again that the functional integral* (14.72) *is only to be understood as a very convenient mnemonic tool, but not as a rigorous mathematical object.*

For a detailed study of the approach to quantum field theory via functional integrals, we refer to Faddeev and Slavnov (1980). In particular, Faddeev and Slavnov compute the S-matrix as the limit of evolution operators by using an operator calculus based on the so-called holomorphic representation, which diagonalizes creation and annihilation operators.

In the next chapter, we will study the operator approach to quantum field theory. We will show there how the S-matrix arises as a quite natural consequence of the Dyson series which is closely related to Lagrange's classical variation-of-the-parameter method.

Problems

The principle of critical action is the most important method in physics in order to derive basic equations in a very effective way.[11] The following problems concern

[11] A compendium of the action principle for classical fields along with Green's functions can be found in Burgess (2002).

important examples. All of the functions are assumed to be smooth if the opposite is not stated explicitly. In addition, the set Ω is tacitly assumed to be nonempty. First recall the following classical definition. For a smooth function $\mathcal{J} : \mathbb{R} \to \mathbb{R}$, the point τ_0 is called a critical point of $\mathcal{J}$ iff $\dot{\mathcal{J}}(\tau_0) = 0$. By definition, the solutions of the problem

$$\mathcal{J}(\tau) = \text{critical!}, \qquad \tau \in \mathbb{R}$$

are precisely the critical points of $\mathcal{J}$. For example, the function $\mathcal{J}(\tau) := \tau^3$ has the unique critical point $\tau_0 = 0$. Let us now generalize this to variational problems. In what follows, we will use the real and the complex variational lemma from Sect. 10.4.1 on page 542.

14.1 *The principle of critical action for the vibrating string.* Let Ω be a bounded open set in $\mathbb{R}^2$. Show that each smooth solution $u : \mathrm{cl}(\Omega) \to \mathbb{R}$ of the variational problem

$$\int_\Omega (u_t^2 - u_x^2 + 2fu) dxdt = \text{critical!} \tag{14.73}$$

with the boundary condition $u =$ fixed on $\partial\Omega$ satisfies the wave equation

$$\boxed{u_{tt} - u_{xx} = f \qquad \text{on} \quad \Omega.}$$

Solution: We argue similarly to the proof of Theorem 10.17 on page 548. Choose a smooth test function $h : \Omega \to \mathbb{R}$ which has compact support, i.e., $h \in \mathcal{D}(\Omega)$. Let u be a solution. Replacing u by $u + \tau h$, we get

$$\mathcal{J}(\tau) := \int_\Omega \{(u_t + \tau h_t)^2 - (u_x + \tau h_x)^2 + 2f(u + \tau h)\}\, dxdt, \qquad \tau \in \mathbb{R}.$$

By definition, the function u is a solution of the variational problem (14.73) iff the function $\mathcal{J}$ has a critical point at $\tau = 0$, for each test function $h \in \mathcal{D}(\Omega)$. Hence $\dot{\mathcal{J}}(0) = 0$. Explicitly,

$$\int_\Omega (u_t h_t - u_x h_x + fh)\, dxdt = 0.$$

Integration by parts yields

$$\int_\Omega (-u_{tt} + u_{xx} + f)h\, dxdt = 0$$

for all $h \in \mathcal{D}(\Omega)$. By the real variational lemma from Sect. 10.4.1, we get $-u_{tt} + u_{xx} + f = 0$ on Ω.

14.2 *General Lagrangian.* Consider the Lagrangian density

$$\mathcal{L} = \mathcal{L}(u, u_x, u_y, x, y).$$

We are given a bounded open subset Ω of $\mathbb{R}^2$. Show that each smooth solution $u : \mathrm{cl}(\Omega) \to \mathbb{R}$ of the variational problem

$$\int_\Omega \mathcal{L}(u(x,y), u_x(x,y), u_y(x,y), x, y)\, dxdy = \text{critical!}$$

with the boundary condition $u =$ fixed on $\partial\Omega$ satisfies the following Euler–Lagrange equation

$$\boxed{\frac{\partial}{\partial x}\mathcal{L}_{u_x} + \frac{\partial}{\partial y}\mathcal{L}_{u_y} = \mathcal{L}_u \qquad \text{on} \quad \Omega.} \tag{14.74}$$

Solution: Choose a test function $h \in \mathcal{D}(\Omega)$. Let u be a solution. Replacing u by $u + \tau h$, we get

$$\mathcal{J}(\tau) := \int_\Omega \mathcal{L}(u + \tau h, u_x + \tau h_x, u_y + \tau h_y, x, y)\, dxdy.$$

By definition, u is a solution of the variational problem iff $\dot{\mathcal{J}}(0) = 0$, for each $h \in \mathcal{D}(\Omega)$. Hence

$$\int_\Omega (\mathcal{L}_u h + \mathcal{L}_{u_x} h_x + \mathcal{L}_{u_y} h_y) dxdy = 0.$$

Integration by parts yields

$$\int_\Omega \left(\mathcal{L}_u - \frac{\partial}{\partial x}\mathcal{L}_{u_x} - \frac{\partial}{\partial y}\mathcal{L}_{u_y} \right) h\, dxdy = 0.$$

Finally, use the variational lemma from Sect. 10.4.1.

14.3 *Lagrangian depending on several field functions.* Replace the Lagrangian density from Problem 14.2 by

$$\mathcal{L} = \mathcal{L}(u, u_x, u_y; v, v_x, v_y, x, y).$$

Show that the Euler–Lagrange equation (14.74) has to be replaced by the system

$$\boxed{\frac{\partial}{\partial x}\mathcal{L}_{u_x} + \frac{\partial}{\partial y}\mathcal{L}_{u_y} = \mathcal{L}_u, \qquad \frac{\partial}{\partial x}\mathcal{L}_{v_x} + \frac{\partial}{\partial y}\mathcal{L}_{v_y} = \mathcal{L}_v \quad \text{on} \quad \Omega.}$$

Solution: Fix the function v (resp. u) and apply the argument from the preceding Problem 14.2 to u (resp. v).

14.4 *The principle of critical action for the real nonlinear Klein–Gordon equation.* We are given a bounded open subset Ω of $\mathbb{R}^4$. Let $F : \mathbb{R} \to \mathbb{R}$ be a smooth function. Show that each smooth solution $u : \mathrm{cl}(\Omega) \to \mathbb{R}$ of the variational problem

$$\int_\Omega (u_t^2 - u_x^2 - u_y^2 - u_z^2 + 2F(u))\, dxdydzdt = \text{critical!}$$

with the boundary condition $u =$ fixed on $\partial\Omega$ satisfies the equation

$$\boxed{u_{tt} - u_{xx} - u_{yy} - u_{zz} = F'(u) \qquad \text{on} \quad \Omega.}$$

Solution: Argue as in Problem 14.2. Each solution u satisfies the Euler–Lagrange equation

$$\frac{\partial}{\partial t}\mathcal{L}_{u_t} + \frac{\partial}{\partial x}\mathcal{L}_{u_x} + \frac{\partial}{\partial y}\mathcal{L}_{u_y} + \frac{\partial}{\partial z}\mathcal{L}_{u_z} = \mathcal{L}_u \qquad \text{on} \quad \Omega.$$

This yields the claim.

14.5 *Equivalent Lagrangian densities.* Modify the variational problem from Problem 14.2 by considering

$$\int_\Omega \mathcal{L}(u, u_x, u_y, x, y) + \mathcal{L}_0(u, u_x, u_y, x, y)\ dxdy = \text{critical!} \qquad (14.75)$$

with the boundary condition $u =$ fixed on $\partial\Omega$. We assume that the additional Lagrangian density $\mathcal{L}_0$ is a null Lagarangian density, that is,

$$\int_\Omega \mathcal{L}_0(u(x,y), u_x(x,y), u_y(x,y), x, y)\ dxdy = \text{const}$$

for all smooth functions $u : \text{cl}(\Omega) \to \mathbb{R}$ with $u =$ fixed on $\partial\Omega$. Show that each smooth solution $u = u(x,y)$ of problem (14.75) satisfies the Euler–Lagrange equation (14.74) which is independent of the choice of the null Lagrangian density $\mathcal{L}_0$. Give examples for null Lagrangian densities.
Solution: Note that each solution of (14.75) is also a solution of the corresponding problem with $\mathcal{L}_0 \equiv 0$. Moreover, for given smooth functions $A, B : \mathbb{R}^3 \to \mathbb{R}$, the function

$$\mathcal{L}_0 := A_x(x,y,u) + A_u(x,y,u)u_x + B_y(x,y,u) + B_u(x,y,u)u_y$$

is a null Lagrangian density. In fact, substituting the function $u = u(x,y)$ into $\mathcal{L}_0$, we get

$$\mathcal{L}_0 := \frac{\partial}{\partial x}A(x,y,u(x,y)) + \frac{\partial}{\partial y}B(x,y,u(x,y)).$$

Integration by parts yields

$$\int_\Omega \mathcal{L}_0 dxdy = \int_{\partial\Omega} A(x,y,u(x,y))dy - B(x,y,u(x,y))dx.$$

This is a boundary integral which only depends on the values of the function u on the boundary $\partial\Omega$. These boundary values of u are fixed.

14.6 *The Euler–Lagrange equation for complex fields.* Choose $-\infty < t_0 < t_1 < \infty$. Let $\varphi : [t_0, t_1] \to \mathbb{C}$ be a complex-valued smooth function which is a solution of the variational problem

$$\int_{t_0}^{t_1} L(\varphi(t), \dot\varphi(t), \varphi(t)^\dagger, \dot\varphi(t)^\dagger, t)\ dt = \text{critical!} \qquad (14.76)$$

with fixed boundary values $\varphi(t_0) = \varphi_0$ and $\varphi(t_1) = \varphi_1$. Show that the function φ satisfies the following Euler–Lagrange equations

$$\boxed{\frac{d}{dt}L_{\dot\varphi} = L_\varphi, \qquad \frac{d}{dt}L_{\dot\varphi^\dagger} = L_{\varphi^\dagger} \qquad \text{on} \quad [t_0, t_1].}$$

Solution: Choose a smooth function $h :]t_0, t_1[\to \mathbb{C}$ which has compact support, that is, $h \in \mathcal{D}(t_0, t_1)$. Let φ be a solution. Replacing φ by $\varphi + \tau h$, we get

$$\mathcal{J}(\tau) := \int_{t_0}^{t_1} L(\varphi + \tau h, \ldots)dt, \qquad \tau \in \mathbb{R}.$$

By definition, the function φ is a solution of the variational problem iff $\dot{\mathcal{J}}(0) = 0$ for all $h \in \mathcal{D}(t_0, t_1)$. Hence

$$\int_\Omega (L_\varphi h + L_{\dot\varphi}\dot h + L_{\varphi^\dagger} h^\dagger + L_{\dot\varphi^\dagger}\dot h^\dagger)\, dt = 0.$$

Integration by parts yields

$$\int_\Omega \left(L_\varphi - \frac{d}{dt}L_{\dot\varphi}\right) h + \left(L_{\varphi^\dagger} - \frac{d}{dt}L_{\dot\varphi^\dagger}\right) h^\dagger \, dt = 0.$$

By the complex variational lemma from Sect. 10.4.1,

$$L_\varphi - \frac{d}{dt}L_{\dot\varphi} = 0, \qquad L_{\varphi^\dagger} - \frac{d}{dt}L_{\dot\varphi^\dagger} = 0.$$

14.7 *General Lagrangian for complex-valued fields.* Consider the Lagrangian density

$$\mathcal{L} = \mathcal{L}(\psi, \psi_x, \psi_y, \psi^\dagger, \psi_x^\dagger, \psi_y^\dagger, x, y).$$

We are given the bounded open subset Ω of $\mathbb{R}^2$. Show that each smooth solution $\psi : \mathrm{cl}(\Omega) \to \mathbb{C}$ of the variational problem

$$\int_\Omega \mathcal{L}\, dxdy = \text{critical!}$$

with the boundary condition $\psi =$ fixed on $\partial\Omega$ satisfies the equations

$$\frac{\partial}{\partial x}\mathcal{L}_{\psi_x} + \frac{\partial}{\partial y}\mathcal{L}_{\psi_y} = \mathcal{L}_\psi \qquad \text{on} \quad \Omega$$

and

$$\frac{\partial}{\partial x}\mathcal{L}_{\psi_x^\dagger} + \frac{\partial}{\partial y}\mathcal{L}_{\psi_y^\dagger} = \mathcal{L}_{\psi^\dagger} \qquad \text{on} \quad \Omega.$$

Solution: Argue as in Problem 14.2 by using the complex variational lemma.

14.8 *The principle of critical action for the complex nonlinear Klein–Gordon equation.* Choose $n = 2, 3, ...$ and a real coupling constant κ. Let Ω be a bounded open set in $\mathbb{R}^4$. Show that each smooth solution $\psi : \mathrm{cl}(\Omega) \to \mathbb{C}$ of the variational problem

$$\int_\Omega (|\psi_t|^2 - |\psi_x|^2 - |\psi_y|^2 - |\psi_z|^2 + \kappa|\psi|^n))\, dxdydzdt = \text{critical!}$$

with the boundary condition $\psi =$ fixed on $\partial\Omega$ satisfies the equations

$$\boxed{\psi_{tt} - \psi_{xx} - \psi_{yy} - \psi_{zz} = \kappa n \psi |\psi|^{n-1} \qquad \text{on} \quad \Omega} \tag{14.77}$$

and

$$\psi_{tt}^\dagger - \psi_{xx}^\dagger - \psi_{yy}^\dagger - \psi_{zz}^\dagger = \kappa n \psi^\dagger |\psi|^{n-1} \qquad \text{on} \quad \Omega. \tag{14.78}$$

Observe that (14.78) is the conjugate-complex equation to (14.77).
Solution: Set

$$\mathcal{L} := \psi_t \psi_t^\dagger - \psi_x \psi_x^\dagger - \psi_y \psi_y^\dagger - \psi_z \psi_z^\dagger + \kappa(\psi\psi^\dagger)^n.$$

By Problem 14.7, the solution ψ satisfies the equations

$$\frac{\partial}{\partial t}\mathcal{L}_{\psi_t^\dagger} + \frac{\partial}{\partial x}\mathcal{L}_{\psi_x^\dagger} + \frac{\partial}{\partial y}\mathcal{L}_{\psi_y^\dagger} + \frac{\partial}{\partial z}\mathcal{L}_{\psi_z^\dagger} = \mathcal{L}_{\psi^\dagger} \qquad \text{on} \quad \Omega,$$

and

$$\frac{\partial}{\partial t}\mathcal{L}_{\psi_t} + \frac{\partial}{\partial x}\mathcal{L}_{\psi_x} + \frac{\partial}{\partial y}\mathcal{L}_{\psi_y} + \frac{\partial}{\partial z}\mathcal{L}_{\psi_z} = \mathcal{L}_\psi \qquad \text{on} \quad \Omega.$$

14.9 *The principle of critical action for the stationary Schrödinger equation.* Let Ω be a bounded open interval on the real line. We are given the continuous real-valued function $U : \mathrm{cl}(\Omega) \to \mathbb{R}$, and the positive mass m. Introduce the Lagrangian density

$$\mathcal{L} := \frac{\varphi_x \varphi_x^\dagger}{2m} + U\varphi\varphi^\dagger.$$

Show that each smooth solution of the variational problem

$$\int_\Omega \mathcal{L}\, dx = \text{critical!}$$

with fixed values of φ on the boundary $\partial\Omega$ satisfies the stationary Schrödinger equation

$$\boxed{-\frac{\varphi_{xx}}{2m} + U\varphi = 0 \qquad \text{on} \quad \Omega,}$$

and the conjugate-complex equation

$$-\frac{\varphi_{xx}^\dagger}{2m} + U\varphi^\dagger = 0 \qquad \text{on} \quad \Omega.$$

Solution: By Problem 14.7, the Euler–Lagrange equations read as

$$\frac{\partial}{\partial x}\mathcal{L}_{\varphi_x^\dagger} = \mathcal{L}_{\varphi^\dagger}, \qquad \frac{\partial}{\partial x}\mathcal{L}_{\varphi_x} = \mathcal{L}_{\varphi}.$$

The first equation coincides with the stationary Schrödinger equation, whereas the second equation is the conjugate-complex equation to the stationary Schrödinger equation.

14.10 *The principle of critical action for the Maxwell equations in classical electrodynamics.* Let Ω be a bounded open subset of $\mathbb{R}^4$. We are given the continuous functions $J^0, J^1, J^2, J^3 : \mathrm{cl}(\Omega) \to \mathbb{R}$. Introduce the Lagrangian density

$$\mathcal{L}_{\mathrm{Maxwell}} := -\tfrac{1}{4} F_{\alpha\beta} F^{\alpha\beta} - J^\alpha A_\alpha$$

where the components of the electromagnetic field tensor $F_{\alpha\beta}$are given by

$$F_{\alpha\beta} := \partial_\alpha A_\beta - \partial_\beta A_\alpha \tag{14.79}$$

in terms of the four-potential A_α, $\alpha = 0, 1, 2, 3$. Show that each smooth solution $A_0, A_1, A_2, A_3 : \mathrm{cl}(\Omega) \to \mathbb{R}$ of the variational problem

$$\int_\Omega \mathcal{L}_{\mathrm{Maxwell}}\, d^4x = \text{critical!}$$

with fixed values of A_0, A_1, A_2, A_3 on the boundary $\partial\Omega$ satisfies the Maxwell equations

$$\boxed{\partial_\alpha F^{\alpha\beta} = J^\beta \qquad \text{on} \quad \Omega, \qquad \beta = 0, 1, 2, 3.}$$

In addition, we have the Bianchi identity

$$\partial_\alpha F_{\beta\gamma} + \partial_\beta F_{\gamma\alpha} + \partial_\gamma F_{\alpha\beta} = 0, \qquad \alpha, \beta, \gamma = 0, 1, 2, 3$$

which follows immediately from (14.79), by antisymmetry.

Solution: Choose smooth test functions $h_\alpha : \mathrm{cl}(\Omega) \to \mathbb{R}$ which have compact support on Ω, $\alpha = 0, 1, 2, 3$. Replacing A_α by $A_\alpha + \tau h_\alpha$, we get

$$\mathcal{J}(\tau) := \int_\Omega -\tfrac{1}{4}(\partial_\alpha A_\beta + \tau\partial_\alpha h_\beta - \partial_\beta A_\alpha - \tau\partial_\beta h_\alpha) \times(\partial^\alpha A^\beta + \tau\partial^\alpha h^\beta - \partial^\beta A^\alpha - \tau\partial^\beta h^\alpha) - J^\beta(A_\beta + \tau h_\beta)\ d^4x$$

for $\tau \in \mathbb{R}$. Suppose that A_0, A_1, A_2, A_3 is a solution of the variational problem. Then $\dot{\mathcal{J}}(0) = 0$. Hence

$$\int_\Omega \{-\tfrac{1}{4}(\partial_\alpha h_\beta - \partial_\beta h_\alpha)F^{\alpha\beta} - \tfrac{1}{4}F_{\alpha\beta}(\partial^\alpha h^\beta - \partial^\beta h^\alpha) - J^\beta h_\beta\}\ d^4x = 0.$$

It follows from $F_{\alpha\beta} = -F_{\beta\alpha}$ and $F_{\alpha\beta}h^{\alpha\beta} = F^{\alpha\beta}h_{\alpha\beta}$ that

$$\int_\Omega \{-F^{\alpha\beta}\partial_\alpha h_\beta - J^\beta h_\beta\}\ d^4x = 0.$$

Integration by parts yields

$$\int_\Omega (\partial_\alpha F^{\alpha\beta} - J^\beta)h_\beta\ d^4x = 0.$$

By the real variational lemma from Sect. 10.4.1, $\partial_\alpha F^{\alpha\beta} - J^\beta = 0$.

14.11 *The principle of critical action for the Dirac equation for the relativistic electron.* Let Ω be a bounded open subset of $\mathbb{R}^4$. We are given the continuous functions $A_0, A_1, A_2, A_3 : \mathrm{cl}(\Omega) \to \mathbb{R}$. Introduce the Lagrangian density

$$\mathcal{L}_{\mathrm{Dirac}} := \overline{\psi}(\mathrm{i}\gamma^\alpha\nabla_\alpha - m_e)\psi$$

with the covariant derivative $\nabla_\alpha := \partial_\alpha - \mathrm{i}eA_\alpha$, and $\nabla_\alpha^- := \partial_\alpha + \mathrm{i}eA_\alpha$. Recall the definition of the Dirac adjoint $\overline{\psi} := \psi^\dagger\gamma^0$. Show that each smooth solution $\psi : \mathrm{cl}(\Omega) \to \mathbb{C}^4$ of the variational problem

$$\int_\Omega \mathcal{L}_{\mathrm{Dirac}}\ d^4x = \text{critical!}$$

along with the boundary condition $\psi =$ fixed on $\partial\Omega$ satisfies the Dirac equation

$$\boxed{\mathrm{i}\gamma^\alpha\nabla_\alpha\psi = m_e\psi \qquad \text{on} \quad \Omega,} \tag{14.80}$$

and the adjoint equation

$$-\mathrm{i}\nabla_\alpha^-\ \overline{\psi}\gamma^\alpha = m_e\overline{\psi} \qquad \text{on} \quad \Omega. \tag{14.81}$$

Prove that (14.81) is a consequence of (14.80).

Solution: Choose smooth test functions $h^\mu : \mathrm{cl}(\Omega) \to \mathbb{C}$ which have compact support on Ω, i.e., $h^\mu \in \mathcal{D}(\Omega)$ for $\mu = 0, 1, 2, 3$. Replacing ψ^μ by $\psi^\mu + \tau h^\mu$, we get

$$\mathcal{J}(\tau) := \int_\Omega \mathcal{L}(\psi + \tau h, \psi^\dagger + \tau h^\dagger)\ d^4x, \qquad \tau \in \mathbb{R}.$$

Suppose that ψ is a solution of the variational problem. Then $\dot{\mathcal{J}}(0) = 0$. Hence

$$\int_\Omega \{h^\dagger\gamma^0(\mathrm{i}\gamma^\alpha\nabla_\alpha - m_e)\psi + \overline{\psi}(\mathrm{i}\gamma^\alpha\nabla_\alpha - m_e)h\}\ d^4x = 0.$$

Integration by parts yields

$$\int_\Omega \{h^\dagger \gamma^0 (\mathrm{i}\gamma^\alpha \nabla_\alpha - m_e)\psi + (-\mathrm{i}\partial_\alpha \overline{\psi}\gamma^\alpha + eA_\alpha \overline{\psi}\gamma^\alpha - \overline{\psi} m_e)h\}\, d^4x = 0.$$

By the complex variational lemma from Sect. 10.4.1,

$$\gamma^0(\mathrm{i}\gamma^\alpha \nabla_\alpha - m_e)\psi = 0, \qquad -\mathrm{i}\partial_\alpha \overline{\psi}\gamma^\alpha + eA_\alpha \overline{\psi}\gamma^\alpha - \overline{\psi} m_e = 0.$$

Since the inverse matrix to γ^0 exists, we get the Dirac equation (14.80) and its adjoint equation (14.81). Finally, let us show that (14.81) is a consequence of (14.80). In fact, equation (14.80) reads as

$$\gamma^\alpha(\mathrm{i}\partial_\alpha \psi + eA_\alpha)\psi = m_e \psi.$$

Applying the operator † to this equation,

$$(-\mathrm{i}\partial_\alpha \psi^\dagger + eA_\alpha \psi^\dagger)\gamma^{\alpha\dagger} = m_e \psi^\dagger.$$

Multiplying this by γ^0 from the right and using

$$\gamma^{0\dagger} = \gamma^0, \qquad \gamma^{j\dagger} = -\gamma^j, \qquad \gamma^j\gamma^0 = -\gamma^j\gamma^0, \qquad j = 1, 2, 3,$$

we obtain

$$(-\mathrm{i}\partial_\alpha \psi^\dagger \gamma^0 + eA_\alpha \psi^\dagger \gamma^0)\gamma^\alpha = m_e \psi^\dagger \gamma^0.$$

This is (14.81).

14.12 *The equations of motion in quantum electrodynamics.* Let Ω be a bounded open subset of $\mathbb{R}^4$. Introduce the Lagrangian density

$$\mathcal{L} = -\tfrac{1}{4} F_{\alpha\beta} F^{\alpha\beta} + \overline{\psi}(\mathrm{i}\partial_\alpha \gamma^\alpha - m_e)\psi + e\mathcal{L}_{\mathrm{int}}$$

with $e\mathcal{L}_{\mathrm{int}} := -J^\alpha_{\mathrm{QED}} A_\alpha$ and $J^\alpha_{\mathrm{QED}} := -e\overline{\psi}\gamma^\alpha\psi$. Show that each smooth solution tuple $A_0, A_1, A_2, A_3 : \mathrm{cl}(\Omega) \to \mathbb{R}$ and $\psi : \mathrm{cl}(\Omega) \to \mathbb{C}^4$ of the variational problem

$$\int_\Omega \mathcal{L}\, d^4x = \text{critical!}$$

with fixed boundary values of A_0, A_1, A_2, A_3 and ψ satisfies the Maxwell–Dirac system

$$\partial_\alpha F^{\alpha\beta} = J^\beta_{\mathrm{QED}}, \qquad \mathrm{i}\gamma^\alpha \nabla_\alpha \psi = m_e \psi \qquad \text{on} \quad \Omega, \qquad \beta = 0, 1, 2, 3.$$

Solution: Use Problems 14.10 and 14.11. Note that $\overline{\psi}(\mathrm{i}\partial_\alpha \gamma^\alpha - m_e)\psi + e\mathcal{L}_{\mathrm{int}}$ is equal to $\overline{\psi}(\mathrm{i}\nabla_\alpha \gamma^\alpha - m_e)\psi$.

14.13 *Proof of Theorem* 14.3. Hint: See Bogoliubov, Logunov, and Todorov (1975), p. 71. One has to use some classical integrals which can be found in Gradshtein and Ryshik (1980).

Further important applications of the principle of critical action to modern physics can be found in Volume III on gauge field theory in physics:

- geodesics, Einstein's theory of general relativity, and the Hilbert action,
- the Ginzburg–Landau equation, the Higgs particle, and the Standard Model in particle physics,
- minimal surfaces, harmonic maps, and string theory,
- the Yang–Mills equation and the Chern–Simons equation,
- the Seiberg–Witten equation,
- supersymmetry and the Wess–Zumino model.

15. The Operator Approach

> Dyson's magic formula for the S-matrix represents a far-reaching generalization of Lagrange's variation-of-the-parameter method in celestial mechanics.
>
> Folklore

In Chap. 14, we have described the approach to quantum field theory which can be traced back to Feynman's approach in the 1940s based on the Feynman rules for Feynman diagrams and the representation of propagators by functional integrals. Typically, this approach does not use operators in Hilbert spaces, that is, the methods of functional analysis do not play any role. Historically, in the 1920s quantum mechanics was first based on operator theory by Heisenberg, Born, Jordan, Dirac, Pauli, and von Neumann. In order to understand Feynman's very effective approach, Dyson related this to operator theory via the magic Dyson formula for the S-matrix.[1] Conceptually, the advantage of operator theory is that the duality between particles and waves is formulated in a very transparent manner.

- The waves appear as solutions of classical wave equations. These equations arise as equations of motion from the classical principle of critical action.
- The particles appear after introducing creation and annihilation operators.
- The free quantum field is a linear combination of creation and annihilation operators where the coefficients are classical wave functions (that is, solutions of the free equations of motion).

The disadvantage of operator theory is the fact that there arise serious mathematical difficulties in applying the rigorous theory of functional analysis to quantum electrodynamics and the Standard Model in particle physics. These difficulties are caused by the interactions which are related to nonlinearities.

Basic strategy. Let us describe the main steps of the approach in this chapter. In what follows, we will use

[1] For the exciting history of Dyson's discovery, see the quotation on page 27.

- a finite number of creation and annihilation operators,
- the quantized finite Fourier series for the free quantum field (Fourier quantization),
- the computation of the 2-point Green's function of the free quantum field via Cauchy's residue theorem,
- the magic Dyson formula for the S-matrix as an axiom,
- the Wick theorem in order to compute S-matrix elements and to represent them as Feynman diagrams which are very close to physical intuition.

Finally, we will show how one can compute cross sections of scattering processes by means of S-matrix elements. In order to get rigorous formulas in each order of perturbation theory, we put the system in a box of finite volume, we consider a finite time interval, and we use a finite lattice in 4-dimensional space (i.e., energy space and 3-dimensional momentum space).

This is the best approach to scattering processes in quantum field theory from the point of view of physical intuition.

This approach has to be complemented by the method of renormalization. This means that we have to study the limit where

- the box goes to $\mathbb{R}^3$,
- the finite time interval goes to $\mathbb{R}$,
- the finite energy interval goes to $\mathbb{R}$, and
- the finite lattice in 4-dimensional momentum space goes to $\mathbb{R}^4$.

Explicitly, we have to add counterterms in order force the existence of this limit.

System of units. In this chapter, we will use the energetic system of units with $\hbar = c = 1$.

15.1 The φ^4-Model

Quantum field theory studies the creation and annihilation of particles.
Folklore

Let us again consider the nonlinear Klein–Gordon equation

$$\boxed{\Box\varphi + m_0^2\varphi + 4\kappa\varphi^3 = 0} \tag{15.1}$$

with the coupling constant $\kappa \geq 0$ and $\mathcal{L}_{\text{int}}(\varphi) := -\varphi^4$. This equation describes an uncharged meson of rest mass $m_0 > 0$. Setting $\kappa = 0$, we get the linearized equation

$$\Box\varphi + m_0^2\varphi = 0 \tag{15.2}$$

which is called the Klein–Gordon equation.

15.1.1 The Lattice Approximation

Be wise and discretize.
Folklore

In what follows we will use the notation for lattices introduced in Sect. 12.1.1 on page 669.

- We put the quantum field in a box $\mathcal{C}(L)$ in 3-dimensional Euclidean position space. The box has side length L and volume $\mathcal{V} = L^3$; the position vectors of $\mathcal{C}(L)$ have the form

$$\mathbf{x} = x^1\mathbf{i} + x^2\mathbf{j} + x^3\mathbf{k}, \qquad x^1, x^2, x^3 \in [-\tfrac{L}{2}, \tfrac{L}{2}]$$

  where $\mathbf{i}, \mathbf{j}, \mathbf{k}$ is a right-handed orthonormal system of vectors.
- We observe the quantum field during the time interval $[-\frac{T}{2}, \frac{T}{2}]$ of length $T > 0$.
- The set $\mathcal{C}(L) \times [-\frac{T}{2}, \frac{T}{2}]$ lies in the 4-dimensional space-time $\mathbb{M}^4$.
- The truncated lattice $\mathcal{G}(N)$ in the 3-dimensional momentum space consists of all the momentum vectors

$$\mathbf{p} = (m^1\mathbf{i} + m^2\mathbf{j} + m^3\mathbf{k}) \cdot \Delta p$$

  where m^1, m^2, m^3 are integers with $|m^1|, |m^2|, |m^3| \le N$ and $\Delta p := 2\pi/L$.

Finally, recall that the discrete Dirac delta function of the lattice $\mathcal{G}(N)$ is given by[2]

$$\boxed{\delta_{\mathcal{G}(N)}(\mathbf{p}) := \frac{\delta_{\mathbf{p},\mathbf{0}}}{\Delta^3\mathbf{p}}, \qquad \mathbf{p} \in \mathcal{G}(N)}$$

where

$$\Delta^3\mathbf{p} := (\Delta p)^3 = \frac{(2\pi)^3}{L^3} = \frac{(2\pi)^3}{\mathcal{V}}.$$

Hence

$$\delta_{\mathcal{G}(N)}(\mathbf{p} - \mathbf{q}) = \frac{\delta_{\mathbf{p},\mathbf{q}}}{\Delta^3\mathbf{p}}, \qquad \mathbf{p}, \mathbf{q} \in \mathcal{G}(N).$$

We also introduce the truncated Dirac delta function in the energy space,

$$\delta_T(E) := \frac{1}{2\pi} \int_{-T/2}^{T/2} \mathrm{e}^{\mathrm{i}Et}\, dt.$$

For all $\mathbf{p}, \mathbf{q} \in \mathcal{G}(N)$, we have the orthogonality relation

$$\frac{1}{\mathcal{V}} \int_{\mathcal{C}(L)} \mathrm{e}^{\mathrm{i}(\mathbf{p}-\mathbf{q})\mathbf{x}}\, d^3\mathbf{x} = \delta_{\mathbf{p},\mathbf{q}}. \tag{15.3}$$

We also introduce the 4-dimensional discrete Dirac delta function

[2] Note that $\delta_{\mathbf{p},\mathbf{q}} := 1$ if $\mathbf{p} = \mathbf{q}$, and $\delta_{\mathbf{p},\mathbf{q}} := 0$ if $\mathbf{p} \neq \mathbf{q}$.

$$\boxed{\delta_{\rm dis}(p) := \delta_{\mathcal{G}(N)}(\mathbf{p})\delta_T(p_0)}$$

where $p = (p_0, \mathbf{p})$ with $\mathbf{p} \in \mathcal{G}(N)$ and $p_0 \in \mathbb{R}$. By (15.3),

$$\delta_{\rm disc}(p) = \frac{1}{(2\pi)^4} \int_{-T/2}^{T/2} dt \int_{\mathcal{C}(L)} d^3\mathbf{x}\, \mathrm{e}^{\mathrm{i}px}. \tag{15.4}$$

Note that $\delta_{\rm dis}(-p) = \delta_{\rm dis}(p)$ for all p.

Key relation for scattering theory. By Sect. 12.1.2 on page 670, for all 4-vectors p,

$$\boxed{\delta_{\rm dis}(p)^2 = \frac{\mathcal{V}T}{(2\pi)^4} \cdot \delta_{\rm dis}(p)(1 + o(1)), \qquad T \to +\infty.} \tag{15.5}$$

Thus, for large time T, we can use the approximation

$$\delta_{\rm dis}(p)^2 = \frac{\mathcal{V}T}{(2\pi)^4} \cdot \delta_{\rm dis}(p).$$

We will frequently use this formula when computing cross sections of scattering processes. Mnemonically, physicists write

$$(\delta^4(p))^2 = \frac{\mathcal{V}T}{(2\pi)^4} \cdot \delta^4(p).$$

Action functional. The action functional on the cube $\mathcal{C}(L)$ is given by

$$S[\varphi] := \int_{\mathcal{C}(L)} -\tfrac{1}{2}\varphi(\Box + m_0^2)\varphi + \kappa\mathcal{L}_{\rm int}(\varphi)\; d^3\mathbf{x}$$

for all smooth functions $\varphi : \mathcal{C}(L) \to \mathbb{C}$ which have the period L with respect to the three position variables x^1, x^2, x^3. Here, $\mathcal{L}_{\rm int}(\varphi) := -\varphi^4$.

Energy. By Sect. 13.8.2 on page 760, the energy of the field φ in the cube $\mathcal{C}(L)$ at time t is equal to

$$H(t) = \int_{\mathcal{C}(L)} \tfrac{1}{2}\dot\varphi(\mathbf{x},t)^2 + \tfrac{1}{2}(\boldsymbol{\partial}\varphi(\mathbf{x},t))^2 + \tfrac{1}{2}m_0^2\varphi(\mathbf{x},t)^2 + \kappa\varphi(\mathbf{x},t)^4\; d^3\mathbf{x}.$$

Truncation of energy. For the length of the momentum vectors of the lattice, we have

$$|\mathbf{p}| \le P_{\max} \qquad \text{for all} \quad \mathbf{p} \in \mathcal{G}(N)$$

where we set $P_{\max} := 3N$ if $L \ge 2\pi$. This yields

$$E_{\max} := \sqrt{m_0^2 + P_{\max}^2} = \sqrt{m_0^2 + 9N^2}$$

which is an upper bound for the energy of a single particle.

15.1.2 Fourier Quantization

The idea of Fourier quantization is to consider solutions of the Klein–Gordon equation in the form of finite Fourier series and to replace the Fourier coefficients by creation and annihilation operators. This way we obtain free quantum fields. Later on, we will use the free quantum field in order to construct the S-matrix operator $\mathsf{S}(T)$ by using Dyson's magic formula. The operator S allows us to compute transition probabilities for scattering processes.

Classical solution of the linearized equation of motion. The following finite sum

$$\boxed{\varphi_{\text{free}}(\mathbf{x},t) = \sum_{\mathbf{p}\in\mathcal{G}(N)} (a_{\mathbf{p}} \mathrm{e}^{\mathrm{i}\mathbf{p}\mathbf{x}}\, \mathrm{e}^{-\mathrm{i}E_{\mathbf{p}}t} + a_{\mathbf{p}}^{\dagger} \mathrm{e}^{-\mathrm{i}\mathbf{p}\mathbf{x}}\, \mathrm{e}^{\mathrm{i}E_{\mathbf{p}}t}) \mathcal{N}_{\mathbf{p}}} \tag{15.6}$$

with the normalization constant

$$\mathcal{N}_{\mathbf{p}} := \frac{1}{\sqrt{2E_{\mathbf{p}}\mathcal{V}}}$$

is a real solution of the Klein–Gordon equation (15.2) which has the period L with respect to x^1, x^2, x^3. Physically, this solution represents a superposition of mesons of momentum $\mathbf{p}$ and energy $E_{\mathbf{p}} := \sqrt{m_0^2 + \mathbf{p}^2}$. We will show below that our choice of the normalization constant $\mathcal{N}_{\mathbf{p}}$ yields the right energy operator of the quantized field.

Quantized solution. We now replace the complex-valued Fourier coefficients $a_{\mathbf{p}}, a_{\mathbf{p}}^{\dagger}$ by operators which satisfy the following crucial commutation relations[3]

$$[a_{\mathbf{p}}, a_{\mathbf{q}}^{\dagger}]_- = \delta_{\mathbf{p},\mathbf{q}} I \qquad \text{for all} \quad \mathbf{p},\mathbf{q} \in \mathcal{G}(N) \tag{15.7}$$

along with

$$[a_{\mathbf{p}}, a_{\mathbf{q}}]_- = 0, \qquad [a_{\mathbf{p}}^{\dagger}, a_{\mathbf{q}}^{\dagger}]_- = 0 \qquad \text{for all} \quad \mathbf{p},\mathbf{q} \in \mathcal{G}(N).$$

Recall that $[A,B]_- := AB - BA$. Explicitly, we get

$$\boxed{a_{\mathbf{p}} a_{\mathbf{q}}^{\dagger} = a_{\mathbf{q}}^{\dagger} a_{\mathbf{p}} + \delta_{\mathbf{p},\mathbf{q}} I, \quad a_{\mathbf{p}} a_{\mathbf{q}} = a_{\mathbf{q}} a_{\mathbf{p}}, \quad a_{\mathbf{p}}^{\dagger} a_{\mathbf{q}}^{\dagger} = a_{\mathbf{q}}^{\dagger} a_{\mathbf{p}}^{\dagger}.} \tag{15.8}$$

We also postulate the existence of a state Φ_0 such that $\langle \Phi_0 | \Phi_0 \rangle = 1$ and

$$\boxed{a_{\mathbf{p}} \Phi_0 = 0 \qquad \text{for all} \quad \mathbf{p} \in \mathcal{G}(N).} \tag{15.9}$$

After this replacement, the operator-valued function φ_{free} from (15.6) describes a free quantum field of meson particles. As we will motivate below,

[3] Recall that the symbol I denotes the identity operator.

- the state Φ_0 is called ground state (or vacuum state) of the free quantum field,
- the operator $a_{\mathbf{p}}$ is called annihilation operator, and
- the adjoint operator $a_{\mathbf{p}}^\dagger$ is called creation operator.

The properties of these operators will be studied thoroughly in Volume II. Such operators arose in the early days of quantum mechanics in connection with the quantization of systems of uncoupled harmonic oscillators. It follows from (15.8) and (15.9) that

$$\langle a_{\mathbf{p}}^\dagger \Phi_0 | a_{\mathbf{p}}^\dagger \Phi_0 \rangle = \langle \Phi_0 | a_{\mathbf{p}} a_{\mathbf{p}}^\dagger \Phi_0 \rangle = \langle \Phi_0 | (I + a_{\mathbf{p}}^\dagger a_{\mathbf{p}}) \Phi_0 \rangle = \langle \Phi_0 | \Phi_0 \rangle = 1.$$

The energy operator. Motivated by Sect. 15.1.1, the energy operator of the free quantum field φ_{free} is given by

$$H := \int_{\mathcal{C}(L)} \tfrac{1}{2} \dot{\varphi}_{\text{free}}(\mathbf{x}, t)^2 + \tfrac{1}{2} (\boldsymbol{\partial} \varphi_{\text{free}}(\mathbf{x}, t))^2 + \tfrac{1}{2} m_0^2 \varphi_{\text{free}}(\mathbf{x}, t)^2 \, d^3\mathbf{x}.$$

Let us show that

$$\boxed{H = \sum_{\mathbf{p} \in \mathcal{G}(N)} E_{\mathbf{p}} a_{\mathbf{p}}^\dagger a_{\mathbf{p}} + \tfrac{1}{2} E_{\mathbf{p}} I.} \tag{15.10}$$

In fact, for the time derivative,

$$\dot{\varphi}_{\text{free}}(\mathbf{x}, t) = \mathrm{i} \sum_{\mathbf{p} \in \mathcal{G}(N)} E_{\mathbf{p}} (-a_{\mathbf{p}} \mathrm{e}^{\mathrm{i}\mathbf{p}\mathbf{x}} \, \mathrm{e}^{-\mathrm{i}E_{\mathbf{p}} t} + a_{\mathbf{p}}^\dagger \mathrm{e}^{-\mathrm{i}\mathbf{p}\mathbf{x}} \, \mathrm{e}^{\mathrm{i}E_{\mathbf{p}} t}) \mathcal{N}_{\mathbf{p}}.$$

Similarly, for the derivative with respect to the position vector $\mathbf{x}$,

$$\boldsymbol{\partial} \varphi_{\text{free}}(\mathbf{x}, t) = \mathrm{i} \sum_{\mathbf{p} \in \mathcal{G}(N)} \mathbf{p} (a_{\mathbf{p}} \mathrm{e}^{\mathrm{i}\mathbf{p}\mathbf{x}} \, \mathrm{e}^{-\mathrm{i}E_{\mathbf{p}} t} - a_{\mathbf{p}}^\dagger \mathrm{e}^{-\mathrm{i}\mathbf{p}\mathbf{x}} \, \mathrm{e}^{\mathrm{i}E_{\mathbf{p}} t}) \mathcal{N}_{\mathbf{p}}.$$

By the orthogonality relation (15.3),

$$H = -\frac{1}{2} \sum_{\mathbf{p} \in \mathcal{G}(N)} \left(a_{\mathbf{p}} a_{-\mathbf{p}} \mathrm{e}^{-2E_{\mathbf{p}} t} + a_{\mathbf{p}}^\dagger a_{-\mathbf{p}}^\dagger \mathrm{e}^{2E_{\mathbf{p}} t} \right) \mathcal{V} \mathcal{N}_{\mathbf{p}}^2 (E_{\mathbf{p}}^2 - \mathbf{p}^2 - m_0^2)$$
$$+ \frac{1}{2} \sum_{\mathbf{p} \in \mathcal{G}(N)} (a_{\mathbf{p}} a_{\mathbf{p}}^\dagger + a_{\mathbf{p}}^\dagger a_{\mathbf{p}}) \cdot \frac{E_{\mathbf{p}}^2 + \mathbf{p}^2 + m_0^2}{2E_{\mathbf{p}}}.$$

Since $E_{\mathbf{p}}^2 = \mathbf{p}^2 + m_0^2$ and $a_{\mathbf{p}} a_{\mathbf{p}}^\dagger = a_{\mathbf{p}}^\dagger a_{\mathbf{p}} + I$, we get the claim (15.10).

The vacuum energy. For the mean energy of the quantum field in a vacuum, it follows from $a_{\mathbf{p}} \Phi_0 = 0$ that

$$\boxed{\overline{E}_{\text{vacuum}} = \langle \Phi_0 | H \Phi_0 \rangle = \frac{1}{2} \sum_{\mathbf{p} \in \mathcal{G}(N)} E_{\mathbf{p}}.}$$

Observe that the mean vacuum energy $\overline{E}_{\text{vacuum}}$ becomes infinite if $N \to +\infty$. That is, if the number of grid points in momentum space goes to infinity. This fact is one of the reasons for the appearance of infinities in quantum field theory.

The Stone–von Neumann theorem. Note the following crucial fact.

It is dangerous to postulate the existence of mathematical objects which have prescribed properties.

The point is that such objects may not exist. To illustrate this, let us first show that the commutation relations for creation and annihilation operators postulated above cannot be realized in a finite-dimensional[4] Hilbert space X. To this end, assume that there are linear operators $a_{\mathbf{p}_j} : X \to X$ which satisfy the commutation relations

$$a_{\mathbf{p}_j} a^\dagger_{\mathbf{p}_k} - a^\dagger_{\mathbf{p}_k} a_{\mathbf{p}_j} = \alpha \delta_{jk} I, \qquad j,k = 1,\dots,n$$

where α is a nonzero complex number. Computing the trace,

$$\alpha \delta_{jk} \operatorname{tr}(I) = \operatorname{tr}(a_{\mathbf{p}_j} a^\dagger_{\mathbf{p}_k}) - \operatorname{tr}(a^\dagger_{\mathbf{p}_k} a_{\mathbf{p}_j}) = 0.$$

This implies $\alpha = 0$, a contradiction.

However, we will show in Volume II that there exists a realization of the finite number of creation and annihilation operators $a_{\mathbf{p}}, a^\dagger_{\mathbf{p}}$ indexed by $\mathbf{p} \in \mathcal{G}(N)$ which satisfies the commutation relations postulated above. These operators are linear and unbounded operators on some infinite-dimensional Hilbert space, and they are uniquely determined up to unitary equivalence. This is the content of the famous Stone–von Neumann theorem.[5]

It was shown by Gårding and Wightman in 1954 that the uniqueness result fails for an infinite number of creation and annihilation operators.[6] This is one of the typical mathematical difficulties in quantum field theory. To circumvent this difficulty, we will use

- the lattice approximation with a finite number of creation and annihilation operators
- combined with a passage to the continuum limit (see Sect. 15.1.8).

If, in contrast to the lattice approach, one starts with the continuum model, then one has to fix the representation of the infinite family of creation and annihilation operators. However, since there exist unitarily inequivalent representations of the commutation relations, there is an element of arbitrariness.

[4] We exclude the trivial space $X = \{0\}$.

[5] J. von Neumann, The uniqueness of the Schrödinger operators (in German), Math. Ann. **104** (1931), 570–578. See also D. Kastler, The C^*-algebra of a free boson field, Commun. Math. Phys. **1** (1965), 14–48.

[6] L. Gårding and A. Wightman, Representations of the commutation relations, Proc. Natl. Acad. Sci. U.S.A. **40**, 622–625. See also I. Gelfand and N. Vilenkin, Generalized Functions, Vol. IV, Sect. 4.5, Academic Press, New York, 1964.

15.1.3 The Free 2-Point Green's Function

Using the free meson quantum field φ from (15.6), we define

$$\boxed{G_{2,\mathcal{G}(N),\mathrm{free}}(\mathbf{x},t;\mathbf{y},s) := \langle 0|\mathcal{T}(\varphi_{\mathrm{free}}(\mathbf{x},t)\varphi_{\mathrm{free}}(\mathbf{y},s))|0\rangle.}$$

This function is called the free 2-point Green's function of the meson quantum field φ_{free} with respect to the lattice $\mathcal{G}(N)$ in momentum space. Let us set $p = (\mathbf{p}, p_0)$. The definition of the chronological operator $\mathcal{T}$ can be found on page 744.

Theorem 15.1 *For all space-time points $x = (\mathbf{x}, t)$ and $y = (\mathbf{y}, s)$ in $\mathbb{M}^4$ with $t \neq s$ and regularization parameter $\varepsilon > 0$,*

$$G_{2,\mathcal{G}(N),\mathrm{free}}(x,y) = \lim_{\varepsilon\to+0} \frac{\mathrm{i}}{(2\pi)^4} \int_{-\infty}^{\infty} dp_0 \sum_{\mathbf{p}\in\mathcal{G}(N)} \frac{e^{-\mathrm{i}p(x-y)}}{p^2 - m_0^2 + \mathrm{i}\varepsilon}\, \Delta^3\mathbf{p}.$$

Proof. (I) Commutation relation. First let $t > s$. Then

$$G_{2,\mathcal{G}(N),\mathrm{free}}(x,y) = \langle \Phi_0|\varphi_{\mathrm{free}}(\mathbf{x},t)\varphi_{\mathrm{free}}(\mathbf{y},s)\Phi_0\rangle.$$

Since $a_{\mathbf{p}}\Phi_0 = 0$, we get $\langle \Phi_0|a_{\mathbf{p}}^\dagger\Phi\rangle = \langle a_{\mathbf{p}}\Phi_0|\Phi\rangle = 0$. Hence

$$\langle \Phi_0|a_{\mathbf{p}}a_{\mathbf{q}}\Phi_0\rangle = \langle \Phi_0|a_{\mathbf{p}}^\dagger a_{\mathbf{q}}\Phi_0\rangle = \langle \Phi_0|a_{\mathbf{p}}^\dagger a_{\mathbf{q}}^\dagger\Phi_0\rangle = 0.$$

Furthermore, it follows from $a_{\mathbf{p}}a_{\mathbf{q}}^\dagger = \delta_{\mathbf{p},\mathbf{q}} I + a_{\mathbf{q}}^\dagger a_{\mathbf{p}}$ and $\langle \Phi_0|\Phi_0\rangle = 1$ that

$$\langle \Phi_0|a_{\mathbf{p}}a_{\mathbf{q}}^\dagger\Phi_0\rangle = \delta_{\mathbf{p},\mathbf{q}}.$$

Thus, by (15.6) along with $\Delta^3\mathbf{p} = (2\pi)^3/\mathcal{V}$,

$$G_{2,\mathcal{G}(N),\mathrm{free}}(x,y) = \frac{1}{2(2\pi)^3} \sum_{\mathbf{p}\in\mathcal{G}(N)} \frac{\mathrm{e}^{\mathrm{i}\mathbf{p}(\mathbf{x}-\mathbf{y})}\, \mathrm{e}^{-\mathrm{i}E_{\mathbf{p}}(t-s)}}{E_{\mathbf{p}}}\, \Delta^3\mathbf{p}.$$

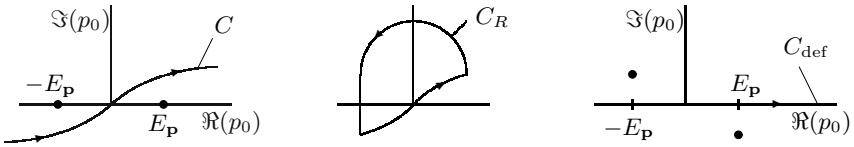
(II) Secondly, let $t < s$. Then

$$G_{2,\mathcal{G}(N),\mathrm{free}}(x,y) = \langle \Phi_0|\varphi_{\mathrm{free}}(\mathbf{y},s)\varphi_{\mathrm{free}}(\mathbf{x},t)\Phi_0\rangle.$$

Interchanging x with y and replacing $\mathbf{p}$ by $-\mathbf{p}$, we get

$$G_{2,\mathcal{G}(N),\mathrm{free}}(x,y) = \frac{1}{2(2\pi)^3} \sum_{\mathbf{p}\in\mathcal{G}(N)} \frac{\mathrm{e}^{\mathrm{i}\mathbf{p}(\mathbf{x}-\mathbf{y})}\, \mathrm{e}^{-\mathrm{i}E_{\mathbf{p}}(s-t)}}{E_{\mathbf{p}}}\, \Delta^3\mathbf{p}.$$

(III) Cauchy's residue theorem. Let $t < s$. Using the curve C pictured in Fig. 15.1(a), we claim that



(a) (b) (c)

Fig. 15.1. Contour of integration

$$\frac{1}{2\pi\mathrm{i}}\int_C \frac{\mathrm{e}^{-\mathrm{i}p_0(t-s)}}{(p_0-E_{\mathbf{p}})(p_0+E_{\mathbf{p}})}\,dp_0 = -\frac{\mathrm{e}^{\mathrm{i}E_{\mathbf{p}}(t-s)}}{2E_{\mathbf{p}}}.$$

To prove this, consider the closed curve C_R pictured in Fig. 15.1(b). Applying Cauchy's residue theorem, we get

$$\frac{1}{2\pi\mathrm{i}}\int_{C_R} \frac{\mathrm{e}^{-\mathrm{i}p_0(t-s)}}{(p_0-E_{\mathbf{p}})(p_0+E_{\mathbf{p}})}\,dp_0 = -\frac{\mathrm{e}^{\mathrm{i}E_{\mathbf{p}}(t-s)}}{2E_{\mathbf{p}}}.$$

Here, the decisive role is played by the pole at the point $p_0 = -E_{\mathbf{p}}$. Finally, letting $R \to +\infty$, the contribution coming from the semicircle of radius R in the upper half-plane goes to zero. In this connection, note that

$$|\mathrm{e}^{-\mathrm{i}p_0(t-s)}| = \mathrm{e}^{-(s-t)\Im(p_0)}$$

where $s-t>0$ and $\Im(p_0)>0$ in the open upper half-plane.

If $t>s$, then we replace the upper semi-circle by a lower semi-circle. This way we get

$$\frac{1}{2\pi\mathrm{i}}\int_C \frac{\mathrm{e}^{-\mathrm{i}p_0(t-s)}}{(p_0-E_{\mathbf{p}})(p_0+E_{\mathbf{p}})}\,dp_0 = -\frac{\mathrm{e}^{-\mathrm{i}E_{\mathbf{p}}(t-s)}}{2E_{\mathbf{p}}}.$$

The contribution to this integral comes from the pole at the point $p_0 = E_{\mathbf{p}}$. Observe that

$$p_0^2 - E_{\mathbf{p}}^2 = p_0^2 - \mathbf{p}^2 - m_0^2 = p^2 - m_0^2.$$

Moreover, $p(x-y) = p_0(t-s) - \mathbf{p}(\mathbf{x}-\mathbf{y})$. Summarizing, we obtain

$$G_{2,\mathcal{G}(N),\mathrm{free}}(x,y) = \frac{\mathrm{i}}{(2\pi)^4}\int_C dp_0 \sum_{\mathbf{p}\in\mathcal{G}(N)} \frac{\mathrm{e}^{-\mathrm{i}p(x-y)}}{p^2-m_0^2}\,\Delta^3\mathbf{p}. \qquad (15.11)$$

(IV) Deformation of the integration contour. The equation $p_0^2 - E_{\mathbf{p}}^2 + \mathrm{i}\varepsilon = 0$ has the two zeros

$$p_0 = \pm E_{\mathbf{p}}\sqrt{1-\frac{\mathrm{i}\varepsilon}{E_{\mathbf{p}}^2}} = \pm E_{\mathbf{p}}\left(1-\frac{\mathrm{i}\varepsilon}{2E_{\mathbf{p}}^2}\right) + o(\varepsilon), \quad \varepsilon \to +0.$$

Consequently, the zero $E_{\mathbf{p}}$ (resp. $-E_{\mathbf{p}}$) moves to the lower (resp. upper) half-plane (Fig. 15.1(c)). Now replace $p^2 - m_0^2$ by $p^2 - m_0^2 + \mathrm{i}\varepsilon$ and the contour C by the real axis. Then the modified integral

$$\frac{\mathrm{i}}{(2\pi)^4} \int_{-\infty}^{\infty} dp_0 \sum_{\mathbf{p}\in\mathcal{G}(N)} \frac{\mathrm{e}^{-\mathrm{i}p(x-y)}}{p^2 - m_0^2 + \mathrm{i}\varepsilon}\, \Delta^3\mathbf{p}$$

can be computed parallel to (III) via Cauchy's residue theorem, and this integral converges to the integral (15.11) as $\varepsilon \to +0$.

□

The truncated Green's function. It is convenient to use the replacement

$$\int_{-\infty}^{\infty} dp_0 \Rightarrow \int_{-E_{\max}}^{E_{\max}} dp_0$$

for the energy integration. This way we obtain the truncated Green's function $G_{2,\mathcal{G}(N),\mathrm{free}}$ used in Table 15.1 on page 831. To simplify notation, we do not introduce a new symbol.

15.1.4 The Magic Dyson Formula for the *S*-Matrix

The S-matrix knows all about scattering processes of elementary particles.
Folklore

The Dyson formula for the S-matrix operator on the finite time-interval $[-\frac{T}{2}, \frac{T}{2}]$ reads as follows:

$$\boxed{\mathsf{S}(T) := \mathcal{T} \exp\left\{-\mathrm{i}\kappa \int_{-T/2}^{T/2} dt : L_{\mathrm{int}}(\varphi_{\mathrm{free}})(t) :\right\}.} \tag{15.12}$$

Here, we set

$$L_{\mathrm{int}}(\varphi_{\mathrm{free}})(t) := \int_{\mathcal{C}(L)} \mathcal{L}_{\mathrm{int}}(\varphi_{\mathrm{free}}(\mathbf{x}, t))\, d^3\mathbf{x}$$

where $\mathcal{L}_{\mathrm{int}}(\varphi) = -\varphi^4$. For the definition of the chronological operator $\mathcal{T}$, we refer to page 744. The free quantum field φ_{free} can be found in (15.6) on page 817. In particular, L_{int} is a polynomial with respect to creation and annihilation operators. The symbol

$$: L_{\mathrm{int}}(\varphi_{\mathrm{free}})(t) :$$

denotes the normal product of $L_{\mathrm{int}}(\varphi_{\mathrm{free}})(t)$. By definition, the normal product is obtained by reordering the terms in such a way that all of the annihilation operators stand on the right of the creation operators. For example,

$$: a_{\mathbf{p}} a_{\mathbf{q}}^\dagger : \; = a_{\mathbf{q}}^\dagger a_{\mathbf{p}}, \qquad : a_{\mathbf{q}}^\dagger a_{\mathbf{p}} : \; = a_{\mathbf{q}}^\dagger a_{\mathbf{p}}, \tag{15.13}$$

and $: a_{\mathbf{p}} a_{\mathbf{q}}^\dagger a_{\mathbf{r}} : \; = a_{\mathbf{q}}^\dagger a_{\mathbf{p}} a_{\mathbf{r}}$. Normal products will be studied in Sect. 15.1.5 below. Observe the following crucial fact.

The S-matrix operator describes interactions of the quantum field in powers of the coupling constant by using nonlinear terms which depend on the known free quantum field.

This way, in the setting of perturbation theory, interactions can be reduced to the interaction-free situation. Formula (15.12) is motivated by the rigorous finite-dimensional formula (7.53) on page 390 which follows from Lagrange's variation-of-the-parameter method.

In the present case, we will use formula (15.12) *as the definition for the crucial S-matrix operator* $\mathsf{S}(T)$.

The Dyson series. Explicitly, formula (15.12) reads as

$$\mathsf{S}(T) = I - \mathrm{i}\kappa \int_{-T/2}^{T/2} dt : L_{\mathrm{int}}(\varphi_{\mathrm{free}})(t) : \tag{15.14}$$
$$+\frac{(-\mathrm{i}\kappa)^2}{2!} \int_{-T/2}^{T/2} dt_1 \int_{-T/2}^{T/2} dt_2 \, \mathcal{T}\{: L_{\mathrm{int}}(\varphi_{\mathrm{free}})(t_1) :: L_{\mathrm{int}}(\varphi_{\mathrm{free}})(t_2) :\} + \dots$$

This is the famous Dyson series in quantum field theory.

Transition probabilities. Let Φ and Ψ be two normalized states. This means that $\langle \Phi | \Phi \rangle = \langle \Psi | \Psi \rangle = 1$. The complex number

$$\langle \Psi |\, \mathsf{S}(T) \Phi \rangle$$

is called the transition amplitude from the state Φ to the state Ψ during the time interval $[-\frac{T}{2}, \frac{T}{2}]$. Furthermore, the nonnegative number

$$\boxed{|\langle \Psi |\, \mathsf{S}(T) \Phi \rangle|^2}$$

is called the transition probability from the state Φ to the state Ψ during the time interval $[-\frac{T}{2}, \frac{T}{2}]$. In what follows, we will discuss

- how to compute transition amplitudes, and hence
- transition probabilities.

We will reduce this to the computation of vacuum expectation values.

Basic tricks for computing vacuum expectation values. The following relations are crucial.

(R1) For the ground state, $\langle \Phi_0 | \Phi_0 \rangle = 1$.
(R2) $\langle \Phi_0 | A_1 A_2 \cdots A_n \Phi_0 \rangle = 0$ if A_1 is a creation operator $a^\dagger_{\mathbf{p}}$ or A_n is an annihilation operator $a_{\mathbf{q}}$.
(R3) $a_{\mathbf{p}} a^\dagger_{\mathbf{q}} = [a_{\mathbf{p}}, a^\dagger_{\mathbf{q}}]_- + a^\dagger_{\mathbf{q}} a_{\mathbf{p}}$.
(R4) $[a_{\mathbf{p}}, a^\dagger_{\mathbf{q}}]_- = \delta_{\mathbf{p},\mathbf{q}} I$.
(R5) $\langle \Phi_0 | a_{\mathbf{p}} a^\dagger_{\mathbf{q}} \Phi_0 \rangle = \langle \Phi_0 |\, [a_{\mathbf{p}}, a^\dagger_{\mathbf{q}}]_- \, \Phi_0 \rangle = \delta_{\mathbf{p},\mathbf{q}}$.

Let $B_1, \ldots, B_n$ be creation or annihilation operators. For computing the vacuum expectation value

$$\langle \Phi_0 | B_1 B_2 \cdots B_n \Phi_0 \rangle,$$

we will use (R3) along with (R4) in order to achieve (R1) or (R2).

Example 15.2 *For all* $\mathbf{p}, \mathbf{q}, \mathbf{r} \in \mathcal{G}(N)$, *we have* $\langle \Phi_0 | a_{\mathbf{p}} a^\dagger_{\mathbf{q}} a^\dagger_{\mathbf{r}} \Phi_0 \rangle = 0$.

Proof. By (R3), $a_{\mathbf{p}} a^\dagger_{\mathbf{q}} a^\dagger_{\mathbf{r}} = [a_{\mathbf{p}}, a^\dagger_{\mathbf{q}}]_- \; a^\dagger_{\mathbf{r}} + a^\dagger_{\mathbf{q}} a_{\mathbf{p}} a^\dagger_{\mathbf{r}}$. Using (R3) again,

$$a_{\mathbf{p}} a^\dagger_{\mathbf{q}} a^\dagger_{\mathbf{r}} = [a_{\mathbf{p}}, a^\dagger_{\mathbf{q}}]_- \; a^\dagger_{\mathbf{r}} + a^\dagger_{\mathbf{q}} [a_{\mathbf{p}}, a^\dagger_{\mathbf{r}}]_- + a^\dagger_{\mathbf{q}} a^\dagger_{\mathbf{r}} a_{\mathbf{p}}.$$

Hence

$$a_{\mathbf{p}} a^\dagger_{\mathbf{q}} a^\dagger_{\mathbf{r}} = \delta_{\mathbf{p},\mathbf{q}} \; a^\dagger_{\mathbf{r}} + \delta_{\mathbf{p},\mathbf{r}} a^\dagger_{\mathbf{q}} + a^\dagger_{\mathbf{q}} a^\dagger_{\mathbf{r}} a_{\mathbf{p}}. \tag{15.15}$$

By (R2), the vacuum expectation values of all the terms vanish. □

Example 15.3 *For all* $\mathbf{s}, \mathbf{p}, \mathbf{q}, \mathbf{r} \in \mathcal{G}(N)$,

$$\langle \Phi_0 | a_{\mathbf{s}} a_{\mathbf{p}} a^\dagger_{\mathbf{q}} a^\dagger_{\mathbf{r}} \Phi_0 \rangle = \delta_{\mathbf{p},\mathbf{q}} \delta_{\mathbf{s},\mathbf{r}} + \delta_{\mathbf{p},\mathbf{r}} \delta_{\mathbf{s},\mathbf{q}}.$$

Proof. It follows from (15.15) that,

$$a_{\mathbf{s}} a_{\mathbf{p}} a^\dagger_{\mathbf{q}} a^\dagger_{\mathbf{r}} = \delta_{\mathbf{p},\mathbf{q}} \; a_{\mathbf{s}} a^\dagger_{\mathbf{r}} + \delta_{\mathbf{p},\mathbf{r}} a_{\mathbf{s}} a^\dagger_{\mathbf{q}} + a_{\mathbf{s}} a^\dagger_{\mathbf{q}} a^\dagger_{\mathbf{r}} a_{\mathbf{p}}.$$

Finally, use (R2) and (R5). □

15.1.5 The Main Wick Theorem

> The main Wick theorem is the basis for computing scattering processes via Feynman diagrams.
>
> Folklore

Recall that, by definition, the normal product

$$: A_1 A_2 \cdots A_n :$$

of creation and annihilation operators is obtained by moving the annihilation operators from left to right. For example,

$$: a^\dagger_{\mathbf{s}} a_{\mathbf{p}} a_{\mathbf{q}} a^\dagger_{\mathbf{r}} : \; = a^\dagger_{\mathbf{s}} a^\dagger_{\mathbf{r}} a_{\mathbf{p}} a_{\mathbf{q}}.$$

Normal products have the following two properties.

(P1) Normal product principle: The vacuum expectation value of a normal product vanishes,

$$\langle \Phi_0 | : A_1 A_2 \cdots A_n : \Phi_0 \rangle = 0.$$

This follows from (R2) above.

(P2) Normal products are invariant under permutations of the factors.

The last statement follows from the commutation relations

$$a_{\mathbf{p}}a_{\mathbf{q}} = a_{\mathbf{q}}a_{\mathbf{p}}, \qquad a^\dagger_{\mathbf{p}}a^\dagger_{\mathbf{q}} = a^\dagger_{\mathbf{q}}a^\dagger_{\mathbf{p}}.$$

Contractions. Let A, B be creation or annihilation operators. The contraction $\mathcal{C}(AB)$ of A with B is defined by

$$\mathcal{C}(AB) := \langle \Phi_0 | AB \Phi_0 \rangle.$$

Explicitly, by (R2) above,

$$\mathcal{C}(a_{\mathbf{p}}a^\dagger_{\mathbf{q}}) = \langle \Phi_0 | (a_{\mathbf{p}}a^\dagger_{\mathbf{q}} - a^\dagger_{\mathbf{q}}a_{\mathbf{p}})\Phi_0 \rangle = \delta_{\mathbf{p},\mathbf{q}}.$$

Furthermore, again by (R2),

$$\mathcal{C}(a^\dagger_{\mathbf{p}}a_{\mathbf{q}}) = \mathcal{C}(a^\dagger_{\mathbf{p}}a^\dagger_{\mathbf{q}}) = \mathcal{C}(a_{\mathbf{p}}a_{\mathbf{q}}) = 0.$$

Paired normal products. It is useful to consider normal products where pairs of factors are replaced by contractions. Generally, we define

$$\mathcal{C}_{kl} : A_1 A_2 \cdots A_n : \; = \mathcal{C}(A_k A_l) : A_1 \cdots A_{k-1} A_{k+1} \ldots A_{l-1} A_{l+1} \cdots A_n : .$$

For example, $\mathcal{C}_{12} : AB : \; = \mathcal{C}(AB) I$, and

$$\mathcal{C}_{34} : ABCD : \; = \mathcal{C}(CD) : AB : .$$

We will also consider iterations of this procedure. For example,

$$\mathcal{C}_{12}\mathcal{C}_{35} : ABCDE : \; = \mathcal{C}(AB)\mathcal{C}(CE) : D : .$$

Trivially, $: D := D$. As we will show below,

$$BC = \; : BC : + \mathcal{C}_{12} : BC : \tag{15.16}$$

and

$$ABC = \; : ABC : + \mathcal{C}_{12} : ABC : + \mathcal{C}_{13} : ABC : + \mathcal{C}_{23} : ABC : . \tag{15.17}$$

Furthermore,

$$\begin{aligned} ABCD = \; & : ABCD : + \mathcal{C}_{12} : ABCD : + \mathcal{C}_{13} : ABCD : \\ & + \mathcal{C}_{14} : ABCD : + \mathcal{C}_{23} : ABCD : + \mathcal{C}_{24} : ABCD : + \mathcal{C}_{34} : ABCD : \\ & + \mathcal{C}_{12}\mathcal{C}_{34} : ABCD : + \mathcal{C}_{13}\mathcal{C}_{24} : ABCD : + \mathcal{C}_{14}\mathcal{C}_{23} : ABCD : . \end{aligned} \tag{15.18}$$

These are special cases of the following first Wick theorem.

Proposition 15.4 *Each product of creation and annihilation operators is the sum of all possible paired normal products.*

Proof. Ad (15.16). Note that $[a_{\mathbf{p}}, a_{\mathbf{q}}^\dagger]_- = \mathcal{C}(a_{\mathbf{p}} a_{\mathbf{q}}^\dagger)\, I$. Hence

$$a_{\mathbf{p}} a_{\mathbf{q}}^\dagger = a_{\mathbf{q}}^\dagger a_{\mathbf{p}} + [a_{\mathbf{p}}, a_{\mathbf{q}}^\dagger]_- = :a_{\mathbf{p}} a_{\mathbf{q}}^\dagger: + \mathcal{C}(a_{\mathbf{p}} a_{\mathbf{q}}^\dagger)\, I.$$

If BC is any one of $a_{\mathbf{p}}^\dagger a_{\mathbf{q}}, a_{\mathbf{p}} a_{\mathbf{q}}$ or $a_{\mathbf{p}}^\dagger a_{\mathbf{q}}^\dagger$, then $BC = :BC:$ and $\mathcal{C}(BC) = 0$.

Ad (15.17). (I) Let $A := a_{\mathbf{p}}^\dagger$. By (15.16),

$$ABC = A:BC: + A\mathcal{C}_{12}:BC:$$

Moreover, $A:BC: = :ABC:$ and $\mathcal{C}(AB) = \mathcal{C}(AC) = 0$.

(II) Let $A := a_{\mathbf{p}}$. Then $:ABC: = :BC:A$. By (15.16),

$$:ABC: = BCA - \mathcal{C}_{12}:BC:A.$$

Hence

$$BCA = :ABC: + \mathcal{C}_{12}:BC:A.$$

Using the commutator, $BCA = B\,[C,A]_- + BAC$. Hence

$$\begin{aligned} BCA &= B\,[C,A]_- + [B,A]_-\, C + ABC \\ &= -\mathcal{C}(AC):B: - \mathcal{C}(AB):C: + ABC. \end{aligned}$$

This yields (15.17). For n factors, we proceed by induction. □

By the normal product principle (P1) on page 824, the vacuum expectation value of a normal product vanishes. Therefore, it follows from (15.18) that

$$\langle \Phi_0 | ABCD\Phi_0 \rangle = \mathcal{C}(AB)\mathcal{C}(CD) + \mathcal{C}(AC)\mathcal{C}(BD) + \mathcal{C}(AD)\mathcal{C}(BD).$$

This is the sum of all possible total pairings. Moreover, it follows from (15.17) that

$$\langle \Phi_0 | ABC\Phi_0 \rangle = 0.$$

This means that the first Wick theorem substantially simplifies the computation of vacuum expectation values by cancelling redundant terms. For the general situation

$$\tau := \langle \Phi_0 | :A_1 A_2 \cdots A_n: \Phi_0 \rangle, \qquad n = 1, 2, \ldots,$$

the same argument yields the following second Wick theorem.

Proposition 15.5 *If n is odd, then $\tau = 0$. If n is even, then τ is equal to the sum of all possible products*

$$\langle \Phi_0 | A_{i_1} A_{i_2} \Phi_0 \rangle \langle \Phi_0 | A_{i_3} A_{i_4} \Phi_0 \rangle \cdots \langle \Phi_0 | A_{i_{n-1}} A_{i_n} \Phi_0 \rangle$$

where $i_1, i_2, \ldots i_n$ is a permutation of $1, 2, \ldots, n$. Moreover, $i_j < i_{j+1}$ for odd indices $j = 1, 3, \ldots n-1$ and $i_1 < i_3 < i_5 < \ldots < i_{n-1}$.

This corresponds to all possible total contractions of the normal product. Total contractions are also called total pairings.

Time-ordered products. Our next goal is the main Wick theorem below which refers to the vacuum expectation values of time-ordered products. We assume the following.

(A) The functions A, B, C, D and $A_j, j = 1, \dots, n$, are finite linear combinations of creation and annihilation operators where the coefficients are complex-valued functions depending on space and time.

For example, the function $A(x) := \varphi_{\text{free}}(x)$ satisfies this assumption. Define the time-ordered contraction between A and B by

$$C(AB)(x, y) := \langle \Phi_0 | \mathcal{T}(A(x)B(y))\Phi_0 \rangle.$$

For example,

$$C(\varphi_{\text{free}}(x)\phi_{\text{free}}(y)) = G_{2,\text{free},\mathcal{G}(N)}(x, y).$$

Moreover, define time-ordered paired normal products by setting

$$C_{kl} : A_1 A_2 \cdots A_n : = C(A_k A_l) : A_1 \cdots A_{k-1} A_{k+1} \dots A_{l-1} A_{l+1} \cdots A_n : .$$

For time-ordered contractions, there holds the relation

$$\mathcal{T}(BC) = \ : BC : + C_{12} : BC : \tag{15.19}$$

along with

$$\mathcal{T}(ABC) = \ : ABC : + C_{12} : ABC : + C_{13} : ABC : + C_{23} : ABC :$$

and

$$\begin{aligned} \mathcal{T}(ABCD) = \ & : ABCD : + C_{12} : ABCD : + C_{13} : ABCD : + C_{14} : ABCD : \\ & + C_{23} : ABCD : + C_{24} : ABCD : + C_{34} : ABCD : \\ & + C_{12}C_{34} : ABCD : + C_{13}C_{24} : ABCD : + C_{14}C_{23} : ABCD : . \end{aligned}$$

Let us prove (15.19). In fact, by (15.16),

$$B(\mathbf{x}, t)C(\mathbf{y}, s) = \ : B(\mathbf{x}, t)C(\mathbf{y}, s) : + \mathcal{C}_{12} : B(\mathbf{x}, t)C(\mathbf{y}, s) :$$

along with

$$C(\mathbf{y}, s)B(\mathbf{x}, t) = \ : C(\mathbf{y}, s)B(\mathbf{x}, t) : + \mathcal{C}_{12} : C(\mathbf{y}, s)B(\mathbf{x}, t) : .$$

Since normal products are invariant under permutations of the factors, we get

$$C(\mathbf{y}, s)B(\mathbf{x}, t) = \ : B(\mathbf{x}, t)C(\mathbf{y}, s) : + \mathcal{C}_{12} : C(\mathbf{y}, s)B(\mathbf{x}, t) : .$$

By the definition of the chronological operator $\mathcal{T}$ on page 744, we obtain (15.19). Similarly, we get the expressions above for $\mathcal{T}(ABC)$ and $\mathcal{T}(ABCD)$.

In general, the first and second Wick theorem remain valid if we replace products (resp. contractions) by time-ordered products (resp. time-ordered contractions). In particular, we obtain the following result.

The main Wick theorem. Consider the vacuum expectation value of the following time-ordered product

$$\tau := \langle \Phi_0 | \mathcal{T}(A_1 A_2 \cdots A_n)\Phi_0\rangle, \qquad n = 1, 2, \ldots$$

where $A_j = A_j(x)$, $j = 1, 2, \ldots, n$, are operator-valued functions which satisfy assumption (A) above.

Theorem 15.6 *If n is odd, then $\tau = 0$. If n is even, then τ is equal to the sum of all possible products*

$$\langle \Phi_0 | \mathcal{T}(A_{i_1} A_{i_2})\Phi_0\rangle \langle \Phi_0 | \mathcal{T}(A_{i_3} A_{i_4})\Phi_0\rangle \cdots \langle \Phi_0 | \mathcal{T}(A_{i_{n-1}} A_{i_n})\Phi_0\rangle$$

where $i_1, i_2, \ldots i_n$ is a permutation of $1, 2, \ldots n$. Moreover, $i_j < i_{j+1}$ for odd indices $j = 1, 3, \ldots n-1$ and $i_1 < i_3 < i_5 < \ldots < i_{n-1}$.

Special notation. In order to best understand the language of Feynman diagrams to be introduced below, let us change the notation. We write

$$\underbrace{AB} := C(AB) = \langle \Phi_0 | \mathcal{T}(AB)\Phi_0\rangle$$

and

$$\langle \Phi_0 | \mathcal{T}(A_1 \cdots \underbrace{A_j \cdots A_k} \cdots A_n)\Phi_0\rangle$$
$$= \underbrace{A_j A_k} \langle \Phi_0 | \mathcal{T}(A_1 \cdots A_{j-1} A_{j+1} \cdots A_{k-1} A_{k+1} \cdots A_n)\Phi_0\rangle.$$

For example, the main Wick theorem (Theorem 15.6) tells us that

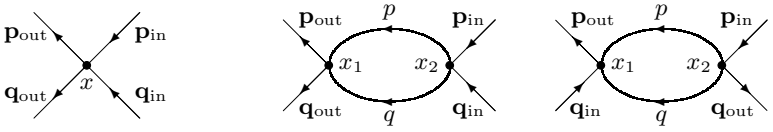
$$\langle \Phi_0 | \mathcal{T}(ABCD)\Phi_0\rangle$$

is equal to the sum of all possible total pairings:

$$\langle \Phi_0 | \mathcal{T}(\underbrace{AB}\,\underbrace{CD})\Phi_0\rangle + \langle \Phi_0 | \mathcal{T}(\underbrace{A\underbrace{BC}D})\Phi_0\rangle + \langle \Phi_0 | \mathcal{T}(\underbrace{\underbrace{AB}C\underbrace{D}})\Phi_0\rangle$$
$$= \underbrace{AB} \cdot \underbrace{CD} + \underbrace{AC} \cdot \underbrace{BD} + \underbrace{AD} \cdot \underbrace{BC}.$$

Such total pairings will be used systematically in the proof of Prop. 15.8 below.

We now want to investigate the fundamental relation of the Wick theorem to transition amplitudes, transition probabilities, and cross sections of scattering processes.



(a) First order (b) Second order

Fig. 15.2. Feynman diagrams

15.1.6 Transition Amplitude

Let us study the following scattering process (Fig. 15.2(a)).

- There are two ingoing particles with momentum vectors $\mathbf{p}_{\text{in}}$ and $\mathbf{q}_{\text{in}}$ and
- two outgoing particles with momentum vectors $\mathbf{p}_{\text{out}}$ and $\mathbf{q}_{\text{out}}$.

We introduce the ingoing particle state

$$\Phi_{\text{in}} := \alpha a^\dagger_{\mathbf{p}_{\text{in}}} a^\dagger_{\mathbf{q}_{\text{in}}} \Phi_0$$

and the outgoing particle state

$$\Phi_{\text{out}} := \beta a^\dagger_{\mathbf{p}_{\text{out}}} a^\dagger_{\mathbf{q}_{\text{out}}} \Phi_0.$$

Here, the unit vector Φ_0 denotes the ground state (vacuum state). Moreover, we set $\alpha := 1$ if $\mathbf{p}_{\text{in}} \neq \mathbf{q}_{\text{in}}$. Otherwise, $\alpha := 1/2$. Similarly, we choose $\beta := 1$ if $\mathbf{p}_{\text{out}} \neq \mathbf{q}_{\text{out}}$. Otherwise, $\beta := 1/2$.

Proposition 15.7 (i) $\langle \Phi_{\text{in}} | \Phi_{\text{in}} \rangle = 1$ *and* $\langle \Phi_{\text{out}} | \Phi_{\text{out}} \rangle = 1$.

(ii) $\langle \Phi_{\text{out}} | \Phi_{\text{in}} \rangle = 0$ *if the momentum vectors of the incoming particles are different from the momentum vectors of the outgoing particles.*

Proof. By Prop. 15.5 on page 826,

$$\langle a^\dagger_{\mathbf{p}} a^\dagger_{\mathbf{q}} \Phi_0 | a^\dagger_{\mathbf{r}} a^\dagger_{\mathbf{s}} \Phi_0 \rangle = \langle \Phi_0 | a_{\mathbf{q}} a_{\mathbf{p}} a^\dagger_{\mathbf{r}} a^\dagger_{\mathbf{s}} \Phi_0 \rangle = \delta_{\mathbf{q},\mathbf{r}} \delta_{\mathbf{p},\mathbf{s}} + \delta_{\mathbf{q},\mathbf{s}} \delta_{\mathbf{p},\mathbf{r}}.$$

For example, if $\mathbf{p} = \mathbf{r}$ and $\mathbf{q} = \mathbf{s}$ along with $\mathbf{r} \neq \mathbf{s}$, then we get the value 1. □

Let us now study the transition amplitude

$$\boxed{\tau := \langle \Phi_{\text{out}} | \, \mathsf{S}(T) \Phi_{\text{in}} \rangle.}$$

By the Dyson series (15.14), we obtain

$$\tau = \tau_0 + \kappa \tau_1 + \kappa^2 \tau_2 + \dots$$

with $\tau_0 := \langle \Phi_{\text{out}} | \Phi_{\text{in}} \rangle$. Explicitly,

$$\tau_1 := \mathrm{i}\int_{-T/2}^{T/2} dt \int_{\mathcal{C}(L)} d^3\mathbf{x}\; f(x)$$

with $x = (\mathbf{x}, t)$ and

$$f(x) := \langle \Phi_0 | a_{\mathbf{p}_{\mathrm{out}}} a_{\mathbf{q}_{\mathrm{out}}} : \varphi_{\mathrm{free}}(x)^4 : a^\dagger_{\mathbf{p}_{\mathrm{in}}} a^\dagger_{\mathbf{q}_{\mathrm{in}}} \Phi_0 \rangle.$$

Furthermore,

$$\tau_2 := -\frac{1}{2}\int_{-T/2}^{T/2} dt_1 \int_{\mathcal{C}(L)} d^3\mathbf{x}_1 \int_{-T/2}^{T/2} dt_2 \int_{\mathcal{C}(L)} d^3\mathbf{x}_2\; g(x_1, x_2)$$

with

$$g(x_1, x_2) := \langle \Phi_0 | a_{\mathbf{p}_{\mathrm{out}}} a_{\mathbf{q}_{\mathrm{out}}} \mathcal{T}\{ : \varphi_{\mathrm{free}}(x_1)^4 :: \varphi_{\mathrm{free}}(x_2)^4 : \}\; a^\dagger_{\mathbf{p}_{\mathrm{in}}} a^\dagger_{\mathbf{q}_{\mathrm{in}}} \Phi_0 \rangle.$$

Theorem 15.8 (i) *The first order approximation reads as*

$$\tau_1 \;= \frac{6(2\pi)^4 \mathrm{i}\; \delta_{\mathrm{disc}}(p_{\mathrm{in}} + q_{\mathrm{in}} - p_{\mathrm{out}} - q_{\mathrm{out}})}{\mathcal{V}^2 \sqrt{E_{\mathbf{p}_{\mathrm{in}}} E_{\mathbf{q}_{\mathrm{in}}} E_{\mathbf{p}_{\mathrm{out}}} E_{\mathbf{q}_{\mathrm{out}}}}}.$$

(ii) *The second order approximation reads as* $\tau_2 = 4\tau_{21} + 4\tau_{22}$. *Here, for j=1,2, we set*

$$\tau_{2j} := \int_{-E_{\max}}^{E_{\max}} dp_0 \int_{-E_{\max}}^{E_{\max}} dq_0 \sum_{\mathbf{p}\in\mathcal{G}(N)} \Delta^3\mathbf{p} \sum_{\mathbf{q}\in\mathcal{G}(N)} \Delta^3\mathbf{p} \cdot \alpha_j$$

along with

$$\alpha_1 := -\frac{(2\pi)^8\; \delta_{\mathrm{disc}}(p + q - p_{\mathrm{out}} - q_{\mathrm{out}})\; \delta_{\mathrm{disc}}(p_{\mathrm{in}} + q_{\mathrm{in}} - p - q)}{8\mathcal{V}^2 (p^2 - m_0^2 + \mathrm{i}\varepsilon)(q^2 - m_0^2 + \mathrm{i}\varepsilon) \sqrt{E_{\mathbf{p}_{\mathrm{in}}} E_{\mathbf{q}_{\mathrm{in}}} E_{\mathbf{p}_{\mathrm{out}}} E_{\mathbf{q}_{\mathrm{out}}}}}$$

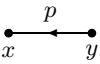
and

$$\alpha_2 := -\frac{(2\pi)^8\; \delta_{\mathrm{disc}}(p + q + q_{\mathrm{in}} - p_{\mathrm{out}})\; \delta_{\mathrm{disc}}(p_{\mathrm{in}} - q_{\mathrm{out}} - p - q)}{8\mathcal{V}^2 (p^2 - m_0^2 + \mathrm{i}\varepsilon)(q^2 - m_0^2 + \mathrm{i}\varepsilon) \sqrt{E_{\mathbf{p}_{\mathrm{in}}} E_{\mathbf{q}_{\mathrm{in}}} E_{\mathbf{p}_{\mathrm{out}}} E_{\mathbf{q}_{\mathrm{out}}}}}.$$

Note that the appearance of the delta functions reflects conservation of momentum and energy for the scattering process in the case of the continuum limit. Here, α_1 (resp. α_2) correspond to the left-hand (resp. right-hand) Feynman diagram from Fig. 15.2(b) on page 829.

Relativistic invariance. Note that the formulas for τ_1 and τ_2 above are not relativistically invariant; they depend on the choice of the inertial system and the corresponding discretization by the lattice in momentum

Table 15.1. Feynman rules

pairing	Feynman diagram
$\underbrace{\varphi_{\text{free}}(x)a_{\mathbf{p}}^{\dagger}} = \mathrm{e}^{-\mathrm{i}px}\mathcal{N}_{\mathbf{p}}$	
$\underbrace{a_{\mathbf{p}}\varphi_{\text{free}}}(x) = \mathrm{e}^{\mathrm{i}px}\mathcal{N}_{\mathbf{p}}$	
$\underbrace{\varphi_{\text{free}}(x)\varphi_{\text{free}}}(y) = G_{2,\mathcal{G}(N),\text{free}}(x,y)$	
$G_{2,\mathcal{G}(N),\text{free}}(x,y) = \lim_{\varepsilon\to+0}\frac{\mathrm{i}}{(2\pi)^4}\int_{-E_{\max}}^{E_{\max}} dp_0 \sum_{\mathbf{p}\in\mathcal{G}(N)} \frac{\mathrm{e}^{-\mathrm{i}px}\mathrm{e}^{\mathrm{i}py}}{p^2-m_0^2+\mathrm{i}\varepsilon}\Delta^3\mathbf{p}$	
$\mathcal{N}_{\mathbf{p}} = \frac{1}{\sqrt{2E_{\mathbf{p}}\mathcal{V}}}$	

space. However, the continuum limit is relativistically invariant, at least on a formal level. To this end, we have to use the replacements

$$\delta_{\text{dis}}(p) \Rightarrow \delta^4(p)$$

and

$$\int_{-T/2}^{T/2} dp_0 \sum_{\mathbf{p}\in\mathcal{G}(N)} \Delta^3\mathbf{p}\ldots \Rightarrow \int_{\mathbb{R}^4} d^4p\ldots$$

Proof of Theorem 15.8. To simplify notation, we write the symbol φ instead of φ_{free}. The basic ideas of the proof are the following ones.

- We apply the main Wick theorem to the S-matrix elements.
- This way, vacuum expectation values of time-ordered products are reduced to products of time-ordered contractions.
- The time-ordered contractions can be represented graphically as the basic elements of Feynman diagrams (see Table 15.1).
- The products of time-ordered contractions can be represented graphically by Feynman diagrams (see Figs. 15.3-15.5 on page 835).

Ad (i). By the main Wick theorem (Theorem 15.6) on page 828,

$$f(x) = \langle\Phi_0|\underbrace{a_{\mathbf{p}_{\text{out}}}\underbrace{a_{\mathbf{q}_{\text{out}}}:\varphi(x)}\underbrace{\varphi(x)\varphi(x)}\underbrace{\varphi(x):a_{\mathbf{p}_{\text{in}}}^{\dagger}}a_{\mathbf{q}_{\text{in}}}^{\dagger}}\Phi_0\rangle + \ldots$$

We have to sum over all possible total pairings. The dots stand for the remaining total pairings. Note the crucial fact that pairings between operators of a normal product vanish. Therefore,

$$\langle \Phi_0 | \ldots : \underbrace{\varphi(x)\varphi(x)}\varphi(x)\varphi(x) : \ldots \Phi_0 \rangle = 0.$$

Consequently, such pairings drop out. Hence

$$f(x) = \underbrace{a_{\mathbf{p}_{\mathrm{out}}}\varphi(x)} \cdot \underbrace{a_{\mathbf{q}_{\mathrm{out}}}\varphi(x)} \cdot \underbrace{\varphi(x)a^\dagger_{\mathbf{p}_{\mathrm{in}}}} \cdot \underbrace{\varphi(x)a^\dagger_{\mathbf{q}_{\mathrm{in}}}} + \ldots$$

Using the expression for the free field φ from (15.6) on page 817,

$$\underbrace{a_{\mathbf{p}}\varphi(x)} = \underbrace{a_{\mathbf{p}}a^\dagger_{\mathbf{p}}}\mathrm{e}^{\mathrm{i}px}\mathcal{N}_{\mathbf{p}} = \mathrm{e}^{\mathrm{i}px}\mathcal{N}_{\mathbf{p}}.$$

Similarly, we obtain the remaining formulas of Table 15.1 on page 831. Hence

$$f(x) = \frac{\mathrm{e}^{-\mathrm{i}x(p_{\mathrm{in}}+q_{\mathrm{in}}-p_{\mathrm{out}}-q_{\mathrm{out}})}}{4\mathcal{V}^2\sqrt{E_{\mathbf{p}_{\mathrm{out}}}E_{\mathbf{q}_{\mathrm{out}}}E_{\mathbf{p}_{\mathrm{in}}}E_{\mathbf{q}_{\mathrm{in}}}}} + \ldots$$

Finally, we have to integrate over t and $\mathbf{x}$,

$$\tau_1 = \mathrm{i}\int_{-T/2}^{T/2} dt \int_{\mathcal{C}(L)} d^3\mathbf{x}\, f(x) + \ldots$$

By (15.4) on page 816,

$$\tau_1 = \frac{(2\pi)^4\mathrm{i}\,\delta_{\mathrm{disc}}(p_{\mathrm{in}} + q_{\mathrm{in}} - p_{\mathrm{out}} - q_{\mathrm{out}})}{4\mathcal{V}^2\sqrt{E_{\mathbf{p}_{\mathrm{in}}}E_{\mathbf{q}_{\mathrm{in}}}E_{\mathbf{p}_{\mathrm{out}}}E_{\mathbf{q}_{\mathrm{out}}}}} + \ldots.$$

This is the claim (i) up to the factor 24. In order to understand the appearance of the symmetry factor 24, we will use the following graphical language which is the prototype of the method of Feynman diagrams.

- Draw a vertex along with four free arms (Fig. 15.2(a) on page 829).
- Equip one of the arms with both the symbol $\mathbf{p}_{\mathrm{in}}$ and an incoming arrow. Mathematically, this corresponds to the contraction[7]

  $$\underbrace{\varphi(x)a^\dagger_{\mathbf{p}_{\mathrm{in}}}}$$

  Physically, this represents an incoming particle with momentum vector $\mathbf{p}_{\mathrm{in}}$.
- Similarly, we equip the remaining arms with the symbols

  $$\mathbf{q}_{\mathrm{in}}, \quad \mathbf{p}_{\mathrm{out}}, \quad \mathbf{q}_{\mathrm{out}}$$

  and one incoming arrow and two outgoing arrows, respectively. Mathematically, this corresponds to the contractions

[7] See Table 15.1 on page 831. Recall that we write φ instead of φ_{free}.

$$\underbrace{\varphi(x)a^\dagger_{\mathbf{q}_{\text{in}}}}, \quad \underbrace{a_{\mathbf{p}_{\text{out}}}\varphi(x)}, \quad \underbrace{a_{\mathbf{q}_{\text{out}}}\varphi(x)},$$

respectively. Physically, this represents an incoming particle and two outgoing particles with momentum vectors $\mathbf{q}_{\text{in}}, \mathbf{p}_{\text{out}}, \mathbf{q}_{\text{out}}$, respectively. For example, the Feynman diagram from Fig. 15.2(a) on page 829 corresponds to the product

$$\underbrace{a_{\mathbf{p}_{\text{out}}}\varphi(x)} \cdot \underbrace{a_{\mathbf{q}_{\text{out}}}\varphi(x)} \cdot \underbrace{\varphi(x)a^\dagger_{\mathbf{p}_{\text{in}}}} \cdot \underbrace{\varphi(x)a^\dagger_{\mathbf{q}_{\text{in}}}}.$$

All of the remaining total pairings are obtained by using all possible permutations of the arms. Finally, each of these $4! = 24$ total pairings yields the same contribution to τ_1.

Ad (ii). We obtain

$$g(x_1, x_2) = a(x_1, x_2) + b(x_1, x_2) + \ldots \tag{15.20}$$

By definition, $a(x_1, x_2)$ is equal to the total pairing

$$\langle \Phi_0 | \mathcal{T}(\underbrace{a_{\mathbf{p}_{\text{out}}}\underbrace{a_{\mathbf{q}_{\text{out}}} : \varphi(x_1)\varphi(x_1)}\underbrace{\varphi(x_1)\varphi(x_1) :: \varphi(x_2)\varphi(x_2)}} \times \underbrace{\varphi(x_2)\underbrace{\varphi(x_2) : a^\dagger_{\mathbf{p}_{\text{in}}}}a^\dagger_{\mathbf{q}_{\text{in}}}})\Phi_0\rangle.$$

Therefore, $a(x_1, x_2)$ is equal to the following product of time-ordered contractions:

$$\underbrace{a_{\mathbf{p}_{\text{out}}}\varphi(x_1)}\underbrace{a_{\mathbf{q}_{\text{out}}}\varphi(x_1)}\underbrace{\varphi(x_1)\varphi(x_2)}\underbrace{\varphi(x_1)\varphi(x_2)}\underbrace{\varphi(x_2)a^\dagger_{\mathbf{p}_{\text{in}}}}\underbrace{\varphi(x_2)a^\dagger_{\mathbf{q}_{\text{in}}}}.$$

This corresponds to the left-hand diagram of Fig. 15.2(b) on page 829. Furthermore, by definition, $b(x_1, x_2)$ is equal to

$$\underbrace{a_{\mathbf{p}_{\text{out}}}\varphi(x_1)}\underbrace{a_{\mathbf{q}_{\text{out}}}\varphi(x_2)}\underbrace{\varphi(x_1)\varphi(x_2)}\underbrace{\varphi(x_1)\varphi(x_2)}\underbrace{\varphi(x_2)a^\dagger_{\mathbf{p}_{\text{in}}}}\underbrace{\varphi(x_1)a^\dagger_{\mathbf{q}_{\text{in}}}}.$$

This corresponds to the right-hand diagram of Fig. 15.2(b). The dots from (15.20) stand for the sum of the remaining total pairings. As in the proof of (i) above, all of the terms vanish which contain pairings inside a normal product. For example,

$$\langle \Phi_0 | \mathcal{T}(\ldots : \underbrace{\varphi(x_1)\varphi(x_1)}\varphi(x_1)\varphi(x_1) :: \varphi(x_2)\varphi(x_2)\varphi(x_2)\varphi(x_2) : \ldots)\Phi_0\rangle = 0.$$

and

$$\langle \Phi_0 | \mathcal{T}(\ldots : \varphi(x_1)\varphi(x_1)\varphi(x_1)\varphi(x_1) :: \underbrace{\varphi(x_2)\underbrace{\varphi(x_2)\varphi(x_2)}\varphi(x_2)} : \ldots)\Phi_0\rangle = 0.$$

Such pairings drop out in (15.20).

(I) Computation of a. By Table 15.1 on page 831, the product of time-ordered contractions

$$\underbrace{a_{\mathbf{p}_{\rm out}}}\varphi(x_1)\underbrace{a_{\mathbf{q}_{\rm out}}}\varphi(x_1)\underbrace{\varphi(x_1)}\varphi(x_2)\underbrace{\varphi(x_1)}\varphi(x_2)\underbrace{\varphi(x_2)}a^\dagger_{\mathbf{p}_{\rm in}}\underbrace{\varphi(x_2)}a^\dagger_{\mathbf{q}_{\rm in}}.$$

is equal to

$$a(x_1,x_2)=\frac{\mathrm{e}^{\mathrm{i}x_1(p_{\rm out}+q_{\rm out})}\ \mathrm{e}^{-\mathrm{i}x_2(p_{\rm in}+q_{\rm in})}}{4\mathcal{V}^2\sqrt{E_{\mathbf{p}_{\rm in}}E_{\mathbf{q}_{\rm in}}E_{\mathbf{p}_{\rm out}}E_{\mathbf{q}_{\rm out}}}}\cdot G_{2,\mathcal{G}(N),\rm free}(x_1,x_2)^2.$$

Hence

$$a(x_1,x_2)=\int_{-E_{\max}}^{E_{\max}}dp_0\int_{-E_{\max}}^{E_{\max}}dq_0\sum_{\mathbf{p}\in\mathcal{G}(N)}\Delta^3\mathbf{p}\sum_{\mathbf{q}\in\mathcal{G}(N)}\Delta^3\mathbf{p}\cdot\mathcal{A}(x_1,x_2)$$

along with

$$\mathcal{A}(x_1,x_2):=\frac{\mathrm{e}^{\mathrm{i}x_1(p_{\rm out}+q_{\rm out}-p-q)}\ \mathrm{e}^{-\mathrm{i}x_2(p_{\rm in}+q_{\rm in}-p-q)}}{4\mathcal{V}^2(p^2-m_0^2+\mathrm{i}\varepsilon)(q^2-m_0^2+\mathrm{i}\varepsilon)\sqrt{E_{\mathbf{p}_{\rm in}}E_{\mathbf{q}_{\rm in}}E_{\mathbf{p}_{\rm out}}E_{\mathbf{q}_{\rm out}}}}.$$

Carrying out the integration

$$\tau_{21}=-\tfrac{1}{2}\int_{-T/2}^{T/2}dt_1\int_{-T/2}^{T/2}dt_2\int_{\mathcal{C}(L)}d^3\mathbf{x}_1\int_{\mathcal{C}(L)}d^3\mathbf{x}_2\cdot a(x_1,x_2),$$

it follows from (15.4) on page 816 that

$$\tau_{21}=\int_{-E_{\max}}^{E_{\max}}dp_0\int_{-E_{\max}}^{E_{\max}}dq_0\sum_{\mathbf{p}\in\mathcal{G}(N)}\Delta^3\mathbf{p}\sum_{\mathbf{q}\in\mathcal{G}(N)}\Delta^3\mathbf{p}\cdot\alpha_1$$

with

$$\alpha_1=-\frac{(2\pi)^8\ \delta_{\rm disc}(p+q-p_{\rm out}-q_{\rm out})\ \delta_{\rm disc}(p_{\rm in}+q_{\rm in}-p-q)}{8\mathcal{V}^2(p^2-m_0^2+\mathrm{i}\varepsilon)(q^2-m_0^2+\mathrm{i}\varepsilon)\sqrt{E_{\mathbf{p}_{\rm in}}E_{\mathbf{q}_{\rm in}}E_{\mathbf{p}_{\rm out}}E_{\mathbf{q}_{\rm out}}}}.$$

This is the claim of Theorem 15.8(ii) for $j=1$, up to the symmetry factor 4 of the term $4\tau_{21}$, which will be discussed below. Graphically, the transition amplitude τ_{21} corresponds to the left-hand diagram of Fig. 15.2(b) on page 829.

(II) Computation of b. Similarly, we get

$$\tau_{22}=-\tfrac{1}{2}\int_{-T/2}^{T/2}dt_1\int_{-T/2}^{T/2}dt_2\int_{\mathcal{C}(L)}d^3\mathbf{x}_1\int_{\mathcal{C}(L)}d^3\mathbf{x}_2\cdot b(x_1,x_2).$$

Computing this integral as above, we get the claim of Theorem 15.8 (ii) for $j=2$, up to the symmetry factor 4 of the term $4\tau_{22}$. Graphically, the transition amplitude τ_{22} corresponds to the right-hand diagram of Fig. 15.2(b).

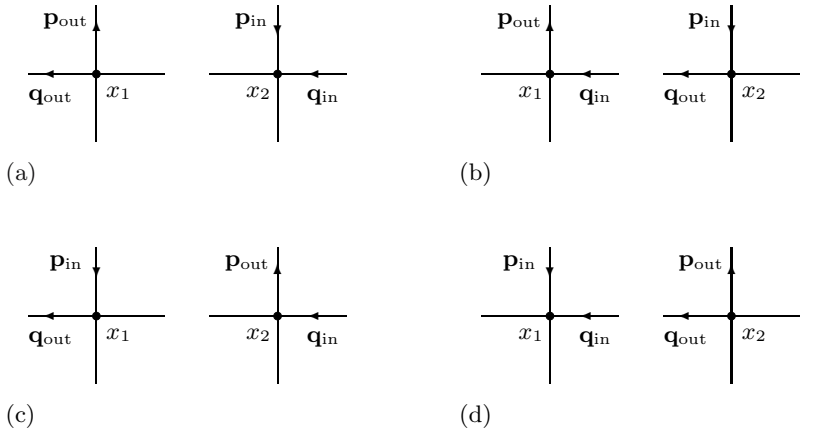


Fig. 15.3. Feynman rules (external lines)

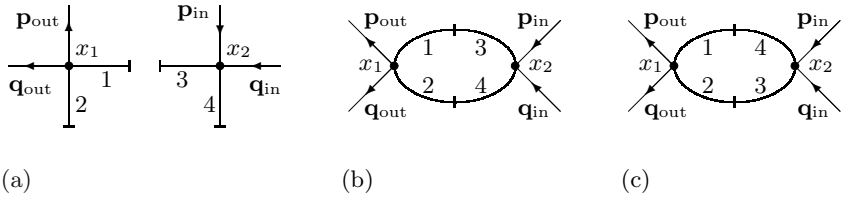


Fig. 15.4. Feynman rules (internal lines)

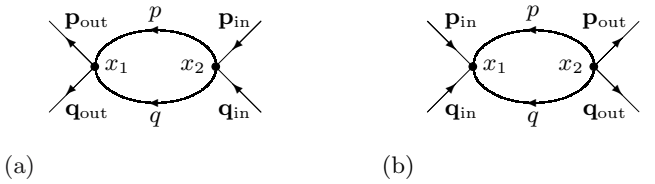


Fig. 15.5. Equivalent Feynman diagrams

(III) Symmetry factors and Feynman diagrams. In Theorem 15.8 (ii) on page 830, we claim that

$$\tau_2 = 4\tau_{21} + 4\tau_{22}.$$

The symmetry factor comes from the remaining total pairings. In order to compute the remaining total pairings from (15.20), we will use the elegant and highly effective graphical language of Feynman diagrams which is pictured in Figs. 15.3–15.5. This language is the key to modern elementary particle physics.

Now to the point. All the possible total pairings can be obtained in the following way.

- Draw two vertices equipped with the symbols x_1 and x_2. Each vertex has four arms.
- Select four arms and equip them with the symbols $\mathbf{p}_{\text{in}}, \mathbf{p}_{\text{out}}, \mathbf{q}_{\text{in}}$, and $\mathbf{q}_{\text{out}}$. This way we get the external lines (Fig. 15.3 on page 835).
- Consider all possible pairs of free arms which belong to different vertices and connect them. This way we get the internal lines. There are two possibilities in Fig. 15.4(b), (c) on page 835.
- Equip the two internal lines with the 4-momentum symbols p and q.
- Introduce arrows. We attribute ingoing (resp. outgoing) arrows to $\mathbf{p}_{\text{in}}, \mathbf{q}_{\text{in}}$ (resp. $\mathbf{p}_{\text{out}}, \mathbf{q}_{\text{out}}$). The arrows of the internal lines point from x_2 to x_1.
- Finally, draw all diagrams that can be constructed by the method described above. Totally, we get 8 diagrams which correspond to all total pairings from (15.20).

The external lines correspond to real particles. The internal lines are called virtual particles, by physicists. Similarly, as in (I) above, we assign the transition amplitude

$$\int_{-E_{\max}}^{E_{\max}} dp_0 \int_{-E_{\max}}^{E_{\max}} dq_0 \sum_{\mathbf{p}\in\mathcal{G}(N)} \Delta^3\mathbf{p} \sum_{\mathbf{q}\in\mathcal{G}(N)} \Delta^3\mathbf{p}\cdot\alpha$$

to each Feynman diagram. Explicitly,

$$\alpha := \frac{\delta_{\text{dis}}\left(\sum_{j=1}^4 \pm p_j^{(1)}\right)\ \delta_{\text{dis}}\left(\sum_{j=1}^4 \pm p_j^{(2)}\right)}{4\mathcal{V}^2(p^2-m_0^2+\mathrm{i}\varepsilon)(q^2-m_0^2+\mathrm{i}\varepsilon)\sqrt{E_{\mathbf{p}_{\text{in}}}E_{\mathbf{q}_{\text{in}}}E_{\mathbf{p}_{\text{out}}}E_{\mathbf{q}_{\text{out}}}}}.$$

Here, $\delta_{\text{dis}}\left(\sum_{j=1}^4 \pm p_j^{(r)}\right)$ corresponds to the vertex x_r with $r = 1, 2$. The symbol

$$\sum_{j=1}^4 \pm p_j^{(r)}$$

represents the sum of the 4-momentum vectors $p_j^{(r)} = (\mathbf{p}_j^{(r)}, p_{0j}^{(r)})$ of the real and virtual particles at the vertex x_r. Ingoing (resp. outgoing) particles are equipped with the positive (resp. negative) sign.

Let us now study the phenomenon of equivalent Feynman diagrams.

The point is that different Feynman diagrams may generate the same transition amplitude. For example, one checks easily that each of the eight possible Feynman diagrams yields the same transition amplitude as one of the two diagrams from Fig. 15.2(b) on page 829. This produces the symmetry factor

4 of $\tau_2 = 4\tau_{21} + 4\tau_{22}$ in Theorem 15.8(ii). For example, the two diagrams from Fig. 15.5 on page 835 yield the following two transition amplitudes

$$\chi_k := \int_{-E_{\max}}^{E_{\max}} dp_0 \int_{-E_{\max}}^{E_{\max}} dq_0 \sum_{\mathbf{p}\in\mathcal{G}(N)} \Delta^3\mathbf{p} \sum_{\mathbf{q}\in\mathcal{G}(N)} \Delta^3\mathbf{p} \cdot \beta_k, \qquad k = 1, 2$$

with

$$\beta_1 := \frac{\delta_{\mathrm{dis}}(p + q - p_{\mathrm{out}} - q_{\mathrm{out}})\ \delta_{\mathrm{dis}}(p_{\mathrm{in}} + q_{\mathrm{in}} - p - q)}{4\mathcal{V}^2(p^2 - m_0^2 + \mathrm{i}\varepsilon)(q^2 - m_0^2 + \mathrm{i}\varepsilon)\sqrt{E_{\mathbf{p}_{\mathrm{in}}} E_{\mathbf{q}_{\mathrm{in}}} E_{\mathbf{p}_{\mathrm{out}}} E_{\mathbf{q}_{\mathrm{out}}}}}$$

and

$$\beta_2 := \frac{\delta_{\mathrm{dis}}(p_{\mathrm{in}} + q_{\mathrm{in}} + p + q)\ \delta_{\mathrm{dis}}(-p - q - p_{\mathrm{out}} - q_{\mathrm{out}})}{4\mathcal{V}^2(p^2 - m_0^2 + \mathrm{i}\varepsilon)(q^2 - m_0^2 + \mathrm{i}\varepsilon)\sqrt{E_{\mathbf{p}_{\mathrm{in}}} E_{\mathbf{q}_{\mathrm{in}}} E_{\mathbf{p}_{\mathrm{out}}} E_{\mathbf{q}_{\mathrm{out}}}}}.$$

Using the substitution, $p \Rightarrow -p$ and $q \Rightarrow -q$, we get

$$\chi_1 = \chi_2.$$

The two diagrams (a) and (b) from Fig. 15.5 on page 835 are called equivalent. They differ by a permutation of the vertices x_1 and x_2. □

15.1.7 Transition Probability

We want to discuss how one passes from transition amplitudes to transition probabilities, which are crucial for the computation of cross sections.

To this end, we will use the quantization of the classical phase space.

Let us consider the scattering process from Sect. 15.1.6 pictured in Fig. 15.2(a) on page 829. In terms of classical mechanics, we have conservation of momentum vectors,

$$\mathbf{p}_{\mathrm{out}} + \mathbf{q}_{\mathrm{out}} = \mathbf{p}_{\mathrm{in}} + \mathbf{q}_{\mathrm{in}}, \tag{15.21}$$

and conservation of energy,

$$E_{\mathbf{p}_{\mathrm{out}}} + E_{\mathbf{q}_{\mathrm{out}}} = E_{\mathbf{p}_{\mathrm{in}}} + E_{\mathbf{q}_{\mathrm{in}}} \tag{15.22}$$

where $E_{\mathbf{p}} := \sqrt{m_0^2 + \mathbf{p}^2}$. Let us consider a proper scattering process. This means that $\mathbf{p}_{\mathrm{out}} \neq \mathbf{p}_{\mathrm{in}}$. This implies $\mathbf{q}_{\mathrm{out}} \neq \mathbf{q}_{\mathrm{in}}$.

The key formula for the transition probability. Consider the momentum vector $\mathbf{p} = p^1\mathbf{i} + p^2\mathbf{j} + p^3\mathbf{k}$, and set

$$\mathbf{p}_{\mathrm{out}} + \Delta^3\mathbf{p} := \{\mathbf{p} :\ p^j_{\mathrm{out}} \leq p^j \leq p^j_{\mathrm{out}} + \Delta p,\ j = 1, 2, 3\}.$$

By definition, the number

$$\boxed{\mathcal{W}(T,\mathcal{V}) := \frac{\mathcal{V}\Delta^3\mathbf{p}}{(2\pi)^3}\cdot\frac{\mathcal{V}\Delta^3\mathbf{p}}{(2\pi)^3}\ |\langle a^\dagger_{\mathbf{p}_{\rm out}}a^\dagger_{\mathbf{q}_{\rm out}}\Phi_0|\mathsf{S}(T)a^\dagger_{\mathbf{p}_{\rm in}}a^\dagger_{\mathbf{q}_{\rm in}}\Phi_0\rangle|^2}\quad (15.23)$$

is the probability for the following event that happens in the large cell $\mathcal{C}(L)$ of volume $\mathcal{V}$ during the large time interval $[-\frac{T}{2},\frac{T}{2}]$:

- Two ingoing particles P and Q having momentum vectors $\mathbf{p}_{\rm in}$ and $\mathbf{q}_{\rm in}$, respectively, are scattered such that
- the momentum vectors of the outgoing particles P and Q are living in the cell $\mathbf{p}_{\rm out}+\Delta^3\mathbf{p}$ and $\mathbf{q}_{\rm out}+\Delta^3\mathbf{p}$, respectively.

Motivation of the key formula. In finite-dimensional Hilbert spaces, the square of the modulus of the transition amplitude yields the corresponding transition probability. However, in the present case, there arises a continuum of states. Therefore, we have to modify the definition of the transition probability. To this end, let $\mathbf{x}=\mathbf{x}_{\rm out}(t)$ and $\mathbf{y}=\mathbf{y}_{\rm out}(t),\ t\in[-\frac{T}{2},\frac{T}{2}]$ be the classical trajectories of the outgoing particles P and Q in position space, respectively. The corresponding trajectory of the particle pair in phase space reads as

$$\mathbf{x}=\mathbf{x}_{\rm out}(t),\ \mathbf{p}=\mathbf{p}_{\rm out}(t),\ \mathbf{y}=\mathbf{y}_{\rm out}(t),\ \mathbf{q}=\mathbf{q}_{\rm out}(t),\qquad t\in[-\tfrac{T}{2},\tfrac{T}{2}].$$

This trajectory lives in the subset

$$\mathcal{P} := (\mathcal{C}(L)\times\mathbf{p}_{\rm out}+\Delta^3\mathbf{p})\times(\mathcal{C}(L)\times\mathbf{q}_{\rm out}+\Delta^3\mathbf{p})$$

of the 12-dimensional phase space. The set $\mathcal{P}$ has the volume

$$\operatorname{meas}\mathcal{P}=\mathcal{V}\Delta^3\mathbf{p}\cdot\mathcal{V}\Delta^3\mathbf{p}.$$

Now to the point. In the SI system, the quantity *length* × *momentum* has the physical dimension of *action*. Therefore, the volume $\operatorname{meas}\mathcal{P}$ has the same physical dimension as h^6 where h denotes the Planck constant. According to semi-classical quantum statistics, the dimensionless quantity

$$\frac{\operatorname{meas}\mathcal{P}}{h^6}=\frac{\mathcal{V}\Delta^3\mathbf{p}}{(2\pi\hbar)^3}\cdot\frac{\mathcal{V}\Delta^3\mathbf{p}}{(2\pi\hbar)^3}$$

is the number of quantum states, which corresponds to the phase volume of the set $\mathcal{P}$. By convention, we use the energetic system in this chapter. Here, $\hbar=1$. This motivates the definition of $\mathcal{W}(T,\mathcal{V})$ from (15.23).

Total probability for scattering. By the key formula (15.23) above, the number

$$\mathcal{W}_{\rm tot}(T,\mathcal{V}) := {\sum}'_{\mathbf{p}_{\rm out}\in\mathcal{G}(N)}\ \frac{\mathcal{V}\Delta^3\mathbf{p}}{(2\pi)^3}\ {\sum}'_{\mathbf{q}_{\rm out}\in\mathcal{G}(N)}\ \frac{\mathcal{V}\Delta^3\mathbf{p}}{(2\pi)^3}$$
$$\times\ |\langle a^\dagger_{\mathbf{p}_{\rm out}}a^\dagger_{\mathbf{q}_{\rm out}}\Phi_0|\mathsf{S}(T)a^\dagger_{\mathbf{p}_{\rm in}}a^\dagger_{\mathbf{q}_{\rm in}}\Phi_0\rangle|^2 \qquad (15.24)$$

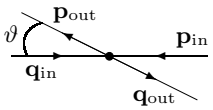


Fig. 15.6. Scattering process in the center-of-mass system

is equal to the probability of the event that the incoming particles P and Q are scattered. The symbol $\sum'$ indicates that the trivial non-scattering case given by $\mathbf{p}_{\text{out}} = \mathbf{p}_{\text{in}}, \mathbf{q}_{\text{out}} = \mathbf{q}_{\text{in}}$ is excluded.

The goal. The total probability $\mathcal{W}_{\text{tot}}(T, \mathcal{V})$ depends on the large length T of the time interval $[-\frac{T}{2}, -\frac{T}{2}]$, the large volume $\mathcal{V}$ of the box $\mathcal{C}(L)$, and the small length Δp of the lattice in momentum space. We are looking for a quantity which describes the asymptotic behavior of $\mathcal{W}_{\text{tot}}(T, \mathcal{V})$ as $T \to \infty$ and $\mathcal{V} \to \infty$. We will show in the next section that in the center-of-mass system, there exists the limit

$$\boxed{\sigma_{\text{tot}} := \lim_{T\to\infty,\, \mathcal{V}\to\infty} \frac{\mathcal{W}_{\text{tot}}(T, \mathcal{V})\mathcal{V}}{v_{\text{rel}} T}}$$

where v_{rel} is the relativistic relative velocity of the two ingoing particles. Since the probability $\mathcal{W}_{\text{tot}}(T, \mathcal{V})$ is dimensionless, the so-called total cross section σ_{tot} has the physical dimension of area.

15.1.8 Scattering Cross Section

Consider a scattering process where a homogeneous stream of ingoing particles P hits a homogeneous stream of ingoing particles Q. Physicists measure the number N_{in} of ingoing particles P and the number N_{out} of outgoing (scattered) particles P. These numbers refer to a box $\mathcal{C}(L)$ of large volume $\mathcal{V}$ and a time interval $[-\frac{T}{2}, \frac{T}{2}]$ of large length T. We will work in the center-of-mass system which will be discussed below (Fig. 15.6).

The first key formula. It is our goal to show that

$$\boxed{N_{\text{out}} = \frac{v_{\text{rel}} N_{\text{int}} T}{\mathcal{V}} \cdot \sigma_{\text{tot}}} \tag{15.25}$$

with the total cross section

$$\sigma_{\text{tot}} := \int_{-\pi/2}^{\pi/2} \frac{d\sigma(\vartheta)}{d\vartheta} \cdot d\vartheta$$

and the so-called differential cross section

$$\frac{d\sigma(\vartheta)}{d\vartheta} = \frac{|\mathcal{M}(\vartheta)|^2}{256\pi^2 E_{\mathbf{p}_{\text{in}}}^2} \cdot 2\pi \cos\vartheta, \qquad \sigma(0) = 0.$$

For a small coupling constant κ, we will get $\mathcal{M} = 24\kappa$ in first order of perturbation theory. The general definition of $\mathcal{M}(\vartheta)$ will be given below.

Formula (15.25) represents an approximation which is valid for large volume $\mathcal{V}$, large length T of the time interval$[-\frac{T}{2}, \frac{T}{2}]$, and small coupling constant κ.

> *The differential cross section is crucial in physics, since it does not depend on the volume $\mathcal{V}$ of the box, the time interval $[-\frac{T}{2}, \frac{T}{2}]$, and the relative velocity $v_{\rm rel}$.*

Using the differential $d\Omega = 2\pi \cos\vartheta \cdot d\vartheta$ of the solid angle, physicists also write

$$d\sigma = \frac{|\mathcal{M}(\vartheta)|^2}{256\pi^2 E_{\mathbf{p}_{\rm in}}^2}\, d\Omega.$$

Recall that the integral

$$\int d\Omega = \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \cos\vartheta \cdot d\varphi\, d\vartheta = \int_{-\pi/2}^{\pi/2} 2\pi \cos\vartheta \cdot\ d\vartheta = 4\pi^2$$

represents the surface measure of the unit sphere.

The second key formula. In greater detail, we obtain

$$\boxed{N_{\rm out}([\vartheta_1, \vartheta_2]) = \frac{v_{\rm rel} N_{\rm int} T}{\mathcal{V}} \int_{\vartheta_1}^{\vartheta_2} \frac{d\sigma(\vartheta)}{d\vartheta} \cdot d\vartheta.} \tag{15.26}$$

Terminology. Let us explain the notation used above. The first (resp. second) homogeneous incoming particle beam consists of particles P (resp. Q) of mass m_0 and momentum vector $\mathbf{p}_{\rm in}$ (resp. $\mathbf{q}_{\rm in}$). After scattering, the particle P (resp. Q) has the momentum vector $\mathbf{p}_{\rm out}$ (resp. $\mathbf{q}_{\rm out}$). In the center-of-mass system, classical conservation of momentum and energy yields

$$\boxed{\mathbf{p}_{\rm out} + \mathbf{q}_{\rm out} = \mathbf{p}_{\rm in} + \mathbf{q}_{\rm in} = 0}$$

and $E_{\mathbf{p}_{\rm out}} + E_{\mathbf{q}_{\rm out}} = E_{\mathbf{p}_{\rm in}} + E_{\mathbf{q}_{\rm in}}$ with $E_{\mathbf{p}} := \sqrt{m_0^2 + \mathbf{p}^2}$. This implies

$$\mathbf{p}_{\rm in} = -\mathbf{q}_{\rm in}, \qquad \mathbf{p}_{\rm out} = -\mathbf{q}_{\rm out}$$

along with $E_{\mathbf{p}_{\rm out}} = E_{\mathbf{q}_{\rm out}} = E_{\mathbf{p}_{\rm in}} = E_{\mathbf{q}_{\rm in}}$. Hence,

$$|\mathbf{p}_{\rm in}| = |\mathbf{q}_{\rm in}| = |\mathbf{p}_{\rm out}| = |\mathbf{q}_{\rm out}|.$$

Let ϑ denote the scattering angle (see Fig. 15.6 on page 839). Explicitly,

$$\mathbf{p}_{\rm in}\mathbf{p}_{\rm out} = |\mathbf{p}_{\rm in}| \cdot |\mathbf{p}_{\rm out}| \cos\vartheta = |\mathbf{p}_{\rm in}|^2 \cdot \cos\vartheta.$$

Furthermore, let us use the following notation for particle numbers.

- Let N_{in} be the number of incoming particle pairs P, Q in the box $\mathcal{C}(L)$ of volume $\mathcal{V} = L^3$ during the time interval $[-\frac{T}{2}, \frac{T}{2}]$. For the particle density of P (resp. Q), we get $\varrho_P = \varrho_Q = N_{\text{in}}/\mathcal{V}$.
- Let N_{out} be the number of scattered particle pairs P, Q in the box $\mathcal{C}(L)$ during the time interval $[-\frac{T}{2}, \frac{T}{2}]$.
- Let $N_{\text{out}}([\vartheta_1, \vartheta_2])$ be the number of scattered particle pairs P, Q, in the box $\mathcal{C}(L)$ during the time interval $[-\frac{T}{2}, \frac{T}{2}]$, whose scattering angle lies in the interval $[\vartheta_1, \vartheta_2]$.

By definition, the relativistic relative velocity between the incoming particles P and Q is equal to

$$v_{\text{rel}} := \frac{\sqrt{p_{\text{in}} q_{\text{in}} - m_0^4}}{E_{\mathbf{p}_{\text{in}}} E_{\mathbf{q}_{\text{in}}}}.$$

Equivalently, $v_{\text{rel}} = 2|\mathbf{p}_{\text{in}}|/\sqrt{m_0^2 + \mathbf{p}_{\text{in}}^2}$. If $|\mathbf{p}_{\text{in}}|$ is small, then we get the approximation $v_{\text{rel}} = 2|\mathbf{p}_{\text{in}}|/m_0$, which is the non-relativistic relative velocity. If we introduce the current density

$$j_{\text{in}} := v_{\text{rel}} \varrho_P = \frac{v_{\text{rel}} N_{\text{in}}}{\mathcal{V}},$$

then it follows from (15.25) that the total cross section is given by

$$\boxed{\sigma_{\text{tot}} = \frac{N_{\text{out}}}{j_{\text{in}} T}.}$$

Motivation of the first key formula. To simplify notation, we introduce the quantity $\mathcal{M}$ by setting

$$\langle a^\dagger_{\mathbf{p}_{\text{out}}} a^\dagger_{\mathbf{q}_{\text{out}}} \Phi_0 | \mathsf{S}(T) a^\dagger_{\mathbf{p}_{\text{in}}} a^\dagger_{\mathbf{q}_{\text{in}}} \Phi_0 \rangle = (2\pi)^4 \mathrm{i} \mathcal{N}_{\mathbf{p}_{\text{in}}} \mathcal{N}_{\mathbf{q}_{\text{in}}} \mathcal{N}_{\mathbf{p}_{\text{out}}} \mathcal{N}_{\mathbf{q}_{\text{out}}} \times \delta_{\text{dis}}(p_{\text{in}} + q_{\text{in}} - p_{\text{out}} - q_{\text{out}})\, \mathcal{M}.$$

Here, we use the normalization factor $\mathcal{N}_{\mathbf{p}} := 1/\sqrt{2\mathcal{V} E_{\mathbf{p}}}$. For example, it follows from Theorem 15.8(i) on page 830 that $\mathcal{M} = 24\kappa$, in first-order perturbation theory. In the general case, $\mathcal{M}$ depends on $\mathbf{q}_{\text{out}}$ and $\mathbf{p}_{\text{out}}$. We write

$$\mathcal{M} = \mathcal{M}(\mathbf{p}_{\text{out}}, \mathbf{q}_{\text{out}}).$$

By (15.24) on page 838,

$$N_{\text{out}} = \mathcal{W}_{\text{tot}}(T, \mathcal{V}) N_{\text{in}}$$

where

$$\mathcal{W}_{\text{tot}}(T, \mathcal{V}) = \sum\nolimits'_{\mathbf{p}_{\text{out}} \in \mathcal{G}(N)} \frac{\mathcal{V} \Delta^3 \mathbf{p}}{(2\pi)^3} \sum\nolimits'_{\mathbf{q}_{\text{out}} \in \mathcal{G}(N)} \frac{\mathcal{V} \Delta^3 \mathbf{p}}{(2\pi)^3} \cdot \gamma$$

along with

$$\gamma := (2\pi)^8 \mathcal{N}^2_{\mathbf{p}_{\rm in}} \mathcal{N}^2_{\mathbf{q}_{\rm in}} \mathcal{N}^2_{\mathbf{p}_{\rm out}} \mathcal{N}^2_{\mathbf{q}_{\rm out}} \cdot \delta_{\rm dis}(p_{\rm in} + q_{\rm in} - p_{\rm out} - q_{\rm out})^2 \, |\mathcal{M}|^2.$$

It remains to compute the discrete integrals.

(I) Square of the discrete Dirac delta function. By (15.5) on page 816,

$$\delta_{\rm dis}(p)^2 = \frac{\mathcal{V}T}{(2\pi)^4} \cdot \delta_{\rm dis}(p) = \frac{\mathcal{V}T}{(2\pi)^4} \cdot \delta_{\mathcal{G}(N)}(\mathbf{p}) \; \delta_T(E_{\mathbf{p}}).$$

Noting that $\mathbf{p}_{\rm in} + \mathbf{q}_{\rm in} = 0$ and $E_{\mathbf{p}_{\rm in}} = E_{\mathbf{q}_{\rm in}}$, we obtain

$$\delta_{\rm dis}(p_{\rm in} + q_{\rm in} - p_{\rm out} - q_{\rm out})^2 = \delta_{\mathcal{G}(N)}(\mathbf{p}_{\rm out} + \mathbf{q}_{\rm out}) \; \delta_T(E_{\mathbf{p}_{\rm out}} + E_{\mathbf{q}_{\rm out}} - 2E_{\mathbf{p}_{\rm in}}).$$

(II) Discrete integration over $\mathbf{q}_{\rm out}$. Note that

$$\sum_{\mathbf{q}_{\rm out} \in \mathcal{G}(N)} \Delta^3\mathbf{p} \;\; f(\mathbf{q}_{\rm out}) \delta_{\mathcal{G}(N)}(\mathbf{p}_{\rm out} + \mathbf{q}_{\rm out}) = f(-\mathbf{p}_{\rm out}).$$

Hence,

$$\mathcal{W}_{\rm tot}(T, \mathcal{V}) = \sum_{\mathbf{p}_{\rm out} \in \mathcal{G}(N)} \Delta^3\mathbf{p} \;\; \frac{T}{64\pi^2 \mathcal{V} E^2_{\mathbf{p}_{\rm out}} E^2_{\mathbf{p}_{\rm in}}} \cdot \delta_T(2E_{\mathbf{p}_{\rm out}} - 2E_{\mathbf{p}_{\rm in}}) \cdot |\mathcal{M}|^2.$$

Here, the symbol $\mathcal{M}$ stands for $\mathcal{M}(\mathbf{p}_{\rm out}, -\mathbf{p}_{\rm out})$.

(III) Discrete integration over $\mathbf{p}_{\rm out}$. Set $r := |\mathbf{p}_{\rm out}|$. Introducing spherical coordinates, we get

$$\Delta^3\mathbf{p} = r^2 \sum_{\Delta\varphi, \Delta\vartheta} \cos\vartheta \; \Delta\varphi \; \Delta\vartheta = 2\pi r^2 \sum_{\Delta\vartheta} \cos\vartheta \; \Delta\vartheta.$$

Recall that $v_{\rm rel} = 2|\mathbf{p}_{\rm in}|/E_{\mathbf{p}_{\rm in}}$. Hence,

$$\frac{\mathcal{W}_{\rm tot}(T, \mathcal{V})\mathcal{V}}{T v_{\rm rel}} = \sum_{\Delta\vartheta, r} g(r, \vartheta) \delta_T(f(r)) \; 2\pi \cos\vartheta \; \Delta\vartheta \; \Delta r$$

along with

$$g(r, \vartheta) := \frac{r^2 |\mathcal{M}(r, \vartheta)|^2}{128\pi^2 |\mathbf{p}_{\rm in}| E_{\mathbf{p}_{\rm out}}(r)^2 E_{\mathbf{p}_{\rm in}}}, \qquad E_{\mathbf{p}_{\rm out}}(r) = \sqrt{m_0^2 + r^2},$$

and $f(r) := 2E_{\mathbf{p}_{\rm out}}(r) - 2E_{\mathbf{p}_{\rm in}}$.

(IV) Continuum limit. We now carry out the limit $T \to \infty, \mathcal{V} \to \infty$ in a formal way. This yields

$$\sigma_{\rm tot} := \lim_{T \to \infty, \mathcal{V} \to \infty} \frac{\mathcal{W}_{\rm tot}(T, \mathcal{V})\mathcal{V}}{T v_{\rm rel}}$$

along with

$$\sigma_{\text{tot}} = \int_{-\pi/2}^{\pi/2} d\vartheta \; 2\pi \cos\vartheta \left(\int_0^\infty dr \; g(r,\vartheta)\delta(f(r)) \right).$$

The equation $f(r) = 0$ has precisely one positive zero r_0 which corresponds to the energy $E_{\mathbf{p}_{\text{out}}} = E_{\mathbf{p}_{\text{in}}}$. Explicitly, $r_0 = |\mathbf{p}_{\text{in}}|$. Using the formal rule

$$\int_0^\infty h(r)\delta(f(r))dr = \frac{h(r_0)}{|f'(r_0)|}$$

for the Dirac delta function, we obtain

$$\sigma_{\text{tot}} = \int_{-\pi/2}^{\pi/2} d\vartheta \; g(r_0,\vartheta) \cdot \frac{\sqrt{m_0^2 + r_0^2}}{2r_0} \cdot 2\pi \cos\vartheta.$$

This yields the first key formula (15.25) on page 839.

Motivation of the second key formula. We only have to replace the integral $\int_{-\pi/2}^{\pi/2} d\vartheta \ldots$ by the integral $\int_{\vartheta_1}^{\vartheta_2} d\vartheta \ldots$

15.1.9 General Feynman Rules for Particle Scattering

> Feynman diagrams represent graphically the Wick theorem.
> Folklore

The arguments used in the proof of Theorem 15.8 on page 830 can be immediately generalized to an arbitrary order of perturbation theory. Let us discuss this. We are given

- k ingoing particles of mass m_0 with pairewise disjoint momentum vectors $\mathbf{p}_{1,\text{in}}, \ldots, \mathbf{p}_{k,\text{in}}$ and
- l outgoing particles of mass m_0 with pairewise disjoint momentum vectors $\mathbf{p}_{1,\text{out}}, \ldots, \mathbf{p}_{l,\text{out}}$. Let $k + l$ be even.

We want to compute the transition amplitude

$$\boxed{\langle \Phi_{\text{out}} | \mathsf{S}(T) \Phi_{\text{in}} \rangle = \sum_{n=1}^{\infty} \kappa^n \tau_n}$$

with the unit states $\Phi_{\text{in}} := a^\dagger_{\mathbf{p}_{1,\text{in}}} a^\dagger_{\mathbf{p}_{2,\text{in}}} \cdots a^\dagger_{\mathbf{p}_{k,\text{in}}} \Phi_0$ and

$$\Phi_{\text{out}} := a^\dagger_{\mathbf{p}_{1,\text{out}}} a^\dagger_{\mathbf{p}_{2,\text{out}}} \cdots a^\dagger_{\mathbf{p}_{l,\text{out}}} \Phi_0.$$

We want to study a proper scattering process. That is, we assume that $\Phi_{\text{in}} \neq \Phi_{\text{out}}$. For fixed $n = 1, 2, \ldots$, it is our goal to determine the coefficient τ_n. Parallel to Sect. 15.1.8, the transition probability can be obtained by multiplying $|\langle \Phi_{\text{out}} | \mathsf{S}(T) \Phi_{\text{in}} \rangle|^2$ with the number of cells in the phase space.

Feynman diagrams. We will use Feynman diagrams with k ingoing particles, l outgoing particles, and m internal lines such that $4n = k + l + 2m$. Proceed as follows.

- Draw n vertices equipped with the symbols $x_1, \dots, x_n$. Each vertex has four arms.
- Select $k + l$ arms and equip them with the symbols

$$\mathbf{p}_{1,\text{in}}, \dots, \mathbf{p}_{k,\text{in}}, \mathbf{p}_{1,\text{out}}, \dots, \mathbf{p}_{l,\text{out}}.$$

  This yields the external lines corresponding to external particles.
- Consider all possible pairs of free arms which belong to different vertices and connect them. This yields the internal lines. To different internal lines, we assign different 4-momentum vectors $p, q, \dots$ In contrast to the external particles, the components of the 4-momentum vectors $p, q, \dots$ are not coupled by the energy relation. For example, p_0 is independent of $\mathbf{p}$. Therefore, we say that the internal lines correspond to virtual particles which violate the usual relation $E_{\mathbf{p}} = \sqrt{m_0^2 + \mathbf{p}^2}$ between energy and momentum of real particles.
- Introduce arrows. We attribute ingoing arrows (resp. outgoing arrows) to the momentum vectors $\mathbf{p}_{1,\text{in}}, \dots, \mathbf{p}_{k,\text{in}}$ (resp. $\mathbf{p}_{1,\text{out}}, \dots, \mathbf{p}_{l,\text{out}}$). The arrows of internal lines point from x_i to x_j where $i > j$.
- Finally, draw all possible diagrams that can be constructed by the method described above.

Feynman rules. To each Feynman diagram with n vertices we assign a product function

$$\alpha := \frac{(-\mathrm{i})^n (2\pi)^{4n}}{n!}\, abc \cdots \tag{15.27}$$

where the factors $a, b, \dots$ depend on the vertices and the internal lines. Integration over the internal lines then yields a complex number β. Finally, the desired coefficient τ_n is the sum of all possible numbers β. The number β is obtained in the following way. Let us first describe the factors $a, b, c, \dots$ from (15.27).

- An ingoing particle with momentum vector $\mathbf{p}_{\text{in}}$ corresponds to the factor

$$\mathcal{N}_{\mathbf{p}_{\text{in}}} := \frac{1}{\sqrt{2\mathcal{V} E_{\mathbf{p}_{\text{in}}}}}$$

  where $E_{\mathbf{p}_{\text{in}}} := \sqrt{m_0^2 + \mathbf{p}_{\text{in}}^2}$.
- An outgoing particle with momentum vector $\mathbf{p}_{\text{out}}$ corresponds to the factor $\mathcal{N}_{\mathbf{p}_{\text{out}}}$.
- Each vertex x_k generates the factor

$$\delta_{\text{dis}} \left(\sum_{j=1}^{4} \pm p_j \right)$$

  where p_1, p_2, p_3, p_4 are the 4-momentum vectors of the particles at the vertex x_k. The positive (resp. negative) sign belongs to ingoing (resp. outgoing) particles.

- Each internal line with 4-momentum vector p corresponds to the factor

$$\frac{1}{p^2 - m_0^2 + \mathrm{i}\varepsilon}$$

with the small regularization parameter $\varepsilon > 0$.

- Finally, we integrate the product function α. Explicitly, to each internal line with 4-momentum vector $p = (p_0, \mathbf{p})$, we assign the 4-dimensional integration

$$\int_{-E_{\max}}^{E_{\max}} dp_0 \sum_{\mathbf{p} \in \mathcal{G}(N)} \Delta^3 \mathbf{p} \ldots$$

Summarizing, concerning the m internal lines, we get an integration of dimension $4m$.

Symmetry factors. Two Feynman diagrams are called equivalent iff they only differ by a permutation of the vertices. Equivalent Feynman diagrams make the same contribution to the transition amplitude. To simplify the approach, we only consider one representative of each class of equivalent Feynman diagrams. We compensate this by cancelling the factor $n!$ from (15.27).

It is possible that additional symmetries of the Feynman diagrams are responsible for the fact that further Feynman diagrams make the same contribution to the transition amplitude. This simplifies the computation by taking additional symmetry factors into account. For example, the $4! = 24$ permutations of the four arms of Fig. 15.2(a) on page 829 generate the symmetry factor 24.

Formal continuum limit. To simplify the approach, physicists pass over to the formal continuum limit. To this end, they use the Feynman rules above by applying the replacements

$$\delta_{\mathrm{dis}}(p) \Rightarrow \delta(p)$$

and

$$\int_{-E_{\max}}^{E_{\max}} dp_0 \sum_{\mathbf{p} \in \mathcal{G}(N)} \Delta^3 \mathbf{p} \ldots \Rightarrow \int_{\mathbb{R}^4} d^4 p \ldots$$

when computing transition amplitudes. For obtaining transition probabilities, physicists use the formal rule

$$\{\delta^4(p)\}^2 = \frac{\mathcal{V}T}{(2\pi)^4}\, \delta^4(p).$$

15.1.10 The Magic Gell-Mann–Low Reduction Formula for Green's Functions

The Dyson formula (15.12) on page 822 for the S-matrix operator S can be used in order to construct the full n-point Green's function for $n = 1, 2, \ldots$. Explicitly,

$$\boxed{G_n(x_1, x_2, \dots, x_n) = \frac{\langle 0|\mathcal{T}\{\varphi_{\text{free}}(x_1)\varphi_{\text{free}}(x_2)\cdots\varphi_{\text{free}}(x_n)\mathsf{S}\}|0\rangle}{\langle 0|\mathsf{S}|0\rangle}.} \tag{15.28}$$

Here, the symbols φ_{free} and $|0\rangle$ denote the free quantum field and the ground state of the free quantum field, respectively. In addition, $\mathcal{T}\{\dots\}$ represents the chronological operator. Formula (15.28) is motivated by the rigorous discrete formula from Theorem 7.33 on page 427, which is a consequence of Wick rotation. Generalizing this theorem in a formal sense, we write

$$G_n(x_1, \dots, x_n) = \lim_{\varepsilon \to +0} \lim_{T \to +\infty} \mathcal{G}_n(x_1, \dots, x_n; T, \varepsilon) \tag{15.29}$$

along with the regularized full n-point Green's function

$$\mathcal{G}_n(x_1, \dots, x_n; T, \varepsilon) := \frac{\langle 0|\mathcal{T}\{\varphi_{\text{free}}(x_1)\cdots\varphi_{\text{free}}(x_n)\mathsf{S}(T(1-\mathrm{i}\varepsilon))\}|0\rangle}{\langle 0|\mathsf{S}(T(1-\mathrm{i}\varepsilon))|0\rangle}.$$

Formula (15.29) is called the Gell-Mann–Low reduction formula for Green's functions. In the sense of Sect. 7.22 on page 424, we will also write

$$G_n(x_1, \dots, x_n) = \langle 0_{\text{int}}|\mathcal{T}(\varphi(x_1)\cdots\varphi(x_n))|0_{\text{int}}\rangle$$

where the symbols φ and $|0_{\text{int}}\rangle$ formally denote the interacting quantum field and the ground state of the interacting quantum field, respectively. A critical discussion of these notions can be found in Sect. 13.5 on page 748.

15.2 A Glance at Quantum Electrodynamics

The application of the operator approach to quantum electrodynamics including Feynman rules and renormalization will be studied in Volume II.

> *We will use creation and annihilation operators for electrons, positrons, and photons.*

Let us only mention the following points concerning renormalization. In quantum electrodynamics, the original bare mass m_e and the bare charge $-e$ of the electron are replaced by the renormalized electron mass m_{ren} and the renormalized electron charge $-e_{\text{ren}}$ which coincide with the values measured in physical experiments. The bare quantities m_e and $-e$ only appear at the very beginning of the renormalization process, but they are eliminated after renormalization. We call this mass and charge renormalization. The physical philosophy behind this reads as follows:

- The original Lagrangian density does not include quantum fluctuations.
- In contrast to the classical theory, mass and electric charge of the electron are the result of complicated quantum interaction processes. This yields the values m_{ren} and e_{ren} which are observed in physical experiments.

- The method of renormalization allows us to introduce renormalized parameters m_{ren} and $-e_{\text{ren}}$ along with so-called radiative corrections to Feynman diagrams. This way, we get close agreement with physical experiments. For example, this allows us to compute the fine structure of the spectrum of the hydrogen atom (the Lamb shift) and the anomalous magnetic moment of both the electron and the muon.

The renormalized parameters m_{ren} and $-e_{\text{ren}}$ are also called effective parameters.

15.3 The Role of Effective Quantities in Physics

> An important task of physics is to compute effective physical constants which average microphysical effects on a macroscopic scale.
>
> Folklore

There is a fundamental phenomenon which appears in all fields of physics:

Physical constants change under interactions.

This means that the interactions are of such a nice type that the typical feature of the theory remains unchanged, but only a few physical constants change their values. We use the term effective (or renormalized) physical constants. As a classical example, consider the propagation of light in a vacuum. The speed of light is given by

$$c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}$$

where ε_0 and μ_0 are the electric and magnetic field constants of vacuum, respectively. There are many materials where the interaction between light and the molecules of the material is of such a type that the speed of light in the material is given by

$$c_{\text{material}} = \frac{1}{\sqrt{\varepsilon \mu}}.$$

Here, $\varepsilon = \varepsilon_{\text{rel}} \varepsilon_0$ and $\mu = \mu_{\text{rel}} \mu_0$ are the electric and magnetic field constants of the material, respectively. More precisely, the Maxwell equations do not change their structure. On a microscopic level, there are complicated interactions between the electromagnetic field and the molecules of the material. Fortunately enough, on a macroscopic level, this can be described by the simple replacement

$$\varepsilon_0 \Rightarrow \varepsilon, \qquad \mu_0 \Rightarrow \mu$$

in the Maxwell equations. The finite factors

$$\varepsilon_{\text{rel}} = 1 + \chi_{\text{el}}, \qquad \mu_{\text{rel}} = 1 + \chi_{\text{magn}}$$

are called renormalization factors. Moreover, χ_{el} and χ_{magn} are called electric and magnetic susceptibility, respectively.[8]

The point is that the renormalization factors are infinite in quantum field theory.

This dramatically complicates the situation and calls for a very careful mathematical approach.

15.4 A Glance at Renormalization

The subject of interacting quantum fields is full of nonsense.[9]
Paul Dirac, 1981

Renormalization theory is a notoriously complicated and technical subject... I want to tell stories with a moral for the earnest student: Renormalization theory has a history of egregious errors by distinguished savants. It has a justified reputation for perversity; a method that works up to 13th order in the perturbation theory fails in the 14th order. Arguments that sound plausible often dissolve into mush when examined closely. The worst that can happen often happens. The prudent student would do well to distinguish sharply between what has been proved and what has been plausible, and in general he should watch out![10]

Artur Wightman, 1976

In 1951 Matthews and Salam[11] formulated a requirement for renormalization procedures that has become popularly known as the *Salam criterion*: "The difficulty, as in all this work, is to find a notation which is both concise and intelligible to at least two persons, of whom one may be an author." Possibly there are many proofs of the renormalizability of quantum electrodynamics which satisfy the Salam criterion. But we must confess that none of us has yet qualified as that other person who is the guarantor of the criterion. While there are today many standard texts which discuss the renormalizability of quantum electrodynamics, we are not aware of any which represents a complete proof and in particular justifies the claim that only gauge invariant counterterms are required. We here submit to you a direct and complete proof and we invite you to judge whether you can vouch for the Salam criterion.[12]

Joel Feldman, Thomas Hurd, Lon Rosen, and Jill Wright, 1988

[8] For a thorough discussion of constitutive laws, we refer to R. Balian, From Microphysics to Macrophysics, Springer, Berlin, 1991.

[9] P. Dirac, Does renormalization make sense? In: D. Duke and J. Owens (Eds.), Perturbative Quantum Chromodynamics, Amer. Institute of Physics, New York, 1981.

[10] A. Wightman, Orientation. In: Renormalization Theory, pp. 1–20. Edited by G. Velo and A. Wightman (Eds.), Reidel, Dordrecht, 1976 (reprinted with permission). We recommend this volume as an introduction to renormalization theory which covers a broad spectrum of topics.

[11] Rev. Mod. Phys. **23** (1951), 311–314.

[12] J. Feldman, T. Hurd, L. Rosen, and J. Wright, QED (Quantum Electrodynamics): A Proof of Renormalizability, Springer, Berlin 1988 (reprinted with permission).

> I shall roughly divide the history of renormalization theory into two main chapters.
> First the structure of the infinities or "divergences" in physical quantum field theory such as electrodynamics was elucidated. A recursive process, due to Bogoliubov and his followers,[13] was found to hide these infinities into unobservable "bare " parameters that describe the fundamental laws of physics at experimentally inaccessible extremely short distances. Although technically very ingenious, this solution left many physicists and probably most mathematicians under the impression that a real difficulty had been "pulled under the rug".
>
> It would be unfortunate however to remain under this impression. Indeed the second chapter of the story, known under the curious and slightly inaccurate name of the "renormalization group", truly solved the difficulty. It was correctly recognized by Wilson and followers that in a quantum theory with many scales involved, the change of parameters from bare to renormalized values is a phenomenon too complex to be described in a single step.
> Just like the trajectory of a complicated dynamical system, it must be instead studied step by step through a local evolution rule. The change of scale in the renormalization group plays the role of time in dynamical systems. This analogy is deep. There is a natural arrow of time, related to the second principle of thermodynamics, and there is similarly a natural arrow for the renormalization group evolution: microscopic laws are expected to determine macroscopic laws, not the converse. The renormalization group erases unnecessary detailed short scale information...
> If we consider the universal character of the action principle both at the classical and quantum level, and observe that the relation between microscopic and macroscopic laws is perhaps the most central of all physical questions, it is probably not an exaggeration to conclude that the renormalization group is in some deep sense the "soul" of physics.[14]
>
> Vincent Rivasseau, 2002

Renormalization theory will play a crucial role in the following volumes. At this point, we only want to discuss a few basic ideas. As an introduction into renormalization theory formulated in the language of physicists, we recommend the textbooks by Nash (1978), Collins (1984), Veltman (1995), Kugo (1997), Ryder (1999), and Zinn–Justin (2004). Renormalization theory in terms of mathematics can be found in Manoukian (1983) and Rivasseau (1991).

[13] This is called the Bogoliubov–Parasiuk–Hepp–Zimmermann (BPHZ) approach to renormalization based on the introduction of counterterms which regularize the divergent integrals by subtracting suitable integrands.

[14] V. Rivasseau, An introduction to renormalization. In: B. Duplantier and V. Rivasseau (Eds.), Poincaré Seminar 2002: Vacuum Energy – Renormalization. Birkhäuser, Basel, 2003, pp. 139–177 (reprinted with permission).

15.4.1 The Trouble with the Continuum Limit

To illustrate the typical situation, let us consider the lattice φ^4-model from Sect. 15.1.2 on page 817. In order to get the continuum model, we have to carry out the following limits.[15]

(i) High energy limit: $N \to +\infty$. This implies $E_{\max} \to +\infty$.
(ii) Low-energy limit: $\mathcal{V} \to +\infty$. This implies $\Delta p \to +0$.
(iii) Large-time limit: $T \to +\infty$.
(iv) Regularization limit: $\varepsilon \to +0$.

This corresponds to a passage from a finite number to an infinite number of degrees of freedom. Since light of low (resp. high) energy is violet (resp. red), the high-energy (resp. the low-energy) limit is also called the ultraviolet (resp. infrared) limit by physicists. The trouble is that, as a rule, these limits do not exist.

> *To overcome this trouble, the main idea is to change the classical Lagrangian by adding counterterms.*

The mathematical prototypes of this technique are the Weierstrass theorem and the Mittag–Leffler theorem considered in Sect. 8.5.1 on page 509. From the physical point of view, the philosophy is that there arise additional quantum fluctuations in a quantum field theory. Such quantum fluctuations can be described by the counterterms of the classical Lagrangian density.

15.4.2 Basic Ideas of Renormalization

The crucial point is that from the physical point of view, renormalization theory allows us to pass from microscopic quantities to macroscopic quantities. In contrast to the microscopic quantities, the macroscopic quantities can be measured in physical experiments which depend on the available scale (e.g., the energy scale). In this context, the following two ideas play the decisive role:

(i) The idea of counterterms.
(ii) The idea of essential and inessential scales.

Let us briefly discuss this.

Counterterms. As a prototype, consider the φ^4-model. Here, we replace the classical Lagrangian density

$$\mathcal{L}(\varphi, \partial\varphi) := \tfrac{1}{2}\varphi(-\Box - m_0^2 + \mathrm{i}\varepsilon)\varphi - \kappa\varphi^4 \tag{15.30}$$

by the renormalized Lagrangian density

[15] Recall that N is the number of lattice points, $\mathcal{V}$ is the volume of the box, and $\Delta p = 2\pi/L$ with $\mathcal{V} = L^3$. Moreover, we use the time interval $[-\frac{T}{2}, \frac{T}{2}]$.

$$\boxed{\mathcal{L}_{\text{ren}} = \tfrac{1}{2}\varphi_{\text{ren}}(-\Box - m_{\text{ren}}^2 + \mathrm{i}\varepsilon)\varphi_{\text{ren}} - \kappa_{\text{ren}}\lambda^\delta \varphi_{\text{ren}}^4 + \mathcal{L}_{\text{counter}}} \quad (15.31)$$

along with the counterterms

$$\begin{aligned}\mathcal{L}_{\text{counter}} := \ & \tfrac{1}{2}(1 - Z_\varphi)\varphi_{\text{ren}}\Box\varphi_{\text{ren}} + \\ & + \tfrac{1}{2}m_{\text{ren}}^2(1 - Z_m)\varphi_{\text{ren}}^2 + (1 - Z_\kappa)\kappa_{\text{ren}}\lambda^\delta\varphi_{\text{ren}}^4.\end{aligned}$$

In order to motivate this, we want to sketch the method of dimensional renormalization. This method was introduced by 't Hooft and Veltman in order to renormalize the electroweak Standard Model in particle physics.[16] We start with the classical Lagrangian density $\mathcal{L}$ in 4-dimensional spacetime (Minkowski space $\mathbb{M}^4$) given in (15.30). In each order of perturbation theory, we use the Feynman rules for the S-matrix elements. This way, we get discrete integrals over the lattice in momentum space. Furthermore, the formal continuum limit

$$N \to +\infty, \quad \mathcal{V} \to +\infty, \quad \Delta p \to +0$$

sends the discrete integrals to integrals over the 4-dimensional Minkowski space. The point is that some of these integrals are divergent. This is caused by the fact that the integrands decrease too slowly at infinity. To overcome this difficulty, physicists proceed in the following two steps.

Step 1: Dimensional regularization of integrals. We pass to the n-dimensional Minkowski space $\mathbb{M}^n$. To this end, we modify the Lagrangian density by setting

$$\mathcal{L}_n(\varphi, \partial\varphi) := \tfrac{1}{2}\varphi(-\Box - m_0^2 + \mathrm{i}\varepsilon)\varphi - \kappa\lambda^{4-n}\varphi^4 \quad (15.32)$$

with the parameter $\lambda > 0$ and $n < 4$. It is convenient to work in the momentum space (Fourier space). For low dimension $n = 1$, the integrals are convergent. The point is that the integrals $I(n)$ lead to functions which can be continued analytically with respect to the complex variable n. For

$$n = 4 - \delta,$$

we get analytic expressions with respect to δ which posses singularities at $\delta = 0$; this corresponds to the physically relevant dimension $n = 4$. A simple example for dimensional regularization of algebraic Feynman integrals can be found in Sect. 11.6.2 on page 636.

Step 2: Compensation for singularities by introducing counterterms into the Lagrangian density. Explicitly, we start with the bare Lagrangian density $\mathcal{L}_n$ from (15.32), and we set

$$m_{\text{ren}} := (m_0 + \Delta m)\sqrt{Z_\varphi}, \qquad \kappa_{\text{ren}} = (\kappa + \Delta\kappa)Z_\varphi^2, \qquad \varphi_{\text{ren}} = \frac{\varphi}{\sqrt{Z_\varphi}}.$$

[16] G. 't Hooft and M. Veltman, Nuclear Physics B **44** (1972), 189–213.

Here, Δm and $\Delta\kappa$ are real parameters which correct the bare mass m_0 and the bare coupling constant κ, respectively. We call $m_{\text{ren}}, \kappa_{\text{ren}}$, and φ_{ren} the renormalized mass, the renormalized coupling constant, and the renormalized quantum field, respectively. Using multipliers, we write

$$m_{\text{ren}} = \frac{m_0\sqrt{Z_\varphi}}{\sqrt{Z_m}}, \qquad \kappa_{\text{ren}} = \frac{\kappa Z_\varphi^2}{Z_\kappa}.$$

Hence

$$\varphi = \varphi_{\text{ren}}\sqrt{Z_\varphi}, \quad m_0^2 = m_{\text{ren}}^2\frac{Z_m}{Z_\varphi}, \quad \kappa = \kappa_{\text{ren}}\frac{Z_\kappa}{Z_\varphi^2}.$$

Using this substitution, the original Lagrangian density $\mathcal{L}_n(\varphi, \partial\varphi, m_0)$ is transformed into the renormalized Lagrangian density

$$\mathcal{L}_{\text{ren}}(\varphi_{\text{ren}}, \partial\varphi_{\text{ren}}, m_{\text{ren}})$$

from (15.31) above. The point is that the counterterms can be chosen in such a way that the divergent terms can be cancelled after carrying out the limits $\delta \to 0$ and $\varepsilon \to 0$. The final result only depends on

- the renormalized meson mass m_{ren} and
- the renormalized coupling constant κ_{ren}.

The renormalized mass m_{ren} and the renormalized coupling constant κ_{ren} are unknown finite quantities. They have to be determined by physical experiments.[17]

Observe the following peculiarity. The multiplicative renormalization constant Z_m depends on δ, and we have the limit $Z_m \to +\infty$ as $\delta \to +0$. This means that $m_0 \to +\infty$. Consequently, it is impossible to determine the parameter m_0 from the renormalized mass m_{ren} measured in a physical experiment. This means that the original bare mass parameter m_0 has no physical meaning at all. The following remarks are in order:

- The counterterms lead to additional Feynman diagrams.
- The procedure of renormalization is not uniquely determined. However, the experience of physicists shows that different renormalization methods yield the same physics.
- There exist rescalings of the renormalized Lagrangian density which correspond to the same S-matrix. Such rescalings form the so-called renormalization group, which is extremely useful.

The method of dimensional renormalizaton method is widely used by physicists in elementary particle physics. Interestingly enough, this method also works extremely well for computing higher-order post-Newtonian approximations in gravitational physics (Einstein's theory of general relativity). Such computations are necessary in order to compute the collision of binary neutron stars or binary black holes. This can be found in the paper by

[17] Explicit computations can be found in Ryder (1999), pp. 308–329.

T. Damour, P. Jaranowski, and G. Schäfer, Dimensional regularization of the gravitational interaction of point masses, Physics Letters B **513** (2001), 147–155.

We also recommend the survey article by G. Schäfer, Gravitation: geometry and dynamics, Ann. Phys. (Leipzig) **14** (2005), 148–164.

Essential and inessential scales. If one wants to pass from microscopic quantities to macroscopic quantities, one has to distinguish between the essential scale and the inessential scale. The behavior of the physical system on the inessential scale can be replaced by suitable averages. This idea is used systematically in the method of the renormalization group.[18]

Methods of renormalization theory. The most important methods in renormalization theory read as follows:

(a) BPHZ renormalization and the Weinberg power-counting theorem.
(b) Pauli–Villars regularization by introducing fictitious masses.
(c) Dimensional regularization.
(d) The BRST symmetry and algebraic renormalization.
(e) The renormalization group approach.
(f) The Epstein–Glaser approach.
(g) The Zimmermann forest formula and the importance of Hopf algebras.
(h) Gauge symmetries of functional integrals and the Ward–Takehashi identities and the Taylor–Slavnov identities for Green's functions.

In what follows, we will discuss some basic ideas.

15.4.3 The BPHZ Renormalization

The important BPHZ renormalization method due to Bogoliubov, Parasiuk, Hepp, and Zimmermann proceeds in the following steps.

(i) Feynman rules: We start with our lattice approach. In each order of perturbation theory, we get well-defined discrete integrals for the S-matrix elements. Carrying out the formal continuum limit, we obtain multi-dimensional integrals over products of the 4-dimensional momentum space. The point is that some of the integrals are divergent, because the integrands decrease too slowly at infinity.
(ii) Regularization of divergent integrals: We subtract regularization terms from the integrands of the divergent integrals. These regularization terms correspond to suitable first terms of the Taylor expansion of the integrands. Prototypes can be found in Sect. 8.5.3 on page 511. The convergence of the regularized integrals is guaranteed by Weinberg's power-counting theorem.
(iii) Renormalization: Add counterterms to the classical Lagrangian density in order to compensate for the regularization terms. This leads to renormalized macroscopic physical parameters.

[18] We recommend Duplantier and Rivasseau (2003).

If this approach works well, the quantum field theory is called renormalizable. The basic papers of BPHZ renormalization are the following:

N. Bogoliubov and O. Parasiuk, On the multiplication of propagators in quantum field theory (in German), Acta Math. **97** (1957), 227–326.

S. Weinberg, High energy behavior in quantum field theory, Phys. Rev. **118** (1969) 838–849 (the power-counting theorem).

K. Hepp, Proof of the Bogoliubov–Parasiuk theorem on renormalization, Commun. Math. Phys. **2** (1966), 301–326.

W. Zimmermann, Convergence of Bogoliubov's method of renormalization in momentum space, Commun. Math. Phys. **15** (1969), 208–234.

G. 't Hooft, Renormalization of massless Yang-Mills fields, Nuclear Phys. B **33** (1971), 173–199.

G. 't Hooft and M. Veltman, Regularization and renormalization of gauge fields, Nucl. Phys. B **44** (1972), 189–213.

G.'t Hooft and M. Veltman, Diagrammar, CERN, Diagrammar, CERN, Report 73/9 (1973).

Internet: `http://doc.cern.ch/yellowrep/1973/1973-009/p1.pdf`

The mathematics of the BPHZ renormalization method can be found in

I. Manoukian, Renormalization, Academic Press, New York, 1983.

We also recommend the lectures given by

P. Cartier, Mathemagics: A tribute to L. Euler and R. Feynman, Séminaire Lotharingien **44** (2000), 1–71.

15.4.4 The Epstein–Glaser Approach

In the BPHZ approach, there arise divergent integrals which have to be regularized. In 1973 Epstein and Glaser developed an alternative approach to quantum field theory which completely avoids divergent integrals for high energies.[19] In terms of mathematics, this is an advantage of the Epstein–Glaser approach.

By using the mathematical theory of tempered generalized functions, ill-defined quantities never appear.

We will study this in Volume II. The basic ideas read as follows.

(i) The iterative method for the generalized S-matrix: As the fundamental object, we choose the generalized S-matrix,

$$S = S(g) \qquad \text{for all} \quad g \in \mathcal{S}(\mathbb{R}^4).$$

This is an operator-valued tempered generalized function. That is, to each test function $g \in \mathcal{S}(\mathbb{R}^4)$, we assign a linear operator

[19] H. Epstein and V. Glaser, The role of locality in perturbation theory, Ann. Inst. Poincaré A **19**(3) (1973), 211–295.

$$S(g) : X \to X$$

on the complex Hilbert space X. Using first physical principles (e.g., causality and pseudo-unitarity), we get an iterative method for computing the S-matrix,

$$\boxed{S(g) = I + \kappa S_1(g) + \kappa^2 S_2(g) + ...,} \tag{15.33}$$

in each order of perturbation theory. Here, κ denotes the coupling constant which measures the strength of the interaction. The first-order approximation $S_1(g)$ is determined by the interaction term of the Lagrangian density of the corresponding classical field theory. It is crucial that

> *All of the higher-order terms $S_2(g), S_3(g), ...$ depend on the first-order term $S_1(g)$ by using the iterative method mentioned above.*

It turns out that the generalized S-matrix $S = S(g)$ contains all the information about the interacting quantum field under consideration. Explicitly, this concerns the computation of the following quantities:

- cross sections for scattering processes of elementary particles,
- decay rates of elementary particles,
- energies of bound states of elementary particles, and
- interacting quantum fields Ψ.

Therefore, it remains to compute the components $S_2, S_3, ...$ of the generalized S-matrix.

(ii) Regularization of tempered distributions: The following observation is crucial for understanding renormalization theory. To emphasize the basic idea, let us simplify the following considerations by dropping out details. Roughly speaking, in each order of perturbation theory, the iterative method determines the generalized S-matrix only up to a finite linear combination of Dirac's delta distribution and its derivatives, that is, we obtain

$$\boxed{S_n = (S_n)_{\text{reg}} + c_{n0}\delta + c_{n1}\delta' + ... + c_{nm}\delta^{(m)}.} \tag{15.34}$$

In other words, first physical principles do not determine the generalized scattering matrix S in a unique manner, but only up to certain generalized functions which are not classical functions. Equivalently, this means that, after Fourier transformation, the S-matrix component S_n is determined up to a polynomial of order m in momentum space. [20] The

[20] In contrast to the classical 1949 Dyson approach, the Epstein–Glaser approach avoids the ill-defined time-ordered product based on the crucial chronological operator $\mathcal{T}$. The singular part of (15.34) is caused by the singularities of the Green's functions.

expression (15.34) is obtained in the following way. We first restrict ourselves to test functions g which vanish outside a small disc D_r of radius r about the origin. This choice of test functions allows us to perform the iterative method in a natural way. We then study the limit $r \to 0$ by using the sophisticated Steinmann renormalization theorem from Sect. 11.4.3 on page 620.

(iii) High-energy (ultraviolet) renormalization: The generalized S-matrix $S(g)$ depends on the free constants $c_{n0}, c_{n1}, ...$ along with physical parameters, e.g., the bare electron mass m_e and the bare electron charge $-e$ in quantum electrodynamics. Our goal is to replace $c_{n0}, ..., c_{nm}, m_e, -e$ by effective quantities that can be measured in a physical experiment, e.g., the renormalized electron mass m_{ren} and the renormalized electron charge $-e_{\text{ren}}$. To this end, we need additional physical information. In this connection, the Ward identities and their generalizations play a crucial role. These identities follow from additional quantum symmetries (gauge symmetries).

(iv) Low-energy (infrared) renormalization: Note that the constant function $g \equiv 1$ does not lie in the space $\mathcal{S}(\mathbb{R}^4)$ of test functions. In terms of physics, this means that the approach (i) above corresponds to quantum fields which are contained in a box of finite volume $\mathcal{V}$. It remains to study the limit $\mathcal{V} \to +\infty$. In terms of test functions, this corresponds to the limit

$$\boxed{\lim_{k\to\infty} g_k(x) = 1 \qquad \text{for all} \quad x \in \mathbb{R}^4}$$

where the test functions $g_1, g_2, ...$ lie in the space $\mathcal{S}(\mathbb{R}^4)$. Finally, we have to study the limit $\lim_{k\to\infty} S(g_k)$. For example, in quantum electrodynamics, infrared renormalization corresponds to the limit

$$\boxed{\lambda \to +\infty}$$

where λ denotes the photon wave length. Such a limit is not always reasonable from the physical point of view. In particular, if the universe has a finite volume, then there exists a natural bound R for the photon wave length,

$$\lambda \leq R,$$

where R denotes the maximal distance within the universe. Here, R is a natural cut-off for the photon wave length λ. In this case, infrared infinities do not appear. It is thinkable that the infrared problem will be settled in the framework of a unified theory for all four fundamental interactions which relates the global structure of the universe to local quantum physics. For concrete physical situations in quantum electrodynamics, infrared divergences do not arise, since they are compensated for by so-called braking radiation (bremsstrahlung) caused by low-energy photons. More precisely, the divergences only vanish after summing over all of the Feynman diagrams which are related to this process.

(v) Computation of quantum fields: Suppose that we know the generalized S-matrix

$$S = S(g, J)$$

as a functional of the test function g and the external source J. Then, the corresponding interacting quantum field Φ can be obtained by means of the following functional derivative,

$$\boxed{\Phi(\chi, J) := S(0, J)^{-1} \frac{\delta S(0, J)}{\delta g}(\chi).} \tag{15.35}$$

This definition dates back to Bogoliubov. In this setting, the interacting quantum field Φ is an operator-valued generalized function. That is, to each test function $\chi \in \mathcal{S}(\mathbb{R}^4)$ and each external source J, we assign the linear operator

$$\Phi(\chi, J) : X \to X$$

on the Hilbert space X.

(vi) The convergence problem in perturbation theory: In 1951 Dyson invented a heuristic physical argument which suggested that the perturbation series (15.33) does not converge for small values of the coupling constant κ. However, the classical 1916 Ritt theorem in complex function theory tells us that each formal power series of the form (15.33) can be viewed as an asymptotic series of some function $f = f(\kappa)$ which is analytic on a circular sector in the complex plane. Here, the sector contains sufficiently small values $\kappa > 0$ of the coupling constant κ. This will be discussed in Sect. 15.5 on page 860.

A detailed application of the Epstein–Glaser approach to quantum electrodynamics including Bogoliubov's formula (15.35) can be found in the monograph by

G. Scharf, Finite Quantum Electrodynamics: the Causal Approach, Springer, New York, 1995.

Summarizing, the Epstein–Glaser approach tells us that renormalization theory is related in a quite natural way to the theory of tempered distributions.

Historical remarks. Concerning the Bogoliubov formula (15.35) for defining full quantum fields in the Epstein–Glaser approach, let us make the following comment. To simplify notation, set $J = 0$ in what follows. In 1929, Heisenberg and Pauli based quantum field theory on the notion of a local operator-valued quantum field $\varphi = \varphi(x)$. Here, to each space-time point $x = (\mathbf{x}, t)$ they assigned an operator

$$\varphi(x) : X \to X$$

on the Hilbert space X. In the framework of canonical quantization, the quantum field φ has to satisfy additional commutation (resp. anticommutation) rules. However, it turns out that such a notion is contradictory from

the mathematical point of view. In 1943 Heisenberg introduced the S-matrix as a substitute for the quantum field φ. In the Epstein–Glaser approach, the quantum field Φ is not a primary object, but it can be derived from the generalized S-matrix. Formally, the relation between φ and Φ is given by

$$\Phi(\chi) = \int_{\mathbb{R}^4} \varphi(x)\chi(x)d^4x \qquad \text{for all} \quad \chi \in \mathcal{S}(\mathbb{R}^4).$$

Intuitively, the quantum field φ is a highly singular object. By forming mean values over test functions χ, we arrive at a less singular mathematical object denoted by Φ. The functional

$$\chi \mapsto \Phi(\chi)$$

on the space $\mathcal{S}(\mathbb{R}^4)$ of test functions with values in a Hilbert space is called a Hilbert space-valued generalized function (or a Wightman functional). The representation of quantum fields by such functionals was studied first by Arthur Wightman[21] in 1956.

15.4.5 Algebraic Renormalization

> In 1974 it was recognized by Becchi, Rouet, and Stora, in their work on the BRST symmetry invariance of gauge theories, that the use of the quantum action principle leads to the possibility of a fully algebraic proof of renormalizability of a theory characterized by a set of local rigid invariances... The quantum action principle allows one to control the breaking of a symmetry induced by a noninvariant subtraction scheme, helping then to give an algebraic answer of restoring the symmetry through the addition of compensating noninvariant local counterterms. It is worthwhile to emphasize that such algebraic proofs do not rely on the existence of a regularization preserving of symmetries.
>
> Olivier Piguet and Silvio Sorella, 1995
> *Algebraic Renormalization:*
> *Perturbative Renormalization, Symmetries, and Anomalies*[22]

As a rule of thumb, quantum field theories related to elementary particles are renormalizable, but the quantized gravitational force is not renormalizable in the usual setting. This is one of the fundamental problems in the foundation of a unified theory for all four fundamental forces in nature. A detailed proof of the renormalizability of the electroweak Standard Model can be found in the following paper:

> E. Kraus, Renormalization of the electroweak standard model to all orders, Ann. Phys. (NY) **262** (1998), 155–259.

[21] Quantum field theories in terms of vacuum expectation values, Phys. Rev. **101** (1956), 860–866.

[22] Springer, Berlin, 1995 (reprinted with permission).

This paper uses the elegant and very effective method of algebraic renormalization. This method combines the quantum action principle (that is, the Dyson–Schwinger equation for the full generating functional) with the method of BRST symmetry to be discussed in Sect. 16.7 on page 890.[23] The method of algebraic renormalization can also be used in order to prove the renormalizability of the simplest supersymmetric variant of the Standard Model in particle physics, called the minimal supersymmetric standard model. We refer to the basic paper by

> W. Hollik, E. Kraus, M. Roth, C. Rupp, K. Sibold, and D. Stöckinger, Renormalization of the minimal supersymmetric standard model, Nuclear Physics **B639** (2002), 3–65.

Unfortunately, renormalization proofs are highly technical and rather involved. For the renomalization of quantum electrodynamics in a Euclidean setting, we recommend

> J. Feldman, T. Hurd, L. Rosen, and J. Wright, QED: A Proof of Renormalizability, Springer, Berlin, 1988.

15.4.6 The Importance of Hopf Algebras

> Behind renormalization there lurks a monster called the motivic Galois group. This monster is responsible for the rich mathematical structure of renormalization theory.
>
> Folklore

It was discovered by Kreimer in 1994 that behind renormalization theory there exists a symmetry encoded into some Hopf algebra.[24] As an introduction, we recommend

> D. Kreimer, Knots and Feynman Diagrams, Cambridge University Press, 2000.

The crucial point is as follows. In the BPHZ renormalization method, it was discovered by Zimmermann that Bogoliubov's iterative method for determing the regularized integrals can be formulated in terms of some global forest formula which elegantly describes the structure of the additional Feynman diagrams.[25] Kreimer discovered that Zimmermann's forest formula can be understood best by using the coinverse (also called the antipode) of a suitable Hopf algebra which is related to Feynman diagrams. Furthermore, Connes

[23] We recommend O. Piguet and S. Sorella, Algebraic Renormalization, Springer, Berlin, 1995.

[24] Note that, roughly speaking, Hopf algebras are dual constructions to algebras. They appear at several places in mathematics (e.g., combinatorics, differential operators, power series expansions, representation theory of groups, and algebraic topology). See Chap. 3 of Vol. II.

[25] See the monograph by Collins (1984).

and Kreimer discovered that the basic mathematical structure of renormalization theory is closely related to the famous Riemann–Hilbert problem. We refer to

> A. Connes and H. Moscovici, Hopf algebras, cyclic cohomology and the transverse index theorem, Commun. Math. Phys. **198** (1998), 199–246.
>
> A. Connes and D. Kreimer, Hopf algebras, renormalization and noncommutative geometry, Comm. Math. Phys. **199** (1998), 203–242.
>
> A. Connes and D. Kreimer, Renormalization in quantum field theory and the Riemann–Hilbert problem I: The Hopf algebra structure of graphs and the main theorem, Commun. Math. Phys. **210** (2000), 249–273.
>
> A. Connes and D. Kreimer, Renormalization in quantum field theory and the Riemann–Hilbert problem II: The beta function, diffeomorphisms, and the renormalization group, Commun. Math. Phys. **216** (2000), 215–241.

As a survey, we recommend

> A. Connes, Symmétries galoisiennes et renormalisation. In: Duplantier and Rivasseau (2003), pp. 241–264.

The final breakthrough can be found in the papers by

> A. Connes and M. Marcolli, Renormalization and motivic Galois theory. Internet 2004: `arXiv:math.NT/0409306`
> A. Connes and M. Marcolli, From number theory to physics via noncommutative geometry, I, II.
> Internet 2005: `http://www.math.dsu.edu/ marcolli/renorm43.pdf`

The basic philosophy is the following one. The experience of physicists shows that renormalization theory is full of symmetries reflected by marvellous relations. The experience of mathematicians shows that symmetries are always governed by an appropriate symmetry group. In terms of renormalization theory, the symmetry group was discovered by Connes and Marcolli in an abstract way by using the theory of categories. This huge group is called the motivic Galois group of renormalization. The renormalization groups used by physicists in specific situations are representations of one-dimensional subgroups of the motivic Galois group.

15.5 The Convergence Problem in Quantum Field Theory

15.5.1 Dyson's No-Go Argument

In 1951 Dyson discovered a simple argument which suggested that the S-matrix series diverges. His reasoning was the following: Suppose one were to calculate a physical observable in a power series in the coupling constant, κ. If this series is convergent for some positive value of κ, it must converge in some circle of radius κ. The series must therefore also converge on some interval of the negative real axis. Now a negative coupling constant κ corresponds to a world in which like charges would attract

> one another and opposite charges would repel each other. However, if κ is negative, then a state which contains a large number N, of electron-positron pairs in which the electrons are clustered together in a region Ω_1 and the positrons are clustered together in another region Ω_2 of space far from Ω_1, would have an energy lower than that of the ground state (vacuum) for N large enough... The ground state is therefore unstable relative to such states. Since the larger the number of pairs, the more pronounced the effect becomes, the higher order terms in the power series expansion must become more and more important so that the series cannot converge... Quantum electrodynamics was over for Dyson in 1951, when he found this heuristic argument that the perturbation theory diverges.
>
> Silvan Schweber
> *QED and the Men Who Made it: Dyson, Feynman, Schwinger, and Tomonaga*[26]

15.5.2 The Power of the Classical Ritt Theorem in Quantum Field Theory

> The question of what conditions a formal power series must satisfy in order to occur as an asymptotic development has a surprisingly simple answer for circular sectors at the origin: there are no such conditions.
>
> Reinhold Remmert, 1991
> *Theory of Complex Functions*[27]

Let $a_0, a_1, ...$ be a sequence of complex numbers. The convergent or divergent power series expansion

$$a_0 + a_1\kappa + a_2\kappa^2 + ...$$

is called a formal power series expansion. The following Ritt theorem shows that each formal power series expansion is the asymptotic series of an appropriate analytic function $f = f(\kappa)$. Observe that the function f is not determined uniquely by the formal power series expansion. In physical applications, the variable κ represents the coupling constant. We are given the circular sector

$$\mathcal{C} := \{\kappa \in \mathbb{C} : 0 < |\kappa| < r,\ -\eta < \arg\kappa < \eta\}$$

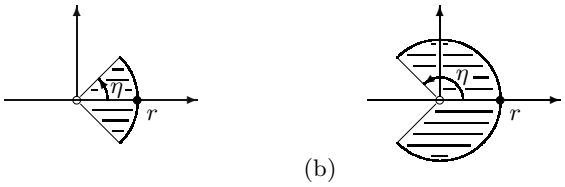
where the radius r is a fixed positive number, and the positive angle η is less than π (Fig. 15.7 on page 862).

Theorem 15.9 *To each formal power series expansion* $\sum_{n=0}^{\infty} a_n\kappa^n$, *there exists a holomorphic function* $f : \mathcal{S} \to \mathbb{C}$ *such that*

$$f(\kappa) \sim \sum_{n=0}^{\infty} a_n\kappa^n \qquad on \quad \mathcal{S},$$

in the sense of an asymptotic power series expansion.

[26] Princeton University Press, 1994 (reprinted with permission).
[27] Springer, New York, 1991 (reprinted with permission).



(a) (b)

Fig. 15.7. Circular sector

Explicitly, this means that we have

$$f(\kappa) = a_0 + a_1\kappa + \ldots + a_n\kappa^n + r_n(\kappa) \qquad \text{for all} \quad \kappa \in \mathcal{S}$$

and each $n = 0, 1, 2, \ldots$ where the remainder is of the form $r_n(\kappa) = \varepsilon_n(\kappa)\kappa^n$ with $\varepsilon_n(\kappa) \to 0$ as $\kappa \to 0$ on $\mathcal{S}$. This theorem was obtained by Ritt in 1916.[28] The proof can be found in Remmert (1991), Vol. 1, p. 300. The idea of the proof is to construct the function

$$f(\kappa) := \sum_{n=0}^{\infty} a_n c_n(\kappa)\kappa^n$$

with convergence factors of the form $c_n(\kappa) := 1 - \mathrm{e}^{-b_n/\sqrt{\kappa}}$. Observe that the function f is not holomorphic in a small neighborhood of the origin $\kappa = 0$, but only on a circular sector. The 1916 Ritt theorem is closely related to Dyson's 'no go' argument from Sect. 15.5.1.[29]

15.6 Rigorous Perspectives

Suggested reading. For a rigorous mathematical approach to the classic φ^4-model, we recommend the following two modern surveys:

P. Federbush, Quantum field theory in ninety minutes, Bull. Amer. Math. Soc. **17** (1) (1987), 93–103.

B. Duplantier and V. Rivasseau (Eds.), Poincaré Seminar 2002: Vacuum Energy – Renormalization, Birkhäuser, Basel, 2003

and the following monographs:

B. Simon, The $P(\varphi)_2$-Euclidean Quantum Field Theory, Princeton University Press, 1974.

J. Glimm and A. Jaffe, Mathematical Methods of Quantum Physics: A Functional Integral Point of View, Springer, New York, 1981.

[28] J. Ritt, On the derivatives of a function at a point, Annals of Math. **18**(2) (1916), 18–23.

[29] As an introduction to Poincaré's theory of asymptotic expansions, we recommend the monographs by Erdélyi (1965), Wasow (1965), and Ramis (1993).

J. Glimm and A. Jaffe, Quantum Field Theory and Statistical Mechanics: Expositions, Birkhäuser, Boston, 1985.

E. Seiler, Gauge Theories as a Problem of Constructive Quantum Field Theory and Statistical Mechanics, Springer, Berlin, 1982.

H. Grosse, Models in Statistical Physics and Quantum Field Theory, Springer, New York, 1988.

R. Balian, From Microphysics to Macrophysics, Springer, New York, 1991.

V. Rivasseau, From Perturbative to Constructive Renormalization, Princeton University Press, 1991.

J. Fröhlich, Non-Perturbative Quantum Field Theory: Mathematical Aspects and Applications (a collection of papers), World Scientific, Singapore, 1992.

J. Fröhlich, Scaling and Self-Similarity in Physics: Renormalization in Statistical Physics, Birkhäuser, Basel, 1993.

R. Fernández, J. Fröhlich, and D. Sokal, Random Walks, Critical Phenomena, and Triviality in Quantum Field Theory, Springer, Berlin, 1992.

G. Benfatto and G. Gallavotti, Renormalization Group, Princeton University Press, 1995.

L. Vázquez, L. Streit, and V. Pérez-García (Eds.), Nonlinear Klein–Gordon and Schrödinger Systems: Theory and Applications, World Scientific, Singapore, 1996.

I. Montvay and G. Münster, Quantum Fields on a Lattice, Cambridge University Press, 1997.

C. Kopper, Renormalization with Flow Equations (in German), Shaker, Aachen, 1998.

G. Battle, Wavelets and Renormalization, World Scientific, Singapore, 1999.

M. Salmhofer, Renormalization: An Introduction, Springer, Berlin, 1999.

J. Klauder, Beyond Conventional Quantization, Cambridge University Press, 2000.

A. Bytsenko, G. Cognola, E. Elizalde, V. Moretti, and S. Zerbini, Analytic Aspects of Quantum Fields, World Scientific, Singapore, 2003.

J. Zinn-Justin, Quantum Field Theory and Critical Phenomena, Clarendon Press, Oxford, 2003.

Much material about two-dimensional models in quantum field theory can be found in:

E. Abdalla, M. Abdalla, and K. Rothe, Non-Perturbative Methods in Two-Dimensional Quantum Field Theory, World Scientific, Singapore, 2001.

S. Lundquist, G. Morandi, Yu Lu (Eds.), Low-Dimensional Quantum Field Theories for Condensed Matter Physicists, World Scientific, Singapore, 1995.

S. Carnip, Quantum Gravitation in 2+1 Dimensions, Cambridge University Press, 1998.

P. Fre and P. Soriani, The $N = 2$ Wonderland: From Calabi–Yau Manifolds to Topological Field Theories, World Scientific, Singapore, 2004.

An impressive series of ten rigorous papers on the statistical physics of two-dimensional Fermi liquids and renormalization is accessible on the Internet:

> J. Feldman, H. Knörrer, and E. Trubowitz, A two-dimensional Fermi liquid, 2003. Internet: `http://www.math.ubc.ca/~feldman/fl.htlm`

Finally, let us mention a pedagogical survey on the relations between classical perturbation theory in celestial mechanics (the Poincaré–Lindstedt series) and modern renormalization theory:

> J. Feldman and E. Trubowitz, Renormalization in classical mechanics and many body quantum field theory, Jerusalem J. d'Analyse Mathématique **52** (1992), 213–247.
> Internet: `http://www.math.ubc.ca.~feldman/research.html`

Minkowskian versus Euclidean models. First we are going to fix the terminology. Let us choose $D = 2, 3, \ldots$ (dimension of the space-time manifold), and let $N = 3, 4, \ldots$ (order of nonlinearity). We will distinguish between the

- Minkowskian case (indefinite metric on the D-dimensional space-time), and
- the Euclidean space (definite metric on the D-dimensional space-time).

The Euclidean case is much simpler to treat mathematically than the Minkowskian case. The Euclidean case allows us to apply the well-elaborated machinery of stochastic processes and probabilistic calculus. The passage from the Minkowskian case to the Euclidean case corresponds to the passage

$$t \mapsto -\mathrm{i}t \tag{15.36}$$

from real time to imaginary time.

(M) The Minkowskian φ^N_D-model: Choose $\sigma := 1$ and $\alpha := 1$. This model is based on the following action functional

$$\boxed{S[\varphi] := \alpha(S_{\mathrm{free}}[\varphi] + \kappa S_{\mathrm{int}}[\varphi])} \tag{15.37}$$

with the free action

$$S_{\mathrm{free}}[\varphi] := \frac{1}{2}\int_{\mathbb{R}^D} \left(-\sigma\varphi_{tt} + \sum_{k=1}^{D-1} \varphi_{x_k x_k} - m_0^2\varphi^2 \right) d^D x,$$

and the interacting part of the action

$$\kappa S_{\mathrm{int}}[\varphi] := \kappa \int_{\mathbb{R}^D} (\lambda\varphi - \varphi^N)\, d^D x.$$

Here, the coupling constant $\kappa \geq 0$ and the real potential parameter λ are fixed. The Euler–Lagrange equation corresponding to the principle of critical action,

$$S[\varphi] = \text{critical!}, \qquad \varphi \in \mathcal{S}(\mathbb{R}^D)$$

reads as

$$\boxed{\sigma\varphi_{tt} - \sum_{k=1}^{D-1} \varphi_{x_k x_k} + m_0^2\varphi + N\kappa\varphi^{N-1} - \kappa\lambda = 0.} \tag{15.38}$$

This is a nonlinear Klein–Gordon equation (meson model).

(E) The Euclidean φ_D^N-model: We use the action functional $S[\varphi]$ along with equation (15.38) by setting $\sigma := -1$. According to the time transformation (15.36), the differential $d^D x$ passes over to $-\mathrm{i}d^D x$. Therefore, we choose now $\alpha := -\mathrm{i}$.

In the general case, the correlation functions read as

$$C_n(x_1, \ldots, x_n) := \frac{\int_{\mathcal{S}(\mathbb{R}^D)} \varphi(x_1) \cdots \varphi(x_n)\, \mathrm{e}^{\mathrm{i}\alpha\kappa S_{\mathrm{int}}[\varphi]} \cdot \mathrm{e}^{\mathrm{i}\alpha S_{\mathrm{free}}[\varphi]}\, \mathcal{D}\varphi}{\int_{\mathcal{S}(\mathbb{R}^D)} \mathrm{e}^{\mathrm{i}\alpha\kappa S_{\mathrm{int}}[\varphi]} \cdot \mathrm{e}^{\mathrm{i}\alpha S_{\mathrm{free}}[\varphi]}\, \mathcal{D}\varphi}$$

where $n = 1, 2, \ldots$ Setting $d\mu := \mathrm{e}^{\mathrm{i}\alpha S_{\mathrm{free}}[\varphi]}\mathcal{D}\varphi$, we get

$$C_n(x_1, \ldots, x_n) := \frac{\int_{\mathcal{S}(\mathbb{R}^D)} \prod_{k=1}^n \varphi(x_k) \cdot \mathrm{e}^{\mathrm{i}\alpha\kappa S_{\mathrm{int}}[\varphi]}\, d\mu(\varphi)}{\int_{\mathcal{S}(\mathbb{R}^D)} \mathrm{e}^{\mathrm{i}\alpha\kappa S_{\mathrm{int}}[\varphi]}\, d\mu(\varphi)}.$$

In the Euclidean case (E), we obtain

$$\boxed{\mathcal{C}_n(x_1, \ldots, x_n) = \frac{\int_{\mathcal{S}(\mathbb{R}^D)} \prod_{k=1}^n \varphi(x_k) \cdot \mathrm{e}^{\kappa S_{\mathrm{int}}[\varphi]}\, d\mu(\varphi)}{\int_{\mathcal{S}(\mathbb{R}^D)} \mathrm{e}^{\kappa S_{\mathrm{int}}[\varphi]}\, d\mu(\varphi)}}$$

along with

$$d\mu(\varphi) := \mathrm{e}^{S_{\mathrm{free}}[\varphi]}\, \mathcal{D}\varphi$$

where

$$S_{\mathrm{free}}[\varphi] = \frac{1}{2} \int_{\mathbb{R}^D} \left(\varphi_{tt} + \sum_{k=1}^{D-1} \varphi_{x_k x_k} - m_0^2\varphi^2 \right) d^D x.$$

In this case, the formula for the correlation function $\mathcal{C}_n(x_1, \ldots, x_n)$ (also called Schwinger function) is well-defined if we use a probabilistic measure μ on the function space $\mathcal{S}(\mathbb{R}^D)$ (or, more generally, on the space $\mathcal{S}'(\mathbb{R}^D)$ of tempered distributions). Such measures generalize the Wiener measure which was introduced by Wiener in 1923 in order to rigorously describe the probabilistic structure of the Brownian motion.[30] Multiplying the function $\mathcal{C}_n$ by test

[30] We refer to Gelfand et al (1964), Vol. IV (measures on linear topological spaces and generalized stochastic processes), Yeh (1973), Simon (1974), (1979), and Glimm, Jaffe (1987).

functions, $f_1, \ldots, f_n \in \mathcal{S}(\mathbb{R}^D)$ and setting $\varphi(f_k) := \int_{\mathbb{R}^D} \varphi(x_k) f(x_k)\, d^D x_k$, we obtain that the integral

$$\mathcal{C}_n[f_1, \ldots, f_n] := \int \mathcal{C}_n(x_1, \ldots, x_n) \prod_{k=1}^{n} f(x_k) \prod_{k=1}^{n} d^D x_k$$

is equal to

$$\frac{\int_{\mathcal{S}(\mathbb{R}^D)} \prod_{k=1}^{n} \varphi(f_k) \cdot \mathrm{e}^{\kappa S_{\mathrm{int}}[\varphi]}\, d\mu(\varphi)}{\int_{\mathcal{S}(\mathbb{R}^D)} \mathrm{e}^{\kappa S_{\mathrm{int}}[\varphi]}\, d\mu(\varphi)}.$$

This motivates the more general definition

$$\mathcal{C}_n[f_1, \ldots, f_n] := \frac{\int_{\mathcal{S}'(\mathbb{R}^D)} \prod_{k=1}^{n} \varphi(f_k)\, d\nu(\varphi)}{\int_{\mathcal{S}'(\mathbb{R}^D)}\, d\nu(\varphi)} \tag{15.39}$$

of correlation distributions. Here, let $n = 1, 2, \ldots$, and let the symbol ν denote a probabilistic measure on the space $\mathcal{S}'(\mathbb{R}^D)$ of tempered distributions, which is an extension of the measure

$$d\nu(\varphi) := \mathrm{e}^{\kappa S_{\mathrm{free}}(\varphi)} d\mu(\varphi)$$

on the space $\mathcal{S}(\mathbb{R}^D)$ of test functions. Concerning (15.39), we integrate over all tempered distributions $\varphi \in \mathcal{S}'(\mathbb{R}^D)$.

15.6.1 Axiomatic Quantum Field Theory

In 1965, Gårding and Wightman formulated axioms for relativistic quantum fields.[31] We will study the axiomatic approach to quantum field theory in Volume IV on quantum mathematics. Essential ingredients of the axiomatic approach to quantum field theory concern

- the 1956 Wightman axioms for vacuum expectation values,
- the Wightman reconstruction theorem for quantum fields,
- the 1965 Gårding–Wightman axioms,
- the 1973 Osterwalder–Schrader axioms on Schwinger distributions,
- the Osterwalder–Schrader reconstruction theorem for quantum fields in the Euclidean setting, and
- the 1964 Haag–Kastler axioms formulated in the language of nets of local C^*-algebras (algebraic quantum field theory).

[31] A. Wightman, Quantum field theory in terms of vacuum expectation values, Phys. Rev. **101** (1956), 860–866.
A. Wightman and L. Gårding, Fields as operator-valued distributions in relativistic quantum field theory, Arkiv för Fysik **28** (1965), 128–184.

The Haag–Kastler approach elaborates the frame of local quantum theory in Minkowski space. We refer to the monographs by Res Jost (1965), Streater and Wightman (1968), Emch (1972), Reed and Simon (1972), Vols. 2, 3, Simon (1974), Glimm and Jaffe (1981), Bogoliubov et al. (1975), (1990), Baumgärtel and Wollenberg (1992), Iagolnitzer (1993), Baumgärtel (1995), Haag (1996), and Araki (1999).

At this point, let us only sketch a few basic ideas. We will start with the notion of Poincaré transformation

$$x' = Ax + a$$

between the space-time coordinates $x := (x^1, x^2, x^3, x^0 = t)$ and x' of two arbitrary inertial systems. Here, the real invertible (4×4)-matrix A is the superposition of a rotation of the space coordinates, a Lorentz boost,

$$x'^1 = \frac{x^1 - vt}{\sqrt{1 - v^2}}, \quad x'^2 = x^2, \quad x'^3 = x^3, \quad t' = \frac{t - vx^1}{\sqrt{1 - v^2}},$$

and space or time reflections. The reflections may drop out. The parameter v with $0 \leq v < 1$ describes the relative velocity between the two inertial systems. Moreover, the map $x \mapsto x + a$ represents a space translation or time translation. Such a Poincaré transformation will be denoted by the symbol (A, a). All of the Poincaré transformations form a group called the Poincaré group.[32] The Gårding–Wightman axioms in the four-dimensional space-time $\mathbb{M}^4$ read as follows.[33]

(GW1) Relativistic symmetry. There exists both a complex Hilbert space X called state space and a unit vector Ω_0 called ground state (or vacuum). Moreover, the relativistic Poincaré group acts as symmetry group on the Hilbert space X. More precisely, there exists a map

$$\boxed{(A, a) \mapsto U(A, a)} \tag{15.40}$$

which assigns a unitary operator $U(A, a) : X \to X$ to each element (A, a) of the Poincaré group. The map (15.40) is a group morphism, that is, it sends products of Poincaré transformations to products of unitary operators.

(GW2) Spectral property. By the Stone theorem on one-parameter unitary groups in Hilbert spaces, there exist self-adjoint operators

$$P_\mu : \mathrm{dom}(P_\mu) \subseteq X \to X, \qquad \mu = 0, 1, 2, 3$$

such that

[32] This fundamental non-compact group and its representations will be studied in Volume III on gauge field theory.

[33] We use basic notions from functional analysis which can be found in Zeidler (1995), Vol. 1.

$$U(I,a) = \mathrm{e}^{\mathrm{i}\sum_{\mu=0}^{3} a^\mu P_\mu}$$

for all $a = (a^0, a^1, a^2, a^3)^d$ in $\mathbb{R}^4$. The operator $H := P_0$ is called the Hamiltonian (or energy operator) of the quantum field. Moreover, the operators P_1, P_2, P_3 are called the momentum operators of the quantum field. We postulate that the spectral points p_0, p_1, p_2, p_3 of the operators P_0, P_1, P_2, P_3, respectively, lie in the following forward cone:

$$p_0^2 - \sum_{j=1}^{3} p_j^2 \geq 0, \qquad p_0 \geq 0.$$

We also postulate that the energy of the ground state is equal to zero, that is,

$$H\Omega_0 = 0.$$

Intuitively, the spectral points p_0, p_1, p_2, p_3 describe the possible energy values $E = p_0$ and the possible momentum vectors

$$\mathbf{p} = p^1\mathbf{e}_1 + p^2\mathbf{e}_2 + p^3\mathbf{e}_3$$

of the quantum field. Here, $p^j := -p_j$ for $j = 1, 2, 3$, and $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ is a right-handed orthonormal system of vectors.

(GW3) The generalized quantum field. There exists a linear map

$$\boxed{f \mapsto \varphi[f]} \tag{15.41}$$

which assigns a linear self-adjoint operator $\varphi[f]$ on the Hilbert space X to each test function $f \in \mathcal{S}(\mathbb{R}^D)$. There exists a dense subset $\mathcal{D}$ of the Hilbert space X such that the domain of definition of each operator $\varphi[f]$ contains the set $\mathcal{D}$. The generalized quantum field is relativistically invariant. That is,

$$U(A,a)^{-1}\varphi[f]\, U(A,a) = \varphi[f_{(A,a)}]$$

for all elements (A, a) of the Poincaré group. By definition, $f_{(A,a)}(x)$ is equal to $f(Ax + a)$ for all $x \in \mathbb{M}^4$. Intuitively, think of

$$\varphi[f] := \int_{\mathbb{M}^4} \varphi(x) f(x)\, d^4x \tag{15.42}$$

for all test function $f \in \mathcal{S}(\mathbb{R}^4)$. Let us briefly discuss this as a key point to quantum field theory. Naively, a quantum field is an operator-valued function

$$x \mapsto \varphi(x)$$

which assigns a self-adjoint operator $\varphi(x)$ to each space-time point $x = (\mathbf{x}, t)$. Unfortunately, this is a highly singular and ill-defined mathematical object. Therefore, we consider first the integral $\varphi[f]$ from (15.42)

which is obtained by averaging over test functions f. This motivates the more general definition where the generalized quantum field (15.41) is a tempered operator-valued distribution; the values $\varphi[f]$ are operators on the Hilbert space X.

(GW4) Density of standard states. The ground state Ω_0 lies in $\mathcal{D}$, and the set of standard states

$$\left(\prod_{k=1}^{N} \varphi[f_k]\right)\Omega_0$$

for all $N = 1, 2, \ldots$ and all test functions $f_1, \ldots, f_N \in \mathcal{S}(\mathbb{R}^4)$ is dense in the Hilbert space X. We say that the ground state Ω_0 is a cyclic vector of the Hilbert space X. Intuitively, the standard states correspond to particle creation and particle annihilation.

(GW5) Locality of the generalized quantum field. If the supports of the two test functions $f, g \in \mathcal{S}(\mathbb{R}^4)$ are space-like separated,[34] then we have the commutation property

$$\big(\varphi[f]\cdot\varphi[g] - \varphi[g]\cdot\varphi[f]\big)\,\theta = 0 \qquad \text{for all} \quad \theta \in \mathcal{D}.$$

Intuitively, the events belonging to the supports of f and g cannot influence each other by physical effects. This locality property is a weak form of causality.

In addition, we distinguish between the following two cases which are important from the physical point of view.

- Uniqueness of the ground state: Up to a complex factor, the ground state Ω_0 is unique. Furthermore, the ground state Ω_0 is relativistically invariant, that is, for each element (A, a) of the Poincaré group, there exists a complex number λ with $|\lambda| = 1$ such that $U(A, a)\Omega_0 = \lambda\Omega_0$.
- A phase transition of the quantum field is possible: There are at least two linearly independent ground states Ω_0 and Ω_1. Intuitively, by a phase transition, we understand a passage from Ω_0 to Ω_1.

The model is called trivial if it behaves like a free model. Finally, the model is called asymptotically free if it behaves like a free model as energy goes to infinity.

In what follows, we will also consider quantum fields on a D-dimensional space-time where $D = 2, 3, 4, \ldots$ In this case, one has to modify the axioms (GW1) through (GW5) above in a natural way. In particular, it is necessary to use the D-dimensional Poincaré group which consists of all the transformations

[34] This means that if $x \in \operatorname{supp} f$ and $y \in \operatorname{supp} g$, then

$$(x-y)^2 := (x^0-y^0)^2 - \sum_{k=1}^{3}(x^k-y)^k < 0.$$

$$x' = Ax + a, \qquad x \in \mathbb{R}^D$$

with arbitrary, but fixed element a of $\mathbb{R}^D$, and a fixed invertible $(D \times D)$-matrix A with the property

$$(Ax) \cdot (Ax) = x \cdot x \qquad \text{for all} \quad x \in \mathbb{R}^d.$$

Here, we write $x \cdot y := x^0 y^0 - \sum_{k=1}^{D-1} x^k y^k$.

15.6.2 The Euclidean Strategy in Constructive Quantum Field Theory

The highly sophisticated mathematical task of constructive quantum field theory is to construct models which satisfy the Gårding–Wightman axioms and which are related to given classical action functionals. The Euclidean strategy reads as follows.

Step 1: Start with a classical action functional. Pass over to the Euclidean version of the action. Discretize the Euclidean action. Study the discretized model with the methods of statistical physics (e.g., study the Ising model which is related to random walks).

Step 2: Carry out the continuum limit of the lattice model, by using estimates based on correlation inequalities. Show that the continuum limit satisfies the Osterwalder–Schrader axioms for the Euclidean correlation functions $\mathcal{C}_1, \mathcal{C}_2, \ldots$ Use the Osterwalder–Schrader reconstruction theorem in order to construct the Wightman functions W_n, by analytic continuation with respect to imaginary time. Intuitively,

$$\mathcal{C}_n(\mathbf{x}_1, t_1; \ldots, \mathbf{x}_n, t_n) = W_n(\mathbf{x}_1, \mathrm{i}t_1; \ldots, \mathbf{x}_n, \mathrm{i}t_n)$$

for $n = 2, 3, 4, \ldots$

Step 3: Prove that the Wightman functions (more precisely, the tempered Wightman distributions) satisfy the Wightman axioms, and use the Wightman reconstruction theorem in order to construct a model which satisfies the Gårding–Wightman axioms mentioned above. Intuitively, the Wightman functions are given by

$$W_n(\mathbf{x}_1, t_1; \ldots; \mathbf{x}_n, t_n) := \langle \Omega_0 | \varphi(\mathbf{x}_1, t_1) \cdots \varphi(\mathbf{x}_n, t_n) \Omega_0 \rangle.$$

That is, the Wightman functions are vacuum expectation values of products of quantum field operators.

Step 4: Study the uniqueness of the ground state or the existence of phase transitions between different ground states.

Step 5 : Study the existence of bound states (e.g., by using the Bethe–Salpeter equation), and the existence of scattering processes (e.g., by using the Haag–Ruelle theory for the S-matrix).

Examples. Let us consider some important examples which were studied in fundamental papers by Glimm and Jaffe around 1970 and by their followers. As an introduction, we recommend the monographs by Simon (1974), Glimm, Jaffe (1981), Grosse (1988), and Fernández, Fröhlich, Sokal (1992). In what follows suppose that the coupling constant $\kappa > 0$ is sufficiently small. Furthermore, we assume that the additional potential parameter vanishes, $\lambda := 0$.

(i) The Minkowskian φ_2^4-model: The Gårding–Wightman axioms are fulfilled.[35]

(ii) The Minkowskian φ_3^4-model: The Gårding–Wightman axioms are fulfilled.

(iii) Triviality of the Euclidean φ_D^4 model with $D > 4$: Roughly speaking, the model behaves like a free model. This crucial fact was proven independently by Aizenman and Fröhlich in 1982. Intuitively, this trivial behavior depends on the fact that random walks do not intersect if the dimension is larger than four.[36]

(iv) Triviality of the Euclidean φ_4^4-model: Based on computer experiments and partial rigorous results, most physicists expect that the φ^4-model in four-dimensional space-time has the surprising property of behaving like a free model (after renormalization).

In addition to the monographs mentioned above, we refer to the following papers:

K. Osterwalder and R. Schrader, Axioms for Euclidean Green's functions, Commun. Math. Phys. **31** (1973), 83–112; **42** (1975), 281–305.

H. Epstein and V. Glaser, The role of locality in perturbation theory, Ann. Inst. Poincaré **A19**(3) (1973), 211–295.

T. Spencer and F. Zirilli, Scattering states and bound states in $P(\varphi)_2$. Commun. Math. Phys. **49** (1975), 1–16.

J. Eckmann, J. Magnen, and R. Sénéor, Decay properties and Borel summability of the Schwinger functions in $P(\varphi)_2$ theories, Commun. Math. Phys. **39** (1975), 251–271.

K. Osterwalder and R. Sénéor, A nontrivial scattering matrix for weakly coupled $P(\varphi)_2$ models, Helv. Phys. Acta **49** (1976), 525–535.

J. Feldman and K. Osterwalder, The Wightman axioms and the mass gap for weakly coupled φ_4^3 quantum field theories, Ann. of Phys. **97** (1976), 80–135.

E. Seiler and B. Simon, On finite mass renormalization in the two-dimensional Yukawa model, J. Math. Phys. **16** (1975), 2289–2293.

J. Fröhlich, B. Simon, and T. Spencer, Infrared bounds, phase transitions, and continuous symmetry breaking, Commun. Math. Phys. **50** (1976), 79–85.

[35] If $\lambda > 0$, then the ground state is unique. See Simon (1974), Sect. IX.4.

[36] For a proof of the triviality of the model based on polymer gases, see the book by Grosse (1988). We also refer to the book by Fernández, Fröhlich, and Sokal (1992). These two books are cited on page 863.

J. Fröhlich, New super selection sectors (soliton states) in two-dimensional Bose quantum field models, Commun. Math. Phys. **47** (1976), 269–310.

D. Brydges, J. Fröhlich, and E. Seiler, On the construction of quantized gauge fields I–III, Ann. Phys. **121** (1979), 237–284, Commun. Math. Phys. **71** (1980), 159–205; **79** (1981), 353–399.

M. Aizenman, Geometric analysis of φ^4 fields and Ising models, Commun. Math. Phys. **86** (1982), 1–48.

J. Fröhlich, On the triviality of $\lambda\varphi^4$ theories and the approach to the critical point in $d \geq 4$ dimensions, Nuclear Phys. B **200** (1982), 281–296.

K. Gawędzki and A. Kupiainen, Massless φ_4^4 theory: Rigorous control of a renormalizable asymptotically free model, Commun. Math. Phys. **99**, (1985), 197–252.

J. Feldman, V. Magnen, V. Rivasseau, and R. Sénéor, A renormalizable field theory: The massive Gross–Neveu model in two dimensions, Commun. Math. Phys. **103** (1986), 67–103.

J. Feldman, V. Magnen, V. Rivasseau, and R. Sénéor, Construction and Borel summability of infrared φ_4^4 by a phase space expansion, Commun. Math. Phys. **109** (1987), 437–460.

Numerical results can be found in

M. Lüscher and P. Weisz, Scaling laws and triviality bounds in the lattice φ^4 theory, Nuclear Phys. B **290** (1987), 25–60.

15.6.3 The Renormalization Group Method

In very rough terms, this modern method in renormalization theory goes as follows. Let us consider the Euclidean φ_4^4-model. Our goal is to construct the Fourier transform of the renormalized Euclidean correlation functions

$$\mathcal{C}_{n,\mathrm{ren}}(p_1,\ldots,p_n) = \sum_{m=0}^{\infty} \mathcal{C}_{n,\mathrm{ren}}^{(m)}(p_1,\ldots,p_n)\kappa^m, \qquad n = 1,2,\ldots$$

as a formal power series expansion with respect to the coupling constant κ. Here, the momenta $p_1,\ldots,p_n$ lie in $\mathbb{R}^4$. Moreover, $p_1+\ldots+p_n = 0$. The idea is to use a limit of the form

$$\boxed{\mathcal{C}_{n,\mathrm{ren}}^{(m)}(p_1,\ldots,p_n) = \lim \ \mathcal{C}_{n,P_{\max},P_{\min}}^{(m)}(p_1,\ldots,p_n)} \tag{15.43}$$

where

- $P_{\max} \to +\infty$ (high-energy limit or ultraviolet limit), and
- $P_{\min} \to +0$ (low-energy limit or infrared limit).

To this end, we fix two positive numbers $P_{\max}$ and $P_{\min}$, and we only consider momenta $p_1,\ldots,p_{n-1} \in \mathbb{R}^4$ which satisfy the following inequality

$$0 < P_{\min}^2 \leq \sum_{\mu=0}^{3} (p_k^\mu)^2 \leq P_{\max}^2, \qquad k = 1,\ldots,n-1.$$

We call this a cut-off in momentum space. Naively, correlation functions are given by functional integrals. Motivated by this, we use the following ingredients.

(i) Finite functional integrals: We construct the approximation

$$\mathcal{C}^{(m)}_{n,P_{\max},P_{\min}}(p_1,\dots,p_n)$$

by using suitable functional integrals on a lattice. The definition of the approximate functional integrals is motivated by physical arguments: in terms of the language of Feynman diagrams, the internal lines correspond to the truncated Euclidean propagator

$$\frac{1}{(2\pi)^2(p^2+m_0^2)}\cdot\left(\mathrm{e}^{-(p^2+m_0^2)/P_{\max}^2}-\mathrm{e}^{-(p^2+m_0^2)/P_{\min}^2}\right),$$

whereas the vertices of the Feynman diagrams are determined by the Lagrangian density of the interaction potential including counterterms.

(ii) The renormalization group equation: We postulate the validity of a first-order partial differential equation for

$$\frac{\partial\mathcal{C}^{(m)}_{n,P_{\max},P_{\min}}}{\partial P_{\max}}.$$

This is the so-called renormalization group equation which determines a semi-flow with respect to the flow parameter $P_{\max}$. Intuitively, this differential equation describes natural rescaling properties of correlation functions under multiplicative renormalization.

(iii) Boundary conditions: We add boundary conditions to equation (ii) at the points $P_{\min}=0$ and $P_{\max}$. For example, these boundary conditions help to fix the counterterms. Therefore, they are also called renormalization conditions.

The idea is to prove the convergence of (15.43) by using estimates coming from an induction argument based on (ii) and (iii). An elegant proof can be found in

C. Kopper, Renormalization with Flow Equations (in German), Shaker, Aachen, 1998, pp. 18–21.

This proof avoids the use of huge, but redundant Feynman diagrams. For the dimensions $D=2,3,4$, a detailed study of the φ^4_D-model is contained in the book by

M. Salmhofer, Renormalization: An Introduction, Springer, Berlin, 1999,

along with a discussion of the relation to Feynman diagrams.

The Polchinski equation. Finally, let us mention that the approach sketched above dates back to the fundamental paper by

J. Polchinski, Renormalization and effective Lagrangians, Nucl. Phys. B **231** (1984), 269–295.

The idea of Polchinski is to consider the space of Lagrangian densities of the φ_4^4-model including counterterms and cut-offs in the momentum space. The goal is to construct a semi-flow in this space which starts with the bare Lagrangian density and ends up with the renormalized Lagrangian density. The flow parameter is equal to $P_{\max}$, along with the high-energy limit

$$P_{\max} \to +\infty.$$

The basic trick is to obtain a partial differential equation called the semi-flow equation (or renormalization group equation) which governs the semi-flow. To this end, Polchinski uses the generating functional

$$Z(J; P_{\max}, P_{\min})$$

for the correlation functions in the form of a functional integral, and he postulates the validity of the so-called Polchinski equation

$$\frac{\partial Z(J; P_{\max}, P_{\min})}{\partial P_{\max}} = 0.$$

This elegantly yields the desired semi-flow equation.[37] Finally, Polchinski simplifies the semi-flow equation by assuming that the low-energy behavior, $P_{\min} \to +0$, has a simple structure. The Polchinski approach avoids the trouble with overlapping divergences in the integrals of the BPHZ approach to renormalization. The ingenious idea of considering semi-flows in the space of physical theories (i.e., the space of Lagrangian densities) is due to Wilson. We refer to

K. Wilson, Renormalization group and critical phenomena, I, II, Phys. Rev. B **4** (1971), 3174–3183, 3184–3205.

K. Wilson and J. Kogut, The renormalization group and the ε-expansion, Physics Reports **12C** (1974), 75–199.

Renormalization in noncommutative geometry. Based on the powerful Polchinski method, the renormalization of the Euclidean φ^4-model on the four-dimensional Moyal plane in noncommutative geometry was proven for the first time by

H. Grosse and R. Wulkenhaar, Renormalisation of φ^4 theory on noncommutative $\mathbb{R}^4$ to all orders. Internet: `hep-th/0403232`

H. Grosse and R. Wulkenhaar, Renormalisation of φ^4 theory on noncommutative $\mathbb{R}^4$ in the matrix base, Commun. Math. Phys. **256** (2005), 305–374. Internet: `hep-th/0401128`

A stronger version of this result can be found in

[37] A simple variant of the Polchinski equation can be found in Sect. 8.5.4.

V. Rivasseau, F. Vignes-Tourneret, and R. Wulkenhaar, Renormalization of noncommutative φ^4 theory by multi-scale analysis.
Internet: `hep-th/0501036`

For noncommutative geometry, we refer to the following monographs:

A. Connes, Noncommutative Geometry, Academic Press, New York, 1994.

J. Gracia-Bondia, J. Várilly, and H. Figueroa, Elements of Noncommutative Geometry, Birkhäuser, Boston, 2001.

F. Scheck, W. Wend, and H. Upmeier, Noncommutative Geometry and the Standard Model of Elementary Particle Physics, Springer, Berlin, 2003.

Yu. Manin, Topics in Noncommutative Geometry, Princeton University Press, 1991.

16. Peculiarities of Gauge Theories

> The interactions between elementary particles in the Standard Model are described by gauge field theories.
>
> Folklore

16.1 Basic Difficulties

Observe the crucial fact that both quantum electrodynamics and the Standard Model in particle physics are gauge field theories. This means that

- the interacting particles (photons, gluons, and vector bosons) are described by gauge potentials which are only determined up to local gauge transformations, and
- the Lagrangian density is invariant under local gauge transformations.

The typical difficulty is that the original response equations are not uniquely solvable because of the local gauge invariance.

> *In order to get uniquely solvable response equations, one has to fix the choice of the gauge potentials by adding side conditions.*

This leads to additional terms in the Lagrangian density which enter the response equations. This is called gauge fixing. The response equations depend on the gauge fixing. We now proceed as follows.

- We fix the response equations by fixing the side conditions (gauge fixing),
- and we apply the universal approach to this special situation.

However, this approach only works well if the following two conditions are fulfilled:

(G) Gauge invariance: The physics does not depend on the choice of the gauge fixing.
(S) Unitarity condition: The S-matrix is unitary.

If the unitarity condition is violated, then the elements of the S-matrix cannot be interpreted as transition probabilities. To guarantee condition (S) in general gauge field theories, one has to introduce additional ghost and antighost fields (see Sect. 16.6).

The general strategy. In the classical calculus of variations, constrained problems are treated by means of Lagrange multipliers (see Sect. 7.28). Physicists have developed several methods in order to generalize the method of Lagrange multipliers to gauge field theories. The main idea is to introduce additional quantities called Lagrange multipliers in mathematics; in physics one speaks of ghosts, antighosts, and antifields. In the next sections, we will sketch the basic ideas related to

- the Faddeev–Popov–De Witt ghost formalism,
- the Becchi–Rouet–Stora–Tyutin (BRST) ghost formalism, and
- the Batalin–Vilkovisky antifield formalism based on a master equation.

Summarizing, the basic ideas behind all of these formalisms read as follows:

- In order to compute physical processes in gauge field theories, the simplest way is to fix the gauge potential by specifying a special gauge condition. Then, one has to show that all of the computations including renormalization via counterterms are independent of the choice of the gauge.
- To guarantee the unitarity of the S-matrix and hence the consistency of transition probabilities for scattering processes, one has to introduce additional fields (Lagrangian multipliers) that are called ghosts, antighosts, or antifields. It is crucial that these additional fields do not appear as real incoming or outgoing particles in scattering processes. This is a consequence of cancellations governed by an additional symmetry related to BRST symmetry or its generalizations.

We will discuss below that

> *Ghosts are closely related to the fundamental concept of cohomology in mathematics.*

As we will show in Volume II on the Gupta–Bleuler formalism in quantum electrodynamics, real physical photons are polarized transversally; in contrast to this, ghosts correspond to photons which possess an unphysical longitudinal polarization. The formalisms mentioned above represent generalizations of the Gupta–Bleuler formalism.

System of units. In this chapter, we use the energetic system of units with $\hbar = c = 1$.

16.2 The Principle of Critical Action

We want to discuss the basic ideas of generalizing quantum electrodynamics to gauge theories having non-commutative gauge groups. The Standard Model in particle physics is of this type. In what follows, let $x \in \mathbb{M}^4$, and choose $\mu, \nu = 0, 1, 2, 3$ along with the notation introduced on page 768. We will use

- a matter field ψ (fundamental particles), and
- a gauge field F (gauge bosons).

The fundamental particles are fermions (e.g., electrons or quarks). The gauge bosons are responsible for the interaction between the fundamental particles (e.g., photons or gluons). The gauge field F has a potential A called gauge potential. Different choices of the potential A correspond to gauge transformations.

> *The theory has to be designed in such a way that it is invariant under gauge transformations.*

This is called the gauge invariance principle

Prototype 1: Quantum electrodynamics with the commutative gauge group $\mathcal{G} = U(1)$ and the gauge Lie algebra $u(1)$. In quantum electrodynamics, the matter field ψ describes electrons, whereas the gauge field F corresponds to photons (electromagnetic field).[1] The covariant derivative reads as

$$\nabla_\mu := \partial_\mu - \mathrm{i}eA_\mu, \qquad \mu = 0, 1, 2, 3$$

where the real valued functions A_0, A_1, A_2, A_3 represent the 4-potential of the electromagnetic field which is given by

$$F_{\mu\nu} := \partial_\mu A_\nu - \partial_\nu A_\mu, \qquad \mu, \nu = 0, 1, 2, 3.$$

If we set

$$\mathcal{A}_\mu := -\mathrm{i}eA_\mu, \qquad \mathcal{F}_{\mu\nu} = -\mathrm{i}eF_{\mu,\nu},$$

then the following hold true for all $\mu, \nu = 0, 1, 2, 3$.

- $\mathcal{A}_\mu \in u(1)$,
- $\nabla_\mu = \partial_\mu + \mathcal{A}_\mu$,
- $\mathcal{F}_{\mu\nu} = \nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu$,
- $\kappa = e$ (coupling constant),
- Dirac equation: $\mathrm{i}\gamma^\mu \nabla_\mu \psi = m_e \psi$ (m_e bare rest mass of the electron).

Here, $-e$ is the bare electric charge of the electron. Note that the bare mass and the bare charge of the electron are used in order to construct the crucial renormalized (effective) masses and charges of the electron which can be measured in physical experiments. To this end, one uses the sophisticated methods of renormalization theory. The complex (4×4)-matrices $\gamma^0, \gamma^1, \gamma^2, \gamma^3$

[1] Recall that the Lie group $U(1)$ consists of all complex numbers z with $|z| = 1$, and the Lie algebra $u(1)$ consists of all purely imaginary numbers $\alpha\mathrm{i}$ where α is a real number.
For $N = 2, 3, \ldots$, the Lie group $SU(N)$ consists of all complex $(N \times N)$-matrices U with $\det U = 1$ and $U^{-1} = U^\dagger$ (special unitary group). Moreover, the corresponding Lie algebra $su(N)$ consists of all complex $(N \times N)$-matrices $\mathcal{A}$ with $\mathcal{A} = -\mathcal{A}^\dagger$ and vanishing trace, $\mathrm{tr}(\mathcal{A}) = 0$.

are the Dirac–Pauli matrices from page 789. Our goal is to replace $U(1)$ by more general gauge groups.

Prototype 2: Quantum chromodynamics with the non-commutative gauge group $\mathcal{G} = SU(3)$ and the gauge Lie algebra $su(3)$. The reader should have in mind the matter field

$$\psi(x) = \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \\ \psi_3(x) \end{pmatrix}$$

of a quark. The components $\psi_1(x), \psi_2(x), \psi_3(x)$ are Dirac 4-spinors as in Sect. 14.3 on page 789. They describe the three different color charges of the quark (red, green, and blue). The gauge bosons correspond to the field $F_{\mu\nu}$ below. Parallel to quantum electrodynamics above, let us introduce the following notions for all $\mu, \nu = 0, 1, 2, 3$.

- $\mathcal{A}_\mu \in su(3)$ (gluon potential)
- $\nabla_\mu = \partial_\mu + \mathcal{A}_\mu$ (covariant derivative),
- $\mathcal{F}_{\mu\nu} = \nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu$ (gluon field tensor),
- Dirac equation: $\mathrm{i}\gamma^\mu \nabla_\mu \psi = m\psi$ (m bare rest mass of the quark).

Explicitly,

$$\boxed{\mathcal{F}_{\mu\nu} = \partial_\mu \mathcal{A}_\nu - \partial_\nu A_\mu + [\mathcal{A}_\mu, \mathcal{A}_\nu]_-, \qquad \mu, \nu = 0, 1, 2, 3.}$$

Recall that $[\mathcal{A}_\mu, \mathcal{A}_\nu]_- := \mathcal{A}_\mu \mathcal{A}_\nu - \mathcal{A}_\nu \mathcal{A}_\mu$. In contrast to quantum electrodynamics, there arises the additional nonlinear term $[A_\mu, A_\nu]_-$ which describes self-interaction of the gluon field. In order to get connect with the terminology used in quantum electrodynamics, we introduce the rescaled gluon potential A_μ and the rescaled gluon field tensor $F_{\mu\nu}$ by setting

$$\mathcal{A}_\mu = -\mathrm{i}\kappa A_\mu, \qquad F_{\mu\nu} = -\mathrm{i}\kappa F_{\mu\nu}$$

where $\kappa > 0$ is the coupling constant for the interaction between quarks and gluons. This way, the Dirac equation reads as

$$\mathrm{i}\gamma^\mu(\partial_\mu - \mathrm{i}\kappa A_\mu(x))\psi(x) = m\psi(x), \qquad x \in \mathbb{M}^4. \tag{16.1}$$

Here, we set

$$A_\mu(x) \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \\ \psi_3(x) \end{pmatrix} = \begin{pmatrix} \varphi_1(x) \\ \varphi_2(x) \\ \varphi_3(x) \end{pmatrix}, \qquad \gamma^\mu A_\mu(x)\psi(x) = \begin{pmatrix} \gamma^\mu \varphi_1(x) \\ \gamma^\mu \varphi_2(x) \\ \gamma^\mu \varphi_3(x) \end{pmatrix}.$$

Introducing Feynman's dagger symbol $\not{\nabla} := \gamma^\mu \nabla_\mu$, the Dirac equation (16.1) can elegantly be written as

$$(\mathrm{i}\,\not\nabla - m)\psi = 0.$$

The real Lie algebra $su(3)$ possesses an 8-dimensional basis $\mathcal{B}_1, \ldots, \mathcal{B}_8$. Therefore,

$$\mathcal{A}_\mu(x) = \kappa(A^1_\mu(x)\mathcal{B}_1 + \ldots + A^8_\mu(x)\mathcal{B}_8).$$

Set $B_s := \mathrm{i}\mathcal{B}_s$. Then

$$A_\mu(x) = A^1_\mu(x)B_1 + \ldots + A^8_\mu(x)B_8.$$

Consequently, the gluon potential can be described by the following eight real four-potentials

$$A^s_0(x), A^s_1(x), A^s_2(x), A^s_3(x), \qquad s = 1, \ldots, 8,$$

in contrast to one real four-potential for the photon in quantum electrodynamics.

This explains why there are eight gluons in quantum chromodynamics.

Equation (16.1) does not fully describe the self-interactions of the quark-gluon field (ψ, A). For formulating the final theory, we will use the principle of critical action along with the Lagrangian density (16.2) below. Let us now discuss this.

The $SU(N)$-gauge field model. We will proceed in the following steps. Let $N = 2, 3, \ldots$ and $\mu, \nu = 0, 1, 2, 3$.

(i) Gauge Lie group $\mathcal{G}$. We choose the Lie group $\mathcal{G} := SU(N)$ as our gauge group.

(ii) Gauge Lie algebra $\mathcal{LG}$. The real Lie algebra $su(N)$ to the Lie group $SU(N)$ is called the gauge Lie algebra $\mathcal{LG}$ corresponding to the gauge group $\mathcal{G}$. The dimension of $su(N)$ is equal to $S := N^2 - 1$.

(iii) Gauge potential $\mathcal{A}$ of gauge bosons (e.g., gluons):

$$\mathcal{A}(x) = (\mathcal{A}_0(x), \mathcal{A}_1(x), \mathcal{A}_2(x), \mathcal{A}_3(x)).$$

The components $\mathcal{A}_\mu(x)$ lie in the gauge Lie algebra $su(N)$. In particular, they are complex $(N \times N)$-matrices, but not complex numbers of the form $\mathcal{A}_\mu(x) = -\mathrm{i}eA_\mu(x)$, as in quantum electrodynamics.

(iv) Covariant derivative: $\nabla_\mu := \partial_\mu + \mathcal{A}_\mu$.

(v) Field tensor $\mathcal{F}$ of gauge bosons: $\mathcal{F}_{\mu\nu} := \nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu$. Hence,

$$\boxed{\mathcal{F}_{\mu\nu} = \partial_\mu\mathcal{A}_\nu - \partial_\nu A_\mu + [\mathcal{A}_\mu, \mathcal{A}_\nu]_-.}$$

(vi) Rescaling: In order to get connect with the notation used in quantum electrodynamics, we introduce the rescaled gauge potential A_μ and the rescaled field tensor $F_{\mu\nu}$ by setting

$$\mathcal{A}_\mu = -\mathrm{i}\kappa A_\mu, \qquad \mathcal{F}_{\mu\nu} = -\mathrm{i}\kappa F_{\mu\nu}.$$

This implies $\nabla_\mu = \partial_\mu - \mathrm{i}\kappa A_\mu$ and

$$F_{\mu\nu}(x) := \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) - \mathrm{i}\kappa\,[A_\mu(x), A_\nu(x)]_-.$$

The constant $\kappa > 0$ is called the coupling constant. The rescaled field tensor $F_{\mu\nu}$ describes particles which are responsible for the interaction between fundamental fermions (quarks and leptons).

(vii) Gauge transformation: For all $x \in \mathbb{M}^4$ and $\mu = 0, 1, 2, 3$,

$$\mathcal{A}^g_\mu(x) = G(x)\mathcal{A}_\mu(x)G(x)^{-1} - \partial_\mu G(x) \cdot G(x)^{-1},$$

where $G(x) \in \mathcal{G}$. Explicitly,

$$G(x) = \exp\left(\sum_{s=1}^{S} \theta^s(x)\mathcal{B}_s\right).$$

Here, $\mathcal{B}_1, \ldots, \mathcal{B}_S$ is a basis of $su(N)$, and we set $\theta = (\theta^1, \ldots, \theta^S)$ with real components θ^s. In order to indicate the dependence of the gauge transformation on the function θ, we also write $\mathcal{A}^\theta$ instead of $\mathcal{A}^g$. For the rescaled gauge potential, the gauge transformation reads as

$$A^g_\mu(x) = G(x)A_\mu(x)G(x)^{-1} - \frac{\mathrm{i}}{\kappa}\,\partial_\mu G(x) \cdot G(x)^{-1}.$$

(viii) Matter field ψ: The field

$$\psi(x) = \begin{pmatrix} \psi_1(x) \\ \vdots \\ \psi_N(x) \end{pmatrix}$$

has N so-called color components. The index $c = 1, \ldots, N$ of ψ_c is called color index. Each color component ψ_c is a Dirac 4-spinor, that is,

$$\psi_c(x) = \begin{pmatrix} \psi^1_c(x) \\ \vdots \\ \psi^4_c(x) \end{pmatrix}, \qquad c = 1, \ldots, N$$

where each $\psi^k_c(x)$ is a complex number. Using the complex (4×4)-Dirac–Pauli matrices $\gamma^0, \gamma^1, \gamma^2, \gamma^3$ from Sect. 14.3, we set $\overline{\psi}_c(x) := \psi_c(x)^\dagger\gamma^0$ along with $\overline{\psi}(x) := (\overline{\psi}_1(x), \ldots, \overline{\psi}_N(x))$.

(ix) Source functional:

$$J[A,\psi] := \int_{\mathbb{M}^4} \{\mathrm{tr}(A_\mu(x)J^\mu(x)) + \overline{\mathcal{J}}(x)\psi(x) + \overline{\psi}(x)\mathcal{J}(x)\}\, d^4x.$$

Here, we assume that $\mathrm{i}J^\mu(x) \in su(N)$ for all $\mu = 0,1,2,3$ and $x \in \mathbb{M}^4$.

(x) Total action:

$$\boxed{S[A,\psi] := \int_{\mathbb{M}^4} \mathcal{L}(A(x), \partial A(x), \psi(x), \partial\psi(x))\, d^4x.}$$

Here we use the Lagrangian density

$$\boxed{\mathcal{L} := \mathcal{L}_{\text{gauge boson}} + \mathcal{L}_{\text{matter}} + \mathcal{L}_{\text{gauge fixing}} + \mathcal{L}_{\text{ghost}}} \tag{16.2}$$

along with the following expressions:[2]

- $\mathcal{L}_{\text{gauge boson}} := -\frac{1}{4}\,\mathrm{tr}(F_{\mu\nu}F^{\mu\nu})$,
- $\mathcal{L}_{\text{matter}} := \overline{\psi}(\mathrm{i}\not{\nabla} - m)\psi$,
- $\mathcal{L}_{\text{gauge fixing}} := \frac{1}{2\xi}\,\mathrm{tr}(\partial^\mu A_\mu \cdot \partial^\nu A_\nu)$,
- $\mathcal{L}_{\text{ghost}} := -\zeta^d \partial^\mu \nabla_\mu \eta$.

Fix the gauge parameter $\xi > 0$. Both the components of the ghost field

$$\eta(x) = \begin{pmatrix} \eta^1(x) \\ \vdots \\ \eta^S(x) \end{pmatrix}$$

and the components of the antighost field $\zeta(x)^d = (\zeta_1(x), \ldots, \zeta_S(x))$ are Grassmann variables which satisfy the anticommuation relations (16.4) on page 890.

Relation to modern differential geometry. In Volume III on gauge theory, we will relate this gauge field approach to the language of modern differential geometry. In this geometric setting, the following hold true.

- The gauge field tensor $\mathcal{F}$ represents the curvature of a principal fiber bundle $\mathcal{P} = \mathbb{M}^4 \times \mathcal{G}$ over the space-time manifold $\mathbb{M}^4$ with structure group $\mathcal{G}$.
- The gauge potential $\mathcal{A}$ is the corresponding connection which describes the parallel transport in the principal fiber bundle $\mathcal{P}$, and which generates the covariant derivative ∇.
- The matter field ψ is the section of a vector bundle $\mathcal{V}$ which is associated to the principal fiber bundle $\mathcal{P}$.
- The gauge transformations correspond to the changing of bundle charts.

[2] The partial action $\int_{\mathbb{M}^4} \mathcal{L}_{\text{gauge boson}} d^4x$ is also called the Yang–Mills action.

16.3 The Language of Physicists

Coordinates. We want to reformulate the theory above in terms of a basis $\mathcal{B}_1, \ldots, \mathcal{B}_S$ of the real gauge Lie algebra $su(N)$. For the gauge potential,

$$\mathcal{A}_\mu = \kappa \sum_{s=1}^{S} A_\mu^s(x) \mathcal{B}_s, \qquad \mu = 0, 1, 2, 3.$$

In order to simplify notation, we will use the following Einstein summation convention: We sum over Latin (resp. Greek) indices from 1 to S (resp. from 0 to 3). For example, we will briefly write

$$\mathcal{A}_\mu(x) = \kappa A_\mu^s(x) \mathcal{B}_s, \qquad \mu = 0, 1, 2, 3.$$

The components $A_\mu^s(x)$ are real numbers for all indices μ, s and all space-time points $x \in \mathbb{M}^4$. Therefore, the theory corresponds to S real four-potentials

$$A_0^s(x), A_1^s(x), A_2^s(x), A_3^s(x), \qquad s = 1, \ldots, S$$

which describe S gauge bosons. The commutation relations for the basis elements $\mathcal{B}_1, \ldots, \mathcal{B}_S$ of the Lie algebra $su(N)$ read as

$$\boxed{[\mathcal{B}_a, \mathcal{B}_b]_- = c_{ab}^s \mathcal{B}_s, \qquad a, b = 1, \ldots, S.}$$

The real numbers c_{ab}^s are called the structure constants of the Lie algebra $su(N)$. In particular, $c_{ab}^s = -c_{ba}^s$ for $a, b, s = 1, \ldots S$. It turns out that the structure constants know all about the Lie algebra and about the local structure of the corresponding Lie group.

> *Observe that most physicists do not use Lie algebras in the sense of mathematics, but they use a slight modification.*

The reason for this is the fact that the elements $\mathcal{B}$ of the Lie algebra $su(N)$ are skew-adjoint matrices. However, observables in quantum theory are self-adjoint matrices. Therefore, physicists pass from $\mathcal{B}$ to $B := \mathrm{i}\mathcal{B}$. Motivated by this, physicists do not use the basis $\mathcal{B}_1, \ldots \mathcal{B}_S$ of the Lie algebra $su(N)$, but they introduce the self-adjoint matrices

$$B_s := \mathrm{i}\mathcal{B}_s, \qquad s = 1, \ldots, S.$$

We then get the commutation relations

$$[B_a, B_b]_- = \mathrm{i} c_{ab}^s B_s, \qquad a, b = 1, \ldots, S.$$

Here, $[B_a, B_b]_- = B_a B_b - B_b B_a$ is a modified Lie product, since B_a, B_b do not lie in the Lie algebra $su(N)$, but in $\mathrm{i} \otimes su(N)$.

To simplify computations, it is important that the matrices $B_1, \ldots, B_S$ can be chosen in such a way that[3]

$$\boxed{\operatorname{tr}(B_s B_r) = \delta_{rs}, \qquad s, r = 1, \ldots, S.}$$

We will also write δ^{rs} instead of δ_{rs}, and we will use the Kronecker symbols δ_{rs} and δ^{rs} in order to lower and lift Latin indices, respectively. For example, $B^s := \delta^{sr} B_r$. Using the matrices $B_1, \ldots, B_S$, we write the rescaled gauge potential as

$$A_\mu(x) = A^s_\mu(x) B_s, \qquad \mu = 0, 1, 2, 3.$$

For the rescaled gauge field tensor $F_{\mu\nu}(x) = F^s_{\mu\nu}(x) B_s$, we get

$$\boxed{F^s_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) + \kappa c^s_{ab} A^a_\mu A^b_\nu(x), \qquad s = 1, \ldots, S.}$$

This means that the strength of the field tensor components $F^s_{\mu\nu}$ depends on the structure constants c^s_{ab} of the gauge Lie algebra $su(N)$. Setting $\psi = 0$, the source functional reads as

$$J[A, 0] := \int_{\mathbb{M}^4} J^\mu_s(x) A^s_\mu(x) d^4x$$

where $J_\mu(x) = J^s_\mu(x) B_s$, and $J^\mu_s(x) := \eta^{\mu\nu} \delta_{sr} J^r_\nu(x)$. The Lagrangian density reads as

- $\mathcal{L}_{\text{gauge boson}} := -\frac{1}{4} F^s_{\mu\nu} F^{\mu\nu}_s$.
- $\mathcal{L}_{\text{matter}} := \overline{\psi} \gamma^\mu (\mathrm{i}\partial_\mu + \kappa A^s_\mu B_s) \psi$.
- $\mathcal{L}_{\text{gauge fixing}} := \frac{1}{2\xi} \partial^\mu A^s_\mu \partial^\nu A^s_\nu$.
- $\mathcal{L}_{\text{ghost}} := -\zeta^d \partial^\mu (\partial_\mu - \mathrm{i}\kappa A^s_\mu B_s) \eta$.

The response function of the gauge bosons: For fixed $\varepsilon > 0$, the response function is given by the formal (divergent) Fourier integral

$$\boxed{R^{rs}_{\mu\nu}(x - y) := \frac{1}{(2\pi)^4} \int_{\mathbb{R}^4} \frac{\mathrm{e}^{-\mathrm{i}p(x-y)}}{p^2 + \varepsilon \mathrm{i}} \left(\eta_{\mu\nu} + (\xi - 1) \frac{p_\mu p_\nu}{p^2} \right) \delta^{rs} \, d^4p.}$$

[3] This follows from a general theorem telling us that the negative Killing form of the Lie algebra $\mathcal{LG}$ to a compact semisimple Lie group $\mathcal{G}$ represents an inner product on the Lie algebra $\mathcal{LG}$. For the proof, see Simon (1996), p. 171. In the special case of the Lie algebra $su(2)$, note that the Pauli matrices $\sigma^1, \sigma^2, \sigma^3$ from page 789 satisfy the commutation relations

$$[\sigma^1, \sigma^2]_- = 2\mathrm{i}\sigma^3, \quad [\sigma^2, \sigma^3]_- = 2\mathrm{i}\sigma^1, \quad [\sigma^3, \sigma^1]_- = 2\mathrm{i}\sigma^2.$$

Furthermore, we have the following anticommutation relations

$$\sigma^r \sigma^s + \sigma^s \sigma^r = 2\delta^{rs}, \qquad r, s = 1, 2, 3$$

which imply $\operatorname{tr}(\sigma^r \sigma^s) = 2\delta^{rs}$. Therefore, we choose $B_s = B^s := \frac{1}{\sqrt{2}} \sigma^s, s = 1, 2, 3$.

The choice $\xi := 1$ (resp. $\xi := 0$) is called the Feynman gauge (resp. the Landau gauge). An elegant approach to the response function can be based on computing a functional integral.

In terms of the operator approach to quantum field theory, the quantized fields $A^s_\mu(x)$ are operators, and we get the following vacuum expectation values for time-ordered products of interacting quantum fields:

$$\boxed{\langle 0_{\text{int}}|\mathcal{T}\{A^r_\mu(x)A^s_\nu(y)\}|0_{\text{int}}\rangle = G^{rs}_{\mu\nu}(x-y).}$$

These functions are also called the full 2-point correlation functions of the gauge bosons (or the full 2-point Green's functions/or the gauge boson propagators). A detailed study along with interesting applications to elementary particles can be found in Volume V on the physics of the Standard Model. For a list of standard references, we refer to page 908.

If we replace the gauge Lie group $SU(N)$ (resp. the gauge Lie algebra $su(N)$ by $U(1)$ (resp. $u(1)$), then our model passes over to quantum electrodynamics.

16.4 The Importance of the Higgs Particle

The Lagrangian density $\mathcal{L}$ above does not contain a quadratic term with respect to the gauge potential A. Such a missing term tells us that the gauge bosons are massless. However, this contradicts experimental experience. In 1967 Weinberg solved this problem by introducing an additional field called Higgs field.[4] This field is responsible for the huge masses of the three gauge bosons W^+, W^-, Z^0 in electroweak interaction (nearly 90 proton masses). The Higgs mechanism will be studied in Volume III on gauge theory. This mechanism is closely related to the theory of phase transitions in solid state physics.

16.5 Integration over Orbit Spaces

Count equivalent physical states only once.
Folklore

The Feynman functional integral sums over all possible physical states. Here, it is crucial that equivalent physical states are counted only once. Therefore, in the case where symmetries are present, one has to sum over the orbits of the symmetry group, but not on the single points of the orbit. This is the basic idea behind the Faddeev–Popov–De Witt ghost approach. To explain this, let us start with a simple example from classical calculus.

[4] S. Weinberg, A model of leptons, Phys. Rev. Lett. **19** (1967), 1264–1266. This is one of the most quoted papers in elementary particle physics.

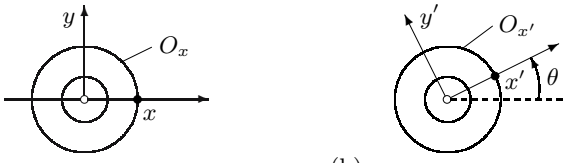


Fig. 16.1. Symmetry and orbits

Prototype of orbit summation. Using rotation symmetry, we want to reduce the 2-dimensional integral

$$\mathcal{J} := \int_{\mathbb{R}^2} \mathrm{e}^{-x^2-y^2}\, dxdy$$

to a one-dimensional integral. Observe that the integrand is invariant under the rotation

$$x' = x\cos\theta + y\sin\theta, \quad y' = y\cos\theta - x\sin\theta$$

of the (x, y)-coordinate system pictured in Fig. 16.1(a). These rotations about the origin are called gauge transformations; they form the group $SO(2)$. If we apply all possible rotations to a given point, say $(x, 0)$, then we obtain the orbit O_x. We regard the points of each orbit as equivalent points.

It is our goal to simplify the given integral $\mathcal{J}$ by eliminating the contributions coming from equivalent points.

To this end, we will proceed in two steps.

(i) Special gauge fixing (Fig. 16.1(a)): Choose first the special gauge angle $\theta := 0$. Using polar coordinates, the integral looks like

$$\mathcal{J} = \int_0^\infty \mathrm{e}^{-r^2} r dr \int_{-\pi}^{\pi} d\varphi.$$

This can be written as

$$\mathcal{J} = \int_0^\infty \mathrm{e}^{-x^2} x \left(\int_{O_x} d\theta \right) dx$$

which is the desired orbit summation. Equivalently,

$$\mathcal{J} = \pi \int_{-\infty}^{\infty} \mathrm{e}^{-x^2} |x|\, dx.$$

Using the Dirac delta function, this rigorous formula can be formally written as[5]

$$\mathcal{J} = \pi \int_{\mathbb{R}^2} \mathrm{e}^{-x^2-y^2} \delta(y)|x|\, dxdy.$$

[5] Note that $\int_{\mathbb{R}} f(x, y)\delta(y)dy = f(x, 0)$.

(ii) General gauge fixing (Fig. 16.1(b)): Now fix the angle θ. Introduce the function

$$y' = f(x, y, \theta) := y \cos\theta - x \sin\theta.$$

The equation

$$\boxed{f(x, y, \theta) = 0}$$

is called the gauge condition; it determines the x'-axis. We claim that

$$\boxed{\mathcal{J} = \pi \int_{\mathbb{R}^2} \mathrm{e}^{-x^2-y^2}\ \delta(f(x, y, \theta))\, |f_\theta(x, y, \theta)|\ dxdy.} \tag{16.3}$$

This is the prototype of the magic Faddeev–Popov formula. Let us justify this in a formal way. Observe first that the partial derivative reads as $f_\theta(x, y, \theta) = -y \sin\theta - x \cos\theta = -x'$. By (i),

$$\mathcal{J} = \pi \int_{\mathbb{R}^2} \mathrm{e}^{-x'^2-y'^2} \delta(y')|x'|\ dx'dy'.$$

Transforming this into the variables x, y and observing that

$$x'^2 + y'^2 = x^2 + y^2, \qquad \frac{\partial(x', y')}{\partial(x, y)} = 1,$$

we get $\mathcal{J} = \pi \int_{R^2} \mathrm{e}^{-x^2-y^2}\ \delta(y')|x'|\ dxdy$. This is the claim (16.3).

16.6 The Magic Faddeev–Popov Formula and Ghosts

Gauge theory. We now consider the generating functional of a gauge theory model

$$\boxed{Z(J) := \int \mathrm{e}^{\mathrm{i}S[A]+\mathrm{i}J[A]}\ \mathcal{D}A}$$

with the action functional

$$S[A] := \int_{\mathbb{M}^4} \mathcal{L}(A(x), \partial A(x)) d^4x$$

and the source functional $J[A] := \int_{\mathbb{M}^4} J^\mu_s(x) A^s_\mu(x) d^4x$. We use the same notation as in Sect. 16.2. In particular, the gauge field $A = A(x)$ has four components $A_0(x), A_1(x), A_2(x), A_3(x)$ depending on the space-time point $x \in \mathbb{M}^4$. Each component $A_\mu(x)$ lies in the S-dimensional real Lie matrix algebra $\mathcal{LG}$ which is related to the Lie matrix group[6] $\mathcal{G}$. The gauge transformations look like

[6] For example, $\mathcal{G} = SU(N)$ and $\mathcal{LG} = su(N)$ with $N = 2, 3, \ldots$. Then $S = N^2 - 1$.

$$A^g_\mu(x) = G(x)A_\mu(x)G(x)^{-1} - \frac{\mathrm{i}}{\kappa}\,\partial_\mu G(x)\cdot G(x)^{-1}$$

for all $\mu = 0, 1, 2, 3$ and $x \in \mathbb{M}^4$. Here, $G(x) \in \mathcal{G}$ for all x. Explicitly,

$$G(x) = \exp\left(\sum_{s=1}^{S} \theta_s(x)\mathcal{B}_s\right)$$

where $\mathcal{B}_1, \ldots, \mathcal{B}_S$ is a basis of the gauge Lie algebra $\mathcal{LG}$. Since the gauge group element $G(x)$ depends on the real functions $\theta^1(x), \ldots, \theta^S(x)$, we write A^θ instead of A^g. We regard the gauge potential $A(x)$ and the transformed gauge potential $A^\theta(x)$ as physically equivalent quantities. It is crucial that

> *The generating functional has to count equivalent gauge field quantities only once. Therefore, we have to modify the generating functional $Z(J)$.*

In 1967 Faddeev and Popov proposed the following modification. The key formulas read as follows:[7]

(i) Gauge condition:[8] $f(A(x)) = 0$ for all $x \in \mathbb{M}^4$.

(ii) The magic Faddeev–Popov formula:

$$\boxed{Z_{\mathrm{FP}}(J) := \int \mathrm{e}^{\mathrm{i}S[A]+\mathrm{i}J[A]}\ \delta(f(A))\ \det\left(\frac{\delta f(A^\theta)}{\delta\theta}\right)_{|\theta=0} \mathcal{D}A.}$$

We replace the original generating functional $Z(J)$ by the Faddeev–Popov generating functional $Z_{\mathrm{FP}}(J)$. The functional integral $Z_{\mathrm{FP}}(J)$ is to be understood as an infinite-dimensional integral of the form

$$Z_{\mathrm{FP}}(J) = \int \mathrm{e}^{\mathrm{i}S[A]+\mathrm{i}J[A]} \prod_{x\in\mathbb{M}^4} \delta(f(A(x)) \det\left(\frac{\delta f(A^\theta)}{\delta\theta(x)}\right)_{|\theta=0} dA(x)$$

with the continuum of integration variables $A(x)$ indexed by the space-time points x.

(iii) Ghosts: Motivated by the determinant trick for Grassmann variables from Sect. 9.5 on page 518, Faddeev and Popov defined the formal determinant appearing in (ii) above by the formula

$$\boxed{\det\left(\frac{\delta f(A^\theta)}{\delta\theta(x)}\right)_{|\theta=0} := \int \exp\left(\zeta(x)^d \frac{\delta f(A^\theta)}{\delta\theta(x)}_{|\theta=0} \eta(x)\right) d\eta(x) d\zeta(x).}$$

[7] As an introduction, we recommend Faddeev and Slavnov (1980), Peskin and Schroeder (1995), and Ryder (1999). We will study this in Volume V on the physics of the Standard Model in particle physics.

[8] For example, $f(A) := \partial^\mu A_\mu$. This is the so-called Lorentz gauge condition.

This determinant is called the Faddeev–Popov determinant. Here, we set

$$\zeta(x)^d := (\zeta_1(x), \dots, \zeta_S(x)), \qquad \eta(x) := \begin{pmatrix} \eta^1(x) \\ \vdots \\ \eta^S(x) \end{pmatrix}$$

where $\eta^1(x), \dots, \eta^S(x)$ and $\zeta_1(x), \dots, \zeta_S(x)$ are Grassmann variables depending on the parameter x. For all $r, s = 1, \dots, S$ and all $x \in \mathbb{M}^4$, we postulate the following anticommutation relations:

$$\eta^r(x)\eta^s(x) = -\eta^s(x)\eta^r(x), \quad \zeta_r(x)\zeta_s(x) = -\zeta_s(x)\zeta_r(x),$$
$$\eta^r(x)\zeta_s(x) + \zeta_s(x)\eta^r(x) = \delta^r_s. \tag{16.4}$$

This way, the additional fields $\eta = \eta(x)$ and $\zeta = \zeta(x)$ are introduced, which are called ghost fields and antighost fields, respectively. The point is that the additional ghost and antighost fields force the unitarity of the S-matrix.

16.7 The BRST Symmetry

Historically, the Lagrangian density (16.2) above was obtained by Faddeev and Popov in 1967. They used the Feynman functional integral in order to derive the Feynman rules for scattering processes in gauge theory. [9] They introduced ghosts and antighosts in order to cancel meaningless infinite factors in the generating functional written as functional integral. We will discuss this in Sect. 16.6. From the physical point of view, the most important quantity is the S-matrix.

The S-matrix has to be unitary.

Otherwise, the entries of the S-matrix cannot be interpreted as transition probabilities (see Sect. 7.15 on page 370). The introduction of the Faddeev–Popov ghosts and antighosts is a formal mathematical trick in order to change the S-matrix in such a way that it becomes a unitary operator.

In 1974 Becchi, Rouet, Stora and independently Tyutin noticed that ghosts and antighosts are related to a new kind of symmetry in quantum field theory that nowadays is called BRST symmetry.[10] The idea is to replace the gauge fixing Lagrangian density $\mathcal{L}_{\text{gaugefixing}}$ from (16.2) by

$$\mathcal{L}_{\text{BRST}} := \tfrac{1}{2}\xi \operatorname{tr}(\chi^2) + \operatorname{tr}(\chi \partial^\mu A_\mu)$$

[9] L. Faddeev and V. Popov, Phys. Lett. **25B** (1967), 29–30.
See also B. De Witt, Phys. Rev. **162** (1967), 1195–1256.

[10] C. Becchi, A. Rouet, and R. Stora, Phys. Lett. **52B** (1974), 344–346; Commun. Math. Phys. **42** (1975), 127–162.
M. Iofa and I. Tyutin, Theor. Math. Phys. **27** (1976), 316–322.

where the additional field $\chi(x) = \chi^s(x)\mathcal{B}_s$ lives in the Lie algebra $\mathcal{LG}$ for each space-time point x in $\mathbb{M}^4$. Therefore, the complete Lagrangian density reads as

$$\mathcal{L} = \mathcal{L}_{\text{gaugeboson}} + \mathcal{L}_{\text{matter}} + \mathcal{L}_{\text{BRST}} + \mathcal{L}_{\text{ghost}}.$$

The point is that this Lagrangian density is invariant under the following infinitesimal transformations:

$$\begin{aligned} \delta A_\mu &= \varepsilon dA_\mu, & \delta\psi &= \varepsilon d\psi, \\ \delta\eta &= \varepsilon d\eta, & \delta\zeta &= \varepsilon d\zeta, & \delta\chi &= \varepsilon d\chi. \end{aligned}$$

Explicitly, the BRST operator d is defined in the following way:

$$\begin{aligned} dA_\mu &:= \nabla_\mu \eta, & d\psi &:= \mathrm{i}\kappa(\eta^s \mathcal{B}_s)\psi, \\ d\eta^s &:= \mathrm{i}\kappa c^s_{ab}\eta^a\eta^b, & d\zeta &:= \chi, & d\chi &:= 0. \end{aligned}$$

The precise meaning of infinitesimal transformations is discussed in (7.69) on page 410. The non-classical parameter ε anticommutes with all of the ghost fields, antighost fields and fermionic matter fields. The point is that the BRST operator d satisfies the relation[11]

$$\boxed{d^2 = 0.}$$

The BRST symmetry will be studied in Volume V on the physics of the Standard Model for elementary particles.

16.8 The Power of Cohomology

We want to show that the BRST symmetry is not an exotic mathematical trick discovered by chance, but it is closely related to one of the deepest tools in modern analysis, geometry, and topology called cohomology. Intuitively, cohomology is rooted in

- the existence of potentials for physical fields,
- the integral theorems of Gauss, Green, and Stokes for physical fields,
- the Gauss–Bonnet theorem on the total curvature of two-dimensional surfaces and its generalization to higher dimensions by Chern,
- the theory of Abelian integrals and their Riemann surfaces,
- Poincaré's dual triangulations of polyhedra, the Betti numbers, and the Euler characteristic,
- the theory of differential forms developed by Élie Cartan and Poincaré at the end of the 19th century and completed by de Rham in the 1930s (de Rham cohomology),

[11] We use the symbol d in order to emphasize the relation to cohomology theory in mathematics. For historical reasons, physicists write Q instead of d.

- the Hodge theory for higher-dimensional Riemann manifolds from the 1930s,
- the Riemann–Roch–Hirzebruch theorem from the 1950s,
- the Atiyah–Singer index theorem from the 1960s,
- and the Ritz combination principle for atomic spectra.

This will be studied in Volume IV on quantum mathematics. In particular, we will show there that cohomology is intimately related to electric circuits (the Kirchhoff rules), the Maxwell equations in electrodynamics, and Dirac's magnetic monopoles. Indeed, the integration of the Maxwell equations via four-potentials is cohomology in action. At this point, we restrict ourselves to sketching a few basic ideas.

Poincaré's boundary operator ∂. In order to study the qualitative (i.e., the topological) properties of geometric objects, Poincaré considered the equation

$$\boxed{S = \partial B.} \tag{16.5}$$

The set S is called the boundary of the set B. Moreover, a set C is called a cycle iff it has no boundary. We write $\partial C = 0$. Typical cycles are circles and spheres.

> *Poincaré studied cycles modulo boundaries. That is, he studied cycles by putting boundaries equal to zero.*

This leads to the concept of homology group which was introduced by Emmy Noether in the 1920s.[12] For example, if B is a ball in 3-dimensional space, then $S = \partial B$ is a sphere, and S is a cycle, that is, $\partial S = 0$. Therefore, we get the crucial relation

$$\partial^2 B = 0.$$

It turns out that operators D with the typical property $D^2 = 0$ appear quite often in mathematics and physics. In such cases, one can apply the methods of homological algebra which lead to deep results.[13]

Élie Cartan's coboundary operator d. As a further example, let us consider the equation

$$\boxed{\omega = d\mu} \tag{16.6}$$

for differential forms ω and μ. We are given the field ω, and we are looking for a potential μ.[14] The crucial Poincaré lemma tells us that

[12] Henri Poincaré (1854–1912), Emmy Noether (1882–1935).

[13] The classical standard text on homological algebra is the monograph by H. Cartan and S. Eilenberg, Homological Algebra, Princeton University Press, 1956. We also recommend G. Bredon, Topology and Geometry, Springer, New York, 1993.

[14] For a simple example, see equation (16.7) on page 895.

$$d^2 = 0.$$

Motivated by the boundary operator above, the field ω is called a cocycle iff $d\omega = 0$. Moreover, the cocycle field ω is called trivial iff it is a coboundary, that is, $\omega = d\mu$.

> *Cohomology theory studies cocycles modulo coboundaries. That is, it studies cocycles by putting coboundaries equal to zero.*

It turns out that there exists a crucial duality between homology and cohomology. In terms of physics, this duality relates the geometry of manifolds (e.g., space-time manifolds) to the analytic structure of the physical fields on manifolds (see Volume IV on quantum mathematics).

The BRST operator Q. In the BRST approach, physicists write Q instead of d, and they use the operator Q in order to eliminate ghosts. Let us discuss this.

16.8.1 Physical States, Unphysical States, and Cohomology

> Use only essential physical states.
> Folklore

Consider a linear space X over $\mathbb{K} = \mathbb{R}$ (real space) or $\mathbb{K} = \mathbb{C}$ (complex space). Let $Q : X \to X$ be a linear operator which has the characteristic property

$$\boxed{Q^2 = 0.}$$

The elements $\omega, \mu, \varrho, \ldots$ of X are called states. The operator Q allows us to classify states in the following way:

(i) Physical state ω: The state ω is called a physical state iff $Q\omega = 0$. Similarly, a state ω is called an unphysical state (or a ghost) iff $Q\omega \neq 0$.

(ii) Trivial physical states ω: Each state of the form $\omega = Q\mu$ for some state μ is a physical state.[15] Such states are called trivial physical states.

(iii) Equivalent physical states: Two physical states ω and ϱ are called equivalent iff the difference $\omega - \varrho$ is a trivial physical state. In other words,

$$\omega \sim \varrho \quad \text{iff} \quad \omega = \varrho + Q\mu \quad \text{for some state } \mu.$$

This equivalence relation respects linear combinations. Explicitly, for physical states $\omega, \omega', \varrho, \varrho'$ and numbers $\alpha, \beta \in \mathbb{K}$,[16]

$$\omega \sim \omega', \ \varrho \sim \varrho' \quad \text{implies} \quad \alpha\omega + \beta\varrho \sim \alpha\omega' + \beta\varrho'.$$

[15] Note that $Q\omega = Q^2\mu = 0$.

[16] In fact, if $\omega = \omega' + Q\mu$ and $\varrho = \varrho' + Q\sigma$, then $\alpha\omega + \beta\varrho = \alpha\omega' + \beta\varrho' + Q(\alpha\mu + \beta\sigma)$.

(iv) Essential physical states $[\omega]$: For each physical state ω, the symbol $[\omega]$ denotes the set of all physical states which are equivalent to ω. In other words,

$$[\omega] := \{\omega + Q\varrho : \ \varrho \in X.\}$$

The equivalence classes $[\omega]$ are called essential physical states.

(v) The essential state space X/Q : The space of all essential physical states $[\omega]$ forms a linear space over $\mathbb{K}$. This space is denoted by X/Q. The linear combinations on X/Q are defined by

$$\alpha[\omega] + \beta[\varrho] := [\alpha\omega + \beta\varrho]$$

for all physical states ω, ϱ and all $\alpha, \beta \in \mathbb{K}$. This definition does not depend on the choice of the representatives.[17]

In order to eliminate ghosts, we replace the original state space X by the essential state space X/Q which is also called the cohomology space of X with respect to the BRST operator Q. From the practical point of view, we work with physical states by simply putting trivial physical states equal to zero. For example, if ω, ϱ, σ are physical states and σ is a trivial physical state, then

$$\alpha\omega + \beta\varrho + \gamma\sigma = \alpha\omega + \beta\varrho \qquad \text{for all} \quad \alpha, \beta \in \mathbb{K}.$$

16.8.2 Forces and Potentials

The calculus of differential forms was introduced by Élie Cartan (1869–1951) at the end of the 19th century. This is the most important tool in modern analysis, geometry, and mathematical physics. Let us discuss the basic ideas. To this end, introduce the potential function $U = U(x, y)$ and the force 1-form

$$F = a(x, y)dx + b(x, y)dy$$

on the Euclidean (x, y)-plane $\mathbb{R}^2$. Using the wedge product

$$dx \wedge dx = dy \wedge dy = 0, \qquad dx \wedge dy = -dy \wedge dx,$$

we define the derivative of differential forms:

- $dU := U_x dx + U_y dy$,
- $da = a_x dx + a_y dy$ and $db = b_x dx + b_y dy$;
- $dF = da \wedge dx + db \wedge dy = c dx \wedge dy$ where we set $c := b_x - a_y$;
- $d(dU) = dU_x \wedge dy + dU_y \wedge dx = (U_{yx} - U_{xy})dx \wedge dy = 0$;
- $d(dF) = dc \wedge dx \wedge dy = c_x dx \wedge dx \wedge dy + c_y dy \wedge dx \wedge dy = 0$.

[17] By (iii), if $\omega \sim \omega'$ and $\varrho \sim \varrho'$, then $\alpha[\omega] + \beta[\varrho] = \alpha[\omega'] + \beta[\varrho']$.

Hence $d^2 = 0$. This is called the Poincaré lemma. The key equation reads as

$$\boxed{F = -dU \qquad \text{on } \Omega} \tag{16.7}$$

where Ω is an open subset of $\mathbb{R}^2$. Explicitly,

$$a = -U_x, \qquad b = -U_y \qquad \text{on } \Omega.$$

We are given the smooth force components $a, b : \Omega \to \mathbb{R}$, and we are looking for the smooth function $U : \Omega \to \mathbb{R}$. Introducing the classical force vector field $\mathbf{F}(x, y) := a(x, y)\mathbf{i} + b(x, y)\mathbf{j}$, equation (16.7) reads as

$$\mathbf{F} = -\,\mathbf{grad}\, U \qquad \text{on } \Omega.$$

In classical mechanics, the function U is called a potential of the given force field $\mathbf{F}$.

Necessary solution condition. If equation (16.7) has a smooth solution $U : \Omega \to \mathbb{R}$, then $dF = -d(dU) = 0$. This means that $\mathbf{curl}\, \mathbf{F} = 0$ on Ω, in the language of vector analysis.

Sufficient solution condition. The point is that the necessary solution condition is not always a sufficient condition. Let us discuss two different situations.

(i) Choose $\Omega := \mathbb{R}^2$. Equation (16.7) has a smooth solution $U : \mathbb{R}^2 \to \mathbb{R}$ iff $\mathbf{curl}\, \mathbf{F} = 0$ on $\mathbb{R}^2$. The general solution then reads as

$$\boxed{U(\mathbf{x}) = \text{const} - \int_{\mathbf{x}_0}^{\mathbf{x}} \mathbf{F}(\mathbf{y}) d\mathbf{y}, \qquad \mathbf{x} \in \mathbb{R}^2.} \tag{16.8}$$

This integral does not depend on the choice of the smooth path in Ω from the fixed initial point $\mathbf{x}_0$ to the final point $\mathbf{x}$. In terms of physics, the integral $\int_{\mathbf{x}_0}^{\mathbf{x}} \mathbf{F}(\mathbf{y}) d\mathbf{y}$ is the work done by the force field $\mathbf{F}$ when moving a particle from $\mathbf{x}_0$ to the point $\mathbf{x}$.

(ii) Choose $\Omega := \mathbb{R}^2 \setminus \{0\}$. Since the force field $\mathbf{F}$ may have a singularity at the origin, the curve integral (16.8) may depend on the choice of the path from $\mathbf{x}_0$ to $\mathbf{x}$ in Ω. In terms of physics, this means that the work done by the force field may be path-dependent. The precise result reads as follows: Equation (16.7) has a smooth solution $U : \Omega \to \mathbb{R}$ iff both the local condition $\mathbf{curl}\, \mathbf{F} = 0$ on Ω and the global condition

$$\int_{\mathbb{S}^1} \mathbf{F}(\mathbf{y}) d\mathbf{y} = 0$$

are satisfied. The general solution is then given by the path-independent integral (16.8).

Let us reformulate this in terms of cocycles and coboundaries. First consider the case where $\Omega := \mathbb{R}^2 \setminus \{0\}$.

- The force field F is a cocycle iff $dF = 0$ on Ω. In terms of polar coordinates φ, r, each cocycle can be represented as

$$F = dU + \gamma d\varphi \qquad \text{on } \Omega$$

  where γ is a real number. Explicitly, $2\pi\gamma = \int_{\mathbb{S}^1} F$.
- The cocycle F is a coboundary iff $\gamma = 0$.
- The essential physical fields are cocycles by putting coboundaries equal to zero. This yields

$$F = \gamma d\varphi, \qquad \gamma \in \mathbb{R}.$$

  Thus, the space $H^1(\Omega)$ of essential 1-forms is isomorphic to $\mathbb{R}$; we call $H^1(\Omega)$ the first de Rham cohomology group of Ω.

The situation changes completely for $\Omega = \mathbb{R}^2$. Then, each cocycle F, that is, $dF = 0$ on $\mathbb{R}^2$, can be represented as

$$F = dU \qquad \text{on } \mathbb{R}^2.$$

The essential physical fields are cocycles by putting coboundaries equal to zero. Hence $F = 0$. This corresponds to the trivial first cohomology group, $H^1(\mathbb{R}^2) = 0$. Roughly speaking, cohomology theory yields the following general result:

> *The topological complexity of a manifold Ω can be measured by the increasing number of independent physical fields on Ω that have no potential.*

16.8.3 The Cohomology of Geometric Objects

Differential topology studies the qualitative structure of smooth geometric objects by considering differential forms (physical fields) on manifolds. The idea is to assign real linear spaces

$$H^k(M), \qquad k = 0, 1, 2, \ldots$$

to a compact manifold M. The space $H^k(M)$ is called the kth cohomology group of M. Furthermore, the dimension of $H^k(M)$,

$$\beta^k := \dim H^k(M),$$

is called the kth Betti number of M. Finally, the number

$$\boxed{\chi(M) = \beta^0 - \beta^1 + \beta^2 - \ldots}$$

is called the Euler characteristic of M.

The cohomology of the unit circle. We want to use the general scheme from Sect. 16.8.1 in order to compute the de Rham cohomology algebra of

the unit circle. As usual in mathematics, we will write the symbol d instead of Q. The points of the unit circle

$$\mathbb{S}^1 := \{z \in \mathbb{C} : \ |z| = 1\}$$

can be uniquely parametrized by the angle φ. Explicitly, we have $z = \mathrm{e}^{\mathrm{i}\varphi}$ where $-\pi < \varphi \leq \pi$. Let $C^\infty_{2\pi}(\mathbb{R})$ denote the set of all smooth functions

$$f : \mathbb{R} \to \mathbb{R}$$

which possess the period 2π. The smooth functions $f : \mathbb{S}^1 \to \mathbb{R}$ can be identified with the functions from $C^\infty_{2\pi}(\mathbb{R})$.

(i) 0-cochains: The functions f from $C^\infty_{2\pi}(\mathbb{R})$ are called the 0-chains of the unit circle $\mathbb{S}^1$. We define $df := f'(\varphi)d\varphi$.

(ii) 1-cochains: The 1-forms $g(\varphi)d\varphi$ with $g \in C^\infty_{2\pi}(\mathbb{R})$ are called 1-cochains. Using $d\varphi \wedge d\varphi = 0$, we define $d(gd\varphi) := g'd\varphi \wedge d\varphi = 0$.

(iii) The Cartan algebra: The set of all differential forms

$$f + gd\varphi, \qquad f, g \in C^\infty_{2\pi}(\mathbb{R}).$$

forms the so-called Cartan algebra of the unit circle $\mathbb{S}^1$. There exist two operations, namely, the sum and the wedge product. For example,

$$(f_1 + g_1 d\varphi) \wedge (f_2 + g_2 d\varphi) = f_1 f_2 + (f_1 g_2 + g_1 f_2)d\varphi.$$

(iv) The de Rham cohomology algebra $H(\mathbb{S}^1)$ of the unit circle: This algebra consists of all the differential forms

$$\boxed{\alpha + \beta d\varphi, \qquad \alpha, \beta \in \mathbb{R}.}$$

This will be computed below.

(v) Betti numbers and the Euler characteristic of the unit circle: Let $H^k(\mathbb{S}^1)$ denote the linear subspace of all the differential forms of $H(\mathbb{S}^1)$ of order $k = 0, 1$, This is called the kth cohomology group of $\mathbb{S}^1$. Explicitly, the first cohomology group $H^1(\mathbb{S}^1)$ consists of all $\beta d\varphi$ with $\beta \in \mathbb{R}$, and $H^0(\mathbb{S}^1)$ consists of all real numbers α. For the Betti numbers of the unit circle, we get

$$\beta^k := \dim H^k(\mathbb{S}^1), \qquad k = 0, 1.$$

Explicitly, $\beta^0 = \beta^1 = 1$. For the Euler characteristic of the unit circle, we have

$$\chi(\mathbb{S}^1) = \beta^0 - \beta^1 = 0.$$

Finally, let us compute $H(\mathbb{S}^1)$. The key to this is the differential equation

$$\boxed{f' = g, \qquad f \in C^\infty_{2\pi}(\mathbb{R})} \tag{16.9}$$

for a given function $g \in C^\infty_{2\pi}(\mathbb{R})$. Integration of (16.9) yields

$$f(\varphi) = \text{const} + \int_0^\varphi g(\theta)d\theta. \tag{16.10}$$

This is a 2π-periodic function iff

$$\int_0^{2\pi} g(\theta)d\theta = 0. \tag{16.11}$$

Therefore, problem (16.9) has a solution iff condition (16.11) is satisfied. Then, the general solution is given by (16.10).

- The 0-form f is called a cocycle iff $df = 0$. This means that f is constant.
- Each 1-form $gd\varphi$ is a cocycle, since $d(gd\varphi) = 0$.
- The 1-form $gd\varphi$ is a coboundary iff the equation

  $$gd\varphi = df, \qquad f \in C^\infty_{2\pi}(\mathbb{R})$$

  has a solution. By (16.9), this is equivalent to (16.11).
- By definition, the cohomology algebra $H(\mathbb{S}^1)$ consists of all cocycles

  $$\alpha + gd\varphi$$

  by putting coboundaries equal to zero. Let us write

  $$\alpha + gd\varphi = \alpha + (g - \beta)d\varphi + \beta d\varphi$$

  with $\beta := \frac{1}{2\pi}\int_0^{2\pi} g(\theta)d\theta$. Hence

  $$\int_0^{2\pi} (g(\theta) - \beta)d\theta = 0.$$

  Thus, $(g - \beta)d\varphi$ is a coboundary. Setting $(g - \beta)d\varphi = 0$, we obtain that $\alpha + gd\varphi$ is equal to $\alpha + \beta d\varphi$. This is the claim from (iv) above.

The cohomology of the 2-dimensional unit sphere $\mathbb{S}^2$. It can be shown that the de Rham cohomology algebra $H(\mathbb{S}^2)$ of the unit sphere $\mathbb{S}^2$ consists of all the differential forms

$$\alpha + \gamma d\varphi \wedge d\vartheta, \qquad \alpha, \gamma \in \mathbb{R}$$

where φ and ϑ represent the geographic length and the geographic latitude of the sphere, respectively. For example, the kth cohomology groups $H^k(\mathbb{S}^2)$ with $k = 0, 1, 2$ consist of

$$\alpha \text{ with } \alpha \in \mathbb{R}; \qquad 0; \qquad \gamma d\varphi \wedge d\vartheta \text{ with } \gamma \in \mathbb{R},$$

respectively. This yields the Betti numbers $\beta^k := \dim H^k(\mathbb{S}^2)$. Explicitly, $\beta^0 = \beta^2 = 1$ and $\beta^1 = 0$. For the Euler characteristic of the unit sphere, we obtain

$$\chi(\mathbb{S}^2) = \beta^0 - \beta^1 + \beta^2 = 2.$$

Intuitive meaning of the Euler characteristic. We have computed the Euler characteristic in terms of differential forms. An important result of topology tells us that for compact finite-dimensional manifolds, the Euler characteristic can be computed by means of triangulations or cell decompositions as in Sect. 5.6.2ff on page 242.

The cohomology of the 2-dimensional torus $\mathbb{T}^2 := \mathbb{S}^1 \times \mathbb{S}^1$. It can be shown that the de Rham cohomology algebra $H(\mathbb{T}^2)$ of the torus $\mathbb{T}^2$ consists of all differential forms

$$\alpha + \beta_1 d\varphi + \beta_2 d\vartheta + \gamma d\varphi \wedge d\vartheta, \qquad \alpha, \beta_1, \beta_2, \gamma \in \mathbb{R}.$$

Here, φ and ϑ represent the geographic length and the geographic latitude of the torus, respectively. For example, the kth cohomology groups $H^k(\mathbb{T}^2)$ with $k = 0, 1, 2$ consist of

$$\alpha \text{ with } \alpha \in \mathbb{R}; \qquad \beta_1 d\varphi + \beta_2 d\vartheta, \ \beta_1, \beta_2 \in \mathbb{R}; \qquad \gamma d\varphi \wedge d\vartheta, \ \gamma \in \mathbb{R},$$

respectively. This yields the Betti numbers $\beta^0 = \beta^2 = 1$ and $\beta^1 = 2$ along with the Euler characteristic

$$\chi(\mathbb{T}^2) := \beta^0 - \beta^1 - \beta^2 = 0.$$

The following fact is crucial.

> *If we replace the unit circle, the unit sphere, and the torus above by diffeomorphic manifolds, then the Betti numbers, and hence the Euler characteristic remain unchanged.*

16.8.4 The Spectra of Atoms and Cohomology

As a prototype, consider the hydrogen atom. By Bohr's model, the electron of the hydrogen atom attains the energy levels $E_1 < E_2 < E_3 < \dots$ If the electron jumps from the upper level E_m to the lower level E_n with $m > n$, then a photon of energy

$$\hbar\omega_{mn} = E_m - E_n \tag{16.12}$$

is emitted.[18] Using

- the energy 0-cochain $E := \{E_1, E_2, \dots\}$ and
- the frequency 1-cochain $\omega := \{\omega_{mn}\}$,

[18] Recall that we are using the energetic system in this chapter. Thus, $\hbar = 1$.

the energy-frequency relation (16.12) can elegantly be written as

$$\boxed{\hbar\omega = dE.}$$

This so-called Ritz combination principle tells us that the frequencies of a radiating atom are not arbitrary, but there is a cohomological structure behind them. Interestingly enough, this structure was used by Heisenberg in order to establish quantum mechanics in 1925. This is one of the reasons why cohomology plays a crucial role in quantum physics.

16.8.5 BRST Symmetry and the Cohomology of Lie Groups

As a further example, we want to show that the classical cohomology of Lie groups is a special model of BRST symmetry. To this end, as in Sect. 16.2, let us choose the gauge Lie group $SU(N)$ with $N = 2, 3, \ldots$ and the corresponding Lie algebra $\mathcal{LG} := su(N)$. In this setting, the following geometric picture is behind BRST symmetry:

- antighosts are left-invariant velocity fields on the gauge group $\mathcal{G}$;
- ghosts are dual objects to antighosts, that is, they are left-invariant covelocity fields on $\mathcal{G}$;
- states are differential forms on $\mathcal{G}$;
- the BRST operator Q coincides with the Cartan derivative d of states; this implies $d^2 = 0$;
- physical states are cocycles with respect to d;
- trivial physical states are coboundaries with respect to d;
- essential physical states are physical states by putting coboundaries equal to zero; the essential physical states are also called the cohomology classes with respect to d.

Let us sketch the basic ideas.[19] Choose a fixed basis $\mathcal{B}_1, \ldots, \mathcal{B}_S$ of the Lie algebra $\mathcal{LG}$. In what follows we will use the following Einstein summation convention: We sum over equal upper and lower Latin indices from 1 to S. For example, the equation

$$[\mathcal{B}_r, \mathcal{B}_s]_- = c_{rs}^k \mathcal{B}_k, \qquad r, s = 1, \ldots, S$$

uniquely determines the real structure constants c_{rs}^k of the Lie algebra $\mathcal{LG}$.

(i) Temperature field on the Lie group $\mathcal{G}$. The symbol $C^\infty(\mathcal{G}, \mathbb{R})$ denotes the set of all smooth functions

$$T : \mathcal{G} \to \mathbb{R}.$$

For example, think of a temperature field on the Lie group $\mathcal{G}$.

[19] A detailed study of the analysis on manifolds and Lie groups can be found in Choquet–Bruhat et al. (1996), Vol. 1. We also refer to Zeidler (2000a), Vol. 2, Chap. 16 (differential calculus on manifolds) and Chap. 17 (Lie groups and Lie algebras).

(ii) Antighosts: For $s = 1, \ldots, S$, define

$$\boxed{\mathbf{v}_s(G) := G\mathcal{B}_s \qquad \text{for all} \quad G \in \mathcal{G}.}$$

Since $t \mapsto G\mathrm{e}^{tB_s}$ is a curve on the Lie group $\mathcal{G}$, and we have the time derivative

$$\mathbf{v}_s(G) = \frac{d}{dt} G\mathrm{e}^{t\mathcal{B}_s}|_{t=0},$$

the vector $\mathbf{v}_s(G)$ is a velocity vector (or tangent vector) of the Lie group $\mathcal{G}$ at the point G. Let I denote the unit element of $\mathcal{G}$. Since

$$\mathbf{v}_s(G) = G\mathbf{v}_s(I) \qquad \text{for all} \quad G \in \mathcal{G}, \quad s = 1, \ldots, S,$$

the velocity field $\mathbf{v}_s$ is called left-invariant. The left-invariant velocity fields $\mathbf{v}_1, \ldots, \mathbf{v}_S$ are called antighosts.
Let the symbol $\mathbf{V}(\mathcal{G})$ denote the set of all smooth, left-invariant velocity fields on the Lie group $\mathcal{G}$. Each vector field $\mathbf{v} \in \mathbf{V}(\mathcal{G})$ can be uniquely represented as

$$\mathbf{v}(G) = v^s(G)\mathbf{v}_s(G)$$

for all points $G \in \mathcal{G}$. Here, the velocity components $v^1, \ldots, v^S$ lie in the space $C^\infty(\mathcal{G}, \mathbb{R})$. The directional derivative of the temperature field $T \in C^\infty(\mathcal{G}, \mathbb{R})$ at the point G with respect to the velocity field $\mathbf{v}_s$ is defined by

$$\frac{\delta T}{\delta \mathbf{v}_s}(G) := \frac{d}{dt} T(G\mathrm{e}^{t\mathcal{B}_s})|_{t=0}.$$

As usual, we write $\mathbf{v}_s(T)$ instead of the functional derivative $\frac{\delta T}{\delta \mathbf{v}_s}$.

(iii) Ghosts: Fix $\mathbf{v} \in \mathbf{V}(\mathcal{G})$. For $s = 1, \ldots, S$, define

$$\boxed{\theta^s(v^1\mathbf{v}_1 + \ldots + v^S\mathbf{v}_S) := v^s.}$$

The linear mappings $\theta^1, \ldots, \theta^S : \mathbf{V}(\mathcal{G}) \to C^\infty(\mathcal{G}, \mathbb{R})$ are called ghosts. In particular, $\theta^s(\mathbf{v}_r) = \delta^s_r$ for all $r, s = 1, \ldots, S$.

(iv) Key relations for ghosts and antighosts: For all $r, s = 1, \ldots, S$,

$$\boxed{\theta^r\theta^s = -\theta^s\theta^r, \quad \mathbf{v}_r\mathbf{v}_s = -\mathbf{v}_s\mathbf{v}_r, \quad \theta^s\mathbf{v}_r + \mathbf{v}_r\theta^s = \delta^r_s.} \tag{16.13}$$

Let us explain this notation. For antighosts, we have the Lie product

$$[\mathbf{v}_r, \mathbf{v}_s](G) := G(v(I)w(I) - w(I)v(I)), \qquad r, s = 1, \ldots, S.$$

To simplify notation, we write $\mathbf{v}_r\mathbf{v}_s$ instead of $[\mathbf{v}_r, \mathbf{v}_s]$. For ghosts, there exists the wedge product

$$(\theta^r \wedge \theta^s)(\mathbf{v}, \mathbf{w}) := \theta^r(\mathbf{v})\theta^s(\mathbf{w}) - \theta^r(\mathbf{w})\theta^s(\mathbf{v})$$

for all $\mathbf{v}, \mathbf{w} \in \mathbf{V}(\mathcal{G})$. We write briefly $\theta^r\theta^s$ instead of $\theta^r \wedge \theta^s$. Finally, set $\theta^r \mathbf{v}_s := \theta^r(\mathbf{v}_s) = \delta^r_s$, and let $\mathbf{v}_r\theta^s$ denote the Lie derivative of θ^s with respect to the velocity field $\mathbf{v}_r$. Since θ^s is constant along the velocity field $\mathbf{v}_r$, we get

$$\mathbf{v}_r\theta^s = 0, \qquad s, r = 1, \dots, S.$$

This yields the last formula from (16.13).

(v) States: By definition, a state is a differential form

$$\alpha + \beta_s\theta^s + \gamma_{rs}\theta^r\theta^s + \mu_{rst}\theta^r\theta^s\theta^t \dots \tag{16.14}$$

on the Lie group $\mathcal{G}$. This is a polynomial in the variables $\theta^1, \theta^2, \dots$ with coefficients $\alpha, \beta_s, \dots$ in $C^\infty(\mathcal{G}, \mathbb{R})$. The products $\theta^r\theta^s, \theta^r\theta^s\theta^t, \dots$ are antisymmetric; that is, they change sign under a transposition of two factors.

(vi) The BRST operator d for ghosts: For each state θ, there exists the Cartan derivative $d\theta$, which can be computed in the following way:

- $df := \mathbf{v}_s(f)\theta^s$ if $f \in C^\infty(\mathcal{G}, \mathbb{R})$.
- Maurer–Cartan structure equation:

$$\boxed{d\theta^s := -\tfrac{1}{2}c^s_{kl}\theta^k\theta^l.} \tag{16.15}$$

- By the product rule, $d(\theta^r\theta^s) = d\theta^r \cdot \theta^s - d\theta^s \cdot \theta^r$. Similarly,

$$d(\theta^r\theta^s\theta^k) = d\theta^r \cdot \theta^s\theta^k - d\theta^s \cdot \theta^r\theta^k + d\theta^k \cdot \theta^r\theta^s,$$

and so on.

- Again by the product rule, $d(\beta_s\theta^s) = d\beta_s \cdot \theta^s + \beta_s d\theta^s$. Similarly,

$$d(\gamma_{rs}\theta^r\theta^s) = d\gamma_{rs} \cdot \theta^r\theta^s + \gamma_{rs}d(\theta^r\theta^s),$$

and so on. It turns out that $d(d\theta) = 0$ for all states θ. Briefly,

$$\boxed{d^2 = 0.}$$

(vii) The BRST operator d for antighosts: We define

$$\boxed{d\mathbf{v}_s := \mathbf{v}_s - c^k_{sr}\theta^r\mathbf{v}_k.} \tag{16.16}$$

The point of this definition is that the operator d mixes ghosts and antighosts.

(viii) Physical states: A state θ is called a physical state iff

$$\boxed{d\theta = 0,}$$

that is, θ is a cocycle. By (16.15), ghosts are not physical states.

(ix) The BRST transformation: Choose a parameter ε with $\varepsilon^2 = 0$ and the property that it anticommutes with ghosts and antighosts. By definition, the infinitesimal BRST transformation for ghosts and antighosts is given by

$$\boxed{\delta\theta^s := \varepsilon d\theta^s, \qquad \delta \mathbf{v}_s := \varepsilon d\mathbf{v}_s, \qquad s = 1, \ldots, S.}$$

Explicitly, $\delta\theta^s := -\frac{1}{2}\varepsilon c^s_{kl}\theta^k\theta^l$ and $\delta \mathbf{v}_s := \varepsilon(\mathbf{v}_s - c^k_{sr}\theta^r \mathbf{v}_k)$.

(x) Cohomology groups $H^k(\mathcal{G})$: By a physical k-state we mean a k-form θ with $d\theta = 0$. The physical k-states form the kth cohomology group of the Lie group $\mathcal{G}$ by putting coboundaries equal to zero, that is, if $\mu = d\varrho$, then $\mu = 0$.

Generalization via representation of the Lie algebra $\mathcal{LG}$. The approach considered above can easily be generalized by passing to states with values in the Lie algebra $\mathcal{LG}$ or, more generally, with values in a representation of $\mathcal{LG}$.

(i) By a state with values in the Lie algebra $\mathcal{LG}$, we mean a symbol of the form

$$\omega^s \otimes \mathcal{B}_s$$

where $\omega^1, \ldots \omega^S$ are differential forms on the Lie group $\mathcal{G}$ as introduced in (16.14) above. The BRST operator d is defined by

$$d(\omega^s \otimes \mathcal{B}_s) := d\omega^s \otimes \mathcal{B}_s.$$

The Lie product of such states is defined by

$$[\omega^s \otimes \mathcal{B}_s, \mu^r \otimes \mathcal{B}_r] := (\omega^s \wedge \mu^r) \otimes [\mathcal{B}_r, \mathcal{B}_s].$$

(ii) Let $\varrho : \mathcal{LG} \to \mathcal{L}$ be a Lie algebra morphism into the Lie algebra $\mathcal{L}$. By a state with values in the Lie algebra $\varrho(\mathcal{LG})$, we mean a symbol of the form

$$\omega^s \otimes \varrho(\mathcal{B}_s).$$

In general, we replace $\mathcal{B}_s$ in (i) by $\varrho(\mathcal{B}_s)$.

The goal of each BRST theory is to show that the S-matrix is unitary and the theory is ghost-free; that is, ghosts are never physical states. Applications of the BRST quantization method to quantum field theory and bosonic strings can be found in Weinberg (1995), Vol. 2 and Jost (2001), respectively.

16.9 The Batalin–Vilkovisky Formalism

There exists a powerful general formalism for quantizing gauge field theories which is called the Batalin–Vilkovisky formalism.[20] This can be viewed as a

[20] I. Batalin and G. Vilkovisky, Quantization of gauge theories with linearly dependent generators, Phys. Rev. **28D** (1983), 2567–2582.
For a review, see J. Gomis, J. Paris, and S. Samuel, Phys. Rep. **295** (1995).

far-reaching generalization of the Lagrange multiplier method. The idea is to introduce additional ghost fields, antighost fields, and antifields to all kind of fields. Furthermore, the original classical action functional S is replaced by the modified action functional

$$S + S_1.$$

Here, the additional term S_1 depends on all kind of fields and antifields. The key of the Batalin–Vilkovisky formalism is the formulation of a master equation for $S + S_1$. This is a functional differential equation which is the consequence of a generalized BRST symmetry. The functional derivatives of $S + S_1$ in the master equation refer to both fields and antifields. Roughly speaking, the master equation encodes crucial relations between fields and antifields. These relations are important for

- an effective renormalization procedure and
- for analyzing the violations of the symmetries of the action by quantum effects (anomalies).

As an introduction to this, we recommend Weinberg (1995), Vol. 2, Sect. 15.9. The geometric meaning of the master equation was clarified in the paper by

> M. Aleksandrov, M. Kontsevich, A. Schwarz, and O. Zaboronsky, Geometry of the master equation, Int. J. Mod. Phys. A **12** (1997), 1405–1430.

In modern literature, one calls this the AKSZ master equation.

16.10 A Glance at Quantum Symmetries

Symmetries play a crucial role in the process of quantizing classical theories. For example, the symmetry properties of functional integrals are very useful for studying quantum fields:

- The invariance of functional integrals under gauge transformations leads to identities for the Green's functions called the Ward–Takehashi identities and the Taylor–Slavnov identities.
- It is possible that the functional integral related to the action of a classical field theory does not possess all the symmetry properties of the classical theory. This leads to the so-called anomalies.

We will study this in Volume V on the physics of the Standard Model. We recommend the following monographs:

> M. Peskin and Schroeder, An Introduction to Quantum Field Theory, Addison-Wesley, Reading, Massachusetts, 1995.
>
> O. Piguet and S. Sorella, Algebraic Renormalization: Perturbative Renormalization, Symmetries, and Anomalies. Springer, Berlin, 1995.

E. Evans and Y. Kawahigashi, Quantum Symmetries on Operator Algebras, Clarendon Press, Oxford, 1998.

A. Zee, Quantum Field Theory in a Nutshell, Princeton University Press, 2003.

K. Fujikawa and H. Suzuki, Path Integrals and Quantum Anomalies, Oxford University Press, 2004.

17. A Panorama of the Literature

> There is no branch of mathematics, however abstract, which may not some day be applied to phenomena of the real world.
>
> Nikolai Lobatchevsky (1792–1856)

17.1 Introduction to Quantum Field Theory

We recommend the following books for first reading in quantum field theory:

- S. Chang, Introduction to Quantum Field Theory, World Scientific, Singapore, 1990.
- W. Greiner and J. Reinhardt, Quantum Electrodynamics, Springer, Berlin, 1996.
- W. Greiner and J. Reinhardt, Field Quantization, Springer, Berlin, 1996.
- K. Huang, Quarks, Leptons, and Gauge Fields, World Scientific, Singapore, 1992.
- K. Huang, Quantum Field Theory: From Operators to Path Integrals, Wiley, New York, 1998.
- L. Ryder, Quantum Field Theory, Cambridge University Press, 1999.
- F. Scheck, Quantized Fields: From Symmetries to Quantum Electrodynamics, Springer, Berlin, 2000 (in German).
- A. Lahiri and B. Pal, A First Book of Quantum Field Theory, Alpha Science International, Pangbourne, India, 2001.
- A. Zee, Quantum Field Theory in a Nutshell, Princeton University Press, 2003.
- N. Straumann, Relativistic Quantum Theory: An Introduction to Quantum Field Theory, Springer, Berlin, 2005 (in German).
- B. Zwiebach, A First Course in String Theory, Cambridge University Press, 2004.
- R. Szabo, An Introduction to String Theory and D-Brane Dynamics, Imperial College Press, London, 2004.
- S. Kar, Non-Commutative Geometry: A Perspective on String and Field Theories, World Scientific, Singapore, 2004.
- S. Abel, A. Faraggi, A. Ibarra, and M. Plümacher (Eds.), String Phenomenology, World Scientific, Singapore, 2002.

Here, the book by Lahiri and Pal is based on the operator approach via Dyson series. The lectures by Zee and Ryder use functional integrals. The two books by Greiner, Reinhardt on field quantization and by Huang on quantum field theory emphasize the interrelationships between different approaches. As an introduction to quantum field theory, we also recommend:

- A. Mandl, Introduction to Quantum Field Theory, Wiley, New York, 1966.
- C. Nash, Relativistic Quantum Fields, Academic Press, New York, 1978.

- L. Faddeev and A. Slavnov, Gauge Fields, Benjamin, Reading, Massachusetts, 1980.
- P. Ramond, Field Theory: A Modern Primer, Addison-Wesley, Reading, Massachusetts, 1990.
- P. Schmüser, Feynman Graphs and Gauge Theories for Experimental Physicists, Springer, Berlin, 1995 (in German).
- M. Veltman, Diagrammatica: the Path to Feynman Diagrams, Cambridge University Press, 1995.
- P. van Baal, A Course in Quantum Field Theory, 2000.
  Internet: `http://rulgm4.leidenuniv.nl/van-baal/FTcourse.html`
- V. Parameswaran Nair, Quantum Field Theory: A Modern Perspective, Springer, New York, 2005.

As an elementary introduction to the Standard Model of particle physics, we recommend:

- W. Cottingham and D. Greenwood, An Introduction to the Standard Model of Particle Physics, Cambridge University Press, 1998.
- K. Sibold, Theory of Elementary Particles, Teubner, Stuttgart, 2001 (in German).

More detailed investigations on the Standard Model of particle physics can be found in the following monographs:

- M. Chaichian and N. Nelipa, Introduction to Gauge Field Theories, Springer, Berlin, 1984.
- O. Nachtmann, Elementary Particle Physics: Concepts and Phenomena, Springer, Berlin, 1990.
- M. Guidry, Gauge Field Theories: An Introduction with Applications, Wiley, New York, 1991.
- W. Greiner and A. Schäfer, Quantum Chromodynamics, Springer, Berlin, 1994.
- W. Greiner and B. Müller, Gauge Theory of Weak Interactions, Springer, New York, 1996.
- M. Böhm, A. Denner, and H. Joos, Gauge Theories of the Strong and Electroweak Interaction, Teubner, Stuttgart, 2001.
- T. Morii, C. Lim, and S. Mukherjee, The Physics of the Standard Model and Beyond, World Scientific, Singapore, 2004.
- S. Narison, QCD (Quantum Chromodynamics) as a Theory of Hadrons: From Partons to Confinement, Cambridge University Press, 2004.

The interrelationships between quantum field theory and many-particle systems at an introductory level are studied in:

- A. Fetter and J. Walecka, Quantum Theory of Many-Particle Systems, McGraw-Hill, New York, 1971.
- H. Haken, Quantum Field Theory of Solids, North-Holland, Amsterdam, 1976.
- J. Kapusta, Quantum Field Theory at Finite Temperature, Cambridge University Press, 1989.
- M. Le Bellac, Quantum and Statistical Field Theory, Clarendon Press, Oxford, 1991.
- R. Feynman, Statistical Physics, 14th edn., Addison Wesley, Reading, Massachusetts, 1998.
- M. Stone, The Physics of Quantum Fields, Springer, New York, 2000.

- J. Asch and A. Joye (Eds.), Mathematical Physics of Quantum Mechanics: Selected Lectures from the Conference *The State of the Art in the Mathematical Physics of Quantum Systems*, QMath9, held in Giens, France, in 2004, Springer, Berlin, 2006 (quantum dynamics, quantum field theory and statistical mechanics, quantum kinetics and Bose–Einstein condensation, disordered systems and random operators, semiclassical analysis and quantum chaos).

As an introduction to quantum mechanics, we recommend:

- P. Hislop and I. Sigal, Introduction to Spectral Theory With Applications to Schrödinger Operators, Springer, New York, 1996.
- S. Gustafson and I. Sigal, Mathematical Concepts of Quantum Mechanics, Springer, Berlin, 2003.
- G. Baym, Lectures on Quantum Mechanics, Benjamin, Menlo Park, California, 1969.
- A. Galindo and P. Pascual, Quantum Mechanics, Vols. 1, 2, Springer, Berlin, 1990.
- A. Peres, Concepts and Methods in Quantum Mechanics, Kluwer, Dordrecht, 1993.
- A. Sudberry, Quantum Mechanics and the Particles of Nature, Cambridge University Press, 1996.
- J. Basdevant, The Quantum–Mechanics Solver: How to Apply Quantum Theory to Modern Physics, Springer, Berlin, 2000.
- J. Basdevant and J. Dalibard, Quantum Mechanics, Springer, Berlin, 2002.
- N. Straumann, Ein Grundkurs über nichtrelativistische Quantentheorie (Quantum mechanics: a basic course of non-relativistic quantum theory), Springer, Berlin, 2002 (in German).
- K. Gottfried and Tung-Mow Yan, Quantum Mechanics: Fundamentals, Springer, New York, 2003.

The most comprehensive textbook in theoretical physics is:

- L. Landau and E. Lifshitz, Course of Theoretical Physics,
  - Vol. 1: Mechanics, Vol. 2: The Classical Theory of Fields,
  - Vol. 3: Quantum Mechanics, Vol. 4: Quantum Electrodynamics,
  - Vol. 5: Statistical Physics, Part 1, Vol. 6: Fluid Mechanics,
  - Vol. 7: Theory of Elasticity, Vol. 8: Electrodynamics of Continuous Media,
  - Vol. 9: Statistical Physics, Part 2, Vol. 10: Physical Kinetics,

  Butterworth–Heinemann, Oxford, 1982.

The most comprehensive textbook on gravitation is:

- C. Misner, K. Thorne, and A. Wheeler, Gravitation, Freeman, San Francisco, 1973.

Modern developments can be found in:

- F. Adams and G. Laughlin, A dying universe: the long-term fate and evolution of astrophysical objects, Rev. Mod. Phys. **69** (1997), 337–372.
- F. Adams and G. Laughlin, The Five Ages of the Universe: Inside the Physics of Eternity, Simon and Schuster, New York, 1999.
- G. Börner, The Early Universe: Facts and Fiction, Springer, Berlin, 2003.
- B. Schutz, Gravity from the Ground Up, Cambridge University Press, 2003.
- N. Straumann, General Relativity with Applications to Astrophysics, Springer, New York, 2004.

- Rendall, A. (1998), Lectures on Nonlinear Hyperbolic Differential Equations (in German), Max-Planck Institute Albert Einstein for Gravitational Physics, Golm/Potsdam, Germany.
  Internet: `http://www.aei-potsdam.mpg.de/rendall/vorlesung1.htlm`
- P. Cruściel and H. Friedrich, The Einstein Equations and the Large Scale Behavior of Gravitational Fields: 50 Years of the Cauchy Problem[1] in General Relativity, Birkhäuser, Boston, 2004.
- D. Liebscher, Cosmology, Springer, Berlin, 2005.
- V. Mukhanov, Physical Foundations of Cosmology, Cambridge University Press, 2005.

For general physics, we recommend:

- R. Feynman, R. Leighton, and M. Sands, The Feynman Lectures in Physics, Addison-Wesley, Reading, Massachusetts, 1963.
- P. Tipler, Physics for Scientists and Engineers, Freeman, New York, 1999.
- C. Gerthsen, Gerthsen Physik. Edited by D. Meschede, Springer, Berlin, 2004 (in German).

17.2 Standard Literature in Quantum Field Theory

The modern approach to quantum field theory is represented in the following two monographs:

- S. Weinberg, Quantum Field Theory, Vols. 1–3, Cambridge University Press, 1995.
- M. Peskin and D. Schroeder, An Introduction to Quantum Field Theory, Addison-Wesley, Reading, Massachusetts, 1999.

Weinberg's treatise emphasizes the general structure of quantum field theories from the physical point of view. The following monographs are classics in quantum field theory:

- N. Bogoliubov and D. Shirkov, Introduction to Quantum Field Theory, 3rd edn., Wiley, New York, 1980 (the first Russian edition was published in 1957).
- S. Schweber, An Introduction to Relativistic Quantum Field Theory, Harper and Row, New York, 1961.
- J. Bjorken and S. Drell, Relativistic Quantum Mechanics, McGraw-Hill, New York, 1964.
- J. Bjorken and S. Drell, Relativistic Quantum Fields, McGraw-Hill, New York, 1965.
- C. Itzykson and J. Zuber, Quantum Field Theory, MacGraw-Hill, New York, 1980.
- J. Zinn-Justin, Quantum Field Theory and Critical Phenomena, Clarendon Press, Oxford, 2004.

We also recommend the following monographs:

[1] The Cauchy problem is also called the initial-value problem. Here, roughly speaking, for given gravitational field at the initial time, we want to determine the gravitational field in the future. This is a highly nontrivial task in the theory of nonlinear hyperbolic partial differential equations (see Vol. III).

- A. Abrikosov, L. Gorkov, and I. Dzyaloshinskii, Methods of Quantum Field Theory in Statistical Physics, Prentice Hall, Englewood Cliffs, New Jersey, 1963.
- A. Barut, The Theory of the Scattering Matrix, MacMillan, New York, 1967.
- A. Fetter and J. Walecka, Quantum Theory of Many-Particle Systems, McGraw-Hill, New York, 1971.
- N. Birell and P. Davies, Quantum Fields in Curved Space, Cambridge University Press, 1982.
- N. Bogoliubov and D. Shirkov, Quantum Fields, Benjamin, Reading, Massachusetts, 1983.
- S. Fulling, Aspects of Quantum Field Theory in Curved Space-Time, Cambridge, University Press, 1989.
- T. Kugo, Gauge Field Theory, Springer, Berlin, 1997 (in German).
- Y. Yang, Solitons in Field Theory and Nonlinear Analysis, Springer, New York, 2001.
- S. Ketov, Quantum Non-Linear Sigma Models: From Quantum Field Theory to Supersymmetry, Conformal Field Theory, Black Holes, and Strings, Springer, Berlin, 2002.
- B. De Witt, The Global Approach to Quantum Field Theory, Vols. 1, 2, Clarendon Press, Oxford, 2003.
- H. Rothe, Lattice Gauge Theories, World Scientific Singapore, 2005.

17.3 Rigorous Approaches to Quantum Field Theory

Since the late 1950s, mathematicians and physicists have tried to give quantum field theory a sound mathematical basis. The basic papers on axiomatic and algebraic quantum field theory are:

- I. Segal, Postulates for general quantum mechanics, Ann. Math. **48** (1947), 930–948.
- A. Wightman, Quantum field theories in terms of, vacuum expectation values, Phys. Rev. **101** (1956), 860–866.
- L. Gårding and A. Wightman, Fields as operator-valued distributions in relativistic quantum field theory, Annals of Physics **16** (1961), 158–176.
- R. Haag and D. Kastler, An algebraic approach to quantum field theory, J. Math. Phys. **5** (1964), 848–861.

In the following volumes, we will come back to this ambitious program. Typical mathematical tools are:

- operator algebras (C^*-algebras, von Neumann algebras),
- local nets of operator algebras,
- the theory of complex-valued analytic functions of several variables, and
- the theory of distributions with values in Hilbert spaces.

At this point, let us recommend the following monographs to the interested reader.

(i) Axiomatic quantum field theory

- R. Streater and A. Wightman, PCT, Spin, Statistics, and All That, Benjamin, New York, 1968.

- M. Reed and B. Simon, Methods of Modern Mathematical Physics (classic), Vol. 1: Functional Analysis, Vol. 2: Fourier Analysis and Self-Adjointness, Vol. 3: Scattering Theory, Vol. 4: Analysis of Operators (Perturbation Theory), Academic Press, New York, 1972.
- B. Simon, The $P(\varphi)_2$-Euclidean Quantum Field Theory, Princeton University Press, 1974.
- N. Bogoliubov, A. Logunov, and I. Todorov, Introduction to Axiomatic Quantum Field Theory, Benjamin, Reading, Massachusetts, 1975.
- N. Bogoliubov, A. Logunov, A. Orsak, and I. Todorov, General Principles of Quantum Field Theory, Kluwer, Dordrecht, 1990 (1200 references).
- J. Glimm and A. Jaffe, Quantum Physics: A Functional Integral Point of View, Springer, New York, 1987.
- D. Iagolnitzer, Scattering in Quantum Field Theory: The Axiomatic and Constructive Approaches, Princeton University Press, 1993.

The relation between quantum theory and Hilbert's theory of syzygies in algebraic geometry can be found in:

- M. Henneaux and C. Teitelboim, Quantization of Gauge Systems, Princeton University Press, 1993 (quantization of constrained systems).
- D. Eisenbud, Commutative Algebra with a View to Algebraic Geometry, Springer, New York, 1994.
- D. Eisenbud, The Geometry of Syzygies: A Second Course in Commutative Algebra and Algebraic Geometry, Springer, New York, 2005.

Furthermore, we recommend the following standard textbooks on algebraic geometry:

- P. Griffith and J. Harris, Principles of Algebraic Geometry, Wiley, New York, 1978.
- I. Shafarevich, Basic Algebraic Geometry, Vols. 1, 2, Springer, Berlin, 1994 (classical methods and modern methods based on Grothendieck's schemes).
- R. Hartshorne, Algebraic Geometry, Springer, New York, 1994 (theory of schemes).
- O. Forster, Lectures on Riemann Surfaces, Springer, Berlin, 1981.

(ii) Operator algebras and algebraic quantum field theory

- R. Haag, Local Quantum Physics: Fields, Particles, Algebras, Springer, New York, 1996 (especially recommended).
- H. Araki, Mathematical Theory of Quantum Fields, Oxford University Press, 1999 (especially recommended).
- G. Emch, Algebraic Methods in Statistical Physics and Quantum Field Theory, Wiley, New York, 1972.
- H. Baumgärtel and M. Wollenberg, Causal Nets of Operator Algebras, Akademie-Verlag, Berlin, 1992.
- H. Baumgärtel, Operator-Algebraic Methods in Quantum Field Theory, Akademie-Verlag, Berlin, 1995.
- R. Verch, The current status of quantum fields in curved space time. Lecture held on the occasion of the 125th anniversary of Einstein's birth, German Physical Society, Ulm 2004. Preprint of the Max-Planck Institute for Mathematics in the Sciences, Leipzig, 2004.
  Internet:`http://www.mis.mpg.de/preprints/`
- R. Kadison and J. Ringrose, Fundamentals of the Theory of Operator Algebras, Vols. 1–4, Academic Press, New York, 1983.

(iii) Applications to models in statistical physics:
- C. Bratelli and D. Robinson, Operator Algebras and Quantum Statistical Mechanics, Vols. 1, 2, Springer, New York, 2002.

(iv) The Epstein–Glaser approach to quantum field theory:
- G. Scharf, Finite Quantum Electrodynamics: the Causal Approach, Springer, New York, 1995.
- G. Scharf, Quantum Gauge Theories: A True Ghost Story, Wiley, New York, 2001.

17.4 The Fascinating Interplay between Modern Physics and Mathematics

In the last years, ideas coming from quantum field theory and string theory have strongly influenced the development of mathematics. Let us summarize some important contributions.

- E. Witten, Physical law and the quest for mathematical understanding, Bull. Amer. Math. Soc. **40** (2003), 21–30.
- M. Atiyah, Mathematics in the 20th Century, Bull. London Math. Soc. **34** (2002), 1–15.
- M. Atiyah, Geometry and Physics of Knots, Cambridge University Press, 1990 (topological quantum field theory).
- M. Atiyah, The impact of Thom's cobordism theory, Bull. Amer. Math. Soc. **41** (3) (2004), 337–340.
- M. Atiyah, Collected Works, Vols. 1–6, Cambridge University Press, 2004.
- F. Hirzebruch, Topological Methods in Algebraic Geometry, Springer, New York, 1966 (first German edition, Springer, Berlin, 1956).
- F. Hirzebruch, Collected Papers, Vols. 1, 2, Springer, Berlin, 1987.
- K. Maurin, Riemann's Legacy: Riemann's Ideas in Mathematics and Physics of the 20th Century, Kluwer, Dordrecht, 1997.

- M. Green, J. Schwarz, and E. Witten, Superstrings, Vols. 1, 2, Cambridge University Press, 1987.
- J. Polchinski, String Theory, Vols. 1, 2, Cambridge University Press, 1998.
- M. Kaku, Strings, Conformal Fields and M-Theory, Springer, New York, 2000.
- B. Zwiebach, A First Course in String Theory, Cambridge University Press, 2004.
- J. Schwarz, String theory: past and present, Séminaire Poincaré 2004, La théorie des cordes (string theory), pp. 23–44, Institut Henri Poincaré, Paris. Internet: `http://www.juisseux.fr.poincare`
- R. Dijkgraaf, The mathematics of string theory, Séminaire Poincaré 2004, La théorie des cordes (string theory), pp. 45–59, Institut Henri Poincaré, Paris. Internet: `http://www.juisseux.fr.poincare`

- P. Deligne, P. Etingof, D. Freed, L. Jeffrey, D. Kazhdan, J. Morgan, D. Morrison, and E. Witten (Eds.), Lectures on Quantum Field Theory: A Course for Mathematicians Given at the Institute for Advanced Study in Princeton, Vols. 1, 2, Amer. Math. Soc., Providence, Rhode Island, 1999.

- E. Witten, A new proof of the positive energy theorem, Commun. Math. Phys. **80** (1981), 381–396.
- E. Witten, Supersymmetry and Morse theory, J. Diff. Geom. **17** (1982), 661–692.
- E. Witten, Topological quantum field theory, Commun. Math. Phys. **117** (1988), 353–386.
- E. Witten, Topological sigma models, Commun. Math. Phys. **118** (1988), 411–449.
- E. Witten, Quantum field theory and the Jones polynomials, Commun. Math. Phys. **212** (1989), 359–399.
- E. Witten, On the structure of the topological phase of two-dimensional gravity, Nucl. Phys. B **340** (1990), 281–332.
- E. Witten, Two-dimensional gravity and intersection theory on moduli spaces, Surv. Diff. Geometry **1** (1991), 243–310.
- N. Seiberg and E. Witten, Electric-magnetic duality, monopole condensation, and confinement in $N = 2$ supersymmetric Yang–Mills theory, Nuclear Phys. B **426** (1994), 19–52.
- N. Seiberg and E. Witten, Monopoles, duality and chiral symmetry breaking in $N = 2$ supersymmetric QCD, Nucl. Physics B **431** (1994), 485–550.
- E. Witten, Monopoles and four-manifolds, Math. Research Letters **1** (1994), 769–796.
- E. Witten, Anti-de Sitter space, thermal phase transition, and confinement in gauge theories, Adv. Theor. Math. Physics, **2** (1998), 505–532
- E. Witten, Lectures on perturbative quantum field theory. In: P. Deligne et al (Eds.) (1999), Vol. 1, pp. 421–509 (see above).
- E. Witten, Lectures on dynamical aspects of quantum field theories. In: P. Deligne et al (Eds.) (1999), Vol. 2, pp. 1119–1424 (see above).
- E. Witten, Lectures on Three-Dimensional Topological Quantum Field Theory. Edited by Sen Hu, World Scientific, Singapore, 1999.
- E. Witten, New perspectives in the quest for unification (1998). Internet: `http://arXiv:hep-th/9812208`

- M. Gromov, Pseudo-holomorphic curves in symplectic manifolds, Invent. Math. **82** (1985), 307-347.
- E. Getzler, The jet-space of a Frobenius manifold and higher-genus Gromov–Witten invariants. In: C. Hertling and M. Marcolli (Eds.), Frobenius Manifolds. Quantum Cohomology and Singularities, Vieweg, Wiesbaden, 2004.

- J. Bourguignon, The unabated vitality of Kählerian geometry. In: E. Kähler, Mathematical Works, pp. 737–766. Edited by R. Berndt and O. Riemenschneider, de Gruyter, Berlin, 2003.
- S. Yau, On the Ricci curvature of a a complex Kähler manifold and the complex Monge–Ampère equation, Comm. Pure Appl. Math. **31** (1978), 339–411.
- D. Morrison, Geometric aspects of mirror symmetry. In: Enquist and Schmid (Eds.), Mathematics Unlimited – 2001 and Beyond, pp. 898–918. Springer, New York, 2001.
- V. Batyrev, Dual polyhedra and mirror symmetry for Calabi–Yau hypersurfaces in toric varieties, J. Alg. Geom. **3**(3) (1994), 493–535.
- V. Batyrev, I. Ciocan-Fontanine, B. Kim, and D. Van Straten, Conifold transitions and mirror-symmetry for Calabi–Yau complete intersections in Grassmannians, Nucl. Phys. B **514**(3) (1998), 640–666.
- V. Batyrev and E. Materov, Toric residues and mirror symmetry, Mosc. Math. J. **2**(3) (2002), 435–475.

- A. Schwarz, Quantum Field Theory and Topology, Springer, Berlin, 1993.
- A. Schwarz, Topology for Physicists, Springer, Berlin, 1994.
- J. Jost, Riemannian Geometry and Geometric Analysis, Springer, Berlin, 2002.
- K. Marathe and G. Martucci, The Mathematical Foundations of Gauge Theories, North-Holland, Amsterdam, 1992.
- K. Marathe, G. Martucci, and M. Francaviglia, Gauge theory, geometry, and topology, Seminario di Matematica dell' Università di Bari **262** (1995), 1–90.
- G. Naber, Topology, Geometry, and Gauge Fields, Springer, New York, 1997.
- B. Felsager, Geometry, Particles, and Fields, Springer, New York, 1997.
- W. Nahm, N. Craigie, and P. Goddard, Monopoles in Quantum Field Theory, World Scientific, Singapore, 1982.
- D. Freed and K. Uhlenbeck, Instantons and Four-Manifolds, Springer, New York, 1984.
- N. Hitchin, G. Segal, and R. Ward, Integrable Systems, Twistors, Loop Groups, and Riemann Surfaces, Oxford University Press, 1999.
- S. Donaldson and P. Kronheimer, The Geometry of Four-Manifolds, Oxford University Press, 1990.
- J. Moore, Lectures on Seiberg–Witten Invariants, Springer, Berlin, 1996 (elementary introduction).
- S. Donaldson, The Seiberg–Witten Equations and 4-Manifold Topology, Bull. Amer. Math. Soc. **33** (1996), 45–70.
- S. Donaldson, Floer Homology Groups, Cambridge University Press, 2002.
- M. Schwarz, Morse Homology, Birkhäuser, Basel, 1993.

- I. Buchbinder and S. Kuzenko, Ideas and Methods of Supersymmetry and Supergravity or A Walk Through Superspace, Institute of Physics, Bristol, 1995.
- G. Tynman, Supermanifolds and Supergroups: Basic Theory, Kluwer, Dordrecht, 2004.
- J. Wess and B. Zumino, Supergauge transformations in four dimensions, Nucl. Phys. B **70** (1974), 39–50 (basic paper on supersymmetry in elementary particle physics).
- J. Wess and J. Bagger, Supersymmetry and Supergravity, Princeton University Press, 1991.
- E. Getzler, A short proof of the local Atiyah–Singer index theorem, Topology **25** (1986), 111–117 (supersymmetric proof).
- N. Berline, E. Getzler, and M. Vergne, Heat Kernels and Dirac Operators, Springer, New York, 1991.
- P. Gilkey, Invariance Theory, the Heat Equation, and the Atiyah–Singer Index Theorem, CRC Press, Boca Raton, 1995.
- K. Kirsten, Spectral Functions in Mathematics and Physics, Chapman, Boca Raton, 2002.
- M. Chaichian and R. Hagedorn, Symmetries in Quantum Mechanics: From Angular Momentum to Supersymmetry, Institute of Physics, Bristol, 1998.
- E. Evans and Y. Kawahigashi, Quantum Symmetries on Operator Algebras, Clarendon Press, Oxford, 1998.

- J. Fuchs, Affine Lie Algebras and Quantum Groups: An Introduction with Applications in Conformal Field Theory, Cambridge University Press, 1992.
- C. Kassel, M. Rosso, and V. Turaev, Quantum Groups and Knot Invariants, Société Mathématique de France, Paris, 1997 (introductory survey).
- V. Turaev, Quantum Invariants of Knots and 3-Manifolds, de Gruyter, Berlin, 1994.

- E. Verlinde, Fusion rules and modular transformations in two-dimensional conformal field theory, Nucl. Phys. B **333** (1988), 261–288.
- G. Faltings, A proof of the Verlinde formula, J. Alg. Geometry, **3** (1994), 347–374.
- A. Klimyk and K. Schmüdgen, Quantum Groups, Springer, Berlin, 1997.
- M. Majid, Foundations of Quantum Group Theory, Cambridge University Press, 1995.
- M. Majid, A Quantum Groups Primer, Cambridge University Press, 2002.

- Yu. Manin, Algebraic aspects of nonlinear differential equations, J. Soviet Math. **11** (1978), 1–122.
- Yu. Manin, Mathematics and Physics, Birkhäuser, Boston, 1981.
- I. Kobzarev and Yu. Manin, Elementary Particles: Mathematics, Physics, and Philosophy, Kluwer, Dordrecht, 1989.
- Yu. Manin, Strings, Math. Intelligencer **11**(2) (1989), 59–65.
- Yu. Manin, Quantum Groups and Noncommutative Geometry, CRM, Université de Montréal, 1988.
- Yu. Manin, Topics in Noncommutative Geometry, Princeton University Press, 1991.
- Yu. Manin, Gauge Fields and Complex Geometry, Springer, Berlin, 1997.
- B. Dubrovin, Geometry of Two-Dimensional Field Theories (foundation of the theory of Frobenius manifolds). In: R. Donagi, B. Dubrovin, E. Frenkel, and E. Previato (Eds.), Integrable Systems and Quantum Groups, pp. 120–348, Springer, Berlin, 1993.
- Yu. Manin, Frobenius Manifolds, Quantum Cohomology and Moduli Spaces, Amer. Math. Soc., Providence, Rhode Island, 1999.

- A. Connes (1985), Noncommutative differential geometry, Publ. Math. IHES **62**, 257-360.
- A. Connes, Noncommutative Geometry, Academic Press, New York, 1994.
- J. Gracia-Bondia, J. Várilly, and H. Figueroa, Elements of Noncommutative Geometry, Birkhäuser, Boston, 2001.
- A. Connes and J. Lott, Particle models and noncommutative geometry, Nucl. Phys. B (Proc. Suppl.) **18** (1990), 29–47.
- A. Connes, Noncommutative geometry and reality, J. Math. Phys. **36** (1995), 6194–6231.
- A. Chamseddine and A. Connes, The spectral action principle, Commun. Math. Phys. **186** (1997), 731–750.
- A. Chamseddine and A. Connes, Scale invariance in the spectral action, 2005. Internet: `http://hep-th/0512169`
- D. Kastler (1996), A detailed account of Connes' version of the standard model in non-commutative geometry I, II, III, Rev. Math. Phys. **5** (1996), 477–532; **8** (1996), 103–165.
- D. Kastler and T. Schücker, A detailed account of Connes' version of the standard model in non-commutative geometry IV, Rev. Math. Phys. **8** (1996), 205–228.
- F. Scheck, W. Wend, and H. Upmeier (Eds.), Noncommutative Geometry and the Standard Model of Elementary Particle Physics, Springer, Berlin, 2003.

- D. Kreimer, Knots and Feynman Diagrams, Cambridge University Press, Cambridge, 2000.
- A. Connes and H. Moscovici, Hopf algebras, cyclic cohomology and the transverse index theorem, Commun. Math. Phys. **198** (1998), 199–246 (generalization of the Atiyah–Singer index theorem to foliations).

- A. Connes and D. Kreimer, Hopf algebras, renormalization and noncommutative geometry, Commun. Math. Phys. **199** (1998), 203–242.
- A. Connes and D. Kreimer, Renormalization in quantum field theory and the Riemann–Hilbert problem I, II, Commun. Math. Phys. **210** (2000), 249–273, **216** (2002), 215–241.
- B. Duplantier and V. Rivasseau (Eds.), Poincaré Seminar 2002: Vacuum Energy – Renormalization, Birkhäuser, Basel, 2003.

- M. Waldschmidt, P. Moussa, J. Luck, and C. Itzykson (Eds.), From Number Theory to Physics, Springer, New York, 1995 (fourteen beautiful survey articles).
- A. Connes, Noncommutative geometry and the Riemann zeta function, Frontieres and perspectives, International Mathematical Union.
- A. Connes and M. Marcolli, Renormalization and motivic Galois theory (2004). Internet: `http://arXiv:math.NT/0409306`
- A. Connes and M. Marcolli, From number theory to physics I, II (2005). Internet: `http://www.math.dsu.edu/~marcolli/renorm43.pdf`
- M. Marcolli and V. Mathai, Towards the fractional Hall-effect: a noncommutative geometry perspective. In: C. Consani and M. Marcolli (Eds.), Noncommutative Geometry and Number Theory: Where Arithmetic meets Geometry and Physics, pp. 235–261, Vieweg, Wiesbaden, 2006.

- A. Weil, Basic Number Theory, Springer, Berlin, 1974 (adèles and idèles).
- J. Bernstein and S. Gelbart, An Introduction to the Langlands program, Birkhäuser, Boston, 2003.
- G. Bachmann, Introduction to p-adic Numbers and Valuation Theory, Academic Press, New York.
- N. Koblitz, p-adic Numbers, p-adic Analysis and Zeta Functions, Springer, New York, 1977.
- J. Serre, A Course in Arithmetic, Springer, Berlin, 1973.
- R. Rammal, G. Toulouse, and M. Virasoro, Ultrametricity for physicists, Rev. Mod. Phys. **58** (1986), 765 (p-adic numbers in physics).
- R. Hain, Classical polylogarithms, AMS Proc. Symp. Pure Math. **55** (1994), Part 2, pp. 3–42.
- D. Zagier, The Bloch–Wigner–Ramakrishnan polylogarithm function, Math. Ann. **286** (1990), 613–624.
- D. Zagier, Polylogarithms, Dedekind zeta functions, and the algebraic K-theory of fields. In: Algebraic Geometry, Progr. Math. **89**, pp. 391–430, Birkhäuser, Basel, 1991.
- D. Kreimer, The residue of quantum field theory – numbers we should know (structural similarities between the polylogarithm and Green functions in quantum field theory). In: C. Consani and M. Marcolli (Eds.), Noncommutative Geometry and Number Theory, pp. 187–203, Vieweg, Wiesbaden, 2006.
- M. Wodzicki, Noncommutative residue, K-theory, arithmetic and geometry. In: Yu. Manin (Ed.), Lecture Notes in Mathematics **1289**, pp. 320–399, Springer, Berlin, 1987.
- Yu. Manin, Strings, Math. Intelligencer **11**(2) (1989), 59–65 (parallels between string theory and arithmetic geometry – the work of Quillen–Faltings).

- J. Kung (Ed.), Gian-Carlo Rota on Combinatorics: Introductory Papers and Commentaries, Birkhäuser, Basel, 1995.
- H. Figueroa and J. Gracia-Bondia, Combinatorial Hopf algebras in quantum field theory I, Rev. Math. Phys. **17** (2005), 881–975. Internet: `http//:arXiv:hep-th/0408145`

- K. Ebrahim-Fard, J. Gracia-Bondia, Li Guo. and J. Várilly, Combinatorics of renormalization as matrix calculus, Phys. Lett. B **632** (2006), 552–558. Internet: `arXiv:hep-th/0508154`
- C. Brouder, B. Fauser, A. Frabetti, and R. Oeckl, Quantum field theory and Hopf algebra cohomology. J. Phys. A: Math. Gen. 37 (2004), 5895–5927. Internet: arXiv:hep-th/0311253

- L. Boutet de Monvel and V. Guillemin, The spectral theory of Toeplitz operators, Princeton University Press, 1981.
- B. Fedosov, Deformation Quantization and Index Theory, Akademie-Verlag, Berlin, 1996.
- N. Woodhouse, Geometric Quantization, Oxford University Press, 1997.
- M. Schlichenmaier, An Introduction to Riemann Surfaces, Algebraic Curves, and Moduli Spaces, Springer, New York, 1989 (lectures given for physicists including nice hints for further reading).
- J. Harris, An introduction to the moduli space of curves. In: S. Yau (Ed.), Mathematical Aspects of String Theory, pp. 285-312, World Scientific, Singapore, 1987.
- J. Harer and D. Zagier, The Euler characteristic of the moduli space of curves, Invent. Math. **85**(3) (1986), 457–495.
- Yu. Manin and D. Zagier, Higher Weil–Peterson volumes of moduli spaces of stable n-pointed curves, Commun. Math. Phys. **181**(3) (1996), 763–787.
- M. Bordemann, E. Meinrenken, and M. Schlichenmaier, Toeplitz quantization of Kähler manifolds and $gl(n), n \to \infty$ limits, Commun. Math. Phys. **165** (1994), 281–296.
- J. Andersen, Deformation quantization and geometric quantization of abelian moduli spaces, Commun. Math. Phys. **255** (2005), 727–745.

- J. Jost, Compact Riemann Surfaces: An Introduction to Contemporary Mathematics, Springer, Berlin, 1997.
- O. Forster, Lectures on Riemann Surfaces, Springer, Berlin, 1981.
- H. Farkas and I. Kra, Riemann Surfaces, Springer, New York, 1992.
- R. Narasimhan, Compact Riemann Surfaces, Birkhäuser, Basel, 1996 (lectures given at the ETH in Zurich, Switzerland).
- R. Godement, Théorie des faiseaux (sheaf theory), Hermann, Paris, 1964.
- G. Bredon, Sheaf Theory, Springer, New York, 1998.
- D. Mumford, Stability of projective varieties, L'Enseign. Math. **23** (1977), 39–110.
- D. Mumford, Curves and their Jacobians, The University of Michigan Press, Ann Arbor, 1975.
- D. Mumford, Geometric Invariant Theory, Springer, Berlin, 1982.
- D. Mumford, Tata Lectures on Theta I, II, Birkhäuser, Basel, 1983.
- E. Arbarello and M. Cornalba, The Picard group of the moduli space of curves, Topology **26** (1987), 153–171.
- H. Rauch, Singularities of moduli spaces, Bull. Amer. Math. Soc. **68** (1962), 390–394.
- A. Polyakov, Gauge Fields and Strings, Harwood, Chur, Switzerland, 1987.
- A. Beilinson and Yu. Manin, The Mumford form and the Polyakov measure in string theory, Commun. Math. Phys. **107** (1986), 359–376.
- A. Belavin and V. Knishnik, Complex geometry and theory of quantum strings, Soviet Phys. JETP **64** (1987), 214.
- D. Quillen, Determinants of Cauchy–Riemann operators over a Riemann surface, Funct. Anal. Appl. **19** (1985), 31–34.

- J. Jost, The Bosonic String: A Mathematical Treatment, International Press, Boston, 2001.

- M. Aleksandrov, M. Kontsevich, A. Schwarz, and O. Zaboronsky, Geometry of the master equation, Int. J. Mod. Phys. **A12** (1997), 1405–1430.
- A. Cattaneo and G. Felder, A path integral approach to the Kontsevich quantization formula, Commun. Math. Phys. **212** (2000), 591–611.
- P. Cartier, Mathemagics, A tribute to L. Euler and R. Feynman, Séminaire Lotharingien **44** (2000), 1–71.
- P. Cartier, A mad day's work: from Grothendieck to Connes and Kontsevich. The evolution of concepts of space and symmetry. Bull. Amer. Math. Soc. **38**(4) (2001), 389–408.

- R. Fernández, J. Fröhlich, and D. Sokal, Random Walks, Critical Phenomena, and Triviality in Quantum Field Theory, Springer, Berlin, 1992.
- C. Grosche and F. Steiner, Handbook of Feynman Path Integrals, Springer, New York, 1998 (950 references).
- G. Johnson and M. Lapidus, The Feynman Integral and Feynman's Operational Calculus, Clarendon Press, Oxford, 2000.
- M. Chaichian and A. Demichev, Path Integrals in Physics, Vol. 1: Stochastic Processes and Quantum Mechanics; Vol. 2: Quantum Field Theory, Statistical Physics, and Other Modern Applications, Institute of Physics, Bristol, 2001.
- J. Klauder, Beyond Conventional Quantization, Cambridge University Press, 2000.

- D. Giulini, C. Kiefer, and C. Lämmerzahl (Eds.), Quantum Gravity: From Theory to Experimental Search, Springer, Berlin, 2003.
- C. Kiefer, Quantum Gravity, Oxford University Press, 2005.
- A. Ashtekar and J. Lewandowski, Background independent quantum gravity: A status report, Class. Quant. Grav. **21** (2004), R53–R152.
- L. Randall and R. Sundrum, Large mass hierarchy from a small extra dimension, Phys. Rev. Lett. **83** (1999), 4690–4693 (d-branes in cosmology).
- P. Mannheim, Brane-Localized Gravity, World Scientific, Singapore, 2005 (models of d-branes in cosmology).
- L. Baulieu, J. de Boer, B. Pioline, and E. Rabinovici (Eds.), String Theory: From Gauge Interactions to Cosmology, Proceedings of the NATO Advanced Study Institute, Cargèse, France, 2004, Springer, Dordrecht, 2006.
- F. Finster, The Fermionic Projector, American Mathematical Society and International Press, Boston, 2006.
- B. Fauser, J. Tolksdorf, and E. Zeidler (Eds.), Quantum Gravity: Status and Ramifications, Birkhäuser, Basel, 2006.

We will study these fascinating topics in later volumes.

17.5 The Monster Group, Vertex Algebras, and Physics

> The finite Monster group brings together modern algebra and conformal quantum field theory.
>
> Folklore

> The study of scattering processes for strings leads to fascinating mathematical structures like moduli spaces of Riemann surfaces, Kähler manifolds, Calabi–Yau manifolds, spectral geometry, zeta functions, Fock spaces, super Lie algebras, Kac–Moody algebras, Heisenberg algebras, Virasoro algebras, vertex operator algebras, modular forms, Hodge invariants and mirror symmetry. This culminates in conformal quantum field theory.
>
> Folklore

The finite Monster group. As an example of the interplay between mathematics and modern physics, let us briefly discuss the monster group. Symmetries in nature are described by groups. There arises the problem of classifying all possible symmetries (i.e., all possible groups). This problem is solved for

(i) all crystal symmetries,
(ii) all simple, simply connected, compact Lie groups,[2] and
(iii) all finite simple groups.

A group G is called simple iff it has only trivial realizations. This means that each group epimorphism $\chi : G \to H$ is an isomorphism or H consists only of the unit element. The study of finite groups dates back to

- the work of Gauss (1777–1855) on cyclotomic fields in 1796 (the construction of the regular 17-polygon with ruler and compass)
- and the work of Galois (1811–1832) on the solvability of algebraic equations in 1831.

Around 1980, problem (iii) above was finally solved by a large group of mathematicians. The full proof contains about ten thousand pages. See the survey article by

- D. Gorenstein, Classifying the finite simple groups, Bull. Amer. Math. Soc. **14** (1986), 1–98.

It turns out that there exist

(a) 16 infinite families of groups of Lie type (matrix groups over finite fields),
(b) the alternating groups of n letters with $n \geq 5$, and
(c) 26 sporadic simple groups.

The largest finite simple sporadic group is called Monster; its number of elements is equal to

$$2^{46} \cdot 3^{20} \cdot 5^9 \cdot 7^6 \cdot 11^2 \cdot 13^3 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71.$$

The existence of the Monster was conjectured by Fischer and Griess in 1973. Nine years later, Griess constructed the Monster group: .

- R. Griess, The friendly giant, Invent. Math. **69** (1982), 1–102.

The monstrous moonshine module. There arises the following question:

[2] For example, the gauge groups $SU(2)$ and $SU(3)$ of the Standard Model in particle physics are of this type. The classification can be found in Simon (1996).

Does the Monster group describe the symmetry of some mathematical object?

This object is called monstrous moonshine. It turns out that the positive answer to this question is closely related to appropriate models in string theory of physicists. The highly sophisticated construction of the monstrous moonshine module V as a vertex algebra can be found in

- I. Frenkel, J. Lepowski, and A. Meurman, Vertex Operator Algebras and the Monster, Academic Press, New York, 1988.

The final, very detailed answer about the action of the Monster group on V was given by

- R. Borcherds, Monstrous moonshine and monstrous Lie super algebras, Invent. Math. **109** (1992), 405–444.

Borcherds was awarded the Fields medal in 1998. Let us sketch the basic ideas.

(M) The monstrous moonshine module V: There exists an infinite direct sum

$$V = \bigoplus_{n=-1}^{\infty} V_n$$

of finite dimensional linear spaces $V_{-1}, V_0, V_1, ...$ such that the Monster group $\mathcal{M}$ acts as symmetry group on V. This means that, for each element g of the Monster group, there exists a linear isomorphism

$$L(g) : V \to V$$

which respects multiplication; that is, $L(gh) = L(g)L(h)$ for all $g, h \in \mathcal{M}$. In addition, each operator $L(g)$ leaves invariant all of the subspaces $V_{-1}, V_0, V_1, ...$

(D) The dimension of the invariant subspaces: If we introduce the generating function

$$\dim(q) := \sum_{n=-1}^{\infty} (\dim V_n) q^n,$$

then

$$\dim(q) = j(\tau) - 744 = q^{-1} + 196\,884q + 21\ \ 492\,760q^2 + ...$$

For example, this yields $\dim V_{-1} = 1, \dim V_0 = 0$, and

$$\dim V_1 = 196\,884.$$

The famous classical modular function $j = j(\tau)$ is defined by

$$j(\tau) := \frac{(1 + 240 \sum_{n=1}^{\infty} \sigma_3(n))^3}{q I\!I_{n=1}^{\infty}(1-q^n)^{24}}$$

for all complex numbers τ with $\Im(\tau) > 0$. Here, we set $q := \mathrm{e}^{2\pi\tau}$, and the function $\sigma_3(n) = \sum_{d|n} d^3$ is the sum of the cubes of the divisors of n. As an introduction to the theory of modular forms, we recommend the survey article by Zagier (1995).

(R) Irreducible representation: There exists a nonzero element $v \in V$ and a decomposition

$$V_1 = \mathrm{span}\{v\} \oplus W$$

such that the monster group leaves the two linear subspaces $\mathrm{span}\{v\}$ and W invariant. In addition, there does not exist any nontrivial linear subspace of W which is invariant under the monster group. In other words, the group of linear isomorphisms of W represents an irreducible representation of the Monster group with $\dim W = 196\,883$.

(d) The characters of the monster representation: Let us introduce the Thomson series

$$T_g(q) := \sum_{n=-1}^{\infty} \mathrm{tr}(L(g)|V_n)$$

where the symbol $\mathrm{tr}(L(g)|V_n)$ denotes the trace of the linear operator $L(g)$ on the linear space V_n. The numbers $\mathrm{tr}(L(g)|V_n)$ are called characters of the representation of the Monster group on V_n.[3] In 1992 Borcherds showed that, for any element g of the Monster group, the Thomson series $T_g(q)$ is a so-called Hauptmodule for a genus zero subgroup of the Lie group $SL(2, \mathbb{R})$.

Borcherds calculated this Thomson series by using the monster Lie algebra. This Lie algebra is constructed as the space of physical states of a bosonic string moving in a $\mathbb{Z}_2$-orbifold $M/\mathbb{Z}_2$ of a 26-dimensional torus M.

Vertex operator algebras. For the investigation of the monstrous moonshine module V, it is crucial that this module carries an additional structure: V is a vertex operator algebra (briefly called vertex algebra). In conformal quantum field theory, vertex algebras are called chiral algebras by physicists. This is related to superstring theory. As an introduction to vertex algebras, we recommend:

- V. Kac, Vertex Algebras for Beginners, Amer. Math. Soc., Providence, Rhode Island, 1996.
- Xiaoping Xu, Introduction to Vertex Operator Superalgebras and Their Modules, Kluwer, Dordrecht, 1998.
- I. Frenkel and D. Ben-Zvi, Vertex Algebras and Algebraic Curves, Amer. Math. Soc., Providence, Rhode Island, 2001.

[3] It is shown in representation theory that the irreducible representations of a finite group are determined by the characters. See Simon (1996).

A look at the history of string theory. In 1968, in an axiomatic way, Veneziano formulated typical properties of the four-point scattering amplitude of elementary particles in strong interaction.

- G. Veneziano, Construction of a crossing-symmetric, Regge-behaved amplitude for linearly rising trajectories, Nuovo Cimento **57A** (1968), 190–197.

He was looking for a function which satisfies his axioms, and he discovered that the Euler beta function

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt, \qquad x,y > 0$$

fits the axioms well. He also introduced vertex operators as a technical tool for describing scattering amplitudes at a vertex of the Feynman diagrams. He used this in order to prove factorization properties of the amplitudes for n scattered particles. In 1970 Nambu and other physicists realized that the scattering amplitudes could be understood in terms of quantized relativistic strings. Around 1973 it was discovered that string theory can incorporate gravitation:

- T. Yoneya, Quantum gravity and the zero-slope limit of the generalized Virasoro model, Nuovo Cimento Lett. **8** (1973), 951–955.
- J. Scherk and J. Schwarz, Dual models for non-hadrons, Nucl. Phys. B **81**, 118–144.

This was the birth of string theory as a candidate for a theory of all fundamental interactions in nature. In 1981 Polyakov applied the idea of the Feynman functional integral to string theory.

- A. Polyakov, Quantum geometry of bosonic strings, Phys. Lett **103B** (1981), 207–211.

He constructed a bridge to the classical mathematical theory of Riemann surfaces. In this setting, the states of a bosonic string are described by Riemann surfaces. Since the Lagrangian of string theory is invariant under conformal transformations, we have to assume that two Riemann surfaces represent the same string state iff they are conformally equivalent. Thus, the Feynman functional integral has to be computed with respect to the conformal equivalence classes of Riemann surfaces. In other words, one has to integrate over Riemann's classical moduli space. In 1984 Belevanin, Polyakov, and Zamolodchikov abstracted from string theory a general setting in terms of conformal invariance.

- A. Belevanin, A. Polyakov, and A. Zamolodchikov, Infinite conformal symmetries in two-dimensional quantum field theory, Nucl. Phys. B **241** (1984), 333–380.

This was the birth of conformal quantum field theory. The nice properties of conformal quantum field theory rely on the rich structure of conformal transformations in two dimensions, which lies at the heart of the classical

theory of complex analytic functions. Nowadays string theory is a rich source for getting beautiful new ideas in mathematics. The questions and conjectures formulated by physicists are very stimulating for recent mathematical research. Important contributions are due to Edward Witten, who says that

> *String theory is twenty-first century physics that fell incidentially into the twentieth century.*

For the history of string theory, we recommend both the book by Brian Greene (1999) (the two string revolutions) and the collection of important papers in string theory edited by John Schwarz (1985).

17.6 Historical Development of Quantum Field Theory

For the history of modern cosmology and the history of the Standard Model in particle physics, we recommend:

- E. Wigner, Philosophical Reflections and Sytheses. Annotated by G. Emch. Edited by J. Mehra and A. Wightman, Springer, Berlin, 1995 (epistemology of quantum mechanics, quantum-mechanical measuring process, consciousness, symmetries, relativity, nuclear physics, broader philosophical essays).
- E. Wigner, The unreasonable effectiveness of mathematics in the natural sciences, Richard Courant Lecture in Mathematical Sciences delivered at New York University, May 11, 1959, Comm. Pure Appl. Math. **13** (1) (1960), 222–237. Reprinted in the volume Wigner (1995) above, pp. 534–549.
- H. Kragh, Quantum Generations: A History of Physics in the Twentieth Century, Princeton University Press, 2000.
- R. Omnès, The Interpretation of Quantum Mechanics, Princeton University Press, Princeton, New Jersey, 1994.
- S. Schweber, QED (Quantum Electrodynamics) and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga, Princeton University Press, 1994 (1300 references).
- L. O'Raifertaigh, The Dawning of Gauge Theory, Princeton University Press, 1997.
- G. Emch, Mathematical and conceptual foundations of 20th century physics, North-Holland, Amsterdam, 1984.
- Tian Yu Cao, Conceptual Developments of 20th Century Field Theories, Cambridge University Press, 1998 (1100 references).
- E. Regis, Who Got Einstein's Office? Eccentricity and Genius at the Institute for Advanced Study in Princeton, Addison-Wesley, Reading, Massachusetts, 1989.
- L. Brown (Ed.), Renormalization: From Lorentz to Landau and Beyond, Springer, New York, 1993.
- G. Smoot and K. Davidson, Wrinkles in Time, Morrow, New York, 1993 (history of modern cosmology).
- K. Thorne, Black Holes & Time Warps, Einstein's Outrageous Legacy, Norton, New York, 1993.
- R. Brennan, Heisenberg Probably Slept Here: The Lives, Times, and Ideas of the Great Physicists of the 20th Century, Wiley, New York, 1997.
- B. Greene, The Elegant Universe: Supersymmetric Strings, Hidden Dimensions, and the Quest for the Ultimate Theory, Norton, New York, 1999.

- M. Veltman, Facts and Mysteries in Elementary Particle Physics, World Scientific, Singapore, 2003.
- D. Kaiser, Drawing Theories Apart: The Dispersion of Feynman Diagrams in Postwar Physics, The University of Chicago Press, 2005.
- R. Laughlin, A Different Universe: Reinventing Physics from the Bottom Down, Basic Books, New York, 2005.

A survey on the most important papers in quantum field theory can be found on the author's homepage: Internet: `http://www.mis.mpg.de`

17.7 General Literature in Mathematics and Physics

The most important literature can be found in the references to the single volumes. A more comprehensive list can be found on the author's homepage in the Internet:

`http://www.mis.mpg.de/`

This list concerns the following broad field of topics:

- general physics,
- theoretical physics,
- mathematical physics,
- scientific computing and computational physics,
- quantum mechanics,
- the Standard Model in particle physics,
- the Standard Model in cosmology,
- quantum fields and strings,
- statistical physics and solid state physics,
- renormalization,
- symmetry and groups,
- algebra,
- number theory,
- geometry,
- topology,
- analysis,
- stochastics,
- collected works,
- history of mathematics,
- the mathematics of the 20th century,
- history of physics,
- philosophy of the sciences.

17.8 Encyclopedias

Let us mention the following encyclopedias:

- Encyclopedia of Physics, Handbuch der Physik edited by S. Flügge, Vols. 1ff, Springer, Berlin, 1956ff.
- Encyclopedia of Mathematics, Vols. 1–10. Edited by M. Hazewinkel, Kluwer, Dordrecht, 1987ff.
- Encyclopedia of Mathematical Sciences, Vols. 1–60ff, Springer, Berlin, New York, 1990ff.
- Encyclopedia Britannica, Vols. 1–32, Chicago, 1987ff.
- Lexikon der Physik, Vols. 1–6, Spektrum, Wiesbaden, 1998 (in German).
- Lexikon der Mathematik, Vols. 1–6, Spektrum, Wiesbaden, 2001 (in German).
- Lexicon of Physics, Vols. 1ff, Wiley, New York, 2000.
- Handbook of Differential Equations, Vols. 1, 2. Edited by C. Dafermos et al., Elsevier, Boston, 2005.
- The Concise Handbook of Algebra. Edited by A. Mikhalev and G. Pilz, Kluwer, Dordrecht, 2002.

Concerning mathematical physics, we draw the attention of the reader to:

- Encyclopedia of Mathematical Physics, Vols. 1–4. Edited by J. Françoise, G. Naber, and T. Tsun, Elsevier, Oxford, 2006.

A glossary of modern physics is contained in the monograph:

- J. Polchinski, String Theory, Vols. 1, 2, Cambridge University Press, 1998.

Translations of technical terms in mathematics and physics from English into French, German, Russian and vice versa can be found in:

- G. Eisenreich and R. Sube, Mathematik A–Z, Verlag Technik, Berlin, 1985 (35 000 entries).
- R. Sube and G. Eisenreich, Physik A–Z, Vols. 1, 2, Verlag Technik, Berlin, 1985 (75 000 entries).

A panorama of modern mathematics and its applications to real life problems is presented in:

- B. Engquist and W. Schmid (Eds.), Mathematics Unlimited – 2001 and Beyond, Springer, New York, 2001 (80 articles written by leading experts).

17.9 Highlights of Physics in the 20th Century

Many of the highly influential papers of 20th century physics along with commentaries can be found in

- H. Stroke (Ed.), The Physical Review: The First 100 Years. A Selection of Seminal Papers and Commentaries, American Institute of Physics, New York, 1995.

This monumental volume contains 1250 pages and one CD with about 1000 papers. The excellent commentaries concern the following fields:

- atomic physics,
- nuclear physics,
- statistical physics,
- gravity physics and cosmology,
- cosmic radiation,
- condensed matter,
- plasma physics,
- elementary particle physics experiments,
- particle theory,
- science and technology.

In December 1900, Max Planck lectured on his famous radiation law, which initiated quantum physics in the 20th century. On the occasion of this anniversary, the German Physical Society organized an outstanding conference entitled

Quantum Theory Centenary in Berlin in December 2000.

The survey lectures of this conference reflect the state of the art; they are collected in the journal *Annalen der Physik* **9** (2000), 11/12 and **10** (2001), 1/2. On the occasion of Heisenberg's 100th birthday, the Max–Planck Institute for Physics Werner Heisenberg in Munich organized a conference. The lectures given by outstanding physicists can be found in

- G. Buschhorn and J. Wess (Eds.), Fundamental Physics: Heisenberg and Beyond, Springer, Berlin, 2004.

Developments of contemporary theoretical and mathematical physics have been summarized in the volumes of the

- Les Houches Lectures, North-Holland, Amsterdam, 1960ff.

Concerning the state of the art in mathematical physics, we refer to the following proceedings:

- Proceedings of the XIIIth International Congress on Mathematical Physics, London, United Kingdom, 2000. Edited by A. Fokal et al., Imperial College Press, River Edge, New York.
- Proceedings of the XIVth International Congress on Mathematical Physics, Lisbon, Portugal, 2003. Edited by J. Zambrini, World Scientific, Singapore.

The International Congress on Mathematical Physics takes place every three years.

17.10 Actual Information

For non-specialists in physics, who want to become familiar with recent developments in physics, we recommend reading the two journals *Scientific American* and *Physics Today*. Specialists like to consult *Physical Review Letters* and *Physics Letters*. The two mainstream journals in elementary particle physics are *The Physical Reviews D* and *Nuclear Physics B*.

For non-specialists in mathematics, we recommend the *Mathematical Intelligencer* and the *Notices of the American Mathematical Society*. The *Bulletin of the American Mathematical Society* contains important survey articles along with nice referee reports about newly published books. These reports emphasize the historical development of the subject. The standard journal on rigorous approaches to questions in quantum field theory is *Communications in Mathematical Physics.*

The *Poincaré Seminar* was founded in Paris in 2002. This seminar takes place twice a year in Paris. The idea is to describe the state of the art of both theoretical and experimental physics. One volume will be published each year. The first volume,

- B. Duplantier and V. Rivasseau (Eds.), Poincaré Seminar 2002: Vacuum Energy – Renormalization. Birkhäuser, Basel,

concerns modern cosmology, dark matter, the cosmological constant problem and vacuum energy, the Casimir effect, the anomalous magnetic moment of the electron and the muon, modern collider physics (the new LHC collider at CERN), renormalization in physics and mathematics, the renormalization group in classical mechanics (KAM theory), statistical physics, and quantum field theory. The volumes published in 2003 and 2004 concern entropy, Bose-Einstein condensation, the quantum Hall effect, and string theory. The articles can be found on the Internet:

`http://www.juisseux.fr.poincare`

Important fundamental constants in physics are up dated by the Task Group on Fundamental Constants of the Committee on Data for Science and Technology (CODATA) of the International Council of Scientific Unions (ICSU). See CODATA Bulletin **63**, November 1986, and E. Cohen and B. Taylor, Review of Modern Physics **54**(4) 1987).[4] Particle properties are up dated by the *Particle Data Group*. The state of the art can be found on the Internet:

`http://pdg.lbl.gov`

A lot of actual information can be obtained from the following web pages of the leading centers in high energy physics:

[4] A list of important physical constants can be found at the end of Volume 1 of the Handbook of Mathematics by Zeidler (2004).

CERN: `http://www.cern.ch` (Geneva, Switzerland),
DESY: `http://www.desy.de` (Hamburg, Germany),
FERMILAB: `http//:www.fnal.gov` (Batavia, Illinois, U.S.A.),
SLAC: `htpp://www.slac.stanford.edu` (Stanford University, Palo Alto, California, U.S.A.),
KEK: `http://www.kek.jp` (Tsukuba, Japan).

Information about Nobel prize laureates in physics can be found on the following Internet address:

`htpp://www.nobel.se./physics/laureates/index.html`

We also recommend:

- Nobel Prize Lectures, Nobel Foundation, Stockholm, 1954ff.
- M. Dardo, Nobel Laureates and Twentieth-Century Physics, Cambridge University Press, 2004.
- Harenberg Lexikon der Nobelpreisträger (Encyclopedia of Nobel prize laureates), Harenberg Verlag, Dortmund, Germany, 2000 (in German).

For the Fields medal in Mathematics and the Wolf prize in mathematics, see

- M. Atiyah and D. Iagolnitzer (Eds.), Fields Medallists' Lectures, World Scientific, Singapore, 2003.
- S. Chern and F. Hirzebruch (Eds.), Wolf Prize in Mathematics, Vols. 1, 2, World Scientific, Singapore, 2001.

Furthermore, we refer to

- M. Atiyah, Mathematics in the 20th century, Bull. London Math. Soc. **34** (2002), 1–15.

Concerning mathematical quantum field theory, we would like to recommend the following web-sites:

- The "local quantum physics" web page by Detlev Buchholz and Klaus Rehren at the University of Göttingen (Germany) reflects actual developments and gives information on related web pages:

  `htpp://www.lqp.uni-goettingen.de`

- The "algebraic quantum theory" web page of Stephen Summers at the University of Florida in Gainesville (U.S.A.) refers to up-dated important literature:

  `htpp://www.math.ufl.edu/~sjs/aqft`

The state of the art in the theory of gravitation including cosmology and quantum gravity can be found in:

- Living Reviews in Relativity, Max-Planck Institute for Gravitational Physics Albert Einstein, Golm, Germany, 2001ff:

  `http://www.livingreviews.org`

Software systems for computing algebraic Feynma integrals in higher orders of perturbation theory (loop computations) can be found on the Internet:

- the 'FORM' website of the NIKHEF (National Institute for Nuclear Physics and High Energy Physics) in Amsterdam, Netherlands:

  `http://www.nikhef.nl/~form/`

- Mathematica package with interface to 'FORM'

  `http://www.feynarts.de/formcalc/`

- Mathematica package 'FEYNCALC' for the Standard model in particle physics

  `http://www.feyncalc.org/`

Important institutions and data banks in the sciences: Go to the first page of the homepage of the Max-Planck Institute for Mathematics in the Sciences in Leipzig (Germany),

`http://www.mis.mpg.de`

and use the proposed links.

How to answer questions. The reader who is not familiar with some notions or historical details in modern physics should consult:

- J. Gribbin, Q is for Quantum: Particle Physics from A–Z, Weidenfeld, London, 1998.
- J. Rohlf, Modern Physics from α to Z^0, Wiley, New York, 1994.
- Encyclopedia of Mathematical Physics, Vols. 1–4. Edited by J. Françoise, G. Naber, and T. Tsun, Elsevier, Oxford, 2006.

Concerning a survey of modern mathematics, we refer to the following handbooks:

- E. Zeidler (Ed.), Oxford User's Guide to Mathematics, Oxford University Press, New York, 2004 (English edition of the Teubner-Taschenbuch der Mathematik, Teubner, Stuttgart/Leipzig, 2003).
- G. Grosche, D. Ziegler, V. Ziegler, and E. Zeidler (Eds.), Teubner-Taschenbuch der Mathematik, Vol. 2, Teubner, Stuttgart/Leipzig, 2003 (English edition in preparation).

Appendix

For the convenience of the reader, let us collect basic material on the following topics:

- notation,
- physical units, the Planck system, and the energetic system of units,
- the Gaussian system of units and the Heaviside system, and
- the method of dimensional analysis – a magic wand of physicists.

A comprehensive table on the units of the most important physical quantities can be found at the end of the Appendix on page 951.

A.1 Notation

Sets and mappings. The abbreviation 'iff' stands for 'if and only if'. To formulate definitions, we use the symbol ':='. For example, we write

$$f(x) := x^2$$

iff the value $f(x)$ of the function f at the point x is equal to x^2, by definition.

The symbol $U \subseteq V$ (resp. $U \subset V$) means that U is a subset (resp. a proper subset) of V. This convention resembles the symbols $x \leq y$ (resp. $x < y$) for real numbers. A map

$$f : X \to Y$$

sends each point x living in the set X to an image point $f(x)$ living in the set Y. The set X is also called the domain of definition, $\mathrm{dom}(f)$, of the map f. By definition, the image, $\mathrm{im}(f)$, of the map f is the set of all image points $f(x)$. Furthermore, the set

$$f(U) := \{f(x) : \ \ x \in U\}$$

is called the image of the set U by the map f. In other words, by definition, the set $f(U)$ contains precisely all the points $f(x)$ with the property that x is an element of the set U. The set

$$f^{-1}(V) := \{x \in X : f(x) \in V\}$$

is called the pre-image of the set V by the map f.

- The map f is called *surjective* iff each point of the set Y is an image point. In this case, we also say that f maps the set X 'onto' the set Y. The French word 'sur' means 'onto'.
- The map f is called *injective* iff $x_1 \neq x_2$ always implies $f(x_1) \neq f(x_2)$. Such maps are also called 'one-to-one'.

- The map f is called *bijective* iff it is both surjective and injective. Precisely in this case, the inverse map $f^{-1} : Y \to X$ exists.

For each given point y in the set Y, consider the equation

$$\boxed{f(x) = y, \qquad x \in X,} \tag{A.1}$$

that is, we are looking for a solution x in the set X. Observe that the map f is surjective (resp. bijective) iff the equation (A.1) has always at least one (resp. precisely one) solution. The map f is injective iff the equation has always at most one solution.

Inverse map. If the map $f : X \to Y$ is bijective, then the inverse map

$$f^{-1} : Y \to X$$

is defined by $f^{-1}(y) := x$ iff $f(x) = y$.

Sets of numbers. The symbol $\mathbb{K}$ always stands either for the set $\mathbb{R}$ of real numbers or the set $\mathbb{C}$ of complex numbers. The real number x is called positive, negative , nonnegative, non-positive iff

$$x > 0, \qquad x < 0, \qquad x \geq 0, \qquad x \leq 0,$$

respectively. The symbols

$$\mathbb{R}^{\times}, \quad \mathbb{R}_{>}, \quad \mathbb{R}_{<}, \quad \mathbb{R}_{\geq}, \quad \mathbb{R}_{\leq}$$

denote the set of nonzero real numbers, positive real numbers, negative real numbers, nonnegative real numbers, non-positive real numbers, respectively.[5] Concerning the sign of a real number, we write $\mathrm{sgn}(x) := 1, -1, 0$ if $x > 0,\ x < 0,\ x = 0$, respectively.

For a given complex number $z = x+y\mathrm{i}$, we introduce both the conjugate complex number $z^{\dagger} := x - y\mathrm{i}$ and the modulus

$$\boxed{|z| := \sqrt{zz^{\dagger}} = \sqrt{x^2 + y^2}.}$$

The real (resp. imaginary) part of z is denoted by $\Re(z) := x$ (resp. $\Im(z) := y$). The definition of the principal argument, $\arg(z)$, of the complex number z can be found on page 209. Traditionally,

- the symbol $\mathbb{Z}$ denotes the set of integers $0, \pm 1, \pm 2, \ldots$,
- the symbol $\mathbb{N}$ denotes the set of nonnegative integers $0, 1, 2, \ldots$ (also called natural numbers),[6]
- the symbol $\mathbb{N}^{\times}$ denotes the set of positive integers $1, 2, \ldots$, and
- the symbol $\mathbb{Q}$ denotes the set of rational numbers.

For closed, open, and half-open intervals, we use the notation

$$[a, b] := \{x \in \mathbb{R} : a \leq x \leq b\}, \qquad]a, b[:= \{x \in \mathbb{R} : a < x < b\},$$

and $]a, b] := \{x \in \mathbb{R} : a < x \leq b\}$, as well as $[a, b[:= \{x \in \mathbb{R} : a \leq x < b\}$.

The Landau symbols. Around 1900 the following symbols were introduced by the number theorist Edmund Landau (1877–1938). We write

[5] For the closed half-line $\mathbb{R}_{\geq}$, one also uses the symbol $\mathbb{R}_{+}$.

[6] For the set $\mathbb{N}$, one also uses the symbol $\mathbb{Z}_{\geq}$.

$$\boxed{f(x) = o(g(x)) \qquad \text{as} \quad x \to a}$$

iff $f(x)/g(x) \to 0$ as $x \to a$. For example, $x^2 = o(x)$ as $x \to 0$. The symbol

$$f(x) = O(g(x)) \qquad \text{as} \quad x \to a \tag{A.2}$$

tells us that $|f(x)| \leq \text{const}\, |g(x)|$ in a sufficiently small, open neighborhood of the point $x = a$. For example, $2x = O(x)$ as $x \to 0$. We write

$$f(x) \simeq g(x), \qquad x \to a$$

iff $f(x)/g(x) \to 1$ as $x \to a$. For example,

$$\sin x \simeq x, \qquad x \to 0.$$

Relativistic physics. In an inertial system, we set

$$x^1 := x, \quad x^2 := y, \quad x^3 := z, \quad x^0 := ct$$

where x, y, z are right-handed Cartesian coordinates, t is time, and c is the velocity of light in a vacuum. Generally,

- Latin indices run from 1 to 3 (e.g., $i, j = 1, 2, 3$), and
- Greek indices run from 0 to 3 (e.g., $\mu, \nu = 0, 1, 2, 3$).

In particular, we use the Kronecker symbols

$$\delta_{ij} = \delta^{ij} = \delta^i_j := \begin{cases} 1 & \text{if} \quad i = j, \\ 0 & \text{if} \quad i \neq j, \end{cases} \tag{A.3}$$

and the Minkowski symbols

$$\eta_{\mu\nu} = \eta^{\mu\nu} := \begin{cases} 1 & \text{if} \quad \mu = \nu = 0, \\ -1 & \text{if} \quad \mu = \nu = 1, 2, 3, \\ 0 & \text{if} \quad \mu \neq \nu. \end{cases} \tag{A.4}$$

Einstein's summation convention. In the Minkowski space-time, we always sum over equal upper and lower Greek (resp. Latin) indices from 0 to 3 (resp. from 1 to 3). For example, for the position vector, we have

$$\mathbf{x} = x^j \mathbf{e}_j := \sum_{j=1}^{3} x^j \mathbf{e}_j,$$

where $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are orthonormal basis vectors of a right-handed orthonormal system. Moreover,

$$\eta_{\mu\nu} x^\nu := \sum_{\nu=0}^{3} \eta_{\mu\nu} x^\nu.$$

Greek indices are lowered and lifted with the help of the Minkowski symbols. That is,

$$x_\mu := \eta_{\mu\nu} x^\nu, \qquad x^\mu = \eta^{\mu\nu} x_\nu.$$

Hence

$$\boxed{x_0 = x^0, \qquad x_j = -x^j, \qquad j = 1, 2, 3.}$$

For the indices $\alpha, \beta, \gamma, \delta = 0, 1, 2, 3$, we introduce the antisymmetric symbol $\epsilon^{\alpha\beta\gamma\delta}$ which is normalized by

$$\epsilon^{0123} := 1, \tag{A.5}$$

and which changes sign if two indices are transposed. In particular, $\epsilon^{\alpha\beta\gamma\delta} = 0$ if two indices coincide. For example, $\epsilon^{0213} = -1$ and $\epsilon^{0113} = 0$. Lowering of indices yields $\epsilon_{\alpha\beta\gamma\delta} := -\epsilon^{\alpha\beta\gamma\delta}$. For example, $\epsilon_{0123} := -1$.

The Minkowski metric. Unfortunately, there exist two different conventions in the literature, namely, the so-called west coast convention (W) which uses the following Minkowski metric,

$$\boxed{\eta_{\mu\nu} x^\mu x^\nu = c^2 t^2 - x^2 - y^2 - z^2,} \tag{A.6}$$

and the east coast convention (E) based on $-c^2t^2 + x^2 + y^2 + z^2$. (This refers to the east and west coast of the United States of America.) From the mathematical point of view, the east coast convention has the advantage that there does not occur any sign change when passing from the Euclidean metric

$$x^2 + y^2 + z^2$$

to the Minkowski metric. From the physical point of view, the west coast convention has the advantage that the Minkowski square of the momentum-energy 4-vector $(\mathbf{p}, E/c)$ is positive,

$$\eta_{\mu\nu} p^\mu p^\nu = \frac{E^2}{c^2} - \mathbf{p}^2 = m_0^2 c^2. \tag{A.7}$$

Here, m_0 denotes the rest mass of the particle. Since most physicists and physics textbooks use the west coast convention, we will follow this tradition, which dates back to Einstein's papers, Dirac's 1930 monograph *Foundations of Quantum Mechanics* and Feynman's papers. Concerning elementary particles, we use the same terminology as in the standard textbook by Peskin and Schroeder (1995). One can easily pass from our convention to the east coast convention by using the replacements

$$\eta_{\mu\nu} \mapsto -\eta_{\mu\nu}, \qquad \gamma^\mu \mapsto -\mathrm{i}\gamma^\mu$$

for the Minkowski metric and the Dirac-Pauli matrices, γ^μ, from the Dirac equation (A.20), respectively.[7]

A.2 The International System of Units

The ultimate goal of physicists is to measure physical quantities in physical experiments. To this end, physicists have to compare the quantity under consideration with appropriate standard quantities. For example, the measurement of the length of a distance can be obtained by comparing the length with the standard length m (meter). This procedure leads to systems of physical units.

The SI system. In the international system of units, SI (for Système International in French), the following basic units are used:

[7] For example, the east coast convention is used in Misner, Thorne, and Wheeler (1973) and in Weinberg (1995).

Table A.1. Prefixes in the SI system

10^{-1}	deci	d	10	deka	D
10^{-2}	centi	c	10^{2}	hecto	H
10^{-3}	milli	m	10^{3}	kilo	K
10^{-6}	micro	μ	10^{6}	mega	M
10^{-9}	nano	n	10^{9}	giga	G
10^{-12}	pico	p	10^{12}	tera	T
10^{-15}	femto	f	10^{15}	peta	P

- length: m (meter),
- time: s (second),
- energy: J (Joule),
- electric charge: C (Coulomb),
- temperature: K (Kelvin).

Each physical quantity q can be uniquely represented as

$$\boxed{q = q_{\mathrm{SI}} \cdot \mathrm{m}^{\alpha}\mathrm{s}^{\beta}\mathrm{J}^{\gamma}\mathrm{C}^{\mu}\mathrm{K}^{\nu}.} \tag{A.8}$$

Here, q_{SI} is a real number, and the exponents $\alpha, \beta, \gamma, \mu, \nu$ are rational numbers. Physicists say that the physical quantity q has the dimension

$$(\text{length})^{\alpha}(\text{time})^{\beta}(\text{energy})^{\gamma}(\text{electric charge})^{\mu}(\text{temperature})^{\nu}.$$

Let us consider a few examples.

- The unit of mass is the kilogram, $\mathrm{kg} := \mathrm{Js}^2\mathrm{m}^{-2}$.
- The unit of force is the Newton, $\mathrm{N} = \mathrm{Jm}^{-1}$.
- The unit of electric current strength is the ampere, $\mathrm{A} := \mathrm{Cs}^{-1}$.

The physical dimensions of the most important physical quantities in the SI system can be found in Table A.4 on page 951. Instead of meter one also uses kilometer, nanometer, femtometer, and so on, which corresponds to

$$1000\mathrm{m}, \quad 10^{-9}\mathrm{m}, \quad 10^{-15}\mathrm{m},$$

respectively (see Table A.1).

The universal character of the SI system. Unfortunately, for historical reasons, there exist many different systems of units used by physicists. In what follows we want to help the reader to understand the relations between the different systems. Let us explain the following.

> *If one knows the physical dimension of some quantity in the SI system, then one can easily pass to every other system used in physics.*

In particular, we will discuss

- the natural SI system,
- the Planck system, and
- the energetic system.

The Planck system has the advantage that the fundamental physical constants $G, \hbar, c, \varepsilon_0, \mu_0, k$ do not appear explicitly in the basic equations (e.g., in elementary particle physics and cosmology). In this system, all the physical quantities are dimensionless.

The energetic system is mainly used in elementary particle physics. In this system, all of the physical quantities are measured in powers of energy, and the physical constants $\hbar, c, \varepsilon_0, \mu_0, k$ do not appear explicitly.

A.3 The Planck System

> All the systems of units which have hitherto been employed owe their origin to the coincidence of accidental circumstances, inasmuch as the choice of the units lying at the base of every system has been made, not according to general points of view, but essentially with reference to the special needs of our terrestrial civilization...
> In contrast with this it might be of interest to note that we have the means of establishing units which are independent of special bodies or substances. The means of determining the units of length, mass, and time are given by the action constant h, together with the magnitude of the velocity of propagation of light in a vacuum c, and that of the constant of gravitation G... These quantities must be found always the same, when measured by the most widely differing intelligences according to the most widely differing methods.
>
> Max Planck, 1906
> *The Theory of Heat Radiation*[8]

Fundamental constants. There exist the following universal constants in nature:

- G (gravitational constant),
- c (velocity of light in a vacuum),
- h (Planck's quantum of action),
- ε_0 (electric field constant of a vacuum),
- k (Boltzmann constant).

The explicit numerical values of these fundamental constants can be found in Table A.3 on page 949. We also use the constants

- $\hbar := h/2\pi$ (reduced Planck's quantum of action), and
- $\mu_0 := 1/\varepsilon_0 c^2$ (magnetic field constant of vacuum).

Basic laws in physics. These universal constants enter the following six basic laws of physics.

(i) Einstein's equivalence between rest mass m_0 and rest energy E of a particle: $E = m_0 c^2$.
(ii) Energy E of a photon with frequency ν: $E = h\nu$.
(iii) Gravitational force F between two masses M_1 and M_2 at distance r:

$$F = \frac{GM_1M_2}{r^2}.$$

[8] M. Planck, Theorie der Wärmestrahlung, Barth, Leipzig 1906. Reprinted by Dover Publications, 1991.

Table A.2. SI system

$1\ \mathrm{m} = 0.63 \cdot 10^{35}\mathbf{m}$	$1\ \mathbf{m} = l = 1.6 \cdot 10^{-35}\mathrm{m}$
$1\ \mathrm{s} = 0.19 \cdot 10^{44}\mathbf{s}$	$1\ \mathbf{s} = 5.3 \cdot 10^{-44}\mathrm{s}$
$1\ \mathrm{J} = 0.51 \cdot 10^{-9}\mathbf{J}$	$1\ \mathbf{J} = 1.97 \cdot 10^{9}\mathrm{J}$
$1\ \mathrm{kg} = 0.48 \cdot 10^{8}\mathbf{kg}$	$1\ \mathbf{kg} = 2.1 \cdot 10^{-8}\mathrm{kg}$
$1\ \mathrm{C} = 0.19 \cdot 10^{19}\mathbf{C}$	$1\ \mathbf{C} = 5.34 \cdot 10^{-19}\mathrm{C}$
$1\ \mathrm{K} = 0.71 \cdot 10^{-32}\mathbf{K}$	$1\ \mathbf{K} = 1.4 \cdot 10^{32}\mathrm{K}$
$1\ \mathrm{GeV} = 10^9\ \mathrm{eV} = 1.602 \cdot 10^{-10}\mathrm{J}$	
$1\ \mathrm{GeV}/c^2 = 1.78 \cdot 10^{-27}\mathrm{kg}$	

(iv) Electric force F between two electric charges Q_1 and Q_2 at distance r:

$$F = \frac{Q_1 Q_2}{4\pi\varepsilon_0 r^2}.$$

(v) Magnetic force F between two parallel electric currents of strength J_1 and J_2 in a wire of length L at distance r:

$$F = \frac{\mu_0 L J_1 J_2}{2\pi r}.$$

(vi) Mean energy E corresponding to one degree of freedom in a many-particle system at temperature T: $E = kT$.

In the SI system, the unit of electric current, called an ampere, is defined in such a way that the magnetic field constant of a vacuum is given by

$$\mu_0 = 4\pi \cdot 10^{-7}\,\frac{\mathrm{N}}{\mathrm{A}^2}.$$

By Table A.1 on prefixes, 1 MeV $=10^6$eV (mega electron volt).

Natural SI units. The five natural constants $G, c, \hbar, \varepsilon_0$, and k can be used to systematically replace the SI units m, s, J, C, K by the following so-called natural SI units:

- Planck length: $\mathbf{m} := l := \sqrt{\hbar G/c^3}$,
- Planck time: $\mathbf{s} := l/c$,
- Planck energy: $\mathbf{J} := \hbar c/l$,
- Planck charge: $\mathbf{C} := \sqrt{c\hbar\varepsilon_0}$,
- Planck temperature: $\mathbf{K} := \hbar c/kl$.

Parallel to $\mathrm{kg} = \mathrm{Js}^2/\mathrm{m}^2$, let us introduce the Planck mass

$$\mathbf{kg} := \mathbf{Js}^2/\mathbf{m}^2 = \hbar/cl.$$

The numerical values can be found in Table A.2. From (A.8) we obtain the representation

$$q = q_{\mathrm{Pl}} \cdot \mathbf{m}^\alpha \mathbf{s}^\beta \mathbf{J}^\gamma \mathbf{C}^\mu \mathbf{K}^\nu \tag{A.9}$$

of the physical quantity q in natural SI units. Hence

$$q = q_{\mathrm{Pl}} \cdot l^{\alpha} \left(\frac{l}{c}\right)^{\beta} \left(\frac{\hbar c}{l}\right)^{\gamma} (c\hbar\varepsilon_0)^{\mu/2} \left(\frac{\hbar c}{kl}\right)^{\nu}.$$

This implies

$$\boxed{q = q_{\mathrm{Pl}} \cdot l^A c^B \hbar^C \varepsilon_0^D k^E.} \tag{A.10}$$

Explicitly,

$$A = \alpha + \beta - \gamma - \nu, \quad B = \gamma + \nu - \beta + \mu/2, \quad C = \gamma + \nu + \mu/2,$$

and $D = \mu/2,\ E = -\nu$.

The Planck system of units. In this system, we set

$$\boxed{l = c = \hbar = \varepsilon_0 = k := 1.}$$

In particular, for the gravitational constant, this implies $G = 1$. By (A.10), $q = q_{\mathrm{Pl}}$.

The Planck system is characterized by the fact that all the physical quantities are dimensionless and their numerical values coincide with the numerical values in natural SI units.

In order to go back from the Planck system to the SI system, one has to replace each physical quantity q by

$$\boxed{q \quad \Rightarrow \quad \frac{q}{l^A c^B \hbar^C \varepsilon_0^D k^E}} \tag{A.11}$$

according to (A.10). The corresponding exponents $A, B, \ldots$ follow from (A.9) and (A.10). These exponents can be found in Table A.4 on page 951.

Example 1. For the proton, we get

$$E = 0.77 \cdot 10^{-19}\mathbf{J} = 1.5 \cdot 10^{-10}\mathrm{J} = 0.938\ \mathrm{GeV} \qquad (\text{rest energy})$$

along with

$$M = E/c^2 = 0.77 \cdot 10^{-19}\mathbf{kg} = 1.67 \cdot 10^{-27}\mathrm{kg} \qquad (\text{rest mass})$$

and

$$e = \sqrt{4\pi\alpha}\ \mathbf{C} = 0.30\mathbf{C} = 1.6 \cdot 10^{-19}\mathrm{C} \qquad (\text{electric charge}).$$

Therefore, $E_{\mathrm{Pl}} = M_{\mathrm{Pl}} = 0.77 \cdot 10^{-19}$, and $e_{\mathrm{Pl}} = 0.30$. In the Planck system, this implies

$$E = M = 0.77 \cdot 10^{-19} \quad \text{and} \quad e = 0.30.$$

Example 2. Consider the Einstein relation

$$E = m_0 c^2 \tag{A.12}$$

between the rest mass m_0 and the rest energy E of a free relativistic particle in the SI system. Letting $c := 1$, we obtain the corresponding equation

$$\boxed{E = m_0} \tag{A.13}$$

in the Planck system. Here, $E = E_{\mathrm{Pl}}$ and $m_0 = M_{\mathrm{Pl}}$. In order to go back from (A.13) to the SI system, one has to observe that

$$E = E_{\mathrm{Pl}} \cdot \mathbf{J}, \qquad m_0 = M_{\mathrm{Pl}} \cdot \mathbf{J}\mathbf{s}^2\mathbf{m}^{-2}$$

in natural SI units, by Table A.4 on page 951. Hence

$$E = E_{\mathrm{Pl}}\frac{\hbar c}{l}, \qquad m_0 = M_{\mathrm{Pl}}\frac{\hbar}{lc}.$$

Thus, we have to replace E and m_0 by

$$\frac{El}{\hbar c} \quad \text{and} \quad \frac{m_0 lc}{\hbar},$$

respectively. This way, we pass over from (A.13) to (A.12).

Example 3. In the SI system, the Maxwell equations in a vacuum are given by

$$\begin{aligned} \operatorname{div}\mathbf{D} &= \varrho, & \operatorname{div}\mathbf{B} &= 0, \\ \mathbf{curl}\,\mathbf{E} &= -\dot{\mathbf{B}}, & \mathbf{curl}\,\mathbf{H} &= \dot{\mathbf{D}} + \mathbf{j} \end{aligned} \tag{A.14}$$

along with $\mathbf{D} = \varepsilon_0\mathbf{E}$ and $\mathbf{B} = \mu_0\mathbf{H}$. Moreover, $c^2 = 1/\varepsilon_0\mu_0$. Alternatively,

$$\begin{aligned} \varepsilon_0 \operatorname{div}\mathbf{E} &= \varrho, & \operatorname{div}\mathbf{B} &= 0, \\ \mathbf{curl}\,\mathbf{E} &= -\dot{\mathbf{B}}, & c^2\,\mathbf{curl}\,\mathbf{B} &= \dot{\mathbf{E}} + \mu_0 c^2\mathbf{j}. \end{aligned} \tag{A.15}$$

Letting $\varepsilon_0 = \mu_0 = c := 1$, we obtain the corresponding Maxwell equations in the Planck system:

$$\begin{aligned} \operatorname{div}\mathbf{E} &= \varrho, & \operatorname{div}\mathbf{B} &= 0, \\ \mathbf{curl}\,\mathbf{E} &= -\dot{\mathbf{B}}, & \mathbf{curl}\,\mathbf{B} &= \dot{\mathbf{E}} + \mathbf{j}. \end{aligned} \tag{A.16}$$

In order to transform equation (A.16) back to the SI system, we replace the quantities $\mathbf{x}, t, \mathbf{E}, \mathbf{B}, \varrho, \mathbf{j}$ by

$$\frac{\mathbf{x}}{\mathbf{m}}, \quad \frac{t}{\mathbf{s}}, \quad \mathbf{E}\cdot\frac{\mathbf{mC}}{\mathbf{J}}, \quad \mathbf{B}\cdot\frac{\mathbf{m}^2}{\mathbf{sJ}}, \quad \varrho\cdot\frac{\mathbf{m}^3}{\mathbf{C}}, \quad \mathbf{j}\cdot\frac{\mathbf{m}^2}{\mathbf{Cs}}, \tag{A.17}$$

respectively, according to Table A.4 on page 951. In addition, the partial derivatives $\partial/\partial x^j$, $\partial/\partial t$ have to be replaced by

$$\mathbf{m}\cdot\frac{\partial}{\partial x^j}, \qquad \mathbf{s}\cdot\frac{\partial}{\partial t},$$

respectively. Finally, we set

$$\mathbf{m} := l, \quad \mathbf{s} := \frac{l}{c}, \quad \mathbf{C} := (c\hbar\varepsilon_0)^{1/2}, \quad \mathbf{J} := \frac{\hbar c}{l}.$$

This way, we get (A.14). In fact, for example, the first Maxwell equation $\operatorname{div}\ \mathbf{E} = \varrho$ from (A.16) means explicitly

$$\partial_j E^j = \varrho,$$

in Cartesian coordinates. Here, $\partial_j = \partial/\partial x^j$, and we sum over $j = 1, 2, 3$. By (A.17), this is transformed into the equation

$$\beta\cdot\partial_j E^j = \varrho,$$

where $\beta := \mathbf{C}^2/\mathbf{Jm}$. Since $\beta = c\hbar\varepsilon_0/c\hbar = \varepsilon_0$, we obtain $\varepsilon_0 \operatorname{div}\mathbf{E} = \varrho$. This is the first Maxwell equation from (A.15).

Example 4. In the SI system, the Schrödinger equation reads as

$$\mathrm{i}\hbar\frac{\partial\psi}{\partial t} = \frac{\hbar^2}{2m_0}\Delta\psi + U\psi. \tag{A.18}$$

Here, m_0 and U denote the mass of the particle and the potential energy, respectively. Letting $\hbar = 1$, we arrive at the Schrödinger equation

$$\boxed{\mathrm{i}\frac{\partial\psi}{\partial t} = \frac{\Delta\psi}{2m_0} + U\psi} \tag{A.19}$$

in the Planck system. In a Cartesian (x, y, z)-system, the Laplacian is defined by

$$\boxed{\Delta := -\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2}.}$$

Note that our sign convention coincides with the use of the Laplacian in modern differential geometry (Riemannian geometry) and string theory.[9]

In order to go back from the Planck system to the SI system[10], we replace the quantities $\mathbf{x}, t, U, m_0, \psi$ by

$$\frac{\mathbf{x}}{\mathbf{m}}, \quad \frac{t}{\mathbf{s}}, \quad \frac{U}{\mathbf{J}}, \quad m_0 \cdot \frac{\mathbf{m}^2}{\mathbf{J}\mathbf{s}^2}, \quad \psi \cdot \mathbf{m}^{3/2},$$

respectively. The Laplacian contains spatial derivatives of second order. Thus, the Laplacian Δ and the partial time derivative $\partial/\partial t$ have to be replaced by

$$\mathbf{m}^2 \cdot \Delta, \qquad \mathbf{s} \cdot \frac{\partial}{\partial t}.$$

Consequently, equation (A.19) is transformed into

$$\mathrm{i}\mathbf{J}\mathbf{s}\psi_t = \frac{(\mathbf{J}\mathbf{s})^{\mathbf{2}}}{2m_0}\Delta\psi + U\psi.$$

Since $\mathbf{J}\mathbf{s} = \hbar$, we get (A.18).

Example 5. Let us start with the Dirac equation

$$\boxed{\mathrm{i}\gamma^\mu\partial_\mu\psi = m_0\psi} \tag{A.20}$$

for the relativistic electron of rest mass m_0 formulated in the Planck system. Here, $\partial_\mu = \partial/\partial x^\mu$. Recall that we sum over μ from 0 to 3, by Einstein's summation convention. The definition of the Dirac–Pauli matrices $\gamma^0, \gamma^1, \gamma^2, \gamma^3$ can be found on page 789. In order to pass over to the SI system, we replace the quantities x^μ, m_0, ψ by

$$\frac{x^\mu}{\mathbf{m}}, \qquad m_0 \cdot \frac{\mathbf{m}^2}{\mathbf{J}\mathbf{s}^2}, \qquad \mathbf{m}^{3/2} \cdot \psi,$$

respectively, according to Table A.4 on page 951. Note that the dimension of the wave function ψ in the SI system is the same as in the case of the Schrödinger equation. Hence

$$\mathrm{i}\gamma^\mu\partial_\mu\psi = \sigma m_0\psi$$

[9] In classic textbooks, one has to replace Δ by $-\Delta$.

[10] The normalization condition $\int_{\mathbb{R}^3}\psi\psi^\dagger d^3x = 1$ implies that the wave function ψ has the dimension $\mathrm{m}^{-3/2}$ in the SI system.

where $\sigma := \mathbf{m}/\mathbf{s}^2\mathbf{J}$. Since $\sigma = c/\hbar$, in the SI system the Dirac equation reads as

$$\boxed{\mathrm{i}\hbar\gamma^\mu\partial_\mu\psi = m_0 c\psi.} \tag{A.21}$$

The quantity $\lambda_e := h/cm_0$ is called the Compton wave length of the electron.

Classical systems of units. In the context of the Maxwell equations, physicists frequently use the Gaussian system or the Heaviside system, for historical reasons. Let us explain the relation of these two systems to the SI system. The idea is to measure all of the physical quantities by meter, second, kilogram, and Kelvin. That is, we do not introduce a specific unit for electric charge. In the Gaussian system, the electric force F between two electric charges Q_1 and Q_2 at distance r (Coulomb law) is given by

$$F = \frac{Q_1 Q_2}{r^2}.$$

Moreover, we use the Gaussian definition of the magnetic field

$$\mathbf{H}_G := c\mathbf{B}.$$

This definition is motivated by the fact that the electric field $\mathbf{E}$ and the Gaussian magnetic field $\mathbf{H}_G$ possess the same physical dimension. In the Heaviside system, we use the Coulomb law

$$F = \frac{Q_1 Q_2}{4\pi r^2}.$$

In contrast to the Gaussian system from the 1830s, the Heaviside system from the 1880s has the advantage that the factor 4π does not appear in the Maxwell equations.

The Heaviside system of units. We use the SI system and set

$$\boxed{\varepsilon_0 := 1.}$$

In the SI system, each physical quantity q can be written as

$$q = q_{\mathrm{Pl}} \cdot l^A c^B \hbar^C \varepsilon_0^D k^E,$$

by (A.10) on page 938. Letting $\varepsilon_0 := 1$, we get

$$q = q_{\mathrm{Pl}} \cdot l^A c^B \hbar^C k^E,$$

in the Heaviside system. Consequently,

$$q = q_{\mathrm{H}} \cdot \mathrm{m}^a \mathrm{s}^b \mathrm{kg}^c \mathrm{K}^d.$$

That is, each physical quantity can be described by powers of meter, second, kilogram, and Kelvin. In the Heaviside system, the Maxwell equations read as follows:

$$\begin{aligned} \operatorname{div}\mathbf{E} &= \varrho, & \operatorname{div}\mathbf{H}_G &= 0, \\ \mathbf{curl}\,\mathbf{E} &= -\frac{1}{c}\frac{\partial \mathbf{H}_G}{\partial t}, & \mathbf{curl}\,\mathbf{H}_G &= \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \frac{\mathbf{j}}{c}. \end{aligned} \tag{A.22}$$

To obtain this, start with the Maxwell equations in the SI system. By (A.15) along with $c^2 = 1/\varepsilon_0\mu_0$,

$$\begin{aligned} \varepsilon_0 \operatorname{div}\mathbf{E} &= \varrho, & \operatorname{div}(c\mathbf{B}) &= 0, \\ \mathbf{curl}\,\mathbf{E} &= -\frac{1}{c}\frac{\partial (c\mathbf{B})}{\partial t}, & \mathbf{curl}(c\mathbf{B}) &= \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \frac{\mathbf{j}}{\varepsilon_0 c}. \end{aligned}$$

Letting $\varepsilon_0 := 1$ and $\mathbf{H}_G := c\mathbf{B}$, we get (A.22).

The Gaussian system of units. Using the rescaling

$$\mathbf{E} \Rightarrow \frac{\mathbf{E}}{4\pi}, \qquad \mathbf{H}_G \Rightarrow \frac{\mathbf{H}_G}{4\pi},$$

the Heaviside system passes over to the Gaussian system. In particular, the Maxwell equations in the Gaussian system read as follows:

$$\begin{aligned} \operatorname{div} \mathbf{E} &= 4\pi\varrho, & \operatorname{div} \mathbf{H}_G &= 0, \\ \mathbf{curl}\, \mathbf{E} &= -\frac{1}{c}\frac{\partial \mathbf{H}_G}{\partial t}, & \mathbf{curl}\, \mathbf{H}_G &= \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi \mathbf{j}}{c}. \end{aligned} \tag{A.23}$$

Observe that the variants (A.22) and (A.23) of the Maxwell equations differ by the factor 4π. The Gaussian system is used in the 10-volume standard textbook on theoretical physics by Landau and Lifshitz (1982).

A.4 The Energetic System

The most important physical quantity in elementary particle physics is given by the energy of a particle accelerator. Therefore, particle physicists like to use energy as basic unit. Let us discuss this. In the SI system, an arbitrary physical quantity can be written as

$$q = q_{\mathrm{Pl}} \cdot l^A c^B \hbar^C \varepsilon_0^D k^E,$$

by (A.10). In the energetic system, we set[11]

$$\boxed{c = \hbar = \varepsilon_0 = k := 1.}$$

Hence

$$q = q_{\mathrm{Pl}} \cdot l^A.$$

Consequently, each physical quantity has the physical dimension of some power of length. In particular, for energy E we get

$$E = E_{\mathrm{Pl}} \cdot l^{-1} \hbar c,$$

in the SI system. Hence

$$E = E_{\mathrm{Pl}} \cdot l^{-1},$$

in the energetic system. That is, energy has the physical dimension of inverse length.

> *Conversely, length has the physical dimension of inverse energy in the energetic system of units.*

In terms of natural SI units, each physical quantity can be written as

$$q = q_{\mathrm{Pl}} \cdot \mathbf{m}^\alpha \mathbf{s}^\beta \mathbf{J}^\gamma \mathbf{C}^\mu \mathbf{K}^\nu.$$

It follows from $c = \hbar = \varepsilon_0 = k := 1$ that

[11] In particular, this implies $\mu_0 = 1$.

$$\boxed{\mathbf{s} = \mathbf{m} = \mathbf{J}^{-1}, \qquad \mathbf{K} = \mathbf{J}, \qquad \mathbf{C} = 1.} \tag{A.24}$$

This implies

$$q = q_{\mathrm{Pl}} \cdot \mathbf{J}^A,$$

where $A = -\alpha - \beta + \gamma + \nu$. This way, each physical quantity can be expressed by powers of the Planck energy $\mathbf{J}$. Using Table A.4 on page 951 along with (A.24), we immediately obtain all the dimensions of important physical quantities in the energetic system. For example, velocity has the dimension

$$v = v_{\mathrm{SI}} \cdot \mathrm{ms}^{-1},$$

in the SI system. Thus, in natural SI units,

$$v = v_{\mathrm{Pl}} \cdot \mathbf{ms}^{-1}.$$

In the energetic system $\mathbf{m} = \mathbf{s}$, by (A.24). Hence

$$v = v_{\mathrm{Pl}},$$

that is, velocity is dimensionless. Note that this follows more simply from the fact that $c := 1$ in the energetic system; that is, the velocity of light is dimensionless. Similarly, using the dimensionless quantities $\hbar = \varepsilon_0 = \mu_0 = k := 1$ along with the basic physical laws (i)-(vi) on page 936, we encounter the following physical dimensions in the energetic system:

- [mass] = [momentum] = [temperature] = [energy],
- [length] = [time] = [energy]$^{-1}$,
- [cross section] = [area] = [length]^2 = [energy]$^{-2}$,
- [electric charge] = [velocity] = [action] = dimensionless,
- [force] = [electric field] = [magnetic field] = [energy]^2,
- [potential] = [vector potential] = [energy],
- the coupling constants of quantum electrodynamics, quantum chromodynamics, and electroweak interaction are dimensionless.

Since the electric charge and the coupling constants of the Standard Model in particle physics are dimensionless in the energetic system, these quantities are independent of the rescaling of energy.

Examples. The Einstein relation $E = m_0 c^2$ reads as

$$E = m_0$$

in the energetic system, since $c := 1$.

> *The Maxwell equations* (A.16)*, the Schrödinger equation* (A.19)*, and the Dirac equation* (A.20) *coincide in the Planck system and in the energetic system.*

In elementary particle physics, physicists like to use GeV (giga electron volt), where

$$\mathbf{J} = 1.98 \cdot 10^{19}\ \mathrm{GeV}.$$

Consequently, from Table A.2 on page 937, we obtain the following conversion formulas between the SI system and the energetic system:

$$
\begin{aligned}
1\,\mathrm{m} &\doteq 5 \cdot 10^{16}\,(\mathrm{GeV})^{-1},\\
1\,\mathrm{s} &\doteq 1.5 \cdot 10^{24}\,(\mathrm{GeV})^{-1},\\
1\,\mathrm{J} &\doteq 6.3 \cdot 10^{9}\,\mathrm{GeV},\\
1\,\mathrm{K} &\doteq 1.4 \cdot 10^{-13}\,\mathrm{GeV},\\
1\,\mathrm{C} &\doteq 1.9 \cdot 10^{18}.
\end{aligned}
$$

Depending on the energy scale, physicists also use mega electron volt, MeV. Here, $1\ \mathrm{GeV} = 10^3\ \mathrm{MeV}$.

The physical dimension of cross sections. Observe that

$$\hbar c = 1.97327 \cdot 10^{-13}\ \mathrm{MeV} \cdot \mathrm{m}.$$

This implies

$$\mathrm{m}^2 = \frac{(\hbar c)^2}{(1.97327)^2} \cdot 10^{26} (\mathrm{MeV})^{-2}.$$

In the SI system, the cross section σ is measured in m^2. Setting

$$\hbar = c := 1,$$

we get the cross section in the energetic system measured in $(\mathrm{MeV})^{-2}$. Conversely, the passage from the energetic system to the SI system can be easily obtained by using the replacement

$$\boxed{\sigma \quad \Rightarrow \quad \frac{\sigma}{(\hbar c)^2}.}$$

In fact, if $\sigma = a$ in the energetic system, then $\sigma = (\hbar c)^2 a$ in the SI system.

A.5 The Beauty of Dimensional Analysis

Physicists use the dimensionality of physical quantities in order to get important information. Let us illustrate this by considering three examples: the pendulum, Newton's gravitational law, and Kolmogorov's law for turbulence.

The pendulum. Consider a pendulum of length l and mass m. We are looking for a formula for the period of oscillation, T, of the pendulum. We expect that T depends on l, m, and the gravitational acceleration g. Thus, we begin with the ansatz

$$T = C \cdot l^{\alpha} m^{\beta} g^{\gamma}$$

where C is a dimensionless constant. Passing to dimensions we get

$$\mathrm{s} = \mathrm{m}^{\alpha} \mathrm{kg}^{\beta} \mathrm{m}^{\gamma} \mathrm{s}^{-2\gamma}.$$

This implies $\beta = 0$, $\gamma = -\frac{1}{2}$, and $\alpha = -\gamma = \frac{1}{2}$, that is,

$$\boxed{T = C\sqrt{\frac{l}{g}}.} \tag{A.25}$$

The constant C has to be determined from experiment. The explicit solution of the problem via elliptic integrals shows that, for small pendulum motions, equation (A.25) is valid with $C = 2\pi$.

Newton's gravitational law. In 1619 Kepler discovered empirically that the motion of a planet satisfies the law

$$\frac{T^2}{a^3} = \text{const}$$

where T is the period of revolution, and a is the great semi-axis of the elliptic orbit. In order to guess Newton's gravitational law from this information, let us make the ansatz

$$m\ddot{\mathbf{x}} = C|\mathbf{x}|^{\mu}\mathbf{x}$$

for the motion $\mathbf{x} = \mathbf{x}(t)$ of the planet. Here, m is the mass of the planet, and C is a constant. We want to show that $\mu = -3$ is the only natural choice. To this end, consider the rescaled motion $\mathbf{y}(t) := \alpha\mathbf{x}(\beta t)$. Then

$$m\ddot{\mathbf{y}} = \beta^2\alpha^{-\mu}C|\mathbf{y}|^{\mu}\mathbf{y}.$$

We postulate that the equation of motion and the third Kepler law are independent of the rescaling. This means that $\beta^2\alpha^{-\mu} = 1$ and

$$(T\beta)^2/(\alpha a)^3 = T^2/a^3.$$

Hence $\mu = -3$. Summarizing, we obtain Newton's gravitational law

$$\boxed{m\ddot{\mathbf{x}} = \frac{C}{|\mathbf{x}|^2} \cdot \frac{\mathbf{x}}{|\mathbf{x}|}.}$$

The Kolmogorov law for energy dissipation in turbulent flows. It is a typical property of turbulent flow that there exist eddies of different diameters λ, where $\lambda_{\min} \leq \lambda \leq \lambda_{\max}$. One may think, for example, of clouds in the air or of nebulas in astronomy. One finds that the large eddies tend to break down into smaller eddies. This way, energy from large eddies flows to smaller eddies. Here, physicists assume that the energy of the smallest eddies with $\lambda = \lambda_{\min}$ is transformed into heat by friction (energy dissipation). Viscosity is of significance only for small eddies. We define

$$\varepsilon := \frac{\text{loss of energy by dissipation}}{\text{mass} \cdot \text{time}}.$$

This is the crucial physical quantity. Note that ε can be measured in experiments; it is equal to the produced heat. Using the method of dimensional analysis, Kolmogorov obtained the law

$$\varepsilon = \int_{\lambda_{\min}}^{\lambda_{\max}} s(\lambda)d\lambda$$

along with the spectral function

$$\boxed{s(\lambda) := C\left(\frac{\lambda}{\lambda_{\min}}\right)\frac{\eta\varepsilon^{2/3}}{\varrho} \cdot \frac{1}{\lambda^{7/3}}.}$$

Here, η and ϱ are viscosity and mass density, respectively. The function C is dimensionless. For values λ near $\lambda_{\min}$, the function C can be approximated by a constant. Therefore, physicists speak of Kolmogorov's 7/3-law. The proof can be found in Zeidler (1986), Vol. IV, p. 514.

It turns out that dimensional analysis represents a magic wand of physicists. In this setting, a minimum of hypotheses provides us a maximum of information.

A.6 The Similarity Principle in Physics

Rescaled SI units. Let us replace the SI units m, s, J, C, K with the rescaled units

$$\mathrm{m}_*, \mathrm{s}_*, \mathrm{J}_*, \mathrm{C}_*, \mathrm{K}_*,$$

where $\mathrm{m}_* = m_+ \cdot \mathrm{m}$, $\mathrm{s}_* = s_+ \cdot \mathrm{s}, \ldots$ with the real numbers $m_+, s_+, \ldots$ Then, each physical quantity q can be represented as

$$\boxed{q = q_* \cdot \mathrm{m}_*^{\alpha} \mathrm{s}_*^{\beta} \mathrm{J}_*^{\gamma} \mathrm{C}_*^{\mu} \mathrm{K}_*^{\nu} = q_* \cdot [q].}$$

The real number q_* is called the numerical value of q, and $[q]$ is called the dimension of q with respect to this system of units. In practice, one chooses $\mathrm{m}_*, \mathrm{s}_*$, ... in such a way that the numerical values of the physical quantities are neither too large nor too small. For example, if we want to study thin layers, then it is convenient to use $\mathrm{m}_* := 10^{-9}\mathrm{m} = 1\mathrm{nm}$ (nanometer). In astronomy, one uses light years for measuring distances, and so on.

The role of small quantities in physics. It is impossible to speak of a small length L in physics. In fact, if

$$L = 1 \text{ meter},$$

then passing to a new length scale, we get

$$L = 10^{15} \text{ femtometer}.$$

Therefore, it makes sense to speak about smallness only for dimensionless quantities. For example, choose the radius r_E of earth and the radius r_p of a proton. Then the dimensionless ratio

$$\frac{r_p}{r_E} = 6 \cdot 10^{-21}$$

is a small quantity compared with 1.

The experience of physicists shows that two different theories are good approximations of each other if suitable dimensionless quantities are small. Let us consider two crucial examples.

(i) Relativistic physics: Let v and c be the velocity of some particle and the velocity of light, respectively. If the dimensionless quotient

$$\frac{v}{c}$$

is sufficiently small, then the relativistic motion of the particle can be described approximately by Newton's classical mechanics. For example, the relativistic mass

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}} = m_0 \left(1 - \frac{v^2}{2c^2} + o\left(\frac{v^2}{c^2}\right)\right), \qquad \frac{v^2}{c^2} \to 0$$

is approximately equal to the rest mass m_0 if the quotient v/c is sufficiently small.

(ii) Quantum mechanics: Let $S = E(t_1 - t_2)$ be the action for the motion of some particle with constant energy E during a fixed reasonable time interval $[t_1, t_2]$, say, one hour. If the dimensionless ratio

$$\frac{S}{\hbar}$$

is small, then the quantum motion of the particle can be approximately described by Newton's classical mechanics.

In (i) and (ii), corrections to classical mechanics can be obtained by perturbation theory if v/c and $S/\hbar$ are small. These are the post-Newtonian approximation and the WKB approximation, respectively.

The fundamental similarity principle in physics. We postulate that

Physical processes are described by equations which are invariant under rescaling of units. Explicitly, we demand that the laws of physics can be written in such a way that, in a fixed system of units, they only depend on the dimensionless quotients

$$\frac{q}{[q]}, \qquad \frac{r}{[r]}, \qquad \dots$$

of all the physical quantities q, r, ...

A special role is played by those physical quantities which are dimensionless in the SI system. We expect that such quantities are related to important physical effects. The experience of physicists confirms this. For example, the so-called fine structure constant

$$\alpha := \frac{e^2}{4\pi\varepsilon_0\hbar c} = \frac{1}{137.04}$$

represents the most important dimensionless quantity that can be constructed from the universal constants. This constant measures the strength of the interaction between electrons, positrons, and photons in quantum electrodynamics. The smallness of α is responsible for the fact that perturbation theory can be successfully applied to quantum electrodynamics.

Example. Consider the Einstein relation

$$E = m_0 c^2$$

between rest mass m_0 and rest energy E of a particle. In any rescaled SI system,

$$E = E_* \cdot \mathrm{J}_*, \qquad m_0 = (m_0)_* \cdot \mathrm{J}_* \mathrm{s}_*^2 \mathrm{m}_*^{-2}, \qquad c = c_* \cdot \mathrm{m}_* \mathrm{s}_*^{-1}.$$

Hence $E_* = (m_0)_* c_*^2$. Moreover, $[E] = J_*$, and

$$[m_0][c]^2 = \mathrm{J}_* \mathrm{s}_*^2 \mathrm{m}_*^{-2} \cdot \mathrm{m}_*^2 \mathrm{s}_*^{-2} = \mathrm{J}_*.$$

This means that

$$\frac{E}{[E]} = \frac{m_0 c^2}{[m_0][c]^2}.$$

Physicists frequently use such dimension tests in order to check the correctness of formulas.

Counterexample. Let x and t denote position and time, respectively. The equation

$$x = \sin t$$

is not allowed in the SI system, since it is not invariant under the rescaling $x \Rightarrow \alpha x$ and $t \Rightarrow \beta t$ for nonzero constants α and β. In contrast to this, the equation

$$\frac{x}{x_0} = \sin\left(\frac{t}{t_0}\right)$$

is admissible in any system of units if x and x_0 as well as t and t_0 possess the same dimensions.

Application to Reynolds numbers in turbulence. The motion of a viscous fluid in a 3-dimensional bounded domain $\mathcal{G}$ is governed by the so-called Navier–Stokes equations[12]

$$\begin{aligned} \varrho \mathbf{v}_t - \nu \Delta_{\mathbf{x}} \mathbf{v} + \varrho(\mathbf{v}\nabla_{\mathbf{x}})\,\mathbf{v} &= \mathbf{f} - \nabla_{\mathbf{x}} p \quad \text{on } \mathcal{G}, \\ \nabla_{\mathbf{x}} \mathbf{v} &= 0 \quad \text{on } \mathcal{G}, \\ \mathbf{v} &= 0 \quad \text{on } \partial\mathcal{G}. \end{aligned}$$

The symbols possess the following physical meaning: $\mathbf{v}$ velocity vector, ϱ mass density, $\mathbf{f}$ force density vector, p pressure, ν viscosity constant, $\mathbf{x}$ position vector, and t time. Set

$$\mathbf{x} = \mathbf{X} \cdot \mathrm{m}_*, \quad t = T \cdot \mathrm{s}_*, \quad \mathbf{v} = \mathbf{u} \cdot \mathrm{m}_* \mathrm{s}_*^{-1}, \quad \varrho = \Omega \cdot \mathrm{J}_* \mathrm{s}_*^2 \mathrm{m}_*^{-5}$$

and

$$\mathbf{f} = \mathbf{F} \cdot \mathrm{J}_* \mathrm{m}_*^{-4}, \quad p = P \cdot \mathrm{J}_* \mathrm{m}_*^{-3}, \quad \nu = N \cdot \mathrm{J}_* \mathrm{m}_*^{-3},$$

where the coefficients $\mathbf{X}, T, \ldots$ are dimensionless. Furthermore, let d and v denote the diameter of the domain $\mathcal{G}$ and a typical velocity of the fluid, respectively. Naturally enough, we choose

$$\mathrm{m}_* := d, \quad \mathrm{s}_* := dv^{-1}, \quad \mathrm{J}_* := \varrho v^2 d^{-3}.$$

This way, we obtain the rescaled dimensionless Navier–Stokes equations

$$\begin{aligned} \mathbf{u}_t - \mathrm{Re}^{-1} \Delta_{\mathbf{X}} \mathbf{u} + (\mathbf{u}\nabla_{\mathbf{X}})\,\mathbf{u} &= \mathbf{F} - \nabla_{\mathbf{X}} P \quad \text{on } \mathcal{H}, \\ \nabla_{\mathbf{X}} \mathbf{u} &= 0 \quad \text{on } \mathcal{H}, \\ \mathbf{u} &= 0 \quad \text{on } \partial\mathcal{H} \end{aligned}$$

with the dimensionless Reynolds number

$$\mathrm{Re} := \frac{\varrho v d}{\nu}.$$

The rescaled domain $\mathcal{H}$ is obtained from the original domain $\mathcal{G}$ by replacing the points $\mathbf{x}$ of $\mathcal{G}$ by $d^{-1}\mathbf{x}$. Physical experiments show that if the Reynolds number Re is sufficiently large, then turbulence occurs.

The rescaled dimensionless Navier–Stokes equations reflect an important similarity principle in hydrodynamics. Explicitly, if two physical situations in different regions are governed by the same rescaled dimensionless Navier–Stokes equations, then the physics is the same up to suitable similarity transformations.

Discovery of errors in physical computations. Physicists use physical dimensions in order to detect errors in their computations. To explain this with a simple example, suppose that we arrive at the equation

$$p = c^3 m_0 \tag{A.26}$$

after finishing some computation. Here, we use the following notation: p momentum, m_0 particle mass, c velocity of light. We want to check this. In the SI system, we have the following dimensions:

$$[p] = \mathrm{kg} \cdot \mathrm{ms}^{-1}, \quad [m_0] = \mathrm{kg}, \quad [c] = \mathrm{ms}^{-1}.$$

Hence $[p] = [c] \cdot [m_0]$. It follows from (A.26) that $[p] = [c]^3 [m_0]$. This implies $[c]^2 = 1$, which is a contradiction. Consequently, our result (A.26) is wrong. The same argument can be used in the energetic system. However, we now have $[c] = 1$, which does not lead to any contradiction. In other words, the energetic system of units is too weak in order to detect that equation (A.26) is wrong, by checking physical dimensions.

[12] Navier (1785–1836), Stokes (1819–1903).

Table A.3. Fundamental constants in nature

fundamental constant	SI units	natural SI units
velocity of light in a vacuum	$c = 2.998 \cdot 10^8$ m/s	$c = \mathbf{m/s}$
Planck's action quantum	$h = 6.626 \cdot 10^{-34}$ Js $\hbar = h/2\pi$	$\hbar = \mathbf{Js}$
gravitational constant	$G = 6.673 \cdot 10^{-11}$ m^5/Js^4	$G = \mathbf{m}^5/\mathbf{Js}^4$ $= l^2c^3/\hbar$
electric field constant	$\varepsilon_0 = 8.854 \cdot 10^{-12}$ C^2/Jm	$\varepsilon_0 = \mathbf{C}^2/\mathbf{Jm}$
magnetic field constant	$\mu_0 = 1/\varepsilon_0 c^2$ $= 4\pi \cdot 10^{-7}$ Js^2/C^{2}m	$\mu_0 = \mathbf{Js}^2/\mathbf{C}^2\mathbf{m}$
Boltzmann constant	$k = 1.380 \cdot 10^{-23}$ J/K	$k = \mathbf{J/K}$
fine structure constant	$\alpha = e^2/4\pi c\hbar\varepsilon_0$ (dimensionless)	$\alpha = 1/137.04$
charge of the proton	$e = 1.602 \cdot 10^{-19}$ C	$e = \sqrt{4\pi\alpha}\,\mathbf{C}$ $= 0.30\,\mathbf{C}$
rest energy of the proton	$E_p = 1.5 \cdot 10^{-10}$ J $= 0.938$ GeV (giga electron volt)	$E_p = 0.77 \cdot 10^{-19}\,\mathbf{J}$
rest mass of the proton	$m_p = 1.672 \cdot 10^{-27}$ kg $= 0.938$ GeV/c^2	$m_p = 0.77 \cdot 10^{-19}\,\mathbf{kg}$
Compton wave length of the proton $\lambda_p = h/m_p c$	$\lambda_p = 1.32 \cdot 10^{-15}$ m $= 1.32$ fm (femtometer)	$\lambda_p = 0.83 \cdot 10^{20}\,\mathbf{m}$
rest energy of the electron	$E_e = 8.16 \cdot 10^{-14}$ J $= 0.511$ MeV (mega electron volt)	$E_e = E_p/1838.1$

Table A.3. (continued)

rest mass of the electron	$m_e = 0.91 \cdot 10^{-30}$ kg	$m_e = m_p/1838.1$
Compton wave length of the electron $\lambda_e = \hbar/m_p c$	$\lambda_e = 2.43 \cdot 10^{-12}$ m	$\lambda_e = 1838.1\,\lambda_p$
radius of the proton	$r_p = 1.3 \cdot 10^{-15}$ m $= 1.3$ fm (femtometer)	$r_p = 0.882 \cdot 10^{20}\,l$
fundamental constant	SI units	natural SI units
Bohr radius of the hydrogen atom	$r_B = 0.529 \cdot 10^{-10}$ m $= 5.29$ nm (nanometer)	$r_B = 40000\,r_p$
Bohr magneton	$\mu_B = -e\hbar/2m_e$ $= -9.27\,10^{-24}$ mC/s	$\mu_B = -\,\mathbf{mC/s}$
magnetic moment of the electron	$\mu_e = \left(1 + \frac{\alpha}{2\pi} - ...\right)\mu_B$	$\mu_e = 1.001\mu_B$
nuclear magneton	$\mu_n = e\hbar/2m_p$ $= 5.05\,10^{-27}$ mC/s	$\mu_B = 1836.1\mu_n$
magnetic moment of the proton	$\mu_p = 2.79\mu_n$	$\mu_p = 2.79\mu_n$

More precise values can be found in CODATA Bull. **63** (1986), and E. Cohen and B. Taylor, Review of Modern Physics **59**(4) (1986). A list of high-precision values can also be found in the Appendix to Zeidler, Oxford User's Guide to Mathematics, Oxford University Press, 2004. In the following Table A.4, observe that the two quantities $\mathbf{E}$ and $c\mathbf{B}$, possess the same physical dimension in the SI system. The same is true for $c\mathbf{D}$ and $\mathbf{H}$. Here, we use the notation:

- $\mathbf{E}$ electric field vector,
- $\mathbf{B}$ magnetic field vector,
- $\mathbf{D}$ electric field intensity vector,
- $\mathbf{H}$ magnetic field intensity vector.

In the literature, the terminology with respect to $\mathbf{E}$, $\mathbf{B}$, $\mathbf{D}$, $\mathbf{H}$ is not uniform, for historical reasons. Since $\mathbf{E}$ and $\mathbf{B}$ generate the electromagnetic field tensor (see (14.50) on page 792), it follows from Einstein's theory of special relativity that the vector fields $\mathbf{E}$ and $\mathbf{B}$ (resp. $\mathbf{D}$ and $\mathbf{H}$) form a unit. The mean magnetic field of earth has the strength $\mathbf{B}_{\text{earth}} = 0.5$ Gauss $= 0.5 \cdot 10^{-4}$ Tesla.

Table A.4. Units of physical quantities

Physical quantity	SI units $\mathrm{m}^\alpha \mathrm{s}^\beta \mathrm{J}^\gamma \mathrm{C}^\mu \mathrm{K}^\nu$	natural SI units $\mathbf{m}^\alpha \mathbf{s}^\beta \mathbf{J}^\gamma \mathbf{C}^\mu \mathbf{K}^\nu = l^A c^B \hbar^C \varepsilon_0^D k^E$
length	m (meter)	$\mathbf{m} = l$ (Planck length)
time	s (second)	$\mathbf{s} = l/c$ (Planck time)
energy, work	J (Joule)	$\mathbf{J} = \hbar c/l$ (Planck energy)
electric charge	C (Coulomb)	$\mathbf{C} = (c\hbar\varepsilon_0)^{\frac{1}{2}}$ (Planck charge)
temperature	K (Kelvin)	$\mathbf{K} = \hbar c/lk$ (Planck temperature)
mass	$\mathrm{kg} = \mathrm{Js}^2/\mathrm{m}^2$ (kilogram)	$\mathbf{Js}^2/\mathbf{m}^2 = \hbar/cl$ (Planck mass)
electric current strength	$\mathrm{A} = \mathrm{C/s}$ (ampere)	$\mathbf{C/s} = c^{\frac{3}{2}}(\hbar\varepsilon_0)^{\frac{1}{2}}/l$
voltage	$\mathrm{V} = \mathrm{J/C}$ (volt)	$\mathbf{J/C} = (c\hbar)^{\frac{1}{2}}/l\varepsilon_0^{\frac{1}{2}}$
action (energy × time)	Js	$\mathbf{Js} = \hbar$
momentum (mass × velocity)	Js/m	$\mathbf{Js/m} = \hbar/l$
power (energy/time)	$\mathrm{W} = \mathrm{J/s}$ (Watt)	$\mathbf{J/s}$
force (energy/length)	$\mathrm{N} = \mathrm{J/m}$ (Newton)	$\mathbf{J/m}$
frequency ν (number of oscillations/time)	1/s	$\mathbf{1/s}$
angular frequency $\omega = 2\pi\nu$	1/s	$\mathbf{1/s}$
pressure (force/area)	$\mathrm{Pa} = \mathrm{N/m}^2 = \mathrm{J/m}^3$	$\mathbf{J/m}^3$
area, cross section	m^2	$\mathbf{m}^2 = l^2$
volume	m^3	$\mathbf{m}^3 = l^3$

Table A.4. (continued)

Physical quantity	SI units $\mathrm{m}^\alpha\mathrm{s}^\beta\mathrm{J}^\gamma\mathrm{C}^\mu\mathrm{K}^\nu$	natural SI units $\mathbf{m}^\alpha\mathbf{s}^\beta\mathbf{J}^\gamma\mathbf{C}^\mu\mathbf{K}^\nu = l^A c^B \hbar^C \varepsilon_0^D k^E$
velocity	$\mathrm{m/s}$	$\mathbf{m/s} = c$
acceleration	$\mathrm{m/s}^2$	$\mathbf{m/s}^2 = c^2/l$
mass density	$\mathrm{kg/m}^3 = \mathrm{Js}^2/\mathrm{m}^5$	$\mathbf{Js}^2/\mathbf{m}^5$
electric charge density ρ (charge/volume)	$\mathrm{C/m}^3$	$\mathbf{C/m}^3$
electric current density vector $\mathbf{j} = \rho\mathbf{v}$	$\mathrm{C/m}^2\mathrm{s}$	$\mathbf{C/m}^2\mathbf{s}$
electric field vector $\mathbf{E}$ (force/charge)	$\mathrm{N/C} = \mathrm{V/m} = \mathrm{J/mC}$	$\mathbf{J/mC}$
magnetic field vector $\mathbf{B}$	$\mathrm{T} = \mathrm{Vs/m}^2 = \mathrm{Js/m}^2\mathrm{C}$ (Tesla)	$\mathbf{Js/m}^2\mathbf{C}$
magnetic flow $\int \mathbf{B}\, d\mathbf{f}$	$\mathrm{Wb} = \mathrm{Vs} = \mathrm{Js/C}$ (Weber)	$\mathbf{Js/C}$
electric field intensity vector $\mathbf{D}$	$\mathrm{C/m}^2$	$\mathbf{C/m}^2$
magnetic field intensity vector $\mathbf{H}$	$\mathrm{A/m} = \mathrm{C/sm}$	$\mathbf{C/sm}$
electric dipole moment	Cm	$\mathbf{Cm}$
magnetic dipole moment	$\mathrm{Am}^2 = \mathrm{m}^2\mathrm{C/s}$	$\mathbf{m}^2\mathbf{C/s}$
polarization $\mathbf{P}$ (electric dipole moment density), $\mathbf{D} = \varepsilon_0\mathbf{E} + \mathbf{P}$	$\mathrm{C/m}^2$	$\mathbf{C/m}^2$
magnetization $\mathbf{M}$ (magnetic dipole moment density), $\mathbf{B} = \mu_0\mathbf{H} + \mathbf{M}$	$\mathrm{C/ms}$	$\mathbf{C/ms}$
scalar potential U ($\mathbf{E} = -\,\mathbf{grad}\, U - \mathbf{A_t}$)	$\mathrm{V} = \mathrm{J/C}$	$\mathbf{J/C}$
vector potential $\mathbf{A}$ ($\mathbf{B} = \mathbf{curl}\, \mathbf{A}$)	$\mathrm{Vs/m} = \mathrm{Js/mC}$	$\mathbf{Js/mC}$
4-potential A^μ ($A^0 = U/c$, $\mathbf{A} = A^j\mathbf{e}_j$)	$\mathrm{Js/mC}$	$\mathbf{Js/mC}$

Table A.4. (continued)

Physical quantity	SI units $\mathrm{m}^{\alpha}\mathrm{s}^{\beta}\mathrm{J}^{\gamma}\mathrm{C}^{\mu}\mathrm{K}^{\nu}$	natural SI units $\mathbf{m}^{\alpha}\mathbf{s}^{\beta}\mathbf{J}^{\gamma}\mathbf{C}^{\mu}\mathbf{K}^{\nu} = l^{A}c^{B}\hbar^{C}\varepsilon_0^{D}k^{E}$
electromagnetic field tensor $F_{\mu\nu}$ $(F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu)$	$\mathrm{Vs/m}^2 = \mathrm{Js/m}^2\mathrm{C}$	$\mathbf{Js/m}^2\mathbf{C}$
electric 4-current j^μ $(j^0 = c\rho,\, \mathbf{j} = j^k\mathbf{e}_k)$	$\mathrm{C/m}^2\mathrm{s}$	$\mathbf{C/m}^2\mathbf{s}$
Schrödinger function ψ in N space dimensions (solution of the Schrödinger equation)	$\mathrm{m}^{-\frac{N}{2}}$	$\mathbf{m}^{-\frac{N}{2}} = l^{-\frac{N}{2}}$
Dirac function ψ (solution of the Dirac equation, electron field, quark field, fermion fields)	$1/\mathrm{ms}^{\frac{1}{2}}$	$\mathbf{1/ms}^{\frac{1}{2}} = c^{\frac{1}{2}}/l^{\frac{3}{2}}$
Lagrangian L in classical mechanics $\text{action} = \int_{t_0}^{t_1} L(q,\dot{q},t)\,dt$	J	$\mathbf{J} = \hbar c/l$
Lagrangian density $\mathcal{L}$ in relativistic field theory $\text{action} = \int_{\mathbb{R}^4} \mathcal{L}(\psi,\partial\psi,x)\,d^4x$	$\mathrm{Js/m}^4$	$\mathbf{Js/m}^4 = \hbar/l^4$
Hamiltonian H in classical mechanics, $H = p\dot{q} - L$	J	$\mathbf{J} = \hbar c/l$
Hamiltonian density $\mathcal{H} = \pi\dot{\psi} - \mathcal{L}$	$\mathrm{Js/m}^4$	$\mathbf{Js/m}^4 = \hbar/l^4$
4-potential B_μ of the gluon field in QCD, $(\mathrm{i}B_\mu \in SU(3))$	Js/m	$\mathbf{Js/m} = \hbar/l$
field tensor $G_{\mu\nu}$ of the gluon field $\big(G_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu + \mathrm{i}g_s[B_\mu, B_\nu]\big)$	$\mathrm{Js/m}^2$	$\mathbf{Js/m}^2 = \hbar/l^2$

Epilogue

Mathematics is the gate and the key to the sciences.
Roger Bacon (1214–1294)

I love mathematics not only because it is applicable to technology but also because it is beautiful.
Rósza Péter (1905–1977)

The perfection of mathematical beauty is such whatsoever is most beautiful is also found to be most useful and excellent.
D'Arcy Wentworth Thompson (1860–1948)

The observation which comes closest to an explanation for the mathematical concepts cropping up in physics which I know is Einstein's statement that the only physical theories we are willing to accept are the beautiful ones.
Eugene Wigner (1902–1995)

A truly realistic mathematics should be conceived, in line with physics, as a branch of the theoretical construction of the one real world, and should adopt the same sober and cautious attitude toward hypothetic extensions of its foundations as is exhibited by physics.
Hermann Weyl (1885–1955)

The interplay between generality and individuality, deduction and construction, logic and imagination – this is the profound essence of live mathematics.
Any one or another of the aspects can be at the center of a given achievement. In a far-reaching development all of them will be involved. Generally speaking, such a development will start from the "concrete ground," then discard ballast by abstraction and rise to the lofty layers of thin air where navigation and observations are easy; after this flight comes the crucial test of landing and reaching specific goals in the newly surveyed low plains of individual "reality."
In brief, the flight into abstract generality must start from and return to the concrete and specific.[13]
Richard Courant (1888–1972)

[13] Mathematics in the modern world, Scientific American **211**(3) (1964), 41–49 (reprinted with permission).

There are mathematicians who reject a binding of mathematics to physics, and who justify mathematical work solely by aesthetical satisfaction which, besides all the difficulty of the material, mathematics is able to offer. Such mathematicians are more likely to regard mathematics as a form of art than science, and this point of view of mathematical unselfishness can be characterized by the slogan "l'art pour l'art".

On the other hand, there are physicists who regret that their science is so much related to mathematics. They fear a loss of intuition in the natural sciences. They consider the intimate relation with nature, the finding of ideas in nature itself, which was given to Goethe (1749–1832) in such a high degree, as being destroyed by mathematics, and their anger or sorrow is the more serious the more they are forced to realize the inevitability of mathematics.

Both points of view deserve serious consideration; because not only people with narrow minds have expressed such opinions. Yes, one can say that such a radical inclination to one side or the other, if not caused by a lack of talent, is sometimes evidence of a deeper perception of science, as if someone is interested in both sciences, but at the same time is satisfied with obvious connections between mathematics and physics...

Mathematics is an organ of knowledge and an infinite refinement of language. It grows from the usual language and world of intuitions as does a plant from the soil, and its roots are the numbers and simple geometrical intuitions. We do not know which kind of content mathematics (as the only adequate language) requires; we cannot imagine into what depths and distances this spiritual eye will lead us.[14]

Erich Kähler (1906–2000)

The most vitally characteristic fact about mathematics, in my opinion, is its quite peculiar relationship to the natural sciences, or more generally, to any science which interprets experience on a higher than purely descriptive level...

I think that this is a relatively good approximation to truth – which is much too complicated to allow anything but approximations – that mathematical ideas originate in *empirical facts*, although the genealogy is sometimes long and obscure. But, once they are so conceived, the subject begins to live a peculiar life of its own and is better compared to a creative one, governed by almost entirely aesthetic motivations, than to anything else and, in particular, to an empirical science.

But there is a grave danger that the subject will develop along the line of least resistance, that the stream, so far from its source, will separate into a multitude of insignificant tributaries, and that the discipline will become a disorganized mass of details and complexities. In other words, at a great distance from its empirical sources or after much *abstract inbreeding*, a mathematical object is in danger of degeneration. At the inception, the style is usually classical; when it shows signs of becoming *baroque*, then the danger signal is up...

Whenever this stage is reached, then the only remedy seems to be a rejuvenating return to the source: the re-injection of more or less directly *empirical ideas.* I am convinced that this is a necessary condition to con-

[14] On the relations of mathematics to physics and astronomy (in German), Jahresberichte der Deutschen Mathematiker-Vereinigung **51** (1941), 52–63 (reprinted with permission).

serve the freshness and the vitality of the subject and that this will remain equally true in the future.[15]

John von Neumann (1903–1957)

I want to say a word about the communication between mathematicians and physicists.
It has been very bad in the past, and some of the blame is doubtless to be laid on the physicist's shoulders. We tend to be very vague, and we don't know what the problem is until we have already seen how to solve it. We drive mathematicians crazy when we try to explain what our problems are. When we write articles we don't do a good enough job of specifying how certain we are about our statements; we do not distinguish guesses from theorems.
On the other hand, since I have said a lot of nice things about mathematics, I have to say that the mathematicians carry an even greater burden of guilt for this communication problem, largely because of their elitism. They often have, it seems to me, as their ideal the savant who is understandable only to a few co-specialists and who writes articles that one has to spend years to try to fathom.
When physicists write articles, they generally start them with a paragraph saying, "Up until now, this has been thought to be the case. Now, so – and – so has pointed out this problem. In this article, we are going to try to suggest a resolution of this difficulty." On the other hand, I have seen books of mathematics, not just articles but books, in which the first sentence in the preface was, "Let H be a nilpotent subgroup of..." These books are written in what I would call a lapidary style. The idea seems to be that there should be no word in the book that is not absolutely necessary, that is inserted merely to help the reader to understand what is going on.

I think this is getting much better. I find it is wonderful how mathematicians these days are willing to explain their field to interested physicists. This situation is improving, partly because as Iz Singer mentioned, we realize now that in certain areas we have much more in common than we had thought, but I think a lot more has to be done. There is still too much mathematics written which is not only not understandable to experimental or theoretical physicists, but is not even understandable to mathematicians who are not the graduate students of the author.[16]

Steven Weinberg (born 1933)

Relations between mathematics and physics vary with time. Right now, and for the past few years, harmony reigns and a honeymoon blossoms. However, I have seen other times, times of divorce and bitter battles, when the sister sciences declared each other as useless – or worse. The following exchange between a famous theoretical physicist and an equally famous mathematician might have been typical, some fifteen or twenty years ago: Says the physicist: "I have no use for mathematics. All the mathematics I ever need, I invent in one week."

[15] The Mathematician. In: The Works of the Mind, Vol. 1, pp. 180–196. Edited by R. Heywood, University of Chicago Press, 1947 (reprinted with permission).
[16] Mathematics: The unifying thread in science: Notices Amer. Math. Soc. **33** (1986), 716–733 (reprinted with permission).

Answers the mathematician: "You must mean the seven days it took the Lord to create the world."
A slightly more reliable document is found in the preface of the first edition of Hermann Weyl's book on group theory and quantum mechanics from 1928. He writes: "I cannot abstain from playing the role of an (often unwelcome) intermediary in this drama between mathematics and physics, which fertilize each other in the dark, and deny and misconstrue one another when face to face."
This dramatic situation, described here by one of the great masters in both sciences, is a result of recent times. At the time of Newton (1643–1727) disharmony between mathematics and physics seemed unthinkable and unnatural, since both were his brainchildren; and close symbiosis persisted through the whole of the eighteenth century. The rift arose around 1800 and was caused by the development of pure mathematics (represented by number theory) on the one hand, and of a new kind of physics, independent of mathematics, which developed out of chemistry, electricity and magnetism on the other. This rift was widened in Germany under the influence of Goethe (1749–1832) and his followers, Schelling (1775–1854) and Hegel (1770–1831) and their "Naturphilosophie".
Our protagonists are Carl Friedrich Gauss (1777–1855), as the creator of modern number theory, and Michael Faraday (1791-1867) as the inventor of physics without mathematics (in the strict sense of the word).
It would be foolish, of course, to claim the nonexistence of number theory before Gauss. An amusing document may illustrate the historical development. Erich Hecke's famous *Lectures on the Theory of Algebraic Numbers* has on its last page a "timetable", which chronologically lists the names and dates of the great number theoreticians, starting with Euclid (300 B.C.) and ending with Hermann Minkowski (1864–1909). As a physicist, I am impressed to find so many familiar names in this Hall of Fame: Fermat (1601–1665), Euler (1707–1783), Lagrange (1736–1813), Legendre (1752–1833), Fourier (1768–1830), and Gauss. In fact, we cannot find a single great number theoretician before Gauss, whom we would not count among the great physicists, provided we disregard antiquity. Specialization starts after 1800 with names like Kummer (1810–1891), Galois (1811–1832), and Eisenstein (1823–1852); who were all under the great influence of Gauss' *Disquisitiones arithmeticae* from 1801. In this specific sense, Gauss' book marks the dividing line between mathematics as a universal science and mathematics as a union of special disciplines, and between the "géomètre" as a universal "savant" in the sense of the eighteenth century and the specialized "mathématicien" of modern times. As is typical for a man of transition, Gauss does not belong to either category, he was universal and specialized. The struggle raged within him – and made him suffer.

Res Jost (1918–1990)
Mathematics and physics since 1800: discord and sympathy[17]

By a particular prerogative, not only does each man advance day by day in the sciences, but all men together make continual progress as the universe ages... Thus, the entire body of mankind as a whole, over many centuries, must be considered as a single man, who lives forever and continues to learn.

Blaise Pascal (1623–1662)

[17] In: R. Jost, The Fairy Tale about the Ivory Tower, essays and lectures. Edited by K. Hepp, W. Hunziker, and W. Kohn, Springer, Berlin, pp. 219–240 (reprinted with permission).

References

Abdalla, E., Abdalla, M., Rothe, K. (2001), Non-Perturbative Methods in Two-Dimensional Quantum Field Theory, World Scientific, Singapore.[1]

Abraham, R., Marsden, J. (1978), Foundations of Mechanics, Addison-Wesley, Reading, Massachusetts.

Abraham, R., Marsden, J., Ratiu, T. (1988), Manifolds, Tensor Analysis, and Applications, Springer, New York.

Abramowitz, M., Stegun, I. (Eds.) (1984), Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, Wiley, New York, and National Bureau of Standards, Washington, DC.

Adams, C. (1994), The Knot Book, Cambridge University Press, Cambridge, United Kingdom.

Adams, F., Laughlin, G. (1997), A dying universe: the long-term fate and evolution of astrophysical objects, Rev. Mod. Phys. **69**, 337–372.

Adams, F., Laughlin, G. (1999), The Five Ages of the Universe: Inside the Physics of Eternity, Simon and Schuster, New York.

Agricola, I., Friedrich, T. (2002), Global Analysis: Differential Forms in Analysis, Geometry and Physics, Amer. Math. Soc., Providence, Rhode Island (translated from German into English).

Albeverio, S., Høegh-Krohn, R. (1975), Mathematical Theory of the Feynman Path Integral, Springer, Berlin.

Albeverio, S. (1986), Nonstandard Methods in Stochastic Analysis and Mathematical Physics, Academic Press, New York.

Albeverio, S. (1988), Solvable Models in Quantum Mechanics, Springer, New York.

Aleksandrov, M., Kontsevich, M., Schwarz, A., Zaboronsky, O. (1997), Geometry of the master equation, Int. J. Mod. Phys. **A12**, 1405–1430.

Ali, S., Engliš, M. (2004), Quantization methods: a guide for physicists and analysts. Internet: `http://arXiv:math-ph/0405065`

Alten, H., Naini, D., Folkerts, M., Schlosser, H., Schlote, K., Wussing, H. (2003), 4000 Jahre Algebra: Geschichte, Kulturen, Menschen (4000 Years of Algebra), Springer, Berlin (in German) (see also Scriba and Schreiber (2003), 5000 Jahre Geometrie).

Amann, H. (1990), Ordinary Differential Equations: An Introduction to Nonlinear Analysis, de Gruyter, Berlin (translated from German into English).

[1] Hints for further reading can be found in Chap. 17. The author's homepage contains a complete list of the references to Volumes I through VI.
Internet: `http://www.mis.mpg.de/`
We also refer to E. Zeidler (Ed.), Oxford User's Guide to Mathematics, Oxford University Press, 2004, which contains a comprehensive list about the standard literature in mathematics.

Amann, H., Escher, J. (2004), Analysis I–III, Birkhäuser, Basel (translated from German into English).
Amelino-Camelia, G., Kowalski-Glikman, J. (Eds.) (2005), Planck-Scale Effects in Astrophysics and Cosmology, Springer, Berlin.
Anosov, D., Bolibruch, A. (1994), The Riemann–Hilbert Problem, Vieweg, Wiesbaden.
Apostol, T. (1986), Introduction to Analytic Number Theory, Springer, New York.
Apostol, T. (1990), Modular Functions and Dirichlet Series in Number Theory, 2nd edn., Springer, New York.
Araki, H. (1999), Mathematical Theory of Quantum Fields, Oxford University Press, Oxford (translated from Japanese into English).
Arnold, V. (1978), Mathematical Theory of Classical Mechanics, Springer, Berlin (translated from Russian into English).
Arnold, V., Gusein-Zade, S., Varchenko, A. (1985), Singularities of Differentiable Maps, Vols. 1, 2, Birkhäuser, Basel (translated from Russian into English).
Arnold, V. (2004), Lectures on Partial Differential Equations, Springer, Berlin (translated from Russian into English).
Arnold, V., Faddeev, L., Filippov, V., Manin, Yu., Tikhomirov, V., Vershik, A., Bolibruch, A., Osipov, Yu., Sinai, Ya. (Eds.) (2006), Mathematical Events of the Twentieth Century, Springer, New York.
Asch, J., Joye, A. (Eds.) (2006), Mathematical Physics of Quantum Mechanics: Selected and Refereed Lectures from *The State of the Art in the Mathematical Physics of Quantum Systems*, QMath9, held in Giens, France, in 2004, Springer, Berlin.
Ashtekhar, A. (1991), Lectures on Non-Perturbative Canonical Gravity, World Scientific, Singapore.
Ashtekhar, A., Lewandowski, J. (1997), Quantum theory of geometry I: Area operators, II: Volume operators, Class. Quant. Grav. **14**, A55–A81, Adv. Theo. Math. Phys. **1**, 388–429.
Ashtekar, A., Bojewald, M., Lewandowski, J. (2003), Mathematical structure of loop quantum cosmology, Adv. Theor. Math. Phys. **7**, 233–268.
Ashtekar, A., Lewandowski, J. (2004), Background independent quantum gravity: A status report, Class. Quant. Grav. **21**, R53–R152.
Atiyah, M. (1970), Resolution of singularities and division of distributions, Commun. Pure Appl. Math. **23**, 145–150.
Atiyah, M. (1979), Geometry of Yang–Mills Fields, Lezioni Fermiani, Academia Nazionale dei Lincei Scuola Normale Superiore, Pisa, Italia.
Atiyah, M., Hitchin, M. (1988), The Geometry and Dynamics of Magnetic Monopoles, Princeton University Press, Princeton, New Jersey.
Atiyah, M. (1990a), The Work of Edward Witten, Proc. Intern. Congr. Math. Kyoto 1990, Math. Soc. Japan 1991. In: Atiyah and Iagolnitzer (Eds.) (2003), pp. 514–518.
Atiyah, M. (1990b), Geometry and Physics of Knots, Cambridge University Press, Cambridge, United Kingdom.
Atiyah, M. (2002), Mathematics in the 20th century, Bull. London Math. Soc. **34**, 1–15.
Atiyah, M., Iagolnitzer, D. (Eds.) (2003), Fields Medallists' Lectures, World Scientific, Singapore.
Atiyah, M. (2004), The impact of Thom's cobordism theory, Bull. Amer. Math. Soc. **41** (3), 337–340.
Atiyah, M. (2005), Collected Works, Vols. 1–6, Cambridge University Press, Cambridge, United Kingdom.

Awramik, M., Czakon, M. (2003a), Complete two loop electroweak contributions to the muon lifetime in the standard model, Phys. Lett. B **568**, 48–54.
Awramik, M., Czakon, M. (2003b), Two loop electroweak corrections to the muon decay lifetime, Nucl. Phys. Proc. Suppl. **116**, 238–242.
Awramik, M., Czakon, M., Freitas, A., Weiglin, G. (2004), Precise prediction for the W-boson mass in the standard model, Phys. Rev. D **69**, 53006–16. Internet: `http://arXiv:hep-ph/0311148`

Bailin, D., Love, A. (1996), Introduction to Gauge Field Theory, 2nd edn., Institute of Physics, Bristol.
Bailin, D., Love, A. (1997), Supersymmetric Gauge Field Theory and String Theory, 2nd edn., Institute of Physics, Bristol.
Baker, A. (2002), Matrix Groups: An Introduction to Lie Group Theory, Springer, New York.
Bandyopadhyay, P. (1996), Geometry, Topology, and Quantization, Kluwer, Dordrecht.
Barenblatt, G. (1996), Scaling, Self-Similarity, and Intermediate Asymptotics, Cambridge University Press, New York (translated from Russian into English).
Barton, G. (1989), Elements of Green's Functions and Propagation: Potentials, Diffusion, and Waves, Clarendon Press, Oxford.
Barut, A. (1967), The Theory of the Scattering Matrix, MacMillan, New York.
Basdevant, J. (2000), The Quantum-Mechanics Solver: How to Apply Quantum Theory to Modern Physics, Springer, Berlin.
Basdevant, J., Dalibard, J. (2002), Quantum Mechanics, Springer, Berlin.
Bateman, H. (Ed.) (1953), Higher Transcendental Functions, Vols. 1–3, MacGraw Hill, New York.
Battle, G. (1999), Wavelets and Renormalization, World Scientific, Singapore.
Baumgärtel, H. (1985), Analytic Perturbation Theory for Matrices and Operators, Birkhäuser, Boston.
Baumgärtel, H., Wollenberg, M. (1992), Causal Nets of Operator Algebras, Akademie-Verlag, Berlin.
Baumgärtel, H. (1995), Operator-Algebraic Methods in Quantum Field Theory, Akademie-Verlag, Berlin.
Baxter, R. (1982), Exactly Solved Models in Statistical Mechanics, Academic Press, New York.
Baym, G. (1969), Lectures on Quantum Mechanics, Benjamin, Menlo Park, California.
Belger, M., Schimming, R., Wünsch, V. (1997), A survey on Huygens' principle, J. for Analysis and its Applications **16**(1), 5–32.
Bell, E. (1986), Men of Mathematics: Biographies of the Greatest Mathematicians of all Times, Simon, New York.
Benfatto, G., Gallavotti, G. (1995), Renormalization Group, Princeton University Press, Princeton, New Jersey.
Berezansky, Yu., Kondratiev, Yu. (1995), Spectral Methods in Infinite-Dimensional Analysis, Vols. 1, 2, Kluwer, Dordrecht (translated from Russian into English).
Berezin, F. (1966), The Method of Second Quantization, Academic Press, New York (translated from Russian into English).
Berezin, F. (1987), Introduction to Superanalysis, Reidel, Dordrecht (translated from Russian into English).
Berezin, F., Shubin, M. (1991), The Schrödinger Equation, Kluwer, Dordrecht (translated from Russian into English).
Berline, N., Getzler, E., Vergne, M. (1991), Heat Kernels and Dirac Operators, Springer, New York.

Bethe, H., Salpeter, E. (1957), Quantum Mechanics of One-and Two-Electron Atoms, Springer, Berlin.
Bethe, H., Bacher, R., Livingstone, M. (1986), Basic Bethe: Seminal Articles on Nuclear Physics 1936–37, American Institute of Physics, New York.
Bethuel, F., Brézis, H., Hélein, F. (1994), Ginzburg–Landau Vortices, Birkhäuser, Basel.
Bjorken, J., Drell, S. (1964), Relativistic Quantum Mechanics, McGraw-Hill, New York.
Bjorken, J., Drell, S. (1965), Relativistic Quantum Fields, McGraw-Hill, New York.
Blau, S., Visser, M., Wipf, A. (1988), Zeta functions and the Casimir energy, Nucl. Phys. B **310**, 163–180.
Bocaletti, D., Pucacco, G. (1998), Theory of Orbits, Vol 1: Integrable Systems and Non-Perturbative Methods, Vol. 2: Perturbative and Geometrical Methods, Springer, Berlin.
Bodanis, D. (2000), $E = mc^2$: A Biography of the World's Most Famous Equation, Walker, New York.
Bogoliubov, N., Shirkov, D. (1956), Charge renormalization group in quantum field theory, Nuovo Cimento **3**, 845–863
Bogoliubov, N., Parasiuk, O. (1957), Über die Multiplikation der Kausalfunktionen in der Quantentheorie der Felder (On the multiplication of propagators in quantum field theory), Acta Math. **97**, 227–326 (in German).
Bogoliubov, N., Medvedev, B., Polivanov, M. (1958), Problems of the Theory of Dispersion Relations, Fizmatgiz, Moscow (in Russian).
Bogoliubov, N. (1967), Lectures on Quantum Statistics, Vols. 1, 2, Gordon and Breach, New York (translated from Russian into English).
Bogoliubov, N., Logunov, A., Todorov, I. (1975), Introduction to Axiomatic Quantum Field Theory, Benjamin, Reading, Massachusetts (translated from Russian into English).
Bogoliubov, N. Shirkov, D. (1980), Introduction to Quantum Field Theory, 3rd edn., Wiley, New York (translated from Russian into English).
Bogoliubov, N., Shirkov, D. (1983), Quantum Fields, Benjamin, Reading, Massachusetts (translated from Russian into English).
Bogoliubov, N., Logunov, A., Orsak, A., Todorov, I. (1990), General Principles of Quantum Field Theory, Kluwer, Dordrecht (translated from Russian into English).[2]
Bogoliubov, N. (1992), Statistical Mechanics and the Theory of Dynamical Systems. Amer. Math. Soc., Providence, Rhode Island (translated from Russian into Engllish).
Böhm, M., Denner, A., Joos, H. (2001), Gauge Theories of the Strong and Electroweak Interaction, 2nd edn., Teubner, Stuttgart.
Borel, A. (1998), Twenty-Fife Years with Nicolas Bourbaki, 1949–1973, Notices Amer. Math. Soc. **45**(3), 373–380.
Born, M. (1969), Physics in My Generation, Springer, New York (translated from German into English).
Born, M., Wolf, E. (1970), Principles of Optics, 4th edn., Pergamon Press, New York.
Born, M. (1977), My Life: Recollection's of a Nobel Laureat, Charles Sribner's Sons, New York (translated from German into English).
Börner, G. (2003), The Early Universe: Facts and Fiction, 4th edn., Springer, Berlin.

[2] 1200 references

Bott, R., Tu, L. (1982), Differential Forms in Algebraic Topology, Springer, New York.
Bott, R. (1994), Collected Papers, Vols. 1–4, Birkhäuser, Boston.
Bouwmeester, D., Ekert, A., Zeilinger, A. (Eds.) (2002), The Physics of Quantum Information: Quantum Cryptography, Quantum Teleportation, Quantum Computation, Springer, Berlin.
Bratelli, C., Robinson, D. (2002), Operator Algebras and Quantum Statistical Mechanics, Vols. 1, 2, 2nd edn., Springer, New York.
Bredon, G. (1993), Topology and Geometry, Springer, New York.
Brennan, R. (1997), Heisenberg Probably Slept Here: The Lives, Times, and Ideas of the Great Physicists of the 20th Century, Wiley, New York.
Brézis, H., Browder, F. (1998), Partial differential equations in the 20th century, Advances in Math. **135**, 76–144.
Brian, D. (2005), Einstein – a Life, Wiley-VCH, Weinheim, Germany.
Bricmont, J., Gawędzki, K., Kupiainen, A. (1999), Kolmogorov–Arnold–Moser (KAM) theorem and quantum field theory, Commun. Math. Phys. **201** (3), 699–727.
Brown, L. (1992), Quantum Field Theory, Cambridge University Press, New York.
Brown, L. (Ed.) (1993), Renormalization: From Lorentz to Landau and Beyond, Springer, New York.
Brunetti, R., Fredenhagen, K., Köhler, M. (1996), The microlocal spectrum condition and the Wick polynomials of free fields, Commun. Math. Phys. **180**, 633–652.
Brunetti, R., Fredenhagen, K. (2000), Micro-local analysis and interacting quantum field theories: renormalization on physical backgrounds, Commun. Math. Phys. **208**, 623–661.
Bühler, W. (1981), Gauss: A Bibliographical Study, Springer, New York.
Burgess, M. (2002), Classical Covariant Fields, Cambridge University Press, Cambridge, United Kingdom.
Buschhorn, G., Wess, J. (Eds.) (2004), Fundamental Physics: Heisenberg and Beyond, Springer, Berlin.
Bytsenko, A., Cognola, G., Elizalde, E., Moretti, V., Zerbini, S. (2003), Analytic Aspects of Quantum Fields, World Scientific, Singapore.

Cameron, R. (1960), A family of integrals serving to connect the Wiener and Feynman integrals, J. of Math. and Phys. Sci. of MIT **39**, 126–140.
Carmichael, H. (1995), Statistical Methods in Quantum Optics: Master Equations and Fokker–Planck Equations, Springer, New York.
Cartier, P. (1995), An introduction to zeta functions. In: Waldschmidt et al. (Eds.) (1995), pp. 1–63.
Cartier, P. (2000), Mathemagics, A tribute to L. Euler and R. Feynman, Séminaire Lotharingien **44**, 1–71.
Cartier, P. (2001), A mad day's work: from Grothendieck to Connes and Kontsevich. The evolution of concepts of space and symmetry. Bull. Amer. Math. Soc. **38**(4), 389–408.
Casimir, H. (1948), On the attraction between two perfectly conducting plates, Proc. Kon. Nederl. Akad. Wetensch. **51**, 793–795.
Cattaneo, A., Felder, G. (2000), A path integral approach to the Kontsevich quantization formula, Commun. Math. Phys. **212**, 591–611.
Chaichian, M., Nelipa, N. (1984), Introduction to Gauge Field Theories, Springer, Berlin.
Chaichian, M., Demichev, A. (1996), Introduction to Quantum Groups, World Scientific, Singapore.

Chaichian, M., Hagedorn, R. (1998), Symmetries in Quantum Mechanics: From Angular Momentum to Supersymmetry, Institute of Physics, Bristol.
Chaichian, M., Demichev, A. (2001), Path Integrals in Physics. Vol. 1: Stochastic Processes and Quantum Mechanics; Vol. 2: Quantum Field Theory, Statistical Physics, and Other Modern Applications, Institute of Physics, Bristol.
Chang, K. (2005), Methods in Nonlinear Analysis, Springer, Berlin.
Chang, S. (1990), Introduction to Quantum Field Theory, World Scientific, Singapore.
Chern, S., Hirzebruch, F. (Eds.) (2001), Wolf Prize in Mathematics, Vols. 1, 2, World Scientific, Singapore.
Chew, G. (1966), The Analytic S-Matrix: A Basis for Nuclear Democracy, Benjamin, New York.
Choquet-Bruhat, Y., DeWitt-Morette, C., Dillard-Bleick, M. (1996), Analysis, Manifolds, and Physics. Vol. 1: Basics; Vol 2: 92 Applications, Elsevier, Amsterdam.
Chung, K., Zhao, Z. (1995), From Brownian Motion to Schrödinger's Equation, Springer, New York.
Collins, J. (1984), Renormalization: An Introduction to Renormalization, the Renormalization Group, and the Operator-Product Expansion, Cambridge University Press, Cambridge, United Kingdom.
Colombeau, J. (1984), New Generalized Functions and Multiplication of Distributions, North-Holland, Amsterdam.
Connes, A. (1985), Noncommutative differential geometry, Publ. Math. IHES **62**, 257–360.
Connes, A. (1988), The action functional in noncommutative geometry, Commun. Math. Phys. **117**, 673–683.
Connes, A., Lott, J. (1990), Particle models and noncommutative geometry, Nucl. Phys. B (Proc. Suppl.) **18**, 29–47.
Connes, A. (1994), Noncommutative Geometry, Academic Press, New York.
Connes, A. (1995), Noncommutative geometry and reality, J. Math. Phys. **36**, 6194–6231.
Connes, A., Kreimer, D. (1998), Hopf algebras, renormalization and noncommutative geometry, Commun. Math. Phys. **199**, 203–242.
Connes, A., Kreimer, D. (2000a), Renormalization in quantum field theory and the Riemann–Hilbert problem I: The Hopf algebra structure of graphs and the main theorem, Commun. Math. Phys. **210**, 249–273.
Connes, A., Kreimer, D. (2000b), Renormalization in quantum field theory and the Riemann–Hilbert problem II: The beta function, diffeomorphisms, and the renormalization group, Commun. Math. Phys. **216**, 215–241.
Connes, A. (2000), Noncommutative Geometry and the Riemann Zeta Function, Frontieres and Perspectives, International Mathematical Union.
Connes, A., Lichnerowicz, A., Schützenberger, M. (2001), Triangle of Thoughts, Amer. Math. Soc., Providence, Rhode Island.
Connes, A. (2003), Symmétries galoisiennes et renormalisation. In: Duplantier and Rivasseau (2003), pp. 241–264.
Connes, A., Marcolli, M. (2004), Renormalization and motivic Galois theory. Internet: `http://arXiv:math.NT/0409306`
Connes, A., Marcolli, M. (2005), From number theory to physics I, II (to appear in the volumes "Frontieres in Number Theory, Physics, and Geometry"). Internet: `http://www.math.dsu.edu/~marcolii/renorm43.pdf`
Consani, C., Marcolli, M. (Eds.) (2006), Noncommutative Geometry and Number Theory: Where Arithmetic meets Geometry and Physics, Vieweg, Wiesbaden.

Cottingham, W., Greenwood, D. (1998), An Introduction to the Standard Model of Particle Physics, Cambridge University Press, Cambridge, United Kingdom.
Coughlan, G., Dood, J. (1991), The Ideas of Particle Physics: An Introduction for Scientists, 2nd edn., Cambridge University Press, Cambridge, United Kingdom.
Courant, R. (1964), Mathematics in the modern world, Scientific American **211**(3), 41–49.
Courant, R., Hilbert, D. (1989), Methods of Mathematical Physics, Vols. 1, 2, Wiley, New York (translated from German into English).
Cox, D., Little, J., O'Shea, D. (1998), Using Algebraic Geometry, Springer, New York.
Coxeter, H. (1995), Projective Geometry, 2nd edn., Springer, New York.
Cycon, R., Froese, R., Kirsch, W., Simon, B. (1986), Schrödinger Operators, Springer, New York.

Dardo, M. (2004), Nobel Laureates and Twentieth-Century Physics, Cambridge University Press, Cambridge, United Kingdom.
Dautray, R., Lions, J. (1988), Mathematical Analysis and Numerical Methods for Science and Technology, Vols. 1–6, Springer, New York (translated from French into English).
Dauxois, T, Peyrard, M. (2006), Physics of Solitons, Cambridge University Press, Cambridge, United Kingdom.
de Gennes, P. (1979), Scaling Concepts in Polymer Physics, Cornell University Press, Ithaca.
de Gennes, P., Prost, J. (1995), The Physics of Liquid Crystals, Clarendon Press, Oxford.
Deligne, P., Etingof, P., Freed, D., Jeffrey, L., Kazhdan, D., Morgan, J., Morrison, D., Witten, E. (Eds.) (1999), Lectures on Quantum Field Theory: A Course for Mathematicians Given at the Institute for Advanced Study in Princeton, Vols. 1, 2, Amer. Math. Soc., Providence, Rhode Island.
de Melo, W., van Strien, S. (1993), One-Dimensional Dynamics, Springer, Berlin.
Dereziński A., Gérard, C. (1997), Scattering Theory of Classical and N-Particle Systems, Springer, New York.
DeSimone, A., Kohn, R., Müller, S., Otto, F. (2002), A reduced theory for thin-film micromagnetics, Comm. Pure Appl. Math. **55**, 1408–1460.
Devlin, K. (2002), The Millenium Problems: The Seven Greatest Unsolved Mathematical Puzzles of Our Time, Basic Books, New York.
Dierkes, U., Hildebrandt, S., Küster, A., Wohlrab, O. (1992), Minmal Surfaces, Vols. 1, 2, Springer, Berlin.
Dieudonné, J. (1968), Foundations of Modern Analysis, Vols. 1–9, Academic Press, New York (translated from French into English). (German edition: Grundzüge der modernen Analysis, Vols. 1–9, Deutscher Verlag der Wissenschaften, Berlin 1972–1987.)
Dieudonné, J. et al. (Eds.) (1978), Abrégé d'histoire des mathématiques, 1700–1900, Vols. 1, 2, Hermann, Paris (in French). (German edition: Geschichte der Mathematik: ein Abriß 1700–1900, Vieweg, Braunschweig, 1985.)
Dieudonné, J. (1981), History of Functional Analysis, 1900–1975, North-Holland, Amsterdam.
Dieudonné, J. (1982), A Panorama of Pure Mathematics as Seen by Bourbaki, Academic Press, New York (translated from French into English).
Dieudonné, J. (1985), History of Algebraic Geometry, 400 B.C.–1985 A.C., Chapman, New York.
Dieudonnè, J. (1989), A History of Algebraic and Differential Topology, 1900–1960, Birkhäuser, Boston.

Dodson, C., Parker, P. (1997), A User's Guide to Algebraic Topology, Kluwer, Dordrecht.
Doetsch, G. (1956), Handbuch der Laplace-Transformation (Handbook of the Laplace transformation), Vols. 1–3, Birkhäuser, Basel (in German).
Dolzmann, G. (2003), Variational Methods for Crystalline Microstructure: Analysis and Computation, Springer, Berlin.
Donaldson, S., Kronheimer, P. (1990), The Geometry of Four-Manifolds, Oxford University Press, Oxford.
Donaldson, S. (1996), The Seiberg–Witten Equations and 4-Manifold Topology, Bull. Amer. Math. Soc. **33**, 45–70.
Donaldson, S. (2002), Floer Homology Groups, Cambridge University Press, Cambridge, United Kingdom.
Dubrovin, B., Fomenko, A., Novikov, S. (1992), Modern Geometry: Methods and Applications, Vols. 1–3, Springer, New York (translated from Russian into English).
Duistermaat, J., Kolk, J. (2000), Lie Groups, Springer, New York.
Dunlap, R. (1997), The Golden Ratio and Fibonacci Numbers, World Scientific, Singapore.
Duplantier, B., Rivasseau, V. (Eds.) (2003), Vacuum Energy – Renormalization, Poincaré Seminar 2002, Birkhäuser, Basel.
Dyson, F. (1952), Divergence of perturbation theory in quantum electrodynamics, Phys. Rev. **85**, 631–632.
Dyson, F. (1979), Disturbing the Universe, Harper and Row, New York.
Dyson, F. (1992), From Eros to Gaia, Pantheon Books, New York.
Dyson, F. (1993), George Green and Physics, Phys. World **6**, August 1993, 33–38.
Dyson, F. (1996), Selected Papers of Freeman Dyson with Commentaries, Amer. Math. Soc., Providence, Rhode Island.
Dyson, F. (1999a), Origins of Life, Cambridge University Press, Cambridge, United Kingdom.
Dyson, F. (1999b), The Sun, the Genome, and the Internet: Tools of Scientific Revolution, Oxford University Press, New York.

Economu, E. (1988), Green's Functions in Quantum Physics, Springer, New York.
Eden, R. et al. (1966), The Analytic S-Matrix Theory, Cambridge University Press, Cambridge, United Kingdom.
Egorov, Yu., Shubin, M. (1991), Partial Differential Equations, Vols. 1–4, Springer, New York (translated from Russian into English).
Egorov, Yu., Shubin, M. (1998), Foundations of the Classical Theory of Partial Differential Equations, New York (translated from Russian into English).
Egorov, Yu., Komech, A., Shubin, M. (1999), Elements of the Modern Theory of Partial Differential Equations, Springer, New York.
Ehlers, J., Falco, E., Schneider, P. (1992), Gravitational Lenses, Springer, New York.
Eichhorn, J. (2006), Global Analysis on Open Manifolds, Nova Science Publishers, New York.
Einstein, A. (1949), Albert Einstein als Philosoph und Naturforscher (Albert Einstein as philosopher and scientist). Edited by P. Schilpp, Kohlhammer Verlag, Stuttgart (in German).
Einstein, A. (1982), Ideas and Opinions, Three Rivers Press, New York.
Eisenbud, D. (1994), Commutative Algebra with a View to Algebraic Geometry, Springer, New York.
Eisenbud, D. (2005), The Geometry of Syzygies: A Second Course in Commutative Algebra and Algebraic Geometry, Springer, New York.

Elizalde, E. (1995), Ten Physical Applications of Spectral Zeta Functions, Springer, Berlin.
Emch, G. (1972), Algebraic Methods in Statistical Physics and Quantum Field Theory, Wiley, New York.
Emch, G. (1984), Mathematical and conceptual foundations of 20th century physics, North-Holland, Amsterdam.
Emch, G., Liu, C. (2002), The Logic of Thermostatistical Physics, Springer, New York.[3]
Encyclopedia of Mathematical Physics (2006), Vols. 1–4. Edited by J. Françoise, G. Naber, and T. Tsun, Elsevier, Oxford.
Engquist, B., Schmid, W. (Eds.) (2001), Mathematics Unlimited – 2001 and Beyond, Springer, New York.
Erdély, A., Magnus, W., Oberhettinger, F., Tricomi, F. (Eds.) (1955), Higher Transcendental Functions, Vols. 1–3, McGraw-Hill, New York.
Erdélyi, A. (1965), Asymptotic Expansions, Dover, New York.
Eschrig, H. (2003), The Fundamentals of Density Functional Theory, 2nd edn., Teubner, Leipzig.
Etingof, P. (2000), A note on dimensional regularization. In: P. Deligne et al. (Eds.) (2000), Vol. 1, pp. 597–607.
Evans, C. (1998), Partial Differential Equations, Amer. Math. Soc., Providence, Rhode Island.
Evans, E., Kawahigashi, Y. (1998), Quantum Symmetries on Operator Algebras, Clarendon Press, Oxford.
Evans, G., Blackledge, J., Yardley, P. (2000), Analytic Methods for Partial Differential Equations, Springer, London.

Faddeev, L. Slavnov, A. (1980), Gauge Fields, Benjamin, Reading, Massachusetts (translated from Russian into English).
Faddeev, L. (1999), Elementary introduction to quantum field theory. In: P. Deligne et al. (Eds.) (1999), pp. 513–550.
Faltings, G. (1994), A proof of the Verlinde formula, J. Alg. Geometry, **3**, 347–374.
Farkas, H., Kra, I. (1992), Riemann Surfaces, Springer, New York.
Farkas, H., Kra, I. (2001), Theta Constants, Riemann Surfaces and the Modular Group: An Introduction with Applications to Uniformization Theorems, Partition Identities and Combinatorial Number Theory, Amer. Math. Soc., Providence, Rhode Island.
Fauser, B., Tolksdorf, J., Zeidler, E. (2006) (Eds.), Quantum Gravity: Status and Ramifications, Birkhäuser, Basel (to appear).
Federbush, P. (1987), Quantum field theory in ninety minutes, Bull. Amer. Math. Soc. **17**(1), 93–103.
Fedosov, B. (1996), Deformation Quantization and Index Theory, Akademie-Verlag, Berlin
Fefferman, C. (1983), The uncertainty principle, Bull. Amer. Math. Soc. **9**(2), 129–206.
Fefferman, C. (1985), The atomic and molecular nature of matter, Revista Mat. Iberoamer. **1**(1), 1–44.
Fefferman, C., de la Llave, R. (1986), Relativistic stability of matter, Revista Mat. Iberoamer. **2**, 119–213.
Fefferman C., Seco, L. (1994), On the Dirac and Schwinger corrections to the ground-state energy of an atom, Advances in Math. **107**(1), 1–185.

[3] 1500 references

Feldman, J., Hurd, T., Rosen, L., Wright, J. (1988): QED: A Proof of Renormalizability, Springer, Berlin.
Feldman, J., E. Trubowitz, E. (1992), Renormalization in classical mechanics and many body quantum field theory, Jerusalem J. d'Analyse Mathématique **52**, 213–247. Internet: `http://www.math.ubc.ca.~feldman/research.html`
Feldman, J., Knörrer, H., Trubowitz, E. (2002), Fermionic Functional Integrals and the Renormalization Group, Amer. Math. Soc., Providence, Rhode Island.
Feldman, J., Knörrer, H., Trubowitz, E. (2003), A two-dimensional Fermi liquid. Internet: `http://www.math.ubc.ca/~feldman/fl.htlm`
Felsager, B. (1997), Geometry, Particles, and Fields, Springer, New York.
Fernández, R., Fröhlich, J., Sokal, D. (1992), Random Walks, Critical Phenomena, and Triviality in Quantum Field Theory, Springer, Berlin.
Ferris, T. (1991), The World Treasury of Physics, Astronomy, and Mathematics, Brown, Boston.
Fetter, A., Walecka, J. (1971), Quantum Theory of Many-Particle Systems, McGraw-Hill, New York.
Feynman, R. (1961), Quantum Electrodynamics, Benjamin, Reading, Massachusetts.
Feynman, F., Leighton, R., Sands, M. (1963), The Feynman Lectures in Physics, Addison-Wesley, Reading, Massachusetts.
Feynman, R., Hibbs, R. (1965), Quantum Mechanics and Path Integrals, McGraw-Hill, New York.
Feynman, R. (1966), The Character of Physical Law, MIT Press, Cambridge, Massachusetts.
Feynman, R. (1985), QED (Quantum Electrodynamics): The Strange Theory of Light and Matter, Mautner Memorial Lectures, Princeton University Press, Princeton, New Jersey.
Feynman, R., Leighton, R. (1985), Surely You're Joking Mr. Feynman: Adventures of a Curious Character, Norton, New York.
Feynman, R. (1995), Feynman Lectures on Gravitation. Edited by B. Hatfield, Addison-Wesley, Boston.
Feynman, R. (1998), Statistical Mechanics: A Set of Lectures, 14th edn., Addison Wesley, Reading, Massachusetts.
Figueroa, H. Gracia-Bondia, J. (2005), Combinatorial Hopf algebras in quantum field theory I, Rev. Math. Phys. **17**, 881–975.
Internet: `http//:arXiv:hep-th/0408145`
Fikhtengol'ts, G. (1965), The Fundamentals of Mathematical Analysis, Vols. 1–3, Pergamon Press, Oxford (translated from Russian into English).
Finster, F. (2006), The Principle of the Fermionic Projector, American Mathematical Society and International Press, Boston.
Flappan, E. (2000), When Topology Meets Chemistry: A Topological Look at Molecular Chirality, Cambridge University Press, Cambridge, United Kingdom.
Ford, L. (1972), Automorphic Functions, 4th edn., Chelsea, New York.
Fowler, A. (1997), Mathematical Models in the Applied Sciences, Cambridge University Press, Cambridge, United Kingdom.
Fradkin, E., Palchik, M. (1996), Conformal Quantum Field Theory in D Dimensions, Kluwer, Dordrecht.
Frank, T. (2005), Nonlinear Fokker–Planck Equation, Springer, Berlin.
Frankel, T. (1999), The Geometry of Physics, Cambridge University Press, Cambridge, United Kingdom.
Frappat, L., Sciarinno, A., Sorba, P. (2000), Dictionary of Lie Algebras and Super Lie Algebras, Academic Press, New York.

Freidlin, M. (1985), Functional Integration and Partial Differential Equations, Princeton University Press, Princeton, New Jersey.
Frenkel, I., Lepowski, J., Meurman, A. (1988), Vertex Operator Algebras and the Monster, Academic Press, New York.
Frenkel, I., Ben-Zvi, D. (2001), Vertex Algebras and Algebraic Curves, American Mathematical Society, Providence, Rhode Island.
Friedrich, T. (Ed.) (1981), Self-Dual Riemannian Geometry and Instantons, Teubner, Leipzig.
Friedrich, T. (2000), Dirac Operators in Riemannian Geometry, Amer. Math. Soc., Providence, Rhode Island (translated from German into English).
Friedrichs, K. (1927), Eine invariante Formulierung des Newtonschen Graviationsgesetzes und des Grenzübergangs vom Einsteinschen zum Newtonschen Gesetz (An invariant formulation of Newton's gravitational law and the passage from Einstein's theory of general relativity to Newton's classical theory), Math. Ann. **98**, 566–575.
Friedrichs, K. (1953), Mathematical Aspects of the Quantum Theory of Fields, Interscience, New York.
Friesecke, G., James, R., Müller, S. (2002), A theorem on geometric rigidity and the derivation of nonlinear plate theory from three-dimensional elasticity, Comm. Pure Appl. Math. **55**, 1461–1506.
Fritzsch, H., Gell-Mann, M., Leutwyler, H. (1973), Advantages of the color octet gluon picture, Phys. Lett. **47B**, 365–368.
Fritzsch, H. (1992), Quarks, Penguin, London (translated from German into English).
Fuchs, J. (1992), Affine Lie Algebras and Quantum Groups: An Introduction with Applications in Conformal Field Theory, Cambridge University Press, United Kingdom.
Fuchs, J., Schweigert, C. (1997), Symmetries, Lie Algebras, and Representations: A Graduate Course for Physicists, Cambridge University Press, Cambridge, United Kingdom.
Fujikawa, K., Suzuki, H. (2004), Path Integrals and Quantum Anomalies, Oxford University Press, Oxford.
Fukugita, M., Yanagita, T. (2003), Physics of Neutrinos and Applications to Astrophysics, Springer, Berlin.
Fulde, P. (1995), Electron Correlations in Molecules and Solids, 3rd edn., Springer, New York.

Galindo, A., Pascual, P. (1990), Quantum Mechanics, Vols. 1, 2, Springer, Berlin (translated from Spanish into English).
Gambini, R., Pulli, J. (2000), Loops, Knots, Gauge Theories and Quantum Gravity, Cambridge University Press, Cambridge, United Kingdom.
Gasquet, C., Witomski, P. (2002), Fourier Analysis and Applications, Springer, New York.
Gauß, C. (1863/1933), Werke (Collected works with commentaries), Vols. 1–12, Göttingen, Germany.
Gelfand, I., Shilov, G., Vilenkin, N. (1964), Generalized Functions, Vols. 1–5, Academic Press, New York (translated from Russian into English).
Gelfand, I. (1989), Collected Papers, Vols. 1–3, Springer, New York.
Gell-Mann, G., Low, F. (1951), Bound states in quantum field theory. Phys. Rev. **84**, 350–354.
Gell-Mann, M., Low, F. (1954), Quantum electrodynamics at small distances, Rev. Phys. **95**(5), 1300–1317.

Genz, H. (2004), Nichts als das Nichts (The vacuum energy), Wiley–VCH, Weinheim, Germany (in German).
Gerthsen, C. (2004), Gerthsen Physik. Edited by D. Meschede, 22nd edn., Springer, Berlin (in German).
Giaquinta, M., Hildebrandt, S. (1995), Calculus of Variations, Vols. 1, 2, Springer, Berlin.
Gilbarg, D., Trudinger, N. (1983), Elliptic Partial Differential Equations of Second Order, Springer, New York.
Gilkey, P. (1995), Invariance Theory, the Heat Equation, and the Atiyah–Singer Index Theorem, 2nd edn., CRC Press, Boca Raton, Florida.
Gilkey, P. (2003), Asymptotic Formulae in Spectral Geometry, Chapman, CRC Press, Boca Raton, Florida.
Giulini, D., Kiefer, C., Lämmerzahl, C. (Eds.) (2003), Quantum Gravity: From Theory to Experimental Search, Springer, Berlin.
Glimm, J., Jaffe, A. (1981), Quantum Physics: A Functional Integral Point of View, Springer, New York.
Glimm, J., Jaffe, A. (1985), Quantum Field Theory and Statistical Mechanics: Expositions, Birkhäuser, Boston.
Goldstein, H. (2002), Classical Mechanics, 3rd edn., Addison-Wesley, Reading Massachusetts.
Golse, F., Saint-Raymond, L. (2001), The Navier–Stokes limit for the Boltzmann equation, Comptes Rendus Acad. Sci. Paris, Ser. I, Math. **333**, 897–902.
Gottfried, K., Tung-Mow Yan (2003), Quantum Mechanics: Fundamentals, 2nd edn., Springer, New York.
Gottwald, S., Ilgauds, H., Schlote, K. (1990), Lexikon bedeutender Mathematiker (Biographies of important mathematicians) (in German), H. Deutsch, Frankfurt/Main.[4]
Gracia-Bondia, J., Várilly, J., Figueroa, H. (2001), Elements of Noncommutative Geometry, Birkhäuser, Boston.
Gradshtein, I., Ryshik, I. (1980), Tables of Integrals, Series, and Products, 4th edn., Academic Press, New York (translated from Russian into English).
Gray, J. (2000a), Linear Differential Equations and Group Theory: From Riemann to Poincaré, Birkhäuser, Boston.
Gray, J. (2000b), The Hilbert Challenge: A Perspective on 20th Century Mathematics, Oxford University Press, Oxford.
Green, M., Schwarz, J., Witten, E. (1987), Superstrings, Vols. 1, 2, Cambridge University Press, Cambridge, United Kingdom.
Greene, B. (1999), The Elegant Universe: Supersymmetric Strings, Hidden Dimensions, and the Quest for the Ultimate Theory, Norton, New York.
Greiner, W. et al. (1996), Course of Modern Theoretical Physics, Vols. 1–13, Springer, New York (translated from German into English).
Greiner, W., Schäfer, A. (1994), Quantum Chromodynamics, Springer, Berlin.
Greiner, W., Müller, B. (1996), Gauge Theory of Weak Interactions, 2nd edn., Springer, New York.
Greiner, W., Reinhardt, J. (1996a), Quantum Electrodynamics, 3rd edn., Springer, Berlin.
Greiner, W., Reinhardt, J. (1996b), Field Quantization, Springer, Berlin.
Greiner, W. (1997), Relativistic Quantum Mechanics: Wave Equations, Springer, Berlin.
Gribbin, J. (1998), Q is for Quantum: Particle Physics from A–Z, Weidenfeld, London.

[4] 1800 biographies

Griffith, R. (1972), Rigorous results and theorems. In: C. Domb and M. Green (Eds.), Phase Transitions and Critical Phenomena, Academic Press, New York, 1972, pp. 9–108.
Griffith, P., Harris, J. (1978), Principles of Algebraic Geometry, Wiley, New York.
Gromov, M. (1985), Pseudo-holomorphic curves in symplectic manifolds, Invent. Math. **82**, 307-347.
Grosche, C., Steiner, F. (1998), Handbook of Feynman Path Integrals, Springer, New York.[5]
Gross, D., Wilczek, F. (1973a), Ultraviolet behavior of non-Abelian gauge theories, Phys. Letters **30**(26), 1343–1346.
Gross, D., Wilczek, F. (1973b), Asymptotically free gauge theories, Phys. Rev. D **9**(4), 980–993.
Gross, F. (1993), Relativistic Quantum Mechanics and Field Theory, Wiley, New York.
Grosse, H. (1988), Models in Statistical Physics and Quantum Field Theory, Springer, New York.
Grosse, H., Wulkenhaar, R. (2005), Renormalisation of φ^4-theory on noncommutative $\mathbb{R}^4$ in the matrix base, Commun. Math. Phys. **256**, 305–374.
Guidry, M. (1991), Gauge Field Theories: An Introduction with Applications, Wiley, New York.
Guillemin, V., Pollack, A. (1974), Differential Topology, Prentice Hall, Englewood Cliffs, New Jersey.
Guillemin, V., Sternberg, S. (1989), Geometric Asymptotics, Amer. Math. Soc. Providence, Rhode Island.
Guillemin, V., Sternberg, S. (1990), Symplectic Techniques in Physics, Cambridge University Press, Cambridge, United Kingdom.
Günter, N. (1934), La théorie du potentiel et ses applications de la physique mathématique, Paris. (German edition: Potentialtheorie, Teubner, Leipzig 1957.)
Günther, P. (1988), Huygens' Principle and Hyperbolic Differential Equations, Academic Press, San Diego.
Gustafson, S., Sigal, I. (2003), Mathematical Concepts of Quantum Mechanics, Springer, Berlin.
Guth, A., Huang, K., Jaffe, A. (Eds.) (1983), Asymptotic Realms of Physics, Essays in Honor of Francis Low, MIT Press, Cambridge, Massachusetts.

Haag, R., Kastler, D. (1964), An algebraic approach to quantum field theory, J. Math. Phys. **5**, 848–861.
Haag, R. (1996), Local Quantum Physics: Fields, Particles, Algebras, 2nd edn., Springer, New York.
Hackbusch, W. (1985), Multigrid Methods and Applications, Springer, Berlin.
Hackbusch, W. (1992), Elliptic Differential Equations: Theory and Numerical Treatment, Springer, New York.
Hahn, Y., Zimmermann, W. (1968), An elementary proof of Dyson's power-counting theorem, Commun. Math. Phys. **10**, 330–342.
Haken, H. (1976), Quantum Field Theory of Solids, North-Holland Amsterdam (translated from German into English). Second German edition: Teubner, Stuttgart, 1993.
Haken, H. (1984), Laser Theory, 2nd edn. In: S. Flügge (Ed.), Handbook of Physics, Vol. XXV/2c, Springer, Berlin.

[5] 950 references

Hall, B. (2003), Lie Groups, Lie Algebras, and Representations: An Elementary Introduction, Springer, New York.
Halzen, F., Martin, A. (1984), Quarks and Leptons, Wiley, New York.
Handbook of Differential Equations (2004), Evolutionary Equations, Vols. 1, 2. Edited by C. Dafermos et al., Elsevier, Boston.
Handbook of Differential Equations (2004), Stationary Partial Differential Equations, Vols. 1, 2. Edited by M. Chipot and P. Quittner, Elsevier, Amsterdam.
Handbook of Differential Equations (2005), Ordinary Differential Equations, Vols. 1, 2. Edited by A. Canada et al., Elsevier, Amsterdam.
Handbook of Dynamical Systems (2002), Vols. 1A, 1B edited by B. Hasselblatt et al.; Vol. 2 edited by B. Fiedler, Elsevier, Amsterdam.
Handbook of Mathematical Physics: see Encyclopedia of Mathematical Physics (2006).
Harenberg Lexikon der Nobelpreisträger (2000) (Encyclopedia of Nobel Prize Laureates), Harenberg Lexikon Verlag, Dortmund, Germany (in German).
Harer, J., Zagier, D. (1986) The Euler characteristic of the moduli space of curves, Invent. Math. **85**(3), 457–495.
Hartshorne, R. (1994), Algebraic Geometry, 3rd ed., Springer, New York.
Hatcher, A. (2002), Algebraic Topology, Cambridge University Press, Cambridge, United Kingdom. Internet: `http://www.math.cornell.edu/~hatcher`
Hatcher, A. (2005a), Spectral Sequences in Algebraic Topology. Internet: `http://www.math.cornell.edu/~hatcher`
Hatcher, A. (2005b), Vector Bundles and K-Theory. Internet: `http://www.math.cornell.edu/~hatcher`
Hauser, H. (2003), The Hironaka theorem on resolution of singularities (or: a proof we always wanted to understand), Bull. Amer. Math. Soc. **40**(3), 323–403.
Hausner, M., Schwartz, J. (1968), Lie Groups and Lie Algebras (lectures held at the Courant Institute, NYU), Gordon and Breach, New York.
Havil, J. (2003), Gamma: Exploring Euler's Constant, Princeton University Press, Princeton, New Jersey.
Heath, H. (1981), A History of Greek Mathematics, Vols. 1, 2, Clarendon Press, Oxford, New York.
Hein, W. (1990), Introduction to Structure and Representation of the Classical Groups (in German), Springer, Berlin.
Heisenberg, W. (1925), Quantum-theoretical re-interpretation of kinematics and mechanical relations (in German), Z. Physik **33**, 879–893. (English translation: see van der Waerden (Ed.) (1968), pp. 261–276.)
Heisenberg, W. (1927), Intuitiver Inhalt der quantentheoretischen Kinematik und Mechanik (The intuitive meaning of quantum-theoretical kinematics and mechanics), Z. Phys. **43**, 172–199.
Heisenberg, W., Pauli, W. (1929), Zur Quantenentheorie der Wellenfelder (On quantum field theory), Z. Phys. **56**, 1–61; **59**, 108–190 (in German).
Heisenberg, W. (1930), The Physical Principles of the Quantum Theory, Chicago University Press, Chicago.
Heisenberg, W. (1943), Die "beobachtbaren Grössen" in der Theorie der Elementarteilchen (The observable quantities in particle physics), Z. Phys. **120** (1943), 513–538, 673–702; **123** (1944), 93–112 (in German).
Heisenberg, W. (1970), Physics and Beyond: Encounters and Conversations, Harper and Row, New York (translated from German into English).
Heisenberg, W. (1984), Gesammelte Abhandlungen (Collected Papers). Edited by P. Dürr et al., Vols. 1–3, Springer, Berlin, New York.
Heiss, D. (Ed.) (2002), Fundamentals of Quantum Information: Quantum Computation, Communication, Decoherence and All That, Springer, Berlin.

Heitler, W. (1936), The Theory of Radiation, Clarendon Press, Oxford.
Helffer, B. (2003), Semiclassical Analysis, World Scientific, Singapore.
Hellagouarch, Y. (2002), Invitation to the Mathematics of Fermat–Wiles, Academic Press, New York.
Henneaux, M., Teitelboim, C. (1993), Quantization of Gauge Systems, Princeton University Press, Princeton, New Jersey.
Hepp, K. (1966), Proof of the Bogoliubov–Parasiuk theorem on renormalization, Commun. Math. Phys. **2**, 301–326.
Hepp, K. (1969), La théorie de la rénormalisation, Springer, Berlin.
Hida, T. (1970), Stationary Stochastic Processes, Princeton University Press, Princeton, New Jersey.
Hida, T. (1980), Brownian Motion, Springer, New York.
Hida, T., Potthoff, J., Streit, L. (1990), White noise analysis and applications. In: Mathematics and Physics: Lectures on Recent Results, Vol. 3, pp. 143–178, World Scientific, Singapore,
Hilbert, D. (1900), Mathematical problems. Lecture delivered before the Second International Congress of Mathematicians at Paris 1900, Bull. Amer. Math. Soc. **8** (1902), 437–479.
Hilbert, D. (1912), Grundzüge einer allgemeinen Theorie der Integralgleichungen (Foundations of the theory of integral equations), Teubner, Leipzig (in German). (Reprinted version: Integralgleichungen und Gleichungen mit unendlich vielen Unbekannten. Edited by A. Pietsch (with historical comments), Teubner, Leipzig, 1989.)
Hilbert, D. (1932), Gesammelte Werke (Collected Works), Vols. 1–3, 12th edn., 1977, Springer, Berlin (in German).
Hildebrandt, S., Tromba, T. (1985), Mathematics and Optimal Form, Scientific American Books, New York.
Hirzebruch, F. (1966), Topological Methods in Algebraic Geometry, 3rd enlarged edn., Springer, New York. (translated from German into English).
Hirzebruch, F. (1987), Gesammelte Abhandlungen (Collected Papers), Vols. 1, 2, Springer, Berlin.
Hislop, P., Sigal, I. (1996), Introduction to Spectral Theory With Applications to Schrödinger Operators, Springer, New York.
Hollik, W., Kraus, E., Roth, M., Rupp, C., Sibold, K., Stöckinger, D. (2002), Renormalization of the minimal supersymmetric standard model, Nuclear Physics B **639**, 3–65.
Honerkamp, J., Römer, H. (1993), Theoretical Physics: A Classical Approach, Springer, Berlin (translated from German into English).
Hörmander, L. (1979), The Weyl calculus for pseudo-differential operators, Comm. Pure Appl. Math. **32**, 359–443.
Hörmander, L. (1983), The Analysis of Linear Partial Differential Operators, Vol. 1: Distribution Theory and Fourier Analysis, Vol. 2: Differential Operators with Constant Coefficients, Vol. 3: Pseudo-Differential Operators, Vol. 4: Fourier Integral Operators, Springer, New York.
Hoste, J., Thistlewaite, M., Weeks, J. (1998), The first 1, 701, 936 knots, Mathematical Intelligencer **20**(4), 33–48.
Huang, K. (1987), Statistical Physics, Wiley, New York.
Huang, K. (1992), Quarks, Leptons, and Gauge Fields, World Scientific, Singapore.
Huang, K. (1998), Quantum Field Theory: From Operators to Path Integrals, Wiley, New York.
Hurwitz, A., Courant, R. (1964), Vorlesungen über allgemeine Funktionentheorie und elliptische Funktionen (Lectures on complex function theory and elliptic functions), 4th edn., Springer, Berlin (in German).

Hwa, R., Teplitz, V. (1966), Homology and Feynman Diagrams, Benjamin, Reading, Massachusetts.

Iagolnitzer, D. (1993), Scattering in Quantum Field Theory, Princeton University Press, Princeton, New Jersey.
I Calle, C. (2001), Superstrings and Other Things: A Guide to Physics, Institute of Physics Publishing, Bristol.
Infeld, L. (1980), Quest: An Autobiography, 2nd edn., Chelsea, New York.
Itzykson, C., Zuber, J. (1980), Quantum Field Theory, MacGraw-Hill, New York.
Ivey, T., Landsberg, J. (2003), Cartan for Beginners: Differential Geometry via Moving Frames and Exterior Differential Systems, Amer. Math. Soc., Providence, Rhode Island.

Jackson, J. (1995), Classical Electrodynamics, 2nd edn., Wiley, New York.
Jacod, P., Protter, P. (2000), Probability Essentials, Springer, Berlin.
Jaffe, A., Taubes, C. (1980), Vortices and Monopoles: The Structure of Static Gauge Fields, Birkhäuser, Boston.
Johnson, G., Lapidus, M. (2000), The Feynman Integral and Feynman's Operational Calculus, Clarendon Press, Oxford.
Jost, J. (1994), Differentialgeometrie und Minimalflächen (Differential geometry and minimal surfaces), Springer, Berlin (in German).
Jost, J. (1997), Compact Riemann Surfaces: An Introduction to Contemporary Mathematics, Springer, Berlin.
Jost, J., Xianqing Li-Jost (1998), Calculus of Variations, Cambridge University Press, Cambridge, United Kingdom.
Jost, J. (2001), The Bosonic String: A Mathematical Treatment, International Press, Boston.
Jost, J. (2002a), Riemannian Geometry and Geometric Analysis, 3rd edn., Springer, Berlin.
Jost, J. (2002b), Partial Differential Equations, Springer, New York.
Jost, J. (2005a), Postmodern Analysis, 3rd edn., Springer, Berlin.
Jost, J. (2005b), Dynamical Systems: Examples of Complex Behavior, Springer, Berlin.
Jost, R. (1965), The Generalized Theory of Quantum Fields, Amer. Math. Soc., Providence, Rhode Island.
Jost, R. (1995), Das Märchen vom elfenbeinernen Turm: Reden und Aufsätze (The fairy tale about the ivory tower: essays and lectures in German and partly in English). Edited by K. Hepp, W. Hunziker, and W. Kohn, Springer, Berlin.

Kac, V. (1996), Vertex Algebras for Beginners, Amer. Math. Soc., Providence, Rhode Island.
Kadanoff, L. (1966), Scaling laws for Ising models near critical temperature, Physics **2**, 263–272.
Kadanoff, L. (2000), Statistical Physics, World Scientific, Singapore.
Kadison, R., Ringrose, J. (1983), Fundamentals of the Theory of Operator Algebras, Vols. 1–4, Academic Press, New York.
Kähler, E. (2003) Mathematische Werke – Mathematical Works. Edited by R. Berndt and Oswald Riemenschneider, de Gruyter, Berlin.
Kaiser, D. (2005), Drawing Theories Apart: The Dispersion of Feynman Diagrams in Postwar Physics, The University of Chicago Press, Chicago.
Kaku, M. (1999), Quantum Field Theory: A Modern Introduction, Oxford University Press, New York, 1999.
Kaku, M. (2000), Strings, Conformal Fields and M-Theory, Springer, New York.

Kalka, H., Soff, G. (1997), Supersymmetrie, Teubner, Stuttgart (in German).
Kane, G. (2000), Supersymmetry: Squarks, Photinos, and the Unveiling of the Ultimate Laws of Nature, Perseus Publishing, Cambridge, Massachusetts.
Kapusta, J. (1989), Quantum Field Theory at Finite Temperature, Cambridge University Press, Cambridge, United Kingdom.
Kassel, C., Rosso, M., Turaev, V. (1997), Quantum Groups and Knot Invariants, Société Mathématique de France, Paris.
Kastler, D. (1992), Algebraic field theory: recollections and thoughts about the future, Rev. Math. Phys. **4**, Special Issue 1 dedicated to Rudolf Haag, pp. 159–166.
Kastler, D. (2003), Rudolf Haag – eighty years, Commun. Math. Phys. **237**, 3–6.
Kato, T. (1966), Perturbation Theory for Linear Operators, Springer, Berlin.
Kaufman, L. (2001), Knots and Physics, 3rd edn., World Scientific, Singapore.
Ketov, S. (2000), Quantum Non-Linear Sigma Models: From From Quantum Field Theory to Supersymmetry, Conformal Field Theory, Black Holes, and Strings, Springer, Berlin.
Kichenassamy, S. (1993), Nonlinear Waves, Pitman, London.
Kiefer, C. (2004), Quantum Gravity, Oxford University Press, Oxford.
Kirsten, K. (2002), Spectral Functions in Mathematics and Physics, Chapman, Boca Raton, Florida.
Kittel, C. (1996), Introduction to Solid State Physics, 7th edn., Wiley, New York.
Klauder, J., Sudarshan, E. (1968), Fundamentals of Quantum Optics, Benjamin, New York.
Klauder, J., Daubchies, I. (1982), Quantum mechanical path integrals with Wiener measures for all polynomial Hamiltonians, Phys. Rev. Lett. **52** (1984), 1161–1164.
Klauder, J., Skagerstam, B. (1985), Coherent States: Applications in Physics and Mathematical Physics, World Scientific, Singapore.
Klauder, J. (1988), Quantization is geometry, after all, Ann. Phys. **188**, 120–141.
Klauder, J. (1989), The regularized Feynman path integral. In: V. Syakanit et al. (Eds.), Path Integrals from meV to MeV, Bangkok, World Scientific, Singapore, 1989, pp. 48–59.
Klauder, J., Onfri, E. (1989), Landau levels and geometric quantization, Int. J. Mod. Phys. **A4**, 3939–3949.
Klauder, J. (2000), Beyond Conventional Quantization, Cambridge University Press, Cambridge, United Kingdom.
Klein, F. (1884), Vorlesungen über das Ikosaeder und die Auflösung der Gleichungen vom fünften Grad, Teubner, Leipzig. Reprinted 1993 by Birkhäuser and supplemented with commentaries by P. Slodowy. (English edition: Lectures on the Icosahedron and the Solution of Fifth-Order Equations, Dover, reprint 1956.)
Klein, F. (1926), Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert, Vols. 1, 2, Springer-Verlag, Berlin. (English edition: Development of Mathematics in the 19th Century, with an appendix by R. Hermann, Math. Sci. Press, New York, 1979.)
Klein, C., Richter, O. (2006), Ernst Equation and Riemann Surfaces, Springer, Berlin.
Kleinert, H. (2004), Path Integrals in Quantum Mechanics, Statistics, and Polymer Physics, 3rd edn., World Scientific, River Edge, New York.
Kleinert, H., Schulte-Frohlinde, V. (2001), Critical Properties of φ^4-Theories, World Scientific, River Edge, New York.
Klein, C. and Richter, O. (2005), Ernst Equation and Riemann Surfaces: Analytical and Numerical Methods, Springer, Berlin.

Klimyk, A., Schmüdgen, K. (1997), Quantum Groups, Springer, Berlin.
Kline, M. (1972), Mathematical Thought from Ancient to Modern Times, Oxford University Press, New York.
Knabner, P., Angermann, L. (2003), Numerical Methods for Elliptic and Parabolic Partial Differential Equations, Springer, New York (translated from German into English).
Knapp, A. (1986), Representation Theory of Semi-Simple Groups, Princeton University Press, Princeton, New Jersey.
Knapp, A. (2002), Lie Groups Beyond an Introduction, 2nd edn., Birkhäuser, Boston.
Knörrer, H. (1996), Geometrie, Vieweg, Wiesbaden (in German).
Kobayashi, S., Nomizu, K. (1963), Foundations of Differential Geometry, Vols. 1, 2, Wiley, New York.
Kobzarev, I., Manin, Yu. (1989), Elementary Particles: Mathematics, Physics, and Philosophy, Kluwer, Dordrecht (translated from Russian into English).
Koch, H. (1986), Einführung in die klassische Mathematik (Introduction to classical mathematics), Akademie-Verlag, Berlin (in German).
Komech, A. (1999), Linear partial differential equations with constant coefficients. In: Egorov, Komech, and Shubin (1999), pp. 121–255.
Kontsevich, M. (2003), Deformation quantization of Poisson manifolds. Lett. Math. Phys. **66**(3), 157–216.
Internet: `http://q-alg/9709040`
Korevaar, J. (2004), Tauberian Theory: A Century of Developments, Springer, Berlin.
Koshy, T. (2001), Fibonnacci and Lucas Numbers with Applications, Wiley, New York.
Kostrikin, A., Manin, Yu. (1989), Linear Algebra and Geometry, Springer, New York (translated from Russian into English).
Kraft, F. (1999), Vorstoß ins Unbekannte: Lexikon großer Naturwissenschaftler (Encyclopedia of great scientists), Wiley, Weinheim, New York (in German).
Kragh, H. (2000), Quantum Generations: A History of Physics in the Twentieth Century, Princeton University Press, Princeton, New Jersey.
Krantz, S., Parks, H. (2002): The Implicit Function Theorem: History, Theory, and Applications, Birkhäuser, Boston.
Kraus, E. (1998), Renormalization of the electroweak standard model to all orders, Ann. Phys. (NY) **262**, 155–259.
Kreimer, D. (2000), Knots and Feynman Diagrams, Cambridge University Press, Cambridge, United Kingdom.
Kufner, A., John, O., Fučik, S. (1977), Function Spaces, Academia, Prague.
Kugo, T. (1997), Eichfeldtheorie (Gauge field theory), Springer, Berlin (translated from Japanese into German).
Kupradse, V. (1956), Randwertaufgaben der Schwingungstheorie und Integralgleichungen (Boundary-value problems for wave problems and integral equations), Deutscher Verlag der Wissenschaften, Berlin (translated from Russian into German).

Lahiri, A., Pal, B. (2001), A First Book of Quantum Field Theory, Alpha Science International, Pangbourne, India.
Landau, L., Lifshitz, E. (1982), Course of Theoretical Physics, Vol. 1: Mechanics, Vol. 2: The Classical Theory of Fields, Vol. 3: Quantum Mechanics, Vol. 4: Quantum Electrodynamics, Vol. 5: Statistical Physics, Part 1, Vol. 6: Fluid Mechanics, Vol. 7: Theory of Elasticity, Vol. 8: Electrodynamics of Contin-

uous Media, Vol. 9: Statistical Physics, Part 2, Vol. 10: Physical Kinetics, Butterworth–Heinemann, Oxford (translated from Russian into English).
Lang, S. (1972), Introduction to Algebraic Geometry, Addison-Wesley, Reading, Massachusetts.
Lang, S. (1973), Elliptic Functions, Addison-Wesley, New York.
Lang, S. (1983), Abelian Varieties, Springer, New York.
Lang, S. (1986), Algebraic Number Theory, Springer, New York.
Lang, S. (1993), Real Analysis and Functional Analysis, Springer, New York.
Lang, S. (1995a), Complex Analysis, 4th edn., Springer, New York.
Lang, S. (1995b), Introduction to Algebraic and Abelian Functions, 2nd edn., Springer, New York.
Lang, S. (1995c), Introduction to Diophantine Approximation, Springer, Berlin.
Lang, S. (1995d), Introduction to Modular Forms, 2nd edn., Springer, New York.
Lang, S. (1999), Fundamentals of Differential Geometry, Springer, New York.
Lang, S. (2002), Algebra, revised 3rd edn., Springer, New York.
Laughlin, R. (2005), A Different Universe: Reinventing Physics from the Bottom Down, Basic Books, New York.
Lax, P. (2002), Functional Analysis, Wiley, New York.
Lax, P. (2003), Jürgen Moser, 1928–1999. In: Ergodic Theory and Dynamical Systems **22**, 1337–1342.
Le Bellac, M. (1991), Quantum and Statistical Field Theory, Clarendon Press, Oxford (translated from French into English).
Lieb, E., Loss, M. (1997), Analysis, Amer. Math. Soc., Providence, Rhode Island.
Lieb, E. (2002a), The Stability of Matter: From Atoms to Stars, Selecta of Elliott Lieb. Edited by W. Thirring, 3rd edn., Springer, New York.
Lieb, E. (2002b), Inequalities, Selecta of Elliott Lieb. Edited by M. Loss, Springer, New York.
Liebscher, D. (2005), Cosmology, Springer, Berlin.
López Dávalos, A., Zanette, D. (1999), Fundamentals of Electromagnetism, Springer, Berlin.
Lurié, D. (1968), Particles and Fields, Wiley, New York.
Lüst, D., Theissen, S. (1989), Lectures on String Theory, Springer, Berlin.

Macke, W. (1959), Lehrbuch der Theoretischen Physik (Textbook on theoretical physics), Vol. 1: Teilchen (particles), Vol. 2: Wellen (waves), Vol. 3 Quanten (quanta), Vol. 4: Elektromagnetische Felder (electromagnetic fields), Vol. 5: Thermodynamik und Statistik (thermodynamics and statistics), Vol. 6: Quanten und Relativität (quanta and relativity), Geest & Portig, Leipzig (in German).
Madsen, I., Tornehave, J. (1997), From Calculus to Cohomology: de Rham Cohomology and Characteristic Classes, Cambridge University Press, Cambridge, United Kingdom.
Magnus, W., Oberhettinger, F., Soni, R. (1966), Formulae and Theorems for the Special Functions of Mathematical Physics, 3rd edn., Berlin, Springer, Berlin.
Majid, M. (1995), Foundations of Quantum Group Theory, Cambridge University Press, Cambridge, United Kingdom.
Mandel, L., Wolf, E. (1995), Optical Coherence and Quantum Optics, Cambridge University Press, Cambridge, United Kingdom.
Manin, Yu. (1978), Algebraic aspects of nonlinear differential equations, J. Soviet Math. **11**, 1–122.
Manin, Yu. (1981), Mathematics and Physics, Birkhäuser, Boston (see also Kobzarev and Manin (1989)).

Manin, Yu. (1988), Quantum Groups and Noncommutative Geometry, CRM, Université de Montréal.
Manin, Yu. (1989), Strings, Math. Intelligencer **11**(2), 59–65.
Manin, Yu. (1991), Topics in Noncommutative Geometry, Princeton University Press.
Manin, Yu., Zagier, D. (1996), Higher Weil–Peterson volumes of moduli spaces of stable n-pointed curves, Commun. Math. Phys. **181**(3), 763–787.
Manin, Yu. (1997), Gauge Fields and Complex Geometry, Springer, Berlin.
Manin, Yu. (1999), Frobenius manifolds, quantum cohomology and moduli spaces, Amer. Math. Soc., Providence, Rhode Island.
Manoukian, I. (1983), Renormalization, Academic Press, New York.
Marathe, K., Martucci, G. (1992), The Mathematical Foundations of Gauge Theories, North-Holland, Amsterdam.
Marathe, K., Martucci, G., Francaviglia, M. (1995), Gauge theory, geometry, and topology, Seminario di Matematica dell' Università di Bari **262**, 1–90.
Marsden, J. (1972), Basic Complex Analysis, Freeman, New York.
Martin, S. (1997), A supersymmetry primer. In: G. Kane, Perspectives on Supersymmetry II, World Scientific Singapore, 1997, pp. 1–98.
Masujima, M. (2000), Path Integral Quantization and Stochastic Quantization, Springer, Berlin.
Mathematical Physics (2000), Proceedings of the XIIIth International Congress on Mathematical Physics, London, United Kingdom, 2000. Edited by A. Fokal et al., Imperial College Press, River Edge, New York.
Mathematical Physics (2003), Proceedings of the XIVth International Congress on Mathematical Physics, Lisbon, Portugal, 2003. Edited by J. Zambrini, World Scientific, Singapore.
Maurin, K. (1968), Generalized Eigenfunction Expansions and Unitary Representations of Topological Groups, Polish Scientific Publishers, Warsaw.
Maurin, K. (1972), Methods of Hilbert Spaces, Polish Scientific Publishers, Warsaw (translated from Polish into English).
Maurin, K. (1976), Analysis I, II, Reidel, Boston (translated from Polish into English).
Maurin, K. (1997), Riemann's Legacy: Riemann's Ideas in Mathematics and Physics of the 20th Century, Kluwer, Dordrecht.
Mehra, J., Milton, K. (2000), Climbing the Mountain: The Scientific Biography of Julian Schwinger, Oxford University Press, New York.
Mehra, J., Rechenberg, H. (2002), The Historical Development of Quantum Mechanics, Vols. 1–6, Springer, New York.
Melic, B., Passek-Kummericki, K., Trampetic, J., Schupp, P., Wohlgenannt, M. (2005a), The standard-model on noncommutative space-time: electroweak currents and Higgs sector. Internet: `http://hep-ph/0502249`
Melic, B., Passek-Kummericki, K., Trampetic, J., Schupp, P., Wohlgenannt, M. (2005b), The standard-model on noncommutative space-time: strong interactions included. Internet: `http://hep-ph/0503064`
Milnor, J. (1963), Morse Theory, Princeton University Press, Princeton, New Jersey.
Milnor, J. (1965), Topology from the Differentiable Viewpoint, Virginia University Press, Charlottesville.
Milnor, J., Stasheff, J. (1974), Characteristic Classes, Princeton University Press, Princeton, New Jersey.
Milnor, J. (2000), Dynamics in One Complex Variable: Introductory Lectures, 2nd edn., Vieweg, Wiesbaden.

Milonni, P. (1994), The Quantum Vacuum: An Introduction to Quantum Electrodynamics, Academic Press, Boston.
Milton, K. (2001), The Casimir Effect: Physical Manifestations of Zero-Point Energy, World Scientific, Singapore.
Minlos, R. (2000), Introduction to Mathematical Statistical Physics, Amer. Math. Soc., Providence, Rhode Island.
Misner, C., Thorne, K., Wheeler, A. (1973), Gravitation, Freeman, San Francisco.
Mitra, A. (Ed.) (2000), Quantum Field Theory: A 20th Century Profile, Indian National Science Academy, Hindustan Book Agency, India.
Moore, J. (1996), Lectures on Seiberg–Witten Invariants, Springer, Berlin.
Morii, T., Lim, C., Mukherjee, S. (2004), The Physics of the Standard Model and Beyond, World Scientific, Singapore.
Morse, P., Feshbach, H. (1953), Methods of Theoretical Physics, Vols. 1, 2, McGraw-Hill, New York.
Mostepanenko, V., Trunov, N. (1997), The Casimir Effect and its Applications, Clarendon Press, Oxford.
Müller, S. (1999), Variational Models for Microstructure and Phase Transitions, In: F. Bethuel, S. Hildebrandt, and M. Struwe (Eds.) Calculus of Variations and Geometric Evolution Problems, Cetraro 1996, pp. 85–210, Springer, Berlin. `http://www.mis.mpg.de/preprints/ln/lecturenote-0298-abstr.html`
Mumford, D. (1975), Curves and their Jacobians, The University of Michigan Press, Ann Arbor.
Mumford, D. (1977), Stability of projective varieties, L'Enseign. Math. **23**, 39–110.
Mumford, D. (1982), Geometric Invariant Theory, Springer, Berlin.
Mumford, D. (1983), Tata Lectures on Theta I, II, Birkhäuser, Basel.

Naber, G. (1980), Topological Methods in Euclidean Spaces, Cambridge University Press, Cambridge, United Kingdom.
Naber, G. (1988), Space-Time and Singularities, Cambridge University Press, Cambridge, United Kingdom.
Naber, G. (1992), The Geometry of Minkowski Space-Time, Springer, New York.
Naber, G. (1997), Topology, Geometry, and Gauge Fields, Springer, New York.
Nachtmann, O. (1990), Elementary Particle Physics: Concepts and Phenomena, Springer, Berlin (translated from German into English).
Nahm, W., Craigie, N., Goddard, P. (1982), Monopoles in Quantum Field Theory, World Scientific, Singapore.
Nakahara, M. (1990), Geometry, Topology and Physics, Adam Hilger, Bristol.
NASA home page, Internet:`http://www.nasa.gov/home/`
Nash, C. (1978), Relativistic Quantum Fields, Academic Press, New York.
Newman; D. (1980), Simple analytic proof of the prime number theorem, American Math. Monthly **87**, 693–696.
Nielsen, M., Chuang, I. (2001) Quantum Computation and Quantum Information, Cambridge University Press, Cambridge, United Kingdom.
Nirenberg, L. (2001), Topics in Nonlinear Functional Analysis, American Mathematical Society, Providence, Rhode Island.
Noack, M., May, R. (2000), Virus Dynamics: Mathematical Principles of Immunology and Virology, Oxford University Press, Oxford.
Nobel Prize Lectures (1954ff), Nobel Foundation, Stockholm.
Nobel prize laureates. Internet: `www.nobel.se/physics/laureats/index.html` (see also Harenberg (2000) and Dardo (2004)).
Nolting, W. (2002), Grundkurs Theoretische Physik (Course in theoretical physics), Vol. 1: Klassische Mechanik, Vol. 2: Analytische Mechanik, Vol. 3: Elektrodynamik, Vol. 4: Spezielle Relativitätstheorie, Thermodynamik, Vol. 5/1

und Vol. 5/2 : Quantenmechanik, Vol. 6: Statistische Physik, Vol. 7: Vielteilchentheorie, Vieweg, Wiesbaden (in German).
Novikov, S. Manakov, S. Pitaevskii, L Zakharov, V. (1984), Solitons: The Inverse Scattering Method, Consultant Bureau, New York (translated from Russian into English).
Novikov, S. (1996), Topology I: General Survey, Encyclopedia of Mathematical Sciences, Vol. 12, Springer, Berlin (translated from Russian into English).
Novozhilov, Yu. (1973), Introduction to Elementary Particle Physics, Pergamon Press, Oxford (translated from Russian into English).

Oberguggenberger, M. (1992), Multiplication of Distributions and Applications to Partial Differential Equations, Longman, Harlow, United Kingdom.
Ocampo, H., Paycha, S., Vargas, A. (Eds.) (2005), Geometric and Topological Methods for Quantum Theory, Springer, Berlin.
Olver, P. (1993), Applications of Lie Groups to Differential Equations, Springer, New York.
Olver, P. (1995), Equivalence, Invariants, and Symmetry, Cambridge University Press, Cambridge, United Kingdom.
Olver, P. (1999), Classical Invariant Theory, Cambridge University Press, Cambridge, United Kingdom.
Omnès, R. (1994), The Interpretation of Quantum Mechanics, Princeton University Press, Princeton, New Jersey.
Ortner, N., Wagner, P. (1997), A survey on explicit representation formulas for fundamental solutions of partial differential equations, Acta Applicandae Mathematicae **47**, 101–124.
Osterwalder, K., Schrader, R. (1973), Axioms for Euclidean Green's functions I, II, Commun. Math. Phys. **31** (1973), 83–112; **42** (1975), 281–305.

Pais, A. (1982), Subtle is the Lord: the Science and the Life of Albert Einstein, Oxford University Press, Oxford.
Pandey, J. (1996), The Hilbert Transform of Schwartz Distributions and Applications, Wiley, New York.
Parameswaran Nair, V. (2005), Quantum Field Theory: A Modern Perspective, Springer, New York.
Particle Data Group. Internet: `http://pdg.lbl.gov`
Patterson, S. (1995), An Introduction to the Theory of the Riemann Zeta Function, Cambridge University Press, Cambridge, United Kingdom.
Pauli, W. (1921), Relativitätstheorie (in German). In: Enzyklopädie der Mathematischen Wissenschaften II, 539–775, Teubner, Leipzig. New edition: Springer, Berlin 2000. (English enlarged edition: Theory of Relativity, Pergamon Press, London, 1958.)
Pauli, W. (1933), Die allgemeinen Prinzipien der Wellenmechanik. In: Handbuch der Physik, Vol. XXIV, edited by H. Geiger and K. Scheel. Reprinted in: Handbuch der Physik, Vol. V, Part I, edited by S. Flügge, Springer, Berlin 1933, 1958 (classic survey article on quantum mechanics).
Pauli, W. (1941), Relativistic field theory of elementary particles, Rev. Modern Physics **13**, 203–204 (classic survey article).
Penrose, R. (2004), The Road to Reality: A Complete Guide to the Laws of the Universe, Jonathan Cape, London.
Peres, A. (1993), Concepts and Methods in Quantum Mechanics, Kluwer, Dordrecht.
Perkins, D. (1987), Introduction to High Energy Physics, Addison Wesley, Menlo Park, California.

Peskin, M., Schroeder, D. (1995), An Introduction to Quantum Field Theory, Addison-Wesley, Reading, Massachusetts.
Piguet, O., Sorella, S. (1995), Algebraic Renormalization: Perturbative Renormalization, Symmetries, and Anomalies. Springer, Berlin, 1995.
Pike, E., Sarkar, S. (1995), The Quantum Theory of Radiation, Clarendon Press, Oxford.
Planck, M. (1948), Wissenschaftliche Selbstbiographie, 5th edn., Barth, Leipzig, 1948/1970. (English edition: Scientific Autobiography, Philosophical Library, New York, 1949.)
Planck, M. (2000), see: Quantum Theory Centenary.
Planck, M. (2001), Vorträge, Reden, Erinnerungen (Lectures, speeches, and recollections). Edited by H. Roos, Springer, Berlin.
Polchinski, J. (1984), Renormalization and effective Lagrangians, Nucl. Phys. B **231**, 269–295.
Poincaré Seminar 2002, Vacuum Energy – Renormalization. Edited by B. Duplantier and V. Rivasseau, Birkhäuser, Basel.
Poincaré Seminar 2003, Bose–Einstein Condenstation/Entropy. Edited by J. Dalibard, B. Duplantier, and V. Rivasseau, Birkhäuser, Basel.
Poincaré Seminar 2004, The Quantum Hall Effect. Edited by B. Douçot, B. Duplantier, V. Rivasseau, and V. Pasquier, Birkhäuser, Basel.
Polchinski, J. (1998), String Theory, Vols. 1, 2, Cambridge University Press, Cambridge, United Kingdom.
Politzer, D. (1973), Reliable perturbative results for strong interactions? Phys. Rev. **30**(2), 1346–1083.
Pontryagin, L. (1965), Foundations of Combinatorial Topology, 5th edn., Rochester, New York.
Pontryagin, L. (1966), Topological Groups, 2nd edn., Gordon and Breach.
Prudnikov, A. Bryhkov, Yu., Manichev, O. (1986), Integrals and Series, Vols. 1–5, Gordon and Breach, New York (translated from Russian into English).

Quantum Theory Centenary (2000), Proceedings of the Symposia Quantum Theory Cetenary, December 2000 in Berlin, Ann. Physik (Leipzig) **9** (2000), 11/12; **10** (2001), 1/2.

Radzikowski, M. (1996), Micro-local approach to the Hadamard condition in quantum field theory on curved space-time, Commun. Math. Phys. **179**, 529–553.
Ramis, J. (1993), Séries divergentes et théories asymptotiques (in French), Société Mathématique de France, Paris.
Ramond, P. (1990), Field Theory: A Modern Primer, Addison-Wesley, Reading, Massachusetts.
Randall, L., Sundrum, R. (1999), Large mass hierarchy from a small extra dimension, Phys. Rev. Lett. **83**, 4690–4693.
Reed, M., Simon, B. (1972), Methods of Modern Mathematical Physics, Vol 1: Functional Analysis, Vol. 2: Fourier Analysis and Self-Adjointness, Vol. 3: Scattering Theory, Vol. 4: Analysis of Operators (Perturbation Theory), Academic Press, New York.
Regis, E. (1989), Who Got Einstein's Office? Eccentricity and Genius at the Institute for Advanced Study in Princeton, Addison-Wesley, Reading, Massachusetts.
Reid, C. (1970), Hilbert, Springer, New York.
Reid, C. (1976), Courant: the Life of an Improbable Mathematician, Springer, New York.

Remmert, R. (1991), Theory of Complex Functions, Springer (translated from German into English).
Remmert, R. (1998), Classical Topics in Complex Function Theory, Springer, New York (translated from German into English).
Rendall, A. (1994), The Newtonian limit for asymptotically flat solutions of the Vlasov–Einstein system, Commun. Math. Phys. **163**, 89–112.
Rendall, A. (1998), Lectures on Nonlinear Hyperbolic Differential Equations (in German), Max-Planck Institute Albert Einstein for Gravitational Physics, Golm/Potsdam, Germany.
Internet: `http://www.aei-potsdam.mpg.de/rendall/vorlesung1.htlm`
Renn, J. (Ed.) (2005a), Einstein's Annalen Papers: The Complete Collection 1901–1922, Wiley-VCH, Weinheim, Germany.
Renn, J. (Ed.) (2005b), Albert Einstein – Engineer of the Universe: 100 Authors for Einstein – Essays, Wiley-VCH, Weinheim, Germany.
Rieffel, M. (1993), Deformation Quantization for Actions of $\mathbb{R}^d$, Mem. Amer. Math. Soc. **106**.
Riemann, B. (1990), Collected Mathematical Works (with commentaries), Springer, New York, and Teubner, Leipzig.
Riesz, F., Nagy, B. (1975), Functional Analysis, Frederyck Ungar, New York (translated from French into English).
Rivasseau, V. (1991), From Perturbative to Constructive Renormalization, Princeton University Press, Princeton, New Jersey.
Roepstorff, G. (1996), Path Integral Approach to Quantum Physics, Springer, New York.
Römer, H. (2005), Theoretical Optics: An Introduction, Wiley-VHC, New York (translated from German into English).
Rogers L., Williams, D. (2000), Diffusion, Markov Processes, and Martingales, Vols., 1, 2, Cambridge University Press, Cambridge, United Kingdom.
Rohlf, R. (1994), Modern Physics from α to Z^0, Wiley, New York.
Rosner, J. (2003), Resource letter SM-1: The standard model and beyond, Am. J. Phys. **71**, 302–318. Internet:`http://arXiv:hep-ph/0206176`
Rothe, H. (2005), Lattice Gauge Theories, 3rd edn., World Scientific Singapore.
Rozanov, Yu. (1969), Introductory Probability Theory, Prentice-Hall, Englewood Cliffs, New Jersey (translated from Russian into English).
Ruelle, D. (1969), Statistical Mechanics: Rigorous Results, Benjamin, New York.
Ryder, L. (1999), Quantum Field Theory, Cambridge University Press, United Kingdom.

Sakurai, J. (1967), Advanced Quantum Mechanics, Reading, Massachusetts.
Sakurai, J., Tuan, S. (1994), Modern Quantum Mechanics, Benjamin, New York.
Salam, A., (Ed.) (1968), From a Life of Physics. Evening Lectures at the International Center for Theoretical Physics, Trieste, Italy, with outstanding contributions by Abdus Salam, Hans Bethe, Paul Dirac, Werner Heisenberg, Eugene Wigner, Oscar Klein, and Eugen Lifshitz, International Atomic Energy Agency, Vienna, Austria.
Salmhofer, M. (1999), Renormalization: An Introduction, Springer, Berlin.
Sauvigny, F. (2004), Partielle Differentialgleichungen der Geometrie und der Physik (Partial differential equations in geometry and physics), Vol. 1: Grundlagen und Integraldarstellungen (basics and integral representations of the solutions), Vol. 2: Funktionalanalytische Lösungsmethoden (functional analytic methods), Springer, Berlin (in German).
Scharf, G. (1994), From Electrostatics to Optics, Springer, Berlin.

Scharf, G. (1995), Finite Quantum Electrodynamics: the Causal Approach, 2nd edn., Springer, New York.
Scharf, G. (2001), Quantum Gauge Theories: A True Ghost Story, Wiley, New York.
Schechter, M. (1982), Operator Methods in Quantum Mechanics, North-Holland, Amsterdam.
Schechter, M. (1986), Spectra of Partial Differential Operators, North-Holland, Amsterdam.
Scheck, F. (2000), Theoretical Physics, Vol. 1: Mechanics: From Newton's Law to Deterministic Chaos, 4th edn., Vol. 2: Non-Relativistic Quantum Theory: From the Hydrogen Atom to Many-Particle Systems, Vol. 3: Classical Field Theory: From Electrodynamics to Gauge Theories, Vol. 4: Quantized Fields: From Symmetries to Quantum Electrodynamics, Vol. 5: Theory of Heat: From the Fundamental Laws in Thermodynamics to Quantum Statistics (to appear), Springers, Berlin. (Vols. 2ff are written in German; the English translations of these volumes are in preparation.)
Scheck, F., Wend, W., Upmeier, H. (Eds.) (2002), Noncommutative Geometry and the Standard Model of Elementary Particle Physics, Springer, Berlin.
Schlichenmaier, M. (1989), An Introduction to Riemann Surfaces, Algebraic Curves, and Moduli Spaces, Springer, New York.
Schmüser, P. (1995), Feynman Graphs and Gauge Theories for Experimental Physicists (in German), 2nd edn., Springer, Berlin.
Schulmann, L. (1981), Techniques and Applications of Path Integration, Wiley, New York.
Schuster, P. (1994), Deterministic Chaos: An Introduction, Physik-Verlag, Weinheim, Germany.
Schutz, B. (1980), Geometrical Methods of Mathematical Physics, Cambridge University Press, Cambridge, United Kingdom.
Schutz, B. (2003), Gravity from the Ground Up, Cambridge University Press, Cambridge, United Kingdom.
Schwabl, F. (2000), Quantum Mechanics, 3rd edn., Springer, Berlin.
Schwabl, F. (2002), Statistical Mechanics, Springer, Berlin.
Schwabl, F. (2003), Advanced Quantum Mechanics, 3rd edn., Springer, Berlin.
Schwartz, L. (1965), Méthodes mathématiques pour les sciences physiques, Hermann, Paris (in French).
Schwartz, L. (1978), Théorie des distributions, 3rd edn., Vols. 1, 2, Hermann, Paris (in French).
Schwarz, A. (1993), Quantum Field Theory and Topology, Springer, Berlin (translated from Russian into English).
Schwarz, A. (1994), Topology for Physicists, Springer, Berlin (translated from Russian into English).
Schwarz, J. (Ed.) (1985), Superstrings: The First 15 Years of Superstring Theory, Vols. 1, 2, World Scientific, Singapore.
Schweber, S. (1994), QED (Quantum Electrodynamics) and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga, Princeton University Press, Princeton, New Jersey.[6]
Schwinger, J. (Ed.) (1958), Quantum Electrodynamics: 34 Selected Articles, Dover, New York.
Schwinger, J. (1970), Particles, Sources, and Fields, Vols. 1–3, Addison-Wesley, Reading, Massachusetts.
Schwinger, J. (1979), Selected Papers. Edited by M. Flato et al., Reidel, Dordrecht.

[6] 1300 references

Schwinger, J. (1998), Classical Electrodynamics, Perseus Books, Reading, Massachusetts.
Schwinger, J. (2001), Quantum Mechanics, Springer, New York.
Scriba, C., Schreiber, P. (2003), 5000 Jahre Geometrie: Geschichte, Kulturen, Menschen (5000 years of geometry), Springer, Berlin (in German) (see also Alten et al. (2003), 4000 years of algebra).
Seiberg, N., Witten, E. (1994a), Electric-magnetic duality, monopole condensation, and confinement in $N = 2$ supersymmetric Yang–Mills theory, Nuclear Phys. B **426**, 19–52.
Seiberg, N., Witten, E. (1994b), Monopoles, duality and chiral symmetry breaking in $N = 2$ supersymmetric QCD, Nucl. Physics B **431**, 485–550.
Seiden, A. (2005), Particle Physics: A Comprehensive Introduction, Addison-Wesley, San Francisco.
Séminaire Poincaré 2002ff, Institut Henri Poincaré, Paris. See also Poincaré Seminar 2002ff. Internet: `http://www.juisseux.fr.poincare`
Sen Hu (1999), Witten's Lectures on Three-Dimensional Topological Quantum Field Theory, World Scientific Singapore.
Shafarevich, I. (1990), Algebra I: Basic Notions of Algebra, Encyclopedia of Mathematical Sciences, Vol. 11, Springer, Berlin (translated from Russian into English).
Shafarevich, I. (1994), Basic Algebraic Geometry, Vols. 1, 2, 2nd edn., Springer, Berlin (translated from Russian into English).
Shiryaev, A. (1996), Probability, 6th edn., Springer, New York (translated from Russian into English).
Shohat J., Tamarkin, J. (1950), The Problem of Moments, American Mathematical Society, New York.
Shore, S. (2003), The Tapestry of Modern Astrophysics, Wiley, New York.
Shuir, Ya. (2005), Magnetic Monopoles, Springer, Berlin.
Sibold, K. (2001), Theorie der Elementarteilchen (Theory of elementary particles), Teubner, Stuttgart (in German).
Siegel, C., Moser, J. (1971), Lectures on Celestial Mechanics, Springer, Berlin.
Simon, B. (1971), Quantum Mechanics for Hamiltonians Defined on Quadratic Forms, Princeton University Press, Princeton, New Jersey.
Simon, B. (1974), The $P(\varphi)_2$-Euclidean Quantum Field Theory, Princeton University Press, Princeton, New Jersey.
Simon, B. (1979), Functional Integration and Quantum Physics, Academic Press, New York.
Simon, B. (1984), Fifteen problems in mathematical physics. In: W. Jäger (Ed.), Perspectives in Mathematics, pp. 423–454, Birkhäuser, Basel.
Simon, B. (1993), The Statistical Theory of Lattice Gases, Princeton University Press, Princeton, New Jersey.
Simon, B. (1996), Representations of Finite and Compact Groups, Amer. Math. Soc., Providence, Rhode Island.
Sinai, Ya. (1982), Theory of Phase Transitions: Rigorous Results, Pergamon Press, Oxford (translated from Russian into English).
Singh, S. (1997), Fermat's Last Theorem: The Story of a Riddle that Confounded the World's Greatest Minds for 358 Years, Fourth Estate, London.
Smirnov, V. (1964), A Course of Higher Mathematics, Vols. 1–5, Pergamon Press, New York (translated from Russian into English).
Smirnov, V. (2004), Evaluating Feynman Integrals, Springer, Berlin.
Smoller, J. (1994), Shock Waves and Reaction-Diffusion Equations, 2nd edn., Springer, New York.

Sommerfeld, A. (1949), Lectures on Theoretical Physics, Vol. 1: Mechanics, Vol. 2: Mechanics of Continua, Vol. 3: Electrodynamics, Vol. 4: Thermodynamics, Vol. 5: Quantum Mechanics, Vol. 6: Partial Differential Equations in Mathematical Physics, Academic Press, New York (translated from German into English).
Spanier, E. (1989), Algebraic Topology, 2nd. edn., Springer, New York.
Stein, E., Shakarchi, R. (2003), Princeton Lectures in Analysis. I: Fourier Analysis, II: Complex Analysis, III: Measure Theory, IV: Selected Topics, Princeton University Press, Princeton, New Jersey.
Steinmann, O. (1971), Perturbation Expansions in Axiomatic Field Theory, Springer, Berlin.
Steinmann, O. (2000), Perturbative Quantum Electrodynamics and Axiomatic Field Theory, Springer, Berlin.
Stephani, H. (1989): Differential Equations: Their Solution Using Symmetries, Cambridge University Press, Cambridge, United Kingdom (translated from German into English).
Stephani, H. (2003), Exact Solutions of Einstein's Field Equations, 2nd edn., Cambridge University Press, Cambridge, United Kingdom (translated from German into English).
Stone, M. (2000), The Physics of Quantum Fields, Springer, New York.
Straumann, N. (2002), Ein Grundkurs über nichtrelativistische Quantentheorie (Quantum mechanics: a basic course of non-relativistic quantum theory), Springer, Berlin (in German).
Straumann, N. (2004), General Relativity with Applications to Astrophysics, Springer, New York.
Straumann, N. (2005), Relativistische Quantentheorie: Eine Einführung in die Quantenfeldtheorie (Relativistic quantum theory: an introduction to quantum field theory), Springer, Berlin (in German).
Streater, R., Wightman, A. (1968), PCT, Spin, Statistics, and All That, 2nd edn., Benjamin, New York.
Stroke, H. (Ed.) (1995), The Physical Review: The First Hundred Years – A Selection of Seminal Papers and Commentaries, American Institute of Physics, New York.[7]
Stueckelberg, E., Rivier, D. (1950), Causalité et structure de la matrice S, Helv. Phys. Acta **23**, 215–222.
Stueckelberg, E., Petermann, A. (1953), La normalisation des constantes dans la théorie des quanta, Helv. Phys. Acta **26**, 499–520.
Sudberry, A. (1986), Quantum Mechanics and the Particles of Nature, Cambridge University Press, Cambridge, United Kingdom.
Sweedler, W. (1969), Hopf Algebras, Benjamin, New York.
Szabo, R. (2000), Equivariant Cohomology and Localization of Path Integrals, Springer, Berlin.
Szabo, R. (2004), An Introduction to String Theory and D-Brane Dynamics, Imperial College Press, London.

Taschner, R. (2005), Der Zahlen gigantische Schatten: Mathematik im Zeichen der Zeit (The role of mathematics in the history of human culture), Vieweg, Wiesbaden (in German).
Taylor, J. (1972), Scattering Theory, Wiley, New York.
Taylor, M. (1997), Partial Differential Equations, Vols. 1–3, Springer, New York.

[7] 14 survey articles on general developments, 200 fundamental articles, and 800 additional articles on CD

't Hooft, G., Veltman, M. (1973), Diagrammar, CERN, Report 73/9.
Internet: http://doc.cern.ch/yellowrep/1973/1973-009/p1.pdf
Thirring, W. (1997), Classical Mathematical Physics: Dynamical Systems and Fields, 3rd edn., Springer, New York (translated from German into English).
Thirring, W. (2002), Quantum Mathematical Physics: Atoms, Molecules, and Large Systems, 2nd edn., Springer, New York (translated from German into English).
Thomas, C. (2004), Representations of Finite and Lie Groups, Imperial College Press, London.
Thorne, K. (1993), Black Holes & Time Warps: Einstein's Outrageous Legacy, Norton, New York.
Thouless, D. (1961), The Quantum Mechanics of Many-Body Systems, Academic Press, New York.
Thouless, D. (Ed.) (1998), Topological Quantum Numbers in Non-Relativistic Physics, World Scientific, Singapore (collection of 40 important articles on superfluidity, quantum Hall effect, phase transitions, etc.)
Tian Yu Cao (1998), Conceptual Developments of 20th Century Field Theories, Cambridge University Press, Cambridge, United Kingdom.
Tian Yu Cao (Ed.) (1999), Conceptual Foundations of Quantum Field Theory (with contributions made by leading physicists), Cambridge University Press, Cambridge, United Kingdom
Tipler, P. (1999), Physics for Scientists and Engineers, 4th edn., Freeman, New York.
Titchmarsh, E. (1946), Eigenfunction Expansions Associated with Second-Order Differential Equations, Vols. 1, 2, Clarendon Press, Oxford.
Titchmarsh, E. (1967), Introduction to the Theory of Fourier Integrals, Clarendon Press, Oxford.
Titchmarsh, E., Heath-Brown, D. (1986), The Theory of the Riemann Zeta-Function (first edition, 1930), Cambridge University Press, Cambridge, United Kingdom.
Todorov, I. (1971), Analytic Properties of Feynman Diagrams in Quantum Field Theory, Pergamon Press, London.
Tomonaga, S. (1971), Scientific Papers, Vols. 1, 2. Edited by T. Miyazima, Tokyo, Japan.
Triebel, H. (1987), Analysis and Mathematical Physics, Kluwer, Dordrecht (translated from German into English).
Triebel, H. (1989), Higher Analysis, Barth, Leipzig (translated from German into English).
Triebel, H. (1992), Theory of Function Spaces, Birkhäuser, Basel.
Tsvelik, A. (2003), Quantum Field Theory in Condensed Matter Physics, 2nd edn., Cambridge University Press, Cambridge, United Kingdom.
Turaev, V. (1994), Quantum Invariants of Knots and 3-Manifolds, de Gruyter, Berlin.

van Baal, P. (2000), A Course in Field Theory,
Internet: http://rulgm4.leidenuniv.nl/van-baal/FTcourse.html
van der Waerden, B. (1930), Moderne Algebra (in German), Vols. 1, 2, 8th edn. 1993, Springer, New York (English edition: Frederyck Ungar, New York 1975).
van der Waerden, B. (1932), Group Theory and Quantum Mechanics, Springer, New York 1974 (translated from the 1932 German edition into English).
van der Waerden, B. (1939), Introduction to Algebraic Geometry, 2nd edn. 1973, Springer, Berlin (in German).

van der Waerden, B. (Ed.) (1968), Sources of Quantum Mechanics 1917–1926, Dover, New York.
van der Waerden, B. (1984), A History of Algebra: From al-Khwarimzi to Emmy Noether, Springer, New York.
Veltman, M. (1995), Diagrammatica: the Path to Feynman Diagrams, Cambridge University Press, Cambridge, United Kingdom.
Veltman, M. (2003), Facts and Mysteries in Elementary Particle Physics, World Scientific, Singapore.
Verch, R. (2004), The current status of quantum fields in curved space time. Lecture held on the occasion of the 125th anniversary of Einstein's birth. Deutsche Physikalische Gesellschaft, Ulm, 2004. Preprint of the Max-Planck Institute for Mathematics in the Sciences, Leipzig.
Internet:`http://www.mis.mpg.de/preprints/`
Vilenkin, N., Klimyk, A. (1991), Special Functions and Representations of Lie Groups, Vols. 1–4, Kluwer, Dordrecht (translated from Russian into English).
Vladimirov, V. (1966), Methods of the Theory of Many Complex Variables, MIT Press, Cambridge, Massachusetts (translated from Russian into English).
Vladimirov, V. (1971), Equations of Mathematical Physics, Marcel Dekker, New York (translated from Russian into English).
von Neumann, J. (1947), The mathematician. In: The Works of the Mind, Vol. 1, pp. 180–196. Edited by R. Heywood, University of Chicago Press.
von Neumann, J. (1955), Mathematical Foundations of Quantum Mechanics, Princeton University Press, New Jersey (translated from the 1932 German edition into English).
von Neumann, J. (1961), Collected Papers, Vols. 1–5, Pergamon Press, New York.

Wald, R. (1992), Space, Time, and Gravity: The Theory of the Big Bang and Black Holes, University of Chicago Press, Chicago.
Waldschmidt, M., Moussa, P., Luck, L., Itzykson, C. (Eds.) (1995), From Number Theory to Physics, 2nd edn., Springer, New York.
Wasow, W. (1965), Asymptotic Expansions for Ordinary Differential Operators, Interscience, New York.
Weil, A. (1974), Basic Number Theory, Springer, Berlin.
Weinberg, S. (1960), High energy behavior in quantum field theory, Phys. Rev. **118**, 838–849.
Weinberg, S. (1972), Gravitation and Cosmology, Wiley, New York.
Weinberg, S. (1977), The First Three Minutes: A Modern View of the Origin of the Universe, Basic Books, New York.
Weinberg, S. (1983), The Discovery of Subatomic Particles, Scientific American Library, New York.
Weinberg, S. (1992), Dreams of a Final Theory, Pantheon Books, New York.
Weinberg, S. (1995), Quantum Field Theory, Vols. 1–3, Cambridge University Press, Cambridge, United Kingdom.
Wess, J., Bagger, J. (1991), Supersymmetry and Supergravity, 2nd edn., Princeton University Press, Princeton, New Jersey.
Weyl, H. (1910), On ordinary differential equations with singularities (in German), Math. Ann. **68**, 220–269.
Weyl, H. (1911), On the asymptotic distribution of eigenvalues (in German), Göttinger Nachr. 1911, 110–117.
Weyl, H. (1913), Die Idee der Riemannschen Fläche (in German). Teubner, Leipzig. New edition with commentaries supervised by R. Remmert, Teubner, Leipzig, 1997. (English edition: The Concept of a Riemann Surface, Addison Wesley, Reading, Massachusetts, 1955.)

Weyl, H. (1918), Raum, Zeit, Materie (in German), 8th edn. 1993, Springer, Berlin. (English edition: Space, Time, Matter, Dover, New York, 1953.)
Weyl, H., Peter, F. (1927), On the completeness of the irreducible representations of compact continuous groups (in German), Math. Ann. **97**, 737–755.
Weyl, H. (1929), Gruppentheorie und Quantenmechanik (in German), Springer, Berlin. (English edition: Theory of Groups and Quantum Mechanics, Dover, New York, 1931.)
Weyl, H. (1938), The Classical Groups: Their Invariants and Representations, Princeton University Press, Princeton, New Jersey (8th edn. 1973).
Weyl, H. (1940), The method of orthogonal projection in potential theory, Duke Math.J. **7**, 414–444.
Weyl, H. (1943), On Hodge's theory of harmonic integrals, Ann. of Math. **44**, 1–6.
Weyl, H. (1944), David Hilbert and its mathematical work, Bull. Amer. Math. Soc. **50**, 612–654.
Weyl, H. (1949), Philosophy of Mathematics and Natural Sciences, Princeton University Press, Princeton, New Jersey.
Weyl, H. (1952), Symmetry, Princeton University Press, Princeton, New Jersey.
Weyl, H. (1968), Gesammelte Werke (Collected Works), Vols. 1–4, Springer, New York.
Whittaker, E., Watson, G. (1979), A Course of Modern Analysis, Cambridge University Press and MacMillan Company, London, 1944; American Mathematical Society, Providence, Rhode Island, 1979.
Widder, D. (1944), The Laplace Transform, Princeton University Press, Princeton, New Jersey.
Wigner, E. (1939), On unitary representations of the inhomogeneous Lorentz group, Ann. of Math. **40**, 149–204.
Wigner, E. (1959), Group Theory and its Applications to the Quantum Mechanics of Atomic Spectra, Academic Press, New York. (German edition: Springer, Berlin, 1931.)
Wigner, E. (1993), Collected Works, Vols. 1ff. Edited by J. Mehra and A. Whightman, Springer, New York.
Wigner, E. (1995), Philosophical Reflections and Syntheses. Annotated by G. Emch. Edited by J. Mehra and A. Wightman, Springer, New York.
Wilson, K., Kogut, J. (1974), The renormalization group and the ε-expansion, Physics Reports **12C**, 75–199.
Witten, E. (1981), A new proof of the positive energy theorem, Commun. Math. Phys. **80**, 381–396.
Witten, E. (1982), Supersymmetry and Morse theory, J. Diff. Geom. **17**, 661–692.
Witten, E. (1988a), Geometry and quantum field theory, Proc. AMS Centennial Symposium, 1988. In: Atiyah and Iagolnitzer (Eds.) (2003), pp. 523–535.
Witten, E. (1988b), Topological quantum field theory, Commun. Math. Phys. **117**, 353–386.
Witten, E. (1988c), Topological sigma models, Commun. Math. Phys. **118**, 411–449.
Witten, E. (1989), Quantum field theory and the Jones polynomials, Commun. Math. Phys. **212**, 359–399.
E. Witten, On the structure of the topological phase of two-dimensional gravity, Nucl. Phys. B **340** (1990), 281–332.
E. Witten, Two-dimensional gravity and intersection theory on moduli spaces, Surv. Diff. Geometry **1** (1991), 243–310.
Witten, E. (1992), Geometry and quantum field theory, AMS Centennial Symposium 1988, Amer. Math. Soc. 1992. Reprinted in Atiyah and Iagolnitzer (2003), pp. 523–535.

Witten, E. (1994), Monopoles and four-manifolds, Math. Research Letters **1**, 769–796.
Witten, E. (1996), Reflections on the fate of space-time, Physics Today **49**, April 1996, 24–30.
Witten, E. (1997), Duality, space-time, and quantum mechanics, Physics Today **50**, May 1997, 28–33.
Witten, E. (1998a), Magic, mystery and matrix (Gibbs Lecture), Notices Amer. Math. Soc. **45**, 1124–1129.
Witten, E. (1998b), New perspectives in the quest for unification. Internet: `arXiv:hep-th/9812208`
Witten, E. (1999a), Lectures on perturbative quantum field theory. In: P. Deligne et al (Eds.) (1999), Vol. 1, 421–509.
Witten, E. (1999b), Lectures on dynamical aspects of quantum field theories. In: P. Deligne et al (Eds.) (1999), Vol. 2, 1119–1424.
Witten, E. (1999c), Lectures on Three-Dimensional Topological Quantum Field Theory. See Sen Hu (1999).
Witten, E. (2003), Physical law and the quest for mathematical understanding, Bull. Amer. Math. Soc. **40**, 21–30.
Woodhouse, N. (1997), Geometric Quantization, 3rd edn., Oxford University Press, New York.
Woodhouse, N. (2003), Special Relativity, Springer, New York.
Wulkenhaar, R. (2005), Lectures on Euclidean quantum field theory and commutative and noncommutative spaces. In: Ocampo, Paycha, and Vargas (Eds.) (2005), pp. 59–100.
Wussing, H. (Ed.) (1983), Geschichte der Naturwissenschaften (History of the natural sciences), Edition Leipzig (in German).

Yandell, B. (2001), The Honors Class: Hilbert's Problems and Their Solvers, Peters, Natick, Massachusetts.
Yang, Y. (2001), Solitons in Field Theory and Nonlinear Analysis, Springer, New York.
Yeh, J. (1973), Stochastic Processes and the Wiener Integral, Marcel Dekker, New York.
Yosida, K. (1960), Lectures on Differential-and Integral Equations, Interscience, New York (translated from Japanese into English).
Yosida, K. (1995), Functional Analysis, 7th edn., Springer, New York.

Zagier, D. (1981), Zetafunktionen und quadratische Zahlkörper: Eine Einführung in die höhere Zahlentheorie (Zeta functions and quadratic number fields: an introduction to advanced number theory), Springer, Berlin (in German).
Zagier, D. (1990), The Bloch–Wigner–Ramakrishnan polylogarithm function, Math. Ann. **286**, 613–624.
Zagier, D. (1995), Introduction to modular forms. In: Waldschmidt et al. (Eds.) (1995), pp. 238–291.
Zagier, D. (1996), Newman's short proof of the prime number theorem, Amer. Math. Monthly **104**, 705–708.
Zagier, D. (2001), Leçon inaugurale, Jeudi 17 Mai 2001, Collège de France, Paris.
Zee, A. (2003), Quantum Field Theory in a Nutshell, Princeton University Press, Princeton, New Jersey.
Zeidler, E. (1986), Nonlinear Functional Analysis and Its Applications. Vol. I: Fixed-Point Theory, 3rd edn. 1998; Vol. IIA: Linear Monotone Operators, 2nd

edn. 1997; Vol. IIB: Nonlinear Monotone Operators; Vol. III: Variational Methods and Optimization; Vol. IV: Applications to Mathematical Physics, 2nd edn. 1995, Springer, New York.

Zeidler, E. (1995a), Applied Functional Analysis, Vol 1: Applications to Mathematical Physics, 2nd edn. 1997, Applied Mathematical Sciences, AMS 108, Springer, New York.

Zeidler, E. (1995b), Applied Functional Analysis, Vol. 2: Main Principles and Their Applications, Applied Mathematical Sciences, AMS 109, Springer, New York.

Zeidler, E. (Ed.) (2002a), Teubner-Taschenbuch der Mathematik (in German), Vol. 1 (English edition: see Zeidler (2004));
Vol. 2 edited by G. Grosche, D. Ziegler, V. Ziegler, and E. Zeidler (English edition in preparation), Teubner, Stuttgart/Leipzig.

Zeidler, E. (2002b), Mathematics – the Cosmic Eye of Humanity.
Internet: `http://www.mis.mpg.de/`

Zeidler, E. (Ed.) (2004), Oxford User's Guide to Mathematics, Oxford University Press, Oxford (translated from German into English).

Zimmermann, W. (1969), Convergence of Bogoliubov's method of renormalization in momentum space, Commun. Math. Phys. **15**, 208–234.

Zinn-Justin, J. (2004), Quantum Field Theory and Critical Phenomena, 4th edn., Clarendon Press, Oxford.

Zorich, V. (2003), Analysis I, II, Springer, New York (translated from Russian into English).

Zwiebach, B. (2004), A First Course in String Theory, Cambridge University Press, Cambridge, United Kingdom.

List of Symbols

$f(x) := x^2$ (definition of f)
$f(x) \simeq g(x),\ x \to a$ (asymptotic equality); this means $\lim_{x\to a} \frac{f(x)}{g(x)} = 1$, 933
$f(x) = o(g(x)),\ x \to a$ (Landau symbol); this means $\lim_{x\to a} \frac{f(x)}{g(x)} = 0$, 933
$f(x) = O(g(x)),\ x \to a$, 933
$f(x) \sim \sum_{n=0}^{\infty} a_n x^n$ (asymptotic expansion), 306, 861

$\mathrm{sgn}(a)$ (sign of the real number a), 933
$[a,b],\ ]a,b[,\ ]a,b]$ (intervals), 933
$\sum_{n=-\infty}^{\infty} b_n$, 213
δ_{ij} (Kronecker symbol), 933
$\delta_{11} := 1,\ \delta_{12} := 0$
$\delta^{ij} = \delta_{ij} = \delta^i_j$, 933
$\delta_{\mathbf{pq}}$, 670
ε_{ij} (skew-symmetric symbol)
$\varepsilon_{12} = -\varepsilon_{21} = 1,\ \varepsilon_{11} = \varepsilon_{22} = 0$, 335

x, y, z (right-handed Cartesian coordinates)
$\mathbf{i}, \mathbf{j}, \mathbf{k}$ (right-handed orthonormal basis)
$\mathbf{x} := x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ (position vector)
$||\mathbf{x}||$ (length (norm) of the vector $\mathbf{x}$)

t (time)
$x^1 := x, x^2 := y, x^3 := z, x^0 := ct$ (space-time point in Minkowski space), 933
$\mu = 0, 1, 2, 3$ (indices for space-time variables in Minkowski space), 933
$j = 1, 2, 3$ (indices for spatial variables in Minkowski space), 933
$\eta_{\mu\nu}$ (Minkowski symbol), $\eta_{00} := 1,\ \eta_{11} := -1,\ \eta_{01} := 0$, 933
$\eta^{\mu\nu} = \eta_{\mu\nu} = \eta^\nu_\mu$, 933
$\epsilon^{\alpha\beta\gamma\delta}, \epsilon_{\alpha\beta\gamma\delta}$ (skew-symmetric symbol)
$\epsilon_{0123} := 1,\ \epsilon_{1023} := -1$, 934

$a_\mu b^\mu := \sum_{\mu=0}^{3} a_\mu b^\mu$ (Einstein's convention in Minkowski space), 934
$\zeta(s)$ (zeta function), 278
B_n (Bernoulli number), 278

$\theta(t)$ (Heaviside function), 90, 577
$\delta(t)$ (Dirac delta function), 590
δ, δ_x (Dirac delta distribution), 609
δ_μ (Dirac delta function with respect to the measure μ), 602
$\delta_{\Delta t}(t)$ (standard discrete Dirac delta function), 580
$\delta_{\Delta^4 x}, \delta_{\mathcal{C}(L)}, \delta_{\mathcal{G}(N)}, \delta_T, \delta_{\mathrm{dis}}$ (discrete Dirac delta functions), 441, 670
$\mathcal{C}(N)$ (cube in position space), 669
$\mathcal{G}(N)$ (grid in momentum space), 669
$\Delta^3 p$, 670
$\mathcal{V}$ (normalization volume), 669
$\delta(x^2 - a^2)$ (special distribution), 595
$\mathcal{P}\left(\frac{1}{x}\right)$ (special distribution), 618
$\frac{1}{x \pm 0_+ \mathrm{i}}$ (special distribution), 620
$\mathcal{P}\left(\ln|x|\right)$ (special distribution), 735

I, id (identity operator)
$x \in U$ (the point x is an element of U)
$U \subseteq V$ (U is a subset of V)
$U \subset V$ (U is a proper subset of V), 931
$U \cup V$ (the union of two given sets U and V)
$U \cap V$ (the intersection of two given sets U and V)
$U \setminus V$ (the difference of two sets U and V, i.e., the set of elements of U not belonging to V)
∂U (boundary of the set U)
$\mathrm{int}(U)$ (interior of U)
$\mathrm{cl}(U) \equiv U \cup \partial U$ (closure of U), 543
$\emptyset$ (empty set)
$\{x : x \text{ has the property } \mathcal{P}\}$ (the set of all things which have the property $\mathcal{P}$)

$f : X \to Y$ (map), 931
$\mathrm{im}(f)$ (image of the map f), 931
$\mathrm{dom}(f)$ (domain of f), 931
$f^{-1} : Y \to X$ (inverse map), 931
$f(U)$ (image of the set U), 931
$f^{-1}(V)$ (pre-image of the set V), 931

$z = x + y\mathrm{i}$ (complex number)
$\Re(z) := x$ (real part of z)
$\Im(z) := y$ (imaginary part of z)
$|z|$ (modulus of z), 209
$\arg(z)$ (principal argument of z), $-\pi < \arg(z) \leq \pi$, 209
$\arg_*(z)$ (argument of z), 209
$z^\dagger := x - y\mathrm{i}$ (conjugate complex number), 336
$\ln z$ (logarithmic function), 220
$\mathrm{res}_z(f)$ (residue of the function f at the point z), 213

$\mathbb{R}$ (set of real numbers)
$\mathbb{C}$ (set of complex numbers)
$\overline{\mathbb{C}}$ (closed complex plane), 217
$\mathbb{C}_>$ (open upper half-plane), 662
$\mathbb{C}_\geq$ (closed upper-half plane)
$\mathbb{C}_<$ (open lower half-plane), 662
$\mathbb{K} = \mathbb{R}, \mathbb{C}$ (set of real or complex numbers)
$\mathbb{Z}$ (set of integers, $0, \pm 1, \pm 2, \ldots$)
$\mathbb{N}$ (set of natural numbers, $0, 1, 2, \ldots$)
$\mathbb{Q}$ (set of rational numbers)
$\mathbb{R}^N, \mathbb{C}^N, \mathbb{K}^N$ ($N = 1, 2, \ldots$), 328
$\mathbb{M}^4$ (Minkowski space), 769,

$\mathbb{R}^\times$ (set of nonzero real numbers)
$\mathbb{N}^\times$ (set of nonzero natural numbers, $1, 2, \ldots$)
$\mathbb{C}^\times$ (set of nonzero complex numbers)
$\mathbb{K}^\times$ (set of nonzero numbers in $\mathbb{K}$)
$\mathbb{R}_\geq$ (set of nonnegative real numbers, $x \geq 0$)
$\mathbb{R}_>$ (set of positive real numbers, $x > 0$)
$\mathbb{R}_\leq$ (set of non-positive real numbers, $x \leq 0$)
$\mathbb{R}_<$ (set of negative real numbers, $x < 0$)
$\mathbb{R}_+$ (additive semigroup of nonnegative real numbers)
$\mathbb{R}_+^\times$ (multiplicative group of positive real numbers)

$\mathbb{B}^2$ (closed unit disc)
$\mathrm{int}(\mathbb{B}^2)$ (open unit disc)
$\mathbb{S}^1 \equiv \partial\mathbb{B}^2$ (unit circle)
$\mathbb{B}^3$ (closed 3-dimensional unit ball)
$\mathrm{int}(\mathbb{B}^3)$ (open 3-dimensional unit ball)
$\mathbb{S}^2 \equiv \partial\mathbb{B}^3$ (2-dimensional unit sphere)
$\mathbb{B}^n$ (closed n-dimensional unit ball), 268
$\mathbb{S}^n \equiv \partial\mathbb{B}^{n+1}$ (n-dimensional unit sphere)

$\bar{A}$ (mean value), 351
ΔA (mean fluctuation), 351

$\dim X$ (dimension of the linear space X), 330
$\mathrm{span}\, S$ (linear hull of the set S), 329
$\langle x|y\rangle$ (inner product), 336
$\langle\varphi|, |\psi\rangle$ (Dirac calculus), 359
$||\varphi||$ (norm), 336, 366
$L(X, Y)$ (space of linear operators), 331
X^d (dual space), 332
$A^\dagger$ (adjoint operator), 357
A^d (dual operator), 357
A^{-1} (inverse operator), 931
A^c (conjugate complex operator); this means $(A^\dagger)^d$
$\mathcal{A}^\dagger$ (adjoint matrix), 341
$\mathcal{A}^d$ (dual or transposed matrix), 341
$\mathcal{A}^c$ (conjugate complex matrix), 341
$\mathcal{A}^{-1}$ (inverse matrix), 931
$[A, B]_- := AB - BA$, 56
$[A, B]_+ := AB + BA$,
$\mathrm{tr}(A)$ (trace), 341, 363
$\det(A)$ (determinant), 333
e^A (exponential function), 345
$\ln A$ (logarithmic function), 346
$\sigma(A)$ (spectrum), 365
$\varrho(A) = \mathbb{C} \setminus \sigma(A)$ (resolvent set), 365

$GL(X), SL(X), U(X), SU(X)$ (Lie groups), 341
$U(1), U(n), SU(n), O(n), SO(n), GL(n, \mathbb{R}), SL(n, \mathbb{R}), GL(n, \mathbb{C}), SL(n, \mathbb{C})$ (matrix Lie groups), 341
$gl(X), sl(X), u(X), su(X)$ (Lie algebras), 342
$u(n), su(n), o(n), so(n), gl(n, \mathbb{R}), sl(n, \mathbb{R}), gl(n, \mathbb{C}), sl(n, \mathbb{C})$ (matrix Lie algebras), 343
$T_x M$ (tangent space), 348

$\lim_{n\to\infty} \varphi_n = \varphi$ (limit), 337
$f(+0)$, 93
$\mathbf{curl}\, \mathbf{E}$ (curl of the vector field $\mathbf{E}$), 170
$\mathrm{div}\, \mathbf{E}$ (divergence of $\mathbf{E}$), 170

$\not\partial$, $\not\nabla$ (Feynman's dagger symbols), 791

c (velocity of light in a vacuum), 949
h (Planck's quantum of action), $\hbar \equiv h/2\pi$, 949
k (Boltzmann constant), 949
G (gravitational constant), 949
ε_0 (electric field constant of a vacuum), 949
μ_0 (magnetic field constant of a vacuum), 949
e (electric charge of the proton), 949
$-e$ (electron charge)
m_e (electron mass), 949
α (fine structure constant), 949
λ_C (Compton wave length), 142
$\not\lambda_C \equiv \lambda_C/2\pi$ (reduced Compton wavelength), 142

m (meter), 934
s (second), 934
J (Joule), 934
C (Coulomb), 934
K (Kelvin), 934
$\mathbf{m}, \mathbf{s}, \mathbf{J}, \mathbf{C}, \mathbf{K}$ (Planck units), 937
eV (electron volt), 937
MeV (mega electron volt), 937
GeV (giga electron volt), 937

$\hat{\varrho}$ (statistical operator), 758
ϱ (density operator), 758
$\beta \equiv 1/kT$, 758

$P(t,s)$ (propagator), 383
$P^+(t,s)$ (retarded propagator), 384
$P^-(t,s)$ (advanced propagator), 384
$S[\varphi]$ (action of the field φ), 752
$\mathsf{S}(s,t)$ (S-matrix on the time interval $[t,s]$), 390
$\mathsf{S}(T)$ (S-matrix on $[-\frac{T}{2}, \frac{T}{2}]$), 822
$|0\rangle, \Phi_0$ (ground state of a free quantum field)
$|0_{\mathrm{int}}\rangle, \Phi_{\mathrm{int}}$ (ground state of a quantum field under interactions)
$\mathcal{T}$ (chronological operator), 388, 744
$:AB:$ (normal product), 822

$Z(J), C_n, \mathcal{C}_n, G_n$ (discrete model of a quantum field), 444
$Z(J,\varphi), Z_{\mathrm{free}}(J,\varphi)$, 448
S_n, 450
φ_{mean}, 456
Z_{red}, 457
V_n, 458
$C_{n,\mathrm{free}}$, 464

$Z(J)$ (full generating functional), 749
$Z_{\mathrm{free}}(J)$ (free generating functional), 749
C_n (full n-point correlation function), 744
$C_{n,\mathrm{free}}$ (free n-point correlation function), 743
G_n (full n-point Green's function), 744
$G_{n,\mathrm{free}}$ (free n-point Green's function), 743

$\frac{1}{p^2 - m_0^2 + 0_+ \mathrm{i}}$ (special distribution), 780
$\mathcal{G}_{F,m_0}$(Feynman propagator for mesons), 775
G_{F,m_0} (Feynman propagator distribution for mesons), 778
$\mathsf{G}_F \equiv \mathsf{G}_{F,m_0=0}$ (Feynman propagator distribution for the wave equation)
$\mathsf{D}_F^{\alpha\beta} \equiv -\eta^{\alpha\beta}\mathsf{G}_F$ (Feynman propagator distribution for photons), 800
$\mathsf{S}_F \equiv (\mathrm{i}\gamma^\alpha \partial_\alpha + m_e)\mathsf{G}_{F,m_e}$ (Feynman propagator distribution for electrons), 800

Index