Ian Postlethwaite
Alistair G. J. MacFarlane

A Complex Variable Approach to the Analysis of Linear Multivariable Feedback Systems

Springer-Verlag
Berlin Heidelberg New York 1979
Authors
Dr. I. Postlethwaite,
Research Fellow, Trinity Hall, Cambridge, and
SRC Postdoctoral Research Fellow,
Engineering Department, University of Cambridge.

Professor A. G. J. MacFarlane,
Engineering Department, University of Cambridge,
Control and Management Systems Division,
Mill Lane,
Cambridge CB2 1RX.
## Contents

**CHAPTER 1**

| Introduction                            | 1 |
| References                              | 6 |

**CHAPTER 2**

| Preliminaries                           | 8 |
| 2.1 System description                   | 8 |
| 2.2 Feedback configuration               | 11 |
| 2.3 Stability                            | 13 |
| 2.3-1 Free systems                       | 13 |
| 2.3-2 Forced systems                     | 16 |
| 2.4 Relationship between open- and closed-loop characteristic polynomials for the general feedback configuration | 17 |
| References                              | 20 |

**CHAPTER 3**

| Characteristic gain functions and characteristic frequency functions | 22 |
| 3.1 Duality between open-loop gain and closed-loop frequency        | 22 |
| 3.2 Algebraic functions: characteristic gain functions and characteristic frequency functions | 25 |
| 3.3 Characteristic gain functions                                   | 27 |
| 3.3-1 Poles and zeros of a characteristic gain function             | 29 |
| 3.3-2 Algebraic definition of poles and zeros for a transfer function matrix | 32 |
| 3.3-3 Relationship between algebraically defined poles/zeros of the open-loop gain matrix G(s) and the poles/zeros of the corresponding set of characteristic gain functions | 36 |
| 3.3-4 Riemann surface of a characteristic gain function             | 39 |
| 3.3-5 Generalized root locus diagrams                               | 46 |
| 3.3-6 Example of frequency surface and characteristic frequency loci | 47 |
| 3.4 Characteristic frequency functions                              | 49 |
| 3.4-1 Generalized Nyquist diagram                                   | 50 |
| 3.4-2 Example of gain surface and characteristic gain loci           | 51 |
| References                                                           | 57 |

**CHAPTER 4**

| A generalized Nyquist stability criterion                           | 58 |
| 4.1 Generalized Nyquist stability criterion                          | 58 |
| 4.2 Proof of the generalized Nyquist stability criterion             | 60 |
| 4.3 Example                                                           | 72 |
| References                                                           | 75 |

**CHAPTER 5**

| A generalized inverse Nyquist stability criterion                    | 77 |
| 5.1 Inverse characteristic gain functions                            | 77 |
| 5.2 Pole-zero relationships                                           | 78 |
1. Introduction

The great success of the optimal control and optimal filtering techniques developed for aerospace work during the late 1950's and early 1960's naturally led to attempts to apply these techniques to a wide range of earth-bound multi-variable industrial processes. In many situations this was less than immediately successful, particularly in cases where the available plant models were not sufficiently accurate or where the performance indices required to stipulate the controlled plant behaviour were much less obvious in form than in the aerospace context. Moreover, the controller which results from a direct application of optimal control and optimal filtering synthesis techniques is in general a complicated one; in fact, if it incorporates a full Kalman-Bucy filter it has a dynamical complexity equal to that of the plant which it is controlling, since the filter essentially consists of a plant model with feedback around it. In contrast, what was needed for many multivariable process control problems was a relatively simple controller which would both stabilize, about an operating point, a plant for which only a very approximate model might be available, and also mitigate the effect of low-frequency disturbances by incorporating integral action. To industrial engineers brought up on frequency-response ideas the sophisticated optimal control methods seemed difficult to use; these engineers essentially relied on a mixture of physical insight and straightforward techniques, such as the use of derivative and integral action, to solve their problems. It became obvious that a huge gap in techniques existed between the classical single-loop frequency-response methods, based on the work of Nyquist [1],
Bode [2] and Evans [3], which were still in use for many industrial applications, and the elegant and powerful multivariable time-response methods developed for aerospace applications.

For these reasons an interest in frequency-response methods slowly began to revive during the mid-1960's. An important first step towards closing the yawning gap between an optimal control approach and the classical frequency-response approach was taken by Kalman [4], who studied the frequency-domain characterization of optimality. A systematic attack on the whole problem of developing a frequency-response analysis and design theory for multivariable systems was begun in a pioneering paper by Rosenbrock [5] which ushered in a decade of increasing interest in a rejuvenated frequency-response approach. Prior to this new point-of-departure some fairly straightforward attacks had been made on the multivariable control problem. Boksenbom and Hood [6] put forward the idea of a non-interacting controller. Their procedure consisted simply of choosing a cascaded compensator such that the overall transfer function matrix of the compensated system had a diagonal form. If such a compensator could be found then the controller design could be finished off using standard single-loop design techniques. The required compensating matrix which usually results from such a procedure is necessarily a complicated one, and the most succinct objection to this approach is simply that it is not essential to go to such drastic lengths merely to reduce interaction. A natural further step in this initial approach to multivariable control was to see what could be achieved by way of standard matrix calculations using rational matrices; papers studying the problem in this way were produced
by Golomb and Usdin [7], Raymond [8], Kavanagh [9], [10],[11], and Freeman [12], [13]. Rosenbrock [14],[15], however, opened up a completely new line of development by seeking to reduce a multi-variable problem to one amenable to classical techniques in a more sophisticated way. In his Inverse Nyquist Array Method [14], [15] the aim was to reduce interaction to an amount which would then enable single-loop techniques to be employed, rather than to eliminate interaction completely. The Rosenbrock approach was based upon a careful use of a specific criterion of partial interaction – the diagonal dominance concept. The success of this method led other investigators to develop ways of seeking to reduce a multivariable control problem to a succession of single-loop problems, as in the Sequential Return Difference approach of Mayne [16].

In the non-interacting, or partially non-interacting, approach to multivariable control the motivation was the eventual deployment of classical single-loop frequency-response techniques during the final stages of a design study. An alternative approach, however, is to investigate the transfer-function matrix representation as a single object in its own right and to ask: how can the key basic concepts of the classical single-loop frequency-response approach be suitably extended? What are the relevant generalizations to the multivariable case of the specific concepts of pole, zero, Nyquist diagram and root locus diagram? It is to questions of this sort that the work presented here is addressed, and it is shown that complex-variable ideas have an important role to play in the study of multivariable feedback systems. An early attempt to extend Nyquist diagram ideas to the multivariable problem was made by Bohn [17], [18]. A generalization of the Nyquist stability criterion was put forward by MacFarlane [19] and, following that heuristic treatment, complex-variable based proofs were supplied by Barman and Katzenelson [20] and MacFarlane and Postlethwaite [21]. This generalization of the Nyquist stability criterion to the multivariable situation was soon followed by complementary generalizations of the
The aim of the work presented in this text is to extend the concepts underlying the techniques of Nyquist, Bode and Evans to multivariable systems. In the two classical approaches to linear feedback system design the Nyquist-Bode approach studies gain as a function of frequency and the Evans' approach studies frequency as a function of gain. In Chapter 3 it is shown how the ideas of studying complex gain as a function of complex frequency and complex frequency as a function of complex gain can be extended to the multivariable case by associating with transfer function matrices (having the same number of rows and columns) a pair of analytic functions: a characteristic gain function and a characteristic frequency function. These are algebraic functions [25] and each is defined on an appropriate Riemann surface [26]. Chapter 2 deals with a number of essential preliminaries such as a description of the type of multivariable feedback system being considered; with basic definitions of stability and related theorems; and with a fundamental relationship between open- and closed-loop behaviour based on the return-difference operator. Chapter 3 also contains a comprehensive discussion of the background to the generalized Nyquist stability criterion for multivariable feedback systems which is presented in Chapter 4. The proof of this criterion is based on the Principle of the Argument applied to an algebraic function defined on an appropriate Riemann surface. In Chapter 5 a generalization of the inverse Nyquist stability criterion to the multivariable case is developed which is complementary to the exposition of the generalized Nyquist criterion given in the previous chapter. Using the material developed in Chapter 3, the Evans root locus approach is extended to multivariable systems.
in Chapter 6; this uses well established results in algebraic function theory. It is also shown how an algebraic-function based approach can be used to find the asymptotic behaviour of the closed-loop poles of a multivariable time-invariant optimal linear regulator as the weight on the input terms of a quadratic performance index approaches zero.

As the work presented progresses it becomes evident that the gain variable used can be considered as a parameter of the system, and consequently that the techniques developed are not only applicable to gain and frequency but to any parameter and frequency. In Chapter 7 the effect of parameter variations on a multivariable feedback system is considered by the introduction of the concepts of 'parametric' root loci and 'parametric' Nyquist loci. This chapter concludes with a few tentative proposals and suggestions for future research.

Information of secondary importance which would unnecessarily break the flow of the text has been placed in appendices. References are listed at the end of each chapter in which they are cited, and also at the end of the text where a bibliography is provided.

References


2. Preliminaries

This text considers the generalization of the classical techniques of Nyquist and Evans to a linear time-invariant dynamical feedback system which consists of several multi-input, multi-output subsystems connected in series. In this chapter a description of the multivariable feedback system under consideration is given. The chapter also includes basic definitions of stability, some associated theorems, and a fundamental relationship between open- and closed-loop behaviour based on the return-difference operator.

2.1 System description

The basic description of a linear time-invariant dynamical system is taken to be the state-space model

\[
\begin{align*}
\dot{x}(t) & = Ax(t) + Bu(t) \\
y(t) & = Cx(t) + Du(t)
\end{align*}
\] (2.1.1)

where \( x(t) \) is the state vector, \( y(t) \) the output vector, \( u(t) \) the input vector; \( \dot{x}(t) \) denotes the derivative of \( x(t) \) with respect to time; \( A, B, C, \) and \( D \) are constant real matrices. For convenience the model will be denoted by \( S(A,B,C,D) \) or \( S \) when the meaning is obvious, and represented diagramatically as shown in figure 1.

In general \( S(A,B,C,D) \) will be considered as being the state-space representation of several subsystems

\[
\begin{align*}
\dot{x}_i(t) & = A_i x_i(t) + B_i u_i(t) \\
y_i(t) & = C_i x_i(t) + D_i u_i(t)
\end{align*}
\] (2.1.2)

\( i=1,2,\ldots,h \)

connected in series, as illustrated in figure 2. If, for example, \( S \) consists of two subsystems \( S_1 \) and \( S_2 \) then the
Figure 1. State-space model

\[ u(t) = u(t) \]

\[ y(t) = y(t) \]

Figure 2. Series connection of subsystems

\[ u(t) = u_1(t) \]

\[ y(t) = y(t) \]

\[ y(t) = y(t) \]
state-space description of $S$ is given by equations (2.1.1) with
\[ x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \text{ the combined states of both subsystems,} \]
\[ u(t) = u_1(t), \text{ the input to } S_1, \]
\[ y(t) = y_2(t), \text{ the output of } S_2, \] (2.1.3)
\[ A = \begin{bmatrix} A_1 & 0 \\ B_2C_1 & A_2 \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2D_1 \end{bmatrix}, \]
\[ C = \begin{bmatrix} D_2C_1 & C_2 \end{bmatrix}, \text{ and } D = D_2D_1. \]

The state-space model for an interconnection of several subsystems can be derived from the above formula by successive application.

The state-space model of a system is often referred to as an **internal** description since it retains a knowledge of the system's internal dynamical structure. An **external** or **input-output** description is obtained if in equations (2.1.1) single-sided Laplace transforms [1] are taken, to give
\[ s\hat{x}(s) - x(0) = A\hat{x}(s) + B\hat{u}(s) \] (2.1.4)
\[ \hat{y}(s) = C\hat{x}(s) + D\hat{u}(s) \]

where $\hat{x}(s)$ denotes the Laplace transform of $x(t)$. If the initial conditions at time $t=0$ are all zero so that $x(0)=0$, then the input and output transform vectors are related by
\[ \hat{y}(s) = G(s) \hat{u}(s) \] (2.1.5)

where
\[ G(s) = C(sI_n - A)^{-1}B + D \] (2.1.6)

$I_n$ is a unit matrix of order $n$ and $(\cdot)^{-1}$ denotes the inverse of a matrix. $G(s)$ is a matrix-valued rational function.
of the complex variable $s$, and is called the **transfer function matrix** for the set of input-output transforms, or the **open-loop gain matrix**. The transfer function matrix $G(s)$ can be regarded as describing a system's response to an exponential input with exponent $s$ \(^2\), and therefore the complex variable $s$ can be considered a complex frequency variable as in the single-input, single-output case.

When $S(A,B,C,D)$ consists of $h$ subsystems
\[ S_i(A_i,B_i,C_i,D_i): i=1,2,\ldots,h \], as shown in figure 2, each subsystem has a transfer function matrix
\[ G_i(s) = C_i(sI - A_i)^{-1}B_i + D_i \]  
(2.1.7)

and the input-output transform vectors of $S$ are related by
\[ \hat{y}(s) = G_h(s)G_{h-1}(s)\cdots G_1(s) \hat{u}(s) \]  
(2.1.8)

with the obvious relationship for the open-loop gain matrix
\[ G(s) = G_h(s)G_{h-1}(s)\cdots G_1(s) \]  
(2.1.9)

For the purpose of connecting outputs back to inputs to form a feedback loop $G(s)$ is assumed to be a square matrix of order $m$.

### 2.2 Feedback configuration

The general feedback configuration that will be considered is shown in figure 3. The output of the feedback system is shown as that of the $h$th subsystem but in practice it may be the output from an earlier subsystem in which case the later subsystems can be thought of as being feedback compensators. The parameter $k$ is a real gain control variable common to all the loops. The system's input and
output are related to the reference input \( r(t) \) by the equations

\[
e(t) = r(t) - y(t) \\
u(t) = ke(t) \quad (2.2.1)
\]

and combining these with equations (2.1.1) the following closed-loop state-space equations are obtained:

\[
\dot{x}(t) = A_c x(t) + B_c r(t) \\
y(t) = C_c x(t) + D_c r(t) \quad (2.2.2)
\]

where

\[
A_c = A - B(k^{-1}I_m + D)^{-1}C \\
B_c = kB - kB(k^{-1}I_m + D)^{-1}D \\
C_c = (I_m + kD)^{-1}C \\
D_c = (k^{-1}I_m + D)^{-1}D
\]

Figure 3. Feedback configuration
2.3 Stability

Stability is the most important single requirement of a feedback system and for general time-dependent nonlinear systems it poses very complex problems. The stability problem for linear time-invariant dynamical systems, however, is much simpler than in the general case. This is because:
(i) all stability properties are constant with respect to time, and
(ii) all stability properties are global, since any solution for the state of the system is proportional to the state at time zero; see equation (2.3.2).

There are many definitions of stability in the literature and broadly speaking these can be divided into two classes. The first class of definitions concerns stability of free systems i.e. those in which there is no input; the second class of definitions concerns the behaviour of forced systems i.e. those in which there is a given input. Both types of stability are discussed below, the definitions and associated theorems following very closely those given by Willems [3].

2.3-1 Free systems

Let us consider the closed-loop dynamical system of figure 3 described by equations (2.2.2) with \( r(t)=0 \) and \( C_c=I \). Then the stability problem reduces to that of considering the free system

\[
\dot{x}(t) = A_c x(t) \quad (2.3.1)
\]

The equilibrium state for equation (2.3.1) is clearly the
origin (assuming $A_c$ is non-singular), and therefore a solution of (2.3.1) which passes through the origin at some time remains there for all subsequent times; this solution is called the null solution. The stability of the origin equilibrium state is characterized using the following definitions.

**Definition 1.** The origin of the free system (2.3.1) is called stable if when the system is perturbed from the origin all subsequent motions remain in a correspondingly small neighbourhood of the origin.

**Definition 2.** The origin of the free system (2.3.1) is called asymptotically stable if when the system is perturbed slightly from the origin all subsequent motions return to the origin.

**Definition 3.** The origin of the free system (2.3.1) is called asymptotically stable in the large, or globally asymptotically stable, if it is stable, and if every motion converges to the origin as $t \to \infty$.

The general solution of equation (2.3.1) is [3]

$$x(t; x(t_0), t_0) = \exp \left[ A_c (t-t_0) \right] x(t_0) \quad (2.3.2)$$

which shows clearly that if the free system is asymptotically stable it is also asymptotically stable in the large. If $J$ is the Jordan canonical form [4] of $A_c$ such that

$$A_c = TJT^{-1} \quad (2.3.3)$$

with

$$J = \begin{bmatrix} J_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & J_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \cdots & J_k \end{bmatrix}$$

where each Jordan block $J_i$ has the form
and $\lambda_i$ is an eigenvalue of $A_C$, then it can be shown [3], that
\[
\exp[A_C(t-t_0)] = T \exp[J(t-t_0)]T^{-1} \tag{2.3.4}
\]
with
\[
\exp[J(t-t_0)] = \begin{bmatrix}
\exp[J_1(t-t_0)] \\
\exp[J_2(t-t_0)] \\
\vdots \\
\exp[J_k(t-t_0)]
\end{bmatrix}
\]
and
\[
\exp[J_i(t-t_0)] = \begin{bmatrix}
1 \pm \frac{t^2}{2!} \ldots \frac{t^{r-1}}{(r-1)!} \\
0 1 \pm \frac{t^{r-2}}{(r-2)!} \\
\vdots \\
0 0 0 1
\end{bmatrix} \exp[\lambda_i(t-t_0)]
\]
where $\bar{t} = (t-t_0)$ and $r$ is the order of the Jordan block $J_i$. The general solution of the free system can therefore be expressed as
\[
x(t; x(t_0), t_0) = T \exp[J(t-t_0)]T^{-1}x(t_0) \tag{2.3.5}
\]
and from this the following theorems can be derived; see [3] for proofs.

**Theorem 1.** The null solution of system (2.3.1) is asymptotically stable if and only if all eigenvalues of the matrix $A_C$ have negative real parts.

**Theorem 2.** The null solution of system (2.3.1) is stable if and only if the matrix $A_C$ has no eigenvalues with positive real parts, and if the eigenvalues with zero real parts correspond to Jordan blocks of order 1.
2.3-2 Forced systems

Let us consider the closed-loop dynamical system (2.2.2) which has the general solution [3],

\[ x(t; x(t_0), t_0) = \exp(A_c t) x(t_0) + \int_0^t \exp(A_c (t-T)) B_c r(T) dT \]  

(2.3.6)

To study the stability properties of this system we need to introduce the concept of input-output stability.

**Definition 4.** A dynamical system is called input-output stable if for any bounded input a bounded output results regardless of the initial state.

By theorem 1, asymptotic stability of the unforced system (2.3.1) implies that all the eigenvalues of \( A_c \) have negative real parts, in which case there exist positive numbers \( P \) and \( a \) such that

\[ \| \exp(A_c t) \| \leq P \exp(-at) \quad \forall t \geq 0 \]  

(2.3.7)

where \( \| . \| \) denotes the Euclidean norm of a matrix or vector [3]. From equations (2.2.2), (2.3.6) and (2.3.7) we then have

\[ \| y(t) \| \leq \| C_c x(t) \| + \| D_c r(t) \| + \| C_c \exp(A_c t) . x(t_0) \| + \int_0^t \| \exp(A_c (t-T)) \| \| B_c r(T) \| dT + c \| x(t_0) \| + cbMP/a \]

where \( b=\| B_c \|, \quad c=\| C_c \|, \quad d=\| D_c \|, \) and \( \| r(t) \| \leq M \quad \forall \ t \geq 0. \)

This result is summarized in the following theorem.

**Theorem 3.** If the null solution of the unforced system (2.3.1) is asymptotically stable, then the forced system (2.2.2) is input-output stable.

Note that input-output stability implies asymptotic stability.
of the equilibrium state at the origin only if the system (2.2.2) is state controllable and state observable; or if all unobservable and/or uncontrollable modes have negative real parts. In the remainder of this book system stability is understood as meaning input-output stability coupled with asymptotic stability of the equilibrium state at the origin.

Theorem 3 is important because it tells us that the stability of a linear time-invariant system can be determined solely from a knowledge of the eigenvalues of the system "A" matrix. The stability conscious eigenvalues corresponding to the closed-loop dynamical system (2.2.2) are values of \(\lambda\) which satisfy the equation

\[
\text{det}(\lambda I_n - A_c) = 0 \quad (2.3.8)
\]

The left-hand side of equation (2.3.8) is called the closed-loop characteristic polynomial, abbreviated as \(\text{CLCP}(\lambda)\) so that

\[
\text{CLCP}(\lambda) \triangleq \text{det}(\lambda I_n - A_c) \quad (2.3.9)
\]

Similarly for the open-loop system \(S(A,B,C,D)\) an open-loop characteristic polynomial, \(\text{OLCP}(\lambda)\), is defined as

\[
\text{OLCP}(\lambda) \triangleq \text{det}(\lambda I_n - A) = \text{det}(\lambda I_{n_1} - A_1)\text{det}(\lambda I_{n_2} - A_2)\ldots \text{det}(\lambda I_{n_h} - A_h) \quad (2.3.10)
\]

In the next section it is shown how the open- and closed-loop characteristic polynomials are related via the return-difference operator [5].

2.4. Relationship between open- and closed-loop characteristic polynomials for the general feedback configuration

Let us suppose that all the feedback loops of the general
Figure 4. Feedback configuration
(a) closed-loop
(b) open-loop
closed-loop configuration are broken and that the subsystems are represented by their transfer function matrices; see figure 4. The corresponding return-difference matrix [5] for this break point is

\[
P(s) - \frac{\hat{A}}{I_m + L(s)}
\]

(2.4.1)

where

\[
L(s) = kG_h(s)G_{h-1}(s)\ldots G_1(s)
\]

(2.4.2)

is called the system return-ratio matrix [5]. A return-difference operator generates the difference between injected and returned signal transforms from the injected signal transform. It plays a major role in feedback theory since the essence of forging a feedback link is making two sets of signals identically equal, thus making the difference between them identically zero. Both \(F(s)\) and \(L(s)\) are matrix-valued rational functions of a complex variable and the key concepts in this text revolve around the properties of such matrices. The importance of the return-difference matrix is emphasized in the relationship between open- and closed-loop characteristic polynomials which is now derived for the general feedback configuration.

If we take determinants of equation (2.4.1) and represent \(G(s)\) by its state-space model, we obtain

\[
\det F(s) = \det \left[ I_m + kC(sI_n - A)^{-1}B + kD \right]
\]

(2.4.3)

which using Schur's formula [6] for the evaluation of partitioned determinants can be rewritten as

\[
\det F(s) = \det \left[ \begin{array}{c|c} sI_n - A & B \\ \hline -kC & I_m + kD \end{array} \right] \cdot \det [sI_n - A]
\]

(2.4.4)

which is equivalent to
\[ \text{det} F(s) = \text{det} \begin{bmatrix} I_n & -B(I_m + kD)^{-1} \\ 0 & I_m \end{bmatrix} \text{det} \begin{bmatrix} sI_n - A & B \\ -kC & I_m + kD \end{bmatrix} \]

\[ = \text{det} \begin{bmatrix} sI_n - A + B(k^{-1}I_m + D)^{-1}C \\ -kC \\ 0 \end{bmatrix} \text{det} \begin{bmatrix} sI_n - A \\ I_m + kD \end{bmatrix} \]

\[ = \frac{\text{det} \left[ sI_n - A + B(k^{-1}I_m + D)^{-1}C \right] \text{det} \left[ I_m + kD \right]}{\text{det} \left[ sI_n - A \right]} \text{det} \left[ I_m + kD \right] \tag{2.4.5} \]

Now from equations (2.2.2) we have

\[ A_c = A - B(k^{-1}I_m + D)^{-1}C \]

and it is obvious from equation (2.4.3) that

\[ \text{det} F(\omega) = \text{det} \left[ I_n + kD \right] \]

and therefore under the assumption that \( \text{det} F(\omega) \neq 0 \) we have

from equation (2.4.5) the following relationship

\[
\frac{\text{det} F(s)}{\text{det} F(\omega)} = \frac{\text{det} \left[ sI_n - A_c \right]}{\text{det} \left[ sI_n - A \right]} \text{det} \left[ sI_n - A \right] \ldots \text{det} \left[ sI_n - A \right] = \frac{\text{CLCP}(s)}{\text{OLCP}(s)} \tag{2.4.6} \]

The zeros of the open- and closed-loop characteristic polynomials, \( \text{OLCP}(s) \) and \( \text{CLCP}(s) \), are known as the open- and closed-loop poles or characteristic frequencies respectively.

Relationship (2.4.6) shows how the matrix-valued rational transfer functions \( F(s) \) and \( G(s) \) are intimately related to the stability of a dynamical feedback system.

The study of such matrices and their eigenvalues opens the way to suitable extensions of the classical techniques of Nyquist \[7\] and Evans \[8;9\]; the results of such a study are given in Chapter 3.

References


In the analysis and design of linear single-loop feedback systems the two classical approaches use complex functions to study open-loop gain as a function of imposed frequency (the Nyquist-Bode approach), and to study closed-loop frequency as a function of imposed gain (the Evans root locus approach). The primary purpose of this chapter is to show how these techniques can be extended to the multi-variable case by associating with appropriate matrix-valued rational functions of a complex variable characteristic gain functions and characteristic frequency functions.

3.1 Duality between open-loop gain and closed-loop frequency

For the general feedback configuration of figure 4 we have from section 2.4 the fundamental relationship

\[
\frac{\det F(s)}{\det F(\omega)} = \frac{\det \left[ sI_n - A_c \right]}{\det \left[ sI_n - A \right]} \quad (3.1.1)
\]

where the return-difference matrix \( F(s) \) is given as

\[
F(s) = I_m + kG(s) \quad (3.1.2)
\]

If we substitute for \( F(s) \) in equation \( (3.1.1) \) we obtain

\[
\frac{\det \left[ sI_n - A_c \right]}{\det \left[ sI_n - A \right]} = \frac{\det \left[ I_m + kG(s) \right]}{\det \left[ I_m + kD \right]} = \frac{\det \left[ k^{-1}I_m + G(s) \right]}{\det \left[ k^{-1}I_m + D \right]} \quad (3.1.3)
\]

and substituting for the gain variable \( k \) using the expression

\[
g = \frac{-1}{k} \quad (3.1.4)
\]

where \( g \) is allowed to be complex i.e. \( g \in \mathbb{C} \) (the complex plane), we have
The closed-loop system matrix $A_c$ is given in equations (2.2.2) as

$$A_c = A - B(k^{-1}I_m+D)^{-1}C$$

and substituting for $k$ from equation (3.1.4) we have

$$A_c = A + B(gI_m-D)^{-1}C$$

The expression (3.1.5) can therefore be rewritten as

$$\frac{\det[sI_n-A_c]}{\det[sI_n-A]} = \frac{\det[gI_m-G(s)]}{\det[gI_m-D]}$$

or

$$\frac{\det[sI_n-S(g)]}{\det[sI_n-S(\infty)]} = \frac{\det[gI_m-G(s)]}{\det[gI_m-G(\infty)]}$$

The form of this relationship shows a striking 'duality' between the complex frequency variable $s$ and the complex gain variable $g$ via their 'parent' matrices $S(g)$ and $G(s)$ respectively. This duality between the roles of frequency and gain forms the basis on which the classical complex variable methods are generalized to the multivariable case. $S(g)$ is called the **closed-loop frequency matrix**: its eigenvalues are the closed-loop characteristic frequencies and are clearly dependent on the gain variable $g$. The eigenvalues of the **open-loop gain matrix** $G(s)$ are called open-loop characteristic gains and are clearly dependent on the frequency variable $s$. The similarity between $G(s)$ and $S(g)$ is stressed if one examines their state-space structures:

$$G(s) = C(sI_n-A)^{-1}B + D$$

$$S(g) = B(gI_m-D)^{-1}C + A$$
In figure 5 the feedback configuration of figure 4a is redrawn with zero reference input, the state-space representation for $G(s)$, and the substitution (3.1.4) for $k$ in order to illustrate explicitly the duality between the closed-loop characteristic frequency variable $s$ and the open-loop characteristic gain variable $g$.

The importance of relationship (3.1.8) is that it shows, for values of $s \not\in \sigma(A)$ and values of $g \not\in \sigma(D)$ (this condition is equivalent to $\det F(\omega) \neq 0$ which has already been assumed), where $\sigma(A)$ denotes the spectrum of $A$, that

$$\det[sI_n - S(g)] = 0 \iff \det[gI_m - G(s)] = 0$$

(3.1.11)
This tells us that a knowledge of the open-loop characteristic gain as a function of frequency is equivalent to a knowledge of closed-loop characteristic frequency as a function of gain. The inference from this is that it ought to be possible to determine the stability of a feedback system from a knowledge of the characteristic gain spectrum of $G(s)$. Note that from equation (3.1.8) we have that

$$CLCP(s) = \frac{\det[gI_n - G(s)]}{\det[gI_m - D]} \cdot OLCP(s)$$

(3.1.12)

and such an expression makes it intuitively obvious that there should be a generalization of Nyquist's stability theorem to loci of the characteristic gains of $G(s)$ as a function of frequency.

3.2 Algebraic functions: characteristic gain functions and characteristic frequency functions.

The characteristic equations for $G(s)$ and $S(g)$ i.e.

$$\Delta(g,s) \triangleq \det[gI_m - G(s)] = 0 \quad (3.2.1)$$

and

$$\nabla(s,g) \triangleq \det[sI_n - S(g)] = 0 \quad (3.2.2)$$

are algebraic equations relating the complex variables $s$ and $g$. Each equation can be considered as a polynomial in $g$ or $s$ with coefficients which are rational functions in $s$ or $g$ respectively, and if irreducible over the field of rational functions each equation defines a pair of algebraic functions [1; appendix 1]:

(i) a characteristic gain function $g(s)$ which gives open-loop characteristic gain as a function of frequency, and

(ii) a characteristic frequency function $s(g)$ which gives
closed-loop characteristic frequency as a function of gain.

In general equations (3.2.1) and (3.2.2) will not be irreducible and each equation will define a set of characteristic gain and characteristic frequency functions. For simplicity of exposition and because this is in any case the usual situation for $G(s)$ and $S(g)$ arising from practical situations, it will normally be assumed that equations (3.2.1) and (3.2.2) are irreducible over the field of rational functions.

Although both equation (3.2.1) and equation (3.2.2) define the same functions $g(s)$ and $s(g)$, equation (3.2.2) will in general contain more information about the system. It is possible under certain circumstances that $\Delta(s,g)$ will contain factors of $s$ independent of $g$ which are not present in $\Delta(g,s)$. These factors occur in the following situations:-

1. When the $A$-matrix of the open-loop system $S(A,B,C,D)$ has eigenvalues which correspond to modes of the system which are unobservable and/or uncontrollable from the point of view of considering the input as that of the first subsystem and the output as that of the $h$th subsystem. Note that if output measurements for earlier subsystems are available then in practice some of the unobservable modes of $S(A,B,C,D)$ may in fact be observable.

2. When the poles and zeros of the open-loop gain matrix $G(s)$ are different from the poles and zeros of the characteristic gain function $g(s)$; see section 3.3-3.

These two conditions, under which equations (3.2.1) and (3.2.2) differ, clearly present problems to the development
of a Nyquist-like stability criterion in terms of loci of the characteristic gains of \( G(s) \). However, by relating the poles and zeros of \( g(s) \) to the poles and zeros of \( G(s) \), and by careful consideration of the unobservable and uncontrollable modes these problems can be overcome; a generalized Nyquist stability criterion is developed in chapter 4.

In the next section a detailed study of the characteristic gain function is given which results in a generalization of the root locus diagram. In section 3.4 a similar study of the characteristic frequency function results in a generalized Nyquist diagram.

3.3 Characteristic gain functions

The natural way to define the characteristic gain function \( g(s) \) is via the characteristic equation

\[
\Delta(g,s) \triangleq \det[g_{Im} - G(s)] = 0 \quad (3.3.1)
\]

In general \( \Delta(g,s) \) will be reducible to the form

\[
\Delta(g,s) = \Delta_1(g,s)\Delta_2(g,s) \cdots \Delta_\ell(g,s) \quad (3.3.2)
\]

where the factors \( \{\Delta_i(g,s) : i=1,2,\ldots,\ell\} \) are polynomials in \( g \) which are irreducible over the field of rational functions in \( s \). Let the irreducible factors \( \Delta_i(g,s) \) have the form

\[
\Delta_i(g,s) = g_1^{t_i} + a_{i1}(s)g_1^{t_i-1} + \cdots + a_{it_i}(s) = 0 \quad (3.3.3)
\]

where \( t_i \) is the degree of the \( i \)-th irreducible polynomial and the coefficients \( \{a_{ij}(s) : i=1,2,\ldots,\ell; j=1,2,\ldots,t_i\} \) are rational functions in \( s \). Then if \( b_{10}(s) \) is the least common denominator of the coefficients \( \{a_{ij}(s) : j=1,2,\ldots,t_i\} \) equation (3.3.3) can be put in the form
where the coefficients \( \{b_{ij}(s) : i=1,2,\ldots,\ell; j=1,2,\ldots,t_i\} \)
are polynomials in \( s \). The function of a complex variable
\( g_i(s) \) defined by equation (3.3.4) is called an algebraic
function \([1]; \text{appendix I}\). Thus associated with an open-loop
gain matrix \( G(s) \) is a set of algebraic functions
\( \{g_i(s) : i=1,2,\ldots,\ell\} \) which are directly related to the
eigenvalues of \( G(s) \). The characteristic gain functions
of \( G(s) \) are defined to be the set of algebraic functions
\( \{g_i(s) : i=1,2,\ldots,\ell\} \).

The problem of finding the irreducible polynomials
\( \{\Delta_i(g,s) : i=1,2,\ldots,\ell\} \) from which the characteristic gain
functions are defined is closely linked to the problem of
finding an appropriate canonical form of \( G(s) \). If \( \Delta(g,s) \)
was reducible to factors linear in \( g \) then \( G(s) \) could be
put into Jordan form \([2]\). In general this will not be the
case and a suitable canonical form is defined as follows.

Let

\[
C(\Delta_i) \triangleq \begin{bmatrix}
0 & 0 & \ldots & 0 & -a_{i t_i}(s) \\
1 & 0 & \ldots & 0 & -a_{i, t_i-1}(s) \\
0 & 1 & \ldots & 0 & -a_{i, t_i-2}(s) \\
\vdots & & \ddots & \ddots & \ddots \\
0 & 0 & \ldots & 1 & -a_{i 1}(s)
\end{bmatrix}
\] (3.3.5)

for \( t_i > 1 \) with

\[
C(\Delta_i) \overset{\Delta}{=} -a_{i 1}(s) \quad \text{if } t_i = 1
\] (3.3.6)
then a transformation matrix $E(s)$ exists such that

$$G(s) = E(s) \cdot Q(s) \cdot E(s)^{-1}$$  \hspace{1cm} (3.3.7)

where $Q(s)$ is a unique block diagonal matrix, which is called the **irreducible rational canonical form** of $G(s)$ and is given by

$$Q(s) \triangleq \begin{bmatrix} \text{diag} [C(A_1), C(A_2), \ldots, C(A_z)] \end{bmatrix}$$  \hspace{1cm} (3.3.8)

It is clear that given $Q(s)$ the irreducible factors $A_i(g,s)$ can easily be obtained. A proposed method for finding $Q(s)$ for any given $G(s)$ is presented in appendix 2.

**3.3-1 Poles and zeros of a characteristic gain function**

Consider the defining equation for a characteristic gain function $g(s)$:

$$\Phi(g,s) = b_0(s)g^t + b_1(s)g^{t-1} + \ldots + b_t(s) = 0$$  \hspace{1cm} (3.3.9)

We will take both

$$b_0(s) \neq 0 \quad \text{and} \quad b_t(s) \neq 0$$

since, if either or both of these polynomial coefficients were to vanish, we could find a reduced-order equation such that both the coefficients of the highest/zeroth powers of $g(s)$ were non-zero; this reduced-order equation would then be taken as defining an appropriate new algebraic function for whose defining equation the supposition would be true.

It may happen however that $b_0(s)$ and $b_t(s)$ share a common factor and thus both vanish together at some specific set of values of $s$. Before looking at the effect of this, consider the situation when $b_0(s)$ and $b_t(s)$ do not share a common factor. The algebraic function will obviously be zero when

$$b_t(s) = 0$$  \hspace{1cm} (3.3.10)
and will tend to infinity as

\[ b_0(s) \to 0 \]  \hspace{1cm} (3.3.11)

For this reason those values of \( s \) which satisfy equation (3.3.10) are defined to be the zeros of the algebraic function \( g(s) \), and those values of \( s \) which satisfy the equation

\[ b_0(s) = 0 \]  \hspace{1cm} (3.3.12)

are defined to be the poles of the algebraic function \( g(s) \).

Unless stated otherwise the terminology 'poles and zeros' should be taken as referring only to finite poles and zeros.

The point \( s = \infty \) requires special attention and is dealt with at the end of this sub-section.

In order to be able to take equations (3.3.10) and (3.3.12) as defining the zeros and poles of \( g(s) \) in the general case, we must show that they remain appropriate when \( b_0(s) \) and \( b_t(s) \) share a common factor. Let us first dispose of the trivial case when all the coefficients \( \{b_i(s) : i=0,2,\ldots,t\} \) share a common factor by saying that such a common factor would simply be divided out to get a new defining equation for an appropriate algebraic function.

Suppose then that \( b_0(s) \) and \( b_t(s) \) have a common factor, but that some non-empty set of coefficients \( \{b_u(s), b_{u+1}(s), \ldots, b_v(s)\} \) do not share this common factor. Then dividing through the left-hand side of equation (3.3.9) by \( b_0(s) \) we get

\[ g^t + \frac{b_1(s)}{b_0(s)} g^{t-1} + \cdots + \frac{b_u(s)}{b_0(s)} g^{t-u} + \cdots + \frac{b_v(s)}{b_0(s)} g^{t-v} + \cdots + \frac{b_t(s)}{b_0(s)} = 0 \]  \hspace{1cm} (3.3.13)

Then, as \( s \to \tilde{s} \) where \( \tilde{s} \) is a zero of the common factor of \( b_0(s) \) and \( b_t(s) \), the moduli of the coefficient set
all become arbitrarily large, and it is obvious that \( g(s) \) will have a pole at \( s = \tilde{s} \).

Again, suppose that \( b_o(s) \) and \( b_t(s) \) have a common factor but that some non-empty set of coefficients \( \{b_j(s), \ldots, b_m(s)\} \) do not. Then as \( s \to \tilde{s} \) where \( \tilde{s} \) is a zero of the common factor, the algebraic equation (3.3.9) may be replaced by

\[
b_j(\tilde{s}) g^{t-j}(\tilde{s}) + \ldots + b_m(\tilde{s}) g^{t-m}(\tilde{s}) = 0	ag{3.3.14}
\]

where

\[
b_j(\tilde{s}) \neq 0, \ldots, b_m(\tilde{s}) \neq 0
\]

so that we must have

\[
g(\tilde{s}) = 0
\]

showing that \( \tilde{s} \) is indeed a zero of the algebraic function \( g(s) \).

We thus conclude that equations (3.3.10) and (3.3.12) may be taken as defining the finite zeros and finite poles of the algebraic function \( g(s) \), and that use of these definitions enables us to cope with the existence of coincident poles and zeros. The pole and zero polynomials of \( g(s) \), denoted by \( p_g(s) \) and \( z_g(s) \), are defined as

\[
p_g(s) \triangleq b_o'(s)
\]

and

\[
z_g(s) \triangleq b_t'(s)
\]

where \( b_o'(s) \) and \( b_t'(s) \) are the monic polynomials obtained from \( b_o(s) \) and \( b_t(s) \) respectively, by dividing each polynomial by its leading coefficient.
For the purpose of considering \( g(s) \) at the point \( s = \infty \), we put

\[
  s = z^{-1} \quad \text{(3.3.16)}
\]

so that

\[
  \mathcal{G}(g,s) = \mathcal{G}(g,z^{-1}) = z^{-q} \mathcal{Y}(g,z) \quad \text{(3.3.17)}
\]

where \( q \) is the number of finite poles of \( g(s) \). In any neighbourhood of the value \( z = 0 \) (the point \( z = 0 \) itself being excluded from it) the equation \( \mathcal{G}(g,s) = 0 \) is equivalent to the equation \( \mathcal{Y}(g,z) = 0 \). Therefore if we consider the equation

\[
  \mathcal{Y}(g,z) = c_0(z)g^t + c_1(z)g^{t-1} + \ldots + c_t(z) = 0 \quad \text{(3.3.18)}
\]

it follows that:

(i) \( s = \infty \) is a pole of the characteristic gain function \( g(s) \) if and only if \( c_0(0) = 0 \)

(ii) \( s = \infty \) is a zero of the characteristic gain function \( g(s) \) if and only if \( c_t(0) = 0 \)

For an open-loop gain matrix \( G(s) \) describing a physically realizable system, which by definition (see section 2.1) we are considering here, it is not possible for \( g(s) \) to have poles at infinity. In fact it is easy to show that for \( s = \infty \) the values of the characteristic gain function \( g(s) \) are simply the eigenvalues of \( D \).

3.3-2 Algebraic definition of poles and zeros for a transfer function matrix

Let \( T(s) \) be an \( m \times \ell \) rational matrix-valued function of the complex variable \( s \). Then there exists a canonical form for \( T(s) \), the Smith-McMillan form \( [3] M(s) \), such that

\[
  T(s) = H(s)M(s)J(s) \quad \text{(3.3.19)}
\]
where the \( m \times m \) matrix \( H(s) \) and the \( \ell \times \ell \) matrix \( J(s) \) are both unimodular (that is having a constant value for their determinants, independent of \( s \)). If \( r \) is the normal rank of \( T(s) \) (that is \( T(s) \) has rank \( r \) for almost all values of \( s \)) then \( M(s) \) has the form

\[
M(s) = \begin{bmatrix}
M^*(s)_{rr} & 0_{r,\ell-r} \\
0_{m-r,r} & O_{m-r,m-r}
\end{bmatrix}
\]

with

\[
M^*(s) = \text{diag} \left[ \frac{\varepsilon_1(s)}{\psi_1(s)}, \frac{\varepsilon_2(s)}{\psi_2(s)}, \ldots, \frac{\varepsilon_r(s)}{\psi_r(s)} \right]
\]

where:

(i) each \( \varepsilon_i(s) \) divides all \( \varepsilon_{i+j}(s) \) and

(ii) each \( \psi_i(s) \) divides all \( \psi_{i-j}(s) \).

With an appropriate partitioning of \( H(s), M(s) \) and \( J(s) \) we therefore have

\[
T(s) = H_1(s) \text{diag} \left[ \frac{\varepsilon_1(s)}{\psi_1(s)}, \frac{\varepsilon_2(s)}{\psi_2(s)}, \ldots, \frac{\varepsilon_r(s)}{\psi_r(s)} \right] J_1(s)
\]

\[
= \sum_{i=1}^{r} h_i(s) \frac{\varepsilon_i(s)}{\psi_i(s)} j_i^+(s)
\]

where:

(i) \( \{ h_i(s) : i = 1,2,\ldots,r \} \) are the columns of the matrix \( H_1(s) \).
(ii) \[ \{ j^t_1(s) : i = 1, 2, \ldots, r \} \] are the rows of the matrix \( J_1(s) \).

We know that
\[ r \leq \min(l, m) \]
and that \( H(s) \) and \( J(s) \) are unimodular matrices of full rank \( m \) and \( k \) respectively for all \( s \). Suppose \( T(s) \) is the transfer function matrix for a system with input transform vector \( \hat{u}(s) \) and output transform vector \( \hat{y}(s) \). Then any input vector \( \hat{u}(s) \) is turned into an output vector \( \hat{y}(s) \) by

\[
\hat{y}(s) = \sum_{i=1}^{r} h_i(s) \frac{\xi_i(s)}{\psi_i(s)} \left[ j^t_1(s) \hat{u}(s) \right] \tag{3.3.24}
\]

For the single-input single-output case where
\[
\hat{y}(s) = k \frac{\xi(s)}{\psi(s)} \hat{u}(s)
\]
with \( k \) a constant, the transfer function
\[
g(s) = \frac{k \xi(s)}{\psi(s)}
\]
is defined as having zeros at those values of \( s \) where \( \xi(s) \) vanishes and poles at those values of \( s \) where \( \psi(s) \) vanishes. Thus for a non-zero \( \hat{u}(s) \) the modulus of \( \hat{y}(s) \) vanishes when \( s \) is a zero of \( g(s) \), and becomes arbitrarily large when \( s \) is a pole of \( g(s) \). A natural way therefore to characterize the zeros and poles of \( T(s) \) is in terms of those values of \( s \) for which \( \| \hat{y}(s) \| \) becomes zero for non-zero \( \| \hat{u}(s) \| \), and arbitrarily large for finite \( \| \hat{u}(s) \| \), where \( \| \cdot \| \) denotes the standard vector norm. This natural extension
of scalar case ideas leads directly to definitions of zeros and poles of $T(s)$ in terms of the Smith-McMillan form quantities

$$\varepsilon_i(s) \quad \{\frac{\varepsilon_i(s)}{\psi_i(s)}\}$$

because of the following pair of simple results.

**Zero lemma:** $\|\hat{y}(s)\|$ vanishes for $\|\hat{u}(s)\| \neq 0$ and $s$ finite if and only if some $\varepsilon_i(s)$ is zero.

**Pole lemma:** $\|\hat{y}(s)\| \rightarrow \infty$ for $\|\hat{u}(s)\| < \infty$ if and only if some $\psi_i(s) \rightarrow 0$.

These considerations lead naturally to the following definitions \[3\].

**Poles of $T(s)$:** The poles of $T(s)$ are defined to be the set of all zeros of the set of polynomials $\{\psi_i(s) : i = 1,2,\ldots,r\}$. In what follows we will usually denote the poles of $T(s)$ by $\{p_1,p_2,\ldots,p_n\}$ and put

$$p_T(s) = (s-p_1)(s-p_2) \cdots (s-p_n)$$  \hspace{1cm} (3.3.25)

where $p_T(s)$ is conveniently referred to as the pole polynomial of $T(s)$ and is given by

$$p_T(s) = \prod_{i=1}^{r} \psi_i(s)$$  \hspace{1cm} (3.3.26)

**Zeros of $T(s)$:** The zeros of $T(s)$ are defined to be the set of all zeros of the set of polynomials $\{\varepsilon_i(s) : i = 1,2,\ldots,r\}$. We will normally denote the zeros of $T(s)$ by $\{z_1,z_2,\ldots,z_\omega\}$ and put

$$z_T(s) = (s-z_1)(s-z_2) \cdots (s-z_\omega)$$  \hspace{1cm} (3.3.27)

where $z_T(s)$ is conveniently referred to as the zero polynomial
of $T(s)$ and is given by

$$z_T(s) = \prod_{i=1}^{r} \varepsilon_i(s)$$

(3.3.28)

It is important to remember that $z_T(s)$ and $p_T(s)$ are not necessarily relatively prime; for this reason it is wrong to simply define $z_T(s)$ and $p_T(s)$ for a square matrix $T(s)$ as the numerator and denominator polynomials of $\det T(s)$.

Rules for calculating pole polynomials and zero polynomials

The route via the Smith-McMillan form is not always convenient for the determination of the poles and zeros of $T(s)$, particularly if the calculation is being done by hand. The following rules [4] can be shown to give the same results as the Smith-McMillan definitions.

Pole polynomial rule: $p_T(s)$ is the monic polynomial obtained from the least common denominator of all non-zero minors of all orders of $T(s)$.

Zero polynomial rule: $z_T(s)$ is the monic polynomial obtained from the greatest common divisor of the numerators of all minors of $T(s)$ of order $r$ ($r$ being the normal rank of $T(s)$) which minors have all been adjusted to have $p_T(s)$ as their common denominator.

3.3-3 Relationship between algebraically defined poles/zeros of the open-loop gain matrix $G(s)$ and the poles/zeros of the corresponding set of characteristic gain functions

As a key step in the establishment of a generalized Nyquist stability criterion, it is crucially important to relate the poles and zeros defined by algebraic means to complex variable theory, and thus to the poles and zeros of
the set of characteristic gain functions.

The coefficients \( a_1(s) \) in the expansion

\[
det [gI_m - G(s)] = g^m + a_1(s)g^{m-1} + a_2(s)g^{m-2} + \ldots + a_m(s) \tag{3.3.29}
\]

are all appropriate sums of minors of \( Q(s) \) since it is well known that:

\[
det [gI_m - G(s)] = g^m - [\text{trace } G(s)]g^{m-1} + [\text{principal minors of } G(s) \text{ of order } 2]g^{m-2} - \ldots + (-1)^m \det G(s) \tag{3.3.30}
\]

and thus the pole polynomial \( p_c'(s) \) is the monic polynomial obtained from the least common denominator of all non-zero principal minors of all orders of \( G(s) \).

Now the pole polynomial \( p_G(s) \) of a square matrix \( G(s) \) is the monic polynomial obtained from the least common denominator of all non-zero minors of all orders of \( G(s) \).

Therefore, if \( e_G(s) \) is the monic polynomial obtained from the least common denominator of all non-zero non-principal minors, with all factors common to \( p_c'(s) \) removed, we have that

\[
p_G(s) = e_G(s)p_c'(s) \tag{3.3.31}
\]

Furthermore since

\[
det G(s) = a_m(s) = \frac{b_m(s)}{b_o(s)} \tag{3.3.32}
\]

and since from the Smith-McMillan form for \( G(s) \)

\[
det G(s) = \alpha. \frac{z_G(s)}{p_G(s)} \tag{3.3.33}
\]

where \( \alpha \) is a scalar quantity independent of \( s \), we must have that
\[ z_G(s) = e_G(s) b_m^I(s) \]  \hspace{1cm} (3.3.34)

In many cases the least common denominator of the non-zero non-principal minors of \( G(s) \) will divide \( b_0(s) \), in which case \( e_G(s) \) will be unity and the pole and zero polynomials for \( G(s) \) will be \( b_0^I(s) \) and \( b_m^I(s) \) respectively. In general a square-matrix-valued function of a complex variable \( G(s) \) will have a set of \( l \) irreducible characteristic gain functions in the form specified by equation (3.3.3) and the general form for the pole and zero polynomials can be written as

\[ p_G(s) = e_G(s) \prod_{i=1}^{l} b_{i,0}^I(s) \]  \hspace{1cm} (3.3.35)

and

\[ z_G(s) = e_G(s) \prod_{i=1}^{l} b_{i,t_i}^I(s) \]  \hspace{1cm} (3.3.36)

where the pole and zero polynomials for the \( j \)th characteristic gain function \( g_j(s) \) are \( b_{j,0}^I(s) \) and \( b_{j,t_j}^I(s) \) respectively.

**Example demonstrating the pole-zero relationships**

Let

\[ G(s) = \frac{1}{(s+1)(s+2)(s-1)} \begin{bmatrix} (s-1)(s+2) & 0 \\ -(s+1)(s+2) & (s-1)(s+1) \end{bmatrix} \]

\[ = \begin{bmatrix} \frac{1}{s+1} & 0 \\ -\frac{1}{s-1} & \frac{1}{s+2} \end{bmatrix} \]

The pole polynomial for \( G(s) \) is obviously

\[ p_G(s) = (s+1)(s+2)(s-1) \]

and consequently the zero polynomial is

\[ z_G(s) = (s-1) \].
The characteristic equation for \( G(s) \) is
\[
\det \left[ gI - G(s) \right] = (g - \frac{1}{s+1})(g - \frac{1}{s+2}) = 0
\]
so that the irreducible characteristic equations are
\[
\Delta_1(g,s) = g - \frac{1}{s+1} = 0
\]
and
\[
\Delta_2(g,s) = g - \frac{1}{s+2} = 0
\]
which may be written as
\[
(s+1)g - 1 = 0
\]
and
\[
(s+2)g - 1 = 0
\]
Therefore the pole and zero polynomials for the characteristic gain functions \( g_1(s) \) and \( g_2(s) \) are
\[
p_{g_1}(s) = b_{10}(s) = (s+1) \quad z_{g_1}(s) = b_{11}(s) = 1
\]
\[
p_{g_2}(s) = b_{20}(s) = (s+2) \quad z_{g_2}(s) = b_{21}(s) = 1
\]
Now for \( G(s) \) the monic polynomial obtained from the least common denominator of all non-zero non-principal minors with all factors common to \( b_{10}'(s) \) \( (=b_{10}'(s)b_{20}'(s)) \) removed is given by
\[
e_{G}(s) = (s-1)
\]
which verifies the relationships
\[
p_{G}(s) = e_{G}(s) \prod_{i=1}^{2} b_{10}'(s)
\]
and
\[
z_{G}(s) = e_{G}(s) \prod_{i=1}^{2} b_{11}'(s)
\]
3.3-4 Riemann surface of a characteristic gain function

A characteristic gain function \( g(s) \) is defined by an irreducible equation of the form
\[ b_0(s)g^t + b_1(s)g^{t-1} + \ldots + b_t(s) = 0 \] (3.3.37)

having in general \( t \) distinct finite roots. An exception occurs only if

(a) \( b_0(s) = 0 \), because the degree of the equation is then lowered, and as \( b_0(s) \to 0 \) one or more of the roots becomes infinite; or if

(b) the equation has multiple roots.

This last situation can occur for finite values of \( s \) if, and only if, an expression, called the discriminant of the equation, vanishes. The discriminant \([5]\) is an entire rational function of the equation coefficients; it will be denoted by \( D_g(s) \), and is discussed in appendix 3.

**Ordinary points of the characteristic gain function**

An ordinary point \([1;6]\) of the characteristic gain function \( g(s) \) is any finite point of the complex plane such that \( b_0(s) \neq 0 \) and \( D_g(s) \neq 0 \).

**Critical points of the characteristic gain function**

A critical point \([1;6]\) of \( g(s) \) is any point of the complex plane at which either

\[ b_0(s) = 0 \quad \text{or} \quad D_g(s) = 0, \]

or both, plus the point \( s = \infty \).

**Branch points of the characteristic function**

Solutions of

\[ D_g(s) = 0 \]

are called finite branch points of the characteristic gain function. The point at infinity is a branch point if the discriminant \( D_g(z) \) of equation (3.3.18) satisfies \( D_g(0) = 0 \). At every ordinary point the equation (3.3.37) defining the
characteristic gain function has \( t \) distinct roots, since the discriminant does not vanish. The theory of algebraic functions [1] then shows that in a simply connected region of the complex plane punctured by the exclusion of the critical points the values of the characteristic gain function \( g(s) \) form a set of analytic functions; each of these analytic functions is called a branch of the characteristic gain function \( g(s) \). Arguments based on standard techniques of analytic continuation, together with the properties of algebraic equations, show that the various branches can be organized into a single entity: the corresponding algebraic function. This is summarized in the following basic theorem of algebraic function theory: an irreducible algebraic equation of the form (3.3.37) defines precisely one \( t \)-valued regular function \( g(s) \) in the punctured plane \([7]\).

Functions defined in this way are called algebraic functions, and can be regarded as natural generalizations of the familiar elementary functions of a complex variable. An elementary function of a complex variable has the set of complex numbers \( \mathbb{C} \) as both its domain and its range. An algebraic function has the complex number set \( \mathbb{C} \) as its range but has a new and appropriately defined domain \( \mathbb{R} \) which is called its Riemann Surface \([8]\). Since the Riemann surface of an algebraic function plays a crucial role in this work it is important to have an intuitive grasp of the ideas underlying its definition and formation, which is therefore now briefly considered.
Figure 6. Analytic continuation

Suppose we have a representation of part of one branch of an algebraic function in the form of a power series; such a representation is usually called a functional element. Imagine its circle of convergence to be cut out of paper and that the individual points of the paper disc are made bearers of the unique functional values of the elements. If now this initial element is analytically continued by means of a second power series, another circle of convergence can be thought of as being cut out and pasted partly over the first, as illustrated by figure 6. The parts pasted together are made bearers of the same functional values and are accordingly treated as a single region covered once with values. If a further analytic continuation is carried out, a further disc is similarly pasted on to the preceding one. Now suppose that, after repeated analytic continuations, one of the discs lies over another disc, not associated with an immediately
preceding analytic continuation, as shown in figure 7. Such an overlapping disc is pasted together with the one it overlaps if and only if both are bearers of the same functional values. If, however, they bear different functional values they are allowed to overlap but remain disconnected. Thus two sheets, which are bearers of different functional values, become superimposed on this part of the complex plane.

![Figure 7. Repeated analytic continuation](image)

Continuing this process for as long as possible, a surface-like configuration is obtained covering $t$ "sheets" of the complex plane, where $t$ is the degree of the algebraic function. To form the Riemann surface these sheets can be joined together in the most varied of ways. This may involve connecting together two sheets which are separated by several other sheets lying between them. Although such a construction cannot be carried out in a three-dimensional space it is not difficult to give a perfectly satisfactory topological description of the process required. This surface-like configuration is called the Riemann surface of the multiple-valued algebraic function. On the Riemann
surface the entire domain of values of the algebraic function is spread out in a completely single-valued manner so that, on every one of the t copies of the complex plane involved, every point is the bearer of one and only one value of the function.

A method for building Riemann surfaces is given in appendix 4. This involves the use of cuts in the complex plane and it may be helpful to say a word about them at this point. Let an algebraic function $g(s)$ have r critical points $\{a_1, a_2, \ldots, a_r\}$. Suppose them to be joined to one another and then to the point at infinity by a line $L$. Any line joining critical points will be called a cut. Let $L$ denote the set of complex numbers defined by the line $L$. We then have that the solutions of equation (3.3.37) define a set of t "distinct" analytic functions $\{\tilde{g}_1(s), \tilde{g}_2(s), \ldots, \tilde{g}_t(s)\}$ in the cut plane $\mathbb{C} - L$. Each of these functions can be analytically continued, by standard procedures, across the cut $L$. Now it follows from the fundamental principles of analytic continuation that if an analytic function satisfies an algebraic equation in one part of its domain of definition, it must satisfy that equation in every region into which it is analytically continued. We must therefore have that:

(i) there are only t "distinct" analytic functions which satisfy the defining algebraic equation in the cut plane $\mathbb{C} - L$;

(ii) each analytic continuation of any of these analytic functions $\{\tilde{g}_i(s) : i = 1, 2, \ldots, t\}$ gives rise to an
analytic function which also satisfies the defining algebraic equation. It follows from this that the set of analytic functions associated with one side of the cut \( L \) must be a simple permutation of the set of analytic functions associated with the other side of the cut. Therefore by identifying and suitably matching up corresponding analytic functions (via their sets of computed values) on opposite sides of the cut \( L \), one can produce an appropriate domain on which a single analytic function may be specified which defines a continuous single-valued mapping from this domain into the complex plane. This function is of course the algebraic function, conceived of as a single entity, and the domain so constructed is its Riemann surface.

It is sufficient for the purposes of understanding this book for the reader to know that a Riemann surface can be constructed for any given algebraic function, on which its values form a single-valued function of position. Many standard relationships and properties of analytic function theory generalize, using the Riemann surface concept, to the algebraic function case and, in particular the Principle of the Argument holds on the Riemann surface; an extension of the Principle of the Argument is developed in appendix 5.

The Riemann surface which is the domain of the characteristic gain function \( g(s) \) will be called the **frequency surface** or **s-surface**. When the open-loop gain matrix \( G(s) \) is \( mxm \) and has a corresponding characteristic equation which is irreducible (i.e. the usual case in practice) the frequency surface is formed out of \( m \) copies of the complex
frequency plane or s-plane.

3.3-5 Generalized root locus diagrams

The characteristic gain function \( g(s) \) is a function of a complex variable whose poles and zeros are located on the frequency surface domain. It is convenient to exhibit the nature of \( g(s) \) by drawing constant phase and constant magnitude contours of \( g(s) \) on the frequency surface. If the computational method outlined in appendix 4 is used to construct the surface then the superposition of constant phase and magnitude contours is clearly a simple process. The frequency surface can be thought of as the set of all possible closed-loop characteristic frequencies associated with all possible values of the complex gain parameter \( g \). When the surface is characterized by constant phase and magnitude contours of \( g(s) \) we have a direct correspondence between a closed-loop characteristic frequency and an open-loop gain, and since the surface is constructed from \( m \) copies of the complex frequency plane, for each value of \( s \) there are \( m \) corresponding characteristic gains.

From equation (3.1.4) we have

\[
g(s) = -\frac{1}{k}
\]

(3.3.38)

so that the variation of the closed-loop poles (characteristic frequencies) with the real control variable \( k \) traces out loci which are equivalent to the 180° phase contours of \( g(s) \). Equation (3.3.38) is a direct generalization of the defining equation for the single-loop root locus diagram. The 180° phase contours of \( g(s) \) are the multivariable root loci i.e. the variation of the closed-loop poles with the gain control variable \( k \). The fact that multivariable root loci 'live'
on a Riemann surface explains their complicated behaviour
as compared with the single-input, single-output case
where the root loci lie on a simple complex plane (a trivial,
i.e. one sheeted, Riemann surface). The multivariable
root loci will sometimes be referred to as the characteristic
frequency loci.

Recall that in section 3.2 it was pointed out that the
characteristic equations for \( G(s) \) and \( S(g) \) are in general
different in that the equation for \( S(g) \) may contain factors
of \( s \) which are independent of \( g \). These factors therefore
correspond to closed-loop poles which are independent of \( g \),
or equivalently independent of the gain control variable \( k \);
and, from the root locus point of view, these factors
correspond to degenerate loci each consisting of a single
point. The degenerate loci are therefore not picked out by
the 180° phase contours of \( g(s) \) on the frequency surface.
In practice the characteristic frequency loci are generated
as the set of loci in a single copy of the complex frequency
plane traced out by the eigenvalues of \( S(g) \) as \( g \) traverses
the negative real axis in the gain plane. This approach
automatically picks out the degenerate loci. In common
with the classical root locus approach of Evans the characteristic
frequency loci are usually calibrated in terms of the gain
control variable \( k=-g^{-1} \).

3.3-6 Example of frequency surface and characteristic
frequency loci

As an illustrative example consider the general multi-
variable feedback configuration of figure 3 with a corresponding
open-loop gain matrix
Figure 8. Sheet 1 of the frequency surface

Figure 9. Sheet 2 of the frequency surface
The matrix is of order two and therefore the appropriate surface will be constructed from two sheets of the complex s-plane. The two sheets are shown characterized by constant phase and magnitude contours of g(s) in figures 8 and 9. The cuts, identifiable by discontinuities in the contours, are represented by thick black lines; and the characteristic frequency loci, which are the 180° phase contours of g(s), are identified by a diamond symbol. The characteristic frequency loci indicate that variation of the gain control parameter k, upwards from zero, causes the system to experience stability, instability and stability again. This phenomenon is clearly linked with the presence of a branch point in the right half-plane (at $s=\frac{1}{24}$).

Note that since we have completely characterized the feedback configuration by its open-loop gain matrix there are no unobservable or uncontrollable modes.

3.4 Characteristic frequency functions

The natural way to define the characteristic frequency function $s(g)$ is via the characteristic equation

$$\nabla(s,g) \triangleq \det[sI_n - S(g)] = 0 \quad (3.4.1)$$

It is an algebraic function and the detailed study of the characteristic gain function presented in the previous section can be applied directly to it with the roles of $s$ and $g$ reversed.

The Riemann surface which is the domain of the characteristic frequency function will be called the gain surface or $g$-surface. It is formed out of $n$ copies of the complex
gain plane or g-plane since there are n values of closed-loop characteristic frequency (closed-loop poles) for every value of g. The gain surface can be thought of as the set of all possible open-loop characteristic gains of the open-loop gain matrix G(s) associated with all possible closed-loop characteristic frequencies. In a similar fashion to the gain function g(s) it is convenient to exhibit the behaviour of s(g) on the gain surface by superimposing constant phase and magnitude contours of s(g) onto the surface. Like g(s) the frequency function s(g) has poles and zeros but their significance is quite different.

3.4-1 Generalized Nyquist diagram

Each 'sheet' of a gain surface characterized by constant phase and magnitude contours of s(g) is divided into regions corresponding to left half-plane and right half-plane closed-loop characteristic frequencies. Therefore given such a calibrated surface one can see at a glance which values of g (or equivalently k) correspond to stable closed-loop poles. The boundary between stable and unstable regions is clearly the ±90° phase contours of s(g). The ±90° phase contours of s(g) are a natural generalization of the single-loop Nyquist diagram and are called characteristic gain loci.

In practice the characteristic gain loci are generated as the loci in the complex gain plane traced out by the eigenvalues of G(s) as s traverses the so called Nyquist D-contour in the s-plane. Suppose that we consider a portion of the imaginary axis. We can then compute a set of loci corresponding to the eigenvalues \( \tilde{g}_1(j\omega), \ldots, \tilde{g}_m(j\omega) \) (where in this context \( j = \sqrt{-1} \))
in the following way:

(i) Select a value of angular frequency, say $\omega_a$.

(ii) Compute the complex matrix $G(j\omega_a)$.

(iii) Use a standard computer algorithm to compute the eigenvalues of $G(j\omega_a)$, which are a set of complex numbers denoted by $\{\tilde{\gamma}_i(j\omega_a)\}$.

(iv) Plot the numbers $\{\tilde{\gamma}_i(j\omega_a)\}$ in the complex plane.

(v) Repeat with further values of angular frequency $\omega_b, \omega_c, \ldots$ etc., and join the resulting plots up into continuous loci using a sorting routine based on the continuity of the various branches of the characteristic functions involved.

For the purpose of developing a generalized Nyquist stability criterion in chapter 4 the Nyquist D-contour is traversed in the standard clockwise direction.

3.4-2 Example of gain surface and characteristic gain loci

As an illustrative example consider the open-loop gain matrix considered in subsection 3.3-5 which has a minimal state-space realization

$$A = \begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0.6 \\ 1 & 0.5 \end{bmatrix}$$

$$C = \begin{bmatrix} 2.4 & -1.6 \\ 4.8 & -4.8 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

The system has two states and therefore the appropriate gain surface will be constructed from two sheets of the complex $g$-plane. The two sheets are shown characterized by constant phase and magnitude contours of $s(g)$ in figures 10 and 11. The characteristic gain loci, which are the $+90^\circ$ phase contours of $s(g)$, are denoted by a series of crosses.
Figure 10. Sheet 1 of the gain surface

Figure 11. Sheet 2 of the gain surface
Figure 12. Sketch of figure 10 emphasizing right half and left half-plane regions.

Figure 13. Sketch of figure 11 emphasizing right half and left half-plane regions.
In figures 12 and 13 sketches of figures 10 and 11 are made to emphasize the right and left half-plane regions of closed-loop poles. From these sketches it is easy to infer bounds on the gain control parameter \( k \) for stability. Since \( g = -k^{-1} \), as we increase \( k \) positively from zero the critical value of \( g \) moves from \( -\infty \) along the real axis towards the origin on each sheet. On sheet 1, \( g \) is in a right half-plane region for \( 1.25 < k < 2.5 \) while on sheet 2, for positive \( k \), \( g \) never moves into a right half-plane region. Therefore the closed-loop system is stable for \( 0 < k < 1.25 \) and \( 2.5 < k < \infty \).

If \( k \) is increased negatively from zero the critical value of \( g \) moves from \( \infty \) along the real axis towards the origin on each sheet. On both sheets the value of \( g \) is in right half-plane regions for \( -\infty < k < -1.875 \). Therefore for negative \( k \), which corresponds to positive feedback, the closed-loop system is stable for \( -1.875 < k < 0 \).

If the calibrated gain surface is projected onto the complex gain plane \( \mathbb{C} \), we have the normal representation of the characteristic gain loci, plus the superposition of contours representing both right half-plane and left half-plane regions. Stability can now be predicted by considering the right half-plane regions in relation to a single critical point \( \left( \frac{1}{k} + j0 \right) \). However, this presentation will in general be difficult to comprehend because of the overlapping of contours. Therefore, although the counting of encirclements in the generalized Nyquist stability criterion [chapter 4] is not fundamental to system stability as pointed out by Saeks [10], it does afford the simplest method of predicting
closed-loop stability in the gain plane $\mathbb{C}$.

From a gain surface plot it is possible to determine the closed-loop poles and hence the relative stability of a closed-loop system. This is now illustrated by finding the dominant closed-loop poles for the example under consideration with unity $k$. It is convenient for this purpose to have the gain surface characterized by constant real and imaginary contours as shown in figures 14 and 15.

Let the dominant closed-loop poles be

$$s_d = \alpha + j\beta$$

Then $\alpha$ is the smallest (in magnitude) negative real contour that passes through any one of the critical, $-1$, values of $g$, and $\beta$ is the corresponding imaginary contour. From figures 14 and 15 we have

$\alpha \approx -0.05$ and $\beta = 0$

so that

$$s_d \approx -0.05$$

By hand calculation the dominant closed-loop pole is

$$s_d = -0.0528$$

Also from a gain surface it is possible to determine the characteristic frequency loci. Apart from possible single-point loci, the root loci are simply the values of $s$ at which the characteristic gain loci has a phase of $180^\circ$. Therefore from a gain surface plot the root loci are determined by the values of the constant contours as they cross the negative real axis on each sheet. Similarly, from a calibrated frequency surface, it is possible to determine the characteristic gain loci from the values of constant contours as they cross the imaginary axes.
Figure 14. Sheet 1 of the gain surface (a) 'complete' sheet (b) small region about -0.2
Figure 15. Sheet 2 of the gain surface

References

4. A generalized Nyquist stability criterion

The Nyquist stability criterion [1] is one of the most fundamental results in the theory of linear feedback systems and its generalization to the multivariable case is of great interest. Such a generalization was put forward by MacFarlane[2] and used as part of a technique called the Characteristic Locus Method [3] for feedback systems analysis and design, but no satisfactory proof was supplied. The proof of a generalized Nyquist stability theorem was undertaken by Barman and Katzenelson [4;5] but their approach ignored certain key algebraic properties of the quantities involved; it also leaned very heavily on the use of cuts in the complex plane; this made their treatment technically complicated, and obscured the essential simplicity of the result. The purpose of this chapter is to give a rigorous proof of a generalized Nyquist-like stability criterion for the general feedback configuration based on a fundamental result in complex variable theory: the Principle of the Argument applied to an algebraic function defined on an appropriate Riemann surface [appendix 5].

Two Nyquist-like stability tests are in fact stated and proved, the usefulness of each depending on how the subsystems are characterized. The two statements of the criterion are given in section 4.1 and proved in section 4.2.

4.1 Generalized Nyquist stability criterion

If the subsystems of the general feedback configuration of figure 3 are each characterized by a state-space model then the following statement of the criterion is applicable.
Statement 1. The general feedback configuration is closed-loop stable if and only if:

1. The net sum of anti-clockwise encirclements of the critical point \((-\frac{1}{k}+j0)\) by the set of characteristic gain loci is equal to the number of right half-plane poles of \(G(s)\);
2. The characteristic gain loci do not pass through the critical point \(-\frac{1}{k}+j0\);
3. The number of branches of the characteristic gain loci passing through infinity is equal to the number of poles of \(G(s)\) on the imaginary axis; and
4. The eigenvalues of the A-matrix, of the open-loop system \(S(A,B,C,D)\), which correspond to modes of the system which are unobservable and/or uncontrollable from the point of view of considering the input as that of the first subsystem and the output as that of the hth subsystem, are all in the left half-plane.

If the subsystems are completely characterized by their transfer function matrices, or if it is known that for each subsystem there are no unobservable and/or uncontrollable modes in the right half-plane including the imaginary axis, then the following statement of the criterion applies.

Statement 2. The general feedback configuration is closed-loop stable if and only if:

1. The net sum of anti-clockwise encirclements of the critical point \((-\frac{1}{k}+j0)\) by the set of characteristic gain loci is equal to the total number of right half-plane poles of \(G_1(s), G_2(s), \ldots, G_n(s)\);
(2) the characteristic gain loci do not pass through the critical point \(-\frac{1}{K} + j0\); and
(3) the number of branches of the characteristic gain loci passing through infinity is equal to the number of poles of \(G(s)\) on the imaginary axis.

Note that if condition (2) and/or condition (3) do not hold the closed-loop system has one or more poles on the imaginary axis and is therefore not input-output stable although the equilibrium state at the origin may be stable.

4.2 Proof of the generalized Nyquist stability criterion

In section 2.4 it was shown how the return-difference operator corresponding to the break point shown in figure 4 is related to the open- and closed-loop characteristic polynomials by the following expression.

\[
\frac{\det F(s)}{\det F(\infty)} = \frac{CLCP(s)}{OLCP(s)}
\]  

(4.2.1)

This fundamental relationship is the foundation on which the proof of the generalized Nyquist stability criterion is based.

The first stage in the proof is to consider the eigenvalue equation of the return-difference matrix \(F(s)\), that is

\[
\det[f_{\text{m}}^* - F(s)] = 0
\]  

(4.2.2)

which in general, as for the characteristic equation of \(G(s)\), can be expressed as a product of irreducible algebraic equations of the form

\[
d_{i0}(s)f_1^{t_i} + d_{i1}(s)f_1^{t_i-1} + \ldots + d_{i\ell}(s) = 0
\]  

(4.2.3)

defining a set of algebraic functions \(\{f_i(s) : i=1,2,\ldots,\ell\}\).
Therefore, as in sub-section 3.3-3 where the poles and zeros of \( G(s) \) are related to the poles and zeros of the characteristic gain functions, the pole and zero polynomials of \( F(s) \) can be related to the pole and zero polynomials of the algebraic functions \( \{f_i(s) : i = 1, 2, \ldots, l \} \) as follows

\[
p_F(s) = e_F(s) \prod_{i=1}^{l} d_{i0}(s)
\]

and

\[
z_F(s) = e_F(s) \prod_{i=1}^{l} d_{it_i}(s)
\]

By definition the open-loop characteristic polynomial of the general feedback configuration is given by

\[
\text{OLCP}(s) \triangleq \text{det}[sI_n - A] = \text{det}[sI_{n_1} - A_1] \text{det}[sI_{n_2} - A_2] \ldots \text{det}[sI_{n_h} - A_h]
\]

and it is easily shown [6] that

\[
\text{det}[sI_{n_i} - A_i] = p_{G_i}(s)p_{d_i}(s)
\]

where \( p_{G_i}(s) \) is the pole polynomial for the transfer function matrix \( G_i(s) \) and the monic polynomial \( p_{d_i}(s) \) has as its zeros the decoupling zeros of the ith subsystem associated with that set of characteristic frequencies (eigenvalues) of \( A_i \) which correspond to modes of the ith subsystem which are uncontrollable and/or unobservable. The open-loop characteristic polynomial can therefore be expressed as

\[
\text{OLCP}(s) = p_{G_1}(s)p_{G_2}(s) \ldots p_{G_h}(s)p_{d_1}(s)p_{d_2}(s) \ldots p_{d_h}(s)
\]

The pole polynomial \( p_G(s) \) for the open-loop gain matrix \( G(s) \) is related to the pole polynomials of the subsystem transfer functions through the relationship
\[ p_G(s)p_X(s) = p_{G_1}(s)p_{G_2}(s) \cdots p_{G_n}(s) \quad (4.2.8) \]

where \( p_X(s) \) has as its zeros those poles of \( G_1(s), G_2(s), \ldots, \) and \( G_n(s) \) which are lost when \( G(s) \) is formed \[7\]. The zeros of \( p_X(s) \) are in fact a subset of the unobservable and uncontrollable modes of the system \( S(A,B,C,D) \) where the input is that of the first subsystem and the output is that of the \( h \)th subsystem. The complete set of unobservable and uncontrollable modes of \( S(A,B,C,D) \) is the set of zeros of the polynomial \( p_d(s) \) where

\[ p_d(s) = p_X(s)p_{d_1}(s)p_{d_2}(s) \cdots p_{d_h}(s) \quad (4.2.9) \]

The open-loop characteristic polynomial can therefore be rewritten as

\[ \text{OLCP}(s) = p_G(s)p_d(s) \quad (4.2.10) \]

or

\[ \text{OLCP}(s) = p_G(s)p_X(s)p_{d_1}(s)p_{d_2}(s) \cdots p_{d_h}(s) \quad (4.2.11) \]

and if we combine expression (4.2.10) with the fundamental relationship (4.2.1) we obtain

\[ \text{CLCP}(s) = p_d(s)p_G(s) \frac{\det F(s)}{\det F(\infty)} \quad (4.2.12) \]

Now from the Smith-McMillan canonical form for \( F(s) \) we have that

\[ \det F(s) = \beta \frac{z_F(s)}{p_F(s)} \quad (4.2.13) \]

where \( \beta \) is a scalar quantity independent of \( s \); and from the structure of \( F(s) \), equation (3.1.2), it is clear that the monic polynomials \( z_F(s) \) and \( p_F(s) \) will be of the same order and hence that

\[ \det F(\infty) = \beta \quad (4.2.14) \]

Therefore combining equations (4.2.12-14) the closed-loop characteristic polynomial is given by
\[ CLCP(s) = p_{d}(s)p_{G}(s) \frac{z_{F}(s)}{p_{F}(s)} \]  

(4.2.15)

If we now substitute from equations (4.2.4) and (3.3.35) this expression can be rewritten as follows:

\[ CLCP(s) = p_{d}(s)e_{G}(s) \prod_{i=1}^{\ell} b'_{io}(s) \prod_{i=1}^{\ell} d'_{it_{i}}(s) \frac{\prod_{i=1}^{\ell} d'_{io}(s)}{\prod_{i=1}^{\ell} d'_{io}(s)} \]

(4.2.16)

But using the eigenvalue shift theorem \[8\], on equation (3.1.2), we have that the characteristic functions \(f_{i}(s): i=1,2,...,\ell\) of F(s) are related to the characteristic gain functions \(g_{i}(s): i=1,2,...,\ell\) by

\[ f_{i}(s) = 1 + kg_{i}(s), \quad i=1,2,...,\ell \]

(4.2.17)

and hence that the polepolynomials for both sets of algebraic functions are identical, that is

\[ \prod_{i=1}^{\ell} d'_{io}(s) = \prod_{i=1}^{\ell} b'_{io}(s) \]

(4.2.18)

and therefore from (4.2.16) we have

\[ CLCP(s) = p_{d}(s)e_{G}(s) \prod_{i=1}^{\ell} d'_{it_{i}}(s) \]

(4.2.19)

Note that the polynomial \(\prod_{i=1}^{\ell} d'_{it_{i}}(s)\) is dependent on the gain control variable \(k\).

The relationship (4.2.19) implies \[ see section 2.3\] that the following conditions are necessary and sufficient for closed-loop stability:

(a) \(e_{o}(s) = 0\) has only left half-plane roots;

(b) \(\prod_{i=1}^{\ell} d'_{it_{i}}(s) = 0\) has only left half-plane roots; and

(c) \(p_{d}(s) = 0\) has only left half-plane roots.
The next step in the proof is to show how condition (b) can be replaced by an encirclement condition similar to that in the classical Nyquist criterion.

For the set of irreducible characteristic equations associated with the return-difference operator $F(s)$ there is a corresponding set of Riemann surfaces on which the appropriate characteristic algebraic functions $\{f_i(s) : i=1,2,...\}$ become single-valued, and mappings from these surfaces on to a corresponding $f_i$-plane are one-to-one and continuous. Let us consider the $j$th equation of the set defined by equations (4.2.3). The degree of the equation is $t_j$ and therefore the corresponding Riemann surface $R_{f_j}$ is formed by piecing together $t_j$ copies of the complex $s$-plane, $\mathbb{C}$. Suppose now that a Nyquist $D$-contour, as shown in figure 16, is drawn on each of the $t_j$ copies of $\mathbb{C}$ before they are pieced together to form $R_{f_j}$. Then when the surface is formed the set of Nyquist $D$-contours combine to form a set of closed Jordan contours enclosing right half-plane regions of $R_{f_j}$. The extended Principle of the Argument [appendix 5] can then be applied to each right half-plane region on $R_{f_j}$. Therefore for a particular right half-plane region, not necessarily simply connected but with a boundary made up from Nyquist $D$-contours, we have that the difference between the number of zeros and poles of the algebraic function $f_j(s)$ in the region, is equal to the number of clockwise encirclements of the origin in $\mathbb{C}$ (the complex $f$-plane) by the image of the boundary curves, under $f_j(s)$ for that particular region. If we therefore consider
all the right half-plane regions on $R_{f_j}$ and apply the extended Argument Principle to each, we have that

$$N(f_j, 0) = Z_{f_j} - P_{f_j} \quad (4.2.20)$$

where:

(i) $N(f_j, 0)$ is the net sum of clockwise encirclements of the origin in $\mathbb{C}$ by the set of image curves, under $f_j(s)$, of the set of curves formed on $R_{f_j}$ by piecing together the appropriate set of Nyquist D-contours when forming $R_{f_j}$;

(ii) $Z_{f_j}$ is the number of right half-plane zeros of $f_j(s)$; and

(iii) $P_{f_j}$ is the number of right half-plane poles of $f_j(s)$.

Figure 16. Theoretical Nyquist D-contour
Note that the indentations on the D-contour are necessary only from a theoretical viewpoint in the development of the Extended Principle of the Argument [appendix 5] and in practice the D-contour is taken as the imaginary axis.

Condition (b) for closed-loop stability is equivalent to saying that there are no zeros of

\[ \{ f_i(s) : i=1,2,\ldots,k \} \quad (4.2.21) \]

in the right half-plane or on the imaginary axis, and can therefore be replaced by

\[ Z_{f_i} = 0 \quad i=1,2,\ldots,k \]

or

\[ N(f_i,0) = -P_{f_i} \quad i=1,2,\ldots,k \quad (4.2.22) \]

plus the condition that the algebraic functions (4.2.21) have no zeros on the imaginary axis. Conditions (4.2.22) imply and are implied by

\[ \sum_{i=1}^{k} N(f_i,0) = -\sum_{i=1}^{k} P_{f_i} \quad (4.2.23) \]

so that the necessary and sufficient conditions for closed-loop stability can be rewritten as

(a') \( e_G(s) = 0 \) has no right half-plane roots;
(b') \( e_G(s) = 0 \) has no roots on the imaginary axis;
(c') \[ \sum_{i=1}^{k} N(f_i,0) = -\sum_{i=1}^{k} P_{f_i} \];
(d') \( \{ f_i(s) : i=1,2,\ldots,k \} \) have no zeros on the imaginary axis; and
(e') \( p_d(s) \) has only left half-plane zeros.

Now, as shown by equations (3.3.35) and (4.2.18), \( e(s) \) together with the pole polynomials for the set of characteristic functions \( \{ f_i(s) : i=1,2,\ldots,k \} \) make up the pole polynomial for \( G(s) \). This leads us to consider combining conditions
(a') and (c') into the single equivalent condition

\[ \sum_{i=1}^{\ell} N(f_i,0) = -P_G \]  \hspace{1cm} (4.2.24)

where \( P_G \) is the number of right half-plane poles of \( G(s) \).

The necessity and sufficiency of condition (4.2.24) for closed-loop stability when conditions (b'), (d'), and (e') are satisfied is proved as follows.

From equations (3.3.35) and (4.2.18) we have that

\[ P_G = e + \sum_{i=1}^{\ell} P_{f_i} \]  \hspace{1cm} (4.2.25)

where \( e \) is the number of right half-plane zeros of \( e_G(s) \), and we also know that

\[ \sum_{i=1}^{\ell} Z_{N(f_i,0)} = \sum_{i=1}^{\ell} Z_{f_i} - \sum_{i=1}^{\ell} P_{f_i} \]  \hspace{1cm} (4.2.26)

so that combining these two expressions we have

\[ \sum_{i=1}^{\ell} N(f_i,0) = e + \sum_{i=1}^{\ell} Z_{f_i} - \sum_{i=1}^{\ell} P_{f_i} \]  \hspace{1cm} (4.2.27)

where \( \sum_{i=1}^{\ell} Z_{f_i}, e \) and \( P_G \) are all positive integers, or zero.

To establish the necessity of condition (4.2.24) suppose that

\[ \sum_{i=1}^{\ell} N(f_i,0) \neq -P_G \]

then from equation (4.2.27) this implies that

\[ \sum_{i=1}^{\ell} Z_{f_i} \neq 0 \]  \hspace{1cm} \text{and/or}  \hspace{1cm} e \neq 0

and hence we conclude that condition (4.2.24) is necessary for closed-loop stability.

For sufficiency suppose that

\[ \sum_{i=1}^{\ell} N(f_i,0) = -P_G \]

then from equation (4.2.27) we must have that

\[ \sum_{i=1}^{\ell} Z_{f_i} = e = 0 \]
and hence the system is closed-loop stable. Thus the sufficiency of condition (4.2.24) is established.

We have therefore shown that the following conditions are necessary and sufficient for closed-loop stability:

(a") \( \sum_{i=1}^{l} N(f_i,0) = -P_G \);

(b") \( \{f_i(s): i=1,2,\ldots,l\} \) have no zeros on the imaginary axis;

(c") \( e_G(s) \) has no zeros on the imaginary axis; and

(d") \( P_d(s) \) has only left half-plane zeros.

From equation (4.2.17) the image curve sets in \( \mathbb{C} \) of the Nyquist D-contour set mapped under \( f_i(s) \) and \( kg_i(s) \) are simply related by a unit shift in \( \mathbb{C} \). The stability condition (a") can consequently be replaced by

\[
\sum_{i=1}^{l} N(kg_i,-1) = -P_G \tag{4.2.28}
\]

where \( N(kg_i,-1) \) is the net sum of clockwise encirclements of the point \((-1+jo)\) in \( \mathbb{C} \) by the characteristic gain loci scaled by \( k \). As in the classical Nyquist criterion the scaling by \( k \) is avoided by counting the number of encirclements that the characteristic gain loci make of the critical point \((\frac{1}{k} + jo)\), and hence replacing condition (4.2.28) by

\[
\sum_{i=1}^{l} N(g_i,\frac{1}{k}) = -P_G \tag{4.2.29}
\]

From equations (4.2.17) we also have that condition (b") is equivalent to

\( g_i(jw) \neq \frac{1}{k}, \quad i=1,2,\ldots,l \) \hspace{1cm} (4.2.30)

and in practice this corresponds to the characteristic gain loci not passing through the critical point \((\frac{1}{k}+jo)\).

Condition (c") can also be replaced by a more practical
The pole polynomial for $G(s)$ is given, equation (3.3.35), as

$$p_G(s) = e_G(s) \prod_{i=1}^{l} b_{i0}^i(s)$$

(4.2.31)

where $b_{i0}^i(s)$ is the pole polynomial for the $j$th characteristic gain function $g_j(s)$. Therefore it is clear that $e_G(s)$ will have no zeros on the imaginary axis if, and only if, the number of infinite branches of the characteristic gain loci is equal to the number of poles of $G(s)$ on the imaginary axis. Note that for a branch of the loci going off to infinity there will also be a branch returning from infinity and care should be taken not to count this as two infinite branches.

The necessary and sufficient conditions for closed-loop stability can therefore be reduced to:

1. \[ \sum_{i=1}^{l} N(g_i, \frac{1}{k}) = -P_G; \]
2. the characteristic gain loci do not pass through the critical point \((\frac{1}{k}+j0)\);
3. the number of branches of the characteristic gain loci passing through infinity is equal to the number of poles of $G(s)$ on the imaginary axis; and
4. $p_d(s)$ has only left half-plane zeros.

This completes the proof of statement 1 of the generalized Nyquist stability criterion.
If the subsystems are completely characterized by their transfer function matrices then their state-space descriptions will each be observable and controllable so that
\[ p_{d_1}(s) = p_{d_2}(s) = \ldots = p_{d_h}(s) = 1 \quad (4.2.32) \]
and hence
\[ p_d(s) = p_x(s) \quad (4.2.32) \]
Condition (4) of statement 1 of the Nyquist criterion is then equivalent to \( p_x(s) \) having only left half-plane zeros. This situation also arises if the subsystems are characterized by their state-space descriptions and we have the additional information that each subsystem has no unobservable or uncontrollable modes in the right half-plane or on the imaginary axis. In these situations conditions (1) and (4) can be combined to give a criterion which needs no information about decoupling zeros.

To show this we will first of all prove that when
\[ (i) \quad p_d(s) = p_x(s) \quad (4.2.34) \]
or
\[ (ii) \quad p_d(s) = p_x(s)p_{d_1}(s)p_{d_2}(s) \ldots p_{d_h}(s) \quad (4.2.35) \]
where the zeros of \( \{p_{d_i}(s) : i=1,2,\ldots,h\} \) are all in the left half-plane,
conditions \((a')\), \((c')\) and \((e')\) are equivalent to
\[ \sum_{i=1}^{l} N(f_i,0) = \sum_{i=1}^{h} -P_{G_i} \quad (4.2.36) \]
where \( P_{G_i} \) is the number of right half-plane poles of \( G_i(s) \).
The necessity and sufficiency of condition (4.2.36) for
closed-loop stability when conditions (b') and (d') are satisfied is proved as follows.

From equations (4.2.9), (4.2.18) and (4.2.31) we have that

\[ h \sum_{i=1}^{l} P_{G_{i}} = e + \sum_{i=1}^{l} P_{f_{i}} + P_{x} \]  \hspace{1cm} (4.2.37)

where \( P_{x} \) is the number of right half-plane zeros of \( P_{x}(s) \), and combining this with equation (4.2.26) we have that

\[ \sum_{i=1}^{l} N(f_{i}, 0) = \sum_{i=1}^{l} z_{f_{i}} + e + P_{x} - \sum_{i=1}^{l} P_{G_{i}} \]  \hspace{1cm} (4.2.38)

Note also that

\[ P_{d}(s) = P_{x}(s)P_{d_{1}}(s) \ldots P_{d_{h}}(s) \]

so that if \( P_{d} \) denotes the number of right half-plane zeros of \( P_{d}(s) \) we have that

\[ P_{d} = P_{x} \]  \hspace{1cm} (4.2.39)

and equation (4.2.38) becomes

\[ \sum_{i=1}^{l} N(f_{i}, 0) = \sum_{i=1}^{l} z_{f_{i}} + e + P_{d} - \sum_{i=1}^{l} P_{G_{i}} \]  \hspace{1cm} (4.2.40)

To establish the necessity of condition (4.2.36) suppose that

\[ \sum_{i=1}^{l} N(f_{i}, 0) \neq - \sum_{i=1}^{l} P_{G_{i}} \]

then from equation (4.2.40) this implies that

\[ \sum_{i=1}^{l} z_{f_{i}} \neq 0, \text{ or } e \neq 0, \text{ or } P_{d} \neq 0 \]

or any combination of these and thus that the system will be closed-loop unstable. Hence we conclude that condition (4.2.36) is necessary for closed-loop stability.

For sufficiency suppose that

\[ \sum_{i=1}^{l} N(f_{i}, 0) = - \sum_{i=1}^{l} P_{G_{i}} \]
then from equation (4.2.40) we must have that
\[ \sum_{i=1}^{l} f_i = e = P_d = 0 \]
and hence the system is closed-loop stable. Thus the sufficiency of condition (4.2.36) is established.

We have therefore shown that when either condition (4.2.34) or condition (4.2.35) is satisfied the following are necessary and sufficient for closed-loop stability.

(a'') \[ \sum_{i=1}^{l} N(f_i, 0) = -\sum_{i=1}^{h} P_G ; \]

(b'') \{ f_i(s) : i=1,2,\ldots,l \} have no zeros on the imaginary axis; and

(c'') e_G(s) has no zeros on the imaginary axis.

But we have already shown, in proving statement 1 of the stability criterion, that
\[ \sum_{i=1}^{l} \frac{1}{f_i} = \sum_{i=1}^{l} \frac{1}{g_i} ; \]
that condition (b'') is equivalent to the characteristic gain loci not passing through the critical point (-\(\frac{1}{k}\)+j0); and that condition (c'') is equivalent to the number of infinite branches of the characteristic gain loci being equal to the number of poles of G(s) on the imaginary axis.

This therefore completes the proof of statement 2 of the generalized Nyquist stability criterion.

4.3 Example

To illustrate the stability criterion consider the example already used in sub-sections 3.3-5 and 3.4-2, where the general feedback configuration is characterized by its open-loop gain matrix.
By definition the system has no unobservable or uncontrollable modes and by virtue of the problem statement we can consider the configuration as consisting of just one subsystem \( G_1(s) = G(s) \). Therefore either statement 1 or statement 2 of the stability criterion is directly applicable.

The pole polynomial for \( G(s) \) is
\[
P_G(s) = (s+1)(s+2)
\]
which has no zeros in the right half-plane or on the imaginary axis and therefore closed-loop stability is ensured if the net sum of anti-clockwise encirclements of the critical point \((-1+j0)\) by the characteristic gain loci is zero, and if the characteristic gain loci do not pass through the critical point and have no infinite branches. The characteristic gain loci are shown in figure 17 from which the following stability conditions are obtained.

(i) For \(-\infty < -\frac{1}{k} < -0.8\) there are no encirclements of, or passage through, the critical point and thus the closed-loop system is stable for \(0 < k < 1.25\).

(ii) For \(-\frac{1}{k} = -0.8\) the characteristic gain loci pass through the critical point and therefore the system is closed-loop unstable for \(k=1.25\).

(iii) For \(-0.8 < -\frac{1}{k} < -0.4\) there is one clockwise encirclement of the critical point and therefore the system is closed-loop unstable for \(1.25 < k < 2.5\).
Figure 17. Characteristic gain loci
(iv) For $\frac{1}{k} = -0.4$ the characteristic gain loci pass through the critical point and therefore the system is closed-loop unstable for $k=2.5$.

(v) For $-0.4 < \frac{1}{k} < 0$ there are no encirclements of, or passage through, the critical point and thus the closed-loop system is stable for $2.5 < k < \infty$.

(vi) For $\frac{1}{k} = 0$ the characteristic gain loci pass through the critical point and therefore the system is closed-loop unstable for $k=\infty$.

(vii) For $0 < \frac{1}{k} < 0.533$ there are two clockwise encirclements of the critical point and therefore the system is closed-loop unstable for $-\infty < k < -1.875$.

(viii) For $\frac{1}{k} = 0.533$ the characteristic gain loci pass through the critical point and therefore the system is closed-loop unstable for $k = -1.875$.

(ix) For $0.533 < \frac{1}{k} < \infty$ there are no encirclements of, or passage through, the critical point and thus the closed-loop system is stable for $-1.875 < k < 0$.

Note that the conditions where $k$ is negative correspond to positive feedback.

References


5. A generalized inverse Nyquist stability criterion

In this chapter a generalization of the inverse Nyquist stability criterion [1] for single-input single-output feedback systems is developed for the general feedback configuration which is complementary to the exposition of the generalized Nyquist stability criterion presented in chapter 4. The development is based on an association of the open-loop gain matrix G(s) with a set of inverse characteristic gain functions, and a corresponding set of inverse Nyquist diagrams which will be termed the inverse characteristic gain loci.

5.1 Inverse characteristic gain functions

A main feature of the generalization of Nyquist's stability criterion, given in chapter 4, is the association of a set of algebraic functions - the characteristic gain functions - with a square transfer function matrix G(s) by means of the characteristic equation

\[ \Delta(g,s) = \det\left[ gI_m - G(s) \right] = 0 \]  \hspace{1cm} (5.1.1)

If G(s) has normal rank m then its inverse function \( G(s)^{-1} \) exists, and has a corresponding characteristic equation given by

\[ \Lambda(^*g,s) = \det[^*gI_m - G(s)^{-1}] = 0 \]  \hspace{1cm} (5.1.2)

If \( \Lambda(^*g,s) \) is regarded as a polynomial in \(^*g\) with coefficients which are rational functions of s then equation (5.1.2) defines a set of algebraic functions \( \{^*g_i(s) : i=1,2,...,\ell\} \) which will be called the inverse characteristic gain functions. If \( \Lambda(^*g,s) \) is also irreducible over the field of rational
functions in $s$, then the inverse characteristic gain function $g^*(s)$, like $g(s)$, has as its domain an appropriate Riemann surface formed out of $m$ copies of the complex $s$-plane suitably joined together. In fact, because the eigenvalues of a matrix are reciprocal to the eigenvalues of the matrix inverse

$$g^*(s) = \frac{1}{g(s)} \quad (5.1.3)$$

and therefore $g^*(s)$ and $g(s)$ have the same branch points, and hence the same Riemann surface domains.

The inverse characteristic gain function is the foundation on which the generalized inverse Nyquist stability criterion is based.

5.2 Pole-zero relationships

In this section a number of pole-zero relationships are derived which will be used in the proof of the generalized inverse Nyquist stability criterion, section 5.4.

Because the eigenvalues of a matrix are the reciprocal of the eigenvalues of the matrix inverse the poles of the set of inverse characteristic gain functions are simply the zeros of the characteristic gain function set, and vice versa. The pole and zero polynomials of $g^*_i(s)$, denoted by $p^*_g(s)$ and $z^*_g(s)$, are therefore expressible as

$$p^*_g (s) = z^*_g (s) \quad (5.1.4)$$

and

$$z^*_g (s) = p^*_g (s) \quad i=1,2,\ldots,\ell \quad (5.1.5)$$

From sub-section 3.3-2 it is clear that for an open-loop gain matrix $G(s)$ there exists a canonical form, the
Smith-McMillan form \([2]\), such that
\[ G(s) = H_G(s)M_G(s)J_G(s) \]  \hspace{1cm} (5.1.6)

where \(H_G(s)\) and \(J_G(s)\) are both \(m \times m\) unimodular matrices and the Smith-McMillan form \(M_G(s)\) is given by
\[ M_G(s) = \text{diag} \left[ \frac{\varepsilon_1(s)}{\psi_1(s)}, \frac{\varepsilon_2(s)}{\psi_2(s)}, \ldots, \frac{\varepsilon_m(s)}{\psi_m(s)} \right] \]  \hspace{1cm} (5.1.7)

Consequently the poles and zeros of \(G(s)\) are defined as
\[ P_G(s) = \prod_{i=1}^{m} \psi_i(s) \]  \hspace{1cm} (5.1.8)
and
\[ Z_G(s) = \prod_{i=1}^{m} \varepsilon_i(s) \]  \hspace{1cm} (5.1.9)

Now if \(G(s)\) is of normal rank \(m\) equation (5.1.6) can be inverted to give
\[ G(s)^{-1} = J_G(s)^{-1}M_G(s)^{-1}H_G(s)^{-1} \]  \hspace{1cm} (5.1.10)

which using the elementary transformation matrix \(E\), of order \(m\), and given by
\[ E = \begin{bmatrix}
0 & 0 & \ldots & 0 & 1 \\
0 & 0 & \ldots & 1 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 1 & \ldots & 0 & 0 \\
1 & 0 & \ldots & 0 & 0
\end{bmatrix} = E^{-1} \]  \hspace{1cm} (5.1.11)

can be rewritten as
\[ G(s)^{-1} = J_G(s)^{-1}EEM_G(s)^{-1}EEH_G(s)^{-1} \]  \hspace{1cm} (5.1.12)
or
\[ G(s)^{-1} = H_G(s)M_G(s)J_G(s) \]  \hspace{1cm} (5.1.13)
where
\[ H^*_G(s) = J_G(s)^{-1}E \] (5.1.14)
\[ J^*_G(s) = EH_G(s)^{-1} \] (5.1.15)

and
\[ M^*_G(s) = EM_G(s)^{-1}E \]
\[ = \text{diag} \left[ \frac{\psi_m(s)}{\varepsilon_m(s)}, \frac{\psi_{m-1}(s)}{\varepsilon_{m-1}(s)}, \ldots, \frac{\psi_1(s)}{\varepsilon_1(s)} \right] \] (5.1.16)
which is the Smith-McMillan form for \( G(s)^{-1} \). The pole-zero definitions for a transfer function matrix [subsection 3.3-2] can now be applied to \( G^{-1}(s) \) with the result that
\[ p^*_G(s) = \prod_{i=1}^{m} \varepsilon_i(s) = z_G(s) \] (5.1.17)
and
\[ z^*_G(s) = \prod_{i=1}^{m} \psi_i(s) = p_G(s) \] (5.1.18)
where \( p^*_G(s) \) and \( z^*_G(s) \) denote the pole and zero polynomials of \( G(s)^{-1} \) respectively.

In sub-section 3.3-3 it is shown that the pole and zero polynomials of \( G(s) \) are related to the poles and zeros of the characteristic gain function set \( \{ g_i(s) : i=1,2,\ldots,l \} \) by
\[ p_G(s) = e_G(s) \prod_{i=1}^{l} b_i'(s) = e_G(s) \prod_{i=1}^{l} p_{g_i}(s) \] (5.1.19)
and
\[ z_G(s) = e_G(s) \prod_{i=1}^{l} b_i'(s) = e_G(s) \prod_{i=1}^{l} z_{g_i}(s) \] (5.1.20)
If the pole and zero polynomials of the complete characteristic gain function set are denoted by \( p_q(s) \) and \( z_q(s) \) respectively, and similarly for the inverse characteristic gain function set.
using \( p_g^*(s) \) and \( z_g^*(s) \), then relationships (5.1.19) and (5.1.20) can be combined with relationships (5.1.17) and (5.1.18) to give

\[
p_G(s) = z_G^*(s) = e_G(s)p_g(s) = e_G(s)z_g^*(s) \tag{5.1.21}
\]

and

\[
z_G(s) = p_G^*(s) = e_G(s)z_g(s) = e_G(s)p_g^*(s) \tag{5.1.22}
\]

These pole-zero relationships will be used in the proof of the inverse Nyquist stability criterion presented later.

5.3 Inverse characteristic gain loci - generalized inverse Nyquist diagrams

The inverse characteristic gain loci are the loci in the complex plane traced out by the reciprocal of each of the eigenvalues of the open-loop gain matrix \( G(s) \) as \( s \) traverses the Nyquist D-contour in the standard (clockwise) direction. The algorithm for computing the loci follows that given for the characteristic gain loci, in sub-section 3.4-1, with the modification that at step (iv) the reciprocal of the numbers \( \tilde{q}_i(j\omega) \) are plotted in the complex plane.

We are now in a position to state and prove the generalized inverse Nyquist stability criterion.

5.4 Generalized inverse Nyquist stability criterion

If the subsystems of the general feedback configuration of figure 3 are each characterized by a state-space model then the following statement of the criterion is applicable.

Statement 1. The general feedback configuration is closed-loop stable if and only if:

(1a) the net sum of anti-clockwise encirclements, of the critical point \((-k+j0)\), by the set of inverse characteristic
gain loci, minus the net sum of anti-clockwise encirclements, of the origin, by the set of inverse characteristic gain loci, is equal to the number of right half-plane poles of $G(s)$;

(2) the inverse characteristic gain loci do not pass through the critical point $(-k+jo)$; and

(3a) the number of branches of the inverse characteristic gain loci that pass through the origin is equal to the number of poles of $G(s)$ on the imaginary axis; and

(4) the eigenvalues of the $A$-matrix, of the open-loop system $S(A,B,C,D)$, which correspond to modes of the system which are unobservable and/or uncontrollable from the point of view of considering the input as that of the first subsystem and the output as that of the $h$th subsystem, are all in the left half-plane.

Alternatively conditions (1a) and (3a) may be replaced by:

(1b) the net sum of anti-clockwise encirclements of the critical point $(-k+jo)$ by the set of inverse characteristic gain loci is equal to the number of right half-plane zeros of $G(s)$;

(3b) the number of branches of the inverse characteristic gain loci passing through infinity is equal to the number of zeros of $G(s)$ on the imaginary axis.

If the subsystems are completely characterized by their transfer function matrices, or if it is known that for each subsystem there are no unobservable and/or uncontrollable modes in the right half-plane including the imaginary axis, then the following statement of the criterion applies.

**Statement 2** The general feedback configuration is closed-loop stable
if and only if:

(1a) the net sum of anti-clockwise encirclements of the critical point \((-k + j\omega)\), by the set of inverse characteristic gain loci, minus the net sum of anti-clockwise encirclements, of the origin, by the set of inverse characteristic gain loci, is equal to the total number of right half-plane poles of \(G_1(s), G_2(s), \ldots,\) and \(G_h(s);\)

(2) the inverse characteristic gain loci do not pass through the critical point \((-k + j\omega)\); and

(3a) the number of branches of the inverse characteristic gain loci that pass through the origin is equal to the number of poles of \(G(s)\) on the imaginary axis.

Alternatively condition (3a) may be replaced by condition (3b), as in statement 1 of the criterion, and if the subsystems are square, i.e. have the same number of outputs as inputs, then condition (1a) may be replaced by:

(1b) the net sum of anti-clockwise encirclements of the critical point \((-k + j\omega)\) by the set of inverse characteristic gain loci is equal to the total number of right half-plane zeros of \(G_1(s), G_2(s), \ldots,\) and \(G_h(s).\)

Note that if condition (2) and/or condition (3a)/(3b) do not hold then the closed-loop system has one or more poles on the imaginary axis and is therefore not input-output stable although the equilibrium state at the origin may be stable.

For strictly proper systems, that is, ones in which the system D-matrix is zero, the inverse characteristic gain loci
approach infinity as $s$ approaches infinity, and $s=\infty$ is a pole of the inverse function. Therefore, in practice, it is necessary to traverse the whole of the $D$-contour, not just the imaginary axis, in order to obtain closed curves. In this way the net sum of encirclements of the critical point and the origin by the inverse characteristic gain loci can be obtained. As an example, the inverse characteristic gain loci for

$$G(s) = \frac{1}{(s+1)^3} \quad (5.4.1)$$

are shown in figure 18.

The traversal of the $D$-contour off the imaginary axis, however, is not necessary if we use conditions (1a), rather than (1b), in both statements of the stability criterion. This is because the extra loci, corresponding to the circular part of the $D$-contour, encircle the origin and the critical point the same number of times, and thereby cancel each other. A useful rule, when using the (1a) conditions, is therefore to join up the corresponding loose ends of the loci via a large semi circle, and to forget any extra encirclements that may actually exist. Whether the ends are joined via the right half-plane or left half-plane is fixed by the fact that the mapping under $g(s)$, from the set of Nyquist curves formed on the Riemann surface of $g(s)$ (by piecing together an appropriate set of Nyquist $D$-contours) to the $g^*$-plane, is conformal. This means that if we imagine two people, $X$ and $Y$, $X$ walking along the domain curve, and $Y$ walking along the corresponding image curve, such that at each step $X$ defines the corresponding image point, then if $X$ turns to his right, $Y$ will turn to $his$ right.
Figure 18. Inverse characteristic gain loci for $G(s) = 1 / (s+1)^3$  
(a) small area around origin  
(b) large area
5.5 **Proof of generalized inverse Nyquist stability criterion**

For the general feedback configuration of figure 3 the closed-loop transfer function matrix \( R(s) \) is given by the relationship

\[
R(s) = kG(s) \left[ I_m + kG(s) \right]^{-1}
\]

(5.5.1)

which after inverting becomes

\[
R(s)^{-1} = \frac{1}{k} G(s)^{-1} + I_m
\]

(5.5.2)

If we now denote the set of algebraic functions defining the eigenvalues of \( R(s)^{-1} \) by \( \{ \xi_i(s) : i = 1,2,\ldots,\ell \} \) we can apply the eigenvalue shift theorem\([3]\) to equation (5.5.2) with the result that

\[
\xi_i(s) = \frac{1}{k} g_i(s) + 1
\]

(5.5.3)

\( i = 1,2,\ldots,\ell \)

Then if the pole and zero polynomials for the complete set of algebraic functions \( \{ r_i^*(s) : i = 1,2,\ldots,\ell \} \) are denoted by the monic polynomials \( p_r^*(s) \) and \( z_r^*(s) \) respectively, we have that

\[
p_r^*(s) = p_g^*(s)
\]

(5.5.4)

Now post-multiplying equation (5.5.2) by \( kG(s) \) we have

\[
kR(s)^{-1}G(s) = kG(s) = \frac{I_m + kG(s)}{k} = F(s)
\]

(5.5.5)

the return difference matrix \([4]\), and taking determinants of this we obtain

\[
\det F(s) = k^m \frac{\det[R(s)^{-1}]}{\det[G(s)^{-1}]}
\]

\[
= \gamma \frac{z_r^*(s)}{p_r^*(s)} \frac{p_g^*(s)}{z_g^*(s)}
\]

(5.5.6)
where \( \gamma \) is a scalar constant independent of \( s \). But

\[
det F(s) = \beta \frac{z_f(s)}{P_f(s)} = \beta \frac{z_f(s)}{p_f(s)}
\]  

(5.5.7)

where \( p_f(s) \) and \( z_f(s) \) are respectively the monic pole and zero polynomials for the set of algebraic functions \( \{f_i(s) : i=1,2,\ldots,\ell\} \)

and therefore

\[
\beta = \gamma
\]  

(5.5.8)

and

\[
z_f(s) = \frac{z_r(s)}{p_r(s)} \cdot \frac{p^*_r(s)}{z^*_r(s)} \cdot p_f(s)
\]  

(5.5.9)

Now from equations (4.2.18) and (5.1.21)

\[
p_f(s) = p_g(s) = \frac{z^*_r(s)}{g}
\]  

(5.5.10)

and hence combining this with equations (5.5.4) and (5.5.9)

we obtain

\[
z_f(s) = \frac{z^*_r(s)}{p_r(s)}
\]  

(5.5.11)

In chapter 4, equation (4.2.19), it was shown that the closed-loop characteristic polynomial can be given by

\[
CLCP(s) = p_d(s)e_G(s) \prod_{i=1}^{\ell} d_i(s)
\]  

(5.5.12)

or equivalently, using the notation presented in this chapter,

\[
CLCP(s) = p_d(s)e_G(s) z_f(s)
\]  

(5.5.13)

and therefore combining this with equation (5.5.11) we have

\[
CLCP(s) = p_d(s)e_G(s) \frac{z^*_r(s)}{p_r(s)}
\]  

(5.5.14)

This implies that the following conditions are necessary and sufficient for closed-loop stability:

(a) \( e_G(s) = 0 \) has only left half-plane roots;

(b) \( z^*_r(s) = 0 \) has only left half-plane roots; and

(c) \( p_d(s) = 0 \) has only left half-plane roots.
Suppose now that a Nyquist D-contour, as shown in figure 16, is drawn on \(m\) copies of the complex plane before they are pieced together to form the Riemann surfaces \(\{\mathbb{R}_i^*: i=1,2,\ldots,\ell\}\) on which the \(\{\mathbb{R}_i^*(s): i=1,2,\ldots,\ell\}\) are defined. Let us consider the \(j\)th surface \(\mathbb{R}_j^*\) corresponding to \(\mathbb{R}_j^*(s)\). When the surface is formed the set of Nyquist D-contours combine to form a set of closed Jordan contours \([5]\) enclosing right half-plane regions on \(\mathbb{R}_j^*\). The extended Principle of the Argument \([appendix 5]\) can be applied to each right half-plane region on \(\mathbb{R}_j^*\). Therefore for a particular right half-plane region, not necessarily simply connected but with a boundary made up from Nyquist D-contours, we have that the difference between the number of zeros and poles of the algebraic function \(\mathbb{R}_j^*(s)\) in the region, is equal to the number of clockwise encirclements of the origin in \(\mathbb{C}\) (the complex \(r\)-plane) by the image of the boundary curves, under \(\mathbb{R}_j^*(s)\), for that particular region. If we therefore consider all the right half-plane regions and apply the extended Argument Principle to each, we have that

\[
N(\mathbb{R}_j^*, 0) = Z_{\mathbb{R}_j^*}^* - P_{\mathbb{R}_j^*}^* \tag{5.5.15}
\]

where:

(i) \(N(\mathbb{R}_j^*, 0)\) is the net sum of clockwise encirclements of the origin in \(\mathbb{C}\) by the set of image curves, under \(\mathbb{R}_j^*(s)\) of the set of curves formed on \(\mathbb{R}_j^*\) by piecing together the appropriate set of Nyquist D-contours when forming \(\mathbb{R}_j^*\);

(ii) \(Z_{\mathbb{R}_j^*}^*\) is the number of right half-plane zeros of \(\mathbb{R}_j^*(s)\);
and

(iii) $P^*_{r_j}$ is the number of right half-plane poles of $r_j(s)$.

If we now consider the corresponding inverse characteristic gain function $g_j^*(s)$ in the same way, we find that

$$N(g_j^*, 0) = Z^*_{g_j} - P^*_{g_j} \quad (5.5.16)$$

where:

(i) $N(g_j^*, 0)$ is the net sum of clockwise encirclements of the origin in $\mathbb{C}$ by the inverse characteristic gain loci corresponding to $g_j^*(s)$;

(ii) $Z^*_{g_j}$ is the number of right half-plane zeros of $g_j^*(s)$; and

(iii) $P^*_{g_j}$ is the number of right half-plane poles of $g_j^*(s)$.

Equations (5.5.15) and (5.5.16) can be combined to give

$$N(r_j^*, 0) - N(g_j^*, 0) = Z^*_{r_j} - P^*_{r_j} - Z^*_{g_j} + P^*_{g_j} \quad (5.5.17)$$

which using equation (5.5.4) becomes

$$N(r_j^*, 0) - N(g_j^*, 0) = Z^*_{r_j} - Z^*_{g_j} \quad (5.5.18)$$

and if we consider the complete set of algebraic functions $\{r_i(s) : i=1,2,\ldots,\ell\}$ and the set $\{g_i(s) : i=1,2,\ldots,\ell\}$ we have

$$\sum_{i=1}^\ell N(r_i^*, 0) - \sum_{i=1}^\ell N(g_i^*, 0) = \sum_{i=1}^\ell Z^*_{r_i} - \sum_{i=1}^\ell Z^*_{g_i} \quad (5.5.19)$$

Now condition (b) for closed-loop stability is equivalent to saying that there are no zeros of $\{r_i(s) : i=1,2,\ldots,\ell\}$ in the right half-plane or on the imaginary axis, and can therefore be replaced by

$$\sum_{i=1}^\ell N(r_i^*, 0) - \sum_{i=1}^\ell N(g_i^*, 0) = - \sum_{i=1}^\ell Z^*_{g_i} \quad (5.5.20)$$
plus the condition that \( \{ r_i^*(s) : i = 1, 2, \ldots, l \} \) have no zeros on the imaginary axis. The necessary and sufficient conditions for closed-loop stability can therefore be rewritten as

(a') \( e_G(s) = 0 \) has no right half-plane roots;

(b') \( e_G(s) = 0 \) has no roots on the imaginary axis;

(c') \( \sum_{i=1}^{l} N(r_i^*, 0) - \sum_{i=1}^{l} N(g_i^*, 0) = -\sum_{i=1}^{l} z_i^* \);

(d') \( \{ r_i^*(s) : i = 1, 2, \ldots, l \} \) have no zeros on the imaginary axis; and

(e') \( p_d(s) \) has only left half-plane zeros.

The pole-zero relationship (5.1.21) now leads us to consider combining conditions (a') and (c') into the single equivalent condition

\[
\sum_{i=1}^{l} N(r_i^*, 0) - \sum_{i=1}^{l} N(g_i^*, 0) = -PG
\]

(5.5.21)

where \( PG \) is the number of right half-plane poles of \( G(s) \).

The necessity and sufficiency of condition (5.5.21) for closed-loop stability when conditions (b'), (d') and (e') are satisfied is proved as follows

From relationship (5.1.21) we have that

\[
P_G = e + \sum_{i=1}^{l} z_i^* \quad (5.5.22)
\]

where \( e \) is the number of right half-plane poles of \( e_G(s) \), and combining this with equation (5.5.19) gives

\[
\sum_{i=1}^{l} N(r_i^*, 0) - \sum_{i=1}^{l} N(g_i^*, 0) = \sum_{i=1}^{l} z_i^* + e - PG \quad (5.5.23)
\]

To establish the necessity of condition (5.5.21) suppose that

\[
\sum_{i=1}^{l} N(r_i^*, 0) - \sum_{i=1}^{l} N(g_i^*, 0) \neq -PG
\]
then from equation (5.5.23) this implies that
\[ \sum_{i=1}^{\ell} z_i^{*} \neq 0 \quad \text{and/or} \quad e \neq 0 \]
and hence we conclude that condition (5.5.21) is necessary for closed-loop stability.

For sufficiency suppose that
\[ \sum_{i=1}^{\ell} N(r_i',0) - \sum_{i=1}^{\ell} N(g_i',0) = -P_G \]
then from equation (5.5.23) this implies that
\[ \sum_{i=1}^{\ell} z_i^{*} = e = 0 \]
and hence the system is closed-loop stable. Thus the sufficiency of condition (5.5.21) is established.

We have therefore shown that the following conditions are necessary and sufficient for closed-loop stability:

(a") \[ \sum_{i=1}^{\ell} N(r_i',0) - \sum_{i=1}^{\ell} N(g_i',0) = -P_G ; \]
(b") \{r_i'(s): i=1,2,...,\ell\} have no zeros on the imaginary axis;
(c") \[ e_G(s)=0 \] has no roots on the imaginary axis; and
(d") \[ \phi_d(s) \] has only left half-plane zeros.

From equation (5.5.3) it is clear that
\[ \sum_{i=1}^{\ell} N(r_i',0) = \sum_{i=1}^{\ell} N(g_i',-k) \quad (5.5.24) \]
and consequently condition (a") is equivalent to condition (1a) in statement 1 of the stability criterion.

Also from equation (5.5.3) we have that \{r_i'(s): i=1,2,...,\ell\} have zeros on the imaginary axis if, and only if the inverse characteristic gain loci pass through the critical point \((-k+j0); \) hence condition (b") is equivalent to condition 2 in statement 1 of the stability criterion.

From the pole-zero relationships, (5.1.21) and (5.1.22),
condition (c") is clearly equivalent to condition (3a) and also condition (3b) in either statement of the stability criterion.

Therefore to complete the proof of statement 1 of the stability criterion all that is left is to show that condition (1b) is equivalent to condition (a"). To do this we will show that when conditions (b"), (c") and (d") are satisfied, the condition

\[ \sum_{i=1}^{l} N(r_i, 0) = \sum_{i=1}^{l} N(g_i, -k) = -Z_G, \]  

(5.5.25)

where \( Z_G \) is the number of right half-plane zeros of \( G(s) \), is necessary and sufficient for closed-loop stability.

From relationship (5.1.22) we have that

\[ Z_G = e + \sum_{i=1}^{l} P_{g_i}, \]  

(5.5.26)

where \( P_{g_i} \) is the number of right half-plane poles of \( g_i(s) \), but from equation (5.5.3) the poles of \( r_i(s) \) and \( g_i(s) \) are the same, and hence

\[ Z_G = e + \sum_{i=1}^{l} P_{r_i}, \]  

(5.5.27)

If we now combine equation (5.5.27) with equation (5.5.15), which is valid for \( \{j=1,2,\ldots,l\} \), we have

\[ \sum_{i=1}^{l} N(r_i, 0) = \sum_{i=1}^{l} Z_{r_i}^{*} + e - Z_G \]  

(5.5.28)

To establish the necessity of condition (5.5.25) suppose that

\[ \sum_{i=1}^{l} N(r_i, 0) \neq -Z_G \]

then from equation (5.5.28) this implies that

\[ \sum_{i=1}^{l} Z_{r_i}^{*} \neq 0 \text{ and/or } e \neq 0 \]
and hence we conclude that condition (5.5.28) is necessary for closed-loop stability.

For sufficiency suppose that

\[ \sum_{i=1}^{\ell} \mathbf{N}(\mathbf{r}_i,0) = -Z_G \]

then from equation (5.5.28) we have that

\[ \sum_{i=1}^{\ell} \mathbf{r}_i = \mathbf{e} = \mathbf{0} \]

and hence that the system is closed-loop stable, providing conditions (b''), (c'') and (d'') are satisfied. Thus the sufficiency of condition (5.5.25) is established.

This completes the proof of statement 1 of the generalized inverse Nyquist stability criterion.

Statement 2 of the stability criterion applies to systems [see chapter 4] in which either

(i) \( P_d(s) = P_x(s) \)

or

(ii) \( P_d(s) = P_x(s)P_{d_1}(s)P_{d_2}(s) \ldots P_{d_h}(s) \)

where the zeros of \( P_{d_i}(s) \) are all in the left half-plane.

In these situations it is possible to combine conditions (1a) and (4), of statement 1, into the single equivalent condition

\[ \sum_{i=1}^{\ell} \mathbf{N}(\mathbf{g}_i, -k) - \sum_{i=1}^{\ell} \mathbf{N}(\mathbf{g}_i,0) = - \sum_{i=1}^{h} \mathbf{P}_G \]

and (5.5.31)
The necessity and sufficiency of condition (5.5.31) for closed-loop stability when conditions (2) and (3) are satisfied is proved as follows.

From equations (4.2.37) and (4.2.39) we have that
\[ h \sum_{i=1}^{\ell} P_{G_i} = e + \sum_{i=1}^{\ell} P + P_d \hspace{1cm} (5.5.32) \]

But the poles of \( \{f_i(s) : i=1,2,\ldots,\ell\} \) are the same as the poles of \( \{g_i(s) : i=1,2,\ldots,\ell\} \) which in turn are the same as the zeros of \( \{g_i^*(s) : i=1,2,\ldots,\ell\} \), and therefore equation (5.5.32) can be rewritten as
\[ h \sum_{i=1}^{\ell} G_i = e + \sum_{i=1}^{\ell} Z^* + P_d \hspace{1cm} (5.5.33) \]

Equation (5.5.33) can now be combined with equation (5.5.19) to give
\[ \sum_{i=1}^{\ell} N(r_i,0) - \sum_{i=1}^{\ell} N(g_i,0) = \sum_{i=1}^{\ell} Z^* + e + P_d - \sum_{i=1}^{\ell} P_{G_i} \hspace{1cm} (5.5.34) \]

which using equation (5.5.24) becomes
\[ \sum_{i=1}^{\ell} N(g_i-k) - \sum_{i=1}^{\ell} N(g_i,0) = \sum_{i=1}^{\ell} Z^* + e + P_d - \sum_{i=1}^{\ell} P_{G_i} \hspace{1cm} (5.5.35) \]

To establish the necessity of condition (5.5.31) suppose that
\[ \sum_{i=1}^{\ell} N(g_i-k) - \sum_{i=1}^{\ell} N(g_i,0) \neq -\sum_{i=1}^{\ell} P_{G_i} \]

then from equation (5.5.35) this implies that
\[ \sum_{i=1}^{\ell} Z^* \neq 0, \text{ or } e \neq 0, \text{ or } P_d \neq 0 \]

or any combination of these, and thus that the system will be closed-loop unstable. Hence we conclude that condition (5.5.31) is necessary for closed-loop stability.
For sufficiency suppose that

\[
\sum_{i=1}^{h} N(g_i,-k) - \sum_{i=1}^{h} N(g_i,0) = - \sum_{i=1}^{h} P_i G_i
\]

then from equation (5.5.35) we must have that

\[
\sum_{i=1}^{l} \sum_{r_i}^{*} = e = P_d = 0
\]

and hence the system is stable providing conditions (2) and (3) are satisfied. Thus the sufficiency of condition (5.5.31) is established.

Therefore to complete the proof of statement 2 of the stability criterion all that is left is to show the equivalence between conditions (1a) and (1b). To do this we will show that when conditions (2) and (3) are satisfied, condition (1b) i.e.

\[
\sum_{i=1}^{l} N(r_i,0) = \sum_{i=1}^{l} N(g_i,-k) = - \sum_{i=1}^{l} P_i G_i
\]

where \( Z_{G_i} \) is the number of right half-plane poles of \( G_i(s) \), is necessary and sufficient for closed-loop stability.

From equation (5.5.15), which is valid for \( \{j=1,2,...,\ell\} \),

\[
\sum_{i=1}^{l} N(r_i,0) = \sum_{i=1}^{l} N(g_i,-k) = \sum_{i=1}^{l} Z_i - \sum_{i=1}^{l} P_i
\]

and since the poles of \( \{r_i(s): i=1,2,...,\ell\} \) are the same as the poles of \( \{g_i(s): i=1,2,...,\ell\} \) equation (5.5.37) can be rewritten as

\[
\sum_{i=1}^{l} N(g_i,-k) = \sum_{i=1}^{l} Z_i - \sum_{i=1}^{l} P_i
\]

If the subsystems \( \{G_i(s): i=1,2,...,h\} \) are each square then \( P_x (=P_d) \) is the number of right half-plane zeros (or poles) which are lost through pole-zero cancellations when \( G(s) \) is formed. Therefore if the number of right half-plane
zeros of $G(s)$ is given [equation (3.3.36)] by

$$Z_G = e + \sum_{i=1}^{h} Z G_i$$  \hspace{1cm} (5.5.39)

then we must have

$$\sum_{i=1}^{h} Z G_i = e + \sum_{i=1}^{h} Z G_i + P_d$$  \hspace{1cm} (5.5.40)

or, since the poles of $g_i(s)$ are identically equal to the zeros, of $g_i(s)$,

$$\sum_{i=1}^{h} Z G_i = e + \sum_{i=1}^{h} P_i^* + P_d$$  \hspace{1cm} (5.5.41)

Consequently equations (5.5.38) and (5.5.41) can be combined to give

$$\sum_{i=1}^{h} Z N(g_i,-k) = \sum_{i=1}^{h} Z R_i^* + e + P_d - \sum_{i=1}^{h} Z G_i$$  \hspace{1cm} (5.5.42)

To establish the necessity of condition (5.5.36) suppose that

$$\sum_{i=1}^{h} Z N(g_i,-k) \neq - \sum_{i=1}^{h} Z G_i$$

then from equation (5.5.42) this implies that

$$\sum_{i=1}^{h} Z R_i^* \neq 0, \text{ or } e \neq 0, \text{ or } P_d \neq 0$$

or any combination of these and thus that the system will be closed-loop unstable. Hence we conclude that condition (5.5.42) is necessary for closed-loop stability.

For sufficiency suppose that

$$\sum_{i=1}^{h} Z N(g_i,-k) = - \sum_{i=1}^{h} Z G_i$$

then from equation (5.5.42) we must have that

$$\sum_{i=1}^{h} Z R_i^* = e = P_d = 0$$

and hence the system is closed-loop stable, providing conditions (2) and (3) are satisfied. Thus the sufficiency
of condition (5.5.36) is established.

This completes the proof of statement 2 of the generalized inverse Nyquist stability criterion.

5.6 Example

To illustrate the stability criteria consider the system example used in chapters 3 and 4 where

\[ G(s) = G_1(s) = \frac{1}{1.25(s+1)(s+2)} \begin{bmatrix} s-1 & s \\ -6 & s-2 \end{bmatrix} \]

The pole and zero polynomials for the open-loop gain matrix \( G(s) \) are

\[ P_G(s) = (s+1)(s+2) \] and \( Z_G(s) = 1 \)

so that

\[ P_G = 0 \quad \text{and} \quad Z_G = 0. \]

The inverse characteristic gain loci are shown in figure 19.

There are no unobservable and/or uncontrollable modes and therefore either statement 1 or statement 2 of the criterion is directly applicable. In fact, because we have effectively only one subsystem, there is no difference between the two statements.

Testing the conditions for stability as outlined by the generalized inverse Nyquist stability criterion we find that the closed-loop system is stable for

\[ -1.875 < k < 0 \]
\[ 0 < k < 1.25 \]
\[ 2.5 < k < \infty \]

These bounds on the gain control variable \( k \) agree with those obtained in section 4.3 using the generalized Nyquist stability criterion.

It is interesting to note that an inverse Nyquist-like
Figure 19. Inverse characteristic gain loci
stability criterion has recently been used by Mees and Rapp [6] to establish stability criteria for multiple-loop nonlinear feedback systems.

References

6. Multivariable root loci

The Evans' root locus approach \([1;2;3]\) is a well established graphical technique used in the analysis and design of linear time-invariant single-input single-output feedback systems for estimating the system closed-loop poles as a function of the gain control variable. Its generalization to the multivariable case has caused considerable interest and steps towards this end have been made with regard to the asymptotic behaviour of the characteristic frequency loci \([4;5]\), and the angles of departure and approach of the characteristic frequency loci \([6]\); both from a state-space point of view. In this chapter a method is developed, using well established results in algebraic function theory \([7;8;9]\), which allows the asymptotic behaviour and the angles of departure and approach of the characteristic frequency loci to be determined from a Laplace transfer function matrix description of the system.

It is also shown how the method can be utilised to find the asymptotic behaviour of the optimal closed-loop poles of a multivariable time-invariant linear regulator as the weight on the input in the performance criterion approaches zero \([10]\).

The method seems particularly useful for systems with small numbers of inputs since the calculations are then simple enough to be carried out by hand.

6.1 Theoretical background

For the general feedback configuration of figure 3 it has been shown in sub-section 3.3-4 that the characteristic
frequency loci (multivariable root loci) are the $180^\circ$ phase contours of the characteristic gain function $g(s)$, where $g(s)$ is defined via the equation
\begin{equation}
\Delta(g,s) \triangleq \det \left[ gI_m - G(s) \right] = 0 \quad (6.1.1)
\end{equation}
or the equation
\begin{equation}
\Phi(g,s) = b_o(s)\Delta(g,s) = 0 \quad (6.1.2)
\end{equation}
where we have multiplied through by the least common denominator of the rational coefficients in $s$.

For simplicity of exposition, and because this is in any case the usual situation for transfer function matrices arising from practical situations, it is assumed that $\Delta(g,s)$ is irreducible over the field of rational functions in $s$. The characteristic gain function $g(s)$ is related to the gain control variable $k$ via the expression
\begin{equation}
g = \frac{1}{k} \quad (6.1.3)
\end{equation}
so that substituting for $g$ in equation (6.1.2) we have
\begin{equation}
\Phi(-k^{-1},s) = k^{-m} \Gamma(k,s) = 0 \quad (6.1.4)
\end{equation}
and solutions of
\begin{equation}
\Gamma(k,s) = 0 \quad (6.1.5)
\end{equation}
for $s$ in terms of positive real $k$, determine the dependence of the closed-loop poles on the gain control variable $k$.

Therefore, apart from possible single-point loci, the graphical description of this dependence constitutes the characteristic frequency loci of the given system. Note that if there are no single-point loci then equation (6.1.5) is directly equivalent to
\begin{equation}
\text{CLCP}(s) \triangleq \det \left[ sI_n - A_c \right] = \det \left[ sI_n - s(-k^{-1}) \right] = 0 \quad (6.1.6)
\end{equation}
and the characteristic frequency loci have \( n \) branches.

Each of the \( n \) branches of the characteristic frequency loci can in theory be represented in the form
\[
s_i(k) = u_i(k) + jv_i(k) \quad i=1,2,\ldots,n \tag{6.1.7}
\]
where \( j=\sqrt{-1} \), and the subscripts \( i \) are labels for the various branches. The tangent to the \( r \)th branch of the loci at a point \( s_0=s_r(k_0) \) is defined as the limiting position of the straight line through \( s_0 \) and another point \( s_1=s_r(k_0+\delta k) \) as \( s_1 \) approaches \( s_0 \) along the branch, that is as \( \delta k \to 0 \). Now the complex number \( s_1-s_0 \) can be represented by the vector from \( s_0 \) to \( s_1 \) (figure 20), and the vector corresponding to \( (s_1-s_0)/\delta k \), where \( \delta k > 0 \), has the same direction as that vector.

![Figure 20. To derive a formula for the tangent to the characteristic frequency loci](image)

It follows therefore that the vector corresponding to
\[
s_r(k_0) = \lim_{\delta k \to 0} \frac{s_1 - s_0}{\delta k}
\]
\[
\lim_{\delta k \to 0} \frac{s_r(k_0 + \delta k) - s_r(k_0)}{\delta k} \quad (6.1.8)
\]
is tangent to the branch at \(s_0\) and the angle \(\theta\) between this vector and the positive real axis is given by
\[
\theta = \arg\{s_r'(k_0)\} \quad (6.1.9)
\]

Formula (6.1.9) is fundamental to the determination of the asymptotic behaviour of the characteristic frequency loci, and the angles of departure and approach of the loci. For \(k = 0\), equation (6.1.3) implies that the characteristic frequency loci start at poles of the open-loop system, and therefore the angle of departure of a branch of the loci from a pole is given by formula (6.1.9) with \(k_0 = 0\). For \(k = \infty\), equation (6.1.3) implies that the characteristic frequency loci terminate at system zeros. Therefore for a branch of the loci terminating at a finite zero, formula (6.1.9), with \(k_0 = \infty\), gives the angle of approach of the branch to the zero. If a branch terminates at an infinite zero then formula (6.1.9), with \(k_0 = \infty\), gives the angle that the asymptote to the branch makes with the positive real axis.

Formula (6.1.9) does not at first seem to be useful, because expressions for the separate branches of the characteristic frequency loci cannot in general be found explicitly. However, in algebraic function theory \([7;8;9]\), there exists a method for the practical construction of series representations for the branches of an algebraic function in the neighbourhood of a given point. The method consists of repeatedly using a "Newton diagram" to find the next most significant terms in the series. Therefore, by using the Newton diagram just once, approximations can be obtained for the branches of the
characteristic frequency loci in the vicinity of a pole or zero (finite or infinite), of the form

\[ s_r(k) \approx a + bk^a \]  \hspace{1cm} (6.1.10)

where \( a \) and \( b \) are complex numbers, \( a \) is a rational real number, and when \( a \) is fractional the principal root of \( k \) is understood. For the negative feedback configuration under consideration \( k \) is real and positive, \( ak^{a-1} \) will always be real, and therefore applying formula (6.1.9) to the approximation (6.1.10) we have

\[ \theta = \text{argument } b \pm \tau 180^\circ \]  \hspace{1cm} (6.1.11)

where

\[ \tau = \begin{cases} 0 \text{ when } a > 0 \\ 1 \text{ when } a < 0 \end{cases} \]

as a result of the differentiation with respect to \( k \).

In the following sections it is shown how to obtain approximations of the form shown in equation (6.1.10) in the vicinity of a pole or zero, and hence to determine the asymptotic behaviour of the characteristic frequency loci and also the angles of departure and approach of the loci.

6.2 Asymptotic behaviour

The Newton diagram is a graphical construction which can be used to find each of the most significant terms in the series representations for the particular branches of an algebraic function \( q(v) \), in the vicinity of the origin, which approach zero as \( v \) approaches zero. Therefore, whether finding the asymptotic behaviour or the angles of departure and approach of the loci, our first aim is always to reduce the problem, by a change of variables in the characteristic
equation, to one of finding approximations for branches of an algebraic function which approach zero as the independent variable approaches zero.

To determine the asymptotic behaviour of the characteristic frequency loci we need to obtain an approximation to the branch (or branches) of the loci about the point \( s = \infty \), as \( k \) approaches infinity. For this purpose we will put

\[
s = z^{-1}
\]

in the characteristic equation (6.1.2) to obtain

\[
\phi(g,s) = \phi(g,z^{-1}) = z^{-q} \psi(g,z)
\]

where \( q \leq n \) is the number of poles of \( g(s) \), so that in any neighbourhood of the value \( z = 0 \) (the point \( z = 0 \) itself is excluded from the region) the equation \( \phi(g,s) = 0 \) is equivalent to the equation

\[
\psi(g,z) = \sum_{x,y} g_x y^y = 0
\]

For a strictly proper system, that is one where \( D \) is zero, or if \( D \) is singular

\[
\psi_\infty = 0
\]

otherwise \( \psi_\infty \) is non-zero which, as we shall see later from the Newton diagram, corresponds to there being no asymptotic behaviour. Note that if \( D \) is non-singular \( g(s) \) has the same number of finite zeros as poles and therefore we would not expect any closed-loop poles to approach infinity.

The next step is to construct the Newton diagram for \( \psi(g,z) \), from which approximations of the form

\[
z \approx_c g^h
\]

(6.2.5)
can be obtained, where \( c \) is a complex number, and \( h \) a rational real number. The procedure for constructing the appropriate Newton diagram is as follows [8].

In a \((g,z)\)-plane we plot the points \((x,y)\) for which
\( \psi_{xy} \neq 0 \) in equation (6.2.3). As an example, the points for
\[
(1-z+2z^2-25z^3+29z^4)g^2 + (z-22z^2+199z^3-210z^4)g
- (33z^3-594z^4) = 0 \quad (6.2.6)
\]
are shown in figure 21. A straight line is then made to coincide with the horizontal axis and rotated clockwise about the smallest \( g \)-axis point \( P_0 \) (the point \( (2,0) \) on figure 21) until this line passes through another point \( P_1 \) of our net. A straight line is then drawn between the points \( P_0 \) and \( P_1 \). A horizontal line through \( P_1 \), and pointing away from \( P_1 \) towards the vertical \( z \)-axis, is then rotated clockwise about the point \( P_1 \) until it passes through another point \( P_2 \) of our net. A straight line is then drawn between the points \( P_1 \) and \( P_2 \). The procedure is repeated until the vertical \( z \)-axis is reached. The complete Newton diagram for equation (6.2.6) is shown in figure 22.

The tangent of the acute angle which the straight line \( P_0P_1 \) makes with the vertical \( z \)-axis determines the first possible value of \( \mu \), and the tangent of the acute angle which the straight line \( P_1P_2 \) makes with the vertical \( z \)-axis determines the second possible value of \( \mu \), etc. For the Newton diagram of figure 22 we have \( \mu_1 = 1 \) and \( \mu_2 = \frac{1}{2} \).

For a particular exponent \( \mu_t \) there may be several approximations of the form
\[
z_{it}^c c_{it}^{\mu_t} \quad (6.2.7)
\]
Note that if some root of \( g \) is implied by \( \mu_t \) i.e. \( \mu_t \) is a fraction, then it is understood that the principal root is being considered. To determine the coefficients \( c_{it} \) it
Figure 21. Points of the Newton diagram for equation (6.2.6)

Figure 22. Complete Newton diagram for equation (6.2.6)
is necessary to substitute \( z = c \mu_t \) in the terms of equation (6.2.3) corresponding to the points of our net lying on the link \( P_{t-1} P_t \), to equate to zero the result of the substitution and to solve the resulting equation. Let the sum of the relevant terms of equation (6.2.3) have the form

\[
\psi x_0 y_0^\sigma z^\sigma + \ldots + \psi x_1 y_1^\sigma z^\sigma = 0 \quad (6.2.8)
\]

where \( y_0 > \ldots > y_1 \), and \( \sigma \) is a positive integer less than the number of infinite zeros of the system. Then because the terms correspond to the same link we have

\[
\frac{x_0 - x_1}{y_0 - y_1} = \frac{x_{\sigma - 1} - x_1}{y_{\sigma - 1} - y_1} = \ldots = \frac{x_2 - x_1}{y_2 - y_1} = \mu_t
\]

Therefore all the terms of the expression (6.2.8) become similar as a result of the substitution \( z = c \mu_t \), and hence for the determination of the coefficient \( c_{it} \) we obtain an equation of the form

\[
\psi x_0 y_0^\sigma + \ldots + \psi x_1 y_1^\sigma = 0 \quad (6.2.9)
\]

or dividing by \( c_1 \) (\( c \neq 0 \)) we obtain

\[
\psi x_0 y_0^{\sigma - 1} + \ldots + \psi x_1 y_1^{\sigma - 1} = 0 \quad (6.2.10)
\]

which has \( y_0^{\sigma - 1} \) solutions. Note that \( y_0^{\sigma - 1} \) is the difference between the ordinates of the points \( P_t \) and \( P_{t-1} \). Consequently it is equal to the projection of the link \( P_{t-1} P_t \) on to the ordinate axis. This is a useful fact since it enables us to see at a glance how many coefficients there are corresponding to the exponent \( \mu_t \). Also the ordinate associated with the point where the final link reaches the vertical axis
must give the total number of coefficients for all the approximations considering all exponents. This vertical axis point on the final link is therefore the number of infinite zeros of the system.

Having obtained the approximations in the form of equation (6.2.7) it is a simple matter of substitution from equations (6.1.3) and (6.2.1) to obtain the required form

\[ s_{it} = b_{it} k^t \]  

(6.2.11)

representing approximations to the characteristic frequency loci about \( s = \omega \), as \( k \to \infty \). If we now apply formula (6.1.9) we obtain the angles which the asymptotes make with the positive real axis.

The asymptotes, as will be shown in sub-section 6.2-1, group into Butterworth configurations. Each pattern has a common intercept which has been termed the "multivariable pivot" by Kouvaritakis and Shaked [4] who also gave a method for calculating the pivot given the state space description of the system. In appendix 6 it is shown how the multivariable pivots can be derived from the characteristic equation \( \Delta(g,s) = 0 \).

6.2-1 Butterworth Patterns

In this sub-section it is proved that the closed-loop poles that go off to infinity do so along asymptotes which group into several Butterworth configurations. The proof is based on well established results in algebraic function theory [8].

Let us consider the closed-loop characteristic polynomial defined by equation (6.1.5) i.e.
\[\Gamma(k,s) = 0 \quad (6.2.12)\]

in which we allow \( k \) to be complex. Also suppose that at a point \( k_c \) the algebraic function \( s(k) \) defined by equation (6.2.12) has a \( p \)-fold root \( s_c \), and let the \( n \) branches of \( s(k) \) in a neighbourhood \( N \) of \( k_c \) be \( \{s_i(k): i=1,2,...,p,...,n\} \), where \( \{s_i(k_c)=s_c: i=1,2,...,p\} \).

If we analytically continue any one of the branches \( \{s_i(k): i=p+1,...,n\} \) around \( k_c \) in \( N \), then after one encirclement we return to the original function. If, however, we analytically continue \( s_1(k) \) around the critical point \( k_c \) then after one encirclement we obtain one of the functions \( \{s_i(k): i=2,3,...,p\} \), \( s_2(k) \) say, and after another encirclement one of the functions \( \{s_i(k): i=3,4,...,p\} \), \( s_3(k) \) say. After a finite number \( v_1 \), of encirclements we return to the original function \( s_1(k) \).

In this case we shall say that the functions
\[s_1(k),...,s_{v_1}(k) \quad (6.2.13)\]
constitute a cyclical system of branches of the algebraic function \( s(k) \) (in the neighbourhood \( N \) of \( k_c \)).

If \( v_1<p \) we take the function \( s_{v_1+1}(k) \) and analytically continue it around \( k_c \) in \( N \). Repeating the previous arguments we obtain a second cyclical system of branches, namely
\[s_{v_1+1}(k),...,s_{v_2}(k) \quad (6.2.14)\]
Finally all the functions \( s_1(k),...,s_p(k) \) will be sorted into cyclical systems.

The functions of each cyclical system pass consecutively into one another according to the cyclical law when \( k \) goes
round the multiple point \( k_c \). Hence every cyclical system of branches defines in the neighbourhood \( N \) of \( k_c \) an analytic function. The number of values given by this analytic function at a point in the neighbourhood \( N \) is equal to the number of branches comprising the cyclical system being considered. When a single branch comprises a cyclical system, the corresponding analytic function will be single-valued and identical with this branch.

Suppose the neighbourhood \( N \) is replaced by a pile of \( n \) such regions, one for each of the functions \( \{ s_i(k) : i=1,2,\ldots,n \} \) and let a point on the \( i \)th sheet represent the pair \( [k,s_i(k)] \). Also suppose that each neighbourhood \( N \) is circular and has a radial line cut from the critical point \( k_c \) to its periphery, as shown in figure 23. Then the neighbourhoods can be joined together along the edges of the cuts so that if we analytically continue the \( j \)th branch \( (j<p) \) around \( k_c \) on the \( j \)th sheet, we move after one encirclement across the cut into the \((j+1)\)th branch on the \((j+1)\)th sheet, etc. eventually returning to the \( j \)th branch. The pile of neighbourhoods obviously falls into cycles containing one or more sheets as shown in figure 24, and each cycle is a domain for one of the analytic functions defined in the neighbourhood \( N \) of \( k_c \).

To denote an analytic function determined by a cyclical system we will use a Roman numerical subscript. For example the analytic function corresponding to the cyclical system \( s_1(k),\ldots,s_{\nu_1}(k) \) will be denoted by the symbol \( s_{\nu_1}(k) \).

When we consider the asymptotic behaviour of the
characteristic frequency loci, we are concerned with the branches of the algebraic function $s(k)$ which approach infinity, as $k$ approaches infinity. Suppose we have $p$ infinite branches for $k=\infty$, then by the previous arguments the $p$ branches will be arranged into several cyclical systems. Each cyclical system is defined by an analytic function, the form of which is given by the following theorem \([8, \text{page } 39]\).

**Theorem 4**

The infinite branches of the algebraic function $s(k)$, constituting a cyclical system in the neighbourhood of its critical point $k_c=\infty$, in their totality constitute an analytic
function given by a series of the form

\[ s_I = k^{\frac{\lambda}{v}} (b_o + b_1 k^{-\frac{1}{v}} + b_2 k^{-\frac{2}{v}} + \ldots) \]  

(6.2.15)

where \( v \) is the number of branches in the cycle, and \( \lambda \) is a positive integer.

Consequently the infinite branches corresponding to a particular cyclical system in the neighbourhood of \( k = \infty \), can be approximated by an analytic function of the form

\[ s_I(k)^{\frac{\lambda}{v}} b_o k^{-\frac{1}{v}} \]  

(6.2.16)

The characteristic frequency loci are values of the algebraic function \( s(k) \) for \( k \) going from zero to infinity along the positive real axis. Therefore the infinite branches of the frequency loci will occur in cyclical systems which can be approximated by equations of the form (6.2.16). Taking the \( \nu \) roots of \( k \) in equation (6.2.16) we find that the asymptotes corresponding to a cyclical system arrange themselves into a Butterworth configuration of order \( \nu \); that is, we have \( \nu \) asymptotes equally spaced by angles of \( \frac{360^\circ}{\nu} \), in the complex plane.

In general, as already discussed, there will be several cyclical systems of branches in the neighbourhood of \( k = \infty \), and therefore the asymptotes of the characteristic frequency loci will arrange themselves into several Butterworth configurations.

6.3 Angles of departure and approach

To determine the angles of departure and approach of the characteristic frequency loci we need to obtain approximations to the branches of the loci at the poles and finite zeros
of the system. We will first consider the angles of departure from the open-loop poles of the system.

Suppose for the characteristic equation
\[ \phi(g, s) = 0 \quad (6.3.1) \]
we have a pole (or multiple pole) at \( s = \beta \). We will make \( s' = s - \beta \) the new independent variable for equation (6.3.1) and also make the substitution \( g = d^{-1} \), so that equation (6.3.1) becomes
\[ \phi(g, s) = \phi(d^{-1}, s' + \beta) = d^{-m} \varphi(d, s') = 0 \quad (6.3.2) \]
The situation we are examining now reduces to the case where \( d = s' = 0 \), and in the neighbourhood of \( s = \beta \) (excluding \( \beta \) itself) equation \( \phi(g, s) = 0 \) is equivalent to the equation
\[ \Xi(d, s') = \sum_{x, y} d^x s'^y = 0, \quad \xi_{\infty} = 0 \quad (6.3.3) \]
If we construct the Newton diagram for the equation \( \Xi(d, s') = 0 \) an approximation (or approximations in the case of multiple poles) of the form
\[ s' \sim e^{\omega} \quad (6.3.4) \]
is obtained, where \( e \) is a complex number and \( \omega \) a rational real number. From equation (6.1.3), the change of variable, and the substitution \( g = d^{-1} \), we therefore arrive at an approximation to the branch (or branches) of the characteristic frequency loci departing from the pole \( \beta \), in the form
\[ s \sim \beta + b_d^u d \quad (6.3.5) \]
If we now apply formula (6.1.9) the angle of departure \( \theta_d \) is given as
\[ \theta_d = \text{argument} \{ b_d \} \quad (6.3.6) \]
We will now consider the angles of approach to the finite
zeros of the system. Suppose for the characteristic equation (6.3.1) we have a zero (or multiple zero) at \( s = \gamma \). We will take \( s' = s - \gamma \) as the new independent variable so that the equation becomes

\[
\Phi(g, s) = \Phi(g, s' + \gamma) = \chi(g, s') = 0 \quad (6.3.7)
\]

The situation reduces to the case where \( g = s' = 0 \), and in the neighbourhood of \( s = \gamma \) (excluding \( \gamma \) itself) equation \( \Phi(g, s) = 0 \) is equivalent to the equation

\[
\chi(g, s') = \Sigma_{xy} g_{xy} s' \gamma_x \Omega_{\infty} = 0, \Omega_{\infty} = 0 \quad (6.3.8)
\]

If we construct the Newton diagram for \( \chi(g, s') \) an approximation of the form

\[
s' \approx pg^n \quad (6.3.9)
\]

is obtained where \( p \) is a complex number and \( n \) is a rational real number. From equation (6.1.3) and the change of variable we therefore arrive at an approximation to the branch (or branches) of the characteristic frequency loci approaching the zero \( s = \gamma \), in the form

\[
s \approx \gamma + b_k a_k \quad (6.3.10)
\]

If we now apply formula (6.1.9) the angle of approach \( \theta_a \) is given as

\[
\theta_a = \text{argument} \{b_k\} + 180^\circ \quad (6.3.11)
\]

Note that \( \theta_a \) will always be negative and hence the presence of the \( 180^\circ \) term in formula (6.3.11). Also note that this definition for the angle of approach differs by \( 180^\circ \) from the usual definition which is the direction you would look, when positioned at the zero, to see the locus arrive.

### 6.4 Example 1

Consider an open-loop gain matrix
which has a characteristic equation

\[ \Phi(g,s) = \left( s^4 + 5s^3 - 2s^2 - 44s + 40 \right) g^2 + \left( s^3 + 116s - 432 \right) g - 12 s^2 - 2s + 2 = 0 \]

(a) To determine the asymptotic behaviour

Putting \( s = z^{-1} \) in the characteristic equation we obtain

\[ \Psi(g,z) = \left( 1 + 5z - 2z^2 - 44z^3 + 40z^4 \right) g^2 + \left( z + 116z^3 - 432z^4 \right) g - 12 \left( z^2 - 2z^3 + 2z^4 \right) = 0 \]

The Newton diagram for \( \Psi(g,z) \) is shown in figure 25 from which we obtain \( \mu = 1 \), and hence the approximation

\[ z \sim cg \]

The coefficient \( c \) is calculated (using the method described in section 6.2) to have the values \( \frac{1}{4} \) and \( \frac{1}{3} \). Therefore, resubstituting for \( z \) and \( g = -k^{-1} \), we have the following approximations for the characteristic frequency loci

\[ s \sim 4k \quad \text{and} \quad s \sim -3k \quad \text{as} \quad k \to \infty \]

Two branches of the characteristic frequency loci therefore move off to infinity at angles of 0° and 180° to the positive real axis.

![Figure 25. Newton diagram for \( \Psi(g,z) \)](image)
(b) To find the angles of departure

The system has four open-loop poles

\[ s = 1, s = 2, s = -4 + 2j, \text{ and } s = -4 - 2j \]

**Pole at \( s = 1 \)**

Putting \( s' = s - 1 \) and \( g = d^{-1} \) in the characteristic equation we obtain

\[ \Xi_1(d,s') = (s'^4 + 9s'^3 + 19s'^2 - 29s') + (s'^3 + 3s'^2 + 119s' - 315)d - 12(s'^2 + 1)d^2 = 0 \]

The Newton diagram for \( \Xi_1(d,s') \) is shown in figure 26 from which we obtain \( \omega_1 = 1 \), and hence the approximation

\[ s' \sim e_1d \]

The coefficient \( e_1 \) is calculated (using the method described in section 6.2) to have the value -10.86, resulting in the following approximation to the characteristic frequency loci

\[ s \sim 1 + 10.86 \pi \]

about the pole \( s = 1 \). Therefore the angle of departure from the pole \( s = 1 \) is 0°.

**Pole at \( s = 2 \)**

Putting \( s' = s - 2 \) and \( g = d^{-1} \) in the characteristic equation we obtain

\[ \Xi_2(d,s') = (s'^4 + 13s'^3 + 52s'^2 + 40s') + (s'^3 + 6s'^2 + 128s' - 192)d - 12(s'^2 + 2s' + 2)d^2 = 0 \]

The Newton diagram for \( \Xi_2(d,s') \) is shown in figure 27 from which we obtain \( \omega_2 = 1 \), and hence the approximation

\[ s' \sim e_2d \]

The coefficient \( e_2 \) is calculated to have the value 4.8, resulting in the following approximation to the characteristic frequency loci
about the pole \( s = 2 \). Therefore the angle of departure from the pole \( s = 2 \) is 180°.

**Pole at \( s = -4+2j \)**

Putting \( s' = s+4-2j \) and \( g = d^{-1} \) in the characteristic equation we obtain

\[
E_3(d,s') = [s'^4+(-11+j8)s'^3+(10-j60)s'^2+(88+j104)s']
+ [s'^3+(-12+j6)s'^2+(152-j48)s'+(-912+j320)]d
-12[s'^2+(-10+j4)s'+(22-j20)]d^2 = 0
\]

The Newton diagram for \( E_3(d,s') \) is shown in figure 28 from which we obtain \( \omega_3 = 1 \), and hence the approximation

\[ s' \sim e_3 d \]

The coefficient \( e_3 \) is calculated to have the value

\[
\frac{912 - j320}{88 + j104}
\]

resulting in the following approximation to the characteristic frequency loci

\[ s \sim -4+2j + \frac{(-912+j320)}{(88+j104)} \; k \]

about the pole \( s = -4+2j \). Therefore the angle of departure from the pole \( s = -4+2j \) is

\[
\text{argument} \left\{ \frac{-912+j320}{88+j104} \right\} = 110.9°
\]

**Pole at \( s = -4-2j \)**

By symmetry the angle of departure from the pole \( s = -4-2j \) is -110.9°.

(c) **To find the angles of approach**

The system has two finite zeros

\[ s = 1+j \quad \text{and} \quad s = 1-j \]
Figure 26. Newton diagram for $\Xi_1(d,s')$

Figure 27. Newton diagram for $\Xi_2(d,s')$

Figure 28. Newton diagram for $\Xi_3(d,s')$

Figure 29. Newton diagram for $\chi(g,s')$

Figure 30. Complete characteristic frequency loci
Zero at $s = 1+j$

Putting $s' = s-1-j$ in the characteristic equation we obtain

$$\chi(g, s') = \left[ s'^4 + (9+j 4) s'^3 + (13+j 29) s'^2 + (-58+j 46) s' + (-18-j 38) \right] g^2$$

$$\left[ s'^3 + (3+j 3) s'^2 + (116+j 8) s' + (-318+j 118) \right] g$$

$$-12 \left[ s'^2 + j 2 s' \right] = 0$$

The Newton diagram for $\chi(g, s')$ is shown in figure 29 from which we obtain $\eta = 1$, and hence the approximation $s' \approx pg$

The coefficient $p$ is calculated to have the value

$$\frac{-318+j 118}{j 24}$$

resulting in the following approximation to the characteristic frequency loci

$$s \approx 1 + j \frac{(318-j 118)}{j 24} k^{-1}$$

about the zero $s = 1+j$. Therefore the angle of approach to the zero $s = 1+j$ is

$$\text{argument} \left\{ \frac{318-j 118}{j 24} \right\} \pm 180\degree \approx 69.64\degree$$

Zero at $s = 1-j$

By symmetry the angle of approach to the zero at $s = 1-j$ is $-69.64\degree$.

To check the results the complete characteristic frequency loci are shown, projected onto a single complex frequency plane, in figure 30. Branches of the loci coincide about the pole $s = 1$ making the movement of the closed-loop poles difficult to comprehend. The frequency surface characterized by constant phase and constant magnitude contours of $g(s)$ is therefore shown in figures 31(a), (b)
Figure 31. Frequency surface (a)sheet1 (b)sheet2
Figure 32. Angle of departure from pole $s = -4 + 2j$

Figure 33. Angle of approach to zero $s = 1 + j$
to give a clearer description of the characteristic frequency loci. Small regions in the neighbourhood of the pole $s = -4 + 2j$ and the zero $s = 1 + j$ are shown in figures 32 and 33 to verify the calculated angles of departure and approach.

6.5 Asymptotic behaviour of optimal closed-loop poles

In this section the "Newton diagram" approach is used to determine the asymptotic behaviour of the optimal closed-loop poles of a multivariable time-invariant linear regulator, as the weight on the input in the performance criterion approaches zero. The method is based on an association of the optimal characteristic frequency loci with the branches of an appropriate algebraic function. Although the procedure is equivalent to that given by Kwakernaak [11] the essential simplicity of the approach is emphasized in the setting of algebraic function theory.

Consider the stabilizable and detectable time-invariant linear system

$$\frac{dx(t)}{dt} = A x(t) + Bu(t) \quad (6.5.1)$$
$$y(t) = C x(t) \quad (6.5.2)$$

and the performance criterion

$$V(\tau) = \int_{0}^{\tau} [y^T(t)Qy(t) + pu^T(t)Ru(t)] \, dt \quad (6.5.3)$$

where $Q$ and $R$ are positive definite symmetric matrices, and the superscript $T$ denotes the transpose of a matrix or vector. Then it is well known (e.g. [12]) that the optimal control action is given by

$$u(t) = -\frac{1}{p} R^{-1}B^T P x(t) \quad (6.5.4)$$

where $P$ is the unique positive semi-definite solution of
the steady state matrix Riccati equation

\[-PA -AT_P + \frac{1}{\rho}PBR^{-1}B^TP = C^TC\]  \hspace{1cm} (6.5.5)

Kwakernaak [11, equation (16)] has related the closed-loop characteristic polynomial (denoted by \(\phi_C(s)\)) for the optimal regulator to the open-loop characteristic polynomial (denoted by \(\phi_O(s)\)) as follows

\[
\phi_C(s)\phi_C(-s) = \phi_O(s)\phi_O(-s) \det \left[ I_m + \frac{1}{\rho}R^{-1}G^T(-s)QG(s) \right]
\]  \hspace{1cm} (6.5.6)

where \(I_m\) is a unit matrix of order \(m\), the number of system inputs, and

\[G(s) = C(sI - A)^{-1}B\]  \hspace{1cm} (6.5.7)

is the open-loop transfer function or gain matrix of the system. It is obvious from (6.5.6) that the poles of the optimal closed-loop system, dependent on the input weighing, are values of \(s\) in the left half-plane which satisfy

\[
\det \left[ I_m + \frac{1}{\rho}H(s) \right] = 0
\]  \hspace{1cm} (6.5.8)

where

\[H(s) \triangleq R^{-1}G^T(-s)QG(s)\]  \hspace{1cm} (6.5.9)

Consider now the characteristic equation defining the eigenvalues of \(H(s)\), that is

\[\Delta(\eta, s) \triangleq \det \left[ \eta I_m - H(s) \right] = 0\]  \hspace{1cm} (6.5.10)

or by expanding the determinant

\[\Delta(\eta, s) = \eta^m + a_1(s^2)\eta^{m-1} + \ldots + a_m(s^2) = 0\]  \hspace{1cm} (6.5.11)

where the coefficients \(\{a_i(s^2) ; i=1,2,\ldots,m\}\) are rational functions in \(s^2\). The coefficients are functions in \(s^2\) because
\[ \Delta(\eta, s) = \det \left[ \eta I_m - R^{-1}G^T(-s)QG(s) \right] \quad (6.5.12) \]
\[ = \det \left[ \eta I_m - R^{-\frac{1}{2}}G^T(-s)QG(s)R^{-\frac{1}{2}} \right] \]

and \( R^{-\frac{1}{2}}G^T(-s)QG(s)R^{-\frac{1}{2}} \) is para-Hermitian implying [11, appendix A] that \( \Delta(\eta, s) = \Delta(\eta, -s) \) which can only be the case if the coefficients are rational functions in \( s^2 \).

If \( b_0(s^2) \) is the least common denominator of the coefficients \( \{ a_i(s^2) ; i=1,2,...,m \} \) then from (6.5.11)

\[ b_0(s^2)\Delta(\eta, s) = b_0(s^2)\eta^m + b_1(s^2)\eta^{m-1} + \ldots + b_m(s^2) = 0 \quad (6.5.13) \]

where the coefficients \( \{ b_i(s^2) ; i=1,2,...,m \} \) are polynomials in \( s^2 \). The function of a complex variable \( \eta(s) \), defined by (6.5.13) is an algebraic function. In general, (6.5.11) and (6.5.13) will be reducible into several irreducible equations over the field of rational functions in \( s \), thereby defining a set of algebraic functions. However, for simplicity of exposition it will be assumed that equations (6.5.11) and (6.5.13) are irreducible over the field of rational functions in \( s \).

If we compare equations (6.5.8) and (6.5.10) we see that the optimal closed-loop poles can be defined as the left half-plane solutions of

\[ \eta(s) = -\rho \quad (6.5.14) \]

i.e. the optimal characteristic frequency loci are the 180° phase contours of the algebraic function \( \eta(s) \) in the left half-plane. If we now consider \( \rho \) as a complex variable and substitute for \( \eta(s) \) in (6.5.13) we obtain

\[ b_0(s^2) (-\rho)^m + b_1(s^2) (-\rho)^{m-1} + \ldots + b_m(s^2) = 0 \quad (6.5.15) \]
which is an algebraic equation defining the algebraic functions \( s(\rho) \) and \( \rho(s) \). The left half-plane branches of the algebraic function \( s(\rho) \), for \( \rho \) real and positive, are the optimal characteristic frequency loci. (Note that equation \((6.5.15)\) can be obtained directly from equation \((6.5.8)\)).

To determine the asymptotic behaviour of the optimal characteristic frequency loci we need approximations to the branches of the algebraic function \( s(\rho) \) in the neighbourhood of \( s = \infty \), as \( \rho \) approaches zero. These can be obtained as the first terms in the series expansions for the branches of \( s(\rho) \) about the point \( s = \infty \), as \( \rho \) approaches zero. For this purpose we put \( s = z^{-1} \) in equation \((6.5.15)\) and proceed as in section 6.2.

6.6 Example 2

To demonstrate the procedure the asymptotic behaviour of the optimal closed-loop poles will be calculated for

\[
G(s) = \frac{1}{(s+1)(s+2)(s+3)(s+4)} \begin{bmatrix} s + 2 & 6 \\ s + 3 & 1 \end{bmatrix},
\]

\( Q = I \)

and

\( \rho R = \rho I \)

From this data we have

\[
H(s) \triangleq R^{-1} G^T(-s) Q G(s)
= \frac{1}{(s^2-1)(s^2-2)(s^2-3)(s^2-4)} \begin{bmatrix} -2s^2 + 13 & -7s + 15 \\ 7s + 15 & 37 \end{bmatrix}
\]

so that
\[ \Delta(n,s) \triangleq \det[nI_2 - H(s)] \]

\[ \eta^2 - \frac{(2s^2 + 50)}{(s^2-1)(s^2-2)(s^2-3)(s^2-4)} \eta + \frac{(-25s^2 + 256)}{(s^2-1)(s^2-2)(s^2-3)(s^2-4)^2} = 0 \]

If we now substitute \( \eta = -\rho \), and multiply throughout by the least common denominator of the coefficients, we obtain

\[ (s^2-1)^2(s^2-2)^2(s^2-3)^2(s^2-4)^2 \rho^2 + (s^2-1)(s^2-2)(s^2-3)(s^2-4)(-2s^2+50) \rho + (-25s^2 + 256) = 0. \]

The substitution \( s = z^{-1} \) now gives

\[ (1-z^2)^2(1-2z^2)^2(1-3z^2)^2(1-4z^2)^2 \rho^2 + z^6(1-z^2)^2(1-2z^2)^2(1-3z^2)^2(1-4z^2)^2 (-2+50z^2) \rho + z^{14}(-25 + 256z^2) = 0 \]

The Newton diagram for this last equation is shown in figure 34, from which we obtain

\[ \nu_1 = \frac{1}{6}, \quad \nu_2 = \frac{1}{8}, \] and hence the approximations \( z \sim e_1 \rho \frac{1}{6} \) and \( z = e_2 \rho \frac{1}{8} \). To find the values of \( e_1 \) we substitute \( z = e_1 \rho \frac{1}{6} \) into \( \rho^2 - 2z^6 \rho = 0 \) i.e. the terms corresponding to the points on the first link of the Newton diagram equated to zero, giving \( e_1 = 6\sqrt{0.5} = 0.891 \exp \left( \frac{j2k\pi}{6} \right) \), \( k=0,1,2,3,4,5 \).

If we substitute \( z = e_2 \rho \frac{1}{8} \) into \( -25z^{14} - 2z^6 \rho = 0 \) i.e. the terms corresponding to the points on the second link of the Newton diagram equated to zero, we obtain

\[ e_2 = 8\sqrt{-0.08} = 0.729 \exp \left( \frac{j(\pi+2k\pi)}{8} \right), \quad k=0,1,2,\ldots,7 \]
Figure 34. Newton diagram
Figure 35. Asymptotic behaviour of the optimal characteristic frequency loci (plus right half-plane image) displayed on the Riemann surface domain of $\eta(s)$
(a) sheet 1 of the surface
(b) sheet 2 of the surface
If we now substitute back for $s$, and take only left half-plane solutions, we obtain the following expressions for the asymptotic behaviour of the optimal characteristic frequency loci

$$s \approx 1.12 \left[ \exp \left( j \frac{k\pi}{3} \right) \right]^{\frac{-1}{6}}, k = 2, 3, 4$$

and

$$s \approx 1.37 \left[ \exp \left( j (\pi + 2k\pi) \right) \right]^{\frac{-1}{8}}, k = 2, 3, 4, 5.$$

Note that using Kwakernaak's notation [11], these define Butterworth patterns of orders 3 and 4. The optimal characteristic frequency loci and their right half-plane image are shown in figure 35. Note that they have been computed as the 180° phase contours of the algebraic function $\eta(s)$ and displayed on its Riemann surface domain.

References


7. On parametric stability and future research

A feedback system is said to be stable if all of its closed-loop poles are in the left half-plane. The stability of a control system is therefore dependent on its associated parameters. Sometimes in a control system the value of a parameter is uncertain perhaps due to ageing, deterioration, or damage; in other instances it may be desirable, for economic reasons, to change a parameter value. In both these cases a technique which predicts the relative stability of a system with respect to a given parameter would be extremely useful.

A dominant theme in the preceding chapters has been the association of a system with two sets of algebraic functions: characteristic gain functions and characteristic frequency functions. In this chapter characteristic parameter functions are introduced, and used to develop the ideas of 'parametric' root loci and 'parametric' Nyquist loci from which the relative stability of a system, with respect to a single parameter, can be determined.

In the final section of this chapter a few tentative proposals and suggestions for future research are made.

7.1 Characteristic frequency and characteristic parameter functions

The feedback configuration considered is shown in figure 36, where \( A(k_2, k_3, \ldots, k_q) \), \( B(k_2, k_3, \ldots, k_q) \), \( C(k_2, k_3, \ldots, k_q) \) and \( D(k_2, k_3, \ldots, k_q) \) are state-space matrices which are dependent on \((q-1)\) real, time-invariant parameters and \( k_1 \) is
a scalar, time-invariant gain parameter common to all the loops.

\[
\text{Figure 36. Feedback configuration for parameter analysis}
\]

The closed-loop poles for this configuration are solutions of

\[
\det [ sI_n - S(k) ] = 0 \quad (7.1.1)
\]

where

\[
S(k) = A(k_2, \ldots, k_q) - B(k_2, \ldots, k_q)(k_1 I_m + D(k_2, \ldots, k_q))^{-1} C(k_2, \ldots, k_q)
\]

is the closed-loop frequency matrix [see section 3.1]. If numerical values for all the parameters except one, \( k_j \) say, are substituted into equation (7.1.1), and \( k_j \) considered as a complex variable, then the resulting algebraic equation

(which for simplicity of exposition will be regarded as irreducible) defines a pair of algebraic functions \[ l \], \( s(k_j) \) and \( k_j(s) \). The algebraic function \( s(k_j) \) is called the characteristic frequency function with respect to \( k_j \), and the algebraic function \( k_j(s) \) is called the characteristic parameter function for \( k_j \). (Note that the characteristic frequency function \( s(g) \) and the characteristic gain function \( g(s) \), introduced in chapter 3, are equivalent to \( s(-k_1^{-1}) \) and \( -k_1(s)^{-1} \) respectively).

The branches of \( s(k_j) \), for \( k_j \) real, clearly define the variation of the closed-loop poles with respect to \( k_j \), and as such are termed parametric root loci. Alternatively, the parametric root loci can be viewed as the 0° phase contours of \( k_j(s) \) on the frequency surface domain for \( k_j(s) \).

Dual to the parametric root loci are the parametric Nyquist loci or characteristic parameter loci which are the branches of \( k_j(s) \) as \( s \) traverses the imaginary axes. Alternatively, the characteristic parameter loci can be viewed as the ±90° phase contours of \( s(k_j) \) on the Riemann surface domain for \( s(k_j) \) which will be called the parameter surface for \( k_j \).

If, for a particular system, we have a set of nominal values for the system parameters we can determine which, if any, are sensitive with respect to stability by looking at the set of parameter surfaces. To help in such an assessment the following generalizations of gain and phase margin are introduced.
7.2 Gain and phase margins

The ± 90° phase contours of $s(k_j)$ on the parameter surface for $k_j$ trace out the boundary between stable and unstable closed-loop poles and therefore we can define parameter gain and phase margins for $k_j$ about a stable operating point $\hat{k}_j$ which give a measure of the relative stability of the system with respect to $k_j$.

**Parameter gain margin.** Parameter gain margin is defined with respect to a stable operating point $\hat{k}_j$ as the smallest change in parameter gain about $\hat{k}_j$ needed to drive the system into instability. Let $d_i$ be the shortest distance along the real axis from a stable operating point $\hat{k}_j$ to the stability boundary (characteristic parameter loci) on the ith sheet of the parameter surface for $k_j$. Then the parameter gain margin is defined as $\min\{d_i: i=1,2,\ldots,n\}$.

**Parameter phase margin.** On each of the n sheets of the parameter surface for $k_j$ imagine that an arc is drawn, centre the origin, from a stable operating point $\hat{k}_j$ until it reaches the stability boundary (characteristic parameter loci). Let $\phi_i$ be the angle subtended at the origin by the corresponding arc on the ith sheet. Then the parameter phase margin is defined as $\min\{\phi_i: i=1,2,\ldots,n\}$.

7.3 Example

In this section an inverted pendulum positioning system (see figure 37) is considered and its stability analysed with respect to one of its parameters, namely the mass of the carriage.
This system has also been used by Kwakernaak and Sivan [2], Cannon [3], and Elgerd [4]. The system can be modelled by the following linearized state differential equation [2]

\[
\begin{bmatrix}
0 & 1 & 0 & 0 \\
0 & -\frac{F}{M} & 0 & 0 \\
0 & 0 & 0 & 1 \\
\frac{F}{L'} & 0 & \frac{F}{L'} & 0
\end{bmatrix}
\begin{bmatrix}
x(t) \\
x(t) \\
x(t) \\
x(t)
\end{bmatrix}
+
\begin{bmatrix}
0 \\
1 \\
0 \\
0
\end{bmatrix}
\begin{bmatrix}
u(t) \\
\frac{1}{M} \\
0 \\
0
\end{bmatrix}
\tag{7.3.1}
\]

where \(u(t)\) is a force exerted on the carriage by a small motor; \(M\) is the mass of the carriage; \(F\) is the friction coefficient associated with the movement of the carriage; and \(L'\) is given by

\[
L' = \frac{J + mL^2}{mL} \tag{7.3.2}
\]
where \( m \) is the mass of the pendulum; \( L \) is the distance from the pivot to the centre of gravity of the pendulum; and \( J \) is the moment of inertia of the pendulum with respect to the centre of gravity.

The system is stabilizable using state feedback of the form

\[
u(t) = -Kx(t) \tag{7.3.3}
\]

and using the numerical values

\[
\begin{align*}
\frac{F}{M} &= 1 \text{ s}^{-1} \\
\frac{1}{M} &= 1 \text{ kg}^{-1} \\
\frac{g}{L'} &= 11.65 \text{ s}^{-2} \\
L' &= 0.842 \text{ m}
\end{align*}
\]

it can be found \([2]\) that

\[
K = \begin{bmatrix} 86.81 & 12.21 & -118.4 & -33.44 \end{bmatrix} \tag{7.3.5}
\]

stabilizes the linearized system placing the closed-loop poles at \(-4.706 + j 1.382\) and \(-1.902 + j 3.420\).

We will now look at the parameter surface for \( M \) to see how variations in the carriage mass, about an operating point of 1 kg, affect the stability of the system. The four sheets of the mass surface, characterized by constant phase and magnitude contours of \( S(M) \), are shown in figures 38-41, from which the following stability margins are obtained:

- parameter (mass) gain margin = 1 kg
- parameter (mass) phase margin = 60°.

The gain margin of 1 kg corresponds to reducing the carriage mass to zero before instability occurs. The parameter
Figure 38. Sheet 1 of parameter (mass) surface

Figure 39. Sheet 2 of parameter (mass) surface
Figure 40. Sheet 3 of parameter (mass) surface

Figure 41. Sheet 4 of parameter (mass) surface
surface also shows that there is a maximum limit to the carriage mass, for stability, of 2.125 kg. The phase margin of 60° indicates adequate damping of the closed-loop system.

7.4 Future research

It is thought that parameter surfaces may prove to be useful in the design of parameter dependent controllers for systems in which a particular parameter suffers large variations during normal operation. For example, the controller of an aircraft engine needs to operate satisfactorily over a wide range of altitudes. A possible design scheme could be

(i) to design real constant controllers at a number of altitudes,
(ii) to obtain an altitude dependent controller by "matrix interpolation", and finally
(iii) to analyse the stability of the system over the whole working range using the "altitude surface".

As indicated in this chapter the methods put forward in this work are not only applicable to gain and frequency, but any single system parameter and frequency. To obtain stability tests in terms of more than one parameter variation is a complicated problem, but one with great practical significance. It is felt that valuable insight into this problem may stem from a study of functions of several complex variables.

In recent years stability tests have been developed for two-dimensional and multi-dimensional digital filters [5]. Two-dimensional digital filters are used widely in many fields, such as image processing, and geophysics for the processing of seismic, gravity and magnetic data. It is expected that higher dimensional filters will also find applications; for example, three-dimensional filters in holography. The stability tests which have emerged are of a Nyquist-type and it is thought that
a deeper understanding of these developments by control engineers may lead to suitable adaption for use in the control field; possibly in the study of stability under multi-parameter variations.

A constant theme throughout this work has been the association of a multivariable system with one or possibly more algebraic functions each of which is defined on an appropriate Riemann surface. A Riemann surface for an algebraic function is topologically equivalent to a sphere with "handles", and the number of handles is known as the genus number of the surface [6]. The genus number may prove to be an important characteristic of a system and it is felt that it may be related to "decoupling", that is, the transformation of a multivariable system to a system which is effectively a set of single-input, single-output systems. A single-input, single-output system has a corresponding trivial (one sheeted) frequency surface of genus zero, so that to decouple a multivariable system would be equivalent to reducing an m-sheeted Riemann surface, with a genus number greater than or equal to zero, to a set of m trivial Riemann surfaces each of zero genus. Consequently the establishment of relationships between branch-point singularities (whose nature and location determine such topological properties as the genus number) and decoupling could be a fruitful line of research. In appendix 7 an interesting relationship is developed which shows that a branch point on a gain surface corresponds to a branch point or stationary point on the corresponding frequency surface and vice versa.
References


References


Appendix 1. Definition of an algebraic function

Let \( \Lambda(q,v) \) be a polynomial in \( q \) of the form
\[
\Lambda(q,v) = f_0(v)q^m + f_1(v)q^{m-1} + \ldots + f_m(v) \quad (A1.1)
\]
where each coefficient \( \{f_i(v): i=1,2,\ldots,m\} \) is itself a polynomial in \( v \) with coefficients in the domain of complex numbers. Then an algebraic function is a function \( q(v) \) defined for values of \( v \) in the complex \( v \)-plane by an equation of the form
\[
\Lambda(q,v) = 0 \quad (A1.2)
\]
The polynomial \( \Lambda(q,v) \) can be rewritten as a polynomial in \( v \) with coefficients which are themselves polynomials in \( q \), and when considered in this way equation (A1.2) defines an algebraic function \( v(q) \).

For a fixed value of \( v \), \( v_0 \) say, equation (A1.2) has \( m \) solutions which are called branches of \( q(v) \), and in the neighbourhood of \( v_0 \) the branches are representable by power series expansions [1].

It is assumed in the above definition that \( \Lambda(q,v) \) is an irreducible polynomial in \( (q,v) \), that is, that \( \Lambda(q,v) \) is not the product of two or more polynomials in \( (q,v) \). If \( \Lambda(q,v) \) was expressible as a product of \( t \) polynomials in \( (q,v) \), equation (A1.2) would then define \( t \) algebraic functions of the form \( q(v) \), or \( t \) algebraic functions of the form \( v(q) \).

Appendix 2. A reduction to the irreducible rational canonical form

In this appendix a proposed method is given for reducing any \( m \)-square matrix to its irreducible rational canonical form. The method used is a variation on that given
by Ayres [2] for finding the rational canonical form of a square matrix. The rational canonical form of a square matrix \( G \) is similar to the irreducible form except that its diagonal blocks correspond to the invariant factors of \( gI-G \), rather than the irréducible factors. The procedure given by Ayres [2] for finding the rational canonical form is outlined below along with some necessary definitions.

**Definitions:** If the vectors

\[
X, GX, G^2X, \ldots, G^{t-1}X
\]

(A2.1)

are linearly independent but

\[
X, GX, G^2X, \ldots, G^{t-1}X, G^tX
\]

(A2.2)

are not then (A2.1) is called a chain of length \( t \) having \( X \) as its leader.

**Procedure:** For a given \( m \)-square matrix \( G \) over any field \( F \):

(i) let \( X_m \) be the leader of a chain \( C_m \) of maximum length for all \( m \)-vectors over \( F \);

(ii) let \( X_{m-1} \) be the leader of a chain \( C_{m-1} \) of maximum length (any member of which is linearly independent of the preceding members and those of \( C_m \)) for all \( m \)-vectors over \( F \) which are linearly independent of the vectors of \( C_m \);

(iii) let \( X_{m-2} \) be the leader of a chain \( C_{m-2} \) of maximum length (any member of which is linearly independent of the preceding members and those of \( C_m \) and \( C_{m-1} \)) for all \( m \)-vectors over \( F \) which are linearly independent of the vectors of \( C_m \) and \( C_{m-1} \);

and so on. Then, for

\[
E = \begin{bmatrix}
X_j, GX_j, \ldots, G^{t_j-1}X_j, X_{j+1}, GX_{j+1}, \ldots, G^{t_{j+1}-1}X_{j+1}; \\
\vdots \\
X_m, GX_m, \ldots, G^{t_m-1}X_m
\end{bmatrix}
\]
we have that $E^{-1}GE$ is the rational canonical form of $G$.

In this approach the chains of maximum length are used to pick out the invariant factors. Now the invariant factors are made up from products of the irreducible characteristic equations which are required in the irreducible rational canonical form. Therefore, if instead of using chains of maximum length those of minimum length are found, the transformation matrix $E$ so formed is that required to give the irreducible rational canonical form $Q$.

The problem with both of these methods is that no indication is given which enables one to know when a chain of maximum or minimum length has been obtained, except that in the contrary case there will appear a chain of longer or shorter length.

It is interesting to note that the chains of maximum and minimum length form bases for the $G$-invariant subspaces of maximum and minimum dimensions respectively. Therefore the problem of finding the chains of minimum length is equivalent to that of finding the invariant subspaces of minimum dimension. It also happens that just as the $l$-dimensional invariant subspace (eigenvector) picks out an eigenvalue, a $t$-dimensional invariant subspace (from the set of those of minimum dimensions) picks out an irreducible equation of degree $t$, defining $t$ eigenvalues.

**Example:** To find the irreducible rational canonical form of

$$G(s) = \begin{bmatrix}
\frac{(s+3)}{(s+1)^2} & \frac{(s+2)}{(s+1)^2} & 0 \\
\frac{(s-3)}{(s+1)^2} & -\frac{2}{(s+1)^2} & 0 \\
-\frac{1}{(s+1)^2} & -\frac{(s+2)}{2(s+1)^2} & \frac{1}{s+1}
\end{bmatrix}$$
Let $X = (0 \ 0 \ 1)^t$, where "t" denotes the transpose,
then $GX = \frac{1}{(s+1)} \ X$
and $X$ is a chain of minimum length.
Let $Y = (0 \ 1 \ 0)^t$
then $GY = \begin{bmatrix} \frac{s+2}{(s+1)^2} & -2 & \frac{-(s+2)}{2(s+1)^2} \\ \frac{s-2}{(s+1)^3} & \frac{s+2}{(s+1)^3} & \frac{-(s+2)}{2(s+1)^3} \end{bmatrix}^t$
and $G^2Y = \begin{bmatrix} \frac{s+2}{(s+1)^3} & \frac{s-2}{(s+1)^3} & \frac{-(s+2)}{2(s+1)^3} \end{bmatrix}^t = \frac{1}{(s+1)} \ GX + \frac{s}{(s+1)^3} \ Y$
No chain of smaller length can be found and therefore $Y, GX$
completes the set of minimum length chains. The transformation
matrix $E(s)$ is therefore given by
$$E(s) = \begin{bmatrix} X & Y & GX \end{bmatrix} = \begin{bmatrix} 0 & 0 & \frac{s+2}{(s+1)^2} \\ 0 & 1 & -\frac{2}{(s+1)^2} \\ 1 & 0 & \frac{-(s+2)}{2(s+1)^2} \end{bmatrix}$$
and
$$Q(s) = E^{-1}(s)G(s)E(s) = \begin{bmatrix} \frac{1}{s+1} & 0 & 0 \\ 0 & 0 & \frac{s}{(s+1)^3} \\ 0 & 1 & \frac{1}{(s+1)} \end{bmatrix}$$
which implies that the irreducible characteristic equations
are
$$\Lambda_1(g, s) = g - \frac{1}{s+1}$$
$$\Lambda_2(g, s) = g^2 - \frac{1}{(s+1)} \ g - \frac{s}{(s+1)^3}$$
Note that these are also obvious from the dependence relations
obtained when finding the chains of minimum length.
Appendix 3. The discriminant

In this appendix two methods are given for finding the discriminant of an equation of the form
\[ \Phi(g,s) = b_0(s)g^t + b_1(s)g^{t-1} + \ldots + b_t(s) = 0 \]

**Method 1 (Barnett [3])**

The resultant \( R[a(g), c(g)] \) of two polynomials \( a(g) \) and \( c(g) \), given by
\[
\begin{align*}
a(g) &= a_0g^n + a_1g^{n-1} + \ldots + a_n \\
c(g) &= c_0g^m + c_1g^{m-1} + \ldots + c_m
\end{align*}
\]
where \( a_i, c_i \in \mathbb{C} \), is the determinant

\[
R[a(g), c(g)] = \begin{vmatrix}
a_0 & a_1 & a_2 & \ldots & a_n & 0 \\
0 & a_0 & a_1 & \ldots & a_{n-1} & a_n \\
& \ddots & \ddots & \ddots & \ddots & \ddots \\
& & \ddots & \ddots & \ddots & \ddots \\
& & & \ddots & \ddots & a_{n-1} & a_n \\
& & & & c_0 & c_1 & \ldots & c_{m-1} & c_m \\
& & & & & c_0 & c_1 & c_2 & \ldots & c_m & 0 \\
& & & & & & c_0 & c_1 & c_2 & \ldots \\
& & & & & & & c_0 & c_1 & c_2 & \ldots
\end{vmatrix}
\]

The polynomials \( a(g) \) and \( c(g) \) have a common factor (of degree greater than zero) if and only if the \((n+m)\)-order determinant \( R[a(g), c(g)] \) is zero, provided that \( a_0 \) and \( c_0 \) are not both zero.

Let the derivative with respect to \( g \) of the polynomial \( a(g) \) be denoted by \( a'(g) \). Then the discriminant of the polynomial \( a(g) \) is the determinant \( D_g(a_0, a_1, \ldots, a_n) \) defined
The polynomial \( a(g) \) has a repeated factor if and only if the discriminant \( D_g(a_0, \ldots, a_n) \) is zero.

Now consider a polynomial of the type
\[
\phi(g,s) = b_0(s)g^t + b_1(s)g^{t-1} + \ldots + b_t(s), \quad t > 0
\]
where the coefficients \( \{b_i(s) : i=1,2,\ldots,t\} \) are all polynomials in \( s \). The discriminant of this polynomial in \( g \) is found from \( D_g(b_0,b_1,\ldots,b_t) \) as defined above by replacing \( b_0,\ldots,b_t \) by \( b_0(s),\ldots,b_t(s) \) respectively. Thus there is a function \( D_g(s) \) again called the discriminant and defined by
\[
D_g(s) = \frac{1}{b_0(s)} \left[ \phi(g,s),\phi'(g,s) \right]
\]
where \( \phi'(g,s) \) is the derivative with respect to \( g \) of \( \phi(g,s) \).

**Method 2** (Sansone and Gerretsen [4])

Consider the polynomial
\[ a(g) = a_0 g^n + a_1 g^{n-1} + \ldots + a_n, \quad n > 0 \]
where \( a_i \in \mathbb{C} \); then the discriminant of \( a(g) \) is given by the expression
\[
D_g(a_0, \ldots, a_n) = a_0^{2n-2} \det P
\]
where \( P \) is a determinant given by
\[
P = \begin{vmatrix}
\sigma_0 & \sigma_1 & \ldots & \sigma_{n-1} \\
\sigma_1 & \sigma_2 & \ldots & \sigma_n \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{n-1} & \sigma_n & \ldots & \sigma_{2n-2}
\end{vmatrix}
\]
and the elements \( \{\sigma_i : i = 1,2,\ldots,2n-2\} \) are functions of
the coefficients \(\{a_i : i = 0,1,...,n\}\), and
\[\sigma_0 = n\]

The elements \(\sigma_1,...,\sigma_{n-1}\) can be found from
\[a_1 + a_0\sigma_1 = 0\]
\[2a_2 + a_1\sigma_1 + a_0\sigma_2 = 0\]
\[
\vdots \\
(n-1)a_{n-1} + a_{n-2}\sigma_1 + ... + a_0\sigma_{n-1} = 0
\]

and \(\sigma_{n},...\), \(\sigma_{2n-2}\) from
\[a_n\sigma_0 + a_{n-1}\sigma_1 + ... + a_0\sigma_n = 0\]
\[a_n\sigma_1 + a_{n-1}\sigma_2 + ... + a_0\sigma_{n+1} = 0\]
\[
\vdots \\
\]
\[a_n\sigma_m + a_{n-1}\sigma_{m+1} + ... + a_0\sigma_{n+m} = 0
\]

Consider now the polynomial \(\phi(g,s)\) given by
\[\phi(g,s) = b_0(s)g^t + b_1(s)g^{t-1} + ... + b_t(s)\]

The discriminant of this polynomial in \(g\) is found from
\[D_g(b_0,...,b_t)\] as defined above by replacing \(b_0,...,b_t\) by \(b_0(s),...,b_t(s)\) respectively. The polynomial \(\phi(g,s)\)
therefore has repeated factors if and only if
\[D_g(s) = b_0^{2t-2}(s)P\]
is zero, where \(P\) is the determinant of a matrix whose
elements are functions of the coefficients \(\{b_i(s) : i = 0,1,2,...,t\}\).
Appendix 4. A method for constructing the Riemann surface domains of the algebraic functions corresponding to an open-loop gain matrix $G(s)$

In sub-section 3.3-4 it was shown that for a characteristic gain function of degree $t$ the corresponding Riemann surface is made up from $t$ sheets of the complex $s$-plane stitched together along cuts made between branch points and infinity. Although the cuts are in some sense arbitrary (i.e. there is no unique set of cuts), it is still a problem to choose a set that is consistent, and then to be able to identify in what order the sheets are connected together. In this appendix a systematic method is given for solving this problem. The method is quite elegant in that the resulting cuts are symmetrical about the real axis and are always parallel to the imaginary axis except for possible cuts along the real axis. Also, the approach uses the open-loop gain matrix directly, so that there is no need to find the characteristic equations, and the resulting non-connected sets of connected sheets represent the Riemann surfaces for the irreducible characteristic equations.

The method is based on finding the eigenvalues of the matrix for a grid of values covering the $s$-plane and then sorting the eigenvalues in a continuous form along certain lines of the grid. If the transfer function is of order $m \times m$, $m$ arrays which will effectively represent the $m$ sheets of the Riemann surfaces are needed to store the eigenvalues. The process is analogous to the analytic continuation procedure described in sub-section 3.3-4 where individual
points of a circular disc are made bearers of unique functional values. Here individual points of the s-plane sheets (arrays) are being made the bearers of functional values.

The first line along which the eigenvalues are calculated and sorted is the real axis, and then the calculation and sorting is carried out along lines parallel to the imaginary axis and emanating from the real axis, as shown in figure 42. The calculation is only necessary in the upper half-plane since the eigenvalues in the lower half-plane are the complex conjugate of those in the upper half. Continuing this process the s-plane is covered, and m arrays of eigenvalues are obtained which are continuous along the real axis and along lines parallel to the imaginary axis but not necessarily crossing the real axis.

![Figure 42. Lines along which eigenvalues are calculated and sorted](image)

By observation of each array or sheet the necessary cuts are obvious. If parallel to the imaginary axis a line of eigenvalues is not continuous with an adjacent line then these must be separated by a cut starting at a branch point and ending at a branch point or infinity as shown in
figures 43 (a) and (b). Eigenvalues corresponding to values of $s$ in the upper half-plane are complex conjugate to those in the lower half-plane and therefore continuity of eigenvalues across the real axis is impossible if the eigenvalues situated on the real axis are complex. Therefore a cut along the real axis is necessary whenever the real axis eigenvalues are complex. Again all the cuts start at a branch point and end at a branch point or infinity, as shown in figure 44.

![Figure 43. Cuts parallel to the imaginary axis](image)

(a) finite (b) infinite

![Figure 44. Cuts along the real axis](image)
The stitching together of the sheets also becomes obvious by matching eigenvalues on one side of a cut on one sheet to those on another. When the matching process is complete, in general, there will be sets of connected sheets each set defining a Riemann surface.

To facilitate the identification of cuts and the matching of sheets it is very useful to draw the constant phase and constant magnitude contours on each sheet; this is demonstrated in the examples given in the main text.

Computationally sorting the eigenvalues along the real axis can be difficult because of the likelihood of real axis poles. This problem is overcome if the first line of calculation and sorting is changed to be a line parallel to, and a "large" distance away from, the real axis. The rest of the calculations and sorting are then carried out along lines parallel to the imaginary axis starting at this new line and finishing at the real axis; as shown in figure 45. With this new approach the only possible cuts are either along the real axis, as before (see figure 44) or between complex conjugate branch points as shown in figure 46.

A Riemann surface construction for

\[ G(s) = \frac{1}{1.25(s+1)(s+2)} \begin{bmatrix} 5s-2 & 2s-1 \\ 3s-18 & s-8 \end{bmatrix} \]

is shown using the original method in figures 47 and 48, and using the modified method in figures 49 and 50, and illustrates the arbitrariness of the cuts.
Figure 45. Lines along which eigenvalues are calculated and sorted (modified method)

Figure 46. Cut parallel to the imaginary axis (modified method)
Figure 47. Sheet 1 of the Riemann surface

Figure 48. Sheet 2 of the Riemann surface
Figure 49. Sheet 1 of the Riemann surface (modified method of construction)

Figure 50. Sheet 2 of the Riemann surface (modified method of construction)
Appendix 5. Extended Principle of the Argument

The required extension of the Argument Principle does not seem to be readily available in the literature, although a suitable statement of the Principle for a general multiple-valued analytic function has recently appeared in a text by Evgrafov [5, page 98]. Therefore, to justify its use in chapters 4 and 5, an appropriately extended Principle of the Argument is developed here.

A5.1 Introduction

The extension required is non-trivial, with two main problems to be overcome. The first of these arises from the fact that, in general, the Riemann surface of an algebraic function will be multiply connected. For an example of this source of difficulty consider a characteristic gain function $g(s)$ associated with a $2 \times 2$ open-loop gain matrix $G(s)$. Suppose that $g(s)$ has four branch points, and that each branch point is associated with a cycle of two sheets [6]; then the genus number [6; 7] of the associated Riemann surface is one. Further suppose that these branch points are disposed in the frequency plane (s-plane) in the way shown in figure 51. Then, taking two copies of the complex number sphere (Riemann number sphere), making cuts between the branch points, and forming the topological equivalent of the Riemann surface in the usual way [7] one obtains a torus, as illustrated in figures 52 and 53. The region $\Omega$, shown shaded on this torus and having a boundary $\partial \Omega$, corresponds to the interiors of the pair of Nyquist D-contours (shown in figure 52) for the original complex number spheres out of which the torus was constructed. To cope with such a situation we must ensure that the extended version of the Argument Principle holds for
Figure 51. Frequency plane

Figure 52. Two copies of the complex number sphere

Figure 53. Riemann surface shown topologically equivalent to a torus
a region of a Riemann surface which is non-simply connected and whose boundary $\Omega$ consists of several distinct closed Jordan contours on the surface. This basic extension of the principle is achieved via a suitable generalization of the Cauchy Residue Theorem and is covered in section A5.2.

The second obstacle to be overcome in a derivation of a suitably extended Argument Principle is more directly associated with the branch points of an algebraic function; in this case with their effect on the calculation of residues via contour integrals taken round the boundary $\Omega$ of a region $\Omega$ having branch points in its interior. This problem is overcome by a change of variable and is treated in section A5.3.

Finally, in section A5.4 the results of sections A5.2 and A5.3 are combined to give the required extended Principle of the Argument.

**A5.2 Generalized form of Cauchy's Residue Theorem**

The basic requirement is a generalization of Cauchy's Residue Theorem to complex functions defined on the Riemann surface of some known algebraic function, where the region of integration may be non-simply connected and have a boundary consisting of one or more closed Jordan contours. Such a generalization may be found in Bliss [6]; the main theorem is given below with slight rephrasing and change of notation for the context of this work. Two observations may be useful in helping the reader unfamiliar with Riemann surface theory to understand this theorem.

(i) Each non-singular place (or point) on the Riemann surface of an algebraic function $f(s)$ is uniquely defined by a
pair of values $(f,s)$ which satisfy a known algebraic equation

$$\Delta(f,s) = 0$$

Thus at each point on the Riemann surface the value of a complex function $\Psi(f,s)$ of the pair of complex variables $s$ and $f$ is defined.

(ii) The Riemann surface for an algebraic function is an orientable surface. Thus one can unambiguously define a positive sense in which a boundary $\partial \Omega$ can be traversed so as to "enclose" some region $\Omega$. The simplest way in which to visualize this procedure is to imagine someone walking round the boundary $\partial \Omega$ on a two-dimensional manifold in which $\Omega$ lies (embedded in a familiar space of three dimensions). The positive sense of traversal may then be taken as that in which someone walking forward round any portion of $\partial \Omega$ will always have $\Omega$ on his right-hand side. In figure 53 the boundary $\Omega$ is shown with a positive orientation.

**Generalized Residue Theorem**

Let $\mathcal{R}$ be the Riemann surface of an algebraic function $f(s)$ and $\Omega$ a portion of $\mathcal{R}$ not necessarily simply connected but having a boundary $\partial \Omega$ consisting of one or more closed Jordan contours not passing through any singular place on $\mathcal{R}$. Then if a function $\Psi(f,s)$ of the places on $\mathcal{R}$ is analytic on $\Omega$ and $\partial \Omega$ except possibly for a finite number of singular points in $\Omega$, we have

$$\frac{1}{2\pi i} \int_{\partial \Omega} \Psi(f,s) \, ds = \Sigma \text{ residues of } \Psi(f,s) \text{ in } \Omega \quad (A5.2.1)$$

where the boundary $\partial \Omega$ is traversed in the positive sense with respect to $\Omega$. 
To derive the extended Argument Principle we will first apply the Generalized Residue Theorem to the function $f'(s)/f(s)$, where $f'(s)$ denotes the derivative of $f(s)$ with respect to $s$, to give

$$\frac{1}{2\pi i} \int_{\partial \Omega} \frac{f'(s)}{f(s)} \, ds = \sum \text{residues of } \frac{f'(s)}{f(s)} \text{ in } \Omega$$

Note that $f(s)$ must have no poles or branch points (either of which are referred to as singular places) on the boundary $\partial \Omega$. For the purposes of deriving an Argument Principle we will now also exclude any zeros of $f(s)$ from $\partial \Omega$. The next step is to calculate the residues of $f'(s)/f(s)$ in $\Omega$.

**A5.3 Calculation of Residues**

The poles, zeros and branch points of $f(s)$ are singularities of $f'(s)/f(s)$, a fact which will become clearer as this section progresses. We will consider all the branch points of $f(s)$ but it turns out that only those which are also poles or zeros of $f(s)$ are relevant.

In the neighbourhood of a pole or a zero (not including those which are associated with a branch point) $f(s)$ can be represented by a series which defines a single-valued function. For a pole $p_1$ of order $t_1$, the algebraic function can be represented by

$$f(s) = \sum_{n=-t_1}^{\infty} a_n (s-p_1)^n, \quad a_{-t_1} \neq 0 \quad (A5.3.1)$$

i.e. $f(s) = \frac{\phi(s)}{(s-p_1)^{t_1}}$ \quad (A5.3.2)

where $\phi(p_1) \neq \infty$ and $\phi(s)$ is analytic in the neighbourhood.
of $p_i$. This gives

$$\frac{f'(s)}{f(s)} = \frac{-t}{s-p_i} + \frac{\phi'(s)}{\phi(s)} \quad (A5.3.3)$$

and since $\frac{\phi'(s)}{\phi(s)}$ is analytic in the neighbourhood of $p_i$

the function $\frac{f'(s)}{f(s)}$ has a simple pole of residue $-t$ at $s = p_i$.

For a zero $s_i$ of order $t$, the algebraic function can be represented by the series

$$f(s) = \sum_{n=t}^{\infty} c_n (s-z_i)^n, \quad c_t \neq 0 \quad (A5.3.4)$$

i.e. $f(s) = (s - z_i)^t \theta(s) \quad (A5.3.5)$

where $\theta(z_i) \neq 0$ and $\theta(s)$ is analytic in the neighbourhood of $z_i$. This gives

$$\frac{f'(s)}{f(s)} = \frac{t}{s-z_i} + \frac{\theta'(s)}{\theta(s)} \quad (A5.3.6)$$

and we see that $\frac{f'(s)}{f(s)}$ has a simple pole of residue $t$ at $s = z_i$.

To determine the residues at a branch point it is necessary to make a change of variable \cite{6, page 81}. The procedure is as follows.

In the neighbourhood of a finite branch point $b_i$ the algebraic function can be represented by a series of the form \cite{8},
where \( r \) is the number of sheets which form a cycle at the branch point. If the branch point is a pole of order \( t^b_{P_i} \) and consists of a cycle of \( r \) sheets we have the series expansion
\[
f(s) = \sum_{n=-\infty}^{\infty} d_n (r/s - b_i)^n, \quad d_{-t^b_{P_i}} \neq 0 \quad (A5.3.8)
\]
Similarly if the branch point is a zero of order \( t^b_{z_i} \) we have that
\[
f(s) = \sum_{n=0}^{\infty} d_n (r/s - b_i)^n, \quad d_{-t^b_{z_i}} \neq 0 \quad (A5.3.9)
\]
otherwise
\[
f(s) = \sum_{n=0}^{\infty} d_n (r/s - b_i)^n, \quad d_o \neq 0 \quad (A5.3.10)
\]
These expansions are multivalued and therefore not suitable for determining residues. If, however, we make the substitution
\[
s = b_i + x^r \quad (A5.3.11)
\]
we find that the series expansions (A5.3.8, 9 and 10) become
\[
f(x) = \sum_{n=-t^b_{P_i}}^{\infty} d_n x^n, \quad d_{-t^b_{P_i}} \neq 0 \quad (A5.3.12)
\]
\[
f(x) = \sum_{n=t^b_{z_i}}^{\infty} d_n x^n, \quad d_{t^b_{z_i}} \neq 0 \quad (A5.3.13)
\]
and
\[ f(x) = \sum_{n=0}^{\infty} d_n x^n, \quad d_0 \neq 0 \quad (A5.3.14) \]

respectively, where it is to be understood that

\[ f(x) \equiv f(b_1 + x^I) \quad (A5.3.15) \]

The substitution has therefore mapped the algebraic function onto the x-plane where it can be represented by a "single-valued" series expansion.

By definition \([6]\) the residue at \(b_1\) of \(f'(s)\) analytic in a neighbourhood of \(b_1\) except possibly at \(b_1\) itself is

\[ (\text{residue})_{b_1} \triangleq \frac{1}{2\pi i} \oint_{C} \frac{f'(s)}{f(s)} \, ds \quad (A5.3.16) \]

where \(C\) is a closed positively oriented Jordan contour on \(\mathbb{R}\) bounding a neighbourhood \(N\) of \(b_1\) in which \(\frac{f'(s)}{f(s)}\) is analytic except possibly at \(b_1\). If we substitute for \(s\) from \((A5.3.11)\) we find that

\[ (\text{residue})_{b_1} \triangleq \frac{1}{2\pi i} \oint_{C} \frac{f'(s)}{f(s)} \, ds \]

\[ = \frac{1}{2\pi i} \oint_{C_x} \frac{f'(x)}{f(x)} \frac{dx}{ds} \, ds \]

\[ = \frac{1}{2\pi i} \oint_{C_x} \frac{f'(x)}{f(x)} \, dx \quad (A5.3.17) \]

where \(C_x\) is a closed positively oriented Jordan contour on the x-plane bounding a neighbourhood of the origin. The residue of \(\frac{f'(s)}{f(s)}\) at a branch point \(b_1\) is therefore equal to the residue of \(\frac{f'(x)}{f(x)}\) at the origin. The residue of the function
\( \frac{f'(x)}{f(x)} \) is easily obtained from its series expansion. Following the same procedure as before we note that \( \frac{f'(x)}{f(x)} \) in the neighbourhood of the origin has:

(i) a simple pole with residue \( -t^b_{pi} \), if the branch point is a pole of order \( t^b_{pi} \); or

(ii) a simple pole with residue \( t^b_{zi} \), if the branch point is a zero of order \( t^b_{zi} \);

otherwise \( \frac{f'(x)}{f(x)} \) is analytic.

**A5.4 Extended Principle of the Argument.**

Combining the results of section A5.3 with that of equation (A5.2.2) we have that

\[
\frac{1}{2\pi i} \oint_{\partial \Omega} \frac{f'(s)}{f(s)} \, ds = \left[ \sum_{i} t^b_{zi} + \sum_{i} t^b_{pi} \right] - \left[ \sum_{i} t^b_{zi} + \sum_{i} t^b_{pi} \right] = \#Z - \#P \tag{A5.4.1}
\]

where \( \#Z \) and \( \#P \) are respectively the total number of poles and zeros of \( f(s) \) in \( \Omega \) enclosed by the boundary \( \partial \Omega \) traversed in the positive sense with respect to \( \Omega \), and where \( \Omega \) lies in the Riemann surface of the algebraic function \( f(s) \). Note that multiple poles and zeros must be counted as many times as their orders indicate.

The left-hand side of equation (4.1) is equivalent to the net sum of the clockwise encirclements of a curve set \( \Gamma \) about the origin of the \( f \)-plane where \( \Gamma \) is the image of \( \partial \Omega \) under \( f(s) \). Denoting this net sum of clockwise encirclements by \( N(\Gamma, \Omega) \), we finally obtain the required extended Principle of the Argument in the form

\[
N(\Gamma, \Omega) = \#Z - \#P \tag{A5.4.2}
\]
Appendix 6. Multivariable pivots from the characteristic

equation \Delta(g,s)=0

Consider the characteristic equation
\begin{equation}
\Delta(g,s) = 0 \quad (A6.1)
\end{equation}
and assume we have found an approximation for a branch of the
characteristic frequency loci about s=m, as k=\infty, of the form
\begin{equation}
s = bk^{\alpha} \quad (A6.2)
\end{equation}
If p is the pivot corresponding to this asymptote then for k=\infty
\begin{equation}
s = p + bk^{\alpha} \quad (A6.3)
\end{equation}
or
\begin{equation}
k = \frac{1}{b} \left( \frac{s-p}{k^{\alpha}} \right) \quad (A6.4)
\end{equation}

The dependence of the closed-loop poles on k is given by
equation (6.1.5), i.e.
\begin{equation}
\Gamma(k,s) = 0 \quad (A6.5)
\end{equation}
Therefore substituting for k or s in equation (A6.5) we
obtain a relationship between s and p or a relationship between
k and p. Since the order of s will in general be much
greater than the order of k in equation (A6.5), we will consider
the case of substituting for k. Substituting for k from
equation (A6.4) into equation (A6.5) we obtain the equation
\begin{equation}
Z(p,s) = 0 \quad (A6.6)
\end{equation}
We require p for s=\infty, and so we put s=z^{-1} in equation (A6.6)
to give
\begin{equation}
Z(p,s) = Z(p,z^{-1}) = z^{-t} \Omega(p,z) = 0 \quad (A6.8)
\end{equation}
where t is the maximum order of s in (A6.6). The equation
\begin{equation}
\Omega(p,0) = 0 \quad (A6.8)
\end{equation}
then gives the multivariable pivot, \( p \).

**Example**

Consider the open-loop gain matrix

\[
G(s) = \begin{bmatrix}
-\frac{1}{s^4} & -s^3 - 2s^2 + 29s + 92 & -20s^2 + 35s + 70 \\
-33s^2 - s - 91 & 33s^2 - 170s + 118
\end{bmatrix}
\]

which has a characteristic equation

\[
\Delta(g,s) = (s^4 - s^3 + 2s^2 - 25s + 29)g^2 - (s^3 + 22s^2 - 199s + 210)g + (-33s + 594) = 0
\]

Using the method described in section 6.2 we find that as \( k \to \infty \) the infinite branches of the characteristic frequency loci are given by the following approximations

\[ s \approx k \text{ and } s \approx -33k \]

To find the pivot corresponding to the second-order Butterworth pattern \( j/(33k) \) we will put

\[ s = p + j(-33k) \]

giving

\[ k = \frac{(s-p)^2}{-33} \]

From the characteristic equation \( \Delta(g,s) = 0 \) we obtain

\[
\Gamma(s,k) = s^4 - s^3 + 2s^2 - 25s + 29 + (-s^3 + 22s^2 - 199s + 210)k + (-33s + 594)k^2 = 0
\]

and substituting for \( k \) in this we obtain

\[
Z(p,s) = 33^2(s^4 - s^3 + 2s^2 - 25s + 29) - 33(-s^3 + 22s^2 - 199s + 210)(-2ps + p^2)
\]

\[
-33s^2(22s^2 - 199s + 210) + (-33s + 594)(-4s^3 + 6s^2p^2 - 4sp^3 + p^4)
\]

\[
+594s^4 = 0
\]

Putting \( s = z^{-\frac{1}{2}} \) we find

\[
\Omega(p,z) = 33^2(1 + z^2 - 25z^3 + 29z^4)
\]

\[
-33(-1 + 22z - 199z^2 + 210z^3)(-2p + z)(-2p^2 + z^2)
\]

\[
-33(22 - 199z + 210z^2)
\]

\[
+(-33 + 594z)(-4p + 6zp^2 - 4z^2p^3 + z^3p^4) + 594 = 0
\]
Therefore
\[ \Omega(p,0) = 33^2 - 66p - 33 \times 22 + 33 \times 4p + 594 = 0 \]
from which we find
\[ p = -14.5 \]

Appendix 7. Association between branch points and stationary points on the gain and frequency surfaces

Let us suppose that from an open-loop gain matrix \( G(s) \), we have obtained an algebraic equation in terms of the complex gain variable \( g \), and the complex frequency variable \( s \), given by

\[ F = f(g,s) = 0 \quad \text{(A7.1)} \]

Let us also consider the following equations involving the derivatives of \( F \):

\[ \left( \frac{\partial F}{\partial g} \right)_s = 0 \quad \text{(A7.2)} \]

and

\[ \left( \frac{\partial F}{\partial s} \right)_g = 0 \quad \text{(A7.3)} \]

The values of \( s \) simultaneously satisfying equation (A7.1 and .2) are the branch points on the frequency surface; the values of \( g \) simultaneously satisfying equations (A7.1 and .3) being the branch points on the gain surface.

If we consider \( g \) as a function of \( s \), the total derivative of \( F \) with respect to \( s \) is

\[ \frac{dF}{ds} = \left( \frac{\partial F}{\partial g} \right)_g + \left( \frac{\partial F}{\partial s} \right)_g \frac{dg}{ds} \quad \text{(A7.4)} \]

Alternatively we can consider \( s \) as a function of \( g \), and obtain

\[ \frac{dF}{dg} = \left( \frac{\partial F}{\partial g} \right)_s + \left( \frac{\partial F}{\partial s} \right)_g \frac{ds}{dg} \quad \text{(A7.5)} \]

Now from equation (A7.1), \( F \) is identically zero and therefore
its total derivatives must also be zero, so that from equations (A7.4 and .5) we have the following:

\[
\frac{\partial F}{\partial s} = -\frac{\partial F}{\partial g} \frac{dg}{ds} \tag{A7.6}
\]

and

\[
\frac{\partial F}{\partial g} = -\frac{\partial F}{\partial s} \frac{ds}{dg} \tag{A7.7}
\]

If we have a branch point on the gain surface equation (A7.3) is satisfied and hence from equation (A7.6) we have that

\[
\left( \frac{\partial F}{\partial g} \right)_s = 0 \quad \text{or} \quad \frac{dg}{ds} = 0
\]

which imply that on the frequency surface we have either a branch point or a stationary point corresponding to the gain surface branch point.

Alternatively if we have a branch point on the frequency surface equation (A7.2) is satisfied and hence from equation (A7.7) we have that

\[
\left( \frac{\partial F}{\partial s} \right)_g = 0 \quad \text{or} \quad \frac{ds}{dg} = 0
\]

which imply that on the gain surface we have either a branch point or a stationary point corresponding to the frequency surface branch point.

References


170


Bibliography


