Third Edition

MODERN CONTROL THEORY

William L. Brogan

Modern Control Theory

Modern Control Theory

Third Edition

WILLIAM L. BROGAN, Ph.D.
Professor of Electrical Engineering
University of Nevada, Las Vegas

PRENTICE HALL, Upper Saddle River, NJ 07458

Library of Congress Cataloging-in-Publication Data

Brogan, William L.
Modern control theory / William L. Brogan. -- 3rd ed.
p. cm.
Includes bibliographical references.
ISBN 0-13-589763-7
1. Control theory. I. Title.
QA402.3.B76 1991
629.8'312--dc20 90-6977
CIP

Editorial/production supervision
and interior design: Patrice Fraccio/Bayani Mendoza de Leon
Cover design: Karen Stephens
Manufacturing buyers: Lori Bulwin/Linda Behrens

Printed in the United States of America
20 19 18 17

ISBN 0-13-589763-7

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall of Canada, Inc., *Toronto*
Prentice-Hall Hispanoamericana, S. A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Prentice-Hall Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

This book is dedicated to my parents Lloyd and Alice,
and to my grandchildren Jesse and Jacque,
representatives of the next generation.

Contents

Preface

This is a text on state variable modelling, analysis and control of dynamical systems. The analytical approach to many control problems consists of three major steps: (1) develop an idealized mathematical representation of the real physical system, (2) apply mathematical analysis and design techniques to the model and (3) interpret the mathematical results in terms of implications on the real, physical system. If the resulting implications are not acceptable or do not seem to match reality or experimental observations, one or more of the above steps may need to be modified and repeated. This book attempts to illustrate these steps and some of the tools that may be involved. While most of the steps are mathematically based, this is intended as an engineering text. No abstract theorems/proofs are included just to enhance the mathematical elegance.

The first two chapters are intended as a review of prerequisite introductory courses on modelling and control of physical systems using primarily a transfer function approach. These are not intended to be complete, stand-alone treatments. Rather, a distilled summary of the essentials is presented. It has been observed that students often get bogged down in the myriad of details of these prerequisite courses, thus losing sight of the big picture. The intent here was to place the major points in perspective before going on to the state variable approach.

State variables, state vectors, state space and linear system matrices are introduced in Chapter 3. Methods of obtaining state variable models from other system descriptions are provided. In previous editions this chapter came after a series of chapters on matrix theory and linear algebra. The present, inverted approach provides early motivation for the need to master the mathematics of matrices and linear algebra. It also allows earlier introduction of control-related examples, which appear throughout the subsequent chapters. Chapters 4 through 8 provide a thorough development of the needed mathematical tools. The treatment has been considerably revised, based on experience gained with previous editions, as well as helpful comments from reviewers.

The depth to which the mathematical topics need to be pursued depends upon the preparation of the reader and the level of understanding needed. My recent experience has been that most students approach this book after having had a first course in linear algebra. It still appears fruitful to cover Chapters 4 through 8 fairly carefully, and students gain new insights from the large number of engineering-motivated examples. Not all of the more abstract topics need to be covered in an undergraduate course, however. A more advance graduate level approach would merely skim the early sections of these chapters, but put more emphasis on the extensions and proof found in the problems.

In addition to reversing the order of presentation mentioned above, this book deleted some interesting but peripheral topics, to make room for new material which is more central to the controls field. Additional topics now covered include QR decomposition of a matrix, which can be used to iteratively solve for eigenvalues and eigenvectors. More importantly, it is useful in finding the Kalman controllable and/or observable canonical forms. These provide a very satisfactory method of determining minimal realizations in Chapter 11. A portion of the material on matrix fraction description of systems, and application to controller/observer design is now included. This Diophantine equation approach supplements and provides an insightful alternative to the state variable approach. The treatment of optimal control has been revised to emphasize the linear quadratic problem and the associated Riccati equations. The question of robustness is addressed in an introductory manner. Two extensions to the LQ theory are introduced; projective controls is a method of designing low order controllers which preserve the dominant modes of a full order optimal controller, and frequency-weighted cost functions which lead to dynamic controllers capable of coping with modelling approximations.

More emphasis is given to stability of time varying linear systems and a new chapter provides some tools for approaching nonlinear system problems.

As in the previous editions, the problems, especially the illustrative problems for which complete solutions are given, should be considered as an integral part of each chapter. Many useful results are derived and presented *only* in these problems.

This third edition has evolved from the first two, and thus all former users who sent comments or filled out review forms for Prentice Hall have contributed to this work. Those individuals who contributed more directly to the preparation of the previous editions have had a lasting impact here too. During the manuscript preparation, student feedback from several classes was very helpful. Steven Crammer, Saeed Karamooz and L. Lane Sanford deserve special thanks. Professors John Boye and George Schade from Nebraska, Hal Tharp of Arizona and Sahjendra Singh of UNLV provided comments or suggestions. Production editors Patrice Fraccio and Bayani Mendoza de Leon who handled the editorial supervision and interior design of the book, and the five anonymous reviewers who provided detailed comments to Prentice Hall editor Tim Bozik, are especially thanked. Finally, I wish to acknowledge Mailliw Nagorb for typing the manuscript.

William L. Brogan

1

Background and Preview

1.1 INTRODUCTION

Control theory is often regarded as a branch of the general, and somewhat more abstract, subject of systems theory [1].‡ The boundaries between these disciplines are often unclear, so a brief section is included to delineate the point of view of this book.

In order to put control theory into practice, a bridge must be built between the real world and the mathematical theory. This bridge is the process of modeling, and a summary review of modeling is included in this chapter [2, 3].

Control theory can be approached from a number of directions. The first systematic method of dealing with what is now called control theory began to emerge in the 1930s. Transfer functions and frequency domain techniques were predominant in these "classical" approaches to control theory. Starting in the late 1950s and early 1960s a time-domain approach using state variable descriptions came into prominence.

For a number of years the state variable approach was synonymous with "modern control theory." At the present time the state variable approach and the various transfer function–based methods are considered on an equal level, and nicely complement each other. Distinctions exist, and the major one appears to be in the kinds of mathematical tools used. The state variable approach uses linear algebra based on the real or complex number field. The newer multivariable transfer function approaches involve the algebra of polynomial matrices and related concepts. By defining the number field properly, the major mathematical tool is once again linear algebra, but on a somewhat less familiar level (see, for example, Sections 6.3.1 and 6.3.2). Some of these concepts are pointed out and used throughout this book. The older classical control theory point of view, using single-input, single-output transfer functions is

‡ Reference citations are given numerically in the text in brackets. The references are listed at the end of each chapter.

reviewed briefly in Chapter 2. However, for the most part this book is devoted to the state variable point of view.

1.2 SYSTEMS, SYSTEMS THEORY, AND CONTROL THEORY

According to the *Encyclopedia Americana,* a system is " . . . an aggregation or assemblage of things so combined by nature or man as to form an integral and complex whole . . . ." Mathematical systems theory is the study of the interactions and behavior of such an assemblage of "things" when subjected to certain conditions or inputs. The abstract nature of systems theory is due to the fact that it is concerned with mathematical properties rather than the physical form of the constitutent parts.

Control theory is more often concerned with physical applications. A control system is considered to be any system which exists for the purpose of regulating or controlling the flow of energy, information, money, or other quantities in some desired fashion. In more general terms, a control system is an interconnection of many components or functional units in such a way as to produce a desired result. In this book control theory is assumed to encompass all questions related to design and analysis of control systems.

Figure 1.1 is a general representation of an *open-loop* control system. The input, or control, $\mathbf{u}(t)$ is selected based on the goals for the system and all available a priori knowledge about the system. The input is in no way influenced by the output of the system, represented by $\mathbf{y}(t)$. If unexpected disturbances act upon an open-loop system, or if its behavior is not completely understood, then the output will not behave precisely as expected.

Another general class of control systems is the *closed-loop,* or *feedback,* control system, as illustrated in Figure 1.2. In the closed-loop system, the control $\mathbf{u}(t)$ is modified in some way by information about the behavior of the system output. A feedback system is often better able to cope with unexpected disturbances and uncertainties about the system's dynamic behavior. However, it need not be true that closed-loop control is always superior to open-loop control. When the measured outputs have errors which are sufficiently large and when unexpected disturbances are relatively unimportant, closed-loop control can have a performance which is inferior to open-loop control.

EXAMPLE 1.1 In order to provide financial security for the retirement years, a person arranges to have \$300 per month invested into an annuity account. The system "input" each month is $u(t) = \$300$. The system output $y(t)$ is the accrued value in the account. Since $u(t)$ is not affected by the current economic climate or by $y(t)$, this is an open-loop system. ■

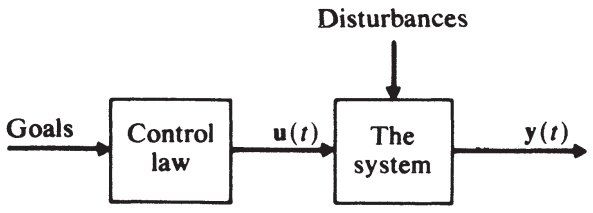


Figure 1.1 An open-loop control system.

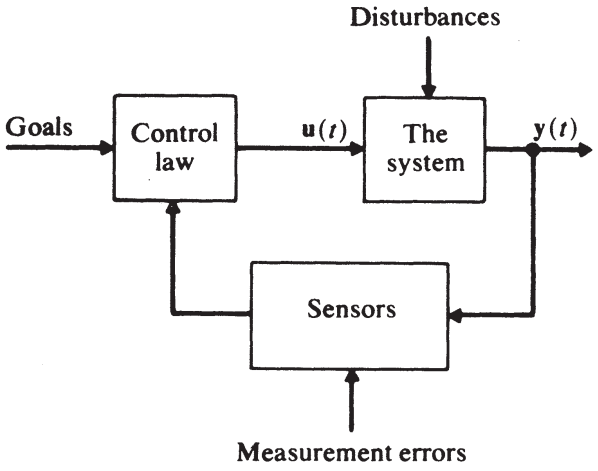


Figure 1.2 A closed-loop control system.

EXAMPLE 1.2 Another person with the same goal of financial security plans to invest in the stock market, by attempting to implement the strategy of buying-low and selling-high. The input $u(t)$ at any given time is influenced by the perceived market conditions, the past success of the stock account, and so forth. This is a feedback or closed-loop system. ■

EXAMPLE 1.3 A typical industrial control system involves components from several engineering disciplines. The automatic control of a machine shown in Figure 1.3 illustrates this. In this example, the desired time history of the carriage motion is patterned into the shape of the cam. As the cam-follower rises and falls, the potentiometer pick-off voltage is proportional to the desired carriage position. This signal is compared with the actual position, as sensed by another potentiometer. This difference, perhaps modified by a tachometer-generated rate signal, gives rise to an error signal at the output of the differential amplifier. The power level of this signal is usually low and must be amplified by a second amplifier before it can be used for corrective action by an electric motor or a servo valve and a hydraulic motor or some other prime mover. The prime mover output would usually be modified by a precise gear train, a lead screw, a chain and sprocket, or some other mechanism. Clearly, mechanical, electrical, electronic, and hydraulic components play important roles in such a system. ■

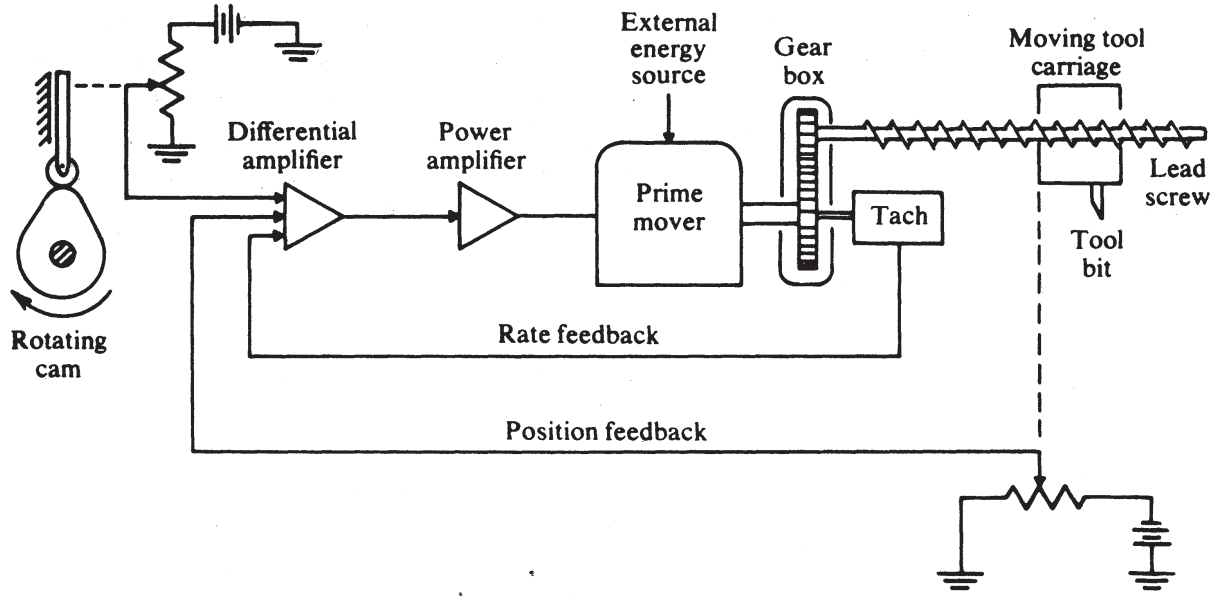


Figure 1.3

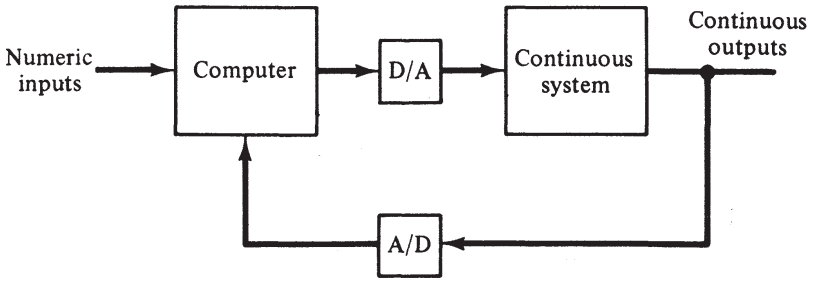


Figure 1.4

EXAMPLE 1.4 The same ultimate purpose of controlling a machine tool could be approached somewhat differently using a small computer in the loop. The continuous-time, or analog, signals for position and velocity must still be controlled. Measurements of these quantities would probably be made directly in the digital domain using some sort of optical pulse counting circuitry. If analog measurements are made, then an analog-to-digital (A/D) conversion is necessary. The desired position and velocity data would be available to the computer in numeric form. The digital measurements would be compared and the differences would constitute inputs into a corrective control algorithm. At the output of the computer a digital-to-analog (D/A) conversion could be performed to obtain the control inputs to the same prime mover, as in Example 1.3. Alternately, a stepper motor may be selected because it can be directly driven by a series of pulses from the computer. Figure 1.4 shows a typical control system with a computer in the loop. ■

1.3 MODELING

Engineers and scientists are frequently confronted with the task of analyzing problems in the real world, synthesizing solutions to these problems, or developing theories to explain them. One of the first steps in any such task is the development of a mathematical model of the phenomenon being studied. This model must not be oversimplified, or conclusions drawn from it will not be valid in the real world. The model should not be so complex as to complicate unnecessarily the analysis.

System models can be developed by two distinct methods. *Analytical modeling* consists of a systematic application of basic physical laws to system components and the interconnection of these components. *Experimental modeling,* or modeling by synthesis, is the selection of mathematical relationships which seem to fit observed input-output data. Analytical modeling is emphasized first. Some aspects of the other approach are presented in Chapter 6 (least-squares data fitting).

1.3.1. Analytical Modeling

An outline of the analytical approach to modeling is presented in Figure 1.5. The steps in this outline are discussed in the following paragraphs.

1. *The intended purposes of the model must be clearly specified.* There is no single model of a complicated system which is appropriate for all purposes. If the purpose is a detailed study of an individual machine tool, the model would be very different from one used to study the dynamics of work flow through an entire factory.

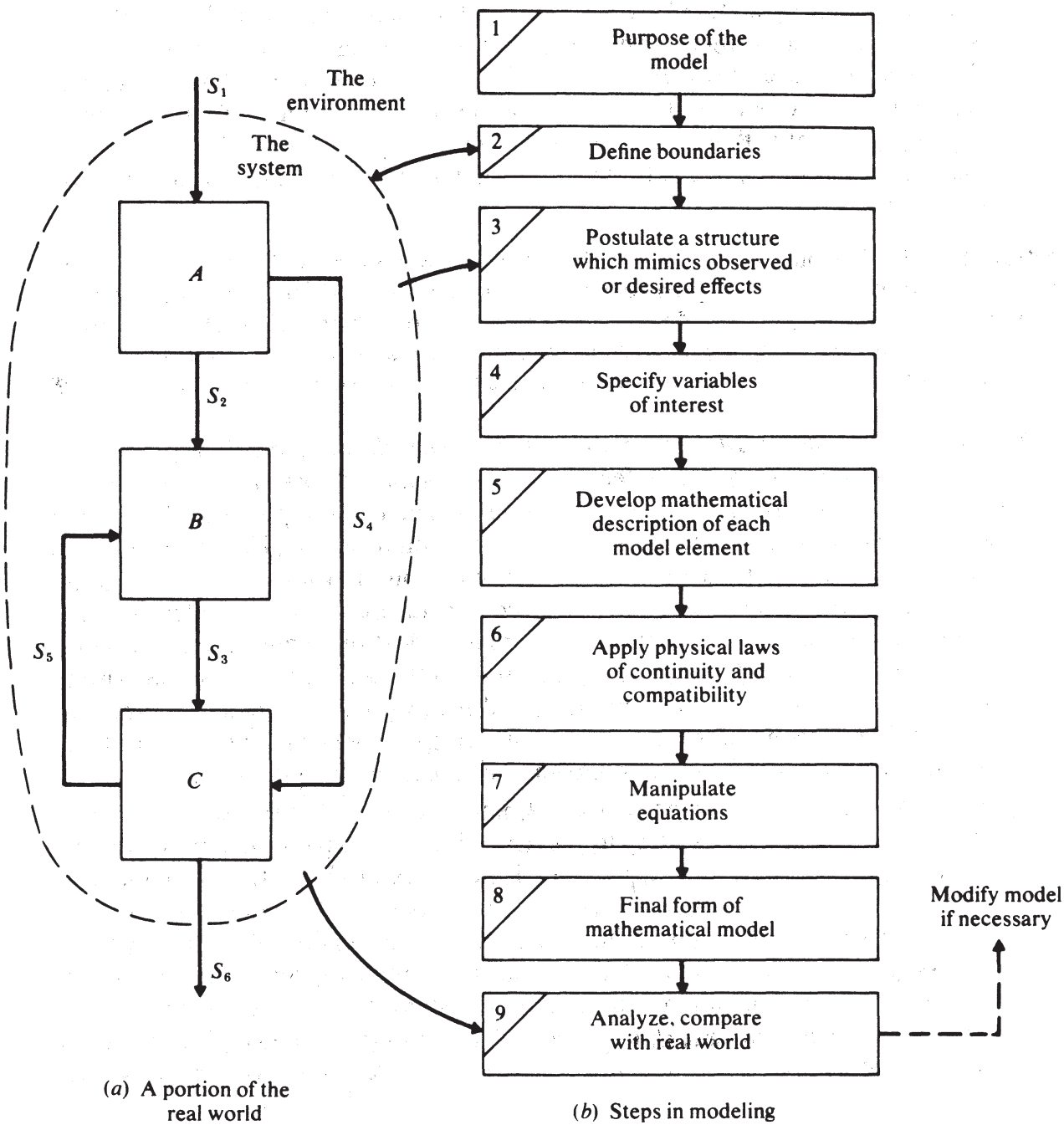


Figure 1.5 Modeling considerations.

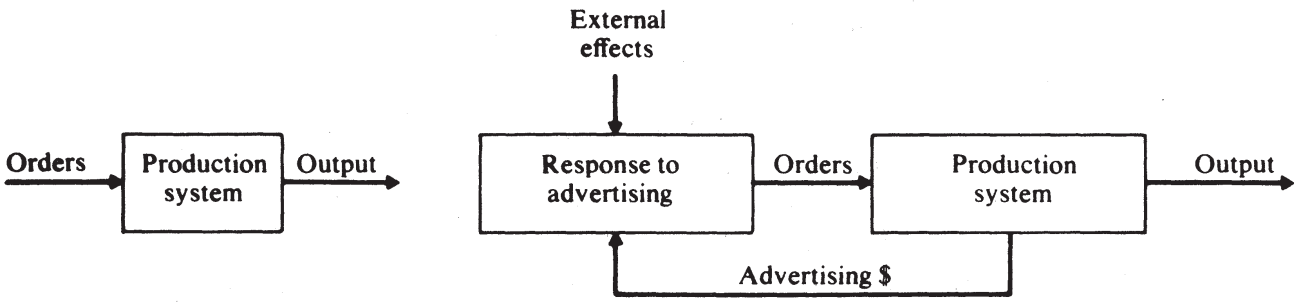
2. *The system boundary is a real or imagined separation of the part of the real world under study, called the system, and the rest of the real world, referred to as the environment.* The system boundary must enclose all components or subsystems of primary interest, such as subsystems *A*, *B*, and *C* in Figure 1.5*a*.

A second requirement on the selection of the boundary is that all causative actions or effects (called signals) crossing the boundary be more or less one-way interactions. The environment can affect the system, and this is represented by the input signal S_1. The system output, represented by the signal S_6, should not

affect the environment, at least not to the extent that it would modify S_1. If there is no interest in subsystem A, then a boundary enclosing B and C, and with inputs S_2 and S_4, could be used. Subsystem C should not be selected as an isolated system because one of its outputs S_5 modifies its input S_3 through subsystem B. The requirement is that all inputs are known, or can be assumed known for the purpose of the study, or can be controlled independently of the internal status of the system.

EXAMPLE 1.5 The purpose of the models of Figure 1.6 is to study the flow of work and information within a production system due to an input rate of orders. These orders could be an input from the environment, as in Model I. If the purpose is to study the effects of an advertising campaign, then orders are determined, at least in part, by a major system variable. In this case the rate of orders should be an internal variable in the feedback system of Model II. ■

3. *All physical systems, whether they are of an electrical, mechanical, fluid, or thermal nature, have mechanisms for storing, dissipating, or transferring energy, or transforming energy from one form to another.* The third step in modeling is one of reducing the actual system to an interconnection of simple, idealized elements which preserve the character of these operations on the various kinds of energy. An electric circuit diagram illustrates such an idealization, with ideal sources representing inputs. In mechanical systems, idealized connections of point masses, springs, and dashpots are often used. In thermal or fluid systems, and to a certain extent in economic, political, and social systems, similar idealizations are possible. This process is referred to as *physical modeling*. The level of detail required depends on the type of information expected from the model.
4. *If the physical model is properly selected, it will exhibit the same major characteristics as the real system.* In order to proceed with development of a mathematical model, variables must be assigned to all attributes of interest. If a quantity of interest does not yet exist and thus cannot be labeled, a modification will be required in Step 3 in order to include it. The classification of system types is discussed in the next section. This book deals mainly with *deterministic lumped-parameter systems*. In all lumped-parameter systems there are basically just two types of variables. They are *through variables* (sometimes called path variables or rate variables) and *across variables* (sometimes called point variables or level



(*a*) Model I (*b*) Model II

Figure 1.6

variables). Through variables flow through two-terminal elements and have the same value at both terminals. Examples are electric current, force or torque, heat flow rate, fluid flow rate, and rate of work flow through a production element. Across variables have different values at the two terminals of a device. Examples are voltage, velocity, temperature, pressure, and inventory level.

5. *Each two-terminal element in the idealized physical model will have one through and one across variable associated with it.* Multiterminal devices such as transformers or controlled sources will have more. In every device, mathematical relationships will exist between the two types of variables. These relationships, called *elemental equations,* must be specified for each element in the model. This step could uncover additional variables that need to be introduced. This would mean a modification of Step 4. Common examples of elemental equations are the current-voltage relationships for resistors, capacitors, and inductors. The form of these relations may be algebraic, differential, or integral expressions, linear or nonlinear, constant or time-varying.
6. *After a system has been reduced to an interconnection of idealized elements, with known elemental equations, equations must be developed to describe the interconnection effects.* Regardless of the physical type of the system, there are just two types of physical laws that are needed for this purpose. The first is a statement of *conservation* or *continuity* of the through variables at each node where two or more elements connect. Examples of this basic law are Kirchhoff's node equations, D'Alembert's version of Newton's second law, conservation of mass in fluid flow problems, and heat balance equations. The second major law is a *compatibility* condition relating across variables. Kirchhoff's voltage law around any closed loop is but one example. Similar laws regarding relative velocities, pressure drops, and temperature drops must also hold. Both of these laws yield linear equations in through or across variables, regardless of whether the elemental equations are linear or nonlinear. This fact is responsible for the name given to linear graphs, an extremely useful tool in applying these two laws.

EXAMPLE 1.6 Consider the system with six elements, including a source v_0, shown in Figure 1.7. Each element is represented as a branch of the linear graph, and the interconnection points are nodes. Each node is identified by an across variable v_i, and each branch has a through variable, called f_i, with the arrow establishing the sign convention for positive flow. Let

b (number of branches) $= 6$
s (number of sources) $= 1$
n (number of nodes) $= 4$

Two unknowns exist for each branch, except source branches have a single unknown. Thus there are $2b - s = 11$ unknowns, and 11 equations are needed. They are $b - s = 5$ elemental equations, $n - 1 = 3$ continuity equations (node 1 is used as a reference and is redundant)

$$f_0 - f_1 - f_2 = 0, \qquad f_2 - f_3 - f_4 = 0, \qquad f_4 - f_5 = 0$$

and $b - (n - 1) = 3$ compatibility (loop) equations. Letting $v_{ab} = v_a - v_b$, these are

$$v_{21} - v_0 = 0, \qquad v_{23} + v_{31} + v_{12} = 0, \qquad v_{34} + v_{41} + v_{13} = 0$$

These $2b - s$ equations can be used to determine all unknowns. ■

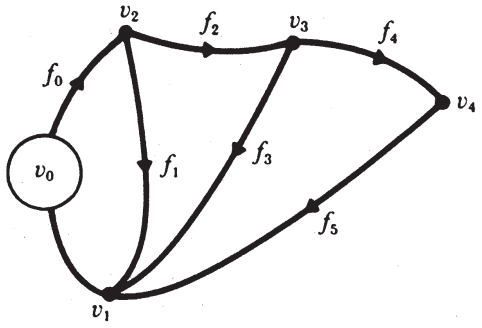


Figure 1.7

7. *This step consists of manipulating the elemental, continuity, and compatibility equations into a desired final form.* A further discussion of this step is presented in Section 1.5.
8. *Item 8 of Figure 1.5 is the end result of the modeling process.* It may be arrived at by a process of iteration, as mentioned in Step 9.
9. *The model developed in the preceding steps should never be confused with the real world system being studied.* Whenever possible, the results produced by the real system should be compared with model results under similar conditions. If unacceptable discrepancies exist, the model is inadequate and should be modified.

1.3.2 Experimental Modeling

Time series models: autoregressive, moving average, and ARMA models. In some experimental modeling situations the system structure may be based solidly on the laws of physics, and perhaps only a few key parameter values are uncertain. Even these unknown parameters may be known to some degree. Upper and lower bounds or the mean and variance or other probabilistic descriptors may be available at the outset.

In other situations, notably in areas of socioeconomic or biological systems, the only thing available is an assumed model form, which is convenient to work with and which does a reasonable job of fitting observations. All the coefficients are usually unknown and must be determined in these cases. When this situation applies, the autoregressive moving average (ARMA) model is frequently used. A brief overview follows.

A large variety of technical applications can be framed in a similar mathematical form. Let $y(k)$ be some variable of interest at a general time t_k. This might be the price of a stock or similar commodity. This is the case of an economic time series problem. In another situation $y(k)$ could be the magnitude of the acoustic waveform of speech, and the interest might be in either speech coding or synthetic speech generation. The major interest here is in the identification of unknown systems, and $y(k)$ is the system output. In all these cases it is assumed that the next value in the time series, $y(k+1)$, is influenced by the present value $y(k)$ and past (lagged) values $y(k-1)$, $y(k-2), \ldots, y(k-n)$. In addition, the signal $y(k+1)$ is normally influenced by one or more current and past input signals $u_i(k+1)$, $u_i(k-j)$ for $j = 0, 1, 2, \ldots, p-1$. It is assumed here that the time series is generated by a linear difference equation with a single input,

$$y(k+1) = a_0 y(k) + a_1 y(k-1) + a_2 y(k-2) + \cdots + a_n y(k-n) + b_0 u(k+1) + b_1 u(k) + \cdots + b_p u(k+1-p) + v(k) \tag{1.1}$$

where $v(k)$ is a random noise term. The identification problem is thus reduced to the estimation of the system coefficients

$$\boldsymbol{\theta} = [a_0 \quad a_1 \cdots a_n \quad b_0 \quad b_1 \cdots b_p]^T \tag{1.2}$$

from a series of measurements of the inputs $u(i)$ and the outputs $y(i)$. Equation (*1.1*) can be recast as

$$\begin{aligned} y(k+1) &= [y(k) \quad y(k-1) \cdots u_1(k+1) \quad u_1(k) \cdots]\boldsymbol{\theta} + v(k) \\ &= \mathbf{C}(k)\boldsymbol{\theta} + v(k) \end{aligned} \tag{1.3}$$

If the measurements of past values of $y(i)$ and $u(i)$ are sufficiently accurate, $\mathbf{C}(k)$ can be assumed known. Equation (*1.3*) is then in a form suitable for recursive least-squares estimation of the unknown parameters $\boldsymbol{\theta}$. A series of equations can be stacked into one larger equation:

$$\begin{bmatrix} y(k+1) \\ y(k+2) \\ \vdots \\ y(k+N) \end{bmatrix} = \begin{bmatrix} \mathbf{C}(k) \\ \mathbf{C}(k+1) \\ \vdots \\ \mathbf{C}(k+N-1) \end{bmatrix} \boldsymbol{\theta} + \begin{bmatrix} v(k) \\ v(k+1) \\ \vdots \\ v(k+N-1) \end{bmatrix} \tag{1.4}$$

Equation (*1.4*) is suitable for use in a batch least-squares estimation procedure for determining approximate values for the unknown constant parameters $\boldsymbol{\theta}$. Both batch and recursive least-squares solutions are presented in Chapter 6.

The performance of such a parameter estimation scheme is dependent upon the input signal. Sometimes, specially selected input signals can be used during the identification process. In other situations only the normal operating signals can be used. It may be tolerable to add a small sinusoidal component, called a *dither signal,* to the input to aid the identification process. It is intuitively clear that the input must excite those modes of the system that are intended for identification. That is, if a constant input is used and if the system has been operating sufficiently long for steady state to be reached, little about the system can be identified, other than its steady-state gain. The input must be "sufficiently exciting" or "sufficiently rich" if the identification is to be successful.

It is informative to take the Z-transform of Eq. (*1.1*). Z-transforms are defined briefly and used in Problems 1.15 through 1.20. For present purposes it suffices to view the variable z^{-1} as a time-delay operator. Then Eq. (*1.1*) can be written in delay operator—i.e., transformed—form as

$$y(z)[z - a_0 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_n z^{-n}] = [b_0 z + b_1 + b_2 z^{-1} + \cdots + b_p z^{-(p-1)}]u(z)$$

Then the input-output transfer function can be written

$$\frac{y(z)}{u(z)} = \frac{b_0 + b_1 z^{-1} + \cdots + b_p z^{-p}}{1 - (a_0 z^{-1} + \cdots + a_n z^{-(n+1)})} = H(z) \tag{1.5}$$

Some commonly used nomenclature [4] is now defined. If $y(k+1)$ depends only on the u terms and not on past y terms, all the a_i coefficients would be zero and the transfer function then would have zeros, but all p poles would be at the origin—i.e., a pure time delay of p units. This is sometimes referred to as an all-zero model, or a *moving average* (MA) model. If the only input is the random term $v(k)$ (or perhaps a single $u(j)$ term), there is at most one non-zero b_i term. This transfer function has poles, but all zeros, if any, are at the origin. It is called an all-pole model, or alternatively, an *autoregressive* (AR) model. The general case involves both poles and zeros and is often referred to as an *autogregressive moving average* (ARMA) model.

In a multiple-input, multiple-output system, $H(z)$ is, of course, a transfer function *matrix*. Scalar transfer functions are used in Chapters 2 and 3. The complete treatment of state variable/transfer function relationships, including the matrix case, begins in Chapter 3 and continues in Chapter 12.

Alternate Model Forms

Matrix fraction description [5]. Let $\mathbf{P}$, $\mathbf{N}$, and $\mathbf{R}$ be finite matrix polynomials in the variable z^{-1}, which for present purposes can be treated as a delay operator. Then the ARMA-type models can be simply expressed as

$$\mathbf{P}(z^{-1})\mathbf{y}(k) = \mathbf{N}(z^{-1})\mathbf{u}(k) + \mathbf{R}(z^{-1})\mathbf{v}(k) \tag{1.6}$$

Of course, with just one input and one output, $\mathbf{P}$, $\mathbf{N}$, and $\mathbf{R}$ are scalar polynomials, and division by the denominator P puts Eq. (*1.6*) into the transfer function form. In the multivariable case these terms are matrices, as is the transfer function $\mathbf{H}(z)$. $\mathbf{P}$ will always be square. Taking its inverse gives the so-called left MFD (matrix fraction description) form for the transfer function,

$$\mathbf{H}(z) = \mathbf{P}^{-1}(z)\mathbf{N}(z)$$

This and the alternative right MFD are discussed in future chapters.

State variables. It will be shown in Chapter 3 that the preceding discrete-time models can be written in state variable form as

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(k) + \mathbf{B}(\boldsymbol{\theta})u(k) \\ y(k+1) &= \mathbf{C}(\boldsymbol{\theta})\mathbf{x}(k+1) + \mathbf{D}(\boldsymbol{\theta})u(k+1) + v(k+1) \end{aligned} \tag{1.7}$$

where

k denotes a discrete time point
$\mathbf{x}$ is the state
u is the input, deterministic or random
y is the output
v is a random noise
$\boldsymbol{\theta}$ is a vector of unknown parameters
$\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are system matrices

Most of the rest of this book will deal with state variable models, under the assumption that the values of $\boldsymbol{\theta}$ have been identified already, and hence $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ will be

assumed known. In some adaptive and self-tuning control systems, the least-squares estimation of model parameters, $\boldsymbol{\theta}$, is carried out in real time. This estimation process constitutes an outer loop. The inner control loops then use the estimated models to carry out control functions. Adaptive control is not treated in this book [6, 7].

Impulse response [4, 8]. Let the inverse Z-transform of $H(z)$ be $h(k)$; then the inverse transform of $y(z)$ can be written as a summation convolution:

$$\begin{aligned} y(k) &= \Sigma h(k-j)u(j) + v(k) \\ &= h(0)u(k) + h(1)u(k-1) + h(2)u(k-2) + h(3)u(k-3) \\ &\quad + \cdots + v(k) \end{aligned} \tag{1.8}$$

Normally, as the time parameter k continues to increase, the preceding sum continues to grow in length. Then the system description is called an infinite impulse response (IIR). The coefficients $h(k)$ are values of the impulse response, and they could also be calculated, at least in the scalar case, by long division of the transfer function. That is, $h(i)$ would be the coefficient of z^{-i} when the transfer function is written as a power series in z^{-1}. For a stable system, $h(i) \to 0$ as $i \to \infty$. This fact means that the power series might be truncated after some finite number of terms. A system model with only a finite number of past input terms is called a finite input response (FIR).

EXAMPLE 1.7 A specific second-order example of Eq. (*1.1*) is

$$y(k+1) = 1.3y(k) - 0.4\,y(k-1) + u(k+1) \tag{1.9}$$

The input-output transfer function relation is

$$y(z) = u(z)/[(1 - 0.5z^{-1})(1 - 0.8z^{-1})] = H(z)u(z) \tag{1.10}$$

The impulse response $h(kT)$ is given by the inverse Z-transform

$$h(kT) = Z^{-1}\{H(z)\}$$

By using partial fraction expansion,

$$H(z) = \frac{\frac{8}{3}z}{z - 0.8} - \frac{\frac{5}{3}z}{z - 0.5}$$

Then $h(kT) = \frac{8}{3}(0.8)^k - \frac{5}{3}(0.5)^k$ for any sample time k. A partial tabulation of this impulse response function follows:

k	$h(kT)$
0	1
1	1.3
2	1.29
3	1.157
4	0.9881
5	0.8217
10	0.2847
20	0.0512
30	0.0055

These $h(kT)$ values could also be evaluated as coefficients of z^{-1} in the infinite series obtained by long division of

$$H(z) = \frac{1}{1 - 1.3z^{-1} + 0.4z^{-2}}$$

Suppose that $h(kT)$ is approximated as zero for $k > 30$. Then an approximate expression for generating the outputs is

$$y(k) = u(k) + 1.3u(k-1) + 1.29u(k-2) + 1.157u(k-3) + \cdots + 0.0055u(k-30) \tag{1.11}$$

The original form of Eq. (*1.10*) (a second-order autoregressive model) has two poles. Equation (*1.11*), a moving-average model, appears to have no nonzero poles but 30 zeros. These types of approximate equivalencies illustrate the difficulty in experimental modeling. The same sequence of measured inputs and outputs could lead to either of these results, or others, depending on the model structure which is assumed. The nonuniqueness of the answer may or may not cause problems, depending on the purpose of the derived model. ■

1.4 CLASSIFICATION OF SYSTEMS

As a result of the modeling discussion of Section 1.3, it can be seen that the types of equations required to describe a system depend on the types of elemental equations and the types of inputs from the environment. System models are classified according to the types of equations used to describe them. The family tree shown in Figure 1.8 illustrates the major system classifications. Combinations of these classes can also occur. The most significant combination is the continuous-time system, digital controller of the type mentioned in Example 1.4. The digital signals are discrete-time in nature, that is, they only change at discrete time points. The most common approach to these problems is to represent the continuous-time part of the system by a discrete-time approximate model and then proceed with a totally discrete problem. The experimentally derived ARMA models of Section 1.3.2 are approximations of this type. Other discrete approximations are given later (Problem 2.18 and Section 9.8).

In Figure 1.8 dashed lines indicate the existence of subdivisions similar to the others shown on the same level.

Distributed parameter systems require partial differential equations [9] for their description, for example, as in the description of currents and voltages at every spatial point along a transmission line. These will not be considered further, but can often be approximated by lumped-parameter models. Lumped-parameter systems are those for which all energy storage or dissipation can be lumped into a finite number of discrete spatial locations. They are described by ordinary difference equations, or in some cases by purely algebraic equations. Discrete component electric circuits fall into this category.

Systems containing parameters or signals (including inputs) which can only be described in a probabilistic fashion (due to ignorance or actual random behavior) are called stochastic, or random, systems. Because random process theory [10] is not an assumed prerequisite for this text, emphasis will be on deterministic (nonrandom)

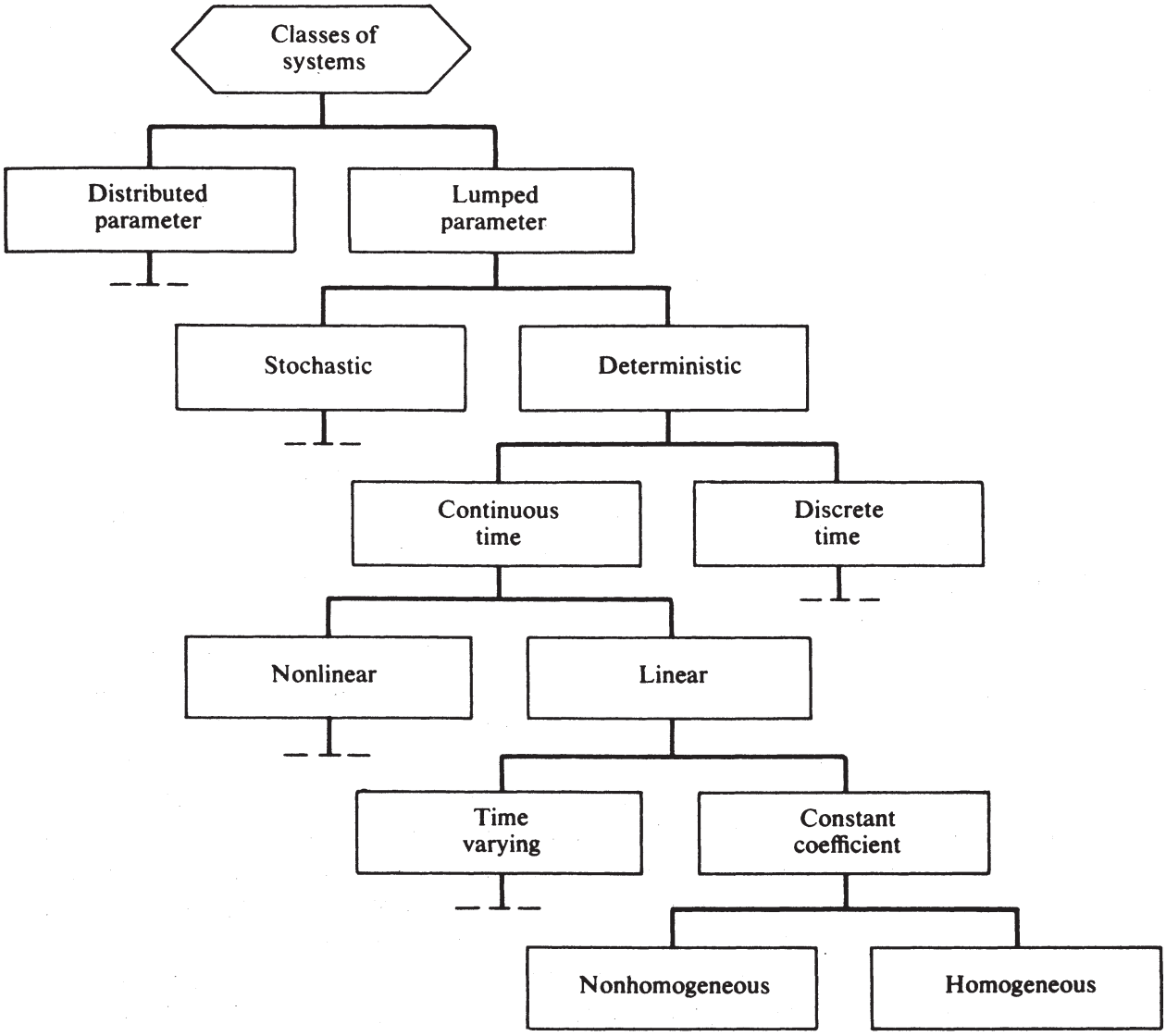


Figure 1.8 Major classes of system equations.

lumped-parameter systems. There will be a few occasions, such as in the discussion of noisy measurements, where the random nature of certain error signals cannot be totally ignored.

If all elemental equations are defined for all time, then the system is a continuous-time system. If, as in sampling or digital systems, some elemental equations are defined or used only at discrete points in time, a discrete-time system is the result. Continuous-time systems are described by differential equations, discrete-time systems by difference equations.

If all elemental equations are linear, so is the system. If one or more elemental equations are nonlinear, as is the case for a diode, then the overall system is nonlinear. When all elemental equations can be described by a set of constant parameter values, as in the familiar *RLC* circuit, the system is said to be stationary or time-invariant or constant coefficient. If one or more parameters, or the very form of an elemental equation, vary in a known fashion with time, the system is said to be time-varying.

Finally, if there are no external inputs and the system behavior is determined entirely by its initial conditions, the system is said to be homogeneous or unforced. With forcing functions acting, a nonhomogeneous system must be considered.

One additional distinction not shown in Figure 1.8 could be made between large-scale (many variables) systems and small-scale systems. The degree of difficulty in analysis varies greatly among these system classifications. These differences have motivated different methods of approach. Modern control theory provides one of the most general approaches.

1.5 MATHEMATICAL REPRESENTATIONS OF SYSTEMS

During the analytical modeling process, equations are developed to describe the behavior of each individual system element and also to describe the interconnections of these elements. These equations, or for that matter the corresponding linear graph, could be taken as the mathematical representation of the system. Normally, however, additional manipulations will be performed before the mathematical representation of the model is in final form.

Many forms are possible, but generally they divide into one of two categories: (*a*) input-output equations, and (*b*) equations which reveal the internal behavior of the system as well as input-output terminal characteristics.

Input-output equations are derived by a process of elimination of all system variables except those constituting inputs and those considered as outputs. For the system shown in Figure 1.5*a,* this would mean expressing signals S_2, S_3, S_4, and S_5 in terms of the input signals S_1 and output signals S_6. This is done using the known dynamic relations of the subsystems *A, B,* and *C.* For example, S_2 and S_4 are related to S_1 through the dynamics of subsystem *A.* The input-output equations could constitute one or more differential or difference equations in any of the classes shown in Figure 1.8. The independent variable is usually time, the dependent variables are the system outputs, and the inputs act as forcing functions. Models developed experimentally from measurements of inputs and outputs are almost invariably of the input-output type. Subsequent conversions to other forms, such as state variable models, are always possible.

When the constituent equations are linear with constant coefficients, Laplace or Z-transforms can be used to define input-output *transfer functions.* When more than one input and more than one output must be treated, matrix notation and the concepts of *transfer matrices* are convenient. It is assumed that the Laplace transform is a tool familiar to most readers. The Z-transform [8] may be less familiar, and so a bare-bones minimum introduction to it is contained in the problems. Although time-domain methods will be stressed in this book, transforms and transfer functions will at times be useful. A reader with no prior exposure should also consult the references.

Integral forms of the input-output equations, using the system *weighting function,* are also widely used. The weighting function, or system impulse response, is obtained from the inverse Laplace or Z-transform of the input-output transfer function.

Equations which reveal the internal behavior of the system, and not just the terminal characteristics, can take several forms. Loop equations, involving all un-

known loop currents (through variables), or node equations, involving all unknown nodal voltages (across variables), can be written. The choice between alternatives may depend on the number of loops versus the number of nodes. If everything is linear, transform techniques lead to the concepts of impedance and admittance matrices. Hybrid combinations of voltage and current equations are also used.

The *state space* approach [1] will be developed extensively in the remainder of this book. At this time it is sufficient to say that state variables consist of some minimum set of variables which are essential for completely describing the internal status, i.e., state of the system.

1.6 MODERN CONTROL THEORY: THE PERSPECTIVE OF THIS BOOK

The most effective control theory makes use of good models of the real-world systems being controlled. A goal of this chapter has been to stress the importance of the modeling process. No attempt was made to present a complete, self-contained theory of modeling. Rather, the intent was to build upon knowledge acquired in prior courses on circuit theory, dynamics, kinematics, and so on. Some background knowledge in introductory feedback control is also assumed. A summary review of this topic is included in Chapter 2.

In some industrial control applications, good results are achieved even though only a rudimentary knowledge of a process model is available. The widely used proportional-integral-derivative (PID) controller can be tuned to give satisfactory performance based solely on knowledge of dominant system time constants. This fact does not violate the dictum that good models are required; it simply reinforces the point that models should be suited to the intended purpose.

In other areas of control practice, systems are successfully designed and built without use of mathematical models or analysis. This artisan, or craftsperson, approach can often work reasonably well when the designer is sufficiently experienced and the system is only incrementally different from previous successful designs. However, when problems do arise, the same ad hoc approach to an attempted solution can sometimes compound the difficulties. One such example occurred with an industrial robotic system. The system exhibited unacceptable signal fluctuations on occasion, which were erroneously diagnosed as a noise problem. Compensating capacitors were added between key signal lines and ground to allow the high-frequency noise terms to bleed off. An after-the-fact model revealed the true problem was caused by very low stability margins. The "compensating" capacitors added additional phase lag and only made the problem worse.

Elegant mathematical results derived from an inappropriate model can likewise yield negative results. Exact models of real systems are extremely rare. Therefore, robustness is an important quality in control design. *Robustness* can be defined in various ways, but generally the word implies the maintenance of adequate stability margins or other performance levels in spite of model errors or deliberate oversimplifications. The ability to operate in the presence of disturbance inputs is also important. The widely used PID controllers owe much of the success to what today would be called their robustness. The linear-quadratic (LQ) optimal controllers of Chapter 14 have certain guaranteed robustness properties. The newer H^∞ design

techniques, which are basically worst-case analyses, represent another approach to robustness in the face of model error. These are not pursued in this book [11].

Many real systems are nonlinear and/or time-varying. Yet, approximately 80% of this book is devoted to linear, constant systems. Perhaps 10% is devoted explicitly to linear time-varying systems and another 10% (primarily Chapter 15) is devoted to nonlinear systems. The purpose of this book is to build a foundation for specialized study that may follow, and linear systems theory is the major part of that foundation. The treatment of nonlinear systems in Chapter 15 is restricted mainly to extensions of the linear theory that follow easily from earlier developments in this book. Several useful approaches to the control of some classes of nonlinear systems are presented.

Chapters 4 through 8 present a large amount of linear algebra for controls rather than control theory per se. An attempt has been made to motivate the linear algebraic developments by bringing in related control topics, even though the same topics may be developed more fully later. Some algorithmic considerations are also included. Liberal use of computer algorithms has been made throughout the book. One advantage of the state variable approach to control problems is that the structure remains the same whether there are 2, 20, or 100 states. The fact is, however, that only the smallest problems can be solved without a computer. Many solutions provided in this book were carried out using code acquired from the literature [12], perhaps with modifications, or with programs developed during the years of teaching this material. There are now several commercially available packages which have the needed capabilities [13, 14].

A large number of diverse problems and examples are included in each chapter. The intention is to show not only how a given problem is worked but why it is worked in a certain way and what the ramifications are. For example, knowing how to compute the feedback gains to achieve certain closed-loop poles is a mathematical result. Additional engineering insight is needed in order to decide whether a pole placement approach should be used and, if so, what constitutes good pole locations. Alternatives are classical feedback design or an optimal control design. In these, intelligent trade-offs must be made between response time, control effort, or disturbance rejection. The problems are intended to give some insight into these issues, in addition to illustrating the mechanics of a given method.

Closely related technical areas include self-tuning and adaptive control [7, 8], learning systems and artificial intelligence [7, 15], neural networks [16], and robotics [17]. A short chapter or two at the end of the present book could not do justice to any of these important topics. Therefore, this book concentrates on developing a basic foundation which would be useful to the broadest class of readers. Those wishing to pursue one of these special topics later will be able to do so more effectively after mastering the material given here.

REFERENCES

1. Zadeh, L. A. and C. A. Desoer: *Linear System Theory, The State Space Approach,* McGraw-Hill, New York, 1963.
2. Cannon, R. H., Jr.: *Dynamics of Physical Systems,* McGraw-Hill, New York, 1967.

3. Shearer, J. A., A. T. Murphy, and H. H. Richardson: *Introduction to System Dynamics,* Addison-Wesley, Reading, Mass., 1967.
4. Makhoul, J.: "Linear Prediction: A Tutorial Review," *Proceedings of the IEEE,* Vol. 63, No. 4, April 1975.
5. Kailath, T.: *Linear Systems,* Prentice-Hall, Englewood Cliffs, N.J., 1980.
6. Harris, C. J. and S. A. Billings: *Self-Tuning and Adaptive Control,* Peter Peregrinus Ltd. (for IEE), London, 1981.
7. Astrom, K. J. and B. Wittenmark: *Adaptive Control,* Addison-Wesley, Reading, Mass., 1989.
8. Franklin, G. F. and J. D. Powell: *Digital Control of Dynamic Systems,* Addison-Wesley, Reading, Mass., 1980.
9. Brogan, W. L.: "Optimal Control Applied to Systems Described by Partial Differential Equations," *Advances in Control Systems,* Vol. 6, C. T. Leondes, Ed., Academic Press, New York, 1968.
10. Papoulis, A.: *Probability, Random Variables and Stochastic Processes,* McGraw-Hill, New York, 1965.
11. Zames, G.: "Feedback and Optimal Sensitivity: Model Reference Transformations, Multiplicative Seminorms and Approximate Inverses," *IEEE Transactions on Automatic Control,* Vol. AC-26, No. 2, April 1981, pp. 301–320.
12. Melsa, J. L. and S. K. Jones: *Computer Programs for Computational Assistance in the Study of Linear Control Theory,* 2d ed., McGraw-Hill, New York, 1973.
13. Herget, C. J. and A. J. Laub, Eds.: "Computer Aided Design of Control Systems Special Issue," *IEEE Control Systems Magazine,* Vol. 2, No. 4, Dec. 1982.
14. Moler, C. "MATLAB Users' Guide," *Tech. Report CS81- 1* (Revised) Dept. Of Computer Science, University of New Mexico, Albuquerque, N.M., Aug. 1982.
15. Grossberg, S.: *The Adaptive Brain I: Cognition, Learning, Reinforcement, and Rhythm,* and *The Adaptive Brain II: Vision Speech, Language and Motor Control,* Elsevier/North Holland, Amsterdam, 1986.
16. Bavarian, B., Ed.: "Special Section on Neural Networks for Systems and Control" (five articles), *IEEE Control Systems Magazine,* Vol. 8, No. 2, Apr. 1988, pp. 3–31. (See also Vol. 9, No. 3, Apr. 1989, pp. 25–59 for five more articles on neural networks in controls.)
17. Klafter, R. D., T. A. Chmielewski, and M. Negin: *Robotic Engineering, an Integrated Approach,* Prentice Hall, Englewood Cliffs, N.J., 1989.
18. Chestnut, H.: "A Systems Approach to the Economic Use of Computers for Controlling Systems in Industry," *General Electric Report no. 70-C-089,* Schenectady, N.Y., Feb. 1970.
19. Kuo, B. C.: *Analysis and Synthesis of Sampled-Data Control Systems,* Prentice-Hall, Englewood Cliffs, N.J., 1963.

ILLUSTRATIVE PROBLEMS

1.1 Explain why it would be inappropriate to consider a single branch of an electric network a system, even if the only variable of interest is the current through that branch.

The current in one branch affects the currents and voltages in other parts of the network. This, in turn, affects the current in the first branch. Because of the two-way coupling, the network must be considered as a whole, and must be solved using simultaneous equations.

1.2 Develop an electromechanical model of the fixed field, armature-controlled dc motor. Consider the voltage supplied to the armature as the input and account for the observed dissipation of electrical energy and mechanical energy.

The dissipation of electrical energy can be accounted for by lumping all armature resist-

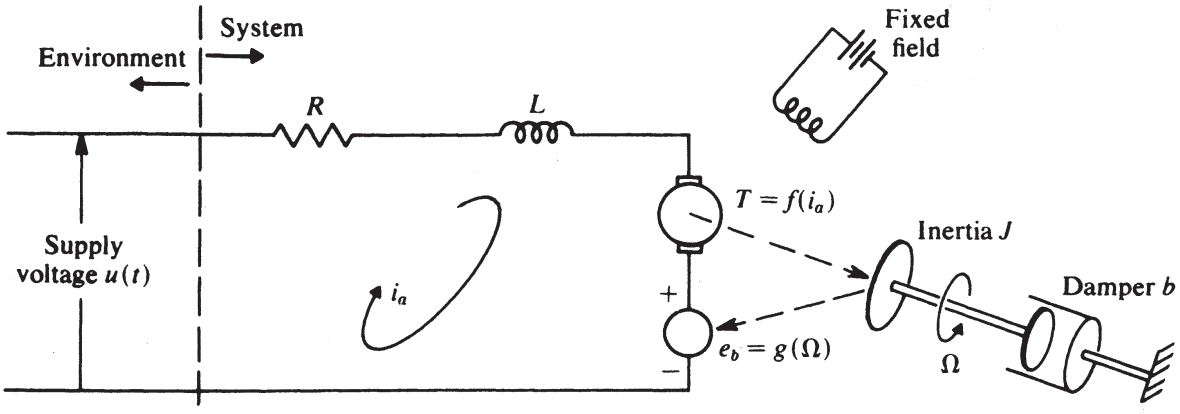


Figure 1.9

ance into a resistor R. A noticeable phase shift between the supply voltage and the current through the armature windings can be accounted for by a lumped inductor L. The load and all rotating parts can be represented as a lumped inertia element J. Mechanical energy losses are accounted for by adding an ideal damper b between the rotating load and some fixed reference. The connection between the electrical and the mechanical aspects is obtained from Maxwell's equations. A moving charge in a magnetic field has a force exerted upon it, so that the armature torque T is a function of the armature current i_a. Likewise, a conductor moving in a magnetic field has a voltage induced in it, the back emf e_b. The model is shown schematically in Figure 1.9.

The torque and the back emf are often approximated by the linear relationships $T = Ki_a$ and $e_b = K\Omega$, where K is a constant for a particular motor. For the linear case the transfer function between the input voltage $u(t)$ and the output angle $y(t) = \int \Omega dt$ is derived as follows. The loop equation for the electrical circuit is $u(t) = L(di_a/dt) + Ri_a + e_b$. The mechanical torque balance is $T(t) = J\ddot{y} + b\dot{y}$. The Laplace transforms of these equations are $u(s) = (Ls + R)i_a(s) + e_b(s)$ and $T(s) = (Js^2 + bs)y(s)$. Solving gives $i_a(s) = [u(s) - e_b(s)]/(R + Ls)$. The electromechanical conversion equation then gives $T(s) = K[u(s) - e_b(s)]/(R + Ls)$. Equating the two forms for $T(s)$ and using $e_b(s) = Ksy(s)$ gives $(Js^2 + bs)y(s) = K[u(s) - Ksy(s)]/(R + Ls)$. The input-output transfer function is found from this:

$$y(s)/u(s) = K/[(Js^2 + bs)(R + Ls) + K^2 s]$$

1.3 Some electronic test gear (Figure 1.10) is mounted near a large tank of liquid gas at −350°F. Develop a simple model which would be useful in estimating the coldest temperature at which the electronic equipment will need to operate.

Because of the insulation material, heat is allowed to flow only in one direction, from the 70° air through the electronic package to the −350° liquid gas. The environment, consisting of the two constant temperatures of 70 and −350, is represented by two ideal sources. There is a single unknown temperature T (across variable), that of the electronic package interior. The

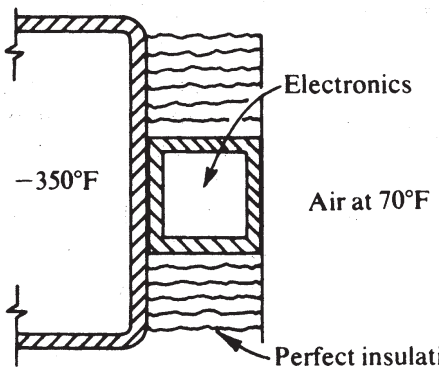


Figure 1.10

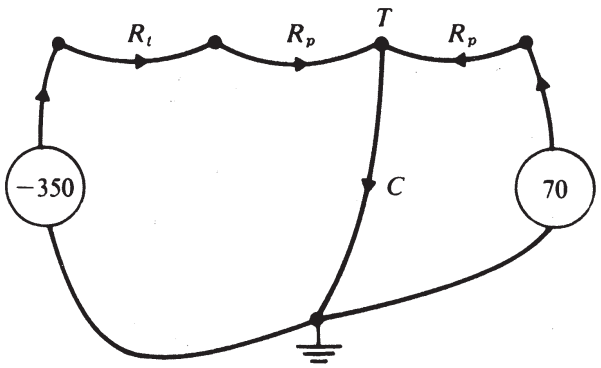


Figure 1.11

package has some thermal capacitance C and its end walls and the tank wall present thermal resistance, R_p and R_t, to heat flow Q. The linear graph is shown in Figure 1.11.

In steady state no heat flows in the branch representing the capacitance. Thus $70 - T = QR_p$ and $T - (-350) = Q(R_p + R_t)$. Eliminating the heat flow Q gives

$$\frac{70 - T}{R_p} = \frac{T + 350}{R_p + R_t} \quad \text{or} \quad T = \frac{70(R_p + R_t) - 350R_p}{2R_p + R_t}$$

If the thermal resistivity R_p of each end of the electronic package is 1°F s/Btu and if the thermal resistivity R_t of the adjacent area of the tank is 2°F s/Btu, then $T = [70(3) - 350(1)]/4 = -35$°F.

1.4 In many ways the flow of work through a factory is similar to fluid flow in a piping network. Figure 1.12 shows such a network. Two pumps deliver fluid at constant pressure P_1 and P_2, respectively. Six lumped approximations for fluid resistance R_i are indicated. They account for the pressure drop in each segment of pipe proportional to the flow Q through the segment. Three fluid capacitances C_i are indicated. The pressure at their base is proportional to the height of the standing fluid, that is, proportional to the integral of the flow into them. Two ideal elements, called fluid inertances I_1 and I_2, are included to account for inertia effects. They cause a pressure drop proportional to the rate of change of flow. a. Draw the linear graph, label all variables; b. write the elemental equations; c. write the continuity equations; d. write the compatibility equations. Neglect all changes in height except in the capacitances, and use atmospheric pressure as the reference node.

(a) Each distinct pressure (across variable) will form a system node. Between each pair of nodes a branch will represent the ideal element that accounts for the pressure change. This allows the construction of the linear graph of Figure 1.13.

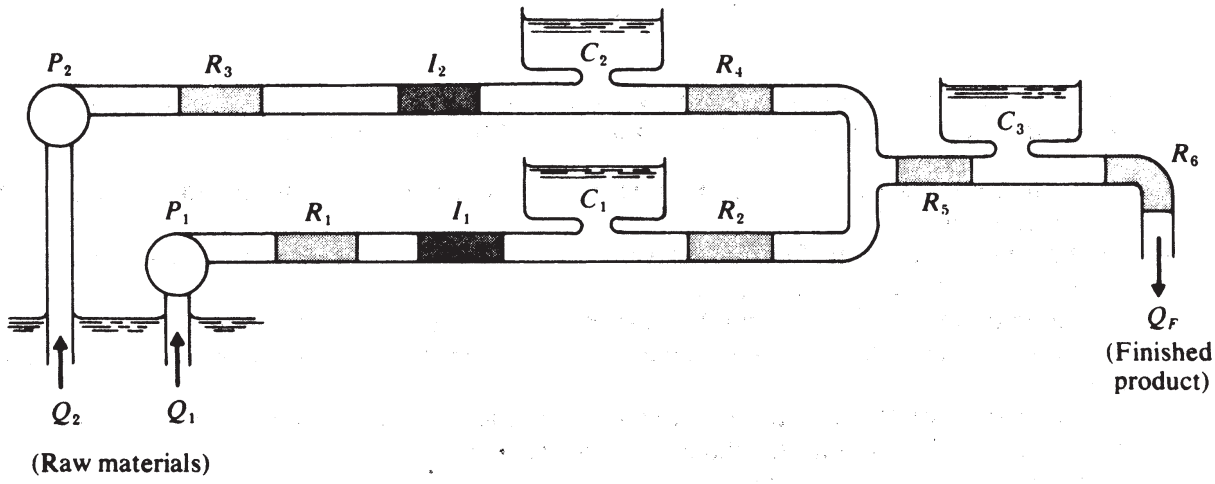


Figure 1.12

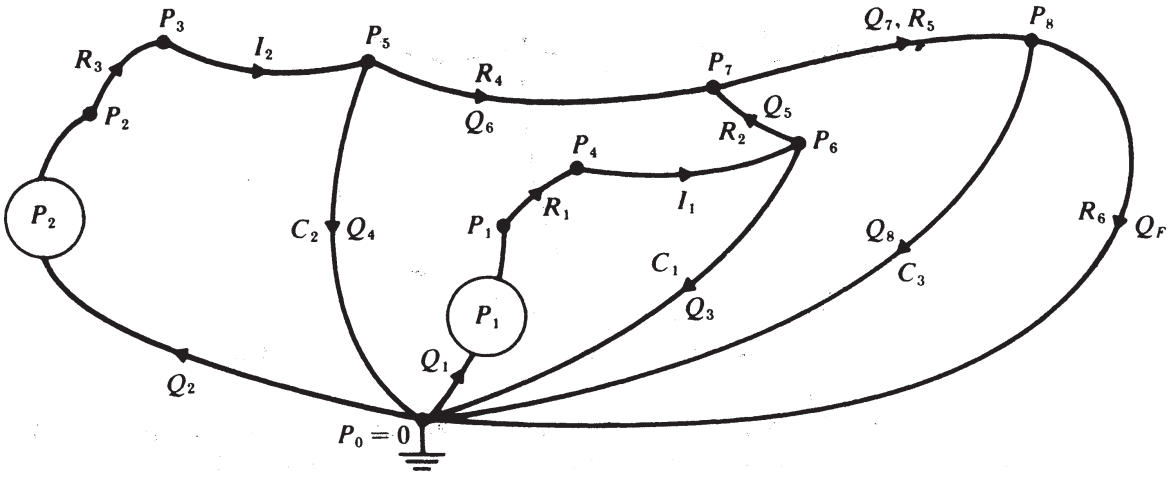


Figure 1.13

(b) There are $n = 9$ nodes, $b = 13$ branches, and $s = 2$ sources. The $b - s = 11$ elemental equations are

$$P_{23} = R_3 Q_2 \qquad P_{14} = R_1 Q_1 \qquad P_{46} = I_1 \frac{dQ_1}{dt} \qquad P_{45} = I_2 \frac{dQ_2}{dt}$$

$$C_2 \frac{dP_{50}}{dt} = Q_4 \qquad C_1 \frac{dP_{60}}{dt} = Q_3 \qquad P_{57} = R_4 Q_6 \qquad P_{67} = R_2 Q_5$$

$$P_{78} = R_5 Q_7 \qquad C_3 \frac{dP_{80}}{dt} = Q_8 \qquad P_{80} = R_6 Q_F$$

(c) The $n - 1 = 8$ continuity equations are

$Q_1 = Q_1$ (twice) since Q_1 flows in three separate branches

$Q_2 = Q_2$ (twice) since Q_2 flows in three separate branches

$$Q_1 - Q_3 - Q_5 = 0 \qquad Q_2 - Q_4 - Q_6 = 0 \qquad Q_5 + Q_6 - Q_7 = 0$$

$$Q_7 - Q_8 - Q_F = 0$$

(d) The $b - (n - 1) = 5$ compatibility equations are

$$P_1 = P_{14} + P_{46} + P_{60} \qquad P_2 = P_{23} + P_{35} + P_{50}$$

$$P_{50} = P_{57} + P_{76} + P_{60} \qquad P_{60} = P_{67} + P_{78} + P_{80}$$

$$P_{08} + P_{80} = 0$$

By eliminating variables in various ways, a set of differential equations involving only flow rates Q_i, or only nodal pressures P_i, or a combination of both could be obtained.

1.5 **(a)** Write equations describing the lumped-parameter approximate model for the transmission line shown in Figure 1.14.

(b) Find the input-output transfer function $y(s)/u(s)$. The input $u(t)$ is the source voltage v_s, and the output $y(t)$ is the load voltage v_L.

(a) For simplicity, the line is segmented into three equal lengths as shown. The leakage conductance G is the reciprocal of the leakage resistance. More segments could be used in the same manner.

The values of R and L are obviously 1/3 that for the entire line, while G and C each have values equal to 1/2 that for the entire line. Since the source voltage is known, there are six unknowns: i_0, i_1, i_2, v_1, v_2, and v_L.

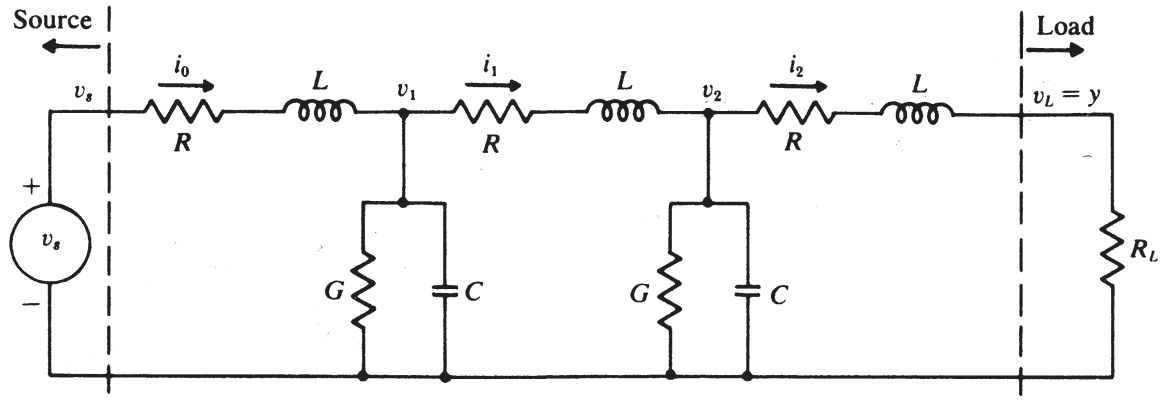


Figure 1.14

(b) Writing loop equations, using the Laplace transforms of the elemental equations, gives

$$v_s = [R + Ls]i_0 + v_1 \qquad v_1 = [R + Ls]i_1 + v_2$$
$$v_2 = [R + Ls]i_2 + v_L \qquad v_L = R_L i_2$$

The nodal equations are $i_0 = (G + Cs)v_1 + i_1$ and $i_1 = (G + Cs)v_2 + i_2$. Letting $A = R + Ls$ and $B = G + Cs$ gives

$$v_s = Ai_0 + v_1 = (AB + 1)v_1 + Ai_1 = (AB + 1)v_2 + (A^2B + 2A)i_1$$

By continuing this process of substitution, a final expression containing only $v_s = u$ and $v_L = y$ is obtained:

$$\frac{y(s)}{u(s)} = \frac{R_L}{(A^3B^2 + 4A^2B + 3A) + R_L(A^2B + 3AB + 1)}$$

1.6 Derive a difference equation for the purely resistive ladder network shown in Figure 1.15 (perhaps a dc version of the lumped approximation for a transmission line).

The difference equation for a typical $(k + 1)$st loop is obtained by writing a loop equation

$$(2r + R)i_{k+1} - ri_k - ri_{k+2} = 0$$

This holds for $1 \leq k + 1 \leq N - 1$. It is a second-order difference equation, and two boundary conditions are needed in order to uniquely specify the solution. The first and the last loops, which do not satisfy the general equation, provide the two necessary conditions:

$$v_s = (R + r)i_0 - ri_1$$
$$0 = (r + R + R_L)i_N - ri_{N-1}$$

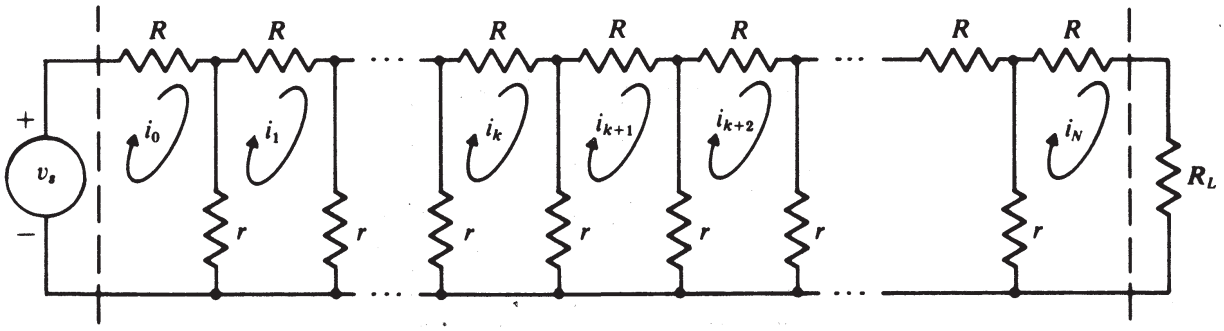


Figure 1.15

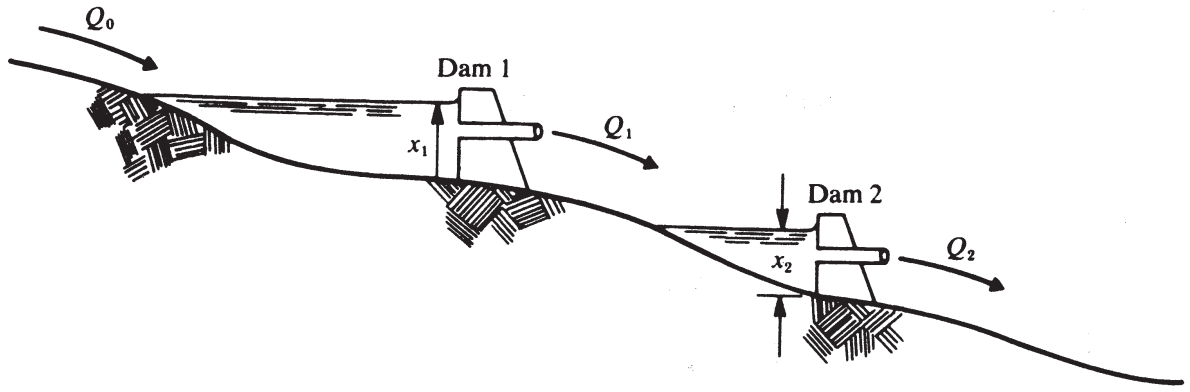


Figure 1.16

1.7 A pair of dams in a flood-control project is shown in Figure 1.16. The water level at dam 1 at a given time t_k is $x_1(k)$, and $x_2(k)$ is the height at dam 2 at the same time. The amount of run-off water collected in reservoir 1 between times t_k and t_{k+1} is $Q_0(k)$. The water released from dams 1 and 2 during this period is denoted by $Q_1(k)$ and $Q_2(k)$. Develop a discrete-time model for this system.

Conservation of flow requires

$$x_1(k+1) = x_1(k) + \alpha[Q_0(k) - Q_1(k)]$$

$$x_2(k+1) = x_2(k) + \beta[Q_1(k) - Q_2(k)]$$

These represent lumped-parameter discrete-time equations. If the amount of controlled spillages Q_1 and Q_2 are selected as functions of the water heights x_1 and x_2, a discrete feedback control system obviously results. Z-transform theory could be used to analyze such a system.

1.8 Draw the linear graph for the ideal transformer circuit of Figure 1.17 noting that the transformer is a four-terminal element.

The linear graph is shown in Figure 1.18*a* with the transformer represented as shown in Figure 1.18*b*.

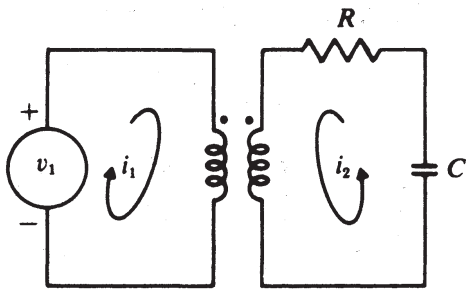
The equations for this circuit are

Node equations: $i_s = i_1, \quad i_3 = -i_2, \quad i_3 = i_4$

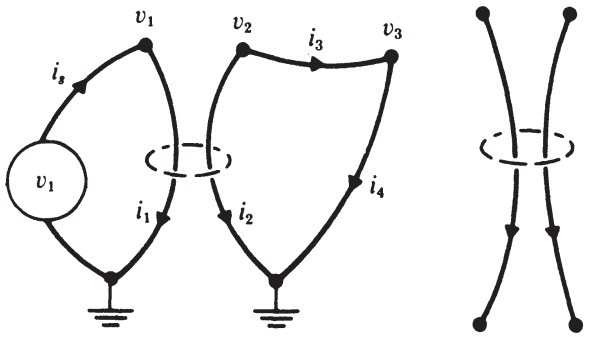
Elemental equations: $v_2 = Nv_1, \quad i_1 = -Ni_2$ (transformer)

$v_2 - v_3 = i_3 R$ (resistance)

$C\dot{v}_3 = i_4$ (capacitance)



Turns ratio = N **Figure 1.17**



(a) (b) **Figure 1.18**

Note that the transformer requires two equations for its specification. Similar multiterminal elements are required whenever energy is transformed from one form to another. Transformers, transducers, and gyrators all have similar representation. Note also that an ideal transformer has zero instantaneous power flow into it, i.e.,

$$v_1 i_1 + v_2 i_2 = v_1(-Ni_2) + (Nv_1)i_2 = 0$$

1.9 Develop a model of an automobile which would be appropriate for studying the effectiveness of the suspension system, tire characteristics, and seat design on passenger comfort.

For simplicity, lateral rolling motions are ignored. An idealized model might be represented as shown in Figure 1.19. The displacements x_1 and x_2 are inputs from the environment (road surface). Masses m_1 and m_2 represent the wheels, whereas M and J represent the mass and pitching inertia of the main car body. The seat and passenger mass are represented by m_p. The elasticity and energy dissipation properties of the tires are represented by k_1, k_2, b_1, and b_2. The suspension system is represented by k_3, k_4, b_3, and b_4. The seat characteristics are represented by k_s. Newton's second law is applied to the wheels, giving

$$m_1\ddot{x}_3 = k_1(x_1 - x_3) + b_1(\dot{x}_1 - \dot{x}_3) + k_3(x_5 - x_3) + b_3(\dot{x}_5 - \dot{x}_3)$$

$$m_2\ddot{x}_4 = k_2(x_2 - x_4) + b_2(\dot{x}_2 - \dot{x}_4) + k_4(x_6 - x_4) + b_4(\dot{x}_6 - \dot{x}_4)$$

Letting l_1 be the distance from the left end to the center of gravity cg and letting l_2 be the distance to the seat mount, the following geometric relations can be obtained. Assuming small angles,

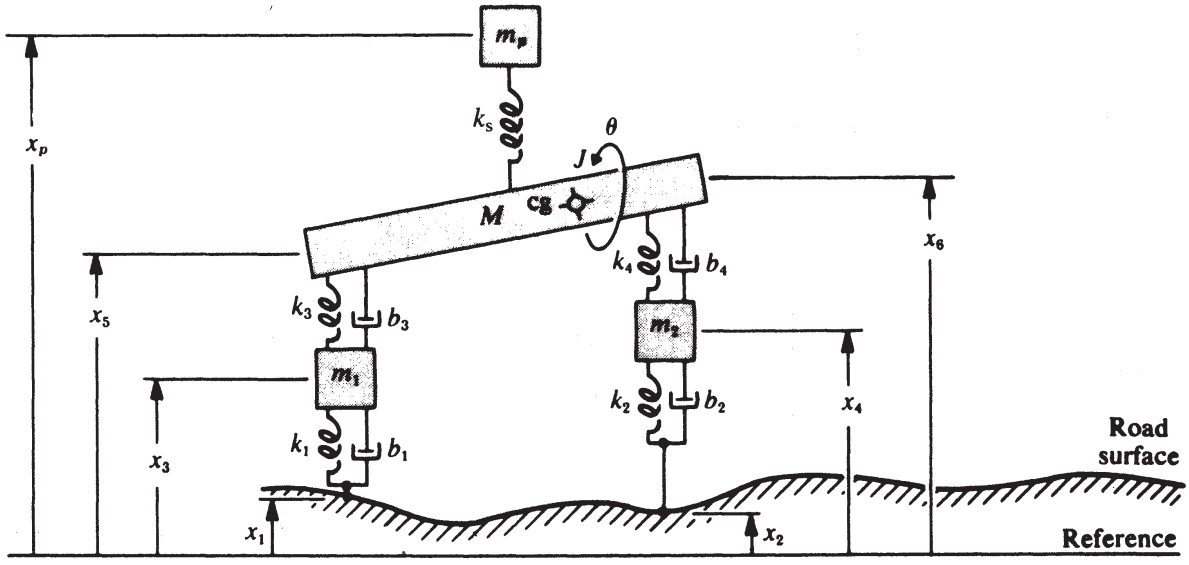


Figure 1.19

$$x_{cg} = x_5 + \frac{l_1}{l}(x_6 - x_5)$$

$$x_s = x_5 + \frac{l_2}{l}(x_6 - x_5) \quad \text{and} \quad \theta = \frac{x_6 - x_5}{l}$$

where l is the total length (wheel base). Summing forces on M gives

$$M\ddot{x}_{cg} = k_3(x_3 - x_5) + k_4(x_4 - x_6) + k_s(x_p - x_s) + b_3(\dot{x}_3 - \dot{x}_5) + b_4(\dot{x}_4 - \dot{x}_6)$$

Summing torques gives

$$J\ddot{\theta} = -l_1 k_3(x_3 - x_5) + (l - l_1)k_4(x_4 - x_6) - (l_1 - l_2)k_s(x_p - x_s)$$
$$-l_1 b_3(\dot{x}_3 - \dot{x}_5) + (l - l_1)b_4(\dot{x}_4 - \dot{x}_6)$$

Finally, summing forces on m_p gives $m_p\ddot{x}_p = k_s(x_s - x_p)$. This set of five coupled second-order differential equations, along with the geometric constraints, constitutes an approximate model for this system.

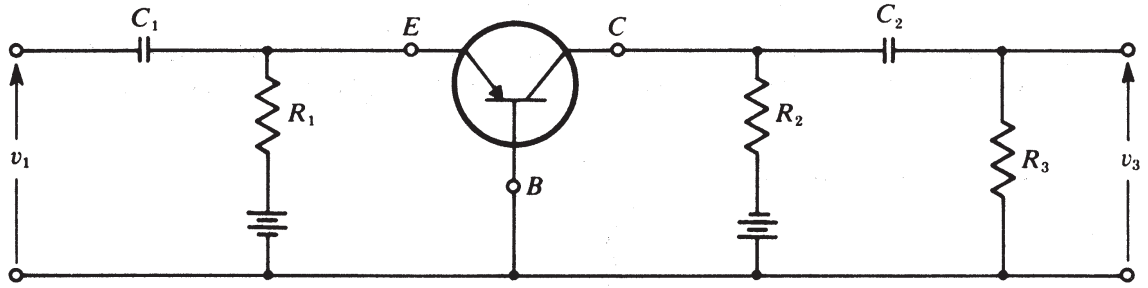
1.10 A typical common base amplifier circuit, using a *pnp* transistor, is shown in Figure 1.20*a*. The h-parameter equivalent circuit for small signals within the amplifier mid-band frequency range is given in Figure 1.20*b*. Draw the linear graph for the amplifier.

The input signal voltage is replaced by an ideal source v_s in series with the source resistance R_s. The three-terminal transistor device is described by the four hybrid parameters h_{ib}, h_{rb}, h_{fb}, and h_{ob}, which are straight line approximations to the various nonlinear device characteristics in the vicinity of the operating point.

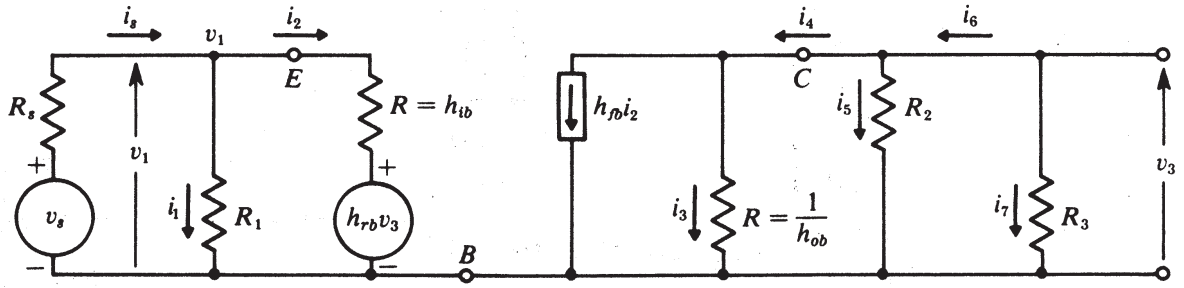
In this example there are two dependent, or controlled, sources, described by $v_d = h_{rb}v_3$ and $i_d = h_{fb}i_2$. Using the equations implied by the linear graph, Figure 1.21,

$$v_1 = i_2 h_{ib} + h_{rb}v_3 \qquad (E\text{-}B \text{ loop equation}) \qquad (1)$$

$$h_{fb}i_2 + i_3 + i_5 + i_7 = 0 \qquad (C \text{ node equation}) \qquad (2)$$



(*a*)



(*b*)

Figure 1.20

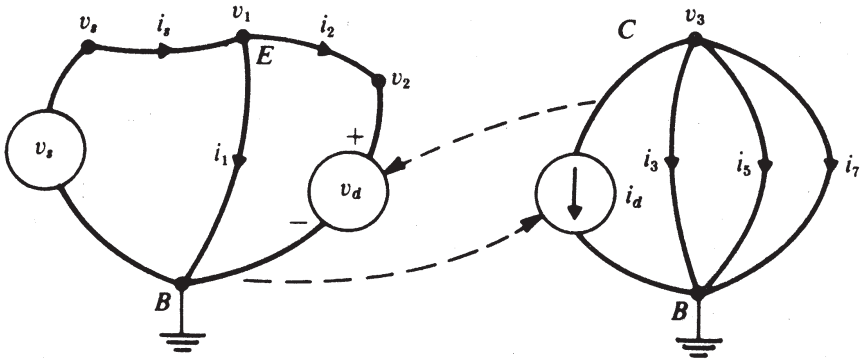


Figure 1.21

or

$$h_{fb}i_2 + v_3 h_{ob} + v_3/R_2 + v_3/R_3 = 0 \qquad \text{(using elemental equations)} \tag{3}$$

From equation (3),

$$i_2 = \frac{-v_3}{h_{fb}}\left(h_{ob} + \frac{1}{R_2} + \frac{1}{R_3}\right) \tag{4}$$

Using equation (4) in equation (1) gives the voltage input-output equation:

$$v_3 = \frac{-h_{fb}v_1}{h_{ib}(h_{ob} + 1/R_2 + 1/R_3) - h_{rb}h_{fb}}$$

1.11 Draw a block diagram for the motor system of Problem 1.2, preserving the individual identity of the electrical, mechanical, and conversion aspects, and illustrating the feedback nature of this system.

The block diagram of Figure 1.22 is drawn directly from the constituent equations of Problem 1.2.

1.12 Classify the systems described by the following equations:

(a) $\ddot{y} + t^2\dot{y} - 6y = u(t)$

(b) $\ddot{y} + \ddot{y}y + 4y = 0$

(c) $\dfrac{\partial^2 y}{\partial t^2} = a\dfrac{\partial^2 y}{\partial x^2}$

(d) $\ddot{y} + a\dot{y} + by^2 = u(t)$

(e) $\dot{y} + ay = u(t)$ if $t < t_1$
$\dot{y} + by = u(t)$ if $t \geq t_1$

(f) $\dot{y} + a \max(0, y) = 0$

(a) Lumped parameter, linear, continuous time, time variable coefficient (t^2), nonhomogeneous.
(b) Lumped parameter, nonlinear (due to $\ddot{y}y$ term), continuous time.
(c) Distributed parameter.
(d) Lumped parameter, nonlinear (due to y^2 term).
(e) Lumped parameter, linear, time variable.
(f) Lumped parameter, nonlinear.

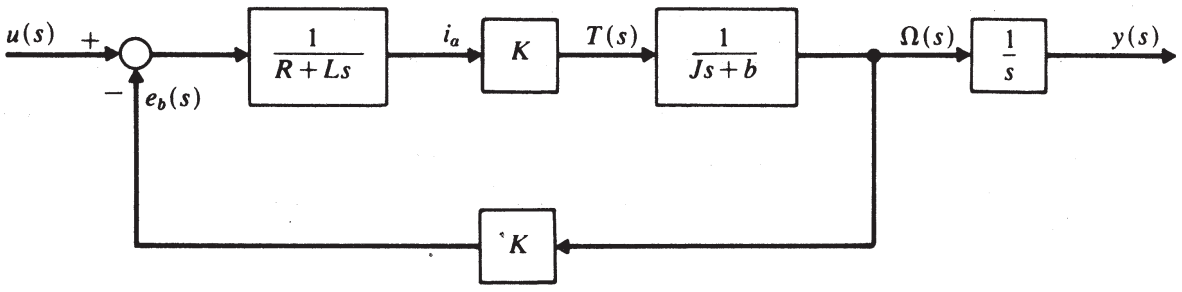


Figure 1.22

Changing of system characteristics, through switching at predetermined times, does not make a system nonlinear. However, if the switching depends on the magnitude of the dependent variable y, the system is nonlinear.

1.13 Use the concepts of Figure 1.5, page 5, to discuss the development of a mathematical model for a rocket vehicle.

1. With such a vague problem statement, many purposes for this model could be considered, such as the structural adequacy of the design, or the temperature history of a component within the vehicle. Suppose that the purpose is to study the trajectory of the vehicle.
2. The boundary of the system is the physical envelope of the vehicle. The inputs from the external environment consist of atmospheric and gravitational effects, as well as a thrust force caused by the gases being expelled across the system boundary. Additional inputs are the mission data, which specify key characteristics that the trajectory should possess. Outputs are the components of position and velocity along the resulting trajectory.
3. The structure of this system consists of several subsystems. One of these is the vehicle dynamics subsystem, which relates forces and torques to the vehicle acceleration. Another is the kinematic subsystem, which relates accelerations to vehicle positions and velocities. A third subsystem is the navigation subsystem, which takes measurements of position, velocity, or acceleration and provides useful signals containing present position and velocity information. A fourth subsystem, the guidance system, accepts the position-velocity data, compares it with mission goals, and computes guidance commands. The final subsystem is the control subsystem. It accepts guidance commands as inputs, and its outputs are commanded body attitude angles or angular rates which will cause the vehicle to steer to the desired trajectory. The control system also turns the thrust on and off. The overall system is shown in Figure 1.23.
4. A wide diversity of models could be developed. If only position and velocity are of interest, the vehicle may be represented as a point mass and the attitude control system might be assumed perfect. If angular attitude information is desired, detailed equations for rotational motion may be required. The description could include elastic vehicle bending, spurious torques due to fuel sloshing, the dynamics of hydraulic control actuators, etc. Each of the other subsystems could be broken down into very fine detail, if required.

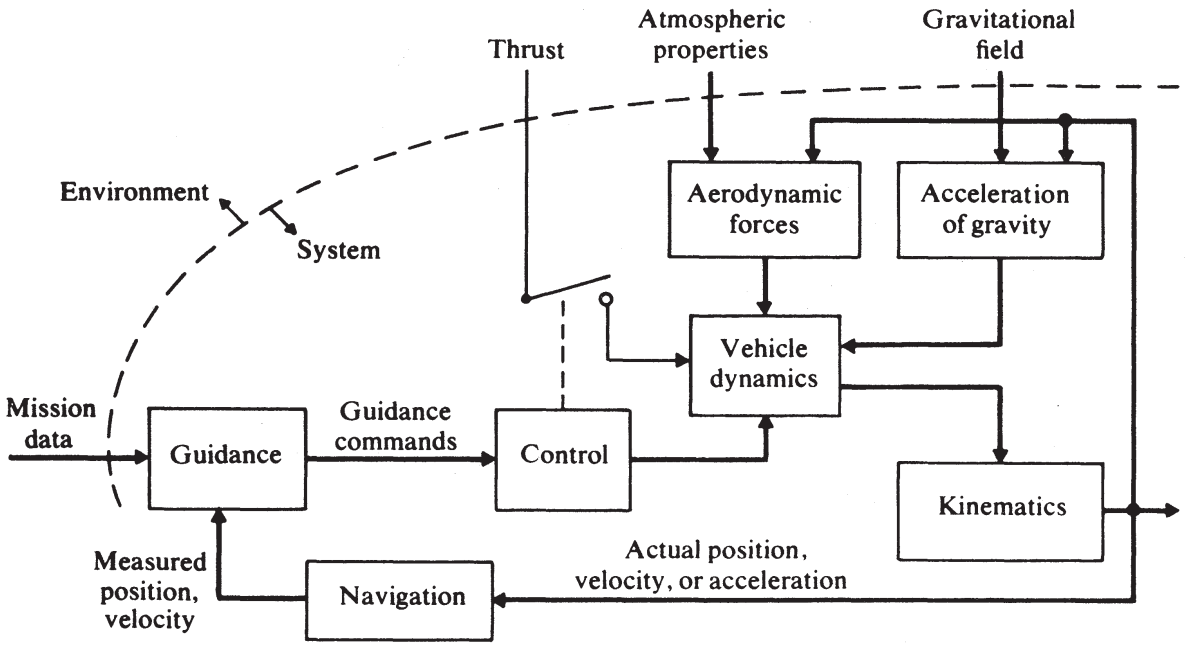


Figure 1.23

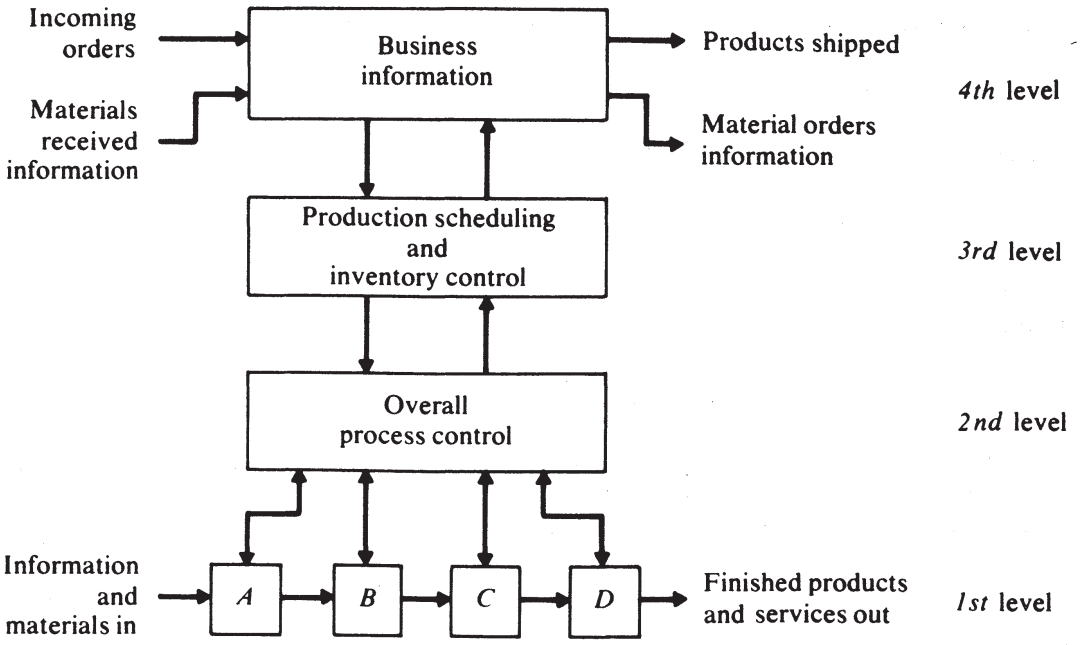


Figure 1.24

5. The remaining steps in Figure 1.5 are relatively straightforward if the preceding steps have been carried out correctly. The resulting equations will be nonlinear and involve many variables, coordinate transformations, etc. Except in certain simple cases, computer simulation will be required.

1.14 Discuss the various levels at which control techniques can be applied in industrial systems.

Harold Chestnut [18] gives the four-level representation to control activities in the business environment shown in Figure 1.24.

Individual operations *A, B, C,* and *D* might represent automatic machine tools, as in Example 1.3. At this level, a very complete model is required to study detailed behavior. The overall sequence *A, B, C, D* could represent a process such as an automated steel mill or chemical plant. The general characteristics of the sequence might be modeled in a way similar to the system of Problem 1.4. The characteristics of the interacting sequence would be described by the level 2 model, but details of individual elements need no longer be apparent. At level 3 a still broader view is taken. The model might be used to determine how work schedules should be set in order to efficiently use the system capability while maintaining optimum inventory, avoiding premium overtime pay, and meeting delivery schedules. At the fourth level, the broadest view is taken. A complete production line might be viewed as a simple time delay. Broader questions regarding market forecasts, new product development, plant expansion, and customer relations become dominant.

According to Mr. Chestnut, "Traditionally the automatic control engineers have focused their attention on the first and second levels of control, which are those associated with the fast control functions in the energy and materials ends of the industrial spectrum. With the current emphasis being developed in the systems aspect of the overall industrial process, more attention is being given to the third and fourth levels of control where significant economies in time and money and resources can and are being realized and can be more readily brought to the attention of the customers." Some of these "big picture" economic systems models are now being facilitated by the various spread sheet and data base management programs which are widely available on management's personal microcomputers.

1.15 Define the Z-transform of a continuous-time signal $y(t)$.

The Z-transform of $y(t)$, written $Z\{y(t)\} = Y(z)$, is defined as the result of a three-step operation:

(i) Modulate $y(t)$ with a periodic train of Dirac delta functions, i.e.,

$$y_s(t) = y(t) \sum_{n=-\infty}^{\infty} \delta(t - nT)$$

where T is the sample period. Since $\delta(\ )$ is zero, except when its argument is zero,

$$y_s(t) = \sum_{n=0}^{\infty} y(nT)\delta(t - nT)$$

assuming $y(t) = 0$ for $t < 0$.

(ii) Laplace transform the impulse-modulated signal

$$Y^*(s) \triangleq \mathcal{L}\{y_s(t)\} = \sum_{n=0}^{\infty} y(nT)e^{-nTs}$$

Note that $y(nT)$ is no longer a function, but just a set of sample values. These act as constants as far as $\mathcal{L}$ is concerned.

(iii) Make a change of variables $z = e^{Ts}$. Thus

$$Y(z) = Y^*(s)|_{z=e^{Ts}} = \sum_{n=0}^{\infty} y(nT)z^{-n}$$

The reason for the change of variables is to allow working with polynomials in z rather than transcendental functions in s.

1.16 What is the significance of the Z-transform as expressed in the previous problem?

The Z-transform of a function can be written in various other forms such as ratios of polynomials in z or z^{-1}. However, if a function's Z-transform can be manipulated into an infinite series in z^{-1}, then we can pick off the function's value at time $t = nT$ as the coefficient multiplying z^{-n}. This series form can be found by long division or by using knowledge of some standard infinite series results.

If $y(t)$ is a unit step, then all $y(nT) = 1$, so

$$Y(z) = \sum_{n=0}^{\infty} z^{-n} = \frac{1}{1 - z^{-1}}$$

Conversely, if $Y(z) = \dfrac{z}{z - 0.5}$, then long division gives

$$Y(z) = 1 + 0.5z^{-1} + 0.25z^{-2} + \cdots$$

From this it is immediately known that

$$y(0) = 1, y(T) = 0.5, y(2T) = 0.25, \ldots, \text{etc.}$$

1.17 What are some other methods of determining the inverse Z-transform?

(i) There are extensive tables of transform pairs available [19]. It should be pointed out that the a function has unique Z-transform, but the inverse transform is not unique. Many functions have the same sample values, but are different between samples.

(ii) A complicated transform expression can often be written as the sum of several simple terms, using a variation of partial fraction expansion. Then each simple term can be inverted.

(iii) The formal definition of the inverse transform is

$$y(nT) = \frac{1}{2\pi j} \oint Y(z)z^{n-1}\,dz$$

The contour integral is around a closed path that encloses all singularities of $Y(z)$. This integral can be evaluated using Cauchy's residue theory of complex variables.

1.18 Why are Z-transforms useful when dealing with constant coefficient difference equations and sampled-data signals?

The Z-transform possesses all the advantages for these systems as does the Laplace transform with differential equations. It allows much of the solution effort to be carried out with only algebraic manipulations in z. After the transform of the desired output variable $Y(z)$ is isolated algebraically, then its inverse can be calculated to give $y(nT)$.

1.19 Analyze the difference equation

$$y(t_k) + a_2 y(t_{k-1}) + a_1 y(t_{k-2}) + a_0 y(t_{k-3}) = b_0 x(t_k) + b_1 x(t_{k-1})$$

$x(t_k)$ is a known input sequence.

Let $Y(z)$ and $X(z)$ be the Z-transforms of y and x, respectively. Since the Z-transform is a linear operator, it can be applied to each individual term in the sum, giving

$$Y(z) + a_2 z^{-1} Y(z) + a_1 z^{-2} Y(z) + a_0 z^{-3} Y(z) = b_0 X(z) + b_1 z^{-1} X(z)$$

The "delay operator" nature of z^{-1} has been used here. As should be apparent from Problems 1.15 and 1.16, a shift of n sample periods in the time domain is achieved by multiplying by z^{-n} in the Z-domain. Thus

$$Y(z) = \left[\frac{b_0 + b_1 z^{-1}}{1 + a_2 z^{-1} + a_1 z^{-2} + a_0 z^{-3}}\right] X(z)$$

The output transform $Y(z)$ is the input transform $X(z)$ multiplied by a rational function of z^{-1} (or z). This rational function is the Z-domain transfer function $H(z)$.

1.20 What is the significance of the poles and zeros of $H(z)$?

Just as in the Laplace s-domain, the behavior of the system depends very heavily on the roots of the denominator of $H(z)$, i.e., the poles. A stable system must not have any s-plane roots with positive real parts. Since $z = e^{Ts}$ this means that in the z-plane all poles of a stable system must be inside the unit circle.

The zeros of the transfer function affect the magnitude of the various terms in the time-domain output. That is, the poles determine the system modes and the zeros help determine how strongly the modes will contribute to the total response. This is evident if $H(z)$ is expanded in partial fraction form.

PROBLEMS

1.21 Derive the elemental equation for the fluid storage tank of Figure 1.25 and show that it is analogous to an electric capacitance. Let Q be the volume flow rate, P the pressure at the base of the tank of cross sectional area A, and h the height of the fluid.

1.22 Show that an inventory storage unit can be modeled by an elemental equation analogous to an electric capacitance. Let the net flow of goods into inventory be Q items per unit time, and let the number of items in inventory at time t be $v(t)$.

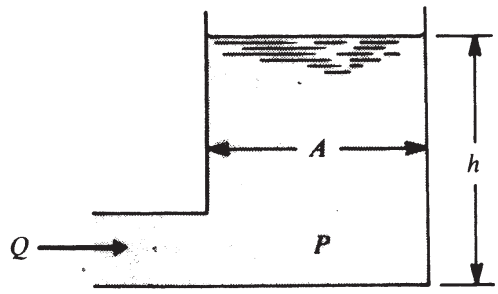


Figure 1.25

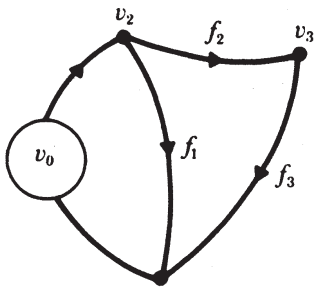


Figure 1.26

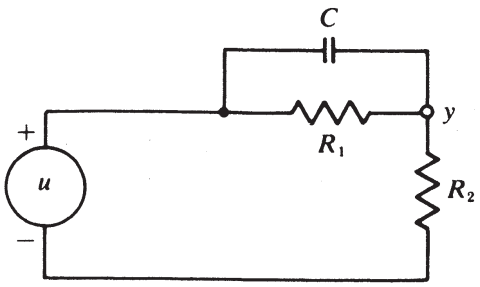


Figure 1.27

1.23 If branch 1 of Figure 1.26 contains an ideal capacitance C, branch 2 an ideal inductance L, and branch 3 an ideal resistance R, find the input-output equation relating v_0 and f_3.

1.24 Derive the input-output differential equation for the network of Figure 1.27. Treat $u(t)$ as the input voltage and $y(t)$ as the output voltage. Also give the input-output transfer function $y(s)/u(s)$.

1.25 A government agency would like a model for studying the effectiveness of its air pollution monitoring and control program. Discuss the factors involved in such a model.

1.26 Derive the Z-transform of $y(t) = e^{-0.1t}$. For $t < 0$, $y(t) = 0$. Use a sample time of $T = 2.0$ s.

1.27 The following values apply to the system in Problem 1.19.

$$a_0 = -0.064, \quad a_1 = 0.56, \quad a_2 = -1.4, \quad b_0 = 10.0, \quad \text{and} \quad b_1 = 5.0$$

Find the poles and zeros of $H(z)$. Does this represent a stable system?

1.28 For the system of Problem 1.19, find $y(nT)$ if the input $x(nT)$ is the sampled version of a unit step function starting at $t = 0$.

2

Highlights of Classical Control Theory

2.1 INTRODUCTION

Classical control theory, at the introductory level, deals primarily with linear, constant coefficient systems. Few real systems are exactly linear over their whole operating range, and few systems have parameter values that are precisely constant forever. But many systems approximately satisfy these conditions over a sufficiently narrow operating range. This chapter reviews the classical methods, which are applicable to linear, constant coefficient systems. More extensive discussions are in References 1 through 5.

2.2 SYSTEM REPRESENTATION

The consideration of linear, stationary systems is greatly simplified by the use of transform techniques and frequency domain methods. For continuous-time systems this means Laplace transforms (or sometimes Fourier transforms) [4]. Z-transforms provide equivalent advantages for discrete-time systems [6, 7]. These methods are basic in classical control systems analysis. Thus algebraic equations in the transformed variables are dealt with rather than the system's differential or difference equations. Manipulation of the algebraic cause and effect relations is facilitated by the use of transfer functions and block diagrams or signal flow graphs [1].

2.3 FEEDBACK

Most systems considered in classical control theory are feedback control systems. A typical single-input, single-output continuous-time (totally analog) system is shown in Figure 2.1*a*. Figure 2.1*b* shows a typical feedback arrangement for controlling a continuous-time process using a digital controller. Notation commonly used in classical

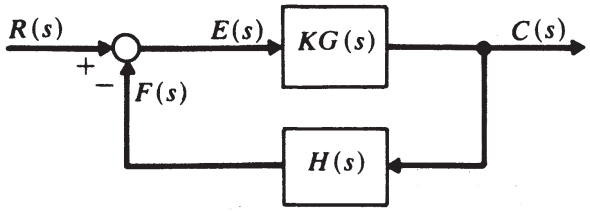


Figure 2.1*a* Elementary feedback control system.

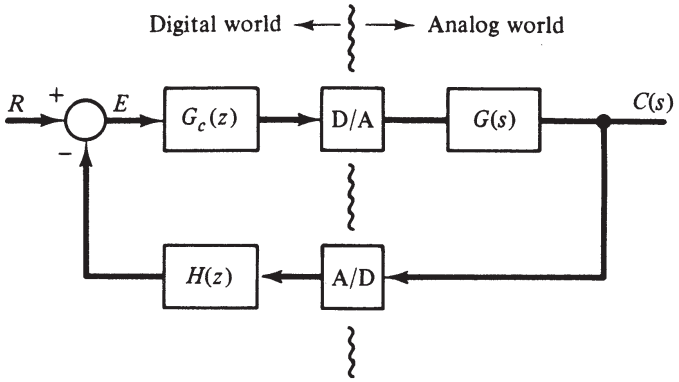


Figure 2.1*b*

control theory will be used in this chapter. In both cases the input or reference signal is R, the output or controlled signal is C, the actuating or error signal is E, and the feedback signal is F. This should cause no confusion with the parameters R and C used in resistance and capacitance networks.

While there are many similarities between these two types of systems, the differences are significant enough to merit a brief separate discussion of each.

Continuous-Time Systems

The forward *transfer function is* $KG(s)$, where K is an adjustable gain. The forward transfer function often consists of two factors $G(s) = G_c(s)G_p(s)$, where $G_p(s)$ is fixed by the nature of the plant or process to be controlled. $G_c(s)$ is a compensation or controller transfer function, which the designer can specify (within certain limits) to achieve desired system behavior. The *feedback transfer function* is $H(s)$. This often represents the dynamics of the instrumentation used to form the feedback signals, but it can also include signal conditioning or compensation networks. The designer may be able to at least partially specify $H(s)$ in some cases, and in other cases it may be totally fixed, or even just unity. In any case, the *open-loop transfer function* is $KG(s)H(s)$. It represents the transfer function around the loop, say from E to F, when the feedback signal is disconnected from the summing junction.

In the feedback system of Figure 2.1*a* and *b* the actuating signal is determined by comparing the feedback signal with the input signal. When $H(s) = 1$, the unity feedback case, the comparison is directly between the output and the input. Then the difference E is truly an error signal.

A major part of classical control theory for continuous-time systems is devoted to the analysis of feedback systems like the one shown in Figure 2.1*a*. Multiple input-output systems and multi-loop systems can also be considered using transfer function techniques (see Problems 4.2 through 4.7), although most of this book is devoted to a

state variable approach instead. It is beneficial to have a thorough understanding of single-input, single-output systems before the multivariable case is considered. This chapter provides a review of the methods used in studying the behavior of C and E as influenced by R.

EXAMPLE 2.1 Relations, in the Laplace transform domain, between the input R and the output C and between R and the error E are derived algebraically as follows. At the summing junction, $R - HC = E$. The relation between E and C is $KGE = C$. Elimination of E gives $KGR - KGHC = C$, so that $C = KGR/(1 + KGH)$. Using $E = C/KG$ gives $E = R/(1 + KGH)$. ■

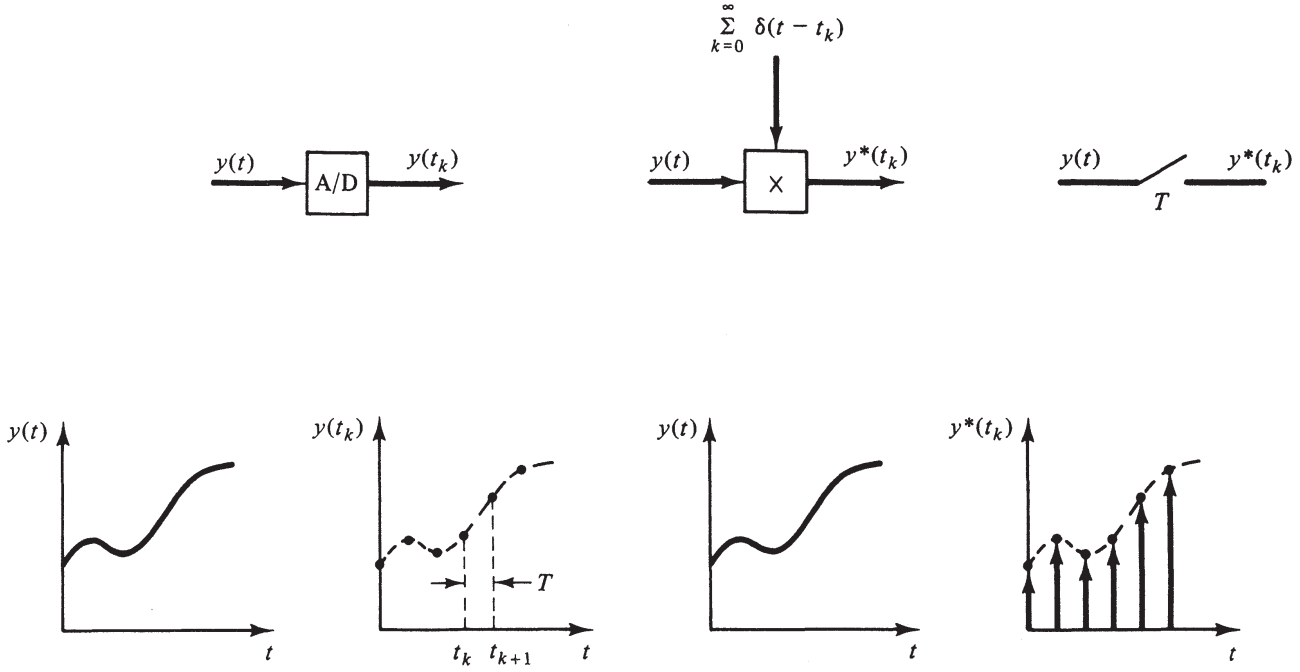
The system of Figure 2.1*a* is the prototype for all continuous system discussions in this chapter. The following terminology will be used frequently. In general, $G(s) = g_n(s)/g_d(s)$ and $H(s) = h_n(s)/h_d(s)$ will be ratios of polynomials in s. The values of s which are roots of the numerator are called *zeros*. Roots of the denominator are called *poles*. In particular, the *open-loop zeros* are values of s which are roots of the numerator of the open-loop transfer function $KG(s)H(s) = Kg_n(s)h_n(s)/[g_d(s)h_d(s)]$. The *open-loop poles* are roots of the denominator of $KG(s)H(s)$. Since the closed-loop transfer function is $C(s)/R(s) = KG(s)/[1 + KG(s)H(s)] = Kg_n(s)h_d(s)/[g_d(s)h_d(s) + Kg_n(s)h_n(s)]$, the *closed-loop zeros* are all the roots of $g_n(s)h_d(s)$. The *closed-loop poles* are roots of $1 + KG(s)H(s) = 0$ or equivalently, roots of $g_d(s)h_d(s) + Kg_n(s)h_n(s) = 0$.

Discrete-Time Systems

There is a richer variety of possibilities when dealing with the digital control of continuous systems. The points of conversion from continuous-time signals to discrete-time signals (A/D) and back again (D/A) can vary from one application to the next. An analog feedback sensor could be used and then its output sampled, or a direct digital measurement may be used. The reference input R could be a continuous-time signal that needs to be sampled before being sent to the control computer, or it might be a direct digital input. Figure 2.1*b* is just one possible arrangement. Other configurations can be analyzed in a similar way. Before proceeding with the analysis, models of the A/D and D/A conversion processes are required. These conversions are also referred to as sampling and desampling or signal reconstruction, respectively.

Sampling. Assume a periodic sampler with period T. A convenient model of the A/D conversion is an impulse modulator, usually shown symbolically as a switch like the one of Figure 2.2, where a general signal $y(t)$ is being sampled. Impulse modulation is not really what physically occurs, since no infinite amplitude signals such as $y^*(t_k)$ actually exist in the system. This series of impulse functions have infinite amplitude at the sample times, but it is their areas or strengths that represent the real signal amplitudes $y(t_k)$ mathematically. This artificial representation is used because

1. It allows the use of Z-transforms, which simplify much of the analysis.
2. The correct answers are obtained (except for quantization effects) as long as it is understood that within the digital portion it is the strengths of the impulses, not their amplitudes, that describe the signals.



(a) Reality (b) Impulse modulation model

Figure 2.2

3. The correct effect on the continuous-time part of the system is obtained provided some sort of "hold" circuit is used on the impulse train before the signal reenters the analog world. There are various versions of hold devices. One function common to them all is an integration, which eliminates the impulses and once again gives a finite amplitude physical signal. This is the desampling function of the D/A.

Desampling. The only model of the D/A process to be considered here is another sampler (perfect time synchronization assumed), followed by a zero order hold (ZOH). The zero order hold integrates the difference between two consecutive impulses in the periodic impulse train shown in Figure 2.2 and repeated in Figure 2.3. Therefore, the output is a piecewise constant signal whose value between t_k and t_{k+1} is clamped at $y(t_k)$ (again, ignoring quantization errors). If the computer made no modification to the signal between the A/D and D/A, the end-to-end effect of this sampling-desampling operation would be to create a piecewise constant approximation to the continuous input signal. From Chapter 1, the Laplace transform of an impulse-modulated signal is (after a change of variable) the Z-transform of the signal. Any linear operation that the computer algorithm performs on the signal samples between the A/D and D/A can be represented by a Z-transform domain transfer function, sometimes called a pulse transfer function. $G_c(z)$ and $H(z)$ in Figure 2.1*b* are examples of this.

Extensive tables of Z-transforms are available [7]. Use of these tables, plus a few simple rules, will allow systems like Figure 2.1*b* to be analyzed almost as easily as, and

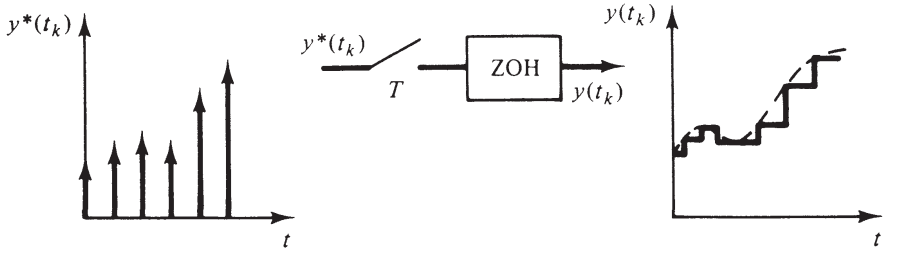
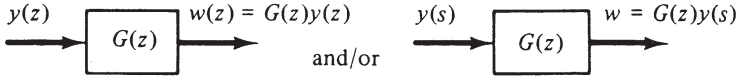


Figure 2.3

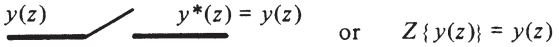
with a great deal of similarity to, those of Figure 2.1*a*. Of course, a complete understanding and appreciation will require a more thorough treatment, as can be found in the references. Some key rules of manipulation are:

1. *As a signal passes through the "switch," it is Z-transformed.*

2. *The transform of the signal out of a transfer function block is the product of that transfer function and the transform of the input signal.*



3. *Sampling a signal that is already sampled does not change it.*

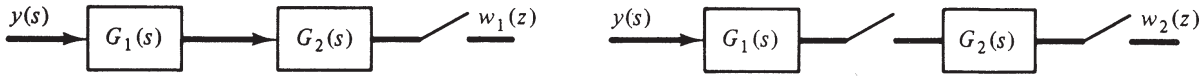


4. *Pulsed or Z-transformed signals (and transfer functions) and s-domain signals (and transfer functions) will appear together in the same expression at times.* The action of a sampler (or a Z-transform) on these mixed signals is illustrated next.

$$Z\{G(z)y(s)\} = G(z)y(z)$$

5. *The Z-transform operator is not associative for products.* The placement of "switches" in a block diagram is important.

$$w_1(z) = Z\{y(s)G_1(s)G_2(s)\} \neq w_2(z) = Z\{G_1(s)y(s)\}G_2(z)$$



6. *When working with closed-loop systems like Figure 2.1b, it is generally best to follow a two-step process.* First, algebraically solve for the variable at the input of a sampler in terms of external inputs and/or outputs of that sampler. Second, "close the loop" by passing through the sampler, i.e., taking the Z-transform. This will give a result, entirely in the Z-domain, which can be used to solve for the system output sequence at the sample times. If the sampling period is small enough compared to the rate of change of the signal, this approximation may be all that is needed to describe the continuous system output.

EXAMPLE 2.2 The system of Figure 2.1*b* is redrawn in Figure 2.4, using symbolism just introduced for A/D and D/A operations. The transfer function for a ZOH is also used, $G_0 = (1 - z^{-1})/s$.

The input to the forward path sampler is first isolated:

$$E_1(z) = G_c(z)[R(z) - F(z)]$$
$$= G_c(z)[R(z) - E_1^*(z)Z\{(1 - z^{-1})G(s)/s\}H(z)]$$

Define the Z-transform of $G(s)/s$ times $(1 - z^{-1})$ as $G'(z)$. Then

$$E_1(z) = G_c(z)[R(z) - E_1(z)G'(z)H(z)]$$

Solving gives

$$E_1(z) = G_c(z)R(z)/[1 + G_c(z)G'(z)H(z)]$$

In this case the input to the selected sampler was already sampled, so the second step of passing through the sampler has no effect, i.e. $E_1^*(z) = E_1(z)$. The full two-step process is better illustrated by selecting the feedback sampler instead.

Before doing that, note that

$$C(s) = E_1(z)[(1 - z^{-1})G(s)/s)]$$

and that

$$C(z) = E_1(z)G'(z)$$
$$= G_c(z)G'(z)R(z)/[1 + G_c(z)G'(z)H(z)],$$

an expression very similar to the continuous system result.

The same example is reworked by isolating the input to the feedback sampler first:

$$C(s) = [R(z) - H(z)C(z)]G_c(z)(1 - z^{-1})G(s)/s$$

Then the traverse around the loop is completed by passing through the sampler to obtain

$$C(z) = [R(z) - H(z)C(z)]G_c(z)G'(z)$$

Solving this for $C(z)$ gives the same result as above.

In order to dispel the idea that a discrete system result can always be written from the continuous system result by substituting all the individual Z-transfer functions for their Laplace transform counterparts, the reader is urged to work through Problems 2.38, 2.39, and 2.41, which have different sampling arrangements. ■

Once $C(z)$ is found, the values of the output $C(t)$ can be determined *at the sample times* t_k by finding the inverse Z-transform. While this will not give the values of $C(t)$

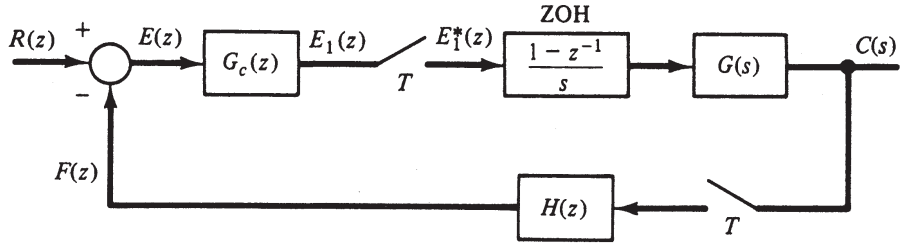


Figure 2.4

between sample points, it often gives a sufficiently accurate representation of the continuous signal.

Control systems like those of Figure 2.1*a* and *b* are used extensively because of the advantages that can be obtained by using feedback. The advantages of feedback control are:

1. The system output can be made to follow or track the specified input function in an automatic fashion. The name automatic control theory is frequently used for this reason.
2. System performance is less sensitive to variations of parameter values (see Problem 2.1).
3. System performance is less sensitive to unwanted disturbances (see Problem 2.4).
4. Use of feedback makes it easier to achieve the desired transient and steady-state response (see Problem 2.5).

The advantages of feedback are gained at the expense of certain disadvantages, the principal ones being:

1. The possibility of instability is introduced and stability becomes a major design concern. Actually, feedback can either stabilize or destabilize a system.
2. There is a loss of system gain, and additional stages of amplification may be required to compensate for this.
3. Additional components of high precision are usually required to provide the feedback signals (see Problems 2.2 and 2.3).

Once the system models are specified in terms of transfer functions and block diagrams, classical control theory is devoted to answering three general questions:

(a) What are appropriate measures of system performance that can be easily applied to feedback control systems?

(b) How can a feedback control system be easily analyzed in terms of these performance measures?

(c) How should the system be modified if its performance is not satisfactory?

2.4 MEASURES OF PERFORMANCE AND METHODS OF ANALYSIS IN CLASSICAL CONTROL THEORY

If the complete solutions for the system output $C(t)$ were available in analytical form for every conceivable input, system performance could be assessed. To obtain an analytical expression for $C(t)$, the inverse Laplace transform of

$$C(s) = \frac{KG(s)R(s)}{1 + KG(s)H(s)} \tag{2.1a}$$

is required. To determine $C(t_k)$ in the discrete-time case or in the mixed discrete-continuous case, the inverse Z-transform of

$$C(z) = \frac{G_c(z)G'(z)R(z)}{1 + G_c(z)G'(z)H(z)} \tag{2.1b}$$

(or a similar expression) is needed. In both cases, if the denominators can be factored, partial fraction expansion can be used to obtain a sum of easily invertible terms. However, the denominator of equation (*2.1*) may be a high-degree polynomial. Also, an infinite number of possible inputs $R(s)$ or $R(z)$ could be considered. Rather than seek complete analytical solutions, classical control theory uses only certain desirable features, which C should possess, in order to evaluate performance. Methods of classical control theory were developed before the widespread availability of digital computers. As a result, all the techniques seek as much information as possible about the behavior of $C(t)$ or $C(t_k)$ without actually solving for them. The methods have been developed for ease of application and stress graphical techniques. They are still useful methods of analysis and design because of the insight they provide.

The problem of infinite variety for possible inputs is dealt with by considering important aperiodic and periodic signals as test inputs. Step functions, ramps, and sinusoids are common examples.

The general characteristics a well-designed control system should possess are (1) stability, (2) steady-state accuracy, (3) satisfactory transient response, (4) satisfactory frequency response, and (5) reduced sensitivity to model parameter variations and disturbance inputs. These requirements are interrelated in various ways and often present conflicting goals. For example, decreasing response times generally requires increasing system bandwidth, which increases suceptibility to high-frequency noise. A key concept in many aspects of feedback system's performance is the so-called return difference function, which happens to be the denominators of Eqs. (*2.1a*) and (*2.1b*). These are both of the form $R_d(\zeta) = 1 + KF(\zeta)$, with the complex variable ζ being either s or z. The name *return difference* arises as follows. If a feedback loop is broken at a given point, the difference between a signal v inserted into the loop at that point and the resulting signal r, which returns to the broken point after traversing the loop, is

$$v - r = [1 + KF(\zeta)]v = R_d(\zeta)v$$

A graphical interpretation of the return difference will be given after polar plots—i.e., Nyquist plots—are introduced.

1. *Stability means that $C(t)$ or $C(t_k)$ must not grow without bound due to a bounded input, initial condition, or unwanted disturbance.* (This intuitive definition is expanded upon in Chapters 10 and 15.) For linear constant coefficient systems, stability depends only on the locations of the roots of the closed-loop characteristic equation. The continuous system's characteristic equation is the denominator of equation (*2.1a*) set to zero. The discrete case uses the denominator of equation (*2.1b*). Both of these are of the form

$$R_d(\zeta) = 1 + KF(\zeta) = 0 \tag{2.2}$$

with the complex variable ζ being either s or z. The principal difference is the stability region. The roots must be in the left-half s-plane for a stable continuous system. Since $z = e^{Ts}$, the entire left-half s-plane maps into the interior of the unit circle in the Z-plane. A stable discrete system must have all roots of its characteristic equation inside the unit circle. Methods of determining stability are discussed below.

a. *Routh's criterion* determines how many roots have positive real parts directly from the coefficients of the characteristic polynomial. The actual root locations are not found. The number of unstable roots of a continuous system are obtained directly. Routh's criterion can also be applied to a discrete system, but first a bilinear transformation $z = (w + 1)/(w - 1)$ is used to map the inside of the unit circle in the Z-plane into the left half of a new complex w-plane. This converts the characteristic equation into a polynomial in w, to which Routh's criterion can be applied.

b. *Root locus* is a graphical means of factoring the characteristic equation (or any algebraic polynomial of similar form). Both continuous and discrete systems can be described simultaneously by using Eq. (*2.2*) as the characteristic equation. The essence of the method is to consider $KF(\zeta) = -1$. Since $F(\zeta)$ is a complex number with a magnitude and a phase angle, this implies two conditions, which are considered separately. They are, assuming the gain is positive and real,

$$\angle F(\zeta) = (1 + 2m)180^\circ \qquad \text{for any integer } m \tag{2.3a}$$

and

$$K|F(\zeta)| = 1 \tag{2.3b}$$

Thus root locus determines the closed-loop roots (and therefore stability) by working with the open-loop transfer function $KF(\zeta)$, which is normally available in factored form.

c. *Bode plots* are another graphical method which provides stability information for *minimum phase systems* (systems with no open-loop poles or zeros in the unstable region). Magnitude and phase angle are considered separately, as in Eqs. (*2.3a*) and (*2.3b*), but the only values of ζ considered are on the stability boundary. This technique is widely used with continuous systems, in which case the stability boundary is defined by $s = j\omega$. This corresponds to the consideration of sinusoidal input functions with frequencies ω. This method is greatly simplified by using decibel units for magnitude and a logarithmic frequency scale for plotting. This allows for rapid construction of straight line asymptotic approximations of the magnitude plot. The critical point for stability, -1, becomes the point of 0 db and -180° phase shift. Bode techniques can also be applied to discrete systems by first using the same bilinear transformation as was mentioned under Routh's criterion. The stability boundary in the w-plane can also be characterized by the purely imaginary values $w = j\omega_w$. This "transformed" frequency ω_w is generally badly distorted from the true sinusoidal frequency, so intuition is of less value in this approach, insofar as stability margins, bandwidths, and similar concepts are concerned. For this reason, Bode methods are probably used less often with discrete systems, and they will not be pursued here. The same is more or less

true for the following two frequency domain methods as well, so they are only discussed for continuous systems.

d. *Polar plots and Nyquist's stability criterion.* Polar plots convey much the same information as Bode plots, but the term $KG(j\omega)H(j\omega)$ is plotted as a locus of phasors with ω as the parameter. The critical point is again -1. Note that the return difference $R_d(j\omega)$ (or $R_d(e^{j\omega T})$ in the discrete case) can be represented as a phasor from the critical -1 point to a point on a plot of the loop transfer function $KG(j\omega)H(j\omega)$. (See Problem 2.1 and, in particular, Figure 2.9*c*.) Nyquist's stability criterion, which applies to nonminimum phase systems as well, states that the number of unstable closed-loop poles is $Z_R = P_R - N$, where N is the number of encirclements of the critical point -1 made by the locus of phasors. Counterclockwise encirclements are considered positive. P_R is the number of open-loop poles in the right-half plane.

e. *Log magnitude versus angle plots* are sometimes used for stability analysis. They contain the same information as Bode plots, but magnitude and angle are combined on a single graph with ω as a parameter.

2. *Steady-state accuracy requires that the signal $E(t)$, which is often an error signal, approach a sufficiently small value for large values of time.* The final value theorem facilitates analyzing the requirement without actually finding inverse transforms. That is, for continuous systems

$$\lim_{t\to\infty}\{E(t)\} = \lim_{s\to 0}\{sE(s)\} \tag{2.4a}$$

For discrete systems, the Z-transform version of the final value theorem is used,

$$\lim_{k\to\infty}\{E(t_k)\} = \lim_{(z\to 1)}\{(z-1)E(z)\} \tag{2.4b}$$

Both versions of the final value theorem are only valid when the indicated limits exist. By considering step, ramp, and parabolic test inputs, the useful parameters called *position, velocity,* and *acceleration* (or step, ramp, and parabolic) *error constants* are developed. These provide direct indications of steady-state accuracy (see Problem 2.8).

3. *Satisfactory transient response means there is no excessive overshoot for abrupt inputs, an acceptable level of oscillation in an acceptable frequency range, and satisfactory speed of response and settling time, among other things.* These are actually questions of relative stability, and depend upon the location of the closed-loop poles in the s-plane or Z-plane and their proximity to the stability boundary. Questions regarding transient response are best studied using root locus, since it is the only classical method which actually determines closed-loop pole locations. Bode, Nyquist, and log-magnitude plot methods also give information regarding transient response, at least indirectly. *Gain margin GM* is a measure of additional gain a system can tolerate with no change in phase, while remaining stable. *Phase margin PM* is the additional phase shift that can be tolerated, with no gain change, while remaining stable. Note that these stability margins are measures of the magnitude of the minimum return differ-

ence phasor. Experience has shown that acceptable transient response will usually require stability margins on the order of

$$PM > 30°, \qquad GM > 6 \text{ db}$$

These frequency domain stability margins can often be used to draw conclusions regarding transient performance, because many control systems have their response characteristics dominated by a pair of underdamped complex poles. For this case known correlations exist between frequency domain and time domain characteristics. A few approximate rules of thumb are

$$\text{damping ratio} \cong 0.01\, PM \text{ (in degrees)}$$

$$\%\text{ overshoot} + PM \cong 75$$

$$(\text{rise time})(\text{closed-loop bandwidth in rad/s}) \cong 0.45\ (2\pi)$$

Other response times have similar inverse relationships with bandwidth. The frequency of 0 db magnitude for the open-loop KGH term has an effect similar to bandwidth. Increasing this crossover frequency increases bandwidth and decreases response times.

4. *Satisfactory frequency response implies such things as satisfactory bandwidth, limits on maximum input-to-output magnification, frequency at which this magnification occurs, as well as gain and phase margin specifications.* Bode, Nyquist, and log-magnitude-angle plots all are frequency response methods, and they deal with the open-loop transfer function. If the closed-loop characteristics, such as closed-loop bandwidth, must be determined, then the *Nichol's chart* [1] can be used. The Nichol's chart is a graphical conversion from open-loop magnitude-phase characteristics to closed-loop characteristics. Normally, one of the open-loop graphical methods is first used and the results are then transferred to a Nichol's chart. From this, the closed-loop frequency response characteristics can be read off directly.

5. *Since perfect models are never available, either because of intentional simplifications or because of unavoidable ignorance, time variations, or noise corruption, a good control system must be at least somewhat forgiving of these errors.* Problems 2.1 through 2.4 briefly review how feedback can lead to reduced sensitivity to external disturbances and internal parameter variations. The concept of return difference plays a prominent role [8, 9].

2.5 METHODS OF IMPROVING SYSTEM PERFORMANCE

Whenever the performance of a feedback control system is not satisfactory, the following possible approaches should be considered.

1. A simple adjustment of the gain parameter K. This could be considered by using any of the analysis methods mentioned in the preceding paragraphs. From a consideration of the system's root locus, it is obvious that gain adjustment can only shift the closed-loop poles along well-defined loci. Perhaps no points on these loci give satisfactory results.

2. Minor changes in the system's structure, such as adding additional measurements to be used as feedback signals. Addition of minor feedback loops can alter the loci of possible pole locations as K is varied. The inclusion of a rate feedback loop, using a tachometer for example, is a common means of improving stability.
3. Major changes in the system's structure or components. A hydraulic motor may perform better than an electric motor in some cases. A higher capacity pump or a more streamlined aerodynamic shape may be the answer in other cases.
4. Addition of compensating networks—i.e., $G_c(s)$ or $H(s)$—or digital algorithms—i.e., $G_c(z)$ or $H(z)$—to alter the root locus or to change the magnitude and phase characteristics in a critical frequency range.

Of these four techniques for improvement, only the second and fourth constitute what are usually referred to as *compensation techniques*. The advantages of root locus, Bode, and Nyquist methods of analysis are that compensating changes in the open-loop transfer function can be rapidly taken into account. The modifications may be made in order to reshape the locus, improve gain or phase margins, or increase the error constants. The classical methods thus constitute design techniques as well as analysis techniques. A process of design by analysis is usually used. That is, a compensating network is selected and then analyzed. However, a little experience gives great insight into the kinds of compensation that are needed. If the major problem is to improve relative stability with less concern for error constants, lead compensation networks are usually tried. If the system has acceptable stability margins, but poor steady-state accuracy, lag compensation networks will usually be appropriate. If a combination of both improvements is needed, a lag-lead network may give the desired results. More complicated networks, such as the bridged-T network, Butterworth filters, and so on, can be used to effectively cancel undesirable left-half-plane poles and replace them with more favorable ones. Cancellation compensation should never be used to eliminate unstable poles, because parameter tolerances will preclude exact cancellation. Even an infinitesimal error in cancellation will leave an unstable closed-loop pole. The form of the desired specifications and the personal preference of the designer will influence the choice of the analysis method. Extra insight can usually be gained by looking at a compensation problem from both the root locus and one of the frequency domain techniques.

An alternative method of design, through synthesis rather than analysis, is also possible. In this approach, the design specifications are translated into a desired closed-loop transfer function which satisfies them. Let the closed-loop transfer function be $M(z)$. This can be related to the compensator $G_c(z)$. For example, the system of Example 2.2 has

$$M(z) = G_c(z)G'(z)/[1 + G_c(z)G'(z)H(z)]$$

which can be solved to give

$$G_c(z) = \frac{M(z)}{[1 - M(z)H(z)]G'(z)} \tag{2.5}$$

Because of this result it is clear that certain restrictions must be imposed on $M(z)$ if the resulting compensator is to be physically realizable. This is discussed in Problems 2.23

and 2.24. More details on this method can be found in References 6 and 7. Since discrete system compensators are just computer algorithms, there is no concern about synthesizing the results in terms of passive electrical components R, C, and maybe an occasional L that dominated classical control compensation in the early years. It is perhaps for this reason that the algebraic synthesis methods seem to be more widely used in discrete system design, although the continuous system version was described years earlier by Truxal [10]. Even in the continuous system domain the definition of what is practical now is quite different from the early years because of progress in technology, such as operational amplifiers and large-scale integrated circuit technology.

One final design parameter in discrete systems is the sampling period T. It can have a profound effect on system performance. Nyquist's sampling theorem tells us that a signal must be sampled at least twice per cycle of the highest frequency present in order to avoid losing information about the signal. The highest frequency present is often interpreted as the highest frequency of interest, and the sampler is generally preceded by a low pass filter. This prevents the high-frequency terms from being *aliased* as low-frequency signals as a result of sampling. *Frequencies of interest* are related to system bandwidth, since that is what determines which frequencies the system is capable of passing or responding to. The "twice" is strictly a theoretical limit based on an unachievable ideal low pass filter, which would be needed to reconstruct the original signal from its sampled version. In reality, a cushion is provided by sampling at a considerably higher rate if possible. Frequently a sampling rate of three to five times the Nyquist rate is more appropriate. The reason for the sampling in the first place might be because of time-shared or multiplexed equipment, so possible T values may be restricted in many cases. In closed-loop systems T has another effect beyond the sampling theorem considerations. The value of T interacts with the loop gain K (and, of course, pole-zero locations, too) to determine system stability.

EXAMPLE 2.3 Investigate the system of Figure 2.5 for stability.

The characteristic equation is

$$1 + \frac{K}{(s-10)(s+20)(s+100)} = 0 \quad \text{or} \quad s^3 + 110s^2 + 800s - 20{,}000 + K = 0$$

The Routhian array is a table with one more row than the highest power of s in the characteristic equation. The first two rows are filled in a sawtooth pattern with the coefficients of the characteristic equation. Each succeeding row is computed from terms in the two rows just above it. The pattern for the computed rows is as follows. Suppose two typical rows with a_i and b_i coefficients are available as in Table 2.1*a*. Then the c_i terms are given by

$$c_1 = (b_1 a_2 - a_1 b_2)/b_1, \qquad c_2 = (b_1 a_3 - a_1 b_3)/b_1, \qquad c_3 = (b_1 a_4 - a_1 b_4)/b_1$$

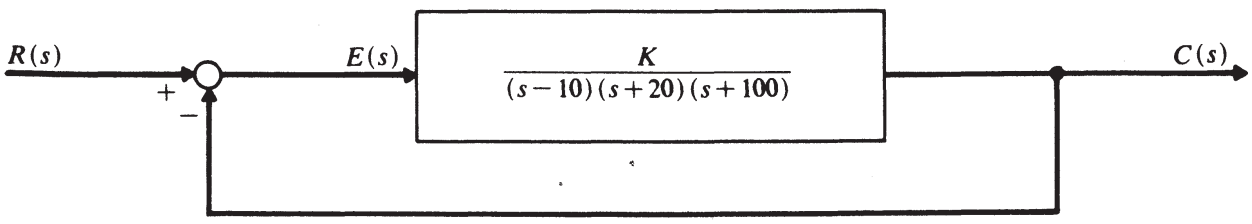


Figure 2.5

TABLE 2.1*a*

a_1	a_2	a_3	a_4
b_1	b_2	b_3	b_4
c_1	c_2	c_3	c_4

TABLE 2.1*b*

s^3	1	800
s^2	110	$K - 20{,}000$
s	$\frac{108{,}000 - K}{110}$	
s^0	$K - 20{,}000$	

Each row is filled in from left to right until all remaining terms are zero. In this case c_4 and all higher c_i terms are zero because of blanks in the a_i and b_i rows.

Table 2.1*b* gives the array for the system of Figure 2.5. Routh's criterion states that the number of sign *changes* in the first column is equal to the number of roots in the right-half plane. For stability the first column must have all entries positive. Therefore, the system is stable if $20{,}000 \leq K \leq 108{,}000$. If $K < 20{,}000$, one sign *change* exists in column one and there will be one unstable root. If $K > 108{,}000$, there are two sign changes and therefore two unstable roots. If $K = 108{,}000$, the s row is zero. Whenever an entire row is zero, the coefficients of the preceding row are used to define the *auxiliary equation*. Roots of the auxiliary equation are also roots of the original characteristic equation. With $K = 108{,}000$, the auxiliary equation is $110s^2 + 88{,}000 = 0$, indicating poles at $s = \pm j\sqrt{800}$. With K at this maximum value, the system oscillates to $\omega = \sqrt{800}$ rad/s. ■

EXAMPLE 2.4 Investigate the steady-state following error for the system of Figure 2.5 if $K = 100{,}000$ and the input is a unit step.

The error is

$$E(s) = \frac{1/s}{1 + \dfrac{K}{(s-10)(s+20)(s+100)}}$$

Using the final value theorem, the steady-state error is

$$E(t)|_{ss} = \frac{1}{1 + \dfrac{K}{(-10)(20)(100)}} = -0.25$$

Since the input is unity, the steady-state output has a 25% error. ■

EXAMPLE 2.5 Add compensation to the previous system in order to achieve a steady-state error of less than 10%. The oscillatory poles should have a damping ratio of $0.7 < \zeta < 0.9$ and a damped natural frequency of about 10 to 20 rad/s.

In order to meet the steady-state error specifications, a gain increase by a factor of 2.2 is required. This would give an unstable system and then the final value theorem cannot be used. Compensation is required, and it will be added in the forward loop. Because of the form of the specifications, root locus will be used. First, the locus of points satisfying Eq. (2.2) is found. The following rules greatly simplify this procedure.

1. The number of branches of the root locus equals the number of open-loop poles. One closed-loop pole will exist on each branch.
2. One branch of the locus starts at each open-loop pole. One branch terminates at each open-loop zero and the remaining branches approach infinity.

3. The part of the locus on the real axis lies to the left of an odd number of poles plus zeros.
4. With $K = 0$, open and closed-loop poles coincide. As K increases, the closed-loop poles move along the loci. As $K \to \infty$ each closed-loop pole approaches either an open-loop zero or infinity.
5. Branches that go to infinity do so along asymptotes with angles given by $\phi_i = 180°(1 + 2k)/(n - m)$ for $k = 0, \pm 1, \pm 2, \ldots$, and where n and m are the number of open-loop poles and zeros respectively.
6. The asymptotes emanate from the *center of gravity* given by cg = [(sum of real parts of all open-loop poles) − (sum of real parts of all open-loop zeros)]/$(n - m)$.
7. The loci are symmetric with respect to the real axis.

By using these rules and by testing the angle criterion at a few additional points off the real axis, the uncompensated root locus of Figure 2.6 is obtained. It is obvious that the locus must be reshaped in order to meet the specifications. The angle criterion at points inside the desired region indicates that an additional 30–70° of phase lead is needed if the locus is to pass through this region. By placing a zero at $s = -20$ and a pole at $s = -45$, adequate phase lead is obtained. However, the lead network transfer function $G_{C1}(s) = (s + 20)/(s + 45)$ introduces a decrease in the error constant by a factor of $\frac{20}{45} = 0.445$. This decrease can be made up, and the additional increase gained by using a lag filter, such as

$$G_{C2} = \frac{(s + 0.1)}{(s + 0.01)}$$

This pole-zero pair near the origin will have only a small effect on the locus in the region of interest since their angle contributions almost cancel each other. Using the compensator transfer function $G_C(s) = (s + 0.1)(s + 20)/[(s + 0.01)(s + 45)]$, the compensated root locus of Figure

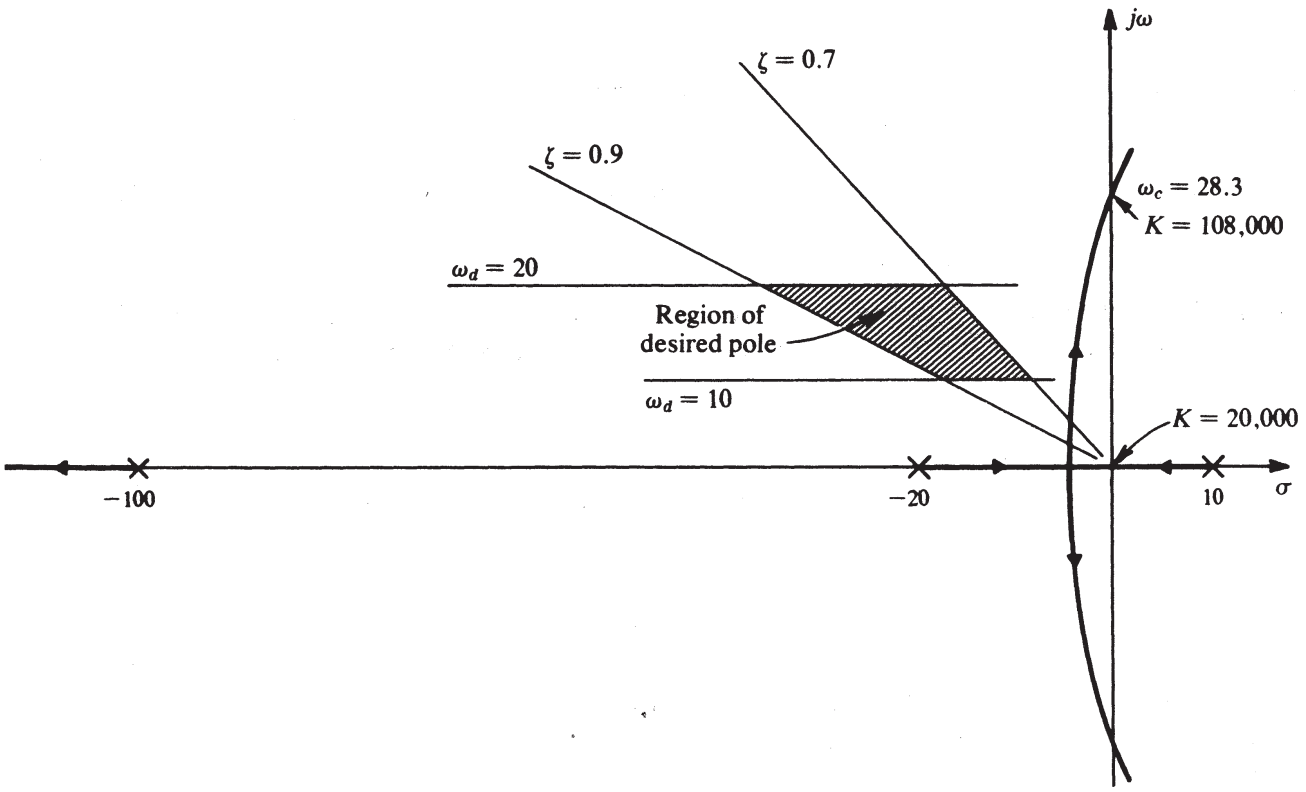


Figure 2.6 Approximate sketch of uncompensated root locus.

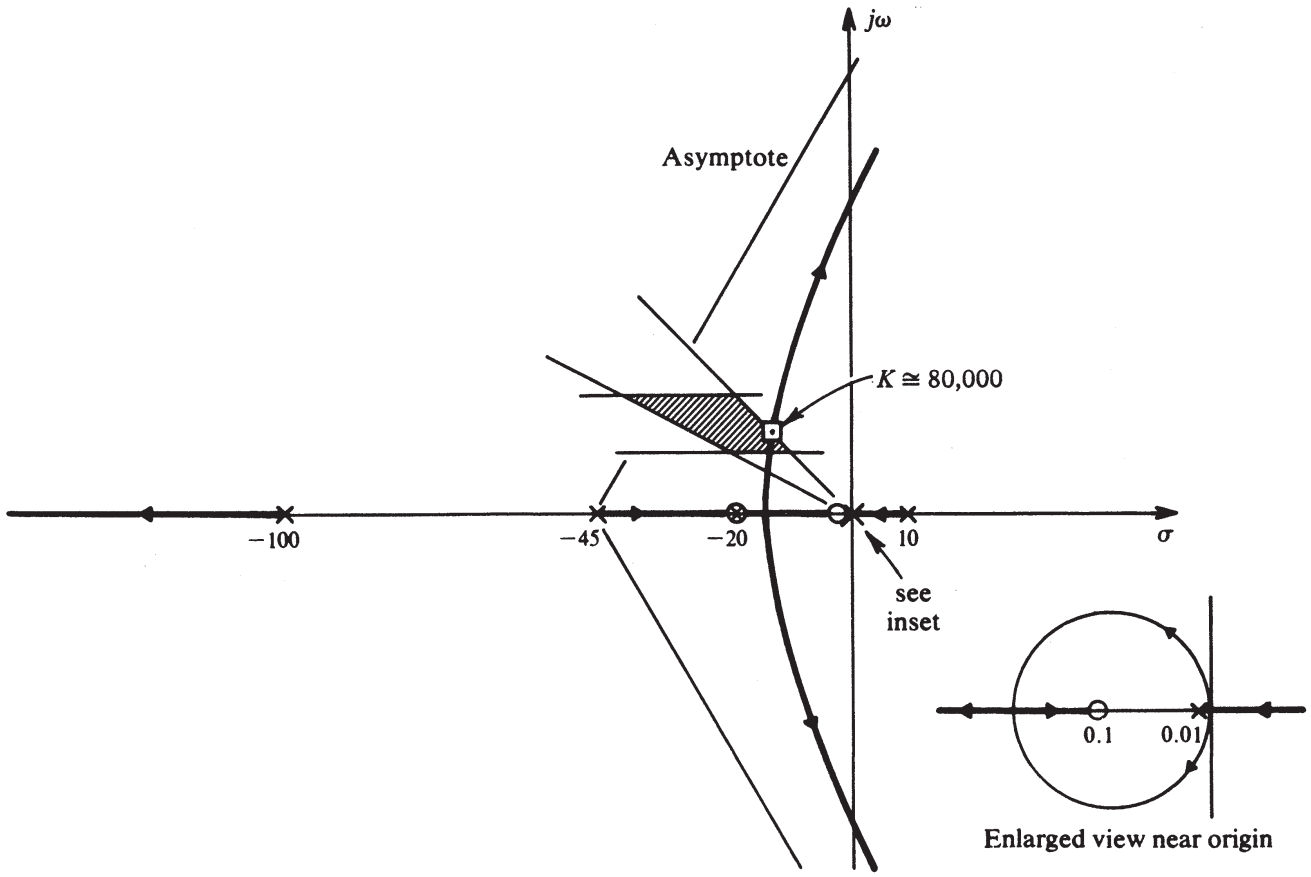


Figure 2.7

2.7 is obtained. Applying Eq. (2.3) at the point ⊡ indicates that the required gain is approximately $K = 80{,}000$. Using this result,

$$E(t)|_{ss} = \frac{1}{1 + \dfrac{K(0.1)}{(-10)(45)(100)(0.01)}} = -0.06 = 6\% \text{ error}$$ ■

EXAMPLE 2.6 Consider the error-sampled system of Figure 2.8, which represents a linearized model of a position control system using an armature controlled dc motor with a time constant $\tau = 2$ seconds.

1. Find expressions for $C(s)$, $C(z)$, and $E(z)$.
2. Show that the steady-state error is zero for a step input and approaches a constant for

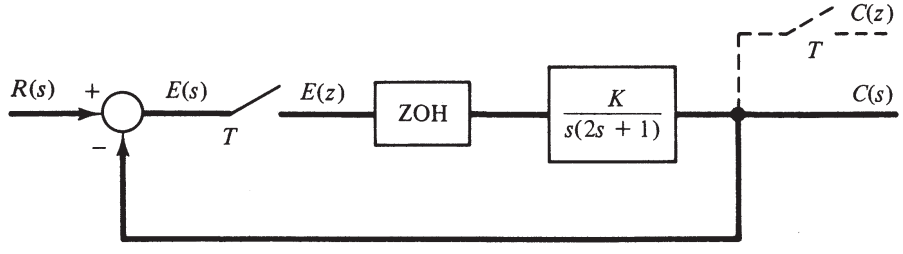


Figure 2.8

ramp inputs, with the constant decreasing inversely with K, so long as the system remains stable.

3. Show that the maximum allowable gain for stability is inversely related to the sampling period T.
4. With $T = 2$ s, find the maximum gain for stability.
5. Set the gain to $K = 0.5$ and find the steady-state error for a ramp input $R(t) = t$. The sampling period is $T = 2$.

1. *Using the two-step procedure of Section 2.3,*

$$E(s) = R(s) - E(z)(1 - z^{-1})K/[s^2(2s + 1)]$$

$$E(z) = R(z) - E(z)(1 - z^{-1})KZ\{1/[s^2(2s + 1)]\}$$

Define $(1 - z^{-1})Z\{.5K/[s^2(s + 0.5)]\} = G'(z)$. Using Z-transform tables,

$$G'(z) = \frac{K\{z[T - 2(1 - e^{-0.5T})] + [2(1 - e^{-0.5T}) - Te^{-0.5T}]\}}{(z - 1)(z - e^{-0.5T})} \tag{2.6}$$

Therefore,

$$E(z) = \frac{R(z)}{1 + G'(z)}$$

$$C(s) = E(z)(1 - z^{-1})K/[s^2(2s + 1)]$$

and

$$\begin{aligned} C(z) &= E(z)G'(z) \\ &= \frac{G'(z)R(z)}{1 + G'(z)} \end{aligned} \tag{2.7}$$

Note that the sampled signal $C(t_k)$ does not exist at any point in this system, so a fictitious sampler is added at the output to facilitate finding it. This sampler does not affect actual system operation.

2. *If $R(t)$ is a unit step, then $R(z) = z/(z - 1)$ and the final value theorem gives*

$$\lim_{k \to \infty} E(t_k) = 1/[1 + \lim_{z \to 1} G'(z)] = 0$$

since $G'(z) \to \infty$ as $z \to 1$.

If $R(t) = t$, then $R(z) = Tz/(z - 1)^2$ and the final value theorem now gives

$$\lim_{k \to \infty} E(t_k) = \lim_{z \to 1} Tz/[(z - 1)G'(z)] = T/[KT] = 1/K$$

This result holds as long as all limits exist, which requires that the system be stable.

3. *Stability requires that the roots of the characteristic equation*

$$1 + G'(z) = 0$$

be inside the unit circle. The characteristic equation can be reduced to

$$F(z) = z^2 + \alpha z + \beta = 0 \tag{2.8}$$

where the α and β coefficients are

$$\alpha = K[T - 2(1 - e^{-0.5T})] - (1 + e^{-0.5T}) \tag{2.9}$$

$$\beta = e^{-0.5T}(1 - KT) + 2K(1 - e^{-0.5T}) \tag{2.10}$$

Rather than factor this quadratic directly, use the bilinear transformation $z = (w + 1)/(w - 1)$ to find a quadratic in w:

$$[1 + \alpha + \beta]w^2 + [2(1 - \beta)]w + [1 + \beta - \alpha] = 0 \tag{2.11}$$

Applying Routh's criterion to the w quadratic shows that for stability,

$$1 + \alpha + \beta > 0, \qquad 2(1 - \beta) > 0, \quad \text{and} \quad 1 + \beta - \alpha > 0$$

are required. In terms of the original z quadratic, these requirements are

$$F(1) > 0, \qquad F(0) < 1, \quad \text{and} \quad F(-1) > 0$$

These requirements are true for any second-order characteristic equation $F(z) = 0$, and are an example of the Schur-Cohn stability test [7]. For this problem $F(1) = KT(1 - e^{-0.5T})$ is positive for all positive T and K.

The second condition states that $\beta < 1$, or

$$K[2 - 2e^{-0.5T} - Te^{-0.5T}] + e^{-0.5T} < 1 \tag{2.12}$$

If $T = 0$, then $\beta = 1$, but T will never be zero. As $T \to \infty$, $\beta \to 2K$, so this condition must be checked in detail when specific values are given in part 4.

Finally, $F(-1) > 0$ leads to

$$K < \frac{2(1 + e^{-0.5T})}{T(1 + e^{-0.5T}) - 4(1 - e^{-0.5T})} \tag{2.13}$$

For very small T this gives $K < 2/T$, and for very large T it gives $K < 2/(T - 4)$. This demonstrates the inverse relationship between K_{max} and T.

4. *With $T = 2$, the requirement of Eq. (2.12) gives $K_{max} = 1.196$, and Eq. (2.13) gives $K < 13.19$.* The most constraining result is the one which is operable.

5. *With $T = 2$*

$$G'(z) = 0.7357588K(z + 0.71828)/[(z - 1)(z - 0.36788)] \tag{2.14}$$

With $K = 0.5$, the steady-state error $\lim_{k \to 1} E(t_k) = 1/K = 2$. This motor control system will follow a commanded ramp in position, but the actual position will be offset by two units from the command. This may not be accurate enough. Even if the gain is increased to near its limit, the error only decreases to around one unit, and the transients will be very slow to die out with the system being that close to the stability limits. This system will need to have some form of compensation if the sampling time cannot be decreased. Note that if $T = 1$, then

$$G'(z) = 0.21306K(z + 0.84675)/[(z - 1)(z - 0.6065)] \tag{2.15}$$

and the maximum allowable stable gain increases to 2.18. The steady-state error due to a ramp input is still $1/K$ independent of T. ■

2.6 EXTENSION OF CLASSICAL TECHNIQUES TO MORE COMPLEX SYSTEMS

When multiple loops or multiple inputs and outputs must be considered, signal flow graph techniques can be used to reduce the problem to one of single-loop analysis. However, the resulting "open-loop" transfer function will usually not be in the convenient factored form. Even the simple techniques can become tedious in this case.

Linear multiple-input, multiple-output systems can be treated systematically using various transfer function matrix representations. Pursuing the subject in that direction quickly leads to the theory of polynomial matrices and the so-called matrix fraction description of systems. In this book, with very few exceptions, the alternate approach of using state space methods is pursued.

Simulation has long served as a supplement and extension to the classical analytical techniques, especially when dealing with complex and nonlinear systems. Analog computers were first historically, then came digital and hybrid methods. At present, the strong trend toward the digital computer continues. Computer-aided design (CAD) tools, which ease or even automate many of the tasks described in this chapter, are now widely available. Further, computers are frequently used as components in the control loop. The controllers or compensators $G_c(z)$ are routinely implemented in microprocessor form.

REFERENCES

1. Dorf, R. C.: *Modern Control Systems, 5th ed., Addison-Wesley, Reading, Mass., 1989.*
2. *Rowland, J. R.: Linear Control Systems,* John Wiley, New York, 1986.
3. Franklin, G. F., J. D. Powell, and A. Emami-Naeini: *Feedback Control of Dynamic Systems,* Addison-Wesley, Reading, Mass., 1986.
4. Kuo, B. C.: *Automatic Control Systems,* 5th ed., Prentice Hall, Englewood Cliffs, N.J., 1987.
5. D'Azzo, J. J. and C. H. Houpis: *Linear Control System Analysis and Design,* Third ed., McGraw-Hill, New York, 1981.
6. Franklin, G. F. and J. D. Powell: *Digital Control of Dynamic Systems,* Addison-Wesley, Reading, Mass., 1980.
7. Kuo, B. C.: *Analysis and Synthesis of Sampled-Data Control Systems,* Prentice Hall, Englewood Cliffs, N.J., 1963.
8. Cruz, J. B., Jr. and W. R. Perkins: "A New Approach to the Sensitivity Problem in Multivariable Feedback Systems," *IEEE Transaction on Automatic Control,* Vol. AC-9, July 1964, pp. 216–223.
9. Cruz, J. B., Jr., Ed.: *Systems Sensitivity Analysis,* Dowden, Hutchinson and Ross, Stroudsburg, Penn., 1973.
10. Truxal, J. G.: *Automatic Control System Synthesis,* McGraw-Hill, New York, 1955.

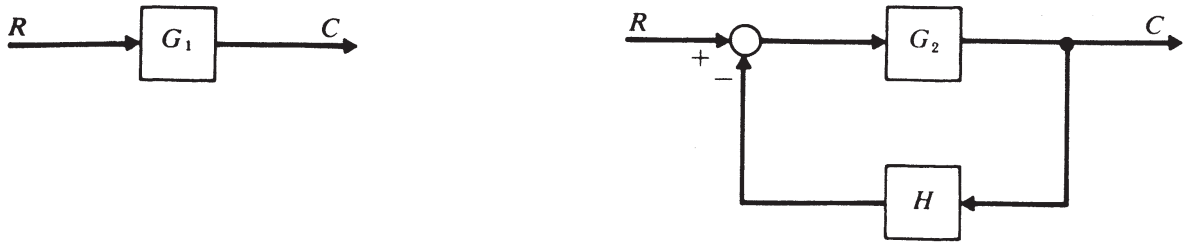
ILLUSTRATIVE PROBLEMS

Properties of Feedback

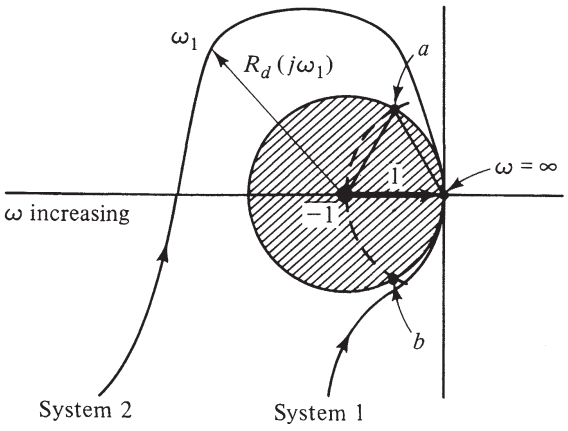
2.1 Compare the open-loop and feedback control systems of Figure 2.9 in terms of the sensitivity of the output C to variations in system parameters.

In the open-loop case $C = G_1 R$ and $\partial C/\partial G_1 = R$, so that a change δG produces a change in the output $\delta C = R\delta G$. System sensitivity S is defined as percentage change in C/R divided by the percentage change in the process transfer function. For the open-loop system,

$$S = \frac{\delta G/G}{\delta G/G} = 1$$



(*a*) Open loop (*b*) Closed loop



(c)

Figure 2.9

For the closed-loop feedback system $C = G_2R/(1 + G_2H)$,

$$\delta G_2 \frac{\partial C}{\partial G_2} = \frac{R\delta G_2}{1 + G_2H} - \frac{G_2RH\delta G_2}{(1 + G_2H)^2} = \frac{R\delta G_2}{(1 + G_2H)^2}$$

The sensitivity is

$$S = \frac{\delta(C/R)}{C/R}\frac{G_2}{\delta G_2} = \frac{\delta G_2}{(1 + G_2H)^2}\frac{(1 + G_2H)}{G_2}\frac{G_2}{\delta G_2} = \frac{1}{1 + G_2H} = \frac{1}{R_d}$$

If the magnitude of the return difference is greater than unity at all frequencies ω, i.e.,

$$|R_d(j\omega)| > 1 \tag{1}$$

then the closed-loop configuration (*b*) is always less sensitive to parameter variations than the open-loop configuration (*a*). Equation (*1*) requires that the polar plot of $G_2(j\omega)H(j\omega)$ never enters the unit-radius disk centered at −1. Figure 2.9*c* shows polar plots for two possible systems which satisfy this requirement. System 1 has an infinite gain margin (increase or decrease). With system 2 the gain can be increased an infinite amount or decreased by at least 50% while remaining stable. Consideration of the equilateral triangles formed by points (*a*, −1, 0) or (*b*, −1, 0) make it clear that systems which do not penetrate the unit disk also have phase margins of at least 60°. Certain linear-quadratic optimal systems are guaranteed to have these desirable properties, as is discussed in Sec. 14.7. For other systems, the design goal is often to shape the polar plot and/or use a high loop gain in order to satisfy Eq. (*1*) at all frequencies within the system bandwidth but perhaps not at all frequencies. Satisfactory reduction of sensitivity to the most problematic parameter variations is thus achieved.

2.2 Show how precise feedback coefficients can give precision feedback control even if gross errors exist in the forward loop system being controlled.

As the loop gain increases, the feedback transfer function becomes

$$\frac{C}{R} = \frac{KG}{1 + KGH} \to \frac{1}{H}$$

Thus for very high loop gain the response is largely determined by H rather than G.

2.3 Show that the sensitivity of the output to errors in H approaches unity for high loop gain.

This could easily be shown by starting with the result of the last problem. Alternatively, define sensitivity to H as

$$S_H = \frac{\delta(C/R)/(C/R)}{\delta H/H} = \frac{-G^2 \delta H}{(1 + GH)^2} \frac{1 + GH}{G} \frac{H}{\delta H} = -\frac{GH}{1 + GH} \to -1 \quad \text{as} \quad |GH| \to \infty$$

This indicates the need for precision components in the feedback loop.

2.4 Compare performance of the systems shown in Figure 2.10*a,b* as degraded by the unwanted disturbance D.

In the open-loop case, $C(s) = KG_1 G_2 R(s) + G_2 D(s)$.

In the feedback case, $C(s) = \dfrac{KG_1 G_2 R(s)}{1 + KG_1 G_2 H} + \dfrac{G_2 D(s)}{1 + KG_1 G_2 H}$.

In the second case the contribution of the disturbance to the output can be made small by increasing the gain K. More generally, this is accomplished by increasing the return difference magnitude over the frequency range of interest, and this should be done by increasing the magnitude of KG_1. Notice that feedback also introduces a loss of useful gain between C and R. For example, if $H = 1$ and G_1 and G_2 are ideal amplifiers with constant gains, the open-loop system gain is $KG_1 G_2$. The feedback system gain $KG_1 G_2/(1 + KG_1 G_2)$ would then be less than unity.

2.5 Use the dc motor of Problem 1.2, page 17, to demonstrate how feedback can favorably improve transient response. Neglect the inductance L.

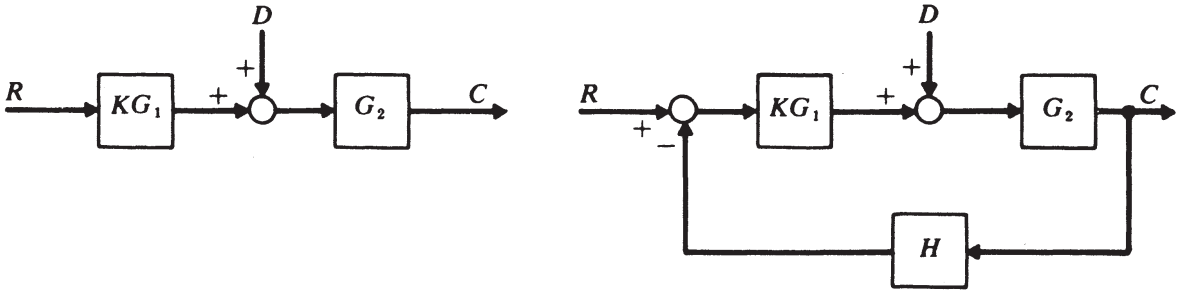
When $L = 0$, the transfer function can be written as

$$\frac{\Omega(s)}{V(s)} = \frac{K/JR}{s + (bR + K^2)/JR}$$

This is considered as the open-loop system. If a step voltage $V(s) = V/s$ is applied, the open-loop speed response is, letting $K' = K/JR$ and $a = (bR + K^2)/JR$,

$$\Omega(s) = \frac{VK'}{s(s + a)} \quad \text{or} \quad \omega(t) = \frac{VK'}{a}(1 - e^{-at})$$

The speed of response is determined by a, which is determined by the load and motor characteristics. If response is too slow for a given load, a new motor with a larger value of K could be



(*a*) Open loop (*b*) Feedback

Figure 2.10

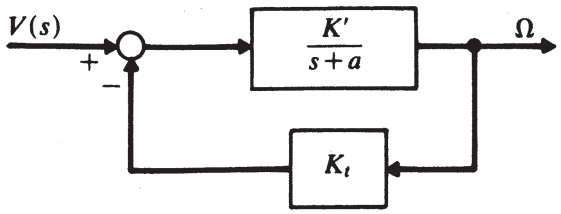


Figure 2.11

installed. Alternatively, consider the tachometer feedback system of Figure 2.11. Here $\Omega(s) = [K'/(s + a + K'K_t)]V(s)$. The system response time is now determined by $a' = a + K'K_t$ and can obviously be improved by proper choice of K_t.

Routh's Criterion

2.6 Use Routh's criterion to determine the number of roots of

$$s^5 + 5s^4 - 2s^3 + 8s^2 + 10s + 3 = 0$$

which have positive real parts.

The Routhian array is given in Table 2.2

Note that any row can be normalized by multiplying or dividing by a positive constant, as in the s^3 row. For the terms in column 1, the sign changes from row s^4 to row s^3 and again from row s^3 to row s^2. Two sign changes indicate there are two right-half-plane roots to this equation.

2.7 Does the following equation have any roots in the right-half plane?

$$s^4 + 2s^3 + 4s^2 + 8s + \alpha = 0$$

The Routhian array is shown in Table 2.3.

Note that the leading term in the s^2 row is zero. Whenever this happens, the zero is replaced by a small number ϵ and the rest of the array is computed as usual. The limiting behavior as $\epsilon \to 0$ is used to determine stability. Here the first column reduces to $\{1, 2, 0, \lim_{\epsilon \to 0}(-2\alpha/\epsilon), \alpha\}$. If $\alpha < 0$, there is just one sign change between the s and s^0 rows, and, therefore, just one right-half-plane root. If $\alpha > 0$, there are two sign changes and two right-half-plane roots.

Steady-State Error and Error Constants

2.8 Derive expressions for the steady-state value of $E(t)$ for the system of Figure 2.1*a* when the input $R(s)$ is a step, ramp, and parabolic function, respectively.

The Laplace transform of the error is $E(s) = R(s)/[1 + KG(s)H(s)]$. For a unit step, $R(s) = 1/s$. Using the final value theorem and assuming a constant steady-state error exists,

TABLE 2.2

s^5	1	−2	10
s^4	5	8	3
s^3	$\frac{5(-2) - (1)(8)}{5} = -18/5 \sim -18$	$\frac{5(10) - 1(3)}{5} = \frac{47}{5} \sim 47$	—
s^2	$\frac{-18(8) - 5(47)}{-18} = \frac{379}{18}$	3	—
s	$\frac{(379/18)(47) - (-18)(3)}{379/18} = \frac{18785}{379}$	—	
s^0	3		

TABLE 2.3

s^4	1	4	α
s^3	2	8	
s^2	$\frac{2(4)-8}{2}=0^{\epsilon}$	α	
s	$\frac{8\epsilon-2\alpha}{\epsilon}$		
s^0	α		

$E_{ss} \triangleq \lim_{t\to\infty}\{E(t)\} = 1/[1 + K\lim_{s\to 0}\{GH\}]$. Similarly, for a ramp input, $R(s) = 1/s^2$ and $E_{ss} = 1/K\lim_{s\to 0}\{sGH\}$. If the input is the parabola $t^2/2$, $R(s) = 1/s^3$ and $E_{ss} = 1/K\lim_{s\to 0}\{s^2GH\}$.

To proceed, we must know the system type. The system type is the number of s terms that factor out of the denominator of $G(s)H(s)$. For a type 0 system there are no such factors, so $K\lim_{s\to 0}\{GH\} = K_b$, the Bode gain. Likewise, for type 0, $K\lim_{s\to 0}\{sGH\} = 0$ and $K\lim_{s\to 0}\{s^2GH\} = 0$. For type 1 systems $K\lim_{s\to 0}\{GH\} = \infty$, $K\lim_{s\to 0}\{sGH\} = K_b$ and $K\lim_{s\to 0}\{s^2GH\} = 0$. Similar results hold for type 2 and higher systems. The three limiting values for each system are called the position, velocity, and acceleration *error constants* K_p, K_v, and K_a. These are summarized in Table 2.4, along with the steady-state error values. Note that larger error constants give smaller steady-state error.

Miscellaneous Methods

2.9 Sketch the root locus for a system with

$$KG(s) = \frac{K}{s(s+8)(s^2+8s+32)}, \qquad H(s) = s + 4$$

The open-loop poles, plotted as $\times$, are located at $s = 0$, -8, $-4 + 4j$, and $-4 - 4j$. The open-loop zero $\odot$ is at $s = -4$. There are four branches of the loci, and Rule 3 gives the real axis portion, shown in Figure 2.12. Three branches must approach $s = \infty$ as $K \to \infty$. One is on the negative real axis, and the others are at $\pm 60°$, since

$$\phi_i = \frac{(1+2k)180°}{3} = (1+2k)60°$$

$$= +60 \quad (k=0), \qquad -60 \quad (k=-1), \quad \text{and} \quad 180° \quad (k=1)$$

Other values of k give multiples of the same three asymptotic angles. Rule 6 gives

$$\text{cg} = \tfrac{1}{3}[(0-8-4-4)-(-4)] = -4$$

TABLE 2.4

	Error constants			Steady-state error		
System type	K_p	K_v	K_a	Step input	Ramp input	Parabolic input
0	K_b	0	0	$1/(1+K_b)$	∞	∞
1	∞	K_b	0	0	$1/K_b$	∞
2	∞	∞	K_b	0	0	$1/K_b$

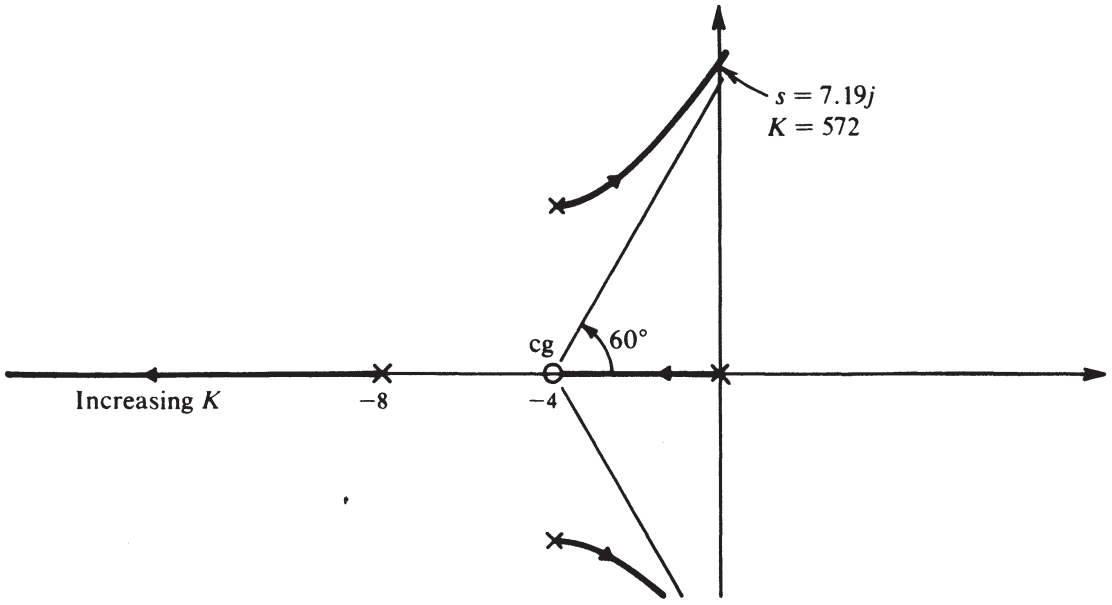


Figure 2.12

In general, the phase angle of the transfer function, equation (*2.2*), can be written as

$$\angle GH(s) = \phi_{z_1} + \phi_{z_2} + \cdots - \phi_{p_1} - \phi_{p_2} = \cdots = (1 + 2k)180°$$

where ϕ_{z_i} is the angle of the line segment from the zero z_i to a point s and ϕ_{p_i} is the angle from pole p_i to s. In order to determine the angle of departure of the locus from a complex pole, a test point s_1 is used, which is infinitesimally close to the pole. The angles of the vectors from all zeros and poles except one are easily measured. The remaining angle, associated with that complex pole, can be computed. This is the angle of departure and in this case it is 0°. With this information, a few more test points allow an accurate sketch of the complete locus.

2.10 For the system shown in Figure 2.13, select K so that the phase margin is greater than 30° and the gain margin is greater than 10 db.

In Bode form,

$$KGH = \frac{50K}{200\,s} \frac{(0.02s + 1)}{(0.1s + 1)(0.05s + 1)}$$

The Bode plots are drawn in Figure 2.14 with the Bode gain $K_b = K/4$ set to unity. At $\omega = 10$, the phase is −150° and the gain is −24 db. This means that K_b could be increased from 0 db to +24 db, and the phase margin specification would just be satisfied. If this were done, then at $\omega = 24$ rad/s, where the phase is −180°, the gain would increase from −39 db to −15 db. This gain margin of 15 db satisfies the specifications, so $K_b = 24$ db, which converts to a real gain of about 15. This means that $K = 4K_b = 60$ can be used to satisfy both specifications.

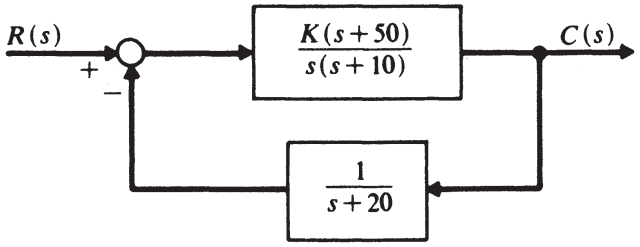


Figure 2.13

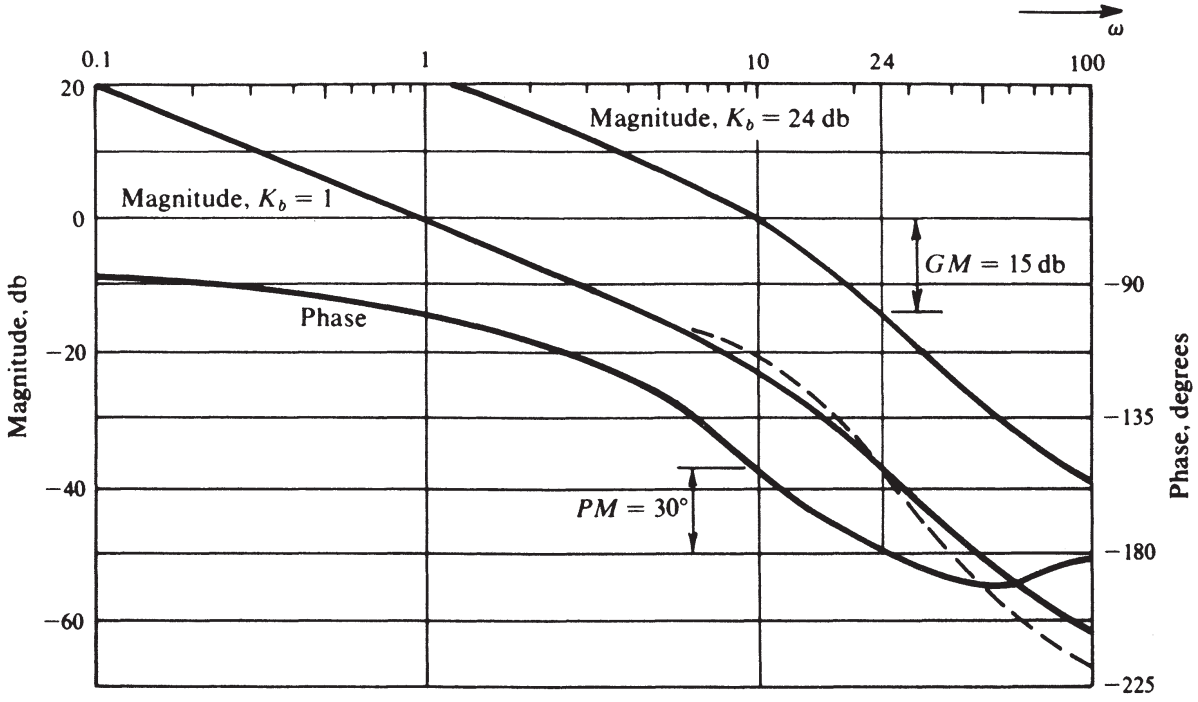


Figure 2.14

2.11 Use root locus to determine the closed-loop poles of Problem 2.10 when $K = 60$ is used.

The upper half of the locus is sketched in Figure 2.15 using the rules of Example 2.4 and a spirule to check Eq. (*2.2*) at a few additional points. The closed-loop poles are shown as ⊡. Notice that the complex poles lie almost exactly on the $\zeta = 0.3$ damping line. For this example, the rule of thumb $\zeta = 0.01\ PM$ is verified.

Polar Plots and Nyquist's Criterion

2.12 What information is readily available from a polar plot of $KG(j\omega)H(j\omega)$?

The number of closed-loop poles in the right half plane can be determined in terms of the number of encirclements of -1, using Nyquist's criterion. The system type is indicated, assuming a minimum phase system, by the phase angle at $\omega = 0$. Type 0 systems have a finite magnitude and zero phase angle. Type 1 systems approach infinite magnitude at an angle of $-90°$, type 2 systems aproach infinite magnitude at an angle of $-180°$, etc. The excess of open-loop poles compared to zeros is indicated by the behavior as $\omega \to \infty$. If there is an equal number of poles and zeros, the magnitude approaches a finite constant. In all other cases the magnitude approaches zero, but if there is one more pole than zero, the approach is along the $-90°$ axis. For two more poles than zeros, it is along the $-180°$ axis, etc. Relative stability, in terms of gain and phase margins, is also readily apparent. For example, in the plot shown in Figure 2.16 the system is type 1, and it has three more poles than zeros. Assuming no open-loop, right-half-plane poles, the system is stable, since the plot does not encircle the point -1. The phase margin is 60° and the gain margin is 1.25, since the gain could be increased by that factor without causing the plot to encircle the -1 point.

2.13 Draw the polar plot for

$$KGH = \frac{K(s + 0.5)(s + 10)}{(s + 1)(s + 2)(s^2 + 2s + 5)}$$

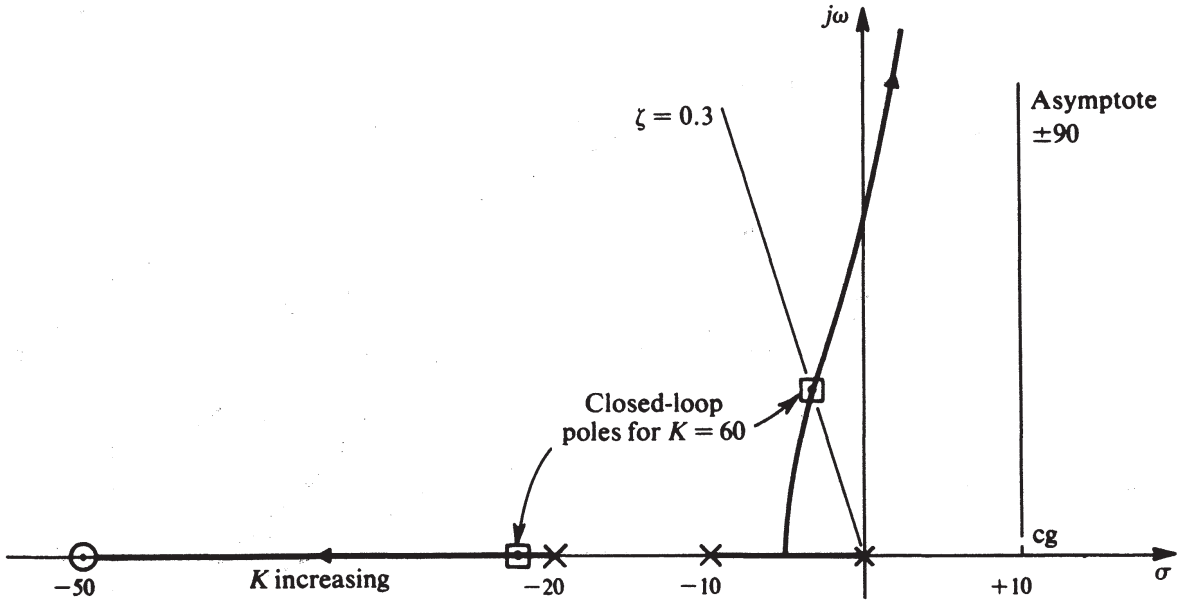


Figure 2.15

Determine the gain and phase margins when $K = 2$.

The polar plot for this type 0 system is given in Figure 2.17. The -1 point is not encircled, so $N = 0$. Since there are no open-loop, right-half-plane poles, $P_R = 0$. Therefore, Nyquist's criterion indicates that there are no unstable closed-loop poles. The phase margin is approximately 50°. The magnitude at 180° phase is 0.32, so the gain margin is $1/0.32 \cong 3.1$.

The smallest magnitude of the return difference, which occurs here at approximately $\omega = 3$, is an excellent measure of stability margins. This is the basis for the constant M-circle concepts in classical control analysis. The larger the minimum return difference is, the more robust the system is to modeling errors and parameter variations, which may cause gain changes or phase shifts.

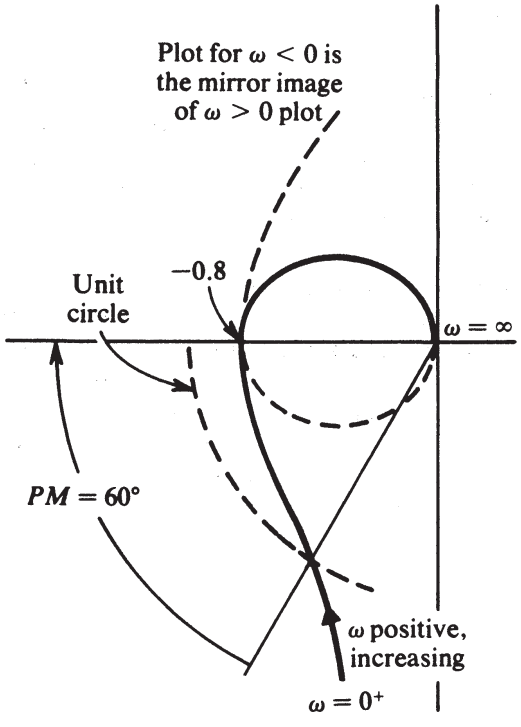


Figure 2.16

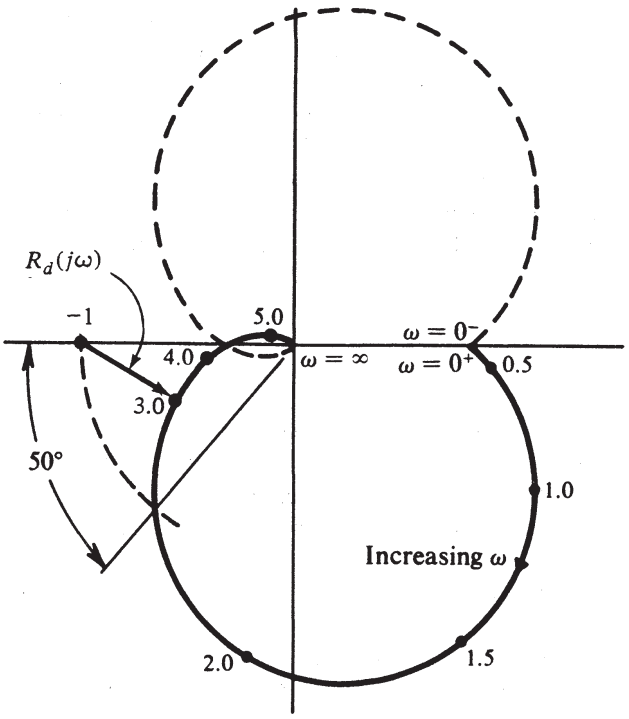


Figure 2.17

2.14 Analyze the polar plot for the nonminimum phase system with

$$KGH = \frac{K(s+10)(s+20)}{s(s-10)(s+40)}$$

Even though this is a type 1 system, the phase angle approaches $-270°$ as $\omega \to 0$ since the minus sign on the unstable pole contributes $-180°$ phase. The general shape of the polar plot is shown in Figure 2.18*a*.

The -1 point could be encircled by either the *a-d-c-b-a* circuit, counterclockwise, or by

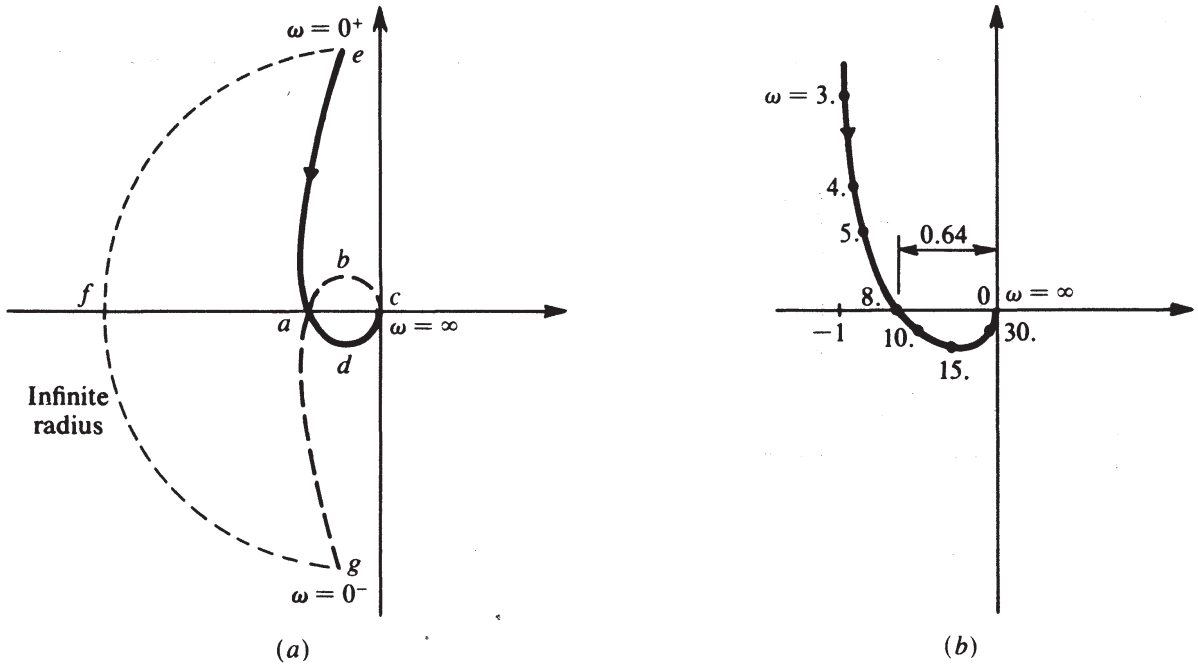


Figure 2.18

the infinite *a-g-f-e-a* circuit, clockwise. Which circumstance prevails depends upon K. If -1 is inside the small circuit, then $N = 1$. Since $P_R = 1$, there would be $Z_R = 1 - 1 = 0$ unstable closed-loop poles. If -1 is to the left of point *a,* then there is one clockwise encirclement so $N = -1$ and, therefore, $Z_R = 1 - (-1) = 2$, indicating two unstable closed-loop poles. An enlarged plot is also given in Figure 2.18*b* for $K = 10$. The magnitude of 0.64 at $-180°$ indicates that the system is unstable for values of $K < 10/0.64 \cong 15.6$, and stable if $K > 15.6$. This problem illustrates that the usual visualization of gain and phase margin is incorrect for nonminimum phase systems. This is why Bode plots should not be used with nonminimum phase transfer functions.

Compensation

2.15 Explain the essence of classical compensation using root locus techniques.

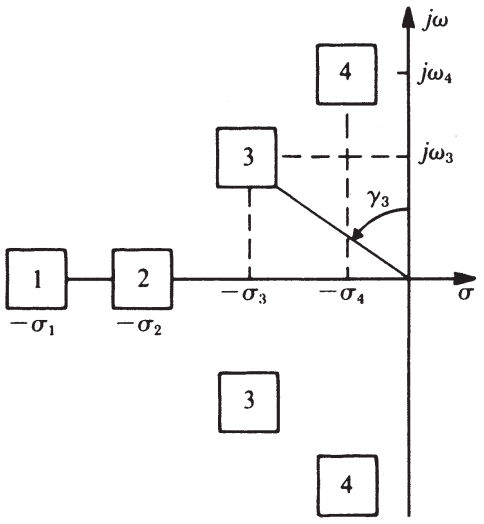
The basis for compensation is a knowledge of the correspondence between closed-loop pole locations and the type of transient time response terms they yield. Some typical closed-loop pole locations are indicated by □ in Figure 2.19, and the corresponding time response terms are shown.

Poles farther to the left of the imaginary axis give terms which die out faster, i.e., faster response times. Complex poles give oscillating terms with a frequency equal to the distance from the real axis and decay time inversely proportional to the real part of the pole, σ. The damping ratio, related to overshoot, is defined in terms of the angle γ as $\zeta = \sin\gamma$. The undamped natural frequency ω_n is the radial distance from the origin, so that $\sigma_i = \zeta\omega_n$ and the damped frequency is $\omega_i = \omega_n\sqrt{1-\zeta^2}$.

Based on these relations, the desired locations of the most dominant closed-loop poles are selected. Checking the root locus angle criterion at that point indicates whether additional lag or lead is needed and how much. Compensating poles and zeros are then selected to provide this phase shift.

Suppose a damping ratio of $\zeta = 0.707$ and a frequency of 10 rad/s are desired. Then a pair of complex closed-loop poles must be located as shown in Figure 2.20. If the angle for the uncompensated *KGH* is $-160°$ at that point, then an additional $-20°$ phase must be provided by compensation. It is common practice to place the compensator zero directly below the desired closed-loop pole. Obviously, a single pole zero pair can provide up to 60–65° phase shift. If a greater shift is needed, more complicated compensation is required.

2.16 Discuss the compensation characteristics of the phase-lag and phase-lead circuits of Figure 2.21 using Bode plots.



1. $c_1 e^{-\sigma_1 t}$
2. $c_2 e^{-\sigma_2 t}$
3. $c_3 e^{-\sigma_3 t}\sin(\omega_3 t + \phi_3)$
4. $c_4 e^{-\sigma_4 t}\sin(\omega_4 t + \phi_4)$

Figure 2.19

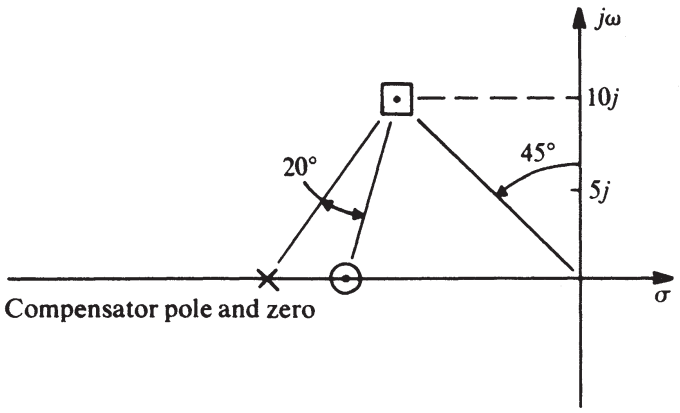


Figure 2.20

The lag circuit lowers the gain at high frequencies and leaves the phase unchanged for $\omega >> 1/\tau_1$ (Figure 2.22*a*). The corner frequencies $1/\tau_1$ and $1/\tau_2$ would be chosen well below the critical crossover frequency. This means the dc gain can be raised in order to improve steady-state accuracy. This increase in gain returns the high frequency gain to its uncompensated level and thus leaves the stability margins relatively unchanged.

The lead circuit Bode plot assumes an additional dc gain has been added to compensate for the τ_2/τ_1 reduction. By proper choice of $1/\tau_1$ and $1/\tau_2$ relative to the cross-over frequencies, the phase lead can be used to increase phase margin. This circuit also delays the 0 db crossover frequency, thus increasing bandwidth and speed of response. It leaves the low frequency characteristics, such as steady-state error, unchanged.

Discrete-Time Problems

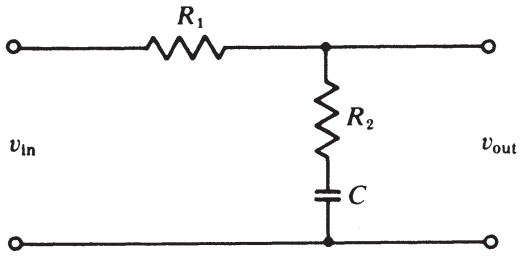
2.17 In the process of designing a digital control system, it was determined that the following compensator was desired:

$$G_c(z) = \frac{10(z + 0.4)(z - 0.5)}{(z + 0.5)(z + 1)}$$

Determine an algorithm which can be coded on the computer to implement $G_c(z)$.

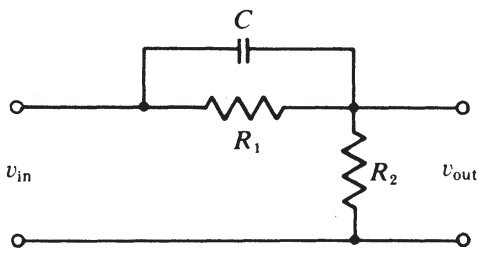
There are several possible answers, two of which follow. Writing G_c in expanded polynomial form gives

$$G_c(z) = \frac{10(z^2 - 0.1z - 0.2)}{z^2 + 1.5z + 0.5} = \frac{10(1 - 0.1z^{-1} - 0.2z^{-2})}{1 + 1.5z^{-1} + 0.5z^{-2}}$$



$$\frac{V_{out}}{V_{in}} = \frac{1 + \tau_1 s}{1 + \tau_2 s} \qquad \tau_2 > \tau_1$$

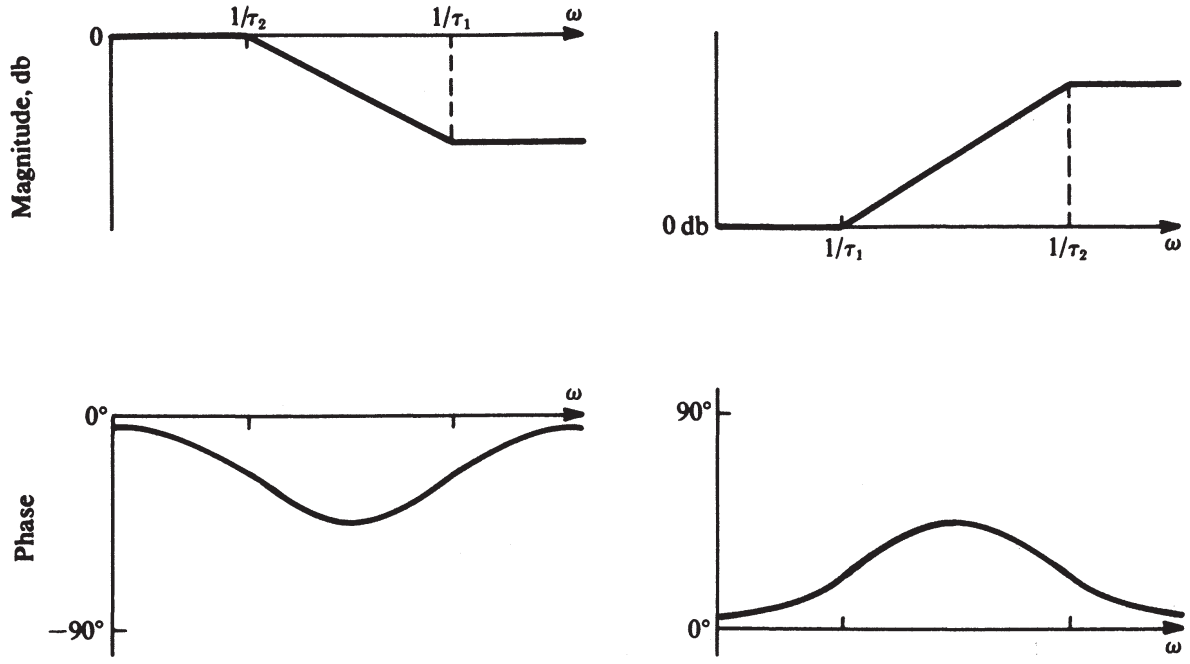
(*a*) Phase-lag circuit



$$\frac{V_{out}}{V_{in}} = \left(\frac{\tau_2}{\tau_1}\right)\frac{1 + \tau_1 s}{1 + \tau_2 s} \qquad \tau_1 > \tau_2$$

(*b*) Phase-lead circuit

Figure 2.21



(*a*) Phase-lag circuit (*b*) Phase-lead circuit

Figure 2.22

If the input to G_c is $E(t_k)$ and the output is $y(t_k)$, then using z^{-1} as the delay operator, cross multiplication gives the so-called direct form realization

$$(1 + 1.5z^{-1} + 0.5z^{-2})Y(z) = (10 - z^{-1} - 2z^{-2})E(z)$$

or, in the time domain

$$y(t_k) = 10E(t_k) - E(t_{k-1}) - 2E(t_{k-2}) - 1.5y(t_{k-1}) - 0.5y(t_{k-2})$$

If $G_c(z)/z$ is expanded in partial fractions and the result is then multiplied by z, one obtains

$$G_c(z) = -4 - 4z/(z + 0.5) + 18z/(z + 1)$$

This represents three separate paths through the compensator, as shown in Figure 2.23. This is the so-called parallel realization.

The algorithm thus consists of

$$y_1(t_k) = 4E(t_k)$$

$$y_2(t_k) = E(t_k) - 0.5y_2(t_{k-1})$$

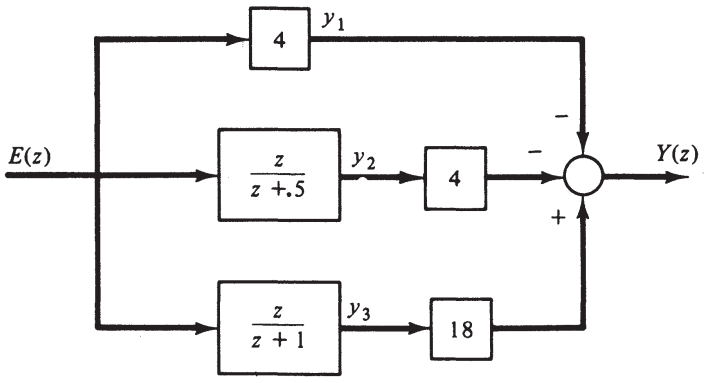


Figure 2.23

$$y_3(t_k) = E(t_k) - y_3(t_{k-1})$$

and

$$y(t_k) = 18y_3(t_k) - y_1(t_k) - 4y_2(t_k)$$

Other forms are possible.

2.18 Derive the transfer function for the device which outputs the piecewise constant approximation to a continuous-time function $E(t)$ as shown in Figure 2.24.

Let the unit step function starting at time $t = 0$ be $u(t)$. Then a shifted unit step function starting at time t_k is $u(t - t_k)$. The piecewise constant function can be written as

$$\begin{aligned} Y^*(t) &= E(0)u(t) + [E(1) - E(0)]u(t - t_1) + [E(2) - E(1)]u(t - t_2) + \cdots \\ &\quad + [E(t_k) - E(t_{k-1})]u(t - t_k) + \cdots \\ &= \sum_{k=0}^{\infty} [E(t_k) - E(t_{k-1})]u(t - t_k) \end{aligned}$$

Since $du(t - t_k)/dt = \delta(t - t_k)$, the derivative of the device output is

$$x(t) = dY^*(t)/dt = \sum_{k=0}^{\infty} [E(t_k) - E(t_{k-1})]\delta(t - t_k)$$

Using the definition of the Z-transform in Problem 1.15, the transform of $x(t)$ is $X(z) = E(z) - z^{-1}E(z) = (1 - z^{-1})E(z)$. The final relationship among these variables is shown in Figure 2.25.

Since $y(t)$ is the integral of $x(t)$, $Y(s) = [(1 - z^{-1})/s]E(z)$. Clearly, then, the transfer function of the zero order hold is $G_0(s) = (1 - z^{-1})/s$.

2.19 Investigate methods of obtaining a Z-transfer function $G'(z)$ which has approximately the same behavior as an s-transfer function $G(s)$. There are several approaches to this question.

(a) One way is to determine the exact Z-transform that corresponds to the product of $G(s)$ and the ZOH transfer function $G_0(s)$, perhaps using transform tables. The ZOH should be included in most cases involving combined continuous and discrete systems, as shown in Figure 2.1*b*. The reason is that Z-transforms are always applied to everything between two samplers (or the same sampler around a complete loop). For most physical systems of interest, a D/A will be included in this segment of the system.

(b) An approximate conversion is obtained by treating s as a derivative operator and using a forward finite difference approximation on the sampled signal:

$$\dot{y}(t) \cong [y(t_{k+1}) - y(t_k)]/T$$

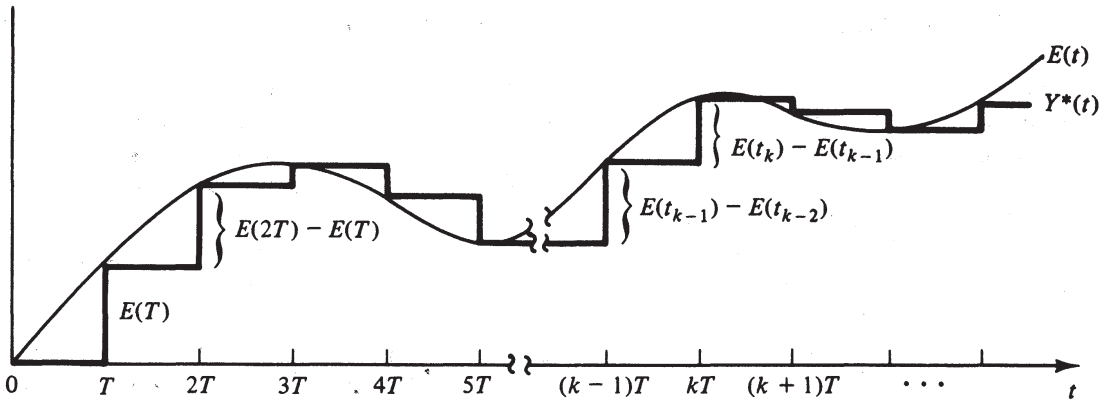


Figure 2.24

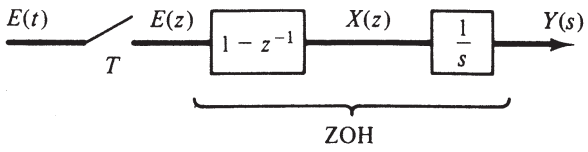


Figure 2.25

In the transform domain this means that $s \cong (z-1)/T$, since z is the advance operator. If this approximation for s is used in $G(s)$, the approximate transfer function $G_b'(z)$ is obtained. Note that this result is also given directly from $z = e^{Ts} \cong 1 + Ts$, which is approximately true for small Ts.

(c) A backward difference approximation is also possible. $\dot{y}(t) \cong [y(t_k) - y(t_{k-1})]/T$ leads to $s \cong (1 - z^{-1})/T$ and to $G_c'(z)$. This result is also obtainable from $z = e^{Ts} = 1/e^{-Ts} \cong 1/[1 - Ts]$.

(d) If z is written as $z = e^{Ts/2}/e^{-Ts/2} \cong [1 + Ts/2]/[1 - Ts/2]$, then s can be solved for as $s \cong (2/T)[z-1]/[z+1]$. Using this gives $G_d'(z)$.

All the above results can also be derived by approximating the integration operator instead of the derivative operator [6].

2.20 Apply the above techniques to determine discrete approximations for $G(s) = 1/(s+a)$.

(a.1) $Z\{(1 - z^{-1})G(s)/s\} = [1 - e^{-aT}]/[a(z - e^{-aT})] = G_a'(z)$

Note that the pole is the exact s- to z-plane mapping of $s = -a$, so stability properties are preserved.

(a.2) For comparison, the transform without the zero order hold is $Z\{G(s)\} = z/[z - e^{-aT}]$. Even though the denominators are the same, there is a one-period delay difference and a gain difference. Unless one is working with true pulsed circuits, form a.1 is the appropriate one to use.

(b) With $s \cong (z-1)/T$, $G_b'(z) = T/[z - 1 + aT]$. The z-plane pole can be in the unstable region (outside the unit circle) even if the s-plane pole is stable.

(c) With $s \cong (z-1)/(Tz)$, $G_c'(z) = [T/(1+aT)]z/\{z - [1/(1+aT)]\}$. Here the z-plane pole is always stable (inside the unit circle) whenever the s-plane pole is stable (and sometimes even when the s-plane pole is unstable).

(d) With $s \cong (2/T)[z-1]/[z+1]$, $G_d'(z) = [T/(aT+1)]z/[z - 1/(1+aT)]$. In this case the z-plane pole is inside the unit circle if $a < 0$, is on the unit circle if $a = 0$, and is outside if $a > 0$. Thus $G_d'(z)$ inherits the exact stability properties of $G(s)$. It can be shown that this is true for all transfer functions formed using approximation (d). That is,

$$z = [1 + Ts/2]/[1 - Ts/2]$$

exactly maps the left-hand s-plane into the interior of the unit circle. Approximations (a) and (d) both preserve stability properties, while (b) and (c) do not. However, the behavior in (c) is preferable to that of (b) in this regard.

2.21 By using the mapping $z = e^{Ts}$, determine how the pole locations discussed in Problem 2.15 map into the Z-plane.

For poles ① and ② or any other on the negative real axis, $z = e^{-\sigma T}$ is a positive number between 0 and 1. Therefore, Z-plane poles on the positive real axis correspond to exponential time functions. These are stable (decaying) exponentials for poles inside the unit circle. Poles on the positive real axis in the s-plane also give positive real axis Z-plane poles that are outside the unit circle. These correspond to unstable (growing) exponentials.

Poles like ③ and ④ give $z = e^{-\sigma T}e^{j\omega T} = e^{-\sigma T}[\cos(\omega T) \pm j \sin(\omega T)]$. These will always occur in conjugate pairs, and are on the unit circle if $\sigma = 0$ and give persistent oscillations. If $\sigma > 0$, they are inside the unit circle and give damped oscillations. For $\sigma < 0$, the poles are outside the unit circle and correspond to growing oscillations. The decay or growth rate depends directly on the radial distance of the poles from the unit circle. The frequency of oscillation is directly related to their angular position ωT. Note that if $\omega T = \pi$, the poles are on the negative real axis. No s-plane pole with $\omega > \pi/T$ will occur if the sampling rate satisfies the Nyquist

sampling theorem. If such higher frequency poles do occur, they map into the Z-plane at points which also correspond to lower frequency poles. This ambiguity is called aliasing, and can be avoided by sampling at least twice per period of the highest frequency pole in the system.

As any s-plane pole's real part approaches $-\infty$, the Z-plane pole approaches the origin. Poles at $s = 0$ map into $z = 1$.

It is informative to consider several familiar contours in the s-plane and see how they map into the Z-plane. In Figure 2.26*a*, a closed contour is considered, with arrows indicating the traverse direction and numbered points showing the correspondence at eight key points. In Figure 2.26*b*, lines of constant frequency, constant settling time, and lines of constant damping ratio are shown for the s- and Z-planes.

2.22 The dc motor controller of Example 2.6 is to be designed using root locus. The sampling period is $T = 1$ s. Closed-loop Z-plane poles at $z = 0.19877 \pm 0.30956j$ are desired. These are the images of $s = -1 \pm j$. They are selected because of the desirable properties they give to continuous-time systems, namely, a settling time of about 4 seconds and the damping ratio of 0.7, which gives about 5% overshoot. The open-loop transfer function is given in Eq. (*2.15*).

The uncompensated root locus is sketched in Figure 2.27*a*. To achieve the desired root locations, compensation is required to reshape the locus. A forward-path cascade

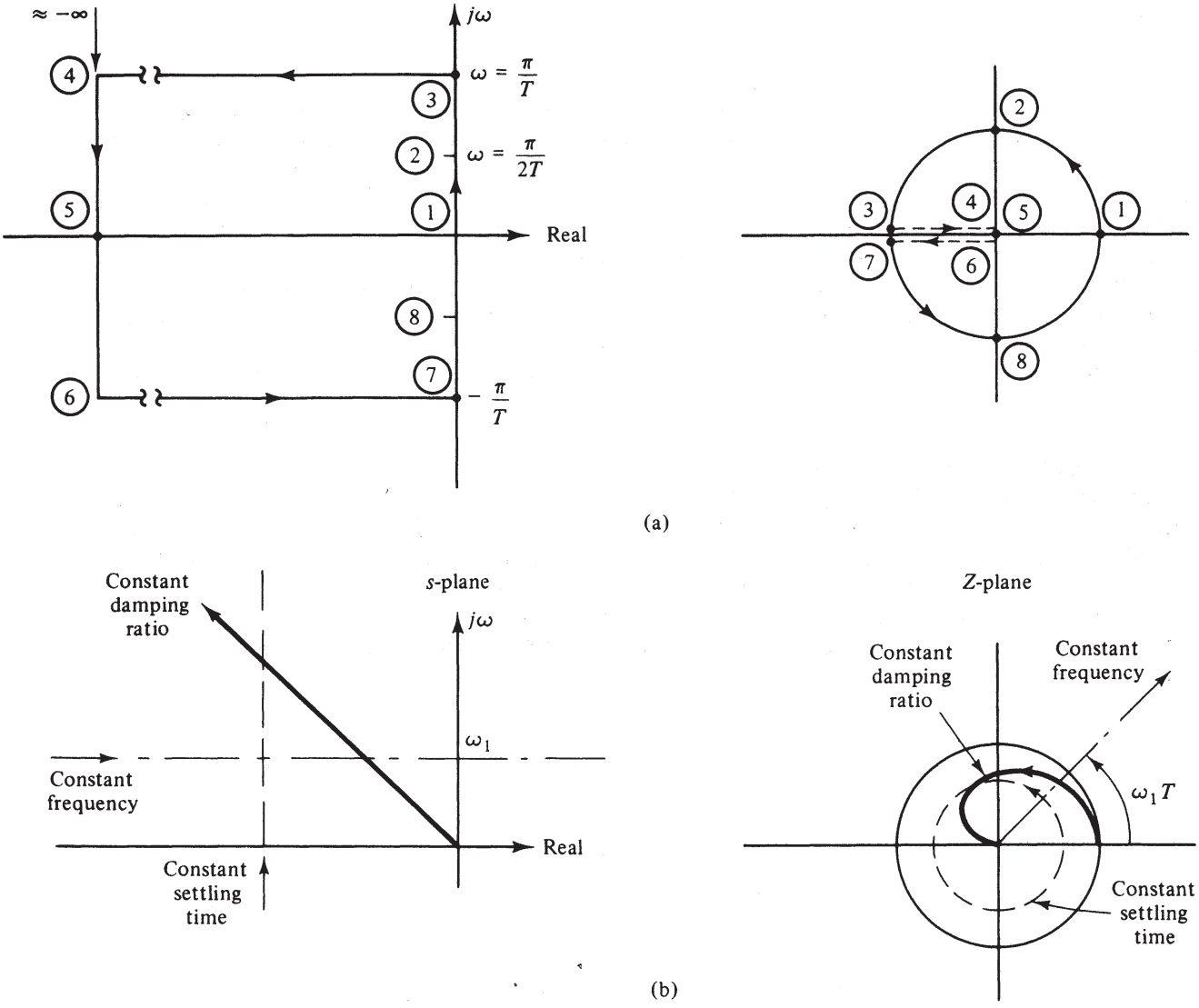
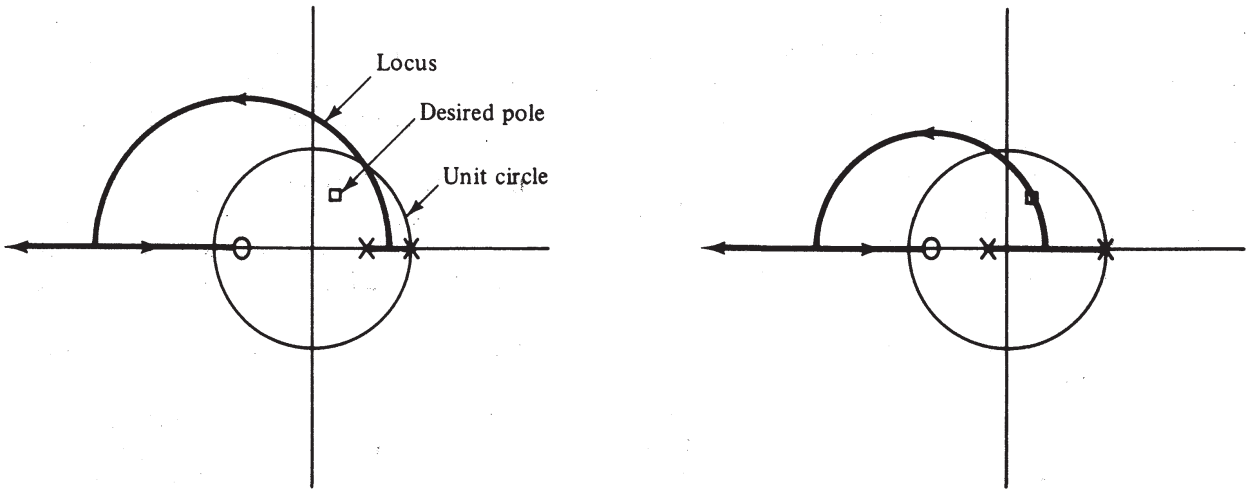


Figure 2.26



(a) Uncompensated root locus

(b) Compensated root locus

Figure 2.27

compensator of the form $G_c(z) = K_c(z - 0.60653)/(z + a)$ is selected. The value of a will be selected so that the locus will be shifted over, as shown in Fig. 2.27*b*. This is a lead-type compensator. As shown in Figure 2.28, if the desired point is to be on the locus, it must be true that $\phi_3 - \phi_1 - \phi_2 = -180$. But $\phi_1 = 180 - \tan^{-1}[0.30956/(1 - 0.19877)] = 158.88°$. Likewise, $\phi_3 = \tan^{-1}[0.30956/(0.84675 + 0.19877)] = 16.49°$. Therefore, ϕ_2 must be 37.61°. This means that the compensator pole must be at $z = -0.203$. The required root locus gain is computed as the product of the vector lengths from the poles, divided by the product of the vector lengths from the zeros to the desired root location ⊡. This gives $K_{RL} = [(0.7378)(0.2572)/1.1889]^{1/2} = 0.3995$. The total root locus gain for this sytem is $K_{RL} = K_c(0.21306K)$, so if, for example, the gain K in the open-loop transfer function has a value of 0.5, then $K_c = 3.7501$, giving the final compensator

$$G_c(z) = 3.7501(z - 0.6065)/(z + 0.2030).$$

The compensated system's response to a step input is shown in Figure 2.29. It shows about 4% overshoot, just what might have been expected from the s-plane roots. This result is not necessarily typical. Quite often the discrete system will have far more overshoot than expected from the s-plane pole positions. This is due to the fact that the sampled system is essentially running open-loop for the intersample periods T. This allows larger overshoot to build up before corrective feedback action can occur. Also, the location of the zero is a major factor in determining the response [6]. The final value theorem shows that this system will have a steady-state error of 3.26 when the input is a ramp.

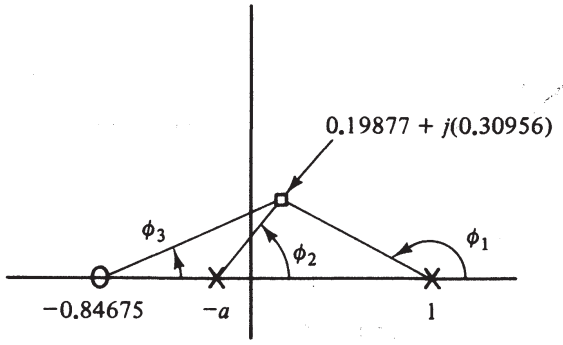


Figure 2.28

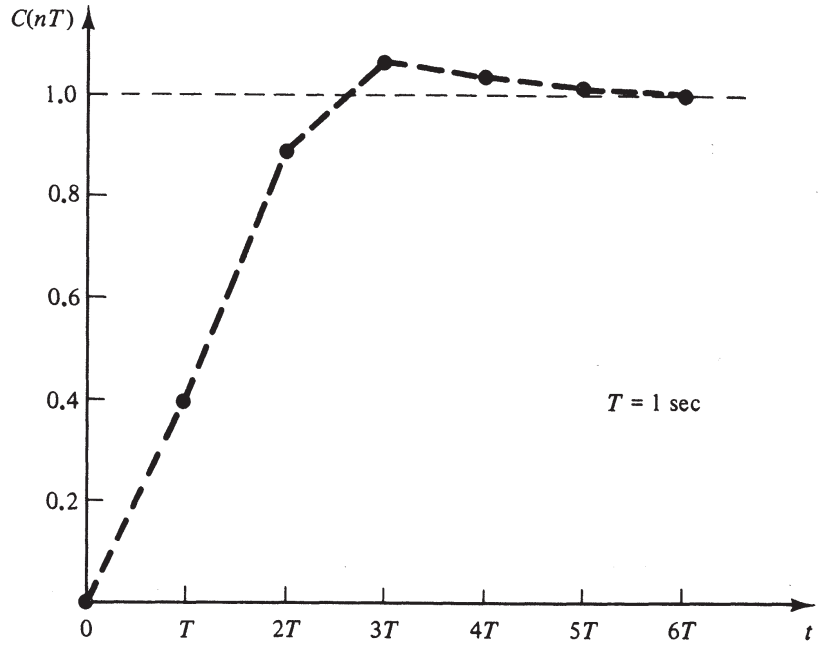


Figure 2.29 Step response.

2.23 Design a cascade compensator for the system of Example 2.6 so that the steady-state value of the error $E(t_k)$ is zero for a ramp input, and such that E goes to zero in the minimum number of sampling periods. This is referred to as the deadbeat response controller.

The transfer function for the error is

$$E(z) = W(z)R(z)$$

where $W(z) = 1/[1 + G_c(z)G'(z)]$.

If $E(t_k)$ is to go to and remain at zero in some finite time, then $W(z)R(z)$ must be a *finite* polynomial in z^{-1}. Since steps, ramps, and other typically used inputs contain a denominator factor of $(1 - z^{-1})^{\alpha}$, $W(z)$ must contain as a factor $(1 - z^{-1})^{\alpha}$. Otherwise, an infinite series in z^{-1} would result for $E(z)$. In this particular case of a ramp input, $\alpha = 2$ and $W(z)$ must have the factor $(1 - z^{-1})^2$. It is generally necessary that $W(z)$ contain another factor $F(z^{-1})$ as well. If $W(z)$ is given, then it is easy to show that $G_c(z) = [1 - W(z)]/[W(z)G'(z)]$. The reason why the extra factor F may be required is that the resulting $G_c(z)$ must be *forced* to be physically realizable if it doesn't come out that way initially. The general rules for doing this are:

1. If $G'(z)$ has no poles or zeros on or outside the unit circle (a single pole at $z = 1$ is acceptable), then $F(z^{-1}) = 1$ is all that is required. A leading 1 is always assumed for the polynomial F, and if any unnecessary extra powers of z^{-1} are included, they only delay the time at which E reaches zero. Otherwise, the following three additional dictums must be met:
2. $F(z^{-1})$ must contain as zeros all the unstable poles of $G'(z)$.
3. $1 - W(z)$ must contain as zeros all the zeros of $G'(z)$ on or outside the unit circle.
4. $1 - W(z)$ must contain z^{-1} as a factor.

In the present case $G'(z)$ has no poles or zeros outside the unit circle so it is permissible to use $F = 1$. Then $W(z) = (1 - z^{-1})^2$ and the compensator is

$$G_c(z) = (2z - 1)(z - 0.6065)/[0.21306K(z + 0.84675)(z - 1)]$$

When this compensator is used, the closed-loop transfer function is $M(z) = 2z^{-1} - z^{-2}$. The responses to step and ramp inputs are as shown in Figs. 2.30 and 2.31, respectively. Note that the design objectives are met for the ramp input, but the step response may not be acceptable since it has 100% overshoot. This illustrates one drawback of deadbeat response controllers: they are

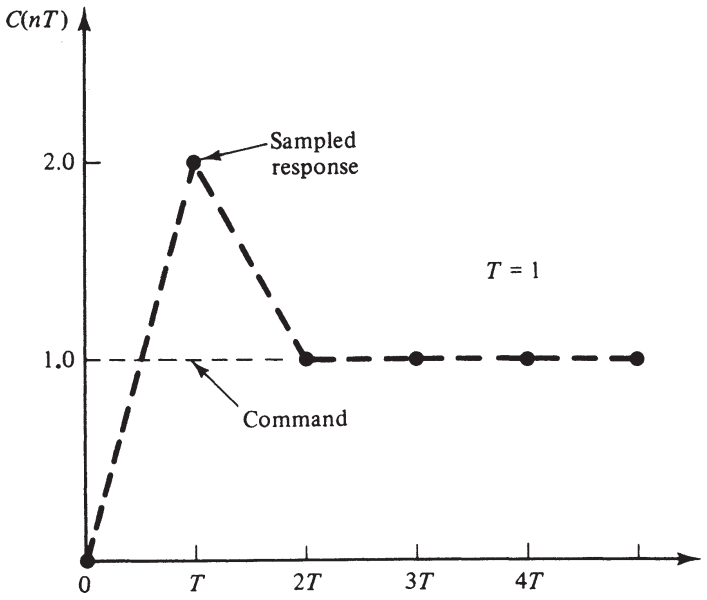


Figure 2.30 Step response.

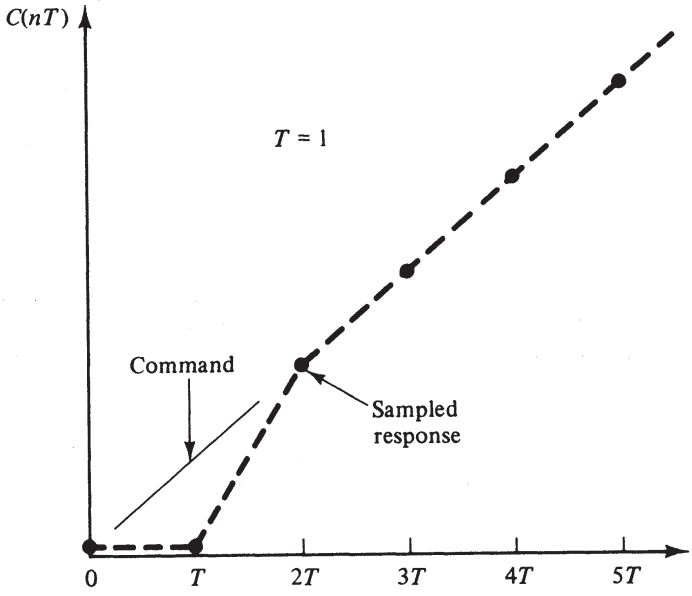


Figure 2.31 Ramp response.

tuned to specific inputs. The other potential drawback is that the intersample error is not necessarily zero just because $E(t_k)$ is zero. There are other methods of suppressing intersample ripple [7].

2.24 Consider the same system as in the previous two problems. This time a cascade compensator is sought which meets the followng specs: (1) closed-loop poles at $z = 0.19877 \pm 0.30956j$, (2) zero steady-state error for a step input, and (3) a velocity error constant K_v of 5 to insure an ability to follow ramp inputs with acceptably small error. This is an example of a direct synthesis method [7], sometimes called the method of Raggazini [6].

A summary of the rules associated with this design method is given before applying them to this specific problem. Let the final closed-loop transfer function be called $M(z)$.

1. In order to ensure that the final compensator is physically realizable, $M(z)$ must have at least as many more poles than zeros as $G'(z)$ does.

2. To ensure a stable closed-loop system, $M(z)$ must contain as zeros all zeros of $G'(z)$ which are on or outside the unit circle. To see this, let the plant transfer function be written $G'(z) = N_{p1}N_{p2}/D_p$, where N_{p2} contains all zeros on or outside the unit circle. Let $G_c = N_c/D_c$. Then

$$M(z) = \frac{N_c N_{p1} N_{p2}}{D_c D_p + N_c N_{p1} N_{p2}}$$

The only way to prevent M from having the zeros in question is to select the compensator denominator to have the factor N_{p2}. But that leaves N_{p2} as a common factor in the denominator of M, thus constituting unstable closed-loop poles. Since perfect cancellation is never possible, this approach is not satisfactory. N_{p2} must be a factor in $M(z)$.

3. Of necessity, if $M(z)$ is to be stable, $1 - M(z)$ must have as zeros all the unstable poles of $G'(z)$. (Those on or outside the unit circle—a single pole at $z = 1$ is not considered unstable.) The reason why this is necessary is clear from a consideration of Eq. (2.5), rearranged as $M = G_c G'(1 - M)$.

4. In order to have zero steady-state error, the final value theorem indicates that $\lim_{z\to 1}\{(z-1)R(z)[1 - M(z)]\} = 0$. The implication of this depends on what $R(z)$ is, but for a step input, since $R(z) = z/(z-1)$, it means that $\lim_{z\to 1} M(z) = 1$.

5. The velocity error coefficient is defined as

$$K_v = (1/T)\lim_{z\to 1}\{(z-1)G_c(z)G'(z)\} = (1/T)\lim_{z\to 1}(z-1)M(z)/[1 - M(z)]$$

For the case where a step input was used in (4), $M(1) = 1$, so the expression for the error constant is indeterminant of the form 0/0. Use of L'Hospital's rule gives $K_v = -1/\{Td[M(z)]/dz\}|_{z=1}$. Solving gives a requirement on $M(z)$ in terms of a specified K_v:

$$\left.\frac{dM(z)}{dz}\right|_{z=1} = \frac{-1}{K_v T}$$

Now for the specific problem here, $M(z)$ must have at least one more pole than zero, since $G'(z)$ does. The open-loop system $G'(z)$ has no poles or zeros outside the unit circle. In view of this and the desired closed-loop poles, a tentative $M(z)$ is selected, which satisfies (1), (2), and (3).

$$M(z) = \frac{A(z + a)}{z^2 - 0.39754z + 0.13534}$$

There are two free design parameters A and a that can be used to satisfy (4) and (5). From (4), $A(1 + a)/0.737797 = 1$ and from (5)

$$A/0.737797 - A(1 + a)(2 - 0.39754)/(0.737797)^2 = -1/5$$

From these two equations the two unknowns are found to be $A = 1.4549$ and $a = -0.492888$. The compensated closed-loop transfer function is

$$M(z) = 1.4549(z - 0.492888)/(z^2 - 0.39754z + 0.135337)$$

Using this, Eq. (2.5) gives the following compensator, after algebraic simplification:

$$G_c(z) = \frac{6.82859(z - 0.60653)(z - 0.492888)}{K(z - 0.85244)(z + 0.84675)}$$

The transient response of the closed-loop system to step and ramp inputs are shown in Figures 2.32 and 2.33, respectively. Note that the maximum percent overshoot (at the sampling times) due to a step input is 45.5% and the settling time is about $4T$, i.e., 4 s. A different set of design specifications may lead to improved performance.

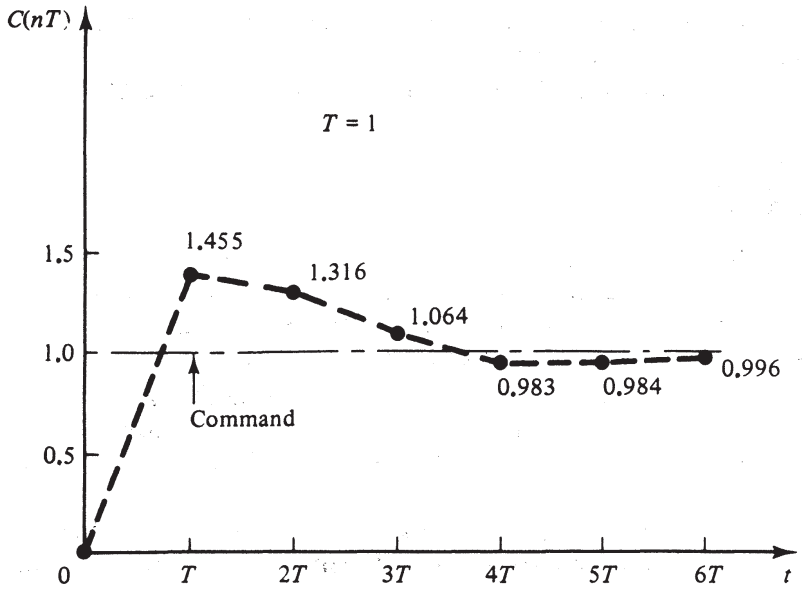


Figure 2.32 Step response.

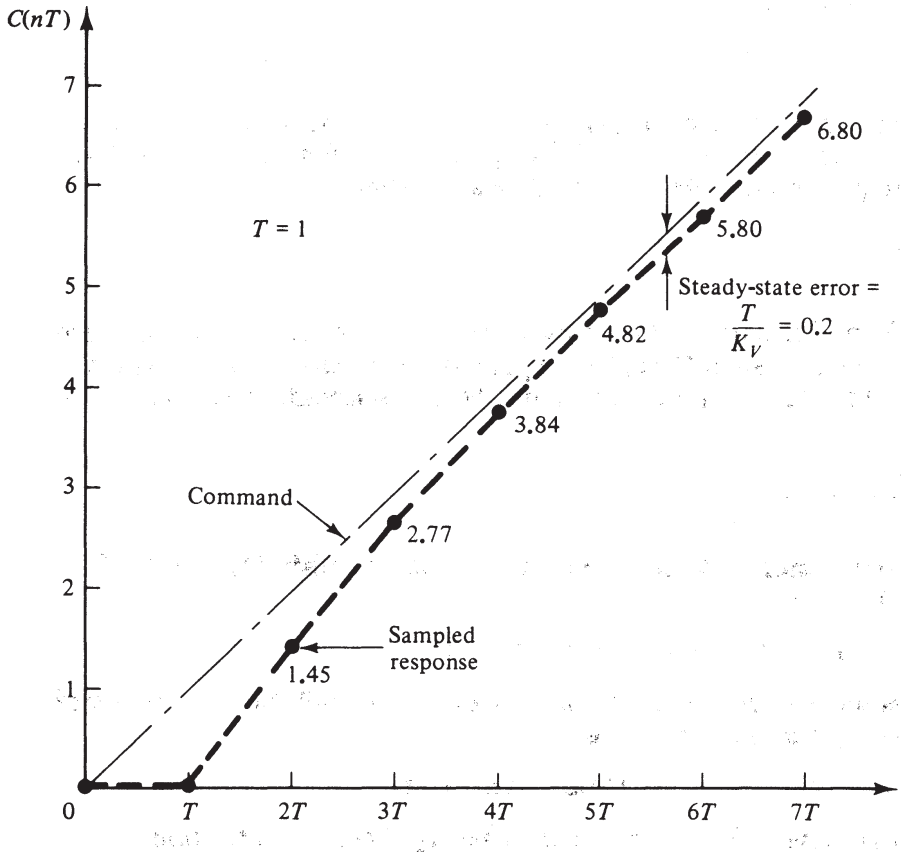


Figure 2.33 Ramp response.

PROBLEMS

2.25 A single-input, single-output system is described by $\dddot{y} + \ddot{y} + 6\dot{y} + (K - 3)y = u(t)$. What is the range of values of K for stability?

2.26 Find the gain K and the frequency ω at which the system of Figure 2.34 becomes unstable. Consider only positive gains.

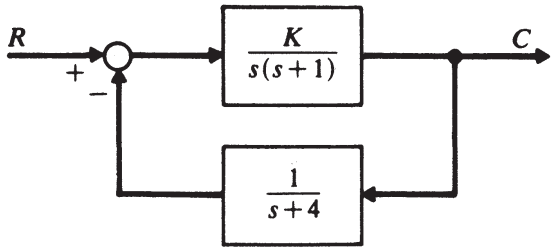


Figure 2.34

2.27 Use Routh's criterion to accurately compute the gain and frequency at the imaginary axis crossover for the system of Problem 2.9.

2.28 The asymptotic gain portions of three Bode plots are shown in Figure 2.35. Identify the system types and their error coefficients.

2.29 Use Bode's method to show that the system with $KGH = K/[s^2(s+1)(s^2+2s+225)]$ is unstable for all positive K.

2.30 Sketch the polar plot and use Nyquist's criterion to investigate the stability of a system with $KGH = K(s+10)/s^2$.

2.31 Sketch the polar plot and use Nyquist's criterion to investigate the stability of a system with $KGH = K(s+10)(s+30)/s^3$.

2.32 Use the polar plot to determine the gain-phase margins for the system described by

$$KGH = \frac{5000(s+2)}{s^2(s+10)(s+30)}$$

2.33 A feedback control system has an open-loop transfer function

$$KGH = \frac{65{,}000K}{s(s+25)(s^2+100s+2600)}$$

Find the value of K such that the exponential envelope of the dominant terms decays to 0.15% of its maximum value in 1 s. Also find the frequency of this damped oscillation.

2.34 Why should Bode plots not be used to infer stability margins for the system in Example 2.3, page 43.

2.35 Give three different algorithms for realizing a digital compensator

$$G_c(z) = (z-0.5)/[z(z+0.5)]$$

whose input and output are E and Y, respectively.

2.36 Determine an expression for the output sequence $C(nT)$, valid for any nonnegative n, if $C(z) = 10z/[(z-1)(z-0.5)(z+0.5)]$.

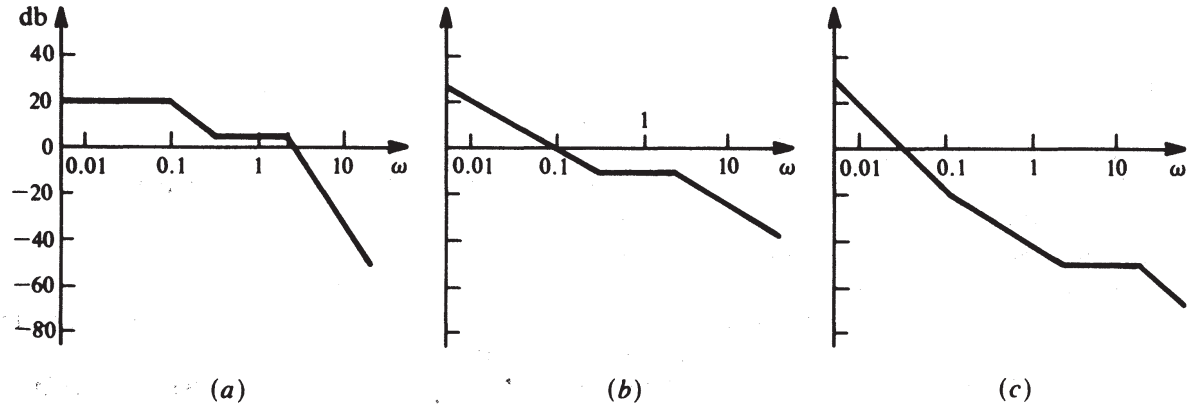


Figure 2.35

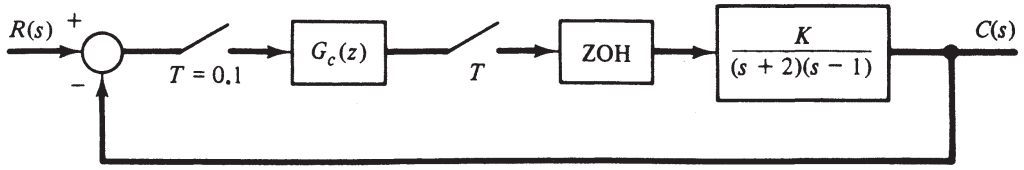


Figure 2.36

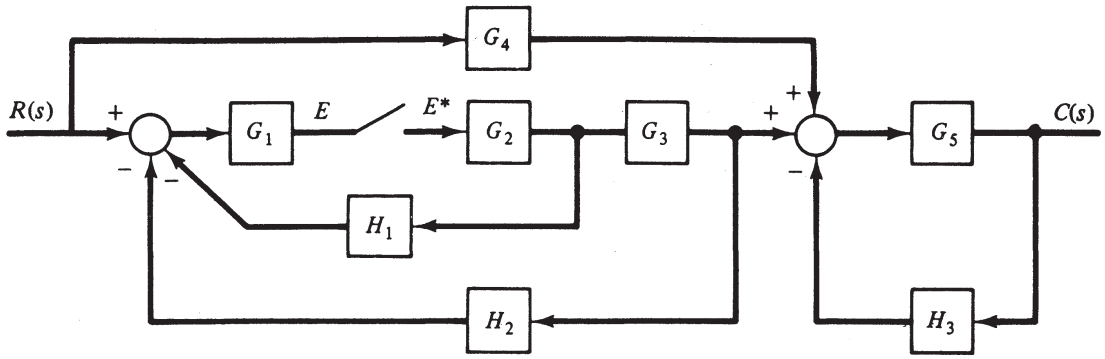


Figure 2.37

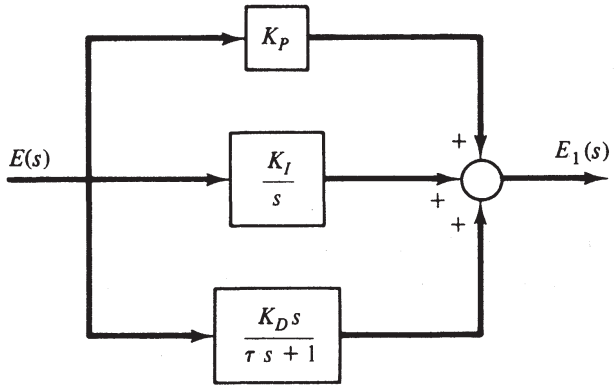


Figure 2.38

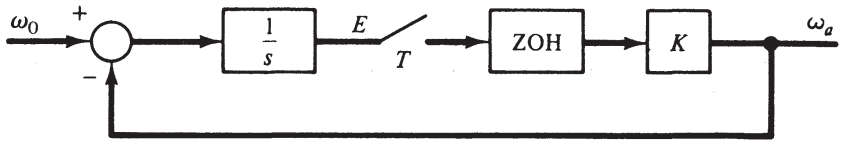


Figure 2.39

2.37 For $C(z) = (z-a)(z-b)/[(z-\alpha)(z-\beta)]$, with a, b, α, and β all distinct:
(a) Find $C(0)$, $C(T)$, and $C(2T)$ using long division.
(b) Find an expression for $C(nT)$ valid for all $n > 0$.
(c) Apply the final value theorem (assume $|\alpha|$ and $|\beta|$ are less than 1), and from its results verify the limiting value of your part (b) answer as $n \to \infty$.

2.38 The system of Figure 2.36 is open-loop unstable.
(a) Determine the closed-loop transfer function $C(z)/R(z)$, assuming that $G_c(z) = 1$.
(b) Sketch the root locus for the uncompensated system. From this sketch, show that the closed-loop system will be stable for some narrow range of K values, but that the resulting system's transient response will not be very desirable. Compensation will probably be required.

2.39 Determine $E(z)$, $C(s)$, and $C(z)$ for the system of Figure 2.37.

2.40 Consider the widely used PID controller (proportional, integral, and derivative control action) shown in Figure 2.38. Use the stable forward difference approximation for s and determine an equivalent digital controller transfer function. Note that a pure derivative does not have a physically realizable transfer function, so a small time-constant term τ is included.

The wide and persistent appeal of PID controllers, especially in the process control industries, can be attributed to their robustness. That is, a properly tuned PID controller can give a good compromise between acceptable time response and disturbance rejection, even with significant model errors present.

2.41 Using the sample and zero-order hold models suggested for A/D and D/A converters, show that the cascade connection of a D/A followed by an A/D has a Z-transfer function of 1, meaning no alteration of a digital signal sequence passing through it.

2.42 A simple one-dimensional model of a digital tracking loop is shown in Figure 2.39. The purpose of the control loop is to keep the angular rate of the antenna ω_a approximately equal to the angular rate ω_0 that the line of sight to the tracked object is making. The integrated angular rate difference is a pointing angle error E. This error angle is sampled, because of a time-shared control computer, and then used to command an antenna rate proportional to the error. The dynamics of the antenna drive system are so rapid that they are neglected, meaning that the actual ω_a is equal to its commanded value. Show that this loop is stable for $0 < KT < 2$.