

CHI-TSONG CHEN

LINEAR SYSTEM THEORY AND DESIGN

INTERNATIONAL
FOURTH EDITION

This version of the text has been adapted and
customized. Not for sale in the U.S.A. or Canada.

OXFORD
UNIVERSITY PRESS

A SUCCINCT AND RIGOROUS INTRODUCTION TO LINEAR AND MULTIVARIABLE SYSTEM DESIGN

"This book offers a superior and accessible presentation of theory."

—Oscar González, *Old Dominion University*

"Chen has an excellent writing style. His book is concise, brief, and to the point.

It is very informative and engaging. Simply perfect for an engineering textbook."

—Amr M. Baz, *University of Maryland, College Park*

"Linear System Theory and Design provides the definitive version of how to present this material."

—Douglas K. Lindner, *Virginia Polytechnic Institute and State University*

Striking a balance between theory and applications, *Linear System Theory and Design*, International Fourth Edition, uses simple and efficient methods to develop results and design procedures that students can readily employ. Ideal for advanced undergraduate courses and first-year graduate courses in linear systems and multivariable system design, it is also a helpful resource for practicing engineers.

NEW TO THIS EDITION

- All examples using MATLAB have been updated using the R2011a version
- New end-of-chapter problems appear throughout the text, contributed by Bahram Shafai, Professor of Electrical and Computer Engineering, Northeastern University
- Chapter 1 provides a brief history and overview of the subject
- Chapter 2 has been restructured in order to discuss four types of equations and to give reasons for focusing on rational transfer functions and state-space equations
- Chapter 4 includes a new section on SISO realizations and expanded discussions of computer computation and real-time processing of state-space equations
- Chapter 7 includes a new section on complete characterization
- PowerPoint-based versions of the figures from the text are available to adopters via the companion website at www.oup.com/us/chen
- An Instructor's Solutions Manual is available to adopters

ABOUT THE AUTHOR

Chi-Tsong Chen is Professor Emeritus of Electrical and Computer Engineering at Stony Brook University, New York.

OXFORD
UNIVERSITY PRESS

www.oup.com/us/he

Cover Design: Kim Howie
Cover Photo: © Shutterstock

ISBN 978-0-19-996454-3



9 780199 964543

CHEN
LINEAR THEORY AND DESIGN
OXFORD

INTERNATIONAL FOURTH EDITION

Linear System Theory and Design

International Fourth Edition

Chi-Tsong Chen

Stony Brook University

with contributions by Bahram Shafai,

Professor of Electrical and Computer Engineering,

Northeastern University

New York Oxford
OXFORD UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Copyright © 2013, 1999, and 1996 by Oxford University Press.

For titles covered by Section 112 of the US Higher Education Opportunity Act, please visit www.oup.com/us/he for the latest information about pricing and alternate formats.

Published by Oxford University Press.
198 Madison Avenue, New York, NY 10016
www.oup.com

Oxford is a registered trademark of Oxford University Press.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data

Chen, Chi-Tsong.

Linear system theory and design / Chi-Tsong Chen, Stony Brook University.—Fourth edition.

pages cm.—(The Oxford series in electrical and computer engineering)

ISBN 978-0-19-996454-3 (hardback)

1. Linear systems. 2. System design. I. Title.

QA402.C44 2012

629.8'32—dc23

2012034220

Linear System Theory and Design



To Jordan, Lauren, Leo, Sommiya, and Zinnia

Preface (International Edition)

This text is intended for use in medium to advanced courses on linear systems in electrical, mechanical, bioengineering, chemical, and aeronautical departments. It may also be useful to practicing engineers because it contains many design procedures. The mathematical background assumed is a working knowledge of linear algebra and the Laplace transform and an elementary knowledge of differential equations. A knowledge of the subject area *Signals and Systems* such as the one in Reference 10 is helpful but not essential.

Linear system theory is a vast subject area. This text studies mainly linear time-invariant lumped systems which are describable by state-space equations and rational transfer functions. The former is an internal description, and the latter is an external description. We study their structures, relationships, and implications in design. As this is an engineering text, we aim to achieve two objectives. The first one is to use simple and efficient methods to develop results and design procedures. Thus the presentation is not exhaustive. For example, we skip many multivariable companion-forms in state-space equations and the Smith–McMillan form in transfer matrices. The second objective is to enable the reader to employ the results to carry out design. Thus most results are discussed with an eye toward numerical computation. All design procedures in the text can be carried out using MATLAB.¹ We adopt the theorem–proof format to cultivate the reader’s ability to think critically and to develop ideas logically.

¹ MATLAB is a registered trademark of the MathWorks, Inc., 24 Prime Park Way, Natick, MA 01760-1500. <http://www.mathworks.com>.

This text also touches upon linear time-invariant distributed systems and linear time-varying lumped systems. We use examples to show that some results in this text are not applicable to those systems. Even so, this text will provide a foundation and a benchmark for studying distributed, time-varying, or nonlinear systems.

The first edition of this text, entitled *Introduction to Linear System Theory*, was published in 1970. Its second edition, renamed *Linear System Theory and Design*, was published in 1984. It expanded to 662 pages from 431 pages of the original edition. The third edition, published in 1999, cut the second edition in half to 332 pages by skipping many topics that are either of only academic interests or of limited practical use. The third edition also introduced the two-parameter (feedforward/feedback) configuration, which is more suitable for practical application.

Before carrying out this revision, Oxford University Press sent out the third edition for reviews. Two reviewers suggested to drop Chapters 7 and 9 because they were not covered in their courses. Indeed, the multi-input and multi-output (MIMO) parts of those two chapters, which are included for completeness, should not be covered; they are more suitable for advanced courses. It is, however, suggested to cover the single-input single-output (SISO) parts because they relate the concept of coprimeness in transfer functions and the concepts of controllability and observability in state-space equations and establish the equivalence of the two descriptions. They also use simpler mathematics to develop results that are more general than those obtained using state-space equations.

New to the Fourth International Edition

The subject area covered in this text is a mature one. Thus this new international edition differs from its preceding one, which was published more than a decade ago, mainly in its presentation as listed in the following:

- A brief history of the subject area is added to Chapter 1.
- Instead of starting from the linear time-varying case, it starts with linear time-invariant lumped systems. It discusses, by restructuring Chapter 2, four types of equations and then give reasons for focusing on rational transfer functions and state-space equations. The concepts *tree*, *loop*, and *link* in graph theory are dropped from this edition.
- Chapter 4 is expanded to discuss computer computation and real-time processing of state-space equations. A new section on SISO realizations is added.
- A new section on complete characterization is added to Chapter 7.
- Some examples are added to provide additional motivation.
- All examples using MATLAB are updated using the R2011a version.
- Many problems with numerical values at the end of each chapter are modified.

It is hoped that this new international edition would be more easily accessible to wider audience.

This text discusses two essentially independent, though related, subject areas. The first area deals with state-space equations, requires all mathematics in Chapter 3, and is discussed in Chapters 4, 5, 6, and 8. The second area deals with rational transfer

functions or polynomial fractions, requires only Section 3.3 of Chapter 3, and is discussed in Chapter 5, 7, and 9.

I am indebted to many people in revising this and previous editions. Professor Imin Kao and Mr. Zhi Chen helped me with MATLAB. Professor Zongli Lin and Mr. T. Anantakrishnan read the whole manuscript and made many valuable suggestions. I am grateful to Dean Yacov Shamash, College of Engineering and Applied Sciences, Stony Brook University, for his encouragement. The third edition of the book was reviewed by Professor B. Ross Barmish, Department of Electrical and Computer Engineering, University of Wisconsin; Professor Harold Broberg, EET Department, Indiana Purdue University; Professor Peyman Givi, Department of Mechanical and Aerospace Engineering, State University of New York at Buffalo; and Professor Mustafa Khamash, Department of Electrical and Computer Engineering, Iowa State University. Their detailed and critical comments prompted me to restructure some sections. I thank them all.

Great thanks also goes to Bahram Shafai, Professor of Northeastern University, who provided a significant portion of the end-of-chapter problems for this international edition.

I am indebted to Patrick Lynch, Dan Pepper, Claire Sullivan, and Carolyn DiTullio of Oxford University Press, who were instrumental for this revision. The people at Oxford University Press, including Pamela Hanley, Christine Mahon, and Deborah Gross were most helpful in this undertaking.

Chi-Tsong Chen

Contents

Preface (International Edition)	xi
CHAPTER 1	
Introduction	1
1.1 Introduction	1
1.2 Overview	2
1.2.1 A Brief History	4
CHAPTER 2	
Mathematical Descriptions of Systems	6
2.1 Introduction	6
2.2 Causality, Lumpedness, and Time Invariance	7
2.2.1 Impulses	9
2.3 Linear Time-Invariant Systems	11
2.3.1 Multi-input Multi-output Case	18
2.4 Linear Time-Varying Systems	19
2.4.1 Linearization	20
2.5 RLC Circuits—Comparisons of Various Descriptions	21
2.6 Mechanical and Hydraulic Systems	30
2.7 Proper Rational Transfer Functions	38
2.8 Discrete-Time Linear Time-Invariant Systems	40
2.9 Concluding Remarks	48
Problems	49

CHAPTER 3

Linear Algebra	56
3.1 Introduction	56
3.2 Basis, Representation, and Orthonormalization	57
3.3 Linear Algebraic Equations	62
3.4 Similarity Transformation	67
3.5 Diagonal Form and Jordan Form	69
3.6 Functions of a Square Matrix	76
3.7 Lyapunov Equation	85
3.8 Some Useful Formulas	87
3.9 Quadratic Form and Positive Definiteness	88
3.10 Singular Value Decomposition	92
3.11 Norms of Matrices	94
Problems	96

CHAPTER 4

State-Space Solutions and Realizations	103
4.1 Introduction	103
4.2 General Solution of CT LTI State-Space Equations	105
4.2.1 Discretization	108
4.2.2 General Solution of DT LTI State-Space Equations	110
4.3 Computer Computation of CT State-Space Equations	112
4.3.1 Real-Time Processing	115
4.3.2 Op-Amp Circuit Implementation	116
4.4 Equivalent State-Space Equations	117
4.4.1 Canonical Forms	122
4.4.2 Magnitude Scaling in Op-Amp Circuits	124
4.5 Realizations	126
4.5.1 Multi-input Multi-output Case	130
4.6 Solution of Linear Time-Varying (LTV) Equations	135
4.6.1 Discrete-Time Case	140
4.7 Equivalent Time-Varying Equations	141
4.8 Time-Varying Realizations	145
Problems	147

CHAPTER 5

Stability	153
5.1 Introduction	153
5.2 Input–Output Stability of LTI Systems	153
5.3 Discrete-Time Case	162

5.4	Internal Stability	167
	5.4.1 Discrete-Time Case	169
5.5	Lyapunov Theorem	170
	5.5.1 Discrete-Time Case	173
5.6	Stability of LTV Systems	176
	Problems	179
CHAPTER 6		
Controllability and Observability		183
6.1	Introduction	183
6.2	Controllability	184
	6.2.1 Controllability Indices	190
6.3	Observability	193
	6.3.1 Observability Indices	197
6.4	Kalman Decomposition	199
6.5	Conditions in Jordan-Form Equations	206
6.6	Discrete-Time State-Space Equations	210
	6.6.1 Controllability to the Origin and Reachability	213
6.7	Controllability after Sampling	214
6.8	LTV State-Space Equations	217
	Problems	222
CHAPTER 7		
Minimal Realizations and Coprime Fractions		227
7.1	Introduction	227
7.2	Implications of Coprimeness	228
	7.2.1 Minimal Realizations	232
	7.2.2 Complete Characterization	236
7.3	Computing Coprime Fractions	238
	7.3.1 QR Decomposition	242
7.4	Balanced Realization	243
7.5	Realizations from Markov Parameters	247
7.6	Degree of Transfer Matrices	252
7.7	Minimal Realizations—Matrix Case	254
7.8	Matrix Polynomial Fractions	256
	7.8.1 Column and Row Reducedness	259
	7.8.2 Computing Matrix Coprime Fractions	262
7.9	Realization from Matrix Coprime Fractions	269
7.10	Realizations from Matrix Markov Parameters	275
7.11	Concluding Remarks	277
	Problems	277

CHAPTER 8

State Feedback and State Estimators	282
8.1 Introduction	282
8.2 State Feedback	283
8.2.1 Solving Lyapunov Equation	290
8.3 Regulation and Tracking	293
8.3.1 Robust Tracking and Disturbance Rejection	295
8.3.2 Stabilization	299
8.4 State Estimator	299
8.4.1 Reduced-Dimensional State Estimator	303
8.5 Feedback from Estimated States	305
8.6 State Feedback—MIMO Case	307
8.6.1 Cyclic Design	308
8.6.2 Lyapunov-Equation Method	311
8.6.3 Controllable-Form Method	312
8.6.4 Effect on Transfer Matrices	314
8.7 State Estimators—MIMO Case	316
8.8 Feedback from Estimated States—MIMO Case Problems	317
	319

CHAPTER 9

Pole Placement and Model Matching	323
9.1 Introduction	323
9.2 Preliminary—Matching Coefficients	325
9.2.1 Compensator Equations—Classical Method	327
9.3 Unity-Feedback Configuration—Pole Placement	328
9.3.1 Regulation and Tracking	331
9.3.2 Robust Tracking and Disturbance Rejection	333
9.3.3 Embedding Internal Models	337
9.4 Implementable Transfer Functions	339
9.4.1 Model Matching—Two-Parameter Configuration	343
9.4.2 Implementation of Two-Parameter Compensators	348
9.5 MIMO Unity Feedback Systems	350
9.5.1 Regulation and Tracking	360
9.5.2 Robust Tracking and Disturbance Rejection	362
9.6 MIMO Model Matching—Two-Parameter Configuration	364
9.6.1 Decoupling	370
9.7 Concluding Remarks	374
Problems	375
References	379
Answers to Selected Problems	381
Index	391

Introduction

1.1 Introduction

The study and design of physical systems can be carried out using empirical methods. We can apply various signals to a physical system and measure its responses. If the performance is not satisfactory, we can adjust some of its parameters or connect to it a compensator to improve its performance. This approach relies heavily on past experience and is carried out by trial and error, and it has succeeded in designing many physical systems.

Empirical methods may become unworkable if physical systems are complex, or they may become too expensive or too dangerous to be experimented on. In these cases, analytical methods become indispensable. The analytical study of physical systems consists of four parts: modeling, development of mathematical descriptions, analysis, and design. We briefly introduce each of these tasks.

The distinction between physical systems and models is important in engineering. For example, circuits or control systems studied in any textbook are models of physical systems. A resistor with a constant resistance is a model; it will burn out if the applied voltage is over a limit. This power limitation is often disregarded in its analytical study. An inductor with a constant inductance is again a model; in reality, the inductance may vary with the amount of current flowing through it. Modeling is a very important problem, for the success of the design depends upon whether the physical system is properly modeled.

A physical system may have different models, depending on the questions asked. It may also be modeled differently in different operational ranges. For example, an electronic amplifier is modeled differently at high and low frequencies. A spaceship can be modeled as a particle in investigating its trajectory; however, it must be modeled as a rigid body in maneuvering. A spaceship may even be modeled as a flexible body

when it is connected to a space station. In order to develop a suitable model for a physical system, a thorough understanding of the physical system and its operational range is essential. In this text, we will call a model of a physical systems simply a *system*. Thus a physical system is a device or a collection of devices existing in the real world; a system is a model of a physical system.

Once a system (or model) is selected for a physical system, the next step is to apply various physical laws to develop mathematical equations to describe the system. For example, we apply Kirchhoff's voltage and current laws to electrical systems and Newton's law to mechanical systems. The equations that describe systems may assume many forms; they may be linear equations, nonlinear equations, integral equations, difference equations, differential equations or others. Depending on the problem under study, one form of equation may be preferable to another in describing the same system. In conclusion, a system may have different mathematical-equation descriptions, just as a physical system may have many different models.

After a mathematical description is obtained, we then carry out analyses—quantitative and qualitative. In quantitative analysis, we are interested in responses of systems excited by certain inputs. In qualitative analysis, we are interested in general properties of systems, such as stability, controllability, and observability. Qualitative analysis is very important, because design techniques may often evolve from this study.

If the response of a system is unsatisfactory, the system must be modified. In some cases, this can be achieved by adjusting some parameters of the system; in other cases, compensators must be introduced. Note that the design is carried out on the model of the physical system. If the model is properly chosen, then the performance of the physical system should be improved by introducing the required adjustments or compensators. If the model is poor, then the performance of the physical system may not improve and the design is useless. Selecting a model that is close enough to a physical system and yet simple enough to be studied analytically is the most difficult and important problem in system design.

1.2 Overview

The study of systems consists of four parts: modeling, setting up mathematical equations, analysis, and design. Developing models for physical systems requires knowledge of the particular field and some measuring devices. For example, to develop models for transistors requires a knowledge of quantum physics and some laboratory setup. Developing models for automobile suspension systems requires actual testing and measurements; it cannot be achieved by use of pencil and paper. Computer simulation certainly helps but cannot replace actual measurements. Thus the modeling problem should be studied in conjunction with the specific field and cannot be properly covered in this text. In this text, we shall assume that models of physical systems are available to us.

The systems to be studied in this text are mostly limited to linear (L), time-invariant (TI), and lumped. This class of systems with input $u(t)$ and output $y(t)$ can be described by

1. Convolution:

$$y(t) = \int_{\tau=0}^t g(t-\tau)u(\tau)d\tau \quad (1.1)$$

where $g(t)$ is called the impulse response.

2. Transfer function:

$$\hat{y}(s) = \hat{g}(s)\hat{u}(s) \quad (1.2)$$

where a variable with a circumflex is the Laplace transform of the variable. The function $\hat{g}(s)$ is called a transfer function and is a rational function of s such as

$$\hat{g}(s) = \frac{3s^2 - 2s + 5}{s^3 + 4s^2 + 2.5s + 3}$$

3. High-order differential equation such as

$$y^{(3)} + 4\ddot{y}(t) + 2.5\dot{y}(t) + 3y(t) = 3\ddot{u}(t) - 2\dot{u}(t) + 5u(t) \quad (1.3)$$

where $y^{(3)}(t) := d^3y(t)/dt^3$, $\ddot{y}(t) := d^2y(t)/dt^2$, and $\dot{y}(t) := dy(t)/dt$.¹ It is a third-order linear differential equation with constant coefficients.

4. State-space equation:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{Ax}(t) + \mathbf{bu}(t) \\ \mathbf{y}(t) &= \mathbf{cx}(t) + \mathbf{du}(t) \end{aligned} \quad (1.4)$$

where $\mathbf{x}(t)$ is a column vector, called the state and \mathbf{A} , \mathbf{b} , \mathbf{c} , and \mathbf{d} are constant matrices of compatible orders.

Because all variables are functions of time, Equations (1.1), (1.3), and (1.4) are called *time-domain* descriptions, whereas (1.2) is called the *transform-domain* description. The first three are called *input-output* or *external descriptions* because they relate only the input u and output y . The state-space equation is called the *internal description* because it also describes the internal variables \mathbf{x} of the system. These four types of equations can be used to describe the same system.

This text will start with the introduction of convolution, because it shows explicitly the use of the concepts of linearity and time invariance. The convolution will be used to develop transfer functions and a stability condition. It, however, will not be used in analysis and design, as we will discuss in the text. Thus its discussion is not exhaustive. We will also show why the differential equation is not studied in this text. We then concentrate on the transfer function, an external description, and the state-space equation, the internal description. We discuss their solutions, properties, and relationships. We also use them to carry out design in feedback systems. We discuss first single-input single-output (SISO) systems and then multi-input multi-output (MIMO) systems. We

¹We use $A := B$ to denote that A , by definition, equals B . We use $A =: B$ to denote that B , by definition, equals A .

start with the continuous-time (CT) case because most physical systems are continuous time and then discuss their discrete-time (DT) counterparts.

We will also touch upon LTI distributed systems described by irrational transfer functions, as well as linear and lumped systems described by time-varying state-space equations. We show that some results for LTI lumped systems may not be applicable to linear time-varying or distributed systems. Thus their study is much more complicated. The study of nonlinear systems is even more so. To study them, this text however will provide a foundation and a benchmark.

1.2.1 A Brief History

This text studies basically rational transfer functions and state-space equations. Thus it is appropriate to discuss the history of their uses in the electrical engineering curriculum. Oliver Heaviside (1850–1925) used p to denote differentiation and introduced resistive operators R , Lp , and $1/Cp$ to study electric circuits in 1887. This ushered the concepts of impedances and transfer functions. Heaviside's operational calculus, however, lacked an inversion formula and was incomplete. It was recognized around 1940 that the Laplace transform, developed by Pierre-Simon Laplace in 1782, encompassed the Heaviside method. Because of its simplicity, the Laplace transform and the transfer or system function had gradually crept into the EE curriculum. By 1960, transfer functions are used in analog filter designs, passive-network syntheses, and circuit analysis.

In order to explain the dancing of flyballs and jerking of machine shafts in steam engines, James Maxwell developed in 1868 a third-order linearized differential equation to raise the issue of stability. This earmarked the beginning of mathematical study of control systems. In 1932, Henry Nyquist developed a graphical method of checking the stability of a feedback system from its open-loop system. The method, however, is fairly complex and is not suitable for design. In 1940, Hendrick Bode simplified the method to use the phase margin and gain margin of the open-loop system to check stability. However the method is applicable only to a small class of open-loop systems. Moreover, the relationship between phase and gain margins and system performances is vague. In 1948, W. R. Evans developed the root-locus method to carry out design of feedback systems. The method is general, but the compensator used is essentially limited to degree 0. The aforementioned methods are all based on transfer functions and constitute the entire bulk of most texts on control systems published before 1970.

State-space equations first appeared in the engineering literature in the early 1960s. The formulation is precise: It first gives definitions, and then it develops conditions and finally establishes theorems. Moreover, its formulation for SISO and MIMO systems are the same, and all results for SISO systems can be extended to MIMO systems. The most celebrated results are: If a state-space equation is controllable, then state feedback can achieve arbitrary eigenvalue-placement. If a state-space equation is observable, then a state estimator with any desired eigenvalues can be constructed. By 1980, state-space equations and designs were introduced into many undergraduate texts on control.

With the impetus of state-space results, researchers took a fresh look in the 1970s into transfer functions. By considering a rational function as a ratio of two

polynomials, the polynomial-fraction approach was born. By so doing, the results in SISO systems can also be extended to MIMO systems. The important concept in this approach is coprimeness. Under the coprimeness assumption, it is possible to achieve pole-placement and model-matching designs. The method is simpler and the results are more general than those based on state-space equations, as we will demonstrate in this text. Even so, state-space equations are indispensable in studying systems because they are used in computer computation and simulation, real-time processing, and op-amp circuit implementation. Thus both rational transfer functions and state-space equations are important in studying systems.

Mathematical Descriptions of Systems

2.1 Introduction

This text models a system as a black box with one or more input terminals and one or more output terminals as shown in Fig. 2.1. We assume that if an excitation or input is applied to the input terminal, a *unique* response or output signal can be measured at the output terminal. This unique relationship between the excitation and response, input and output, or cause and effect is essential in defining a system. A system with only one input terminal and only one output terminal is called a single-input single-output (SISO) system. A system with two or more input terminals and two or more output terminals is called a multi-input multi-output (MIMO) system. Likewise, a single-input multi-output (SIMO) system has only one input terminal and two or more output terminals. A MISO system has multi-input terminals and single-output terminal.

A signal is called a *continuous-time* (CT) signal if it is defined at every instant of time. A system is called a CT system if it accepts CT signals as its input and generates CT signals as its output. The input will be denoted by lowercase italic $u(t)$ for single input or by boldface $\mathbf{u}(t)$ for multiple inputs. If a system has p input terminals, then $\mathbf{u}(t)$ is a $p \times 1$ vector or $\mathbf{u} = [u_1 \ u_2 \ \dots \ u_p]'$, where the prime denotes the transpose. Similarly, the output will be denoted by $y(t)$ or $\mathbf{y}(t)$. The time t is mostly assumed to range from $-\infty$ to ∞ .

A signal is called a *discrete-time* (DT) signal if it is defined only at discrete instants of time. A system is called a DT system if it accepts DT signals as its input and generates DT signals as its output. All DT signals in a system will be assumed to have the same sampling period T . The input and output will be denoted by $u[k] := u(kT)$ and $y[k] := y(kT)$, where the integer k , ranging mostly from $-\infty$ to ∞ , is called the *time index* and kT denotes discrete time instant. For multiple inputs and multiple outputs, we use boldface $\mathbf{u}[n]$ and $\mathbf{y}[n]$.

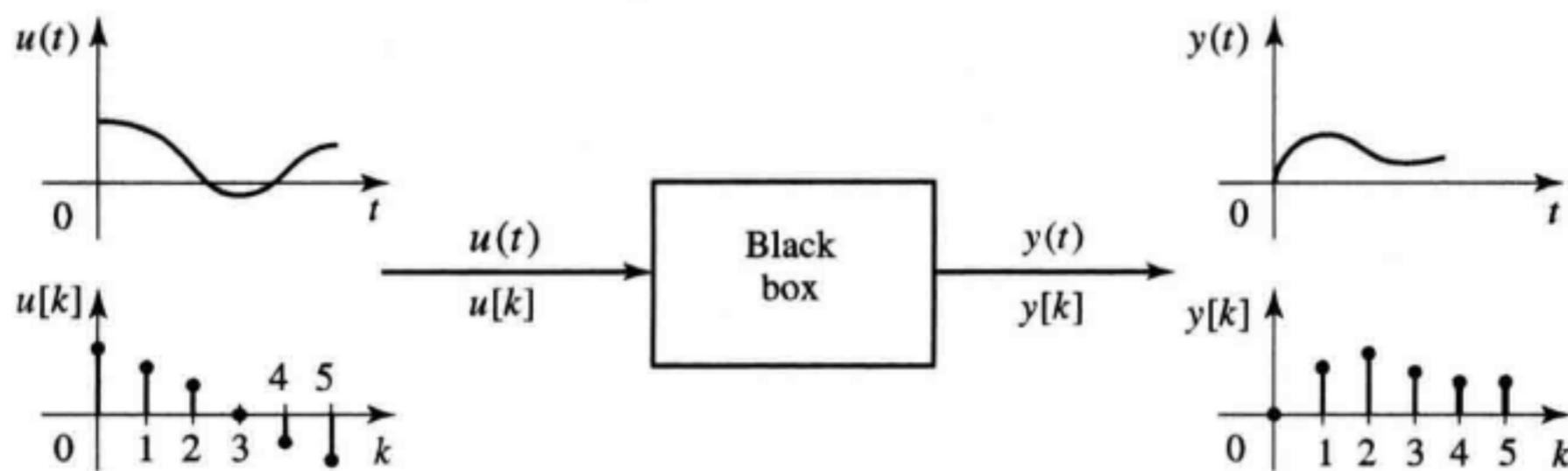


FIGURE 2.1 System.

2.2 Causality, Lumpedness, and Time Invariance

A system is called a *memoryless system* if its output $y(t_0)$ depends only on the input applied at t_0 ; it is independent of the input applied before or after t_0 . This will be stated succinctly as: Current output of a memoryless system depends only on current input; it is independent of past and future inputs. A circuit that consists of only resistors is a memoryless system. Operational amplifiers are often modeled as memoryless. See Reference 10.

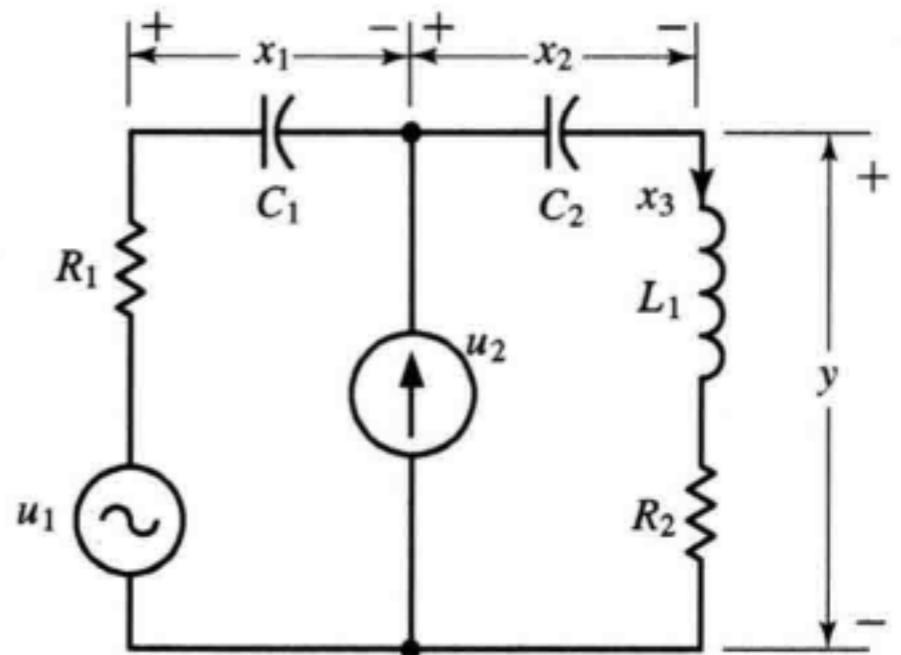
Most systems, however, have memory. By this, we mean that the output at t_0 depends on $\mathbf{u}(t)$ for $t < t_0$, $t = t_0$, and $t > t_0$. That is, current output of a system with memory may depend on past, current, and future inputs.

A system is called a *causal* or *nonanticipatory* system if its current output depends on past and current inputs but not on future input. If a system is not causal, then its current output will depend on future input. In other words, a noncausal system can *predict* or *anticipate* what will be applied in the future. No physical system has such capability. Therefore, every physical system is causal, and causality is a necessary condition for a system to be built or implemented in the real world. This text studies only causal systems.

Current output of a causal system is affected by past input. How far back in time will the past input affect the current output? Generally, the time should go all the way back to minus infinity. In other words, the input from $-\infty$ to time t_0 has an effect on $y(t_0)$. Tracking $\mathbf{u}(t)$ from $t = -\infty$ onward is, if not impossible, very inconvenient. The concept of state can deal this problem.

Definition 2.1 The state $\mathbf{x}(t_0)$ of a system at time t_0 is the information at t_0 that, together with the input $\mathbf{u}(t)$, for $t \geq t_0$, determines uniquely the output $\mathbf{y}(t)$ for all $t \geq t_0$.

By definition, if we know the state at t_0 , there is no more need to know the input $\mathbf{u}(t)$ applied before t_0 in determining the output $\mathbf{y}(t)$ after t_0 . Thus in some sense, the state summarizes the effect of past input on future output. For the circuit shown in Fig. 2.2, if we know the voltages $x_1(t_0)$ and $x_2(t_0)$ across the two capacitors and the current $x_3(t_0)$ passing through the inductor, then for any input applied on and after t_0

FIGURE 2.2 Circuit with three state variables.

we can determine uniquely the output for $t \geq t_0$. Thus the state of the circuit at time t_0 is

$$\mathbf{x}(t_0) = \begin{bmatrix} x_1(t_0) \\ x_2(t_0) \\ x_3(t_0) \end{bmatrix}$$

It is a 3×1 vector. The entries of \mathbf{x} are called *state variables*. Thus, in general, we may consider the state simply as a set of initial conditions. We call $\mathbf{x}(t_0)$ the *initial state*.

Using the initial state at t_0 , we can express the input and output of a system as

$$\left. \begin{array}{l} \mathbf{x}(t_0) \\ \mathbf{u}(t), t \geq t_0 \end{array} \right\} \rightarrow \mathbf{y}(t), \quad t \geq t_0 \quad (2.1)$$

It means that the output is partly excited by the initial state at t_0 and partly by the input applied at and after t_0 . In using (2.1), there is no more need to know the input applied before t_0 all the way back to $-\infty$. Thus (2.1) is easier to track and will be called a state-input-output pair.

A system is said to be *lumped* if its number of state variables is finite or its state is a finite vector. The circuit in Fig. 2.2 is clearly a lumped system; its state consists of three numbers. A system is called a *distributed* system if its state has infinitely many state variables. The transmission line is the most well known distributed system. We give one more example.

Example 2.2.1 Consider the unit-time delay system defined by

$$y(t) = u(t - 1)$$

The output is simply the input delayed by one second. In order to determine $\{y(t), t \geq t_0\}$ from $\{u(t), t \geq t_0\}$, we need the information $\{u(t), t_0 - 1 \leq t < t_0\}$. Therefore, the initial state of the system is $\{u(t), t_0 - 1 \leq t < t_0\}$. There are infinitely many points in $\{t_0 - 1 \leq t < t_0\}$. Thus the unit-time delay system is a distributed system.

A system is said to be *time invariant* if its characteristics do not change with time. For example, the circuit in Fig. 2.2 is time invariant if R_i , C_i and L_i are constants, independent of time. For such a system, no matter at what time we apply an input, the output waveform will always be the same. This property can be expressed using

state-input-output pairs as follows. Let

$$\left. \begin{array}{l} \mathbf{x}(t_0) \\ \mathbf{u}(t), t \geq t_0 \end{array} \right\} \rightarrow \mathbf{y}(t), \quad t \geq t_0$$

be any pair of a system. If the system is time invariant, then we have, for any t_1 ,

$$\left. \begin{array}{l} \mathbf{x}(t_0 + t_1) \\ \mathbf{u}(t - t_1), t \geq t_0 + t_1 \end{array} \right\} \rightarrow \mathbf{y}(t - t_1), \quad t \geq t_0 + t_1 \quad (\text{time shifting})$$

where $\mathbf{u}(t - t_1)$ and $\mathbf{y}(t - t_1)$ are shifting, respectively, of $\mathbf{u}(t)$ and $\mathbf{y}(t)$ from t_0 to $t_0 + t_1$. The equation means that if the initial state is shifted to time $t_0 + t_1$ and the same input waveform is applied from $t_0 + t_1$ instead of from t_0 , then the output waveform will be the same except that it starts to appear from time $t_0 + t_1$. In other words, if the initial state and the input are the same, no matter at what time they are applied, the output waveform will always be the same. Therefore, for time-invariant systems, we can always assume, without loss of generality, that $t_0 = 0$. Note that the initial time $t_0 = 0$ is not absolute; it can be selected by us. It may be the instant we start to study a system. If we select $t_0 = 0$, then the time interval of interest will be $[0, \infty)$. If a system is not time invariant, it is said to be *time varying*.

Some physical systems must be modeled as time-varying systems. For example, a burning rocket is a time-varying system, because its mass decreases rapidly with time. Although the performance of an automobile or a TV set may deteriorate over a long period of time, its characteristics do not change appreciable in the first couple of years. Thus a large number of physical systems can be modeled as time invariant over a limited time period.

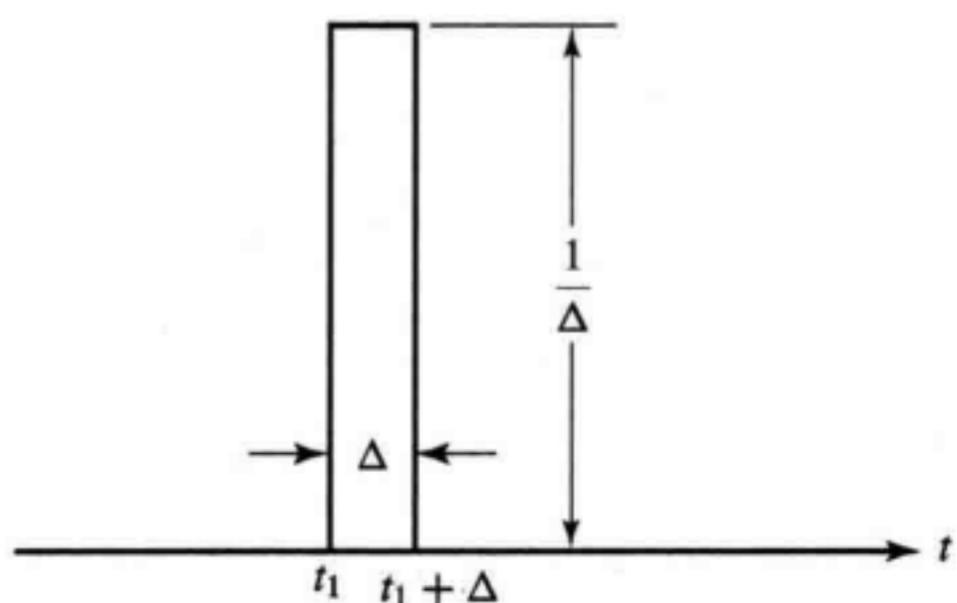
2.2.1 Impulses

We need the concept of impulses to develop some mathematical equations. Consider the pulse defined by

$$\delta_a(t - t_1) := \begin{cases} 1/\Delta & t_1 \leq t < t_0 + \Delta \\ 0 & t < t_1 \text{ and } t \geq t_1 + \Delta \end{cases}$$

and shown in Fig. 2.3. It is located at time t_1 and has width Δ and height $1/\Delta$. Its area or any integration covering the pulse equals 1 for any $\Delta > 0$. The *impulse* at $t = t_1$ is

FIGURE 2.3 Pulse at t_1 .



then defined as

$$\delta(t - t_1) := \lim_{\Delta \rightarrow 0} \delta_\Delta(t - t_1)$$

It equals 0 for all $t \neq t_1$ and ∞ at $t = t_1$. Its area or any integration covers the impulse equals 1 such as

$$\int_{t=-\infty}^{\infty} \delta(t - t_1) dt = \int_{t=t_1}^{t_1+} \delta(t - t_1) dt = \int_{t=t_1}^{t_1} \delta(t - t_1) dt = 1$$

where t_1+ is infinitesimally larger than t_1 . For convenience, we drop the use of t_1+ as in the last integration and assume that whenever an integration touches an impulse, the integration covers the entire impulse. Note that if an integration interval does not cover nor touch an impulse, then the integration is zero such as

$$\int_{t=-\infty}^2 \delta(t - 2.5) dt = 0 \quad \text{and} \quad \int_{t=2.501}^{\infty} \delta(t - 2.5) dt = 0$$

where the impulse is located at $t = 2.5$. The impulse has the following *sifting* property

$$\begin{aligned} \int_{t=-\infty}^{\infty} f(t) \delta(t - t_1) dt &= \int_{t=t_1}^{t_1} f(t) \delta(t - t_1) dt \\ &= f(t)|_{t=t_1=0} = f(t)|_{t=t_1} = f(t_1) \end{aligned}$$

for any function $f(t)$ that is continuous at t_1 . The impulse at t_1 sifts out the value of $f(t)$ at $t = t_1$. In general, whenever a function is multiplied by an impulse in an integration, we simply move the function outside the integration and then replace the integration variable by the variable obtained by equating the argument of the impulse to zero.

We discuss an application. Consider the signal shown in Fig. 2.4. It can be approximated by a staircase function formed from a sequence of pulses as shown. We call Δ the *step size*. The pulse in Fig. 2.3 has height $1/\Delta$; its multiplication by Δ or $\delta_\Delta(t - t_i)\Delta$ has height 1. The leftmost pulse in Fig. 2.4 has height $u(t_i)$, thus it can be expressed as

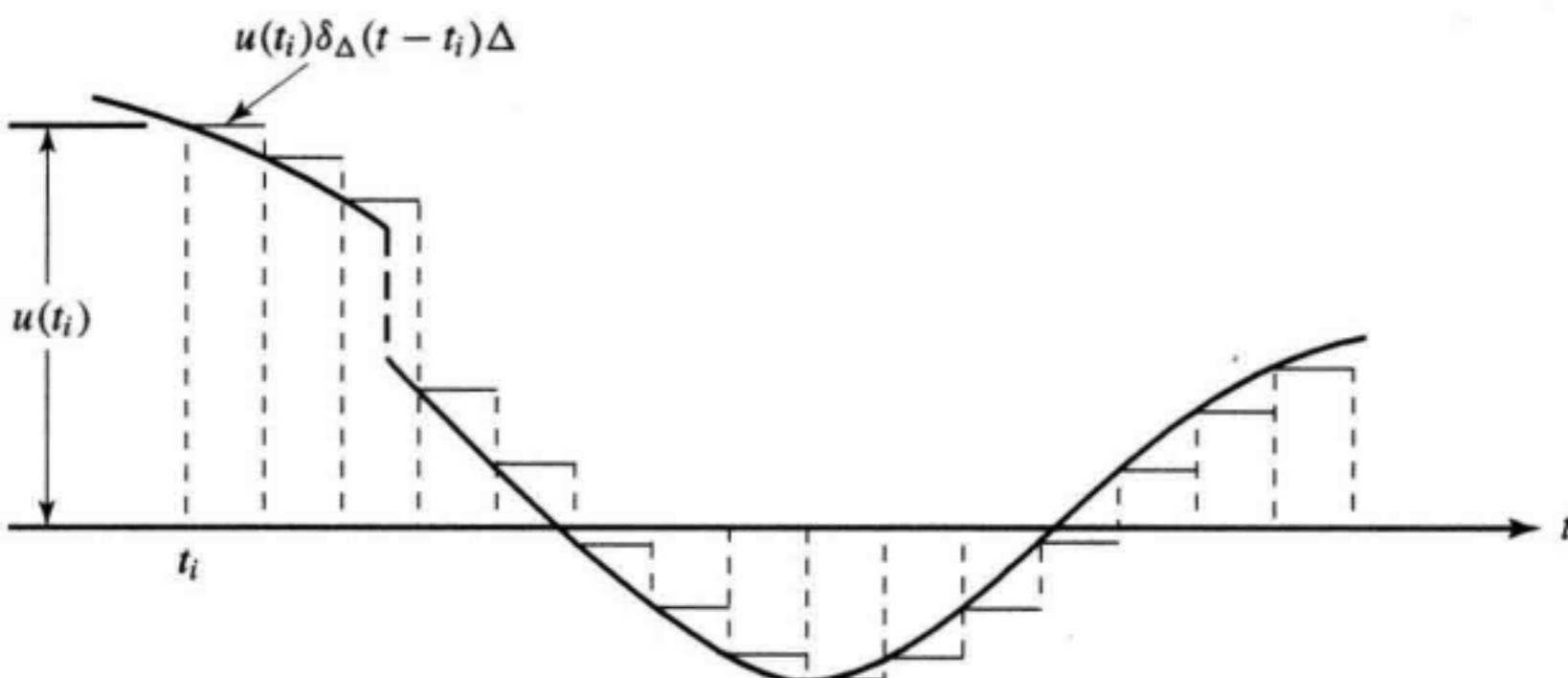


FIGURE 2.4 Approximation of input signal.

$u(t_i)\delta_\Delta(t - t_i)\Delta$. Consequently, the input $u(t)$ can be expressed approximately as

$$u(t) \approx \sum_i u(t_i)\delta_\Delta(t - t_i)\Delta$$

where the summation is an exact description of the staircase function. As $\Delta \rightarrow 0$, the pulse $\delta_\Delta(t - t_i)$ becomes an impulse, t_i becomes a continuous variable and will be denoted by τ , and Δ can be written as $d\tau$. Moreover, the summation becomes an integration and the approximation becomes an equality. Thus, as $\Delta \rightarrow 0$, we have

$$u(t) = \int_{\tau=-\infty}^{\infty} u(\tau)\delta(t - \tau)d\tau \quad (2.2)$$

This is actually the sifting property discussed earlier. Thus the staircase function becomes the original function as $\Delta \rightarrow 0$ and is a mathematically sound approximation of $u(t)$.

2.3 Linear Time-Invariant Systems

A time-invariant system is defined to be *linear* if for any two state-input-output pairs

$$\left. \begin{array}{l} \mathbf{x}_i(0) \\ \mathbf{u}_i(t), t \geq 0 \end{array} \right\} \rightarrow \mathbf{y}_i(t), \quad t \geq 0$$

for $i = 1, 2$, we have

$$\left. \begin{array}{l} \mathbf{x}_1(0) + \mathbf{x}_2(0) \\ \mathbf{u}_1(t) + \mathbf{u}_2(t), t \geq 0 \end{array} \right\} \rightarrow \mathbf{y}_1(t) + \mathbf{y}_2(t), \quad t \geq 0 \quad (\text{additivity})$$

and

$$\left. \begin{array}{l} \alpha \mathbf{x}_1(0) \\ \alpha \mathbf{u}_1(t), t \geq 0 \end{array} \right\} \rightarrow \alpha \mathbf{y}_1(t), \quad t \geq 0 \quad (\text{homogeneity})$$

for any real constant α , where we have assumed $t_0 = 0$. The first property is called the *additivity* property, the second, the *homogeneity* property. These two properties can be combined as

$$\left. \begin{array}{l} \alpha_1 \mathbf{x}_1(0) + \alpha_2 \mathbf{x}_2(0) \\ \alpha_1 \mathbf{u}_1(t) + \alpha_2 \mathbf{u}_2(t), t \geq 0 \end{array} \right\} \rightarrow \alpha_1 \mathbf{y}_1(t) + \alpha_2 \mathbf{y}_2(t), \quad t \geq 0$$

for any real constants α_1 and α_2 and is called the *superposition property*. A system is called a nonlinear system if the superposition property does not hold.

If the input $\mathbf{u}(t)$ is identically zero for $t \geq t_0 = 0$, then the output will be excited exclusively by the initial state $\mathbf{x}(0)$. This output is called the *zero-input response* and will be denoted by \mathbf{y}_{zi} or

$$\left. \begin{array}{l} \mathbf{x}(0) \\ \mathbf{u}(t) \equiv \mathbf{0}, \quad t \geq 0 \end{array} \right\} \rightarrow \mathbf{y}_{zi}(t), \quad t \geq 0$$

If the initial state $\mathbf{x}(0)$ is zero, then the output will be excited exclusively by the input. This output is called the *zero state response* and will be denoted by \mathbf{y}_{zs} or

$$\left. \begin{array}{l} \mathbf{x}(0) = \mathbf{0} \\ \mathbf{u}(t), \quad t \geq 0 \end{array} \right\} \rightarrow \mathbf{y}_{zs}(t), \quad t \geq 0$$

The additivity property implies

$$\text{Output excited by } \left\{ \begin{array}{l} \mathbf{x}(0) \\ \mathbf{u}(t), \quad t \geq 0 \end{array} \right\} = \text{output excited by } \left\{ \begin{array}{l} \mathbf{x}(0) \\ \mathbf{u}(t) \equiv \mathbf{0}, \quad t \geq 0 \end{array} \right\} + \text{output excited by } \left\{ \begin{array}{l} \mathbf{x}(0) = \mathbf{0} \\ \mathbf{u}(t), \quad t \geq 0 \end{array} \right\}$$

or

$$\text{Response} = \text{zero-input response} + \text{zero-state response}$$

Thus the response of every linear time-invariant system can be decomposed into the zero-state response and the zero-input response. Furthermore, the two responses can be studied separately, and their sum yields the total response. For nonlinear time-invariant systems, the total response can be very different from the sum of the zero-input response and zero-state response. Therefore, we cannot separate the zero-input and zero-state responses in studying nonlinear time-invariant systems.

If a time-invariant system is linear, then the additivity and homogeneity properties apply to zero-state responses. To be more specific, if $\mathbf{x}(0) = \mathbf{0}$, then the output will be excited exclusively by the input and the state-input-output equation can be simplified as $\{u_i(t) \rightarrow y_i(t)\}$. If the system is time invariant, then $\{u_i(t - t_1) \rightarrow y_i(t - t_1)\}$, for any t_1 . If the system is linear, then we have $\{u_1 + u_2 \rightarrow y_1 + y_2\}$ and $\{\alpha u_i \rightarrow \alpha y_i\}$ for all α and all u_i . A similar remark applies to zero-input responses of any linear time-invariant system.

Convolution

We develop a mathematical equation to describe zero-state responses of linear time-invariant systems. In this study the initial state $\mathbf{x}(0)$ is assumed to be zero, and the output $y(t)$ for $t \geq 0$ is excited exclusively by the input $u(t)$ for $t \geq 0$. If $\mathbf{x}(0) = \mathbf{0}$, the system is said to be *initially relaxed* at $t = 0$. Note that if $u(t) = 0$ for all $t < 0$, then the system is initially relaxed at $t = 0$.

Consider an input $u(t)$ applied from $t = 0$ onward. Then it can be approximated by

$$u(t) \approx \sum_{i=0}^{\infty} u(t_i) \delta_{\Delta}(t - t_i) \Delta$$

as discussed in the preceding subsection. Let $g_\Delta(t)$ be the output at time t excited by the pulse $u(t) = \delta_\Delta(t - 0) = \delta_\Delta(t)$ applied at time $t_0 = 0$. Then we have

$$\begin{aligned} \delta_\Delta(t) &\rightarrow g_\Delta(t) && \text{(definition)} \\ \delta_\Delta(t - t_i) &\rightarrow g_\Delta(t - t_i) && \text{(time shifting)} \\ \delta_\Delta(t - t_i)u(t_i)\Delta &\rightarrow g_\Delta(t - t_i)u(t_i)\Delta && \text{(homogeneity)} \\ \sum_{i=0}^{\infty} \delta_\Delta(t - t_i)u(t_i)\Delta &\rightarrow \sum_{i=0}^{\infty} g_\Delta(t - t_i)u(t_i)\Delta && \text{(additivity)} \end{aligned}$$

Thus the output $y(t)$, for $t \geq 0$, excited by the input $u(t)$, for $t \geq 0$, can be approximated by

$$y(t) \approx \sum_{i=0}^{\infty} g_\Delta(t - t_i)u(t_i)\Delta \quad (2.3)$$

As Δ approaches zero, the pulse $\delta_\Delta(t)$ becomes an *impulse* at $t = 0$ and $g_\Delta(t)$ becomes $g(t)$, which is the output excited by an impulse applied at $t = 0$. We call $g(t)$ the *impulse response*. As Δ approaches zero, the approximation in (2.3) becomes an equality, the summation becomes an integration, the discrete t_i becomes a continuous variable and will be denoted by τ , and Δ can be written as $d\tau$. Thus (2.3) becomes, as $\Delta \rightarrow 0$,

$$y(t) = \int_{\tau=0}^{\infty} g(t - \tau)u(\tau)d\tau \quad (2.4)$$

This describes the zero-state response of an LTI system.

If a system is causal, the output will not appear before an input is applied. The impulse response is the output excited by $\delta(t)$ which is zero for all $t < 0$. Thus we have

$$\text{causal} \iff g(t) = 0 \quad \text{for all } t < 0$$

This is in fact the necessary and sufficient condition for a system to be causal. If $g(t) = 0$ for all $t < 0$, then $g(t - \tau) = 0$ for all $\tau > t$. Thus if a system is causal, the upper limit ∞ in the integration of (2.4) can be replaced by t and (2.4) becomes

$$y(t) = \int_{\tau=0}^t g(t - \tau)u(\tau)d\tau$$

This is called the *integral convolution*. We develop a different form. Define a new integration variable $\bar{\tau}$ by $\bar{\tau} := t - \tau$, where t is fixed. Then we have $d\tau = -d\bar{\tau}$ and

$$y(t) = \int_{\bar{\tau}=t}^0 g(\bar{\tau})u(t - \bar{\tau})(-d\bar{\tau}) = \int_{\bar{\tau}=0}^t g(\bar{\tau})u(t - \bar{\tau})(d\bar{\tau})$$

Thus we have, by renaming the dummy variable $\bar{\tau}$ as τ ,

$$y(t) = \int_{\tau=0}^t g(t - \tau)u(\tau)d\tau = \int_{\tau=0}^t g(\tau)u(t - \tau)d\tau \quad (2.5)$$

Thus the *integral convolution* has two forms or the commutative property in which the role of g and u can be interchanged. Every linear time-invariant (LTI) causal system can be described by such an equation, but it describes only the zero-state response. In other words, in using (2.5), the system must be initially relaxed.

In the derivation of (2.5), the condition of lumpedness is not used. Therefore, any lumped or distributed LTI system has such a description. This description is developed using only the time-shifting, additivity and homogeneity properties; therefore, every LTI system—be it an electrical system, a mechanical system, a chemical process, or any other system—has such a description.

Example 2.3.1 Consider the unit-time delay system studied in Example 2.2.1. It is time-invariant and linear. Thus it can be described by a convolution. If we apply $u(t) = \delta(t)$ to the system, then the output is $\delta(t - 1)$. Thus the impulse response of the system is $g(t) = \delta(t - 1)$ and the system can be described by

$$y(t) = \int_{\tau=0}^t g(t - \tau)u(\tau)d\tau = \int_{\tau=0}^t \delta(t - 1 - \tau)u(\tau)d\tau$$

which reduces to, using the sifting property of impulses,

$$y(t) = u(\tau)|_{t-\tau-1=0} = u(\tau)|_{\tau=t-1} = u(t - 1)$$

Thus the convolution description reduces to simply the definition of the system. Recall that the convolution describes only zero-state responses. The state at $t = 0$ of the unit-time delay system is, as discussed in Example 2.2.1, $u(t)$ for all t in $[-1, 0]$. Thus the unit-time delay system is initially relaxed at $t = 0$ if $u(t) = 0$ for all t in $[-1, 0]$. Note that the input $u(t)$ for $t < -1$ is irrelevant; it can be zero or nonzero. For convenience, in using a convolution, we may simply assume $u(t) = 0$ for all $t < 0$.

Example 2.3.2 Consider the unity feedback system shown in Fig. 2.5(a). It consists of a multiplier with gain a and a unit-time delay element. It is a SISO system. Let $r(t)$, called the *reference input*, be the input of the feedback system. If $r(t) = \delta(t)$, then the output is the impulse response of the feedback system and equals

$$g_f(t) = a\delta(t - 1) + a^2\delta(t - 2) + a^3\delta(t - 3) + \dots = \sum_{i=1}^{\infty} a^i\delta(t - i) \quad (2.6)$$

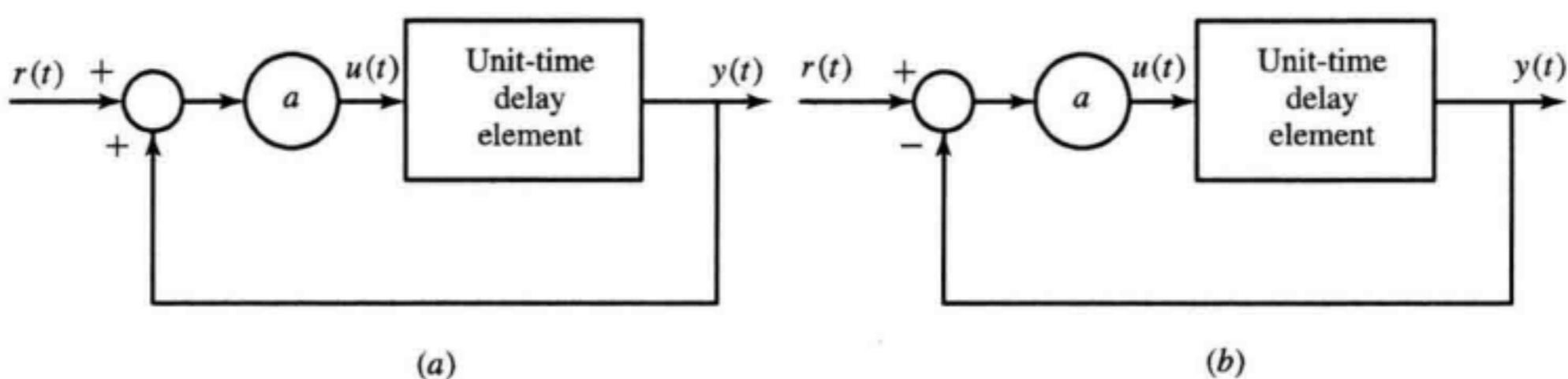


FIGURE 2.5 Positive and negative feedback systems.

Let $r(t)$ be any input with $r(t) = 0$ for all $t < 0$, then the output is given by

$$\begin{aligned} y(t) &= \int_0^t g_f(t-\tau)r(\tau)d\tau = \sum_{i=1}^{\infty} a^i \int_0^t \delta(t-\tau-i)r(\tau)d\tau \\ &= \sum_{i=1}^{\infty} a^i r(\tau) \Big|_{\tau=t-i} = \sum_{i=1}^{\infty} a^i r(t-i) \end{aligned}$$

Because the unit-time delay system is distributed, so is the feedback system.

Transfer function

The Laplace transform is an important tool in the study of linear time-invariant (LTI) systems. Let $\hat{y}(s)$ be the Laplace transform of $y(t)$, that is,

$$\hat{y}(s) := \mathcal{L}[y(t)] := \int_0^{\infty} y(t)e^{-st} dt$$

Throughout this text, we use a variable with a circumflex to denote the Laplace transform of the variable. Substituting (2.4) and interchanging the order of integrations yield

$$\begin{aligned} \hat{y}(s) &= \int_{t=0}^{\infty} \left(\int_{\tau=0}^{\infty} g(t-\tau)u(\tau)d\tau \right) e^{-s(t-\tau)} e^{-st} dt \\ &= \int_{\tau=0}^{\infty} \left(\int_{t=0}^{\infty} g(t-\tau)e^{-s(t-\tau)} dt \right) u(\tau)e^{-s\tau} d\tau \end{aligned}$$

which becomes, after introducing the new variable $v := t - \tau$ for the inner integration where τ is fixed,

$$\hat{y}(s) = \int_{\tau=0}^{\infty} \left(\int_{v=-\tau}^{\infty} g(v)e^{-sv} dv \right) u(\tau)e^{-s\tau} d\tau$$

Using the causality condition to replace the lower integration limit inside the parentheses from $v = -\tau$ to $v = 0$, the integration becomes independent of τ and the double integration becomes

$$\hat{y}(s) = \left(\int_{v=0}^{\infty} g(v)e^{-sv} dv \right) \left(\int_{\tau=0}^{\infty} u(\tau)e^{-s\tau} d\tau \right)$$

or

$$\hat{y}(s) = \hat{g}(s)\hat{u}(s) \tag{2.7}$$

where

$$\hat{g}(s) = \mathcal{L}[g(t)] = \int_0^{\infty} g(t)e^{-st} dt$$

is called the *transfer function* of the system. Thus the transfer function is the Laplace transform of the impulse response and, conversely, the impulse response is the inverse

Laplace transform of the transfer function. We see that the Laplace transform transforms the integral convolution into the algebraic equation in (2.7).

The transfer function is defined as the Laplace transform of the impulse response. Using (2.7), we can also define it as

$$\hat{g}(s) = \frac{\hat{y}(s)}{\hat{u}(s)} = \left. \frac{\mathcal{L}[\text{output}]}{\mathcal{L}[\text{input}]} \right|_{\text{initially relaxed}} \quad (2.8)$$

Most transfer functions will be computed using this definition.

Example 2.3.3 Consider the unit-time delay system studied in Example 2.2.1. Its impulse response is $\delta(t - 1)$. Therefore, its transfer function is

$$\hat{g}(s) = \mathcal{L}[\delta(t - 1)] = \int_0^\infty \delta(t - 1)e^{-st} dt = e^{-st} \Big|_{t=1} = e^{-s}$$

This transfer function is an irrational function of s . In the Laplace transform domain, the input and output are related by $\hat{y}(s) = \hat{g}(s)\hat{u}(s) = e^{-s}\hat{u}(s)$.

Example 2.3.4 Consider the feedback system shown in Fig. 2.5(a). Its impulse response was computed in (2.6) as

$$g_f(t) = \sum_{i=1}^{\infty} a^i \delta(t - i)$$

Because $\mathcal{L}[\delta(t - i)] = e^{-is}$, the Laplace transform of $g_f(t)$ is

$$\hat{g}_f(s) = \mathcal{L}[g_f(t)] = \sum_{i=1}^{\infty} a^i e^{-is} = ae^{-s} \sum_{i=0}^{\infty} (ae^{-s})^i$$

Using $\sum_{i=0}^{\infty} b^i = 1/(1 - b)$, for $|b| < 1$, we can express the infinite series in closed form as

$$\hat{g}_f(s) = \frac{ae^{-s}}{1 - ae^{-s}}$$

Next we use (2.8) to compute the transfer function. From Fig. 2.5(a), we have, in the Laplace transform domain,

$$\hat{u}(s) = a(\hat{r}(s) + \hat{y}(s)) = a\hat{r}(s) + ae^{-s}\hat{u}(s)$$

where we have used $\hat{y}(s) = e^{-s}\hat{u}(s)$. Simple manipulation yields

$$\hat{u}(s) = \frac{a}{1 - ae^{-s}} \hat{r}(s)$$

and

$$\hat{y}(s) = e^{-s}\hat{u}(s) = \frac{ae^{-s}}{1 - ae^{-s}} \hat{r}(s)$$

Thus the transfer function, from r to y , of the feedback system is

$$\hat{g}_f(s) = \frac{\hat{y}(s)}{\hat{r}(s)} = \frac{ae^{-s}}{1 - ae^{-s}}$$

which is the same as the one computed earlier.

State-space equation

Every linear time-invariant lumped system with input $u(t)$ and output $y(t)$ can be described by a set of equations of form

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t) \quad (2.9)$$

$$y(t) = \mathbf{c}\mathbf{x}(t) + du(t) \quad (2.10)$$

where $\dot{\mathbf{x}}(t) := d\mathbf{x}(t)/dt$. If the system has n state variables, then \mathbf{x} is an $n \times 1$ vector. In order for the matrices in (2.9) and (2.10) to be compatible, \mathbf{A} , \mathbf{b} , \mathbf{c} , and d must be $n \times n$, $n \times 1$, $1 \times n$, and 1×1 matrices. All entries of the four matrices are real numbers, independent of time. Equation (2.9), called a *state equation*, actually consists of a set of n first-order differential equations. Equation (2.10), called an *output equation*, is an algebraic equation. The constant d is called the *direct transmission part*. The set of two equations is called an n -dimensional *state-space (ss)* equation. Linear time invariant distributed systems which have an infinite number of state variables cannot be described by a state-space equation of the form shown.

The convolution description was developed from the conditions of time-invariance and the linearity. The development of the state-space equation from the conditions, however, is not as simple and will not be attempted. We will simply accept it as a fact.

Applying the Laplace transform to (2.9) and (2.10) yields¹

$$s\hat{\mathbf{x}}(s) - \mathbf{x}(0) = \mathbf{A}\hat{\mathbf{x}}(s) + \mathbf{b}\hat{u}(s)$$

$$\hat{y}(s) = \mathbf{c}\hat{\mathbf{x}}(s) + d\hat{u}(s)$$

which implies

$$s\hat{\mathbf{x}}(s) - \mathbf{A}\hat{\mathbf{x}}(s) = (s\mathbf{I} - \mathbf{A})\hat{\mathbf{x}}(s) = \mathbf{x}(0) + \mathbf{b}\hat{u}(s)$$

where \mathbf{I} is the unit matrix of the same order as \mathbf{A} and has the property $\hat{\mathbf{x}}(s) = \mathbf{I}\hat{\mathbf{x}}(s)$. Without introducing \mathbf{I} , $(s - \mathbf{A})$ is not defined. Premultiplying $(s\mathbf{I} - \mathbf{A})^{-1}$ to the preceding equation yields

$$\hat{\mathbf{x}}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}(0) + (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}\hat{u}(s)$$

and

$$\hat{y}(s) = \mathbf{c}\hat{\mathbf{x}}(s) + d\hat{u}(s) = \mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}(0) + \mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}\hat{u}(s) + d\hat{u}(s) \quad (2.11)$$

¹If $\hat{x}(s) = \mathcal{L}[x(t)]$, then $\mathcal{L}[\dot{x}(t)] = s\hat{x}(s) - x(0)$.

They are algebraic equations. Equation (2.11) reveals the fact that the response of an LTI system can be decomposed as the zero-state response and the zero-input response. If the initial state $\mathbf{x}(0)$ is zero, then (2.11) reduces to

$$\hat{\mathbf{y}}(s) = [\mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} + d]\hat{\mathbf{u}}(s)$$

Comparing this with (2.7) yields

$$\hat{\mathbf{g}}(s) = \mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} + d \quad (2.12)$$

This relates the transfer function (an external description) and the state-space equation (the internal description).

2.3.1 Multi-input Multi-output Case

We extend the discussion in the preceding section for the single-input single-out (SISO) case to the multi-input multi-output (MIMO) case. If an LTI system has p input terminals and q output terminals, then (2.5) can be extended to

$$\mathbf{y}(t) = \int_{\tau=0}^t \mathbf{G}(t-\tau)\mathbf{u}(\tau) d\tau \quad (2.13)$$

where $\mathbf{y}(t)$ is a $q \times 1$ output vector, $\mathbf{u}(t)$ is a $p \times 1$ input vector, and

$$\mathbf{G}(t) = \begin{bmatrix} g_{11}(t) & g_{12}(t) & \cdots & g_{1p}(t) \\ g_{21}(t) & g_{22}(t) & \cdots & g_{2p}(t) \\ \vdots & \vdots & & \vdots \\ g_{q1}(t) & g_{q2}(t) & \cdots & g_{qp}(t) \end{bmatrix}$$

The function $g_{ij}(t)$ is the response at time t at the i th output terminal excited by an impulse applied at time 0 at the j th input terminal, with the inputs at other terminals being identically zero. That is, $g_{ij}(t)$ is the impulse response between the j th input terminal and the i th output terminal. Thus \mathbf{G} is called the *impulse response matrix* of the system. We stress once again that if a system is described by (2.13), the system is linear, time-invariant, causal, and initially relaxed at $t_0 = 0$.

The application of the Laplace transform to (2.13) yields

$$\hat{\mathbf{y}}(s) = \hat{\mathbf{G}}(s)\hat{\mathbf{u}}(s) \quad (2.14)$$

or

$$\begin{bmatrix} \hat{y}_1(s) \\ \hat{y}_2(s) \\ \vdots \\ \hat{y}_q(s) \end{bmatrix} = \begin{bmatrix} \hat{g}_{11}(s) & \hat{g}_{12}(s) & \cdots & \hat{g}_{1p}(s) \\ \hat{g}_{21}(s) & \hat{g}_{22}(s) & \cdots & \hat{g}_{2p}(s) \\ \vdots & \vdots & & \vdots \\ \hat{g}_{q1}(s) & \hat{g}_{q2}(s) & \cdots & \hat{g}_{qp}(s) \end{bmatrix} \begin{bmatrix} \hat{u}_1(s) \\ \hat{u}_2(s) \\ \vdots \\ \hat{u}_p(s) \end{bmatrix}$$

where $\hat{g}_{ij}(s)$ is the transfer function from the j th input to the i th output. The $q \times p$ matrix $\hat{\mathbf{G}}(s)$ is called the *transfer-function matrix* or, simply, *transfer matrix* of the system.

Equations (2.13) and (2.14) are applicable to any LTI lumped or distributed system. If a system is lumped as well, then it can also be described by a state-space equation of form

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{Ax}(t) + \mathbf{Bu}(t) \\ \mathbf{y}(t) &= \mathbf{Cx}(t) + \mathbf{Du}(t)\end{aligned}\quad (2.15)$$

For a system with p inputs, q outputs, and n state variables, \mathbf{A} , \mathbf{B} , \mathbf{C} , and \mathbf{D} are, respectively, $n \times n$, $n \times p$, $q \times n$, and $q \times p$ constant matrices.

The MIMO case of (2.12) is

$$\hat{\mathbf{G}}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D} \quad (2.16)$$

Its derivation is identical to the SISO case and is omitted.

2.4 Linear Time-Varying Systems

We now modify (2.5) so that it can be applied to linear, causal but time-varying systems. For a time-invariant system, if we know the response excited by an impulse applied at time 0, then we know the response excited by an impulse applied at any t . This is not the case for the time-varying case. The response excited by an impulse applied at time t_1 is generally different from the response excited by an impulse applied at $t_2 \neq t_1$. Thus the impulse response of a time-varying system must be a function of two variables. Let $g(t, \tau)$ be the output at time t excited by an impulse applied at time τ of a linear time-varying system. If the system is causal, then we have $g(t, \tau) = 0$, for all $t < \tau$. Then the system with input $u(t)$ and output $y(t)$ can be described by

$$y(t) = \int_{\tau=t_0}^t g(t, \tau)u(\tau) d\tau$$

assuming that the system is initially relaxed at t_0 . Note that we can no longer assume $t_0 = 0$. For a MIMO linear time-varying system, we have

$$\mathbf{y}(t) = \int_{\tau=t_0}^t \mathbf{G}(t, \tau)\mathbf{u}(\tau) d\tau \quad (2.17)$$

where $g_{ij}(t, \tau)$, the ij th entry of \mathbf{G} , is the impulse response between the j th input terminal and the i th output terminal.

To modify (2.15) to describe linear time-varying lumped system is fairly simple. Changing constant matrices into time-varying matrices yields

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}(t)\mathbf{x}(t) + \mathbf{D}(t)\mathbf{u}(t)\end{aligned}\quad (2.18)$$

Thus the use of state-space equations to study time-varying systems is much simpler than the use of convolutions.

The Laplace transform is an important tool in studying time-invariant systems but is not used in studying time-varying systems. The Laplace transform of $g(t, \tau)$ is a

function of two variables and $\mathcal{L}[\mathbf{A}(t)\mathbf{x}(t)] \neq \mathcal{L}[\mathbf{A}(t)]\mathcal{L}[\mathbf{x}(t)]$, thus the Laplace transform does not offer any advantage and is not used in studying time-varying systems.

2.4.1 Linearization

Most physical systems are nonlinear and time varying. Some of them can be described by the nonlinear differential equation of form

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{h}(\mathbf{x}(t), \mathbf{u}(t), t) \\ \mathbf{y}(t) &= \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t)\end{aligned}\quad (2.19)$$

where \mathbf{h} and \mathbf{f} are nonlinear functions. The behavior of such equations can be very complicated, and its study is beyond the scope of this text.

Some nonlinear equations, however, can be approximated by linear equations under certain conditions. Suppose for some input function $\mathbf{u}_o(t)$ and some initial state, $\mathbf{x}_o(t)$ is the solution of (2.19), that is,

$$\dot{\mathbf{x}}_o(t) = \mathbf{h}(\mathbf{x}_o(t), \mathbf{u}_o(t), t) \quad (2.20)$$

Now suppose the input is perturbed slightly to become $\mathbf{u}_o(t) + \bar{\mathbf{u}}(t)$ and the initial state is also perturbed only slightly. For some nonlinear equations, the corresponding solution may differ from $\mathbf{x}_o(t)$ only slightly. In this case, the solution can be expressed as $\mathbf{x}_o(t) + \bar{\mathbf{x}}(t)$ with $\bar{\mathbf{x}}(t)$ small for all t .² Under this assumption, we can expand (2.19) as

$$\begin{aligned}\dot{\mathbf{x}}_o(t) + \dot{\bar{\mathbf{x}}}(t) &= \mathbf{h}(\mathbf{x}_o(t) + \bar{\mathbf{x}}(t), \mathbf{u}_o(t) + \bar{\mathbf{u}}(t), t) \\ &= \mathbf{h}(\mathbf{x}_o(t), \mathbf{u}_o(t), t) + \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \bar{\mathbf{x}}(t) + \frac{\partial \mathbf{h}}{\partial \mathbf{u}} \bar{\mathbf{u}}(t) + \dots\end{aligned}\quad (2.21)$$

where, for $\mathbf{h} = [h_1 \ h_2 \ h_3]'$, $\mathbf{x} = [x_1 \ x_2 \ x_3]'$, and $\mathbf{u} = [u_1 \ u_2]'$, we have

$$\begin{aligned}\mathbf{A}(t) := \frac{\partial \mathbf{h}}{\partial \mathbf{x}} &:= \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \frac{\partial h_1}{\partial x_3} \\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} & \frac{\partial h_2}{\partial x_3} \\ \frac{\partial h_3}{\partial x_1} & \frac{\partial h_3}{\partial x_2} & \frac{\partial h_3}{\partial x_3} \end{bmatrix} \\ \mathbf{B}(t) := \frac{\partial \mathbf{h}}{\partial \mathbf{u}} &:= \begin{bmatrix} \frac{\partial h_1}{\partial u_1} & \frac{\partial h_1}{\partial u_2} \\ \frac{\partial h_2}{\partial u_1} & \frac{\partial h_2}{\partial u_2} \\ \frac{\partial h_3}{\partial u_1} & \frac{\partial h_3}{\partial u_2} \end{bmatrix}\end{aligned}$$

They are called *Jacobians*. Because \mathbf{A} and \mathbf{B} are computed along the two time functions $\mathbf{x}_o(t)$ and $\mathbf{u}_o(t)$, they are, in general, functions of t . Using (2.20) and neglecting higher powers of $\bar{\mathbf{x}}$ and $\bar{\mathbf{u}}$, we can reduce (2.21) to

$$\dot{\bar{\mathbf{x}}}(t) = \mathbf{A}(t)\bar{\mathbf{x}}(t) + \mathbf{B}(t)\bar{\mathbf{u}}(t)$$

This is a linear time-varying state equation. The equation $\mathbf{y}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t)$ can be similarly linearized. This linearization technique is often used in practice to obtain linear equations.

²This is not true in general. For some nonlinear equations, very small difference in initial states will generate completely different solutions, yielding the phenomenon of *chaos*.

2.5 RLC Circuits—Comparisons of Various Descriptions

In this section, we discuss how to develop a state-space equation, a high-order differential equation, a transfer function, and a convolution to describe an RLC circuit. We also compare the four descriptions. Before proceeding, we mention that most systems are built by interconnecting a number of components. If every component is linear and time invariant, then so is the overall system. However, no physical component is linear and time invariant in the mathematical sense. For example, a resistor with resistance R is not linear because if the applied voltage is very large, the resistor may burn out. However, within its power limitation, the resistor with its voltage $v(t)$ and current $i(t)$ related by $v(t) = Ri(t)$ is a linear element. Likewise, disregarding saturation, the charge $Q(t)$ stored in a capacitor is related to the applied voltage $v(t)$ by $Q(t) = Cv(t)$, where C is the capacitance, which implies $i(t) = Cdv(t)/dt$; the flux $F(t)$ generated by an inductor is related to its current $i(t)$ by $F(t) = Li(t)$, where L is the inductance, which implies $v(t) = Ldi(t)/dt =: L\dot{i}(t)$. They are linear elements. The resistance R , capacitance C , and inductance L probably will change their values after 100 years, thus they are not time-invariant. However, their values will remain constant within a number of years. Thus they may be considered as linear and time-invariant. If all resistors, capacitors, and inductors of RLC circuits such as the one in Fig. 2.6(a) are so modeled, then the circuits are linear and time invariant. In this section, we will develop four types of equations to describe the circuit in Fig. 2.6(a).

State-space equation

Consider the RLC circuit shown in Figure 2.6(a) with a voltage source as its input and the voltage across the capacitor as its output. It consists of one resistor with resistance 3 ohms (Ω), two inductors with inductances 5 and 4 henrys (H), and one capacitor with capacitance 2 farads (F). To develop a state-space equation, we must first select state variables. For RLC circuits, state variables are associated with energy-storage elements. A resistor is a memoryless element and cannot store energy;³ thus its variable (current or voltage) cannot be selected as a state variable. A capacitor can store energy in its electric field, and its variable can be selected as a state variable. If we select the

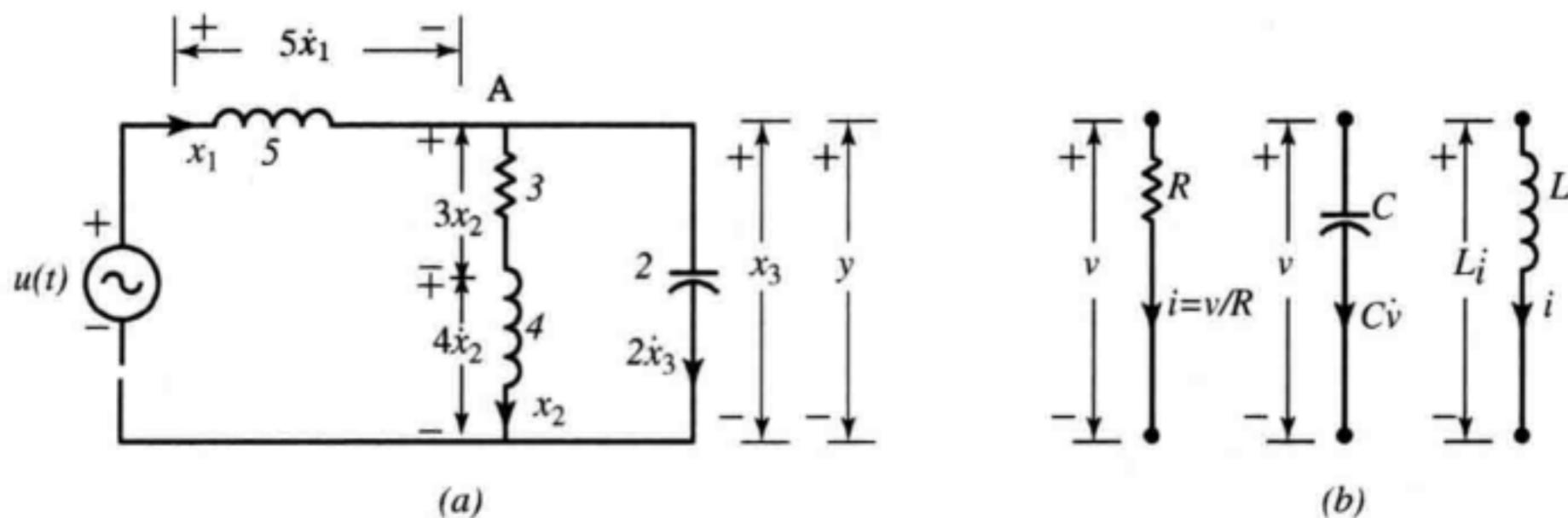


FIGURE 2.6 (a) an RLC circuit. (b) Linear elements.

³ Its energy is all dissipated into heat.

capacitor voltage $v(t)$ as a state variable, then its current is $Cdv(t)/dt =: C\dot{v}(t)$ as shown in Fig. 2.6(b). If we select the capacitor current as a state variable, then its voltage is an integration and is not used. An inductor can store energy in its magnetic field, and its variable can be selected as a state variable. If we select the inductor current $i(t)$ as a state variable, then its voltage is $Ldi(t)/dt = Li(t)$ as shown in Fig. 2.6(b). In the assignment, it is important to assign the polarity of a voltage and the direction of a current, otherwise the assignment is not complete.

Procedure of developing state-space equations for RLC circuits

1. Assign all capacitor voltages and all inductor currents as state variables.⁴ If the voltage of a capacitor with capacitance C is assigned as $x_i(t)$, then its current is $C\dot{x}_i(t)$. If the current of an inductor with inductance L is assigned as $x_j(t)$, then its voltage is $L\dot{x}_j(t)$.
2. Use Kirchhoff's current or voltage law to express the current or voltage of every resistor in terms of state variables (but not their derivatives) and, if necessary, the input. If the expression is for the current (voltage), then its multiplication (division) by its resistance yields the voltage (current).
3. Use Kirchhoff's current or voltage law to express each $\dot{x}_i(t)$ in terms of state variables and input.

For the circuit shown in Fig. 2.6(a), we assign the 5-H inductor current as $x_1(t)$. Then its voltage is $5\dot{x}_1(t)$ as shown. Next we assign the 4-H inductor current as $x_2(t)$. Then its voltage is $4\dot{x}_2(t)$. Finally, we assign the 2-F capacitor voltage as $x_3(t)$. Then its current is $2\dot{x}_3(t)$. Thus the network has three state variables. Next we express the 3Ω resistor's voltage or current in terms of state variables. Because the resistor is in series connection with the 4-H inductor, its current is $x_2(t)$. Thus the voltage across the 3Ω resistor is $3x_2(t)$. Note that finding first the voltage across the resistor will be more complicated. This completes the first two steps of the procedure.

Next applying Kirchhoff's voltage law along the outer loop of the circuit yields

$$5\dot{x}_1(t) + x_3(t) - u(t) = 0$$

which implies

$$\dot{x}_1(t) = -0.2x_3(t) + 0.2u(t) \quad (2.22)$$

Applying Kirchhoff's voltage law along the right-hand-side loop yields

$$x_3(t) - 4\dot{x}_2(t) - 3x_2(t) = 0$$

which implies

$$\dot{x}_2(t) = -0.75x_2(t) + 0.25x_3(t) \quad (2.23)$$

⁴There are some exceptions. See Problems 2.11–2.13. In essence, all state variables selected must be able to vary independently.

Finally, applying Kirchhoff's current law to the node denoted by A yields

$$x_1(t) - x_2(t) - 2\dot{x}_3(t) = 0$$

which implies

$$\dot{x}_3(t) = 0.5x_1(t) - 0.5x_2(t) \quad (2.24)$$

From Figure 2.6(a), we have

$$y(t) = x_3(t) \quad (2.25)$$

The equations from (2.22) through (2.25) can be arranged in matrix form as

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & -0.2 \\ 0 & -0.75 & 0.25 \\ 0.5 & -0.5 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} 0.2 \\ 0 \\ 0 \end{bmatrix} u(t) \quad (2.26)$$

$$y(t) = [0 \ 0 \ 1] \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + 0 \times u(t) \quad (2.27)$$

or, using matrix symbols,

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t) \quad (2.28)$$

$$y(t) = \mathbf{c}\mathbf{x}(t) + du(t) \quad (2.29)$$

with $\mathbf{x} = [x_1 \ x_2 \ x_3]'$ and

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & -0.2 \\ 0 & -0.75 & 0.25 \\ 0.5 & -0.5 & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0.2 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{c} = [0 \ 0 \ 1], \quad d = 0$$

This state-space of dimension 3 describes the circuit in Fig. 2.6(a).

High-order differential equation

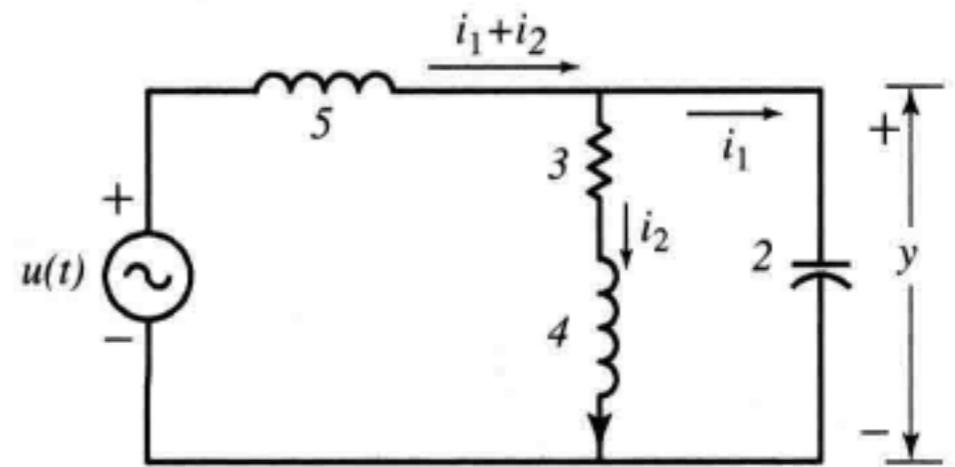
Consider the circuit in Fig. 2.6(a) and replotted in Fig. 2.7.⁵ Let the current passing through the 2-F capacitor be denoted by $i_1(t)$, and let the current passing through the series connection of the 3-Ω resistor and the 4-H inductor be denoted by $i_2(t)$. Then we have

$$i_1(t) = 2\dot{y}(t) \quad (2.30)$$

$$y(t) = 3i_2(t) + 4\dot{i}_2(t) \quad (2.31)$$

The current passing through the 5-H inductor is $i_1(t) + i_2(t)$. Thus the voltage across the inductor is $5i_1(t) + 5\dot{i}_2(t)$. Applying Kirchhoff's voltage law around the outer loop

⁵This part may be skipped without loss of continuity.

FIGURE 2.7 RLC circuit.

of Fig. 2.7 yields

$$5\dot{i}_1(t) + 5\dot{i}_2(t) + y(t) - u(t) = 0 \quad (2.32)$$

In order to develop a differential equation to relate $u(t)$ and $y(t)$, we must eliminate $i_1(t)$ and $i_2(t)$ from (2.30) through (2.32). First we substitute (2.30) into (2.32) to yield

$$10\ddot{y}(t) + 5\dot{i}_2(t) + y(t) = u(t) \quad (2.33)$$

where $\ddot{y}(t) := d^2y(t)/dt^2$, and then differentiate it to yield

$$10y^{(3)}(t) + 5\ddot{i}_2(t) + \dot{y}(t) = \dot{u}(t) \quad (2.34)$$

where $y^{(3)}(t) := d^3y(t)/dt^3$. The summation of (2.33) multiplied by 3 and (2.34) multiplied by 4 yields

$$40y^{(3)}(t) + 5(4\ddot{i}_2(t) + 3\dot{i}_2(t)) + 30\ddot{y}(t) + 4\dot{y}(t) + 3y(t) = 4\dot{u}(t) + 3u(t)$$

which becomes, after substituting the derivative of (2.31),

$$40y^{(3)}(t) + 30\ddot{y}(t) + 9\dot{y}(t) + 3y(t) = 4\dot{u}(t) + 3u(t) \quad (2.35)$$

This is a third-order linear differential equation with constant coefficients. It describes the circuit in Fig. 2.7 or 2.6(a). Note that (2.35) can also be developed using a different method as we will discuss later.

We now compare high-order differential equations and state-space equations:

1. For a simple system that has only one or two state variables, there is not much difference in developing a differential equation or a state-space equation to describe it. However, for a system with three or more state variables, it is generally simpler to develop a state-space equation than a single high-order differential equation because the latter requires us to eliminate intermediate variables as shown in the preceding example. Furthermore, the form of state-space equations is more compact than the form of differential equations.
2. A high-order differential equation is an external description. A state-space equation is an internal description. It describes not only the relationship between the input and output but also the internal variables.
3. High-order differential equations are not suitable for computer computation because of the difficulties in discretizing second, third, and higher-order derivatives. State-space equations involve only the discretization of the first derivative,

thus they are more suitable for computer computation, as we will discuss in Section 4.3.

4. State-space equations can be more easily extended than high-order differential equations to describe nonlinear and time-varying systems.

In view of the preceding reasons, there seem to be no reasons to develop and to study high-order differential equations in this text.

Transfer function

The transfer function of the circuit in Fig. 2.6(a) can be obtained, using (2.12), from its state-space equation in (2.26) and (2.27). It is, however, instructive and simpler to develop directly the transfer function using the concept of impedances as we discuss next.

The voltage $v(t)$ and current $i(t)$ of a resistor with resistance R , a capacitor with capacitance C , and an inductor with inductance L are related, respectively, by

$$v(t) = Ri(t), \quad i(t) = C \frac{dv(t)}{dt}, \quad v(t) = L \frac{di(t)}{dt}$$

as shown in Fig. 2.6(b). Applying the Laplace transform and assuming zero initial conditions, we obtain

$$\hat{v}(s) = R\hat{i}(s), \quad \hat{i}(s) = Cs\hat{v}(s), \quad \hat{v}(s) = Ls\hat{i}(s) \quad (2.36)$$

where $\hat{v}(s)$ and $\hat{i}(s)$ are the Laplace transforms of $v(t)$ and $i(t)$. If we consider the current as the input and the excited voltage as the output, then the transfer functions of R , C , and L are R , $1/Cs$, and Ls , respectively. They are called *transform impedances* or, simply, *impedances*. If we consider the voltage as the input and the current as its output, then their transfer functions are called *admittances*.

Using impedances, the voltage and current of every circuit element can be written as $\hat{v}(s) = Z(s)\hat{i}(s)$ with $Z(s) = R$ for resistors, $Z(s) = 1/Cs$ for capacitors, and $Z(s) = Ls$ for inductors. They involves only multiplications. In other words, the relationship between the input and output of a circuit element is algebraic in the (Laplace) transform domain, whereas it is calculus (differentiation or integration) in the time domain. Thus the former is much simpler. Consequently, the manipulation of impedances is just like the manipulation of resistances. For example, the resistance of the series connection of R_1 and R_2 is $R_1 + R_2$. The resistance of the parallel connection of R_1 and R_2 is $R_1R_2/(R_1 + R_2)$. Likewise, the impedance of the series connection of $Z_1(s)$ and $Z_2(s)$ is $Z_1(s) + Z_2(s)$. The impedance of the parallel connection of $Z_1(s)$ and $Z_2(s)$ is

$$\frac{Z_1(s)Z_2(s)}{Z_1(s) + Z_2(s)}$$

Using these two simple rules, we can readily obtain the transfer function of the circuit in Fig. 2.6(a). Before proceeding, we draw the circuit in Fig. 2.8(a) in the Laplace transform domain or, equivalently, using impedances. The impedance of the series connection of the resistor with impedance 3 and the inductor with impedance $4s$ is $3 + 4s$. The impedance $Z_{AB}(s)$ between the two nodes A and B shown in Fig. 2.8(a) is the parallel

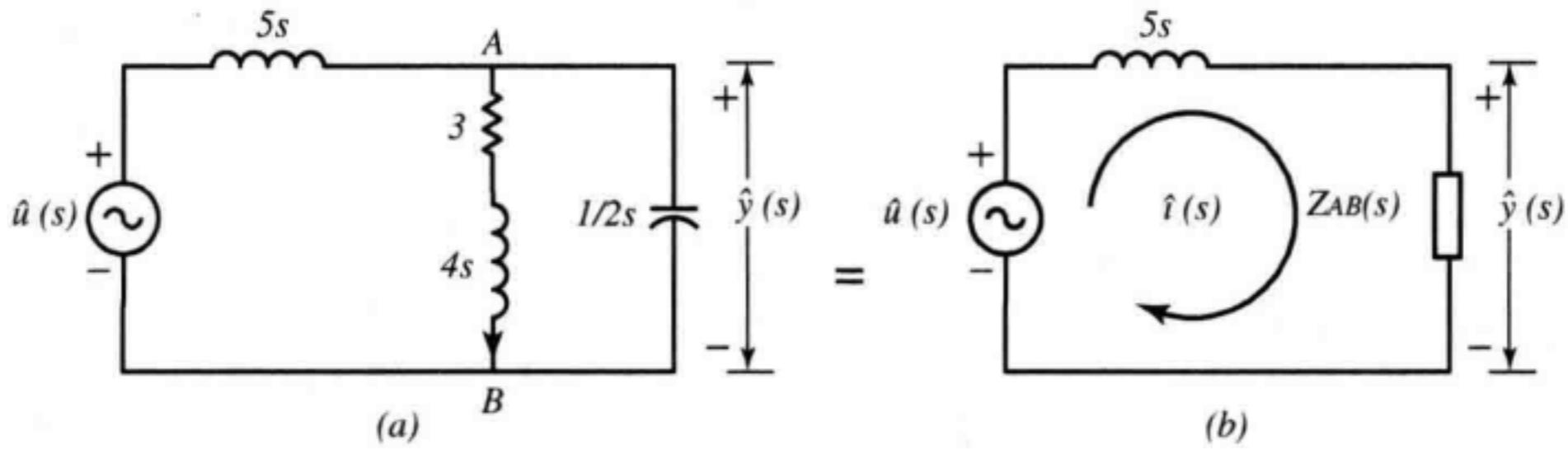


FIGURE 2.8 (a) RLC circuit in the transform domain. (b) Equivalent circuit.

connection of $(4s + 3)$ and the capacitor with impedance $1/2s$ or

$$Z_{AB}(s) = \frac{(4s + 3)(1/2s)}{4s + 3 + 1/2s} = \frac{4s + 3}{8s^2 + 6s + 1}$$

Using $Z_{AB}(s)$, we can simplify the circuit in Fig. 2.8(a) as in Fig. 2.8(b). The current $\hat{i}(s)$ around the loop is given by

$$\hat{i}(s) = \frac{\hat{u}(s)}{5s + Z_{AB}(s)}$$

Thus the voltage $\hat{y}(s)$ across $Z_{AB}(s)$ is

$$\hat{y}(s) = Z_{AB}(s)\hat{i}(s) = \frac{Z_{AB}(s)}{5s + Z_{AB}(s)}\hat{u}(s)$$

and the transfer function from $u(t)$ to $y(t)$ is

$$\hat{g}(s) = \frac{\hat{y}(s)}{\hat{u}(s)} = \frac{Z_{AB}(s)}{5s + Z_{AB}(s)}$$

or

$$\begin{aligned} \hat{g}(s) &= \frac{\frac{4s+3}{8s^2+6s+1}}{5s + \frac{4s+3}{8s^2+6s+1}} = \frac{4s+3}{5s(8s^2+6s+1)+4s+3} \\ &= \frac{4s+3}{40s^3+30s^2+9s+3} \end{aligned} \tag{2.37}$$

This transfer function describes the circuit in Fig. 2.6(a). It is a ratio of two polynomials and is called a *rational transfer function*.

We mention that once we have a transfer function of a system, we can readily obtain its differential equation. Recall that if $\hat{y}(s) = \mathcal{L}[y(t)]$, then $\mathcal{L}[\dot{y}(t)] = s\hat{y}(s) - y(0)$, $\mathcal{L}[\ddot{y}(t)] = s^2\hat{y}(s) - sy(0) - \dot{y}(0)$, and so forth. If all initial conditions are zero, then we have $\mathcal{L}[\dot{y}(t)] = s\hat{y}(s)$, $\mathcal{L}[\ddot{y}(t)] = s^2\hat{y}(s)$, and so forth. We see that one differentiation in the time domain is equivalent to the multiplication by one s in the transform domain,

and the k th derivative in the time domain is equivalent to the multiplication by s^k in the transform domain, that is,

$$s^k \longleftrightarrow \frac{d^k}{dt^k} \quad (2.38)$$

for $k = 1, 2, \dots$. Using this, we can transform a transfer function into a differential equation and vice versa.

Consider the transfer function in (2.37). We write it as

$$(40s^3 + 30s^2 + 9s + 3)\hat{y}(s) = (4s + 3)\hat{u}(s)$$

or

$$40s^3\hat{y}(s) + 30s^2\hat{y}(s) + 9s\hat{y}(s) + 3\hat{y}(s) = 4s\hat{u}(s) + 3\hat{u}(s)$$

which becomes, in the time domain,

$$40y^{(3)}(t) + 30\ddot{y}(t) + 9\dot{y}(t) + 3y(t) = 4\dot{u}(t) + 3u(t)$$

This is the differential equation in (2.35). Thus once a transfer function is obtained, we can readily obtain its differential equation. Conversely, we can also readily obtain a transfer function from a differential equation.

MATLAB contains the functions `ss2tf` and `tf2ss`. The former transforms a state-space (ss) equation into a transfer function (tf), and the latter carries out the converse. For example, for the state-space equation in (2.26) and (2.27), typing in MATLAB

```
a=[ 0 0 -0.2; 0 -0.75 0.25; 0.5 -0.5 0];
b=[ 0.2; 0; 0]; c=[ 0 0 1]; d=0;
[n,de]=ss2tf(a,b,c,d)
```

will yield

$$\begin{aligned} n &= 0 & -0.0000 & 0.1000 & 0.0750 \\ de &= 1.0000 & 0.7500 & 0.2250 & 0.0750 \end{aligned}$$

or

$$\hat{g}(s) = \frac{0.1s + 0.075}{s^3 + 0.75s^2 + 0.225s + 0.075}$$

which is the same as the transfer function in (2.37).

To transform the transfer function in (2.37) into a state-space equation, we type in MATLAB

```
n=[ 4 3]; de=[ 40 30 9 3]; [a,b,c,d]=tf2ss(n,de)
```

which yields

$$\begin{aligned} a &= \begin{bmatrix} -0.7500 & -0.2250 & -0.0750 \\ 1.0000 & 0 & 0 \\ 0 & 1.0000 & 0 \end{bmatrix} \\ b &= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\ c &= [0 \ 0.10000 \ 0.0750] \\ d &= 0 \end{aligned}$$

or

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \begin{bmatrix} -0.75 & -0.225 & -0.075 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u(t) \\ y(t) &= [0 \ 0.1 \ 0.075] \mathbf{x}(t) + 0 \times u(t) \end{aligned} \quad (2.39)$$

This state-space equation, however, is different from the one in (2.26) and (2.27). This is because many different state-space equations may have the same transfer functions, as we will discuss in Chapter 4.

Convolution

We now develop a convolution to describe the circuit in Fig. 2.6(a). To do so, we must first find its impulse response. Although it is possible to obtain an impulse response by measurement in theory; it is not possible to do so in practice. Therefore we compute it analytically. It is the inverse Laplace transform of the transfer function in (2.37). Using the MATLAB function `residue`, we can expand $\hat{g}(s)$ in (2.37) as

$$\hat{g}(s) = \frac{0.04}{s + 0.58} + \frac{-0.02 - 0.18j}{s + 0.08 - 0.35j} + \frac{-0.02 + 0.18j}{s + 0.08 + 0.35j}$$

Thus the impulse response of the circuit is, using $\mathcal{L}^{-1}[1/(s + a)] = e^{-at}$, for any real or complex a ,

$$g(t) = 0.04e^{-0.58t} + (-0.02 - 0.18j)e^{(-0.08+0.35j)t} + (-0.02 + 0.18j)e^{(-0.08-0.35j)t}$$

for $t \geq 0$. Thus the convolution description of the circuit is

$$\begin{aligned} y(t) &= \int_{\tau=0}^t \left[0.04e^{-0.58(t-\tau)} + (-0.02 - 0.18j)e^{(-0.08+0.35j)(t-\tau)} \right. \\ &\quad \left. + (-0.02 + 0.18j)e^{(-0.08-0.35j)(t-\tau)} \right] u(\tau) d\tau \end{aligned}$$

It is very complicated. Even though the expression can be simplified by combining the two complex-valued function into one real-valued function, it is still much more complicated than the other three descriptions. Moreover, its computer computation

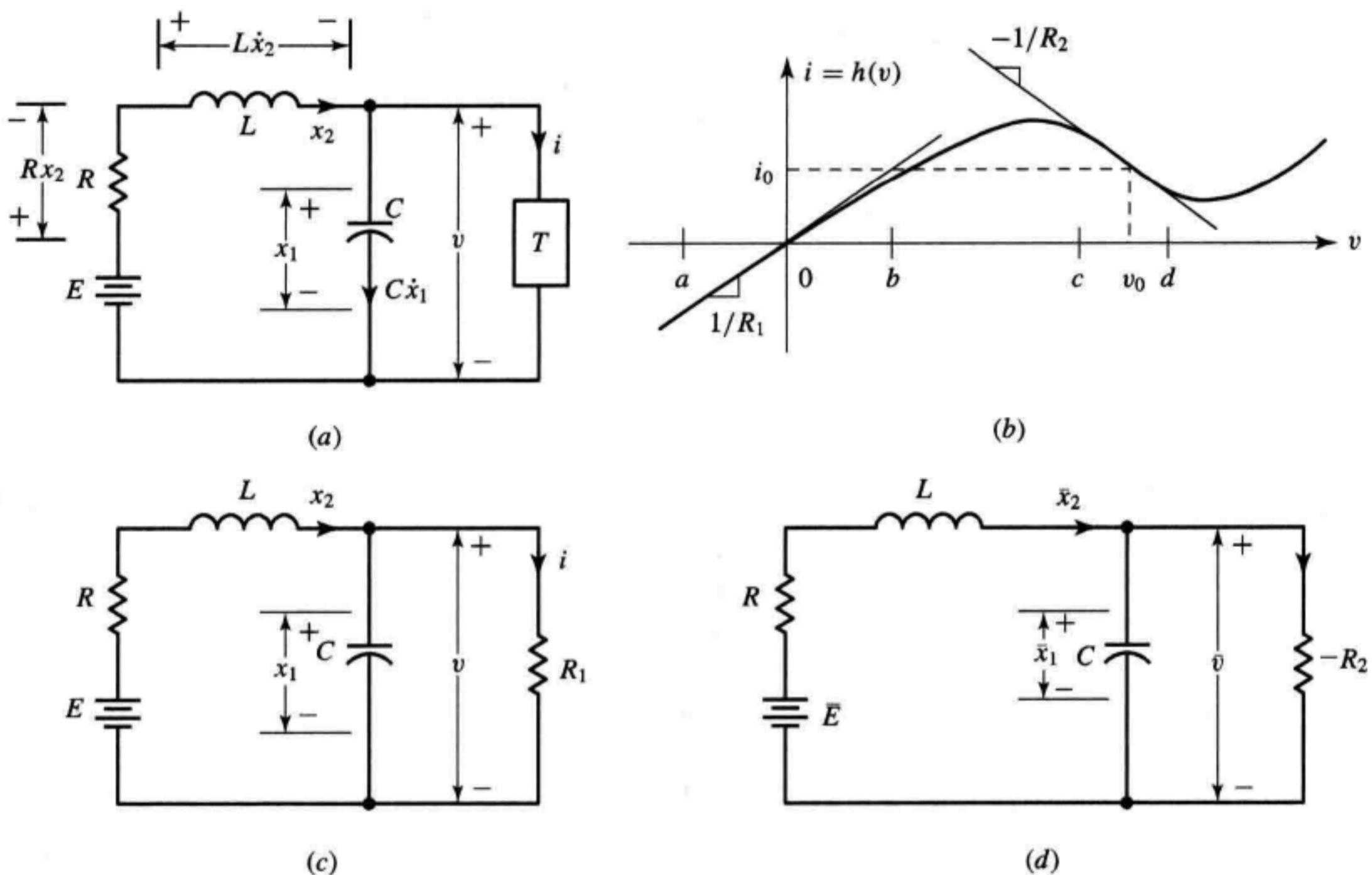


FIGURE 2.9 Circuit with a tunnel diode.

requires a very large number of operations. See Reference 10. Thus we study in the remainder of this text mostly transfer functions and state-space equations.

The next example illustrates the process of linearization.

Example 2.5.1 Consider the circuit shown in Fig. 2.9(a), where T is a tunnel diode with the characteristic shown in Fig. 2.9(b). Let x_1 be the voltage across the capacitor and x_2 be the current through the inductor. Then we have $v = x_1$ and

$$\begin{aligned} x_2(t) &= C\dot{x}_1(t) + i(t) = C\dot{x}_1(t) + h(x_1(t)) \\ L\dot{x}_2(t) &= E - Rx_2(t) - x_1(t) \end{aligned}$$

They can be arranged as

$$\begin{aligned} \dot{x}_1(t) &= \frac{-h(x_1(t))}{C} + \frac{x_2(t)}{C} \\ \dot{x}_2(t) &= \frac{-x_1(t) - Rx_2(t)}{L} + \frac{E}{L} \end{aligned} \quad (2.40)$$

This set of nonlinear equations describes the circuit. Now if $x_1(t)$ is known to lie only inside the range (a, b) shown in Fig. 2.9(b), then $h(x_1(t))$ can be approximated by $h(x_1(t)) = x_1(t)/R_1$. In this case, the circuit can be reduced to the one in Fig. 2.9(c) and

can be described by

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} -1/CR_1 & 1/C \\ -1/L & -R/L \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1/L \end{bmatrix} E$$

This is an LTI state equation. Now if $x_1(t)$ is known to lie only inside the range (c, d) shown in Fig. 2.9(b), we may introduce the variables $\bar{x}_1(t) = x_1(t) - v_o$, $\bar{x}_2(t) = x_2(t) - i_o$ and approximate $h(x_1(t))$ as $i_o - \bar{x}_1(t)/R_2$. Substituting these into (2.40) yields

$$\begin{bmatrix} \dot{\bar{x}}_1(t) \\ \dot{\bar{x}}_2(t) \end{bmatrix} = \begin{bmatrix} 1/CR_2 & 1/C \\ -1/L & -R/L \end{bmatrix} \begin{bmatrix} \bar{x}_1(t) \\ \bar{x}_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1/L \end{bmatrix} \bar{E}$$

where $\bar{E} = E - v_o - Ri_o$. This equation is obtained by shifting the operating point from $(0,0)$ to (v_o, i_o) and by linearization at (v_o, i_o) . Because the two linearized equations are identical if $-R_2$ is replaced by R_1 and \bar{E} by E , we can readily obtain its equivalent circuit shown in Fig. 2.9(d). Note that it is not obvious how to obtain the equivalent circuit from the original circuit without developing first the state equation.

2.6 Mechanical and Hydraulic Systems

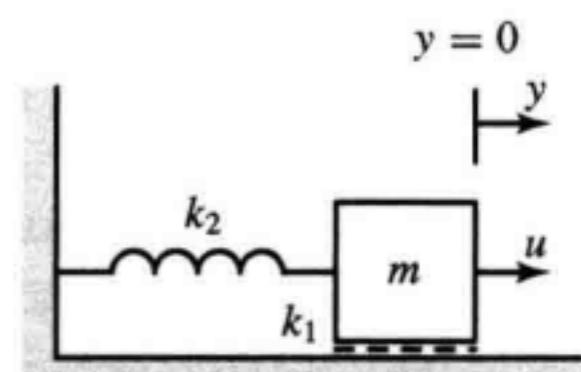
In this section we give more examples to develop transfer functions and state-space equations for physical systems.

Example 2.6.1 Consider the mechanical system shown in Fig. 2.10. It consists of a block with mass m connected to a wall through a spring. We consider the applied force u to be the input and displacement y from the equilibrium to be the output. The friction between the floor and the block generally consists of three distinct parts: static friction, Coulomb friction, and viscous friction, as shown in Fig. 2.11. Note that the horizontal coordinate is velocity $\dot{y} = dy/dt$. The friction is clearly not a linear function of the velocity. To simplify the analysis, we disregard the static and Coulomb frictions and consider only the viscous friction. Then the friction becomes linear and can be expressed as $k_1\dot{y}(t)$, where k_1 is the *viscous friction coefficient*. The characteristic of the spring is shown in Fig. 2.12; it is not linear. However, if the displacement is limited to (y_1, y_2) as shown, then the spring can be considered to be linear and the spring force equals k_2y , where k_2 is the *spring constant*. Thus the mechanical system can be modeled as an LTI system under linearization and simplification.

We apply Newton's law to develop an equation to describe the system. The applied force u must overcome the friction and the spring force. The remainder is used to accelerate the block. Thus we have

$$m\ddot{y}(t) = u(t) - k_1\dot{y}(t) - k_2y(t) \quad (2.41)$$

FIGURE 2.10 Mechanical system.



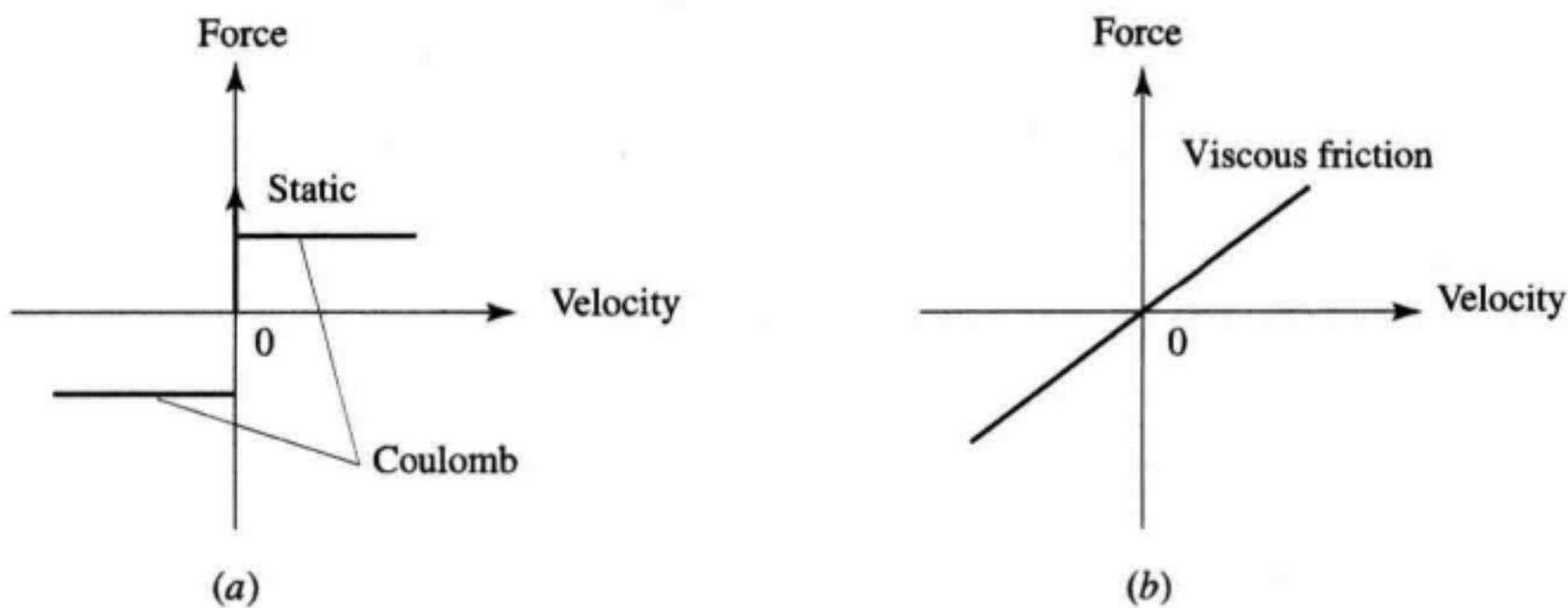


FIGURE 2.11 (a) Static and Coulomb frictions. (b) Viscous friction.

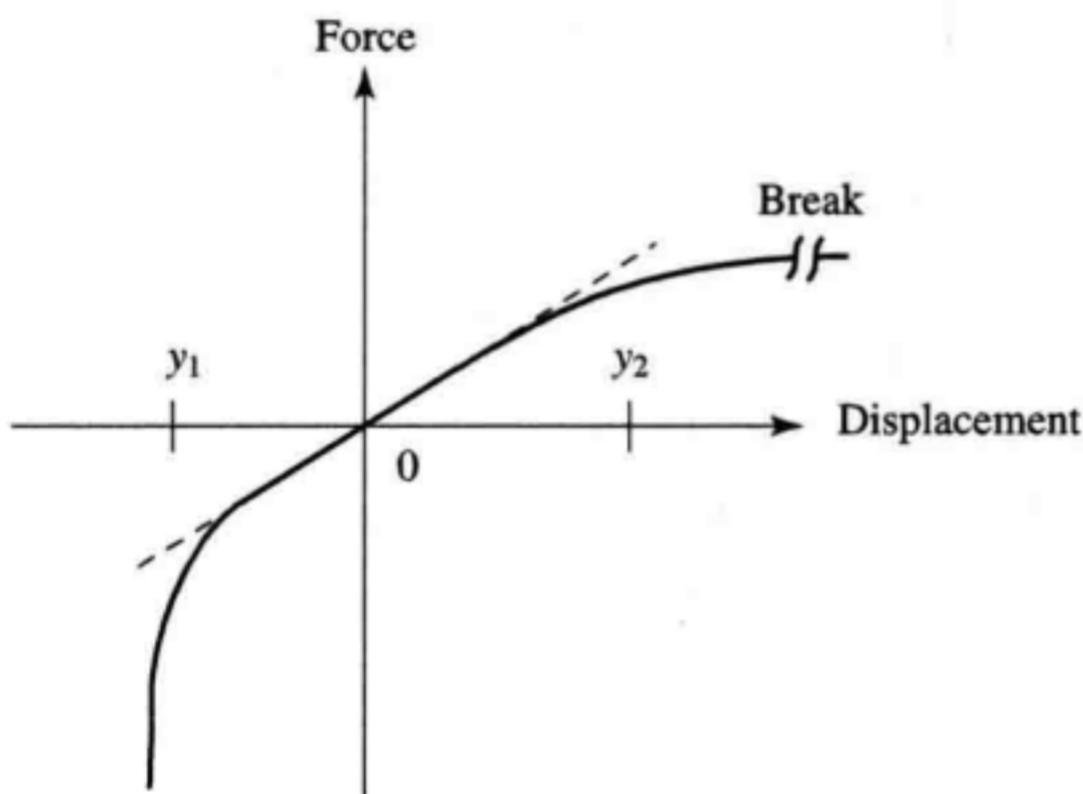


FIGURE 2.12 Characteristic of spring.

where $\ddot{y} = d^2y(t)/dt^2$ and $\dot{y} = dy(t)/dt$. Applying the Laplace transform and assuming zero initial conditions yield

$$ms^2\hat{y}(s) = \hat{u}(s) - k_1s\hat{y}(s) - k_2\hat{y}(s)$$

which implies

$$\hat{y}(s) = \frac{1}{ms^2 + k_1s + k_2}\hat{u}(s)$$

Thus the transfer function of the system is $1/(ms^2 + k_1s + k_2)$.

Next we develop a state-space equation to describe the system. Let us select the displacement and velocity of the block as state variables; that is, $x_1(t) = y(t)$, $x_2(t) = \dot{y}(t)$. Then we have, using (2.41),

$$\dot{x}_1(t) = x_2(t), \quad m\dot{x}_2(t) = u(t) - k_1x_2(t) - k_2x_1(t)$$

They can be expressed in matrix form as

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -k_2/m & -k_1/m \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1/m \end{bmatrix} u(t)$$

$$y(t) = [1 \ 0] \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

which describes the system.

Example 2.6.2 Consider the system shown in Fig. 2.13. It consists of two blocks, with mass m_1 and m_2 , connected by three springs with spring constants k_i , $i = 1, 2, 3$. To simplify the discussion, we assume that there is no friction between the blocks and the floor. The applied force u_1 must overcome the spring forces, and the remainder is used to accelerate the block; thus we have

$$u_1(t) - k_1 y_1(t) - k_2(y_1(t) - y_2(t)) = m_1 \ddot{y}_1(t)$$

or

$$m_1 \ddot{y}_1(t) + (k_1 + k_2)y_1(t) - k_2 y_2(t) = u_1(t) \quad (2.42)$$

For the second block, we have

$$m_2 \ddot{y}_2(t) - k_2 y_1(t) + (k_1 + k_2)y_2(t) = u_2(t) \quad (2.43)$$

They can be combined as

$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{bmatrix} \ddot{y}_1(t) \\ \ddot{y}_2(t) \end{bmatrix} + \begin{bmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_1 + k_2 \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$$

This is a standard equation in studying vibration and is said to be in normal form. See Reference 21. Let us define

$$x_1(t) := y_1(t), \quad x_2(t) := \dot{y}_1(t), \quad x_3(t) := y_2(t), \quad x_4(t) := \dot{y}_2(t)$$

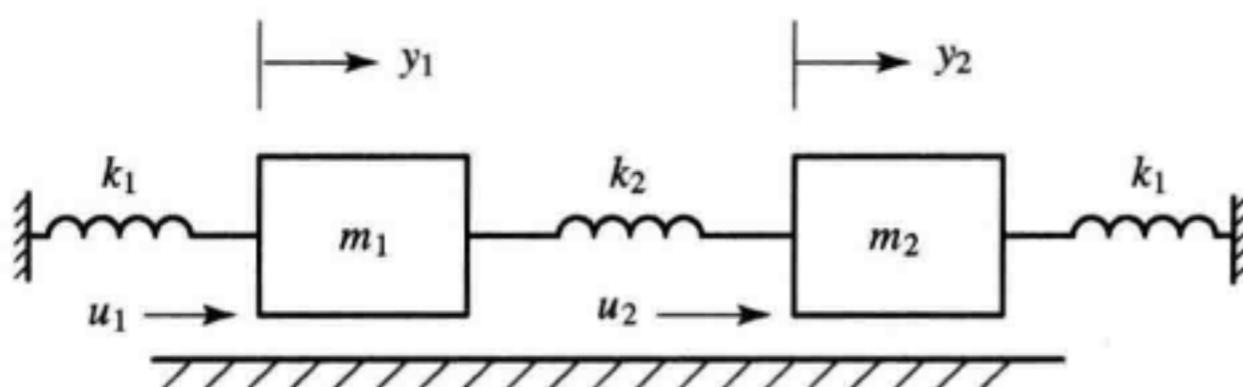


FIGURE 2.13 Spring–mass system

Then we can readily obtain

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \\ \dot{x}_4(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{-(k_1+k_2)}{m_1} & 0 & \frac{k_2}{m_1} & 0 \\ 0 & 0 & 0 & 1 \\ \frac{k_2}{m_2} & 0 & \frac{-(k_1+k_2)}{m_1} & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \frac{1}{m_1} & 0 \\ 0 & 0 \\ 0 & \frac{1}{m_2} \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$$

$$\mathbf{y}(t) := \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{x}(t)$$

This two-input two-output state-space equation describes the system in Fig. 2.13.

To develop its input–output description, we apply the Laplace transform to (2.42) and (2.43) and assume zero initial conditions to yield

$$m_1 s^2 \hat{y}_1(s) + (k_1 + k_2) \hat{y}_1(s) - k_2 \hat{y}_2(s) = \hat{u}_1(s)$$

$$m_2 s^2 \hat{y}_2(s) - k_2 \hat{y}_1(s) + (k_1 + k_2) \hat{y}_2(s) = \hat{u}_2(s)$$

From these two equations, we can obtain

$$\begin{bmatrix} \hat{y}_1(s) \\ \hat{y}_2(s) \end{bmatrix} = \begin{bmatrix} \frac{m_2 s^2 + k_1 + k_2}{d(s)} & \frac{k_2}{d(s)} \\ \frac{k_2}{d(s)} & \frac{m_1 s^2 + k_1 + k_2}{d(s)} \end{bmatrix} \begin{bmatrix} \hat{u}_1(s) \\ \hat{u}_2(s) \end{bmatrix}$$

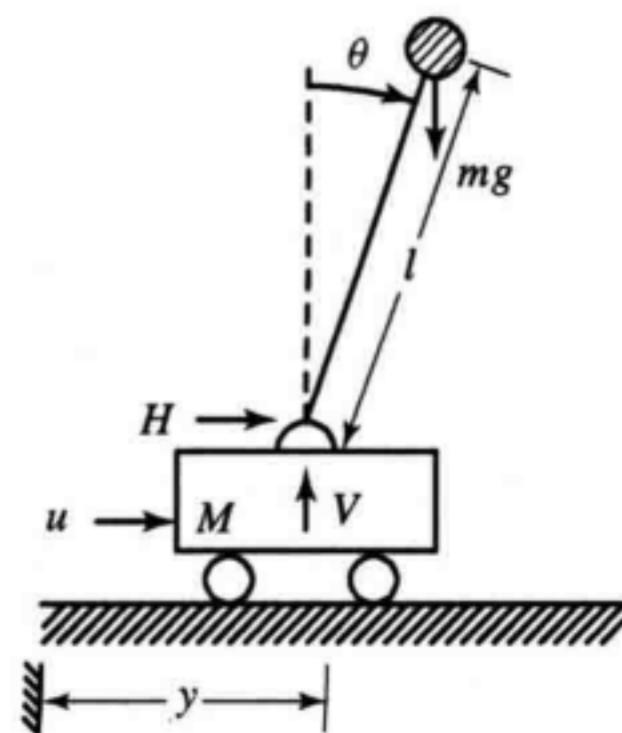
where

$$d(s) := (m_1 s^2 + k_1 + k_2)(m_2 s^2 + k_1 + k_2) - k_2^2$$

This is the transfer-matrix description of the system. Thus what we will discuss in this text can be directly applied to study vibration.

Example 2.6.3 Consider a cart with an inverted pendulum hinged on top of it as shown in Fig. 2.14. For simplicity, the cart and the pendulum are assumed to move in only one direction, and the friction, the mass of the stick, and the gust of wind are disregarded. The problem is to maintain the pendulum at the vertical position. For example, if the inverted pendulum is falling in the direction shown, the cart moves

FIGURE 2.14 Cart with inverted pendulum.



to the right and exerts a force, through the hinge, to push the pendulum back to the vertical position. This simple mechanism can be used as a model of a space vehicle on takeoff.

Let H and V be, respectively, the horizontal and vertical forces exerted by the cart to the pendulum as shown. The application of Newton's law to the linear movements yields

$$\begin{aligned} M \frac{d^2y}{dt^2} &= u - H \\ H &= m \frac{d^2}{dt^2}(y + l \sin \theta) = m\ddot{y} + ml\ddot{\theta} \cos \theta - ml(\dot{\theta})^2 \sin \theta \\ mg - V &= m \frac{d^2}{dt^2}(l \cos \theta) = ml[-\ddot{\theta} \sin \theta - (\dot{\theta})^2 \cos \theta] \end{aligned}$$

The application of Newton's law to the rotational movement of the pendulum around the hinge yields

$$mgl \sin \theta = ml\ddot{\theta} \cdot l + m\ddot{y}l \cos \theta$$

They are nonlinear equations. Because the design objective is to maintain the pendulum at the vertical position, it is reasonable to assume θ and $\dot{\theta}$ to be small. Under this assumption, we can use the approximation $\sin \theta = \theta$ and $\cos \theta = 1$. By retaining only the linear terms in θ and $\dot{\theta}$ or, equivalently, dropping the terms with θ^2 , $(\dot{\theta})^2$, $\theta\dot{\theta}$, and $\theta\ddot{\theta}$, we obtain $V = mg$ and

$$\begin{aligned} M\ddot{y} &= u - m\ddot{y} - ml\ddot{\theta} \\ g\theta &= l\ddot{\theta} + \ddot{y} \end{aligned}$$

which imply

$$M\ddot{y} = u - mg\theta \quad (2.44)$$

$$Ml\ddot{\theta} = (M+m)g\theta - u \quad (2.45)$$

Using these linearized equations, we now can develop the input-output and state-space descriptions. Applying the Laplace transform to (2.44) and (2.45) and assuming zero initial conditions yield

$$Ms^2\hat{y}(s) = \hat{u}(s) - mg\hat{\theta}(s)$$

$$Mls^2\hat{\theta}(s) = (M+m)g\hat{\theta}(s) - \hat{u}(s)$$

From these equations, we can readily compute the transfer function $\hat{g}_{yu}(s)$ from u to y and the transfer function $\hat{g}_{\theta u}(s)$ from u to θ as

$$\hat{g}_{yu}(s) = \frac{s^2 - g}{s^2[Ms^2 - (M+m)g]}$$

$$\hat{g}_{\theta u}(s) = \frac{-1}{Ms^2 - (M+m)g}$$

To develop a state-space equation, we select state variables as $x_1(t) = y(t)$, $x_2(t) = \dot{y}(t)$, $x_3(t) = \theta(t)$, and $x_4(t) = \dot{\theta}(t)$. Then from this selection, (2.44), and (2.45) we can readily obtain

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \\ \dot{x}_4(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -mg/M & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & (M+m)g/Ml & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1/M \\ 0 \\ -1/Ml \end{bmatrix} u(t)$$

$$y(t) = [1 \ 0 \ 0 \ 0] \mathbf{x}(t) \quad (2.46)$$

This state-space equation has dimension 4 and describes the system when θ and $\dot{\theta}$ are very small.

Example 2.6.4 A communication satellite of mass m orbiting around the earth is shown in Fig. 2.15. The altitude of the satellite is specified by $r(t)$, $\theta(t)$, and $\phi(t)$ as shown. The orbit can be controlled by three orthogonal thrusts; $u_r(t)$, $u_\theta(t)$, and $u_\phi(t)$. The

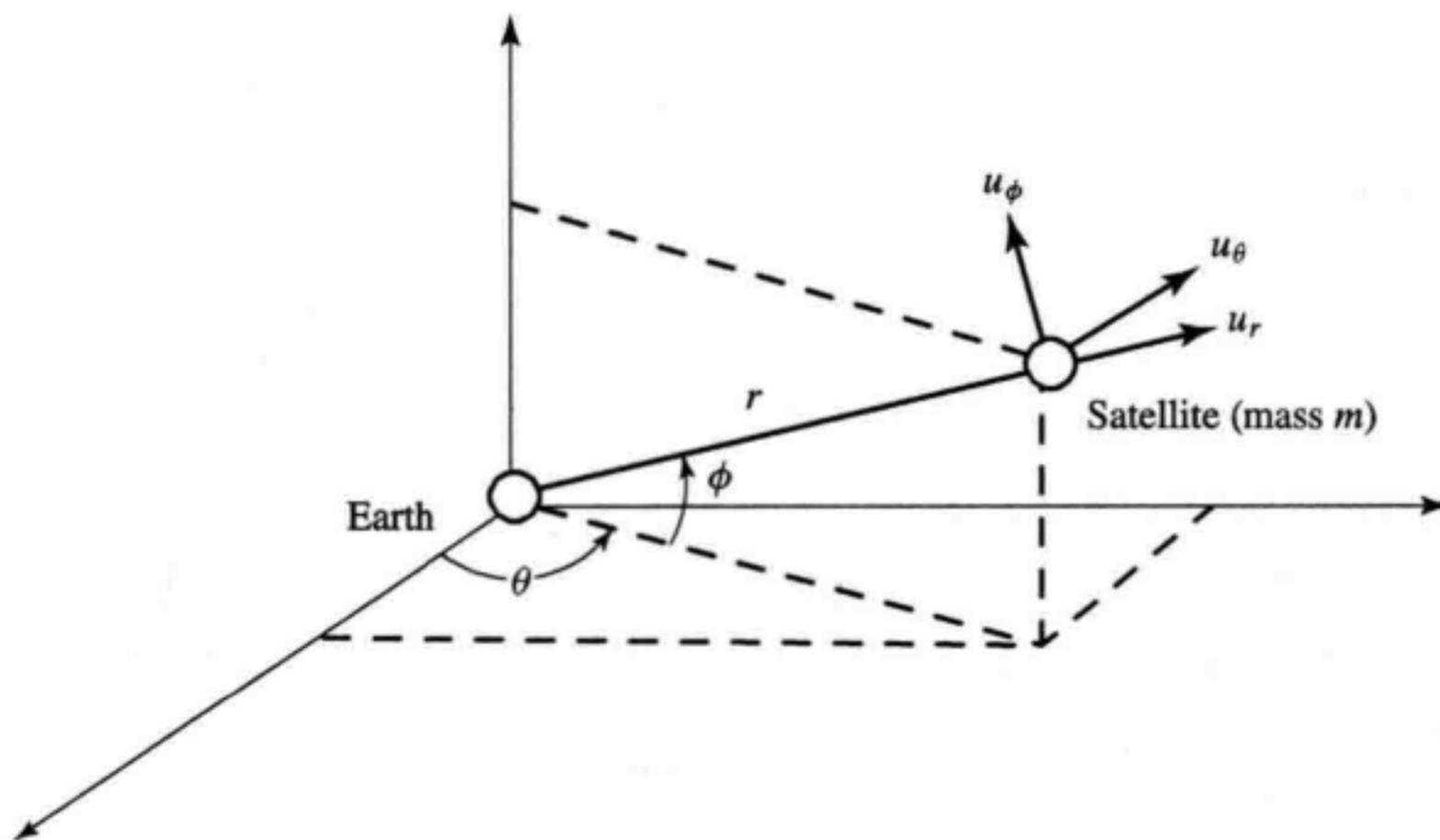


FIGURE 2.15 Satellite in orbit.

state, input, and output of the system are chosen as

$$\mathbf{x}(t) = \begin{bmatrix} r(t) \\ \dot{r}(t) \\ \theta(t) \\ \dot{\theta}(t) \\ \phi(t) \\ \dot{\phi}(t) \end{bmatrix}, \quad \mathbf{u}(t) = \begin{bmatrix} u_r(t) \\ u_\theta(t) \\ u_\phi(t) \end{bmatrix}, \quad \mathbf{y}(t) = \begin{bmatrix} r(t) \\ \theta(t) \\ \phi(t) \end{bmatrix}$$

Then the system can be shown to be described by

$$\dot{\mathbf{x}} = \mathbf{h}(\mathbf{x}, \mathbf{u}) = \begin{bmatrix} \dot{r} \\ r\dot{\theta}^2 \cos^2 \phi + r\dot{\phi}^2 - k/r^2 + u_r/m \\ \dot{\theta} \\ -2\dot{r}\dot{\theta}/r + 2\dot{\theta}\dot{\phi} \sin \phi / \cos \phi + u_\theta / mr \cos \phi \\ \dot{\phi} \\ -\dot{\theta}^2 \cos \phi \sin \phi - 2\dot{r}\dot{\phi}/r + u_\phi / mr \end{bmatrix} \quad (2.47)$$

One solution which corresponds to a circular equatorial orbit is given by

$$\mathbf{x}_o(t) = [r_o \ 0 \ \omega_o t \ \omega_o \ 0 \ 0]', \quad b\mathbf{u}_o \equiv \mathbf{0}$$

with $r_o^3 \omega_o^2 = k$, a known physical constant. Once the satellite reaches the orbit, it will remain in the orbit as long as there are no disturbances. If the satellite deviates from the orbit, thrusts must be applied to push it back to the orbit. Define

$$\mathbf{x}(t) = \mathbf{x}_o(t) + \bar{\mathbf{x}}(t) \quad b\mathbf{u}(t) = \mathbf{u}_o(t) + \bar{\mathbf{u}}(t) \quad b\mathbf{y}(t) = \mathbf{y}_o + \bar{\mathbf{y}}(t)$$

If the perturbation is very small, then (2.47) can be linearized as

$$\dot{\bar{\mathbf{x}}}(t) = \begin{bmatrix} 0 & 1 & 0 & 0 & \vdots & 0 & 0 \\ 3\omega_o^2 & 0 & 0 & 2\omega_o r_o & \vdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \vdots & 0 & 0 \\ 0 & \frac{-2\omega_o}{r_o} & 0 & 0 & \vdots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \vdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \vdots & -\omega_o^2 & 0 \end{bmatrix} \bar{\mathbf{x}}(t)$$

$$\begin{aligned}
 & + \begin{bmatrix} 0 & 0 & \vdots & 0 \\ \frac{1}{m} & 0 & \vdots & 0 \\ 0 & 0 & \vdots & 0 \\ 0 & \frac{1}{mr_o} & \vdots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \vdots & 0 \\ 0 & 0 & \vdots & \frac{1}{mr_o} \end{bmatrix} \bar{\mathbf{u}}(t) \\
 \bar{\mathbf{y}}(t) = & \begin{bmatrix} 1 & 0 & 0 & 0 & \vdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \vdots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \vdots & 1 & 0 \end{bmatrix} \bar{\mathbf{x}}(t) \quad (2.48)
 \end{aligned}$$

This six-dimensional state-space equation describes the system. In this equation, **A**, **B**, and **C** happen to be constant. If the orbit is an elliptic one, then they will be time varying. We note that the three matrices are all block diagonal. Thus the equation can be decomposed into two uncoupled parts: one involving r and θ , the other involving ϕ . Studying these two parts separately can simplify analysis and design.

Example 2.6.5 In chemical plants, it is often necessary to maintain the levels of liquids. A simplified model of a connection of two tanks is shown in Fig. 2.16. It is assumed that under normal operation, the inflows and outflows of both tanks all equal Q and their liquid levels equal H_1 and H_2 . Let u be inflow perturbation of the first tank, which will cause variations in liquid level x_1 and outflow y_1 as shown. These variations will cause level variation x_2 and outflow variation y in the second tank. It is assumed that

$$y_1 = \frac{x_1 - x_2}{R_1} \quad \text{and} \quad y = \frac{x_2}{R_2}$$

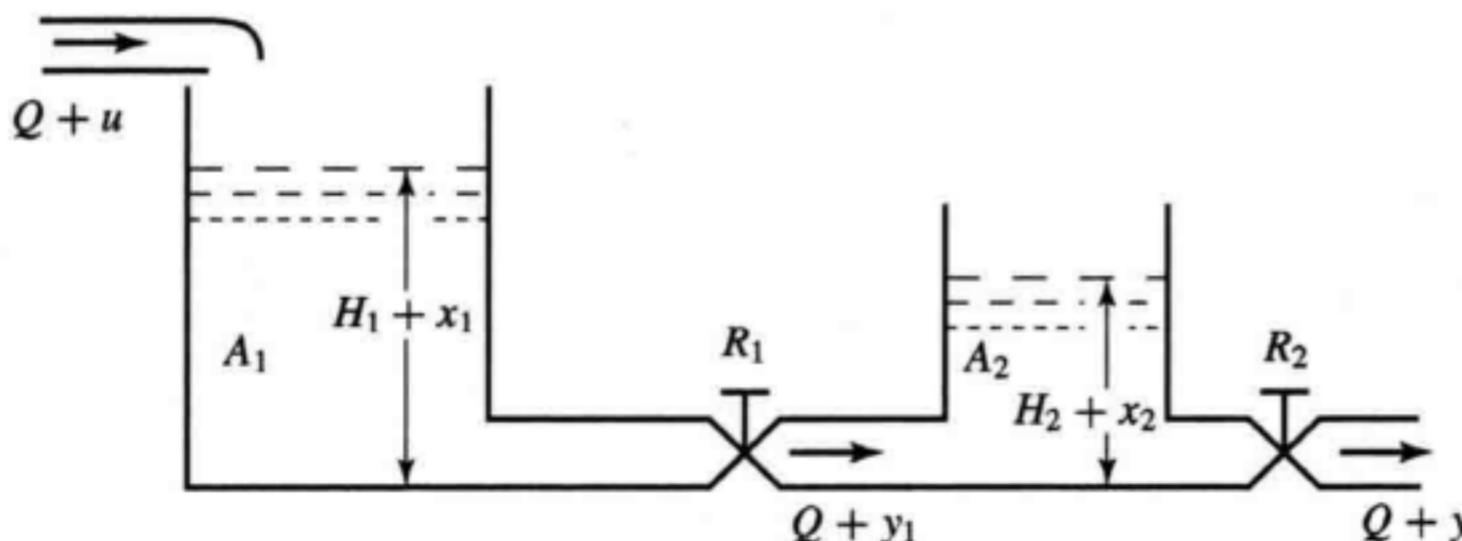


FIGURE 2.16 Hydraulic tanks.

where R_i are the flow resistances and depend on the normal heights H_1 and H_2 . They can also be controlled by the valves. Changes of liquid levels are governed by

$$A_1 dx_1 = (u - y_1) dt \quad \text{and} \quad A_2 dx_2 = (y_1 - y) dt$$

where A_i are the cross sections of the tanks. From these equations, we can readily obtain

$$\begin{aligned}\dot{x}_1 &= \frac{u}{A_1} - \frac{x_1 - x_2}{A_1 R_1} \\ \dot{x}_2 &= \frac{x_1 - x_2}{A_2 R_1} - \frac{x_2}{A_2 R_2}\end{aligned}$$

Thus the state-space description of the system is given by

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} -1/A_1 R_1 & 1/A_1 R_1 \\ 1/A_2 R_1 & -(1/A_2 R_1 + 1/A_2 R_2) \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 1/A_1 \\ 0 \end{bmatrix} u(t)$$

$$y(t) = [0 \quad 1/R_2] \mathbf{x}(t)$$

Its transfer function can be computed as

$$\hat{g}(s) = \frac{1}{A_1 A_2 R_1 R_2 s^2 + (A_1 R_1 + A_1 R_2 + A_2 R_2)s + 1}$$

2.7 Proper Rational Transfer Functions

The transfer functions encountered in Examples 2.3.3 and 2.3.4 are irrational functions of s . The rest, however, are all rational functions of s . In fact, we can also use transfer functions to define systems to be lumped or distributed. A system is lumped if its transfer function is a rational function of s and is distributed if its transfer function is not. In other words, the transfer function of a distributed system is an irrational function of s or cannot be expressed in a closed form. We study mostly lumped systems, thus the transfer functions we will encounter are mostly rational functions of s .

Every rational transfer function can be expressed as $\hat{g}(s) = N(s)/D(s)$ where $N(s)$ and $D(s)$ are polynomials of s . Let us use \deg to denote the degree of a polynomial. Then $\hat{g}(s)$ can be classified as

- $\hat{g}(s)$ proper $\iff \deg D(s) \geq \deg N(s) \iff \hat{g}(\infty) = \text{zero or nonzero constant}$
- $\hat{g}(s)$ strictly proper $\iff \deg D(s) > \deg N(s) \iff \hat{g}(\infty) = 0$
- $\hat{g}(s)$ biproper $\iff \deg D(s) = \deg N(s) \iff \hat{g}(\infty) = \text{nonzero constant}$
- $\hat{g}(s)$ improper $\iff \deg D(s) < \deg N(s) \iff |\hat{g}(\infty)| = \infty$

For example, the rational functions

$$\frac{s^3 - 2s - 5}{s^2 + 1}, \quad s^2 + 2s + 1 = \frac{s^2 + 2s + 1}{1}$$

are improper and the rational functions

$$\frac{s^2 + 1}{s^3 - 2s - 5}, \quad \frac{s + 1}{s^{10}}, \quad \frac{s^2 + 1}{2s^2 + s - 5}, \quad 10 = \frac{10}{1}$$

are proper. The first two are also strictly proper and the last two are also biproper. Thus proper includes both strictly proper and biproper. Note that if $\hat{g}(s)$ is biproper, so is $1/\hat{g}(s)$. Improper rational transfer functions will amplify high-frequency noise which often exists in the real world, especially in electrical systems; therefore, improper rational transfer functions rarely arise in electrical systems. In this text, we study only proper rational transfer functions.

The roots of a polynomial $D(s)$ are defined as all solutions of the polynomial equation $D(s) = 0$. If $D(s)$ has degree m , then $D(s)$ has m real or complex roots. If $D(s)$ has only real coefficients, then its complex-conjugate roots must appear in pairs. The roots of a polynomial can be computed in MATLAB using the function `roots`. A polynomial is represented in MATLAB as a row vector with its descending power coefficients as entries, separated by commas or spaces, and bounded by a pair of brackets. For example, for the polynomial

$$D(s) = s^5 + 6s^4 + 29.25s^3 + 93.25s^2 + 134s + 65 \quad (2.49)$$

typing in a MATLAB command window the following

```
d=[1 6 29.25 93.25 134 65]; roots(d)
```

will yield its five roots as

```
-0.5+4i -0.5-4i -2+0i -2-0i -1
```

It has a pair of complex-conjugate roots at $-0.5 \pm 4i$ and three real roots $-2, -2, -1$. The complex roots and -1 are called *simple roots*, and -2 is called a *repeated root* or a root with multiplicity 2. Note that in MATLAB, both i and j denote $\sqrt{-1}$.

The coefficient of the highest power of a polynomial is called the *leading coefficient*. A polynomial is called *monic* if its leading coefficient is 1. For example, $D(s)$ in (2.49) is monic but

$$D_1(s) = -3s^3 - 12s^2 - 39s + 150$$

is not. Its monic polynomial is $D_2(s) := D_1(s)/(-3) = s^3 + 4s^2 + 13s - 50$. Typing in MATLAB `roots([-3 -12 -39 150])` and `roots([1 4 13 -50])` will yield the same three roots 2 and $-3 \pm j4$. By definition, the roots of $D_1(s)$ are the solutions of $D_1(s) = 0$. Thus $kD_1(s)$, for any nonzero number k , has the same set of roots as $D_1(s)$. Consequently, in factoring $D_1(s)$, we must include its leading coefficient as

$$D_1(s) = -3s^3 - 12s^2 - 39s + 150 = -3(s - 2)(s + 3 - j4)(s + 3 + j4)$$

A real or complex number λ is called a *pole* of the proper transfer function $\hat{g}(s) = N(s)/D(s)$ if $\hat{g}(\lambda) = \infty$ or $-\infty$; a *zero* if $\hat{g}(\lambda) = 0$. Two polynomials $N(s)$ and $D(s)$ are defined to be *coprime* if they have no common root. If $N(s)$ and $D(s)$ are not coprime or have the common root a , then $\hat{g}(a) = N(a)/D(a) = 0/0$, which is not defined. If $N(s)$ and $D(s)$ are coprime and if $D(a) = 0$, then $N(a) \neq 0$. In this case, $\hat{g}(a) = N(a)/D(a) = N(a)/0$ which is ∞ or $-\infty$ and a is a pole of $\hat{g}(s)$. Likewise if $N(b) = 0$, then $D(b) \neq 0$ and $\hat{g}(b) = 0/D(b) = 0$ and b is a zero. In conclusion, if $N(s)$

and $D(s)$ are coprime, then every root of $N(s)$ is a zero of $\hat{g}(s)$, and every root of $D(s)$ is a pole of $\hat{g}(s)$. In terms of poles and zeros, a transfer function can be expressed as

$$\hat{g}(s) = k \frac{(s - z_1)(s - z_2) \cdots (s - z_m)}{(s - p_1)(s - p_2) \cdots (s - p_n)}$$

This is called the *zero-pole-gain* form. In MATLAB, this form can be obtained from the transfer function by calling $[z, p, k] = \text{tf2zp}(\text{num}, \text{den})$. For example, consider

$$\hat{g}(s) = \frac{4s^4 - 16s^3 + 8s^2 - 48s + 180}{s^5 + 6s^4 + 29.25s^3 + 93.25s^2 + 134s + 65}$$

Typing in the command window of MATLAB

```
n=[4 -16 8 -48 180]; d=[1 6 29.25 93.25 134 65];
[z, p, k]=tf2zp(n, d)
```

will yield $z = [-1+2i -1-2i 3 3]$; $p = [-0.5+4i -0.5-4i -2+0i -2-0i -1]$; $k=4$. It means that the transfer function can be expressed as

$$H(s) = \frac{4(s + 1 - 2j)(s + 1 + 2j)(s - 3)^2}{(s + 0.5 - 4j)(s + 0.5 + 4j)(s + 2)^2(s + 1)}$$

It has zeros and poles as shown and gain 4. Note that zeros and poles do not specify uniquely a transfer function. We must also specify the gain k .

A rational matrix $\hat{\mathbf{G}}(s)$ is said to be proper if its every entry is proper or if $\hat{\mathbf{G}}(\infty)$ is a zero or nonzero constant matrix. It is strictly proper if its every entry is strictly proper or if $\hat{\mathbf{G}}(\infty)$ is a zero matrix. If a rational matrix $\hat{\mathbf{G}}(s)$ is square and if both $\hat{\mathbf{G}}(s)$ and $\hat{\mathbf{G}}^{-1}(s)$ are proper, then $\hat{\mathbf{G}}(s)$ is said to be biproper. We call λ a pole of $\hat{\mathbf{G}}(s)$ if it is a pole of some entry of $\hat{\mathbf{G}}(s)$. Thus every pole of every entry of $\hat{\mathbf{G}}(s)$ is a pole of $\hat{\mathbf{G}}(s)$. There are a number of ways of defining zeros for $\hat{\mathbf{G}}(s)$. We call λ a *blocking zero* if it is a zero of every nonzero entry of $\hat{\mathbf{G}}(s)$. A more useful definition is the *transmission zero*, which will be introduced in Chapter 9.

2.8 Discrete-Time Linear Time-Invariant Systems

This section develops the discrete-time counterpart of continuous-time systems. Because most concepts in continuous-time systems can be directly applied to the discrete-time case, the discussion will be brief.

The input and output of every discrete-time (DT) system will be assumed to have the same sampling period T and will be denoted by $u[k] := u(kT)$, $y[k] := y(kT)$, where k , called the time index, is an integer ranging from $-\infty$ to $+\infty$. A discrete-time system is causal if current output depends on current and past inputs. The state at time k_0 , denoted by $\mathbf{x}[k_0]$, is the information at time instant k_0 , which together with $u[k]$, $k \geq k_0$, determines uniquely the output $y[k]$, $k \geq k_0$. The entries of \mathbf{x} are called state variables. If the number of state variables is finite, the discrete-time system is lumped; otherwise, it is distributed. Every continuous-time system involving time delay, as the ones in Examples 2.3.1 and 2.3.2, is a distributed system. In a discrete-time system, if

the time delay is an integer multiple of the sampling period T , then the discrete-time system is lumped.

A DT system is time-invariant if its characteristics do not change with time. For such a system, no matter at what time instant we apply an input, the output sequence will always be the same. Thus we may assume $k_0 = 0$ as the initial time instant. The instant $k_0 = 0$ is a relative one and is selected by us. For DT time-invariant systems, the time index k is generally limited to $k \geq 0$.

A DT time-invariant system is linear if the additivity and homogeneity properties hold. The response of every DT LTI system can be decomposed as

$$\text{Response} = \text{zero-state response} + \text{zero-input response}$$

and the zero-state responses satisfy the superposition property. So do the zero-input responses. Before proceeding, we introduce the concept of impulse sequences.

Let $\delta_d[k]$ be the *impulse sequence* defined as

$$\delta_d[k - m] = \begin{cases} 1 & \text{if } k = m \\ 0 & \text{if } k \neq m \end{cases}$$

where k is the time index and m is a fixed integer.⁶ Thus we have $\delta_d[0] = 0$, $\delta_d[4] = 0$, and $\delta_d[56] = 0$. The impulse sequence is the discrete counterpart of the impulse $\delta(t - t_1)$. The impulse $\delta(t - t_1)$ has zero width and infinite height and cannot be generated in practice, whereas the impulse sequence $\delta_d[k - m]$ can be easily generated.

Consider a sequence $u[k]$. Its value at $k = m$ equals $u[m]$ or $u[m]\delta_d[k - m]$ for all k . Thus $u[k]$ can be expressed as the sum of impulse sequences as

$$u[k] = \sum_{m=-\infty}^{\infty} u[m]\delta_d[k - m]$$

For example, if $k = 10$, the infinite summation has all its terms equal to zero except $m = 10$ and therefore reduces to $u[10]$. The equation is the DT counterpart of (2.2).

Convolution

Consider a DT LTI system with input sequence $u[k]$ and output sequence $y[k]$. We develop an equation to describe its zero-state responses. That is, the output excited by an input sequence applied from $k_0 = 0$ onward with the system initially relaxed at $k_0 = 0$. If $u[k] = 0$ for all $k < 0$, then the system is initially relaxed.

Consider an input $u[k]$ applied from $k = 0$ onward. Then it can be express as

$$u[k] = \sum_{m=0}^{\infty} u[m]\delta_d[k - m]$$

⁶We may also consider m to be the time index and k is fixed.

as discussed earlier. Let $g[k]$ be the output at time instant k excited by the impulse sequence applied at time instant $k_0 = 0$. Then we have

$$\begin{aligned}\delta_d[k] &\rightarrow g[k] \quad (\text{definition}) \\ \delta_d[k-m] &\rightarrow g[k-m] \quad (\text{time shifting}) \\ \delta_d[k-m]u[m] &\rightarrow g[k-m]u[m] \quad (\text{homogeneity}) \\ \sum_m \delta_d[k-m]u[m] &\rightarrow \sum_m g[k-m]u[m] \quad (\text{additivity})\end{aligned}$$

Thus the output $y[k]$ excited by the input $u[k]$, for $k \geq 0$, equals

$$y[k] = \sum_{m=0}^{\infty} g[k-m]u[m] \quad (2.50)$$

for $k \geq 0$. If the system is causal as well, then

$$g[k] = 0 \quad \text{for all } k < 0$$

which implies $g[k-m] = 0$, for all $m > k$. Thus the upper limit of the summation in (2.50) can be changed from ∞ to k , and (2.50) reduces to

$$y[k] = \sum_{m=0}^k g[k-m]u[m] = \sum_{m=0}^k g[m]u[k-m] \quad (2.51)$$

This is called the *discrete convolution*. The second equation is obtained from the first by defining $\bar{m} := k - m$ and then renaming \bar{m} as m . It is the discrete counterpart of (2.5). Its derivation, however, is much simpler because it does not involve any approximation and a limiting process. The sequence $g[k]$ is called the *impulse response sequence*.

Transfer function

The z -transform is an important tool in the study of DT LTI systems. Let $\hat{y}(z)$ be the z -transform of $y[k]$ defined as

$$\hat{y}(z) := \mathcal{Z}[y[k]] := \sum_{k=0}^{\infty} y[k]z^{-k} \quad (2.52)$$

Substituting (2.50) and interchanging the order of summation yield

$$\begin{aligned}\hat{y}(z) &= \sum_{k=0}^{\infty} \left(\sum_{m=0}^{\infty} g[k-m]u[m] \right) z^{-(k-m)} z^{-m} \\ &= \sum_{m=0}^{\infty} \left(\sum_{k=0}^{\infty} g[k-m]z^{-(k-m)} \right) u[m] z^{-m}\end{aligned}$$

which becomes, after introducing the new variable $l = k - m$ for the inner summation where m is fixed,

$$\hat{g}(z) = \sum_{m=0}^{\infty} \left(\sum_{l=-m}^{\infty} g[l] z^{-l} \right) u[m] z^{-m}$$

Using the causality condition $g[l] = 0$ for $l < 0$ to replace the lower summation limit from $-m$ to 0, the inner summation becomes independent of m and the double summation becomes

$$\hat{y}(z) = \left(\sum_{l=0}^{\infty} g[l] z^{-l} \right) \left(\sum_{m=0}^{\infty} u[m] z^{-m} \right)$$

or

$$\hat{y}(z) = \hat{g}(z) \hat{u}(z) \quad (2.53)$$

where

$$\hat{g}(z) := \mathcal{Z}[g[k]] = \sum_{k=0}^{\infty} g[k] z^{-k}$$

is called the *discrete transfer function*. It is the z -transform of the impulse response sequence. Using (2.53), we can also define the discrete transfer function as

$$\hat{g}(z) = \frac{\hat{y}(z)}{\hat{u}(z)} = \frac{\mathcal{Z}[\text{output}]}{\mathcal{Z}[\text{input}]} \Big|_{\text{initially relaxed}}$$

This is the DT counterpart of (2.8). We stress once again that the discrete convolution and transfer function describe only zero-state responses. They are applicable to LTI lumped and distributed systems.

Example 2.8.1 Consider the unit-sampling-period delay system defined by

$$y[k] = u[k - 1]$$

The output equals the input delayed by one sampling period. Its impulse response sequence is $g[k] = \delta_d[k - 1]$ and its discrete transfer function is

$$\hat{g}(z) = \mathcal{Z}[\delta_d[k - 1]] = z^{-1} = \frac{1}{z}$$

It is a rational function of z . Note that every continuous-time system involving time delay is a distributed system. This is not so in discrete-time systems.

Using the transfer function, the input and output are related by $\hat{y}(z) = \hat{g}(z) \hat{u}(z) = \hat{y}(z) = z^{-1} \hat{u}(z)$. Thus the delay of one sampling period is equivalent to the multiplication

of z^{-1} in the z -transform domain. Thus z^{-1} is often called the *unit-sampling-period delay operator*. Thus if $x[k] = 0$, for all $k < 0$, and if $\hat{x}(z) = \mathcal{Z}[x[n]]$, then we have

$$\mathcal{Z}[x[k - m]] = z^{-m}\hat{x}(z)$$

for any positive integer m .

Example 2.8.2 Consider the discrete-time feedback system shown in Fig. 2.17(a). It is the discrete counterpart of Fig. 2.5(a). Its impulse response sequence is, as in (2.6),

$$g_f[k] = a\delta_d[k - 1] + a^2\delta_d[k - 2] + \dots = \sum_{m=1}^{\infty} a^m\delta_d[k - m]$$

The z -transform of $\delta_d[k - m]$ is z^{-m} . Thus the transfer function of the feedback system is

$$\begin{aligned}\hat{g}_f(z) &= \mathcal{Z}[g_f[k]] = az^{-1} + a^2z^{-2} + a^3z^{-3} + \dots \\ &= az^{-1} \sum_{m=0}^{\infty} (az^{-1})^m = \frac{az^{-1}}{1 - az^{-1}}\end{aligned}$$

It is a rational function of z , in contrast to the CT irrational transfer function in Example 2.3.4.

If the unit-sampling-time delay element is replaced by its transfer function z^{-1} , then the block diagram becomes the one in Fig. 2.17(b). Using the algebraic method as in Example 2.3.4, we can compute the transfer function from r to y as

$$\hat{g}(z) = \frac{\hat{y}(z)}{\hat{r}(z)} = \frac{az^{-1}}{1 - az^{-1}} = \frac{a}{z - a}$$

This is the same as the one computed earlier.

The discrete transfer functions in the preceding two examples are rational functions of z . This may not be so in general. For example, consider a system with the impulse response sequence

$$g[k] = \begin{cases} 0 & \text{for } m \leq 0 \\ 1/k & \text{for } k = 1, 2, \dots \end{cases}$$

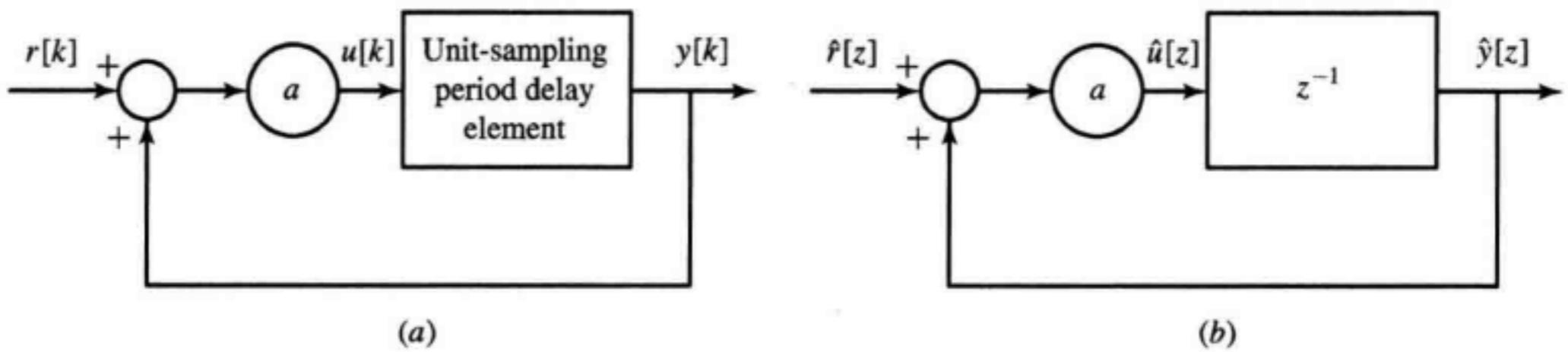


FIGURE 2.17 Discrete-time feedback system.

Using the Maclaurin series

$$\ln(1-x) = -\sum_{m=1}^{\infty} \frac{x^m}{m}$$

we can compute the transfer function or the z -transform of $g[k] = 1/k$ as

$$\hat{g}(z) = \sum_{k=1}^{\infty} \frac{z^{-k}}{k} = -\ln(1-z^{-1}) \quad (2.54)$$

It is not a rational function of z . Such a system is a distributed system. We study in this text only lumped systems, and their discrete transfer functions are all rational functions of z .

Discrete rational transfer functions can be proper or improper. If a transfer function is improper such as $\hat{g}(z) = (z^2 + 2z - 1)/(z - 0.5)$, then

$$\frac{\hat{y}(z)}{\hat{u}(z)} = \frac{z^2 + 2z - 1}{z - 0.5} = \frac{1 + 2z^{-1} - z^{-2}}{z^{-1} - 0.5z^{-2}}$$

which implies

$$z^{-1}\hat{y}(z) - 0.5z^{-2}\hat{y}(z) = \hat{u}(z) + 2z^{-1}\hat{u}(z) - z^{-2}\hat{u}(z)$$

or, in the time domain,

$$y[k-1] - 0.5y[k-2] = u[k] + 2u[k-1] - u[k-2]$$

Thus we have

$$y[k-1] = 0.5y[k-2] + u[k] + 2u[k-1] - u[k-2]$$

for all integers $k \geq 1$. Note that we have implicitly assumed that $y[k] = 0$ for all $k < 0$. If $k = 1$, then $y[0]$ depends on $u[1]$, a future input, and the system is not causal. In general, a discrete-time system described by an improper transfer function is not causal. We study only causal systems. Thus all discrete rational transfer functions will be proper. We mentioned earlier that we also study only proper rational transfer functions of s in the continuous-time case. The reason, however, is different. Consider $\hat{g}(s) = s$ or $y(t) = du(t)/dt$. It is a pure differentiator. If we define the differentiation as

$$y(t) = \frac{du(t)}{dt} = \lim_{\Delta \rightarrow 0} \frac{u(t + \Delta) - u(t)}{\Delta}$$

where $\Delta > 0$, then the output $y(t)$ depends on future input $u(t + \Delta)$ and the differentiator is not causal. However, if we define the differentiation as

$$y(t) = \frac{du(t)}{dt} = \lim_{\Delta \rightarrow 0} \frac{u(t) - u(t - \Delta)}{\Delta}$$

then the output $y(t)$ does not depend on future input and the differentiator is causal. Therefore in continuous-time systems, it is open to argument whether an improper transfer function represents a noncausal system. However, improper transfer functions of s

will amplify high frequency noise which often exists in electrical systems. Therefore, improper transfer functions are avoided in electrical systems.

State-space equations

Every DT SISO linear time-invariant lumped system can be described by

$$\begin{aligned}\mathbf{x}[k+1] &= \mathbf{Ax}[k] + \mathbf{bu}[k] \\ y[k] &= \mathbf{cx}[k] + du[k]\end{aligned}\quad (2.55)$$

where \mathbf{A} , \mathbf{b} , \mathbf{c} , and d are independent of the time index k . Let $\hat{\mathbf{x}}(z)$ be the z -transform of $\mathbf{x}[k]$ or

$$\hat{\mathbf{x}}(z) = \mathcal{Z}[\mathbf{x}[k]] := \sum_{k=0}^{\infty} \mathbf{x}[k] z^{-k}$$

Then we have

$$\begin{aligned}\mathcal{Z}[\mathbf{x}[k+1]] &= \sum_{k=0}^{\infty} \mathbf{x}[k+1] z^{-k} = z \sum_{k=0}^{\infty} \mathbf{x}[k+1] z^{-(k+1)} \\ &= z \left[\sum_{l=1}^{\infty} \mathbf{x}[l] z^{-l} + \mathbf{x}[0] - \mathbf{x}[0] \right] = z(\hat{\mathbf{x}}(z) - \mathbf{x}[0])\end{aligned}$$

Applying the z -transform to (2.55) yields

$$\begin{aligned}z\hat{\mathbf{x}}(z) - z\mathbf{x}[0] &= \mathbf{A}\hat{\mathbf{x}}(z) + \mathbf{b}\hat{u}(z) \\ \hat{y}(z) &= \mathbf{c}\hat{\mathbf{x}}(z) + d\hat{u}(z)\end{aligned}$$

which implies

$$\hat{\mathbf{x}}(z) = (z\mathbf{I} - \mathbf{A})^{-1} z\mathbf{x}[0] + (z\mathbf{I} - \mathbf{A})^{-1} \mathbf{b}\hat{u}(z) \quad (2.56)$$

and

$$\hat{y}(z) = \mathbf{c}(z\mathbf{I} - \mathbf{A})^{-1} z\mathbf{x}[0] + \mathbf{c}(z\mathbf{I} - \mathbf{A})^{-1} \mathbf{b}\hat{u}(z) + d\hat{u}(z) \quad (2.57)$$

which is the discrete counterparts of (2.11). Note that there is an extra z in front of $\mathbf{x}[0]$. If $\mathbf{x}[0] = \mathbf{0}$, then (2.57) reduces to

$$\hat{y}(z) = [\mathbf{c}(z\mathbf{I} - \mathbf{A})^{-1} \mathbf{b} + d]\hat{u}(z) \quad (2.58)$$

Comparing this with $\hat{y}(z) = \hat{g}(z)\hat{u}(z)$ yields

$$\hat{g}(z) = \mathbf{c}(z\mathbf{I} - \mathbf{A})^{-1} \mathbf{b} + d \quad (2.59)$$

This is the discrete counterpart of (2.12). If the Laplace transform variable s is replaced by the z -transform variable z , then the two equations are identical.

Example 2.8.3 Consider a money market account in a brokerage firm. Let $u[k]$ be the amount of money deposited into or withdrawn from the account in the k th day and let $y[k]$ be the total amount of money in the account at the end of the k th day. Then the account can be considered as a discrete-time system with input $u[k]$ and output $y[k]$.

The system is nonlinear if the interest rate depends on the amount of money in the account. If the interest rate is the same no matter how much money in the account, then it is a linear system. It is a time-varying system if the interest rate changes with time; a time-invariant system if the interest rate is fixed. We consider here only the LTI case with interest rate $r = 0.015\%$ per day and compounded daily.

If we deposit one dollar on the first day (that is, $u[0] = 1$) and nothing thereafter ($u[k] = 0$, $k = 1, 2, \dots$), then $y[0] = u[0] = 1$ and $y[1] = 1 + 0.00015 = 1.00015$. Because the money is compounded daily, we have

$$y[2] = y[1] + y[1] \cdot 0.00015 = y[1] \cdot 1.00015 = (1.00015)^2$$

and, in general,

$$y[k] = (1.00015)^k$$

Because the input $\{1, 0, 0, \dots\}$ is an impulse sequence, the output is, by definition, the impulse response sequence or

$$g[k] = (1.00015)^k$$

and the input-output description of the account is

$$y[k] = \sum_{m=0}^k g[k-m]u[m] = \sum_{m=0}^k (1.00015)^{k-m}u[m] \quad (2.60)$$

The discrete transfer function is the z -transform of the impulse response sequence or

$$\begin{aligned} \hat{g}(z) &= \mathcal{Z}[g[k]] = \sum_{k=0}^{\infty} (1.00015)^k z^{-k} = \sum_{k=0}^{\infty} (1.00015 z^{-1})^k \\ &= \frac{1}{1 - 1.00015 z^{-1}} = \frac{z}{z - 1.00015} \end{aligned} \quad (2.61)$$

Whenever we use (2.60) or (2.61), the initial state must be zero, that is, there is initially no money in the account.

Next we develop a state-space equation to describe the account. Suppose $y[k]$ is the total amount of money at the end of the k th day. Then we have

$$y[k+1] = y[k] + 0.00015y[k] + u[k+1] = 1.00015y[k] + u[k+1] \quad (2.62)$$

If we define the state variable as $x[k] := y[k]$, then

$$\begin{aligned}x[k+1] &= 1.00015x[k] + u[k+1] \\y[k] &= x[k]\end{aligned}\tag{2.63}$$

Because of $u[k+1]$, (2.63) is not in the standard form of (2.55). Thus we cannot select $x[k] := y[k]$ as a state variable. Next we select a different state variable as

$$x[k] := y[k] - u[k]$$

Substituting $y[k+1] = x[k+1] + u[k+1]$ and $y[k] = x[k] + u[k]$ into (2.62) yields

$$\begin{aligned}x[k+1] &= 1.00015x[k] + 1.00015u[k] \\y[k] &= x[k] + u[k]\end{aligned}\tag{2.64}$$

This is in the standard form and describes the money market account.

We discuss so far only DT LTI SISO systems. Its extension to the MIMO case is similar to the CT case and will not be repeated.

2.9 Concluding Remarks

We introduced in this chapter the concepts of causality, lumpedness, time invariance (TI), and linearity (L). We then discussed the following four types of equations to describe LTI-lumped causal systems:

1. Convolutions
2. Transfer functions
3. State-space equations
4. High-order differential equations

The convolution is developed using explicitly the conditions of time-invariance and linearity, thus its derivation is instructive. Applying the Laplace transform to a convolution leads to the concept of transfer functions. The convolution will be used once again in Chapter 5 to establish an important stability condition. Thus its introduction is important. On the other hand, its actual employment is very complicated as shown in page 28. Thus the convolution is not used in either analysis or design. High-order differential equations are generally difficult to develop and are not suitable for computer computation. Thus we study in this text mostly transfer functions and state-space equations.

From the examples in this chapter, we saw that if a system contains nonlinear elements, we must first develop a set of first-order nonlinear differential equations and then linearize it to become a state-space equation. We can then use (2.12) or (2.16) to compute its transfer function. Developing directly the transfer function would be difficult or impossible. Thus the formulation of state-space equations is very important.

Problems

- 2.1** Consider the memoryless systems with characteristics shown in Fig. 2.18, in which u denotes the input and y the output. Which of them is a linear system? Is it possible to introduce a new output so that the system in Fig. 2.18(b) is linear?

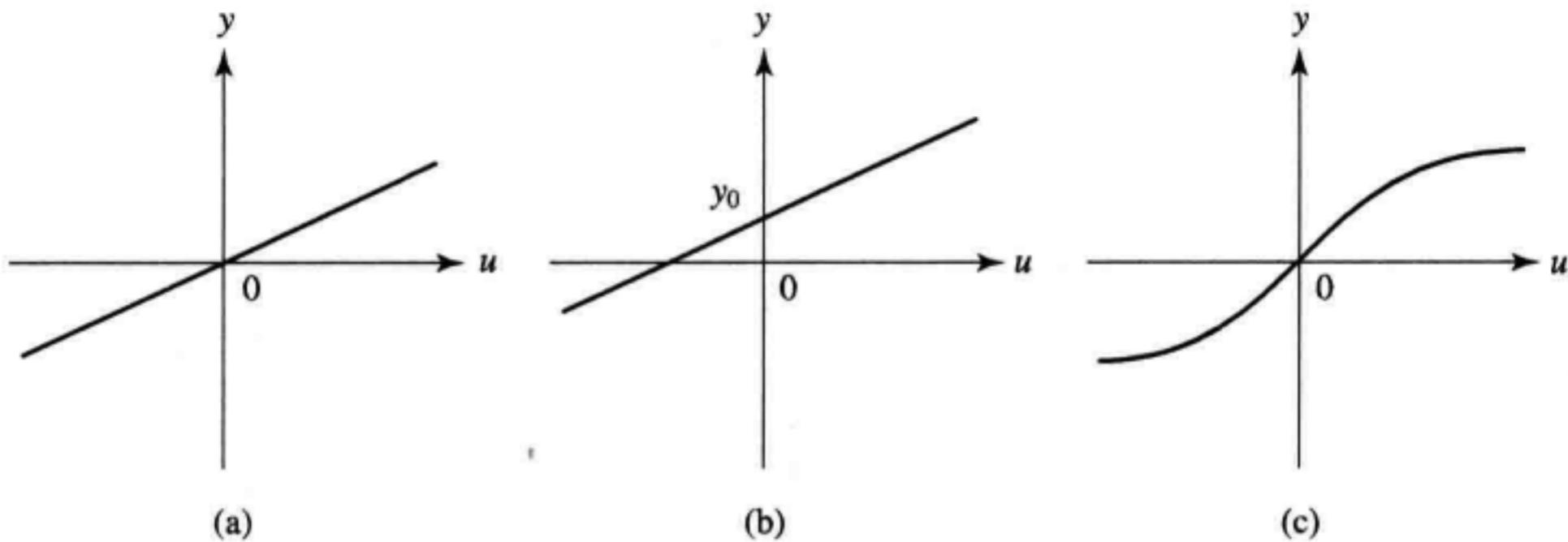


FIGURE 2.18

- 2.2** The impulse response of an ideal lowpass filter is given by

$$g(t) = 2\omega \frac{\sin 2\omega(t - t_0)}{2\omega(t - t_0)}$$

for all t , where ω and t_0 are constants. Is the ideal lowpass filter causal? Is it possible to build the filter in the real world?

- 2.3** Consider a system whose input u and output y are related by

$$y(t) = (P_\alpha u)(t) := \begin{cases} u(t) & \text{for } t \leq \alpha \\ 0 & \text{for } t > \alpha \end{cases}$$

where α is a fixed constant. The system is called a *truncation operator*, which chops off the input after time α . Is the system linear? Is it time-invariant? Is it causal?

- 2.4** The input and output of an initially relaxed system can be denoted by $y = Hu$, where H is some mathematical operator. Show that if the system is causal, then

$$P_\alpha y = P_\alpha Hu = P_\alpha HP_\alpha u$$

where P_α is the truncation operator defined in Problem 2.3. Is it true $P_\alpha Hu = HP_\alpha u$?

- 2.5** Consider a system with input u and output y . Three experiments are performed on the system using the input $u_1(t)$, $u_2(t)$, and $u_3(t)$ for $t \geq 0$. In each case, the initial state $\mathbf{x}(0)$ at time $t = 0$ is the same. The corresponding outputs are denoted by y_1 , y_2 and y_3 . Which of the following statements are correct if $\mathbf{x}(0) \neq 0$?

1. If $u_3 = u_1 + u_2$, then $y_3 = y_1 + y_2$.
2. If $u_3 = 0.5(u_1 + u_2)$, then $y_3 = 0.5(y_1 + y_2)$.
3. If $u_3 = u_1 - u_2$, then $y_3 = y_1 - y_2$.

Which are correct if $\mathbf{x}(0) = \mathbf{0}$?

- 2.6** Consider a system whose input and output are related by

$$y(t) = \begin{cases} u^2(t)/u(t-1) & \text{if } u(t-1) \neq 0 \\ 0 & \text{if } u(t-1) = 0 \end{cases}$$

for all t . Show that the system satisfies the homogeneity property but not the additivity property.

- 2.7** Show that if the additivity property holds, then the homogeneity property holds for all rational number α . Thus if a system has some “continuity” property, then additivity implies homogeneity.
- 2.8** Find a state-space equation to describe the circuit shown in Fig. 2.2. Compute also its transfer matrix.
- 2.9** Find a state-space equation to describe the circuit shown in Fig. 2.19. Find also its transfer function.

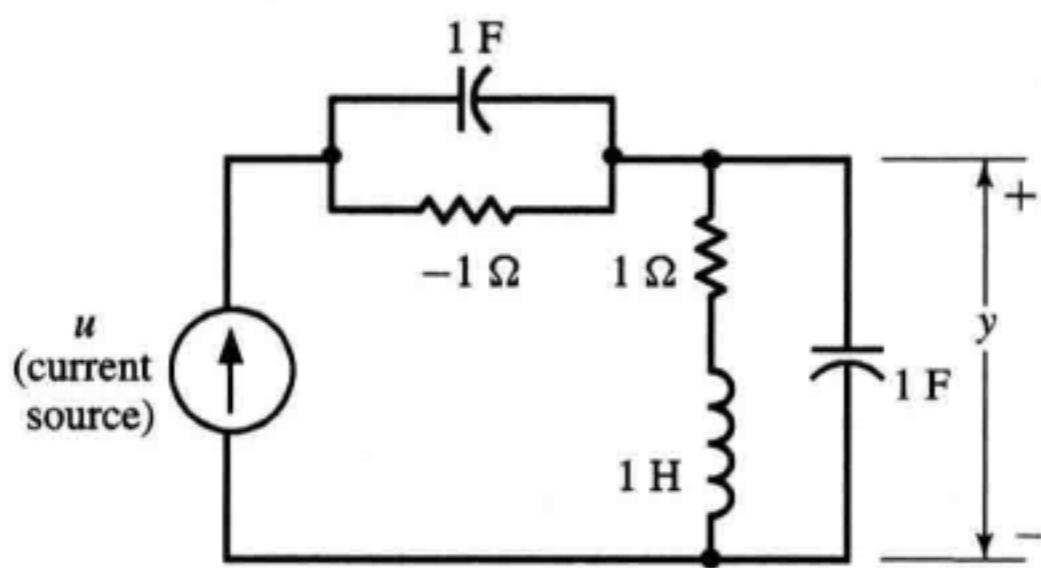


FIGURE 2.19

- 2.10** Find a state-space equation to describe the circuit shown in Fig. 2.20. Find also its transfer function.

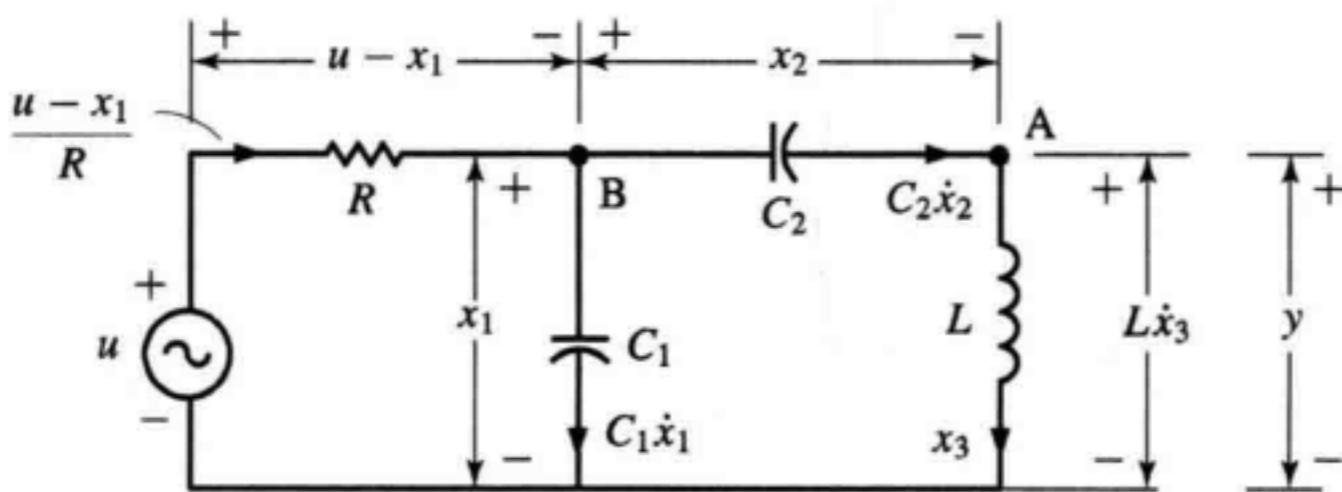


FIGURE 2.20

- 2.11** Develop a one-dimensional state-space equation and a two-dimensional state-space equation to describe the circuit shown in Fig. 2.21(a). Also compute the transfer function of the circuit using (2.12). See Problem 7.4.
- 2.12** Develop a one-dimensional state-space equation and a two-dimensional state-space equation to describe the circuit shown in Fig. 2.21(b). Also compute the transfer function of the circuit using impedances. See Problem 7.5.
- 2.13** Consider the circuit shown in Fig. 2.22. (a) Verify that if we assign the 1-F capacitor voltage as $x_1(t)$, the inductor current as $x_2(t)$, and the 2-F capacitor voltage as $x_3(t)$, then

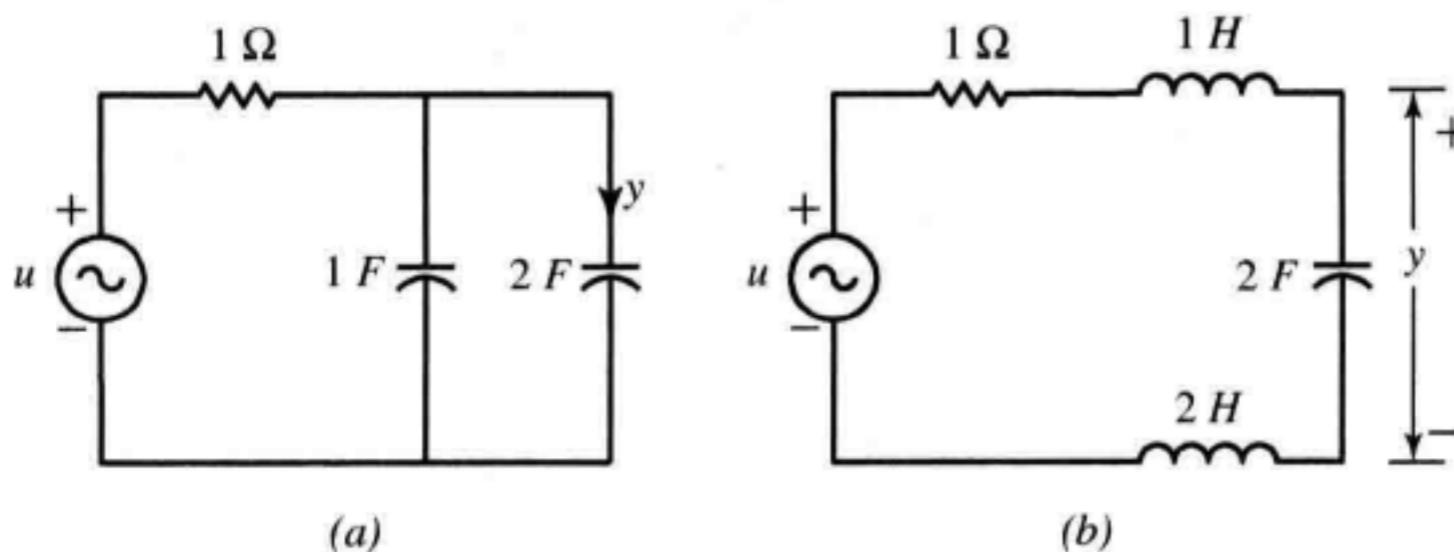


FIGURE 2.21 (a) Circuit in which not all capacitor voltages are assigned as state variables. (b) Circuit in which not all inductor currents are assigned as state variables.

the circuit can be described by

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1/3 & 0 \\ 0 & -2 & 1 \\ 0 & -1/3 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 2/3 \\ 0 \\ 1/3 \end{bmatrix} u(t)$$

$$y(t) = [0 \quad -2 \quad 1] \mathbf{x}(t)$$

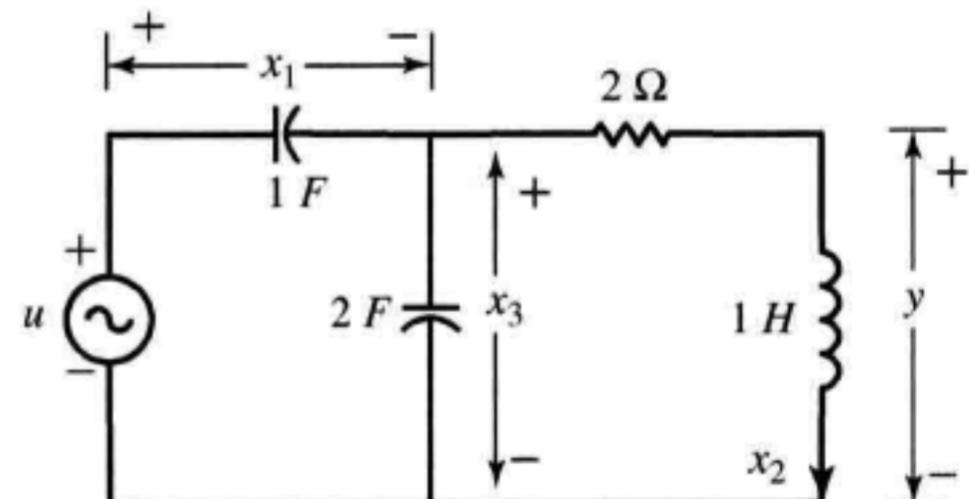
where $\mathbf{x} = [x_1 \ x_2 \ x_3]'$. Is it in the standard state-space equation form? (b) Verify that if we assign only $x_1(t)$ and $x_2(t)$ (without assign $x_3(t)$), then the circuit can be described by

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1/3 \\ -1 & -2 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t) + \begin{bmatrix} 2/3 \\ 0 \end{bmatrix} \dot{u}(t)$$

$$y(t) = [-1 \quad -2] \mathbf{x}(t) + u(t)$$

where $\mathbf{x} = [x_1 \ x_2]'$. Is it in the standard state-space equation form? Note that $x_1(t) + x_3(t) = u(t)$ for all t ; thus the 2-F capacitor is, in some sense, redundant. See Problem 7.6.

FIGURE 2.22 Circuit which cannot be described by a standard state-space equation.



2.14 Consider a system described by

$$\ddot{y} + 2\dot{y} - 3y = \dot{u} - u$$

What are the transfer function and the impulse response of the system?

2.15 Let $\bar{y}(t)$ be the unit step response of a linear time-invariant system. Show that the impulse response of the system equals $d\bar{y}(t)/dt$.

- 2.16** Consider a two-input and two-output system described by

$$D_{11}(p)y_1(t) + D_{12}(p)y_2(t) = N_{11}(p)u_1(t) + N_{12}(p)u_2(t)$$

$$D_{21}(p)y_1(t) + D_{22}(p)y_2(t) = N_{21}(p)u_1(t) + N_{22}(p)u_2(t)$$

- where N_{ij} and D_{ij} are polynomials of $p := d/dt$. What is the transfer matrix of the system?
- 2.17** Consider the feedback systems shown in Fig. 2.5. Show that the unit-step responses of the positive-feedback system are as shown in Fig. 2.23(a) for $a = 1$ and in Fig. 2.23(b) for $a = 0.5$. Show also that the unit-step response of the negative-feedback system are as shown in Figs. 2.21(c) and 2.21(d), respectively, for $a = 1$ and $a = 0.5$.

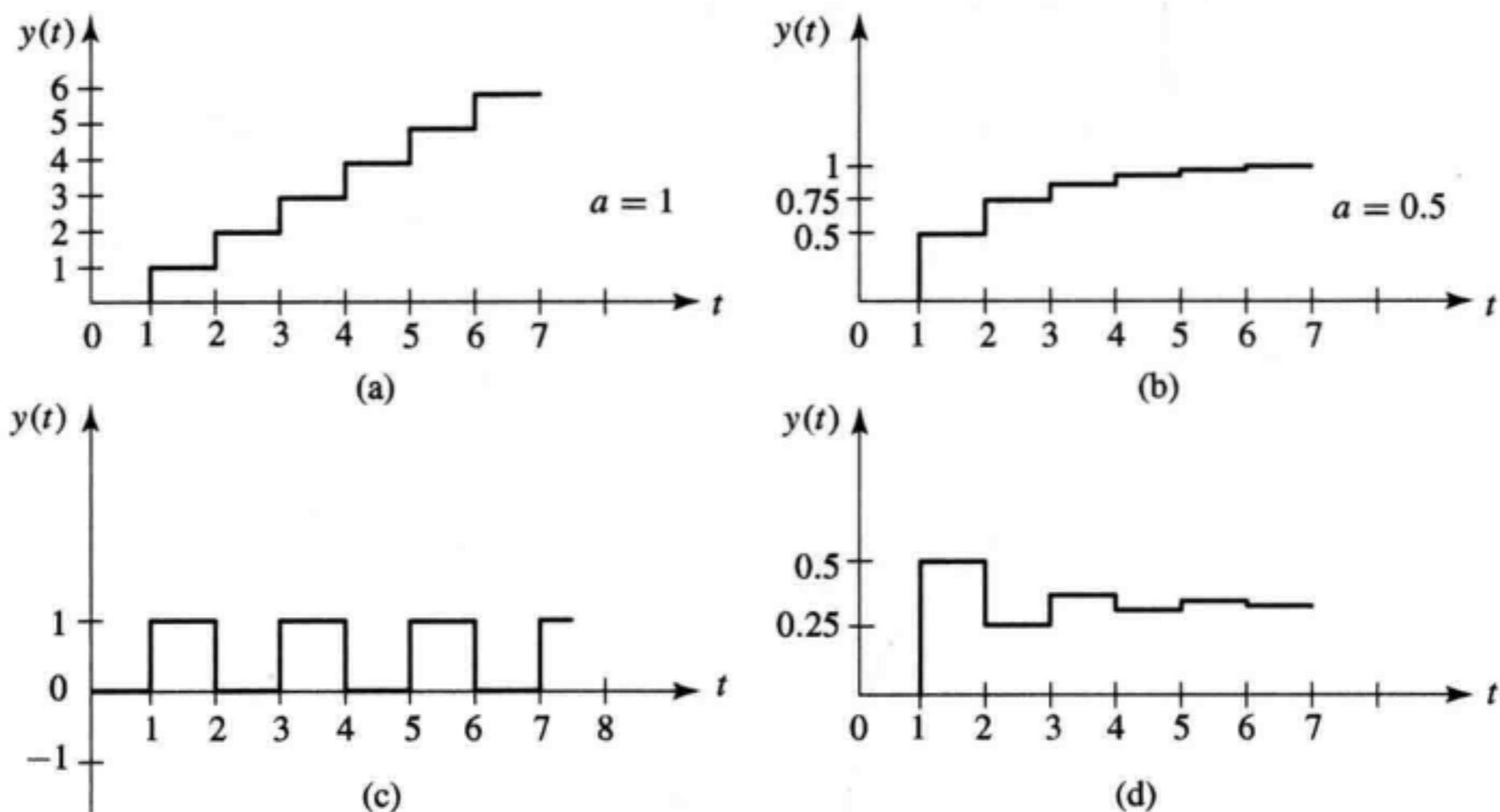


FIGURE 2.23

- 2.18** The transfer function of a second order system is usually described as $\hat{g}(s) = \frac{k\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2}$ where ξ , ω_n , and k are damping factor, natural frequency and dc gain, respectively.

1. Write a state-space equation to describe the system.
2. Suppose the numerator of $\hat{g}(s)$ is changed to $s + a$. Write a state-space equation for this case.

- 2.19** Consider the system described by the coupled differential equations,

$$\begin{aligned}\ddot{y}_1 + k_1\dot{y}_1 + k_2y_1 &= u_1 + k_3u_2 \\ \ddot{y}_2 + k_4y_2 + k_5\dot{y}_1 &= k_6u_1\end{aligned}$$

where u_1 and u_2 are inputs and y_1 and y_2 are outputs.

Define the state variables as the outputs and the derivatives of the outputs, as needed, and write the state-space equation of the system.

- 2.20** Obtain a state-space equation for the following time-varying system

$$\ddot{y} + e^{-2t}\dot{y} + e^{-t}y = u$$

- 2.21** Write a state-space equation for the following discrete-time system represented by the difference equation,

$$y(k+3) + 2y(k+2) + 3y(k+1) + y(k) = u(k)$$

- Obtain the transfer function of the system.
- 2.22** Consider the state equation of a discrete-time system given by,

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 0.4 \\ 0.2 & 0.3 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(k)$$

- Obtain its corresponding difference equation.
- 2.23** Find state-space equations to describe the pendulum systems in Fig. 2.24. The systems are useful to model one- or two-link robotic manipulators. If θ , θ_1 , and θ_2 are very small, can you consider the two systems as linear?

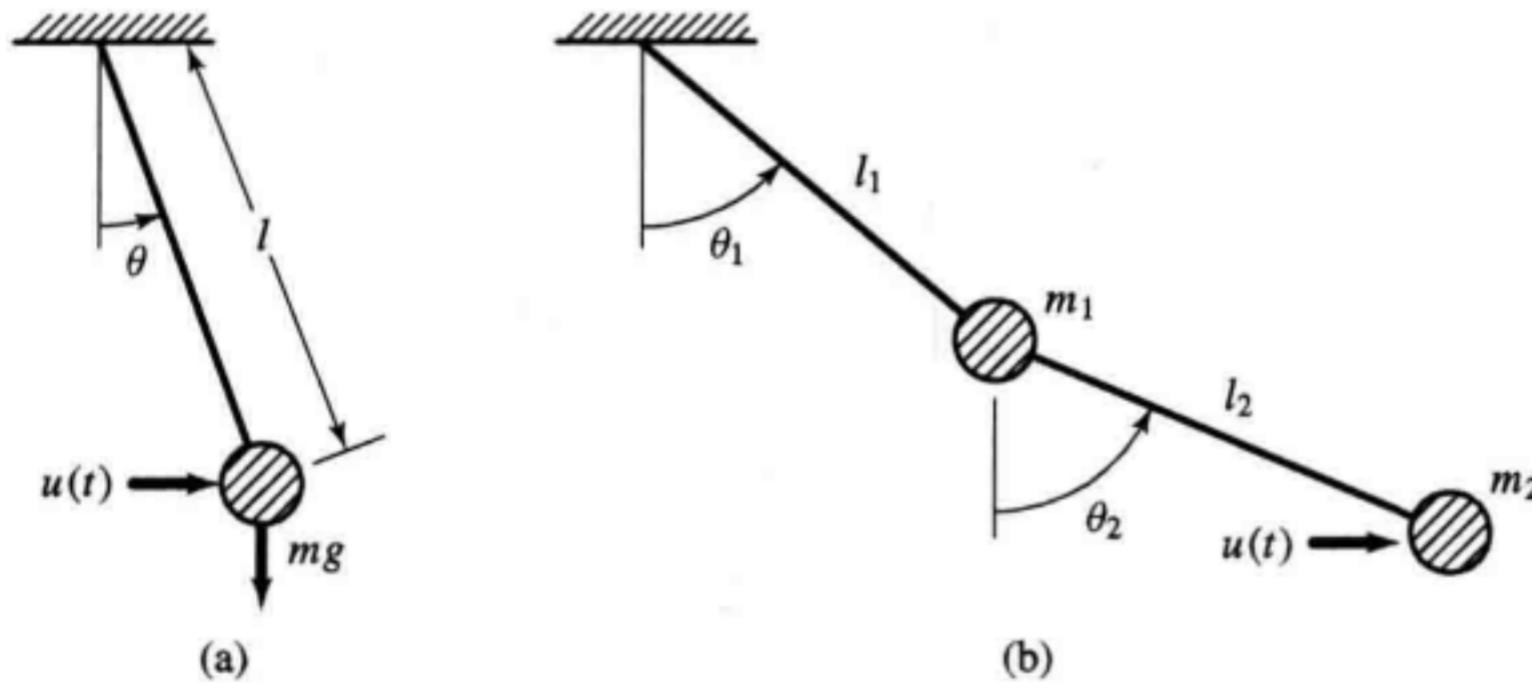


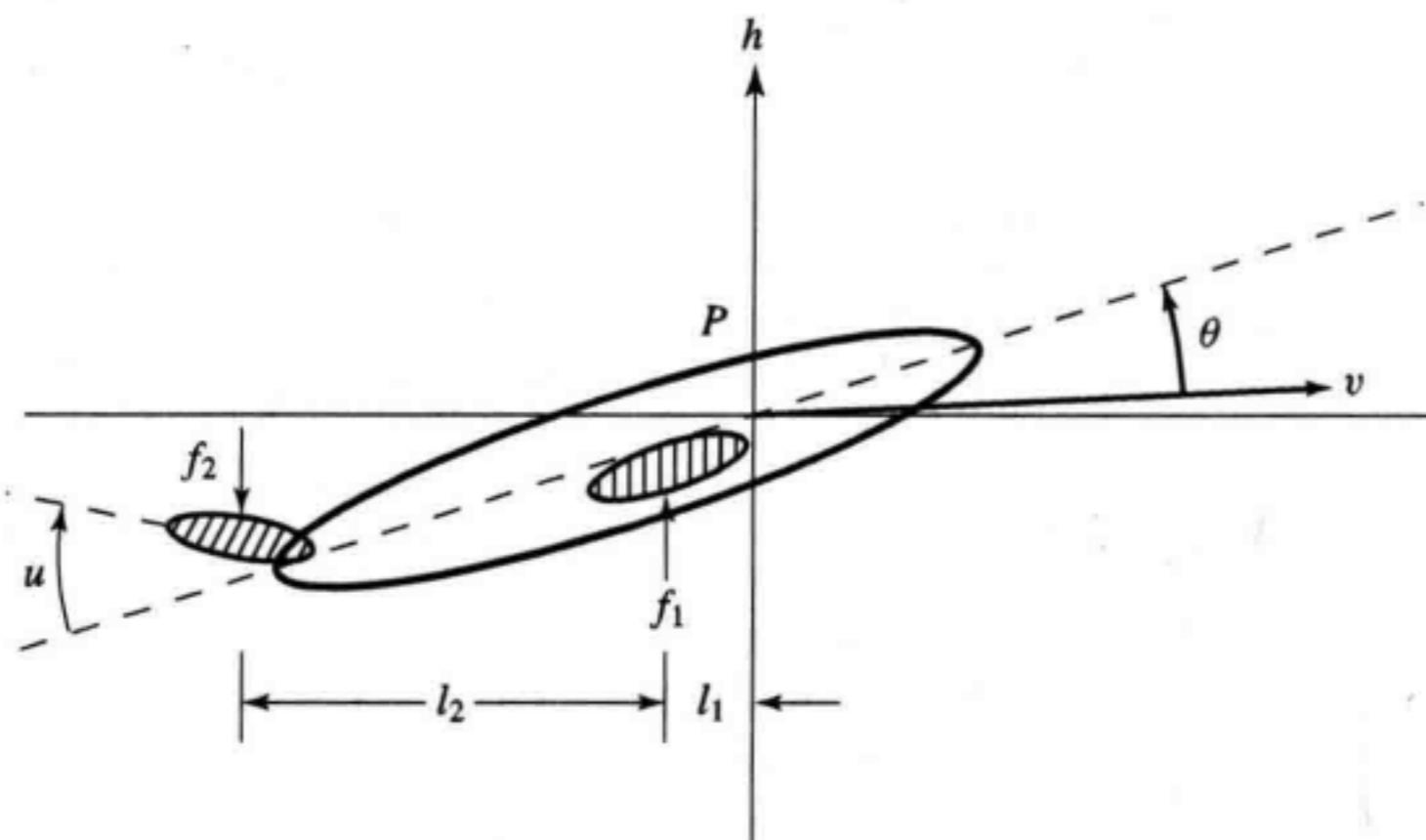
FIGURE 2.24

- 2.24** Consider the pendulum shown in Fig 2.24(a). Suppose there is a friction force resisting the motion, which is assumed to be proportional to the speed of the pendulum bob with a coefficient of friction k .

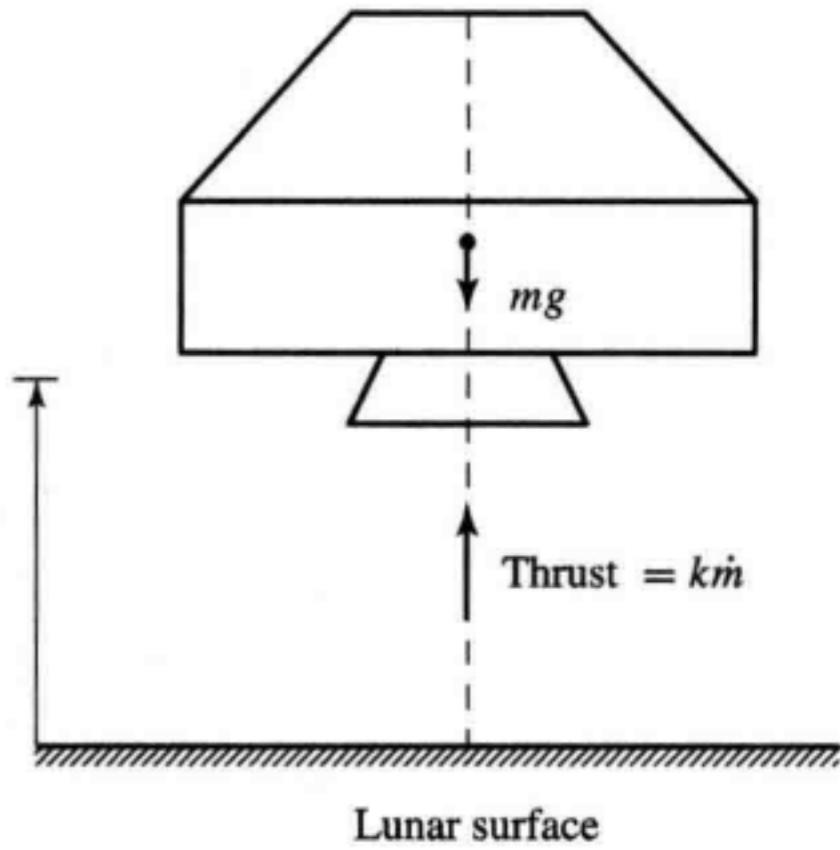
1. Write the equation of motion in the tangential direction when θ is not small. Is the system linear?
2. Define the state variables as $\theta = x_1$ and $\dot{\theta} = \dot{x}_1 = x_2$ and write the state-space equation of the system when θ is large.
3. A more general version of the pendulum equation arises when a torque τ is applied to the pendulum. Write the state-space equation for both large and small θ .

- 2.25** Consider the simplified model of an aircraft shown in Fig. 2.25. It is assumed that the aircraft is in an equilibrium state at the pitched angle θ_0 , elevator angle u_0 , altitude h_0 , and cruising speed v_0 . It is assumed that small deviations of θ and u from θ_0 and u_0 generate forces $f_1 = k_1\theta$ and $f_2 = k_2u$ as shown in the figure. Let m be the mass of the aircraft, I the moment of inertia about the center of gravity P , $b\dot{\theta}$ the aerodynamic damping, and h the deviation of the altitude from h_0 . Find a state-space equation to describe the system. Show also that the transfer function from u to h , by neglecting the effect of I , is

$$\hat{g}(s) = \frac{\hat{h}(s)}{\hat{u}(s)} = \frac{k_1 k_2 l_2 - k_2 b s}{m s^2 (b s + k_1 l_1)}$$

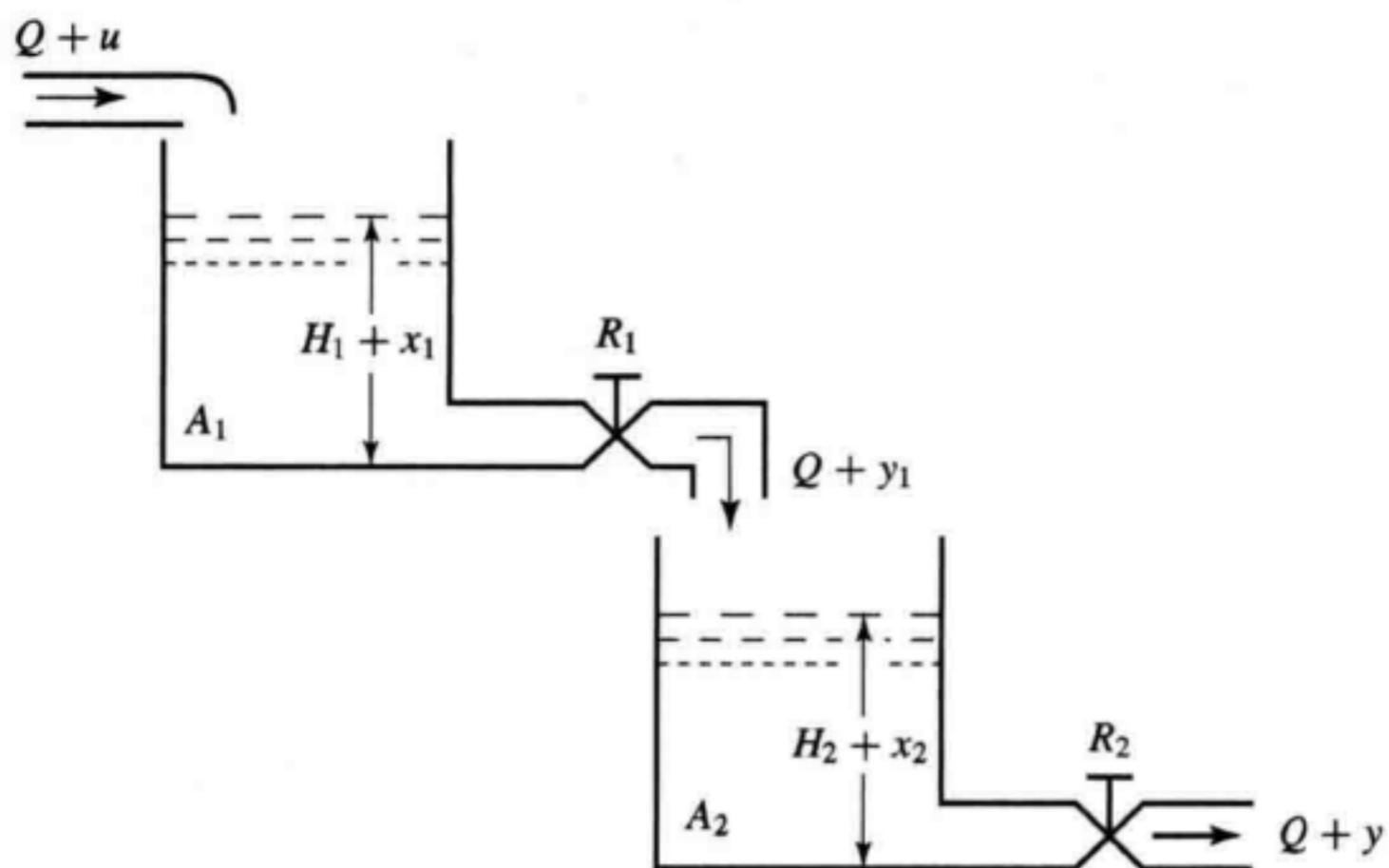
**FIGURE 2.25**

- 2.26** The soft landing phase of a lunar module descending on the moon can be modeled as shown in Fig. 2.26. The thrust generated is assumed to be proportional to \dot{m} , where m is

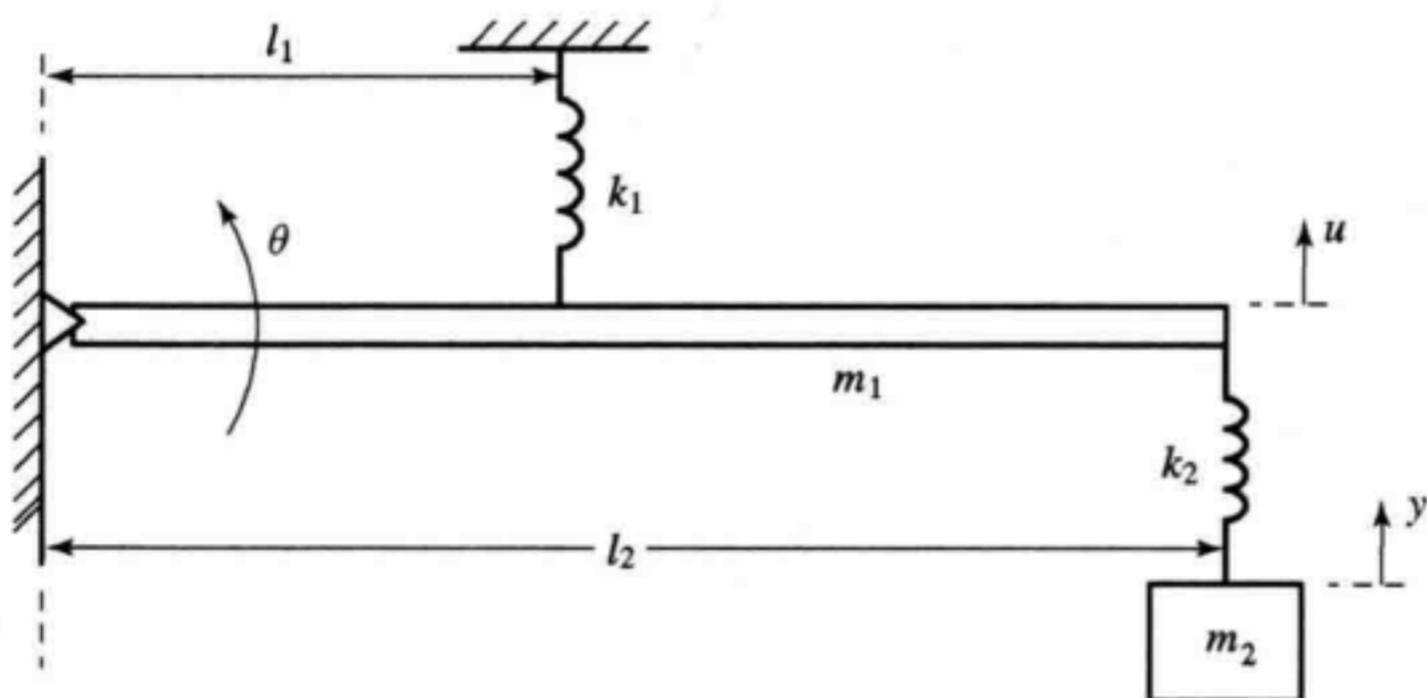
FIGURE 2.26

the mass of the module. Then the system can be described by $m\ddot{y} = -km - mg$, where g is the gravity constant on the lunar surface. Define state variables of the system as $x_1 = y$, $x_2 = \dot{y}$, $x_3 = m$, and $u = \dot{m}$. Find a state-space equation to describe the system.

- 2.27** Find the transfer functions from u to y_1 and from y_1 to y of the hydraulic tank system shown in Fig. 2.27. Does the transfer function from u to y equal the product of the two transfer functions? Is this also true for the system shown in Fig. 2.16? [Answer: No, because of the *loading* problem in the two tanks in Fig. 2.16. The loading problem is an important issue in developing mathematical equations to describe composite systems. See Reference 7.]

**FIGURE 2.27**

- 2.28** Consider the mechanical system shown in Fig. 2.28. Let I denote the moment of inertia of the bar and block about the hinge. It is assumed that the angular displacement θ is very small. An external force u is applied to the bar as shown. Let y be the displacement of the block, with mass m_2 , from the equilibrium. Find a state-space equation to describe the system. Find also the transfer function from u to y .

**FIGURE 2.28**

Linear Algebra

3.1 Introduction

This chapter reviews a number of concepts and results in linear algebra encountered in this text. Most results are developed intuitively in order for the reader to better grasp the ideas. They are stated as theorems for easy reference in later chapters. However, no formal proofs are given. This text studies transfer functions and state-space equations. To study the former, we need only Section 3.3; to study the latter, we need the entire chapter.

As we saw in the preceding chapter, all parameters that arise in the real world are real numbers. Therefore, we deal only with real numbers, unless stated otherwise, throughout this text. Let \mathbf{A} , \mathbf{B} , \mathbf{C} , and \mathbf{D} be, respectively, $n \times m$, $m \times r$, $l \times n$, and $r \times p$ real matrices. Let \mathbf{a}_i be the i th column of \mathbf{A} , and let \mathbf{b}_j be the j th row of \mathbf{B} . Then we have

$$\mathbf{AB} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_m] \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_m \end{bmatrix} = \mathbf{a}_1\mathbf{b}_1 + \mathbf{a}_2\mathbf{b}_2 + \dots + \mathbf{a}_m\mathbf{b}_m \quad (3.1)$$

$$\mathbf{CA} = \mathbf{C}[\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_m] = [\mathbf{Ca}_1 \ \mathbf{Ca}_2 \ \dots \ \mathbf{Ca}_m] \quad (3.2)$$

and

$$\mathbf{BD} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_m \end{bmatrix} \mathbf{D} = \begin{bmatrix} \mathbf{b}_1\mathbf{D} \\ \mathbf{b}_2\mathbf{D} \\ \vdots \\ \mathbf{b}_m\mathbf{D} \end{bmatrix} \quad (3.3)$$

These identities can be easily verified. Note that $\mathbf{a}_i \mathbf{b}_i$ is an $n \times r$ matrix; it is the product of an $n \times 1$ column vector and a $1 \times r$ row vector. The product $\mathbf{b}_i \mathbf{a}_i$ is not defined unless $n = r$; it becomes a scalar if $n = r$.

3.2 Basis, Representation, and Orthonormalization

Consider an n -dimensional real linear space, denoted by \mathcal{R}^n . Every vector in \mathcal{R}^n is an n -tuple of real numbers such as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

To save space, we write it as $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]'$, where the prime denotes the transpose. We call $\mathbf{x} = [0 \ 0 \ \dots \ 0]' =: \mathbf{0}$ the *zero vector*.

The set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ in \mathcal{R}^n is said to be *linearly dependent* if there exist real numbers $\alpha_1, \alpha_2, \dots, \alpha_m$, not all zero, such that

$$\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \dots + \alpha_m \mathbf{x}_m = \mathbf{0} \quad (3.4)$$

If the only set of α_i for which (3.4) holds is $\alpha_1 = 0, \alpha_2 = 0, \dots, \alpha_m = 0$, then the set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ is said to be *linearly independent*.

If the set of vectors in (3.4) is linearly dependent, then there exists at least one α_i , say α_1 , which is different from zero. Then (3.4) implies

$$\begin{aligned} \mathbf{x}_1 &= \frac{-1}{\alpha_1} [\alpha_2 \mathbf{x}_2 + \alpha_3 \mathbf{x}_3 + \dots + \alpha_m \mathbf{x}_m] \\ &=: \beta_2 \mathbf{x}_2 + \beta_3 \mathbf{x}_3 + \dots + \beta_m \mathbf{x}_m \end{aligned}$$

where $\beta_i = -\alpha_i/\alpha_1$. Such an expression is called a linear combination. In conclusion, if a set of vectors is linearly dependent, then the set contains at least one vector that can be expressed as a linear combination of the remaining vectors and is said to be linearly dependent in the set. If a set of vectors is linearly independent, then no vector in the set can be expressed as a linear combination of the remaining vectors. Note that the coefficients β_i in the preceding equation can all be zero. Thus if a set contains a zero vector, it is linearly dependent.

The dimension of a linear space can be defined as the maximum number of linearly independent vectors in the space. Thus in \mathcal{R}^n , we can find at most n linearly independent vectors. For example, consider \mathcal{R}^2 or a plane. We first select a point on the plane as the origin. Then every point, say A, on the plane is a vector. The vector can be denoted by a straight line emitting from the origin and entering into A with an arrow head. If point A coincides with the origin, it is the zero vector. In \mathcal{R}^2 , any *nonzero* vector is linearly independent by itself. Any set that consists of two nonzero vectors not on the same straight line is linearly independent. We cannot find three or more linearly independent vectors in \mathcal{R}^2 .

Basis and representation

A set of linearly independent vectors in \mathcal{R}^n is called a *basis* if every vector in \mathcal{R}^n can be expressed as a unique linear combination of the set. In \mathcal{R}^n , any set of n linearly independent vectors can be used as a basis. Let $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$ be such a set. Then every vector \mathbf{x} can be expressed uniquely as

$$\mathbf{x} = \alpha_1 \mathbf{q}_1 + \alpha_2 \mathbf{q}_2 + \cdots + \alpha_n \mathbf{q}_n \quad (3.5)$$

Define the $n \times n$ square matrix

$$\mathbf{Q} := [\mathbf{q}_1 \ \mathbf{q}_2 \ \cdots \ \mathbf{q}_n] \quad (3.6)$$

Then (3.5) can be written as

$$\mathbf{x} = \mathbf{Q} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} =: \mathbf{Q}\bar{\mathbf{x}} \quad (3.7)$$

We call $\bar{\mathbf{x}} = [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n]'$ the *representation* of the vector \mathbf{x} with respect to the basis $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$.

We will associate with every \mathcal{R}^n the following *orthonormal basis*

$$\mathbf{i}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{i}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \dots, \mathbf{i}_{n-1} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{i}_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad (3.8)$$

With respect to this basis, we have

$$\mathbf{x} := \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \mathbf{i}_1 + x_2 \mathbf{i}_2 + \cdots + x_n \mathbf{i}_n = \mathbf{I}_n \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

where \mathbf{I}_n is the $n \times n$ unit matrix. In other words, the representation of any vector \mathbf{x} with respect to the orthonormal basis in (3.8) equals itself. A vector has different representations with respect to different sets of basis, as the next example shows.

Example 3.2.1 Consider the vector $\mathbf{x} = [1 \ 3]'$ in \mathcal{R}^2 as shown in Fig. 3.1. The two vectors $\mathbf{q}_1 = [3 \ 1]'$ and $\mathbf{q}_2 = [2 \ 2]'$ are clearly linearly independent and can be used as a basis. Let us draw from \mathbf{x} a line, in parallel with the \mathbf{q}_2 -line, which intersects the

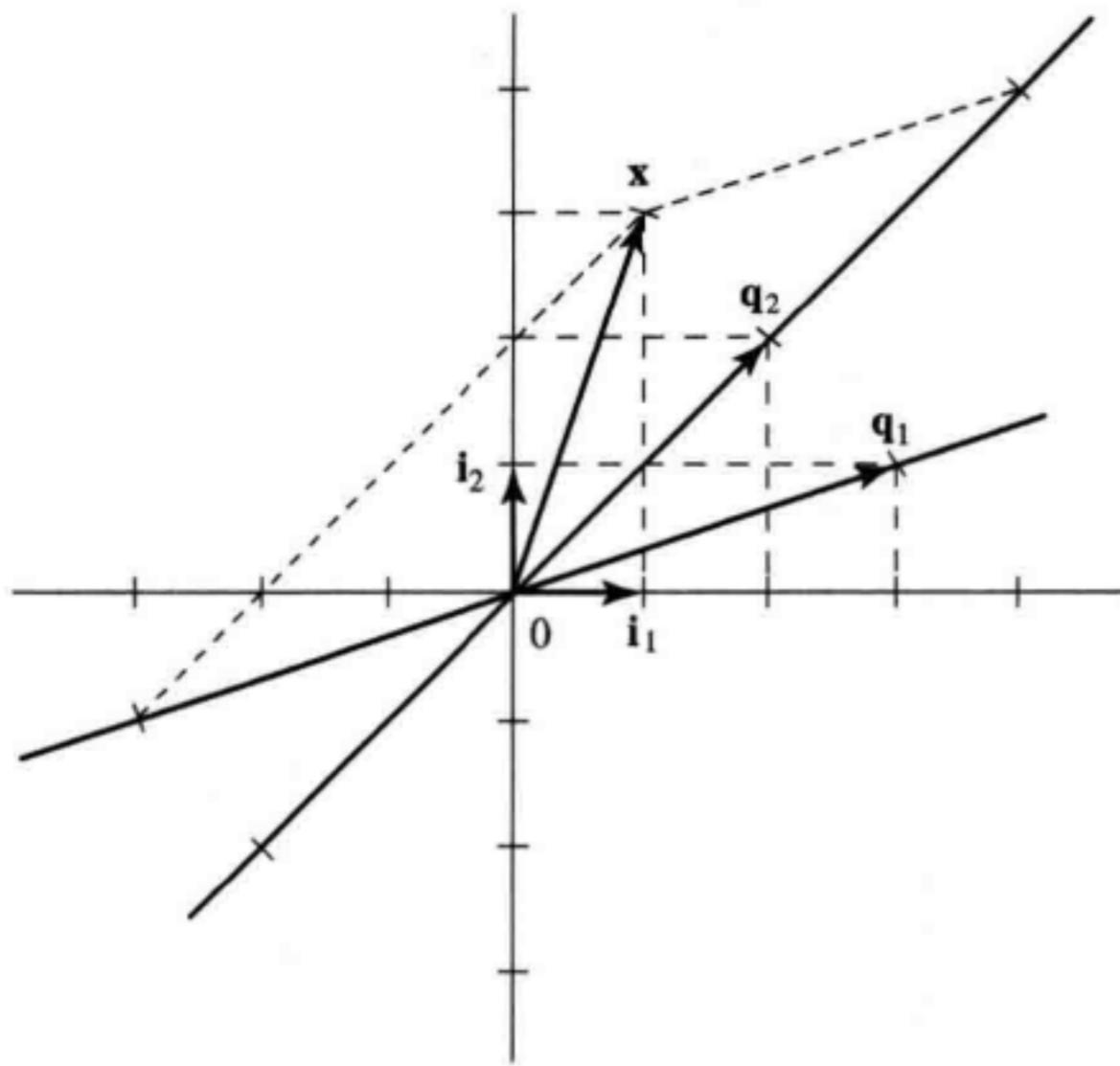


FIGURE 3.1 Different representations of vector \mathbf{x} .

\mathbf{q}_1 -line at $-\mathbf{q}_1$ as shown and a line, in parallel with the \mathbf{q}_1 -line, which intersects the \mathbf{q}_2 -line at $2\mathbf{q}_2$ as shown. Then the representation of \mathbf{x} with respect to $\{\mathbf{q}_1, \mathbf{q}_2\}$ is $[-1 \ 2]'$. This can be verified as

$$\mathbf{x} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} = [\mathbf{q}_1 \ \mathbf{q}_2] \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

To find the representation of \mathbf{x} with respect to the basis $\{\mathbf{q}_2, \mathbf{i}_2\}$, we draw from \mathbf{x} two lines in parallel with \mathbf{i}_2 and \mathbf{q}_2 . They intersect at $0.5\mathbf{q}_2$ and $2\mathbf{i}_2$. Thus the representation of \mathbf{x} with respect to $\{\mathbf{q}_2, \mathbf{i}_2\}$ is $[0.5 \ 2]'$ (Verify).

Norms of vectors

The concept of *norm* is a generalization of length or magnitude. Any real-valued function of \mathbf{x} , denoted by $\|\mathbf{x}\|$, can be defined as a norm if it has the following properties:

1. $\|\mathbf{x}\| \geq 0$ for every \mathbf{x} and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$.
2. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$, for any real α .
3. $\|\mathbf{x}_1 + \mathbf{x}_2\| \leq \|\mathbf{x}_1\| + \|\mathbf{x}_2\|$ for every \mathbf{x}_1 and \mathbf{x}_2 .

The last inequality is called the *triangular inequality*.

Let $\mathbf{x} = [x_1 \ x_2 \ \cdots \ x_n]'$. Then the norm of \mathbf{x} can be chosen as any one of the following

$$\|\mathbf{x}\|_1 := \sum_{i=1}^n |x_i|$$

$$\|\mathbf{x}\|_2 := \sqrt{\mathbf{x}'\mathbf{x}} = \left(\sum_{i=1}^n |x_i|^2 \right)^{1/2}$$

$$\|\mathbf{x}\|_\infty := \max_i |x_i|$$

They are called, respectively, 1-norm, 2- or Euclidean norm, and infinite norm. The 2-norm is the length of the vector from the origin. We use exclusively, unless stated otherwise, the Euclidean norm, and the subscript 2 will be dropped. For example, for the vector $\mathbf{x} = [2 \ -4]'$, we have $\|\mathbf{x}\|_1 = 2 + |-4| = 6$, $\|\mathbf{x}\| = \|\mathbf{x}\|_2 = \sqrt{4+16} = \sqrt{20} = 4.47$, and $\|\mathbf{x}\|_\infty = 4$.

In MATLAB, the norms just introduced can be obtained by using the functions `norm(x, 1)`, `norm(x, 2) = norm(x)`, and `norm(x, inf)`.

Orthonormalization

A vector \mathbf{x} is said to be normalized if its Euclidean norm is 1 or $\mathbf{x}'\mathbf{x} = 1$. Note that $\mathbf{x}'\mathbf{x}$ is scalar and $\mathbf{x}\mathbf{x}'$ is $n \times n$. Two vectors, \mathbf{x}_1 and \mathbf{x}_2 , are said to be *orthogonal* if $\mathbf{x}_1'\mathbf{x}_2 = \mathbf{x}_2'\mathbf{x}_1 = 0$. A set of vectors $\mathbf{x}_i, i = 1, 2, \dots, m$, is said to be *orthonormal* if

$$\mathbf{x}_i'\mathbf{x}_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

Given a set of linearly independent vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m$. We can obtain an orthonormal set using the procedure that follows:

$$\begin{array}{ll} \mathbf{u}_1 := \mathbf{e}_1 & \mathbf{q}_1 := \mathbf{u}_1 / \|\mathbf{u}_1\| \\ \mathbf{u}_2 := \mathbf{e}_2 - (\mathbf{q}_1'\mathbf{e}_2)\mathbf{q}_1 & \mathbf{q}_2 := \mathbf{u}_2 / \|\mathbf{u}_2\| \\ \vdots & \\ \mathbf{u}_m := \mathbf{e}_m - \sum_{k=1}^{m-1} (\mathbf{q}_k'\mathbf{e}_m)\mathbf{q}_k & \mathbf{q}_m := \mathbf{u}_m / \|\mathbf{u}_m\| \end{array}$$

The first equation normalizes the vector \mathbf{e}_1 to have norm 1. The vector $(\mathbf{q}_1'\mathbf{e}_2)\mathbf{q}_1$ is the projection of the vector \mathbf{e}_2 along \mathbf{q}_1 . Its subtraction from \mathbf{e}_2 yields the vertical part \mathbf{u}_2 . It is then normalized to 1 as shown in Fig. 3.2. Using this procedure, we can obtain a set of orthonormal set. This is called the *Schmidt orthonormalization procedure*.

Let $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_m]$ be an $n \times m$ matrix with $m \leq n$. If all columns of \mathbf{A} or $\{\mathbf{a}_i, i = 1, 2, \dots, m\}$ are orthonormal, then

$$\mathbf{A}'\mathbf{A} = \begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \vdots \\ \mathbf{a}_m' \end{bmatrix} [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_m] = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \mathbf{I}_m$$

where \mathbf{I}_m is the unit matrix of order m . Note that if $m < n$, we have $\mathbf{A}\mathbf{A}' \neq \mathbf{I}_n$. See Problem 3.5.

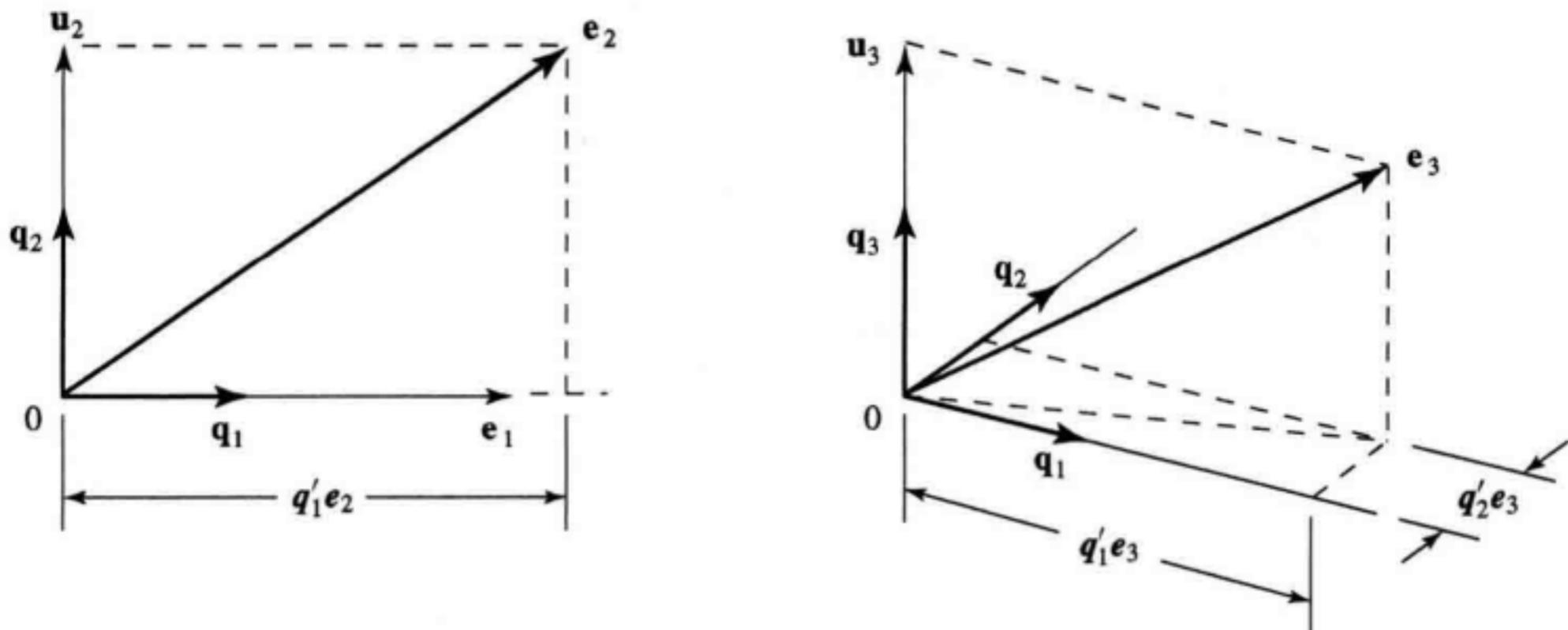


FIGURE 3.2 Schmidt orthonormalization procedure.

Example 3.2.2 Consider the matrix

$$\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_3] = \begin{bmatrix} 6 & 3 & 10 \\ 0 & 2 & 0 \\ 8 & 4 & 5 \end{bmatrix}$$

The three column vectors are linearly independent and can be used as a basis of \mathcal{R}^3 . We develop an orthonormal basis. The vector \mathbf{a}_1 has 2-norm

$$\|\mathbf{a}_1\| = \sqrt{6^2 + 0^2 + 8^2} = \sqrt{100} = 10$$

and can be normalized as

$$\mathbf{q}_1 := \mathbf{a}_1 / \|\mathbf{a}_1\| = \begin{bmatrix} 0.6 \\ 0 \\ 0.8 \end{bmatrix} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_3] \begin{bmatrix} 0.1 \\ 0 \\ 0 \end{bmatrix}$$

where we have also expressed \mathbf{q}_1 in terms of the basis \mathbf{a}_i . Next we define

$$\bar{\mathbf{q}}_2 := \mathbf{a}_2 - 0.5\mathbf{a}_1 = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$$

Because $\mathbf{q}_1' \bar{\mathbf{q}}_2 = 0.6 \times 0 + 0 \times 2 + 0.8 \times 0 = 0$, $\bar{\mathbf{q}}_2$ is orthogonal to \mathbf{q}_1 . We normalize $\bar{\mathbf{q}}_2$ as

$$\mathbf{q}_2 := \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = 0.5\bar{\mathbf{q}}_2 = -0.25\mathbf{a}_1 + 0.5\mathbf{a}_2 = [\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_3] \begin{bmatrix} -0.25 \\ 0.5 \\ 0 \end{bmatrix}$$

Finally, we define

$$\bar{\mathbf{q}}_3 := \mathbf{a}_3 - \mathbf{a}_1 = \begin{bmatrix} 4 \\ 0 \\ -3 \end{bmatrix}$$

We can readily verify that the three vectors \mathbf{q}_1 , \mathbf{q}_2 , and $\bar{\mathbf{q}}_3$ are mutually orthogonal. For example, we have $\mathbf{q}_1' \bar{\mathbf{q}}_3 = 0.6 \times 4 + 0 \times 0 + 0.8 \times (-3) = 0$. The vector $\bar{\mathbf{q}}_3$ has norm $\sqrt{16+9}=5$ and can be normalized as

$$\mathbf{q}_3 := \bar{\mathbf{q}}_3/5 = \begin{bmatrix} 0.8 \\ 0 \\ -0.6 \end{bmatrix} = -0.2\mathbf{a}_1 + 0.2\mathbf{a}_3 = [\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_3] \begin{bmatrix} -0.2 \\ 0 \\ 0.2 \end{bmatrix}$$

The set of three vectors \mathbf{q}_i , for $i = 1, 2, 3$, forms an orthonormal bases. They can be expressed as

$$\begin{aligned} \mathbf{Q} := [\mathbf{q}_1 \ \mathbf{q}_2 \ \mathbf{q}_3] &= \begin{bmatrix} 0.6 & 0 & 0.8 \\ 0 & 1 & 0 \\ 0.8 & 0 & -0.6 \end{bmatrix} \\ &= [\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_3] \begin{bmatrix} 0.1 & -0.25 & -0.2 \\ 0 & -0.5 & 0 \\ 0 & 0 & 0.2 \end{bmatrix} \end{aligned}$$

This equation is intimately related to the QR decomposition which will be used in Chapter 7.

A real *square* matrix \mathbf{A} is called an *orthogonal matrix* if its columns consist of an orthonormal basis or $\mathbf{A}'\mathbf{A} = \mathbf{I}$. It turns out that such a matrix also has the property $\mathbf{A}\mathbf{A}' = \mathbf{I}$. This can be verified for the orthogonal matrix \mathbf{Q} in the preceding example and will be established later. Note that if \mathbf{A} is not square such as 3×2 , then $\mathbf{A}'\mathbf{A} = \mathbf{I}_2$ but $\mathbf{A}\mathbf{A}' \neq \mathbf{I}_3$.

3.3 Linear Algebraic Equations

Consider the set of linear algebraic equations

$$\mathbf{Ax} = \mathbf{y} \tag{3.9}$$

where \mathbf{A} and \mathbf{y} are, respectively, $m \times n$ and $m \times 1$ real matrices and \mathbf{x} is an $n \times 1$ vector. The matrices \mathbf{A} and \mathbf{y} are given, and \mathbf{x} is the unknown to be solved. Thus the set actually consists of m equations and n unknowns. The number of equations can be larger than, equal to, or smaller than the number of unknowns.

We discuss the existence condition and general form of solutions of (3.9). The *range space* of \mathbf{A} is defined as all possible linear combinations of all columns of \mathbf{A} . The *rank* of \mathbf{A} is defined as the dimension of the range space or, equivalently, the number of linearly independent columns in \mathbf{A} . A vector \mathbf{x} is called a *null vector* of \mathbf{A} if $\mathbf{Ax} = \mathbf{0}$. The

null space of \mathbf{A} consists of all its null vectors. The *nullity* is defined as the maximum number of linearly independent null vectors of \mathbf{A} and is related to the rank by

$$\text{Nullity}(\mathbf{A}) = \text{No. of columns of } \mathbf{A} - \text{rank}(\mathbf{A}) \quad (3.10)$$

Example 3.3.1 Consider the matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 2 \\ 1 & 2 & 3 & 4 \\ 2 & 0 & 2 & 0 \end{bmatrix} =: [\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_3 \ \mathbf{a}_4] \quad (3.11)$$

where \mathbf{a}_i denotes the i th column of \mathbf{A} . Clearly, \mathbf{a}_1 and \mathbf{a}_2 are linearly independent. The third column is the sum of the first two columns or $\mathbf{a}_1 + \mathbf{a}_2 - \mathbf{a}_3 = \mathbf{0}$. The last column is twice of the second column, or $2\mathbf{a}_2 - \mathbf{a}_4 = \mathbf{0}$. Thus \mathbf{A} has two linearly independent columns and has rank 2. The set $\{\mathbf{a}_1, \mathbf{a}_2\}$ can be used as a basis of the range space of \mathbf{A} .

Equation (3.10) implies that the nullity of \mathbf{A} is 2. It can be readily verified that the two vectors

$$\mathbf{n}_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \quad \mathbf{n}_2 = \begin{bmatrix} 0 \\ 2 \\ 0 \\ -1 \end{bmatrix} \quad (3.12)$$

meet $\mathbf{A}\mathbf{n}_i = \mathbf{0}$. Because the two vectors are linearly independent, they form a basis of the null space.

The rank of \mathbf{A} is defined as the number of linearly independent columns. It also equals the number of linearly independent rows. Because of this fact, if \mathbf{A} is $m \times n$, then

$$\text{rank}(\mathbf{A}) \leq \min(m, n)$$

In MATLAB, the range space, null space and rank can be obtained by calling the functions `orth`, `null` and `rank`. For example, for the matrix in (3.11), we type

```
a=[0 1 1 2;1 2 3 4;2 0 2 0];
rank(a)
```

which yields 2. Note that MATLAB computes ranks by using singular value decomposition (svd) which will be introduced later. The svd algorithm also yields the range and null spaces of the matrix. The MATLAB function `R=orth(a)` yields

Ans R=

$$\begin{array}{cc} -0.3782 & 0.3084 \\ -0.8877 & 0.1468 \\ -0.2627 & -0.9379 \end{array} \quad (3.13)$$

The two columns of \mathbf{R} is a set of orthonormal basis of the range space. To check the orthonormality, we type $\mathbf{R}' * \mathbf{R}$ which yields the unity matrix of order 2. The two columns in \mathbf{R} are not obtained from the basis $\{\mathbf{a}_1, \mathbf{a}_2\}$ in (3.11) by using the Schmidt

orthonormalization procedure; they are a byproduct of svd. However, the two sets of basis should span the same range space. This can be verified by typing

```
rank([a1 a2 R])
```

which yields 2. This confirms that $\{\mathbf{a}_1, \mathbf{a}_2\}$ span the same space as the two vectors of \mathbf{R} . We mention that the rank of a matrix is very sensitive to roundoff errors and imprecise data. For example, if we use the 5-digit display of \mathbf{R} in (3.13), the rank of $[\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{R}]$ is 3. The rank is 2 if we use the \mathbf{R} stored in MATLAB which uses 16 digits plus exponent.

The null space of (3.11) can be obtained by typing `null(a)`, which yields

Ans N=

$$\begin{array}{cc} -0.6194 & -0.2662 \\ 0.2293 & -0.8750 \\ -0.6194 & 0.2662 \\ -0.4249 & 0.3044 \end{array} \quad (3.14)$$

The two columns are a set of orthonormal basis of the null space spanned by the two vectors $\{\mathbf{n}_1, \mathbf{n}_2\}$ in (3.12). All discussion for the range space applies here. That is, `rank([n1 n2 N])` yields 3 if we use the five-digit display in (3.14). The rank is 2 if we use the \mathbf{N} stored in MATLAB.

With this background, we are ready to discuss solutions of (3.9). We use ρ to denote the rank of a matrix.

► Theorem 3.1

- Given an $m \times n$ matrix \mathbf{A} and an $m \times 1$ vector \mathbf{y} , an $n \times 1$ solution \mathbf{x} exists in $\mathbf{Ax} = \mathbf{y}$ if and only if \mathbf{y} lies in the range space of \mathbf{A} or, equivalently,

$$\rho(\mathbf{A}) = \rho([\mathbf{A} \ \mathbf{y}])$$

where $[\mathbf{A} \ \mathbf{y}]$ is an $m \times (n+1)$ matrix with \mathbf{y} appended to \mathbf{A} as an additional column.

- Given \mathbf{A} , a solution \mathbf{x} exists in $\mathbf{Ax} = \mathbf{y}$ for every \mathbf{y} , if and only if \mathbf{A} has rank m (full row rank).

The first statement follows directly from the definition of the range space. If \mathbf{A} has full row rank, then the rank condition in (1) is always satisfied for every \mathbf{y} . This establishes the second statement.

► Theorem 3.2 (Parameterization of all solutions)

Given an $m \times n$ matrix \mathbf{A} and an $m \times 1$ vector \mathbf{y} . Let \mathbf{x}_p be a solution of $\mathbf{Ax} = \mathbf{y}$ and let $k := n - \rho(\mathbf{A})$ be the nullity of \mathbf{A} . If \mathbf{A} has rank n (full column rank) or $k = 0$, then the solution \mathbf{x}_p is unique. If $k > 0$, then for every real α_i , $i = 1, 2, \dots, k$, the vector

$$\mathbf{x} = \mathbf{x}_p + \alpha_1 \mathbf{n}_1 + \cdots + \alpha_k \mathbf{n}_k \quad (3.15)$$

is a solution of $\mathbf{Ax} = \mathbf{y}$, where $\{\mathbf{n}_1, \dots, \mathbf{n}_k\}$ is a basis of the null space of \mathbf{A} .

Substituting (3.15) into $\mathbf{Ax} = \mathbf{y}$ yields

$$\mathbf{Ax}_p + \sum_{i=1}^k \alpha_i \mathbf{An}_i = \mathbf{Ax}_p + \mathbf{0} = \mathbf{y}$$

Thus, for every α_i , (3.15) is a solution. Let $\bar{\mathbf{x}}$ be a solution of $\mathbf{A}\bar{\mathbf{x}} = \mathbf{y}$. Subtracting this from $\mathbf{Ax}_p = \mathbf{y}$ yields

$$\mathbf{A}(\bar{\mathbf{x}} - \mathbf{x}_p) = \mathbf{0}$$

which implies that $\bar{\mathbf{x}} - \mathbf{x}_p$ is in the null space. Thus $\bar{\mathbf{x}}$ can be expressed as in (3.15). This establishes Theorem 3.2.

Example 3.3.2 Consider the equation

$$\mathbf{Ax} = \begin{bmatrix} 0 & 1 & 1 & 2 \\ 1 & 2 & 3 & 4 \\ 2 & 0 & 2 & 0 \end{bmatrix} \mathbf{x} =: [\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_3 \ \mathbf{a}_4] \mathbf{x} = \begin{bmatrix} -4 \\ -8 \\ 0 \end{bmatrix} = \mathbf{y} \quad (3.16)$$

This \mathbf{y} clearly lies in the range space of \mathbf{A} and $\mathbf{x}_p = [0 \ -4 \ 0 \ 0]'$ is a solution. A basis of the null space of \mathbf{A} was shown in (3.12). Thus the general solution of (3.16) can be expressed as

$$\mathbf{x} = \mathbf{x}_p + \alpha_1 \mathbf{n}_1 + \alpha_2 \mathbf{n}_2 = \begin{bmatrix} 0 \\ -4 \\ 0 \\ 0 \end{bmatrix} + \alpha_1 \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 \\ 2 \\ 0 \\ -1 \end{bmatrix} \quad (3.17)$$

for any real α_1 and α_2 .

In application, we will also encounter $\mathbf{x}\mathbf{A} = \mathbf{y}$, where the $m \times n$ matrix \mathbf{A} and the $1 \times n$ vector \mathbf{y} are given, and the $1 \times m$ vector \mathbf{x} is to be solved. Applying Theorems 3.1 and 3.2 to the transpose of the equation, we can readily obtain the following result.

► Corollary 3.2

1. Given an $m \times n$ \mathbf{A} , a solution \mathbf{x} exists in $\mathbf{x}\mathbf{A} = \mathbf{y}$, for any \mathbf{y} , if and only if \mathbf{A} has full column rank.
2. Given an $m \times n$ matrix \mathbf{A} and an $1 \times n$ vector \mathbf{y} . Let \mathbf{x}_p be a solution of $\mathbf{x}\mathbf{A} = \mathbf{y}$ and let $k = m - \rho(\mathbf{A})$. If $k = 0$, the solution \mathbf{x}_p is unique. If $k > 0$, then for any α_i , $i = 1, 2, \dots, k$, the vector

$$\mathbf{x} = \mathbf{x}_p + \alpha_1 \mathbf{n}_1 + \cdots + \alpha_k \mathbf{n}_k$$

is a solution of $\mathbf{x}\mathbf{A} = \mathbf{y}$, where $\mathbf{n}_i \mathbf{A} = \mathbf{0}$ and the set $\{\mathbf{n}_1, \dots, \mathbf{n}_k\}$ is linearly independent.

In MATLAB, the solution of $\mathbf{Ax} = \mathbf{y}$ can be obtained by typing $\mathbf{A}\backslash\mathbf{y}$. Note the use of backslash which denotes matrix left division. For example, for the equation in

(3.16), typing

```
a=[0 1 1 2;1 2 3 4;2 0 2 0];y=[-4;-8;0];
a\y
```

yields $[0 \ 0 \ 0 \ -2]'$. The solution of $\mathbf{x}\mathbf{A} = \mathbf{y}$ can be obtained by typing \mathbf{y}/\mathbf{A} . Here we use slash which denotes matrix right division.

Determinant and inverse of square matrices

The rank of a matrix is defined as the number of linearly independent columns or rows. It can also be defined using the determinant. The determinant of a 1×1 matrix is defined as itself. For $n = 2, 3, \dots$, the determinant of $n \times n$ square matrix $\mathbf{A} = [a_{ij}]$ is defined recursively as, for any chosen j ,

$$\det \mathbf{A} = \sum_i^n a_{ij} c_{ij} \quad (3.18)$$

where a_{ij} denotes the entry at the i th row and j th column of \mathbf{A} . Equation (3.18) is called the *Laplace expansion*. The number c_{ij} is the *cofactor* corresponding to a_{ij} and equals $(-1)^{i+j} \det M_{ij}$, where M_{ij} is the $(n-1) \times (n-1)$ submatrix of \mathbf{A} by deleting its i th row and j th column. If \mathbf{A} is diagonal or triangular, then $\det \mathbf{A}$ equals the product of all diagonal entries.

The determinant of any $r \times r$ submatrix of \mathbf{A} is called a *minor* of order r . Then the rank can be defined as the largest order of all nonzero minors of \mathbf{A} . In other words, if \mathbf{A} has rank r , then there is at least one nonzero minor of order r , and every minor of order larger than r is zero. This definition applies to \mathbf{A} square or not. If \mathbf{A} is square of order n , and if its determinant is nonzero, then \mathbf{A} has rank n (full rank) and is said to be *nonsingular*. If the determinant is zero, \mathbf{A} is said to be *singular* and has a rank less than n . For a nonsingular square matrix, all columns (rows) are linearly independent.

The *inverse* of a nonsingular square matrix $\mathbf{A} = [a_{ij}]$ is denoted by \mathbf{A}^{-1} . The inverse has the property $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ and can be computed as

$$\mathbf{A}^{-1} = \frac{\text{Adj } \mathbf{A}}{\det \mathbf{A}} = \frac{1}{\det \mathbf{A}} [c_{ij}]' \quad (3.19)$$

where $[c_{ij}]'$ is called the *adjoint* and is denoted by $\text{Adj } \mathbf{A}$. If \mathbf{A} is 2×2 , the adjoint can be easily computed and (3.19) reduces to

$$\mathbf{A}^{-1} := \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \quad (3.20)$$

Thus the inverse of a 2×2 matrix is very simple: interchanging diagonal entries and changing the sign of off diagonal entries (without changing position) and dividing the resulting matrix by the determinant of \mathbf{A} . In general, using (3.19) to compute the inverse is complicated. If \mathbf{A} is triangular, it is simpler to compute its inverse by solving $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. Note that the inverse of a triangular matrix is again triangular. The MATLAB function `inv` computes the inverse of \mathbf{A} . Note that an orthogonal matrix \mathbf{A} has a full rank and its inverse exists. Postmultiplying \mathbf{A}^{-1} to $\mathbf{A}'\mathbf{A} = \mathbf{I}$ yields $\mathbf{A}'\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}\mathbf{A}^{-1}$, which implies $\mathbf{A}' = \mathbf{A}^{-1}$. Consequently we have $\mathbf{A}\mathbf{A}' = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. Thus an orthogonal

matrix \mathbf{A} has the properties $\mathbf{A}'\mathbf{A} = \mathbf{AA}' = \mathbf{I}$ and $\mathbf{A}^{-1} = \mathbf{A}'$. We next give a special case of Theorem 3.2 where $n = m$.

Theorem 3.3

Consider $\mathbf{Ax} = \mathbf{y}$, where \mathbf{A} is a square matrix of order n , and \mathbf{x} and \mathbf{y} are $n \times 1$ vectors.

1. If \mathbf{A} is nonsingular or $\rho(\mathbf{A}) = n$, then the equation has a unique solution for every \mathbf{y} and the solution equals $\mathbf{A}^{-1}\mathbf{y}$. In particular, the only solution of $\mathbf{Ax} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$.
2. The homogeneous equation $\mathbf{Ax} = \mathbf{0}$ has nonzero solutions if and only if \mathbf{A} is singular or $\rho(\mathbf{A}) < n$. The number of linearly independent solutions equals $n - \rho(\mathbf{A})$, the nullity of \mathbf{A} .

This is an important theorem and will be used repeatedly in this text.

3.4 Similarity Transformation

Consider an $n \times n$ matrix \mathbf{A} . It maps \mathcal{R}^n into itself. If we associate with \mathcal{R}^n the orthonormal basis $\{\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_n\}$ in (3.8), then the i th column of \mathbf{A} is the representation of \mathbf{Ai}_i with respect to the orthonormal basis. Now if we select a different set of basis $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$, then the matrix \mathbf{A} has a different representation $\bar{\mathbf{A}}$. This is first illustrated by the example that follows the i th column of $\bar{\mathbf{A}}$, as will be verified shortly, is the representation of \mathbf{Aq}_i with respect to the basis $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$.

Example 3.4.1 Consider the matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & -1 \\ -2 & 1 & 0 \\ 4 & 3 & 1 \end{bmatrix} \quad (3.21)$$

Let $\mathbf{b} = [0 \ 0 \ 1]'$. Then we have

$$\mathbf{Ab} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{A}^2\mathbf{b} = \mathbf{A}(\mathbf{Ab}) = \begin{bmatrix} -4 \\ 2 \\ -3 \end{bmatrix}, \quad \mathbf{A}^3\mathbf{b} = \mathbf{A}(\mathbf{A}^2\mathbf{b}) = \begin{bmatrix} -5 \\ 10 \\ -13 \end{bmatrix}$$

It can be verified that the following relation holds:

$$\mathbf{A}^3\mathbf{b} = 17\mathbf{b} - 15\mathbf{Ab} + 5\mathbf{A}^2\mathbf{b} \quad (3.22)$$

Because the three vectors \mathbf{b} , \mathbf{Ab} , and $\mathbf{A}^2\mathbf{b}$ are linearly independent, they can be used as a basis. We now compute the representation of \mathbf{A} with respect to the basis. It is clear

that

$$\begin{aligned}\mathbf{A}(\mathbf{b}) &= [\mathbf{b} \ \mathbf{Ab} \ \mathbf{A}^2\mathbf{b}] \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \\ \mathbf{A}(\mathbf{Ab}) &= [\mathbf{b} \ \mathbf{Ab} \ \mathbf{A}^2\mathbf{b}] \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\ \mathbf{A}(\mathbf{A}^2\mathbf{b}) &= [\mathbf{b} \ \mathbf{Ab} \ \mathbf{A}^2\mathbf{b}] \begin{bmatrix} 17 \\ -15 \\ 5 \end{bmatrix}\end{aligned}$$

where the last equation is obtained from (3.22). Thus the representation of \mathbf{A} with respect to the basis $\{\mathbf{b}, \mathbf{Ab}, \mathbf{A}^2\mathbf{b}\}$ is

$$\bar{\mathbf{A}} = \begin{bmatrix} 0 & 0 & 17 \\ 1 & 0 & -15 \\ 0 & 1 & 5 \end{bmatrix} \quad (3.23)$$

The preceding discussion can be extended to the general case. Let \mathbf{A} be an $n \times n$ matrix. If there exists an $n \times 1$ vector \mathbf{b} such that the n vectors $\mathbf{b}, \mathbf{Ab}, \dots, \mathbf{A}^{n-1}\mathbf{b}$ are linearly independent and if

$$\mathbf{A}^n\mathbf{b} = \beta_1\mathbf{b} + \beta_2\mathbf{Ab} + \dots + \beta_n\mathbf{A}^{n-1}\mathbf{b}$$

then the representation of \mathbf{A} with respect to the basis $\{\mathbf{b}, \mathbf{Ab}, \dots, \mathbf{A}^{n-1}\mathbf{b}\}$ is

$$\bar{\mathbf{A}} = \begin{bmatrix} 0 & 0 & \cdots & 0 & \beta_1 \\ 1 & 0 & \cdots & 0 & \beta_2 \\ 0 & 1 & \cdots & 0 & \beta_3 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \beta_{n-1} \\ 0 & 0 & \cdots & 1 & \beta_n \end{bmatrix} \quad (3.24)$$

This matrix is said to be in a *companion* form.

We next develop an equation to relate \mathbf{A} and $\bar{\mathbf{A}}$. Consider the equation

$$\mathbf{Ax} = \mathbf{y} \quad (3.25)$$

The square matrix \mathbf{A} maps \mathbf{x} in \mathcal{R}^n into \mathbf{y} in \mathcal{R}^n . With respect to the basis $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$, the equation becomes

$$\bar{\mathbf{A}}\bar{\mathbf{x}} = \bar{\mathbf{y}} \quad (3.26)$$

where $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ are the representations of \mathbf{x} and \mathbf{y} with respect to the basis $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$. As discussed in (3.7), they are related by

$$\mathbf{x} = \mathbf{Q}\bar{\mathbf{x}} \quad \mathbf{y} = \mathbf{Q}\bar{\mathbf{y}}$$

with

$$\mathbf{Q} = [\mathbf{q}_1 \ \mathbf{q}_2 \ \dots \ \mathbf{q}_n] \quad (3.27)$$

an $n \times n$ nonsingular matrix. Substituting these into (3.25) yields

$$\mathbf{A}\mathbf{Q}\bar{\mathbf{x}} = \mathbf{Q}\bar{\mathbf{y}} \quad \text{or} \quad \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}\bar{\mathbf{x}} = \bar{\mathbf{y}} \quad (3.28)$$

Comparing this with (3.26) yields

$$\bar{\mathbf{A}} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} \quad \text{or} \quad \mathbf{A} = \mathbf{Q}\bar{\mathbf{A}}\mathbf{Q}^{-1} \quad (3.29)$$

This is called the *similarity transformation* and \mathbf{A} and $\bar{\mathbf{A}}$ are said to be *similar*. We write (3.29) as

$$\mathbf{A}\mathbf{Q} = \mathbf{Q}\bar{\mathbf{A}}$$

or

$$\mathbf{A}[\mathbf{q}_1 \ \mathbf{q}_2 \ \dots \ \mathbf{q}_n] = [\mathbf{A}\mathbf{q}_1 \ \mathbf{A}\mathbf{q}_2 \ \dots \ \mathbf{A}\mathbf{q}_n] = [\mathbf{q}_1 \ \mathbf{q}_2 \ \dots \ \mathbf{q}_n]\bar{\mathbf{A}}$$

This shows that the i th column of $\bar{\mathbf{A}}$ is indeed the representation of $\mathbf{A}\mathbf{q}_i$ with respect to the basis $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$.

3.5 Diagonal Form and Jordan Form

A square matrix \mathbf{A} has different representations with respect to different sets of basis. In this section, we introduce a set of basis so that the representation will be diagonal or block diagonal.

A real or complex number λ is called an *eigenvalue* of the $n \times n$ real matrix \mathbf{A} if there exists a nonzero vector \mathbf{x} such that $\mathbf{Ax} = \lambda\mathbf{x}$. Any nonzero vector \mathbf{x} satisfying $\mathbf{Ax} = \lambda\mathbf{x}$ is called a (right) *eigenvector* of \mathbf{A} associated with eigenvalue λ . In order to find the eigenvalue of \mathbf{A} , we write $\mathbf{Ax} = \lambda\mathbf{x} = \lambda\mathbf{Ix}$ as

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0} \quad (3.30)$$

where \mathbf{I} is the unit matrix of order n . This is a homogeneous equation. If the matrix $(\mathbf{A} - \lambda\mathbf{I})$ is nonsingular, then the only solution of (3.30) is $\mathbf{x} = \mathbf{0}$ (Theorem 3.3). Thus in order for (3.30) to have a nonzero solution \mathbf{x} , the matrix $(\mathbf{A} - \lambda\mathbf{I})$ must be singular or has determinant 0. We define

$$\Delta(\lambda) = \det(\lambda\mathbf{I} - \mathbf{A})$$

It is a polynomial of degree n with real coefficients and is called the *characteristic polynomial* of \mathbf{A} . If λ is a root of the polynomial, then the determinant of $(\mathbf{A} - \lambda\mathbf{I})$ is 0 and (3.30) has at least one nonzero solution. Thus every root of $\Delta(\lambda)$ is an eigenvalue

of \mathbf{A} . Because $\Delta(\lambda)$ has degree n , the $n \times n$ matrix \mathbf{A} has n eigenvalues (not necessarily all distinct).

We mention that the matrices

$$\begin{bmatrix} 0 & 0 & 0 & -\alpha_4 \\ 1 & 0 & 0 & -\alpha_3 \\ 0 & 1 & 0 & -\alpha_2 \\ 0 & 0 & 1 & -\alpha_1 \end{bmatrix} \quad \begin{bmatrix} -\alpha_1 & -\alpha_2 & -\alpha_3 & -\alpha_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

and their transposes

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\alpha_4 & -\alpha_3 & -\alpha_2 & -\alpha_1 \end{bmatrix} \quad \begin{bmatrix} -\alpha_1 & 1 & 0 & 0 \\ -\alpha_2 & 0 & 1 & 0 \\ -\alpha_3 & 0 & 0 & 1 \\ -\alpha_4 & 0 & 0 & 0 \end{bmatrix}$$

all have the following characteristic polynomial:

$$\Delta(\lambda) = \lambda^4 + \alpha_1\lambda^3 + \alpha_2\lambda^2 + \alpha_3\lambda + \alpha_4$$

These matrices can be easily formed from the coefficients of $\Delta(\lambda)$ and are called *companion-form* matrices. The companion-form matrices will arise repeatedly later. The matrix in (3.24) is in such a form.

Eigenvalues of \mathbf{A} are all distinct

Let λ_i , $i = 1, 2, \dots, n$, be the eigenvalues of \mathbf{A} , and let all be distinct. Let \mathbf{q}_i be an eigenvector of \mathbf{A} associated with λ_i , that is, $\mathbf{A}\mathbf{q}_i = \lambda_i\mathbf{q}_i$. Then the set of eigenvectors $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$ is linearly independent (see Reference 6, pp. 34–35) and can be used as a basis. Let $\hat{\mathbf{A}}$ be the representation of \mathbf{A} with respect to this basis. Then the first column of $\hat{\mathbf{A}}$ is the representation of $\mathbf{A}\mathbf{q}_1 = \lambda_1\mathbf{q}_1$ with respect to $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$. From

$$\mathbf{A}\mathbf{q}_1 = \lambda_1\mathbf{q}_1 = [\mathbf{q}_1 \ \mathbf{q}_2 \ \cdots \ \mathbf{q}_n] \begin{bmatrix} \lambda_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

we conclude that the first column of $\hat{\mathbf{A}}$ is $[\lambda_1 \ 0 \ \cdots \ 0]'$. The second column of $\hat{\mathbf{A}}$ is the representation of $\mathbf{A}\mathbf{q}_2 = \lambda_2\mathbf{q}_2$ with respect to $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$, that is, $[0 \ \lambda_2 \ 0 \ \cdots \ 0]'$. Proceeding forward, we can establish

$$\hat{\mathbf{A}} = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} \quad (3.31)$$

This is a diagonal matrix. Thus we conclude that every matrix with distinct eigenvalues has a diagonal matrix representation by using its eigenvectors as a basis. Different orderings of eigenvectors will yield different diagonal matrices for the same \mathbf{A} .

If we define

$$\mathbf{Q} = [\mathbf{q}_1 \ \mathbf{q}_2 \ \cdots \ \mathbf{q}_n] \quad (3.32)$$

then the matrix $\hat{\mathbf{A}}$ equals

$$\hat{\mathbf{A}} = \mathbf{Q}^{-1} \mathbf{A} \mathbf{Q} \quad (3.33)$$

as derived in (3.29). Computing (3.33) is not simple by hand because the need of computing the inverse of \mathbf{Q} . However, if we know $\hat{\mathbf{A}}$, then we can verify (3.33) by checking $\mathbf{Q}\hat{\mathbf{A}} = \mathbf{A}\mathbf{Q}$.

Example 3.5.1 Consider the matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix}$$

Its characteristic polynomial is

$$\begin{aligned} \Delta(\lambda) &= \det(\lambda\mathbf{I} - \mathbf{A}) = \det \begin{bmatrix} \lambda & 0 & 0 \\ -1 & \lambda & -2 \\ 0 & -1 & \lambda - 1 \end{bmatrix} \\ &= \lambda[\lambda(\lambda - 1) - 2] = (\lambda - 2)(\lambda + 1)\lambda \end{aligned}$$

Thus \mathbf{A} has eigenvalues 2, -1, and 0. The eigenvector associated with $\lambda = 2$ is any nonzero solution of

$$(\mathbf{A} - 2\mathbf{I})\mathbf{q}_1 = \begin{bmatrix} -2 & 0 & 0 \\ 1 & -2 & 2 \\ 0 & 1 & -1 \end{bmatrix} \mathbf{q}_1 = \mathbf{0}$$

Thus $\mathbf{q}_1 = [0 \ 1 \ 1]'$ is an eigenvector associated with $\lambda = 2$. Note that the eigenvector is not unique; $[0 \ \alpha \ \alpha]'$ for any nonzero real α can also be chosen as an eigenvector. The eigenvector associated with $\lambda = -1$ is any nonzero solution of

$$(\mathbf{A} - (-1)\mathbf{I})\mathbf{q}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 2 \\ 0 & 1 & 2 \end{bmatrix} \mathbf{q}_2 = \mathbf{0}$$

which yields $\mathbf{q}_2 = [0 \ -2 \ 1]'$. Similarly, the eigenvector associated with $\lambda = 0$ can be computed as $\mathbf{q}_3 = [2 \ 1 \ -1]'$. Thus, the representation of \mathbf{A} with respect to $\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$ is

$$\hat{\mathbf{A}} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (3.34)$$

It is a diagonal matrix with eigenvalues on the diagonal. This matrix can also be obtained by computing

$$\hat{\mathbf{A}} = \mathbf{Q}^{-1} \mathbf{A} \mathbf{Q}$$

with

$$\mathbf{Q} = [\mathbf{q}_1 \ \mathbf{q}_2 \ \mathbf{q}_3] = \begin{bmatrix} 0 & 0 & 2 \\ 1 & -2 & 1 \\ 1 & 1 & -1 \end{bmatrix} \quad (3.35)$$

However, it is simpler to verify $\mathbf{Q}\hat{\mathbf{A}} = \mathbf{A}\mathbf{Q}$ or

$$\begin{bmatrix} 0 & 0 & 2 \\ 1 & -2 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 2 \\ 1 & -2 & 1 \\ 1 & 1 & -1 \end{bmatrix}$$

The result in this example can be easily obtained using MATLAB. Typing

```
a=[0 0 0;1 0 2;0 1 1]; [q,d]=eig(a)
```

yields

$$q = \begin{bmatrix} 0 & 0 & 0.8186 \\ 0.7071 & 0.8944 & 0.4082 \\ 0.7071 & -0.4472 & -0.4082 \end{bmatrix}, \quad d = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

where d is the diagonal matrix in (3.34). The matrix q is different from the \mathbf{Q} in (3.35); but their corresponding columns differ only by a constant. This is due to nonuniqueness of eigenvectors and every column of q is normalized to have norm 1 in MATLAB. If we type `eig(a)` without the left-hand-side argument, then MATLAB generates only the three eigenvalues 2, -1, 0. We mention that eigenvalues in MATLAB are *not* computed from the characteristic polynomial. Computing the characteristic polynomial using the Laplace expansion and then computing its roots are not numerically reliable, especially when there are repeated roots. Eigenvalues are computed in MATLAB directly from the matrix by using similarity transformations. Once all eigenvalues are computed, the characteristic polynomial equals $\prod(\lambda - \lambda_i)$. In MATLAB, typing `r=eig(a); poly(r)` yields the characteristic polynomial.

Example 3.5.2 Consider the matrix

$$\mathbf{A} = \begin{bmatrix} -1 & 1 & 1 \\ 0 & 4 & -13 \\ 0 & 1 & 0 \end{bmatrix}$$

Its characteristic polynomial is $(\lambda + 1)(\lambda^2 - 4\lambda + 13)$. Thus \mathbf{A} has eigenvalues $-1, 2 \pm 3j$. Note that complex conjugate eigenvalues must appear in pairs because \mathbf{A} has only real coefficients. The eigenvectors associated with -1 and $2 + 3j$ are, respectively, $[1 \ 0 \ 0]'$ and $[j \ -3 + 2j \ j]'$. The eigenvector associated with $\lambda = 2 - 3j$ is $[-j \ -3 - 2j \ -j]'$, the complex conjugate of the eigenvector associated with $\lambda = 2 + 3j$. Thus we have

$$\mathbf{Q} = \begin{bmatrix} 1 & j & -j \\ 0 & -3+2j & -3-2j \\ 0 & j & j \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{A}} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2+3j & 0 \\ 0 & 0 & 2-3j \end{bmatrix} \quad (3.36)$$

The MATLAB function `[q, d]=eig(a)` yields

$$q = \begin{bmatrix} 1 & 0.1432 - 0.2148j & 0.1432 + 0.2148j \\ 0 & 0.9309 & 0.9309 \\ 0 & 0.1432 - 0.2148j & 0.1432 + 0.2148j \end{bmatrix}$$

and

$$d = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2+3j & 0 \\ 0 & 0 & 2-3j \end{bmatrix}$$

All discussion in the preceding example applies here.

Complex eigenvalues

Even though the data we encounter in practice are all real numbers, complex numbers may arise when we compute eigenvalues and eigenvectors. To deal with this problem, we must extend real linear spaces into complex linear spaces and permit all scalars such as α_i in (3.4) to assume complex numbers. To see the reason, we consider

$$\mathbf{Av} = \begin{bmatrix} 1 & 1+j \\ 1-j & 2 \end{bmatrix} \mathbf{v} = \mathbf{0} \quad (3.37)$$

If we restrict \mathbf{v} to real vectors, then (3.37) has no nonzero solution and the two columns of \mathbf{A} are linearly independent. However, if \mathbf{v} is permitted to assume complex numbers, then $\mathbf{v} = [-2 \ 1-j]'$ is a nonzero solution of (3.37). Thus the two columns of \mathbf{A} are linearly dependent and \mathbf{A} has rank 1. This is the rank obtained in MATLAB. Therefore, whenever complex eigenvalues arise, we consider complex linear spaces and complex scalars and transpose is replaced by complex-conjugate transpose. By so doing, all concepts and results developed for real vectors and matrices can be applied to complex vectors and matrices. Incidentally, the diagonal matrix with complex eigenvalues in (3.36) can be transformed into a useful real matrix, as we will discuss in Section 4.3.1.

Eigenvalues of \mathbf{A} are not all distinct

An eigenvalue with multiplicity 2 or higher is called a *repeated* eigenvalue. In contrast, an eigenvalue with multiplicity 1 is called a *simple* eigenvalue. If \mathbf{A} has only simple eigenvalues, it always has a diagonal-form representation. If \mathbf{A} has repeated eigenvalues, then it may not have such a representation. However, it has a Jordan-form representation as we discuss next.

We first discuss a special matrix. Consider

$$\mathbf{J} := \begin{bmatrix} \lambda_1 & 1 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 \\ 0 & 0 & \lambda_1 & 1 \\ 0 & 0 & 0 & \lambda_1 \end{bmatrix} \quad (3.38)$$

Its characteristic polynomial is

$$\Delta(\lambda) = \det(\lambda\mathbf{I} - \mathbf{J}) = \det \begin{bmatrix} (\lambda - \lambda_1) & -1 & 0 & 0 \\ 0 & (\lambda - \lambda_1) & -1 & 0 \\ 0 & 0 & (\lambda - \lambda_1) & -1 \\ 0 & 0 & 0 & (\lambda - \lambda_1) \end{bmatrix} = (\lambda - \lambda_1)^4$$

Thus the matrix \mathbf{J} has eigenvalue λ_1 with multiplicity 4. Such a matrix which has its eigenvalue on the diagonal and 1 on the superdiagonal is called a *Jordan block* of order 4. We can have Jordan blocks of various order. A Jordan block of order 1 is simply its eigenvalue.

Consider a 5×5 matrix \mathbf{A} with repeated eigenvalue λ_1 with multiplicity 4 and simple eigenvalue λ_2 . Then there exists a nonsingular matrix \mathbf{Q} such that the similarity matrix

$$\hat{\mathbf{A}} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$$

assumes one of the following forms

$$\hat{\mathbf{A}}_1 = \begin{bmatrix} \lambda_1 & 1 & 0 & 0 & \vdots & 0 \\ 0 & \lambda_1 & 1 & 0 & \vdots & 0 \\ 0 & 0 & \lambda_1 & 1 & \vdots & 0 \\ 0 & 0 & 0 & \lambda_1 & \vdots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \vdots & \lambda_2 \end{bmatrix} \quad \hat{\mathbf{A}}_2 = \begin{bmatrix} \lambda_1 & 1 & 0 & \vdots & 0 & 0 \\ 0 & \lambda_1 & 1 & \vdots & 0 & 0 \\ 0 & 0 & \lambda_1 & \vdots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \vdots & \lambda_1 & \vdots & 0 \\ & & & & \dots & \dots & \dots \\ 0 & 0 & 0 & & 0 & \vdots & \lambda_2 \end{bmatrix}$$

$$\hat{\mathbf{A}}_3 = \begin{bmatrix} \lambda_1 & 1 & \vdots & 0 & 0 & 0 \\ 0 & \lambda_1 & \vdots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \vdots & \lambda_1 & 1 & \vdots & 0 \\ 0 & 0 & \vdots & 0 & \lambda_1 & \vdots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & & 0 & 0 & \vdots & \lambda_2 \end{bmatrix} \quad \hat{\mathbf{A}}_4 = \begin{bmatrix} \lambda_1 & 1 & \vdots & 0 & 0 & 0 \\ 0 & \lambda_1 & \vdots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \vdots & \lambda_1 & 1 & \vdots & 0 \\ 0 & 0 & \vdots & 0 & \lambda_1 & \vdots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & & 0 & 0 & \vdots & \lambda_2 \end{bmatrix}$$

$$\hat{\mathbf{A}}_5 = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & 0 & \lambda_1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{bmatrix} \quad (3.39)$$

For the simple eigenvalue λ_2 , there is only one Jordan block of order 1 in all preceding matrices. For the repeated eigenvalue λ_1 of multiplicity 4, the first matrix has one Jordan block of order 4. The second matrix has two Jordan blocks of order 3 and 1, respectively. The third matrix has two Jordan blocks both of order 2. The fourth matrix has three Jordan blocks of order 2, 1, and 1 respectively. The last matrix has four Jordan blocks of order 1 associated with λ_1 . These matrices are all said to be of *Jordan form*. Which form a matrix assumes depends on the nullity of the matrix $(\mathbf{A} - \lambda_1 \mathbf{I})$. If the nullity is 4, we can find four linearly independent eigenvectors associated with λ_1 . Using the four eigenvectors and the eigenvector associated with λ_2 as a basis, we can obtain $\hat{\mathbf{A}}_5$ and \mathbf{A} is said to be diagonalizable. If the nullity of $(\mathbf{A} - \lambda_1 \mathbf{I})$ is 1, we can find only one (ordinary) eigenvector associated with λ_1 . In this case, we must find a chain of *generalized* eigenvectors as a basis to yield the form in $\hat{\mathbf{A}}_1$. See Reference 6. If the nullity is 2 or 3, then we will obtain the other forms. Different ordering of the basis will result in different ordering of Jordan blocks; it may also result in a Jordan block with 1 in subdiagonal rather than in superdiagonal. In conclusion, computing Jordan-form matrices is complicated and will not be discussed further. We will simply accept the fact that every matrix can be transformed into a Jordan form using a similarity transformation. Note that the earlier versions of MATLAB contained the function `jordan` but the version R2011a does not. The function `[q, d] = eig(a)` applies to all matrices but will yield a correct result only if a matrix is diagonalizable.

Jordan-form matrices are triangular and block diagonal and can be used to establish many general properties of matrices. For example, because $\det(\mathbf{CD}) = \det \mathbf{C} \det \mathbf{D}$ and $\det \mathbf{Q} \det \mathbf{Q}^{-1} = \det \mathbf{I} = 1$, from $\mathbf{A} = \mathbf{Q} \hat{\mathbf{A}} \mathbf{Q}^{-1}$, we have

$$\det \mathbf{A} = \det \mathbf{Q} \det \hat{\mathbf{A}} \det \mathbf{Q}^{-1} = \det \hat{\mathbf{A}}$$

The determinant of $\hat{\mathbf{A}}$ is the product of all diagonal entries or, equivalently, all eigenvalues of \mathbf{A} . Thus we have

$$\det \mathbf{A} = \text{product of all eigenvalues of } \mathbf{A}$$

which implies that \mathbf{A} is nonsingular if and only if it has no zero eigenvalue.

We discuss a useful property of Jordan blocks to conclude this section. Consider the Jordan block in (3.38) with order 4. Then we have

$$\begin{aligned} (\mathbf{J} - \lambda \mathbf{I}) &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, & (\mathbf{J} - \lambda \mathbf{I})^2 &= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ (\mathbf{J} - \lambda \mathbf{I})^3 &= \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{aligned} \quad (3.40)$$

and $(\mathbf{J} - \lambda \mathbf{I})^k = \mathbf{0}$ for $k \geq 4$. This is called *nilpotent*.

3.6 Functions of a Square Matrix

This section studies functions of a square matrix. We use Jordan form extensively because many properties of functions can almost be visualized in terms of Jordan form. We study first polynomials and then general functions of a square matrix.

Polynomials of a square matrix

Let \mathbf{A} be a square matrix. If k is a positive integer, we define

$$\mathbf{A}^k := \mathbf{A}\mathbf{A} \cdots \mathbf{A} \quad (k \text{ terms})$$

and $\mathbf{A}^0 = \mathbf{I}$. Let $f(\lambda)$ be a polynomial such as $f(\lambda) = \lambda^3 + 2\lambda^2 - 6$ or $(\lambda + 2)(4\lambda - 3)$. Then $f(\mathbf{A})$ is defied as

$$f(\mathbf{A}) = \mathbf{A}^3 + 2\mathbf{A}^2 - 6\mathbf{I} \quad \text{or} \quad f(\mathbf{A}) = (\mathbf{A} + 2\mathbf{I})(4\mathbf{A} - 3\mathbf{I})$$

If \mathbf{A} is block diagonal such as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{bmatrix}$$

where \mathbf{A}_1 and \mathbf{A}_2 are square matrices of any order, then it is straightforward to verify

$$\mathbf{A}^k = \begin{bmatrix} \mathbf{A}_1^k & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2^k \end{bmatrix} \quad \text{and} \quad f(\mathbf{A}) = \begin{bmatrix} f(\mathbf{A}_1) & \mathbf{0} \\ \mathbf{0} & f(\mathbf{A}_2) \end{bmatrix} \quad (3.41)$$

Consider the similarity transformation $\hat{\mathbf{A}} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$ or $\mathbf{A} = \mathbf{Q}\hat{\mathbf{A}}\mathbf{Q}^{-1}$. Because

$$\mathbf{A}^k = (\mathbf{Q}\hat{\mathbf{A}}\mathbf{Q}^{-1})(\mathbf{Q}\hat{\mathbf{A}}\mathbf{Q}^{-1}) \cdots (\mathbf{Q}\hat{\mathbf{A}}\mathbf{Q}^{-1}) = \mathbf{Q}\hat{\mathbf{A}}^k\mathbf{Q}^{-1}$$

we have

$$f(\mathbf{A}) = \mathbf{Q}f(\hat{\mathbf{A}})\mathbf{Q}^{-1} \quad \text{or} \quad f(\hat{\mathbf{A}}) = \mathbf{Q}^{-1}f(\mathbf{A})\mathbf{Q} \quad (3.42)$$

A *monic polynomial* is a polynomial with 1 as its leading coefficient. The *minimal polynomial* of \mathbf{A} is defined as the monic polynomial $\psi(\lambda)$ of least degree such that $\psi(\mathbf{A}) = \mathbf{0}$. Note that the $\mathbf{0}$ is a zero matrix of the same order as \mathbf{A} . A direct consequence of (3.42) is that $f(\mathbf{A}) = \mathbf{0}$ if and only if $f(\hat{\mathbf{A}}) = \mathbf{0}$. Thus, \mathbf{A} and $\hat{\mathbf{A}}$ or, more generally, all similarity matrices have the same minimal polynomial. Computing the minimal polynomial directly from \mathbf{A} is not simple (see Problem 3.30); however, if the Jordan-form representation of \mathbf{A} is available, the minimal polynomial can be read out by inspection.

Let λ_i be an eigenvalue of \mathbf{A} with multiplicity n_i . That is, the characteristic polynomial of \mathbf{A} is

$$\Delta(\lambda) = \det(\lambda\mathbf{I} - \mathbf{A}) = \prod_i (\lambda - \lambda_i)^{n_i}$$

Suppose the Jordan form of \mathbf{A} is known. Associated with each eigenvalue, there may be one or more Jordan blocks. The *index* of λ_i , denoted by \bar{n}_i , is defined as the largest order of all Jordan blocks associated with λ_i . Clearly, we have $\bar{n}_i \leq n_i$. For example, the multiplicities of λ_1 in all five matrices in (3.39) are 4; their indices are, respectively, 4, 3, 2, 2, and 1. The multiplicities and indices of λ_2 in all five matrices in (3.39) are all 1. Using the indices of all eigenvalues, the minimal polynomial can be expressed as

$$\psi(\lambda) = \prod_i (\lambda - \lambda_i)^{\bar{n}_i}$$

with degree $\bar{n} = \sum \bar{n}_i \leq \sum n_i = n$ = dimension of \mathbf{A} . For example, the minimal polynomials of the five matrices in (3.39) are

$$\begin{aligned}\psi_1 &= (\lambda - \lambda_1)^4(\lambda - \lambda_2), & \psi_2 &= (\lambda - \lambda_1)^3(\lambda - \lambda_2) \\ \psi_3 &= (\lambda - \lambda_1)^2(\lambda - \lambda_2), & \psi_4 &= (\lambda - \lambda_1)^2(\lambda - \lambda_2) \\ \psi_5 &= (\lambda - \lambda_1)(\lambda - \lambda_2)\end{aligned}$$

Their characteristic polynomials, however, all equal

$$\Delta(\lambda) = (\lambda - \lambda_1)^4(\lambda - \lambda_2)$$

We see that the minimal polynomial is a factor of the characteristic polynomial, and its degree is less than or equal to the degree of the characteristic polynomial. Clearly, if all eigenvalues of \mathbf{A} are distinct, then the minimal polynomial equals the characteristic polynomial.

Using the nilpotent property in (3.40), we can show that

$$\psi(\mathbf{A}) = \mathbf{0}$$

and that no polynomial of lesser degree meets the condition. Thus, $\psi(\lambda)$ as defined is the minimal polynomial.

► Theorem 3.4 (Cayley–Hamilton theorem)

Let

$$\Delta(\lambda) = \det(\lambda\mathbf{I} - \mathbf{A}) = \lambda^n + \alpha_1\lambda^{n-1} + \cdots + \alpha_{n-1}\lambda + \alpha_n$$

be the characteristic polynomial of \mathbf{A} . Then

$$\Delta(\mathbf{A}) = \mathbf{A}^n + \alpha_1 \mathbf{A}^{n-1} + \cdots + \alpha_{n-1} \mathbf{A} + \alpha_n \mathbf{I} = \mathbf{0} \quad (3.43)$$

In words, a matrix satisfies its own characteristic polynomial. Because $n_i \geq \bar{n}_i$, the characteristic polynomial contains the minimal polynomial as a factor or $\Delta(\lambda) = \psi(\lambda)h(\lambda)$ for some polynomial $h(\lambda)$. Because $\psi(\mathbf{A}) = \mathbf{0}$, we have $\Delta(\mathbf{A}) = \psi(\mathbf{A})h(\mathbf{A}) = \mathbf{0} \cdot h(\mathbf{A}) = \mathbf{0}$. This establishes the theorem. The Cayley–Hamilton theorem implies that \mathbf{A}^n can be written as a linear combination of $\{\mathbf{I}, \mathbf{A}, \dots, \mathbf{A}^{n-1}\}$. Multiplying (3.43) by \mathbf{A} yields

$$\mathbf{A}^{n+1} + \alpha_1 \mathbf{A}^n + \cdots + \alpha_{n-1} \mathbf{A}^2 + \alpha_n \mathbf{A} = \mathbf{0} \cdot \mathbf{A} = \mathbf{0}$$

which implies that \mathbf{A}^{n+1} can be written as a linear combination of $\{\mathbf{A}, \mathbf{A}^2, \dots, \mathbf{A}^n\}$, which, in turn, can be written as a linear combination of $\{\mathbf{I}, \mathbf{A}, \dots, \mathbf{A}^{n-1}\}$. Proceeding forward, we conclude that, for any polynomial $f(\lambda)$, no matter how large its degree is, $f(\mathbf{A})$ can always be expressed as

$$f(\mathbf{A}) = \beta_0 \mathbf{I} + \beta_1 \mathbf{A} + \cdots + \beta_{n-1} \mathbf{A}^{n-1} \quad (3.44)$$

for some β_i . In other words, every polynomial of an $n \times n$ matrix \mathbf{A} can be expressed as a linear combination of $\{\mathbf{I}, \mathbf{A}, \dots, \mathbf{A}^{n-1}\}$. If the minimal polynomial of \mathbf{A} with degree \bar{n} is available, then every polynomial of \mathbf{A} can be expressed as a linear combination of $\{\mathbf{I}, \mathbf{A}, \dots, \mathbf{A}^{\bar{n}-1}\}$. This is a better result. However, because \bar{n} may not be available, we discuss in the following only (3.44) with the understanding that all discussion applies to \bar{n} .

One way to compute (3.44) is to use long division to express $f(\lambda)$ as

$$f(\lambda) = q(\lambda) \Delta(\lambda) + h(\lambda) \quad (3.45)$$

where $q(\lambda)$ is the quotient and $h(\lambda)$ is the remainder with degree less than n . Then we have

$$f(\mathbf{A}) = q(\mathbf{A}) \Delta(\mathbf{A}) + h(\mathbf{A}) = q(\mathbf{A}) \mathbf{0} + h(\mathbf{A}) = h(\mathbf{A})$$

Long division is not convenient to carry out if the degree of $f(\lambda)$ is much larger than the degree of $\Delta(\lambda)$. In this case, we may solve $h(\lambda)$ directly from (3.45). Let

$$h(\lambda) := \beta_0 + \beta_1 \lambda + \cdots + \beta_{n-1} \lambda^{n-1}$$

where the n unknowns β_i are to be solved. If all n eigenvalues of \mathbf{A} are distinct, these β_i can be solved from the n equations

$$f(\lambda_i) = q(\lambda_i) \Delta(\lambda_i) + h(\lambda_i) = h(\lambda_i)$$

for $i = 1, 2, \dots, n$. If \mathbf{A} has repeated eigenvalues, then (3.45) must be differentiated to yield additional equations. This is stated as a theorem.

► **Theorem 3.5**

Given $f(\lambda)$ and an $n \times n$ matrix \mathbf{A} with characteristic polynomial

$$\Delta(\lambda) = \prod_{i=1}^m (\lambda - \lambda_i)^{n_i}$$

where $n = \sum_{i=1}^m n_i$. Define

$$h(\lambda) := \beta_0 + \beta_1 \lambda + \cdots + \beta_{n-1} \lambda^{n-1}$$

It is a polynomial of degree $n - 1$ with n unknown coefficients. These n unknowns are to be solved from the following set of n equations:

$$f^{(l)}(\lambda_i) = h^{(l)}(\lambda_i) \quad \text{for } l = 0, 1, \dots, n_i - 1 \quad \text{and } i = 1, 2, \dots, m$$

where

$$f^{(l)}(\lambda_i) := \left. \frac{d^l f(\lambda)}{d\lambda} \right|_{\lambda=\lambda_i}$$

and $h^{(l)}(\lambda_i)$ is similarly defined. Then

$$f(\mathbf{A}) = h(\mathbf{A})$$

and $h(\lambda)$ is said to equal $f(\lambda)$ on the spectrum of \mathbf{A} .

Example 3.6.1 Compute \mathbf{A}^{100} with

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix}$$

In other words, given $f(\lambda) = \lambda^{100}$, compute $f(\mathbf{A})$. The characteristic polynomial of \mathbf{A} is $\Delta(\lambda) = \lambda^2 + 2\lambda + 1 = (\lambda + 1)^2$. Let $h(\lambda) = \beta_0 + \beta_1 \lambda$. On the spectrum of \mathbf{A} , we have

$$\begin{aligned} f(-1) &= h(-1): & (-1)^{100} &= \beta_0 - \beta_1 \\ f'(-1) &= h'(-1): & 100 \cdot (-1)^{99} &= \beta_1 \end{aligned}$$

Thus we have $\beta_1 = -100$, $\beta_0 = 1 + \beta_1 = -99$, $h(\lambda) = -99 - 100\lambda$, and

$$\begin{aligned} \mathbf{A}^{100} &= \beta_0 \mathbf{I} + \beta_1 \mathbf{A} = -99 \mathbf{I} - 100 \mathbf{A} \\ &= -99 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - 100 \begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} -99 & -100 \\ 100 & 101 \end{bmatrix} \end{aligned}$$

Clearly, \mathbf{A}^{100} can also be obtained by multiplying \mathbf{A} 100 times. However, it is simpler to use Theorem 3.5.

Functions of a square matrix

Let $f(\lambda)$ be any function, not necessarily a polynomial. One way to define $f(\mathbf{A})$ is to use Theorem 3.5. Let $h(\lambda)$ be a polynomial of degree $n - 1$, where n is the order of \mathbf{A} .

We solve the coefficients of $h(\lambda)$ by equating $f(\lambda) = h(\lambda)$ on the spectrum of \mathbf{A} . Then $f(\mathbf{A})$ is defined as $h(\mathbf{A})$.

Example 3.6.2 Let

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 0 & -2 \\ 0 & 1 & 0 \\ 1 & 0 & 3 \end{bmatrix}$$

Compute $e^{\mathbf{A}_1 t}$. Or, equivalently, if $f(\lambda) = e^{\lambda t}$, what is $f(\mathbf{A}_1)$?

The characteristic polynomial of \mathbf{A}_1 is $(\lambda - 1)^2(\lambda - 2)$. Let $h(\lambda) = \beta_0 + \beta_1\lambda + \beta_2\lambda^2$. Then

$$\begin{aligned} f(1) = h(1): \quad e^t &= \beta_0 + \beta_1 + \beta_2 \\ f'(1) = h'(1): \quad te^t &= \beta_1 + 2\beta_2 \\ f(2) = h(2): \quad e^{2t} &= \beta_0 + 2\beta_1 + 4\beta_2 \end{aligned}$$

Note that in the second equation, the differentiation is with respect to λ , not t . Solving these equations yields $\beta_0 = -2te^t + e^{2t}$, $\beta_1 = 3te^t + 2e^t - 2e^{2t}$ and $\beta_2 = e^{2t} - e^t - te^t$. Thus we have

$$\begin{aligned} e^{\mathbf{A}_1 t} = h(\mathbf{A}_1) &= (-2te^t + e^{2t})\mathbf{I} + (3te^t + 2e^t - 2e^{2t})\mathbf{A}_1 \\ &\quad + (e^{2t} - e^t - te^t)\mathbf{A}_1^2 = \begin{bmatrix} 2e^t - e^{2t} & 0 & 2e^t - 2e^{2t} \\ 0 & e^t & 0 \\ e^{2t} - e^t & 0 & 2e^{2t} - e^t \end{bmatrix} \end{aligned}$$

Example 3.6.3 Let

$$\mathbf{A}_2 = \begin{bmatrix} 0 & 2 & -2 \\ 0 & 1 & 0 \\ 1 & -1 & 3 \end{bmatrix}$$

Compute $e^{\mathbf{A}_2 t}$. The characteristic polynomial of \mathbf{A}_2 is $(\lambda - 1)^2(\lambda - 2)$, which is the same as the one of \mathbf{A}_1 . Hence we have the same $h(\lambda)$ as in Example 3.6.2. Consequently, we have

$$e^{\mathbf{A}_2 t} = h(\mathbf{A}_2) = \begin{bmatrix} 2e^t - e^{2t} & 2te^t & 2e^t - 2e^{2t} \\ 0 & e^t & 0 \\ e^{2t} - e^t & -te^t & 2e^{2t} - e^t \end{bmatrix}$$

Example 3.6.4 Consider the Jordan block of order 4:

$$\hat{\mathbf{A}} = \begin{bmatrix} \lambda_1 & 1 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 \\ 0 & 0 & \lambda_1 & 1 \\ 0 & 0 & 0 & \lambda_1 \end{bmatrix} \quad (3.46)$$

Its characteristic polynomial is $(\lambda - \lambda_1)^4$. Although we can select $h(\lambda)$ as $\beta_0 + \beta_1\lambda + \beta_2\lambda^2 + \beta_3\lambda^3$, it is computationally simpler to select $h(\lambda)$ as

$$h(\lambda) = \beta_0 + \beta_1(\lambda - \lambda_1) + \beta_2(\lambda - \lambda_1)^2 + \beta_3(\lambda - \lambda_1)^3$$

This selection is permitted because $h(\lambda)$ has degree $(n-1)=3$ and $n=4$ independent unknowns. The condition $f(\lambda) = h(\lambda)$ on the spectrum of $\hat{\mathbf{A}}$ yields immediately

$$\beta_0 = f(\lambda_1), \beta_1 = f'(\lambda_1), \beta_2 = \frac{f''(\lambda_1)}{2!}, \beta_3 = \frac{f^{(3)}(\lambda_1)}{3!}$$

Thus we have

$$f(\hat{\mathbf{A}}) = f(\lambda_1)\mathbf{I} + \frac{f'(\lambda_1)}{1!}(\hat{\mathbf{A}} - \lambda_1\mathbf{I}) + \frac{f''(\lambda_1)}{2!}(\hat{\mathbf{A}} - \lambda_1\mathbf{I})^2 + \frac{f^{(3)}(\lambda_1)}{3!}(\hat{\mathbf{A}} - \lambda_1\mathbf{I})^3$$

Using the special forms of $(\hat{\mathbf{A}} - \lambda_1\mathbf{I})^k$ as discussed in (3.40), we can readily obtain

$$f(\hat{\mathbf{A}}) = \begin{bmatrix} f(\lambda_1) & f'(\lambda_1)/1! & f''(\lambda_1)/2! & f^{(3)}(\lambda_1)/3! \\ 0 & f(\lambda_1) & f'(\lambda_1)/1! & f''(\lambda_1)/2! \\ 0 & 0 & f(\lambda_1) & f'(\lambda_1)/1! \\ 0 & 0 & 0 & f(\lambda_1) \end{bmatrix} \quad (3.47)$$

If $f(\lambda) = e^{\lambda t}$, then

$$e^{\hat{\mathbf{A}}t} = \begin{bmatrix} e^{\lambda_1 t} & te^{\lambda_1 t} & t^2 e^{\lambda_1 t}/2! & t^3 e^{\lambda_1 t}/3! \\ 0 & e^{\lambda_1 t} & te^{\lambda_1 t} & t^2 e^{\lambda_1 t}/2! \\ 0 & 0 & 0 & e^{\lambda_1 t} \end{bmatrix} \quad (3.48)$$

Because functions of \mathbf{A} are defined through polynomials of \mathbf{A} , Equations (3.41) and (3.42) are applicable to functions.

Example 3.6.5 Consider

$$\mathbf{A} = \begin{bmatrix} \lambda_1 & 1 & 0 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 & 1 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{bmatrix}$$

It is block diagonal and contains two Jordan blocks. If $f(\lambda) = e^{\lambda t}$, then (3.41) and (3.48) imply

$$e^{\mathbf{A}t} = \begin{bmatrix} e^{\lambda_1 t} & te^{\lambda_1 t} & t^2 e^{\lambda_1 t}/2! & 0 & 0 \\ 0 & e^{\lambda_1 t} & te^{\lambda_1 t} & 0 & 0 \\ 0 & 0 & e^{\lambda_1 t} & 0 & 0 \\ 0 & 0 & 0 & e^{\lambda_2 t} & te^{\lambda_2 t} \\ 0 & 0 & 0 & 0 & e^{\lambda_2 t} \end{bmatrix}$$

If $f(\lambda) = (s - \lambda)^{-1}$, then (3.41) and (3.47) imply

$$(s\mathbf{I} - \mathbf{A})^{-1} = \begin{bmatrix} \frac{1}{(s-\lambda_1)} & \frac{1}{(s-\lambda_1)^2} & \frac{1}{(s-\lambda_1)^3} & 0 & 0 \\ 0 & \frac{1}{(s-\lambda_1)} & \frac{1}{(s-\lambda_1)^2} & 0 & 0 \\ 0 & 0 & \frac{1}{(s-\lambda_1)} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{(s-\lambda_2)} & \frac{1}{(s-\lambda_2)^2} \\ 0 & 0 & 0 & 0 & \frac{1}{(s-\lambda_2)} \end{bmatrix} \quad (3.49)$$

Using power series

The function of \mathbf{A} was defined using a polynomial of finite degree. We now give an alternative definition by using an infinite power series. Suppose $f(\lambda)$ can be expressed as the power series

$$f(\lambda) = \sum_{i=0}^{\infty} \beta_i \lambda^i$$

with the radius of convergence ρ . If all eigenvalues of \mathbf{A} have magnitudes less than ρ , then $f(\mathbf{A})$ can be defined as

$$f(\mathbf{A}) = \sum_{i=0}^{\infty} \beta_i \mathbf{A}^i \quad (3.50)$$

Instead of proving the equivalence of this definition and the definition based on Theorem 3.5, we use (3.50) to derive (3.47).

Example 3.6.6 Consider the Jordan form matrix $\hat{\mathbf{A}}$ in (3.46). Let $f(\lambda)$ be a function expanded in the Taylor series as

$$f(\lambda) = f(\lambda_1) + f'(\lambda_1)(\lambda - \lambda_1) + \frac{f''(\lambda_1)}{2!}(\lambda - \lambda_1)^2 + \dots$$

where λ_1 is the eigenvalue of $\hat{\mathbf{A}}$. Then we have

$$f(\hat{\mathbf{A}}) = f(\lambda_1)\mathbf{I} + f'(\lambda_1)(\hat{\mathbf{A}} - \lambda_1\mathbf{I}) + \dots + \frac{f^{(n-1)}(\lambda_1)}{(n-1)!}(\hat{\mathbf{A}} - \lambda_1\mathbf{I})^{n-1} + \dots$$

Because $(\hat{\mathbf{A}} - \lambda_1\mathbf{I})^k = \mathbf{0}$ for $k \geq n = 4$ as discussed in (3.40), the infinite series reduces immediately to (3.47). Thus the two definitions lead to the same function of a matrix.

The most important function of \mathbf{A} is the exponential function $e^{\mathbf{At}}$. The function $e^{\lambda t}$ has the following Taylor series with respect to $\lambda_1 = 0$, often called a Maclaurin

series,

$$e^{\lambda t} = 1 + \lambda t + \frac{\lambda^2 t^2}{2!} + \cdots + \frac{\lambda^n t^n}{n!} + \cdots$$

The series converges for all finite λ and t . Thus we have

$$e^{\mathbf{A}t} = \mathbf{I} + t\mathbf{A} + \frac{t^2}{2!}\mathbf{A}^2 + \cdots = \sum_{k=0}^{\infty} \frac{1}{k!} t^k \mathbf{A}^k \quad (3.51)$$

This series involves only multiplications and additions, and may converge rapidly; therefore, it is suitable for computer computation. We list in the following the program in MATLAB that computes (3.51) for $t = 1$:

```
Function E=expmdemo2 (A)
E=zeros (size(A));
F=eye (size(A));
k=1;
while norm(E+F-E, 1)>0
    E=E+F;
    F=A.*F/k;
    k=k+1;
end
```

In the program, E denotes the partial sum and F is the next term to be added to E . The first line defines the function. The next two lines initialize E and F . Let c_k denote the k th term of (3.51) with $t = 1$. Then we have $c_{k+1} = (\mathbf{A}/k)c_k$ for $k = 1, 2, \dots$. Thus we have $F = \mathbf{A} * F / k$. The computation stops if the 1-norm of $E+F-E$, denoted by $\text{norm}(E+F-E, 1)$, is rounded to 0 in computers. Because the algorithm compares F and E , not F and 0, the algorithm uses $\text{norm}(E+F-E, 1)$ instead of $\text{norm}(F, 1)$. Note that $\text{norm}(a, 1)$ is the 1-norm discussed in Section 3.2 for vectors and will be discussed in Section 3.11 for matrices. We see that the series can indeed be easily programmed. To improve the computed result, the techniques of scaling and squaring can be used. In MATLAB, the function `expmdemo2` uses (3.51). The function `expm` or `expmdemo1`, however, uses the so-called Padé approximation. It yields comparable results as `expmdemo2` but requires only about half of the computing time. Thus, `expm` is preferred to `expmdemo2`. The function `expmdemo3` uses Jordan form, but it will yield incorrect solution if a matrix is not diagonalizable. If a closed-form solution of $e^{\mathbf{A}t}$ is needed, we must use Theorem 3.5 or Jordan form to compute $e^{\mathbf{A}t}$.

We derive some important properties of $e^{\mathbf{A}t}$ to conclude this section. Using (3.51), we can readily verify the next two equalities

$$e^{\mathbf{0}} = \mathbf{I} \quad (3.52)$$

$$e^{\mathbf{A}(t_1+t_2)} = e^{\mathbf{A}t_1} e^{\mathbf{A}t_2} \quad (3.53)$$

$$[e^{\mathbf{A}t}]^{-1} = e^{-\mathbf{A}t} \quad (3.54)$$

To show (3.54), we set $t_2 = -t_1$. Then (3.53) and (3.52) imply

$$e^{\mathbf{A}t_1} e^{-\mathbf{A}t_1} = e^{\mathbf{A} \cdot 0} = e^{\mathbf{0}} = \mathbf{I}$$

which implies (3.54). Thus the inverse of $e^{\mathbf{A}t}$ can be obtained by simply changing the sign of t . Differentiating term by term of (3.51) yields

$$\begin{aligned} \frac{d}{dt} e^{\mathbf{A}t} &= \sum_{k=1}^{\infty} \frac{1}{(k-1)!} t^{k-1} \mathbf{A}^k \\ &= \mathbf{A} \left(\sum_{k=0}^{\infty} \frac{1}{k!} t^k \mathbf{A}^k \right) = \left(\sum_{k=0}^{\infty} \frac{1}{k!} t^k \mathbf{A}^k \right) \mathbf{A} \end{aligned}$$

Thus we have

$$\frac{d}{dt} e^{\mathbf{A}t} = \mathbf{A} e^{\mathbf{A}t} = e^{\mathbf{A}t} \mathbf{A} \quad (3.55)$$

This is an important equation. We mention that

$$e^{(\mathbf{A}+\mathbf{B})t} \neq e^{\mathbf{A}t} e^{\mathbf{B}t} \quad (3.56)$$

The equality holds only if \mathbf{A} and \mathbf{B} commute or $\mathbf{AB} = \mathbf{BA}$. This can be verified by direct substituting of (3.51).

The Laplace transform of a function $f(t)$ is defined as

$$\hat{f}(s) := \mathcal{L}[f(t)] = \int_0^{\infty} f(t) e^{-st} dt$$

It can be shown that

$$\mathcal{L}\left[\frac{t^k}{k!}\right] = s^{-(k+1)}$$

Taking the Laplace transform of (3.51) yields

$$\mathcal{L}[e^{\mathbf{A}t}] = \sum_{k=0}^{\infty} s^{-(k+1)} \mathbf{A}^k = s^{-1} \sum_{k=0}^{\infty} (s^{-1} \mathbf{A})^k$$

Because the infinite series

$$\sum_{k=0}^{\infty} (s^{-1} \lambda)^k = 1 + s^{-1} \lambda + s^{-2} \lambda^2 + \dots = (1 - s^{-1} \lambda)^{-1}$$

converges for $|s^{-1}\lambda| < 1$, we have

$$\sum_{k=0}^{\infty} (s^{-1}\mathbf{A})^k = \mathbf{I} + s^{-1}\mathbf{A} + s^{-2}\mathbf{A}^2 + \dots = (\mathbf{I} - s^{-1}\mathbf{A})^{-1} \quad (3.57)$$

and

$$\mathcal{L}[e^{\mathbf{At}}] = s^{-1}(\mathbf{I} - s^{-1}\mathbf{A})^{-1} = [s(\mathbf{I} - s^{-1}\mathbf{A})]^{-1} = (s\mathbf{I} - \mathbf{A})^{-1} \quad (3.58)$$

Although in this derivation we require s to be sufficiently large so that all eigenvalues of $s^{-1}\mathbf{A}$ have magnitudes less than 1, Equation (3.58) actually holds for all s except at the eigenvalues of \mathbf{A} . Equation (3.58) can also be established from (3.55). Because $\mathcal{L}[df(t)/dt] = s\mathcal{L}[f(t)] - f(0)$, applying the Laplace transform to (3.55) yields

$$s\mathcal{L}[e^{\mathbf{At}}] - e^0 = \mathbf{A}\mathcal{L}[e^{\mathbf{At}}]$$

or

$$(s\mathbf{I} - \mathbf{A})\mathcal{L}[e^{\mathbf{At}}] = e^0 = \mathbf{I}$$

which implies (3.58).

3.7 Lyapunov Equation

Consider the equation

$$\mathbf{AM} + \mathbf{MB} = \mathbf{C} \quad (3.59)$$

where \mathbf{A} and \mathbf{B} are, respectively, $n \times n$ and $m \times m$ constant matrices. In order for the equation to be meaningful, the matrices \mathbf{M} and \mathbf{C} must be of order $n \times m$. The equation is called the *Lyapunov* equation.

The equation can be written as a set of standard linear algebraic equation. To see this, we assume $n = 3$ and $m = 2$ and write (3.59) explicitly as

$$\begin{aligned} & \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \\ m_{31} & m_{32} \end{bmatrix} + \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \\ m_{31} & m_{32} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \\ &= \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \end{bmatrix} \end{aligned}$$

Multiplying them out and then equating the corresponding entries on both sides of the equality yield

$$\begin{bmatrix} a_{11} + b_{11} & a_{12} & a_{13} & b_{21} & 0 & 0 \\ a_{21} & a_{22} + b_{11} & a_{23} & 0 & b_{21} & 0 \\ a_{31} & a_{32} & a_{33} + b_{11} & 0 & 0 & b_{21} \\ b_{12} & 0 & 0 & a_{11} + b_{22} & a_{12} & a_{13} \\ 0 & b_{12} & 0 & a_{21} & a_{22} + b_{22} & a_{23} \\ 0 & 0 & b_{12} & a_{31} & a_{32} & a_{33} + b_{22} \end{bmatrix} \times \begin{bmatrix} m_{11} \\ m_{21} \\ m_{31} \\ m_{12} \\ m_{22} \\ m_{32} \end{bmatrix} = \begin{bmatrix} c_{11} \\ c_{21} \\ c_{31} \\ c_{12} \\ c_{22} \\ c_{32} \end{bmatrix} \quad (3.60)$$

This is indeed a standard linear algebraic equation. The left-hand-side matrix is a square matrix of order $n \times m = 3 \times 2 = 6$.

Let us define $\mathcal{A}(\mathbf{M}) := \mathbf{AM} + \mathbf{MB}$. Then the Lyapunov equation can be written as $\mathcal{A}(\mathbf{M}) = \mathbf{C}$. It maps an nm -dimensional linear space into itself. A scalar η is called an eigenvalue of \mathcal{A} if there exists a nonzero \mathbf{M} such that

$$\mathcal{A}(\mathbf{M}) = \eta\mathbf{M}$$

Because \mathcal{A} can be considered as a square matrix of order nm , it has nm eigenvalues η_k for $k = 1, 2, \dots, nm$. It turns out

$$\eta_k = \lambda_i + \mu_j \quad \text{for } i = 1, 2, \dots, n; j = 1, 2, \dots, m$$

where $\lambda_i, i = 1, 2, \dots, n$, and $\mu_j, j = 1, 2, \dots, m$, are, respectively, the eigenvalues of \mathbf{A} and \mathbf{B} . In other words, the eigenvalues of \mathcal{A} are all possible sums of the eigenvalues of \mathbf{A} and \mathbf{B} .

We show intuitively why this is the case. Let \mathbf{u} be an $n \times 1$ right eigenvector of \mathbf{A} associated with λ_i , that is, $\mathbf{Au} = \lambda_i \mathbf{u}$. Let \mathbf{v} be a $1 \times m$ left eigenvector of \mathbf{B} associated with μ_j , that is, $\mathbf{vB} = \mathbf{v}\mu_j$. Applying \mathcal{A} to the $n \times m$ matrix \mathbf{uv} yields

$$\mathcal{A}(\mathbf{uv}) = \mathbf{Auv} + \mathbf{uvB} = \lambda_i \mathbf{uv} + \mathbf{uv}\mu_j = (\lambda_i + \mu_j) \mathbf{uv}$$

Because both \mathbf{u} and \mathbf{v} are nonzero, so is the matrix \mathbf{uv} . Thus $(\lambda_i + \mu_j)$ is an eigenvalue of \mathcal{A} .

The determinant of a square matrix is the product of all its eigenvalues. Thus a matrix is nonsingular if and only if it has no zero eigenvalue. If there are no i and j such that $\lambda_i + \mu_j = 0$, then the square matrix in (3.60) is nonsingular and for every \mathbf{C} , there exists a unique \mathbf{M} satisfying the equation. In this case, the Lyapunov equation is said to be nonsingular. If $\lambda_i + \mu_j = 0$ for some i and j , then for a given \mathbf{C} , solutions may or may not exist. If \mathbf{C} lies in the range space of \mathcal{A} , then solutions exist and are not unique. See Problem 3.37.

The MATLAB function `m=lyap(a,b,-c)` computes the solution of the Lyapunov equation in (3.59).

3.8 Some Useful Formulas

This section discusses some formulas which will be needed later. Let \mathbf{A} and \mathbf{B} be $m \times n$ and $n \times p$ constant matrices. Then we have

$$\rho(\mathbf{AB}) \leq \min(\rho(\mathbf{A}), \rho(\mathbf{B})) \quad (3.61)$$

where ρ denotes the rank. This can be argued as follows. Let $\rho(\mathbf{B}) = \alpha$. Then \mathbf{B} has α linearly independent rows. In \mathbf{AB} , \mathbf{A} operates on the rows of \mathbf{B} . Thus the rows of \mathbf{AB} are linear combinations of the rows of \mathbf{B} . As a result, \mathbf{AB} has at most α linearly independent rows. In \mathbf{AB} , \mathbf{B} operates on the columns of \mathbf{A} . Thus if \mathbf{A} has β linearly independent columns, then \mathbf{AB} has at most β linearly independent columns. This establishes (3.61). Consequently, if $\mathbf{A} = \mathbf{B}_1 \mathbf{B}_2 \mathbf{B}_3 \dots$, then the rank of \mathbf{A} is equal to or smaller than the smallest rank of \mathbf{B}_i .

Let \mathbf{A} be $m \times n$ and let \mathbf{C} and \mathbf{D} be any $n \times n$ and $m \times m$ nonsingular matrices. Then we have

$$\rho(\mathbf{AC}) = \rho(\mathbf{A}) = \rho(\mathbf{DA}) \quad (3.62)$$

In words, the rank of a matrix will not change after pre- or postmultiplying by a nonsingular matrix. To show (3.62), we define

$$\mathbf{P} := \mathbf{AC} \quad (3.63)$$

Because $\rho(\mathbf{A}) \leq \min(m, n)$ and $\rho(\mathbf{C}) = n$, we have $\rho(\mathbf{A}) \leq \rho(\mathbf{C})$. Thus (3.61) implies

$$\rho(\mathbf{P}) \leq \min(\rho(\mathbf{A}), \rho(\mathbf{C})) \leq \rho(\mathbf{A})$$

Next we write (3.63) as $\mathbf{A} = \mathbf{PC}^{-1}$. Using the same argument, we have $\rho(\mathbf{A}) \leq \rho(\mathbf{P})$. Thus we conclude $\rho(\mathbf{P}) = \rho(\mathbf{A})$. A consequence of (3.62) is that the rank of a matrix will not change by elementary operations. Elementary operations are (1) multiplying a row or a column by a nonzero number, (2) interchanging two rows or two columns, and (3) adding the product of one row (column) and a number to another row (column). These operations are the same as multiplying nonsingular matrices. See Reference 6, p. 542.

Let \mathbf{A} be $m \times n$ and \mathbf{B} be $n \times m$. Then we have

$$\det(\mathbf{I}_m + \mathbf{AB}) = \det(\mathbf{I}_n + \mathbf{BA}) \quad (3.64)$$

where \mathbf{I}_m is the unit matrix of order m . To show (3.64), let us define

$$\mathbf{N} = \begin{bmatrix} \mathbf{I}_m & \mathbf{A} \\ \mathbf{0} & \mathbf{I}_n \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ -\mathbf{B} & \mathbf{I}_n \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} \mathbf{I}_m & -\mathbf{A} \\ \mathbf{B} & \mathbf{I}_n \end{bmatrix}$$

We compute

$$\mathbf{NP} = \begin{bmatrix} \mathbf{I}_m + \mathbf{AB} & \mathbf{0} \\ \mathbf{B} & \mathbf{I}_n \end{bmatrix}$$

and

$$\mathbf{QP} = \begin{bmatrix} \mathbf{I}_m & -\mathbf{A} \\ \mathbf{0} & \mathbf{I}_n + \mathbf{BA} \end{bmatrix}$$

Because \mathbf{N} and \mathbf{Q} are block triangular, their determinants equal the products of the determinant of their block diagonal matrices or

$$\det \mathbf{N} = \det \mathbf{I}_m \cdot \det \mathbf{I}_n = 1 = \det \mathbf{Q}$$

Likewise, we have

$$\det(\mathbf{NP}) = \det(\mathbf{I}_m + \mathbf{AB}), \quad \det(\mathbf{QP}) = \det(\mathbf{I}_n + \mathbf{BA})$$

Because

$$\det(\mathbf{NP}) = \det \mathbf{N} \det \mathbf{P} = \det \mathbf{P}$$

and

$$\det(\mathbf{QP}) = \det \mathbf{Q} \det \mathbf{P} = \det \mathbf{P}$$

we conclude $\det(\mathbf{I}_m + \mathbf{AB}) = \det(\mathbf{I}_n + \mathbf{BA})$.

In \mathbf{N} , \mathbf{Q} , and \mathbf{P} , if \mathbf{I}_n , \mathbf{I}_m , and \mathbf{B} are replaced, respectively, by $\sqrt{s}\mathbf{I}_n$, $\sqrt{s}\mathbf{I}_m$, and $-\mathbf{B}$, then we can readily obtain

$$s^n \det(s\mathbf{I}_m - \mathbf{AB}) = s^m \det(s\mathbf{I}_n - \mathbf{BA}) \quad (3.65)$$

which implies, for $n = m$ or for $n \times n$ square matrices \mathbf{A} and \mathbf{B} ,

$$\det(s\mathbf{I}_n - \mathbf{AB}) = \det(s\mathbf{I}_n - \mathbf{BA}) \quad (3.66)$$

They are useful formula.

3.9 Quadratic Form and Positive Definiteness¹

An $n \times n$ real matrix \mathbf{M} is said to be *symmetric* if its transpose equals itself. The scalar function $\mathbf{x}'\mathbf{M}\mathbf{x}$, where \mathbf{x} is an $n \times 1$ real vector and $\mathbf{M}' = \mathbf{M}$, is called a *quadratic form*. We show that all eigenvalues of symmetric \mathbf{M} are real.

The eigenvalues and eigenvectors of real matrices can be complex, as shown in Example 3.5.2. Therefore, we must allow \mathbf{x} to assume complex numbers for the time being and consider the scalar function $\mathbf{x}^*\mathbf{M}\mathbf{x}$, where \mathbf{x}^* is the complex conjugate transpose of \mathbf{x} . Taking the complex-conjugate transpose of $\mathbf{x}^*\mathbf{M}\mathbf{x}$ yields

$$(\mathbf{x}^*\mathbf{M}\mathbf{x})^* = \mathbf{x}^*\mathbf{M}^*\mathbf{x} = \mathbf{x}^*\mathbf{M}'\mathbf{x} = \mathbf{x}^*\mathbf{M}\mathbf{x} \quad (3.67)$$

where we have used the fact that the complex conjugate transpose of a real \mathbf{M} reduces to simply the transpose. Thus $\mathbf{x}^*\mathbf{M}\mathbf{x}$ is real for any complex \mathbf{x} . This assertion is not true

¹This section is used only in Sections 5.5 and 3.10. Thus its study may be tied to those sections.

if \mathbf{M} is not symmetric. Let λ be an eigenvalue of \mathbf{M} and \mathbf{v} be its eigenvector, that is, $\mathbf{M}\mathbf{v} = \lambda\mathbf{v}$. Because

$$\mathbf{v}^*\mathbf{M}\mathbf{v} = \mathbf{v}^*\lambda\mathbf{v} = \lambda(\mathbf{v}^*\mathbf{v}) \quad (3.68)$$

and because both $\mathbf{v}^*\mathbf{M}\mathbf{v}$ and $\mathbf{v}^*\mathbf{v}$ are real, the eigenvalue λ must be real. This shows that all eigenvalues of symmetric \mathbf{M} are real. After establishing this fact, we can return our study to exclusively real vector \mathbf{x} .

The eigenvalues of a symmetric matrix \mathbf{M} are all real values; they can be simple or repeated. Suppose \mathbf{M} has a repeated eigenvalue with multiplicity m , then there exist m linearly independent eigenvectors associated with the eigenvalue, as we demonstrate in Example 3.9.1 for $m = 2$. Thus it has no Jordan block of order 2 or higher. In other words, \mathbf{M} is diagonalizable as in the last matrix in (3.39). Thus for a symmetric matrix \mathbf{M} with simple or repeated eigenvalues, there exists a nonsingular matrix \mathbf{Q} such that

$$\mathbf{D} = \mathbf{Q}^{-1}\mathbf{M}\mathbf{Q} \quad \text{or} \quad \mathbf{M} = \mathbf{Q}\mathbf{D}\mathbf{Q}^{-1} \quad (3.69)$$

where \mathbf{D} is a diagonal matrix with the real eigenvalues of \mathbf{M} on the diagonal. Taking the transpose of (3.69) yields, using $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$,

$$\mathbf{M}' = (\mathbf{Q}\mathbf{D}\mathbf{Q}^{-1})' = (\mathbf{Q}^{-1})'\mathbf{D}'\mathbf{Q}'$$

which becomes, using $\mathbf{M}' = \mathbf{M}$ (by assumption) and $\mathbf{D}' = \mathbf{D}$ (because \mathbf{D} is diagonal),

$$\mathbf{M} = (\mathbf{Q}^{-1})'\mathbf{D}\mathbf{Q}'$$

Its comparison with (3.69) yields $\mathbf{Q}^{-1} = \mathbf{Q}'$. Thus \mathbf{Q} is an orthogonal matrix. This is stated as a theorem.

► Theorem 3.6

For every real symmetric matrix \mathbf{M} , there exists an orthogonal matrix \mathbf{Q} such that

$$\mathbf{M} = \mathbf{Q}\mathbf{D}\mathbf{Q}' \quad \text{or} \quad \mathbf{D} = \mathbf{Q}'\mathbf{M}\mathbf{Q}$$

where \mathbf{D} is a diagonal matrix with the eigenvalues of \mathbf{M} , which are all real, on the diagonal.

Example 3.9.1 Consider the 3×3 symmetric matrix

$$\mathbf{M} = \begin{bmatrix} 3 & 2 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

We compute its characteristic polynomial:

$$\begin{aligned} \det(\lambda\mathbf{I} - \mathbf{M}) &= \det \begin{bmatrix} \lambda - 3 & -2 & 0 \\ -2 & \lambda & 0 \\ 0 & 0 & \lambda - 4 \end{bmatrix} \\ &= (\lambda^2 - 3\lambda - 4)(\lambda - 4) = (\lambda + 1)(\lambda - 4)^2 \end{aligned}$$

Thus \mathbf{M} has three real eigenvalues $\lambda_1 = -1$, $\lambda_2 = 4$, and $\lambda_3 = 4$.

We compute an eigenvector of \mathbf{M} associated with $\lambda_1 = -1$:

$$(\mathbf{M} - \lambda_1 \mathbf{I})\mathbf{q}_1 = \begin{bmatrix} 4 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix} \mathbf{q}_1 = \mathbf{0}$$

Clearly $\bar{\mathbf{q}}_1 = [1 \ -2 \ 0]'$ is a solution. Because $\bar{\mathbf{q}}_1' \bar{\mathbf{q}}_1 = 5$, we divide every entry of $\bar{\mathbf{q}}_1$ by $\sqrt{5}$ to yield $\mathbf{q}_1 = [1/\sqrt{5} \ -2/\sqrt{5} \ 0]'$ which is a normalized eigenvector of \mathbf{M} associated with λ_1 . For the repeated eigenvalue $\lambda_2 = \lambda_3 = 4$, we solve

$$(\mathbf{M} - \lambda_2 \mathbf{I})\mathbf{q}_2 = \begin{bmatrix} -1 & 2 & 0 \\ 2 & -4 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{q}_2 = \mathbf{0}$$

The matrix has rank 1 and nullity 2. Thus we can find two linearly independent solutions (Theorem 3.3). They are $\mathbf{q}_2 = [0 \ 0 \ 1]'$ and $\bar{\mathbf{q}}_3 = [2 \ 1 \ 0]'$ or $\mathbf{q}_3 = [2/\sqrt{5} \ 1/\sqrt{5} \ 0]'$ after normalization. The three vectors \mathbf{q}_i for $i = 1, 2, 3$ clearly are orthonormal. Thus the orthogonal matrix

$$\mathbf{Q} = \begin{bmatrix} 1/\sqrt{5} & 0 & 2/\sqrt{5} \\ -2/\sqrt{5} & 0 & 1/\sqrt{5} \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{Q}^{-1} = \mathbf{Q}' = \begin{bmatrix} 1/\sqrt{5} & -2/\sqrt{5} & 0 \\ 0 & 0 & 1 \\ 2/\sqrt{5} & 1/\sqrt{5} & 0 \end{bmatrix}$$

will transform \mathbf{M} into the diagonal matrix $\mathbf{D} = \text{diag}[-1, 4, 4]$. We verify $\mathbf{M} = \mathbf{Q}\mathbf{D}\mathbf{Q}'$ or, equivalently, $\mathbf{MQ} = \mathbf{QD}$. We compute

$$\mathbf{MQ} = \begin{bmatrix} 3 & 2 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 0 & 2/\sqrt{5} \\ -2/\sqrt{5} & 0 & 1/\sqrt{5} \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} -1/\sqrt{5} & 0 & 8/\sqrt{5} \\ 2/\sqrt{5} & 0 & 4/\sqrt{5} \\ 0 & 4 & 0 \end{bmatrix}$$

and

$$\mathbf{QD} = \begin{bmatrix} 1/\sqrt{5} & 0 & 2/\sqrt{5} \\ -2/\sqrt{5} & 0 & 1/\sqrt{5} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} = \begin{bmatrix} -1/\sqrt{5} & 0 & 8/\sqrt{5} \\ 2/\sqrt{5} & 0 & 4/\sqrt{5} \\ 0 & 4 & 0 \end{bmatrix}$$

They indeed equal each other. Note that the MATLAB function `[q, d] = eig(m)` will also generate \mathbf{Q} and \mathbf{D} .

A symmetric matrix \mathbf{M} is said to be *positive definite*, denoted by $\mathbf{M} > 0$, if $\mathbf{x}' \mathbf{M} \mathbf{x} > 0$ for every nonzero \mathbf{x} . It is *positive semidefinite*, denoted by $\mathbf{M} \geq 0$, if $\mathbf{x}' \mathbf{M} \mathbf{x} \geq 0$ for every nonzero \mathbf{x} . If $\mathbf{M} > 0$, then $\mathbf{x}' \mathbf{M} \mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$. If \mathbf{M} is positive semidefinite, then there exists a nonzero \mathbf{x} such that $\mathbf{x}' \mathbf{M} \mathbf{x} = 0$. This property will be repeatedly used later.

► Theorem 3.7

A symmetric $n \times n$ matrix \mathbf{M} is positive definite (positive semidefinite) if and only if any one of the following conditions holds.

1. Every eigenvalue of \mathbf{M} is positive (zero or positive).

2. All the *leading* principal minors of \mathbf{M} are positive (All the principal minors of \mathbf{M} are zero or positive).
3. There exists an $n \times n$ nonsingular matrix \mathbf{N} (an $n \times n$ singular matrix \mathbf{N} or an $m \times n$ matrix \mathbf{N} with $m < n$) such that $\mathbf{M} = \mathbf{N}'\mathbf{N}$.

Condition 1 can be readily proved by using Theorem 3.6. Next we consider Condition 3. If $\mathbf{M} = \mathbf{N}'\mathbf{N}$, then

$$\mathbf{x}'\mathbf{M}\mathbf{x} = \mathbf{x}'\mathbf{N}'\mathbf{N}\mathbf{x} = (\mathbf{Nx})'(\mathbf{Nx}) = \|\mathbf{Nx}\|_2^2 \geq 0$$

for any \mathbf{x} . If \mathbf{N} is nonsingular, the only \mathbf{x} to make $\mathbf{Nx} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$. Thus \mathbf{M} is positive definite. If \mathbf{N} is singular, there exists a nonzero \mathbf{x} to make $\mathbf{Nx} = \mathbf{0}$. Thus \mathbf{M} is positive semidefinite. For a proof of Condition 2, see Reference 13.

We use an example to illustrate the principal minors and leading principal minors. Consider

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}$$

Its principal minors are m_{11}, m_{22}, m_{33} ,

$$\det \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}, \quad \det \begin{bmatrix} m_{11} & m_{13} \\ m_{31} & m_{33} \end{bmatrix}, \quad \det \begin{bmatrix} m_{22} & m_{23} \\ m_{32} & m_{33} \end{bmatrix},$$

and $\det \mathbf{M}$. Thus the principal minors are the determinants of all submatrices of \mathbf{M} whose diagonals coincide with the diagonal of \mathbf{M} . The leading principal minors of \mathbf{M} are $m_{11}, \det \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$, and $\det \mathbf{M}$. Thus the leading principal minors of \mathbf{M} are the determinants of the submatrices of \mathbf{M} obtained by deleting the last k columns and last k rows for $k = 2, 1, 0$.

► Theorem 3.8

1. An $m \times n$ matrix \mathbf{H} , with $m \geq n$, has rank n , if and only if the $n \times n$ matrix $\mathbf{H}'\mathbf{H}$ has rank n or $\det(\mathbf{H}'\mathbf{H}) \neq 0$.
2. An $m \times n$ matrix \mathbf{H} , with $m \leq n$, has rank m , if and only if the $m \times m$ matrix $\mathbf{H}\mathbf{H}'$ has rank m or $\det(\mathbf{H}\mathbf{H}') \neq 0$.

The symmetric matrix $\mathbf{H}'\mathbf{H}$ is always positive semidefinite. It becomes positive definite if $\mathbf{H}'\mathbf{H}$ is nonsingular. We give a proof of this theorem. The argument in the proof will be used to establish the main results in Chapter 6; therefore, the proof is spelled out in detail.



Proof: *Necessity:* The condition $\rho(\mathbf{H}'\mathbf{H}) = n$ implies $\rho(\mathbf{H}) = n$. We show this by contradiction. Suppose $\rho(\mathbf{H}'\mathbf{H}) = n$ but $\rho(\mathbf{H}) < n$. Then there exists a nonzero vector \mathbf{v} such that $\mathbf{H}\mathbf{v} = \mathbf{0}$ which implies $\mathbf{H}'\mathbf{H}\mathbf{v} = \mathbf{0}$. This contradicts $\rho(\mathbf{H}'\mathbf{H}) = n$. Thus $\rho(\mathbf{H}'\mathbf{H}) = n$ implies $\rho(\mathbf{H}) = n$.

Sufficiency: The condition $\rho(\mathbf{H}) = n$ implies $\rho(\mathbf{H}'\mathbf{H}) = n$. Suppose not, or $\rho(\mathbf{H}'\mathbf{H}) < n$, then there exists a nonzero vector \mathbf{v} such that $\mathbf{H}'\mathbf{H}\mathbf{v} = \mathbf{0}$ which implies $\mathbf{v}'\mathbf{H}'\mathbf{H}\mathbf{v} = 0$ or

$$0 = \mathbf{v}'\mathbf{H}'\mathbf{H}\mathbf{v} = (\mathbf{H}\mathbf{v})'(\mathbf{H}\mathbf{v}) = \|\mathbf{H}\mathbf{v}\|_2^2$$

Thus we have $\mathbf{H}\mathbf{v} = \mathbf{0}$. This contradicts the hypotheses that $\mathbf{v} \neq 0$ and $\rho(\mathbf{H}) = n$. Thus $\rho(\mathbf{H}) = n$ implies $\rho(\mathbf{H}'\mathbf{H}) = n$. This establishes the first part of Theorem 3.8. The second part can be similarly established. Q.E.D.

We discuss the relationship between the eigenvalues of $\mathbf{H}'\mathbf{H}$ and those of $\mathbf{H}\mathbf{H}'$. Because both $\mathbf{H}'\mathbf{H}$ and $\mathbf{H}\mathbf{H}'$ are symmetric and positive semidefinite, their eigenvalues are real and nonnegative (zero or positive). If \mathbf{H} is $m \times n$, then $\mathbf{H}'\mathbf{H}$ has n eigenvalues and $\mathbf{H}\mathbf{H}'$ has m eigenvalues. Let $\mathbf{A} = \mathbf{H}$, and $\mathbf{B} = \mathbf{H}'$. Then (3.65) becomes

$$\det(s\mathbf{I}_m - \mathbf{H}\mathbf{H}') = s^{m-n} \det(s\mathbf{I}_n - \mathbf{H}'\mathbf{H}) \quad (3.70)$$

This implies that the characteristic polynomials of $\mathbf{H}\mathbf{H}'$ and $\mathbf{H}'\mathbf{H}$ differ only by s^{m-n} . Thus we conclude that $\mathbf{H}\mathbf{H}'$ and $\mathbf{H}'\mathbf{H}$ have the same nonzero eigenvalues but may have different numbers of zero eigenvalues. Furthermore, they have at most $\bar{n} := \min(m, n)$ number of nonzero eigenvalues.

3.10 Singular Value Decomposition

The singular value decomposition is a reliable numerical method of computing the rank, range, and null space of a matrix, as we saw in the preceding sections. Thus its familiarity is useful. Its theoretical result however will be used only in Section 7.10, thus its study may be tied to that section.

Let \mathbf{H} be an $m \times n$ real matrix. Define $\mathbf{M} := \mathbf{H}'\mathbf{H}$. Clearly, \mathbf{M} is $n \times n$, symmetric and semidefinite. Thus all eigenvalues of \mathbf{M} are real and nonnegative (zero or positive). Let r be the number of its positive eigenvalues. Then the eigenvalues of $\mathbf{M} = \mathbf{H}'\mathbf{H}$ can be arranged as

$$\lambda_1^2 \geq \lambda_2^2 \geq \cdots \lambda_r^2 > 0 = \lambda_{r+1} = \cdots = \lambda_n$$

Let $\bar{n} := \min(m, n)$. Then the set

$$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_{\bar{n}}$$

is called the *singular values* of \mathbf{H} . The singular values are usually arranged in descending order in magnitude.

Example 3.10.1 Consider the 2×3 matrix

$$\mathbf{H} = \begin{bmatrix} -4 & -1 & 2 \\ 2 & 0.5 & -1 \end{bmatrix}$$

We compute

$$\mathbf{M} = \mathbf{H}'\mathbf{H} = \begin{bmatrix} 20 & 5 & -10 \\ 5 & 1.25 & -2.5 \\ -10 & -2.5 & 5 \end{bmatrix}$$

and compute its characteristic polynomial as

$$\det(\lambda\mathbf{I} - \mathbf{M}) = \lambda^3 - 26.25\lambda^2 = \lambda^2(\lambda - 26.25)$$

Thus the eigenvalues of $\mathbf{H}'\mathbf{H}$ are 26.25, 0, and 0, and the singular values of \mathbf{H} are $\sqrt{26.25} = 5.1235$ and 0. Note that the number of singular values equals $\min(n, m)$.

In view of (3.70), we can also compute the singular values of \mathbf{H} from the eigenvalues of $\mathbf{H}\mathbf{H}'$. Indeed, we have

$$\bar{\mathbf{M}} := \mathbf{H}\mathbf{H}' = \begin{bmatrix} 21 & -10.5 \\ -10.5 & 5.25 \end{bmatrix}$$

and

$$\det(\lambda\mathbf{I} - \bar{\mathbf{M}}) = \lambda^2 - 26.25\lambda = \lambda(\lambda - 26.25)$$

Thus the eigenvalues of $\mathbf{H}\mathbf{H}'$ are 26.25 and 0 and the singular values of \mathbf{H}' are 5.1235 and 0. We see that the eigenvalues of $\mathbf{H}'\mathbf{H}$ differ from those of $\mathbf{H}\mathbf{H}'$ only in the number of zero eigenvalues, and the singular values of \mathbf{H} equal the singular values of \mathbf{H}' .

For $\mathbf{M} = \mathbf{H}'\mathbf{H}$, there exists, following Theorem 3.6, an orthogonal matrix \mathbf{Q} such that

$$\mathbf{Q}'\mathbf{H}'\mathbf{H}\mathbf{Q} = \mathbf{D} =: \mathbf{S}'\mathbf{S} \quad (3.71)$$

where \mathbf{D} is an $n \times n$ diagonal matrix with λ_i^2 on the diagonal. The matrix \mathbf{S} is $m \times n$ with the singular values λ_i on the diagonal. Manipulation on (3.71) will lead eventually to the theorem that follows.

► Theorem 3.9 (Singular value decomposition)

Every $m \times n$ matrix \mathbf{H} can be transformed into the form

$$\mathbf{H} = \mathbf{RSQ}'$$

with $\mathbf{R}'\mathbf{R} = \mathbf{RR}' = \mathbf{I}_m$, $\mathbf{Q}'\mathbf{Q} = \mathbf{QQ}' = \mathbf{I}_n$ and \mathbf{S} is $m \times n$ with the singular values of \mathbf{H} on the diagonal.

The columns of \mathbf{Q} are orthonormalized eigenvectors of $\mathbf{H}'\mathbf{H}$ and the columns of \mathbf{R} are orthonormalized eigenvectors of $\mathbf{H}\mathbf{H}'$. Once \mathbf{R} , \mathbf{S} , and \mathbf{Q} are computed, the rank of \mathbf{H} equals the number of nonzero singular values. If the rank of \mathbf{H} is r , the first r columns of \mathbf{R} is an orthonormal basis of the range space of \mathbf{H} . The last $(n - r)$ columns of \mathbf{Q} is an orthonormal basis of the null space of \mathbf{H} . Although computing singular value decomposition is time-consuming, it is very reliable and gives a quantitative measure of the rank. Thus, it is used in MATLAB to compute the rank, range space, and null space. In MATLAB, the singular values of \mathbf{H} can be obtained by typing `s=svd(H)`. Typing

$[R, S, Q] = \text{svd}(H)$ yields the three matrices in the theorem. Typing $\text{orth}(H)$ and $\text{null}(H)$ yields, respectively, orthonormal bases of the range space and null space of H . The function null will be repeatedly used in Chapter 7.

Example 3.10.2 Consider the matrix in (3.11). We type

```
a=[0 1 1 2;1 2 3 4;2 0 2 0];
[r,s,q]=svd(a)
```

which yield

$$r = \begin{bmatrix} -0.3782 & -0.3084 & 0.8729 \\ -0.8877 & -0.1468 & -0.4364 \\ -0.2627 & 0.9399 & 0.2182 \end{bmatrix}, \quad s = \begin{bmatrix} 6.1568 & 0 & 0 & 0 \\ 0 & 2.4686 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$q = \begin{bmatrix} -0.2295 & 0.7020 & -0.6194 & -0.2662 \\ -0.3498 & -0.2439 & 0.2293 & -0.8750 \\ -0.5793 & 0.4581 & 0.6194 & 0.2662 \\ -0.6996 & -0.4877 & -0.4243 & -0.3044 \end{bmatrix}$$

Thus the singular values of the matrix A in (3.11) are 6.1568, 2.4686, and 0. The matrix has two nonzero singular values, thus its rank is 2 and, consequently, its nullity is $4 - \rho(A) = 2$. The first two columns of r are essentially the orthonormal basis in (3.13), and the last two columns of q are the orthonormal basis in (3.14).

3.11 Norms of Matrices

The concept of norms for vectors can be extended to matrices. This concept is needed in Chapter 5. Let A be an $m \times n$ matrix. The norm of A can be defined as

$$\|A\| = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|} = \sup_{\|\mathbf{x}\|=1} \|\mathbf{Ax}\| \quad (3.72)$$

where \sup stands for supremum or the least upper bound. This norm is defined through the norm of \mathbf{x} and is, therefore, called an *induced norm*. For different $\|\mathbf{x}\|$, we have different $\|A\|$. For example, if the 1-norm $\|\mathbf{x}\|_1$ is used, then

$$\|A\|_1 = \max_j \left(\sum_{i=1}^m |a_{ij}| \right) = \text{largest column absolute sum}$$

where a_{ij} is the ij th element of A . If the Euclidean norm $\|\mathbf{x}\|_2$ is used, then

$$\begin{aligned} \|A\|_2 &= \text{largest singular value of } A \\ &= (\text{largest eigenvalue value of } A' A)^{1/2} \end{aligned}$$

If the infinite-norm $\|\mathbf{x}\|_\infty$ is used, then

$$\|\mathbf{A}\|_\infty = \max_i \left(\sum_{j=1}^n |a_{ij}| \right) = \text{largest row absolute sum}$$

These norms are all different for the same \mathbf{A} . For example, if

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ -1 & 0 \end{bmatrix}$$

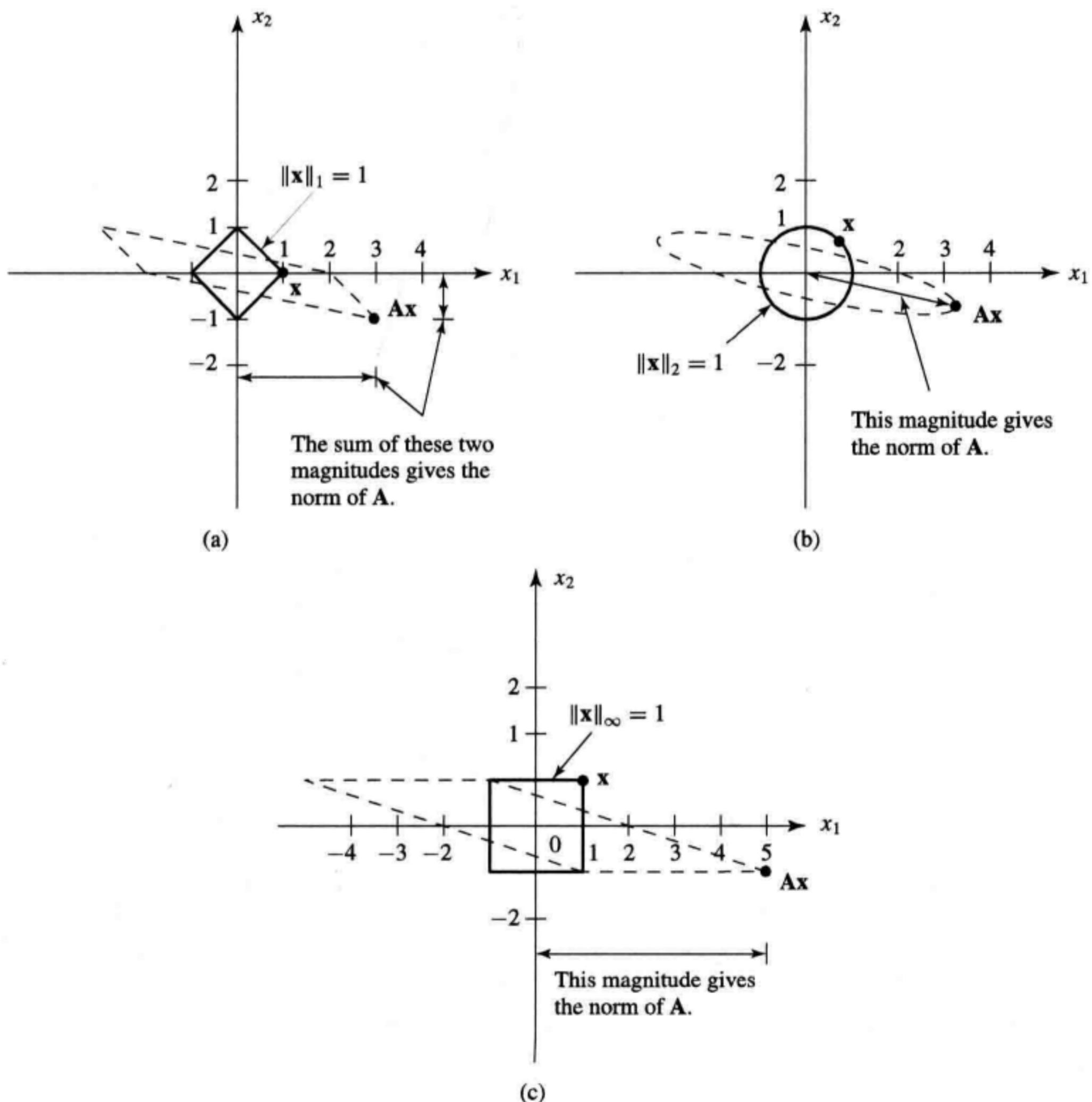


FIGURE 3.3 Different norms of \mathbf{A} .

then $\|\mathbf{A}\|_1 = 3 + |-1| = 4$, $\|\mathbf{A}\|_2 = 3.7$, and $\|\mathbf{A}\|_\infty = 3 + 2 = 5$ as shown in Fig. 3.3. The MATLAB functions `norm(a, 1)`, `norm(a, 2) = norm(a)`, and `norm(a, inf)` compute the three norms.

The norm of matrices has the following properties:

$$\begin{aligned} \|\mathbf{Ax}\| &\leq \|\mathbf{A}\| \|\mathbf{x}\| \\ \|\mathbf{A} + \mathbf{B}\| &\leq \|\mathbf{A}\| + \|\mathbf{B}\| \\ \|\mathbf{AB}\| &\leq \|\mathbf{A}\| \|\mathbf{B}\| \end{aligned}$$

Problems²

- 3.1** Consider Fig. 3.1. What is the representation of the vector \mathbf{x} with respect to the basis $\{\mathbf{q}_1, \mathbf{i}_2\}$? What is the representation of \mathbf{q}_1 with respect to $\{\mathbf{i}_2, \mathbf{q}_2\}$?
- 3.2** What are the 1-, 2-, and infinite norm of the vectors

$$\mathbf{x}_1 = \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

- 3.3** Find two orthonormal vectors that span the same space as the two vectors in Problem 3.2.
- 3.4** If two vectors $x_1 = [3 \ -4 \ 5]'$ and $x_2 = [\alpha \ 1 \ 2]'$ are orthogonal, find α . With the obtained value of α find the 1-, 2-, and infinite norm of the vectors.
- 3.5** Construct an orthogonal matrix \mathbf{Q} from the matrix

$$\mathbf{A} = \begin{bmatrix} 5 & 0 & 5 \\ 0 & 2 & 2 \\ 5 & 0 & -5 \end{bmatrix}$$

and then express \mathbf{Q} as a product of \mathbf{A} and an upper triangular matrix. Verify $\mathbf{Q}'\mathbf{Q} = \mathbf{QQ}' = \mathbf{I}$. Let $\bar{\mathbf{Q}}$ consist of the first two columns of \mathbf{Q} . Verify $\bar{\mathbf{Q}}'\bar{\mathbf{Q}} = \mathbf{I}_2$ and $\bar{\mathbf{Q}}\bar{\mathbf{Q}}' \neq \mathbf{I}_3$.

- 3.6** Find the ranks and nullities of the following matrices:

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 4 & 1 & 1 \\ 3 & 2 & 0 \\ 1 & 1 & 0 \end{bmatrix}, \quad \mathbf{A}_3 = \begin{bmatrix} 1 & 2 & -3 & 4 \\ 0 & -1 & 2 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- 3.7** Find bases of the range spaces and null spaces of the matrices in Problem 3.6.
- 3.8** Find a third column of the matrix A so that it becomes orthogonal.

$$A = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & * \\ 1/\sqrt{3} & 0 & * \\ 1/\sqrt{3} & -1/\sqrt{2} & * \end{bmatrix}$$

²It is suggested that the reader solves all numerical problems by hand and then check the results using MATLAB.

- 3.9** If u is a unit vector, show that $Q = I - 2uu'$ is an orthogonal matrix. Compute Q when $u = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix}'$.
- 3.10** Find a basis for the subspace of \mathbb{R}^4 in which the three components of the basis vector satisfy $x_1 = x_2 = x_3$.
- 3.11** Answer the following as True or False with appropriate reasoning.
1. If the vectors v_1, v_2, \dots, v_m span a subspace S , then $\dim S = m$.
 2. If matrix A is $m \times n$ and matrix B is $n \times m$ with $n < m$, then AB is singular.
 3. $\text{Rank}(A + B) \leq \text{Rank}(A) + \text{Rank}(B)$.

- 3.12** Consider the linear algebraic equation

$$\begin{bmatrix} 2 & -1 \\ -3 & 3 \\ -1 & 2 \end{bmatrix} \mathbf{x} = \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix} =: \mathbf{y}$$

It has three equations and two unknowns. Does a solution \mathbf{x} exist in the equation? Is the solution unique? Does a solution exist if $\mathbf{y} = [1 \ 1 \ 1]'$?

- 3.13** Find the general solution of

$$\begin{bmatrix} 1 & 2 & -3 & 4 \\ 0 & -1 & 2 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

How many parameters do you have?

- 3.14** Find the solution in Example 3.3.2 that has the smallest Euclidean norm.
3.15 Find the solution in Problem 3.13 that has the smallest Euclidean norm.
3.16 Consider the equation

$$\mathbf{x}[n] = \mathbf{A}^n \mathbf{x}[0] + \mathbf{A}^{n-1} \mathbf{b} u[0] + \mathbf{A}^{n-2} \mathbf{b} u[1] + \dots + \mathbf{A} b u[n-2] + \mathbf{b} u[n-1]$$

where \mathbf{A} is an $n \times n$ matrix and \mathbf{b} is an $n \times 1$ column vector. Under what conditions on \mathbf{A} and \mathbf{b} will there exist $u[0], u[1], \dots, u[n-1]$ to meet the equation for any $\mathbf{x}[n]$ and $\mathbf{x}[0]$?
Hint: Write the equation in the form

$$\mathbf{x}[n] - \mathbf{A}^n \mathbf{x}[0] = [\mathbf{b} \ \mathbf{Ab} \ \dots \ \mathbf{A}^{n-1} \mathbf{b}] \begin{bmatrix} u[n-1] \\ u[n-2] \\ \vdots \\ u[0] \end{bmatrix}$$

- 3.17** Consider

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{b}_1 = \begin{bmatrix} 0 \\ 0 \\ 2 \\ -3 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 2 \end{bmatrix}, \quad \mathbf{b}_3 = \begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$

Show that the square matrix of order 4

$$\mathbf{C}_i := [\mathbf{b}_i \ \mathbf{Ab}_i \ \mathbf{A}^2 \mathbf{b}_i \ \mathbf{A}^3 \mathbf{b}_i]$$

has rank 4 for $i = 1, 2$ and rank 3 for $i = 3$. What are the representations of \mathbf{A} with respect to the basis $\{\mathbf{b}_1, \mathbf{Ab}_1, \mathbf{A}^2\mathbf{b}_1, \mathbf{A}^3\mathbf{b}_1\}$ and the basis $\{\mathbf{b}_2, \mathbf{Ab}_2, \mathbf{A}^2\mathbf{b}_2, \mathbf{A}^3\mathbf{b}_2\}$, respectively? (Note that the representations are the same!)

- 3.18** Find Jordan-form representations of the following matrices:

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 4 & 10 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & -4 & -3 \end{bmatrix}$$

$$\mathbf{A}_3 = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad \mathbf{A}_4 = \begin{bmatrix} 0 & 4 & 3 \\ 0 & 20 & 16 \\ 0 & -25 & -20 \end{bmatrix}$$

Note that all except \mathbf{A}_4 can be diagonalized.

- 3.19** Consider the companion-form matrix

$$\mathbf{A} = \begin{bmatrix} -\alpha_1 & -\alpha_2 & -\alpha_3 & -\alpha_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Show that its characteristic polynomial is given by

$$\Delta(\lambda) = \lambda^4 + \alpha_1\lambda^3 + \alpha_2\lambda^2 + \alpha_3\lambda + \alpha_4$$

Show also that if λ_i is an eigenvalue of \mathbf{A} or a solution of $\Delta(\lambda) = 0$, then $[\lambda_i^3 \ \lambda_i^2 \ \lambda_i \ 1]'$ is an eigenvector of \mathbf{A} associated with λ_i .

- 3.20** Show that the *Vandermonde* determinant

$$\begin{bmatrix} \lambda_1^3 & \lambda_2^3 & \lambda_3^3 & \lambda_4^3 \\ \lambda_1^2 & \lambda_2^2 & \lambda_3^2 & \lambda_4^2 \\ \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

equals $\prod_{1 \leq i < j \leq 4} (\lambda_j - \lambda_i)$. Thus we conclude that the matrix is nonsingular or, equivalently, the eigenvectors are linearly independent if all eigenvalues are distinct.

- 3.21** Show that the companion-form matrix in Problem 3.19 is nonsingular if and only if $\alpha_4 \neq 0$. Under this assumption, show that its inverse equals

$$\mathbf{A}^{-1} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1/\alpha_4 & -\alpha_1/\alpha_4 & -\alpha_2/\alpha_4 & -\alpha_3/\alpha_4 \end{bmatrix}$$

- 3.22** Consider

$$\mathbf{A} = \begin{bmatrix} \lambda & \lambda T & \lambda T^2/2 \\ 0 & \lambda & \lambda T \\ 0 & 0 & \lambda \end{bmatrix}$$

with $\lambda \neq 0$ and $T > 0$. Show that $[0 \ 0 \ 1]'$ is a generalized eigenvector of grade 3 and the three columns of

$$\mathbf{Q} = \begin{bmatrix} \lambda^2 T^2 & \lambda T^2 & 0 \\ 0 & \lambda T & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

constitute a chain of generalized eigenvectors of length 3. Verify

$$\mathbf{Q}^{-1} \mathbf{A} \mathbf{Q} = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}$$

- 3.23** Find the characteristic polynomials and the minimal polynomials of the following matrices:

$$\begin{bmatrix} \lambda_1 & 1 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 \\ 0 & 0 & \lambda_1 & 0 \\ 0 & 0 & 0 & \lambda_2 \end{bmatrix} \quad \begin{bmatrix} \lambda_1 & 1 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 \\ 0 & 0 & \lambda_1 & 0 \\ 0 & 0 & 0 & \lambda_1 \end{bmatrix}$$

$$\begin{bmatrix} \lambda_1 & 1 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 \\ 0 & 0 & 0 & \lambda_1 \end{bmatrix} \quad \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 \\ 0 & 0 & 0 & \lambda_1 \end{bmatrix}$$

- 3.24** Show that if λ is an eigenvalue of \mathbf{A} with eigenvector \mathbf{x} , then $f(\lambda)$ is an eigenvalue of $f(\mathbf{A})$ with the same eigenvector \mathbf{x} .
- 3.25** Show that an $n \times n$ matrix has the property $\mathbf{A}^k = \mathbf{0}$ for $k \geq m$ if and only if \mathbf{A} has eigenvalues 0 with multiplicity n and index m or less. Such a matrix is called a *nilpotent* matrix.
- 3.26** Consider

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

- Compute \mathbf{A}^{10} , \mathbf{A}^{103} , and $e^{\mathbf{At}}$.
- 3.27** Use two different methods to compute $e^{\mathbf{At}}$ for \mathbf{A}_1 and \mathbf{A}_4 in Problem 3.18.
- 3.28** Show that functions of the same matrix commute; that is

$$f(\mathbf{A})g(\mathbf{A}) = g(\mathbf{A})f(\mathbf{A})$$

- Consequently we have $\mathbf{A}e^{\mathbf{At}} = e^{\mathbf{At}}\mathbf{A}$.
- 3.29** Let

$$\mathbf{C} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

Find a matrix \mathbf{B} such that $e^{\mathbf{B}} = \mathbf{C}$. Show that if $\lambda_i = 0$ for some i , then \mathbf{B} does not exist. Let

$$\mathbf{C} = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}$$

Find a \mathbf{B} such that $e^{\mathbf{B}} = \mathbf{C}$. Is it true that for any nonsingular \mathbf{C} , there exists a matrix \mathbf{B} such that $e^{\mathbf{B}} = \mathbf{C}$?

3.30 Let

$$(s\mathbf{I} - \mathbf{A})^{-1} = \frac{1}{\Delta(s)} \text{Adj}(s\mathbf{I} - \mathbf{A})$$

and let $m(s)$ be the monic greatest common divisor of all entries of $\text{Adj}(s\mathbf{I} - \mathbf{A})$. Verify for the matrix \mathbf{A}_3 in Problem 3.18 that the minimal polynomial of \mathbf{A} equals $\Delta(s)/m(s)$.

3.31 Define

$$(s\mathbf{I} - \mathbf{A})^{-1} := \frac{1}{\Delta(s)} \left[\mathbf{R}_0 s^{n-1} + \mathbf{R}_1 s^{n-2} + \cdots + \mathbf{R}_{n-2} s + \mathbf{R}_{n-1} \right]$$

where

$$\Delta(s) := \det(s\mathbf{I} - \mathbf{A}) := s^n + \alpha_1 s^{n-1} + \alpha_2 s^{n-2} + \cdots + \alpha_n$$

and \mathbf{R}_i are constant matrices. This definition is valid because the degree in s of the adjoint of $(s\mathbf{I} - \mathbf{A})$ is at most $n - 1$. Verify

$$\begin{aligned} \alpha_1 &= -\frac{\text{tr}(\mathbf{A}\mathbf{R}_0)}{1}, & \mathbf{R}_0 &= \mathbf{I} \\ \alpha_2 &= -\frac{\text{tr}(\mathbf{A}\mathbf{R}_1)}{2}, & \mathbf{R}_1 &= \mathbf{A}\mathbf{R}_0 + \alpha_1 \mathbf{I} = \mathbf{A} + \alpha_1 \mathbf{I} \\ \alpha_3 &= -\frac{\text{tr}(\mathbf{A}\mathbf{R}_2)}{3}, & \mathbf{R}_2 &= \mathbf{A}\mathbf{R}_1 + \alpha_2 \mathbf{I} = \mathbf{A}^2 + \alpha_1 \mathbf{A} + \alpha_2 \mathbf{I} \\ &\vdots & & \\ \alpha_{n-1} &= -\frac{\text{tr}(\mathbf{A}\mathbf{R}_{n-2})}{n-1}, & \mathbf{R}_{n-1} &= \mathbf{A}\mathbf{R}_{n-2} + \alpha_{n-1} \mathbf{I} = \mathbf{A}^{n-1} + \alpha_1 \mathbf{A}^{n-2} \\ &&&+ \cdots + \alpha_{n-2} \mathbf{A} + \alpha_{n-1} \mathbf{I} \\ \alpha_n &= -\frac{\text{tr}(\mathbf{A}\mathbf{R}_{n-1})}{n}, & \mathbf{0} &= \mathbf{A}\mathbf{R}_{n-1} + \alpha_n \mathbf{I} \end{aligned}$$

where *tr* stands for the *trace* of a matrix and is defined as the sum of all its diagonal entries. This process of computing α_i and \mathbf{R}_i is called the *Leverrier algorithm*.

3.32 Use Problem 3.31 to prove the Cayley–Hamilton theorem.

3.33 Use Problem 3.31 to show

$$\begin{aligned} (s\mathbf{I} - \mathbf{A})^{-1} &= \frac{1}{\Delta(s)} \left[\mathbf{A}^{n-1} + (s + \alpha_1) \mathbf{A}^{n-2} + (s^2 + \alpha_1 s + \alpha_2) \mathbf{A}^{n-3} \right. \\ &\quad \left. + \cdots + (s^{n-1} + \alpha_1 s^{n-2} + \cdots + \alpha_{n-1}) \mathbf{I} \right] \end{aligned}$$

- 3.34** Let all eigenvalues of \mathbf{A} be distinct and let \mathbf{q}_i be a right eigenvector of \mathbf{A} associated with λ_i , that is, $\mathbf{A}\mathbf{q}_i = \lambda_i\mathbf{q}_i$. Define $\mathbf{Q} = [\mathbf{q}_1 \ \mathbf{q}_2 \ \cdots \ \mathbf{q}_n]$ and define

$$\mathbf{P} := \mathbf{Q}^{-1} =: \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ \mathbf{p}_n \end{bmatrix}$$

where \mathbf{p}_i is the i th row of \mathbf{P} . Show that \mathbf{p}_i is a left eigenvector of \mathbf{A} associated with λ_i , that is, $\mathbf{p}_i \mathbf{A} = \lambda_i \mathbf{p}_i$.

- 3.35** Show that if all eigenvalues of \mathbf{A} are distinct, then $(s\mathbf{I} - \mathbf{A})^{-1}$ can be expressed as

$$(s\mathbf{I} - \mathbf{A})^{-1} = \sum \frac{1}{s - \lambda_i} \mathbf{q}_i \mathbf{p}_i$$

where \mathbf{q}_i and \mathbf{p}_i are right and left eigenvectors of \mathbf{A} associated with λ_i .

- 3.36** Find the \mathbf{M} to meet the Lyapunov equation in (3.59) with

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -2 & -2 \end{bmatrix}, \quad \mathbf{B} = 3, \quad \mathbf{C} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

What are the eigenvalues of the Lyapunov equation? Is the Lyapunov equation singular? Is the solution unique?

- 3.37** Consider the Lyapunov equation (a variation of 3.59).

$$\mathbf{A}' \mathbf{P} + \mathbf{P} \mathbf{A} = -\mathbf{Q}$$

Show that this equation can be reduced to an algebraic equation,

$$\mathbf{M} \mathbf{p} = -\mathbf{q}$$

where $\mathbf{M} = \mathbf{A}' \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{A}'$ and \mathbf{p}, \mathbf{q} are vectors generated by the components of the matrices \mathbf{P} and \mathbf{Q} formed from their rows respectively taken in order.

- 3.38** Check positive definite or semidefinite of the following matrices

$$\begin{bmatrix} 2 & 3 & 2 \\ 3 & 1 & 0 \\ 2 & 0 & 2 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 2 \end{bmatrix} \quad \begin{bmatrix} a_1 a_1 & a_1 a_2 & a_1 a_3 \\ a_2 a_1 & a_2 a_2 & a_2 a_3 \\ a_3 a_1 & a_3 a_2 & a_3 a_3 \end{bmatrix}$$

- 3.39** Compute the singular values of the following matrices

$$\begin{bmatrix} -1 & 0 & 1 \\ 2 & -1 & 0 \end{bmatrix} \quad \begin{bmatrix} -1 & 2 \\ 2 & 4 \end{bmatrix}$$

- 3.40** If \mathbf{A} is symmetric, what is the relationship between its eigenvalues and singular values?

- 3.41** Verify

$$\det \left(\mathbf{I}_n + \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} [b_1 \ b_2 \ \cdots \ b_n] \right) = 1 + \sum_{m=1}^n a_m b_m$$

- 3.42** Verify Equation (3.65).

- 3.43** Consider $\mathbf{Ax} = \mathbf{y}$, where \mathbf{A} is $m \times n$ and has rank m . Is $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}$ a solution? If not, under what condition will it be a solution? Is $\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{y}$ a solution?
- 3.44** Let A be a square matrix. Prove that

$$\rho(A) \leq \|A\|_2$$

where $\rho(A)$ is the spectral radius of A defined by

$$\rho(A) = \max \{ |\lambda| : \lambda \text{ is an eigenvalue of } A \}$$

Furthermore, if A is a symmetric matrix, show that the above is satisfied with equality.

- 3.45** Prove the following identities for partitioned matrices:

$$1. \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}\Delta^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}\Delta^{-1} \\ -\Delta^{-1}A_{21}A_{11}^{-1} & \Delta^{-1} \end{bmatrix} \text{ where } A$$

and A_{11} are assumed to be nonsingular with $\Delta = A_{22} - A_{21}A_{11}^{-1}A_{12}$. Note that A is nonsingular if and only if Δ is nonsingular. Furthermore, show that $\det A = \det A_{11} \det \Delta$.

$$2. \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \bar{\Delta}^{-1} & -\bar{\Delta}^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}\bar{\Delta}^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}\bar{\Delta}^{-1}A_{12}A_{22}^{-1} \end{bmatrix} \text{ where } A$$

and A_{22} are assumed to be nonsingular with $\bar{\Delta} = A_{11} - A_{12}A_{22}^{-1}A_{21}$. Note that A is nonsingular if and only if $\bar{\Delta}$ is nonsingular. Furthermore, show that $\det A = \det A_{22} \det \bar{\Delta}$.

- 3.46** Consider two vectors $a_1 = [1 \ 1 \ 0]'$ and $a_2 = [2 \ 5 \ 0]'$, which span a plane in three dimensional space, and a vector b located outside of this plane given by $b = [4 \ 3 \ 9]'$. Find the projection of b onto the plane spanned by a_1 and a_2 .

- 3.47** Five medical tubes are connected to a container for medical delivery. The percentage amount devoted to each tube for delivery is x_i , so $\sum_{i=1}^5 x_i = 100$. The tube 1 delivers four times more than tube 2 during the process of delivery. The percentage amount required by tube 5 is the sum of the ones required by tube 3 and 4. Find the best possible percentage of medication associated with each tube.

State-Space Solutions and Realizations

4.1 Introduction

With the mathematical background in Chapter 3, we are ready to study analytically state-space equations. Before proceeding, we mention that analytical study of rational transfer functions is usually covered in a sophomore/junior course. See, for example, Reference 10. We review here what will be needed in this text.

Consider $\hat{y}(s) = \hat{g}(s)\hat{u}(s)$, where $\hat{g}(s)$ is a strictly proper rational transfer function such as

$$\hat{g}(s) = \frac{N(s)}{D(s)} = \frac{N(s)}{(s - p_1)^{m_1} \bar{D}(s)}$$

It has a pole at p_1 with multiplicity m_1 . Then its impulse response is of the form¹

$$g(t) = a_0 e^{p_1 t} + a_1 t e^{p_1 t} + \cdots + a_{m_1-1} t^{m_1-1} e^{p_1 t} + \text{terms due to other poles} \quad (4.1)$$

for any numerator polynomial $N(s)$ of a degree less than the degree of $D(s)$. The form of the impulse response is determined by the poles alone and is independent of zeros. The zeros do affect the coefficients a_i . The coefficients a_i are real if the pole p_i is real, and complex if p_1 is complex.²

¹If $\hat{g}(s)$ is biproper, then its impulse response contains an impulse.

²If p_1 is complex, its complex conjugate is also a pole. By combining the responses of the pair of complex-conjugate poles, the coefficients can be reduced to real.

Consider a system with the proper rational transfer function

$$\hat{g}(s) = \frac{s^2 - 2s + 10}{(s+1)(s+2)(s+3)} = \frac{s^2 - 2s + 10}{s^3 + 6s^2 + 11s + 6}$$

Typing

```
n=[1 -2 10]; d=[1 6 11 6]; step(n, d)
```

in the command window of MATLAB will yield the step response of the system. We mention that even though we use the coefficients of the transfer function in the program, the response is not computed from the equation $\hat{y}(s) = \hat{g}(s)\hat{u}(s)$. The equation is not useful in computing analytically the response if $\hat{u}(s)$ is not a rational function of s , as is usually the case in practice. Even if $\hat{u}(s)$ is a rational function such as $\hat{u}(s) = 1/s$ for a step input, the equation is still not used. Solving $\hat{y}(s) = \hat{g}(s)\hat{u}(s)$ requires computing the poles of $\hat{g}(s)$, carrying out a partial fraction expansion, and a table look-up. Developing such a computer program is complicated. Moreover, computing the roots of a polynomial is very sensitive to its parameter variations. For example, the roots of

$$D(s) = s^4 + 7s^3 + 18s^2 + 20s + 8$$

are $-1, -2, -2$, and -2 . However, the roots of

$$\bar{D}(s) = s^4 + 7.001s^3 + 17.999s^2 + 20s + 8$$

are $-0.998, -2.2357$, and $-1.8837 \pm 0.1931j$. We see that the two coefficients of $D(s)$ change less than 0.1%, but all roots of $D(s)$ change their values greatly. Thus any procedure that requires computing roots is not desirable in computer computation. In conclusion, transfer functions are not used in computer computation. In fact, MATLAB uses exclusively state-space equations to compute responses, as we will discuss in this chapter.

This chapter studies mainly state-space equations. We study first their analytical or closed-form solutions. These results are needed for developing general properties of equations in this and later chapters. We then show that state-space equations can readily be used for computer computation, real-time processing, and op-amp circuit implementation. We then introduce equivalent and zero-state equivalent state-space equations and discuss their practical applications. Because transfer functions are not suitable, as discussed earlier, for computer computation, we discuss how to transform them into state-space equations, called the realization problem. Using its state-space-equation realization, a transfer function can then be computed digitally or implemented using an op-amp circuit.

This chapter discusses first time-invariant linear state-space equations and then the time varying case. Excluding the realization problem, the derivations for SISO and MIMO state-space equations are almost identical, thus we discuss only the latter. For the realization problem, we discuss first the SISO case and then the MIMO case.

4.2 General Solution of CT LTI State-Space Equations

Consider the continuous-time (CT) linear time-invariant (LTI) state-space equation

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \quad (4.2)$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) \quad (4.3)$$

where \mathbf{A} , \mathbf{B} , \mathbf{C} , and \mathbf{D} are, respectively, $n \times n$, $n \times p$, $q \times n$, and $q \times p$ constant matrices. The problem is to find the solution excited by the initial state $\mathbf{x}(0)$ and the input $\mathbf{u}(t)$ for $t \geq 0$. The solution hinges on the exponential function of \mathbf{A} studied in Section 3.6. In particular, we need the property in (3.55) or

$$\frac{d}{dt} e^{\mathbf{At}} = \mathbf{A}e^{\mathbf{At}} = e^{\mathbf{At}}\mathbf{A}$$

to develop the solution. Premultiplying $e^{-\mathbf{At}}$ on both side of (4.2) yields

$$e^{-\mathbf{At}}\dot{\mathbf{x}}(t) - e^{-\mathbf{At}}\mathbf{A}\mathbf{x}(t) = e^{-\mathbf{At}}\mathbf{B}\mathbf{u}(t)$$

which implies

$$\frac{d}{dt} \left(e^{-\mathbf{At}}\mathbf{x}(t) \right) = e^{-\mathbf{At}}\mathbf{B}\mathbf{u}(t)$$

Its integration from 0 to t yields

$$e^{-\mathbf{A}\tau}\mathbf{x}(\tau) \Big|_{\tau=0}^t = \int_0^t e^{-\mathbf{A}\tau}\mathbf{B}\mathbf{u}(\tau) d\tau$$

Thus we have

$$e^{-\mathbf{A}t}\mathbf{x}(t) - e^0\mathbf{x}(0) = \int_0^t e^{-\mathbf{A}\tau}\mathbf{B}\mathbf{u}(\tau) d\tau \quad (4.4)$$

Because the inverse of $e^{-\mathbf{At}}$ is $e^{\mathbf{At}}$ and $e^0 = \mathbf{I}$ as discussed in (3.54) and (3.52), (4.4) implies

$$\mathbf{x}(t) = e^{\mathbf{At}}\mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau) d\tau \quad (4.5)$$

This is the solution of (4.2).

It is instructive to verify that (4.5) is the solution of (4.2). To verify this, we must show that (4.5) satisfies (4.2) and the initial condition $\mathbf{x}(t) = \mathbf{x}(0)$ at $t = 0$. Indeed, at $t = 0$, (4.5) reduces to

$$\mathbf{x}(0) = e^{\mathbf{A}\cdot 0}\mathbf{x}(0) = e^0\mathbf{x}(0) = \mathbf{I}\mathbf{x}(0) = \mathbf{x}(0)$$

Thus (4.5) satisfies the initial condition. We need the equation

$$\frac{\partial}{\partial t} \int_{t_0}^t f(t, \tau) d\tau = \int_{t_0}^t \left(\frac{\partial}{\partial t} f(t, \tau) \right) d\tau + f(t, \tau) \Big|_{\tau=t} \quad (4.6)$$

to show that (4.5) satisfies (4.2). Differentiating (4.5) and using (4.6) yield

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \frac{d}{dt} \left[e^{\mathbf{A}t} \mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-\tau)} \mathbf{B} \mathbf{u}(\tau) d\tau \right] \\ &= \mathbf{A} e^{\mathbf{A}t} \mathbf{x}(0) + \int_0^t \mathbf{A} e^{\mathbf{A}(t-\tau)} \mathbf{B} \mathbf{u}(\tau) d\tau + e^{\mathbf{A}(t-\tau)} \mathbf{B} \mathbf{u}(\tau) \Big|_{\tau=t} \\ &= \mathbf{A} \left[e^{\mathbf{A}t} \mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-\tau)} \mathbf{B} \mathbf{u}(\tau) d\tau \right] + e^{\mathbf{A} \cdot 0} \mathbf{B} \mathbf{u}(t)\end{aligned}$$

which becomes, after substituting (4.5),

$$\dot{\mathbf{x}}(t) = \mathbf{A} \mathbf{x}(t) + \mathbf{B} \mathbf{u}(t)$$

Thus (4.5) meets (4.2) and the initial condition $\mathbf{x}(0)$ and is the solution of (4.2).

The solution of (4.3) can be obtained by substituting (4.5) into (4.3) as

$$\mathbf{y}(t) = \mathbf{C} e^{\mathbf{A}t} \mathbf{x}(0) + \mathbf{C} \int_0^t e^{\mathbf{A}(t-\tau)} \mathbf{B} \mathbf{u}(\tau) d\tau + \mathbf{D} \mathbf{u}(t) \quad (4.7)$$

This solution and (4.5) are computed directly in the time domain. We can also compute the solutions by using the Laplace transform. Applying the Laplace transform to (4.2) and (4.3) yields, as derived in (2.11) and its preceding equation,

$$\begin{aligned}\hat{\mathbf{x}}(s) &= (s\mathbf{I} - \mathbf{A})^{-1} [\mathbf{x}(0) + \mathbf{B} \hat{\mathbf{u}}(s)] \\ \hat{\mathbf{y}}(s) &= \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1} [\mathbf{x}(0) + \mathbf{B} \hat{\mathbf{u}}(s)] + \mathbf{D} \hat{\mathbf{u}}(s)\end{aligned}$$

Once $\hat{\mathbf{x}}(s)$ and $\hat{\mathbf{y}}(s)$ are computed algebraically, their inverse Laplace transforms yield the time-domain solutions.

We now give some remarks concerning the computation of $e^{\mathbf{A}t}$. We discussed in Section 3.6 three methods of computing functions of a matrix. They can all be used to compute $e^{\mathbf{A}t}$:

1. Using Theorem 3.5: First compute the eigenvalues of \mathbf{A} ; next, find a polynomial $h(\lambda)$ of degree $n - 1$ that equals $e^{\lambda t}$ on the spectrum of \mathbf{A} ; then $e^{\mathbf{A}t} = h(\mathbf{A})$.
2. Using Jordan form of \mathbf{A} : Let $\mathbf{A} = \mathbf{Q} \hat{\mathbf{A}} \mathbf{Q}^{-1}$; then $e^{\mathbf{A}t} = \mathbf{Q} e^{\hat{\mathbf{A}}t} \mathbf{Q}^{-1}$, where $\hat{\mathbf{A}}$ is in Jordan form and $e^{\hat{\mathbf{A}}t}$ can be readily obtained by using (3.48).
3. Using the infinite power series in (3.51); Although the series will not, in general, yield a closed-form solution, it is suitable for computer computation, as discussed following (3.51).

In addition, we can use (3.58) to compute $e^{\mathbf{A}t}$, that is,

$$e^{\mathbf{A}t} = \mathcal{L}^{-1}(s\mathbf{I} - \mathbf{A})^{-1} \quad (4.8)$$

The inverse of $(s\mathbf{I} - \mathbf{A})$ is a function of \mathbf{A} ; therefore, again we have many methods to compute it:

1. Taking the inverse of $(s\mathbf{I} - \mathbf{A})$.
2. Using Theorem 3.5.
3. Using $(s\mathbf{I} - \mathbf{A})^{-1} = \mathbf{Q}(s\mathbf{I} - \hat{\mathbf{A}})^{-1}\mathbf{Q}^{-1}$ and (3.49).
4. Using the infinite power series in (3.57).
5. Using the Leverrier algorithm discussed in Problem 3.31.

Example 4.2.1 We use Methods 1 and 2 to compute $(s\mathbf{I} - \mathbf{A})^{-1}$ where

$$\mathbf{A} = \begin{bmatrix} 0 & -1 \\ 1 & -2 \end{bmatrix}$$

Method 1: We use (3.20) to compute

$$\begin{aligned} (s\mathbf{I} - \mathbf{A})^{-1} &= \begin{bmatrix} s & 1 \\ -1 & s+2 \end{bmatrix}^{-1} = \frac{1}{s^2 + 2s + 1} \begin{bmatrix} s+2 & -1 \\ 1 & s \end{bmatrix} \\ &= \begin{bmatrix} (s+2)/(s+1)^2 & -1/(s+1)^2 \\ 1/(s+1)^2 & s/(s+1)^2 \end{bmatrix} \end{aligned}$$

Method 2: The eigenvalues of \mathbf{A} are $-1, -1$. Let $h(\lambda) = \beta_0 + \beta_1\lambda$. If $h(\lambda)$ equals $f(\lambda) := (s - \lambda)^{-1}$ on the spectrum of \mathbf{A} , then

$$\begin{aligned} f(-1) = h(-1) : \quad (s+1)^{-1} &= \beta_0 - \beta_1 \\ f'(-1) = h'(-1) : \quad (s+1)^{-2} &= \beta_1 \end{aligned}$$

Thus we have

$$h(\lambda) = [(s+1)^{-1} + (s+1)^{-2}] + (s+1)^{-2}\lambda$$

and

$$\begin{aligned} (s\mathbf{I} - \mathbf{A})^{-1} &= h(\mathbf{A}) = [(s+1)^{-1} + (s+1)^{-2}]\mathbf{I} + (s+1)^{-2}\mathbf{A} \\ &= \begin{bmatrix} (s+2)/(s+1)^2 & -1/(s+1)^2 \\ 1/(s+1)^2 & s/(s+1)^2 \end{bmatrix} \end{aligned}$$

Example 4.2.2 Consider the equation

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & -1 \\ 1 & -2 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t)$$

Its solution is

$$\mathbf{x}(t) = e^{\mathbf{A}t} \mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-\tau)} \mathbf{B} u(\tau) d\tau$$

The matrix function $e^{\mathbf{A}t}$ is the inverse Laplace transform of $(s\mathbf{I} - \mathbf{A})^{-1}$ which was computed in the preceding example. Thus we have

$$e^{\mathbf{A}t} = \mathcal{L}^{-1} \begin{bmatrix} \frac{s+2}{(s+1)^2} & \frac{-1}{(s+1)^2} \\ \frac{1}{(s+1)^2} & \frac{s}{(s+1)^2} \end{bmatrix} = \begin{bmatrix} (1+t)e^{-t} & -te^{-t} \\ te^{-t} & (1-t)e^{-t} \end{bmatrix}$$

and

$$\mathbf{x}(t) = \begin{bmatrix} (1+t)e^{-t} & -te^{-t} \\ te^{-t} & (1-t)e^{-t} \end{bmatrix} \mathbf{x}(0) + \begin{bmatrix} \int_0^t -(t-\tau)e^{-(t-\tau)} u(\tau) d\tau \\ \int_0^t [1 - (t-\tau)] e^{-(t-\tau)} u(\tau) d\tau \end{bmatrix}$$

We discuss a general property of the zero-input response $e^{\mathbf{A}t} \mathbf{x}(0)$. Consider the second matrix in (3.39). Then we have

$$e^{\mathbf{A}t} = \mathbf{Q} \begin{bmatrix} e^{\lambda_1 t} & te^{\lambda_1 t} & t^2 e^{\lambda_1 t}/2 & 0 & 0 \\ 0 & e^{\lambda_1 t} & te^{\lambda_1 t} & 0 & 0 \\ 0 & 0 & e^{\lambda_1 t} & 0 & 0 \\ 0 & 0 & 0 & e^{\lambda_1 t} & 0 \\ 0 & 0 & 0 & 0 & e^{\lambda_2 t} \end{bmatrix} \mathbf{Q}^{-1}$$

Every entry of $e^{\mathbf{A}t}$ and, consequently, of the zero-input response is a linear combination of terms $\{e^{\lambda_1 t}, te^{\lambda_1 t}, t^2 e^{\lambda_1 t}/2, e^{\lambda_2 t}\}$. These terms are dictated by the eigenvalues and their indices. In general, if \mathbf{A} has eigenvalue λ_1 with index \bar{n}_1 , then every entry of $e^{\mathbf{A}t}$ is a linear combination of

$$e^{\lambda_1 t} \quad te^{\lambda_1 t} \quad \dots \quad t^{\bar{n}_1-1} e^{\lambda_1 t}$$

Every such term is *analytic* in the sense that it is infinitely differentiable and can be expanded in a Taylor series at every t . This is a nice property and will be used in Chapter 6.

If every eigenvalue, simple or repeated, of \mathbf{A} has a negative real part, then every zero-input response will approach zero as $t \rightarrow \infty$. If \mathbf{A} has an eigenvalue, simple or repeated, with a positive real part, then most zero-input responses will grow unbounded as $t \rightarrow \infty$. If \mathbf{A} has some eigenvalues with zero real part and all with index 1 and the remaining eigenvalues all have negative real parts, then no zero-input response will grow unbounded. However, if the index is 2 or higher, then some zero-input response may become unbounded. For example, if \mathbf{A} has eigenvalue 0 with index 2, then $e^{\mathbf{A}t}$ contains the terms $\{e^{0 \cdot 0 \cdot t}, te^{0 \cdot 0 \cdot t}\} = \{1, t\}$. If a zero-input response contains the term t , then it will grow unbounded.

4.2.1 Discretization

Consider the continuous-time state-space equation

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \quad (4.9)$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) \quad (4.10)$$

If they are to be computed on a digital computer, they must be discretized. The simplest way is to use the approximation

$$\dot{\mathbf{x}}(t) \approx \frac{\mathbf{x}(t+T) - \mathbf{x}(t)}{T} \quad (4.11)$$

where $T > 0$ and is called the *step size*. The smaller the step size, the better the approximation. Using the approximation, we can write (4.9) as

$$\mathbf{x}(t+T) = \mathbf{x}(t) + \mathbf{A}\mathbf{x}(t)T + \mathbf{B}\mathbf{u}(t)T = (\mathbf{I} + \mathbf{A}T)\mathbf{x}(t) + \mathbf{B}T\mathbf{u}(t)$$

where we have used $\mathbf{x} = \mathbf{x}\mathbf{I}$ and \mathbf{I} is a unit matrix. If we compute $\mathbf{x}(t)$ and $\mathbf{y}(t)$ only at $t = kT$ for $k = 0, 1, \dots$, then

$$\begin{aligned} \mathbf{x}((k+1)T) &= (\mathbf{I} + T\mathbf{A})\mathbf{x}(kT) + T\mathbf{B}\mathbf{u}(kT) \\ \mathbf{y}(kT) &= \mathbf{C}\mathbf{x}(kT) + \mathbf{D}\mathbf{u}(kT) \end{aligned}$$

This is a discrete-time (DT) state-space equation. The discretization scheme in (4.11) is the simplest and yields the least accurate result for a given T . However, it can yield a result as accurate as any method if we select T to be sufficiently small. We next discuss a different discretization.

If an input $\mathbf{u}(t)$ is generated by a digital computer followed by a digital-to-analog converter using a zero-order hold, then $\mathbf{u}(t)$ will be a staircase function. This situation often arises in computer control of control systems. Let

$$\mathbf{u}(t) = \mathbf{u}(kT) =: \mathbf{u}[k] \quad \text{for } kT \leq t < (k+1)T \quad (4.12)$$

and for $k = 0, 1, 2, \dots$. This input changes values only at discrete-time instants. For this input, the solution of (4.9) is still given by (4.5). Computing (4.5) at $t = kT$ and $t = (k+1)T$ yields

$$\mathbf{x}[k] := \mathbf{x}(kT) = e^{\mathbf{A}kT}\mathbf{x}(0) + \int_0^{kT} e^{\mathbf{A}(kT-\tau)}\mathbf{B}\mathbf{u}(\tau)d\tau \quad (4.13)$$

and

$$\mathbf{x}[k+1] := \mathbf{x}((k+1)T) = e^{\mathbf{A}(k+1)T}\mathbf{x}(0) + \int_0^{(k+1)T} e^{\mathbf{A}((k+1)T-\tau)}\mathbf{B}\mathbf{u}(\tau)d\tau \quad (4.14)$$

Equation (4.14) can be written as

$$\begin{aligned} \mathbf{x}[k+1] &= e^{\mathbf{A}T} \left[e^{\mathbf{A}kT}\mathbf{x}(0) + \int_0^{kT} e^{\mathbf{A}(kT-\tau)}\mathbf{B}\mathbf{u}(\tau)d\tau \right] \\ &\quad + \int_{kT}^{(k+1)T} e^{\mathbf{A}(kT+T-\tau)}\mathbf{B}\mathbf{u}(\tau)d\tau \end{aligned}$$

which becomes, after substituting (4.12) and (4.13) and introducing the new variable $\alpha := kT + T - \tau$,

$$\mathbf{x}[k+1] = e^{\mathbf{A}T} \mathbf{x}[k] + \left(\int_0^T e^{\mathbf{A}\alpha} d\alpha \right) \mathbf{B} \mathbf{u}[k]$$

Thus, if an input changes value only at discrete time instants kT and if we compute only the responses at $t = kT$, then (4.9) and (4.10) become

$$\mathbf{x}[k+1] = \mathbf{A}_d \mathbf{x}[k] + \mathbf{B}_d \mathbf{u}[k] \quad (4.15)$$

$$\mathbf{y}[k] = \mathbf{C}_d \mathbf{x}[k] + \mathbf{D}_d \mathbf{u}[k] \quad (4.16)$$

with

$$\mathbf{A}_d = e^{\mathbf{A}T}, \quad \mathbf{B}_d = \left(\int_0^T e^{\mathbf{A}\tau} d\tau \right) \mathbf{B}, \quad \mathbf{C}_d = \mathbf{C}, \quad \mathbf{D}_d = \mathbf{D} \quad (4.17)$$

This is a discrete-time state-space equation. Note that there is no approximation involved in this derivation, and (4.15) yields the exact solution of (4.9) at $t = kT$ if the input is staircase.

We discuss the computation of \mathbf{B}_d . Using (3.51), we have

$$\begin{aligned} \mathbf{B}_d &= \int_0^T \left[\mathbf{I} + \mathbf{A}\tau + \mathbf{A}^2 \frac{\tau^2}{2!} + \dots \right] d\tau \\ &= T\mathbf{I} + \frac{T^2}{2}\mathbf{A} + \frac{T^3}{3!}\mathbf{A}^2 + \frac{T^4}{4!}\mathbf{A}^3 + \dots \end{aligned}$$

This power series can be computed recursively as in computing (3.51). If \mathbf{A} is nonsingular, then the series can be written as, using (3.51),

$$\mathbf{A}^{-1} \left[T\mathbf{A} + \frac{T^2}{2}\mathbf{A}^2 + \frac{T^3}{3!}\mathbf{A}^3 + \dots + \mathbf{I} - \mathbf{I} \right] = \mathbf{A}^{-1}(e^{\mathbf{A}T} - \mathbf{I})$$

Thus we have

$$\mathbf{B}_d = \mathbf{A}^{-1}(\mathbf{A}_d - \mathbf{I})\mathbf{B} \quad (\text{If } \mathbf{A} \text{ is nonsingular.})$$

Using this formula, we can avoid computing an infinite series.

The MATLAB function `[ad, bd] = c2d(a, b, T)` transforms the continuous-time state equation in (4.9) into the discrete-time state equation in (4.15).

4.2.2 General Solution of DT LTI State-Space Equations

Consider the discrete-time (DT) state-space equation

$$\mathbf{x}[k+1] = \mathbf{Ax}[k] + \mathbf{Bu}[k]$$

$$\mathbf{y}[k] = \mathbf{Cx}[k] + \mathbf{Du}[k]$$

where the subscript d has been dropped. It is understood that if the equation is obtained from a continuous equation, then the four matrices must be computed from (4.17). The two DT equations are algebraic, only involving additions and multiplications. Once $\mathbf{x}[0]$ and $\mathbf{u}[k]$, $k = 0, 1, \dots$, are given, the response can be computed recursively from the equations.

In order to discuss the general behavior of discrete-time state-space equations, we develop a general form of solutions. We compute

$$\begin{aligned}\mathbf{x}[1] &= \mathbf{Ax}[0] + \mathbf{Bu}[0] \\ \mathbf{x}[2] &= \mathbf{Ax}[1] + \mathbf{Bu}[1] = \mathbf{A}^2\mathbf{x}[0] + \mathbf{ABu}[0] + \mathbf{Bu}[1]\end{aligned}$$

Proceeding forward, we can readily obtain, for $k > 0$,

$$\begin{aligned}\mathbf{x}[k] &= \mathbf{A}^k\mathbf{x}[0] + \sum_{m=0}^{k-1} \mathbf{A}^{k-1-m}\mathbf{Bu}[m] \\ \mathbf{y}[k] &= \mathbf{CA}^k\mathbf{x}[0] + \sum_{m=0}^{k-1} \mathbf{CA}^{k-1-m}\mathbf{Bu}[m] + \mathbf{Du}[k]\end{aligned}$$

They are the discrete counterparts of (4.5) and (4.7). Their derivations are considerably simpler than the continuous-time case.

We discuss a general property of the zero-input response $\mathbf{A}^k\mathbf{x}[0]$. Suppose \mathbf{A} has eigenvalue λ_1 with multiplicity 4 and eigenvalue λ_2 with multiplicity 1 and suppose its Jordan form is as shown in the second matrix in (3.39). In other words, λ_1 has index 3 and λ_2 has index 1. Then we have, using (3.47),

$$\mathbf{A}^k = \mathbf{Q} \begin{bmatrix} \lambda_1^k & k\lambda_1^{k-1} & k(k-1)\lambda_1^{k-2}/2 & 0 & 0 \\ 0 & \lambda_1^k & k\lambda_1^{k-1} & 0 & 0 \\ 0 & 0 & \lambda_1^k & 0 & 0 \\ 0 & 0 & 0 & \lambda_1^k & 0 \\ 0 & 0 & 0 & 0 & \lambda_2^k \end{bmatrix} \mathbf{Q}^{-1}$$

which implies that every entry of the zero-input response is a linear combination of $\{\lambda_1^k, k\lambda_1^{k-1}, k^2\lambda_1^{k-2}, \lambda_2^k\}$. These terms are dictated by the eigenvalues and their indices.

If every eigenvalue, simple or repeated, of \mathbf{A} has magnitude less than 1, then every zero-input response will approach zero as $k \rightarrow \infty$. If \mathbf{A} has an eigenvalue, simple or repeated, with magnitude larger than 1, then most zero-input responses will grow unbounded as $k \rightarrow \infty$. If \mathbf{A} has some eigenvalues with magnitude 1 and all with index 1 and the remaining eigenvalues all have magnitude less than 1, then no zero-input response will grow unbounded. However, if the index is 2 or higher, then some zero-state response may become unbounded. For example, if \mathbf{A} has eigenvalue 1 with index 2, then \mathbf{A}^k contains the terms $\{1^k, k \cdot 1^k\} = \{1, k\}$. If a zero-input response contains the term k , then it will grow unbounded as $k \rightarrow \infty$.

4.3 Computer Computation of CT State-Space Equations

The subject of computer computation of continuous-time differential and integral equations is a vast one. It involves many issues: discretization, complexity of an algorithm, efficiency, and numerical stability. The best discretization scheme should yield the most accurate result for a given step size. A good algorithm should be simple to develop, requires a small number of operations (additions and multiplications), and should not be sensitive to parameter variations. These topics are beyond the scope of this text. We discuss in this section only the basic idea and the use of MATLAB functions.

Consider the CT state-space equation

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t) \\ y(t) &= \mathbf{c}\mathbf{x}(t) + du(t)\end{aligned}$$

Suppose we are asked to compute the output $y(t)$ excited by the input $u(t)$ for t in $[0, t_f]$ and the initial state $\mathbf{x}(0)$. The first step is to select a step size T and then discretize the CT state-space equation into a DT state-space equation as

$$y(kT) = \mathbf{c}_d\mathbf{x}(kT) + d_d u(kT) \quad (4.18)$$

$$\mathbf{x}((k+1)T) = \mathbf{A}_d\mathbf{x}(kT) + \mathbf{b}_d u(kT) \quad (4.19)$$

If we use the discretization scheme in (4.11), then we have $\mathbf{A}_d = \mathbf{I} + T\mathbf{A}$, $\mathbf{b}_d = T\mathbf{b}$, $\mathbf{c}_d = \mathbf{c}$, and $d_d = d$. The set of two equations can be computed recursively. We use $\mathbf{x}(0)$ and $u(0)$ to compute $y(0)$ from (4.18) and $\mathbf{x}(T)$ from (4.19). We then use the computed $\mathbf{x}(T)$ and $u(T)$ to compute $y(T)$ and $\mathbf{x}(2T)$. Proceeding forward, we can compute $y(kT)$, for $k = 0, 1, 2, \dots$. We then carry out interpolation to generate a CT signal $y(t)$ from the DT sequence $y(kT)$. This computation involves only additions and multiplications; it does not encounter the sensitivity issue discussed in Section 4.1. Thus state-space equations are most suited for computer computation.

The remaining question is how to select a step size T . Clearly, the smaller a step size, the more accurate the result. However, it will also require more computation. Thus, in practice, we will not select an unnecessarily small T . The sampling theorem states that if a signal $y(t)$ is bandlimited to ω_{max} and if the sampling period T is selected to be smaller than π/ω_{max} , then $y(t)$ can be recovered exactly from its sampled sequence $y(kT)$. To employ the theorem, we must know the frequency spectrum of $y(t)$, which is generally not available, and use the ideal interpolator, which is not physically realizable. Thus it is not simple to employ the sampling theorem. The easiest way of selecting a T is by trial and error. We select arbitrarily T_1 and $T_2 < T_1$ and carry out the computation. If the two results differ considerably, T_1 is not small enough; whether T_2 is small enough is to be determined by comparing its result with the one computed using a smaller T . We repeat the computation until the results of two subsequent T_i are indistinguishable. If the results of using the original T_1 and T_2 are indistinguishable, T_1 is small enough. In order to find a largest acceptable T , we increase T_1 until its result starts to differ from the preceding computation.

We use an example to illustrate the procedure. Consider the CT state-space equation

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & -0.2 \\ 0 & -0.75 & 0.25 \\ 0.5 & -0.5 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} 0.2 \\ 0 \\ 0 \end{bmatrix} u(t)$$

$$y(t) = [0 \ 0 \ 1] \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + 0 \times u(t)$$

We use the MATLAB function `lsim`, abbreviation for *linear simulation*, to compute the output of the system excited by the input $u(t) = \sin 2t$ for $t \geq 0$, and the initial state $\mathbf{x}(0) = [0 \ 0.2 \ -0.1]'$, where the prime denotes the transpose. We type in an edit window the following:

```
% Program 4.1 (f41.m)
a=[0 0 -0.2;0 -0.75 0.25;0.5 -0.5 0];b=[0.2;0;0];
c=[0 0 1];d=0;
dog=ss(a,b,c,d);
t1=0:1:50;u1=sin(2*t1);x=[0;0.2;-0.1];
y1=lsim(dog,u1,t1,x);

t2=0:0.1:50;u2=sin(2*t2);
y2=lsim(dog,u2,t2,x);

t3=0:0.01:50;u3=sin(2*t3);
y3=lsim(dog,u3,t3,x);
plot(t1,y1,t2,y2,:',t3,y3,'-.',[0 50],[0 0])
xlabel('Time (s)'), ylabel('Amplitude')
```

The first two lines express $\{\mathbf{A}, \mathbf{b}, \mathbf{c}, d\}$ in MATLAB format. The third line defines the system. We call the system ‘dog’, which is defined using the state-space model, denoted by `ss`. The fourth line $t1=0:1:50$ indicates the time interval $[0, 50]$ to be computed with increment 1 or, equivalently, with step size $T_1 = 1$ and the corresponding input $u1=\sin(2*t1)$. We then use `lsim` to compute the output $y1=lsim(dog,u1,t1,x)$. The solid line in Fig. 4.1 is generated by `plot(t1,y1)` which also carries out linear interpolation (connecting every pair of neighboring points by a straight line). We then repeat the computation by selecting $T_2 = 0.1$ as its step size and the result $y2$ is plotted in Fig. 4.1 with a dotted line using `plot(t2,y2,':')`. It is quite different from the solid line. Thus the result obtained using $T_1 = 1$ is not acceptable. At this point we don’t know whether the result using $T_2 = 0.1$ will be acceptable. We next select $T_3 = 0.01$ and repeat the computation. We

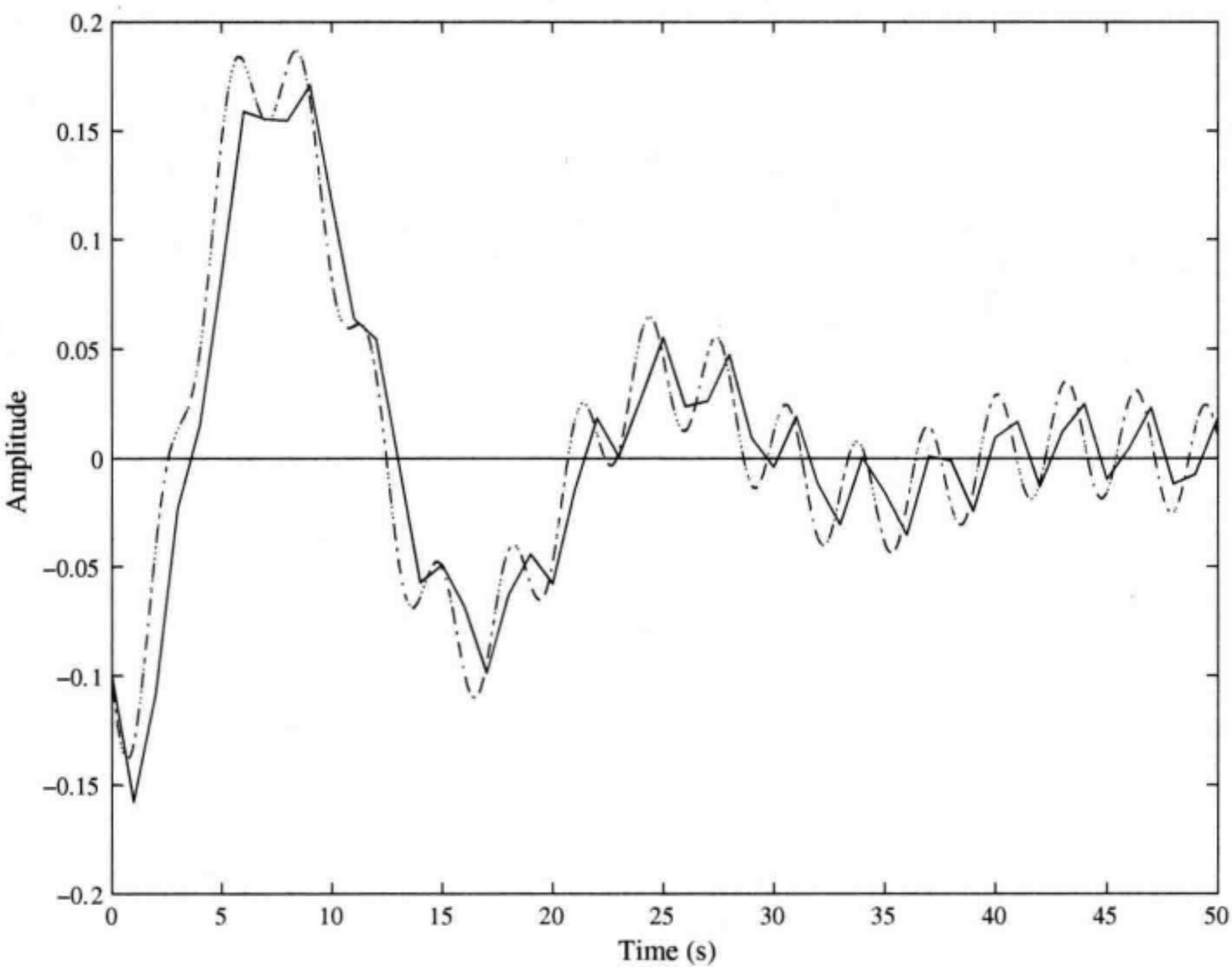


FIGURE 4.1 Outputs computed using $T_1 = 1$ (solid line), using $T_2 = 0.1$ (dotted line), and using $T_3 = 0.01$ (dash-and-dotted line).

then use `plot(t3, y3, '-.')` to plot the result in Fig. 4.1 using the dash-and-dotted line. It is indistinguishable from the one using $T_2 = 0.1$. Thus we conclude that the response of the state-space equation can be computed from its discretized equation if the step size is selected to be 0.1 or smaller. Note that the preceding three plot functions can be combined into one as in the second line from the bottom in Program 4.1. The combined plot function also plots the horizontal axis which is generated by `plot([0 50], [0 0])`. We save Program 4.1 as an m-file named `f41.m`. Typing in the command window `f41` will yield Fig. 4.1 in a figure window.

MATLAB contains the function `step(a, b, c, d)` and `impulse(a, b, c, d)`. The function `step` computes the output of a system, initially relaxed, excited by a step input and `impulse` computes the output excited by an impulse input. In using both functions, there is no need to specify the step size and time interval to be computed as in using `lsim`. Note that both `step` and `impulse` are based on `lsim`. However, they require adaptive selection of step sizes and automatically stop the computation when the responses hardly change.

It is not possible to generate an impulse in any computer. Then how is an impulse response generated? Consider

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t)$$

Its integration from $t = 0$ to $t = 0+$, where $0+$ is infinitesimally larger than 0, with $u(t) = \delta(t)$, yields

$$\mathbf{x}(0+) - \mathbf{x}(0) = \mathbf{A} \int_0^{0+} \mathbf{x}(t) dt + \mathbf{b} \int_0^{0+} \delta(t) dt = \mathbf{A} \cdot \mathbf{0} + \mathbf{b} \cdot 1 = \mathbf{b}$$

where the first integration is zero because $\mathbf{x}(t)$ contains no impulse and the second integration is 1. Thus the impulse input will transfer the initial state from $\mathbf{x}(0) = \mathbf{0}$ to $\mathbf{x}(0+) = \mathbf{b}$. Consequently, the impulse response (a zero-state response) can be generated as a zero-input response excited by the initial state \mathbf{b} . Indeed, this is the way MATLAB generates impulse responses. However, `impulse` generates a correct response only for a state-space equation with $d = 0$ or, more precisely, simply ignores $d \neq 0$.

MATLAB functions also apply to transfer functions such as `step(n, de)` and `impulse(n, de)`, where n and de denote the numerator's coefficients and denominator's coefficients of a transfer function. Internally, MATLAB first uses the function `tf2ss` to transform a transfer function into a state-space equation and then carries out computation. In Program 4.1, we used the state-space model to define the system as `dog=ss(a, b, c, d)`. We may also define the system as `dog=tf(n, de)` using the transfer-function model. Recall that transfer functions describe only zero-state responses. Therefore when we use the transfer-function model, the initial state is automatically assumed to be zero. If we use the transfer-function model in Program 4.1, then the result will be different because its initial state is different from zero. If the initial state is zero, there is no difference in using the `ss` model or `tf` model.

4.3.1 Real-Time Processing

Consider the response denoted by the dash-and-dotted line shown in Fig. 4.1. It is the output of the system excited by an input of 50 s long. The elapsed time of computing and plotting the output in Fig. 4.1 takes only 0.07 s, clocked by the MATLAB functions `tic` and `toc`. Thus the time scale in Fig. 4.1 has nothing to do with real time. This type of processing is called a non-real-time processing. All computer computation and simulation are non-real-time processing because time and sampling period do not play any role. Only orders of time sequences are important.

Telephone conversation is carried out in real time. So is audio CD. As soon as a CD player starts to spin, the music starts to appear; it does not wait until the player scans the entire CD. Digital processing of a CT signal always involves a sampling period. For example, the sampling period is $T = 1/8000 = 0.000125$ s or $125 \mu\text{s}$ for telephone transmission and $T = 1/44100 = 0.0000226$ s or $22.6 \mu\text{s}$ for audio CD. If the sampling period is larger than the amount of time needed to compute each output $y(kT)$, the specialized digital hardware will store the output in memory and then deliver it at time instant kT . Because the input $u(kT)$ and the output $y(kT)$ appear at the same real-time instant, it is called a real-time processing. The state-space equation is suited for such a processing. Indeed, consider the state-space equation in (4.18) and (4.19). If it is of dimension 3, then computing each $y(kT)$ requires no more than 20 additions and 20 multiplications. Suppose each addition requires 20 ns (20×10^{-9} s) and each multiplication requires 30 ns , then computing each $y(kT)$ requires at most $1 \mu\text{s}$ (10^{-6} s). If $d = 0$, and $\mathbf{x}(0) = \mathbf{0}$, then $y(0)$ is automatically 0 and can be delivered at the instant as

$u(0)$ arrives. Once $u(0)$ arrives, we can start computing $x(T)$ using (4.19) and then $y(T)$ using (4.18). If the sampling period is $22.6 \mu\text{s}$, once $y(T)$ is computed (it takes less than $1 \mu\text{s}$), it is stored in memory. It can be delivered at the same instant as $u(T)$ arrives. Once $u(T)$ arrives, we start to compute $x(2T)$ and then $y(2T)$ and store them in memory. The output $y(2T)$ can be delivered at the instant that $u(2T)$ arrives. Proceeding forward, the output $y(kT)$ can appear at the same instant as $u(kT)$. This is a real-time processing. If $d \neq 0$ and $\mathbf{x}(0) \neq \mathbf{0}$, the delivery of $y(kT)$ must be delayed by one sampling period T because it takes time to multiply $u(kT)$ by d and then to add the product to $\mathbf{c}\mathbf{x}(kT)$. Because T is generally very small, $y(kT)$ appears practically at the same instant as $u(kT)$. In conclusion, state-space equations can be used in computer computation and in real-time processing. It can also be used in op-amp circuit implementations as we discuss in the next subsection. Note that transfer functions cannot be used in real-time processing because we can compute the Laplace transform of $u(t)$ only after the entire $u(t)$ for all t in $[0, \infty)$ is available.

4.3.2 Op-Amp Circuit Implementation

Every linear time-invariant (LTI) state-space equation can be implemented using an operational amplifier (op-amp) circuit. Figure 4.2 shows two standard op-amp circuit elements. All inputs are connected, through resistors, to the inverting terminal. Not shown are the grounded noninverting terminal and power supplies. If the feedback branch is a resistor as shown in Fig. 4.2(a), then the output of the element is $-(ax_1 + bx_2 + cx_3)$. If the feedback branch is a capacitor with capacitance C and $RC = 1$ as shown in Fig. 4.2(b); and if the output is assigned as x , then $\dot{x} = -(av_1 + bv_2 + cv_3)$. We call the first element an *adder*; the second element, an *integrator*. Actually, the adder functions also as multipliers and the integrator functions also as multipliers and adder. If we use only one input, say x_1 , in Fig. 4.2(a), then the output equals $-ax_1$, and the element can be used as an *inverter* with gain a . Now we use an example to show that every LTI state-space equation can be implemented using the two types of elements in Fig. 4.2.

Consider the state-space equation

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 2 & -0.3 \\ 1 & -8 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} -2 \\ 0 \end{bmatrix} u(t) \quad (4.20)$$

$$y(t) = [-2 \ 3] \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + 5u(t) \quad (4.21)$$

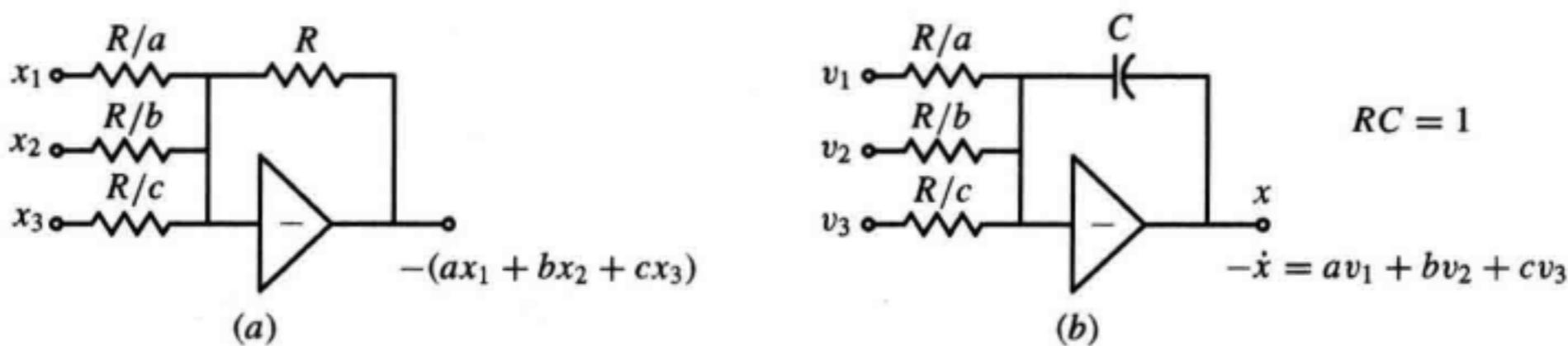


FIGURE 4.2 Two op-amp circuit elements.

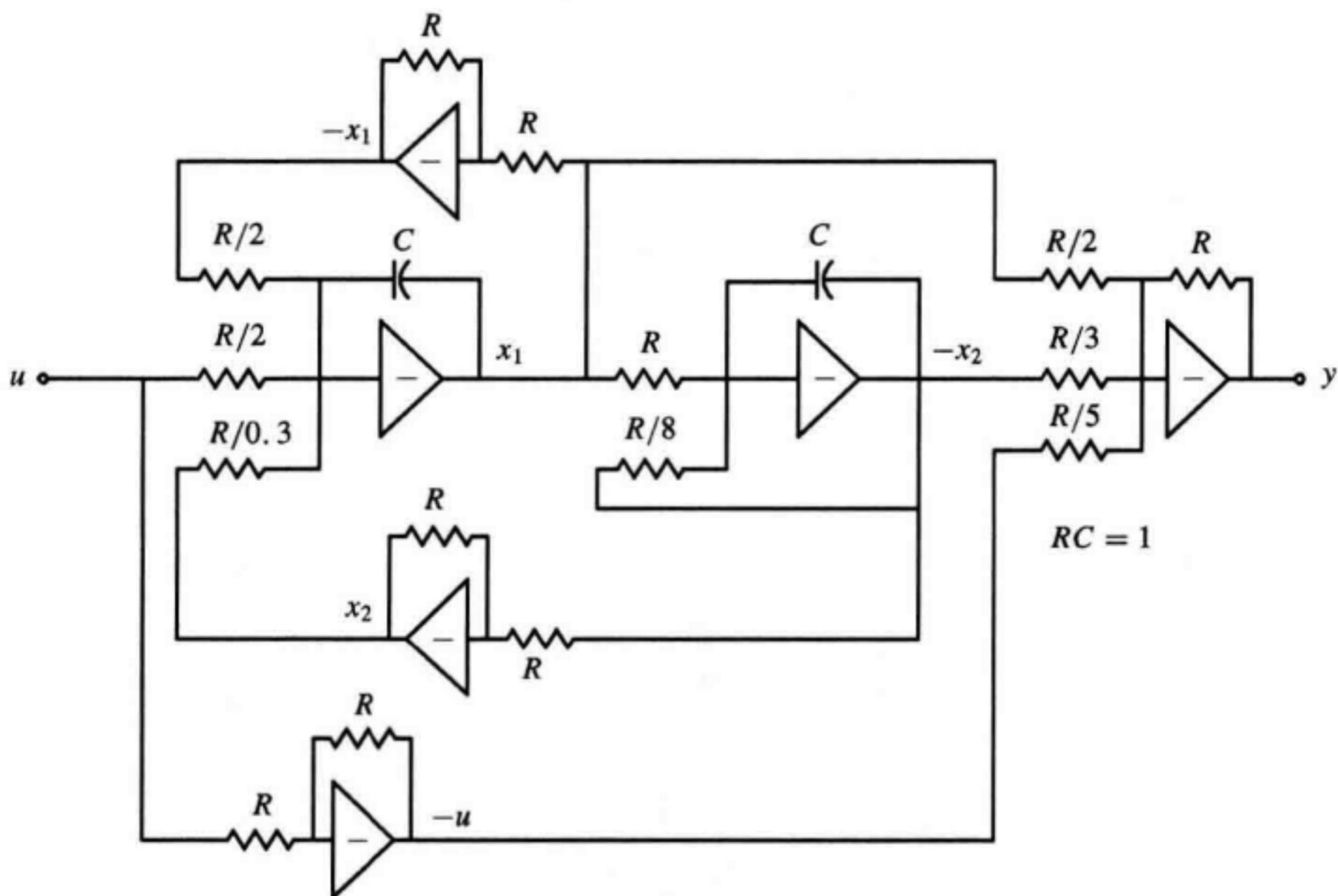


FIGURE 4.3 Op-amp implementation of (4.20) and (4.21).

It has dimension 2 and we need two integrators to implement it. We have the freedom in choosing the output of each integrator as $+x_i$ or $-x_i$. Suppose we assign the output of the left-hand-side (LHS) integrator as x_1 and assign the output of the right-hand-side (RHS) integrator as $-x_2$, as shown in Fig. 4.3. Then the input of the LHS integrator should be, from the first equation of (4.20), $-\dot{x}_1 = -2x_1 + 0.3x_2 + 2u$ and is connected as shown. The input of the RHS integrator should be $\dot{x}_2 = x_1 - 8x_2$ and is connected as shown. If the output of the adder is chosen as y , then its input should equal $-y = 2x_1 - 3x_2 - 5u$ and is connected as shown. Thus the state-space equation in (4.20) and (4.21) can be implemented as shown in Fig. 4.3. Note that there are many ways to implement the same equation. For example, if we assign the outputs of the two integrators in Fig. 4.3 as x_1 and x_2 , instead of x_1 and $-x_2$, then we will obtain a different implementation. See References 7 and 10.

In actual operational amplifier circuits, the range of signals is limited by the supplied voltages. If any signal grows outside the range, the circuit will saturate or burn out and the circuit will not behave as the equation does. There is, however, a way to deal with this problem, as we will discuss in a later section.

4.4 Equivalent State-Space Equations

The example that follows provides a motivation for studying equivalent state-space equations.

Example 4.4.1 Consider the circuit shown in Fig. 4.4. It consists of one capacitor, one inductor, one resistor, and one voltage source. First we select the inductor current x_1

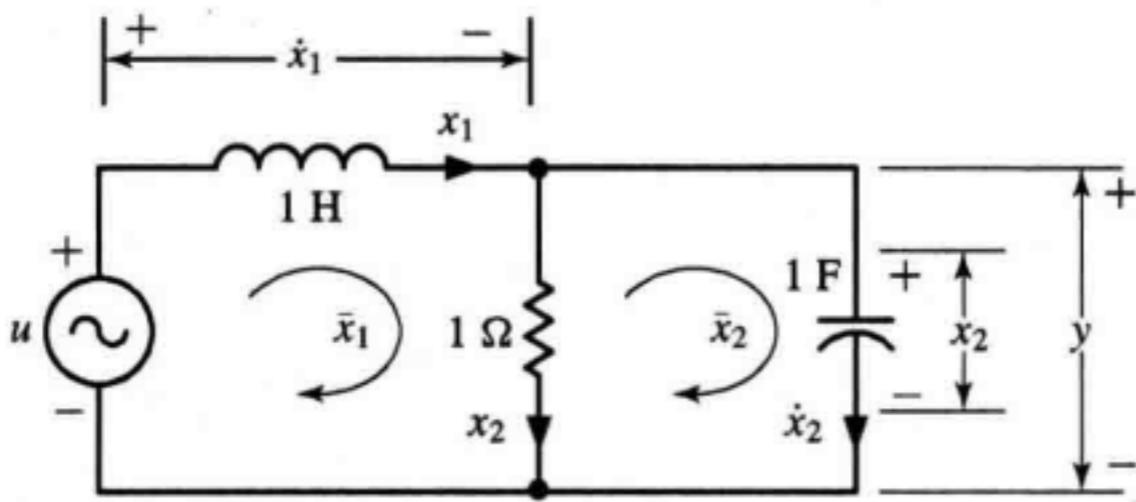


FIGURE 4.4 Circuit with two different sets of state variables.

and capacitor voltage x_2 as state variables as shown. The voltage across the inductor is \dot{x}_1 and the current through the capacitor is \dot{x}_2 . The voltage across the resistor is x_2 ; thus its current is $x_2/1 = x_2$. Clearly we have $x_1 = x_2 + \dot{x}_2$ and $\dot{x}_1 + x_2 - u = 0$. Thus the circuit is described by the following state-space equation:

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t) \quad (4.22)$$

$$y(t) = [0 \quad 1] \mathbf{x}(t)$$

If, instead, we select the loop currents \bar{x}_1 and \bar{x}_2 shown as state variables, then the voltage across the inductor is $\dot{\bar{x}}_1$ and the voltage across the resistor is $(\bar{x}_1 - \bar{x}_2) \cdot 1$. From the left-hand-side loop, we have

$$u = \dot{\bar{x}}_1 + \bar{x}_1 - \bar{x}_2 \quad \text{or} \quad \dot{\bar{x}}_1 = -\bar{x}_1 + \bar{x}_2 + u$$

The voltage across the capacitor is the same as the one across the resistor which is $\bar{x}_1 - \bar{x}_2$. Thus the current through the capacitor is $\dot{\bar{x}}_1 - \dot{\bar{x}}_2$, which equals \dot{x}_2 . Using $\dot{\bar{x}}_1 - \dot{\bar{x}}_2 = \dot{x}_2$ and the computed $\dot{\bar{x}}_1$, we have

$$\dot{\bar{x}}_2 = \dot{\bar{x}}_1 - \bar{x}_2 = -\bar{x}_1 + u$$

Thus the circuit is also described by the state-space equations

$$\begin{bmatrix} \dot{\bar{x}}_1(t) \\ \dot{\bar{x}}_2(t) \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \bar{x}_1(t) \\ \bar{x}_2(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u(t) \quad (4.23)$$

$$y(t) = [1 \quad -1] \bar{\mathbf{x}}(t)$$

The state-space equations in (4.22) and (4.23) describe the same circuit; therefore, they must be closely related. In fact, they are equivalent as will be established shortly.

Consider the n -dimensional state-space equation

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) \end{aligned} \quad (4.24)$$

where \mathbf{A} is an $n \times n$ constant matrix mapping an n -dimensional real space \mathcal{R}^n into itself. The state \mathbf{x} is a vector in \mathcal{R}^n for all t ; thus the real space is also called the state

space. The state-space equation in (4.24) can be considered to be associated with the orthonormal basis in (3.8). Now we study the effect on the equation by choosing a different set of basis.

Definition 4.1 Let \mathbf{P} be an $n \times n$ real nonsingular matrix and let $\bar{\mathbf{x}} = \mathbf{Px}$. Then the state-space equation

$$\begin{aligned}\dot{\bar{\mathbf{x}}}(t) &= \bar{\mathbf{A}}\bar{\mathbf{x}}(t) + \bar{\mathbf{B}}\mathbf{u}(t) \\ \mathbf{y}(t) &= \bar{\mathbf{C}}\bar{\mathbf{x}}(t) + \bar{\mathbf{D}}\mathbf{u}(t)\end{aligned}\quad (4.25)$$

where

$$\bar{\mathbf{A}} = \mathbf{PAP}^{-1} \quad \bar{\mathbf{B}} = \mathbf{PB} \quad \bar{\mathbf{C}} = \mathbf{CP}^{-1} \quad \bar{\mathbf{D}} = \mathbf{D} \quad (4.26)$$

is said to be (*algebraically*) *equivalent* to (4.24) and $\bar{\mathbf{x}} = \mathbf{Px}$ is called an *equivalence transformation*.

Equation (4.26) is obtained from (4.24) by substituting $\mathbf{x}(t) = \mathbf{P}^{-1}\bar{\mathbf{x}}(t)$ and $\dot{\mathbf{x}}(t) = \mathbf{P}^{-1}\dot{\bar{\mathbf{x}}}(t)$. In this substitution, we have changed, as in Equation (3.7), the basis vectors of the state space from the orthonormal basis to the columns of $\mathbf{P}^{-1} =: \mathbf{Q}$. Clearly \mathbf{A} and $\bar{\mathbf{A}}$ are similar and $\bar{\mathbf{A}}$ is simply a different representation of \mathbf{A} . To be precise, let $\mathbf{Q} = \mathbf{P}^{-1} = [\mathbf{q}_1 \ \mathbf{q}_2 \ \cdots \ \mathbf{q}_n]$. Then the i th column of $\bar{\mathbf{A}}$ is, as discussed in Section 3.4, the representation of $\mathbf{A}\mathbf{q}_i$ with respect to the basis $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$. From the equation $\bar{\mathbf{B}} = \mathbf{PB}$ or $\mathbf{B} = \mathbf{P}^{-1}\bar{\mathbf{B}} = [\mathbf{q}_1 \ \mathbf{q}_2 \ \cdots \ \mathbf{q}_n]\bar{\mathbf{B}}$, we see that the i th column of $\bar{\mathbf{B}}$ is the representation of the i th column of \mathbf{B} with respect to the basis $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$. The matrix $\bar{\mathbf{C}}$ is to be computed from \mathbf{CP}^{-1} . The matrix \mathbf{D} , called the *direct transmission part* between the input and output, has nothing to do with the state space and is not affected by the equivalence transformation.

We show that (4.24) and (4.25) have the same set of eigenvalues and the same transfer matrix. Indeed, we have, using $\det(\mathbf{P})\det(\mathbf{P}^{-1}) = 1$,

$$\begin{aligned}\bar{\Delta}(\lambda) &= \det(\lambda\mathbf{I} - \bar{\mathbf{A}}) = \det(\lambda\mathbf{PP}^{-1} - \mathbf{PAP}^{-1}) = \det[\mathbf{P}(\lambda\mathbf{I} - \mathbf{A})\mathbf{P}^{-1}] \\ &= \det(\mathbf{P})\det(\lambda\mathbf{I} - \mathbf{A})\det(\mathbf{P}^{-1}) = \det(\lambda\mathbf{I} - \mathbf{A}) = \Delta(\lambda)\end{aligned}$$

and

$$\begin{aligned}\hat{\mathbf{G}}(s) &= \bar{\mathbf{C}}(s\mathbf{I} - \bar{\mathbf{A}})^{-1}\bar{\mathbf{B}} + \bar{\mathbf{D}} = \mathbf{CP}^{-1}[\mathbf{P}(s\mathbf{I} - \mathbf{A})\mathbf{P}^{-1}]^{-1}\mathbf{PB} + \mathbf{D} \\ &= \mathbf{CP}^{-1}\mathbf{P}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{P}^{-1}\mathbf{PB} = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D} = \hat{\mathbf{G}}(s)\end{aligned}$$

Thus equivalent state-space equations have the same characteristic polynomial and, consequently, the same set of eigenvalues and same transfer matrix. In fact, all properties of (4.24) are preserved or invariant under any equivalence transformation.

Consider again the circuit shown in Fig. 4.4, which can be described by (4.22) and (4.23). We show that the two equations are equivalent. From Fig. 4.4, we have $x_1 = \bar{x}_1$. Because the voltage across the resistor is x_2 , its current is $x_2/1$ and equals

$\bar{x}_1 - \bar{x}_2$. Thus we have

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}$$

or

$$\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (4.27)$$

Note that for this \mathbf{P} , its inverse happens to equal itself. It is straightforward to verify that (4.22) and (4.23) are related by the equivalence transformation in (4.27).

The MATLAB function $[ab, bb, cb, db] = ss2ss(a, b, c, d, p)$ carries out equivalence transformations.

Two state-space equations are said to be *zero-state equivalent* if they have the same transfer matrix or

$$\mathbf{D} + \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} = \bar{\mathbf{D}} + \bar{\mathbf{C}}(s\mathbf{I} - \bar{\mathbf{A}})^{-1}\bar{\mathbf{B}}$$

This becomes, after substituting (3.57),

$$\begin{aligned} & \mathbf{D} + \mathbf{C}\mathbf{B}s^{-1} + \mathbf{C}\mathbf{A}\mathbf{B}s^{-2} + \mathbf{C}\mathbf{A}^2\mathbf{B}s^{-3} + \dots \\ &= \bar{\mathbf{D}} + \bar{\mathbf{C}}\bar{\mathbf{B}}s^{-1} + \bar{\mathbf{C}}\bar{\mathbf{A}}\bar{\mathbf{B}}s^{-2} + \bar{\mathbf{C}}\bar{\mathbf{A}}^2\bar{\mathbf{B}}s^{-3} + \dots \end{aligned}$$

Thus we have the theorem that follows.

► Theorem 4.1

Two linear time-invariant state-space equations $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ and $\{\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}}, \bar{\mathbf{D}}\}$ are zero-state equivalent or have the same transfer matrix if and only if $\mathbf{D} = \bar{\mathbf{D}}$ and

$$\mathbf{C}\mathbf{A}^m\mathbf{B} = \bar{\mathbf{C}}\bar{\mathbf{A}}^m\bar{\mathbf{B}}, \quad m = 0, 1, 2, \dots$$

It is clear that (algebraic) equivalence implies zero-state equivalence. In order for two state-space equations to be equivalent, they must have the same dimension. This is, however, not the case for zero-state equivalence as the next example shows.

Example 4.4.2 Consider the two circuits shown in Fig. 4.5. The one in Fig. 4.5(a) has two inductors and the one in Fig. 4.5(b) has only one inductor. We use impedances to compute their transfer functions. The impedance of the parallel connection of the $1-\Omega$ resistor and the $1-H$ inductor in Fig. 4.5(a) is $s/(s+1)$. Thus the transfer function from u in y in Fig. 4.5(a) is

$$\hat{g}_1(s) = \frac{\hat{y}(s)}{\hat{u}(s)} = \frac{\frac{s}{s+1}}{s + \frac{s}{s+1}} = \frac{s}{s^2 + 2s} = \frac{1}{s+2}$$

The transfer function from u to y in Fig. 4.5(b) is

$$\hat{g}_2(s) = \frac{\hat{y}(s)}{\hat{u}(s)} = \frac{1}{s+1+1} = \frac{1}{s+2}$$

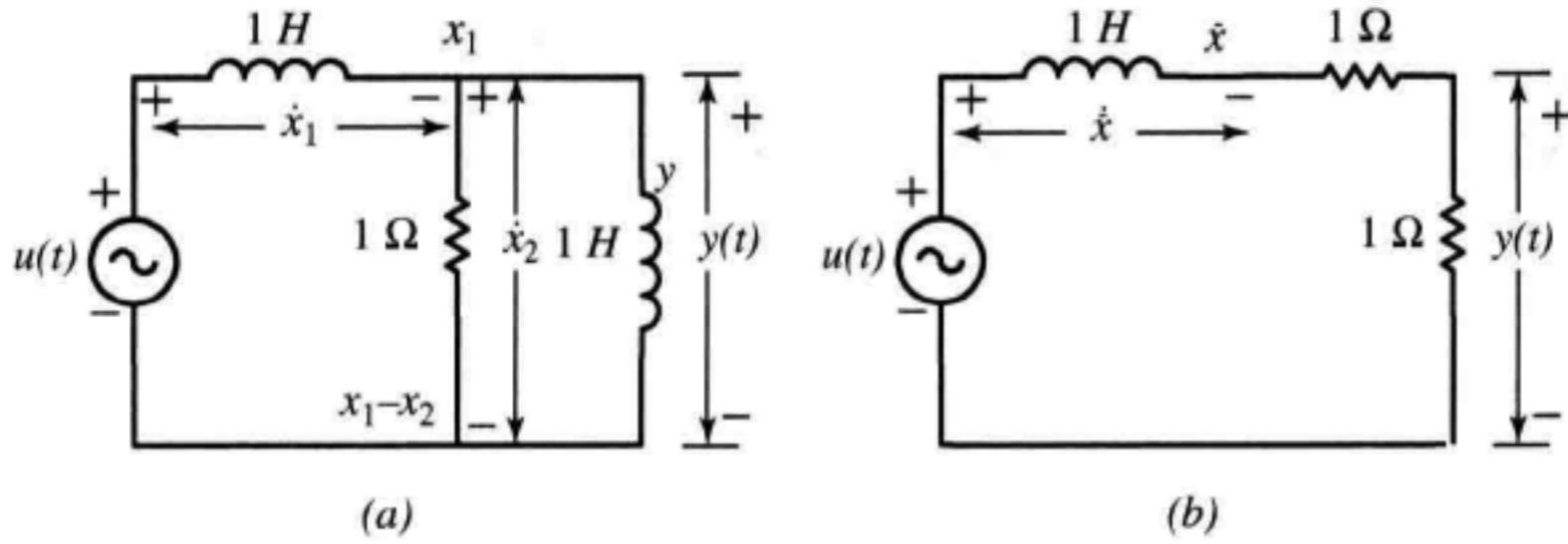


FIGURE 4.5 Two zero-state equivalent circuits.

which equals $\hat{g}_1(s)$. Thus the two circuits in Fig. 4.5 are zero-state equivalent.

To find a state-space equation to describe Fig. 4.5(a), we assign the current of the $1-H$ inductor in series with the voltage source as $x_1(t)$ and the current in the other inductor as $x_2(t)$. Then their voltages are, respectively, $\dot{x}_1(t)$ and $\dot{x}_2(t)$. The $1-\Omega$ resistor has current $x_1(t) - x_2(t)$ as shown and consequently has voltage $1 \times (x_1(t) - x_2(t))$. From the left-hand-side loop of Fig. 4.5(a), we have

$$\dot{x}_1(t) = u(t) - (x_1(t) - x_2(t)) = -x_1(t) + x_2(t) + u(t)$$

From the right-hand-side loop of Fig. 4.5(a), we have

$$\dot{x}_2(t) = x_1(t) - x_2(t)$$

and $y(t) = x_1(t) - x_2(t)$. They can be arranged in matrix form as

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t) =: \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t) \\ y(t) &= [1 \quad -1] \mathbf{x}(t) =: \mathbf{c}\mathbf{x}(t) + 0 \cdot u(t)\end{aligned}$$

This two-dimensional state-space equation describes the circuit in Fig. 4.5(a).

To find a state-space equation to describe the circuit in Fig. 4.5(b), we assign the current of the $1-H$ inductor as $\bar{x}(t)$. Then the voltage across the inductor is $\dot{\bar{x}}(t)$. The same current passes through the two $1-\Omega$ resistors, thus we have $u(t) = \dot{\bar{x}}(t) + \bar{x}(t) + \bar{x}(t)$ and $y(t) = \bar{x}(t)$ or

$$\begin{aligned}\dot{\bar{x}}(t) &= -2\bar{x}(t) + u(t) =: \bar{A}\bar{x}(t) + \bar{b}u(t) \\ y(t) &= \bar{x}(t) =: \bar{c}\bar{x}(t) + \bar{d}u(t)\end{aligned}$$

with $\bar{A} = -2$, $\bar{b} = 1$, $\bar{c} = 1$, $\bar{d} = 0$. This one-dimensional state-space equation describes the circuit in Fig. 4.5(b).

Next we use Theorem 4.1 to show that the two-dimensional and one-dimensional state-space equations are zero-state equivalent. For the one-dimensional state-space equation, we have $\bar{d} = 0$ and $\bar{c}\bar{A}^m\bar{b} = 1 \cdot (-2)^m \cdot 1 = (-2)^m$. For the two-dimensional state-space equation, we first use Theorem 3.5 to compute \mathbf{A}^m . The eigenvalues of \mathbf{A} is

$$\det(\lambda\mathbf{I} - \mathbf{A}) = \det \begin{bmatrix} \lambda + 1 & -1 \\ -1 & \lambda + 1 \end{bmatrix} = (\lambda + 1)^2 - 1 = \lambda^2 + 2\lambda = \lambda(\lambda + 2)$$

Thus \mathbf{A} has eigenvalues 0 and -2 . Let $f(\lambda) = \lambda^m$ and $h(\lambda) = \beta_0 + \beta_1\lambda$. On the spectrum of \mathbf{A} , we have $f(0) = 0 = h(0) = \beta_0$ and $f(-2) = (-2)^m = h(-2) = 0 + \beta_1(-2)$, which implies $\beta_1 = (-2)^{m-1}$. Thus we have

$$\mathbf{A}^m = \beta_0 \mathbf{I} + \beta_1 \mathbf{A} = (-2)^{m-1} \mathbf{A}$$

for any positive integer m . Consequently we have

$$\begin{aligned} \mathbf{c}\mathbf{A}^m\mathbf{b} &= [1 \ -1] \cdot (-2)^{m-1} \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = (-2)^{m-1}[1 \ -1] \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\ &= (-2)^{m-1}(-2) = (-2)^m \end{aligned}$$

We see that $d = \bar{d} = 0$ and $\mathbf{c}\mathbf{A}^m\mathbf{b} = \bar{c}\bar{\mathbf{A}}^m\bar{\mathbf{b}} = (-2)^m$ for $m = 0, 1, \dots$. Thus the two state-space equations are zero-state equivalent. Its practical implication will be discussed in Subsection 7.2.2.

4.4.1 Canonical Forms

MATLAB contains the function $[ab, bb, cb, db, P] = \text{canon}(a, b, c, d, \text{'type'})$. If $\text{type}=\text{companion}$, the function will generate an equivalent state-space equation with $\bar{\mathbf{A}}$ in the companion form in (3.24). This function works only if $\mathbf{Q} := [\mathbf{b}_1 \ \mathbf{Ab}_1 \ \cdots \ \mathbf{A}^{n-1}\mathbf{b}_1]$ is nonsingular, where \mathbf{b}_1 is the first column of \mathbf{B} . This condition is the same as $\{\mathbf{A}, \mathbf{b}_1\}$ controllable, as we will discuss in Chapter 6. The \mathbf{P} that the function `canon` generates equals \mathbf{Q}^{-1} . See the discussion in Section 3.4. The next example shows the use of `canon` and its practical implication.

Example 4.4.3 Consider the state-space equation in (4.20) and (4.21). In its op-amp circuit implementation, entry 0 is no connection and 1 is direct connection. Therefore the number of components needed in its implementation will be proportional to the total number of entries which differ from 0 or 1. The total number of such entries in (4.20) and (4.21) is 7.

Typing in MATLAB

```
a=[2 -0.3;1 -8];b=[-2;0];c=[1 -8];d=5;
[ab,bb,cb,db]=canon(a,b,c,d,'companion')
```

will yield

$$\begin{bmatrix} \dot{\bar{x}}_1(t) \\ \dot{\bar{x}}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 15.7 \\ 1 & -6 \end{bmatrix} \begin{bmatrix} \bar{x}_1(t) \\ \bar{x}_2(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t)$$

$$y(t) = [4 \ 2]\bar{\mathbf{x}}(t) + 5u(t)$$

This state-space equation is equivalent to the one in (4.20) and (4.21) and its A-matrix is of the companion form shown in (3.24). The equation has 5 entries which are different from 0 or 1. If we use the equation to carry out an op-amp circuit implementation, then the number of components used can be reduced by $(7 - 5)/7 = 28.5\%$.

We next discuss a different canonical form. Suppose \mathbf{A} has two real eigenvalues and two complex eigenvalues. Because \mathbf{A} has only real coefficients, the two complex

eigenvalues must be complex conjugate. Let $\lambda_1, \lambda_2, \alpha + j\beta$, and $\alpha - j\beta$ be the eigenvalues and let $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$, and \mathbf{q}_4 be the corresponding eigenvectors, where $\lambda_1, \lambda_2, \alpha, \beta, \mathbf{q}_1$, and \mathbf{q}_2 are all real and \mathbf{q}_4 equals the complex conjugate of \mathbf{q}_3 . Define $\mathbf{Q} = [\mathbf{q}_1 \mathbf{q}_2 \mathbf{q}_3 \mathbf{q}_4]$. Then we have

$$\mathbf{J} := \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \alpha + j\beta & 0 \\ 0 & 0 & 0 & \alpha - j\beta \end{bmatrix} = \mathbf{Q}^{-1} \mathbf{A} \mathbf{Q}$$

Note that \mathbf{Q} and \mathbf{J} can be obtained from $[\mathbf{q}, \mathbf{j}] = \text{eig}(\mathbf{a})$ in MATLAB as shown in Examples 3.5.1 and 3.5.2. This form is useless in practice but can be transformed into a real matrix by the following similarity transformation

$$\begin{aligned} \bar{\mathbf{Q}}^{-1} \mathbf{J} \bar{\mathbf{Q}} &:= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & j & -j \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \alpha + j\beta & 0 \\ 0 & 0 & 0 & \alpha - j\beta \end{bmatrix} \\ &\cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.5 & -0.5j \\ 0 & 0 & 0.5 & 0.5j \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \alpha & \beta \\ 0 & 0 & -\beta & \alpha \end{bmatrix} =: \bar{\mathbf{A}} \end{aligned}$$

We see that this transformation transforms the complex eigenvalues on the diagonal into a block with the real part of the eigenvalues on the diagonal and the imaginary part on the off-diagonal. This new \mathbf{A} -matrix is said to be in *modal* form. The MATLAB function $[\mathbf{ab}, \mathbf{bb}, \mathbf{cb}, \mathbf{db}, \mathbf{P}] = \text{canon}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \text{'modal'})$ or $\text{canon}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d})$ with no type specified will yield an equivalent state-space equation with $\bar{\mathbf{A}}$ in modal form. Note that there is no need to transform \mathbf{A} into a diagonal form and then to a modal form. The two transformations can be combined into one as

$$\begin{aligned} \mathbf{P}^{-1} &= \mathbf{Q} \bar{\mathbf{Q}} = [\mathbf{q}_1 \mathbf{q}_2 \mathbf{q}_3 \mathbf{q}_4] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.5 & -0.5j \\ 0 & 0 & 0.5 & 0.5j \end{bmatrix} \\ &= [\mathbf{q}_1 \mathbf{q}_2 \text{Re}(\mathbf{q}_3) \text{Im}(\mathbf{q}_3)] \end{aligned}$$

where Re and Im stand, respectively, for the real part and imaginary part and we have used in the last equality the fact that \mathbf{q}_4 is the complex conjugate of \mathbf{q}_3 . We give one more example. The modal form of a matrix with real eigenvalue λ_1 and two pairs of distinct complex conjugate eigenvalues $\alpha_i \pm j\beta_i$, for $i = 1, 2$ is

$$\bar{\mathbf{A}} = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \alpha_1 & \beta_1 & 0 & 0 \\ 0 & -\beta_1 & \alpha_1 & 0 & 0 \\ 0 & 0 & 0 & \alpha_2 & \beta_2 \\ 0 & 0 & 0 & -\beta_2 & \alpha_2 \end{bmatrix} \quad (4.28)$$

It is block diagonal and can be obtained by the similarity transformation

$$\mathbf{P}^{-1} = [\mathbf{q}_1 \operatorname{Re}(\mathbf{q}_2) \operatorname{Im}(\mathbf{q}_2) \operatorname{Re}(\mathbf{q}_4) \operatorname{Im}(\mathbf{q}_4)]$$

where \mathbf{q}_1 , \mathbf{q}_2 , and \mathbf{q}_4 are, respectively, eigenvectors associated with λ_1 , $\alpha_1 + j\beta_1$, and $\alpha_2 + j\beta_2$. This form is useful in state-space design and op-amp circuit implementation.

4.4.2 Magnitude Scaling in Op-Amp Circuits

As discussed in Subsection 4.3.1, every LTI state-space equation can be implemented using an op-amp circuit.³ In actual op-amp circuits, all signals are limited by power supplies. If we use ± 15 -volt power supplies, then all signals are roughly limited to ± 13 volts. If any signal goes outside the range, the circuit will saturate and will not behave according to the state-space equation. Therefore, saturation is an important issue in actual op-amp circuit implementation. We first use an example to demonstrate this point.

Example 4.4.4 Consider the state-space equation

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \begin{bmatrix} -0.1 & 2 \\ 0 & -1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 10 \\ 0.1 \end{bmatrix} u(t) \\ y(t) &= [0.2 \quad -1] \mathbf{x}(t)\end{aligned}\tag{4.29}$$

Suppose the input is a step function of various amplitude and the equation is to be implemented using an op-amp circuit in which all signals must be limited to ± 10 . First we use MATLAB to find its unit step response. Typing in MATLAB

```
a=[-0.1 2; 0 -1]; b=[10; 0.1]; c=[0.2 -1]; d=0;
[y, x, t]=step(a, b, c, d);
plot(t, y, t, x)
```

yields the plot in Fig. 4.6(a). We see that the largest magnitude of all variables is 100. Thus if we apply a step input with amplitude 1, the circuit will saturate and cannot function as designed. Because the equation is linear, if we decrease the amplitude of the step input by a factor, all responses will decrease by the same factor. In other words, if we apply a step input with amplitude 0.1, then the largest magnitude of all variables will be 10 and the circuit will not saturate. In conclusion, the op-amp circuit built from (4.29) can function as the equation indicates only if the applied step input has magnitude 0.1 or less.

It is generally true that every practical system can function as linear only for a limited input range. For the system described by (4.29), we now use an equivalence transformation to expand the admissible input range. Any equivalence transformation will not affect the relationship between the input $u(t)$ and the output $y(t)$, but we can select an equivalence transformation to achieve

$$|x_i(t)| \leq |y|_{max}$$

³This subsection may be skipped without loss of continuity.

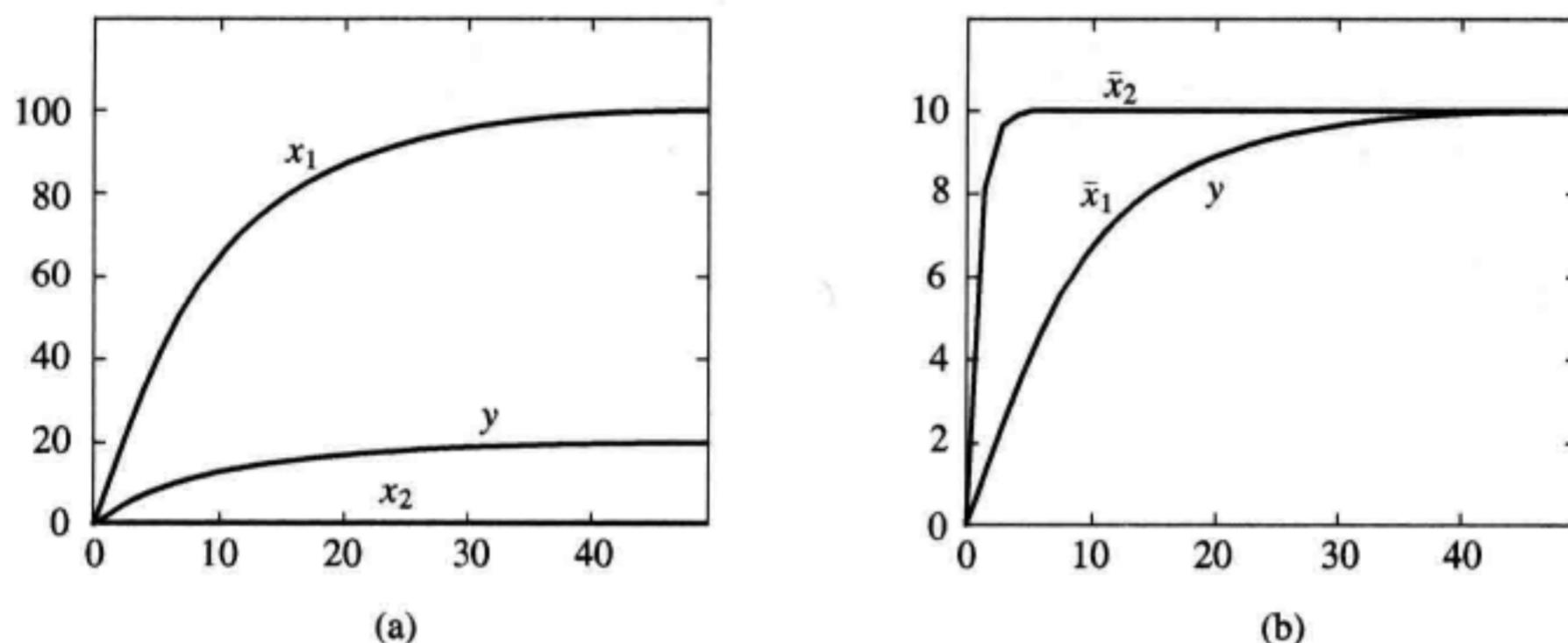


FIGURE 4.6 Time responses.

for all i and for all t . Under this assumption, if the output does not saturate, then all state variables will not saturate. From the response shown in Fig. 4.6(a), we see that if the applied step input has magnitude 0.5 or less, then $|y|_{max} = 10$ and the output will not saturate.

We now find an equivalence transformation to achieve $|x_i(t)| \leq |y|_{max}$. From Fig. 4.6(a), we have $|x_1|_{max} = 100$, $|y|_{max} = 20$, and $|x_2| \ll |y|_{max}$. The state variable x_2 is hardly visible and its largest magnitude is found to be 0.1 by plotting it separatively (not shown). Let us select

$$\bar{x}_1 = \frac{20}{100}x_1 = 0.2x_1, \quad \bar{x}_2 = \frac{20}{0.1}x_2 = 200x_2$$

With this transformation, the maximum magnitudes of $\bar{x}_1(t)$ and $\bar{x}_2(t)$ will equal $|y|_{max}$. Note that the selection is not unique; we may select $\bar{x}_1 = 0.1x_1$ and $\bar{x}_2 = x_2$ or others. The transformation can be expressed as $\bar{\mathbf{x}} = \mathbf{P}\mathbf{x}$ with

$$\mathbf{P} = \begin{bmatrix} 0.2 & 0 \\ 0 & 200 \end{bmatrix}, \quad \mathbf{P}^{-1} = \begin{bmatrix} 5 & 0 \\ 0 & 0.005 \end{bmatrix}$$

Then its equivalent state-space equation can be readily computed from (4.29) as

$$\begin{aligned} \dot{\bar{\mathbf{x}}}(t) &= \begin{bmatrix} -0.1 & 0.002 \\ 0 & -1 \end{bmatrix} \bar{\mathbf{x}}(t) + \begin{bmatrix} 2 \\ 20 \end{bmatrix} u(t) \\ y(t) &= [1 \quad -0.005] \bar{\mathbf{x}}(t) \end{aligned}$$

Its responses excited by $u(t) = 0.5$ for $t \geq 0$ are plotted in Fig. 4.6(b). We see that all signals lie inside the range ± 10 and occupy the full scale. Thus the op-amp circuit based on the equivalence equation will function properly if the applied step input has a magnitude 0.5 or less. This is a big improvement over the op-amp circuit based on (4.29), where the step input is required to have magnitude 0.1 or less. Although we discuss only step inputs, the idea is applicable to any input but will be more complicated.

Analog computers are essentially op-amp circuits. In using an analog computer, the most difficult part is to avoid saturation. The process of adjusting parameters to avoid saturation is called *magnitude scaling* and is carried out by trial and error. Using

a digital computer and an equivalence transformation, we can find, as illustrated above, a state-space equation with largest permissible input range. This is a practical application of similarity or equivalence transformations.

4.5 Realizations

Every SISO linear time-invariant (LTI) system can be described by the input–output description

$$\hat{y}(s) = \hat{g}(s)\hat{u}(s)$$

and, if the system is lumped as well, by the state-space equation description

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t) \\ y(t) &= \mathbf{c}\mathbf{x}(t) + du(t)\end{aligned}$$

If the state-space equation is known, the transfer matrix can be computed as $\hat{g}(s) = \mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} + d$. The computed transfer function is unique. Now we study the converse problem, that is, to find a state-space equation from a given transfer function. This is called the *realization* problem. This terminology is justified by the fact that by using the state-space equation, we can build an op-amp circuit for the transfer function.

A transfer function $\hat{g}(s)$ is said to be *realizable* if there exists a finite-dimensional state-space equation or, simply, $\{\mathbf{A}, \mathbf{b}, \mathbf{c}, d\}$ such that

$$\hat{g}(s) = \mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} + d$$

and $\{\mathbf{A}, \mathbf{b}, \mathbf{c}, d\}$ is called a *realization* of $\hat{g}(s)$. An LTI distributed system can be described by an transfer function, but not by a finite-dimensional state equation. Thus not every $\hat{g}(s)$ is realizable. If $\hat{g}(s)$ is realizable, then it has infinitely many realizations, not necessarily of the same dimension. Thus the realization problem is fairly complex. We study here only the realizability condition. The other issues will be studied in later chapters.

► Theorem 4.2

A transfer function $\hat{g}(s)$ is realizable if and only if $\hat{g}(s)$ is a proper rational function.

We use (3.19) to write

$$\hat{g}_{sp}(s) := \mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} = \frac{1}{\det(s\mathbf{I} - \mathbf{A})}\mathbf{c}[\text{Adj}(s\mathbf{I} - \mathbf{A})]\mathbf{b} \quad (4.30)$$

If \mathbf{A} is $n \times n$, then $\det(s\mathbf{I} - \mathbf{A})$ has degree n . Every entry of $\text{Adj}(s\mathbf{I} - \mathbf{A})$ is the determinant of an $(n - 1) \times (n - 1)$ submatrix of $(s\mathbf{I} - \mathbf{A})$; thus it has at most degree $(n - 1)$. Their linear combinations again have at most degree $(n - 1)$. Thus we conclude that $\mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}$ is a strictly proper rational function. If d is a nonzero constant, then $\mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} + d$ is proper. This shows that if $\hat{g}(s)$ is realizable, then it is a proper rational function. Note that we have

$$\hat{g}(\infty) = d$$

Next we show the converse: that is, if $\hat{g}(s)$ is a proper rational function, then there exists a realization. We use the following proper rational transfer function to verify the assertion:

$$\hat{g}(s) = \frac{\bar{N}(s)}{\bar{D}(s)} = \frac{\bar{b}_1 s^4 + \bar{b}_2 s^3 + \bar{b}_3 s^2 + \bar{b}_4 s + \bar{b}_5}{\bar{a}_1 s^4 + \bar{a}_2 s^3 + \bar{a}_3 s^2 + \bar{a}_4 s + \bar{a}_5} \quad (4.31)$$

with $\bar{a}_1 \neq 0$. The realization procedure, however, is applicable to any proper rational transfer function.

The first step in realization is to write (4.31) as

$$\hat{g}(s) = \frac{b_1 s^3 + b_2 s^2 + b_3 s + b_4}{s^4 + a_2 s^3 + a_3 s^2 + a_4 s + a_5} + d =: \frac{N(s)}{D(s)} + d \quad (4.32)$$

where $D(s) := s^4 + a_2 s^3 + a_3 s^2 + a_4 s + a_5$ with $a_1 = 1$. Note that $N(s)/D(s)$ is strictly proper and $D(s)$ is monic. This can be achieved by dividing the numerator and denominator of (4.31) by \bar{a}_1 and then carrying out a direct division. Now we claim that the following state-space equation realizes (4.32) or, equivalently, (4.31):

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \begin{bmatrix} -a_2 & -a_3 & -a_4 & -a_5 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} u(t) \\ y(t) &= [b_1 \ b_2 \ b_3 \ b_4] \mathbf{x}(t) + du(t) \end{aligned} \quad (4.33)$$

with $\mathbf{x}(t) = [x_1(t) \ x_2(t) \ x_3(t) \ x_4(t)]'$. The number of state variables equals the degree of the denominator of $\hat{g}(s)$. This state-space equation can be obtained directly from the coefficients in (4.32). We place the denominator's coefficients, except for its leading coefficient 1, with sign reversed in the first row of \mathbf{A} , and we place the numerator's coefficients, without changing sign, directly as \mathbf{c} . The constant d in (4.33) is the constant term in (4.32). The rest of the state-space equation have fixed patterns. The second row of \mathbf{A} is $[1 \ 0 \ 0 \dots]$. The third row of \mathbf{A} is $[0 \ 1 \ 0 \dots]$ and so forth. The column vector \mathbf{b} is all zero except for its first entry, which is 1.

To show that (4.33) is a realization of (4.32), we must compute its transfer function. We first write the matrix equation explicitly as

$$\begin{aligned} \dot{x}_1(t) &= -a_2 x_1(t) - a_3 x_2(t) - a_4 x_3(t) - a_5 x_4(t) + u(t) \\ \dot{x}_2(t) &= x_1(t) \\ \dot{x}_3(t) &= x_2(t) \\ \dot{x}_4(t) &= x_3(t) \end{aligned} \quad (4.34)$$

We see that the four-dimensional state equation in (4.33) actually consists of four first-order differential equations as shown in (4.34). Applying the Laplace transform and

assuming zero initial conditions yield

$$\begin{aligned}s\hat{x}_1(s) &= -a_2\hat{x}_1(s) - a_3\hat{x}_2(s) - a_4\hat{x}_3(s) - a_5\hat{x}_4(s) + \hat{u}(s) \\ s\hat{x}_2(s) &= \hat{x}_1(s) \\ s\hat{x}_3(s) &= \hat{x}_2(s) \\ s\hat{x}_4(s) &= \hat{x}_3(s)\end{aligned}$$

From the second to the last equations, we can obtain

$$\hat{x}_2(s) = \frac{\hat{x}_1(s)}{s}, \quad \hat{x}_3(s) = \frac{\hat{x}_2(s)}{s} = \frac{\hat{x}_1(s)}{s^2}, \quad \hat{x}_4(s) = \frac{\hat{x}_1(s)}{s^3} \quad (4.35)$$

Substituting these into the first equation yields

$$\left[s + a_2 + \frac{a_3}{s} + \frac{a_4}{s^2} + \frac{a_5}{s^3} \right] \hat{x}_1(s) = \hat{u}(s)$$

or

$$\left[\frac{s^4 + a_2s^3 + a_3s^2 + a_4s + a_5}{s^3} \right] \hat{x}_1(s) = \hat{u}(s)$$

which implies

$$\hat{x}_1(s) = \frac{s^3}{s^4 + a_2s^3 + a_3s^2 + a_4s + a_5} U(s) =: \frac{s^3}{D(s)} \hat{u}(s) \quad (4.36)$$

Substituting (4.35) and (4.36) into the following Laplace transform of the output equation in (4.33) yields

$$\begin{aligned}\hat{y}(s) &= b_1\hat{x}_1(s) + b_2\hat{x}_2(s) + b_3\hat{x}_3(s) + b_4\hat{x}_4(s) + d\hat{u}(s) \\ &= \left[\frac{b_1s^3}{D(s)} + \frac{b_2s^3}{D(s)s} + \frac{b_3s^3}{D(s)s^2} + \frac{b_4s^3}{D(s)s^3} \right] \hat{u}(s) + d\hat{u}(s) \\ &= \left[\frac{b_1s^3 + b_2s^2 + b_3s + b_4}{D(s)} + d \right] \hat{u}(s)\end{aligned}$$

This shows that the transfer function of (4.33) equals (4.32). Thus (4.33) is a realization of (4.32) or (4.31). The state-space equation in (4.33) is said to be in *controllable form* for the reason to be given in Chapter 7. It will also be developed again using a different method there.

For matrices \mathbf{P} and \mathbf{Q} and a scalar d , we have $(\mathbf{PQ})' = \mathbf{Q}'\mathbf{P}'$ and $d' = d$, where the prime denotes the transpose. Applying the transpose to $\hat{g}(s) = \mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} + d$ yields

$$\hat{g}(s) = [\hat{g}(s)]' = [\mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} + d]' = \mathbf{b}'(s'\mathbf{I}' - \mathbf{A}')^{-1}\mathbf{c}' + d' = \mathbf{b}'(s\mathbf{I} - \mathbf{A}')^{-1}\mathbf{c}' + d$$

Thus $\{\mathbf{A}', \mathbf{c}', \mathbf{b}', d\}$ is a different realization of $\hat{g}(s)$. From (4.33), we obtain a different realization of (4.32) as

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \begin{bmatrix} -a_2 & 1 & 0 & 0 \\ -a_3 & 0 & 1 & 0 \\ -a_4 & 0 & 0 & 1 \\ -a_5 & 0 & 0 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} u(t) \\ y(t) &= [1 \ 0 \ 0 \ 0] \mathbf{x}(t) + du(t)\end{aligned}\quad (4.37)$$

This is called the *observable-form* realization.

Before proceeding, we mention that if $\hat{g}(s)$ is strictly proper, then no direct division is needed and we have $d = 0$. In other words, there is no direct transmission part from u to y .

Example 4.5.1 Consider the transfer function

$$\hat{g}(s) = \frac{3s^4 + 5s^3 + 24s^2 + 23s - 5}{2s^4 + 6s^3 + 15s^2 + 12s + 5} \quad (4.38)$$

We first divide the numerator and denominator by 2 to yield

$$\hat{g}(s) = \frac{1.5s^4 + 2.5s^3 + 12s^2 + 11.5s - 2.5}{s^4 + 3s^3 + 7.5s^2 + 6s + 2.5}$$

which can be written as, using direct division,

$$\hat{g}(s) = \frac{-2s^3 + 0.75s^2 + 2.5s - 6.25}{s^4 + 3s^3 + 7.5s^2 + 6s + 2.5} + 1.5$$

Thus its controllable-form realization is, using (4.33),

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \begin{bmatrix} -3 & -7.5 & -6 & -2.5 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} u(t) \\ y(t) &= [-2 \ 0.75 \ 2.5 \ -6.25] \mathbf{x}(t) + 1.5u(t)\end{aligned}$$

and its observable-form realization is, using (4.37)

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \begin{bmatrix} -3 & 1 & 0 & 0 \\ -7.5 & 0 & 1 & 0 \\ -6 & 0 & 0 & 1 \\ -2.5 & 0 & 0 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} -2 \\ 0.75 \\ 2.5 \\ -6.25 \end{bmatrix} u(t) \\ y(t) &= [1 \ 0 \ 0 \ 0] \mathbf{x}(t) + 1.5u(t)\end{aligned}$$

We see that the realizations can be read out from the coefficients of the transfer function.

Example 4.5.2 Consider the circuit in Fig. 2.8(a). Its transfer function was computed in (2.37) as

$$\hat{g}(s) = \frac{4s + 3}{40s^3 + 30s^2 + 9s + 3}$$

We first normalize its denominator's leading coefficient to 1 to yield

$$H(s) = \frac{0.1s + 0.075}{s^3 + 0.75s^2 + 0.225s + 0.075}$$

It is strictly proper and $d = 0$. Thus its controllable-form realization is

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \begin{bmatrix} -0.75 & -0.225 & -0.075 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u(t) \\ \mathbf{y}(t) &= [0 \ 0.1 \ 0.075] \mathbf{x}(t) + 0 \cdot u(t)\end{aligned}$$

This is actually the state-space equation in (2.39) obtained in using the MATLAB function `tf2ss`. Although it differs from the state-space equation description of the circuit developed in (2.26), the two equations are equivalent.

4.5.1 Multi-input Multi-output Case

A transfer matrix $\hat{\mathbf{G}}(s)$ is called realizable if there exists a state-space equation

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t)\end{aligned}$$

which has $\hat{\mathbf{G}}(s)$ as its transfer matrix. As in the SISO case, $\hat{\mathbf{G}}(s)$ is realizable if and only if $\hat{\mathbf{G}}(s)$ is a proper rational matrix. If $\hat{\mathbf{G}}(s)$ has a realization, then using the matrix version of (4.30) and the same argument, we can show that every entry of $\hat{\mathbf{G}}(s)$ is a proper rational function. Now given a $q \times p$ proper rational transfer matrix $\hat{\mathbf{G}}(s)$, we discuss a number of ways of obtaining realizations.

Let us decompose $\hat{\mathbf{G}}(s)$ as

$$\hat{\mathbf{G}}(s) = \hat{\mathbf{G}}(\infty) + \hat{\mathbf{G}}_{sp}(s) \quad (4.39)$$

where $\hat{\mathbf{G}}_{sp}$ is the strictly proper part of $\hat{\mathbf{G}}(s)$. Let

$$d(s) = s^r + \alpha_1 s^{r-1} + \cdots + \alpha_{r-1} s + \alpha_r \quad (4.40)$$

be the least common denominator of all entries of $\hat{\mathbf{G}}_{sp}(s)$. Here we require $d(s)$ to be monic, that is, its leading coefficient is 1. Then $\hat{\mathbf{G}}_{sp}(s)$ can be expressed as

$$\hat{\mathbf{G}}_{sp}(s) = \frac{1}{d(s)} [\mathbf{N}(s)] = \frac{1}{d(s)} [\mathbf{N}_1 s^{r-1} + \mathbf{N}_2 s^{r-2} + \cdots + \mathbf{N}_{r-1} s + \mathbf{N}_r] \quad (4.41)$$

where \mathbf{N}_i are $q \times p$ constant matrices. Now we claim that the equation

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -\alpha_1 \mathbf{I}_p & -\alpha_2 \mathbf{I}_p & \cdots & -\alpha_{r-1} \mathbf{I}_p & -\alpha_r \mathbf{I}_p \\ \mathbf{I}_p & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_p & \mathbf{0} \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} \mathbf{I}_p \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \mathbf{u}(t) \quad (4.42)$$

$$\mathbf{y}(t) = [\mathbf{N}_1 \ \mathbf{N}_2 \ \cdots \ \mathbf{N}_{r-1} \ \mathbf{N}_r] \mathbf{x}(t) + \hat{\mathbf{G}}(\infty) \mathbf{u}(t)$$

is a realization of $\hat{\mathbf{G}}(s)$. The matrix \mathbf{I}_p is the $p \times p$ unit matrix and every $\mathbf{0}$ is a $p \times p$ zero matrix. The A-matrix is said to be in block companion form; it consists of r rows and r columns of $p \times p$ matrices; thus the A-matrix has order $rp \times rp$. The B-matrix has order $rp \times p$. Because the C-matrix consists of r number of \mathbf{N}_i , each of order $q \times p$, the C-matrix has order $q \times rp$. The realization has dimension rp and is said to be in *block controllable form*.

We show that (4.42) is a realization of $\hat{\mathbf{G}}(s)$ in (4.41). Let us define

$$\mathbf{Z} := \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_r \end{bmatrix} := (s\mathbf{I} - \mathbf{A})^{-1} \mathbf{B} \quad (4.43)$$

where \mathbf{Z}_i is $p \times p$ and \mathbf{Z} is $rp \times p$. Then the transfer matrix of (4.42) equals

$$\mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1} \mathbf{B} + \hat{\mathbf{G}}(\infty) = \mathbf{N}_1 \mathbf{Z}_1 + \mathbf{N}_2 \mathbf{Z}_2 + \cdots + \mathbf{N}_r \mathbf{Z}_r + \hat{\mathbf{G}}(\infty) \quad (4.44)$$

We write (4.43) as $(s\mathbf{I} - \mathbf{A})\mathbf{Z} = \mathbf{B}$ or

$$s\mathbf{Z} = \mathbf{A}\mathbf{Z} + \mathbf{B} \quad (4.45)$$

Using the shifting property of the companion form of \mathbf{A} , from the second to the last block equations of (4.45), we can readily obtain

$$s\mathbf{Z}_2 = \mathbf{Z}_1, \quad s\mathbf{Z}_3 = \mathbf{Z}_2, \quad \dots, \quad s\mathbf{Z}_r = \mathbf{Z}_{r-1}$$

which implies

$$\mathbf{Z}_2 = \frac{1}{s} \mathbf{Z}_1, \quad \mathbf{Z}_3 = \frac{1}{s^2} \mathbf{Z}_1, \quad \dots, \quad \mathbf{Z}_r = \frac{1}{s^{r-1}} \mathbf{Z}_1$$

Substituting these into the first block equation of (4.45) yields

$$\begin{aligned} s\mathbf{Z}_1 &= -\alpha_1 \mathbf{Z}_1 - \alpha_2 \mathbf{Z}_2 - \cdots - \alpha_r \mathbf{Z}_r + \mathbf{I}_p \\ &= -\left(\alpha_1 + \frac{\alpha_2}{s} + \cdots + \frac{\alpha_r}{s^{r-1}}\right) \mathbf{Z}_1 + \mathbf{I}_p \end{aligned}$$

or, using (4.40),

$$\left(s + \alpha_1 + \frac{\alpha_2}{s} + \cdots + \frac{\alpha_r}{s^{r-1}}\right) \mathbf{Z}_1 = \frac{d(s)}{s^{r-1}} \mathbf{Z}_1 = \mathbf{I}_p$$

Thus we have

$$\mathbf{Z}_1 = \frac{s^{r-1}}{d(s)} \mathbf{I}_p, \quad \mathbf{Z}_2 = \frac{s^{r-2}}{d(s)} \mathbf{I}_p, \quad \dots, \quad \mathbf{Z}_r = \frac{1}{d(s)} \mathbf{I}_p$$

Substituting these into (4.44) yields

$$\mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \hat{\mathbf{G}}(\infty) = \frac{1}{d(s)}[\mathbf{N}_1 s^{r-1} + \mathbf{N}_2 s^{r-2} + \dots + \mathbf{N}_r] + \hat{\mathbf{G}}(\infty)$$

This equals $\hat{\mathbf{G}}(s)$ in (4.39) and (4.41). This shows that (4.42) is a realization of $\hat{\mathbf{G}}(s)$.

Example 4.5.3 Consider the proper rational matrix

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} \frac{4s-10}{2s+1} & \frac{3}{s+2} \\ \frac{1}{(2s+1)(s+2)} & \frac{s+1}{(s+2)^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \frac{-12}{2s+1} & \frac{3}{s+2} \\ \frac{1}{(2s+1)(s+2)} & \frac{s+1}{(s+2)^2} \end{bmatrix} \quad (4.46)$$

where we have decomposed $\hat{\mathbf{G}}(s)$ into the sum of a constant matrix and a strictly proper rational matrix $\hat{\mathbf{G}}_{sp}(s)$. The monic least common denominator of $\hat{\mathbf{G}}_{sp}(s)$ is $d(s) = (s + 0.5)(s + 2)^2 = s^3 + 4.5s^2 + 6s + 2$. Thus we have

$$\begin{aligned} \hat{\mathbf{G}}_{sp}(s) &= \frac{1}{s^3 + 4.5s^2 + 6s + 2} \begin{bmatrix} -6(s+2)^2 & 3(s+2)(s+0.5) \\ 0.5(s+2) & (s+1)(s+0.5) \end{bmatrix} \\ &= \frac{1}{d(s)} \left(\begin{bmatrix} -6 & 3 \\ 0 & 1 \end{bmatrix} s^2 + \begin{bmatrix} -24 & 7.5 \\ 0.5 & 1.5 \end{bmatrix} s + \begin{bmatrix} -24 & 3 \\ 1 & 0.5 \end{bmatrix} \right) \end{aligned}$$

and a realization of (4.46) is

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \begin{bmatrix} -4.5 & 0 & \vdots & -6 & 0 & \vdots & -2 & 0 \\ 0 & -4.5 & \vdots & 0 & -6 & \vdots & 0 & -2 \\ \dots & \dots \\ 1 & 0 & \vdots & 0 & 0 & \vdots & 0 & 0 \\ 0 & 1 & \vdots & 0 & 0 & \vdots & 0 & 0 \\ \dots & \dots \\ 0 & 0 & \vdots & 1 & 0 & \vdots & 0 & 0 \\ 0 & 0 & \vdots & 0 & 1 & \vdots & 0 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \dots & \dots \\ 0 & 0 \\ 0 & 0 \\ \dots & \dots \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} \\ \mathbf{y}(t) &= \begin{bmatrix} -6 & 3 & \vdots & -24 & 7.5 & \vdots & -24 & 3 \\ 0 & 1 & \vdots & 0.5 & 1.5 & \vdots & 1 & 0.5 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} \end{aligned} \quad (4.47)$$

This is a six-dimensional realization.

We discuss a special case of (4.39) and (4.42) in which $p = 1$. To save space, we assume $r = 4$ and $q = 2$. However, the discussion applies to any positive integers r and

q. Consider the 2×1 proper rational matrix

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} + \frac{1}{s^4 + \alpha_1 s^3 + \alpha_2 s^2 + \alpha_3 s + \alpha_4} \cdot \begin{bmatrix} \beta_{11}s^3 + \beta_{12}s^2 + \beta_{13}s + \beta_{14} \\ \beta_{21}s^3 + \beta_{22}s^2 + \beta_{23}s + \beta_{24} \end{bmatrix} \quad (4.48)$$

Then its realization can be obtained directly from (4.42) as

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -\alpha_1 & -\alpha_2 & -\alpha_3 & -\alpha_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} u(t)$$

$$\mathbf{y}(t) = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} \\ \beta_{21} & \beta_{22} & \beta_{23} & \beta_{24} \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} u(t) \quad (4.49)$$

This controllable-form realization can be read out from the coefficients of $\hat{\mathbf{G}}(s)$ in (4.48).

There are many other ways to realize a proper transfer matrix. For example, Problem 4.13 gives a different realization of (4.41) with dimension rq . Let $\hat{\mathbf{G}}_{ci}(s)$ be the i th column of $\hat{\mathbf{G}}(s)$ and let u_i be the i th component of the input vector \mathbf{u} . Then $\hat{\mathbf{y}}(s) = \hat{\mathbf{G}}(s)\hat{\mathbf{u}}(s)$ can be expressed as

$$\hat{\mathbf{y}}(s) = \hat{\mathbf{G}}_{c1}(s)\hat{u}_1(s) + \hat{\mathbf{G}}_{c2}(s)\hat{u}_2(s) + \dots =: \hat{\mathbf{y}}_{c1}(s) + \hat{\mathbf{y}}_{c2}(s) + \dots$$

as shown in Fig. 4.7(a). Thus we can realize each column of $\hat{\mathbf{G}}(s)$ and then combine them to yield a realization of $\hat{\mathbf{G}}(s)$. Let $\hat{\mathbf{G}}_{ri}(s)$ be the i th row of $\hat{\mathbf{G}}(s)$ and let y_i be the i th component of the output vector \mathbf{y} . Then $\hat{\mathbf{y}}(s) = \hat{\mathbf{G}}(s)\hat{\mathbf{u}}(s)$ can be expressed as

$$\hat{y}_i(s) = \hat{\mathbf{G}}_{ri}(s)\hat{\mathbf{u}}(s)$$

as shown in Fig. 4.7(b). Thus we can realize each row of $\hat{\mathbf{G}}(s)$ and then combine them to obtain a realization of $\hat{\mathbf{G}}(s)$. Clearly, we can also realize each entry of $\hat{\mathbf{G}}(s)$ and then combine them to obtain a realization of $\hat{\mathbf{G}}(s)$. See Reference 6, pp. 158–160.

The MATLAB function `[a, b, c, d] = tf2ss (num, den)` generates the controllable-form realization shown in (4.49) for any single-input multiple-output (SIMO) transfer matrix $\hat{\mathbf{G}}(s)$. In its employment, there is no need to decompose $\hat{\mathbf{G}}(s)$ as in (4.39). But we must compute its least common denominator, not necessarily monic. The next example will apply `tf2ss` to each column of $\hat{\mathbf{G}}(s)$ in (4.46) and then combine them to form a realization of $\hat{\mathbf{G}}(s)$.

Example 4.5.4 Consider the proper rational matrix in (4.46). Its first column is

$$\hat{\mathbf{G}}_{c1}(s) = \begin{bmatrix} \frac{4s-10}{2s+1} \\ 1 \\ \frac{1}{(2s+1)(s+2)} \end{bmatrix} = \begin{bmatrix} \frac{(4s-10)(s+2)}{(2s+1)(s+2)} \\ 1 \\ \frac{1}{2s^2+5s+2} \end{bmatrix} = \begin{bmatrix} \frac{4s^2-2s-20}{2s^2+5s+2} \\ 1 \\ \frac{1}{2s^2+5s+2} \end{bmatrix}$$

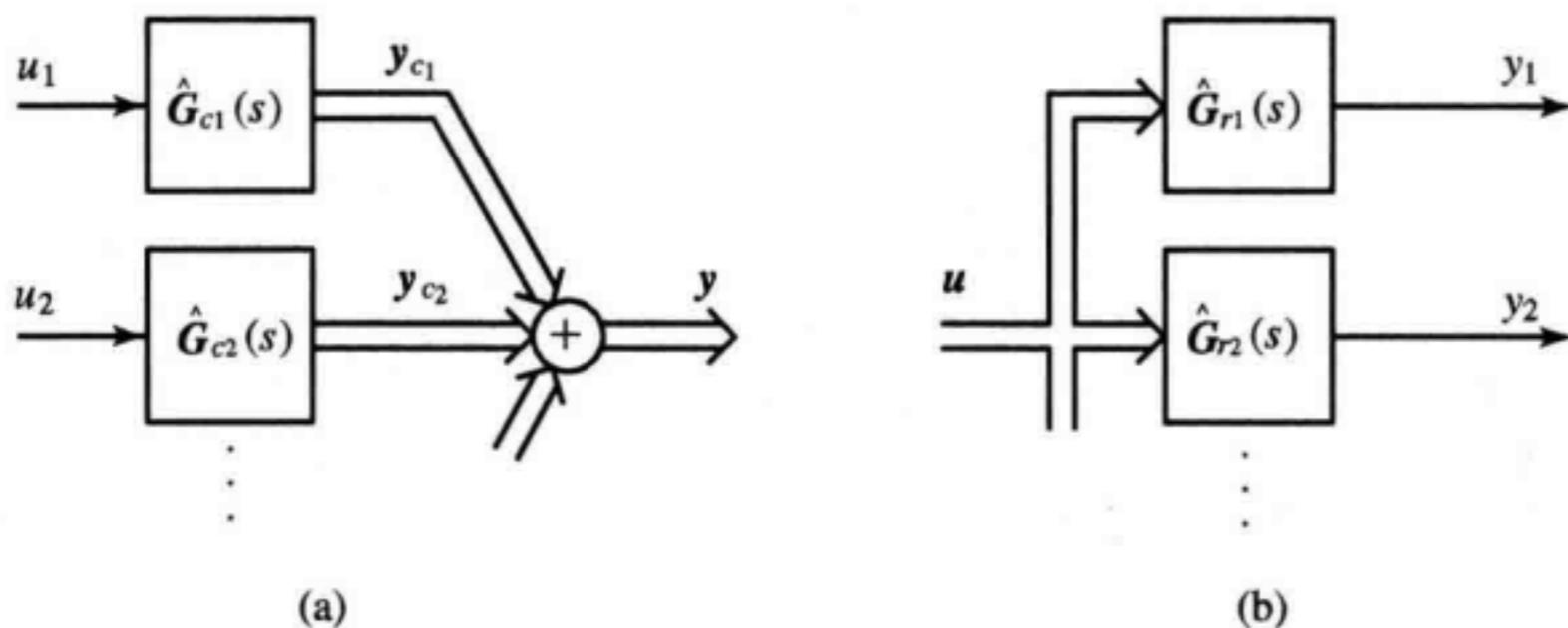


FIGURE 4.7 Realizations of $\hat{\mathbf{G}}(s)$ by columns and by rows.

Typing

```
n1=[4 -2 -20;0 0 1];d1=[2 5 2]; [a,b,c,d]=tf2ss(n1,d1)
```

yields the following realization for the first column of $\hat{\mathbf{G}}(s)$:

$$\begin{aligned}\dot{\mathbf{x}}_1(t) &= \mathbf{A}_1 \mathbf{x}_1(t) + \mathbf{b}_1 u_1(t) = \begin{bmatrix} -2.5 & -1 \\ 1 & 0 \end{bmatrix} \mathbf{x}_1(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u_1(t) \\ \mathbf{y}_{c1}(t) &= \mathbf{C}_1 \mathbf{x}_1(t) + \mathbf{d}_1 u_1(t) = \begin{bmatrix} -6 & -12 \\ 0 & 0.5 \end{bmatrix} \mathbf{x}_1(t) + \begin{bmatrix} 2 \\ 0 \end{bmatrix} u_1(t)\end{aligned}\quad (4.50)$$

Similarly, the function `tf2ss` can generate the following realization for the second column of $\hat{\mathbf{G}}(s)$:

$$\begin{aligned}\dot{\mathbf{x}}_2(t) &= \mathbf{A}_2 \mathbf{x}_2(t) + \mathbf{b}_2 u_2(t) = \begin{bmatrix} -4 & -4 \\ 1 & 0 \end{bmatrix} \mathbf{x}_2(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u_2(t) \\ \mathbf{y}_{c2}(t) &= \mathbf{C}_2 \mathbf{x}_2(t) + \mathbf{d}_2 u_2(t) = \begin{bmatrix} 3 & 6 \\ 1 & 1 \end{bmatrix} \mathbf{x}_2(t) + \begin{bmatrix} 0 \\ 0 \end{bmatrix} u_2(t)\end{aligned}\quad (4.51)$$

These two realizations can be combined as

$$\begin{bmatrix} \dot{\mathbf{x}}_1(t) \\ \dot{\mathbf{x}}_2(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} + \begin{bmatrix} \mathbf{b}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{b}_2 \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$$

$$\mathbf{y}(t) = \mathbf{y}_{c1} + \mathbf{y}_{c2} = [\mathbf{C}_1 \quad \mathbf{C}_2] \mathbf{x}(t) + [\mathbf{d}_1 \quad \mathbf{d}_2] \mathbf{u}(t)$$

or

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \begin{bmatrix} -2.5 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & -4 & -4 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{u}(t) \\ \mathbf{y}(t) &= \begin{bmatrix} -6 & -12 & 3 & 6 \\ 0 & 0.5 & 1 & 1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \mathbf{u}(t)\end{aligned}\quad (4.52)$$

This is a different realization of $\hat{\mathbf{G}}(s)$ in (4.46). This realization has dimension 4, 2 less than the one in (4.47).

The two state-space equations in (4.47) and (4.52) are zero-state equivalent because they have the same transfer matrix. They are, however, not algebraically equivalent. More will be said in Chapter 7 regarding realizations. We mention that all discussion, including `tf2ss`, in this section applies without any modification to the discrete-time case.

4.6 Solution of Linear Time-Varying (LTV) Equations

Consider the linear time-varying (LTV) state-space equation

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \quad (4.53)$$

$$\mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{D}(t)\mathbf{u}(t) \quad (4.54)$$

It is assumed that for every initial state $\mathbf{x}(t_0)$ and any input $\mathbf{u}(t)$, the state-space equation has a unique solution. A sufficient condition for such assumption is that every entry of $\mathbf{A}(t)$ is a continuous function of t . Before considering the general case, we discuss first the solutions of $\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t)$ and the reasons why the approach taken in the time-invariant case cannot be used here.

The solution of the time-invariant equation $\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t)$ can be extended from the scalar equation $\dot{x}(t) = ax(t)$. Indeed, the solution of $\dot{x}(t) = ax(t)$ is $x(t) = e^{at}x(0)$ with $d(e^{at})/dt = ae^{at} = e^{at}a$. Likewise, the solution of $\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t)$ is $\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0)$ with

$$\frac{d}{dt}e^{\mathbf{A}t} = \mathbf{A}e^{\mathbf{A}t} = e^{\mathbf{A}t}\mathbf{A}$$

In this extension, the commutative property $\mathbf{A}e^{\mathbf{A}t} = e^{\mathbf{A}t}\mathbf{A}$ is crucial. Note that in general, we have $\mathbf{AB} \neq \mathbf{BA}$ and $e^{(\mathbf{A}+\mathbf{B})t} \neq e^{\mathbf{At}}e^{\mathbf{Bt}}$.

The solution of the scalar time-varying equation $\dot{x}(t) = a(t)x(t)$ due to $x(0)$ is

$$x(t) = e^{\int_0^t a(\tau)d\tau}x(0)$$

with

$$\frac{d}{dt}e^{\int_0^t a(\tau)d\tau} = a(t)e^{\int_0^t a(\tau)d\tau} = e^{\int_0^t a(\tau)d\tau}a(t)$$

Its extension to the matrix case becomes

$$\mathbf{x}(t) = e^{\int_0^t \mathbf{A}(\tau)d\tau}\mathbf{x}(0) \quad (4.55)$$

where, using (3.51),

$$e^{\int_0^t \mathbf{A}(\tau)d\tau} = \mathbf{I} + \int_0^t \mathbf{A}(\tau)d\tau + \frac{1}{2} \left(\int_0^t \mathbf{A}(\tau)d\tau \right) \left(\int_0^t \mathbf{A}(s)ds \right) + \dots$$

This extension, however, is not valid because

$$\begin{aligned} \frac{d}{dt} e^{\int_0^t \mathbf{A}(\tau) d\tau} &= \mathbf{A}(t) + \frac{1}{2} \mathbf{A}(t) \left(\int_0^t \mathbf{A}(s) ds \right) + \frac{1}{2} \left(\int_0^t \mathbf{A}(\tau) d\tau \right) \mathbf{A}(t) + \dots \\ &\neq \mathbf{A}(t) e^{\int_0^t \mathbf{A}(\tau) d\tau} \end{aligned} \quad (4.56)$$

Thus, in general, (4.55) is not a solution of $\dot{\mathbf{x}} = \mathbf{A}(t)\mathbf{x}$. In conclusion, we cannot extend the solution of scalar time-varying equations to the matrix case and must use a different approach to develop the solution.

Consider

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) \quad (4.57)$$

where $\mathbf{A}(t)$ is $n \times n$ with continuous functions of t as its entries. Then for every initial state $\mathbf{x}_i(t_0)$, there exists a unique solution $\mathbf{x}_i(t)$ for $i = 1, 2, \dots, n$. We can arrange these n solutions as $\mathbf{X}(t) = [\mathbf{x}_1(t) \ \mathbf{x}_2(t) \ \dots \ \mathbf{x}_n(t)]$, a square matrix of order n for every t . Because every $\mathbf{x}_i(t)$ satisfies (4.57), we have

$$\dot{\mathbf{X}}(t) = \mathbf{A}(t)\mathbf{X}(t) \quad (4.58)$$

If $\mathbf{X}(t_0)$ is nonsingular or the n initial states are linearly independent, then $\mathbf{X}(t)$ is called a *fundamental matrix* of (4.57). Because the initial states can be arbitrarily chosen so long as they are linearly independent, the fundamental matrix is not unique.

Example 4.6.1 Consider the homogeneous equation

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 0 \\ t & 0 \end{bmatrix} \mathbf{x}(t) \quad (4.59)$$

or

$$\dot{x}_1(t) = 0, \quad \dot{x}_2(t) = tx_1(t)$$

The solution of $\dot{x}_1(t) = 0$ for $t_0 = 0$ is $x_1(t) = x_1(0)$; the solution of $\dot{x}_2(t) = tx_1(t) = tx_1(0)$ is

$$x_2(t) = \int_0^t \tau x_1(0) d\tau + x_2(0) = 0.5t^2 x_1(0) + x_2(0)$$

Thus we have

$$\mathbf{x}(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Rightarrow \mathbf{x}(t) = \begin{bmatrix} 1 \\ 0.5t^2 \end{bmatrix}$$

and

$$\mathbf{x}(0) = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \Rightarrow \mathbf{x}(t) = \begin{bmatrix} 1 \\ 0.5t^2 + 2 \end{bmatrix}$$

The two initial states are linearly independent, thus

$$\mathbf{X}(t) = \begin{bmatrix} 1 & 1 \\ 0.5t^2 & 0.5t^2 + 2 \end{bmatrix} \quad (4.60)$$

is a fundamental matrix.

A very important property of the fundamental matrix is that $\mathbf{X}(t)$ is nonsingular for all t . For example, $\mathbf{X}(t)$ in (4.60) has determinant $0.5t^2 + 2 - 0.5t^2 = 2$; thus it is nonsingular for all t . We argue intuitively why this is the case. If $\mathbf{X}(t)$ is singular at some t_1 , then there exists a nonzero vector \mathbf{v} such that $\mathbf{x}(t_1) := \mathbf{X}(t_1)\mathbf{v} = \mathbf{0}$. Recall that (4.57) is assumed to have a unique solution for every initial condition. Clearly, $\mathbf{x}(t) \equiv \mathbf{0}$ for all t is the unique solution of (4.57) with $\mathbf{x}(t_1) = \mathbf{0}$. Thus we have $\bar{\mathbf{x}}(t) = \mathbf{X}(t)\mathbf{v} = \mathbf{0}$ for all t , in particular, at $t = t_0$. This contradicts the assumption that $\mathbf{X}(t_0)$ is nonsingular. Thus $\mathbf{X}(t)$ is nonsingular for all t .

Definition 4.2 Let $\mathbf{X}(t)$ be any fundamental matrix of $\dot{\mathbf{x}} = \mathbf{A}(t)\mathbf{x}$. Then

$$\Phi(t, t_0) := \mathbf{X}(t)\mathbf{X}^{-1}(t_0)$$

is called the *state transition matrix* of $\dot{\mathbf{x}} = \mathbf{A}(t)\mathbf{x}$. The transition matrix is also the unique solution of

$$\frac{\partial}{\partial t} \Phi(t, t_0) = \mathbf{A}(t)\Phi(t, t_0) \quad (4.61)$$

with the initial condition $\Phi(t_0, t_0) = \mathbf{I}$.

Because $\mathbf{X}(t)$ is nonsingular for all t , its inverse is well defined. Equation (4.61) follows directly from (4.58). From the definition, we have the following important properties of the state transition matrix:

$$\Phi(t, t) = \mathbf{I} \quad (4.62)$$

$$\Phi^{-1}(t, t_0) = [\mathbf{X}(t)\mathbf{X}^{-1}(t_0)]^{-1} = \mathbf{X}(t_0)\mathbf{X}^{-1}(t) = \Phi(t_0, t) \quad (4.63)$$

$$\Phi(t, t_0) = \Phi(t, t_1)\Phi(t_1, t_0) \quad (4.64)$$

for every t , t_0 , and t_1 .

Example 4.6.2 Consider the homogeneous equation in Example 4.6.1. Its fundamental matrix was computed as

$$\mathbf{X}(t) = \begin{bmatrix} 1 & 1 \\ 0.5t^2 & 0.5t^2 + 2 \end{bmatrix}$$

Its inverse is, using (3.20),

$$\mathbf{X}^{-1}(t) = \begin{bmatrix} 0.25t^2 + 1 & -0.5 \\ -0.25t^2 & 0.5 \end{bmatrix}$$

Thus the state transition matrix is given by

$$\begin{aligned}\Phi(t, t_0) &= \begin{bmatrix} 1 & 1 \\ 0.5t^2 & 0.5t^2 + 2 \end{bmatrix} \begin{bmatrix} 0.25t_0^2 + 1 & -0.5 \\ -0.25t_0^2 & 0.5 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0.5(t^2 - t_0^2) & 1 \end{bmatrix}\end{aligned}$$

It is straightforward to verify that this transition matrix satisfies (4.61) and has the three properties listed in (4.62) through (4.64).

Now we claim that the solution of (4.53) excited by the initial state $\mathbf{x}(t_0) = \mathbf{x}_0$ and the input $\mathbf{u}(t)$ is given by

$$\mathbf{x}(t) = \Phi(t, t_0)\mathbf{x}_0 + \int_{t_0}^t \Phi(t, \tau)\mathbf{B}(\tau)\mathbf{u}(\tau) d\tau \quad (4.65)$$

$$= \Phi(t, t_0) \left[\mathbf{x}_0 + \int_{t_0}^t \Phi(t_0, \tau)\mathbf{B}(\tau)\mathbf{u}(\tau) d\tau \right] \quad (4.66)$$

where $\Phi(t, \tau)$ is the state transition matrix of $\dot{\mathbf{x}} = \mathbf{A}(t)\mathbf{x}$. Equation (4.66) follows from (4.65) by using $\Phi(t, \tau) = \Phi(t, t_0)\Phi(t_0, \tau)$. We show that (4.65) satisfies the initial condition and the state equation. At $t = t_0$, we have

$$\mathbf{x}(t_0) = \Phi(t_0, t_0)\mathbf{x}_0 + \int_{t_0}^{t_0} \Phi(t, \tau)\mathbf{B}(\tau)\mathbf{u}(\tau) d\tau = \mathbf{I}\mathbf{x}_0 + \mathbf{0} = \mathbf{x}_0$$

Thus (4.65) satisfies the initial condition. Using (4.61) and (4.6), we have

$$\begin{aligned}\frac{d}{dt}\mathbf{x}(t) &= \frac{\partial}{\partial t}\Phi(t, t_0)\mathbf{x}_0 + \frac{\partial}{\partial t} \int_{t_0}^t \Phi(t, \tau)\mathbf{B}(\tau)\mathbf{u}(\tau) d\tau \\ &= \mathbf{A}(t)\Phi(t, t_0)\mathbf{x}_0 + \int_{t_0}^t \left(\frac{\partial}{\partial t}\Phi(t, \tau)\mathbf{B}(\tau) \right) d\tau + \Phi(t, t)\mathbf{B}(t)\mathbf{u}(t) \\ &= \mathbf{A}(t)\Phi(t, t_0)\mathbf{x}_0 + \int_{t_0}^t \mathbf{A}(t)\Phi(t, \tau)\mathbf{B}(\tau)\mathbf{u}(\tau) d\tau + \mathbf{B}(t)\mathbf{u}(t) \\ &= \mathbf{A}(t) \left[\Phi(t, t_0)\mathbf{x}_0 + \int_{t_0}^t \Phi(t, \tau)\mathbf{B}(\tau)\mathbf{u}(\tau) d\tau \right] + \mathbf{B}(t)\mathbf{u}(t) \\ &= \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t)\end{aligned}$$

Thus (4.65) is the solution. Substituting (4.65) into (4.54) yields

$$\mathbf{y}(t) = \mathbf{C}(t)\Phi(t, t_0)\mathbf{x}_0 + \mathbf{C}(t) \int_{t_0}^t \Phi(t, \tau)\mathbf{B}(\tau)\mathbf{u}(\tau) d\tau + \mathbf{D}(t)\mathbf{u}(t) \quad (4.67)$$

If the input is identically zero, then Equation (4.65) reduces to

$$\mathbf{x}(t) = \Phi(t, t_0)\mathbf{x}_0$$

This is the zero-input response. Thus the state transition matrix governs the unforced propagation of the state vector. If the initial state is zero, then (4.67) reduces to

$$\begin{aligned}\mathbf{y}(t) &= \mathbf{C}(t) \int_{t_0}^t \Phi(t, \tau) \mathbf{B}(\tau) \mathbf{u}(\tau) d\tau + \mathbf{D}(t) \mathbf{u}(t) \\ &= \int_{t_0}^t [\mathbf{C}(t) \Phi(t, \tau) \mathbf{B}(\tau) + \mathbf{D} \delta(t - \tau)] \mathbf{u}(\tau) d\tau\end{aligned}\quad (4.68)$$

where $\delta(t - t_1)$ is defined in Fig. 2.3 with $\Delta \rightarrow 0$. This is the zero-state response. As discussed in (2.17), the zero-state response can be described by

$$\mathbf{y}(t) = \int_{t_0}^t \mathbf{G}(t, \tau) \mathbf{u}(\tau) d\tau \quad (4.69)$$

where $\mathbf{G}(t, \tau)$ is the impulse response matrix and is the output at time t excited by an impulse input applied at time τ . Comparing (4.68) and (4.69) yields

$$\begin{aligned}\mathbf{G}(t, \tau) &= \mathbf{C}(t) \Phi(t, \tau) \mathbf{B}(\tau) + \mathbf{D}(t) \delta(t - \tau) \\ &= \mathbf{C}(t) \mathbf{X}(t) \mathbf{X}^{-1}(\tau) \mathbf{B}(\tau) + \mathbf{D}(t) \delta(t - \tau)\end{aligned}\quad (4.70)$$

This relates the input-output and state-space descriptions.

The solutions in (4.65) and (4.67) hinge on solving (4.57) or (4.61). If $\mathbf{A}(t)$ is triangular such as

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} a_{11}(t) & 0 \\ a_{21}(t) & a_{22}(t) \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

we can solve the scalar equation $\dot{x}_1(t) = a_{11}(t)x_1(t)$ and then substitute it into

$$\dot{x}_2(t) = a_{22}(t)x_2(t) + a_{21}(t)x_1(t)$$

Because $x_1(t)$ has been solved, the preceding scalar equation can be solved for $x_2(t)$. This is what we did in Example 4.6.1. If $\mathbf{A}(t)$, such as $\mathbf{A}(t)$ diagonal or constant, has the following commutative property:

$$\mathbf{A}(t) \left(\int_{t_0}^t \mathbf{A}(\tau) d\tau \right) = \left(\int_{t_0}^t \mathbf{A}(\tau) d\tau \right) \mathbf{A}(t)$$

for all t_0 and t , then the solution of (4.61) can be shown to be

$$\Phi(t, t_0) = e^{\int_{t_0}^t \mathbf{A}(\tau) d\tau} = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\int_{t_0}^t \mathbf{A}(\tau) d\tau \right)^k \quad (4.71)$$

For $\mathbf{A}(t)$ constant, (4.71) reduces to

$$\Phi(t, \tau) = e^{\mathbf{A}(t-\tau)} = \Phi(t - \tau)$$

and $\mathbf{X}(t) = e^{\mathbf{A}t}$. Other than the preceding special cases, computing state transition matrices is generally difficult.

4.6.1 Discrete-Time Case

Consider the discrete-time state-space equation

$$\mathbf{x}[k+1] = \mathbf{A}[k]\mathbf{x}[k] + \mathbf{B}[k]\mathbf{u}[k] \quad (4.72)$$

$$\mathbf{y}[k] = \mathbf{C}[k]\mathbf{x}[k] + \mathbf{D}[k]\mathbf{u}[k] \quad (4.73)$$

They are algebraic equations and their solutions can be computed recursively once the initial state $\mathbf{x}[k_0]$ and the input $\mathbf{u}[k]$, for $k \geq k_0$, are given. The situation here is much simpler than the continuous-time case.

As in the continuous-time case, we can define the discrete state transition matrix as the solution of

$$\Phi[k+1, k_0] = \mathbf{A}[k]\Phi[k, k_0] \quad \text{with } \Phi[k_0, k_0] = \mathbf{I}$$

for $k = k_0, k_0 + 1, \dots$. This is the discrete counterpart of (4.61) and its solution can be obtained directly as

$$\Phi[k, k_0] = \mathbf{A}[k-1]\mathbf{A}[k-2] \cdots \mathbf{A}[k_0] \quad (4.74)$$

for $k > k_0$ and $\Phi[k_0, k_0] = \mathbf{I}$. We discuss a significant difference between the continuous- and discrete-time cases. Because the fundamental matrix in the continuous-time case is nonsingular for all t , the state transition matrix $\Phi(t, t_0)$ is defined for $t \geq t_0$ and $t < t_0$ and can govern the propagation of the state vector in the positive-time and negative-time directions. In the discrete-time case, the \mathbf{A} -matrix can be singular; thus the inverse of $\Phi[k, k_0]$ may not be defined. Thus $\Phi[k, k_0]$ is defined only for $k \geq k_0$ and governs the propagation of the state vector in only the positive-time direction. Therefore, the discrete counterpart of (4.64) or

$$\Phi[k, k_0] = \Phi[k, k_1]\Phi[k_1, k_0]$$

holds only for $k \geq k_1 \geq k_0$.

Using the discrete state transition matrix, we can express the solutions of (4.72) and (4.73) as, for $k > k_0$,

$$\begin{aligned} \mathbf{x}[k] &= \Phi[k, k_0]\mathbf{x}_0 + \sum_{m=k_0}^{k-1} \Phi[k, m+1]\mathbf{B}[m]\mathbf{u}[m] \\ \mathbf{y}[k] &= \mathbf{C}[k]\Phi[k, k_0]\mathbf{x}_0 + \mathbf{C}[k] \sum_{m=k_0}^{k-1} \Phi[k, m+1]\mathbf{B}[m]\mathbf{u}[m] \\ &\quad + \mathbf{D}[k]\mathbf{u}[k] \end{aligned} \quad (4.75)$$

Their derivations are similar to those in the preceding section and will not be repeated.

If the initial state is zero, Equation (4.75) reduces to

$$\mathbf{y}[k] = \mathbf{C}[k] \sum_{m=k_0}^{k-1} \Phi[k, m+1]\mathbf{B}[m]\mathbf{u}[m] + \mathbf{D}[k]\mathbf{u}[k] \quad (4.76)$$

for $k > k_0$. This describes the zero-state response of (4.73). If we define $\Phi[k, m] = \mathbf{0}$ for $k < m$, then (4.76) can be written as

$$\mathbf{y}[k] = \sum_{m=k_0}^k (\mathbf{C}[k]\Phi[k, m+1]\mathbf{B}[m] + \mathbf{D}[m]\delta_d[k-m])\mathbf{u}[m]$$

where the impulse sequence $\delta_d[k-m]$ equals 1 if $k=m$ and 0 if $k \neq m$. Comparing this with the DT version of (2.17), we have

$$\mathbf{G}[k, m] = \mathbf{C}[k]\Phi[k, m+1]\mathbf{B}[m] + \mathbf{D}[m]\delta_d[k-m]$$

for $k \geq m$. This relates the impulse response sequence and the state-space equation and is the discrete counterpart of (4.70).

4.7 Equivalent Time-Varying Equations

This section extends the equivalent state-space equations discussed in Section 4.4 to the time-varying case. Consider the n -dimensional linear time-varying state-space equation

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \quad (4.77)$$

$$\mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{D}(t)\mathbf{u}(t)$$

Let $\mathbf{P}(t)$ be an $n \times n$ matrix. It is assumed that $\mathbf{P}(t)$ is nonsingular and both $\mathbf{P}(t)$ and $\dot{\mathbf{P}}(t)$ are continuous for all t . Let $\bar{\mathbf{x}} = \mathbf{P}(t)\mathbf{x}$. Then the state-space equation

$$\dot{\bar{\mathbf{x}}}(t) = \bar{\mathbf{A}}(t)\bar{\mathbf{x}}(t) + \bar{\mathbf{B}}(t)\mathbf{u}(t) \quad (4.78)$$

$$\mathbf{y}(t) = \bar{\mathbf{C}}(t)\bar{\mathbf{x}}(t) + \bar{\mathbf{D}}(t)\mathbf{u}(t)$$

where

$$\begin{aligned}\bar{\mathbf{A}}(t) &= [\mathbf{P}(t)\mathbf{A}(t) + \dot{\mathbf{P}}(t)]\mathbf{P}^{-1}(t) \\ \bar{\mathbf{B}}(t) &= \mathbf{P}(t)\mathbf{B}(t) \\ \bar{\mathbf{C}}(t) &= \mathbf{C}(t)\mathbf{P}^{-1}(t) \\ \bar{\mathbf{D}}(t) &= \mathbf{D}(t)\end{aligned}$$

is said to be (algebraically) equivalent to (4.77) and $\mathbf{P}(t)$ is called an *(algebraic) equivalence* transformation.

Equation (4.78) is obtained from (4.77) by substituting $\bar{\mathbf{x}} = \mathbf{P}(t)\mathbf{x}$ and $\dot{\bar{\mathbf{x}}} = \mathbf{P}(t)\dot{\mathbf{x}} + \dot{\mathbf{P}}(t)\mathbf{x}$. Let \mathbf{X} be a fundamental matrix of (4.77). Then we claim that

$$\bar{\mathbf{X}}(t) := \mathbf{P}(t)\mathbf{X}(t) \quad (4.79)$$

is a fundamental matrix of (4.78). By definition, $\dot{\bar{\mathbf{X}}}(t) = \bar{\mathbf{A}}(t)\bar{\mathbf{X}}(t)$ and $\mathbf{X}(t)$ is nonsingular for all t . Because the rank of a matrix will not change by multiplying a nonsingular

matrix, the matrix $\mathbf{P}(t)\mathbf{X}(t)$ is also nonsingular for all t . Now we show that $\mathbf{P}(t)\mathbf{X}(t)$ satisfies the equation $\dot{\bar{\mathbf{x}}} = \bar{\mathbf{A}}(t)\bar{\mathbf{x}}$. Indeed, we have

$$\begin{aligned}\frac{d}{dt}[\mathbf{P}(t)\mathbf{X}(t)] &= \dot{\mathbf{P}}(t)\mathbf{X}(t) + \mathbf{P}(t)\dot{\mathbf{X}}(t) = \dot{\mathbf{P}}(t)\mathbf{X}(t) + \mathbf{P}(t)\mathbf{A}(t)\mathbf{X}(t) \\ &= [\dot{\mathbf{P}}(t) + \mathbf{P}(t)\mathbf{A}(t)][\mathbf{P}^{-1}(t)\mathbf{P}(t)]\mathbf{X}(t) = \bar{\mathbf{A}}(t)[\mathbf{P}(t)\mathbf{X}(t)]\end{aligned}$$

Thus, $\mathbf{P}(t)\mathbf{X}(t)$ is a fundamental matrix of $\dot{\bar{\mathbf{x}}}(t) = \bar{\mathbf{A}}(t)\bar{\mathbf{x}}(t)$.

Theorem 4.3

Let \mathbf{A}_o be an arbitrary constant matrix. Then there exists an equivalence transformation that transforms (4.77) into (4.78) with $\bar{\mathbf{A}}(t) = \mathbf{A}_o$.

Proof: Let $\mathbf{X}(t)$ be a fundamental matrix of $\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t)$. The differentiation of $\mathbf{X}^{-1}(t)\mathbf{X}(t) = \mathbf{I}$ yields

$$\dot{\mathbf{X}}^{-1}(t)\mathbf{X}(t) + \mathbf{X}^{-1}\dot{\mathbf{X}}(t) = \mathbf{0}$$

which implies

$$\dot{\mathbf{X}}^{-1}(t) = -\mathbf{X}^{-1}(t)\mathbf{A}(t)\mathbf{X}(t)\mathbf{X}^{-1}(t) = -\mathbf{X}^{-1}(t)\mathbf{A}(t) \quad (4.80)$$

Because $\bar{\mathbf{A}}(t) = \mathbf{A}_o$ is a constant matrix, $\bar{\mathbf{X}}(t) = e^{\mathbf{A}_o t}$ is a fundamental matrix of $\dot{\bar{\mathbf{x}}}(t) = \bar{\mathbf{A}}(t)\bar{\mathbf{x}}(t) = \mathbf{A}_o\bar{\mathbf{x}}(t)$. Following (4.79), we define

$$\mathbf{P}(t) := \bar{\mathbf{X}}(t)\mathbf{X}^{-1}(t) = e^{\mathbf{A}_o t}\mathbf{X}^{-1}(t) \quad (4.81)$$

and compute

$$\begin{aligned}\bar{\mathbf{A}}(t) &= [\mathbf{P}(t)\mathbf{A}(t) + \dot{\mathbf{P}}(t)]\mathbf{P}^{-1}(t) \\ &= [e^{\mathbf{A}_o t}\mathbf{X}^{-1}(t)\mathbf{A}(t) + \mathbf{A}_o e^{\mathbf{A}_o t}\mathbf{X}^{-1}(t) + e^{\mathbf{A}_o t}\dot{\mathbf{X}}^{-1}(t)]\mathbf{X}(t)e^{-\mathbf{A}_o t}\end{aligned}$$

which becomes, after substituting (4.80),

$$\bar{\mathbf{A}}(t) = \mathbf{A}_o e^{\mathbf{A}_o t}\mathbf{X}^{-1}(t)\mathbf{X}(t)e^{-\mathbf{A}_o t} = \mathbf{A}_o$$

This establishes the theorem. Q.E.D.

If \mathbf{A}_o is chosen as a zero matrix, then $\bar{\mathbf{X}}(t) = I$ and $\mathbf{P}(t) = \mathbf{X}^{-1}(t)$. Thus (4.78) reduces to

$$\bar{\mathbf{A}}(t) = \mathbf{0}, \quad \bar{\mathbf{B}}(t) = \mathbf{X}^{-1}(t)\mathbf{B}(t), \quad \bar{\mathbf{C}}(t) = \mathbf{C}(t)\mathbf{X}(t), \quad \bar{\mathbf{D}}(t) = \mathbf{D}(t) \quad (4.82)$$

The block diagrams of (4.77) with $\mathbf{A}(t) \neq \mathbf{0}$ and $\mathbf{A}(t) = \mathbf{0}$ are plotted in Fig. 4.8. The block diagram with $\mathbf{A}(t) = \mathbf{0}$ has no feedback and is considerably simpler. Every time-varying state-space equation can be transformed into such a block diagram. However, in order to do so, we need its fundamental matrix, which, however, is generally not available.

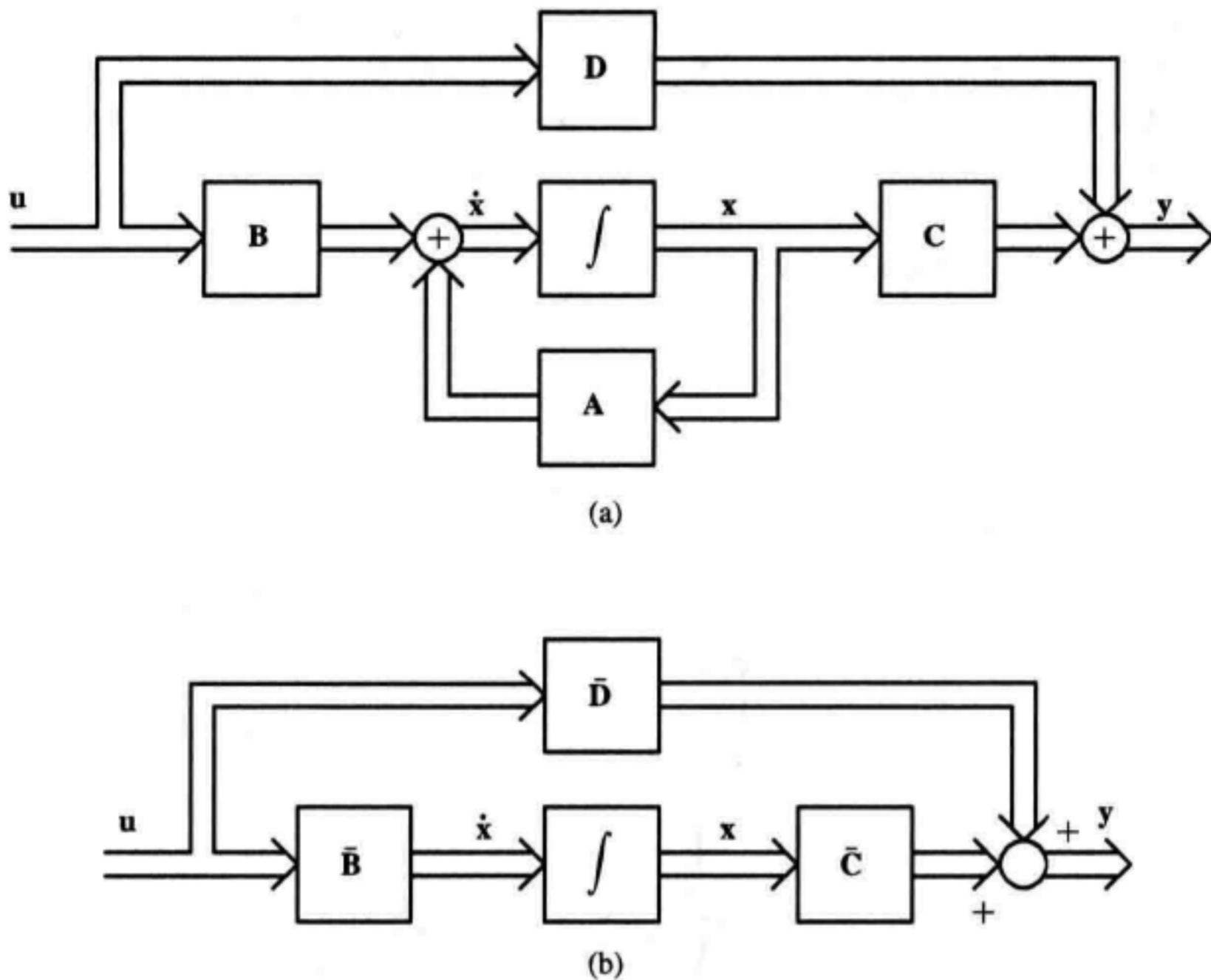


FIGURE 4.8 Block diagrams with feedback and without feedback.

The impulse response matrix of (4.77) is given in (4.70). The impulse response matrix of (4.78) is, using (4.79) and (4.80),

$$\begin{aligned}\bar{\mathbf{G}}(t, \tau) &= \bar{\mathbf{C}}(t)\bar{\mathbf{X}}(t)\bar{\mathbf{X}}^{-1}(\tau)\bar{\mathbf{B}}(\tau) + \bar{\mathbf{D}}(t)\delta(t - \tau) \\ &= \mathbf{C}(t)\mathbf{P}^{-1}(t)\mathbf{P}(t)\mathbf{X}(t)\mathbf{X}^{-1}(\tau)\mathbf{P}^{-1}(\tau)\mathbf{P}(\tau)\mathbf{B}(\tau) + \mathbf{D}(t)\delta(t - \tau) \\ &= \mathbf{C}(t)\mathbf{X}(t)\mathbf{X}^{-1}(\tau)\mathbf{B}(\tau) + \mathbf{D}(t)\delta(t - \tau) = \mathbf{G}(t, \tau)\end{aligned}$$

Thus the impulse response matrix is invariant under any equivalence transformation. The property of the A-matrix, however, may not be preserved in equivalence transformations. For example, every A-matrix can be transformed, as shown in Theorem 4.3, into a constant or a zero matrix. Clearly the zero matrix does not have any property of $\mathbf{A}(t)$. In the time-invariant case, algebraic transformation will preserve all properties of the original state-space equation. Thus the algebraic transformation in the time-invariant case is not a special case of the time-varying case.

Definition 4.3 A matrix $\mathbf{P}(t)$ is called a *Lyapunov transformation* if $\mathbf{P}(t)$ is nonsingular, $\mathbf{P}(t)$ and $\dot{\mathbf{P}}(t)$ are continuous, and $\mathbf{P}(t)$ and $\mathbf{P}^{-1}(t)$ are bounded for all t . Equations (4.77) and (4.78) are said to be *Lyapunov equivalent* if $\mathbf{P}(t)$ is a Lyapunov transformation.

It is clear that if $\mathbf{P}(t) = \mathbf{P}$ is a constant matrix, then it is a Lyapunov transformation. Thus the (algebraic) transformation in the time-invariant case is a special case of the

Lyapunov transformation. If $\mathbf{P}(t)$ is required to be a Lyapunov transformation, then Theorem 4.3 does not hold in general. In other words, not every time-varying state-space equation can be Lyapunov equivalent to a state-space equation with a constant A-matrix. However, this is true if $\mathbf{A}(t)$ is periodic.

Periodic state-space equations

Consider the linear time-varying state-space equation in (4.77). It is assumed that

$$\mathbf{A}(t+T) = \mathbf{A}(t)$$

for all t and for some positive constant T . That is, $\mathbf{A}(t)$ is periodic with period T . Let $\mathbf{X}(t)$ be a fundamental matrix of $\dot{\mathbf{x}} = \mathbf{A}(t)\mathbf{x}$ or $\dot{\mathbf{X}}(t) = \mathbf{A}(t)\mathbf{X}(t)$ with $\mathbf{X}(0)$ nonsingular. Then we have

$$\dot{\mathbf{X}}(t+T) = \mathbf{A}(t+T)\mathbf{X}(t+T) = \mathbf{A}(t)\mathbf{X}(t+T)$$

Thus $\mathbf{X}(t+T)$ is also a fundamental matrix. Furthermore, it can be expressed as

$$\mathbf{X}(t+T) = \mathbf{X}(t)\mathbf{X}^{-1}(0)\mathbf{X}(T) \quad (4.83)$$

This can be verified by direct substitution. Let us define $\mathbf{Q} = \mathbf{X}^{-1}(0)\mathbf{X}(T)$. It is a constant nonsingular matrix. For this \mathbf{Q} there exists a constant matrix $\bar{\mathbf{A}}$ such that $e^{\bar{\mathbf{A}}T} = \mathbf{Q}$ (Problem 3.29). Thus (4.83) can be written as

$$\mathbf{X}(t+T) = \mathbf{X}(t)e^{\bar{\mathbf{A}}T} \quad (4.84)$$

Define

$$\mathbf{P}(t) := e^{\bar{\mathbf{A}}t}\mathbf{X}^{-1}(t) \quad (4.85)$$

We show that $\mathbf{P}(t)$ is periodic with period T :

$$\begin{aligned} \mathbf{P}(t+T) &= e^{\bar{\mathbf{A}}(t+T)}\mathbf{X}^{-1}(t+T) = e^{\bar{\mathbf{A}}t}e^{\bar{\mathbf{A}}T}[e^{-\bar{\mathbf{A}}T}\mathbf{X}^{-1}(t)] \\ &= e^{\bar{\mathbf{A}}t}\mathbf{X}^{-1}(t) = \mathbf{P}(t) \end{aligned}$$

Theorem 4.4

Consider (4.77) with $\mathbf{A}(t) = \mathbf{A}(t+T)$ for all t and some $T > 0$. Let $\mathbf{X}(t)$ be a fundamental matrix of $\dot{\mathbf{x}} = \mathbf{A}(t)\mathbf{x}$. Let $\bar{\mathbf{A}}$ be the constant matrix computed from $e^{\bar{\mathbf{A}}T} = \mathbf{X}^{-1}(0)\mathbf{X}(T)$. Then (4.77) is Lyapunov equivalent to

$$\begin{aligned} \dot{\bar{\mathbf{x}}}(t) &= \bar{\mathbf{A}}\bar{\mathbf{x}}(t) + \mathbf{P}(t)\mathbf{B}(t)\mathbf{u}(t) \\ \bar{\mathbf{y}}(t) &= \mathbf{C}(t)\mathbf{P}^{-1}(t)\bar{\mathbf{x}}(t) + \mathbf{D}(t)\mathbf{u}(t) \end{aligned}$$

where $\mathbf{P}(t) = e^{\bar{\mathbf{A}}t}\mathbf{X}^{-1}(t)$.

The matrix $\mathbf{P}(t)$ in (4.85) satisfies all conditions in Definition 4.3, thus it is a Lyapunov transformation. The rest of the theorem follows directly from Theorem 4.3. The homogeneous part of Theorem 4.4 is the so-called *theory of Floquet*. It states that

if $\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t)$ and if $\mathbf{A}(t+T) = \mathbf{A}(t)$ for all t , then its fundamental matrix is of the form $\mathbf{P}^{-1}(t)e^{\tilde{\mathbf{A}}t}$, where $\mathbf{P}^{-1}(t)$ is a periodic function. Furthermore, $\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t)$ is Lyapunov equivalent to $\dot{\bar{\mathbf{x}}}(t) = \tilde{\mathbf{A}}\bar{\mathbf{x}}(t)$.

4.8 Time-Varying Realizations

We studied in Section 4.5 the realization problem for linear time-invariant systems. In this section, we study the corresponding problem for linear time-varying systems. The Laplace transform cannot be used here; thus, we study the problem directly in the time domain.

Every linear time-varying system can be described by the input–output description

$$\mathbf{y}(t) = \int_{t_0}^t \mathbf{G}(t, \tau) \mathbf{u}(\tau) d\tau$$

and, if the system is lumped as well, by the state-space equation

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}(t)\mathbf{x}(t) + \mathbf{D}(t)\mathbf{u}(t)\end{aligned}\tag{4.86}$$

If the state-space equation is available, the impulse response matrix can be computed from

$$\mathbf{G}(t, \tau) = \mathbf{C}(t)\mathbf{X}(t)\mathbf{X}^{-1}(\tau)\mathbf{B}(\tau) + \mathbf{D}(t)\delta(t - \tau) \quad \text{for } t \geq \tau\tag{4.87}$$

where $\mathbf{X}(t)$ is a fundamental matrix of $\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t)$. The converse problem is to find a state-space equation from a given impulse response matrix. An impulse response matrix $\mathbf{G}(t, \tau)$ is said to be *realizable* if there exists $\{\mathbf{A}(t), \mathbf{B}(t), \mathbf{C}(t), \mathbf{D}(t)\}$ to meet (4.87).

Theorem 4.5

An $q \times p$ impulse response matrix $\mathbf{G}(t, \tau)$ is realizable if and only if $\mathbf{G}(t, \tau)$ can be decomposed as

$$\mathbf{G}(t, \tau) = \mathbf{M}(t)\mathbf{N}(\tau) + \mathbf{D}(t)\delta(t - \tau)\tag{4.88}$$

for all $t \geq \tau$, where \mathbf{M} , \mathbf{N} , and \mathbf{D} , are, respectively, $q \times n$, $n \times p$, and $q \times p$ matrices for some integer n .



Proof: If $\mathbf{G}(t, \tau)$ is realizable, there exist a realization that meets (4.87). Identifying $\mathbf{M}(t) = \mathbf{C}(t)\mathbf{X}(t)$ and $\mathbf{N}(\tau) = \mathbf{X}^{-1}(\tau)\mathbf{B}(\tau)$ establishes the necessary part of the theorem.

If $\mathbf{G}(t, \tau)$ can be decomposed as in (4.88), then the n -dimensional state-space equation

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{N}(t)\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{M}(t)\mathbf{x}(t) + \mathbf{D}(t)\mathbf{u}(t)\end{aligned}\tag{4.89}$$

is a realization. Indeed, a fundamental matrix of $\dot{\mathbf{x}} = \mathbf{0} \cdot \mathbf{x}$ is $\mathbf{X}(t) = \mathbf{I}$. Thus the impulse response matrix of (4.89) is, using (4.87),

$$\mathbf{M}(t)\mathbf{I} \cdot \mathbf{I}^{-1}\mathbf{N}(\tau) + \mathbf{D}(t)\delta(t - \tau)$$

which equals $\mathbf{G}(t, \tau)$. This shows the sufficiency of the theorem. Q.E.D.

Although Theorem 4.5 can also be applied to time-invariant systems, the result is not satisfactory as the next example illustrates.

Example 4.8.1 Consider $g(t) = te^{\lambda t}$ or

$$g(t, \tau) = g(t - \tau) = (t - \tau)e^{\lambda(t-\tau)}$$

It is straightforward to verify

$$g(t - \tau) = [e^{\lambda t} \quad te^{\lambda t}] \begin{bmatrix} -\tau e^{-\lambda \tau} \\ e^{-\lambda \tau} \end{bmatrix}$$

Thus the two-dimensional time-varying state-space equation

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} -te^{-\lambda t} \\ e^{-\lambda t} \end{bmatrix} u(t) \\ \mathbf{y}(t) &= [e^{\lambda t} \quad te^{\lambda t}] \mathbf{x}(t)\end{aligned}\tag{4.90}$$

is a realization of the impulse response $g(t) = te^{\lambda t}$.

The Laplace transform of the impulse response is

$$\hat{g}(s) = \mathcal{L}[te^{\lambda t}] = \frac{1}{(s - \lambda)^2} = \frac{1}{s^2 - 2\lambda s + \lambda^2}$$

Using (4.33), we can readily obtain

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \begin{bmatrix} 2\lambda & -\lambda^2 \\ 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t) \\ \mathbf{y}(t) &= [0 \quad 1] \mathbf{x}(t)\end{aligned}\tag{4.91}$$

This LTI state-space equation is a different realization of the same impulse response. This realization is clearly more desirable because it can be readily implemented using an op-amp circuit. The implementation of (4.90) is much more difficult in practice.

Problems

- 4.1** An oscillation can be generated by

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \mathbf{x}(t)$$

Show that its solution is

$$\mathbf{x}(t) = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix} \mathbf{x}(0)$$

- 4.2** Use two different methods to find the unit step response of

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \begin{bmatrix} 0 & 1 \\ -2 & -2 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u(t) \\ y(t) &= [2 \quad 3] \mathbf{x}(t)\end{aligned}$$

- 4.3** Find the solution of the following state equations with arbitrary initial conditions.

1. $\dot{x}(t) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} x(t)$

2. $\dot{x}(t) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 2 \\ 0 & 1 & -3 \end{bmatrix} x(t)$

3. $x(k+1) = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} x(k)$

- 4.4** Discretize the state-space equation in Problem 4.2 for $T = 1$ and $T = \pi$.

- 4.5** Find the companion-form and modal-form equivalent equations of

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \begin{bmatrix} -2 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & -2 & -2 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} u(t) \\ y(t) &= [1 \quad -1 \quad 0] \mathbf{x}(t)\end{aligned}$$

- 4.6** Consider the following system given in its state-space representation,

$$\dot{x}(t) = \begin{bmatrix} -2 & -2 & 0 \\ 0 & 0 & 1 \\ 0 & -3 & -4 \end{bmatrix} x(t) + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u(t)$$

1. Obtain the equivalent modal form of the above system.

2. Find the response of the system with $x(0) = [1 \quad 1 \quad 1]'$ and the input as the step command using (a).

3. Use MATLAB to sketch the response shape of the system based on (b).

- 4.7** Find an equivalent state-space equation of the equation in Problem 4.5 so that all state variables have largest magnitudes roughly equal to the largest magnitude of the output.

If all signals are required to lie inside ± 10 volts and if the input is a step function with magnitude a , what is the permissible largest a ?

- 4.8** Consider the following state equation of a linear system

$$\begin{aligned}\dot{x}(t) &= \begin{bmatrix} 0 & 1 \\ -5 & -2 \end{bmatrix}x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix}u(t) \\ y(t) &= [1 \ 0]x(t)\end{aligned}$$

1. Find the zero-input response of the system with $x(0) = [1 \ 1]'$.
2. Find the transfer function of the system.
3. Find an equivalent modal representation of the system. Is the modal representation physically realizable? If the answer is No, find a physically realizable one.

- 4.9** The discretization method of section 4.2.1 is analogous to the sampled-data system in digital control for converting continuous-time systems to discrete-time. Suppose the transfer function of a continuous-time system is given by

$$\hat{g}(s) = \frac{K}{s(s+1)}$$

in a sampled-data system with $T = 0.1\text{s}$. Obtain its discrete state-space representation and compute $\hat{g}(z)$. Let $K = 10$ and obtain the response of both continuous-time and discrete-time systems to step command. Plot the step responses and compare the results.

- 4.10** Consider

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix}\mathbf{x}(t) + \begin{bmatrix} b_1 \\ \bar{b}_1 \end{bmatrix}u(t), \quad y(t) = [c_1 \ \bar{c}_1]\mathbf{x}(t)$$

where the overbar denotes complex conjugate. Verify that the equation can be transformed into

$$\dot{\bar{\mathbf{x}}}(t) = \bar{\mathbf{A}}\bar{\mathbf{x}}(t) + \bar{\mathbf{b}}u(t), \quad y(t) = \bar{\mathbf{c}}\bar{\mathbf{x}}(t)$$

with

$$\bar{\mathbf{A}} = \begin{bmatrix} 0 & 1 \\ -\lambda\bar{\lambda} & \lambda + \bar{\lambda} \end{bmatrix}, \quad \bar{\mathbf{b}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \bar{\mathbf{c}}_1 = [-2\operatorname{Re}(\bar{\lambda}b_1c_1) \ 2\operatorname{Re}(b_1c_1)]$$

by using the transformation $\mathbf{x} = \mathbf{Q}\bar{\mathbf{x}}$ with

$$\mathbf{Q}_1 = \begin{bmatrix} -\bar{\lambda}b_1 & b_1 \\ -\lambda\bar{b}_1 & \bar{b}_1 \end{bmatrix}$$

- 4.11** Verify that the Jordan-form equation

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \begin{bmatrix} \lambda & 1 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & \bar{\lambda} & 1 & 0 \\ 0 & 0 & 0 & 0 & \bar{\lambda} & 1 \\ 0 & 0 & 0 & 0 & 0 & \bar{\lambda} \end{bmatrix}\mathbf{x}(t) + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \bar{b}_1 \\ \bar{b}_2 \\ \bar{b}_3 \end{bmatrix}u(t) \\ y(t) &= [c_1 \ c_2 \ c_3 \ \bar{c}_1 \ \bar{c}_2 \ \bar{c}_3]\mathbf{x}(t)\end{aligned}$$

can be transformed into

$$\dot{\bar{\mathbf{x}}}(t) = \begin{bmatrix} \bar{\mathbf{A}} & \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{A}} & \mathbf{I}_2 \\ \mathbf{0} & \mathbf{0} & \bar{\mathbf{A}} \end{bmatrix} \bar{\mathbf{x}}(t) + \begin{bmatrix} \bar{\mathbf{b}} \\ \bar{\mathbf{b}} \\ \bar{\mathbf{b}} \end{bmatrix} u(t)$$

$$y(t) = [\bar{\mathbf{c}}_1 \quad \bar{\mathbf{c}}_2 \quad \bar{\mathbf{c}}_3] \bar{\mathbf{x}}(t)$$

where $\bar{\mathbf{A}}$, $\bar{\mathbf{b}}$, and $\bar{\mathbf{c}}_i$ are defined in Problem 4.10 and \mathbf{I}_2 is the unit matrix of order 2. [Hint: Change the order of the state variables from $[x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6]'$ to $[x_1 \ x_4 \ x_2 \ x_5 \ x_3 \ x_6]'$ and then apply the equivalence transformation $\mathbf{x} = \mathbf{Q}\bar{\mathbf{x}}$ with $\mathbf{Q} = \text{diag}(\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3)$.]

- 4.12** Are the two sets of state-space equations

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 2 & 1 & 2 \\ 0 & 2 & 2 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} u(t), \quad y(t) = [1 \ -1 \ 0] \mathbf{x}(t)$$

and

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & -1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} u(t), \quad y(t) = [1 \ -1 \ 0] \mathbf{x}(t)$$

equivalent? zero-state equivalent?

- 4.13** Verify that the transfer matrix in (4.41) has the following realization:

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -\alpha_1 \mathbf{I}_q & \mathbf{I}_q & \mathbf{0} & \cdots & \mathbf{0} \\ -\alpha_2 \mathbf{I}_q & \mathbf{0} & \mathbf{I}_q & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & & \vdots \\ -\alpha_{r-1} \mathbf{I}_q & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_q \\ -\alpha_r \mathbf{I}_q & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} \mathbf{N}_1 \\ \mathbf{N}_2 \\ \vdots \\ \mathbf{N}_{r-1} \\ \mathbf{N}_r \end{bmatrix} \mathbf{u}(t)$$

$$\mathbf{y}(t) = [\mathbf{I}_q \ \mathbf{0} \ \mathbf{0} \ \cdots \ \mathbf{0}] \mathbf{x}(t)$$

This is called a *block observable form realization* and has dimension rq . It is dual to (4.42).

- 4.14** Consider the 1×2 proper rational matrix:

$$\hat{\mathbf{G}}(s) = [d_1 \ d_2] + \frac{1}{s^4 + \alpha_1 s^3 + \alpha_2 s^2 + \alpha_3 s + \alpha_4}$$

$$\times [\beta_{11} s^3 + \beta_{21} s^2 + \beta_{31} s + \beta_{41} \quad \beta_{12} s^3 + \beta_{22} s^2 + \beta_{32} s + \beta_{42}]$$

Show that its observable form realization can be reduced from Problem 4.13 as

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -\alpha_1 & 1 & 0 & 0 \\ -\alpha_2 & 0 & 1 & 0 \\ -\alpha_3 & 0 & 0 & 1 \\ -\alpha_4 & 0 & 0 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{13} & \beta_{32} \\ \beta_{41} & \beta_{42} \end{bmatrix} \mathbf{u}(t)$$

$$y(t) = [1 \ 0 \ 0 \ 0] \mathbf{x}(t) + [d_1 \ d_2] \mathbf{u}(t)$$

4.15 Find a realization for the proper rational matrix:

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} \frac{2}{s+1} & \frac{2s-3}{(s+1)(s+2)} \\ \frac{s-2}{s+1} & \frac{s}{s+2} \end{bmatrix}$$

4.16 Find a realization for each column of $\hat{\mathbf{G}}(s)$ in Problem 4.15 and then connect them, as shown in Fig. 4.7(a), to obtain a realization of $\hat{\mathbf{G}}(s)$. What is the dimension of this realization? Compare this dimension with the one in Problem 4.15.

4.17 Find a realization for each row of $\hat{\mathbf{G}}(s)$ in Problem 4.15 and then connect them, as shown in Fig. 4.7(b), to obtain a realization of $\hat{\mathbf{G}}(s)$. What is the dimension of this realization? Compare this dimension with the ones in Problems 4.15 and 4.16.

4.18 Find a realization for

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} -(12s+6) & 22s+23 \\ 3s+34 & 3s+34 \end{bmatrix}$$

4.19 Consider the n -dimensional state-space equation

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t), \quad y(t) = \mathbf{c}\mathbf{x}(t)$$

Let $\hat{g}(s)$ be its transfer function. Show that $\hat{g}(s)$ has m zeros or, equivalently, the numerator of $\hat{g}(s)$ has degree m if and only if

$$\mathbf{c}\mathbf{A}^i\mathbf{b} = 0 \quad \text{for } i = 0, 1, 2, \dots, n-m-2$$

and $\mathbf{c}\mathbf{A}^{n-m-1}\mathbf{b} \neq 0$. Or, equivalently, the difference between the degrees of the denominator and numerator of $\hat{g}(s)$ is $\alpha = n-m$ if and only if

$$\mathbf{c}\mathbf{A}^{\alpha-1}\mathbf{b} \neq 0 \quad \text{and} \quad \mathbf{c}\mathbf{A}^i\mathbf{b} = 0$$

for $i = 0, 1, 2, \dots, \alpha-2$.

4.20 Find fundamental matrices and state transition matrices for

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 \\ 0 & t \end{bmatrix} \mathbf{x}(t)$$

and

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -1 & e^{2t} \\ 0 & -1 \end{bmatrix} \mathbf{x}(t)$$

4.21 Consider a system with proper transfer function $\hat{g}(s)$ and its irreducible realization $\{A, b, c, d\}$. Show that if the input $u(t)$ applied to the system is of the form $e^{\lambda t}$, where λ is not a pole of $\hat{g}(s)$, then the output due to the initial state $x(0) = -(A - \lambda I)^{-1} b$ and the input $u(t) = e^{\lambda t}$ is equal to $y(t) = \hat{g}(\lambda) e^{\lambda t}$ for $t \geq 0$. Hint: use the following identity to prove the result,

$$(sI - A)^{-1} (s - \lambda)^{-1} = (\lambda I - A)^{-1} (s - \lambda)^{-1} + (sI - A)^{-1} (A - \lambda I)^{-1}$$

- 4.22** For the following system find the initial conditions such that for the input $u(t) = e^{-4t}$ the output $y(t)$ will be of the form e^{-4t} without containing any transient.

$$\begin{aligned}\dot{x}(t) &= \begin{bmatrix} -1 & 1 \\ -0.5 & 0 \end{bmatrix}x(t) + \begin{bmatrix} 0 \\ 0.5 \end{bmatrix}u(t) \\ y(t) &= [1 \ 0]x(t)\end{aligned}$$

Hint: Use the result of Problem 4.5 (New Problem).

- 4.23** Find the solution to the following time-varying system,

$$\dot{x}(t) = \begin{bmatrix} 1 & \cos t \\ 0 & 0 \end{bmatrix}x(t), \quad x(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

- 4.24** Show that

$$\frac{d}{dt} \left[\Phi^{-1}(t, t_0) \right] = -\Phi^{-1}(t, t_0) \left[\frac{d}{dt} \Phi(t, t_0) \right] \Phi^{-1}(t, t_0)$$

- 4.25** Find the state transition matrix of following time-varying equation

$$\begin{aligned}\dot{x}_1(t) &= tx_1(t) \\ \dot{x}_2(t) &= 2tx_1(t) - tx_2(t)\end{aligned}$$

and find the zero-input response of the system if $x_1(0) = x_2(0) = 1$.

- 4.26** Show that $\partial \Phi(t_0, t)/\partial t = -\Phi(t_0, t)\mathbf{A}(t)$.

- 4.27** Given

$$\mathbf{A}(t) = \begin{bmatrix} a_{11}(t) & a_{12}(t) \\ a_{21}(t) & a_{22}(t) \end{bmatrix}$$

show that

$$\det \Phi(t, t_0) = \exp \left[\int_{t_0}^t (a_{11}(\tau) + a_{22}(\tau)) d\tau \right]$$

- 4.28** Let

$$\Phi(t, t_0) = \begin{bmatrix} \Phi_{11}(t, t_0) & \Phi_{12}(t, t_0) \\ \Phi_{21}(t, t_0) & \Phi_{22}(t, t_0) \end{bmatrix}$$

be the state transition matrix of

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} \mathbf{A}_{11}(t) & \mathbf{A}_{12}(t) \\ \mathbf{0} & \mathbf{A}_{22}(t) \end{bmatrix} \mathbf{x}(t)$$

Show that $\Phi_{21}(t, t_0) = \mathbf{0}$ for all t and t_0 and that $(\partial/\partial t)\Phi_{ii}(t, t_0) = \mathbf{A}_{ii}\Phi_{ii}(t, t_0)$, for $i = 1, 2$.

- 4.29** Find the state transition matrix of

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -\sin t & 0 \\ 0 & -\cos t \end{bmatrix} \mathbf{x}(t)$$

- 4.30** Verify that $\mathbf{X}(t) = e^{\mathbf{A}t} \mathbf{C} e^{\mathbf{B}t}$ is the solution of

$$\dot{\mathbf{X}}(t) = \mathbf{A}\mathbf{X}(t) + \mathbf{X}(t)\mathbf{B}, \quad \mathbf{X}(0) = \mathbf{C}$$

- 4.31 Show that if $\dot{\mathbf{A}}(t) = \mathbf{A}_1 \mathbf{A}(t) - \mathbf{A}(t) \mathbf{A}_1$, then

$$\mathbf{A}(t) = e^{\mathbf{A}_1 t} \mathbf{A}(0) e^{-\mathbf{A}_1 t}$$

Show also that the eigenvalues of $\mathbf{A}(t)$ is independent of t .

- 4.32 Find an equivalent time-invariant state-space equation of the equation in Problem 4.29.
- 4.33 Transform a time-invariant $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ into $(\mathbf{0}, \bar{\mathbf{B}}(t), \bar{\mathbf{C}}(t))$ by a time-varying equivalence transformation.
- 4.34 Find a time-varying realization and a time-invariant realization of the impulse response $g(t) = t^2 e^{\lambda t}$.
- 4.35 Find a realization of $g(t, \tau) = \sin t (e^{-(t-\tau)}) \cos \tau$. Is it possible to find a time-invariant state-space equation realization?

Stability

5.1 Introduction

Systems are designed to perform some tasks or to process signals. If a system is not stable, the system may burn out, disintegrate, or saturate when a signal, no matter how small, is applied. Therefore, an unstable system is useless in practice and stability is a basic requirement for all systems. In addition to stability, systems must meet other requirements, such as to track desired signals or to suppress noise, to be really useful in practice.

The response of linear systems can always be decomposed as the zero-state response and the zero-input response. It is customary to study stabilities of these two responses separately. We will introduce BIBO (bounded-input bounded-output) stability for zero-state responses and marginal and asymptotic stabilities for zero-input responses. In the study of zero-state responses, we study first SISO systems and then apply the result to MIMO systems. In the study of zero-input responses, there is no difference between the two types of systems. Thus we study directly MIMO systems. We study first the time-invariant case and then the time-varying case.

5.2 Input–Output Stability of LTI Systems

Consider a SISO linear time-invariant (LTI) system described by, as derived in (2.5),

$$y(t) = \int_0^t g(t-\tau)u(\tau) d\tau = \int_0^t g(\tau)u(t-\tau) d\tau \quad (5.1)$$

where $g(t)$ is the impulse response or the output excited by an impulse input applied at $t = 0$. Recall that in order to be describable by (5.1), the system must be linear, time-invariant and causal. In addition, the system must be initially relaxed at $t = 0$.

An input $u(t)$ is said to be *bounded* if $u(t)$ does not grow to positive or negative infinity or, equivalently, there exists a constant u_m such that

$$|u(t)| \leq u_m < \infty \quad \text{for all } t \geq 0$$

A system is said to be *BIBO stable* (bounded-input bounded output stable) if every bounded input excites a bounded output. This stability is defined for zero-state responses and is applicable only if a system is initially relaxed.

► **Theorem 5.1**

A SISO system described by (5.1) is BIBO stable if and only if $g(t)$ is absolutely integrable in $[0, \infty)$ or

$$\int_0^\infty |g(t)| dt \leq M < \infty$$

for some constant M .

⇒ **Proof:** First we show that if $g(t)$ is absolutely integrable, then every bounded input excites a bounded output. Let $u(t)$ be an arbitrary input with $|u(t)| \leq u_m < \infty$ for all $t \geq 0$. Then we have

$$|y(t)| = \left| \int_0^t g(\tau)u(t-\tau) d\tau \right| \leq \int_0^t |g(\tau)||u(t-\tau)| d\tau$$

The inequality follows from the fact that the integrand in the first integration may be positive or negative and its integration may cancel out, whereas the integrand in the second integration is all positive and there is no more cancellation in its integration. Substituting $|u(t)| \leq u_m$ for all $t \geq 0$ yields

$$|y(t)| \leq u_m \int_0^t |g(\tau)| d\tau \leq u_m \int_0^\infty |g(\tau)| d\tau \leq u_m M$$

for all $t \geq 0$. Thus the output is bounded. Next we show that if $g(t)$ is not absolutely integrable, then the system is not BIBO stable. If $g(t)$ is not absolutely integrable, then for any arbitrarily large N , there exists a t_1 such that

$$\int_0^{t_1} |g(\tau)| d\tau > N$$

Consider the input $u(t)$, for t in $[0, t_1]$, defined by

$$u(t_1 - t) = \begin{cases} 1 & \text{if } g(t) \geq 0 \\ -1 & \text{if } g(t) < 0 \end{cases}$$

It is bounded. The output at t_1 excited by this input equals

$$y(t_1) = \int_0^{t_1} g(\tau)u(t_1 - \tau) d\tau = \int_0^{t_1} |g(\tau)| d\tau > N$$

Because $y(t_1)$ can be arbitrarily large, we conclude that the bounded input excites an unbounded output. This completes the proof of Theorem 5.1. Q.E.D.

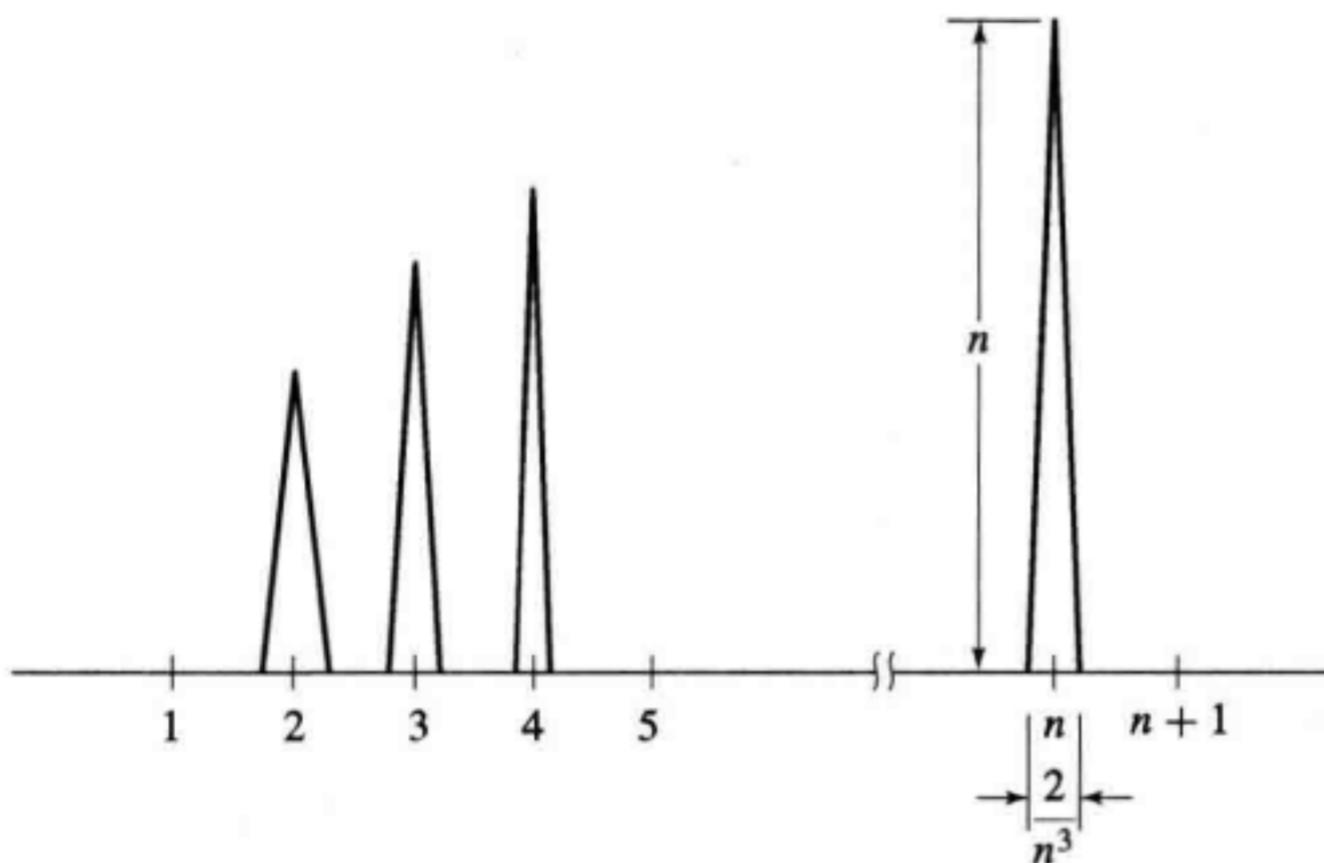


FIGURE 5.1 Function.

A function that is absolutely integrable may not be bounded or may not approach zero as $t \rightarrow \infty$. Indeed, consider the function defined by

$$f(t-n) = \begin{cases} n + (t-n)n^4 & \text{for } (-1/n^3) < (t-n) \leq 0 \\ n - (t-n)n^4 & \text{for } 0 \leq (t-n) \leq (1/n^3) \end{cases}$$

for $n = 2, 3, \dots$ and plotted in Fig. 5.1. The area under each triangle is $1/n^2$. Thus the absolute integration of the function equals $\sum_{n=2}^{\infty} (1/n^2) < \infty$. This function is absolutely integrable, but is not bounded and does not approach zero as $t \rightarrow \infty$.

Example 5.2.1 Consider the positive feedback system shown in Fig. 2.5(a). Its impulse response was computed in (2.6) as

$$g(t) = \sum_{i=1}^{\infty} a^i \delta(t-i)$$

where the gain a can be positive or negative. The impulse is defined as the limit of the pulse in Fig. 2.3 and can be considered to be positive. Thus we have

$$|g(t)| = \sum_{0}^{\infty} |a|^i \delta(t-i)$$

and, using $\int_0^{\infty} \delta(t-i) dt = 1$ for all $i \geq 1$,

$$\int_0^{\infty} |g(t)| dt = \sum_{0}^{\infty} |a|^i = \begin{cases} \infty & \text{if } |a| \geq 1 \\ 1/(1-|a|) < \infty & \text{if } |a| < 1 \end{cases}$$

Thus we conclude that the positive feedback system in Fig. 2.5(a) is BIBO stable if and only if the gain a has a magnitude less than 1.

The transfer function of the system was computed in Example 2.3.4 as

$$\hat{g}(s) = \frac{ae^{-s}}{1 - ae^{-s}} \quad (5.2)$$

It is an irrational function of s and the system is distributed.

► Theorem 5.2

If a system with impulse response $g(t)$ is BIBO stable, then, as $t \rightarrow \infty$,

1. the output excited by $u(t) = a$, for all $t \geq 0$, approaches $\hat{g}(0) \cdot a$,
2. the output excited by $u(t) = \cos \omega_0 t$, for all $t \geq 0$, approaches

$$|\hat{g}(j\omega_0)| \cos(\omega_0 t + \arg \hat{g}(j\omega_0))$$

3. the output excited by $u(t) = \sin \omega_0 t$, for all $t \geq 0$, approaches

$$|\hat{g}(j\omega_0)| \sin(\omega_0 t + \arg \hat{g}(j\omega_0))$$

where $\hat{g}(s)$ is the Laplace transform of $g(t)$ or

$$\hat{g}(s) := \mathcal{L}[g(t)] := \int_{t=0}^{\infty} g(t)e^{-st} dt = \int_{\tau=0}^{\infty} g(\tau)e^{-s\tau} d\tau \quad (5.3)$$



Proof: If $u(t) = a$ for all $t \geq 0$, then (5.1) becomes

$$y(t) = \int_{\tau=0}^t g(\tau)u(t-\tau) d\tau = a \int_{\tau=0}^t g(\tau) d\tau$$

which implies, as $t \rightarrow \infty$,

$$y(t) \rightarrow a \int_{\tau=0}^{\infty} g(\tau) d\tau = a\hat{g}(0)$$

where we have used (5.3) with $s = 0$. This establishes the first part of Theorem 5.2.

Although the second and third parts can be directly proved, it is much simpler to prove them using $u(t) = e^{j\omega_0 t}$. Note that $u(t) = e^{j\omega_0 t}$ is a complex-valued function and all discussion is still applicable. If $u(t) = e^{j\omega_0 t}$ for all $t \geq 0$, then (5.1) becomes

$$y(t) = \int_0^t g(\tau)u(t-\tau) d\tau = \int_0^t g(\tau)e^{j\omega_0(t-\tau)} d\tau = e^{j\omega_0 t} \int_0^t g(\tau)e^{-j\omega_0 \tau} d\tau$$

which implies, as $t \rightarrow \infty$,

$$y(t) \rightarrow e^{j\omega_0 t} \int_0^{\infty} g(\tau)e^{-j\omega_0 \tau} d\tau = \hat{g}(j\omega_0)e^{j\omega_0 t}$$

where we have used (5.3) with $s = j\omega_0$. Note that even though $g(t)$ is real-valued, $\hat{g}(j\omega_0)$ is generally complex-valued and can be expressed as $\hat{g}(j\omega_0) =$

$A(\omega_0)e^{j\theta(\omega_0)}$, where $A(\omega_0) := |\hat{g}(j\omega_0)|$ and $\theta(\omega_0) := \arg \hat{g}(j\omega_0)$. Now if $e^{j\omega_0 t} = \cos(\omega_0 t) + j \sin(\omega_0 t)$ for all $t \geq 0$, then the output approaches, as $t \rightarrow \infty$,

$$\begin{aligned} y(t) &\rightarrow \hat{g}(j\omega_0)e^{j\omega_0 t} = A(\omega_0)e^{j\theta(\omega_0)}e^{j\omega_0 t} = A(\omega_0)e^{j(\omega_0 t + \theta(\omega_0))} \\ &= A(\omega_0)\cos(\omega_0 + \theta(\omega_0)) + jA(\omega_0)\sin(\omega_0 + \theta(\omega_0)) \end{aligned}$$

The real part of $e^{j\omega_0 t}$ excites the real part of $y(t)$, which is the second part of the theorem. The third part follows from the imaginary parts. Q.E.D.

In this theorem, the stability condition is essential. For example, if $g(t) = 4e^{2t}$, for $t \geq 0$, then $g(t)$ is not absolutely integrable. Thus the system is not BIBO stable and Theorem 5.2 is not applicable. Note that the Laplace transform of $g(t) = 4e^{2t}$ is well defined and equals $\hat{g}(s) = 4/(s - 2)$. If we apply $u(t) = 1$, then its output will grow unbounded rather than approach $\hat{g}(0) = 4/(-2) = -2$. If we apply $\sin 2t$, its output will also grow unbounded rather than approach $4|\hat{g}(j2)| \sin(2t + \arg \hat{g}(j2))$, where

$$\hat{g}(j2) = \frac{4}{j2 - 2} = \frac{2}{-1 + j1} = \frac{2}{1.414e^{j3\pi/4}} = 1.414e^{-j3\pi/4}$$

with $|\hat{g}(j2)| = 1.414$ and $\arg \hat{g}(j2) = -3\pi/4$. On the other hand, if an LTI system is BIBO stable, and if we apply a step input with amplitude a , the output will approach a step function with amplitude $a\hat{g}(0)$. If we apply a sinusoidal input, then the output will approach a sinusoidal function of the *same* frequency with its amplitude and phase modified by $\hat{g}(j\omega_0)$.

Consider a system with transfer function $\hat{g}(s)$. We call $\hat{g}(j\omega)$ the *frequency response* of the system, we call its magnitude $|\hat{g}(j\omega)|$ the *magnitude response*, and we call its phase $\arg \hat{g}(j\omega)$ the *phase response*. If the system is not BIBO stable, its frequency response has no physical meaning. If the system is BIBO stable, then a sinusoidal input with frequency ω_0 will eventually excite a sinusoidal function of frequency ω_0 with its magnitude and phase modified by the frequency response at ω_0 . Filter design is essentially based on Theorem 5.2: To find a BIBO system with a desired or specified frequency response.

Theorem 5.1 is applicable to LTI lumped or distributed systems. If an LTI system is lumped as well, then it can also be described by a rational transfer function. We now state the BIBO stability condition in terms of proper rational transfer functions. We first divide the complex s -plane into three parts: right half-plane (RHP), left half-plane (LHP), and the imaginary or $j\omega$ axis. Note that the right and left half-planes do not include the $j\omega$ axis.

► Theorem 5.3

A SISO system with proper rational transfer function $\hat{g}(s)$ is BIBO stable if and only if every pole of $\hat{g}(s)$ has a negative real part or, equivalently, lies inside the left half- s -plane.



Proof: We first establish that the response (inverse Laplace transform) of a pole, real or complex, simple or repeated, approaches 0 as $t \rightarrow \infty$ and is absolutely integrable if and only if the pole has a negative real part. If $\hat{g}(s)$ has a pole p_i with multiplicity m_i , then the inverse Laplace transform of $\hat{g}(s)$ or the impulse

response contains the factors

$$e^{p_i t}, \quad te^{p_i t}, \dots, t^{m_i-1} e^{p_i t} \quad (5.4)$$

We first assume p_i to be real. If p_i is 0, then the first term equals 1 for all $t \geq 0$, and the rest grow unbounded. If p_i is positive, then every term in (5.4) grows unbounded. In conclusion, if $p_i \geq 0$, then every term in (5.4) will not approach zero as $t \rightarrow \infty$ and is not absolutely integrable in $[0, \infty)$. Next consider p_i to be real and negative such as $p_i = -0.01$. Then we have $e^{-0.01t} \rightarrow 0$ as $t \rightarrow \infty$ and

$$\begin{aligned} \int_0^\infty |e^{-0.01t}| dt &= \int_0^\infty e^{-0.01t} dt = \frac{1}{-0.01} e^{-0.01t} \Big|_{t=0}^\infty = \frac{1}{-0.01} [0 - 1] \\ &= \frac{-1}{-0.01} = 100 \end{aligned}$$

For $te^{-0.01t}$, we have, using l'Hôpital's rule,

$$\lim_{t \rightarrow \infty} \left[te^{-0.01t} = \frac{t}{e^{0.01t}} = \frac{1}{0.01e^{0.01t}} \right] = \frac{1}{\infty} = 0$$

and, using an integration table,

$$\int_0^\infty |te^{-0.01t}| dt = \int_0^\infty te^{-0.01t} dt = \frac{(-0.01t - 1)e^{-0.01t}}{(-0.01)^2} \Big|_{t=0}^\infty = \frac{0 - (-1)}{10^{-4}} = 10^4$$

Proceeding forward, we can conclude that, for any positive integer m_i and any negative real number p_i , the function $t^{m_i} e^{p_i t}$ approaches 0 as $t \rightarrow \infty$ and is absolutely integrable. This is so because the exponential function $e^{p_i t}$ approaches zero much faster than the rate of the polynomial t^{m_i} approaching ∞ . Thus as $t \rightarrow \infty$, the exponential function dominates over the polynomial.

The preceding discussion is in fact applicable to complex poles. Let $p_i = \alpha_i + j\beta_i$, where α_i and β_i are real numbers. Because $|e^{j\beta_i t}| = 1$ for any real β_i and all t , we have, for any positive integer m and all $t \geq 0$,

$$|t^m e^{p_i t}| = t^m |e^{(\alpha_i + j\beta_i)t}| = t^m |e^{\alpha_i t}| |e^{j\beta_i t}| = t^m e^{\alpha_i t}$$

where we have used $e^{\alpha_i t} \geq 0$ for any positive or negative real α_i . Thus the absolute integrability of a complex pole is determined by the real part of the pole and is independent of its imaginary part. Because the preceding term has the form in (5.4), all discussion for real poles is applicable to complex poles if we consider only their real parts.¹ Thus the response of a pole, real or complex, simple or repeated, approaches 0 as $t \rightarrow \infty$ and is absolutely integrable if the pole has a negative real part or lies inside the left half- s -plane.

¹Here we permit complex numbers to simplify the discussion. Otherwise we must combine the response of a complex pole with the response of its complex conjugate. The response of a pair of complex-conjugate poles $(\alpha_i \pm j\beta_i)$ with multiplicity $m+1$ can be expressed as $t^m e^{\alpha_i t} \sin(\beta_i t)$, which implies $|t^m e^{\alpha_i t} \sin(\beta_i t)| \leq t^m e^{\alpha_i t}$, for all $t \geq 0$. Thus all discussion still applies.

If $\hat{g}(s)$ has a number of poles, then its impulse response $g(t)$ is a linear combination of the impulse responses $g_i(t)$ of all poles or

$$g(t) = \sum_i a_i g_i(t) + b\delta(t)$$

for some constants a_i and b . The constant b is 0 if $\hat{g}(s)$ is strictly proper and nonzero if $\hat{g}(s)$ is biproper. The term $b\delta(t)$ is absolutely integrable and is not associated with any pole. Thus it will be dropped in the next equation. Because

$$\int_0^\infty |g(t)| dt = \int_0^\infty \left| \sum_i a_i g_i(t) \right| dt \leq \int_0^\infty \sum_i |a_i g_i(t)| dt = \sum_i |a_i| \int_0^\infty |g_i(t)| dt$$

if every $g_i(t)$ is absolutely integrable, so is $g(t)$. Consequently the system is BIBO stable following Theorem 5.1. This completes the proof of Theorem 5.3. Q.E.D.

The poles of a rational transfer function $\hat{g}(s) = N(s)/D(s)$ are the roots of $D(s)$ if $N(s)$ and $D(s)$ are coprime, therefore we check only the roots of a polynomial in employing Theorem 5.3. A polynomial is defined to be *CT stable* if all its roots have negative real parts. If the denominator $D(s)$ is CT stable, then the transfer function is BIBO stable. Whether or not $D(s)$ is CT stable can be checked by computing all its roots by calling the MATLAB function `roots`. It can also be checked using the *Routh test* without computing the roots. See References 7 and 10.

► Corollary 5.3

A SISO system with proper rational transfer function $\hat{g}(s)$ is BIBO stable if and only if its impulse response $g(t)$ approaches 0 as $t \rightarrow \infty$.

The impulse response of a proper rational transfer function is a linear combination of the responses of all poles. It was shown in the proof of Theorem 5.3 that the responses approach 0 as $t \rightarrow \infty$ if and only if every pole of $\hat{g}(s)$ lies inside the left half-s-plane or, equivalently, the system is BIBO stable. Thus Theorem 5.3 implies directly Corollary 5.3.

We discuss some differences between Theorems 5.1 and 5.3. We have defined λ as a pole of the proper rational transfer function $\hat{g}(s)$ if $\hat{g}(\lambda) = \infty$ or $-\infty$. Can we use the same definition for irrational transfer functions? For $\hat{g}(s)$ in (5.2) with $a = 1$, if $s = jm(2\pi)$ for all integers m , then

$$\hat{g}(jm(2\pi)) = \frac{e^{-s}}{1 - e^{-s}} \Big|_{s=jm(2\pi)} = \frac{1}{1 - 1} = \infty$$

Thus the irrational transfer function in (5.2) with $a = 1$ has an infinite number of poles if we adopt the preceding definition. A pole p_i in a rational transfer function will generate the time function $e^{p_i t}$ in its impulse response. It is, however, not clear what is the implication of a pole in an irrational transfer function. Theorem 5.3 states that a proper rational transfer function is BIBO stable if and only if every pole lies inside the left half-s-plane or, equivalently, if and only if there is no pole inside the right half-s-plane or on the $j\omega$ axis. The irrational transfer function in (5.2) with $a = 1$ has poles on the

$j\omega$ axis and is not BIBO stable. For this example, Theorem 5.3 is applicable. But it is not true in general. For example, consider an LTI system with impulse response

$$g_1(t) = \sin(t^2/2)$$

for $t \geq 0$ and $g_1(t) = 0$ for $t < 0$. Its Laplace transform $\hat{g}_1(s)$ is not a rational function of s , and the system is distributed. Computing $\hat{g}_1(s)$ is difficult, but it is shown in Reference 17 that $\hat{g}_1(s)$ meets some differential equation and is analytic (has no singular point or pole) inside the right half- s -plane and on the $j\omega$ axis. Thus the system would be BIBO stable if Theorem 5.3 were applicable. The system is actually not BIBO stable according to Theorem 5.1 because $g_1(t)$ is not absolutely integrable.

Consider an LTI system with impulse response

$$g_2(t) = \frac{1}{t+1}$$

for $t \geq 0$. It approaches 0 as $t \rightarrow \infty$. The system is BIBO stable according to Corollary 5.3 but is actually not BIBO stable because

$$\int_0^\infty |g_2(t)| dt = \int_0^\infty \frac{1}{t+1} dt = \ln(t+1)|_0^\infty = \infty$$

(Theorem 5.1). There is no inconsistency here because Corollary 5.3 is actually not applicable because the Laplace transform of $1/(t+1)$ is not a rational function of s . In conclusion, the concept of poles for rational transfer functions (LTI lumped systems) may not be extendable to irrational transfer functions (LTI distributed systems). Nor may Theorem 5.3 and its corollary be applicable to the latter. Thus the study of LTI distributed systems is much more complicated than the study of LTI lumped systems.

Corollary 5.3 has an important implication; it can be used to check the BIBO stability of a system by one measurement. Note that it is not possible to generate an impulse as an input. Let us apply the input $u(t) = 1$, for t in $[0, t_0]$ with an arbitrary $t_0 > 0$, and $u(t) = 0$, for all $t > t_0$. This input will excite *all* poles of a proper rational transfer function. See Problem 5.7. Once the input is removed, the output consists of only the responses of the poles. Thus the system is BIBO stable if and only if the output approaches zero as $t \rightarrow \infty$. See also Problem 5.8. In general, if we apply an arbitrary input of finite duration to an LTI lumped system, then the system is BIBO stable if and only if the response vanishes.

For MIMO systems, we have the following results.

► Theorem 5.M1

A MIMO system with impulse response matrix $\mathbf{G}(t) = [g_{ij}(t)]$ is BIBO stable if and only if every $g_{ij}(t)$ is absolutely integrable in $[0, \infty)$.

► Theorem 5.M3

A MIMO system with proper rational transfer matrix $\hat{\mathbf{G}}(s) = [\hat{g}_{ij}(s)]$ is BIBO stable if and only if every pole of every $\hat{g}_{ij}(s)$ has a negative real part.

We now discuss the BIBO stability of state-space equations. Consider

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t)\end{aligned}\tag{5.5}$$

Its transfer matrix is

$$\hat{\mathbf{G}}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$$

Thus Equation (5.5) or, to be more precise, the zero-state response of (5.5) is BIBO stable if and only if every pole of $\hat{\mathbf{G}}(s)$ has a negative real part. Recall that every pole of every entry of $\hat{\mathbf{G}}(s)$ is called a pole of $\hat{\mathbf{G}}(s)$.

We discuss the relationship between the poles of $\hat{\mathbf{G}}(s)$ and the eigenvalues of \mathbf{A} . Because of

$$\hat{\mathbf{G}}(s) = \frac{1}{\det(s\mathbf{I} - \mathbf{A})} \mathbf{C}[\text{Adj}(s\mathbf{I} - \mathbf{A})]\mathbf{B} + \mathbf{D}\tag{5.6}$$

every pole of $\hat{\mathbf{G}}(s)$ is an eigenvalue of \mathbf{A} . Thus if every eigenvalue of \mathbf{A} has a negative real part, then every pole has a negative real part and (5.5) is BIBO stable. On the other hand, because of possible cancelation in (5.6), not every eigenvalue is a pole. Thus, even if \mathbf{A} has some eigenvalues with zero or positive real part, (5.5) may still be BIBO stable, as the next example shows.

Example 5.2.2 Consider the state-space equation

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \begin{bmatrix} -2 & 5 \\ 0 & 3 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 4 \\ 0 \end{bmatrix} \mathbf{u}(t) \\ \mathbf{y}(t) &= [7 \quad 8] \mathbf{x}(t) + 1.5\mathbf{u}(t)\end{aligned}$$

The matrix \mathbf{A} is upper triangular and its eigenvalues are simply its diagonal entries or -2 and 3 . We compute

$$[s\mathbf{I} - \mathbf{A}]^{-1} = \begin{bmatrix} s+2 & -5 \\ 0 & s-3 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{s+2} & \frac{5}{(s+2)(s-3)} \\ 0 & \frac{1}{s-3} \end{bmatrix}$$

Thus the transfer function of the state-space equation is

$$\begin{aligned}\hat{g}(s) &= \mathbf{c}[s\mathbf{I} - \mathbf{A}]^{-1}\mathbf{b} + d = [7 \quad 8] \begin{bmatrix} \frac{1}{s+2} & \frac{5}{(s+2)(s-3)} \\ 0 & \frac{1}{s-3} \end{bmatrix} \begin{bmatrix} 4 \\ 0 \end{bmatrix} + 1.5 \\ &= [7 \quad 8] \begin{bmatrix} \frac{4}{s+2} \\ 0 \end{bmatrix} + 1.5 = \frac{28}{s+2} + 1.5 = \frac{1.5s+31}{s+2}\end{aligned}$$

It has only one pole at -2 ; thus the state-space equation is BIBO stable even though \mathbf{A} has eigenvalue 3 which lies inside the right half- s -plane. We see that not every eigenvalue of \mathbf{A} appears as a pole of the transfer function. Its reason will be given in the next chapter.

5.3 Discrete-Time Case

Consider a discrete-time SISO system described by

$$y[k] = \sum_{m=0}^k g[k-m]u[m] = \sum_{m=0}^k g[m]u[k-m] \quad (5.7)$$

where $g[k]$ is the impulse response sequence or the output sequence excited by an impulse sequence applied at $k = 0$. Recall that in order to be describable by (5.7), the discrete-time system must be linear, time-invariant, and causal. In addition, the system must be initially relaxed at $k = 0$.

An input sequence $u[k]$ is said to be *bounded* if $u[k]$ does not grow to positive or native infinity or there exists a constant u_m such that

$$|u[k]| \leq u_m < \infty \quad \text{for } k = 0, 1, 2, \dots$$

A system is said to be *BIBO stable* (bounded-input bounded output stable) if every bounded input sequence excites a bounded output sequence. This stability is defined for the zero-state response and is applicable only if the system is initially relaxed.

► Theorem 5.D1

A discrete-time SISO system described by (5.7) is BIBO stable if and only if $g[k]$ is absolutely summable in $[0, \infty)$ or

$$\sum_{k=0}^{\infty} |g[k]| \leq M < \infty$$

for some constant M .



Proof: First we show that if $g[k]$ is absolutely summable, then every bounded input excites a bounded output. Let $u[k]$ be an arbitrary input sequence with $|u[k]| \leq u_m < \infty$ for all $k \geq 0$. Then we have

$$|y[k]| = \left| \sum_{m=0}^k g[m]u[k-m] \right| \leq \sum_{m=0}^k |g[m]| |u[k-m]|$$

The inequality follows from the fact that $g[m]u[k-m]$ may be positive or negative and their summation may cancel out, whereas $|g[m]| |u[k-m]|$ are all positive and there is no more cancelation in their summation. Substituting $|u[k]| \leq u_m$, for all $k \geq 0$ yields

$$|y[k]| \leq u_m \sum_{m=0}^k |g[m]| \leq u_m \sum_{m=0}^{\infty} |g[m]| \leq u_m M$$

Thus the output is bounded. Next we show that if $g[k]$ is not absolutely summable, then the system is not BIBO stable. If $g[k]$ is not absolutely summable, then for

any arbitrarily large N , there exists a k_1 such that

$$\sum_{m=0}^{k_1} |g[m]| > N$$

Consider the input $u[k]$, for k in $[0, k_1]$, defined by

$$u[k_1 - k] = \begin{cases} 1 & \text{if } g[k] \geq 0 \\ -1 & \text{if } g[k] < 0 \end{cases}$$

It is bounded. The output at k_1 excited by this input equals

$$y[k_1] = \sum_{m=0}^{k_1} g[m]u[k_1 - m] = \sum_{m=0}^{k_1} |g[m]| > N$$

Because $y[k_1]$ can be arbitrarily large, we conclude that the bounded input excites an unbounded output. This completes the proof of Theorem 5.D1. Q.E.D.

We see that the proof of Theorem 5.D1 is similar to the one of Theorem 5.1. There is, however, a significant difference between the CT and DT case. A CT function $g(t)$ which is absolutely integrable may not be bounded and may not approach 0 as $t \rightarrow \infty$ as shown in Fig. 5.1. Every DT sequence $g[k]$ which is absolutely summable, however, is bounded and approaches 0 as $k \rightarrow \infty$. See Reference 10, Sec. 3.5.

► Theorem 5.D2

If a discrete-time system with impulse response sequence $g[k]$ is BIBO stable, then, as $k \rightarrow \infty$,

1. the output excited by $u[k] = a$ for all $k \geq 0$ approaches $\hat{g}(1) \cdot a$,
2. the output excited by $u[k] = \cos \omega_0 k$ for all $k \geq 0$ approaches

$$|\hat{g}(e^{j\omega_0})| \cos(\omega_0 k + \hat{\phi}(e^{j\omega_0}))$$

3. the output excited by $u[k] = \sin \omega_0 k$ for all $k \geq 0$ approaches

$$|\hat{g}(e^{j\omega_0})| \sin(\omega_0 k + \hat{\phi}(e^{j\omega_0}))$$

where $\hat{g}(z)$ is the z -transform of $g[k]$ or

$$\hat{g}(z) = \mathcal{Z}[g[k]] = \sum_{k=0}^{\infty} g[k]z^{-k} = \sum_{m=0}^{\infty} g[m]z^{-m} \quad (5.8)$$

⇒ Proof: If $u[k] = a$ for all $k \geq 0$, then (5.7) becomes

$$y[k] = \sum_{m=0}^k g[m]u[k-m] = a \sum_{m=0}^k g[m]$$

which implies, as $k \rightarrow \infty$,

$$y[k] \rightarrow a \sum_{m=0}^{\infty} g[m] = a\hat{g}(1)$$

where we have used (5.8) with $z = 1$. This establishes the first part of Theorem 5.D2. Although the second and third parts can be directly established, it is simpler to derive them using the complex exponential sequence $u[k] = e^{j\omega_0 k}$, for all $k \geq 0$. Substituting this into (5.7) yields

$$\begin{aligned} y[k] &= \sum_{m=0}^k g[m]u[k-m] = \sum_{m=0}^k g(m)e^{j\omega_0(k-m)} \\ &= e^{j\omega_0 k} \sum_{m=0}^k g[m]e^{-j\omega_0 m} \end{aligned}$$

which implies, as $k \rightarrow \infty$,

$$y[k] \rightarrow e^{j\omega_0 k} \sum_{m=0}^{\infty} g[m]e^{-j\omega_0 m} = e^{j\omega_0 k} \hat{g}(e^{j\omega_0})$$

where we have used (5.8) with $z = e^{j\omega_0}$. Note that even though $g[k]$ is real-valued, $\hat{g}(e^{j\omega_0})$ is generally complex-valued and can be expressed as $\hat{g}(e^{j\omega_0}) = A(\omega_0)e^{j\theta(\omega_0)}$, where $A(\omega_0) := |\hat{g}(e^{j\omega_0})|$ and $\theta(\omega_0) := \arg(\hat{g}(e^{j\omega_0}))$. Thus if $e^{j\omega_0 k} = \cos(\omega_0 k) + j \sin(\omega_0 k)$ for all $k \geq 0$, then the output approaches, as $k \rightarrow \infty$,

$$\begin{aligned} y[k] &\rightarrow \hat{g}(e^{j\omega_0})e^{j\omega_0 k} = A(\omega_0)e^{j\theta(\omega_0)}e^{j\omega_0 k} = A(\omega_0)e^{j(\omega_0 k + \theta(\omega_0))} \\ &= A(\omega_0)\cos(\omega_0 k + \theta(\omega_0)) + jA(\omega_0)\sin(\omega_0 k + \theta(\omega_0)) \end{aligned}$$

The real part of $e^{j\omega_0 k}$ excites the real part of $y[k]$ which is the second part of the theorem. The third part follows from the imaginary parts. Q.E.D.

In this theorem, the stability condition is essential. For example, if $g[k] = 4 \cdot 2^k$, for $k \geq 0$, then $g[k]$ is not absolutely summable. Thus the system is not BIBO stable and Theorem 5.D2 is not applicable. Note that the z -transform of $g[k] = 4 \cdot 2^k$ is well-defined and equals $\hat{g}(z) = 4z/(z - 2)$. If we apply $u[k] = 1$, then its output will grow unbounded rather than approaches $\hat{g}(1) = 4/(-1) = -4$. If we apply $\sin 2k$, its output will also grow unbounded rather than approaches $4|\hat{g}(e^{j2})| \sin(2k + \arg(\hat{g}(e^{j2})))$ where

$$\hat{g}(e^{j2}) = \frac{4e^{j2}}{e^{j2} - 2} = 1.55e^{-0.78j}$$

with $|\hat{g}(e^{j2})| = 1.55$ and $\arg(\hat{g}(e^{j2})) = -0.78$. On the other hand, if a discrete-time LTI system is BIBO stable, and if we apply a step input with amplitude a , the output will approaches a step sequence with amplitude $a\hat{g}(1)$. If we apply a sinusoidal input, then the output will approach a sinusoidal sequence of the *same* frequency with its amplitude and phase modified by $\hat{g}(e^{j\omega})$.

Consider a system with transfer function $\hat{g}(z)$. We call $\hat{g}(e^{j\omega})$ the *frequency response* of the system, we call its magnitude $|\hat{g}(e^{j\omega})|$ the *magnitude response*, and we call its phase $\angle \hat{g}(e^{j\omega})$ the *phase response*. If the system is not BIBO stable, its frequency response has no physical meaning. If the system is BIBO stable, then a sinusoidal input sequence with frequency ω_0 will eventually excite a sinusoidal sequence of frequency ω_0 with its magnitude and phase modified by the frequency response at ω_0 . Digital filter design is essentially based on Theorem 5.D2: To find a discrete-time BIBO system with a desired or specified frequency response.

Theorem 5.D1 is applicable to LTI lumped or distributed systems. If an LTI system is lumped as well, then it can also be described by a rational transfer function. We now state the BIBO stability condition in terms of proper rational transfer functions. We first divide the complex z -plane into three parts: inside the unit circle, the unit circle, and outside the unit circle.

► Theorem 5.D3

A discrete-time SISO system with proper rational transfer function $\hat{g}(z)$ is BIBO stable if and only if every pole of $\hat{g}(z)$ has a magnitude less than 1 or, equivalently, lies inside the unit circle of the z -plane.

In the proof of Theorem 5.3, we showed that the response of a pole, real or complex, simple or repeated, of a continuous-time proper rational transfer function approaches 0 as $t \rightarrow \infty$ and is absolutely integrable if and only if the pole lies inside the left half- s -plane or has a negative real part. To establish Theorem 5.D3, we will show that the response of a pole, real or complex, simple or repeated, of a discrete-time proper rational transfer function approaches 0 as $k \rightarrow \infty$ and is absolutely summable if and only if the pole lies inside the unit circle on the z -plane or has magnitude less than 1. The proof is similar to the continuous-time case but is much simpler.

Let p_i be a pole of $\hat{g}(z)$ expressed in polar form as $p_i = \alpha_i e^{j\theta_i}$, where $\alpha_i \geq 0$ and θ_i are real. If $\hat{g}(z)$ has a pole p_i with multiplicity m_i , then the inverse z -transform of $\hat{g}(z)$ or the impulse response sequence contains the factors

$$p_i^k = \alpha_i^k e^{jk\theta_i}, \quad kp_i^k = k\alpha_i^k e^{jk\theta_i}, \quad \dots, \quad k^{m_i-1} p_i^k = k^{m_i-1} \alpha_i^k e^{jk\theta_i} \quad (5.9)$$

for all integers $k \geq 0$. Because $|e^{jk\theta_i}| = 1$ for all k and θ_i , we have

$$|k^m \alpha_i^k e^{jk\theta_i}| = k^m \alpha_i^k$$

for all positive integers k and m and any $\alpha_i \geq 0$. Thus the absolute summability of each term in (5.9) is governed only by the magnitude of the pole (independent of the phase). This is in contrast to the continuous-time case where the absolute integrability is governed only by the real part of the pole (independent of the imaginary part). A necessary condition of a sequence to be absolutely summable is that the sequence approaches 0 as $k \rightarrow \infty$. If $\alpha_i \geq 1$, then every term in (5.9) does not approach zero as $k \rightarrow \infty$ and is not absolutely summable. If $0 \leq \alpha_i < 1$, using $|e^{jk\theta_i}| = 1$, $|e^{j(k+1)\theta_i}| = 1$, and the ratio test that follows

$$\lim_{k \rightarrow \infty} \frac{|(k+1)^m \alpha_i^{k+1} e^{j(k+1)\theta_i}|}{|k^m \alpha_i^k e^{jk\theta_i}|} = \lim_{k \rightarrow \infty} \left(\frac{k+1}{k} \right)^m \alpha_i = \alpha_i < 1$$

for any positive integer m , we conclude that every term in (5.9) is absolutely summable and consequently approaches 0 as $k \rightarrow \infty$. Note that α_i^k is an exponentially decreasing sequence for $\alpha_i < 1$, because there exists a negative real number λ such that $\alpha_i^k = e^{\lambda k}$, for all $k \geq 0$. For example, we have $\alpha_i^k = 0.9^k = e^{\lambda k} = e^{-0.1054k}$. The ratio test states that every sequence in (5.9) is dominated by an exponentially decreasing sequence as $k \rightarrow \infty$. Thus every sequence in (5.9) is absolutely summable for $\alpha_i < 1$. The rest of the argument is similar to that of Theorem 5.3 and is omitted.

Before proceeding, we discuss the employment of Theorem 5.D3. A polynomial is defined to be *DT stable* if all its roots have magnitudes less than 1. If the denominator $D(z)$ of a proper rational transfer function is DT stable, then the discrete transfer function is BIBO stable. Whether or not $D(z)$ is DT stable can be checked by computing all its roots using the MATLAB function `roots`. It can also be checked using the *Jury test* without computing the roots. See Reference 10.

► Corollary 5.D3

A discrete-time SISO system with proper rational transfer function $\hat{g}(z)$ is BIBO stable if and only if its impulse response approaches 0 as $k \rightarrow \infty$.

Unlike the CT case, the impulse response of a DT system can be easily generated. If a DT system is known to be LTI and lumped, its BIBO stability can be easily checked by measurement or testing. We apply an impulse sequence input, then the system is stable if and only if the response vanishes. Note that this may not be true if a system is distributed as the next example shows.

Example 5.3.1 Consider a discrete-time LTI system with impulse response sequence $g[k] = 1/k$, for $k = 1, 2, \dots$, and $g[0] = 0$. The impulse response approaches 0 as $k \rightarrow \infty$. Before we use Corollary 5.D3, we must compute its transfer function. Its transfer function was computed in (2.54) as

$$\hat{g}(z) = -\ln(1 - z^{-1})$$

It is not a rational function of z . Thus Corollary 5.D3 is not applicable.

Theorem 5.D1 is applicable to LTI lumped or distributed systems. We check whether or not $g[k]$ is absolutely summable. We compute

$$\begin{aligned} S &:= \sum_{k=0}^{\infty} |g[k]| = \sum_{k=1}^{\infty} \frac{1}{k} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots \\ &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \dots + \frac{1}{8}\right) + \left(\frac{1}{9} + \dots + \frac{1}{16}\right) + \dots \end{aligned}$$

The sum of the terms inside every pair of parentheses is 1/2 or larger. Thus we have

$$S > 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \dots = \infty$$

In other words, the impulse response sequence is not absolutely summable. Thus the discrete-time system is not BIBO stable according to Theorem 5.D1.

For MIMO discrete-time systems, we have the following results.

► **Theorem 5.MD1**

A MIMO discrete-time system with impulse response sequence matrix $\mathbf{G}[k] = [g_{ij}[k]]$ is BIBO stable if and only if every $g_{ij}[k]$ is absolutely summable.

► **Theorem 5.MD3**

A MIMO discrete-time system with discrete proper rational transfer matrix $\hat{\mathbf{G}}(z) = [\hat{g}_{ij}(z)]$ is BIBO stable if and only if every pole of every $\hat{g}_{ij}(z)$ has a magnitude less than 1.

We now discuss the BIBO stability of discrete-time state-space equations. Consider

$$\begin{aligned}\mathbf{x}[k+1] &= \mathbf{Ax}[k] + \mathbf{Bu}[k] \\ \mathbf{y}[k] &= \mathbf{Cx}[k] + \mathbf{Du}[k]\end{aligned}\tag{5.10}$$

Its discrete transfer matrix is

$$\hat{\mathbf{G}}(z) = \mathbf{C}(z\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$$

Thus Equation (5.10) or, to be more precise, the zero-state response of (5.10) is BIBO stable if and only if every pole of $\hat{\mathbf{G}}(z)$ has a magnitude less than 1.

We discuss the relationship between the poles of $\hat{\mathbf{G}}(z)$ and the eigenvalues of \mathbf{A} . Because of

$$\hat{\mathbf{G}}(z) = \frac{1}{\det(z\mathbf{I} - \mathbf{A})} \mathbf{C}[\text{Adj}(z\mathbf{I} - \mathbf{A})]\mathbf{B} + \mathbf{D}$$

every pole of $\hat{\mathbf{G}}(z)$ is an eigenvalue of \mathbf{A} . Thus if every eigenvalue of \mathbf{A} has a magnitude less than 1, then (5.10) is BIBO stable. On the other hand, even if \mathbf{A} has some eigenvalues with magnitude 1 or larger, (5.10) may, as in the continuous-time case, still be BIBO stable.

5.4 Internal Stability

The BIBO stability is defined for zero-state responses. We now discuss stability of zero-input responses. If $u(t)$ is identically zero for all $t \geq 0$, then the state-space equation reduces to $\dot{\mathbf{x}}(t) = \mathbf{Ax}(t)$, $y(t) = \mathbf{c}\mathbf{x}(t)$. Because the output is a linear combination of all state variables, if the latter have some property, then the output has the same property. However, the converse is not true because it involves the issue of observability to be discussed in the next chapter.² To simplify the discussion, we study here only the stability of

$$\dot{\mathbf{x}}(t) = \mathbf{Ax}(t)\tag{5.11}$$

²In the extreme case, if $\mathbf{c} = 0$, then $y(t) = 0$ for all t , but all state variables can be nonzero.

excited by nonzero initial state \mathbf{x}_o . Clearly, the solution of (5.11) is

$$\mathbf{x}(t) = e^{\mathbf{A}t} \mathbf{x}_o \quad (5.12)$$

Definition 5.1 The response of $\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t)$ is *marginally stable* or *stable in the sense of Lyapunov* if every finite initial state \mathbf{x}_o excites a bounded response. It is *asymptotically stable* if every finite initial state excites a bounded response which, in addition, approaches $\mathbf{0}$ as $t \rightarrow \infty$.

We mention that this definition is applicable only to linear systems. The definition that is applicable to both linear and nonlinear systems must be defined using the concept of equivalence states and can be found, for example, in Reference 6, pp. 401–403. This text studies only linear systems; therefore, we use the simplified Definition 5.1.

► Theorem 5.4

1. The equation $\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t)$ is marginally stable if and only if all eigenvalues of \mathbf{A} have zero or negative real parts and those with zero real parts are simple roots of the minimal polynomial of \mathbf{A} .
2. The equation $\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t)$ is asymptotically stable if and only if all eigenvalues of \mathbf{A} have negative real parts.

We first argue that any equivalence transformation will not alter the stability of a state-space equation. Consider $\bar{\mathbf{x}}(t) = \mathbf{P}\mathbf{x}(t)$, where \mathbf{P} is a nonsingular matrix. Then $\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t)$ is equivalent to $\dot{\bar{\mathbf{x}}}(t) = \bar{\mathbf{A}}\bar{\mathbf{x}}(t) = \mathbf{P}\mathbf{A}\mathbf{P}^{-1}\bar{\mathbf{x}}(t)$. Because every entry of $\bar{\mathbf{x}}$ is a linear combination of all entries of \mathbf{x} as implied by $\bar{\mathbf{x}}(t) = \mathbf{P}\mathbf{x}(t)$, if $\mathbf{x}(t)$ is bounded, so is $\bar{\mathbf{x}}(t)$. If $\mathbf{x}(t)$ approaches zero, so does $\bar{\mathbf{x}}(t)$. Using $\mathbf{x}(t) = \mathbf{P}^{-1}\bar{\mathbf{x}}(t)$, we can show the converse. Thus, the stability of $\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t)$ is invariant under any equivalence transformation. Note that \mathbf{A} and $\bar{\mathbf{A}}$ have, as shown in Section 4.4, the same set of eigenvalues.

The response of $\dot{\bar{\mathbf{x}}} = \bar{\mathbf{A}}\bar{\mathbf{x}}$ excited by $\bar{\mathbf{x}}(0)$ equals $\bar{\mathbf{x}}(t) = e^{\bar{\mathbf{A}}t}\bar{\mathbf{x}}(0)$. It is clear that the response is bounded if and only if every entry of $e^{\bar{\mathbf{A}}t}$ is bounded for all $t \geq 0$. If $\bar{\mathbf{A}}$ is in Jordan form, then $e^{\bar{\mathbf{A}}t}$ is of the form shown in (3.48). Using (3.48), we can show that if an eigenvalue has a negative real part, then every entry of (3.48) is bounded and approaches 0 as $t \rightarrow \infty$. If an eigenvalue has zero real part and has no Jordan block of order 2 or higher, then the corresponding entry in (3.48) is a constant or sinusoidal for all t and is, therefore, bounded. This establishes the sufficiency of the first part of Theorem 5.4. If $\bar{\mathbf{A}}$ has an eigenvalue with a positive real part, then every entry in (3.48) will grow without bound. If $\bar{\mathbf{A}}$ has an eigenvalue with zero real part and its Jordan block has order 2 or higher, then (3.48) has at least one entry that grows unbounded. This completes the proof of the first part. To be asymptotically stable, every entry of (3.48) must approach zero as $t \rightarrow \infty$. Thus no eigenvalue with zero real part is permitted. This establishes the second part of the theorem.

Example 5.4.1 Consider

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \mathbf{x}(t)$$

Its characteristic polynomial is $\Delta(\lambda) = \lambda^2(\lambda + 1)$ and its minimal polynomial is $\psi(\lambda) = \lambda(\lambda + 1)$. The matrix has eigenvalues 0, 0 and -1. The eigenvalue 0 is a simple root of the minimal polynomial. Thus the equation is marginally stable. The equation

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \mathbf{x}(t)$$

however, is not marginally stable because its minimal polynomial is $\lambda^2(\lambda + 1)$ and $\lambda = 0$ is not a simple root of the minimal polynomial.

As discussed earlier, every pole of the transfer matrix

$$\hat{\mathbf{G}}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$$

is an eigenvalue of \mathbf{A} . Thus asymptotic stability implies BIBO stability. Note that asymptotic stability is defined for the zero-input response, whereas BIBO stability is defined for the zero-state response. The system in Example 5.2.2 has eigenvalue 3 and is not asymptotically stable; however, it is BIBO stable. Thus BIBO stability, in general, does not imply asymptotic stability. We mention that marginal stability is useful only in the design of oscillators. Other than oscillators, every physical system is designed to be asymptotically stable or BIBO stable.

5.4.1 Discrete-Time Case

This subsection studies the internal stability of discrete-time systems or the stability of

$$\mathbf{x}[k+1] = \mathbf{A}\mathbf{x}[k] \quad (5.13)$$

excited by nonzero initial state \mathbf{x}_o . The solution of (5.13) is, as derived in Subsection 4.2.2,

$$\mathbf{x}[k] = \mathbf{A}^k \mathbf{x}_o \quad (5.14)$$

Equation (5.13) is said to be *marginally stable* or *stable in the sense of Lyapunov* if every finite initial state \mathbf{x}_o excites a bounded response. It is *asymptotically stable* if every finite initial state excites a bounded response that, in addition, approaches $\mathbf{0}$ as $k \rightarrow \infty$. These definitions are identical to the continuous-time case.

► Theorem 5.D4

1. The equation $\mathbf{x}[k+1] = \mathbf{A}\mathbf{x}[k]$ is marginally stable if and only if all eigenvalues of \mathbf{A} have magnitudes less than or equal to 1, and those equal to 1 are simple roots of the minimal polynomial of \mathbf{A} .

2. The equation $\mathbf{x}[k+1] = \mathbf{A}\mathbf{x}[k]$ is asymptotically stable if and only if all eigenvalues of \mathbf{A} have magnitudes less than 1.

As in the continuous-time case, any (algebraic) equivalent transformation will not alter the stability of a state-space equation. Thus we can use the Jordan form to establish the theorem. The proof is similar to the continuous-time case and will not be repeated. Asymptotic stability implies BIBO stability but not conversely. We mention that marginal stability is useful only in the design of oscillators. Other than oscillators, every physical system is designed to be asymptotically stable or BIBO stable.

5.5 Lyapunov Theorem³

This section introduces a method of checking the asymptotic stability of $\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t)$. For convenience, we call \mathbf{A} CT stable if every eigenvalue of \mathbf{A} has a negative real part. The eigenvalues of \mathbf{A} are the roots of its characteristic polynomial $\Delta(s) := \det(s\mathbf{I} - \mathbf{A})$. If $\Delta(s)$ is a CT stable polynomial, which can be checked by computing its roots or by using the Routh test without computing the roots, then \mathbf{A} is CT stable. We next discuss a method of checking whether or not \mathbf{A} is CT stable without computing its eigenvalues.

► Theorem 5.5

The matrix \mathbf{A} is CT stable if and only if for any given positive definite symmetric matrix \mathbf{N} , the *Lyapunov* equation

$$\mathbf{A}'\mathbf{M} + \mathbf{M}\mathbf{A} = -\mathbf{N} \quad (5.15)$$

has a unique symmetric solution \mathbf{M} and \mathbf{M} is positive definite.

► Corollary 5.5

The $n \times n$ matrix \mathbf{A} is CT stable if and only if for any given $m \times n$ matrix $\bar{\mathbf{N}}$ with $m < n$ and with the property

$$\text{rank } \mathcal{O} := \text{rank} \begin{bmatrix} \bar{\mathbf{N}} \\ \bar{\mathbf{N}}\mathbf{A} \\ \vdots \\ \bar{\mathbf{N}}\mathbf{A}^{n-1} \end{bmatrix} = n \quad (\text{full column rank}) \quad (5.16)$$

where \mathcal{O} is an $nm \times n$ matrix, the Lyapunov equation

$$\mathbf{A}'\mathbf{M} + \mathbf{M}\mathbf{A} = -\bar{\mathbf{N}}'\bar{\mathbf{N}} =: -\mathbf{N} \quad (5.17)$$

has a unique symmetric solution \mathbf{M} , and \mathbf{M} is positive definite.

For any $\bar{\mathbf{N}}$, the matrix \mathbf{N} in (5.17) is positive semidefinite (Theorem 3.7). Theorem 5.5 and its corollary are valid for any given \mathbf{N} ; therefore, we shall use the simplest possible \mathbf{N} . Even so, using them to check the CT stability of \mathbf{A} is not simple. It is much simpler to compute, using MATLAB, the eigenvalues of \mathbf{A} and then check their real parts. Thus the importance of Theorem 5.5 and its corollary is not in checking the

³This and its subsection may be skipped without loss of continuity. The results therein are needed in using the Lyapunov method to study nonlinear systems.

CT stability of \mathbf{A} but rather in stability study of nonlinear systems. They are essential in using the so-called second method of Lyapunov. We mention that Corollary 5.5 can be used to prove the Routh test. See Reference 6, pp. 417–419.



Proof: Proof of Theorem 5.5 Necessity: Equation (5.15) is a special case of (3.59) with $\mathbf{A} = \mathbf{A}'$ and $\mathbf{B} = \mathbf{A}$. Because \mathbf{A} and \mathbf{A}' have the same set of eigenvalues, if \mathbf{A} is stable, \mathbf{A} has no two eigenvalues such that $\lambda_i + \lambda_j = 0$. Thus the Lyapunov equation is nonsingular and has a unique solution \mathbf{M} for any \mathbf{N} . We claim that the solution can be expressed as

$$\mathbf{M} = \int_0^\infty e^{\mathbf{A}'t} \mathbf{N} e^{\mathbf{A}t} dt \quad (5.18)$$

Indeed, substituting (5.18) into (5.15) yields

$$\begin{aligned} \mathbf{A}'\mathbf{M} + \mathbf{M}\mathbf{A} &= \int_0^\infty \mathbf{A}'e^{\mathbf{A}'t} \mathbf{N} e^{\mathbf{A}t} dt + \int_0^\infty e^{\mathbf{A}'t} \mathbf{N} e^{\mathbf{A}t} \mathbf{A} dt \\ &= \int_0^\infty \frac{d}{dt} (e^{\mathbf{A}'t} \mathbf{N} e^{\mathbf{A}t}) dt = e^{\mathbf{A}'t} \mathbf{N} e^{\mathbf{A}t} \Big|_{t=0}^\infty \\ &= \mathbf{0} - \mathbf{N} = -\mathbf{N} \end{aligned} \quad (5.19)$$

where we have used the fact $e^{\mathbf{A}t} = \mathbf{0}$ at $t = \infty$ for stable \mathbf{A} . This shows that the \mathbf{M} in (5.18) is the solution. It is clear that if \mathbf{N} is symmetric, so is \mathbf{M} . Let us decompose \mathbf{N} as $\mathbf{N} = \bar{\mathbf{N}}'\bar{\mathbf{N}}$, where $\bar{\mathbf{N}}$ is nonsingular (Theorem 3.7), and consider

$$\mathbf{x}'\mathbf{M}\mathbf{x} = \int_0^\infty \mathbf{x}' e^{\mathbf{A}'t} \bar{\mathbf{N}}' \bar{\mathbf{N}} e^{\mathbf{A}t} \mathbf{x} dt = \int_0^\infty \|\bar{\mathbf{N}} e^{\mathbf{A}t} \mathbf{x}\|_2^2 dt \quad (5.20)$$

Because both $\bar{\mathbf{N}}$ and $e^{\mathbf{A}t}$ are nonsingular, for any nonzero \mathbf{x} , the integrand of (5.20) is positive for every t . Thus $\mathbf{x}'\mathbf{M}\mathbf{x}$ is positive for any $\mathbf{x} \neq \mathbf{0}$. This shows the positive definiteness of \mathbf{M} . **Sufficiency:** We show that if \mathbf{N} and \mathbf{M} are positive definite, then \mathbf{A} is stable. Let λ be an eigenvalue of \mathbf{A} and $\mathbf{v} \neq \mathbf{0}$ be a corresponding eigenvector, that is, $\mathbf{Av} = \lambda\mathbf{v}$. Even though \mathbf{A} is a real matrix, its eigenvalue and eigenvector can be complex as shown in Example 3.5.2. Taking the complex conjugate transpose of $\mathbf{Av} = \lambda\mathbf{v}$ yields $\mathbf{v}^*\mathbf{A}^* = \mathbf{v}^*\mathbf{A}' = \lambda^*\mathbf{v}^*$, where the asterisk denotes complex-conjugate transpose. Premultiplying \mathbf{v}^* and postmultiplying \mathbf{v} to (5.15) yields

$$\begin{aligned} -\mathbf{v}^*\mathbf{N}\mathbf{v} &= \mathbf{v}^*\mathbf{A}'\mathbf{M}\mathbf{v} + \mathbf{v}^*\mathbf{M}\mathbf{A}\mathbf{v} \\ &= (\lambda^* + \lambda)\mathbf{v}^*\mathbf{M}\mathbf{v} = 2\operatorname{Re}(\lambda)\mathbf{v}^*\mathbf{M}\mathbf{v} \end{aligned} \quad (5.21)$$

Because $\mathbf{v}^*\mathbf{M}\mathbf{v}$ and $\mathbf{v}^*\mathbf{N}\mathbf{v}$ are, as discussed in Section 3.9, both real and positive, (5.21) implies $\operatorname{Re}(\lambda) < 0$. This shows that every eigenvalue of \mathbf{A} has a negative real part. Q.E.D.

The proof of Corollary 5.5 follows the proof of Theorem 5.5 with some modification. We discuss only where the proof of Theorem 5.5 is not applicable. Consider (5.20). Now $\bar{\mathbf{N}}$ is $m \times n$ with $m < n$ and $\mathbf{N} = \bar{\mathbf{N}}'\bar{\mathbf{N}}$ is positive semidefinite. Even so, \mathbf{M} in (5.18) can still be positive definite if the integrand of (5.20) is not identically zero

for all t . Suppose the integrand of (5.20) is identically zero or $\bar{\mathbf{N}}e^{\mathbf{A}t}\mathbf{x} \equiv \mathbf{0}$. Then its derivative with respect to t yields $\bar{\mathbf{N}}\mathbf{A}e^{\mathbf{A}t}\mathbf{x} = \mathbf{0}$. Proceeding forward, we can obtain

$$\begin{bmatrix} \bar{\mathbf{N}} \\ \bar{\mathbf{N}}\mathbf{A} \\ \vdots \\ \bar{\mathbf{N}}\mathbf{A}^{n-1} \end{bmatrix} e^{\mathbf{A}t}\mathbf{x} = \mathbf{0} \quad (5.22)$$

This equation implies that, because of (5.16) and the nonsingularity of $e^{\mathbf{A}t}$, the only \mathbf{x} meeting (5.22) is $\mathbf{0}$. Thus, the integrand of (5.20) cannot be identically zero for any $\mathbf{x} \neq \mathbf{0}$. Thus \mathbf{M} is positive definite under the condition in (5.16). This shows the necessity of Corollary 5.5. Next we consider (5.21) with $\mathbf{N} = \bar{\mathbf{N}}'\bar{\mathbf{N}}$ or⁴

$$2\operatorname{Re}(\lambda)\mathbf{v}^*\mathbf{M}\mathbf{v} = -\mathbf{v}^*\bar{\mathbf{N}}'\bar{\mathbf{N}}\mathbf{v} = -\|\bar{\mathbf{N}}\mathbf{v}\|_2^2 \quad (5.23)$$

We show that $\bar{\mathbf{N}}\mathbf{v}$ is nonzero under (5.16). Because of $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, we have $\mathbf{A}^2\mathbf{v} = \lambda\mathbf{A}\mathbf{v} = \lambda^2\mathbf{v}, \dots, \mathbf{A}^{n-1}\mathbf{v} = \lambda^{n-1}\mathbf{v}$. Consider

$$\begin{bmatrix} \bar{\mathbf{N}} \\ \bar{\mathbf{N}}\mathbf{A} \\ \vdots \\ \bar{\mathbf{N}}\mathbf{A}^{n-1} \end{bmatrix} \mathbf{v} = \begin{bmatrix} \bar{\mathbf{N}}\mathbf{v} \\ \bar{\mathbf{N}}\mathbf{A}\mathbf{v} \\ \vdots \\ \bar{\mathbf{N}}\mathbf{A}^{n-1}\mathbf{v} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{N}}\mathbf{v} \\ \lambda\bar{\mathbf{N}}\mathbf{v} \\ \vdots \\ \lambda^{n-1}\bar{\mathbf{N}}\mathbf{v} \end{bmatrix}$$

If $\bar{\mathbf{N}}\mathbf{v} = \mathbf{0}$, the rightmost matrix is zero; the leftmost matrix, however, is nonzero under the conditions of (5.16) and $\mathbf{v} \neq \mathbf{0}$. This is a contradiction. Thus $\bar{\mathbf{N}}\mathbf{v}$ is nonzero and (5.23) implies $\operatorname{Re}(\lambda) < 0$. This completes the proof of Corollary 5.5.

In the proof of Theorem 5.5, we have established the following result. For easy reference, we state it as a theorem.

► Theorem 5.6

If the matrix \mathbf{A} is CT stable, then the Lyapunov equation

$$\mathbf{A}'\mathbf{M} + \mathbf{M}\mathbf{A} = -\mathbf{N}$$

has a unique solution for every \mathbf{N} , and the solution can be expressed as

$$\mathbf{M} = \int_0^\infty e^{\mathbf{A}'t}\mathbf{N}e^{\mathbf{A}t} dt \quad (5.24)$$

Because of the importance of this theorem, we give a different proof of the uniqueness of the solution. Suppose there are two solutions \mathbf{M}_1 and \mathbf{M}_2 . Then we have

$$\mathbf{A}'(\mathbf{M}_1 - \mathbf{M}_2) + (\mathbf{M}_1 - \mathbf{M}_2)\mathbf{A} = \mathbf{0}$$

⁴Note that if \mathbf{x} is a complex vector, then the Euclidean norm defined in Section 3.2 must be modified as $\|\mathbf{x}\|_2^2 = \mathbf{x}^*\mathbf{x}$, where \mathbf{x}^* is the complex conjugate transpose of \mathbf{x} .

which implies

$$e^{\mathbf{A}'t}[\mathbf{A}'(\mathbf{M}_1 - \mathbf{M}_2) + (\mathbf{M}_1 - \mathbf{M}_2)\mathbf{A}]e^{\mathbf{A}t} = \frac{d}{dt}[e^{\mathbf{A}'t}(\mathbf{M}_1 - \mathbf{M}_2)e^{\mathbf{A}t}] = \mathbf{0}$$

Its integration from 0 to ∞ yields

$$[e^{\mathbf{A}'t}(\mathbf{M}_1 - \mathbf{M}_2)e^{\mathbf{A}t}] \Big|_0^\infty = \mathbf{0}$$

or, using $e^{\mathbf{A}t} \rightarrow \mathbf{0}$ as $t \rightarrow \infty$,

$$\mathbf{0} - (\mathbf{M}_1 - \mathbf{M}_2) = \mathbf{0}$$

This shows the uniqueness of \mathbf{M} . Although the solution can be expressed as in (5.24), the integration is not used in computing the solution. It is simpler to arrange the Lyapunov equation, after some transformations, into a standard linear algebraic equation as in (3.60) and then solve the equation. Note that even if \mathbf{A} is not CT stable, a unique solution still exists if \mathbf{A} has no two eigenvalues such that $\lambda_i + \lambda_j = 0$. The solution however cannot be expressed as in (5.24); the integration will diverge and is meaningless. If \mathbf{A} is singular or, equivalently, has at least one zero eigenvalue, then the Lyapunov equation is always singular and solutions may or may not exist depending on whether or not \mathbf{N} lies in the range space of the equation.

5.5.1 Discrete-Time Case

Before discussing the discrete counterpart of Theorems 5.5 and 5.6, we discuss the discrete counterpart of the Lyapunov equation in (3.59). Consider

$$\mathbf{M} - \mathbf{AMB} = \mathbf{C} \quad (5.25)$$

where \mathbf{A} and \mathbf{B} are, respectively, $n \times n$ and $m \times m$ matrices, and \mathbf{M} and \mathbf{C} are $n \times m$ matrices. As (3.60), Equation (5.25) can be expressed as $\mathbf{Ym} = \mathbf{c}$, where \mathbf{Y} is an $nm \times nm$ matrix; \mathbf{m} and \mathbf{c} are $nm \times 1$ column vectors with the m columns of \mathbf{M} and \mathbf{C} stacked in order. Thus (5.25) is essentially a set of linear algebraic equations. Let η_k be an eigenvalue of \mathbf{Y} or of (5.25). Then we have

$$\eta_k = 1 - \lambda_i \mu_j \quad \text{for } i = 1, 2, \dots, n; j = 1, 2, \dots, m$$

where λ_i and μ_j are, respectively, the eigenvalues of \mathbf{A} and \mathbf{B} . This can be intuitively established as follows. Let us define $\mathcal{A}(\mathbf{M}) := \mathbf{M} - \mathbf{AMB}$. Then (5.25) can be written as $\mathcal{A}(\mathbf{M}) = \mathbf{C}$. A scalar η is an eigenvalue of \mathcal{A} if there exists a nonzero \mathbf{M} such that $\mathcal{A}(\mathbf{M}) = \eta\mathbf{M}$. Let \mathbf{u} be an $n \times 1$ right eigenvector of \mathbf{A} associated with λ_i , that is, $\mathbf{Au} = \lambda_i \mathbf{u}$. Let \mathbf{v} be a $1 \times m$ left eigenvector of \mathbf{B} associated with μ_j , that is, $\mathbf{vB} = \mathbf{v}\mu_j$. Applying \mathcal{A} to the $n \times m$ nonzero matrix \mathbf{uv} yields

$$\mathcal{A}(\mathbf{uv}) = \mathbf{uv} - \mathbf{AuvB} = (1 - \lambda_i \mu_j)\mathbf{uv}$$

Thus the eigenvalues of (5.25) are $1 - \lambda_i \mu_j$ for all i and j . If there are no i and j such that $\lambda_i \mu_j = 1$, then (5.25) is nonsingular and for any \mathbf{C} a unique solution \mathbf{M} exists in (5.25).

If $\lambda_i \mu_j = 1$ for some i and j , then (5.25) is singular and for a given \mathbf{C} , solutions may or may not exist. The situation here is similar to what was discussed in Section 3.7.

We discuss a method of checking the asymptotic stability of $\dot{\mathbf{x}}[k+1] = \mathbf{A}\mathbf{x}[k]$. For convenience, we call \mathbf{A} DT stable if every eigenvalue of \mathbf{A} has a magnitude less than 1. The eigenvalues of \mathbf{A} are the roots of its characteristic polynomial $\Delta(z) := \det(z\mathbf{I} - \mathbf{A})$. If $\Delta(z)$ is a DT stable polynomial, which can be checked by computing its roots or by using the Jury test without computing the roots, then \mathbf{A} is DT stable. We next discuss a method of checking whether or not \mathbf{A} is DT stable without computing its eigenvalues.

Theorem 5.D5

An $n \times n$ matrix \mathbf{A} is DT stable if and only if for any given positive definite symmetric matrix \mathbf{N} or for $\mathbf{N} = \tilde{\mathbf{N}}'\tilde{\mathbf{N}}$, where $\tilde{\mathbf{N}}$ is any given $m \times n$ matrix with $m < n$ and with the property in (5.16), the discrete Lyapunov equation

$$\mathbf{M} - \mathbf{A}'\mathbf{M}\mathbf{A} = \mathbf{N} \quad (5.26)$$

has a unique symmetric solution \mathbf{M} and \mathbf{M} is positive definite.

We sketch briefly its proof for $\mathbf{N} > 0$. If all eigenvalues of \mathbf{A} and, consequently, of \mathbf{A}' have magnitudes less than 1, then we have $|\lambda_i \lambda_j| < 1$ for all i and j . Thus $\lambda_i \lambda_j \neq 1$ and (5.26) is nonsingular. Therefore, for any \mathbf{N} , a unique solution exists in (5.26). We claim that the solution can be expressed as

$$\mathbf{M} = \sum_{m=0}^{\infty} (\mathbf{A}')^m \mathbf{N} \mathbf{A}^m \quad (5.27)$$

Because $|\lambda_i| < 1$ for all i , the infinite series converges and is well-defined. Substituting (5.27) into (5.26) yields

$$\begin{aligned} & \sum_{m=0}^{\infty} (\mathbf{A}')^m \mathbf{N} \mathbf{A}^m - \mathbf{A}' \left(\sum_{m=0}^{\infty} (\mathbf{A}')^m \mathbf{N} \mathbf{A}^m \right) \mathbf{A} \\ &= \mathbf{N} + \sum_{m=1}^{\infty} (\mathbf{A}')^m \mathbf{N} \mathbf{A}^m - \sum_{m=1}^{\infty} (\mathbf{A}')^m \mathbf{N} \mathbf{A}^m = \mathbf{N} \end{aligned}$$

Thus (5.27) is the solution. If \mathbf{N} is symmetric, so is \mathbf{M} . If \mathbf{N} is positive definite, so is \mathbf{M} . This establishes the necessity. To show sufficiency, let λ be an eigenvalue of \mathbf{A} and $\mathbf{v} \neq \mathbf{0}$ be a corresponding eigenvector, that is, $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Then we have

$$\begin{aligned} \mathbf{v}^* \mathbf{N} \mathbf{v} &= \mathbf{v}^* \mathbf{M} \mathbf{v} - \mathbf{v}^* \mathbf{A}' \mathbf{M} \mathbf{A} \mathbf{v} \\ &= \mathbf{v}^* \mathbf{M} \mathbf{v} - \lambda^* \mathbf{v}^* \mathbf{M} \mathbf{v} \lambda = (1 - |\lambda|^2) \mathbf{v}^* \mathbf{M} \mathbf{v} \end{aligned}$$

Because both $\mathbf{v}^* \mathbf{N} \mathbf{v}$ and $\mathbf{v}^* \mathbf{M} \mathbf{v}$ are real and positive, we conclude $(1 - |\lambda|^2) > 0$ or $|\lambda|^2 < 1$. This establishes the theorem for $\mathbf{N} > 0$. The case $\mathbf{N} \geq 0$ can be similarly established.

► **Theorem 5.D6**

If \mathbf{A} is DT stable, then the discrete Lyapunov equation

$$\mathbf{M} - \mathbf{A}'\mathbf{M}\mathbf{A} = \mathbf{N}$$

has a unique solution for every \mathbf{N} , and the solution can be expressed as

$$\mathbf{M} = \sum_{m=0}^{\infty} (\mathbf{A}')^m \mathbf{N} \mathbf{A}^m$$

It is important to mention that even if \mathbf{A} has one or more eigenvalues with magnitudes larger than 1, a unique solution still exists in the discrete Lyapunov equation if $\lambda_i \lambda_j \neq 1$ for all i and j . In this case, the solution cannot be expressed as in (5.27) but can be computed from a set of linear algebraic equations.

Let us discuss the relationships between the continuous-time and discrete-time Lyapunov equations. The CT stability condition for continuous-time systems is that all eigenvalues lie inside the left half- s -plane. The DT stability condition for discrete-time systems is that all eigenvalues lie inside the unit circle on the z -plane. These conditions can be related by the bilinear transformation

$$s = \frac{z-1}{z+1}, \quad z = \frac{1+s}{1-s} \quad (5.28)$$

which maps the left half- s -plane into the interior of the unit circle on the z -plane and vice versa. See Reference 7, p. 520 with $T = 2$. To differentiate the continuous-time and discrete-time cases, we write

$$\mathbf{A}'\mathbf{M} + \mathbf{M}\mathbf{A} = -\mathbf{N} \quad (5.29)$$

and

$$\mathbf{M}_d - \mathbf{A}'_d \mathbf{M}_d \mathbf{A}_d = \mathbf{N}_d \quad (5.30)$$

These two equations can be related by, following (5.28),

$$\mathbf{A} = (\mathbf{A}_d + \mathbf{I})^{-1}(\mathbf{A}_d - \mathbf{I}), \quad \mathbf{A}_d = (\mathbf{I} + \mathbf{A})(\mathbf{I} - \mathbf{A})^{-1}$$

Substituting the right-hand-side equation into (5.30) and simple manipulation yield

$$\mathbf{A}'\mathbf{M}_d + \mathbf{M}_d\mathbf{A} = -0.5(\mathbf{I} - \mathbf{A}')\mathbf{N}_d(\mathbf{I} - \mathbf{A})$$

Comparing this with (5.29) yields

$$\mathbf{A} = (\mathbf{A}_d + \mathbf{I})^{-1}(\mathbf{A}_d - \mathbf{I}), \quad \mathbf{M} = \mathbf{M}_d, \quad \mathbf{N} = 0.5(\mathbf{I} - \mathbf{A}')\mathbf{N}_d(\mathbf{I} - \mathbf{A}) \quad (5.31)$$

These relates (5.29) and (5.30).

The MATLAB function `lyap` computes the Lyapunov equation in (5.29) and `dlyap` computes the discrete Lyapunov equation in (5.30). The function `dlyap` transforms (5.30) into (5.29) using (5.31) and then calls `lyap`. The result yields $\mathbf{M} = \mathbf{M}_d$.

5.6 Stability of LTV Systems

Consider a SISO linear time-varying (LTV) system described by

$$\mathbf{y}(t) = \int_{t_0}^t g(t, \tau) u(\tau) d\tau \quad (5.32)$$

The system is said to be BIBO stable if every bounded input excites a bounded output. The condition for (5.32) to be BIBO stable is that there exist a finite constant M such that

$$\int_{t_0}^t |g(t, \tau)| d\tau \leq M < \infty \quad (5.33)$$

for all t and t_0 with $t \geq t_0$. The proof in the time-invariant case applies here with only minor modification.

For the MIMO case, (5.32) becomes

$$\mathbf{y}(t) = \int_{t_0}^t \mathbf{G}(t, \tau) \mathbf{u}(\tau) d\tau \quad (5.34)$$

The condition for (5.34) to be BIBO stable is that every entry of $\mathbf{G}(t, \tau)$ meet the condition in (5.33). For MIMO systems, we can also express the condition in terms of norms. Any norm discussed in Section 3.11 can be used. However, the infinite norm

$$\|\mathbf{u}\|_\infty = \max_i |u_i|, \quad \|\mathbf{G}\|_\infty = \text{largest row absolute sum}$$

is probably the most convenient to use in stability study. For convenience, no subscript will be attached to any norm. The necessary and sufficient condition for (5.34) to be BIBO stable is that there exist a finite constant M such that

$$\int_{t_0}^t \|\mathbf{G}(t, \tau)\| d\tau \leq M < \infty$$

for all t and t_0 with $t \geq t_0$.

The impulse response matrix of

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}(t)\mathbf{x}(t) + \mathbf{D}(t)\mathbf{u}(t) \end{aligned} \quad (5.35)$$

is

$$\mathbf{G}(t, \tau) = \mathbf{C}(t)\Phi(t, \tau)\mathbf{B}(\tau) + \mathbf{D}(t)\delta(t - \tau)$$

and the zero-state response is

$$\mathbf{y}(t) = \int_{t_0}^t \mathbf{C}(t)\Phi(t, \tau)\mathbf{B}(\tau)\mathbf{u}(\tau) d\tau + \mathbf{D}(t)\mathbf{u}(t)$$

Thus (5.35) or, more precisely, the zero-state response of (5.35) is BIBO stable if and only if there exist constants M_1 and M_2 such that

$$\|\mathbf{D}(t)\| \leq M_1 < \infty$$

and

$$\int_{t_0}^t \|\mathbf{G}(t, \tau)\| d\tau \leq M_2 < \infty$$

for all t and t_0 with $t \geq t_0$.

Next we study the stability of $\dot{\mathbf{x}} = \mathbf{A}(t)\mathbf{x}$. As in the time-invariant case, we define the equation to be marginally stable if every finite initial state excites a bounded response. Because the response is governed by

$$\mathbf{x}(t) = \Phi(t, t_0)\mathbf{x}(t_0) \quad (5.36)$$

we conclude that the response is marginally stable if and only if there exists a finite constant M such that

$$\|\Phi(t, t_0)\| \leq M < \infty \quad (5.37)$$

for all t_0 and for all $t \geq t_0$. The equation $\dot{\mathbf{x}} = \mathbf{A}(t)\mathbf{x}$ is asymptotically stable if the response excited by every finite initial state is bounded and approaches zero as $t \rightarrow \infty$. The asymptotic stability conditions are the boundedness condition in (5.37) and

$$\|\Phi(t, t_0)\| \rightarrow 0 \quad \text{as } t \rightarrow \infty \quad (5.38)$$

A great deal can be said regarding these definitions and conditions. Does the constant M in (5.37) depend on t_0 ? What is the rate for the state transition matrix to approach 0 in (5.38)? The interested reader is referred to References 4 and 19.

A time-invariant equation $\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t)$ is asymptotically stable if all eigenvalues of \mathbf{A} have negative real parts. Is this also true for the time-varying case? The answer is negative as the next example shows.

Example 5.6.1 Consider the linear time-varying equation

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) = \begin{bmatrix} -1 & e^{2t} \\ 0 & -1 \end{bmatrix} \mathbf{x}(t) \quad (5.39)$$

The characteristic polynomial of $\mathbf{A}(t)$ is

$$\det(\lambda\mathbf{I} - \mathbf{A}(t)) = \det \begin{bmatrix} \lambda + 1 & -e^{2t} \\ 0 & \lambda + 1 \end{bmatrix} = (\lambda + 1)^2$$

Thus $\mathbf{A}(t)$ has eigenvalues -1 and -1 for all t . It can be directly verified that the equation

$$\Phi(t, 0) = \begin{bmatrix} e^{-t} & 0.5(e^t - e^{-t}) \\ 0 & e^{-t} \end{bmatrix}$$

Imeets (4.61) and is, therefore, the state transition matrix of (5.39). See also Problem 4.20. Because the (1,2)-th entry of Φ grows without bound, the equation is not asymptotically stable nor marginally stable. This example shows that even though the eigenvalues can be defined for $\mathbf{A}(t)$ at every t , the concept of eigenvalues is not useful in the time-varying case.

All stability properties in the time-invariant case are invariant under any equivalence transformation. In the time-varying case, BIBO stability is also invariant because any equivalence transformation will not alter, as discussed in Section 4.7, the impulse response. An equivalence transformation can transform, as shown in Theorem 4.3, any $\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t)$ into $\dot{\bar{\mathbf{x}}}(t) = \mathbf{A}_o\bar{\mathbf{x}}(t)$, where \mathbf{A}_o is any constant matrix including the zero matrix. Therefore, in the time-varying case, marginal and asymptotic stabilities are not invariant under any equivalence transformation.

► Theorem 5.7

Marginal and asymptotic stabilities of $\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t)$ are invariant under any Lyapunov transformation.

As discussed in Section 4.6, if $\mathbf{P}(t)$ and $\dot{\mathbf{P}}(t)$ are continuous, and $\mathbf{P}(t)$ is nonsingular for all t , then $\bar{\mathbf{x}} = \mathbf{P}(t)\mathbf{x}$ is an equivalence transformation. If, in addition, $\mathbf{P}(t)$ and $\mathbf{P}^{-1}(t)$ are bounded for all t , then $\bar{\mathbf{x}} = \mathbf{P}(t)\mathbf{x}$ is a Lyapunov transformation. The fundamental matrix $\mathbf{X}(t)$ of $\dot{\mathbf{x}} = \mathbf{A}(t)\mathbf{x}$ and the fundamental matrix $\bar{\mathbf{X}}(t)$ of $\dot{\bar{\mathbf{x}}} = \bar{\mathbf{A}}(t)\bar{\mathbf{x}}$ are related by, as derived in (4.79),

$$\bar{\mathbf{X}}(t) = \mathbf{P}(t)\mathbf{X}(t)$$

which implies

$$\begin{aligned}\bar{\Phi}(t, \tau) &= \bar{\mathbf{X}}(t)\bar{\mathbf{X}}^{-1}(\tau) = \mathbf{P}(t)\mathbf{X}(t)\mathbf{X}^{-1}(\tau)\mathbf{P}^{-1}(\tau) \\ &= \mathbf{P}(t)\Phi(t, \tau)\mathbf{P}^{-1}(\tau)\end{aligned}\quad (5.40)$$

Because both $\mathbf{P}(t)$ and $\mathbf{P}^{-1}(t)$ are bounded, if $\|\Phi(t, \tau)\|$ is bounded, so is $\|\bar{\Phi}(t, \tau)\|$; if $\|\Phi(t, \tau)\| \rightarrow 0$ as $t \rightarrow \infty$, so is $\|\bar{\Phi}(t, \tau)\|$. This establishes Theorem 5.7.

In the time-invariant case, asymptotic stability of zero-input responses always implies BIBO stability of zero-state responses. This is not necessary so in the time-varying case. A time-varying equation is asymptotically stable if

$$\|\Phi(t, t_0)\| \rightarrow 0 \quad \text{as } t \rightarrow \infty \quad (5.41)$$

for all t, t_0 with $t \geq t_0$. It is BIBO stable if

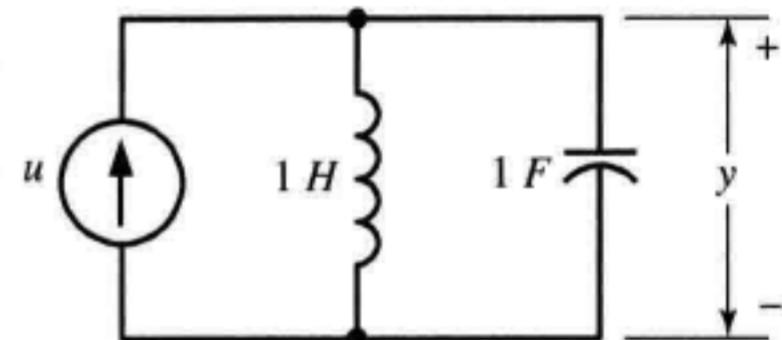
$$\int_{t_0}^t \|\mathbf{C}(t)\Phi(t, \tau)\mathbf{B}(\tau)\| d\tau < \infty \quad (5.42)$$

for all t, t_0 with $t \geq t_0$. A function that approaches 0 as $t \rightarrow \infty$ may not be absolutely integrable. Thus asymptotic stability may not imply BIBO stability in the time-varying case. However, if $\|\Phi(t, \tau)\|$ decreases to zero rapidly—in particular, exponentially—and if $\mathbf{C}(t)$ and $\mathbf{B}(t)$ are bounded for all t , then asymptotic stability does imply BIBO stability. See References 4, 6, and 19.

Problems

- 5.1** Is the network shown in Fig. 5.2 BIBO stable? If not, find a bounded input that will excite an unbounded output.

FIGURE 5.2



- 5.2** Consider a system with an irrational transfer function $\hat{g}(s)$. Show that a necessary condition for the system to be BIBO stable is that $|\hat{g}(s)|$ is finite for all $\text{Re } s \geq 0$.
- 5.3** Is a system with impulse response $g(t) = 1/(t + 1)$ for $t \geq 0$ BIBO stable? How about $g(t) = te^{-t}$ for $t \geq 0$?
- 5.4** Is a system with transfer function $\hat{g}(s) = e^{-2s}/(s + 1)$ BIBO stable?
- 5.5** Show that the negative-feedback system shown in Fig. 2.5(b) is BIBO stable if and only if the gain a has a magnitude less than 1. For $a = 1$, find a bounded input $r(t)$ that will excite an unbounded output.
- 5.6** Consider a system with transfer function $\hat{g}(s) = (s - 2)/(s + 1)$. What are the steady-state responses excited by $u(t) = 3$, for $t \geq 0$, and by $u(t) = \sin 2t$, for $t \geq 0$?
- 5.7** Consider a system with transfer function $\hat{g}(s) = 1/(s - 1)$ and an input whose Laplace transform equals $\hat{u}(s) = (s - 1)/s(s + 1)$. Is the system BIBO stable? Is the input bounded? What is the output of the system excited by the input? Will the response of the pole appear in the output? We see from this example that a pole of a system will not be excited by an input only if the Laplace transform of the input has a zero to cancel out the pole. This rarely happens in an arbitrarily generated input. Thus most inputs will excite every pole of a system.
- 5.8** Show that a system with a proper rational transfer function is BIBO stable if and only if its step response approaches a zero or nonzero constant.
- 5.9** Consider

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \begin{bmatrix} -1 & 5 \\ 0 & 2 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 2 \\ 0 \end{bmatrix} u(t) \\ y(t) &= [-2 \quad 4] \mathbf{x}(t) - 2u(t)\end{aligned}$$

Is it BIBO stable?

- 5.10** Consider a discrete-time system with impulse response sequence

$$g[k] = k(0.9)^k \quad \text{for } k \geq 0$$

Is the system BIBO stable?

- 5.11** Is the state equation in Problem 5.9 marginally stable? asymptotically stable?
5.12 Is the homogeneous state equation

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -1 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{x}(t)$$

marginally stable? asymptotically stable?

5.13 Is the homogeneous state equation

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -1 & 0 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{x}(t)$$

marginally stable? asymptotically stable?

5.14 Investigate the stability of the following system.

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} u(t)$$

$$y(t) = [1 \ 0 \ 1 \ 0 \ 0 \ 1] \mathbf{x}(t)$$

5.15 Given a system described by

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -3 & -2k & -3 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 0 \\ k \end{bmatrix} u(t)$$

Determine the range of k which the system is asymptotically stable. Find k for marginally stable case and obtain the frequency at which the system oscillates.

5.16 Consider the system represented by state equation

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u(t)$$

$$y(t) = [1 \ 1 \ 1] \mathbf{x}(t)$$

Find all equilibrium states of the system. Is every state stable in the sense of Lyapunov? Is it asymptotically stable? Is the zero state response BIBO stable?

5.17 Consider the system represented by the transfer function.

$$\hat{g}(s) = (s + k_1) / (s^3 + 2s^2 + k_2s + 4)$$

Find the range of k_1 and k_2 such that the system is BIBO stable. Suppose the system is connected in a conventional unity feedback system. Find the range of k_1 and k_2 for BIBO stability of feedback system.

5.18 Consider a state-space representation of a stable continuous-time system,

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + Bu(t)$$

$$y(t) = C\mathbf{x}(t)$$

Let the matrix A be replaced by (a) A' , (b) A^{-1} , (c) A^r , (d) kA , (e) $A + kI$, where r and k are positive integers. Investigate the stability status of the system for each case. Repeat your investigation for a stable discrete-time system.

- 5.19** Is the discrete-time homogeneous state equation

$$\mathbf{x}[k+1] = \begin{bmatrix} 0.9 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}[k]$$

marginally stable? asymptotically stable?

- 5.20** Is the discrete-time homogeneous state equation

$$\mathbf{x}[k+1] = \begin{bmatrix} 0.9 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}[k]$$

marginally stable? asymptotically stable?

- 5.21** Construct polynomials of degree 3 which are

1. CT stable and DT stable.
2. CT stable but not DT stable.
3. DT stable but not CT stable.
4. CT unstable and DT unstable.

- 5.22** Use Theorem 5.5 to show that the matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -0.5 & -1 \end{bmatrix}$$

is CT stable.

- 5.23** Use Theorem 5.D5 to show that the matrix \mathbf{A} in Problem 5.22 is DT stable.

- 5.24** For any distinct negative real λ_i and any nonzero real a_i , show that the matrix

$$\mathbf{M} = \begin{bmatrix} -\frac{a_1^2}{2\lambda_1} & -\frac{a_1 a_2}{\lambda_1 + \lambda_2} & -\frac{a_1 a_3}{\lambda_1 + \lambda_3} \\ -\frac{a_2 a_1}{\lambda_2 + \lambda_1} & -\frac{a_2^2}{2\lambda_2} & -\frac{a_2 a_3}{\lambda_2 + \lambda_3} \\ -\frac{a_3 a_1}{\lambda_1 + \lambda_3} & -\frac{a_3 a_2}{\lambda_2 + \lambda_3} & -\frac{a_3^2}{2\lambda_3} \end{bmatrix}$$

is positive definite. (*Hint:* Use Corollary 5.5 and $\mathbf{A} = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$.)

- 5.25** A real matrix \mathbf{M} (not necessarily symmetric) is defined to be positive definite if $\mathbf{x}'\mathbf{M}\mathbf{x} > 0$ for any nonzero \mathbf{x} . Is it true that the matrix \mathbf{M} is positive definite if all eigenvalues of \mathbf{M} are real and positive or if all its leading principal minors are positive? If not, how do you check its positive definiteness? (*Hint:* Try

$$\begin{bmatrix} 0 & 1 \\ -2 & 3 \end{bmatrix} \quad \begin{bmatrix} 2 & 1 \\ 1.9 & 1 \end{bmatrix})$$

- 5.26** Show that all eigenvalues of \mathbf{A} have real parts less than $-\mu < 0$ if and only if, for any given positive definite symmetric matrix \mathbf{N} , the equation

$$\mathbf{A}'\mathbf{M} + \mathbf{M}\mathbf{A} + 2\mu\mathbf{M} = -\mathbf{N}$$

has a unique symmetric solution \mathbf{M} and \mathbf{M} is positive definite.

- 5.27** Show that all eigenvalues of \mathbf{A} have magnitudes less than ρ if and only if, for any given positive definite symmetric matrix \mathbf{N} , the equation

$$\rho^2 \mathbf{M} - \mathbf{A}' \mathbf{M} \mathbf{A} = \rho^2 \mathbf{N}$$

has a unique symmetric solution \mathbf{M} and \mathbf{M} is positive definite.

- 5.28** Use the Lyapunov stability theorem to show that if A is asymptotically stable, then

$$\int_0^\infty x'(t) N x(t) dt = x'(0) M x(0)$$

- 5.29** Consider a set of symmetric stable matrices given by $S = \{A_1, A_2, \dots, A_k\}$. Show that a common quadratic Lyapunov function for this set is given by $V(x) = (\|x\|_2)^2$.

- 5.30** Is a system with impulse response $g(t, \tau) = e^{-2|t| - |\tau|}$ for $t \geq \tau$ BIBO stable? How about $g(t, \tau) = \sin t (e^{-(t-\tau)}) \cos \tau$?

- 5.31** Consider the time-varying equation

$$\dot{x}(t) = 2tx(t) + u(t), \quad y(t) = e^{-t^2} x(t)$$

Is the equation BIBO stable? marginally stable? asymptotically stable?

- 5.32** Show that the equation in Problem 5.26 can be transformed by using $\bar{x} = P(t)x$, with $P(t) = e^{-t^2}$, into

$$\dot{\bar{x}}(t) = 0 \cdot \bar{x}(t) + e^{-t^2} u(t), \quad y(t) = \bar{x}(t)$$

Is the equation BIBO stable? marginally stable? asymptotically stable? Is the transformation a Lyapunov transformation?

- 5.33** Is the homogeneous equation

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -1 & 0 \\ -e^{-3t} & 0 \end{bmatrix} \mathbf{x}(t)$$

for $t_0 \geq 0$ marginally stable? asymptotically stable?

- 5.34** Investigate the stability status of time-varying systems of Problems 4.16 and 4.20.

Controllability and Observability

6.1 Introduction

This chapter introduces the concepts of controllability and observability. Controllability is concerned with whether or not the state of a state-space equation can be controlled from the input, and observability is concerned with whether or not the initial state can be observed from the output. These concepts can be illustrated using the circuit shown in Fig. 6.1. The circuit has two state variables. Let x_i be the voltage across the capacitor with capacitance C_i , for $i = 1, 2$. The input u is a current source and the output y is the voltage shown. From the circuit, we see that, because of the open circuit across y , the input has no effect on x_2 or cannot control x_2 . The current passing through the $2\text{-}\Omega$ resistor always equals the current source u ; therefore, the response excited by the initial state x_1 will not appear in y . Thus the initial state x_1 cannot be observed from the output. Thus the equation describing the circuit cannot be controllable and observable.

These concepts are essential in discussing the internal structure of linear systems. They are also needed in studying control and filtering problems. We study first continuous-time linear time-invariant (LTI) state-space equations and then the

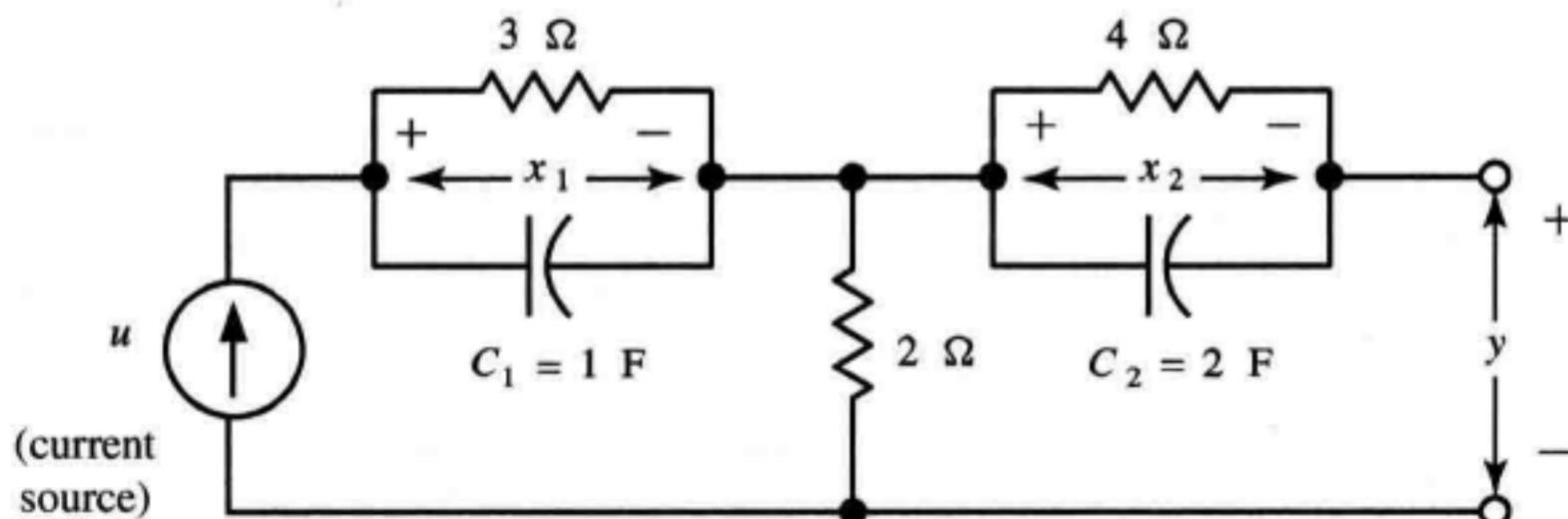


FIGURE 6.1 Circuit.

discrete-time case. Finally, we study the time-varying case. Because the mathematics involved is essentially the same for MIMO and SISO systems, we study first MIMO systems and then mention the results for SISO systems.

6.2 Controllability

Consider the n -dimensional p -input state equation

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \quad (6.1)$$

where \mathbf{A} and \mathbf{B} are, respectively, $n \times n$ and $n \times p$ real constant matrices. Because the output does not play any role in controllability, we will disregard the output equation in this study.

Definition 6.1 The state equation (6.1) or the pair (\mathbf{A}, \mathbf{B}) is said to be *controllable* if for any initial state $\mathbf{x}(0) = \mathbf{x}_o$ and any final state \mathbf{x}_1 , there exists an input that transfers \mathbf{x}_o to \mathbf{x}_1 in a finite time. Otherwise (6.1) or (\mathbf{A}, \mathbf{B}) is said to be *uncontrollable*.

This definition requires only that the input be capable of moving any state in the state space to any other state in a finite time; what trajectory the state should take is not specified. Furthermore, there is no constraint imposed on the input; its magnitude can be as large as desired. We give an example to illustrate the concept.

Example 6.2.1 Consider the circuit shown in Fig. 6.2(a). Its state variable x is the voltage across the capacitor. If $x(0) = 0$, then $x(t) = 0$ for all $t \geq 0$ no matter what input is applied. This is due to the symmetry of the circuit, and the input has no effect on the voltage across the capacitor. Thus the system or, more precisely, the state equation that describes the system is not controllable.

Next we consider the circuit shown in Fig. 6.2(b). It has two state variables x_1 and x_2 as shown. The input can transfer x_1 or x_2 to any value; but it cannot transfer x_1 and x_2 to any values. For example, if $x_1(0) = x_2(0) = 0$, then no matter what input

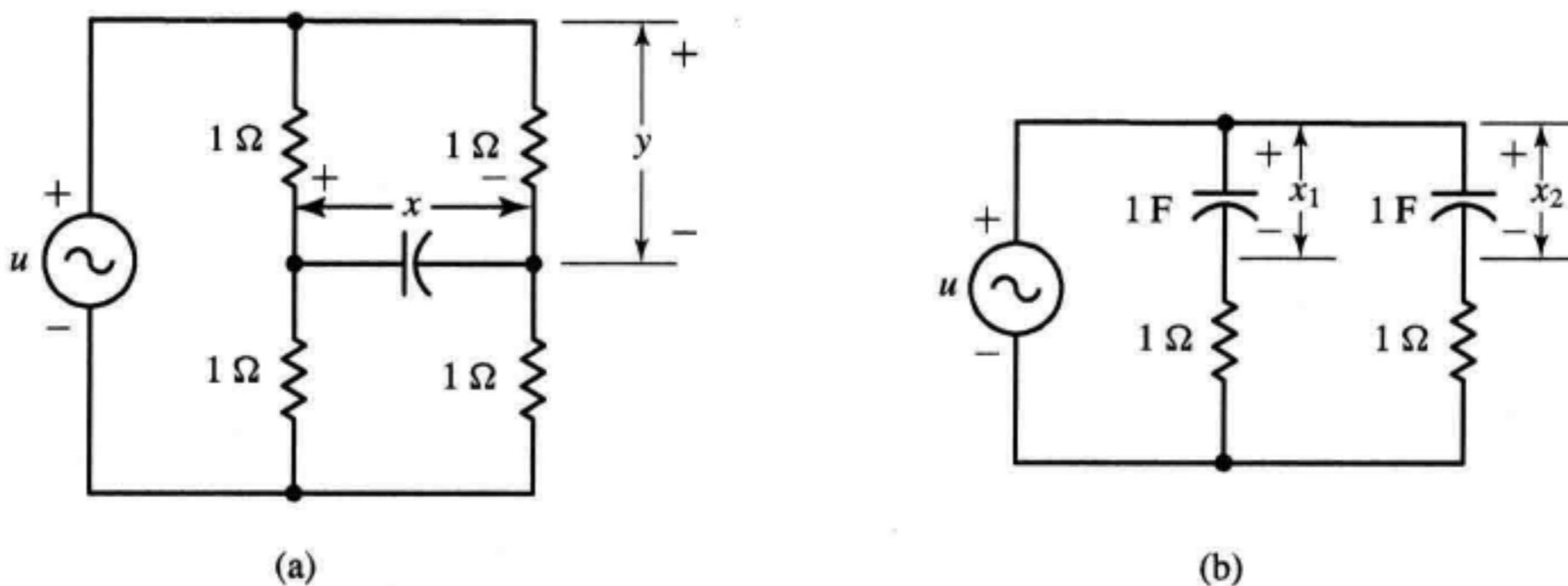


FIGURE 6.2 Uncontrollable circuits.

is applied, $x_1(t)$ always equals $x_2(t)$ for all $t \geq 0$. Thus the equation that describes the circuit is not controllable.

► Theorem 6.1

The following statements are equivalent.

1. The n -dimensional pair (\mathbf{A}, \mathbf{B}) is controllable.
2. The $n \times n$ matrix

$$\mathbf{W}_c(t) = \int_0^t e^{\mathbf{A}\tau} \mathbf{B} \mathbf{B}' e^{\mathbf{A}'\tau} d\tau = \int_0^t e^{\mathbf{A}(t-\tau)} \mathbf{B} \mathbf{B}' e^{\mathbf{A}'(t-\tau)} d\tau \quad (6.2)$$

is nonsingular for any $t > 0$.

3. The $n \times np$ controllability matrix

$$\mathcal{C} = [\mathbf{B} \ \mathbf{AB} \ \mathbf{A}^2 \mathbf{B} \cdots \mathbf{A}^{n-1} \mathbf{B}] \quad (6.3)$$

has rank n (full row rank).

4. The $n \times (n+p)$ matrix $[\mathbf{A} - \lambda \mathbf{I} \ \mathbf{B}]$ has full row rank at every eigenvalue, λ , of \mathbf{A} .¹
5. If, in addition, all eigenvalues of \mathbf{A} have negative real parts, then the unique solution of

$$\mathbf{AW}_c + \mathbf{W}_c \mathbf{A}' = -\mathbf{BB}' \quad (6.4)$$

is positive definite. The solution is called the *controllability Gramian* and can be expressed as

$$\mathbf{W}_c = \int_0^\infty e^{\mathbf{A}\tau} \mathbf{B} \mathbf{B}' e^{\mathbf{A}'\tau} d\tau \quad (6.5)$$



Proof: (1) \Leftrightarrow (2): First we show the equivalence of the two forms in (6.2). Let us introduce the new integration variable $\bar{\tau} := t - \tau$, where t is fixed. Then we have

$$\begin{aligned} \int_{\tau=0}^t e^{\mathbf{A}(t-\tau)} \mathbf{B} \mathbf{B}' e^{\mathbf{A}'(t-\tau)} d\tau &= \int_{\bar{\tau}=t}^0 e^{\mathbf{A}\bar{\tau}} \mathbf{B} \mathbf{B}' e^{\mathbf{A}'\bar{\tau}} (-d\bar{\tau}) \\ &= \int_{\bar{\tau}=0}^t e^{\mathbf{A}\bar{\tau}} \mathbf{B} \mathbf{B}' e^{\mathbf{A}'\bar{\tau}} d\bar{\tau} \end{aligned}$$

It becomes the first form of (6.2) after replacing $\bar{\tau}$ by τ . Because of the form of the integrand, $\mathbf{W}_c(t)$ is always positive semidefinite; it is positive definite if and only if it is nonsingular. See Section 3.9.

First we show that if $\mathbf{W}_c(t)$ is nonsingular, then (6.1) is controllable. The response of (6.1) at time t_1 was derived in (4.5) as

$$\mathbf{x}(t_1) = e^{\mathbf{At}_1} \mathbf{x}(0) + \int_0^{t_1} e^{\mathbf{A}(t_1-\tau)} \mathbf{B} \mathbf{u}(\tau) d\tau \quad (6.6)$$

¹If λ is complex, then we must use complex numbers as scalars in checking the rank. See the discussion regarding (3.37).

We claim that for any $\mathbf{x}(0) = \mathbf{x}_0$ and any $\mathbf{x}(t_1) = \mathbf{x}_1$, the input

$$\mathbf{u}(t) = -\mathbf{B}' e^{\mathbf{A}'(t_1-t)} \mathbf{W}_c^{-1}(t_1) [e^{\mathbf{A}t_1} \mathbf{x}_0 - \mathbf{x}_1] \quad (6.7)$$

will transfer \mathbf{x}_0 to \mathbf{x}_1 at time t_1 . Indeed, substituting (6.7) into (6.6) yields

$$\begin{aligned} \mathbf{x}(t_1) &= e^{\mathbf{A}t_1} \mathbf{x}_0 - \left(\int_0^{t_1} e^{\mathbf{A}(t_1-\tau)} \mathbf{B} \mathbf{B}' e^{\mathbf{A}'(t_1-\tau)} d\tau \right) \mathbf{W}_c^{-1}(t_1) [e^{\mathbf{A}t_1} \mathbf{x}_0 - \mathbf{x}_1] \\ &= e^{\mathbf{A}t_1} \mathbf{x}_0 - \mathbf{W}_c(t_1) \mathbf{W}_c^{-1}(t_1) [e^{\mathbf{A}t_1} \mathbf{x}_0 - \mathbf{x}_1] = \mathbf{x}_1 \end{aligned}$$

This shows that if \mathbf{W}_c is nonsingular, then the pair (\mathbf{A}, \mathbf{B}) is controllable. We show the converse by contradiction. Suppose the pair is controllable but $\mathbf{W}_c(t_1)$ is not positive definite for some t_1 . Then there exists an $n \times 1$ nonzero vector \mathbf{v} such that

$$\begin{aligned} \mathbf{v}' \mathbf{W}_c(t_1) \mathbf{v} &= \int_0^{t_1} \mathbf{v}' e^{\mathbf{A}(t_1-\tau)} \mathbf{B} \mathbf{B}' e^{\mathbf{A}'(t_1-\tau)} \mathbf{v} d\tau \\ &= \int_0^{t_1} \|\mathbf{B}' e^{\mathbf{A}'(t_1-\tau)} \mathbf{v}\|^2 d\tau = 0 \end{aligned}$$

which implies

$$\mathbf{B}' e^{\mathbf{A}'(t_1-\tau)} \mathbf{v} \equiv \mathbf{0} \quad \text{or} \quad \mathbf{v}' e^{\mathbf{A}(t_1-\tau)} \mathbf{B} \equiv \mathbf{0} \quad (6.8)$$

for all τ in $[0, t_1]$. If (6.1) is controllable, there exists an input that transfers the initial state $\mathbf{x}(0) = e^{-\mathbf{A}t_1} \mathbf{v}$ to $\mathbf{x}(t_1) = \mathbf{0}$ and (6.6) becomes

$$\mathbf{0} = \mathbf{v} + \int_0^{t_1} e^{\mathbf{A}(t_1-\tau)} \mathbf{B} \mathbf{u}(\tau) d\tau$$

Its premultiplication by \mathbf{v}' yields

$$0 = \mathbf{v}' \mathbf{v} + \int_0^{t_1} \mathbf{v}' e^{\mathbf{A}(t_1-\tau)} \mathbf{B} \mathbf{u}(\tau) d\tau = \|\mathbf{v}\|^2 + 0$$

which contradicts $\mathbf{v} \neq \mathbf{0}$. This establishes the equivalence of (1) and (2).

(2) \Leftrightarrow (3): Because every entry of $e^{\mathbf{A}t} \mathbf{B}$ is, as discussed at the end of Section 4.2, an analytical function of t , if $\mathbf{W}_c(t)$ is nonsingular for some t , then it is nonsingular for all t in $(-\infty, \infty)$. See Reference 6, p. 554. Because of the equivalence of the two forms in (6.2), (6.8) implies that $\mathbf{W}_c(t)$ is nonsingular if and only if there exists no $n \times 1$ nonzero vector \mathbf{v} such that

$$\mathbf{v}' e^{\mathbf{A}t} \mathbf{B} = \mathbf{0} \quad \text{for all } t \quad (6.9)$$

Now we show that if $\mathbf{W}_c(t)$ is nonsingular, then the controllability matrix \mathcal{C} has full row rank. Suppose \mathcal{C} does not have full row rank, then there exists an $n \times 1$ nonzero vector \mathbf{v} such that $\mathbf{v}' \mathcal{C} = \mathbf{0}$ or

$$\mathbf{v}' \mathbf{A}^k \mathbf{B} = \mathbf{0} \quad \text{for } k = 0, 1, 2, \dots, n-1$$

Because $e^{\mathbf{A}t} \mathbf{B}$ can be expressed as a linear combination of $\{\mathbf{B}, \mathbf{AB}, \dots, \mathbf{A}^{n-1} \mathbf{B}\}$ (Theorem 3.5), we conclude $\mathbf{v}' e^{\mathbf{A}t} \mathbf{B} = \mathbf{0}$. This contradicts the nonsingularity assumption of $\mathbf{W}_c(t)$. Thus Condition (2) implies Condition (3). To show

the converse, suppose \mathcal{C} has full row rank but $\mathbf{W}_c(t)$ is singular. Then there exists a nonzero \mathbf{v} such that (6.9) holds. Setting $t = 0$ yields $\mathbf{v}'\mathbf{B} = \mathbf{0}$. Differentiating (6.9) and then setting $t = 0$ yield $\mathbf{v}'\mathbf{AB} = \mathbf{0}$. Proceeding forward yields $\mathbf{v}'\mathbf{A}^k\mathbf{B} = \mathbf{0}$ for $k = 0, 1, 2, \dots$. They can be arranged as

$$\mathbf{v}'[\mathbf{B} \ \mathbf{AB} \ \cdots \ \mathbf{A}^{n-1}\mathbf{B}] = \mathbf{v}'\mathcal{C} = \mathbf{0}$$

This contradicts the hypothesis that \mathcal{C} has full row rank. This shows the equivalence of (2) and (3).

(3) \Leftrightarrow (4): If \mathcal{C} has full row rank, then the matrix $[\mathbf{A} - \lambda_1\mathbf{I} \ \mathbf{B}]$ has full row rank at every eigenvalue of \mathbf{A} . If not, there exists an eigenvalue λ_1 and a $1 \times n$ vector $\mathbf{q} \neq \mathbf{0}$ such that

$$\mathbf{q}[\mathbf{A} - \lambda_1\mathbf{I} \ \mathbf{B}] = \mathbf{0}$$

which implies $\mathbf{q}\mathbf{A} = \lambda_1\mathbf{q}$ and $\mathbf{q}\mathbf{B} = \mathbf{0}$. Thus \mathbf{q} is a left eigenvector of \mathbf{A} . We compute

$$\mathbf{q}\mathbf{A}^2 = (\mathbf{q}\mathbf{A})\mathbf{A} = (\lambda_1\mathbf{q})\mathbf{A} = \lambda_1^2\mathbf{q}$$

Proceeding forward, we have $\mathbf{q}\mathbf{A}^k = \lambda_1^k\mathbf{q}$. Thus we have

$$\mathbf{q}[\mathbf{B} \ \mathbf{AB} \ \cdots \ \mathbf{A}^{n-1}\mathbf{B}] = [\mathbf{q}\mathbf{B} \ \lambda_1\mathbf{q}\mathbf{B} \ \cdots \ \lambda_1^{n-1}\mathbf{q}\mathbf{B}] = \mathbf{0}$$

This contradicts the hypothesis that \mathcal{C} has full row rank.

In order to show that $\rho(\mathcal{C}) < n$ implies $\rho([\mathbf{A} - \lambda_1\mathbf{I} \ \mathbf{B}]) < n$ at some eigenvalue λ_1 of \mathbf{A} , we need Theorems 6.2 and 6.6, which will be established later. Theorem 6.2 states that controllability is invariant under any equivalence transformation. Therefore we may show $\rho([\bar{\mathbf{A}} - \lambda_1\bar{\mathbf{I}} \ \bar{\mathbf{B}}]) < n$ at some eigenvalue of $\bar{\mathbf{A}}$, where $(\bar{\mathbf{A}}, \bar{\mathbf{B}})$ is equivalent to (\mathbf{A}, \mathbf{B}) . Theorem 6.6 states that if the rank of \mathcal{C} is less than n or $\rho(\mathcal{C}) = n - m$, for some integer $m \geq 1$, then there exists a nonsingular matrix \mathbf{P} such that

$$\bar{\mathbf{A}} = \mathbf{P}\mathbf{A}\mathbf{P}^{-1} = \begin{bmatrix} \bar{\mathbf{A}}_c & \bar{\mathbf{A}}_{12} \\ \mathbf{0} & \bar{\mathbf{A}}_{\bar{c}} \end{bmatrix}, \quad \bar{\mathbf{B}} = \mathbf{P}\mathbf{B} = \begin{bmatrix} \bar{\mathbf{B}}_c \\ \mathbf{0} \end{bmatrix}$$

where $\bar{\mathbf{A}}_{\bar{c}}$ is $m \times m$. Let λ_1 be an eigenvalue of $\bar{\mathbf{A}}_{\bar{c}}$ and \mathbf{q}_1 be a corresponding $1 \times m$ nonzero left eigenvector of $\bar{\mathbf{A}}_{\bar{c}} = \lambda_1\mathbf{q}_1$. Then we have $\mathbf{q}_1(\bar{\mathbf{A}}_{\bar{c}} - \lambda_1\mathbf{I}) = \mathbf{0}$. Now we form the $1 \times n$ vector $\mathbf{q} := [\mathbf{0} \ \mathbf{q}_1]$. We compute

$$\mathbf{q}[\bar{\mathbf{A}} - \lambda_1\mathbf{I} \ \bar{\mathbf{B}}] = [\mathbf{0} \ \mathbf{q}_1] \begin{bmatrix} \bar{\mathbf{A}}_c - \lambda_1\mathbf{I} & \bar{\mathbf{A}}_{12} & \bar{\mathbf{B}}_c \\ \mathbf{0} & \bar{\mathbf{A}}_{\bar{c}} - \lambda_1\mathbf{I} & \mathbf{0} \end{bmatrix} = \mathbf{0} \quad (6.10)$$

which implies $\rho([\bar{\mathbf{A}} - \lambda_1\bar{\mathbf{I}} \ \bar{\mathbf{B}}]) < n$ and, consequently, $\rho([\mathbf{A} - \lambda_1\mathbf{I} \ \mathbf{B}]) < n$ at some eigenvalue of \mathbf{A} . This establishes the equivalence of (3) and (4).

(2) \Leftrightarrow (5): If \mathbf{A} is stable, then the unique solution of (6.4) can be expressed as in (6.5) (Theorem 5.6). The Gramian \mathbf{W}_c is always positive semidefinite. It is positive definite if and only if \mathbf{W}_c is nonsingular. This establishes the equivalence of (2) and (5). Q.E.D.

Example 6.2.2 Consider the inverted pendulum studied in Example 2.6.3. Its state-space equation was developed in (2.46). Suppose for a given pendulum, the equation becomes

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 5 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 1 \\ 0 \\ -2 \end{bmatrix} u(t) \quad (6.11)$$

$$y(t) = [1 \ 0 \ 0 \ 0] \mathbf{x}(t)$$

We compute

$$\mathcal{C} = [\mathbf{B} \ \mathbf{AB} \ \mathbf{A}^2\mathbf{B} \ \mathbf{A}^3\mathbf{B}] = \begin{bmatrix} 0 & 1 & 0 & 2 \\ 1 & 0 & 2 & 0 \\ 0 & -2 & 0 & -10 \\ -2 & 0 & -10 & 0 \end{bmatrix}$$

This matrix can be shown to have rank 4; thus the system is controllable. Therefore, if $x_3 = \theta$ deviates from zero slightly, we can find a control u to push it back to zero. In fact, a control exists to bring $x_1 = y$, x_3 and their derivatives back to zero. This is consistent with our experience of balancing a broom on our palm.

The MATLAB functions `ctrb` and `gram` will generate the controllability matrix and controllability Gramian. Note that the controllability Gramian is not computed from (6.5); it is obtained by solving a set of linear algebraic equations. Whether a state equation is controllable can then be determined by computing the rank of the controllability matrix or Gramian using the MATLAB function `rank`.

Example 6.2.3 Consider the platform system shown in Fig. 6.3; it can be used to study suspension systems of automobiles. The system consists of one platform; both ends of the platform are supported on the ground by means of springs and dashpots which provide viscous frictions. The mass of the platform is assumed to be zero; thus the movements of the two spring systems are independent, and half of the force is applied to each spring system. The spring constants of both springs are assumed to be 1, and the viscous friction coefficients are assumed to be 2 and 1 as shown. If the displacements of the two spring systems from the equilibrium are chosen as state variables x_1 and x_2 ,

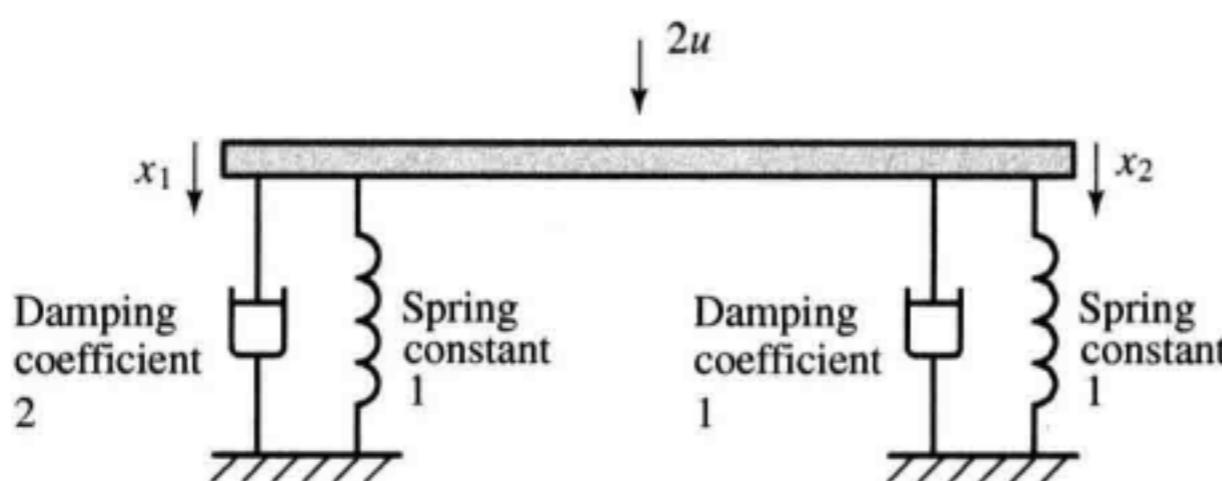


FIGURE 6.3 Platform system.

then we have $x_1 + 2\dot{x}_1 = u$ and $x_2 + \dot{x}_2 = u$ or

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -0.5 & 0 \\ 0 & -1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} u(t) \quad (6.12)$$

This state equation describes the system.

Now if the initial displacements are different from zero and if no force is applied, the platform will return to zero exponentially. In theory, it will take an infinite time for x_i to equal 0 exactly. Now we pose the problem: If $x_1(0) = 10$ and $x_2(0) = -1$, can we apply a force to bring the platform to the equilibrium in 2 seconds. The answer does not seem to be obvious because the *same* force is applied to the two spring systems.

For the equation in (6.12), we compute

$$\rho([\mathbf{B} \ \mathbf{AB}]) = \rho \begin{bmatrix} 0.5 & -0.25 \\ 1 & -1 \end{bmatrix} = 2$$

Thus the equation is controllable and, for any $\mathbf{x}(0)$, there exists an input that transfers $\mathbf{x}(0)$ to $\mathbf{0}$ in 2 seconds or in any finite time. We compute (6.2) and (6.7) for this system at $t_1 = 2$:

$$\begin{aligned} \mathbf{W}_c(2) &= \int_0^2 \left(\begin{bmatrix} e^{-0.5\tau} & 0 \\ 0 & e^{-\tau} \end{bmatrix} \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} [0.5 \ 1] \begin{bmatrix} e^{-0.5\tau} & 0 \\ 0 & e^{-\tau} \end{bmatrix} \right) d\tau \\ &= \begin{bmatrix} 0.2162 & 0.3167 \\ 0.3167 & 0.4908 \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} u_1(t) &= -[0.5 \ 1] \begin{bmatrix} e^{-0.5(2-t)} & 0 \\ 0 & e^{-(2-t)} \end{bmatrix} \mathbf{W}_c^{-1}(2) \begin{bmatrix} e^{-1} & 0 \\ 0 & e^{-2} \end{bmatrix} \begin{bmatrix} 10 \\ -1 \end{bmatrix} \\ &= -58.82e^{0.5t} + 27.96e^t \end{aligned}$$

for t in $[0, 2]$. This input force will transfer $\mathbf{x}(0) = [10 \ -1]'$ to $[0 \ 0]'$ in 2 seconds as shown in Fig. 6.4(a), in which the input is also plotted. It is obtained by using the MATLAB function `lsim`, acronym for *linear simulation*. The largest magnitude of the input is about 45.

Figure 6.4(b) plots the input u_2 that transfers $\mathbf{x}(0) = [10 \ -1]'$ to $\mathbf{0}$ in 4 seconds. We see that the smaller the time interval, the larger the input magnitude. If no restriction is imposed on the input, we can transfer $\mathbf{x}(0)$ to zero in an arbitrarily small time interval; however, the input magnitude may become very large. If some restriction is imposed on the input magnitude, then we cannot achieve the transfer as fast as desired. For example, if we require $|u(t)| < 9$ for all t , in Example 6.2.3, then we cannot transfer $\mathbf{x}(0)$ to $\mathbf{0}$ in less than 4 seconds. We remark that the input $\mathbf{u}(t)$ in (6.7) is called the *minimal energy control* in the sense that for any other input $\bar{\mathbf{u}}(t)$ that achieves the same transfer, we have

$$\int_{t_0}^{t_1} \bar{\mathbf{u}}'(t) \bar{\mathbf{u}}(t) dt \geq \int_{t_0}^{t_1} \mathbf{u}'(t) \mathbf{u}(t) dt$$

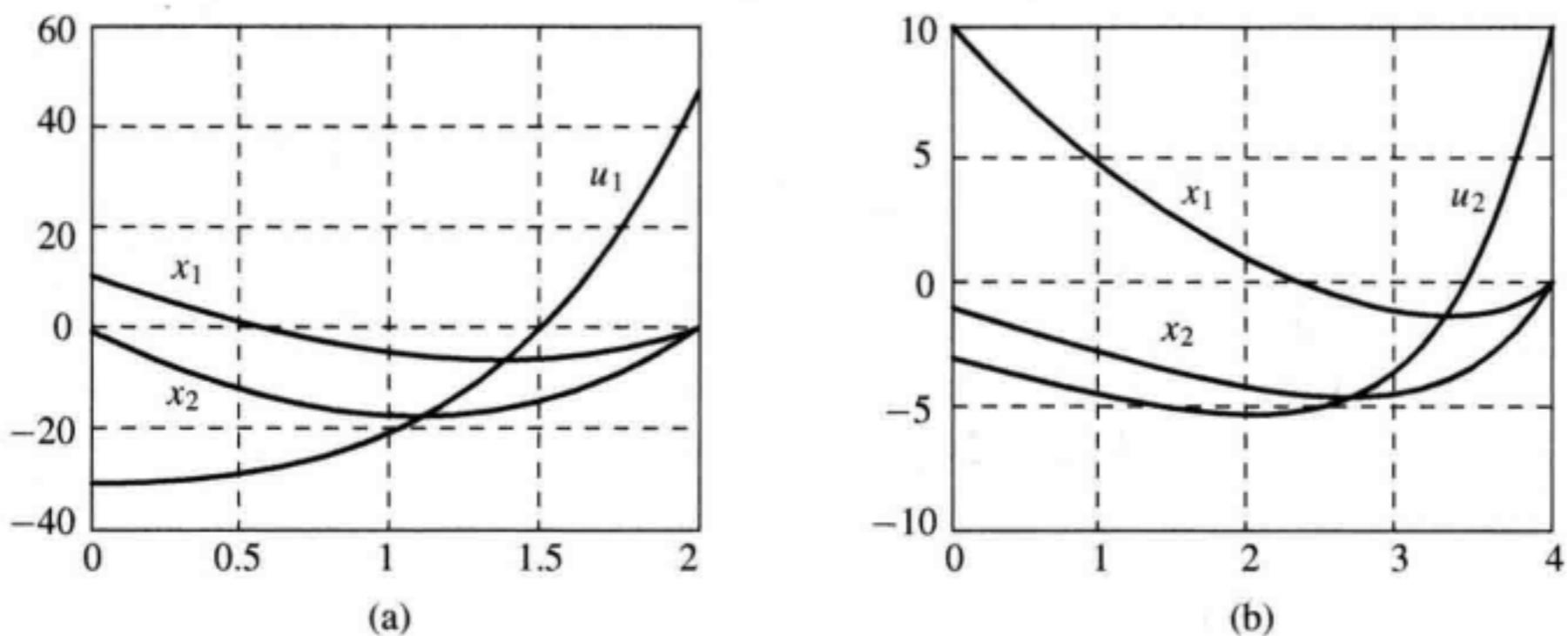


FIGURE 6.4 Transfer $\mathbf{x}(0) = [10 \ -1]'$ to $[0 \ 0]'$.

Its proof can be found in Reference 6, pp. 556–558.

Example 6.2.4 Consider again the platform system shown in Fig. 6.3. We now assume that the viscous friction coefficients and spring constants of both spring systems all equal 1. Then the state equation that describes the system becomes

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u(t)$$

Clearly we have

$$\rho(\mathcal{C}) = \rho \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} = 1$$

and the state equation is not controllable. If $x_1(0) \neq x_2(0)$, no input can transfer $\mathbf{x}(0)$ to zero in a finite time.

6.2.1 Controllability Indices²

Let \mathbf{A} and \mathbf{B} be $n \times n$ and $n \times p$ constant matrices. We assume that \mathbf{B} has rank p or full column rank. If \mathbf{B} does not have full column rank, there is a redundancy in inputs. For example, if the second column of \mathbf{B} equals the first column of \mathbf{B} , then the effect of the second input on the system can be generated from the first input. Thus the second input is redundant. In conclusion, deleting linear dependent columns of \mathbf{B} and the corresponding inputs will not affect the control of the system. Thus it is reasonable to assume that \mathbf{B} has full column rank.

If (\mathbf{A}, \mathbf{B}) is controllable, its controllability matrix \mathcal{C} has rank n and, consequently, n linearly independent columns. Note that there are np columns in \mathcal{C} ; therefore, it is possible to find many sets of n linearly independent columns in \mathcal{C} . We discuss in the following the most important way of searching these columns; the search also happens

²This subsection may be skipped without loss of continuity.

to be most natural. Let \mathbf{b}_i be the i th column of \mathbf{B} . Then \mathcal{C} can be written explicitly as

$$\mathcal{C} = [\mathbf{b}_1 \cdots \mathbf{b}_p : \mathbf{Ab}_1 \cdots \mathbf{Ab}_p : \cdots : \mathbf{A}^{n-1}\mathbf{b}_1 \cdots \mathbf{A}^{n-1}\mathbf{b}_p] \quad (6.13)$$

Let us search linearly independent columns of \mathcal{C} from left to right. Because of the pattern of \mathcal{C} , if $\mathbf{A}^i\mathbf{b}_m$ depends on its left-hand-side (LHS) columns, then $\mathbf{A}^{i+1}\mathbf{b}_m$ will also depend on its LHS columns. It means that once a column associated with \mathbf{b}_m becomes linearly dependent, then all columns associated with \mathbf{b}_m thereafter are linearly dependent. Let μ_m be the number of the linearly independent columns associated with \mathbf{b}_m in \mathcal{C} . That is, the columns

$$\mathbf{b}_m, \mathbf{Ab}_m, \dots, \mathbf{A}^{\mu_m-1}\mathbf{b}_m$$

are linearly independent in \mathcal{C} and $\mathbf{A}^{\mu_m+i}\mathbf{b}_m$ for $i = 0, 1, \dots$, are linearly dependent. It is clear that if \mathcal{C} has rank n , then

$$\mu_1 + \mu_2 + \cdots + \mu_p = n \quad (6.14)$$

The set $\{\mu_1, \mu_2, \dots, \mu_p\}$ is called the *controllability indices* and

$$\mu = \max(\mu_1, \mu_2, \dots, \mu_p)$$

is called the *controllability index* of (\mathbf{A}, \mathbf{B}) . Or, equivalently, if (\mathbf{A}, \mathbf{B}) is controllable, the controllability index μ is the least integer such that

$$\rho(\mathcal{C}_\mu) = \rho([\mathbf{B} \ \mathbf{AB} \ \cdots \ \mathbf{A}^{\mu-1}\mathbf{B}]) = n \quad (6.15)$$

Now we give a range of μ . If $\mu_1 = \mu_2 = \cdots = \mu_p$, then $n/p \leq \mu$. If all μ_m , except one, equal 1, then $\mu = n - (p - 1)$; this is the largest possible μ . Let \bar{n} be the degree of the minimal polynomial of \mathbf{A} . Then, by definition, there exist α_i such that

$$\mathbf{A}^{\bar{n}} = \alpha_1 \mathbf{A}^{\bar{n}-1} + \alpha_2 \mathbf{A}^{\bar{n}-2} + \cdots + \alpha_{\bar{n}} \mathbf{I}$$

which implies that $\mathbf{A}^{\bar{n}}\mathbf{B}$ can be written as a linear combination of $\{\mathbf{B}, \mathbf{AB}, \dots, \mathbf{A}^{\bar{n}-1}\mathbf{B}\}$. Thus we conclude

$$n/p \leq \mu \leq \min(\bar{n}, n - p + 1) \quad (6.16)$$

where $\rho(\mathbf{B}) = p$. Because of (6.16), in checking controllability, it is unnecessary to check the $n \times np$ matrix \mathcal{C} . It is sufficient to check a matrix of lesser columns. Because the degree of the minimal polynomial is generally not available, whereas the rank of \mathbf{B} can be readily computed, we can use the following corollary to check controllability. The second part of the corollary follows Theorem 3.8. We mention that the controllability index simply equals n for any SISO or SIMO n -dimensional controllable state-space equation.

► **Corollary 6.1**

The n -dimensional pair (\mathbf{A}, \mathbf{B}) is controllable if and only if the matrix

$$\mathcal{C}_{n-p+1} := [\mathbf{B} \ \mathbf{AB} \ \cdots \ \mathbf{A}^{n-p}\mathbf{B}] \quad (6.17)$$

where $\rho(\mathbf{B}) = p$, has rank n or the $n \times n$ matrix $\mathcal{C}_{n-p+1}\mathcal{C}'_{n-p+1}$ is nonsingular.

Example 6.2.5 Consider the satellite system studied in Fig. 2.15. Its linearized state-space equation was developed in (2.48). From the equation, we can see that the control of the first four state variables by the first two inputs and the control of the last two state variables by the last input are decoupled; therefore, we can consider only the following subequation of (2.48):

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 3 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & -2 & 0 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{u}(t) \\ \mathbf{y}(t) &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{x}(t) \end{aligned} \quad (6.18)$$

where we have assumed, for simplicity, $\omega_0 = m = r_o = 1$. The controllability matrix of (6.18) is of order 4×8 . If we use Corollary 6.1, then we can check its controllability by using the following 4×6 matrix

$$[\mathbf{B} \ \mathbf{AB} \ \mathbf{A}^2\mathbf{B}] = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 2 \\ 1 & 0 & 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 0 \\ 0 & 1 & -2 & 0 & 0 & -4 \end{bmatrix} \quad (6.19)$$

It has rank 4. Thus (6.18) is controllable. From (6.19), we can readily verify that the controllability indices are 2 and 2, and the controllability index is 2.

► **Theorem 6.2**

The controllability property is invariant under any equivalence transformation.



Proof: Consider the pair (\mathbf{A}, \mathbf{B}) with controllability matrix

$$\mathcal{C} = [\mathbf{B} \ \mathbf{AB} \ \cdots \ \mathbf{A}^{n-1}\mathbf{B}]$$

and its equivalent pair $(\bar{\mathbf{A}}, \bar{\mathbf{B}})$ with $\bar{\mathbf{A}} = \mathbf{P}\mathbf{A}\mathbf{P}^{-1}$ and $\bar{\mathbf{B}} = \mathbf{P}\mathbf{B}$, where \mathbf{P} is a nonsingular matrix. The controllability matrix of $(\bar{\mathbf{A}}, \bar{\mathbf{B}})$ is

$$\begin{aligned} \bar{\mathcal{C}} &= [\bar{\mathbf{B}} \ \bar{\mathbf{A}}\bar{\mathbf{B}} \ \cdots \ \bar{\mathbf{A}}^{n-1}\bar{\mathbf{B}}] \\ &= [\mathbf{PB} \ \mathbf{PAP}^{-1}\mathbf{PB} \ \cdots \ \mathbf{PA}^{n-1}\mathbf{P}^{-1}\mathbf{PB}] \\ &= [\mathbf{PB} \ \mathbf{PAB} \ \cdots \ \mathbf{PA}^{n-1}\mathbf{B}] \\ &= \mathbf{P}[\mathbf{B} \ \mathbf{AB} \ \cdots \ \mathbf{A}^{n-1}\mathbf{B}] = \mathbf{PC} \end{aligned} \quad (6.20)$$

Because \mathbf{P} is nonsingular, we have $\rho(\mathcal{C}) = \rho(\bar{\mathcal{C}})$ (see Equation (3.62)). This establishes Theorem 6.2. Q.E.D.

► Theorem 6.3

The set of the controllability indices of (\mathbf{A}, \mathbf{B}) is invariant under any equivalence transformation and any reordering of the columns of \mathbf{B} .



Proof: Let us define

$$\mathcal{C}_k = [\mathbf{B} \ \mathbf{AB} \ \cdots \ \mathbf{A}^{k-1}\mathbf{B}] \quad (6.21)$$

Then we have, following the proof of Theorem 6.2,

$$\rho(\mathcal{C}_k) = \rho(\bar{\mathcal{C}}_k)$$

for $k = 0, 1, 2, \dots$. Thus the set of controllability indices is invariant under any equivalence transformation.

The rearrangement of the columns of \mathbf{B} can be achieved by

$$\hat{\mathbf{B}} = \mathbf{BM}$$

where \mathbf{M} is a $p \times p$ nonsingular permutation matrix. It is straightforward to verify

$$\hat{\mathcal{C}}_k := [\hat{\mathbf{B}} \ \hat{\mathbf{AB}} \ \cdots \ \hat{\mathbf{A}}^{k-1}\hat{\mathbf{B}}] = \mathcal{C}_k \text{diag}(\mathbf{M}, \mathbf{M}, \dots, \mathbf{M})$$

Because $\text{diag}(\mathbf{M}, \mathbf{M}, \dots, \mathbf{M})$ is nonsingular, we have $\rho(\hat{\mathcal{C}}_k) = \rho(\mathcal{C}_k)$ for $k = 0, 1, \dots$. Thus the set of controllability indices is invariant under any reordering of the columns of \mathbf{B} . Q.E.D.

Because the set of the controllability indices is invariant under any equivalence transformation and any rearrangement of the inputs, it is an intrinsic property of the system that the state-space equation describes. The physical significance of the controllability index is not transparent here, but it becomes obvious in the discrete-time case. As we will discuss in later chapters, the controllability index can also be computed from transfer matrices and dictates the minimum degree required to achieve pole placement and model matching.

6.3 Observability

The concept of observability is dual to that of controllability. Roughly speaking, controllability studies the possibility of steering the state from the input; observability studies the possibility of estimating the state from the output. These two concepts are defined under the assumption that the state-space equation or, equivalently, all \mathbf{A} , \mathbf{B} , \mathbf{C} , and \mathbf{D} are known. Thus the problem of observability is different from the problem of realization or identification, which is to determine or estimate \mathbf{A} , \mathbf{B} , \mathbf{C} , and \mathbf{D} from the information collected at the input and output terminals.

Consider the n -dimensional p -input q -output state-space equation

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t)\end{aligned}\quad (6.22)$$

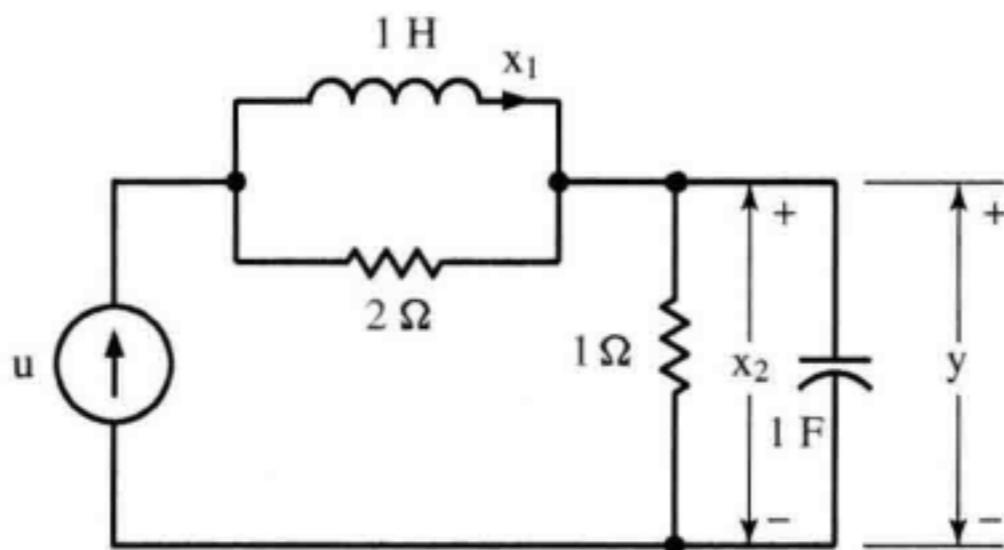
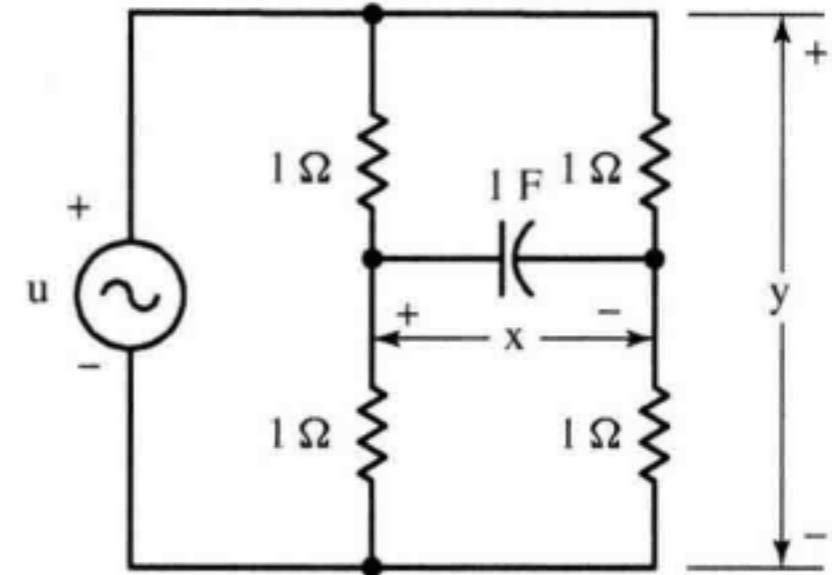
where \mathbf{A} , \mathbf{B} , \mathbf{C} , and \mathbf{D} are, respectively, $n \times n$, $n \times p$, $q \times n$, and $q \times p$ constant matrices.

Definition 6.01 The state-space equation (6.22) is said to be *observable* if for any unknown initial state $\mathbf{x}(0)$, there exists a finite $t_1 > 0$ such that the knowledge of the input \mathbf{u} and the output \mathbf{y} over $[0, t_1]$ suffices to determine uniquely the initial state $\mathbf{x}(0)$. Otherwise, the equation is said to be *unobservable*.

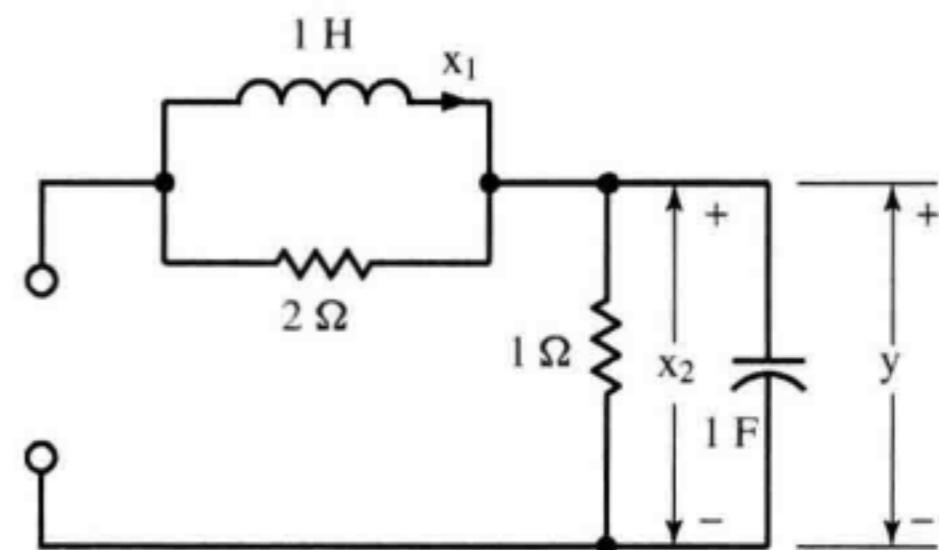
Example 6.3.1 Consider the circuit shown in Fig. 6.5. If the input is zero, no matter what the initial voltage across the capacitor is, the output is identically zero because of the symmetry of the four resistors. We know the input and output (both are identically zero), but we cannot determine uniquely the initial state. Thus the circuit or, more precisely, the state-space equation that describes the circuit is not observable.

Example 6.3.2 Consider the circuit shown in Fig. 6.6(a). The circuit has two state variables: the current x_1 through the inductor and the voltage x_2 across the capacitor. The input u is a current source. If $u = 0$, the circuit reduces to the one shown in Fig. 6.6(b). If $x_1(0) = a \neq 0$ and $x_2 = 0$, then the output is identically zero. Any $\mathbf{x}(0) = [a \ 0]'$ and

FIGURE 6.5 Unobservable circuit.



(a)



(b)

FIGURE 6.6 Unobservable circuit.

$\mathbf{u}(t) \equiv 0$ yield the same output $\mathbf{y}(t) \equiv 0$. Thus there is no way to determine the initial state $[a \ 0]'$ uniquely and the equation that describes the circuit is not observable.

The response of (6.22) excited by the initial state $\mathbf{x}(0)$ and the input $\mathbf{u}(t)$ was derived in (4.7) as

$$\mathbf{y}(t) = \mathbf{C}e^{\mathbf{A}t}\mathbf{x}(0) + \mathbf{C} \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)d\tau + \mathbf{D}\mathbf{u}(t) \quad (6.23)$$

In the study of observability, the output \mathbf{y} and the input \mathbf{u} are assumed to be known, the initial state $\mathbf{x}(0)$ is the only unknown; thus we can write (6.23) as

$$\mathbf{C}e^{\mathbf{A}t}\mathbf{x}(0) = \bar{\mathbf{y}}(t) \quad (6.24)$$

where

$$\bar{\mathbf{y}}(t) := \mathbf{y}(t) - \mathbf{C} \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)d\tau - \mathbf{D}\mathbf{u}(t)$$

is a known function. Thus the observability problem reduces to solving $\mathbf{x}(0)$ from (6.24). If $\mathbf{u} \equiv \mathbf{0}$, then $\bar{\mathbf{y}}(t)$ reduces to the zero-input response $\mathbf{C}e^{\mathbf{A}t}\mathbf{x}(0)$. Thus Definition 6.O1 can be modified as: Equation (6.22) is observable if and only if the initial state $\mathbf{x}(0)$ can be uniquely determined from its zero-input response over a finite time interval.

Next we discuss how to solve $\mathbf{x}(0)$ from (6.24). For a fixed t , $\mathbf{C}e^{\mathbf{A}t}$ is a $q \times n$ constant matrix, $\bar{\mathbf{y}}(t)$ is a $q \times 1$ constant vector. Thus (6.24) is a set of linear algebraic equations with n unknowns. Because of the way it is developed, for every fixed t , $\bar{\mathbf{y}}(t)$ is in the range space of $\mathbf{C}e^{\mathbf{A}t}$ and solutions always exist in (6.24). The only question is whether the solution is unique. If $q < n$ as is the case in general, the $q \times n$ matrix $\mathbf{C}e^{\mathbf{A}t}$ has rank at most q and, consequently, has nullity $n - q$ or larger. Thus solutions are not unique (Theorem 3.2). In conclusion, we cannot find a unique $\mathbf{x}(0)$ from (6.24) at an isolated t . In order to determine $\mathbf{x}(0)$ uniquely from (6.24), we must use the knowledge of $\mathbf{u}(t)$ and $\mathbf{y}(t)$ over a nonzero time interval as stated in the next theorem.

Theorem 6.4

The state-space equation (6.22) is observable if and only if the $n \times n$ matrix

$$\mathbf{W}_o(t) = \int_0^t e^{\mathbf{A}'\tau} \mathbf{C}' \mathbf{C} e^{\mathbf{A}\tau} d\tau \quad (6.25)$$

is nonsingular for any $t > 0$.

⇒ **Proof:** We premultiply (6.24) by $e^{\mathbf{A}'t}\mathbf{C}'$ and then integrate it over $[0, t_1]$ to yield

$$\left(\int_0^{t_1} e^{\mathbf{A}'t} \mathbf{C}' \mathbf{C} e^{\mathbf{A}t} dt \right) \mathbf{x}(0) = \int_0^{t_1} e^{\mathbf{A}'t} \mathbf{C}' \bar{\mathbf{y}}(t) dt$$

If $\mathbf{W}_o(t_1)$ is nonsingular, then

$$\mathbf{x}(0) = \mathbf{W}_o^{-1}(t_1) \int_0^{t_1} e^{\mathbf{A}'t} \mathbf{C}' \bar{\mathbf{y}}(t) dt \quad (6.26)$$

This yields a unique $\mathbf{x}(0)$. This shows that if $\mathbf{W}_o(t)$, for any $t > 0$, is nonsingular, then (6.22) is observable. Next we show that if $\mathbf{W}_o(t_1)$ is singular or, equivalently, positive semidefinite for all t_1 , then (6.22) is not observable. If $\mathbf{W}_o(t_1)$ is positive semidefinite, there exists an $n \times 1$ nonzero constant vector \mathbf{v} such that

$$\begin{aligned}\mathbf{v}'\mathbf{W}_o(t_1)\mathbf{v} &= \int_0^{t_1} \mathbf{v}' e^{\mathbf{A}'\tau} \mathbf{C}' \mathbf{C} e^{\mathbf{A}\tau} \mathbf{v} d\tau \\ &= \int_0^{t_1} \|\mathbf{C} e^{\mathbf{A}\tau} \mathbf{v}\|^2 d\tau = 0\end{aligned}$$

which implies

$$\mathbf{C} e^{\mathbf{A}t} \mathbf{v} \equiv \mathbf{0} \quad (6.27)$$

for all t in $[0, t_1]$. If $\mathbf{u} \equiv \mathbf{0}$, then $\mathbf{x}_1(0) = \mathbf{v} \neq \mathbf{0}$ and $\mathbf{x}_2(0) = \mathbf{0}$ both yield the same

$$\mathbf{y}(t) = \mathbf{C} e^{\mathbf{A}t} \mathbf{x}_i(0) \equiv \mathbf{0}$$

Two different initial states yield the same zero-input response; therefore, we cannot uniquely determine $\mathbf{x}(0)$. Thus (6.22) is not observable. This completes the proof of Theorem 6.4. Q.E.D.

We see from this theorem that observability depends only on \mathbf{A} and \mathbf{C} . This can also be deduced from Definition 6.01 by choosing $\mathbf{u}(t) \equiv \mathbf{0}$. Thus observability is a property of the pair (\mathbf{A}, \mathbf{C}) and is independent of \mathbf{B} and \mathbf{D} . As in the controllability part, if $\mathbf{W}_o(t)$ is nonsingular for some t , then it is nonsingular for every t and the initial state can be computed from (6.26) by using any nonzero time interval.

► Theorem 6.5 (Theorem of duality)

The pair (\mathbf{A}, \mathbf{B}) is controllable if and only if the pair $(\mathbf{A}', \mathbf{B}')$ is observable.



Proof: The pair (\mathbf{A}, \mathbf{B}) is controllable if and only if

$$\mathbf{W}_c(t) = \int_0^t e^{\mathbf{A}\tau} \mathbf{B} \mathbf{B}' e^{\mathbf{A}'\tau} d\tau$$

is nonsingular for any t . The pair $(\mathbf{A}', \mathbf{B}')$ is observable if and only if, by replacing \mathbf{A} by \mathbf{A}' and \mathbf{C} by \mathbf{B}' in (6.25),

$$\mathbf{W}_o(t) = \int_0^t e^{\mathbf{A}\tau} \mathbf{B} \mathbf{B}' e^{\mathbf{A}'\tau} d\tau$$

is nonsingular for any t . The two conditions are identical and the theorem follows. Q.E.D.

We list in the following the observability counterpart of Theorem 6.1. It can be proved either directly or by applying Theorem of duality.

► Theorem 6.01

The following statements are equivalent.

1. The n -dimensional pair (\mathbf{A}, \mathbf{C}) is observable.
2. The $n \times n$ matrix

$$\mathbf{W}_o(t) = \int_0^t e^{\mathbf{A}'\tau} \mathbf{C}' \mathbf{C} e^{\mathbf{A}\tau} d\tau \quad (6.28)$$

is nonsingular for any $t > 0$.

3. The $nq \times n$ observability matrix

$$\mathcal{O} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{n-1} \end{bmatrix} \quad (6.29)$$

has rank n (full column rank). This matrix can be generated by calling `obsv` in MATLAB.

4. The $(n+q) \times n$ matrix $\begin{bmatrix} \mathbf{A} - \lambda \mathbf{I} \\ \mathbf{C} \end{bmatrix}$ has full column rank at every eigenvalue, λ , of \mathbf{A} .
5. If, in addition, all eigenvalues of \mathbf{A} have negative real parts, then the unique solution of

$$\mathbf{A}' \mathbf{W}_o + \mathbf{W}_o \mathbf{A} = -\mathbf{C}' \mathbf{C} \quad (6.30)$$

is positive definite. The solution is called the *observability Gramian* and can be expressed as

$$\mathbf{W}_o = \int_0^\infty e^{\mathbf{A}'\tau} \mathbf{C}' \mathbf{C} e^{\mathbf{A}\tau} d\tau \quad (6.31)$$

6.3.1 Observability Indices³

Let \mathbf{A} and \mathbf{C} be $n \times n$ and $q \times n$ constant matrices. We assume that \mathbf{C} has rank q (full row rank). If \mathbf{C} does not have full row rank, then the output at some output terminal can be expressed as a linear combination of other outputs. Thus the output does not offer any new information regarding the system and the terminal can be eliminated. By deleting the corresponding row, the reduced \mathbf{C} will then have full row rank.

If (\mathbf{A}, \mathbf{C}) is observable, its observability matrix \mathcal{O} has rank n and, consequently, n linearly independent rows. Let \mathbf{c}_i be the i th row of \mathbf{C} . Let us search linearly independent rows of \mathcal{O} in order from top to bottom. Dual to the controllability part, if a row associated with \mathbf{c}_m becomes linearly dependent on its upper rows, then all rows associated with \mathbf{c}_m thereafter will also be dependent. Let v_m be the number of the linearly independent rows associated with \mathbf{c}_m . It is clear that if \mathcal{O} has rank n , then

$$v_1 + v_2 + \cdots + v_q = n \quad (6.32)$$

³This subsection may be skipped without loss of continuity.

The set $\{v_1, v_2, \dots, v_q\}$ is called the *observability indices* and

$$v = \max(v_1, v_2, \dots, v_q) \quad (6.33)$$

is called the *observability index* of (A, C) . If (A, C) is observable, it is the least integer such that

$$\rho(\mathcal{O}_v) := \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{v-1} \end{bmatrix} = n$$

Dual to the controllability part, we have

$$n/q \leq v \leq \min(\bar{n}, n - q + 1) \quad (6.34)$$

where $\rho(C) = q$ and \bar{n} is the degree of the minimal polynomial. We mention that the observability index simply equals n for any SISO or MISO n -dimensional observable state-space equation.

Corollary 6.01

The n -dimensional pair (A, C) is observable if and only if the matrix

$$\mathcal{O}_{n-q+1} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-q} \end{bmatrix} \quad (6.35)$$

where $\rho(C) = q$, has rank n or the $n \times n$ matrix $\mathcal{O}'_{n-q+1} \mathcal{O}_{n-q+1}$ is nonsingular.

Theorem 6.02

The observability property is invariant under any equivalence transformation.

Theorem 6.03

The set of the observability indices of (A, C) is invariant under any equivalence transformation and any reordering of the rows of C .

Before concluding this section, we discuss a different way of solving (6.24). Differentiating (6.24) repeatedly and setting $t = 0$, we can obtain

$$\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{v-1} \end{bmatrix} \mathbf{x}(0) = \begin{bmatrix} \bar{\mathbf{y}}(0) \\ \dot{\bar{\mathbf{y}}}(0) \\ \vdots \\ \ddot{\bar{\mathbf{y}}}^{(v-1)}(0) \end{bmatrix}$$

or

$$\mathcal{O}_{v-1} \mathbf{x}(0) = \tilde{\mathbf{y}}(0) \quad (6.36)$$

where $\bar{\mathbf{y}}^{(i)}(t)$ is the i th derivative of $\bar{\mathbf{y}}(t)$, and $\tilde{\mathbf{y}}(0) := [\bar{\mathbf{y}}'(0) \quad \dot{\bar{\mathbf{y}}}'(0) \quad \dots \quad (\bar{\mathbf{y}}^{(v-1)})']'$. Equation (6.36) is a set of linear algebraic equations. Because of the way it is developed, $\tilde{\mathbf{y}}(0)$ must lie in the range space of \mathcal{O}_{v-1} . Thus a solution $\mathbf{x}(0)$ exists in (6.36). If (\mathbf{A}, \mathbf{C}) is observable, then \mathcal{O}_{v-1} has full column rank and following Theorem 3.2, the solution is unique. Premultiplying (6.36) by \mathcal{O}_{v-1}' and then using Theorem 3.8, we can obtain the solution as

$$\mathbf{x}(0) = [\mathcal{O}_{v-1} \mathcal{O}_{v-1}]^{-1} \mathcal{O}_{v-1}' \tilde{\mathbf{y}}(0) \quad (6.37)$$

We mention that in order to obtain $\dot{\bar{\mathbf{y}}}(0), \ddot{\bar{\mathbf{y}}}(0), \dots$, we need the knowledge of $\bar{\mathbf{y}}(t)$ in the neighborhood of $t = 0$. This is consistent with the earlier assertion that we need the knowledge of $\bar{\mathbf{y}}(t)$ over a nonzero time interval in order to determine $\mathbf{x}(0)$ uniquely from (6.24). In conclusion, the initial state can be computed using (6.26) or (6.37).

The output $\mathbf{y}(t)$ measured in practice is often corrupted by high-frequency noise. Because

- differentiation will amplify high-frequency noise and
- integration will suppress or smooth high frequency noise,

the result obtained from (6.36) or (6.37) may differ greatly from the actual initial state. Thus (6.26) is preferable to (6.36) in computing initial states.

The physical significance of the observability index can be seen from (6.36). It is the smallest integer in order to determine $\mathbf{x}(0)$ uniquely from (6.36) or (6.37). It also dictates the minimum degree required to achieve pole placement and model matching, as we will discuss in Chapter 9.

6.4 Kalman Decomposition

This section discusses canonical decomposition of state-space equations. This fundamental result will be used to establish the relationship between the state-space description and the transfer-matrix description. Consider

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{Ax}(t) + \mathbf{Bu}(t) \\ \mathbf{y}(t) &= \mathbf{Cx}(t) + \mathbf{Du}(t) \end{aligned} \quad (6.38)$$

Let $\bar{\mathbf{x}} = \mathbf{Px}$, where \mathbf{P} is a nonsingular matrix. Then the state equation

$$\begin{aligned} \dot{\bar{\mathbf{x}}}(t) &= \bar{\mathbf{A}}\bar{\mathbf{x}}(t) + \bar{\mathbf{B}}\mathbf{u}(t) \\ \mathbf{y}(t) &= \bar{\mathbf{C}}\bar{\mathbf{x}}(t) + \bar{\mathbf{D}}\mathbf{u}(t) \end{aligned} \quad (6.39)$$

with $\bar{\mathbf{A}} = \mathbf{PAP}^{-1}$, $\bar{\mathbf{B}} = \mathbf{PB}$, $\bar{\mathbf{C}} = \mathbf{CP}^{-1}$ and $\bar{\mathbf{D}} = \mathbf{D}$ is equivalent to (6.38). All properties of (6.38), including stability, controllability, and observability, are preserved in (6.39). We also have

$$\bar{\mathcal{C}} = \mathbf{PC}, \quad \bar{\mathcal{O}} = \mathcal{O}\mathbf{P}^{-1}$$

► **Theorem 6.6**

Consider the n -dimensional state-space equation in (6.38) with

$$\rho(\mathcal{C}) = \rho([\mathbf{B} \ \mathbf{AB} \ \cdots \ \mathbf{A}^{n-1}\mathbf{B}]) = n_1 < n$$

We form the $n \times n$ matrix

$$\mathbf{P}^{-1} := [\mathbf{q}_1 \ \cdots \ \mathbf{q}_{n_1} \ \cdots \ \mathbf{q}_n]$$

where the first n_1 columns are any n_1 linearly independent columns of \mathcal{C} , and the remaining columns can be arbitrarily chosen so long as \mathbf{P} is nonsingular. Then the equivalence transformation $\bar{\mathbf{x}} = \mathbf{Px}$ or $\mathbf{x} = \mathbf{P}^{-1}\bar{\mathbf{x}}$ will transform (6.38) into

$$\begin{aligned} \begin{bmatrix} \dot{\bar{\mathbf{x}}}_c(t) \\ \dot{\bar{\mathbf{x}}}_{\bar{c}}(t) \end{bmatrix} &= \begin{bmatrix} \bar{\mathbf{A}}_c & \bar{\mathbf{A}}_{12} \\ \mathbf{0} & \bar{\mathbf{A}}_{\bar{c}} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}_c(t) \\ \bar{\mathbf{x}}_{\bar{c}}(t) \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{B}}_c \\ \mathbf{0} \end{bmatrix} \mathbf{u}(t) \\ \mathbf{y}(t) &= [\bar{\mathbf{C}}_c \ \bar{\mathbf{C}}_{\bar{c}}] \begin{bmatrix} \bar{\mathbf{x}}_c(t) \\ \bar{\mathbf{x}}_{\bar{c}}(t) \end{bmatrix} + \mathbf{Du}(t) \end{aligned} \quad (6.40)$$

where $\bar{\mathbf{A}}_c$ is $n_1 \times n_1$ and $\bar{\mathbf{A}}_{\bar{c}}$ is $(n - n_1) \times (n - n_1)$, and the following n_1 -dimensional subequation of (6.40)

$$\begin{aligned} \dot{\bar{\mathbf{x}}}_c(t) &= \bar{\mathbf{A}}_c \bar{\mathbf{x}}_c(t) + \bar{\mathbf{B}}_c \mathbf{u}(t) \\ \bar{\mathbf{y}}(t) &= \bar{\mathbf{C}}_c \bar{\mathbf{x}}_c(t) + \mathbf{Du}(t) \end{aligned} \quad (6.41)$$

is controllable and has the same transfer matrix as (6.38).



Proof: As discussed in Section 4.4, the transformation $\mathbf{x} = \mathbf{P}^{-1}\bar{\mathbf{x}}$ changes the basis of the state space from the orthonormal basis in (3.8) to the columns of $\mathbf{Q} := \mathbf{P}^{-1}$ or $\{\mathbf{q}_1, \dots, \mathbf{q}_{n_1}, \dots, \mathbf{q}_n\}$. The i th column of $\bar{\mathbf{A}}$ is the representation of $\mathbf{A}\mathbf{q}_i$ with respect to $\{\mathbf{q}_1, \dots, \mathbf{q}_{n_1}, \dots, \mathbf{q}_n\}$. Now the vector $\mathbf{A}\mathbf{q}_i$, for $i = 1, 2, \dots, n_1$, are linearly dependent on the set $\{\mathbf{q}_1, \dots, \mathbf{q}_{n_1}\}$; they are linearly independent of $\{\mathbf{q}_{n_1+1}, \dots, \mathbf{q}_n\}$. Thus the matrix $\bar{\mathbf{A}}$ has the form shown in (6.40). The columns of $\bar{\mathbf{B}}$ are the representation of the columns of \mathbf{B} with respect to $\{\mathbf{q}_1, \dots, \mathbf{q}_{n_1}, \dots, \mathbf{q}_n\}$. Now the columns of \mathbf{B} depend only on $\{\mathbf{q}_1, \dots, \mathbf{q}_{n_1}\}$; thus $\bar{\mathbf{B}}$ has the form shown in (6.40). We mention that if the $n \times p$ matrix \mathbf{B} has rank p and if its columns are chosen as the first p columns of \mathbf{P}^{-1} , then the upper part of $\bar{\mathbf{B}}$ is the unit matrix of order p .

Let $\bar{\mathcal{C}}$ be the controllability matrix of (6.40). Then we have $\rho(\mathcal{C}) = \rho(\bar{\mathcal{C}}) = n_1$. It is straightforward to verify

$$\begin{aligned} \bar{\mathcal{C}} &= \begin{bmatrix} \bar{\mathbf{B}}_c & \bar{\mathbf{A}}_c \bar{\mathbf{B}}_c & \cdots & \bar{\mathbf{A}}_c^{n_1} \bar{\mathbf{B}}_c & \cdots & \bar{\mathbf{A}}_c^{n-1} \bar{\mathbf{B}}_c \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \\ &= \begin{bmatrix} \bar{\mathcal{C}}_c & \bar{\mathbf{A}}_c^{n_1} \bar{\mathbf{B}}_c & \cdots & \bar{\mathbf{A}}_c^{n-1} \bar{\mathbf{B}}_c \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \end{aligned}$$

where $\bar{\mathcal{C}}_c$ is the controllability matrix of $(\bar{\mathbf{A}}_c, \bar{\mathbf{B}}_c)$. Because the columns of $\bar{\mathbf{A}}_c^k \bar{\mathbf{B}}_c$, for $k \geq n_1$, are linearly dependent on the columns of $\bar{\mathcal{C}}_c$, the condition $\rho(\mathcal{C}) = n_1$

implies $\rho(\bar{\mathcal{C}}_c) = n_1$. Thus the n_1 -dimensional state-space equation in (6.41) is controllable.

Next we show that (6.41) has the same transfer matrix as (6.38). Because (6.38) and (6.40) have the same transfer matrix, we need to show only that (6.40) and (6.41) have the same transfer matrix. By direct verification, we can show

$$\begin{bmatrix} s\mathbf{I} - \bar{\mathbf{A}}_c & -\bar{\mathbf{A}}_{12} \\ \mathbf{0} & s\mathbf{I} - \bar{\mathbf{A}}_{\bar{c}} \end{bmatrix}^{-1} = \begin{bmatrix} (s\mathbf{I} - \bar{\mathbf{A}}_c)^{-1} & \mathbf{M} \\ \mathbf{0} & (s\mathbf{I} - \bar{\mathbf{A}}_{\bar{c}})^{-1} \end{bmatrix} \quad (6.42)$$

where

$$\mathbf{M} = (s\mathbf{I} - \bar{\mathbf{A}}_c)^{-1} \bar{\mathbf{A}}_{12} (s\mathbf{I} - \bar{\mathbf{A}}_{\bar{c}})^{-1}$$

Thus the transfer matrix of (6.40) is

$$\begin{aligned} & [\bar{\mathbf{C}}_c \quad \bar{\mathbf{C}}_{\bar{c}}] \begin{bmatrix} s\mathbf{I} - \bar{\mathbf{A}}_c & -\bar{\mathbf{A}}_{12} \\ \mathbf{0} & s\mathbf{I} - \bar{\mathbf{A}}_{\bar{c}} \end{bmatrix}^{-1} \begin{bmatrix} \bar{\mathbf{B}}_c \\ \mathbf{0} \end{bmatrix} + \mathbf{D} \\ &= [\bar{\mathbf{C}}_c \quad \bar{\mathbf{C}}_{\bar{c}}] \begin{bmatrix} (s\mathbf{I} - \bar{\mathbf{A}}_c)^{-1} & \mathbf{M} \\ \mathbf{0} & (s\mathbf{I} - \bar{\mathbf{A}}_{\bar{c}})^{-1} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{B}}_c \\ \mathbf{0} \end{bmatrix} + \mathbf{D} \\ &= \bar{\mathbf{C}}_c (s\mathbf{I} - \bar{\mathbf{A}}_c)^{-1} \bar{\mathbf{B}}_c + \mathbf{D} \end{aligned}$$

which is the transfer matrix of (6.41). This completes the proof of Theorem 6.6. Q.E.D.

In the equivalence transformation $\bar{\mathbf{x}} = \mathbf{P}\mathbf{x}$, the n -dimensional state space is divided into two subspaces. One is the n_1 -dimensional subspace that consists of all vectors of form $[\bar{\mathbf{x}}'_c \quad \mathbf{0}']'$; the other is the $(n - n_1)$ -dimensional subspace that consists of all vectors of form $[\mathbf{0}' \quad \bar{\mathbf{x}}'_{\bar{c}}]'$. Because (6.41) is controllable, the input \mathbf{u} can transfer $\bar{\mathbf{x}}_c$ from any state to any other state. However, the input \mathbf{u} cannot control $\bar{\mathbf{x}}_{\bar{c}}$ because, as we can see from (6.40), \mathbf{u} does not affect $\bar{\mathbf{x}}_{\bar{c}}$ directly, nor does it affect $\bar{\mathbf{x}}_{\bar{c}}$ indirectly through the state $\bar{\mathbf{x}}_c$. By dropping the uncontrollable state vector, we obtain a controllable state equation of lesser dimension that is zero-state equivalent to the original equation.

Example 6.4.1 Consider the three-dimensional state-space equation

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} u(t), \quad y(t) = [1 \ 1 \ 1] \mathbf{x}(t) \quad (6.43)$$

The rank of \mathbf{B} is 2; therefore, we can use $\mathcal{C}_2 = [\mathbf{B} \ \mathbf{AB}]$, instead of $\mathcal{C} = [\mathbf{B} \ \mathbf{AB} \ \mathbf{A}^2 \mathbf{B}]$, to check the controllability of (6.43) (Corollary 6.1). Because

$$\rho(\mathcal{C}_2) = \rho([\mathbf{B} \ \mathbf{AB}]) = \rho \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} = 2 < 3$$

the state equation in (6.43) is not controllable. Let us choose

$$\mathbf{P}^{-1} = \mathbf{Q} := \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

The first two columns of \mathbf{Q} are the first two linearly independent columns of \mathcal{C}_2 ; the last column is chosen arbitrarily to make \mathbf{Q} nonsingular. Let $\bar{\mathbf{x}} = \mathbf{Px}$. We compute

$$\begin{aligned}\bar{\mathbf{A}} &= \mathbf{PAP}^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & \vdots & 0 \\ 1 & 1 & \vdots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \vdots & 1 \end{bmatrix} \\ \bar{\mathbf{B}} &= \mathbf{PB} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \dots & \dots \\ 0 & 0 \end{bmatrix} \\ \bar{\mathbf{C}} &= \mathbf{CP}^{-1} = [1 \ 1 \ 1] \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = [1 \ 2 : 1]\end{aligned}$$

Note that the 1×2 submatrix $\bar{\mathbf{A}}_{21}$ of $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}_c$ are zero as expected. The 2×1 submatrix $\bar{\mathbf{A}}_{12}$ happens to be zero; it could be nonzero. The upper part of $\bar{\mathbf{B}}$ is a unit matrix because the columns of \mathbf{B} are the first two columns of \mathbf{Q} . Thus (6.43) can be reduced to

$$\dot{\bar{\mathbf{x}}}_c(t) = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \bar{\mathbf{x}}_c(t) + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{u}(t), \quad y(t) = [1 \ 2] \bar{\mathbf{x}}_c(t)$$

This equation is controllable and has the same transfer matrix as (6.43).

The MATLAB function `ctrbf` transforms (6.38) into (6.40) except that the order of the columns in \mathbf{P}^{-1} is reversed. Thus the resulting equation has the form

$$\begin{bmatrix} \bar{\mathbf{A}}_c & \mathbf{0} \\ \bar{\mathbf{A}}_{21} & \bar{\mathbf{A}}_c \end{bmatrix} \quad \begin{bmatrix} \mathbf{0} \\ \bar{\mathbf{B}}_c \end{bmatrix}$$

Theorem 6.6 is established from the controllability matrix. In actual computation, it is unnecessary to form the controllability matrix. The result can be obtained by carrying out a sequence of similarity transformations to transform $[\mathbf{B} \ \mathbf{A}]$ into a Hessenberg form.

See Reference 6, pp. 220–222. This procedure is efficient and numerically stable and should be used in actual computation.

Dual to Theorem 6.6, we have the following theorem for unobservable state-space equations.

► Theorem 6.06

Consider the n -dimensional state-space equation in (6.38) with

$$\rho(\mathcal{O}) = \rho \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{n-1} \end{bmatrix} = n_2 < n$$

We form the $n \times n$ matrix

$$\mathbf{P} = \begin{bmatrix} \mathbf{p}_1 \\ \vdots \\ \mathbf{p}_{n_2} \\ \vdots \\ \mathbf{p}_n \end{bmatrix}$$

where the first n_2 rows are any n_2 linearly independent rows of \mathcal{O} , and the remaining rows can be arbitrarily chosen so long as \mathbf{P} is nonsingular. Then the equivalence transformation $\bar{\mathbf{x}} = \mathbf{Px}$ will transform (6.38) into

$$\begin{aligned} \begin{bmatrix} \dot{\bar{\mathbf{x}}}_o(t) \\ \dot{\bar{\mathbf{x}}}_{\bar{o}}(t) \end{bmatrix} &= \begin{bmatrix} \bar{\mathbf{A}}_o & \mathbf{0} \\ \bar{\mathbf{A}}_{21} & \bar{\mathbf{A}}_{\bar{o}} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}_o(t) \\ \bar{\mathbf{x}}_{\bar{o}}(t) \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{B}}_o \\ \bar{\mathbf{B}}_{\bar{o}} \end{bmatrix} \mathbf{u}(t) \\ \mathbf{y}(t) &= [\bar{\mathbf{C}}_o \ \mathbf{0}] \begin{bmatrix} \bar{\mathbf{x}}_o(t) \\ \bar{\mathbf{x}}_{\bar{o}}(t) \end{bmatrix} + \mathbf{D}\mathbf{u}(t) \end{aligned} \quad (6.44)$$

where $\bar{\mathbf{A}}_o$ is $n_2 \times n_2$ and $\bar{\mathbf{A}}_{\bar{o}}$ is $(n - n_2) \times (n - n_2)$, and the following n_2 -dimensional subequation of (6.44)

$$\begin{aligned} \dot{\bar{\mathbf{x}}}_o(t) &= \bar{\mathbf{A}}_o \bar{\mathbf{x}}_o(t) + \bar{\mathbf{B}}_o \mathbf{u}(t) \\ \bar{\mathbf{y}}(t) &= \bar{\mathbf{C}}_o \bar{\mathbf{x}}_o(t) + \mathbf{D}\mathbf{u}(t) \end{aligned}$$

is observable and has the same transfer matrix as (6.38).

In the equivalence transformation $\bar{\mathbf{x}} = \mathbf{Px}$, the n -dimensional state space is divided into two subspaces. One is the n_2 -dimensional subspace that consists of all vectors of form $[\bar{\mathbf{x}}'_o \ \mathbf{0}']'$; the other is the $(n - n_2)$ -dimensional subspace consisting of all vectors of form $[\mathbf{0}' \ \bar{\mathbf{x}}'_{\bar{o}}]'$. The state $\bar{\mathbf{x}}_o$ can be detected from the output. However, $\bar{\mathbf{x}}_{\bar{o}}$ cannot be detected from the output because, as we can see from (6.44), it is not connected to the output directly, nor is it connected indirectly through the state $\bar{\mathbf{x}}_o$. By dropping the unobservable state vector, we obtain an observable state equation of lesser dimension that is zero-state equivalent to the original equation. The MATLAB function `obsvf` is the counterpart of `ctrbf`. Combining Theorems 6.6 and 6.06, we have the following *Kalman decomposition theorem*.

► **Theorem 6.7**

Every state-space equation can be transformed, by an equivalence transformation, into the following canonical form:

$$\begin{bmatrix} \dot{\bar{\mathbf{x}}}_{co}(t) \\ \dot{\bar{\mathbf{x}}}_{c\bar{o}}(t) \\ \dot{\bar{\mathbf{x}}}_{\bar{c}o}(t) \\ \dot{\bar{\mathbf{x}}}_{\bar{c}\bar{o}}(t) \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{A}}_{co} & \mathbf{0} & \bar{\mathbf{A}}_{13} & \mathbf{0} \\ \bar{\mathbf{A}}_{21} & \bar{\mathbf{A}}_{c\bar{o}} & \bar{\mathbf{A}}_{23} & \bar{\mathbf{A}}_{24} \\ \mathbf{0} & \mathbf{0} & \bar{\mathbf{A}}_{\bar{c}o} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \bar{\mathbf{A}}_{43} & \bar{\mathbf{A}}_{\bar{c}\bar{o}} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}_{co}(t) \\ \bar{\mathbf{x}}_{c\bar{o}}(t) \\ \bar{\mathbf{x}}_{\bar{c}o}(t) \\ \bar{\mathbf{x}}_{\bar{c}\bar{o}}(t) \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{B}}_{co} \\ \bar{\mathbf{B}}_{c\bar{o}} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \mathbf{u}(t)$$

$$\mathbf{y}(t) = [\bar{\mathbf{C}}_{co} \ \mathbf{0} \ \bar{\mathbf{C}}_{\bar{c}o} \ \mathbf{0}] \bar{\mathbf{x}}(t) + \mathbf{D}\mathbf{u}(t) \quad (6.45)$$

where the vector $\bar{\mathbf{x}}_{co}$ is controllable and observable, $\bar{\mathbf{x}}_{c\bar{o}}$ is controllable but not observable, $\bar{\mathbf{x}}_{\bar{c}o}$ is observable but not controllable, and $\bar{\mathbf{x}}_{\bar{c}\bar{o}}$ is neither controllable nor observable. Furthermore, the state-space equation is zero-state equivalent to the controllable and observable state-space equation

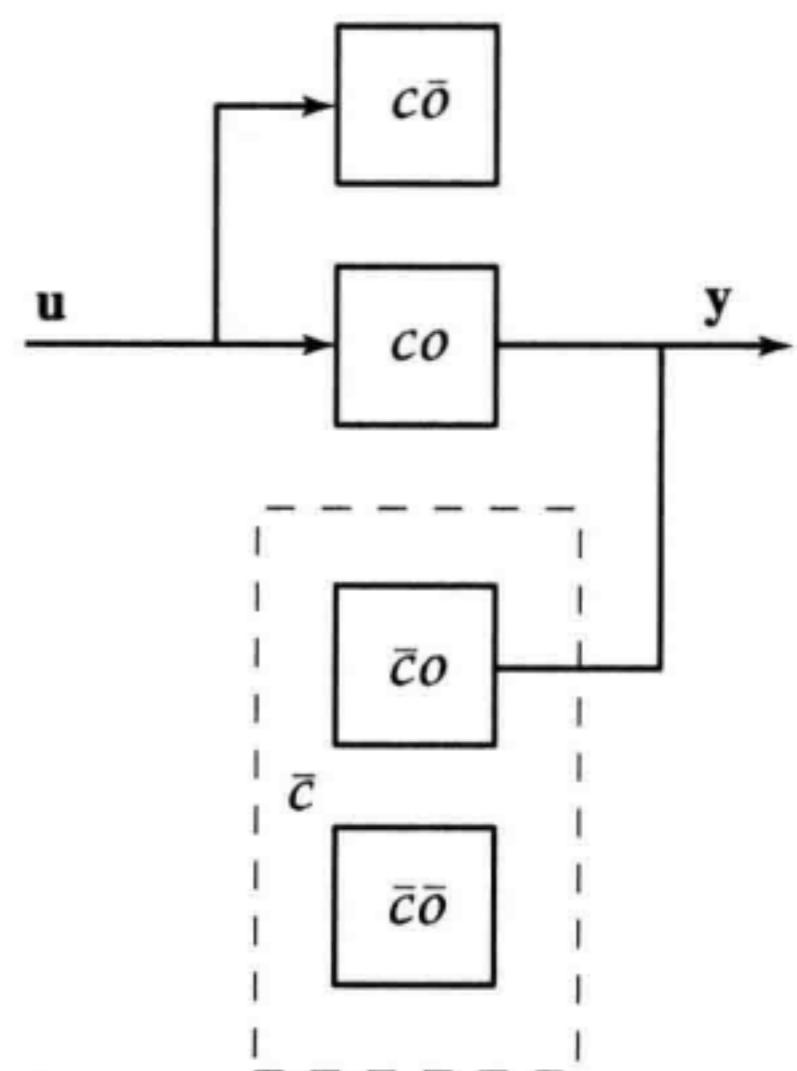
$$\begin{aligned} \dot{\bar{\mathbf{x}}}_{co}(t) &= \bar{\mathbf{A}}_{co}\bar{\mathbf{x}}_{co}(t) + \bar{\mathbf{B}}_{co}\mathbf{u}(t) \\ \mathbf{y}(t) &= \bar{\mathbf{C}}_{co}\bar{\mathbf{x}}_{co}(t) + \mathbf{D}\mathbf{u}(t) \end{aligned} \quad (6.46)$$

and has the transfer matrix

$$\hat{\mathbf{G}}(s) = \bar{\mathbf{C}}_{co}(s\mathbf{I} - \bar{\mathbf{A}}_{co})^{-1}\bar{\mathbf{B}}_{co} + \mathbf{D}$$

This theorem can be illustrated symbolically as shown in Fig. 6.7. The equation is first decomposed, using Theorem 6.6, into controllable and uncontrollable subequations. We then decompose each subequation, using Theorem 6.O6, into observable and unobservable parts. From the figure, we see that only the controllable and observable part is connected to both the input and output terminals. Thus the transfer

FIGURE 6.7 Kalman decomposition.



matrix describes only this part of the system. This is the reason that the transfer-function description and the state-space description are not necessarily equivalent. For example, if any A-matrix other than \bar{A}_{co} has an eigenvalue with a positive real part, then some state variable may grow without bound and the system may burn out. This phenomenon, however, cannot be detected from the transfer matrix.

The MATLAB function `minreal`, acronym for *minimal realization*, can reduce (6.45) to (6.46). The reason for calling it minimal realization will be given in the next chapter.

Example 6.4.2 Consider the circuit shown in Fig. 6.8(a). Because the input is a current source, responses due to the initial conditions in C_1 and L_1 will not appear at the output. Thus the state variables associated with C_1 and L_1 are not observable; whether or not they are controllable is immaterial in subsequent discussion. Similarly, the state variable associated with L_2 is not controllable. Because of the symmetry of the four 1Ω resistors, the state variable associated with C_2 is neither controllable nor observable. By dropping the state variables that are either uncontrollable or unobservable, the circuit in Fig. 6.8(a) can be reduced to the one in Fig. 6.8(b). The current in each branch is $u/2$; thus the output y equals $2 \cdot (u/2)$ or $y = u$. Thus the transfer function of the circuit in Fig. 6.8(a) is $\hat{g}(s) = 1$.

If we assign state variables as shown, then the circuit can be described by

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & -0.5 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & -0.5 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0.5 \\ 0 \\ 0 \\ 0 \end{bmatrix} u(t)$$

$$y(t) = [0 \ 0 \ 0 \ 1] \mathbf{x}(t) + u(t)$$

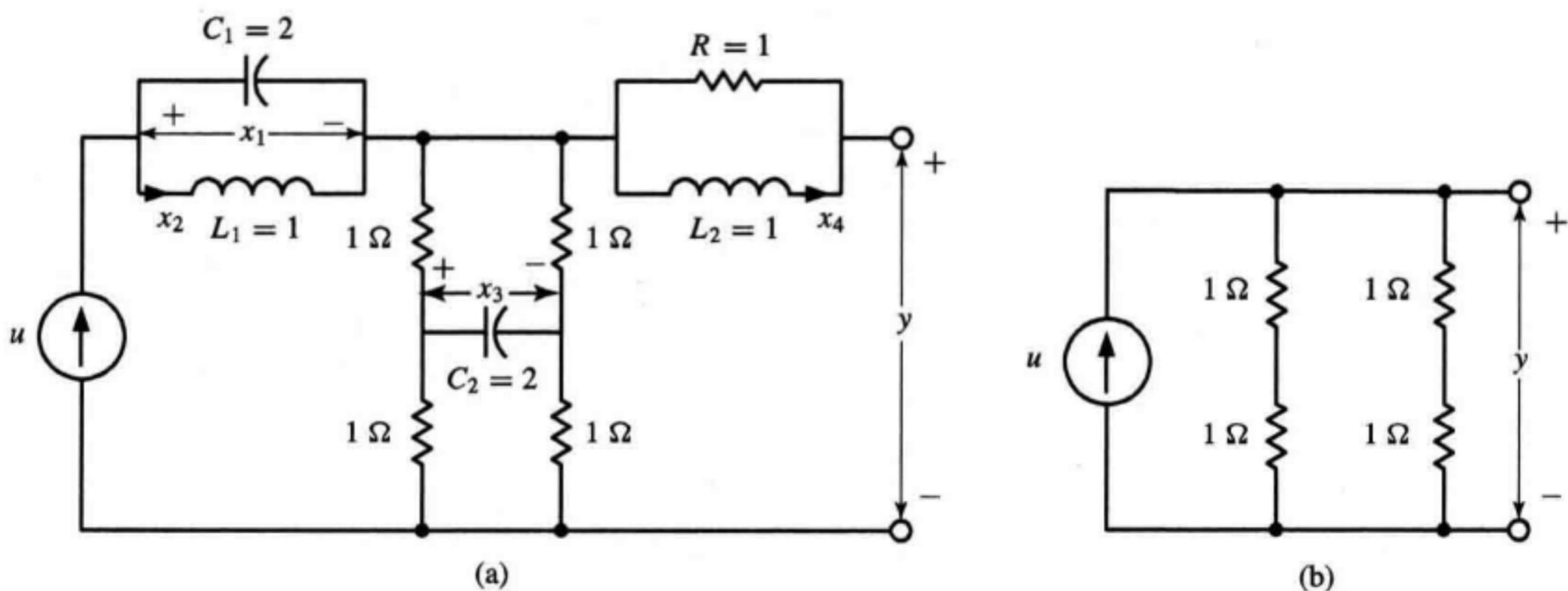


FIGURE 6.8 Circuits.

Because the equation is already of the form shown in (6.40), it can be reduced to the following controllable state-space equation:

$$\begin{aligned}\dot{\mathbf{x}}_c(t) &= \begin{bmatrix} 0 & -0.5 \\ 1 & 0 \end{bmatrix} \mathbf{x}_c(t) + \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} u(t) \\ y(t) &= [0 \ 0] \mathbf{x}_c(t) + u(t)\end{aligned}$$

The output is independent of \mathbf{x}_c , thus the equation can be further reduced to $y = u$. This is what we will obtain by using the MATLAB function `minreal`.

6.5 Conditions in Jordan-Form Equations

Controllability and observability are invariant under any equivalence transformation. If a state-space equation is transformed into Jordan form, then the controllability and observability conditions become very simple and can often be checked by inspection. Consider the state-space equation

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{J}\mathbf{x}(t) + \mathbf{B}\mathbf{x}(t) \\ y(t) &= \mathbf{C}\mathbf{x}(t)\end{aligned}\tag{6.47}$$

where \mathbf{J} is in Jordan form. To simplify the discussion, we assume that \mathbf{J} has only two distinct eigenvalues λ_1 and λ_2 and can be written as

$$\mathbf{J} = \text{diag}(\mathbf{J}_1, \mathbf{J}_2)$$

where \mathbf{J}_1 consists of all Jordan blocks associated with λ_1 and \mathbf{J}_2 consists of all Jordan blocks associated with λ_2 . Again to simplify the discussion, we assume that \mathbf{J}_1 has three Jordan blocks and \mathbf{J}_2 has two Jordan blocks or

$$\mathbf{J}_1 = \text{diag}(\mathbf{J}_{11}, \mathbf{J}_{12}, \mathbf{J}_{13}) \quad \mathbf{J}_2 = \text{diag}(\mathbf{J}_{21}, \mathbf{J}_{22})$$

The row of \mathbf{B} corresponding to the *last row* of \mathbf{J}_{ij} is denoted by \mathbf{b}_{lij} . The column of \mathbf{C} corresponding to the *first column* of \mathbf{J}_{ij} is denoted by \mathbf{c}_{fij} .

Theorem 6.8

1. The state-space equation in (6.47) is controllable if and only if the three row vectors $\{\mathbf{b}_{l11}, \mathbf{b}_{l12}, \mathbf{b}_{l13}\}$ are linearly independent and the two row vectors $\{\mathbf{b}_{l21}, \mathbf{b}_{l22}\}$ are linearly independent.
2. The state-space equation in (6.47) is observable if and only if the three column vectors $\{\mathbf{c}_{f11}, \mathbf{c}_{f12}, \mathbf{c}_{f13}\}$ are linearly independent and the two column vectors $\{\mathbf{c}_{f21}, \mathbf{c}_{f22}\}$ are linearly independent.

We discuss first the implications of this theorem. If a state-space equation is in Jordan form, then the controllability of the state variables associated with one eigenvalue can be checked independently from those associated with different eigenvalues. The controllability of the state variables associated with the same eigenvalue depends only on the rows of \mathbf{B} corresponding to the last row of all Jordan blocks associated with

the eigenvalue. All other rows of \mathbf{B} play no role in determining the controllability. Similar remarks apply to the observability part except that the columns of \mathbf{C} corresponding to the first column of all Jordan blocks determine the observability. We use an example to illustrate the use of Theorem 6.8.

Example 6.5.1 Consider the Jordan-form state-space equation

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \begin{bmatrix} \lambda_1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_2 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 3 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \mathbf{u}(t) \\ \mathbf{y}(t) &= \begin{bmatrix} 1 & 1 & 2 & 0 & 0 & 2 & 1 \\ 1 & 0 & 1 & 2 & 0 & 1 & 1 \\ 1 & 0 & 2 & 3 & 0 & 2 & 0 \end{bmatrix} \mathbf{x}(t)\end{aligned}\quad (6.48)$$

The matrix \mathbf{J} has two distinct eigenvalues λ_1 and λ_2 . There are three Jordan blocks, with order 2, 1, and 1, associated with λ_1 . The rows of \mathbf{B} corresponding to the last row of the three Jordan blocks are $[1 \ 0 \ 0]$, $[0 \ 1 \ 0]$, and $[1 \ 1 \ 1]$. The three rows are linearly independent. There is only one Jordan block, with order 3, associated with λ_2 . The row of \mathbf{B} corresponding to the last row of the Jordan block is $[1 \ 1 \ 1]$, which is nonzero and is, therefore, linearly independent. Thus we conclude that the state equation in (6.48) is controllable.

The conditions for (6.48) to be observable are the three columns $[1 \ 1 \ 1]', [2 \ 1 \ 2]',$ and $[0 \ 2 \ 3]'$ are linearly independent (they are) and the one column $[0 \ 0 \ 0]'$ is linearly independent (it is not). Therefore the state-space equation is not observable.

Before proving Theorem 6.8, we draw a block diagram to show how the conditions in the theorem arise. The inverse of $(s\mathbf{I} - \mathbf{J})$ is of the form shown in (3.49) whose entries consist of only $1/(s - \lambda_i)^k$. Using (3.49), we can draw a block diagram for (6.48) as shown in Fig. 6.9. Each chain of blocks corresponds to one Jordan block in the equation. Because (6.48) has four Jordan blocks, the figure has four chains. The output of each block can be assigned as a state variable as shown in Fig. 6.10. Let us consider the last chain in Fig. 6.9. If $\mathbf{b}_{l21} = \mathbf{0}$, the state variable x_{l21} is not connected to the input and is not controllable no matter what values \mathbf{b}_{221} and \mathbf{b}_{121} assume. On the other hand, if \mathbf{b}_{l21} is nonzero, then all state variables in the chain are controllable. If there are two or more chains associated with the same eigenvalue, then we require the linear independence of the first gain vectors of those chains. The chains associated with different eigenvalues can be checked separately. All discussion applies to the the observability part except that the column vector \mathbf{c}_{fj} plays the role of the row vector \mathbf{b}_{lij} .



Proof: Proof of Theorem 6.8: We prove the theorem by using the condition that the matrix $[\mathbf{A} - s\mathbf{I} \ \mathbf{B}]$ or $[s\mathbf{I} - \mathbf{A} \ \mathbf{B}]$ has full row rank at every eigenvalue of \mathbf{A} . In order not to be overwhelmed by notation, we assume $[s\mathbf{I} - \mathbf{J} \ \mathbf{B}]$ to be

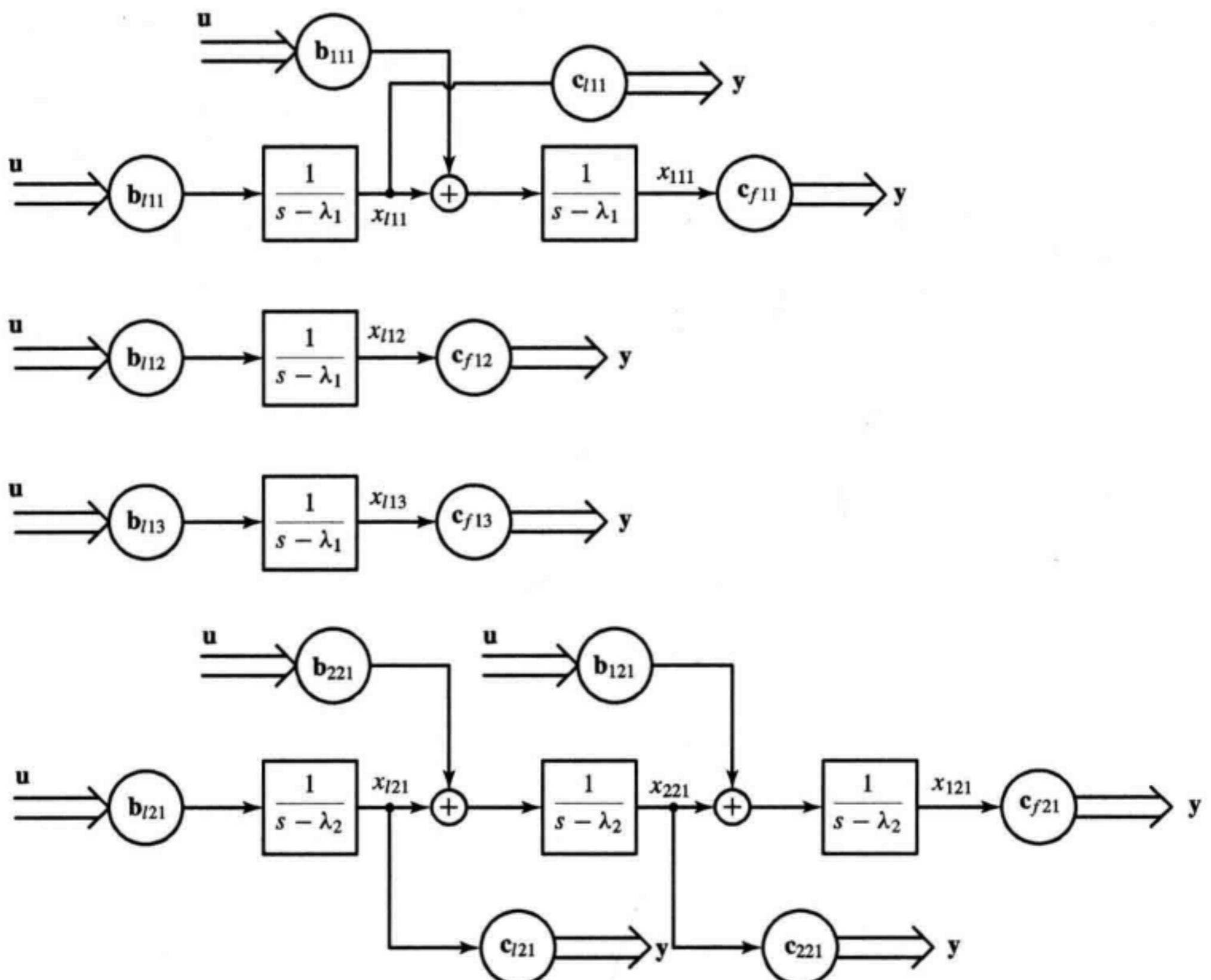
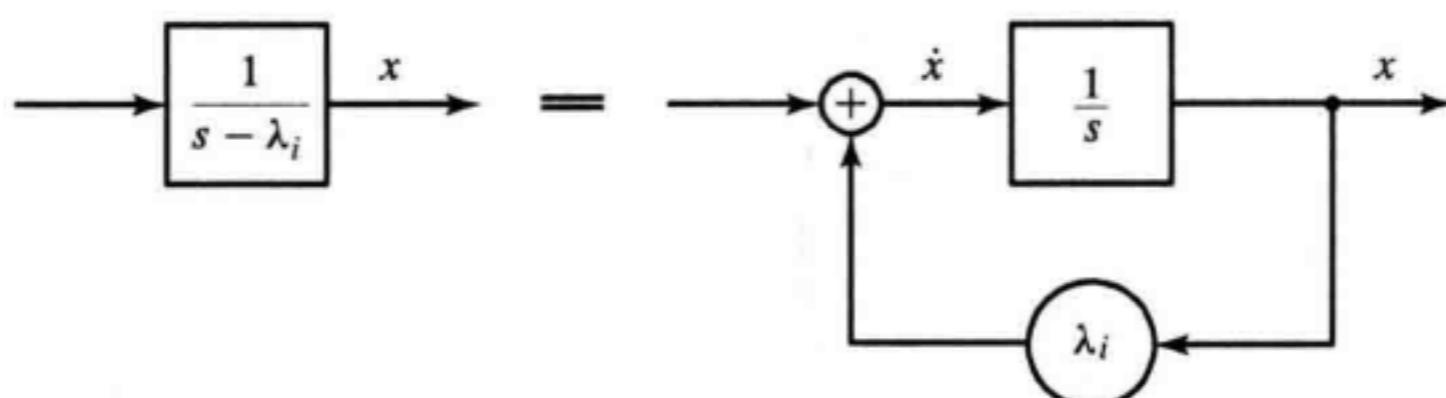


FIGURE 6.9 Block diagram of (6.48).

FIGURE 6.10 Internal structure of $1/(s - \lambda_i)$

of the form

$$\begin{bmatrix} s - \lambda_1 & -1 & 0 & 0 & 0 & 0 & 0 & \mathbf{b}_{111} \\ 0 & s - \lambda_1 & -1 & 0 & 0 & 0 & 0 & \mathbf{b}_{211} \\ 0 & 0 & s - \lambda_1 & 0 & 0 & 0 & 0 & \mathbf{b}_{l11} \\ 0 & 0 & 0 & s - \lambda_1 & -1 & 0 & 0 & \mathbf{b}_{112} \\ 0 & 0 & 0 & 0 & s - \lambda_1 & 0 & 0 & \mathbf{b}_{l12} \\ 0 & 0 & 0 & 0 & 0 & s - \lambda_2 & -1 & \mathbf{b}_{121} \\ 0 & 0 & 0 & 0 & 0 & 0 & s - \lambda_2 & \mathbf{b}_{l21} \end{bmatrix} \quad (6.49)$$

The Jordan-form matrix \mathbf{J} has two distinct eigenvalues λ_1 and λ_2 . There are two Jordan blocks associated with λ_1 and one associated with λ_2 . If $s = \lambda_1$, (6.49) becomes

$$\begin{bmatrix} 0 & -1 & 0 & 0 & 0 & 0 & 0 & \mathbf{b}_{111} \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & \mathbf{b}_{211} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{b}_{l11} \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & \mathbf{b}_{112} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{b}_{l12} \\ 0 & 0 & 0 & 0 & 0 & \lambda_1 - \lambda_2 & -1 & \mathbf{b}_{121} \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_1 - \lambda_2 & \mathbf{b}_{l21} \end{bmatrix} \quad (6.50)$$

The rank of the matrix will not change by elementary column operations. We add the product of the second column of (6.50) by \mathbf{b}_{111} to the last block column. Repeating the process for the third and fifth columns, we can obtain

$$\begin{bmatrix} 0 & -1 & 0 & 0 & 0 & 0 & 0 & \mathbf{0} \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & \mathbf{0} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{b}_{l11} \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & \mathbf{0} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{b}_{l12} \\ 0 & 0 & 0 & 0 & 0 & \lambda_1 - \lambda_2 & -1 & \mathbf{b}_{121} \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_1 - \lambda_2 & \mathbf{b}_{l21} \end{bmatrix}$$

Because λ_1 and λ_2 are distinct, $\lambda_1 - \lambda_2$ is nonzero. We add the product of the seventh column and $-\mathbf{b}_{l21}/(\lambda_1 - \lambda_2)$ to the last column and then use the sixth column to eliminate its right-hand-side entries to yield

$$\begin{bmatrix} 0 & -1 & 0 & 0 & 0 & 0 & 0 & \mathbf{0} \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & \mathbf{0} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{b}_{l11} \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & \mathbf{0} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{b}_{l12} \\ 0 & 0 & 0 & 0 & 0 & \lambda_1 - \lambda_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_1 - \lambda_2 & 0 \end{bmatrix} \quad (6.51)$$

It is clear that the matrix in (6.51) has full row rank if and only if \mathbf{b}_{l11} and \mathbf{b}_{l12} are linearly independent. Proceeding similarly for each eigenvalue, we can establish Theorem 6.8. Q.E.D.

Consider an n -dimensional Jordan-form state-space equation with p inputs and q outputs. If there are m , with $m > p$, Jordan blocks associated with the same eigenvalue, then m number of $1 \times p$ row vectors can never be linearly independent and the state-space equation can never be controllable. Thus a necessary condition for the state-space equation to be controllable is $m \leq p$. Similarly, a necessary condition for the state-space equation to be observable is $m \leq q$. For the single-input case, we then have the following corollaries.

► **Corollary 6.8**

A single-input Jordan-form state-space equation is controllable if and only if there is only one Jordan block associated with each distinct eigenvalue and every entry of \mathbf{B} corresponding to the last row of each Jordan block is different from zero.

► **Corollary 6.08**

A single-output Jordan-form state-space equation is observable if and only if there is only one Jordan block associated with each distinct eigenvalue and every entry of \mathbf{C} corresponding to the first column of each Jordan block is different from zero.

Example 6.5.2 Consider the state-space equation

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 10 \\ 9 \\ 0 \\ 1 \end{bmatrix} u(t) \quad (6.52)$$

$$y(t) = [1 \ 0 \ 0 \ 2] \mathbf{x}(t)$$

There are two Jordan blocks, one with order 3 and associated with eigenvalue 0, the other with order 1 and associated with eigenvalue -2 . The entry of \mathbf{B} corresponding to the last row of the first Jordan block is zero; thus the state equation is not controllable. The two entries of \mathbf{C} corresponding to the first column of both Jordan blocks are different from zero, thus the state-space equation is observable.

6.6 Discrete-Time State-Space Equations

Consider the n -dimensional p -input q -output state-space equation

$$\begin{aligned} \mathbf{x}[k+1] &= \mathbf{Ax}[k] + \mathbf{Bu}[k] \\ \mathbf{y}[k] &= \mathbf{Cx}[k] \end{aligned} \quad (6.53)$$

where \mathbf{A} , \mathbf{B} , and \mathbf{C} are, respectively, $n \times n$, $n \times p$, and $q \times n$ real constant matrices.

Definition 6.D1 The discrete-time state equation (6.53) or the pair (\mathbf{A}, \mathbf{B}) is said to be *controllable* if for any initial state $\mathbf{x}(0) = \mathbf{x}_0$ and any final state \mathbf{x}_1 , there exists an input sequence of finite length that transfers \mathbf{x}_0 to \mathbf{x}_1 . Otherwise the equation or (\mathbf{A}, \mathbf{B}) is said to be *uncontrollable*.

► **Theorem 6.D1**

The following statements are equivalent.

1. The n -dimensional pair (\mathbf{A}, \mathbf{B}) is controllable.
2. The $n \times n$ matrix

$$\mathbf{W}_{dc}[n-1] = \sum_{m=0}^{n-1} (\mathbf{A})^m \mathbf{B} \mathbf{B}' (\mathbf{A}')^m \quad (6.54)$$

is nonsingular.

3. The $n \times np$ controllability matrix

$$\mathcal{C}_d = [\mathbf{B} \ \mathbf{AB} \ \mathbf{A}^2\mathbf{B} \ \cdots \ \mathbf{A}^{n-1}\mathbf{B}] \quad (6.55)$$

has rank n (full row rank). The matrix can be generated by calling `ctrb` in MATLAB.

- 4.** The $n \times (n+p)$ matrix $[\mathbf{A} - \lambda \mathbf{I} \ \mathbf{B}]$ has full row rank at every eigenvalue, λ , of \mathbf{A} .
- 5.** If, in addition, all eigenvalues of \mathbf{A} have magnitudes less than 1, then the unique solution of

$$\mathbf{W}_{dc} - \mathbf{AW}_{dc}\mathbf{A}' = \mathbf{BB}' \quad (6.56)$$

is positive definite. The solution is called the discrete *controllability Gramian* and can be obtained by using the MATLAB function `dgram`. The discrete Gramian can be expressed as

$$\mathbf{W}_{dc} = \sum_{m=0}^{\infty} \mathbf{A}^m \mathbf{BB}' (\mathbf{A}')^m \quad (6.57)$$

The solution of (6.53) at $k = n$ was derived in Subsection 4.2.2 as

$$x[n] = \mathbf{A}^n \mathbf{x}[0] + \sum_{m=0}^{n-1} \mathbf{A}^{n-1-m} \mathbf{B} \mathbf{u}[m]$$

which can be written as

$$\mathbf{x}[n] - \mathbf{A}^n \mathbf{x}[0] = [\mathbf{B} \ \mathbf{AB} \ \cdots \ \mathbf{A}^{n-1}\mathbf{B}] \begin{bmatrix} \mathbf{u}[n-1] \\ \mathbf{u}[n-2] \\ \vdots \\ \mathbf{u}[0] \end{bmatrix} \quad (6.58)$$

It follows from Theorem 3.1 that for any $\mathbf{x}[0]$ and $\mathbf{x}[n]$, an input sequence exists if and only if the controllability matrix has full row rank. This shows the equivalence of (1) and (3). The matrix $\mathbf{W}_{dc}[n-1]$ can be written as

$$\mathbf{W}_{dc}[n-1] = [\mathbf{B} \ \mathbf{AB} \ \cdots \ \mathbf{A}^{n-1}\mathbf{B}] \begin{bmatrix} \mathbf{B}' \\ \mathbf{B}'\mathbf{A}' \\ \vdots \\ \mathbf{B}'(\mathbf{A}')^{n-1} \end{bmatrix}$$

The equivalence of (2) and (3) then follows Theorem 3.8. Note that $\mathbf{W}_{dc}[m]$ is always positive semidefinite. If it is nonsingular or, equivalently, positive definite, then (6.53) is controllable. The proof of the equivalence of (3) and (4) is identical to the continuous-time case. Condition (5) follows Condition (2) and Theorem 5.D6. We see that establishing Theorem 6.D1 is considerably simpler than establishing Theorem 6.1.

There is one important difference between the continuous- and discrete-time cases. If a continuous-time state-space equation is controllable, the input can transfer any state to any other state in any nonzero time interval, no matter how small. If a discrete-time state-space equation is controllable, an input sequence of length n can transfer any state to any other state. If we compute the controllability index μ as defined

in (6.15), then the transfer can be achieved using an input sequence of length μ . If an input sequence is shorter than μ , it is not possible to transfer any state to any other state.

Definition 6.D2 The discrete-time state-space equation (6.53) or the pair (\mathbf{A}, \mathbf{C}) is said to be *observable* if for any unknown initial state $\mathbf{x}[0]$, there exists a finite integer $k_1 > 0$ such that the knowledge of the input sequence $\mathbf{u}[k]$ and output sequence $\mathbf{y}[k]$ from $k = 0$ to k_1 suffices to determine uniquely the initial state $\mathbf{x}[0]$. Otherwise, the equation is said to be *unobservable*.

► Theorem 6.D01

The following statements are equivalent.

1. The n -dimensional pair (\mathbf{A}, \mathbf{C}) is observable.
2. The $n \times n$ matrix

$$\mathbf{W}_{do}[n-1] = \sum_{m=0}^{n-1} (\mathbf{A}')^m \mathbf{C}' \mathbf{C} \mathbf{A}^m \quad (6.59)$$

is nonsingular or, equivalently, positive definite.

3. The $nq \times n$ *observability matrix*

$$\mathcal{O}_d = \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A} \\ \vdots \\ \mathbf{C}\mathbf{A}^{n-1} \end{bmatrix} \quad (6.60)$$

has rank n (full column rank). The matrix can be generated by calling `obsv` in MATLAB.

4. The $(n+q) \times n$ matrix $\begin{bmatrix} \mathbf{A} - \lambda \mathbf{I} \\ \mathbf{B} \end{bmatrix}$ has full column rank at every eigenvalue, λ , of \mathbf{A} .
5. If, in addition, all eigenvalues of \mathbf{A} have magnitudes less than 1, then the unique solution of

$$\mathbf{W}_{do} - \mathbf{A}' \mathbf{W}_{do} \mathbf{A} = \mathbf{C}' \mathbf{C} \quad (6.61)$$

is positive definite. The solution is called the discrete *observability Gramian* and can be expressed as

$$\mathbf{W}_{do} = \sum_{m=0}^{\infty} (\mathbf{A}')^m \mathbf{C}' \mathbf{C} \mathbf{A}^m \quad (6.62)$$

This can be proved directly or indirectly using the duality theorem. We mention that all other properties, such as controllability, and observability indices, Kalman decomposition, Jordan-form controllability, and observability conditions, discussed for the continuous-time case apply to the discrete-time case without any modification. The controllability index and observability index, however, have simple interpretation in the discrete-time case. The controllability index is the shortest input sequence that can transfer any state to any other state. The observability index is the shortest input and output sequences needed to determine the initial state uniquely.

6.6.1 Controllability to the Origin and Reachability

In the literature, there are three different controllability definitions:

1. Transfer any state to any other state as adopted in Definition 6.D1.
2. Transfer any state to the zero state, called controllability to the origin.
3. Transfer the zero state to any state, called controllability from the origin or, more often, *reachability*.

In the continuous-time case, because $e^{\mathbf{A}t}$ is nonsingular, the three definitions are equivalent. In the discrete-time case, if \mathbf{A} is nonsingular, the three definitions are again equivalent. But if \mathbf{A} is singular, then (1) and (3) are equivalent, but not (2) and (3). The equivalence of (1) and (3) can be easily seen from (6.58). We use examples to discuss the difference between (2) and (3). Consider

$$\mathbf{x}[k+1] = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{x}[k] + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} u[k] \quad (6.63)$$

Its controllability matrix has rank 0 and the equation is not controllable as defined in (1) or not reachable as defined in (3). The matrix \mathbf{A} has the form shown in (3.40) and has the property $\mathbf{A}^k = \mathbf{0}$ for $k \geq 3$. Thus we have

$$\mathbf{x}[3] = \mathbf{A}^3 \mathbf{x}[0] = \mathbf{0}$$

for any initial state $\mathbf{x}[0]$. Thus every state propagates to the zero state whether or not an input sequence is applied. Thus the equation is controllable to the origin. A different example follows. Consider

$$\mathbf{x}[k+1] = \begin{bmatrix} 2 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x}[k] + \begin{bmatrix} -1 \\ 0 \end{bmatrix} u[k] \quad (6.64)$$

Its controllability matrix

$$\begin{bmatrix} -1 & -2 \\ 0 & 0 \end{bmatrix}$$

has rank 1 and the equation is not reachable. However, for any $x_1[0] = \alpha$ and $x_2[0] = \beta$, the input $u[0] = 2\alpha + \beta$ transfers $\mathbf{x}[0]$ to $\mathbf{x}[1] = \mathbf{0}$. Thus the equation is controllable to the origin. Note that the A-matrices in (6.63) and (6.64) are both singular. The definition adopted in Definition 6.D1 encompasses the other two definitions and make the discussion simple. For a through discussion of the three definitions, see Reference 4.

6.7 Controllability after Sampling

Consider a continuous-time state equation

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \quad (6.65)$$

If the input is piecewise constant or

$$u[k] := u(kT) = u(t) \quad \text{for } kT \leq t < (k+1)T$$

then the equation can be described, as developed in (4.17), by

$$\bar{\mathbf{x}}[k+1] = \bar{\mathbf{A}}\bar{\mathbf{x}}[k] + \bar{\mathbf{B}}\mathbf{u}[k] \quad (6.66)$$

with

$$\bar{\mathbf{A}} = e^{\mathbf{AT}}, \quad \bar{\mathbf{B}} = \left(\int_0^T e^{\mathbf{At}} dt \right) \mathbf{B} =: \mathbf{MB} \quad (6.67)$$

The question is: If (6.65) is controllable, will its sampled equation in (6.66) be controllable? This problem is important in designing so-called dead-beat sampled-data systems and in computer control of continuous-time systems. The answer to the question depends on the sampling period T and the location of the eigenvalues of \mathbf{A} . Let λ_i and $\bar{\lambda}_i$ be, respectively, the eigenvalues of \mathbf{A} and $\bar{\mathbf{A}}$. We use Re and Im to denote the real part and imaginary part. Then we have the following theorem.

► Theorem 6.9

Suppose (6.65) is controllable. A sufficient condition for its discretized equation in (6.66), with sampling period T , to be controllable is that $|\text{Im}[\lambda_i - \lambda_j]| \neq 2\pi m/T$ for $m = 1, 2, \dots$, whenever $\text{Re}[\lambda_i - \lambda_j] = 0$. For the single-input case, the condition is necessary as well.

First we remark on the conditions. If \mathbf{A} has only real eigenvalues, then the discretized equation with any sampling period $T > 0$ is always controllable. Suppose \mathbf{A} has complex conjugate eigenvalues $\alpha \pm j\beta$. If the sampling period T does not equal any integer multiple of π/β , then the discretized state equation is controllable. If $T = m\pi/\beta$ for some integer m , then the discretized equation *may not* be controllable. The reason is as follows: Because $\bar{\mathbf{A}} = e^{\mathbf{AT}}$, if λ_i is an eigenvalue of \mathbf{A} , then $\bar{\lambda}_i := e^{\lambda_i T}$ is an eigenvalue of $\bar{\mathbf{A}}$ (Problem 3.24). If $T = m\pi/\beta$, the two distinct eigenvalues $\lambda_1 = \alpha + j\beta$ and $\lambda_2 = \alpha - j\beta$ of \mathbf{A} become a repeated eigenvalue $-e^{\alpha T}$ or $e^{\alpha T}$ of $\bar{\mathbf{A}}$. This will cause the discretized equation to be uncontrollable, as we will see in the proof. We show Theorem 6.9 by assuming \mathbf{A} to be in Jordan form. This is permitted because controllability is invariant under any equivalence transformation.



Proof: Proof of Theorem 6.9 To simplify discussion, we assume \mathbf{A} to be of the form:

$$\mathbf{A} = \text{diag}(\mathbf{A}_{11}, \mathbf{A}_{12}, \mathbf{A}_{21}) = \begin{bmatrix} \lambda_1 & 1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 & 1 \\ 0 & 0 & 0 & 0 & 0 & \lambda_2 \end{bmatrix} \quad (6.68)$$

In other words, \mathbf{A} has two distinct eigenvalues λ_1 and λ_2 . There are two Jordan blocks, one with order 3 and one with order 1, associated with λ_1 and only one Jordan block of order 2 associated with λ_2 . Using (3.48), we have

$$\begin{aligned} \bar{\mathbf{A}} &= \text{diag}(\bar{\mathbf{A}}_{11}, \bar{\mathbf{A}}_{12}, \bar{\mathbf{A}}_{21}) \\ &= \begin{bmatrix} e^{\lambda_1 T} & Te^{\lambda_1 T} & T^2 e^{\lambda_1 T}/2 & 0 & 0 & 0 \\ 0 & e^{\lambda_1 T} & Te^{\lambda_1 T} & 0 & 0 & 0 \\ 0 & 0 & e^{\lambda_1 T} & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{\lambda_1 T} & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{\lambda_2 T} & Te^{\lambda_2 T} \\ 0 & 0 & 0 & 0 & 0 & e^{\lambda_2 T} \end{bmatrix} \end{aligned} \quad (6.69)$$

This is not in Jordan form. Because we will use Theorem 6.8, which is also applicable to the discrete-time case without any modification, to prove Theorem 6.9, we must transform $\bar{\mathbf{A}}$ in (6.69) into Jordan form. It turns out that the Jordan form of $\bar{\mathbf{A}}$ equals the one in (6.68) if λ_i is replaced by $\bar{\lambda}_i := e^{\lambda_i T}$ (Problem 3.22). In other words, there exists a nonsingular upper triangular \mathbf{P} such that the transformation $\tilde{\mathbf{x}} = \mathbf{P}\bar{\mathbf{x}}$ will transform (6.66) into

$$\tilde{\mathbf{x}}[k+1] = \mathbf{P}\bar{\mathbf{A}}\mathbf{P}^{-1}\tilde{\mathbf{x}}[k] + \mathbf{PMBu}[k] \quad (6.70)$$

with $\mathbf{P}\bar{\mathbf{A}}\mathbf{P}^{-1}$ in the Jordan form in (6.68) with λ_i replaced by $\bar{\lambda}_i$. Now we are ready to establish Theorem 6.9.

First we show that \mathbf{M} in (6.67) is nonsingular. If \mathbf{A} is of the form shown in (6.68), then \mathbf{M} is upper triangular. Its diagonal entry is of form

$$m_{ii} := \int_0^T e^{\lambda_i \tau} d\tau = \begin{cases} (e^{\lambda_i T} - 1)/\lambda_i & \text{if } \lambda_i \neq 0 \\ T & \text{if } \lambda_i = 0 \end{cases} \quad (6.71)$$

Let $\lambda_i = \alpha_i + j\beta_i$ with $\beta_i > 0$. The only way for $m_{ii} = 0$ is $\alpha_i = 0$ and $\beta_i T = 2\pi\bar{m}$, for some positive integer \bar{m} . The theorem however requires that $2\beta_i T \neq 2\pi m$, for all positive integers m . Thus we conclude, under the condition of the theorem, that $m_{ii} \neq 0$ and \mathbf{M} is nonsingular.

If \mathbf{A} is of the form shown in Fig. (6.68), then it is controllable if and only if the third and fourth rows of \mathbf{B} are linearly independent and the last row of \mathbf{B} is nonzero (Theorem 6.8). Under the condition in Theorem 6.9, the two eigenvalues $\bar{\lambda}_1 = e^{\lambda_1 T}$ and $\bar{\lambda}_2 = e^{\lambda_2 T}$ of $\bar{\mathbf{A}}$ are distinct. Thus (6.70) is controllable if and only if the third and fourth rows of \mathbf{PMB} are linearly independent and the last row of \mathbf{PMB}

is nonzero. This is so because those rows of \mathbf{PMB} equal the corresponding rows of \mathbf{B} multiplied by nonzero constants. This shows the sufficiency of the theorem. If the condition in Theorem 6.9 is not met, then $\bar{\lambda}_1 = \bar{\lambda}_2$. In this case, (6.70) is controllable if the third, fourth and last rows of \mathbf{PMB} are linearly independent. This is still possible if \mathbf{B} has three or more columns. Thus the condition is not necessary. In the single-input case, if $\bar{\lambda}_1 = \bar{\lambda}_2$, then (6.70) has two or more Jordan blocks associated with the same eigenvalue and (6.70) is, following Corollary 6.8, not controllable. This establishes the theorem. Q.E.D.

In the proof of Theorem 6.9, we have essentially established the theorem that follows.

► **Theorem 6.10**

If a continuous-time linear time-invariant state equation is not controllable, then its discretized state equation, with any sampling period, is not controllable.

This theorem is intuitively obvious. If a state equation is not controllable by using any input, it is certainly not controllable by using any piecewise constant input.

Example 6.7.1 Consider the system shown in Fig. 6.11. Its input is sampled every T seconds and then kept constant using a zero-order hold circuit. The transfer function of the system is given as

$$\hat{g}(s) = \frac{s+2}{s^3 + 3s^2 + 7s + 5} = \frac{s+2}{(s+1)(s+1+j2)(s+1-j2)} \quad (6.72)$$

Using (4.33), we can readily obtain the following state-space equation

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \begin{bmatrix} -3 & -7 & -5 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u(t) \\ y(t) &= [0 \ 1 \ 2] \mathbf{x}(t) \end{aligned} \quad (6.73)$$

to describe the system. It is a controllable-form realization and is clearly controllable. The eigenvalues of \mathbf{A} are $-1, -1 \pm j2$ and are plotted in Fig. 6.11. The three eigenvalues

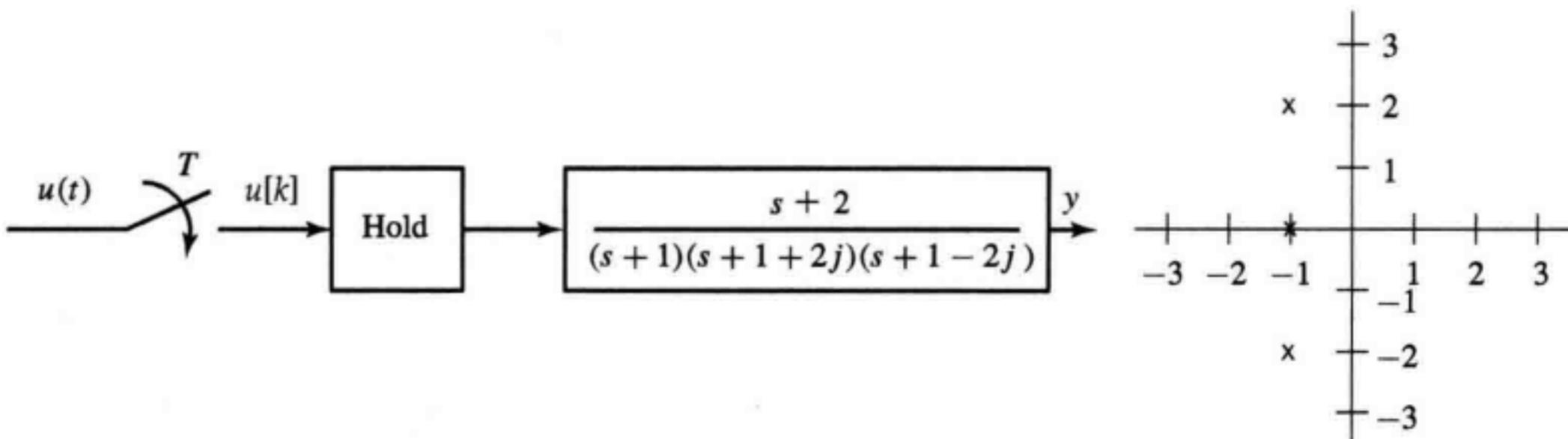


FIGURE 6.11 System with piecewise constant input.

have the same real part; their differences in imaginary parts are 2 and 4. Thus the discretized state equation is controllable if and only if

$$T \neq \frac{2\pi m}{2} = \pi m \quad \text{and} \quad T \neq \frac{2\pi m}{4} = 0.5\pi m$$

for $m = 1, 2, \dots$. The second condition includes the first condition. Thus we conclude that the discretized equation of (6.73) is controllable if and only if $T \neq 0.5m\pi$ for any positive integer m .

We use MATLAB to check the result for $m = 1$ or $T = 0.5\pi$. Typing

```
a=[-3 -7 -5;1 0 0;0 1 0];b=[1;0;0];
[ad,bd]=c2d(a,b,pi/2)
```

yields the discretized state equation as

$$\bar{\mathbf{x}}[k+1] = \begin{bmatrix} -0.1039 & 0.2079 & 0.5197 \\ -0.1039 & -0.4158 & -0.5197 \\ 0.1039 & 0.2079 & 0.3118 \end{bmatrix} \bar{\mathbf{x}}[k] + \begin{bmatrix} -0.1039 \\ 0.1039 \\ 0.1376 \end{bmatrix} u[k] \quad (6.74)$$

Its controllability matrix can be obtained by typing `ctrb(ad, bd)`, which yields

$$\mathcal{C}_d = \begin{bmatrix} -0.1039 & 0.1039 & -0.0045 \\ 0.1039 & -0.1039 & 0.0045 \\ 0.1376 & 0.0539 & 0.0059 \end{bmatrix}$$

Its first two rows are clearly linearly dependent. Thus \mathcal{C}_d does not have full row rank, and (6.74) is not controllable as predicted by Theorem 6.9. Typing `rank(ctrb(ad, bd))` also yields 2. Thus the discretized state equation is not controllable.

What has been discussed is also applicable to the observability part. In other words, under the conditions in Theorem 6.9, if a continuous-time state-space equation is observable, its discretized equation is also observable.

6.8 LTV State-Space Equations

Consider the n -dimensional p -input q -output state-space equation

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}(t)\mathbf{x}(t) \end{aligned} \quad (6.75)$$

The state-space equation is said to be controllable at t_0 , if there exists a finite $t_1 > t_0$ such that for any $\mathbf{x}(t_0) = \mathbf{x}_0$ and any \mathbf{x}_1 , there exists an input that transfers \mathbf{x}_0 to \mathbf{x}_1 at time t_1 . Otherwise the state-space equation is uncontrollable at t_0 . In the time-invariant case, if a state equation is controllable, then it is controllable at every t_0 and for every $t_1 > t_0$; therefore, there is no need to specify t_0 and t_1 . In the time-varying case, the specification of t_0 and t_1 is crucial.

► **Theorem 6.11**

The n -dimensional pair $(\mathbf{A}(t), \mathbf{B}(t))$ is controllable at time t_0 if and only if there exists a finite $t_1 > t_0$ such that the $n \times n$ matrix

$$\mathbf{W}_c(t_0, t_1) = \int_{t_0}^{t_1} \Phi(t_1, \tau) \mathbf{B}(\tau) \mathbf{B}'(\tau) \Phi'(t_1, \tau) d\tau \quad (6.76)$$

where $\Phi(t, \tau)$ is the state transition matrix of $\dot{\mathbf{x}} = \mathbf{A}(t)\mathbf{x}$, is nonsingular.

⇒ **Proof:** We first show that if $\mathbf{W}_c(t_0, t_1)$ is nonsingular, then (6.75) is controllable. The response of (6.75) at t_1 was computed in (4.65) as

$$\mathbf{x}(t_1) = \Phi(t_1, t_0) \mathbf{x}_0 + \int_{t_0}^{t_1} \Phi(t_1, \tau) \mathbf{B}(\tau) \mathbf{u}(\tau) d\tau \quad (6.77)$$

We claim that the input

$$\mathbf{u}(t) = -\mathbf{B}'(t) \Phi'(t_1, t) \mathbf{W}_c^{-1}(t_0, t_1) [\Phi(t_1, t_0) \mathbf{x}_0 - \mathbf{x}_1] \quad (6.78)$$

will transfer \mathbf{x}_0 at time t_0 to \mathbf{x}_1 at time t_1 . Indeed, substituting (6.78) into (6.77) yields

$$\begin{aligned} \mathbf{x}(t_1) &= \Phi(t_1, t_0) \mathbf{x}_0 - \int_{t_0}^{t_1} \Phi(t_1, \tau) \mathbf{B}(\tau) \mathbf{B}'(\tau) \Phi'(t_1, \tau) d\tau \\ &\quad \cdot \mathbf{W}_c^{-1}(t_0, t_1) [\Phi(t_1, t_0) \mathbf{x}_0 - \mathbf{x}_1] \\ &= \Phi(t_1, t_0) \mathbf{x}_0 - \mathbf{W}_c(t_0, t_1) \mathbf{W}_c^{-1}(t_0, t_1) [\Phi(t_1, t_0) \mathbf{x}_0 - \mathbf{x}_1] = \mathbf{x}_1 \end{aligned}$$

Thus the equation is controllable at t_0 . We show the converse by contradiction. Suppose (6.75) is controllable at t_0 but $\mathbf{W}_c(t_0, t)$ is singular or, positive semidefinite, for all $t_1 > t_0$. Then there exists an $n \times 1$ nonzero constant vector \mathbf{v} such that

$$\begin{aligned} \mathbf{v}' \mathbf{W}_c(t_0, t_1) \mathbf{v} &= \int_{t_0}^{t_1} \mathbf{v}' \Phi(t_1, \tau) \mathbf{B}(\tau) \mathbf{B}'(\tau) \Phi'(t_1, \tau) \mathbf{v} d\tau \\ &= \int_{t_0}^{t_1} \|\mathbf{B}'(\tau) \Phi'(t_1, \tau) \mathbf{v}\|^2 d\tau = 0 \end{aligned}$$

which implies

$$\mathbf{B}'(\tau) \Phi'(t_1, \tau) \mathbf{v} \equiv \mathbf{0} \quad \text{or} \quad \mathbf{v}' \Phi(t_1, \tau) \mathbf{B}(\tau) \equiv \mathbf{0} \quad (6.79)$$

for all τ in $[t_0, t_1]$. If (6.75) is controllable, there exists an input that transfers the initial state $\mathbf{x}_0 = \Phi(t_0, t_1) \mathbf{v}$ at t_0 to $\mathbf{x}(t_1) = \mathbf{0}$. Then (6.77) becomes

$$\mathbf{0} = \Phi(t_1, t_0) \Phi(t_0, t_1) \mathbf{v} + \int_{t_0}^{t_1} \Phi(t_1, \tau) \mathbf{B}(\tau) \mathbf{u}(\tau) d\tau \quad (6.80)$$

Its premultiplication by \mathbf{v}' yields

$$0 = \mathbf{v}'\mathbf{v} + \mathbf{v}' \int_{t_0}^{t_1} \Phi(t_1, \tau) \mathbf{B}(\tau) \mathbf{u}(\tau) d\tau = ||\mathbf{v}||^2 + 0$$

This contradicts the hypothesis $\mathbf{v} \neq \mathbf{0}$. Thus if $(\mathbf{A}(t), \mathbf{B}(t))$ is controllable at t_0 , $\mathbf{W}_c(t_0, t_1)$ must be nonsingular for some finite $t_1 > t_0$. This establishes Theorem 6.12. **Q.E.D.**

In order to apply Theorem 6.11, we need the knowledge of the state transition matrix which, however, may not be available. Therefore, it is desirable to develop a controllability condition without involving $\Phi(t, \tau)$. This is possible if we have additional conditions on $\mathbf{A}(t)$ and $\mathbf{B}(t)$. Recall that we have assumed $\mathbf{A}(t)$ and $\mathbf{B}(t)$ to be continuous. Now we require them to be $(n - 1)$ times continuously differentiable. Define $\mathbf{M}_0(t) = \mathbf{B}(t)$. We then define recursively a sequence of $n \times p$ matrices $\mathbf{M}_m(t)$ as

$$\mathbf{M}_{m+1}(t) := -\mathbf{A}(t)\mathbf{M}_m(t) + \frac{d}{dt}\mathbf{M}_m(t) \quad (6.81)$$

for $m = 0, 1, \dots, n - 1$. Clearly, we have

$$\Phi(t_2, t)\mathbf{B}(t) = \Phi(t_2, t)\mathbf{M}_0(t)$$

for any fixed t_2 . Using

$$\frac{\partial}{\partial t}\Phi(t_2, t) = -\Phi(t_2, t)\mathbf{A}(t)$$

(Problem 4.26), we compute

$$\begin{aligned} \frac{\partial}{\partial t}[\Phi(t_2, t)\mathbf{B}(t)] &= \frac{\partial}{\partial t}[\Phi(t_2, t)]\mathbf{B}(t) + \Phi(t_2, t)\frac{d}{dt}\mathbf{B}(t) \\ &= \Phi(t_2, t)[-A(t)\mathbf{M}_0(t) + \frac{d}{dt}\mathbf{M}_0(t)] = \Phi(t_2, t)\mathbf{M}_1(t) \end{aligned}$$

Proceeding forward, we have

$$\frac{\partial^m}{\partial t^m}\Phi(t_2, t)\mathbf{B}(t) = \Phi(t_2, t)\mathbf{M}_m(t) \quad (6.82)$$

for $m = 0, 1, 2, \dots$. The following theorem is sufficient but not necessary for (6.75) to be controllable.

► Theorem 6.12

Let $\mathbf{A}(t)$ and $\mathbf{B}(t)$ be $(n - 1)$ times continuously differentiable. Then the n -dimensional pair $(\mathbf{A}(t), \mathbf{B}(t))$ is controllable at t_0 if there exists a finite $t_1 > t_0$ such that

$$\text{rank}[\mathbf{M}_0(t_1) \mathbf{M}_1(t_1) \cdots \mathbf{M}_{n-1}(t_1)] = n \quad (6.83)$$



Proof: We show that if (6.83) holds, then $\mathbf{W}_c(t_0, t)$ is nonsingular for all $t \geq t_1$. Suppose not, that is, $\mathbf{W}_c(t_0, t)$ is singular or positive semidefinite for some $t_2 \geq t_1$. Then there exists an $n \times 1$ nonzero constant vector \mathbf{v} such that

$$\begin{aligned}\mathbf{v}' \mathbf{W}_c(t_0, t_2) \mathbf{v} &= \int_{t_0}^{t_2} \mathbf{v}' \Phi(t_2, \tau) \mathbf{B}(\tau) \mathbf{B}'(\tau) \Phi'(t_2, \tau) \mathbf{v} d\tau \\ &= \int_{t_0}^{t_2} \|\mathbf{B}'(\tau) \Phi'(t_2, \tau) \mathbf{v}\|^2 d\tau = 0\end{aligned}$$

which implies

$$\mathbf{B}'(\tau) \Phi'(t_2, \tau) \mathbf{v} \equiv \mathbf{0} \quad \text{or} \quad \mathbf{v}' \Phi(t_2, \tau) \mathbf{B}(\tau) \equiv \mathbf{0} \quad (6.84)$$

for all τ in $[t_0, t_2]$. Its differentiations with respect to τ yield, as derived in (6.82),

$$\mathbf{v}' \Phi(t_2, \tau) \mathbf{M}_m(\tau) \equiv \mathbf{0}$$

for $m = 0, 1, 2, \dots, n-1$, and all τ in $[t_0, t_2]$, in particular, at t_1 . They can be arranged as

$$\mathbf{v}' \Phi(t_2, t_1) [\mathbf{M}_0(t_1) \ \mathbf{M}_1(t_1) \ \cdots \ \mathbf{M}_{n-1}(t_1)] = \mathbf{0} \quad (6.85)$$

Because $\Phi(t_2, t_1)$ is nonsingular, $\mathbf{v}' \Phi(t_2, t_1)$ is nonzero. Thus (6.85) contradicts (6.83). Therefore, under the condition in (6.83), $\mathbf{W}_c(t_0, t_2)$, for any $t_2 \geq t_1$, is nonsingular and $(\mathbf{A}(t), \mathbf{B}(t))$ is, following Theorem 6.11, controllable at t_0 . Q.E.D.

Example 6.8.1 Consider

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} t & -1 & 0 \\ 0 & -t & t \\ 0 & 0 & t \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} u(t) \quad (6.86)$$

We have $\mathbf{M}_0(t) = [0 \ 1 \ 1]'$ and compute

$$\begin{aligned}\mathbf{M}_1(t) &= -\mathbf{A}(t)\mathbf{M}_0(t) + \frac{d}{dt}\mathbf{M}_0(t) = \begin{bmatrix} 1 \\ 0 \\ -t \end{bmatrix} \\ \mathbf{M}_2(t) &= -\mathbf{A}(t)\mathbf{M}_1(t) + \frac{d}{dt}\mathbf{M}_1(t) = \begin{bmatrix} -t \\ t^2 \\ t^2 - 1 \end{bmatrix}\end{aligned}$$

The determinant of the matrix

$$[\mathbf{M}_0(t) \ \mathbf{M}_1(t) \ \mathbf{M}_2(t)] = \begin{bmatrix} 0 & 1 & -t \\ 1 & 0 & t^2 \\ 1 & -t & t^2 - 1 \end{bmatrix}$$

is $t^2 + 1$ which is nonzero for all t . Thus the state equation in (6.86) is controllable at every t .

Example 6.8.2 Consider

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u(t) \quad (6.87)$$

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} e^t \\ e^{2t} \end{bmatrix} u(t) \quad (6.88)$$

Equation (6.87) is a time-invariant equation and is controllable according to Corollary 6.8. Equation (6.88) is a time-varying equation; the two entries of its B-matrix are nonzero for all t and one might be tempted to conclude that (6.88) is controllable. Let us check this by using Theorem 6.11. Its state transition matrix is

$$\Phi(t, \tau) = \begin{bmatrix} e^{t-\tau} & 0 \\ 0 & e^{2(t-\tau)} \end{bmatrix}$$

and

$$\Phi(t, \tau) \mathbf{B}(\tau) = \begin{bmatrix} e^{t-\tau} & 0 \\ 0 & e^{2(t-\tau)} \end{bmatrix} \begin{bmatrix} e^\tau \\ e^{2\tau} \end{bmatrix} = \begin{bmatrix} e^t \\ e^{2t} \end{bmatrix}$$

We compute

$$\begin{aligned} \mathbf{W}_c(t_0, t) &= \int_{t_0}^t \begin{bmatrix} e^\tau \\ e^{2\tau} \end{bmatrix} [e^t \ e^{2t}] d\tau = \begin{bmatrix} \int_{t_0}^t e^{2\tau} d\tau & \int_{t_0}^t e^{3\tau} d\tau \\ \int_{t_0}^t e^{3\tau} d\tau & \int_{t_0}^t e^{4\tau} d\tau \end{bmatrix} \\ &= \begin{bmatrix} e^{2t}(t - t_0) & e^{3t}(t - t_0) \\ e^{3t}(t - t_0) & e^{4t}(t - t_0) \end{bmatrix} \end{aligned}$$

Its determinant is identically zero for all t_0 and t . Thus (6.88) is not controllable at any t_0 . From this example, we see that, in applying a theorem, every condition should be carefully checked; otherwise, we might obtain an erroneous conclusion.

We now discuss the observability part. The linear time-varying state-space equation in (6.75) is observable at t_0 if there exists a finite t_1 such that for any state $\mathbf{x}(t_0) = \mathbf{x}_0$, the knowledge of the input and output over the time interval $[t_0, t_1]$ suffices to determine uniquely the initial state \mathbf{x}_0 . Otherwise, the state-space equation is said to be unobservable at t_0 .

Theorem 6.011

The pair $(\mathbf{A}(t), \mathbf{C}(t))$ is observable at time t_0 if and only if there exists a finite $t_1 > t_0$ such that the $n \times n$ matrix

$$\mathbf{W}_o(t_0, t_1) = \int_{t_0}^{t_1} \Phi'(\tau, t_0) \mathbf{C}'(\tau) \mathbf{C}(\tau) \Phi(\tau, t_0) d\tau \quad (6.89)$$

where $\Phi(t, \tau)$ is the state transition matrix of $\dot{\mathbf{x}} = \mathbf{A}(t)\mathbf{x}$, is nonsingular.

► **Theorem 6.O12**

Let $\mathbf{A}(t)$ and $\mathbf{C}(t)$ be $(n - 1)$ times continuously differentiable. Then the n -dimensional pair $(\mathbf{A}(t), \mathbf{C}(t))$ is observable at t_0 if there exists a finite $t_1 > t_0$ such that

$$\text{rank} \begin{bmatrix} \mathbf{N}_0(t_1) \\ \mathbf{N}_1(t_1) \\ \vdots \\ \mathbf{N}_{n-1}(t_1) \end{bmatrix} = n \quad (6.90)$$

where

$$\mathbf{N}_{m+1}(t) = \mathbf{N}_m(t)\mathbf{A}(t) + \frac{d}{dt}\mathbf{N}_m(t), \quad m = 0, 1, \dots, n-1$$

with

$$\mathbf{N}_0 = \mathbf{C}(t)$$

We mention that the duality theorem in Theorem 6.5 for time-invariant systems is not applicable to time-varying systems. It must be modified. See Problems 6.28 and 6.29.

Problems

6.1 Is the state-space equation

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \begin{bmatrix} -2 & 3 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u(t) \\ y(t) &= [0 \quad 1 \quad 3] \mathbf{x}(t) \end{aligned}$$

controllable? Observable?

6.2 Is the state-space equation

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 2 & -1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 & 2 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} u(t) \\ y(t) &= [1 \quad 0 \quad -2] \mathbf{x}(t) \end{aligned}$$

controllable? Observable?

6.3 Is it true that the rank of $[\mathbf{B} \ \mathbf{AB} \ \dots \ \mathbf{A}^{n-1}\mathbf{B}]$ equals the rank of $[\mathbf{AB} \ \mathbf{A}^2\mathbf{B} \ \dots \ \mathbf{A}^n\mathbf{B}]$? If not, under what condition will it be true?

6.4 Show that if the state equation

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{0} \end{bmatrix} u(t)$$

is controllable, then the pair $(\mathbf{A}_{22}, \mathbf{A}_{21})$ is controllable.

- 6.5** Find a state-space equation to describe the circuit shown in Fig. 6.1, and then check its controllability and observability.
- 6.6** Find the controllability index and observability index of the state-space equations in Problems 6.1 and 6.2.
- 6.7** What is the controllability index of the state equation

$$\dot{\mathbf{x}}(t) = \mathbf{Ax}(t) + \mathbf{Iu}(t)$$

where \mathbf{I} is the unit matrix?

- 6.8** Reduce the state-space equation

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -1 & 4 \\ 4 & -1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u(t), \quad y(t) = [1 \ 1] \mathbf{x}(t)$$

to a controllable one. Is the reduced equation observable?

- 6.9** Reduce the state-space equation in Problem 6.5 to a controllable and observable equation.
- 6.10** Reduce the state-space equation

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} \lambda_1 & 1 & 0 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 & 1 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} u(t)$$

$$y(t) = [0 \ 1 \ 1 \ 0 \ 1] \mathbf{x}(t)$$

to a controllable and observable equation.

- 6.11** Consider the n -dimensional state-space equation

$$\dot{\mathbf{x}}(t) = \mathbf{Ax}(t) + \mathbf{Bu}(t)$$

$$\mathbf{y}(t) = \mathbf{Cx}(t) + \mathbf{Du}(t)$$

The rank of its controllability matrix is assumed to be $n_1 < n$. Let \mathbf{Q}_1 be an $n \times n_1$ matrix whose columns are any n_1 linearly independent columns of the controllability matrix. Let \mathbf{P}_1 be an $n_1 \times n$ matrix such that $\mathbf{P}_1 \mathbf{Q}_1 = \mathbf{I}_{n_1}$, where \mathbf{I}_{n_1} is the unit matrix of order n_1 . Show that the following n_1 -dimensional state-space equation

$$\dot{\bar{\mathbf{x}}}_1(t) = \mathbf{P}_1 \mathbf{A} \mathbf{Q}_1 \bar{\mathbf{x}}_1(t) + \mathbf{P}_1 \mathbf{B} \mathbf{u}(t)$$

$$\bar{\mathbf{y}}(t) = \mathbf{C} \mathbf{Q}_1 \bar{\mathbf{x}}_1(t) + \mathbf{D} \mathbf{u}(t)$$

is controllable and has the same transfer matrix as the original state-space equation.

- 6.12** In Problem 6.11, the reduction procedure reduces to solving for \mathbf{P}_1 in $\mathbf{P}_1 \mathbf{Q}_1 = \mathbf{I}$. How do you solve \mathbf{P}_1 ?
- 6.13** Develop a similar statement as in Problem 6.11 for an unobservable state-space equation.

6.14 Is the Jordan-form state-space equation controllable and observable?

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 2 & 1 & 1 \\ 2 & 1 & 1 \\ 1 & 1 & -1 \\ 3 & 2 & 1 \\ -1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & 2 \end{bmatrix} \mathbf{u}(t)$$

$$\mathbf{y}(t) = \begin{bmatrix} 2 & 2 & -1 & 3 & -1 & -1 & 1 \\ 1 & 3 & -1 & 2 & 0 & 0 & 0 \\ 0 & -4 & -1 & 1 & 1 & 1 & 0 \end{bmatrix} \mathbf{x}(t)$$

6.15 Is it possible to find a set of b_{ij} and a set of c_{ij} such that the state-space equation

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \\ b_{41} & b_{42} \\ b_{51} & b_{52} \end{bmatrix} \mathbf{u}(t)$$

$$\mathbf{y}(t) = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} \\ c_{21} & c_{22} & c_{23} & c_{24} & c_{25} \\ c_{31} & c_{32} & c_{33} & c_{34} & c_{35} \end{bmatrix} \mathbf{x}(t)$$

is controllable? Observable?

6.16 Consider the state-space equation

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \alpha_1 & \beta_1 & 0 & 0 \\ 0 & -\beta_1 & \alpha_1 & 0 & 0 \\ 0 & 0 & 0 & \alpha_2 & \beta_2 \\ 0 & 0 & 0 & -\beta_2 & \alpha_2 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} b_1 \\ b_{11} \\ b_{12} \\ b_{21} \\ b_{22} \end{bmatrix} \mathbf{u}(t)$$

$$\mathbf{y}(t) = [c_1 \ c_{11} \ c_{12} \ c_{21} \ c_{22}] \mathbf{x}(t)$$

It is the modal form discussed in (4.28). It has one real eigenvalue and two pairs of complex conjugate eigenvalues. It is assumed that they are distinct. Show that the state-space equation is controllable if and only if $b_1 \neq 0$; $b_{i1} \neq 0$ or $b_{i2} \neq 0$ for $i = 1, 2$. It is Observable if and only if $c_1 \neq 0$; $c_{i1} \neq 0$ or $c_{i2} \neq 0$ for $i = 1, 2$.

6.17 Check the controllability and observability of the Problem 5.1.

6.18 Investigate the controllability and observability of a system consisting of the cascade connection of $\hat{g}_1(s) = (s+a)/(s+b)$ and $\hat{g}_2(s) = 1/(s+c)$. Furthermore, suppose the cascaded system is connected in parallel with a third transfer function $\hat{g}_3(s) = 1/(s+d)$. Investigate the controllability and observability of the overall system in this case. Note that the parameters $\{a, b, c, d\}$ are all assumed to be non-zero.

6.19 Let a system be represented by the triple $\{A, B, C\}$ and suppose the pair $\{A, B\}$ is controllable. Find a transformation matrix P to transform the system to the input identifiable form,

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} I \\ 0 \end{bmatrix} \mathbf{u}(t)$$

Show that $\{A, B\}$ is controllable if and only if $\{A_{22}, A_{21}\}$ is controllable. In a similar fashion perform the same task for observable pair $\{A, C\}$.

- 6.20** Show that $\{A, C\}$ is observable if and only if $\{A, C'C\}$ is observable.
- 6.21** Find a mechanical or electrical system to demonstrate the concept of controllability (uncontrollability) and observability (unobservability).
- 6.22** Consider the following systems,

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ e^{-t} & 0 & 1 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u(t)$$

$$y(t) = [0 \ 0 \ 1] x(t)$$

$$\dot{x}(t) = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -2 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} u(t)$$

$$y(t) = [1 \ 1 \ 1] x(t)$$

- (a) Check the controllability and observability of the above systems.
- (b) Is it possible to choose an initial condition state at $t = 0$ such that zero input response of the second system becomes $y(t) = t e^{-t}$? If yes, find it.
- 6.23** Find 2- and 3-dimensional state-space equations to describe the circuit shown in Fig. 6.12. Discuss their controllability and observability.

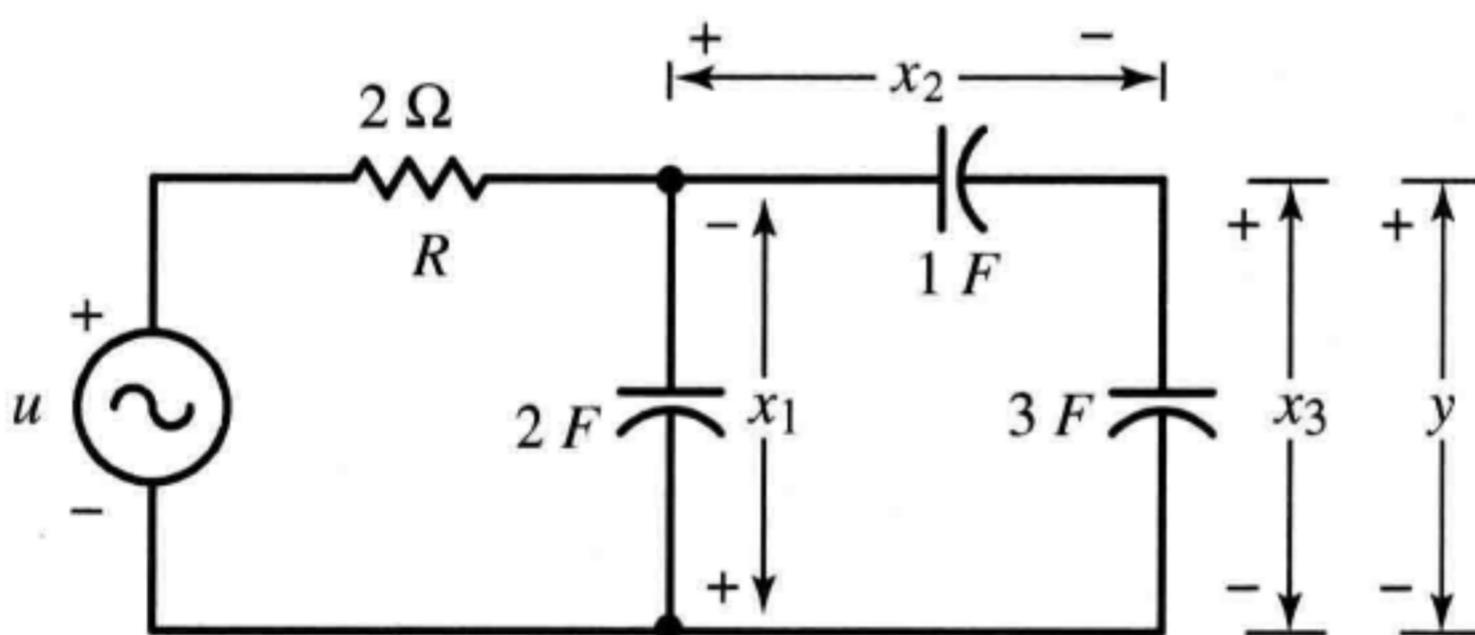


FIGURE 6.12

- 6.24** Develop a state-space equation to describe the circuit in Fig. 2.19 and then check its controllability and observability. Can you obtain the same conclusion directly from the circuit?
- 6.25** Consider the continuous-time state-space equation in Problem 4.2 and its discretized equations in Problem 4.4 with sampling period $T = 1$ and π . Discuss controllability and observability of the discretized equations.
- 6.26** Check controllability and observability of

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 \\ 0 & t \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t), \quad y(t) = [0 \ 1] \mathbf{x}(t)$$

6.27 Check controllability and observability of

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ e^{-t} \end{bmatrix} u(t), \quad y(t) = [0 \ e^{-t}] \mathbf{x}(t)$$

- 6.28** Show that $(\mathbf{A}(t), \mathbf{B}(t))$ is controllable at t_0 if and only if $(-\mathbf{A}'(t), \mathbf{B}'(t))$ is observable at t_0 .
- 6.29** For time-invariant systems, show that (\mathbf{A}, \mathbf{B}) is controllable if and only if $(-\mathbf{A}, \mathbf{B})$ is controllable. Is this true for time-varying systems?

Minimal Realizations and Coprime Fractions

7.1 Introduction

In the remaining three chapters we study only linear time-invariant systems. In this chapter we study further the realization problem discussed in Section 4.5. Recall that a transfer function $\hat{g}(s)$ is said to be realizable if there exists a state-space equation

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t) \\ y(t) &= \mathbf{c}\mathbf{x}(t) + du(t)\end{aligned}$$

that has $\hat{g}(s)$ as its transfer function. This is an important problem for the following reasons: First, many design methods and computational algorithms are developed for state-space equations. In order to apply these methods and algorithms, transfer functions must be realized into state-space equations. As an example, computing the response of a transfer function in MATLAB is achieved by first transforming the transfer function into a state-space equation. Second, once a transfer function is realized into a state-space equation, the transfer function can be implemented using op-amp circuits as discussed in Section 4.3.2. We showed in Theorem 4.2 that a transfer function is realizable if and only if it is a proper rational function.

If a transfer function is realizable, then it has infinitely many realizations, not necessarily of the same dimension, as shown in Subsections 4.5.1 and 7.2.1. An important question is then raised: What is the smallest possible dimension? Realizations with the smallest possible dimension are called *minimal-dimensional* or *minimal* realizations. If we use a minimal realization to implement a transfer function, then the number of integrators used in an op-amp circuit will be minimum. Thus minimal realizations are of practical importance.

Sections 7.2 through 7.5 study SISO systems and Section 7.6 onward study MIMO systems. In the first reading, it is suggested to study Sections 7.2 through 7.2.2 and then Sections 7.6 and 7.7. Section 7.2 and its subsections discuss the relationship between coprimeness in transfer functions and controllability and observability in state-space equations and their implications in practical systems. Section 7.6 extends the concept of degree from scalar transfer functions to transfer matrices. It shows the complexity in the latter case. Section 7.7 shows how to obtain a minimal realization, using the results in Chapter 6, from a nonminimal one.

Section 7.3 discusses a method of computing coprime fractions and is needed for carrying out the SISO system design in Chapter 9. Unlike the SISO case, the concept of coprimeness in MIMO systems is fairly complex and is discussed in Section 7.8 and its subsections. Its study may be skipped if one is interested in only the SISO part of Chapter 9. The rest of Chapter 7 (balanced and Markov-parameter realizations) will not be used in the remainder of this text.

7.2 Implications of Coprimeness

Consider a system with proper transfer function $\hat{g}(s)$. We decompose it as

$$\hat{g}(s) = \hat{g}(\infty) + \hat{g}_{sp}(s)$$

where $\hat{g}_{sp}(s)$ is strictly proper and $\hat{g}(\infty)$ is a constant. The constant $\hat{g}(\infty)$ yields the direct transmission part of every realization and will not play any role in what will be discussed. Therefore, we consider in this section only strictly proper rational functions. Consider

$$\hat{g}(s) = \frac{N(s)}{D(s)} = N(s)D^{-1}(s) = \frac{\beta_1 s^3 + \beta_2 s^2 + \beta_3 s + \beta_4}{s^4 + \alpha_1 s^3 + \alpha_2 s^2 + \alpha_3 s + \alpha_4} \quad (7.1)$$

To simplify the discussion, we have assumed that the denominator $D(s)$ has degree 4 and is monic (has 1 as its leading coefficient). In Section 4.5, we introduced for (7.1) the realization in (4.33) without any discussion of its state variables. Now we will redevelop (4.33) by first defining a set of state variables and then discuss the implications of the coprimeness of $D(s)$ and $N(s)$.

Consider $\hat{y}(s) = \hat{g}(s)\hat{u}(s)$ or

$$\hat{y}(s) = N(s)D^{-1}(s)\hat{u}(s) \quad (7.2)$$

Let us introduce a new variable $v(t)$ defined by $\hat{v}(s) = D^{-1}(s)\hat{u}(s)$. Then we have

$$D(s)\hat{v}(s) = \hat{u}(s) \quad (7.3)$$

$$\hat{y}(s) = N(s)\hat{v}(s) \quad (7.4)$$

Define state variables as

$$\mathbf{x}(t) := \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix} := \begin{bmatrix} v^{(3)}(t) \\ \ddot{v}(t) \\ \dot{v}(t) \\ v(t) \end{bmatrix} \quad \text{or} \quad \hat{\mathbf{x}}(s) = \begin{bmatrix} \hat{x}_1(s) \\ \hat{x}_2(s) \\ \hat{x}_3(s) \\ \hat{x}_4(s) \end{bmatrix} = \begin{bmatrix} s^3 \\ s^2 \\ s \\ 1 \end{bmatrix} \hat{v}(s) \quad (7.5)$$

Recall that if $\hat{v}(s) = \mathcal{L}[v(t)]$ and if $v(0) = \dot{v}(0) = \ddot{v}(0) = \dots = 0$, then $\mathcal{L}[\dot{v}(t)] = s\hat{v}(s)$, $\mathcal{L}[\ddot{v}(t)] = s^2\hat{v}(s)$, and so forth. In other words, differentiation in the time domain is equivalent to the multiplication by s in the transform domain. From (7.5), we have

$$\dot{x}_2(t) = x_1(t), \quad \dot{x}_3(t) = x_2(t), \quad \text{and} \quad \dot{x}_4(t) = x_3(t) \quad (7.6)$$

They are independent of (7.1) and follow directly from the definition in (7.5). In order to develop an equation for \dot{x}_1 , we substitute (7.5) into (7.3) or

$$(s^4 + \alpha_1 s^3 + \alpha_2 s^2 + \alpha_3 s + \alpha_4) \hat{v}(s) = \hat{u}(s)$$

to yield

$$s\hat{x}_1(s) = -\alpha_1 \hat{x}_1(s) - \alpha_2 \hat{x}_2(s) - \alpha_3 \hat{x}_3(s) - \alpha_4 \hat{x}_4(s) + \hat{u}(s)$$

which becomes, in the time domain,

$$\dot{x}_1(t) = [-\alpha_1 \ -\alpha_2 \ -\alpha_3 \ -\alpha_4] \mathbf{x}(t) + 1 \cdot u(t) \quad (7.7)$$

Substituting (7.5) into (7.4) yields

$$\begin{aligned} \hat{y}(s) &= (\beta_1 s^3 + \beta_2 s^2 + \beta_3 s + \beta_4) \hat{v}(s) \\ &= \beta_1 \hat{x}_1(s) + \beta_2 \hat{x}_2(s) + \beta_3 \hat{x}_3(s) + \beta_4 \hat{x}_4(s) \\ &= [\beta_1 \ \beta_2 \ \beta_3 \ \beta_4] \hat{\mathbf{x}}(s) \end{aligned}$$

which becomes, in the time domain,

$$y(t) = [\beta_4 \ \beta_3 \ \beta_2 \ \beta_1] \mathbf{x}(t) \quad (7.8)$$

Equations (7.6), (7.7) and (7.8) can be combined as

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t) = \begin{bmatrix} -\alpha_1 & -\alpha_2 & -\alpha_3 & -\alpha_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} u(t) \\ y(t) &= \mathbf{c}\mathbf{x}(t) = [\beta_1 \ \beta_2 \ \beta_3 \ \beta_4] \mathbf{x}(t) \end{aligned} \quad (7.9)$$

This is a realization of (7.1) and was developed in (4.33) by direct verification.

Before proceeding, we mention that if $N(s)$ in (7.1) is 1, then $y(t) = v(t)$ and the output $y(t)$ and its derivatives can be chosen as state variables. However, if $N(s)$ is a polynomial of degree 1 or higher and if we choose the output and its derivatives as state variables, then its realization will be of the form

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t) + d_1 \dot{u}(t) + d_2 \ddot{u}(t) + \dots \\ y(t) &= \mathbf{c}\mathbf{x}(t) + du(t) \end{aligned}$$

This equation requires differentiations of u and is not used. Therefore, in general, we cannot select the output y and its derivatives as state variables¹. We must define state variables by using $v(t)$. Thus $v(t)$ is called a *pseudo state*.

Now we check the controllability and observability of (7.9). Its controllability matrix can be readily computed as

$$\mathcal{C} = \begin{bmatrix} 1 & -\alpha_1 & \alpha_1^2 - \alpha_2 & -\alpha_1^3 + 2\alpha_1\alpha_2 - \alpha_3 \\ 0 & 1 & -\alpha_1 & \alpha_1^2 - \alpha_2 \\ 0 & 0 & 1 & -\alpha_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7.10)$$

Its determinant is 1 for any α_i . Thus the controllability matrix \mathcal{C} has full row rank and the state equation is always controllable. This is the reason that (7.9) is called a *controllable-form realization*.

Next we check its observability. It turns out that it depends on whether or not $N(s)$ and $D(s)$ are *coprime* or have no common roots as stated in the next theorem.

► Theorem 7.1

The controllable-form equation in (7.9) is observable if and only if $D(s)$ and $N(s)$ in (7.1) are coprime.



Proof: We first show that if (7.9) is observable, then $D(s)$ and $N(s)$ are coprime. We show this by contradiction. If $D(s)$ and $N(s)$ are not coprime, then there exists a λ_1 such that

$$N(\lambda_1) = \beta_1\lambda_1^3 + \beta_2\lambda_1^2 + \beta_3\lambda_1 + \beta_4 = 0 \quad (7.11)$$

$$D(\lambda_1) = \lambda_1^4 + \alpha_1\lambda_1^3 + \alpha_2\lambda_1^2 + \alpha_3\lambda_1 + \alpha_4 = 0 \quad (7.12)$$

Let us define $\mathbf{v} := [\lambda_1^3 \ \lambda_1^2 \ \lambda_1 \ 1]'$; it is a 4×1 nonzero vector. Then (7.11) can be written as $N(\lambda_1) = \mathbf{cv} = 0$, where \mathbf{c} is defined in (7.9). Using (7.12) and the shifting property of the companion form, we can readily verify

$$\mathbf{Av} = \begin{bmatrix} -\alpha_1 & -\alpha_2 & -\alpha_3 & -\alpha_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1^3 \\ \lambda_1^2 \\ \lambda_1 \\ 1 \end{bmatrix} = \begin{bmatrix} \lambda_1^4 \\ \lambda_1^3 \\ \lambda_1^2 \\ \lambda_1 \end{bmatrix} = \lambda_1 \mathbf{v} \quad (7.13)$$

Thus we have $\mathbf{A}^2 \mathbf{v} = \mathbf{A}(\mathbf{Av}) = \lambda_1 \mathbf{Av} = \lambda_1^2 \mathbf{v}$ and $\mathbf{A}^3 \mathbf{v} = \lambda_1^3 \mathbf{v}$. We compute, using $\mathbf{cv} = 0$,

$$\mathcal{O}\mathbf{v} = \begin{bmatrix} \mathbf{c} \\ \mathbf{c}\mathbf{A} \\ \mathbf{c}\mathbf{A}^2 \\ \mathbf{c}\mathbf{A}^3 \end{bmatrix} \mathbf{v} = \begin{bmatrix} \mathbf{cv} \\ \mathbf{c}\mathbf{Av} \\ \mathbf{c}\mathbf{A}^2 \mathbf{v} \\ \mathbf{c}\mathbf{A}^3 \mathbf{v} \end{bmatrix} = \begin{bmatrix} \mathbf{cv} \\ \lambda_1 \mathbf{cv} \\ \lambda_1^2 \mathbf{cv} \\ \lambda_1^3 \mathbf{cv} \end{bmatrix} = \mathbf{0}$$

¹See also Example 2.8.3—in particular, (2.63).

which implies that the observability matrix does not have a full column rank. This contradicts the hypothesis that (7.9) is observable. Thus if (7.9) is observable, then $D(s)$ and $N(s)$ are coprime.

Next we show the converse; that is, if $D(s)$ and $N(s)$ are coprime, then (7.9) is observable. We show this by contradiction. Suppose (7.9) is not observable, then Theorem 6.O1 implies that there exists an eigenvalue λ_1 of \mathbf{A} and a nonzero vector \mathbf{v} such that

$$\begin{bmatrix} \mathbf{A} - \lambda_1 \mathbf{I} \\ \mathbf{c} \end{bmatrix} \mathbf{v} = \mathbf{0}$$

or

$$\mathbf{Av} = \lambda_1 \mathbf{v} \quad \text{and} \quad \mathbf{cv} = 0$$

Thus \mathbf{v} is an eigenvector of \mathbf{A} associated with eigenvalue λ_1 . From (7.13), we see that $\mathbf{v} = [\lambda_1^3 \lambda_1^2 \lambda_1 1]'$ is an eigenvector. Substituting this \mathbf{v} into $\mathbf{cv} = 0$ yields

$$N(\lambda_1) = \beta_1 \lambda_1^3 + \beta_2 \lambda_1^2 + \beta_3 \lambda_1 + \beta_4 = 0$$

Thus, λ_1 is a root of $N(s)$. The eigenvalue of \mathbf{A} is a root of its characteristic polynomial which, because of the companion form of \mathbf{A} , equals $D(s)$. Thus we also have $D(\lambda_1) = 0$, and $D(s)$ and $N(s)$ have the same root λ_1 . This contradicts the hypothesis that $D(s)$ and $N(s)$ are coprime. Thus if $D(s)$ and $N(s)$ are coprime, then (7.9) is observable. This establishes the theorem. Q.E.D.

If (7.9) is a realization of $\hat{g}(s)$, then we have, by definition,

$$\hat{g}(s) = \mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}$$

Taking its transpose yields

$$\hat{g}'(s) = \hat{g}(s) = [\mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}]' = \mathbf{b}'(s\mathbf{I} - \mathbf{A}')^{-1}\mathbf{c}'$$

Thus the state-space equation

$$\dot{\mathbf{x}}(t) = \mathbf{A}'\mathbf{x}(t) + \mathbf{c}'u(t) = \begin{bmatrix} -\alpha_1 & 1 & 0 & 0 \\ -\alpha_2 & 0 & 1 & 0 \\ -\alpha_3 & 0 & 0 & 1 \\ -\alpha_4 & 0 & 0 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} u(t) \quad (7.14)$$

$$y(t) = \mathbf{b}'\mathbf{x}(t) = [1 \ 0 \ 0 \ 0]\mathbf{x}(t)$$

is a different realization of (7.1). This state-space equation is always observable and is called an *observable-form* realization. Dual to Theorem 7.1, Equation (7.14) is controllable if and only if $D(s)$ and $N(s)$ are coprime.

We mention that the equivalence transformation $\bar{\mathbf{x}} = \mathbf{P}\mathbf{x}$ with

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad (7.15)$$

will transform (7.9) into

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\alpha_4 & -\alpha_3 & -\alpha_2 & -\alpha_1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} u(t)$$

$$y(t) = [\beta_4 \ \beta_3 \ \beta_2 \ \beta_1] \mathbf{x}(t)$$

This is also called a controllable-form equation. It is obtained from (7.9) by reordering the state variables. Similarly, (7.15) will transform (7.14) into

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 0 & 0 & -\alpha_4 \\ 1 & 0 & 0 & -\alpha_3 \\ 0 & 1 & 0 & -\alpha_2 \\ 0 & 0 & 1 & -\alpha_1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} \beta_4 \\ \beta_3 \\ \beta_2 \\ \beta_1 \end{bmatrix} u(t)$$

$$y(t) = [0 \ 0 \ 0 \ 1] \mathbf{x}(t)$$

This is a different observable-form equation.

7.2.1 Minimal Realizations

Consider the transfer function

$$\hat{g}(s) = \frac{2s^2 - 2s - 4}{s^3 + 3s^2 + 5s + 3} \quad (7.16)$$

It is strictly proper with a monic denominator. Its controllable-form realization is

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -3 & -5 & -3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u(t)$$

$$y(t) = [2 \ -2 \ -4] \mathbf{x}(t) \quad (7.17)$$

Using similarity transformations, we can obtain different three-dimensional realizations. If we multiply $s - 1$ to the denominator and numerator of $\hat{g}(s)$ to yield

$$\hat{g}(s) = \frac{(2s^2 - 2s - 4)(s - 1)}{(s^3 + 3s^2 + 5s + 3)(s - 1)} = \frac{2s^3 - 4s^2 - 2s + 4}{s^4 + 2s^3 + 2s^2 - 2s - 3}$$

then we can obtain the following four-dimensional controllable-form realization

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -2 & -2 & 2 & 3 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} u(t)$$

$$y(t) = [2 \ -4 \ -2 \ 4] \mathbf{x}(t)$$

and other realizations. Proceeding forward, we can obtain five-, six-, ..., dimensional realizations. Note that these realizations are all zero-state equivalent because they have

the same transfer function. In conclusion, a realizable transfer function has infinitely many realizations, not necessarily of the same dimension. An important question is then raised: What is the smallest possible dimension? Realizations with the smallest possible dimension are called *minimal-dimensional* or *minimal* realizations. This issue will be studied in this subsection. Before proceeding, we need the concept of degree for proper rational functions.

A rational function $\hat{g}(s) = N(s)/D(s)$ will also be called a *polynomial fraction* or, simply, a *fraction*. Because

$$\hat{g}(s) = \frac{N(s)}{D(s)} = \frac{N(s)Q(s)}{D(s)Q(s)}$$

for any polynomial $Q(s)$, fractions are not unique. If $N(s)$ and $D(s)$ have common roots, after canceling out the common roots, we have

$$\hat{g}(s) = \frac{N(s)}{D(s)} = \frac{\bar{N}(s)}{\bar{D}(s)}$$

where $\bar{N}(s)$ and $\bar{D}(s)$ are coprime. We call $\bar{N}(s)/\bar{D}(s)$ a *coprime fraction*. Then the *degree* of $\hat{g}(s)$ is defined as the degree of $\bar{D}(s)$. Note that we study only the case where $\deg \bar{D}(s) \geq \deg \bar{N}(s)$.

Is the transfer function in (7.16) a coprime fraction? Using the MATLAB function `roots` or `tfe2zp`, we can factor (7.16) as

$$\hat{g}(s) = \frac{2(s-2)(s+1)}{(s^2 + 2s + 3)(s+1)} = \frac{2(s-2)}{s^2 + 2s + 3}$$

Thus (7.16) is not a coprime fraction. This can also be verified by showing that (7.17) is not observable. Using its coprime fraction $(2s-4)/(s^2 + 2s + 3)$, we can obtain a realization as

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \begin{bmatrix} -2 & -3 \\ 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t) \\ y(t) &= [2 \quad -4] \mathbf{x}(t)\end{aligned}$$

It has dimension 2 which equals the degree of (7.16). Is this a minimal realization? This will be answered in the theorem that follows.

► Theorem 7.2

A state-space equation $(\mathbf{A}, \mathbf{b}, \mathbf{c}, d)$ is a minimal realization of a proper rational function $\hat{g}(s)$ if and only if (\mathbf{A}, \mathbf{b}) is controllable and (\mathbf{A}, \mathbf{c}) is observable or if and only if

$$\dim \mathbf{A} = \deg \hat{g}(s)$$



Proof: If (\mathbf{A}, \mathbf{b}) is not controllable; or if (\mathbf{A}, \mathbf{c}) is not observable, then the state-space equation can be reduced to a lesser dimensional equation that has the same transfer function (Theorems 6.6 and 6.O6). Thus $(\mathbf{A}, \mathbf{b}, \mathbf{c}, d)$ is not a minimal realization. This shows the necessity of the theorem.

To show the sufficiency, consider the n -dimensional controllable and observable state-space equation

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t) \\ y(t) &= \mathbf{c}\mathbf{x}(t) + du(t)\end{aligned}\tag{7.18}$$

Clearly, its $n \times n$ controllability matrix

$$\mathcal{C} = [\mathbf{b} \ \mathbf{Ab} \ \cdots \ \mathbf{A}^{n-1}\mathbf{b}]$$

and its $n \times n$ observability matrix

$$\mathcal{O} = \begin{bmatrix} \mathbf{c} \\ \mathbf{c}\mathbf{A} \\ \vdots \\ \mathbf{c}\mathbf{A}^{n-1} \end{bmatrix}$$

both have rank n . We show that (7.18) is a minimal realization by contradiction. Suppose the following \bar{n} -dimensional state-space equation, with $\bar{n} < n$,

$$\begin{aligned}\dot{\bar{\mathbf{x}}}(t) &= \bar{\mathbf{A}}\bar{\mathbf{x}}(t) + \bar{\mathbf{b}}u(t) \\ y(t) &= \bar{\mathbf{c}}\bar{\mathbf{x}}(t) + \bar{d}u(t)\end{aligned}\tag{7.19}$$

is a realization of $\hat{g}(s)$. Then Theorem 4.1 implies $d = \bar{d}$ and

$$\mathbf{c}\mathbf{A}^m\mathbf{b} = \bar{\mathbf{c}}\bar{\mathbf{A}}^m\bar{\mathbf{b}} \quad \text{for } m = 0, 1, 2, \dots\tag{7.20}$$

Let us consider the product

$$\begin{aligned}\mathcal{OC} &= \begin{bmatrix} \mathbf{c} \\ \mathbf{c}\mathbf{A} \\ \vdots \\ \mathbf{c}\mathbf{A}^{n-1} \end{bmatrix} [\mathbf{b} \ \mathbf{Ab} \ \cdots \ \mathbf{A}^{n-1}\mathbf{b}] \\ &= \begin{bmatrix} \mathbf{cb} & \mathbf{cAb} & \mathbf{cA}^2\mathbf{b} & \cdots & \mathbf{cA}^{n-1}\mathbf{b} \\ \mathbf{cAb} & \mathbf{cA}^2\mathbf{b} & \mathbf{cA}^3\mathbf{b} & \cdots & \mathbf{cA}^n\mathbf{b} \\ \mathbf{cA}^2\mathbf{b} & \mathbf{cA}^3\mathbf{b} & \mathbf{cA}^4\mathbf{b} & \cdots & \mathbf{cA}^{n+1}\mathbf{b} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{cA}^{n-1}\mathbf{b} & \mathbf{cA}^n\mathbf{b} & \mathbf{cA}^{n+1}\mathbf{b} & \cdots & \mathbf{cA}^{2(n-1)}\mathbf{b} \end{bmatrix}\end{aligned}\tag{7.21}$$

Using (7.20), we can replace every $\mathbf{c}\mathbf{A}^m\mathbf{b}$ by $\bar{\mathbf{c}}\bar{\mathbf{A}}^m\bar{\mathbf{b}}$. Thus we have

$$\mathcal{OC} = \bar{\mathcal{O}}_n \bar{\mathcal{C}}_n\tag{7.22}$$

where $\bar{\mathcal{O}}_n$ is defined as in (6.21) for the \bar{n} -dimensional state-space equation in (7.19) and $\bar{\mathcal{C}}_n$ is similarly defined. Because (7.18) is controllable and observable, we have $\rho(\mathcal{O}) = n$ and $\rho(\mathcal{C}) = n$. Thus (3.62) implies $\rho(\mathcal{OC}) = n$. Now $\bar{\mathcal{O}}_n$ and

$\bar{\mathcal{C}}_n$ are, respectively, $n \times \bar{n}$ and $\bar{n} \times n$; thus (3.61) implies that the matrix $\bar{\mathcal{O}}_n \bar{\mathcal{C}}_n$ has rank at most \bar{n} . This contradicts $\rho(\bar{\mathcal{O}}_n \bar{\mathcal{C}}_n) = \rho(\mathcal{OC}) = n$. Thus $(\mathbf{A}, \mathbf{b}, \mathbf{c}, d)$ is minimal. This establishes the first part of the theorem.

Consider a proper rational transfer function $\hat{g}(s) = N(s)/D(s)$ with its realization in (7.18). Then we have

$$\frac{N(s)}{D(s)} = \mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} + d = \frac{1}{\det(s\mathbf{I} - \mathbf{A})}\mathbf{c}[\text{Adj}(s\mathbf{I} - \mathbf{A})]\mathbf{b} + d$$

If $\{\mathbf{A}, \mathbf{b}\}$ is in controllable form, then we have

$$D(s) = k \det(s\mathbf{I} - \mathbf{A})$$

for some nonzero constant k . Note that $k = 1$ if $D(s)$ is monic. Now if $N(s)$ and $D(s)$ are coprime, then $\deg D(s) = \deg \hat{g}(s) = \dim \mathbf{A}$, and the controllable-form realization in (7.18) is also observable and, consequently, minimal. Because all minimal realizations are equivalent, as will be established immediately, we conclude that every realization is minimal if and only if $\dim \mathbf{A} = \deg \hat{g}(s)$. This establishes the theorem. Q.E.D.

To complete the proof of Theorem 7.2, we need the following theorem.

► Theorem 7.3

All minimal realizations of $\hat{g}(s)$ are equivalent.

⇒ **Proof:** Let $(\mathbf{A}, \mathbf{b}, \mathbf{c}, d)$ and $(\bar{\mathbf{A}}, \bar{\mathbf{b}}, \bar{\mathbf{c}}, \bar{d})$ be minimal realizations of $\hat{g}(s)$. Then we have $\bar{d} = d$ and, following (7.22),

$$\mathcal{OC} = \bar{\mathcal{O}}\bar{\mathcal{C}} \quad (7.23)$$

Multiplying \mathcal{OAC} out explicitly and then using (7.20), we can show

$$\mathcal{OAC} = \bar{\mathcal{O}}\bar{\mathbf{A}}\bar{\mathcal{C}} \quad (7.24)$$

Note that the controllability and observability matrices are all nonsingular square matrices. Let us define

$$\mathbf{P} := \bar{\mathcal{O}}^{-1}\mathcal{O}$$

Then (7.23) implies

$$\mathbf{P} = \bar{\mathcal{O}}^{-1}\mathcal{O} = \bar{\mathcal{C}}\mathcal{C}^{-1} \quad \text{and} \quad \mathbf{P}^{-1} = \mathcal{O}^{-1}\bar{\mathcal{O}} = \mathcal{C}\bar{\mathcal{C}}^{-1} \quad (7.25)$$

From (7.23), we have $\bar{\mathcal{C}} = \bar{\mathcal{O}}^{-1}\mathcal{OC} = \mathbf{PC}$. The first columns on both side of the equality yield $\bar{\mathbf{b}} = \mathbf{Pb}$. Again from (7.23), we have $\bar{\mathcal{O}} = \mathcal{OC}\bar{\mathcal{C}}^{-1} = \mathcal{OP}^{-1}$. The first rows on both sides of the equality yield $\bar{\mathbf{c}} = \mathbf{cP}^{-1}$. Equation (7.24) implies

$$\bar{\mathbf{A}} = \bar{\mathcal{O}}^{-1}\mathcal{OAC}\bar{\mathcal{C}}^{-1} = \mathbf{PAP}^{-1}$$

Thus $(\mathbf{A}, \mathbf{b}, \mathbf{c}, d)$ and $(\bar{\mathbf{A}}, \bar{\mathbf{b}}, \bar{\mathbf{c}}, \bar{d})$ are, following Definition 4.1, equivalent. This establishes the theorem. Q.E.D.

Theorems 7.2 and 7.3 have many important implications. Given a state-space equation, if we compute its transfer function and then its degree, then the minimality of the state-space equation can be readily determined without checking its controllability and observability. Thus the theorem provides an alternative way of checking controllability and observability. Conversely, given a rational function. If we compute first its common factors and reduce it to a coprime fraction, then the state-space equations obtained using its coefficients as shown in (7.9) and (7.14) will be automatically controllable and observable.

Consider a state-space equation $(\mathbf{A}, \mathbf{b}, \mathbf{c}, d)$ and its transfer function $\hat{g}(s)$. If the equation is controllable and observable, then every eigenvalue of \mathbf{A} is a pole of $\hat{g}(s)$ and every pole of $\hat{g}(s)$ is an eigenvalue of \mathbf{A} . See Example 5.2.2. In this case, we have

$$\text{Asymptotic stability} \iff \text{BIBO stability}$$

Note that the former is concerned with zero-input responses or internal stability and the latter is concerned with zero-state responses or external stability.

7.2.2 Complete Characterization

A state-space equation describes not only the relationship between the input and output of a system but also its internal variables, thus it is an internal description. Whereas a transfer function is an external description because it does not reveal anything inside the system. Moreover, a transfer function describes only zero-state responses, whereas a state-space equation describes zero-state as well as zero-input responses. Thus it is natural to ask: Does the transfer function describe fully a system or is there any information of the system missing from the transfer function? This is discussed next.

Consider the circuit shown in Fig. 6.1. Let us assign the voltage across the 1-F capacitor as x_1 as shown. Then its current is \dot{x}_1 and the current passing through the 3- Ω resistor is $x_1/3$. Their sum equals u . Thus we have

$$\dot{x}_1(t) + x_1(t)/3 = u(t) \quad \text{or} \quad \dot{x}_1(t) = -x_1(t)/3 + u(t)$$

If we assign the voltage across the 2-F capacitor as x_2 , then its current is $2\dot{x}_2$ and the current passing through the 4- Ω resistor is $x_2(t)/4$. Because of the open circuit at the output, we have $2\dot{x}_2(t) = -x_2(t)/4$ and the current passing through the 2- Ω resistor is u . Thus we have $y = 2u - x_2$. They can be arranged as

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} -1/3 & 0 \\ 0 & -1/8 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t)$$

$$y(t) = [0 \quad -1] \mathbf{x}(t) + 2u(t)$$

This two-dimensional state-space equation describes the circuit in Fig. 6.1.

We next compute its transfer function as

$$\begin{aligned}\hat{g}(s) &= \mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} + d = [0 \quad -1] \begin{bmatrix} 1/(s+1/3) & 0 \\ 0 & 1/(s+1/8) \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \\ &= 0 + 2 = 2\end{aligned}$$

This transfer function does not describe fully the circuit because the responses in the two RC loops cannot be detected from $\hat{g}(s) = 2$. On the other hand, the circuit is completely described by its state-space equation. For example, if $x_2(0) \neq 0$, then the state equation will generate the response $x_2(t) = e^{-t/8}x_2(0)$. In conclusion, state-space equations are generally more descriptive than transfer functions.

Even though the state-space equation is more descriptive, the equation is not both controllable and observable because its dimension is different from the degree of its transfer function $\hat{g}(s) = 2$ which is zero. We can also verify directly that the equation is not controllable nor observable (Problem 6.5). To discuss its physical significance, we next introduce a new concept.

Consider a system with transfer function $\hat{g}(s)$. The system is defined to be *completely characterized* by its transfer function if the number of energy storage elements in the system equals the degree of $\hat{g}(s)$. For an RLC circuit, because an inductor can store energy in its magnetic field and a capacitor can store energy in its electric field, the number of energy storage elements in the circuit is the total number of inductors and capacitors. Note that a resistor is not an energy storage element because all its energy is dissipated into heat. For example, the circuit in Fig. 6.1 has two energy-storage elements, but its transfer function $\hat{g}(s) = 2$ has degree 0. Thus the circuit is not completely characterized by its transfer function.

In practice, we try to design a simplest possible system to achieve a given task. As far as the input and output are concerned, the circuit in Fig. 6.1 can be replaced by a $2\text{-}\Omega$ resistor. In other words, the circuit has some redundancies or some unnecessary components, and we should not design such a system. It turns out that such redundancies can be detected from their transfer functions. If the number of energy storage elements in a system is larger than the degree of its transfer function, then the system has some redundancy. Most practical systems, unless inadvertently designed, have no redundant components and are completely characterized by their transfer functions.²

The redundancy of a system can also be checked from its state-space equation. From the Kalman decomposition shown in Fig. 6.7, we can conclude that a system has no redundancy if and only if its state-space equation is controllable and observable. However this way of checking is generally more complicated than the one based on transfer functions. For example, the two circuits in Fig. 4.5 both have transfer function $1/(s+2)$ of degree 1 as computed in Example 4.4.2. The circuit in Fig. 4.5(a) has two energy-storage elements and is therefore not completely characterized by its transfer function, whereas the circuit in Fig. 4.5(b) has only one energy-storage element and is completely characterized by its transfer function. There is no need to develop state-space equations and then to check their controllability and observability.

²The redundancy in this section is different from the redundancy purposely introduced for safety reasons, such as to use two or three identical systems to achieve the same control in the space station.

Most systems designed in practice do not have unnecessary components. For such a system, will its state-space equation be more descriptive than its transfer function? The answer is negative. If a system has no redundant components and if we develop a state-space equation and a transfer function to describe the system, then we have

$$\begin{aligned}\text{Number of energy-storage elements} &= \text{Degree of its transfer function} \\ &= \text{Dimension of its state-space equation}\end{aligned}$$

In this case, there is no difference in using its transfer function or state-space equation to describe the system. To be more specific, all internal characteristics of the system will appear in the transfer function. The response excited by any set of initial conditions can be generated from its transfer function by applying some input. In other words, every zero-input response of the system can be generated as a zero-state response. Thus nothing is lost in using transfer functions. In conclusion, *controllable and observable state-space equations and coprime fractions contain essentially the same information, and either description can be used to carry out analysis and design.*

Most designs in filters and control systems are carried out using transfer functions. Once a satisfactory transfer function is designed, we then use its minimal realization to carry out op-amp circuit implementation. Then the resulting system has no redundant components.

To conclude this subsection, we mention that complete characterization deals only with energy-storage elements; it pays no attention to non-energy storage elements such as resistors.

7.3 Computing Coprime Fractions

The importance of coprime fractions and degrees was demonstrated in the preceding section. In this section, we discuss how to compute coprime fractions. Consider a proper rational function

$$\hat{g}(s) = \frac{N(s)}{D(s)}$$

where $N(s)$ and $D(s)$ are polynomials. If we use the MATLAB function `roots` to compute their roots and then to cancel their common roots, we will obtain a coprime fraction. We may also use the Euclidean algorithm to compute their common factors or roots. In this section, we introduce a different method by solving a set of linear algebraic equations. The method does not offer any advantages over the aforementioned methods for scalar rational functions. However, it can be readily extended to the matrix case. More importantly, the method will be used to carry out design in Chapter 9.

Consider $N(s)/D(s)$. To simplify the discussion, we assume $\deg N(s) \leq \deg D(s) = n = 4$. Let us write

$$\frac{N(s)}{D(s)} = \frac{\bar{N}(s)}{\bar{D}(s)}$$

which implies

$$D(s)(-\bar{N}(s)) + N(s)\bar{D}(s) = 0 \quad (7.26)$$

It is clear that $D(s)$ and $N(s)$ are not coprime if and only if there exist polynomials $\bar{N}(s)$ and $\bar{D}(s)$ with $\deg \bar{N}(s) \leq \deg \bar{D}(s) < n = 4$ to meet (7.26). The condition $\deg \bar{D}(s) < n$ is crucial; otherwise, (7.26) has infinitely many solutions $\bar{N}(s) = N(s)R(s)$ and $\bar{D}(s) = D(s)R(s)$ for any polynomial $R(s)$. Thus the coprimeness problem can be reduced to solving the polynomial equation in (7.26).

Instead of solving (7.26) directly, we will change it into solving a set of linear algebraic equations. We write

$$\begin{aligned} D(s) &= D_0 + D_1s + D_2s^2 + D_3s^3 + D_4s^4 \\ N(s) &= N_0 + N_1s + N_2s^2 + N_3s^3 + N_4s^4 \\ \bar{D}(s) &= \bar{D}_0 + \bar{D}_1s + \bar{D}_2s^2 + \bar{D}_3s^3 \\ \bar{N}(s) &= \bar{N}_0 + \bar{N}_1s + \bar{N}_2s^2 + \bar{N}_3s^3 \end{aligned} \quad (7.27)$$

where $D_4 \neq 0$ and the remaining D_i , N_i , \bar{D}_i , and \bar{N}_i can be zero or nonzero. Note that the subscript of D and N corresponds to the power of s . Substituting these into (7.26) and equating to zero the coefficients associated with s^k , for $k = 0, 1, \dots, 7$, yield

$$\mathbf{Sm} := \left[\begin{array}{cccccc|cc|c} D_0 & N_0 & : & 0 & 0 & : & 0 & 0 & : & 0 & 0 \\ D_1 & N_1 & : & D_0 & N_0 & : & 0 & 0 & : & 0 & 0 \\ D_2 & N_2 & : & D_1 & N_1 & : & D_0 & N_0 & : & 0 & 0 \\ D_3 & N_3 & : & D_2 & N_2 & : & D_1 & N_1 & : & D_0 & N_0 \\ D_4 & N_4 & : & D_3 & N_3 & : & D_2 & N_2 & : & D_1 & N_1 \\ 0 & 0 & : & D_4 & N_4 & : & D_3 & N_3 & : & D_2 & N_2 \\ 0 & 0 & : & 0 & 0 & : & D_4 & N_4 & : & D_3 & N_3 \\ 0 & 0 & : & 0 & 0 & : & 0 & 0 & : & D_4 & N_4 \end{array} \right] \left[\begin{array}{c} -\bar{N}_0 \\ \bar{D}_0 \\ \vdots \\ -\bar{N}_1 \\ \bar{D}_1 \\ \vdots \\ -\bar{N}_2 \\ \bar{D}_2 \\ \vdots \\ -\bar{N}_3 \\ \bar{D}_3 \end{array} \right] = \mathbf{0} \quad (7.28)$$

This is a homogeneous linear algebraic equation. The first block column of \mathbf{S} consists of two columns formed from the coefficients of $D(s)$ and $N(s)$ arranged in ascending power of s . The second block column is the first block column shifted down one position. Repeating the process until \mathbf{S} is a square matrix of order $2n = 8$. The square matrix \mathbf{S} is called the *Sylvester resultant*. If the Sylvester resultant is singular, nonzero solutions exist in (7.28) (Theorem 3.3). This means that polynomials $\bar{N}(s)$ and $\bar{D}(s)$ of degree 3 or less exist to meet (7.26). Thus $D(s)$ and $N(s)$ are not coprime. If the Sylvester resultant is nonsingular, no nonzero solutions exist in (7.28) or, equivalently, no polynomials $\bar{N}(s)$ and $\bar{D}(s)$ of degree 3 or less exist to meet (7.26). Thus $D(s)$ and $N(s)$ are coprime. In conclusion, $D(s)$ and $N(s)$ are coprime if and only if the Sylvester resultant is nonsingular.

If the Sylvester resultant is singular, then $N(s)/D(s)$ can be reduced to

$$\frac{N(s)}{D(s)} = \frac{\bar{N}(s)}{\bar{D}(s)}$$

where $\bar{N}(s)$ and $\bar{D}(s)$ are coprime. We discuss how to obtain a coprime fraction directly from (7.28). Let us search linearly independent columns of \mathbf{S} in order from left to right. We call columns formed from D_i D -columns and formed from N_i N -columns. Then every D -column is linearly independent of its left-hand-side (LHS) columns. Indeed, because $D_4 \neq 0$, the first D -column is linearly independent. The second D -column is also linearly independent of its LHS columns because the LHS entries of D_4 are all zero. Proceeding forward, we conclude that all D -columns are linearly independent of their LHS columns. On the other hand, an N -column can be dependent or independent of its LHS columns. Because of the repetitive pattern of \mathbf{S} , if an N -column becomes linearly dependent on its LHS columns, then all subsequent N -columns are linearly dependent on their LHS columns. Let μ denote the number of linearly independent N -columns in \mathbf{S} . Then the $(\mu + 1)$ th N -column is the first N -column to become linearly dependent on its LHS columns and will be called the *primary dependent N -column*. Let us use \mathbf{S}_1 to denote the submatrix of \mathbf{S} that consists of the primary dependent N -column and all its LHS columns. That is, \mathbf{S}_1 consists of $(\mu + 1)$ D -columns (all of them are linearly independent) and $(\mu + 1)$ N -columns (the last one is dependent). Thus \mathbf{S}_1 has $2(\mu + 1)$ columns but rank $2\mu + 1$. In other words, \mathbf{S}_1 has nullity 1 and, consequently, has one independent null vector. Note that if $\bar{\mathbf{n}}$ is a null vector, so is $\alpha\bar{\mathbf{n}}$ for any nonzero α . Although any null vector can be used, we will use exclusively the null vector with 1 as its last entry to develop $\bar{N}(s)$ and $\bar{D}(s)$. For convenience, we call such a null vector a *monic null vector*. If we use the MATLAB function `null` to generate a null vector, then the null vector must be divided by its last entry to yield a monic null vector. This is illustrated in the next example.

Example 7.3.1 Consider

$$\frac{N(s)}{D(s)} = \frac{6s^3 + s^2 + 3s - 20}{2s^4 + 7s^3 + 15s^2 + 16s + 10} \quad (7.29)$$

We have $n = 4$ and its Sylvester resultant \mathbf{S} is 8×8 . The fraction is coprime if and only if \mathbf{S} is nonsingular or has rank 8. We use MATLAB to check the rank of \mathbf{S} . Because it is simpler to key in the transpose of \mathbf{S} , we type

```
d=[10 16 15 7 2];n=[-20 3 1 6 0];
s=[d 0 0 0;n 0 0 0;0 d 0 0;0 n 0 0;...
    0 0 d 0;0 0 n 0;0 0 0 d;0 0 0 n]';
m=rank(s)
```

The answer is 6; thus $D(s)$ and $N(s)$ are not coprime. Because all four D -columns of \mathbf{S} are linearly independent, we conclude that \mathbf{S} has only two linearly independent N -columns and $\mu = 2$. The third N -column is the primary dependent N -column, and all its LHS columns are linearly independent. Let \mathbf{S}_1 denote the first six columns of \mathbf{S} , an 8×6 matrix. The submatrix \mathbf{S}_1 has three D -columns (all linearly independent)

and two linearly independent N -columns, thus it has rank 5 and nullity 1. Because all entries of the last row of \mathbf{S}_1 are zero, they can be skipped in forming \mathbf{S}_1 . We type

```
s1=[d 0 0;n 0 0;0 d 0;0 n 0;0 0 d;0 0 n]';
z=null(s1)
```

which yields

```
ans z= [-0.6860 -0.3430 0.5145 -0.3430 -0.0000
-0.1715 ]'
```

This null vector does not have 1 as its last entry. We divide it by the last entry or the sixth entry of z by typing

```
zb=z/z(6)
```

which yields

```
ans zb= [4 2 -3 2 0 1]'
```

This monic null vector equals $[-\bar{N}_0 \bar{D}_0 -\bar{N}_1 \bar{D}_1 -\bar{N}_2 \bar{D}_2]'$. Thus we have

$$\bar{N}(s) = -4 + 3s + 0 \cdot s^2, \quad \bar{D}(s) = 2 + 2s + s^2$$

and

$$\frac{6s^3 + s^2 + 3s - 20}{2s^4 + 7s^3 + 15s^2 + 16s + 10} = \frac{3s - 4}{s^2 + 2s + 2}$$

Because the null vector is computed from the first linearly dependent N -column, the computed $\bar{N}(s)$ and $\bar{D}(s)$ have the smallest possible degrees to meet (7.26) and, therefore, are coprime. This completes the reduction of $N(s)/D(s)$ to a coprime fraction.

The preceding procedure can be summarized as a theorem.

► **Theorem 7.4**

Consider $\hat{g}(s) = N(s)/D(s)$. We use the coefficients of $D(s)$ and $N(s)$ to form the Sylvester resultant \mathbf{S} in (7.28) and search its linearly independent columns in order from left to right. Then we have

$$\deg \hat{g}(s) = \text{no. of linearly independent } N\text{-columns} =: \mu$$

and the coefficients of a coprime fraction $\hat{g}(s) = \bar{N}(s)/\bar{D}(s)$ or

$$[-\bar{N}_0 \bar{D}_0 -\bar{N}_1 \bar{D}_1 \cdots -\bar{N}_\mu \bar{D}_\mu]'$$

equal the monic null vector of the submatrix that consists of the primary dependent N -column and all its LHS linearly independent columns of \mathbf{S} .

We mention that if D - and N -columns in \mathbf{S} are arranged in descending powers of s , then it is not true that all D -columns are linearly independent of their LHS columns and that the degree of $\hat{g}(s)$ equals the number of linearly independent N -columns. See Problem 7.9. Thus it is essential to arrange the D - and N -columns in ascending powers of s in \mathbf{S} .

7.3.1 QR Decomposition

As discussed in the preceding section, a coprime fraction can be obtained by searching linearly independent columns of the Sylvester resultant in order from left to right. It turns out the widely available QR decomposition can be used to achieve this searching.

Consider an $n \times m$ matrix \mathbf{M} . Then there exist an $n \times n$ orthogonal matrix \mathbf{Q} and an $n \times m$ upper triangular matrix \mathbf{R} such that

$$\mathbf{M} = \mathbf{QR}$$

This is called the QR decomposition. There are a number of ways to carry out the decomposition. See Wikipedia. One way is to use the Schmidt orthonormalization procedure discussed in Section 3.2. In fact, Example 3.2.2 showed $\mathbf{Q} = \mathbf{A}\bar{\mathbf{R}}$, for a nonsingular \mathbf{A} with \mathbf{Q} orthogonal and $\bar{\mathbf{R}}$ upper triangular. Let $\mathbf{R} := \bar{\mathbf{R}}^{-1}$. Then $\mathbf{Q} = \mathbf{A}\bar{\mathbf{R}}$ implies $\mathbf{A} = \mathbf{Q}\bar{\mathbf{R}}^{-1} = \mathbf{QR}$. Because the inverse of an upper triangular matrix is an upper triangular matrix, the example showed a QR decomposition. In MATLAB, \mathbf{Q} and \mathbf{R} can be obtained by typing `[q, r] = qr(m)`.

We now discuss the use of the decomposition. Because \mathbf{Q} is orthogonal, we have $\mathbf{Q}^{-1} = \mathbf{Q}' =: \bar{\mathbf{Q}}$. Premultiplication of \mathbf{Q}^{-1} to $\mathbf{M} = \mathbf{QR}$ yields

$$\bar{\mathbf{Q}}\mathbf{M} = \mathbf{R}$$

where \mathbf{R} is an upper triangular matrix of the same dimensions as \mathbf{M} . Because $\bar{\mathbf{Q}}$ operates on the rows of \mathbf{M} , the linear independence of the columns of \mathbf{M} is preserved in the columns of \mathbf{R} . In other words, if a column of \mathbf{R} is linearly dependent on its left-hand-side (LHS) columns, so is the corresponding column of \mathbf{M} . Now because \mathbf{R} is in upper triangular form, its m th column is linearly independent of its LHS columns if and only if its m th entry at the diagonal position is nonzero. Thus using \mathbf{R} , the linearly independent columns of \mathbf{M} , in order from left to right, can be obtained by inspection.

Let us apply QR decomposition to the resultant in Example 7.3.1. We type

```
d=[10 16 15 7 2];n=[-20 3 1 6 0];
s=[d 0 0 0;n 0 0 0;0 d 0 0;0 n 0 0;...
    0 0 d 0;0 0 n 0;0 0 0 d;0 0 0 n]';
[q, r]=qr(s)
```

Because \mathbf{Q} is not needed, we show only \mathbf{R} :

$$r = \begin{bmatrix} -25.1 & 3.7 & -20.6 & 10.1 & -11.6 & 11.0 & -4.1 & 5.3 \\ 0 & -20.7 & -10.3 & 4.3 & -7.2 & 2.1 & -3.6 & 6.7 \\ 0 & 0 & -10.2 & -15.6 & -20.3 & 0.8 & -16.8 & 9.6 \\ 0 & 0 & 0 & 8.9 & -3.5 & -17.9 & -11.2 & 7.3 \\ 0 & 0 & 0 & 0 & -5.0 & 0 & -12.0 & -15.0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -2.0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -4.6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We see that the matrix is upper triangular. Because the sixth column has 0 as its sixth entry (diagonal position), it is linearly dependent on its LHS columns. So is the last

column. To determine whether a column is linearly dependent, we need to know only whether the diagonal entry is zero or not. Thus the matrix can be simplified as

$$r = \begin{bmatrix} d & x & x & x & x & x & x & x \\ 0 & n & x & x & x & x & x & x \\ 0 & 0 & d & x & x & x & x & x \\ 0 & 0 & 0 & n & x & x & x & x \\ 0 & 0 & 0 & 0 & d & 0 & x & x \\ 0 & 0 & 0 & 0 & 0 & 0 & x & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & d & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

where d, n and x denote nonzero entries and d also denotes D -column and n denotes N -column. We see that every D -column is linearly independent of its LHS columns, and there are only two linearly independent N -columns. Thus by employing QR decomposition, we obtain immediately μ and the primary dependent N -column. In scalar transfer functions, we can use either `rank` or `qr` to find μ . In the matrix case, using `rank` is very inconvenient; we will use QR decomposition.

7.4 Balanced Realization³

Every transfer function has infinitely many minimal realizations. Among these realizations, it is of interest to see which realizations are more suitable for practical implementation. If we use the controllable or observable form, then the A -matrix and b - or c -vector have many zero entries, and its implementation will use a small number of components. However, either form is very sensitive to parameter variations; therefore, both forms should be avoided if sensitivity is an important issue. If all eigenvalues of A are distinct, we can transform A , using an equivalence transformation, into a diagonal form (if all eigenvalues are real), or into the modal form discussed in Section 4.4.1 (if some eigenvalues are complex). The diagonal or modal form has many zero entries in A and will use a small number of components in its implementation. More importantly, the diagonal and modal forms are least sensitive to parameter variations among all realizations; thus they are good candidates for practical implementation.

We discuss next a different minimal realization, called a balanced realization. However, the discussion is applicable only to stable A . Consider

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t) \\ y(t) &= cx(t) \end{aligned} \tag{7.30}$$

It is assumed that A is stable or all its eigenvalues have negative real parts. Then the controllability Gramian W_c and the observability Gramian W_o are, respectively, the unique solutions of

$$AW_c + W_c A' = -bb' \tag{7.31}$$

³This section may be skipped without loss of continuity.

and

$$\mathbf{A}'\mathbf{W}_o + \mathbf{W}_o\mathbf{A} = -\mathbf{c}'\mathbf{c} \quad (7.32)$$

They are positive definite if (7.30) is controllable and observable.

Different minimal realizations of the same transfer function have different controllability and observability Gramians. For example, the state-space equation, taken from Reference 25,

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \begin{bmatrix} -1 & -4/\alpha \\ 4\alpha & -2 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 2\alpha \end{bmatrix} u(t) \\ y(t) &= [-1 \ 2/\alpha] \mathbf{x}(t)\end{aligned} \quad (7.33)$$

for any nonzero α , has transfer function $\hat{g}(s) = (3s + 18)/(s^2 + 3s + 18)$ and is controllable and observable. Its controllability and observability Gramians can be computed as

$$\mathbf{W}_c = \begin{bmatrix} 0.5 & 0 \\ 0 & \alpha^2 \end{bmatrix} \quad \text{and} \quad \mathbf{W}_o = \begin{bmatrix} 0.5 & 0 \\ 0 & 1/\alpha^2 \end{bmatrix} \quad (7.34)$$

We see that different α yields different minimal realization and different controllability and observability Gramians. Even though the controllability and observability Gramians will change; their product, however, remains the same as $\text{diag}(0.25, 1)$ for all α .

► Theorem 7.5

Let $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ and $(\bar{\mathbf{A}}, \bar{\mathbf{b}}, \bar{\mathbf{c}})$ be minimal and equivalent and let $\mathbf{W}_c \mathbf{W}_o$ and $\bar{\mathbf{W}}_c \bar{\mathbf{W}}_o$ be the products of their controllability and observability Gramians. Then $\mathbf{W}_c \mathbf{W}_o$ and $\bar{\mathbf{W}}_c \bar{\mathbf{W}}_o$ are similar and their eigenvalues are all real and positive.

⇒ **Proof:** Let $\bar{\mathbf{x}} = \mathbf{P}\mathbf{x}$, where \mathbf{P} is an nonsingular constant matrix. Then we have

$$\bar{\mathbf{A}} = \mathbf{P}\mathbf{A}\mathbf{P}^{-1}, \quad \bar{\mathbf{b}} = \mathbf{P}\mathbf{b}, \quad \bar{\mathbf{c}} = \mathbf{c}\mathbf{P}^{-1} \quad (7.35)$$

The controllability Gramian $\bar{\mathbf{W}}_c$ and observability Gramian $\bar{\mathbf{W}}_o$ of $(\bar{\mathbf{A}}, \bar{\mathbf{b}}, \bar{\mathbf{c}})$ are, respectively, the unique solutions of

$$\bar{\mathbf{A}}\bar{\mathbf{W}}_c + \bar{\mathbf{W}}_c\bar{\mathbf{A}}' = -\bar{\mathbf{b}}\bar{\mathbf{b}}' \quad (7.36)$$

and

$$\bar{\mathbf{A}}'\bar{\mathbf{W}}_o + \bar{\mathbf{W}}_o\bar{\mathbf{A}} = -\bar{\mathbf{c}}'\bar{\mathbf{c}} \quad (7.37)$$

Substituting $\bar{\mathbf{A}} = \mathbf{P}\mathbf{A}\mathbf{P}^{-1}$ and $\bar{\mathbf{b}} = \mathbf{P}\mathbf{b}$ into (7.36) yields

$$\mathbf{P}\mathbf{A}\mathbf{P}^{-1}\bar{\mathbf{W}}_c + \bar{\mathbf{W}}_c(\mathbf{P}')^{-1}\mathbf{A}'\mathbf{P}' = -\mathbf{P}\mathbf{b}\mathbf{b}'\mathbf{P}'$$

which implies

$$\mathbf{A}\mathbf{P}^{-1}\bar{\mathbf{W}}_c(\mathbf{P}')^{-1} + \mathbf{P}^{-1}\bar{\mathbf{W}}_c(\mathbf{P}')^{-1}\mathbf{A}' = -\mathbf{b}\mathbf{b}'$$

Comparing this with (7.31) yields

$$\mathbf{W}_c = \mathbf{P}^{-1} \bar{\mathbf{W}}_c (\mathbf{P}')^{-1} \text{ or } \bar{\mathbf{W}}_c = \mathbf{P} \mathbf{W}_c \mathbf{P}' \quad (7.38)$$

Similarly, we can show

$$\mathbf{W}_o = \mathbf{P}' \bar{\mathbf{W}}_o \mathbf{P} \text{ or } \bar{\mathbf{W}}_o = (\mathbf{P}')^{-1} \mathbf{W}_o \mathbf{P}^{-1} \quad (7.39)$$

Thus we have

$$\mathbf{W}_c \mathbf{W}_o = \mathbf{P}^{-1} \bar{\mathbf{W}}_c (\mathbf{P}')^{-1} \mathbf{P}' \bar{\mathbf{W}}_o \mathbf{P} = \mathbf{P}^{-1} \bar{\mathbf{W}}_c \bar{\mathbf{W}}_o \mathbf{P}$$

This shows that all $\mathbf{W}_c \mathbf{W}_o$ are similar and, consequently, have the same set of eigenvalues.

Next we show that all eigenvalues of $\mathbf{W}_c \mathbf{W}_o$ are real and positive. Note that both \mathbf{W}_c and \mathbf{W}_o are symmetric, but their product may not be. Therefore, Theorem 3.6 is not directly applicable to $\mathbf{W}_c \mathbf{W}_o$. Now we apply Theorem 3.6 to \mathbf{W}_c :

$$\mathbf{W}_c = \mathbf{Q}' \mathbf{D} \mathbf{Q} = \mathbf{Q}' \mathbf{D}^{1/2} \mathbf{D}^{1/2} \mathbf{Q} =: \mathbf{R}' \mathbf{R} \quad (7.40)$$

where \mathbf{D} is a diagonal matrix with the eigenvalues of \mathbf{W}_c on the diagonal. Because \mathbf{W}_c is symmetric and positive definite, all its eigenvalues are real and positive. Thus we can express \mathbf{D} as $\mathbf{D}^{1/2} \mathbf{D}^{1/2}$, where $\mathbf{D}^{1/2}$ is diagonal with positive square roots of the diagonal entries of \mathbf{D} as its diagonal entries. Note that \mathbf{Q} is orthogonal or $\mathbf{Q}^{-1} = \mathbf{Q}'$. The matrix $\mathbf{R} = \mathbf{D}^{1/2} \mathbf{Q}$ is not orthogonal but is nonsingular.

Consider $\mathbf{R} \mathbf{W}_o \mathbf{R}'$; it is clearly symmetric and positive definite. Thus its eigenvalues are all real and positive. Using (7.40) and (3.66), we have

$$\det(\sigma^2 \mathbf{I} - \mathbf{W}_c \mathbf{W}_o) = \det(\sigma^2 \mathbf{I} - \mathbf{R}' \mathbf{R} \mathbf{W}_o) = \det(\sigma^2 \mathbf{I} - \mathbf{R} \mathbf{W}_o \mathbf{R}') \quad (7.41)$$

which implies that $\mathbf{W}_c \mathbf{W}_o$ and $\mathbf{R} \mathbf{W}_o \mathbf{R}'$ have the same set of eigenvalues. Thus we conclude that all eigenvalues of $\mathbf{W}_c \mathbf{W}_o$ are real and positive. Q.E.D.

Let us define

$$\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n) \quad (7.42)$$

where σ_i are positive square roots of the eigenvalues of $\mathbf{W}_c \mathbf{W}_o$. For convenience, we arrange them in descending order in magnitude or

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n > 0$$

These eigenvalues are called the *Hankel singular values*. The product $\mathbf{W}_c \mathbf{W}_o$ of any minimal realization is similar to Σ^2 .

► Theorem 7.6

For any n -dimensional minimal state-space equation $(\mathbf{A}, \mathbf{b}, \mathbf{c})$, there exists an equivalence transformation $\tilde{\mathbf{x}} = \mathbf{P}\mathbf{x}$ such that the controllability Gramian $\bar{\mathbf{W}}_c$ and observability Gramian $\bar{\mathbf{W}}_o$ of its equivalent state-space equation have the property

$$\bar{\mathbf{W}}_c = \bar{\mathbf{W}}_o = \Sigma \quad (7.43)$$

This is called a *balanced realization*.

⇒ **Proof:** We first compute $\mathbf{W}_c = \mathbf{R}'\mathbf{R}$ as in (7.40). We then apply singular value decomposition to $\mathbf{RW}_o\mathbf{R}'$ to yield

$$\mathbf{RW}_o\mathbf{R}' = \mathbf{U}\Sigma^2\mathbf{U}'$$

where \mathbf{U} is orthogonal or $\mathbf{U}'\mathbf{U} = \mathbf{I}$. Let

$$\mathbf{P}^{-1} = \mathbf{R}'\mathbf{U}\Sigma^{-1/2} \text{ or } \mathbf{P} = \Sigma^{1/2}\mathbf{U}'(\mathbf{R}')^{-1}$$

Then (7.38) and $\mathbf{W}_c = \mathbf{R}'\mathbf{R}$ imply

$$\bar{\mathbf{W}}_c = \Sigma^{1/2}\mathbf{U}'(\mathbf{R}')^{-1}\mathbf{W}_c\mathbf{R}^{-1}\mathbf{U}\Sigma^{1/2} = \Sigma$$

and (7.39) and $\mathbf{RW}_o\mathbf{R}' = \mathbf{U}\Sigma^2\mathbf{U}'$ imply

$$\bar{\mathbf{W}}_o = \Sigma^{-1/2}\mathbf{U}'\mathbf{R}\mathbf{W}_o\mathbf{R}'\mathbf{U}\Sigma^{-1/2} = \Sigma$$

This establishes the theorem. Q.E.D.

By selecting a different \mathbf{P} , it is possible to find an equivalent state equation with $\bar{\mathbf{W}}_c = \mathbf{I}$ and $\bar{\mathbf{W}}_o = \Sigma^2$. Such a state-space equation is called the *input-normal* realization. Similarly, we can have a state-space equation with $\bar{\mathbf{W}}_c = \Sigma^2$ and $\bar{\mathbf{W}}_o = \mathbf{I}$ which is called the *output-normal* realization. The balanced realization in Theorem 7.5 can be used in system reduction. More specifically, suppose

$$\begin{bmatrix} \dot{\mathbf{x}}_1(t) \\ \dot{\mathbf{x}}_2(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} + \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} u(t)$$

$$y(t) = [\mathbf{c}_1 \quad \mathbf{c}_2] \mathbf{x}(t) \tag{7.44}$$

is a balanced minimal realization of a stable $\hat{g}(s)$ with

$$\mathbf{W}_c = \mathbf{W}_o = \text{diag}(\Sigma_1, \Sigma_2)$$

where the A -, b -, and c -matrices are partitioned according to the order of Σ_i . If the Hankel singular values of Σ_1 and Σ_2 are disjoint, then the reduced state-space equation

$$\begin{aligned} \dot{\mathbf{x}}_1(t) &= \mathbf{A}_{11}\mathbf{x}_1(t) + \mathbf{b}_1 u(t) \\ y(t) &= \mathbf{c}_1 \mathbf{x}_1(t) \end{aligned} \tag{7.45}$$

is balanced and \mathbf{A}_{11} is stable. If the singular values of Σ_2 are much smaller than those of Σ_1 , then the transfer function of (7.45) will be close to $\hat{g}(s)$. See Reference 25.

The MATLAB function `balreal` will transform $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ into a balanced state-space equation. The reduced equation in (7.45) can be obtained by using `balred`.

7.5 Realizations from Markov Parameters⁴

Consider the strictly proper rational function

$$\hat{g}(s) = \frac{\beta_1 s^{n-1} + \beta_2 s^{n-2} + \cdots + \beta_{n-1} s + \beta_n}{s^n + \alpha_1 s^{n-1} + \alpha_2 s^{n-2} + \cdots + \alpha_{n-1} s + \alpha_n} \quad (7.46)$$

We expand it into an infinite power series as

$$\hat{g}(s) = h(0) + h(1)s^{-1} + h(2)s^{-2} + \cdots \quad (7.47)$$

If $\hat{g}(s)$ is strictly proper as assumed in (7.46), then $h(0) = 0$. The coefficients $h(m)$, $m = 1, 2, \dots$ are called *Markov parameters*. Let $g(t)$ be the inverse Laplace transform of $\hat{g}(s)$ or, equivalently, the impulse response of the system. Then we have

$$h(m) = \left. \frac{d^{m-1}}{dt^{m-1}} g(t) \right|_{t=0}$$

for $m = 1, 2, 3, \dots$. This method of computing Markov parameters is impractical because it requires repetitive differentiations and differentiations are susceptible to noise.⁵ Equating (7.46) and (7.47) yields

$$\begin{aligned} & \beta_1 s^{n-1} + \beta_2 s^{n-2} + \cdots + \beta_n \\ &= (s^n + \alpha_1 s^{n-1} + \alpha_2 s^{n-2} + \cdots + \alpha_n)(h(1)s^{-1} + h(2)s^{-2} + \cdots) \end{aligned}$$

From this equation, we can obtain the Markov parameters recursively as

$$\begin{aligned} h(1) &= \beta_1 \\ h(2) &= -\alpha_1 h(1) + \beta_2 \\ h(3) &= -\alpha_1 h(2) - \alpha_2 h(1) + \beta_3 \\ &\vdots \\ h(n) &= -\alpha_1 h(n-1) - \alpha_2 h(n-2) - \cdots - \alpha_{n-1} h(1) + \beta_n \end{aligned} \quad (7.48)$$

$$\begin{aligned} h(m) &= -\alpha_1 h(m-1) - \alpha_2 h(m-2) - \cdots - \alpha_{n-1} h(m-n+1) \\ &\quad - \alpha_n h(m-n) \end{aligned} \quad (7.49)$$

for $m = n+1, n+2, \dots$

⁴This section may be skipped without loss of continuity.

⁵In the discrete-time case, if we apply an impulse sequence to a system, then the output sequence yields directly Markov parameters. Thus Markov parameters can be easily generated in discrete-time systems.

Next we use the Markov parameters to form the $\alpha \times \beta$ matrix

$$\mathbf{T}(\alpha, \beta) = \begin{bmatrix} h(1) & h(2) & h(3) & \cdots & h(\beta) \\ h(2) & h(3) & h(4) & \cdots & h(\beta+1) \\ h(3) & h(4) & h(5) & \cdots & h(\beta+2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h(\alpha) & h(\alpha+1) & h(\alpha+2) & \cdots & h(\alpha+\beta-1) \end{bmatrix} \quad (7.50)$$

It is called a *Hankel matrix*. It is important to mention that even if $h(0) \neq 0$, $h(0)$ does not appear in the Hankel matrix.

► Theorem 7.7

A strictly proper rational function $\hat{g}(s)$ has degree n if and only if

$$\rho\mathbf{T}(n, n) = \rho\mathbf{T}(n+k, n+l) = n \quad \text{for every } k, l = 1, 2, \dots \quad (7.51)$$

where ρ denotes the rank.



Proof: We first show that if $\deg \hat{g}(s) = n$, then $\rho\mathbf{T}(n, n) = \rho\mathbf{T}(n+1, n) = \rho\mathbf{T}(\infty, n)$. If $\deg \hat{g}(s) = n$, then (7.49) holds, and n is the smallest integer having the property. Because of (7.49), the $(n+1)$ th row of $\mathbf{T}(n+1, n)$ can be written as a linear combination of the first n rows. Thus we have $\rho\mathbf{T}(n, n) = \rho\mathbf{T}(n+1, n)$. Again, because of (7.49), the $(n+2)$ th row of $\mathbf{T}(n+2, n)$ depends on its previous n rows and, consequently, on the first n rows. Proceeding forward, we can establish $\rho\mathbf{T}(n, n) = \rho\mathbf{T}(\infty, n)$. Now we claim $\rho\mathbf{T}(\infty, n) = n$. If not, there would be an integer $\bar{n} < n$ having the property (7.49). This contradicts the hypothesis that $\deg \hat{g}(s) = n$. Thus we have $\rho\mathbf{T}(n, n) = \rho\mathbf{T}(\infty, n) = n$. Applying (7.49) to the columns of \mathbf{T} yields (7.51).

Now we show that if (7.51) holds, then $\hat{g}(s) = h(1)s^{-1} + h(2)s^{-2} + \dots$ can be expressed as a strictly proper rational function of degree n . From the condition $\rho\mathbf{T}(n+1, \infty) = \rho\mathbf{T}(n, \infty) = n$, we can compute $\{\alpha_i, i = 1, 2, \dots, n\}$ to meet (7.49). We then use (7.48) to compute $\{\beta_i, i = 1, 2, \dots, n\}$. Hence we have

$$\begin{aligned} \hat{g}(s) &= h(1)s^{-1} + h(2)s^{-2} + h(3)s^{-3} \dots \\ &= \frac{\beta_1 s^{n-1} + \beta_2 s^{n-2} + \dots + \beta_{n-1} s + \beta_n}{s^n + \alpha_1 s^{n-1} + \alpha_2 s^{n-2} + \dots + \alpha_{n-1} s + \alpha_n} \end{aligned}$$

Because the n is the smallest integer having the property in (7.51), we have $\deg \hat{g}(s) = n$. This completes the proof of the theorem. Q.E.D.

With this preliminary, we are ready to discuss the realization problem. Consider a strictly proper transfer function $\hat{g}(s)$ expressed as

$$\hat{g}(s) = h(1)s^{-1} + h(2)s^{-2} + h(3)s^{-3} + \dots$$

If the triplet $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ is a realization of $\hat{g}(s)$, then

$$\hat{g}(s) = \mathbf{c}(\mathbf{sI} - \mathbf{A})^{-1} \mathbf{b} = \mathbf{c}[\mathbf{s}(\mathbf{I} - \mathbf{s}^{-1}\mathbf{A})]^{-1} \mathbf{b}$$

which becomes, using (3.57),

$$\hat{g}(s) = \mathbf{c}\mathbf{b}s^{-1} + \mathbf{c}\mathbf{A}\mathbf{b}s^{-2} + \mathbf{c}\mathbf{A}^2\mathbf{b}s^{-3} + \dots$$

Thus we conclude that $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ is a realization of $\hat{g}(s)$ if and only if

$$h(m) = \mathbf{c}\mathbf{A}^{m-1}\mathbf{b} \quad \text{for } m = 1, 2, \dots \quad (7.52)$$

Substituting (7.52) into the Hankel matrix $\mathbf{T}(n,n)$ yields

$$\mathbf{T}(n,n) = \begin{bmatrix} \mathbf{c}\mathbf{b} & \mathbf{c}\mathbf{Ab} & \mathbf{c}\mathbf{A}^2\mathbf{b} & \dots & \mathbf{c}\mathbf{A}^{n-1}\mathbf{b} \\ \mathbf{c}\mathbf{Ab} & \mathbf{c}\mathbf{A}^2\mathbf{b} & \mathbf{c}\mathbf{A}^3\mathbf{b} & \dots & \mathbf{c}\mathbf{A}^n\mathbf{b} \\ \mathbf{c}\mathbf{A}^2\mathbf{b} & \mathbf{c}\mathbf{A}^3\mathbf{b} & \mathbf{c}\mathbf{A}^4\mathbf{b} & \dots & \mathbf{c}\mathbf{A}^{n+1}\mathbf{b} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{c}\mathbf{A}^{n-1}\mathbf{b} & \mathbf{c}\mathbf{A}^n\mathbf{b} & \mathbf{c}\mathbf{A}^{n+1}\mathbf{b} & \dots & \mathbf{c}\mathbf{A}^{2(n-1)}\mathbf{b} \end{bmatrix}$$

which implies, as shown in (7.21),

$$\mathbf{T}(n,n) = \mathcal{O}\mathcal{C} \quad (7.53)$$

where \mathcal{O} and \mathcal{C} are, respectively, the $n \times n$ observability and controllability matrices of $(\mathbf{A}, \mathbf{b}, \mathbf{c})$. Define

$$\tilde{\mathbf{T}}(n,n) = \begin{bmatrix} h(2) & h(3) & h(4) & \dots & h(n+1) \\ h(3) & h(4) & h(5) & \dots & h(n+2) \\ h(4) & h(5) & h(6) & \dots & h(n+3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h(n+1) & h(n+2) & h(n+3) & \dots & h(2n) \end{bmatrix} \quad (7.54)$$

It is the submatrix of $\mathbf{T}(n+1,n)$ by deleting the first row or the submatrix of $\mathbf{T}(n,n+1)$ by deleting the first column. Then as (7.53), we can readily show

$$\tilde{\mathbf{T}}(n,n) = \mathcal{O}\mathbf{A}\mathcal{C} \quad (7.55)$$

Using (7.53) and (7.55), we can obtain many different realizations. We discuss here only a companion-form and a balanced-form realization.

Companion form

There are many ways to decompose $\mathbf{T}(n,n)$ into $\mathcal{O}\mathcal{C}$. The simplest is to select $\mathcal{O} = \mathbf{I}$ or $\mathcal{C} = \mathbf{I}$. If we select $\mathcal{O} = \mathbf{I}$, then (7.53) and (7.55) imply $\mathcal{C} = \mathbf{T}(n,n)$ and $\mathbf{A} = \tilde{\mathbf{T}}(n,n)\mathbf{T}^{-1}(n,n)$. The state-space equation corresponding to $\mathcal{O} = \mathbf{I}$, $\mathcal{C} = \mathbf{T}(n,n)$ and $\mathbf{A} = \tilde{\mathbf{T}}(n,n)\mathbf{T}^{-1}(n,n)$ is

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ -\alpha_1 & -\alpha_2 & -\alpha_3 & \dots & -\alpha_{n-1} & -\alpha_n \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} h(1) \\ h(2) \\ \vdots \\ h(n-1) \\ h(n) \end{bmatrix} u(t) \quad (7.56)$$

$$y(t) = [1 \ 0 \ 0 \ \dots \ 0 \ 0] \mathbf{x}(t)$$

Indeed, the first row of $\mathcal{O} = \mathbf{I}$ and the first column of $\mathcal{C} = \mathbf{T}(n,n)$ yield the \mathbf{c} and \mathbf{b} in (7.56). Instead of showing $\mathbf{A} = \tilde{\mathbf{T}}(n,n)\mathbf{T}^{-1}(n,n)$, we show

$$\mathbf{AT}(n,n) = \tilde{\mathbf{T}}(n,n) \quad (7.57)$$

Using the shifting property of the companion-form matrix in (7.56), we can readily verify

$$\mathbf{A} \begin{bmatrix} h(1) \\ h(2) \\ \vdots \\ h(n) \end{bmatrix} = \begin{bmatrix} h(2) \\ h(3) \\ \vdots \\ h(n+1) \end{bmatrix}, \quad \mathbf{A} \begin{bmatrix} h(2) \\ h(3) \\ \vdots \\ h(n+1) \end{bmatrix} = \begin{bmatrix} h(3) \\ h(4) \\ \vdots \\ h(n+2) \end{bmatrix}, \dots \quad (7.58)$$

We see that the Markov parameters of a column are shifted up one position if the column is premultiplied by \mathbf{A} . Using this property, we can readily establish (7.57). Thus $\mathcal{O} = \mathbf{I}$, $\mathcal{C} = \mathbf{T}(n,n)$ and $\mathbf{A} = \tilde{\mathbf{T}}(n,n)\mathbf{T}^{-1}(n,n)$ generate the realization in (7.56). It is a companion-form realization. Now we use (7.52) to show that (7.56) is indeed a realization. Because of the form of \mathbf{c} , $\mathbf{c}\mathbf{A}^m\mathbf{b}$ equals simply the top entry of $\mathbf{A}^m\mathbf{b}$ or

$$\mathbf{cb} = h(1), \quad \mathbf{cAb} = h(2), \quad \mathbf{cA}^2\mathbf{b} = h(3), \dots$$

Thus (7.56) is a realization of $\hat{g}(s)$. The state-space equation is always observable because $\mathcal{O} = \mathbf{I}$ has full rank. It is controllable if $\mathcal{C} = \mathbf{T}(n,n)$ has rank n .

Example 7.5.1 Consider

$$\begin{aligned} \hat{g}(s) &= \frac{4s^2 - 2s - 6}{2s^4 + 2s^3 + 2s^2 + 3s + 1} \\ &= 0 \cdot s^{-1} + 2s^{-2} - 3s^{-3} - 2s^{-4} + 2s^{-5} + 3.5s^{-6} + \dots \end{aligned} \quad (7.59)$$

We form $\mathbf{T}(4,4)$ and compute its rank. The rank is 3; thus $\hat{g}(s)$ in (7.59) has degree 3 and its numerator and denominator have a common factor of degree 1. There is no need to cancel first the common factor in the expansion in (7.59). From the preceding derivation, we have

$$\mathbf{A} = \begin{bmatrix} 2 & -3 & -2 \\ -3 & -2 & 2 \\ -2 & 2 & 3.5 \end{bmatrix} \begin{bmatrix} 0 & 2 & -3 \\ 2 & -3 & -2 \\ -3 & -2 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -0.5 & -1 & 0 \end{bmatrix} \quad (7.60)$$

and

$$\mathbf{b} = [0 \ 2 \ -3]', \quad \mathbf{c} = [1 \ 0 \ 0]$$

The triplet $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ is a minimal realization of $\hat{g}(s)$ in (7.59).

We mention that the matrix \mathbf{A} in (7.60) can be obtained without computing $\tilde{\mathbf{T}}(n,n)\mathbf{T}^{-1}(n,n)$. Using (7.49) we can show

$$\mathbf{T}(3,4)\mathbf{a} := \begin{bmatrix} 0 & 2 & -3 & -2 \\ 2 & -3 & -2 & 2 \\ -3 & -2 & 2 & 3.5 \end{bmatrix} \begin{bmatrix} \alpha_3 \\ \alpha_2 \\ \alpha_1 \\ 1 \end{bmatrix} = \mathbf{0}$$

Thus \mathbf{a} is a null vector of $\mathbf{T}(3,4)$. The MATLAB function

```
t=[0 2 -3 -2; 2 -3 -2 2; -3 -2 2 3.5]; a=null(t)
```

yields $\mathbf{a} = [0.3333 \ 0.6667 \ -0.0000 \ 0.6667]'$. We normalize the last entry of \mathbf{a} to 1 by typing $\mathbf{a}/\mathbf{a}(4)$ which yields $[0.5 \ 1 \ -0 \ 1]'$. The first three entries, with sign reversed, are the last row of \mathbf{A} .

Balanced form

Next we discuss a different decomposition of $\mathbf{T}(n,n) = \mathcal{O}\mathcal{C}$ which will yield a realization with the property

$$\mathcal{C}\mathcal{C}' = \mathcal{O}'\mathcal{O}$$

First we use singular-value decomposition to express $\mathbf{T}(n,n)$ as

$$\mathbf{T}(n,n) = \mathbf{K}\Lambda\mathbf{L}' = \mathbf{K}\Lambda^{1/2}\Lambda^{1/2}\mathbf{L}' \quad (7.61)$$

where \mathbf{K} and \mathbf{L} are orthogonal matrices and $\Lambda^{1/2}$ is diagonal with the singular values of $\mathbf{T}(n,n)$ on the diagonal. Let us select

$$\mathcal{O} = \mathbf{K}\Lambda^{1/2} \quad \text{and} \quad \mathcal{C} = \Lambda^{1/2}\mathbf{L}' \quad (7.62)$$

Then we have

$$\mathcal{O}^{-1} = \Lambda^{-1/2}\mathbf{K}' \quad \text{and} \quad \mathcal{C}^{-1} = \mathbf{L}\Lambda^{-1/2} \quad (7.63)$$

For this selection of \mathcal{C} and \mathcal{O} , the matrices triplet

$$\mathbf{A} = \mathcal{O}^{-1}\tilde{\mathbf{T}}(n,n)\mathcal{C}^{-1} \quad (7.64)$$

$$\mathbf{b} = \text{first column of } \mathcal{C} \quad (7.65)$$

$$\mathbf{c} = \text{first row of } \mathcal{O} \quad (7.66)$$

comprise a minimal realization of $\hat{g}(s)$. For this realization, we have

$$\mathcal{C}\mathcal{C}' = \Lambda^{1/2}\mathbf{L}'\mathbf{L}\Lambda^{1/2} = \Lambda$$

and

$$\mathcal{O}'\mathcal{O} = \Lambda^{1/2}\mathbf{K}'\mathbf{K}\Lambda^{1/2} = \Lambda = \mathcal{C}\mathcal{C}'$$

Thus it is called a *balanced realization*. This balanced realization is different from the balanced realization discussed in Section 7.4. It is not clear the relationships between them.

Example 7.5.2 Consider the transfer function in Example 7.5.1. Now we will find a balanced realization from Hankel matrices. We type

```
t=[0 2 -3;2 -3 -2;-3 -2 2];tt=[2 -3 -2;-3 -2 2;-2 2 3.5];
[k,s,l]=svd(t);
s1=sqrt(s);
O=k*s1;C=s1*l';
a=inv(O)*tt*inv(C),
b=[C(1,1);C(2,1);C(3,1)],c=[O(1,1) O(1,2) O(1,3)]
```

This yields the following balanced realization

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0.4003 & -1.0024 & 0.4805 \\ 1.0024 & -0.3121 & -0.3209 \\ -0.4805 & -0.3209 & -0.0882 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} -1.2883 \\ 0.7303 \\ 1.0614 \end{bmatrix} u(t)$$

$$y(t) = [-1.2883 \ -0.7303 \ -1.0614] \mathbf{x}(t) + 0 \cdot u(t)$$

To check the correctness of this result, we type $[n, d] = ss2tf(a, b, c, 0)$, which yields

$$\hat{g}(s) = \frac{2s - 3}{s^3 + s + 0.5}$$

This equals $\hat{g}(s)$ in (7.59) after canceling common factor $2(s + 1)$.

7.6 Degree of Transfer Matrices

The degree of a proper rational function is defined as the degree of its denominator if its denominator and numerator are coprime. We can also define a degree for proper rational matrices. One way is to express a proper rational matrix $\hat{\mathbf{G}}(s)$ as $\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)$, where $\mathbf{N}(s)$ and $\mathbf{D}(s)$ are polynomial matrices and then develop the concept of coprimeness for two polynomial matrices. This will be discussed in latter sections. In this section, we will define the degree directly from a rational matrix. This approach is simpler and will reveal the significant difference between scalar and matrix rational functions. However, the result can be used only for discussing minimal realizations but not for developing design methods.

Consider a proper rational matrix $\hat{\mathbf{G}}(s)$. It is assumed that every entry of $\hat{\mathbf{G}}(s)$ is a proper coprime fraction; that is, its numerator and denominator have no common root and its numerator's degree equals at most its denominator's degree. This will be a standing assumption throughout the remainder of this text. We call the determinant of any $r \times r$ submatrix of $\hat{\mathbf{G}}(s)$ a minor of order r .

Definition 7.1 The *characteristic polynomial* of a proper rational matrix $\hat{\mathbf{G}}(s)$ is defined as the least common denominator of all minors of $\hat{\mathbf{G}}(s)$. The degree of the characteristic polynomial is defined as the *McMillan degree* or, simply, the *degree* of $\hat{\mathbf{G}}(s)$ and is denoted by $\delta\hat{\mathbf{G}}(s)$.

Example 7.6.1 Consider the rational matrices

$$\hat{\mathbf{G}}_1(s) = \begin{bmatrix} \frac{1}{s+1} & \frac{1}{s+1} \\ \frac{1}{s+1} & \frac{1}{s+1} \end{bmatrix}, \quad \hat{\mathbf{G}}_2(s) = \begin{bmatrix} \frac{2}{s+1} & \frac{1}{s+1} \\ \frac{1}{s+1} & \frac{1}{s+1} \end{bmatrix}$$

The matrix $\hat{\mathbf{G}}_1(s)$ has $1/(s+1)$, $1/(s+1)$, $1/(s+1)$, and $1/(s+1)$ as its minors of order 1, and 0 as its minor of order 2. Thus the characteristic polynomial of $\hat{\mathbf{G}}_1(s)$ is $s+1$ and $\delta\hat{\mathbf{G}}_1(s) = 1$. The matrix $\hat{\mathbf{G}}_2(s)$ has $1/(s+1)$, $1/(s+1)$, $1/(s+1)$, and $1/(s+1)$ as its minors of order 1, and $1/(s+1)^2$ as its minor of order 2. Thus the characteristic polynomial of $\hat{\mathbf{G}}_2(s)$ is $(s+1)^2$ and $\delta\hat{\mathbf{G}}_2(s) = 2$.

From this example, we see that the characteristic polynomial of $\hat{\mathbf{G}}(s)$ is, in general, different from the denominator of the determinant of $\hat{\mathbf{G}}(s)$ [if $\hat{\mathbf{G}}(s)$ is square] and different from the least common denominator of all entries of $\hat{\mathbf{G}}(s)$.

Example 7.6.2 Consider the 2×3 rational matrix

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} \frac{s}{s+1} & \frac{1}{(s+1)(s+2)} & \frac{1}{s+3} \\ \frac{-1}{s+1} & \frac{1}{(s+1)(s+2)} & \frac{1}{s} \end{bmatrix}$$

Its minors of order 1 are the six entries of $\hat{\mathbf{G}}(s)$. The matrix has the following three minors of order 2:

$$\begin{aligned} \frac{s}{(s+1)^2(s+2)} + \frac{1}{(s+1)^2(s+2)} &= \frac{s+1}{(s+1)^2(s+2)} = \frac{1}{(s+1)(s+2)} \\ \frac{s}{s+1} \cdot \frac{1}{s} + \frac{1}{(s+1)(s+3)} &= \frac{s+4}{(s+1)(s+3)} \\ \frac{1}{(s+1)(s+2)s} - \frac{1}{(s+1)(s+2)(s+3)} &= \frac{3}{s(s+1)(s+2)(s+3)} \end{aligned}$$

The least common denominator of all these minors is $s(s+1)(s+2)(s+3)$. Thus the degree of $\hat{\mathbf{G}}(s)$ is 4.

In computing the characteristic polynomial, every minor must be reduced to a coprime fraction as we did in the preceding example; otherwise, we will obtain an erroneous result. We discuss two special cases. If $\hat{\mathbf{G}}(s)$ is $1 \times p$ or $q \times 1$, then there are no minors of order 2 or higher. Thus the characteristic polynomial equals the least common denominator of all entries of $\hat{\mathbf{G}}(s)$. In particular, if $\hat{\mathbf{G}}(s)$ is scalar, then the characteristic polynomial equals its denominator. If every entry of $q \times p$ $\hat{\mathbf{G}}(s)$ has poles that differ from those of all other entries such as

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} \frac{1}{(s+1)^2(s+2)} & \frac{s+2}{s^2} \\ \frac{s-2}{s+3} & \frac{s}{(s+5)(s-3)} \end{bmatrix}$$

then its minors contain no poles with multiplicities higher than those of each entry. Thus the characteristic polynomial equals the product of the denominators of all entries of $\hat{\mathbf{G}}(s)$.

To conclude this section, we mention two important properties. Let $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ be a controllable and observable realization of $\hat{\mathbf{G}}(s)$. Then we have

- Monic least common denominator of all *minors* of $\hat{\mathbf{G}}(s)$ = characteristic polynomial of \mathbf{A} .
- Monic least common denominator of all *entries* of $\hat{\mathbf{G}}(s)$ = minimal polynomial of \mathbf{A} .

For their proofs, see Reference 4, pp. 302–304.

7.7 Minimal Realizations—Matrix Case

We introduced in Section 7.2.1 minimal realizations for scalar transfer functions. Now we discuss the matrix case.

► Theorem 7.M2

A state-space equation $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ is a minimal realization of a proper rational matrix $\hat{\mathbf{G}}(s)$ if and only if (\mathbf{A}, \mathbf{B}) is controllable and (\mathbf{A}, \mathbf{C}) is observable or if and only if

$$\dim \mathbf{A} = \deg \hat{\mathbf{G}}(s)$$

⇒ **Proof:** The proof of the first part is similar to the proof of Theorem 7.2. If (\mathbf{A}, \mathbf{B}) is not controllable or if (\mathbf{A}, \mathbf{C}) is not observable, then the state-space equation is zero-state equivalent to a lesser-dimensional state-space equation and is not minimal. If $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ is of dimension n and is controllable and observable, and if the \bar{n} -dimensional state-space equation $(\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}}, \bar{\mathbf{D}})$, with $\bar{n} < n$, is a realization of $\hat{\mathbf{G}}(s)$. Then Theorem 4.1 implies $\mathbf{D} = \bar{\mathbf{D}}$ and

$$\mathbf{C}\mathbf{A}^m\mathbf{B} = \bar{\mathbf{C}}\bar{\mathbf{A}}^m\bar{\mathbf{B}} \quad \text{for } m = 0, 1, 2, \dots$$

Thus, as in (7.22), we have

$$\mathcal{OC} = \bar{\mathcal{O}}_n \bar{\mathcal{C}}_n$$

Note that \mathcal{O} , \mathcal{C} , $\bar{\mathcal{O}}_n$, and $\bar{\mathcal{C}}_n$ are, respectively, $nq \times n$, $n \times np$, $nq \times \bar{n}$, and $\bar{n} \times np$ matrices. Using the following Sylvester inequality

$$\rho(\mathcal{O}) + \rho(\mathcal{C}) - n \leq \rho(\mathcal{OC}) \leq \min(\rho(\mathcal{O}), \rho(\mathcal{C}))$$

which is proved in Reference 6, p. 31, and $\rho(\mathcal{O}) = \rho(\mathcal{C}) = n$, we have $\rho(\mathcal{OC}) = n$. Similarly, we have $\rho(\bar{\mathcal{O}}_n \bar{\mathcal{C}}_n) = \bar{n} < n$. This contradicts $\rho(\mathcal{OC}) = \rho(\bar{\mathcal{O}}_n \bar{\mathcal{C}}_n)$. Thus every controllable and observable state-space equation is a minimal realization.

To show that $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ is minimal if and only if $\dim \mathbf{A} = \deg \hat{\mathbf{G}}(s)$ is much more complex and will be established in the remainder of this chapter. Q.E.D.

► Theorem 7.M3

All minimal realizations of $\hat{\mathbf{G}}(s)$ are equivalent.



Proof: The proof follows closely the proof of Theorem 7.3. Let $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ and $(\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}}, \bar{\mathbf{D}})$ be any two n -dimensional minimal realizations of a $q \times p$ proper rational matrix $\hat{\mathbf{G}}(s)$. As in (7.23) and (7.24), we have

$$\mathcal{O}\mathcal{C} = \bar{\mathcal{O}}\bar{\mathcal{C}} \quad (7.67)$$

and

$$\mathcal{O}\mathbf{A}\mathcal{C} = \bar{\mathcal{O}}\bar{\mathbf{A}}\bar{\mathcal{C}} \quad (7.68)$$

In the scalar case, $\mathcal{O}, \mathcal{C}, \bar{\mathcal{O}}$, and $\bar{\mathcal{C}}$ are all $n \times n$ nonsingular matrices and their inverses are well-defined. Here \mathcal{O} and $\bar{\mathcal{O}}$ are $nq \times n$ matrices of rank n ; \mathcal{C} and $\bar{\mathcal{C}}$ are $n \times np$ matrices of rank n . They are not square and their inverses are not defined. Let us define the $n \times nq$ matrix

$$\mathcal{O}^+ := (\mathcal{O}'\mathcal{O})^{-1}\mathcal{O}' \quad (7.69)$$

Because \mathcal{O}' is $n \times nq$ and \mathcal{O} is $nq \times n$, the matrix $\mathcal{O}'\mathcal{O}$ is $n \times n$ and is, following Theorem 3.8, nonsingular. Clearly, we have

$$\mathcal{O}^+\mathcal{O} = (\mathcal{O}'\mathcal{O})^{-1}\mathcal{O}'\mathcal{O} = \mathbf{I}$$

Thus \mathcal{O}^+ is called the *pseudoinverse* or *left inverse* of \mathcal{O} . Note that $\mathcal{O}\mathcal{O}^+$ is $nq \times nq$ and does not equal a unit matrix. Similarly, we define

$$\mathcal{C}^+ := \mathcal{C}'(\mathcal{C}\mathcal{C}')^{-1} \quad (7.70)$$

It is an $np \times n$ matrix and has the property

$$\mathcal{C}\mathcal{C}^+ = \mathcal{C}\mathcal{C}'(\mathcal{C}\mathcal{C}')^{-1} = \mathbf{I}$$

Thus \mathcal{C}^+ is called the *pseudoinverse* or *right inverse* of \mathcal{C} . In the scalar case, the equivalence transformation is defined in (7.25) as $\mathbf{P} = \bar{\mathcal{O}}^{-1}\mathcal{O} = \bar{\mathcal{C}}\mathcal{C}^{-1}$. Now we replace inverses by pseudoinverses to yield

$$\mathbf{P} := \bar{\mathcal{O}}^+\mathcal{O} = (\bar{\mathcal{O}}'\bar{\mathcal{O}})^{-1}\bar{\mathcal{O}}'\mathcal{O} \quad (7.71)$$

$$= \bar{\mathcal{C}}\mathcal{C}^+ = \bar{\mathcal{C}}\mathcal{C}'(\mathcal{C}\mathcal{C}')^{-1} \quad (7.72)$$

This equality can be directly verified by premultiplying $(\bar{\mathcal{O}}'\bar{\mathcal{O}})$ and postmultiplying $(\mathcal{C}\mathcal{C}')$ and then using (7.67). The inverse of \mathbf{P} in the scalar case is $\mathbf{P}^{-1} = \mathcal{O}^{-1}\bar{\mathcal{O}} = \mathcal{C}\mathcal{C}^{-1}$. In the matrix case, it becomes

$$\mathbf{P}^{-1} := \mathcal{O}^+\bar{\mathcal{O}} = (\mathcal{O}'\mathcal{O})^{-1}\mathcal{O}'\bar{\mathcal{O}} \quad (7.73)$$

$$= \mathcal{C}\bar{\mathcal{C}}^+ = \mathcal{C}\bar{\mathcal{C}}'(\bar{\mathcal{C}}\bar{\mathcal{C}}')^{-1} \quad (7.74)$$

This again can be verified using (7.67). From $\bar{\mathcal{O}}\bar{\mathcal{C}} = \mathcal{O}\mathcal{C}$, we have

$$\bar{\mathcal{C}} = (\bar{\mathcal{O}}'\bar{\mathcal{O}})^{-1}\bar{\mathcal{O}}'\mathcal{O}\mathcal{C} = \mathbf{P}\mathbf{C}$$

$$\bar{\mathcal{O}} = \mathcal{O}\mathcal{C}\bar{\mathcal{C}}'(\bar{\mathcal{C}}\bar{\mathcal{C}}')^{-1} = \mathcal{O}\mathbf{P}^{-1}$$

Their first p columns and first q rows are $\bar{\mathbf{B}} = \mathbf{PB}$ and $\bar{\mathbf{C}} = \mathbf{CP}^{-1}$. The equation $\bar{\mathcal{O}}\bar{\mathbf{A}}\bar{\mathcal{C}} = \mathcal{O}\mathbf{AC}$ implies

$$\bar{\mathbf{A}} = (\bar{\mathcal{O}}'\bar{\mathcal{O}})^{-1}\bar{\mathcal{O}}'\mathcal{O}\mathbf{AC}\bar{\mathcal{C}}'(\bar{\mathcal{C}}\bar{\mathcal{C}}')^{-1} = \mathbf{PAP}^{-1}$$

This shows that all minimal realizations of the same transfer matrix are equivalent. Q.E.D.

We see from the proof of Theorem 7.M3 that the results in the scalar case can be directly extended to the matrix case if inverses are replaced by pseudoinverses. In MATLAB, the function `pinv` generates the pseudoinverse. For the balanced realization discussed in Section 7.4, because the controllability and observability Gramians are square for the SISO as well as MIMO cases, all discussion in Section 7.4 is applicable to the MIMO case without any modification.

Example 7.7.1 Consider the transfer matrix in Example 4.5.3 or

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} \frac{4s-10}{2s+1} & \frac{3}{s+2} \\ \frac{1}{(2s+1)(s+2)} & \frac{1}{(s+2)^2} \end{bmatrix} \quad (7.75)$$

Its characteristic polynomial can be computed as $(2s + 1)(s + 2)^2$. Thus the rational matrix has degree 3. Its six-dimensional realization in (4.47) and four-dimensional realization in (4.52) are clearly not minimal realizations. They can be reduced to minimal realizations by calling the MATLAB function `minreal`. For example, for (4.47), we type

```
a=[-4.5 0 -6 0 -2 0;0 -4.5 0 -6 0 -2;1 0 0 0 0 0;...
    0 1 0 0 0 0;0 0 1 0 0 0;0 0 0 1 0 0];
b=[1 0;0 1;0 0;0 0;0 0;0 0];
c=[-6 3 -24 7.5 -24 3;0 1 0.5 1.5 1 0.5];d=[2 0;0 0];
[am,bm,cm,dm]=minreal(a,b,c,d)
```

then it will yield

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -1.3387 & 0.2212 & -1.6000 \\ 0.25335 & -1.1682 & 4.8352 \\ 0.0036 & 0.0012 & -1.9931 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} -0.2666 & 0.2026 \\ 0.2513 & -0.6125 \\ 0.0004 & 0.3473 \end{bmatrix} \mathbf{u}(t)$$

$$\mathbf{y}(t) = \begin{bmatrix} 32.7210 & 10.8198 & 8.6323 \\ -0.8143 & -0.8663 & 1.8266 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \mathbf{u}(t)$$

Its dimension equals the degree of $\hat{\mathbf{G}}(s)$; thus it is controllable and observable and is a minimal realization of $\hat{\mathbf{G}}(s)$ in (7.75).

7.8 Matrix Polynomial Fractions

Two polynomials are said to be *coprime* if they have no common root. In order to extend this concept to the matrix case, we reformulate the definition. A polynomial $R(s)$ is called a common factor or a *common divisor* of $D(s)$ and $N(s)$ if both $D(s)$

and $N(s)$ can be divided by $R(s)$ without any remainder. A polynomial $R(s)$ is called a *greatest common divisor* (gcd) of $D(s)$ and $N(s)$ if (1) it is a common divisor of $D(s)$ and $N(s)$ and (2) it can be divided without remainder by every other common divisor of $D(s)$ and $N(s)$. Note that if $R(s)$ is a gcd, so is $\alpha R(s)$ for any nonzero constant α . Thus gcd's are not unique.⁶ In terms of gcd, the polynomials $D(s)$ and $N(s)$ are coprime if their gcd $R(s)$ is a nonzero constant, a polynomial of degree 0; they are not coprime if their gcd has degree 1 or higher. If $R(s)$ is a gcd of $N(s)$ and $D(s)$ of degree 1 or higher, we can write $N(s) = \bar{N}(s)R(s)$ and $D(s) = \bar{D}(s)R(s)$. Then the polynomial fraction $\bar{N}(s)/\bar{D}(s) = \bar{N}(s)\bar{D}^{-1}(s)$ is coprime.

For any two polynomials $N(s)$ and $D(s)$, we have $N(s)D^{-1}(s) = D^{-1}(s)N(s)$. For polynomial matrices, the situation is more complex. First, the inverse of a polynomial matrix is defined only if it is square. Moreover, for any two square polynomial matrices $\mathbf{N}(s)$ and $\mathbf{D}(s)$, generally we have $\mathbf{N}(s)\mathbf{D}^{-1}(s) \neq \mathbf{D}^{-1}(s)\mathbf{N}(s)$. Because of this noncommutative property, the order or position of polynomial matrices becomes critical in the subsequent discussion.

Every $q \times p$ proper rational matrix $\hat{\mathbf{G}}(s)$ can be expressed as

$$\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s) \quad (7.76)$$

where $\mathbf{N}(s)$ and $\mathbf{D}(s)$ are $q \times p$ and $p \times p$ polynomial matrices. For example, the 2×3 rational matrix in Example 7.6.2 can be expressed as

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} s & 1 & s \\ -1 & 1 & s+3 \end{bmatrix} \begin{bmatrix} s+1 & 0 & 0 \\ 0 & (s+1)(s+2) & 0 \\ 0 & 0 & s(s+3) \end{bmatrix}^{-1} \quad (7.77)$$

The three diagonal entries of $\mathbf{D}(s)$ in (7.77) are the least common denominators of the three columns of $\hat{\mathbf{G}}(s)$. The fraction in (7.76) or (7.77) is called a *right polynomial fraction* or, simply, *right fraction*. Dual to (7.76), the expression

$$\hat{\mathbf{G}}(s) = \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)$$

where $\bar{\mathbf{D}}(s)$ and $\bar{\mathbf{N}}(s)$ are $q \times q$ and $q \times p$ polynomial matrices, is called a *left polynomial fraction* or, simply, *left fraction*.

Let $\mathbf{R}(s)$ be any $p \times p$ polynomial matrix. Then we have

$$\begin{aligned} \hat{\mathbf{G}}(s) &= [\mathbf{N}(s)\mathbf{R}(s)][\mathbf{D}(s)\mathbf{R}(s)]^{-1} \\ &= \mathbf{N}(s)\mathbf{R}(s)\mathbf{R}^{-1}(s)\mathbf{D}^{-1}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s) \end{aligned}$$

Thus right fractions are not unique. So are left fractions. We introduce in the following right coprime fractions.

Consider $\mathbf{A}(s) = \mathbf{B}(s)\mathbf{C}(s)$, where $\mathbf{A}(s)$, $\mathbf{B}(s)$, and $\mathbf{C}(s)$ are polynomial of compatible orders. We call $\mathbf{C}(s)$ a *right divisor* of $\mathbf{A}(s)$ and $\mathbf{A}(s)$ a *left multiple* of $\mathbf{C}(s)$. Similarly, we call $\mathbf{B}(s)$ a *left divisor* of $\mathbf{A}(s)$ and $\mathbf{A}(s)$ a *right multiple* of $\mathbf{B}(s)$.

⁶If we require $R(s)$ to be monic, then the gcd is unique.

Consider two polynomial matrices $\mathbf{D}(s)$ and $\mathbf{N}(s)$ with the same number of columns. A square polynomial matrix $\mathbf{R}(s)$ is called a common *right divisor* of $\mathbf{D}(s)$ and $\mathbf{N}(s)$ if there exist polynomial matrices $\hat{\mathbf{D}}(s)$ and $\hat{\mathbf{N}}(s)$ such that

$$\mathbf{D}(s) = \hat{\mathbf{D}}(s)\mathbf{R}(s) \quad \text{and} \quad \mathbf{N}(s) = \hat{\mathbf{N}}(s)\mathbf{R}(s)$$

Note that $\mathbf{D}(s)$ and $\mathbf{N}(s)$ can have different numbers of rows.

Definition 7.2 A square polynomial matrix $\mathbf{M}(s)$ is called a *unimodular matrix* if its determinant is nonzero and independent of s .

The following polynomial matrices are all unimodular matrices:

$$\begin{bmatrix} 2s & s^2 + s + 1 \\ 2 & s + 1 \end{bmatrix}, \quad \begin{bmatrix} -2 & s^{10} + s + 1 \\ 0 & 3 \end{bmatrix}, \quad \begin{bmatrix} s & s + 1 \\ s - 1 & s \end{bmatrix}$$

Products of unimodular matrices are clearly unimodular matrices. Consider

$$\det \mathbf{M}(s) \det \mathbf{M}^{-1}(s) = \det[\mathbf{M}(s)\mathbf{M}^{-1}(s)] = \det \mathbf{I} = 1$$

which implies that if the determinant of $\mathbf{M}(s)$ is a nonzero constant, so is the determinant of $\mathbf{M}^{-1}(s)$. Thus the inverse of a unimodular matrix $\mathbf{M}(s)$ is again a unimodular matrix.

Definition 7.3 A square polynomial matrix $\mathbf{R}(s)$ is a *greatest common right divisor* (gcrd) of $\mathbf{D}(s)$ and $\mathbf{N}(s)$ if (i) $\mathbf{R}(s)$ is a common right divisor of $\mathbf{D}(s)$ and $\mathbf{N}(s)$ and (ii) $\mathbf{R}(s)$ is a left multiple of every common right divisor of $\mathbf{D}(s)$ and $\mathbf{N}(s)$. If a gcrd is a unimodular matrix, then $\mathbf{D}(s)$ and $\mathbf{N}(s)$ are said to be *right coprime*.

Dual to this definition, a square polynomial matrix $\bar{\mathbf{R}}(s)$ is a *greatest common left divisor* (gcld) of $\bar{\mathbf{D}}(s)$ and $\bar{\mathbf{N}}(s)$ if (i) $\bar{\mathbf{R}}(s)$ is a common left divisor of $\bar{\mathbf{D}}(s)$ and $\bar{\mathbf{N}}(s)$ and (ii) $\bar{\mathbf{R}}(s)$ is a right multiple of every common left divisor of $\bar{\mathbf{D}}(s)$ and $\bar{\mathbf{N}}(s)$. If a gcld is a unimodular matrix, then $\bar{\mathbf{D}}(s)$ and $\bar{\mathbf{N}}(s)$ are said to be *left coprime*.

Definition 7.4 Consider a proper rational matrix $\hat{\mathbf{G}}(s)$ factored as

$$\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s) = \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)$$

where $\mathbf{N}(s)$ and $\mathbf{D}(s)$ are right coprime, and $\bar{\mathbf{N}}(s)$ and $\bar{\mathbf{D}}(s)$ are left coprime. Then the characteristic polynomial of $\hat{\mathbf{G}}(s)$ is defined as

$$\det \mathbf{D}(s) \quad \text{or} \quad \det \bar{\mathbf{D}}(s)$$

and the degree of $\hat{\mathbf{G}}(s)$ is defined as

$$\deg \hat{\mathbf{G}}(s) = \deg \det \mathbf{D}(s) = \deg \det \bar{\mathbf{D}}(s)$$

Consider

$$\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s) = [\mathbf{N}(s)\mathbf{R}(s)][\mathbf{D}(s)\mathbf{R}(s)]^{-1} \quad (7.78)$$

Clearly we have

$$\det \mathbf{D}(s) = \det[\mathbf{D}(s)\mathbf{R}(s)] = \det \mathbf{D}(s) \det \mathbf{R}(s)$$

which implies

$$\deg \det \mathbf{D}(s) = \deg \det \mathbf{D}(s) + \deg \det \mathbf{R}(s)$$

Thus we may conclude that if $\mathbf{N}(s)\mathbf{D}^{-1}(s)$ is a coprime fraction, then $\mathbf{D}(s)$ has the smallest possible determinantal degree and the degree is defined as the degree of the transfer matrix. Thus a coprime fraction can also be defined as a polynomial matrix fraction with the smallest denominator's determinantal degree. From (7.78), we can see that coprime fractions are not unique; they can differ by a unimodular matrix. Every pole of every entry of $\hat{\mathbf{G}}(s)$ was defined as a pole of $\hat{\mathbf{G}}(s)$. If we express $\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)$, then every root of $\det \mathbf{D}(s)$ is a pole of $\hat{\mathbf{G}}(s)$ if and only if $\mathbf{N}(s)$ and $\mathbf{D}(s)$ are coprime. This is identical to the scalar case.

We have introduced Definitions 7.1 and 7.4 to define the degree of rational matrices. Their equivalence can be established by using the Smith–McMillan form and will not be discussed here. The interested reader is referred to, for example, Reference 13.

7.8.1 Column and Row Reducedness

In order to apply Definition 7.4, we must compute the determinant of a polynomial matrix. This can be avoided if coprime fractions have some additional property as we will discuss next.

The degree of a polynomial vector is defined as the highest power of s in all entries of the vector. Consider a polynomial matrix $\mathbf{M}(s)$. We define

$$\delta_{ci}\mathbf{M}(s) = \text{degree of } i\text{th column of } \mathbf{M}(s)$$

$$\delta_{ri}\mathbf{M}(s) = \text{degree of } i\text{th row of } \mathbf{M}(s)$$

and call δ_{ci} the *column degree* and δ_{ri} the *row degree*. For example, the matrix

$$\mathbf{M}(s) = \begin{bmatrix} s+1 & s^3 - 2s + 5 & -1 \\ s-1 & s^2 & 0 \end{bmatrix}$$

has $\delta_{c1} = 1$, $\delta_{c2} = 3$, $\delta_{c3} = 0$, $\delta_{r1} = 3$, and $\delta_{r2} = 2$.

Definition 7.5 A nonsingular polynomial matrix $\mathbf{M}(s)$ is *column reduced* if

$$\deg \det \mathbf{M}(s) = \text{sum of all column degrees}$$

It is *row reduced* if

$$\deg \det \mathbf{M}(s) = \text{sum of all row degrees}$$

A matrix can be column reduced but not row reduced and vice versa. For example, the matrix

$$\mathbf{M}(s) = \begin{bmatrix} 3s^2 + 2s & 2s + 1 \\ s^2 + s - 3 & s \end{bmatrix} \quad (7.79)$$

has determinant $s^3 - s^2 + 5s + 3$. Its degree equals the sum of its column degrees 2 and 1. Thus $\mathbf{M}(s)$ in (7.79) is column reduced. The row degrees of $\mathbf{M}(s)$ are 2 and 2; their sum is larger than 3. Thus $\mathbf{M}(s)$ is not row reduced. A diagonal polynomial matrix is always both column and row reduced. If a square polynomial matrix is not column reduced, then the degree of its determinant is less than the sum of its column degrees.

Let $\delta_{ci}\mathbf{M}(s) = k_{ci}$ and define $\mathbf{H}_c(s) = \text{diag}(s^{k_{c1}}, s^{k_{c2}}, \dots)$. Then the polynomial matrix $\mathbf{M}(s)$ can be expressed as

$$\mathbf{M}(s) = \mathbf{M}_{hc}\mathbf{H}_c(s) + \mathbf{M}_{lc}(s) \quad (7.80)$$

For example, the $\mathbf{M}(s)$ in (7.79) has column degrees 2 and 1 and can be expressed as

$$\mathbf{M}(s) = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} s^2 & 0 \\ 0 & s \end{bmatrix} + \begin{bmatrix} 2s & 1 \\ s-3 & 0 \end{bmatrix}$$

The constant matrix \mathbf{M}_{hc} is called the *column-degree coefficient matrix*; its i th column is the coefficients of the i th column of $\mathbf{M}(s)$ associated with $s^{k_{ci}}$. The polynomial matrix $\mathbf{M}_{lc}(s)$ contains the remaining terms and its i th column has degree less than k_{ci} . If $\mathbf{M}(s)$ is expressed as in (7.80), then it can be verified that $\mathbf{M}(s)$ is column reduced if and only if its column-degree coefficient matrix \mathbf{M}_{hc} is nonsingular. Dual to (7.80), we can express $\mathbf{M}(s)$ as

$$\mathbf{M}(s) = \mathbf{H}_r(s)\mathbf{M}_{hr} + \mathbf{M}_{lr}(s)$$

where $\delta_{ri}\mathbf{M}(s) = k_{ri}$ and $\mathbf{H}_r(s) = \text{diag}(s^{k_{r1}}, s^{k_{r2}}, \dots)$. The matrix \mathbf{M}_{hr} is called the *row-degree coefficient matrix*. Then $\mathbf{M}(s)$ is row reduced if and only if \mathbf{M}_{hr} is nonsingular.

Consider $\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)$. If $\hat{\mathbf{G}}(s)$ is strictly proper, then $\delta_{ci}\mathbf{N}(s) < \delta_{ci}\mathbf{D}(s)$, for $i = 1, 2, \dots, p$, that is, the column degrees of $\mathbf{N}(s)$ are less than the corresponding column degrees of $\mathbf{D}(s)$. If $\hat{\mathbf{G}}(s)$ is proper, then $\delta_{ci}\mathbf{N}(s) \leq \delta_{ci}\mathbf{D}(s)$, for $i = 1, 2, \dots, p$. The converse however is not necessarily true. For example, consider

$$\mathbf{N}(s)\mathbf{D}^{-1}(s) = [1 \ 2] \begin{bmatrix} s^2 & s-1 \\ s+1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} -2s-1 & 2s^2-s+1 \\ 1 & 1 \end{bmatrix}$$

Although $\delta_{ci}\mathbf{N}(s) < \delta_{ci}\mathbf{D}(s)$ for $i = 1, 2$, $\mathbf{N}(s)\mathbf{D}^{-1}(s)$ is not strictly proper. The reason is that $\mathbf{D}(s)$ is not column reduced.

► Theorem 7.8

Let $\mathbf{N}(s)$ and $\mathbf{D}(s)$ be $q \times p$ and $p \times p$ polynomial matrices, and let $\mathbf{D}(s)$ be column reduced. Then the rational matrix $\mathbf{N}(s)\mathbf{D}^{-1}(s)$ is proper (strictly proper) if and only if

$$\delta_{ci}\mathbf{N}(s) \leq \delta_{ci}\mathbf{D}(s) \quad [\delta_{ci}\mathbf{N}(s) < \delta_{ci}\mathbf{D}(s)]$$

for $i = 1, 2, \dots, p$.



Proof: The necessity part of the theorem follows from the preceding example. We show the sufficiency. Following (7.80), we express

$$\begin{aligned} \mathbf{D}(s) &= \mathbf{D}_{hc}\mathbf{H}_c(s) + \mathbf{D}_{lc}(s) = [\mathbf{D}_{hc} + \mathbf{D}_{lc}(s)\mathbf{H}_c^{-1}(s)]\mathbf{H}_c(s) \\ \mathbf{N}(s) &= \mathbf{N}_{hc}\mathbf{H}_c(s) + \mathbf{N}_{lc}(s) = [\mathbf{N}_{hc} + \mathbf{N}_{lc}(s)\mathbf{H}_c^{-1}(s)]\mathbf{H}_c(s) \end{aligned}$$

Then we have

$$\hat{\mathbf{G}}(s) := \mathbf{N}(s)\mathbf{D}^{-1}(s) = [\mathbf{N}_{hc} + \mathbf{N}_{lc}(s)\mathbf{H}_c^{-1}(s)][\mathbf{D}_{hc} + \mathbf{D}_{lc}(s)\mathbf{H}_c^{-1}(s)]^{-1}$$

Because $\mathbf{D}_{lc}(s)\mathbf{H}_c^{-1}(s)$ and $\mathbf{N}_{lc}(s)\mathbf{H}_c^{-1}(s)$ both approach zero as $s \rightarrow \infty$, we have

$$\lim_{s \rightarrow \infty} \hat{\mathbf{G}}(s) = \mathbf{N}_{hc}\mathbf{D}_{hc}^{-1}$$

where \mathbf{D}_{hc} is nonsingular by assumption. Now if $\delta_{ci}\mathbf{N}(s) \leq \delta_{ci}\mathbf{D}(s)$, then \mathbf{N}_{hc} is a nonzero constant matrix. Thus $\hat{\mathbf{G}}(\infty)$ is a nonzero constant and $\hat{\mathbf{G}}(s)$ is proper. If $\delta_{ci}\mathbf{N}(s) < \delta_{ci}\mathbf{D}(s)$, then \mathbf{N}_{hc} is a zero matrix. Thus $\hat{\mathbf{G}}(\infty) = \mathbf{0}$ and $\hat{\mathbf{G}}(s)$ is strictly proper. This establishes the theorem. Q.E.D.

We state the dual of Theorem 7.8 without proof.

► **Corollary 7.8**

Let $\bar{\mathbf{N}}(s)$ and $\bar{\mathbf{D}}(s)$ be $q \times p$ and $q \times q$ polynomial matrices, and let $\bar{\mathbf{D}}(s)$ be row reduced. Then the rational matrix $\bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)$ is proper (strictly proper) if and only if

$$\delta_{ri}\bar{\mathbf{N}}(s) \leq \delta_{ri}\bar{\mathbf{D}}(s) \quad [\delta_{ri}\bar{\mathbf{N}}(s) < \delta_{ri}\bar{\mathbf{D}}(s)]$$

for $i = 1, 2, \dots, q$.

7.8.2 Computing Matrix Coprime Fractions

Given a right fraction $\mathbf{N}(s)\mathbf{D}^{-1}(s)$. One way to reduce it to a right coprime fraction is to compute its gcd. This can be achieved by applying a sequence of elementary operations. Once a gcd $\mathbf{R}(s)$ is computed, we compute $\hat{\mathbf{N}}(s) = \mathbf{N}(s)\mathbf{R}^{-1}(s)$ and $\hat{\mathbf{D}}(s) = \mathbf{D}(s)\mathbf{R}^{-1}(s)$. Then $\mathbf{N}(s)\mathbf{D}^{-1}(s) = \hat{\mathbf{N}}(s)\hat{\mathbf{D}}^{-1}(s)$ and $\hat{\mathbf{N}}(s)\hat{\mathbf{D}}^{-1}(s)$ is a right coprime fraction. This procedure will not be discussed here. The interested reader is referred to Reference 6, pp. 590–591.

We now extend the method of computing scalar coprime fractions in Section 7.3 to the matrix case. Consider a $q \times p$ proper rational matrix $\hat{\mathbf{G}}(s)$ expressed as

$$\hat{\mathbf{G}}(s) = \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s) \quad (7.81)$$

In this section, we use variables with an overbar to denote left fractions, without an overbar to denote right fractions. Clearly, (7.81) implies

$$\bar{\mathbf{N}}(s)\mathbf{D}(s) = \bar{\mathbf{D}}(s)\mathbf{N}(s)$$

and

$$\bar{\mathbf{D}}(s)(-\mathbf{N}(s)) + \bar{\mathbf{N}}(s)\mathbf{D}(s) = \mathbf{0} \quad (7.82)$$

We shall show that given a left fraction $\bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)$, not necessarily left coprime, we can obtain a right coprime fraction $\mathbf{N}(s)\mathbf{D}^{-1}(s)$ by solving the polynomial matrix equation in (7.82). Instead of solving (7.82) directly, we shall change it into solving sets of linear algebraic equations. As in (7.27), we express the polynomial matrices as, assuming the highest degree to be 4 to simplify writing,

$$\begin{aligned} \bar{\mathbf{D}}(s) &= \bar{\mathbf{D}}_0 + \bar{\mathbf{D}}_1 s + \bar{\mathbf{D}}_2 s^2 + \bar{\mathbf{D}}_3 s^3 + \bar{\mathbf{D}}_4 s^4 \\ \bar{\mathbf{N}}(s) &= \bar{\mathbf{N}}_0 + \bar{\mathbf{N}}_1 s + \bar{\mathbf{N}}_2 s^2 + \bar{\mathbf{N}}_3 s^3 + \bar{\mathbf{N}}_4 s^4 \\ \mathbf{D}(s) &= \mathbf{D}_0 + \mathbf{D}_1 s + \mathbf{D}_2 s^2 + \mathbf{D}_3 s^3 \\ \mathbf{N}(s) &= \mathbf{N}_0 + \mathbf{N}_1 s + \mathbf{N}_2 s^2 + \mathbf{N}_3 s^3 \end{aligned}$$

where $\bar{\mathbf{D}}_i$, $\bar{\mathbf{N}}_i$, \mathbf{D}_i , and \mathbf{N}_i , are, respectively, $q \times q$, $q \times p$, $p \times p$, and $q \times p$ constant matrices. The constant matrices $\bar{\mathbf{D}}_i$ and $\bar{\mathbf{N}}_i$ are known, and \mathbf{D}_i and \mathbf{N}_i are to be solved. Substituting these into (7.82) and equating to zero the constant matrices associated with

s^k , for $k = 0, 1, \dots$, yield

$$\mathbf{SM} := \begin{bmatrix} \bar{\mathbf{D}}_0 & \bar{\mathbf{N}}_0 & : & \mathbf{0} & \mathbf{0} & : & \mathbf{0} & \mathbf{0} & : & \mathbf{0} & \mathbf{0} \\ \bar{\mathbf{D}}_1 & \bar{\mathbf{N}}_1 & : & \bar{\mathbf{D}}_0 & \bar{\mathbf{N}}_0 & : & \mathbf{0} & \mathbf{0} & : & \mathbf{0} & \mathbf{0} \\ \bar{\mathbf{D}}_2 & \bar{\mathbf{N}}_2 & : & \bar{\mathbf{D}}_1 & \bar{\mathbf{N}}_1 & : & \bar{\mathbf{D}}_0 & \bar{\mathbf{N}}_0 & : & \mathbf{0} & \mathbf{0} \\ \bar{\mathbf{D}}_3 & \bar{\mathbf{N}}_3 & : & \bar{\mathbf{D}}_2 & \bar{\mathbf{N}}_2 & : & \bar{\mathbf{D}}_1 & \bar{\mathbf{N}}_1 & : & \bar{\mathbf{D}}_0 & \bar{\mathbf{N}}_0 \\ \bar{\mathbf{D}}_4 & \bar{\mathbf{N}}_4 & : & \bar{\mathbf{D}}_3 & \bar{\mathbf{N}}_3 & : & \bar{\mathbf{D}}_2 & \bar{\mathbf{N}}_2 & : & \bar{\mathbf{D}}_1 & \bar{\mathbf{N}}_1 \\ \mathbf{0} & \mathbf{0} & : & \bar{\mathbf{D}}_4 & \bar{\mathbf{N}}_4 & : & \bar{\mathbf{D}}_3 & \bar{\mathbf{N}}_3 & : & \bar{\mathbf{D}}_2 & \bar{\mathbf{N}}_2 \\ \mathbf{0} & \mathbf{0} & : & \mathbf{0} & \mathbf{0} & : & \bar{\mathbf{D}}_4 & \bar{\mathbf{N}}_4 & : & \bar{\mathbf{D}}_3 & \bar{\mathbf{N}}_3 \\ \mathbf{0} & \mathbf{0} & : & \mathbf{0} & \mathbf{0} & : & \mathbf{0} & \mathbf{0} & : & \bar{\mathbf{D}}_4 & \bar{\mathbf{N}}_4 \end{bmatrix} \begin{bmatrix} -\mathbf{N}_0 \\ \mathbf{D}_0 \\ \dots \\ -\mathbf{N}_1 \\ \mathbf{D}_1 \\ \dots \\ -\mathbf{N}_2 \\ \mathbf{D}_2 \\ \dots \\ -\mathbf{N}_3 \\ \mathbf{D}_3 \end{bmatrix} = \mathbf{0} \quad (7.83)$$

This equation is the matrix version of (7.28) and the matrix \mathbf{S} will be called a *generalized resultant*. Note that every \bar{D} -block column has q \bar{D} -columns and every \bar{N} -block column has p \bar{N} -columns. The generalized resultant \mathbf{S} as shown has four pairs of \bar{D} - and \bar{N} -block columns; thus it has a total of $4(q + p)$ columns. It has eight block rows; each block row has q rows. Thus the resultant has $8q$ rows.

We now discuss some general properties of \mathbf{S} under the assumption that linearly independent columns of \mathbf{S} from left to right have been found. It turns out that every \bar{D} -column in every \bar{D} -block column is linearly independent of its left-hand-side (LHS) columns. The situation for \bar{N} -columns, however, is different. Recall that there are p \bar{N} -columns in each \bar{N} -block column. We use $\bar{N}i$ -column to denote the i th \bar{N} -column in each \bar{N} -block column. It turns out that if the $\bar{N}i$ -column in some \bar{N} -block column is linearly dependent on its LHS columns, then all subsequent $\bar{N}i$ -column will be, because of the repetitive structure of \mathbf{S} , linearly dependent on its LHS columns. Let μ_i , $i = 1, 2, \dots, p$, be the number of linearly independent $\bar{N}i$ -columns in \mathbf{S} . They are called the *column indices* of $\hat{\mathbf{G}}(s)$. The first $\bar{N}i$ -column which becomes linearly dependent on its LHS columns is called the *primary dependent $\bar{N}i$ -column*. It is clear that the $(\mu_i + 1)$ th $\bar{N}i$ -column is the primary dependent column.

Corresponding to each primary dependent $\bar{N}i$ -column, we compute the monic null vector (its last entry equals 1) of the submatrix that consists of the primary dependent $\bar{N}i$ -column and all its LHS linearly independent columns. There are totally p such monic null vectors. From these monic null vectors, we can then obtain a right fraction. This fraction is right coprime because we use the least possible μ_i and the resulting $\mathbf{D}(s)$ has the smallest possible column degrees. The next example illustrates the procedure.

Example 7.8.1 Find a right coprime fraction of the transfer matrix in (7.75) or

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} \frac{4s-10}{2s+1} & \frac{3}{s+2} \\ \frac{1}{(2s+1)(s+2)} & \frac{s+2}{(s+2)^2} \end{bmatrix} \quad (7.84)$$

First we must find a left fraction, not necessarily left coprime. Using the least common denominator of each row, we can readily obtain

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} (2s+1)(s+2) & 0 \\ 0 & (2s+1)(s+2)^2 \end{bmatrix}^{-1} \times \begin{bmatrix} (4s-10)(s+2) & 3(2s+1) \\ s+2 & (s+1)(2s+1) \end{bmatrix} =: \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)$$

Thus we have

$$\begin{aligned} \bar{\mathbf{D}}(s) &= \begin{bmatrix} 2s^2 + 5s + 2 & 0 \\ 0 & 2s^3 + 9s^2 + 12s + 4 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} + \begin{bmatrix} 5 & 0 \\ 0 & 12 \end{bmatrix}s + \begin{bmatrix} 2 & 0 \\ 0 & 9 \end{bmatrix}s^2 + \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}s^3 \end{aligned}$$

and

$$\begin{aligned} \bar{\mathbf{N}}(s) &= \begin{bmatrix} 4s^2 - 2s - 20 & 6s + 3 \\ s + 2 & 2s^2 + 3s + 1 \end{bmatrix} \\ &= \begin{bmatrix} -20 & 3 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} -2 & 6 \\ 1 & 3 \end{bmatrix}s + \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}s^2 + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}s^3 \end{aligned}$$

We form the generalized resultant and then use the QR decomposition discussed in Section 7.3.1 to search its linearly independent columns in order from left to right. Because it is simpler to key in the transpose of \mathbf{S} , we type

```
d1=[2 0 5 0 2 0 0 0];d2=[0 4 0 12 0 9 0 2];
n1=[-20 2 -2 1 4 0 0 0];n2=[3 1 6 3 0 2 0 0];
s=[d1 0 0 0;d2 0 0 0;n1 0 0 0 0;n2 0 0 0 0;...
    0 0 d1 0 0;0 0 d2 0 0;0 0 n1 0 0;0 0 n2 0 0;...
    0 0 0 0 d1;0 0 0 0 d2;0 0 0 0 n1;0 0 0 0 n2]';
[q,r]=qr(s)
```

We need only r ; therefore, the matrix q will not be shown. As discussed in Section 7.3.1, we need to know whether or not entries of r are zero in determining linear independence of columns; therefore, all nonzero entries are represented by x, di , and ni . The result is

$$r = \begin{bmatrix} d1 & 0 & x & x & x & x & x & x & x & 0 & x & x \\ 0 & d2 & x & x & x & x & x & x & 0 & x & x & x \\ 0 & 0 & n1 & x & x & x & x & x & x & x & x & x \\ 0 & 0 & 0 & n2 & x & x & x & x & x & x & x & x \\ 0 & 0 & 0 & 0 & d1 & x & x & x & x & x & x & x \\ 0 & 0 & 0 & 0 & 0 & d2 & x & x & x & x & x & x \\ 0 & 0 & 0 & 0 & 0 & 0 & n1 & x & x & x & x & x \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x & x & x & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & d1 & x & x & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & d2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We see that all D -columns are linearly independent of their LHS columns. There are two linearly independent \bar{N}_1 -columns and one linearly independent \bar{N}_2 -column. Thus we have $\mu_1 = 2$ and $\mu_2 = 1$. The eighth column of \mathbf{S} is the primary dependent \bar{N}_2 -column. We compute a null vector of the submatrix that consists of the primary dependent \bar{N}_2 -column and all its LHS linearly independent columns as

```
z2=null([d1 0 0;d2 0 0;n1 0 0;n2 0 0;...
          0 0 d1;0 0 d2;0 0 n1;0 0 n2]');
```

and then normalize the last entry to 1 by typing

```
z2b=z2/z2(8)
```

which yields the first monic null vector as

```
ans=[7 -1 1 2 -4 0 2 1]'
```

The eleventh column of \mathbf{S} is the primary dependent \bar{N}_1 -column. We compute a null vector of the submatrix that consists of the primary dependent \bar{N}_1 -column and all its LHS linearly independent columns (that is, deleting the eighth column) as

```
z1=null([d1 0 0 0 0;d2 0 0 0 0;n1 0 0 0 0;n2 0 0 0 0;...
          0 0 d1 0 0;0 0 d2 0 0;0 0 n1 0 0;...
          0 0 0 0 d1;0 0 0 0 d2;0 0 0 0 n1]');
```

and then normalize the last entry to 1 by typing

```
z1b=z1/z1(10)
```

which yields the second monic null vector as

```
ans=[10 -0.5 1 0 1 0 2.5 -2 0 1]'
```

Thus we have

$$\begin{bmatrix} -\mathbf{N}_0 \\ \dots \\ \mathbf{D}_0 \\ \dots \\ -\mathbf{N}_1 \\ \dots \\ \mathbf{D}_1 \\ \dots \\ -\mathbf{N}_2 \\ \dots \\ \mathbf{D}_2 \end{bmatrix} = \begin{bmatrix} -n_0^{11} & -n_0^{12} \\ -n_0^{21} & -n_0^{22} \\ \dots & \dots \\ d_0^{11} & d_0^{12} \\ d_0^{21} & d_0^{22} \\ \dots & \dots \\ -n_1^{11} & -n_1^{12} \\ -n_1^{21} & -n_1^{22} \\ \dots & \dots \\ d_1^{11} & d_1^{12} \\ d_1^{21} & d_1^{22} \\ \dots & \dots \\ -n_2^{11} & -n_2^{12} \\ -n_2^{21} & -n_2^{22} \\ \dots & \dots \\ d_2^{11} & d_2^{12} \\ d_2^{21} & d_2^{22} \end{bmatrix} = \begin{bmatrix} 10 & 7 \\ -0.5 & -1 \\ \dots & \dots \\ 1 & 1 \\ 0 & 2 \\ \dots & \dots \\ 1 & -4 \\ 0 & 0 \\ \dots & \dots \\ 2.5 & 2 \\ 1 & \\ \dots & \dots \\ -2 & \\ 0 & \\ \dots & \dots \\ 1 & \end{bmatrix} \quad (7.85)$$

where we have written out \mathbf{N}_i and \mathbf{D}_i explicitly with the superscripts ij denoting the ij th entry and the subscript denoting the degree. The two monic null vectors are arranged as shown. The order of the two null vectors can be interchanged as we will discuss shortly. The empty entries are to be filled up with zero. Note that the empty entry at the (8×1) th location is due to deleting the second $\bar{N}2$ linearly dependent column in computing the second null vector. By equating the corresponding entries in (7.85), we can readily obtain

$$\begin{aligned} \mathbf{D}(s) &= \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} + \begin{bmatrix} 2.5 & 2 \\ 0 & 1 \end{bmatrix}s + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}s^2 \\ &= \begin{bmatrix} s^2 + 2.5s + 1 & 2s + 1 \\ 0 & s + 2 \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} \mathbf{N}(s) &= \begin{bmatrix} -10 & -7 \\ 0.5 & 1 \end{bmatrix} + \begin{bmatrix} -1 & 4 \\ 0 & 0 \end{bmatrix}s + \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}s^2 \\ &= \begin{bmatrix} s^2 - s - 10 & 4s - 7 \\ 0.5 & 1 \end{bmatrix} \end{aligned}$$

Thus $\hat{\mathbf{G}}(s)$ in (7.84) has the following right coprime fraction:

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} (2s-5)(s+2) & 4s-7 \\ 0.5 & 1 \end{bmatrix} \begin{bmatrix} (s+2)(s+0.5) & 2s+1 \\ 0 & s+2 \end{bmatrix}^{-1} \quad (7.86)$$

The $\mathbf{D}(s)$ in (7.86) is column reduced with column degrees $\mu_1 = 2$ and $\mu_2 = 1$. Thus we have $\deg \det \mathbf{D}(s) = 2 + 1 = 3$ and the degree of $\hat{\mathbf{G}}(s)$ in (7.84) is 3. The degree was computed in Example 7.7.1 as 3 by using Definition 7.1.

In general, if the generalized resultant has μ_i linearly independent \bar{N}_i -columns, then $\mathbf{D}(s)$ is column reduced with column degrees μ_i . Thus we have

$$\begin{aligned}\deg \hat{\mathbf{G}}(s) &= \deg \det \mathbf{D}(s) = \sum \mu_i \\ &= \text{Total number of linearly independent } \bar{N} \text{-columns in } \mathbf{S}\end{aligned}$$

We next show that the order of column degrees is immaterial. In other words, the order of the columns of $\mathbf{N}(s)$ and $\mathbf{D}(s)$ can be changed. For example, consider the permutation matrix

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

and $\hat{\mathbf{D}}(s) = \mathbf{D}(s)\mathbf{P}$ and $\hat{\mathbf{N}}(s) = \mathbf{N}(s)\mathbf{P}$. Then the first, second, and third columns of $\mathbf{D}(s)$ and $\mathbf{N}(s)$ become the second, third, and first columns of $\hat{\mathbf{D}}(s)$ and $\hat{\mathbf{N}}(s)$. However, we have

$$\hat{\mathbf{G}}(s) = \hat{\mathbf{N}}(s)\hat{\mathbf{D}}^{-1}(s) = [\mathbf{N}(s)\mathbf{P}][\mathbf{D}(s)\mathbf{P}]^{-1} = \mathbf{N}(s)\mathbf{D}^{-1}(s)$$

This shows that the columns of $\mathbf{D}(s)$ and $\mathbf{N}(s)$ can be arbitrarily permuted. This is the same as permuting the order of the null vectors in (7.83). Thus the set of column degrees is an intrinsic property of a system just as the set of controllability indices is (Theorem 6.3). What has been discussed can be stated as a theorem. It is an extension of Theorem 7.4 to the matrix case.

► Theorem 7.M4

Let $\hat{\mathbf{G}}(s) = \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)$ be a left fraction, not necessarily left coprime. We use the coefficient matrices of $\bar{\mathbf{D}}(s)$ and $\bar{\mathbf{N}}(s)$ to form the generalized resultant \mathbf{S} shown in (7.83) and search its linearly independent columns from left to right. Let $\mu_i, i = 1, 2, \dots, p$, be the numbers of linearly independent \bar{N}_i -columns. Then we have

$$\deg \hat{\mathbf{G}}(s) = \mu_1 + \mu_2 + \dots + \mu_p \quad (7.87)$$

and a right coprime fraction $\mathbf{N}(s)\mathbf{D}^{-1}(s)$ can be obtained by computing p monic null vectors of the p matrices formed from each primary dependent \bar{N}_i -column and all its LHS linearly independent columns.

The right coprime fraction obtained by solving the equation in (7.83) has one additional important property. After permutation, the column-degree-coefficient matrix \mathbf{D}_{hc} can always become an unit upper triangular matrix (an upper triangular matrix with 1 as its diagonal entries). Such a $\mathbf{D}(s)$ is said to be in *column echelon form*.

See References 6, pp. 610–612; 13, pp. 483–487. For the $\mathbf{D}(s)$ in (7.86), its column-degree-coefficient matrix is

$$\mathbf{D}_{hc} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

It is unit upper triangular; thus the $\mathbf{D}(s)$ is in column echelon form. Although we need only column reducedness in subsequent discussion, if $\mathbf{D}(s)$ is in column echelon form, then the result in the next section will be nicer.

Dual to the preceding discussion, we can compute a left coprime fraction from a right fraction $\mathbf{N}(s)\mathbf{D}^{-1}(s)$ which is not necessarily right coprime. Then similar to (7.83), we form

$$[-\bar{\mathbf{N}}_0 \bar{\mathbf{D}}_0 : -\bar{\mathbf{N}}_1 \bar{\mathbf{D}}_1 : -\bar{\mathbf{N}}_2 \bar{\mathbf{D}}_2 : -\bar{\mathbf{N}}_3 \bar{\mathbf{D}}_3] \mathbf{T} = \mathbf{0} \quad (7.88)$$

with

$$\mathbf{T} := \left[\begin{array}{ccccccccc} \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \mathbf{D}_3 & \mathbf{D}_4 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{N}_0 & \mathbf{N}_1 & \mathbf{N}_2 & \mathbf{N}_3 & \mathbf{N}_4 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \dots & \dots \\ \mathbf{0} & \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \mathbf{D}_3 & \mathbf{D}_4 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_0 & \mathbf{N}_1 & \mathbf{N}_2 & \mathbf{N}_3 & \mathbf{N}_4 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \dots & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \mathbf{D}_3 & \mathbf{D}_4 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{N}_0 & \mathbf{N}_1 & \mathbf{N}_2 & \mathbf{N}_3 & \mathbf{N}_4 & \mathbf{0} & \mathbf{0} \\ \dots & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \mathbf{D}_3 & \mathbf{D}_4 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{N}_0 & \mathbf{N}_1 & \mathbf{N}_2 & \mathbf{N}_3 & \mathbf{N}_4 & \mathbf{0} \end{array} \right] \quad (7.89)$$

We search linearly independent rows in order from top to bottom. Then all D-rows are linearly independent. Let Ni-row denote the i -th N-row in each N block-row. If an Ni-row becomes linearly dependent, then all Ni-rows in subsequent N block-rows are linearly dependent on their preceding rows. The first Ni-row which becomes linearly dependent is called a primary dependent Ni-row. Let v_i , $i = 1, 2, \dots, q$, be the number of linearly independent Ni-rows. They are called the *row indices* of $\hat{\mathbf{G}}(s)$. Then dual to Theorem 7.M4, we have the following corollary.

► Corollary 7.M4

Let $\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)$ be a right fraction, not necessarily right coprime. We use the coefficient matrices of $\mathbf{D}(s)$ and $\mathbf{N}(s)$ to form the generalized resultant \mathbf{T} shown in (7.89) and search its linearly independent rows in order from top to bottom. Let v_i for $i = 1, 2, \dots, q$ be the numbers of linearly independent Ni-rows in \mathbf{T} . Then we have

$$\deg \hat{\mathbf{G}}(s) = v_1 + v_2 + \dots + v_q$$

and a left coprime fraction $\bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)$ can be obtained by computing q monic left null vectors of the q matrices formed from each primary dependent Ni-row and all its preceding linearly independent rows.

The polynomial matrix $\bar{\mathbf{D}}(s)$ obtained in Corollary 7.M4 is row reduced with row degrees $\{v_i, i = 1, 2, \dots, q\}$. In fact, it is in *row echelon form*, that is, its row-degree-coefficient matrix, after some row permutation, is a unit lower triangular matrix.

7.9 Realization from Matrix Coprime Fractions

Not to be overwhelmed by notations, we discuss a realization of a 2×2 strictly proper rational matrix $\hat{\mathbf{G}}(s)$ expressed as

$$\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s) \quad (7.90)$$

where $\mathbf{N}(s)$ and $\mathbf{D}(s)$ are right coprime and $\mathbf{D}(s)$ is in column echelon form.⁷ We further assume that the column degrees of $\mathbf{D}(s)$ are $\mu_1 = 4$ and $\mu_2 = 2$. First we define

$$\mathbf{H}(s) := \begin{bmatrix} s^{\mu_1} & 0 \\ 0 & s^{\mu_2} \end{bmatrix} = \begin{bmatrix} s^4 & 0 \\ 0 & s^2 \end{bmatrix} \quad (7.91)$$

and

$$\mathbf{L}(s) := \begin{bmatrix} s^{\mu_1-1} & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & s^{\mu_2-1} \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} s^3 & 0 \\ s^2 & 0 \\ s & 0 \\ 1 & 0 \\ 0 & s \\ 0 & 1 \end{bmatrix} \quad (7.92)$$

The procedure for developing a realization for

$$\hat{\mathbf{y}}(s) = \hat{\mathbf{G}}(s)\hat{\mathbf{u}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)\hat{\mathbf{u}}(s)$$

follows closely the scalar case from (7.3) through (7.9). First we introduce a new variable $\mathbf{v}(t)$ defined by $\hat{\mathbf{v}}(s) = \mathbf{D}^{-1}(s)\hat{\mathbf{u}}(s)$. Note that $\hat{\mathbf{v}}(s)$, called a pseudo state, is a 2×1 column vector. Then we have

$$\mathbf{D}(s)\hat{\mathbf{v}}(s) = \hat{\mathbf{u}}(s) \quad (7.93)$$

$$\hat{\mathbf{y}}(s) = \mathbf{N}(s)\hat{\mathbf{v}}(s) \quad (7.94)$$

⁷All discussion is still applicable if $\mathbf{D}(s)$ is column reduced but not in echelon form.

Let us define state variables as

$$\begin{aligned}\hat{\mathbf{x}}(s) &= \mathbf{L}(s)\hat{\mathbf{v}}(s) = \begin{bmatrix} s^{\mu_1-1} & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & s^{\mu_2-1} \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{v}_1(s) \\ \hat{v}_2(s) \end{bmatrix} \\ &= \begin{bmatrix} s^3\hat{v}_1(s) \\ s^2\hat{v}_1(s) \\ s\hat{v}_1(s) \\ \hat{v}_1(s) \\ s\hat{v}_2(s) \\ \hat{v}_2(s) \end{bmatrix} =: \begin{bmatrix} x_1(s) \\ x_2(s) \\ x_3(s) \\ x_4(s) \\ x_5(s) \\ x_6(s) \end{bmatrix} \quad (7.95)\end{aligned}$$

or, in the time domain,

$$\begin{aligned}x_1(t) &= v_1^{(3)}(t), & x_2(t) &= \ddot{v}_1(t), & x_3(t) &= \dot{v}_1(t), & x_4(t) &= v_1(t) \\ x_5(t) &= \dot{v}_2(t), & x_6(t) &= v_2(t)\end{aligned}$$

This state vector has dimension $\mu_1 + \mu_2 = 6$. These definitions imply immediately

$$\dot{x}_2 = x_1, \quad \dot{x}_3 = x_2, \quad \dot{x}_4 = x_3, \quad \dot{x}_6 = x_5 \quad (7.96)$$

Next we use (7.93) to develop equations for \dot{x}_1 and \dot{x}_5 . First we express $\mathbf{D}(s)$ as

$$\mathbf{D}(s) = \mathbf{D}_{hc}\mathbf{H}(s) + \mathbf{D}_{lc}\mathbf{L}(s) \quad (7.97)$$

where $\mathbf{H}(s)$ and $\mathbf{L}(s)$ are defined in (7.91) and (7.92). Note that \mathbf{D}_{hc} and \mathbf{D}_{lc} are constant matrices and the column-degree-coefficient matrix \mathbf{D}_{hc} is a unit upper triangular matrix. Substituting (7.97) into (7.93) yields

$$[\mathbf{D}_{hc}\mathbf{H}(s) + \mathbf{D}_{lc}\mathbf{L}(s)]\hat{\mathbf{v}}(s) = \hat{\mathbf{u}}(s)$$

or

$$\mathbf{H}(s)\hat{\mathbf{v}}(s) + \mathbf{D}_{hc}^{-1}\mathbf{D}_{lc}\mathbf{L}(s)\hat{\mathbf{v}}(s) = \mathbf{D}_{hc}^{-1}\hat{\mathbf{u}}(s)$$

Thus we have, using (7.95),

$$\mathbf{H}(s)\hat{\mathbf{v}}(s) = -\mathbf{D}_{hc}^{-1}\mathbf{D}_{lc}\hat{\mathbf{x}}(s) + \mathbf{D}_{hc}^{-1}\hat{\mathbf{u}}(s) \quad (7.98)$$

Let

$$\mathbf{D}_{hc}^{-1}\mathbf{D}_{lc} =: \begin{bmatrix} \alpha_{111} & \alpha_{112} & \alpha_{113} & \alpha_{114} & \alpha_{121} & \alpha_{122} \\ \alpha_{211} & \alpha_{212} & \alpha_{213} & \alpha_{214} & \alpha_{221} & \alpha_{222} \end{bmatrix} \quad (7.99)$$

and

$$\mathbf{D}_{hc}^{-1} =: \begin{bmatrix} 1 & b_{12} \\ 0 & 1 \end{bmatrix} \quad (7.100)$$

Note that the inverse of a unit upper triangular matrix is again a unit upper triangular matrix. Substituting (7.99) and (7.100) into (7.98) and using $s\hat{x}_1(s) = s^4\hat{v}_1(s)$ and $s\hat{x}_5(s) = s^2\hat{v}_2(s)$ yield

$$\begin{bmatrix} s\hat{x}_1(s) \\ s\hat{x}_5(s) \end{bmatrix} = -\begin{bmatrix} \alpha_{111} & \alpha_{112} & \alpha_{113} & \alpha_{114} & \alpha_{121} & \alpha_{122} \\ \alpha_{211} & \alpha_{212} & \alpha_{213} & \alpha_{214} & \alpha_{221} & \alpha_{222} \end{bmatrix}\hat{\mathbf{x}}(s) + \begin{bmatrix} 1 & b_{12} \\ 0 & 1 \end{bmatrix}\hat{\mathbf{u}}(s)$$

which becomes, in the time domain,

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_5(t) \end{bmatrix} = -\begin{bmatrix} \alpha_{111} & \alpha_{112} & \alpha_{113} & \alpha_{114} & \alpha_{121} & \alpha_{122} \\ \alpha_{211} & \alpha_{212} & \alpha_{213} & \alpha_{214} & \alpha_{221} & \alpha_{222} \end{bmatrix}\mathbf{x}(t) + \begin{bmatrix} 1 & b_{12} \\ 0 & 1 \end{bmatrix}\mathbf{u}(t) \quad (7.101)$$

If $\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)$ is strictly proper, then the column degrees of $\mathbf{N}(s)$ is less than the corresponding column degrees of $\mathbf{D}(s)$. Thus we can express $\mathbf{N}(s)$ as

$$\mathbf{N}(s) = \begin{bmatrix} \beta_{111} & \beta_{112} & \beta_{113} & \beta_{114} & \beta_{121} & \beta_{122} \\ \beta_{211} & \beta_{212} & \beta_{213} & \beta_{214} & \beta_{221} & \beta_{222} \end{bmatrix}\mathbf{L}(s) \quad (7.102)$$

Substituting this into (7.94) and using $\hat{\mathbf{x}}(s) = \mathbf{L}(s)\hat{\mathbf{v}}(s)$ yield

$$\hat{\mathbf{y}}(s) = \begin{bmatrix} \beta_{111} & \beta_{112} & \beta_{113} & \beta_{114} & \beta_{121} & \beta_{122} \\ \beta_{211} & \beta_{212} & \beta_{213} & \beta_{214} & \beta_{221} & \beta_{222} \end{bmatrix}\hat{\mathbf{x}}(s) \quad (7.103)$$

Combining (7.96), (7.101), and (7.103) yields the following realization for $\hat{\mathbf{G}}(s)$:

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \left[\begin{array}{cccccc} -\alpha_{111} & -\alpha_{112} & -\alpha_{113} & -\alpha_{114} & \vdots & -\alpha_{121} & -\alpha_{122} \\ 1 & 0 & 0 & 0 & \vdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \vdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \vdots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -\alpha_{211} & -\alpha_{212} & -\alpha_{213} & -\alpha_{214} & \vdots & -\alpha_{221} & -\alpha_{222} \\ 0 & 0 & 0 & 0 & \vdots & 1 & 0 \end{array} \right] \mathbf{x}(t) \\ &+ \left[\begin{array}{cc} 1 & b_{12} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \dots & \dots \\ 0 & 1 \\ 0 & 0 \end{array} \right] \mathbf{u}(t) \\ \mathbf{y}(t) &= \left[\begin{array}{cccccc} \beta_{111} & \beta_{112} & \beta_{113} & \beta_{114} & \vdots & \beta_{121} & \beta_{122} \\ \beta_{211} & \beta_{212} & \beta_{213} & \beta_{214} & \vdots & \beta_{221} & \beta_{222} \end{array} \right] \mathbf{x}(t)\end{aligned}\tag{7.104}$$

This is a $(\mu_1 + \mu_2)$ -dimensional state-space equation. The A-matrix has two companion-form diagonal blocks; one with order $\mu_1 = 4$ and the other with order $\mu_2 = 2$. The off-diagonal blocks are zeros except their first rows. This state-space equation is a generalization of the state-space equation in (7.9) to two-input two-output transfer matrices. We can readily show that (7.104) is always controllable and is called a *controllable-form* realization. Furthermore, the controllability indices equal $\mu_1 = 4$ and $\mu_2 = 2$. As in (7.9), the state-space equation in (7.104) is observable if and only if $\mathbf{D}(s)$ and $\mathbf{N}(s)$ are right coprime. For its proof, see Reference 6, p. 282. Because we start with a coprime fraction $\mathbf{N}(s)\mathbf{D}^{-1}(s)$, the realization in (7.104) is observable as well. In conclusion, the realization in (7.104) is controllable and observable and its dimension equals $\mu - 1 + \mu_2$ and, following Theorem 7.M4, the degree of $\hat{\mathbf{G}}(s)$. This establishes the second part of Theorem 7.M2; that is, a state-space equation is minimal or controllable and observable if and only if its dimension equals the degree of its transfer matrix.

Example 7.9.1 Consider the transfer matrix in Example 7.7.1. We gave there a minimal realization that is obtained by reducing the nonminimal realization in (4.47). Now we will develop directly a minimal realization by using a coprime fraction. We first write

the transfer matrix as

$$\begin{aligned}\hat{\mathbf{G}}(s) &= \begin{bmatrix} \frac{4s-10}{2s+1} & \frac{3}{s+1} \\ \frac{1}{(2s+1)(s+2)} & \frac{s+2}{(s+2)^2} \end{bmatrix} =: \hat{\mathbf{G}}(\infty) + \hat{\mathbf{G}}_{sp}(s) \\ &= \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \frac{-12}{2s+1} & \frac{3}{s+1} \\ \frac{1}{(2s+1)(s+2)} & \frac{s+2}{(s+2)^2} \end{bmatrix}\end{aligned}$$

As in Example 7.8.1, we can find a right coprime fraction for the strictly proper part of $\hat{\mathbf{G}}(s)$ as

$$\hat{\mathbf{G}}_{sp}(s) = \begin{bmatrix} -6s-12 & -9 \\ 0.5 & 1 \end{bmatrix} \begin{bmatrix} s^2+2.5s+1 & 2s+1 \\ 0 & s+2 \end{bmatrix}^{-1}$$

Note that its denominator matrix is the same as the one in (7.86). See Problem 7.24. Clearly we have $\mu_1 = 2$ and $\mu_2 = 1$. We define

$$\mathbf{H}(s) = \begin{bmatrix} s^2 & 0 \\ 0 & s \end{bmatrix}, \quad \mathbf{L}(s) = \begin{bmatrix} s & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Thus we have

$$\mathbf{D}(s) = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \mathbf{H}(s) + \begin{bmatrix} 2.5 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix} \mathbf{L}(s)$$

and

$$\mathbf{N}(s) = \begin{bmatrix} -6 & -12 & -9 \\ 0 & 0.5 & 1 \end{bmatrix} \mathbf{L}(s)$$

We compute

$$\mathbf{D}_{hc}^{-1} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$$

and

$$\mathbf{D}_{hc}^{-1} \mathbf{D}_{lc} = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2.5 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 2.5 & 1 & -3 \\ 0 & 0 & 2 \end{bmatrix}$$

Thus a minimal realization of $\hat{\mathbf{G}}(s)$ is

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \begin{bmatrix} -2.5 & -1 & \vdots & 3 \\ 1 & 0 & \vdots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \vdots & -2 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 & -2 \\ 0 & 0 \\ \dots & \dots \\ 0 & 1 \end{bmatrix} \mathbf{u}(t) \\ \mathbf{y}(t) &= \begin{bmatrix} -6 & -12 & \vdots & -9 \\ 0 & 0.5 & \vdots & 1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \mathbf{u}(t)\end{aligned}\quad (7.105)$$

This A-matrix has two companion-form diagonal blocks: one with order 2 and the other with order 1. This three dimensional realization is a minimal realization and is in controllable form. Comparing with the minimal realization obtained in Example 7.7.1, the controllable-form realization in (7.105) consists of many zeros and ones, thus its op-amp circuit realization will use less number of components. Thus the realization using a coprime fraction is an important one.

Dual to the preceding minimal realization, if we use $\hat{\mathbf{G}}(s) = \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)$ where $\bar{\mathbf{D}}(s)$ and $\bar{\mathbf{N}}(s)$ are left coprime and $\bar{\mathbf{D}}(s)$ is in row echelon form with row degrees $\{v_i, i = 1, 2, \dots, q\}$, then we can obtain an observable-form realization with observability indices $\{v_i, i = 1, 2, \dots, q\}$. This will not be repeated.

We summarize the main results in the following. As in the SISO case, an n -dimensional MIMO state-space equation is controllable and observable if its transfer matrix has degree n . If a proper transfer matrix is expressed as a right coprime fraction with column reducedness, then the realization obtained by using the preceding procedure will automatically be controllable and observable.

Let $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ be a minimal realization of $\hat{\mathbf{G}}(s)$ and let $\hat{\mathbf{G}}(s) = \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s) = \mathbf{N}(s)\bar{\mathbf{D}}^{-1}(s)$ be coprime fractions: $\bar{\mathbf{D}}(s)$ is row reduced, and $\mathbf{D}(s)$ is column reduced. Then we have

$$\mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D} = \mathbf{N}(s)\bar{\mathbf{D}}^{-1}(s) = \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)$$

which implies

$$\begin{aligned}\frac{1}{\det(s\mathbf{I} - \mathbf{A})}\mathbf{C}[\text{Adj}(s\mathbf{I} - \mathbf{A})]\mathbf{B} + \mathbf{D} &= \frac{1}{\det \mathbf{D}(s)}\mathbf{N}(s)[\text{Adj}(\mathbf{D}(s))] \\ &= \frac{1}{\det \bar{\mathbf{D}}(s)}[\text{Adj}(\bar{\mathbf{D}}(s))]\bar{\mathbf{N}}(s)\end{aligned}$$

Because the three polynomials $\det(s\mathbf{I} - \mathbf{A})$, $\det \mathbf{D}(s)$, and $\det \bar{\mathbf{D}}(s)$ have the same degree, they must denote the same polynomial except possibly different leading coefficients. Thus we conclude the following:

- $\deg \hat{\mathbf{G}}(s) = \deg \det \mathbf{D}(s) = \deg \det \bar{\mathbf{D}}(s) = \dim \mathbf{A}$.
- Characteristic polynomial of $\hat{\mathbf{G}}(s) = k_1 \det \mathbf{D}(s) = k_2 \det \bar{\mathbf{D}}(s) = k_3 \det(s\mathbf{I} - \mathbf{A})$ for some nonzero constants k_i .
- The set of column degrees of $\mathbf{D}(s)$ equals the set of controllability indices of (\mathbf{A}, \mathbf{B}) .
- The set of row degrees of $\bar{\mathbf{D}}(s)$ equals the set of observability indices of (\mathbf{A}, \mathbf{C}) .

We see that coprime fractions and controllable and observable state-space equations contain essentially the same information. Thus either description can be used in analysis and design.

7.10 Realizations from Matrix Markov Parameters

Consider a $q \times p$ strictly proper rational matrix $\hat{\mathbf{G}}(s)$. We expand it as

$$\hat{\mathbf{G}}(s) = \mathbf{H}(1)s^{-1} + \mathbf{H}(2)s^{-2} + \mathbf{H}(3)s^{-3} + \dots \quad (7.106)$$

where $\mathbf{H}(m)$ are $q \times p$ constant matrices. Let r be the degree of the least common denominator of all entries of $\hat{\mathbf{G}}(s)$. We form

$$\mathbf{T} = \begin{bmatrix} \mathbf{H}(1) & \mathbf{H}(2) & \mathbf{H}(3) & \cdots & \mathbf{H}(r) \\ \mathbf{H}(2) & \mathbf{H}(3) & \mathbf{H}(4) & \cdots & \mathbf{H}(r+1) \\ \mathbf{H}(3) & \mathbf{H}(4) & \mathbf{H}(5) & \cdots & \mathbf{H}(r+2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}(r) & \mathbf{H}(r+1) & \mathbf{H}(r+2) & \cdots & \mathbf{H}(2r-1) \end{bmatrix} \quad (7.107)$$

$$\tilde{\mathbf{T}} = \begin{bmatrix} \mathbf{H}(2) & \mathbf{H}(3) & \mathbf{H}(4) & \cdots & \mathbf{H}(r+1) \\ \mathbf{H}(3) & \mathbf{H}(4) & \mathbf{H}(5) & \cdots & \mathbf{H}(r+2) \\ \mathbf{H}(4) & \mathbf{H}(5) & \mathbf{H}(6) & \cdots & \mathbf{H}(r+3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}(r+1) & \mathbf{H}(r+2) & \mathbf{H}(r+3) & \cdots & \mathbf{H}(2r) \end{bmatrix} \quad (7.108)$$

Note that \mathbf{T} and $\tilde{\mathbf{T}}$ have r block columns and r block rows and, consequently, have dimension $rq \times rp$. As in (7.53) and (7.55), we have

$$\mathbf{T} = \mathcal{O}_{r-1} \mathcal{C}_{r-1} \quad \text{and} \quad \tilde{\mathbf{T}} = \mathcal{O}_{r-1} \mathbf{A} \mathcal{C}_{r-1} \quad (7.109)$$

where \mathcal{O}_{r-1} and \mathcal{C}_{r-1} are the observability and controllability matrices of orders $rq \times n$ and $n \times rp$. Note that n is not yet determined. Because r equals the degree of the minimal polynomial of any minimal realization of $\hat{\mathbf{G}}(s)$ and because of (6.16) and (6.34), the matrix \mathbf{T} is sufficiently large to have rank n . This is stated as a theorem.

► **Theorem 7.M7**

A strictly proper rational matrix $\hat{\mathbf{G}}(s)$ has degree n if and only if the matrix \mathbf{T} in (7.107) has rank n .

The singular-value decomposition method discussed in Section 7.5 can be applied to the matrix case with some modification. This is discussed in the following. First we use singular-value decomposition to express \mathbf{T} as

$$\mathbf{T} = \mathbf{K} \begin{bmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{L}' \quad (7.110)$$

where \mathbf{K} and \mathbf{L} are orthogonal matrices and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, where λ_i are the positive square roots of the positive eigenvalues of $\mathbf{T}'\mathbf{T}$. Clearly, n is the rank of \mathbf{T} . Let $\bar{\mathbf{K}}$ denote the first n columns of \mathbf{K} and $\bar{\mathbf{L}}'$ denote the first n rows of \mathbf{L}' . Then we can write \mathbf{T} as

$$\mathbf{T} = \bar{\mathbf{K}} \Lambda \bar{\mathbf{L}}' = \bar{\mathbf{K}} \Lambda^{1/2} \Lambda^{1/2} \bar{\mathbf{L}}' =: \mathcal{O} \mathcal{C} \quad (7.111)$$

where

$$\mathcal{O} = \bar{\mathbf{K}} \Lambda^{1/2} \quad \text{and} \quad \mathcal{C} = \Lambda^{1/2} \bar{\mathbf{L}}'$$

Note that \mathcal{O} is $nq \times n$ and \mathcal{C} is $n \times np$. They are not square and their inverses are not defined. However, their pseudoinverses are defined. The pseudoinverse of \mathcal{O} is, as defined in (7.69),

$$\mathcal{O}^+ = [(\Lambda^{1/2})' \bar{\mathbf{K}}' \bar{\mathbf{K}} \Lambda^{1/2}]^{-1} (\Lambda^{1/2})' \bar{\mathbf{K}}'$$

Because \mathbf{K} is orthogonal, we have $\bar{\mathbf{K}}' \bar{\mathbf{K}} = \mathbf{I}$ and because $\Lambda^{1/2}$ is symmetric, the pseudoinverse \mathcal{O}^+ reduces to

$$\mathcal{O}^+ = \Lambda^{-1/2} \bar{\mathbf{K}}' \quad (7.112)$$

Similarly, we have

$$\mathcal{C}^+ = \bar{\mathbf{L}} \Lambda^{-1/2} \quad (7.113)$$

Then, as in (7.64) through (7.66), the triplet

$$\mathbf{A} = \mathcal{O}^+ \tilde{\mathbf{T}} \mathcal{C}^+ \quad (7.114)$$

$$\mathbf{B} = \text{first } p \text{ columns of } \mathcal{C} \quad (7.115)$$

$$\mathbf{C} = \text{first } q \text{ rows of } \mathcal{O} \quad (7.116)$$

is a minimal realization of $\hat{\mathbf{G}}(s)$. This realization has the property

$$\mathcal{O}' \mathcal{O} = \Lambda^{1/2} \bar{\mathbf{K}}' \bar{\mathbf{K}} \Lambda^{1/2} = \Lambda$$

and, using $\bar{\mathbf{L}}' \bar{\mathbf{L}} = \mathbf{I}$,

$$\mathcal{C} \mathcal{C}' = \Lambda^{1/2} \bar{\mathbf{L}}' \bar{\mathbf{L}} \Lambda^{1/2} = \Lambda$$

Thus the realization is a balanced realization. The MATLAB procedure in Example 7.5.2 can be directly applied to the matrix case if the function inverse (`inv`) is replaced by the function pseudoinverse (`pinv`). We see once again that the procedure in the scalar case can be directly extended to the matrix case. We also mention that if we decompose $\mathbf{T} = \mathcal{O}\mathcal{C}$ in (7.111) differently, we will obtain different realization. This will not be discussed.

7.11 Concluding Remarks

In addition to a number of minimal realizations, we introduced in this chapter coprime fractions (right fractions with column-reducedness and left fractions with row-reducedness). These fractions can be obtained by searching linearly independent vectors of generalized resultants and then solving monic null vectors of their submatrices. A fundamental result of this chapter is that controllable and observable state-space equations are essentially equivalent to polynomial coprime fractions, denoted as

$$\begin{gathered} \text{controllable \& observable state-space equations} \\ \Updownarrow \\ \text{coprime polynomial fractions} \end{gathered}$$

Thus either description can be used to describe a system. We use the former in the next chapter and the latter in Chapter 9 to carry out various designs.

A great deal more can be said regarding coprime polynomial fractions. For example, it is possible to show that all coprime fractions are related by unimodular matrices. Controllability and observability conditions can also be expressed in terms of coprimeness conditions. See References 4, 6, 16, and 24. The objectives of this chapter are to discuss a numerical method to compute coprimes fractions and to introduce just enough background to carry out designs in Chapter 9.

In addition to polynomial fractions, it is possible to express transfer functions as stable rational function fractions. See References 12 and 23. Stable rational function fractions can be developed without discussing polynomial fractions; however, polynomial fractions can provide an efficient way of computing rational fractions. Thus this chapter is also useful in studying rational fractions.

Problems

7.1 Given

$$\hat{g}(s) = \frac{s+2}{(s^2 - 1)(s+2)}$$

- Find a three-dimensional controllable realization. Check its observability.
- 7.2** Find a three-dimensional observable realization for the transfer function in Problem 7.1. Check its controllability.
- 7.3** Find an uncontrollable and unobservable realization for the transfer function in Problem 7.1. Find also a minimal realization.

- 7.4** Consider the circuit shown in Fig. 2.21(a). Is its two-dimensional state-space equation developed in Problem 2.11 controllable and observable? Is the circuit completely characterized by its transfer function? Find a simpler circuit which is completely characterized by the transfer function.
- 7.5** Consider the circuit shown in Fig. 2.21(b). Is its three-dimensional state-space equation developed in Problem 2.12 controllable and observable? Is the circuit completely characterized by its transfer function? Find a simpler circuit which is completely characterized by the transfer function.
- 7.6** Consider the circuit shown in Fig. 2.22 and studied in Problem 2.13. Compute its transfer function using impedances. Is the circuit completely characterized by its transfer function? Find a simpler circuit which is completely characterized by the transfer function.
- 7.7** Use the Sylvester resultant to find the degree of the transfer function in Problem 7.1.
- 7.8** Use the Sylvester resultant to reduce $(2s - 1)/(4s^2 - 1)$ to a coprime fraction.
- 7.9** Form the Sylvester resultant of $\hat{g}(s) = (s + 2)/(s^2 + 2s)$ by arranging the coefficients of $N(s)$ and $D(s)$ in descending power of s and then search linearly independent columns in order from left to right. Is it true that all D -columns are linearly independent of their LHS columns? Is it true that the degree of $\hat{g}(s)$ equals the number of linearly independent N -columns?
- 7.10** Consider

$$\hat{g}(s) = \frac{\beta_1 s + \beta_2}{s^2 + \alpha_1 s + \alpha_2} =: \frac{N(s)}{D(s)}$$

and its realization

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -\alpha_1 & -\alpha_2 \\ 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t), \quad y(t) = [\beta_1 \ \beta_2] \mathbf{x}(t)$$

Show that the state-space equation is observable if and only if the Sylvester resultant of $D(s)$ and $N(s)$ is nonsingular.

- 7.11** Repeat Problem 7.10 for a transfer function of degree 3 and its controllable-form realization.
- 7.12** Verify Theorem 7.7 for $\hat{g}(s) = 1/(s + 1)^2$.
- 7.13** Use the Markov parameters of $\hat{g}(s) = 1/(s + 1)^2$ to find an irreducible companion-form realization.
- 7.14** Use the Markov parameters of $\hat{g}(s) = 1/(s + 1)^2$ to find an irreducible balanced-form realization.
- 7.15** Consider the two state-space equations

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t), \quad y(t) = [2 \ 2] \mathbf{x}(t)$$

and

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t), \quad y(t) = [2 \ -2] \mathbf{x}(t)$$

Are they minimal realizations? Are they zero-state equivalent? Are they algebraically equivalent?

- 7.16** A simple algorithm was developed by Laub et al. (1987) for computing state-space balancing transformation directly from a state space realization $\{A, B, C\}$. Consult this

reference or the method of section 7.4 to obtain a balanced realization for the following system.

$$\begin{aligned}\dot{x}(t) &= \begin{bmatrix} 0 & 0 & 0 & -150 \\ 1 & 0 & 0 & -245 \\ 0 & 1 & 0 & -113 \\ 0 & 0 & 1 & -19 \end{bmatrix} x(t) + \begin{bmatrix} 4 \\ 1 \\ 0 \\ 0 \end{bmatrix} u(t) \\ y(t) &= [0 \ 0 \ 0 \ 1] x(t)\end{aligned}$$

- 7.17** Find a reduced-order model using the balanced truncation method for the system of the Problem 7.1 (**New Problem**). Obtain the step response of a full order model and three balanced reduced-order models. Which one of the reduced-order model best match the response of the original model? Hint: Use MATLAB functions *balreal* and *balred* as well as its step response.
- 7.18** A reduced order model can also be constructed from a partition balanced realization (7.44) by using

$$\dot{\bar{x}}(t) = \bar{A}\bar{x}(t) + \bar{B}u(t), \quad y(t) = \bar{C}\bar{x}(t)$$

where

$$\bar{A} = A_{11} - A_{12}A_{22}^{-1}A_{21}$$

$$\bar{B} = B_1 - A_{12}A_{22}^{-1}B_2$$

$$\bar{C} = C_1 - C_2A_{22}^{-1}A_{21}$$

Show that if $\{A, B, C\}$ is balanced, then $\{\bar{A}, \bar{B}, \bar{C}\}$ is also balanced.

Apply this technique to the system of Problem 7.1 (**New Problem**) and repeat the task of Problem 7.2 (**New Problem**) for comparison.

- 7.19** Find the characteristic polynomials and the degrees of the following proper rational matrices

$$\hat{G}_1(s) = \begin{bmatrix} \frac{1}{s} & \frac{s+3}{s+1} \\ \frac{1}{s+3} & \frac{s}{s+1} \end{bmatrix}, \quad \hat{G}_2(s) = \begin{bmatrix} \frac{1}{(s+1)^2} & \frac{1}{(s+1)(s+2)} \\ \frac{1}{s+2} & \frac{1}{(s+1)(s+2)} \end{bmatrix}$$

and

$$\hat{G}_3(s) = \begin{bmatrix} \frac{1}{(s+1)^2} & \frac{s+3}{s+2} & \frac{1}{s+5} \\ \frac{1}{(s+3)^2} & \frac{s+1}{s+4} & \frac{1}{s} \end{bmatrix}$$

Note that each entry of $\hat{G}_3(s)$ has different poles from other entries.

- 7.20** Use the left fraction

$$\hat{G}(s) = \begin{bmatrix} s & 1 \\ -s & s \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

to form a generalized resultant as in (7.83), and then search its linearly independent columns in order from left to right. What is the number of linearly independent N -columns? What is the degree of $\hat{G}(s)$? Find a right coprime fraction of $\hat{G}(s)$. Is the given left fraction coprime?

- 7.21** Are all D -columns in the generalized resultant in Problem 7.20 linearly independent of their LHS columns? Now in forming the generalized resultant, the coefficient matrices of $D(s)$ and $N(s)$ are arranged in descending power of s , instead of ascending power of s as in Problem 7.20. Is it true that all D -columns are linearly independent of their LHS

columns? Does the degree of $\hat{\mathbf{G}}(s)$ equal the number of linearly independent N -columns? Does Theorem 7.M4 hold?

- 7.22** Use the right coprime fraction of $\hat{\mathbf{G}}(s)$ obtained in Problem 7.20 to form a generalized resultant as in (7.89), search its linearly independent rows in order from top to bottom, and then find a left coprime fraction of $\hat{\mathbf{G}}(s)$.
- 7.23** Find a right coprime fraction of

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} \frac{s^2+1}{s^3} & \frac{2s+1}{s^2} \\ \frac{s+2}{s^2} & \frac{2}{s} \end{bmatrix}$$

and then a minimal realization.

- 7.24** Verify that the coprime fraction in Example 7.9.1 or

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -6s - 12 & -9 \\ 0.5 & 1 \end{bmatrix} \begin{bmatrix} s^2 + 2.5s + 1 & 2s + 1 \\ 0 & s + 2 \end{bmatrix}^{-1}$$

equals the coprime fraction in (7.86).

- 7.25** Find a gcd of

$$N(s) = [-1 \quad s^2 + 2s - 1], \quad D(s) = \begin{bmatrix} s & 3s + 1 \\ -1 & s^2 + s - 2 \end{bmatrix}$$

corresponding to MFD representation of a system given by $\hat{G}(s) = N(s)D^{-1}(s)$ and obtain the coprime factors $\{\bar{N}(s), \bar{D}(s)\}$.

- 7.26** Consider the following transfer function matrices

$$\hat{G}_1(s) = \begin{bmatrix} \frac{s}{s+2} & 0 & \frac{s+1}{s+2} \\ 0 & \frac{s+1}{s^2} & \frac{1}{s} \end{bmatrix}, \quad \hat{G}_2(s) = \begin{bmatrix} \frac{s-1}{s^2-1} & \frac{1}{s-1} \\ \frac{s^2}{s^2-1} & \frac{2}{s-1} \end{bmatrix}$$

(a) Find a left MFD for $\hat{G}_1(s)$ and a right MFD for $\hat{G}_2(s)$ such that $D_1(s)$ and $D_2(s)$ are diagonal.

(b) Suppose the term $s - 1$ in (1,1) element of $\hat{G}_2(s)$ is cancelled before a right MFD is obtained. What is the difference between two MFDs?

- 7.27** The transfer function matrix of a system is given by $\hat{G}(s) = N(s)D^{-1}(s)$. Let the gcd relationship of $N(s)$ and $D(s)$ be given by

$$\begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} \begin{bmatrix} D \\ N \end{bmatrix} = \begin{bmatrix} R \\ 0 \end{bmatrix} \text{ or } \begin{bmatrix} D \\ N \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} \begin{bmatrix} R \\ 0 \end{bmatrix}$$

Prove that u_{22} is nonsingular and $\hat{G}(s)$ can be represented by an irreducible MFD as

$$\hat{G}(s) = \bar{D}(s)^{-1} \bar{N}(s) = -u_{22}^{-1}(s) u_{21}(s)$$

- 7.28** Find the poles of the transfer function matrix

$$\hat{G}(s) = \begin{bmatrix} \frac{s}{s+1} & \frac{1}{s+2} & 0 \\ 1 & \frac{s}{(s+1)(s+2)} & s \\ 0 & 0 & 1 \end{bmatrix}$$

- 7.29** The poles and zeros of transfer function matrix $\hat{G}(s)$ can also be obtained by smith-McMillan form of $\hat{G}(s)$. Consult the literature and obtain poles and zeros of the system given in Problem 7.7 (**New Problem**).

- 7.30** Find a controller type realization of

$$\begin{aligned}\hat{G}(s) &= \begin{bmatrix} \frac{1}{s+1} & \frac{2}{s+1} \\ -1 & \frac{1}{s+2} \end{bmatrix} \\ &= \begin{bmatrix} s+2 & 2(s+2) \\ -1 & s+1 \end{bmatrix} \begin{bmatrix} (s+1)(s+2) & 0 \\ 0 & (s+1)(s+2) \end{bmatrix}^{-1}\end{aligned}$$

- 7.31** Consider the transfer function $\hat{G}(s) = N(s) D^{-1}(s)$ of Problem 7.9 (**New Problem**). Show that a gcd by $N(s)$ and $D(s)$ is given by

$$R(s) = \begin{bmatrix} 1 & 1 \\ 0 & s+2 \end{bmatrix}$$

and obtain an irreducible realization for $\hat{G}(s)$.

State Feedback and State Estimators

8.1 Introduction

The concepts of controllability and observability were used in the preceding two chapters to study the internal structure of systems and to establish the relationships between the internal and external descriptions. Now we discuss their implications in the design of feedback control systems.

Most control systems can be formulated as shown in Fig. 8.1, in which the *plant* and the reference signal $r(t)$ are given. The input $u(t)$ of the plant is called the *actuating signal* or *control signal*. The output $y(t)$ of the plant is called the *plant output* or *controlled signal*. The problem is to design an overall system so that the plant output will track as closely as possible the reference signal $r(t)$. There are two types of control. If the actuating signal $u(t)$ depends only on the reference signal and is independent of the plant output, the control is called an *open-loop* control. If the actuating signal depends on both the reference signal and the plant output, the control is called a *closed-loop* or *feedback* control. The open-loop control is, in general, not satisfactory if there are plant parameter variations and/or there are noise and disturbance around the system.

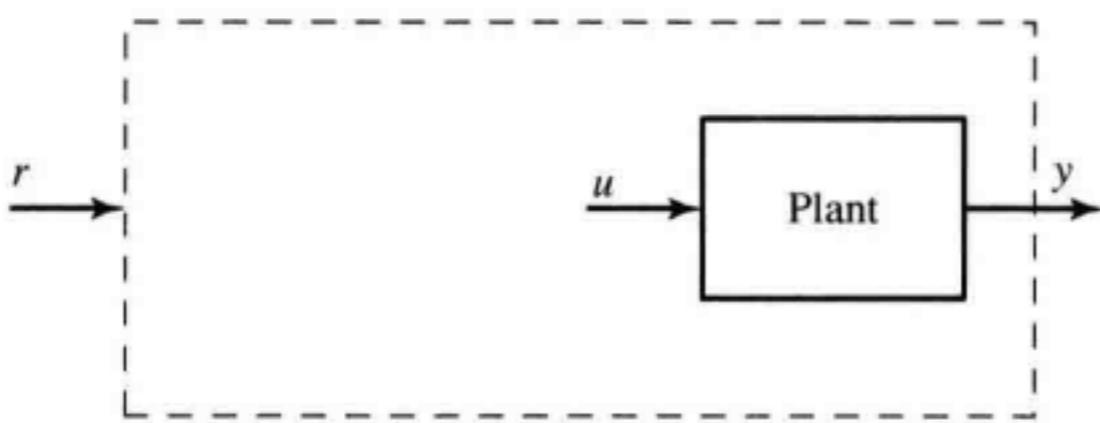


FIGURE 8.1 Design of control systems.

A properly designed feedback system, on the other hand, can reduce the effect of parameter variations and suppress noise and disturbance. Therefore, feedback control is widely used in practice.

This chapter studies designs using state-space equations. Designs using coprime fractions will be studied in the next chapter. We study first SISO systems and then MIMO systems. We study only linear time-invariant and lumped systems.

8.2 State Feedback

Consider the n -dimensional SISO state-space equation

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t) \\ y(t) &= \mathbf{c}\mathbf{x}(t)\end{aligned}\quad (8.1)$$

where we have assumed $d = 0$ to simplify the discussion. In state feedback, the input u is given by

$$u(t) = r(t) - \mathbf{k}\mathbf{x}(t) = r(t) - [k_1 \ k_2 \ \cdots \ k_n]\mathbf{x}(t) = r(t) - \sum_{i=1}^n k_i x_i(t) \quad (8.2)$$

as shown in Fig. 8.2. Each feedback gain k_i is a real constant. This is called the *constant gain negative state feedback* or, simply, *state feedback*. Substituting (8.2) into (8.1) yields

$$\begin{aligned}\dot{\mathbf{x}}(t) &= (\mathbf{A} - \mathbf{b}\mathbf{k})\mathbf{x}(t) + \mathbf{b}r(t) \\ y(t) &= \mathbf{c}\mathbf{x}(t)\end{aligned}\quad (8.3)$$

Theorem 8.1

The pair $(\mathbf{A} - \mathbf{b}\mathbf{k}, \mathbf{b})$, for any $1 \times n$ real constant vector \mathbf{k} , is controllable if and only if (\mathbf{A}, \mathbf{b}) is controllable.



Proof: We show the theorem for $n = 4$. Define

$$\mathcal{C} = [\mathbf{b} \ \mathbf{Ab} \ \mathbf{A}^2\mathbf{b} \ \mathbf{A}^3\mathbf{b}]$$

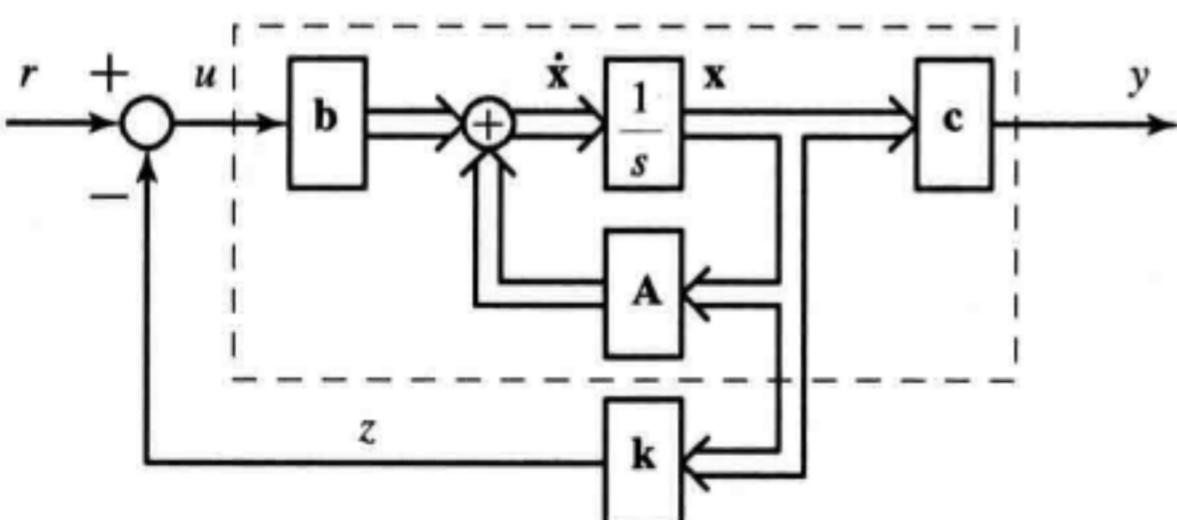


FIGURE 8.2 State feedback.

and

$$\mathcal{C}_f = [\mathbf{b} \ (\mathbf{A} - \mathbf{b}\mathbf{k})\mathbf{b} \ (\mathbf{A} - \mathbf{b}\mathbf{k})^2\mathbf{b} \ (\mathbf{A} - \mathbf{b}\mathbf{k})^3\mathbf{b}]$$

They are the controllability matrices of (8.1) and (8.3). It is straightforward to verify

$$\mathcal{C}_f = \mathcal{C} \begin{bmatrix} 1 & -\mathbf{k}\mathbf{b} & -\mathbf{k}(\mathbf{A} - \mathbf{b}\mathbf{k})\mathbf{b} & -\mathbf{k}(\mathbf{A} - \mathbf{b}\mathbf{k})^2\mathbf{b} \\ 0 & 1 & -\mathbf{k}\mathbf{b} & -\mathbf{k}(\mathbf{A} - \mathbf{b}\mathbf{k})\mathbf{b} \\ 0 & 0 & 1 & -\mathbf{k}\mathbf{b} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (8.4)$$

Note that \mathbf{k} is $1 \times n$ and \mathbf{b} is $n \times 1$. Thus $\mathbf{k}\mathbf{b}$ is scalar; so is every entry in the rightmost matrix in (8.4). Because the right-most matrix is nonsingular for any \mathbf{k} , the rank of \mathcal{C}_f equals the rank of \mathcal{C} . Thus (8.3) is controllable if and only if (8.1) is controllable.

This theorem can also be established directly from the definition of controllability. Let \mathbf{x}_0 and \mathbf{x}_1 be two arbitrary states. If (8.1) is controllable, there exists an input u_1 that transfers \mathbf{x}_0 to \mathbf{x}_1 in a finite time. Now if we choose $r_1 = u_1 + \mathbf{k}\mathbf{x}$, then the input r_1 of the state feedback system will transfer \mathbf{x}_0 to \mathbf{x}_1 . Thus we conclude that if (8.1) is controllable, so is (8.3).

We see from Fig. 8.2 that the input r does not control the state \mathbf{x} directly; it generates u to control \mathbf{x} . Therefore, if u cannot control \mathbf{x} , neither can r . This establishes once again the theorem. Q.E.D.

Although the controllability property is invariant under any state feedback, the observability property is not. This is demonstrated by the example that follows.

Example 8.2.1 Consider the state-space equation

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t) \\ y(t) &= [1 \ 2] \mathbf{x}(t) \end{aligned}$$

The state-space equation can be readily shown to be controllable and observable. Now we introduce the state feedback

$$u(t) = r(t) - [3 \ 1] \mathbf{x}(t)$$

Then the state feedback equation becomes

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} r(t) \\ y(t) &= [1 \ 2] \mathbf{x}(t) \end{aligned}$$

Its controllability matrix is

$$\mathcal{C}_f = \begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix}$$

which is nonsingular. Thus the state feedback equation is controllable. Its observability matrix is

$$\mathcal{O}_f = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$$

which is singular. Thus the state feedback equation is not observable. The reason that the observability property may not be preserved in state feedback will be given later.

We use an example to discuss what can be achieved by state feedback.

Example 8.2.2 Consider a plant described by

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t)$$

The A-matrix has characteristic polynomial

$$\Delta(s) = (s - 1)^2 - 9 = s^2 - 2s - 8 = (s - 4)(s + 2)$$

and, consequently, eigenvalues 4 and -2 . It is unstable. Let us introduce state feedback $u = r - [k_1 \ k_2]\mathbf{x}$. Then the state feedback system is described by

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \left(\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} - \begin{bmatrix} k_1 & k_2 \\ 0 & 0 \end{bmatrix} \right) \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} r(t) \\ &= \begin{bmatrix} 1 - k_1 & 3 - k_2 \\ 3 & 1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} r(t) \end{aligned}$$

This new A-matrix has characteristic polynomial

$$\begin{aligned} \Delta_f(s) &= (s - 1 + k_1)(s - 1) - 3(3 - k_2) \\ &= s^2 + (k_1 - 2)s + (3k_2 - k_1 - 8) \end{aligned}$$

It is clear that the roots of $\Delta_f(s)$ or, equivalently, the eigenvalues of the state feedback system can be placed in any positions by selecting appropriate k_1 and k_2 . For example, if the two eigenvalues are to be placed at $-1 \pm j2$, then the desired characteristic polynomial is $(s + 1 + j2)(s + 1 - j2) = s^2 + 2s + 5$. Equating $k_1 - 2 = 2$ and $3k_2 - k_1 - 8 = 5$ yields $k_1 = 4$ and $k_2 = 17/3$. Thus the state feedback gain $[4 \ 17/3]$ will shift the eigenvalues from $4, -2$ to $-1 \pm j2$.

This example shows that state feedback can be used to place eigenvalues in any desired positions. Moreover, the feedback gain can be computed by direct substitution. This approach, however, can become very involved for three or higher dimensional state-space equations. More seriously, the approach will not reveal how the controllability condition comes into the design. Therefore a more systematic approach is desirable. Before proceeding, we need the following theorem. We state the theorem for $n = 4$; the theorem, however, holds for every positive integer n .

► **Theorem 8.2**

Consider the state-space equation in (8.1) with $n = 4$ and the characteristic polynomial

$$\Delta(s) = \det(s\mathbf{I} - \mathbf{A}) = s^4 + \alpha_1 s^3 + \alpha_2 s^2 + \alpha_3 s + \alpha_4 \quad (8.5)$$

If (8.1) is controllable, then it can be transformed by the transformation $\bar{\mathbf{x}} = \mathbf{Px}$ with

$$\mathbf{Q} := \mathbf{P}^{-1} = [\mathbf{b} \ \mathbf{Ab} \ \mathbf{A}^2\mathbf{b} \ \mathbf{A}^3\mathbf{b}] \begin{bmatrix} 1 & \alpha_1 & \alpha_2 & \alpha_3 \\ 0 & 1 & \alpha_1 & \alpha_2 \\ 0 & 0 & 1 & \alpha_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (8.6)$$

into the controllable form

$$\begin{aligned} \dot{\bar{\mathbf{x}}}(t) &= \bar{\mathbf{A}}\bar{\mathbf{x}}(t) + \bar{\mathbf{b}}u(t) = \begin{bmatrix} -\alpha_1 & -\alpha_2 & -\alpha_3 & -\alpha_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \bar{\mathbf{x}}(t) + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} u(t) \\ y(t) &= \bar{\mathbf{c}}\bar{\mathbf{x}} = [\beta_1 \ \beta_2 \ \beta_3 \ \beta_4]\bar{\mathbf{x}}(t) \end{aligned} \quad (8.7)$$

Furthermore, the transfer function of (8.1) with $n = 4$ equals

$$\hat{g}(s) = \frac{\beta_1 s^3 + \beta_2 s^2 + \beta_3 s + \beta_4}{s^4 + \alpha_1 s^3 + \alpha_2 s^2 + \alpha_3 s + \alpha_4} \quad (8.8)$$

⇒ **Proof:** Let \mathcal{C} and $\bar{\mathcal{C}}$ be the controllability matrices of (8.1) and (8.7). In the SISO case, both \mathcal{C} and $\bar{\mathcal{C}}$ are square. If (8.1) is controllable or \mathcal{C} is nonsingular, so is $\bar{\mathcal{C}}$. And they are related by $\bar{\mathcal{C}} = \mathbf{P}\mathcal{C}$ (Theorem 6.2 and Equation (6.20)). Thus we have

$$\mathbf{P} = \bar{\mathcal{C}}\bar{\mathcal{C}}^{-1} \quad \text{or} \quad \mathbf{Q} := \mathbf{P}^{-1} = \mathcal{C}\bar{\mathcal{C}}^{-1}$$

The controllability matrix $\bar{\mathcal{C}}$ of (8.7) was computed in (7.10). Its inverse turns out to be

$$\bar{\mathcal{C}}^{-1} = \begin{bmatrix} 1 & \alpha_1 & \alpha_2 & \alpha_3 \\ 0 & 1 & \alpha_1 & \alpha_2 \\ 0 & 0 & 1 & \alpha_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (8.9)$$

This can be verified by multiplying (8.9) with (7.10) to yield a unit matrix. Note that the constant term α_4 of (8.5) does not appear in (8.9). Substituting (8.9) into $\mathbf{Q} = \mathcal{C}\bar{\mathcal{C}}^{-1}$ yields (8.6). As shown in Section 7.2, the state-space equation in (8.7) is a realization of (8.8). Thus the transfer function of (8.7) and, consequently, of (8.1) equals (8.8). This establishes the theorem. Q.E.D.

With this theorem, we are ready to discuss eigenvalue assignment by state feedback.

► **Theorem 8.3**

If the n -dimensional state-space equation in (8.1) is controllable, then by state feedback $u(t) = r(t) - \mathbf{k}\mathbf{x}(t)$, where \mathbf{k} is a $1 \times n$ real constant vector, the eigenvalues of $\mathbf{A} - \mathbf{b}\mathbf{k}$ can be arbitrarily assigned, provided that complex conjugate eigenvalues are assigned in pairs.



Proof: We again prove the theorem for $n = 4$. If (8.1) is controllable, it can be transformed into the controllable form in (8.7). Let $\bar{\mathbf{A}}$ and $\bar{\mathbf{b}}$ denote the matrices in (8.7). Then we have $\bar{\mathbf{A}} = \mathbf{P}\mathbf{A}\mathbf{P}^{-1}$ and $\bar{\mathbf{b}} = \mathbf{P}\mathbf{b}$. Substituting $\bar{\mathbf{x}} = \mathbf{P}\mathbf{x}$ into the state feedback yields

$$u(t) = r(t) - \mathbf{k}\mathbf{x}(t) = r(t) - \mathbf{k}\mathbf{P}^{-1}\bar{\mathbf{x}}(t) =: r(t) - \bar{\mathbf{k}}\bar{\mathbf{x}}(t)$$

where $\bar{\mathbf{k}} := \mathbf{k}\mathbf{P}^{-1}$. Because $\bar{\mathbf{A}} - \bar{\mathbf{b}}\bar{\mathbf{k}} = \mathbf{P}(\mathbf{A} - \mathbf{b}\mathbf{k})\mathbf{P}^{-1}$, $\mathbf{A} - \mathbf{b}\mathbf{k}$ and $\bar{\mathbf{A}} - \bar{\mathbf{b}}\bar{\mathbf{k}}$ have the same set of eigenvalues. From any set of desired eigenvalues, we can readily form

$$\Delta_f(s) = s^4 + \bar{\alpha}_1 s^3 + \bar{\alpha}_2 s^2 + \bar{\alpha}_3 s + \bar{\alpha}_4 \quad (8.10)$$

If $\bar{\mathbf{k}}$ is chosen as

$$\bar{\mathbf{k}} = [\bar{\alpha}_1 - \alpha_1 \quad \bar{\alpha}_2 - \alpha_2 \quad \bar{\alpha}_3 - \alpha_3 \quad \bar{\alpha}_4 - \alpha_4] \quad (8.11)$$

then the state feedback equation becomes

$$\dot{\bar{\mathbf{x}}}(t) = (\bar{\mathbf{A}} - \bar{\mathbf{b}}\bar{\mathbf{k}})\bar{\mathbf{x}}(t) + \bar{\mathbf{b}}r(t) = \begin{bmatrix} -\bar{\alpha}_1 & -\bar{\alpha}_2 & -\bar{\alpha}_3 & -\bar{\alpha}_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \bar{\mathbf{x}}(t) + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} r(t)$$

$$y(t) = [\beta_1 \quad \beta_2 \quad \beta_3 \quad \beta_4]\bar{\mathbf{x}}(t) \quad (8.12)$$

Because of the companion form, the characteristic polynomial of $(\bar{\mathbf{A}} - \bar{\mathbf{b}}\bar{\mathbf{k}})$ and, consequently, of $(\mathbf{A} - \mathbf{b}\mathbf{k})$ equals (8.10). Thus the state feedback equation has the set of desired eigenvalues. The feedback gain \mathbf{k} can be computed from

$$\mathbf{k} = \bar{\mathbf{k}}\mathbf{P} = \bar{\mathbf{k}}\bar{\mathcal{C}}\mathcal{C}^{-1} \quad (8.13)$$

with $\bar{\mathbf{k}}$ in (8.11), $\bar{\mathcal{C}}^{-1}$ in (8.9), and $\mathcal{C} = [\mathbf{b} \quad \mathbf{Ab} \quad \mathbf{A}^2\mathbf{b} \quad \mathbf{A}^3\mathbf{b}]$. Q.E.D.

We give an alternative derivation of the formula in (8.11). We compute

$$\begin{aligned} \Delta_f(s) &= \det(s\mathbf{I} - \mathbf{A} + \mathbf{b}\mathbf{k}) = \det((s\mathbf{I} - \mathbf{A})[\mathbf{I} + (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}\mathbf{k}]) \\ &= \det(s\mathbf{I} - \mathbf{A})\det[\mathbf{I} + (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}\mathbf{k}] \end{aligned}$$

which becomes, using (8.5) and (3.64),

$$\Delta_f(s) = \Delta(s)[1 + \mathbf{k}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}]$$

Thus we have

$$\Delta_f(s) - \Delta(s) = \Delta(s)\mathbf{k}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} = \Delta(s)\bar{\mathbf{k}}(s\mathbf{I} - \bar{\mathbf{A}})^{-1}\bar{\mathbf{b}} \quad (8.14)$$

Let z be the output of the feedback gain shown in Fig. 8.2 and let $\bar{\mathbf{k}} = [\bar{k}_1 \bar{k}_2 \bar{k}_3 \bar{k}_4]$. Because the transfer function from u to y in Fig. 8.2 equals

$$\bar{\mathbf{c}}(s\mathbf{I} - \bar{\mathbf{A}})^{-1}\bar{\mathbf{b}} = \frac{\beta_1 s^3 + \beta_2 s^2 + \beta_3 s + \beta_4}{\Delta(s)}$$

the transfer function from u to z should equal

$$\bar{\mathbf{k}}(s\mathbf{I} - \bar{\mathbf{A}})^{-1}\bar{\mathbf{b}} = \frac{\bar{k}_1 s^3 + \bar{k}_2 s^2 + \bar{k}_3 s + \bar{k}_4}{\Delta(s)} \quad (8.15)$$

Substituting (8.15), (8.5), and (8.10) into (8.14) yields

$$(\bar{\alpha}_1 - \alpha_1)s^3 + (\bar{\alpha}_2 - \alpha_2)s^2 + (\bar{\alpha}_3 - \alpha_3)s + (\bar{\alpha}_4 - \alpha_4) = \bar{k}_1 s^3 + \bar{k}_2 s^2 + \bar{k}_3 s + \bar{k}_4$$

This yields (8.11).

Feedback transfer function

Consider a plant described by $(\mathbf{A}, \mathbf{b}, \mathbf{c})$. If (\mathbf{A}, \mathbf{b}) is controllable, $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ can be transformed into the controllable form in (8.7) and its transfer function can then be read out as, for $n = 4$,

$$\hat{g}(s) = \mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} = \frac{\beta_1 s^3 + \beta_2 s^2 + \beta_3 s + \beta_4}{s^4 + \alpha_1 s^3 + \alpha_2 s^2 + \alpha_3 s + \alpha_4} \quad (8.16)$$

After state feedback, the state-space equation becomes $(\mathbf{A} - \mathbf{b}\mathbf{k}, \mathbf{b}, \mathbf{c})$ and is still of the controllable form as shown in (8.12). Thus the feedback transfer function from r to y is

$$\hat{g}_f(s) = \mathbf{c}(s\mathbf{I} - \mathbf{A} + \mathbf{b}\mathbf{k})^{-1}\mathbf{b} = \frac{\beta_1 s^3 + \beta_2 s^2 + \beta_3 s + \beta_4}{s^4 + \bar{\alpha}_1 s^3 + \bar{\alpha}_2 s^2 + \bar{\alpha}_3 s + \bar{\alpha}_4} \quad (8.17)$$

We see that the numerators of (8.16) and (8.17) are the same. In other words, state feedback does not affect the zeros of the plant transfer function. This is actually a general property of feedback: *Feedback can shift the poles of a plant but has no effect on the zeros*. This can be used to explain why a state feedback may alter the observability property of a state equation. If one or more poles are shifted to coincide with zeros of $\hat{g}(s)$, then the numerator and denominator of $\hat{g}_f(s)$ in (8.17) are not coprime. Thus the state-space equation in (8.12) and, equivalently, $(\mathbf{A} - \mathbf{b}\mathbf{k}, \mathbf{c})$ are not observable (Theorem 7.1).

Example 8.2.3 Consider the inverted pendulum studied in Example 6.2.2. Its state-space equation is, as derived in (6.11),

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 5 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 1 \\ 0 \\ -2 \end{bmatrix} u(t) \quad (8.18)$$

$$y(t) = [1 \ 0 \ 0 \ 0] \mathbf{x}(t)$$

It is controllable; thus its eigenvalues can be arbitrarily assigned. Because the A-matrix is block triangular, its characteristic polynomial can be obtained by inspection as

$$\Delta(s) = s^2(s^2 - 5) = s^4 + 0 \cdot s^3 - 5s^2 + 0 \cdot s + 0$$

First we compute \mathbf{P} that will transform (8.18) into the controllable canonical form. Using (8.6), we have

$$\begin{aligned}\mathbf{P}^{-1} &= \mathcal{C}\bar{\mathcal{C}}^{-1} = \begin{bmatrix} 0 & 1 & 0 & 2 \\ 1 & 0 & 2 & 0 \\ 0 & -2 & 0 & -10 \\ -2 & 0 & -10 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & -5 & 0 \\ 0 & 1 & 0 & -5 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 & 0 & -3 \\ 1 & 0 & -3 & 0 \\ 0 & -2 & 0 & 0 \\ -2 & 0 & 0 & 0 \end{bmatrix}\end{aligned}$$

Its inverse is

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 0 & -1/2 \\ 0 & 0 & -1/2 & 0 \\ 0 & -1/3 & 0 & -1/6 \\ -1/3 & 0 & -1/6 & 0 \end{bmatrix}$$

Let the desired eigenvalues be $-1.5 \pm 0.5j$ and $-1 \pm j$. Then we have

$$\begin{aligned}\Delta_f(s) &= (s + 1.5 - 0.5j)(s + 1.5 + 0.5j)(s + 1 - j)(s + 1 + j) \\ &= s^4 + 5s^3 + 10.5s^2 + 11s + 5\end{aligned}$$

Thus we have, using (8.11),

$$\bar{\mathbf{k}} = [5 - 0 \ 10.5 + 5 \ 11 - 0 \ 5 - 0] = [5 \ 15.5 \ 11 \ 5]$$

and

$$\mathbf{k} = \bar{\mathbf{k}}\mathbf{P} = [-5/3 \ -11/3 \ -103/12 \ -13/3] \quad (8.19)$$

This state feedback gain will shift the eigenvalues of the plant from $\{0, 0, \pm j\sqrt{5}\}$ to $\{-1.5 \pm 0.5j, -1 \pm j\}$.

The MATLAB function `place` computes state feedback gains for eigenvalue placement or assignment. For the example, we type

```
a=[0 1 0 0; 0 0 -1 0; 0 0 0 1; 0 0 5 0]; b=[0; 1; 0; -2];
p=[-1.5+0.5j -1.5-0.5j -1+j -1-j];
k=place(a,b,p)
```

which yields $[-1.6667 \ -3.6667 \ -8.5833 \ -4.3333]$. This is the gain in (8.19).

One may wonder at this point how to select a set of desired eigenvalues. This depends on the performance criteria, such as rise time, settling time, overshoot, and

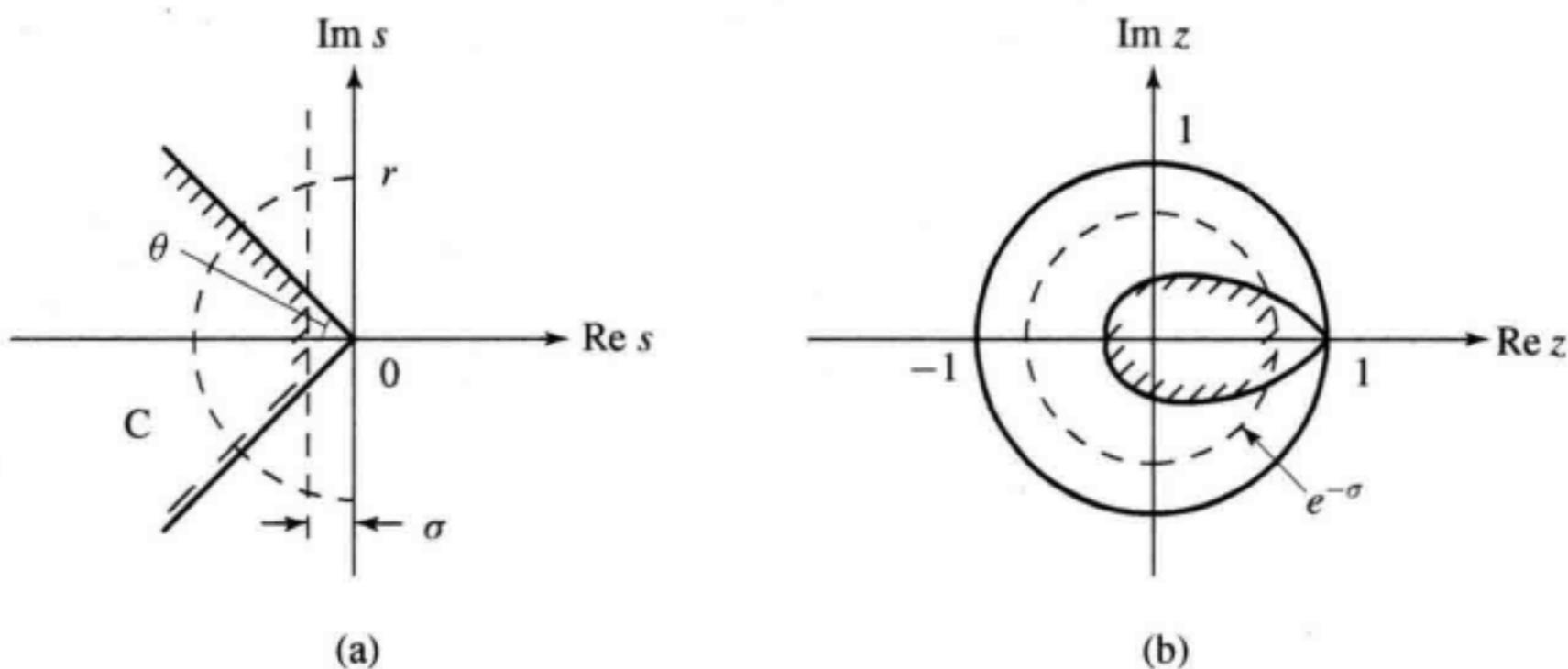


FIGURE 8.3 Desired eigenvalue location.

others, used in the design. Because the response of a system depends not only on poles but also on zeros, the zeros of the plant will also affect the selection. In addition, most physical systems will saturate or burn out if the magnitude of the actuating signal is very large. This will again affect the selection of desired poles. As a guide, we may place all eigenvalues inside the region denoted by C in Fig. 8.3(a). The region is bounded on the right by a vertical line. The larger the distance of the vertical line from the imaginary axis, the faster the response. The region is also bounded by two straight lines emitting from the origin with angle θ . The larger the angle, the larger the overshoot. See Reference 7. If we place all eigenvalues at one point or group them in a very small region, then usually the response will be slow and the actuating signal will be large. Therefore, it is better to place all eigenvalues evenly around a circle with radius r inside the sector as shown. The larger the radius, the faster the response; however, the actuating signal will also be larger. Furthermore, the bandwidth of the feedback system will be larger and the resulting system will be more susceptible to noise. Therefore a final selection may involve compromises among many conflicting requirements. One way to proceed is by computer simulation. Another way is to find the state feedback gain \mathbf{k} to minimize the quadratic performance index

$$J = \int_0^\infty [\mathbf{x}'(t)\mathbf{Q}\mathbf{x}(t) + \mathbf{u}'(t)\mathbf{R}\mathbf{u}(t)] dt$$

See Reference 1. However, selecting \mathbf{Q} and \mathbf{R} requires trial and error. In conclusion, how to select a set of desired eigenvalues is not a simple problem.

We mention that Theorems 8.1 through 8.3—in fact, all theorems to be introduced later in this chapter—apply to the discrete-time case without any modification. The only difference is that the region in Fig. 8.3(a) must be replaced by the one in Fig. 8.3(b) which is obtained by the transformation $z = e^s$.

8.2.1 Solving Lyapunov Equation

This subsection discusses a different method of computing state feedback gain for eigenvalue assignment. The method, however, has the restriction that the selected eigenvalues cannot contain any eigenvalues of \mathbf{A} .

► Procedure 8.1

Consider controllable (\mathbf{A}, \mathbf{b}) where \mathbf{A} is $n \times n$ and \mathbf{b} is $n \times 1$. Find a $1 \times n$ real \mathbf{k} such that $(\mathbf{A} - \mathbf{b}\mathbf{k})$ has any set of desired eigenvalues which contain no eigenvalues of \mathbf{A} .

1. Select an $n \times n$ matrix \mathbf{F} with the set of desired eigenvalues. The form of \mathbf{F} can be arbitrarily chosen and will be discussed later.
2. Select an arbitrary $1 \times n$ vector $\bar{\mathbf{k}}$ such that $(\mathbf{F}, \bar{\mathbf{k}})$ is observable.
3. Solve the unique \mathbf{T} in the Lyapunov equation $\mathbf{AT} - \mathbf{TF} = \bar{\mathbf{b}}\bar{\mathbf{k}}$.
4. Compute the feedback gain $\mathbf{k} = \bar{\mathbf{k}}\mathbf{T}^{-1}$.

We justify first the procedure. If \mathbf{T} is nonsingular, then $\bar{\mathbf{k}} = \mathbf{kT}$ and the Lyapunov equation $\mathbf{AT} - \mathbf{TF} = \bar{\mathbf{b}}\bar{\mathbf{k}}$ implies

$$(\mathbf{A} - \mathbf{b}\mathbf{k})\mathbf{T} = \mathbf{TF} \quad \text{or} \quad \mathbf{A} - \mathbf{b}\mathbf{k} = \mathbf{TF}\mathbf{T}^{-1}$$

Thus $(\mathbf{A} - \mathbf{b}\mathbf{k})$ and \mathbf{F} are similar and have the same set of eigenvalues. Thus the eigenvalues of $(\mathbf{A} - \mathbf{b}\mathbf{k})$ can be arbitrarily assigned except those of \mathbf{A} . As discussed in Section 3.7, if \mathbf{A} and \mathbf{F} have no eigenvalues in common, then a solution \mathbf{T} exists in $\mathbf{AT} - \mathbf{TF} = \bar{\mathbf{b}}\bar{\mathbf{k}}$ for any $\bar{\mathbf{k}}$ and is unique. If \mathbf{A} and \mathbf{F} have common eigenvalues, a solution \mathbf{T} may or may not exist depending on $\bar{\mathbf{b}}\bar{\mathbf{k}}$. To remove this uncertainty, we require \mathbf{A} and \mathbf{F} to have no eigenvalues in common. What remains to be proved is the nonsingularity of \mathbf{T} .

► Theorem 8.4

If \mathbf{A} and \mathbf{F} have no eigenvalues in common, then the unique solution \mathbf{T} of $\mathbf{AT} - \mathbf{TF} = \bar{\mathbf{b}}\bar{\mathbf{k}}$ is nonsingular if and only if (\mathbf{A}, \mathbf{b}) is controllable and $(\mathbf{F}, \bar{\mathbf{k}})$ is observable.



Proof: We prove the theorem for $n = 4$. Let the characteristic polynomial of \mathbf{A} be

$$\Delta(s) = s^4 + \alpha_1 s^3 + \alpha_2 s^2 + \alpha_3 s + \alpha_4 \quad (8.20)$$

Then we have

$$\Delta(\mathbf{A}) = \mathbf{A}^4 + \alpha_1 \mathbf{A}^3 + \alpha_2 \mathbf{A}^2 + \alpha_3 \mathbf{A} + \alpha_4 \mathbf{I} = \mathbf{0}$$

(Cayley–Hamilton theorem). Let us consider

$$\Delta(\mathbf{F}) := \mathbf{F}^4 + \alpha_1 \mathbf{F}^3 + \alpha_2 \mathbf{F}^2 + \alpha_3 \mathbf{F} + \alpha_4 \mathbf{I} \quad (8.21)$$

If $\bar{\lambda}_i$ is an eigenvalue of \mathbf{F} , then $\Delta(\bar{\lambda}_i)$ is an eigenvalue of $\Delta(\mathbf{F})$ (Problem 3.24). Because \mathbf{A} and \mathbf{F} have no eigenvalues in common, we have $\Delta(\bar{\lambda}_i) \neq 0$ for all eigenvalues of \mathbf{F} . Because the determinant of a matrix equals the product of all its eigenvalues, we have

$$\det \Delta(\mathbf{F}) = \prod_i \Delta(\bar{\lambda}_i) \neq 0$$

Thus $\Delta(\mathbf{F})$ is nonsingular.

Substituting $\mathbf{AT} = \mathbf{TF} + \bar{\mathbf{bk}}$ into $\mathbf{A}^2\mathbf{T} - \mathbf{AF}^2$ yields

$$\begin{aligned}\mathbf{A}^2\mathbf{T} - \mathbf{TF}^2 &= \mathbf{A}(\mathbf{TF} + \bar{\mathbf{bk}}) - \mathbf{TF}^2 = \mathbf{Abk} + (\mathbf{AT} - \mathbf{TF})\mathbf{F} \\ &= \mathbf{Abk} + \bar{\mathbf{bk}}\mathbf{F}\end{aligned}$$

Proceeding forward, we can obtain the following set of equations:

$$\begin{aligned}\mathbf{IT} - \mathbf{TI} &= \mathbf{0} \\ \mathbf{AT} - \mathbf{TF} &= \bar{\mathbf{bk}} \\ \mathbf{A}^2\mathbf{T} - \mathbf{TF}^2 &= \mathbf{Abk} + \bar{\mathbf{bk}}\mathbf{F} \\ \mathbf{A}^3\mathbf{T} - \mathbf{TF}^3 &= \mathbf{A}^2\bar{\mathbf{bk}} + \mathbf{Abk}\mathbf{F} + \bar{\mathbf{bk}}\mathbf{F}^2 \\ \mathbf{A}^4\mathbf{T} - \mathbf{TF}^4 &= \mathbf{A}^3\bar{\mathbf{bk}} + \mathbf{A}^2\bar{\mathbf{bk}}\mathbf{F} + \mathbf{Abk}\mathbf{F}^2 + \bar{\mathbf{bk}}\mathbf{F}^3\end{aligned}$$

We multiply the first equation by α_4 , second equation by α_3 , third equation by α_2 , fourth equation by α_1 and the last equation by 1, and then sum them up. After some manipulation, we finally obtain

$$\begin{aligned}\Delta(\mathbf{A})\mathbf{T} - \mathbf{T}\Delta(\mathbf{F}) &= -\mathbf{T}\Delta(\mathbf{F}) \\ &= [\mathbf{b} \mathbf{Ab} \mathbf{A}^2\mathbf{b} \mathbf{A}^3\mathbf{b}] \begin{bmatrix} \alpha_3 & \alpha_2 & \alpha_1 & 1 \\ \alpha_2 & \alpha_1 & 1 & 0 \\ \alpha_1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{\mathbf{k}} \\ \bar{\mathbf{k}}\mathbf{F} \\ \bar{\mathbf{k}}\mathbf{F}^2 \\ \bar{\mathbf{k}}\mathbf{F}^3 \end{bmatrix} \quad (8.22)\end{aligned}$$

where we have used $\Delta(\mathbf{A}) = \mathbf{0}$. If (\mathbf{A}, \mathbf{b}) is controllable and $(\mathbf{F}, \bar{\mathbf{k}})$ is observable, then all three matrices after the last equality are nonsingular. Thus (8.22) and the nonsingularity of $\Delta(\mathbf{F})$ imply that \mathbf{T} is nonsingular. If (\mathbf{A}, \mathbf{b}) is uncontrollable and/or $(\mathbf{F}, \bar{\mathbf{k}})$ is unobservable, then the product of the three matrices is singular. Therefore, \mathbf{T} is singular. This establishes the theorem. Q.E.D.

We now discuss the selection of \mathbf{F} and $\bar{\mathbf{k}}$. Given a set of desired eigenvalues, there are infinitely many \mathbf{F} that have the set of eigenvalues. If we form a polynomial from the set, we can use its coefficients to form a companion-form matrix \mathbf{F} as shown in (7.14). For this \mathbf{F} , we can select $\bar{\mathbf{k}}$ as $[1 \ 0 \ \dots \ 0]$ and $(\mathbf{F}, \bar{\mathbf{k}})$ is observable. If the desired eigenvalues are all distinct, we can also use the modal form discussed in Section 4.4.1. For example, if $n = 5$ and if the five distinct desired eigenvalues are selected as $\lambda_1, \alpha_1 \pm j\beta_1$ and $\alpha_2 \pm j\beta_2$, then we can select \mathbf{F} as

$$\mathbf{F} = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \alpha_1 & \beta_1 & 0 & 0 \\ 0 & -\beta_1 & \alpha_1 & 0 & 0 \\ 0 & 0 & 0 & \alpha_2 & \beta_2 \\ 0 & 0 & 0 & -\beta_2 & \alpha_2 \end{bmatrix} \quad (8.23)$$

It is a block diagonal matrix. For this \mathbf{F} , if $\bar{\mathbf{k}}$ has at least one nonzero entry associated with each diagonal block such as $\bar{\mathbf{k}} = [1 \ 1 \ 0 \ 1 \ 0]$, $\bar{\mathbf{k}} = [1 \ 1 \ 0 \ 0 \ 1]$, or $\bar{\mathbf{k}} = [1 \ 1 \ 1 \ 1 \ 1]$, then $(\mathbf{F}, \bar{\mathbf{k}})$ is observable (Problem 6.16). Thus the first two steps of Procedure 8.1 are

very simple. Once \mathbf{F} and $\bar{\mathbf{k}}$ are selected, we may use the MATLAB function `lyap` to solve the Lyapunov equation in step 3. Thus Procedure 8.1 is easy to carry out as the next example illustrates.

Example 8.2.4 Consider the inverted pendulum studied in Example 8.2.3. The plant state-space equation is given in (8.18) and the desired eigenvalues were chosen as $-1 \pm j$ and $-1.5 \pm 0.5j$. We select \mathbf{F} in modal form as

$$\mathbf{F} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & 0 & -1.5 & 0.5 \\ 0 & 0 & -0.5 & -1.5 \end{bmatrix}$$

and $\bar{\mathbf{k}} = [1 \ 0 \ 1 \ 0]$. We type

```
a=[0 1 0 0;0 0 -1 0;0 0 0 1;0 0 5 0];b=[0;1;0;-2];
f=[-1 1 0 0;-1 -1 0 0;0 0 -1.5 0.5;0 0 -0.5 -1.5];
kb=[1 0 1 0];t=lyap(a,-f,-b*kb);
k=kb*inv(t)
```

The answer is $[-1.6667 \ -3.6667 \ -8.5833 \ -4.3333]$, which is the same as the one obtained by using function `place`. If we use a different $\bar{\mathbf{k}} = [1 \ 1 \ 1 \ 1]$, we will obtain the same \mathbf{k} . Note that the feedback gain is unique for the SISO case.

8.3 Regulation and Tracking

Consider the state feedback system shown in Fig. 8.2. Suppose the reference signal r is zero, and the response of the system is caused by some nonzero initial conditions. The problem is to find a state feedback gain so that the response will die out at a desired rate. This is called a *regulator problem*. This problem may arise when an aircraft is cruising at a fixed altitude H_0 . Now, because of turbulence or other factors, the aircraft may deviate from the desired altitude. Bringing the deviation to zero is a regulator problem. This problem also arises in maintaining the liquid level in Fig. 2.16 at equilibrium.

A closely related problem is the tracking problem. Suppose the reference signal r is a constant or $r(t) = a$ for all $t \geq 0$. The problem is to design an overall system so that $y(t)$ approaches $r(t) = a$ as t approaches infinity. This is called *asymptotic tracking* of a step reference input. It is clear that if $r(t) = a = 0$, then the tracking problem reduces to the regulator problem. Why do we then study these two problems separately? Indeed, if the same state-space equation is valid for all r , designing a system to track asymptotically a step reference input will automatically achieve regulation. However, a linear time-invariant state-space equation is often obtained by shifting to an operating point and then by linearization. The linearized equation is valid only for r very small or zero, thus the study of the regulator problem is needed. We mention that a step reference input can be set by the position of a potentiometer and is, therefore, often referred to as set point. Maintaining a chamber at a desired temperature is often said to regulate the temperature; it is actually to track the desired temperature. Therefore, no sharp distinction is made in practice between regulation and tracking a step reference

input. Tracking a nonconstant reference signal is called a *servomechanism* problem and is a much more difficult problem.

Consider a plant described by $(\mathbf{A}, \mathbf{b}, \mathbf{c})$. If all eigenvalues of \mathbf{A} lie inside the sector shown in Fig. 8.3, then the response caused by any initial conditions will rapidly decay to zero and no state feedback is needed. If \mathbf{A} is stable but some eigenvalues are outside the sector, then the decay may be slow or too oscillatory. If \mathbf{A} is unstable, then the response excited by any nonzero initial conditions will grow unbounded. In these situations, we may introduce state feedback to improve the behavior of the system. Let $u = r - \mathbf{kx}$. Then the state feedback equation becomes $(\mathbf{A} - \mathbf{bk}, \mathbf{b}, \mathbf{c})$ and the response caused by $\mathbf{x}(0)$ is

$$y(t) = \mathbf{c}e^{(\mathbf{A}-\mathbf{bk})t}\mathbf{x}(0)$$

If all eigenvalues of $(\mathbf{A} - \mathbf{bk})$ lie inside the sector in Fig. 8.3, then the output will rapidly decay to zero. Thus, regulation can be easily achieved by introducing state feedback.

The tracking problem is slightly more complex. In general, in addition to state feedback, we need a feedforward gain p as

$$u(t) = pr(t) - \mathbf{kx}(t)$$

Then the transfer function from r to y differs from the one in (8.17) only by the feedforward gain p . Thus we have

$$\hat{g}_f(s) = \frac{\hat{y}(s)}{\hat{r}(s)} = p \frac{\beta_1 s^3 + \beta_2 s^2 + \beta_3 s + \beta_4}{s^4 + \bar{\alpha}_1 s^3 + \bar{\alpha}_2 s^2 + \bar{\alpha}_3 s + \bar{\alpha}_4} \quad (8.24)$$

If (\mathbf{A}, \mathbf{b}) is controllable, all eigenvalues of $(\mathbf{A} - \mathbf{bk})$ or, equivalently, all poles of $\hat{g}_f(s)$ can be arbitrarily assigned, in particular, assigned to lie inside the sector in Fig. 8.3. Under this assumption, if the reference input is a step function with amplitude a , then the output $y(t)$ will approach the constant $\hat{g}_f(0) \cdot a$ as $t \rightarrow \infty$ (Theorem 5.2). Thus in order for $y(t)$ to track asymptotically any step reference input, we need

$$1 = \hat{g}_f(0) = p \frac{\beta_4}{\bar{\alpha}_4} \quad \text{or} \quad p = \frac{\bar{\alpha}_4}{\beta_4} \quad (8.25)$$

which requires $\beta_4 \neq 0$. From (8.16) and (8.17), we see that β_4 is the numerator constant term of the plant transfer function. Thus $\beta_4 \neq 0$ if and only if the plant transfer function $\hat{g}(s)$ has no zero at $s = 0$. In conclusion, if a plant transfer function has no zero at the origin, we can introduce state feedback to stabilize the system and then introduce a feedforward gain as in (8.25). Then the resulting system will track asymptotically any step reference input.

We summarize the preceding discussion. Given $(\mathbf{A}, \mathbf{b}, \mathbf{c})$. If (\mathbf{A}, \mathbf{b}) is controllable, we may introduce state feedback to place the eigenvalues of $(\mathbf{A} - \mathbf{bk})$ in any desired positions and the resulting system will achieve regulation. If (\mathbf{A}, \mathbf{b}) is controllable and if $\mathbf{c}(\mathbf{sI} - \mathbf{A})^{-1}\mathbf{b}$ has no zero at $s = 0$, then after state feedback, we may introduce a feedforward gain as in (8.25). Then the resulting system can track asymptotically any step reference input.

8.3.1 Robust Tracking and Disturbance Rejection*

The state-space equation and transfer function developed to describe a plant may change due to change of load, environment, or aging. Thus plant parameter variations often occur in practice. The equation used in the design is often called the *nominal equation*. The feedforward gain p in (8.25), computed for the nominal plant transfer function, may not yield $\hat{g}_f(0) = 1$ for non-nominal plant transfer functions. Then the output will not track asymptotically any step reference input. Such a tracking is said to be *nonrobust*.

In this subsection we discuss a different design that can achieve robust tracking and disturbance rejection. Consider a plant described by (8.1). We now assume that a constant disturbance w with unknown amplitude enters at the plant input as shown in Fig. 8.4(a). Then the state-space equation must be modified as

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{Ax}(t) + \mathbf{bu}(t) + \mathbf{bw}(t) \\ y(t) &= \mathbf{cx}(t)\end{aligned}\quad (8.26)$$

The problem is to design an overall system so that the output $y(t)$ will track asymptotically any step reference input even with the presence of a disturbance $w(t)$ and with plant parameter variations. This is called *robust tracking and disturbance rejection*. In order to achieve this design, in addition to introducing state feedback, we will introduce an integrator and a unity feedback from the output as shown in Fig. 8.4(a). Let the output of the integrator be denoted by $x_a(t)$, an augmented state variable. Then the

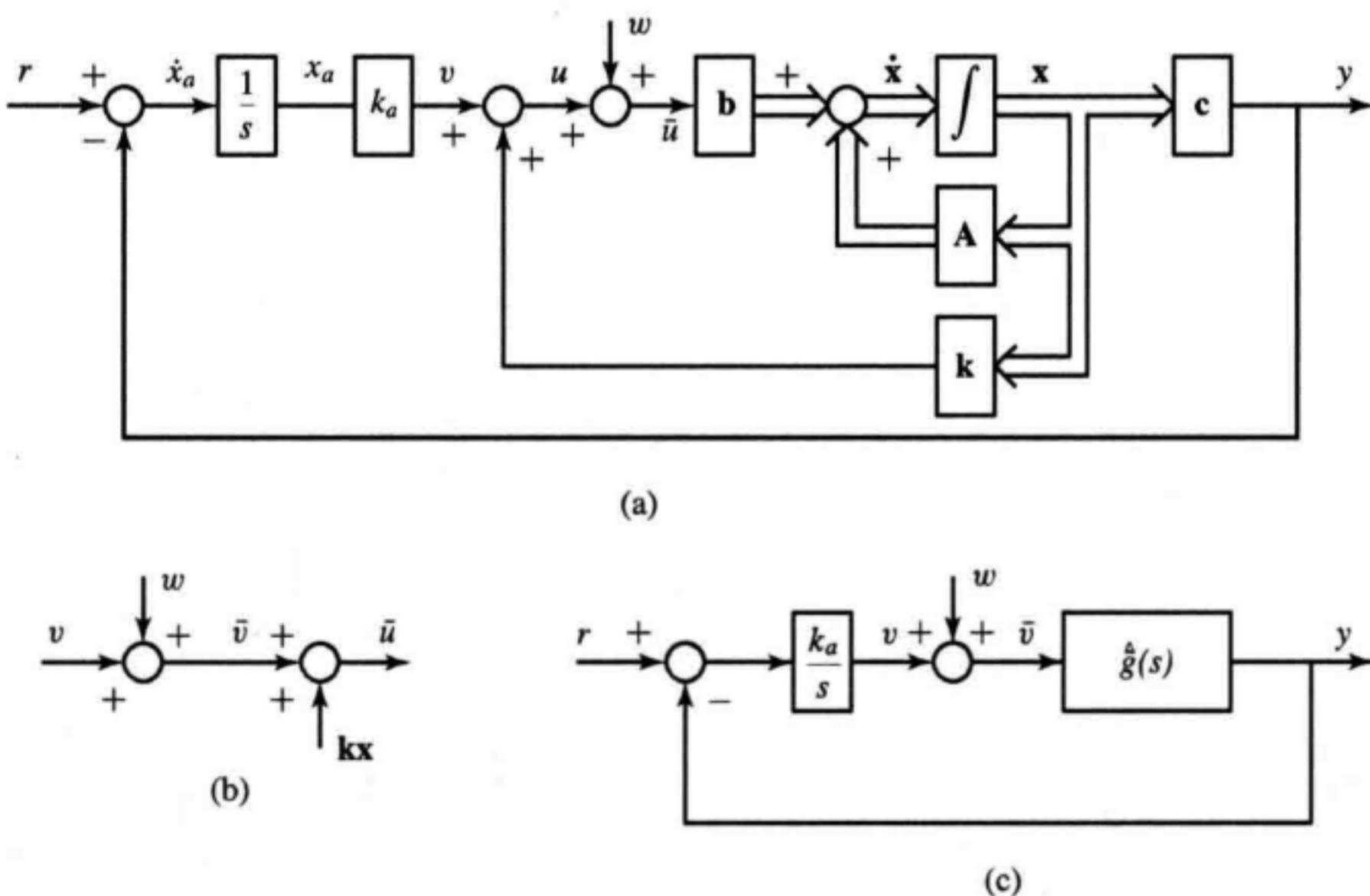


FIGURE 8.4 (a) State feedback with internal model. (b) Interchange of two summers. (c) Transfer-function block diagram.

*This section may be skipped without loss of continuity.

system has the augmented state vector $[\mathbf{x}' \ x_a]'$. From Fig. 8.4(a), we have

$$\dot{x}_a(t) = r(t) - y(t) = r(t) - \mathbf{c}\mathbf{x}(t) \quad (8.27)$$

$$u(t) = [\mathbf{k} \ k_a] \begin{bmatrix} \mathbf{x}(t) \\ x_a(t) \end{bmatrix} \quad (8.28)$$

For convenience, the state is fed back positively to u as shown. Substituting these into (8.26) yields

$$\begin{bmatrix} \dot{\mathbf{x}}(t) \\ \dot{x}_a(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A} + \mathbf{b}\mathbf{k} & \mathbf{b}k_a \\ -\mathbf{c} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ x_a(t) \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} r(t) + \begin{bmatrix} \mathbf{b} \\ 0 \end{bmatrix} w(t) \quad (8.29)$$

$$y(t) = [\mathbf{c} \ 0] \begin{bmatrix} \mathbf{x}(t) \\ x_a(t) \end{bmatrix}$$

This describes the system in Fig. 8.4(a).

► Theorem 8.5

If (\mathbf{A}, \mathbf{b}) is controllable and if $\hat{g}(s) = \mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}$ has no zero at $s = 0$, then all eigenvalues of the \mathbf{A} -matrix in (8.29) can be arbitrarily assigned by selecting a feedback gain $[\mathbf{k} \ k_a]$.



Proof: We show the theorem for $n = 4$. We assume that \mathbf{A} , \mathbf{b} and \mathbf{c} have been transformed into the controllable form in (8.7) and its transfer function equals (8.8). Then the plant transfer function has no zero at $s = 0$ if and only if $\beta_4 \neq 0$. We now show that the pair

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ -\mathbf{c} & 0 \end{bmatrix} \quad \begin{bmatrix} \mathbf{b} \\ 0 \end{bmatrix} \quad (8.30)$$

is controllable if and only if $\beta_4 \neq 0$. Note that we have assumed $n = 4$, thus the dimension of (8.30) is five because of the additional augmented state variable x_a . The controllability matrix of (8.30) is

$$\begin{bmatrix} \mathbf{b} & \mathbf{Ab} & \mathbf{A}^2\mathbf{b} & \mathbf{A}^3\mathbf{b} & \mathbf{A}^4\mathbf{b} \\ 0 & -\mathbf{cb} & -\mathbf{cAb} & -\mathbf{cA}^2\mathbf{b} & -\mathbf{cA}^3\mathbf{b} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -\alpha_1 & \alpha_1^2 - \alpha_2 & -\alpha_1(\alpha_1^2 - \alpha_2) + \alpha_2\alpha_1 - \alpha_3 & a_{15} \\ 0 & 1 & -\alpha_1 & \alpha_1^2 - \alpha_2 & a_{25} \\ 0 & 0 & 1 & -\alpha_1 & a_{35} \\ 0 & 0 & 0 & 1 & a_{45} \\ 0 & -\beta_1 & \beta_1\alpha_1 - \beta_2 & -\beta_1(\alpha_1^2 - \alpha_2) + \beta_2\alpha_1 - \beta_3 & a_{55} \end{bmatrix}$$

where the last column is not written out to save space. The rank of a matrix will not change by elementary operations. Adding the second row multiplied by β_1 to the last row, and adding the third row multiplied by β_2 to the last row, and adding

the fourth row multiplied by β_3 to the last row yield

$$\left[\begin{array}{cccccc} 1 & -\alpha_1 & \alpha_1^2 - \alpha_2 & -\alpha_1(\alpha_1^2 - \alpha_2) + \alpha_2\alpha_1 - \alpha_3 & a_{15} \\ 0 & 1 & -\alpha_1 & \alpha_1^2 - \alpha_2 & a_{25} \\ 0 & 0 & 1 & -\alpha_1 & a_{35} \\ 0 & 0 & 0 & 1 & a_{45} \\ 0 & 0 & 0 & 0 & -\beta_4 \end{array} \right] \quad (8.31)$$

Its determinant is $-\beta_4$. Thus the matrix is nonsingular if and only if $\beta_4 \neq 0$. In conclusion, if (\mathbf{A}, \mathbf{b}) is controllable and if $\hat{g}(s)$ has no zero at $s = 0$, then the pair in (8.30) is controllable. It follows from Theorem 8.3 that all eigenvalues of the \mathbf{A} -matrix in (8.29) can be arbitrarily assigned by selecting a feedback gain $[\mathbf{k} \ k_a]$. Q.E.D.

We mention that the controllability of the pair in (8.30) can also be explained from pole-zero cancelations. If the plant transfer function has a zero at $s = 0$, then the tandem connection of the integrator, which has transfer function $1/s$, and the plant will involve the pole-zero cancelation of s and the state-space equation describing the connection will not be controllable. On the other hand, if the plant transfer function has no zero at $s = 0$, then there is no pole-zero cancelation and the connection will be controllable.

Consider again (8.29). We assume that a set of $(n + 1)$ desired stable eigenvalues or, equivalently, a desired polynomial $\Delta_f(s)$ of degree $n + 1$ has been selected and the feedback gain $[\mathbf{k} \ k_a]$ has been found such that

$$\Delta_f(s) = \det \begin{bmatrix} s\mathbf{I} - \mathbf{A} - \mathbf{b}\mathbf{k} & -\mathbf{b}k_a \\ \mathbf{c} & s \end{bmatrix} \quad (8.32)$$

Now we show that the output y will track asymptotically and robustly any step reference input $r = a$ and reject any step disturbance with unknown amplitude. Instead of establishing the assertion directly from (8.29), we will develop an equivalent block diagram of Fig. 8.4(a) and then establish the assertion. First we interchange the two summers between v and \bar{u} as shown in Fig. 8.4(b). This is permitted because of $\bar{u} = v + \mathbf{kx} + w$ before and after the interchange. The transfer function from \bar{v} to y is

$$\hat{\bar{g}}(s) := \frac{\bar{N}(s)}{\bar{D}(s)} := \mathbf{c}(s\mathbf{I} - \mathbf{A} - \mathbf{b}\mathbf{k})^{-1}\mathbf{b} \quad (8.33)$$

with $\bar{D}(s) = \det(s\mathbf{I} - \mathbf{A} - \mathbf{b}\mathbf{k})$. Thus Fig. 8.4(a) can be redrawn as shown in Fig. 8.4(c). We next establish the relationship between $\Delta_f(s)$ in (8.32) and $\hat{\bar{g}}(s)$ in (8.33). It is straightforward to verify the following equality:

$$\begin{aligned} & \left[\begin{array}{cc} \mathbf{I} & \mathbf{0} \\ -\mathbf{c}(s\mathbf{I} - \mathbf{A} - \mathbf{b}\mathbf{k})^{-1} & 1 \end{array} \right] \left[\begin{array}{cc} s\mathbf{I} - \mathbf{A} - \mathbf{b}\mathbf{k} & -\mathbf{b}k_a \\ \mathbf{c} & s \end{array} \right] \\ &= \left[\begin{array}{cc} s\mathbf{I} - \mathbf{A} - \mathbf{b}\mathbf{k} & -\mathbf{b}k_a \\ 0 & s + \mathbf{c}(s\mathbf{I} - \mathbf{A} - \mathbf{b}\mathbf{k})^{-1}\mathbf{b}k_a \end{array} \right] \end{aligned}$$

Taking its determinants and using (8.32) and (8.33) yield

$$1 \cdot \Delta_f(s) = \bar{D}(s) \left(s + \frac{\bar{N}(s)}{\bar{D}(s)} k_a \right)$$

which implies

$$\Delta_f(s) = s\bar{D}(s) + k_a\bar{N}(s)$$

This is a key equation.

From Fig. 8.4(c), the transfer function from w to y can be readily computed as

$$\hat{g}_{yw} = \frac{\frac{\bar{N}(s)}{\bar{D}(s)}}{1 + \frac{k_a\bar{N}(s)}{s\bar{D}(s)}} = \frac{s\bar{N}(s)}{s\bar{D}(s) + k_a\bar{N}(s)} = \frac{s\bar{N}(s)}{\Delta_f(s)}$$

If the disturbance is $w(t) = \bar{w}$ for all $t \geq 0$, where \bar{w} is an unknown constant, then $\hat{w}(s) = \bar{w}/s$ and the corresponding output is given by

$$\hat{y}_w(s) = \frac{s\bar{N}(s)}{\Delta_f(s)} \frac{\bar{w}}{s} = \frac{\bar{w}\bar{N}(s)}{\Delta_f(s)} \quad (8.34)$$

Because the pole at $s = 0$ in (8.34) is canceled, all remaining poles of $\hat{y}_w(s)$ are stable poles. Therefore the corresponding time response, for any \bar{w} , will die out as $t \rightarrow \infty$. The only condition to achieve the disturbance rejection is that $\hat{y}_w(s)$ has only stable poles. Thus the rejection still holds even if there are plant parameter variations and variations in the feedforward gain k_a and feedback gain \mathbf{k} so long as the overall system remains stable. Thus the disturbance is suppressed at the output both asymptotically and robustly.

The transfer function from r to y is

$$\hat{g}_{yr}(s) = \frac{\frac{k_a}{s} \frac{\bar{N}(s)}{\bar{D}(s)}}{1 + \frac{k_a}{s} \frac{\bar{N}(s)}{\bar{D}(s)}} = \frac{k_a\bar{N}(s)}{s\bar{D}(s) + k_a\bar{N}(s)} = \frac{k_a\bar{N}(s)}{\Delta_f(s)}$$

We see that

$$\hat{g}_{yr}(0) = \frac{k_a\bar{N}(0)}{0 \cdot \bar{D}(0) + k_a\bar{N}(0)} = \frac{k_a\bar{N}(0)}{k_a\bar{N}(0)} = 1 \quad (8.35)$$

Equation (8.35) holds even there are parameter perturbations in the plant transfer function and the gains. Thus asymptotic tracking of any step reference input is robust. Note that this robust tracking holds even for very large parameter perturbations so long as the overall system remains stable.

We see that the design is achieved by inserting an integrator as shown in Fig. 8.4. The integrator is in fact a model of the step reference input and constant disturbance. Thus it is called the *internal model principle*. This will be discussed further in the next chapter.

8.3.2 Stabilization

If a state equation is controllable, all eigenvalues can be arbitrarily assigned by introducing state feedback. We now discuss the case when the state equation is not controllable. Every uncontrollable state equation can be transformed into

$$\begin{bmatrix} \dot{\bar{\mathbf{x}}}_c \\ \dot{\bar{\mathbf{x}}}_{\bar{c}} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{A}}_c & \bar{\mathbf{A}}_{12} \\ \mathbf{0} & \bar{\mathbf{A}}_{\bar{c}} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}_c \\ \bar{\mathbf{x}}_{\bar{c}} \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{b}}_c \\ \mathbf{0} \end{bmatrix} u \quad (8.36)$$

where $(\mathbf{A}_c, \mathbf{b}_c)$ is controllable (Theorem 6.6). Because the A-matrix is block triangular, the eigenvalues of the original A-matrix is the union of the eigenvalues of \mathbf{A}_c and $\bar{\mathbf{A}}_{\bar{c}}$. If we introduce the state feedback

$$u(t) = r(t) - \mathbf{k}\mathbf{x}(t) = r(t) - \bar{\mathbf{k}}\bar{\mathbf{x}}(t) = r(t) - [\bar{\mathbf{k}}_1 \bar{\mathbf{k}}_2] \begin{bmatrix} \bar{\mathbf{x}}_c(t) \\ \bar{\mathbf{x}}_{\bar{c}}(t) \end{bmatrix}$$

where we have partitioned $\bar{\mathbf{k}}$ as in $\bar{\mathbf{x}}$, then (8.36) becomes

$$\begin{bmatrix} \dot{\bar{\mathbf{x}}}_c(t) \\ \dot{\bar{\mathbf{x}}}_{\bar{c}}(t) \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{A}}_c - \bar{\mathbf{b}}_c \bar{\mathbf{k}}_1 & \bar{\mathbf{A}}_{12} - \bar{\mathbf{b}}_c \bar{\mathbf{k}}_2 \\ \mathbf{0} & \bar{\mathbf{A}}_{\bar{c}} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}_c(t) \\ \bar{\mathbf{x}}_{\bar{c}}(t) \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{b}}_c \\ \mathbf{0} \end{bmatrix} r(t) \quad (8.37)$$

we see that $\bar{\mathbf{A}}_{\bar{c}}$ and, consequently, its eigenvalues are not affected by the state feedback. Thus we conclude that the controllability condition of (\mathbf{A}, \mathbf{b}) in Theorem 8.3 is not only sufficient but also necessary to assign *all* eigenvalues of $(\mathbf{A} - \mathbf{b}\mathbf{k})$ to any desired positions.

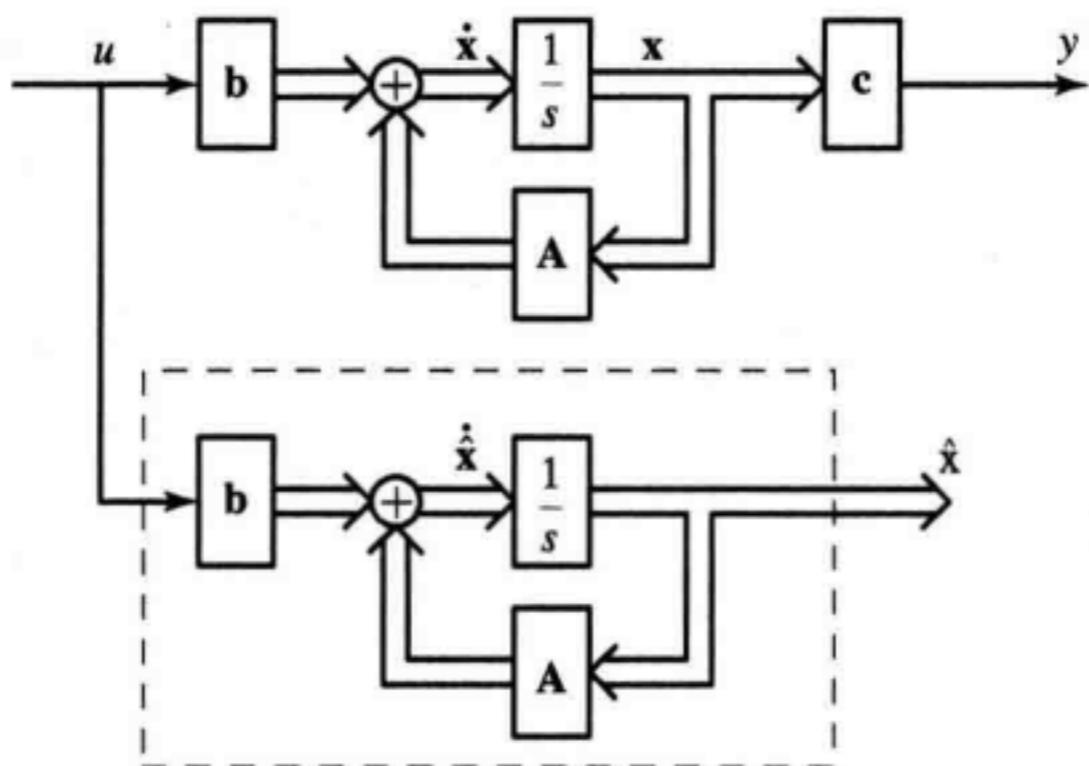
Consider again the state equation in (8.36). If $\bar{\mathbf{A}}_{\bar{c}}$ is stable and if $(\bar{\mathbf{A}}_c, \bar{\mathbf{b}}_c)$ is controllable, then (8.36) is said to be *stabilizable*. We mention that the controllability condition for tracking and disturbance rejection can be replaced by the weaker condition of stabilizability. But in this case, we do not have complete control of the rate of tracking and rejection. If the uncontrollable stable eigenvalues have large imaginary part and are close to the imaginary axis, then the tracking and rejection may not be satisfactory.

8.4 State Estimator

In the preceding sections we introduced state feedback under the implicit assumption that all state variables are available for feedback. This assumption may not hold in practice either because the state variables are not accessible for direct connection or because sensing devices or transducers are not available or very expensive. In this case, in order to apply state feedback, we must design a device, called a *state estimator* or *state observer*, so that the output of the device will generate an estimate of the state. In this section, we introduce full-dimensional state estimators. We use the circumflex over a variable to denote an estimate of the variable. For example, $\hat{\mathbf{x}}$ is an estimate of \mathbf{x} and $\hat{\mathbf{x}}$ is an estimate of $\bar{\mathbf{x}}$.

Consider the n -dimensional state-space equation

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{Ax}(t) + \mathbf{bu}(t) \\ \mathbf{y}(t) &= \mathbf{cx}(t) \end{aligned} \quad (8.38)$$

**FIGURE 8.5** Open-loop state estimator.

where \mathbf{A} , \mathbf{b} , \mathbf{c} are given and the input $u(t)$ and the output $y(t)$ are available to us. The state \mathbf{x} , however, is not available to us. The problem is to estimate \mathbf{x} from u and y with the knowledge of \mathbf{A} , \mathbf{b} , and \mathbf{c} . If we know \mathbf{A} and \mathbf{b} , we can duplicate the original system as

$$\dot{\hat{\mathbf{x}}}(t) = \mathbf{A}\hat{\mathbf{x}}(t) + \mathbf{b}u \quad (8.39)$$

as shown in Fig. 8.5. Note that the original system could be an electromechanical system and the duplicated system could be an op-amp circuit. The duplication will be called an *open-loop* estimator. Now if (8.38) and (8.39) have the same initial state, then for any input we have $\hat{\mathbf{x}}(t) = \mathbf{x}(t)$ for all $t \geq 0$. Therefore, the remaining question is how to find the initial state of (8.38) and then set the initial state of (8.39) to that state. If (8.38) is observable, its initial state $\mathbf{x}(0)$ can be computed from u and y over any time interval, say, $[0, t_1]$. We can then compute the state at t_2 and set $\hat{\mathbf{x}}(t_2) = \mathbf{x}(t_2)$. Then we have $\hat{\mathbf{x}}(t) = \mathbf{x}(t)$ for all $t \geq t_2$. Thus if (8.38) is observable, an open-loop estimator can be used to generate the state vector.

There are, however, two disadvantages in using an open-loop estimator. First, the initial state must be computed and set each time we use the estimator. This is very inconvenient. Second, and more seriously, if the matrix \mathbf{A} has eigenvalues with positive real parts, then even for a very small difference between $\mathbf{x}(t_0)$ and $\hat{\mathbf{x}}(t_0)$ for some t_0 , which may be caused by disturbance or imperfect estimation of the initial state, the difference between $\mathbf{x}(t)$ and $\hat{\mathbf{x}}(t)$ will grow with time. Therefore, the open-loop estimator is, in general, not satisfactory.

We see from Fig. 8.5 that even though the input and output of (8.38) are available, we use only the input to drive the open-loop estimator. Now we shall modify the estimator in Fig. 8.5 to the one in Fig. 8.6, in which the output $y(t) = \mathbf{c}\mathbf{x}(t)$ of (8.38) is compared with $\mathbf{c}\hat{\mathbf{x}}(t)$. Their difference, passing through an $n \times 1$ constant gain vector \mathbf{l} , is used as a correcting term. If the difference is zero, no correction is needed. If the difference is nonzero and if the gain \mathbf{l} is properly designed, the difference will drive the estimated state to the actual state. Such an estimator is called a *closed-loop* or an *asymptotic* estimator or, simply, an estimator.

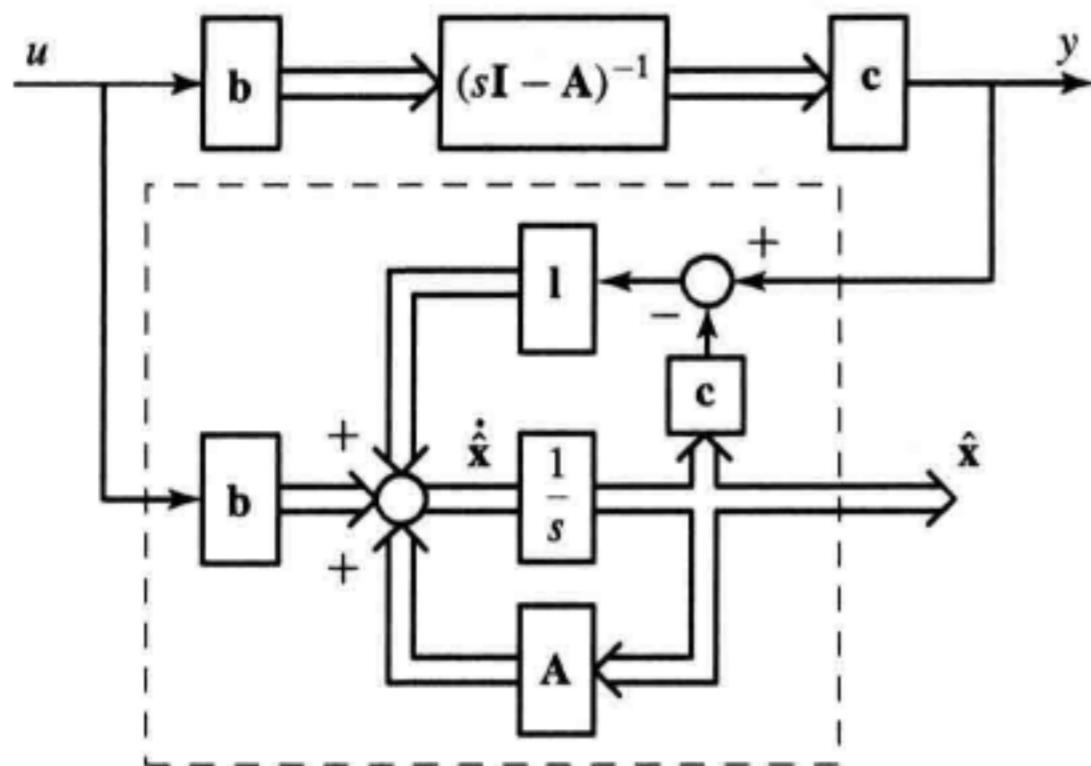


FIGURE 8.6 Closed-loop state estimator.

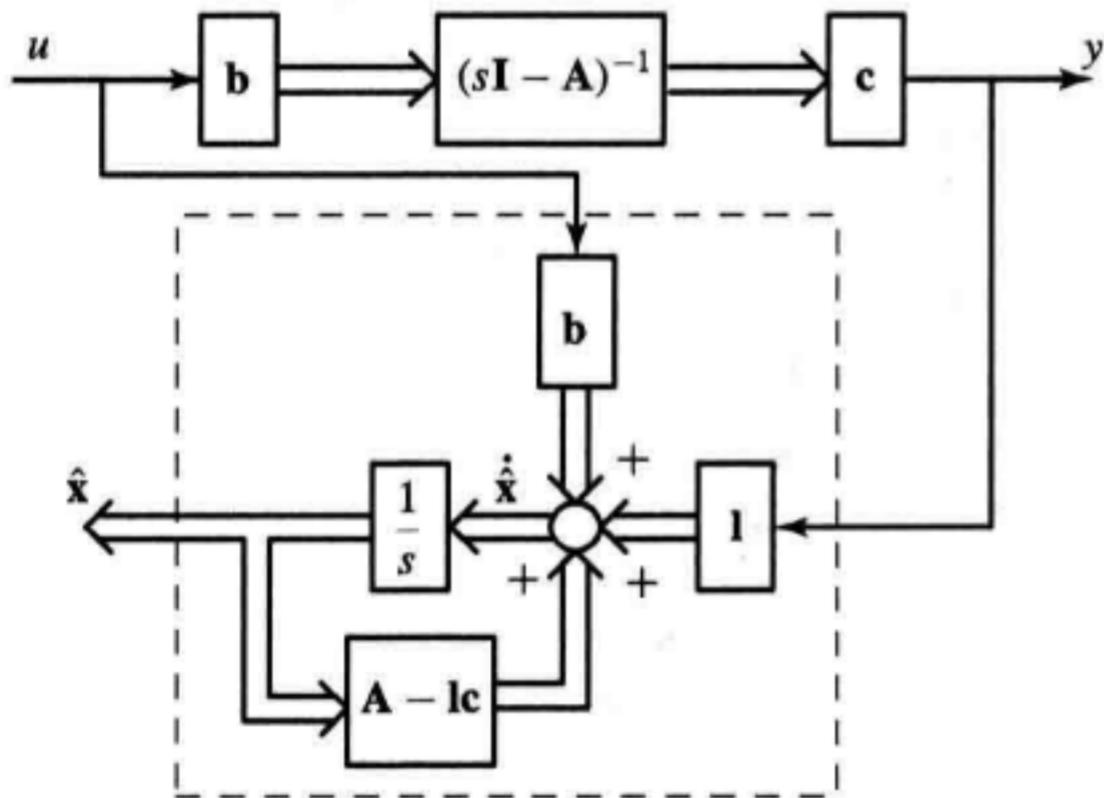


FIGURE 8.7 Closed-loop state estimator.

The open-loop estimator in (8.39) is now modified as, following Fig. 8.6,

$$\dot{\hat{x}}(t) = \mathbf{A}\hat{x}(t) + \mathbf{b}u(t) + \mathbf{l}(y(t) - \mathbf{c}\hat{x}(t))$$

which can be written as

$$\dot{\hat{x}}(t) = (\mathbf{A} - \mathbf{l}\mathbf{c})\hat{x}(t) + \mathbf{b}u(t) + \mathbf{l}y(t) \quad (8.40)$$

and is shown in Fig. 8.7. It has two inputs u and y and its output yields an estimated state \hat{x} . Let us define

$$\mathbf{e}(t) := \mathbf{x}(t) - \hat{x}(t)$$

It is the error between the actual state and the estimated state. Differentiating \mathbf{e} and then substituting (8.38) and (8.40) into it yield

$$\begin{aligned} \dot{\mathbf{e}}(t) &= \dot{\mathbf{x}}(t) - \dot{\hat{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t) - (\mathbf{A} - \mathbf{l}\mathbf{c})\hat{x}(t) - \mathbf{b}u(t) - \mathbf{l}(\mathbf{c}\mathbf{x}(t)) \\ &= (\mathbf{A} - \mathbf{l}\mathbf{c})\mathbf{x}(t) - (\mathbf{A} - \mathbf{l}\mathbf{c})\hat{x}(t) = (\mathbf{A} - \mathbf{l}\mathbf{c})(\mathbf{x}(t) - \hat{x}(t)) \end{aligned}$$

or

$$\dot{\mathbf{e}}(t) = (\mathbf{A} - \mathbf{lc})\mathbf{e}(t) \quad (8.41)$$

This equation governs the estimation error. If all eigenvalues of $(\mathbf{A} - \mathbf{lc})$ can be arbitrarily assigned, then we can control the rate for $\mathbf{e}(t)$ to approach zero or, equivalently, for the estimated state to approach the actual state. For example, if all eigenvalues of $(\mathbf{A} - \mathbf{lc})$ have negative real parts smaller than $-\sigma$, then all entries of \mathbf{e} will approach zero at rates faster than $e^{-\sigma t}$. Thus, even if there is a large error between $\hat{\mathbf{x}}(t_0)$ and $\mathbf{x}(t_0)$ at initial time t_0 , the estimated state will approach the actual state rapidly. Consequently, there is no need to compute the initial state of the original state-space equation. In conclusion, if all eigenvalues of $(\mathbf{A} - \mathbf{lc})$ are properly assigned, a closed-loop estimator is much more desirable than an open-loop estimator.

As in the state feedback, what constitutes the best eigenvalues is not a simple problem. Probably, they should be placed evenly along a circle inside the sector shown in Fig. 8.3. If an estimator is to be used in state feedback, then the estimator eigenvalues should be faster than the desired eigenvalues of the state feedback. Again, saturation and noise problem will impose constraints on the selection. One way to carry out the selection is by computer simulation.

► Theorem 8.03

Consider the pair (\mathbf{A}, \mathbf{c}) . All eigenvalues of $(\mathbf{A} - \mathbf{lc})$ can be arbitrarily assigned by selecting a real constant vector \mathbf{l} if and only if (\mathbf{A}, \mathbf{c}) is observable.

This theorem can be established directly or indirectly by using the duality theorem. The pair (\mathbf{A}, \mathbf{c}) is observable if and only if $(\mathbf{A}', \mathbf{c}')$ is controllable. If $(\mathbf{A}', \mathbf{c}')$ is controllable, all eigenvalues of $(\mathbf{A}' - \mathbf{c}'\mathbf{k})$ can be arbitrarily assigned by selecting a constant gain vector \mathbf{k} . The transpose of $(\mathbf{A}' - \mathbf{c}'\mathbf{k})$ is $(\mathbf{A} - \mathbf{k}'\mathbf{c})$. Thus we have $\mathbf{l} = \mathbf{k}'$. In conclusion, the procedure for computing state feedback gains can be used to compute the gain \mathbf{l} in state estimators.

Solving Lyapunov equation

We discuss a different method of designing a state estimator for the n -dimensional state-space equation

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{Ax}(t) + \mathbf{bu}(t) \\ y(t) &= \mathbf{cx}(t)\end{aligned} \quad (8.42)$$

The method is dual to Procedure 8.1 in Section 8.2.1.

► Procedure 8.01

1. Select an arbitrary $n \times n$ stable matrix \mathbf{F} that has no eigenvalues in common with those of \mathbf{A} .
2. Select an arbitrary $n \times 1$ vector \mathbf{l} such that (\mathbf{F}, \mathbf{l}) is controllable.
3. Solve the unique \mathbf{T} in the Lyapunov equation $\mathbf{TA} - \mathbf{FT} = \mathbf{lc}$. This \mathbf{T} is nonsingular following the dual of Theorem 8.4.

4. Then the state-space equation

$$\dot{\mathbf{z}}(t) = \mathbf{F}\mathbf{z}(t) + \mathbf{T}bu(t) + \mathbf{I}y(t) \quad (8.43)$$

$$\hat{\mathbf{x}}(t) = \mathbf{T}^{-1}\mathbf{z}(t) \quad (8.44)$$

generates an estimate of \mathbf{x} .

We first justify the procedure. Let us define

$$\mathbf{e}(t) := \mathbf{z}(t) - \mathbf{T}\mathbf{x}(t)$$

Then we have, replacing $\mathbf{T}\mathbf{A}$ by $\mathbf{FT} + \mathbf{I}\mathbf{c}$,

$$\begin{aligned}\dot{\mathbf{e}}(t) &= \dot{\mathbf{z}}(t) - \mathbf{T}\dot{\mathbf{x}}(t) = \mathbf{F}\mathbf{z}(t) + \mathbf{T}bu(t) + \mathbf{I}\mathbf{c}\mathbf{x}(t) - \mathbf{T}\mathbf{A}\mathbf{x}(t) - \mathbf{T}bu(t) \\ &= \mathbf{F}\mathbf{z}(t) + \mathbf{I}\mathbf{c}\mathbf{x}(t) - (\mathbf{FT} + \mathbf{I}\mathbf{c})\mathbf{x}(t) = \mathbf{F}(\mathbf{z}(t) - \mathbf{T}\mathbf{x}(t)) = \mathbf{Fe}(t)\end{aligned}$$

If \mathbf{F} is stable, for any $\mathbf{e}(0)$, the error vector $\mathbf{e}(t)$ approaches zero as $t \rightarrow \infty$. Thus $\mathbf{z}(t)$ approaches $\mathbf{T}\mathbf{x}(t)$ or, equivalently, $\mathbf{T}^{-1}\mathbf{z}(t)$ is an estimate of $\mathbf{x}(t)$. All discussion in Section 8.2.1 applies here and will not be repeated.

8.4.1 Reduced-Dimensional State Estimator

Consider the state-space equation in (8.42). If it is observable, then it can be transformed, dual to Theorem 8.2, into the observable form in (7.14). We see that y equals x_1 , the first state variable. Therefore, it is sufficient to construct an $(n - 1)$ -dimensional state estimator to estimate x_i for $i = 2, 3, \dots, n$. This $(n - 1)$ -dimensional estimator with the output equation can then be used to estimate all n state variables.

Reduced-dimensional estimators can be designed by using equivalence transformations or by solving Lyapunov equations. The latter approach is considerably simpler and will be discussed next. For the former approach, the interested reader is referred to Reference pp. 6, 361–363.

Procedure 8.R1

1. Select an arbitrary $(n - 1) \times (n - 1)$ CT stable matrix \mathbf{F} that has no eigenvalues in common with those of \mathbf{A} .
2. Select an arbitrary $(n - 1) \times 1$ vector \mathbf{l} such that (\mathbf{F}, \mathbf{l}) is controllable.
3. Solve the unique \mathbf{T} in the Lyapunov equation $\mathbf{T}\mathbf{A} - \mathbf{FT} = \mathbf{I}\mathbf{c}$. Note that \mathbf{T} is an $(n - 1) \times n$ matrix.
4. Then the $(n - 1)$ -dimensional state-space equation

$$\dot{\mathbf{z}}(t) = \mathbf{F}\mathbf{z}(t) + \mathbf{T}bu(t) + \mathbf{I}y(t) \quad (8.45)$$

$$\hat{\mathbf{x}}(t) = \begin{bmatrix} \mathbf{c} \\ \mathbf{T} \end{bmatrix}^{-1} \begin{bmatrix} y(t) \\ \mathbf{z}(t) \end{bmatrix} \quad (8.46)$$

is an estimate of $\mathbf{x}(t)$.

We first justify the procedure. We write (8.46) as

$$\begin{bmatrix} y(t) \\ z(t) \end{bmatrix} = \begin{bmatrix} \mathbf{c} \\ \mathbf{T} \end{bmatrix} \hat{\mathbf{x}}(t) =: \mathbf{P}\hat{\mathbf{x}}(t)$$

which implies $y(t) = \mathbf{c}\hat{\mathbf{x}}(t)$ and $\mathbf{z}(t) = \mathbf{T}\hat{\mathbf{x}}(t)$. Clearly $y(t)$ is an estimate of $\mathbf{c}\mathbf{x}(t)$. We now show that $\mathbf{z}(t)$ is an estimate of $\mathbf{T}\mathbf{x}(t)$. Define

$$\mathbf{e}(t) = \mathbf{z}(t) - \mathbf{T}\mathbf{x}(t)$$

Then we have

$$\dot{\mathbf{e}}(t) = \dot{\mathbf{z}}(t) - \mathbf{T}\dot{\mathbf{x}}(t) = \mathbf{F}\mathbf{z}(t) + \mathbf{T}\mathbf{b}u(t) + \mathbf{l}\mathbf{c}\mathbf{x}(t) - \mathbf{T}\mathbf{A}\mathbf{x}(t) - \mathbf{T}\mathbf{b}u(t) = \mathbf{F}\mathbf{e}(t)$$

Now if \mathbf{F} is CT stable or every eigenvalue of \mathbf{F} has a negative real part, then $\mathbf{e}(t) \rightarrow \mathbf{0}$ as $t \rightarrow \infty$. Thus \mathbf{z} is an estimate of $\mathbf{T}\mathbf{x}$.

► Theorem 8.6

If \mathbf{A} and \mathbf{F} have no common eigenvalues, then the square matrix

$$\mathbf{P} = \begin{bmatrix} \mathbf{c} \\ \mathbf{T} \end{bmatrix}$$

where \mathbf{T} is the unique solution of $\mathbf{T}\mathbf{A} - \mathbf{F}\mathbf{T} = \mathbf{l}\mathbf{c}$, is nonsingular if and only if (\mathbf{A}, \mathbf{c}) is observable and (\mathbf{F}, \mathbf{l}) is controllable.



Proof: We prove the theorem for $n = 4$. The first part of the proof follows closely the proof of Theorem 8.4. Let

$$\Delta(s) = \det(s\mathbf{I} - \mathbf{A}) = s^4 + \alpha_1 s^3 + \alpha_2 s^2 + \alpha_3 s + \alpha_4$$

Then, dual to (8.22), we have

$$-\mathbf{T}\Delta(\mathbf{F}) = [\mathbf{l} \ \mathbf{F} \ \mathbf{l} \ \mathbf{F}^2 \ \mathbf{l} \ \mathbf{F}^3 \ \mathbf{l}] \begin{bmatrix} \alpha_3 & \alpha_2 & \alpha_1 & 1 \\ \alpha_2 & \alpha_1 & 1 & 0 \\ \alpha_1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \mathbf{c}\mathbf{A} \\ \mathbf{c}\mathbf{A}^2 \\ \mathbf{c}\mathbf{A}^3 \end{bmatrix} \quad (8.47)$$

and $\Delta(\mathbf{F})$ is nonsingular if \mathbf{A} and \mathbf{F} have no common eigenvalues. Note that if \mathbf{A} is 4×4 , then \mathbf{F} is 3×3 . The rightmost matrix in (8.47) is the observability matrix of (\mathbf{A}, \mathbf{c}) and will be denoted by \mathcal{O} . The first matrix after the equality is the controllability matrix of (\mathbf{F}, \mathbf{l}) with one extra column and will be denoted by \mathcal{C}_4 . The middle matrix will be denoted by Λ and is always nonsingular. Using these notations, we write \mathbf{T} as $-\Delta^{-1}(\mathbf{F})\mathcal{C}_4\Lambda\mathcal{O}$ and \mathbf{P} becomes

$$\begin{aligned} \mathbf{P} &= \begin{bmatrix} \mathbf{c} \\ \mathbf{T} \end{bmatrix} = \begin{bmatrix} \mathbf{c} \\ -\Delta^{-1}(\mathbf{F})\mathcal{C}_4\Lambda\mathcal{O} \end{bmatrix} \\ &= \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & -\Delta^{-1}(\mathbf{F}) \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \mathcal{C}_4\Lambda\mathcal{O} \end{bmatrix} \end{aligned} \quad (8.48)$$

Note that if $n = 4$, then \mathbf{P} , \mathcal{O} , and Λ are 4×4 ; \mathbf{T} and \mathcal{C}_4 are 3×4 and $\Delta(\mathbf{F})$ is 3×3 . If (\mathbf{F}, \mathbf{I}) is not controllable, \mathcal{C}_4 has rank at most 2. Thus \mathbf{T} has rank at most 2 and \mathbf{P} is singular. If (\mathbf{A}, \mathbf{c}) is not observable, then there exists a nonzero 4×1 vector \mathbf{r} such that $\mathcal{O}\mathbf{r} = \mathbf{0}$ which implies $\mathbf{c}\mathbf{r} = \mathbf{0}$ and $\mathbf{P}\mathbf{r} = \mathbf{0}$. Thus \mathbf{P} is singular. This shows the necessity of the theorem.

Next we show the sufficiency by contradiction. Suppose \mathbf{P} is singular, then there exists a nonzero vector \mathbf{r} such that $\mathbf{P}\mathbf{r} = \mathbf{0}$ which implies

$$\begin{bmatrix} \mathbf{c} \\ \mathcal{C}_4\Lambda\mathcal{O} \end{bmatrix}\mathbf{r} = \begin{bmatrix} \mathbf{c}\mathbf{r} \\ \mathcal{C}_4\Lambda\mathcal{O}\mathbf{r} \end{bmatrix} = \mathbf{0} \quad (8.49)$$

Define $\mathbf{a} := \Lambda\mathcal{O}\mathbf{r} = [a_1 \ a_2 \ a_3 \ a_4]' =: [\bar{\mathbf{a}} \ a_4]',$ where $\bar{\mathbf{a}}$ represents the first three entries of $\mathbf{a}.$ Expressing it explicitly yields

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} \alpha_3 & \alpha_2 & \alpha_1 & 1 \\ \alpha_2 & \alpha_1 & 1 & 0 \\ \alpha_1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{c}\mathbf{r} \\ \mathbf{c}\mathbf{A}\mathbf{r} \\ \mathbf{c}\mathbf{A}^2\mathbf{r} \\ \mathbf{c}\mathbf{A}^3\mathbf{r} \end{bmatrix} = \begin{bmatrix} x \\ x \\ x \\ \mathbf{c}\mathbf{r} \end{bmatrix}$$

where x denotes entries that are not needed in subsequent discussion. Thus we have $a_4 = \mathbf{c}\mathbf{r}.$ Clearly (8.49) implies $a_4 = \mathbf{c}\mathbf{r} = 0.$ Substituting $a_4 = 0$ into the lower part of (8.49) yields

$$\mathcal{C}_4\Lambda\mathcal{O}\mathbf{r} = \mathcal{C}_4\mathbf{a} = \mathcal{C}\bar{\mathbf{a}} = \mathbf{0} \quad (8.50)$$

where \mathcal{C} is 3×3 and is the controllability matrix of (\mathbf{F}, \mathbf{I}) and $\bar{\mathbf{a}}$ is the first three entries of $\mathbf{a}.$ If (\mathbf{F}, \mathbf{I}) is controllable, then $\mathcal{C}\bar{\mathbf{a}} = \mathbf{0}$ implies $\bar{\mathbf{a}} = \mathbf{0}.$ In conclusion, (8.49) and the controllability of (\mathbf{F}, \mathbf{I}) implies $\mathbf{a} = \mathbf{0}.$

Consider $\Lambda\mathcal{O}\mathbf{r} = \mathbf{a} = \mathbf{0}.$ The matrix Λ is always nonsingular. If (\mathbf{A}, \mathbf{c}) is observable, then \mathcal{O} is nonsingular and $\Lambda\mathcal{O}\mathbf{r} = \mathbf{0}$ implies $\mathbf{r} = \mathbf{0}.$ This contradicts the hypothesis that \mathbf{r} is nonzero. Thus if (\mathbf{A}, \mathbf{c}) is observable and (\mathbf{F}, \mathbf{I}) is controllable, then \mathbf{P} is nonsingular. This establishes Theorem 8.6. Q.E.D.

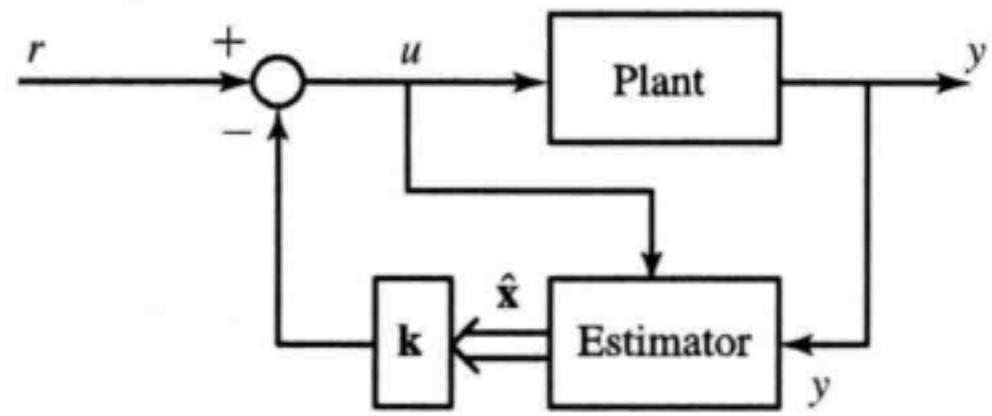
Designing state estimators by solving Lyapunov equations is convenient because the same procedure can be used to design full-dimensional and reduced-dimensional estimators. As we shall see in a later section, the same procedure can also be used to design estimators for MIMO systems.

8.5 Feedback from Estimated States

Consider a plant described by the n -dimensional state-space equation

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{Ax}(t) + \mathbf{bu}(t) \\ y(t) &= \mathbf{cx}(t) \end{aligned} \quad (8.51)$$

If (\mathbf{A}, \mathbf{b}) is controllable, state feedback $u = r - \mathbf{kx}$ can place the eigenvalues of $(\mathbf{A} - \mathbf{bk})$ in any desired positions. If the state variables are not available for feedback, we can design a state estimator. If (\mathbf{A}, \mathbf{c}) is observable, a full- or reduced-dimensional estimator

FIGURE 8.8 Controller-estimator configuration.

with arbitrary eigenvalues can be constructed. We discuss here only full-dimensional estimators. Consider the n -dimensional state estimator

$$\dot{\hat{x}}(t) = (\mathbf{A} - \mathbf{lc})\hat{x}(t) + \mathbf{bu}(t) + \mathbf{ly}(t) \quad (8.52)$$

The estimated state in (8.52) can approach the actual state in (8.51) with any rate by selecting the vector \mathbf{l} .

The state feedback is designed for the state in (8.51). If \mathbf{x} is not available, it is natural to apply the feedback gain to the estimated state as

$$u(t) = r(t) - \mathbf{k}\hat{x}(t) \quad (8.53)$$

as shown in Fig. 8.8. The connection is called the *controller-estimator* configuration. Three questions may be raised in this connection: (1) The eigenvalues of $(\mathbf{A} - \mathbf{bk})$ is obtained from $u = r - \mathbf{kx}$. Do we still have the same set of eigenvalues in using $u = r - \mathbf{k}\hat{x}$? (2) Will the eigenvalues of the estimator \mathbf{F} be affected by the connection? (3) What is the effect of the estimator on the transfer function from r to y ? To answer these questions, we must develop a state-space equation to describe the overall system in Fig. 8.8. Substituting (8.53) into (8.51) and (8.52) yields

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{Ax}(t) - \mathbf{bk}\hat{\mathbf{x}}(t) + \mathbf{br}(t) \\ \dot{\hat{\mathbf{x}}}(t) &= (\mathbf{A} - \mathbf{lc})\hat{\mathbf{x}}(t) + \mathbf{b}(r(t) - \mathbf{k}\hat{\mathbf{x}}(t)) + \mathbf{lc}\mathbf{x}(t) \end{aligned}$$

They can be combined as

$$\begin{aligned} \begin{bmatrix} \dot{\mathbf{x}}(t) \\ \dot{\hat{\mathbf{x}}}(t) \end{bmatrix} &= \begin{bmatrix} \mathbf{A} & -\mathbf{bk} \\ \mathbf{lc} & \mathbf{A} - \mathbf{lc} - \mathbf{bk} \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \hat{\mathbf{x}}(t) \end{bmatrix} + \begin{bmatrix} \mathbf{b} \\ \mathbf{b} \end{bmatrix} r(t) \\ \mathbf{y}(t) &= [\mathbf{c} \ 0] \begin{bmatrix} \mathbf{x}(t) \\ \hat{\mathbf{x}}(t) \end{bmatrix} \end{aligned} \quad (8.54)$$

This $2n$ -dimensional state-space equation describes the feedback system in Fig. 8.8. It is not easy to answer the posed questions from this equation. Let us introduce the following equivalence transformation:

$$\begin{bmatrix} \mathbf{x}(t) \\ \mathbf{e}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{x}(t) - \hat{\mathbf{x}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{I} & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \hat{\mathbf{x}}(t) \end{bmatrix} =: \mathbf{P} \begin{bmatrix} \mathbf{x}(t) \\ \hat{\mathbf{x}}(t) \end{bmatrix}$$

Computing \mathbf{P}^{-1} , which happens to equal \mathbf{P} , and then using (4.26), we can obtain the following equivalent state-space equation

$$\begin{bmatrix} \dot{\mathbf{x}}(t) \\ \dot{\mathbf{e}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A} - \mathbf{b}\mathbf{k} & \mathbf{b}\mathbf{k} \\ \mathbf{0} & \mathbf{A} - \mathbf{l}\mathbf{c} \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{e}(t) \end{bmatrix} + \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} r(t) \quad (8.55)$$

$$y(t) = [\mathbf{c} \ \mathbf{0}] \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{e}(t) \end{bmatrix}$$

The \mathbf{A} -matrix in (8.55) is block triangular; therefore, its eigenvalues are the union of those of $(\mathbf{A} - \mathbf{b}\mathbf{k})$ and $(\mathbf{A} - \mathbf{l}\mathbf{c})$. Thus inserting the state estimator does not affect the eigenvalues of the original state feedback, nor are the eigenvalues of the state estimator affected by the connection. Thus the design of state feedback and the design of state estimator can be carried out independently. This is called the *separation property*.

The state-space equation in (8.55) is of the form shown in (6.40); thus (8.55) is not controllable and the transfer function of (8.55) equals the transfer function of the reduced equation

$$\dot{\mathbf{x}}(t) = (\mathbf{A} - \mathbf{b}\mathbf{k})\mathbf{x}(t) + \mathbf{b}r(t), \quad y(t) = \mathbf{c}\mathbf{x}(t)$$

or

$$\hat{g}_f(s) = \mathbf{c}(s\mathbf{I} - \mathbf{A} + \mathbf{b}\mathbf{k})^{-1}\mathbf{b}$$

(Theorem 6.6). This is the transfer function of the original state feedback system without using a state estimator. Thus the estimator is completely canceled in the transfer function from r to y . This has a simple explanation. In computing transfer functions, all initial states are assumed to be zero. Consequently, we have $\mathbf{x}(0) = \hat{\mathbf{x}}(0) = \mathbf{0}$, which implies $\mathbf{x}(t) = \hat{\mathbf{x}}(t)$ for all t . Thus, as far as the transfer function from r to y is concerned, there is no difference whether a state estimator is employed or not.

8.6 State Feedback—MIMO Case

This section extends state feedback to MIMO systems. Consider a plant described by the n -dimensional p -input state-space equation

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{Ax}(t) + \mathbf{Bu}(t) \\ \mathbf{y}(t) &= \mathbf{Cx}(t) \end{aligned} \quad (8.56)$$

In state feedback, the input \mathbf{u} is given by

$$\mathbf{u}(t) = \mathbf{r}(t) - \mathbf{Kx}(t) \quad (8.57)$$

where \mathbf{K} is a $p \times n$ real constant matrix and \mathbf{r} is a reference signal. Substituting (8.57) into (8.56) yields

$$\begin{aligned} \dot{\mathbf{x}}(t) &= (\mathbf{A} - \mathbf{BK})\mathbf{x}(t) + \mathbf{Br}(t) \\ \mathbf{y}(t) &= \mathbf{Cx}(t) \end{aligned} \quad (8.58)$$

► **Theorem 8.M1**

The pair $(A - BK, B)$, for any $p \times n$ real constant matrix K , is controllable if and only if (A, B) is controllable.

The proof of this theorem follows closely the proof of Theorem 8.1. The only difference is that we must modify (8.4) as

$$\mathcal{C}_f = \mathcal{C} \begin{bmatrix} \mathbf{I}_p & -\mathbf{KB} & -\mathbf{K}(A - BK)\mathbf{B} & -\mathbf{K}(A - BK)^2\mathbf{B} \\ \mathbf{0} & \mathbf{I}_p & -\mathbf{KB} & -\mathbf{K}(A - BK)\mathbf{B} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_p & -\mathbf{KB} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_p \end{bmatrix}$$

where \mathcal{C}_f and \mathcal{C} are $n \times np$ controllability matrices with $n = 4$ and \mathbf{I}_p is the unit matrix of order p . Because the rightmost $(4p \times 4p)$ matrix is nonsingular, \mathcal{C}_f has rank n if and only if \mathcal{C} has rank n . Thus the controllability property is preserved in any state feedback. As in the SISO case, the observability property, however, may not be preserved. Next we extend Theorem 8.3 to the matrix case.

► **Theorem 8.M3**

All eigenvalues of $(A - BK)$ can be arbitrarily assigned (provided complex conjugate eigenvalues are assigned in pairs) by selecting a real constant K if and only if (A, B) is controllable.

If (A, B) is not controllable, then (A, B) can be transformed into the form shown in (8.36) and the eigenvalues of $\bar{A}_{\bar{c}}$ will not be affected by any state feedback. This shows the necessity of the theorem. The sufficiency will be established constructively in the next three subsections.

8.6.1 Cyclic Design

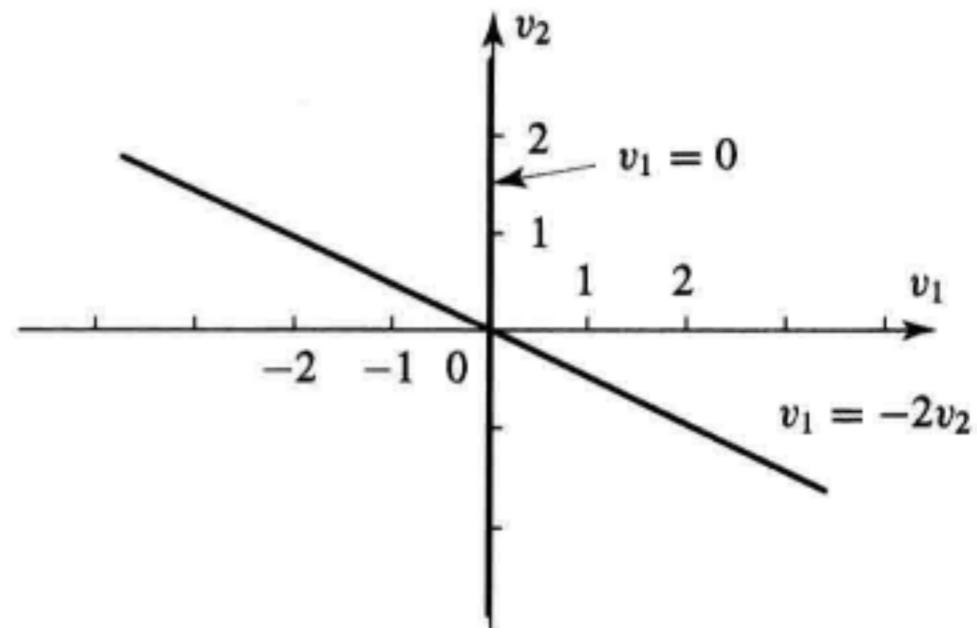
In this design, we change the multi-input problem into a single-input problem and then apply Theorem 8.3. A matrix A is called *cyclic* if its characteristic polynomial equals its minimal polynomial. From the discussion in Section 3.6, we can conclude that A is cyclic if and only if the Jordan form of A has one and only one Jordan block associated with each distinct eigenvalue.

► **Theorem 8.7**

If the n -dimensional p -input pair (A, B) is controllable and if A is cyclic, then for almost any $p \times 1$ vector v , the single-input pair (A, Bv) is controllable.

We argue intuitively the validity of this theorem. Controllability is invariant under any equivalence transformation, thus we may assume A to be in Jordan form. To see the basic idea, we use the following example:

$$A = \begin{bmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 2 \\ 4 & 3 \\ 1 & 0 \end{bmatrix}, \quad Bv = B \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} x \\ x \\ \alpha \\ x \\ \beta \end{bmatrix} \quad (8.59)$$

FIGURE 8.9 Two-dimensional real space.

There is only one Jordan block associated with each distinct eigenvalue; thus \mathbf{A} is cyclic. The condition for (\mathbf{A}, \mathbf{B}) to be controllable is that the third and the last rows of \mathbf{B} are nonzero (Theorem 6.8).

The necessary and sufficient condition for $(\mathbf{A}, \mathbf{B}\mathbf{v})$ to be controllable is $\alpha \neq 0$ and $\beta \neq 0$ in (8.59). Because $\alpha = v_1 + 2v_2$ and $\beta = v_1$, either α or β is zero if and only if $v_1/v_2 = -2/1$ or $v_1 = 0$. Thus any \mathbf{v} other than $v_1 = 0$ and $v_1 = -2v_2$ will make $(\mathbf{A}, \mathbf{B}\mathbf{v})$ controllable. The vector \mathbf{v} can assume any values in the two-dimensional real space shown in Fig. 8.9. The conditions $v_1 = 0$ and $v_1 = -2v_2$ constitute two straight lines as shown. The probability for an arbitrarily selected \mathbf{v} to lie on either straight line is zero. This establishes Theorem 8.6. The cyclicity assumption in this theorem is essential. For example, the pair

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

is controllable (Theorem 6.8). However, there is no \mathbf{v} such that $(\mathbf{A}, \mathbf{B}\mathbf{v})$ is controllable (Corollary 6.8).

If all eigenvalues of \mathbf{A} are distinct, then there is only one Jordan block associated with each eigenvalue. Thus a sufficient condition for \mathbf{A} to be cyclic is that all eigenvalues of \mathbf{A} are distinct.

► Theorem 8.8

If (\mathbf{A}, \mathbf{B}) is controllable, then for almost any $p \times n$ real constant matrix \mathbf{K} , the matrix $(\mathbf{A} - \mathbf{B}\mathbf{K})$ has only distinct eigenvalues and is, consequently, cyclic.

We show intuitively the theorem for $n = 4$. Let the characteristic polynomial of $\mathbf{A} - \mathbf{B}\mathbf{K}$ be

$$\Delta_f(s) = s^4 + a_1s^3 + a_2s^2 + a_3s + a_4$$

where a_i are functions of the entries of \mathbf{K} . The differentiation of $\Delta_f(s)$ with respect to s yields

$$\Delta'_f(s) = 4s^3 + 3a_1s^2 + 2a_2s + a_3$$

If $\Delta_f(s)$ has repeated roots, then $\Delta_f(s)$ and $\Delta'_f(s)$ are not coprime. The necessary and sufficient condition for them to be not coprime is that their Sylvester resultant is singular or

$$\det \begin{bmatrix} a_4 & a_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ a_3 & 2a_2 & a_4 & a_3 & 0 & 0 & 0 & 0 \\ a_2 & 3a_1 & a_3 & 2a_2 & a_4 & a_3 & 0 & 0 \\ a_1 & 4 & a_2 & 3a_1 & a_3 & 2a_2 & a_4 & a_3 \\ 1 & 0 & a_1 & 4 & a_2 & 3a_1 & a_3 & 2a_2 \\ 0 & 0 & 1 & 0 & a_1 & 4 & a_2 & 3a_1 \\ 0 & 0 & 0 & 0 & 1 & 0 & a_1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} = b(k_{ij}) = 0$$

See (7.28). It is clear that all possible solutions of $b(k_{ij}) = 0$ constitute a very small subset of all real k_{ij} . Thus if we select an arbitrary \mathbf{K} , the probability for its entries to meet $b(k_{ij}) = 0$ is 0. Thus all eigenvalues of $\mathbf{A} - \mathbf{BK}$ will be distinct. This establishes the theorem.

With these two theorems, we can now find a \mathbf{K} to place all eigenvalues of $(\mathbf{A} - \mathbf{BK})$ in any desired positions. If \mathbf{A} is not cyclic, we introduce $\mathbf{u} = \mathbf{w} - \mathbf{K}_1 \mathbf{x}$, as shown in Fig. 8.10, such that $\bar{\mathbf{A}} := \mathbf{A} - \mathbf{BK}_1$ in

$$\dot{\mathbf{x}}(t) = (\mathbf{A} - \mathbf{BK}_1)\mathbf{x}(t) + \mathbf{B}\mathbf{w}(t) =: \bar{\mathbf{A}}\mathbf{x}(t) + \mathbf{B}\mathbf{w}(t) \quad (8.60)$$

is cyclic. Because (\mathbf{A}, \mathbf{B}) is controllable, so is $(\bar{\mathbf{A}}, \mathbf{B})$. Thus there exists a $p \times 1$ real vector \mathbf{v} such that $(\bar{\mathbf{A}}, \mathbf{B}\mathbf{v})$ is controllable.¹ Next we introduce another state

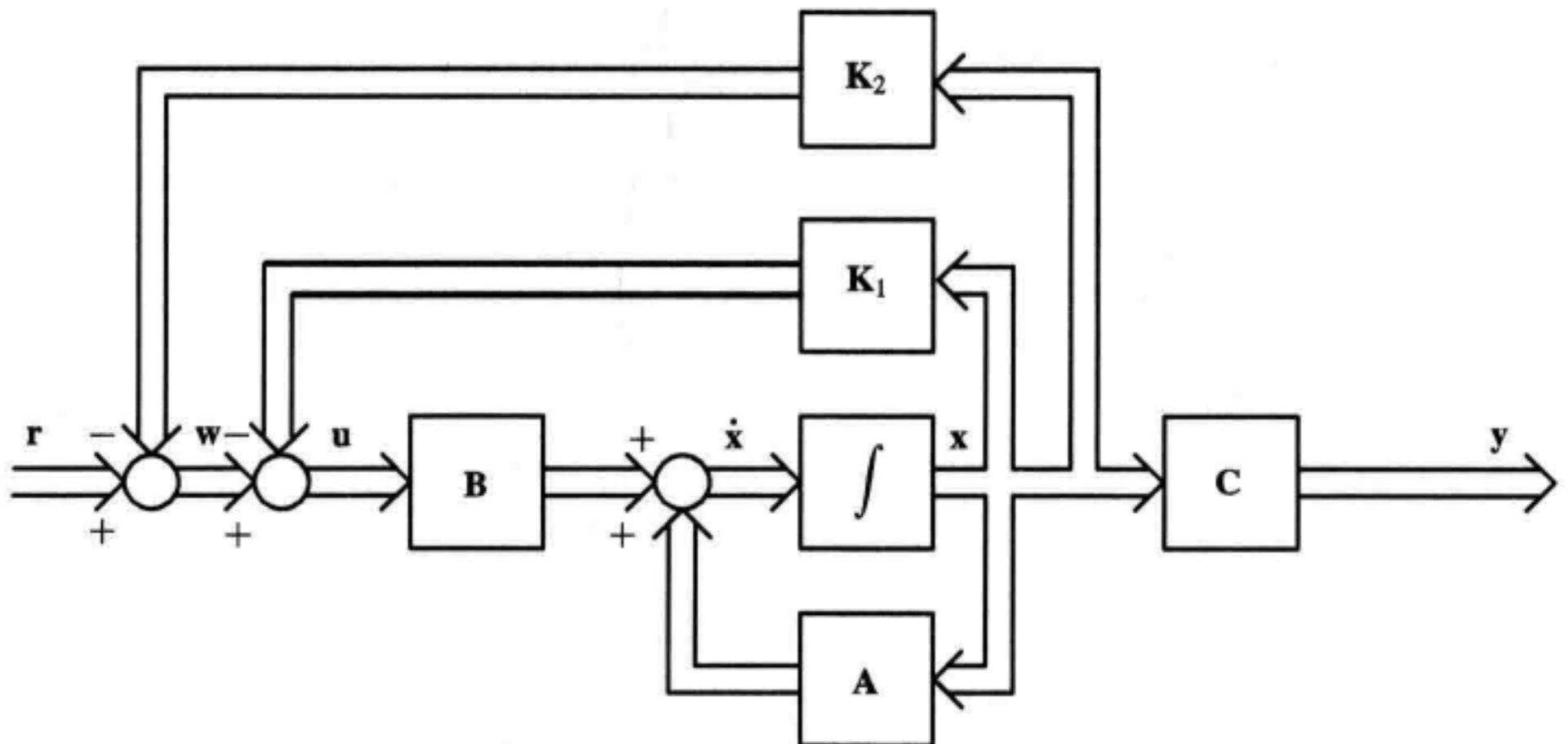


FIGURE 8.10 State feedback by cyclic design.

¹The choices of \mathbf{K}_1 and \mathbf{v} are not unique. They can be chosen arbitrarily, and the probability is 1 that they will meet the requirements. In Theorem 7.5 of Reference 6, a procedure is given to choose \mathbf{K}_1 and \mathbf{v} without any uncertainty. The computation, however, is complicated.

feedback $\mathbf{w} = \mathbf{r} - \mathbf{K}_2\mathbf{x}$ with $\mathbf{K}_2 = \mathbf{v}\mathbf{k}$ where \mathbf{k} is a $1 \times n$ real vector. Then (8.60) becomes

$$\dot{\mathbf{x}}(t) = (\bar{\mathbf{A}} - \mathbf{B}\mathbf{K}_2)\mathbf{x}(t) + \mathbf{Br}(t) = (\bar{\mathbf{A}} - \mathbf{Bv}\mathbf{k})\mathbf{x}(t) + \mathbf{Br}(t)$$

Because the single-input pair $(\bar{\mathbf{A}}, \mathbf{Bv})$ is controllable, the eigenvalues of $(\bar{\mathbf{A}} - \mathbf{Bv}\mathbf{k})$ can be arbitrarily assigned by selecting a \mathbf{k} (Theorem 8.3). Combining the two state-feedback $\mathbf{u} = \mathbf{w} - \mathbf{K}_1\mathbf{x}$ and $\mathbf{w} = \mathbf{r} - \mathbf{K}_2\mathbf{x}$ as

$$\mathbf{u}(t) = \mathbf{r}(t) - (\mathbf{K}_1 + \mathbf{K}_2)\mathbf{x}(t) =: \mathbf{r}(t) - \mathbf{Kx}(t)$$

we obtain a $\mathbf{K} := \mathbf{K}_1 + \mathbf{K}_2$ that achieves arbitrary eigenvalue assignment. This establishes Theorem 8.M3.

8.6.2 Lyapunov-Equation Method

This section extends the procedure of computing feedback gain in Section 8.2.1 to the MIMO case. Consider an n -dimensional p -input pair (\mathbf{A}, \mathbf{B}) . Find a $p \times n$ real constant matrix \mathbf{K} so that $(\mathbf{A} - \mathbf{BK})$ has any set of desired eigenvalues so long as the set does not contain any eigenvalue of \mathbf{A} .

► **Procedure 8.M1**

1. Select an $n \times n$ matrix \mathbf{F} with a set of desired eigenvalues that contains no eigenvalues of \mathbf{A} .
2. Select an arbitrary $p \times n$ matrix $\bar{\mathbf{K}}$ such that $(\mathbf{F}, \bar{\mathbf{K}})$ is observable.
3. Solve the unique \mathbf{T} in the Lyapunov equation $\mathbf{AT} - \mathbf{TF} = \mathbf{B}\bar{\mathbf{K}}$.
4. If \mathbf{T} is singular, select a different $\bar{\mathbf{K}}$ and repeat the process. If \mathbf{T} is nonsingular, we compute $\mathbf{K} = \bar{\mathbf{K}}\mathbf{T}^{-1}$, and $(\mathbf{A} - \mathbf{BK})$ has the set of desired eigenvalues.

If \mathbf{T} is nonsingular, the Lyapunov equation and $\mathbf{KT} = \bar{\mathbf{K}}$ imply

$$(\mathbf{A} - \mathbf{BK})\mathbf{T} = \mathbf{TF} \text{ or } \mathbf{A} - \mathbf{BK} = \mathbf{TFT}^{-1}$$

Thus $(\mathbf{A} - \mathbf{BK})$ and \mathbf{F} are similar and have the same set of eigenvalues. Unlike the SISO case where \mathbf{T} is always nonsingular, the \mathbf{T} here may not be nonsingular even if (\mathbf{A}, \mathbf{B}) is controllable and $(\mathbf{F}, \bar{\mathbf{K}})$ is observable. In other words, the two conditions are necessary but not sufficient for \mathbf{T} to be nonsingular.

► **Theorem 8.M4**

If \mathbf{A} and \mathbf{F} have no eigenvalues in common, then the unique solution \mathbf{T} of $\mathbf{AT} - \mathbf{TF} = \mathbf{B}\bar{\mathbf{K}}$ is nonsingular only if (\mathbf{A}, \mathbf{B}) is controllable and $(\mathbf{F}, \bar{\mathbf{K}})$ is observable.

The proof of Theorem 8.4 applies here except that (8.22) must be modified as, for $n = 4$,

$$-\mathbf{T}\Delta(\mathbf{F}) = [\mathbf{B} \mathbf{A} \mathbf{B} \mathbf{A}^2 \mathbf{B} \mathbf{A}^3 \mathbf{B}] \begin{bmatrix} \alpha_3 \mathbf{I} & \alpha_2 \mathbf{I} & \alpha_1 \mathbf{I} & \mathbf{I} \\ \alpha_2 \mathbf{I} & \alpha_1 \mathbf{I} & \mathbf{I} & \mathbf{0} \\ \alpha_1 \mathbf{I} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{K}} \\ \bar{\mathbf{K}}\mathbf{F} \\ \bar{\mathbf{K}}\mathbf{F}^2 \\ \bar{\mathbf{K}}\mathbf{F}^3 \end{bmatrix}$$

or

$$-\mathbf{T}\Delta(\mathbf{F}) = \mathcal{C}\Sigma\mathcal{O} \quad (8.61)$$

where $\Delta(\mathbf{F})$ is nonsingular and \mathcal{C} , Σ , and \mathcal{O} are, respectively, $n \times np$, $np \times np$ and $np \times n$. If \mathcal{C} or \mathcal{O} has rank less than n , then \mathbf{T} is singular following (3.61). However, the conditions that \mathcal{C} and \mathcal{O} have rank n do not imply the nonsingularity of \mathbf{T} . Thus the controllability of (\mathbf{A}, \mathbf{B}) and observability of $(\mathbf{F}, \bar{\mathbf{K}})$ are only necessary conditions for \mathbf{T} to be nonsingular. This establishes Theorem 8.M3.

Given a controllable (\mathbf{A}, \mathbf{B}) , it is possible to construct an observable $(\mathbf{F}, \bar{\mathbf{K}})$ so that the \mathbf{T} in Theorem 8.M4 is singular. However, after selecting \mathbf{F} , if $\bar{\mathbf{K}}$ is selected randomly and if $(\mathbf{F}, \bar{\mathbf{K}})$ is observable, it is believed that the probability for \mathbf{T} to be nonsingular is 1. Therefore, solving the Lyapunov equation is a viable method of computing a feedback gain matrix to achieve arbitrary eigenvalue assignment. As in the SISO case, we may choose \mathbf{F} in companion form or in modal form as shown in (8.23). If \mathbf{F} is chosen as in (8.23), then we can select $\bar{\mathbf{K}}$ as

$$\bar{\mathbf{K}} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad \text{or} \quad \bar{\mathbf{K}} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

(see Problem 6.16). Once \mathbf{F} and $\bar{\mathbf{K}}$ are chosen, we can then use the MATLAB function `lyap` to solve the Lyapunov equation. Thus the procedure can be easily carried out.

8.6.3 Controllable-Form Method

We introduced in the preceding subsections two methods of computing a feedback gain matrix to achieve arbitrary eigenvalue assignment. The methods are relatively simple; however, they will not reveal the structure of the resulting feedback system. In this subsection, we discuss a different design that will reveal the effect of state feedback on the transfer matrix. We also give a transfer matrix interpretation of state feedback.

In this design, we must transform (\mathbf{A}, \mathbf{B}) into a controllable form. It is an extension of Theorem 8.2 to the MIMO case. Although the basic idea is the same, the procedure can become very involved. Therefore, we will skip the detail and present the final result. To simplify the discussion, we assume that (8.56) has dimension 6, two inputs and two outputs. We first search linearly independent columns of $\mathcal{C} = [\mathbf{B} \mathbf{A}\mathbf{B} \cdots \mathbf{A}^5\mathbf{B}]$ in order

from left to right. It is assumed that its controllability indices are $\mu_1 = 4$ and $\mu_2 = 2$. Then there exists a nonsingular matrix \mathbf{P} and $\bar{\mathbf{x}} = \mathbf{Px}$ will transform (8.56) into the controllable form

$$\dot{\bar{\mathbf{x}}}(t) = \begin{bmatrix} -\alpha_{111} & -\alpha_{112} & -\alpha_{113} & -\alpha_{114} & \vdots & -\alpha_{121} & -\alpha_{122} \\ 1 & 0 & 0 & 0 & \vdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \vdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \vdots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -\alpha_{211} & -\alpha_{212} & -\alpha_{213} & -\alpha_{214} & \vdots & -\alpha_{221} & -\alpha_{222} \\ 0 & 0 & 0 & 0 & \vdots & 1 & 0 \end{bmatrix} \bar{\mathbf{x}}(t)$$

$$+ \begin{bmatrix} 1 & b_{12} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \dots & \dots \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{u}(t)$$

$$\mathbf{y}(t) = \begin{bmatrix} \beta_{111} & \beta_{112} & \beta_{113} & \beta_{114} & \beta_{121} & \beta_{122} \\ \beta_{211} & \beta_{212} & \beta_{213} & \beta_{214} & \beta_{221} & \beta_{222} \end{bmatrix} \bar{\mathbf{x}}(t) \quad (8.62)$$

Note that this form is identical to the one in (7.104).

We now discuss how to find a feedback gain matrix to achieve arbitrary eigenvalue assignment. From a given set of six desired eigenvalues, we can form

$$\Delta_f(s) = (s^4 + \bar{\alpha}_{111}s^3 + \bar{\alpha}_{112}s^2 + \bar{\alpha}_{113}s + \bar{\alpha}_{114})(s^2 + \bar{\alpha}_{221}s + \bar{\alpha}_{222}) \quad (8.63)$$

Let us select $\bar{\mathbf{K}}$ as

$$\bar{\mathbf{K}} = \begin{bmatrix} 1 & b_{12} \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \bar{\alpha}_{111} - \alpha_{111} & \bar{\alpha}_{112} - \alpha_{112} & \bar{\alpha}_{113} - \alpha_{113} \\ \bar{\alpha}_{211} - \alpha_{211} & \bar{\alpha}_{212} - \alpha_{212} & \bar{\alpha}_{213} - \alpha_{213} \\ \bar{\alpha}_{114} - \alpha_{114} & -\alpha_{121} & -\alpha_{122} \\ \bar{\alpha}_{214} - \alpha_{214} & \bar{\alpha}_{221} - \alpha_{221} & \bar{\alpha}_{222} - \alpha_{222} \end{bmatrix} \quad (8.64)$$

Then it is straightforward to verify the following

$$\bar{\mathbf{A}} - \bar{\mathbf{B}}\bar{\mathbf{K}} = \begin{bmatrix} -\bar{\alpha}_{111} & -\bar{\alpha}_{112} & -\bar{\alpha}_{113} & -\bar{\alpha}_{114} & \vdots & 0 & 0 \\ 1 & 0 & 0 & 0 & \vdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \vdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \vdots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -\bar{\alpha}_{211} & -\bar{\alpha}_{212} & -\bar{\alpha}_{213} & -\bar{\alpha}_{214} & \vdots & -\bar{\alpha}_{221} & -\bar{\alpha}_{222} \\ 0 & 0 & 0 & 0 & \vdots & 1 & 0 \end{bmatrix} \quad (8.65)$$

Because $(\bar{\mathbf{A}} - \bar{\mathbf{B}}\bar{\mathbf{K}})$ is block triangular for any $\bar{\alpha}_{21i}$, $i = 1, 2, 3, 4$, its characteristic polynomial equals the product of the characteristic polynomials of the two diagonal blocks of orders 4 and 2. Because the diagonal blocks are of companion form, we conclude that the characteristic polynomial of $(\bar{\mathbf{A}} - \bar{\mathbf{B}}\bar{\mathbf{K}})$ equals the one in (8.63). If $\mathbf{K} = \bar{\mathbf{K}}\mathbf{P}$, then $(\bar{\mathbf{A}} - \bar{\mathbf{B}}\bar{\mathbf{K}}) = \mathbf{P}(\mathbf{A} - \mathbf{B}\mathbf{K})\mathbf{P}^{-1}$. Thus the feedback gain $\mathbf{K} = \bar{\mathbf{K}}\mathbf{P}$ will place the eigenvalues of $(\mathbf{A} - \mathbf{B}\mathbf{K})$ in desired locations. This establishes once again Theorem 8.M3.

Unlike the single-input case where the feedback gain is unique, the feedback gain matrix in the multi-input case is not unique. For example, the $\bar{\mathbf{K}}$ in (8.64) yields a lower block triangular matrix in $(\bar{\mathbf{A}} - \bar{\mathbf{B}}\bar{\mathbf{K}})$. It is possible to select a different $\bar{\mathbf{K}}$ to yield an upper block triangular matrix or a block diagonal matrix. Furthermore, different grouping of (8.63) will again yield different $\bar{\mathbf{K}}$.

8.6.4 Effect on Transfer Matrices²

In the SISO case, state feedback can shift the poles of a plant transfer function $\hat{g}(s)$ to any positions and yet has no effect on the zeros. Or, equivalently, state feedback can change the denominator coefficients, except the leading coefficient 1, to any values but has no effect on the numerator coefficients. Although we can establish a similar result for the MIMO case from (8.62) and (8.65), it is instructive to do so by using the result in Section 7.9. Following the notations in Section 7.9, we express $\hat{\mathbf{G}}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}$ as

$$\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s) \quad (8.66)$$

or

$$\hat{\mathbf{y}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)\hat{\mathbf{u}}(s) \quad (8.67)$$

where $\mathbf{N}(s)$ and $\mathbf{D}(s)$ are right coprime, $\mathbf{D}(s)$ is column reduced. Define

$$\mathbf{D}(s)\hat{\mathbf{v}}(s) = \hat{\mathbf{u}}(s) \quad (8.68)$$

²This subsection may be skipped without loss of continuity. The material in Section 7.9 is needed to study this subsection.

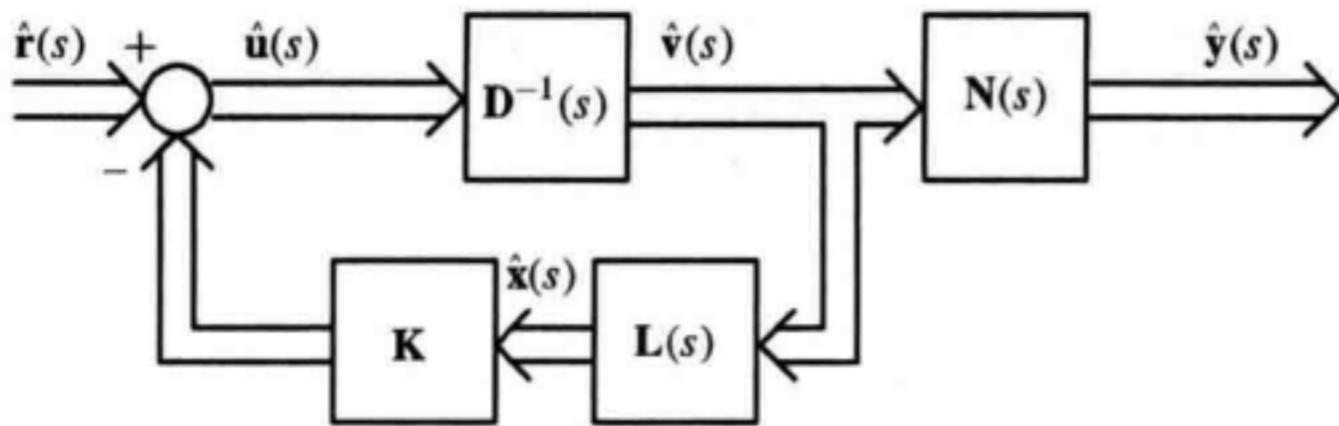


FIGURE 8.11 Transfer matrix interpretation of state feedback.

as in (7.93). Then we have

$$\hat{y}(s) = \mathbf{N}(s)\hat{v}(s) \quad (8.69)$$

Let $\mathbf{H}(s)$ and $\mathbf{L}(s)$ be defined as in (7.91) and (7.92). Then the state vector in (8.62) is

$$\hat{\mathbf{x}}(s) = \mathbf{L}(s)\hat{v}(s)$$

Thus the state feedback becomes, in the Laplace-transform domain,

$$\hat{\mathbf{u}}(s) = \hat{\mathbf{r}}(s) - \mathbf{K}\hat{\mathbf{x}}(s) = \hat{\mathbf{r}}(s) - \mathbf{KL}(s)\hat{v}(s) \quad (8.70)$$

and can be represented as shown in Fig. 8.11.

Let us express $\mathbf{D}(s)$ as

$$\mathbf{D}(s) = \mathbf{D}_{hc}\mathbf{H}(s) + \mathbf{D}_{lc}\mathbf{L}(s) \quad (8.71)$$

Substituting (8.71) and (8.70) into (8.68) yields

$$[\mathbf{D}_{hc}\mathbf{H}(s) + \mathbf{D}_{lc}\mathbf{L}(s)]\hat{v}(s) = \hat{\mathbf{r}}(s) - \mathbf{KL}(s)\hat{v}(s)$$

which implies

$$[\mathbf{D}_{hc}\mathbf{H}(s) + (\mathbf{D}_{lc} + \mathbf{K})\mathbf{L}(s)]\hat{v}(s) = \hat{\mathbf{r}}(s)$$

Substituting this into (8.69) yields

$$\hat{y}(s) = \mathbf{N}(s)[\mathbf{D}_{hc}\mathbf{H}(s) + (\mathbf{D}_{lc} + \mathbf{K})\mathbf{L}(s)]^{-1}\hat{\mathbf{r}}(s)$$

Thus the transfer matrix from \mathbf{r} to \mathbf{y} is

$$\hat{\mathbf{G}}_f(s) = \mathbf{N}(s)[\mathbf{D}_{hc}\mathbf{H}(s) + (\mathbf{D}_{lc} + \mathbf{K})\mathbf{L}(s)]^{-1} \quad (8.72)$$

The state feedback changes the plant transfer matrix $\mathbf{N}(s)\mathbf{D}^{-1}(s)$ to the one in (8.72). We see that the numerator matrix $\mathbf{N}(s)$ is not affected by the state feedback, nor are the column degree $\mathbf{H}(s)$ and the column-degree-coefficient matrix \mathbf{D}_{hc} affected by the state feedback. However, all coefficients associated with $\mathbf{L}(s)$ can be arbitrarily assigned by selecting a \mathbf{K} . This is similar to the SISO case.

It is possible to extend the robust tracking and disturbance rejection discussed in Section 8.3 to the MIMO case. It is, however, simpler to do so by using coprime fractions; therefore, it will not be discussed here.

8.7 State Estimators—MIMO Case

All discussion for state estimators in the SISO case applies to the MIMO case; therefore, the discussion will be brief. Consider the n -dimensional p -input q -output state-space equation

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{Ax}(t) + \mathbf{Bu}(t) \\ \mathbf{y}(t) &= \mathbf{Cx}(t)\end{aligned}\quad (8.73)$$

The problem is to use available input \mathbf{u} and output \mathbf{y} to drive a system whose output gives an estimate of the state \mathbf{x} . We extend (8.40) to the MIMO case as

$$\dot{\hat{\mathbf{x}}}(t) = (\mathbf{A} - \mathbf{LC})\hat{\mathbf{x}}(t) + \mathbf{Bu}(t) + \mathbf{Ly}(t) \quad (8.74)$$

This is a full-dimensional state estimator. Let us define the error vector as

$$\mathbf{e}(t) := \mathbf{x}(t) - \hat{\mathbf{x}}(t) \quad (8.75)$$

Then we have, as in (8.41),

$$\dot{\mathbf{e}}(t) = (\mathbf{A} - \mathbf{LC})\mathbf{e}(t) \quad (8.76)$$

If (\mathbf{A}, \mathbf{C}) is observable, then all eigenvalues of $(\mathbf{A} - \mathbf{LC})$ can be arbitrarily assigned by choosing an \mathbf{L} . Thus the convergence rate for the estimated state $\hat{\mathbf{x}}$ to approach the actual state \mathbf{x} can be as fast as desired. As in the SISO case, the three methods of computing state feedback gain \mathbf{K} can be applied here to compute \mathbf{L} .

Next we discuss reduced-dimensional state estimators. The next procedure is an extension of Procedure 8.R1 to the MIMO case.

► Procedure 8.MR1

Consider the n -dimensional q -output observable pair (\mathbf{A}, \mathbf{C}) . It is assumed that \mathbf{C} has rank q .

1. Select an arbitrary $(n - q) \times (n - q)$ stable matrix \mathbf{F} which has no eigenvalues in common with those of \mathbf{A} .
2. Select an arbitrary $(n - q) \times q$ matrix \mathbf{L} such that (\mathbf{F}, \mathbf{L}) is controllable.
3. Solve the unique $(n - q) \times n$ \mathbf{T} in the Lyapunov equation $\mathbf{TA} - \mathbf{FT} = \mathbf{LC}$.
4. If the square matrix of order n

$$\mathbf{P} = \begin{bmatrix} \mathbf{C} \\ \mathbf{T} \end{bmatrix} \quad (8.77)$$

is singular, go back to Step (2) and repeat the process. If \mathbf{P} is nonsingular, then the $(n - q)$ -dimensional state-space equation

$$\dot{\mathbf{z}}(t) = \mathbf{Fz}(t) + \mathbf{TBu}(t) + \mathbf{Ly}(t) \quad (8.78)$$

$$\hat{\mathbf{x}}(t) = \begin{bmatrix} \mathbf{C} \\ \mathbf{T} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{z}(t) \end{bmatrix} \quad (8.79)$$

generates an estimate of $\mathbf{x}(t)$.

We first justify the procedure. We write (8.79) as

$$\begin{bmatrix} \mathbf{y}(t) \\ \mathbf{z}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{C} \\ \mathbf{T} \end{bmatrix} \hat{\mathbf{x}}(t)$$

which implies $\mathbf{y} = \mathbf{C}\hat{\mathbf{x}}$ and $\mathbf{z} = \mathbf{T}\hat{\mathbf{x}}$. Clearly, \mathbf{y} is an estimate of \mathbf{Cx} . We now show that \mathbf{z} is an estimate of \mathbf{Tx} . Let us define

$$\mathbf{e}(t) := \mathbf{z}(t) - \mathbf{T}\mathbf{x}(t)$$

Then we have

$$\begin{aligned} \dot{\mathbf{e}}(t) &= \dot{\mathbf{z}}(t) - \mathbf{T}\dot{\mathbf{x}}(t) = \mathbf{F}\mathbf{z}(t) + \mathbf{T}\mathbf{B}\mathbf{u}(t) + \mathbf{L}\mathbf{C}\mathbf{x}(t) - \mathbf{T}\mathbf{A}\mathbf{x}(t) - \mathbf{T}\mathbf{B}\mathbf{u}(t) \\ &= \mathbf{F}\mathbf{z}(t) + (\mathbf{LC} - \mathbf{TA})\mathbf{x}(t) = \mathbf{F}(\mathbf{z}(t) - \mathbf{T}\mathbf{x}(t)) = \mathbf{Fe}(t) \end{aligned}$$

If \mathbf{F} is stable, then $\mathbf{e}(t) \rightarrow \mathbf{0}$ as $t \rightarrow \infty$. Thus $\mathbf{z}(t)$ is an estimate of $\mathbf{Tx}(t)$.

► Theorem 8.M6

If \mathbf{A} and \mathbf{F} have no common eigenvalues, then the square matrix

$$\mathbf{P} := \begin{bmatrix} \mathbf{C} \\ \mathbf{T} \end{bmatrix}$$

where \mathbf{T} is the unique solution of $\mathbf{TA} - \mathbf{FT} = \mathbf{LC}$, is nonsingular only if (\mathbf{A}, \mathbf{C}) is observable and (\mathbf{F}, \mathbf{L}) is controllable.

This theorem can be proved by combining the proofs of Theorems 8.M4 and 8.6. Unlike Theorem 8.6, where the conditions are necessary and sufficient for \mathbf{P} to be nonsingular, the conditions here are only necessary. Given (\mathbf{A}, \mathbf{C}) , it is possible to construct a controllable pair (\mathbf{F}, \mathbf{L}) so that \mathbf{P} is singular. However, after selecting \mathbf{F} , if \mathbf{L} is selected randomly and if (\mathbf{F}, \mathbf{L}) is controllable, it is believed that the probability for \mathbf{P} to be nonsingular is 1.

8.8 Feedback from Estimated States—MIMO Case

This section extends the separation property discussed in Section 8.5 to the MIMO case. We use the reduced-dimensional state estimator; therefore, the development is more complex.

Consider the n -dimensional state-space equation

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{Ax}(t) + \mathbf{Bu}(t) \\ \mathbf{y}(t) &= \mathbf{Cx}(t) \end{aligned} \tag{8.80}$$

and the $(n - q)$ -dimensional state estimator in (8.78) and (8.79). First we compute the inverse of \mathbf{P} in (8.77) and then partition it as $[\mathbf{Q}_1 \ \mathbf{Q}_2]$, where \mathbf{Q}_1 is $n \times q$ and \mathbf{Q}_2 is $n \times (n - q)$, that is

$$[\mathbf{Q}_1 \ \mathbf{Q}_2] \begin{bmatrix} \mathbf{C} \\ \mathbf{T} \end{bmatrix} = \mathbf{Q}_1 \mathbf{C} + \mathbf{Q}_2 \mathbf{T} = \mathbf{I} \tag{8.81}$$

Then the $(n - q)$ -dimensional state estimator in (8.78) and (8.79) can be written as

$$\dot{\mathbf{z}}(t) = \mathbf{F}\mathbf{z}(t) + \mathbf{T}\mathbf{B}\mathbf{u}(t) + \mathbf{L}\mathbf{y}(t) \quad (8.82)$$

$$\hat{\mathbf{x}}(t) = \mathbf{Q}_1\mathbf{y}(t) + \mathbf{Q}_2\mathbf{z}(t) \quad (8.83)$$

If the original state is not available for state feedback, we apply the feedback gain matrix to $\hat{\mathbf{x}}$ to yield

$$\mathbf{u}(t) = \mathbf{r}(t) - \mathbf{K}\hat{\mathbf{x}}(t) = \mathbf{r}(t) - \mathbf{K}\mathbf{Q}_1\mathbf{y}(t) - \mathbf{K}\mathbf{Q}_2\mathbf{z}(t) \quad (8.84)$$

Substituting this into (8.80) and (8.82) yields

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}(\mathbf{r}(t) - \mathbf{K}\mathbf{Q}_1\mathbf{C}\mathbf{x}(t) - \mathbf{K}\mathbf{Q}_2\mathbf{z}(t)) \\ &= (\mathbf{A} - \mathbf{B}\mathbf{K}\mathbf{Q}_1\mathbf{C})\mathbf{x}(t) - \mathbf{B}\mathbf{K}\mathbf{Q}_2\mathbf{z}(t) + \mathbf{B}\mathbf{r}(t) \end{aligned} \quad (8.85)$$

$$\begin{aligned} \dot{\mathbf{z}}(t) &= \mathbf{F}\mathbf{z}(t) + \mathbf{T}\mathbf{B}(\mathbf{r}(t) - \mathbf{K}\mathbf{Q}_1\mathbf{C}\mathbf{x}(t) - \mathbf{K}\mathbf{Q}_2\mathbf{z}(t)) + \mathbf{L}\mathbf{C}\mathbf{x}(t) \\ &= (\mathbf{LC} - \mathbf{T}\mathbf{B}\mathbf{K}\mathbf{Q}_1\mathbf{C})\mathbf{x}(t) + (\mathbf{F} - \mathbf{T}\mathbf{B}\mathbf{K}\mathbf{Q}_2)\mathbf{z}(t) + \mathbf{T}\mathbf{B}\mathbf{r}(t) \end{aligned} \quad (8.86)$$

They can be combined as

$$\begin{bmatrix} \dot{\mathbf{x}}(t) \\ \dot{\mathbf{z}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A} - \mathbf{B}\mathbf{K}\mathbf{Q}_1\mathbf{C} & -\mathbf{B}\mathbf{K}\mathbf{Q}_2 \\ \mathbf{LC} - \mathbf{T}\mathbf{B}\mathbf{K}\mathbf{Q}_1\mathbf{C} & \mathbf{F} - \mathbf{T}\mathbf{B}\mathbf{K}\mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{z}(t) \end{bmatrix} + \begin{bmatrix} \mathbf{B} \\ \mathbf{T}\mathbf{B} \end{bmatrix} \mathbf{r}(t) \quad (8.87)$$

$$\mathbf{y}(t) = [\mathbf{C} \ \mathbf{0}] \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{z}(t) \end{bmatrix}$$

This $(2n - q)$ -dimensional state-space equation describes the feedback system in Fig. 8.8. As in the SISO case, let us carry out the following equivalence transformation

$$\begin{bmatrix} \mathbf{x}(t) \\ \mathbf{e}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{z}(t) - \mathbf{T}\mathbf{x}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ -\mathbf{T} & \mathbf{I}_{n-q} \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{z}(t) \end{bmatrix}$$

After some manipulation and using $\mathbf{T}\mathbf{A} - \mathbf{F}\mathbf{T} = \mathbf{LC}$ and (8.81), we can finally obtain the following equivalent state-space equation:

$$\begin{bmatrix} \dot{\mathbf{x}}(t) \\ \dot{\mathbf{e}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A} - \mathbf{B}\mathbf{K} & -\mathbf{B}\mathbf{K}\mathbf{Q}_2 \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{e}(t) \end{bmatrix} + \begin{bmatrix} \mathbf{B} \\ \mathbf{0} \end{bmatrix} \mathbf{r}(t) \quad (8.88)$$

$$\mathbf{y}(t) = [\mathbf{C} \ \mathbf{0}] \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{e}(t) \end{bmatrix}$$

This equation is similar to (8.55) for the SISO case. Therefore all the discussion there applies, without any modification, to the MIMO case. In other words, the design of state feedback and the design of state estimator can be carried out independently. This is the *separation property*. Furthermore, all eigenvalues of \mathbf{F} are not controllable from \mathbf{r} and the transfer matrix from \mathbf{r} to \mathbf{y} equals

$$\hat{\mathbf{G}}_f(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A} + \mathbf{B}\mathbf{K})^{-1}\mathbf{B}$$

Problems

8.1 Given

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 2 \end{bmatrix} u(t), \quad y(t) = [1 \ 1] \mathbf{x}(t)$$

Find the state feedback gain \mathbf{k} so that the state feedback system has -1 and -2 as its eigenvalues. Compute \mathbf{k} directly without using any equivalence transformation.

8.2 Repeat Problem 8.1 by using (8.13).

8.3 Repeat Problem 8.1 by solving a Lyapunov equation.

8.4 Find the state feedback gain for the state equation

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 1 & 1 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} u(t)$$

so that the resulting system has eigenvalues -2 and $-1 \pm j1$. Use the method you think is the simplest by hand to carry out the design.

8.5 Consider the unstable single-input single-output system

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \begin{bmatrix} -1 & 1 & -1 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} u(t) \\ y(t) &= [1 \ 0 \ 0] \mathbf{x}(t) \end{aligned}$$

Design a state feedback control law to shift the eigenvalues to $-1, -2, -3$ by transforming the above system to controllable form.

8.6 Consider a system with transfer function

$$\hat{g}(s) = \frac{(s-1)(s+2)}{(s+1)(s-2)(s+3)}$$

Is it possible to change the transfer function to

$$\hat{g}_f(s) = \frac{s-1}{(s+2)(s+3)}$$

by state feedback? Is the resulting system BIBO stable? asymptotically stable?

8.7 Consider a system with transfer function

$$\hat{g}(s) = \frac{(s-1)(s+2)}{(s+1)(s-2)(s+3)}$$

Is it possible to change the transfer function to

$$\hat{g}_f(s) = \frac{1}{s+3}$$

by state feedback? Is the resulting system BIBO stable? asymptotically stable?

- 8.8** Consider the continuous-time state-space equation

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 1 & 1 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} u(t)$$

$$y(t) = [2 \ 0 \ 0] \mathbf{x}(t)$$

Let $u(t) = pr(t) - \mathbf{kx}(t)$. Find the feedforward gain p and state feedback gain \mathbf{k} so that the resulting system has eigenvalues -2 and $-1 \pm j1$ and will track asymptotically any step reference input.

- 8.9** Consider the discrete-time state-space equation

$$\mathbf{x}[k+1] = \begin{bmatrix} 1 & 1 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}[k] + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} u[k]$$

$$y[k] = [2 \ 0 \ 0] \mathbf{x}[k]$$

Find the state feedback gain so that the resulting system has all eigenvalues at $z = 0$. Show that for any initial state, the zero-input response of the feedback system becomes identically zero for $k \geq 3$.

- 8.10** Consider the discrete-time state-space equation in Problem 8.9. Let $u[k] = pr[k] - \mathbf{kx}[k]$, where p is a feedforward gain. For the \mathbf{k} in Problem 8.9, find a gain p so that the output will track any step reference input. Show also that $y[k] = r[k]$ for $k \geq 3$. Thus exact tracking is achieved in a finite number of sampling periods instead of asymptotically. This is possible if all poles of the resulting system are placed at $z = 0$. This is called the *dead beat* design.

- 8.11** Consider the uncontrollable state equation

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} u(t)$$

Is it possible to find a gain \mathbf{k} so that the equation with state feedback $u = r - \mathbf{kx}$ has eigenvalues $-2, -2, -1, -1$? Is it possible to have eigenvalues $-2, -2, -2, -1$? How about $-2, -2, -2, -2$? Is the equation stabilizable?

- 8.12** Design a full-dimensional and a reduced-dimensional state estimator for the state-space equation in Problem 8.1. Select the eigenvalues of the estimators from $\{-3, -2 \pm j2\}$

- 8.13** Consider the state-space equation in Problem 8.1. Compute the transfer function from r to y of the state feedback system. Compute the transfer function from r to y if the feedback gain is applied to the estimated state of the full-dimensional estimator designed in Problem 8.12. Compute the transfer function from r to y if the feedback gain is applied to the estimated state of the reduced-dimensional state estimator also designed in Problem 8.12. Are the three overall transfer functions the same?

- 8.14** Let

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -3 & 1 & 2 & 3 \\ 2 & 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 2 \\ 0 & 2 \end{bmatrix}$$

Find two different constant matrices \mathbf{K} such that $(\mathbf{A} - \mathbf{B}\mathbf{K})$ has eigenvalues $-4 \pm 3j$ and $-5 \pm 4j$.

- 8.15** Consider the unstable multi-input multi-output system

$$\begin{aligned}\dot{x}(t) &= \begin{bmatrix} -1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}x(t) + \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}u(t) \\ y(t) &= \begin{bmatrix} 1 & 3 & -1 \\ 0 & 1 & 0 \end{bmatrix}x(t)\end{aligned}$$

Use controllable-form method to design a state feedback control law to shift the eigenvalues to $-1, -2, -3$.

- 8.16** Consider the chemical reactor system discussed in Munro (1972)

$$\begin{aligned}\dot{x}(t) &= \begin{bmatrix} 1.3800 & -0.2077 & 6.7150 & -5.6760 \\ -0.5814 & -4.2900 & 0 & 0.6750 \\ 1.0670 & 4.2730 & -6.6540 & 5.8930 \\ 0.0480 & 4.2730 & 1.3430 & -2.1040 \end{bmatrix}x(t) + \begin{bmatrix} 0 & 0 \\ 5.679 & 0 \\ 1.136 & -3.146 \\ 1.136 & 0 \end{bmatrix}u(t) \\ y(t) &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}x(t)\end{aligned}$$

Design a reduced-order observer-based feedback control law to shift the eigenvalues of system to $-5, -2, -1, -9$ by selecting the observer eigenvalues at -10 and -12 .

- 8.17** Consider the following system

$$\dot{x}(t) = \begin{bmatrix} -1 & 1 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}x(t) + \begin{bmatrix} 1 \\ 2+\sqrt{2} \\ 1 \end{bmatrix}u(t)$$

Is it possible to shift two eigenvalues of the system to $-\sqrt{2} + 1$ and the third one to $-\sqrt{2} - 1$ by using state feedback? If yes, find the feedback gain; if no, why not?

- 8.18** Consider a system with the transfer function

$$G(s) = \frac{s+1}{(s-2)(s+3)}$$

- (a) Design a full-order observer and a reduced-order observer for the above system such that all eigenvalues of the observers are placed at -3 .
 - (b) Observer is usually designed in connection with state feedback control law to stabilize an unstable system such as the one given above. However, it is desirable to avoid observer design, if one can stabilize the system with static output feedback control law of the form $u = v - Hy$. Is it possible to do this for the above system? If yes, find the range of values for H such that the closed-loop system is stable.
 - (c) If the answer to part(b) is positive, show a procedure to track asymptotically a step reference input.
- 8.19** Design a full and a reduced order observer for Problem 8.7. Choose the observer eigenvalues from the set $\{-1, -2, -3\}$.

8.20 Consider a system represented by

$$\begin{aligned}\dot{x}(t) &= A x(t) + B u(t) \\ y(t) &= C x(t) + D u(t)\end{aligned}$$

and a dynamic compensator constructed in a feedback configuration described by

$$\begin{aligned}\dot{z}(t) &= F z(t) + G y(t) \\ w(t) &= M z(t) + N y(t)\end{aligned}$$

Show that the overall system can be transformed to a static output feedback problem.

Pole Placement and Model Matching

9.1 Introduction

We first give reasons for introducing this chapter. Chapter 6 introduces state-space analysis (controllability and observability) and Chapter 8 introduces design (state feedback and state estimators). We introduced in Chapter 7 coprime fractions. Therefore, it is logical to discuss in this chapter their applications in design.

Most control systems, as discussed in Section 8.1, can be formulated as shown in Fig. 8.1. That is, given a plant with input u and output y and a reference signal r , design an overall system so that the output y will track the reference signal r as closely as possible. The plant input u is also called the actuating signal and the plant output y , the controlled signal. If the actuating signal u depends only on the reference signal r as shown in Fig. 9.1(a), it is called an open-loop control. If u depends on r and y , then it is called a closed-loop or feedback control. There are many possible feedback configurations. The simplest is the unity-feedback configuration shown in Fig. 9.1(b) in which the constant gain p and the compensator with transfer function $C(s)$ are to be designed. Clearly we have

$$\hat{u}(s) = C(s)[p\hat{r}(s) - \hat{y}(s)] \quad (9.1)$$

Because p is a constant, the reference signal r and the plant output y drive essentially the same compensator to generate an actuating signal. Thus the configuration is said to have one degree of freedom. Clearly, the open-loop configuration also has one degree of freedom.

The connection of state feedback and state estimator in Fig. 8.8 can be redrawn as shown in Fig. 9.1(c). Simple manipulation yields

$$\hat{u}(s) = \frac{1}{1 + C_1(s)}\hat{r}(s) - \frac{C_2(s)}{1 + C_1(s)}\hat{y}(s)$$

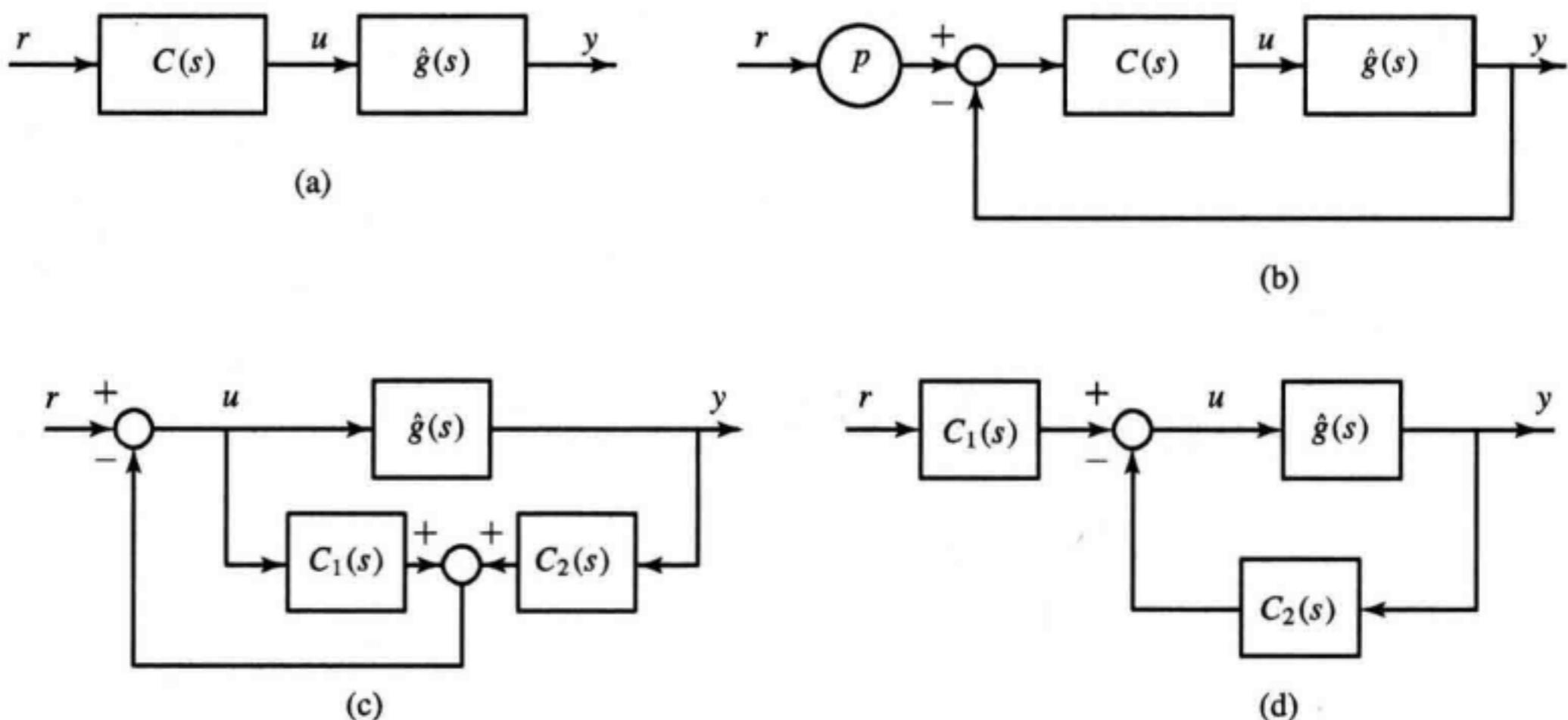


FIGURE 9.1 Control configurations.

We see that r and y drive two independent compensators to generate a u . Thus the configuration is said to have two degrees of freedom.

A more natural two-degree-of-freedom configuration can be obtained by modifying (9.1) as

$$\hat{u}(s) = C_1(s)\hat{r}(s) - C_2(s)\hat{y}(s) \quad (9.2)$$

and is plotted in Fig. 9.1(d). This is the most general control signal because each of r and y drives a compensator which we have freedom in designing. Thus no configuration has three degrees of freedom. There are many possible two-degree-of-freedom configurations, see, for example, Reference 15. We call the one in Fig. 9.1(d) the *two-parameter configuration* and call the one in Fig. 9.1(c) the *controller-estimator* or *plant-input-output-feedback* configuration. Because the two-parameter configuration seems to be more natural and more suitable for practical application, we study only the configuration in this chapter. For designs using the plant-input-output-feedback configuration, see Reference 6.

The design to be introduced in this chapter is based on coprime polynomial fractions. We need the result in Section 7.3 for SISO systems, and we need the results in Sections 7.6 through 7.8.2 for MIMO systems. The rest of Chapter 7 and the entire Chapter 8, however, are not needed here. In this chapter, we will change the design problem into solving sets of linear algebraic equations. Thus the method is called the *linear algebraic* method in Reference 7.

For convenience, we first introduce some terminology. Every transfer function $\hat{g}(s) = N(s)/D(s)$ is assumed to be a coprime fraction. Then every root of $D(s)$ is a pole of $\hat{g}(s)$ and every root of $N(s)$ is a zero of $\hat{g}(s)$. A pole is called a CT stable pole if it has a negative real part; a CT unstable pole if it has a zero or positive real part.

A polynomial is called a *CT stable polynomial* if all its roots have negative real parts. We also define the following:

- Minimum-phase zeros: zeros with negative real parts.
- Nonminimum-phase zeros: zeros with zero or positive real parts.

Although some texts call them stable and unstable zeros, they have nothing to do with stability. A transfer function with only minimum-phase zeros has the smallest phase among all transfer functions with the same magnitude responses. See Reference 7, pp. 284–285. Thus we use the aforementioned terminology.

9.2 Preliminary—Matching Coefficients

In this section, we use a simple example to discuss the basic idea to be employed in the design. Before proceeding, we discuss the effects of feedback on poles and zeros. Consider the unity feedback system in Fig. 9.1(b). Let $p = 1$, $\hat{g}(s) = N(s)/D(s)$, and $C(s) = B(s)/A(s)$. Then the overall transfer function $\hat{g}_o(s)$ from r to y is

$$\begin{aligned}\hat{g}_o(s) &= \frac{C(s)\hat{g}(s)}{1 + C(s)\hat{g}(s)} = \frac{\frac{B(s)}{A(s)} \cdot \frac{N(s)}{D(s)}}{1 + \frac{B(s)}{A(s)} \cdot \frac{N(s)}{D(s)}} \\ &= \frac{B(s)N(s)}{A(s)D(s) + B(s)N(s)}\end{aligned}\quad (9.3)$$

We see that the zeros of the plant remain as the zeros of the feedback system. In other words, feedback has no effect on the zeros. The poles of the feedback system or the roots of $A(s)D(s) + B(s)N(s)$ are different from the poles of the plant or the roots of $D(s)$. Thus the poles of the plant are shifted to new locations. In conclusion, feedback affects poles but not zeros. The situation is identical to the state feedback discussed in (8.17).

The effect on poles clearly depends on the degree of a compensator. We use an example to show that if the degree of a compensator is large enough, then all poles of the feedback system can be arbitrarily assigned. Consider the plant transfer function

$$\hat{g}(s) = \frac{s-2}{s^2-1} = \frac{s-2}{(s+1)(s-1)} =: \frac{N(s)}{D(s)}$$

We first consider a compensator of degree 0 or $C(s) = k$, where k is a real constant. Then (9.3) becomes

$$\begin{aligned}\hat{g}_o(s) &= \frac{C(s)\hat{g}(s)}{1 + C(s)\hat{g}(s)} = \frac{k \cdot \frac{s-2}{s^2-1}}{1 + k \cdot \frac{s-2}{s^2-1}} = \frac{k(s-2)}{s^2 - 1 + k(s-2)} \\ &= \frac{k(s-2)}{s^2 + ks - (2k+1)}\end{aligned}$$

We see that for any real k , the zero of $\hat{g}_o(s)$ remains as 2 but the poles will vary. For example, if $k = 4$, the poles of $\hat{g}(s)$ are shifted from ± 1 to the roots of $s^2 + 4s - 9$ or

-1.6 and 5.6. The feedback system is not stable because it has a CT unstable pole at 5.6. In fact, for any real-valued k , $\hat{g}_o(s)$ remains unstable. Thus we cannot design a BIBO stable system using a compensator of degree 0.

Next we consider a proper compensator of degree 1. Instead of using $C(s) = (bs + c)/(s + a)$, we select, for convenience of subsequent development,

$$C(s) = \frac{B_1 s + B_0}{A_1 s + A_0} =: \frac{B(s)}{A(s)}$$

where B_i and A_i are real constants and $A_1 \neq 0$. If $A_1 = 0$, the compensator is improper. With this compensator, we have, using (9.3),

$$\begin{aligned}\hat{g}_o(s) &= \frac{B(s)N(s)}{A(s)D(s) + B(s)N(s)} = \frac{(B_1 s + B_0)(s - 2)}{(A_1 s + A_0)(s^2 - 1) + (B_1 s + B_0)(s - 2)} \\ &= \frac{(B_1 s + B_0)(s - 2)}{A_1 s^3 + (A_0 + B_1)s^2 + (B_0 - A_1 - 2B_1)s + (-A_0 - 2B_0)}\end{aligned}$$

Again the zero of $\hat{g}(s)$ at 2 is not affected by the feedback and remains as a zero of $\hat{g}_o(s)$. Because of the feedback, the four parameters of the compensator appear in each coefficient of the denominator of $\hat{g}_o(s)$ as

$$F(s) := A_1 s^3 + (A_0 + B_1)s^2 + (B_0 - A_1 - 2B_1)s + (-A_0 - 2B_0)$$

Thus it seems possible to place the poles in any position. For example, if the three poles are selected as $-2, -1 \pm j1$, then $\hat{g}_o(s)$ should have

$$F(s) = (s + 2)(s + 1 - j1)(s + 1 + j1) = (s + 2)(s^2 + 2s + 2) = s^3 + 4s^2 + 6s + 4$$

as its denominator. Matching the coefficients of like powers yields

$$\begin{aligned}A_1 &= 1 \\ A_0 + B_1 &= 4 \\ B_0 - A_1 - 2B_1 &= 6 \\ -A_0 - 2B_0 &= 4\end{aligned}$$

Substituting the first equation into the third and summing up the second and fourth equations yield, respectively,

$$\begin{aligned}B_0 - 2B_1 &= 7 \\ -2B_0 + B_1 &= 8\end{aligned}$$

Adding the first equation multiplied by 2 to the second equation yields $-3B_1 = 22$ or $B_1 = -22/3$. Consequently we have $B_0 = 7 + 2B_1 = -23/3$ and $A_0 = 4 - B_1 = 34/3$. Thus we have

$$A(s) = A_1 s + A_0 = s + (34/3) \quad \text{and} \quad B(s) = B_1 s + B_0 = (-22/3)s - (23/3)$$

and the compensator is

$$C(s) = \frac{B_1 s + B_0}{A_1 s + A_0} = \frac{-22s/3 + (-23/3)}{1 \cdot s + 34/3} = \frac{-22s - 23}{3s + 34}$$

For the given $\hat{g}(s) = N(s)/D(s)$ and the preceding compensator $C(s) = B(s)/A(s)$, (9.3) becomes

$$\begin{aligned}\hat{g}_o(s) &= \frac{B(s)N(s)}{A(s)D(s) + B(s)N(s)} = \frac{B(s)N(s)}{F(s)} \\ &= \frac{((-22/3)s - (23/3))(s - 2)}{s^3 + 4s^2 + 6s + 4}\end{aligned}$$

It has poles at $-2, -1 \pm j1$. This is called the *pole-placement* design.

We see that design involves only matching coefficients of the polynomial equation $A(s)D(s) + B(s)N(s) = F(s)$, thus the procedure is very simple. In the remainder of this chapter, we will formalize the design procedure and then extend it to design model matching.

9.2.1 Compensator Equations—Classical Method

The pole-place problem involves solving the polynomial equation

$$A(s)D(s) + B(s)N(s) = F(s) \quad (9.4)$$

where $D(s)$, $N(s)$, and $F(s)$ are given polynomials and $A(s)$ and $B(s)$ are unknown polynomials to be solved. Mathematically speaking, this problem is equivalent to the problem that given integers D , N , and F , find integer solutions A and B in $AD + BN = F$. This is a very old mathematical problem and has been associated with mathematicians Diophantus, Bezout, Aryabhatta, and others.¹ To avoid controversy, we follow Reference 4 and call it a *compensator equation*.

We first discuss the existence condition and general solutions of the equation. What will be discussed, however, is not needed in subsequent sections, and the reader may glance through this subsection.

► Theorem 9.1

Given polynomials $D(s)$ and $N(s)$, polynomial solutions $A(s)$ and $B(s)$ exist in (9.4) for any polynomial $F(s)$ if and only if $D(s)$ and $N(s)$ are coprime.

Suppose $D(s)$ and $N(s)$ are not coprime and contain the same root a or the same factor $s - a$. Then the factor $s - a$ will appear in $F(s)$. Thus if $F(s)$ does not contain the factor, no solutions exist in (9.4). This shows the necessity of the theorem.

If $D(s)$ and $N(s)$ are coprime, there exist polynomials $\bar{A}(s)$ and $\bar{B}(s)$ such that

$$\bar{A}(s)D(s) + \bar{B}(s)N(s) = 1 \quad (9.5)$$

¹See Reference 23, the last page of Preface.

Its matrix version is called the *Bezout identity* in Reference 16. The polynomials $\bar{A}(s)$ and $\bar{B}(s)$ can be obtained by the Euclidean algorithm and will not be discussed here. See Reference 6, pp. 578–580. For example, if $D(s) = s^2 - 1$ and $N(s) = s - 2$, then $\bar{A}(s) = 1/3$ and $\bar{B}(s) = -(s + 2)/3$ meet (9.5). For any polynomial $F(s)$, (9.5) implies

$$F(s)\bar{A}(s)D(s) + F(s)\bar{B}(s)N(s) = F(s) \quad (9.6)$$

Thus $A(s) = F(s)\bar{A}(s)$ and $B(s) = F(s)\bar{B}(s)$ are solutions. This shows the sufficiency of the theorem.

Next we discuss general solutions. For any $D(s)$ and $N(s)$, there exist two polynomials $\hat{A}(s)$ and $\hat{B}(s)$ such that

$$\hat{A}(s)D(s) + \hat{B}(s)N(s) = 0 \quad (9.7)$$

Obviously $\hat{A}(s) = -N(s)$ and $\hat{B}(s) = D(s)$ are such solutions. Then for any polynomial $Q(s)$, the equations

$$A(s) = \bar{A}(s)F(s) + Q(s)\hat{A}(s), \quad B(s) = \bar{B}(s)F(s) + Q(s)\hat{B}(s) \quad (9.8)$$

are general solutions of (9.4). This can be readily verified by substituting (9.8) into (9.4) and using (9.5) and (9.7).

Example 9.2.1 Consider $D(s) = s^2 - 1$ and $N(s) = s - 2$. They are coprime and have $\bar{A}(s) = 1/3$ and $\bar{B}(s) = -(s + 2)/3$ to meet (9.5). If $F(s) = s^3 + 4s^2 + 6s + 4$, then

$$\begin{aligned} A(s) &= \frac{1}{3}(s^3 + 4s^2 + 6s + 4) + Q(s)(-s + 2) \\ B(s) &= -\frac{1}{3}(s + 2)(s^3 + 4s^2 + 6s + 4) + Q(s)(s^2 - 1) \end{aligned} \quad (9.9)$$

for any polynomial $Q(s)$, are solutions of (9.4).

Although the classical method can yield general solutions, the solutions are not necessarily convenient to use in design. For example, we may be interested in solving $A(s)$ and $B(s)$ with least degrees to meet (9.4). In this case, after some manipulation, we find that if $Q(s) = (s^2 + 6s + 15)/3$, then (9.9) reduces to

$$A(s) = s + 34/3, \quad B(s) = (-22s - 23)/3 \quad (9.10)$$

They are the ones obtained in the preceding section by matching coefficients. In this chapter, instead of solving the compensator equation directly as discussed above, we will change it into solving a set of linear algebraic equations as in Section 7.3. By so doing, we can bypass some polynomial theorems.

9.3 Unity-Feedback Configuration—Pole Placement

Consider the unity-feedback system shown in Fig. 9.1(b). The plant transfer function $\hat{g}(s)$ is assumed to be strictly proper and of degree n . If $\hat{g}(s)$ is biproper, the discussion must be modified slightly. See Reference 6. The problem is to design a proper

compensator $C(s)$ of least possible degree m so that the resulting overall system has any set of $n + m$ desired poles. Because all transfer functions are required to have real coefficients, complex conjugate poles must be assigned in pairs. This will be a standing assumption throughout this chapter.

Let $\hat{g}(s) = N(s)/D(s)$ and $C(s) = B(s)/A(s)$. Then the overall transfer function from r to y in Fig. 9.1(b) is, as derived in (9.3) with $p = 1$,

$$\hat{g}_o(s) = \frac{pB(s)N(s)}{A(s)D(s) + B(s)N(s)} \quad (9.11)$$

In pole assignment, we are interested in assigning all poles of $\hat{g}_o(s)$ or, equivalently, all roots of $A(s)D(s) + B(s)N(s)$. In this design, nothing is said about the zeros of $\hat{g}_o(s)$. As we can see from (9.11), the design not only has no effect on the plant zeros (roots of $N(s)$), but also introduces new zeros (roots of $B(s)$) into the overall transfer function. On the other hand, the poles of the plant and compensator are shifted from $D(s)$ and $A(s)$ to the roots of $A(s)D(s) + B(s)N(s)$.

Given a set of desired poles, we can readily form a polynomial $F(s)$ that has the desired poles as its roots. Then the pole-placement problem becomes solving the polynomial equation

$$A(s)D(s) + B(s)N(s) = F(s) \quad (9.12)$$

Instead of solving (9.12) directly, we will transform it into solving a set of linear algebraic equations. Let $\deg N(s) < \deg D(s) = n$ and $\deg B(s) \leq \deg A(s) = m$. Then $F(s)$ in (9.12) has degree at most $n+m$. Let us write

$$\begin{aligned} D(s) &= D_0 + D_1s + D_2s^2 + \cdots + D_ns^n, & D_n &\neq 0 \\ N(s) &= N_0 + N_1s + N_2s^2 + \cdots + N_ns^n \\ A(s) &= A_0 + A_1s + A_2s^2 + \cdots + A_ms^m \\ B(s) &= B_0 + B_1s + B_2s^2 + \cdots + B_ms^m \\ F(s) &= F_0 + F_1s + F_2s^2 + \cdots + F_{n+m}s^{n+m} \end{aligned}$$

where all coefficients are real constants, not necessarily nonzero. Note that the subscript of a coefficient equals the power of s . Substituting these into (9.12) and matching the coefficients of like power of s yield

$$\begin{aligned} A_0D_0 + B_0N_0 &= F_0 \\ A_0D_1 + B_0N_1 + A_1D_0 + B_1N_0 &= F_1 \\ &\vdots \\ A_mD_n + B_mN_n &= F_{n+m} \end{aligned}$$

There are a total of $(n+m+1)$ equations. They can be arranged in matrix form as

$$[A_0 \ B_0 \ A_1 \ B_1 \ \cdots \ A_m \ B_m] \mathbf{S}_m = [F_0 \ F_1 \ F_2 \ \cdots \ F_{n+m}] \quad (9.13)$$

with

$$\mathbf{S}_m := \begin{bmatrix} D_0 & D_1 & \cdots & D_n & 0 & \cdots & 0 \\ N_0 & N_1 & \cdots & N_n & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & D_0 & \cdots & D_{n-1} & D_n & \cdots & 0 \\ 0 & N_0 & \cdots & N_{n-1} & N_n & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & D_0 & \cdots & D_n \\ 0 & 0 & \cdots & 0 & N_0 & \cdots & N_n \end{bmatrix} \quad (9.14)$$

If we take the transpose of (9.13), then it becomes the standard form studied in Theorems 3.1 and 3.2. We use the form in (9.13), because it can be directly extended to the matrix case. The matrix \mathbf{S}_m has $2(m+1)$ rows and $(n+m+1)$ columns and is formed from the coefficients of $D(s)$ and $N(s)$. The first two rows are simply the coefficients of $D(s)$ and $N(s)$ arranged in ascending power of s . The next two rows are the first two rows shifted to the right by one position. We repeat the process until we have $(m+1)$ sets of coefficients. The left-hand-side row vector of (9.13) consists of the coefficients of the compensator $C(s)$ to be solved. If $C(s)$ has degree m , then the row vector has $2(m+1)$ entries. The right-hand-side row vector of (9.13) consists of the coefficients of $F(s)$. Now solving the compensator equation in (9.12) becomes solving the linear algebraic equation in (9.13).

Applying Corollary 3.2, we conclude that (9.13) has a solution for any $F(s)$ if and only if \mathbf{S}_m has full column rank. A necessary condition for \mathbf{S}_m to have full column rank is that \mathbf{S}_m is square or has more rows than columns, that is,

$$2(m+1) \geq n+m+1 \quad \text{or} \quad m \geq n-1$$

If $m < n-1$, then \mathbf{S}_m does not have full column rank and solutions may exist for some $F(s)$, but not for every $F(s)$. Thus if the degree of the compensator is less than $n-1$, it is not possible to achieve arbitrary pole placement.

If $m = n-1$, \mathbf{S}_{n-1} becomes a square matrix of order $2n$ and is called the Sylvester resultant. As discussed in Section 7.3, \mathbf{S}_{n-1} is nonsingular if and only if $D(s)$ and $N(s)$ are coprime. Thus if $D(s)$ and $N(s)$ are coprime, then \mathbf{S}_{n-1} has rank $2n$ (full column rank). Now if m increases by 1, the number of columns increases by 1 but the number of rows increases by 2. Because $D_n \neq 0$, the new D row is linearly independent of its preceding rows. Thus the $2(n+1) \times (2n+1)$ matrix \mathbf{S}_n has rank $(2n+1)$ (full column rank). Repeating the argument, we conclude that if $D(s)$ and $N(s)$ are coprime and if $m \geq n-1$, then the matrix \mathbf{S}_m in (9.14) has full column rank.

Theorem 9.2

Consider the unity-feedback system shown in Fig. 9.2(b). The plant is described by a strictly proper transfer function $\hat{g}(s) = N(s)/D(s)$ with $N(s)$ and $D(s)$ coprime and $\deg N(s) < \deg D(s) = n$. Let $m \geq n-1$. Then for any polynomial $F(s)$ of degree $(n+m)$, there exists a proper

compensator $C(s) = B(s)/A(s)$ of degree m such that the overall transfer function equals

$$\hat{g}_o(s) = \frac{pN(s)B(s)}{A(s)D(s) + B(s)N(s)} = \frac{pN(s)B(s)}{F(s)}$$

Furthermore, the compensator can be obtained by solving the linear algebraic equation in (9.13).

As discussed earlier, the matrix \mathbf{S}_m has full column rank for $m \geq n - 1$; therefore, for any $(n+m)$ desired poles or, equivalently, for any $F(s)$ of degree $(n+m)$, solutions exist in (9.13). Next we show that $B(s)/A(s)$ is proper or $A_m \neq 0$. If $N(s)/D(s)$ is strictly proper, then $N_n = 0$ and the last equation of (9.13) reduces to

$$A_m D_n + B_m N_n = D_n A_m = F_{n+m}$$

Because $F(s)$ has degree $n+m$, we have $F_{n+m} \neq 0$ and, consequently, $A_m \neq 0$. This establishes the theorem. If $m = n - 1$, the compensator is unique; if $m > n - 1$, compensators are not unique and free parameters can be used to achieve other design objectives, as we will discuss later.

9.3.1 Regulation and Tracking

Pole-placement can be used to achieve the regulation and tracking discussed in Section 8.3. In the regulator problem, we have $r = 0$ and the problem is to design a compensator $C(s)$ so that the response excited by any nonzero initial state will vanish at a desired rate. For this problem, if all poles of $\hat{g}_o(s)$ are selected to have negative real parts, then for any gain p , in particular $p = 1$ (no feedforward gain is needed), the overall system will achieve regulation.

We discuss next the tracking problem. Let the reference signal be a step function with amplitude a . Then $\hat{r}(s) = a/s$ and the output $\hat{y}(s)$ equals

$$\hat{y}(s) = \hat{g}_o(s)\hat{r}(s) = \hat{g}_o(s)\frac{a}{s}$$

If $\hat{g}_o(s)$ is BIBO stable, the output will approach the constant $\hat{g}_o(0)a$ (Theorem 5.2). This can also be obtained by employing the final-value theorem of the Laplace transform as

$$\lim_{t \rightarrow \infty} y(t) = \lim_{s \rightarrow 0} s\hat{y}(s) = \hat{g}_o(0)a$$

Thus in order to track asymptotically any step reference input, $\hat{g}_o(s)$ must be BIBO stable and $\hat{g}_o(0) = 1$. The transfer function from r to y in Fig. 9.1(b) is $\hat{g}_o(s) = pN(s)B(s)/F(s)$. Thus we have

$$\hat{g}_o(0) = p \frac{N(0)B(0)}{F(0)} = p \frac{B_0 N_0}{F_0}$$

which implies

$$p = \frac{F_0}{B_0 N_0} \quad (9.15)$$

Thus in order to track any step reference input, we require $B_0 \neq 0$ and $N_0 \neq 0$. The constant B_0 is a coefficient of the compensator and can be designed to be nonzero. The coefficient N_0 is the constant term of the plant numerator. Thus if the plant transfer function has a zero at $s = 0$, then $N_0 = 0$ and the plant cannot be designed to track any step reference input. This is consistent with the discussion in Section 8.3.

If the reference signal is a ramp function or $r(t) = at$, for all $t \geq 0$ and some constant a , then, using the similar argument, we can show that the overall transfer function $\hat{g}_o(s)$ must be BIBO stable and has the properties $\hat{g}_o(0) = 1$ and $\hat{g}'_o(0) = 0$ (Problems 9.13 and 9.14). This is summarized in the following.

- Regulation $\Leftrightarrow \hat{g}_o(s)$ BIBO stable.
- Tracking any step reference input $\Leftrightarrow \hat{g}_o(s)$ BIBO stable and $\hat{g}_o(0) = 1$.
- Tracking any ramp reference input $\Leftrightarrow \hat{g}_o(s)$ BIBO stable, $\hat{g}_o(0) = 1$, and $\hat{g}'_o(0) = 0$.

Example 9.3.1 Given a plant with transfer function $\hat{g}(s) = (s - 2)/(s^2 - 1)$. Find a proper compensator $C(s)$ and a gain p in the unity-feedback configuration in Fig. 9.1(b) so that the output y will track asymptotically any step reference input.

The plant transfer function has degree $n = 2$. Thus if we choose $m = 1$, then all three poles of the overall system can be arbitrarily assigned. Let the three poles be selected arbitrarily as $-2, -1 \pm j1$. Then we have

$$F(s) = (s + 2)(s + 1 + j1)(s + 1 - j1) = (s + 2)(s^2 + 2s + 2) = s^3 + 4s^2 + 6s + 4$$

We use the coefficients of $D(s) = -1 + 0 \cdot s + 1 \cdot s^2$ and $N(s) = -2 + 1 \cdot s + 0 \cdot s^2$ to form (9.13) as

$$[A_0 \ B_0 \ A_1 \ B_1] \begin{bmatrix} -1 & 0 & 1 & 0 \\ -2 & 1 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & -1 & 0 & 1 \\ 0 & -2 & 1 & 0 \end{bmatrix} = [4 \ 6 \ 4 \ 1]$$

It actually consists of the following four equations:

$$A_0 \cdot (-1) + B_0 \cdot (-2) + A_1 \cdot 0 + B_1 \cdot 0 = 4 \quad \text{or} \quad -A_0 - 2B_0 = 4$$

$B_0 - A_1 - 2B_1 = 6$, $A_0 + B_1 = 4$, and $A_1 = 1$. These are the four equations developed in Section 9.2; and their solutions were obtained, by eliminating variables, as,

$$A_1 = 1 \quad A_0 = 34/3, \quad B_1 = -22/3, \quad B_0 = -23/3$$

It is important to point out that the solution can also be obtained using MATLAB as

```
S=[-1 0 1 0;-2 1 0 0;0 -1 0 1;0 -2 1 0];F=[4 6 4 1];
F/S
```

which yield

$$[A_0 \ B_0 \ A_1 \ B_1] = [11.333 \ -7.6667 \ 1.0000 \ -7.3333]$$

It is basically the same as the one obtained by hand. Thus we have

$$A(s) = s + 34/3, \quad B(s) = (-22/3)s - 23/3$$

and the compensator

$$C(s) = \frac{B(s)}{A(s)} = \frac{-(22/3)s - (22/3)}{s + 34/3} \quad (9.16)$$

will yield $\hat{g}_o(s)$ as

$$\hat{g}_o(s) = \frac{p(-(22/3)s - 23/3)(s - 2)}{s^3 + 4s^2 + 6s + 4}$$

with its three poles at -2 and $-1 \pm j1$. If the system is designed to achieve regulation, we set $p = 1$ (no feedforward gain is needed) and the design is completed.

To track any step reference input, we require $\hat{g}_o(0) = 1$ or

$$\frac{p(-22/3)(-2)}{4} = 1$$

which implies

$$p = \frac{4}{(-23/3)(-2)} = \frac{6}{23} \quad (9.17)$$

Thus the overall transfer function from r to y that follows

$$\hat{g}_o(s) = \frac{(6/23)(-(22/3)s - (23/3))(s - 2)}{(s^3 + 4s^2 + 6s + 4)} = \frac{-2(22s + 23)(s - 2)}{23(s^3 + 4s^2 + 6s + 4)} \quad (9.18)$$

will track any step reference input.

9.3.2 Robust Tracking and Disturbance Rejection²

Consider the design problem in Example 9.3.1. Suppose after the design is completed, the plant transfer function $\hat{g}(s)$ changes, due to load variations, to $\hat{g}(s) = (s - 2.1)/(s^2 - 0.95)$. Then the overall transfer function becomes

$$\begin{aligned} \hat{\hat{g}}_o(s) &= \frac{pC(s)\hat{g}(s)}{1 + C(s)\hat{g}(s)} = \frac{6}{23} \frac{\frac{-22s-23}{3s+34} \frac{s-2.1}{s-0.95}}{1 + \frac{-22s-23}{3s+34} \frac{s-2.1}{s-0.95}} \\ &= \frac{-6(22s + 23)(s - 2.1)}{23(s^3 + 12s^2 + 20.35s + 16)} \end{aligned} \quad (9.19)$$

This $\hat{\hat{g}}_o(s)$ is still BIBO stable, but $\hat{\hat{g}}_o(0) = (6 \cdot 23 \cdot 2.1) / (23 \cdot 16) = 0.7875 \neq 1$. If the reference input is a unit step function, the output will approach 0.7875 as $t \rightarrow \infty$. There

²This and next subsections may be skipped without loss of continuity. It is recommended to do so in the first reading.

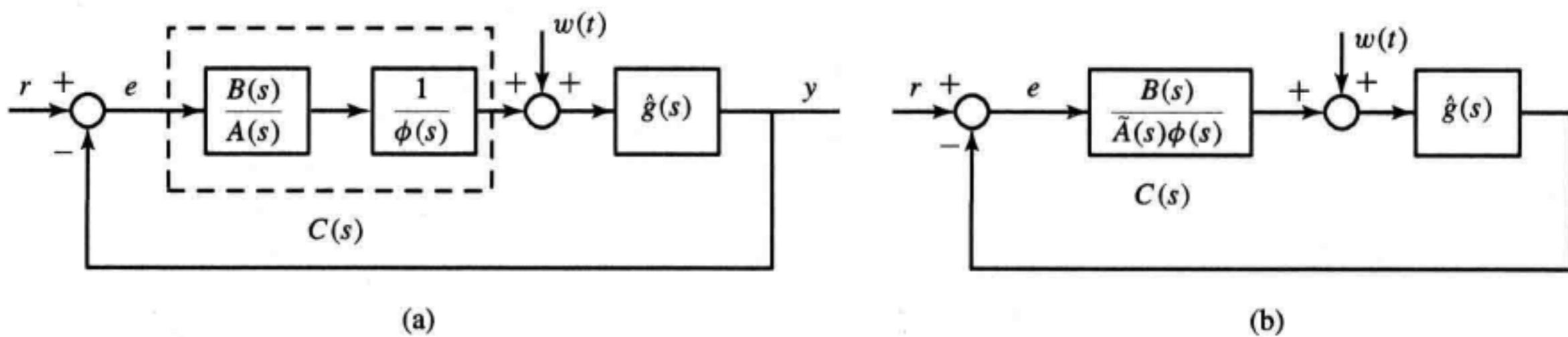


FIGURE 9.2 Robust tracking and disturbance rejection.

is a tracking error of over 20%. Thus the overall system will no longer track any step reference input after plant parameter variations, and the design is said to be not robust.

In this subsection, we discuss a design that can achieve robust tracking and disturbance rejection. Consider the system shown in Fig. 9.2, in which a disturbance enters at the plant input as shown. The problem is to design an overall system so that the plant output y will track asymptotically a class of reference signal r even with the presence of the disturbance and with plant parameter variations. This is called *robust tracking and disturbance rejection*.

Before proceeding, we discuss the nature of the reference signal $r(t)$ and the disturbance $w(t)$. If both $r(t)$ and $w(t)$ approach zero as $t \rightarrow \infty$, then the design is automatically achieved if the overall system in Fig. 9.2 is designed to be BIBO stable. To exclude this trivial case, we assume that $r(t)$ and $w(t)$ do not approach zero as $t \rightarrow \infty$. If we have no knowledge whatsoever about the nature of $r(t)$ and $w(t)$, it is not possible to achieve asymptotic tracking and disturbance rejection. Therefore we need some information of $r(t)$ and $w(t)$ before carrying out the design. We assume that the Laplace transforms of $r(t)$ and $w(t)$ are given by

$$\hat{r}(s) = \mathcal{L}[r(t)] = \frac{N_r(s)}{D_r(s)}, \quad \hat{w}(s) = \mathcal{L}[w(t)] = \frac{N_w(s)}{D_w(s)} \quad (9.20)$$

where $D_r(s)$ and $D_w(s)$ are known polynomials; however, $N_r(s)$ and $N_w(s)$ are unknown to us. For example, if $r(t)$ is a step function with unknown amplitude a , then $\hat{r}(s) = a/s$. Suppose the disturbance is $w(t) = b + c \sin(\omega_0 t + d)$; it consists of a constant biasing with unknown amplitude b and a sinusoid with known frequency ω_0 but unknown amplitude c and phase d . Then we have $\hat{w}(s) = N_w(s)/s(s^2 + \omega_0^2)$. Let $\phi(s)$ be the least common denominator of the unstable poles of $\hat{r}(s)$ and $\hat{w}(s)$. The stable poles are excluded because they have no effect on y as $t \rightarrow \infty$. Thus all roots of $\phi(s)$ have zero or positive real parts. For the examples just discussed, we have $\phi(s) = s(s^2 + \omega_0^2)$.

Theorem 9.3

Consider the unity-feedback system shown in Fig. 9.2(a) with a strictly proper plant transfer function $\hat{g}(s) = N(s)/D(s)$. It is assumed that $D(s)$ and $N(s)$ are coprime. The reference signal $r(t)$ and disturbance $w(t)$ are modeled as $\hat{r}(s) = N_r(s)/D_r(s)$ and $\hat{w}(s) = N_w(s)/D_w(s)$. Let $\phi(s)$ be the least common denominator of the unstable poles of $\hat{r}(s)$ and $\hat{w}(s)$. If no root of $\phi(s)$ is a zero of $\hat{g}(s)$, then there exists a proper compensator such that the overall system will track $r(t)$ and reject $w(t)$, both asymptotically and robustly.



Proof: If no root of $\phi(s)$ is a zero of $\hat{g}(s) = N(s)/D(s)$, then $D(s)\phi(s)$ and $N(s)$ are coprime. Thus there exists a proper compensator $B(s)/A(s)$ such that the polynomial $F(s)$ in

$$A(s)D(s)\phi(s) + B(s)N(s) = F(s)$$

has any desired roots, in particular, has all roots lying inside the sector shown in Fig. 8.3. We claim that the compensator

$$C(s) = \frac{B(s)}{A(s)\phi(s)}$$

as shown in Fig. 9.2(a) will achieve the design. Let us compute the transfer function from w to y :

$$\begin{aligned}\hat{g}_{yw}(s) &= \frac{N(s)/D(s)}{1 + (B(s)/A(s)\phi(s))(N(s)/D(s))} \\ &= \frac{N(s)A(s)\phi(s)}{A(s)D(s)\phi(s) + B(s)N(s)} = \frac{N(s)A(s)\phi(s)}{F(s)}\end{aligned}$$

Thus the output excited by $w(t)$ equals

$$\hat{y}_w(s) = \hat{g}_{yw}\hat{w}(s) = \frac{N(s)A(s)\phi(s)}{F(s)} \frac{N_w(s)}{D_w(s)} \quad (9.21)$$

Because all unstable roots of $D_w(s)$ are canceled by $\phi(s)$, all poles of $\hat{y}_w(s)$ have negative real parts. Thus we have $y_w(t) \rightarrow 0$ as $t \rightarrow \infty$ and the response excited by $w(t)$ is asymptotically suppressed at the output.

Next we compute the output $\hat{y}_r(s)$ excited by $\hat{r}(s)$:

$$\hat{y}_r(s) = \hat{g}_{yr}(s)\hat{r}(s) = \frac{B(s)N(s)}{A(s)D(s)\phi(s) + B(s)N(s)}\hat{r}(s) \quad (9.22)$$

Thus we have

$$\begin{aligned}\hat{e}(s) &:= \hat{r}(s) - \hat{y}_r(s) = (1 - \hat{g}_{yr}(s))\hat{r}(s) \\ &= \frac{A(s)D(s)\phi(s)}{F(s)} \frac{N_r(s)}{D_r(s)}\end{aligned} \quad (9.23)$$

Again all unstable roots of $D_r(s)$ are canceled by $\phi(s)$ in (9.23). Thus we conclude $r(t) - y_r(t) \rightarrow 0$ as $t \rightarrow \infty$. Because of linearity, we have $y(t) = y_w(t) + y_r(t)$ and $r(t) - y(t) \rightarrow 0$ as $t \rightarrow \infty$. This shows asymptotic tracking and disturbance rejection. From (9.21) and (9.23), we see that even if the parameters of $D(s)$, $N(s)$, $A(s)$ and $B(s)$ change, so long as the overall system remains BIBO stable and the unstable roots of $D_r(s)$ and $D_w(s)$ are canceled by $\phi(s)$, the system still achieves tracking and rejection. Thus the design is robust. Q.E.D.

This robust design consists of two steps: First find a model $1/\phi(s)$ of the reference signal and disturbance and then carry out pole-placement design. Inserting the model

inside the loop is referred to as the *internal model principle*. If the model $1/\phi(s)$ is not located in the forward path from w to y and from r to e , then $\phi(s)$ will appear in the numerators of $\hat{g}_{yw}(s)$ and $\hat{g}_{er}(s)$ (see Problem 9.7) and cancel the unstable poles of $\hat{w}(s)$ and $\hat{r}(s)$, as shown in (9.21) and (9.22). Thus the design is achieved by unstable pole-zero cancellations of $\phi(s)$. It is important to mention that there are no unstable pole-zero cancellations in the pole-placement design and that the resulting unity-feedback system is totally stable, which will be defined in Section 9.3. Thus the internal model principle can be used in practical design.

In classical control system design, if a plant transfer function or a compensator transfer function is of type 1 (has one pole at $s = 0$), and if the unity feedback system is designed to be BIBO stable, then the overall system will track asymptotically and robustly any step reference input. This is a special case of the internal model principle because the plant or compensator contains the model of any step reference input.

Example 9.3.2 Consider the plant in Example 9.3.1 or $\hat{g}(s) = (s - 2)/(s^2 - 1)$. Design a unity feedback system with a set of desired poles to track robustly any step reference input.

First we introduce the internal model $\phi(s) = 1/s$. Then $B(s)/A(s)$ in Fig. 9.2(a) can be solved from

$$A(s)D(s)\phi(s) + B(s)N(s) = F(s)$$

Because $\tilde{D}(s) := D(s)\phi(s)$ has degree 3, we may select $A(s)$ and $B(s)$ to have degree 2. Then $F(s)$ has degree 5. If we select five desired poles as $-2, -2 \pm j1$, and $-1 \pm j2$, then we have

$$\begin{aligned} F(s) &= (s + 2)(s^2 + 4s + 5)(s^2 + 2s + 5) \\ &= s^5 + 8s^4 + 30s^3 + 66s^2 + 85s + 50 \end{aligned}$$

Using the coefficients of $\tilde{D}(s) = (s^2 - 1)s = 0 - s + 0 \cdot s^2 + s^3$ and $N(s) = -2 + s + 0 \cdot s^2 + 0 \cdot s^3$, we form

$$[A_0 \ B_0 \ A_1 \ B_1 \ A_2 \ B_2] \begin{bmatrix} 0 & -1 & 0 & 1 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & -2 & 1 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & -2 & 1 & 0 & 0 \end{bmatrix} = [50 \ 85 \ 66 \ 30 \ 8 \ 1]$$

Its solution is $[127.3 \ -25 \ 8 \ -118.7 \ 1 \ -96.3]$. Thus we have

$$\frac{B(s)}{A(s)} = \frac{-96.3s^2 - 118.7s - 25}{s^2 + 8s + 127.3}$$

and the compensator is

$$C(s) = \frac{B(s)}{A(s)\phi(s)} = \frac{-96.3s^2 - 118.7s - 25}{(s^2 + 8s + 127.3)s}$$

Using this compensator of degree 3, the unity feedback system in Fig. 9.2(a) will track robustly any step reference input and has the set of desired poles.

9.3.3 Embedding Internal Models³

The design in the preceding subsection was achieved by first introducing an internal model $1/\phi(s)$ and then design a proper $B(s)/A(s)$. Thus the compensator $B(s)/A(s)\phi(s)$ is always strictly proper. In this subsection, we discuss a method of designing a biproper compensator whose denominator will include the internal model as a factor as shown in Fig. 9.2(b). By so doing, the degree of compensators can be reduced.

Consider

$$A(s)D(s) + B(s)N(s) = F(s)$$

If $\deg D(s) = n$ and if $\deg A(s) = n - 1$, then the solution $A(s)$ and $B(s)$ is unique. If we increase the degree of $A(s)$ by one, then solutions are not unique, and there is one free parameter we can select. Using the free parameter, we may be able to include an internal model in the compensator as the next example illustrates.

Example 9.3.3 Consider again the design problem in Example 9.3.2. The degree of $D(s)$ is 2. If $A(s)$ has degree 1, then the solution is unique. Let us select $A(s)$ to have degree 2. Then $F(s)$ must have degree 4 and can be selected as

$$F(s) = (s^2 + 4s + 5)(s^2 + 2s + 5) = s^4 + 6s^3 + 18s^2 + 30s + 25$$

We form

$$[A_0 \ B_0 \ A_1 \ B_1 \ A_2 \ B_2] \begin{bmatrix} -1 & 0 & 1 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & -1 & 0 & 1 & 0 \\ 0 & -2 & 1 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & -2 & 1 & 0 \end{bmatrix} = [25 \ 30 \ 18 \ 6 \ 1]$$

Using Corollary 3.2, we can express its general solution as

$$\begin{aligned} [A_0 \ B_0 \ A_1 \ B_1 \ A_2 \ B_2] \\ = [1 \ -13 \ 34.3 \ -38.7 \ 1 \ -28.3] + \alpha[2 \ -1 \ -1 \ 0 \ 0 \ 1] \end{aligned}$$

³This subsection may be skipped without loss of continuity.

with one free parameter α . In order for the proper compensator

$$C(s) = \frac{B_0 + B_1 s + B_2 s^2}{A_0 + A_1 s + A_2 s^2}$$

to have $1/s$ as a factor, we require $A_0 = 0$. This can be achieved by selecting $\alpha = -0.5$. Then the solution becomes

$$[A_0 \ B_0 \ A_1 \ B_1 \ A_2 \ B_2] = [0 \ -12.5 \ 34.8 \ -38.7 \ 1 \ -28.8]$$

and the compensator is

$$C(s) = \frac{B(s)}{A(s)} = \frac{-28.8s^2 - 38.7s - 12.5}{s^2 + 34.8s}$$

This biproper compensator can achieve robust tracking. This compensator has degree 2, one less than the one obtained in Example 9.3.2. Thus this is a better design.

We give one more example and discuss a different method of embedding $\phi(s)$ in the compensator.

Example 9.3.4 Consider the unity feedback system in Fig. 9.2(b) with $\hat{g}(s) = 1/s$. Design a proper compensator $C(s) = B(s)/A(s)$ so that the system will track asymptotically any step reference input and reject disturbance $w(t) = a \sin(2t + \theta)$ with unknown a and θ .

In order to achieve the design, the polynomial $A(s)$ must contain the disturbance model $(s^2 + 4)$. Note that the reference model s is not needed because the plant already contains the factor. Consider

$$A(s)D(s) + B(s)N(s) = F(s)$$

For this equation, we have $\deg D(s) = n = 1$. If $m = n - 1 = 0$, then the solution is unique and we have no freedom in assigning $A(s)$. If $m = 2$, then we have two free parameters which can be used to assign $A(s)$. Let

$$A(s) = \tilde{A}_0(s^2 + 4), \quad B(s) = B_0 + B_1 s + B_2 s^2$$

Define

$$\tilde{D}(s) := D(s)(s^2 + 4) = \tilde{D}_0 + \tilde{D}_1 s + \tilde{D}_2 s^2 + \tilde{D}_3 s^3 = 0 + 4s + 0 \cdot s^2 + s^3$$

We write $A(s)D(s) + B(s)N(s) = F(s)$ as

$$\tilde{A}_0 \tilde{D}(s) + B(s)N(s) = F(s)$$

Equating its coefficients, we obtain

$$[\tilde{A}_0 \ B_0 \ B_1 \ B_2] \begin{bmatrix} \tilde{D}_0 & \tilde{D}_1 & \tilde{D}_2 & \tilde{D}_3 \\ N_0 & N_1 & 0 & 0 \\ 0 & N_0 & N_1 & 0 \\ 0 & 0 & N_0 & N_1 \end{bmatrix} = [F_0 \ F_1 \ F_2 \ F_3]$$

For this example, if we select

$$F(s) = (s + 2)(s^2 + 2s + 2) = s^3 + 4s^2 + 6s + 4$$

then the equation becomes

$$[\tilde{A}_0 \ B_0 \ B_1 \ B_2] \begin{bmatrix} 0 & 4 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = [4 \ 6 \ 4 \ 1]$$

Its solution is [1 4 2 4]. Thus the compensator is

$$C(s) = \frac{B(s)}{A(s)} = \frac{4s^2 + 2s + 4}{1 \times (s^2 + 4)} = \frac{4s^2 + 2s + 4}{s^2 + 4}$$

This biproper compensator will place the poles of the unity feedback system in the assigned positions, track any step reference input, and reject the disturbance $a \sin(2t + \theta)$, both asymptotically and robustly.

9.4 Implementable Transfer Functions

Consider again the design problem posed in Fig. 8.1 with a given plant transfer function $\hat{g}(s)$. Now the problem is: given a desired overall transfer function $\hat{g}_o(s)$, find a feedback configuration and compensators so that the transfer function from r to y equals $\hat{g}_o(s)$. This is called the *model matching* problem. This problem is clearly different from the pole-placement problem. In pole placement, we specify only poles; its design will introduce some zeros which we have no control. In model matching, we specify not only poles but also zeros. Thus model matching can be considered as pole-and-zero placement and should yield a better design. See Reference 7.

Given a proper plant transfer function $\hat{g}(s)$. We claim that $\hat{g}_o(s) = 1$ is the best possible overall transfer function we can design. Indeed, if $\hat{g}_o(s) = 1$, then $y(t) = r(t)$ for all $t \geq 0$ and for any $r(t)$. Thus the overall system can track immediately (not asymptotically) any reference input no matter how erratic $r(t)$ is. Note that although $y(t) = r(t)$, the power levels at the reference input and plant output may be different. The reference signal may be provided by turning a knob by hand; the plant output may be the angular position of an antenna with weight over several tons.

Although $\hat{g}_o(s) = 1$ is the best overall transfer function, we may not be able to match it for a given plant. The reason is that in matching or implementation, there are some physical constraints which every overall system should meet. These constraints are listed in the following:

1. All compensators used have proper rational transfer functions.
2. The configuration selected has *no plant leakage* in the sense that all forward paths from r to y pass through the plant.
3. The closed-loop transfer function of every possible input-output pair is proper and BIBO stable.

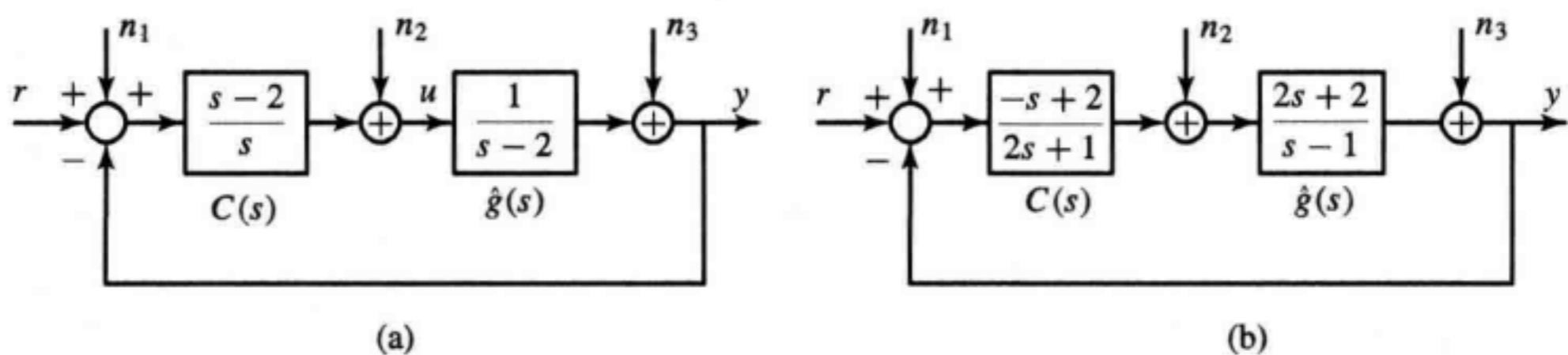


FIGURE 9.3 Feedback systems.

Every compensator with a proper rational transfer function can be implemented using the op-amp circuit elements shown in Fig. 4.2. If a compensator has an improper transfer function, then its implementation requires the use of pure differentiators which will amplify high-frequency noise. Thus in practice we often use only compensators with proper transfer functions. The second constraint requires that all power pass through the plant and no compensator be introduced in parallel with the plant. All configurations in Fig. 9.1 meet this constraint. In practice, noise and disturbance may exist in every component. For example, noise may be generated in using potentiometers because of brush jumps and wire irregularity. The load of an antenna may change because of gusting or air turbulence. These will be modeled as exogenous inputs entering the input and output terminals of every block as shown in Fig. 9.3. Clearly we cannot disregard the effects of these exogenous inputs on the system. Although the plant output is the signal we want to control, we should be concerned with all variables inside the system. For example, suppose the closed-loop transfer function from r to u is not BIBO stable, then any r will excite an unbounded u and the system will either saturate or burn out. If the closed-loop transfer function from n_1 to u is improper and if n_1 contains high-frequency noise, then the noise will be greatly amplified at u and make the system useless. Thus, the closed-loop transfer function of every possible input-output pair of the overall system should be proper and BIBO stable. An overall system is said to be *well posed* if the closed-loop transfer function of every possible input-output pair is proper; *totally stable* if the closed-loop transfer function of every possible input-output pair is BIBO stable.

Total stability can be readily met in design. If the overall transfer function from r to y is BIBO stable and if there is no unstable pole-zero cancelation in the system, then the overall system is totally stable. For example, consider the system shown in Fig. 9.3(a). The overall transfer function from r to y is

$$\hat{g}_{yr}(s) = \frac{C(s)\hat{g}(s)}{1 + C(s)\hat{g}(s)} = \frac{\frac{s-2}{s} \cdot \frac{1}{s-2}}{1 + \frac{s-2}{s} \cdot \frac{1}{s-2}} = \frac{\frac{1}{s}}{1 + \frac{1}{s}} = \frac{1}{s+1}$$

which is BIBO stable. However, the system is not totally stable because it involves an unstable pole-zero cancelation of $(s - 2)$. The closed-loop transfer function from n_2 to y can be computed as $s/(s - 2)(s + 1)$, which is not BIBO stable. Thus the output will grow unbounded if noise n_2 , even very small, enters the system. Thus we require not only the BIBO stability of $\hat{g}_o(s)$ but also of every possible closed-loop transfer function. Note that whether or not $\hat{g}(s)$ and $C(s)$ are BIBO stable is immaterial.

The condition for the unity feedback configuration in Fig. 9.3 to be well posed is $C(\infty)\hat{g}(\infty) \neq -1$ (Problem 9.9). This can be established using Mason's formula. See Reference 7, pp. 200–201. For example, for the unity feedback system in Fig. 9.3(b), we have $C(\infty)\hat{g}(\infty) = (-1/2) \times 2 = -1$. Thus the system is not well-posed. Indeed, the closed-loop transfer function from r to y is

$$\begin{aligned}\hat{g}_{yr}(s) &= \frac{C(s)\hat{g}(s)}{1 + C(s)\hat{g}(s)} = \frac{\frac{-s+2}{2s+1} \cdot \frac{2s+2}{s-1}}{1 + \frac{-s+2}{2s+1} \cdot \frac{2s+2}{s-1}} = \frac{(-s+2)(2s+2)}{(2s+1)(s-1) + (-s+2)(2s+2)} \\ &= \frac{(-s+2)(2s+2)}{2s^2 - s - 1 - 2s^2 + 2s + 4} = \frac{(-s+2)(2s+2)}{s+3}\end{aligned}$$

which is improper. The condition for the two-parameter configuration in Fig. 9.1(d) to be well-posed is $\hat{g}(\infty)C_2(\infty) \neq -1$. In the unity feedback and two-parameter configurations, if $\hat{g}(s)$ is strictly proper or $\hat{g}(\infty) = 0$, then $\hat{g}(\infty)C(\infty) = 0 \neq -1$ for any proper $C(s)$ and the overall systems will be automatically well-posed. In conclusion, total stability and well-posedness can be easily met in design. Nevertheless, they do impose some restrictions on $\hat{g}_o(s)$.

Definition 9.1 Given a plant with proper transfer function $\hat{g}(s)$. An overall transfer function $\hat{g}_o(s)$ is said to be *implementable* if there exists a no-plant-leakage configuration and proper compensators so that the transfer function from r to y in Fig. 8.1 equals $\hat{g}_o(s)$ and the overall system is well posed and totally stable.

If an overall transfer function $\hat{g}_o(s)$ is not implementable, then no matter what configuration is used to implement it, the design will violate at least one of the aforementioned constraints. Therefore, in model matching, the selected $\hat{g}_o(s)$ must be implementable; otherwise, it is not possible to implement it in practice.

► Theorem 9.4

Consider a plant with proper transfer function $\hat{g}(s)$. Then $\hat{g}_o(s)$ is implementable if and only if $\hat{g}_o(s)$ and

$$\hat{t}(s) := \frac{\hat{g}_o(s)}{\hat{g}(s)} \quad (9.24)$$

are proper and BIBO stable.

We show the necessity of Theorem 9.4. For any configuration that has no plant leakage, if the closed-loop transfer function from r to y is $\hat{g}_o(s)$, then we have

$$\hat{y}(s) = \hat{g}_o(s)\hat{r}(s) = \hat{g}(s)\hat{u}(s)$$

which implies

$$\hat{u}(s) = \frac{\hat{g}_o(s)}{\hat{g}(s)}\hat{r}(s) = \hat{t}(s)\hat{r}(s)$$

Thus the closed-loop transfer function from r to u is $\hat{t}(s)$. Total stability requires every closed-loop transfer function to be BIBO stable. Thus $\hat{g}_o(s)$ and $\hat{t}(s)$ must be BIBO stable. Well-posedness requires every closed-loop transfer function to be proper. Thus $\hat{g}_o(s)$ and $\hat{t}(s)$ must be proper. This shows the necessity of the theorem. The sufficiency will be established constructively in the next subsection.

We give a simple example. For the plant transfer function $\hat{g}(s) = 1/(s - 1)$, is $\hat{g}_o(s) = 1$ implementable? Although $\hat{g}_o(s) = 1$ is proper and BIBO stable, the transfer function

$$\hat{t}(s) = \frac{\hat{g}_o(s)}{\hat{g}(s)} = \frac{1}{1/(s - 1)} = \frac{s - 1}{1}$$

is not proper. Thus $\hat{g}_o(s) = 1$ is not implementable. In other words, for the plant $\hat{g}(s) = 1/(s - 1)$, it is not possible to design an overall system to have $\hat{g}_o(s) = 1$ as its transfer function without violating the aforementioned physical constraints.

► Corollary 9.4

Consider a plant with proper transfer function $\hat{g}(s) = N(s)/D(s)$. Then $\hat{g}_o(s) = E(s)/F(s)$ is implementable if and only if the following are true:

1. All roots of $F(s)$ have negative real parts ($F(s)$ is a CT stable polynomial).
2. $\deg F(s) - \deg E(s) \geq \deg D(s) - \deg N(s)$ (pole-zero excess inequality).
3. All zeros of $N(s)$ with zero or positive real parts are retained in $E(s)$ (retainment of nonminimum-phase zeros).

Note that $N(s)$ and $D(s)$ are assumed to be coprime. So are $E(s)$ and $F(s)$. We first develop Corollary 9.4 from Theorem 9.4. If $\hat{g}_o(s) = E(s)/F(s)$ is BIBO stable, then all roots of $F(s)$ have negative real parts. This is condition 1. We write (9.24) as

$$\hat{t}(s) = \frac{\hat{g}_o(s)}{\hat{g}(s)} = \frac{E(s)D(s)}{F(s)N(s)}$$

The condition for $\hat{t}(s)$ to be proper is

$$\deg F(s) + \deg N(s) \geq \deg E(s) + \deg D(s)$$

which implies 2. In order for $\hat{t}(s)$ to be BIBO stable, all roots of $N(s)$ with zero or positive real parts must be canceled by the roots of $E(s)$. Thus $E(s)$ must contain the nonminimum-phase zeros of $N(s)$. This is condition 3. Thus Corollary 9.4 follows directly from Theorem 9.4.

We give an example. Consider $\hat{g}(s) = (s + 2)(s - 1)/s(s - 2)(s + 3)$. It has one extra pole and the nonminimum-phase zero $s = 1$. Thus an implementable transfer function must have one or more extra poles and contain $(s - 1)$ as its zero. Clearly we have the following:

1. $\hat{g}_o(s) = 1$ is not implementable because it violates conditions 2 and 3.
2. $\hat{g}_o(s) = 1/(s + 1)$ is not implementable because it violates (3). Note that it meets 1 and 2.

3. $\hat{g}_o(s) = (s+2)/(s+1)(s+4)$ is not implementable because it violates condition 3. Note that it meets 1 and 2.
4. $\hat{g}_o(s) = k(s-1)/(s+1)(s+4)$ is implementable for any nonzero real constant k . If $k = -4$, then $\hat{g}_o(0) = 1$ and $\hat{g}_o(s)$ will track any step reference input without error.
5. $\hat{g}_o(s) = k(s+\alpha)(s-1)/(s+1)(s+4)$ is not implementable because it violates condition 2.
6. $\hat{g}_o(s) = k(s+\alpha)(s-1)/(s+1)(s+4)(s+5)$ is implementable for any real $k \neq 0$ and α . Let us expand $\hat{g}_o(s)$ as

$$\hat{g}_o(s) = \frac{ks^2 + k(\alpha - 1)s - k\alpha}{s^3 + 10s^2 + 29s + 20}$$

If we select $k(\alpha - 1) = 29$ and $-k\alpha = 20$ or $k = -49$ and $\alpha = 20/49$, then $\hat{g}_o(0) = 1$ and $\hat{g}'_o(0) = 0$, and $\hat{g}_o(s)$ will track any step and ramp reference inputs without error.

In pole placement, the design will always introduce some zeros which we have no control. In model matching, other than retaining nonminimum-phase zeros and meeting the pole-zero excess inequality, we have complete freedom in selecting poles and zeros: any pole inside the left half- s -plane and any zero in the entire s -plane. Thus model matching can be considered as pole-and-zero placement and should yield a better overall system than pole-placement design.

Given a plant transfer function $\hat{g}(s)$, how to select an implementable model $\hat{g}_o(s)$ is not a simple problem. For a discussion of this problem, see Reference 7, Chapter 9.

9.4.1 Model Matching—Two-Parameter Configuration

This section discusses the implementation problem. Consider a proper $\hat{g}(s)$ and an implementable $\hat{g}_o(s)$. If we use the open-loop configuration in Fig. 9.1(a) to implement $\hat{g}_o(s)$, then we have $\hat{g}(s)C(s) = \hat{g}_o(s)$ which implies

$$C(s) = \hat{g}_o(s)/\hat{g}(s) = \hat{t}(s)$$

where $\hat{t}(s)$ is defined in (9.24). The compensator is proper and the open-loop system is well posed. However if $\hat{g}(s)$ has CT unstable poles, the implementation will involve unstable pole-zero cancelations. In this case, the system is not totally stable and is not acceptable. If $\hat{g}(s)$ is BIBO stable, the implementation will be totally stable. Even so, the open-loop configuration can be very sensitive to plant parameter variations. Thus the open-loop configuration is generally not desirable. The unity feedback configuration in Fig. 9.1(b) can be used to achieve every pole placement; it, however, cannot be used to achieve every model matching as the next example shows.

Example 9.4.1 Consider a plant with transfer function $\hat{g}(s) = (s-2)/(s^2 - 1)$. We can readily show that

$$\hat{g}_o(s) = \frac{-(s-2)}{s^2 + 2s + 2} \quad (9.25)$$

is implementable. Because $\hat{g}_o(0) = 1$, the plant output will track asymptotically any step reference input. Suppose we use the unity feedback configuration in Fig. 9.1(b) with $p = 1$ to implement $\hat{g}_o(s)$. Then from

$$\hat{g}_o(s) = \frac{C(s)\hat{g}(s)}{1 + C(s)\hat{g}(s)}$$

we can compute the compensator as

$$C(s) = \frac{\hat{g}_o(s)}{\hat{g}(s)[1 - \hat{g}_o(s)]} = \frac{-(s^2 - 1)}{s(s + 3)}$$

This compensator is proper. However, the tandem connection of $C(s)$ and $\hat{g}(s)$ involves the pole-zero cancelation of $(s^2 - 1) = (s + 1)(s - 1)$. The cancelation of the stable pole $s + 1$ will not cause any serious problem in the overall system. However, the cancelation of the unstable pole $s - 1$ will make the overall system not totally stable. Thus the implementation is not acceptable.

Model matching in general involves some pole-zero cancelations. The same situation arises in state-feedback state-estimator design; all eigenvalues of the estimator are not controllable from the reference input and are canceled in the overall transfer function. However, because we have complete freedom in selecting the eigenvalues of the estimator, if we select them properly, the cancelation will not cause any problem in design. In using the unity feedback configuration in model matching, as we saw in the preceding example, the canceled poles are dictated by the plant transfer function. Thus, if a plant transfer function has unstable poles, the unity-feedback configuration cannot be used in model matching.

The open-loop and the unity-feedback configurations in Figs. 9.1(a) and 9.1(b) have one degree of freedom and cannot be used to achieve every model matching. The configurations in Figs. 9.1(c) and 9.1(d) both have two degrees of freedom. In using either configuration, we have complete freedom in assigning canceled poles; therefore, both can be used to achieve every model matching. Because the two-parameter configuration in Fig. 9.1(d) seems to be more natural and more suitable for practical implementation, we discuss only the configuration here. For model matching using the configuration in Fig. 9.1(c), see Reference 6.

Consider the two-parameter configuration in Fig. 9.1(d). Let

$$C_1(s) = \frac{L(s)}{A_1(s)}, \quad C_2(s) = \frac{M(s)}{A_2(s)}$$

where $L(s)$, $M(s)$, $A_1(s)$, and $A_2(s)$ are polynomials. We call $C_1(s)$ *feedforward compensator* and $C_2(s)$ *feedback compensator*. In general, $A_1(s)$ and $A_2(s)$ need not be the same. It turns out that even if they are chosen to be the same, the configuration can still be used to achieve every model matching. Furthermore, a simple design procedure can be developed. Therefore, we assume $A_1(s) = A_2(s) = A(s)$ and the compensators become

$$C_1(s) = \frac{L(s)}{A(s)}, \quad C_2(s) = \frac{M(s)}{A(s)} \quad (9.26)$$

The transfer function from r to y in Fig. 9.1(d) then becomes

$$\begin{aligned}\hat{g}_o(s) &= C_1(s) \frac{\hat{g}(s)}{1 + \hat{g}(s)C_2(s)} = \frac{L(s)}{A(s)} \frac{\frac{N(s)}{D(s)}}{1 + \frac{N(s)}{D(s)} \frac{M(s)}{A(s)}} \\ &= \frac{L(s)N(s)}{A(s)D(s) + M(s)N(s)}\end{aligned}\quad (9.27)$$

Thus in model matching, we search proper $L(s)/A(s)$ and $M(s)/A(s)$ to meet

$$\hat{g}_o(s) = \frac{E(s)}{F(s)} = \frac{L(s)N(s)}{A(s)D(s) + M(s)N(s)} \quad (9.28)$$

Note that the two-parameter configuration has no plant leakage. If the plant transfer function $\hat{g}(s)$ is strictly proper and if $C_2(s) = M(s)/A(s)$ is proper, then the overall system is automatically well posed. The question of total stability will be discussed in the next subsection. Note that if $\hat{g}(s)$ is biproper, the discussion must be modified slightly. See Reference 6.

► Problem

Given $\hat{g}(s) = N(s)/D(s)$, where $N(s)$ and $D(s)$ are coprime and $\deg N(s) < \deg D(s) = n$, and given an implementable $\hat{g}_o(s) = E(s)/F(s)$, find proper $L(s)/A(s)$ and $M(s)/A(s)$ to meet (9.28).

► Procedure 9.1

1. Compute

$$\frac{\hat{g}_o(s)}{N(s)} = \frac{E(s)}{F(s)N(s)} =: \frac{\bar{E}(s)}{\bar{F}(s)} \quad (9.29)$$

where $\bar{E}(s)$ and $\bar{F}(s)$ are coprime. Since $E(s)$ and $F(s)$ are implicitly assumed to be coprime, common factors may exist only between $E(s)$ and $N(s)$. Cancel all common factors between them and denote the rest $\bar{E}(s)$ and $\bar{F}(s)$. Note that if $E(s) = N(s)$, then $\bar{F}(s) = F(s)$ and $\bar{E}(s) = 1$. Using (9.29), we rewrite (9.28) as

$$\hat{g}_o(s) = \frac{\bar{E}(s)N(s)}{\bar{F}(s)} = \frac{L(s)N(s)}{A(s)D(s) + M(s)N(s)} \quad (9.30)$$

From this equation, we may be tempted to set $L(s) = \bar{E}(s)$ and solve for $A(s)$ and $M(s)$ from $\bar{F}(s) = A(s)D(s) + M(s)N(s)$. However, the computed $C_2(s) = M(s)/A(s)$ may be improper. This violates the requirement that all compensators be proper. Thus we need some additional manipulation.

2. Introduce an arbitrary CT stable polynomial $\hat{F}(s)$ such that the degree of $\bar{F}(s)\hat{F}(s)$ is $2n - 1$ or larger. In other words, if $\deg \bar{F}(s) = p$, then $\deg \hat{F}(s) \geq 2n - 1 - p$. Because the polynomial $\hat{F}(s)$ will be canceled in the design, its roots should be chosen to lie inside the sector shown in Fig. 8.3.
3. Rewrite (9.30) as

$$\hat{g}_o(s) = \frac{\bar{E}(s)\hat{F}(s)N(s)}{\bar{F}(s)\hat{F}(s)} = \frac{L(s)N(s)}{A(s)D(s) + M(s)N(s)} \quad (9.31)$$

Now we set

$$L(s) = \bar{E}(s)\hat{F}(s) \quad (9.32)$$

and solve $A(s)$ and $M(s)$ from

$$A(s)D(s) + M(s)N(s) = \hat{F}(s)\bar{F}(s) \quad (9.33)$$

If we write

$$\begin{aligned} A(s) &= A_0 + A_1 s + A_2 s^2 + \cdots + A_m s^m \\ M(s) &= M_0 + M_1 s + M_2 s^2 + \cdots + M_m s^m \\ \bar{F}(s)\hat{F}(s) &= F_0 + F_1 s + F_2 s^2 + \cdots + F_{n+m} s^{n+m} \end{aligned}$$

with $m \geq n - 1$, then $A(s)$ and $M(s)$ can be obtained by solving

$$[A_0 \ M_0 \ A_1 \ M_1 \ \cdots \ A_m \ M_m] \mathbf{S}_m = [F_0 \ F_1 \ F_2 \ \cdots \ F_{n+m}] \quad (9.34)$$

with

$$\mathbf{S}_m := \begin{bmatrix} D_0 & D_1 & \cdots & D_n & 0 & \cdots & 0 \\ N_0 & N_1 & \cdots & N_n & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & D_0 & \cdots & D_{n-1} & D_n & \cdots & 0 \\ 0 & N_0 & \cdots & N_{n-1} & N_n & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & D_0 & \cdots & D_n \\ 0 & 0 & \cdots & 0 & N_0 & \cdots & N_n \end{bmatrix}$$

The computed compensators $L(s)/A(s)$ and $M(s)/A(s)$ are proper.

We justify the procedure. By introducing $\hat{F}(s)$, the degree of $\bar{F}(s)\hat{F}(s)$ is $2n - 1$ or larger and, following Theorem 9.2, solution $A(s)$ and $M(s)$ with $\deg M(s) \leq \deg A(s) = m$ and $m \geq n - 1$ exists in (9.34) for any $\bar{F}(s)\hat{F}(s)$. Thus the compensator $M(s)/A(s)$ is proper. Note that if we do not introduce $\hat{F}(s)$, proper compensator $M(s)/A(s)$ may not exist in (9.34).

Next we show $\deg L(s) \leq \deg A(s)$. Applying the pole-zero excess inequality to (9.31) and using (9.32) yield

$$\deg(\bar{F}(s)\hat{F}(s)) - \deg N(s) - \deg L(s) \geq \deg D(s) - \deg N(s)$$

which implies

$$\deg L(s) \leq \deg(\bar{F}(s)\hat{F}(s)) - \deg D(s) = \deg A(s)$$

Thus the compensator $L(s)/A(s)$ is proper.

Example 9.4.2 Consider the model matching problem studied in Example 9.4.1. That is, given $\hat{g}(s) = (s - 2)/(s^2 - 1)$, match $\hat{g}_o(s) = -(s - 2)/(s^2 + 2s + 2)$. We implement it using the two-parameter configuration shown in Fig. 9.1(d). First we compute

$$\frac{\hat{g}_o(s)}{N(s)} = \frac{-(s - 2)}{(s^2 + 2s + 2)(s - 2)} = \frac{-1}{s^2 + 2s + 2} =: \frac{\bar{E}(s)}{\bar{F}(s)}$$

where $\bar{E}(s) = -1$ and $\bar{F}(s) = s^2 + 2s + 2$ are coprime. Because the degree of $\bar{F}(s)$ is 2, we select arbitrarily $\hat{F}(s) = s + 4$ so that the degree of $\bar{F}(s)\hat{F}(s)$ is $3 = 2n - 1$. Thus we have

$$L(s) = \bar{E}(s)\hat{F}(s) = -(s + 4) \quad (9.35)$$

and $A(s)$ and $M(s)$ can be solved from

$$\begin{aligned} A(s)D(s) + M(s)N(s) &= \hat{F}(s)\bar{F}(s) = (s^2 + 2s + 2)(s + 4) \\ &= s^3 + 6s^2 + 10s + 8 \end{aligned}$$

or

$$[A_0 \ M_0 \ A_1 \ M_1] \begin{bmatrix} -1 & 0 & 1 & 0 \\ -2 & 1 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & -1 & 0 & 1 \\ 0 & -2 & 1 & 0 \end{bmatrix} = [8 \ 10 \ 6 \ 1]$$

The solution is $A_0 = 18$, $A_1 = 1$, $M_0 = -13$ and $M_1 = -12$. Thus we have $A(s) = 18 + s$ and $M(s) = -13 - 12s$ and the compensators are

$$C_1(s) = \frac{L(s)}{A(s)} = \frac{-(s + 4)}{s + 18}, \quad C_2(s) = \frac{M(s)}{A(s)} = \frac{-(12s + 13)}{s + 18}$$

This completes the design. Note that, because $\hat{g}_o(0) = 1$, the output of the feedback system will track any step reference input.

The design in this example can also be carried out using state-space equations. We find a two-dimensional controllable-form realization of the plant. Its state feedback gain can then be readily obtained. We then design a one-dimensional state estimator by solving a Lyapunov equation. Applying the feedback gain to the outputs of the plant and the estimator and introducing the feedforward gain $p = -1$, we can complete the design. However, the design is fairly complex. Even worse, for the design problem in the next example, it is not obvious how to carry it out using state-space equations.

Example 9.4.3 Given $\hat{g}(s) = (s - 2)/(s^2 - 1)$. Match

$$\hat{g}_o(s) = \frac{-(s - 2)(4s + 2)}{(s^2 + 2s + 2)(s + 2)} = \frac{-4s^2 + 6s + 4}{s^3 + 4s^2 + 6s + 4}$$

This $\hat{g}_o(s)$ is BIBO stable and has the property $\hat{g}_o(0) = 1$ and $\hat{g}'_o(0) = 0$, thus the overall system will track asymptotically not only any step reference input but also any

ramp input. See Problems 9.13 and 9.14. This $\hat{g}_o(s)$ meets all the three conditions in Corollary 9.4, thus it is implementable. We use the two-parameter configuration. First we compute

$$\frac{\hat{g}_o(s)}{N(s)} = \frac{-(s-2)(4s+2)}{(s^2+2s+2)(s+2)(s-2)} = \frac{-(4s+2)}{s^3+4s^2+6s+4} =: \frac{\bar{E}(s)}{\bar{F}(s)}$$

Because the degree of $\bar{F}(s)$ is 3, which equals $2n - 1 = 3$, there is no need to introduce $\hat{F}(s)$ and we set $\hat{F}(s) = 1$. Thus we have

$$L(s) = \hat{F}(s)\bar{E}(s) = -(4s+2)$$

and $A(s)$ and $M(s)$ can be solved from

$$[A_0 \ M_0 \ A_1 \ M_1] \begin{bmatrix} -1 & 0 & 1 & 0 \\ -2 & 1 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & -1 & 0 & 1 \\ 0 & -2 & 1 & 0 \end{bmatrix} = [4 \ 6 \ 4 \ 1]$$

as $A_1 = 1$, $A_0 = 34/3 = 11.33$, $M_0 = -23/3 = -7.67$, and $M_1 = -22/3 = -7.33$. Thus the compensators are

$$C_1(s) = \frac{-(4s+2)}{s+11.33}, \quad C_2(s) = \frac{-7.33s-7.67}{s+11.33}$$

This completes the design. Note that this design does not involve any pole-zero cancelation because $\hat{F}(s) = 1$.

9.4.2 Implementation of Two-Parameter Compensators

Given a plant with transfer function $\hat{g}(s)$ and an implementable model $\hat{g}_o(s)$, we can implement the model in the two-parameter configuration shown in Fig. 9.1(d) and redrawn in Fig. 9.4(a). The compensators $C_1(s) = L(s)/A(s)$ and $C_2(s) = M(s)/A(s)$ can be obtained by using Procedure 9.1. To complete the design, the compensators must be built or implemented. This is discussed in this subsection.

Consider the configuration in Fig. 9.4(a). The denominator $A(s)$ of $C_1(s)$ is obtained by solving the compensator equation in (9.33) and may or may not be a CT stable polynomial. See Problem 9.12. If it is not a CT stable polynomial and if we implement $C_1(s)$ as shown in Fig. 9.4(a), then the output of $C_1(s)$ will grow unbounded and the overall system is not totally stable. Therefore, in general, we should not implement the two compensators as shown in Fig. 9.4(a). If we move $C_2(s)$ outside the loop as shown in Fig. 9.4(b), then the design will involve the cancelation of $M(s)$. Because $M(s)$ is also obtained by solving (9.33), we have no direct control of $M(s)$. Thus the design is in general not acceptable. If we move $C_1(s)$ inside the loop, then the configuration becomes the one shown in Fig. 9.4(c). We see that the connection involves the pole-zero cancelation of $L(s) = \hat{F}(s)\bar{E}(s)$. We have freedom in selecting $\hat{F}(s)$. The polynomial $\bar{E}(s)$ is part of $E(s)$ which, other than the nonminimum-phase zeros of

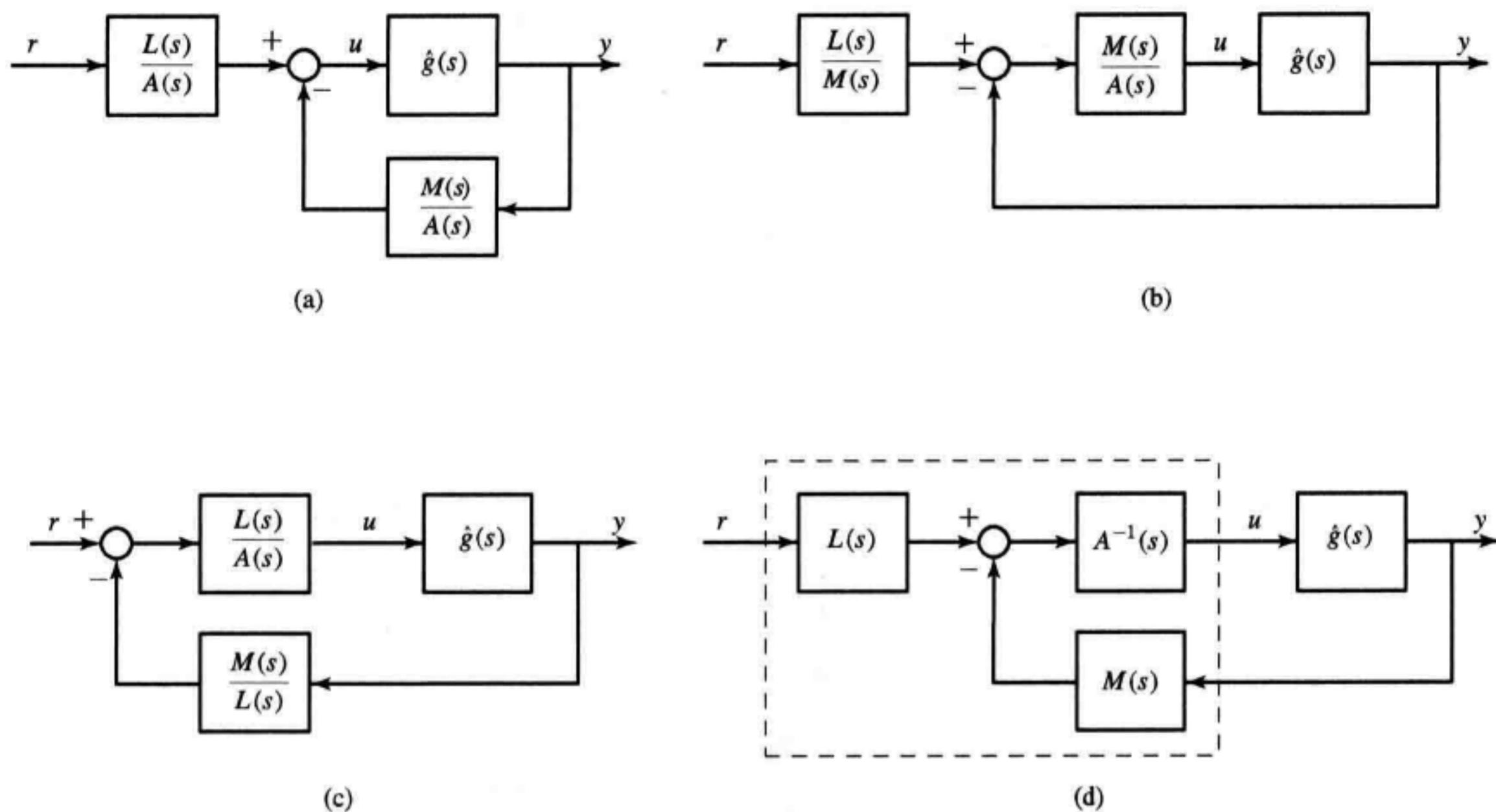


FIGURE 9.4 Two-degree-of-freedom configurations.

$N(s)$, we can also select. The nonminimum-phase zeros, however, are completely canceled in $\bar{E}(s)$. Thus $L(s)$ can be CT stable⁴ and the implementation in Fig. 9.4(c) can be totally stable and is acceptable. However, because the two compensators $L(s)/A(s)$ and $M(s)/L(s)$ have different denominators, their implementations require a total of $2m$ integrators. We discuss next a better implementation that requires only m integrators and involves only the cancelation of $\hat{F}(s)$.

Consider

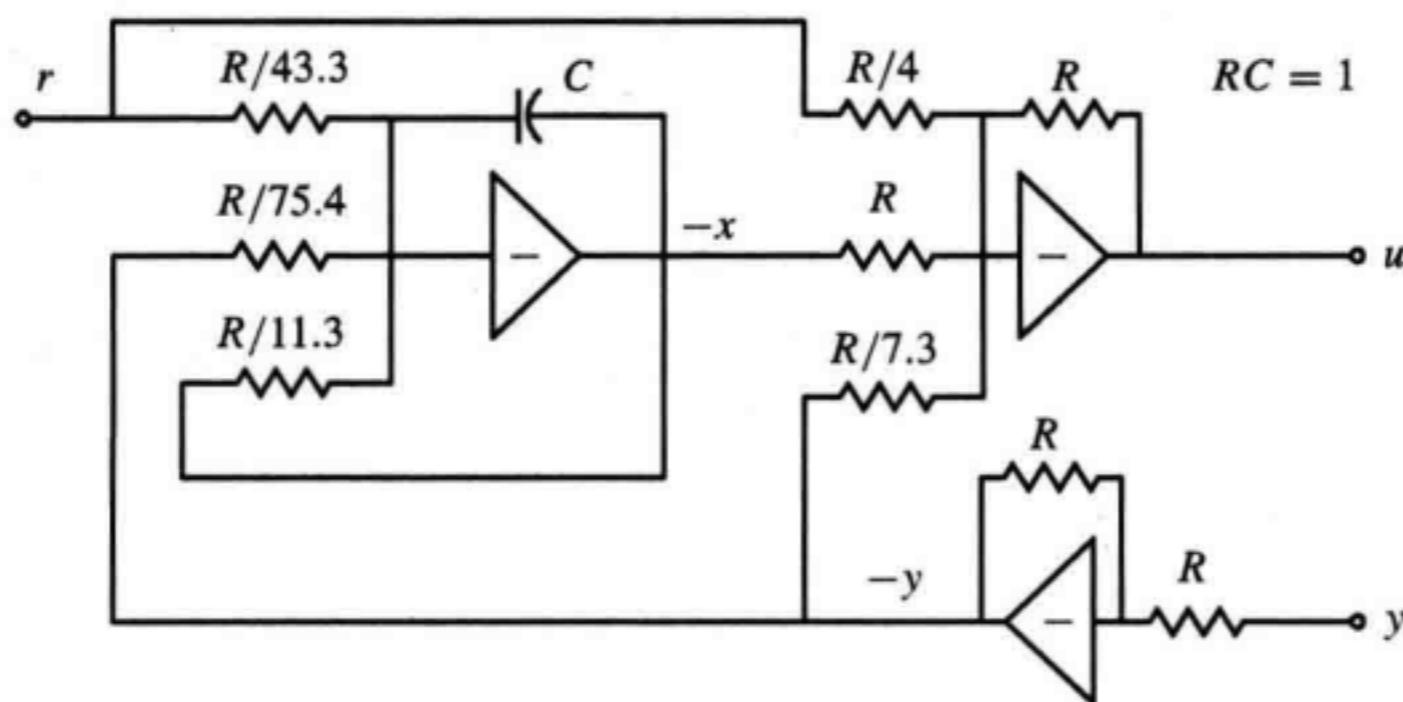
$$\begin{aligned}\hat{u}(s) &= C_1(s)\hat{r}(s) - C_2(s)\hat{y}(s) = \frac{L(s)}{A(s)}\hat{r}(s) - \frac{M(s)}{A(s)}\hat{y}(s) \\ &= A^{-1}(s)[L(s) - M(s)] \begin{bmatrix} \hat{r}(s) \\ \hat{y}(s) \end{bmatrix}\end{aligned}$$

This can be plotted as shown in Fig. 9.4(d). Thus we can consider the two compensators as a single compensator with two inputs r and y and one output u with transfer matrix

$$\mathbf{C}(s) = [C_1(s) \quad -C_2(s)] = A^{-1}(s)[L(s) - M(s)] \quad (9.36)$$

If we find a minimal realization of (9.36), then its dimension is m and the two compensators can be implemented using only m integrators. As we can see from (9.31), the design involves only the cancelation of $\hat{F}(s)$. Thus the implementation in Fig. 9.4(d) is

⁴This may not be true in the MIMO case.

**FIGURE 9.5** Op-amp circuit implementation.

superior to the one in Fig. 9.4(c). In conclusion, the four configurations in Fig. 9.4 all have two degrees of freedom and are mathematically equivalent, they however, can be different in actual implementation.

Example 9.4.4 Implement the compensators in Example 9.4.3 using an op-amp circuit. We write

$$\begin{aligned}\hat{u}(s) &= C_1(s)\hat{r}(s) - C_2(s)\hat{y}(s) = \begin{bmatrix} -(4s+2) & 7.33s+7.67 \\ s+11.33 & s+11.33 \end{bmatrix} \begin{bmatrix} \hat{r}(s) \\ \hat{y}(s) \end{bmatrix} \\ &= \left([-4 \ 7.33] + \frac{1}{s+11.33} [43.33 \ -75.38] \right) \begin{bmatrix} \hat{r}(s) \\ \hat{y}(s) \end{bmatrix}\end{aligned}$$

Its state-space realization is, using the formula in Problem 4.14,

$$\begin{aligned}\dot{x} &= -11.33x + [43.33 \ -75.38] \begin{bmatrix} r \\ y \end{bmatrix} \\ u &= x + [-4 \ 7.33] \begin{bmatrix} r \\ y \end{bmatrix}\end{aligned}$$

(See Problem 4.18.) This equation can be implemented using an op-amp circuit as shown in Fig. 9.5. Note that the adder with positive and negative signs in Fig. 9.4(d) is no longer recognizable in Fig. 9.5. Nor are the feedforward and feedback compensators.

9.5 MIMO Unity Feedback Systems

This section extends the pole placement discussed in Section 9.3 to the MIMO case. Consider the unity-feedback system shown in Fig. 9.6. The plant has p inputs and q outputs and is described by a $q \times p$ strictly proper rational matrix $\hat{\mathbf{G}}(s)$. We assume that $\hat{\mathbf{G}}(s)$ has full rank in the sense that it has a $q \times q$ or $p \times p$ submatrix with a nonzero determinant. If $\hat{\mathbf{G}}(s)$ is square, then its determinant is nonzero or its inverse exists. This is equivalent to the assumption that if $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is a minimal realization of the transfer matrix, then \mathbf{B} has full column rank and \mathbf{C} has full row rank.

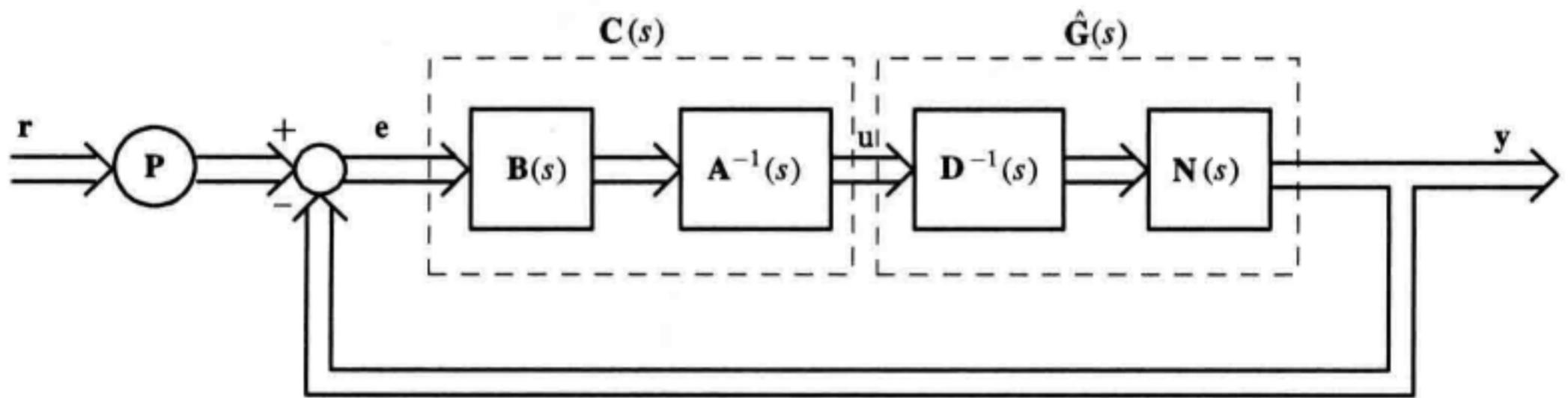


FIGURE 9.6 MIMO unity feedback system with $\mathbf{P} = \mathbf{I}_q$.

The design problem is to find a compensator to achieve pole placement. The compensator $\mathbf{C}(s)$ to be designed must have q inputs and p outputs in order for the connection to be possible. Thus $\mathbf{C}(s)$ is required to be a $p \times q$ proper rational matrix. The matrix \mathbf{P} is a $q \times q$ constant gain matrix. For the time being, we assume $\mathbf{P} = \mathbf{I}_q$. Let the transfer matrix from \mathbf{r} to \mathbf{y} be denoted by $\hat{\mathbf{G}}_o(s)$, a $q \times q$ matrix. Then we have

$$\begin{aligned}\hat{\mathbf{G}}_o(s) &= [\mathbf{I}_q + \hat{\mathbf{G}}(s)\mathbf{C}(s)]^{-1}\hat{\mathbf{G}}(s)\mathbf{C}(s) \\ &= \hat{\mathbf{G}}(s)\mathbf{C}(s)[\mathbf{I}_q + \hat{\mathbf{G}}(s)\mathbf{C}(s)]^{-1} \\ &= \hat{\mathbf{G}}(s)[\mathbf{I}_p + \mathbf{C}(s)\hat{\mathbf{G}}(s)]^{-1}\mathbf{C}(s)\end{aligned}\quad (9.37)$$

The first equality is obtained from $\hat{\mathbf{y}}(s) = \hat{\mathbf{G}}(s)\mathbf{C}(s)[\hat{\mathbf{r}}(s) - \hat{\mathbf{y}}(s)]$; the second one from $\hat{\mathbf{e}}(s) = \hat{\mathbf{r}}(s) - \hat{\mathbf{G}}(s)\mathbf{C}(s)\hat{\mathbf{e}}(s)$; and the third one from $\hat{\mathbf{u}}(s) = \mathbf{C}(s)[\hat{\mathbf{r}}(s) - \hat{\mathbf{G}}(s)\hat{\mathbf{u}}(s)]$. They can also be directly verified. For example, pre- and postmultiplying $[\mathbf{I}_q + \hat{\mathbf{G}}(s)\mathbf{C}(s)]$ to the first two equations yield

$$\hat{\mathbf{G}}(s)\mathbf{C}(s)[\mathbf{I}_q + \hat{\mathbf{G}}(s)\mathbf{C}(s)] = [\mathbf{I}_q + \hat{\mathbf{G}}(s)\mathbf{C}(s)]\hat{\mathbf{G}}(s)\mathbf{C}(s)$$

which is an identity. This establishes the second equality. The third equality can be similarly established.

Let $\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)$ be a right coprime fraction and let $\mathbf{C}(s) = \mathbf{A}^{-1}(s)\mathbf{B}(s)$ be a left fraction to be designed. Then (9.37) implies

$$\begin{aligned}\hat{\mathbf{G}}_o(s) &= \mathbf{N}(s)\mathbf{D}^{-1}(s)[\mathbf{I} + \mathbf{A}^{-1}(s)\mathbf{B}(s)\mathbf{N}(s)\mathbf{D}^{-1}(s)]^{-1}\mathbf{A}^{-1}(s)\mathbf{B}(s) \\ &= \mathbf{N}(s)\mathbf{D}^{-1}(s) \left\{ \mathbf{A}^{-1}(s)[\mathbf{A}(s)\mathbf{D}(s) \right. \\ &\quad \left. + \mathbf{B}(s)\mathbf{N}(s)]\mathbf{D}^{-1}(s) \right\}^{-1}\mathbf{A}^{-1}(s)\mathbf{B}(s) \\ &= \mathbf{N}(s)[\mathbf{A}(s)\mathbf{D}(s) + \mathbf{B}(s)\mathbf{N}(s)]^{-1}\mathbf{B}(s) \\ &= \mathbf{N}(s)\mathbf{F}^{-1}(s)\mathbf{B}(s)\end{aligned}\quad (9.38)$$

where

$$\mathbf{A}(s)\mathbf{D}(s) + \mathbf{B}(s)\mathbf{N}(s) = \mathbf{F}(s) \quad (9.39)$$

It is a polynomial matrix equation. Thus the design problem becomes: Given $p \times p$ $\mathbf{D}(s)$ and $q \times p$ $\mathbf{N}(s)$ and an arbitrary $p \times p$ $\mathbf{F}(s)$, find $p \times p$ $\mathbf{A}(s)$ and $p \times q$ $\mathbf{B}(s)$ to meet (9.39). This is the matrix version of the polynomial compensator equation in (9.12).

► Theorem 9.M1

Given polynomial matrices $\mathbf{D}(s)$ and $\mathbf{N}(s)$, polynomial matrix solutions $\mathbf{A}(s)$ and $\mathbf{B}(s)$ exist in (9.39) for any polynomial matrix $\mathbf{F}(s)$ if and only if $\mathbf{D}(s)$ and $\mathbf{N}(s)$ are right coprime.

Suppose $\mathbf{D}(s)$ and $\mathbf{N}(s)$ are not right coprime, then there exists a nonunimodular polynomial matrix $\mathbf{R}(s)$ such that $\mathbf{D}(s) = \hat{\mathbf{D}}(s)\mathbf{R}(s)$ and $\mathbf{N}(s) = \hat{\mathbf{N}}(s)\mathbf{R}(s)$. Then $\mathbf{F}(s)$ in (9.39) must be of the form $\hat{\mathbf{F}}(s)\mathbf{R}(s)$, for some polynomial matrix $\hat{\mathbf{F}}(s)$. Thus if $\mathbf{F}(s)$ cannot be expressed in such a form, then no solutions exist in (9.39). This shows the necessity of the theorem. If $\mathbf{D}(s)$ and $\mathbf{N}(s)$ are right coprime, there exist polynomial matrices $\bar{\mathbf{A}}(s)$ and $\bar{\mathbf{B}}(s)$ such that

$$\bar{\mathbf{A}}(s)\mathbf{D}(s) + \bar{\mathbf{B}}(s)\mathbf{N}(s) = \mathbf{I}$$

The polynomial matrices $\bar{\mathbf{A}}(s)$ and $\bar{\mathbf{B}}(s)$ can be obtained by a sequence of elementary operations. See Reference 6, pp. 587–595. Thus $\mathbf{A}(s) = \mathbf{F}(s)\bar{\mathbf{A}}(s)$ and $\mathbf{B}(s) = \mathbf{F}(s)\bar{\mathbf{B}}(s)$ are solutions of (9.39) for any $\mathbf{F}(s)$. This establishes Theorem 9.M1. As in the scalar case, it is possible to develop general solutions for (9.39). However, the general solutions are not convenient to use in our design. Thus it will not be discussed.

Next we will change solving (9.39) into solving a set of linear algebraic equations. Consider $\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)$ where $\mathbf{D}(s)$ and $\mathbf{N}(s)$ are right coprime and $\mathbf{D}(s)$ is column reduced. Let μ_i be the degree of the i th column of $\mathbf{D}(s)$. Then we have, as discussed in Section 7.8.2,

$$\deg \hat{\mathbf{G}}(s) = \deg \det \mathbf{D}(s) = \mu_1 + \mu_2 + \cdots + \mu_p =: n \quad (9.40)$$

Let $\mu := \max(\mu_1, \mu_2, \dots, \mu_p)$. Then we can express $\mathbf{D}(s)$ and $\mathbf{N}(s)$ as

$$\begin{aligned} \mathbf{D}(s) &= \mathbf{D}_0 + \mathbf{D}_1 s + \mathbf{D}_2 s^2 + \cdots + \mathbf{D}_\mu s^\mu \quad \mathbf{D}_\mu \neq \mathbf{0} \\ \mathbf{N}(s) &= \mathbf{N}_0 + \mathbf{N}_1 s + \mathbf{N}_2 s^2 + \cdots + \mathbf{N}_\mu s^\mu \end{aligned}$$

Note that \mathbf{D}_μ is singular unless $\mu_1 = \mu_2 = \cdots = \mu_p$. Note also that $\mathbf{N}_\mu = \mathbf{0}$ following the strict properness assumption of $\hat{\mathbf{G}}(s)$. We also express $\mathbf{A}(s)$, $\mathbf{B}(s)$, and $\mathbf{F}(s)$ as

$$\begin{aligned} \mathbf{A}(s) &= \mathbf{A}_0 + \mathbf{A}_1 s + \mathbf{A}_2 s^2 + \cdots + \mathbf{A}_m s^m \\ \mathbf{B}(s) &= \mathbf{B}_0 + \mathbf{B}_1 s + \mathbf{B}_2 s^2 + \cdots + \mathbf{B}_m s^m \\ \mathbf{F}(s) &= \mathbf{F}_0 + \mathbf{F}_1 s + \mathbf{F}_2 s^2 + \cdots + \mathbf{F}_{\mu+m} s^{\mu+m} \end{aligned}$$

Substituting these into (9.39) and matching the coefficients of like power of s yield

$$[\mathbf{A}_0 \ \mathbf{B}_0 \ \mathbf{A}_1 \ \mathbf{B}_1 \ \cdots \ \mathbf{A}_m \ \mathbf{B}_m] \mathbf{S}_m = [\mathbf{F}_0 \ \mathbf{F}_1 \ \cdots \ \mathbf{F}_{\mu+m}] =: \bar{\mathbf{F}} \quad (9.41)$$

where

$$\mathbf{S}_m := \left[\begin{array}{ccccccccc} \mathbf{D}_0 & \mathbf{D}_1 & \cdots & \mathbf{D}_\mu & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \cdots & \cdots \\ \mathbf{N}_0 & \mathbf{N}_1 & \cdots & \mathbf{N}_\mu & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{D}_0 & \cdots & \mathbf{D}_{\mu-1} & \mathbf{D}_\mu & \mathbf{0} & \cdots & \mathbf{0} \\ \cdots & \cdots \\ \mathbf{0} & \mathbf{N}_0 & \cdots & \mathbf{N}_{\mu-1} & \mathbf{N}_\mu & \mathbf{0} & \cdots & \mathbf{0} \\ \hline \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{D}_0 & \mathbf{D}_1 & \cdots & \mathbf{D}_\mu \\ \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{N}_0 & \mathbf{N}_1 & \cdots & \mathbf{N}_\mu \end{array} \right] \quad (9.42)$$

The matrix \mathbf{S}_m has $m+1$ block rows; each block row consists of $p D$ -rows and $q N$ -rows. Thus \mathbf{S}_m has $(m+1)(p+q)$ number of rows. Let us search linearly independent rows of \mathbf{S}_m in order from top to bottom. It turns out that if $\mathbf{N}(s)\mathbf{D}^{-1}(s)$ is proper, then all D -rows are linearly independent of their previous rows. An N -row can be linearly independent of its previous rows. However, if an N -row becomes linearly dependent, then, because of the structure of \mathbf{S}_m , the same N -rows in subsequent block rows will be linearly dependent. Let v_i be the number of linearly independent i th N -rows and let

$$v := \max\{v_1, v_2, \dots, v_q\}$$

It is called the *row index* of $\hat{\mathbf{G}}(s)$. Then all $q N$ -rows in the last block row of \mathbf{S}_v are linearly dependent on their previous rows. Thus \mathbf{S}_{v-1} contains all linearly independent N -rows and its total number equals, as discussed in Section 7.8.2, the degree of $\hat{\mathbf{G}}(s)$, that is,

$$v_1 + v_2 + \cdots + v_q = n \quad (9.43)$$

Because all D -rows are linearly independent and there are a total of $p v$ D -rows in \mathbf{S}_{v-1} , we conclude that \mathbf{S}_{v-1} has $n + p v$ independent rows or rank $n + p v$.

Let us consider

$$\mathbf{S}_0 = \left[\begin{array}{ccccc} \mathbf{D}_0 & \mathbf{D}_1 & \cdots & \mathbf{D}_{\mu-1} & \mathbf{D}_\mu \\ \mathbf{N}_0 & \mathbf{N}_1 & \cdots & \mathbf{N}_{\mu-1} & \mathbf{N}_\mu \end{array} \right]$$

It has $p(\mu+1)$ number of columns; however, it has at least a total of $\sum_{i=1}^p (\mu - \mu_i)$ zero columns. In the matrix \mathbf{S}_1 , some new zero columns will appear in the rightmost block column. However, some zero columns in \mathbf{S}_0 will not be zero columns in \mathbf{S}_1 . Thus the number of zero columns in \mathbf{S}_1 remains to be

$$\alpha := \sum_{i=1}^p (\mu - \mu_i) = p\mu - (\mu_1 + \mu_2 + \cdots + \mu_p) = p\mu - n \quad (9.44)$$

In fact, this is the number of zero columns in \mathbf{S}_m , $m = 2, 3, \dots$. Let $\tilde{\mathbf{S}}_{\mu-1}$ be the matrix $\mathbf{S}_{\mu-1}$ after deleting these zero columns. Because the number of columns in \mathbf{S}_m is $p(\mu + 1 + m)$, the number of columns in $\tilde{\mathbf{S}}_{\mu-1}$ is

$$\beta := p(\mu + 1 + v - 1) - (p\mu - n) = pv + n \quad (9.45)$$

The rank of $\tilde{\mathbf{S}}_{\mu-1}$ clearly equals the rank of $\mathbf{S}_{\mu-1}$ or $pv + n$. Thus $\tilde{\mathbf{S}}_{\mu-1}$ has full column rank. Now if m increases by 1, the rank and the number of the columns of $\tilde{\mathbf{S}}_\mu$ both increase by p (because the p new D -rows are all linearly independent of their previous rows); thus $\tilde{\mathbf{S}}_\mu$ still has full column rank. Proceeding forward, we conclude that $\tilde{\mathbf{S}}_m$, for $m \geq \mu - 1$, has full column rank.

Let us define

$$\mathbf{H}_c(s) := \text{diag}(s^{\mu_1}, s^{\mu_2}, \dots, s^{\mu_p}) \quad (9.46)$$

and

$$\mathbf{H}_r(s) = \text{diag}(s^{m_1}, s^{m_2}, \dots, s^{m_p}) \quad (9.47)$$

Then we have the following matrix version of Theorem 9.2.

► **Theorem 9.M2**

Consider the unity feedback system shown in Fig. 9.6. The plant is described by a $q \times p$ strictly proper rational matrix $\hat{\mathbf{G}}(s)$. Let $\hat{\mathbf{G}}(s)$ be factored as $\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)$, where $\mathbf{D}(s)$ and $\mathbf{N}(s)$ are right coprime and $\mathbf{D}(s)$ is column reduced with column degrees μ_i , $i = 1, 2, \dots, p$. Let v be the row index of $\hat{\mathbf{G}}(s)$ and let $m_i \geq v - 1$ for $i = 1, 2, \dots, p$. Then for any $p \times p$ polynomial matrix $\mathbf{F}(s)$ such that

$$\lim_{s \rightarrow \infty} \mathbf{H}_r^{-1}(s)\mathbf{F}(s)\mathbf{H}_c^{-1}(s) = \mathbf{F}_h \quad (9.48)$$

is a nonsingular constant matrix, there exists a $p \times q$ proper compensator $\mathbf{A}^{-1}(s)\mathbf{B}(s)$, where $\mathbf{A}(s)$ is row reduced with row degrees m_i , such that the transfer matrix from \mathbf{r} to \mathbf{y} equals

$$\hat{\mathbf{G}}_o(s) = \mathbf{N}(s)\mathbf{F}^{-1}(s)\mathbf{B}(s)$$

Furthermore, the compensator can be obtained by solving sets of linear algebraic equations in (9.41).

⇒ **Proof:** Let $m = \max(m_1, m_2, \dots, m_p)$. Consider the constant matrix

$$\bar{\mathbf{F}} := [\mathbf{F}_0 \ \mathbf{F}_1 \ \mathbf{F}_2 \ \dots \ \mathbf{F}_{m+\mu}]$$

It is formed from the coefficient matrices of $\mathbf{F}(s)$ and has order $p \times (m + \mu + 1)$. Clearly, $\mathbf{F}(s)$ has column degrees at most $m + \mu_i$. Thus $\bar{\mathbf{F}}$ has at least α number of zero columns, where α is given in (9.44). Furthermore, the positions of these zero columns coincide with those of \mathbf{S}_m . Let $\tilde{\mathbf{F}}$ be the constant matrix $\bar{\mathbf{F}}$ after deleting these zero columns. Now consider

$$[\mathbf{A}_0 \ \mathbf{B}_0 \ \mathbf{A}_1 \ \mathbf{B}_1 \ \dots \ \mathbf{A}_m \ \mathbf{B}_m]\tilde{\mathbf{S}}_m = \tilde{\mathbf{F}} \quad (9.49)$$

It is obtained from (9.41) by deleting α number of zero columns in \mathbf{S}_m and the corresponding zero columns in $\tilde{\mathbf{F}}$. Now because $\tilde{\mathbf{S}}_m$ has full column rank if $m \geq v - 1$, we conclude that for any $\mathbf{F}(s)$ of column degrees at most $m + \mu_i$, solutions \mathbf{A}_i and \mathbf{B}_i exist in (9.49). Or, equivalently, polynomial matrices $\mathbf{A}(s)$ and $\mathbf{B}(s)$ of row degree m or less exist in (9.49). Note that generally $\tilde{\mathbf{S}}_m$ has more rows than columns; therefore solutions of (9.49) are not unique.

Next we show that $\mathbf{A}^{-1}(s)\mathbf{B}(s)$ is proper. Note that \mathbf{D}_μ is, in general, singular and the method of proving Theorem 9.2 cannot be used here. Using $\mathbf{H}_r(s)$ and $\mathbf{H}_c(s)$, we write, as in (7.80),

$$\begin{aligned}\mathbf{D}(s) &= [\mathbf{D}_{hc} + \mathbf{D}_{lc}(s)\mathbf{H}_c^{-1}(s)]\mathbf{H}_c(s) \\ \mathbf{N}(s) &= [\mathbf{N}_{hc} + \mathbf{N}_{lc}(s)\mathbf{H}_c^{-1}(s)]\mathbf{H}_c(s) \\ \mathbf{A}(s) &= \mathbf{H}_r(s)[\mathbf{A}_{hr} + \mathbf{H}_r^{-1}(s)\mathbf{A}_{lr}(s)] \\ \mathbf{B}(s) &= \mathbf{H}_r(s)[\mathbf{B}_{hr} + \mathbf{H}_r^{-1}(s)\mathbf{B}_{lr}(s)] \\ \mathbf{F}(s) &= \mathbf{H}_r(s)[\mathbf{F}_h + \mathbf{H}_r^{-1}(s)\mathbf{F}_l(s)\mathbf{H}_c^{-1}(s)]\mathbf{H}_c(s)\end{aligned}$$

where $\mathbf{D}_{lc}(s)\mathbf{H}_c^{-1}(s)$, $\mathbf{N}_{lc}(s)\mathbf{H}_c^{-1}(s)$, $\mathbf{H}_r^{-1}(s)\mathbf{A}_{lr}(s)$, $\mathbf{H}_r^{-1}(s)\mathbf{B}_{lr}(s)$, and $\mathbf{H}_r^{-1}(s)\mathbf{F}_l(s)\mathbf{H}_c^{-1}(s)$ are all strictly proper rational functions. Substituting the above into (9.39) yields, at $s = \infty$,

$$\mathbf{A}_{hr}\mathbf{D}_{hc} + \mathbf{B}_{hr}\mathbf{N}_{hc} = \mathbf{F}_h$$

which reduces to, because $\mathbf{N}_{hc} = \mathbf{0}$ following strict properness of $\hat{\mathbf{G}}(s)$,

$$\mathbf{A}_{hr}\mathbf{D}_{hc} = \mathbf{F}_h$$

Because $\mathbf{D}(s)$ is column reduced, \mathbf{D}_{hc} is nonsingular. The constant matrix \mathbf{F}_h is nonsingular by assumption, thus $\mathbf{A}_{hr} = \mathbf{F}_h\mathbf{D}_{hc}^{-1}$ is nonsingular and $\mathbf{A}(s)$ is row reduced. Therefore $\mathbf{A}^{-1}(s)\mathbf{B}(s)$ is proper (Corollary 7.8). This establishes the theorem. Q.E.D.

A polynomial matrix $\mathbf{F}(s)$ meeting (9.48) is said to be *row-column reduced* with row degrees m_i and column degrees μ_i . If $m_1 = m_2 = \dots = m_p = m$, then the row-column reducedness is the same as column reducedness with column degrees $m + \mu_i$. In application, we can select $\mathbf{F}(s)$ to be diagonal or triangular with polynomials with desired roots as its diagonal entries. Then $\mathbf{F}^{-1}(s)$ and, consequently, $\hat{\mathbf{G}}_o(s)$ have the desired roots as their poles.

Consider again \mathbf{S}_{v-1} . It is of order $(p+q)v \times (\mu+v)p$. It has $\alpha = p\mu - n$ number of zero columns. Thus the matrix $\tilde{\mathbf{S}}_{v-1}$ is of order $(p+q)v \times [(\mu+v)p - (p\mu - n)]$ or $(p+q)v \times (vp + n)$. The matrix $\tilde{\mathbf{S}}_{v-1}$ contains $p\nu$ linearly independent D -rows but contains only $v_1 + \dots + v_q = n$ linearly independent N -rows. Thus $\tilde{\mathbf{S}}_{v-1}$ contains

$$\gamma := (p+q)v - p\nu - n = qv - n$$

linearly dependent N -rows. Let $\check{\mathbf{S}}_{v-1}$ be the matrix $\tilde{\mathbf{S}}_{v-1}$ after deleting these linearly dependent N -rows. Then the matrix $\check{\mathbf{S}}_{v-1}$ is of order

$$[(p+q)v - (qv - n)] \times (vp + n) = (vp + n) \times (vp + n)$$

Thus $\check{\mathbf{S}}_{v-1}$ is square and nonsingular.

Consider (9.49) with $m = v - 1$:

$$\mathbf{K}\tilde{\mathbf{S}}_{v-1} := [\mathbf{A}_0 \ \mathbf{B}_0 \ \mathbf{A}_1 \ \mathbf{B}_1 \ \cdots \ \mathbf{A}_{v-1} \ \mathbf{B}_{v-1}] \tilde{\mathbf{S}}_{v-1} = \tilde{\mathbf{F}}$$

It actually consists of the following p sets of linear algebraic equations

$$\mathbf{k}_i \tilde{\mathbf{S}}_{v-1} = \tilde{\mathbf{f}}_i, \quad i = 1, 2, \dots, p \quad (9.50)$$

where \mathbf{k}_i and $\tilde{\mathbf{f}}_i$ are the i th row of \mathbf{K} and $\tilde{\mathbf{F}}$. Because $\tilde{\mathbf{S}}_{v-1}$ has full column rank, for any $\tilde{\mathbf{f}}_i$, solutions \mathbf{k}_i exist in (9.50). Because $\tilde{\mathbf{S}}_{v-1}$ has more γ rows than columns, the general solution of (9.50) contains γ free parameters (Corollary 3.2). It m in \mathbf{S}_m increases by 1 from $v - 1$ to v , then the number of rows of $\tilde{\mathbf{S}}_v$ increases by $(p + q)$ but the rank of $\tilde{\mathbf{S}}_v$ increases only by p . In this case, the number of free parameters will increase from γ to $\gamma + q$. Thus in the MIMO case, we have a great deal of freedom in carrying out the design.

We discuss a special case of (9.50). The matrix $\tilde{\mathbf{S}}_{v-1}$ has γ linearly dependent N -rows. If we delete these linearly dependent N -rows from $\tilde{\mathbf{S}}_{v-1}$ and assign the corresponding columns in \mathbf{B}_i as zero, then (9.50) becomes

$$[\mathbf{A}_0 \ \bar{\mathbf{B}}_0 \ \cdots \ \mathbf{A}_{v-1} \bar{\mathbf{B}}_{v-1}] \check{\mathbf{S}}_{v-1} = \tilde{\mathbf{F}}$$

where $\check{\mathbf{S}}_{v-1}$ is, as discussed earlier, square and nonsingular. Thus the solution is unique. This is illustrated in the next example.

Example 9.5.1 Consider a plant with the strictly proper transfer matrix

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} 1/s^2 & 1/s \\ 0 & 1/s \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} s^2 & 0 \\ 0 & s \end{bmatrix}^{-1} =: \mathbf{N}(s)\mathbf{D}^{-1}(s) \quad (9.51)$$

The fraction is right coprime and $\mathbf{D}(s)$ is column reduced with column degrees $\mu_1 = 2$ and $\mu_2 = 1$. We write

$$\mathbf{D}(s) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}s + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}s^2$$

and

$$\mathbf{N}(s) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}s + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}s^2$$

We use MATLAB to compute the row index. The qr-decomposition discussed in Section 7.3.1 can reveal linearly independent columns, from left to right, of a matrix. Here we need linearly independent rows, from top to bottom, of \mathbf{S}_m ; therefore, we will apply qr-decomposition to the transpose of \mathbf{S}_m . We type

```
d1=[0 0 0 0 1 0];d2=[0 0 0 1 0 0];
n1=[1 1 0 0 0 0];n2=[0 1 0 0 0 0];
s1=[d1 0 0;d2 0 0;n1 0 0;n2 0 0;...
     0 0 d1;0 0 d2;0 0 n1;0 0 n2];
[q,r]=qr(s1')
```

which yields, as in Example 7.8.1,

$$r = \begin{bmatrix} d1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & d2 & 0 & 0 & 0 & 0 & x & x \\ 0 & 0 & n1 & x & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & n2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & d1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & d2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & n1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The matrix q is not needed and is not typed. In the matrix r , we use x, di and ni to denote nonzero entries. The nonzero diagonal entries of r yield the linearly independent columns of \mathbf{S}'_1 or, equivalently, linearly independent rows of \mathbf{S}_1 . We see that there are two linearly independent $N1$ -rows and one linearly independent $N2$ -row. The degree of $\hat{\mathbf{G}}(s)$ is 3 and we have found three linearly independent N -rows. Therefore there is no need to search further and we have $v_1 = 2$ and $v_2 = 1$. Thus the row index is $v = 2$. We select $m_1 = m_2 = m = v - 1 = 1$. Thus for any column-reduced $\mathbf{F}(s)$ of column degrees $m + \mu_1 = 3$ and $m + \mu_2 = 2$, we can find a proper compensator such that the resulting unity feedback system has $\mathbf{F}(s)$ as its denominator matrix. Let us choose arbitrarily

$$\begin{aligned} \mathbf{F}(s) &= \begin{bmatrix} (s^2 + 4s + 5)(s + 3) & 0 \\ 0 & s^2 + 2s + 5 \end{bmatrix} \\ &= \begin{bmatrix} 15 + 17s + 7s^2 + s^3 & 0 \\ 0 & 5 + 2s + s^2 \end{bmatrix} \end{aligned} \quad (9.52)$$

and form (9.41) with $m = v - 1 = 1$:

$$\begin{aligned}
 & [\mathbf{A}_0 \ \mathbf{B}_0 \ \mathbf{A}_1 \ \mathbf{B}_1] \left[\begin{array}{ccccccc} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \dots & \dots \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \\
 & = \left[\begin{array}{ccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \dots & \dots \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{array} \right] = \bar{\mathbf{F}} \quad (9.53)
 \end{aligned}$$

The α in (9.44) is 1 for this problem. Thus $\tilde{\mathbf{S}}_1$ and $\bar{\mathbf{F}}$ both have one zero column as we can see in (9.53). After deleting the zero column, the remaining $\tilde{\mathbf{S}}_1$ has full column rank and for any $\tilde{\mathbf{F}}$, solutions exist in (9.53). The matrix $\tilde{\mathbf{S}}_1$ has order 8×7 and solutions are not unique. In searching the row index, we knew that the last row of $\tilde{\mathbf{S}}_1$ is a linearly dependent row. If we delete the row, then we must assign the second column of \mathbf{B}_1 as $\mathbf{0}$ and the solution will be unique. We type

```

d1=[0 0 0 0 1 0];d2=[0 0 0 1 0 0];
n1=[1 1 0 0 0 0];n2=[0 1 0 0 0 0];
d1t=[0 0 0 0 1];d2t=[0 0 0 1 0];n1t=[1 1 0 0 0];
s1t=[d1 0;d2 0;n1 0;n2 0;0 0 d1t;0 0 d2t;0 0 n1t];
f1t=[15 0 17 0 7 0 1];
f1t/s1t

```

which yields $[7 \ -17 \ 15 \ -15 \ 1 \ 0 \ 17]$. Computing once again for the second row of $\bar{\mathbf{F}}$, we can finally obtain

$$[\mathbf{A}_0 \ \mathbf{B}_0 \ \mathbf{A}_1 \ \mathbf{B}_1] = \left[\begin{array}{ccccccc} 7 & -17 & 15 & -15 & 1 & 0 & 17 & 0 \\ 0 & 2 & 0 & 5 & 0 & 1 & 0 & 0 \end{array} \right]$$

Note that MATLAB yields the first 7 columns; the last column $\mathbf{0}$ is assigned by us (due to deleting the last row of $\tilde{\mathbf{S}}_1$). Thus we have

$$\mathbf{A}(s) = \begin{bmatrix} 7+s & -17 \\ 0 & 2+s \end{bmatrix}, \quad \mathbf{B}(s) = \begin{bmatrix} 15+17s & -15 \\ 0 & 5 \end{bmatrix}$$

and the proper compensator

$$\mathbf{C}(s) = \begin{bmatrix} s+7 & -17 \\ 0 & s+2 \end{bmatrix}^{-1} \begin{bmatrix} 17s+15 & -15 \\ 0 & 5 \end{bmatrix}$$

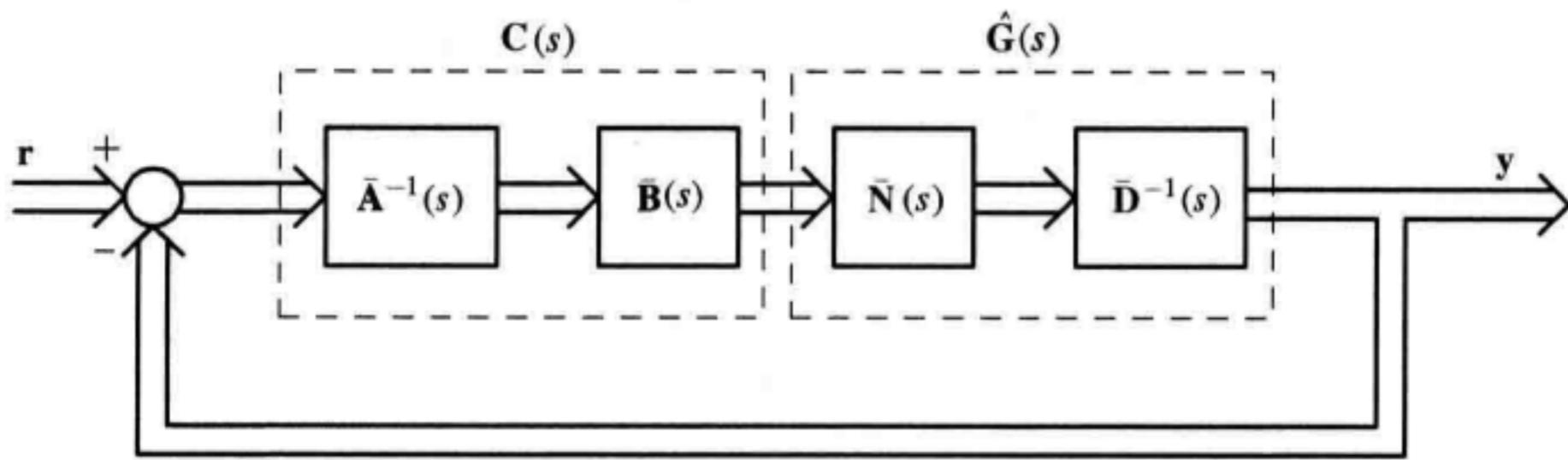


FIGURE 9.7 Unity feedback with $\hat{\mathbf{G}}(s) = \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)$.

will yield the overall transfer matrix

$$\begin{aligned}\hat{\mathbf{G}}_o(s) &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} (s^2 + 4s + 5)(s + 3) & 0 \\ 0 & s^2 + 2s + 5 \end{bmatrix}^{-1} \\ &\times \begin{bmatrix} 17s + 15 & -15 \\ 0 & 5 \end{bmatrix}\end{aligned}\quad (9.54)$$

This transfer matrix has the desired poles. This completes the design.

The design in Theorem 9.M2 is carried out using a right coprime fraction of $\hat{\mathbf{G}}(s)$. We state next the result using a left coprime fraction of $\hat{\mathbf{G}}(s)$.

► Corollary 9.M2

Consider the unity feedback system shown in Fig. 9.7. The plant is described by a $q \times p$ strictly proper rational matrix $\hat{\mathbf{G}}(s)$. Let $\hat{\mathbf{G}}(s)$ be factored as $\hat{\mathbf{G}}(s) = \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)$, where $\bar{\mathbf{D}}(s)$ and $\bar{\mathbf{N}}(s)$ are left coprime and $\bar{\mathbf{D}}(s)$ is row reduced with row degrees v_i , $i = 1, 2, \dots, q$. Let μ be the column index of $\hat{\mathbf{G}}(s)$ and let $m_i \geq \mu - 1$. Then for any $q \times q$ row-column reduced polynomial matrix $\bar{\mathbf{F}}(s)$ such that

$$\lim_{s \rightarrow \infty} \text{diag}(s^{-v_1}, s^{-v_2}, \dots, s^{-v_q})\bar{\mathbf{F}}(s)\text{diag}(s^{-m_1}, s^{-m_2}, \dots, s^{-m_q}) = \bar{\mathbf{F}}_h$$

is a nonsingular constant matrix, there exists a $p \times q$ proper compensator $\mathbf{C}(s) = \bar{\mathbf{B}}(s)\bar{\mathbf{A}}^{-1}(s)$, where $\bar{\mathbf{A}}(s)$ is column reduced with column degrees m_i , to meet

$$\bar{\mathbf{D}}(s)\bar{\mathbf{A}}(s) + \bar{\mathbf{N}}(s)\bar{\mathbf{B}}(s) = \bar{\mathbf{F}}(s) \quad (9.55)$$

and the transfer matrix from \mathbf{r} to \mathbf{y} equals

$$\hat{\mathbf{G}}_o(s) = \mathbf{I} - \bar{\mathbf{A}}(s)\bar{\mathbf{F}}^{-1}(s)\bar{\mathbf{D}}(s) \quad (9.56)$$

Substituting $\hat{\mathbf{G}}(s) = \bar{\mathbf{D}}^{-1}\bar{\mathbf{N}}(s)$ and $\mathbf{C}(s) = \bar{\mathbf{B}}(s)\bar{\mathbf{A}}^{-1}(s)$ into the first equation in (9.37) yields

$$\begin{aligned}\hat{\mathbf{G}}_o(s) &= [\mathbf{I} + \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)\bar{\mathbf{B}}(s)\bar{\mathbf{A}}^{-1}(s)]^{-1}\bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)\bar{\mathbf{B}}(s)\bar{\mathbf{A}}^{-1}(s) \\ &= \bar{\mathbf{A}}(s)[\bar{\mathbf{D}}(s)\bar{\mathbf{A}}(s) + \bar{\mathbf{N}}(s)\bar{\mathbf{B}}(s)]^{-1}\bar{\mathbf{N}}(s)\bar{\mathbf{B}}(s)\bar{\mathbf{A}}^{-1}(s)\end{aligned}$$

which becomes, after substituting (9.55),

$$\begin{aligned}\hat{\mathbf{G}}_o(s) &= \bar{\mathbf{A}}(s)\bar{\mathbf{F}}^{-1}(s)[\bar{\mathbf{F}}(s) - \bar{\mathbf{D}}(s)\bar{\mathbf{A}}(s)]\bar{\mathbf{A}}^{-1}(s) \\ &= \mathbf{I} - \bar{\mathbf{A}}(s)\bar{\mathbf{F}}^{-1}(s)\bar{\mathbf{D}}(s)\end{aligned}$$

This establishes the transfer matrix from \mathbf{r} to \mathbf{y} in the theorem. The design in Corollary 9.M2 hinges on solving (9.55). Note that the transpose of (9.55) becomes (9.39); left coprime and row reducedness become right coprime and column reducedness. Thus the linear algebraic equation in (9.41) can be used to solve the transpose of (9.55). We can also solve (9.55) directly by forming

$$\begin{aligned}\mathbf{T}_m \begin{bmatrix} \bar{\mathbf{B}}_0 \\ \bar{\mathbf{A}}_0 \\ \bar{\mathbf{B}}_1 \\ \bar{\mathbf{A}}_1 \\ \vdots \\ \bar{\mathbf{B}}_m \\ \bar{\mathbf{A}}_m \end{bmatrix} &= \begin{bmatrix} \bar{\mathbf{D}}_0 & \bar{\mathbf{N}}_0 & \vdots & \mathbf{0} & \mathbf{0} & \vdots & \dots & \vdots & \mathbf{0} & \mathbf{0} \\ \bar{\mathbf{D}}_1 & \bar{\mathbf{N}}_1 & \vdots & \bar{\mathbf{D}}_0 & \bar{\mathbf{N}}_0 & \vdots & \dots & \vdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots \\ \bar{\mathbf{D}}_n & \bar{\mathbf{N}}_n & \vdots & \bar{\mathbf{D}}_{n-1} & \bar{\mathbf{N}}_{n-1} & \vdots & \dots & \vdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \vdots & \bar{\mathbf{D}}_n & \bar{\mathbf{N}}_n & \vdots & \dots & \vdots & \bar{\mathbf{D}}_0 & \bar{\mathbf{N}}_0 \\ \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \vdots & \mathbf{0} & \mathbf{0} & \vdots & \dots & \vdots & \bar{\mathbf{D}}_n & \bar{\mathbf{N}}_n \end{bmatrix} \\ \times \begin{bmatrix} \bar{\mathbf{B}}_0 \\ \bar{\mathbf{A}}_0 \\ \bar{\mathbf{B}}_1 \\ \bar{\mathbf{A}}_1 \\ \vdots \\ \bar{\mathbf{B}}_m \\ \bar{\mathbf{A}}_m \end{bmatrix} &= \begin{bmatrix} \bar{\mathbf{F}}_0 \\ \bar{\mathbf{F}}_1 \\ \bar{\mathbf{F}}_2 \\ \vdots \\ \bar{\mathbf{F}}_{n+m} \end{bmatrix} \quad (9.57)\end{aligned}$$

We search linearly independent columns of \mathbf{T}_m in order from left to right. Let μ be the column index of $\hat{\mathbf{G}}(s)$ or the least integer such that $\mathbf{T}_{\mu-1}$ contains n linearly independent \bar{N} -columns. Then the compensator can be solved from (9.57) with $m = \mu - 1$.

9.5.1 Regulation and Tracking

As in the SISO case, pole placement can be used to achieve regulation and tracking in MIMO systems. In the regulator problem, we have $\mathbf{r} \equiv \mathbf{0}$ and if all poles of the overall system are assigned to have negative real parts, then the responses caused by any nonzero initial state will vanish as $t \rightarrow \infty$. Furthermore, the rate of vanishing can be controlled by the locations of the poles; the larger the negative real parts, the faster the vanishing.

Next we discuss tracking of any step reference input. In this design, we generally require a feedforward constant gain matrix \mathbf{P} . Suppose the compensator in Fig. 9.6 has been designed using Theorem 9.M2. Then the $q \times q$ transfer matrix from \mathbf{r} to \mathbf{y} is

given by

$$\hat{\mathbf{G}}_o(s) = \mathbf{N}(s)\mathbf{F}^{-1}(s)\mathbf{B}(s)\mathbf{P} \quad (9.58)$$

If $\hat{\mathbf{G}}_o(s)$ is BIBO stable, then the steady-state response excited by $\mathbf{r}(t) = \mathbf{d}$, for $t \geq 0$, or $\hat{\mathbf{r}}(s) = \mathbf{d}s^{-1}$, where \mathbf{d} is an arbitrary $q \times 1$ constant vector, can be computed as, using the final-value theorem of the Laplace transform,

$$\lim_{t \rightarrow \infty} \mathbf{y}(t) = \lim_{s \rightarrow 0} s\hat{\mathbf{G}}_o(s)\mathbf{d}s^{-1} = \hat{\mathbf{G}}_o(0)\mathbf{d}$$

Thus we conclude that in order for $\mathbf{y}(t)$ to track asymptotically any step reference input, we need, in addition to BIBO stability,

$$\hat{\mathbf{G}}_o(0) = \mathbf{N}(0)\mathbf{F}^{-1}(0)\mathbf{B}(0)\mathbf{P} = \mathbf{I}_q \quad (9.59)$$

Before discussing the conditions for meeting (9.59), we need the concept of transmission zeros.

Transmission zeros

Consider $q \times p \hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)$, where $\mathbf{N}(s)$ and $\mathbf{D}(s)$ are right coprime. A number λ , real or complex, is called a *transmission zero* of $\hat{\mathbf{G}}(s)$ if the rank of $\mathbf{N}(\lambda)$ is smaller than $\min(p, q)$.

Example 9.5.2 Consider

$$\hat{\mathbf{G}}_1(s) = \begin{bmatrix} \frac{s}{s+2} & 0 \\ 0 & \frac{s+1}{s^2} \\ \frac{s+1}{s+2} & \frac{1}{s} \end{bmatrix} = \begin{bmatrix} s & 0 \\ 0 & s+1 \\ s+1 & s \end{bmatrix} \begin{bmatrix} s+2 & 0 \\ 0 & s^2 \end{bmatrix}^{-1}$$

This $\mathbf{N}(s)$ has rank 2 at every s , thus $\hat{\mathbf{G}}_1(s)$ has no transmission zero. Consider

$$\hat{\mathbf{G}}_2(s) = \begin{bmatrix} \frac{s}{s+2} & 0 \\ 0 & \frac{s+2}{s} \end{bmatrix} = \begin{bmatrix} s & 0 \\ 0 & s+2 \end{bmatrix} \begin{bmatrix} s+2 & 0 \\ 0 & s \end{bmatrix}^{-1}$$

This $\mathbf{N}(s)$ has rank 1 at $s = 0$ and $s = -2$. Thus $\hat{\mathbf{G}}_2(s)$ has two transmission zeros at 0 and -2 . Note that $\hat{\mathbf{G}}_2(s)$ has poles also at 0 and -2 .

From this example, we see that transmission zeros cannot be defined directly from $\hat{\mathbf{G}}(s)$; it must be defined from its coprime fraction. Either a right coprime or left coprime fraction can be used and they yield the same set of transmission zeros. Note that if $\hat{\mathbf{G}}(s)$ is square and if $\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s) = \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)$, where $\mathbf{N}(s)$ and $\mathbf{D}(s)$ are right coprime and $\bar{\mathbf{D}}(s)$ and $\bar{\mathbf{N}}(s)$ are left coprime, then the transmission zeros of $\hat{\mathbf{G}}(s)$ are the roots of $\det \mathbf{N}(s)$ or the roots of $\det \bar{\mathbf{N}}(s)$. Transmission zeros can also be defined from a minimal realization of $\hat{\mathbf{G}}(s)$. Let $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ be any n -dimensional minimal realization of a $q \times p$ proper rational matrix $\hat{\mathbf{G}}(s)$. Then the transmission zeros

are those λ such that

$$\text{rank} \begin{bmatrix} \lambda \mathbf{I} - \mathbf{A} & \mathbf{B} \\ -\mathbf{C} & \mathbf{D} \end{bmatrix} < n + \min(p, q)$$

This is used in the MATLAB function `tzero` to compute transmission zeros. For a more detailed discussion of transmission zeros, see Reference 6, pp. 623–635.

Now we are ready to discuss the conditions for achieving tracking or for meeting (9.59). Note that $\mathbf{N}(s)$, $\mathbf{F}(s)$ and $\mathbf{B}(s)$ are $q \times p$, $p \times p$ and $p \times q$. Because \mathbf{I}_q has rank q , a necessary condition for (9.59) to hold is that the $q \times p$ matrix $\mathbf{N}(0)$ has rank q . Necessary conditions for $\rho(\mathbf{N}(0)) = q$ are $p \geq q$ and $\hat{\mathbf{G}}(s)$ having no transmission zero at $s = 0$. Thus we conclude that in order for the unity-feedback configuration in Fig. 9.6 to achieve asymptotic tracking, the plant must have the following two properties:

- The plant has the same or more number of inputs than outputs.
- The plant transfer function has no transmission zero at $s = 0$.

Under these conditions, $\mathbf{N}(0)$ has rank q . Because we have freedom in selecting $\mathbf{F}(s)$, we can select it such that $\mathbf{B}(0)$ has rank q and the $q \times q$ constant matrix $\mathbf{N}(0)\mathbf{F}^{-1}(0)\mathbf{B}(0)$ is nonsingular. Under these conditions, the constant gain matrix \mathbf{P} can be computed as

$$\mathbf{P} = [\mathbf{N}(0)\mathbf{F}^{-1}(0)\mathbf{B}(0)]^{-1} \quad (9.60)$$

Then we have $\hat{\mathbf{G}}_o(0) = \mathbf{I}_q$, and the unity feedback system in Fig. 9.6 with \mathbf{P} in (9.60) will track asymptotically any step reference input.

9.5.2 Robust Tracking and Disturbance Rejection

As in the SISO case, the asymptotic tracking design in the preceding section is not robust. In this section, we discuss a different design. To simplify the discussion, we study only plants with equal number of input terminals and output terminals or $p = q$. Consider the unity feedback system shown in Fig. 9.8. The plant is described by a $p \times p$ strictly proper transfer matrix factored in left coprime fraction as $\hat{\mathbf{G}}(s) = \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)$. It is assumed that $p \times 1$ disturbance $\mathbf{w}(t)$ enters the plant input as shown. The problem is to design a compensator so that the output $\mathbf{y}(t)$ will track asymptotically a class of $p \times 1$ reference signal $\mathbf{r}(t)$ even with the presence of the disturbance $\mathbf{w}(t)$ and with plant parameter variations. This is called *robust tracking and disturbance rejection*.

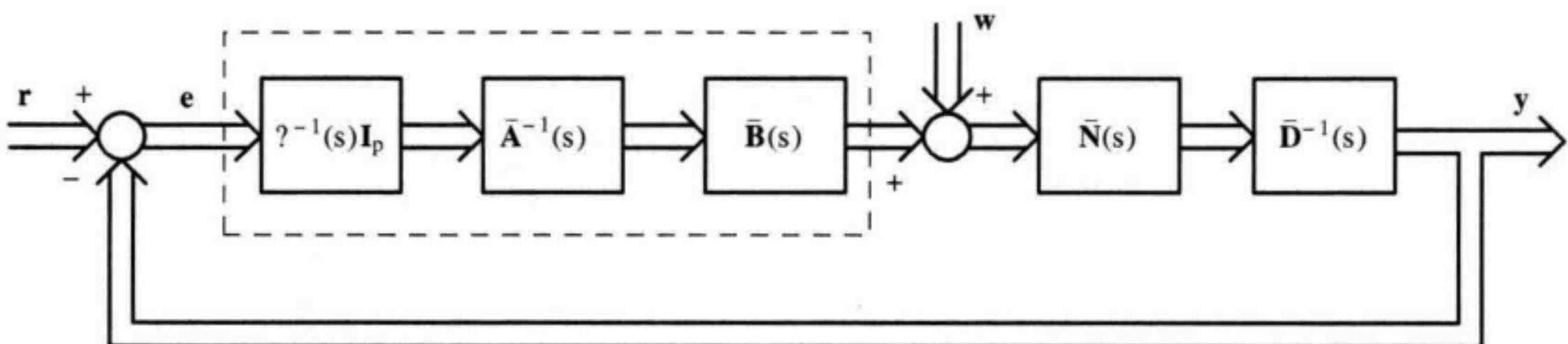


FIGURE 9.8 Robust tracking and disturbance rejection.

As in the SISO case, we need some information of $\mathbf{r}(t)$ and $\mathbf{w}(t)$ before carrying out the design. We assume that the Laplace transforms of $\mathbf{r}(t)$ and $\mathbf{w}(t)$ are given by

$$\hat{\mathbf{r}}(s) = \mathcal{L}[\mathbf{r}(t)] = \mathbf{N}_r(s)D_r^{-1}(s) \quad \hat{\mathbf{w}}(s) = \mathcal{L}[\mathbf{w}(t)] = \mathbf{N}_w(s)D_w^{-1}(s) \quad (9.61)$$

where $D_r(s)$ and $D_w(s)$ are known polynomials; however, $\mathbf{N}_r(s)$ and $\mathbf{N}_w(s)$ are unknown to us. Let $\phi(s)$ be the least common denominator of the unstable poles of $\hat{\mathbf{r}}(s)$ and $\hat{\mathbf{w}}(s)$. The stable poles are excluded because they have no effect on $\mathbf{y}(t)$ as $t \rightarrow \infty$. We introduce the internal model $\phi^{-1}(s)\mathbf{I}_p$ as shown as shown in Fig. 9.8. If $\bar{\mathbf{D}}(s)$ and $\bar{\mathbf{N}}(s)$ are left coprime and if no root of $\phi(s)$ is a transmission zero of $\hat{\mathbf{G}}(s)$ or, equivalently, $\phi(s)$ and $\det \bar{\mathbf{N}}(s)$ are coprime, then it can be shown that $\bar{\mathbf{D}}(s)\phi(s)$ and $\bar{\mathbf{N}}(s)$ are left coprime. See Reference 6, p. 443. Then Corollary 9.M2 implies that there exists a proper compensator $\mathbf{C}(s) = \bar{\mathbf{B}}(s)\bar{\mathbf{A}}^{-1}(s)$ such that

$$\phi(s)\bar{\mathbf{D}}(s)\bar{\mathbf{A}}(s) + \bar{\mathbf{N}}(s)\bar{\mathbf{B}}(s) = \bar{\mathbf{F}}(s) \quad (9.62)$$

for any $\bar{\mathbf{F}}(s)$ meeting the condition in Corollary 9.M2. Clearly $\bar{\mathbf{F}}(s)$ can be chosen to be diagonal with roots of its diagonal entries lying inside the sector shown in Fig. 8.3. The unity feedback system in Fig. 9.8 so designed will track asymptotically and robustly the reference signal $\mathbf{r}(t)$ and reject the disturbance $\mathbf{w}(t)$. This is stated as a theorem.

► Theorem 9.M3

Consider the unity-feedback system shown in Fig. 9.8 where the plant has a $p \times p$ strictly proper transfer matrix $\hat{\mathbf{G}}(s) = \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)$. It is assumed that $\bar{\mathbf{D}}(s)$ and $\bar{\mathbf{N}}(s)$ are left coprime and $\bar{\mathbf{D}}(s)$ is row reduced with row degrees v_i , $i = 1, 2, \dots, p$. The reference signal $\mathbf{r}(t)$ and disturbance $\mathbf{w}(t)$ are modeled as $\hat{\mathbf{r}}(s) = \mathbf{N}_r(s)D_r^{-1}(s)$ and $\hat{\mathbf{w}}(s) = \mathbf{N}_w(s)D_w^{-1}(s)$. Let $\phi(s)$ be the least common denominator of the unstable poles of $\hat{\mathbf{r}}(s)$ and $\hat{\mathbf{w}}(s)$. If no root of $\phi(s)$ is a transmission zero of $\hat{\mathbf{G}}(s)$, then there exists a proper compensator $\mathbf{C}(s) = \bar{\mathbf{B}}(s)(\bar{\mathbf{A}}(s)\phi(s))^{-1}$ such that the overall system will robustly and asymptotically track the reference signal $\mathbf{r}(t)$ and reject the disturbance $\mathbf{w}(t)$.



Proof: First we show that the system will reject the disturbance at the output. Let us assume $\mathbf{r} = 0$ and compute the output $\hat{\mathbf{y}}_w(s)$ excited by $\hat{\mathbf{w}}(s)$. Clearly we have

$$\hat{\mathbf{y}}_w(s) = \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)[\hat{\mathbf{w}}(s) - \bar{\mathbf{B}}(s)\bar{\mathbf{A}}^{-1}(s)\phi^{-1}(s)\hat{\mathbf{y}}_w(s)]$$

which implies

$$\begin{aligned} \hat{\mathbf{y}}_w(s) &= [\mathbf{I} + \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)\bar{\mathbf{B}}(s)\bar{\mathbf{A}}^{-1}(s)\phi^{-1}(s)]^{-1}\bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)\hat{\mathbf{w}}(s) \\ &= [\bar{\mathbf{D}}^{-1}(s)[\bar{\mathbf{D}}(s)\phi(s)\bar{\mathbf{A}}(s) + \bar{\mathbf{N}}(s)\bar{\mathbf{B}}(s)]\bar{\mathbf{A}}^{-1}(s)\phi^{-1}(s)]^{-1} \\ &\quad \times \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)\hat{\mathbf{w}}(s) \\ &= \phi(s)\bar{\mathbf{A}}(s)[\bar{\mathbf{D}}(s)\phi(s)\bar{\mathbf{A}}(s) + \bar{\mathbf{N}}(s)\bar{\mathbf{B}}(s)]^{-1}\bar{\mathbf{N}}(s)\hat{\mathbf{w}}(s) \end{aligned}$$

Thus we have, using (9.61) and (9.62),

$$\hat{\mathbf{y}}_w(s) = \bar{\mathbf{A}}(s)\bar{\mathbf{F}}^{-1}(s)\bar{\mathbf{N}}(s)\phi(s)\mathbf{N}_w(s)D_w^{-1}(s) \quad (9.63)$$

Because all unstable roots of $D_w(s)$ are canceled by $\phi(s)$, all poles of $\hat{\mathbf{y}}_w(s)$ have negative real parts. Thus we have $\mathbf{y}_w(t) \rightarrow 0$ as $t \rightarrow \infty$ and the response excited by $\mathbf{w}(t)$ is suppressed asymptotically at the output.

Next we compute the error $\hat{\mathbf{e}}_r(s)$ excited by the reference signal $\hat{\mathbf{r}}(s)$:

$$\hat{\mathbf{e}}_r(s) = \hat{\mathbf{r}}(s) - \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)\bar{\mathbf{B}}(s)\bar{\mathbf{A}}^{-1}(s)\phi^{-1}(s)\hat{\mathbf{e}}_r(s)$$

which implies

$$\begin{aligned}\hat{\mathbf{e}}_r(s) &= [\mathbf{I} + \bar{\mathbf{D}}^{-1}(s)\bar{\mathbf{N}}(s)\bar{\mathbf{B}}(s)\bar{\mathbf{A}}^{-1}(s)\phi^{-1}(s)]^{-1}\hat{\mathbf{r}}(s) \\ &= \phi(s)\bar{\mathbf{A}}(s)[\bar{\mathbf{D}}(s)\phi(s)\bar{\mathbf{A}}(s) + \bar{\mathbf{N}}(s)\bar{\mathbf{B}}(s)]^{-1}\bar{\mathbf{D}}(s)\hat{\mathbf{r}}(s) \\ &= \bar{\mathbf{A}}(s)\bar{\mathbf{F}}^{-1}(s)\bar{\mathbf{D}}(s)\phi(s)\mathbf{N}_r(s)D_r^{-1}(s)\end{aligned}\quad (9.64)$$

Because all unstable roots of $D_r(s)$ are canceled by $\phi(s)$, the error vector $\hat{\mathbf{e}}_r(s)$ has only stable poles. Thus its time response approaches zero as $t \rightarrow \infty$. Consequently, the output $\mathbf{y}(t)$ will track asymptotically the reference signal $\mathbf{r}(t)$. The tracking and disturbance rejection is accomplished by inserting the internal model $\phi^{-1}(s)\mathbf{I}_p$. If there is no perturbation in the internal model, the tracking property holds for any plant and compensator parameter perturbations, even large ones, so long as the unity feedback system remains BIBO stable. Thus the design is robust. This establishes the theorem. Q.E.D.

In the robust design, because of the internal model, $\phi(s)$ becomes zeros of every nonzero entry of the transfer matrices from \mathbf{w} to \mathbf{y} and from \mathbf{r} to \mathbf{e} . Such zeros are called *blocking zeros*. These blocking zeros cancel all unstable poles of $\hat{\mathbf{w}}(s)$ and $\hat{\mathbf{r}}(s)$; thus, the responses due to these unstable poles are completely blocked at the output. It is clear that every blocking zero is a transmission zero. The converse however is not true. To conclude this section, we mention that if we use a right coprime fraction for $\hat{\mathbf{G}}(s)$, insert an internal model and stabilize it, this writer is able to show only disturbance rejection. Because of the noncommutative property of matrices, this writer is not able to establish robust tracking. However, it is believed that the system still achieves robust tracking. The design discussed in Section 9.3.3 can also be extended to the MIMO case; the design, however, will be more complex and will not be discussed.

9.6 MIMO Model Matching—Two-Parameter Configuration

In this section, we extend the SISO model matching to the MIMO case. We study only plants with square and nonsingular strictly proper transfer matrices. As in the SISO case, given a plant transfer matrix $\hat{\mathbf{G}}(s)$, a model $\hat{\mathbf{G}}_o(s)$ is said to be *implementable* if there exists a no-plant-leakage configuration and proper compensators so that the resulting system has the overall transfer matrix $\hat{\mathbf{G}}_o(s)$ and is totally stable and well-posed. The next theorem extends Theorem 9.4 to the matrix case.

► **Theorem 9.M4**

Consider a plant with $p \times p$ strictly proper transfer matrix $\hat{\mathbf{G}}(s)$. Then a $p \times p$ transfer matrix $\hat{\mathbf{G}}_o(s)$ is implementable if and only if $\hat{\mathbf{G}}_o(s)$ and

$$\hat{\mathbf{T}}(s) := \hat{\mathbf{G}}^{-1}(s)\hat{\mathbf{G}}_o(s) \quad (9.65)$$

are proper and BIBO stable.⁵

For any no-plant-leakage configuration, the closed-loop transfer matrix from \mathbf{r} to \mathbf{u} is $\hat{\mathbf{T}}(s)$. Thus well-posedness and total stability require $\hat{\mathbf{G}}_o(s)$ and $\hat{\mathbf{T}}(s)$ to be proper and BIBO stable. This shows the necessity of Theorem 9.M4. Let us write (9.65) as

$$\hat{\mathbf{G}}_o(s) = \hat{\mathbf{G}}(s)\hat{\mathbf{T}}(s) \quad (9.66)$$

Then (9.66) can be implemented in the open-loop configuration in Fig. 9.1(a) with $\mathbf{C}(s) = \hat{\mathbf{T}}(s)$. This design, however, is not acceptable either because it is not totally stable or because it is very sensitive to plant parameter variations. If we implement it in the unity feedback configuration, we have no freedom in assigning canceled poles. Thus the configuration may not be acceptable. In the unity feedback configuration, we have

$$\hat{\mathbf{u}}(s) = \mathbf{C}(s)[\hat{\mathbf{r}}(s) - \hat{\mathbf{y}}(s)]$$

Now we extend it to

$$\hat{\mathbf{u}}(s) = \mathbf{C}_1(s)\hat{\mathbf{r}}(s) - \mathbf{C}_2(s)\hat{\mathbf{y}}(s) \quad (9.67)$$

This is a two-degree-of-freedom configuration. As in the SISO case, we may select $\mathbf{C}_1(s)$ and $\mathbf{C}_2(s)$ to have the same denominator matrix as

$$\mathbf{C}_1(s) = \mathbf{A}^{-1}(s)\mathbf{L}(s), \quad \mathbf{C}_2(s) = \mathbf{A}^{-1}(s)\mathbf{M}(s) \quad (9.68)$$

Then the two-parameter configuration can be plotted as shown in Fig. 9.9. From the

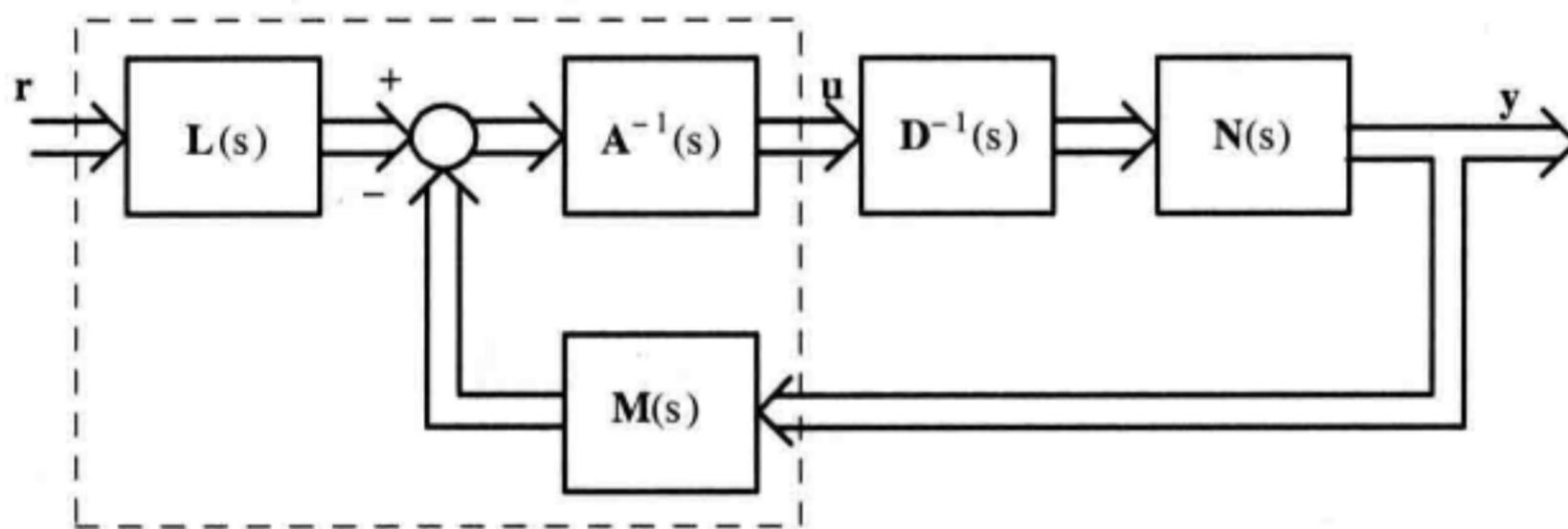


FIGURE 9.9 Two-parameter configuration.

⁵If $\hat{\mathbf{G}}(s)$ is not square, then $\hat{\mathbf{G}}_o(s)$ is implementable if and only if $\hat{\mathbf{G}}_o(s)$ is proper, is BIBO stable, and can be expressed as $\hat{\mathbf{G}}_o(s) = \hat{\mathbf{G}}(s)\hat{\mathbf{T}}(s)$, where $\hat{\mathbf{T}}(s)$ is proper and BIBO stable. See Reference 6, pp. 517–523.

figure, we have

$$\hat{\mathbf{u}}(s) = \mathbf{A}^{-1}(s)[\mathbf{L}(s)\hat{\mathbf{r}}(s) - \mathbf{M}(s)\mathbf{N}(s)\mathbf{D}^{-1}(s)\hat{\mathbf{u}}(s)]$$

which implies

$$\begin{aligned}\hat{\mathbf{u}}(s) &= [\mathbf{I} + \mathbf{A}^{-1}(s)\mathbf{M}(s)\mathbf{N}(s)\mathbf{D}^{-1}(s)]^{-1}\mathbf{A}^{-1}(s)\mathbf{L}(s)\hat{\mathbf{r}}(s) \\ &= \mathbf{D}(s)[\mathbf{A}(s)\mathbf{D}(s) + \mathbf{M}(s)\mathbf{N}(s)]^{-1}\mathbf{L}(s)\hat{\mathbf{r}}(s)\end{aligned}$$

Thus we have

$$\hat{\mathbf{y}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)\hat{\mathbf{u}}(s) = \mathbf{N}(s)[\mathbf{A}(s)\mathbf{D}(s) + \mathbf{M}(s)\mathbf{N}(s)]^{-1}\mathbf{L}(s)\hat{\mathbf{r}}(s)$$

and the transfer matrix from \mathbf{r} to \mathbf{y} is

$$\hat{\mathbf{G}}_o(s) = \mathbf{N}(s)[\mathbf{A}(s)\mathbf{D}(s) + \mathbf{M}(s)\mathbf{N}(s)]^{-1}\mathbf{L}(s) \quad (9.69)$$

Thus model matching becomes solving $\mathbf{A}(s)$, $\mathbf{M}(s)$ and $\mathbf{L}(s)$ from (9.69).

► Problem

Given $p \times p$ strictly proper $\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)$, where $\mathbf{N}(s)$ and $\mathbf{D}(s)$ are right coprime and $\mathbf{D}(s)$ is column reduced with column degrees μ_i , $i = 1, 2, \dots, p$, and given any implementable $\hat{\mathbf{G}}_o(s)$, find proper compensators $\mathbf{A}^{-1}(s)\mathbf{L}(s)$ and $\mathbf{A}^{-1}(s)\mathbf{M}(s)$ in Fig. 9.9 to implement $\hat{\mathbf{G}}_o(s)$.

► Procedure 9.M1

1. Compute

$$\mathbf{N}^{-1}(s)\hat{\mathbf{G}}_o(s) = \bar{\mathbf{F}}^{-1}(s)\bar{\mathbf{E}}(s) \quad (9.70)$$

where $\bar{\mathbf{F}}(s)$ and $\bar{\mathbf{E}}(s)$ are left coprime and $\bar{\mathbf{F}}(s)$ is row reduced.

2. Compute the row index v of $\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)$. This can be achieved by using *qr*-decomposition.

3. Select

$$\hat{\mathbf{F}}(s) = \text{diag}(\alpha_1(s), \alpha_2(s), \dots, \alpha_p(s)) \quad (9.71)$$

where $\alpha_i(s)$ are arbitrary CT stable polynomials, such that $\hat{\mathbf{F}}(s)\bar{\mathbf{F}}(s)$ is row-column reduced with column degrees μ_i and row degrees m_i with

$$m_i \geq v - 1 \quad (9.72)$$

for $i = 1, 2, \dots, p$.

4. Set

$$\mathbf{L}(s) = \hat{\mathbf{F}}(s)\bar{\mathbf{E}}(s) \quad (9.73)$$

and solve $\mathbf{A}(s)$ and $\mathbf{M}(s)$ from

$$\mathbf{A}(s)\mathbf{D}(s) + \mathbf{M}(s)\mathbf{N}(s) = \hat{\mathbf{F}}(s)\bar{\mathbf{F}}(s) =: \mathbf{F}(s) \quad (9.74)$$

Then proper compensators $\mathbf{A}^{-1}(s)\mathbf{L}(s)$ and $\mathbf{A}^{-1}(s)\mathbf{M}(s)$ can be obtained to achieve the model matching.

This procedure reduces to Procedure 9.1 if $\hat{\mathbf{G}}(s)$ is scalar. We first justify the procedure. Substituting (9.73) and (9.74) into (9.69) yields

$$\hat{\mathbf{G}}_o(s) = \mathbf{N}(s)[\hat{\mathbf{F}}(s)\bar{\mathbf{F}}(s)]^{-1}\hat{\mathbf{F}}(s)\bar{\mathbf{E}}(s) = \mathbf{N}(s)\bar{\mathbf{F}}^{-1}(s)\bar{\mathbf{E}}(s)$$

This is (9.70). Thus the compensators implement $\hat{\mathbf{G}}_o(s)$. Define

$$\mathbf{H}_c(s) = \text{diag}(s^{\mu_1}, s^{\mu_2}, \dots, s^{\mu_p}) \quad \mathbf{H}_r(s) = \text{diag}(s^{m_1}, s^{m_2}, \dots, s^{m_p})$$

By assumption, the matrix

$$\lim_{s \rightarrow \infty} \mathbf{H}_r^{-1}(s)\mathbf{F}(s)\mathbf{H}_c^{-1}(s) =: \mathbf{F}_h$$

is a nonsingular constant matrix. Thus, solutions $\mathbf{A}(s)$, with row degrees m_i and row reduced, and $\mathbf{M}(s)$, with row degrees m_i or less, exist in (9.74) (Theorem 9.M2). Thus $\mathbf{A}^{-1}(s)\mathbf{M}(s)$ is proper. To show that $\mathbf{A}^{-1}(s)\mathbf{L}(s)$ is proper, we consider

$$\begin{aligned} \hat{\mathbf{T}}(s) &= \hat{\mathbf{G}}^{-1}(s)\hat{\mathbf{G}}_o(s) = \mathbf{D}(s)\mathbf{N}^{-1}(s)\hat{\mathbf{G}}_o(s) \\ &= \mathbf{D}(s)\bar{\mathbf{F}}^{-1}(s)\bar{\mathbf{E}}(s) = \mathbf{D}(s)[\hat{\mathbf{F}}(s)\bar{\mathbf{F}}(s)]^{-1}\hat{\mathbf{F}}(s)\bar{\mathbf{E}}(s) \\ &= \mathbf{D}(s)[\mathbf{A}(s)\mathbf{D}(s) + \mathbf{M}(s)\mathbf{N}(s)]^{-1}\mathbf{L}(s) \\ &= \mathbf{D}(s)\left\{\mathbf{A}(s)[\mathbf{I} + \mathbf{A}^{-1}(s)\mathbf{M}(s)\mathbf{N}(s)\mathbf{D}^{-1}(s)]\mathbf{D}(s)\right\}^{-1}\mathbf{L}(s) \\ &= [\mathbf{I} + \mathbf{A}^{-1}(s)\mathbf{M}(s)\hat{\mathbf{G}}(s)]^{-1}\mathbf{A}^{-1}(s)\mathbf{L}(s) \end{aligned}$$

which implies, because $\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)$ is strictly proper and $\mathbf{A}^{-1}(s)\mathbf{M}(s)$ is proper,

$$\lim_{s \rightarrow \infty} \hat{\mathbf{T}}(s) = \lim_{s \rightarrow \infty} \mathbf{A}^{-1}(s)\mathbf{L}(s)$$

Because $\mathbf{T}(\infty)$ is finite by assumption, we conclude that $\mathbf{A}^{-1}(s)\mathbf{L}(s)$ is proper. If $\hat{\mathbf{G}}(s)$ is strictly proper and if all compensators are proper, then the two-parameter configuration is automatically well posed. The design involves the pole-zero cancelation of $\hat{\mathbf{F}}(s)$ which we can select. If $\hat{\mathbf{F}}(s)$ is selected as diagonal with CT stable polynomials as its entries, then pole-zero cancelations involve only stable poles and the system is totally stable. This completes the justification of the design procedure.

Example 9.6.1 Consider a plant with transfer matrix

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} 1/s^2 & 1/s \\ 0 & 1/s \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} s^2 & 0 \\ 0 & s \end{bmatrix}^{-1} \quad (9.75)$$

It has column degrees $\mu_1 = 2$ and $\mu_2 = 1$. Let us select a model as

$$\hat{\mathbf{G}}_o(s) = \begin{bmatrix} \frac{2}{s^2+2s+2} & 0 \\ 0 & \frac{2}{s^2+2s+2} \end{bmatrix} \quad (9.76)$$

which is proper and BIBO stable. To check whether or not $\hat{\mathbf{G}}_o(s)$ is implementable, we compute

$$\hat{\mathbf{T}}(s) := \hat{\mathbf{G}}^{-1}(s)\hat{\mathbf{G}}_o(s) = \begin{bmatrix} s^2 & -s^2 \\ 0 & s \end{bmatrix} \hat{\mathbf{G}}_o(s) = \begin{bmatrix} \frac{2s^2}{s^2+2s+2} & \frac{-2s^2}{s^2+2s+2} \\ 0 & \frac{2s}{s^2+2s+2} \end{bmatrix}$$

which is proper and BIBO stable. Thus $\hat{\mathbf{G}}_o(s)$ is implementable. We compute

$$\begin{aligned} \mathbf{N}^{-1}(s)\hat{\mathbf{G}}_o(s) &= \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{2}{s^2+2s+2} & 0 \\ 0 & \frac{2}{s^2+2s+2} \end{bmatrix} \\ &= \begin{bmatrix} \frac{2}{s^2+2s+2} & \frac{-2}{s^2+2s+2} \\ 0 & \frac{2}{s^2+2s+2} \end{bmatrix} \\ &= \begin{bmatrix} s^2 + 2s + 2 & 0 \\ 0 & s^2 + 2s + 2 \end{bmatrix}^{-1} \begin{bmatrix} 2 & -2 \\ 0 & 2 \end{bmatrix} \\ &=: \bar{\mathbf{F}}^{-1}(s)\bar{\mathbf{E}}(s) \end{aligned}$$

For this example, the degree of $\mathbf{N}^{-1}(s)\hat{\mathbf{G}}_o(s)$ can be readily computed as 4. The determinant of $\bar{\mathbf{F}}(s)$ has degree 4; thus the pair $\bar{\mathbf{F}}(s)$ and $\bar{\mathbf{E}}(s)$ are left coprime. It is clear that $\bar{\mathbf{F}}(s)$ is row reduced with row degrees $r_1 = r_2 = 2$. The row index of $\hat{\mathbf{G}}(s)$ was computed in Example 9.5.1 as $v = 2$. Let us select

$$\hat{\mathbf{F}}(s) = \text{diag}((s+2), 1)$$

Then we have

$$\begin{aligned} \hat{\mathbf{F}}(s)\bar{\mathbf{F}}(s) &= \begin{bmatrix} (s^2 + 2s + 2)(s+2) & 0 \\ 0 & s^2 + 2s + 2 \end{bmatrix} \\ &= \begin{bmatrix} 4 + 6s + 4s^2 + s^3 & 0 \\ 0 & 2 + 2s + s^2 \end{bmatrix} \quad (9.77) \end{aligned}$$

It is row-column reduced with row degrees $\{m_1 = m_2 = 1 = v - 1\}$ and column degrees $\{\mu_1 = 2, \mu_2 = 1\}$. Note that without introducing $\hat{\mathbf{F}}(s)$, proper compensators may not exist. We set

$$\begin{aligned} \mathbf{L}(s) = \hat{\mathbf{F}}(s)\bar{\mathbf{E}}(s) &= \begin{bmatrix} s+2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & -2 \\ 0 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 2(s+2) & -2(s+2) \\ 0 & 2 \end{bmatrix} \quad (9.78) \end{aligned}$$

and solve $\mathbf{A}(s)$ and $\mathbf{M}(s)$ from

$$\mathbf{A}(s)\mathbf{D}(s) + \mathbf{M}(s)\mathbf{N}(s) = \hat{\mathbf{F}}(s)\bar{\mathbf{F}}(s) =: \mathbf{F}(s) \quad (9.79)$$

From the coefficient matrices of $\mathbf{D}(s)$ and $\mathbf{N}(s)$ and the coefficient matrices of (9.77), we can readily form

$$[\mathbf{A}_0 \ \mathbf{M}_0 \ \mathbf{A}_1 \ \mathbf{M}_1] \left[\begin{array}{ccccccc|c} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \dots & \dots \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] = \left[\begin{array}{ccccccc|c} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \dots & \dots \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{array} \right] = \left[\begin{array}{ccccccc|c} 4 & 0 & 6 & 0 & 4 & 0 & 1 & 0 \\ 0 & 2 & 0 & 2 & 0 & 1 & 0 & 0 \end{array} \right] \quad (9.80)$$

As discussed in Example 9.5.1, if we delete the last column of $\tilde{\mathbf{S}}_1$, then the remaining $\tilde{\mathbf{S}}_1$ has full column rank and for any $\mathbf{F}(s)$, after deleting the last zero column, solutions exist in (9.80). Now if we delete the last N -row of $\tilde{\mathbf{S}}_1$, which is linearly dependent of its previous row, the set of solutions is unique and can be obtained, using MATLAB, as

$$[\mathbf{A}_0 \ \mathbf{M}_0 \ \mathbf{A}_1 \ \mathbf{M}_1] = \left[\begin{array}{ccccccc|c} 4 & -6 & 4 & -4 & 1 & 0 & 6 & 0 \\ 0 & 2 & 0 & 2 & 0 & 1 & 0 & 0 \end{array} \right] \quad (9.81)$$

Note that the last zero column is by assignment because of the deletion of the last N -row in $\tilde{\mathbf{S}}_1$. Thus we have

$$\mathbf{A}(s) = \left[\begin{array}{cc} 4+s & -6 \\ 0 & 2+s \end{array} \right], \quad \mathbf{M}(s) = \left[\begin{array}{cc} 4+6s & -4 \\ 0 & 2 \end{array} \right] \quad (9.82)$$

The two compensators $\mathbf{A}^{-1}(s)\mathbf{M}(s)$ and $\mathbf{A}^{-1}(s)\mathbf{L}(s)$ are clearly proper. This completes the design. As a check, we compute

$$\begin{aligned} \hat{\mathbf{G}}_o(s) &= \mathbf{N}(s)\mathbf{F}^{-1}(s)\mathbf{L}(s) = \left[\begin{array}{cc} \frac{2(s+2)}{s^3+4s^2+6s+4} & 0 \\ 0 & \frac{2}{s^2+2s+2} \end{array} \right] \\ &= \left[\begin{array}{cc} \frac{2}{s^2+2s+2} & 0 \\ 0 & \frac{2}{s^2+2s+2} \end{array} \right] \end{aligned}$$

This is the desired model. Note that the design involves the cancelation of $(s+2)$, which we can select. Thus the design is satisfactory.

Let us discuss a special case of model matching. Given $\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)$. Let us select $\hat{\mathbf{T}}(s) = \mathbf{D}(s)\mathbf{D}_f^{-1}(s)$, where $\mathbf{D}_f(s)$ has the same column degrees and the same column-degree-coefficient matrix as $\mathbf{D}(s)$. Then $\hat{\mathbf{T}}(s)$ is proper and $\hat{\mathbf{G}}_o(s) = \hat{\mathbf{G}}(s)\hat{\mathbf{T}}(s) = \mathbf{N}(s)\mathbf{D}_f^{-1}(s)$. This is the feedback transfer matrix discussed in (8.72). Thus the state feedback design discussed in Chapter 8 can also be carried out by using Procedure 9.M1.

9.6.1 Decoupling

Consider $p \times p \hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)$. We have assumed that $\hat{\mathbf{G}}(s)$ is nonsingular. Thus $\hat{\mathbf{G}}^{-1}(s) = \mathbf{D}(s)\mathbf{N}^{-1}(s)$ is well-defined; however, it is in general improper. Let us select $\hat{\mathbf{T}}(s)$ as

$$\hat{\mathbf{T}}(s) = \hat{\mathbf{G}}^{-1}(s)\text{diag}(d_1^{-1}(s), d_2^{-1}(s), \dots, d_p^{-1}(s)) \quad (9.83)$$

where $d_i(s)$ are CT stable polynomials of least degrees to make $\hat{\mathbf{T}}(s)$ proper. Define

$$\Sigma(s) = \text{diag}(d_1(s), d_2(s), \dots, d_p(s)) \quad (9.84)$$

Then we can write $\hat{\mathbf{T}}(s)$ as

$$\hat{\mathbf{T}}(s) = \mathbf{D}(s)\mathbf{N}^{-1}(s)\Sigma^{-1}(s) = \mathbf{D}(s)[\Sigma(s)\mathbf{N}(s)]^{-1} \quad (9.85)$$

If all transmission zeros of $\hat{\mathbf{G}}(s)$ or, equivalently, all roots of $\det \mathbf{N}(s)$ have negative real parts, then $\hat{\mathbf{T}}(s)$ is proper and BIBO stable. Thus the overall transfer matrix

$$\begin{aligned} \hat{\mathbf{G}}_o(s) &= \hat{\mathbf{G}}(s)\hat{\mathbf{T}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)\mathbf{D}(s)[\Sigma(s)\mathbf{N}(s)]^{-1} \\ &= \mathbf{N}(s)[\Sigma(s)\mathbf{N}(s)]^{-1} = \Sigma^{-1}(s) \end{aligned} \quad (9.86)$$

is implementable. This overall transfer matrix is a diagonal matrix and is said to be *decoupled*. The design in Example 9.6.1 is in fact of this design.

If $\hat{\mathbf{G}}(s)$ has nonminimum-phase transmission zeros or transmission zeros with zero or positive real parts, then the preceding design cannot be employed. However, with some modification, it is still possible to design a decoupled overall system. Consider a $p \times p \hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s)$. We factor $\mathbf{N}(s)$ as

$$\mathbf{N}(s) = \mathbf{N}_1(s)\mathbf{N}_2(s)$$

with

$$\mathbf{N}_1(s) = \text{diag}(\beta_{11}(s), \beta_{12}(s), \dots, \beta_{1p}(s))$$

where $\beta_{1i}(s)$ is the greatest common divisor of the i th row of $\mathbf{N}(s)$. Let us compute $\mathbf{N}_2^{-1}(s)$, and let $\beta_{2i}(s)$ be the least common denominator of the *unstable* poles of the i th column of $\mathbf{N}_2^{-1}(s)$. Define

$$\mathbf{N}_{2d} := \text{diag}(\beta_{21}(s), \beta_{22}(s), \dots, \beta_{2p}(s))$$

Then the matrix

$$\bar{\mathbf{N}}_2(s) := \mathbf{N}_2^{-1}(s)\mathbf{N}_{2d}(s)$$

has no unstable poles. Now we select $\hat{\mathbf{T}}(s)$ as

$$\hat{\mathbf{T}}(s) = \mathbf{D}(s)\bar{\mathbf{N}}_2(s)\Sigma^{-1}(s) \quad (9.87)$$

with

$$\Sigma(s) = \text{diag}(d_1(s), d_2(s), \dots, d_p(s))$$

where $d_i(s)$ are CT stable polynomials of least degrees to make $\hat{\mathbf{T}}(s)$ proper. Because $\bar{\mathbf{N}}_2(s)$ has only stable poles, and $d_i(s)$ are CT stable, $\hat{\mathbf{T}}(s)$ is BIBO stable. Consider

$$\begin{aligned} \hat{\mathbf{G}}_o(s) &= \hat{\mathbf{G}}(s)\hat{\mathbf{T}}(s) = \mathbf{N}_1(s)\bar{\mathbf{N}}_2(s)\mathbf{D}^{-1}(s)\mathbf{D}(s)\bar{\mathbf{N}}_2(s)\Sigma^{-1}(s) \\ &= \mathbf{N}_1(s)\bar{\mathbf{N}}_{2d}(s)\Sigma^{-1}(s) \\ &= \text{diag}\left(\frac{\beta_1(s)}{d_1(s)}, \frac{\beta_2(s)}{d_2(s)}, \dots, \frac{\beta_p(s)}{d_p(s)}\right) \end{aligned} \quad (9.88)$$

where $\beta_i(s) = \beta_{1i}(s)\beta_{2i}(s)$. It is proper because both $\mathbf{T}(s)$ and $\hat{\mathbf{G}}(s)$ are proper. It is clearly BIBO stable. Thus $\hat{\mathbf{G}}_o(s)$ is implementable and is a decoupled system.

Example 9.6.2 Consider

$$\hat{\mathbf{G}}(s) = \mathbf{N}(s)\mathbf{D}^{-1}(s) = \begin{bmatrix} s & 1 \\ s-1 & s-1 \end{bmatrix} \begin{bmatrix} s^3+1 & 1 \\ 0 & s^2 \end{bmatrix}^{-1}$$

We compute $\det \mathbf{N}(s) = (s-1)(s-1) = (s-1)^2$. The plant has two nonminimum-phase transmission zeros. We factor $\mathbf{N}(s)$ as

$$\mathbf{N}(s) = \mathbf{N}_1(s)\mathbf{N}_2(s) = \begin{bmatrix} 1 & 0 \\ 0 & s-1 \end{bmatrix} \begin{bmatrix} s & 1 \\ 1 & 1 \end{bmatrix}$$

with $\mathbf{N}_1(s) = \text{diag}(1, (s-1))$, and compute

$$\mathbf{N}_2^{-1}(s) = \frac{1}{(s-1)} \begin{bmatrix} 1 & -1 \\ -1 & s \end{bmatrix}$$

If we select

$$\mathbf{N}_{2d} = \text{diag}((s-1), (s-1)) \quad (9.89)$$

then the rational matrix

$$\bar{\mathbf{N}}_2(s) = \mathbf{N}_2^{-1}(s)\mathbf{N}_{2d}(s) = \begin{bmatrix} 1 & -1 \\ -1 & s \end{bmatrix}$$

has no unstable poles. We compute

$$\begin{aligned}\mathbf{D}(s)\bar{\mathbf{N}}_2(s) &= \begin{bmatrix} s^3 + 1 & 1 \\ 0 & s^2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & s \end{bmatrix} \\ &= \begin{bmatrix} s^3 & -s^3 + s - 1 \\ -s^2 & s^3 \end{bmatrix}\end{aligned}$$

If we choose

$$\Sigma(s) = \text{diag}((s^2 + 2s + 2)(s + 2), (s^2 + 2s + 2)(s + 2))$$

then

$$\hat{\mathbf{T}}(s) = \mathbf{D}(s)\bar{\mathbf{N}}_2(s)\Sigma^{-1}(s)$$

is proper. Thus the overall transfer matrix

$$\hat{\mathbf{G}}_o(s) = \hat{\mathbf{G}}(s)\hat{\mathbf{T}}(s) = \text{diag}\left(\frac{s-1}{(s^2+2s+2)(s+2)}, \frac{(s-1)^2}{(s^2+2s+2)(s+2)}\right)$$

is implementable. This is a decoupled system. This decoupled system will not track any step reference input. Thus we modify it as

$$\hat{\mathbf{G}}_o(s) = \text{diag}\left(\frac{-4(s-1)}{(s^2+2s+2)(s+2)}, \frac{4(s-1)^2}{(s^2+2s+2)(s+2)}\right) \quad (9.90)$$

which has $\hat{\mathbf{G}}_o(0) = \mathbf{I}$ and will track any step reference input.

Next we implement (9.90) in the two-parameter configuration. We follow Procedure 9.M1. To save space, we define $d(s) := (s^2 + 2s + 2)(s + 2)$. First we compute

$$\begin{aligned}\mathbf{N}^{-1}(s)\hat{\mathbf{G}}_o(s) &= \begin{bmatrix} s & 1 \\ s-1 & s-1 \end{bmatrix}^{-1} \begin{bmatrix} \frac{-4(s-1)}{d(s)} & 0 \\ 0 & \frac{4(s+1)^2}{d(s)} \end{bmatrix} \\ &= \frac{1}{(s-1)^2} \begin{bmatrix} s-1 & -1 \\ 1-s & s \end{bmatrix} \begin{bmatrix} \frac{-4(s-1)}{d(s)} & 0 \\ 0 & \frac{4(s+1)^2}{d(s)} \end{bmatrix} \\ &= \begin{bmatrix} \frac{-4}{d(s)} & \frac{-4}{d(s)} \\ \frac{4}{d(s)} & \frac{4s}{d(s)} \end{bmatrix} = \begin{bmatrix} d(s) & 0 \\ 0 & d(s) \end{bmatrix}^{-1} \begin{bmatrix} -4 & -4 \\ 4 & 4s \end{bmatrix} \\ &=: \bar{\mathbf{F}}^{-1}(s)\bar{\mathbf{E}}(s)\end{aligned}$$

It is a left-coprime fraction.

From the plant transfer matrix, we have

$$\mathbf{D}(s) = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}s + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}s^2 + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}s^3$$

and

$$\mathbf{N}(s) = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}s + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}s^2 + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}s^3$$

We use qr-decomposition to compute the row index of $\hat{\mathbf{G}}(s)$. We type

```
d1=[1 1 0 0 0 0 1 0];d2=[0 0 0 0 0 1 0 0];
n1=[0 1 1 0 0 0 0 0];n2=[-1 -1 1 1 0 0 0 0];
s2=[d1 0 0 0 0;d2 0 0 0 0;n1 0 0 0 0;n2 0 0 0 0;...
     0 0 d1 0 0;0 0 d2 0 0;0 0 n1 0 0;0 0 n2 0 0;...
     0 0 0 0 d1;0 0 0 0 d2;0 0 0 0 n1;0 0 0 0 n2];
[q,r]=qr(s2')
```

From the matrix r (not shown), we can see that there are three linearly independent N_1 -rows and two linearly independent N_2 -rows. Thus we have $v_1 = 3$ and $v_2 = 2$ and the row index equals $v = 3$. If we select

$$\hat{\mathbf{F}}(s) = \text{diag}((s+3)^2, (s+3))$$

then $\hat{\mathbf{F}}(s)\bar{\mathbf{F}}(s)$ is row-column reduced with row degrees $\{2, 2\}$ and column degrees $\{3, 2\}$. We set

$$\mathbf{L}(s) = \hat{\mathbf{F}}(s)\bar{\mathbf{E}}(s) = \begin{bmatrix} -4(s+3)^2 & -4(s+3)^2 \\ 4(s+3) & 4s(s+3) \end{bmatrix} \quad (9.91)$$

and solve $\mathbf{A}(s)$ and $\mathbf{M}(s)$ from

$$\mathbf{A}(s)\mathbf{D}(s) + \mathbf{M}(s)\mathbf{N}(s) = \hat{\mathbf{F}}(s)\bar{\mathbf{F}}(s) := \mathbf{F}(s)$$

Using MATLAB, they can be solved as

$$\mathbf{A}(s) = \begin{bmatrix} s^2 + 10s + 329 & 100 \\ -46 & s^2 + 7s + 6 \end{bmatrix} \quad (9.92)$$

and

$$\mathbf{M}(s) = \begin{bmatrix} -290s^2 - 114s - 36 & 189s + 293 \\ 46s^2 + 34s + 12 & -34s - 46 \end{bmatrix} \quad (9.93)$$

The compensators $\mathbf{A}^{-1}(s)\mathbf{M}(s)$ and $\mathbf{A}^{-1}(s)\mathbf{L}(s)$ are clearly proper. This completes the design.

The model matching discussed can be modified in several ways. For example, if stable roots of $\det \mathbf{N}_2(s)$ are not inside the sector shown in Fig. 8.3, they can be included in β_{2i} . Then they will be retained in $\hat{\mathbf{G}}_o(s)$ and will not be canceled in the design. Instead of decoupling the plant for each pair of input and output, we may decouple it for a group of inputs and a group of outputs. In this case, the resulting overall transfer matrix is a block diagonal matrix. These modifications are straightforward and will not be discussed.

9.7 Concluding Remarks

In this chapter, we used coprime fractions to carry out designs to achieve pole placement or model matching. For pole placement, the unity feedback configuration shown in Fig. 9.1(a), a one-degree-of-freedom configuration, can be used. If a plant has degree n , then any pole placement can be achieved by using a compensator of degree $n - 1$ or larger. If the degree of a compensator is larger than minimum required, the extra degrees can be used to achieve robust tracking, disturbance rejection, or other design objectives.

Model matching generally involves pole-zero cancelations. One-degree-of-freedom configurations cannot be used here because we have no freedom in selecting canceled poles. Any two-degree-of-freedom configuration can be used because we have freedom in selecting canceled poles. This text discusses only the two-parameter configuration shown in Fig. 9.4.

All designs in this chapter are achieved by solving sets of linear algebraic equations. The basic idea and procedure are the same for both the SISO and MIMO cases. All discussion in this chapter can be applied, without any modification, to discrete-time systems; the only difference is that CT stable polynomials must be replaced by DT stable polynomials which have all their roots lying inside the unit circle of the z -plane.

This chapter studies only strictly proper $\hat{G}(s)$. If $\hat{G}(s)$ is biproper, the basic idea and procedure are still applicable, but the degree of compensators must be increased to insure properness of compensators and well-posedness of overall systems. See Reference 6. The model matching in Section 9.6 can also be extended to nonsquare plant transfer matrices. See also Reference 6.

The controller-estimator design in Chapter 8 can also be carried out using polynomial fractions. See References 6 (pp. 506–514) and 7 (pp. 461–465). Conversely, because of the equivalence of controllable and observable state-space equations and coprime fractions, we should be able to use state-space equations to carry out all designs in this chapter. It is, however, believed that state-space design will be more complex and less transparent as we can see by comparing the designs in Sections 8.3.1 and 9.3.2.

The state-space approach first appeared in the 1960s and by the 1980s, the concepts of controllability and observability and controller-estimator design were integrated into most undergraduate control texts. The polynomial fraction approach in transfer functions was developed in the 1970s; its underlying concept of coprimeness, however, is an ancient one. Even though the concepts and design procedures of the coprime fraction approach are as simple as, if not simpler than, the state-space approach, the approach appears to be less widely known. It is hoped that this chapter has demonstrated its simplicity and usefulness and will help its dissemination.

Although the design of linear time-invariant lumped systems can be more easily carried out using transfer functions, for computer computation and simulation, real-time processing, and op-amp circuit implementation, state-space equations are indispensable. Moreover, in developing mathematical equations, it is generally simpler to develop sets of first-order differential equations, not necessarily linear and time-invariant, to

describe systems. After linearization and approximation, we can then obtain (LTI lumped) state-space equations. Only then can we compute their rational transfer functions.⁶ Thus both transfer functions and state-space equations are important in the study of systems.

Problems

- 9.1** In solving the compensator equation in (9.4), we often require $\deg B(s) \leq \deg A(s)$ to insure the compensator $C(s) = B(s)/A(s)$ to be proper. For $D(s)$ and $N(s)$ in Example 9.2.1, if $F(s) = s^2 + 3s + 4$, can you find $B(s)$ and $A(s)$ to meet the degree inequality? Verify that if $F(s)$ is of degree 3, then such solutions always exist.
- 9.2** Given a plant with transfer function $\hat{g}(s) = (s - 1)/(s^2 - 4)$. Find a compensator in the unity feedback configuration so that the overall system has desired poles at -2 and $-1 \pm j1$. Also find a feedforward gain so that the resulting system will track any step reference input.
- 9.3** Suppose the plant transfer function in Problem 9.2 changes to $\hat{g}(s) = (s - 0.9)/(s^2 - 4.1)$ after the design is completed. Can the overall system still track asymptotically any step reference input? If not, give two different designs, one with a compensator of degree 3 and another with degree 2, that will track asymptotically and robustly any step reference input. Do you need additional desired poles? If yes, place them at -3 .
- 9.4** Repeat Problem 9.2 for a plant with transfer function $\hat{g}(s) = (s - 1)/s(s - 2)$. Do you need a feedforward gain to achieve tracking of any step reference input? Give your reason?
- 9.5** Suppose the plant transfer function in Problem 9.4 changes to $\hat{g}(s) = (s - 0.9)/s(s - 2.1)$ after the design is completed. Can the overall system still track any step reference input? Is the design robust?
- 9.6** Consider a plant with transfer function $\hat{g}(s) = 1/(s - 1)$. Suppose a disturbance of form $w(t) = a \sin(2t + \theta)$, with unknown amplitude a and phase θ , enters the plant as shown in Fig. 9.2. Design a biproper compensator of degree 3 in the feedback system so that the output will track asymptotically any step reference input and reject the disturbance. Place the desired poles at $-1 \pm j2$ and $-2 \pm j1$.
- 9.7** Consider the unity feedback system shown in Fig. 9.10. The plant transfer function is $\hat{g}(s) = 2/s(s + 1)$. Can the output track robustly any step reference input? Can the output reject any step disturbance $w(t) = a$? Why?

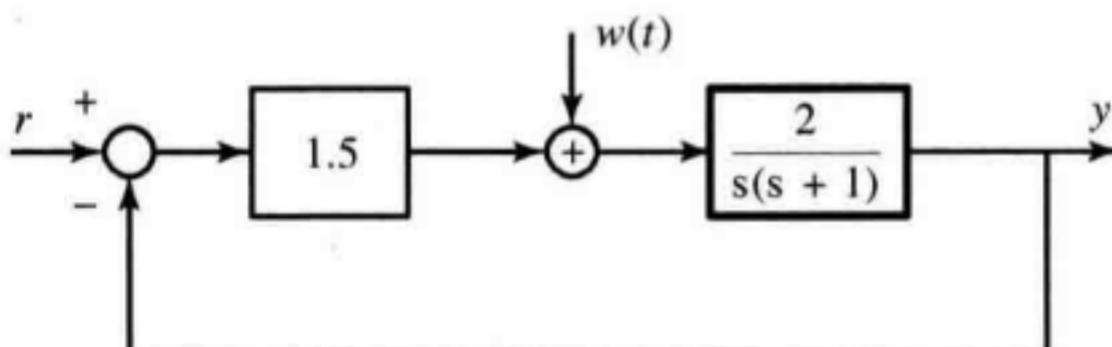


FIGURE 9.10

- 9.8** Consider the unity feedback system shown in Fig. 9.11(a). Is the transfer function from r to y BIBO stable? Is the system totally stable? If not, find an input-output pair whose closed-loop transfer function is not BIBO stable?

⁶Transfer functions can also be obtained directly by measuring frequency responses of some physical systems. See Reference 7. Such an approach, however, is not possible for systems such as the ones in Figs. 2.14 ~ 2.16.

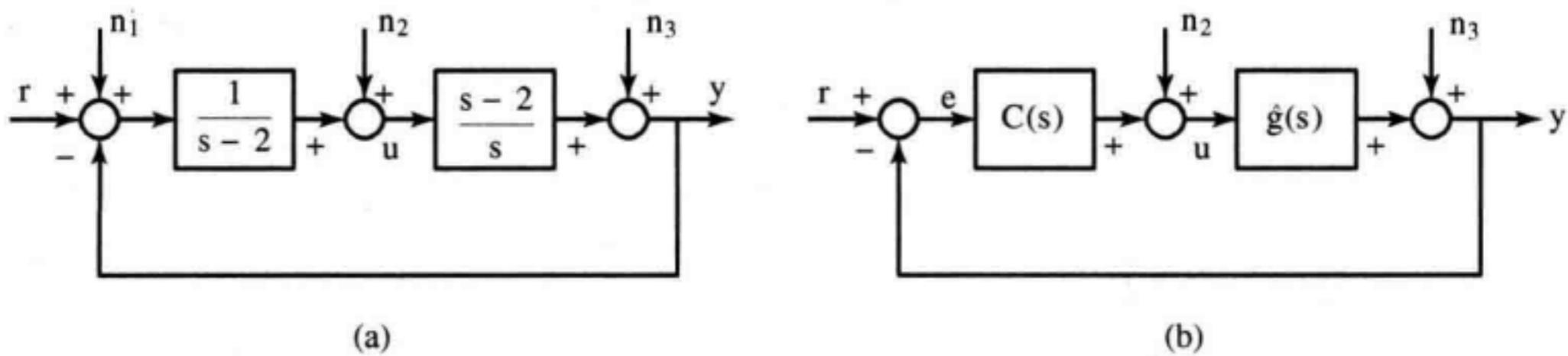


FIGURE 9.11

- 9.9** Consider the unity feedback system shown in Fig. 9.11(b). (1) Show that the closed-loop transfer function of every possible input-output pair contains the factor $(1 + C(s)\hat{g}(s))^{-1}$. (2) Show that $(1 + C(s)\hat{g}(s))^{-1}$ is proper if and only if

$$1 + C(\infty)\hat{g}(\infty) \neq 0$$

(3) Show that if $C(s)$ and $\hat{g}(s)$ are proper, and if $C(\infty)\hat{g}(\infty) \neq -1$, then the system is well posed.

- 9.10** Given $\hat{g}(s) = (s^2 - 1)/(s^3 + a_1s^2 + a_2s + a_3)$, which of the following $\hat{g}_o(s)$

$$\begin{aligned} & \frac{s-1}{(s+1)^2} \quad \frac{s+1}{(s+2)(s+3)} \quad \frac{s^2-1}{(s-2)^3} \\ & \frac{(s^2-1)}{(s+2)^2} \quad \frac{(s-1)(b_0s+b_1)}{(s+2)^2(s^2+2s+2)} \quad \frac{1}{1} \end{aligned}$$

for any a_i and b_i , are implementable?

- 9.11** Given $\hat{g}(s) = (s-1)/s(s-2)$, implement the model $\hat{g}_o(s) = -2(s-1)/(s^2+2s+2)$ in the open-loop and the unity feedback configurations. Are they totally stable? Can the implementations be used in practice?

- 9.12** Implement Problem 9.11 in the two-parameter configuration. Select the poles to be canceled at $s = -3$. Is $A(s)$ a CT stable polynomial? Can you implement the two compensators as shown in Fig. 9.4(a)? Implement the two compensators in Fig. 9.4(d) and draw its op-amp circuit.

- 9.13** Given a BIBO stable $\hat{g}_o(s)$. Show that the steady-state response $y_{ss}(t) := \lim_{t \rightarrow \infty} y(t)$ excited by the ramp reference input $r(t) = at$ for $t \geq 0$ is given by

$$y_{ss}(t) = \hat{g}_o(0)at + \hat{g}'_o(0)a$$

Thus if the output is to track asymptotically the ramp reference input we require $\hat{g}_o(0) = 1$ and $\hat{g}'_o(0) = 0$.

- 9.14** Given a BIBO stable

$$\hat{g}_o(s) = \frac{b_0 + b_1s + \dots + b_ms^m}{a_0 + a_1s + \dots + a_ns^n}$$

with $n \geq m$. Show that $\hat{g}_o(0) = 1$ and $\hat{g}'_o(0) = 0$ if and only if $a_0 = b_0$ and $a_1 = b_1$.

- 9.15** Given a plant with transfer function $\hat{g}(s) = (s+3)(s-2)/(s^3+2s-1)$. (1) Find conditions on b_1 , b_0 and a so that

$$\hat{g}_o(s) = \frac{b_1s + b_0}{s^2 + 2s + a}$$

is implementable. (2) Is

$$\hat{g}_o(s) = \frac{(s-2)(b_1s+b_0)}{(s+2)(s^2+2s+2)}$$

implementable? Find conditions on b_1 and b_2 so that the overall transfer function will track any ramp reference input.

- 9.16** Consider a plant with transfer matrix

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} \frac{s+1}{s(s-1)} \\ \frac{1}{s^2-1} \end{bmatrix}$$

Find a compensator in the unity feedback configuration so that the poles of the overall system are located at $-2, -1 \pm j$ and the rest at $s = -3$. Can you find a feedforward gain so that the overall system will track asymptotically any step reference input?

- 9.17** Repeat Problem 9.16 for a plant with transfer matrix

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} \frac{s+1}{s(s-1)} & \frac{1}{s^2-1} \end{bmatrix}$$

- 9.18** Repeat Problem 9.16 for a plant with transfer matrix

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} \frac{s-2}{s^2-1} & \frac{1}{s-1} \\ \frac{1}{s} & \frac{2}{s-1} \end{bmatrix}$$

- 9.19** Given a plant with the transfer matrix given in Problem 9.18. Is

$$\hat{\mathbf{G}}_o(s) = \begin{bmatrix} \frac{4(s^2-4s+1)}{(s^2+2s+2)(s+2)} & 0 \\ 0 & \frac{4(s^2-4s+1)}{(s^2+2s+2)(s+2)} \end{bmatrix}$$

implementable. If yes, implement it.

- 9.20** Diagonalize a plant with transfer matrix

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} 1 & s \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & s^2+1 \\ s & 0 \end{bmatrix}^{-1}$$

Set the denominator of each diagonal entry of the resulting overall system as $s^2 + 2s + 2$.

References

We list only the references used in preparing this edition. The list does not indicate original sources of the results presented. For a more extensive references, see Reference 2, 6, 16, and 18.

1. Anderson, B. D. O., and Moore, J. B., *Optimal Control—Linear Quadratic Methods*, Englewood Cliffs, NJ: Prentice Hall, 1990.
2. Antsaklis, A. J., and Michel, A. N., *Linear Systems*, New York: McGraw-Hill, 1997.
3. Callier, F. M., and Desoer, C. A., *Multivariable Feedback Systems*, New York: Springer-Verlag, 1982.
4. Callier, F. M., and Desoer, C. A., *Linear System Theory*, New York: Springer-Verlag, 1991.
5. Chen, C. T., *Introduction to Linear System Theory*, New York: Holt, Rinehart & Winston, 1970.
6. Chen, C. T., *Linear System Theory and Design*, New York: Oxford University Press, 1984.
7. Chen, C. T., *Analog and Digital Control System Design: Transfer Function, State-Space, and Algebraic Methods*, New York: Oxford University Press, 1993.
8. Chen, C. T., *Digital Signal Processing: Spectral Computation and Filter Design*, New York: Oxford University Press, 2001.
9. Chen, C. T., *Signals and Systems*, 3rd ed., New York: Oxford University Press, 2004.
10. Chen, C. T., *Signals and Systems: A Fresh Look*, CreateSpace, an amazon.com company, 2011.
11. Chen, C. T., and Liu, C. S., “Design of Control Systems: A Comparative Study,” *IEEE Control System Magazine*, vol. 14, no. 5, pp. 47–51, 1994.
12. Doyle, J. C., Francis, B. A., and Tannenbaum, A. R., *Feedback Control Theory*, New York: Macmillan, 1992.
13. Gantmacher, F. R., *The Theory of Matrices*, vols. 1 and 2. New York: Chelsea, 1959.
14. Golub, G. H., and Van Loan, C. F., *Matrix Computations*, 3rd ed. Baltimore: Johns Hopkins University Press, 1996.
15. Howze, J. W., and Bhattacharyya, S. P., “Robust tracking, error feedback, and two-degree-of-freedom controllers,” *IEEE Trans. on Automatic Control*, vol. 42, pp. 980–983, 1997.
16. Kailath, T., *Linear Systems*, Englewood Cliffs, NJ: Prentice Hall, 1980.
17. Körner, T. W., *Fourier Analysis*, Cambridge, UK: Cambridge University Press, 1988.

18. Kurcera, V., *Analysis and Design of Discrete Linear Control Systems*, London: Prentice-Hall, 1991.
19. Rugh, W., *Linear System Theory*, 2nd ed., Upper Saddle River, NJ: Prentice Hall, 1996.
20. Strang, G., *Linear Algebra and Its Application*, 3rd ed., San Diego: Harcourt, Brace, Janovich, 1988.
21. Tongue, B. H., *Principles of Vibration*, New York: Oxford University Press, 1996.
22. Vardulakis, A. I. G., *Linear Multivariable Control*, Chichester: John Wiley, 1991.
23. Vidyasagar, M., *Control System Synthesis: A Factorization Approach*, Cambridge, MA: MIT Press, 1985.
24. Wolovich, W. Q., *Linear Multivariable Systems*, New York: Springer-Verlag, 1974.
25. Zhou, K., Doyle, J. C., and Glover, K., *Robust and Optimal Control*, Upper Saddle River, NJ: Prentice Hall, 1995.

Answers to Selected Problems

Chapter 2

2.1 (a) Linear. (b) and (c) Nonlinear. In (b), shift the operating point to $(0, y_o)$, then (u, \bar{y}) is linear, where $\bar{y} = y - y_o$.

2.3 Linear, time varying, causal.

2.5 No, yes, no for $\mathbf{x}(0) \neq \mathbf{0}$. Yes, yes, yes for $\mathbf{x}(0) = \mathbf{0}$. The reason is that the superposition property must also apply to the initial states.

2.8

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 0 & 1/C_1 \\ 0 & 0 & 1/C_2 \\ -1/L_1 & -1/L_1 & -(R_1 + R_2)/L_1 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 & -1/C_1 \\ 0 & 0 \\ 1/L_1 & R_1/L_1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$y = [-1 \ -1 \ -R_1] \mathbf{x} + [1 \ R_1] \mathbf{u}$$

$$\hat{g}_1(s) = \frac{\hat{y}(s)}{\hat{u}_1(s)} = \frac{s^2 + (R_2/L_1)s}{s^2 + \left(\frac{R_1+R_2}{L_1}\right)s + \left(\frac{1}{C_1} + \frac{1}{C_2}\right)\frac{1}{L_1}}$$

$$\hat{g}_2(s) = \frac{\hat{y}(s)}{\hat{u}_2(s)} = \frac{(R_1s + (1/C_1))(s + (R_2/L_1))}{s^2 + \left(\frac{R_1+R_2}{L_1}\right)s + \left(\frac{1}{C_1} + \frac{1}{C_2}\right)\frac{1}{L_1}}$$

$$\hat{y}(s) = \hat{g}_1(s)\hat{u}_1(s) + \hat{g}_2(s)\hat{u}_2(s) = [\hat{g}_1(s) \ \hat{g}_2(s)] \begin{bmatrix} \hat{u}_1(s) \\ \hat{u}_2(s) \end{bmatrix}$$

2.10

$$\dot{\mathbf{x}} = \begin{bmatrix} -1/RC_1 & 0 & -1/C_1 \\ 0 & 0 & 1/C_2 \\ 1/L & -1/L & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 1/RC_1 \\ 0 \\ 0 \end{bmatrix} u$$

$$y = [1 \ -1 \ 0] \mathbf{x}$$

$$\hat{g}(s) = \frac{LC_2 s}{RCC_1 C_2 s^3 + LC_2 s^2 + R(C_1 + C_2)s + 1}$$

2.12

$$\dot{\mathbf{x}} = \begin{bmatrix} -1/3 & 1/3 & 0 \\ 1/2 & 0 & 0 \\ -1/3 & -1/2 & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 1/3 \\ 0 \\ 1/3 \end{bmatrix} u$$

$$y = [0 \ 1 \ 0] \mathbf{x}$$

$$\hat{g}(s) = \frac{1}{6s^2 + 2s + 1}$$

2.14 $\hat{g}(s) = 1/(s+3)$, $g(t) = e^{-3t}$ for $t \geq 0$.

2.16

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} D_{11}(s) & D_{12}(s) \\ D_{21}(s) & D_{22}(s) \end{bmatrix}^{-1} \begin{bmatrix} N_{11}(s) & N_{12}(s) \\ N_{21}(s) & N_{22}(s) \end{bmatrix}$$

2.23 (a) Define $x_1 = \theta$ and $x_2 = \dot{\theta}$. Then $\dot{x}_1 = x_2$, $\dot{x}_2 = -(g/l) \sin x_1 - (u/ml) \cos x_1$. If θ is small, then

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 \\ -g/l & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ -1/mg \end{bmatrix} u$$

It is a linear state equation.

2.25 Define $x_1 = h$, $x_2 = \dot{h}$, $x_3 = \theta$, $x_4 = \dot{\theta}$. Then

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & k_1/m & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -l_1 k_1/I & -b/I \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ -k_2/m \\ 0 \\ (l_1 + l_2)k_2 \end{bmatrix} u$$

$$y = [1 \ 0 \ 0 \ 0] \mathbf{x}$$

If $I = 0$, then

$$\hat{g}(s) = \frac{k_1 k_2 l_2 - k_2 b s}{m s^2 (b s + k_1 l_1)}$$

2.27 $\hat{g}_1(s) = \hat{y}_1(s)/\hat{u}(s) = 1/(A_1 R_1 s + 1)$, $\hat{g}_2(s) = \hat{y}(s)/\hat{y}_1(s) = 1/(A_2 R_2 s + 1)$. Yes, $\hat{y}(s)/\hat{u}(s) = \hat{g}_1(s)\hat{g}_2(s)$.

Chapter 3**3.1** $[1/3 \ 8/3]', [-2 \ 1.5]'$.**3.3**

$$\mathbf{q}_1 = \frac{1}{3.74} \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}, \quad \mathbf{q}_2 = \frac{1}{1.732} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

3.6 $\rho(\mathbf{A}_1) = 2$, nullity(\mathbf{A}_1) = 1; 3, 0; 3, 1.**3.12** $\mathbf{x} = [-1 \ -1]$ is a solution. Unique. No solution if $\mathbf{y} = [1 \ 1 \ 1]'$.**3.14** $\alpha_1 = 4/11$, $\alpha_2 = 16/11$. The solution

$$\mathbf{x} = [4/11 \ -8/11 \ -4/11 \ -16/11]'$$

has the smallest 2-norm.

3.17 The representations of \mathbf{A} with respect to $\{\mathbf{b}_1 \ \mathbf{Ab}_i \ \mathbf{A}^2\mathbf{b}_i \ \mathbf{A}^3\mathbf{b}_i\}$, for $i = 1, 2$, both equal

$$\bar{\mathbf{A}} = \begin{bmatrix} 0 & 0 & 0 & -8 \\ 1 & 0 & 0 & 20 \\ 0 & 1 & 0 & -18 \\ 0 & 0 & 1 & 7 \end{bmatrix}$$

3.18

$$\mathbf{Q}_3 = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \hat{\mathbf{A}}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$\mathbf{Q}_4 = \begin{bmatrix} 5 & 4 & 0 \\ 0 & 20 & 1 \\ 0 & -25 & 0 \end{bmatrix}, \quad \hat{\mathbf{A}}_4 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

3.23

$$\Delta_1(\lambda) = (\lambda - \lambda_1)^3(\lambda - \lambda_2), \quad \psi_1(\lambda) = \Delta_1(\lambda)$$

$$\Delta_2(\lambda) = (\lambda - \lambda_1)^4, \quad \psi_2(\lambda) = (\lambda - \lambda_1)^3$$

$$\Delta_3(\lambda) = (\lambda - \lambda_1)^4, \quad \psi_3(\lambda) = (\lambda - \lambda_1)^2$$

$$\Delta_4(\lambda) = (\lambda - \lambda_1)^4, \quad \psi_4(\lambda) = (\lambda - \lambda_1)$$

3.26

$$\mathbf{A}^{10} = \begin{bmatrix} 1 & 1 & 9 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{A}^{103} = \begin{bmatrix} 1 & 1 & 102 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$e^{\mathbf{At}} = \begin{bmatrix} e^t & e^t - 1 & te^t - e^t + 1 \\ 0 & 1 & e^t - 1 \\ 0 & 0 & e^t \end{bmatrix}$$

3.27

$$e^{\mathbf{A}_4 t} = \begin{bmatrix} 1 & 4t + 2.5t^2 & 3t + 2t^2 \\ 0 & 1 + 20t & 16t \\ 0 & -25t & 1 - 20t \end{bmatrix}$$

3.29

$$\mathbf{B} = \begin{bmatrix} \ln \lambda_1 & 0 & 0 \\ 0 & \ln \lambda_2 & 0 \\ 0 & 0 & \ln \lambda_3 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} \ln \lambda & 1/\lambda & 0 \\ 0 & \ln \lambda & 0 \\ 0 & 0 & \ln \lambda \end{bmatrix}$$

3.37 Eigenvalues: 0, 0. No solution for \mathbf{C}_1 . For any m_1 , $[m_1 \ 3 - m_1]'$ is a solution for \mathbf{C}_2 .

3.39 $\sqrt{6}$, 1, 4.7, 1.7.

Chapter 4

4.2 $y(t) = 5e^{-t} \sin t$ for $t \geq 0$.

4.4 For $T = \pi$,

$$\mathbf{x}[k+1] = \begin{bmatrix} -0.0432 & 0 \\ 0 & -0.0432 \end{bmatrix} \mathbf{x}[k] + \begin{bmatrix} 1.5648 \\ -1.0432 \end{bmatrix} u[k]$$

$$y[k] = [2 \ 3] \mathbf{x}[k]$$

4.7 MATLAB yields $|y|_{max} = 0.55$, $|x_1|_{max} = 0.5$, $|x_2|_{max} = 1.05$ and $|x_3|_{max} = 0.52$ for unit step input. Define $\bar{x}_1 = x_1$, $\bar{x}_2 = 0.5x_2$ and $\bar{x}_3 = x_3$. Then

$$\dot{\bar{\mathbf{x}}} = \begin{bmatrix} -2 & 0 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & -4 & -2 \end{bmatrix} \bar{\mathbf{x}} + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} u, \quad y = [1 \ -2 \ 0] \bar{\mathbf{x}}$$

The largest permissible a is $10/0.55 = 18.2$.

4.12 They are not equivalent but are zero-state equivalent.

4.15 Using (4.34), we have

$$\dot{\mathbf{x}} = \begin{bmatrix} -3 & 0 & -2 & 0 \\ 0 & -3 & 0 & -2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} u$$

$$\mathbf{y} = \begin{bmatrix} 2 & 2 & 4 & -3 \\ -3 & -2 & -6 & -2 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} u$$

4.17

$$\dot{\mathbf{x}} = \begin{bmatrix} -3 & 1 & 0 & 0 \\ -2 & 0 & 0 & 0 \\ 0 & 0 & -3 & 1 \\ 0 & 0 & -2 & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 2 & 2 \\ 4 & -3 \\ -3 & -2 \\ -6 & -2 \end{bmatrix} u$$

$$\mathbf{y} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} u$$

Both have dimension 4.

4.20

$$\mathbf{X}(t) = \begin{bmatrix} 1 & \int_0^t e^{0.5\tau^2} d\tau \\ 0 & e^{0.5t^2} \end{bmatrix}$$

$$\Phi(t, t_0) = \begin{bmatrix} 1 & -e^{0.5t^2} \int_{t_0}^t e^{0.5\tau^2} d\tau \\ 0 & e^{0.5(t^2 - t_0^2)} \end{bmatrix}$$

$$\mathbf{X}(t) = \begin{bmatrix} e^{-t} & e^t \\ 0 & 2e^{-t} \end{bmatrix}$$

$$\Phi(t, t_0) = \begin{bmatrix} e^{-(t-t_0)} & 0.5(e^t e^{t_0} - e^{-t} e^{t_0}) \\ 0 & e^{-(t-t_0)} \end{bmatrix}$$

4.29

$$\Phi(t, t_0) = \begin{bmatrix} e^{\cos t - \cos t_0} & 0 \\ 0 & e^{-\sin t + \sin t_0} \end{bmatrix}$$

4.32 Let $\bar{\mathbf{x}}(t) = \mathbf{P}(t)\mathbf{x}(t)$ with

$$\mathbf{P}(t) = \begin{bmatrix} e^{-\cos t} & 0 \\ 0 & e^{\sin t} \end{bmatrix}$$

Then $\dot{\bar{\mathbf{x}}}(t) = \mathbf{0} \cdot \bar{\mathbf{x}} = \mathbf{0}$.**4.34**

$$\dot{\mathbf{x}} = \mathbf{0} \cdot \mathbf{x} + \begin{bmatrix} t^2 e^{-\lambda t} \\ -2t e^{-\lambda t} \\ e^{-\lambda t} \end{bmatrix} u \quad y = [e^{\lambda t} \ te^{\lambda t} \ t^2 e^{\lambda t}] \mathbf{x}$$

$$\dot{\mathbf{x}} = \begin{bmatrix} 3\lambda & -3\lambda^2 & \lambda^3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u, \quad y = [0 \ 0 \ 2] \mathbf{x}$$

Chapter 5

5.1 The bounded input $u(t) = \sin t$ excites $y(t) = 0.5t \sin t$ which is not bounded. Thus the system is not BIBO stable.

5.3 No. Yes.**5.6** If $u(t) = 3$, then $y(t) \rightarrow -6$. If $u(t) = \sin 2t$, then $y(t) \rightarrow 1.26 \sin(2t + 1.25)$.**5.9** Its transfer function is $\hat{g}(s) = -4/(s+1)$. BIBO stable.**5.10** Yes.**5.12** Not asymptotically stable. Its minimal polynomial can be found from its Jordan form as $\psi(\lambda) = \lambda(\lambda + 1)$. $\lambda = 0$ is a simple eigenvalue, thus the equation is marginally stable.**5.20** Not asymptotically stable. Its minimal polynomial can be fund from its Jordan form as $\psi(\lambda) = (\lambda - 1)^2(\lambda + 1)$. $\lambda = 1$ is a repeated eigenvalue, thus the equation is not marginally stable.**5.22** If $\mathbf{N} = \mathbf{I}$, then

$$\mathbf{M} = \begin{bmatrix} 1075 & 1 \\ 1 & 1.5 \end{bmatrix}$$

It is positive definite, thus both eigenvalues have negative real parts.

5.23 If $\mathbf{N} = \mathbf{I}$, then

$$\mathbf{M} = \begin{bmatrix} 2.2 & 1.6 \\ 1.6 & 4.8 \end{bmatrix}$$

It is positive definite, thus both eigenvalues have magnitudes less than 1.

5.25 Because $\mathbf{x}'\mathbf{M}\mathbf{x} = \mathbf{x}'[0.5(\mathbf{M} + \mathbf{M}')]\mathbf{x}$, the only way to check positive definiteness of nonsymmetric \mathbf{M} is to check positive definiteness of symmetric $0.5(\mathbf{M} + \mathbf{M}')$.**5.28** Both are BIBO stable.**5.30** BIBO stable, marginally stable, not asymptotically stable. $P(t)$ is not a Lyapunov transformation.

Chapter 6

6.2 Controllable, observable.**6.5**

$$\dot{\mathbf{x}} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u, \quad y = [0 \ -1] \mathbf{x} + 2u$$

Not controllable, not observable.

6.7 $\mu_i = 1$ for all i and $\mu = 1$.**6.9** $y = 2u$.**6.14** Controllable, not observable.**6.23** Using x_1 and x_2 yields

$$\dot{\mathbf{x}} = \begin{bmatrix} -2/11 & 0 \\ 3/22 & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} -2/11 \\ 3/22 \end{bmatrix} u, \quad y = [-1 \ -1] \mathbf{x}$$

This 2-dim. equation is not controllable but observable.

Using x_1, x_2 , and x_3 yields

$$\dot{\mathbf{x}} = \begin{bmatrix} -2/11 & 0 & 0 \\ 3/22 & 0 & 0 \\ 1/22 & 0 & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} -2/11 \\ 3/22 \\ 1/22 \end{bmatrix} u, \quad y = [0 \ 0 \ 1] \mathbf{x}$$

This three-dimensional equation is not controllable and not observable.

6.25 For $T = \pi$, not controllable and not observable.

6.27 Not controllable at any t . Observable at every t .

Chapter 7

7.1

$$\dot{\mathbf{x}} = \begin{bmatrix} -2 & 1 & 2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u, \quad y = [0 \ 1 \ 2] \mathbf{x}$$

Not observable.

7.3

$$\dot{\mathbf{x}} = \begin{bmatrix} -2 & 1 & 2 & a_1 \\ 1 & 0 & 0 & a_2 \\ 0 & 1 & 0 & a_3 \\ 0 & 0 & 0 & a_4 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} u, \quad y = [0 \ 1 \ -1 \ c_4] \mathbf{x}$$

For any a_i and c_4 , it is an uncontrollable and unobservable realization.

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u, \quad y = [0 \ 1] \mathbf{x}$$

A controllable and observable realization.

7.8 Solving the monic null vector of

$$\begin{bmatrix} -1 & -1 & 0 & 0 \\ 0 & 2 & -1 & -1 \\ 4 & 0 & 0 & 2 \\ 0 & 0 & 4 & 0 \end{bmatrix} \begin{bmatrix} -N_0 \\ D_0 \\ -N_1 \\ D_1 \end{bmatrix} = \mathbf{0}$$

as $[-0.5 \ 0.5 \ 0 \ 1]'$. Thus $0.5/(0.5+s) = 1/(2s+1)$.

7.13

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 \\ -2 & -1 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \quad y = [1 \ 0] \mathbf{x}$$

7.19

$$\Delta_1(s) = s(s+1)(s+3), \quad \deg = 3$$

$$\Delta_2(s) = (s+1)^3(s+2)^2, \quad \deg = 5$$

$$\Delta_3(s) = (s+1)^2(s+3)^2(s+2)(s+4)(s+5)s, \quad \deg = 8$$

7.23 Its right coprime fraction can be computed from any left fraction as

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} 2.5 & s+0.5 \\ 2.5 & s+2.5 \end{bmatrix} \begin{bmatrix} 0.5s & s^2+0.5s \\ s-0.5 & -0.5 \end{bmatrix}^{-1}$$

or, by interchanging their two columns,

$$\hat{\mathbf{G}}(s) = \begin{bmatrix} s+0.5 & 2.5 \\ s+2.5 & 2.5 \end{bmatrix} \begin{bmatrix} s^2+0.5s & 0.5s \\ -0.5 & s-0.5 \end{bmatrix}^{-1}$$

Using the latter, we can obtain

$$\dot{\mathbf{x}} = \begin{bmatrix} -0.5 & -0.25 & -0.25 \\ 1 & 0 & 0 \\ 0 & 0.5 & 0.5 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 1 & -0.5 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{u}$$

$$\mathbf{y} = \begin{bmatrix} 1 & 0.5 & 2.5 \\ 1 & 2.5 & 2.5 \end{bmatrix} \mathbf{x}$$

Chapter 8

8.1 $\mathbf{k} = [4 \ 1]$.

8.3 For

$$\mathbf{F} = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} \quad \bar{\mathbf{k}} = [1 \ 1]$$

we have

$$\mathbf{T} = \begin{bmatrix} 0 & 1/13 \\ 1 & 9/13 \end{bmatrix} \quad \mathbf{k} = \bar{\mathbf{k}}\mathbf{T}^{-1} = [4 \ 1]$$

8.6 Yes, yes, yes.

8.8 $u = pr - \mathbf{kx}$ with $\mathbf{k} = [15 \ 47 \ -8]$ and $p = 0.5$.

8.10 $u[k] = pr[k] - \mathbf{kx}[k]$ with $\mathbf{k} = [1 \ 5 \ 2]$ and $p = 0.5$. The overall transfer function is

$$\hat{g}_f(z) = \frac{0.5(2z^2 - 8z + 8)}{z^3}$$

and the output excited by $r[k] = a$ is $y[0] = 0$, $y[1] = a$, $y[2] = -3a$ and $y[k] = a$ for $k \geq 3$.

8.12 Two-dimensional estimator:

$$\dot{\mathbf{z}} = \begin{bmatrix} -2 & 2 \\ -2 & -2 \end{bmatrix} \mathbf{z} + \begin{bmatrix} 0.6282 \\ -0.3105 \end{bmatrix} u + \begin{bmatrix} 1 \\ 0 \end{bmatrix} y$$

$$\hat{\mathbf{x}} = \begin{bmatrix} -12 & -27.5 \\ 19 & 32 \end{bmatrix} \mathbf{z}$$

One-dimensional estimator:

$$\dot{z} = -3z + (13/21)u + y$$

$$\hat{\mathbf{x}} = \begin{bmatrix} -4 & 21 \\ 5 & -21 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix}$$

8.14 Select

$$\mathbf{F} = \begin{bmatrix} -4 & 3 & 0 & 0 \\ -3 & 4 & 0 & 0 \\ 0 & 0 & -5 & 4 \\ 0 & 0 & -4 & -5 \end{bmatrix}$$

$$\text{If } \bar{\mathbf{K}} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \text{then } \mathbf{K} = \begin{bmatrix} 62.5 & 147 & 20 & 515.5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\text{If } \bar{\mathbf{K}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \text{then } \mathbf{K} = \begin{bmatrix} -606.2 & -168 & -14.2 & -2 \\ 371.1 & 119.2 & 14.9 & 2.2 \end{bmatrix}$$

Chapter 9

9.2 $C(s) = (10s + 20)/(s - 6)$, $p = -0.2$.

9.4 $C(s) = (22s - 4)/(s - 16)$, $p = 1$. There is no need for a forward gain because $\hat{g}(s)$ contains $1/s$.

9.6 $C(s) = (7s^3 + 14s^2 + 34s + 25)/(s(s^2 + 4))$.

9.8 Yes, no. The transfer function from r to u is $\hat{g}_{ur}(s) = s/(s + 1)(s - 2)$, which is not BIBO stable.

9.10 Yes, no, no, no, yes, no. (row-wise)

9.12 $C_1(s) = -2(s + 3)/(s - 21)$, $C_2(s) = (28s - 6)/(s - 21)$. $A(s) = s - 21$ is not CT stable. Its implementation in Fig. 9.4(a) will not be totally stable. A minimal realization of $[C_1(s) \ - C_2(s)]$ is

$$\dot{x} = 21x + [-48 \ -582] \begin{bmatrix} r \\ y \end{bmatrix}, \quad y = x + [-2 \ -28] \begin{bmatrix} r \\ y \end{bmatrix}$$

From which an op-amp circuit can be drawn.

9.15 (1) $a > 0$ and $b_0 = -2b_1$. (2) Yes, $b_0 = -2$, $b_1 = -4$.

9.16 The 1×2 compensator

$$\mathbf{C}(s) = \frac{1}{s + 3.5} [3.5s + 12 \ -2]$$

will place the closed-loop poles at $-2, -1 \pm j, -3$. No.

9.18 If we select

$$\mathbf{F}(s) = \text{diag}((s + 2)(s^2 + 2s + 2)(s + 3), (s^2 + 2s + 2))$$

then we cannot find a feedforward gain to achieve tracking. If

$$\mathbf{F}(s) = \begin{bmatrix} (s+2)(s^2+2s+2)(s+3) & 0 \\ 1 & s^2+2s+2 \end{bmatrix}$$

then the compensator

$$\mathbf{C}(s) = \mathbf{A}^{-1}(s)\mathbf{B}(s) = \begin{bmatrix} s-4.7 & -53.7 \\ -3.3 & s-4.3 \end{bmatrix}^{-1} \begin{bmatrix} -30.3s-29.7 & 4.2s-12 \\ -0.7s-0.3 & 4s-1 \end{bmatrix}$$

and the feedforward gain matrix

$$\mathbf{P} = \begin{bmatrix} 0.92 & 0 \\ -4.28 & 1 \end{bmatrix}$$

will achieve the design.

9.20 The diagonal transfer matrix

$$\hat{\mathbf{G}}_o(s) = \begin{bmatrix} \frac{-2(s-1)}{s^2+2s+2} & 0 \\ 0 & \frac{-2(s-1)}{s^2+2s+2} \end{bmatrix}$$

is implementable. The proper compensators $\mathbf{A}^{-1}(s)\mathbf{L}(s)$ and $\mathbf{A}^{-1}(s)\mathbf{M}(s)$ with

$$\mathbf{A}(s) = \begin{bmatrix} s+5 & -14 \\ 0 & s+4 \end{bmatrix}, \quad \mathbf{L}(s) = \begin{bmatrix} -2(s+3) & 2(s+3) \\ 2 & -2s \end{bmatrix}$$

$$\mathbf{M}(s) = \begin{bmatrix} -6 & 13s+1 \\ 2 & -2s \end{bmatrix}$$

in the two-parameter configuration will achieve the design.

Index

Actuating signal, 282, 323
Adder, 116
Additivity, 11, 42
Adjoint, 66, 126
Admittance, 25
Asymptotic stability, 168, 169
Asymptotic tracking, 293, 331,
 360
 of ramp input, 332
 of step input, 332

Balanced realization, 245
Basis, 58
 change of, 58, 67
 orthonormal, 58
BIBO stability, 154, 162
Bilinear transformation,
 175

Black box, 6
Blocking zero, 40

Causality, 7, 40
Cayley-Hamilton theorem, 77
Characteristic polynomial,
 of constant matrix, 69
 of rational matrix, 252
Companion-form matrix, 68, 70,
 98
 characteristic polynomial of,
 70, 98
 eigenvectors of, 98
Compensator, 325
 feedback, 334
 feedforward, 334
Compensator equation, 95
Complete characterization, 237

- Control configuration,
 - closed-loop, 282, 323
 - controller-estimator, 324
 - feedback, 282, 323
 - no-plant leakage, 339
 - one-degree-of-freedom, 323
 - open-loop, 282, 323
 - plant-input-output feedback, 324
 - two-degree-of-freedom, 324
 - two-parameter, 324, 365
 - unity-feedback, 323, 359
- Controllability, 184, 210
 - after sampling, 214
 - from the origin, 213
 - Gramian, 185, 188, 211
 - in Jordan-form, 206
 - index, 191
 - matrix, 185, 188, 257
 - to the origin, 213
- Convolution, integral, 13
 - discrete, 42
- Coprimeness, 39, 239, 257
- Cyclic matrix, 308
- Dead-beat design, 322
- Decoupling, 370
- Degree of rational function, 233
 - McMillan, 252
 - of rational matrix, 252
- Descriptions of systems,
 - external, 3
 - convolution, 3, 13, 42
 - differential equation, 3
- input-output, 3
- internal, 3
- time domain, 3
- state-space (ss) equation, 3
- transfer function (tf), 3
- transform domain, 3
- Determinant, 66
- Direct transmission part, 17, 129
- Discretization, 108
- Distributed systems, 8, 38
- Disturbance rejection, 295, 334, 362
- Divisor, 256
 - greatest common (gcd), 257
 - left, 257
 - right, 257
- Duality, theorem of, 196
- Eigenvalue, 69
 - assignment of, 287
 - index of, 77, 108, 111
 - multiplicity of, 74, 77
- Eigenvector, 69
 - generalized, 75
- Equivalence, 119
 - algebraic, 119, 178
 - Lyapunov, 143, 178
 - zero-state, 120
- Equivalence transformation, 119, 141, 178
- Fraction, polynomial, 233
 - coprime, 233

- matrix 257
Frequency response, 157, 165
Friction
 Coulomb, 31
 static, 31
 viscous, 31
Fundamental matrix, 136
- gcd, 257
 gcld, 258
 gcrd, 258
Gramian, controllability, 185,
 188, 211
observability, 197, 212
- Hankel matrix, 248
Hankel singular value, 245
Homogeneity, 11, 42
- Impedance, 25
Implementable transfer function,
 341
Impulse, 9
 sifting property, 10
Impulse response, 13, 18, 114,
 139
Impulse response sequence, 42,
 162
Impulse sequence, 41
Initial conditions, 8
Initial relaxedness, 12
Integrator, 116
- Internal model principal, 246,
 336
Inverter, 116
- Jacobian, 20
Jordan block, 74
Jordan-form matrix, 75
Jury test, 166
- Kalman decomposition theorem,
 204
- Laplace expansion, 66, 156
Laplace transform, 15, 84
Leading coefficient, 39
Leverrier algorithm, 100
Linear algebraic equation, 62
Linear simulation (lsim), 113,
 189
- Linear space, 57
 basis of, 58
 change of basis, 58, 67
 dimension, 57
 orthonomal basis, 58
- Linearity, 11, 41
Linearization, 20
Lumped systems, 8, 38
Lyapunov equation, 85, 170,
 291, 302, 311
 discrete, 173, 175
Lyapunov theorem, 170, 174
Lyapunov transformation, 143,
 178

- Magnitude response, 157, 165
 Magnitude scaling, 125
 Markov parameter, 247
 Matrix,
 cofactor of, 66
 companion form, 68, 98
 CT stable, 170
 cyclic, 308
 determinant of, 66
 digonalizable, 71, 75, 89
 DT stable, 174
 function of, 79
 fundamental, 136
 inverse of, 66
 Jordan form, 75
 left inverse of, 255
 minor of, 66
 modal form, 123
 nilpotent, 76
 nonsingular, 66
 norm of, 94
 orthogonal, 62, 66
 polynomial of, 76
 principal minor, 91
 leading, 91
 pseudoinverse of, 255
 rank of, 62
 right inverse of, 255
 singular, 66
 singular value of, 92
 state transition, 137
 symmetric, 88
 trace of, 100
 unit, 55
 McMillan degree, 252
 Minimal polynomial, 77
 Minimal realization, 126, 145, 227, 254
 Minimum energy control, 189
 Minor, 66, 252
 Modal form, 123
 Model matching, 339, 365
 Model reduction, 246
 Monic polynomial, 39, 77

 Nilpotent matrix, 76
 Nominal equation, 295
 Non-real-time processing, 115
 Norm, 59
 Euclidean, 60
 induced, 94
 of matrix, 94
 of vector, 59
 Null vector, 62, 240
 monic, 240
 Nullity, 63

 Observability, 194, 210
 Gramian, 197, 212
 in Jordan form, 206
 index, 198, 212
 matrix, 197, 212
 Observer, *see* State estimator,
 Op-amp circuit, 116, 350
 magnitude scaling, 125
 Orthogonal matrix, 62, 66, 89, 242

- Orthogonality of vectors, 60
Orthonormal, 60
Output equation, 17
- Parameterization, 64
Periodic state equation, 144
Phase response, 157, 165
Pole, 39
 repeated, 39
 simple, 39
Pole placement, 327
Pole-zero cancelation, 340
 unstable, 340, 343
Pole-zero excess inequality, 342
Polynomial, 39
 CT stable, 159, 325
 DT stable, 166
 leading coefficient of, 39
 minimal, 77
 monic, 39, 77
 roots of, 39
Polynomial fraction, 233
Polynomial matrix,
 column degree, 259
 column-degree-coefficient
 matrix,
 260
 column reduced, 260
 echelon form, 267
 left divisor, 258
 left multiple, 258
 right divisor, 258
 right multiple, 258
 row degree, 259
- row-degree-coefficient matrix,
 260
row reduced, 260
unimodular, 258
Positive definite matrix, 90
Positive semidefinite matrix, 90
Primary dependent column, 263
Principal minor, 91
 leading 91
Pseudoinverse, 255
Pseudostate, 230
- QR decomposition, 242, 264, 357
Quadratic form, 88
- Range space, 62
Rank, of matrix, 62, 66
 of product of matrices, 87
Rational function, 38, 233
 biproper, 38
 coprime, 39
 degree, 233
 improper, 38
 proper, 38
 strictly proper, 38
Rational matrix, 40
 biproper, 40
 characteristic polynomial, 252
 degree of, 252
 McMillan degree, 252
 proper, 40
 strictly proper, 40
Reachability, 213

- Realization, 126, 145, 227
 - balanced, 245, 251
 - companion-form, 249
 - controllable-form, 128, 131, 230
 - input-normal, 246
 - minimal, 227, 233
 - observable-form, 129, 149
 - output-normal, 246
 - time-varying, 145
- Real-time processing, 116
- Reference signal, 282, 323
- Regulator problem, 293, 332, 360
- Relaxedness, 12, 14
- Responses,
 - impulse, 13, 18, 114, 162
 - zero-input, 12
 - zero-state, 12
- Resultant, generalized, 263
 - Sylvester, 239
- Robust design, 335
- Robust tracking, 295, 333, 362
- Root, 39
 - repeated, 39
 - simple, 39
- Sampling theorem, 112
- Savings account, 47
- Schmidt orthonormalization, 60
- Separation property, 307
- Servomechanism, 294
- Similar matrices, 69
- Similarity transformation, 69
- Singular value, 92
- Singular value decomposition, 93
- Spring constant, 30
- Stability,
 - asymptotic, 168, 169
 - BIBO, 154, 162
 - in the sense of Lyapunov, 168, 169
 - marginal, 168, 169
 - total, 340
- Stabilization, 299
- State, 7
 - initial, 8
 - pseudo-, 230
 - variable, 8
- State equation, 17, 110
- State estimator, 299, 316
 - asymptotic, 302
 - closed-loop, 300
 - full-dimensional, 299
 - open-loop, 300
 - reduced-dimensional, 303
- State feedback, 283, 307, 314
- State-space (ss) equation, 3, 17, 46
 - computation of, 112
 - non-real-time, 115
 - real-time, 116
- controllable-form, 128, 131, 230, 232, 286
 - CT, 105
 - DT, 46
 - equivalent, 119, 141
 - general solutions, 105, 111, 138, 140

- observable-form, 129, 149, 231, 232
- op-amp circuit
 - implementation, 116
- periodic, 144
- RLC circuits, 21
- time-invariant, 17, 46
- time-varying, 19, 135, 140, 217
- State transition matrix, 137, 140
- Step response, 114
- Step size, 10, 109, 112
- Superposition property, 11
- Sylvester resultant, 239
- Sylvester inequality, 254
- System, 2
 - causal, 7, 40
 - continuous-time (CT), 6, 11, 153
 - discrete-time (DT), 6, 40, 162
 - distributed, 8, 14, 159, 170
 - linear, 11
 - lumped, 8
 - memoryless, 7
 - MIMO, 6
 - nonlinear, 11
 - SISO, 6
 - time-invariant, 8, 72
 - time-varying, 9, 19
- Time domain, 3, 25
- Time index, 6
- Time invariance, 9, 42
- Time shifting, 8, 41
- Total stability, 340
- Trace of matrix, 100
- Transfer function, 15, 26, 42
 - amplitude response, 157, 165
 - discrete, 43
 - frequency response, 157, 165
 - implementable, 341
 - phase response, 157, 165
 - pole of, 39
 - see also* rational function, zero of, 39
 - zero-pole-gain form, 40
- Transfer matrix, 18, 40
 - blocking zero, 40
 - characteristic polynomial, 252
 - see also* rational matrix, transmission zero, 361
- Transform domain, 3, 25
- Transmission zero, 361
- Trancation operator, 49
- Unimodular matrix, 258
- Unit-sampling-period-delay element, 43
- Unit-time-delay element, 8, 14, 16
- Vandermonde determinant, 88
- Vector, 57
 - Euclidean norm of, 60
 - Infinite norm of, 60
 - 1-norm of, 60

- Vandermonde determinant
 (*Cont.*)
nomalization, 60
orthogonal, 60
2-norm of, 60
- Viscous friction coefficient, 30
- Well-posedness, 340
- Zero, 39
 blocking, 40
minimum-phase, 325, 342
nonminimum-phase, 325
transmission, 361
- Zero-input response, 12, 41
- Zero-state response, 12, 41
- z*-transform, 42, 163