MODERN CONTROL ENGINEERING

FIFTH EDITION

Katsuhiko Ogata

Modern Control Engineering

Fifth Edition

Katsuhiko Ogata

Prentice Hall
Boston Columbus Indianapolis New York San Francisco Upper Saddle River
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montreal Toronto
Delhi Mexico City Sao Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

VP/Editorial Director, Engineering/Computer Science: Marcia J. Horton
Assistant/Supervisor: Dolores Mars
Senior Editor: Andrew Gilfillan
Associate Editor: Alice Dworkin
Editorial Assistant: William Opaluch
Director of Marketing: Margaret Waples
Senior Marketing Manager: Tim Galligan
Marketing Assistant: Mack Patterson
Senior Managing Editor: Scott Disanno
Art Editor: Greg Dulles
Senior Operations Supervisor: Alan Fischer
Operations Specialist: Lisa McDowell
Art Director: Kenny Beck
Cover Designer: Carole Anson
Media Editor: Daniel Sandin

Credits and acknowledgments borrowed from other sources and reproduced, with permission, in this textbook appear on appropriate page within text.

Library of Congress Cataloging-in-Publication Data on File

10 9 8 7 6 5 4 3 2 1

Prentice Hall
is an imprint of

PEARSON

www.pearsonhighered.com

ISBN 10: 0-13-615673-8
ISBN 13: 978-0-13-615673-4

Contents

Chapter 3 Mathematical Modeling of Mechanical Systems and Electrical Systems 63

Chapter 4 Mathematical Modeling of Fluid Systems and Thermal Systems 100

Chapter 5 Transient and Steady-State Response Analyses 159

Chapter 6 Control Systems Analysis and Design by the Root-Locus Method 269

Chapter 7 Control Systems Analysis and Design by the Frequency-Response Method 398

Chapter 8 PID Controllers and Modified PID Controllers 567

Chapter 9 Control Systems Analysis in State Space 648

Chapter 10 Control Systems Design in State Space 722

Preface

This book introduces important concepts in the analysis and design of control systems. Readers will find it to be a clear and understandable textbook for control system courses at colleges and universities. It is written for senior electrical, mechanical, aerospace, or chemical engineering students. The reader is expected to have fulfilled the following prerequisites: introductory courses on differential equations, Laplace transforms, vector-matrix analysis, circuit analysis, mechanics, and introductory thermodynamics.

The main revisions made in this edition are as follows:

- The use of MATLAB for obtaining responses of control systems to various inputs has been increased.
- The usefulness of the computational optimization approach with MATLAB has been demonstrated.
- New example problems have been added throughout the book.
- Materials in the previous edition that are of secondary importance have been deleted in order to provide space for more important subjects. Signal flow graphs were dropped from the book. A chapter on Laplace transform was deleted. Instead, Laplace transform tables, and partial-fraction expansion with MATLAB are presented in Appendix A and Appendix B, respectively.
- A short summary of vector-matrix analysis is presented in Appendix C; this will help the reader to find the inverses of n x n matrices that may be involved in the analysis and design of control systems.

This edition of *Modern Control Engineering* is organized into ten chapters. The outline of this book is as follows: Chapter 1 presents an introduction to control systems. Chapter 2 deals with mathematical modeling of control systems. A linearization technique for nonlinear mathematical models is presented in this chapter. Chapter 3 derives mathematical models of mechanical systems and electrical systems. Chapter 4 discusses mathematical modeling of fluid systems (such as liquid-level systems, pneumatic systems, and hydraulic systems) and thermal systems.

Chapter 5 treats transient response and steady-state analyses of control systems. MATLAB is used extensively for obtaining transient response curves. Routh's stability criterion is presented for stability analysis of control systems. Hurwitz stability criterion is also presented.

Chapter 6 discusses the root-locus analysis and design of control systems, including positive feedback systems and conditionally stable systems Plotting root loci with MATLAB is discussed in detail. Design of lead, lag, and lag-lead compensators with the root-locus method is included.

Chapter 7 treats the frequency-response analysis and design of control systems. The Nyquist stability criterion is presented in an easily understandable manner. The Bode diagram approach to the design of lead, lag, and lag-lead compensators is discussed.

Chapter 8 deals with basic and modified PID controllers. Computational approaches for obtaining optimal parameter values for PID controllers are discussed in detail, particularly with respect to satisfying requirements for step-response characteristics.

Chapter 9 treats basic analyses of control systems in state space. Concepts of controllability and observability are discussed in detail.

Chapter 10 deals with control systems design in state space. The discussions include pole placement, state observers, and quadratic optimal control. An introductory discussion of robust control systems is presented at the end of Chapter 10.

The book has been arranged toward facilitating the student's gradual understanding of control theory. Highly mathematical arguments are carefully avoided in the presentation of the materials. Statement proofs are provided whenever they contribute to the understanding of the subject matter presented.

Special effort has been made to provide example problems at strategic points so that the reader will have a clear understanding of the subject matter discussed. In addition, a number of solved problems (A-problems) are provided at the end of each chapter, except Chapter 1. The reader is encouraged to study all such solved problems carefully; this will allow the reader to obtain a deeper understanding of the topics discussed. In addition, many problems (without solutions) are provided at the end of each chapter, except Chapter 1. The unsolved problems (B-problems) may be used as homework or quiz problems.

If this book is used as a text for a semester course (with 56 or so lecture hours), a good portion of the material may be covered by skipping certain subjects. Because of the abundance of example problems and solved problems (A-problems) that might answer many possible questions that the reader might have, this book can also serve as a self-study book for practicing engineers who wish to study basic control theories.

I would like to thank the following reviewers for this edition of the book: Mark Campbell, Cornell University; Henry Sodano, Arizona State University; and Atul G. Kelkar, Iowa State University. Finally, I wish to offer my deep appreciation to Ms. Alice Dworkin, Associate Editor, Mr. Scott Disanno, Senior Managing Editor, and all the people involved in this publishing project, for the speedy yet superb production of this book.

Katsuhiko Ogata

1

Introduction to Control Systems

1–1 INTRODUCTION

Control theories commonly used today are classical control theory (also called conventional control theory), modern control theory, and robust control theory. This book presents comprehensive treatments of the analysis and design of control systems based on the classical control theory and modern control theory. A brief introduction of robust control theory is included in Chapter 10.

Automatic control is essential in any field of engineering and science. Automatic control is an important and integral part of space-vehicle systems, robotic systems, modern manufacturing systems, and any industrial operations involving control of temperature, pressure, humidity, flow, etc. It is desirable that most engineers and scientists are familiar with theory and practice of automatic control.

This book is intended to be a text book on control systems at the senior level at a college or university. All necessary background materials are included in the book. Mathematical background materials related to Laplace transforms and vector-matrix analysis are presented separately in appendixes.

Brief Review of Historical Developments of Control Theories and Practices. The first significant work in automatic control was James Watt's centrifugal governor for the speed control of a steam engine in the eighteenth century. Other significant works in the early stages of development of control theory were due to

Minorsky, Hazen, and Nyquist, among many others. In 1922, Minorsky worked on automatic controllers for steering ships and showed how stability could be determined from the differential equations describing the system. In 1932, Nyquist developed a relatively simple procedure for determining the stability of closed-loop systems on the basis of open-loop response to steady-state sinusoidal inputs. In 1934, Hazen, who introduced the term *servomechanisms* for position control systems, discussed the design of relay servomechanisms capable of closely following a changing input.

During the decade of the 1940s, frequency-response methods (especially the Bode diagram methods due to Bode) made it possible for engineers to design linear closed-loop control systems that satisfied performance requirements. Many industrial control systems in 1940s and 1950s used PID controllers to control pressure, temperature, etc. In the early 1940s Ziegler and Nichols suggested rules for tuning PID controllers, called Ziegler–Nichols tuning rules. From the end of the 1940s to the 1950s, the root-locus method due to Evans was fully developed.

The frequency-response and root-locus methods, which are the core of classical control theory, lead to systems that are stable and satisfy a set of more or less arbitrary performance requirements. Such systems are, in general, acceptable but not optimal in any meaningful sense. Since the late 1950s, the emphasis in control design problems has been shifted from the design of one of many systems that work to the design of one optimal system in some meaningful sense.

As modern plants with many inputs and outputs become more and more complex, the description of a modern control system requires a large number of equations. Classical control theory, which deals only with single-input, single-output systems, becomes powerless for multiple-input, multiple-output systems. Since about 1960, because the availability of digital computers made possible time-domain analysis of complex systems, modern control theory, based on time-domain analysis and synthesis using state variables, has been developed to cope with the increased complexity of modern plants and the stringent requirements on accuracy, weight, and cost in military, space, and industrial applications.

During the years from 1960 to 1980, optimal control of both deterministic and stochastic systems, as well as adaptive and learning control of complex systems, were fully investigated. From 1980s to 1990s, developments in modern control theory were centered around robust control and associated topics.

Modern control theory is based on time-domain analysis of differential equation systems. Modern control theory made the design of control systems simpler because the theory is based on a model of an actual control system. However, the system's stability is sensitive to the error between the actual system and its model. This means that when the designed controller based on a model is applied to the actual system, the system may not be stable. To avoid this situation, we design the control system by first setting up the range of possible errors and then designing the controller in such a way that, if the error of the system stays within the assumed range, the designed control system will stay stable. The design method based on this principle is called robust control theory. This theory incorporates both the frequency-response approach and the time-domain approach. The theory is mathematically very complex.

Because this theory requires mathematical background at the graduate level, inclusion of robust control theory in this book is limited to introductory aspects only. The reader interested in details of robust control theory should take a graduate-level control course at an established college or university.

Definitions. Before we can discuss control systems, some basic terminologies must be defined.

Controlled Variable and Control Signal or Manipulated Variable. The *controlled* variable is the quantity or condition that is measured and controlled. The *control signal* or *manipulated* variable is the quantity or condition that is varied by the controller so as to affect the value of the controlled variable. Normally, the controlled variable is the output of the system. *Control* means measuring the value of the controlled variable of the system and applying the control signal to the system to correct or limit deviation of the measured value from a desired value.

In studying control engineering, we need to define additional terms that are necessary to describe control systems.

Plants. A plant may be a piece of equipment, perhaps just a set of machine parts functioning together, the purpose of which is to perform a particular operation. In this book, we shall call any physical object to be controlled (such as a mechanical device, a heating furnace, a chemical reactor, or a spacecraft) a plant.

Processes. The *Merriam–Webster Dictionary* defines a process to be a natural, progressively continuing operation or development marked by a series of gradual changes that succeed one another in a relatively fixed way and lead toward a particular result or end; or an artificial or voluntary, progressively continuing operation that consists of a series of controlled actions or movements systematically directed toward a particular result or end. In this book we shall call any operation to be controlled a *process.* Examples are chemical, economic, and biological processes.

Systems. A system is a combination of components that act together and perform a certain objective. A system need not be physical. The concept of the system can be applied to abstract, dynamic phenomena such as those encountered in economics. The word system should, therefore, be interpreted to imply physical, biological, economic, and the like, systems.

Disturbances. A disturbance is a signal that tends to adversely affect the value of the output of a system. If a disturbance is generated within the system, it is called *internal,* while an *external* disturbance is generated outside the system and is an input.

Feedback Control. Feedback control refers to an operation that, in the presence of disturbances, tends to reduce the difference between the output of a system and some reference input and does so on the basis of this difference. Here only unpredictable disturbances are so specified, since predictable or known disturbances can always be compensated for within the system.

1–2 EXAMPLES OF CONTROL SYSTEMS

In this section we shall present a few examples of control systems.

Speed Control System. The basic principle of a Watt's speed governor for an engine is illustrated in the schematic diagram of Figure 1–1. The amount of fuel admitted to the engine is adjusted according to the difference between the desired and the actual engine speeds.

The sequence of actions may be stated as follows: The speed governor is adjusted such that, at the desired speed, no pressured oil will flow into either side of the power cylinder. If the actual speed drops below the desired value due to disturbance, then the decrease in the centrifugal force of the speed governor causes the control valve to move downward, supplying more fuel, and the speed of the engine increases until the desired value is reached. On the other hand, if the speed of the engine increases above the desired value, then the increase in the centrifugal force of the governor causes the control valve to move upward. This decreases the supply of fuel, and the speed of the engine decreases until the desired value is reached.

In this speed control system, the plant (controlled system) is the engine and the controlled variable is the speed of the engine. The difference between the desired speed and the actual speed is the error signal. The control signal (the amount of fuel) to be applied to the plant (engine) is the actuating signal. The external input to disturb the controlled variable is the disturbance. An unexpected change in the load is a disturbance.

Temperature Control System. Figure 1–2 shows a schematic diagram of temperature control of an electric furnace. The temperature in the electric furnace is measured by a thermometer, which is an analog device. The analog temperature is converted

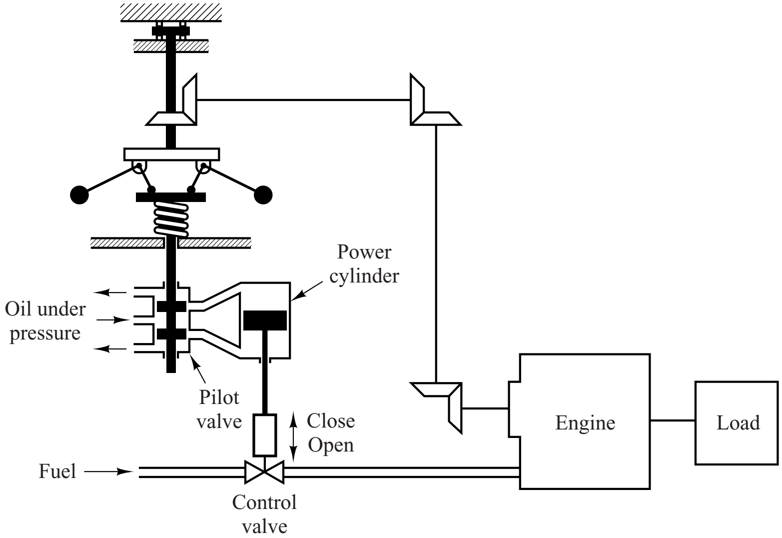


Figure 1–1
Speed control system.

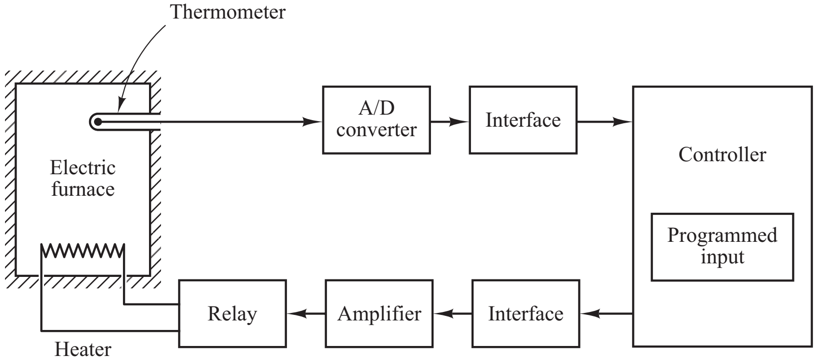


Figure 1–2 Temperature control system.

to a digital temperature by an A/D converter. The digital temperature is fed to a controller through an interface. This digital temperature is compared with the programmed input temperature, and if there is any discrepancy (error), the controller sends out a signal to the heater, through an interface, amplifier, and relay, to bring the furnace temperature to a desired value.

Business Systems. A business system may consist of many groups. Each task assigned to a group will represent a dynamic element of the system. Feedback methods of reporting the accomplishments of each group must be established in such a system for proper operation. The cross-coupling between functional groups must be made a minimum in order to reduce undesirable delay times in the system. The smaller this cross-coupling, the smoother the flow of work signals and materials will be.

A business system is a closed-loop system. A good design will reduce the managerial control required. Note that disturbances in this system are the lack of personnel or materials, interruption of communication, human errors, and the like.

The establishment of a well-founded estimating system based on statistics is mandatory to proper management. It is a well-known fact that the performance of such a system can be improved by the use of lead time, or *anticipation.*

To apply control theory to improve the performance of such a system, we must represent the dynamic characteristic of the component groups of the system by a relatively simple set of equations.

Although it is certainly a difficult problem to derive mathematical representations of the component groups, the application of optimization techniques to business systems significantly improves the performance of the business system.

Consider, as an example, an engineering organizational system that is composed of major groups such as management, research and development, preliminary design, experiments, product design and drafting, fabrication and assembling, and tesing. These groups are interconnected to make up the whole operation.

Such a system may be analyzed by reducing it to the most elementary set of components necessary that can provide the analytical detail required and by representing the dynamic characteristics of each component by a set of simple equations. (The dynamic performance of such a system may be determined from the relation between progressive accomplishment and time.)

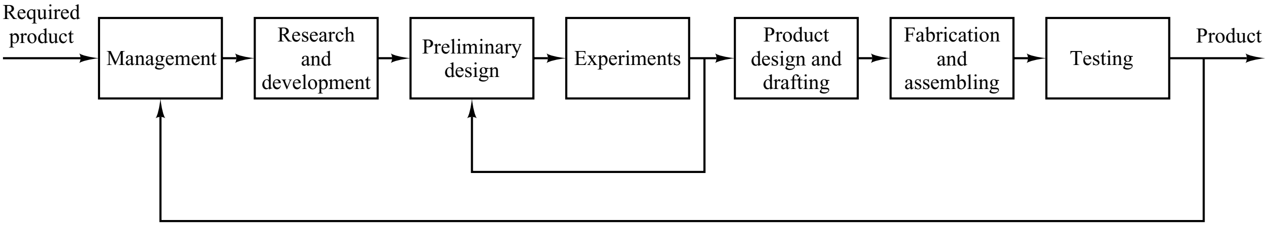


Figure 1–3
Block diagram of an engineering organizational system.

A functional block diagram may be drawn by using blocks to represent the functional activities and interconnecting signal lines to represent the information or product output of the system operation. Figure 1–3 is a possible block diagram for this system.

Robust Control System. The first step in the design of a control system is to obtain a mathematical model of the plant or control object. In reality, any model of a plant we want to control will include an error in the modeling process. That is, the actual plant differs from the model to be used in the design of the control system.

To ensure the controller designed based on a model will work satisfactorily when this controller is used with the actual plant, one reasonable approach is to assume from the start that there is an uncertainty or error between the actual plant and its mathematical model and include such uncertainty or error in the design process of the control system. The control system designed based on this approach is called a robust control system.

Suppose that the actual plant we want to control is $\widetilde{G}(s)$ and the mathematical model of the actual plant is $G(s)$, that is,

$\widetilde{G}(s)$ = actual plant model that has uncertainty $\Delta(s)$

$G(s)$ = nominal plant model to be used for designing the control system

$\widetilde{G}(s)$ and $G(s)$ may be related by a multiplicative factor such as

$$\widetilde{G}(s) = G(s)[1 + \Delta(s)]$$

or an additive factor

$$\widetilde{G}(s) = G(s) + \Delta(s)$$

or in other forms.

Since the exact description of the uncertainty or error $\Delta(s)$ is unknown, we use an estimate of $\Delta(s)$ and use this estimate, $W(s)$, in the design of the controller. $W(s)$ is a scalar transfer function such that

$$\|\Delta(s)\|_\infty < \|W(s)\|_\infty = \max_{0 \le \omega \le \infty} |W(j\omega)|$$

where $\|W(s)\|_\infty$ is the maximum value of $|W(j\omega)|$ for $0 \le \omega \le \infty$ and is called the H infinity norm of $W(s)$.

Using the small gain theorem, the design procedure here boils down to the determination of the controller $K(s)$ such that the inequality

$$\left\| \frac{W(s)}{1 + K(s)G(s)} \right\|_\infty < 1$$

is satisfied, where $G(s)$ is the transfer function of the model used in the design process, $K(s)$ is the transfer function of the controller, and $W(s)$ is the chosen transfer function to approximate $\Delta(s)$. In most practical cases, we must satisfy more than one such inequality that involves $G(s)$, $K(s)$, and $W(s)$'s. For example, to guarantee robust stability and robust performance we may require two inequalities, such as

$$\left\| \frac{W_m(s)K(s)G(s)}{1 + K(s)G(s)} \right\|_\infty < 1 \quad \text{for robust stability}$$

$$\left\| \frac{W_s(s)}{1 + K(s)G(s)} \right\|_\infty < 1 \quad \text{for robust performance}$$

be satisfied. (These inequalities are derived in Section 10–9.) There are many different such inequalities that need to be satisfied in many different robust control systems. (Robust stability means that the controller $K(s)$ guarantees internal stability of all systems that belong to a group of systems that include the system with the actual plant. Robust performance means the specified performance is satisfied in all systems that belong to the group.) In this book all the plants of control systems we discuss are assumed to be known precisely, except the plants we discuss in Section 10–9 where an introductory aspect of robust control theory is presented.

1–3 CLOSED-LOOP CONTROL VERSUS OPEN-LOOP CONTROL

Feedback Control Systems. A system that maintains a prescribed relationship between the output and the reference input by comparing them and using the difference as a means of control is called a *feedback control system.* An example would be a room-temperature control system. By measuring the actual room temperature and comparing it with the reference temperature (desired temperature), the thermostat turns the heating or cooling equipment on or off in such a way as to ensure that the room temperature remains at a comfortable level regardless of outside conditions.

Feedback control systems are not limited to engineering but can be found in various nonengineering fields as well. The human body, for instance, is a highly advanced feedback control system. Both body temperature and blood pressure are kept constant by means of physiological feedback. In fact, feedback performs a vital function: It makes the human body relatively insensitive to external disturbances, thus enabling it to function properly in a changing environment.

Closed-Loop Control Systems. Feedback control systems are often referred to as *closed-loop control* systems. In practice, the terms feedback control and closed-loop control are used interchangeably. In a closed-loop control system the actuating error signal, which is the difference between the input signal and the feedback signal (which may be the output signal itself or a function of the output signal and its derivatives and/or integrals), is fed to the controller so as to reduce the error and bring the output of the system to a desired value. The term closed-loop control always implies the use of feedback control action in order to reduce system error.

Open-Loop Control Systems. Those systems in which the output has no effect on the control action are called *open-loop control systems.* In other words, in an open-loop control system the output is neither measured nor fed back for comparison with the input. One practical example is a washing machine. Soaking, washing, and rinsing in the washer operate on a time basis. The machine does not measure the output signal, that is, the cleanliness of the clothes.

In any open-loop control system the output is not compared with the reference input. Thus, to each reference input there corresponds a fixed operating condition; as a result, the accuracy of the system depends on calibration. In the presence of disturbances, an open-loop control system will not perform the desired task. Open-loop control can be used, in practice, only if the relationship between the input and output is known and if there are neither internal nor external disturbances. Clearly, such systems are not feedback control systems. Note that any control system that operates on a time basis is open loop. For instance, traffic control by means of signals operated on a time basis is another example of open-loop control.

Closed-Loop versus Open-Loop Control Systems. An advantage of the closed-loop control system is the fact that the use of feedback makes the system response relatively insensitive to external disturbances and internal variations in system parameters. It is thus possible to use relatively inaccurate and inexpensive components to obtain the accurate control of a given plant, whereas doing so is impossible in the open-loop case.

From the point of view of stability, the open-loop control system is easier to build because system stability is not a major problem. On the other hand, stability is a major problem in the closed-loop control system, which may tend to overcorrect errors and thereby can cause oscillations of constant or changing amplitude.

It should be emphasized that for systems in which the inputs are known ahead of time and in which there are no disturbances it is advisable to use open-loop control. Closed-loop control systems have advantages only when unpredictable disturbances and/or unpredictable variations in system components are present. Note that the output power rating partially determines the cost, weight, and size of a control system. The number of components used in a closed-loop control system is more than that for a corresponding open-loop control system. Thus, the closed-loop control system is generally higher in cost and power. To decrease the required power of a system, open-loop control may be used where applicable. A proper combination of open-loop and closed-loop controls is usually less expensive and will give satisfactory overall system performance.

Most analyses and designs of control systems presented in this book are concerned with closed-loop control systems. Under certain circumstances (such as where no disturbances exist or the output is hard to measure) open-loop control systems may be desired. Therefore, it is worthwhile to summarize the advantages and disadvantages of using open-loop control systems.

The major advantages of open-loop control systems are as follows:

1. Simple construction and ease of maintenance.
2. Less expensive than a corresponding closed-loop system.
3. There is no stability problem.
4. Convenient when output is hard to measure or measuring the output precisely is economically not feasible. (For example, in the washer system, it would be quite expensive to provide a device to measure the quality of the washer's output, cleanliness of the clothes.)

The major disadvantages of open-loop control systems are as follows:

1. Disturbances and changes in calibration cause errors, and the output may be different from what is desired.
2. To maintain the required quality in the output, recalibration is necessary from time to time.

1–4 DESIGN AND COMPENSATION OF CONTROL SYSTEMS

This book discusses basic aspects of the design and compensation of control systems. Compensation is the modification of the system dynamics to satisfy the given specifications. The approaches to control system design and compensation used in this book are the root-locus approach, frequency-response approach, and the state-space approach. Such control systems design and compensation will be presented in Chapters 6, 7, 9 and 10. The PID-based compensational approach to control systems design is given in Chapter 8.

In the actual design of a control system, whether to use an electronic, pneumatic, or hydraulic compensator is a matter that must be decided partially based on the nature of the controlled plant. For example, if the controlled plant involves flammable fluid, then we have to choose pneumatic components (both a compensator and an actuator) to avoid the possibility of sparks. If, however, no fire hazard exists, then electronic compensators are most commonly used. (In fact, we often transform nonelectrical signals into electrical signals because of the simplicity of transmission, increased accuracy, increased reliability, ease of compensation, and the like.)

Performance Specifications. Control systems are designed to perform specific tasks. The requirements imposed on the control system are usually spelled out as performance specifications. The specifications may be given in terms of transient response requirements (such as the maximum overshoot and settling time in step response) and of steady-state requirements (such as steady-state error in following ramp input) or may be given in frequency-response terms. The specifications of a control system must be given before the design process begins.

For routine design problems, the performance specifications (which relate to accuracy, relative stability, and speed of response) may be given in terms of precise numerical values. In other cases they may be given partially in terms of precise numerical values and partially in terms of qualitative statements. In the latter case the specifications may have to be modified during the course of design, since the given specifications may never be satisfied (because of conflicting requirements) or may lead to a very expensive system.

Generally, the performance specifications should not be more stringent than necessary to perform the given task. If the accuracy at steady-state operation is of prime importance in a given control system, then we should not require unnecessarily rigid performance specifications on the transient response, since such specifications will require expensive components. Remember that the most important part of control system design is to state the performance specifications precisely so that they will yield an optimal control system for the given purpose.

System Compensation. Setting the gain is the first step in adjusting the system for satisfactory performance. In many practical cases, however, the adjustment of the gain alone may not provide sufficient alteration of the system behavior to meet the given specifications. As is frequently the case, increasing the gain value will improve the steady-state behavior but will result in poor stability or even instability. It is then necessary to redesign the system (by modifying the structure or by incorporating additional devices or components) to alter the overall behavior so that the system will behave as desired. Such a redesign or addition of a suitable device is called *compensation.* A device inserted into the system for the purpose of satisfying the specifications is called a *compensator.* The compensator compensates for deficient performance of the original system.

Design Procedures. In the process of designing a control system, we set up a mathematical model of the control system and adjust the parameters of a compensator. The most time-consuming part of the work is the checking of the system performance by analysis with each adjustment of the parameters. The designer should use MATLAB or other available computer package to avoid much of the numerical drudgery necessary for this checking.

Once a satisfactory mathematical model has been obtained, the designer must construct a prototype and test the open-loop system. If absolute stability of the closed loop is assured, the designer closes the loop and tests the performance of the resulting closed-loop system. Because of the neglected loading effects among the components, nonlinearities, distributed parameters, and so on, which were not taken into consideration in the original design work, the actual performance of the prototype system will probably differ from the theoretical predictions. Thus the first design may not satisfy all the requirements on performance. The designer must adjust system parameters and make changes in the prototype until the system meets the specificications. In doing this, he or she must analyze each trial, and the results of the analysis must be incorporated into the next trial. The designer must see that the final system meets the performance apecifications and, at the same time, is reliable and economical.

1–5 OUTLINE OF THE BOOK

This text is organized into 10 chapters. The outline of each chapter may be summarized as follows:

Chapter 1 presents an introduction to this book.

Chapter 2 deals with mathematical modeling of control systems that are described by linear differential equations. Specifically, transfer function expressions of differential equation systems are derived. Also, state-space expressions of differential equation systems are derived. MATLAB is used to transform mathematical models from transfer functions to state-space equations and vice versa. This book treats linear systems in detail. If the mathematical model of any system is nonlinear, it needs to be linearized before applying theories presented in this book. A technique to linearize nonlinear mathematical models is presented in this chapter.

Chapter 3 derives mathematical models of various mechanical and electrical systems that appear frequently in control systems.

Chapter 4 discusses various fluid systems and thermal systems, that appear in control systems. Fluid systems here include liquid-level systems, pneumatic systems, and hydraulic systems. Thermal systems such as temperature control systems are also discussed here. Control engineers must be familiar with all of these systems discussed in this chapter.

Chapter 5 presents transient and steady-state response analyses of control systems defined in terms of transfer functions. MATLAB approach to obtain transient and steady-state response analyses is presented in detail. MATLAB approach to obtain three-dimensional plots is also presented. Stability analysis based on Routh's stability criterion is included in this chapter and the Hurwitz stability criterion is briefly discussed.

Chapter 6 treats the root-locus method of analysis and design of control systems. It is a graphical method for determining the locations of all closed-loop poles from the knowledge of the locations of the open-loop poles and zeros of a closed-loop system as a parameter (usually the gain) is varied from zero to infinity. This method was developed by W. R. Evans around 1950. These days MATLAB can produce root-locus plots easily and quickly. This chapter presents both a manual approach and a MATLAB approach to generate root-locus plots. Details of the design of control systems using lead compensators, lag compensators, are lag–lead compensators are presented in this chapter.

Chapter 7 presents the frequency-response method of analysis and design of control systems. This is the oldest method of control systems analysis and design and was developed during 1940–1950 by Nyquist, Bode, Nichols, Hazen, among others. This chapter presents details of the frequency-response approach to control systems design using lead compensation technique, lag compensation technique, and lag–lead compensation technique. The frequency-response method was the most frequently used analysis and design method until the state-space method became popular. However, since H-infinity control for designing robust control systems has become popular, frequency response is gaining popularity again.

Chapter 8 discusses PID controllers and modified ones such as multidegrees-of-freedom PID controllers. The PID controller has three parameters; proportional gain, integral gain, and derivative gain. In industrial control systems more than half of the controllers used have been PID controllers. The performance of PID controllers depends on the relative magnitudes of those three parameters. Determination of the relative magnitudes of the three parameters is called tuning of PID controllers.

Ziegler and Nichols proposed so-called "Ziegler–Nichols tuning rules" as early as 1942. Since then numerous tuning rules have been proposed. These days manufacturers of PID controllers have their own tuning rules. In this chapter we present a computer optimization approach using MATLAB to determine the three parameters to satisfy given transient response characteristics. The approach can be expanded to determine the three parameters to satisfy any specific given characteristics.

Chapter 9 presents basic analysis of state-space equations. Concepts of controllability and observability, most important concepts in modern control theory, due to Kalman are discussed in full. In this chapter, solutions of state-space equations are derived in detail.

Chapter 10 discusses state-space designs of control systems. This chapter first deals with pole placement problems and state observers. In control engineering, it is frequently desirable to set up a meaningful performance index and try to minimize it (or maximize it, as the case may be). If the performance index selected has a clear physical meaning, then this approach is quite useful to determine the optimal control variable. This chapter discusses the quadratic optimal regulator problem where we use a performance index which is an integral of a quadratic function of the state variables and the control variable. The integral is performed from $t = 0$ to $t = \infty$. This chapter concludes with a brief discussion of robust control systems.

2

Mathematical Modeling of Control Systems

2–1 INTRODUCTION

In studying control systems the reader must be able to model dynamic systems in mathematical terms and analyze their dynamic characteristics. A mathematical model of a dynamic system is defined as a set of equations that represents the dynamics of the system accurately, or at least fairly well. Note that a mathematical model is not unique to a given system. A system may be represented in many different ways and, therefore, may have many mathematical models, depending on one's perspective.

The dynamics of many systems, whether they are mechanical, electrical, thermal, economic, biological, and so on, may be described in terms of differential equations. Such differential equations may be obtained by using physical laws governing a particular system—for example, Newton's laws for mechanical systems and Kirchhoff's laws for electrical systems. We must always keep in mind that deriving reasonable mathematical models is the most important part of the entire analysis of control systems.

Throughout this book we assume that the principle of causality applies to the systems considered. This means that the current output of the system (the output at time $t = 0$) depends on the past input (the input for $t < 0$) but does not depend on the future input (the input for $t > 0$).

Mathematical Models. Mathematical models may assume many different forms. Depending on the particular system and the particular circumstances, one mathematical model may be better suited than other models. For example, in optimal control problems, it is advantageous to use state-space representations. On the other hand, for the transient-response or frequency-response analysis of single-input, single-output, linear, time-invariant systems, the transfer-function representation may be more convenient than any other. Once a mathematical model of a system is obtained, various analytical and computer tools can be used for analysis and synthesis purposes.

Simplicity Versus Accuracy. In obtaining a mathematical model, we must make a compromise between the simplicity of the model and the accuracy of the results of the analysis. In deriving a reasonably simplified mathematical model, we frequently find it necessary to ignore certain inherent physical properties of the system. In particular, if a linear lumped-parameter mathematical model (that is, one employing ordinary differential equations) is desired, it is always necessary to ignore certain nonlinearities and distributed parameters that may be present in the physical system. If the effects that these ignored properties have on the response are small, good agreement will be obtained between the results of the analysis of a mathematical model and the results of the experimental study of the physical system.

In general, in solving a new problem, it is desirable to build a simplified model so that we can get a general feeling for the solution. A more complete mathematical model may then be built and used for a more accurate analysis.

We must be well aware that a linear lumped-parameter model, which may be valid in low-frequency operations, may not be valid at sufficiently high frequencies, since the neglected property of distributed parameters may become an important factor in the dynamic behavior of the system. For example, the mass of a spring may be neglected in low-frequency operations, but it becomes an important property of the system at high frequencies. (For the case where a mathematical model involves considerable errors, robust control theory may be applied. Robust control theory is presented in Chapter 10.)

Linear Systems. A system is called linear if the principle of superposition applies. The principle of superposition states that the response produced by the simultaneous application of two different forcing functions is the sum of the two individual responses. Hence, for the linear system, the response to several inputs can be calculated by treating one input at a time and adding the results. It is this principle that allows one to build up complicated solutions to the linear differential equation from simple solutions.

In an experimental investigation of a dynamic system, if cause and effect are proportional, thus implying that the principle of superposition holds, then the system can be considered linear.

Linear Time-Invariant Systems and Linear Time-Varying Systems. A differential equation is linear if the coefficients are constants or functions only of the independent variable. Dynamic systems that are composed of linear time-invariant lumped-parameter components may be described by linear time-invariant differential equations—that is, constant-coefficient differential equations. Such systems are called *linear time-invariant* (or *linear constant-coefficient*) systems. Systems that are represented by differential equations whose coefficients are functions of time are called *linear time-varying* systems. An example of a time-varying control system is a spacecraft control system. (The mass of a spacecraft changes due to fuel consumption.)

Outline of the Chapter. Section 2–1 has presented an introduction to the mathematical modeling of dynamic systems. Section 2–2 presents the transfer function and impulse-response function. Section 2–3 introduces automatic control systems and Section 2–4 discusses concepts of modeling in state space. Section 2–5 presents state-space representation of dynamic systems. Section 2–6 discusses transformation of mathematical models with MATLAB. Finally, Section 2–7 discusses linearization of nonlinear mathematical models.

2–2 TRANSFER FUNCTION AND IMPULSE-RESPONSE FUNCTION

In control theory, functions called transfer functions are commonly used to characterize the input-output relationships of components or systems that can be described by linear, time-invariant, differential equations. We begin by defining the transfer function and follow with a derivation of the transfer function of a differential equation system. Then we discuss the impulse-response function.

Transfer Function. The *transfer function* of a linear, time-invariant, differential equation system is defined as the ratio of the Laplace transform of the output (response function) to the Laplace transform of the input (driving function) under the assumption that all initial conditions are zero.

Consider the linear time-invariant system defined by the following differential equation:

$$\begin{aligned} a_0 \overset{(n)}{y} + a_1 \overset{(n-1)}{y} + \cdots + a_{n-1}\dot{y} + a_n y \\ = b_0 \overset{(m)}{x} + b_1 \overset{(m-1)}{x} + \cdots + b_{m-1}\dot{x} + b_m x \qquad (n \geq m) \end{aligned}$$

where y is the output of the system and x is the input. The transfer function of this system is the ratio of the Laplace transformed output to the Laplace transformed input when all initial conditions are zero, or

$$\begin{aligned} \text{Transfer function} = G(s) &= \left.\frac{\mathscr{L}[\text{output}]}{\mathscr{L}[\text{input}]}\right|_{\text{zero initial conditions}} \\ &= \frac{Y(s)}{X(s)} = \frac{b_0 s^m + b_1 s^{m-1} + \cdots + b_{m-1}s + b_m}{a_0 s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n} \end{aligned}$$

By using the concept of transfer function, it is possible to represent system dynamics by algebraic equations in s. If the highest power of s in the denominator of the transfer function is equal to n, the system is called an *nth-order system*.

Comments on Transfer Function. The applicability of the concept of the transfer function is limited to linear, time-invariant, differential equation systems. The transfer function approach, however, is extensively used in the analysis and design of such systems. In what follows, we shall list important comments concerning the transfer function. (Note that a system referred to in the list is one described by a linear, time-invariant, differential equation.)

1. The transfer function of a system is a mathematical model in that it is an operational method of expressing the differential equation that relates the output variable to the input variable.
2. The transfer function is a property of a system itself, independent of the magnitude and nature of the input or driving function.
3. The transfer function includes the units necessary to relate the input to the output; however, it does not provide any information concerning the physical structure of the system. (The transfer functions of many physically different systems can be identical.)
4. If the transfer function of a system is known, the output or response can be studied for various forms of inputs with a view toward understanding the nature of the system.
5. If the transfer function of a system is unknown, it may be established experimentally by introducing known inputs and studying the output of the system. Once established, a transfer function gives a full description of the dynamic characteristics of the system, as distinct from its physical description.

Convolution Integral. For a linear, time-invariant system the transfer function $G(s)$ is

$$G(s) = \frac{Y(s)}{X(s)}$$

where $X(s)$ is the Laplace transform of the input to the system and $Y(s)$ is the Laplace transform of the output of the system, where we assume that all initial conditions involved are zero. It follows that the output $Y(s)$ can be written as the product of $G(s)$ and $X(s)$, or

$$Y(s) = G(s)X(s) \tag{2–1}$$

Note that multiplication in the complex domain is equivalent to convolution in the time domain (see Appendix A), so the inverse Laplace transform of Equation (2–1) is given by the following convolution integral:

$$\begin{aligned} y(t) &= \int_0^t x(\tau)g(t - \tau)\,d\tau \\ &= \int_0^t g(\tau)x(t - \tau)\,d\tau \end{aligned}$$

where both $g(t)$ and $x(t)$ are 0 for $t < 0$.

Impulse-Response Function. Consider the output (response) of a linear time-invariant system to a unit-impulse input when the initial conditions are zero. Since the Laplace transform of the unit-impulse function is unity, the Laplace transform of the output of the system is

$$Y(s) = G(s) \tag{2–2}$$

The inverse Laplace transform of the output given by Equation (2–2) gives the impulse response of the system. The inverse Laplace transform of $G(s)$, or

$$\mathscr{L}^{-1}\left[G(s)\right] = g(t)$$

is called the impulse-response function. This function $g(t)$ is also called the weighting function of the system.

The impulse-response function $g(t)$ is thus the response of a linear time-invariant system to a unit-impulse input when the initial conditions are zero. The Laplace transform of this function gives the transfer function. Therefore, the transfer function and impulse-response function of a linear, time-invariant system contain the same information about the system dynamics. It is hence possible to obtain complete information about the dynamic characteristics of the system by exciting it with an impulse input and measuring the response. (In practice, a pulse input with a very short duration compared with the significant time constants of the system can be considered an impulse.)

2–3 AUTOMATIC CONTROL SYSTEMS

A control system may consist of a number of components. To show the functions performed by each component, in control engineering, we commonly use a diagram called the *block diagram*. This section first explains what a block diagram is. Next, it discusses introductory aspects of automatic control systems, including various control actions. Then, it presents a method for obtaining block diagrams for physical systems, and, finally, discusses techniques to simplify such diagrams.

Block Diagrams. A *block diagram* of a system is a pictorial representation of the functions performed by each component and of the flow of signals. Such a diagram depicts the interrelationships that exist among the various components. Differing from a purely abstract mathematical representation, a block diagram has the advantage of indicating more realistically the signal flows of the actual system.

In a block diagram all system variables are linked to each other through functional blocks. The *functional* block or simply *block* is a symbol for the mathematical operation on the input signal to the block that produces the output. The transfer functions of the components are usually entered in the corresponding blocks, which are connected by arrows to indicate the direction of the flow of signals. Note that the signal can pass only in the direction of the arrows. Thus a block diagram of a control system explicitly shows a unilateral property.

Figure 2–1 shows an element of the block diagram. The arrowhead pointing toward the block indicates the input, and the arrowhead leading away from the block represents the output. Such arrows are referred to as *signals*.

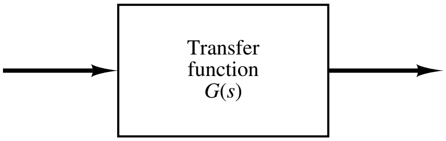


Figure 2–1 Element of a block diagram.

Note that the dimension of the output signal from the block is the dimension of the input signal multiplied by the dimension of the transfer function in the block.

The advantages of the block diagram representation of a system are that it is easy to form the overall block diagram for the entire system by merely connecting the blocks of the components according to the signal flow and that it is possible to evaluate the contribution of each component to the overall performance of the system.

In general, the functional operation of the system can be visualized more readily by examining the block diagram than by examining the physical system itself. A block diagram contains information concerning dynamic behavior, but it does not include any information on the physical construction of the system. Consequently, many dissimilar and unrelated systems can be represented by the same block diagram.

It should be noted that in a block diagram the main source of energy is not explicitly shown and that the block diagram of a given system is not unique. A number of different block diagrams can be drawn for a system, depending on the point of view of the analysis.

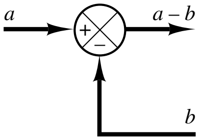


Figure 2–2
Summing point.

Summing Point. Referring to Figure 2–2, a circle with a cross is the symbol that indicates a summing operation. The plus or minus sign at each arrowhead indicates whether that signal is to be added or subtracted. It is important that the quantities being added or subtracted have the same dimensions and the same units.

Branch Point. A *branch point* is a point from which the signal from a block goes concurrently to other blocks or summing points.

Block Diagram of a Closed-Loop System. Figure 2–3 shows an example of a block diagram of a closed-loop system. The output $C(s)$ is fed back to the summing point, where it is compared with the reference input $R(s)$. The closed-loop nature of the system is clearly indicated by the figure. The output of the block, $C(s)$ in this case, is obtained by multiplying the transfer function $G(s)$ by the input to the block, $E(s)$. Any linear control system may be represented by a block diagram consisting of blocks, summing points, and branch points.

When the output is fed back to the summing point for comparison with the input, it is necessary to convert the form of the output signal to that of the input signal. For example, in a temperature control system, the output signal is usually the controlled temperature. The output signal, which has the dimension of temperature, must be converted to a force or position or voltage before it can be compared with the input signal. This conversion is accomplished by the feedback element whose transfer function is $H(s)$, as shown in Figure 2–4. The role of the feedback element is to modify the output before it is compared with the input. (In most cases the feedback element is a sensor that measures

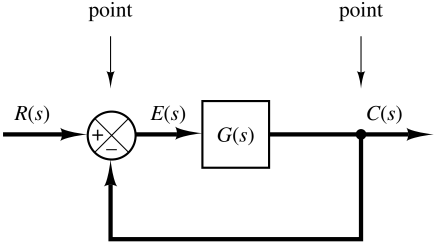


Figure 2–3
Block diagram of a closed-loop system.

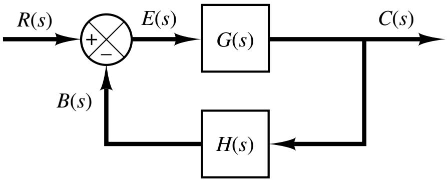


Figure 2–4
Closed-loop system.

the output of the plant. The output of the sensor is compared with the system input, and the actuating error signal is generated.) In the present example, the feedback signal that is fed back to the summing point for comparison with the input is $B(s) = H(s)C(s)$.

Open-Loop Transfer Function and Feedforward Transfer Function. Referring to Figure 2–4, the ratio of the feedback signal $B(s)$ to the actuating error signal $E(s)$ is called the *open-loop transfer function*. That is,

$$\text{Open-loop transfer function} = \frac{B(s)}{E(s)} = G(s)H(s)$$

The ratio of the output $C(s)$ to the actuating error signal $E(s)$ is called the *feedforward transfer function*, so that

$$\text{Feedforward transfer function} = \frac{C(s)}{E(s)} = G(s)$$

If the feedback transfer function $H(s)$ is unity, then the open-loop transfer function and the feedforward transfer function are the same.

Closed-Loop Transfer Function. For the system shown in Figure 2–4, the output $C(s)$ and input $R(s)$ are related as follows: since

$$\begin{aligned} C(s) &= G(s)E(s) \\ E(s) &= R(s) - B(s) \\ &= R(s) - H(s)C(s) \end{aligned}$$

eliminating $E(s)$ from these equations gives

$$C(s) = G(s)\left[R(s) - H(s)C(s)\right]$$

or

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)H(s)} \tag{2–3}$$

The transfer function relating $C(s)$ to $R(s)$ is called the *closed-loop transfer function*. It relates the closed-loop system dynamics to the dynamics of the feedforward elements and feedback elements.

From Equation (2–3), $C(s)$ is given by

$$C(s) = \frac{G(s)}{1 + G(s)H(s)} R(s)$$

Thus the output of the closed-loop system clearly depends on both the closed-loop transfer function and the nature of the input.

Obtaining Cascaded, Parallel, and Feedback (Closed-Loop) Transfer Functions with MATLAB. In control-systems analysis, we frequently need to calculate the cascaded transfer functions, parallel-connected transfer functions, and feedback-connected (closed-loop) transfer functions. MATLAB has convenient commands to obtain the cascaded, parallel, and feedback (closed-loop) transfer functions.

Suppose that there are two components $G_1(s)$ and $G_2(s)$ connected differently as shown in Figure 2–5 (a), (b), and (c), where

$$G_1(s) = \frac{\text{num1}}{\text{den1}}, \qquad G_2(s) = \frac{\text{num2}}{\text{den2}}$$

To obtain the transfer functions of the cascaded system, parallel system, or feedback (closed-loop) system, the following commands may be used:

```
[num, den] = series(num1,den1,num2,den2)
[num, den] = parallel(num1,den1,num2,den2)
[num, den] = feedback(num1,den1,num2,den2)
```

As an example, consider the case where

$$G_1(s) = \frac{10}{s^2 + 2s + 10} = \frac{\text{num1}}{\text{den1}}, \qquad G_2(s) = \frac{5}{s + 5} = \frac{\text{num2}}{\text{den2}}$$

MATLAB Program 2–1 gives $C(s)/R(s) = \text{num/den}$ for each arrangement of $G_1(s)$ and $G_2(s)$. Note that the command

```
printsys(num,den)
```

displays the num/den $\left[\text{that is, the transfer function } C(s)/R(s)\right]$ of the system considered.

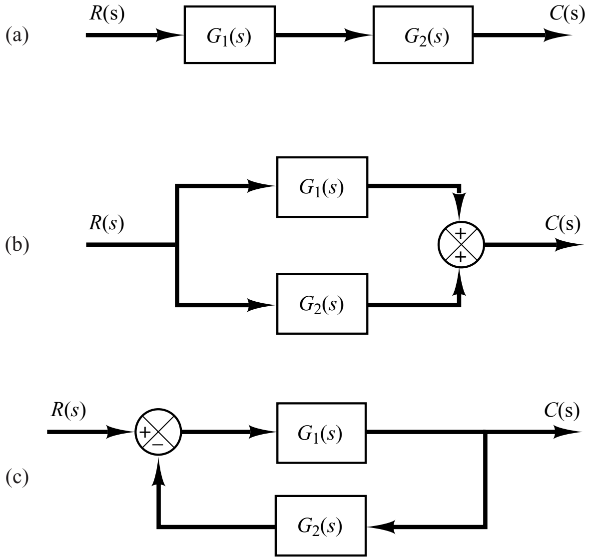


Figure 2–5
(a) Cascaded system;
(b) parallel system;
(c) feedback (closed-loop) system.

MATLAB Program 2–1

```
num1 = [10];
den1 = [1  2  10];
num2 = [5];
den2 = [1  5];
[num, den] = series(num1,den1,num2,den2);
printsys(num,den)

num/den =

             50
   ---------------------
   s^3 + 7s^2 + 20s + 50

[num, den] = parallel(num1,den1,num2,den2);
printsys(num,den)

num/den =

     5s^2 + 20s + 100
   ---------------------
   s^3 + 7s^2 + 20s + 50

[num, den] = feedback(num1,den1,num2,den2);
printsys(num,den)

num/den =

          10s + 50
   ----------------------
   s^3 + 7s^2 + 20s + 100
```

Automatic Controllers. An automatic controller compares the actual value of the plant output with the reference input (desired value), determines the deviation, and produces a control signal that will reduce the deviation to zero or to a small value. The manner in which the automatic controller produces the control signal is called the *control action*. Figure 2–6 is a block diagram of an industrial control system, which

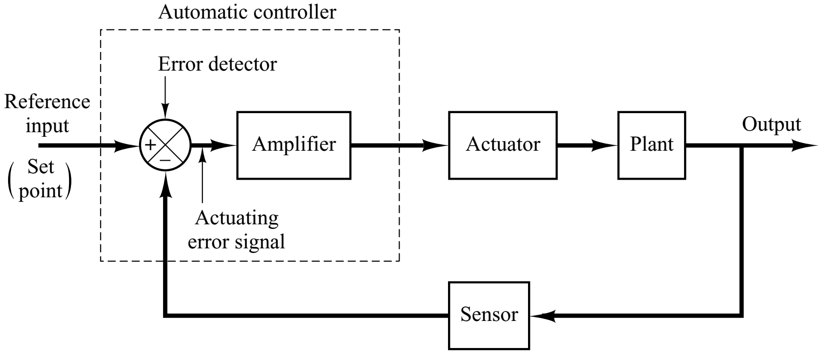


Figure 2–6
Block diagram of an industrial control system, which consists of an automatic controller, an actuator, a plant, and a sensor (measuring element).

consists of an automatic controller, an actuator, a plant, and a sensor (measuring element). The controller detects the actuating error signal, which is usually at a very low power level, and amplifies it to a sufficiently high level. The output of an automatic controller is fed to an actuator, such as an electric motor, a hydraulic motor, or a pneumatic motor or valve. (The actuator is a power device that produces the input to the plant according to the control signal so that the output signal will approach the reference input signal.)

The sensor or measuring element is a device that converts the output variable into another suitable variable, such as a displacement, pressure, voltage, etc., that can be used to compare the output to the reference input signal. This element is in the feedback path of the closed-loop system. The set point of the controller must be converted to a reference input with the same units as the feedback signal from the sensor or measuring element.

Classifications of Industrial Controllers. Most industrial controllers may be classified according to their control actions as:

1. Two-position or on–off controllers
2. Proportional controllers
3. Integral controllers
4. Proportional-plus-integral controllers
5. Proportional-plus-derivative controllers
6. Proportional-plus-integral-plus-derivative controllers

Most industrial controllers use electricity or pressurized fluid such as oil or air as power sources. Consequently, controllers may also be classified according to the kind of power employed in the operation, such as pneumatic controllers, hydraulic controllers, or electronic controllers. What kind of controller to use must be decided based on the nature of the plant and the operating conditions, including such considerations as safety, cost, availability, reliability, accuracy, weight, and size.

Two-Position or On–Off Control Action. In a two-position control system, the actuating element has only two fixed positions, which are, in many cases, simply on and off. Two-position or on–off control is relatively simple and inexpensive and, for this reason, is very widely used in both industrial and domestic control systems.

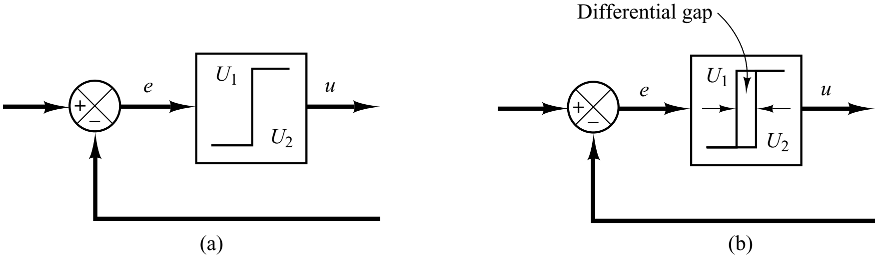
Let the output signal from the controller be $u(t)$ and the actuating error signal be $e(t)$. In two-position control, the signal $u(t)$ remains at either a maximum or minimum value, depending on whether the actuating error signal is positive or negative, so that

$$\begin{aligned} u(t) &= U_1, \qquad \text{for } e(t) > 0 \\ &= U_2, \qquad \text{for } e(t) < 0 \end{aligned}$$

where U_1 and U_2 are constants. The minimum value U_2 is usually either zero or $-U_1$. Two-position controllers are generally electrical devices, and an electric solenoid-operated valve is widely used in such controllers. Pneumatic proportional controllers with very high gains act as two-position controllers and are sometimes called pneumatic two-position controllers.

Figures 2–7(a) and (b) show the block diagrams for two-position or on–off controllers. The range through which the actuating error signal must move before the switching occurs

Figure 2–7
(a) Block diagram of an on–off controller; (b) block diagram of an on–off controller with differential gap.



is called the *differential gap*. A differential gap is indicated in Figure 2–7(b). Such a differential gap causes the controller output $u(t)$ to maintain its present value until the actuating error signal has moved slightly beyond the zero value. In some cases, the differential gap is a result of unintentional friction and lost motion; however, quite often it is intentionally provided in order to prevent too-frequent operation of the on–off mechanism.

Consider the liquid-level control system shown in Figure 2–8(a), where the electromagnetic valve shown in Figure 2–8(b) is used for controlling the inflow rate. This valve is either open or closed. With this two-position control, the water inflow rate is either a positive constant or zero. As shown in Figure 2–9, the output signal continuously moves between the two limits required to cause the actuating element to move from one fixed position to the other. Notice that the output curve follows one of two exponential curves, one corresponding to the filling curve and the other to the emptying curve. Such output oscillation between two limits is a typical response characteristic of a system under two-position control.

Figure 2–8
(a) Liquid-level control system; (b) electromagnetic valve.

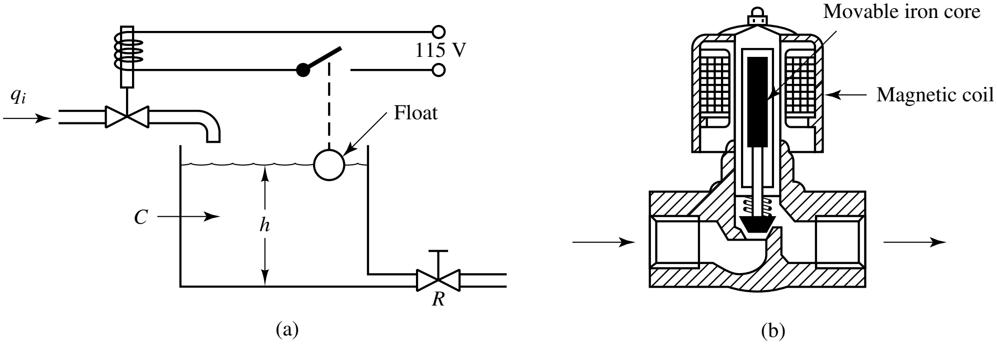
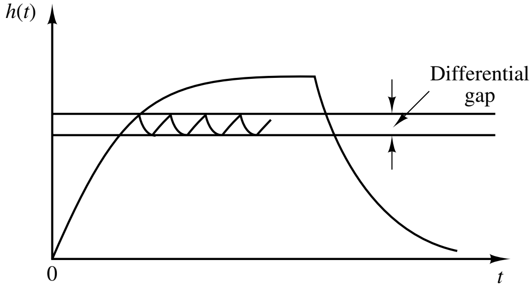


Figure 2–9
Level $h(t)$-versus-t curve for the system shown in Figure 2–8(a).



From Figure 2–9, we notice that the amplitude of the output oscillation can be reduced by decreasing the differential gap. The decrease in the differential gap, however, increases the number of on–off switchings per minute and reduces the useful life of the component. The magnitude of the differential gap must be determined from such considerations as the accuracy required and the life of the component.

Proportional Control Action. For a controller with proportional control action, the relationship between the output of the controller $u(t)$ and the actuating error signal $e(t)$ is

$$u(t) = K_p e(t)$$

or, in Laplace-transformed quantities,

$$\frac{U(s)}{E(s)} = K_p$$

where K_p is termed the proportional gain.

Whatever the actual mechanism may be and whatever the form of the operating power, the proportional controller is essentially an amplifier with an adjustable gain.

Integral Control Action. In a controller with integral control action, the value of the controller output $u(t)$ is changed at a rate proportional to the actuating error signal $e(t)$. That is,

$$\frac{du(t)}{dt} = K_i e(t)$$

or

$$u(t) = K_i \int_0^t e(t)\,dt$$

where K_i is an adjustable constant. The transfer function of the integral controller is

$$\frac{U(s)}{E(s)} = \frac{K_i}{s}$$

Proportional-Plus-Integral Control Action. The control action of a proportional-plus-integral controller is defined by

$$u(t) = K_p e(t) + \frac{K_p}{T_i} \int_0^t e(t)\,dt$$

or the transfer function of the controller is

$$\frac{U(s)}{E(s)} = K_p\left(1 + \frac{1}{T_i s}\right)$$

where T_i is called the *integral time*.

Proportional-Plus-Derivative Control Action. The control action of a proportional-plus-derivative controller is defined by

$$u(t) = K_p e(t) + K_p T_d \frac{de(t)}{dt}$$

and the transfer function is

$$\frac{U(s)}{E(s)} = K_p(1 + T_d s)$$

where T_d is called the *derivative time*.

Proportional-Plus-Integral-Plus-Derivative Control Action. The combination of proportional control action, integral control action, and derivative control action is termed proportional-plus-integral-plus-derivative control action. It has the advantages of each of the three individual control actions. The equation of a controller with this combined action is given by

$$u(t) = K_p e(t) + \frac{K_p}{T_i}\int_0^t e(t)\,dt + K_p T_d \frac{de(t)}{dt}$$

or the transfer function is

$$\frac{U(s)}{E(s)} = K_p\left(1 + \frac{1}{T_i s} + T_d s\right)$$

where K_p is the proportional gain, T_i is the integral time, and T_d is the derivative time. The block diagram of a proportional-plus-integral-plus-derivative controller is shown in Figure 2–10.

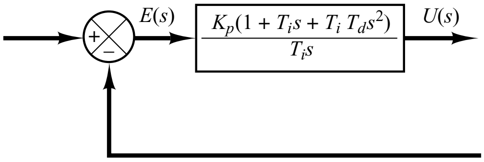


Figure 2–10
Block diagram of a proportional-plus-integral-plus-derivative controller.

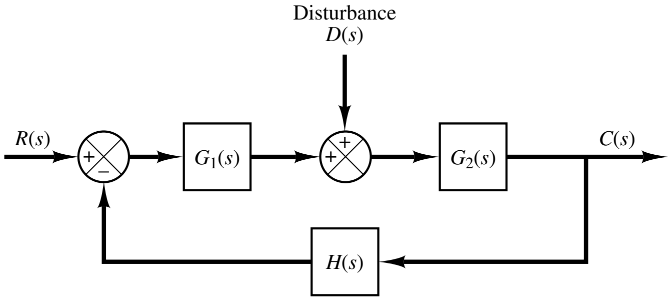


Figure 2–11
Closed-loop system subjected to a disturbance.

Closed-Loop System Subjected to a Disturbance. Figure 2–11 shows a closed-loop system subjected to a disturbance. When two inputs (the reference input and disturbance) are present in a linear time-invariant system, each input can be treated independently of the other; and the outputs corresponding to each input alone can be added to give the complete output. The way each input is introduced into the system is shown at the summing point by either a plus or minus sign.

Consider the system shown in Figure 2–11. In examining the effect of the disturbance $D(s)$, we may assume that the reference input is zero; we may then calculate the response $C_D(s)$ to the disturbance only. This response can be found from

$$\frac{C_D(s)}{D(s)} = \frac{G_2(s)}{1 + G_1(s)G_2(s)H(s)}$$

On the other hand, in considering the response to the reference input $R(s)$, we may assume that the disturbance is zero. Then the response $C_R(s)$ to the reference input $R(s)$ can be obtained from

$$\frac{C_R(s)}{R(s)} = \frac{G_1(s)G_2(s)}{1 + G_1(s)G_2(s)H(s)}$$

The response to the simultaneous application of the reference input and disturbance can be obtained by adding the two individual responses. In other words, the response $C(s)$ due to the simultaneous application of the reference input $R(s)$ and disturbance $D(s)$ is given by

$$\begin{aligned} C(s) &= C_R(s) + C_D(s) \\ &= \frac{G_2(s)}{1 + G_1(s)G_2(s)H(s)}\left[G_1(s)R(s) + D(s)\right] \end{aligned}$$

Consider now the case where $|G_1(s)H(s)| \gg 1$ and $|G_1(s)G_2(s)H(s)| \gg 1$. In this case, the closed-loop transfer function $C_D(s)/D(s)$ becomes almost zero, and the effect of the disturbance is suppressed. This is an advantage of the closed-loop system.

On the other hand, the closed-loop transfer function $C_R(s)/R(s)$ approaches $1/H(s)$ as the gain of $G_1(s)G_2(s)H(s)$ increases. This means that if $|G_1(s)G_2(s)H(s)| \gg 1$, then the closed-loop transfer function $C_R(s)/R(s)$ becomes independent of $G_1(s)$ and $G_2(s)$ and inversely proportional to $H(s)$, so that the variations of $G_1(s)$ and $G_2(s)$ do not affect the closed-loop transfer function $C_R(s)/R(s)$. This is another advantage of the closed-loop system. It can easily be seen that any closed-loop system with unity feedback, $H(s) = 1$, tends to equalize the input and output.

Procedures for Drawing a Block Diagram. To draw a block diagram for a system, first write the equations that describe the dynamic behavior of each component. Then take the Laplace transforms of these equations, assuming zero initial conditions, and represent each Laplace-transformed equation individually in block form. Finally, assemble the elements into a complete block diagram.

As an example, consider the *RC* circuit shown in Figure 2–12(a). The equations for this circuit are

$$i = \frac{e_i - e_o}{R} \tag{2–4}$$

$$e_o = \frac{\int i\,dt}{C} \tag{2–5}$$

The Laplace transforms of Equations (2–4) and (2–5), with zero initial condition, become

$$I(s) = \frac{E_i(s) - E_o(s)}{R} \tag{2–6}$$

$$E_o(s) = \frac{I(s)}{Cs} \tag{2–7}$$

Equation (2–6) represents a summing operation, and the corresponding diagram is shown in Figure 2–12(b). Equation (2–7) represents the block as shown in Figure 2–12(c). Assembling these two elements, we obtain the overall block diagram for the system as shown in Figure 2–12(d).

Block Diagram Reduction. It is important to note that blocks can be connected in series only if the output of one block is not affected by the next following block. If there are any loading effects between the components, it is necessary to combine these components into a single block.

Any number of cascaded blocks representing nonloading components can be replaced by a single block, the transfer function of which is simply the product of the individual transfer functions.

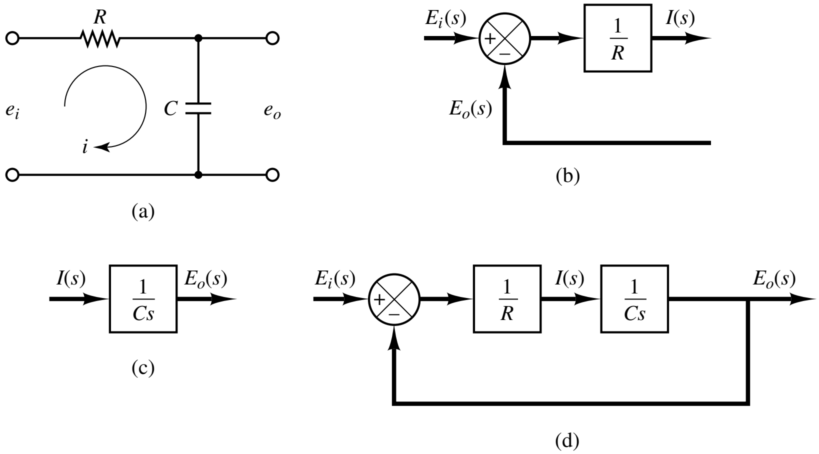


Figure 2–12
(a) *RC* circuit;
(b) block diagram representing Equation (2–6);
(c) block diagram representing Equation (2–7);
(d) block diagram of the *RC* circuit.

A complicated block diagram involving many feedback loops can be simplified by a step-by-step rearrangement. Simplification of the block diagram by rearrangements considerably reduces the labor needed for subsequent mathematical analysis. It should be noted, however, that as the block diagram is simplified, the transfer functions in new blocks become more complex because new poles and new zeros are generated.

EXAMPLE 2–1 Consider the system shown in Figure 2–13(a). Simplify this diagram.

By moving the summing point of the negative feedback loop containing H_2 outside the positive feedback loop containing H_1, we obtain Figure 2–13(b). Eliminating the positive feedback loop, we have Figure 2–13(c). The elimination of the loop containing H_2/G_1 gives Figure 2–13(d). Finally, eliminating the feedback loop results in Figure 2–13(e).

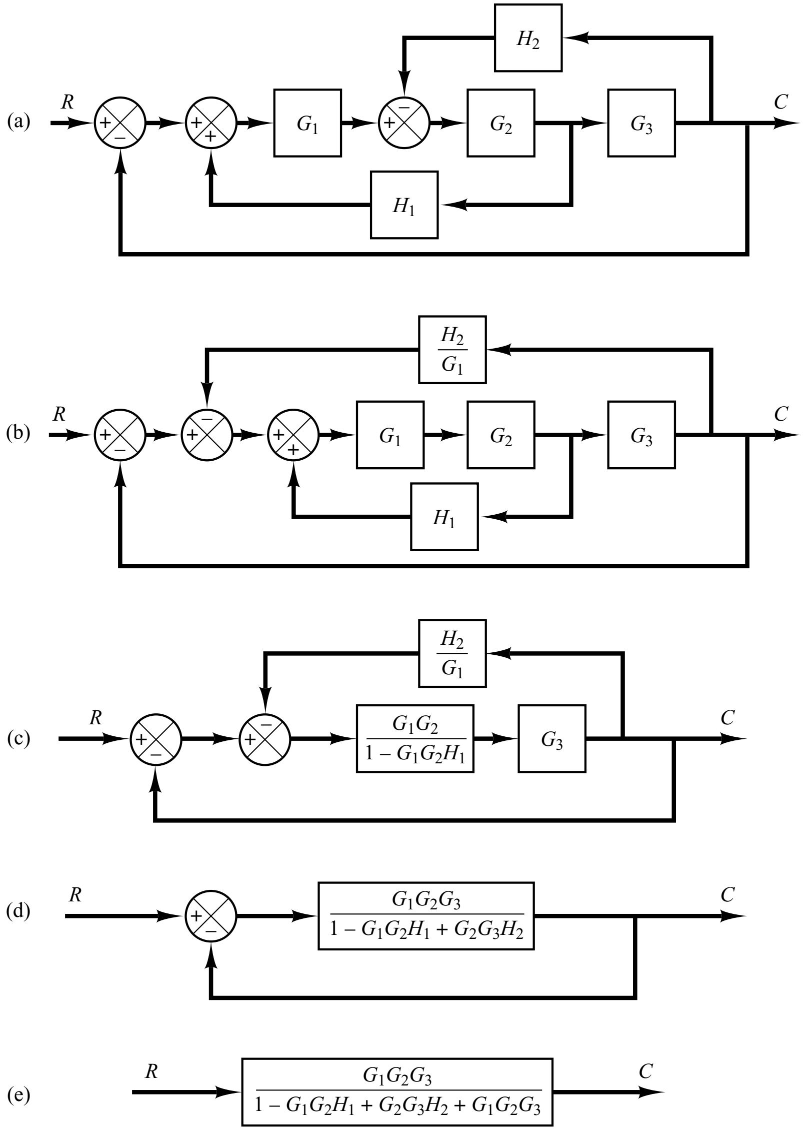


Figure 2–13
(a) Multiple-loop system;
(b)–(e) successive reductions of the block diagram shown in (a).

Notice that the numerator of the closed-loop transfer function $C(s)/R(s)$ is the product of the transfer functions of the feedforward path. The denominator of $C(s)/R(s)$ is equal to

$$1 + \sum \text{(product of the transfer functions around each loop)}$$

$$= 1 + \left(-G_1 G_2 H_1 + G_2 G_3 H_2 + G_1 G_2 G_3\right)$$

$$= 1 - G_1 G_2 H_1 + G_2 G_3 H_2 + G_1 G_2 G_3$$

(The positive feedback loop yields a negative term in the denominator.)

2–4 MODELING IN STATE SPACE

In this section we shall present introductory material on state-space analysis of control systems.

Modern Control Theory. The modern trend in engineering systems is toward greater complexity, due mainly to the requirements of complex tasks and good accuracy. Complex systems may have multiple inputs and multiple outputs and may be time varying. Because of the necessity of meeting increasingly stringent requirements on the performance of control systems, the increase in system complexity, and easy access to large scale computers, modern control theory, which is a new approach to the analysis and design of complex control systems, has been developed since around 1960. This new approach is based on the concept of state. The concept of state by itself is not new, since it has been in existence for a long time in the field of classical dynamics and other fields.

Modern Control Theory Versus Conventional Control Theory. Modern control theory is contrasted with conventional control theory in that the former is applicable to multiple-input, multiple-output systems, which may be linear or nonlinear, time invariant or time varying, while the latter is applicable only to linear time-invariant single-input, single-output systems. Also, modern control theory is essentially time-domain approach and frequency domain approach (in certain cases such as H-infinity control), while conventional control theory is a complex frequency-domain approach. Before we proceed further, we must define state, state variables, state vector, and state space.

State. The state of a dynamic system is the smallest set of variables (called *state variables*) such that knowledge of these variables at $t = t_0$, together with knowledge of the input for $t \geq t_0$, completely determines the behavior of the system for any time $t \geq t_0$.

Note that the concept of state is by no means limited to physical systems. It is applicable to biological systems, economic systems, social systems, and others.

State Variables. The state variables of a dynamic system are the variables making up the smallest set of variables that determine the state of the dynamic system. If at least n variables $x_1, x_2, \ldots, x_n$ are needed to completely describe the behavior of a dynamic system (so that once the input is given for $t \geq t_0$ and the initial state at $t = t_0$ is specified, the future state of the system is completely determined), then such n variables are a set of state variables.

Note that state variables need not be physically measurable or observable quantities. Variables that do not represent physical quantities and those that are neither measurable nor observable can be chosen as state variables. Such freedom in choosing state variables is an advantage of the state-space methods. Practically, however, it is convenient to choose easily measurable quantities for the state variables, if this is possible at all, because optimal control laws will require the feedback of all state variables with suitable weighting.

State Vector. If n state variables are needed to completely describe the behavior of a given system, then these n state variables can be considered the n components of a vector $\mathbf{x}$. Such a vector is called a *state vector*. A state vector is thus a vector that determines uniquely the system state $\mathbf{x}(t)$ for any time $t \geq t_0$, once the state at $t = t_0$ is given and the input $u(t)$ for $t \geq t_0$ is specified.

State Space. The n-dimensional space whose coordinate axes consist of the x_1 axis, x_2 axis, $\ldots$, x_n axis, where $x_1, x_2, \ldots, x_n$ are state variables, is called a *state space*. Any state can be represented by a point in the state space.

State-Space Equations. In state-space analysis we are concerned with three types of variables that are involved in the modeling of dynamic systems: input variables, output variables, and state variables. As we shall see in Section 2–5, the state-space representation for a given system is not unique, except that the number of state variables is the same for any of the different state-space representations of the same system.

The dynamic system must involve elements that memorize the values of the input for $t \geq t_1$. Since integrators in a continuous-time control system serve as memory devices, the outputs of such integrators can be considered as the variables that define the internal state of the dynamic system. Thus the outputs of integrators serve as state variables. The number of state variables to completely define the dynamics of the system is equal to the number of integrators involved in the system.

Assume that a multiple-input, multiple-output system involves n integrators. Assume also that there are r inputs $u_1(t), u_2(t), \ldots, u_r(t)$ and m outputs $y_1(t), y_2(t), \ldots, y_m(t)$. Define n outputs of the integrators as state variables: $x_1(t), x_2(t), \ldots, x_n(t)$ Then the system may be described by

$$\begin{aligned} \dot{x}_1(t) &= f_1(x_1, x_2, \ldots, x_n; u_1, u_2, \ldots, u_r; t) \\ \dot{x}_2(t) &= f_2(x_1, x_2, \ldots, x_n; u_1, u_2, \ldots, u_r; t) \\ &\cdot \\ &\cdot \\ &\cdot \\ \dot{x}_n(t) &= f_n(x_1, x_2, \ldots, x_n; u_1, u_2, \ldots, u_r; t) \end{aligned} \tag{2–8}$$

The outputs $y_1(t), y_2(t), \ldots, y_m(t)$ of the system may be given by

$$\begin{aligned} y_1(t) &= g_1(x_1, x_2, \ldots, x_n; u_1, u_2, \ldots, u_r; t) \\ y_2(t) &= g_2(x_1, x_2, \ldots, x_n; u_1, u_2, \ldots, u_r; t) \\ &\cdot \\ &\cdot \\ &\cdot \\ y_m(t) &= g_m(x_1, x_2, \ldots, x_n; u_1, u_2, \ldots, u_r; t) \end{aligned} \tag{2–9}$$

If we define

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \cdot \\ \cdot \\ \cdot \\ x_n(t) \end{bmatrix}, \quad \mathbf{f}(\mathbf{x}, \mathbf{u}, t) = \begin{bmatrix} f_1(x_1, x_2, \ldots, x_n; u_1, u_2, \ldots, u_r; t) \\ f_2(x_1, x_2, \ldots, x_n; u_1, u_2, \ldots, u_r; t) \\ \cdot \\ \cdot \\ \cdot \\ f_n(x_1, x_2, \ldots, x_n; u_1, u_2, \ldots, u_r; t) \end{bmatrix},$$

$$\mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \cdot \\ \cdot \\ \cdot \\ y_m(t) \end{bmatrix}, \quad \mathbf{g}(\mathbf{x}, \mathbf{u}, t) = \begin{bmatrix} g_1(x_1, x_2, \ldots, x_n; u_1, u_2, \ldots, u_r; t) \\ g_2(x_1, x_2, \ldots, x_n; u_1, u_2, \ldots, u_r; t) \\ \cdot \\ \cdot \\ \cdot \\ g_m(x_1, x_2, \ldots, x_n; u_1, u_2, \ldots, u_r; t) \end{bmatrix}, \quad \mathbf{u}(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \\ \cdot \\ \cdot \\ \cdot \\ u_r(t) \end{bmatrix}$$

then Equations (2–8) and (2–9) become

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}, \mathbf{u}, t) \tag{2–10}$$

$$\mathbf{y}(t) = \mathbf{g}(\mathbf{x}, \mathbf{u}, t) \tag{2–11}$$

where Equation (2–10) is the state equation and Equation (2–11) is the output equation. If vector functions **f** and/or **g** involve time t explicitly, then the system is called a time-varying system.

If Equations (2–10) and (2–11) are linearized about the operating state, then we have the following linearized state equation and output equation:

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \tag{2–12}$$

$$\mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{D}(t)\mathbf{u}(t) \tag{2–13}$$

where $\mathbf{A}(t)$ is called the state matrix, $\mathbf{B}(t)$ the input matrix, $\mathbf{C}(t)$ the output matrix, and $\mathbf{D}(t)$ the direct transmission matrix. (Details of linearization of nonlinear systems about

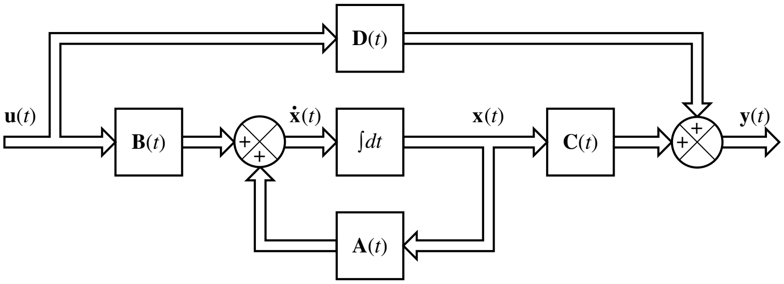


Figure 2–14 Block diagram of the linear, continuous-time control system represented in state space.

the operating state are discussed in Section 2–7.) A block diagram representation of Equations (2–12) and (2–13) is shown in Figure 2–14.

If vector functions **f** and **g** do not involve time *t* explicitly then the system is called a time-invariant system. In this case, Equations (2–12) and (2–13) can be simplified to

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \tag{2–14}$$

$$\dot{\mathbf{y}}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) \tag{2–15}$$

Equation (2–14) is the state equation of the linear, time-invariant system and Equation (2–15) is the output equation for the same system. In this book we shall be concerned mostly with systems described by Equations (2–14) and (2–15).

In what follows we shall present an example for deriving a state equation and output equation.

EXAMPLE 2–2 Consider the mechanical system shown in Figure 2–15. We assume that the system is linear. The external force $u(t)$ is the input to the system, and the displacement $y(t)$ of the mass is the output. The displacement $y(t)$ is measured from the equilibrium position in the absence of the external force. This system is a single-input, single-output system.

From the diagram, the system equation is

$$m\ddot{y} + b\dot{y} + ky = u \tag{2–16}$$

This system is of second order. This means that the system involves two integrators. Let us define state variables $x_1(t)$ and $x_2(t)$ as

$$x_1(t) = y(t)$$

$$x_2(t) = \dot{y}(t)$$

Then we obtain

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = \frac{1}{m}\left(-ky - b\dot{y}\right) + \frac{1}{m}u$$

or

$$\dot{x}_1 = x_2 \tag{2–17}$$

$$\dot{x}_2 = -\frac{k}{m}x_1 - \frac{b}{m}x_2 + \frac{1}{m}u \tag{2–18}$$

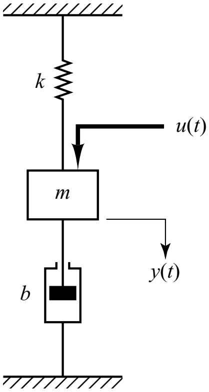


Figure 2–15 Mechanical system.

The output equation is

$$y = x_1 \tag{2–19}$$

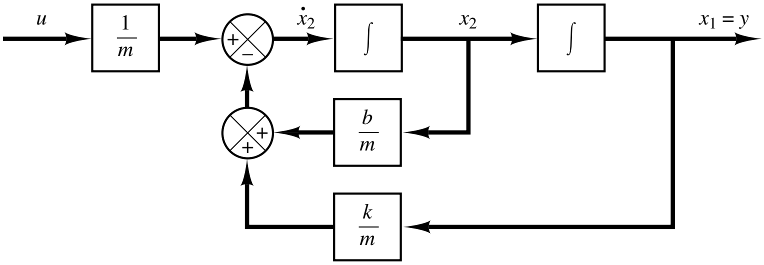


Figure 2–16 Block diagram of the mechanical system shown in Figure 2–15.

In a vector-matrix form, Equations (2–17) and (2–18) can be written as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix} u \tag{2–20}$$

The output equation, Equation (2–19), can be written as

$$y = [1 \quad 0] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{2–21}$$

Equation (2–20) is a state equation and Equation (2–21) is an output equation for the system. They are in the standard form:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u$$
$$y = \mathbf{C}\mathbf{x} + Du$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix}, \qquad \mathbf{C} = [1 \quad 0], \qquad D = 0$$

Figure 2–16 is a block diagram for the system. Notice that the outputs of the integrators are state variables.

Correlation Between Transfer Functions and State-Space Equations. In what follows we shall show how to derive the transfer function of a single-input, single-output system from the state-space equations.

Let us consider the system whose transfer function is given by

$$\frac{Y(s)}{U(s)} = G(s) \tag{2–22}$$

This system may be represented in state space by the following equations:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u \tag{2–23}$$

$$y = \mathbf{C}\mathbf{x} + Du \tag{2–24}$$

where $\mathbf{x}$ is the state vector, u is the input, and y is the output. The Laplace transforms of Equations (2–23) and (2–24) are given by

$$s\mathbf{X}(s) - \mathbf{x}(0) = \mathbf{A}\mathbf{X}(s) + \mathbf{B}U(s) \tag{2–25}$$

$$Y(s) = \mathbf{C}\mathbf{X}(s) + DU(s) \tag{2–26}$$

Since the transfer function was previously defined as the ratio of the Laplace transform of the output to the Laplace transform of the input when the initial conditions were zero, we set $\mathbf{x}(0)$ in Equation (2–25) to be zero. Then we have

$$s\mathbf{X}(s) - \mathbf{A}\mathbf{X}(s) = \mathbf{B}U(s)$$

or

$$(s\mathbf{I} - \mathbf{A})\mathbf{X}(s) = \mathbf{B}U(s)$$

By premultiplying $(s\mathbf{I} - \mathbf{A})^{-1}$ to both sides of this last equation, we obtain

$$\mathbf{X}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}U(s) \tag{2–27}$$

By substituting Equation (2–27) into Equation (2–26), we get

$$Y(s) = \left[\mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + D\right]U(s) \tag{2–28}$$

Upon comparing Equation (2–28) with Equation (2–22), we see that

$$G(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + D \tag{2–29}$$

This is the transfer-function expression of the system in terms of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and D.

Note that the right-hand side of Equation (2–29) involves $(s\mathbf{I} - \mathbf{A})^{-1}$. Hence $G(s)$ can be written as

$$G(s) = \frac{Q(s)}{|s\mathbf{I} - \mathbf{A}|}$$

where $Q(s)$ is a polynomial in s. Notice that $|s\mathbf{I} - \mathbf{A}|$ is equal to the characteristic polynomial of $G(s)$. In other words, the eigenvalues of $\mathbf{A}$ are identical to the poles of $G(s)$.

EXAMPLE 2–3 Consider again the mechanical system shown in Figure 2–15. State-space equations for the system are given by Equations (2–20) and (2–21). We shall obtain the transfer function for the system from the state-space equations.

By substituting $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and D into Equation (2–29), we obtain

$$\begin{aligned} G(s) &= \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + D \\ &= [1 \quad 0]\left\{\begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ -\dfrac{k}{m} & -\dfrac{b}{m} \end{bmatrix}\right\}^{-1}\begin{bmatrix} 0 \\ \dfrac{1}{m} \end{bmatrix} + 0 \\ &= [1 \quad 0]\begin{bmatrix} s & -1 \\ \dfrac{k}{m} & s + \dfrac{b}{m} \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ \dfrac{1}{m} \end{bmatrix} \end{aligned}$$

Note that

$$\begin{bmatrix} s & -1 \\ \frac{k}{m} & s + \frac{b}{m} \end{bmatrix}^{-1} = \frac{1}{s^2 + \frac{b}{m}s + \frac{k}{m}} \begin{bmatrix} s + \frac{b}{m} & 1 \\ -\frac{k}{m} & s \end{bmatrix}$$

(Refer to Appendix C for the inverse of the 2 × 2 matrix.)
Thus, we have

$$\begin{aligned} G(s) &= [1 \quad 0] \frac{1}{s^2 + \frac{b}{m}s + \frac{k}{m}} \begin{bmatrix} s + \frac{b}{m} & 1 \\ -\frac{k}{m} & s \end{bmatrix} \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix} \\ &= \frac{1}{ms^2 + bs + k} \end{aligned}$$

which is the transfer function of the system. The same transfer function can be obtained from Equation (2–16).

Transfer Matrix. Next, consider a multiple-input, multiple-output system. Assume that there are r inputs $u_1, u_2, \ldots, u_r$, and m outputs $y_1, y_2, \ldots, y_m$. Define

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_m \end{bmatrix}, \qquad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ \cdot \\ u_r \end{bmatrix}$$

The transfer matrix $\mathbf{G}(s)$ relates the output $\mathbf{Y}(s)$ to the input $\mathbf{U}(s)$, or

$$\mathbf{Y}(s) = \mathbf{G}(s)\mathbf{U}(s)$$

where $\mathbf{G}(s)$ is given by

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

[The derivation for this equation is the same as that for Equation (2–29).] Since the input vector **u** is r dimensional and the output vector **y** is m dimensional, the transfer matrix **G**(s) is an $m \times r$ matrix.

2–5 STATE-SPACE REPRESENTATION OF SCALAR DIFFERENTIAL EQUATION SYSTEMS

A dynamic system consisting of a finite number of lumped elements may be described by ordinary differential equations in which time is the independent variable. By use of vector-matrix notation, an nth-order differential equation may be expressed by a first-order vector-matrix differential equation. If n elements of the vector are a set of state variables, then the vector-matrix differential equation is a *state* equation. In this section we shall present methods for obtaining state-space representations of continuous-time systems.

State-Space Representation of *n*th-Order Systems of Linear Differential Equations in which the Forcing Function Does Not Involve Derivative Terms. Consider the following nth-order system:

$$\overset{(n)}{y} + a_1\overset{(n-1)}{y} + \cdots + a_{n-1}\dot{y} + a_n y = u \tag{2–30}$$

Noting that the knowledge of $y(0), \dot{y}(0), \ldots, \overset{(n-1)}{y}(0)$, together with the input $u(t)$ for $t \geq 0$, determines completely the future behavior of the system, we may take $y(t), \dot{y}(t), \ldots, \overset{(n-1)}{y}(t)$ as a set of n state variables. (Mathematically, such a choice of state variables is quite convenient. Practically, however, because higher-order derivative terms are inaccurate, due to the noise effects inherent in any practical situations, such a choice of the state variables may not be desirable.)

Let us define

$$\begin{aligned}
x_1 &= y \\
x_2 &= \dot{y} \\
&\cdot \\
&\cdot \\
&\cdot \\
x_n &= \overset{(n-1)}{y}
\end{aligned}$$

Then Equation (2–30) can be written as

$$\begin{aligned}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= x_3 \\
&\cdot \\
&\cdot \\
&\cdot \\
\dot{x}_{n-1} &= x_n \\
\dot{x}_n &= -a_n x_1 - \cdots - a_1 x_n + u
\end{aligned}$$

or

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u \tag{2–31}$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix}, \qquad \mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 1 \end{bmatrix}$$

The output can be given by

$$y = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix}$$

or

$$y = \mathbf{Cx} \tag{2–32}$$

where

$$\mathbf{C} = [1 \quad 0 \quad \cdots \quad 0]$$

[Note that D in Equation (2–24) is zero.] The first-order differential equation, Equation (2–31), is the state equation, and the algebraic equation, Equation (2–32), is the output equation.

Note that the state-space representation for the transfer function system

$$\frac{Y(s)}{U(s)} = \frac{1}{s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n}$$

is given also by Equations (2–31) and (2–32).

State-Space Representation of *n*th-Order Systems of Linear Differential Equations in which the Forcing Function Involves Derivative Terms. Consider the differential equation system that involves derivatives of the forcing function, such as

$$\overset{(n)}{y} + a_1 \overset{(n-1)}{y} + \cdots + a_{n-1}\dot{y} + a_n y = b_0 \overset{(n)}{u} + b_1 \overset{(n-1)}{u} + \cdots + b_{n-1}\dot{u} + b_n u \tag{2–33}$$

The main problem in defining the state variables for this case lies in the derivative terms of the input u. The state variables must be such that they will eliminate the derivatives of u in the state equation.

One way to obtain a state equation and output equation for this case is to define the following n variables as a set of n state variables:

$$\begin{aligned} x_1 &= y - \beta_0 u \\ x_2 &= \dot{y} - \beta_0 \dot{u} - \beta_1 u = \dot{x}_1 - \beta_1 u \\ x_3 &= \ddot{y} - \beta_0 \ddot{u} - \beta_1 \dot{u} - \beta_2 u = \dot{x}_2 - \beta_2 u \\ &\cdot \\ &\cdot \\ &\cdot \\ x_n &= \overset{(n-1)}{y} - \beta_0 \overset{(n-1)}{u} - \beta_1 \overset{(n-2)}{u} - \cdots - \beta_{n-2}\dot{u} - \beta_{n-1}u = \dot{x}_{n-1} - \beta_{n-1}u \end{aligned} \tag{2–34}$$

where $\beta_0, \beta_1, \beta_2, \ldots, \beta_{n-1}$ are determined from

$$\begin{aligned} \beta_0 &= b_0 \\ \beta_1 &= b_1 - a_1\beta_0 \\ \beta_2 &= b_2 - a_1\beta_1 - a_2\beta_0 \\ \beta_3 &= b_3 - a_1\beta_2 - a_2\beta_1 - a_3\beta_0 \\ &\;\cdot \\ &\;\cdot \\ &\;\cdot \\ \beta_{n-1} &= b_{n-1} - a_1\beta_{n-2} - \cdots - a_{n-2}\beta_1 - a_{n-1}\beta_0 \end{aligned} \tag{2–35}$$

With this choice of state variables the existence and uniqueness of the solution of the state equation is guaranteed. (Note that this is not the only choice of a set of state variables.) With the present choice of state variables, we obtain

$$\begin{aligned} \dot{x}_1 &= x_2 + \beta_1 u \\ \dot{x}_2 &= x_3 + \beta_2 u \\ &\;\cdot \\ &\;\cdot \\ &\;\cdot \\ \dot{x}_{n-1} &= x_n + \beta_{n-1} u \\ \dot{x}_n &= -a_n x_1 - a_{n-1} x_2 - \cdots - a_1 x_n + \beta_n u \end{aligned} \tag{2–36}$$

where β_n is given by

$$\beta_n = b_n - a_1\beta_{n-1} - \cdots - a_{n-1}\beta_1 - a_{n-1}\beta_0$$

[To derive Equation (2–36), see Problem **A–2–6**.] In terms of vector-matrix equations, Equation (2–36) and the output equation can be written as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \dot{x}_{n-1} \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_{n-1} \\ x_n \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_{n-1} \\ \beta_n \end{bmatrix} u$$

$$y = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} + \beta_0 u$$

or

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u \tag{2–37}$$

$$y = \mathbf{C}\mathbf{x} + Du \tag{2–38}$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_{n-1} \\ x_n \end{bmatrix}, \qquad \mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_{n-1} \\ \beta_n \end{bmatrix}, \qquad \mathbf{C} = [1 \quad 0 \quad \cdots \quad 0], \qquad D = \beta_0 = b_0$$

In this state-space representation, matrices **A** and **C** are exactly the same as those for the system of Equation (2–30). The derivatives on the right-hand side of Equation (2–33) affect only the elements of the **B** matrix.

Note that the state-space representation for the transfer function

$$\frac{Y(s)}{U(s)} = \frac{b_0 s^n + b_1 s^{n-1} + \cdots + b_{n-1} s + b_n}{s^n + a_1 s^{n-1} + \cdots + a_{n-1} s + a_n}$$

is given also by Equations (2–37) and (2–38).

There are many ways to obtain state-space representations of systems. Methods for obtaining canonical representations of systems in state space (such as controllable canonical form, observable canonical form, diagonal canonical form, and Jordan canonical form) are presented in Chapter 9.

MATLAB can also be used to obtain state-space representations of systems from transfer-function representations, and vice versa. This subject is presented in Section 2–6.

2–6 TRANSFORMATION OF MATHEMATICAL MODELS WITH MATLAB

MATLAB is quite useful to transform the system model from transfer function to state space, and vice versa. We shall begin our discussion with transformation from transfer function to state space.

Let us write the closed-loop transfer function as

$$\frac{Y(s)}{U(s)} = \frac{\text{numerator polynomial in } s}{\text{denominator polynomial in } s} = \frac{\text{num}}{\text{den}}$$

Once we have this transfer-function expression, the MATLAB command

```
[A,B,C,D] = tf2ss(num,den)
```

will give a state-space representation. It is important to note that the state-space representation for any system is not unique. There are many (infinitely many) state-space representations for the same system. The MATLAB command gives one possible such state-space representation.

Transformation from Transfer Function to State Space Representation. Consider the transfer-function system

$$\frac{Y(s)}{U(s)} = \frac{s}{(s + 10)(s^2 + 4s + 16)}$$

$$= \frac{s}{s^3 + 14s^2 + 56s + 160} \qquad (2\text{–}39)$$

There are many (infinitely many) possible state-space representations for this system. One possible state-space representation is

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -160 & -56 & -14 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ -14 \end{bmatrix} u$$

$$y = [1 \quad 0 \quad 0] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + [0]u$$

Another possible state-space representation (among infinitely many alternatives) is

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -14 & -56 & -160 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u \qquad (2\text{–}40)$$

$$y = [0 \quad 1 \quad 0]\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + [0]u \tag{2–41}$$

MATLAB transforms the transfer function given by Equation (2–39) into the state-space representation given by Equations (2–40) and (2–41). For the example system considered here, MATLAB Program 2–2 will produce matrices **A**, **B**, **C**, and *D*.

MATLAB Program 2–2

```
num = [1    0];
den = [1   14   56   160];
[A,B,C,D] = tf2ss(num,den)

A =

   -14   -56  -160
     1     0     0
     0     1     0

B =

     1
     0
     0

C =

     0     1     0

D =

     0
```

Transformation from State Space Representation to Transfer Function. To obtain the transfer function from state-space equations, use the following command:

```
[num,den] = ss2tf(A,B,C,D,iu)
```

`iu` must be specified for systems with more than one input. For example, if the system has three inputs ($u1$, $u2$, $u3$), then `iu` must be either 1, 2, or 3, where 1 implies $u1$, 2 implies $u2$, and 3 implies $u3$.

If the system has only one input, then either

```
[num,den] = ss2tf(A,B,C,D)
```

or

```
[num,den] = ss2tf(A,B,C,D,1)
```

may be used. For the case where the system has multiple inputs and multiple outputs, see Problem **A–2–12**.

EXAMPLE 2–4 Obtain the transfer function of the system defined by the following state-space equations:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -5 & -25 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 25 \\ -120 \end{bmatrix} u$$

$$y = [1 \quad 0 \quad 0] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

MATLAB Program 2-3 will produce the transfer function for the given system. The transfer function obtained is given by

$$\frac{Y(s)}{U(s)} = \frac{25s + 5}{s^3 + 5s^2 + 25s + 5}$$

MATLAB Program 2–3

```
A = [0   1   0;   0   0   1;   -5   -25   -5];
B = [0; 25; -120];
C = [1   0     0];
D = [0];
[num,den] = ss2tf(A,B,C,D)

num =

   0   0.0000   25.0000  5.0000

den

   1.0000   5.0000   25.0000  5.0000

% ***** The same result can be obtained by entering the following command: *****

[num,den] = ss2tf(A,B,C,D,1)

num =

   0   0.0000   25.0000  5.0000

den =

   1.0000   5.0000   25.0000  5.0000
```

2–7 LINEARIZATION OF NONLINEAR MATHEMATICAL MODELS

Nonlinear Systems. A system is nonlinear if the principle of superposition does not apply. Thus, for a nonlinear system the response to two inputs cannot be calculated by treating one input at a time and adding the results.

Although many physical relationships are often represented by linear equations, in most cases actual relationships are not quite linear. In fact, a careful study of physical systems reveals that even so-called "linear systems" are really linear only in limited operating ranges. In practice, many electromechanical systems, hydraulic systems, pneumatic systems, and so on, involve nonlinear relationships among the variables. For example, the output of a component may saturate for large input signals. There may be a dead space that affects small signals. (The dead space of a component is a small range of input variations to which the component is insensitive.) Square-law nonlinearity may occur in some components. For instance, dampers used in physical systems may be linear for low-velocity operations but may become nonlinear at high velocities, and the damping force may become proportional to the square of the operating velocity.

Linearization of Nonlinear Systems. In control engineering a normal operation of the system may be around an equilibrium point, and the signals may be considered small signals around the equilibrium. (It should be pointed out that there are many exceptions to such a case.) However, if the system operates around an equilibrium point and if the signals involved are small signals, then it is possible to approximate the nonlinear system by a linear system. Such a linear system is equivalent to the nonlinear system considered within a limited operating range. Such a linearized model (linear, time-invariant model) is very important in control engineering.

The linearization procedure to be presented in the following is based on the expansion of nonlinear function into a Taylor series about the operating point and the retention of only the linear term. Because we neglect higher-order terms of the Taylor series expansion, these neglected terms must be small enough; that is, the variables deviate only slightly from the operating condition. (Otherwise, the result will be inaccurate.)

Linear Approximation of Nonlinear Mathematical Models. To obtain a linear mathematical model for a nonlinear system, we assume that the variables deviate only slightly from some operating condition. Consider a system whose input is $x(t)$ and output is $y(t)$. The relationship between $y(t)$ and $x(t)$ is given by

$$y = f(x) \tag{2–42}$$

If the normal operating condition corresponds to $\bar{x}$, $\bar{y}$, then Equation (2–42) may be expanded into a Taylor series about this point as follows:

$$\begin{aligned} y &= f(x) \\ &= f(\bar{x}) + \frac{df}{dx}(x - \bar{x}) + \frac{1}{2!}\frac{d^2f}{dx^2}(x - \bar{x})^2 + \cdots \end{aligned} \tag{2–43}$$

where the derivatives $df/dx, d^2f/dx^2, \ldots$ are evaluated at $x = \bar{x}$. If the variation $x - \bar{x}$ is small, we may neglect the higher-order terms in $x - \bar{x}$. Then Equation (2–43) may be written as

$$y = \bar{y} + K(x - \bar{x}) \tag{2–44}$$

where

$$\bar{y} = f(\bar{x})$$

$$K = \left.\frac{df}{dx}\right|_{x=\bar{x}}$$

Equation (2–44) may be rewritten as

$$y - \bar{y} = K(x - \bar{x}) \tag{2–45}$$

which indicates that $y - \bar{y}$ is proportional to $x - \bar{x}$. Equation (2–45) gives a linear mathematical model for the nonlinear system given by Equation (2–42) near the operating point $x = \bar{x}, y = \bar{y}$.

Next, consider a nonlinear system whose output y is a function of two inputs x_1 and x_2, so that

$$y = f(x_1, x_2) \tag{2–46}$$

To obtain a linear approximation to this nonlinear system, we may expand Equation (2–46) into a Taylor series about the normal operating point $\bar{x}_1, \bar{x}_2$. Then Equation (2–46) becomes

$$\begin{aligned} y = f(\bar{x}_1, \bar{x}_2) &+ \left[\frac{\partial f}{\partial x_1}(x_1 - \bar{x}_1) + \frac{\partial f}{\partial x_2}(x_2 - \bar{x}_2)\right] \\ &+ \frac{1}{2!}\left[\frac{\partial^2 f}{\partial x_1^2}(x_1 - \bar{x}_1)^2 + 2\frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)\right. \\ &\left. + \frac{\partial^2 f}{\partial x_2^2}(x_2 - \bar{x}_2)^2\right] + \cdots \end{aligned}$$

where the partial derivatives are evaluated at $x_1 = \bar{x}_1, x_2 = \bar{x}_2$. Near the normal operating point, the higher-order terms may be neglected. The linear mathematical model of this nonlinear system in the neighborhood of the normal operating condition is then given by

$$y - \bar{y} = K_1(x_1 - \bar{x}_1) + K_2(x_2 - \bar{x}_2)$$

where

$$\bar{y} = f(\bar{x}_1, \bar{x}_2)$$

$$K_1 = \left.\frac{\partial f}{\partial x_1}\right|_{x_1=\bar{x}_1,\, x_2=\bar{x}_2}$$

$$K_2 = \left.\frac{\partial f}{\partial x_2}\right|_{x_1=\bar{x}_1,\, x_2=\bar{x}_2}$$

The linearization technique presented here is valid in the vicinity of the operating condition. If the operating conditions vary widely, however, such linearized equations are not adequate, and nonlinear equations must be dealt with. It is important to remember that a particular mathematical model used in analysis and design may accurately represent the dynamics of an actual system for certain operating conditions, but may not be accurate for other operating conditions.

EXAMPLE 2–5 Linearize the nonlinear equation

$$z = xy$$

in the region $5 \le x \le 7$, $10 \le y \le 12$. Find the error if the linearized equation is used to calculate the value of z when $x = 5$, $y = 10$.

Since the region considered is given by $5 \le x \le 7$, $10 \le y \le 12$, choose $\bar{x} = 6$, $\bar{y} = 11$. Then $\bar{z} = \bar{x}\bar{y} = 66$. Let us obtain a linearized equation for the nonlinear equation near a point $\bar{x} = 6$, $\bar{y} = 11$.

Expanding the nonlinear equation into a Taylor series about point $x = \bar{x}$, $y = \bar{y}$ and neglecting the higher-order terms, we have

$$z - \bar{z} = a(x - \bar{x}) + b(y - \bar{y})$$

where

$$a = \left.\frac{\partial(xy)}{\partial x}\right|_{x=\bar{x},\, y=\bar{y}} = \bar{y} = 11$$

$$b = \left.\frac{\partial(xy)}{\partial y}\right|_{x=\bar{x},\, y=\bar{y}} = \bar{x} = 6$$

Hence the linearized equation is

$$z - 66 = 11(x - 6) + 6(y - 11)$$

or

$$z = 11x + 6y - 66$$

When $x = 5$, $y = 10$, the value of z given by the linearized equation is

$$z = 11x + 6y - 66 = 55 + 60 - 66 = 49$$

The exact value of z is $z = xy = 50$. The error is thus $50 - 49 = 1$. In terms of percentage, the error is 2%.

EXAMPLE PROBLEMS AND SOLUTIONS

A–2–1. Simplify the block diagram shown in Figure 2–17.

Solution. First, move the branch point of the path involving H_1 outside the loop involving H_2, as shown in Figure 2–18(a). Then eliminating two loops results in Figure 2–18(b). Combining two blocks into one gives Figure 2–18(c).

A–2–2. Simplify the block diagram shown in Figure 2–19. Obtain the transfer function relating $C(s)$ and $R(s)$.

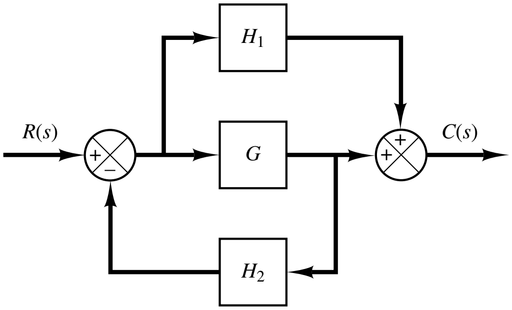


Figure 2–17 Block diagram of a system.

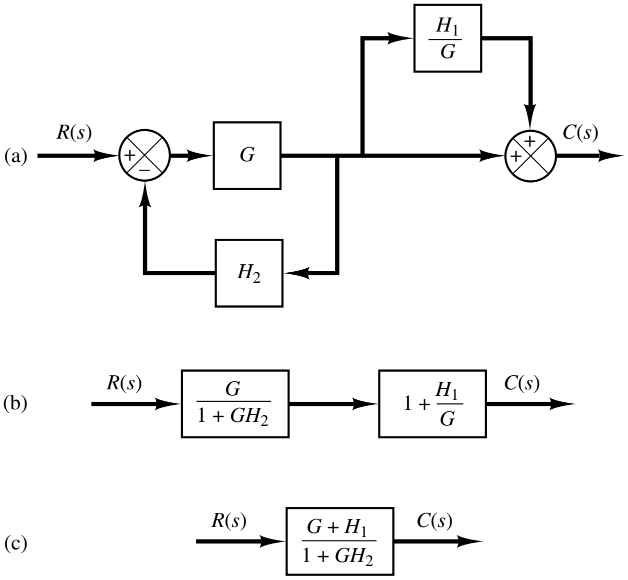


Figure 2–18 Simplified block diagrams for the system shown in Figure 2–17.

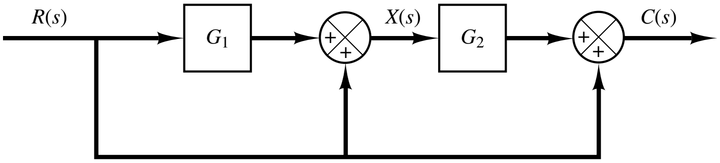


Figure 2–19 Block diagram of a system.

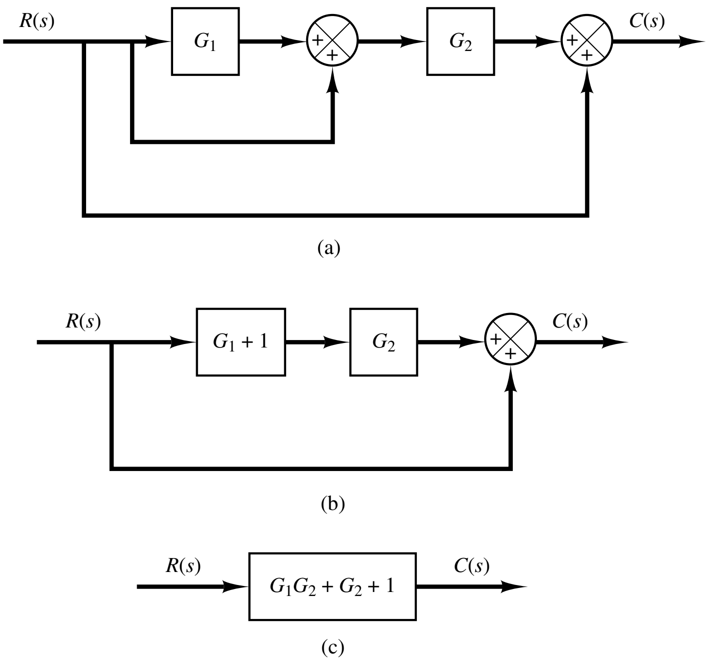


Figure 2–20 Reduction of the block diagram shown in Figure 2–19.

Solution. The block diagram of Figure 2–19 can be modified to that shown in Figure 2–20(a). Eliminating the minor feedforward path, we obtain Figure 2–20(b), which can be simplified to Figure 2–20(c). The transfer function $C(s)/R(s)$ is thus given by

$$\frac{C(s)}{R(s)} = G_1G_2 + G_2 + 1$$

The same result can also be obtained by proceeding as follows: Since signal $X(s)$ is the sum of two signals $G_1R(s)$ and $R(s)$, we have

$$X(s) = G_1R(s) + R(s)$$

The output signal $C(s)$ is the sum of $G_2X(s)$ and $R(s)$. Hence

$$C(s) = G_2X(s) + R(s) = G_2\left[G_1R(s) + R(s)\right] + R(s)$$

And so we have the same result as before:

$$\frac{C(s)}{R(s)} = G_1G_2 + G_2 + 1$$

A–2–3. Simplify the block diagram shown in Figure 2–21. Then obtain the closed-loop transfer function $C(s)/R(s)$.

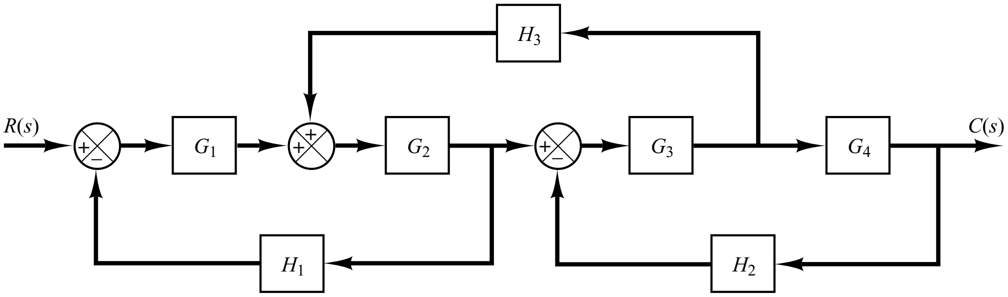


Figure 2–21 Block diagram of a system.

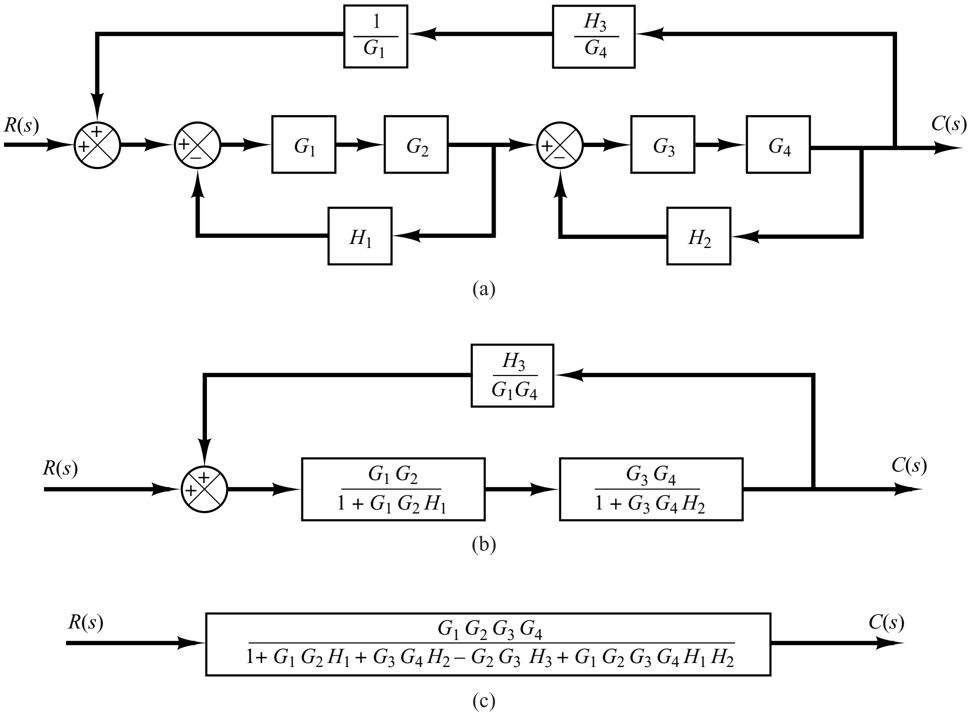


Figure 2–22 Successive reductions of the block diagram shown in Figure 2–21.

Solution. First move the branch point between G_3 and G_4 to the right-hand side of the loop containing G_3, G_4, and H_2. Then move the summing point between G_1 and G_2 to the left-hand side of the first summing point. See Figure 2–22(a). By simplifying each loop, the block diagram can be modified as shown in Figure 2–22(b). Further simplification results in Figure 2–22(c), from which the closed-loop transfer function $C(s)/R(s)$ is obtained as

$$\frac{C(s)}{R(s)} = \frac{G_1G_2G_3G_4}{1 + G_1G_2H_1 + G_3G_4H_2 - G_2G_3H_3 + G_1G_2G_3G_4H_1H_2}$$

A–2–4. Obtain transfer functions $C(s)/R(s)$ and $C(s)/D(s)$ of the system shown in Figure 2–23.

Solution. From Figure 2–23 we have

$$U(s) = G_f R(s) + G_c E(s) \tag{2–47}$$

$$C(s) = G_p\left[D(s) + G_1 U(s)\right] \tag{2–48}$$

$$E(s) = R(s) - HC(s) \tag{2–49}$$

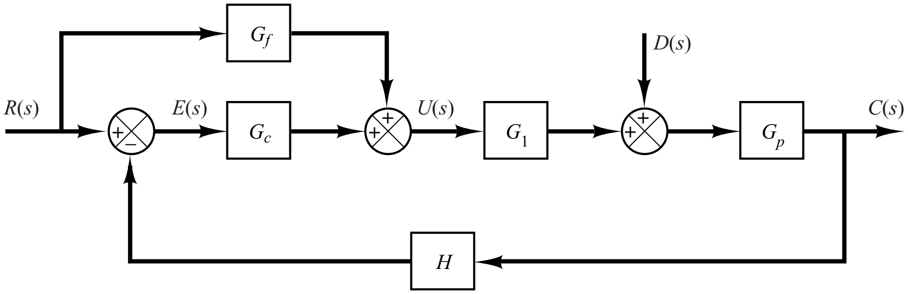


Figure 2–23 Control system with reference input and disturbance input.

By substituting Equation (2–47) into Equation (2–48), we get

$$C(s) = G_p D(s) + G_1 G_p \left[G_f R(s) + G_c E(s) \right] \tag{2–50}$$

By substituting Equation (2–49) into Equation (2–50), we obtain

$$C(s) = G_p D(s) + G_1 G_p \left\{ G_f R(s) + G_c \left[R(s) - HC(s) \right] \right\}$$

Solving this last equation for $C(s)$, we get

$$C(s) + G_1 G_p G_c HC(s) = G_p D(s) + G_1 G_p \left(G_f + G_c \right) R(s)$$

Hence

$$C(s) = \frac{G_p D(s) + G_1 G_p \left(G_f + G_c \right) R(s)}{1 + G_1 G_p G_c H} \tag{2–51}$$

Note that Equation (2–51) gives the response $C(s)$ when both reference input $R(s)$ and disturbance input $D(s)$ are present.

To find transfer function $C(s)/R(s)$, we let $D(s) = 0$ in Equation (2–51). Then we obtain

$$\frac{C(s)}{R(s)} = \frac{G_1 G_p \left(G_f + G_c \right)}{1 + G_1 G_p G_c H}$$

Similarly, to obtain transfer function $C(s)/D(s)$, we let $R(s) = 0$ in Equation (2–51). Then $C(s)/D(s)$ can be given by

$$\frac{C(s)}{D(s)} = \frac{G_p}{1 + G_1 G_p G_c H}$$

A–2–5. Figure 2–24 shows a system with two inputs and two outputs. Derive $C_1(s)/R_1(s)$, $C_1(s)/R_2(s)$, $C_2(s)/R_1(s)$, and $C_2(s)/R_2(s)$. (In deriving outputs for $R_1(s)$, assume that $R_2(s)$ is zero, and vice versa.)

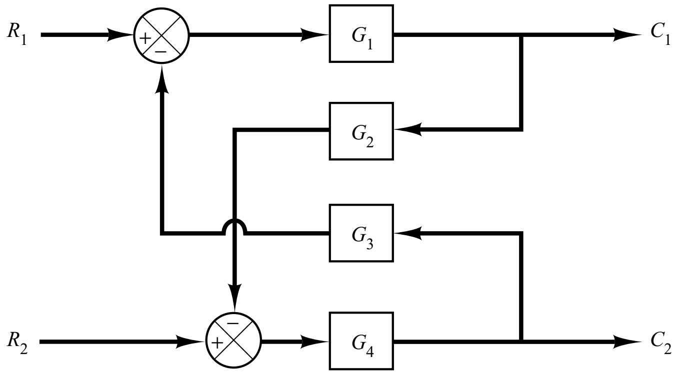


Figure 2–24 System with two inputs and two outputs.

Solution. From the figure, we obtain

$$C_1 = G_1(R_1 - G_3C_2) \tag{2–52}$$

$$C_2 = G_4(R_2 - G_2C_1) \tag{2–53}$$

By substituting Equation (2–53) into Equation (2–52), we obtain

$$C_1 = G_1[R_1 - G_3G_4(R_2 - G_2C_1)] \tag{2–54}$$

By substituting Equation (2–52) into Equation (2–53), we get

$$C_2 = G_4[R_2 - G_2G_1(R_1 - G_3C_2)] \tag{2–55}$$

Solving Equation (2–54) for C_1, we obtain

$$C_1 = \frac{G_1R_1 - G_1G_3G_4R_2}{1 - G_1G_2G_3G_4} \tag{2–56}$$

Solving Equation (2–55) for C_2 gives

$$C_2 = \frac{-G_1G_2G_4R_1 + G_4R_2}{1 - G_1G_2G_3G_4} \tag{2–57}$$

Equations (2–56) and (2–57) can be combined in the form of the transfer matrix as follows:

$$\begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} \dfrac{G_1}{1 - G_1G_2G_3G_4} & -\dfrac{G_1G_3G_4}{1 - G_1G_2G_3G_4} \\ -\dfrac{G_1G_2G_4}{1 - G_1G_2G_3G_4} & \dfrac{G_4}{1 - G_1G_2G_3G_4} \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \end{bmatrix}$$

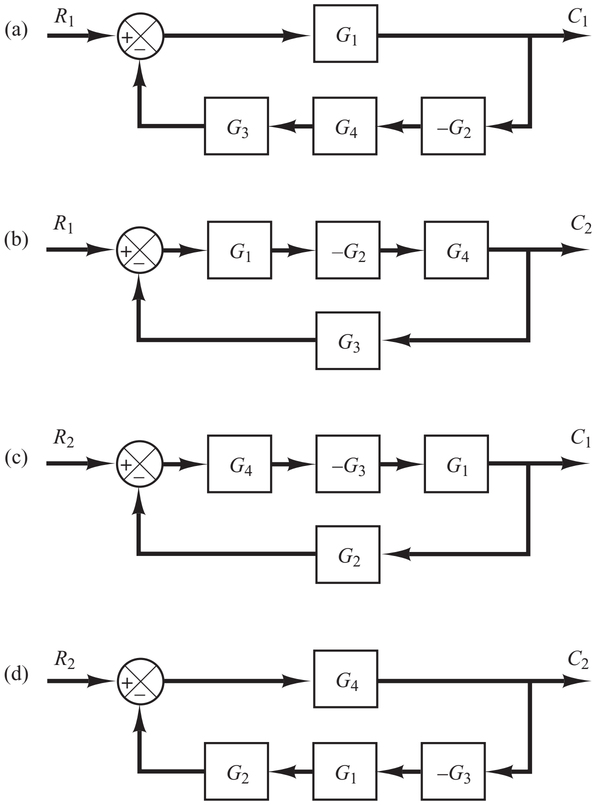
Then the transfer functions $C_1(s)/R_1(s)$, $C_1(s)/R_2(s)$, $C_2(s)/R_1(s)$ and $C_2(s)/R_2(s)$ can be obtained as follows:

$$\frac{C_1(s)}{R_1(s)} = \frac{G_1}{1 - G_1G_2G_3G_4}, \qquad \frac{C_1(s)}{R_2(s)} = -\frac{G_1G_3G_4}{1 - G_1G_2G_3G_4}$$

$$\frac{C_2(s)}{R_1(s)} = -\frac{G_1G_2G_4}{1 - G_1G_2G_3G_4}, \qquad \frac{C_2(s)}{R_2(s)} = \frac{G_4}{1 - G_1G_2G_3G_4}$$

Note that Equations (2–56) and (2–57) give responses C_1 and C_2, respectively, when both inputs R_1 and R_2 are present.

Notice that when $R_2(s) = 0$, the original block diagram can be simplified to those shown in Figures 2–25(a) and (b). Similarly, when $R_1(s) = 0$, the original block diagram can be simplified to those shown in Figures 2–25(c) and (d). From these simplified block diagrams we can also obtain $C_1(s)/R_1(s)$, $C_2(s)/R_1(s)$, $C_1(s)/R_2(s)$, and $C_2(s)/R_2(s)$, as shown to the right of each corresponding block diagram.



$$\frac{C_1}{R_1} = \frac{G_1}{1 - G_1 G_2 G_3 G_4}$$

$$\frac{C_2}{R_1} = \frac{- G_1 G_2 G_4}{1 - G_1 G_2 G_3 G_4}$$

$$\frac{C_1}{R_2} = \frac{- G_1 G_3 G_4}{1 - G_1 G_2 G_3 G_4}$$

$$\frac{C_2}{R_2} = \frac{G_4}{1 - G_1 G_2 G_3 G_4}$$

Figure 2–25 Simplified block diagrams and corresponding closed-loop transfer functions.

A–2–6. Show that for the differential equation system

$$\dddot{y} + a_1\ddot{y} + a_2\dot{y} + a_3 y = b_0\dddot{u} + b_1\ddot{u} + b_2\dot{u} + b_3 u \tag{2–58}$$

state and output equations can be given, respectively, by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_3 & -a_2 & -a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} u \tag{2–59}$$

and

$$y = [1 \quad 0 \quad 0] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \beta_0 u \tag{2–60}$$

where state variables are defined by

$$\begin{aligned} x_1 &= y - \beta_0 u \\ x_2 &= \dot{y} - \beta_0\dot{u} - \beta_1 u = \dot{x}_1 - \beta_1 u \\ x_3 &= \ddot{y} - \beta_0\ddot{u} - \beta_1\dot{u} - \beta_2 u = \dot{x}_2 - \beta_2 u \end{aligned}$$