# Modern Control Engineering

*Third Edition*

**Katsuhiko Ogata**

*University of Minnesota*

Prentice Hall, Upper Saddle River, New Jersey 07458

**Library of Congress Cataloging-in-Publication Data**

Ogata, Katsuhiko.
Modern control engineering / Katsuhiko Ogata. — 3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN: 0-13-227307-1
1. Automatic control. 2. Control theory. I. Title.
TJ213.028 1997
629.8—dc20 96-2345
CIP

Publisher: Tom Robbins
Associate editor: Alice Dworkin
Production editor: Ann Marie Longobardo
Cover designer: Bruce Kenselaar
Manufacturing Buyer: Donna Sullivan



Printed in the United States of America

10 9 8

ISBN: 0-13-227307-1

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



The MathWorks, Inc.
24 Prime Park Way
Natick, MA 01760-1500
Phone: (508) 647-7000
Fax: (508) 647-7001
E-mail: info@mathworks.com
WWW: http://www.mathworks.com

# Contents

**Preface** ix

**Chapter 1 Introduction to Control Systems** 1

1-1 Introduction 1
1-2 Examples of Control Systems 3
1-3 Closed-Loop Control Versus Open-Loop Control 6
1-4 Design of Control Systems 8
1-5 Outline of the Book 9
Example Problems and Solutions 10
Problems 11

**Chapter 2 The Laplace Transform** 13

2-1 Introduction 13
2-2 Review of Complex Variables and Complex Functions 14
2-3 Laplace Transformation 17
2-4 Laplace Transform Theorems 27
2-5 Inverse Laplace Transformation 35
2-6 Partial-Fraction Expansion with MATLAB 41
2-7 Solving Linear, Time-Invariant, Differential Equations 44

Example Problems and Solutions 46
Problems 55

Chapter 3 Mathematical Modeling of Dynamic Systems 57

3-1 Introduction 57
3-2 Transfer Function and Impulse-Response Function 60
3-3 Block Diagrams 63
3-4 Modeling in State Space 70
3-5 State-Space Representation of Dynamic Systems 76
3-6 Mechanical Systems 81
3-7 Electrical Systems 87
3-8 Liquid-Level Systems 92
3-9 Thermal Systems 96
3-10 Linearization of Nonlinear Mathematical Models 100
Example Problems and Solutions 105
Problems 129

Chapter 4 Transient-Response Analysis 134

4-1 Introduction 134
4-2 First-Order Systems 136
4-3 Second-Order Systems 141
4-4 Transient-Response Analysis with MATLAB 160
4-5 An Example Problem Solved with MATLAB 178
Example Problems and Solutions 187
Problems 207

Chapter 5 Basic Control Actions and Response of Control Systems 211

5-1 Introduction 211
5-2 Basic Control Actions 212
5-3 Effects of Integral and Derivative Control Actions on System Performance 219
5-4 Higher-Order Systems 228
5-5 Routh's Stability Criterion 232
5-6 Pneumatic Controllers 238
5-7 Hydraulic Controllers 255
5-8 Electronic Controllers 262
5-9 Phase Lead and Phase Lag in Sinusoidal Response 269
5-10 Steady-State Errors in Unity-Feedback Control Systems 274

Example Problems and Solutions 282
Problems 309

Chapter 6 Root-Locus Analysis 317

6-1 Introduction 317
6-2 Root-Locus Plots 319
6-3 Summary of General Rules for Constructing Root Loci 330
6-4 Root-Locus Plots with MATLAB 338
6-5 Special Cases 348
6-6 Root-Locus Analysis of Control Systems 357
6-7 Root Loci for Systems with Transport Lag 360
6-8 Root-Contour Plots 364
Example Problems and Solutions 368
Problems 400

Chapter 7 Control Systems Design by the Root-Locus Method 404

7-1 Introduction 404
7-2 Preliminary Design Considerations 407
7-3 Lead Compensation 409
7-4 Lag Compensation 418
7-5 Lag–Lead Compensation 427
Example Problems and Solutions 439
Problems 467

Chapter 8 Frequency-Response Analysis 471

8-1 Introduction 471
8-2 Bode Diagrams 473
8-3 Plotting Bode Diagrams with MATLAB 492
8-4 Polar Plots 504
8-5 Drawing Nyquist Plots with MATLAB 512
8-6 Log-Magnitude versus Phase Plots 519
8-7 Nyquist Stability Criterion 521
8-8 Stability Analysis 532
8-9 Relative Stability 542
8-10 Closed-Loop Frequency Response 556
8-11 Experimental Determination of Transfer Functions 567
Example Problems and Solutions 573
Problems 605

**Chapter 9 Control Systems Design By Frequency Response** 609

9-1 Introduction 609
9-2 Lead Compensation 612
9-3 Lag Compensation 621
9-4 Lag–Lead Compensation 630
9-5 Concluding Comments 636
Example Problems and Solutions 639
Problems 667

**Chapter 10 PID Controls and Introduction to Robust Control** 669

10-1 Introduction 669
10-2 Tuning Rules for PID Controllers 670
10-3 Modifications of PID Control Schemes 679
10-4 Two-Degrees-of-Freedom Control 683
10-5 Design Considerations for Robust Control 685
Example Problems and Solutions 690
Problems 703

**Chapter 11 Analysis of Control Systems in State Space** 710

11-1 Introduction 710
11-2 State-Space Representations of Transfer-Function Systems 711
11-3 Transformation of System Models with MATLAB 718
11-4 Solving The Time-Invariant State Equation 722
11-5 Some Useful Results in Vector-Matrix Analysis 729
11-6 Controllability 737
11-7 Observability 743
Example Problems and Solutions 749
Problems 783

**Chapter 12 Design of Control Systems in State Space** 786

12-1 Introduction 786
12-2 Pole Placement 787
12-3 Solving Pole-Placement Problems with MATLAB 798
12-4 Design of Regulator-Type Systems by Pole Placement 803
12-5 State Observers 813
12-6 Design of State Observers with MATLAB 837
12-7 Design of Servo Systems 843

12-8 Example of Control System Design with MATLAB 852
Example Problems and Solutions 864
Problems 893

**Chapter 13 Liapunov Stability Analysis and Quadratic Optimal Control** **896**

13-1 Introduction 896
13-2 Liapunov Stability Analysis 897
13-3 Liapunov Stability Analysis of Linear, Time-Invariant Systems 907
13-4 Model-Reference Control Systems 912
13-5 Quadratic Optimal Control 915
13-6 Solving Quadratic Optimal Control Problems with MATLAB 925
Example Problems and Solutions 935
Problems 958

**Appendix Background Materials Necessary for the Effective Use of MATLAB** **960**

A-1 Introduction 960
A-2 Plotting Response Curves 965
A-3 Computing Matrix Functions 967
A-4 Mathematical Models of Linear Systems 977

**References** **983**

**Index** **987**

# Preface

This book is written at the level of the senior engineering student and is intended to be used as a text for the first course in control systems. It presents a comprehensive treatment of the analysis and design of continuous-time control systems. It is assumed that the reader has had courses on introductory differential equations, introductory vector-matrix analysis, introductory circuit analysis, and mechanics.

In this third edition, MATLAB® is integrated into the text. All computational problems are solved with MATLAB. Also, design aspects have been strengthened, and new topics, examples and problems are added.

The text is organized into 13 chapters and an appendix. The outline of the book is as follows: Chapter 1 presents introductory materials on control systems. Chapter 2 gives the Laplace transforms of commonly encountered time functions and basic Laplace transform theorems. (If the students have an adequate background on the Laplace transform, this chapter may be skipped.) Chapter 3 treats mathematical modeling of dynamic systems and develops transfer-function models and state-space models. Chapter 4 gives transient-response analysis of first- and second-order systems. This chapter includes a computational analysis of transient response by use of MATLAB. Chapter 5 presents basic control actions in industrial automatic controllers and discusses pneumatic, hydraulic, and electronic controllers. This chapter also discusses the response of higher-order systems and Routh's stability criterion.

Chapter 6 treats the root-locus analysis. The MATLAB approach to plotting root loci is presented in this chapter. Chapter 7 presents the design of lead, lag and lag–lead compensators by the root-locus method. Chapter 8 deals with the frequency-response analysis of control systems. Bode diagrams, polar plots, the Nyquist stability criterion, and closed-loop frequency response are discussed, including the MATLAB approach to obtain frequency-response plots. Chapter 9 covers the design and compensation techniques using frequency-response methods. Specifically, the Bode diagram approach to the design of lead, lag, and lag–lead compensators is discussed in detail in this chapter. Chapter 10 deals with the basic and modified PID controls. This chapter gives discussions of two-degrees-of-freedom controls and design considerations for robust control.

Chapter 11 presents a basic analysis of control systems in state space. Concepts of controllability and observability are given here. The transformation of system models (from transfer-function model to state-space model, and vice versa) by the use of MATLAB is included in this chapter. Chapter 12 treats the design of control systems in state space. This chapter begins with pole-placement design problems, followed by the design of state observers. A design of a type 1 servo system based on the pole-placement approach is presented, including a computational solution with MATLAB. Chapter 13 begins with Liapunov stability analysis, followed by design of a model-reference control system, where the conditions for Liapunov stability are formulated first and then the system is designed within these limitations. Then quadratic optimal control problems are treated. Here the Liapunov stability equation is used to lead into quadratic optimal control theory. A MATLAB solution to the quadratic optimal control problem is also presented.

No prior knowledge of MATLAB is assumed in this book. If the reader is not yet familiar with MATLAB, it is suggested that he or she read the appendix first and then study MATLAB as presented in the text.

Throughout the book the basic concepts involved are emphasized and highly mathematical arguments are carefully avoided in the presentation of the materials. Mathematical proofs are provided when they contribute to the understanding of the subjects presented. All the material has been organized toward a gradual development of control theory.

Examples are presented at strategic points throughout the book so that the reader will have a better understanding of the subject matter discussed. In addition, a number of solved problems (A-problems) are provided at the end of each chapter. These problems constitute an integral part of the text. It is suggested that the reader study all of these problems carefully to obtain a deeper understanding of the topics discussed. In addition, many unsolved problems (B-problems) are provided for use as homework or quiz problems. An instructor using this text can obtain a complete solutions manual (for B-problems) from the publisher.

Most of the materials presented in this book have been class tested in senior and first-year graduate-level courses on control systems at the University of Minnesota.

If this book is used as a text for a four-hour quarter course (with 40 lecture hours) or a three-hour semester course (with 42 lecture hours), most of the materials in the first 10 chapters may be covered. (The first 10 chapters cover all basic materials of control systems normally required in a first course on control systems.) If this book is used as a text for a four-hour semester course (with 52 lecture hours), a good part of the book may be covered with flexibility in omitting certain subjects. For a two-quarter course sequence (with 60 or more lecture hours), the entire book may be covered. This book can also serve as a self-study book for practicing engineers who wish to study basic materials of control theory.

I would like to express my sincere appreciation to Professor Suhada Jayasuriya of Texas A & M University, who reviewed the final manuscript and made many constructive comments. Appreciation is also due to Linda Ratts Engelman for her enthusiasm in publishing the third edition, to the anonymous reviewers who made valuable suggestions at the early stage of the revision process, and to my former students who solved many of the A-problems and B-problems included in this book.

*Katsuhiko Ogata*

# 1

# Introduction to Control Systems

## 1-1 INTRODUCTION

Automatic control has played a vital role in the advance of engineering and science. In addition to its extreme importance in space-vehicle systems, missile-guidance systems, robotic systems, and the like, automatic control has become an important and integral part of modern manufacturing and industrial processes. For example, automatic control is essential in the numerical control of machine tools in the manufacturing industries, in the design of autopilot systems in the aerospace industries, and in the design of cars and trucks in the automobile industries. It is also essential in such industrial operations as controlling pressure, temperature, humidity, viscosity, and flow in the process industries.

Since advances in the theory and practice of automatic control provide the means for attaining optimal performance of dynamic systems, improving productivity, relieving the drudgery of many routine repetitive manual operations, and more, most engineers and scientists must now have a good understanding of this field.

**Historical review.** 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 made it possible for engineers to design linear closed-loop control systems that satisfied performance requirements. From the end of the 1940s to early 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 1980 to the present, developments in modern control theory centered around robust control, $H_\infty$ control, and associated topics.

Now that digital computers have become cheaper and more compact, they are used as integral parts of control systems. Recent applications of modern control theory include such nonengineering systems as biological, biomedical, economic, and socioeconomic systems.

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

***Controlled Variable and Manipulated Variable.*** The *controlled* variable is the quantity or condition that is measured and controlled. The *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 manipulated variable 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 artifical 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 is not limited to physical ones. 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 that 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 several 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

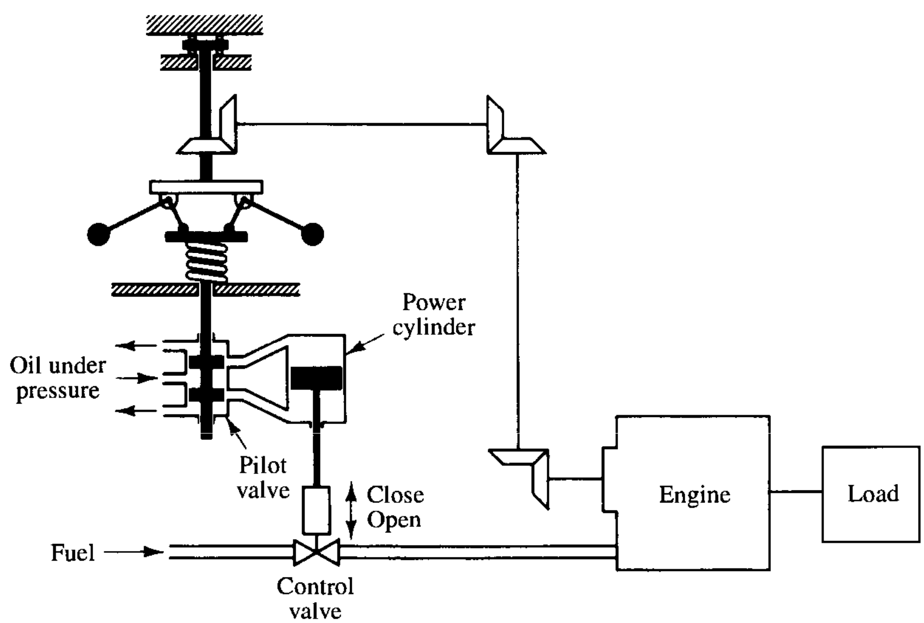



**Figure 1–1**
Speed control system.

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.

**Robot control system.** Industrial robots are frequently used in industry to improve productivity. The robot can handle monotonous jobs as well as complex jobs without errors in operation. The robot can work in an environment intolerable to human operators. For example, it can work in extreme temperatures (both high and low) or in a high- or low-pressure environment or under water or in space. There are special robots for fire fighting, underwater exploration, and space exploration, among many others.

The industrial robot must handle mechanical parts that have particular shapes and weights. Hence, it must have at least an arm, a wrist, and a hand. It must have sufficient power to perform the task and the capability for at least limited mobility. In fact, some robots of today are able to move freely by themselves in a limited space in a factory.

The industrial robot must have some sensory devices. In low-level robots, microswitches are installed in the arms as sensory devices. The robot first touches an object and then, through the microswitches, confirms the existence of the object in space and proceeds in the next step to grasp it.

In a high-level robot, an optical means (such as a television system) is used to scan the background of the object. It recognizes the pattern and determines the presence and orientation of the object. A computer is necessary to process signals in the pattern-recognition process (see Figure 1–2). In some applications, the computerized robot recognizes the presence and orientation of each mechanical part by a pattern-recognition process that consists of reading the code numbers attached to it. Then the robot picks up the part and moves it to an appropriate place for assembling, and there it assembles several parts into a component. A well-programmed digital computer acts as a controller.

**Temperature control system.** Figure 1–3 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 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

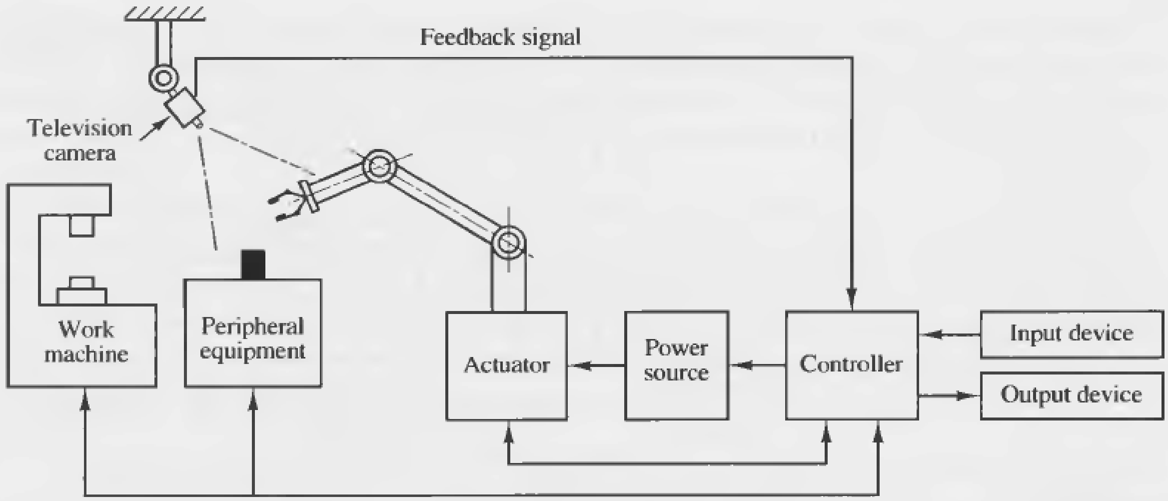



**Figure 1–2**
Robot using a pattern-recognition process.

the heater, through an interface, amplifier, and relay, to bring the furnace temperature to a desired value.

**Temperature control of the passenger compartment of a car.** Figure 1–4 shows a functional diagram of temperature control of the passenger compartment of a car. The desired temperature, converted to a voltage, is the input to the controller. The actual temperature of the passenger compartment is converted to a voltage through a sensor and is fed back to the controller for comparison with the input. The ambient temperature and radiation heat transfer from the sun, which are not constant while the car is driven, act as disturbances. This system employs both feedback control and feedforward control. (Feedforward control gives corrective action before the disturbances affect the output.)

The temperature of the passenger car compartment differs considerably depending on the place where it is measured. Instead of using multiple sensors for temperature measurement and averaging the measured values, it is economical to install a small suction blower at the place where passengers normally sense the temperature. The temperature of the air from the suction blower is an indication of the passenger compartment temperature and is considered the output of the system.

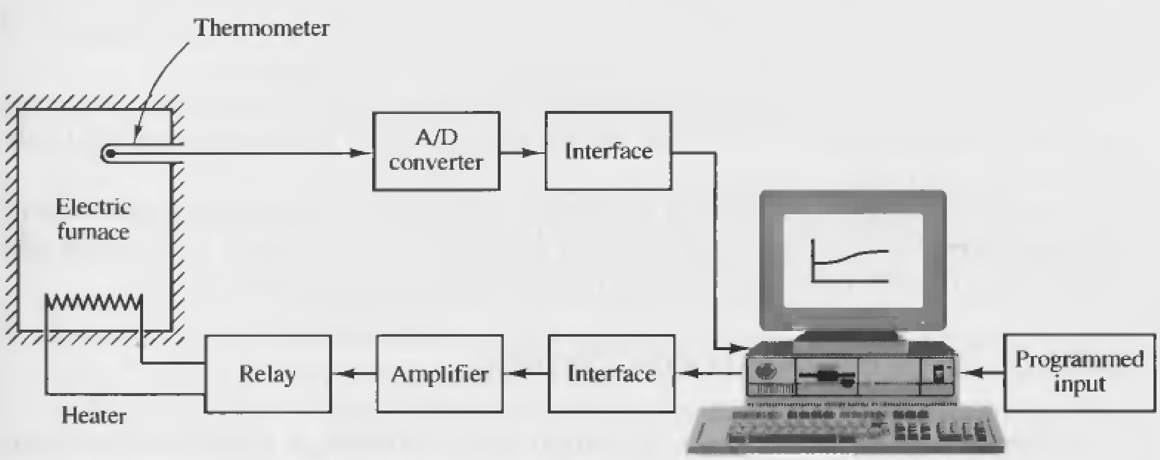



**Figure 1–3**
Temperature control system.

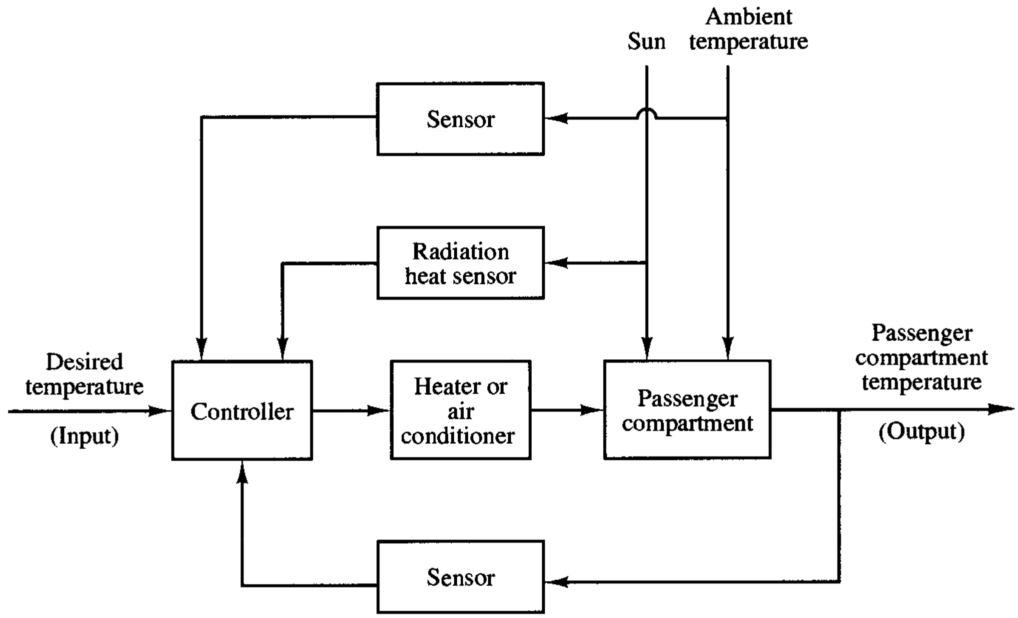



**Figure 1–4** Temperature control of passenger compartment of a car.

The controller receives the input signal, output signal, and signals from sensors from disturbance sources. The controller sends out an optimal control signal to the air conditioner or heater to control the amount of cooling air or warm air so that the passenger compartment temperature is about the desired temperature.

**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. (Note that 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.

## 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 that 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.

## 1–4 DESIGN OF CONTROL SYSTEMS

Actual control systems are generally nonlinear. However, if they can be approximated by linear mathematical models, we may use one or more of the well-developed design methods. In a practical sense, the performance specifications given to the particular system suggest which method to use. If the performance specifications are given in terms of transient-response characteristics and/or frequency-domain performance measures, then we have no choice but to use a conventional approach based on the root-locus and/or frequency-response methods. (These methods are presented in Chapters 6 through 9.) If the performance specifications are given as performance indexes in terms of state variables, then modern control approaches should be used. (These approaches are presented in Chapters 11 through 13.)

While control system design via the root-locus and frequency-response approaches is an engineering endeavor, system design in the context of modern control theory (state-space methods) employs mathematical formulations of the problem and applies mathematical theory to design problems in which the system can have multiple inputs and multiple outputs and can be time varying. By applying modern control theory, the designer is able to start from a performance index, together with constraints imposed on the system, and to proceed to design a stable system by a completely analytical procedure. The advantage of design based on such modern control theory is that it enables the designer to produce a control system that is optimal with respect to the performance index considered.

The systems that may be designed by a conventional approach are usually limited to single-input–single-output, linear time-invariant systems. The designer seeks to satisfy all performance specifications by means of educated trial-and-error repetition. After a system is designed, the designer checks to see if the designed system satisfies all the performance specifications. If it does not, then he repeats the design process by adjusting parameter settings or by changing the system configuration until the given specifications are met. Although the design is based on a trial-and-error procedure, the ingenuity and know-how of the designer will play an important role in a successful design. An experienced designer may be able to design an acceptable system without using many trials.

It is generally desirable that the designed system should exhibit as small errors as possible in responding to the input signal. In this regard, the damping of the system should be reasonable. The system dynamics should be relatively insensitive to small changes in system parameters. The undesirable disturbances should be well attenuated. [In general, the high-frequency portion should attenuate fast so that high-frequency noises (such as sensor noises) can be attenuated. If the noise or disturbance frequencies are known, notch filters may be used to attenuate these specific frequencies.] If the design of the system is boiled down to a few candidates, an optimal choice among them may be made from such considerations as projected overall performance, cost, space, and weight.

## 1–5 OUTLINE OF THE BOOK

In what follows we shall briefly present the arrangements and contents of the book.

Chapter 1 has given introductory materials on control systems. Chapter 2 presents basic Laplace transform theory necessary for understanding the control theory presented in this book. Chapter 3 deals with mathematical modeling of dynamic systems in terms of transfer functions and state-space equations. This chapter includes discussions of linearization of nonlinear systems. Chapter 4 treats transient-response analyses of first- and second-order systems. This chapter also gives details of transient-response analysis with MATLAB. Chapter 5 first presents basic control actions and then discusses pneumatic, hydraulic, and electronic controllers. This chapter also discusses Routh's stability criterion.

Chapter 6 gives a root-locus analysis of control systems. General rules for constructing root loci are presented. Detailed discussions for plotting root loci with MATLAB are included. Chapter 7 deals with the design of control systems via the root-locus method. Specifically, root-locus approaches to the design of lead compensators, lag compensators, and lag–lead compensators are discussed in detail. Chapter 8 gives the frequency-response analysis of control systems. Bode diagrams, polar plots, Nyquist stability criterion, and closed-loop frequency response are discussed. Chapter 9 treats control systems design via the frequency-response approach. Here Bode diagrams are used to design lead compensators, lag compensators, and lag–lead compensators. Chapter 10 discusses the basic and modified PID controls. Topics included are tuning rules for PID controllers, modifications of PID control schemes, two-degrees-of-freedom control, and design considerations for robust control.

Chapter 11 presents basic materials for the state-space analysis of control systems. The solution of the time-invariant state equation is derived and concepts of controllability and observability are discussed. Chapter 12 treats the design of control systems in state space. This chapter begins with the pole-placement problems, followed by the design of state observers, and concludes with the design of type 1 servo systems. MATLAB is utilized in solving pole-placement problems, design of state observers, and design of servo systems. Chapter 13, the final chapter, presents the Liapunov stability analysis and the quadratic optimal control. This chapter begins with the Liapunov stability analysis. Then the Liapunov stability approach is used for designing model-reference control systems. Finally, quadratic optimal control problems are discussed in detail. Here the Liapunov stability approach is utilized to derive the Riccati equation for quadratic optimal control. MATLAB solutions to quadratic optimal control problems are included.

The appendix summarizes background materials necessary for the effective use of MATLAB. This appendix is specifically provided for those readers who are as yet unfamiliar with MATLAB.

## EXAMPLE PROBLEMS AND SOLUTIONS

**A-1-1.** List the major advantages and disadvantages of open-loop control systems.

**Solution.** The 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 economically not feasible. (For example, in the washer system, it would be quite expensive to provide a device to measure the quality of the output, cleanliness of the clothes, of the washer.)

The 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.

**A-1-2.** Figure 1–5(a) is a schematic diagram of a liquid-level control system. Here the automatic controller maintains the liquid level by comparing the actual level with a desired level and correcting any error by adjusting the opening of the pneumatic valve. Figure 1–5(b) is a block diagram of the control system. Draw the corresponding block diagram for a human-operated liquid-level control system.

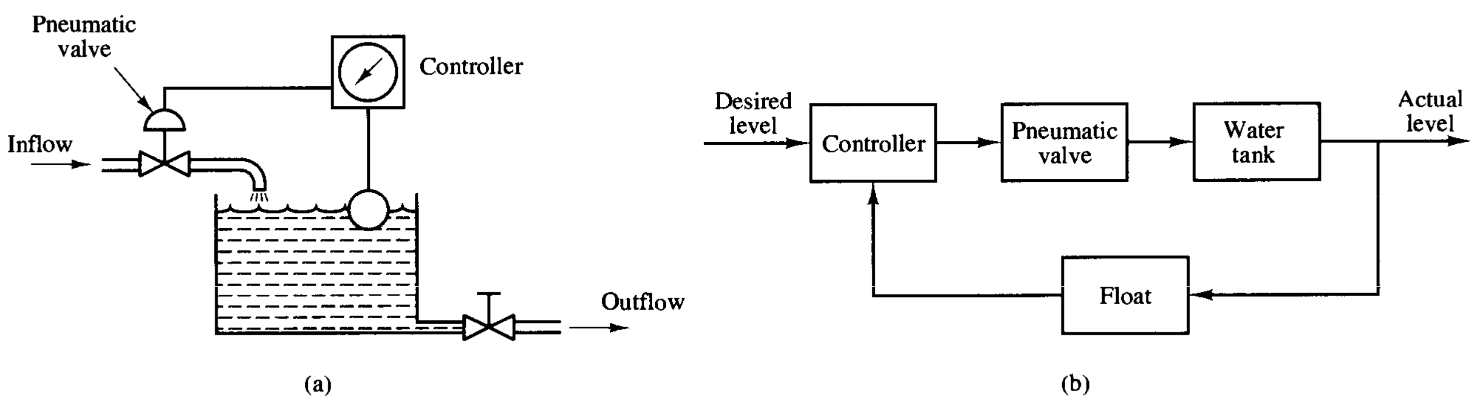



**Figure 1–5**
(a) Liquid-level control system; (b) block diagram.

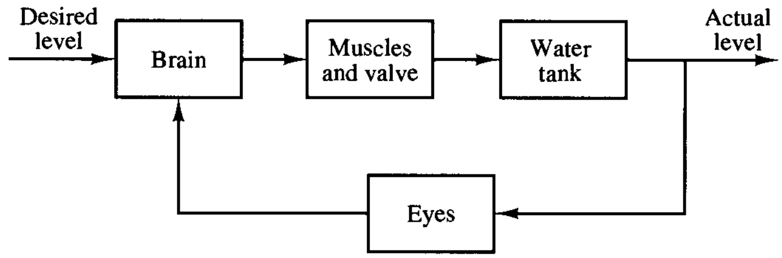



**Figure 1–6**
Block diagram of human-operated liquid-level control system.

**Solution.** In the human-operated system, the eyes, brain, and muscles correspond to the sensor, controller, and pneumatic valve, respectively. A block diagram is shown in Figure 1–6.

**A-1-3.** An engineering organizational system is composed of major groups, such as management, research and development, preliminary design, experiments, product design and drafting, fabrication and assembling, and testing. These groups are interconnected to make up the whole operation.

The 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.)

Draw a functional block diagram showing an engineering organizational system.

**Solution.** A functional block diagram can 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. A possible block diagram is shown in Figure 1–7.

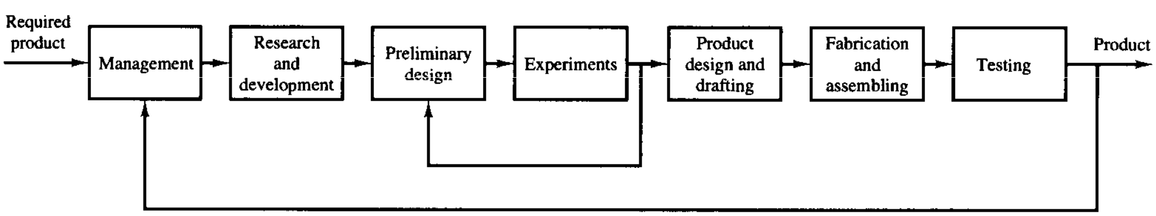



**Figure 1–7**
Block diagram of an engineering organizational system.

## PROBLEMS

**B-1-1.** Many closed-loop and open-loop control systems may be found in homes. List several examples and describe them.

**B-1-2.** Give two examples of feedback control systems in which a human acts as a controller.

**B-1-3.** Figure 1–8 shows a tension control system. Explain the sequence of control actions when the feed speed is suddenly changed for a short time.

**B-1-4.** Many machines, such as lathes, milling machines, and grinders, are provided with tracers to reproduce the contour of templates. Figure 1–9 shows a schematic diagram of a tracing system in which the tool duplicates the shape of the template on the workpiece. Explain the operation of this system.

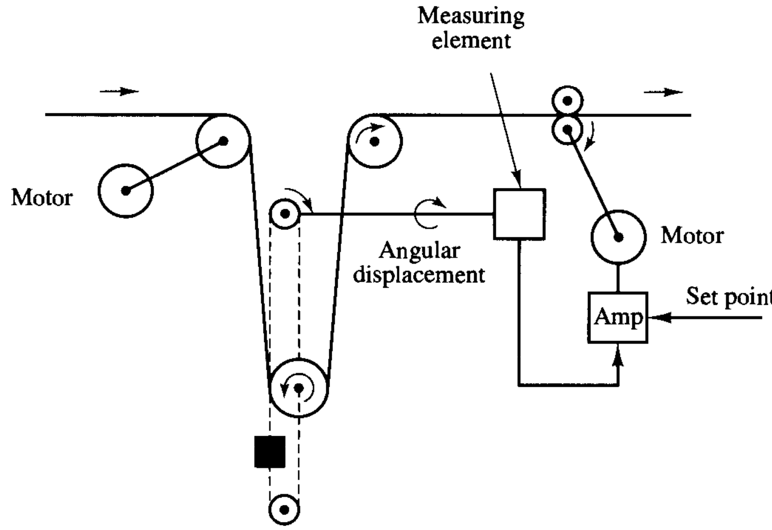



**Figure 1–8**
Tension control system.

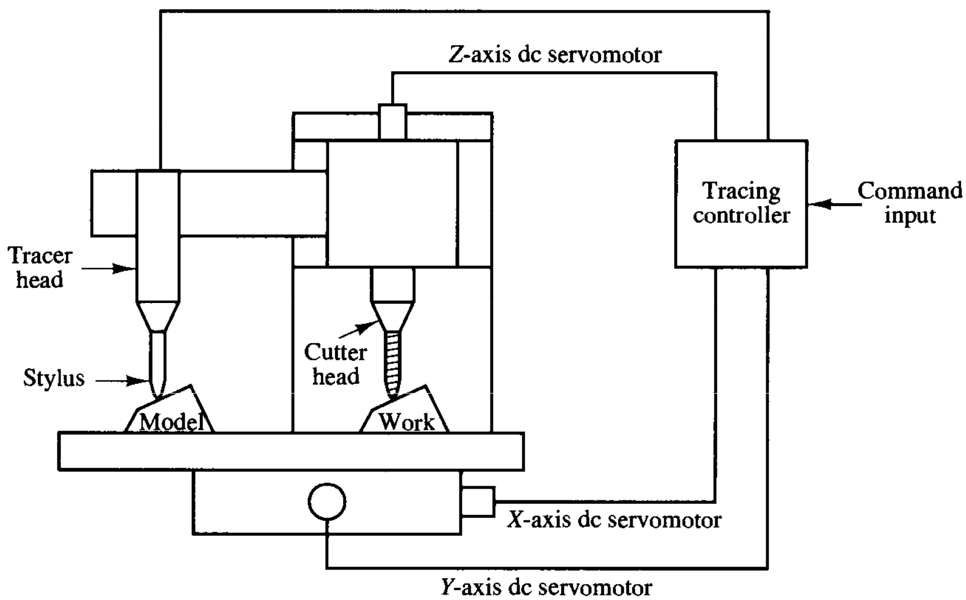



**Figure 1–9**
Schematic diagram of a tracing system.

# 2

# The Laplace Transform*

## 2–1 INTRODUCTION

The Laplace transform method is an operational method that can be used advantageously for solving linear differential equations. By use of Laplace transforms, we can convert many common functions, such as sinusoidal functions, damped sinusoidal functions, and exponential functions, into algebraic functions of a complex variable $s$. Operations such as differentiation and integration can be replaced by algebraic operations in the complex plane. Thus, a linear differential equation can be transformed into an algebraic equation in a complex variable $s$. If the algebraic equation in $s$ is solved for the dependent variable, then the solution of the differential equation (the inverse Laplace transform of the dependent variable) may be found by use of a Laplace transform table or by use of the partial-fraction expansion technique, which is presented in Section 2–5.

An advantage of the Laplace transform method is that it allows the use of graphical techniques for predicting the system performance without actually solving system differential equations. Another advantage of the Laplace transform method is that, when we solve the differential equation, both the transient component and steady-state component of the solution can be obtained simultaneously.

**Outline of the chapter.** Section 2–1 presents introductory remarks. Section 2–2 briefly reviews complex variables and complex functions. Section 2–3 derives Laplace

*This chapter may be skipped if the student is already familiar with Laplace transforms.

transforms of time functions that are frequently used in control engineering. Section 2–4 presents useful theorems of Laplace transforms, and Section 2–5 treats the inverse Laplace transformation. Section 2–6 presents the MATLAB approach to obtain partial-fraction expansion of $B(s)/A(s)$, where $A(s)$ and $B(s)$ are polynomials in $s$. Finally, Section 2–7 deals with solutions of linear time-invariant differential equations by the Laplace transform approach.

## 2–2 REVIEW OF COMPLEX VARIABLES AND COMPLEX FUNCTIONS

Before we present the Laplace transformation, we shall review the complex variable and complex function. We shall also review Euler's theorem, which relates the sinusoidal functions to exponential functions.

**Complex variable.** A complex number has a real part and an imaginary part, both of which are constant. If the real part and/or imaginary part are variables, a complex number is called a *complex variable.* In the Laplace transformation we use the notation $s$ as a complex variable; that is,

$$s = \sigma + j\omega$$

where $\sigma$ is the real part and $\omega$ is the imaginary part.

**Complex function.** A complex function $F(s)$, a function of $s$, has a real part and an imaginary part or

$$F(s) = F_x + jF_y$$

where $F_x$ and $F_y$ are real quantities. The magnitude of $F(s)$ is $\sqrt{F_x^2 + F_y^2}$, and the angle $\theta$ of $F(s)$ is $\tan^{-1}(F_y/F_x)$. The angle is measured counterclockwise from the positive real axis. The complex conjugate of $F(s)$ is $\bar{F}(s) = F_x - jF_y$.

Complex functions commonly encountered in linear control systems analysis are single-valued functions of $s$ and are uniquely determined for a given value of $s$.

A complex function $G(s)$ is said to be *analytic* in a region if $G(s)$ and all its derivatives exist in that region. The derivative of an analytic function $G(s)$ is given by

$$\frac{d}{ds}G(s) = \lim_{\Delta s \to 0} \frac{G(s + \Delta s) - G(s)}{\Delta s} = \lim_{\Delta s \to 0} \frac{\Delta G}{\Delta s}$$

Since $\Delta s = \Delta\sigma + j\Delta\omega$, $\Delta s$ can approach zero along an infinite number of different paths. It can be shown, but is stated without a proof here, that if the derivatives taken along two particular paths, that is, $\Delta s = \Delta\sigma$ and $\Delta s = j\Delta\omega$, are equal, then the derivative is unique for any other path $\Delta s = \Delta\sigma + j\Delta\omega$ and so the derivative exists.

For a particular path $\Delta s = \Delta\sigma$ (which means that the path is on the real axis),

$$\frac{d}{ds}G(s) = \lim_{\Delta\sigma \to 0}\left(\frac{\Delta G_x}{\Delta\sigma} + j\frac{\Delta G_y}{\Delta\sigma}\right) = \frac{\partial G_x}{\partial\sigma} + j\frac{\partial G_y}{\partial\sigma}$$

For another particular path $\Delta s = j\Delta\omega$ (which means that the path is on the imaginary axis),

$$\frac{d}{ds}G(s) = \lim_{j\Delta\omega \to 0}\left(\frac{\Delta G_x}{j\Delta\omega} + j\frac{\Delta G_y}{j\Delta\omega}\right) = -j\frac{\partial G_x}{\partial\omega} + \frac{\Delta G_y}{\partial\omega}$$

If these two values of the derivative are equal,

$$\frac{\partial G_x}{\partial\sigma} + j\frac{\partial G_y}{\partial\sigma} = \frac{\partial G_y}{\partial\omega} - j\frac{\partial G_x}{\partial\omega}$$

or if the following two conditions

$$\frac{\partial G_x}{\partial\sigma} = \frac{\partial G_y}{\partial\omega} \quad \text{and} \quad \frac{\partial G_y}{\partial\sigma} = -\frac{\partial G_x}{\partial\omega}$$

are satisfied, then the derivative $dG(s)/ds$ is uniquely determined. These two conditions are known as the Cauchy–Riemann conditions. If these conditions are satisfied, the function $G(s)$ is analytic.

As an example, consider the following $G(s)$:

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

Then

$$G(\sigma + j\omega) = \frac{1}{\sigma + j\omega + 1} = G_x + jG_y$$

where

$$G_x = \frac{\sigma + 1}{(\sigma+1)^2 + \omega^2} \quad \text{and} \quad G_y = \frac{-\omega}{(\sigma+1)^2 + \omega^2}$$

It can be seen that, except at $s = -1$ (that is, $\sigma = -1, \omega = 0$), $G(s)$ satisfies the Cauchy–Riemann conditions:

$$\frac{\partial G_x}{\partial\sigma} = \frac{\partial G_y}{\partial\omega} = \frac{\omega^2 - (\sigma+1)^2}{[(\sigma+1)^2 + \omega^2]^2}$$

$$\frac{\partial G_y}{\partial\sigma} = -\frac{\partial G_x}{\partial\omega} = \frac{2\omega(\sigma+1)}{[(\sigma+1)^2 + \omega^2]^2}$$

Hence $G(s) = 1/(s+1)$ is analytic in the entire $s$ plane except at $s = -1$. The derivative $dG(s)/ds$, except at $s = 1$, is found to be

$$\frac{d}{ds}G(s) = \frac{\partial G_x}{\partial\sigma} + j\frac{\partial G_y}{\partial\sigma} = \frac{\partial G_y}{\partial\omega} - j\frac{\partial G_x}{\partial\omega}$$

$$= -\frac{1}{(\sigma + j\omega + 1)^2} = -\frac{1}{(s+1)^2}$$

Note that the derivative of an analytic function can be obtained simply by differentiating $G(s)$ with respect to $s$. In this example,

$$\frac{d}{ds}\left(\frac{1}{s+1}\right) = -\frac{1}{(s+1)^2}$$

Points in the $s$ plane at which the function $G(s)$ is analytic are called *ordinary* points, while points in the $s$ plane at which the function $G(s)$ is not analytic are called *singular* points. Singular points at which the function $G(s)$ or its derivatives approach infinity are called *poles*. In the previous example, $s = -1$ is a singular point and is a pole of the function $G(s)$.

If $G(s)$ approaches infinity as $s$ approaches $-p$ and if the function

$$G(s)(s+p)^n, \qquad \text{for } n = 1, 2, 3, \ldots$$

has a finite, nonzero value at $s = -p$, then $s = -p$ is called a pole of order $n$. If $n = 1$, the pole is called a simple pole. If $n = 2, 3, \ldots$, the pole is called a second-order pole, a third-order pole, and so on. Points at which the function $G(s)$ equals zero are called *zeros*.

To illustrate, consider the complex function

$$G(s) = \frac{K(s+2)(s+10)}{s(s+1)(s+5)(s+15)^2}$$

$G(s)$ has zeros at $s = -2, s = -10$, simple poles at $s = 0, s = -1, s = -5$, and a double pole (multiple pole of order 2) at $s = -15$. Note that $G(s)$ becomes zero at $s = \infty$. Since for large values of $s$

$$G(s) \doteqdot \frac{K}{s^3}$$

$G(s)$ possesses a triple zero (multiple zero of order 3) at $s = \infty$. If points at infinity are included, $G(s)$ has the same number of poles as zeros. To summarize, $G(s)$ has five zeros ($s = -2, s = -10, s = \infty, s = \infty, s = \infty$) and five poles ($s = 0, s = -1, s = -5, s = -15, s = -15$).

**Euler's theorem.** The power series expansions of $\cos\theta$ and $\sin\theta$ are, respectively,

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots$$

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots$$

And so

$$\cos\theta + j\sin\theta = 1 + (j\theta) + \frac{(j\theta)^2}{2!} + \frac{(j\theta)^3}{3!} + \frac{(j\theta)^4}{4!} + \cdots$$

Since

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

we see that

$$\cos\theta + j\sin\theta = e^{j\theta} \tag{2–1}$$

This is known as *Euler's theorem.*

By using Euler's theorem, we can express sine and cosine in terms of an exponential function. Noting that $e^{-j\theta}$ is the complex conjugate of $e^{j\theta}$ and that

$$e^{j\theta} = \cos\theta + j\sin\theta$$

$$e^{-j\theta} = \cos\theta - j\sin\theta$$

we find, after adding and subtracting these two equations, that

$$\cos\theta = \frac{1}{2}(e^{j\theta} + e^{-j\theta}) \tag{2–2}$$

$$\sin\theta = \frac{1}{2j}(e^{j\theta} - e^{-j\theta}) \tag{2–3}$$

## 2–3 LAPLACE TRANSFORMATION

We shall first present a definition of the Laplace transformation and a brief discussion of the condition for the existence of the Laplace transform and then provide examples for illustrating the derivation of Laplace transforms of several common functions.

Let us define

$f(t)$ = a function of time $t$ such that $f(t) = 0$ for $t < 0$

$s$ = a complex variable

$\mathscr{L}$ = an operational symbol indicating that the quantity that it prefixes is to be transformed by the Laplace integral $\int_0^\infty e^{-st}\,dt$

$F(s)$ = Laplace transform of $f(t)$

Then the Laplace transform of $f(t)$ is given by

$$\mathscr{L}[f(t)] = F(s) = \int_0^\infty e^{-st}\,dt[f(t)] = \int_0^\infty f(t)e^{-st}\,dt$$

The reverse process of finding the time function $f(t)$ from the Laplace transform $F(s)$ is called the *inverse Laplace transformation.* The notation for the inverse Laplace transformation is $\mathscr{L}^{-1}$, and the inverse Laplace transform can be found from $F(s)$ by the following inversion integral:

$$\mathscr{L}^{-1}[F(s)] = f(t) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F(s)e^{st}\,ds, \qquad \text{for } t > 0 \tag{2–4}$$

where $c$, the abscissa of convergence, is a real constant and is chosen larger than the real parts of all singular points of $F(s)$. Thus, the path of integration is parallel to the $j\omega$ axis and is displaced by the amount $c$ from it. This path of integration is to the right of all singular points.

Evaluating the inversion integral appears complicated. In practice, we seldom use this integral for finding $f(t)$. There are simpler methods for finding $f(t)$. We shall discuss such simpler methods in Section 2–5.

It is noted that in this book the time function $f(t)$ is always assumed to be zero for negative time; that is,

$$f(t) = 0, \qquad \text{for } t < 0$$

**Existence of Laplace transform.** The Laplace transform of a function $f(t)$ exists if the Laplace integral converges. The integral will converge if $f(t)$ is sectionally continuous in every finite interval in the range $t > 0$ and if it is of exponential order as $t$ approaches infinity. A function $f(t)$ is said to be of exponential order if a real, positive constant $\sigma$ exists such that the function

$$e^{-\sigma t}|f(t)|$$

approaches zero as $t$ approaches infinity. If the limit of the function $e^{-\sigma t}|f(t)|$ approaches zero for $\sigma$ greater than $\sigma_c$ and the limit approaches infinity for $\sigma$ less than $\sigma_c$, the value $\sigma_c$ is called the *abscissa of convergence.*

For the function $f(t) = Ae^{-\alpha t}$

$$\lim_{t\to\infty} e^{-\sigma t}|Ae^{-\alpha t}|$$

approaches zero if $\sigma > -\alpha$. The abscissa of convergence in this case is $\sigma_c = -\alpha$. The integral $\int_0^\infty f(t)e^{-st}\,dt$ converges only if $\sigma$, the real part of $s$, is greater than the abscissa of convergence $\sigma_c$. Thus the operator $s$ must be chosen as a constant such that this integral converges.

In terms of the poles of the function $F(s)$, the abscissa of convergence $\sigma_c$ corresponds to the real part of the pole located farthest to the right in the $s$ plane. For example, for the following function $F(s)$,

$$F(s) = \frac{K(s+3)}{(s+1)(s+2)}$$

the abscissa of convergence $\sigma_c$ is equal to $-1$. It can be seen that for such functions as $t$, $\sin \omega t$, and $t \sin \omega t$ the abscissa of convergence is equal to zero. For functions like $e^{-ct}$, $te^{-ct}$, $e^{-ct}\sin \omega t$, and so on, the abscissa of convergence is equal to $-c$. For functions that increase faster than the exponential function, however, it is impossible to find suitable values of the abcissa of convergence. Therefore, such functions as $e^{t^2}$ and $te^{t^2}$ do not possess Laplace transforms.

The reader should be cautioned that although $e^{t^2}$ (for $0 \le t \le \infty$) does not possess a Laplace transform, the time function defined by

$$\begin{aligned} f(t) &= e^{t^2}, & &\text{for } 0 \le t \le T < \infty \\ &= 0, & &\text{for } t < 0,\ T < t \end{aligned}$$

does possess a Laplace transform since $f(t) = e^{t^2}$ for only a limited time interval $0 \le t \le T$ and not for $0 \le t \le \infty$. Such a signal can be physically generated. Note that the signals that we can physically generate always have corresponding Laplace transforms.

If a function $f(t)$ has a Laplace transform, then the Laplace transform of $Af(t)$, where $A$ is a constant, is given by

$$\mathcal{L}[Af(t)] = A\mathcal{L}[f(t)]$$

This is obvious from the definition of the Laplace transform. Similarly, if functions $f_1(t)$ and $f_2(t)$ have Laplace transforms, then the Laplace transform of the function $f_1(t) + f_2(t)$ is given by

$$\mathscr{L}[f_1(t) + f_2(t)] = \mathscr{L}[f_1(t)] + \mathscr{L}[f_2(t)]$$

Again the proof of this relationship is evident from the definition of the Laplace transform.

In what follows, we derive Laplace transforms of a few commonly encountered functions.

**Exponential function.** Consider the exponential function

$$\begin{aligned} f(t) &= 0, & \text{for } t < 0 \\ &= Ae^{-\alpha t}, & \text{for } t \geq 0 \end{aligned}$$

where $A$ and $\alpha$ are constants. The Laplace transform of this exponential function can be obtained as follows:

$$\mathscr{L}[Ae^{-\alpha t}] = \int_0^\infty Ae^{-\alpha t}e^{-st}\,dt = A\int_0^\infty e^{-(\alpha+s)t}\,dt = \frac{A}{s+\alpha}$$

It is seen that the exponential function produces a pole in the complex plane.

In deriving the Laplace transform of $f(t) = Ae^{-\alpha t}$, we required that the real part of $s$ be greater than $-\alpha$ (the abscissa of convergence). A question may immediately arise as to whether or not the Laplace transform thus obtained is valid in the range where $\sigma < -\alpha$ in the $s$ plane. To answer this question, we must resort to the theory of complex variables. In the theory of complex variables, there is a theorem known as the analytic extension theorem. It states that, if two analytic functions are equal for a finite length along any arc in a region in which both are analytic, then they are equal everywhere in the region. The arc of equality is usually the real axis or a portion of it. By using this theorem the form of $F(s)$ determined by an integration in which $s$ is allowed to have any real positive value greater than the abscissa of convergence holds for any complex values of $s$ at which $F(s)$ is analytic. Thus, although we require the real part of $s$ to be greater than the abscissa of convergence to make the $\int_0^\infty f(t)e^{-st}\,dt$ absolutely convergent, once the Laplace transform $F(s)$ is obtained, $F(s)$ can be considered valid throughout the entire $s$ plane except at the poles of $F(s)$.

**Step function.** Consider the step function

$$\begin{aligned} f(t) &= 0, & \text{for } t < 0 \\ &= A, & \text{for } t > 0 \end{aligned}$$

where $A$ is a constant. Note that it is a special case of the exponential function $Ae^{-\alpha t}$, where $\alpha = 0$. The step function is undefined at $t = 0$. Its Laplace transform is given by

$$\mathscr{L}[A] = \int_0^\infty Ae^{-st}\,dt = \frac{A}{s}$$

In performing this integration, we assumed that the real part of $s$ was greater than zero (the abscissa of convergence) and therefore that $\lim_{t\to\infty} e^{-st}$ was zero. As stated earlier, the Laplace transform thus obtained is valid in the entire $s$ plane except at the pole $s = 0$.

The step function whose height is unity is called *unit-step* function. The unit-step function that occurs at $t = t_0$ is frequently written as $1(t - t_0)$. The step function of height $A$ that occurs at $t = 0$ can then be written as $f(t) = A1(t)$. The Laplace transform of the unit-step function, which is defined by

$$\begin{aligned} 1(t) &= 0, \qquad \text{for } t < 0 \\ &= 1, \qquad \text{for } t > 0 \end{aligned}$$

is $1/s$, or

$$\mathscr{L}[1(t)] = \frac{1}{s}$$

Physically, a step function occurring at $t = 0$ corresponds to a constant signal suddenly applied to the system at time $t$ equals zero.

**Ramp function.** Consider the ramp function

$$\begin{aligned} f(t) &= 0, \qquad \text{for } t < 0 \\ &= At, \qquad \text{for } t \geq 0 \end{aligned}$$

where $A$ is a constant. The Laplace transform of this ramp function is obtained as

$$\begin{aligned} \mathscr{L}[At] &= \int_0^\infty Ate^{-st}\,dt = At\frac{e^{-st}}{-s}\bigg|_0^\infty - \int_0^\infty \frac{Ae^{-st}}{-s}\,dt \\ &= \frac{A}{s}\int_0^\infty e^{-st}\,dt = \frac{A}{s^2} \end{aligned}$$

**Sinusoidal function.** The Laplace transform of the sinusoidal function

$$\begin{aligned} f(t) &= 0, \qquad \text{for } t < 0 \\ &= A\sin\omega t, \qquad \text{for } t \geq 0 \end{aligned}$$

where $A$ and $\omega$ are constants, is obtained as follows. Referring to Equation (2–3), $\sin \omega t$ can be written

$$\sin\omega t = \frac{1}{2j}(e^{j\omega t} - e^{-j\omega t})$$

Hence

$$\begin{aligned} \mathscr{L}[A\sin\omega t] &= \frac{A}{2j}\int_0^\infty (e^{j\omega t} - e^{-j\omega t})e^{-st}\,dt \\ &= \frac{A}{2j}\frac{1}{s - j\omega} - \frac{A}{2j}\frac{1}{s + j\omega} = \frac{A\omega}{s^2 + \omega^2} \end{aligned}$$

Similarly, the Laplace transform of $A \cos \omega t$ can be derived as follows:

$$\mathscr{L}[A\cos\omega t] = \frac{As}{s^2 + \omega^2}$$

**Comments.** The Laplace transform of any Laplace transformable function $f(t)$ can be found by multiplying $f(t)$ by $e^{-st}$ and then integrating the product from $t = 0$ to $t = \infty$. Once we know the method of obtaining the Laplace transform, however, it is not necessary to derive the Laplace transform of $f(t)$ each time. Laplace transform tables can conveniently be used to find the transform of a given function $f(t)$. Table 2–1 shows Laplace transforms of time functions that will frequently appear in linear control systems analysis.

In the following discussion we present Laplace transforms of functions as well as theorems on the Laplace transformation that are useful in the study of linear control systems.

**Translated function.** Let us obtain the Laplace transform of the translated function $f(t-\alpha)1(t-\alpha)$, where $\alpha \geq 0$. This function is zero for $t < \alpha$. The functions $f(t)1(t)$ and $f(t-\alpha)1(t-\alpha)$ are shown in Figure 2–1.

By definition, the Laplace transform of $f(t-\alpha)1(t-\alpha)$ is

$$\mathscr{L}[f(t-\alpha)1(t-\alpha)] = \int_0^\infty f(t-\alpha)1(t-\alpha)e^{-st}\,dt$$

By changing the independent variable from $t$ to $\tau$, where $\tau = t - \alpha$, we obtain

$$\int_0^\infty f(t-\alpha)1(t-\alpha)e^{-st}\,dt = \int_{-\alpha}^\infty f(\tau)1(\tau)e^{-s(\tau+\alpha)}\,d\tau$$

Since in this book we always assume that $f(t) = 0$ for $t < 0$, $f(\tau)1(\tau) = 0$ for $\tau < 0$. Hence we can change the lower limit of integration from $-\alpha$ to 0. Thus

$$\begin{aligned}\int_{-\alpha}^\infty f(\tau)1(\tau)e^{-s(\tau+\alpha)}\,d\tau &= \int_0^\infty f(\tau)1(\tau)e^{-s(\tau+\alpha)}\,d\tau \\ &= \int_0^\infty f(\tau)e^{-s\tau}e^{-\alpha s}\,d\tau \\ &= e^{-\alpha s}\int_0^\infty f(\tau)e^{-s\tau}\,d\tau = e^{-\alpha s}F(s)\end{aligned}$$

where

$$F(s) = \mathscr{L}[f(t)] = \int_0^\infty f(t)e^{-st}\,dt$$

And so

$$\mathscr{L}[f(t-\alpha)1(t-\alpha)] = e^{-\alpha s}F(s), \qquad \text{for } \alpha \geq 0$$

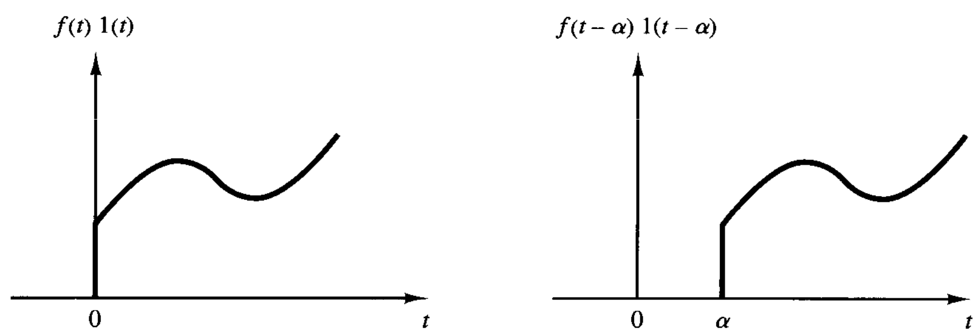



**Figure 2–1** Function $f(t)1(t)$ and translated function $f(t-\alpha)1(t-\alpha)$.

**Table 2–1** Laplace Transform Pairs

| | $f(t)$ | $F(s)$ |
|---|---|---|
| 1 | Unit impulse $\delta(t)$ | $1$ |
| 2 | Unit step $1(t)$ | $\frac{1}{s}$ |
| 3 | $t$ | $\frac{1}{s^2}$ |
| 4 | $\frac{t^{n-1}}{(n-1)!} \quad (n = 1, 2, 3, \ldots)$ | $\frac{1}{s^n}$ |
| 5 | $t^n \quad (n = 1, 2, 3, \ldots)$ | $\frac{n!}{s^{n+1}}$ |
| 6 | $e^{-at}$ | $\frac{1}{s+a}$ |
| 7 | $te^{-at}$ | $\frac{1}{(s+a)^2}$ |
| 8 | $\frac{1}{(n-1)!} t^{n-1}e^{-at} \quad (n = 1, 2, 3, \ldots)$ | $\frac{1}{(s+a)^n}$ |
| 9 | $t^n e^{-at} \quad (n = 1, 2, 3, \ldots)$ | $\frac{n!}{(s+a)^{n+1}}$ |
| 10 | $\sin \omega t$ | $\frac{\omega}{s^2+\omega^2}$ |
| 11 | $\cos \omega t$ | $\frac{s}{s^2+\omega^2}$ |
| 12 | $\sinh \omega t$ | $\frac{\omega}{s^2-\omega^2}$ |
| 13 | $\cosh \omega t$ | $\frac{s}{s^2-\omega^2}$ |
| 14 | $\frac{1}{a}(1 - e^{-at})$ | $\frac{1}{s(s+a)}$ |
| 15 | $\frac{1}{b-a}(e^{-at} - e^{-bt})$ | $\frac{1}{(s+a)(s+b)}$ |
| 16 | $\frac{1}{b-a}(be^{-bt} - ae^{-at})$ | $\frac{s}{(s+a)(s+b)}$ |
| 17 | $\frac{1}{ab}\left[1 + \frac{1}{a-b}(be^{-at} - ae^{-bt})\right]$ | $\frac{1}{s(s+a)(s+b)}$ |

**Table 2–1** (*Continued*)

| | | |
|---|---|---|
| 18 | $\frac{1}{a^2}(1 - e^{-at} - ate^{-at})$ | $\frac{1}{s(s+a)^2}$ |
| 19 | $\frac{1}{a^2}(at - 1 + e^{-at})$ | $\frac{1}{s^2(s+a)}$ |
| 20 | $e^{-at} \sin \omega t$ | $\frac{\omega}{(s+a)^2 + \omega^2}$ |
| 21 | $e^{-at} \cos \omega t$ | $\frac{s+a}{(s+a)^2 + \omega^2}$ |
| 22 | $\frac{\omega_n}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t} \sin \omega_n \sqrt{1-\zeta^2}\, t$ | $\frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$ |
| 23 | $-\frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t} \sin(\omega_n \sqrt{1-\zeta^2}\, t - \phi)$<br>$\phi = \tan^{-1} \frac{\sqrt{1-\zeta^2}}{\zeta}$ | $\frac{s}{s^2 + 2\zeta\omega_n s + \omega_n^2}$ |
| 24 | $1 - \frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t} \sin(\omega_n \sqrt{1-\zeta^2}\, t + \phi)$<br>$\phi = \tan^{-1} \frac{\sqrt{1-\zeta^2}}{\zeta}$ | $\frac{\omega_n^2}{s(s^2 + 2\zeta\omega_n s + \omega_n^2)}$ |
| 25 | $1 - \cos \omega t$ | $\frac{\omega^2}{s(s^2 + \omega^2)}$ |
| 26 | $\omega t - \sin \omega t$ | $\frac{\omega^3}{s^2(s^2 + \omega^2)}$ |
| 27 | $\sin \omega t - \omega t \cos \omega t$ | $\frac{2\omega^3}{(s^2 + \omega^2)^2}$ |
| 28 | $\frac{1}{2\omega} t \sin \omega t$ | $\frac{s}{(s^2 + \omega^2)^2}$ |
| 29 | $t \cos \omega t$ | $\frac{s^2 - \omega^2}{(s^2 + \omega^2)^2}$ |
| 30 | $\frac{1}{\omega_2^2 - \omega_1^2}(\cos \omega_1 t - \cos \omega_2 t) \qquad (\omega_1^2 \neq \omega_2^2)$ | $\frac{s}{(s^2 + \omega_1^2)(s^2 + \omega_2^2)}$ |
| 31 | $\frac{1}{2\omega}(\sin \omega t + \omega t \cos \omega t)$ | $\frac{s^2}{(s^2 + \omega^2)^2}$ |

This last equation states that the translation of the time function $f(t)1(t)$ by $\alpha$ (where $\alpha \geq 0$) corresponds to the multiplication of the transform $F(s)$ by $e^{-\alpha s}$.

**Pulse function.** Consider the pulse function

$$
\begin{aligned}
f(t) &= \frac{A}{t_0}, && \text{for } 0 < t < t_0 \\
&= 0, && \text{for } t < 0,\ t_0 < t
\end{aligned}
$$

where $A$ and $t_0$ are constants.

The pulse function here may be considered a step function of height $A/t_0$ that begins at $t = 0$ and that is superimposed by a negative step function of height $A/t_0$ beginning at $t = t_0$; that is,

$$
f(t) = \frac{A}{t_0} 1(t) - \frac{A}{t_0} 1(t - t_0)
$$

Then the Laplace transform of $f(t)$ is obtained as

$$
\begin{aligned}
\mathcal{L}[f(t)] &= \mathcal{L}\left[\frac{A}{t_0} 1(t)\right] - \mathcal{L}\left[\frac{A}{t_0} 1(t - t_0)\right] \\
&= \frac{A}{t_0 s} - \frac{A}{t_0 s} e^{-st_0} \\
&= \frac{A}{t_0 s}(1 - e^{-st_0}) \qquad (2\text{–}5)
\end{aligned}
$$

**Impulse function.** The impulse function is a special limiting case of the pulse function. Consider the impulse function

$$
\begin{aligned}
g(t) &= \lim_{t_0 \to 0} \frac{A}{t_0}, && \text{for } 0 < t < t_0 \\
&= 0, && \text{for } t < 0,\ t_0 < t
\end{aligned}
$$

Since the height of the impulse function is $A/t_0$ and the duration is $t_0$, the area under the impulse is equal to $A$. As the duration $t_0$ approaches zero, the height $A/t_0$ approaches infinity, but the area under the impulse remains equal to $A$. Note that the magnitude of the impulse is measured by its area.

Referring to Equation (2–5), the Laplace transform of this impulse function is shown to be

$$
\begin{aligned}
\mathcal{L}[g(t)] &= \lim_{t_0 \to 0} \left[\frac{A}{t_0 s}(1 - e^{-st_0})\right] \\
&= \lim_{t_0 \to 0} \frac{\dfrac{d}{dt_0}[A(1 - e^{-st_0})]}{\dfrac{d}{dt_0}(t_0 s)} = \frac{As}{s} = A
\end{aligned}
$$

Thus the Laplace transform of the impulse function is equal to the area under the impulse.

The impulse function whose area is equal to unity is called the *unit-impulse function* or the *Dirac delta function.* The unit-impulse function occurring at $t = t_0$ is usually denoted by $\delta(t - t_0)$. $\delta(t - t_0)$ satisfies the following:

$$\delta(t - t_0) = 0, \qquad \text{for } t \neq t_0$$

$$\delta(t - t_0) = \infty, \qquad \text{for } t = t_0$$

$$\int_{-\infty}^{\infty} \delta(t - t_0)dt = 1$$

It should be mentioned that an impulse that has an infinite magnitude and zero duration is mathematical fiction and does not occur in physical systems. If, however, the magnitude of a pulse input to a system is very large and its duration is very short compared to the system time constants, then we can approximate the pulse input by an impulse function. For instance, if a force or torque input $f(t)$ is applied to a system for a very short time duration, $0 < t < t_0$, where the magnitude of $f(t)$ is sufficiently large so that the integral $\int_0^{t_0} f(t)\, dt$ is not negligible, then this input can be considered an impulse input. (Note that when we describe the impulse input the area or magnitude of the impulse is most important, but the exact shape of the impulse is usually immaterial.) The impulse input supplies energy to the system in an infinitesimal time.

The concept of the impulse function is quite useful in differentiating discontinuous functions. The unit-impulse function $\delta(t - t_0)$ can be considered the derivative of the unit-step function $1(t - t_0)$ at the point of discontinuity $t = t_0$ or

$$\delta(t - t_0) = \frac{d}{dt} 1(t - t_0)$$

Conversely, if the unit-impulse function $\delta(t - t_0)$ is integrated, the result is the unit-step function $1(t - t_0)$. With the concept of the impulse function we can differentiate a function containing discontinuities, giving impulses, the magnitudes of which are equal to the magnitude of each corresponding discontinuity.

**Multiplication of $f(t)$ by $e^{-\alpha t}$.** If $f(t)$ is Laplace transformable, its Laplace transform being $F(s)$, then the Laplace transform of $e^{-\alpha t}f(t)$ is obtained as

$$\mathscr{L}[e^{-\alpha t}f(t)] = \int_0^{\infty} e^{-\alpha t}f(t)e^{-st}\, dt = F(s + \alpha) \tag{2–6}$$

We see that the multiplication of $f(t)$ by $e^{-\alpha t}$ has the effect of replacing $s$ by $(s + \alpha)$ in the Laplace transform. Conversely, changing $s$ to $(s + \alpha)$ is equivalent to multiplying $f(t)$ by $e^{-\alpha t}$. (Note that $\alpha$ may be real or complex.)

The relationship given by Equation (2–6) is useful in finding the Laplace transforms of such functions as $e^{-\alpha t} \sin \omega t$ and $e^{-\alpha t} \cos \omega t$. For instance, since

$$\mathscr{L}[\sin \omega t] = \frac{\omega}{s^2 + \omega^2} = F(s), \qquad \mathscr{L}[\cos \omega] = \frac{s}{s^2 + \omega^2} = G(s)$$

it follows from Equation (2–6) that the Laplace transforms of $e^{-\alpha t} \sin \omega t$ and $e^{-\alpha t} \cos \omega t$ are given, respectively, by

$$\mathcal{L}[e^{-\alpha t}\sin\omega t] = F(s+\alpha) = \frac{\omega}{(s+\alpha)^2+\omega^2}$$

$$\mathcal{L}[e^{-\alpha t}\cos\omega t] = G(s+\alpha) = \frac{s+\alpha}{(s+\alpha)^2+\omega^2}$$

**Change of time scale.** In analyzing physical systems, it is sometimes desirable to change the time scale or normalize a given time function. The result obtained in terms of normalized time is useful because it can be applied directly to different systems having similar mathematical equations.

If $t$ is changed into $t/\alpha$, where $\alpha$ is a positive constant, then the function $f(t)$ is changed into $f(t/\alpha)$. If we denote the Laplace transform of $f(t)$ by $F(s)$, then the Laplace transform of $f(t/\alpha)$ may be obtained as follows:

$$\mathcal{L}\left[f\left(\frac{t}{\alpha}\right)\right] = \int_0^\infty f\left(\frac{t}{\alpha}\right)e^{-st}\,dt$$

Letting $t/\alpha = t_1$ and $\alpha s = s_1$, we obtain

$$\begin{aligned}\mathcal{L}\left[f\left(\frac{t}{\alpha}\right)\right] &= \int_0^\infty f(t_1)e^{-s_1t_1}\,d(\alpha t_1)\\ &= \alpha\int_0^\infty f(t_1)e^{-s_1t_1}\,dt_1\\ &= \alpha F(s_1)\end{aligned}$$

or

$$\mathcal{L}\left[f\left(\frac{t}{\alpha}\right)\right] = \alpha F(\alpha s)$$

As an example, consider $f(t) = e^{-t}$ and $f(t/5) = e^{-0.2t}$. We obtain

$$\mathcal{L}[f(t)] = \mathcal{L}[e^{-t}] = F(s) = \frac{1}{s+1}$$

Hence

$$\mathcal{L}\left[f\left(\frac{t}{5}\right)\right] = \mathcal{L}[e^{-0.2t}] = 5F(5s) = \frac{5}{5s+1}$$

This result can be verified easily by taking the Laplace transform of $e^{-0.2t}$ directly as follows:

$$\mathcal{L}[e^{-0.2t}] = \frac{1}{s+0.2} = \frac{5}{5s+1}$$

**Comments on the lower limit of the Laplace integral.** In some cases, $f(t)$ possesses an impulse function at $t = 0$. Then the lower limit of the Laplace integral must be clearly specified as to whether it is $0-$ or $0+$, since the Laplace transforms of $f(t)$ differ for these two lower limits. If such a distinction of the lower limit of the Laplace integral is necessary, we use the notations

$$\mathcal{L}_+[f(t)] = \int_{0+}^{\infty} f(t)e^{-st}\,dt$$

$$\mathcal{L}_-[f(t)] = \int_{0-}^{\infty} f(t)e^{-st}\,dt = \mathcal{L}_+[f(t)] + \int_{0-}^{0+} f(t)e^{-st}\,dt$$

If $f(t)$ involves an impulse function at $t = 0$, then

$$\mathcal{L}_+[f(t)] \neq \mathcal{L}_-[f(t)]$$

since

$$\int_{0-}^{0+} f(t)e^{-st}\,dt \neq 0$$

for such a case. Obviously, if $f(t)$ does not possess an impulse function at $t = 0$ (that is, if the function to be transformed is finite between $t = 0-$ and $t = 0+$), then

$$\mathcal{L}_+[f(t)] = \mathcal{L}_-[f(t)]$$

## 2–4 LAPLACE TRANSFORM THEOREMS

This section presents several theorems on Laplace transformation that are important in control engineering.

**Real differentiation theorem.** The Laplace transform of the derivative of a function $f(t)$ is given by

$$\mathcal{L}\left[\frac{d}{dt}f(t)\right] = sF(s) - f(0) \tag{2–7}$$

where $f(0)$ is the initial value of $f(t)$ evaluated at $t = 0$.

For a given function $f(t)$, the values of $f(0+)$ and $f(0-)$ may be the same or different, as illustrated in Figure 2–2. The distinction between $f(0+)$ and $f(0-)$ is important when $f(t)$ has a discontinuity at $t = 0$ because in such a case $df(t)/dt$ will involve an impulse function at $t = 0$. If $f(0+) \neq f(0-)$, Equation (2–7) must be modified to

$$\mathcal{L}_+\left[\frac{d}{dt}f(t)\right] = sF(s) - f(0+)$$

$$\mathcal{L}_-\left[\frac{d}{dt}f(t)\right] = sF(s) - f(0-)$$

To prove the real differentiation theorem, Equation (2–7), we proceed as follows. Integrating the Laplace integral by parts gives

$$\int_0^{\infty} f(t)e^{-st}\,dt = f(t)\frac{e^{-st}}{-s}\bigg|_0^{\infty} - \int_0^{\infty}\left[\frac{d}{dt}f(t)\right]\frac{e^{-st}}{-s}\,dt$$

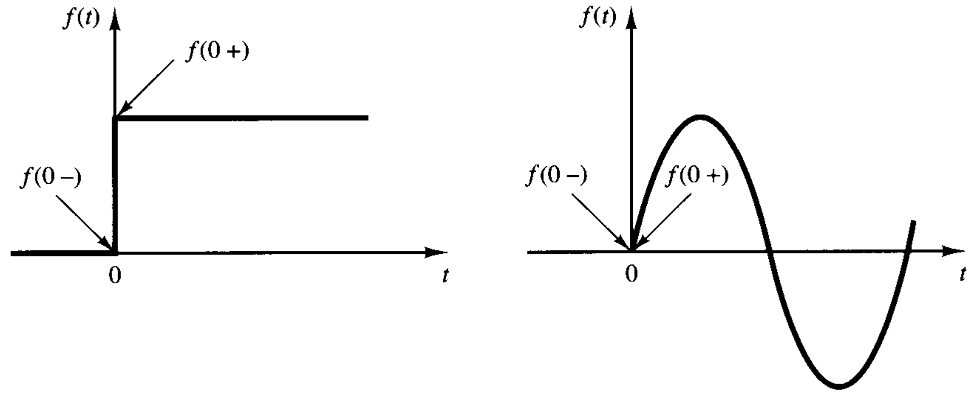



**Figure 2–2**
Step function and sine function indicating initial values at $t = 0-$ and $t = 0+$.

Hence

$$F(s) = \frac{f(0)}{s} + \frac{1}{s}\mathscr{L}\left[\frac{d}{dt}f(t)\right]$$

It follows that

$$\mathscr{L}\left[\frac{d}{dt}f(t)\right] = sF(s) - f(0)$$

Similarly, we obtain the following relationship for the second derivative of $f(t)$:

$$\mathscr{L}\left[\frac{d^2}{dt^2}f(t)\right] = s^2F(s) - sf(0) - \dot{f}(0)$$

where $\dot{f}(0)$ is the value of $df(t)/dt$ evaluated at $t = 0$. To derive this equation, define

$$\frac{d}{dt}f(t) = g(t)$$

Then

$$\begin{aligned}
\mathscr{L}\left[\frac{d^2}{dt^2}f(t)\right] &= \mathscr{L}\left[\frac{d}{dt}g(t)\right] = s\mathscr{L}[g(t)] - g(0) \\
&= s\mathscr{L}\left[\frac{d}{dt}f(t)\right] - \dot{f}(0) \\
&= s^2F(s) - sf(0) - \dot{f}(0)
\end{aligned}$$

Similarly, for the $n$th derivative of $f(t)$, we obtain

$$\mathscr{L}\left[\frac{d^n}{dt^n}f(t)\right] = s^nF(s) - s^{n-1}f(0) - s^{n-2}\dot{f}(0) - \cdots - s\overset{(n-2)}{f}(0) - \overset{(n-1)}{f}(0)$$

where $f(0), \dot{f}(0), \ldots, \overset{(n-1)}{f}(0)$ represent the values of $f(t), df(t)/dt, \ldots, d^{n-1}f(t)/dt^{n-1}$, respectively, evaluated at $t = 0$. If the distinction between $\mathscr{L}_+$ and $\mathscr{L}_-$ is necessary, we substitute $t = 0+$ or $t = 0-$ into $f(t), df(t)/dt, \ldots, d^{n-1}f(t)/dt^{n-1}$, depending on whether we take $\mathscr{L}_+$ or $\mathscr{L}_-$.

Note that, in order for Laplace transforms of derivatives of $f(t)$ to exist, $d^nf(t)/dt^n$ $(n = 1, 2, 3, \ldots)$ must be Laplace transformable.

Note also that, if all the initial values of $f(t)$ and its derivatives are equal to zero, then the Laplace transform of the $n$th derivative of $f(t)$ is given by $s^nF(s)$.

**EXAMPLE 2–1** Consider the cosine function.

$$g(t) = 0, \qquad \text{for } t < 0$$
$$= \cos \omega t, \qquad \text{for } t \geq 0$$

The Laplace transform of this cosine function can be obtained directly as in the case of the sinusoidal function considered earlier. The use of the real differentiation theorem, however, will be demonstrated here by deriving the Laplace transform of the cosine function from the Laplace transform of the sine function. If we define

$$f(t) = 0, \qquad \text{for } t < 0$$
$$= \sin \omega t, \qquad \text{for } t \geq 0$$

then

$$\mathscr{L}[\sin \omega t] = F(s) = \frac{\omega}{s^2 + \omega^2}$$

The Laplace transform of the cosine function is obtained as

$$\mathscr{L}[\cos \omega t] = \mathscr{L}\left[\frac{1}{\omega}\left(\frac{d}{dt}\sin \omega t\right)\right] = \frac{1}{\omega}[sF(s) - f(0)]$$
$$= \frac{1}{\omega}\left[\frac{s\omega}{s^2 + \omega^2} - 0\right] = \frac{s}{s^2 + \omega^2}$$

**Final-value theorem.** The final-value theorem relates the steady-state behavior of $f(t)$ to the behavior of $sF(s)$ in the neighborhood of $s = 0$. This theorem, however, applies if and only if $\lim_{t\to\infty} f(t)$ exists [which means that $f(t)$ settles down to a definite value for $t \to \infty$]. If all poles of $sF(s)$ lie in the left half $s$ plane, $\lim_{t\to\infty} f(t)$ exists. But if $sF(s)$ has poles on the imaginary axis or in the right half $s$ plane, $f(t)$ will contain oscillating or exponentially increasing time functions, respectively, and $\lim_{t\to\infty} f(t)$ will not exist. The final-value theorem does not apply in such cases. For instance, if $f(t)$ is the sinusoidal function $\sin \omega t$, $sF(s)$ has poles at $s = \pm j\omega$ and $\lim_{t\to\infty} f(t)$ does not exist. Therefore, this theorem is not applicable to such a function.

The final-value theorem may be stated as follows. If $f(t)$ and $df(t)/dt$ are Laplace transformable, if $F(s)$ is the Laplace transform of $f(t)$, and if $\lim_{t\to\infty} f(t)$ exists, then

$$\lim_{t\to\infty} f(t) = \lim_{s\to 0} sF(s)$$

To prove the theorem, we let $s$ approach zero in the equation for the Laplace transform of the derivative of $f(t)$ or

$$\lim_{s\to 0}\int_0^\infty \left[\frac{d}{dt} f(t)\right] e^{-st}\, dt = \lim_{s\to 0}[sF(s) - f(0)]$$

Since $\lim_{s\to 0} e^{-st} = 1$, we obtain

$$\int_0^\infty \left[\frac{d}{dt} f(t)\right] dt = f(t)\Big|_0^\infty = f(\infty) - f(0)$$
$$= \lim_{s\to 0} sF(s) - f(0)$$

from which

$$f(\infty) = \lim_{t\to\infty} f(t) = \lim_{s\to 0} sF(s)$$

The final-value theorem states that the steady-state behavior of $f(t)$ is the same as the behavior of $sF(s)$ in the neighborhood of $s = 0$. Thus, it is possible to obtain the value of $f(t)$ at $t = \infty$ directly from $F(s)$.

**EXAMPLE 2–2** Given

$$\mathcal{L}[f(t)] = F(s) = \frac{1}{s(s+1)}$$

what is $\lim_{t\to\infty} f(t)$?

Since the pole of $sF(s) = 1/(s + 1)$ lies in the left half $s$ plane, $\lim_{t\to\infty} f(t)$ exists. So the final-value theorem is applicable in this case.

$$\lim_{t\to\infty} f(t) = f(\infty) = \lim_{s\to 0} sF(s) = \lim_{s\to 0} \frac{s}{s(s+1)} = \lim_{s\to 0} \frac{1}{s+1} = 1$$

In fact, this result can easily be verified, since

$$f(t) = 1 - e^{-t}, \qquad \text{for } t \geq 0$$

**Initial-value theorem.** The initial-value theorem is the counterpart of the final-value theorem. By using this theorem, we are able to find the value of $f(t)$ at $t = 0+$ directly from the Laplace transform of $f(t)$. The initial-value theorem does not give the value of $f(t)$ at exactly $t = 0$ but at a time slightly greater than zero.

The initial-value theorem may be stated as follows: If $f(t)$ and $df(t)/dt$ are both Laplace transformable and if $\lim_{s\to\infty} sF(s)$ exists, then

$$f(0+) = \lim_{s\to\infty} sF(s)$$

To prove this theorem, we use the equation for the $\mathcal{L}_+$ transform of $df(t)/dt$:

$$\mathcal{L}_+\left[\frac{d}{dt} f(t)\right] = sF(s) - f(0+)$$

For the time interval $0+ \leq t \leq \infty$, as $s$ approaches infinity, $e^{-st}$ approaches zero. (Note that we must use $\mathcal{L}_+$ rather than $\mathcal{L}_-$ for this condition.) And so

$$\lim_{s\to\infty} \int_{0+}^{\infty} \left[\frac{d}{dt} f(t)\right] e^{-st}\, dt = \lim_{s\to\infty} [sF(s) - f(0+)] = 0$$

or

$$f(0+) = \lim_{s\to\infty} sF(s)$$

In applying the initial-value theorem, we are not limited as to the locations of the poles of $sF(s)$. Thus the initial-value theorem is valid for the sinusoidal function.

It should be noted that the initial-value theorem and the final-value theorem provide a convenient check on the solution, since they enable us to predict the system behavior in the time domain without actually transforming functions in $s$ back to time functions.

**Real-integration theorem.** If $f(t)$ is of exponential order, then the Laplace transform of $\int f(t)\,dt$ exists and is given by

$$\mathcal{L}\left[\int f(t)\,dt\right] = \frac{F(s)}{s} + \frac{f^{-1}(0)}{s} \tag{2–8}$$

where $F(s) = \mathcal{L}[f(t)]$ and $f^{-1}(0) = \int f(t)\,dt$, evaluated at $t = 0$.

Note that if $f(t)$ involves an impulse function at $t = 0$, then $f^{-1}(0+) \neq f^{-1}(0-)$. So if $f(t)$ involves an impulse function at $t = 0$, we must modify Equation (2–8) as follows:

$$\mathcal{L}_+\left[\int f(t)\,dt\right] = \frac{F(s)}{s} + \frac{f^{-1}(0+)}{s}$$

$$\mathcal{L}_-\left[\int f(t)\,dt\right] = \frac{F(s)}{s} + \frac{f^{-1}(0-)}{s}$$

The real-integration theorem given by Equation (2–8) can be proved in the following way. Integration by parts yields

$$\begin{aligned}
\mathcal{L}\left[\int f(t)\,dt\right] &= \int_0^\infty \left[\int f(t)\,dt\right] e^{-st}\,dt \\
&= \left[\int f(t)\,dt\right] \frac{e^{-st}}{-s}\Bigg|_0^\infty - \int_0^\infty f(t)\frac{e^{-st}}{-s}\,dt \\
&= \frac{1}{s}\int f(t)\,dt\Bigg|_{t=0} + \frac{1}{s}\int_0^\infty f(t)e^{-st}\,dt \\
&= \frac{f^{-1}(0)}{s} + \frac{F(s)}{s}
\end{aligned}$$

and the theorem is proved.

We see that integration in the time domain is converted into division in the $s$ domain. If the initial value of the integral is zero, the Laplace transform of the integral of $f(t)$ is given by $F(s)/s$.

The preceding real-integration theorem given by Equation (2–8) can be modified slightly to deal with the definite integral of $f(t)$. If $f(t)$ is of exponential order, the Laplace transform of the definite integral $\int_0^t f(t)\,dt$ is given by

$$\mathcal{L}\left[\int_0^t f(t)\,dt\right] = \frac{F(s)}{s} \tag{2–9}$$

where $F(s) = \mathcal{L}[f(t)]$. This is also referred to as the real-integration theorem. Note that if $f(t)$ involves an impulse function at $t = 0$ then $\int_{0+}^t f(t)\,dt \neq \int_{0-}^t f(t)\,dt$, and the following distinction must be observed:

$$\mathcal{L}_+\left[\int_{0+}^t f(t)\,dt\right] = \frac{\mathcal{L}_+[f(t)]}{s}$$

$$\mathcal{L}_-\left[\int_{0-}^t f(t)\,dt\right] = \frac{\mathcal{L}_-[f(t)]}{s}$$

To prove Equation (2–9), first note that

$$\int_0^t f(t)\,dt = \int f(t)\,dt - f^{-1}(0)$$

where $f^{-1}(0)$ is equal to $\int f(t)\,dt$ evaluated at $t = 0$ and is a constant. Hence

$$\mathscr{L}\left[\int_0^t f(t)\,dt\right] = \mathscr{L}\left[\int f(t)\,dt\right] - \mathscr{L}[f^{-1}(0)]$$

Noting that $f^{-1}(0)$ is a constant so that

$$\mathscr{L}[f^{-1}(0)] = \frac{f^{-1}(0)}{s}$$

we obtain

$$\mathscr{L}\left[\int_0^t f(t)\,dt\right] = \frac{F(s)}{s} + \frac{f^{-1}(0)}{s} - \frac{f^{-1}(0)}{s} = \frac{F(s)}{s}$$

**Complex-differentiation theorem.** If $f(t)$ is Laplace transformable, then, except at poles of $F(s)$,

$$\mathscr{L}[tf(t)] = -\frac{d}{ds}F(s)$$

where $F(s) = \mathscr{L}[f(t)]$. This is known as the complex-differentiation theorem. Also,

$$\mathscr{L}[t^2 f(t)] = \frac{d^2}{ds^2}F(s)$$

In general,

$$\mathscr{L}[t^n f(t)] = (-1)^n \frac{d^n}{ds^n}F(s), \qquad \text{for } n = 1, 2, 3, \ldots$$

To prove the complex-differentiation theorem, we proceed as follows:

$$\begin{aligned}
\mathscr{L}[tf(t)] &= \int_0^\infty tf(t)e^{-st}\,dt = -\int_0^\infty f(t)\frac{d}{ds}(e^{-st})\,dt \\
&= -\frac{d}{ds}\int_0^\infty f(t)e^{-st}\,dt = -\frac{d}{ds}F(s)
\end{aligned}$$

Hence the theorem. Similarly, by defining $tf(t) = g(t)$, the result is

$$\begin{aligned}
\mathscr{L}[t^2 f(t)] &= \mathscr{L}[tg(t)] = -\frac{d}{ds}G(s) = -\frac{d}{ds}\left[-\frac{d}{ds}F(s)\right] \\
&= (-1)^2\frac{d^2}{ds^2}F(s) = \frac{d^2}{ds^2}F(s)
\end{aligned}$$

Repeating the same process, we obtain

$$\mathscr{L}[t^n f(t)] = (-1)^n \frac{d^n}{ds^n} F(s), \qquad \text{for } n = 1, 2, 3, \ldots$$

**Convolution integral.** Consider the Laplace transform of

$$\int_0^t f_1(t - \tau) f_2(\tau)\, d\tau$$

This integral is often written as

$$f_1(t) * f_2(t)$$

The mathematical operation $f_1(t)*f_2(t)$ is called *convolution.* Note that if we put $t - \tau = \xi$, then

$$\begin{aligned}\int_0^t f_1(t - \tau) f_2(\tau)\, d\tau &= -\int_t^0 f_1(\xi) f_2(t - \xi)\, d\xi \\ &= \int_0^t f_1(\tau) f_2(t - \tau)\, d\tau\end{aligned}$$

Hence

$$\begin{aligned}f_1(t)*f_2(t) &= \int_0^t f_1(t - \tau) f_2(\tau)\, d\tau \\ &= \int_0^t f_1(\tau) f_2(t - \tau)\, d\tau \\ &= f_2(t)*f_1(t)\end{aligned}$$

If $f_1(t)$ and $f_2(t)$ are piecewise continuous and of exponential order, then the Laplace transform of

$$\int_0^t f_1(t - \tau) f_2(\tau)\, d\tau$$

can be obtained as follows:

$$\mathscr{L}\left[\int_0^t f_1(t - \tau) f_2(\tau)\, d\tau\right] = F_1(s) F_2(s) \tag{2–10}$$

where

$$F_1(s) = \int_0^\infty f_1(t) e^{-st}\, dt = \mathscr{L}[f_1(t)]$$

$$F_2(s) = \int_0^\infty f_2(t) e^{-st}\, dt = \mathscr{L}[f_2(t)]$$

To prove Equation (2–10) note that $f_1(t - \tau)1(t - \tau) = 0$ for $\tau > t$. Hence

$$\int_0^t f_1(t - \tau) f_2(\tau)\, d\tau = \int_0^\infty f_1(t - \tau) 1(t - \tau) f_2(\tau)\, d\tau$$

Then

$$\mathscr{L}\left[\int_0^t f_1(t-\tau)f_2(\tau)\,d\tau\right] = \mathscr{L}\left[\int_0^\infty f_1(t-\tau)1(t-\tau)f_2(\tau)\,d\tau\right]$$

$$= \int_0^\infty e^{-st}\left[\int_0^\infty f_1(t-\tau)1(t-\tau)f_2(\tau)\,d\tau\right]dt$$

Substituting $t - \tau = \lambda$ in this last equation and changing the order of integration, which is valid in this case because $f_1(t)$ and $f_2(t)$ are Laplace transformable, we obtain

$$\mathscr{L}\left[\int_0^t f_1(t-\tau)f_2(\tau)\,d\tau\right] = \int_0^\infty f_1(t-\tau)1(t-\tau)e^{-st}\,dt \int_0^\infty f_2(\tau)\,d\tau$$

$$= \int_0^\infty f_1(\lambda)e^{-s(\lambda+\tau)}\,d\lambda \int_0^\infty f_2(\tau)\,d\tau$$

$$= \int_0^\infty f_1(\lambda)e^{-s\lambda}\,d\lambda \int_0^\infty f_2(\tau)e^{-s\tau}\,d\tau$$

$$= F_1(s)F_2(s)$$

This last equation gives the Laplace transform of the convolution integral. Conversely, if the Laplace transform of a function is given by a product of two Laplace transform functions, $F_1(s)F_2(s)$, then the corresponding time function (the inverse Laplace transform) is given by the convolution integral $f_1(t)*f_2(t)$.

**Laplace transform of product of two time functions.** The Laplace transform of the product of two Laplace transformable functions $f(t)$ and $g(t)$ can be given by

$$\mathscr{L}[f(t)g(t)] = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F(p)G(s-p)\,dp \qquad (2\text{–}11)$$

To show this, we may proceed as follows: The Laplace transform of the product of $f(t)$ and $g(t)$ can be written as

$$\mathscr{L}[f(t)g(t)] = \int_0^\infty f(t)g(t)e^{-st}\,dt \qquad (2\text{–}12)$$

Note that the inversion integral is

$$f(t) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F(s)e^{st}\,ds, \qquad \text{for } t > 0$$

where $c$ is the abscissa of convergence for $F(s)$. Thus,

$$\mathscr{L}[f(t)g(t)] = \frac{1}{2\pi j}\int_0^\infty \int_{c-j\infty}^{c+j\infty} F(p)e^{pt}\,dp\,g(t)e^{-st}\,dt$$

Because of the uniform convergence of the integrals considered, we may invert the order of integration:

$$\mathscr{L}[f(t)g(t)] = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F(p)\, dp \int_0^\infty g(t) e^{-(s-p)t}\, dt$$

Noting that

$$\int_0^\infty g(t) e^{-(s-p)t}\, dt = G(s-p)$$

we obtain

$$\mathscr{L}[f(t)g(t)] = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F(p) G(s-p)\, dp \tag{2–13}$$

**Summary.** Table 2–2 summarizes properties and theorems of the Laplace transforms. Most of them have been derived or proved in this section.

## 2-5 INVERSE LAPLACE TRANSFORMATION

As noted earlier, the inverse Laplace transform can be obtained by use of the inversion integral given by Equation (2–4). However, the inversion integral is complicated and, therefore, its use is not recommended for finding inverse Laplace transforms of commonly encountered functions in control engineering.

A convenient method for obtaining inverse Laplace transforms is to use a table of Laplace transforms. In this case, the Laplace transform must be in a form immediately recognizable in such a table. Quite often the function in question may not appear in tables of Laplace transforms available to the engineer. If a particular transform $F(s)$ cannot be found in a table, then we may expand it into partial fractions and write $F(s)$ in terms of simple functions of $s$ for which the inverse Laplace transforms are already known.

Note that these simpler methods for finding inverse Laplace transforms are based on the fact that the unique correspondence of a time function and its inverse Laplace transform holds for any continuous time function.

**Partial-fraction expansion method for finding inverse Laplace transforms.** For problems in control systems analysis, $F(s)$, the Laplace transform of $f(t)$, frequently occurs in the form

$$F(s) = \frac{B(s)}{A(s)}$$

where $A(s)$ and $B(s)$ are polynomials in $s$. In the expansion of $F(s) = B(s)/A(s)$ into a partial-fraction form, it is important that the highest power of $s$ in $A(s)$ be greater than the highest power of $s$ in $B(s)$. If such is not the case, the numerator $B(s)$ must be divided by the denominator $A(s)$ in order to produce a polynomial in $s$ plus a remainder (a ratio of polynomials in $s$ whose numerator is of lower degree than the denominator).

**Table 2–2** Properties of Laplace Transforms

| | |
|---|---|
| 1 | $\mathscr{L}[Af(t)] = AF(s)$ |
| 2 | $\mathscr{L}[f_1(t) \pm f_2(t)] = F_1(s) \pm F_2(s)$ |
| 3 | $\mathscr{L}_\pm\left[\dfrac{d}{dt} f(t)\right] = sF(s) - f(0\pm)$ |
| 4 | $\mathscr{L}_\pm\left[\dfrac{d^2}{dt^2} f(t)\right] = s^2F(s) - sf(0\pm) - \dot{f}(0\pm)$ |
| 5 | $\mathscr{L}_\pm\left[\dfrac{d^n}{dt^n} f(t)\right] = s^nF(s) - \sum_{k=1}^{n} s^{n-k} \overset{(k-1)}{f}(0\pm)$ where $\overset{(k-1)}{f}(t) = \dfrac{d^{k-1}}{dt^{k-1}} f(t)$ |
| 6 | $\mathscr{L}_\pm\left[\int f(t)\,dt\right] = \dfrac{F(s)}{s} + \dfrac{1}{s}\left[\int f(t)\,dt\right]_{t=0\pm}$ |
| 7 | $\mathscr{L}_\pm\left[\int \cdots \int f(t)(dt)^n\right] = \dfrac{F(s)}{s^n} + \sum_{k=1}^{n} \dfrac{1}{s^{n-k+1}}\left[\int \cdots \int f(t)(dt)^k\right]_{t=0\pm}$ |
| 8 | $\mathscr{L}\left[\int_0^t f(t)\,dt\right] = \dfrac{F(s)}{s}$ |
| 9 | $\int_0^\infty f(t)\,dt = \lim_{s\to 0} F(s) \quad$ if $\int_0^\infty f(t)\,dt$ exists |
| 10 | $\mathscr{L}[e^{-at} f(t)] = F(s + a)$ |
| 11 | $\mathscr{L}[f(t - \alpha)1(t - \alpha)] = e^{-\alpha s}F(s) \qquad \alpha \geq 0$ |
| 12 | $\mathscr{L}[tf(t)] = -\dfrac{dF(s)}{ds}$ |
| 13 | $\mathscr{L}[t^2f(t)] = \dfrac{d^2}{ds^2} F(s)$ |
| 14 | $\mathscr{L}[t^nf(t)] = (-1)^n \dfrac{d^n}{ds^n} F(s) \qquad n = 1, 2, 3, \ldots$ |
| 15 | $\mathscr{L}\left[\dfrac{1}{t} f(t)\right] = \int_s^\infty F(s)\,ds \quad$ if $\lim_{t\to 0} \dfrac{1}{t} f(t)$ exists |
| 16 | $\mathscr{L}\left[f\left(\dfrac{t}{a}\right)\right] = aF(as)$ |
| 17 | $\mathscr{L}\left[\int_0^t f_1(t - \tau)f_2(\tau)\,d\tau\right] = F_1(s)F_2(s)$ |
| 18 | $\mathscr{L}[f(t)g(t)] = \dfrac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F(p)G(s - p)\,dp$ |

If $F(s)$ is broken up into components,

$$F(s) = F_1(s) + F_2(s) + \cdots + F_n(s)$$

and if the inverse Laplace transforms of $F_1(s), F_2(s), \ldots, F_n(s)$ are readily available, then

$$\begin{aligned}\mathcal{L}^{-1}[F(s)] &= \mathcal{L}^{-1}[F_1(s)] + \mathcal{L}^{-1}[F_2(s)] + \cdots + \mathcal{L}^{-1}[F_n(s)] \\ &= f_1(t) + f_2(t) + \cdots + f_n(t)\end{aligned}$$

where $f_1(t), f_2(t), \ldots, f_n(t)$ are the inverse Laplace transforms of $F_1(s), F_2(s), \ldots, F_n(s)$, respectively. The inverse Laplace transform of $F(s)$ thus obtained is unique except possibly at points where the time function is discontinuous. Whenever the time function is continuous, the time function $f(t)$ and its Laplace transform $F(s)$ have a one-to-one correspondence.

The advantage of the partial-fraction expansion approach is that the individual terms of $F(s)$, resulting from the expansion into partial-fraction form, are very simple functions of $s$; consequently, it is not necessary to refer to a Laplace transform table if we memorize several simple Laplace transform pairs. It should be noted, however, that in applying the partial-fraction expansion technique in the search for the inverse Laplace transform of $F(s) = B(s)/A(s)$ the roots of the denominator polynomial $A(s)$ must be obtained in advance. That is, this method does not apply until the denominator polynomial has been factored.

**Partial-fraction expansion when $F(s)$ involves distinct poles only.** Consider $F(s)$ written in the factored form

$$F(s) = \frac{B(s)}{A(s)} = \frac{K(s + z_1)(s + z_2)\cdots(s + z_m)}{(s + p_1)(s + p_2)\cdots(s + p_n)}, \qquad \text{for } m < n$$

where $p_1, p_2, \ldots, p_n$ and $z_1, z_2, \ldots, z_m$ are either real or complex quantities, but for each complex $p_i$ or $z_i$ there will occur the complex conjugate of $p_i$ or $z_i$, respectively. If $F(s)$ involves distinct poles only, then it can be expanded into a sum of simple partial fractions as follows:

$$F(s) = \frac{B(s)}{A(s)} = \frac{a_1}{s + p_1} + \frac{a_2}{s + p_2} + \cdots + \frac{a_n}{s + p_n} \tag{2–14}$$

where $a_k$ $(k = 1, 2, \ldots, n)$ are constants. The coefficient $a_k$ is called the *residue* at the pole at $s = -p_k$. The value of $a_k$ can be found by multiplying both sides of Equation (2–14) by $(s + p_k)$ and letting $s = -p_k$, which gives

$$\begin{aligned}\left[(s + p_k)\frac{B(s)}{A(s)}\right]_{s=-p_k} &= \left[\frac{a_1}{s + p_1}(s + p_k) + \frac{a_2}{s + p_2}(s + p_k)\right. \\ &\qquad \left. + \cdots + \frac{a_k}{s + p_k}(s + p_k) + \cdots + \frac{a_n}{s + p_n}(s + p_k)\right]_{s=-p_k} \\ &= a_k\end{aligned}$$

We see that all the expanded terms drop out with the exception of $a_k$. Thus the residue $a_k$ is found from

$$a_k = \left[(s + p_k)\frac{B(s)}{A(s)}\right]_{s=-p_k} \tag{2–15}$$

Note that, since $f(t)$ is a real function of time, if $p_1$ and $p_2$ are complex conjugates, then the residues $a_1$ and $a_2$ are also complex conjugates. Only one of the conjugates, $a_1$ or $a_2$, needs to be evaluated because the other is known automatically.

Since

$$\mathscr{L}^{-1}\left[\frac{a_k}{s + p_k}\right] = a_k e^{-p_k t}$$

$f(t)$ is obtained as

$$f(t) = \mathscr{L}^{-1}[F(s)] = a_1 e^{-p_1 t} + a_2 e^{-p_2 t} + \cdots + a_n e^{-p_n t}, \qquad \text{for } t \geq 0$$

**EXAMPLE 2–3** Find the inverse Laplace transform of

$$F(s) = \frac{s + 3}{(s + 1)(s + 2)}$$

The partial-fraction expansion of $F(s)$ is

$$F(s) = \frac{s + 3}{(s + 1)(s + 2)} = \frac{a_1}{s + 1} + \frac{a_2}{s + 2}$$

where $a_1$ and $a_2$ are found by using Equation (2–15):

$$a_1 = \left[(s + 1)\frac{s + 3}{(s + 1)(s + 2)}\right]_{s=-1} = \left[\frac{s + 3}{s + 2}\right]_{s=-1} = 2$$

$$a_2 = \left[(s + 2)\frac{s + 3}{(s + 1)(s + 2)}\right]_{s=-2} = \left[\frac{s + 3}{s + 1}\right]_{s=-2} = -1$$

Thus

$$\begin{aligned} f(t) &= \mathscr{L}^{-1}[F(s)] \\ &= \mathscr{L}^{-1}\left[\frac{2}{s + 1}\right] + \mathscr{L}^{-1}\left[\frac{-1}{s + 2}\right] \\ &= 2e^{-t} - e^{-2t}, \qquad \text{for } t \geq 0 \end{aligned}$$

**EXAMPLE 2–4** Obtain the inverse Laplace transform of

$$G(s) = \frac{s^3 + 5s^2 + 9s + 7}{(s + 1)(s + 2)}$$

Here, since the degree of the numerator polynomial is higher than that of the denominator polynomial, we must divide the numerator by the denominator.

$$G(s) = s + 2 + \frac{s + 3}{(s + 1)(s + 2)}$$

Note that the Laplace transform of the unit-impulse function $\delta(t)$ is 1 and that the Laplace transform of $d\delta(t)/dt$ is $s$. The third term on the right-hand side of this last equation is $F(s)$ in Example 2–3. So the inverse Laplace transform of $G(s)$ is given as

$$g(t) = \frac{d}{dt}\delta(t) + 2\delta(t) + 2e^{-t} - e^{-2t}, \qquad \text{for } t \geq 0-$$

**EXAMPLE 2–5** Find the inverse Laplace transform of

$$F(s) = \frac{2s + 12}{s^2 + 2s + 5}$$

Notice that the denominator polynomial can be factored as

$$s^2 + 2s + 5 = (s + 1 + j2)(s + 1 - j2)$$

If the function $F(s)$ involves a pair of complex-conjugate poles, it is convenient not to expand $F(s)$ into the usual partial fractions but to expand it into the sum of a damped sine and a damped cosine function.

Noting that $s^2 + 2s + 5 = (s + 1)^2 + 2^2$ and referring to the Laplace transforms of $e^{-\alpha t} \sin \omega t$ and $e^{-\alpha t} \cos \omega t$, rewritten thus,

$$\mathcal{L}[e^{-\alpha t} \sin \omega t] = \frac{\omega}{(s + \alpha)^2 + \omega^2}$$

$$\mathcal{L}[e^{-\alpha t} \cos \omega t] = \frac{s + \alpha}{(s + \alpha)^2 + \omega^2}$$

the given $F(s)$ can be written as a sum of a damped sine and a damped cosine function.

$$F(s) = \frac{2s + 12}{s^2 + 2s + 5} = \frac{10 + 2(s + 1)}{(s + 1)^2 + 2^2}$$
$$= 5\frac{2}{(s + 1)^2 + 2^2} + 2\frac{s + 1}{(s + 1)^2 + 2^2}$$

It follows that

$$f(t) = \mathcal{L}^{-1}[F(s)]$$
$$= 5\mathcal{L}^{-1}\left[\frac{2}{(s + 1)^2 + 2^2}\right] + 2\mathcal{L}^{-1}\left[\frac{s + 1}{(s + 1)^2 + 2^2}\right]$$
$$= 5e^{-t} \sin 2t + 2e^{-t} \cos 2t, \qquad \text{for } t \geq 0$$

**Partial-fraction expansion when $F(s)$ involves multiple poles.** Instead of discussing the general case, we shall use an example to show how to obtain the partial-fraction expansion of $F(s)$. (See also Problem A-2-16.)

Consider the following $F(s)$:

$$F(s) = \frac{s^2 + 2s + 3}{(s + 1)^3}$$

The partial-fraction expansion of this $F(s)$ involves three terms,

$$F(s) = \frac{B(s)}{A(s)} = \frac{b_1}{s + 1} + \frac{b_2}{(s + 1)^2} + \frac{b_3}{(s + 1)^3}$$

where $b_3$, $b_2$, and $b_1$ are determined as follows. By multiplying both sides of this last equation by $(s + 1)^3$, we have

$$(s + 1)^3 \frac{B(s)}{A(s)} = b_1(s + 1)^2 + b_2(s + 1) + b_3 \tag{2–16}$$

Then letting $s = -1$, Equation (2–16) gives

$$\left[(s + 1)^3 \frac{B(s)}{A(s)}\right]_{s=-1} = b_3$$

Also, differentiation of both sides of Equation (2–16) with respect to $s$ yields

$$\frac{d}{ds}\left[(s + 1)^3 \frac{B(s)}{A(s)}\right] = b_2 + 2b_1(s + 1) \tag{2–17}$$

If we let $s = -1$ in equation (2–17), then

$$\frac{d}{ds}\left[(s + 1)^3 \frac{B(s)}{A(s)}\right]_{s=-1} = b_2$$

By differentiating both sides of equation (2–17) with respect to $s$, the result is

$$\frac{d^2}{ds^2}\left[(s + 1)^3 \frac{B(s)}{A(s)}\right] = 2b_1$$

From the preceding analysis it can be seen that the values of $b_3$, $b_2$, and $b_1$ are found systematically as follows:

$$\begin{aligned}
b_3 &= \left[(s + 1)^3 \frac{B(s)}{A(s)}\right]_{s=-1} \\
&= (s^2 + 2s + 3)_{s=-1} \\
&= 2 \\
b_2 &= \left\{\frac{d}{ds}\left[(s + 1)^3 \frac{B(s)}{A(s)}\right]\right\}_{s=-1} \\
&= \left[\frac{d}{ds}(s^2 + 2s + 3)\right]_{s=-1} \\
&= (2s + 2)_{s=-1} \\
&= 0 \\
b_1 &= \frac{1}{2!}\left\{\frac{d^2}{ds^2}\left[(s + 1)^3 \frac{B(s)}{A(s)}\right]\right\}_{s=-1} \\
&= \frac{1}{2!}\left[\frac{d^2}{ds^2}(s^2 + 2s + 3)\right]_{s=-1} \\
&= \frac{1}{2}(2) = 1
\end{aligned}$$

We thus obtain

$$
\begin{aligned}
f(t) &= \mathcal{L}^{-1}[F(s)] \\
&= \mathcal{L}^{-1}\left[\frac{1}{s+1}\right] + \mathcal{L}^{-1}\left[\frac{0}{(s+1)^2}\right] + \mathcal{L}^{-1}\left[\frac{2}{(s+1)^3}\right] \\
&= e^{-t} + 0 + t^2 e^{-t} \\
&= (1 + t^2)e^{-t}, \qquad \text{for } t \geq 0
\end{aligned}
$$

**Comments.** For complicated functions with denominators involving higher-order polynomials, partial-fraction expansion may be quite time consuming. In such a case, use of MATLAB is recommended. (See Section 2–6.)

## 2–6 PARTIAL-FRACTION EXPANSION WITH MATLAB

MATLAB has a command to obtain the partial-fraction expansion of $B(s)/A(s)$.

Consider the transfer function

$$
\frac{B(s)}{A(s)} = \frac{\text{num}}{\text{den}} = \frac{b_0 s^n + b_1 s^{n-1} + \cdots + b_n}{s^n + a_1 s^{n-1} + \cdots + a_n}
$$

where some of $a_i$ and $b_j$ may be zero. In MATLAB row vectors num and den specify the coefficients of the numerator and denominator of the transfer function. That is,

```
num = [b0  b1  ...  bn]
den = [1   a1  ...  an]
```

The command

```
[r,p,k] = residue(num,den)
```

finds the residues, poles, and direct terms of a partial-fraction expansion of the ratio of two polynomials $B(s)$ and $A(s)$.

The partial-fraction expansion of $B(s)/A(s)$ is given by

$$
\frac{B(s)}{A(s)} = \frac{r(1)}{s - p(1)} + \frac{r(2)}{s - p(2)} + \cdots + \frac{r(n)}{s - p(n)} + k(s) \tag{2–18}
$$

Comparing Equations (2–14) and (2–18), we note that $p(1) = -p_1, p(2) = -p_2, \ldots, p(n) = -p_n$; $r(1) = a_1, r(2) = a_2, \ldots, r(n) = a_n$. [$k(s)$ is a direct term.]

**EXAMPLE 2–6** Consider the following transfer function:

$$
\frac{B(s)}{A(s)} = \frac{2s^3 + 5s^2 + 3s + 6}{s^3 + 6s^2 + 11s + 6}
$$

For this function.

```
num = [2   5   3   6]
den = [1   6   11   6]
```

The command

```
[r,p,k] = residue(num,den)
```

gives the following result:

```
[r,p,k] = residue(num,den)

r =

   -6.0000
   -4.0000
    3.0000

p =

   -3.0000
   -2.0000
   -1.0000

k =

    2
```

(Note that the residues are returned in column vector r, the pole locations in column vector p, and the direct term in row vector k.) This is the MATLAB representation of the following partial-fraction expansion of $B(s)/A(s)$;

$$\frac{B(s)}{A(s)} = \frac{2s^3 + 5s^2 + 3s + 6}{s^3 + 6s^2 + 11s + 6}$$

$$= \frac{-6}{s+3} + \frac{-4}{s+2} + \frac{3}{s+1} + 2$$

The command

```
[num,den] = residue(r,p,k)
```

where r, p, and k are as given in the previous MATLAB output, converts the partial-fraction expansion back to the polynomial ratio $B(s)/A(s)$, as follows:

```
[num,den] = residue(r,p,k)

num =

    2.0000    5.0000    3.0000    6.0000

den =

    1.0000    6.0000   11.0000    6.0000
```

Note that if $p(j) = p(j+1) = \cdots = p(j+m-1)$ [that is, $p_j = p_{j+1} = \ldots = p_{j+m-1}$], the pole $p(j)$ is a pole of multiplicity $m$. In such a case, the expansion includes terms of the form

$$\frac{r(j)}{s - p(j)} + \frac{r(j+1)}{[s - p(j)]^2} + \cdots + \frac{r(j+m-1)}{[s - p(j)]^m}$$

For details, see Example 2–7.

**EXAMPLE 2–7** Expand the following $B(s)/A(s)$ into partial-fractions with MATLAB.

$$\frac{B(s)}{A(s)} = \frac{s^2 + 2s + 3}{(s+1)^3} = \frac{s^2 + 2s + 3}{s^3 + 3s^2 + 3s + 1}$$

For this function, we have

```
num = [0  1  2  3]
den = [1  3  3  1]
```

The command

```
[r,p,k] = residue(num,den)
```

gives the result shown on the next page. It is the MATLAB representation of the following partial-fraction expansion of $B(s)/A(s)$:

$$\frac{B(s)}{A(s)} = \frac{1}{s+1} + \frac{0}{(s+1)^2} + \frac{2}{(s+1)^3}$$

Note that the direct term k is zero.

```
num = [0   1   2   3];
den = [1   3   3   1];
[r,p,k] = residue(num,den)

r =

    1.0000
    0.0000
    2.0000

p =

   -1.0000
   -1.0000
   -1.0000

k =

    []
```

## 2-7 SOLVING LINEAR, TIME-INVARIANT, DIFFERENTIAL EQUATIONS

In this section we are concerned with the use of the Laplace transform method in solving linear, time-invariant, differential equations.

The Laplace transform method yields the complete solution (complementary solution and particular solution) of linear, time-invariant, differential equations. Classical methods for finding the complete solution of a differential equation require the evaluation of the integration constants from the initial conditions. In the case of the Laplace transform method, however, this requirement is unnecessary because the initial conditions are automatically included in the Laplace transform of the differential equation.

If all initial conditions are zero, then the Laplace transform of the differential equation is obtained simply by replacing $d/dt$ with $s$, $d^2/dt^2$ with $s^2$, and so on.

In solving linear, time-invariant, differential equations by the Laplace transform method, two steps are involved.

1. By taking the Laplace transform of each term in the given differential equation, convert the differential equation into an algebraic equation in $s$ and obtain the

expression for the Laplace transform of the dependent variable by rearranging the algebraic equation.

2. The time solution of the differential equation is obtained by finding the inverse Laplace transform of the dependent variable.

In the following discussion, two examples are used to demonstrate the solution of linear, time-invariant, differential equations by the Laplace transform method.

**EXAMPLE 2–8** Find the solution $x(t)$ of the differential equation

$$\ddot{x} + 3\dot{x} + 2x = 0, \qquad x(0) = a, \qquad \dot{x}(0) = b$$

where $a$ and $b$ are constants.

By writing the Laplace transform of $x(t)$ as $X(s)$ or

$$\mathscr{L}[x(t)] = X(s)$$

we obtain

$$\mathscr{L}[\dot{x}] = sX(s) - x(0)$$

$$\mathscr{L}[\ddot{x}] = s^2X(s) - sx(0) - \dot{x}(0)$$

And so the given differential equation becomes

$$[s^2X(s) - sx(0) - \dot{x}(0)] + 3[sX(s) - x(0)] + 2X(s) = 0$$

By substituting the given initial conditions into this last equation, we obtain

$$[s^2X(s) - as - b] + 3[sX(s) - a] + 2X(s) = 0$$

or

$$(s^2 + 3s + 2)X(s) = as + b + 3a$$

Solving for $X(s)$, we have

$$X(s) = \frac{as + b + 3a}{s^2 + 3s + 2} = \frac{as + b + 3a}{(s + 1)(s + 2)} = \frac{2a + b}{s + 1} - \frac{a + b}{s + 2}$$

The inverse Laplace transform of $X(s)$ gives

$$\begin{aligned} x(t) &= \mathscr{L}^{-1}[X(s)] = \mathscr{L}^{-1}\left[\frac{2a + b}{s + 1}\right] - \mathscr{L}^{-1}\left[\frac{a + b}{s + 2}\right] \\ &= (2a + b)e^{-t} - (a + b)e^{-2t}, \qquad \text{for } t \geq 0 \end{aligned}$$

which is the solution of the given differential equation. Notice that the initial conditions $a$ and $b$ appear in the solution. Thus $x(t)$ has no undetermined constants.

**EXAMPLE 2–9** Find the solution $x(t)$ of the differential equation

$$\ddot{x} + 2\dot{x} + 5x = 3, \qquad x(0) = 0, \qquad \dot{x}(0) = 0$$

Noting that $\mathscr{L}[3] = 3/s$, $x(0) = 0$, and $\dot{x}(0) = 0$, the Laplace transform of the differential equation becomes

$$s^2X(s) + 2sX(s) + 5X(s) = \frac{3}{s}$$

Solving for $X(s)$, we find

$$X(s) = \frac{3}{s(s^2 + 2s + 5)} = \frac{3}{5}\frac{1}{s} - \frac{3}{5}\frac{s + 2}{s^2 + 2s + 5}$$

$$= \frac{3}{5}\frac{1}{s} - \frac{3}{10}\frac{2}{(s + 1)^2 + 2^2} - \frac{3}{5}\frac{s + 1}{(s + 1)^2 + 2^2}$$

Hence the inverse Laplace transform becomes

$$x(t) = \mathcal{L}^{-1}[X(s)]$$

$$= \frac{3}{5}\mathcal{L}^{-1}\left[\frac{1}{s}\right] - \frac{3}{10}\mathcal{L}^{-1}\left[\frac{2}{(s + 1)^2 + 2^2}\right] - \frac{3}{5}\mathcal{L}^{-1}\left[\frac{s + 1}{(s + 1)^2 + 2^2}\right]$$

$$= \frac{3}{5} - \frac{3}{10}e^{-t}\sin 2t - \frac{3}{5}e^{-t}\cos 2t, \qquad \text{for } t \geq 0$$

which is the solution of the given differential equation.

## EXAMPLE PROBLEMS AND SOLUTIONS

**A-2-1.** Find the poles of the following $F(s)$:

$$F(s) = \frac{1}{1 - e^{-s}}$$

**Solution.** The poles are found from

$$e^{-s} = 1$$

or

$$e^{-(\sigma + j\omega)} = e^{-\sigma}(\cos\omega - j\sin\omega) = 1$$

From this it follows that $\sigma = 0$, $\omega = \pm 2n\pi$ $(n = 0,1,2,\ldots)$. Thus, the poles are located at

$$s = \pm j2n\pi \qquad (n = 0, 1, 2, \ldots)$$

**A-2-2.** Find the Laplace transform of $f(t)$ defined by

$$\begin{aligned} f(t) &= 0, && \text{for } t < 0 \\ &= te^{-3t}, && \text{for } t \geq 0 \end{aligned}$$

**Solution.** Since

$$\mathcal{L}[t] = G(s) = \frac{1}{s^2}$$

referring to Equation (2–6), we obtain

$$F(s) = \mathcal{L}[te^{-3t}] = G(s + 3) = \frac{1}{(s + 3)^2}$$

**A-2-3.** What is the Laplace transform of

$$\begin{aligned} f(t) &= 0, && \text{for } t < 0 \\ &= \sin(\omega t + \theta), && \text{for } t \geq 0 \end{aligned}$$

where $\theta$ is a constant?

**Solution.** Noting that

$$\sin(\omega t + \theta) = \sin \omega t \cos \theta + \cos \omega t \sin \theta$$

we have

$$\begin{aligned}
\mathscr{L}[\sin(\omega t + \theta)] &= \cos\theta\, \mathscr{L}[\sin \omega t] + \sin\theta\, \mathscr{L}[\cos \omega t] \\
&= \cos\theta \frac{\omega}{s^2 + \omega^2} + \sin\theta \frac{s}{s^2 + \omega^2} \\
&= \frac{\omega \cos\theta + s \sin\theta}{s^2 + \omega^2}
\end{aligned}$$

**A-2-4.** Find the Laplace transform $F(s)$ of the function $f(t)$ shown in Figure 2–3. Also find the limiting value of $F(s)$ as $a$ approaches zero.

**Solution.** The function $f(t)$ can be written

$$f(t) = \frac{1}{a^2} 1(t) - \frac{2}{a^2} 1(t - a) + \frac{1}{a^2} 1(t - 2a)$$

Then

$$\begin{aligned}
F(s) &= \mathscr{L}[f(t)] \\
&= \frac{1}{a^2} \mathscr{L}[1(t)] - \frac{2}{a^2} \mathscr{L}[1(t - a)] + \frac{1}{a^2} \mathscr{L}[1(t - 2a)] \\
&= \frac{1}{a^2} \frac{1}{s} - \frac{2}{a^2} \frac{1}{s} e^{-as} + \frac{1}{a^2} \frac{1}{s} e^{-2as} \\
&= \frac{1}{a^2 s}(1 - 2e^{-as} + e^{-2as})
\end{aligned}$$

As $a$ approaches zero, we have

$$\begin{aligned}
\lim_{a\to 0} F(s) &= \lim_{a\to 0} \frac{1 - 2e^{-as} + e^{-2as}}{a^2 s} = \lim_{a\to 0} \frac{\dfrac{d}{da}(1 - 2e^{-as} + e^{-2as})}{\dfrac{d}{da}(a^2 s)} \\
&= \lim_{a\to 0} \frac{2se^{-as} - 2se^{-2as}}{2as} = \lim_{a\to 0} \frac{e^{-as} - e^{-2as}}{a} \\
&= \lim_{a\to 0} \frac{\dfrac{d}{da}(e^{-as} - e^{-2as})}{\dfrac{d}{da}(a)} = \lim_{a\to 0} \frac{-se^{-as} + 2se^{-2as}}{1} \\
&= -s + 2s = s
\end{aligned}$$

**A-2-5.** Find the initial value of $df(t)/dt$ when the Laplace transform of $f(t)$ is given by

$$F(s) = \mathscr{L}[f(t)] = \frac{2s + 1}{s^2 + s + 1}$$

**Solution.** Using the initial-value theorem,

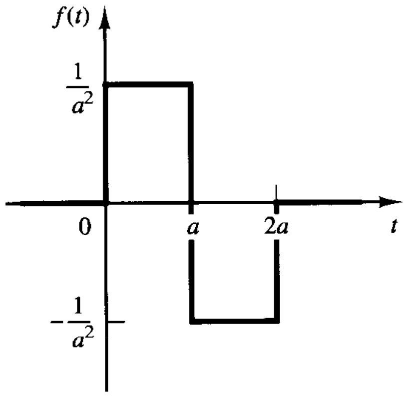



**Figure 2–3**
Function $f(t)$.

$$\lim_{t\to 0+} f(t) = f(0+) = \lim_{s\to\infty} sF(s) = \lim_{s\to\infty} \frac{s(2s+1)}{s^2+s+1} = 2$$

Since the $\mathscr{L}_+$ transform of $df(t)/dt = g(t)$ is given by

$$\begin{aligned}\mathscr{L}_+[g(t)] &= sF(s) - f(0+) \\ &= \frac{s(2s+1)}{s^2+s+1} - 2 = \frac{-s-2}{s^2+s+1}\end{aligned}$$

the initial value of $df(t)/dt$ is obtained as

$$\begin{aligned}\lim_{t\to 0+} \frac{df(t)}{dt} &= g(0+) = \lim_{s\to\infty} s[sF(s) - f(0+)] \\ &= \lim_{s\to\infty} \frac{-s^2-2s}{s^2+s+1} = -1\end{aligned}$$

**A-2-6.** The derivative of the unit-impulse function $\delta(t)$ is called a *unit-doublet* function. (Thus, the integral of the unit-doublet function is the unit-impulse function.) Mathematically, an example of the unit-doublet function, which is usually denoted by $u_2(t)$, may be given by

$$u_2(t) = \lim_{t_0\to 0} \frac{1(t) - 2[1(t-t_0)] + 1(t-2t_0)}{t_0^2}$$

Obtain the Laplace transform of $u_2(t)$.

**Solution.** The Laplace transform of $u_2(t)$ is given by

$$\begin{aligned}\mathscr{L}[u_2(t)] &= \lim_{t_0\to 0} \frac{1}{t_0^2}\left(\frac{1}{s} - \frac{2}{s}e^{-t_0 s} + \frac{1}{s}e^{-2t_0 s}\right) \\ &= \lim_{t_0\to 0} \frac{1}{t_0^2 s}\left[1 - 2\left(1 - t_0 s + \frac{t_0^2 s^2}{2} + \cdots\right) + \left(1 - 2t_0 s + \frac{4t_0^2 s^2}{2} + \cdots\right)\right] \\ &= \lim_{t_0\to 0} \frac{1}{t_0^2 s}[t_0^2 s^2 + (\text{higher-order terms in } t_0 s)] = s\end{aligned}$$

**A-2-7.** Find the Laplace transform of $f(t)$ defined by

$$\begin{aligned}f(t) &= 0, && \text{for } t < 0 \\ &= t^2 \sin \omega t, && \text{for } t \geq 0\end{aligned}$$

**Solution.** Since

$$\mathscr{L}[\sin \omega t] = \frac{\omega}{s^2 + \omega^2}$$

applying the complex-differentiation theorem

$$\mathscr{L}[t^2 f(t)] = \frac{d^2}{ds^2} F(s)$$

to this problem, we have

$$\mathscr{L}[f(t)] = \mathscr{L}[t^2 \sin \omega t] = \frac{d^2}{ds^2}\left[\frac{\omega}{s^2 + \omega^2}\right] = \frac{-2\omega^3 + 6\omega s^2}{(s^2 + \omega^2)^3}$$

**A-2-8.** Prove that if $f(t)$ is of exponential order and if $\int_0^\infty f(t)\, dt$ exists [which means that $\int_0^\infty f(t)\, dt$ assumes a definite value] then

$$\int_0^\infty f(t)\, dt = \lim_{s \to 0} F(s)$$

where $F(s) = \mathscr{L}[f(t)]$.

**Solution.** Note that

$$\int_0^\infty f(t)\, dt = \lim_{t \to \infty} \int_0^t f(t)\, dt$$

Referring to equation (2–9),

$$\mathscr{L}\left[\int_0^t f(t)\, dt\right] = \frac{F(s)}{s}$$

Since $\int_0^\infty f(t)\, dt$ exists, by applying the final-value theorem to this case,

$$\lim_{t \to \infty} \int_0^t f(t)\, dt = \lim_{s \to 0} s \frac{F(s)}{s}$$

or

$$\int_0^\infty f(t)\, dt = \lim_{s \to 0} F(s)$$

**A-2-9.** Determine the Laplace transform of the convolution integral

$$f_1(t) * f_2(t) = \int_0^t \tau[1 - e^{-(t-\tau)}]\, d\tau = \int_0^t (t - \tau)(1 - e^{-\tau})\, d\tau$$

where

$$\begin{aligned} f_1(t) &= f_2(t) = 0, && \text{for } t < 0 \\ f_1(t) &= t, && \text{for } t \geq 0 \\ f_2(t) &= 1 - e^{-t}, && \text{for } t \geq 0 \end{aligned}$$

**Solution.** Note that

$$\mathcal{L}[t] = F_1(s) = \frac{1}{s^2}$$

$$\mathcal{L}[1 - e^{-t}] = F_2(s) = \frac{1}{s} - \frac{1}{s+1}$$

The Laplace transform of the convolution integral is given by

$$\mathcal{L}[f_1(t)*f_2(t)] = F_1(s)F_2(s) = \frac{1}{s^2}\left(\frac{1}{s} - \frac{1}{s+1}\right)$$

$$= \frac{1}{s^3} - \frac{1}{s^2(s+1)} = \frac{1}{s^3} - \frac{1}{s^2} + \frac{1}{s} - \frac{1}{s+1}$$

To verify that it is indeed the Laplace transform of the convolution integral, let us first perform integration of the convolution integral and then take its Laplace transform.

$$f_1(t)*f_2(t) = \int_0^t \tau[1 - e^{-(t-\tau)}]\,d\tau = \int_0^t (t-\tau)(1 - e^{-\tau})\,d\tau$$

$$= \frac{t^2}{2} - t + 1 - e^{-t}$$

And so

$$\mathcal{L}\left[\frac{t^2}{2} - t + 1 - e^{-t}\right] = \frac{1}{s^3} - \frac{1}{s^2} + \frac{1}{s} - \frac{1}{s+1}$$

**A-2-10.** Prove that if $f(t)$ is a periodic function with period $T$ then

$$\mathcal{L}[f(t)] = \frac{\int_0^T f(t)e^{-st}\,dt}{1 - e^{-Ts}}$$

**Solution.**

$$\mathcal{L}[f(t)] = \int_0^\infty f(t)e^{-st}\,dt = \sum_{n=0}^{\infty} \int_{nT}^{(n+1)T} f(t)e^{-st}\,dt$$

By changing the independent variable from $t$ to $\tau$, where $\tau = t - nT$,

$$\mathcal{L}[f(t)] = \sum_{n=0}^{\infty} e^{-nTs} \int_0^T f(\tau)e^{-s\tau}\,d\tau$$

Noting that

$$\sum_{n=0}^{\infty} e^{-nTs} = 1 + e^{-Ts} + e^{-2Ts} + \cdots$$

$$= 1 + e^{-Ts}(1 + e^{-Ts} + e^{-2Ts} + \cdots)$$

$$= 1 + e^{-Ts}\left(\sum_{n=0}^{\infty} e^{-nTs}\right)$$

we obtain

$$\sum_{n=0}^{\infty} e^{-nTs} = \frac{1}{1 - e^{-Ts}}$$

It follows that

$$\mathscr{L}[f(t)] = \frac{\int_0^T f(t)e^{-st}\,dt}{1 - e^{-Ts}}$$

**A-2-11.** What is the Laplace transform of the periodic function shown in Figure 2–4?

**Solution.** Note that

$$\int_0^T f(t)e^{-st}\,dt = \int_0^{T/2} e^{-st}\,dt + \int_{T/2}^{T} (-1)e^{-st}\,dt$$

$$= \frac{e^{-st}}{-s}\bigg|_0^{T/2} - \frac{e^{-st}}{-s}\bigg|_{T/2}^{T}$$

$$= \frac{e^{-(1/2)Ts} - 1}{-s} + \frac{e^{-Ts} - e^{-(1/2)Ts}}{s}$$

$$= \frac{1}{s}[e^{-Ts} - 2e^{-(1/2)Ts} + 1]$$

$$= \frac{1}{s}[1 - e^{-(1/2)Ts}]^2$$

Referring to Problem A-2-10, we have

$$F(s) = \frac{\int_0^T f(t)e^{-st}\,dt}{1 - e^{-Ts}} = \frac{(1/s)[1 - e^{-(1/2)Ts}]^2}{1 - e^{-Ts}}$$

$$= \frac{1 - e^{-(1/2)Ts}}{s[1 + e^{-(1/2)Ts}]} = \frac{1}{s}\tanh\frac{Ts}{4}$$

**A-2-12.** Find the inverse Laplace transform of $F(s)$, where

$$F(s) = \frac{1}{s(s^2 + 2s + 2)}$$

**Solution.** Since

$$s^2 + 2s + 2 = (s + 1 + j1)(s + 1 - j1)$$

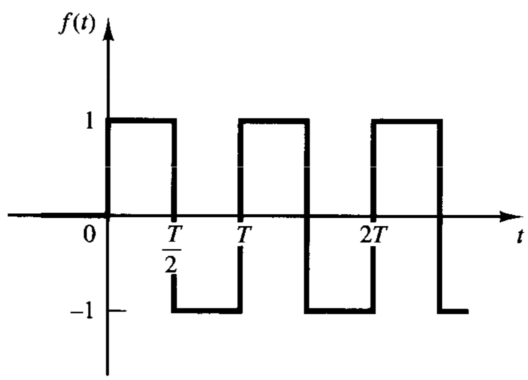



**Figure 2–4**
Periodic function (square wave).

we notice that $F(s)$ involves a pair of complex-conjugate poles, and so we expand $F(s)$ into the form

$$F(s) = \frac{1}{s(s^2 + 2s + 2)} = \frac{a_1}{s} + \frac{a_2 s + a_3}{s^2 + 2s + 2}$$

where $a_1$, $a_2$, and $a_3$ are determined from

$$1 = a_1(s^2 + 2s + 2) + (a_2 s + a_3)s$$

By comparing coefficients of $s^2$, $s$, and $s^0$ terms on both sides of this last equation, respectively, we obtain

$$a_1 + a_2 = 0, \qquad 2a_1 + a_3 = 0, \qquad 2a_1 = 1$$

from which

$$a_1 = \frac{1}{2}, \qquad a_2 = -\frac{1}{2}, \qquad a_3 = -1$$

Therefore,

$$\begin{aligned} F(s) &= \frac{1}{2}\frac{1}{s} - \frac{1}{2}\frac{s+2}{s^2 + 2s + 2} \\ &= \frac{1}{2}\frac{1}{s} - \frac{1}{2}\frac{1}{(s+1)^2 + 1^2} - \frac{1}{2}\frac{s+1}{(s+1)^2 + 1^2} \end{aligned}$$

The inverse Laplace transform of $F(s)$ gives

$$f(t) = \frac{1}{2} - \frac{1}{2}e^{-t}\sin t - \frac{1}{2}e^{-t}\cos t, \qquad \text{for } t \geq 0$$

**A-2-13.** Obtain the inverse Laplace transform of

$$F(s) = \frac{5(s+2)}{s^2(s+1)(s+3)}$$

**Solution.**

$$F(s) = \frac{5(s+2)}{s^2(s+1)(s+3)} = \frac{b_1}{s} + \frac{b_2}{s^2} + \frac{a_1}{s+1} + \frac{a_2}{s+3}$$

where

$$a_1 = \frac{5(s+2)}{s^2(s+3)}\bigg|_{s=-1} = \frac{5}{2}$$

$$a_2 = \frac{5(s+2)}{s^2(s+1)}\bigg|_{s=-3} = \frac{5}{18}$$

$$b_2 = \frac{5(s+2)}{(s+1)(s+3)}\bigg|_{s=0} = \frac{10}{3}$$

$$\begin{aligned} b_1 &= \frac{d}{ds}\left[\frac{5(s+2)}{(s+1)(s+3)}\right]_{s=0} \\ &= \frac{5(s+1)(s+3) - 5(s+2)(2s+4)}{(s+1)^2(s+3)^2}\bigg|_{s=0} = -\frac{25}{9} \end{aligned}$$

Thus

$$F(s) = -\frac{25}{9}\frac{1}{s} + \frac{10}{3}\frac{1}{s^2} + \frac{5}{2}\frac{1}{s+1} + \frac{5}{18}\frac{1}{s+3}$$

The inverse Laplace transform of $F(s)$ is

$$f(t) = -\frac{25}{9} + \frac{10}{3}t + \frac{5}{2}e^{-t} + \frac{5}{18}e^{-3t}, \qquad \text{for } t \geq 0$$

**A-2-14.** Find the inverse Laplace transform of

$$F(s) = \frac{s^4 + 2s^3 + 3s^2 + 4s + 5}{s(s+1)}$$

**Solution.** Since the numerator polynomial is of higher degree than the denominator polynomial, by dividing the numerator by the denominator until the remainder is a fraction, we obtain

$$F(s) = s^2 + s + 2 + \frac{2s+5}{s(s+1)} = s^2 + s + 2 + \frac{a_1}{s} + \frac{a_2}{s+1}$$

where

$$a_1 = \left.\frac{2s+5}{s+1}\right|_{s=0} = 5$$

$$a_2 = \left.\frac{2s+5}{s}\right|_{s=-1} = -3$$

It follows that

$$F(s) = s^2 + s + 2 + \frac{5}{s} - \frac{3}{s+1}$$

The inverse Laplace transform of $F(s)$ is

$$f(t) = \mathcal{L}^{-1}[F(s)] = \frac{d^2}{dt^2}\,\delta(t) + \frac{d}{dt}\,\delta(t) + 2\,\delta(t) + 5 - 3e^{-t}, \qquad \text{for } t \geq 0-$$

**A-2-15.** Derive the inverse Laplace transform of

$$F(s) = \frac{1}{s(s^2 + \omega^2)}$$

**Solution.**

$$F(s) = \frac{1}{s(s^2+\omega^2)} = \frac{1}{\omega^2}\frac{1}{s} - \frac{1}{\omega^2}\frac{s}{s^2+\omega^2}$$

Hence the inverse Laplace transform of $F(s)$ is obtained as

$$f(t) = \mathcal{L}^{-1}[F(s)] = \frac{1}{\omega^2}(1 - \cos\omega t), \qquad \text{for } t \geq 0$$

**A-2-16.** Obtain the inverse Laplace transform of the following $F(s)$:

$$F(s) = \frac{B(s)}{A(s)} = \frac{B(s)}{(s+p_1)^r(s+p_{r+1})(s+p_{r+2})\cdots(s+p_n)}$$

where the degree of polynomial $B(s)$ is lower than that of polynomial $A(s)$.

**Solution.** The partial-fraction expansion of $F(s)$ is

$$F(s) = \frac{B(s)}{A(s)} = \frac{b_1}{s + p_1} + \frac{b_2}{(s + p_1)^2} + \cdots + \frac{b_{r-1}}{(s + p_1)^{r-1}} + \frac{b_r}{(s + p_1)^r}$$

$$+ \frac{a_{r+1}}{s + p_{r+1}} + \frac{a_{r+2}}{s + p_{r+2}} + \cdots + \frac{a_n}{s + p_n} \tag{2–19}$$

where $b_r, b_{r-1}, \ldots, b_1$ are given by

$$b_r = \left[(s + p_1)^r \frac{B(s)}{A(s)}\right]_{s=-p_1}$$

$$b_{r-1} = \left\{\frac{d}{ds}\left[(s + p_1)^r \frac{B(s)}{A(s)}\right]\right\}_{s=-p_1}$$

$$\vdots$$

$$b_{r-j} = \frac{1}{j!}\left\{\frac{d^j}{ds^j}\left[(s + p_1)^r \frac{B(s)}{A(s)}\right]\right\}_{s=-p_1}$$

$$\vdots$$

$$b_1 = \frac{1}{(r-1)!}\left\{\frac{d^{r-1}}{ds^{r-1}}\left[(s + p_1)^r \frac{B(s)}{A(s)}\right]\right\}_{s=-p_1}$$

The foregoing relationships for the $b$'s may be obtained as follows: By multiplying both sides of Equation (2–19) by $(s + p_1)^r$ and letting $s$ approach $-p_1$, we obtain

$$b_r = \left[(s + p_1)^r \frac{B(s)}{A(s)}\right]_{s=-p_1}$$

If we multiply both sides of Equation (2–19) by $(s + p_1)^r$ and then differentiate with respect to $s$,

$$\frac{d}{ds}\left[(s + p_1)^r \frac{B(s)}{A(s)}\right] = b_r \frac{d}{ds}\left[\frac{(s + p_1)^r}{(s + p_1)^r}\right] + b_{r-1}\frac{d}{ds}\left[\frac{(s + p_1)^r}{(s + p_1)^{r-1}}\right]$$

$$+ \cdots + b_1 \frac{d}{ds}\left[\frac{(s + p_1)^r}{s + p_1}\right] + a_{r+1}\frac{d}{ds}\left[\frac{(s + p_1)^r}{s + p_{r+1}}\right]$$

$$+ \cdots + a_n \frac{d}{ds}\left[\frac{(s + p_1)^r}{s + p_n}\right]$$

The first term on the right-hand side of this last equation vanishes. The second term becomes $b_{r-1}$. Each of the other terms contains some power of $(s + p_1)$ as a factor, with the result that, when $s$ is allowed to approach $-p_1$, these terms drop out. Hence

$$b_{r-1} = \lim_{s \to -p_1} \frac{d}{ds}\left[(s + p_1)^r \frac{B(s)}{A(s)}\right]$$

$$= \left\{\frac{d}{ds}\left[(s + p_1)^r \frac{B(s)}{A(s)}\right]\right\}_{s=-p_1}$$

Similarly, by performing successive differentiations with respect to $s$ and by letting $s$ approach $-p_1$, we obtain equations for the $b_{r-j}$, where $j = 2, 3, \ldots, r - 1$.

Note that the inverse Laplace transform of $1/(s + p_1)^n$ is given by

$$\mathcal{L}^{-1}\left[\frac{1}{(s + p_1)^n}\right] = \frac{t^{n-1}}{(n-1)!} e^{-p_1 t}$$

The constants $a_{r+1}, a_{r+2}, \ldots, a_n$ in Equation (2–19) are determined from

$$a_k = \left[(s + p_k)\frac{B(s)}{A(s)}\right]_{s=-p_k}, \qquad \text{for } k = r + 1, r + 2, \ldots, n$$

The inverse Laplace transform of $F(s)$ is then obtained as follows:

$$f(t) = \mathcal{L}^{-1}[F(s)] = \left[b_1 + b_2 t + \cdots + \frac{b_{r-1}}{(r-2)!}t^{r-2} + \frac{b_r}{(r-1)!}t^{r-1}\right]e^{-p_1 t}$$
$$+ a_{r+1}e^{-p_{r+1}t} + a_{r+2}e^{-p_{r+2}t} + \cdots + a_n e^{-p_n t}, \qquad \text{for } t \geq 0$$

**A-2-17.** Find the Laplace transform of the following differential equation:

$$\ddot{x} + 3\dot{x} + 6x = 0, \qquad x(0) = 0, \qquad \dot{x}(0) = 3$$

Taking the inverse Laplace transform of $X(s)$, obtain the time solution $x(t)$.

**Solution.** The Laplace transform of the differential equation is

$$s^2X(s) - sx(0) - \dot{x}(0) + 3sX(s) - 3x(0) + 6X(s) = 0$$

Substituting the initial conditions and solving for $X(s)$,

$$X(s) = \frac{3}{s^2 + 3s + 6} = \frac{2\sqrt{3}}{\sqrt{5}} \frac{\frac{\sqrt{15}}{2}}{(s + 1.5)^2 + \left(\frac{\sqrt{15}}{2}\right)^2}$$

The inverse Laplace transform of $X(s)$ is

$$x(t) = \frac{2\sqrt{3}}{\sqrt{5}} e^{-1.5t} \sin\left(\frac{\sqrt{15}}{2}t\right)$$

## PROBLEMS

**B-2-1.** Find the Laplace transforms of the following functions:

(a)
$$f_1(t) = 0, \qquad \text{for } t < 0$$
$$= e^{-0.4t}\cos 12t, \qquad \text{for } t \geq 0$$

(b)
$$f_2(t) = 0, \qquad \text{for } t < 0$$
$$= \sin\left(4t + \frac{\pi}{3}\right), \qquad \text{for } t \geq 0$$

**B-2-2.** Find the Laplace transforms of the following functions:

(a)
$$f_1(t) = 0, \qquad \text{for } t < 0$$
$$= 3\sin(5t + 45°) \qquad \text{for } t \geq 0$$

(b)
$$f_2(t) = 0, \qquad \text{for } t < 0$$
$$= 0.03(1 - \cos 2t) \qquad \text{for } t \geq 0$$

**B-2-3.** Obtain the Laplace transform of the function defined by

$$f(t) = 0, \qquad \text{for } t < 0$$
$$= t^2e^{-at} \qquad \text{for } t \geq 0$$

**B-2-4.** Obtain the Laplace transform of the function defined by

$$\begin{aligned} f(t) &= 0, && \text{for } t < 0 \\ &= \cos 2\omega t \cdot \cos 3\omega t, && \text{for } t \geq 0 \end{aligned}$$

**B-2-5.** What is the Laplace transform of the function $f(t)$ shown in Figure 2–5?

**B-2-6.** Obtain the Laplace transform of the function $f(t)$ shown in Figure 2–6.

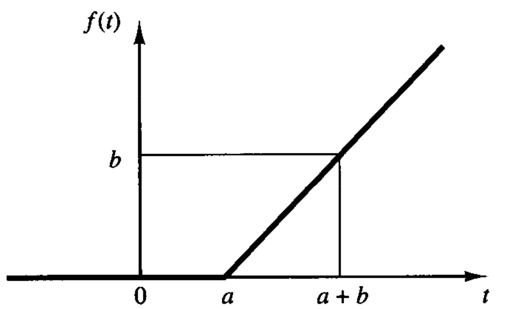



**Figure 2–5**
Function $f(t)$.

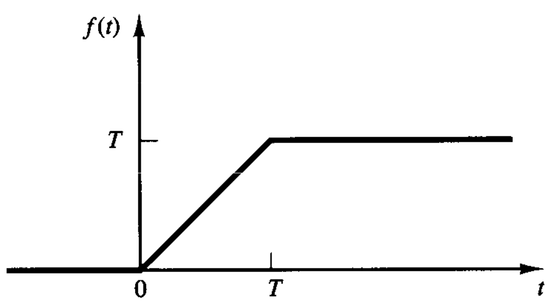



**Figure 2–6**
Function $f(t)$.

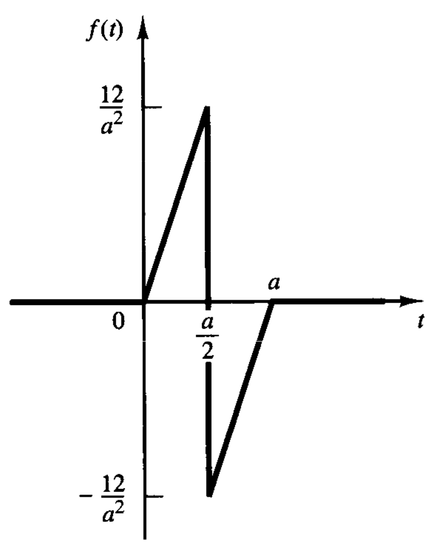



**Figure 2–7**
Function $f(t)$.

**B-2-7.** Find the Laplace transform of the function $f(t)$ shown in Figure 2–7. Also, find the limiting value of $\mathscr{L}[f(t)]$ as $a$ approaches zero.

**B-2-8.** By applying the final-value theorem, find the final value of $f(t)$ whose Laplace transform is given by

$$F(s) = \frac{10}{s(s+1)}$$

Verify this result by taking the inverse Laplace transform of $F(s)$ and letting $t \to \infty$.

**B-2-9.** Given

$$F(s) = \frac{1}{(s+2)^2}$$

determine the values of $f(0+)$ and $\dot{f}(0+)$. (Use the initial-value theorem.)

**B-2-10.** Find the inverse Laplace transform of

$$F(s) = \frac{s+1}{s(s^2+s+1)}$$

**B-2-11.** Find the inverse Laplace transforms of the following functions:

(a) $$F_1(s) = \frac{6s+3}{s^2}$$

(b) $$F_2(s) = \frac{5s+2}{(s+1)(s+2)^2}$$

**B-2-12.** Find the inverse Laplace transform of

$$F(s) = \frac{1}{s^2(s^2+\omega^2)}$$

**B-2-13.** What is the solution of the following differential equation?

$$2\ddot{x} + 7\dot{x} + 3x = 0, \qquad x(0) = 3, \qquad \dot{x}(0) = 0$$

**B-2-14.** Solve the differential equation

$$\dot{x} + 2x = \delta(t), \qquad x(0-) = 0$$

**B-2-15.** Solve the following differential equation:

$$\ddot{x} + 2\zeta\omega_n\dot{x} + \omega_n^2 x = 0, \qquad x(0) = a, \qquad \dot{x}(0) = b$$

where $a$ and $b$ are constants.

**B-2-16.** Obtain the solution of the differential equation

$$\dot{x} + ax = A \sin \omega t, \qquad x(0) = b$$

# 3

# Mathematical Modeling of Dynamic Systems

## 3–1 INTRODUCTION

In studying control systems the reader must be able to model dynamic systems and analyze 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 a reasonable mathematical model is the most important part of the entire analysis.

**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.** It is possible to improve the accuracy of a mathematical model by increasing its complexity. In some cases, we include hundreds of equations to describe a complete system. In obtaining a mathematical model, however, we must make a compromise between the simplicity of the model and the accuracy of the results of the analysis. If extreme accuracy is not needed, however, it is preferable to obtain only a reasonably simplified model. In fact, we are generally satisfied if we can obtain a mathematical model that is adequate for the problem under consideration. It is important to note, however, that the results obtained from the analysis are valid only to the extent that the model approximates a given dynamic system.

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 is, ones giving rise to partial differential equations) 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, we find it desirable first 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 complete analysis.

We must be well aware of the fact 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.

**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 (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.)

**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. Examples of nonlinear differential equations are

$$\frac{d^2x}{dt^2} + \left(\frac{dx}{dt}\right)^2 + x = A \sin \omega t$$

$$\frac{d^2x}{dt^2} + (x^2 - 1)\frac{dx}{dt} + x = 0$$

$$\frac{d^2x}{dt^2} + \frac{dx}{dt} + x + x^3 = 0$$

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. Examples of characteristic curves for these nonlinearities are shown in Figure 3–1.

Note that some important control systems are nonlinear for signals of any size. For example, in on–off control systems, the control action is either on or off, and there is no linear relationship between the input and output of the controller.

Procedures for finding the solutions of problems involving such nonlinear systems, in general, are extremely complicated. Because of this mathematical difficulty attached to nonlinear systems, one often finds it necessary to introduce "equivalent" linear systems in place of nonlinear ones. Such equivalent linear systems are valid for only a limited range of operation. Once a nonlinear system is approximated by a linear mathematical model, a number of linear tools may be applied for analysis and design purposes.

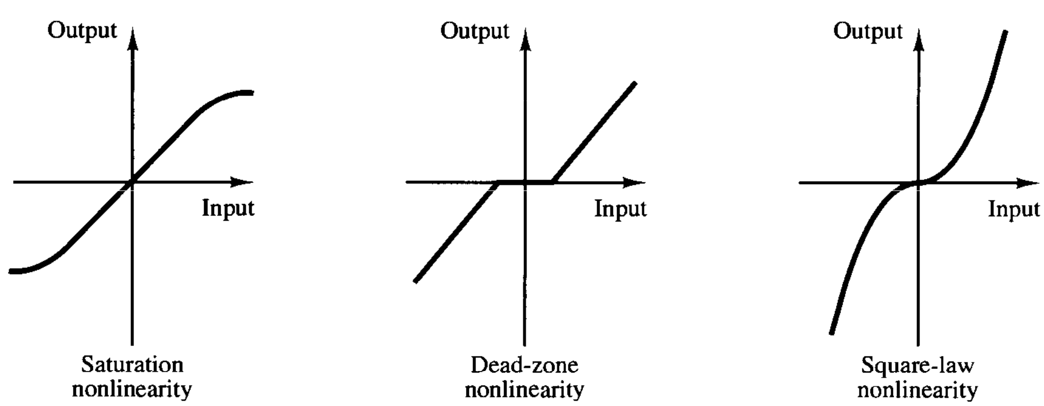



**Figure 3–1** Characteristic curves for various nonlinearities.

**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. We shall discuss a linearization technique in Section 3–10.

**Outline of the chapter.** Section 3–1 has presented an introduction to the mathematical modeling of dynamic systems, including discussions of linear and nonlinear systems. Section 3–2 presents the transfer function and impulse-response function. Section 3–3 introduces block diagrams and Section 3–4 discusses concepts of modeling in state space. Section 3–5 presents state-space representation of dynamic systems. Section 3–6 treats mathematical modeling of mechanical systems. We discuss Newton's approach to modeling mechanical systems. Section 3–7 deals with mathematical modeling of electrical circuits, Section 3–8 treats liquid-level systems, and Section 3–9 presents mathematical modeling of thermal systems. Finally, Section 3–10 discusses the linearization of nonlinear mathematical models. (Mathematical modeling of other types of systems is treated throughout the remaining chapters of the book.)

## 3–2 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 mechanical 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:

$$a_0 \overset{(n)}{y} + a_1 \overset{(n-1)}{y} + \cdots + a_{n-1}\dot{y} + a_n y$$
$$= b_0 \overset{(m)}{x} + b_1 \overset{(m-1)}{x} + \cdots + b_{m-1}\dot{x} + b_m x \qquad (n \geq m) \qquad (3\text{–}1)$$

where $y$ is the output of the system and $x$ is the input. The transfer function of this system is obtained by taking the Laplace transforms of both sides of Equation (3–1), under the assumption that all initial conditions are zero, or

$$\text{Transfer function} = G(s) = \left.\frac{\mathscr{L}[\text{output}]}{\mathscr{L}[\text{input}]}\right|_{\text{zero initial conditions}}$$

$$= \frac{Y(s)}{X(s)} = \frac{b_0 s^m + b_1 s^{m-1} + \cdots + b_{m-1}s + b_m}{a_0 s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n} \tag{3–2}$$

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 in the list a system referred to 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.

**Mechanical system.** Consider the satellite attitude control system shown in Figure 3–2. The diagram shows the control of only the yaw angle $\theta$. (In the actual system there are controls about three axes.) Small jets apply reaction forces to rotate the satellite body into the desired attitude. The two skew symmetrically placed jets denoted by $A$ or $B$ operate in pairs. Assume that each jet thrust is $F/2$ and a torque $T = Fl$ is applied to the system. The jets are applied for a certain time duration and thus the torque can be written as $T(t)$. The moment of inertia about the axis of rotation at the center of mass is $J$.

Let us obtain the transfer function of this system by assuming that torque $T(t)$ is the input, and the angular displacement $\theta(t)$ of the satellite is the output. (We consider the motion only in the plane of the page.)

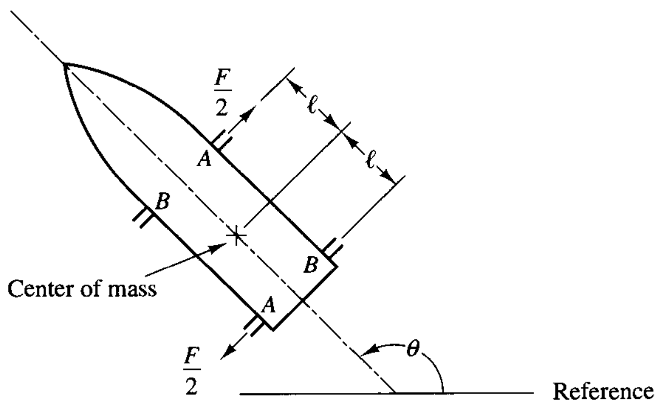



**Figure 3–2**
Schematic diagram of a satellite attitude control system.

To derive the transfer function, we proceed according to the following steps.

1. Write the differential equation for the system.
2. Take the Laplace transform of the differential equation, assuming all initial conditions are zero.
3. Take the ratio of the output $\Theta(s)$ to the input $T(s)$. This ratio is the transfer function.

Applying Newton's second law to the present system and noting that there is no friction in the environment of the satellite, we have

$$J\frac{d^2\theta}{dt^2} = T$$

Taking the Laplace transform of both sides of this last equation and assuming all initial conditions to be zero yields

$$Js^2\Theta(s) = T(s)$$

where $\Theta(s) = \mathscr{L}[\theta(t)]$ and $T(s) = \mathscr{L}[T(t)]$. The transfer function of the system is thus obtained as

$$\text{Transfer function} = \frac{\Theta(s)}{T(s)} = \frac{1}{Js^2}$$

**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 and $Y(s)$ is the Laplace transform of the output, 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

$$Y(s) = G(s)X(s) \tag{3–3}$$

Note that multiplication in the complex domain is equivalent to convolution in the time domain, so the inverse Laplace transform of Equation (3–3) is given by the following convolution integral:

$$y(t) = \int_0^t x(\tau)g(t - \tau)\, d\tau$$
$$= \int_0^t g(\tau)x(t - \tau)\, d\tau \tag{3–4}$$

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

**Impulse-response function.** Consider the output (response) of a 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

$$Y(s) = G(s) \tag{3–5}$$

The inverse Laplace transform of the output given by Equation (3–5) gives the impulse response of the system. The inverse Laplace transform of $G(s)$, or

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

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 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.)

## 3–3 BLOCK DIAGRAMS

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 explains what a block diagram is, 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

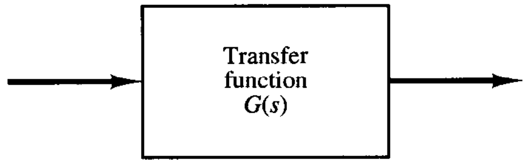



**Figure 3–3**
Element of a block diagram.

in the direction of the arrows. Thus a block diagram of a control system explicitly shows a unilateral property.

Figure 3–3 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*.

Note that the dimensions of the output signal from the block are the dimensions of the input signal multiplied by the dimensions of the transfer function in the block.

The advantages of the block diagram representation of a system lie in the fact 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.

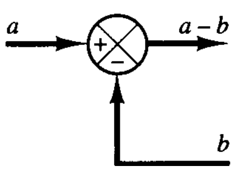



**Figure 3–4**
Summing point.

***Summing Point.*** Referring to Figure 3–4, 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 3–5 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

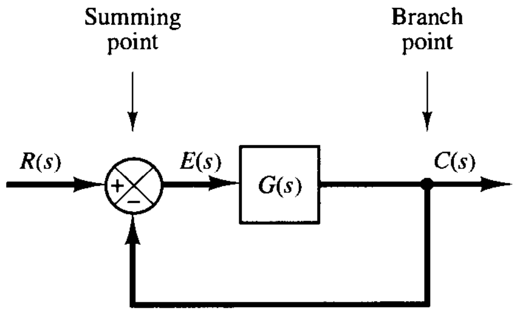



**Figure 3–5**
Block diagram of a closed-loop system.

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 3–6. 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 the output of the plant. The output of the sensor is compared with the 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 3–6, 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 3–6, the output $C(s)$ and input $R(s)$ are related as follows:

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

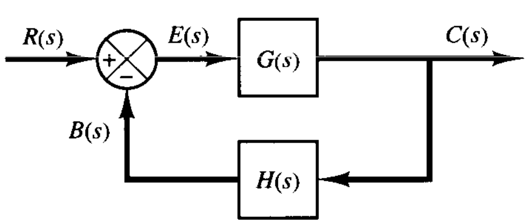



**Figure 3–6**
Closed-loop system.

Eliminating $E(s)$ from these equations gives

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

or

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)H(s)} \tag{3–6}$$

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

From Equation (3–6), $C(s)$ 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.

**Closed-loop system subjected to a disturbance.** Figure 3–7 shows a closed-loop system subjected to a disturbance. When two inputs (the reference input and disturbance) are present in a linear 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 3–7. In examining the effect of the disturbance $D(s)$, we may assume that the system is at rest initially with zero error; 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

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

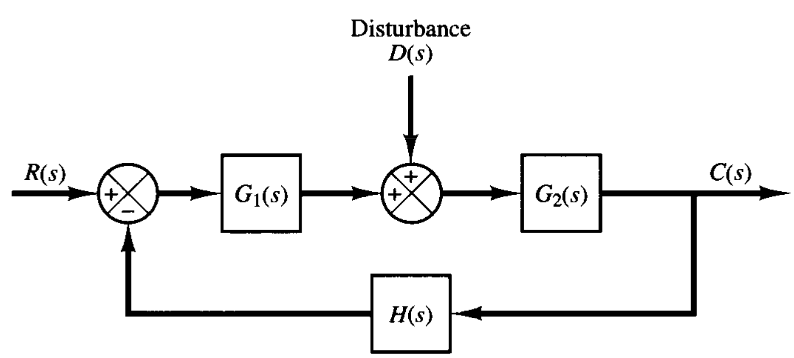



**Figure 3–7**
Closed-loop system subjected to a disturbance.

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

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

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 becomes 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 3–8(a). The equations for this circuit are

$$
i = \frac{e_i - e_o}{R} \tag{3–7}
$$

$$
e_o = \frac{\int i\,dt}{C} \tag{3–8}
$$

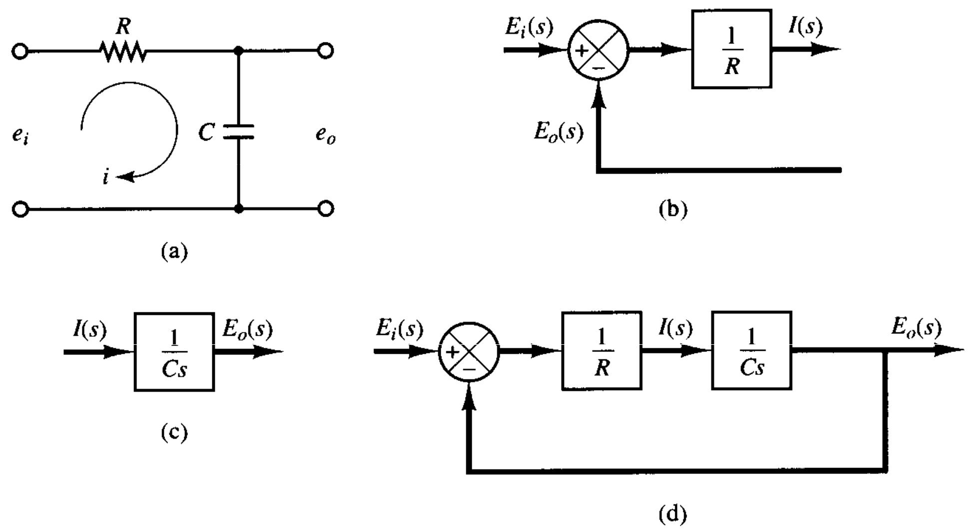



**Figure 3–8**
(a) $RC$ circuit; (b) block diagram representing Equation (3–9); (c) block diagram representing Equation (3–10); (d) block diagram of the $RC$ circuit.

The Laplace transforms of Equations (3–7) and (3–8), with zero initial condition, become

$$I(s) = \frac{E_i(s) - E_o(s)}{R} \tag{3–9}$$

$$E_o(s) = \frac{I(s)}{Cs} \tag{3–10}$$

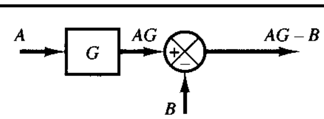
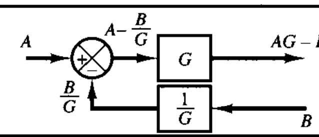
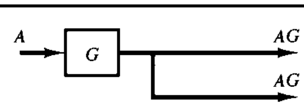
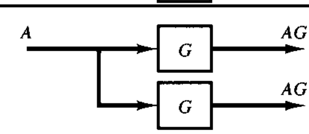
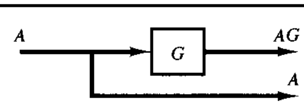
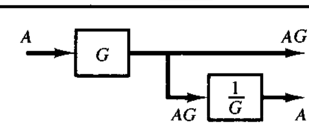
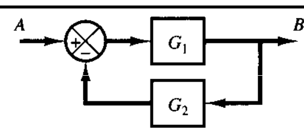
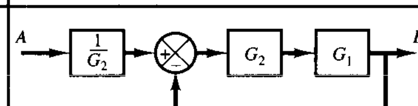
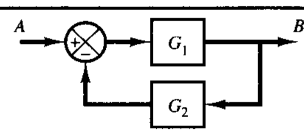
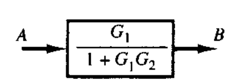
Equation (3–9) represents a summing operation, and the corresponding diagram is shown in Figure 3–8(b). Equation (3–10) represents the block as shown in Figure 3–8(c). Assembling these two elements, we obtain the overall block diagram for the system as shown in Figure 3–8(d).

**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.

A complicated block diagram involving many feedback loops can be simplified by a step-by-step rearrangement, using rules of block diagram algebra. Some of these important rules are given in Table 3–1. They are obtained by writing the same equation in

**Table 3–1** Rules of Block Diagram Algebra

| | Original Block Diagrams | Equivalent Block Diagrams |
|---|---|---|
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |

a different way. Simplification of the block diagram by rearrangements and substitutions 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.

In simplifying a block diagram, remember the following.

1. The product of the transfer functions in the feedforward direction must remain the same.
2. The product of the transfer functions around the loop must remain the same.

**EXAMPLE 3-1** Consider the system shown in Figure 3-9(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 3-9(b). Eliminating the positive feedback

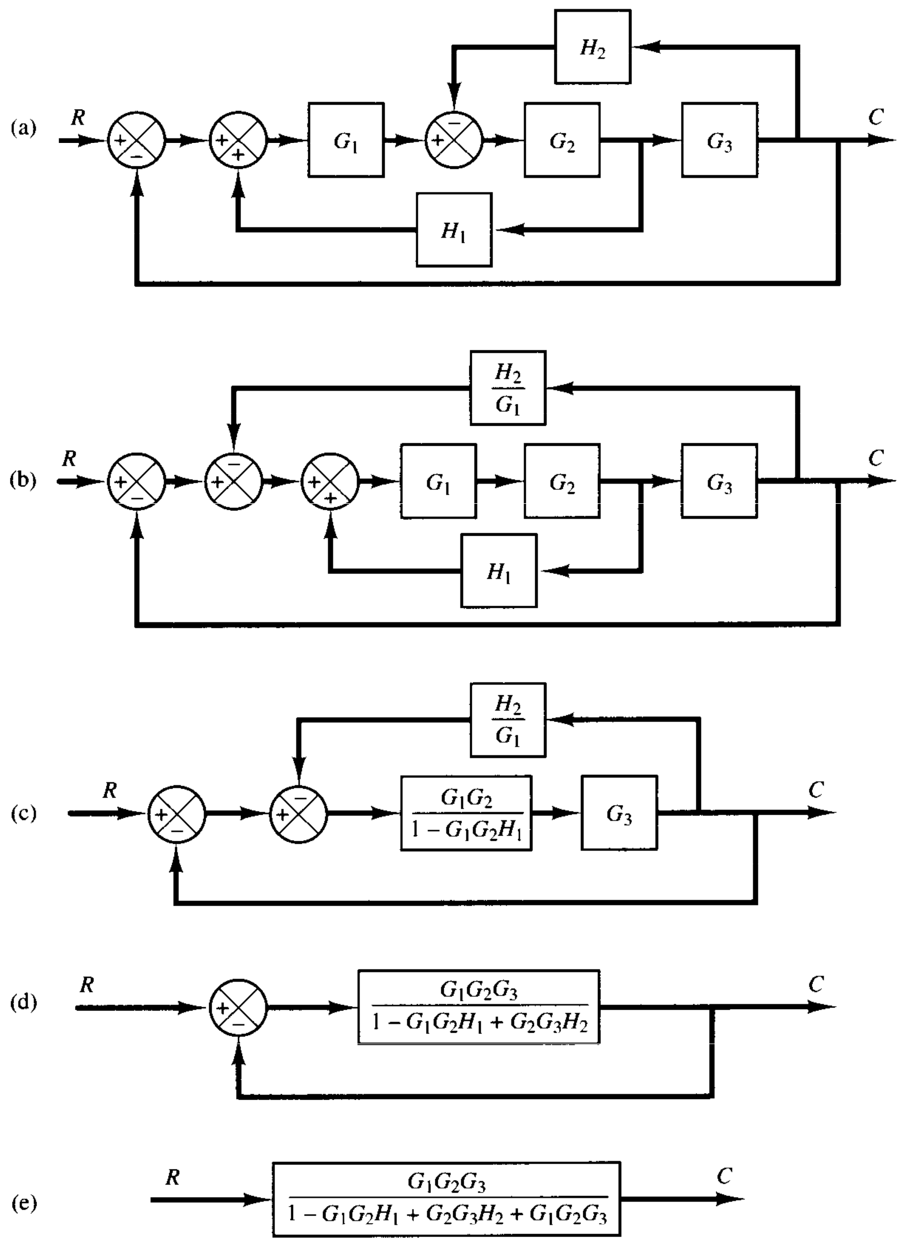



**Figure 3-9**
(a) Multiple-loop system; (b)–(e) successive reductions of the block diagram shown in (a).

loop, we have Figure 3–9(c). The elimination of the loop containing $H_2/G_1$ gives Figure 3–9(d). Finally, eliminating the feedback loop results in Figure 3–9(e).

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

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

$$= 1 - (G_1G_2H_1 - G_2G_3H_2 - G_1G_2G_3)$$
$$= 1 - G_1G_2H_1 + G_2G_3H_2 + G_1G_2G_3$$

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

## 3–4 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 a time-domain approach, 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 the knowledge of these variables at $t = t_0$, together with the 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 $\mathbf{u}(t)$ 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, . . . , $x_n$ axis 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 3–5, 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 $u_1, u_2(t), \ldots, u_r(t)$ 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

$$
\begin{aligned}
\dot{x}_1(t) &= f_1(x_1, x_2, \ldots, x_n; u_1, u_2, \ldots, u_r; t) \\
\dot{x}_2(t) &= f_2(x_1, x_2, \ldots, x_n; u_1, u_2, \ldots, u_r; t) \\
&\;\;\vdots \\
\dot{x}_n(t) &= f_n(x_1, x_2, \ldots, x_n; u_1, u_2, \ldots, u_r; t)
\end{aligned}
\qquad (3\text{–}11)
$$

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

$$
\begin{aligned}
y_1(t) &= g_1(x_1, x_2, \ldots, x_n; u_1, u_2, \ldots, u_r; t) \\
y_2(t) &= g_2(x_1, x_2, \ldots, x_n; u_1, u_2, \ldots, u_r; t) \\
&\;\;\vdots \\
y_m(t) &= g_m(x_1, x_2, \ldots, x_n; u_1, u_2, \ldots, u_r; t)
\end{aligned}
\tag{3–12}
$$

If we define

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

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

then Equations (3–11) and (3–12) become

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}, \mathbf{u}, t) \tag{3–13}$$

$$\mathbf{y}(t) = \mathbf{g}(\mathbf{x}, \mathbf{u}, t) \tag{3–14}$$

where Equation (3–13) is the state equation and Equation (3–14) is the output equation. If vector functions $\mathbf{f}$ and/or $\mathbf{g}$ involve time $t$ explicitly, then the system is called a time-varying system.

If Equations (3–13) and (3–14) are linearized about the operating state, then we have the following linearized state equation and output equation:

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \tag{3–15}$$

$$\mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{D}(t)\mathbf{u}(t) \tag{3–16}$$

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 the operating state are discussed in Section 3–10.) A block diagram representation of Equations (3–15) and (3–16) is shown in Figure 3–10.

If vector functions $\mathbf{f}$ and $\mathbf{g}$ do not involve time $t$ explicitly then the system is called a time-invariant system. In this case, Equations (3–15) and (3–16) can be simplified to

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \tag{3–17}$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) \tag{3–18}$$

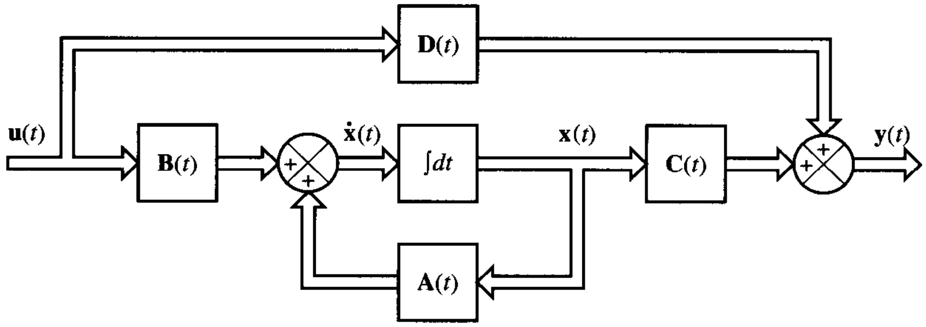



**Figure 3–10**
Block diagram of the linear continuous-time control system represented in state space.

Equation (3–17) is the state equation of the linear, time-invariant system. Equation (3–18) is the output equation for the same system. In this book we shall be concerned mostly with systems described by Equations (3–17) and (3–18).

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

## EXAMPLE 3–2

Consider the mechanical system shown in Figure 3–11. 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

$$m\ddot{y} + b\dot{y} + ky = u \tag{3–19}$$

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

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

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

or

$$\dot{x}_1 = x_2 \tag{3–20}$$

$$\dot{x}_2 = -\frac{k}{m}x_1 - \frac{b}{m}x_2 + \frac{1}{m}u \tag{3–21}$$

The output equation is

$$y = x_1 \tag{3–22}$$

In a vector-matrix form, Equations (3–20) and (3–21) can be written as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix} u \tag{3–23}$$

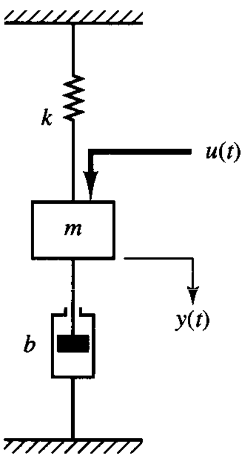



**Figure 3–11**
Mechanical system.

The output equation, Equation (3–22), can be written as

$$y = [1 \quad 0] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{3–24}$$

Equation (3–23) is a state equation and Equation (3–24) is an output equation for the system. Equations (3–23) and (3–24) are in the standard form:

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

$$y = \mathbf{C}\mathbf{x} + Du$$

where

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

Figure 3–12 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

$$\frac{Y(s)}{U(s)} = G(s) \tag{3–25}$$

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

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u \tag{3–26}$$

$$y = \mathbf{C}\mathbf{x} + Du \tag{3–27}$$

where $\mathbf{x}$ is the state vector, $u$ is the input, and $y$ is the output. The Laplace transforms of Equations (3–26) and (3–27) are given by

$$s\mathbf{X}(s) - \mathbf{x}(0) = \mathbf{A}\mathbf{X}(s) + \mathbf{B}U(s) \tag{3–28}$$

$$Y(s) = \mathbf{C}\mathbf{X}(s) + DU(s) \tag{3–29}$$

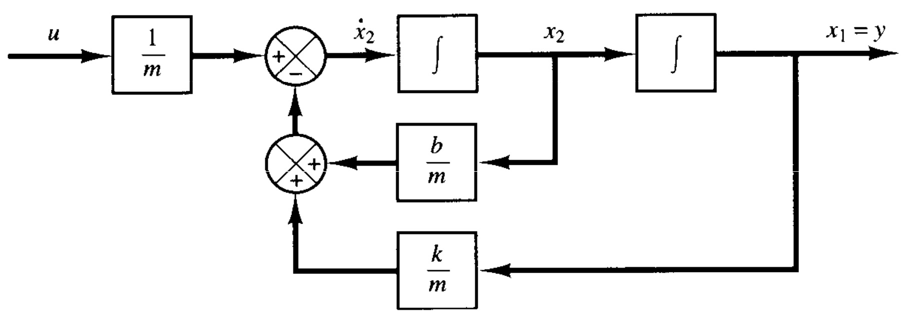



**Figure 3–12** Block diagram of the mechanical system shown in Figure 3–11.

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 assume that $\mathbf{x}(0)$ in Equation (3–28) is 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

$$\mathbf{X}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}U(s) \tag{3–30}$$

By substituting Equation (3–30) into Equation (3–29), we get

$$Y(s) = [\mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + D]U(s) \tag{3–31}$$

Upon comparing Equation (3–31) with Equation (3–25), we see that

$$G(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + D \tag{3–32}$$

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

Note that the right-hand side of equation (3–32) 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 $s$. Therefore, $|s\mathbf{I} - \mathbf{A}|$ 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 3–3** Consider again the mechanical system shown in Figure 3–11. State-space equations for the system are given by Equations (3–23) and (3–24). 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 (3–32), we obtain

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

Since

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

we have

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

$$= \frac{1}{ms^2 + bs + k}$$

which is the transfer function of the system. The same transfer function can be obtained from Equation (3–19).

**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

$$\mathbf{Y}(s) = \mathbf{G}(s)\mathbf{U}(s) \tag{3–33}$$

Since the input vector $\mathbf{u}$ is $r$ dimensional and the output vector $\mathbf{y}$ is $m$ dimensional, the transfer matrix is an $m \times r$ matrix.

## 3–5 STATE-SPACE REPRESENTATION OF DYNAMIC 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 $n$th-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 $n$th-order system:

$$\overset{(n)}{y} + a_1 \overset{(n-1)}{y} + \cdots + a_{n-1}\dot{y} + a_n y = u \tag{3–34}$$

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

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

Then Equation (3–34) can be written as

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

or

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u \tag{3–35}$$

where

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

The output can be given by

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

or

$$y = \mathbf{C}\mathbf{x} \tag{3–36}$$

where

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

[Note that $D$ in Equation (3–27) is zero.] The first-order differential equation, Equation (3–35), is the state equation, and the algebraic equation, Equation (3–36), is the output equation. A block diagram realization of the state equation and output equation given by Equations (3–35) and (3–36), respectively, is shown in Figure 3–13.

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 (3–35) and (3–36).

**State-space representation of *n*th-order systems of linear differential equations in which the forcing function involves derivative terms.** If the differential equation of the system involves derivatives of the forcing function, such as

$$\overset{(n)}{y} + a_1 \overset{(n-1)}{y} + \cdots + a_{n-1}\dot{y} + a_n y = b_0 \overset{(n)}{u} + b_1 \overset{(n-1)}{u} + \ldots + b_{n-1}\dot{u} + b_n u \qquad (3\text{–}37)$$

then the set of $n$ variables $y, \dot{y}, \ddot{y}, \ldots, \overset{(n-1)}{y}$ does not qualify as a set of state variables, and the straightforward method previously employed cannot be used. This is because $n$ first-order differential equations

$$\begin{aligned}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= x_3 \\
&\vdots \\
\dot{x}_n &= -a_n x_1 - a_{n-1}x_2 - \cdots - a_1 x_n + b_0 \overset{(n)}{u} + b_1 \overset{(n-1)}{u} + \cdots + b_n u
\end{aligned}$$

where $x_1 = y$, may not yield a unique solution.

The main problem in defining the state variables for this case lies in the derivative terms on the right-hand side of the last of the preceding $n$ equations. 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 is to define the following $n$ variables as a set of $n$ state variables:

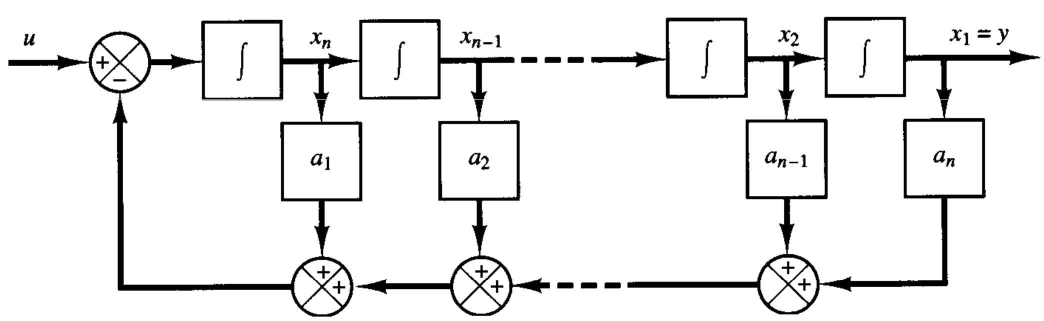



**Figure 3–13** Block diagram realization of state equation and output equation given by Equations (3–35) and (3–36), respectively.

$$
\begin{aligned}
x_1 &= y - \beta_0 u \\
x_2 &= \dot{y} - \beta_0 \dot{u} - \beta_1 u = \dot{x}_1 - \beta_1 u \\
x_3 &= \ddot{y} - \beta_0 \ddot{u} - \beta_1 \dot{u} - \beta_2 u = \dot{x}_2 - \beta_2 u \\
&\ \ \vdots \\
x_n &= \overset{(n-1)}{y} - \beta_0 \overset{(n-1)}{u} - \beta_1 \overset{(n-2)}{u} - \cdots - \beta_{n-2}\dot{u} - \beta_{n-1} u = \dot{x}_{n-1} - \beta_{n-1} u
\end{aligned}
\tag{3-38}
$$

where $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ are determined from

$$
\begin{aligned}
\beta_0 &= b_0 \\
\beta_1 &= b_1 - a_1\beta_0 \\
\beta_2 &= b_2 - a_1\beta_1 - a_2\beta_0 \\
\beta_3 &= b_3 - a_1\beta_2 - a_2\beta_1 - a_3\beta_0 \\
&\ \ \vdots \\
\beta_n &= b_n - a_1\beta_{n-1} - \cdots - a_{n-1}\beta_1 - a_n\beta_0
\end{aligned}
\tag{3-39}
$$

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

$$
\begin{aligned}
\dot{x}_1 &= x_2 + \beta_1 u \\
\dot{x}_2 &= x_3 + \beta_2 u \\
&\ \ \vdots \\
\dot{x}_{n-1} &= x_n + \beta_{n-1} u \\
\dot{x}_n &= -a_n x_1 - a_{n-1} x_2 - \cdots - a_1 x_n + \beta_n u
\end{aligned}
\tag{3-40}
$$

[To derive Equation (3–40), see Problem A-3-3.] In terms of vector-matrix equations, Equation (3–40) and the output equation can be written as

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

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

or

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u \tag{3–41}$$

$$y = \mathbf{C}\mathbf{x} + Du \tag{3–42}$$

where

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

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

The initial condition $\mathbf{x}(0)$ may be determined by use of Equation (3–38).

In this state-space representation, matrices $\mathbf{A}$ and $\mathbf{C}$ are exactly the same as those for the system of Equation (3–34). The derivatives on the right-hand side of Equation (3–37) affect only the elements of the $\mathbf{B}$ matrix.

Note that the state-space representation for the transfer function

$$\frac{Y(s)}{U(s)} = \frac{b_0 s^n + b_1 s^{n-1} + \cdots + b_{n-1}s + b_n}{s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n} \tag{3–43}$$

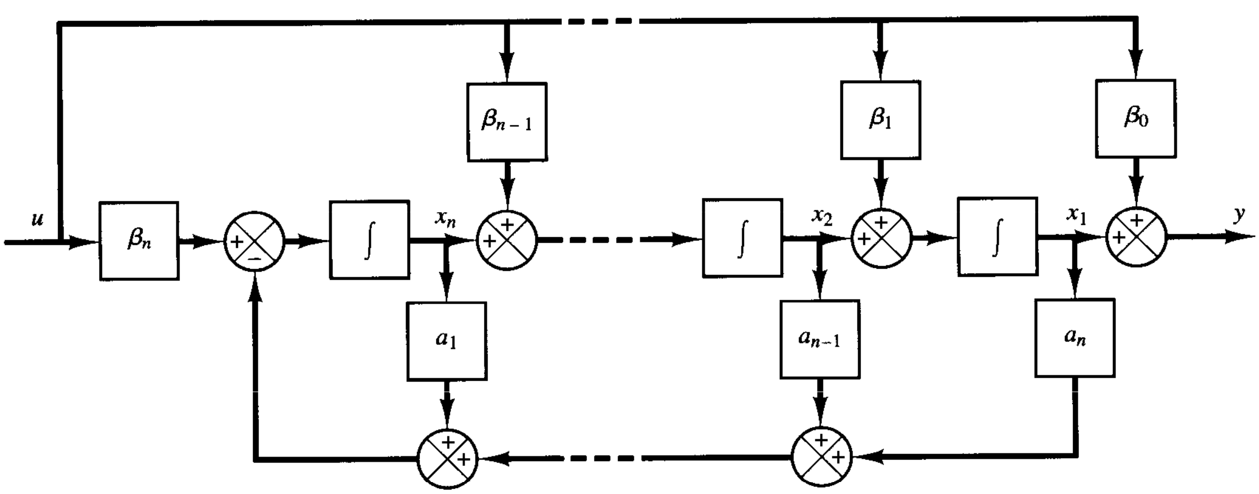



**Figure 3–14**
Block diagram realization of state equation and output equation given by Equations (3–41) and (3–42), respectively.

is given also by Equations (3–41) and (3–42). Figure 3–14 is a block diagram realization of the state equation and output equation given by Equations (3–41) and (3–42), respectively.

There are many ways to obtain state-space representations of systems. Some of them are presented in Problems A-3-4 through A-3-7. 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 11.

## 3–6 MECHANICAL SYSTEMS

In this section we shall discuss mathematical modeling of mechanical systems. The fundamental law governing mechanical systems is Newton's second law. It can be applied to any mechanical system. In this section we shall derive mathematical models of two mechanical systems. (Mathematical models of additional mechanical systems will be derived and analyzed throughout the remaining chapters.) Before we discuss mechanical systems, let us review definitions of mass, force, and unit systems.

**Mass.** The *mass* of a body is the quantity of matter in it, which is assumed to be constant. Physically, mass is the property of a body that gives it inertia, that is, resistance to starting and stopping. A body is attracted by Earth, and the magnitude of the force that Earth exerts on it is called its *weight.*

In practical situations, we know the weight $w$ of a body but not the mass $m$. We calculate mass $m$ from

$$m = \frac{w}{g}$$

where $g$ is the gravitational acceleration constant. The value of $g$ varies slightly from point to point on Earth's surface. As a result, the weight of a body varies slightly at different points on Earth's surface, but its mass remains constant. For engineering purposes, $g$ is taken as

$$g = 9.81 \text{ m/s}^2 = 981 \text{ cm/s}^2 = 32.2 \text{ ft/s}^2 = 386 \text{ in./s}^2$$

Far out in space, a body becomes weightless. Yet its mass remains constant and so the body possesses inertia.

The units of mass are kg, g, lb, $\text{kg}_f\text{-s}^2/\text{m}$, and slug, as shown in Table 3–2. If mass is expressed in units of kilogram (or pound), we call it kilogram mass (or pound mass) to distinguish it from the unit of force, which is termed kilogram force (or pound force). In this book kg is used to denote a kilogram mass and $\text{kg}_f$ a kilogram force. Similarly, lb denotes a pound mass and $\text{lb}_f$ a pound force.

A slug is a unit of mass such that, when acted on by 1-pound force, a 1-slug mass accelerates at 1 $\text{ft/s}^2$ (slug = $\text{lb}_f\text{-s}^2/\text{ft}$). In other words, if a mass of 1 slug is acted on by 32.2 pounds force, it accelerates at 32.2 $\text{ft/s}^2$ (= $g$). Hence the mass of a body weighing 32.2 $\text{lb}_f$ at the earth's surface is 1 slug or

$$m = \frac{w}{g} = \frac{32.2 \text{ lb}_f}{32.2 \text{ ft/s}^2} = 1 \text{ slug}$$

**Table 3–2** Systems of Units

| Quantity \ Systems of units | Absolute systems | | | Gravitational systems | |
|---|---|---|---|---|---|
| | Metric | | | Metric engineering | British engineering |
| | SI | mks | cgs | | |
| Length | m | m | cm | m | ft |
| Mass | kg | kg | g | $\frac{kg_f\text{-}s^2}{m}$ | slug $=\frac{lb_f\text{-}s^2}{ft}$ |
| Time | s | s | s | s | s |
| Force | N $=\frac{kg\text{-}m}{s^2}$ | N $=\frac{kg\text{-}m}{s^2}$ | dyn $=\frac{g\text{-}cm}{s^2}$ | $kg_f$ | $lb_f$ |
| Energy | J = N-m | J = N-m | erg = dyn-cm | $kg_f$-m | ft-$lb_f$ or Btu |
| Power | W $=\frac{N\text{-}m}{s}$ | W $=\frac{N\text{-}m}{s}$ | $\frac{dyn\text{-}cm}{s}$ | $\frac{kg_f\text{-}m}{s}$ | $\frac{ft\text{-}lb_f}{s}$ or hp |

**Force.** *Force* can be defined as the cause that tends to produce a change in motion of a body on which it acts. To move a body, force must be applied to it. Two types of forces are capable of acting on a body: *contact* forces and *field* forces. Contact forces are those that come into direct contact with a body, whereas field forces, such as gravitational force and magnetic force, act on a body but do not come into contact with it.

The units of force are newton (N), dyne (dyn), $kg_f$, and $lb_f$. In SI units and the mks system (a metric absolute system) of units the force unit is the newton. The newton is the force that will give a 1-kilogram mass an acceleration of 1 m/s$^2$ or

$$1 \text{ N} = 1 \text{ kg-m/s}^2$$

This means that 9.81 newtons will give a kilogram mass an acceleration of 9.81 m/s$^2$. Since the gravitational acceleration is $g$ = 9.81 m/s$^2$ (as stated earlier, for engineering calculations, the value of $g$ may be taken as 9.81 m/s$^2$ or 32.2 ft/s$^2$), a mass of 1 kilogram will produce a force on its support of 9.81 newtons.

The force unit in the cgs system (a metric absolute system) is the dyne, which will give a gram mass an acceleration of 1 cm/s$^2$ or

$$1 \text{ dyn} = 1 \text{ g-cm/s}^2$$

The force unit in the metric engineering (gravitational) system is $kg_f$, which is a primary dimension in the system. Similarly, in the British engineering system the force unit is $lb_f$. It is also a primary dimension in this system of units.

**Comments.** SI units for force, mass, and length are the newton (N), kilogram mass (kg), and meter (m). The mks units for force, mass, and length are the same as the SI units. Similarly, the cgs units for force, mass, and length are the dyne (dyn), gram (g), and centimeter (cm), and those for BES units are pound force ($lb_f$), slug, and foot (ft). Each of the unit systems is consistent in that the unit of force accelerates the unit of mass 1 unit of length per second per second.

In the systems of units shown in Table 3–2, "s" is used for the second. In engineering papers and books, however, "sec" is commonly used. Therefore, in this book we use "sec", rather than "s", for the second.

**Mechanical system.** Consider the spring–mass–dashpot system mounted on a massless cart as shown in Figure 3–15. A dashpot is a device that provides viscous friction, or damping. It consists of a piston and oil-filled cylinder. Any relative motion between the piston rod and the cylinder is resisted by the oil because the oil must flow around the piston (or through orifices provided in the piston) from one side of the piston to the other. The dashpot essentially absorbs energy. This absorbed energy is dissipated as heat, and the dashpot does not store any kinetic or potential energy. The dashpot is also called a *damper*.

Let us obtain a mathematical model of this spring–mass–dashpot system mounted on a cart by assuming that the cart is standing still for $t < 0$. In this system, $u(t)$ is the displacement of the cart and is the input to the system. At $t = 0$, the cart is moved at a constant speed, or $\dot{u} =$ constant. The displacement $y(t)$ of the mass is the output. (The displacement is relative to the ground.) In this system, $m$ denotes the mass, $b$ denotes the viscous friction coefficient, and $k$ denotes the spring constant. We assume that the friction force of the dashpot is proportional to $\dot{y} - \dot{u}$ and that the spring is a linear spring; that is, the spring force is proportional to $y - u$.

For translational systems, Newton's second law states that

$$ma = \sum F$$

where $m$ is a mass, $a$ is the acceleration of the mass, and $\Sigma F$ is the sum of the forces acting on the mass. Applying Newton's second law to the present system and noting that the cart is massless, we obtain

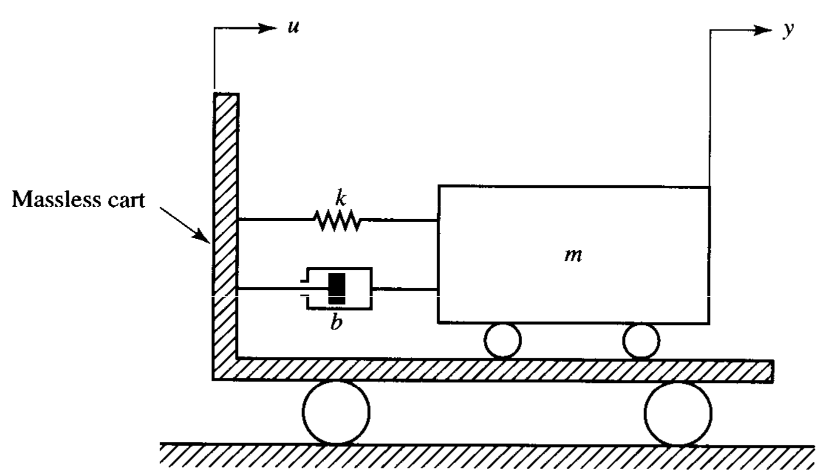



**Figure 3–15** Spring–mass–dashpot system mounted on a cart.

$$m \frac{d^2y}{dt^2} = -b\left(\frac{dy}{dt} - \frac{du}{dt}\right) - k(y - u)$$

or

$$m \frac{d^2y}{dt^2} + b \frac{dy}{dt} + ky = b \frac{du}{dt} + ku \tag{3-44}$$

Equation (3–44) gives a mathematical model of the system considered.

A transfer function model is another way of representing a mathematical model of a linear, time-invariant system. For the present mechanical system, the transfer function model can be obtained as follows: Taking the Laplace transform of each term of Equation (3–44) gives

$$\mathscr{L}\left[m \frac{d^2y}{dt^2}\right] = m[s^2Y(s) - sy(0) - \dot{y}(0)]$$

$$\mathscr{L}\left[b \frac{dy}{dt}\right] = b[sY(s) - y(0)]$$

$$\mathscr{L}[ky] = kY(s)$$

$$\mathscr{L}\left[b \frac{du}{dt}\right] = b[sU(s) - u(0)]$$

$$\mathscr{L}[ku] = kU(s)$$

If we set the initial conditions equal to zero, or set $y(0) = 0$, $\dot{y}(0) = 0$, and $u(0) = 0$, the Laplace transform of Equation (3–44) can be written as

$$(ms^2 + bs + k)Y(s) = (bs + k)U(s)$$

Taking the ratio of $Y(s)$ to $U(s)$, we find the transfer function of the system to be

$$\text{Transfer function} = G(s) = \frac{Y(s)}{U(s)} = \frac{bs + k}{ms^2 + bs + k}$$

Such a transfer-function representation of a mathematical model is used very frequently in control engineering. It should be noted, however, that transfer-function models apply only to linear, time-invariant systems, since the transfer functions are defined only for such systems.

Next we shall obtain a state-space model of this system. We shall first compare the differential equation for this system

$$\ddot{y} + \frac{b}{m} \dot{y} + \frac{k}{m} y = \frac{b}{m} \dot{u} + \frac{k}{m} u$$

with the standard form

$$\ddot{y} + a_1\dot{y} + a_2y = b_0\ddot{u} + b_1\dot{u} + b_2u$$

and identify $a_1, a_2, b_0, b_1$, and $b_2$ as follows:

$$a_1 = \frac{b}{m}, \qquad a_2 = \frac{k}{m}, \qquad b_0 = 0, \qquad b_1 = \frac{b}{m}, \qquad b_2 = \frac{k}{m}$$

Referring to Equation (3–39), we have

$$\beta_0 = b_0 = 0$$

$$\beta_1 = b_1 - a_1\beta_0 = \frac{b}{m}$$

$$\beta_2 = b_2 - a_1\beta_1 - a_2\beta_0 = \frac{k}{m} - \left(\frac{b}{m}\right)^2$$

Then, referring to Equation (3–38), define

$$x_1 = y - \beta_0 u = y$$

$$x_2 = \dot{x}_1 - \beta_1 u = \dot{x}_1 - \frac{b}{m}u$$

From Equation (3–40) we have

$$\dot{x}_1 = x_2 + \beta_1 u = x_2 + \frac{b}{m}u$$

$$\dot{x}_2 = -a_2x_1 - a_1x_2 + \beta_2 u = -\frac{k}{m}x_1 - \frac{b}{m}x_2 + \left[\frac{k}{m} - \left(\frac{b}{m}\right)^2\right]u$$

and the output equation becomes

$$y = x_1$$

or

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \frac{b}{m} \\ \frac{k}{m} - \left(\frac{b}{m}\right)^2 \end{bmatrix} u \tag{3–45}$$

and

$$y = [1 \quad 0]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{3–46}$$

Equations (3–45) and (3–46) give a state-space representation of the system. (Note that this is not the only state-space representation. There are infinitely many state-space representations for the system.)

**EXAMPLE 3–4** An inverted pendulum mounted on a motor-driven cart is shown in Figure 3–16(a). This is a model of the attitude control of a space booster on takeoff. (The objective of the attitude control problem is to keep the space booster in a vertical position.) The inverted pendulum is unstable in that it may fall over any time in any direction unless a suitable control force is applied. Here we consider only a two-dimensional problem in which the pendulum moves only in the plane of the page. The control force $u$ is applied to the cart. Assume that the center of gravity of the pendulum rod is at its geometric center. Obtain a mathematical model for the system. Assume that the mass $m$ of the pendulum rod is 0.1 kg, the mass $M$ of the cart is 2 kg, and the length $2l$ of the pendulum rod is 1 m, or

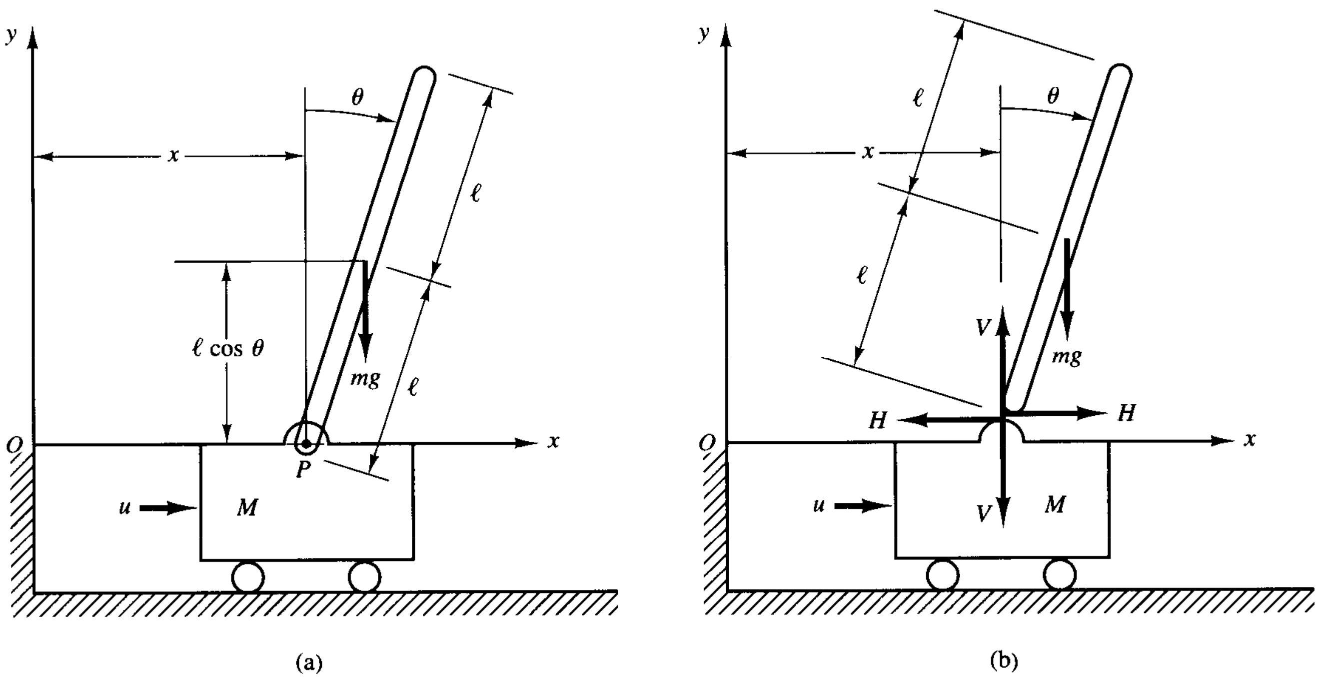



**Figure 3–16**
(a) Inverted pendulum system; (b) free-body diagram.

$$m = 0.1 \text{ kg}, \qquad M = 2 \text{ kg}, \qquad 2l = 1 \text{ m}$$

Define the angle of the rod from the vertical line as $\theta$. Define also the $(x, y)$ coordinates of the center of gravity of the pendulum rod as $(x_G, y_G)$. Then

$$x_G = x + l \sin \theta$$

$$y_G = l \cos \theta$$

To derive the equations of motion for the system, consider the free-body diagram shown in Figure 3–16(b). The rotational motion of the pendulum rod about its center of gravity can be described by

$$I\ddot{\theta} = Vl \sin \theta - Hl \cos \theta \tag{3–47}$$

where $I$ is the moment of inertia of the rod about its center of gravity.

The horizontal motion of center of gravity of pendulum rod is given by

$$m \frac{d^2}{dt^2}(x + l \sin \theta) = H \tag{3–48}$$

The vertical motion of center of gravity of pendulum rod is

$$m \frac{d^2}{dt^2}(l \cos \theta) = V - mg \tag{3–49}$$

The horizontal motion of cart is described by

$$M \frac{d^2 x}{dt^2} = u - H \tag{3–50}$$

Equations (3–47) through (3–50) describe the motion of the inverted-pendulum-on-the-cart system. Because these equations involve $\sin\theta$ and $\cos\theta$, they are nonlinear equations.

If we assume angle $\theta$ to be small, Equations (3–47) through (3–50) may be linearized as follows:

$$I\ddot{\theta} = Vl\theta - Hl \tag{3–51}$$

$$m(\ddot{x} + l\ddot{\theta}) = H \tag{3–52}$$

$$0 = V - mg \tag{3–53}$$

$$M\ddot{x} = u - H \tag{3–54}$$

From Equations (3–52) and (3–54), we obtain

$$(M + m)\ddot{x} + ml\ddot{\theta} = u \tag{3–55}$$

From Equations (3–51) and (3–53), we have

$$\begin{aligned} I\ddot{\theta} &= mgl\theta - Hl \\ &= mgl\theta - l(m\ddot{x} + ml\ddot{\theta}) \end{aligned}$$

or

$$(I + ml^2)\ddot{\theta} + ml\ddot{x} = mgl\theta \tag{3–56}$$

Equations (3–55) and (3–56) describe the motion of the inverted-pendulum-on-the-cart system. They constitute a mathematical model of the system. (Later in Chapters 12 and 13, we design controllers to keep the pendulum upright in the presence of disturbances.)

## 3–7 ELECTRICAL SYSTEMS

In this section we shall deal with electrical circuits involving resistors, capacitors, and inductors.

Basic laws governing electrical circuits are Kirchhoff's current law and voltage law. Kirchhoff's current law (node law) states that the algebraic sum of all currents entering and leaving a node is zero. (This law can also be stated as follows: The sum of currents entering a node is equal to the sum of currents leaving the same node.) Kirchhoff's voltage law (loop law) states that at any given instant the algebraic sum of the voltages around any loop in an electrical circuit is zero. (This law can also be stated as follows: The sum of the voltage drops is equal to the sum of the voltage rises around a loop.) A mathematical model of an electrical circuit can be obtained by applying one or both of Kirchhoff's laws to it.

This section deals with simple electrical circuits. Mathematical modeling of operational amplifier systems is presented in Chapter 5.

***LRC* circuit.** Consider the electrical circuit shown in Figure 3–17. The circuit consists of an inductance $L$ (henry), a resistance $R$ (ohm), and a capacitance $C$ (farad). Applying Kirchhoff's voltage law to the system, we obtain the following equations:

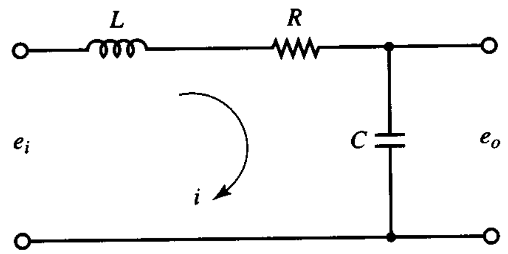



**Figure 3–17**
Electrical circuit.

$$L\frac{di}{dt} + Ri + \frac{1}{C}\int i\,dt = e_i \tag{3–57}$$

$$\frac{1}{C}\int i\,dt = e_o \tag{3–58}$$

Equations (3–57) and (3–58) give a mathematical model of the circuit.

A transfer function model of the circuit can also be obtained as follows: Taking the Laplace transforms of Equations (3–57) and (3–58), assuming zero initial conditions, we obtain

$$LsI(s) + RI(s) + \frac{1}{C}\frac{1}{s}I(s) = E_i(s)$$

$$\frac{1}{C}\frac{1}{s}I(s) = E_o(s)$$

If $e_i$ is assumed to be the input and $e_o$ the output, then the transfer function of this system is found to be

$$\frac{E_o(s)}{E_i(s)} = \frac{1}{LCs^2 + RCs + 1} \tag{3–59}$$

**Complex impedances.** In driving transfer functions for electrical circuits, we frequently find it convenient to write the Laplace-transformed equations directly, without writing the differential equations. Consider the system shown in Figure 3–18(a). In this system, $Z_1$ and $Z_2$ represent complex impedances. The complex impedance $Z(s)$ of a two-terminal circuit is the ratio of $E(s)$, the Laplace transform of the voltage across the terminals, to $I(s)$, the Laplace transform of the current through the element, under the assumption that the initial conditions are zero, so that $Z(s) = E(s)/I(s)$. If the two-terminal elements is a resistance $R$, capacitance $C$, or inductance $L$, then the complex impedance is given by $R$, $1/Cs$, or $Ls$, respectively. If complex impedances are connected in series, the total impedance is the sum of the individual complex impedances.

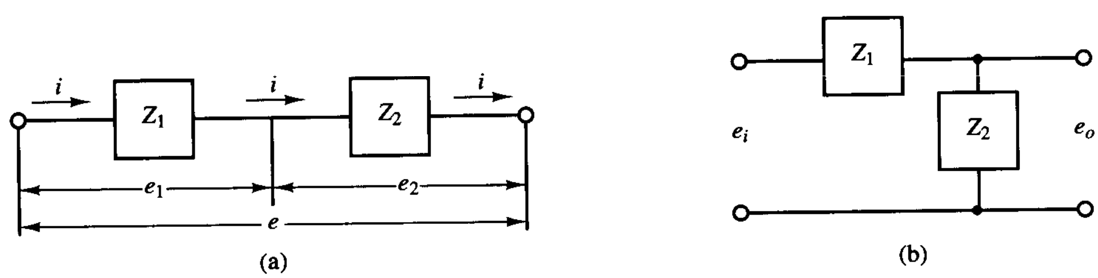



**Figure 3–18**
Electrical circuits.

Remember that the impedance approach is valid only if the initial conditions involved are all zeros. Since the transfer function requires zero initial conditions, the impedance approach can be applied to obtain the transfer function of the electrical circuit. This approach greatly simplifies the derivation of transfer functions of electrical circuits.

Consider the circuit shown in Figure 3–18(b). Assume that the voltages $e_i$ and $e_o$ are the input and output of the circuit, respectively. Then the transfer function of this circuit is

$$\frac{E_o(s)}{E_i(s)} = \frac{Z_2(s)}{Z_1(s) + Z_2(s)}$$

For the system shown in Figure 3–17,

$$Z_1 = Ls + R, \qquad Z_2 = \frac{1}{Cs}$$

Hence the transfer function $E_o(s)/E_i(s)$ can be found as follows:

$$\frac{E_o(s)}{E_i(s)} = \frac{\dfrac{1}{Cs}}{Ls + R + \dfrac{1}{Cs}} = \frac{1}{LCs^2 + RCs + 1}$$

which is, of course, identical to Equation (3–59).

**State-space representation.** A state-space model of the system shown in Figure 3–17 may be obtained as follows: First, note that the differential equation for the system can be obtained from Equation (3–59) as

$$\ddot{e}_o + \frac{R}{L}\dot{e}_o + \frac{1}{LC}e_o = \frac{1}{LC}e_i$$

Then by defining state variables by

$$x_1 = e_o$$
$$x_2 = \dot{e}_o$$

and the input and output variables by

$$u = e_i$$
$$y = e_o = x_1$$

we obtain

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\dfrac{1}{LC} & -\dfrac{R}{L} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \dfrac{1}{LC} \end{bmatrix} u$$

and

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

These two equations give a mathematical model of the system in state space.

**Transfer functions of cascaded elements.** Many feedback systems have components that load each other. Consider the system shown in Figure 3–19. Assume that $e_i$ is the input and $e_o$ is the output. In this system the second stage of the circuit ($R_2C_2$ portion) produces a loading effect on the first stage ($R_1C_1$ portion). The equations for this system are

$$\frac{1}{C_1}\int (i_1 - i_2)\,dt + R_1 i_1 = e_i \tag{3–60}$$

and

$$\frac{1}{C_1}\int (i_2 - i_1)\,dt + R_2 i_2 + \frac{1}{C_2}\int i_2\,dt = 0 \tag{3–61}$$

$$\frac{1}{C_2}\int i_2\,dt = e_o \tag{3–62}$$

Taking the Laplace transforms of Equations (3–60) through (3–62), respectively, assuming zero initial conditions, we obtain

$$\frac{1}{C_1 s}[I_1(s) - I_2(s)] + R_1 I_1(s) = E_i(s) \tag{3–63}$$

$$\frac{1}{C_1 s}[I_2(s) - I_1(s)] + R_2 I_2(s) + \frac{1}{C_2 s} I_2(s) = 0 \tag{3–64}$$

$$\frac{1}{C_2 s} I_2(s) = E_o(s) \tag{3–65}$$

Eliminating $I_1(s)$ from Equations (3–63) and (3–64) and writing $E_i(s)$ in terms of $I_2(s)$, we find the transfer function between $E_o(s)$ and $E_i(s)$ to be

$$\begin{aligned}\frac{E_o(s)}{E_i(s)} &= \frac{1}{(R_1C_1s + 1)(R_2C_2s + 1) + R_1C_2s} \\ &= \frac{1}{R_1C_1R_2C_2s^2 + (R_1C_1 + R_2C_2 + R_1C_2)s + 1}\end{aligned} \tag{3–66}$$

The term $R_1C_2s$ in the denominator of the transfer function represents the interaction of two simple $RC$ circuits. Since $(R_1C_1 + R_2C_2 + R_1C_2)^2 > 4R_1C_1R_2C_2$, the two roots of the denominator of Equation (3–66) are real.

The present analysis shows that, if two $RC$ circuits are connected in cascade so that the output from the first circuit is the input to the second, the overall transfer function

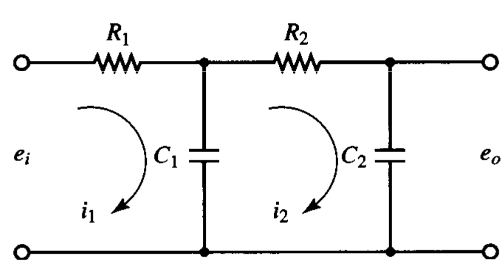



**Figure 3–19**
Electrical system.

is not the product of $1/(R_1C_1s + 1)$ and $1/(R_2C_2s + 1)$. The reason for this is that, when we derive the transfer function for an isolated circuit, we implicitly assume that the output is unloaded. In other words, the load impedance is assumed to be infinite, which means that no power is being withdrawn at the output. When the second circuit is connected to the output of the first, however, a certain amount of power is withdrawn, and thus the assumption of no loading is violated. Therefore, if the transfer function of this system is obtained under the assumption of no loading, then it is not valid. The degree of the loading effect determines the amount of modification of the transfer function.

**Transfer functions of nonloading cascaded elements.** The transfer function of a system consisting of two nonloading cascaded elements can be obtained by eliminating the intermediate input and output. For example, consider the system shown in Figure 3–20(a). The transfer functions of the elements are

$$G_1(s) = \frac{X_2(s)}{X_1(s)} \quad \text{and} \quad G_2(s) = \frac{X_3(s)}{X_2(s)}$$

If the input impedance of the second element is infinite, the output of the first element is not affected by connecting it to the second element. Then the transfer function of the whole system becomes

$$G(s) = \frac{X_3(s)}{X_1(s)} = \frac{X_2(s)X_3(s)}{X_1(s)X_2(s)} = G_1(s)G_2(s)$$

The transfer function of the whole system is thus the product of the transfer functions of the individual elements. This is shown in Figure 3–20(b).

As an example, consider the system shown in Figure 3–21. The insertion of an isolating amplifier between the circuits to obtain nonloading characteristics is frequently used in combining circuits. Since amplifiers have very high input impedances, an isolation amplifier inserted between the two circuits justifies the nonloading assumption.

The two simple $RC$ circuits, isolated by an amplifier as shown in Figure 3–21, have negligible loading effects, and the transfer function of the entire circuit equals the product of the individual transfer functions. Thus, in this case,

$$\begin{aligned}\frac{E_o(s)}{E_i(s)} &= \left(\frac{1}{R_1C_1s + 1}\right)(K)\left(\frac{1}{R_2C_2s + 1}\right) \\ &= \frac{K}{(R_1C_1s + 1)(R_2C_2s + 1)}\end{aligned}$$


**Figure 3–20**
(a) System consisting of two nonloading cascaded elements; (b) an equivalent system.

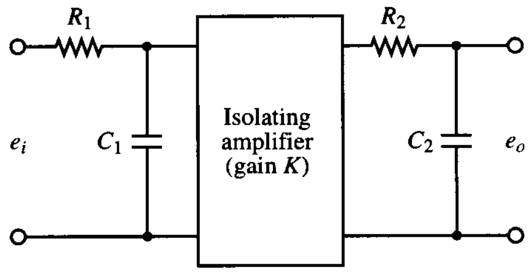



**Figure 3–21**
Electrical system.

## 3–8 LIQUID-LEVEL SYSTEMS

In analyzing systems involving fluid flow, we find it necessary to divide flow regimes into laminar flow and turbulent flow, according to the magnitude of the Reynolds number. If the Reynolds number is greater than about 3000 to 4000, then the flow is turbulent. The flow is laminar if the Reynolds number is less than about 2000. In the laminar case, fluid flow occurs in streamlines with no turbulence. Systems involving turbulent flow often have to be represented by nonlinear differential equations, while systems involving laminar flow may be represented by linear differential equations. (Industrial processes often involve flow of liquids through connecting pipes and tanks. The flow in such processes is often turbulent and not laminar.)

In this section we shall derive mathematical models of liquid-level systems. By introducing the concept of resistance and capacitance for such liquid-level systems, it is possible to describe the dynamic characteristics of such systems in simple forms.

**Resistance and capacitance of liquid-level systems.** Consider the flow through a short pipe connecting two tanks. The resistance $R$ for liquid flow in such a pipe or restriction is defined as the change in the level difference (the difference of the liquid levels of the two tanks) necessary to cause a unit change in flow rate; that is,

$$R = \frac{\text{change in level difference, m}}{\text{change in flow rate, m}^3\text{/sec}}$$

Since the relationship between the flow rate and level difference differs for the laminar flow and turbulent flow, we shall consider both cases in the following.

Consider the liquid-level system shown in Figure 3–22(a). In this system the liquid spouts through the load valve in the side of the tank. If the flow through this restriction is laminar, the relationship between the steady-state flow rate and steady-state head at the level of the restriction is given by

$$Q = KH$$

where $Q$ = steady-state liquid flow rate, $\text{m}^3$/sec
$K$ = coefficient, $\text{m}^2$/sec
$H$ = steady-state head, m

Notice that the law governing laminar flow is analogous to Coulomb's law, which states that the current is directly proportional to the potential difference.

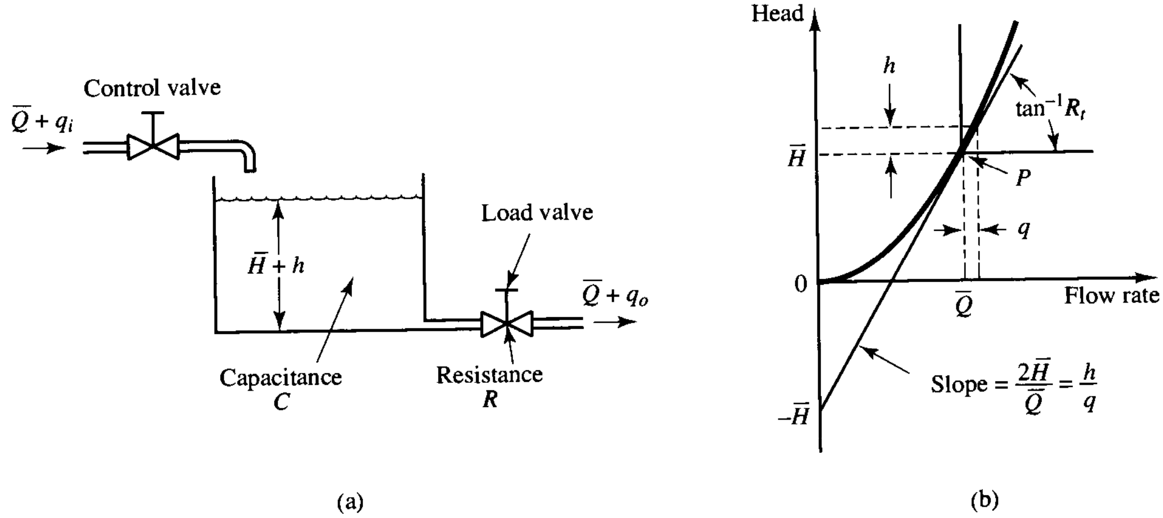



**Figure 3–22**
(a) Liquid-level system; (b) head versus flow rate curve.

For laminar flow, the resistance $R_l$ is obtained as

$$R_l = \frac{dH}{dQ} = \frac{H}{Q}$$

The laminar-flow resistance is constant and is analogous to the electrical resistance.

If the flow through the restriction is turbulent, the steady-state flow rate is given by

$$Q = K\sqrt{H} \tag{3–67}$$

where $Q$ = steady-state liquid flow rate, $m^3/sec$
$K$ = coefficient, $m^{2.5}/sec$
$H$ = steady-state head, m

The resistance $R_t$ for turbulent flow is obtained from

$$R_t = \frac{dH}{dQ}$$

Since from Equation (3–67) we obtain

$$dQ = \frac{K}{2\sqrt{H}} dH$$

we have

$$\frac{dH}{dQ} = \frac{2\sqrt{H}}{K} = \frac{2\sqrt{H}\sqrt{H}}{Q} = \frac{2H}{Q}$$

Thus,

$$R_t = \frac{2H}{Q} \tag{3–68}$$

The value of the turbulent-flow resistance $R_t$ depends on the flow rate and the head. The value of $R_t$, however, may be considered constant if the changes in head and flow rate are small.

By use of the turbulent-flow resistance, the relationship between $Q$ and $H$ can be given by

$$Q = \frac{2H}{R_t}$$

Such linearization is valid, provided that changes in the head and flow rate from their respective steady-state values are small.

In many practical cases, the value of the coefficient $K$ in Equation (3–67), which depends on the flow coefficient and the area of restriction, is not known. Then the resistance may be determined by plotting the head versus flow rate curve based on experimental data and measuring the slope of the curve at the operating condition. An example of such a plot is shown in Figure 3–22(b). In the figure, point $P$ is the steady-state operating point. The tangent line to the curve at point $P$ intersects the ordinate at point $(-\bar{H}, 0)$. Thus, the slope of this tangent line is $2\bar{H}/\bar{Q}$. Since the resistance $R_t$ at the operating point $P$ is given by $2\bar{H}/\bar{Q}$, the resistance $R_t$ is the slope of the curve at the operating point.

Consider the operating condition in the neighborhood of point $P$. Define a small deviation of the head from the steady-state value as $h$ and the corresponding small change of the flow rate as $q$. Then the slope of the curve at point $P$ can be given by

$$\text{Slope of curve at point } P = \frac{h}{q} = \frac{2\bar{H}}{\bar{Q}} = R_t$$

The linear approximation is based on the fact that the actual curve does not differ much from its tangent line if the operating condition does not vary too much.

The capacitance $C$ of a tank is defined to be the change in quantity of stored liquid necessary to cause a unit change in the potential (head). (The potential is the quantity that indicates the energy level of the system.)

$$C = \frac{\text{change in liquid stored, m}^3}{\text{change in head, m}}$$

It should be noted that the capacity ($m^3$) and the capacitance ($m^2$) are different. The capacitance of the tank is equal to its cross-sectional area. If this is constant, the capacitance is constant for any head.

**Liquid-level systems.** Consider the system shown in Figure 3–22(a). The variables are defined as follows:

$\bar{Q}$ = steady-state flow rate (before any change has occurred), $m^3$/sec

$q_i$ = small deviation of inflow rate from its steady-state value, $m^3$/sec

$q_o$ = small deviation of outflow rate from its steady-state value, $m^3$/sec

$\bar{H}$ = steady-state head (before any change has occurred), m

$h$ = small deviation of head from its steady-state value, m

As stated previously, a system can be considered linear if the flow is laminar. Even if the flow is turbulent, the system can be linearized if changes in the variables are kept small. Based on the assumption that the system is either linear or linearized, the differential equation of this system can be obtained as follows: Since the inflow minus outflow during the small time interval $dt$ is equal to the additional amount stored in the tank, we see that

$$C\,dh = (q_i - q_o)\,dt$$

From the definition of resistance, the relationship between $q_o$ and $h$ is given by

$$q_o = \frac{h}{R}$$

The differential equation for this system for a constant value of $R$ becomes

$$RC\frac{dh}{dt} + h = Rq_i \tag{3–69}$$

Note that $RC$ is the time constant of the system. Taking the Laplace transforms of both sides of Equation (3–69), assuming the zero initial condition, we obtain

$$(RCs + 1)H(s) = RQ_i(s)$$

where

$$H(s) = \mathscr{L}[h] \qquad \text{and} \qquad Q_i(s) = \mathscr{L}[q_i]$$

If $q_i$ is considered the input and $h$ the output, the transfer function of the system is

$$\frac{H(s)}{Q_i(s)} = \frac{R}{RCs + 1}$$

If, however, $q_o$ is taken as the output, the input being the same, then the transfer function is

$$\frac{Q_o(s)}{Q_i(s)} = \frac{1}{RCs + 1}$$

where we have used the relationship

$$Q_o(s) = \frac{1}{R}H(s)$$

**Liquid-level systems with interaction.** Consider the system shown in Figure 3–23. In this system, the two tanks interact. Thus the transfer function of the system is not the product of two first-order transfer functions.

In the following, we shall assume only small variations of the variables from the steady-state values. Using the symbols as defined in Figure 3–23, we can obtain the following equations for this system:

$$\frac{h_1 - h_2}{R_1} = q_1 \tag{3–70}$$

$$C_1\frac{dh_1}{dt} = q - q_1 \tag{3–71}$$

$$\frac{h_2}{R_2} = q_2 \tag{3–72}$$

$$C_2\frac{dh_2}{dt} = q_1 - q_2 \tag{3–73}$$

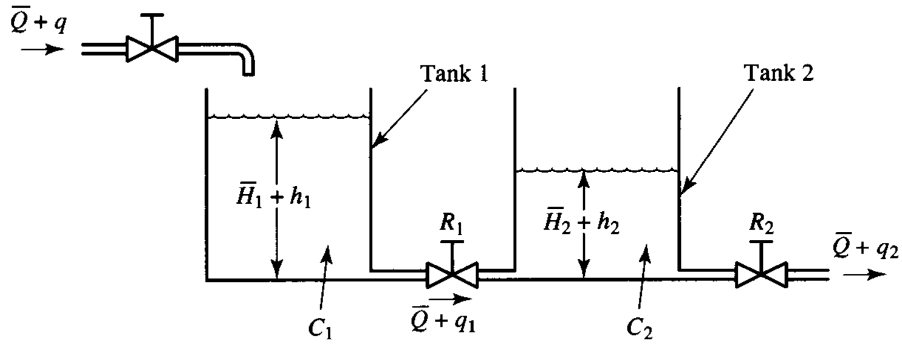



**Figure 3–23** Liquid-level system with interaction.

$\bar{Q}$ : Steady-state flow rate
$\bar{H}_1$ : Steady-state liquid level of tank 1
$\bar{H}_2$ : Steady-state liquid level of tank 2

If $q$ is considered the input and $q_2$ the output, the transfer function of the system is

$$\frac{Q_2(s)}{Q(s)} = \frac{1}{R_1C_1R_2C_2s^2 + (R_1C_1 + R_2C_2 + R_2C_1)s + 1} \tag{3–74}$$

It is instructive to obtain Equation (3–74), the transfer function of the interacted system, by block diagram reduction. From Equations (3–70) through (3–73), we obtain the elements of the block diagram, as shown in Figure 3–24(a). By connecting signals properly, we can construct a block diagram, as shown in Figure 3–24(b). This block diagram can be simplified, as shown in Figure 3–24(c). Further simplifications result in Figures 3–24(d) and (e). Figure 3–24(e) is equivalent to Equation (3–74).

Notice the similarity and difference between the transfer function given by Equation (3–74) and that given by Equation (3–66). The term $R_2C_1s$ that appears in the denominator of Equation (3–74) exemplifies the interaction between the two tanks. Similarly, the term $R_1C_2s$ in the denominator of Equation (3–66) represents the interaction between the two $RC$ circuits shown in Figure 3–19.

## 3–9 THERMAL SYSTEMS

Thermal systems are those that involve the transfer of heat from one substance to another. Thermal systems may be analyzed in terms of resistance and capacitance, although the thermal capacitance and thermal resistance may not be represented accurately as lumped parameters since they are usually distributed throughout the substance. For precise analysis, distributed-parameter models must be used. Here, however, to simplify the analysis we shall assume that a thermal system can be represented by a lumped-parameter model, that substances that are characterized by resistance to heat flow have negligible heat capacitance, and that substances that are characterized by heat capacitance have negligible resistance to heat flow.

There are three different ways heat can flow from one substance to another: conduction, convection, and radiation. Here we consider only conduction and convection. (Radiation heat transfer is appreciable only if the temperature of the emitter is very high compared to that of the receiver. Most thermal processes in process control systems do not involve radiation heat transfer.)

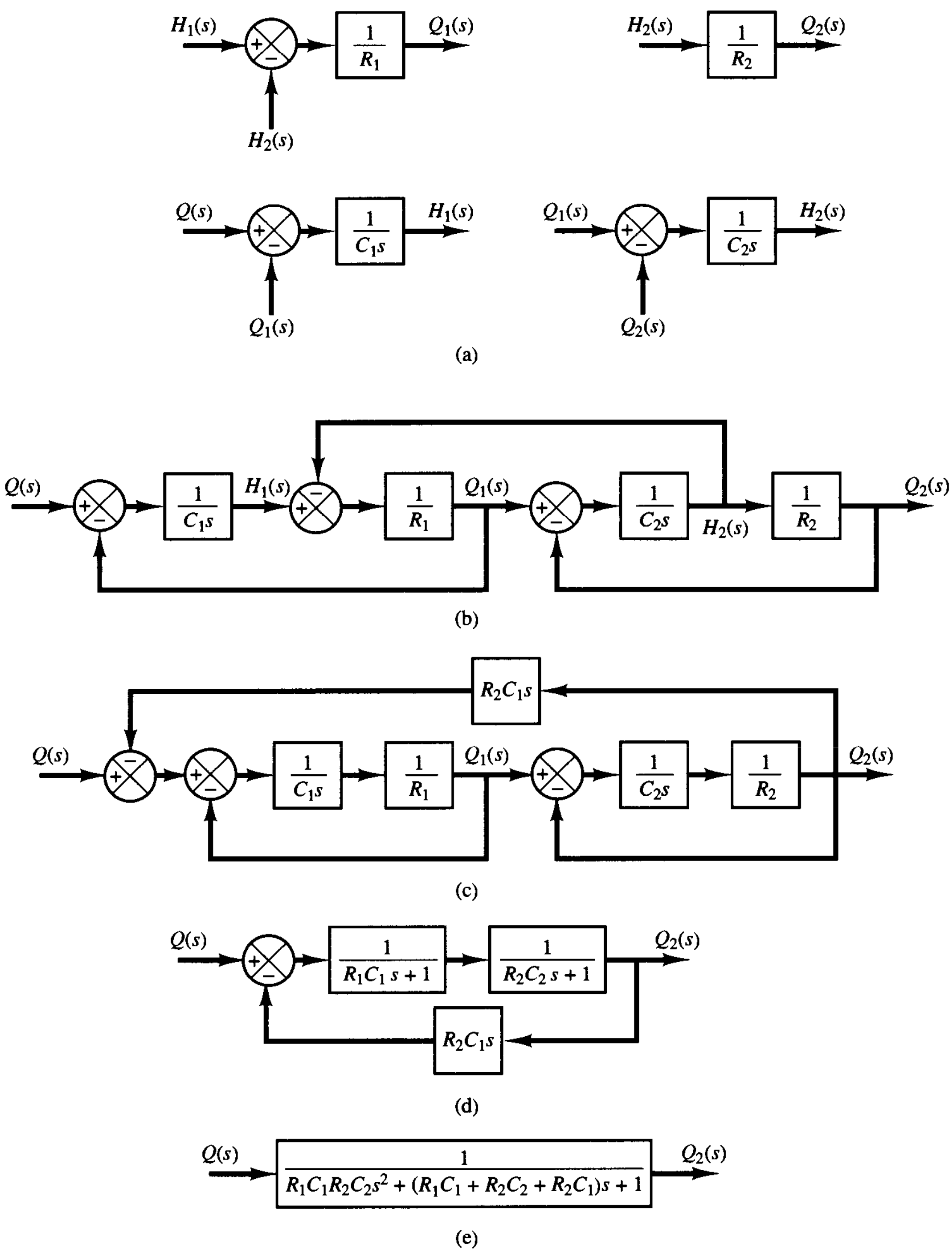



**Figure 3–24**
(a) Elements of the block diagram of the system shown in Figure 3–23; (b) block diagram of the system; (c)–(e) successive reductions of the block diagram.

For conduction or convection heat transfer,

$$q = K\,\Delta\theta$$

where $q$ = heat flow rate, kcal/sec
$\Delta\theta$ = temperature difference, °C
$K$ = coefficient, kcal/sec °C

The coefficient $K$ is given by

$$K = \frac{kA}{\Delta X}, \qquad \text{for conduction}$$
$$= HA, \qquad \text{for convection}$$

where $k$ = thermal conductivity, kcal/m sec °C
$A$ = area normal to heat flow, $m^2$
$\Delta X$ = thickness of conductor, m
$H$ = convection coefficient, kcal/$m^2$ sec °C

**Thermal resistance and thermal capacitance.** The thermal resistance $R$ for heat transfer between two substances may be defined as follows:

$$R = \frac{\text{change in temperature difference, °C}}{\text{change in heat flow rate, kcal/sec}}$$

The thermal resistance for conduction or convection heat transfer is given by

$$R = \frac{d(\Delta\theta)}{dq} = \frac{1}{K}$$

Since the thermal conductivity and convection coefficients are almost constant, the thermal resistance for either conduction or convection is constant.

The thermal capacitance $C$ is defined by

$$C = \frac{\text{change in heat stored, kcal}}{\text{change in temperature, °C}}$$

or

$$C = mc$$

where $m$ = mass of substance considered, kg
$c$ = specific heat of substance, kcal/kg °C

**Thermal systems.** Consider the system shown in Figure 3–25(a). It is assumed that the tank is insulated to eliminate heat loss to the surrounding air. It is also assumed that there is no heat storage in the insulation and that the liquid in the tank is perfectly mixed so that it is at a uniform temperature. Thus, a single temperature is used to describe the temperature of the liquid in the tank and of the outflowing liquid.

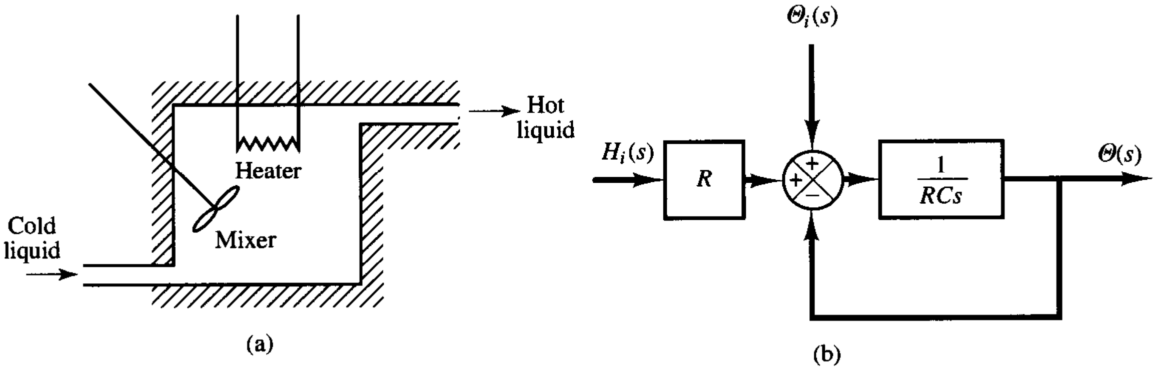



**Figure 3–25**
(a) Thermal system; (b) block diagram of the system.

Let us define

$\bar{\Theta}_i$ = steady-state temperature of inflowing liquid, °C

$\bar{\Theta}_o$ = steady-state temperature of outflowing liquid, °C

$G$ = steady-state liquid flow rate, kg/sec

$M$ = mass of liquid in tank, kg

$c$ = specific heat of liquid, kcal/kg °C

$R$ = thermal resistance, °C sec/kcal

$C$ = thermal capacitance, kcal/°C

$\bar{H}$ = steady-state heat input rate, kcal/sec

Assume that the temperature of the inflowing liquid is kept constant and that the heat input rate to the system (heat supplied by the heater) is suddenly changed from $\bar{H}$ to $\bar{H} + h_i$, where $h_i$ represents a small change in the heat input rate. The heat outflow rate will then change gradually from $\bar{H}$ to $\bar{H} + h_o$. The temperature of the outflowing liquid will also be changed from $\bar{\Theta}_o$ to $\bar{\Theta}_o + \theta$. For this case, $h_o$, $C$, and $R$ are obtained, respectively, as

$$h_o = Gc\theta$$

$$C = Mc$$

$$R = \frac{\theta}{h_o} = \frac{1}{Gc}$$

The differential equation for this system is

$$C\frac{d\theta}{dt} = h_i - h_o$$

which may be rewritten as

$$RC\frac{d\theta}{dt} + \theta = Rh_i$$

Note that the time constant of the system is equal to $RC$ or $M/G$ seconds. The transfer function relating $\theta$ and $h_i$ is given by

$$\frac{\Theta(s)}{H_i(s)} = \frac{R}{RCs + 1}$$

where $\Theta(s) = \mathscr{L}[\theta(t)]$ and $H_i(s) = \mathscr{L}[h_i(t)]$.

In practice, the temperature of the inflowing liquid may fluctuate and may act as a load disturbance. (If a constant outflow temperature is desired, an automatic controller may be installed to adjust the heat inflow rate to compensate for the fluctuations in the temperature of the inflowing liquid.) If the temperature of the inflowing liquid is suddenly changed from $\bar{\Theta}_i$ to $\bar{\Theta}_i + \theta_i$ while the heat input rate $H$ and the liquid flow rate $G$ are kept constant, then the heat outflow rate will be changed from $\bar{H}$ to $\bar{H} + h_o$, and the temperature of the outflowing liquid will be changed from $\bar{\Theta}_o$ to $\bar{\Theta}_o + \theta$. The differential equation for this case is

$$C\frac{d\theta}{dt} = Gc\theta_i - h_o$$

which may be rewritten

$$RC\frac{d\theta}{dt} + \theta = \theta_i$$

The transfer function relating $\theta$ and $\theta_i$ is given by

$$\frac{\Theta(s)}{\Theta_i(s)} = \frac{1}{RCs + 1}$$

where $\Theta(s) = \mathscr{L}[\theta(t)]$ and $\Theta_i(s) = \mathscr{L}[\theta_i(t)]$.

If the present thermal system is subjected to changes in both the temperature of the inflowing liquid and the heat input rate, while the liquid flow rate is kept constant, the change $\theta$ in the temperature of the outflowing liquid can be given by the following equation:

$$RC\frac{d\theta}{dt} + \theta = \theta_i + Rh_i$$

A block diagram corresponding to this case is shown in Figure 3–25(b). Notice that the system involves two inputs.

## 3–10 LINEARIZATION OF NONLINEAR MATHEMATICAL MODELS

In this section we present a linearization technique that is applicable to many nonlinear systems. The process of linearizing nonlinear systems is important, for by linearizing nonlinear equations, it is possible to apply numerous linear analysis methods that will produce information on the behavior of nonlinear systems. The linearization procedure presented here is based on the expansion of the 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 Taylor series expansion, these neglected terms must be small enough; that is, the variables deviate only slightly from the operating condition.

In what follows we shall first present mathematical aspects of the linearization technique and then apply the technique to a hydraulic servo system to obtain a linear model for the system.

**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

$$y = f(x) \tag{3–75}$$

If the normal operating condition corresponds to $\bar{x}, \bar{y}$, then Equation (3–75) may be expanded into a Taylor series about this point as follows:

$$\begin{aligned} y &= f(x) \\ &= f(\bar{x}) + \frac{df}{dx}(x - \bar{x}) + \frac{1}{2!}\frac{d^2f}{dx^2}(x - \bar{x})^2 + \cdots \end{aligned} \tag{3–76}$$

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 (3–76) may be written as

$$y = \bar{y} + K(x - \bar{x}) \tag{3–77}$$

where

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

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

Equation (3–77) may be rewritten as

$$y - \bar{y} = K(x - \bar{x}) \tag{3–78}$$

which indicates that $y - \bar{y}$ is proportional to $x - \bar{x}$. Equation (3–78) gives a linear mathematical model for the nonlinear system given by Equation (3–75) 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

$$y = f(x_1, x_2) \tag{3–79}$$

To obtain a linear approximation to this nonlinear system, we may expand Equation (3–79) into a Taylor series about the normal operating point $\bar{x}_1, \bar{x}_2$. Then Equation (3–79) becomes

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

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.

**Linearization of a hydraulic servo system.** Figure 3–26(a) shows a hydraulic servomotor. It is essentially a pilot-valve-controlled hydraulic power amplifier and actuator. The pilot valve is a balanced valve, in the sense that the pressure forces acting on it are all balanced. A very large power output can be controlled by a pilot valve, which can be positioned with very little power.

In practice, the ports shown in Figure 3–26(a) are often made wider than the corresponding valves. In such a case, there is always leakage through the valves. Such leakage improves both the sensitivity and the linearity of the hydraulic servomotor. In the following analysis we shall make the assumption that the ports are made wider than the valves, that is, the valves are underlapped. [Note that sometimes a dither signal, a high-frequency signal of very small amplitude (with respect to the maximum displacement of the valve), is superimposed on the motion of the pilot valve. This also improves the sensitivity and linearity. In this case also there is leakage through the valve.]

We shall apply the linearization technique just presented to obtain a linearized mathematical model of the hydraulic servomotor. We assume that the valve is underlapped and symmetrical and admits hydraulic fluid under high pressure into a power cylinder that contains a large piston, so that a large hydraulic force is established to

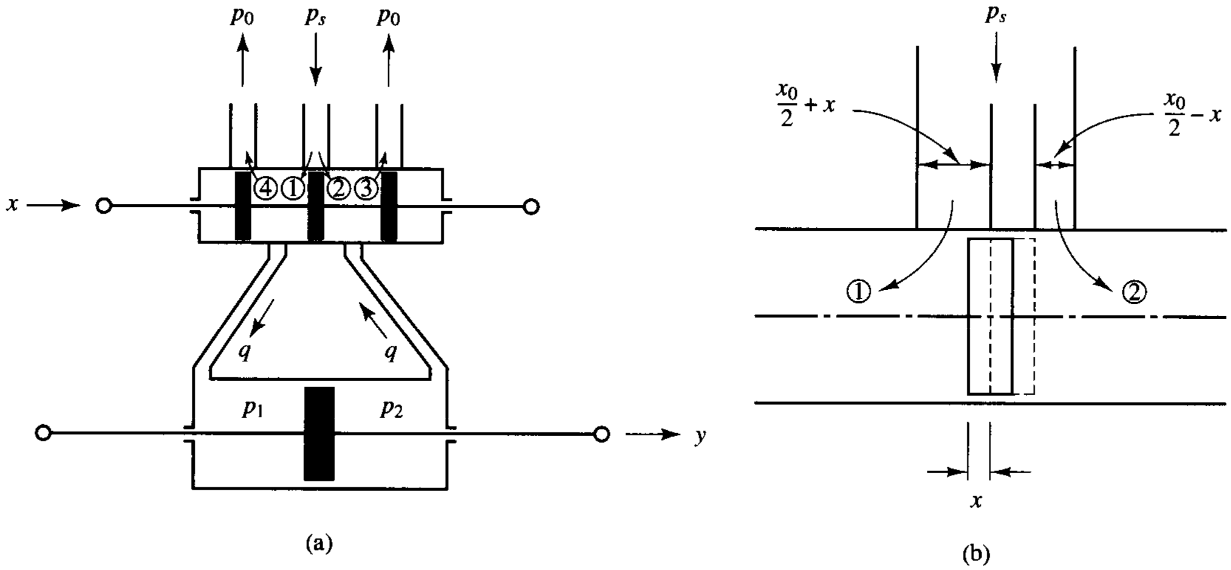



**Figure 3–26**
(a) Hydraulic servo system; (b) enlarged diagram of the valve orifice area.

move a load. We assume that the load inertia and friction are small compared to the large hydraulic force. In the present analysis, the hydraulic fluid is assumed to be incompressible and the inertia force of the power piston negligible. We also assume that, as is usually the case, the orifice area (the width of the slot in the valve sleeve) at each port is proportional to the valve displacement $x$.

In Figure 3–26(b) we have an enlarged diagram of the valve orifice area. Let us define the valve orifice areas of ports 1, 2, 3, 4 as $A_1, A_2, A_3, A_4$, respectively. Also, define the flow rates through ports 1, 2, 3, 4 as $q_1, q_2, q_3, q_4$, respectively. Note that, since the valve is symmetrical, $A_1 = A_3$ and $A_2 = A_4$. Assuming the displacement $x$ to be small, we obtain

$$A_1 = A_3 = k\left(\frac{x_0}{2} + x\right)$$

$$A_2 = A_4 = k\left(\frac{x_0}{2} - x\right)$$

where $k$ is a constant.

Furthermore, we shall assume that the return pressure $p_0$ in the return line is small and thus can be neglected. Then, referring to Figure 3–26(a), flow rates through valve orifices are

$$q_1 = c_1 A_1 \sqrt{\frac{2g}{\gamma}(p_s - p_1)} = C_1 \sqrt{p_s - p_1}\left(\frac{x_0}{2} + x\right)$$

$$q_2 = c_2 A_2 \sqrt{\frac{2g}{\gamma}(p_s - p_2)} = C_2 \sqrt{p_s - p_2}\left(\frac{x_0}{2} - x\right)$$

$$q_3 = c_1 A_3 \sqrt{\frac{2g}{\gamma}(p_2 - p_0)} = C_1 \sqrt{p_2 - p_0}\left(\frac{x_0}{2} + x\right) = C_1 \sqrt{p_2}\left(\frac{x_0}{2} + x\right)$$

$$q_4 = c_2 A_4 \sqrt{\frac{2g}{\gamma}(p_1 - p_0)} = C_2 \sqrt{p_1 - p_0}\left(\frac{x_0}{2} - x\right) = C_2 \sqrt{p_1}\left(\frac{x_0}{2} - x\right)$$

where $C_1 = c_1 k\sqrt{2g/\gamma}$ and $C_2 = c_2 k\sqrt{2g/\gamma}$, and $\gamma$ is the specific weight and is given by $\gamma = \varrho g$, where $\varrho$ is mass density and $g$ is the acceleration of gravity. The flow rate $q$ to the left-hand side of the power piston is

$$q = q_1 - q_4 = C_1\sqrt{p_s - p_1}\left(\frac{x_0}{2} + x\right) - C_2\sqrt{p_1}\left(\frac{x_0}{2} - x\right) \tag{3–80}$$

The flow rate from the right-hand side of the power piston to the drain is the same as this $q$ and is given by

$$q = q_3 - q_2 = C_1\sqrt{p_2}\left(\frac{x_0}{2} + x\right) - C_2\sqrt{p_s - p_2}\left(\frac{x_0}{2} - x\right)$$

Note that the fluid is incompressible and that the valve is symmetrical. So we have $q_1 = q_3$ and $q_2 = q_4$. By equating $q_1$ and $q_3$, we obtain

$$p_s - p_1 = p_2$$

or

$$p_s = p_1 + p_2$$

If we define the pressure difference across the power piston as $\Delta p$ or

$$\Delta p = p_1 - p_2$$

then

$$p_1 = \frac{p_s + \Delta p}{2}, \qquad p_2 = \frac{p_s - \Delta p}{2}$$

For the symmetrical valve shown in Figure 3–26(a), the pressure in each side of the power piston is $\frac{1}{2}p_s$ when no load is applied, or $\Delta p = 0$. As the spool valve is displaced, the pressure in one line increases as the pressure in the other line decreases by the same amount.

In terms of $p_s$ and $\Delta p$, we can rewrite the flow rate $q$ given by Equation (3–80) as

$$q = q_1 - q_4 = C_1\sqrt{\frac{p_s - \Delta p}{2}}\left(\frac{x_0}{2} + x\right) - C_2\sqrt{\frac{p_s + \Delta p}{2}}\left(\frac{x_0}{2} - x\right)$$

Noting that the supply pressure $p_s$ is constant, the flow rate $q$ can be written as a function of the valve displacement $x$ and pressure difference $\Delta p$, or

$$q = C_1\sqrt{\frac{p_s - \Delta p}{2}}\left(\frac{x_0}{2} + x\right) - C_2\sqrt{\frac{p_s + \Delta p}{2}}\left(\frac{x_0}{2} - x\right) = f(x, \Delta p)$$

By applying the linearization technique presented earlier in this section to this case, the linearized equation about point $x = \bar{x}$, $\Delta p = \Delta\bar{p}$, $q = \bar{q}$ is

$$q - \bar{q} = a(x - \bar{x}) + b(\Delta p - \Delta\bar{p}) \tag{3–81}$$

where

$$\bar{q} = f(\bar{x}, \Delta\bar{p})$$

$$a = \frac{\partial f}{\partial x}\bigg|_{x=\bar{x},\,\Delta p=\Delta\bar{p}} = C_1\sqrt{\frac{p_s - \Delta\bar{p}}{2}} + C_2\sqrt{\frac{p_s + \Delta\bar{p}}{2}}$$

$$b = \frac{\partial f}{\partial \Delta p}\bigg|_{x=\bar{x},\Delta p=\Delta\bar{p}} = -\left[\frac{C_1}{2\sqrt{2}\sqrt{p_s - \Delta\bar{p}}}\left(\frac{x_0}{2} + \bar{x}\right) + \frac{C_2}{2\sqrt{2}\sqrt{p_s + \Delta\bar{p}}}\left(\frac{x_0}{2} - \bar{x}\right)\right] < 0$$

Coefficients $a$ and $b$ here are called *valve coefficients*. Equation (3–81) is a linearized mathematical model of the spool valve near an operating point $x = \bar{x}$, $\Delta p = \Delta\bar{p}$, $q = \bar{q}$. The values of valve coefficients $a$ and $b$ vary with the operating point. Note that $\partial f/\partial \Delta p$ is negative and so $b$ is negative.

Since the normal operating point is the point where $\bar{x} = 0$, $\Delta\bar{p} = 0$, $\bar{q} = 0$, near the normal operating point, Equation (3–81) becomes

$$q = K_1 x - K_2 \Delta p \tag{3–82}$$

where

$$K_1 = (C_1 + C_2)\sqrt{\frac{p_s}{2}} > 0$$

$$K_2 = (C_1 + C_2)\frac{x_0}{4\sqrt{2}\sqrt{p_s}} > 0$$

Equation (3–82) is a linearized mathematical model of the spool valve near the origin ($\bar{x} = 0, \Delta\bar{p} = 0, \bar{q} = 0$). Note that the region near the origin is most important in this kind of system, because the system operation usually occurs near this point. (For a derivation of a mathematical model of a hydraulic servo system when the load reactive forces are not negligible, see Problem A-3-20.)

## EXAMPLE PROBLEMS AND SOLUTIONS

**A-3-1.** Simplify the block diagram shown in Figure 3–27.

**Solution.** First, move the branch point of the path involving $H_1$ outside the loop involving $H_2$, as shown in Figure 3–28(a). Then eliminating two loops results in Figure 3–28(b). Combining two blocks into one gives Figure 3–28(c).

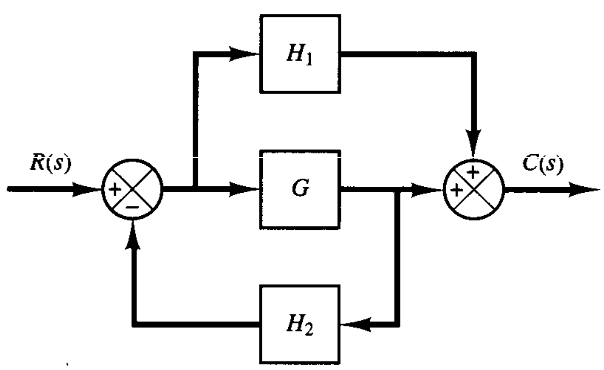



**Figure 3–27**
Block diagram of a system.

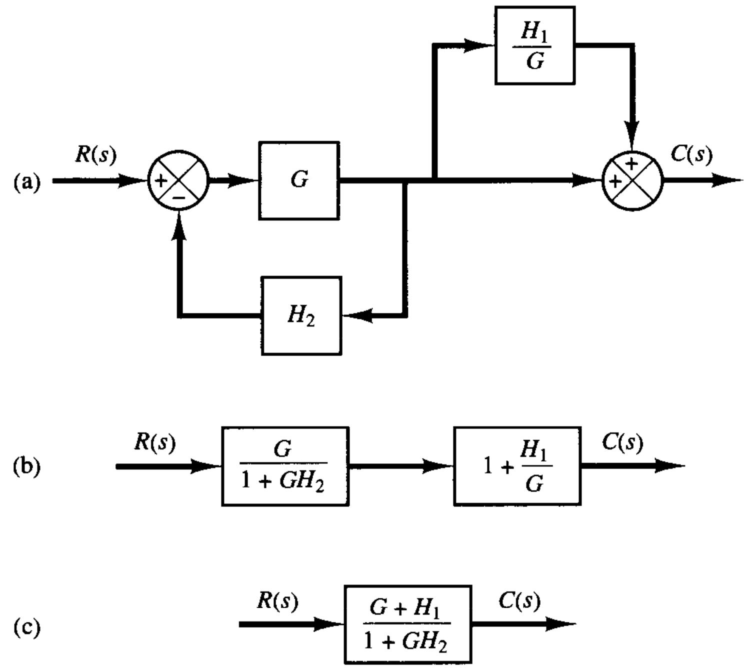



**Figure 3–28** Simplified block diagrams for the system shown in Figure 3–27.

**A-3-2.** Simplify the block diagram shown in Figure 3–29. Obtain the transfer function relating $C(s)$ and $R(s)$.

**Solution.** The block diagram of Figure 3–29 can be modified to that shown in Figure 3–30(a). Eliminating the minor feedforward path, we obtain Figure 3–30(b), which can be simplified to that shown in Figure 3–30(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

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

And so we have the same result as before:

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

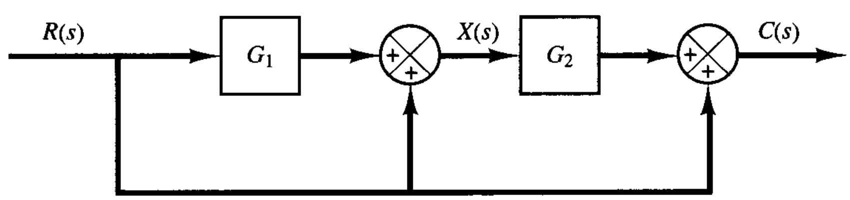



**Figure 3–29** Block diagram of a system.

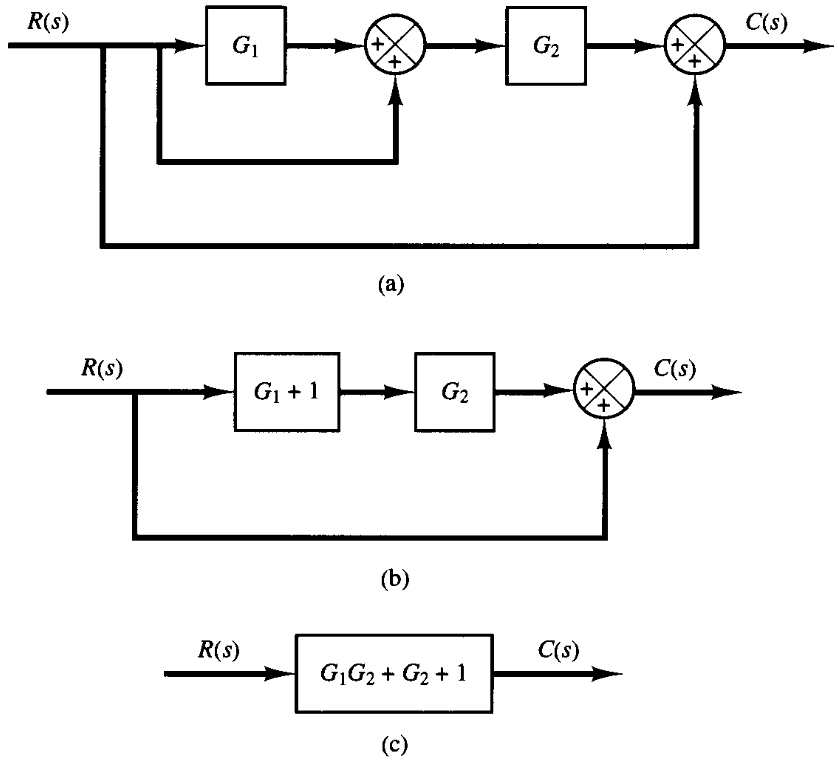



**Figure 3–30** Reduction of the block diagram shown in Figure 3–29.

**A-3-3.** Show that for the differential equation system

$$\dddot{y} + a_1\ddot{y} + a_2\dot{y} + a_3 y = b_0\dddot{u} + b_1\ddot{u} + b_2\dot{u} + b_3 u \tag{3–83}$$

state and output equations can be given, respectively, by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_3 & -a_2 & -a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} u \tag{3–84}$$

and

$$y = [1 \quad 0 \quad 0] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \beta_0 u \tag{3–85}$$

where state variables are defined by

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

and

$$\beta_0 = b_0$$
$$\beta_1 = b_1 - a_1\beta_0$$
$$\beta_2 = b_2 - a_1\beta_1 - a_2\beta_0$$
$$\beta_3 = b_3 - a_1\beta_2 - a_2\beta_1 - a_3\beta_0$$

**Solution.** From the definition of state variables $x_2$ and $x_3$, we have

$$\dot{x}_1 = x_2 + \beta_1 u \tag{3–86}$$

$$\dot{x}_2 = x_3 + \beta_2 u \tag{3–87}$$

To derive the equation for $\dot{x}_3$, we first note from Equation (3–83) that

$$\dddot{y} = -a_1\ddot{y} - a_2\dot{y} - a_3 y + b_0\dddot{u} + b_1\ddot{u} + b_2\dot{u} + b_3 u$$

Since

$$x_3 = \ddot{y} - \beta_0\ddot{u} - \beta_1\dot{u} - \beta_2 u$$

we have

$$\begin{aligned}
\dot{x}_3 &= \dddot{y} - \beta_0\dddot{u} - \beta_1\ddot{u} - \beta_2\dot{u} \\
&= (-a_1\ddot{y} - a_2\dot{y} - a_3 y) + b_0\dddot{u} + b_1\ddot{u} + b_2\dot{u} + b_3 u - \beta_0\dddot{u} - \beta_1\ddot{u} - \beta_2\dot{u} \\
&= -a_1(\ddot{y} - \beta_0\ddot{u} - \beta_1\dot{u} - \beta_2 u) - a_1\beta_0\ddot{u} - a_1\beta_1\dot{u} - a_1\beta_2 u \\
&\quad -a_2(\dot{y} - \beta_0\dot{u} - \beta_1 u) - a_2\beta_0\dot{u} - a_2\beta_1 u - a_3(y - \beta_0 u) - a_3\beta_0 u \\
&\quad + b_0\dddot{u} + b_1\ddot{u} + b_2\dot{u} + b_3 u - \beta_0\dddot{u} - \beta_1\ddot{u} - \beta_2\dot{u} \\
&= -a_1x_3 - a_2x_2 - a_3x_1 + (b_0 - \beta_0)\dddot{u} + (b_1 - \beta_1 - a_1\beta_0)\ddot{u} \\
&\quad + (b_2 - \beta_2 - a_1\beta_1 - a_2\beta_0)\dot{u} + (b_3 - a_1\beta_2 - a_2\beta_1 - a_3\beta_0)u \\
&= -a_1x_3 - a_2x_2 - a_3x_1 + (b_3 - a_1\beta_2 - a_2\beta_1 - a_3\beta_0)u \\
&= -a_1x_3 - a_2x_2 - a_3x_1 + \beta_3 u
\end{aligned}$$

Hence, we get

$$\dot{x}_3 = -a_3x_1 - a_2x_2 - a_1x_3 + \beta_3 u \tag{3–88}$$

Combining Equations (3–86), (3–87), and (3–88) into a vector-matrix differential equation, we obtain Equation (3–84). Also, from the definition of state variable $x_1$, we get the output equation given by Equation (3–85).

**A-3-4** Obtain a state-space model of the system shown in Figure 3–31.

**Solution.** The system involves one integrator and two delayed integrators. The output of each integrator or delayed integrator can be a state variable. Let us define the output of the plant as $x_1$, the output of the controller as $x_2$, and the output of the sensor as $x_3$. Then we obtain

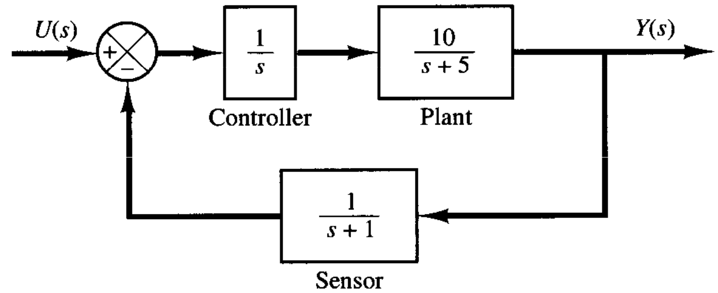



**Figure 3–31**
Control system.

$$\frac{X_1(s)}{X_2(s)} = \frac{10}{s+5}$$

$$\frac{X_2(s)}{U(s) - X_3(s)} = \frac{1}{s}$$

$$\frac{X_3(s)}{X_1(s)} = \frac{1}{s+1}$$

$$Y(s) = X_1(s)$$

which can be rewritten as

$$sX_1(s) = -5X_1(s) + 10X_2(s)$$

$$sX_2(s) = -X_3(s) + U(s)$$

$$sX_3(s) = X_1(s) - X_3(s)$$

$$Y(s) = X_1(s)$$

By taking the inverse Laplace transforms of the preceding four equations, we obtain

$$\dot{x}_1 = -5x_1 + 10x_2$$

$$\dot{x}_2 = -x_3 + u$$

$$\dot{x}_3 = x_1 - x_3$$

$$y = x_1$$

Thus, a state-space model of the system in the standard form is given by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -5 & 10 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} u$$

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

It is important to note that this is not the only state-space representation of the system. Many other state-space representations are possible. However, the number of state variables is the same in any state-space representation of the same system. In the present system, the number of state variables is three, regardless of what variables are chosen as state variables.

**A-3-5.** Obtain a state-space model for the system shown in Figure 3–32(a).

**Solution.** First, notice that $(as + b)/s^2$ involves a derivative term. Such a derivative term may be avoided if we modify $(as + b)/s^2$ as

$$\frac{as + b}{s^2} = \left(a + \frac{b}{s}\right)\frac{1}{s}$$

Using this modification, the block diagram of Figure 3–32(a) can be modified to that shown in Figure 3–32(b).

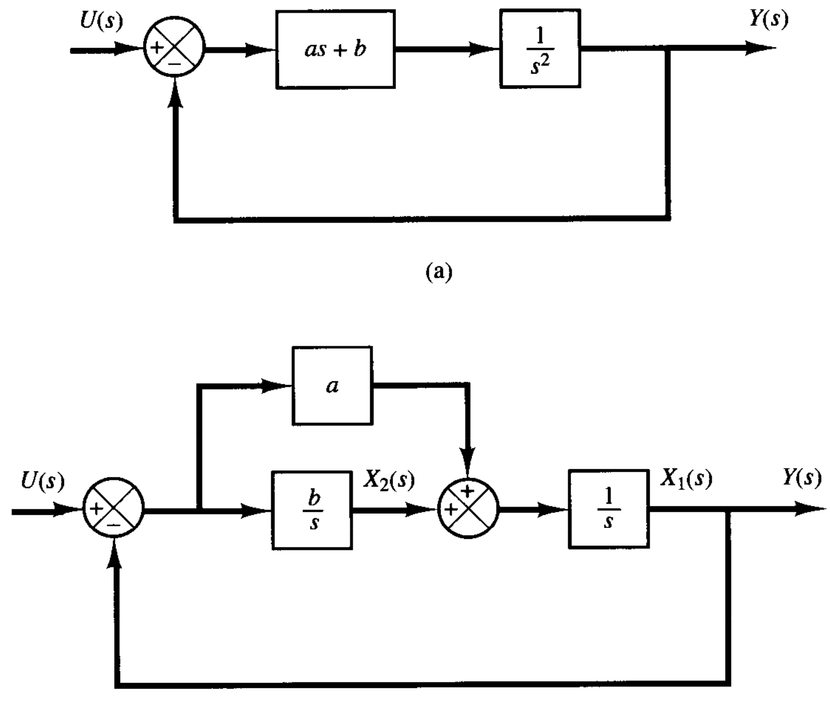



**Figure 3–32**
(a) Control system;
(b) modified block diagram.

Define the outputs of the integrators as state variables, as shown in Figure 3–32(b). Then from Figure 3–32(b) we obtain

$$\frac{X_1(s)}{X_2(s) + a[U(s) - X_1(s)]} = \frac{1}{s}$$

$$\frac{X_2(s)}{U(s) - X_1(s)} = \frac{b}{s}$$

$$Y(s) = X_1(s)$$

which may be modified to

$$sX_1(s) = X_2(s) + a[U(s) - X_1(s)]$$

$$sX_2(s) = -bX_1(s) + bU(s)$$

$$Y(s) = X_1(s)$$

Taking the inverse Laplace transforms of the preceding three equations, we obtain

$$\dot{x}_1 = -ax_1 + x_2 + au$$

$$\dot{x}_2 = -bx_1 + bu$$

$$y = x_1$$

Rewriting the state and output equations in the standard vector-matrix form, we obtain

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -a & 1 \\ -b & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} a \\ b \end{bmatrix} u$$

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

**A-3-6.** Obtain a state-space representation of the system shown in Figure 3–33(a).

**Solution.** In this problem, first expand $(s + z)/(s + p)$ into partial fractions.

$$\frac{s+z}{s+p} = 1 + \frac{z-p}{s+p}$$

Next covert $K/[s(s + a)]$ into the product of $K/s$ and $1/(s + a)$. Then redraw the block diagram, as shown in Figure 3–33(b). Defining a set of state variables, as shown in Figure 3–33(b), we obtain the following equations:

$$\dot{x}_1 = -ax_1 + x_2$$
$$\dot{x}_2 = -Kx_1 + Kx_3 + Ku$$
$$\dot{x}_3 = -(z - p)x_1 - px_3 + (z - p)u$$
$$y = x_1$$

Rewriting gives

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -a & 1 & 0 \\ -K & 0 & K \\ -(z-p) & 0 & -p \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ K \\ z-p \end{bmatrix} u$$

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

Notice that the output of the integrator and the outputs of the first-order delayed integrators $[1/(s + a)$ and $(z - p)/(s + p)]$ are chosen as state variables. It is important to remember that the output of the block $(s + z)/(s + p)$ in Figure 3–33(a) cannot be a state variable, because this block involves a derivative term, $s + z$.

**A-3-7.** Gyros for sensing angular motion are commonly used in inertial guidance systems, autopilot systems, and the like. Figure 3–34(a) shows a single-degree-of-freedom gyro. The spinning wheel is

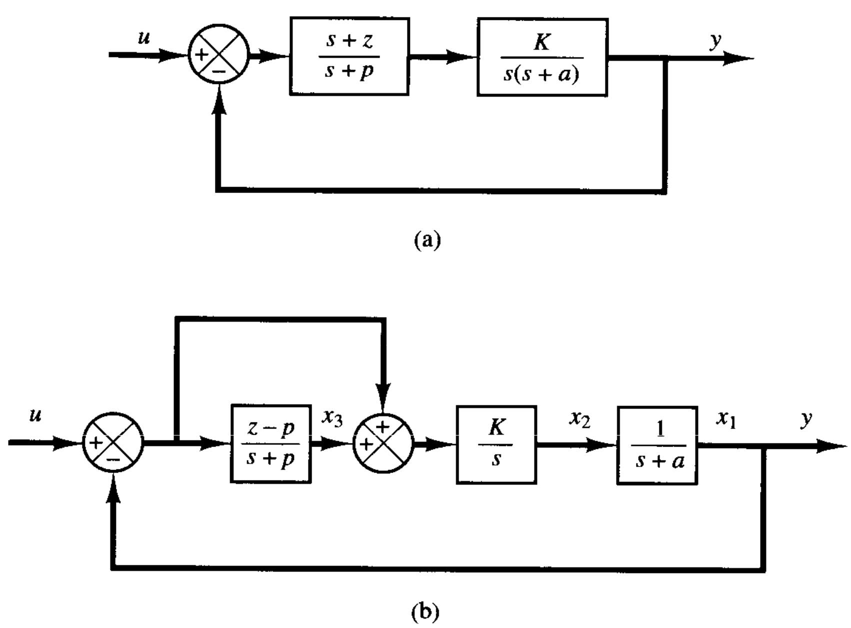



**Figure 3–33**
(a) Control system; (b) block diagram defining state variables for the system.

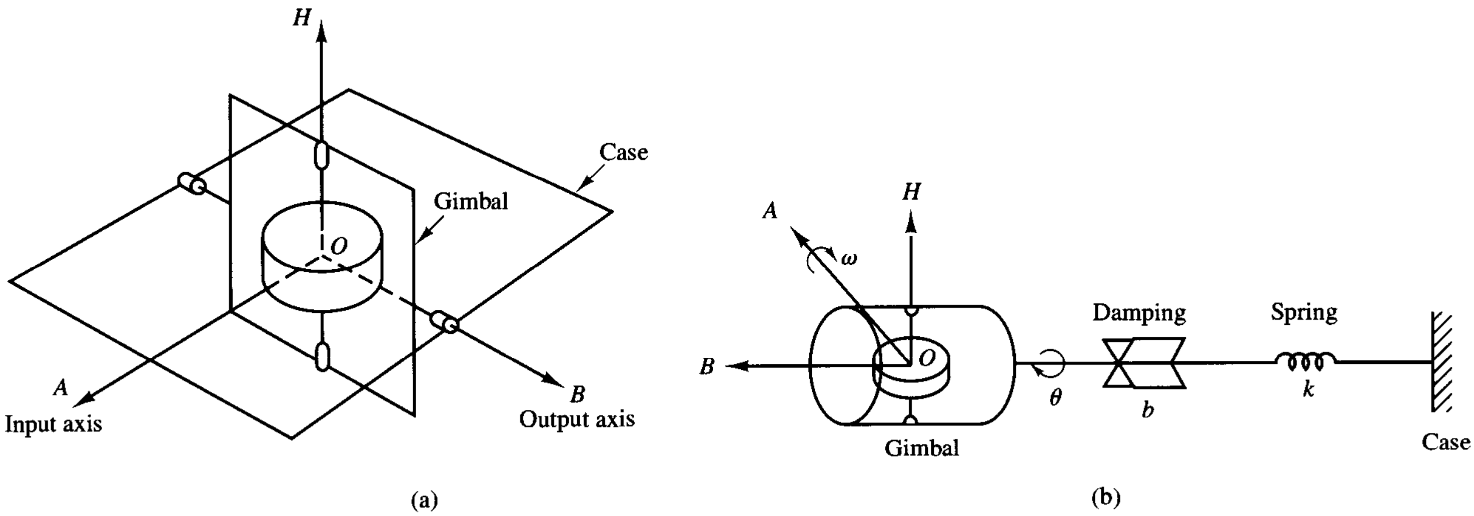



**Figure 3–34**
(a) Schematic diagram of a single-degree-of-freedom gyro; (b) functional diagram of the gyro shown in part (a).

mounted in a movable gimbal, which is, in turn, mounted in a gyro case. The gimbal is free to move relative to the case about the output axis $OB$. Note that the output axis is perpendicular to the wheel spin axis. The input axis around which a turning rate, or angle, is measured is perpendicular to both the output and spin axes. Information on the input signal (the turning rate or angle around the input axis) is obtained from the resulting motion of the gimbal about the output axis, relative to the case.

Figure 3–34(b) shows a functional diagram of the gyro system. The equation of motion about the output axis can be obtained by equating the rate of change of angular momentum to the sum of the external torques.

The change in angular momentum about axis $OB$ consists of two parts: $I\ddot{\theta}$, the change due to acceleration of the gimbal around axis $OB$, and $-H\omega\cos\theta$, the change due to the turning of the wheel angular-momentum vector around axis $OA$. The external torques consist of $-b\dot{\theta}$, the damping torque, and $-k\theta$, the spring torque. Thus the equation of the gyro system is

$$I\ddot{\theta} - H\omega\cos\theta = -b\dot{\theta} - k\theta$$

or

$$I\ddot{\theta} + b\dot{\theta} + k\theta = H\omega\cos\theta \tag{3–89}$$

In practice, $\theta$ is a very small angle, usually not more than $\pm 2.5°$.

Obtain a state-space representation of the gyro system.

**Solution.** In this system, $\theta$ and $\dot{\theta}$ may be chosen as state variables. The input variable is $\omega$ and the output variable is $\theta$. Let us define

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \theta \\ \dot{\theta} \end{bmatrix}, \qquad u = \omega, \qquad y = \theta$$

Then Equation (3–89) can be written as follows:

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

$$\dot{x}_2 = -\frac{k}{I}x_1 - \frac{b}{I}x_2 + \frac{H}{I}u\cos x_1$$

or

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, u)$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad \mathbf{f}(\mathbf{x}, u) = \begin{bmatrix} f_1(\mathbf{x}, u) \\ f_2(\mathbf{x}, u) \end{bmatrix} = \begin{bmatrix} x_2 \\ -\dfrac{k}{I}x_1 - \dfrac{b}{I}x_2 + \dfrac{H}{I}u\cos x_1 \end{bmatrix}$$

Clearly, $f_2(\mathbf{x}, u)$ involves a nonlinear term in $x_1$ and $u$. By expanding $\cos x_1$ into its series representation,

$$\cos x_1 = 1 - \frac{1}{2}x_1^2 + \cdots$$

and noting that $x_1$ is a very small angle, we may approximate $\cos x_1$ by unity to obtain the following linearized state equation:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\dfrac{k}{I} & -\dfrac{b}{I} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \dfrac{H}{I} \end{bmatrix} u$$

The output equation is

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

**A-3-8.** Consider a system defined by the following state-space equations:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -5 & -1 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 2 \\ 5 \end{bmatrix} u$$

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

Obtain the transfer function $G(s)$ of the system.

**Solution.** Referring to Equation (3–32), the transfer function of the system can be obtained as follows (note that $D = 0$ in this case):

$$\begin{aligned} G(s) &= \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} \\ &= [1 \quad 2] \begin{bmatrix} s+5 & 1 \\ -3 & s+1 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 5 \end{bmatrix} \\ &= [1 \quad 2] \begin{bmatrix} \dfrac{s+1}{(s+2)(s+4)} & \dfrac{-1}{(s+2)(s+4)} \\ \dfrac{3}{(s+2)(s+4)} & \dfrac{s+5}{(s+2)(s+4)} \end{bmatrix} \begin{bmatrix} 2 \\ 5 \end{bmatrix} \\ &= \frac{12s + 59}{(s+2)(s+4)} \end{aligned}$$

**A-3-9.** Figure 3–35(a) shows a schematic diagram of an automobile suspension system. As the car moves along the road, the vertical displacements at the tires act as the motion excitation to the automobile suspension system. The motion of this system consists of a translational motion of the center of mass and a rotational motion about the center of mass. Mathematical modeling of the complete system is quite complicated.

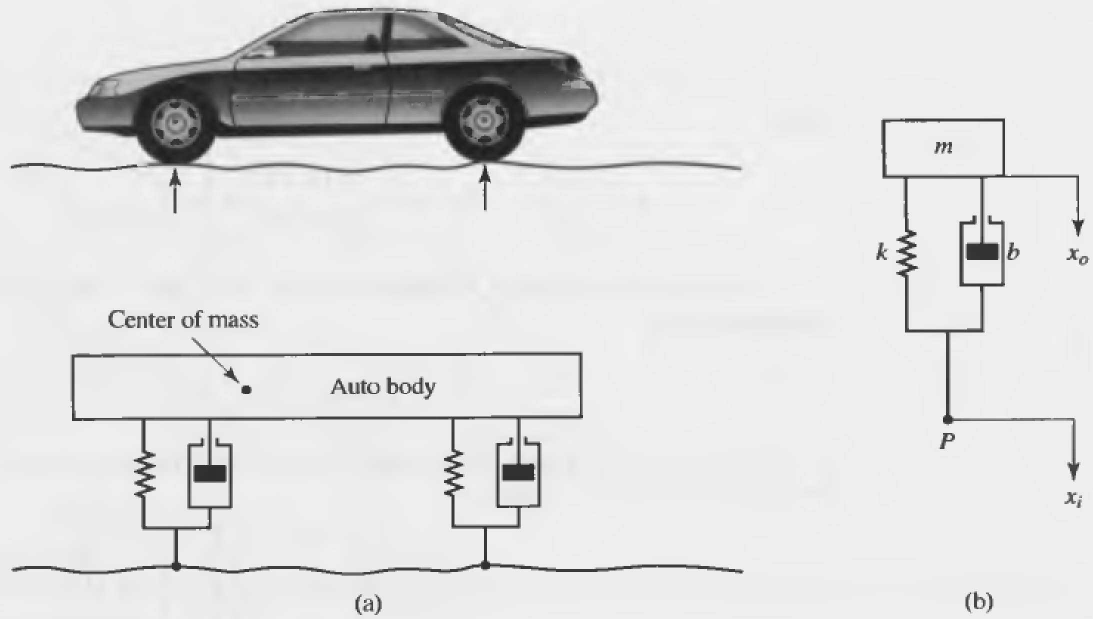



**Figure 3–35**
(a) Automobile suspension system;
(b) simplified suspension system.

A very simplified version of the suspension system is shown in Figure 3–35(b). Assuming that the motion $x_i$ at point $P$ is the input to the system and the vertical motion $x_o$ of the body is the output, obtain the transfer function $X_o(s)/X_i(s)$. (Consider the motion of the body only in the vertical direction.) Displacement $x_o$ is measured from the equilibrium position in the absence of input $x_i$.

**Solution.** The equation of motion for the system shown in Figure 3–35(b) is

$$m\ddot{x}_o + b(\dot{x}_o - \dot{x}_i) + k(x_o - x_i) = 0$$

or

$$m\ddot{x}_o + b\dot{x}_o + kx_o = b\dot{x}_i + kx_i$$

Taking the Laplace transform of this last equation, assuming zero initial conditions, we obtain

$$(ms^2 + bs + k)X_o(s) = (bs + k)X_i(s)$$

Hence the transfer function $X_o(s)/X_i(s)$ is given by

$$\frac{X_o(s)}{X_i(s)} = \frac{bs + k}{ms^2 + bs + k}$$

**A-3-10.** Obtain the transfer function $Y(s)/U(s)$ of the system shown in Figure 3–36. (Similar to the system of Problem A-3-9, this is also a simplified version of an automobile or motorcycle suspension system.)

**Solution.** Applying the Newton's second law to the system, we obtain

$$m_1\ddot{x} = k_2(y - x) + b(\dot{y} - \dot{x}) + k_1(u - x)$$
$$m_2\ddot{y} = -k_2(y - x) - b(\dot{y} - \dot{x})$$

Hence, we have

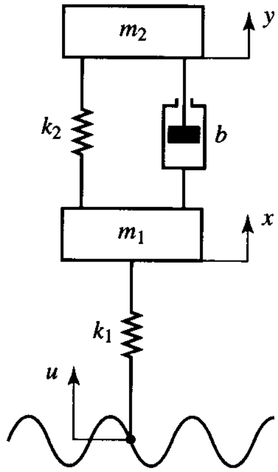



**Figure 3–36**
Suspension system.

$$m_1\ddot{x} + b\dot{x} + (k_1 + k_2)x = b\dot{y} + k_2y + k_1u$$

$$m_2\ddot{y} + b\dot{y} + k_2y = b\dot{x} + k_2x$$

Taking Laplace transforms of these two equations, assuming zero initial conditions, we obtain

$$[m_1s^2 + bs + (k_1 + k_2)]X(s) = (bs + k_2)Y(s) + k_1U(s)$$

$$[m_2s^2 + bs + k_2]Y(s) = (bs + k_2)X(s)$$

Eliminating $X(s)$ from the last two equations, we have

$$(m_1s^2 + bs + k_1 + k_2)\frac{m_2s^2 + bs + k_2}{bs + k_2}Y(s) = (bs + k_2)Y(s) + k_1U(s)$$

which yields

$$\frac{Y(s)}{U(s)} = \frac{k_1(bs + k_2)}{m_1m_2s^4 + (m_1 + m_2)bs^3 + [k_1m_2 + (m_1 + m_2)k_2]s^2 + k_1bs + k_1k_2}$$

**A-3-11.** Consider the electrical circuit shown in Figure 3–37. Obtain the transfer function $E_o(s)/E_i(s)$ by use of the block diagram approach.

**Solution.** Equations for the circuits are

$$\frac{1}{C_1}\int (i_1 - i_2)\,dt + R_1i_1 = e_i \tag{3–90}$$

$$\frac{1}{C_1}\int (i_2 - i_1)\,dt + R_2i_2 + \frac{1}{C_2}\int i_2\,dt = 0 \tag{3–91}$$

$$\frac{1}{C_2}\int i_2\,dt = e_o \tag{3–92}$$

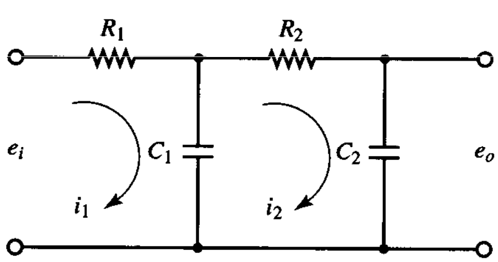



**Figure 3–37**
Electrical circuit.

The Laplace transforms of Equations (3–90), (3–91), and (3–92), with zero initial conditions, give

$$\frac{1}{C_1 s}[I_1(s) - I_2(s)] + R_1 I_1(s) = E_i(s) \tag{3–93}$$

$$\frac{1}{C_1 s}[I_2(s) - I_1(s)] + R_2 I_2(s) + \frac{1}{C_2 s} I_2(s) = 0 \tag{3–94}$$

$$\frac{1}{C_2 s} I_2(s) = E_o(s) \tag{3–95}$$

Equation (3–93) can be rewritten as

$$C_1 s[E_i(s) - R_1 I_1(s)] = I_1(s) - I_2(s) \tag{3–96}$$

Equation (3–96) gives the block diagram shown in Figure 3–38(a). Equation (3–94) can be modified to

$$I_2(s) = \frac{C_2 s}{R_2 C_2 s + 1} \frac{1}{C_1 s}[I_1(s) - I_2(s)] \tag{3–97}$$

Equation (3–97) yields the block diagram shown in Figure 3–38(b). Also, Equation (3–95) gives the block diagram shown in Figure 3–38(c). Combining the block diagrams of Figures 3–38(a), (b), and (c), we obtain Figure 3–39(a). This block diagram can be successively modified as shown in Figures 3–39(b) through (e). Thus we obtained the transfer function $E_o(s)/E_i(s)$ of the system. [This is the same as that we derived earlier for the same electrical circuit. See Equation (3–66).]

**A-3-12.** Obtain the transfer function of the mechanical system shown in Figure 3–40(a). Also obtain the transfer function of the electrical system shown in Figure 3–40(b). Show that the transfer functions of the two systems are of identical form and thus they are analogous systems.

**Solution.** The equations of motion for the mechanical system shown in Figure 3–40(a) are

$$b_1(\dot{x}_i - \dot{x}_o) + k_1(x_i - x_o) = b_2(\dot{x}_o - \dot{y})$$

$$b_2(\dot{x}_o - \dot{y}) = k_2 y$$

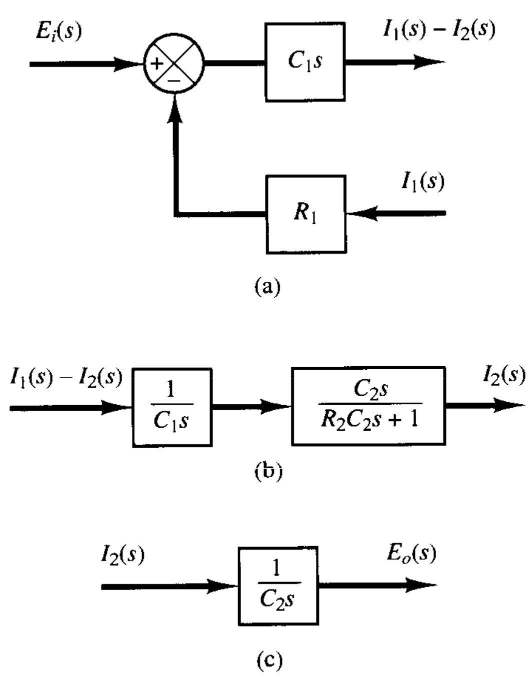



**Figure 3–38**
Block diagrams: (a) corresponding to Equation (3–96); (b) corresponding to Equation (3–97); (c) corresponding to Equation (3–95).

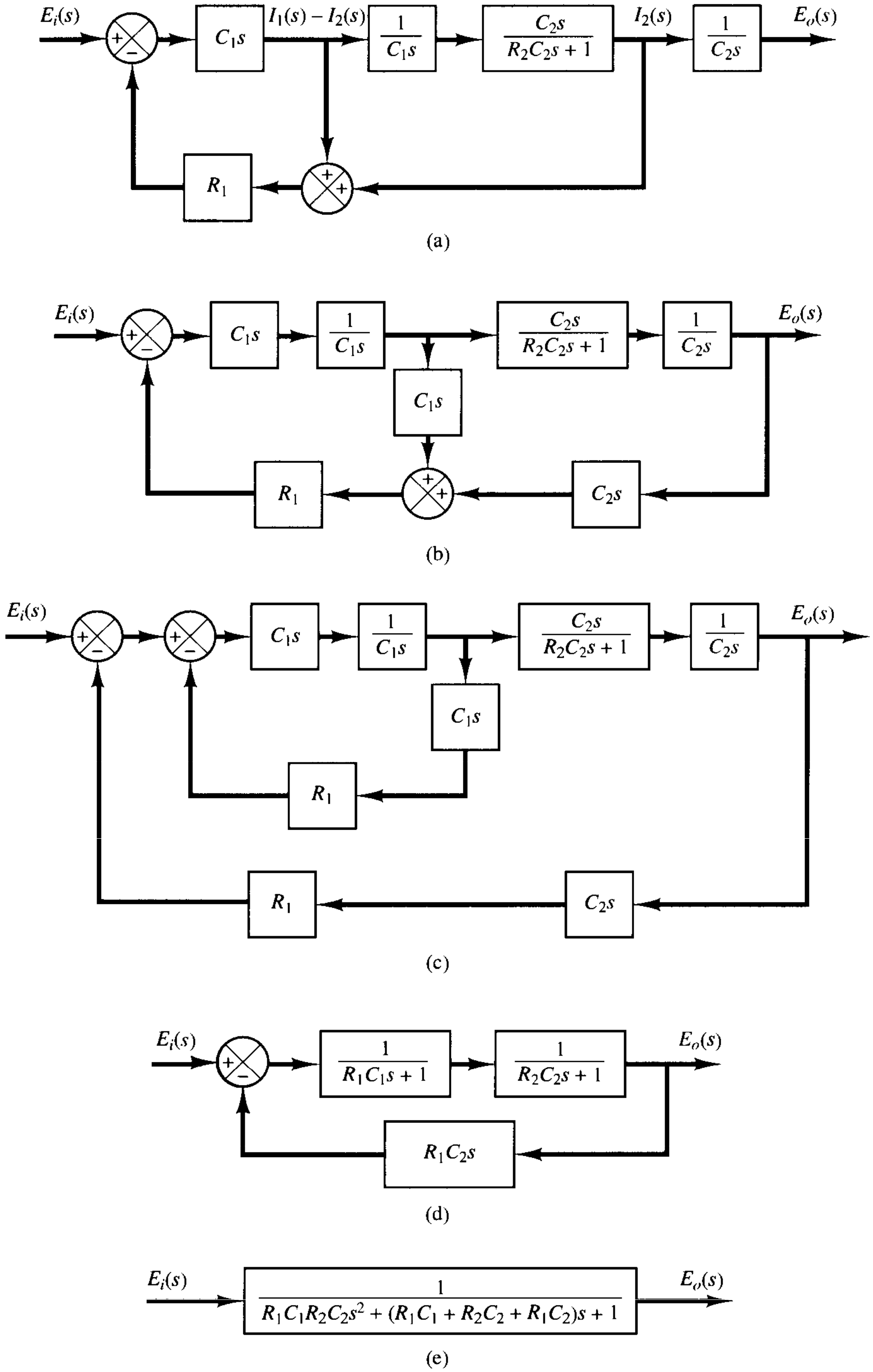



**Figure 3–39** Block diagrams for the system shown in Figure 3–37. (a) through (e) show successive simplifications of block diagrams.

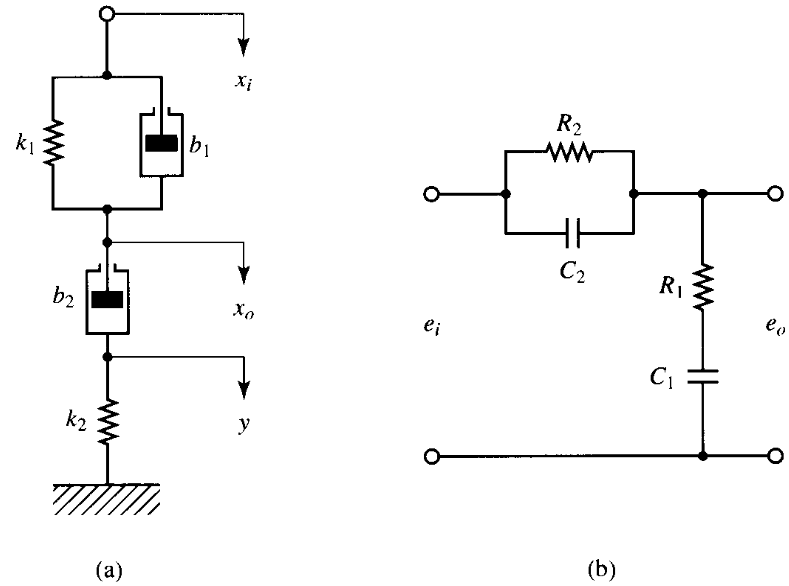



**Figure 3–40**
(a) Mechanical system; (b) analogous electrical system.

By taking the Laplace transforms of these two equations, assuming zero initial conditions, we have

$$b_1[sX_i(s) - sX_o(s)] + k_1[X_i(s) - X_o(s)] = b_2[sX_o(s) - sY(s)]$$

$$b_2[sX_o(s) - sY(s)] = k_2Y(s)$$

If we eliminate $Y(s)$ from the last two equations, then we obtain

$$b_1[sX_i(s) - sX_o(s)] + k_1[X_i(s) - X_o(s)] = b_2sX_o(s) - b_2s\frac{b_2sX_o(s)}{b_2s + k_2}$$

or

$$(b_1s + k_1)X_i(s) = \left(b_1s + k_1 + b_2s - b_2s\frac{b_2s}{b_2s + k_2}\right)X_o(s)$$

Hence the transfer function $X_o(s)/X_i(s)$ can be obtained as

$$\frac{X_o(s)}{X_i(s)} = \frac{\left(\frac{b_1}{k_1}s + 1\right)\left(\frac{b_2}{k_2}s + 1\right)}{\left(\frac{b_1}{k_1}s + 1\right)\left(\frac{b_2}{k_2}s + 1\right) + \frac{b_2}{k_1}s}$$

For the electrical system shown in Figure 3–40(b), the transfer function $E_o(s)/E_i(s)$ is found to be

$$\frac{E_o(s)}{E_i(s)} = \frac{R_1 + \frac{1}{C_1s}}{\frac{1}{(1/R_2) + C_2s} + R_1 + \frac{1}{C_1s}}$$

$$= \frac{(R_1C_1s + 1)(R_2C_2s + 1)}{(R_1C_1s + 1)(R_2C_2s + 1) + R_2C_1s}$$

A comparison of the transfer functions shows that the systems shown in Figures 3–40(a) and (b) are analogous.

**A-3-13.** In the liquid-level system of Figure 3–41, assume that the outflow rate $Q$ m³/sec through the outflow valve is related to the head $H$ m by

$$Q = K\sqrt{H} = 0.01\sqrt{H}$$

Assume also that when the inflow rate $Q_i$ is 0.015 m³/sec the head stays constant. At $t = 0$ the inflow valve is closed and so there is no inflow for $t \geq 0$. Find the time necessary to empty the tank to half the original head. The capacitance $C$ of the tank is 2 m².

**Solution.** When the head is stationary, the inflow rate equals the outflow rate. Thus head $H_o$ at $t = 0$ is obtained from

$$0.015 = 0.01\sqrt{H_o}$$

or

$$H_o = 2.25 \text{ m}$$

The equation for the system for $t > 0$ is

$$-C\,dH = Q\,dt$$

or

$$\frac{dH}{dt} = -\frac{Q}{C} = \frac{-0.01\sqrt{H}}{2}$$

Hence

$$\frac{dH}{\sqrt{H}} = -0.005\,dt$$

Assume that, at $t = t_1$, $H = 1.125$ m. Integrating both sides of this last equation, we obtain

$$\int_{2.25}^{1.125} \frac{dH}{\sqrt{H}} = \int_0^{t_1} (-0.005)\,dt = -0.005t_1$$

It follows that

$$2\sqrt{H}\,\Big|_{2.25}^{1.125} = 2\sqrt{1.125} - 2\sqrt{2.25} = -0.005t_1$$

or

$$t_1 = 175.7$$

Thus, the head becomes half the original value (2.25 m) in 175.7 sec.

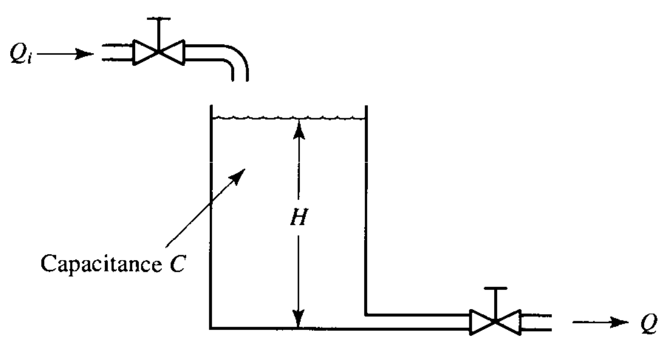



**Figure 3–41**
Liquid-level system.

**A-3-14.** Consider the liquid-level system shown in Figure 3–42. At steady state, the inflow rate and outflow rate are both $\bar{Q}$ and the flow rate between the tanks is zero. The heads of tanks 1 and 2 are both $\bar{H}$. At $t = 0$, the inflow rate is changed from $\bar{Q}$ to $\bar{Q} + q$, where $q$ is a small change in the inflow rate. The resulting changes in the heads ($h_1$ and $h_2$) and flow rates ($q_1$ and $q_2$) are assumed to be small. The capacitances of tanks 1 and 2 are $C_1$ and $C_2$, respectively. The resistance of the valve between the tanks is $R_1$ and that of the outflow valve is $R_2$.

Derive mathematical models for the system when (a) $q$ is the input and $h_2$ the output, (b) $q$ is the input and $q_2$ the output, and (c) $q$ is the input and $h_1$ the output.

**Solution.** (a) For tank 1, we have

$$C_1\, dh_1 = q_1\, dt$$

where

$$q_1 = \frac{h_2 - h_1}{R_1}$$

Consequently,

$$R_1 C_1 \frac{dh_1}{dt} + h_1 = h_2 \tag{3–98}$$

For tank 2, we get

$$C_2\, dh_2 = (q - q_1 - q_2)\, dt$$

where

$$q_1 = \frac{h_2 - h_1}{R_1}, \qquad q_2 = \frac{h_2}{R_2}$$

It follows that

$$R_2 C_2 \frac{dh_2}{dt} + \frac{R_2}{R_1} h_2 + h_2 = R_2 q + \frac{R_2}{R_1} h_1 \tag{3–99}$$

By eliminating $h_1$ from Equations (3–98) and (3–99), we have

$$R_1 C_1 R_2 C_2 \frac{d^2 h_2}{dt^2} + (R_1 C_1 + R_2 C_2 + R_2 C_1) \frac{dh_2}{dt} + h_2 = R_1 R_2 C_1 \frac{dq}{dt} + R_2 q \tag{3–100}$$

In terms of the transfer function, we have

$$\frac{H_2(s)}{Q(s)} = \frac{R_2(R_1 C_1 s + 1)}{R_1 C_1 R_2 C_2 s^2 + (R_1 C_1 + R_2 C_2 + R_2 C_1)s + 1}$$

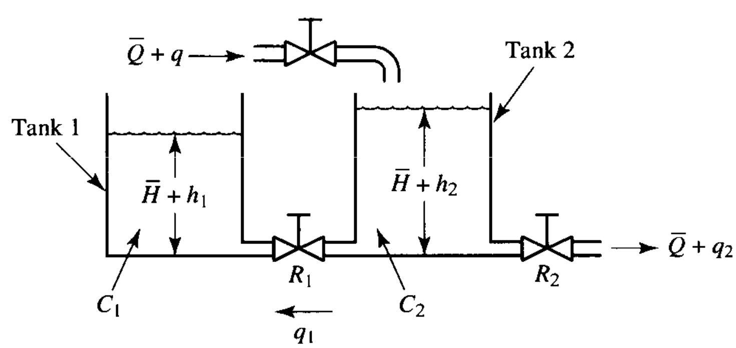



**Figure 3–42**
Liquid-level system.

This is the desired mathematical model in which $q$ is considered the input and $h_2$ is the output.
(b) Substitution of $h_2 = R_2q_2$ into Equation (3–100) gives

$$R_1C_1R_2C_2\frac{d^2q_2}{dt^2} + (R_1C_1 + R_2C_2 + R_2C_1)\frac{dq_2}{dt} + q_2 = R_1C_1\frac{dq}{dt} + q$$

This equation is a mathematical model of the system when $q$ is considered the input and $q_2$ is the output. In terms of the transfer function, we obtain

$$\frac{Q_2(s)}{Q(s)} = \frac{R_1C_1s + 1}{R_1C_1R_2C_2s^2 + (R_1C_1 + R_2C_2 + R_2C_1)s + 1}$$

(c) Elimination of $h_2$ from Equations (3–98) and (3–99) yields

$$R_1C_1R_2C_2\frac{d^2h_1}{dt^2} + (R_1C_1 + R_2C_2 + R_2C_1)\frac{dh_1}{dt} + h_1 = R_2q$$

which is a mathematical model of the system in which $q$ is considered the input and $h_1$ is the output. In terms of the transfer function, we get

$$\frac{H_1(s)}{Q(s)} = \frac{R_2}{R_1C_1R_2C_2s^2 + (R_1C_1 + R_2C_2 + R_2C_1)s + 1}$$

**A-3-15.** Consider the liquid-level system shown in Figure 3–43. In the system, $\bar{Q}_1$ and $\bar{Q}_2$ are steady-state inflow rates and $\bar{H}_1$ and $\bar{H}_2$ are steady-state heads. The quantities $q_{i1}, q_{i2}, h_1, h_2, q_1$, and $q_o$ are considered small. Obtain a state-space representation for the system when $h_1$ and $h_2$ are the outputs and $q_{i1}$ and $q_{i2}$ are the inputs.

**Solution.** The equations for the system are

$$C_1dh_1 = (q_{i1} - q_1)\,dt \tag{3–101}$$

$$\frac{h_1 - h_2}{R_1} = q_1 \tag{3–102}$$

$$C_2dh_2 = (q_1 + q_{i2} - q_o)\,dt \tag{3–103}$$

$$\frac{h_2}{R_2} = q_o \tag{3–104}$$

Elimination of $q_1$ from Equation (3–101) using Equation (3–102) results in

$$\frac{dh_1}{dt} = \frac{1}{C_1}\left(q_{i1} - \frac{h_1 - h_2}{R_1}\right) \tag{3–105}$$

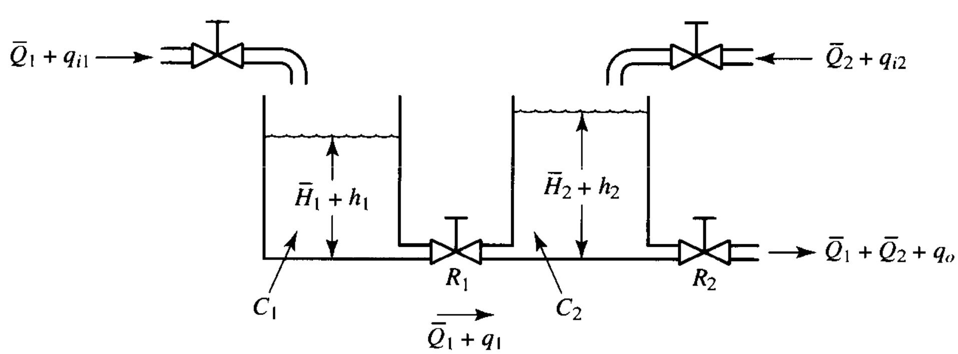



**Figure 3–43**
Liquid-level system.

Eliminating $q_1$ and $q_o$ from Equation (3–103) by using Equations (3–102) and (3–104) gives

$$\frac{dh_2}{dt} = \frac{1}{C_2}\left(\frac{h_1 - h_2}{R_1} + q_{i2} - \frac{h_2}{R_2}\right) \tag{3–106}$$

Define state variables $x_1$ and $x_2$ by

$$x_1 = h_1$$
$$x_2 = h_2$$

the input variables $u_1$ and $u_2$ by

$$u_1 = q_{i1}$$
$$u_2 = q_{i2}$$

and the output variables $y_1$ and $y_2$ by

$$y_1 = h_1 = x_1$$
$$y_2 = h_2 = x_2$$

Then Equations (3–105) and (3–106) can be written as

$$\dot{x}_1 = -\frac{1}{R_1C_1}x_1 + \frac{1}{R_1C_1}x_2 + \frac{1}{C_1}u_1$$

$$\dot{x}_2 = \frac{1}{R_1C_2}x_1 - \left(\frac{1}{R_1C_2} + \frac{1}{R_2C_2}\right)x_2 + \frac{1}{C_2}u_2$$

In the form of the standard vector-matrix representation, we have

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -\dfrac{1}{R_1C_1} & \dfrac{1}{R_1C_1} \\ \dfrac{1}{R_1C_2} & -\left(\dfrac{1}{R_1C_2} + \dfrac{1}{R_2C_2}\right) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \dfrac{1}{C_1} & 0 \\ 0 & \dfrac{1}{C_2} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

which is the state equation, and

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

which is the output equation.

**A-3-16.** Considering small deviations from steady-state operation, draw a block diagram of the air heating system shown in Figure 3–44. Assume that the heat loss to the surroundings and the heat capacitance of the metal parts of the heater are negligible.

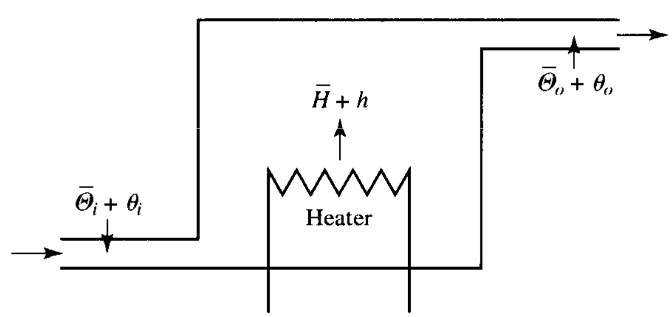



**Figure 3–44**
Air heating system.

**Solution.** Let us define

$\bar{\Theta}_i$ = steady-state temperature of inlet air, °C

$\bar{\Theta}_o$ = steady-state temperature of outlet air, °C

$G$ = mass flow rate of air through the heating chamber, kg/sec

$M$ = mass of air contained in the heating chamber, kg

$c$ = specific heat of air, kcal/kg °C

$R$ = thermal resistance, °C sec/kcal

$C$ = thermal capacitance of air contained in the heating chamber = $Mc$, kcal/°C

$\bar{H}$ = steady-state heat input, kcal/sec

Let us assume that the heat input is suddenly changed from $\bar{H}$ to $\bar{H} + h$ and the inlet air temperature is suddenly changed from $\bar{\Theta}_i$ to $\bar{\Theta}_i + \theta_i$. Then the outlet air temperature will be changed from $\bar{\Theta}_o$ to $\bar{\Theta}_o + \theta_o$.

The equation describing the system behavior is

$$C\,d\theta_o = [h + Gc(\theta_i - \theta_o)]\,dt$$

or

$$C\frac{d\theta_o}{dt} = h + Gc(\theta_i - \theta_o)$$

Noting that

$$Gc = \frac{1}{R}$$

we obtain

$$C\frac{d\theta_o}{dt} = h + \frac{1}{R}(\theta_i - \theta_o)$$

or

$$RC\frac{d\theta_o}{dt} + \theta_o = Rh + \theta_i$$

Taking the Laplace transforms of both sides of this last equation and substituting the initial condition that $\theta_o(0) = 0$, we obtain

$$\Theta_o(s) = \frac{R}{RCs + 1}H(s) + \frac{1}{RCs + 1}\Theta_i(s)$$

The block diagram of the system corresponding to this equation is shown in Figure 3–45.

**A-3-17.** Consider the thin, glass-wall, mercury thermometer system shown in Figure 3–46. Assume that the thermometer is at a uniform temperature $\bar{\Theta}$°C (ambient temperature) and that at $t = 0$ it is immersed in a bath of temperature $\bar{\Theta} + \theta_b$°C, where $\theta_b$ is the bath temperature (which may be constant or changing) measured from the ambient temperature $\bar{\Theta}$. Define the instantaneous thermometer temperature by $\bar{\Theta} + \theta$°C so that $\theta$ is the change in the thermometer temperature satisfying the condition that $\theta(0) = 0$. Obtain a mathematical model for the system. Also obtain an electrical analog of the thermometer system.

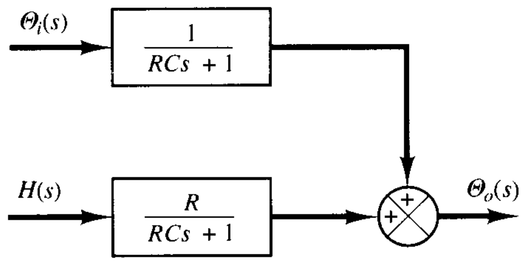



**Figure 3–45**
Block diagram of the air heating system shown in Figure 3–44.

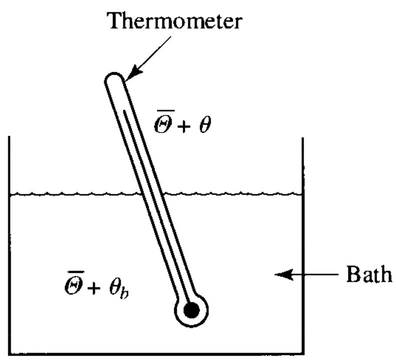



**Figure 3–46**
Thin, glass-wall, mercury thermometer system.

**Solution.** A mathematical model for the system can be derived by considering heat balance as follows: The heat entering the thermometer during $dt$ sec is $q\,dt$, where $q$ is the heat flow rate to the thermometer. This heat is stored in the thermal capacitance $C$ of the thermometer, thereby raising its temperature by $d\theta$. Thus the heat-balance equation is

$$C\,d\theta = q\,dt \tag{3–107}$$

Since thermal resistance $R$ may be written as

$$R = \frac{d(\Delta\theta)}{dq} = \frac{\Delta\theta}{q}$$

heat flow rate $q$ may be given, in terms of thermal resistance $R$, as

$$q = \frac{(\bar{\Theta} + \theta_b) - (\bar{\Theta} + \theta)}{R} = \frac{\theta_b - \theta}{R}$$

where $\bar{\Theta} + \theta_b$ is the bath temperature and $\bar{\Theta} + \theta$ is the thermometer temperature. Hence, we can rewrite Equation (3–107) as

$$C\frac{d\theta}{dt} = \frac{\theta_b - \theta}{R}$$

or

$$RC\frac{d\theta}{dt} + \theta = \theta_b \tag{3–108}$$

Equation (3–108) is a mathematical model of the thermometer system.

Referring to Equation (3–108), an electrical analog for the thermometer system can be written as

$$RC\frac{de_o}{dt} + e_o = e_i$$

An electrical circuit represented by this last equation is shown in Figure 3–47.

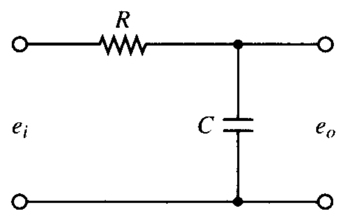



**Figure 3–47**
Electrical analog of the thermometer system shown in Figure 3–46.

**A-3-18.** 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 $x = 5, y = 10$.

**Solution.** 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%.

**A-3-19.** Consider the liquid-level system shown in Figure 3–48. At steady state the inflow rate is $Q_i = \bar{Q}$, the outflow rate is $Q_o = \bar{Q}$, and head is $H = \bar{H}$. If the flow is turbulent, then we have

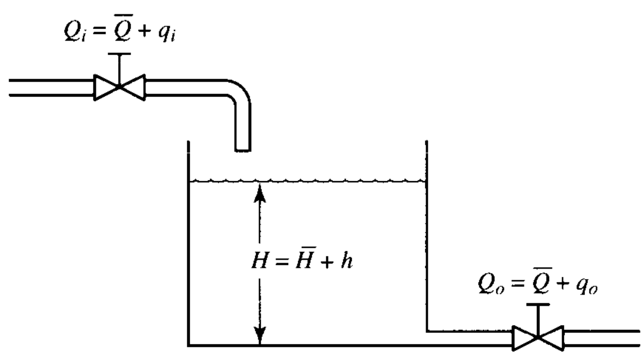



**Figure 3–48**
Liquid-level system.

$$\bar{Q} = K\sqrt{\bar{H}}$$

Assume that at $t = 0$ the inflow rate is changed from $Q_i = \bar{Q}$ to $Q_i = \bar{Q} + q_i$. This change causes the head to change from $H = \bar{H}$ to $H = \bar{H} + h$, which, in turn, causes the outflow rate to change from $Q_o = \bar{Q}$ to $Q_o = \bar{Q} + q_o$. For this system we have

$$C\frac{dH}{dt} = Q_i - Q_o = Q_i - K\sqrt{H}$$

where $C$ is the capacitance of the tank. Let us define

$$\frac{dH}{dt} = f(H, Q_i) = \frac{1}{C}Q_i - \frac{K\sqrt{H}}{C} \tag{3–109}$$

Note that the steady-state operating condition is $(\bar{H}, \bar{Q})$ and $H = \bar{H} + h$, $Q_i = \bar{Q} + q_i$. Since at steady-state operation $dH/dt = 0$, we have $f(\bar{H}, \bar{Q}) = 0$.

Linearize Equation (3–109) near the operating point $(\bar{H}, \bar{Q})$.

**Solution.** Using the linearization technique presented in Section 3–10, a linearized equation for Equation (3–109) can be obtained as follows:

$$\frac{dH}{dt} - f(\bar{H}, \bar{Q}) = \frac{\partial f}{\partial H}(H - \bar{H}) + \frac{\partial f}{\partial Q_i}(Q_i - \bar{Q}) \tag{3–110}$$

where

$$f(\bar{H}, \bar{Q}) = 0$$

$$\left.\frac{\partial f}{\partial H}\right|_{H=\bar{H},\, Q_i=\bar{Q}} = -\frac{K}{2C\sqrt{\bar{H}}} = -\frac{\bar{Q}}{\sqrt{\bar{H}}}\frac{1}{2C\sqrt{\bar{H}}} = -\frac{\bar{Q}}{2C\bar{H}} = -\frac{1}{RC}$$

where we used the resistance $R$ defined by

$$R = \frac{2\bar{H}}{\bar{Q}}$$

Also,

$$\left.\frac{\partial f}{\partial Q_i}\right|_{H=\bar{H},\, Q_i=\bar{Q}} = \frac{1}{C}$$

Then Equation (3–110) can be written as

$$\frac{dH}{dt} = -\frac{1}{RC}(H - \bar{H}) + \frac{1}{C}(Q_i - \bar{Q}) \tag{3–111}$$

Since $H - \bar{H} = h$ and $Q_i - \bar{Q} = q_i$, Equation (3–111) can be written as

$$\frac{dh}{dt} = -\frac{1}{RC}h + \frac{1}{C}q_i$$

or

$$RC\frac{dh}{dt} + h = Rq_i$$

which is the linearized equation for the liquid-level system and is the same as Equation (3–69) that we obtained in Section 3–8.

**A-3-20.** Consider the hydraulic servo system shown in Figure 3–49. Assuming that the load reaction forces are not negligible, derive a mathematical model of the system. Assume also that the mass of the power piston is included in the load mass $m$.

**Solution.** In deriving a mathematical model of the system when the load reactive forces are not negligible, such effects as the pressure drop across the orifice, the leakage of oil around the valve and around the piston, and the compressibility of the oil must be considered.

The pressure drop across the orifice is a function of the supply pressure $p_s$ and the pressure difference $\Delta p = p_1 - p_2$. Thus the flow rate $q$ is a nonlinear function of valve displacement $x$ and pressure difference $\Delta p$ or

$$q = f(x, \Delta p)$$

Linearizing this nonlinear equation about the origin ($x = 0, \Delta p = 0, q = 0$), we obtain, referring to Equation (3–82),

$$q = K_1 x - K_2 \Delta p \tag{3–112}$$

The flow rate $q$ can be considered as consisting of three parts

$$q = q_0 + q_L + q_C \tag{3–113}$$

where $q_0$ = useful flow rate to the power cylinder causing power piston to move, kg/sec
$q_L$ = leakage flow rate, kg/sec
$q_C$ = equivalent compressibility flow rate, kg/sec

Let us obtain specific expressions for $q_0$, $q_L$, and $q_C$. The flow $q_0\,dt$ to the left-hand side of the power piston causes the piston to move to the right by $dy$. So we have

$$A\varrho\, dy = q_0\, dt$$

where $A$ ($\mathrm{m}^2$) is the power piston area, $\varrho$ ($\mathrm{kg/m}^3$) the density of oil, and $dy$ (m) the displacement of the power piston. Then

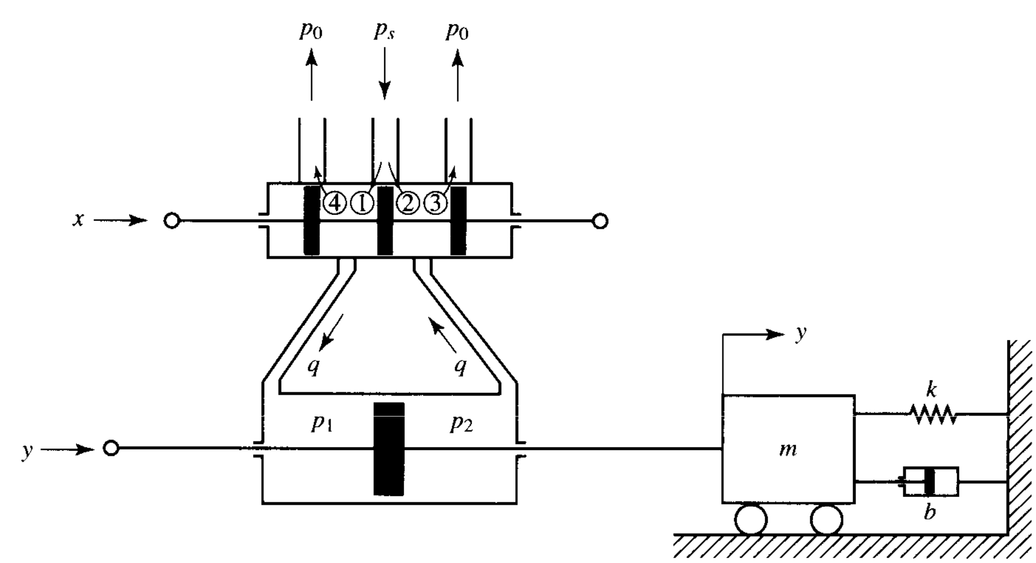



**Figure 3–49** Hydraulic servo system.

$$q_0 = A\varrho \frac{dy}{dt} \tag{3–114}$$

The leakage component $q_L$ can be written

$$q_L = L\,\Delta p \tag{3–115}$$

where $L$ is the leakage coefficient of the system.

The equivalent compressibility flow rate $q_C$ can be expressed in terms of the effective bulk modulus $K$ of oil (including the effects of entrapped air, expansion of pipes, etc.), where

$$K = \frac{d\,\Delta p}{-dV/V}$$

(Here $dV$ is negative and so $-dV$ is positive.) Rewriting this last equation gives

$$-dV = \frac{V}{K}\,d\,\Delta p$$

or

$$\varrho \frac{-dV}{dt} = \frac{\varrho V}{K} \frac{d\,\Delta p}{dt}$$

Noting that $q_C = \varrho(-dV)/dt$, we find

$$q_C = \frac{\varrho V}{K} \frac{d\,\Delta p}{dt} \tag{3–116}$$

where $V$ is the effective volume of oil under compression (that is, approximately half the total power cylinder volume).

Using Equations (3–112) through (3–116),

$$q = K_1 x - K_2\,\Delta p = A\varrho \frac{dy}{dt} + L\,\Delta p + \frac{\varrho V}{K} \frac{d\,\Delta p}{dt}$$

or

$$A\varrho \frac{dy}{dt} + \frac{\varrho V}{K} \frac{d\Delta p}{dt} + (L + K_2)\,\Delta p = K_1 x \tag{3–117}$$

The force developed by the power piston is $A\,\Delta p$, and this force is applied to the load elements. Thus

$$m \frac{d^2 y}{dt^2} + b \frac{dy}{dt} + ky = A\,\Delta p \tag{3–118}$$

Eliminating $\Delta p$ from Equations (3–117) and (3–118) results in

$$\frac{\varrho V m}{KA} \frac{d^3 y}{dt^3} + \left[\frac{\varrho V b}{KA} + \frac{(L + K_2)m}{A}\right] \frac{d^2 y}{dt^2} + \left[A\varrho + \frac{\varrho V k}{KA} + \frac{(L + K_2)b}{A}\right] \frac{dy}{dt} + \frac{(L + K_2)k}{A}\, y = K_1 x$$

This is a mathematical model of the system relating the valve spool displacement $x$ and the power piston displacement $y$ when the load reactive forces are not negligible.

**B-3-1.** Simplify the block diagram shown in Figure 3–50 and obtain the closed-loop transfer function $C(s)/R(s)$.

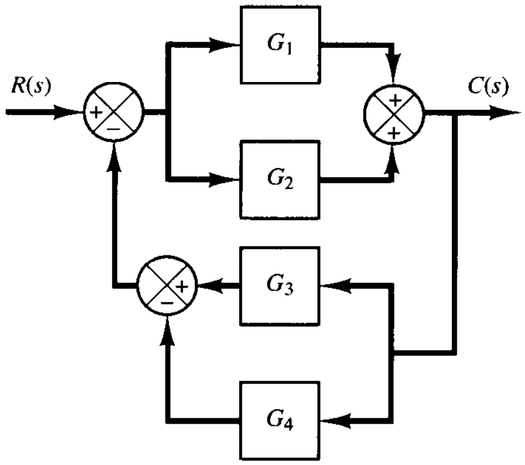



**Figure 3–50** Block diagram of a system.

**B-3-2.** Simplify the block diagram shown in Figure 3–51 and obtain the transfer function $C(s)/R(s)$.

**B-3-3.** Simplify the block diagram shown in Figure 3–52 and obtain the closed-loop transfer function $C(s)/R(s)$.

**B-3-4.** Obtain a state-space representation of the system shown in Figure 3–53.

**B-3-5.** Consider the system described by

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

Derive a state-space representation of the system.

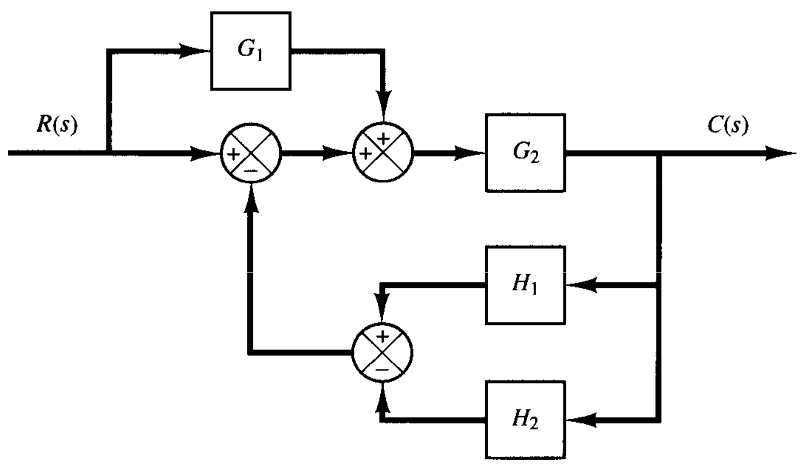



**Figure 3–51** Block diagram of a system.

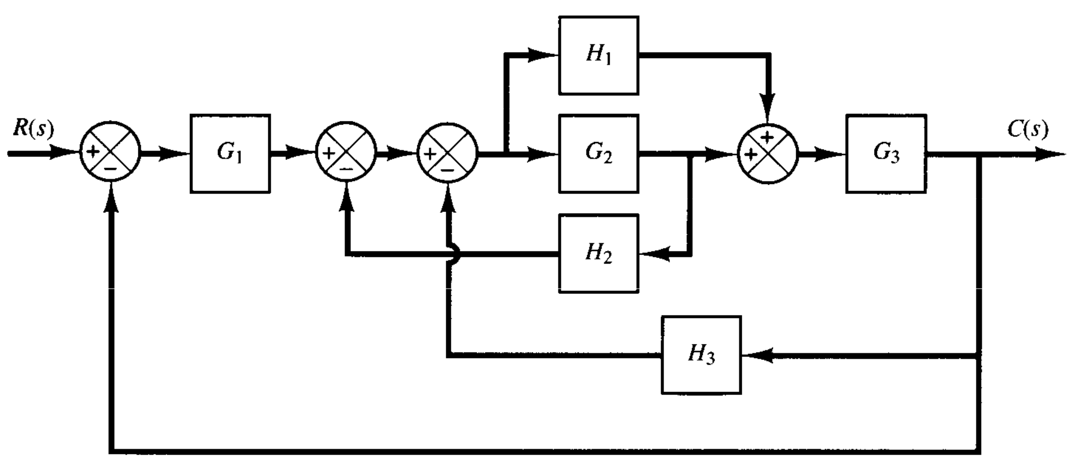



**Figure 3–52** Block diagram of a system.

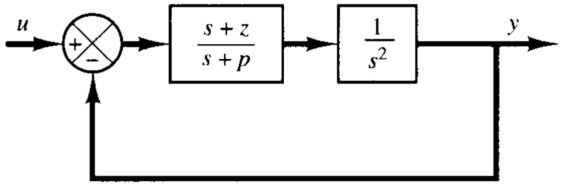



**Figure 3–53** Control system.

**B-3-6.** Consider the system described by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -4 & -1 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u$$

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

Obtain the transfer function of the system.

**B-3-7.** Obtain the transfer function $X_o(s)/X_i(s)$ of each of the three mechanical systems shown in Figure 3–54. In the diagrams, $x_i$ denotes the input displacement and $x_o$ denotes the output displacement. (Each displacement is measured from its equilibrium position.)

**B-3-8.** Obtain mathematical models of the mechanical systems shown in Figures 3–55(a) and (b).

**B-3-9.** Obtain a state-space representation of the mechanical system shown in Figure 3–56, where $u_1$ and $u_2$ are the inputs and $y_1$ and $y_2$ are the outputs.

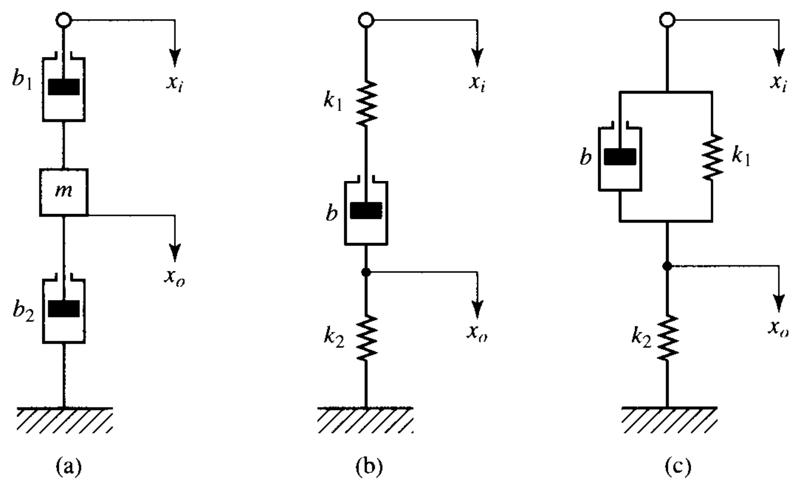



**Figure 3–54** Mechanical systems.

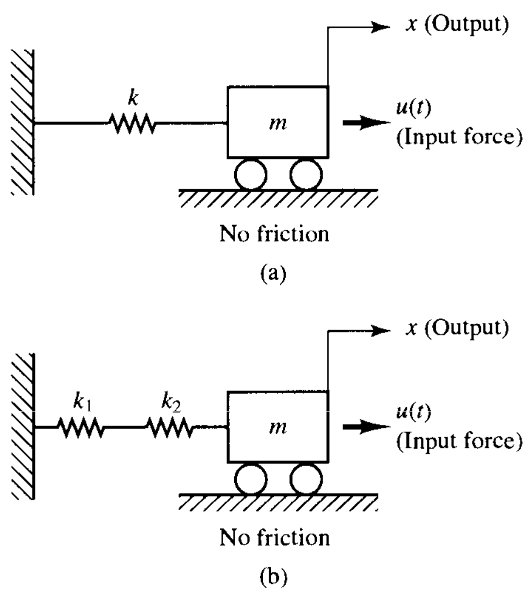



**Figure 3–55** Mechanical systems.

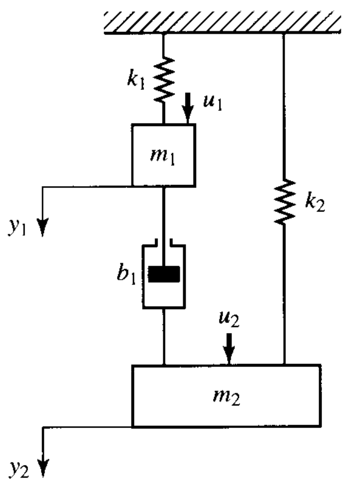



**Figure 3–56**
Mechanical system.

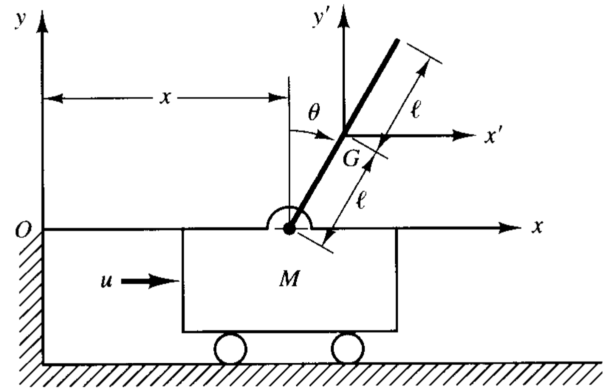



**Figure 3–58** Inverted pendulum system.

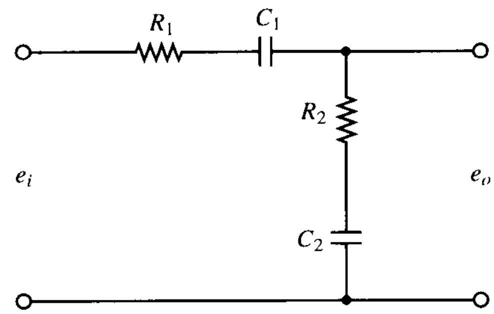



**Figure 3–59** Electrical system.

**B-3-10.** Consider the spring-loaded pendulum system shown in Figure 3–57. Assume that the spring force acting on the pendulum is zero when the pendulum is vertical, or $\theta = 0$. Assume also that the friction involved is negligible and the angle of oscillation $\theta$ is small. Obtain a mathematical model of the system.

**B-3-11.** Referring to Example 3–4, consider the inverted pendulum system shown in Figure 3–58. Assume that the mass of the inverted pendulum is $m$ and is evenly distributed along the length of the rod. (The center of gravity of the pendulum is located at the center of the rod.) Assuming that $\theta$ is small, derive mathematical models for the system in the forms of differential equations, transfer functions, and state-space equations.

**B-3-12.** Derive the transfer function of the electrical system shown in Figure 3–59. Draw a schematic diagram of an analogous mechanical system.

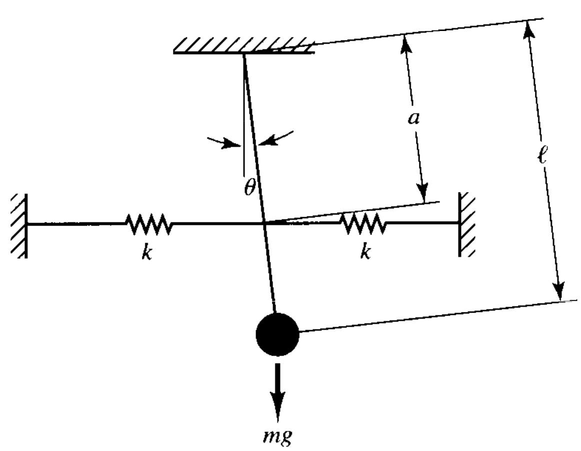



**Figure 3–57** Spring-loaded pendulum system.

**B-3-13.** Consider the liquid-level system shown in Figure 3–60. Assuming that $\bar{H} = 3$ m, $\bar{Q} = 0.02$ m$^3$/sec, and the cross-sectional area of the tank is equal to 5 m$^2$, obtain the time constant of the system at the operating point $(\bar{H}, \bar{Q})$. Assume that the flow through the valve is turbulent.

**B-3-14.** Consider the conical water tank system shown in Figure 3–61. The flow through the valve is turbulent and is related to the head $H$ by

$$Q = 0.005\sqrt{H}$$

where $Q$ is the flow rate measured in m$^3$/sec and $H$ is in meters.

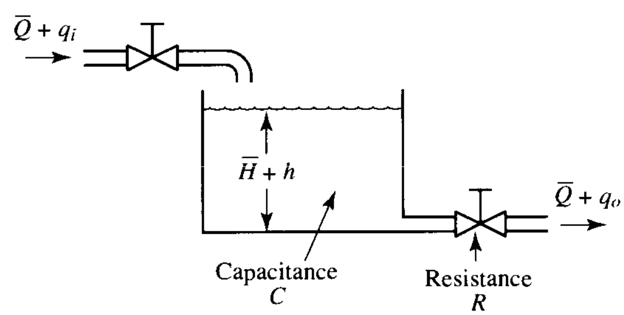



**Figure 3–60** Liquid-level system.

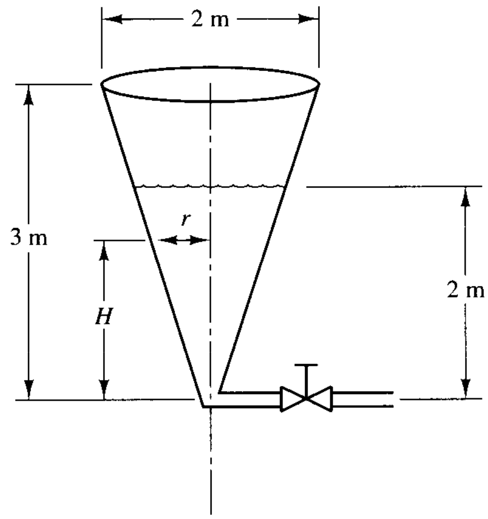



**Figure 3–61** Conical water tank system.

Suppose that the head is 2 m at $t = 0$. What will be the head at $t = 60$ sec?

**B-3-15.** Consider the liquid-level system shown in Figure 3–62. At steady state the inflow rate is $\bar{Q}$ and the outflow rate is also $\bar{Q}$. Assume that at $t = 0$ the inflow rate is changed from $\bar{Q}$ to $\bar{Q} + q_i$, where $q_i$ is a small quantity. The disturbance input is $q_d$, which is also a small quantity. Draw a block diagram of the system and simplify it to obtain $H_2(s)$ as a function of $Q_i(s)$ and $Q_d(s)$, where $H_2(s) = \mathscr{L}[h_2(t)]$, $Q_i(s) = \mathscr{L}[q_i(t)]$, and $Q_d(s) = \mathscr{L}[q_d(t)]$. The capacitances of tanks 1 and 2 are $C_1$ and $C_2$, respectively.

**B-3-16.** A thermocouple has a time constant of 2 sec. A thermal well has a time constant of 30 sec. When the thermocouple is inserted into the well, this temperature-measuring device can be considered a two-capacitance system.

Determine the time constants of the combined thermocouple–thermal well system. Assume that the weight of the thermocouple is 8 g and the weight of the thermal well is 40 g. Assume also that the specific heats of the thermocouple and thermal well are the same.

**B-3-17.** Suppose that the flow rate $Q$ and head $H$ in a liquid-level system are related by

$$Q = 0.002\sqrt{H}$$

Obtain a linearized mathematical model relating the flow rate and head near the steady-state operating point $(\bar{H}, \bar{Q})$, where $\bar{H} = 2.25$ m and $\bar{Q} = 0.003$ m³/sec.

**B-3-18.** Find a linearized equation for

$$y = 0.2x^3$$

about a point $\bar{x} = 2$.

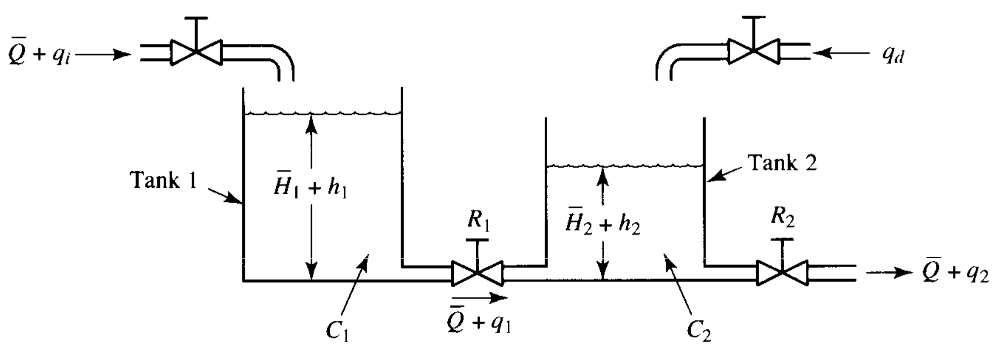



**Figure 3–62** Liquid-level system.

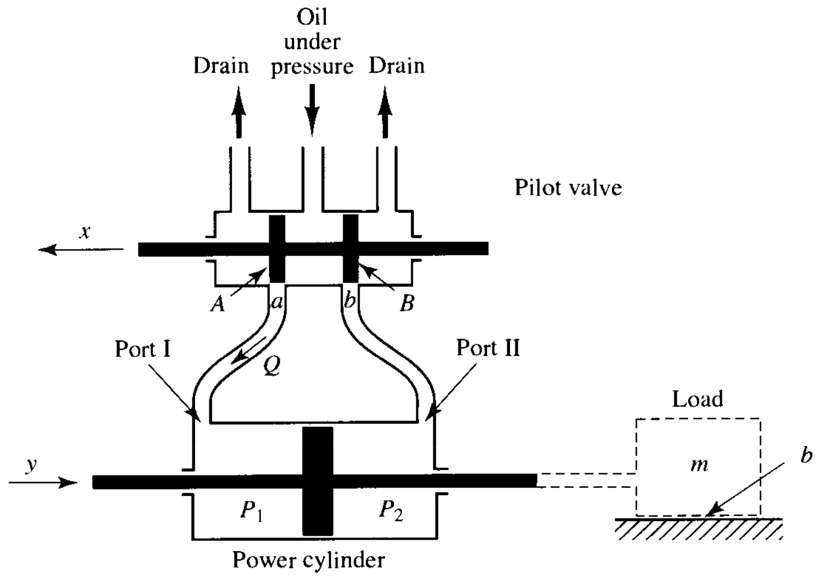


**Figure 3–63** Schematic diagram of a hydraulic servomotor.

**B-3-19.** Linearize the nonlinear equation

$$z = x^2 + 4xy + 6y^2$$

in the region defined by $8 \leq x \leq 10, 2 \leq y \leq 4$.

**B-3-20.** Consider the hydraulic servomotor shown in Figure 3–63. Derive the transfer function $Y(s)/X(s)$. Assume that the inertia force due to the mass of power piston and shaft is negligible compared with the inertia force due to the load mass $m$ and viscous friction force $b\dot{y}$.

# 4

# Transient-Response Analysis

## 4–1 INTRODUCTION

It was stated in Chapter 3 that the first step in analyzing a control system was to derive a mathematical model of the system. Once such a model is obtained, various methods are available for the analysis of system performance.

In practice, the input signal to a control system is not known ahead of time but is random in nature, and the instantaneous input cannot be expressed analytically. Only in some special cases is the input signal known in advance and expressible analytically or by curves, such as in the case of the automatic control of cutting tools.

In analyzing and designing control systems, we must have a basis of comparison of performance of various control systems. This basis may be set up by specifying particular test input signals and by comparing the responses of various systems to these input signals.

Many design criteria are based on such signals or on the response of systems to changes in initial conditions (without any test signals). The use of test signals can be justified because of a correlation existing between the response characteristics of a system to a typical test input signal and the capability of the system to cope with actual input signals.

**Typical test signals.** The commonly used test input signals are those of step functions, ramp functions, acceleration functions, impulse functions, sinusoidal functions, and the like. With these test signals, mathematical and experimental analyses of control systems can be carried out easily since the signals are very simple functions of time.

Which of these typical input signals to use for analyzing system characteristics may be determined by the form of the input that the system will be subjected to most frequently under normal operation. If the inputs to a control system are gradually changing functions of time, then a ramp function of time may be a good test signal. Similarly, if a system is subjected to sudden disturbances, a step function of time may be a good test signal; and for a system subjected to shock inputs, an impulse function may be best. Once a control system is designed on the basis of test signals, the performance of the system in response to actual inputs is generally satisfactory. The use of such test signals enables one to compare the performance of all systems on the same basis.

**Transient response and steady-state response.** The time response of a control system consists of two parts: the transient and the steady-state response. By transient response, we mean that which goes from the initial state to the final state. By steady-state response, we mean the manner in which the system output behaves as $t$ approaches infinity.

**Absolute stability, relative stability, and steady-state error.** In designing a control system, we must be able to predict the dynamic behavior of the system from a knowledge of the components. The most important characteristic of the dynamic behavior of a control system is absolute stability, that is, whether the system is stable or unstable. A control system is in equilibrium if, in the absence of any disturbance or input, the output stays in the same state. A linear time-invariant control system is stable if the output eventually comes back to its equilibrium state when the system is subjected to an initial condition. A linear time-invariant control system is critically stable if oscillations of the output continue forever. It is unstable if the output diverges without bound from its equilibrium state when the system is subjected to an initial condition. Actually, the output of a physical system may increase to a certain extent but may be limited by mechanical "stops," or the system may break down or become nonlinear after the output exceeds a certain magnitude so that the linear differential equations no longer apply.

Important system behavior (other than absolute stability) to which we must give careful consideration includes relative stability and steady-state error. Since a physical control system involves energy storage, the output of the system, when subjected to an input, cannot follow the input immediately but exhibits a transient response before a steady state can be reached. The transient response of a practical control system often exhibits damped oscillations before reaching a steady state. If the output of a system at steady state does not exactly agree with the input, the system is said to have steady-state error. This error is indicative of the accuracy of the system. In analyzing a control system, we must examine transient-response behavior and steady-state behavior.

**Outline of the chapter.** This chapter is concerned with system responses to aperiodic signals (such as step, ramp, acceleration, and impulse functions of time). The outline of the chapter is as follows: Section 4–1 has presented introductory material for the chapter. Section 4–2 treats the response of first-order systems to aperiodic inputs. Section 4–3 deals with the transient response of the second-order systems. Detailed analyses of the step response, ramp response, and impulse response of the second-order systems are presented. (The transient response analysis of higher-order systems is discussed in Chapter 5.) Section 4–4 gives an introduction to the MATLAB approach to the solution of transient response. Section 4–5 presents an example of a transient-response problem solved with MATLAB.

## 4–2 FIRST-ORDER SYSTEMS

Consider the first-order system shown in Figure 4–1(a). Physically, this system may represent an $RC$ circuit, thermal system, or the like. A simplified block diagram is shown in Figure 4–1(b). The input–output relationship is given by

$$\frac{C(s)}{R(s)} = \frac{1}{Ts + 1} \tag{4–1}$$

In the following, we shall analyze the system responses to such inputs as the unit-step, unit-ramp, and unit-impulse functions. The initial conditions are assumed to be zero.

Note that all systems having the same transfer function will exhibit the same output in response to the same input. For any given physical system, the mathematical response can be given a physical interpretation.

**Unit-step response of first-order systems.** Since the Laplace transform of the unit-step function is $1/s$, substituting $R(s) = 1/s$ into Equation (4–1), we obtain

$$C(s) = \frac{1}{Ts + 1}\frac{1}{s}$$

Expanding $C(s)$ into partial fractions gives

$$C(s) = \frac{1}{s} - \frac{T}{Ts + 1} = \frac{1}{s} - \frac{1}{s + (1/T)} \tag{4–2}$$

Taking the inverse Laplace transform of Equation (4–2), we obtain

$$c(t) = 1 - e^{-t/T}, \qquad \text{for } t \geq 0 \tag{4–3}$$

Equation (4–3) states that initially the output $c(t)$ is zero and finally it becomes unity. One important characteristic of such an exponential response curve $c(t)$ is that at $t = T$ the value of $c(t)$ is 0.632, or the response $c(t)$ has reached 63.2% of its total change. This may be easily seen by substituting $t = T$ in $c(t)$. That is,

$$c(T) = 1 - e^{-1} = 0.632$$

Note that the smaller the time constant $T$, the faster the system response. Another important characteristic of the exponential response curve is that the slope of the tangent line at $t = 0$ is $1/T$, since

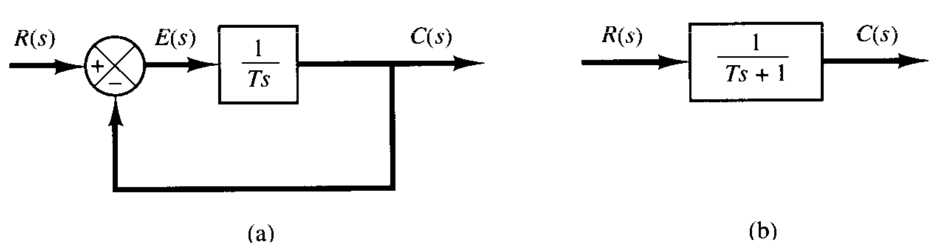



**Figure 4–1**
(a) Block diagram of a first-order system; (b) simplified block diagram.

$$\frac{dc}{dt} = \frac{1}{T} e^{-t/T} \bigg|_{t=0} = \frac{1}{T} \tag{4–4}$$

The output would reach the final value at $t = T$ if it maintained its initial speed of response. From Equation (4–4) we see that the slope of the response curve $c(t)$ decreases monotonically from $1/T$ at $t = 0$ to zero at $t = \infty$.

The exponential response curve $c(t)$ given by Equation (4–3) is shown in Figure 4–2. In one time constant, the exponential response curve has gone from 0 to 63.2% of the final value. In two time constants, the response reaches 86.5% of the final value. At $t = 3T$, $4T$, and $5T$, the response reaches 95%, 98.2%, and 99.3%, respectively, of the final value. Thus, for $t \geq 4T$, the response remains within 2% of the final value. As seen from Equation (4–3), the steady state is reached mathematically only after an infinite time. In practice, however, a reasonable estimate of the response time is the length of time the response curve needs to reach the 2% line of the final value, or four time constants.

Consider the system shown in Figure 4–3. To determine experimentally whether or not the system is of first order, plot the curve $\log |c(t) - c(\infty)|$, where $c(t)$ is the system output, as a function of $t$. If the curve turns out to be a straight line, the system is of first order. The time constant $T$ can be read from the graph as the time $T$ that satisfies the following equation:

$$c(T) - c(\infty) = 0.368\,[c(0) - c(\infty)]$$

Note that instead of plotting $\log |c(t) - c(\infty)|$ versus $t$ it is convenient to plot $|c(t) - c(\infty)|/|c(0) - c(\infty)|$ versus $t$ on semilog paper, as shown in Figure 4–4.

**Unit-ramp response of first-order systems.** Since the Laplace transform of the unit-ramp function is $1/s^2$, we obtain the output of the system of Figure 4–1(a) as

$$C(s) = \frac{1}{Ts + 1} \frac{1}{s^2}$$

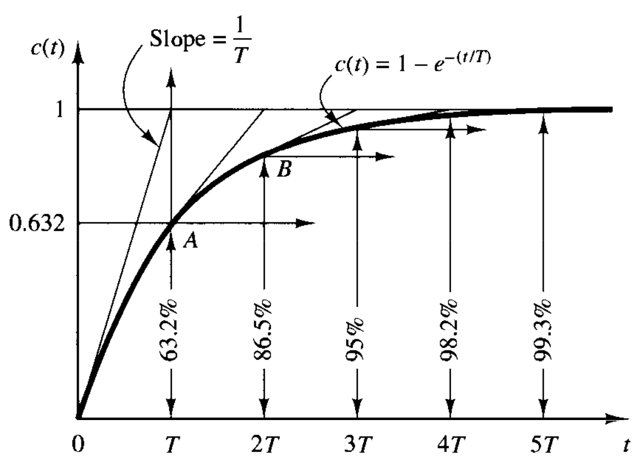



**Figure 4–2**
Exponential response curve.

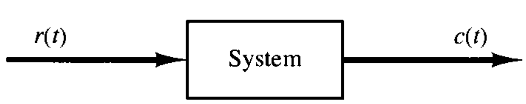



**Figure 4–3**
A general system.

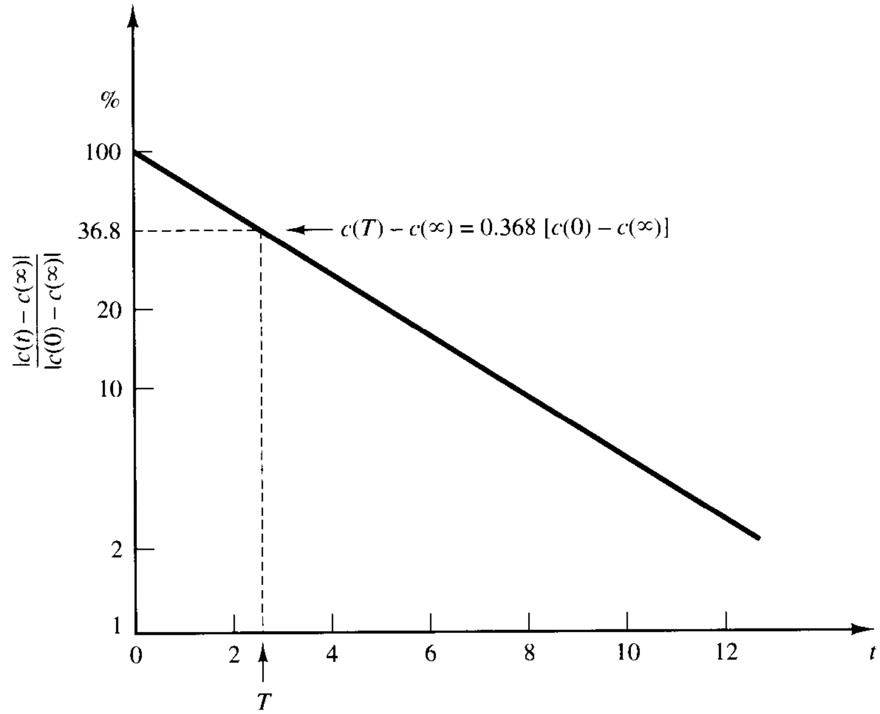



**Figure 4–4**
Plot of $|c(t) - c(\infty)|$ / $|c(0) - c(\infty)|$ versus $t$ on semilog paper.

Expanding $C(s)$ into partial fractions gives

$$C(s) = \frac{1}{s^2} - \frac{T}{s} + \frac{T^2}{Ts + 1} \tag{4–5}$$

Taking the inverse Laplace transform of Equation (4–5), we obtain

$$c(t) = t - T + Te^{-t/T}, \qquad \text{for } t \geq 0$$

The error signal $e(t)$ is then

$$\begin{aligned} e(t) &= r(t) - c(t) \\ &= T(1 - e^{-t/T}) \end{aligned}$$

As $t$ approaches infinity, $e^{-t/T}$ approaches zero, and thus the error signal $e(t)$ approaches $T$ or

$$e(\infty) = T$$

The unit-ramp input and the system output are shown in Figure 4–5. The error in following the unit-ramp input is equal to $T$ for sufficiently large $t$. The smaller the time constant $T$, the smaller the steady-state error in following the ramp input.

**Unit-impulse response of first-order systems.** For the unit-impulse input, $R(s) = 1$ and the output of the system of Figure 4–1(a) can be obtained as

$$C(s) = \frac{1}{Ts + 1}$$

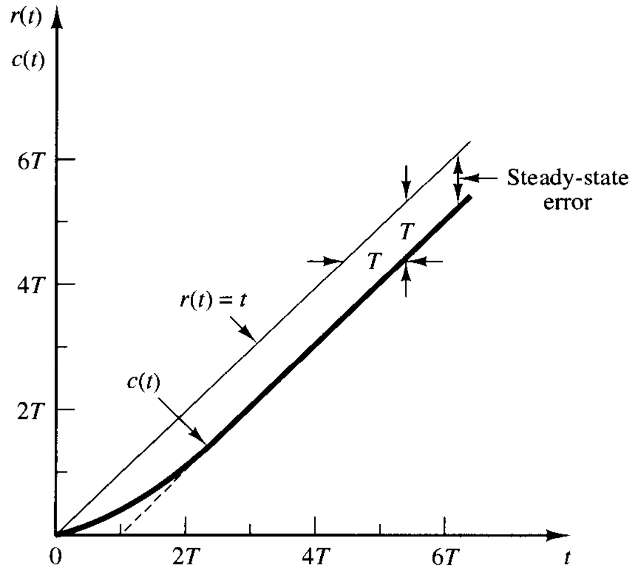



**Figure 4–5**
Unit-ramp response of the system shown in Figure 4–1(a).

or

$$c(t) = \frac{1}{T} e^{-t/T}, \qquad \text{for } t \geq 0 \tag{4–6}$$

The response curve given by Equation (4–6) is shown in Figure 4–6.

**An important property of linear time-invariant systems.** In the analysis above, it has been shown that for the unit-ramp input the output $c(t)$ is

$$c(t) = t - T + Te^{-t/T}, \qquad \text{for } t \geq 0$$

For the unit-step input, which is the derivative of unit-ramp input, the output $c(t)$ is

$$c(t) = 1 - e^{-t/T}, \qquad \text{for } t \geq 0$$

Finally, for the unit-impulse input, which is the derivative of unit-step input, the output $c(t)$ is

$$c(t) = \frac{1}{T} e^{-t/T}, \qquad \text{for } t \geq 0$$

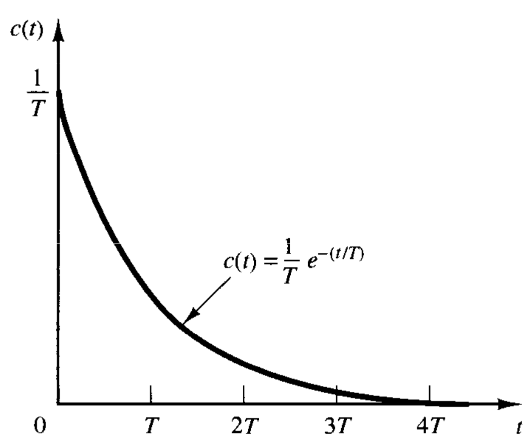



**Figure 4–6**
Unit-impulse response of the system shown in Figure 4–1(a).

Comparison of the system responses to these three inputs clearly indicates that the response to the derivative of an input signal can be obtained by differentiating the response of the system to the original signal. It can also be seen that the response to the integral of the original signal can be obtained by integrating the response of the system to the original signal and by determining the integration constants from the zero output initial condition. This is a property of linear time-invariant systems. Linear time-varying systems and nonlinear systems do not possess this property.

**EXAMPLE 4–1** Consider the liquid-level control system shown in Figure 4–7(a). (The controller is assumed to be a proportional controller; that is, the output of the controller is proportional to the input of the controller.) We assume that all the variables, $r$, $q_i$, $h$, and $q_o$ are measured from their respective steady-state values $\bar{R}$, $\bar{Q}$, $\bar{H}$, and $\bar{Q}$. We also assume that the magnitudes of the variables $r$, $q_i$, $h$, and $q_o$ are sufficiently small so that the system can be approximated by a linear mathematical model.

Referring to Section 3–8, we can obtain the transfer function of the liquid-level system as

$$\frac{H(s)}{Q_i(s)} = \frac{R}{RCs + 1}$$

Since the controller is a proportional controller, the change in inflow $q_i$ is proportional to the actuating error $e$ so that $q_i = K_p K_v e$, where $K_p$ is the gain of the controller and $K_v$ is the gain of the control valve. In terms of Laplace-transformed quantities,

$$Q_i(s) = K_p K_v E(s)$$

A block diagram of this system is shown in Figure 4–7(b). A simplified block diagram is given in Figure 4–7(c), where $X(s) = (1/K_b)R(s)$, $K = K_p K_v R K_b$, and $T = RC$.

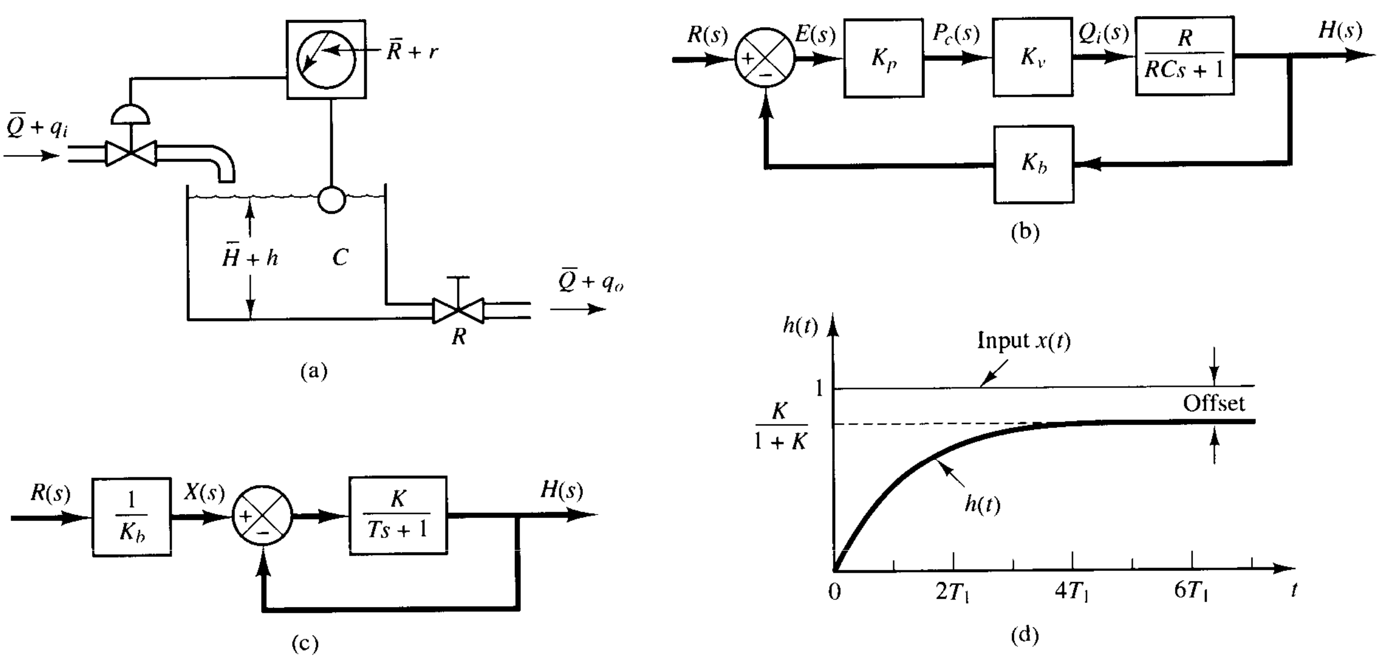



**Figure 4–7**
(a) Liquid-level control system; (b) block diagram; (c) simplified block diagram; (d) curve $h(t)$ versus $t$.

In what follows we shall investigate the response $h(t)$ to a change in the reference input. We shall assume a unit-step change in $x(t)$, where $x(t) = (1/K_b)r(t)$. The closed-loop transfer function between $H(s)$ and $X(s)$ is given by

$$\frac{H(s)}{X(s)} = \frac{K}{Ts + 1 + K} \tag{4-7}$$

Since the Laplace transform of the unit-step function is $1/s$, substituting $X(s) = 1/s$ into Equation (4–7) gives

$$H(s) = \frac{K}{Ts + 1 + K}\frac{1}{s}$$

Expanding $H(s)$ into partial fractions gives

$$H(s) = \frac{K}{1 + K}\frac{1}{s} - \frac{K}{1 + K}\frac{1}{s + (1 + K)/T}$$

Taking the inverse Laplace transforms of both sides of this last equation, we obtain the following time solution $h(t)$:

$$h(t) = \frac{K}{1 + K}(1 - e^{-t/T_1}), \qquad \text{for } t \geq 0 \tag{4-8}$$

where

$$T_1 = \frac{T}{1 + K}$$

The response curve $h(t)$ is plotted in Figure 4–7(d). From Equation (4–8), notice that the time constant $T_1$ of the closed-loop system is different from the time constant $T$ of the feed-forward block.

From Equation (4–8), we see that as $t$ approaches infinity the value of $h(t)$ approaches $K/(1 + K)$, or

$$h(\infty) = \frac{K}{1 + K}$$

Since $x(\infty) = 1$, there is a steady-state error of $1/(1 + K)$. Such an error is called *offset*. The value of the offset becomes smaller as the gain $K$ becomes larger.

Offset is a characteristic of the proportional control of a plant whose transfer function does not possess an integrating element. (In such a case we need a nonzero error to provide a nonzero output.) To eliminate such offset, we must add integral control action. (Refer to Section 5–3.)

## 4–3 SECOND-ORDER SYSTEMS

In this section, we shall obtain the response of a typical second-order control system to a step input, ramp input, and impulse input. Here we consider a dc servomotor as an example of a second-order system. Conventional dc motors use mechanical brushes and commutators that require regular maintenance. Due to improvements that have been made in the brushes and commutators, however, many dc motors used in servo systems can be operated almost maintenance free. Some dc motors use electronic commutation. They are called brushless dc motors.

**DC servomotors.** There are many types of dc motors in use in industries. DC motors that are used in servo systems are called dc servomotors. In dc servomotors, the rotor inertias have been made very small, with the result that motors with very high torque-to-inertia ratios are commercially available. Some dc servomotors have extremely small time constants. DC servomotors with relatively small power ratings are used in instruments and computer-related equipment such as disk drives, tape drives, printers, and word processors. DC servomotors with medium and large power ratings are used in robot systems, numerically controlled milling machines, and so on.

In dc servomotors, the field windings may be connected in series with the armature or the field windings may be separate from the armature. (That is, the magnetic field is produced by a separate circuit.) In the latter case, where the field is excited separately, the magnetic flux is independent of the armature current. In some dc servomotors, the magnetic field is produced by a permanent magnet and, therefore, the magnetic flux is constant. Such dc servomotors are called permanent magnet dc servomotors. DC servomotors with separately excited fields, as well as permanent magnet dc servomotors, can be controlled by the armature current. Such a scheme to control the output of the dc servomotor by the armature current is called armature control of dc servomotors.

In the case where the armature current is maintained constant and the speed is controlled by the field voltage, the dc motor is called a field-controlled dc motor. (Some speed control systems use field-controlled dc motors.) The requirement of constant armature current, however, is a serious disadvantage. (Providing a constant current source is much more difficult than providing a constant voltage source.) The time constants of the field-controlled dc motor are generally large compared with the time constants of a comparable armature-controlled dc motor.

A dc servomotor may also be driven by an electronic motion controller, frequently called a servodriver, as a motor–driver combination. The servodriver controls the motion of a dc servomotor and operates in various modes. Some of the features are point-to-point positioning, velocity profiling, and programmable acceleration. The use of an electronic motion controller using a pulse-width-modulated driver to control a dc servomotor is frequently seen in robot control systems, numerical control systems, and other position and/or speed control systems.

In what follows we shall discuss armature control of dc servomotors.

**A servo system.** Consider the servo system shown in Figure 4–8(a). The objective of this system is to control the position of the mechanical load in accordance with the reference position. The operation of this system is as follows: A pair of potentiometers acts as an error-measuring device. They convert the input and output positions into proportional electric signals. The command input signal determines the angular position $r$ of the wiper arm of the input potentiometer. The angular position $r$ is the reference input to the system, and the electric potential of the arm is proportional to the angular position of the arm. The output shaft position determines the angular position $c$ of the wiper arm of the output potentiometer. The difference between the input angular position $r$ and the output angular position $c$ is the error signal $e$, or

$$e = r - c$$

The potential difference $e_r - e_c = e_v$ is the error voltage, where $e_r$ is proportional to $r$ and $e_c$ is proportional to $c$; that is, $e_r = K_0 r$ and $e_c = K_0 c$, where $K_0$ is a proportionality

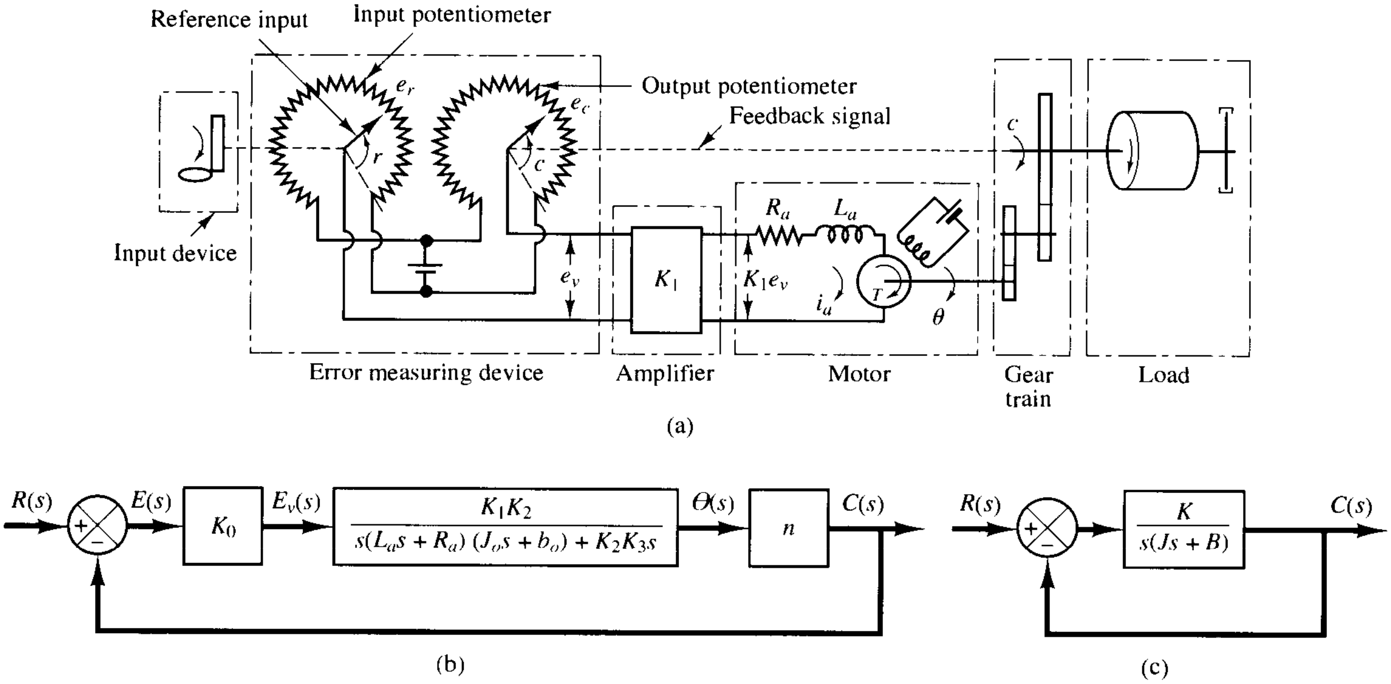



**Figure 4–8**
(a) Schematic diagram of servo system; (b) block diagram for the system; (c) simplified block diagram.

constant. The error voltage that appears at the potentiometer terminals is amplified by the amplifier whose gain constant is $K_1$. The output voltage of this amplifier is applied to the armature circuit of the dc motor. (The amplifier must have very high input impedance because the potentiometers are essentially high impedance circuits and do not tolerate current drain. At the same time, the amplifier must have low output impedance since it feeds into the armature circuit of the motor.) A fixed voltage is applied to the field winding. If an error exists, the motor develops a torque to rotate the output load in such a way as to reduce the error to zero. For constant field current, the torque developed by the motor is

$$T = K_2 i_a$$

where $K_2$ is the motor torque constant and $i_a$ is the armature current.

Notice that if the sign of the current $i_a$ is reversed the sign of the torque $T$ will be reversed, which will result in the reversion of the direction of rotor rotation.

When the armature is rotating, a voltage proportional to the product of the flux and angular velocity is induced in the armature. For a constant flux, the induced voltage $e_b$ is directly proportional to the angular velocity $d\theta/dt$, or

$$e_b = K_3 \frac{d\theta}{dt} \tag{4–9}$$

where $e_b$ is the back emf, $K_3$ is the back emf constant of the motor, and $\theta$ is the angular displacement of the motor shaft.

The speed of an armature-controlled dc servomotor is controlled by the armature voltage $e_a$. (The armature voltage $e_a = K_1 e_v$ is the output of the amplifier.) The differential equation for the armature circuit is

$$L_a \frac{di_a}{dt} + R_a i_a + e_b = e_a$$

or

$$L_a \frac{di_a}{dt} + R_a i_a + K_3 \frac{d\theta}{dt} = K_1 e_v \tag{4-10}$$

The equation for torque equilibrium is

$$J_0 \frac{d^2\theta}{dt^2} + b_0 \frac{d\theta}{dt} = T = K_2 i_a \tag{4-11}$$

where $J_0$ is the inertia of the combination of the motor, load, and gear train referred to the motor shaft and $b_0$ is the viscous-friction coefficient of the combination of the motor, load, and gear train referred to the motor shaft. The transfer function between the motor shaft displacement and the error voltage is obtained from Equations (4–10) and (4–11) as follows:

$$\frac{\Theta(s)}{E_v(s)} = \frac{K_1 K_2}{s(L_a s + R_a)(J_0 s + b_0) + K_2 K_3 s} \tag{4-12}$$

where $\Theta(s) = \mathscr{L}[\theta(t)]$ and $E_v(s) = \mathscr{L}[e_v(t)]$. We assume that the gear ratio of the gear train is such that the output shaft rotates $n$ times for each revolution of the motor shaft. Thus,

$$C(s) = n\Theta(s) \tag{4-13}$$

where $C(s) = \mathscr{L}[c(t)]$ and $c(t)$ is the angular displacement of the output shaft. The relationship among $E_v(s)$, $R(s)$, and $C(s)$ is

$$E_v(s) = K_0[R(s) - C(s)] = K_0 E(s) \tag{4-14}$$

where $R(s) = \mathscr{L}[r(t)]$. The block diagram of this system can be constructed from Equations (4–12), (4–13), and (4–14), as shown in Figure 4–8(b). The transfer function in the feedforward path of this system is

$$G(s) = \frac{C(s)}{\Theta(s)} \frac{\Theta(s)}{E_v(s)} \frac{E_v(s)}{E(s)} = \frac{K_0 K_1 K_2 n}{s[(L_a s + R_a)(J_0 s + b_0) + K_2 K_3]}$$

Since $L_a$ is usually small, it can be neglected, and the transfer function $G(s)$ in the feedforward path becomes

$$\begin{aligned} G(s) &= \frac{K_0 K_1 K_2 n}{s[R_a(J_0 s + b_0) + K_2 K_3]} \\ &= \frac{K_0 K_1 K_2 n / R_a}{J_0 s^2 + \left(b_0 + \dfrac{K_2 K_3}{R_a}\right)s} \end{aligned} \tag{4-15}$$

The term $[b_0 + (K_2K_3/R_a)]s$ indicates that the back emf of the motor effectively increases the viscous friction of the system. The inertia $J_0$ and viscous friction coefficient $b_0 + (K_2K_3/R_a)$ are referred to the motor shaft. When $J_0$ and $b_0 + (K_2K_3/R_a)$ are multiplied by $1/n^2$, the inertia and viscous-friction coefficient are expressed in terms of the output shaft. Introducing new parameters defined by

$J = J_0/n^2 =$ moment of inertia referred to the output shaft

$B = [b_0 + (K_2K_3/R_a)]/n^2 =$ viscous-friction coefficient referred to the output shaft

$K = K_0K_1K_2/nR_a$

the transfer function $G(s)$ given by Equation (4–15) can be simplified, yielding

$$G(s) = \frac{K}{Js^2 + Bs}$$

or

$$G(s) = \frac{K_m}{s(T_m s + 1)} \tag{4–16}$$

where

$$K_m = \frac{K}{B}, \qquad T_m = \frac{J}{B} = \frac{R_a J_0}{R_a b_0 + K_2 K_3}$$

The block diagram of the system shown in Figure 4–8(b) can thus be simplified as shown in Figure 4–8(c).

In the following, we shall investigate the dynamic responses of this system to unit-step, unit-ramp, and unit-impulse inputs.

From Equations (4–15) and (4–16), it can be seen that the transfer functions involve the term $1/s$. Thus, this system possesses an integrating property. In Equation (4–16), notice that the time constant of the motor is smaller for a smaller $R_a$ and smaller $J_0$. With small $J_0$, as the resistance $R_a$ is reduced, the motor time constant approaches zero, and the motor acts as an ideal integrator.

**Effect of load on servomotor dynamics.** Most important among the characteristics of the servomotor is the maximum acceleration obtainable. For a given available torque, the rotor moment of inertia must be a minimum. Since the servomotor operates under continuously varying conditions, acceleration and deceleration of the rotor occur from time to time. The servomotor must be able to absorb mechanical energy as well as to generate it. The performance of the servomotor when used as a brake should be satisfactory.

Let $J_m$ and $b_m$ be, respectively, the moment of inertia and viscous-friction coefficient of the rotor, and let $J_L$ and $b_L$ be, respectively, the moment of inertia and viscous-friction coefficient of the load on the output shaft. Assume that the moment of inertia and viscous-friction coefficient of the gear train are either negligible or included in $J_L$ and $b_L$, respectively. Then, the equivalent moment of inertia $J_{eq}$ referred to the motor shaft and equivalent viscous-friction coefficient $b_{eq}$ referred to the motor shaft can be written as (for details, refer to Problem A-4-4)

$$J_{eq} = J_m + n^2 J_L$$
$$b_{eq} = b_m + n^2 f_L$$

where $n(n < 1)$ is the gear ratio between the motor and load. If the gear ratio $n$ is small and $J_m \gg n^2 J_L$, then the moment of inertia of the load referred to the motor shaft is negligible with respect to the rotor moment of inertia. A similar argument applies to the load friction. In general, when the gear ratio $n$ is small, the transfer function of the electric servomotor may be obtained without taking into account the load moment of inertia and friction. If neither $J_m$ nor $n^2 J_L$ is negligibly small compared with the other, however, then the equivalent moment of inertia $J_{eq}$ must be used for evaluating the transfer function of the motor–load combination.

**Step response of second-order systems.** The closed-loop transfer function of the system shown in Figure 4–8(c) is

$$\frac{C(s)}{R(s)} = \frac{K}{Js^2 + Bs + K} \tag{4–17}$$

which can be rewritten as

$$\frac{C(s)}{R(s)} = \frac{\frac{K}{J}}{\left[s + \frac{B}{2J} + \sqrt{\left(\frac{B}{2J}\right)^2 - \frac{K}{J}}\right]\left[s + \frac{B}{2J} - \sqrt{\left(\frac{B}{2J}\right)^2 - \frac{K}{J}}\right]} \tag{4–18}$$

The closed-loop poles are complex if $B^2 - 4JK < 0$, and they are real if $B^2 - 4JK \geq 0$. In transient-response analysis, it is convenient to write

$$\frac{K}{J} = \omega_n^2, \qquad \frac{B}{J} = 2\zeta\omega_n = 2\sigma$$

where $\sigma$ is called the *attenuation;* $\omega_n$, the *undamped natural frequency;* and $\zeta$, the *damping ratio* of the system. The damping ratio $\zeta$ is the ratio of the actual damping $B$ to the critical damping $B_c = 2\sqrt{JK}$ or

$$\zeta = \frac{B}{B_c} = \frac{B}{2\sqrt{JK}}$$

In terms of $\zeta$ and $\omega_n$, the system shown in Figure 4–8(c) can be modified to that shown in Figure 4–9, and the closed-loop transfer function $C(s)/R(s)$ given by Equation (4–18) can be written

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{4–19}$$

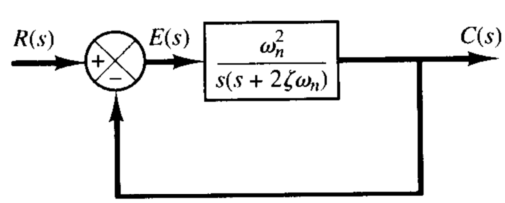



**Figure 4–9**
Second-order system.

The dynamic behavior of the second-order system can then be described in terms of two parameters $\zeta$ and $\omega_n$. If $0 < \zeta < 1$, the closed-loop poles are complex conjugates and lie in the left-half $s$ plane. The system is then called underdamped, and the transient response is oscillatory. If $\zeta = 1$, the system is called critically damped. Overdamped systems correspond to $\zeta > 1$. The transient response of critically damped and overdamped systems do not oscillate. If $\zeta = 0$, the transient response does not die out.

We shall now solve for the response of the system shown in Figure 4–9 to a unit-step input. We shall consider three different cases: the underdamped ($0 < \zeta < 1$), critically damped ($\zeta = 1$), and overdamped ($\zeta > 1$) cases.

(1) Underdamped case ($0 < \zeta < 1$): In this case, $C(s)/R(s)$ can be written

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{(s + \zeta\omega_n + j\omega_d)(s + \zeta\omega_n - j\omega_d)}$$

where $\omega_d = \omega_n\sqrt{1 - \zeta^2}$. The frequency $\omega_d$ is called the *damped natural frequency*. For a unit-step input, $C(s)$ can be written

$$C(s) = \frac{\omega_n^2}{(s^2 + 2\zeta\omega_n s + \omega_n^2)s} \tag{4–20}$$

The inverse Laplace transform of Equation (4–20) can be obtained easily if $C(s)$ is written in the following form:

$$\begin{aligned} C(s) &= \frac{1}{s} - \frac{s + 2\zeta\omega_n}{s^2 + 2\zeta\omega_n s + \omega_n^2} \\ &= \frac{1}{s} - \frac{s + \zeta\omega_n}{(s + \zeta\omega_n)^2 + \omega_d^2} - \frac{\zeta\omega_n}{(s + \zeta\omega_n)^2 + \omega_d^2} \end{aligned}$$

In Chapter 2 it was shown that

$$\mathcal{L}^{-1}\left[\frac{s + \zeta\omega_n}{(s + \zeta\omega_n)^2 + \omega_d^2}\right] = e^{-\zeta\omega_n t}\cos\omega_d t$$

$$\mathcal{L}^{-1}\left[\frac{\omega_d}{(s + \zeta\omega_n)^2 + \omega_d^2}\right] = e^{-\zeta\omega_n t}\sin\omega_d t$$

Hence the inverse Laplace transform of Equation (4–20) is obtained as

$$\begin{aligned} \mathcal{L}^{-1}[C(s)] &= c(t) \\ &= 1 - e^{-\zeta\omega_n t}\left(\cos\omega_d t + \frac{\zeta}{\sqrt{1 - \zeta^2}}\sin\omega_d t\right) \\ &= 1 - \frac{e^{-\zeta\omega_n t}}{\sqrt{1 - \zeta^2}}\sin\left(\omega_d t + \tan^{-1}\frac{\sqrt{1 - \zeta^2}}{\zeta}\right), \qquad \text{for } t \geq 0 \end{aligned} \tag{4–21}$$

This result can be obtained directly by using a table of Laplace transforms. From Equation (4–21), it can be seen that the frequency of transient oscillation is the damped natural frequency $\omega_d$ and thus varies with the damping ratio $\zeta$. The error signal for this system is the difference between the input and output and is

$$
\begin{aligned}
e(t) &= r(t) - c(t) \\
&= e^{-\zeta\omega_n t}\left(\cos \omega_d t + \frac{\zeta}{\sqrt{1-\zeta^2}} \sin \omega_d t\right) \qquad \text{for } t \geq 0
\end{aligned}
$$

This error signal exhibits a damped sinusoidal oscillation. At steady state, or at $t = \infty$, no error exists between the input and output.

If the damping ratio $\zeta$ is equal to zero, the response becomes undamped and oscillations continue indefinitely. The response $c(t)$ for the zero damping case may be obtained by substituting $\zeta = 0$ in Equation (4–21), yielding

$$
c(t) = 1 - \cos \omega_n t, \qquad \text{for } t \geq 0 \tag{4–22}
$$

Thus, from Equation (4–22), we see that $\omega_n$ represents the undamped natural frequency of the system. That is, $\omega_n$ is that frequency at which the system would oscillate if the damping were decreased to zero. If the linear system has any amount of damping, the undamped natural frequency cannot be observed experimentally. The frequency that may be observed is the damped natural frequency $\omega_d$, which is equal to $\omega_n\sqrt{1-\zeta^2}$. This frequency is always lower than the undamped natural frequency. An increase in $\zeta$ would reduce the damped natural frequency $\omega_d$. If $\zeta$ is increased beyond unity, the response becomes overdamped and will not oscillate.

(2) Critically damped case ($\zeta = 1$): If the two poles of $C(s)/R(s)$ are nearly equal, the system may be approximated by a critically damped one.

For a unit-step input, $R(s) = 1/s$ and $C(s)$ can be written

$$
C(s) = \frac{\omega_n^2}{(s + \omega_n)^2 s} \tag{4–23}
$$

The inverse Laplace transform of Equation (4–23) may be found as

$$
c(t) = 1 - e^{-\omega_n t}(1 + \omega_n t), \qquad \text{for } t \geq 0 \tag{4–24}
$$

This result can be obtained by letting $\zeta$ approach unity in Equation (4–21) and by using the following limit:

$$
\lim_{\zeta \to 1} \frac{\sin \omega_d t}{\sqrt{1-\zeta^2}} = \lim_{\zeta \to 1} \frac{\sin \omega_n \sqrt{1-\zeta^2}\, t}{\sqrt{1-\zeta^2}} = \omega_n t
$$

(3) Overdamped case ($\zeta > 1$): In this case, the two poles of $C(s)/R(s)$ are negative real and unequal. For a unit-step input, $R(s) = 1/s$ and $C(s)$ can be written

$$
C(s) = \frac{\omega_n^2}{(s + \zeta\omega_n + \omega_n\sqrt{\zeta^2-1})(s + \zeta\omega_n - \omega_n\sqrt{\zeta^2-1})s} \tag{4–25}
$$

The inverse Laplace transform of Equation (4–25) is

$$
\begin{aligned}
c(t) = 1 &+ \frac{1}{2\sqrt{\zeta^2-1}(\zeta + \sqrt{\zeta^2-1})} e^{-(\zeta+\sqrt{\zeta^2-1})\omega_n t} \\
&- \frac{1}{2\sqrt{\zeta^2-1}(\zeta - \sqrt{\zeta^2-1})} e^{-(\zeta-\sqrt{\zeta^2-1})\omega_n t}
\end{aligned}
$$

$$= 1 + \frac{\omega_n}{2\sqrt{\zeta^2 - 1}}\left(\frac{e^{-s_1 t}}{s_1} - \frac{e^{-s_2 t}}{s_2}\right), \qquad \text{for } t \geq 0 \tag{4-26}$$

where $s_1 = (\zeta + \sqrt{\zeta^2 - 1})\omega_n$ and $s_2 = (\zeta - \sqrt{\zeta^2 - 1})\omega_n$. Thus, the response $c(t)$ includes two decaying exponential terms.

When $\zeta$ is appreciably greater than unity, one of the two decaying exponentials decreases much faster than the other, so the faster decaying exponential term (which corresponds to a smaller time constant) may be neglected. That is, if $-s_2$ is located very much closer to the $j\omega$ axis than $-s_1$ (which means $|s_2| \ll |s_1|$), then for an approximate solution we may neglect $-s_1$. This is permissible because the effect of $-s_1$ on the response is much smaller than that of $-s_2$, since the term involving $s_1$ in Equation (4–26) decays much faster than the term involving $s_2$. Once the faster decaying exponential term has disappeared, the response is similar to that of a first-order system, and $C(s)/R(s)$ may be approximated by

$$\frac{C(s)}{R(s)} = \frac{\zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1}}{s + \zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1}} = \frac{s_2}{s + s_2}$$

This approximate form is a direct consequence of the fact that the initial values and final values of both the original $C(s)/R(s)$ and the approximate one agree with each other.

With the approximate transfer function $C(s)/R(s)$, the unit-step response can be obtained as

$$C(s) = \frac{\zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1}}{(s + \zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1})s}$$

The time response $c(t)$ is then

$$c(t) = 1 - e^{-(\zeta - \sqrt{\zeta^2 - 1})\omega_n t}, \qquad \text{for } t \geq 0$$

This gives an approximate unit-step response when one of the poles of $C(s)/R(s)$ can be neglected.

A family of curves $c(t)$ with various values of $\zeta$ is shown in Figure 4–10, where the abscissa is the dimensionless variable $\omega_n t$. The curves are functions only of $\zeta$. These

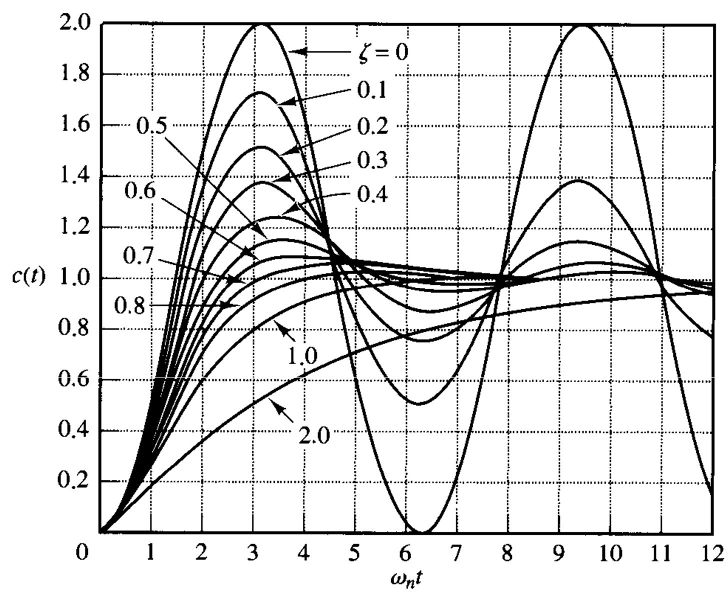



**Figure 4–10**
Unit-step response curves of the system shown in Figure 4–9.

curves are obtained from Equations (4–21), (4–24), and (4–26). The system described by these equations was initially at rest.

Note that two second-order systems having the same $\zeta$ but different $\omega_n$ will exhibit the same overshoot and the same oscillatory pattern. Such systems are said to have the same relative stability.

It is important to note that, for second-order systems whose closed-loop transfer functions are different from that given by Equation (4–19), the step-response curves may look quite different from those shown in Figure 4–10.

From Figure 4–10, we see that an underdamped system with $\zeta$ between 0.5 and 0.8 gets close to the final value more rapidly than a critically damped or overdamped system. Among the systems responding without oscillation, a critically damped system exhibits the fastest response. An overdamped system is always sluggish in responding to any inputs.

**Definitions of transient-response specifications.** In many practical cases, the desired performance characteristics of control systems are specified in terms of time-domain quantities. Systems with energy storage cannot respond instantaneously and will exhibit transient responses whenever they are subjected to inputs or disturbances.

Frequently, the performance characteristics of a control system are specified in terms of the transient response to a unit-step input since it is easy to generate and is sufficiently drastic. (If the response to a step input is known, it is mathematically possible to compute the response to any input.)

The transient response of a system to a unit-step input depends on the initial conditions. For convenience in comparing transient responses of various systems, it is a common practice to use the standard initial condition that the system is at rest initially with output and all time derivatives thereof zero. Then the response characteristics can be easily compared.

The transient response of a practical control system often exhibits damped oscillations before reaching steady state. In specifying the transient-response characteristics of a control system to a unit-step input, it is common to specify the following:

1. Delay time, $t_d$
2. Rise time, $t_r$
3. Peak time, $t_p$
4. Maximum overshoot, $M_p$
5. Settling time, $t_s$

These specifications are defined in what follows and are shown graphically in Figure 4–11.

1. Delay time, $t_d$: The delay time is the time required for the response to reach half the final value the very first time.
2. Rise time, $t_r$: The rise time is the time required for the response to rise from 10% to 90%, 5% to 95%, or 0% to 100% of its final value. For underdamped second-order systems, the 0% to 100% rise time is normally used. For overdamped systems, the 10% to 90% rise time is commonly used.

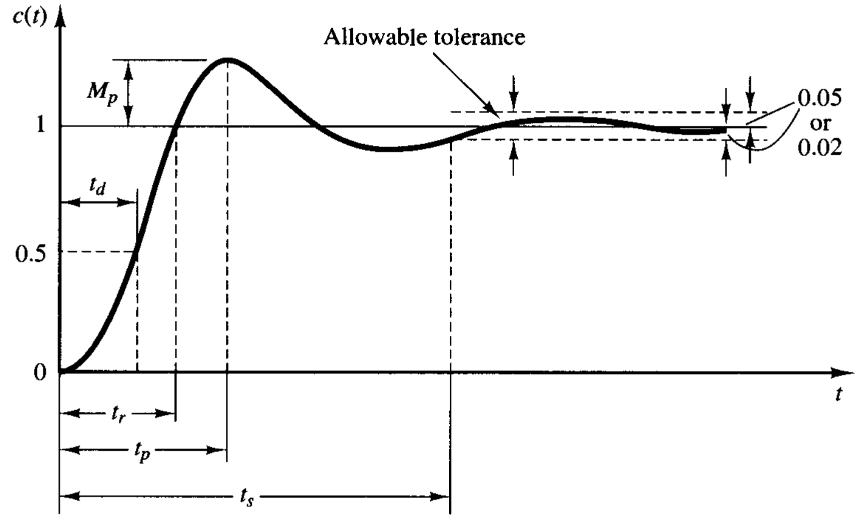



**Figure 4–11** Unit-step response curve showing $t_d$, $t_r$, $t_p$, $M_p$, and $t_s$.

3. Peak time, $t_p$: The peak time is the time required for the response to reach the first peak of the overshoot.
4. Maximum (percent) overshoot, $M_p$: The maximum overshoot is the maximum peak value of the response curve measured from unity. If the final steady-state value of the response differs from unity, then it is common to use the maximum percent overshoot. It is defined by

$$\text{Maximum percent overshoot} = \frac{c(t_p) - c(\infty)}{c(\infty)} \times 100\%$$

   The amount of the maximum (percent) overshoot directly indicates the relative stability of the system.
5. Settling time, $t_s$: The settling time is the time required for the response curve to reach and stay within a range about the final value of size specified by absolute percentage of the final value (usually 2% or 5%). The settling time is related to the largest time constant of the control system. Which percentage error criterion to use may be determined from the objectives of the system design in question.

The time-domain specifications just given are quite important since most control systems are time-domain systems; that is, they must exhibit acceptable time responses. (This means that the control system must be modified until the transient response is satisfactory.) Note that if we specify the values of $t_d$, $t_r$, $t_p$, $t_s$, and $M_p$, then the shape of the response curve is virtually determined. This may be seen clearly from Figure 4–12.

Note that not all these specifications necessarily apply to any given case. For example, for an overdamped system, the terms peak time and maximum overshoot do not apply. (For systems that yield steady-state errors for step inputs, this error must be kept within a specified percentage level. Detailed discussions of steady-state errors are postponed until Section 5–10.)

**A few comments on transient-response specifications.** Except for certain applications where oscillations cannot be tolerated, it is desirable that the transient

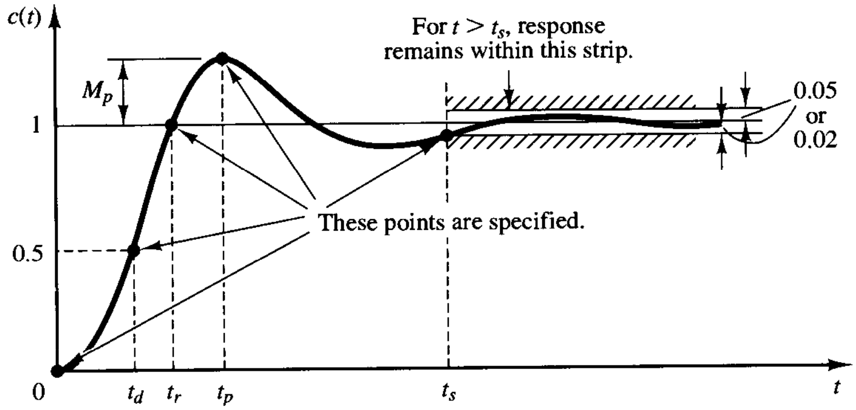



**Figure 4–12** Transient-response specifications.

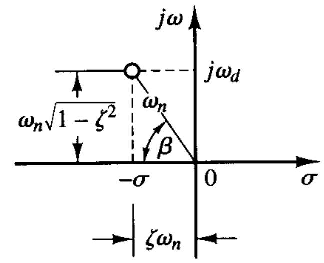



**Figure 4–13** Definition of the angle $\beta$.

response be sufficiently fast and be sufficiently damped. Thus, for a desirable transient response of a second-order system, the damping ratio must be between 0.4 and 0.8. Small values of $\zeta$ ($\zeta < 0.4$) yield excessive overshoot in the transient response, and a system with a large value of $\zeta$ ($\zeta > 0.8$) responds sluggishly.

We shall see later that the maximum overshoot and the rise time conflict with each other. In other words, both the maximum overshoot and the rise time cannot be made smaller simultaneously. If one of them is made smaller, the other necessarily becomes larger.

**Second-order systems and transient-response specifications.** In the following, we shall obtain the rise time, peak time, maximum overshoot, and settling time of the second-order system given by Equation (4–19). These values will be obtained in terms of $\zeta$ and $\omega_n$. The system is assumed to be underdamped.

Rise time $t_r$: Referring to Equation (4–21), we obtain the rise time $t_r$ by letting $c(t_r) = 1$ or

$$c(t_r) = 1 = 1 - e^{-\zeta\omega_n t_r}\left(\cos \omega_d t_r + \frac{\zeta}{\sqrt{1-\zeta^2}} \sin \omega_d t_r\right) \tag{4–27}$$

Since $e^{-\zeta\omega_n t_r} \neq 0$, we obtain from Equation (4–27) the following equation:

$$\cos \omega_d t_r + \frac{\zeta}{\sqrt{1-\zeta^2}} \sin \omega_d t_r = 0$$

or

$$\tan \omega_d t_r = -\frac{\sqrt{1-\zeta^2}}{\zeta} = -\frac{\omega_d}{\sigma}$$

Thus, the rise time $t_r$ is

$$t_r = \frac{1}{\omega_d} \tan^{-1}\left(\frac{\omega_d}{-\sigma}\right) = \frac{\pi - \beta}{\omega_d} \tag{4–28}$$

where $\beta$ is defined in Figure 4–13. Clearly, for a small value of $t_r$, $\omega_d$ must be large.

Peak time $t_p$: Referring to Equation (4–21), we may obtain the peak time by differentiating $c(t)$ with respect to time and letting this derivative equal zero. Since

$$\frac{dc}{dt} = \zeta\omega_n e^{-\zeta\omega_n t}\left(\cos \omega_d t + \frac{\zeta}{\sqrt{1-\zeta^2}} \sin \omega_d t\right)$$
$$+ e^{-\zeta\omega_n t}\left(\omega_d \sin \omega_d t - \frac{\zeta\omega_d}{\sqrt{1-\zeta^2}} \cos \omega_d t\right)$$

and the cosine terms in this last equation cancel each other, $dc/dt$, evaluated at $t = t_p$, can be simplified to

$$\left.\frac{dc}{dt}\right|_{t=t_p} = (\sin \omega_d t_p) \frac{\omega_n}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t_p} = 0$$

This last equation yields the following equation:

$$\sin \omega_d t_p = 0$$

or

$$\omega_d t_p = 0, \pi, 2\pi, 3\pi, \ldots$$

Since the peak time corresponds to the first peak overshoot, $\omega_d t_p = \pi$. Hence

$$t_p = \frac{\pi}{\omega_d} \tag{4-29}$$

The peak time $t_p$ corresponds to one-half cycle of the frequency of damped oscillation.

Maximum overshoot $M_p$: The maximum overshoot occurs at the peak time or at $t = t_p = \pi/\omega_d$. Thus, from Equation (4–21), $M_p$ is obtained as

$$\begin{aligned} M_p &= c(t_p) - 1 \\ &= -e^{-\zeta\omega_n(\pi/\omega_d)}\left(\cos \pi + \frac{\zeta}{\sqrt{1-\zeta^2}} \sin \pi\right) \\ &= e^{-(\sigma/\omega_d)\pi} = e^{-(\zeta/\sqrt{1-\zeta^2})\pi} \end{aligned} \tag{4-30}$$

The maximum percent overshoot is $e^{-(\sigma/\omega_d)\pi} \times 100\%$.

Settling time $t_s$: For an underdamped second-order system, the transient response is obtained from Equation (4–21) as

$$c(t) = 1 - \frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}} \sin\left(\omega_d t + \tan^{-1}\frac{\sqrt{1-\zeta^2}}{\zeta}\right), \qquad \text{for } t \geq 0$$

The curves $1 \pm (e^{-\zeta\omega_n t}/\sqrt{1-\zeta^2})$ are the envelope curves of the transient response for a unit-step input. The response curve $c(t)$ always remains within a pair of the envelope curves, as shown in Figure 4–14. The time constant of these envelope curves is $1/\zeta\omega_n$.

The speed of decay of the transient response depends on the value of the time constant $1/\zeta\omega_n$. For a given $\omega_n$, the settling time $t_s$ is a function of the damping ratio $\zeta$. From Figure 4–10, we see that for the same $\omega_n$ and for a range of $\zeta$ between 0 and 1 the settling time $t_s$ for a very lightly damped system is larger than that for a properly damped system. For an overdamped system, the settling time $t_s$ becomes large because of the sluggish start of the response.

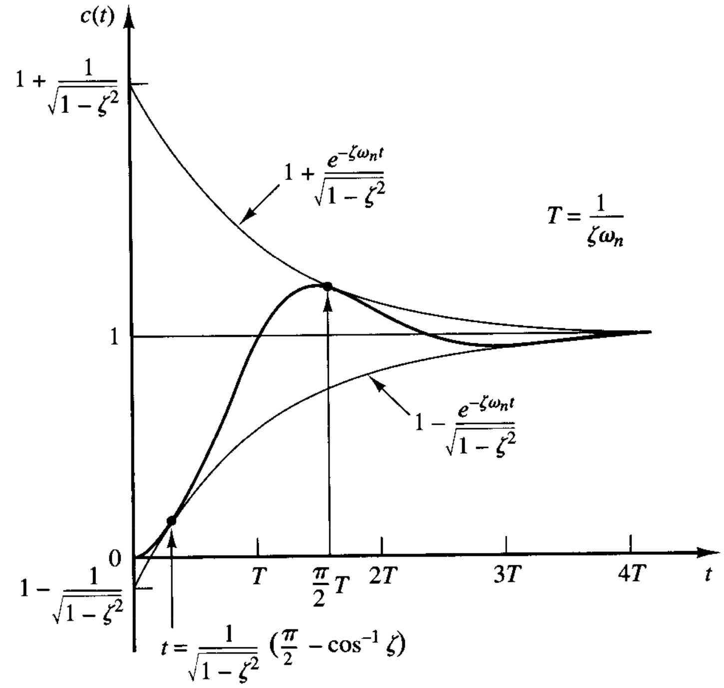



**Figure 4–14**
Pair of envelope curves for the unit-step response curve of the system shown in Figure 4–9.

The settling time corresponding to a $\pm 2\%$ or $\pm 5\%$ tolerance band may be measured in terms of the time constant $T = 1/\zeta\omega_n$ from the curves of Figure 4–10 for different values of $\zeta$. The results are shown in Figure 4–15. For $0 < \zeta < 0.9$, if the 2% criterion is used $t_s$ is approximately four times the time constant of the system. If the 5% criterion is used, then $t_s$ is approximately three times the time constant. Note that the settling time reaches a minimum value around $\zeta = 0.76$ (for the 2% criterion) or $\zeta = 0.68$ (for the 5% criterion) and then increases almost linearly for large values of $\zeta$. The discontinuities in the curves of Figure 4–15 arise because an infinitesimal change in the value of $\zeta$ can cause a finite change in the settling time.

For convenience in comparing the responses of systems, we commonly define the settling time $t_s$ to be

$$t_s = 4T = \frac{4}{\sigma} = \frac{4}{\zeta\omega_n} \qquad \text{(2\% criterion)} \tag{4–31}$$

or

$$t_s = 3T = \frac{3}{\sigma} = \frac{3}{\zeta\omega_n} \qquad \text{(5\% criterion)} \tag{4–32}$$

Note that the settling time is inversely proportional to the product of the damping ratio and the undamped natural frequency of the system. Since the value of $\zeta$ is usually determined from the requirement of permissible maximum overshoot, the settling time is determined primarily by the undamped natural frequency $\omega_n$. This means that the duration of the transient period may be varied, without changing the maximum overshoot, by adjusting the undamped natural frequency $\omega_n$.

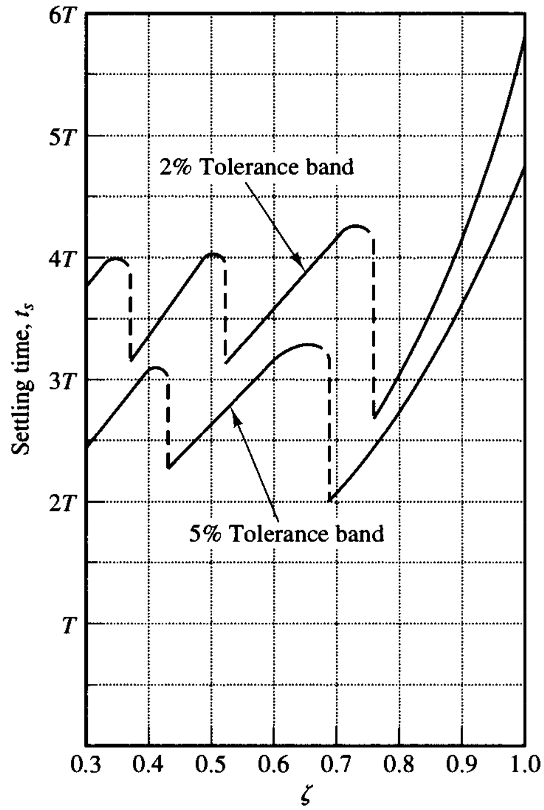



**Figure 4–15**
Settling time $t_s$ versus $\zeta$ curves.

From the preceding analysis, it is evident that for rapid response $\omega_n$ must be large. To limit the maximum overshoot $M_p$ and to make the settling time small, the damping ratio $\zeta$ should not be too small. The relationship between the maximum percent overshoot $M_p$ and the damping ratio $\zeta$ is presented in Figure 4–16. Note that if the damping ratio is between 0.4 and 0.8 then the maximum percent overshoot for step response is between 25% and 2.5%.

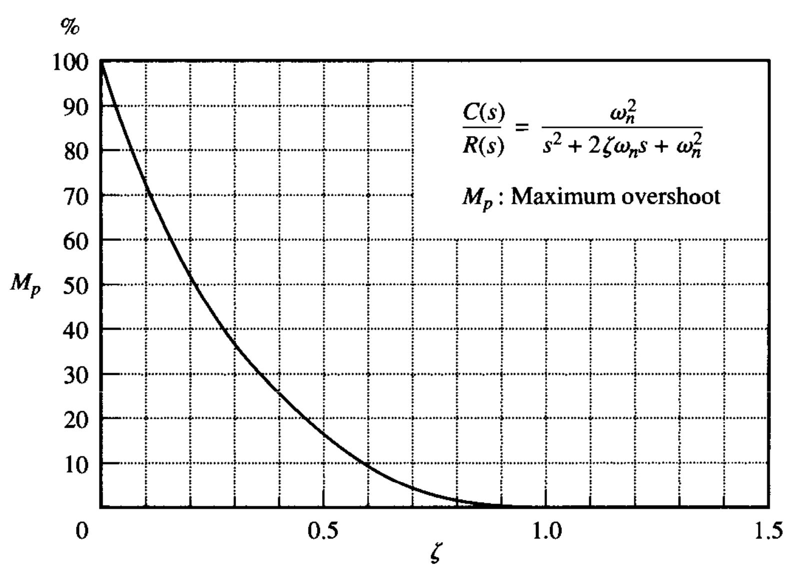



**Figure 4–16**
$M_p$ versus $\zeta$ curve.

**EXAMPLE 4–2** Consider the system shown in Figure 4–9, where $\zeta = 0.6$ and $\omega_n = 5$ rad/sec. Let us obtain the rise time $t_r$, peak time $t_p$, maximum overshoot $M_p$, and settling time $t_s$ when the system is subjected to a unit-step input.

From the given values of $\zeta$ and $\omega_n$, we obtain $\omega_d = \omega_n\sqrt{1-\zeta^2} = 4$ and $\sigma = \zeta\omega_n = 3$.

Rise time $t_r$: The rise time is

$$t_r = \frac{\pi - \beta}{\omega_d} = \frac{3.14 - \beta}{4}$$

where $\beta$ is given by

$$\beta = \tan^{-1}\frac{\omega_d}{\sigma} = \tan^{-1}\frac{4}{3} = 0.93 \text{ rad}$$

The rise time $t_r$ is thus

$$t_r = \frac{3.14 - 0.93}{4} = 0.55 \text{ sec}$$

Peak time $t_p$: The peak time is

$$t_p = \frac{\pi}{\omega_d} = \frac{3.14}{4} = 0.785 \text{ sec}$$

Maximum overshoot $M_p$: The maximum overshoot is

$$M_p = e^{-(\sigma/\omega_d)\pi} = e^{-(3/4)\times 3.14} = 0.095$$

The maximum percent overshoot is thus 9.5%

Settling time $t_s$: For the 2% criterion, the settling time is

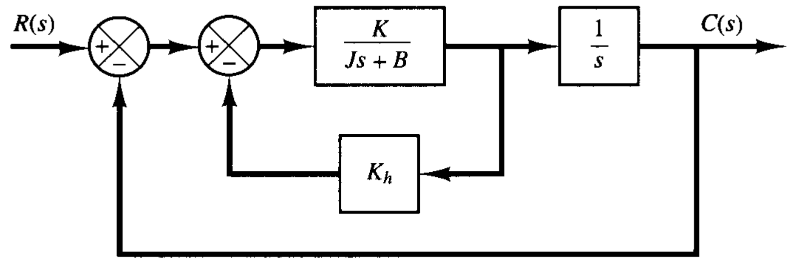
$$t_s = \frac{4}{\sigma} = \frac{4}{3} = 1.33 \text{ sec}$$

For the 5% criterion,

$$t_s = \frac{3}{\sigma} = \frac{3}{3} = 1 \text{ sec}$$

**Servo system with velocity feedback.** The derivative of the output signal can be used to improve system performance. In obtaining the derivative of the output position signal, it is desirable to use a tachometer instead of physically differentiating the output signal. (Note that the differentiation amplifies noise effects. In fact, if discontinuous noises are present, differentiation amplifies the discontinuous noises more than the useful signal. For example, the output of a potentiometer is a discontinuous voltage signal because, as the potentiometer brush is moving on the windings, voltages are induced in the switchover turns and thus generate transients. The output of the potentiometer therefore should not be followed by a differentiating element.)

Consider the servo system shown in Figure 4–17(a). In this device, the velocity signal, together with the positional signal, is fed back to the input to produce the actuating error signal. In any servo system, such a velocity signal can be easily generated by a



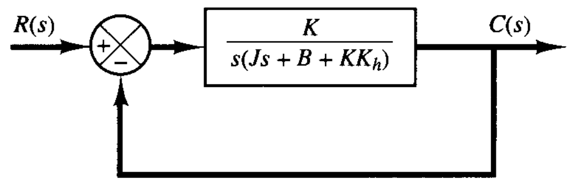




**Figure 4–17**
(a) Block diagram of a servo system; (b) simplified block diagram.

tachometer. The block diagram shown in Figure 4–17(a) can be simplified, as shown in Figure 4–17(b), giving

$$\frac{C(s)}{R(s)} = \frac{K}{Js^2 + (B + KK_h)s + K} \tag{4–33}$$

Comparing Equation (4–33) with Equation (4–17), notice that the velocity feedback has the effect of increasing damping. The damping ratio $\zeta$ becomes

$$\zeta = \frac{B + KK_h}{2\sqrt{KJ}} \tag{4–34}$$

The undamped natural frequency $\omega_n = \sqrt{K/J}$ is not affected by velocity feedback. Noting that the maximum overshoot for a unit-step input can be controlled by controlling the value of the damping ratio $\zeta$, we can reduce the maximum overshoot by adjusting the velocity feedback constant $K_h$ so that $\zeta$ is between 0.4 and 0.7.

Remember that velocity feedback has the effect of increasing the damping ratio without affecting the undamped natural frequency of the system.

**EXAMPLE 4–3**

For the system shown in Figure 4–17(a), determine the values of gain $K$ and velocity feedback constant $K_h$ so that the maximum overshoot in the unit-step response is 0.2 and the peak time is 1 sec. With these values of $K$ and $K_h$, obtain the rise time and settling time. Assume that $J = 1$ kg-m$^2$ and $B = 1$ N-m/rad/sec.

Determination of the values of $K$ and $K_h$: The maximum overshoot $M_p$ is given by Equation (4–30) as

$$M_p = e^{-(\zeta/\sqrt{1-\zeta^2})\pi}$$

This value must be 0.2. Thus,

$$e^{-(\zeta/\sqrt{1-\zeta^2})\pi} = 0.2$$

or

$$\frac{\zeta\pi}{\sqrt{1-\zeta^2}} = 1.61$$

which yields

$$\zeta = 0.456$$

The peak time $t_p$ is specified as 1 sec; therefore, from Equation (4–29),

$$t_p = \frac{\pi}{\omega_d} = 1$$

or

$$\omega_d = 3.14$$

Since $\zeta$ is 0.456, $\omega_n$ is

$$\omega_n = \frac{\omega_d}{\sqrt{1-\zeta^2}} = 3.53$$

Since the natural frequency $\omega_n$ is equal to $\sqrt{K/J}$,

$$K = J\omega_n^2 = \omega_n^2 = 12.5 \text{ N-m}$$

Then, $K_h$ is, from Equation (4–34),

$$K_h = \frac{2\sqrt{KJ}\zeta - B}{K} = \frac{2\sqrt{K}\zeta - 1}{K} = 0.178 \text{ sec}$$

Rise time $t_r$: From Equation (4–28), the rise time $t_r$ is

$$t_r = \frac{\pi - \beta}{\omega_d}$$

where

$$\beta = \tan^{-1}\frac{\omega_d}{\sigma} = \tan^{-1} 1.95 = 1.10$$

Thus, $t_r$ is

$$t_r = 0.65 \text{ sec}$$

Settling time $t_s$: For the 2% criterion,

$$t_s = \frac{4}{\sigma} = 2.48 \text{ sec}$$

For the 5% criterion,

$$t_s = \frac{3}{\sigma} = 1.86 \text{ sec}$$

**Impulse response of second-order systems.** For a unit-impulse input $r(t)$, the corresponding Laplace transform is unity, or $R(s) = 1$. The unit-impulse response $C(s)$ of the second-order system shown in Figure 4–9 is

$$C(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

The inverse Laplace transform of this equation yields the time solution for the response $c(t)$ as follows:

For $0 \le \zeta < 1$,

$$c(t) = \frac{\omega_n}{\sqrt{1 - \zeta^2}} e^{-\zeta\omega_n t} \sin \omega_n \sqrt{1 - \zeta^2}\, t, \qquad \text{for } t \ge 0 \tag{4–35}$$

For $\zeta = 1$,

$$c(t) = \omega_n^2 t e^{-\omega_n t}, \qquad \text{for } t \ge 0 \tag{4–36}$$

For $\zeta > 1$,

$$c(t) = \frac{\omega_n}{2\sqrt{\zeta^2 - 1}} e^{-(\zeta - \sqrt{\zeta^2-1})\omega_n t} - \frac{\omega_n}{2\sqrt{\zeta^2 - 1}} e^{-(\zeta + \sqrt{\zeta^2-1})\omega_n t}, \qquad \text{for } t \ge 0 \tag{4–37}$$

Note that without taking the inverse Laplace transform of $C(s)$ we can also obtain the time response $c(t)$ by differentiating the corresponding unit-step response since the unit-impulse function is the time derivative of the unit-step function. A family of unit-impulse response curves given by Equations (4–35) and (4–36) with various values of $\zeta$ is shown in Figure 4–18. The curves $c(t)/\omega_n$ are plotted against the dimensionless variable $\omega_n t$, and thus they are functions only of $\zeta$. For the critically damped and overdamped cases, the unit-impulse response is always positive or zero; that is, $c(t) \ge 0$. This can be seen from Equations (4–36) and (4–37). For the underdamped case, the unit-impulse response $c(t)$ oscillates about zero and takes both positive and negative values.

From the foregoing analysis, we may conclude that if the impulse response $c(t)$ does not change sign, the system is either critically damped or overdamped, in which case the

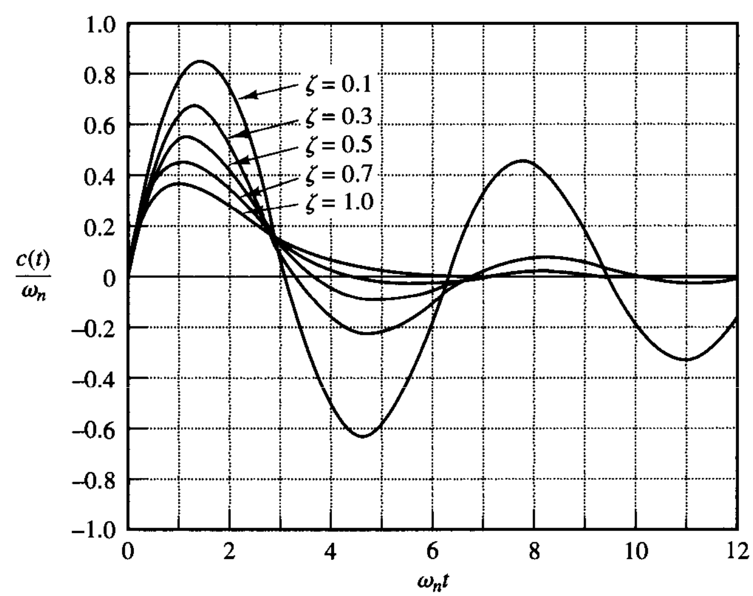



**Figure 4–18** Unit-impulse response curves of the system shown in Figure 4–9.

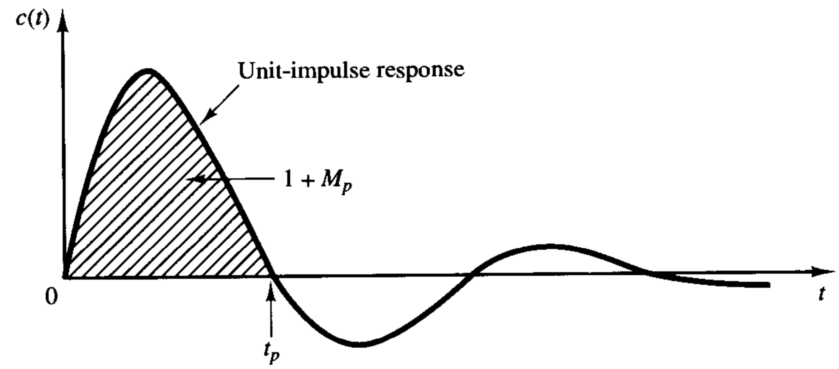



**Figure 4–19**
Unit-impulse response curve of the system shown in Figure 4–9.

corresponding step response does not overshoot but increases or decreases monotonically and approaches a constant value.

The maximum overshoot for the unit-impulse response of the underdamped system occurs at

$$t = \frac{\tan^{-1}\dfrac{\sqrt{1-\zeta^2}}{\zeta}}{\omega_n\sqrt{1-\zeta^2}}, \qquad \text{where } 0 < \zeta < 1$$

and the maximum overshoot is

$$c(t)_{\max} = \omega_n \exp\left(-\frac{\zeta}{\sqrt{1-\zeta^2}}\tan^{-1}\frac{\sqrt{1-\zeta^2}}{\zeta}\right), \qquad \text{where } 0 < \zeta < 1$$

Since the unit-impulse response function is the time derivative of the unit-step response function, the maximum overshoot $M_p$ for the unit-step response can be found from the corresponding unit-impulse response. That is, the area under the unit-impulse response curve from $t = 0$ to the time of the first zero, as shown in Figure 4–19, is $1 + M_p$, where $M_p$ is the maximum overshoot (for the unit-step response) given by Equation (4–30). The peak time $t_p$ (for the unit-step response) given by Equation (4–29) corresponds to the time that the unit-impulse response first crosses the time axis.

## 4–4 TRANSIENT-RESPONSE ANALYSIS WITH MATLAB

**Introduction.** In this section we present the computational approach to the transient-response analysis with MATLAB. Those readers who are as yet unfamiliar with MATLAB may wish to read Appendix before studying this section.

As stated earlier in this chapter, transient responses (such as the step response, impulse response, and ramp response) are used frequently to investigate the time-domain characteristics of control systems.

**MATLAB representation of linear systems.** The transfer function of a system is represented by two arrays of numbers. Consider the system

$$\frac{C(s)}{R(s)} = \frac{25}{s^2 + 4s + 25} \tag{4–38}$$

This system is represented as two arrays each containing the coefficients of the polynomials in decreasing powers of $s$ as follows:

```
num = [0  0  25]
den = [1  4  25]
```

Note that zeros are padded where necessary.

If num and den (the numerator and denominator of the closed-loop transfer function) are known, commands such as

```
step(num,den),      step(num,den,t)
```

will generate plots of unit-step responses. (t in the step command is the user-specified time.)

For a control system defined in a state-space form, where state matrix **A**, control matrix **B**, output matrix **C**, and direct transmission matrix **D** of state-space equations are known, the command

```
step(A,B,C,D)
```

will generate plots of unit-step responses. The time vector is automatically determined when t is not explicitly included in the step commands.

Note that when step commands have left-hand arguments such as

```
[y,x,t] = step(num,den,t)
[y,x,t] = step(A,B,C,D,iu)
[y,x,t] = step(A,B,C,D,iu,t)                    (4–39)
```

no plot is shown on the screen. Hence it is necessary to use a *plot* command to see the response curves. The matrices y and x contain the output and state response of the system, respectively, evaluated at the computation time points t. (y has as many columns as outputs and one row for each element in t. x has as many columns as states and one row for each element in t.)

Note in Equation (4–39) that the scalar iu is an index into the inputs of the system and specifies which input is to be used for the response, and t is the user-specified time. If the system involves multiple inputs and multiple outputs, the step command, such as given by Equation (4–39), produces a series of step response plots, one for each input and output combination of

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu}$$

$$\mathbf{y} = \mathbf{Cx} + \mathbf{Du}$$

(For details, see Example 4–4.)

**Obtaining the unit-step response of the transfer-function system.** Let us consider the unit-step response of the system given by Equation (4–38). MATLAB Program 4–1 will yield a plot of the unit-step response of this system. A plot of the unit-step response curve is shown in Figure 4–20.

```
MATLAB Program 4-1

% ---------Unit-step response----------

% ***** Enter the numerator and denominator of the transfer
% function *****

num = [0    0    25];
den = [1    4    25];

% ***** Enter the following step-response command *****

step(num,den)

% ***** Enter grid and title of the plot *****

grid
title ('Unit-Step Response of G(s) = 25/(s^2 + 4s + 25)')
```

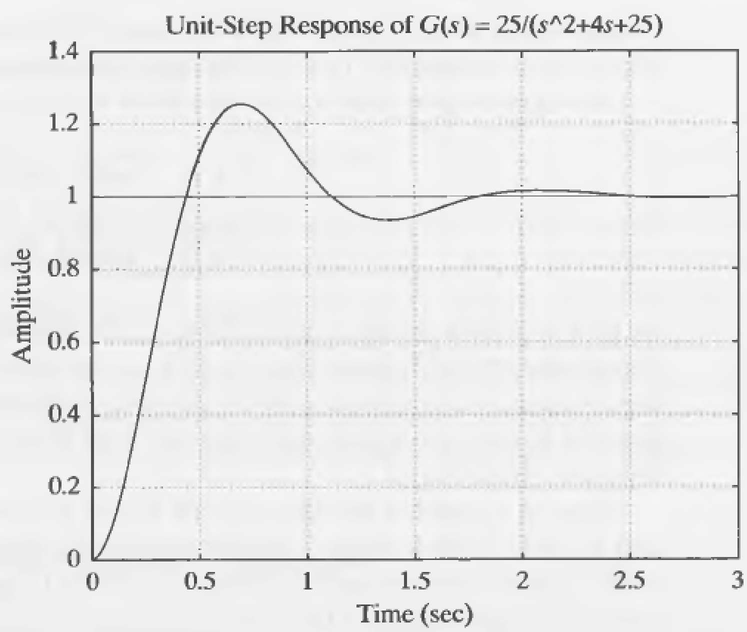



**Figure 4–20**
Unit-step response curve.

**EXAMPLE 4–4** Consider the following system:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ 6.5 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

Obtain the unit-step response curves.

Although it is not necessary to obtain the transfer function expression for the system to obtain the unit-step response curves with MATLAB, we shall derive such an expression for reference. For the system defined by

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

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}$$

the transfer matrix $\mathbf{G}(s)$ is a matrix that relates $\mathbf{Y}(s)$ and $\mathbf{U}(s)$ as follows:

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

Taking Laplace transforms of the state-space equations, we obtain

$$s\mathbf{X}(s) - \mathbf{x}(0) = \mathbf{A}\mathbf{X}(s) + \mathbf{B}\mathbf{U}(s) \tag{4-40}$$

$$\mathbf{Y}(s) = \mathbf{C}\mathbf{X}(s) + \mathbf{D}\mathbf{U}(s) \tag{4-41}$$

In deriving the transfer matrix, we assume that $\mathbf{x}(0) = \mathbf{0}$. Then, from Equation (4–40), we get

$$\mathbf{X}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{U}(s) \tag{4-42}$$

Substituting Equation (4–42) into Equation (4–41), we obtain

$$\mathbf{Y}(s) = [\mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}]\mathbf{U}(s)$$

Thus the transfer matrix $\mathbf{G}(s)$ is given by

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

The transfer matrix $\mathbf{G}(s)$ for the given system becomes

$$\begin{aligned}
\mathbf{G}(s) &= \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} s+1 & 1 \\ -6.5 & s \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \\
&= \frac{1}{s^2 + s + 6.5} \begin{bmatrix} s & -1 \\ 6.5 & s+1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \\
&= \frac{1}{s^2 + s + 6.5} \begin{bmatrix} s-1 & s \\ s+7.5 & 6.5 \end{bmatrix}
\end{aligned}$$

Hence

$$\begin{bmatrix} Y_1(s) \\ Y_2(s) \end{bmatrix} = \begin{bmatrix} \dfrac{s-1}{s^2+s+6.5} & \dfrac{s}{s^2+s+6.5} \\ \dfrac{s+7.5}{s^2+s+6.5} & \dfrac{6.5}{s^2+s+6.5} \end{bmatrix} \begin{bmatrix} U_1(s) \\ U_2(s) \end{bmatrix}$$

Since this system involves two inputs and two outputs, four transfer functions may be defined depending on which signals are considered as input and output. Note that, when considering the signal $u_1$ as the input, we assume that signal $u_2$ is zero, and vice versa. The four transfer functions are

$$\frac{Y_1(s)}{U_1(s)} = \frac{s-1}{s^2+s+6.5}, \qquad \frac{Y_2(s)}{U_1(s)} = \frac{s+7.5}{s^2+s+6.5}$$

$$\frac{Y_1(s)}{U_2(s)} = \frac{s}{s^2+s+6.5}, \qquad \frac{Y_2(s)}{U_2(s)} = \frac{6.5}{s^2+s+6.5}$$

The four individual step-response curves can be plotted by use of the command

```
step(A,B,C,D)
```

MATLAB Program 4–2 produces four such step-response curves. The curves are shown in Figure 4–21.

| MATLAB Program 4–2 |
|---|
| A = [−1 −1;6.5 0]; |
| B = [1 1;1 0]; |
| C = [1 0;0 1]; |
| D = [0 0;0 0]; |
| step(A,B,C,D) |

To plot two step-response curves for the input $u_1$ in one diagram and two step-response curves for the input $u_2$ in another diagram, we may use the commands

step(A,B,C,D,1)

and

step(A,B,C,D,2)

respectively. MATLAB Program 4–3 is a program to plot two step-response curves for the input $u_1$ in one diagram and two step-response curves for the input $u_2$ in another diagram. Figure 4–22 shows the two diagrams, each consisting of two step-response curves.

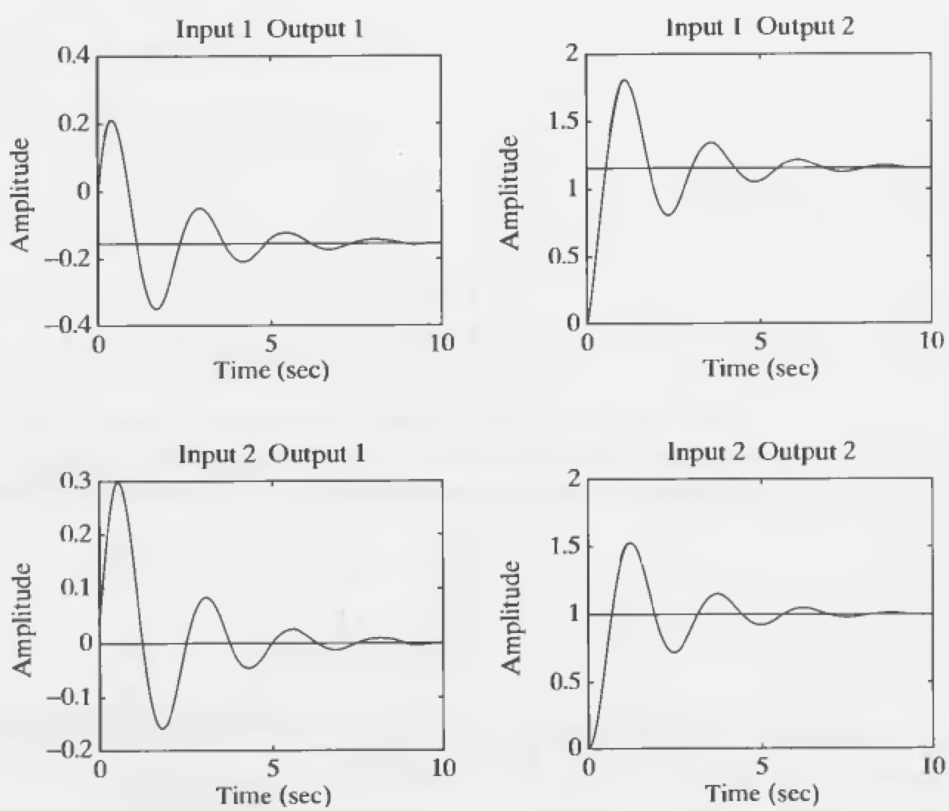



**Figure 4–21**
Unit-step response curves.

MATLAB Program 4–3

```
% ---------- Step-response curves for system defined in state
% space ----------

% ***** In this program we plot step-response curves of a system
% having two inputs (u1 and u2) and two outputs (y1 and y2) *****

% ***** We shall first plot step-response curves when the input is
% u1. Then we shall plot step-response curves when the input is
% u2 *****

% ***** Enter matrices A, B, C, and D *****

A = [-1   -1;6.5   0];
B = [1   1;1   0];
C = [1   0;0   1];
D = [0   0;0   0];

% ***** To plot step-response curves when the input is u1, enter
% the command 'step(A,B,C,D,1)' *****

step(A,B,C,D,1)
grid
title ('Step-Response Plots: Input = u1 (u2 = 0)')
text(3.4,-0.06,'Y1')
text(3.4,1.4,'Y2')

% ***** Next, we shall plot step-response curves when the input
% is u2. Enter the command 'step(A,B,C,D,2)' *****

step(A,B,C,D,2);
grid
title('Step-Response Plots: Input = u2 (u1 = 0)')
text(3,0.14,'Y1')
text(2.8,1.1,'Y2')
```

**Writing text on the graphics screen.** To write text on the graphics screen, enter, for example, the following statements:

```
text(3.4,-0.06,'Y1')
```

and

```
text(3.4,1.4,'Y2')
```

The first statement tells the computer to write 'Y1' beginning at the coordinates $x = 3.4$, $y = -0.06$. Similarly, the second statement tells the computer to write 'Y2' beginning at the coordinates $x = 3.4$, $y = 1.4$. [See MATLAB Program 4–3 and Figure 4–22(a).]

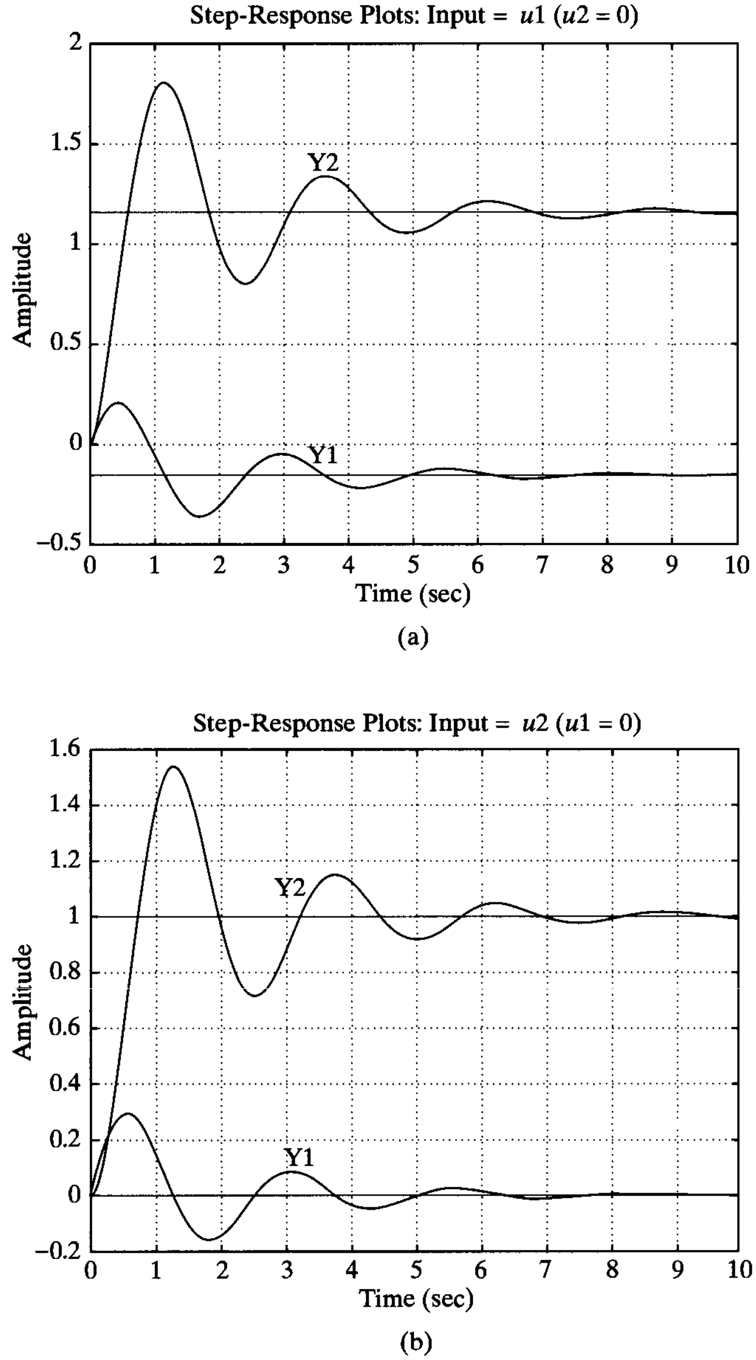



**Figure 4–22** Unit-step response curves. (a) $u_1$ is the input ($u_2 = 0$); (b) $u_2$ is the input ($u_1 = 0$).

**Impulse response.** The unit-impulse response of a control system may be obtained by use of one of the following MATLAB commands:

```
impulse(num,den)
impulse(A,B,C,D)
[y,x,t] = impulse (num,den)
[y,x,t] = impulse(num,den,t)                    (4–43)
```

```
[y,x,t] = impulse(A,B,C,D)
[y,x,t] = impulse(A,B,C,D,iu)                    (4-44)
[y,x,t] = impulse (A,B,C,D,iu,t)                 (4-45)
```

The command "impulse(num,den)" plots the unit-impulse response on the screen. The command "impulse(A,B,C,D)" produces a series of unit-impulse-response plots, one for each input and output combination of the system

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

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}$$

with the time vector automatically determined. Note that in Equations (4–44) and (4–45) the scalar iu is an index into the inputs of the system and specifies which input to be used for the impulse response.

Note also that in Equations (4–43) and (4–45) t is the user-supplied time vector. The vector t specifies the times at which the impulse response is to be computed.

If MATLAB is invoked with the left-hand argument [y,x,t], such as in the case of [y,x,t] = impulse(A,B,C,D), the command returns the output and state responses of the system and the time vector t. No plot is drawn on the screen. The matrices y and x contain the output and state responses of the system evaluated at the time points t. (y has as many columns as outputs and one row for each element in t. x has as many columns as state variables and one row for each element in t.)

**EXAMPLE 4–5** Obtain the unit-impulse response of the following system:

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

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

A possible MATLAB program is shown in MATLAB Program 4–4. The resulting response curve is shown in Figure 4–23.

MATLAB Program 4–4

```
A = [0   1;-1   -1];
B = [0;1];
C = [1   0];
D = [0];
impulse(A,B,C,D);
grid;
title('Unit-Impulse Response')
```

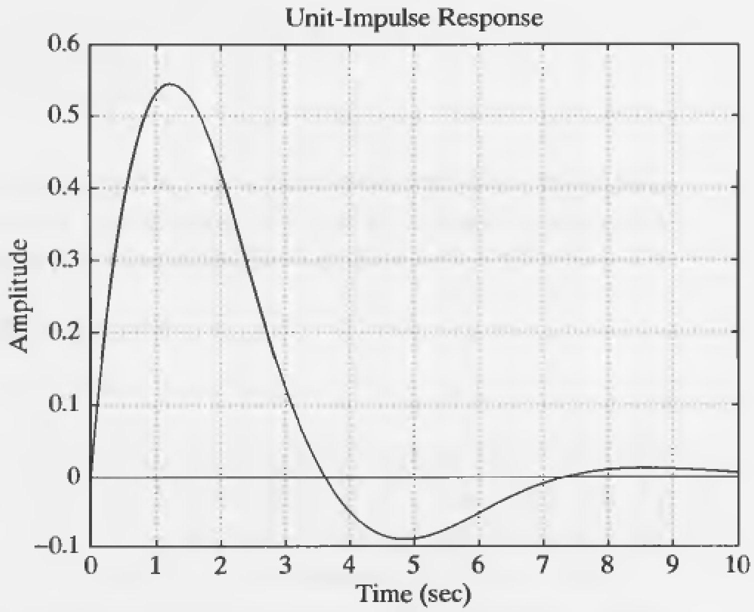



**Figure 4–23** Unit-impulse response curve.

**EXAMPLE 4–6** Obtain the unit-impulse response of the following system:

$$\frac{C(s)}{R(s)} = G(s) = \frac{1}{s^2 + 0.2s + 1}$$

MATLAB Program 4–5 will produce the unit-impulse response. The resulting plot is shown in Figure 4–24.

MATLAB Program 4–5

```
num = [0   0   1];
den = [1   0.2   1];
impulse(num,den);
grid
title('Unit-Impulse Response of G(s) = 1/(s^2 + 0.2s + l)')
```

**Alternative approach to obtain impulse response.** Note that when the initial conditions are zero the unit-impulse response of $G(s)$ is the same as the unit-step response of $sG(s)$.

Consider the unit-impulse response of the system considered in Example 4–6. Since $R(s) = 1$ for the unit-impulse input, we have

$$\frac{C(s)}{R(s)} = C(s) = G(s) = \frac{1}{s^2 + 0.2s + 1}$$

$$= \frac{s}{s^2 + 0.2s + 1}\frac{1}{s}$$

We can thus convert the unit-impulse response of $G(s)$ to the unit-step response of $sG(s)$.

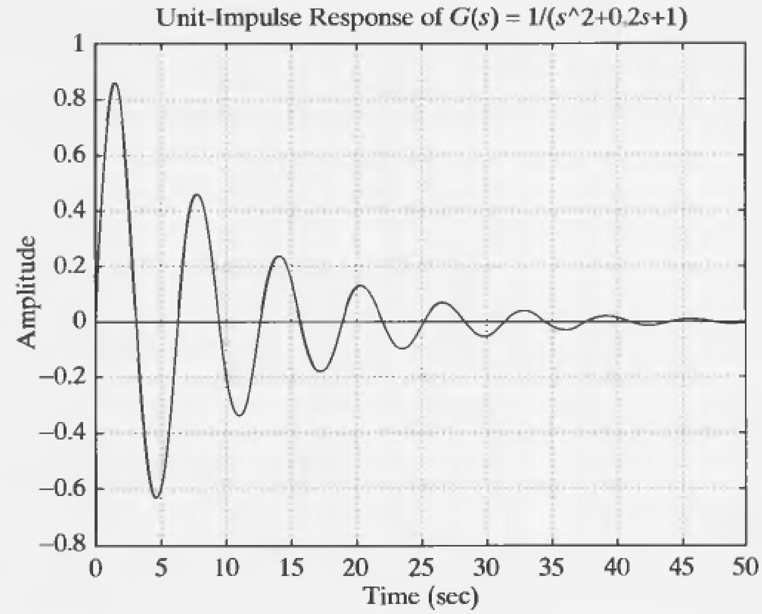



**Figure 4–24** Unit-impulse response curve.

If we enter the following num and den into MATLAB,

```
num = [0   1   0]
den = [1   0.2   1]
```

and use the step-response command, as given in MATLAB Program 4–6, we obtain a plot of the unit-impulse response of the system as shown in Figure 4–25.

MATLAB Program 4–6

```
num = [0   1   0];
den = [1   0.2   1];
step(num,den);
grid;
title('Unit-Step Response of sG(s) = s/(s^2 + 0.2s + 1)')
```

Notice in Figure 4–25 (and many others) that the $x$ axis and $y$ axis labels are automatically determined. If it is desired to label the $x$ axis and $y$ axis differently, we need to modify the step command. For example, if it is desired to label the $x$ axis as 't Sec' and the $y$ axis as 'Input and Output,' then use step-response commands with left-hand arguments, such as

```
c = step(num,den,t)
```

or, more generally,

```
[y,x,t] = step(num,den,t)
```

See MATLAB Program 4–7.

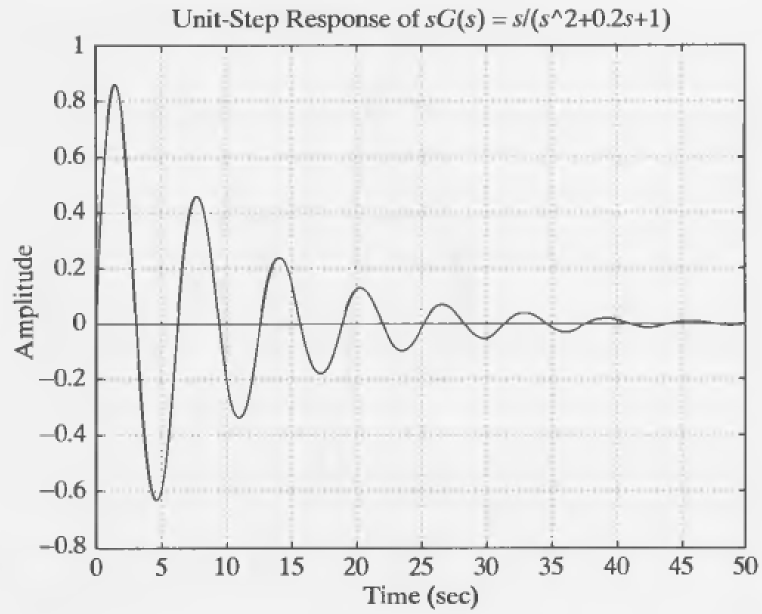



**Figure 4–25** Unit-impulse response curve obtained as the unit-step response of $sG(s) = s/(s^2 + 0.2s + 1)$.

MATLAB Program 4–7

```
% --------- Unit-ramp response ----------

% ***** The unit-ramp response is obtained as the unit-step
% response of G(s)/s *****

% ***** Enter the numerator and denominator of G(s)/s *****

num = [0   0   0   1];
den = [1   1   1   0];

% ***** Specify the computing time points (such as t = 0:0.1:7)
% and then enter step-response command: c = step(num,den,t) *****

t = 0:0.1:7;
c = step(num,den,t);

% ***** In plotting the ramp-response curve, add the reference
% input to the plot. The reference input is t. Add to the
% argument of the plot command with the following: t,t,'-'. Thus
% the plot command becomes as follows: plot(t,c,'o',t,t,'-') *****

plot(t,c,'o',t,t,'-')

% ***** Add grid, title, xlabel, and ylabel *****

grid
title('Unit-Ramp Response Curve for System G(s) = 1/(s^2 + s + 1)')
xlabel('t Sec')
ylabel('Input and Output')
```

**Ramp response.** There is no ramp command in MATLAB. Therefore, we need to use the step command to obtain the ramp response. Specifically, to obtain the ramp response of the transfer-function system $G(s)$, divide $G(s)$ by $s$ and use the step-response command. For example, consider the closed-loop system

$$\frac{C(s)}{R(s)} = \frac{1}{s^2 + s + 1}$$

For a unit-ramp input, $R(s) = 1/(s^2)$. Hence

$$C(s) = \frac{1}{s^2 + s + 1}\frac{1}{s^2} = \frac{1}{(s^2 + s + 1)s}\frac{1}{s}$$

To obtain the unit-ramp response of this system, enter the following numerator and denominator into the MATLAB program,

```
num = [0  0  0  1];
den = [1  1  1  0];
```

and use the step-response command. See MATLAB Program 4–7. The plot obtained by using this program is shown in Figure 4–26.

**Unit-ramp response of a system defined in state space.** Next, we shall treat the unit-ramp response of the system in state-space form. Consider the system described by

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{B}u$$

$$y = \mathbf{Cx} + Du$$

In what follows, we shall consider a simple example to explain the method. Let us assume that

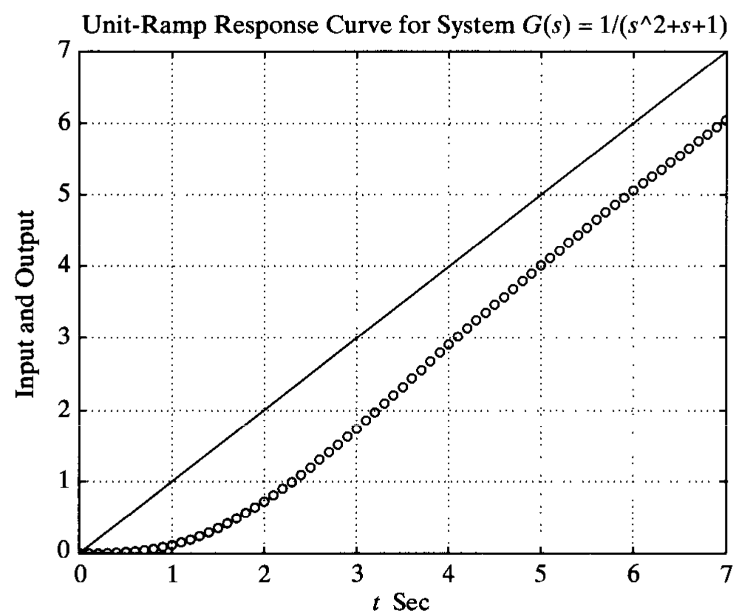



**Figure 4–26** Unit-ramp response curve.

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{x}(0) = \mathbf{0}$$

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

When the initial conditions are zeros, the unit-ramp response is the integral of the unit-step response. Hence the unit-ramp response can be given by

$$z = \int_0^t y\,dt \tag{4-46}$$

From Equation (4–46), we obtain

$$\dot{z} = y = x_1 \tag{4-47}$$

Let us define

$$z = x_3$$

Then Equation (4–47) becomes

$$\dot{x}_3 = x_1 \tag{4-48}$$

Combining Equation (4–48) with the original state-space equation, we obtain

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

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

which can be written as

$$\dot{\mathbf{x}} = \mathbf{AAx} + \mathbf{BB}u$$

$$z = \mathbf{CCx} + DDu$$

where

$$\mathbf{AA} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \left[\begin{array}{c|c} \mathbf{A} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \hline \mathbf{C} & 0 \end{array}\right]$$

$$\mathbf{BB} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{B} \\ \hline 0 \end{bmatrix}, \quad \mathbf{CC} = [0 \quad 0 \quad 1], \quad DD = [0]$$

Note that $x_3$ is the third element of $\mathbf{x}$. A plot of the unit-ramp response curve $z(t)$ can be obtained by entering MATLAB Program 4–8 into the computer. A plot of

MATLAB Program 4–8

```
% ---------- Unit-ramp response ----------

% ***** The unit-ramp response is obtained by adding a new
% state variable x3. The dimension of the state equation
% is enlarged by one *****

% ***** Enter matrices A, B, C, and D of the original state
% equation and output equation *****

A = [0   1;-1   -1];
B = [0;1];
C = [1   0];
D = [0];

% ***** Enter matrices AA, BB, CC, and DD of the new,
% enlarged state equation and output equation *****

AA = [A   zeros(2,1);C   0];
BB = [B;0];
CC = [0   0   1];
DD = [0];

% ***** Enter step-response command: [z,x,t] = step(AA,BB,CC,DD) *****

[z,x,t] = step(AA,BB,CC,DD);

% ***** In plotting x3 add the unit-ramp input t in the plot
% by entering the following command: plot(t,x3,'o',t,t,'-') *****

x3 = [0   0   1]*x'; plot(t,x3,'o',t,t,'-')
grid
title('Unit-Ramp Response')
xlabel('t Sec')
ylabel('Input and Output')
```

the unit-ramp response curve obtained from this MATLAB program is shown in Figure 4–27.

**Response to initial condition (transfer-function approach).** In what follows we shall present a method for obtaining the response to an initial condition by use of an example.

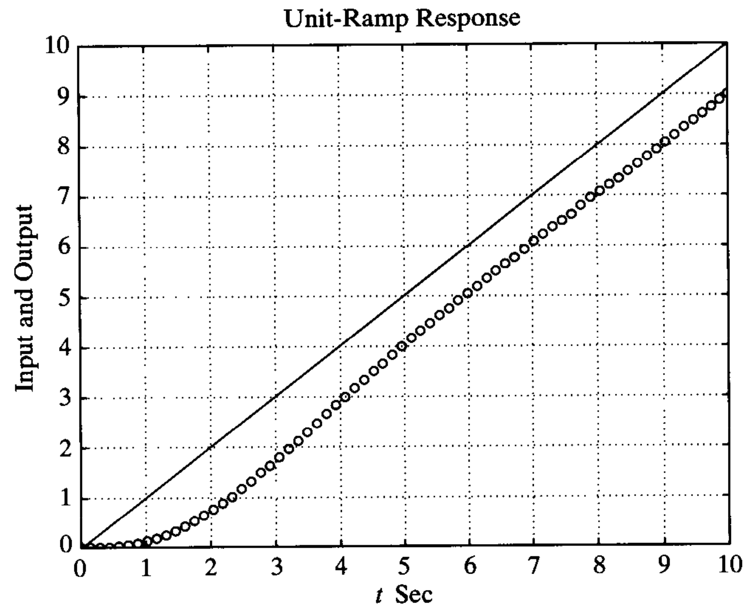



**Figure 4–27**
Unit-ramp response curve.

**EXAMPLE 4–7** In this example, we shall consider a system subjected only to an initial condition.

Consider the mechanical system shown in Figure 4–28, where $m = 1$ kg, $b = 3$ N-sec/m, and $k = 2$ N/m. Assume that at $t = 0$ the mass $m$ is pulled downward such that $x(0) = 0.1$ m and $\dot{x}(0) = 0.05$ m/sec. Obtain the motion of the mass subjected to the initial condition. (Assume no external forcing function.)

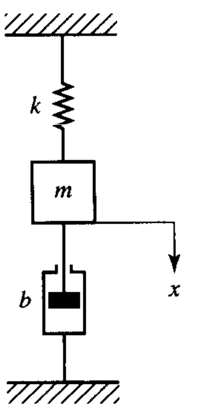



**Figure 4–28**
Mechanical system.

The system equation is

$$m\ddot{x} + b\dot{x} + kx = 0$$

with the initial conditions $x(0) = 0.1$ m and $\dot{x}(0) = 0.05$ m/sec. The Laplace transform of the system equation gives

$$m[s^2X(s) - sx(0) - \dot{x}(0)] + b[sX(s) - x(0)] + kX(s) = 0$$

or

$$(ms^2 + bs + k)X(s) = mx(0)s + m\dot{x}(0) + bx(0)$$

Solving this last equation for $X(s)$ and substituting the given numerical values, we obtain

$$X(s) = \frac{mx(0)s + m\dot{x}(0) + bx(0)}{ms^2 + bs + k}$$

$$= \frac{0.1s + 0.35}{s^2 + 3s + 2}$$

This equation can be written as

$$X(s) = \frac{0.1s^2 + 0.35s}{s^2 + 3s + 2}\frac{1}{s}$$

Hence the motion of the mass $m$ may be obtained as the unit-step response of the following system:

$$G(s) = \frac{0.1s^2 + 0.35s}{s^2 + 3s + 2}$$

MATLAB Program 4–9 will give a plot of the motion of the mass. The plot is shown in Figure 4–29.

```
MATLAB Program 4-9

% ---------- Response to initial conditions ----------

% ***** System response to initial conditions is converted to
% a unit-step response by modifying the numerator polynomial *****

% ***** Enter the numerator and denominator of the transfer
% function G(s) *****

num = [0.1   0.35   0];
den = [1   3   2];

% ***** Enter the following step-response command *****

step(num,den)

% ***** Enter grid and title of the plot *****

grid
title('Response of Spring-Mass-Damper System to Initial Conditions')
```

**Response to initial condition (state-space approach, case 1).** Consider the system defined by

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}, \qquad \mathbf{x}(0) = \mathbf{x}_0 \tag{4–49}$$

Let us obtain the response $\mathbf{x}(t)$ when the initial condition $\mathbf{x}(0)$ is specified. (There is no external input function acting on this system.) Assume that $\mathbf{x}$ is an $n$-vector.

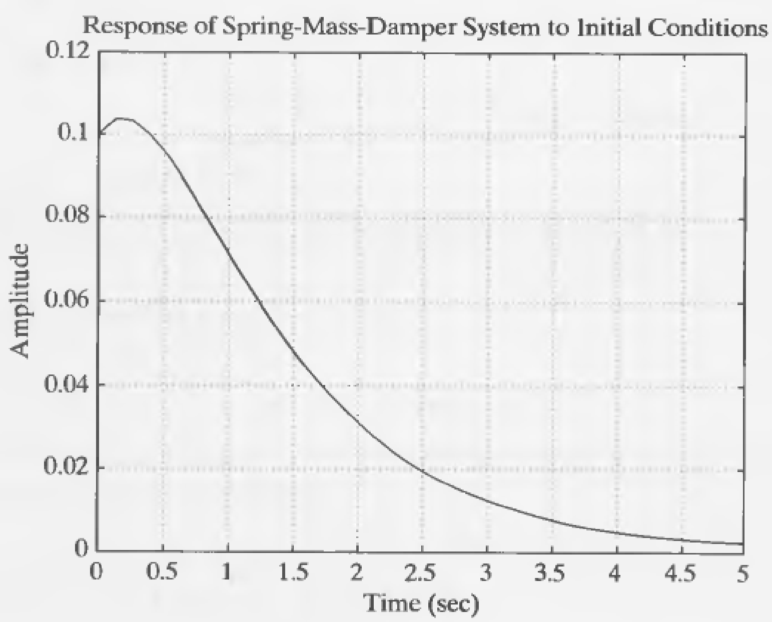



**Figure 4–29** Response of the mechanical system considered in Example 4–7.

First, take Laplace transforms of both sides of Equation (4–49).

$$s\mathbf{X}(s) - \mathbf{x}(0) = \mathbf{A}\mathbf{X}(s)$$

This equation can be rewritten as

$$s\mathbf{X}(s) = \mathbf{A}\mathbf{X}(s) + \mathbf{x}(0) \tag{4–50}$$

Taking the inverse Laplace transform of Equation (4–50), we obtain

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{x}(0)\,\delta(t) \tag{4–51}$$

(Notice that by taking the Laplace transform of a differential equation and then by taking the inverse Laplace transform of the Laplace-transformed equation we generate a differential equation that involves the initial condition.)

Now define

$$\dot{\mathbf{z}} = \mathbf{x} \tag{4–52}$$

Then Equation (4–51) can be written as

$$\ddot{\mathbf{z}} = \mathbf{A}\dot{\mathbf{z}} + \mathbf{x}(0)\,\delta(t) \tag{4–53}$$

By integrating Equation (4–53) with respect to $t$, we obtain

$$\dot{\mathbf{z}} = \mathbf{A}\mathbf{z} + \mathbf{x}(0)1(t) = \mathbf{A}\mathbf{z} + \mathbf{B}u \tag{4–54}$$

where

$$\mathbf{B} = \mathbf{x}(0), \qquad u = 1(t)$$

Referring to Equation (4–52), the state $\mathbf{x}(t)$ is given by $\dot{\mathbf{z}}(t)$. Thus,

$$\mathbf{x} = \dot{\mathbf{z}} = \mathbf{A}\mathbf{z} + \mathbf{B}u \tag{4–55}$$

Equation (4–55) gives the response to the initial condition.

Summarizing, the response of Equation (4–49) to the initial condition $\mathbf{x}(0)$ is obtained by solving the following state-space equations:

$$\dot{\mathbf{z}} = \mathbf{A}\mathbf{z} + \mathbf{B}u$$

$$\mathbf{x} = \mathbf{A}\mathbf{z} + \mathbf{B}u$$

where

$$\mathbf{B} = \mathbf{x}(0), \qquad u = 1(t)$$

MATLAB commands to obtain the response curves in one diagram are given next.

```
[x,z,t] = step(A,B,A,B);
x1 = [1   0   0 . . . 0]*x';
x2 = [0   1   0 . . . 0]*x';
      .
      .
      .

xn = [0   0   0 . . . 1]*x';
plot(t,xl,t,x2, . . . ,t,xn)
```

**Response to initial condition (state-space approach, case 2).** Consider the system defined by

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}, \qquad \mathbf{x}(0) = \mathbf{x}_0 \tag{4–56}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} \tag{4–57}$$

(Assume that $\mathbf{x}$ is an $n$-vector and $\mathbf{y}$ is an $m$-vector.)

Similar to case 1, by defining

$$\dot{\mathbf{z}} = \mathbf{x}$$

we can obtain the following equation:

$$\dot{\mathbf{z}} = \mathbf{A}\mathbf{z} + \mathbf{x}(0)1(t) = \mathbf{A}\mathbf{z} + \mathbf{B}u \tag{4–58}$$

where

$$\mathbf{B} = \mathbf{x}(0), \qquad u = 1(t)$$

Noting that $\mathbf{x} = \dot{\mathbf{z}}$, Equation (4–57) can be written as

$$\mathbf{y} = \mathbf{C}\dot{\mathbf{z}} \tag{4–59}$$

By substituting Equation (4–58) into Equation (4–59), we obtain

$$\mathbf{y} = \mathbf{C}(\mathbf{A}\mathbf{z} + \mathbf{B}u) = \mathbf{C}\mathbf{A}\mathbf{z} + \mathbf{C}\mathbf{B}u \tag{4–60}$$

The solution of Equations (4–58) and (4–60) gives the response of the system to a given initial condition. MATLAB commands to obtain the response curves (output curves y1 versus t, y2 versus t, . . . , ym versus t) are shown next.

```
[y,z,t] = step(A,B,C*A,C*B)
y1 = [1   0   0 . . . 0]*y';
y2 = [0   1   0 . . . 0]*y';
  .
  .
  .
ym = [0   0   0 . . . 1]*y';
plot(t,y1,t,y2 . . . ,t,ym)
```

**EXAMPLE 4–8** Obtain the response of the system subjected to the given initial condition.

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -10 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

or

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}, \qquad \mathbf{x}(0) = \mathbf{x}_0$$

Obtaining the response of the system to the given initial condition becomes that of solving the unit-step response of the following system:

$$\dot{\mathbf{z}} = \mathbf{A}\mathbf{z} + \mathbf{B}u$$

$$\mathbf{x} = \mathbf{A}\mathbf{z} + \mathbf{B}u$$

where

$$\mathbf{B} = \mathbf{x}(0), \qquad u = 1(t)$$

Hence a possible MATLAB program for obtaining the response may be given as shown in MATLAB Program 4–10. The resulting response curves are shown in Figure 4–30.

MATLAB Program 4–10

```
A = [0   1;-10   -5];
B = [2;1];
[x,z,t] = step(A,B,A,B);
x1 = [1   0]*x';
x2 = [0   1]*x';
plot(t,x1,'o',t,x2,'-')
grid
title('Response to Initial Condition')
xlabel('t Sec')
ylabel('x1   x2')
```

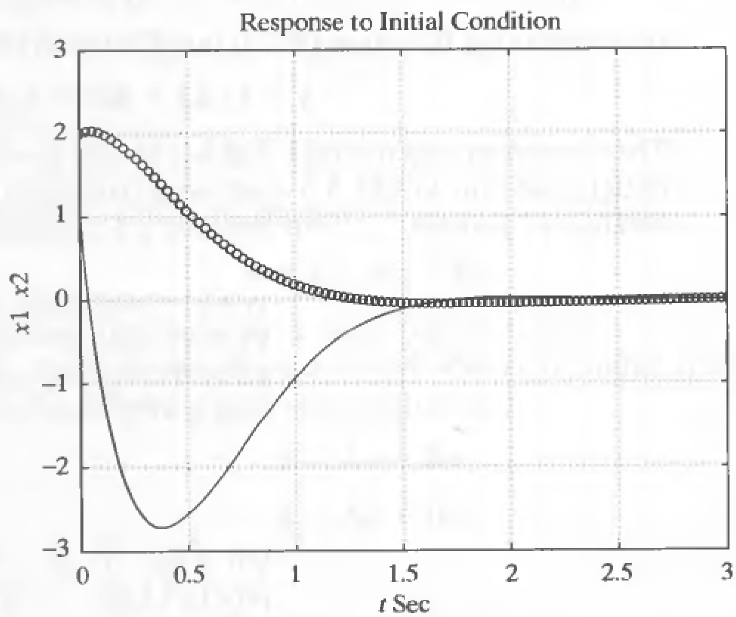



**Figure 4–30** Response of system in Example 4–8 to initial condition.

## 4–5 AN EXAMPLE PROBLEM SOLVED WITH MATLAB

The purpose of this section is to present a MATLAB solution of the response of a mechanical vibratory system. The mathematical model of the system is first developed, then the system is simulated using MATLAB for a continuous-time and a discrete-time approach, and response curves are generated for each approach.

**Mechanical vibratory system.** Consider the mechanical vibratory system shown in Figure 4–31(a). A wheel has a spring–mass–damper system hanging from it. The wheel is in a track that contains a flat (horizontal) portion, a slanted (downward at 45°)

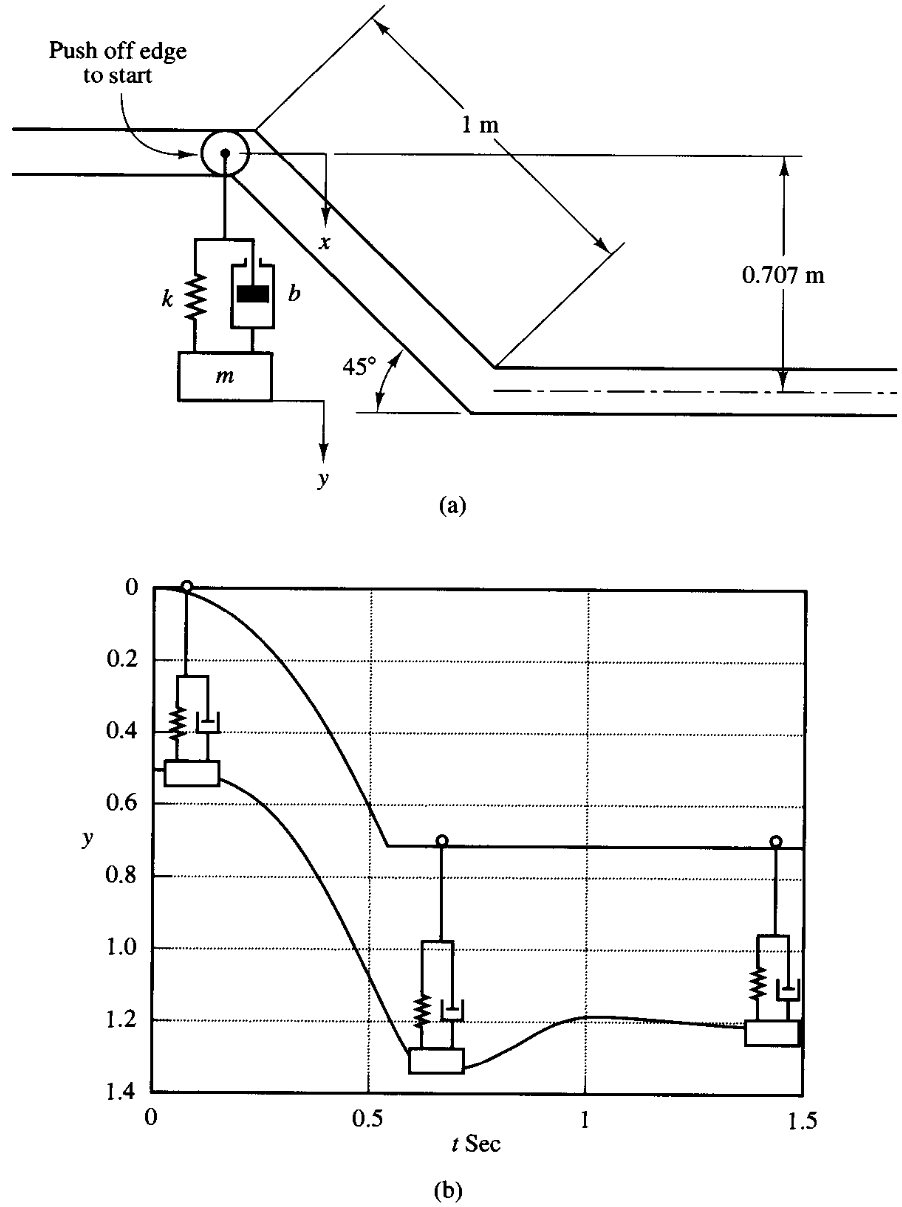



**Figure 4–31**
(a) Wheel with hanging mass-damper system; (b) dynamic response of the system.

portion, and another flat (horizontal) portion. We start the motion of the system by nudging the wheel over the edge of the ramp. As the wheel drops down the ramp for a total of 0.707 m (vertically measured), the mass $m$ hanging from the spring and damper drops with it, and the mass gains momentum that dissipates gradually. In this problem the wheel is assumed to slide on the slanted portion of the track without friction. On the second flat portion of the track, the wheel slides and rolls. The wheel continues to move on the flat portion of the track until it is stopped by an external means.

Assume the following numerical values for $m$, $b$, and $k$:

$$m = 4 \text{ kg}, \qquad b = 40 \text{ N-sec/m}, \qquad k = 400 \text{ N/m}$$

Assume also that the mass $m_p$ of the wheel is negligible compared with the mass $m$. Obtain $x(t)$, the vertical motion of the wheel. Then obtain $Y(s)$, the Laplace transform of $y(t)$, which represents the up and down motion of mass $m$. The coordinate $y$ is attached to the spring–mass–damper system as shown in Figure 4–31 and is measured from the equilibrium position of the system. The initial conditions are that $y(0) = 0$ and $\dot{y}(0) = 0$. Note that in this problem we are interested only in the vertical motions of the spring–mass–damper system. Note also that the system is frictionless with the exception of the damper, which relies on viscosity for its operation.

As the spring–mass–damper component travels down the ramp, it will undergo an acceleration produced as a result of the gravity force. When the spring–mass–damper reaches the level region at the bottom of the ramp, a shock will immediately be imposed on the spring–mass–damper component. It will, however, eventually come to a state of equilibrium following the impact due to the settling effects of the damper and spring. The dynamic response of this system is shown in Figure 4–31(b).

***Determination of $x(t)$.*** The system starts with zero initial velocity and follows the track. The input to the system is the vertical position $x$ along the track, and the output is the vertical position $y$ of the mass. Since we assume no sliding friction, referring to Figure 4–32(a) we have in the $z$ direction the following equation:

$$m\ddot{z} = mg \sin 45^\circ$$

or

$$\ddot{z} = 9.81 \times 0.707 = 6.9357$$

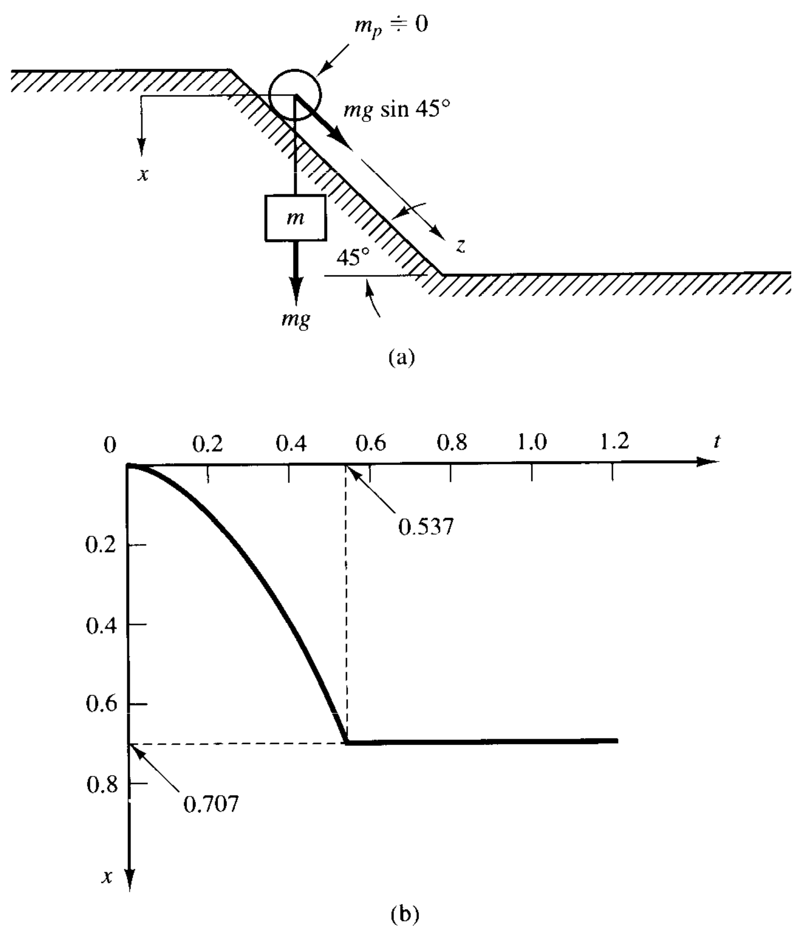



**Figure 4–32**
(a) Wheel with mass $m$ slides on inclined plane; (b) curve $x(t)$ versus $t$.

Let us define the time it takes for the wheel to move from $z = 0$ to $z = 1$ m as $t_1$. Then

$$z = 6.9357 \frac{t_1^2}{2} = 1$$

which yields

$$t_1 = 0.537 \text{ sec}$$

Thus $x(t)$ can be given as follows:

$$\begin{aligned} x(t) = 0.707z = 0.707 \times 3.4678t^2 &= 2.452t^2, && \text{for } 0 \leq t \leq 0.537 \\ &= 0.707, && \text{for } 0.537 < t \end{aligned}$$

It follows that from $t = 0.537$ sec to $t = \infty$ we have an input defined by a constant of 0.707. The position $x$ at the end of the ramp is 0.707 and it takes approximately 0.537 sec to get there. A curve $x(t)$ versus $t$ is shown in Figure 4–32(b). Note that the positive direction of $x(t)$ is vertically downward.

To get a better picture of the events taking place in the system, we need to look at the input, shown in Figure 4–32(b). The effects of gravity do not allow us to model the behavior of the system with an ordinary ramp, but rather a parabolic function, which is followed by a constant input.

***Determination of Transfer Function $Y(s)/X(s)$.*** Next, we shall first obtain the equation of motion for the system and then the transfer function $Y(s)/X(s)$. Since $y$ is measured from the equilibrium position, the system equation becomes

$$m\ddot{y} + b(\dot{y} - \dot{x}) + k(y - x) = 0$$

or

$$m\ddot{y} + b\dot{y} + ky = b\dot{x} + kx$$

where $x$ is the input to the system and $y$ is the output. By substituting the given numerical values for $m$, $b$, and $k$, we obtain

$$4\ddot{y} + 40\dot{y} + 400y = 40\dot{x} + 400x$$

or

$$\ddot{y} + 10\dot{y} + 100y = 10\dot{x} + 100x \tag{4–61}$$

The transfer function for the system can now be given by

$$\frac{Y(s)}{X(s)} = \frac{10s + 100}{s^2 + 10s + 100} \tag{4–62}$$

where the input $x(t)$ is given by

$$\begin{aligned} x(t) &= 2.452t^2, && 0 \leq t \leq 0.537 \\ &= 0.707, && 0.537 < t \end{aligned} \tag{4–63}$$

Our problem here is to use MATLAB to find the inverse Laplace transform of $Y(s)$ given by Equation (4–62). In what follows we consider two approaches. One is to work in the continuous-time domain using the step command. The other is to work in the discrete-time domain using the filter command. We shall first present the continuous-time approach and then the discrete-time approach.

**Computer simulation (continuous-time approach).** In the continuous-time approach we separate the time region into two parts; $0 \le t \le 0.537$ and $0.537 < t$.

For $0 \le t \le 0.537$:

$$x_1(t) = 2.452t^2$$

Hence

$$X_1(s) = \frac{2.452 \times 2}{s^3} = \frac{4.904}{s^3}$$

The output $Y(s)$ can then be given by

$$\begin{aligned} Y(s) &= \frac{10s + 100}{s^2 + 10s + 100}\frac{4.904}{s^3} \\ &= \frac{49.04s + 490.4}{s^4 + 10s^3 + 100s^2}\frac{1}{s} \end{aligned} \tag{4–64}$$

For $0.537 < t$:

$$x_2(t) = 0.707$$

Since

$$\frac{Y_2(s)}{X_2(s)} = \frac{10s + 100}{s^2 + 10s + 100}$$

the corresponding differential equation becomes

$$\ddot{y}_2 + 10\dot{y}_2 + 100y_2 = 10\dot{x}_2 + 100x_2$$

The Laplace transform of this last equation becomes

$$\begin{aligned} [s^2Y_2(s) - sy_2(0) - \dot{y}_2(0)] + 10[sY_2(s) - y_2(0)] + 100Y_2(s) \\ = 10[sX_2(s) - x_2(0)] + 100X_2(s) \end{aligned}$$

or

$$\begin{aligned} (s^2 + 10s + 100)Y_2(s) = (10s + 100)X_2(s) + sy_2(0) \\ + \dot{y}_2(0) + 10y_2(0) - 10x_2(0) \end{aligned}$$

Hence

$$Y_2(s) = \frac{10s + 100}{s^2 + 10s + 100} X_2(s) + \frac{sy_2(0) + \dot{y}_2(0) + 10y_2(0) - 10x_2(0)}{s^2 + 10s + 100}$$

The initial conditions are found from $y_2(0) = y_1(0.537)$ and $\dot{y}_2(0) = \dot{y}_1(0.537)$. Therefore,

$$Y_2(s) = \frac{10s + 100}{s^2 + 10s + 100} \frac{0.707}{s} + \frac{s^2[y_1(0.537)] + [\dot{y}_1(0.537) + 10y_1(0.537) - 10(0.707)]s}{s^2 + 10s + 100} \frac{1}{s}$$

or

$$Y_2(s) = \frac{10s + 100}{s^2 + 10s + 100} \frac{0.707}{s} + \frac{s^2[\,y1(537)] + [\,y1\mathrm{dot}(537) + 10y1(537) - 7.07]s}{s^2 + 10s + 100} \frac{1}{s} \tag{4-65}$$

where

$$y1(537) = y_1(0.537), \qquad y1\mathrm{dot}(537) = \dot{y}_1(0.537)$$

A MATLAB program to obtain the response $y(t)$ based on the continuous-time approach is given in MATLAB Program 4–11. The resulting response curve $y(t)$ versus $t$, as well as the input $x(t)$ versus $t$, is shown in Figure 4–33.

MATLAB Program 4–11

```
% ---------- Continuous-time approach ----------

% ***** Obtain y1 and y1dot *****

num1 = [0   0   0   49.04   490.4];
den1 = [1   10   100   0   0];
t1 = 0:0.001:0.537;
y1 = step(num1,den1,t1);
num2 = [0   0   49.04   490.4];
den2 = [1   10   100   0];
y1dot = step(num2,den2,t1);

% ***** Determine the initial values of y1(537) and y1dot(537)
% for the second part. The initial values for the second
% part are output y2_ini = y1(537) and y2dot_ini = y1dot(537) *****
```

```
y2_ini = y1(537);
y2dot_ini = y1dot(537);
t2 = 0.538:0.001:1.5;
num3 = [0   7.07   70.7];
den3 = [1   10   100];
num4 = [y1(537)   y1dot(537)+10*y1(537)-7.07   0];
y2o = step(num3,den3,t2);
y2i = step(num4,den3,t2);
y2 = y2o + y2i;
y = [y1'   y2'];
t = [t1   t2];
plot(t,-y,'.')

hold

Current plot held

x1 = 2.452*t1.^2;
x2 = 0.707*ones(size(t2));
x = [x1   x2];
plot(t,-x,'-')
grid
title('Response of System (Continuous-Time Approach)')
xlabel('t Sec')
ylabel('Input x and Output y')
text(0.2,-0.54,'Input x')
text(0.47,-0.25,'Output y')
```

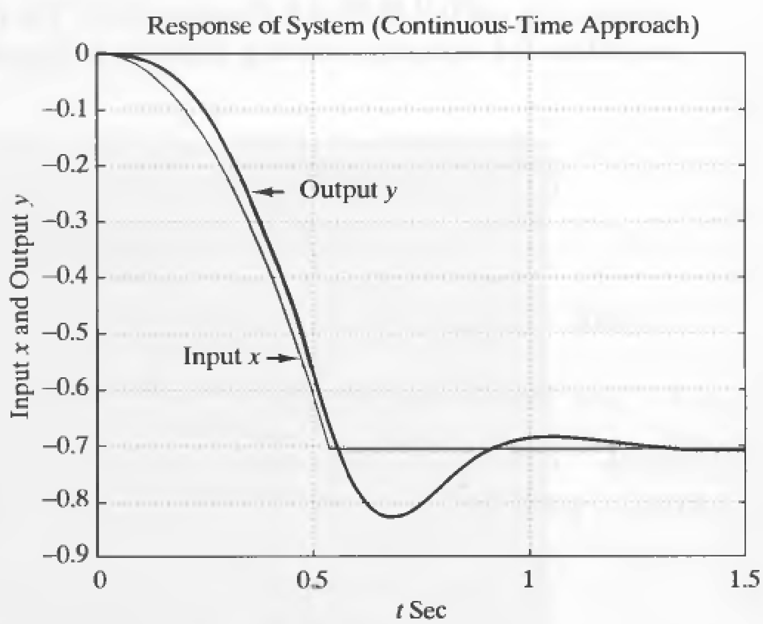



**Figure 4–33** Input $x(t)$ and output $y(t)$ obtained by the continuous-time approach.

It is noted that when plotting multiple curves on one diagram we may use the "hold" command. If we enter the command "hold" in the computer, the screen will show

```
hold

Current plot held
```

To release the held plot, enter the command "hold" again. Then the current plot will be released as shown next.

```
hold

Current plot held

hold

Current plot released
```

**Computer simulation (discrete-time approach).** The continuous-time transfer function may be converted to a pulse transfer function (discrete-time transfer function) using general formulas. The simpler method is to convert the continuous-time transfer function to a pulse transfer function using MATLAB commands. The first step is to convert the continuous-time transfer function to a continuous-time set of state-space equations using the MATLAB command [A,B,C,D] = tf2ss(num,den). The state-space equations can then be converted from continuous-time to discrete-time using the command [G,H] = c2d(A,B,Ts), where Ts is the desired time step (sampling period). The discrete-time state-space equations are converted to a pulse transfer function with the command [numz,denz] = ss2tf(G,H,C,D).

In the present case we choose $T = 0.001$ sec. The input function $x(t)$ must first be discretized. The continuous-time input function was determined to be

$$x(t) = 2.452t^2, \qquad \text{for } 0 \leqslant t \leqslant 0.537$$

$$x(t) = 0.707, \qquad \text{for } 0.537 < t$$

Note that we define $x$ as an array of points in MATLAB. This array initially follows $x(t) = 2.452t^2$ and, after $t = 0.537$ sec, follows $x(t) = 0.707$. We assume that the time region is $0 \leqslant t \leqslant 1.5$.

The acceleration input in the first part can be written as

```
k1 = 0:537;
x1 = [2.452*(0.001*k1).^2]
```

where k1 represents a time count and x1 is the first part of the complete input function. (There are 538 calculation points from the initial position until the input reaches

0.707 m.) For the second part of the input, we need a step function with magnitude 0.707. After time 0.537 sec,

```
k2 = 538:1500;
x2 = [0.707*ones(size(k2))]
```

(There are 963 points from 0.538 sec through 1.5 sec, inclusive.) The next step is to transform both inputs to one complete input:

```
x = [x1    x2];
```

(The two input equations are transformed into a single vector in order to appear as a single entry in the filter command argument.)

Now we can use the filter command, assigning a variable y,

```
y = filter(numz,denz,x);
```

and plot the response y(t), as well as the original input itself, x(t), taking care with the time intervals using t:

```
t = 0:1500;
plot(t/1000,-y,'.',t/1000,-x,'-')
```

(We divide t by 1000 because the time step is 0.001 sec.) Note also that the plotted input and output functions are negated. (Otherwise, we would have a positive acceleration input and response, which would be incorrect.)

A possible MATLAB program using the discrete-time approach is shown in MATLAB Program 4–12. The resulting response curves $x(t)$ versus $t$ and $y(t)$ versus $t$ are shown in Figure 4–34.

MATLAB Program 4–12

```
% ---------- Discrete-time approach ----------

% ***** Convert continuous-time transfer function to discrete-time
% transfer function (pulse transfer function) by choosing the time
% step (sampling period) to be 0.001 sec *****

num = [0   10   100];
den = [1   10   100];
[A,B,C,D] = tf2ss(num,den);
[G,H] = c2d(A,B,0.001);
[numz,denz] = ss2tf(G,H,C,D);

% ***** Enter the inputs x1 and x2 *****
```

```
k1 = 0:537;
x1 = [2.452*(0.001*k1).^2];
k2 = 538:1500;
x2 = [0.707*ones(size(k2))];
x = [x1    x2];

% ***** Using the filter command, obtain the response y *****

y = filter(numz,denz,x);
t = 0:1500;
plot(t/1000,-y,'.',t/1000,-x,'-')
grid
title('Response of System (Discrete-Time Approach)')
xlabel('t Sec')
ylabel('Input x  and  Output y')
text(0.2,-0.54,'Input x')
text(0.47,-0.25, 'Output y')
```

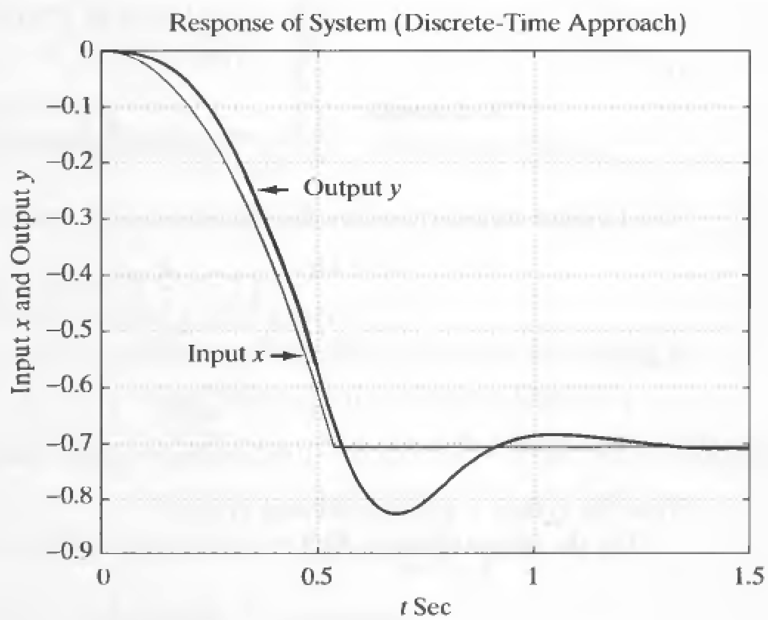



**Figure 4–34** Input $x(t)$ and output $y(t)$ obtained by the discrete-time approach.

## EXAMPLE PROBLEMS AND SOLUTIONS

**A-4-1.** In the system of Figure 4–35, $x(t)$ is the input displacement and $\theta(t)$ is the output angular displacement. Assume that the masses involved are negligibly small and that all motions are restricted to be small; therefore, the system can be considered linear. The initial conditions for $x$ and $\theta$ are zeros, or $x(0-) = 0$ and $\theta(0-) = 0$. Show that this system is a differentiating element. Then obtain the response $\theta(t)$ when $x(t)$ is a unit-step input.

**Solution.** The equation for the system is

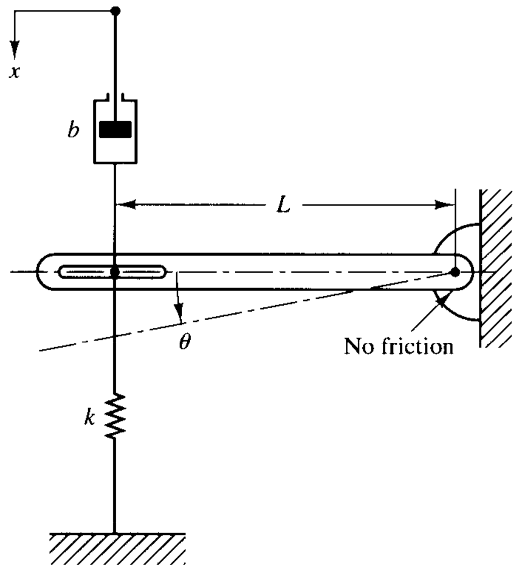



**Figure 4–35**
Mechanical system.

$$b(\dot{x} - L\dot{\theta}) = kL\theta$$

or

$$L\dot{\theta} + \frac{k}{b}L\theta = \dot{x}$$

The Laplace transform of this last equation, using zero initial conditions, gives

$$\left(Ls + \frac{k}{b}L\right)\Theta(s) = sX(s)$$

And so

$$\frac{\Theta(s)}{X(s)} = \frac{1}{L}\frac{s}{s + (k/b)}$$

Thus the system is a differentiating system.

For the unit-step input $X(s) = 1/s$, the output $\Theta(s)$ becomes

$$\Theta(s) = \frac{1}{L}\frac{1}{s + (k/b)}$$

The inverse Laplace transform of $\Theta(s)$ gives

$$\theta(t) = \frac{1}{L}e^{-(k/b)t}$$

Note that if the value of $k/b$ is large the response $\theta(t)$ approaches a pulse signal as shown in Figure 4–36.

**A-4-2.** Consider the mechanical system shown in Figure 4–37. Suppose that the system is at rest initially $[x(0) = 0, \dot{x}(0) = 0]$, and at $t = 0$ it is set into motion by a unit-impulse force. Obtain a mathematical model for the system. Then find the motion of the system.

**Solution.** The system is excited by a unit-impulse input. Hence

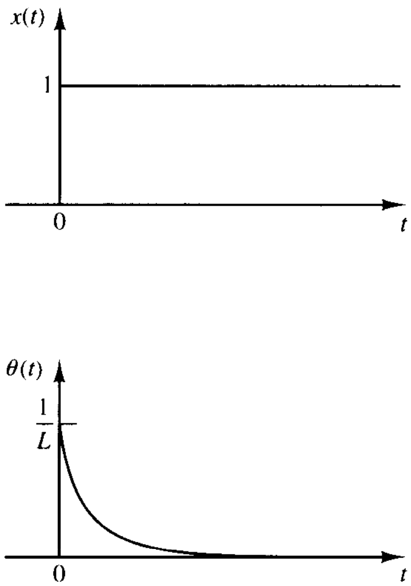



**Figure 4–36**
Unit-step input and the response of the mechanical system shown in Figure 4–35.

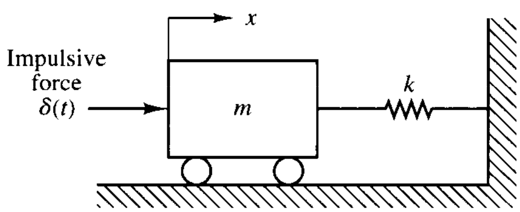



**Figure 4–37**
Mechanical system.

$$m\ddot{x} + kx = \delta(t)$$

This is a mathematical model for the system.

Taking the Laplace transform of both sides of this last equation gives

$$m[s^2X(s) - sx(0) - \dot{x}(0)] + kX(s) = 1$$

By substituting the initial conditions $x(0) = 0$ and $\dot{x}(0) = 0$ into this last equation and solving for $X(s)$, we obtain

$$X(s) = \frac{1}{ms^2 + k}$$

The inverse Laplace transform of $X(s)$ becomes

$$X(t) = \frac{1}{\sqrt{mk}} \sin \sqrt{\frac{k}{m}}\, t$$

The oscillation is a simple harmonic motion. The amplitude of the oscillation is $1/\sqrt{mk}$.

**A-4-3.** Obtain the closed-loop transfer function for the positional servo system shown in Figure 4–38. Assume that the input and output of the system are the input shaft position and the output shaft position, respectively. Assume the following numerical values for system constants:

$r$ = angular displacement of the reference input shaft, radians
$c$ = angular displacement of the output shaft, radians
$\theta$ = angular displacement of the motor shaft, radians
$K_0$ = gain of the potentiometric error detector = $24/\pi$ V/rad

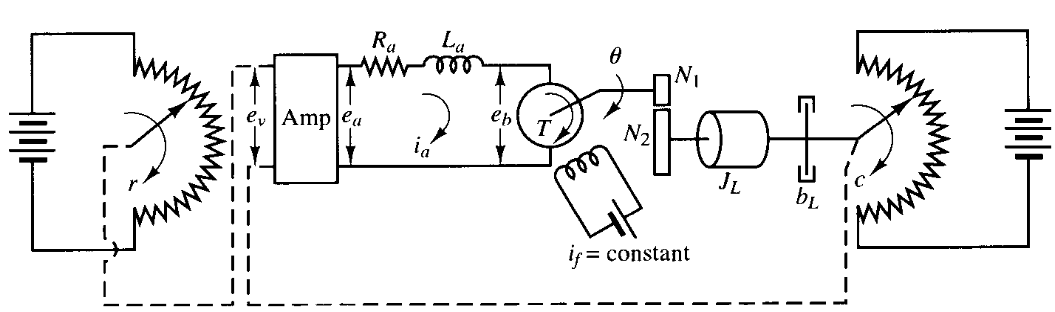



**Figure 4–38**
Positional servo system.

$K_1$ = amplifier gain = 10 V/V
$e_a$ = armature voltage, V
$e_b$ = back emf, V
$R_a$ = armature-winding resistance = 0.2 Ω
$L_a$ = armature-winding inductance = negligible
$i_a$ = armature-winding current, A
$K_3$ = back emf constant = $5.5 \times 10^{-2}$ V-sec/rad
$K_2$ = motor torque constant = $6 \times 10^{-5}$ N-m/A
$J_m$ = moment of inertia of the motor referred to the motor shaft = $1 \times 10^{-5}$ kg-m$^2$
$b_m$ = viscous-friction coefficient of the motor referred to the motor shaft = negligible
$J_L$ = moment of inertia of the load referred to the output shaft = $4.4 \times 10^{-3}$ kg-m$^2$
$b_L$ = viscous-friction coefficient of the load referred to the output shaft = $4 \times 10^{-2}$N-m/rad/sec
$n$ = gear ratio $N_1/N_2 = \frac{1}{10}$

**Solutiou.** The equivalent moment of inertia $J_0$ and equivalent viscous friction coefficient $b_0$ referred to the motor shaft are, respectively,

$$J_0 = J_m + n^2 J_L$$
$$= 1 \times 10^{-5} + 4.4 \times 10^{-5} = 5.4 \times 10^{-5}$$
$$b_0 = b_m + n^2 b_L$$
$$= 4 \times 10^{-4}$$

Referring to Equation (4–16), we obtain

$$\frac{C(s)}{E(s)} = \frac{K_m}{s(T_m s + 1)}$$

where

$$K_m = \frac{K_0 K_1 K_2 n}{R_a b_0 + K_2 K_3} = \frac{7.64 \times 10 \times 6 \times 10^{-5} \times 0.1}{(0.2)(4 \times 10^{-4}) + (6 \times 10^{-5})(5.5 \times 10^{-2})} = 5.5$$
$$T_m = \frac{R_a J_0}{R_a b_0 + K_2 K_3} = \frac{(0.2)(5.4 \times 10^{-5})}{(0.2)(4 \times 10^{-4}) + (6 \times 10^{-5})(5.5 \times 10^{-2})} = 0.13$$

Thus,

$$\frac{C(s)}{E(s)} = \frac{5.5}{s(0.13s + 1)} \tag{4–66}$$

Using Equation (4–66), we can draw the block diagram of the system as shown in Figure 4–39. The closed-loop transfer function of the system is

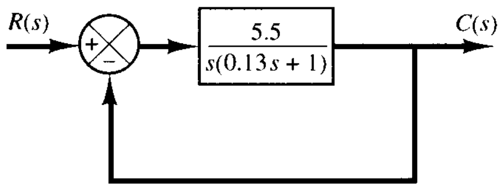



**Figure 4–39**
Block diagram of the system shown in Figure 4–38.

$$\frac{C(s)}{R(s)} = \frac{5.5}{0.13s^2 + s + 5.5} = \frac{42.3}{s^2 + 7.69s + 42.3}$$

**A-4-4.** Gear trains are often used in servo systems to reduce speed, to magnify torque, or to obtain the most efficient power transfer by matching the driving member to the given load.

Consider the gear train system shown in Figure 4–40. In this system, a load is driven by a motor through the gear train. Assuming that the stiffness of the shafts of the gear train is infinite (there is neither backlash nor elastic deformation) and that the number of teeth on each gear is proportional to the radius of the gear, obtain the equivalent moment of inertia and equivalent viscous-friction coefficient referred to the motor shaft and referred to the load shaft.

In Figure 4–40 the numbers of teeth on gears 1, 2, 3, and 4 are $N_1, N_2, N_3$, and $N_4$, respectively. The angular displacements of shafts, 1, 2, and 3 are $\theta_1, \theta_2$, and $\theta_3$, respectively. Thus, $\theta_2/\theta_1 = N_1/N_2$ and $\theta_3/\theta_2 = N_3/N_4$. The moment of inertia and viscous-friction coefficient of each gear train component are denoted by $J_1, b_1$; $J_2, b_2$; and $J_3, b_3$; respectively. ($J_3$ and $b_3$ include the moment of inertia and friction of the load.)

**Solution.** For this gear train system, we can obtain the following three equations: For shaft 1,

$$J_1\ddot{\theta}_1 + b_1\dot{\theta}_1 + T_1 = T_m \tag{4-67}$$

where $T_m$ is the torque developed by the motor and $T_1$ is the load torque on gear 1 due to the rest of the gear train. For shaft 2,

$$J_2\ddot{\theta}_2 + b_2\dot{\theta}_2 + T_3 = T_2 \tag{4-68}$$

where $T_2$ is the torque transmitted to gear 2 and $T_3$ is the load torque on gear 3 due to the rest of the gear train. Since the work done by gear 1 is equal to that of gear 2,

$$T_1\theta_1 = T_2\theta_2 \qquad \text{or} \qquad T_2 = T_1\frac{N_2}{N_1}$$

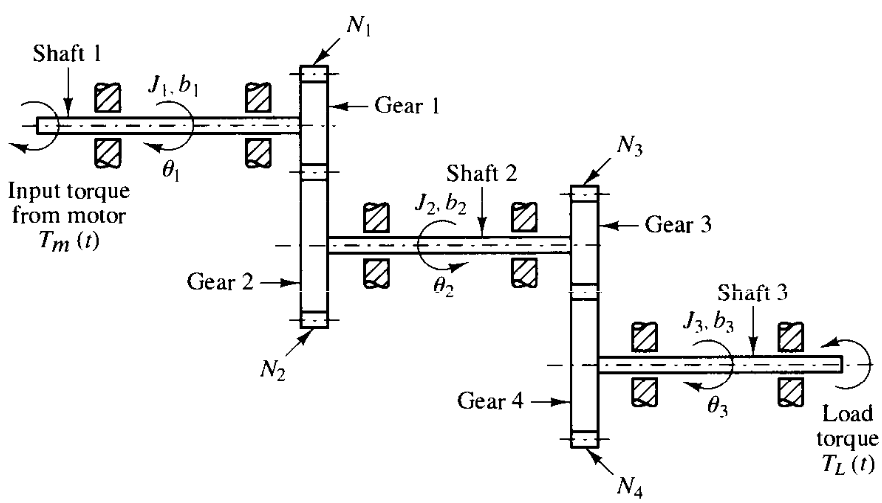



**Figure 4–40**
Gear train system.

If $N_1/N_2 < 1$, the gear ratio reduces the speed as well as magnifies the torque. For the third shaft,

$$J_3\ddot{\theta}_3 + b_3\dot{\theta}_3 + T_L = T_4 \tag{4-69}$$

where $T_L$ is the load torque and $T_4$ is the torque transmitted to gear 4. $T_3$ and $T_4$ are related by

$$T_4 = T_3\frac{N_4}{N_3}$$

and $\theta_3$ and $\theta_1$ are related by

$$\theta_3 = \theta_2\frac{N_3}{N_4} = \theta_1\frac{N_1}{N_2}\frac{N_3}{N_4}$$

Elimination of $T_1$, $T_2$, $T_3$, and $T_4$ from Equations (4–67), (4–68), and (4–69) yields

$$J_1\ddot{\theta}_1 + b_1\dot{\theta}_1 + \frac{N_1}{N_2}(J_2\ddot{\theta}_2 + b_2\dot{\theta}_2) + \frac{N_1N_3}{N_2N_4}(J_3\ddot{\theta}_3 + b_3\dot{\theta}_3 + T_L) = T_m$$

Eliminating $\theta_2$ and $\theta_3$ from this last equation and writing the resulting equation in terms of $\theta_1$ and its time derivatives, we obtain

$$\begin{aligned}\left[J_1 + \left(\frac{N_1}{N_2}\right)^2 J_2 + \left(\frac{N_1}{N_2}\right)^2\left(\frac{N_3}{N_4}\right)^2 J_3\right]\ddot{\theta}_1 \\ + \left[b_1 + \left(\frac{N_1}{N_2}\right)^2 b_2 + \left(\frac{N_1}{N_2}\right)^2\left(\frac{N_3}{N_4}\right)^2 b_3\right]\dot{\theta}_1 + \left(\frac{N_1}{N_2}\right)\left(\frac{N_3}{N_4}\right)T_L = T_m\end{aligned} \tag{4-70}$$

Thus, the equivalent moment of inertia and viscous-friction coefficient of the gear train referred to shaft 1 are given, respectively, by

$$J_{1\text{eq}} = J_1 + \left(\frac{N_1}{N_2}\right)^2 J_2 + \left(\frac{N_1}{N_2}\right)^2\left(\frac{N_3}{N_4}\right)^2 J_3$$

$$b_{1\text{eq}} = b_1 + \left(\frac{N_1}{N_2}\right)^2 b_2 + \left(\frac{N_1}{N_2}\right)^2\left(\frac{N_3}{N_4}\right)^2 b_3$$

Similarly, the equivalent moment of inertia and viscous-friction coefficient of the gear train referred to the load shaft (shaft 3) are given, respectively, by

$$J_{3\text{eq}} = J_3 + \left(\frac{N_4}{N_3}\right)^2 J_2 + \left(\frac{N_2}{N_1}\right)^2\left(\frac{N_4}{N_3}\right)^2 J_1$$

$$b_{3\text{eq}} = b_3 + \left(\frac{N_4}{N_3}\right)^2 b_2 + \left(\frac{N_2}{N_1}\right)^2\left(\frac{N_4}{N_3}\right)^2 b_1$$

The relationship between $J_{1\text{eq}}$ and $J_{3\text{eq}}$ is thus

$$J_{1\text{eq}} = \left(\frac{N_1}{N_2}\right)^2\left(\frac{N_3}{N_4}\right)^2 J_{3\text{eq}}$$

and that between $b_{1\text{eq}}$ and $b_{3\text{eq}}$ is

$$b_{1\text{eq}} = \left(\frac{N_1}{N_2}\right)^2\left(\frac{N_3}{N_4}\right)^2 b_{3\text{eq}}$$

The effect of $J_2$ and $J_3$ on an equivalent moment of inertia is determined by the gear ratios $N_1/N_2$ and $N_3/N_4$. For speed-reducing gear trains, the ratios $N_1/N_2$ and $N_3/N_4$ are usually less than unity.

If $N_1/N_2 \ll 1$ and $N_3/N_4 \ll 1$, then the effect of $J_2$ and $J_3$ on the equivalent moment of inertia $J_{1eq}$ is negligible. Similar comments apply to the equivalent viscous-friction coefficient $b_{1eq}$ of the gear train. In terms of the equivalent moment of inertia $J_{1eq}$ and equivalent viscous-friction coefficient $b_{1eq}$, Equation (4–70) can be simplified to give

$$J_{1eq}\ddot{\theta}_1 + b_{1eq}\dot{\theta}_1 + nT_L = T_m$$

where

$$n = \frac{N_1}{N_2}\frac{N_3}{N_4}$$

**A-4-5.** Show that the torque-to-inertia ratios referred to the motor shaft and to the load shaft differ from each other by a factor of $n$. Show also that the torque squared-to-inertia ratios referred to the motor shaft and to the load shaft are the same.

**Solution.** Suppose that $T_{max}$ is the maximum torque that can be produced on the motor shaft. Then the torque-to-inertia ratio referred to the motor shaft is

$$\frac{T_{max}}{J_m + n^2 J_L}$$

where $J_m$ = moment of inertia of the rotor
$J_L$ = moment of inertia of the load
$n$ = gear ratio

The torque-to-inertia ratio referred to the load shaft is

$$\frac{\dfrac{T_{max}}{n}}{J_L + \dfrac{J_m}{n^2}} = \frac{nT_{max}}{J_m + n^2 J_L}$$

Clearly, they differ from each other by a factor of $n$. Hence, in comparing torque-to-inertia ratios of motors, we find it necessary to specify which shaft is the reference.

Note that the ratio of torque squared to inertia referred to the motor shaft is

$$\frac{T_{max}^2}{J_m + n^2 J_L}$$

and that referred to the load shaft is

$$\frac{\dfrac{T_{max}^2}{n^2}}{J_L + \dfrac{J_m}{n^2}} = \frac{T_{max}^2}{J_m + n^2 J_L}$$

These two ratios are clearly the same.

**A-4-6.** When the system shown in Figure 4–41(a) is subjected to a unit-step input, the system output responds as shown in Figure 4–41(b). Determine the values of $K$ and $T$ from the response curve.

**Solution.** The maximum overshoot of 25.4% corresponds to $\zeta = 0.4$. From the response curve we have

$$t_p = 3$$

Consequently,

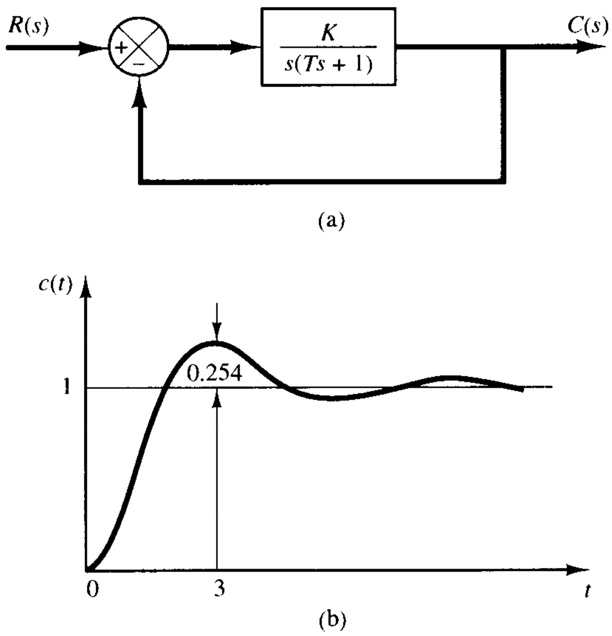



**Figure 4–41**
(a) Closed-loop system; (b) unit-step response curve.

$$t_p = \frac{\pi}{\omega_d} = \frac{\pi}{\omega_n\sqrt{1-\zeta^2}} = \frac{\pi}{\omega_n\sqrt{1-0.4^2}} = 3$$

It follows that

$$\omega_n = 1.14$$

From the block diagram we have

$$\frac{C(s)}{R(s)} = \frac{K}{Ts^2 + s + K}$$

from which

$$\omega_n = \sqrt{\frac{K}{T}}, \qquad 2\zeta\omega_n = \frac{1}{T}$$

Therefore, the values of $T$ and $K$ are determined as

$$T = \frac{1}{2\zeta\omega_n} = \frac{1}{2 \times 0.4 \times 1.14} = 1.09$$

$$K = \omega_n^2 T = 1.14^2 \times 1.09 = 1.42$$

**A-4-7.** Determine the values of $K$ and $k$ of the closed-loop system shown in Figure 4–42 so that the maximum overshoot in unit-step response is 25% and the peak time is 2 sec. Assume that $J = 1$ kg-m$^2$.

**Solution.** The closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{K}{Js^2 + Kks + K}$$

By substituting $J = 1$ kg-m$^2$ into this last equation, we have

$$\frac{C(s)}{R(s)} = \frac{K}{s^2 + Kks + K}$$

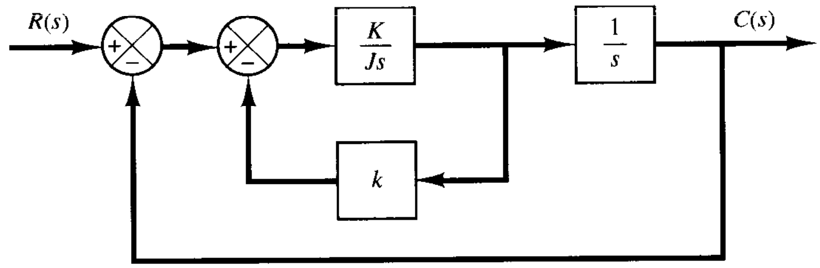



**Figure 4–42**
Closed-loop system.

Note that

$$\omega_n = \sqrt{K}, \qquad 2\zeta\omega_n = Kk$$

The maximum overshoot $M_p$ is

$$M_p = e^{-\zeta\pi/\sqrt{1-\zeta^2}}$$

which is specified as 25%. Hence

$$e^{-\zeta\pi/\sqrt{1-\zeta^2}} = 0.25$$

from which

$$\frac{\zeta\pi}{\sqrt{1-\zeta^2}} = 1.386$$

or

$$\zeta = 0.404$$

The peak time $t_p$ is specified as 2 sec. And so

$$t_p = \frac{\pi}{\omega_d} = 2$$

or

$$\omega_d = 1.57$$

Then the undamped natural frequency $\omega_n$ is

$$\omega_n = \frac{\omega_d}{\sqrt{1-\zeta^2}} = \frac{1.57}{\sqrt{1-0.404^2}} = 1.72$$

Therefore, we obtain

$$K = \omega_n^2 = 1.72^2 = 2.95 \text{ N-m}$$

$$k = \frac{2\zeta\omega_n}{K} = \frac{2 \times 0.404 \times 1.72}{2.95} = 0.471 \text{ sec}$$

**A-4-8.** What is the unit-step response of the system shown in Figure 4–43?

**Solution.** The closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{10s + 10}{s^2 + 10s + 10}$$

For the unit-step input $[R(s) = 1/s]$, we have

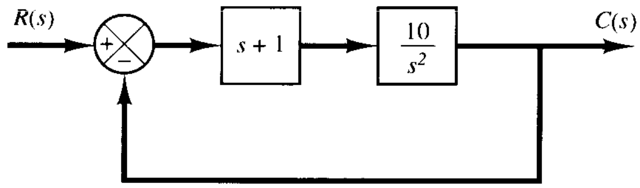



**Figure 4–43**
Closed-loop system.

$$C(s) = \frac{10s + 10}{s^2 + 10s + 10}\frac{1}{s}$$

$$= \frac{10s + 10}{(s + 5 + \sqrt{15})(s + 5 - \sqrt{15})s}$$

$$= \frac{-4 - \sqrt{15}}{3 + \sqrt{15}}\frac{1}{s + 5 + \sqrt{15}} + \frac{-4 + \sqrt{15}}{3 - \sqrt{15}}\frac{1}{s + 5 - \sqrt{15}} + \frac{1}{s}$$

The inverse Laplace transform of $C(s)$ gives

$$c(t) = -\frac{4 + \sqrt{15}}{3 + \sqrt{15}}e^{-(5+\sqrt{15})t} + \frac{4 - \sqrt{15}}{-3 + \sqrt{15}}e^{-(5-\sqrt{15})t} + 1$$

$$= -1.1455e^{-8.87t} + 0.1455e^{-1.13t} + 1$$

Clearly, the output will not exhibit any oscillation. The response curve exponentially approaches the final value $c(\infty) = 1$.

**A-4-9.** Figure 4–44(a) shows a mechanical vibratory system. When 2 lb of force (step input) is applied to the system, the mass oscillates, as shown in Figure 4–44(b). Determine $m$, $b$, and $k$ of the system from this response curve. The displacement $x$ is measured from the equilibrium position.

**Solution.** The transfer function of this system is

$$\frac{X(s)}{P(s)} = \frac{1}{ms^2 + bs + k}$$

Since

$$P(s) = \frac{2}{s}$$

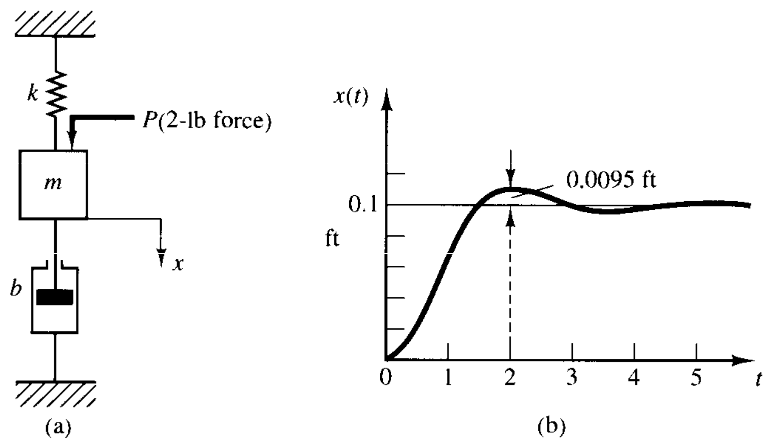



**Figure 4–44**
(a) Mechanical vibratory system; (b) step-response curve.

we obtain

$$X(s) = \frac{2}{s(ms^2 + bs + k)}$$

It follows that the steady-state value of $x$ is

$$x(\infty) = \lim_{s \to 0} sX(s) = \frac{2}{k} = 0.1 \text{ ft}$$

Hence

$$k = 20 \text{ lb}_f/\text{ft}$$

Note that $M_p = 9.5\%$ corresponds to $\zeta = 0.6$. The peak time $t_p$ is given by

$$t_p = \frac{\pi}{\omega_d} = \frac{\pi}{\omega_n \sqrt{1 - \zeta^2}} = \frac{\pi}{0.8\omega_n}$$

The experimental curve shows that $t_p = 2$ sec. Therefore,

$$\omega_n = \frac{3.14}{2 \times 0.8} = 1.96 \text{ rad/sec}$$

Since $\omega_n^2 = k/m = 20/m$, we obtain

$$m = \frac{20}{\omega_n^2} = \frac{20}{1.96^2} = 5.2 \text{ slugs} = 166 \text{ lb}$$

(Note that 1 slug = 1 $\text{lb}_f\text{-sec}^2/\text{ft}$.) Then $b$ is determined from

$$2\zeta\omega_n = \frac{b}{m}$$

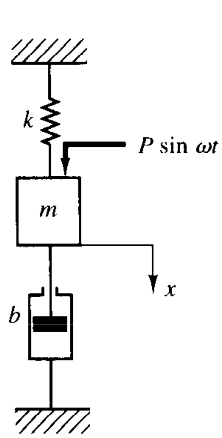
or

$$b = 2\zeta\omega_n m = 2 \times 0.6 \times 1.96 \times 5.2 = 12.2 \text{ lb}_f/\text{ft/sec}$$

**A-4-10.** Assuming that the mechanical system shown in Figure 4–45 is at rest before excitation force $P \sin \omega t$ is given, derive the complete solution $x(t)$ and the steady-state solution $x_{ss}(t)$. The displacement $x$ is measured from the equilibrium position. Assume that the system is underdamped.

**Solution.** The equation of motion for the system is

$$m\ddot{x} + b\dot{x} + kx = P \sin \omega t$$



**Fignre 4–45**
Mechanical system.

Noting that $x(0) = 0$ and $\dot{x}(0) = 0$, the Laplace transform of this equation is

$$(ms^2 + bs + k)X(s) = P \frac{\omega}{s^2 + \omega^2}$$

or

$$X(s) = \frac{P\omega}{(s^2 + \omega^2)} \frac{1}{(ms^2 + bs + k)}$$

Since the system is underdamped, $X(s)$ can be written as follows:

$$X(s) = \frac{P\omega}{m} \frac{1}{s^2 + \omega^2} \frac{1}{s^2 + 2\zeta\omega_n s + \omega_n^2}, \qquad \text{where } 0 < \zeta < 1$$

where $\omega_n = \sqrt{k/m}$ and $\zeta = b/(2\sqrt{mk})$. $X(s)$ can be expanded as

$$X(s) = \frac{P\omega}{m}\left(\frac{as + c}{s^2 + \omega^2} + \frac{-as + d}{s^2 + 2\zeta\omega_n s + \omega_n^2}\right)$$

By simple calculations it can be found that

$$a = \frac{-2\zeta\omega_n}{(\omega_n^2 - \omega^2)^2 + 4\zeta^2\omega_n^2\omega^2}, \qquad c = \frac{(\omega_n^2 - \omega^2)}{(\omega_n^2 - \omega^2)^2 + 4\zeta^2\omega_n^2\omega^2}, \qquad d = \frac{4\zeta^2\omega_n^2 - (\omega_n^2 - \omega^2)}{(\omega_n^2 - \omega^2)^2 + 4\zeta^2\omega_n^2\omega^2}$$

Hence

$$X(s) = \frac{P\omega}{m}\frac{1}{(\omega_n^2 - \omega^2)^2 + 4\zeta^2\omega_n^2\omega^2}\left[\frac{-2\zeta\omega_n s + (\omega_n^2 - \omega^2)}{s^2 + \omega^2} + \frac{2\zeta\omega_n(s + \zeta\omega_n) + 2\zeta^2\omega_n^2 - (\omega_n^2 - \omega^2)}{s^2 + 2\zeta\omega_n s + \omega_n^2}\right]$$

The inverse Laplace transform of $X(s)$ gives

$$x(t) = \frac{P\omega}{m[(\omega_n^2 - \omega^2)^2 + 4\zeta^2\omega_n^2\omega^2]}\left[-2\zeta\omega_n \cos\omega t + \frac{(\omega_n^2 - \omega^2)}{\omega}\sin\omega t + 2\zeta\omega_n e^{-\zeta\omega_n t}\cos\omega_n\sqrt{1 - \zeta^2}\,t + \frac{2\zeta^2\omega_n^2 - (\omega_n^2 - \omega^2)}{\omega_n\sqrt{1 - \zeta^2}}e^{-\zeta\omega_n t}\sin\omega_n\sqrt{1 - \zeta^2}\,t\right]$$

At steady state ($t \to \infty$) the terms involving $e^{-\zeta\omega_n t}$ approach zero. Thus at steady state

$$\begin{aligned} x_{ss}(t) &= \frac{P\omega}{m[(\omega_n^2 - \omega^2)^2 + 4\zeta^2\omega_n^2\omega^2]}\left(-2\zeta\omega_n\cos\omega t + \frac{\omega_n^2 - \omega^2}{\omega}\sin\omega t\right) \\ &= \frac{P\omega}{(k - m\omega^2)^2 + b^2\omega^2}\left(-b\cos\omega t + \frac{k - m\omega^2}{\omega}\sin\omega t\right) \\ &= \frac{P}{\sqrt{(k - m\omega^2)^2 + b^2\omega^2}}\sin\left(\omega t - \tan^{-1}\frac{b\omega}{k - m\omega^2}\right) \end{aligned}$$

**A-4-11.** Consider the unit-step response of the second-order system

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

The amplitude of the exponentially damped sinusoid changes as a geometric series. At time $t = t_p = \pi/\omega_d$, the amplitude is equal to $e^{-(\sigma/\omega_d)\pi}$. After one oscillation, or at $t = t_p + 2\pi/\omega_d = 3\pi/\omega_d$, the amplitude is equal to $e^{-(\sigma/\omega_d)3\pi}$; after another cycle of oscillation, the amplitude is $e^{-(\sigma/\omega_d)5\pi}$. The logarithm of the ratio of successive amplitudes is called the *logarithmic decrement.* Determine the logarithmic decrement for this second-order system. Describe a method for experimental determination of the damping ratio from the rate of decay of the oscillation.

**Solution.** Let us define the amplitude of the output oscillation at $t = t_i$ to be $x_i$, where $t_i = t_p + (i - 1)T$ ($T$ = period of oscillation). The amplitude ratio per one period of damped oscillation is

$$\frac{x_1}{x_2} = \frac{e^{-(\sigma/\omega_d)\pi}}{e^{-(\sigma/\omega_d)3\pi}} = e^{2(\sigma/\omega_d)\pi} = e^{2\zeta\pi/\sqrt{1-\zeta^2}}$$

Thus, the logarithmic decrement $\delta$ is

$$\delta = \ln \frac{x_1}{x_2} = \frac{2\zeta\pi}{\sqrt{1 - \zeta^2}}$$

It is a function only of the damping ratio $\zeta$. Thus, the damping ratio $\zeta$ can be determined by use of the logarithmic decrement.

In the experimental determination of the damping ratio $\zeta$ from the rate of decay of the oscillation, we measure the amplitude $x_1$ at $t = t_p$ and amplitude $x_n$ at $t = t_p + (n - 1)T$. Note that it is necessary to choose $n$ large enough so that the ratio $x_1/x_n$ is not near unity. Then

$$\frac{x_1}{x_n} = e^{(n-1)2\zeta\pi/\sqrt{1-\zeta^2}}$$

or

$$\ln \frac{x_1}{x_n} = (n - 1)\frac{2\zeta\pi}{\sqrt{1 - \zeta^2}}$$

Hence

$$\zeta = \frac{\dfrac{1}{n-1}\left(\ln \dfrac{x_1}{x_n}\right)}{\sqrt{4\pi^2 + \left[\dfrac{1}{n-1}\left(\ln \dfrac{x_1}{x_n}\right)\right]^2}}$$

**A-4-12.** In the system shown in Figure 4–46, the numerical values of $m$, $b$, and $k$ are given as $m = 1$ kg, $b = 2$ N-sec/m, and $k = 100$ N/m. The mass is displaced 0.05 m and released without initial velocity. Find the frequency observed in the vibration. In addition, find the amplitude four cycles later. The displacement $x$ is measured from the equilibrium position.

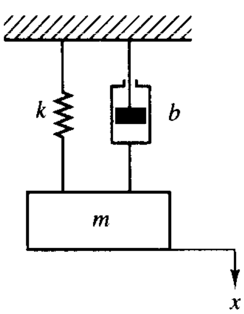



**Figure 4–46** Spring–mass–damper system.

**Solution.** The equation of motion for the system is

$$m\ddot{x} + b\dot{x} + kx = 0$$

Substituting the numerical values for $m$, $b$, and $k$ into this equation gives

$$\ddot{x} + 2\dot{x} + 100x = 0$$

where the initial conditions are $x(0) = 0.05$ and $\dot{x}(0) = 0$. From this last equation the undamped natural frequency $\omega_n$ and the damping ratio $\zeta$ are found to be

$$\omega_n = 10, \qquad \zeta = 0.1$$

The frequency actually observed in the vibration is the damped natural frequency $\omega_d$.

$$\omega_d = \omega_n\sqrt{1 - \zeta^2} = 10\sqrt{1 - 0.01} = 9.95 \text{ rad/sec}$$

In the present analysis, $\dot{x}(0)$ is given as zero. Thus, solution $x(t)$ can be written as

$$x(t) = x(0)e^{-\zeta\omega_n t}\left(\cos \omega_d t + \frac{\zeta}{\sqrt{1 - \zeta^2}} \sin \omega_d t\right)$$

It follows that at $t = nT$, where $T = 2\pi/\omega_d$,

$$x(nT) = x(0)e^{-\zeta\omega_n nT}$$

Consequently, the amplitude four cycles later becomes

$$x(4T) = x(0)e^{-\zeta\omega_n 4T} = x(0)e^{-(0.1)(10)(4)(0.6315)}$$
$$= 0.05e^{-2.526} = 0.05 \times 0.07998 = 0.004 \text{ m}$$

**A-4-13.** Consider a system whose closed-loop poles and closed-loop zero are located in the $s$ plane on a line parallel to the $j\omega$ axis, as shown in Figure 4–47. Show that the impulse response of such a system is a damped cosine function.

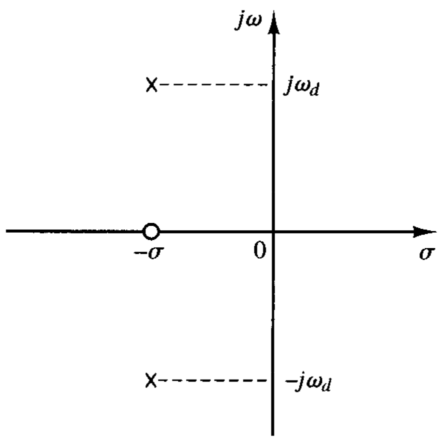



**Figure 4–47**
Closed-loop pole–zero configuration of system whose impulse response is a damped cosine function.

**Solution.** The closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{K(s+\sigma)}{(s+\sigma+j\omega_d)(s+\sigma-j\omega_d)}$$

For a unit-impulse input, $R(s) = 1$ and

$$C(s) = \frac{K(s+\sigma)}{(s+\sigma)^2 + \omega_d^2}$$

The inverse Laplace transform of $C(s)$ is

$$c(t) = Ke^{-\sigma t}\cos\omega_d t, \qquad \text{for } t \geq 0$$

which is a damped cosine function.

**A-4-14.** Consider the liquid-level control system shown in Figure 4–48. The controller is of the proportional type. The set point of the controller is fixed.

Draw a block diagram of the system, assuming that changes in the variables are small. Obtain the transfer function between the level of the second tank and the disturbance input $q_d$. Obtain the steady-state error when the disturbance $q_d$ is a unit-step function.

**Solution.** Figure 4–49(a) is a block diagram of this system when changes in the variables are small. Since the set point of the controller is fixed, $r = 0$. (Note that $r$ is the change in set point.)

To investigate the response of the level of the second tank subjected to a unit-step disturbance $q_d$, we find it convenient to modify the block diagram of Figure 4–49(a) to the one shown in Figure 4–49(b).

The transfer function between $H_2(s)$ and $Q_d(s)$ can be obtained as

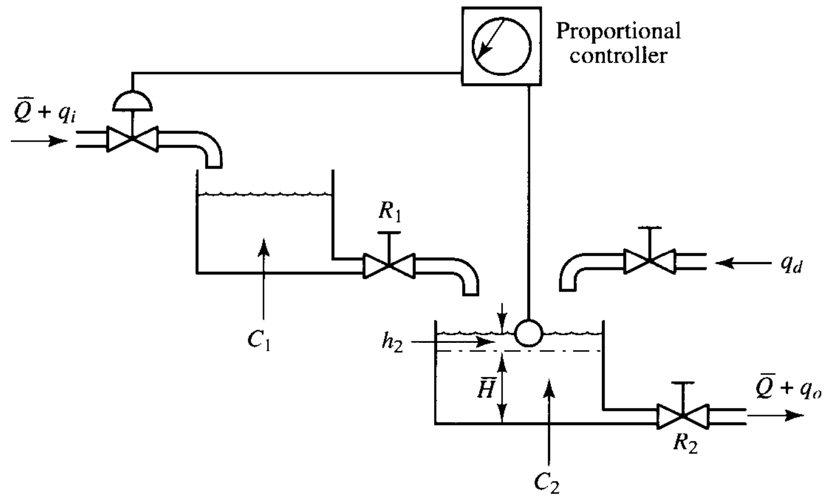



**Figure 4–48**
Liquid-level control system.

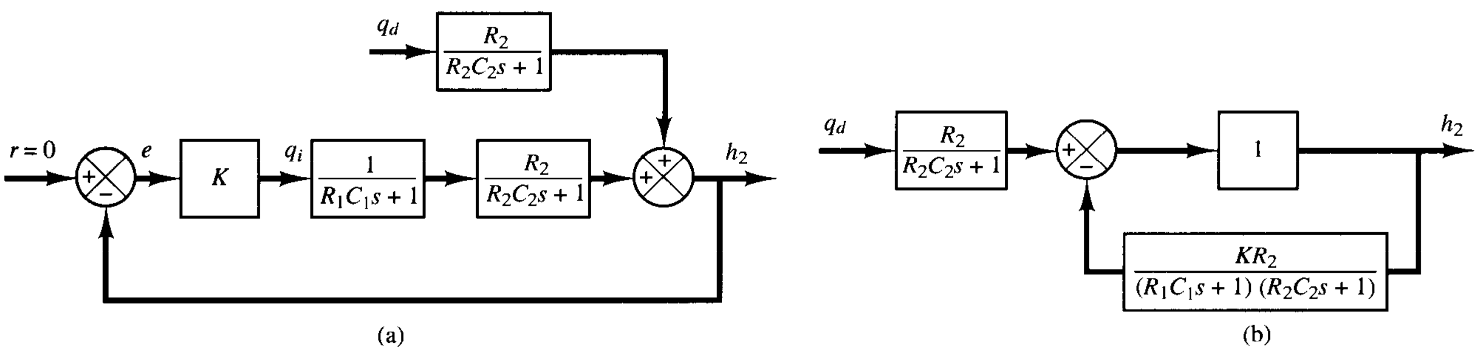



**Figure 4–49**
(a) Block diagram of the system shown in Figure 4–48; (b) modified block diagram.

$$\frac{H_2(s)}{Q_d(s)} = \frac{R_2(R_1C_1s + 1)}{(R_1C_1s + 1)(R_2C_2s + 1) + KR_2}$$

From this equation, the response $H_2(s)$ to the disturbance $Q_d(s)$ can be found. The effect of the controller is seen by the presence of $K$ in the denominator of this last equation.

For the unit-step disturbance $Q_d(s)$, we obtain

$$h_2(\infty) = \frac{R_2}{1 + KR_2}$$

or

$$\text{Steady-state error} = -\frac{R_2}{1 + KR_2}$$

The system exhibits offset in the response to a unit-step disturbance.

Note that the characteristic equation for the disturbance input and that for the reference input are the same. The characteristic equation for this system is

$$(R_1C_1s + 1)(R_2C_2s + 1) + KR_2 = 0$$

which can be modified to

$$s^2 + \left(\frac{R_1C_1 + R_2C_2}{R_1C_1R_2C_2}\right)s + \frac{1 + KR_2}{R_1C_1R_2C_2} = 0$$

The undamped natural frequency $\omega_n$ and the damping ratio $\zeta$ are given by

$$\omega_n = \sqrt{\frac{1 + KR_2}{R_1C_1R_2C_2}}, \qquad \zeta = \frac{R_1C_1 + R_2C_2}{2\sqrt{R_1C_1R_2C_2}\sqrt{1 + KR_2}}$$

Both the undamped natural frequency and the damping ratio depend on the value of the gain $K$. This gain must be adjusted so that the transient responses to both the reference input and disturbance input show reasonable damping and reasonable speed.

**A-4-15.** Consider the liquid-level control system shown in Figure 4–50. The inlet valve is controlled by a hydraulic integral controller. Assume that the steady-state inflow rate is $\bar{Q}$ and steady-state outflow rate is also $\bar{Q}$, the steady-state head is $\bar{H}$, steady-state pilot valve displacement is $\bar{X} = 0$, and steady-state valve position is $\bar{Y}$. We assume that the set point $\bar{R}$ corresponds to the steady-state head $\bar{H}$. The set point is fixed. Assume also that the disturbance inflow rate $q_d$, which is a small quantity, is applied to the water tank at $t = 0$. This disturbance causes the head to change from $\bar{H}$ to $\bar{H} + h$. This change results in a change in the outflow rate by $q_o$. Through the hydraulic controller, the change in head causes a change in the inflow rate from $\bar{Q}$ to $\bar{Q} + q_i$. (The integral controller tends to keep the head constant as much as possible in the presence of disturbances.) We assume that all changes are of small quantities.

Assuming the following numerical values for the system,

$$C = 2 \text{ m}^2, \qquad R = 0.5 \text{ sec/m}^2, \qquad K_v = 1 \text{ m}^2\text{/sec}$$
$$a = 0.25 \text{ m}, \qquad b = 0.75 \text{ m}, \qquad K_1 = 4 \text{ sec}^{-1}$$

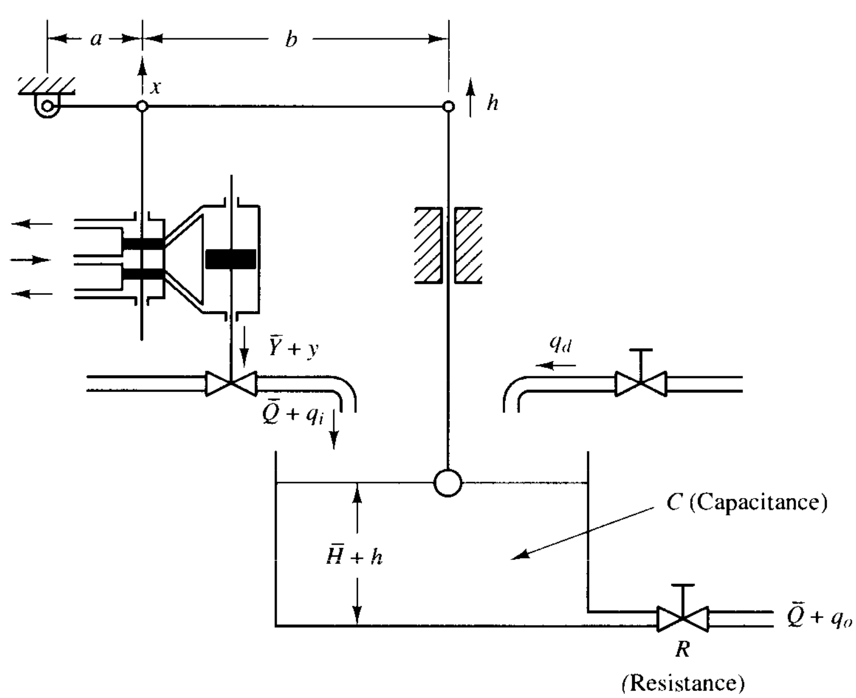



**Figure 4–50** Liquid-level control system.

obtain the response $h(t)$ when the disturbance input $q_d$ is a unit-step function. Also obtain this response $h(t)$ with MATLAB.

**Solution.** Since the increase of water in the tank during $dt$ seconds is equal to the net inflow to the tank during the same $dt$ seconds, we have

$$C\,dh = (q_i - q_o + q_d)\,dt \tag{4–71}$$

where

$$q_o = \frac{h}{R} \tag{4–72}$$

For the feedback lever mechanism, we have

$$x = \frac{a}{a+b}h \tag{4–73}$$

We assume that the velocity of the power piston (valve) is proportional to pilot valve displacement $x$, or

$$\frac{dy}{dt} = K_1 x \tag{4–74}$$

where $K_1$ is a positive constant. We also assume that the change in the inflow rate $q_i$ is negatively proportional to the change in the valve opening $y$, or

$$q_i = -K_v y \tag{4–75}$$

where $K_v$ is a positive constant.

Now we obtain the equations for the system as follows: From Equations (4–71), (4–72), and (4–75), we get

$$C\frac{dh}{dt} = -K_v y - \frac{h}{R} + q_d \tag{4–76}$$

From Equations (4–73) and (4–74), we have

$$\frac{dy}{dt} = \frac{K_1 a}{a+b}h \tag{4–77}$$

By substituting the given numerical values into Equations (4–76) and (4–77), we obtain

$$2\frac{dh}{dt} = -y - 2h + q_d$$

$$\frac{dy}{dt} = h$$

Taking the Laplace transforms of the preceding two equations, assuming zero initial conditions, we obtain

$$2sH(s) = -Y(s) - 2H(s) + Q_d(s)$$

$$sY(s) = H(s)$$

By eliminating $Y(s)$ from the last two equations and noting that the disturbance input is a unit-step function, or $Q_d(s) = 1/s$, we get

$$H(s) = \frac{s}{2s^2 + 2s + 1}\frac{1}{s} = \frac{0.5}{(s + 0.5)^2 + 0.5^2}$$

The inverse Laplace transform of $H(s)$ gives the time response $h(t)$.

$$h(t) = e^{-0.5t} \sin 0.5t$$

Notice that the unit-step disturbance input $q_d$ caused a transient error in the head which becomes zero at steady state. The integral controller thus eliminated the error caused by the disturbance input $q_d$.

**Plotting the response curve $h(t)$ with MATLAB.** Since the response $H(s)$ is given by

$$H(s) = \frac{s}{2s^2 + 2s + 1}\frac{1}{s}$$

MATLAB Program 4–13 may be used to obtain the response to the unit-step disturbance input. The resulting response curve is shown in Figure 4–51.

MATLAB Program 4–13

```
% ---------- Unit-step response ----------

% ***** Enter numerator and denominator of the transfer
% function *****

num = [0   1   0];
den = [2   2   1];

% ***** Enter step command *****

step(num,den)
grid
title('Unit-Step Response')
```

**A-4-16.** Consider the impulse response of the standard second-order system defined by

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

For a unit-impulse input, $R(s) = 1$. Thus

$$C(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} = \frac{\omega_n^2 s}{s^2 + 2\zeta\omega_n s + \omega_n^2}\frac{1}{s}$$

Consider the normalized system where $\omega_n = 1$. Then

$$C(s) = \frac{s}{s^2 + 2\zeta s + 1}\frac{1}{s}$$

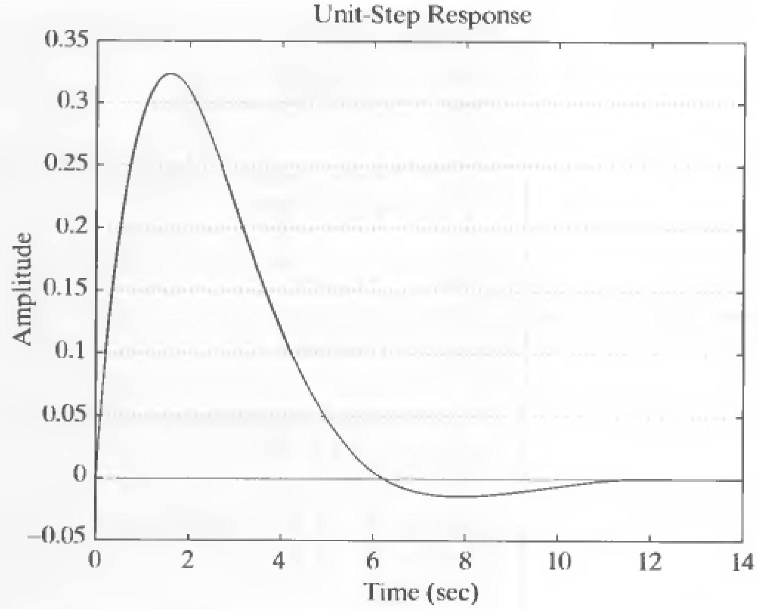



**Figure 4–51**
Response to unit-step disturbance input.

Consider five different values of zeta: $\zeta = 0.1, 0.3, 0.5, 0.7$, and $1.0$. Obtain the unit-impulse response curves for each zeta with MATLAB.

**Solution.** A MATLAB program for plotting the five unit-impulse response curves in one diagram is given in MATLAB Program 4–14. The resulting diagram is shown in Figure 4–52.

MATLAB Program 4–14

```
% ---------- Unit-impulse response ----------

% ***** Unit-impulse response curves for the normalized
% second-order system G(s) = 1/[s^2 + 2(zeta)s + 1] *****

% ***** The unit-impulse response is obtained as the
% unit-step response of sG(s) *****

% ***** The values of zeta considered here are 0.1, 0.3,
% 0.5, 0.7, and 1.0 *****

% ***** Enter the numerator and denominator of sG(s) for
% zeta = 0.1 *****

num = [0   1   0];
den1 = [1   0.2   1];

% ***** Specify the computing time points (such as t = 0:0.1:10).
% Then enter the step-response command step(num,den,t) and text
% command text(, ,' ') *****
```

```
t = 0:0.1:10;
step(num,den1,t);
text(2.2,0.88,'Zeta = 0.1')

% ***** Hold this plot and add other unit-impulse response
% curves to it *****

hold

Current plot held

% ***** Enter denominators of sG(s) for zeta = 0.3, 0.5,
% 0.7, and 1.0 *****

den2 = [1   0.6   1]; den3 = [1   1   1]; den4 = [1   1.4   1];
den5 = [1   2   1];

% ***** Superimpose on the held plot the unit-impulse response
% curves for zeta = 0.3, 0.5, 0.7, and 1.0 by entering
% successively the step-response command step(num,den,t)
% and text command text(, ,' ') *****

step(num,den2,t);
text(1.33,0.72,'0.3')
step(num,den3,t);
text(1.15,0.58,'0.5')
step(num,den4,t);
text(1.1,0.46,'0.7')
step(num,den5,t);
text(0.8,0.28,'1.0')

% ***** Enter grid and title of the plot *****

grid
title('Impulse-Response Curves for G(s) = 1/[s^2 + 2(zeta)s + 1]')

% ***** Clear hold on graphics *****

hold

Current plot released
```

From the unit-impulse response curves for different values of zeta, we may conclude that if the impulse response $c(t)$ does not change sign the system is either critically damped or overdamped, in which case the corresponding step response does not overshoot, but increases or decreases monotonically and approaches a constant value.

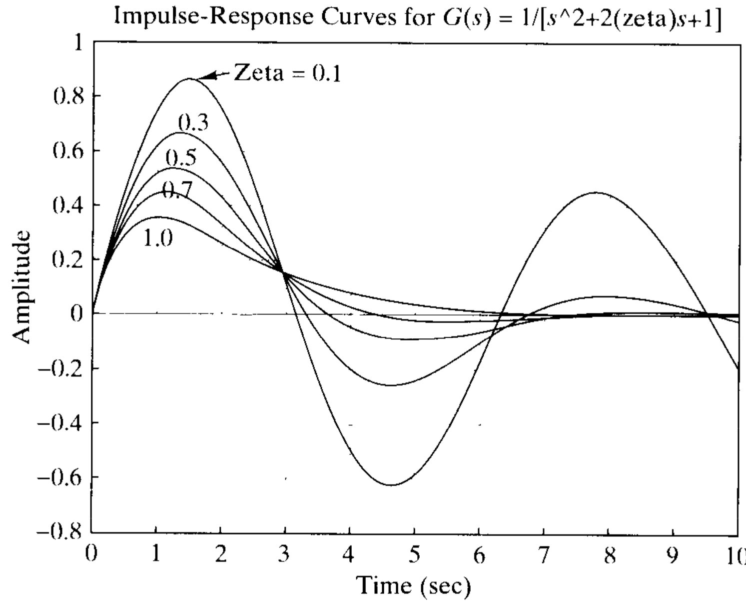



**Figure 4–52**
Unit-impulse response curves.

## PROBLEMS

**B-4-1.** A thermometer requires 1 min to indicate 98% of the response to a step input. Assuming the thermometer to be a first-order system, find the time constant.

If the thermometer is placed in a bath, the temperature of which is changing linearly at a rate of 10°/min, how much error does the thermometer show?

**B-4-2.** Consider the system shown in Figure 4–53. An armature-controlled dc servomotor drives a load consisting of the moment of inertia $J_L$. The torque developed by the motor is $T$. The angular displacements of the motor rotor and the load element are $\theta_m$ and $\theta$, respectively. The gear ratio is $n = \theta/\theta_m$. Obtain the transfer function $\Theta(s)/E_i(s)$.

**B-4-3.** Consider the system shown in Figure 4–54(a). The damping ratio of this system is 0.158 and the undamped natural frequency is 3.16 rad/sec. To improve the relative stability, we employ tachometer feedback. Figure 4–54(b) shows such a tachometer-feedback system.

Determine the value of $K_h$ so that the damping ratio of the system is 0.5. Draw unit-step response curves of both the original and tachometer-feedback systems. Also draw the error-versus-time curves for the unit-ramp response of both systems.

**B-4-4.** Obtain the unit-step response of a unity-feedback system whose open-loop transfer function is

$$G(s) = \frac{4}{s(s+5)}$$

**B-4-5.** Consider the unit-step response of a unity-feedback control system whose open-loop transfer function is

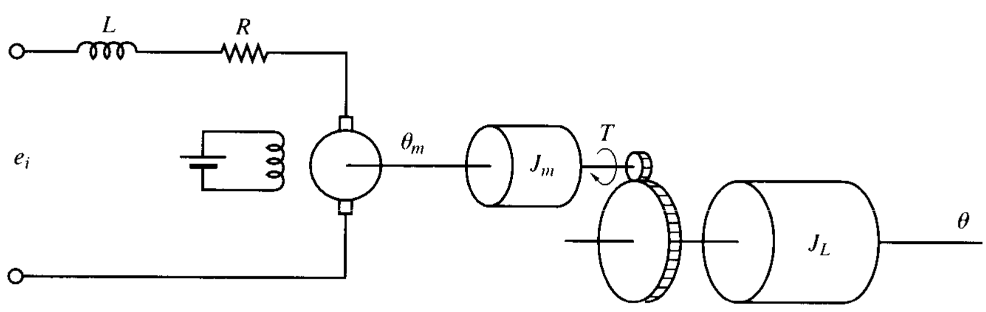



**Figure 4–53**
Armature-controlled dc servomotor system.

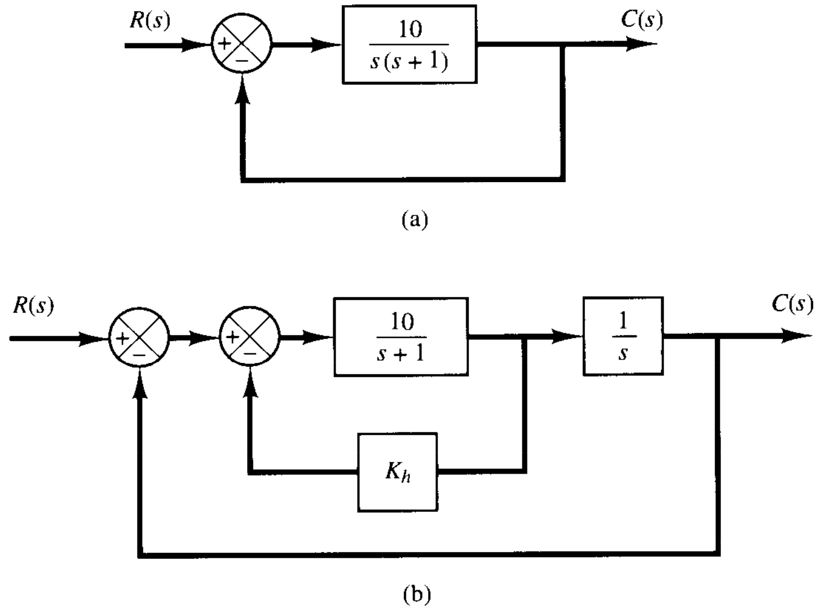



**Figure 4–54**
(a) Control system; (b) control system with tachometer feedback.

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

Obtain the rise time, peak time, maximum overshoot, and settling time.

**B-4-6.** Consider the closed-loop system given by

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

Determine the values of $\zeta$ and $\omega_n$ so that the system responds to a step input with approximately 5% overshoot and with a settling time of 2 sec. (Use the 2% criterion.)

**B-4-7.** Figure 4–55 is a block diagram of a space-vehicle attitude-control system. Assuming the time constant $T$ of the controller to be 3 sec and the ratio of torque to inertia $K/J$ to be $\frac{2}{9}$ rad²/sec², find the damping ratio of the system.

**B-4-8.** Consider the system shown in Figure 4–56. The system is initially at rest. Suppose that the cart is set into motion by an impulsive force whose strength is unity. Can it be stopped by another such impulsive force?

**B-4-9.** Obtain the unit-impulse response and the unit-step response of a unity-feedback system whose open-loop transfer function is

$$G(s) = \frac{2s+1}{s^2}$$

**B-4-10.** Consider the system shown in Figure 4–57. Show that the transfer function $Y(s)/X(s)$ has a zero in the right-

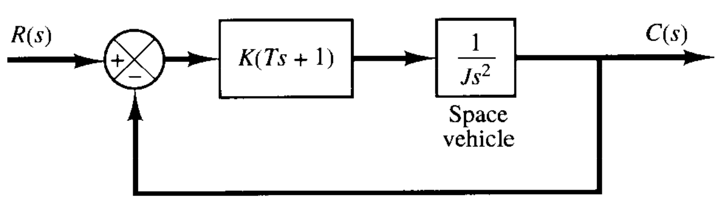



**Figure 4–55**
Space-vehicle attitude-control system.

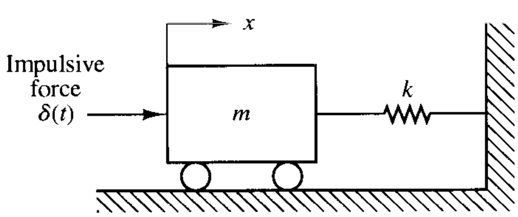



**Figure 4–56**
Mechanical system.

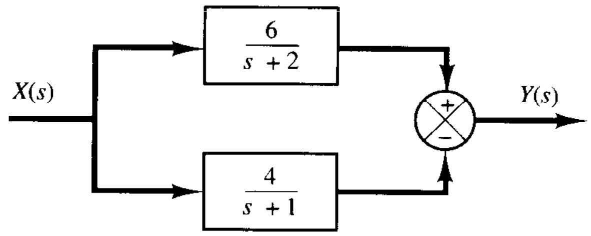



**Figure 4–57**
System with zero in the right-half $s$ plane.

half $s$ plane. Then obtain $y(t)$ when $x(t)$ is a unit step. Plot $y(t)$ versus $t$.

**B-4-11.** An oscillatory system is known to have a transfer function of the following form:

$$G(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

Assume that a record of a damped oscillation is available as shown in Figure 4–58. Determine the damping ratio $\zeta$ of the system from the graph.

**B-4-12.** Referring to the system shown in Figure 4–59, determine the values of $K$ and $k$ such that the system has a damping ratio $\zeta$ of 0.7 and an undamped natural frequency $\omega_n$ of 4 rad/sec.

**B-4-13.** Consider the system shown in Figure 4–60. Determine the value of $k$ such that the damping ratio $\zeta$ is 0.5. Then obtain the rise time $t_r$, peak time $t_p$, maximum overshoot $M_p$, and settling time $t_s$ in the unit-step response.

**B-4-14.** Using MATLAB, obtain the unit-step response, unit-ramp response, and unit-impulse response of the following system:

$$\frac{C(s)}{R(s)} = \frac{10}{s^2 + 2s + 10}$$

where $R(s)$ and $C(s)$ are Laplace transforms of the input $r(t)$ and output $c(t)$, respectively.

**B-4-15.** Using MATLAB, obtain the unit-step response, unit-ramp response, and unit-impulse response of the

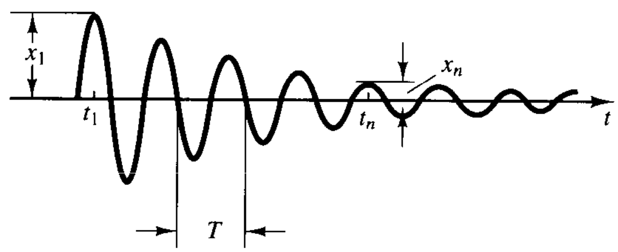



**Figure 4–58**
Decaying oscillation.

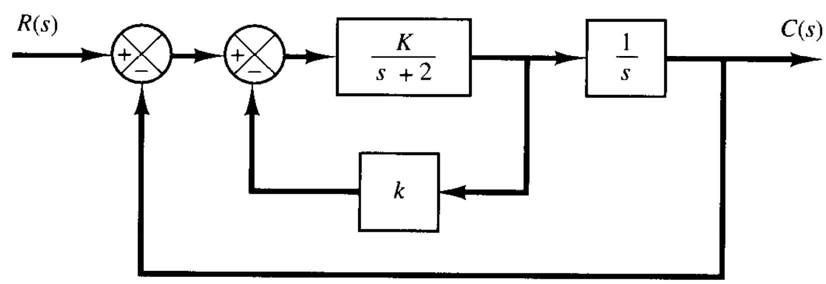



**Figure 4–59**
Closed-loop system.

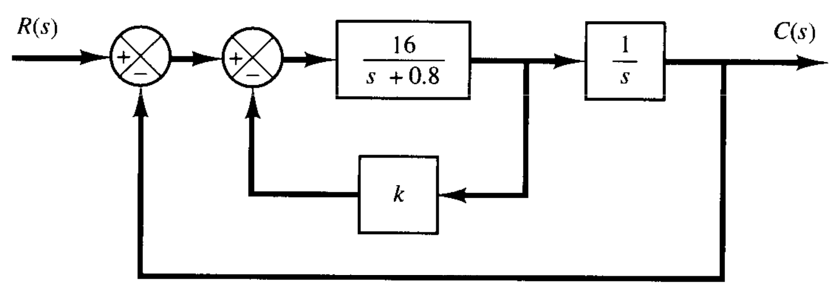



**Figure 4–60**
Block diagram of a system.

following system:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -1 & -0.5 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} u$$

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

where $u$ is the input and $y$ is the output.

**B-4-16.** Consider the same problem as discussed in Problem A-4-16. It is desired to use different marks for different curves (such as 'o', '×', '--', '-', ':'). Modify MATLAB Program 4–14 for this purpose.

# 5

# Basic Control Actions and Response of Control Systems

## 5-1 INTRODUCTION

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*.

In this chapter we shall first discuss the basic control actions used in industrial control systems. Then we shall discuss the effects of integral and derivative control actions on the system response. We shall next consider the response of higher-order systems. Any physical system will become unstable if any of the closed-loop poles lies in the right-half $s$ plane. To check the existence or nonexistence of such right-half plane poles, the Routh stability criterion is useful. We shall include discussions of this stability criterion in this chapter.

Many industrial automatic controllers are electronic, hydraulic, pneumatic, or their combinations. In this chapter we present principles of pneumatic controllers, hydraulic controllers, and electronic controllers.

The outline of this chapter follows: Section 5–1 has presented introductory material. Section 5–2 gives the basic control actions commonly used in industrial automatic controllers. Section 5–3 discusses the effects of integral and derivative control actions on system performance. Section 5–4 deals with higher-order systems, and Section 5–5 treats Routh's stability criterion. Sections 5–6 and 5–7 discuss pneumatic controllers and hydraulic controllers, respectively. Here we introduce the principle of operation of pneumatic and hydraulic controllers and methods for generating various control actions.

Section 5–8 treats electronic controllers using operational amplifiers. Section 5–9 discusses phase lead and phase lag in sinusoidal response. We derive the sinusoidal transfer function and show phase lead and phase lag that may occur in the sinusoidal response. Finally, in Section 5–10 we treat steady-state errors in system responses.

## 5–2 BASIC CONTROL ACTIONS

In this section we shall discuss the details of basic control actions used in industrial analog controllers. We shall begin with classifications of industrial analog controllers.

**Classifications of industrial controllers.** 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. 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.

**Automatic controller, actuator, and sensor (measuring element).** Figure 5–1 is a block diagram of an industrial control system, which 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 a pneumatic motor or valve, a hydraulic motor, or an electric motor. (The actuator is

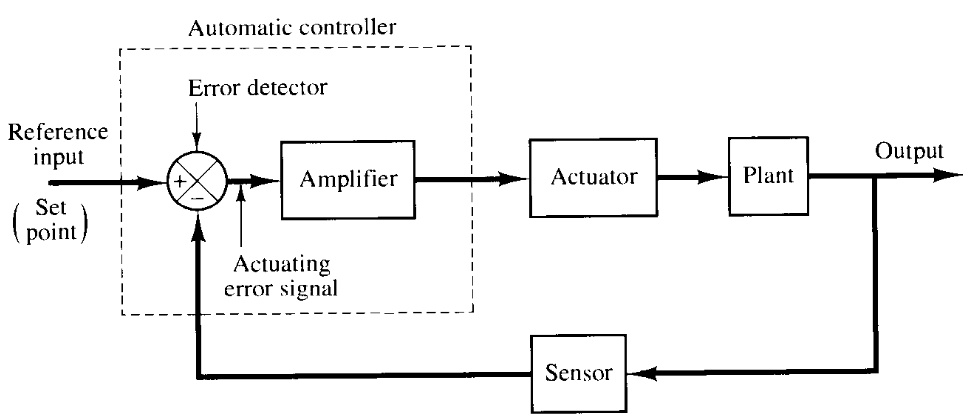



**Figure 5–1** Block diagram of an industrial control system, which consists of an automatic controller, an actuator, a plant, and a sensor (measuring element).

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, or voltage, 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.

**Self-operated controllers.** In most industrial automatic controllers, separate units are used for the measuring element and for the actuator. In a very simple one, however, such as a self-operated controller, these elements are assembled in one unit. Self-operated controllers utilize power developed by the measuring element and are very simple and inexpensive. An example of such a self-operated controller is shown in Figure 5–2. The set point is determined by the adjustment of the spring force. The controlled pressure is measured by the diaphragm. The actuating error signal is the net force acting on the diaphragm. Its position determines the valve opening.

The operation of the self-operated controller is as follows: Suppose that the output pressure is lower than the reference pressure, as determined by the set point. Then the downward spring force is greater than the upward pressure force, resulting in a downward movement of the diaphragm. This increases the flow rate and raises the output pressure. When the upward pressure force equals the downward spring force, the valve plug stays stationary and the flow rate is constant. Conversely, if the output pressure is higher than the reference pressure, the valve opening becomes small and reduces the flow rate through the valve opening. Such a self-operated controller is widely used for water and gas pressure control.

**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

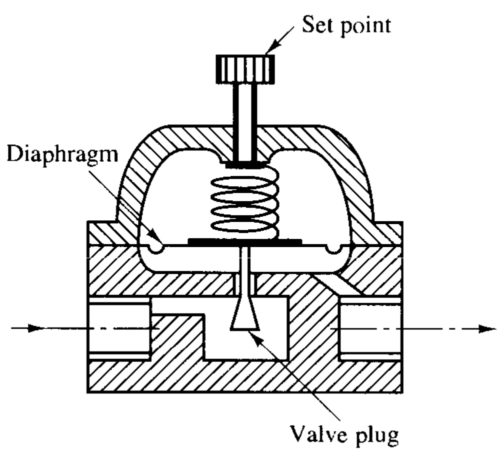



**Figure 5–2**
Self-operated controller.

$$u(t) = U_1, \qquad \text{for } e(t) > 0$$
$$\quad = U_2, \qquad \text{for } e(t) < 0$$

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 5–3 (a) and (b) show the block diagrams for two-position controllers. The range through which the actuating error signal must move before the switching occurs is called the *differential gap*. A differential gap is indicated in Figure 5–3(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 one–off mechanism.

Consider the liquid-level control system shown in Figure 5–4(a), where the electromagnetic valve shown in Figure 5–4(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 5–5, the output signal continuously moves between the two limits required to cause the actuating element to move from one fixed

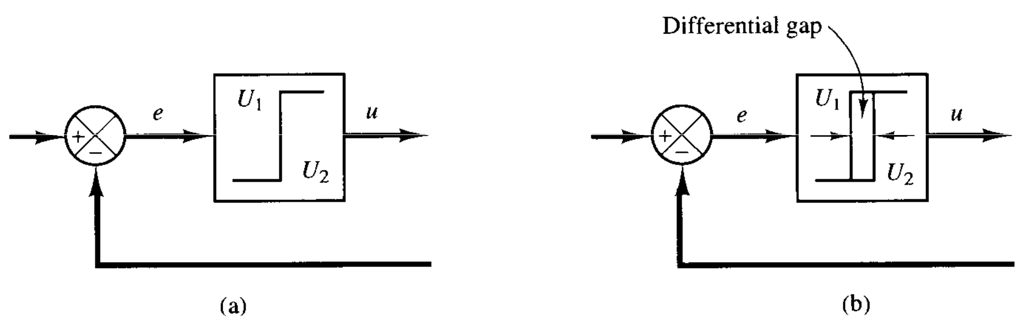



**Figure 5–3**
(a) Block diagram of an on-off controller; (b) block diagram of an on-off controller with differential gap.

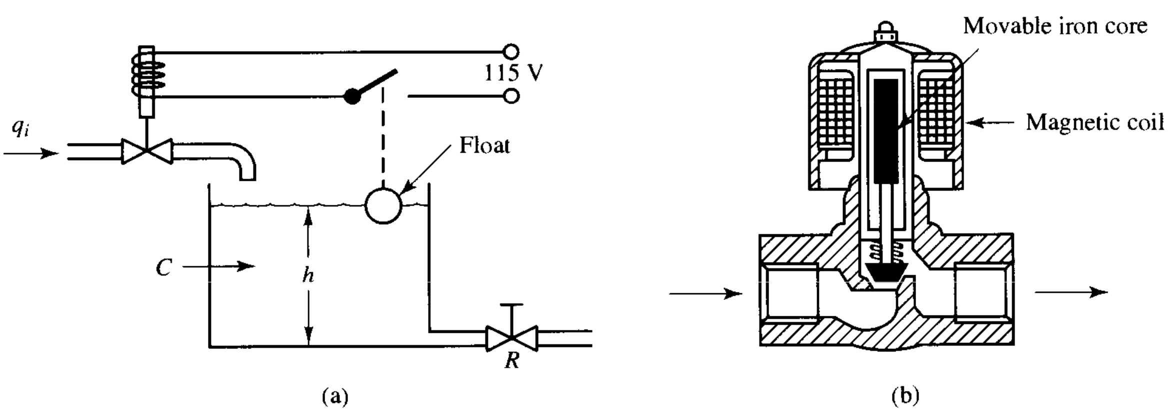



**Figure 5–4**
(a) Liquid-level control system; (b) electromagnetic valve.

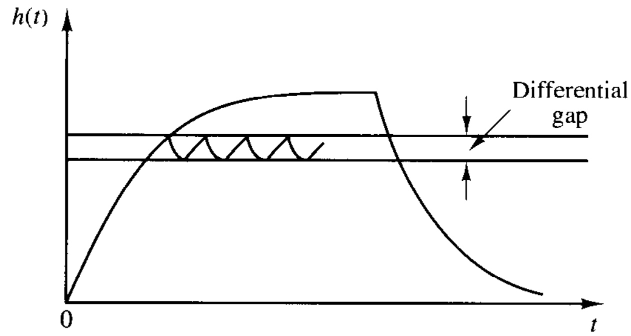



**Figure 5–5**
Level $h(t)$ versus $t$ curve for the system shown in Figure 5–4(a).

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.

From Figure 5–5, 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. A block diagram of such a controller is shown in Figure 5–6.

**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,

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

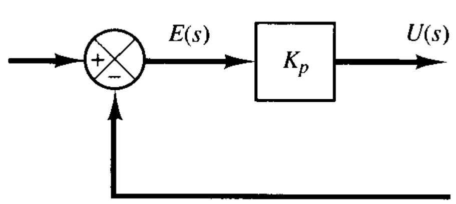



**Figure 5–6**
Block diagram of a proportional controller.

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

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

If the value of $e(t)$ is doubled, then the value of $u(t)$ varies twice as fast. For zero actuating error, the value of $u(t)$ remains stationary. The integral control action is sometimes called reset control. Figure 5–7 shows a block diagram of such a controller.

**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 $K_p$ is the proportional gain, and $T_i$ is called the *integral time.* Both $K_p$ and $T_i$ are adjustable. The integral time adjusts the integral control action, while a change in the value of $K_p$ affects both the proportional and integral parts of the control action. The inverse of the integral time $T_i$ is called the *reset rate.* The reset rate is the number of times per minute that the proportional part of the control action is duplicated. Reset rate is measured in terms of repeats per minute. Figure 5–8 (a) shows a block diagram of a proportional-plus-integral controller. If the actuating error signal $e(t)$ is a unit-step function as shown in Figure 5–8(b), then the controller output $u(t)$ becomes as shown in Figure 5–8(c).

**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

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

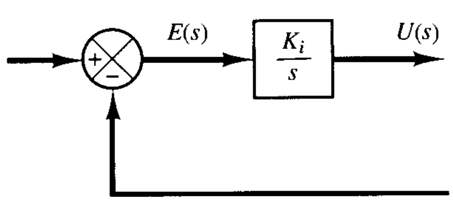



**Figure 5–7**
Block diagram of an integral controller.

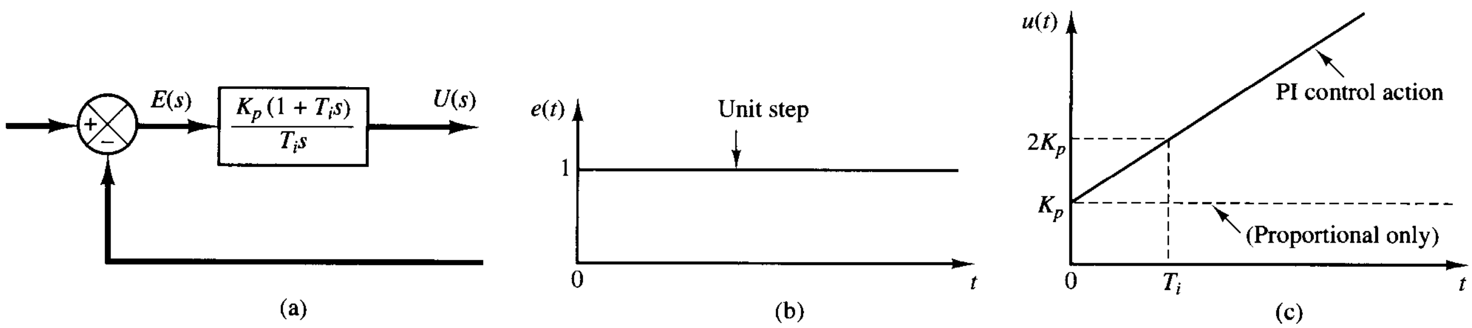



**Figure 5–8**
(a) Block diagram of a proportional-plus-integral controller; (b) and (c) diagrams depicting a unit-step input and the controller output.

where $K_p$ is the proportional gain and $T_d$ is a constant called the *derivative time*. Both $K_p$ and $T_d$ are adjustable. The derivative control action, sometimes called *rate control*, is where the magnitude of the controller output is proportional to the rate of change of the actuating error signal. The derivative time $T_d$ is the time interval by which the rate action advances the effect of the proportional control action. Figure 5–9(a) shows a block diagram of a proportional-plus-derivative controller. If the actuating error signal $e(t)$ is a unit-ramp function as shown in Figure 5–9(b), then the controller output $u(t)$ becomes as shown in Figure 5–9(c). As may be seen from Figure 5–9(c), the derivative control action has an anticipatory character. As a matter of course, however, derivative control action can never anticipate any action that has not yet taken place.

While derivative control action has the advantage of being anticipatory, it has the disadvantages that it amplifies noise signals and may cause a saturation effect in the actuator.

Note that derivative control action can never be used alone because this control action is effective only during transient periods.

**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. This combined action

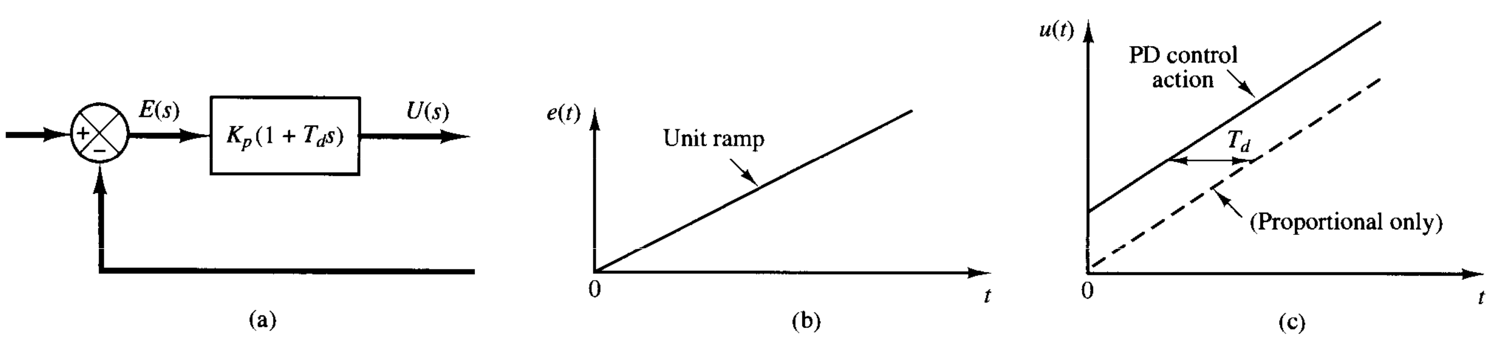



**Figure 5–9**
(a) Block diagram of a proportional-plus-derivative controller; (b) and (c) diagrams depicting a unit-ramp input and the controller output.

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 5–10(a). If $e(t)$ is a unit-ramp function as shown in Figure 5–10(b), then the controller output $u(t)$ becomes as shown in Figure 5–10(c).

**Effects of the sensor (measuring element) on system performance.** Since the dynamic and static characteristics of the sensor or measuring element affect the indication of the actual value of the output variable, the sensor plays an important role in determining the overall performance of the control system. The sensor usually determines the transfer function in the feedback path. If the time constants of a sensor are negligibly small compared with other time constants of the control system, the transfer function of the sensor simply becomes a constant. Figures 5–11(a), (b), and (c) show block diagrams of automatic controllers having a first-order sensor, an overdamped second-order sensor, and an underdamped second-order sensor, respectively. The response of a thermal sensor is often of the overdamped second-order type.

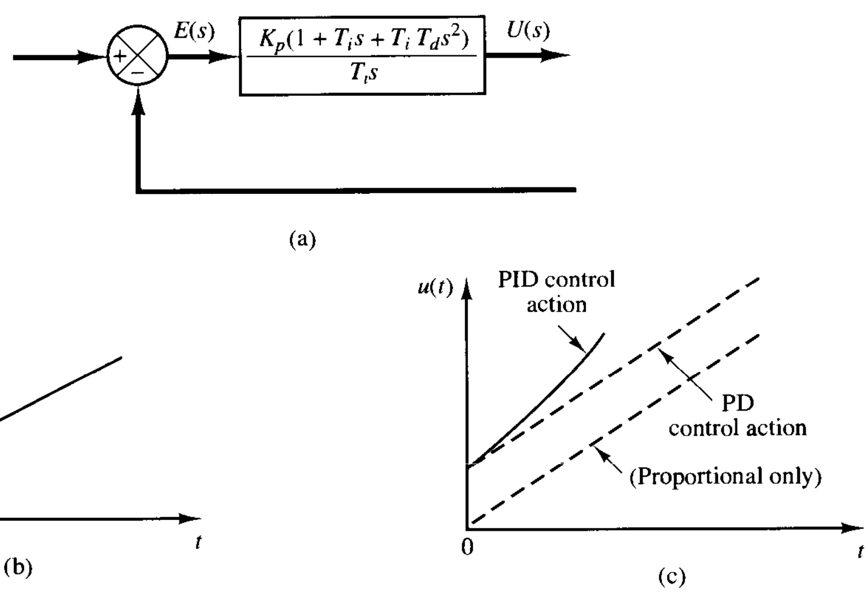



**Figure 5–10**
(a) Block diagram of a proportional-plus-integral-plus-derivative controller; (b) and (c) diagrams depicting a unit-ramp input and the controller output.

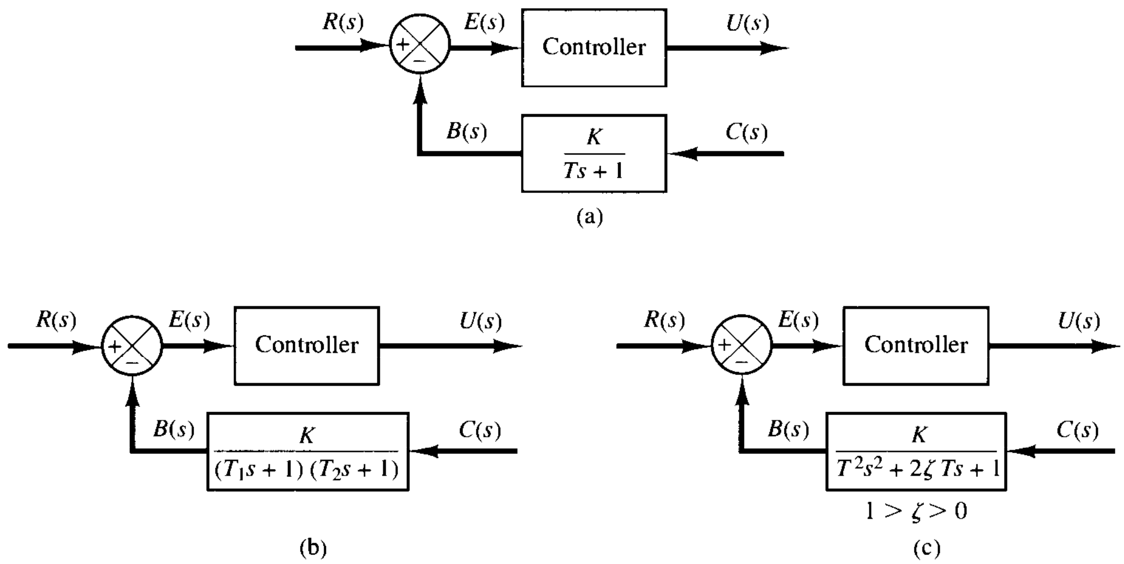



**Figure 5–11**
Block diagrams of automatic controllers with (a) first-order sensor; (b) overdamped second-order sensor; (c) underdamped second-order sensor.

## 5–3 EFFECTS OF INTEGRAL AND DERIVATIVE CONTROL ACTIONS ON SYSTEM PERFORMANCE

In this section, we shall investigate the effects of integral and derivative control actions on the system performance. Here we shall consider only simple systems so that the effects of integral and derivative control actions on system performance can be clearly seen.

**Integral control action.** In the proportional control of a plant whose transfer function does not possess an integrator $1/s$, there is a steady-state error, or offset, in the response to a step input. Such an offset can be eliminated if the integral control action is included in the controller.

In the integral control of a plant, the control signal, the output signal from the controller, at any instant is the area under the actuating error signal curve up to that instant. The control signal $u(t)$ can have a nonzero value when the actuating error signal $e(t)$ is zero, as shown in Figure 5–12(a). This is impossible in the case of the proportional controller since a nonzero control signal requires a nonzero actuating error signal. (A nonzero actuating error signal at steady state means that there is an offset.) Figure 5–12(b) shows the curve $e(t)$ versus $t$ and the corresponding curve $u(t)$ versus $t$ when the controller is of the proportional type.

Note that integral control action, while removing offset or steady-state error, may lead to oscillatory response of slowly decreasing amplitude or even increasing amplitude, both of which are usually undesirable.

**Integral control of liquid-level control systems.** In Section 4–2, we found that the proportional control of a liquid-level system will result in a steady-state error with a step input. We shall now show that such an error can be eliminated if integral control action is included in the controller.

**Figure 5–12**
(a) Plots of $e(t)$ and $u(t)$ curves showing nonzero control signal when the actuating error signal is zero (integral control); (b) plots of $e(t)$ and $u(t)$ curves showing zero control signal when the actuating error signal is zero (proportional control).

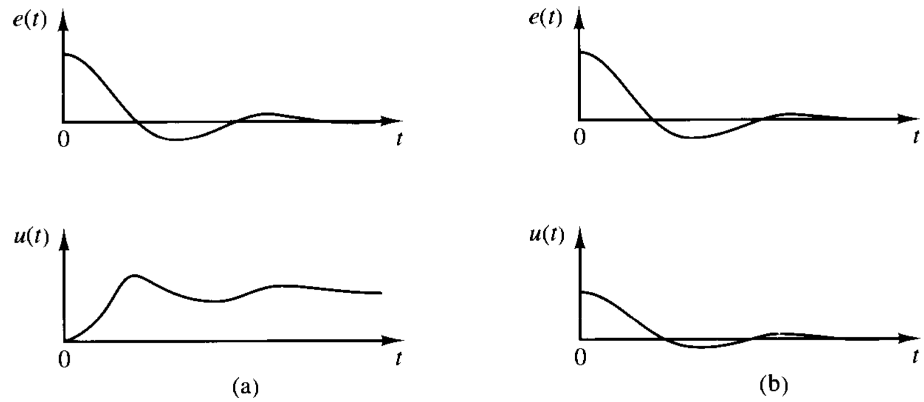



Figure 5–13(a) shows a liquid-level control system. We assume that the controller is an integral controller. We also assume that the variables $x$, $q_i$, $h$, and $q_o$, which are measured from their respective steady-state values $\bar{X}$, $\bar{Q}$, $\bar{H}$, and $\bar{Q}$, are small quantities so that the system can be considered linear. Under these assumptions, the block diagram of the system can be obtained as shown in Figure 5–13(b). From Figure 5–13(b), the closed-loop transfer function between $H(s)$ and $X(s)$ is

$$\frac{H(s)}{X(s)} = \frac{KR}{RCs^2 + s + KR}$$

Hence

$$\frac{E(s)}{X(s)} = \frac{X(s) - H(s)}{X(s)}$$

$$= \frac{RCs^2 + s}{RCs^2 + s + KR}$$

Since the system is stable, the steady-state error for the unit-step response is obtained by applying the final-value theorem, as follows:

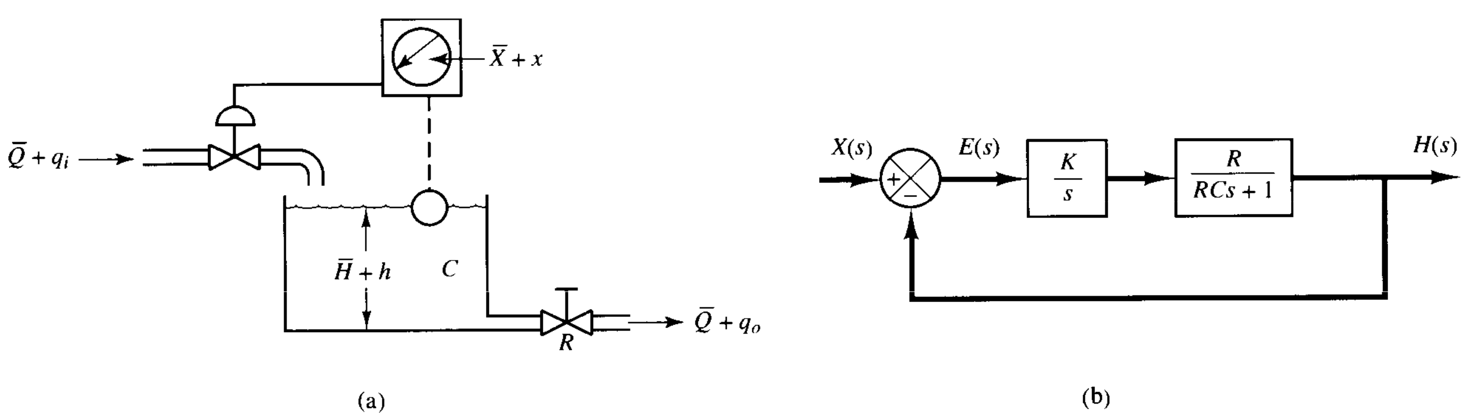



**Figure 5–13**
(a) Liquid-level control system; (b) block diagram of the system.

$$
\begin{aligned}
e_{ss} &= \lim_{s \to 0} sE(s) \\
&= \lim_{s \to 0} \frac{s(RCs^2 + s)}{RCs^2 + s + KR} \frac{1}{s} \\
&= 0
\end{aligned}
$$

Integral control of the liquid-level system thus eliminates the steady-state error in the response to the step input. This is an important improvement over the proportional control alone, which gives offset.

**Response to torque disturbances (proportional control).** Let us investigate the effect of a torque disturbance occurring at the load element. Consider the system shown in Figure 5–14. The proportional controller delivers torque $T$ to position the load element, which consists of moment of inertia and viscous friction. Torque disturbance is denoted by $D$.

Assuming that the reference input is zero or $R(s) = 0$, the transfer function between $C(s)$ and $D(s)$ is given by

$$
\frac{C(s)}{D(s)} = \frac{1}{Js^2 + bs + K_p}
$$

Hence

$$
\frac{E(s)}{D(s)} = -\frac{C(s)}{D(s)} = -\frac{1}{Js^2 + bs + K_p}
$$

The steady-state error due to a step disturbance torque of magnitude $T_d$ is given by

$$
\begin{aligned}
e_{ss} &= \lim_{s \to 0} sE(s) \\
&= \lim_{s \to 0} \frac{-s}{Js^2 + bs + K_p} \frac{T_d}{s} \\
&= -\frac{T_d}{K_p}
\end{aligned}
$$

At steady state, the proportional controller provides the torque $-T_d$, which is equal in magnitude but opposite in sign to the disturbance torque $T_d$. The steady-state output due to the step disturbance torque is

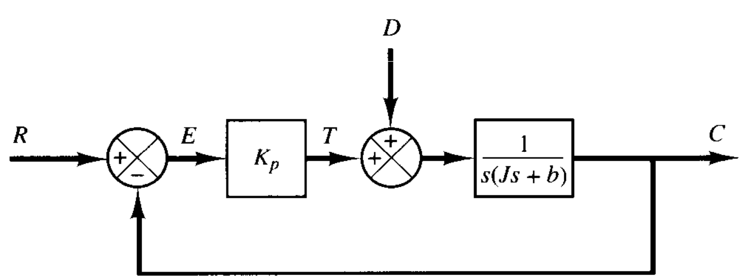



**Figure 5–14**
Control system with a torque disturbance.

$$c_{ss} = -e_{ss} = \frac{T_d}{K_p}$$

The steady-state error can be reduced by increasing the value of the gain $K_p$. Increasing this value, however, will cause the system response to be more oscillatory.

**Obtaining responses with MATLAB.** In what follows, we shall obtain the response curves of the system shown in Figure 5–14 when it is subjected to a unit-step disturbance. Specifically, we shall obtain step-response curves for a small value of $K_p$ and a large value of $K_p$.

Consider two cases:

Case 1: $J = 1, b = 0.5, K_p = 1$ (system 1):

$$\frac{C(s)}{D(s)} = \frac{1}{s^2 + 0.5s + 1}$$

Case 2: $J = 1, b = 0.5, K_p = 4$ (system 2):

$$\frac{C(s)}{D(s)} = \frac{1}{s^2 + 0.5s + 4}$$

Note that for system 1

```
num1 = [0   0   1]
den1 = [1   0.5   1]
```

For system 2

```
num2 = [0   0   1]
den2 = [1   0.5   4]
```

In MATLAB Program 5–1 we have used notations y1 and y2 for the response. y1 is the response $c(t)$ of system 1, and y2 is the response $c(t)$ of system 2.

In MATLAB Program 5–1, note that we have used the plot command with multiple arguments, rather than using the *hold* command. (We get the same result either way.) To use the *plot* command with multiple arguments, the sizes of the y1 and y2 vectors need not be the same. However, it is convenient if the two vectors are of the same length. Hence, we specify the same number of computing points by specifying the computing time points (such as t = 0:0.1:20). The *step* command must include this user-specified time t. Thus, in MATLAB Program 5–1 we have used the following *step* command:

```
[y, x, t] = step(num, den,t)
```

The unit-step response curves obtained by use of MATLAB Program 5–1 are shown in Figure 5–15.

MATLAB Program 5–1

```
% ---------- Plotting two step-response curves on one
% diagram ----------

% *****Enter numerators and denominators of two
% transfer functions*****

num1 = [0   0   1];
den1 = [1   0.5   1];
num2 = [0   0   1];
den2 = [1   0.5   4];

% ***** To plot two step-response curves y1 versus t
% and y2 versus t on one diagram and write texts
% 'System 1' and 'System 2' to distinguish two curves,
% enter the following commands *****

t = 0:0.1:20;
[y1,x1,t] = step(num1,den1,t);
[y2,x2,t] = step(num2,den2,t);
plot(t,y1,t,y2)
grid
text(11,0.75,'System 1'), text(11.2,0.16,'System 2')

% ***** Add title of the plot, xlabel, and ylabel *****

title('Step Responses of Two Systems')
xlabel('t Sec')
ylabel('Outputs y1 and y2')
```

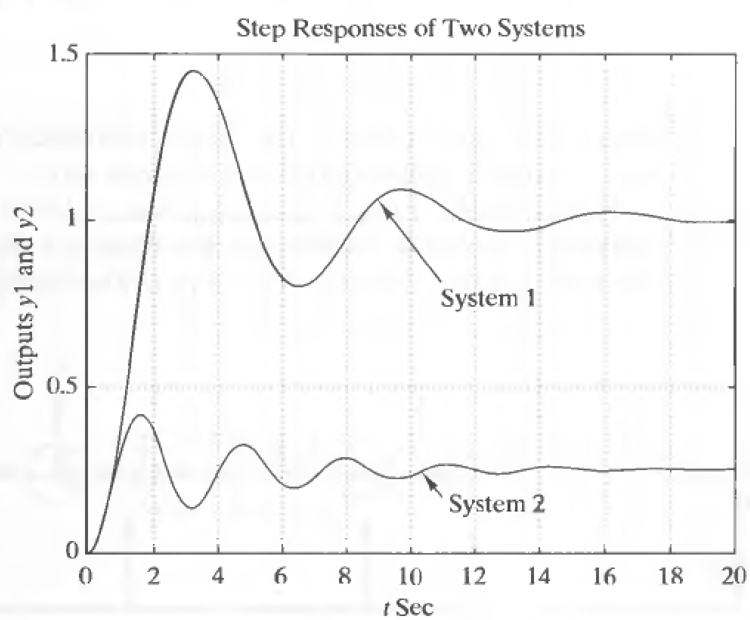



**Figure 5–15** Unit-step response curves.

**Response to torque disturbances (proportional-plus-integral control).** To eliminate offset due to torque disturbance, the proportional controller may be replaced by a proportional-plus-integral controller.

If integral control action is added to the controller, then, as long as there is an error signal, a torque is developed by the controller to reduce this error, provided the control system is a stable one.

Figure 5–16 shows the proportional-plus-integral control of the load element, consisting of moment of inertia and viscous friction.

The closed-loop transfer function between $C(s)$ and $D(s)$ is

$$\frac{C(s)}{D(s)} = \frac{s}{Js^3 + bs^2 + K_p s + \dfrac{K_p}{T_i}}$$

In the absence of the reference input, or $r(t) = 0$, the error signal is obtained from

$$E(s) = -\frac{s}{Js^3 + bs^2 + K_p s + \dfrac{K_p}{T_i}} D(s)$$

If this control system is stable, that is, if the roots of the characteristic equation

$$Js^3 + bs^2 + K_p s + \frac{K_p}{T_i} = 0$$

have negative real parts, then the steady-state error in the response to a unit-step disturbance torque can be obtained by applying the final-value theorem as follows:

$$\begin{aligned} e_{ss} &= \lim_{s\to 0} sE(s) \\ &= \lim_{s\to 0} \frac{-s^2}{Js^3 + bs^2 + K_p s + \dfrac{K_p}{T_i}} \frac{1}{s} \\ &= 0 \end{aligned}$$

Thus steady-state error to the step disturbance torque can be eliminated if the controller is of the proportional-plus-integral type.

Note that the integral control action added to the proportional controller has converted the originally second-order system to a third-order one. Hence the control system may become unstable for a large value of $K_p$ since the roots of the characteristic

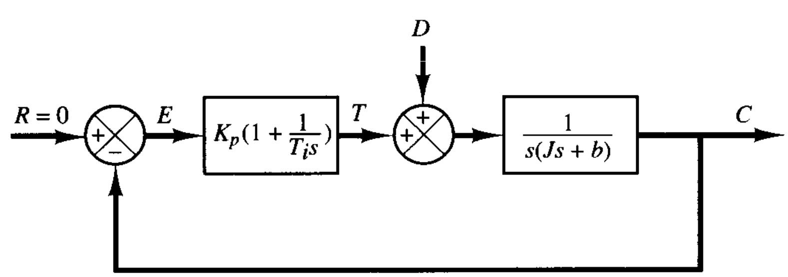



**Figure 5–16** Proportional-plus-integral control of a load element consisting of moment of inertia and viscous friction.

equation may have positive real parts. (The second-order system is always stable if the coefficients in the system differential equation are all positive.)

It is important to point out that if the controller were an integral controller, as in Figure 5–17, then the system always becomes unstable because the characteristic equation

$$Js^3 + bs^2 + K = 0$$

will have roots with positive real parts. Such an unstable system cannot be used in practice.

Note that in the system of Figure 5–16 the proportional control action tends to stabilize the system, while the integral control action tends to eliminate or reduce steady-state error in response to various inputs.

**Derivative control action.** Derivative control action, when added to a proportional controller, provides a means of obtaining a controller with high sensitivity. An advantage of using derivative control action is that it responds to the rate of change of the actuating error and can produce a significant correction before the magnitude of the actuating error becomes too large. Derivative control thus anticipates the actuating error, initiates an early corrective action, and tends to increase the stability of the system.

Although derivative control does not affect the steady-state error directly, it adds damping to the system and thus permits the use of a larger value of the gain $K$, which will result in an improvement in the steady-state accuracy.

Because derivative control operates on the rate of change of the actuating error and not the actuating error itself, this mode is never used alone. It is always used in combination with proportional or proportional-plus-integral control action.

**Proportional control of systems with inertia load.** Before we discuss the effect of derivative control action on system performance, we shall consider the proportional control of an inertia load.

Consider the system shown in Figure 5–18(a). The closed-loop transfer function is obtained as

$$\frac{C(s)}{R(s)} = \frac{K_p}{Js^2 + K_p}$$

Since the roots of the characteristic equation

$$Js^2 + K_p = 0$$

are imaginary, the response to a unit-step input continues to oscillate indefinitely, as shown in Figure 5–18(b).

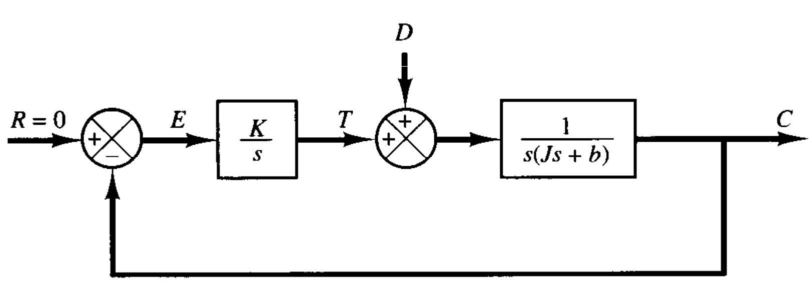



**Figure 5–17** Integral control of a load element consisting of moment of inertia and viscous friction.

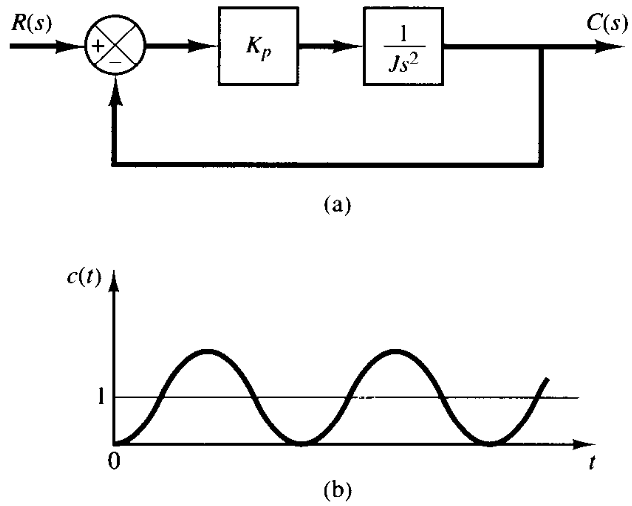



**Figure 5–18**
(a) Proportional control of a system with inertia load; (b) response to a unit-step input.

Control systems exhibiting such response characteristics are not desirable. We shall see that the addition of derivative control will stabilize the system.

**Proportional-plus-derivative control of a system with inertia load.** Let us modify the proportional controller to a proportional-plus-derivative controller whose transfer function is $K_p(1 + T_d s)$. The torque developed by the controller is proportional to $K_p(e + T_d \dot{e})$. Derivative control is essentially anticipatory, measures the instantaneous error velocity, and predicts the large overshoot ahead of time and produces an appropriate counteraction before too large an overshoot occurs.

Consider the system shown in Figure 5–19(a). The closed-loop transfer function is given by

$$\frac{C(s)}{R(s)} = \frac{K_p(1 + T_d s)}{Js^2 + K_p T_d s + K_p}$$

The characteristic equation

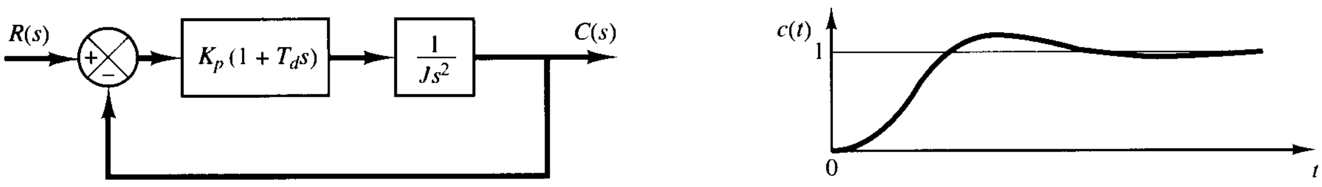
$$Js^2 + K_p T_d s + K_p = 0$$

now has two roots with negative real parts for positive values of $J$, $K_p$, and $T_d$. Thus derivative control introduces a damping effect. A typical response curve $c(t)$ to a unit-step input is shown in Figure 5–19(b). Clearly, the response curve shows a marked improvement over the original response curve shown in Figure 5–18(b).

**Proportional-plus-derivative control of second-order systems.** A compromise between acceptable transient-response behavior and acceptable steady-state behavior may be achieved by use of proportional-plus-derivative control action.

Consider the system shown in Figure 5–20. The closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{K_p + K_d s}{Js^2 + (B + K_d)s + K_p}$$




(a) (b)

**Figure 5–19**
(a) Proportional-plus-derivative control of a system with inertia load; (b) response to a unit-step input.

The steady-state error for a unit-ramp input is

$$e_{ss} = \frac{B}{K_p}$$

The characteristic equation is

$$Js^2 + (B + K_d)s + K_p = 0$$

The effective damping coefficient of this system is thus $B + K_d$ rather than $B$. Since the damping ratio $\zeta$ of this system is

$$\zeta = \frac{B + K_d}{2\sqrt{K_p J}}$$

it is possible to make both the steady-state error $e_{ss}$ for a ramp input and the maximum overshoot for a step input small by making $B$ small, $K_p$ large, and $K_d$ large enough so that $\zeta$ is between 0.4 and 0.7.

In the following, we shall examine the unit-step response of the system shown in Figure 5–20. Let us define

$$\omega_n = \sqrt{\frac{K_p}{J}}, \qquad z = \frac{K_p}{K_d}$$

The closed-loop transfer function can then be written

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{z} \frac{s + z}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

When a second-order system has a zero near the closed-loop poles, the transient-response behavior becomes considerably different from that of a second-order system without a zero.

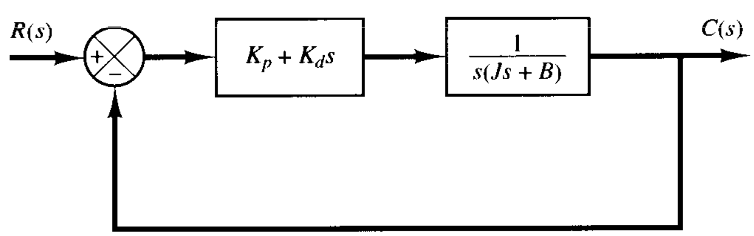



**Figure 5–20**
Control system.

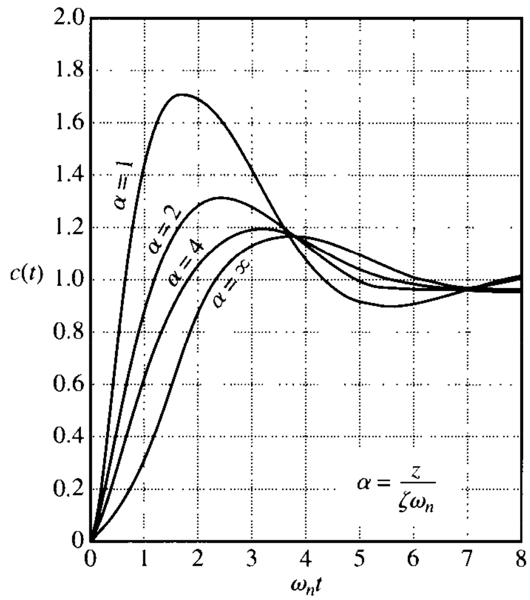



**Figure 5–21**
Unit-step response curves of the second-order system

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{z}\,\frac{s+z}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

$$\zeta = 0.5$$

If the zero at $s = -z$ is located close to the $j\omega$ axis, the effect of the zero on the unit-step response is quite significant. Typical step-response curves of this system with $\zeta = 0.5$ and various values of $z/(\zeta\omega_n)$ are shown in Figure 5–21.

## 5–4 HIGHER-ORDER SYSTEMS

In this section, we shall first discuss the unit-step response of a particular type of higher-order system. We shall then present a transient response analysis of higher-order systems in general terms. Finally, we shall present a discussion of stability analysis in the complex plane.

**Transient response of higher-order systems.** Consider the system shown in Figure 5–22. The closed loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)H(s)} \tag{5–1}$$

In general, $G(s)$ and $H(s)$ are given as ratios of polynomials in $s$, or

$$G(s) = \frac{p(s)}{q(s)} \qquad \text{and} \qquad H(s) = \frac{n(s)}{d(s)}$$

where $p(s)$, $q(s)$, $n(s)$, and $d(s)$ are polynomials in $s$. The closed-loop transfer function given by Equation (5–1) may then be written

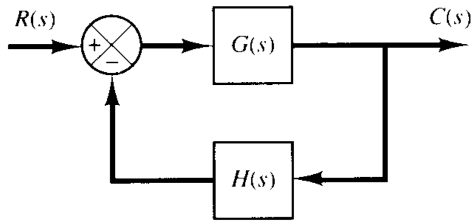



**Figure 5–22**
Control system.

$$\frac{C(s)}{R(s)} = \frac{p(s)d(s)}{q(s)d(s) + p(s)n(s)}$$

$$= \frac{b_0 s^m + b_1 s^{m-1} + \cdots + b_{m-1}s + b_m}{a_0 s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n} \qquad (m \leq n)$$

The transient response of this system to any given input can be obtained by a computer simulation (see Section 4–4). If an analytical expression for the transient response is desired, then it is necessary to factor the denominator polynomial. [MATLAB may be used for finding the roots of the denominator polynomial. Use the command roots(den).] Once the numerator and the denominator have been factored, $C(s)/R(s)$ can be written as

$$\frac{C(s)}{R(s)} = \frac{K(s + z_1)(s + z_2) \cdots (s + z_m)}{(s + p_1)(s + p_2) \cdots (s + p_n)} \tag{5–2}$$

Let us examine the response behavior of this system to a unit-step input. Consider first the case where the closed-loop poles are all real and distinct. For a unit-step input, Equation (5–2) can be written

$$C(s) = \frac{a}{s} + \sum_{i=1}^{n} \frac{a_i}{s + p_i} \tag{5–3}$$

where $a_i$ is the residue of the pole at $s = -p_i$.

If all closed-loop poles lie in the left-half $s$ plane, the relative magnitudes of the residues determine the relative importance of the components in the expanded form of $C(s)$. If there is a closed-loop zero close to a closed-loop pole, then the residue at this pole is small and the coefficient of the transient-response term corresponding to this pole becomes small. A pair of closely located poles and zeros will effectively cancel each other. If a pole is located very far from the origin, the residue at this pole may be small. The transients corresponding to such a remote pole are small and last a short time. Terms in the expanded form of $C(s)$ having very small residues contribute little to the transient response, and these terms may be neglected. If this is done, the higher-order system may be approximated by a lower-order one. (Such an approximation often enables us to estimate the response characteristics of a higher-order system from those of a simplified one.)

Next, consider the case where the poles of $C(s)$ consist of real poles and pairs of complex-conjugate poles. A pair of complex-conjugate poles yields a second-order term in $s$. Since the factored form of the higher-order characteristic equation consists of first- and second-order terms, Equation (5–3) can be rewritten

$$C(s) = \frac{K \prod_{i=1}^{m} (s + z_i)}{s \prod_{j=1}^{q} (s + p_j) \prod_{k=1}^{r} (s^2 + 2\zeta_k \omega_k s + \omega_k^2)} \tag{5–4}$$

where $q + 2r = n$. If the closed-loop poles are distinct, Equation (5–4) can be expanded into partial fractions as follows:

$$C(s) = \frac{a}{s} + \sum_{j=1}^{q} \frac{a_j}{s + p_j} + \sum_{k=1}^{r} \frac{b_k(s + \zeta_k \omega_k) + c_k \omega_k \sqrt{1 - \zeta_k^2}}{s^2 + 2\zeta_k \omega_k s + \omega_k^2}$$

From this last equation, we see that the response of a higher-order system is composed of a number of terms involving the simple functions found in the responses of first- and second-order systems. The unit-step response $c(t)$, the inverse Laplace transform of $C(s)$, is then

$$\begin{aligned} c(t) = a &+ \sum_{j=1}^{q} a_j e^{-p_j t} + \sum_{k=1}^{r} b_k e^{-\zeta_k \omega_k t} \cos \omega_k \sqrt{1 - \zeta_k^2}\, t \\ &+ \sum_{k=1}^{r} c_k e^{-\zeta_k \omega_k t} \sin \omega_k \sqrt{1 - \zeta_k^2}\, t, \qquad \text{for } t \geq 0 \end{aligned} \tag{5–5}$$

Thus the response curve of a stable higher-order system is the sum of a number of exponential curves and damped sinusoidal curves.

If all closed-loop poles lie in the left-half $s$ plane, then the exponential terms and the damped exponential terms in Equation (5–5) will approach zero as time $t$ increases. The steady-state output is then $c(\infty) = a$.

Let us assume that the system considered is a stable one. Then the closed-loop poles that are located far from the $j\omega$ axis have large negative real parts. The exponential terms that correspond to these poles decay very rapidly to zero. (Note that the horizontal distance from a closed-loop pole to the $j\omega$ axis determines the settling time of transients due to that pole. The smaller the distance is, the longer the settling time.)

Remember that the type of transient response is determined by the closed-loop poles, while the shape of the transient response is primarily determined by the closed-loop zeros. As we have seen earlier, the poles of the input $R(s)$ yield the steady-state response terms in the solution, while the poles of $C(s)/R(s)$ enter into the exponential transient-response terms and/or damped sinusoidal transient-response terms. The zeros of $C(s)/R(s)$ do not affect the exponents in the exponential terms, but they do affect the magnitudes and signs of the residues.

**Dominant closed-loop poles.** The relative dominance of closed-loop poles is determined by the ratio of the real parts of the closed-loop poles, as well as by the relative magnitudes of the residues evaluated at the closed-loop poles. The magnitudes of the residues depend on both the closed-loop poles and zeros.

If the ratios of the real parts exceed 5 and there are no zeros nearby, then the closed-loop poles nearest the $j\omega$ axis will dominate in the transient-response behavior because these poles correspond to transient-response terms that decay slowly. Those closed-loop poles that have dominant effects on the transient-response behavior are called *dominant closed-loop* poles. Quite often the dominant closed-loop poles occur in the form of a complex-conjugate pair. The dominant closed-loop poles are most important among all closed-loop poles.

The gain of a higher-order system is often adjusted so that there will exist a pair of dominant complex-conjugate closed-loop poles. The presence of such poles in a stable system reduces the effect of such nonlinearities as dead zone, backlash, and coulomb friction.

Remember that, although the concept of dominant closed-loop poles is useful for estimating the dynamic behavior of a closed-loop system, we must be careful to see that the underlying assumptions are met before using it.

**Stability analysis in the complex plane.** The stability of a linear closed-loop system can be determined from the location of the closed-loop poles in the $s$ plane. If any of these poles lie in the right-half $s$ plane, then with increasing time they give rise to the dominant mode, and the transient response increases monotonically or oscillates with increasing amplitude. This represents an unstable system. For such a system, as soon as the power is turned on, the output may increase with time. If no saturation takes place in the system and no mechanical stop is provided, then the system may eventually be subjected to damage and fail since the response of a real physical system cannot increase indefinitely. Therefore, closed-loop poles in the right-half $s$ plane are not permissible in the usual linear control system. If all closed-loop poles lie to the left of the $j\omega$ axis, any transient response eventually reaches equilibrium. This represents a stable system.

Whether a linear system is stable or unstable is a property of the system itself and does not depend on the input or driving function of the system. The poles of the input, or driving function, do not affect the property of stability of the system, but they contribute only to steady-state response terms in the solution. Thus, the problem of absolute stability can be solved readily by choosing no closed-loop poles in the right-half $s$ plane, including the $j\omega$ axis. (Mathematically, closed-loop poles on the $j\omega$ axis will yield oscillations, the amplitude of which is neither decaying nor growing with time. In practical cases, where noise is present, however, the amplitude of oscillations may increase at a rate determined by the noise power level. Therefore, a control system should not have closed-loop poles on the $j\omega$ axis.)

Note that the mere fact that all closed-loop poles lie in the left-half $s$ plane does not guarantee satisfactory transient-response characteristics. If dominant complex-conjugate closed-loop poles lie close to the $j\omega$ axis, the transient response may exhibit excessive oscillations or may be very slow. Therefore, to guarantee fast, yet well-damped, transient-response characteristics, it is necessary that the closed-loop poles of the system lie in a particular region in the complex plane, such as the region bounded by the shaded area in Figure 5–23.

Since the relative stability and transient performance of a closed-loop control system are directly related to the closed-loop pole–zero configuration in the $s$ plane, it is frequently necessary to adjust one or more system parameters in order to obtain suitable configurations. The effects of varying system parameters on the closed-loop poles will be discussed in detail in Chapter 6.

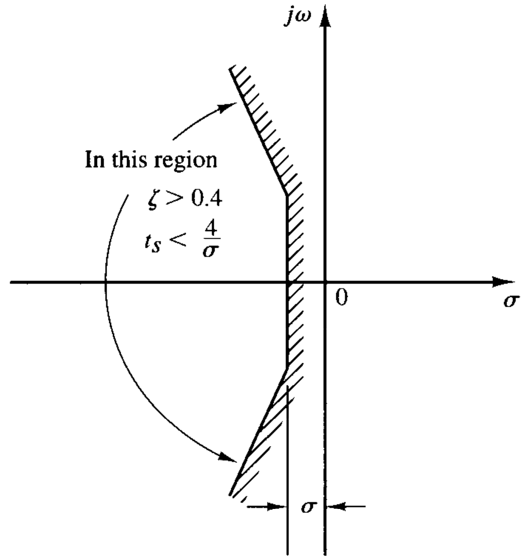



**Figure 5–23**
Region in the complex plane satisfying the conditions $\zeta > 0.4$ and $t_s < 4/\sigma$.

## 5–5 ROUTH'S STABILITY CRITERION

The most important problem in linear control systems concerns stability. That is, under what conditions will a system become unstable? If it is unstable, how should we stabilize the system? In Section 5–4 it was stated that a control system is stable if and only if all closed-loop poles lie in the left-half $s$ plane. Since most linear closed-loop systems have closed-loop transfer functions of the form

$$\frac{C(s)}{R(s)} = \frac{b_0 s^m + b_1 s^{m-1} + \cdots + b_{m-1}s + b_m}{a_0 s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n} = \frac{B(s)}{A(s)}$$

where the $a$'s and $b$'s are constants and $m \le n$, we must first factor the polynomial $A(s)$ in order to find the closed-loop poles. A simple criterion, known as Routh's stability criterion, enables us to determine the number of closed-loop poles that lie in the right-half $s$ plane without having to factor the polynomial.

**Routh's stability criterion.** Routh's stability criterion tells us whether or not there are unstable roots in a polynomial equation without actually solving for them. This stability criterion applies to polynomials with only a finite number of terms. When the criterion is applied to a control system, information about absolute stability can be obtained directly from the coefficients of the characteristic equation.

The procedure in Routh's stability criterion is as follows:

**1.** Write the polynomial in $s$ in the following form:

$$a_0 s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n = 0 \tag{5–6}$$

where the coefficients are real quantities. We assume that $a_n \neq 0$; that is, any zero root has been removed.

**2.** If any of the coefficients are zero or negative in the presence of at least one positive coefficient, there is a root or roots that are imaginary or that have positive real parts.

Therefore, in such a case, the system is not stable. If we are interested in only the absolute stability, there is no need to follow the procedure further. Note that all the coefficients must be positive. This is a necessary condition, as may be seen from the following argument: A polynomial in $s$ having real coefficients can always be factored into linear and quadratic factors, such as $(s + a)$ and $(s^2 + bs + c)$, where $a$, $b$, and $c$ are real. The linear factors yield real roots and the quadratic factors yield complex roots of the polynomial. The factor $(s^2 + bs + c)$ yields roots having negative real parts only if $b$ and $c$ are both positive. For all roots to have negative real parts, the constants $a$, $b$, $c$, and so on, in all factors must be positive. The product of any number of linear and quadratic factors containing only positive coefficients always yields a polynomial with positive coefficients. It is important to note that the condition that all the coefficients be positive is not sufficient to assure stability. The necessary but not sufficient condition for stability is that the coefficients of Equation (5–6) all be present and all have a positive sign. (If all $a$'s are negative, they can be made positive by multiplying both sides of the equation by $-1$.)

**3.** If all coefficients are positive, arrange the coefficients of the polynomial in rows and columns according to the following pattern:

$$
\begin{array}{lllllll}
s^n & a_0 & a_2 & a_4 & a_6 & \cdot & \cdot \quad \cdot \\
s^{n-1} & a_1 & a_3 & a_5 & a_7 & \cdot & \cdot \quad \cdot \\
s^{n-2} & b_1 & b_2 & b_3 & b_4 & \cdot & \cdot \quad \cdot \\
s^{n-3} & c_1 & c_2 & c_3 & c_4 & \cdot & \cdot \quad \cdot \\
s^{n-4} & d_1 & d_2 & d_3 & d_4 & \cdot & \cdot \quad \cdot \\
\cdot & \cdot & \cdot & & & & \\
\cdot & \cdot & \cdot & & & & \\
\cdot & \cdot & \cdot & & & & \\
s^2 & e_1 & e_2 & & & & \\
s^1 & f_1 & & & & & \\
s^0 & g_1 & & & & &
\end{array}
$$

The coefficients $b_1$, $b_2$, $b_3$, and so on, are evaluated as follows:

$$
b_1 = \frac{a_1 a_2 - a_0 a_3}{a_1}
$$

$$
b_2 = \frac{a_1 a_4 - a_0 a_5}{a_1}
$$

$$
b_3 = \frac{a_1 a_6 - a_0 a_7}{a_1}
$$

$$
\vdots
$$

The evaluation of the $b$'s is continued until the remaining ones are all zero. The same pattern of cross-multiplying the coefficients of the two previous rows is followed in evaluating the $c$'s, $d$'s, $e$'s, and so on. That is,

$$c_1 = \frac{b_1 a_3 - a_1 b_2}{b_1}$$

$$c_2 = \frac{b_1 a_5 - a_1 b_3}{b_1}$$

$$c_3 = \frac{b_1 a_7 - a_1 b_4}{b_1}$$

$$\vdots$$

and

$$d_1 = \frac{c_1 b_2 - b_1 c_2}{c_1}$$

$$d_2 = \frac{c_1 b_3 - b_1 c_3}{c_1}$$

$$\vdots$$

This process is continued until the $n$th row has been completed. The complete array of coefficients is triangular. Note that in developing the array an entire row may be divided or multiplied by a positive number in order to simplify the subsequent numerical calculation without altering the stability conclusion.

Routh's stability criterion states that the number of roots of Equation (5–6) with positive real parts is equal to the number of changes in sign of the coefficients of the first column of the array. It should be noted that the exact values of the terms in the first column need not be known; instead, only the signs are needed. The necessary and sufficient condition that all roots of Equation (5–6) lie in the left-half $s$ plane is that all the coefficients of Equation (5–6) be positive and all terms in the first column of the array have positive signs.

**EXAMPLE 5–1** Let us apply Routh's stability criterion to the following third-order polynomial:

$$a_0 s^3 + a_1 s^2 + a_2 s + a_3 = 0$$

where all the coefficients are positive numbers. The array of coefficients becomes

$$\begin{array}{lcc} s^3 & a_0 & a_2 \\ s^2 & a_1 & a_3 \\ s^1 & \dfrac{a_1 a_2 - a_0 a_3}{a_1} & \\ s^0 & a_3 & \end{array}$$

The condition that all roots have negative real parts is given by

$$a_1 a_2 > a_0 a_3$$

**EXAMPLE 5–2** Consider the following polynomial:

$$s^4 + 2s^3 + 3s^2 + 4s + 5 = 0$$

Let us follow the procedure just presented and construct the array of coefficients. (The first two rows can be obtained directly from the given polynomial. The remaining terms are obtained from these. If any coefficients are missing, they may be replaced by zeros in the array.)

$$
\begin{array}{lrrr}
s^4 & 1 & 3 & 5 \\
s^3 & 2 & 4 & 0 \\
 & & & \\
s^2 & 1 & 5 & \\
s^1 & -6 & & \\
s_0 & 5 & &
\end{array}
\quad\Bigg\Vert\quad
\begin{array}{lrrr}
s^4 & 1 & 3 & 5 \\
s^3 & \cancel{2} & \cancel{4} & \cancel{0} \\
 & 1 & 2 & 0 \\
s^2 & 1 & 5 & \\
s^1 & -3 & & \\
s^0 & 5 & &
\end{array}
\qquad
\begin{array}{l}
\\
\text{The second row is divided} \\
\text{by 2.}
\end{array}
$$

In this example, the number of changes in sign of the coefficients in the first column is 2. This means that there are two roots with positive real parts. Note that the result is unchanged when the coefficients of any row are multiplied or divided by a positive number in order to simplify the computation.

**Special cases.** If a first-column term in any row is zero, but the remaining terms are not zero or there is no remaining term, then the zero term is replaced by a very small positive number $\epsilon$ and the rest of the array is evaluated. For example, consider the following equation:

$$s^3 + 2s^2 + s + 2 = 0 \tag{5–7}$$

The array of coefficients is

$$
\begin{array}{lcc}
s^3 & 1 & 1 \\
s^2 & 2 & 2 \\
s^1 & 0 \approx \epsilon & \\
s^0 & 2 &
\end{array}
$$

If the sign of the coefficient above the zero ($\epsilon$) is the same as that below it, it indicates that there are a pair of imaginary roots. Actually, Equation (5–7) has two roots at $s = \pm j$.

If, however, the sign of the coefficient above the zero ($\epsilon$) is opposite that below it, it indicates that there is one sign change. For example, for the equation

$$s^3 - 3s + 2 = (s - 1)^2(s + 2) = 0$$

the array of coefficients is

$$
\begin{array}{rlcc}
 & s^3 & 1 & -3 \\
\text{One sign change:} & s^2 & 0 \approx \epsilon & 2 \\
\text{One sign change:} & s^1 & -3 - \dfrac{2}{\epsilon} & \\
 & s^0 & 2 &
\end{array}
$$

There are two sign changes of the coefficients in the first column. This agrees with the correct result indicated by the factored form of the polynomial equation.

If all the coefficients in any derived row are zero, it indicates that there are roots of equal magnitude lying radially opposite in the $s$ plane, that is, two real roots with equal magnitudes and opposite signs and/or two conjugate imaginary roots. In such a case, the evaluation of the rest of the array can be continued by forming an auxiliary polynomial with the coefficients of the last row and by using the coefficients of the derivative of this polynomial in the next row. Such roots with equal magnitudes and lying radially opposite in the $s$ plane can be found by solving the auxiliary polynomial, which is always even. For a $2n$-degree auxiliary polynomial, there are $n$ pairs of equal and opposite roots. For example, consider the following equation:

$$s^5 + 2s^4 + 24s^3 + 48s^2 - 25s - 50 = 0$$

The array of coefficients is

$$\begin{array}{llll} s^5 & 1 & 24 & -25 \\ s^4 & 2 & 48 & -50 \quad \longleftarrow \text{Auxiliary polynomial } P(s) \\ s^3 & 0 & 0 & \end{array}$$

The terms in the $s^3$ row are all zero. The auxiliary polynomial is then formed from the coefficients of the $s^4$ row. The auxiliary polynomial $P(s)$ is

$$P(s) = 2s^4 + 48s^2 - 50$$

which indicates that there are two pairs of roots of equal magnitude and opposite sign. These pairs are obtained by solving the auxiliary polynomial equation $P(s) = 0$. The derivative of $P(s)$ with respect to $s$ is

$$\frac{dP(s)}{ds} = 8s^3 + 96s$$

The terms in the $s^3$ row are replaced by the coefficients of the last equation, that is, 8 and 96. The array of coefficients then becomes

$$\begin{array}{llll} s^5 & 1 & 24 & -25 \\ s^4 & 2 & 48 & -50 \\ s^3 & 8 & 96 & \quad \leftarrow \text{Coefficients of } dP(s)/ds \\ s^2 & 24 & -50 & \\ s^1 & 112.7 & 0 & \\ s^0 & -50 & & \end{array}$$

We see that there is one change in sign in the first column of the new array. Thus, the original equation has one root with a positive real part. By solving for roots of the auxiliary polynomial equation,

$$2s^4 + 48s^2 - 50 = 0$$

we obtain

$$s^2 = 1, \qquad s^2 = -25$$

or

$$s = \pm 1, \qquad s = \pm j5$$

These two pairs of roots are a part of the roots of the original equation. As a matter of fact, the original equation can be written in factored form as follows:

$$(s + 1)(s - 1)(s + j5)(s - j5)(s + 2) = 0$$

Clearly, the original equation has one root with a positive real part.

**Relative stability analysis.** Routh's stability criterion provides the answer to the question of absolute stability. This, in many practical cases, is not sufficient. We usually require information about the relative stability of the system. A useful approach for examining relative stability is to shift the $s$-plane axis and apply Routh's stability criterion. That is, we substitute

$$s = \hat{s} - \sigma \qquad (\sigma = \text{constant})$$

into the characteristic equation of the system, write the polynomial in terms of $\hat{s}$, and apply Routh's stability criterion to the new polynomial in $\hat{s}$. The number of changes of sign in the first column of the array developed for the polynomial in $\hat{s}$ is equal to the number of roots that are located to the right of the vertical line $s = -\sigma$. Thus, this test reveals the number of roots that lie to the right of the vertical line $s = -\sigma$.

**Application of Routh's stability criterion to control system analysis.** Routh's stability criterion is of limited usefulness in linear control system analysis mainly because it does not suggest how to improve relative stability or how to stabilize an unstable system. It is possible, however, to determine the effects of changing one or two parameters of a system by examining the values that cause instability. In the following, we shall consider the problem of determining the stability range of a parameter value.

Consider the system shown in Figure 5–24. Let us determine the range of $K$ for stability. The closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{K}{s(s^2 + s + 1)(s + 2) + K}$$

The characteristic equation is

$$s^4 + 3s^3 + 3s^2 + 2s + K = 0$$

The array of coefficients becomes

$$\begin{array}{llll} s^4 & 1 & 3 & K \\ s^3 & 3 & 2 & 0 \\ s^2 & \frac{7}{3} & K & \\ s^1 & 2 - \frac{9}{7}K & & \\ s^0 & K & & \end{array}$$

For stability, $K$ must be positive, and all coefficients in the first column must be positive. Therefore,

$$\tfrac{14}{9} > K > 0$$

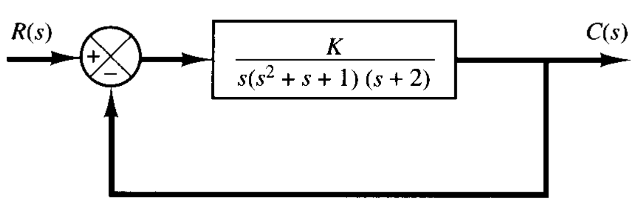



**Figure 5–24**
Control system.

When $K = \frac{14}{9}$, the system becomes oscillatory and, mathematically, the oscillation is sustained at constant amplitude.

## 5–6 PNEUMATIC CONTROLLERS

As the most versatile medium for transmitting signals and power, fluids, either as liquids or gases, have wide usage in industry. Liquids and gases can be distinguished basically by their relative incompressibilities and the fact that a liquid may have a free surface, whereas a gas expands to fill its vessel. In the engineering field the term *pneumatic* describes fluid systems that use air or gases and *hydraulic* applies to those using oil.

Pneumatic systems are extensively used in the automation of production machinery and in the field of automatic controllers. For instance, pneumatic circuits that convert the energy of compressed air into mechanical energy enjoy wide usage, and various types of pneumatic controllers are found in industry.

Since pneumatic systems and hydraulic systems are often compared, in what follows we shall give a brief comparison of these two kinds of systems.

**Comparison between pneumatic systems and hydraulic systems.** The fluid generally found in pneumatic systems is air; in hydraulic systems it is oil. And it is primarily the different properties of the fluids involved that characterize the differences between the two systems. These differences can be listed as follows:

1. Air and gases are compressible, whereas oil is incompressible.
2. Air lacks lubricating property and always contains water vapor. Oil functions as a hydraulic fluid as well as a lubricator.
3. The normal operating pressure of pneumatic systems is very much lower than that of hydraulic systems.
4. Output powers of pneumatic systems are considerably less than those of hydraulic systems.
5. Accuracy of pneumatic actuators is poor at low velocities, whereas accuracy of hydraulic actuators may be made satisfactory at all velocities.
6. In pneumatic systems, external leakage is permissible to a certain extent, but internal leakage must be avoided because the effective pressure difference is rather small. In hydraulic systems internal leakage is permissible to a certain extent, but external leakage must be avoided.
7. No return pipes are required in pneumatic systems when air is used, whereas they are always needed in hydraulic systems.
8. Normal operating temperature for pneumatic systems is 5° to 60°C (41° to 140°F). The pneumatic system, however, can be operated in the 0° to 200°C (32° to 392°F) range. Pneumatic systems are insensitive to temperature changes, in contrast to hydraulic systems, in which fluid friction due to viscosity depends greatly on temperature. Normal operating temperature for hydraulic systems is 20° to 70°C(68° to 158°F).
9. Pneumatic systems are fire- and explosion-proof, whereas hydraulic systems are not.

In what follows we begin with a mathematical modeling of pneumatic systems. Then we shall present pneumatic proportional controllers. We shall illustrate the fact that proportional controllers utilize the principle of negative feedback in themselves. We shall give a detailed discussion of the principle by which proportional controllers operate. Finally, we shall treat methods for obtaining derivative and integral control actions. Throughout the discussions, we shall place emphasis on the fundamental principles, rather than on the details of the operation of the actual mechanisms.

**Pneumatic systems.** The past decades have seen a great development in low-pressure pneumatic controllers for industrial control systems, and today they are used extensively in industrial processes. Reasons for their broad appeal include an explosion-proof character, simplicity, and ease of maintenance.

**Resistance and capacitance of pressure systems.** Many industrial processes and pneumatic controllers involve the flow of a gas or air through connected pipelines and pressure vessels.

Consider the pressure system shown in Figure 5–25(a). The gas flow through the restriction is a function of the gas pressure difference $p_i - p_o$. Such a pressure system may be characterized in terms of a resistance and a capacitance.

The gas flow resistance $R$ may be defined as follows:

$$R = \frac{\text{change in gas pressure difference, lb}_f/\text{ft}^2}{\text{change in gas flow rate, lb/sec}}$$

or

$$R = \frac{d\,(\Delta P)}{dq} \tag{5–8}$$

where $d(\Delta P)$ is a small change in the gas pressure difference and $dq$ is a small change in the gas flow rate. Computation of the value of the gas flow resistance $R$ may be quite time consuming. Experimentally, however, it can be easily determined from a plot of the pressure difference versus flow rate by calculating the slope of the curve at a given operating condition, as shown in Figure 5–25(b).

The capacitance of the pressure vessel may be defined by

$$C = \frac{\text{change in gas stored, lb}}{\text{change in gas pressure, lb}_f/\text{ft}^2}$$

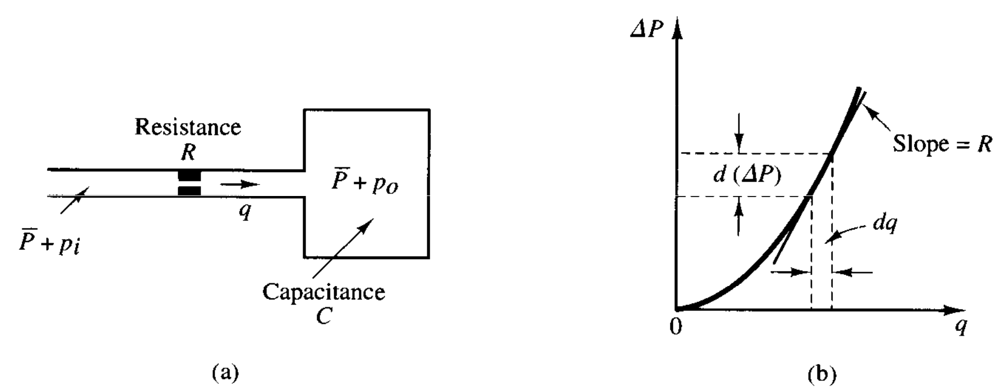



**Figure 5–25**
(a) Schematic diagram of a pressure system; (b) pressure difference versus flow rate curve.

or

$$C = \frac{dm}{dp} = V\frac{d\rho}{dp} \tag{5–9}$$

where $C$ = capacitance, lb-ft$^2$/lb$_f$
$m$ = mass of gas in vessel, lb
$p$ = gas pressure, lb$_f$/ft$^2$
$V$ = volume of vessel, ft$^3$
$\rho$ = density, lb/ft$^3$

The capacitance of the pressure system depends on the type of expansion process involved. The capacitance can be calculated by use of the ideal gas law. If the gas expansion process is polytropic and the change of state of the gas is between isothermal and adiabatic, then

$$p\left(\frac{V}{m}\right)^n = \frac{p}{\rho^n} = \text{constant} \tag{5–10}$$

where $n$ = polytropic exponent.

For ideal gases,

$$p\bar{v} = \bar{R}T \qquad \text{or} \qquad pv = \frac{\bar{R}}{M}T$$

where $p$ = absolute pressure, lb$_f$/ft$^2$
$\bar{v}$ = volume occupied by 1 mole of a gas, ft$^3$/lb-mole
$\bar{R}$ = universal gas constant, ft-lb$_f$/lb-mole °R
$T$ = absolute temperature, °R
$v$ = specific volume of gas, ft$^3$/lb
$M$ = molecular weight of gas per mole, lb/lb-mole

Thus

$$pv = \frac{p}{\rho} = \frac{\bar{R}}{M}T = R_{\text{gas}}T \tag{5–11}$$

where $R_{\text{gas}}$ = gas constant, ft-lb$_f$/lb °R.

The polytropic exponent $n$ is unity for isothermal expansion. For adiabatic expansion, $n$ is equal to the ratio of specific heats $c_p/c_v$, where $c_p$ is the specific heat at constant pressure and $c_v$ is the specific heat at constant volume. In many practical cases, the value of $n$ is approximately constant, and thus the capacitance may be considered constant. The value of $d\rho/dp$ is obtained from Equations (5–10) and (5–11) as

$$\frac{d\rho}{dp} = \frac{1}{nR_{\text{gas}}T}$$

The capacitance is then obtained as

$$C = \frac{V}{nR_{\text{gas}}T} \tag{5–12}$$

The capacitance of a given vessel is constant if the temperature stays constant. (In many practical cases, the polytropic exponent $n$ is approximately $1.0 \sim 1.2$ for gases in uninsulated metal vessels.)

**Pressure systems.** Consider the system shown in Figure 5–25(a). If we assume only small deviations in the variables from their respective steady-state values, then this system may be considered linear.

Let us define

$\bar{P}$ = gas pressure in the vessel at steady state (before changes in pressure have occurred), $\mathrm{lb_f/ft^2}$

$p_i$ = small change in inflow gas pressure, $\mathrm{lb_f/ft^2}$

$p_o$ = small change in gas pressure in the vessel, $\mathrm{lb_f/ft^2}$

$V$ = volume of the vessel, $\mathrm{ft^3}$

$m$ = mass of gas in vessel, lb

$q$ = gas flow rate, lb/sec

$\rho$ = density of gas, $\mathrm{lb/ft^3}$

For small values of $p_i$ and $p_o$, the resistance $R$ given by Equation (5–8) becomes constant and may be written as

$$R = \frac{p_i - p_o}{q}$$

The capacitance $C$ is given by Equation (5–9), or

$$C = \frac{dm}{dp}$$

Since the pressure change $dp_o$ times the capacitance $C$ is equal to the gas added to the vessel during $dt$ seconds, we obtain

$$C\,dp_o = q\,dt$$

or

$$C\frac{dp_o}{dt} = \frac{p_i - p_o}{R}$$

which can be written as

$$RC\frac{dp_o}{dt} + p_o = p_i$$

If $p_i$ and $p_o$ are considered the input and output, respectively, then the transfer function of the system is

$$\frac{P_o(s)}{P_i(s)} = \frac{1}{RCs + 1}$$

where $RC$ has the dimension of time and is the time constant of the system.

**Pneumatic nozzle–flapper amplifiers.** A schematic diagram of a pneumatic nozzle–flapper amplifier is shown in Figure 5–26(a). The power source for this amplifier is a supply of air at constant pressure. The nozzle–flapper amplifier converts small changes in the position of the flapper into large changes in the back pressure in the nozzle. Thus a large power output can be controlled by the very little power that is needed to position the flapper.

In Figure 5–26(a), pressurized air is fed through the orifice, and the air is ejected from the nozzle toward the flapper. Generally, the supply pressure $P_s$ for such a controller is 20 psig (1.4 $kg_f/cm^2$ gage). The diameter of the orifice is on the order of 0.01 in. (0.25 mm) and that of the nozzle is on the order of 0.016 in. (0.4 mm). To ensure proper functioning of the amplifier, the nozzle diameter must be larger than the orifice diameter.

In operating this system, the flapper is positioned against the nozzle opening. The nozzle back pressure $P_b$ is controlled by the nozzle–flapper distance $X$. As the flapper approaches the nozzle, the opposition to the flow of air through the nozzle increases, with the result that the nozzle back pressure $P_b$ increases. If the nozzle is completely closed by the flapper, the nozzle back pressure $P_b$ becomes equal to the supply pressure $P_s$. If the flapper is moved away from the nozzle, so that the nozzle–flapper distance is wide (on the order of 0.01 in.), then there is practically no restriction to flow, and the nozzle back pressure $P_b$ takes on a minimum value that depends on the nozzle–flapper device. (The lowest possible pressure will be the ambient pressure $P_a$.)

Note that, because the air jet puts a force against the flapper, it is necessary to make the nozzle diameter as small as possible.

A typical curve relating the nozzle back pressure $P_b$ to the nozzle–flapper distance $X$ is shown in Figure 5–26(b). The steep and almost linear part of the curve is utilized in the actual operation of the nozzle–flapper amplifier. Because the range of flapper displacements is restricted to a small value, the change in output pressure is also small, unless the curve is very steep.

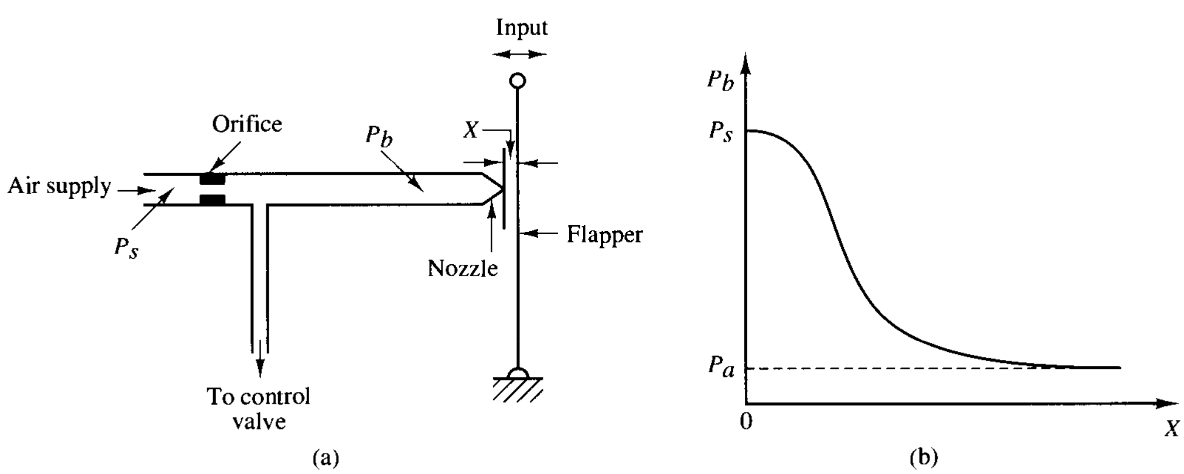



**Figure 5–26**
(a) Schematic diagram of a pneumatic nozzle–flapper amplifier; (b) characteristic curve relating nozzle back pressure and nozzle–flapper distance.

The nozzle–flapper amplifier converts displacement into a pressure signal. Since industrial process control systems require large output power to operate large pneumatic actuating valves, the power amplification of the nozzle–flapper amplifier is usually insufficient. Consequently, a pneumatic relay often serves as a power amplifier in connection with the nozzle–flapper amplifier.

**Pneumatic relays.** In practice, in a pneumatic controller, a nozzle–flapper amplifier acts as the first-stage amplifier and a pneumatic relay as the second-stage amplifier. The pneumatic relay is capable of handling a large quantity of airflow.

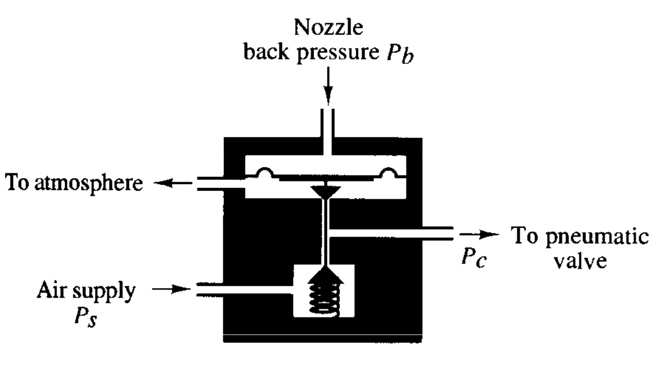
A schematic diagram of a pneumatic relay is shown in Figure 5–27(a). As the nozzle back pressure $P_b$ increases, the diaphragm valve moves downward. The opening to the atmosphere decreases and the opening to the pneumatic valve increases, thereby increasing the control pressure $P_c$. When the diaphragm valve closes the opening to the atmosphere, the control pressure $P_c$ becomes equal to the supply pressure $P_s$. When the nozzle back pressure $P_b$ decreases and the diaphragm valve moves upward and shuts off the air supply, the control pressure $P_c$ drops to the ambient pressure $P_a$. The control pressure $P_c$ can thus be made to vary from 0 psig to full supply pressure, usually 20 psig.

The total movement of the diaphragm valve is very small. In all positions of the valve, except at the position to shut off the air supply, air continues to bleed into the atmosphere, even after the equilibrium condition is attained between the nozzle back pressure and the control pressure. Thus the relay shown in Figure 5–27(a) is called a bleed-type relay.

There is another type of relay, the nonbleed type. In this one the air bleed stops when the equilibrium condition is obtained and, therefore, there is no loss of pressurized air at steady-state operation. Note, however, that the nonbleed-type relay must have an atmospheric relief to release the control pressure $P_c$ from the pneumatic actuating valve. A schematic diagram of a nonbleed-type relay is shown in Figure 5–27(b).

In either type of relay, the air supply is controlled by a valve, which is in turn controlled by the nozzle back pressure. Thus, the nozzle back pressure is converted into the control pressure with power amplification.

Since the control pressure $P_c$ changes almost instantaneously with changes in the nozzle back pressure $P_b$, the time constant of the pneumatic relay is negligible compared with the other larger time constants of the pneumatic controller and the plant.



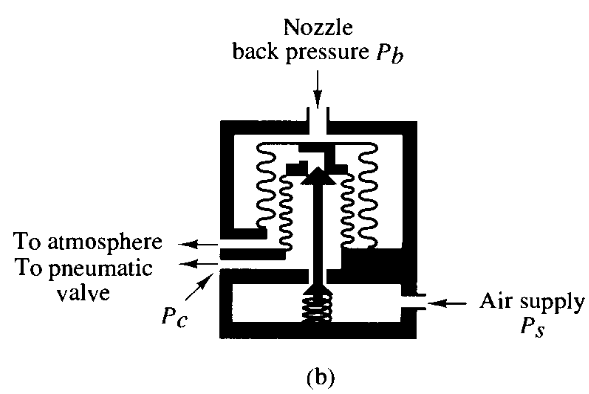




**Figure 5–27**
(a) Schematic diagram of a bleed-type relay; (b) schematic diagram of a nonbleed-type relay.

It is noted that some pneumatic relays are reverse acting. For example, the relay shown in Figure 5–28 is a reverse-acting relay. Here, as the nozzle back pressure $P_b$ increases, the ball valve is forced toward the lower seat, thereby decreasing the control pressure $P_c$. Thus, this relay is a reverse-acting relay.

**Pneumatic proportional controllers (force–distance type).** Two types of pneumatic controllers, one called the force–distance type and the other the force–balance type, are used extensively in industry. Regardless of how differently industrial pneumatic controllers may appear, careful study will show the close similarity in the functions of the pneumatic circuit. Here we shall consider the force–distance type of pneumatic controllers.

Figure 5–29(a) shows a schematic diagram of such a proportional controller. The nozzle–flapper amplifier constitutes the first-stage amplifier, and the nozzle back pressure is controlled by the nozzle–flapper distance. The relay-type amplifier constitutes the second-stage amplifier. The nozzle back pressure determines the position of the diaphragm valve for the second-stage amplifier, which is capable of handling a large quantity of airflow.

In most pneumatic controllers, some type of pneumatic feedback is employed. Feedback of the pneumatic output reduces the amount of actual movement of the flapper. Instead of mounting the flapper on a fixed point, as shown in Figure 5–29(b), it is often pivoted on the feedback bellows, as shown in Figure 5–29(c). The amount of feedback can be regulated by introducing a variable linkage between the feedback bellows and the flapper connecting point. The flapper then becomes a floating link. It can be moved by both the error signal and the feedback signal.

The operation of the controller shown in Figure 5–29(a) is as follows. The input signal to the two-stage pneumatic amplifier is the actuating error signal. Increasing the actuating error signal moves the flapper to the left. This move will, in turn, increase the nozzle back pressure, and the diaphragm valve moves downward. This results in an increase of the control pressure. This increase will cause bellows $F$ to expand and move the flapper to the right, thus opening the nozzle. Because of this feedback, the nozzle–flapper displacement is very small, but the change in the control pressure can be large.

It should be noted that proper operation of the controller requires that the feedback bellows move the flapper less than that movement caused by the error signal alone. (If these two movements were equal, no control action would result.)

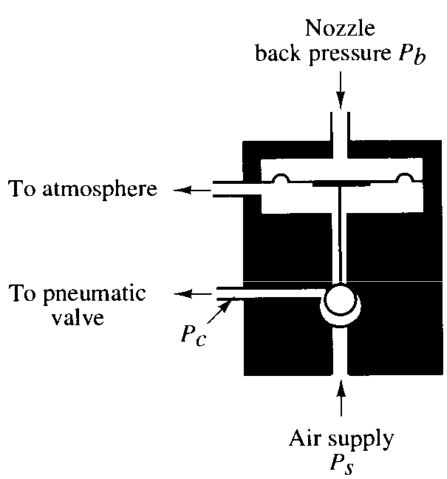



**Figure 5–28**
Reverse-acting relay.

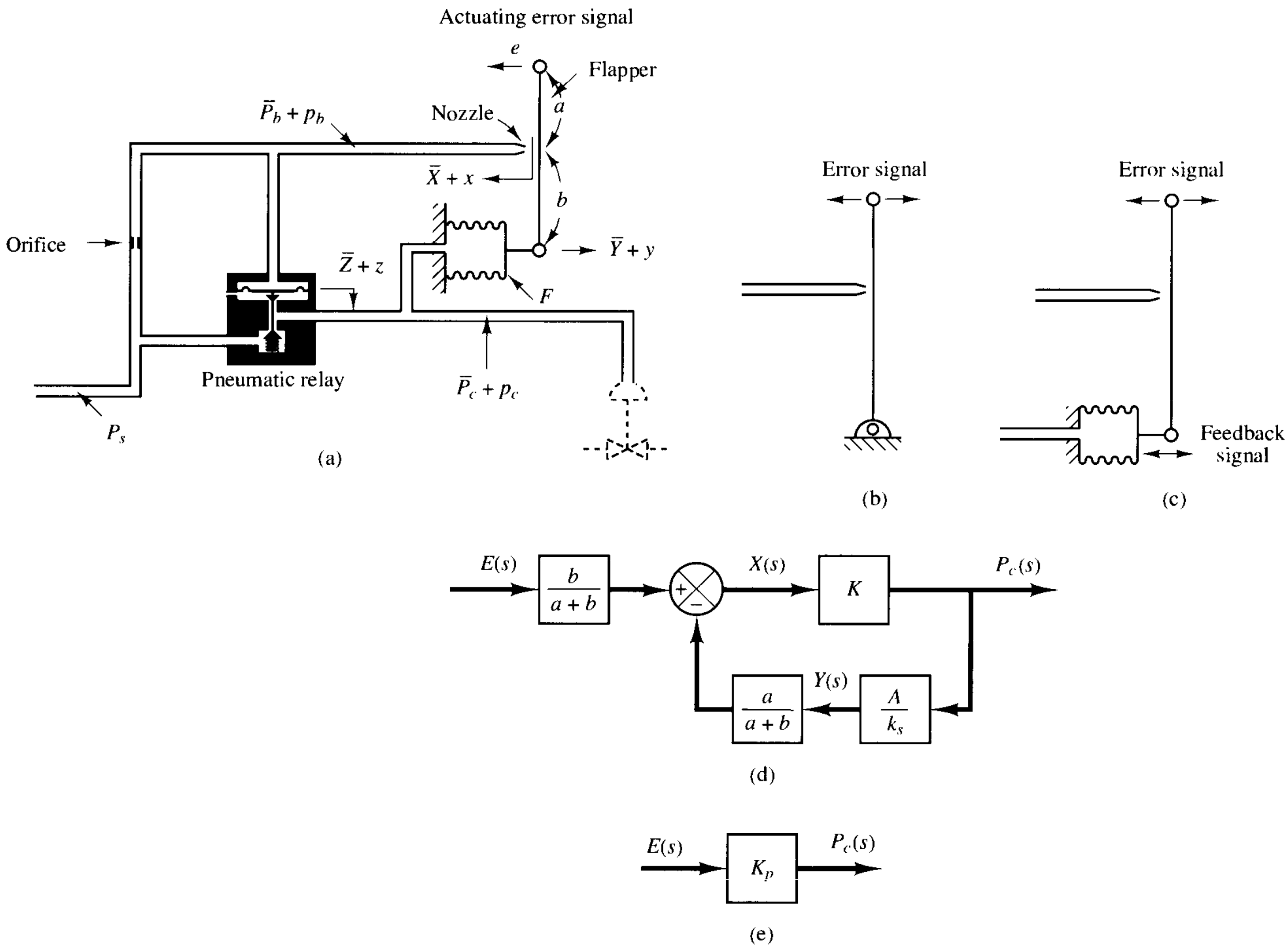



**Figure 5–29**
(a) Schematic diagram of a force–distance type of pneumatic proportional controller; (b) flapper mounted on a fixed point; (c) flapper mounted on a feedback bellows; (d) block diagram for the controller; (e) simplified block diagram for the controller.

Equations for this controller can be derived as follows. When the actuating error is zero, or $e = 0$, an equilibrium state exists with the nozzle–flapper distance equal to $\bar{X}$, the displacement of bellows equal to $\bar{Y}$, the displacement of the diaphragm equal to $\bar{Z}$, the nozzle back pressure equal to $\bar{P}_b$, and the control pressure equal to $\bar{P}_c$. When an actuating error exists, the nozzle–flapper distance, the displacement of the bellows, the displacement of the diaphragm, the nozzle back pressure, and the control pressure deviate from their respective equilibrium values. Let these deviations be $x$, $y$, $z$, $p_b$, and $p_c$, respectively. (The positive direction for each displacement variable is indicated by an arrowhead in the diagram.)

Assuming that the relationship between the variation in the nozzle back pressure and the variation in the nozzle–flapper distance is linear, we have

$$p_b = K_1 x \tag{5–13}$$

where $K_1$ is a positive constant. For the diaphragm valve,

$$p_b = K_2 z \tag{5–14}$$

where $K_2$ is a positive constant. The position of the diaphragm valve determines the control pressure. If the diaphragm valve is such that the relationship between $p_c$ and $z$ is linear, then

$$p_c = K_3 z \tag{5–15}$$

where $K_3$ is a positive constant. From Equations (5–13), (5–14), and (5–15), we obtain

$$p_c = \frac{K_3}{K_2} p_b = Kx \tag{5–16}$$

where $K = K_1 K_3 / K_2$ is a positive constant. For the flapper movement, we have

$$x = \frac{b}{a+b} e - \frac{a}{a+b} y \tag{5–17}$$

The bellows acts like a spring, and the following equation holds true:

$$A p_c = k_s y \tag{5–18}$$

where $A$ is the effective area of the bellows and $k_s$ is the equivalent spring constant, that is, the stiffness due to the action of the corrugated side of the bellows.

Assuming that all variations in the variables are within a linear range, we can obtain a block diagram for this system from Equations (5–16), (5–17), and (5–18) as shown in Figure 5–29(d). From Figure 5–29(d), it can be clearly seen that the pneumatic controller shown in Figure 5–29(a) itself is a feedback system. The transfer function between $p_c$ and $e$ is given by

$$\frac{P_c(s)}{E(s)} = \frac{\dfrac{b}{a+b} K}{1 + K \dfrac{a}{a+b} \dfrac{A}{k_s}} = K_p \tag{5–19}$$

A simplified block diagram is shown in Figure 5–29(e). Since $p_c$ and $e$ are proportional, the pneumatic controller shown in Figure 5–29(a) is called a *pneumatic proportional controller*. As seen from Equation (5–19), the gain of the pneumatic proportional controller can be widely varied by adjusting the flapper connecting linkage. [The flapper connecting linkage is not shown in Figure 5–29(a).] In most commercial proportional controllers an adjusting knob or other mechanism is provided for varying the gain by adjusting this linkage.

As noted earlier, the actuating error signal moved the flapper in one direction, and the feedback bellows moved the flapper in the opposite direction, but to a smaller degree. The effect of the feedback bellows is thus to reduce the sensitivity of the controller. The principle of feedback is commonly used to obtain wide proportional-band controllers.

Pneumatic controllers that do not have feedback mechanisms [which means that one end of the flapper is fixed, as shown in Figure 5–30(a)] have high sensitivity and are called *pneumatic two-position controllers* or *pneumatic on–off controllers*. In such a controller, only a small motion between the nozzle and the flapper is required to give a com-

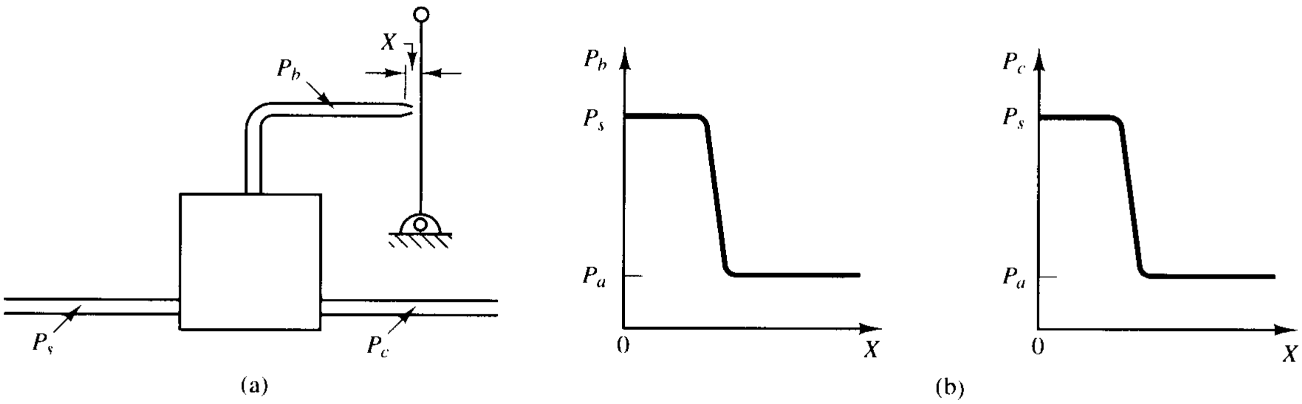



**Figure 5–30**
(a) Pneumatic controller without a feedback mechanism; (b) curves $P_b$ versus $X$ and $P_c$ versus $X$.

plete change from the maximum to the minimum control pressure. The curves relating $P_b$ to $X$ and $P_c$ to $X$ are shown in Figure 5–30(b). Notice that a small change in $X$ can cause a large change in $P_b$, which causes the diaphragm valve to be completely open or completely closed.

**Pneumatic proportional controllers (force–balance type).** Figure 5–31 shows a schematic diagram of a force-balance pneumatic proportional controller. Force-balance controllers are in extensive use in industry. Such controllers are called stack controllers. The basic principle of operation does not differ from that of the force-distance controller. The main advantage of the force-balance controller is that it eliminates many mechanical linkages and pivot joints, thereby reducing the effects of friction.

In what follows, we shall consider the principle of the force-balance controller. In the controller shown in Figure 5–31, the reference input pressure $P_r$ and the output pressure $P_o$ are fed to large diaphragm chambers. Note that a force-balance pneumatic controller operates only on pressure signals. Therefore, it is necessary to convert the reference input and system output to corresponding pressure signals.

As in the case of the force-distance controller, this controller employs a flapper, nozzle, and orifices. In Figure 5–31, the drilled opening in the bottom chamber is the nozzle. The diaphragm just above the nozzle acts as a flapper.

The operation of the force-balance controller shown in Figure 5–31 may be summarized as follows: 20-psig air from an air supply flows through an orifice, causing a reduced

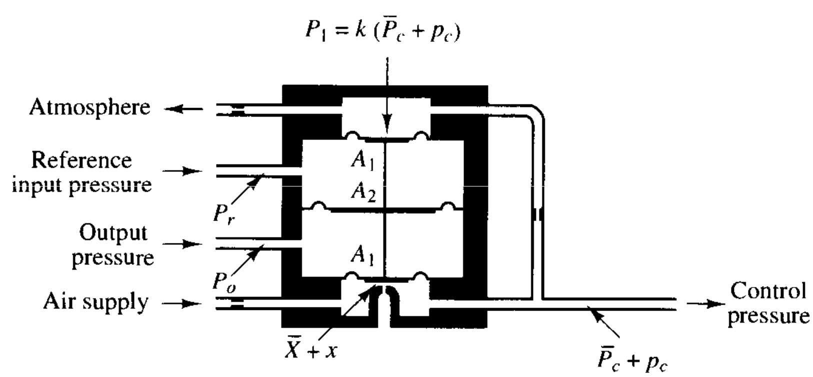



**Figure 5–31**
Schematic diagram of a force-balance type of pneumatic proportional controller.

pressure in the bottom chamber. Air in this chamber escapes to the atmosphere through the nozzle. The flow through the nozzle depends on the gap and the pressure drop across it. An increase in the reference input pressure $P_r$, while the output pressure $P_o$ remains the same, causes the valve stem to move down, decreasing the gap between the nozzle and the flapper diaphragm. This causes the control pressure $P_c$ to increase. Let

$$p_e = P_r - P_o \tag{5–20}$$

If $p_c = 0$, there is an equilibrium state with the nozzle–flapper distance equal to $\bar{X}$ and the control pressure equal to $\bar{P}_c$. At this equilibrium state, $P_1 = \bar{P}_c k$ (where $k < 1$) and

$$\bar{X} = \alpha(\bar{P}_c A_1 - \bar{P}_c k A_1) \tag{5–21}$$

where $\alpha$ is a constant.

Let us assume that $p_e \neq 0$ and define small variations in the nozzle–flapper distance and control pressure as $x$ and $p_c$, respectively. Then we obtain the following equation:

$$\bar{X} + x = \alpha[(\bar{P}_c + p_c)A_1 - (\bar{P}_c + p_c)kA_1 - p_e(A_2 - A_1)] \tag{5–22}$$

From Equations (5–21) and (5–22), we obtain

$$x = \alpha[p_c(1 - k)A_1 - p_e(A_2 - A_1)] \tag{5–23}$$

At this point, we must examine the quantity $x$. In the design of pneumatic controllers, the nozzle–flapper distance is made quite small. In view of the fact that $x/\alpha$ is a higher-order term than $p_c(1 - k)A_1$ or $p_e(A_2 - A_1)$, that is, for $p_c \neq 0$,

$$\frac{x}{\alpha} \ll p_c(1 - k)A_1$$

$$\frac{x}{\alpha} \ll p_e(A_2 - A_1)$$

we may neglect the term $x$ in our analysis. Equation (5–23) can then be rewritten to reflect this assumption as follows:

$$p_c(1 - k)A_1 = p_e(A_2 - A_1)$$

and the transfer function between $p_c$ and $p_e$ becomes

$$\frac{P_c(s)}{P_e(s)} = \frac{A_2 - A_1}{A_1} \frac{1}{1 - k} = K_p$$

where $p_e$ is defined by Equation (5–20). The controller shown in Figure 5–31 is a proportional controller. The value of gain $K_p$ increases as $k$ approaches unity. Note that the value of $k$ depends on the diameters of the orifices in the inlet and outlet pipes of the feedback chamber. (The value of $k$ approaches unity as the resistance to flow in the orifice of the inlet pipe is made smaller.)

**Pneumatic actuating valves.** One characteristic of pneumatic controls is that they almost exclusively employ pneumatic actuating valves. A pneumatic actuating valve can provide a large power output. (Since a pneumatic actuator requires a large power input to produce a large power output, it is necessary that a sufficient quantity of pressurized air be available.) In practical pneumatic actuating valves, the valve characteristics may not be linear; that is, the flow may not be directly proportional to the valve stem position, and also there may be other nonlinear effects, such as hysteresis.

Consider the schematic diagram of a pneumatic actuating valve shown in Figure 5–32. Assume that the area of the diaphragm is $A$. Assume also that when the actuating error is zero the control pressure is equal to $\bar{P}_c$ and the valve displacement is equal to $\bar{X}$.

In the following analysis, we shall consider small variations in the variables and linearize the pneumatic actuating valve. Let us define the small variation in the control pressure and the corresponding valve displacement to be $p_c$ and $x$, respectively. Since a small change in the pneumatic pressure force applied to the diaphragm repositions the load, consisting of the spring, viscous friction, and mass, the force balance equation becomes

$$Ap_c = m\ddot{x} + b\dot{x} + kx \tag{5–24}$$

where $m$ = mass of the valve and valve stem
$b$ = viscous-friction coefficient
$k$ = spring constant

If the force due to the mass and viscous friction are negligibly small, then Equation (5–24) can be simplified to:

$$Ap_c = kx$$

The transfer function between $x$ and $p_c$ thus becomes

$$\frac{X(s)}{P_c(s)} = \frac{A}{k} = K_c$$

where $X(s) = \mathscr{L}[x]$ and $P_c(s) = \mathscr{L}[p_c]$. If $q_i$, the change in flow through the pneumatic actuating valve, is proportional to $x$, the change in the valve-stem displacement, then

$$\frac{Q_i(s)}{X(s)} = K_q$$

where $Q_i(s) = \mathscr{L}[q_i]$ and $K_q$ is a constant. The transfer function between $q_i$ and $p_c$ becomes

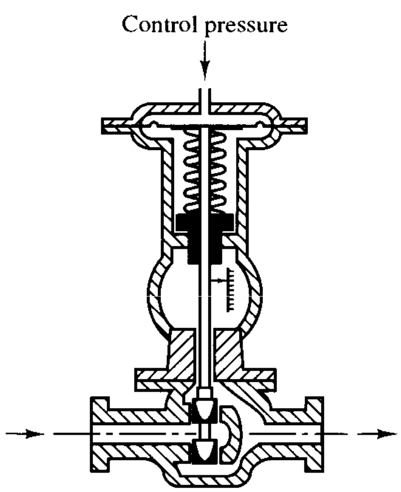



**Figure 5–32**
Schematic diagram of a pneumatic actuating valve.

$$\frac{Q_i(s)}{P_c(s)} = K_c K_q = K_v$$

where $K_v$ is a constant.

The standard control pressure for this kind of a pneumatic actuating valve is between 3 and 15 psig. The valve-stem displacement is limited by the allowable stroke of the diaphragm and is only a few inches. If a longer stroke is needed, a piston–spring combination may be employed.

In pneumatic actuating valves, the static-friction force must be limited to a low value so that excessive hysteresis does not result. Because of the compressibility of air, the control action may not be positive; that is, an error may exist in the valve-stem position. The use of a valve positioner results in improvements in the performance of a pneumatic actuating valve.

**Basic principle for obtaining derivative control action.** We shall now present methods for obtaining derivative control action. We shall again place the emphasis on the principle and not on the details of the actual mechanisms.

The basic principle for generating a desired control action is to insert the inverse of the desired transfer function in the feedback path. For the system shown in Figure 5–33, the closed-loop transfer function is

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

If $|G(s)H(s)| \gg 1$, then $C(s)/R(s)$ can be modified to

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

Thus, if proportional-plus-derivative control action is desired, we insert an element having the transfer function $1/(Ts + 1)$ in the feedback path.

Consider the pneumatic controller shown in Figure 5–34(a). Considering small changes in the variables, we can draw a block diagram of this controller as shown in Figure 5–34(b). From the block diagram we see that the controller is of proportional type.

We shall now show that the addition of a restriction in the negative feedback path will modify the proportional controller to a proportional-plus-derivative controller, commonly called a PD controller.

Consider the pneumatic controller shown in Figure 5–35(a). Assuming again small changes in the actuating error, nozzle–flapper distance, and control pressure, we can

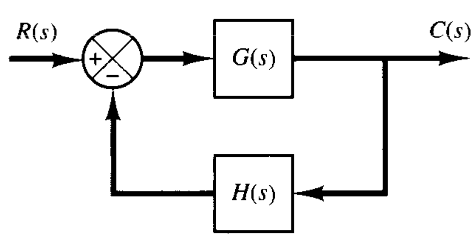



**Figure 5–33**
Control system.

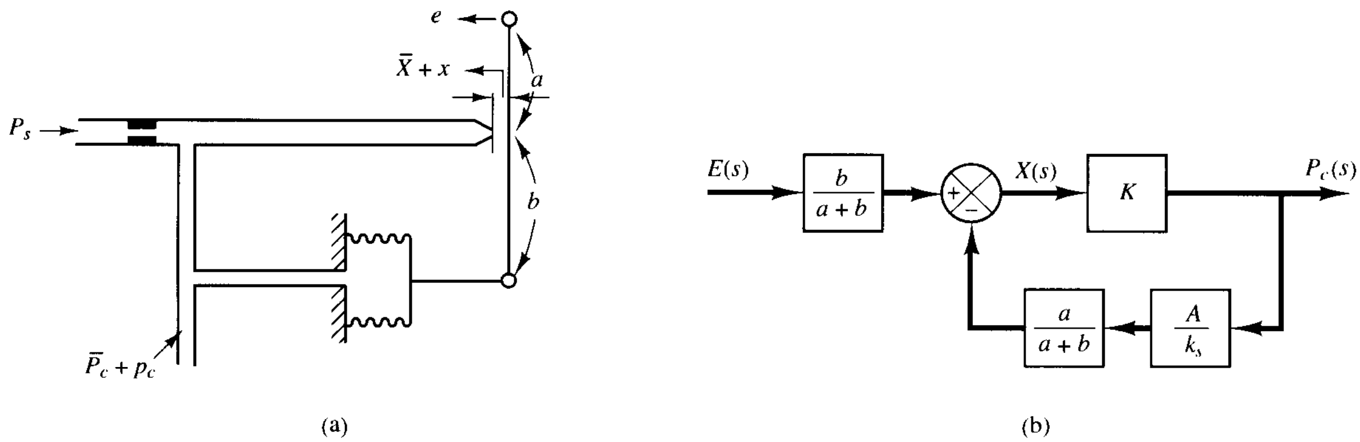



**Figure 5–34**
(a) Pneumatic proportional controller; (b) block diagram of the controller.

summarize the operation of this controller as follows: Let us first assume a small step change in $e$. Then the change in the control pressure $p_c$ will be instantaneous. The restriction $R$ will momentarily prevent the feedback bellows from sensing the pressure change $p_c$. Thus the feedback bellows will not respond momentarily, and the pneumatic actuating valve will feel the full effect of the movement of the flapper. As time goes on, the feedback bellows will expand or contract. The change in the nozzle-flapper distance $x$ and the change in the control pressure $p_c$ can be plotted against time $t$, as shown in Figure 5–35(b). At steady state, the feedback bellows acts like an ordinary feedback

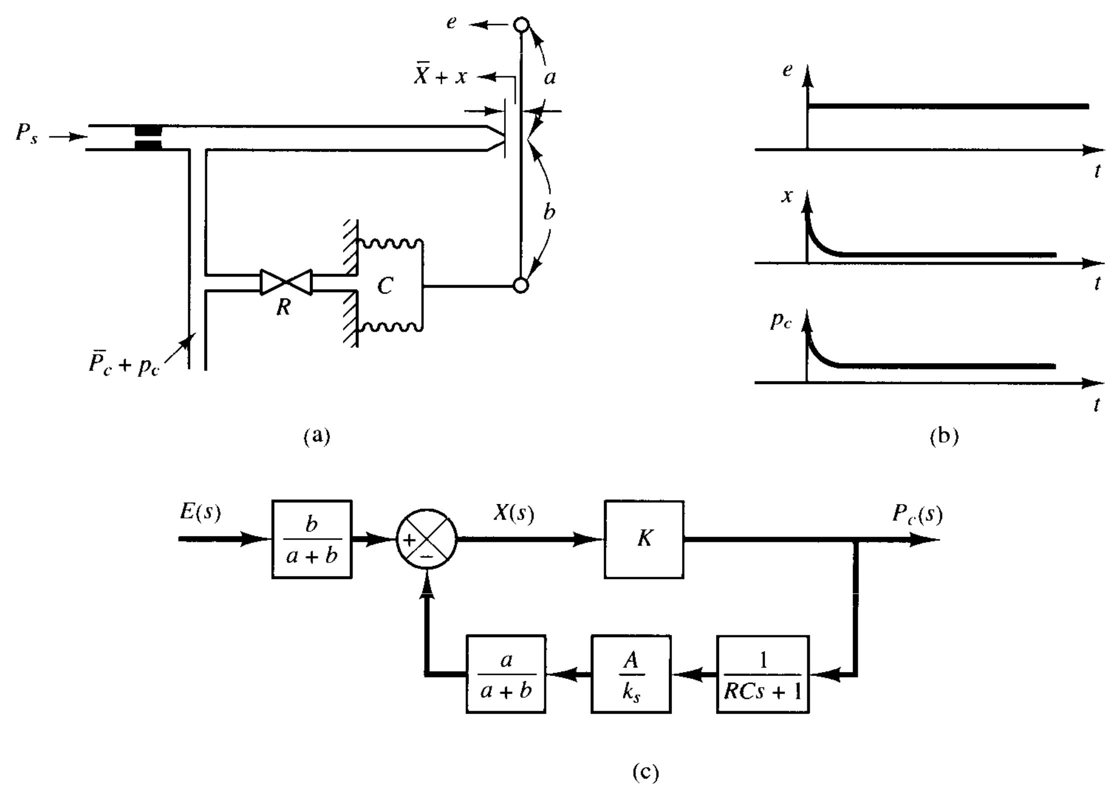



**Figure 5–35**
(a) Pneumatic proportional-plus-derivative controller; (b) step change in $e$ and the corresponding changes in $x$ and $p_c$ plotted versus $t$; (c) block diagram of the controller.

mechanism. The curve $p_c$ versus $t$ clearly shows that this controller is of the proportional-plus-derivative type.

A block diagram corresponding to this pneumatic controller is shown in Figure 5–35(c). In the block diagram, $K$ is a constant, $A$ is the area of the bellows, and $k_s$ is the equivalent spring constant of the bellows. The transfer function between $p_c$ and $e$ can be obtained from the block diagram as follows:

$$\frac{P_c(s)}{E(s)} = \frac{\dfrac{b}{a+b}K}{1 + \dfrac{Ka}{a+b}\dfrac{A}{k_s}\dfrac{1}{RCs+1}}$$

In such a controller the loop gain $|KaA/[(a+b)k_s(RCs+1)]|$ is normally very much greater than unity. Thus the transfer function $P_c(s)/E(s)$ can be simplified to give

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

where

$$K_p = \frac{bk_s}{aA}, \qquad T_d = RC$$

Thus, delayed negative feedback, or the transfer function $1/(RCs+1)$ in the feedback path, modifies the proportional controller to a proportional-plus-derivative controller.

Note that if the feedback valve is fully opened the control action becomes proportional. If the feedback valve is fully closed, the control action becomes narrow-band proportional (on–off).

**Obtaining pneumatic proportional-plus-integral control action.** Consider the proportional controller shown in Figure 5–34(a). Considering small changes in the variables, we can show that the addition of delayed positive feedback will modify this proportional controller to a proportional-plus-integral controller, commonly called a PI controller.

Consider the pneumatic controller shown in Figure 5–36(a). The operation of this controller is as follows: The bellows denoted by I is connected to the control pressure source without any restriction. The bellows denoted by II is connected to the control pressure source through a restriction. Let us assume a small step change in the actuating error. This will cause the back pressure in the nozzle to change instantaneously. Thus a change in the control pressure $p_c$ also occurs instantaneously. Due to the restriction of the valve in the path to bellows II, there will be a pressure drop across the valve. As time goes on, air will flow across the valve in such a way that the change in pressure in bellows II attains the value $p_c$. Thus bellows II will expand or contract as time elapses in such a way as to move the flapper an additional amount in the direction of the original

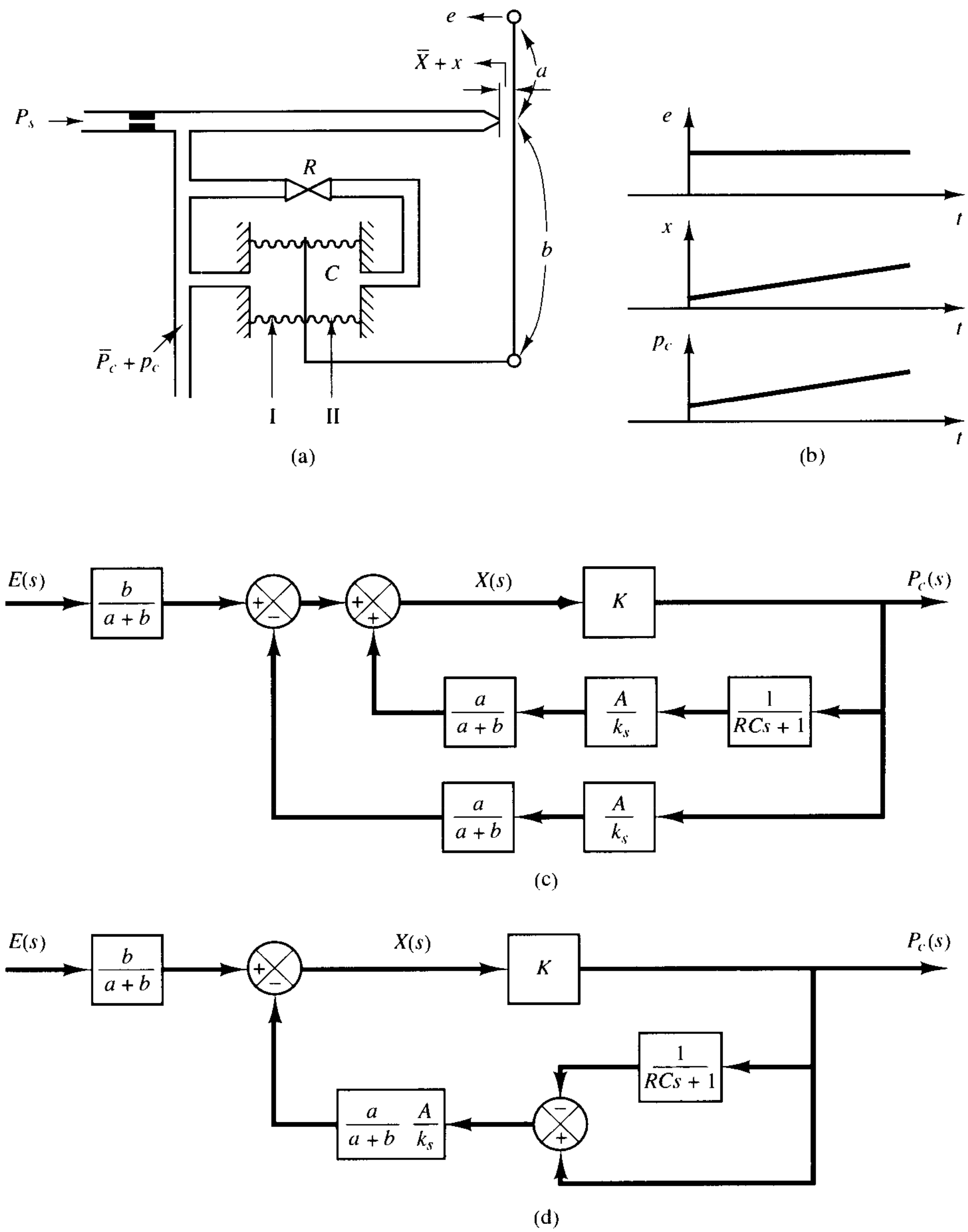



**Figure 5–36**
(a) Pneumatic proportional-plus-integral controller; (b) step change in $e$ and the corresponding changes in $x$ and $p_c$ plotted versus $t$; (c) block diagram of the controller; (d) simplified block diagram.

displacement $e$. This will cause the back pressure $p_c$ in the nozzle to change continuously, as shown in Figure 5–36(b).

Note that the integral control action in the controller takes the form of slowly canceling the feedback that the proportional control originally provided.

A block diagram of this controller under the assumption of small variations in the variables is shown in Figure 5–36(c). A simplification of this block diagram yields Figure 5–36(d). The transfer function of this controller is

$$\frac{P_c(s)}{E(s)} = \frac{\dfrac{b}{a+b}K}{1 + \dfrac{Ka}{a+b}\dfrac{A}{k_s}\left(1 - \dfrac{1}{RCs+1}\right)}$$

where $K$ is a constant, $A$ is the area of the bellows, and $k_s$ is the equivalent spring constant of the combined bellows. If $|KaARCs/[(a+b)k_s(RCs+1)]| \gg 1$, which is usually the case, the transfer function can be simplified to

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

where

$$K_p = \frac{bk_s}{aA}, \qquad T_i = RC$$

**Obtaining pneumatic proportional-plus-integral-plus-derivative control action.** A combination of the pneumatic controllers shown in Figures 5–35(a) and 5–36(a) yields a proportional-plus-integral-plus-derivative controller, commonly called a PID controller. Figure 5–37(a) shows a schematic diagram of such a controller. Figure 5–37(b) shows a block diagram of this controller under the assumption of small variations in the variables.

The transfer function of this controller is

$$\frac{P_c(s)}{E(s)} = \frac{\dfrac{bK}{a+b}}{1 + \dfrac{Ka}{a+b}\dfrac{A}{k_s}\dfrac{(R_iC - R_dC)s}{(R_dCs+1)(R_iCs+1)}}$$

By defining

$$T_i = R_iC, \qquad T_d = R_dC$$

and noting that under normal operation $|KaA(T_i - T_d)s/[(a+b)k_s(T_ds+1)(T_is+1)]| \gg 1$ and $T_i \gg T_d$, we obtain

$$\begin{aligned}\frac{P_c(s)}{E(s)} &\doteq \frac{bk_s}{aA}\frac{(T_ds+1)(T_is+1)}{(T_i - T_d)s} \\ &\doteq \frac{bk_s}{aA}\frac{T_dT_is^2 + T_is + 1}{T_is} \\ &= K_p\left(1 + \frac{1}{T_is} + T_ds\right) \qquad (5\text{–}25)\end{aligned}$$

where

$$K_p = \frac{bk_s}{aA}$$

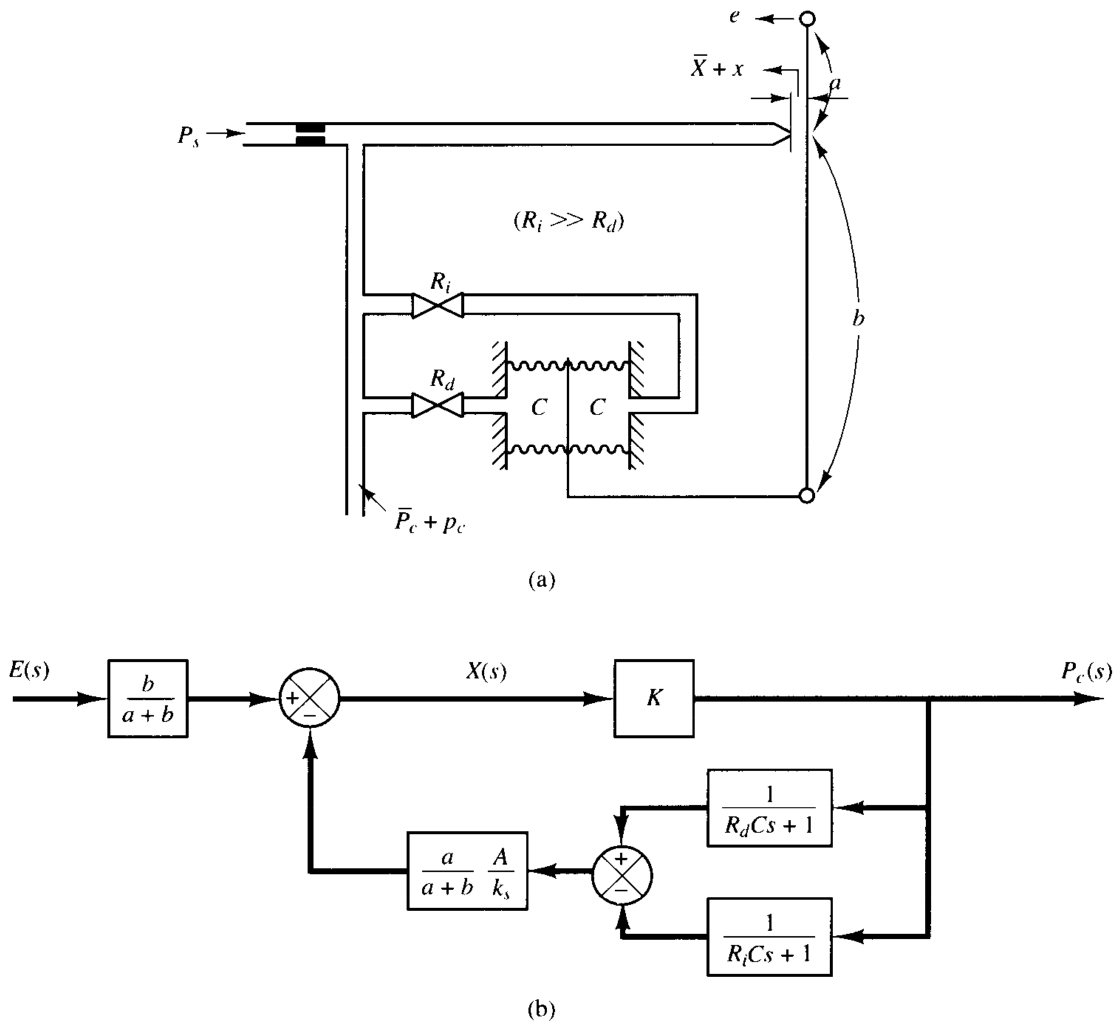



**Figure 5–37**
(a) Pneumatic proportional-plus-integral-plus-derivative controller; (b) block diagram of the controller.

Equation (5–25) indicates that the controller shown in Figure 5–37(a) is a proportional-plus-integral-plus-derivative controller (a PID controller).

## 5–7 HYDRAULIC CONTROLLERS

Except for low-pressure pneumatic controllers, compressed air has seldom been used for the continuous control of the motion of devices having significant mass under external load forces. For such a case, hydraulic controllers are generally preferred.

**Hydraulic systems.** The widespread use of hydraulic circuitry in machine tool applications, aircraft control systems, and similar operations occurs because of such factors as positiveness, accuracy, flexibility, high horsepower-to-weight ratio, fast starting, stopping, and reversal with smoothness and precision, and simplicity of operations.

The operating pressure in hydraulic systems is somewhere between 145 and 5000 $\text{lb}_f/\text{in.}^2$ (between 1 and 35 MPa). In some special applications, the operating pressure may go up to 10,000 $\text{lb}_f/\text{in.}^2$ (70 MPa). For the same power requirement, the weight and size of the hydraulic unit can be made smaller by increasing the supply pressure. With high-pressure hydraulic systems, very large force can be obtained. Rapid-acting, accurate positioning of heavy loads is possible with hydraulic systems. A combination of electronic and hydraulic systems is widely used because it combines the advantages of both electronic control and hydraulic power.

**Advantages and disadvantages of hydraulic systems.** There are certain advantages and disadvantages in using hydraulic systems rather than other systems. Some of the advantages are the following:

1. Hydraulic fluid acts as a lubricant, in addition to carrying away heat generated in the system to a convenient heat exchanger.
2. Comparatively small sized hydraulic actuators can develop large forces or torques.
3. Hydraulic actuators have a higher speed of response with fast starts, stops, and speed reversals.
4. Hydraulic actuators can be operated under continuous, intermittent, reversing, and stalled conditions without damage.
5. Availability of both linear and rotary actuators gives flexibility in design.
6. Because of low leakages in hydraulic actuators, speed drop when loads are applied is small.

On the other hand, several disadvantages tend to limit their use.

1. Hydraulic power is not readily available compared to electric power.
2. Cost of a hydraulic system may be higher than a comparable electrical system performing a similar function.
3. Fire and explosion hazards exist unless fire-resistant fluids are used.
4. Because it is difficult to maintain a hydraulic system that is free from leaks, the system tends to be messy.
5. Contaminated oil may cause failure in the proper functioning of a hydraulic system.
6. As a result of the nonlinear and other complex characteristics involved, the design of sophisticated hydraulic systems is quite involved.
7. Hydraulic circuits have generally poor damping characteristics. If a hydraulic circuit is not designed properly, some unstable phenomena may occur or disappear, depending on the operating condition.

**Comments.** Particular attention is necessary to ensure that the hydraulic system is stable and satisfactory under all operating conditions. Since the viscosity of hydraulic fluid can greatly affect damping and friction effects of the hydraulic circuits, stability tests must be carried out at the highest possible operating temperature.

Note that most hydraulic systems are nonlinear. Sometimes, however, it is possible to linearize nonlinear systems so as to reduce their complexity and permit solutions that are sufficiently accurate for most purposes. A useful linearization technique for dealing with nonlinear systems was presented in Section 3–10.

**Hydraulic integral controllers.** The hydraulic servomotor shown in Figure 5–38 is essentially a pilot-valve-controlled hydraulic power amplifier and actuator. The pilot valve is a balanced valve in the sense that the pressure forces acting on it are all bal-

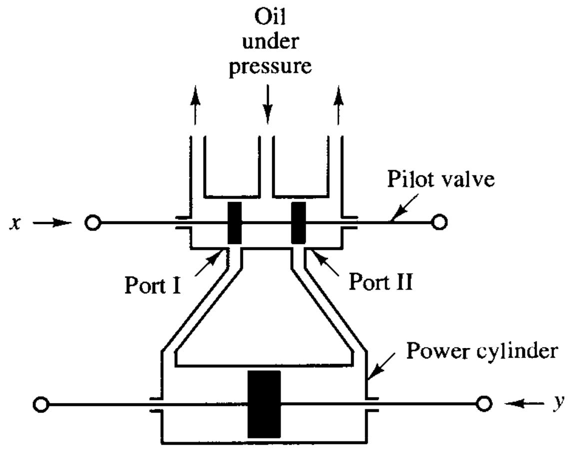



**Figure 5–38**
Hydraulic servomotor.

anced. A very large power output can be controlled by a pilot valve, which can be positioned with very little power.

It will be shown in the following that for negligibly small load mass the servomotor shown in Figure 5–38 acts as an integrator or an integral controller. Such a servomotor constitutes the basis of the hydraulic control circuit.

In the hydraulic servomotor shown in Figure 5–38, the pilot valve (a four-way valve) has two lands on the spool. If the width of the land is smaller than the port in the valve sleeve, the valve is said to be *underlapped. Overlapped* valves have a land width greater than the port width. A *zero-lapped* valve has a land width that is identical to the port width. (If the pilot valve is not a zero-lapped valve, analyses of hydraulic servomotors become very complicated.)

In the present analysis, we assume that hydraulic fluid is incompressible and that the inertia force of the power piston and load is negligible compared to the hydraulic force at the power piston. We also assume that the pilot valve is a zero-lapped valve, and the oil flow rate is proportional to the pilot valve displacement.

Operation of this hydraulic servomotor is as follows. If input $x$ moves the pilot valve to the right, port II is uncovered, and so high-pressure oil enters the right side of the power piston. Since port I is connected to the drain port, the oil in the left side of the power piston is returned to the drain. The oil flowing into the power cylinder is at high pressure; the oil flowing out from the power cylinder into the drain is at low pressure. The resulting difference in pressure on both sides of the power piston will cause it to move to the left.

Note that the rate of flow of oil $q$ (kg/sec) times $dt$ (sec) is equal to the power piston displacement $dy$(m) times the piston area $A$(m$^2$) times the density of oil $\rho$ (kg/m$^3$). Therefore,

$$A\rho\, dy = q\, dt \tag{5–26}$$

Because of the assumption that the oil flow rate $q$ is proportional to the pilot valve displacement $x$, we have

$$q = K_1 x \tag{5–27}$$

where $K_1$ is a positive constant. From Equations (5–26) and (5–27) we obtain

$$A\rho \frac{dy}{dt} = K_1 x$$

The Laplace transform of this last equation, assuming a zero initial condition, gives

$$A\rho sY(s) = K_1X(s)$$

or

$$\frac{Y(s)}{X(s)} = \frac{K_1}{A\rho s} = \frac{K}{s}$$

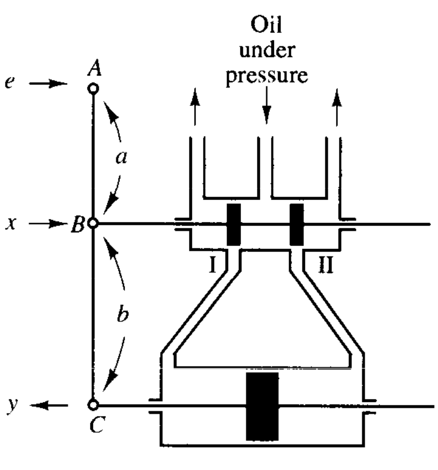
where $K = K_1/(A\rho)$. Thus the hydraulic servomotor shown in Figure 5–38 acts as an integral controller.

**Hydraulic proportional controllers.** It has been shown that the servomotor in Figure 5–38 acts as an integral controller. This servomotor can be modified to a proportional controller by means of a feedback link. Consider the hydraulic controller shown in Figure 5–39(a). The left side of the pilot valve is joined to the left side of the power piston by a link $ABC$. This link is a floating link rather than one moving about a fixed pivot.

The controller here operates in the following way. If input $e$ moves the pilot valve to the right, port II will be uncovered and high-pressure oil will flow through port II into the right side of the power piston and force this piston to the left. The power piston, in moving to the left, will carry the feedback link $ABC$ with it, thereby moving the pilot valve to the left. This action continues until the pilot piston again covers ports I and II. A block diagram of the system can be drawn as in Figure 5–39(b). The transfer function between $Y(s)$ and $E(s)$ is given by

$$\frac{Y(s)}{E(s)} = \frac{\dfrac{b}{a+b}\dfrac{K}{s}}{1 + \dfrac{K}{s}\dfrac{a}{a+b}}$$

$$= \frac{bK}{s(a+b) + Ka}$$



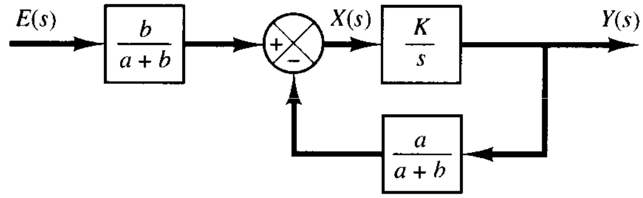




**Figure 5–39**
(a) Servomotor that acts as a proportional controller; (b) block diagram of the servomotor.

Noting that under the normal operating conditions we have $|Ka/[s(a + b)]| \gg 1$, this last equation can be simplified to

$$\frac{Y(s)}{E(s)} = \frac{b}{a} = K_p$$

The transfer function between $y$ and $e$ becomes a constant. Thus, the hydraulic controller shown in Figure 5–39(a) acts as a proportional controller, the gain of which is $K_p$. This gain can be adjusted by effectively changing the lever ratio $b/a$. (The adjusting mechanism is not shown in the diagram.)

We have thus seen that the addition of a feedback link will cause the hydraulic servomotor to act as a proportional controller.

**Dashpots.** The dashpot (also called a damper) shown in Figure 5–40(a) acts as a differentiating element. Suppose that we introduce a step displacement to the piston position $x$. Then the displacement $y$ becomes equal to $x$ momentarily. Because of the spring force, however, the oil will flow through the resistance $R$ and the cylinder will come back to the original position. The curves $x$ versus $t$ and $y$ versus $t$ are shown in Figure 5–40(b).

Let us derive the transfer function between the displacement $y$ and displacement $x$. Define the pressures existing on the right and left sides of the piston as $P_1$ (lb/in.$^2$) and $P_2$ (lb/in.$^2$), respectively. Suppose that the inertia force involved is negligible. Then the force acting on the piston must balance the spring force. Thus

$$A(P_1 - P_2) = ky$$

where $A$ = piston area, in.$^2$
$k$ = spring constant, $\text{lb}_f$/in.

The flow rate $q$ is given by

$$q = \frac{P_1 - P_2}{R}$$

where $q$ = flow rate through the restriction, lb/sec
$R$ = resistance to flow at the restriction, $\text{lb}_f$-sec/in.$^2$-lb

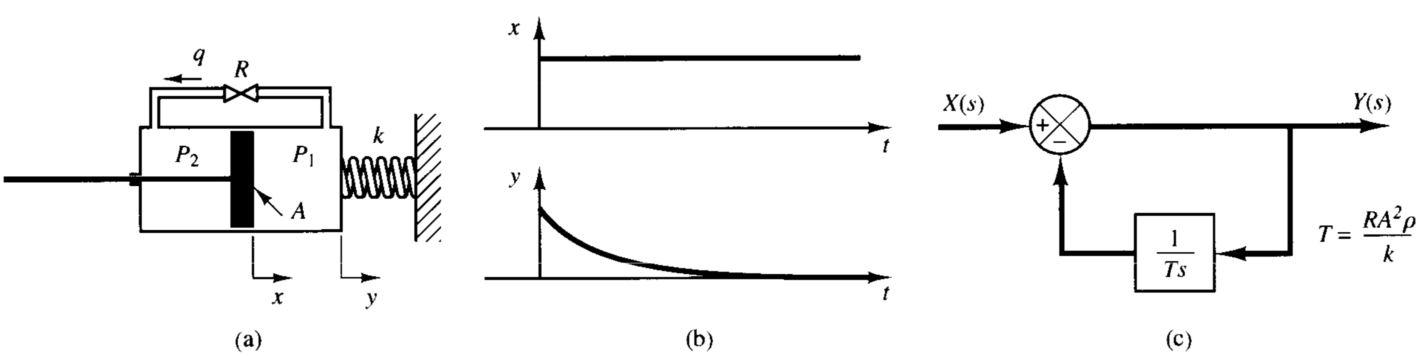



**Figure 5–40**
(a) Dashpot; (b) step change in $x$ and the corresponding change in $y$ plotted versus $t$; (c) block diagram of the dashpot.

Since the flow through the restriction during $dt$ seconds must equal the change in the mass of oil to the left of the piston during the same $dt$ seconds, we obtain

$$q\,dt = A\rho(dx - dy)$$

where $\rho$ = density, lb/in.$^3$. (We assume that the fluid is incompressible or $\rho$ = constant.) This last equation can be rewritten as

$$\frac{dx}{dt} - \frac{dy}{dt} = \frac{q}{A\rho} = \frac{P_1 - P_2}{RA\rho} = \frac{ky}{RA^2\rho}$$

or

$$\frac{dx}{dt} = \frac{dy}{dt} + \frac{ky}{RA^2\rho}$$

Taking the Laplace transforms of both sides of this last equation, assuming zero initial conditions, we obtain

$$sX(s) = sY(s) + \frac{k}{RA^2\rho}Y(s)$$

The transfer function of this system thus becomes

$$\frac{Y(s)}{X(s)} = \frac{s}{s + \dfrac{k}{RA^2\rho}}$$

Let us define $RA^2\rho/k = T$. Then

$$\frac{Y(s)}{X(s)} = \frac{Ts}{Ts + 1} = \frac{1}{1 + \dfrac{1}{Ts}}$$

Figure 5–40(c) shows a block diagram representation for this system.

**Obtaining hydraulic proportional-plus-integral control action.** Figure 5–41(a) shows a schematic diagram of a hydraulic proportional-plus-integral controller. A block diagram of this controller is shown in Figure 5–41(b). The transfer function $Y(s)/E(s)$ is given by

$$\frac{Y(s)}{E(s)} = \frac{\dfrac{b}{a+b}\dfrac{K}{s}}{1 + \dfrac{Ka}{a+b}\dfrac{T}{Ts+1}}$$

In such a controller, under normal operation $|KaT/[(a + b)(Ts + 1)]| \gg 1$, with the result that

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

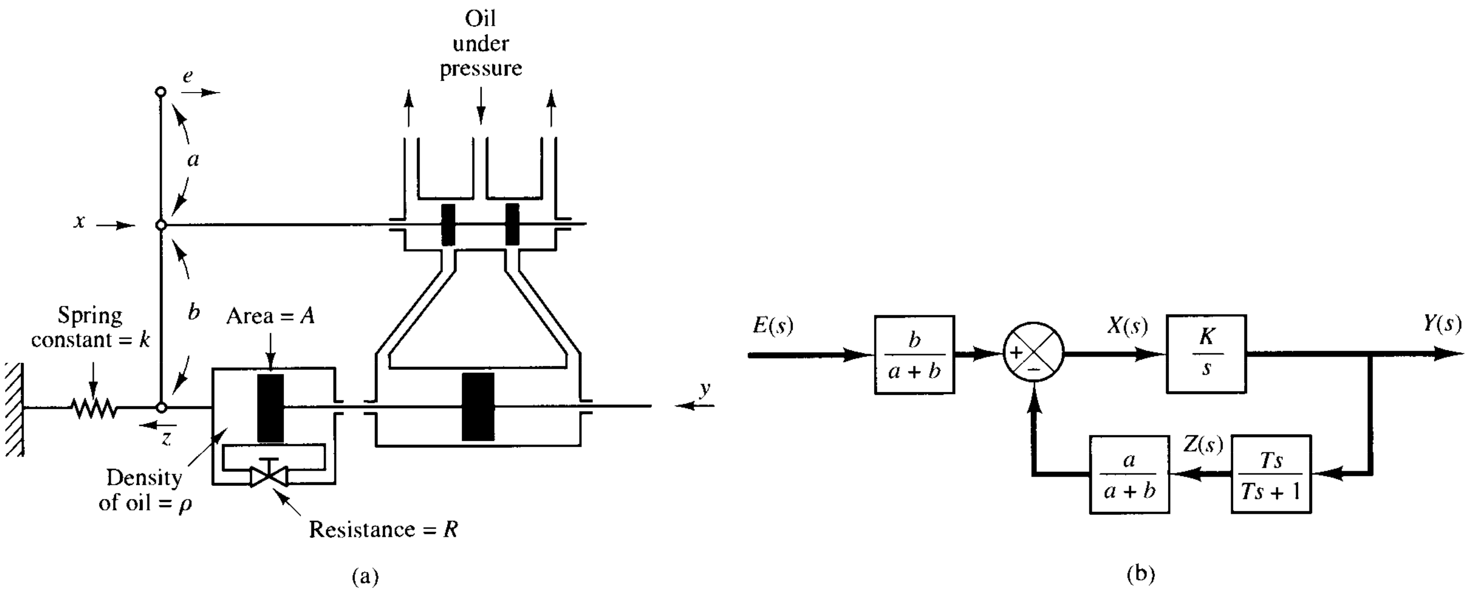



**Figure 5–41**
(a) Schematic diagram of a hydraulic proportional-plus-integral controller; (b) block diagram of the controller.

where

$$K_p = \frac{b}{a}, \qquad T_i = T = \frac{RA^2\rho}{k}$$

Thus the controller shown in Figure 5–52(a) is a proportional-plus-integral controller (a PI controller.)

**Obtaining hydraulic proportional-plus-derivative control action.** Figure 5–42(a) shows a schematic diagram of a hydraulic proportional-plus-derivative

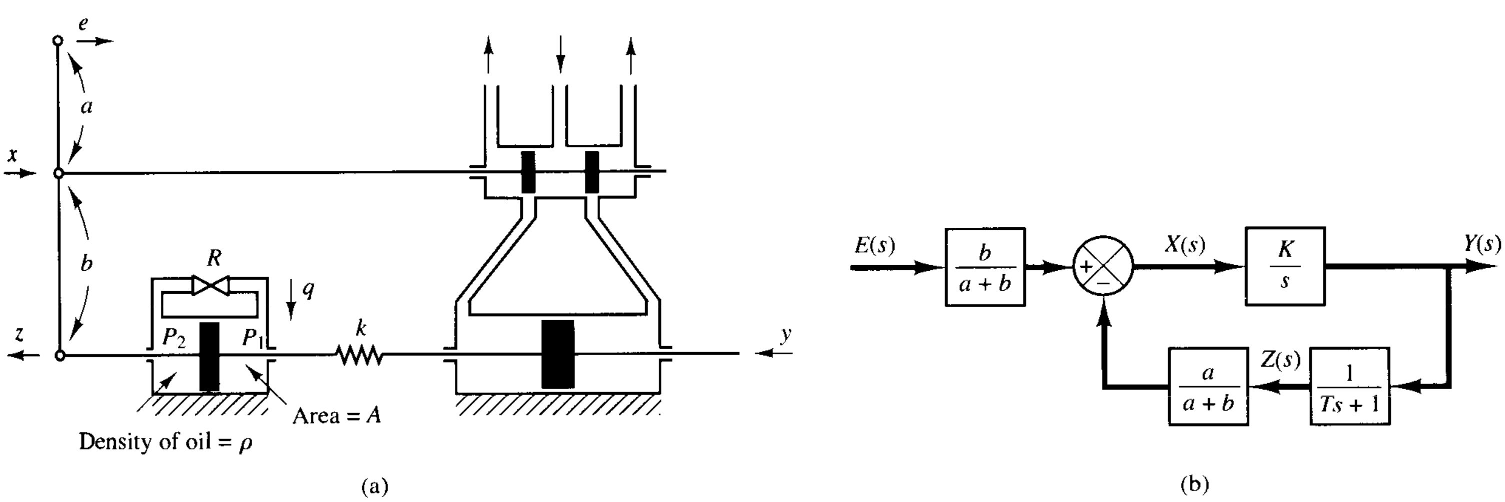



**Figure 5–42**
(a) Schematic diagram of a hydraulic proportional-plus-derivative controller; (b) block diagram of the controller.

controller. The cylinders are fixed in space and the pistons can move. For this system, notice that

$$k(y - z) = A(P_2 - P_1)$$

$$q = \frac{P_2 - P_1}{R}$$

$$q\,dt = \rho A\,dz$$

Hence

$$y = z + \frac{A}{k}qR = z + \frac{RA^2\rho}{k}\frac{dz}{dt}$$

or

$$\frac{Z(s)}{Y(s)} = \frac{1}{Ts + 1}$$

where

$$T = \frac{RA^2\rho}{k}$$

A block diagram for this system is shown in Figure 5–42(b). From the block diagram the transfer function $Y(s)/E(s)$ can be obtained as

$$\frac{Y(s)}{E(s)} = \frac{\dfrac{b}{a+b}\dfrac{K}{s}}{1 + \dfrac{a}{a+b}\dfrac{K}{s}\dfrac{1}{Ts+1}}$$

Under normal operation we have $|aK/[(a + b)s(Ts + 1)]| \gg 1$. Hence

$$\frac{Y(s)}{E(s)} = K_p(1 + Ts)$$

where

$$K_p = \frac{b}{a}, \qquad T = \frac{RA^2\rho}{k}$$

Thus the controller shown in Figure 5–42(a) is a proportional-plus-derivative controller (a PD controller).

## 5–8 ELECTRONIC CONTROLLERS

This section discusses electronic controllers using operational amplifiers. We begin by deriving the transfer functions of simple operational-amplifier circuits. Then we derive the transfer functions of some of the operational-amplifier controllers. Finally, we give operational-amplifier controllers and their transfer functions in the form of a table.

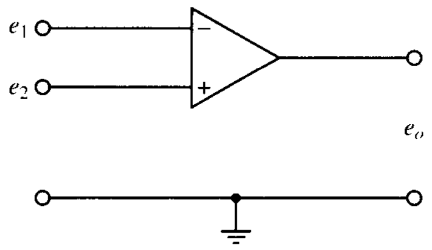



**Figure 5–43**
Operational amplifier.

**Operational amplifiers.** Operational amplifiers, often called op amps, are frequently used to amplify signals in sensor circuits. Op amps are also frequently used in filters used for compensation purposes. Figure 5–43 shows an op amp. It is a common practice to choose the ground as 0 volt and measure the input voltages $e_1$ and $e_2$ relative to the ground. The input $e_1$ to the minus terminal of the amplifier is inverted, and the input $e_2$ to the plus terminal is not inverted. The total input to the amplifier thus becomes $e_2 - e_1$. Hence, for the circuit shown in Figure 5–54, we have

$$e_o = K(e_2 - e_1) = -K(e_1 - e_2)$$

where the inputs $e_1$ and $e_2$ may be dc or ac signals and $K$ is the differential gain or voltage gain. The magnitude of $K$ is approximately $10^5 \sim 10^6$ for dc signals and ac signals with frequencies less than approximately 10 Hz. (The differential gain $K$ decreases with the signal frequency and becomes about unity for frequencies of 1 MHz ~ 50 MHz.) Note that the op amp amplifies the difference in voltages $e_1$ and $e_2$. Such an amplifier is commonly called a differential amplifier. Since the gain of the op amp is very high, it is necessary to have a negative feedback from the output to the input to make the amplifier stable. (The feedback is made from the output to the inverted input so that the feedback is a negative feedback.)

In the ideal op amp, no current flows into the input terminals, and the output voltage is not affected by the load connected to the output terminal. In other words, the input impedance is infinity and the output impedance is zero. In an actual op amp, a very small (almost negligible) current flows into an input terminal and the output cannot be loaded too much. In our analysis here, we make the assumption that the op amps are ideal.

**Inverting amplifier.** Consider the operational amplifier circuit shown in Figure 5–44. Let us obtain the output voltage $e_o$.

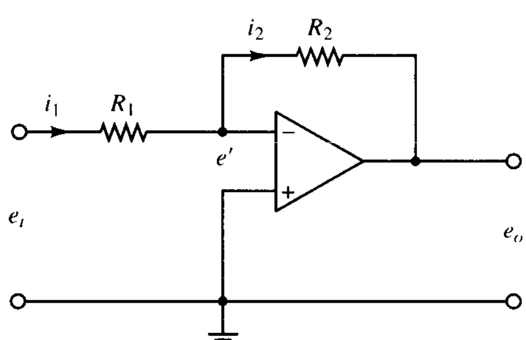



**Figure 5–44**
Inverting amplifier.

The equation for this circuit can be obtained as follows: Define

$$i_1 = \frac{e_i - e'}{R_1}, \qquad i_2 = \frac{e' - e_o}{R_2}$$

Since only a negligible current flows into the amplifier, the current $i_1$ must be equal to current $i_2$. Thus

$$\frac{e_i - e'}{R_1} = \frac{e' - e_o}{R_2}$$

Since $K(0 - e') = e_0$ and $K \gg 1$, $e'$ must be almost zero, or $e' \doteqdot 0$. Hence we have

$$\frac{e_i}{R_1} = \frac{-e_o}{R_2}$$

or

$$e_o = -\frac{R_2}{R_1} e_i$$

Thus the circuit shown is an inverting amplifier. If $R_1 = R_2$, then the op-amp circuit shown acts as a sign inverter.

**Noninverting amplifier.** Figure 5–45(a) shows a noninverting amplifier. A circuit equivalent to this one is shown in Figure 5–45(b). For the circuit of Figure 5–45(b), we have

$$e_o = K\left(e_i - \frac{R_1}{R_1 + R_2} e_o\right)$$

where $K$ is the differential gain of the amplifier. From this last equation, we get

$$e_i = \left(\frac{R_1}{R_1 + R_2} + \frac{1}{K}\right) e_o$$

Since $K \gg 1$, if $R_1/(R_1 + R_2) \gg 1/K$, then

$$e_o = \left(1 + \frac{R_2}{R_1}\right) e_i$$

This equation gives the output voltage $e_o$. Since $e_o$ and $e_i$ have the same signs, the op-amp circuit shown in Figure 5–45(a) is noninverting.

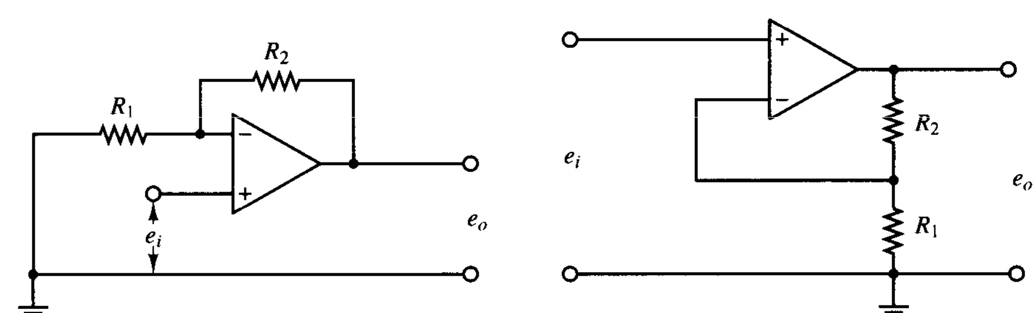



**Figure 5–45**
(a) Noninverting operational amplifier; (b) equivalent circuit.

**EXAMPLE 5–3** Figure 5–46 shows an electrical circuit involving an operational amplifier. Obtain the output $e_o$.

Let us define

$$i_1 = \frac{e_i - e'}{R_1}, \qquad i_2 = C\frac{d(e' - e_o)}{dt}, \qquad i_3 = \frac{e' - e_o}{R_2}$$

Noting that the current flowing into the amplifier is negligible, we have

$$i_1 = i_2 + i_3$$

Hence

$$\frac{e_i - e'}{R_1} = C\frac{d(e' - e_o)}{dt} + \frac{e' - e_o}{R_2}$$

Since $e' \doteqdot 0$, we have

$$\frac{e_i}{R_1} = -C\frac{de_o}{dt} - \frac{e_o}{R_2}$$

Taking the Laplace transform of this last equation, assuming the zero initial condition, we have

$$\frac{E_i(s)}{R_1} = -\frac{R_2Cs + 1}{R_2}E_o(s)$$

which can be written as

$$\frac{E_o(s)}{E_i(s)} = -\frac{R_2}{R_1}\frac{1}{R_2Cs + 1}$$

The op-amp circuit shown in Figure 5–46 is a first-order lag circuit. (Several other circuits involving op amps are shown in Table 5–1 together with their transfer functions.)

**Impedance approach for obtaining transfer functions.** Consider the op-amp circuit shown in Figure 5–47. Similar to the case of electrical circuits we discussed earlier, the impedance approach can be applied to op-amp circuits to obtain their transfer functions. For the circuit shown in Figure 5–47, we have

$$E_i(s) = Z_1(s)I(s), \qquad E_o(s) = -Z_2(s)I(s)$$

Hence, the transfer function for the circuit is obtained as

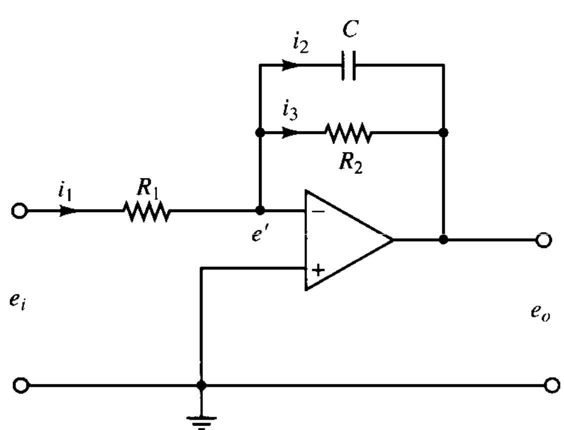



**Figure 5–46**
First-order lag circuit using operational amplifier.

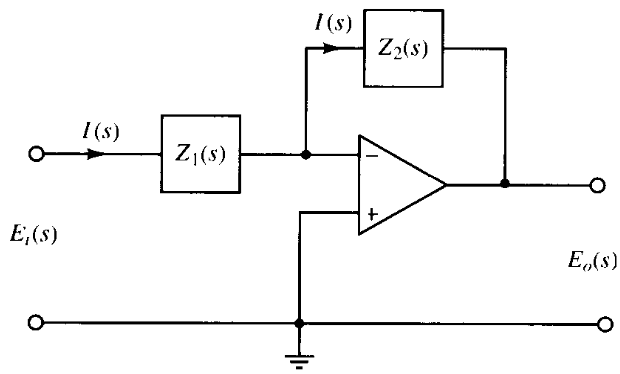



**Figure 5–47**
Operational amplifier circuit.

$$\frac{E_o(s)}{E_i(s)} = -\frac{Z_2(s)}{Z_1(s)}$$

**EXAMPLE 5–4** Referring to the op-amp circuit shown in Figure 5–46, obtain the transfer function $E_o(s)/E_i(s)$ by use of the impedance approach.

The complex impedances $Z_1(s)$ and $Z_2(s)$ for this circuit are

$$Z_1(s) = R_1 \qquad \text{and} \qquad Z_2(s) = \frac{1}{Cs + \dfrac{1}{R_2}} = \frac{R_2}{R_2Cs + 1}$$

Hence, $E_i(s)$ and $E_o(s)$ are obtained as

$$E_i(s) = R_1 I(s), \qquad E_o(s) = -\frac{R_2}{R_2Cs + 1} I(s)$$

The transfer function $E_o(s)/E_i(s)$ is, therefore, obtained as

$$\frac{E_o(s)}{E_i(s)} = -\frac{R_2}{R_1}\frac{1}{R_2Cs + 1}$$

which is, of course, the same as that obtained in Example 5–3.

**Lead or lag networks using operational amplifiers.** Figure 5–48(a) shows an electronic circuit using an operational amplifier. The transfer function for this circuit

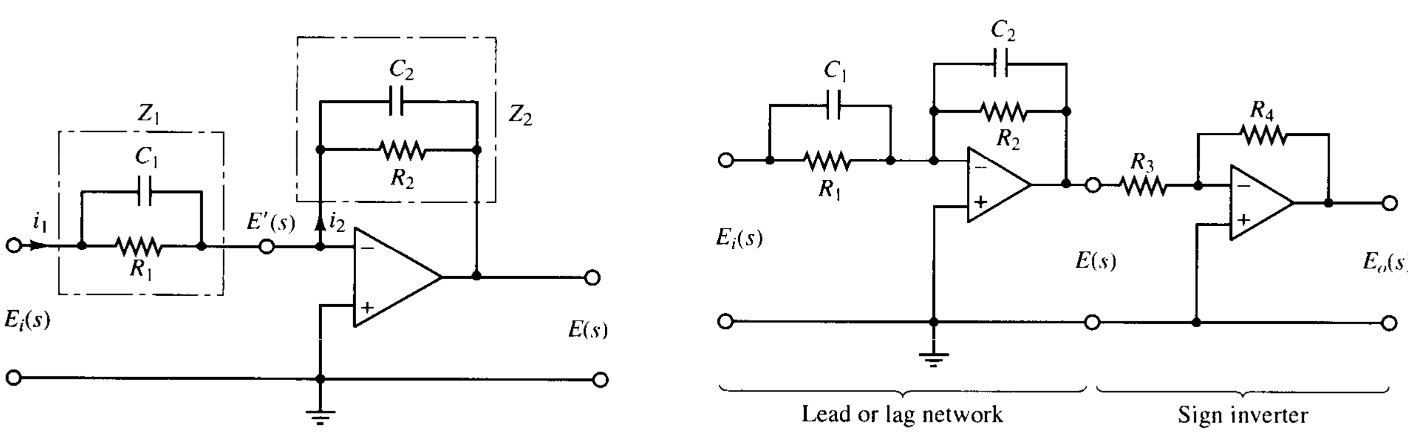



**Figure 5–48**
(a) Operational-amplifier circuit; (b) operational-amplifier circuit used as a lead or lag compensator.

can be obtained as follows: Define the input impedance and feedback impedance as $Z_1$ and $Z_2$, respectively. Then

$$Z_1 = \frac{R_1}{R_1C_1s + 1}, \qquad Z_2 = \frac{R_2}{R_2C_2s + 1}$$

Since the current flowing into the amplifier is negligible, current $i_1$ is equal to current $i_2$. Thus $i_1 = i_2$, or

$$\frac{E_i(s) - E'(s)}{Z_1} = \frac{E'(s) - E(s)}{Z_2}$$

Since $E'(s) \doteqdot 0$, we have

$$\frac{E(s)}{E_i(s)} = -\frac{Z_2}{Z_1} = -\frac{R_2}{R_1}\frac{R_1C_1s + 1}{R_2C_2s + 1} = -\frac{C_1}{C_2}\frac{s + \dfrac{1}{R_1C_1}}{s + \dfrac{1}{R_2C_2}} \tag{5–28}$$

Notice that the transfer function in Equation (5–28) contains a minus sign. Thus, this circuit is sign inverting. If such a sign inversion is not convenient in the actual application, a sign inverter may be connected to either the input or the output of the circuit of Figure 5–48(a). An example is shown in Figure 5–48(b). The sign inverter has the transfer function of

$$\frac{E_o(s)}{E(s)} = -\frac{R_4}{R_3}$$

The sign inverter has the gain of $-R_4/R_3$. Hence the network shown in Figure 5–48(b) has the following transfer function:

$$\frac{E_o(s)}{E_i(s)} = \frac{R_2R_4}{R_1R_3}\frac{R_1C_1s + 1}{R_2C_2s + 1} = \frac{R_4C_1}{R_3C_2}\frac{s + \dfrac{1}{R_1C_1}}{s + \dfrac{1}{R_2C_2}}$$

$$= K_c\alpha\frac{Ts + 1}{\alpha Ts + 1} = K_c\frac{s + \dfrac{1}{T}}{s + \dfrac{1}{\alpha T}} \tag{5–29}$$

where

$$T = R_1C_1, \qquad \alpha T = R_2C_2, \qquad K_c = \frac{R_4C_1}{R_3C_2}$$

Notice that

$$K_c\alpha = \frac{R_4C_1}{R_3C_2}\frac{R_2C_2}{R_1C_1} = \frac{R_2R_4}{R_1R_3}, \qquad \alpha = \frac{R_2C_2}{R_1C_1}$$

This network has a dc gain of $K_c\alpha = R_2R_4/(R_1R_3)$.

Referring to Equation (5–29), this network is a lead network if $R_1C_1 > R_2C_2$, or $a < 1$. It is a lag network if $R_1C_1 < R_2C_2$. (For the definitions of lead and lag networks, refer to Section 5–9.)

**PID controller using operational amplifiers.** Figure 5–49 shows an electronic proportional-plus-integral-plus-derivative controller (a PID controller) using operational amplifiers. The transfer function $E(s)/E_i(s)$ is given by

$$\frac{E(s)}{E_i(s)} = -\frac{Z_2}{Z_1}$$

where

$$Z_1 = \frac{R_1}{R_1C_1s + 1}, \qquad Z_2 = \frac{R_2C_2s + 1}{C_2s}$$

Thus

$$\frac{E(s)}{E_i(s)} = -\left(\frac{R_2C_2s + 1}{C_2s}\right)\left(\frac{R_1C_1s + 1}{R_1}\right)$$

Noting that

$$\frac{E_o(s)}{E(s)} = -\frac{R_4}{R_3}$$

we have

$$\begin{aligned}
\frac{E_o(s)}{E_i(s)} = \frac{E_o(s)}{E(s)}\frac{E(s)}{E_i(s)} &= \frac{R_4R_2}{R_3R_1}\frac{(R_1C_1s + 1)(R_2C_2s + 1)}{R_2C_2s} \\
&= \frac{R_4R_2}{R_3R_1}\left(\frac{R_1C_1 + R_2C_2}{R_2C_2} + \frac{1}{R_2C_2s} + R_1C_1s\right) \\
&= \frac{R_4(R_1C_1 + R_2C_2)}{R_3R_1C_2}\left[1 + \frac{1}{(R_1C_1 + R_2C_2)s} + \frac{R_1C_1R_2C_2}{R_1C_1 + R_2C_2}s\right] \qquad (5\text{–}30)
\end{aligned}$$

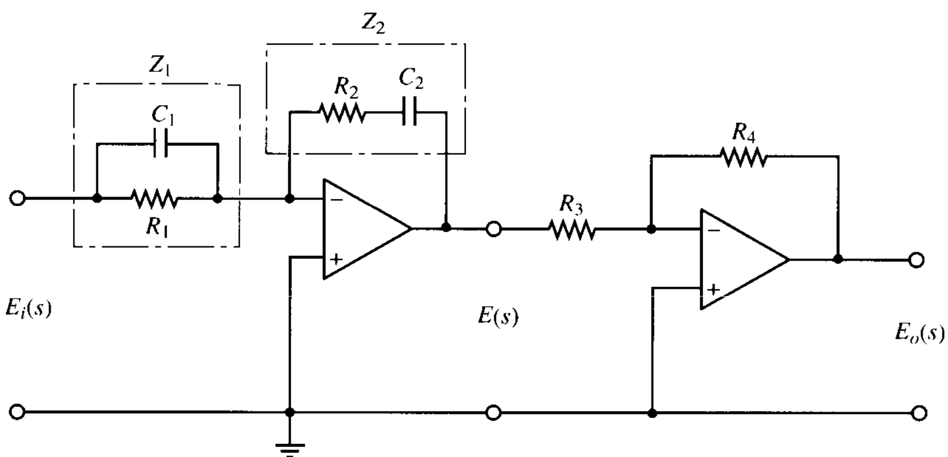



**Figure 5–49** Electronic PID controller.

Thus

$$K_p = \frac{R_4(R_1C_1 + R_2C_2)}{R_3R_1C_2}$$

$$T_i = R_1C_1 + R_2C_2$$

$$T_d = \frac{R_1C_1R_2C_2}{R_1C_1 + R_2C_2}$$

In terms of the proportional gain, integral gain, and derivative gain, we have

$$K_p = \frac{R_4(R_1C_1 + R_2C_2)}{R_3R_1C_2}$$

$$K_i = \frac{R_4}{R_3R_1C_2}$$

$$K_d = \frac{R_4R_2C_1}{R_3}$$

Notice that the second operational-amplifier circuit acts as a sign inverter as well as a gain adjuster.

Table 5–1 shows a list of operational-amplifier circuits that may be used as controllers or compensators.

## 5–9 PHASE LEAD AND PHASE LAG IN SINUSOIDAL RESPONSE

For a sinusoidal input, the steady-state output of a linear time-invariant system is sinusoidal with a phase shift that is a function of the input frequency. This phase angle varies as the frequency is increased from zero to infinity. If the steady-state sinusoidal output of a network leads (lags) the input sinusoid, it is called a lead (lag) network. We shall first derive the steady-state output of a linear, time-invariant network to a sinusoidal input.

**Obtaining steady-state outputs to sinusoidal inputs.** We shall show that the steady-state output of a transfer function system can be obtained directly from the sinusoidal transfer function, that is, the transfer function in which $s$ is replaced by $j\omega$, where $\omega$ is frequency.

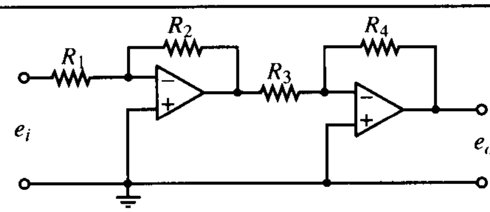
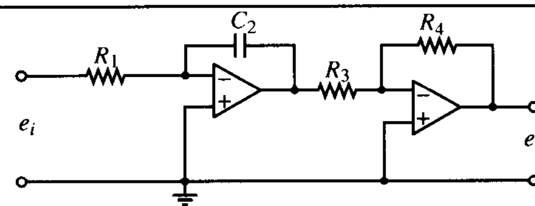
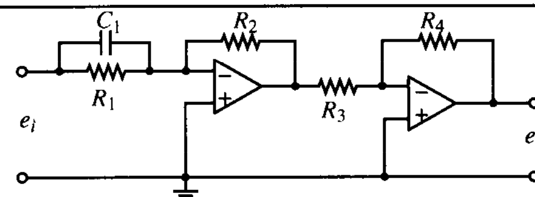
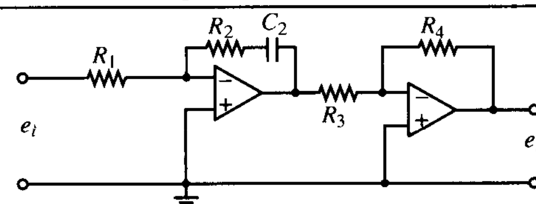
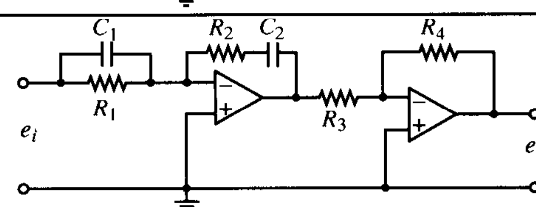
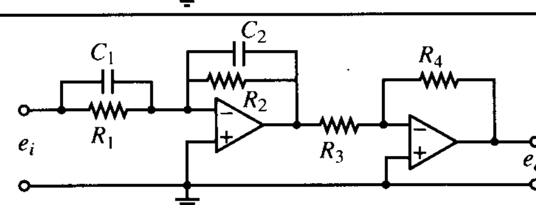
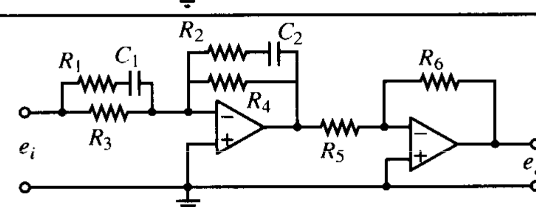
Consider the stable, linear, time-invariant system shown in Figure 5–50. The input and output of the system, whose transfer function is $G(s)$, are denoted by $x(t)$ and $y(t)$, respectively. If the input $x(t)$ is a sinusoidal signal, the steady-state output will also be a sinusoidal signal of the same frequency but with possibly different magnitude and phase angle.

Let us assume that the input signal is given by

$$x(t) = X \sin \omega t$$

Suppose that the transfer function $G(s)$ can be written as a ratio of two polynomials in $s$; that is,

**Table 5–1** Operational-Amplifier Circuits That May Be Used as Compensators

| | Control Action | $G(s) = \frac{E_o(s)}{E_i(s)}$ | Operational Amplifier Circuits |
|---|---|---|---|
| 1 | P | $\frac{R_4}{R_3}\frac{R_2}{R_1}$ |  |
| 2 | I | $\frac{R_4}{R_3}\frac{1}{R_1C_2s}$ |  |
| 3 | PD | $\frac{R_4}{R_3}\frac{R_2}{R_1}(R_1C_1s+1)$ |  |
| 4 | PI | $\frac{R_4}{R_3}\frac{R_2}{R_1}\frac{R_2C_2s+1}{R_2C_2s}$ |  |
| 5 | PID | $\frac{R_4}{R_3}\frac{R_2}{R_1}\frac{(R_1C_1s+1)(R_2C_2s+1)}{R_2C_2s}$ |  |
| 6 | Lead or lag | $\frac{R_4}{R_3}\frac{R_2}{R_1}\frac{R_1C_1s+1}{R_2C_2s+1}$ |  |
| 7 | Lag–lead | $\frac{R_6}{R_5}\frac{R_4}{R_3}\frac{[(R_1+R_3)C_1s+1](R_2C_2s+1)}{(R_1C_1s+1)[(R_2+R_4)C_2s+1]}$ |  |

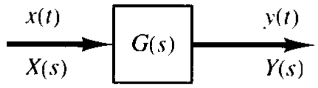



**Figure 5–50**
Stable, linear, time-invariant system.

$$G(s) = \frac{p(s)}{q(s)} = \frac{p(s)}{(s + s_1)(s + s_2)\cdots(s + s_n)}$$

The Laplace-transformed output $Y(s)$ is then

$$Y(s) = G(s)X(s) = \frac{p(s)}{q(s)}X(s) \tag{5–31}$$

where $X(s)$ is the Laplace transform of the input $x(t)$.

It will be shown that after waiting until steady-state conditions are reached the frequency response can be calculated by replacing $s$ in the transfer function by $j\omega$. It will also be shown that the steady-state response can be given by

$$G(j\omega) = Me^{j\phi} = M\underline{/\phi}$$

where $M$ is the amplitude ratio of the output and input sinusoids and $\phi$ is the phase shift between the input sinusoid and the output sinusoid. In the frequency-response test, the input frequency $\omega$ is varied until the entire frequency range of interest is covered.

The steady-state response of a stable, linear, time-invariant system to a sinusoidal input does not depend on the initial conditions. (Thus, we can assume the zero initial condition.) If $Y(s)$ has only distinct poles, then the partial fraction expansion of Equation (5–31) yields

$$\begin{aligned} Y(s) &= G(s)X(s) = G(s)\frac{\omega X}{s^2 + \omega^2} \\ &= \frac{a}{s + j\omega} + \frac{\bar{a}}{s - j\omega} + \frac{b_1}{s + s_1} + \frac{b_2}{s + s_2} + \cdots + \frac{b_n}{s + s_n} \end{aligned} \tag{5–32}$$

where $a$ and the $b_i$ (where $i = 1, 2, \ldots, n$) are constants and $\bar{a}$ is the complex conjugate of $a$. The inverse Laplace transform of Equation (5–32) gives

$$y(t) = ae^{-j\omega t} + \bar{a}e^{j\omega t} + b_1e^{-s_1t} + b_2e^{-s_2t} + \cdots + b_ne^{-s_nt} \qquad (t \geq 0) \tag{5–33}$$

For a stable system, $-s_1, -s_2, \ldots, -s_n$ have negative real parts. Therefore, as $t$ approaches infinity, the terms $e^{-s_1t}, e^{-s_2t}, \ldots,$ and $e^{-s_nt}$ approach zero. Thus, all the terms on the right-hand side of Equation (5–33), except the first two, drop out at steady state.

If $Y(s)$ involves multiple poles $s_j$ of multiplicity $m_j$, then $y(t)$ will involve terms such as $t^{h_j}e^{-s_jt}$ ($h_j = 0, 1, 2, \ldots, m_j - 1$). For a stable system, the terms $t^{h_j}e^{-s_jt}$ approach zero as $t$ approaches infinity.

Thus, regardless of whether the system is of the distinct-pole type, the steady-state response becomes

$$y_{ss}(t) = ae^{-j\omega t} + \bar{a}e^{j\omega t} \tag{5–34}$$

where the constant $a$ can be evaluated from Equation (5–32) as follows:

$$a = G(s)\frac{\omega X}{s^2 + \omega^2}(s + j\omega)\bigg|_{s=-j\omega} = -\frac{XG(-j\omega)}{2j}$$

Note that

$$\bar{a} = G(s)\frac{\omega X}{s^2 + \omega^2}(s - j\omega)\bigg|_{s=j\omega} = \frac{XG(j\omega)}{2j}$$

Since $G(j\omega)$ is a complex quantity, it can be written in the following form:

$$G(j\omega) = |G(j\omega)|e^{j\phi}$$

where $|(Gj\omega)|$ represents the magnitude and $\phi$ represents the angle of $G(j\omega)$; that is,

$$\phi = \underline{/G(j\omega)} = \tan^{-1}\left[\frac{\text{imaginary part of } G(j\omega)}{\text{real part of } G(j\omega)}\right]$$

The angle $\phi$ may be negative, positive, or zero. Similarly, we obtain the following expression for $G(-j\omega)$:

$$G(-j\omega) = |G(-j\omega)|e^{-j\phi} = |G(j\omega)|e^{-j\phi}$$

Then, noting that

$$a = -\frac{X|G(j\omega)|e^{-j\phi}}{2j}, \qquad \bar{a} = \frac{X|G(j\omega)|e^{j\phi}}{2j}$$

Equation (5–34) can be written

$$\begin{aligned} y_{ss}(t) &= X|G(j\omega)|\frac{e^{j(\omega t + \phi)} - e^{-j(\omega t + \phi)}}{2j} \\ &= X|G(j\omega)|\sin(\omega t + \phi) \\ &= Y\sin(\omega t + \phi) \end{aligned} \tag{5–35}$$

where $Y = X|G(j\omega)|$. We see that a stable, linear, time-invariant system subjected to a sinusoidal input will, at steady state, have a sinusoidal output of the same frequency as the input. But the amplitude and phase of the output will, in general, be different from those of the input. In fact, the amplitude of the output is given by the product of that of the input and $|G(j\omega)|$, while the phase angle differs from that of the input by the amount $\phi = \underline{/G(j\omega)}$. An example of input and output sinusoidal signals is shown in Figure 5–51.

On the basis of this, we obtain this important result: For sinusoidal inputs,

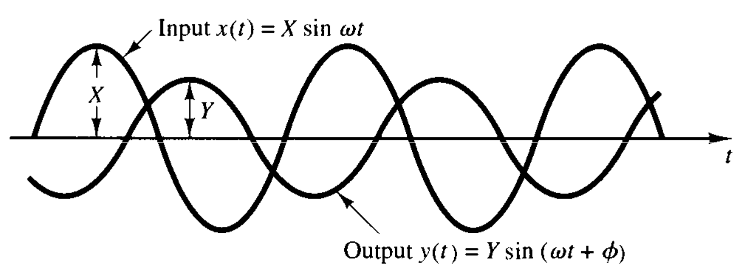



**Figure 5–51** Input and output sinusoidal signals.

$$|G(j\omega)| = \left|\frac{Y(j\omega)}{X(j\omega)}\right| = \text{amplitude ratio of the output sinuisoid to the input sinusoid}$$

$$\angle G(j\omega) = \angle \frac{Y(j\omega)}{X(j\omega)} = \text{phase shift of the output sinusoid with respect to the input sinusoid}$$

Hence, the response characteristics of a system to a sinusoidal input can be obtained directly from

$$\frac{Y(j\omega)}{X(j\omega)} = G(j\omega)$$

The function $G(j\omega)$ is called the *sinusoidal transfer function*. It is the ratio of $Y(j\omega)$ to $X(j\omega)$, is a complex quantity, and can be represented by the magnitude and phase angle with frequency as a parameter. (A negative phase angle is called *phase lag*, and a positive phase angle is called *phase lead*.) The sinusoidal transfer function of any linear system is obtained by substituting $j\omega$ for $s$ in the transfer function of the system.

A network that has phase-lead characteristics is commonly called a lead network. Similarly, a network that has phase-lag characteristics is called a lag network.

**EXAMPLE 5–5** Consider the system shown in Figure 5–52. The transfer function $G(s)$ is

$$G(s) = \frac{K}{Ts + 1}$$

For the sinusoidal input $x(t) = X \sin \omega t$, the steady-state output $y_{ss}(t)$ can be found as follows: Substituting $j\omega$ for $s$ in $G(s)$ yields

$$G(j\omega) = \frac{K}{jT\omega + 1}$$

The amplitude ratio of the output to input is

$$|G(j\omega)| = \frac{K}{\sqrt{1 + T^2\omega^2}}$$

while the phase angle $\phi$ is

$$\phi = \angle G(j\omega) = -\tan^{-1} T\omega$$

Thus, for the input $x(t) = X \sin \omega t$, the steady-state output $y_{ss}(t)$ can be obtained from Equation (5–35) as follows:

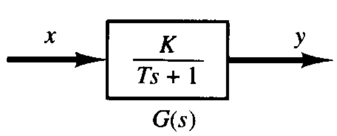



**Figure 5–52**
First-order system.

$$y_{ss}(t) = \frac{XK}{\sqrt{1 + T^2\omega^2}} \sin (\omega t - \tan^{-1} T\omega) \tag{5-36}$$

From Equation (5–36) it can be seen that for small $\omega$ the amplitude of the steady-state output $y_{ss}(t)$ is almost equal to $K$ times the amplitude of the input. The phase shift of the output is small for small $\omega$. For large $\omega$, the amplitude of the output is small and almost inversely proportional to $\omega$. The phase shift approaches $-90°$ as $\omega$ approaches infinity. This is a phase-lag network.

**EXAMPLE 5–6** Consider the network given by

$$G(s) = \frac{s + \frac{1}{T_1}}{s + \frac{1}{T_2}}$$

Find if this network is a lead network or lag network.

For the sinusoidal input $x(t) = X \sin \omega t$, the steady-state output $y_{ss}(t)$ can be found as follows: Since

$$G(j\omega) = \frac{j\omega + \frac{1}{T_1}}{j\omega + \frac{1}{T_2}} = \frac{T_2(1 + T_1 j\omega)}{T_1(1 + T_2 j\omega)}$$

we have

$$|G(j\omega)| = \frac{T_2\sqrt{1 + T_1^2\omega^2}}{T_1\sqrt{1 + T_2^2\omega^2}}$$

and

$$\phi = \underline{/G(j\omega)} = \tan^{-1} T_1\omega - \tan^{-1} T_2\omega$$

Thus the steady-state output is

$$y_{ss}(t) = \frac{XT_2\sqrt{1 + T_1^2\omega^2}}{T_1\sqrt{1 + T_2^2\omega^2}} \sin (\omega t + \tan^{-1} T_1\omega - \tan^{-1} T_2\omega)$$

From this expression, we find that if $T_1 > T_2$ then $\tan^{-1} T_1\omega - \tan^{-1} T_2\omega > 0$. Thus, if $T_1 > T_2$, then the network is a lead network. If $T_1 < T_2$, then the network is a lag network.

## 5–10 STEADY-STATE ERRORS IN UNITY-FEEDBACK CONTROL SYSTEMS

Errors in a control system can be attributed to many factors. Changes in the reference input will cause unavoidable errors during transient periods and may also cause steady-state errors. Imperfections in the system components, such as static friction, backlash, and amplifier drift, as well as aging or deterioration, will cause errors at steady state. In this section, however, we shall not discuss errors due to imperfections in the system components. Rather, we shall investigate a type of steady-state error that is caused by the incapability of a system to follow particular types of inputs.

Any physical control system inherently suffers steady-state error in response to certain types of inputs. A system may have no steady-state error to a step input, but the same system may exhibit nonzero steady-state error to a ramp input. (The only way we may be able to eliminate this error is to modify the system structure.) Whether a given system will exhibit steady-state error for a given type of input depends on the type of open-loop transfer function of the system, to be discussed in what follows.

**Classification of control systems.** Control systems may be classified according to their ability to follow step inputs, ramp inputs, parabolic inputs, and so on. This is a reasonable classification scheme because actual inputs may frequently be considered combinations of such inputs. The magnitudes of the steady-state errors due to these individual inputs are indicative of the goodness of the system.

Consider the unity-feedback control system with the following open-loop transfer function $G(s)$:

$$G(s) = \frac{K(T_a s + 1)(T_b s + 1)\cdots(T_m s + 1)}{s^N(T_1 s + 1)(T_2 s + 1)\cdots(T_p s + 1)}$$

It involves the term $s^N$ in the denominator, representing a pole of multiplicity $N$ at the origin. The present classification scheme is based on the number of integrations indicated by the open-loop transfer function. A system is called type 0, type 1, type 2, ..., if $N = 0, N = 1, N = 2, \ldots$, respectively. Note that this classification is different from that of the order of a system. As the type number is increased, accuracy is improved; however, increasing the type number aggravates the stability problem. A compromise between steady-state accuracy and relative stability is always necessary. In practice, it is rather exceptional to have type 3 or higher systems because we find it generally difficult to design stable systems having more than two integrations in the feedforward path.

We shall see later that, if $G(s)$ is written so that each term in the numerator and denominator, except the term $s^N$, approaches unity as $s$ approaches zero, then the open-loop gain $K$ is directly related to the steady-state error.

**Steady-state errors.** Consider the system shown in Figure 5–53. The closed-loop transfer function is

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

The transfer function between the error signal $e(t)$ and the input signal $r(t)$ is

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

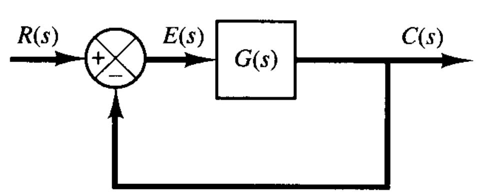



**Figure 5–53**
Control system.

where the error $e(t)$ is the difference between the input signal and the output signal.

The final-value theorem provides a convenient way to find the steady-state performance of a stable system. Since $E(s)$ is

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

the steady-state error is

$$e_{ss} = \lim_{t\to\infty} e(t) = \lim_{s\to 0} sE(s) = \lim_{s\to 0} \frac{sR(s)}{1 + G(s)}$$

The static error constants defined in the following are figures of merit of control systems. The higher the constants, the smaller the steady-state error. In a given system, the output may be the position, velocity, pressure, temperature, or the like. The physical form of the output, however, is immaterial to the present analysis. Therefore, in what follows, we shall call the output "position," the rate of change of the output "velocity," and so on. This means that in a temperature control system "position" represents the output temperature, "velocity" represents the rate of change of the output temperature, and so on.

**Static position error constant $K_p$.** The steady-state error of the system for a unit-step input is

$$\begin{aligned} e_{ss} &= \lim_{s\to 0} \frac{s}{1 + G(s)} \frac{1}{s} \\ &= \frac{1}{1 + G(0)} \end{aligned}$$

The static position error constant $K_p$ is defined by

$$K_p = \lim_{s\to 0} G(s) = G(0)$$

Thus, the steady-state error in terms of the static position error constant $K_p$ is given by

$$e_{ss} = \frac{1}{1 + K_p}$$

For a type 0 system,

$$K_p = \lim_{s\to 0} \frac{K(T_a s + 1)(T_b s + 1)\cdots}{(T_1 s + 1)(T_2 s + 1)\cdots} = K$$

For a type 1 or higher system,

$$K_p = \lim_{s\to 0} \frac{K(T_a s + 1)(T_b s + 1)\cdots}{s^N(T_1 s + 1)(T_2 s + 1)\cdots} = \infty, \qquad \text{for } N \geq 1$$

Hence, for a type 0 system, the static position error constant $K_p$ is finite, while for a type 1 or higher system, $K_p$ is infinite.

For a unit-step input, the steady-state error $e_{ss}$ may be summarized as follows:

$$e_{ss} = \frac{1}{1 + K}, \qquad \text{for type 0 systems}$$

$$e_{ss} = 0, \qquad \text{for type 1 or higher systems}$$

From the foregoing analysis, it is seen that the response of a feedback control system to a step input involves a steady-state error if there is no integration in the feedforward path. (If small errors for step inputs can be tolerated, then a type 0 system may be permissible, provided that the gain $K$ is sufficiently large. If the gain $K$ is too large, however, it is difficult to obtain reasonable relative stability.) If zero steady-state error for a step input is desired, the type of the system must be one or higher.

**Static velocity error constant $K_v$.** The steady-state error of the system with a unit-ramp input is given by

$$e_{ss} = \lim_{s\to 0} \frac{s}{1 + G(s)} \frac{1}{s^2}$$

$$= \lim_{s\to 0} \frac{1}{sG(s)}$$

The static velocity error constant $K_v$ is defined by

$$K_v = \lim_{s\to 0} sG(s)$$

Thus, the steady-state error in terms of the static velocity error constant $K_v$ is given by

$$e_{ss} = \frac{1}{K_v}$$

The term *velocity error* is used here to express the steady-state error for a ramp input. The dimension of the velocity error is the same as the system error. That is, velocity error is not an error in velocity, but it is an error in position due to a ramp input.

For a type 0 system,

$$K_v = \lim_{s\to 0} \frac{sK(T_a s + 1)(T_b s + 1)\cdots}{(T_1 s + 1)(T_2 s + 1)\cdots} = 0$$

For a type 1 system,

$$K_v = \lim_{s\to 0} \frac{sK(T_a s + 1)(T_b s + 1)\cdots}{s(T_1 s + 1)(T_2 s + 1)\cdots} = K$$

For a type 2 or higher system,

$$K_v = \lim_{s\to 0} \frac{sK(T_a s + 1)(T_b s + 1)\cdots}{s^N(T_1 s + 1)(T_2 s + 1)\cdots} = \infty, \qquad \text{for } N \geq 2$$

The steady-state error $e_{ss}$ for the unit-ramp input can be summarized as follows:

$$e_{ss} = \frac{1}{K_v} = \infty, \qquad \text{for type 0 systems}$$

$$e_{ss} = \frac{1}{K_v} = \frac{1}{K}, \qquad \text{for type 1 systems}$$

$$e_{ss} = \frac{1}{K_v} = 0, \qquad \text{for type 2 or higher systems}$$

The foregoing analysis indicates that a type 0 system is incapable of following a ramp input in the steady state. The type 1 system with unity feedback can follow the ramp input with a finite error. In steady-state operation, the output velocity is exactly the same as the input velocity, but there is a positional error. This error is proportional to the velocity of the input and is inversely proportional to the gain $K$. Figure 5–54 shows an example of the response of a type 1 system with unity feedback to a ramp input. The type 2 or higher system can follow a ramp input with zero error at steady state.

**Static acceleration error constant $K_a$.** The steady-state error of the system with a unit-parabolic input (acceleration input), which is defined by

$$\begin{aligned} r(t) &= \frac{t^2}{2}, \qquad &\text{for } t \geq 0 \\ &= 0, \qquad &\text{for } t < 0 \end{aligned}$$

is given by

$$\begin{aligned} e_{ss} &= \lim_{s \to 0} \frac{s}{1 + G(s)} \frac{1}{s^3} \\ &= \frac{1}{\lim_{s \to 0} s^2 G(s)} \end{aligned}$$

The static acceleration error constant $K_a$ is defined by the equation

$$K_a = \lim_{s \to 0} s^2 G(s)$$

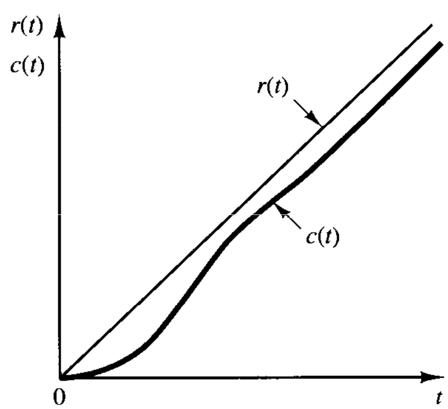



**Figure 5–54**
Response of a type 1 unity-feedback system to a ramp input.

The steady-state error is then

$$e_{ss} = \frac{1}{K_a}$$

Note that the acceleration error, the steady-state error due to a parabolic input, is an error in position.

The values of $K_a$ are obtained as follows:

For a type 0 system,

$$K_a = \lim_{s\to 0} \frac{s^2K(T_as + 1)(T_bs + 1)\cdots}{(T_1s + 1)(T_2s + 1)\cdots} = 0$$

For a type 1 system,

$$K_a = \lim_{s\to 0} \frac{s^2K(T_as + 1)(T_bs + 1)\cdots}{s(T_1s + 1)(T_2s + 1)\cdots} = 0$$

For a type 2 system,

$$K_a = \lim_{s\to 0} \frac{s^2K(T_as + 1)(T_bs + 1)\cdots}{s^2(T_1s + 1)(T_2s + 1)\cdots} = K$$

For a type 3 or higher system,

$$K_a = \lim_{s\to 0} \frac{s^2K(T_as + 1)(T_bs + 1)\cdots}{s^N(T_1s + 1)(T_2s + 1)\cdots} = \infty, \qquad \text{for } N \geq 3$$

Thus, the steady-state error for the unit parabolic input is

$$e_{ss} = \infty, \qquad \text{for type 0 and type 1 systems}$$

$$e_{ss} = \frac{1}{K}, \qquad \text{for type 2 systems}$$

$$e_{ss} = 0, \qquad \text{for type 3 or higher systems}$$

Note that both type 0 and type 1 systems are incapable of following a parabolic input in the steady state. The type 2 system with unity feedback can follow a parabolic input with a finite error signal. Figure 5–55 shows an example of the response of a type 2 system with unity feedback to a parabolic input. The type 3 or higher system with unity feedback follows a parabolic input with zero error at steady state.

**Summary.** Table 5–2 summarizes the steady-state errors for type 0, type 1, and type 2 systems when they are subjected to various inputs. The finite values for steady-state errors appear on the diagonal line. Above the diagonal, the steady-state errors are infinity; below the diagonal, they are zero.

Remember that the terms *position error*, *velocity error*, and *acceleration error* mean steady-state deviations in the output position. A finite velocity error implies that after transients have died out the input and output move at the same velocity but have a finite position difference.

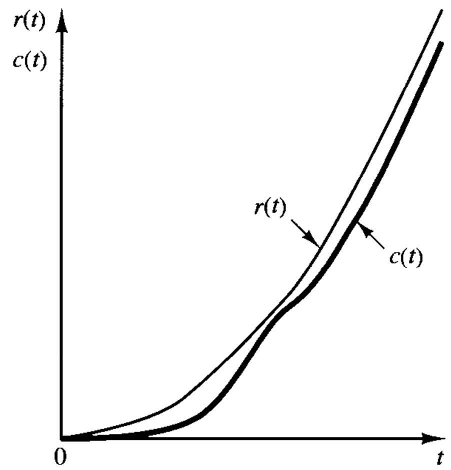



**Figure 5–55**
Response of a type 2 unity-feedback system to a parabolic input.

The error constants $K_p$, $K_v$, and $K_a$ describe the ability of a unity-feedback system to reduce or eliminate steady-state error. Therefore, they are indicative of the steady-state performance. It is generally desirable to increase the error constants, while maintaining the transient response within an acceptable range. If there is any conflict between the static velocity error constant and the static acceleration error constant, then the latter may be considered less important than the former. It is noted that to improve the steady-state performance we can increase the type of the system by adding an integrator or integrators to the feedforward path. This, however, introduces an additional stability problem. The design of a satisfactory system with more than two integrators in series in the feedforward path is generally difficult.

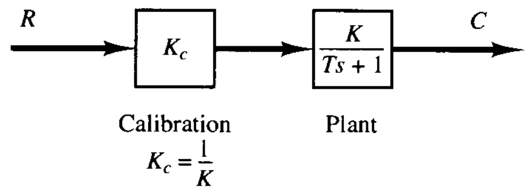
**Comparison of steady-state errors in open-loop control system and closed-loop control system.** Consider the open-loop control system and closed-loop control system shown in Figure 5–56. In the open loop one, gain $K_c$ is calibrated so that $K_c = 1/K$. Thus, the transfer function of the open-loop control system is

$$G_0(s) = \frac{1}{K}\frac{K}{Ts+1} = \frac{1}{Ts+1}$$

In the closed-loop control system, gain $K_p$ of the controller is set so that $K_pK \gg 1$.

**Table 5–2** Steady-State Error in Terms of Gain $K$

| | Step Input $r(t) = 1$ | Ramp Input $r(t) = t$ | Acceleration Input $r(t) = \frac{1}{2}t^2$ |
|---|---|---|---|
| Type 0 system | $\frac{1}{1+K}$ | $\infty$ | $\infty$ |
| Type 1 system | 0 | $\frac{1}{K}$ | $\infty$ |
| Type 2 system | 0 | 0 | $\frac{1}{K}$ |



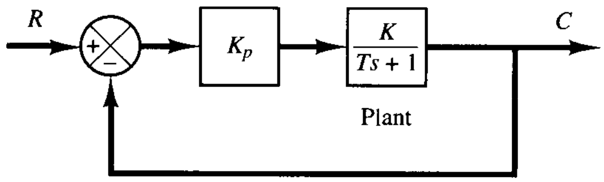




**Figure 5–56**
Block diagrams of an open-loop control system and a closed-loop control system.

Assuming a unit-step input, let us compare the steady-state errors for these control systems. For the open-loop control system, the error signal is

$$e(t) = r(t) - c(t)$$

or

$$\begin{aligned} E(s) &= R(s) - C(s) \\ &= [1 - G_0(s)]R(s) \end{aligned}$$

The steady-state error in the unit-step response is

$$\begin{aligned} e_{ss} &= \lim_{s \to 0} sE(s) \\ &= \lim_{s \to 0} s[1 - G_0(s)]\frac{1}{s} \\ &= 1 - G_0(0) \end{aligned}$$

If $G_o(0)$, the dc gain of the open-loop control system, is equal to unity, then the steady-state error is zero. Due to environmental changes and aging of components, however, the dc gain $G_o(0)$ will drift from unity as time elapses, and the steady-state error will no longer be equal to zero. Such steady-state error in an open-loop control system will remain until the system is recalibrated.

For the closed-loop control system, the error signal is

$$\begin{aligned} E(s) &= R(s) - C(s) \\ &= \frac{1}{1 + G(s)}R(s) \end{aligned}$$

where

$$G(s) = \frac{K_p K}{Ts + 1}$$

The steady-state error in the unit-step response is

$$
\begin{aligned}
e_{ss} &= \lim_{s\to 0} s\left[\frac{1}{1+G(s)}\right]\frac{1}{s} \\
&= \frac{1}{1+G(0)} \\
&= \frac{1}{1+K_pK}
\end{aligned}
$$

In the closed-loop control system, gain $K_p$ is set at a large value compared with $1/K$. Thus the steady-state error can be made small, although not exactly zero.

Let us assume the following variation in the transfer function of the plant, assuming $K_c$ and $K_p$ constant:

$$
\frac{K+\Delta K}{Ts+1}
$$

For simplicity, let us assume that $K = 10, \Delta K = 1$, or $\Delta K/K = 0.1$. Then the steady-state error in the unit-step response for the open-loop control system becomes

$$
\begin{aligned}
e_{ss} &= 1 - \frac{1}{K}(K+\Delta K) \\
&= 1 - 1.1 = -0.1
\end{aligned}
$$

For the closed-loop control system, if $K_p$ is set at $100/K$, then the steady-state error in the unit-step response becomes

$$
\begin{aligned}
e_{ss} &= \frac{1}{1+G(0)} \\
&= \frac{1}{1+\dfrac{100}{K}(K+\Delta K)} \\
&= \frac{1}{1+110} = 0.009
\end{aligned}
$$

Thus, the closed-loop control system is superior to the open-loop control system in the presence of environmental changes, aging of components, and the like, which definitely affect the steady-state performance.

## EXAMPLE PROBLEMS AND SOLUTIONS

**A-5-1.** Explain why the proportional control of a plant that does not possess an integrating property (which means that the plant transfer function does not include the factor $1/s$) suffers offset in response to step inputs.

**Solution.** Consider, for example, the system shown in Figure 5–57. At steady state, if $c$ were equal to a nonzero constant $r$, then $e = 0$ and $u = Ke = 0$, resulting in $c = 0$, which contradicts the assumption that $c = r =$ nonzero constant.

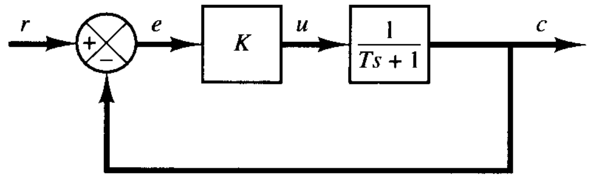



**Figure 5–57**
Control system.

A nonzero offset must exist for proper operation of such a control system. In other words, at steady state, if $e$ were equal to $r/(1 + K)$, then $u = Kr/(1 + K)$ and $c = Kr/(1 + K)$, which results in the assumed error signal $e = r/(1 + K)$. Thus the offset of $r/(1 + K)$ must exist in such a system.

**A-5-2.** Consider the system shown in Figure 5–58. Show that the steady-state error in following the unit-ramp input is $B/K$. This error can be made smaller by choosing $B$ small and/or $K$ large. However, making $B$ small and/or $K$ large would have the effect of making the damping ratio small, which is normally not desirable. Describe a method or methods to make $B/K$ small and yet make the damping ratio have reasonable value ($0.5 < \zeta < 0.7$).

**Solution.** From Figure 5–58 we obtain

$$E(s) = R(s) - C(s) = \frac{Js^2 + Bs}{Js^2 + Bs + K} R(s)$$

The steady-state error for the unit-ramp response can be obtained as follows: For the unit-ramp input, the steady-state error $e_{ss}$ is

$$\begin{aligned} e_{ss} &= \lim_{s \to 0} sE(s) \\ &= \lim_{s \to 0} s \frac{Js^2 + Bs}{Js^2 + Bs + K} \frac{1}{s^2} \\ &= \frac{B}{K} \\ &= \frac{2\zeta}{\omega_n} \end{aligned}$$

where

$$\zeta = \frac{B}{2\sqrt{KJ}}, \qquad \omega_n = \sqrt{\frac{K}{J}}$$

To assure acceptable transient response and acceptable steady-state error in following a ramp input, $\zeta$ must not be too small and $\omega_n$ must be sufficiently large. It is possible to make the steady-state error $e_{ss}$ small by making the value of the gain $K$ large. (A large value of $K$ has an additional advantage of suppressing undesirable effects caused by dead zone, backlash, coulomb friction,

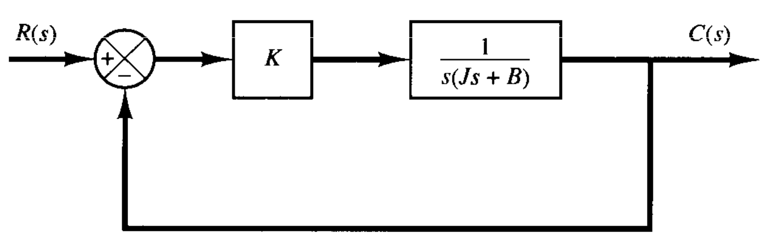



**Figure 5–58**
Control system.

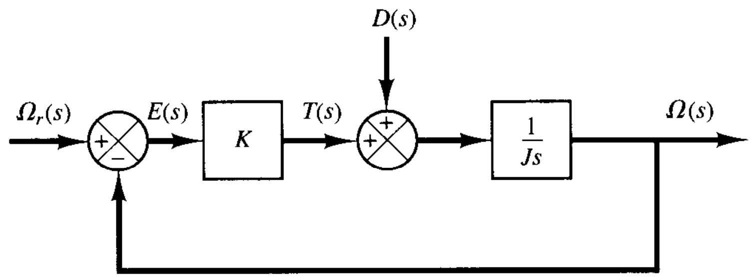



**Figure 5–59**
Block diagram of a speed control system.

and the like.) A large value of $K$ would, however, make the value of $\zeta$ small and increase the maximum overshoot, which is undesirable.

It is therefore necessary to compromise between the magnitude of the steady-state error to a ramp input and the maximum overshoot to a unit-step input. In the system shown in Figure 5–58, a reasonable compromise may not be reached easily. It is then desirable to consider other types of control action that may improve both the transient-response and steady-state behavior. Two schemes to improve both the transient-response and steady-state behavior are available. One scheme is to use a proportional-plus-derivative controller and the other is to use tachometer feedback.

**A-5-3.** The block diagram of Figure 5–59 shows a speed control system in which the output member of the system is subject to a torque disturbance. In the diagram, $\Omega_r(s)$, $\Omega(s)$, $T(s)$, and $D(s)$ are the Laplace transforms of the reference speed, output speed, driving torque, and disturbance torque, respectively. In the absence of a disturbance torque, the output speed is equal to the reference speed.

Investigate the response of this system to a unit-step disturbance torque. Assume that the reference input is zero, or $\Omega_r(s) = 0$.

**Solution.** Figure 5–60 is a modified block diagram convenient for the present analysis. The closed-loop transfer function is

$$\frac{\Omega_D(s)}{D(s)} = \frac{1}{Js + K}$$

where $\Omega_D(s)$ is the Laplace transform of the output speed due to the disturbance torque. For a unit-step disturbance torque, the steady-state output velocity is

$$\begin{aligned}\omega_D(\infty) &= \lim_{s \to 0} s\Omega_D(s) \\ &= \lim_{s \to 0} \frac{s}{Js + K} \frac{1}{s} \\ &= \frac{1}{K}\end{aligned}$$

From this analysis, we conclude that, if a step disturbance torque is applied to the output member of the system, an error speed will result so that the ensuing motor torque will exactly

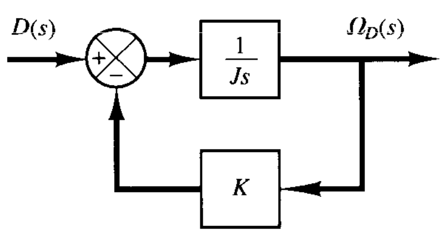



**Figure 5–60**
Block diagram of the speed control system of Figure 5–59 when $\Omega_r(s) = 0$.

cancel the disturbance torque. To develop this motor torque, it is necessary that there be an error in speed so that nonzero torque will result.

**A-5-4.** In the system considered in Problem A-5-3, it is desired to eliminate as much as possible the speed errors due to torque disturbances.

Is it possible to cancel the effect of a disturbance torque at steady state so that a constant disturbance torque applied to the output member will cause no speed change at steady state?

**Solution.** Suppose that we choose a suitable controller whose transfer function is $G_c(s)$, as shown in Figure 5–61. Then in the absence of the reference input the closed-loop transfer function between the output velocity $\Omega_D(s)$ and the disturbance torque $D(s)$ is

$$\frac{\Omega_D(s)}{D(s)} = \frac{\frac{1}{Js}}{1 + \frac{1}{Js}G_c(s)}$$

$$= \frac{1}{Js + G_c(s)}$$

The steady-state output speed due to a unit-step disturbance torque is

$$\omega_D(\infty) = \lim_{s \to 0} s\Omega_D(s)$$

$$= \lim_{s \to 0} \frac{s}{Js + G_c(s)} \frac{1}{s}$$

$$= \frac{1}{G_c(0)}$$

To satisfy the requirement that

$$\omega_D(\infty) = 0$$

we must choose $G_c(0) = \infty$. This can be realized if we choose

$$G_c(s) = \frac{K}{s}$$

Integral control action will continue to correct until the error is zero. This controller, however, presents a stability problem because the characteristic equation will have two imaginary roots.

One method of stabilizing such a system is to add a proportional mode to the controller or choose

$$G_c(s) = K_p + \frac{K}{s}$$

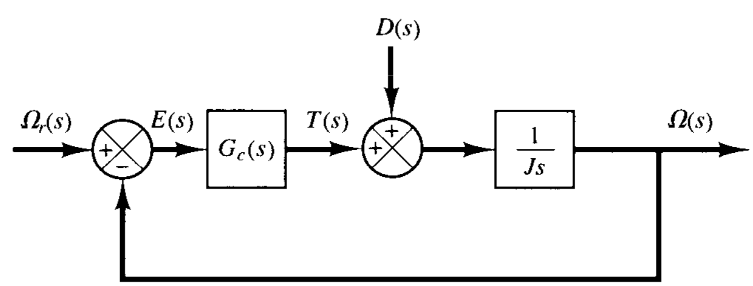



**Figure 5–61** Block diagram of a speed control system.

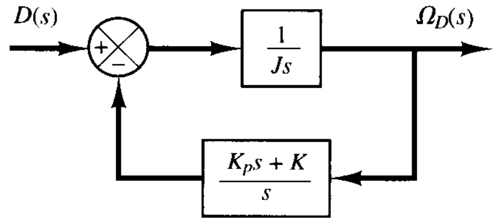


**Figure 5–62**
Block diagram of the speed control system of Figure 5–61 when $G_c(s) = K_p + (K/s)$ and $\Omega_r(s) = 0$.

With this controller, the block diagram of Figure 5–61 in the absence of the reference input can be modified to that of Figure 5–62. The closed-loop transfer function $\Omega_D(s)/D(s)$ becomes

$$\frac{\Omega_D(s)}{D(s)} = \frac{s}{Js^2 + K_p s + K}$$

For a unit-step disturbance torque, the steady-state output speed is

$$\omega_D(\infty) = \lim_{s\to 0} s\Omega_D(s) = \lim_{s\to 0} \frac{s^2}{Js^2 + K_p s + K}\frac{1}{s} = 0$$

Thus, we see that the proportional-plus-integral controller eliminates speed error at steady state.

The use of integral control action has increased the order of the system by 1. (This tends to produce an oscillatory response.)

In the present system, a step disturbance torque will cause a transient error in the output speed, but the error will become zero at steady state. The integrator provides a nonzero output with zero error. (The nonzero output of the integrator produces a motor torque that exactly cancels the disturbance torque.)

Note that the integrator in the transfer function of the plant does not eliminate the steady-state error due to a step disturbance torque. To eliminate this, we must have an integrator before the point where the disturbance torque enters.

**A-5-5.** Consider the system shown in Figure 5–63(a). The steady-state error to a unit-ramp input is $e_{ss} = 2\zeta/\omega_n$. Show that the steady-state error for following a ramp input may be eliminated if the input is introduced to the system through a proportional-plus-derivative filter, as shown in Figure 5–63(b), and the value of $k$ is properly set. Note that the error $e(t)$ is given by $r(t) - c(t)$.

**Solution.** The closed-loop transfer function of the system shown in Figure 5–63(b) is

$$\frac{C(s)}{R(s)} = \frac{(1 + ks)\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

Then

$$R(s) - C(s) = \left(\frac{s^2 + 2\zeta\omega_n s - \omega_n^2 ks}{s^2 + 2\zeta\omega_n s + \omega_n^2}\right) R(s)$$

If the input is a unit ramp, then the steady-state error is

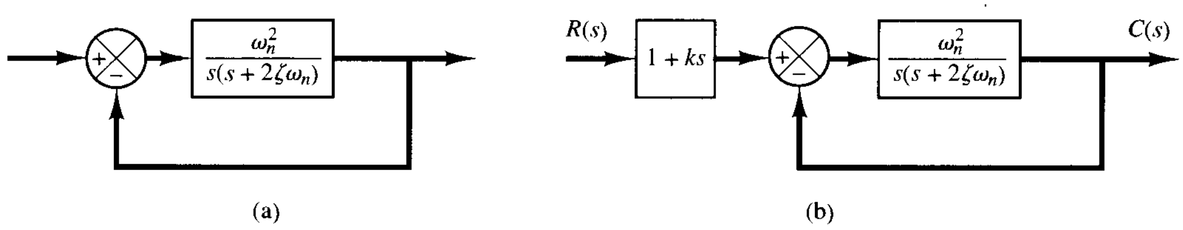


**Figure 5–63**
(a) Control system;
(b) control system with input filter.

$$e(\infty) = r(\infty) - c(\infty)$$

$$= \lim_{s\to 0} s\left(\frac{s^2 + 2\zeta\omega_n s - \omega_n^2 ks}{s^2 + 2\zeta\omega_n s + \omega_n^2}\right)\frac{1}{s^2}$$

$$= \frac{2\zeta\omega_n - \omega_n^2 k}{\omega_n^2}$$

Therefore, if $k$ is chosen as

$$k = \frac{2\zeta}{\omega_n}$$

then the steady-state error for following a ramp input can be made equal to zero. Note that, if there are any variations in the values of $\zeta$ and/or $\omega_n$ due to environmental changes or aging, then a nonzero steady-state error for a ramp response may result.

**A-5-6.** Consider the liquid-level control system shown in Figure 5–64. Assume that the set point of the controller is fixed. Assuming a step disturbance of magnitude $D_0$, determine the error. Assume that $D_0$ is small and the variations in the variables from their respective steady-state values are also small. The controller is proportional.

If the controller is not proportional, but integral, what is the steady-state error?

**Solution.** Figure 5–65 is a block diagram of the system when the controller is proportional with gain $K_p$. (We assume the transfer function of the pneumatic valve to be unity.) Since the set point is fixed, the variation in the set point is zero, or $X(s) = 0$. The Laplace transform of $h(t)$ is

$$H(s) = \frac{K_p R}{RCs + 1}E(s) + \frac{R}{RCs + 1}D(s)$$

Then

$$E(s) = -H(s) = -\frac{K_p R}{RCs + 1}E(s) - \frac{R}{RCs + 1}D(s)$$

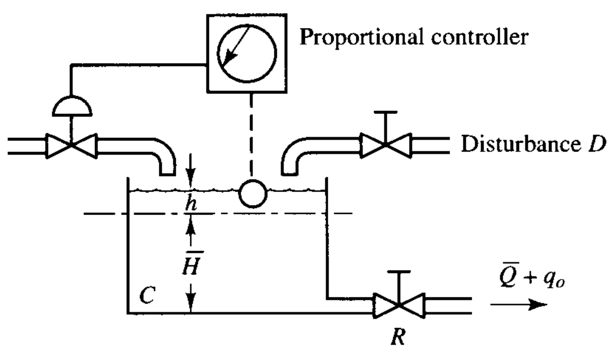



**Figure 5–64**
Liquid-level control system.

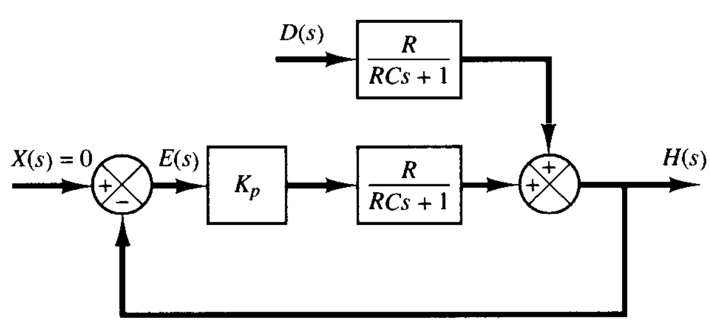



**Figure 5–65**
Block diagram of the liquid-level control system shown in Figure 5–64.

Hence

$$E(s) = -\frac{R}{RCs + 1 + K_pR} D(s)$$

Since

$$D(s) = \frac{D_0}{s}$$

we obtain

$$E(s) = -\frac{R}{RCs + 1 + K_pR} \frac{D_0}{s}$$
$$= \frac{RD_0}{1 + K_pR} \left( \frac{1}{s + \dfrac{1 + K_pR}{RC}} \right) - \frac{RD_0}{1 + K_pR} \frac{1}{s}$$

The time solution for $t > 0$ is

$$e(t) = \frac{RD_0}{1 + K_pR} \left[ \exp\left( -\frac{1 + K_pR}{RC} t \right) - 1 \right]$$

Thus, the time constant is $RC/(1 + K_pR)$. (In the absence of the controller, the time constant is equal to $RC$.) As the gain of the controller is increased, the time constant is decreased. The steady-state error is

$$e(\infty) = -\frac{RD_0}{1 + K_pR}$$

As the gain $K_p$ of the controller is increased, the steady-state error, or offset, is reduced. Thus, mathematically, the larger the gain $K_p$ is, the smaller the offset and time constant are. In practical systems, however, if the gain $K_p$ of the proportional controller is increased to a very large value, oscillation may result in the output since in our analysis all the small lags and small time constants that may exist in the actual control system are neglected. (If these small lags and time constants are included in the analysis, the transfer function becomes higher order, and for very large values of $K_p$ the possibility of oscillation or even instability may occur.)

If the controller is integral, then assuming the transfer function of the controller to be

$$G_c = \frac{K}{s}$$

we obtain

$$E(s) = -\frac{Rs}{RCs^2 + s + KR} D(s)$$

The steady-state error for a step disturbance $D(s) = D_0/(s)$ is

$$e(\infty) = \lim_{s \to 0} sE(s)$$
$$= \lim_{s \to 0} \frac{-Rs^2}{RCs^2 + s + KR} \frac{D_0}{s}$$
$$= 0$$

Thus, an integral controller eliminates steady-state error or offset due to the step disturbance. (The value of $K$ must be chosen so that the transient response due to the command input and/or disturbance damps out with a reasonable speed.)

**A-5-7.** Obtain both analytical and computational solutions of the unit-step response of a unity-feedback system whose open-loop transfer function is

$$G(s) = \frac{5(s + 20)}{s(s + 4.59)(s^2 + 3.41s + 16.35)}$$

**Solution.** The closed-loop transfer function is

$$\begin{aligned} \frac{C(s)}{R(s)} &= \frac{5(s + 20)}{s(s + 4.59)(s^2 + 3.41s + 16.35) + 5(s + 20)} \\ &= \frac{5s + 100}{s^4 + 8s^3 + 32s^2 + 80s + 100} \\ &= \frac{5(s + 20)}{(s^2 + 2s + 10)(s^2 + 6s + 10)} \end{aligned}$$

The unit-step response of this system is then

$$\begin{aligned} C(s) &= \frac{5(s + 20)}{s(s^2 + 2s + 10)(s^2 + 6s + 10)} \\ &= \frac{1}{s} + \frac{\frac{3}{8}(s + 1) - \frac{17}{8}}{(s + 1)^2 + 3^2} + \frac{-\frac{11}{8}(s + 3) - \frac{13}{8}}{(s + 3)^2 + 1^2} \end{aligned}$$

The time response $c(t)$ can be found by taking the inverse Laplace transform of $C(s)$ as follows:

$$c(t) = 1 + \tfrac{3}{8}e^{-t}\cos 3t - \tfrac{17}{24}e^{-t}\sin 3t - \tfrac{11}{8}e^{-3t}\cos t - \tfrac{13}{8}e^{-3t}\sin t, \qquad \text{for } t \geq 0$$

A MATLAB program to obtain the unit-step response of this system is shown in MATLAB Program 5–2. The resulting unit-step response curve is shown in Figure 5–66.

MATLAB Program 5–2

```
% ---------- Unit-step-response ----------

num = [0   0   0   5   100];
den = [1   8   32   80   100];
step(num,den)
grid
title('Unit-Step Response of C(s)/R(s) = (5s + 100)/(s^4 + 8s^3 + 32s^2 + 80s + 100)')
```

**A-5-8.** Consider the following characteristic equation:

$$s^4 + Ks^3 + s^2 + s + 1 = 0$$

Determine the range of $K$ for stability.

**Solution.** The Routh array of coefficients is

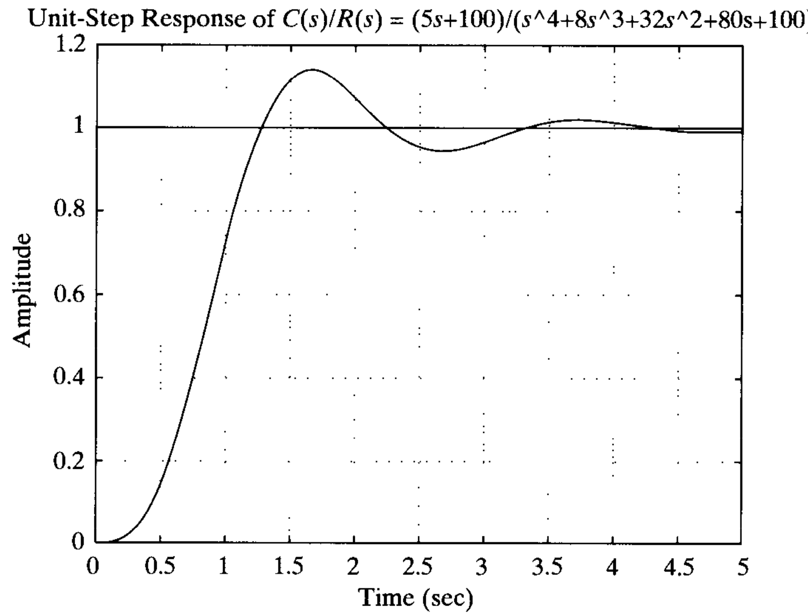



**Figure 5–66**
Unit-step response curve.

$$
\begin{array}{lccc}
s^4 & 1 & 1 & 1 \\
s^3 & K & 1 & 0 \\
s^2 & \dfrac{K-1}{K} & 1 & 0 \\
s^1 & 1 - \dfrac{K^2}{K-1} & 0 & \\
s^0 & 1 & &
\end{array}
$$

For stability, we require that

$$K > 0$$

$$\frac{K-1}{K} > 0$$

$$1 - \frac{K^2}{K-1} > 0$$

From the first and second conditions, $K$ must be greater than 1. For $K > 1$, notice that the term $1 - [K^2/(K-1)]$ is always negative, since

$$\frac{K - 1 - K^2}{K - 1} = \frac{-1 + K(1 - K)}{K - 1} < 0$$

Thus, the three conditions cannot be fulfilled simultaneously. Therefore, there is no value of $K$ that allows stability of the system.

**A-5-9.** Consider the characteristic equation given by

$$a_0 s^n + a_1 s^{n-1} + a_2 s^{n-2} + \cdots + a_{n-1} s + a_n = 0 \tag{5–37}$$

The Hurwitz stability criterion, given next, gives conditions for all the roots to have negative real parts in terms of the coefficients of the polynomial. As stated in the discussions of Routh's stability criterion in Section 5–5, for all the roots to have negative real parts, all the coefficients $a$'s must be positive. This is a necessary condition but not a sufficient condition. If this condition is not satisfied, it indicates that some of the roots have positive real parts or are imaginary or zero. A sufficient condition for all the roots to have negative real parts is given in the following Hurwitz stability criterion: If all the coefficients of the polynomial are positive, arrange these coefficients in the following determinant:

$$\Delta_n = \begin{vmatrix} a_1 & a_3 & a_5 & \cdots & 0 & 0 & 0 \\ a_0 & a_2 & a_4 & \cdots & \cdot & \cdot & \cdot \\ 0 & a_1 & a_3 & \cdots & a_n & 0 & 0 \\ 0 & a_0 & a_2 & \cdots & a_{n-1} & 0 & 0 \\ \cdot & \cdot & \cdot & & a_{n-2} & a_n & 0 \\ \cdot & \cdot & \cdot & & a_{n-3} & a_{n-1} & 0 \\ 0 & 0 & 0 & \cdots & a_{n-4} & a_{n-2} & a_n \end{vmatrix}$$

where we substituted zero for $a_s$ if $s > n$. For all the roots to have negative real parts, it is necessary and sufficient that successive principal minors of $\Delta_n$ be positive. The successive principal minors are the following determinants:

$$\Delta_i = \begin{vmatrix} a_1 & a_3 & \cdots & a_{2i-1} \\ a_0 & a_2 & \cdots & a_{2i-2} \\ 0 & a_1 & \cdots & a_{2i-3} \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & a_i \end{vmatrix} \qquad (i = 1, 2, \ldots, n - 1)$$

where $a_s = 0$ if $s > n$. (It is noted that some of the conditions for the lower-order determinants are included in the conditions for the higher-order determinants.) If all these determinants are positive, and $a_0 > 0$ as already assumed, the equilibrium state of the system whose characteristic equation is given by Equation (5–37) is asymptotically stable. Note that exact values of determinants are not needed; instead, only signs of these determinants are needed for the stability criterion.

Now consider the following characteristic equation:

$$a_0 s^4 + a_1 s^3 + a_2 s^2 + a_3 s + a_4 = 0$$

Obtain the condition for stability using the Hurwitz stability criterion.

**Solution.** The conditions for stability are that all the $a$'s be positive and that

$$\Delta_2 = \begin{vmatrix} a_1 & a_3 \\ a_0 & a_2 \end{vmatrix} = a_1 a_2 - a_0 a_3 > 0$$

$$\begin{aligned} \Delta_3 &= \begin{vmatrix} a_1 & a_3 & 0 \\ a_0 & a_2 & a_4 \\ 0 & a_1 & a_3 \end{vmatrix} \\ &= a_1(a_2 a_3 - a_1 a_4) - a_0 a_3^2 \\ &= a_3(a_1 a_2 - a_0 a_3) - a_1^2 a_4 > 0 \end{aligned}$$

It is clear that, if all the $a$'s are positive and if the condition $\Delta_3 > 0$ is satisfied, the condition $\Delta_2 > 0$ is also satisfied. Therefore, for all the roots of the given characteristic equation to have negative real parts, it is necessary and sufficient that all the coefficients $a$'s are positive and $\Delta_3 > 0$.

**A-5-10.** Show that the Routh's stability criterion and Hurwitz stability criterion are equivalent.

**Solution.** If we write Hurwitz determinants in the triangular form

$$\Delta_i = \begin{vmatrix} a_{11} & & & & * \\ & a_{22} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & a_{ii} \end{vmatrix} \qquad (i = 1, 2, \ldots, n)$$

where the elements below the diagonal line are all zeros and the elements above the diagonal line any numbers, then the Hurwitz conditions for asymptotic stability become

$$\Delta_i = a_{11}a_{22}\cdots a_{ii} > 0 \qquad (i = 1, 2, \ldots, n)$$

which are equivalent to the conditions

$$a_{11} > 0, \qquad a_{22} > 0, \qquad \ldots, \qquad a_{nn} > 0$$

We shall show that these conditions are equivalent to

$$a_1 > 0, \qquad b_1 > 0, \qquad c_1 > 0, \qquad \ldots$$

where $a_1, b_1, c_1, \ldots$ are the elements of the first column in the Routh array.

Consider, for example, the following Hurwitz determinant, which corresponds to $n = 4$:

$$\Delta_4 = \begin{vmatrix} a_1 & a_3 & a_5 & a_7 \\ a_0 & a_2 & a_4 & a_6 \\ 0 & a_1 & a_3 & a_5 \\ 0 & a_0 & a_2 & a_4 \end{vmatrix}$$

The determinant is unchanged if we subtract from the $i$th row $k$ times the $j$th row. By subtracting from the second row $a_0/a_1$ times the first row, we obtain

$$\Delta_4 = \begin{vmatrix} a_{11} & a_3 & a_5 & a_7 \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & a_1 & a_3 & a_5 \\ 0 & a_0 & a_2 & a_4 \end{vmatrix}$$

where

$$a_{11} = a_1$$

$$a_{22} = a_2 - \frac{a_0}{a_1} a_3$$

$$a_{23} = a_4 - \frac{a_0}{a_1} a_5$$

$$a_{24} = a_6 - \frac{a_0}{a_1} a_7$$

Similarly, subtracting from the fourth row $a_0/a_1$ times the third row yields

$$\Delta_4 = \begin{vmatrix} a_{11} & a_3 & a_5 & a_7 \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & a_1 & a_3 & a_5 \\ 0 & 0 & \hat{a}_{43} & \hat{a}_{44} \end{vmatrix}$$

where

$$\hat{a}_{43} = a_2 - \frac{a_0}{a_1} a_3$$

$$\hat{a}_{44} = a_4 - \frac{a_0}{a_1} a_5$$

Next, subtracting from the third row $a_1/a_{22}$ times the second row yields

$$\Delta_4 = \begin{vmatrix} a_{11} & a_3 & a_5 & a_7 \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & \hat{a}_{43} & \hat{a}_{44} \end{vmatrix}$$

where

$$a_{33} = a_3 - \frac{a_1}{a_{22}} a_{23}$$

$$a_{34} = a_5 - \frac{a_1}{a_{22}} a_{24}$$

Finally, subtracting from the last row $\hat{a}_{43}/a_{33}$ times the third row yields

$$\Delta_4 = \begin{vmatrix} a_{11} & a_3 & a_5 & a_7 \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{vmatrix}$$

where

$$a_{44} = \hat{a}_{44} - \frac{\hat{a}_{43}}{a_{33}} a_{34}$$

From this analysis, we see that

$$\Delta_4 = a_{11}a_{22}a_{33}a_{44}$$

$$\Delta_3 = a_{11}a_{22}a_{33}$$

$$\Delta_2 = a_{11}a_{22}$$

$$\Delta_1 = a_{11}$$

The Hurwitz conditions for asymptotic stability

$$\Delta_1 > 0, \qquad \Delta_2 > 0, \qquad \Delta_3 > 0, \qquad \Delta_4 > 0, \qquad \ldots$$

reduce to the conditions

$$a_{11} > 0, \qquad a_{22} > 0, \qquad a_{33} > 0, \qquad a_{44} > 0, \qquad \ldots$$

The Routh array for the polynomial

$$a_0 s^4 + a_1 s^3 + a_2 s^2 + a_3 s + a_4 = 0$$

where $a_0 > 0$, is given by

$$\begin{matrix} a_0 & a_2 & a_4 \\ a_1 & a_3 & \\ b_1 & b_2 & \\ c_1 & & \\ d_1 & & \end{matrix}$$

From this Routh array, we see that

$$a_{11} = a_1$$

$$a_{22} = a_2 - \frac{a_0}{a_1} a_3 = b_1$$

$$a_{33} = a_3 - \frac{a_1}{a_{22}} a_{23} = \frac{a_3 b_1 - a_1 b_2}{b_1} = c_1$$

$$a_{44} = \hat{a}_{44} - \frac{\hat{a}_{43}}{a_{33}} a_{34} = \frac{b_2 c_1 - b_1 c_2}{c_1} = d_1$$

Hence the Hurwitz conditions for asymptotic stability become

$$a_1 > 0, \qquad b_1 > 0, \qquad c_1 > 0, \qquad d_1 > 0, \qquad \ldots$$

Thus we have demonstrated that Hurwitz conditions for asymptotic stability can be reduced to Routh's conditions for asymptotic stability. The same argument can be extended to Hurwitz determinants of any order, and the equivalence of Routh's stability criterion and Hurwitz stability criterion can be established.

**A-5-11.** Show that the first column of the Routh array of

$$s^n + a_1 s^{n-1} + a_2 s^{n-2} + \cdots + a_{n-1} s + a_n = 0$$

is given by

$$1, \qquad \Delta_1, \qquad \frac{\Delta_2}{\Delta_1}, \qquad \frac{\Delta_3}{\Delta_2}, \qquad \ldots, \qquad \frac{\Delta_n}{\Delta_{n-1}}$$

where

$$\Delta_r = \begin{vmatrix} a_1 & 1 & 0 & 0 & \cdots & 0 \\ a_3 & a_2 & a_1 & 1 & \cdots & 0 \\ a_5 & a_4 & a_3 & a_2 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ a_{2r-1} & \cdot & \cdot & \cdot & \cdots & a_r \end{vmatrix}$$

$$a_k = 0 \qquad \text{if } k > n$$

**Solution.** The Routh array of coefficients has the form

$$\begin{array}{llllll}
1 & a_2 & a_4 & a_6 & \cdots & a_n \\
a_1 & a_3 & a_5 & \cdots & & \\
b_1 & b_2 & b_3 & \cdots & & \\
c_1 & c_2 & \cdot & & & \\
\cdot & \cdot & \cdot & & & \\
\cdot & \cdot & \cdot & & & \\
\cdot & \cdot & \cdot & & &
\end{array}$$

The first term in the first column of the Routh array is 1. The next term in the first column is $a_1$, which is equal to $\Delta_1$. The next term is $b_1$, which is equal to

$$\frac{a_1a_2 - a_3}{a_1} = \frac{\Delta_2}{\Delta_1}$$

The next term in the first column is $c_1$, which is equal to

$$\begin{aligned}
\frac{b_1a_3 - a_1b_2}{b_1} &= \frac{\left[\dfrac{a_1a_2 - a_3}{a_1}\right]a_3 - a_1\left[\dfrac{a_1a_4 - a_5}{a_1}\right]}{\left[\dfrac{a_1a_2 - a_3}{a_1}\right]} \\
&= \frac{a_1a_2a_3 - a_3^2 - a_1^2a_4 + a_1a_5}{a_1a_2 - a_3} \\
&= \frac{\Delta_3}{\Delta_2}
\end{aligned}$$

In a similar manner the remaining terms in the first column of the Routh array can be found.

The Routh array has the property that the last nonzero terms of any columns are the same; that is, if the array is given by

$$\begin{array}{llll}
a_0 & a_2 & a_4 & a_6 \\
a_1 & a_3 & a_5 & a_7 \\
b_1 & b_2 & b_3 & \\
c_1 & c_2 & c_3 & \\
d_1 & d_2 & & \\
e_1 & e_2 & & \\
f_1 & & & \\
g_1 & & &
\end{array}$$

then

$$a_7 = c_3 = e_2 = g_1$$

and if the array is given by

$$\begin{array}{llll}
a_0 & a_2 & a_4 & a_6 \\
a_1 & a_3 & a_5 & 0 \\
b_1 & b_2 & b_3 & \\
c_1 & c_2 & 0 & \\
d_1 & d_2 & & \\
e_1 & 0 & & \\
f_1 & & &
\end{array}$$

then

$$a_6 = b_3 = d_2 = f_1$$

In any case, the last term of the first column is equal to $a_n$, or

$$a_n = \frac{\Delta_{n-1}a_n}{\Delta_{n-1}} = \frac{\Delta_n}{\Delta_{n-1}}$$

For example, if $n = 4$, then

$$\Delta_4 = \begin{vmatrix} a_1 & 1 & 0 & 0 \\ a_3 & a_2 & a_1 & 1 \\ a_5 & a_4 & a_3 & a_2 \\ a_7 & a_6 & a_5 & a_4 \end{vmatrix} = \begin{vmatrix} a_1 & 1 & 0 & 0 \\ a_3 & a_2 & a_1 & 1 \\ 0 & a_4 & a_3 & a_2 \\ 0 & 0 & 0 & a_4 \end{vmatrix} = \Delta_3 a_4$$

Thus it has been shown that the first column of the Routh array is given by

$$1, \quad \Delta_1, \quad \frac{\Delta_2}{\Delta_1}, \quad \frac{\Delta_3}{\Delta_2}, \quad \ldots, \quad \frac{\Delta_n}{\Delta_{n-1}}$$

**A-5-12.** The value of the gas constant for any gas may be determined from accurate experimental observations of simultaneous values of $p$, $v$, and $T$.

Obtain the gas constant $R_{air}$ for air. Note that at 32°F and 14.7 psia the specific volume of air is 12.39 ft³/lb. Then obtain the capacitance of a 20-ft³ pressure vessel that contains air at 160°F. Assume that the expansion process is isothermal.

**Solntion.**

$$R_{\text{air}} = \frac{pv}{T} = \frac{14.7 \times 144 \times 12.39}{460 + 32} = 53.3 \text{ ft-lb}_f/\text{lb °R}$$

Referring to Equation (5–12), the capacitance of a 20-ft³ pressure vessel is

$$C = \frac{V}{nR_{\text{air}}T} = \frac{20}{1 \times 53.3 \times 620} = 6.05 \times 10^{-4} \frac{\text{lb}}{\text{lb}_f/\text{ft}^2}$$

Note that in terms of SI units, $R_{air}$ is given by

$$R_{\text{air}} = 287 \text{ N-m/kg K}$$

**A-5-13.** Figure 5–67 is a schematic diagram of a pneumatic diaphragm valve. At steady state the control pressure from a controller is $\bar{P}_c$, the pressure in the valve is also $\bar{P}_c$, and the valve-stem displacement is $\bar{X}$. Assume that at $t = 0$ the control pressure is changed from $\bar{P}_c$ to $\bar{P}_c + p_c$. Then the valve pressure will be changed from $\bar{P}_c$ to $\bar{P}_c + p_v$. The change in valve pressure $p_v$ will cause the valve-stem displacement to change from $\bar{X}$ to $\bar{X} + x$. Find the transfer function between the change in valve-stem displacement $x$ and the change in control pressure $p_c$.

**Solntion.** Let us define the airflow rate to the diaphragm valve through resistance $R$ as $q$. Then

$$q = \frac{p_c - p_v}{R}$$

For the air chamber in the diaphragm valve, we have

$$C\,dp_v = q\,dt$$

Consequently,

$$C\frac{dp_v}{dt} = q = \frac{p_c - p_v}{R}$$

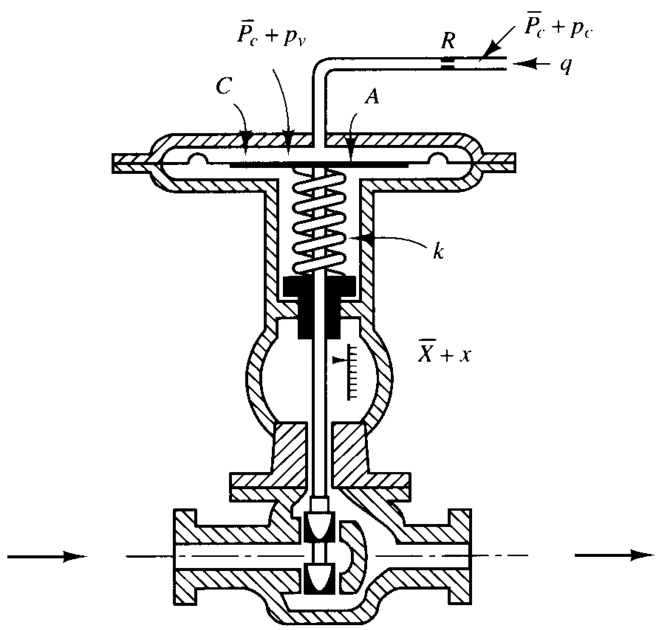



**Figure 5–67**
Pneumatic diaphragm valve.

from which

$$RC\frac{dp_v}{dt} + p_v = p_c$$

Noting that

$$Ap_v = kx$$

we have

$$\frac{k}{A}\left(RC\frac{dx}{dt} + x\right) = p_c$$

The transfer function between $x$ and $p_c$ is

$$\frac{X(s)}{P_c(s)} = \frac{A/k}{RCs + 1}$$

**A-5-14.** In the pneumatic pressure system of Figure 5–68(a), assume that, for $t < 0$, the system is at steady state and that the pressure of the entire system is $\bar{P}$. Also, assume that the two bellows are identical. At $t = 0$, the input pressure is changed from $\bar{P}$ to $\bar{P} + p_i$. Then the pressures in bellows 1 and 2 will change from $\bar{P}$ to $\bar{P} + p_1$ and from $\bar{P}$ to $\bar{P} + p_2$, respectively. The capacity (volume) of each bellows is $5 \times 10^{-4}$ m$^3$, and the operating pressure difference $\Delta p$ (difference between $p_i$ and $p_1$ or difference between $p_i$ and $p_2$) is between $-0.5 \times 10^5$ N/m$^2$ and $0.5 \times 10^5$ N/m$^2$. The corresponding mass flow rates (kg/sec) through the valves are shown in Figure 5–68(b). Assume that the bellows expand or contract linearly with the air pressures applied to them, that the equivalent spring constant of the bellows system is $k = 1 \times 10^5$ N/m, and that each bellows has area $A = 15 \times 10^{-4}$ m$^2$.

Defining the displacement of the midpoint of the rod that connects two bellows as $x$, find the transfer function $X(s)/P_i(s)$. Assume that the expansion process is isothermal and that the temperature of the entire system stays at 30°C.

**Solution.** Referring to Section 5–6, transfer function $P_1(s)/P_i(s)$ can be obtained as

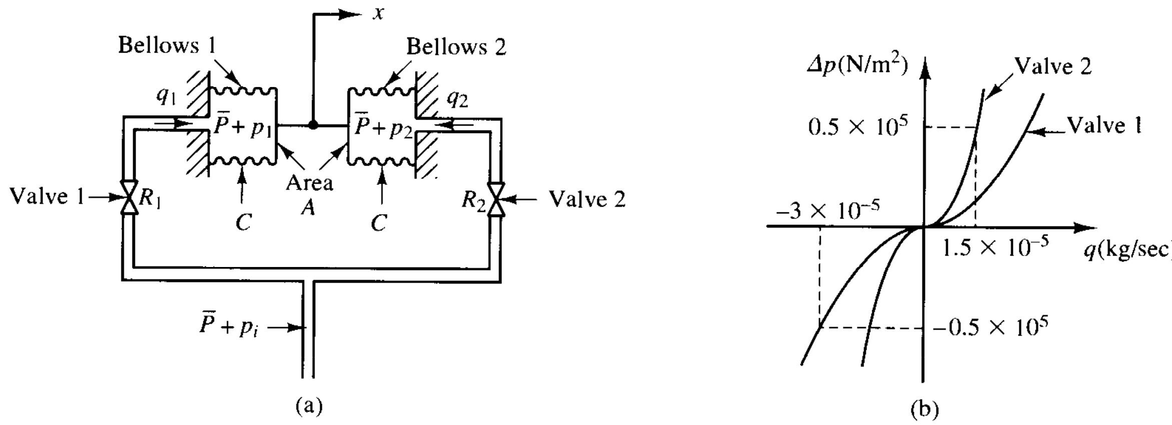



**Figure 5–68**
(a) Pneumatic pressure system; (b) pressure difference versus mass flow rate curve.

$$\frac{P_1(s)}{P_i(s)} = \frac{1}{R_1Cs + 1} \tag{5-38}$$

Similarly, transfer function $P_2(s)/P_i(s)$ is

$$\frac{P_2(s)}{P_i(s)} = \frac{1}{R_2Cs + 1} \tag{5-39}$$

The force acting on bellows 1 in the $x$ direction is $A(\bar{P} + p_1)$, and the force acting on bellows 2 in the negative $x$ direction is $A(\bar{P} + p_2)$. The resultant force balances with $kx$, the equivalent spring force of the corrugated side of the bellows.

$$A(p_1 - p_2) = kx$$

or

$$A[P_1(s) - P_2(s)] = kX(s) \tag{5-40}$$

Referring to Equations (5–38) and (5–39), we see that

$$\begin{aligned} P_1(s) - P_2(s) &= \left(\frac{1}{R_1Cs + 1} - \frac{1}{R_2Cs + 1}\right)P_i(s) \\ &= \frac{R_2Cs - R_1Cs}{(R_1Cs + 1)(R_2Cs + 1)}P_i(s) \end{aligned}$$

By substituting this last equation into Equation (5–40) and rewriting, the transfer function $X(s)/P_i(s)$ is obtained as

$$\frac{X(s)}{P_i(s)} = \frac{A}{k}\frac{(R_2C - R_1C)s}{(R_1Cs + 1)(R_2Cs + 1)} \tag{5-41}$$

The numerical values of average resistances $R_1$ and $R_2$ are

$$R_1 = \frac{d\,\Delta p}{dq_1} = \frac{0.5 \times 10^5}{3 \times 10^{-5}} = 0.167 \times 10^{10}\,\frac{\text{N/m}^2}{\text{kg/sec}}$$

$$R_2 = \frac{d\,\Delta p}{dq_2} = \frac{0.5 \times 10^5}{1.5 \times 10^{-5}} = 0.333 \times 10^{10}\,\frac{\text{N/m}^2}{\text{kg/sec}}$$

The numerical value of capacitance $C$ of each bellows is

$$C = \frac{V}{nR_{\text{air}}T} = \frac{5 \times 10^{-4}}{1 \times 287 \times (273 + 30)} = 5.75 \times 10^{-9}\,\frac{\text{kg}}{\text{N/m}^2}$$

where $R_{air} = 287$ N-m/kg K. (See Problem A-5-12.) Consequently,

$$R_1C = 0.167 \times 10^{10} \times 5.75 \times 10^{-9} = 9.60 \text{ sec}$$

$$R_2C = 0.333 \times 10^{10} \times 5.75 \times 10^{-9} = 19.2 \text{ sec}$$

By substituting the numerical values for $A, k, R_1C$, and $R_2C$ into Equation (5–41), we obtain

$$\frac{X(s)}{P_i(s)} = \frac{1.44 \times 10^{-7}s}{(9.6s + 1)(19.2s + 1)}$$

**A-5-15.** Draw a block diagram of the pneumatic controller shown in Figure 5–69. Then derive the transfer function of this controller.

If the resistance $R_d$ is removed (replaced by the line-sized tubing), what control action do we get? If the resistance $R_i$ is removed (replaced by the line-sized tubing), what control action do we get?

**Solution.** Let us assume that when $e = 0$ the nozzle–flapper distance is equal to $\bar{X}$ and the control pressure is equal to $\bar{P}_c$. In the present analysis, we shall assume small deviations from the respective reference values as follows:

$e$ = small error signal

$x$ = small change in the nozzle-flapper distance

$p_c$ = small change in the control pressure

$p_{\text{I}}$ = small pressure change in bellows I due to small change in the control pressure

$p_{\text{II}}$ = small pressure change in bellows II due to small change in the control pressure

$y$ = small displacement at the lower end of the flapper

In this controller, $p_c$ is transmitted to bellows I through the resistance $R_d$. Similarly, $p_c$ is transmitted to bellows II through the series of resistances $R_d$ and $R_i$. An approximate relationship between $p_{\text{I}}$ and $p_c$ is

$$\frac{P_{\text{I}}(s)}{P_c(s)} = \frac{1}{R_dCs + 1} = \frac{1}{T_ds + 1}$$

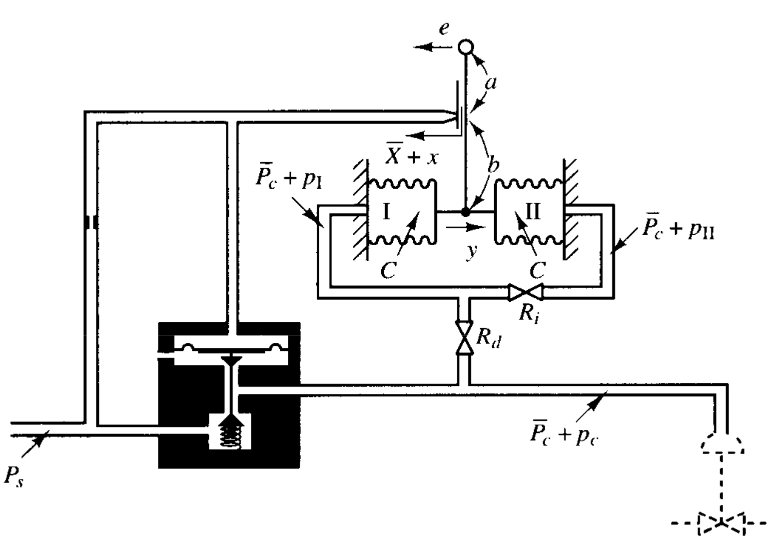



Figure 5–69
Schematic diagram of a pneumatic controller.

where $T_d = R_dC$ = derivative time. Similarly, $p_{II}$ and $p_I$ are related by the transfer function

$$\frac{P_{II}(s)}{P_I(s)} = \frac{1}{R_iCs + 1} = \frac{1}{T_is + 1}$$

where $T_i = R_iC$ = integral time. The force–balance equation for the two bellows is

$$(p_I - p_{II})A = k_s y$$

where $k_s$ is the stiffness of the two connected bellows and $A$ is the cross-sectional area of the bellows. The relationship among the variables $e$, $x$, and $y$ is

$$x = \frac{b}{a+b}e - \frac{a}{a+b}y$$

The relationship between $p_c$ and $x$ is

$$p_c = Kx \qquad (K > 0)$$

From the equations just derived, a block diagram of the controller can be drawn, as shown in Figure 5–70(a). Simplification of this block diagram results in Figure 5–70(b).

The transfer function between $P_c(s)$ and $E(s)$ is

$$\frac{P_c(s)}{E(s)} = \frac{\frac{b}{a+b}K}{1 + K\frac{a}{a+b}\frac{A}{k_s}\left(\frac{T_is}{T_is+1}\right)\left(\frac{1}{T_ds+1}\right)}$$

For a practical controller, under normal operation $|KaAT_is/[(a + b)k_s(T_is + 1)(T_ds + 1)]|$ is very much greater than unity and $T_i \gg T_d$. Therefore, the transfer function can be simplified as follows:

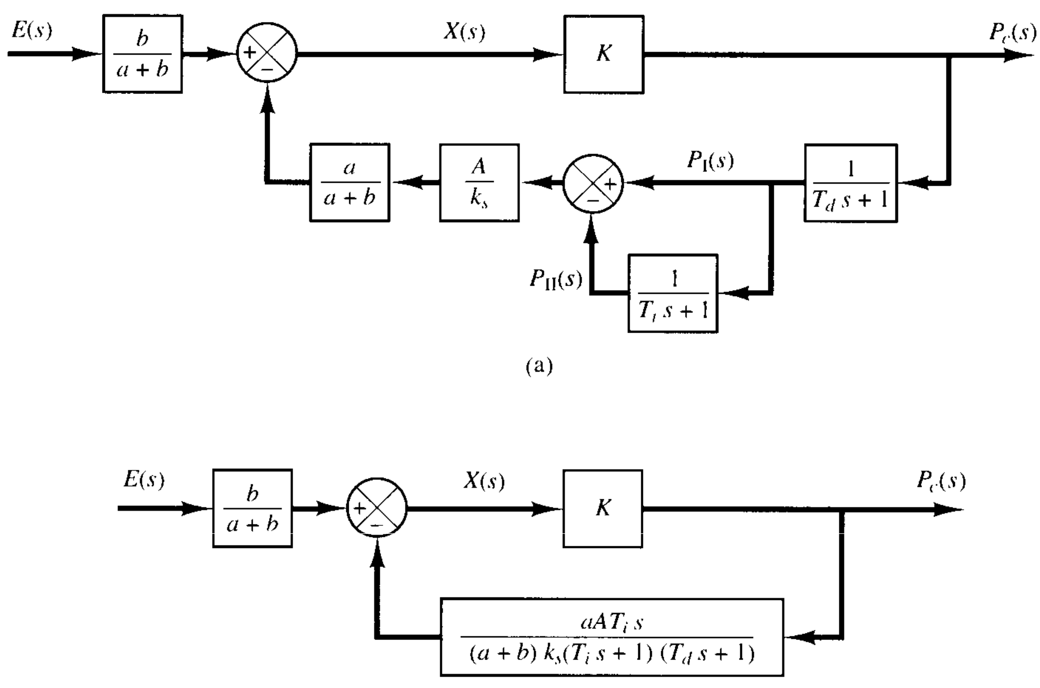



**Figure 5–70**
(a) Block diagram of the pneumatic controller shown in Figure 5–69;
(b) simplified block diagram.

$$\frac{P_c(s)}{E(s)} \doteq \frac{bk_s(T_i s + 1)(T_d s + 1)}{aAT_i s}$$

$$= \frac{bk_s}{aA}\left(\frac{T_i + T_d}{T_i} + \frac{1}{T_i s} + T_d s\right)$$

$$\doteq K_p\left(1 + \frac{1}{T_i s} + T_d s\right)$$

where

$$K_p = \frac{bk_s}{aA}$$

Thus the controller shown in Figure 5–69 is a proportional-plus-integral-plus-derivative one.

If the resistance $R_d$ is removed, or $R_d = 0$, the action becomes that of a proportional-plus-integral controller.

If the resistance $R_i$ is removed, or $R_i = 0$, the action becomes that of a narrow-band proportional, or two-position, controller. (Note that the actions of two feedback bellows cancel each other, and there is no feedback.)

**A-5-16.** Actual spool valves are either overlapped or underlapped because of manufacturing tolerances. Consider the overlapped and underlapped spool valves shown in Figure 5–71(a) and (b). Sketch curves relating the uncovered port area $A$ versus displacement $x$.

**Solution.** For the overlapped valve, a dead zone exists between $-\frac{1}{2}x_0$ and $\frac{1}{2}x_0$, or $-\frac{1}{2}x_0 < x < \frac{1}{2}x_0$. The uncovered port area $A$ versus displacement $x$ curve is shown in Figure 5–72(a). Such an overlapped valve is unfit as a control valve.

For the underlapped valve, the port area $A$ versus displacement $x$ curve is shown in Figure 5–72(b). The effective curve for the underlapped region has a higher slope, meaning a higher sensitivity. Valves used for controls are usually underlapped.

**A-5-17.** Figure 5–73 shows a hydraulic jet-pipe controller. Hydraulic fluid is ejected from the jet pipe. If the jet pipe is shifted to the right from the neutral position, the power piston moves to the left, and vice versa. The jet pipe valve is not used as much as the flapper valve because of large null flow, slower response, and rather unpredictable characteristics. Its main advantage lies in its insensitivity to dirty fluids.

Suppose that the power piston is connected to a light load so that the inertia force of the load element is negligible compared to the hydraulic force developed by the power piston. What type of control action does this controller produce?

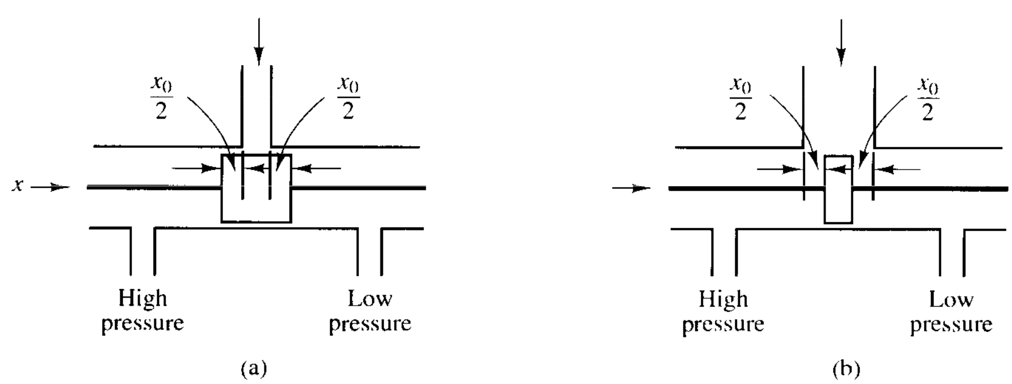
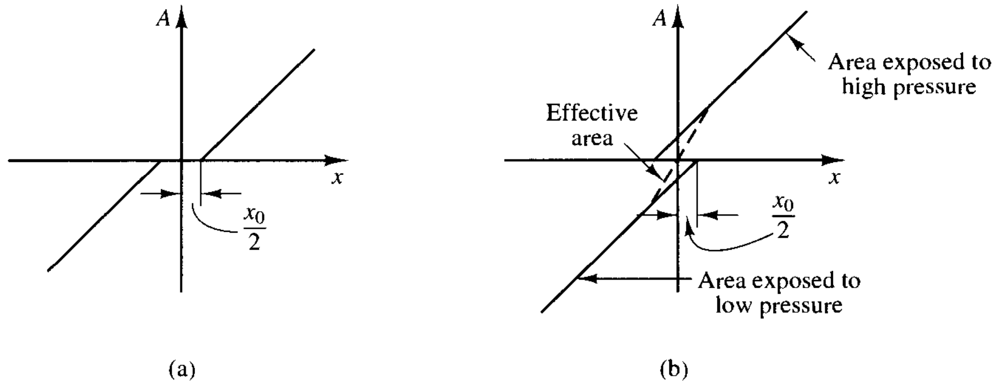



**Figure 5–71**
(a) Overlapped spool valve; (b) underlapped spool valve.

**Figure 5–72**
(a) Uncovered port area $A$ versus displacement $x$ curve for the overlapped valve; (b) uncovered port area $A$ versus displacement $x$ curve for the underlapped valve.



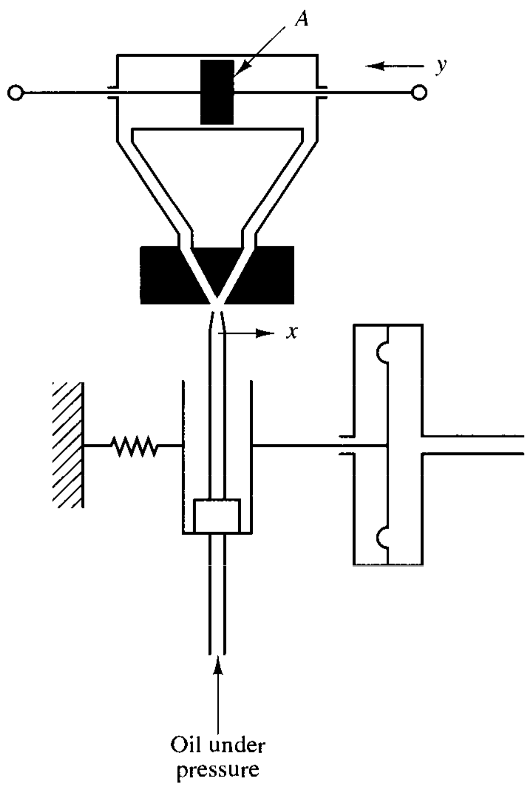




**Figure 5–73**
Hydraulic jet-pipe controller.

**Solution.** Define the displacement of the jet nozzle from the neutral position as $x$ and the displacement of the power piston as $y$. If the jet nozzle is moved to the right by a small displacement $x$, the oil flows to the right side of the power piston, and the oil in the left side of the power piston is returned to the drain. The oil flowing into the power cylinder is at high pressure; the oil flowing out from the power cylinder into the drain is at low pressure. The resulting pressure difference causes the power piston to move to the left.

For a small jet nozzle displacement $x$, the flow rate $q$ to the power cylinder is proportional to $x$; that is,

$$q = K_1 x$$

For the power cylinder,

$$A\rho\, dy = q\, dt$$

where $A$ is the power piston area and $\rho$ is the density of oil. Hence

$$\frac{dy}{dt} = \frac{q}{A\rho} = \frac{K_1}{A\rho}x = Kx$$

where $K = K_1/(A\rho)$ = constant. The transfer function $Y(s)/X(s)$ is thus

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

The controller produces the integral control action.

**A-5-18.** Figure 5–74 shows a hydraulic jet-pipe applied to a flow control system. The jet-pipe controller governs the position of the butterfly valve. Discuss the operation of this system. Plot a possible curve relating the displacement $x$ of the nozzle to the total force $F$ acting on the power piston.

**Solution.** The operation of this system is as follows: The flow rate is measured by the orifice, and the pressure difference produced by this orifice is transmitted to the diaphragm of the pressure-measuring device. The diaphragm is connected to the free swinging nozzle, or jet pipe, through a linkage. High-pressure oil ejects from the nozzle all the time. When the nozzle is at a neutral position, no oil flows through either of the pipes to move the power piston. If the nozzle is displaced by the motion of the balance arm to one side, the high-pressure oil flows through the corresponding pipe, and the oil in the power cylinder flows back to the sump through the other pipe.

Assume that the system is initially at rest. If the reference input is changed suddenly to a higher flow rate, then the nozzle is moved in such a direction as to move the power piston and open the butterfly valve. Then the flow rate will increase, the pressure difference across the orifice becomes larger, and the nozzle will move back to the neutral position. The movement of the power piston stops when $x$, the displacement of the nozzle, comes back to and stays at the neutral position. (The jet pipe controller thus possesses an integrating property.)

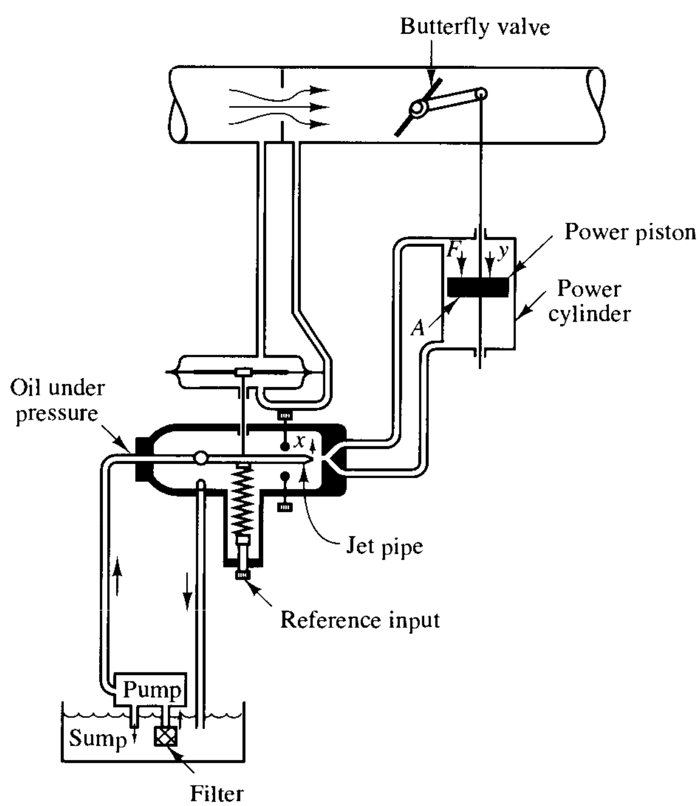



**Figure 5–74** Schematic diagram of a flow control system using a hydraulic jet-pipe controller.

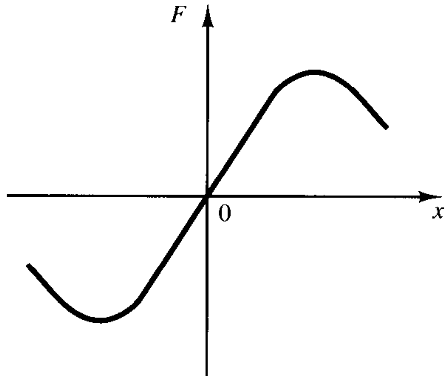



**Figure 5–75**
Force versus diplacement curve.

The relationship between the total force $F$ acting on the power piston and the displacement $x$ of the nozzle is shown in Figure 5–75. The total force is equal to the pressure difference $\Delta P$ across the piston times the area $A$ of the power piston. For a small displacement $x$ of the nozzle, the total force $F$ and displacement $x$ may be considered proportional.

**A-5-19.** Explain the operation of the speed control system shown in Figure 5–76.

**Solution.** If the engine speed increases, the sleeve of the fly-ball governor moves upward. This movement acts as the input to the hydraulic controller. A positive error signal (upward motion of the sleeve) causes the power piston to move downward, reduces the fuel-valve opening, and decreases the engine speed. A block diagram for the system is shown in Figure 5–77.

From the block diagram the transfer function $Y(s)/E(s)$ can be obtained as

$$\frac{Y(s)}{E(s)} = \frac{a_2}{a_1 + a_2} \frac{\frac{K}{s}}{1 + \frac{a_1}{a_1 + a_2} \frac{bs}{bs + k} \frac{K}{s}}$$

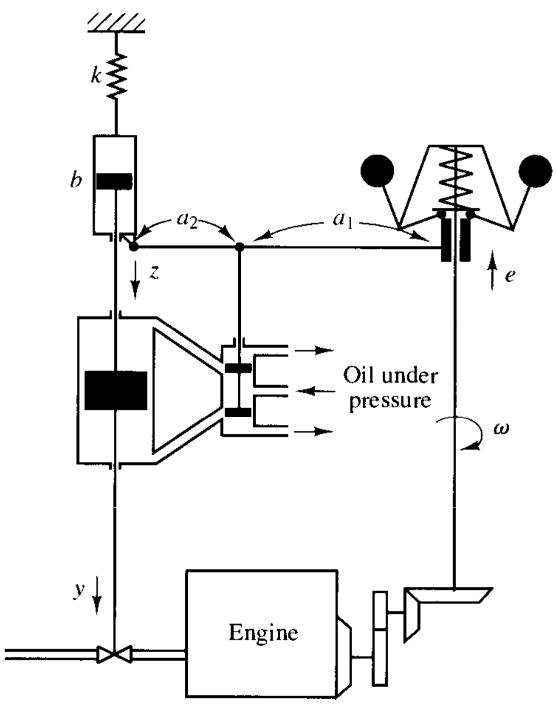



**Figure 5–76**
Speed control system.

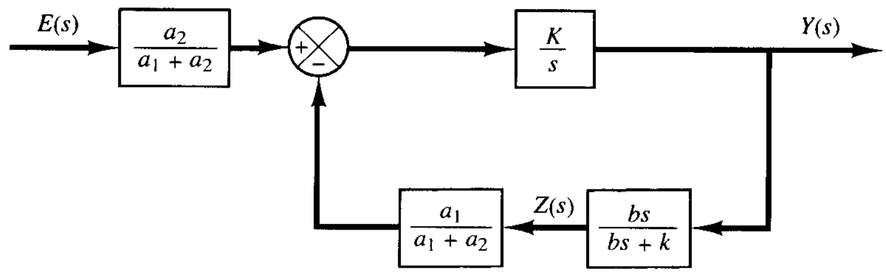



**Figure 5–77**
Block diagram for the speed control system shown in Figure 5–76.

If the following condition applies,

$$\left| \frac{a_1}{a_1 + a_2} \frac{bs}{bs + k} \frac{K}{s} \right| \gg 1$$

the transfer function $Y(s)/E(s)$ becomes

$$\frac{Y(s)}{E(s)} \doteq \frac{a_2}{a_1 + a_2} \frac{a_1 + a_2}{a_1} \frac{bs + k}{bs} = \frac{a_2}{a_1}\left(1 + \frac{k}{bs}\right)$$

The speed controller has proportional-plus-integral control action.

**A-5-20.** Consider the hydraulic servo system shown in Figure 5–78. Assuming that signal $e(t)$ is the input and power piston displacement $y(t)$ the output, find the transfer function $Y(s)/E(s)$.

**Solution.** A block diagram for the system can be drawn as shown in Figure 5–79. Assuming that $|K_1a_1/[s(a_1 + a_2)]| \gg 1$ and $|K_2b_1/[s(b_1 + b_2)]| \gg 1$, we obtain

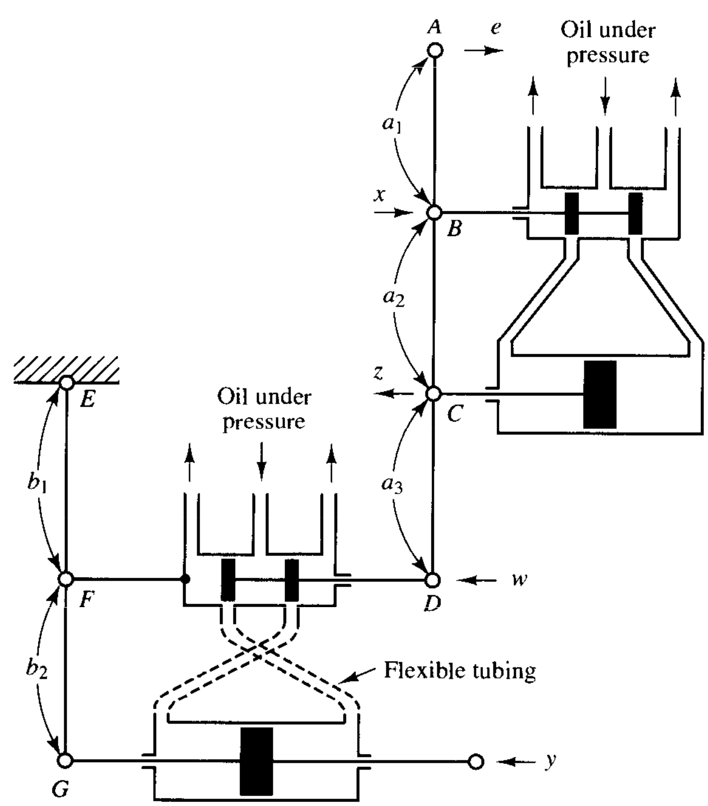



**Figure 5–78**
Hydraulic servo system.

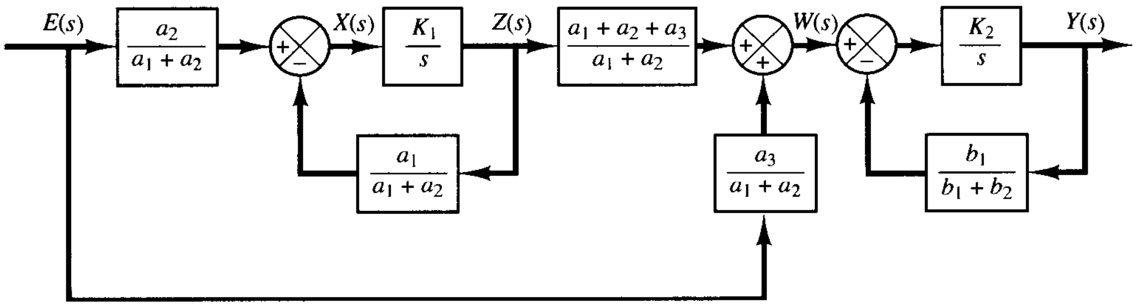



**Figure 5–79**
Block diagram for the system shown in Figure 5–78.

$$\frac{Z(s)}{E(s)} = \frac{\dfrac{a_2}{a_1 + a_2} \cdot \dfrac{K_1}{s}}{1 + \dfrac{K_1}{s} \cdot \dfrac{a_1}{a_1 + a_2}} \doteqdot \frac{a_2}{a_1 + a_2} \cdot \frac{a_1 + a_2}{a_1} = \frac{a_2}{a_1}$$

$$\frac{W(s)}{E(s)} = \frac{a_1 + a_2 + a_3}{a_1 + a_2} \cdot \frac{Z(s)}{E(s)} + \frac{a_3}{a_1 + a_2} = \frac{a_2 + a_3}{a_1}$$

$$\frac{Y(s)}{W(s)} = \frac{\dfrac{K_2}{s}}{1 + \dfrac{b_1}{b_1 + b_2} \dfrac{K_2}{s}} \doteqdot \frac{b_1 + b_2}{b_1}$$

Hence

$$\frac{Y(s)}{E(s)} = \frac{Y(s)}{W(s)} \cdot \frac{W(s)}{E(s)} = \frac{(a_2 + a_3)(b_1 + b_2)}{a_1 b_1}$$

This servo system is a proportional controller.

**A-5-21.** Obtain the transfer function $E_o(s)/E_i(s)$ of the op-amp circuit shown in Figure 5–80.

**Solution.** Define the voltage at point $A$ as $e_A$. Then

$$\frac{E_A(s)}{E_i(s)} = \frac{R_1}{\dfrac{1}{Cs} + R_1} = \frac{R_1 Cs}{R_1 Cs + 1}$$

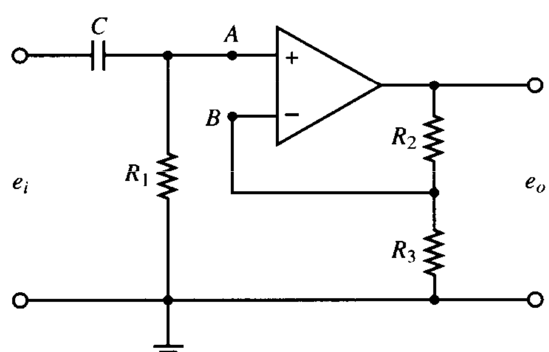



**Figure 5–80**
Operational-amplifier circuit.

Define the voltage at point $B$ as $e_B$. Then

$$E_B(s) = \frac{R_3}{R_2 + R_3} E_o(s)$$

Noting that

$$[E_A(s) - E_B(s)]K = E_o(s)$$

and $K \gg 1$, we must have

$$E_A(s) = E_B(s)$$

Hence

$$E_A(s) = \frac{R_1Cs}{R_1Cs + 1} E_i(s) = E_B(s) = \frac{R_3}{R_2 + R_3} E_o(s)$$

from which we obtain

$$\frac{E_o(s)}{E_i(s)} = \frac{R_2 + R_3}{R_3} \frac{R_1Cs}{R_1Cs + 1} = \frac{\left(1 + \frac{R_2}{R_3}\right)s}{s + \frac{1}{R_1C}}$$

**A-5-22.** Obtain the transfer function $E_o(s)/E_i(s)$ of the op-amp circuit shown in Figure 5–81.

**Solution.** The voltage at point $A$ is

$$e_A = \frac{1}{2}(e_i - e_o) + e_o$$

The Laplace-transformed version of this last equation is

$$E_A(s) = \frac{1}{2}[E_i(s) + E_o(s)]$$

The voltage at point $B$ is

$$E_B(s) = \frac{\frac{1}{Cs}}{R_2 + \frac{1}{Cs}} E_i(s) = \frac{1}{R_2Cs + 1} E_i(s)$$

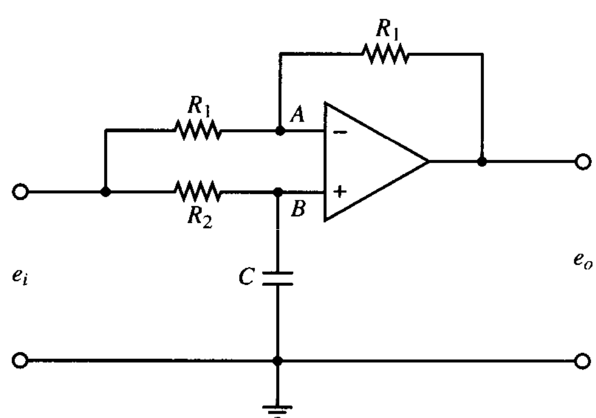



**Figure 5–81**
Operational-amplifier circuit.

Since $[E_B(s) - E_A(s)]K = E_o(s)$ and $K \gg 1$, we must have $E_A(s) = E_B(s)$. Thus

$$\frac{1}{2}[E_i(s) + E_o(s)] = \frac{1}{R_2Cs + 1}E_i(s)$$

Hence

$$\frac{E_o(s)}{E_i(s)} = -\frac{R_2Cs - 1}{R_2Cs + 1} = -\frac{s - \dfrac{1}{R_2C}}{s + \dfrac{1}{R_2C}}$$

**A-5-23.** Consider the stable unity-feedback control system with feedforward transfer function $G(s)$. Suppose that the closed-loop transfer function can be written

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)} = \frac{(T_as + 1)(T_bs + 1)\cdots(T_ms + 1)}{(T_1s + 1)(T_2s + 1)\cdots(T_ns + 1)} \qquad (m \leq n)$$

Show that

$$\int_0^\infty e(t)\,dt = (T_1 + T_2 + \cdots + T_n) - (T_a + T_b + \cdots + T_m)$$

where $e(t)$ is the error in the unit-step response. Show also that

$$\frac{1}{\lim_{s\to 0} sG(s)} = (T_1 + T_2 + \cdots + T_n) - (T_a + T_b + \cdots + T_m)$$

**Solution.** Let us define

$$(T_as + 1)(T_bs + 1)\cdots(T_ms + 1) = P(s)$$

and

$$(T_1s + 1)(T_2s + 1)\cdots(T_ns + 1) = Q(s)$$

Then

$$\frac{C(s)}{R(s)} = \frac{P(s)}{Q(s)}$$

and

$$E(s) = \frac{Q(s) - P(s)}{Q(s)}R(s)$$

For a unit-step input, $R(s) = 1/s$ and

$$E(s) = \frac{Q(s) - P(s)}{sQ(s)}$$

Since the system is stable, $\int_0^\infty e(t)\,dt$ converges to a constant value. Referring to Table 2–2 (item 9), we have

$$\int_0^\infty e(t)\,dt = \lim_{s\to 0} s\frac{E(s)}{s} = \lim_{s\to 0} E(s)$$

Hence

$$\int_0^\infty e(t)\,dt = \lim_{s\to 0} \frac{Q(s) - P(s)}{sQ(s)}$$

$$= \lim_{s\to 0} \frac{Q'(s) - P'(s)}{Q(s) + sQ'(s)}$$

$$= \lim_{s\to 0} [Q'(s) - P'(s)]$$

Since

$$\lim_{s\to 0} P'(s) = T_a + T_b + \cdots + T_m$$

$$\lim_{s\to 0} Q'(s) = T_1 + T_2 + \cdots + T_n$$

we have

$$\int_0^\infty e(t)\,dt = (T_1 + T_2 + \cdots + T_n) - (T_a + T_b + \cdots + T_m)$$

For a unit-step input $r(t)$, since

$$\int_0^\infty e(t)\,dt = \lim_{s\to 0} E(s) = \lim_{s\to 0} \frac{1}{1 + G(s)} R(s) = \lim_{s\to 0} \frac{1}{1 + G(s)} \frac{1}{s} = \frac{1}{\lim_{s\to 0} sG(s)} = \frac{1}{K_v}$$

we have

$$\frac{1}{\lim_{s\to 0} sG(s)} = (T_1 + T_2 + \cdots + T_n) - (T_a + T_b + \cdots + T_m)$$

Note that zeros in the left half-plane (that is, positive $T_a, T_b, \ldots, T_m$) will improve $K_v$. Poles close to the origin cause low velocity-error constants unless there are zeros nearby.

## PROBLEMS

**B-5-1.** If the feedforward path of a control system contains at least one integrating element, then the output continues to change as long as an error is present. The output stops when the error is precisely zero. If an external disturbance enters the system, it is desirable to have an integrating element between the error-measuring element and the point where the disturbance enters so that the effect of the external disturbance may be made zero at steady state.

Show that, if the disturbance is a ramp function, then the steady-state error due to this ramp disturbance may be eliminated only if two integrators precede the point where the disturbance enters.

**B-5-2.** Consider industrial automatic controllers whose control actions are proportional, integral, proportional-plus-integral, proportional-plus-derivative, and proportional-plus-integral-plus-derivative. The transfer functions of these controllers can be given, respectively, by

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

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

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

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

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

where $U(s)$ is the Laplace transform of $u(t)$, the controller output, and $E(s)$ the Laplace transform of $e(t)$, the actuating error signal. Sketch $u(t)$ versus $t$ curves for each of the five types of controllers when the actuating error signal is

(a) $e(t)$ = unit-step function

(b) $e(t)$ = unit-ramp function

In sketching curves, assume that the numerical values of $K_p$, $K_i$, $T_i$, and $T_d$ are given as

$$K_p = \text{proportional gain} = 4$$

$$K_i = \text{integral gain} = 2$$

$$T_i = \text{integral time} = 2 \text{ sec}$$

$$T_d = \text{derivative time} = 0.8 \text{ sec}$$

**B-5-3.** Consider a unity-feedback control system whose open-loop transfer function is

$$G(s) = \frac{K}{s(Js + B)}$$

Discuss the effects that varying the values of $K$ and $B$ has on the steady-state error in unit-ramp response. Sketch typical unit-ramp response curves for a small value, medium value, and large value of $K$.

**B-5-4.** Figure 5–82 shows three systems. System I is a positional servo system. System II is a positional servo system with PD control action. System III is a positional servo system with velocity feedback. Compare the unit-step, unit-impulse, and unit-ramp responses of the three systems. Which system is best with respect to the speed of response and maximum overshoot in the step response?

**B-5-5.** Consider the position control system shown in Figure 5–83. Write a MATLAB program to obtain a unit-step response and a unit-ramp response of the system. Plot curves $x_1(t)$ versus $t$, $x_2(t)$ versus $t$, $x_3(t)$ versus $t$, and $e(t)$ versus $t$ [where $e(t) = r(t) - x_1(t)$] for both the unit-step response and the unit-ramp response.

**B-5-6.** Determine the range of $K$ for stability of a unity-feedback control system whose open-loop transfer function is

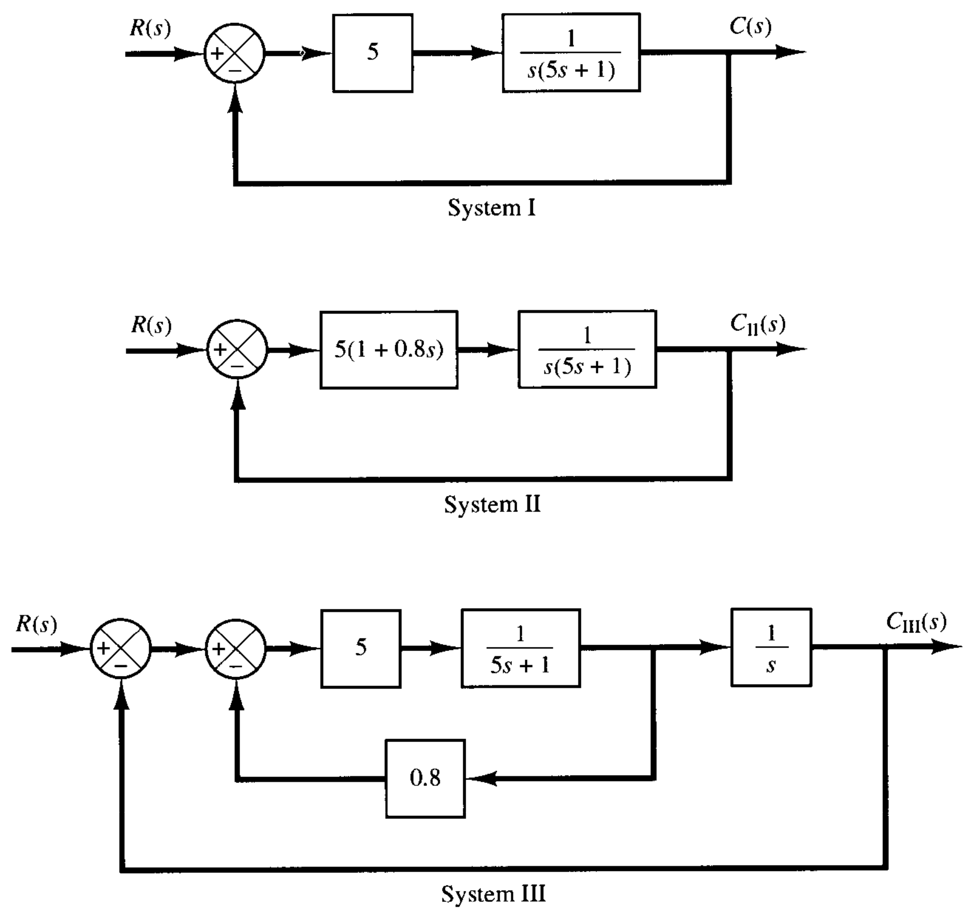



**Figure 5–82**
(a) Positional servo system; (b) positional servo system with PD control action; (c) positional servo system with velocity feedback.

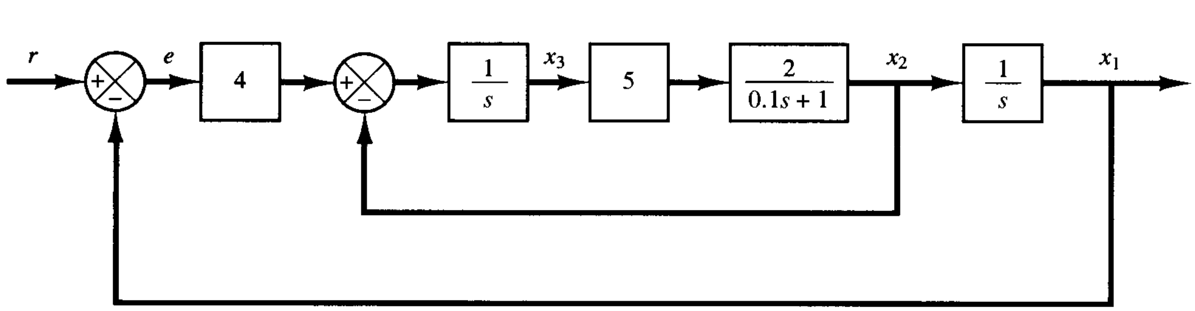



**Figure 5–83** Position control system.

$$G(s) = \frac{K}{s(s+1)(s+2)}$$

**B-5-7.** Consider the unity-feedback control system with the following open-loop transfer function:

$$G(s) = \frac{10}{s(s-1)(2s+3)}$$

Is this system stable?

**B-5-8.** Consider the system

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$$

where matrix $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ -b_3 & 0 & 1 \\ 0 & -b_2 & -b_1 \end{bmatrix}$$

($\mathbf{A}$ is called Schwarz matrix.) Show that the first column of the Routh's array of the characteristic equation $|s\mathbf{I} - \mathbf{A}| = 0$ consists of 1, $b_1$, $b_2$, and $b_1 b_3$.

**B-5-9.** Consider the pneumatic system shown in Figure 5–84. Obtain the transfer function $X(s)/P_i(s)$.

**B-5-10.** Figure 5–85 shows a pneumatic controller. What kind of control action does this controller produce? Derive the transfer function $P_c(s)/E(s)$.

**B-5-11.** Consider the pneumatic controller shown in Figure 5–86. Assuming that the pneumatic relay has the characteristics that $p_c = Kp_b$ (where $K > 0$), determine the control action of this controller. The input to the controller is $e$ and the output is $p_c$.

**B-5-12.** Figure 5–87 shows a pneumatic controller. The signal $e$ is the input and the change in the control pressure $p_c$ is the output. Obtain the transfer function $P_c(s)/E(s)$. Assume that the pneumatic relay has the characteristics that $p_c = Kp_b$, where $K > 0$.

**B-5-13.** Consider the pneumatic controller shown in Figure 5–88. What control action does this controller produce? Assume that the pneumatic relay has the characteristics that $p_c = Kp_b$, where $K > 0$.

**B-5-14.** Figure 5–89 shows an electric-pneumatic transducer. Show that the change in the output pressure is proportional to the change in the input current.

**B-5-15.** Figure 5–90 shows a flapper valve. It is placed between two opposing nozzles. If the flapper is moved slightly to the right, the pressure unbalance occurs in the nozzles and the power piston moves to the left, and vice versa. Such a device is frequently used in hydraulic servos as the first-stage valve in two-stage servovalves. This usage occurs because considerable force may be needed to stroke larger spool valves that result from the steady-state flow force. To reduce or compensate this force, two-stage valve configuration is often employed; a flapper valve or jet pipe is used as the first-stage valve to provide a necessary force to stroke the second-stage spool valve.

Figure 5–91 shows a schematic diagram of a hydraulic servomotor in which the error signal is amplified in two stages using a jet pipe and a pilot valve. Draw a block diagram of the system of Figure 5–91 and then find the trans-

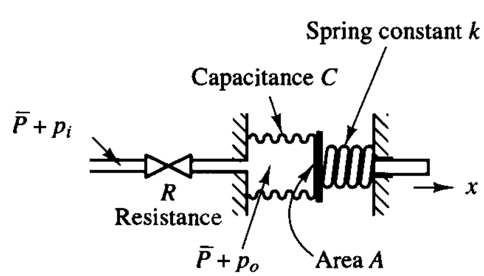



**Figure 5–84** Pneumatic system.

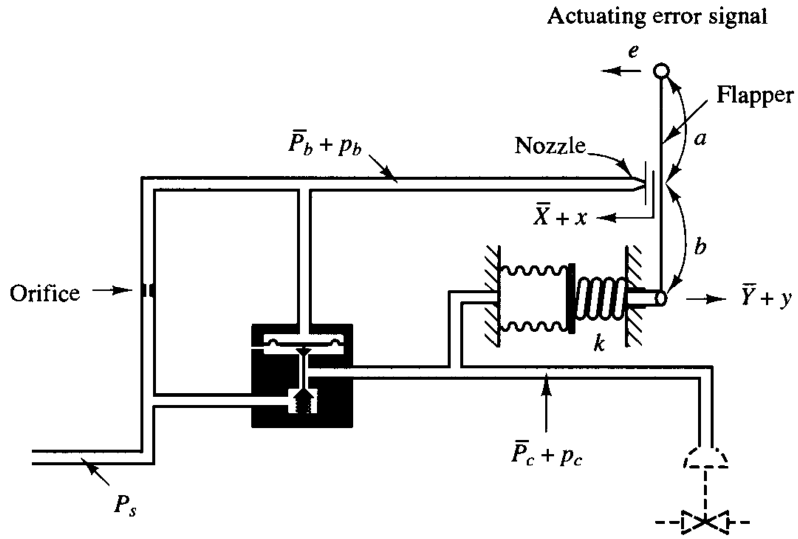



**Figure 5–85** Pneumatic controller.

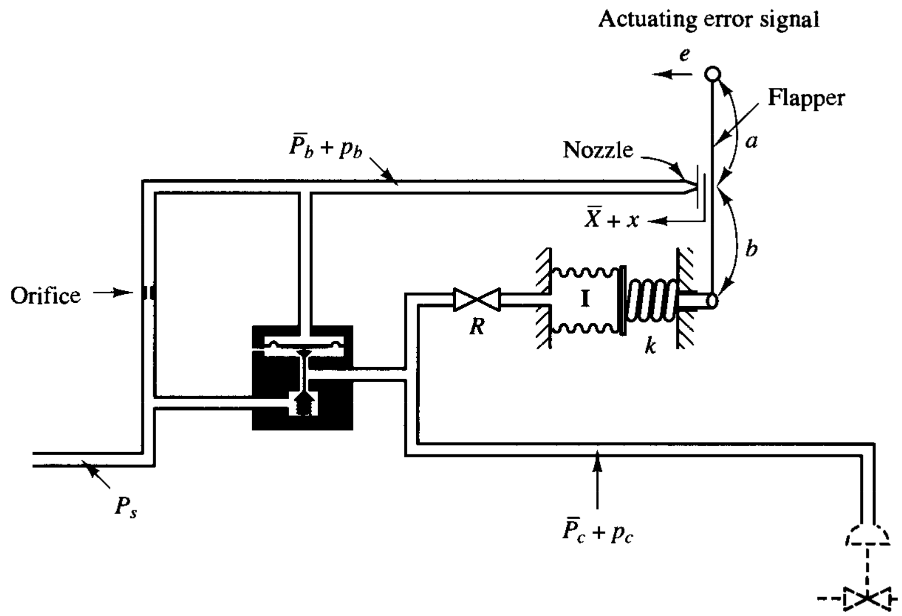



**Figure 5–86** Pneumatic controller.

fer function between $y$ and $x$, where $x$ is the air pressure and $y$ is the displacement of the power piston.

**B-5-16.** Figure 5–92 is a schematic diagram of an aircraft elevator control system. The input to the system is the deflection angle $\theta$ of the control lever, and the output is the elevator angle $\phi$. Assume that angles $\theta$ and $\phi$ are relatively small. Show that for each angle $\theta$ of the control lever there is a corresponding (steady-state) elevator angle $\phi$.

**B-5-17.** Consider the controller shown in Figure 5–93. The input is the air pressure $p_i$ and the output is the displacement $y$ of the power piston. Obtain the transfer function $Y(s)/P_i(s)$.

**B-5-18.** Obtain the transfer function $E_o(s)/E_i(s)$ of the op-amp circuit shown in Figure 5–94.

**B-5-19.** Obtain the transfer function $E_o(s)/E_i(s)$ of the op-amp circuit shown in Figure 5–95.

**B-5-20.** Consider a unity-feedback control system with the closed-loop transfer function

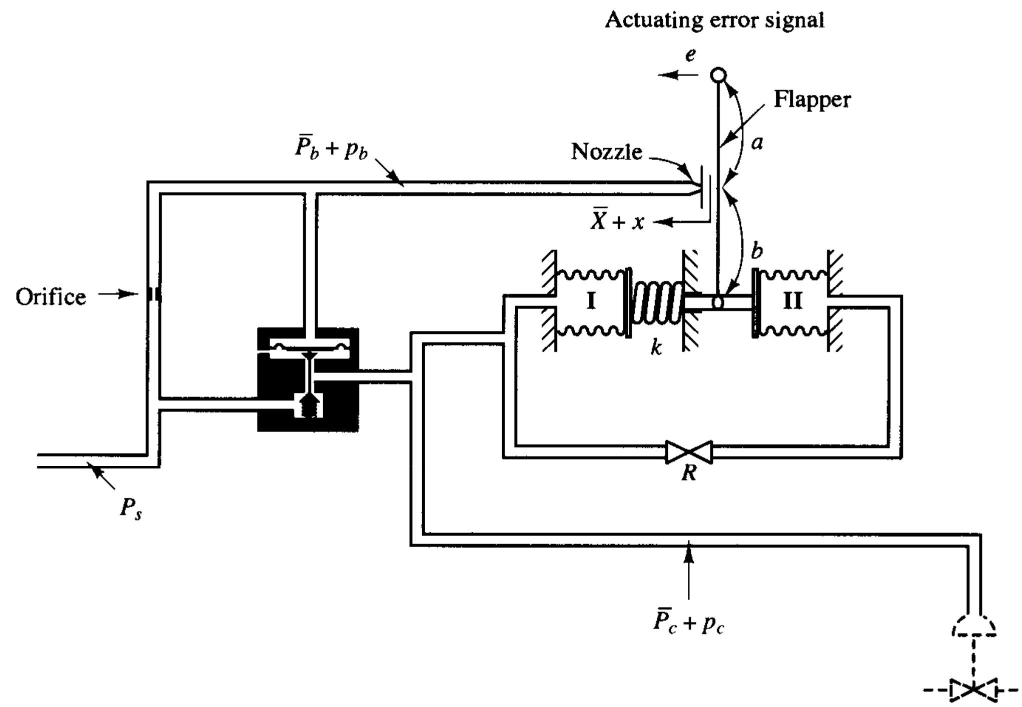



**Figure 5–87** Pneumatic controller.

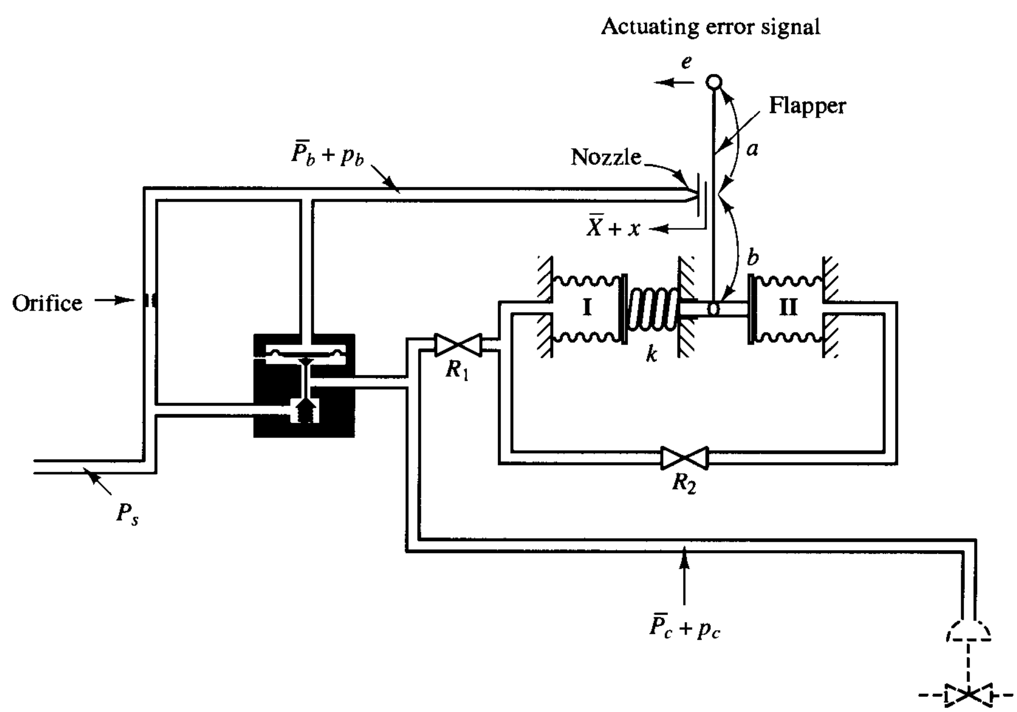



**Figure 5–88** Pneumatic controller.

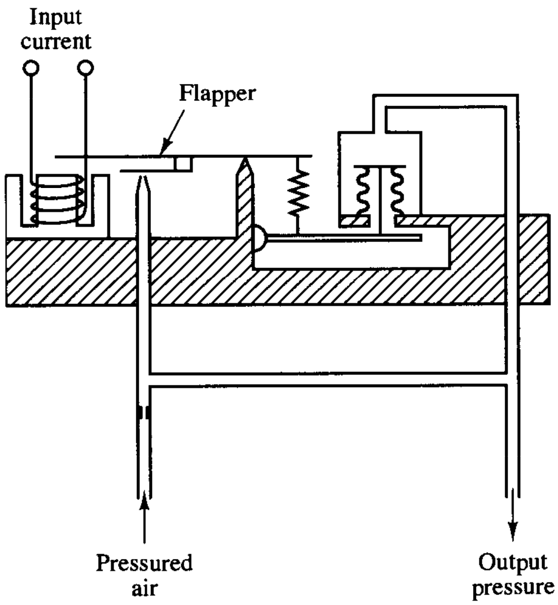



**Figure 5–89**
Electric-pneumatic transducer.

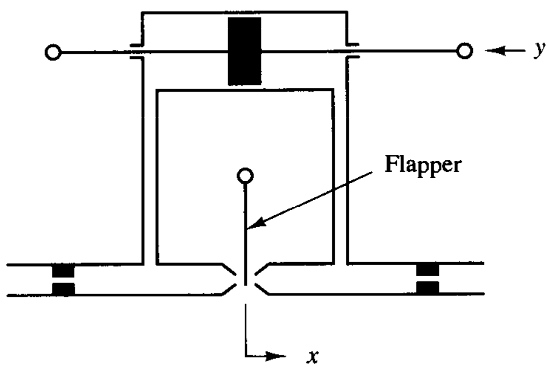



**Figure 5–90**
Flapper valve.

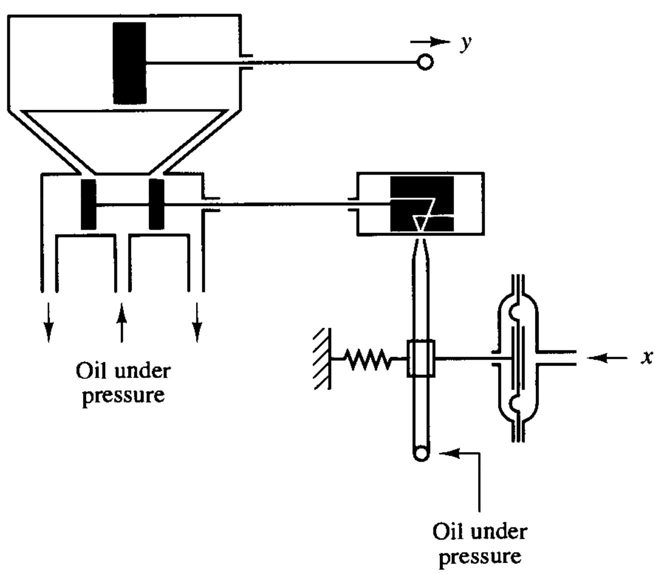



**Figure 5–91**
Schematic diagram of a hydraulic servomotor.

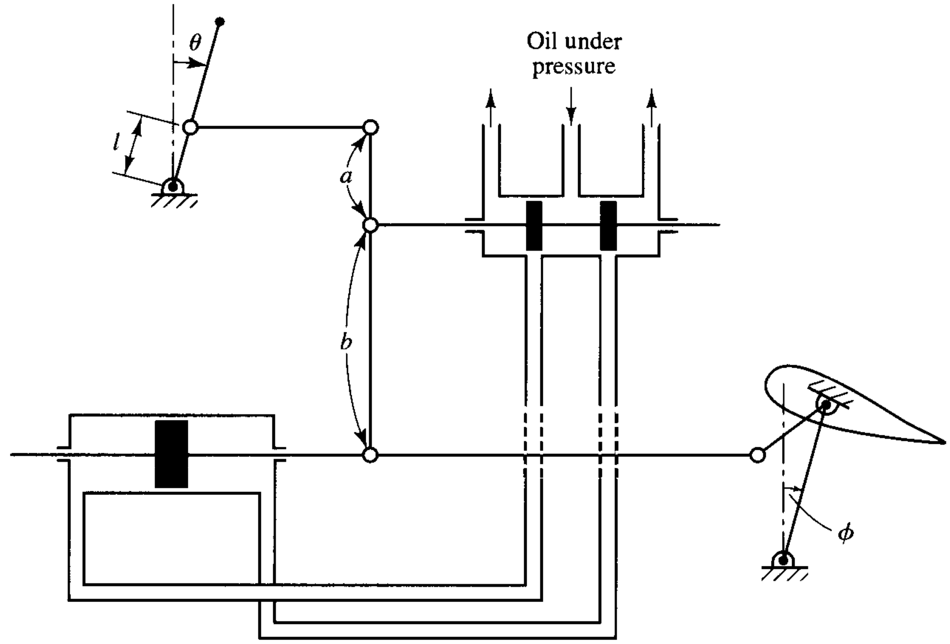



**Figure 5–92**
Aircraft elevator control system.

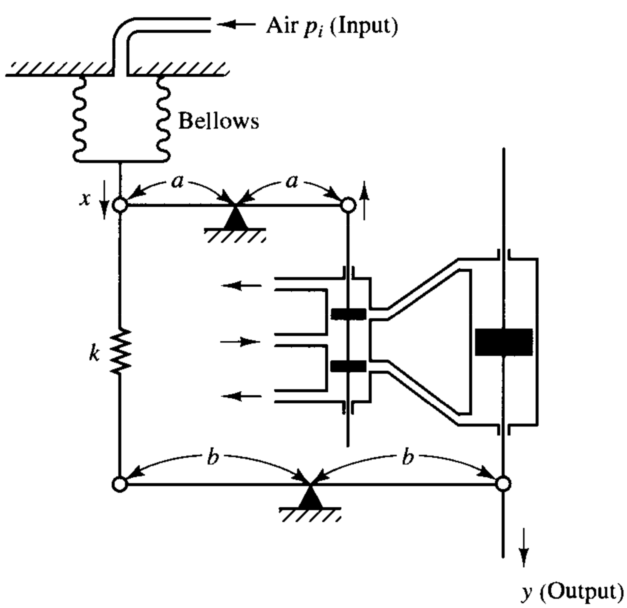



**Figure 5–93**
Controller.

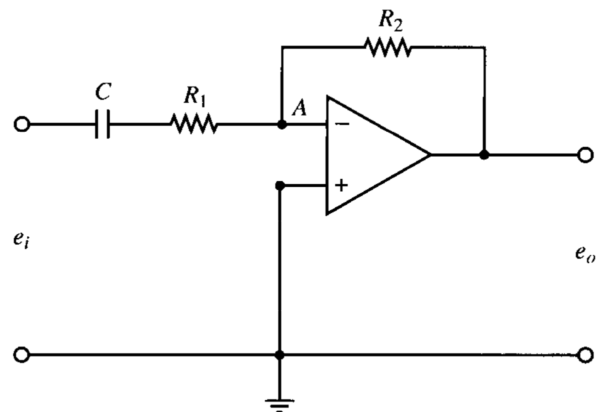



**Figure 5–94**
Operational-amplifier circuit.

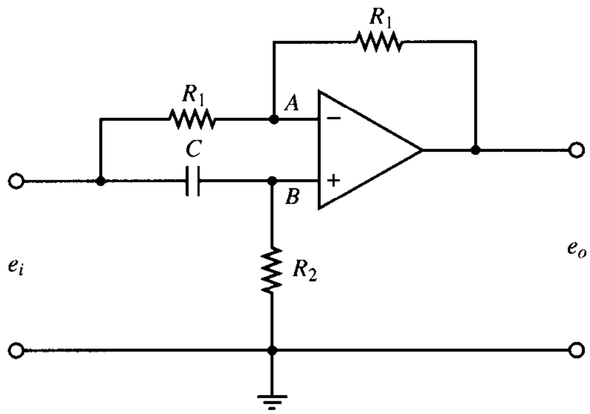



**Figure 5–95**
Operational amplifier circuit.

$$\frac{C(s)}{R(s)} = \frac{Ks + b}{s^2 + as + b}$$

Determine the open-loop transfer function $G(s)$.

Show that the steady-state error in the unit-ramp response is given by

$$e_{ss} = \frac{1}{K_v} = \frac{a - K}{b}$$

**B-5-21.** Show that the steady-state error in the response to ramp inputs can be made zero if the closed-loop transfer function is given by

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