Process Systems Analysis and Control

Third Edition

Donald R. Coughanowr

Steven E. LeBlanc

PROCESS SYSTEMS ANALYSIS
AND CONTROL

McGraw-Hill Chemical Engineering Series

The Founding of a Discipline:
The McGraw-Hill Companies, Inc. Series in Chemical Engineering

Over 80 years ago, 15 prominent chemical engineers met in New York to plan a continuing literature for their rapidly growing profession. From industry came such pioneer practitioners as Leo H. Baekeland, Arthur D. Little, Charles L. Reese, John V. N. Dorr, M. C. Whitaker, and R. S. McBride. From the universities came such eminent educators as William H. Walker, Alfred H. White, D. D. Jackson, J. H. James, Warren K. Lewis, and Harry A. Curtis. H. C. Parmlee, then editor of *Chemical and Metallurgical Engineering,* served as chairman and was joined subsequently by S. D. Kirkpatrick as consulting editor.

After several meetings, this committee submitted its report to the McGraw-Hill Book Company in September 1925. In the report were detailed specifications for a correlated series of more than a dozen texts and reference books which became the McGraw-Hill Series in Chemical Engineering—and in turn became the cornerstone of the chemical engineering curricula.

From this beginning, a series of texts has evolved, surpassing the scope and longevity envisioned by the founding Editorial Board. The McGraw-Hill Series in Chemical Engineering stands as a unique historical record of the development of chemical engineering education and practice. In the series one finds milestones of the subject's evolution: industrial chemistry, stoichiometry, unit operations and processes, thermodynamics, kinetics, and transfer operations.

Textbooks such as McCabe et al., *Unit Operations of Chemical Engineering,* Smith et al., *Introduction to Chemical Engineering Thermodynamics,* and Peters et al., *Plant Design and Economics for Chemical Engineers* have taught to generations of students the principles that are key to success in chemical engineering.

Chemical engineering is a dynamic profession, and its literature continues to grow. McGraw-Hill, with its in-house editors and consulting editors Eduardo Glandt (Dean, University of Pennsylvania), Michael Klein (Dean, Rutgers University), and Thomas Edgar (Professor, University of Texas at Austin), remains committed to a publishing policy that will serve the needs of the global chemical engineering profession throughout the years to come.

PROCESS SYSTEMS ANALYSIS AND CONTROL

THIRD EDITION

Steven E. LeBlanc

Chemical Engineering University of Toledo

Donald R. Coughanowr

Emeritus Professor, Chemical Engineering Drexel University

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 0 DOC/DOC 0 9 8

www.mhhe.com

Dedication

For Molly, my children, and grandchildren . . .

CONTENTS

PREFACE TO THE THIRD EDITION

It has been over 17 years since the second edition of this book was published. The second edition, which was written by Dr. Donald R. Coughanowr in 1991, contained many changes and new topics to bring the book up to date at the time of publication. The third edition has been a number of years in the making. I would like to thank Dr. Coughanowr for the opportunity to work on this project and help update this excellent work, which he first published in 1965 with Dr. Lowell B. Koppel. As an undergraduate, I learned process control from the first edition of this text over 30 years ago. It was an excellent book then, and it still is. I've used a number of other books over the years as a student and as a professor, but I kept coming back to this one. I felt that it was the best book to learn from. Is it an all-encompassing, totally comprehensive process dynamics and control book? No, but it is not intended to be. It is a clearly written book that is geared toward students in a first process dynamics and control course. Many control books on the market contain more material than one could ever hope to cover in a standard undergraduate semester-long class. They can be overwhelming and difficult to learn from. I have always felt that one of the strengths of this book, from both the student's and professor's point of view, was the relatively short, easy-to-read chapters that can be covered in one to two lectures. An additional strength of this text has been its unique ability to be a teaching and learning text. I hope that in this current revision, I have been able to retain that style and flavor, while introducing some new material and examples to update the text.

OBJECTIVES AND USES OF THE TEXT

This text is intended for use in an introductory one-semester-long undergraduate process dynamics and control course. It is intended to be not a comprehensive treatise on process control, but rather a textbook that provides students with the tools to learn the basic material and be in a position to continue their studies in the area if they so choose. Students are expected to have a background in mathematics through differential equations, material and energy balance concepts, and unit operations. After the first 13 chapters, the instructor may select from the remaining chapters to fit a course of particular duration and scope. A typical one-semester 15-week course, for example, may include Chapters 1 through 19 and 26.

Features of the third edition

- A capsule summary of the important points at the end of each chapter
- Restructuring of the initial chapters to reduce the impression that students frequently have regarding control classes—that this is just another mathematics course disguised as an engineering course
- Integration of MATLAB,® Simulink,® and Excel throughout the text:
  - To reduce the tedium of solving problems so that students may concentrate more on the concepts of dynamics and control and not get bogged down in the mathematical complexities of each problem

- To give students the tools to be able to ask (and more easily answer) "what if . . .?" type of questions
- To allow students to explore more difficult problems than would otherwise be possible in the time available for the course material

ACKNOWLEDGMENTS

We would like to thank the following reviewers of the third edition for their helpful comments and suggestions: Thomas F. Edgar, University of Texas–Austin; John Erjavec, University of North Dakota; Duane Johnson, University of Alabama; Costas Maranas, Penn State University; Michael Nikolaou, University of Houston; F. Joseph Schork, Georgia Institute of Technology; Delmar Timm, University of Nebraska; and William A. Weigand, University of Maryland. We especially acknowledge the helpful suggestions from Susan Montgomery of the University of Michigan and thank her for her thoroughness and useful comments to help make the text more student-friendly.

I would like to thank McGraw-Hill for having confidence in this project and providing the opportunity to revise and update the text. Special thanks go to Lorraine Buczek, Developmental Editor, and Melissa Leick, Project Manager, for their help in the final stages of this revision.

I would also like to thank my students and my University of Toledo colleague Sasidhar Varanasi for his help in using manuscript drafts when he taught the Process Control course to "field-test" the revisions. I am also grateful to my friend and colleague Dean Nagi Naganathan, of the College of Engineering at the University of Toledo, for his general support and his willingness to allow me the time required to complete this work. I especially want to thank my wife, Molly, for her love and continuing encouragement and support over the course of the writing and revising.

Dr. Steven E. LeBlanc
University of Toledo

RESOURCES FOR INSTRUCTORS AND STUDENTS:

For instructors, the solutions to the end-of-chapter problems are available at the text's website: www.mhhe.com/coughanowr-leblanc

ELECTRONIC TEXTBOOK OPTIONS

This text is offered through CourseSmart for both instructors and students. CourseSmart is an online browser where students can purchase access to this and other McGraw-Hill textbooks in a digital format. Through their browser, students can access the complete text online for one year at almost half the cost of a traditional text. Purchasing the eTextbook also allows students to take advantage of CourseSmart's Web tools for learning, which include full text search, notes and highlighting, and e-mail tools for sharing notes between classmates. To learn more about CourseSmart options, contact your sales representative or visit www.CourseSmart.com.

HISTORY OF PROCESS SYSTEMS ANALYSIS AND CONTROL (FROM THE SECOND EDITION PREFACE)

Since the first edition of this book was published in 1965, many changes have taken place in process control. Nearly all undergraduate students in chemical engineering are now required to take a course in process dynamics and control. The purpose of this book is to take the student from the basic mathematics to a variety of design applications in a clear, concise manner.

The most significant change since the first edition is the use of the digital computer in complex problem solving and in process control instrumentation. However, the fundamentals of process control, which remain the same, must be acquired before one can appreciate the advanced topics of control.

In its present form, this book represents a major revision of the first edition. The material for this book evolved from courses taught at Purdue University and Drexel University. The first 17 chapters on fundamentals are quite close to the first 20 chapters of the first edition. The remaining 18 chapters contain many new topics, which were considered very advanced when the first edition was published.

Knowledge of calculus, unit operations, and complex numbers is presumed on the part of the student. In certain later chapters, more advanced mathematical preparation is useful. Some examples would include partial differential equations in Chap. 21, linear algebra in Chaps. 28 through 30, and Fourier series in Chap. 33.

Analog computation and pneumatic controllers in the first edition have been replaced by digital computation and microprocessor-based controllers in Chaps. 34 and 35. The student should be assigned material from these chapters at the appropriate time in the development of the fundamentals. For example, the transient response for a system containing a transport lag can be obtained easily only with the use of computer simulation of the transport lag. Some of the software now available for solving control problems should be available to the student; such software is described in Chap. 34. To understand the operation of modern microprocessor-based controllers, the student should have hands-on experience with these instruments in a laboratory.

Chapter 1 is intended to meet one of the problems consistently faced in presenting this material to chemical engineering students, that is, one of perspective. The methods of analysis used in the control area are so different from the previous experiences of students that the material comes to be regarded as a sequence of special mathematical techniques, rather than an integrated design approach to a class of real and practically significant industrial problems. Therefore, this chapter presents an overall, albeit superficial, look at a simple control system design problem. The body of the text covers the following topics: Laplace transforms, Chaps. 2. to 4; transfer functions and responses of open-loop systems, Chaps. 5 to 8; basic techniques of closed-loop control, Chaps. 9 to 13; stability, Chap. 14; root locus methods, Chap. 15; frequency response methods and design, Chaps. 16 and 17; advanced control strategies (cascade, feedforward, Smith predictor, internal model control), Chap. 18; controller tuning and

process identification, Chap. 19; control valves, Chap. 20; advanced process dynamics, Chap. 21; sampled-data control, Chaps. 22 to 27; state-space methods and multivariable control, Chaps. 28 to 30; nonlinear control, Chaps. 31 to 33; digital computer simulation, Chap. 34; microprocessor-based controllers, Chap. 35.

It has been my experience that the book covers sufficient material for a one-semester (15-week) undergraduate course and an elective undergraduate course or part of a graduate course. In a lecture course meeting 3 hours per week during a 10-week term, I have covered the following chapters: 1 to 10, 12 to 14, 16, 17, 20, 34, and 35.

After the first 14 chapters, the instructor may select the remaining chapters to fit a course of particular duration and scope. The chapters on the more advanced topics are written in a logical order; however, some can be skipped without creating a gap in understanding.

I gratefully acknowledge the support and encouragement of the Drexel University Department of Chemical Engineering for fostering the evolution of this text in its curriculum and for providing clerical staff and supplies for several editions of class notes. I want to acknowledge Dr. Lowell B. Koppel's important contribution as coauthor of the first edition of this book. I also want to thank my colleague Dr. Rajakannu Mutharasan for his most helpful discussions and suggestions and for his sharing of some of the new problems. For her assistance in typing, I want to thank Dorothy Porter. Helpful suggestions were also provided by Drexel students, in particular Russell Anderson, Joseph Hahn, and Barbara Hayden. I also want to thank my wife Effie for helping me check the page proofs by reading to me the manuscript, the subject matter of which is far removed from her specialty of Greek and Latin.

McGraw-Hill and I would like to thank Ali Cinar, Illinois Institute of Technology; Joshua S. Dranoff, Northwestern University; H. R. Heichelheim, Texas Tech University; and James H. McMicking, Wayne State University, for their many helpful comments and suggestions in reviewing this second edition.

Dr. Donald R. Coughanowr

ABOUT THE AUTHORS

Steven E. LeBlanc is Associate Dean for Academic Affairs and professor of chemical engineering at the University of Toledo. He received a B.S. degree in chemical engineering from the University of Toledo and his M.S. and Ph.D. in chemical engineering from the University of Michigan. He joined the faculty at the University of Toledo in 1980. He served as the department chair for the Department of Chemical and Environmental Engineering from 1993 to 2003, when he became an Associate Dean in the College of Engineering.

Dr. LeBlanc's industrial experience includes power plant process system design and review for Toledo Edison Company (now a division of First Energy). He has taught the Process Dynamics and Control course numerous times, and was responsible for a major revamp of laboratory activities associated with the course.

He is a member of the American Institute of Chemical Engineers (AIChE) and the American Society for Engineering Education (ASEE). He has served as an ABET chemical engineering program evaluator for AIChE since 1998. He chaired the national ASEE Chemical Engineering Education Division and cochaired the 2007 ASEE Chemical Engineering Summer School for Faculty. He coauthored and judged the 1992 AIChE Senior Design Project competition. He is also coauthor of a textbook on *Strategies for Creative Problem Solving* with H. Scott Fogler of the University of Michigan.

Donald R. Coughanowr is Emeritus Professor of Chemical Engineering at Drexel University. In 1991 he wrote the second edition of *Process Systems Analysis and Control* which contained many changes and new topics in order to bring the book up to date at the time of publication. He received a Ph.D. in chemical engineering from the University of Illinois in 1956, an M.S. degree in chemical engineering from the University of Pennsylvania in 1951, and a B.S. degree in chemical engineering from the Rose-Hulman Institute of Technology in 1949. He joined the faculty at Drexel University in 1967 as department head, a position he held until 1988. Before going to Drexel, he was a faculty member of the School of Chemical Engineering at Purdue University for 11 years.

At Drexel and Purdue he taught a wide variety of courses, which include material and energy balances, thermodynamics, unit operations, transport phenomena, petroleum refinery engineering, environmental engineering, chemical engineering laboratory, applied mathematics, and process dynamics and control. At Purdue,

he developed a new course and laboratory in process control and collaborated with Dr. Lowell B. Koppel on the writing of the first edition of *Process Systems Analysis and Control.*

His research interests included environmental engineering, diffusion with chemical reaction, and process dynamics and control. Much of his research in control emphasized the development and evaluation of new control algorithms for the processes that cannot be controlled easily by conventional control; some of the areas investigated were time-optimal control, adaptive pH control, direct digital control, and batch control of fermentors. He reported on his research in numerous publications and received support for research projects from the National Science Foundation and industry. He spent sabbatical leaves teaching and writing at Case-Western Reserve University, the Swiss Federal Institute, the University of Canterbury, the University of New South Wales, the University of Queensland, and Lehigh University.

Dr. Coughanowr's industrial experience included process design and pilot plant at Standard Oil Co. (Indiana) and summer employment at Electronic Associates and Dow Chemical Company.

He is a member of the American Institute of Chemical Engineers. He has served the AIChE by participating in accreditation visits to departments of chemical engineering for ABET and by chairing sessions of the Department Heads Forum at the annual meetings of AIChE.

INTRODUCTORY CONCEPTS

In this chapter we examine the concept of chemical process control and introduce several examples to illustrate the necessity for process modeling as we begin our study of process dynamics and control.

1.1 WHY PROCESS CONTROL?

As competition becomes stiffer in the chemical marketplace and processes become more complicated to operate, it is advantageous to make use of some form of automatic control. Automatic control of a process offers many advantages, including

- Enhanced process safety
- Satisfying environmental constraints
- Meeting ever-stricter product quality specifications
- More efficient use of raw materials and energy
- Increased profitability

Considering all the benefits that can be realized through process control, it is well worth the time and effort required to become familiar with the concepts and practices used in the field.

1.2 CONTROL SYSTEMS

Control systems are used to maintain process conditions at their desired values by manipulating certain process variables to adjust the variables of interest. A common example of a control system from everyday life is the cruise control on an automobile. The purpose of a cruise control is to maintain the speed of the vehicle (the controlled variable) at the desired value (the set point) despite variations in terrain, hills, etc.

(disturbances) by adjusting the throttle, or the fuel flow to the engine (the manipulated variable). Another common example is the home hot water heater. The control system on the hot water heater attempts to maintain the temperature in the tank at the desired value by manipulating the fuel flow to the burner (for a gas heater) or the electrical input to the heater in the face of disturbances such as the varying demand on the heater early in the morning, as it is called upon to provide water for the daily showers. A third example is the home thermostat. This control system is designed to maintain the temperature in the home at a comfortable value by manipulating the fuel flow or electrical input to the furnace. The furnace control system must deal with a variety of disturbances to maintain temperature in the house, such as heat losses, doors being opened and hopefully closed, and leaky inefficient windows. The furnace must also be able to respond to a request to raise the desired temperature if necessary. For example, we might desire to raise the temperature by 5°, and we'd like the system to respond smoothly and efficiently. From these examples, we can deduce that there are several common attributes of control systems:

- The ablity to maintain the process variable at its desired value in spite of disturbances that might be experienced (this is termed *disturbance rejection*)
- The ability to move the process variable from one setting to a new desired setting (this is termed *set point tracking*)

Conceptually we can view the control systems we've discussed in the following general manner (Fig. 1–1).

The controller compares the measurement signal of the controlled variable to the set point (the desired value of the controlled variable). The difference between the two values is called the *error*.

$$\text{Error} = (\text{Set point value}) - (\text{Measurement signal of controlled variable})$$

Depending upon the magnitude and sign of the error, the controller takes appropriate action by sending a signal to the final control element, which provides an input to the process to return the controlled variable to the set point. The concept of using information
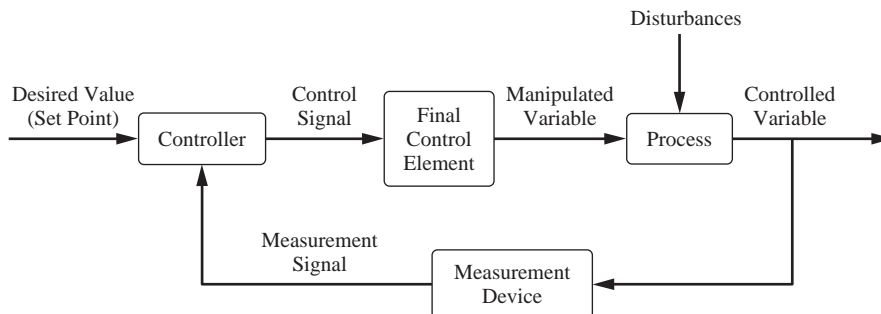
FIGURE 1–1
Generalized process control system.

about the deviation of the system from its desired state to control the system is called *feedback* control. Information about the state of the system is "fed back" to a controller, which utilizes this information to change the system in some way.

The type of control system shown in Fig. 1–1 is termed a *closed-loop* feedback control system. *Closed-loop* refers to the fact that the controller automatically acts to return the controlled variable to its desired value. In contrast, an *open-loop* system would have the measurement signal disconnected from the controller, and the controller output would have to be manually adjusted to change the value of the controlled variable. An open-loop system is sometimes said to be in manual mode as opposed to automatic mode (closed-loop). Negative feedback is the most common type of signal feedback. *Negative* refers to the fact that the error signal is computed from the difference between the set point and the measured signal. The *negative* value of the measured signal is "fed back" to the controller and added to the set point to compute the *error.*

Example 1.1.   Hot water tank control system. As a specific example, let us consider a hot water heater for a home (Fig. 1–2) and examine its control system, using the same type of diagram (Fig. 1–3).

The desired hot water temperature is selected by the homeowner, and typically it is in the neighborhood of 120 to 140°F. Let us assume that the set point is 130°F. The thermocouple measures the temperature of the water in the tank and sends a signal to the thermostat indicating the temperature. The thermostat (controller) determines the error as

$$\text{Error} = T_{\text{set point}} - T_{\text{measured}}$$

If the error is positive (> 0), the measured temperature is lower than desired and the thermostat opens the fuel valve to the burner which adds heat to the tank. If the error is zero or negative (≤ 0), the thermostat closes the fuel valve and no heat is added to the tank. Disturbances to the system, which decrease the temperature of the water in the tank, include ambient heat losses and hot water demand by the household which is replaced with a cold water feed.

FIGURE 1–2
Physical drawing of a hot water heater.

Types of Controllers

The thermostat on the hot water heater is called an "on/off" type of controller. Depending on the value of the error signal, the output from the controller is
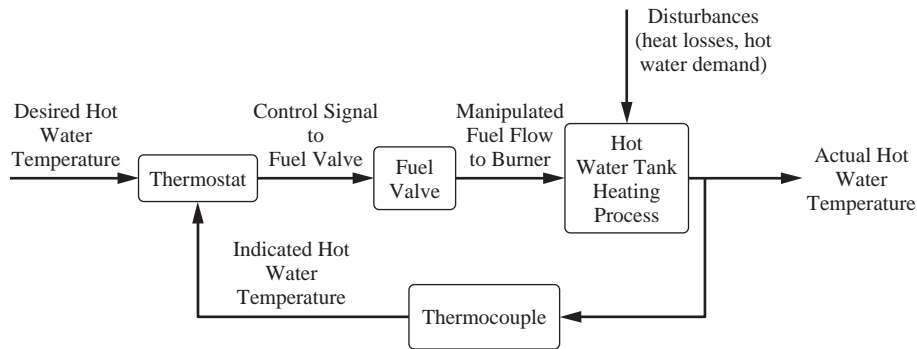
FIGURE 1–3
Block diagram of a hot water heater control system.

either "full on" or "full off" and the fuel valve is full open or full closed; there are no intermediate values of the output. Many other types of controllers that we will study can modulate their output based on the magnitude of the error signal, how long the error signal has persisted, and even how rapidly the error appears to be changing.

Clearly, the larger the error, the less we are satisfied with the present state of affairs and vice versa. In fact, we are completely satisfied only when the error is exactly zero. Based on these considerations, it is natural to suggest that the controller should change the heat input by an amount *proportional* to the error. This is called *proportional control.* In effect, the controller is instructed to maintain the heat input at the steady-state design value as long as the error is zero. If the tank temperature deviates from the set point, causing an error, the controller is to use the magnitude of the error to change the heat input proportionally. We shall reserve the right to vary the proportionality constant to suit our needs. This degree of freedom forms a part of our instructions to the controller. As we will see shortly during the course of our studies, the larger we make the proportionality constant for the proportional controller (called the controller gain), the smaller the steady-state error will become. We will also see that it is impossible to completely eliminate the error through the use of a proportional controller. For example, if the set point is 130°F and a disturbance occurs that drops the temperature to 120°F, if we use only a proportional controller, then we will never be able to get the tank temperature to *exactly* 130°F. Once the sytem stabilizes again, the temperature will not be exactly 130°F, but perhaps 127°F or 133°F. There will always be some residual steady-state error (called *offset*). For a home water heater, this is probably good enough; the exact temperature is not that critical. In an industrial process, this may not be adequate, and we have to resort to a bit more complicated controller to drive the error to zero.

Considerable improvement may be obtained over proportional control by adding integral control. The controller is now instructed to change the heat input by an additional amount proportional to the time integral of the error. This type of control system has two adjustable parameters: a multiplier for the error and a multiplier for the integral of the error. If this type of controller is used, the steady-state error will be zero. From this standpoint, the response is clearly superior to that of the system with proportional control only. One price we pay for this improvement is the tendency for the system to be more

oscillatory. The system will tend to overshoot its final steady-state value before slowly settling out at the desired set point. So what is the best control system to use for a particular application? This and related questions will be addressed in subsequent chapters.

Some Further Complications

At this point, it would appear that the problem has been solved in some sense. A little further probing will shatter this illusion.

It has been assumed that the controller receives instantaneous information about the tank temperature. From a physical standpoint, some measuring device such as a thermocouple will be required to measure this temperature. The temperature of a thermocouple inserted in the tank may or may not be the same as the temperature of the fluid in the tank. This can be demonstrated by placing a mercury thermometer in a beaker of hot water. The thermometer does not instantaneously rise to the water temperature. Rather, it takes a bit of time to respond. Since the controller will receive measured values of the temperature, rather than the actual values, it will be acting upon the *apparent error,* rather than the actual error. The effect of the thermocouple delay in transmission of the temperature to the controller is primarily to make the response of the system somewhat more oscillatory than if the response were instantaneous. If we increase the controller gain (the proportionality constants), the tank temperature will eventually oscillate with *increasing* amplitude and will continue to do so until the physical limitations of the heating system are reached. In this case, the control system has actually caused a deterioration in performance, and this type of reponse is referred to as an *unstable response.*

This problem of *stability* of response will be a major concern for obvious reasons. At present, it is sufficient to note that extreme care must be exercised in specifying control systems. In the case considered, the proportional and integral controllers described above will perform satisfactorily if the gain is kept lower than some particular value. However, it is not difficult to construct examples of systems for which the addition of *any* amount of integral control will cause an unstable response. Since integral control usually has the desirable feature of eliminating steady-state error, it is extremely important that we develop means for predicting the occurrence of unstable response in the design of any control system.

Block Diagram

A good overall picture of the relationships among variables in the heated-tank control system may be obtained by preparing a *block diagram* as shown in Fig. 1–1. It indicates the flow of information around the control system and the function of each part of the system. Much more will be said about block diagrams later, but the reader can undoubtedly form a good intuitive notion about them by comparing Fig. 1–1 with the physical description of the process. Particularly significant is the fact that each component of the system is represented by a block, with little regard for the actual physical characteristics of the represented component (e.g., the tank or controller). The major interest is in (1) the relationship between the signals entering and leaving the block and (2) the manner in which information flows around the system.

SUMMARY

We have taken an overall look at a typical control problem and some of its ramifications. At present, the reader has been asked to accept the results on faith and to concentrate on obtaining a physical understanding of the transient behavior of the heated tank. In the forthcoming chapters we develop tools for determining the response of such systems. As this new material is presented, the reader may find it helpful to refer to this chapter to place the material in proper perspective to the overall control problem.

PROBLEMS

1.1. Draw a block diagram for the control system generated when a human being steers an automobile.

1.2. Draw a block diagram for the control system generated when a human being shoots a bow and arrow.

1.3. Draw a block diagram for an automobile cruise control system.

1.4. Draw a block diagram for the control system that maintains the water level in a toilet tank.

1.5. Draw a block diagram for a security lighting system that activates at dusk and turns off at dawn.

1.6. Draw a block diagram for the control system for a home oven.

CAPSULE SUMMARY

DEFINITIONS

Block diagram—Diagram that indicates the flow of information around the control system and the function of each part of the system.

Closed loop—In closed loop, the measured value of the controlled variable is fed back to the controller.

Controlled variable—The process variable that we want to maintain at a particular value.

Controller—A device that outputs a signal to the process based on the magnitude of the error signal. A *proportional* controller outputs a signal proportional to the error.

Disturbance rejection—One goal of a control system, which is to enable the system to "reject" the effect of disturbance changes changes and maintain the controlled variable at the set point.

Disturbances—Any process variables that can cause the controlled variable to change. In general, disturbances are variables that we have no control over.

Error—The difference between the values of the set point and the measured variable.

Manipulated variable—Process variable that is adjusted to bring the controlled variable back to the set point.

Negative feedback—In negative feedback, the error is the difference between the set point and the measured variable (this is usually the desired configuration).

Offset—The steady-state value of the error.

Open loop—In open loop, the measured value of the controlled variable is not fed back to the controller.

Positive feedback—In positive feedback, the measured temperature is added to the set point. (This is usually an undesirable situation and frequently leads to instability.)

Set point—The desired value of the controlled variable.

Set point tracking—One goal of a control system, which is to force the system to follow or "track" requested set point changes.

PART
I

MODELING FOR
PROCESS DYNAMICS

MODELING TOOLS FOR PROCESS DYNAMICS

U nderstanding process dynamics (how process variables change with time) will be very important to our studies of process control. In the examples in Chap.1, we saw some of the implications of process dynamics and their relationship to process control. In this chapter we explore process dynamics further and review some mathematical tools for solving the resulting process models.

2.1 PROCESS DYNAMICS—A CHEMICAL MIXING SCENARIO

Consider the following chemical mixing example (Fig. 2–1). Two process streams are mixed to produce one of the feeds for our chemical reactor. After mixing, the blended stream is fed to a heating vessel before being sent to the reactor.

Stream 3
C_{a3}=? g/L
Mixing tee v_3=30 L/min

Stream 1
C_{a1}=1 g/L
v_1=10 L/min

V=150 L

Stream 2
C_{a2}=4 g/L
v_2=20 L/min

C_a

To reactor

Heater
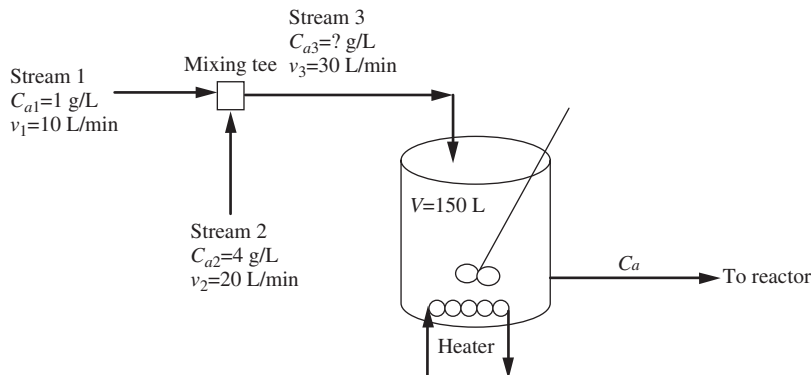
FIGURE 2–1
Chemical mixing process flow diagram.

The process is running along at steady state. The concentration of A in stream 1 is 1 g/L and in stream 2 is 4 g/L. At 3:00 P.M. the shift changes at the plant. The new operator on our unit misreads the flowmeters for the process and switches the flow rates of the two streams. Stream 1 is switched to 20 L/min, and stream 2 is switched to 10 L/min. At 3:30 P.M. the shift supervisor hurries to the control room to determine the source of the problem now being experiencing with the reactor. Use your knowledge of chemical engineering to determine what has happened to the exit concentration from the heating vessel over the first half-hour of the shift.

We can model the mixing tee and the blending tank using an unsteady-state mass balance to predict the behavior of this part of the process since the shift change and the unfortunate error by the new operator.

A balance on component A around the mixing tee before and after the change will yield information on how the feed concentration to the heating vessel changes. The component A balance around the mixing tee is

$$\begin{pmatrix} \text{Rate of } A \text{ into} \\ \text{mixing tee} \\ \text{in stream 1 (g/min)} \end{pmatrix} + \begin{pmatrix} \text{Rate of } A \text{ into} \\ \text{mixing tee} \\ \text{in stream 2 (g/min)} \end{pmatrix} = \begin{pmatrix} \text{Rate of } A \text{ leaving} \\ \text{mixing tee} \\ \text{in stream 3 (g/min)} \end{pmatrix}$$

$$v_1 C_{a1} + v_2 C_{a2} = v_3 C_{a3}$$

Before the change, we can calculate the original steady-state concentration into the heating vessel:

$$\left(10\ \frac{\text{L}}{\text{min}} \right)\left(1\ \frac{\text{g}}{\text{L}} \right) + \left(20\ \frac{\text{L}}{\text{min}} \right)\left(4\ \frac{\text{g}}{\text{L}} \right) = \left(30\ \frac{\text{L}}{\text{min}} \right)\left(C_{a3}\ \frac{\text{g}}{\text{L}} \right)$$

$$C_{a3} = 3\ \frac{\text{g}}{\text{L}}$$

Since the process has been running along at steady state for a long time under these conditions, the concentration in the heating vessel is also the same.

After the change, the new feed concentration to the heating vessel is

$$\left(20\ \frac{\text{L}}{\text{min}} \right)\left(1\ \frac{\text{g}}{\text{L}} \right) + \left(10\ \frac{\text{L}}{\text{min}} \right)\left(4\ \frac{\text{g}}{\text{L}} \right) = \left(30\ \frac{\text{L}}{\text{min}} \right)\left(C_{a3}\ \frac{\text{g}}{\text{L}} \right)$$

$$C_{a3} = 2\ \frac{\text{g}}{\text{L}}$$

So the net result of the operator error is to decrease the feed concentration to the heating vessel from 3 to 2 g/L. After we analyze the process for a moment, it is apparent that the exit concentration from the heating vessel will eventually also fall from 3 to 2 g/L if the process is left in its current configuration for a long enough time. So, what we know of the situation is shown in Fig. 2–2.

To analyze how the exit from the heating vessel (the feed to the reactor) varies with time, we must perform an unsteady mass balance on component A around the heating vessel.
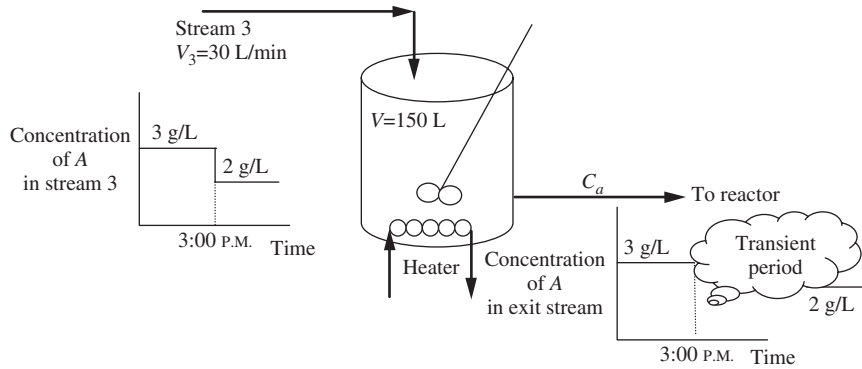
FIGURE 2–2

Chemical mixing process flow diagram showing operator-induced transient.

$$\underbrace{\left(\begin{array}{c}\text{Rate of } A \text{ into}\\\text{heating vessel (g/min)}\end{array}\right)}_{\text{in}} - \underbrace{\left(\begin{array}{c}\text{Rate of } A \text{ leaving}\\\text{heating vessel (g/min)}\end{array}\right)}_{\text{out}} = \underbrace{\left(\begin{array}{c}\text{Accumulation of } A \text{ in}\\\text{heating vessel (g/min)}\end{array}\right)}_{\text{accumulation}}$$

$$v_3 C_{a3} - v_3 C_a = \frac{d}{dt}(V C_a)$$

Note that the volumetric flow rate v is constant into and out of the heating vessel at v_3. Thus the volume of fluid in the tank V is constant. We can rearrange this equation to the following form:

$$\underbrace{\frac{V}{v_3}}_{\tau} \frac{dC_a}{dt} + C_a = C_{a3} \qquad (2.1)$$

The coefficient of the derivative term is the residence time of the heating vessel τ, which in this process is 5 min. Substituting the numbers for this scenario yields

$$5\frac{dC_a}{dt} + C_a = 2 \qquad C_a(0) = 3\frac{g}{L}$$

We can rearrange and solve this equation as follows.

$$\int_3^{C_a} \frac{dC_a}{2 - C_a} = \int_0^t \frac{1}{5} dt$$

$$-\ln(2 - C_a)\big|_3^{C_a} = \frac{t}{5}$$

$$\ln\left(\frac{2 - C_a}{2 - 3}\right) = \frac{-t}{5}$$

$$C_a = 2 + e^{-t/5} \qquad (2.2)$$

A plot of the exit concentration from the heating vessel is shown in Fig. 2–3. As expected, the concentration starts at the original steady-state concentration of 3 g/L and exponentially decreases to 2 g/L.

Modeling the mixing process enables us to determine the concentration of component A in the stream being fed to the reactor. Being able to determine or predict the dynamic behavior of a process is crucial to being able to design a control system for it.

As another modeling example, consider the energy balance for the mixing process described above. Prior to 3 P.M. the process conditions are depicted as in Fig. 2–4.

Stream 1 (at 25°C) mixes with stream 2 (at 55°C), producing stream 3, the feed to the heating vessel. The heater adds energy to the vessel to bring the outlet stream to 80°C. Before we look at the effect of the disturbance caused by the operator, it is necessary to determine the steady-state process conditions prior to the upset. An energy balance around the mixing tee will enable us to calculate the steady-state feed temperature to the heating vessel T_3.

$$
\begin{pmatrix}
\text{Rate of} \\
\text{enthalpy into} \\
\text{mixing tee} \\
\text{with stream 1}
\end{pmatrix}
+
\begin{pmatrix}
\text{Rate of} \\
\text{enthalpy into} \\
\text{mixing tee} \\
\text{with stream 2}
\end{pmatrix}
=
\begin{pmatrix}
\text{Rate of} \\
\text{enthalpy leaving} \\
\text{mixing tee} \\
\text{with stream 3}
\end{pmatrix}
$$

$$
\rho v_1 C_p (T_1 - T_{\text{ref}}) + \rho v_2 C_p (T_2 - T_{\text{ref}}) = \rho v_3 C_p (T_3 - T_{\text{ref}})
$$

We have assumed that the stream density ρ (g/L) and specific heat Cp [cal/(g · °C)] remain constant, independent of the concentration of component A in the stream. We

Heating Vessel Exit Concentration
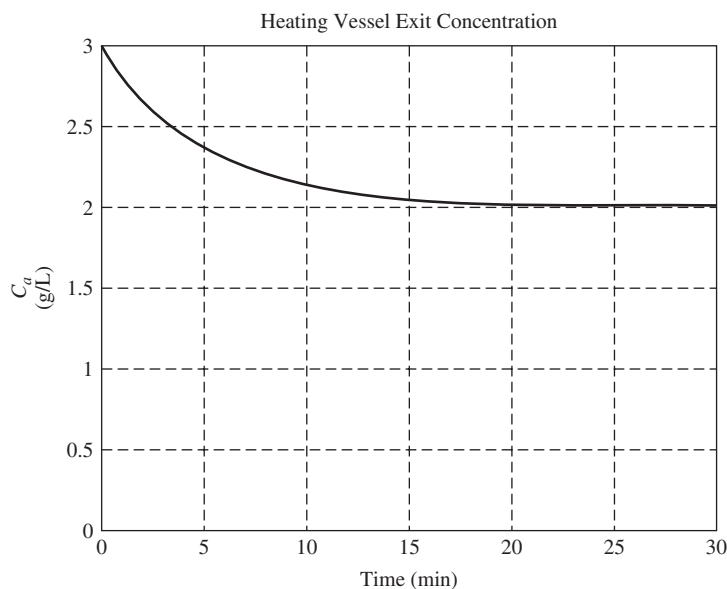
FIGURE 2–3
Outlet concentration from the heating vessel as a function of time.
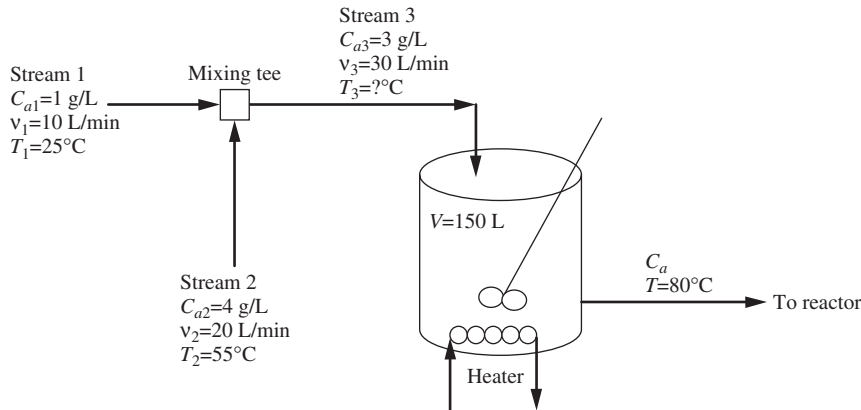
FIGURE 2–4
Chemical mixing process flow diagram showing initial temperatures.

have also defined a reference temperature T_{ref} for the enthalpy calculation. The energy balance can be simplified to

$$v_1 T_1 + v_2 T_2 = v_3 T_3$$

Note that we have made use of the relation $(v_1 + v_2)T_{\text{ref}} = v_3 T_{\text{ref}}$ to eliminate some terms. Solving for T_3 yields

$$T_3 = \frac{v_1 T_1 + v_2 T_2}{v_3} = \frac{(10)(25) + (20)(55)}{30} = 45°C$$

So, the steady-state inlet temperature to the heating vessel is 45°C. We can now determine the steady-state heat input required from the heater by performing a steady-state energy balance around the heating vessel.

$$\begin{pmatrix} \text{Rate of} \\ \text{enthalpy into} \\ \text{heating vessel} \\ \text{with stream 3} \end{pmatrix} + \begin{pmatrix} \text{Rate of} \\ \text{enthalpy into} \\ \text{heating vessel} \\ \text{from heater} \end{pmatrix} = \begin{pmatrix} \text{Rate of} \\ \text{enthalpy leaving} \\ \text{heating vessel} \end{pmatrix}$$

$$\rho v_3 C_p (T_3 - T_{\text{ref}}) + \qquad Q \qquad = \qquad \rho v_3 C_p (T - T_{\text{ref}})$$

Solving for the heater input gives

$$Q = \rho v_3 C_p (T - T_3) = \left(1000 \ \frac{\text{g}}{\text{L}} \right)\left(30 \ \frac{\text{L}}{\text{min}} \right)\left(1 \ \frac{\text{cal}}{\text{g} \cdot °\text{C}} \right)(80°\text{C} - 45°\text{C})$$

$$= 1.05 \times 10^6 \ \frac{\text{cal}}{\text{min}} = 73.2 \ \text{kW}$$

The energy balance for the original steady-state case is summarized in Fig. 2–5.

The inlet temperature to the heating vessel after the 3:00 P.M. disturbance can be determined from the steady-state energy balance around the mixing tee using the new flow rates (Fig. 2–6).
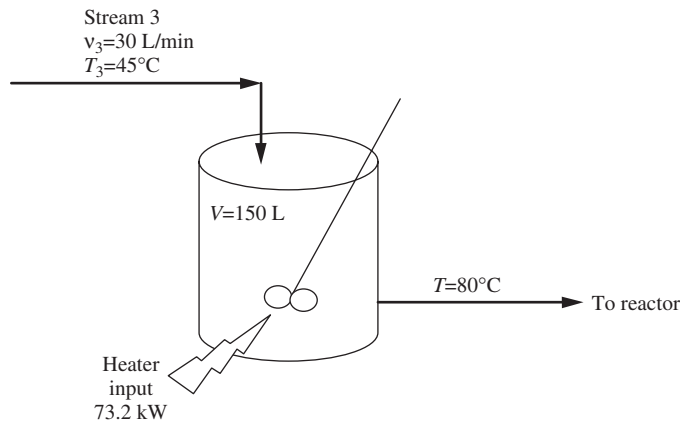
FIGURE 2–5
Chemical mixing process flow diagram summarizing initial temperatures.

$$T_3 = \frac{v_1 T_1 + v_2 T_2}{v_3} = \frac{(20)(25) + (10)(55)}{30} = 35°C$$
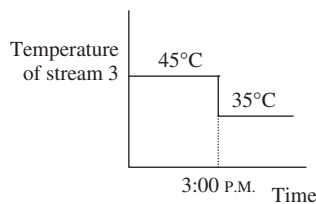
FIGURE 2–6
Inlet temperature disturbance.

To determine the effect of this inlet temperature disturbance on the feed to the reactor, an unsteady-state energy balance on the heating vessel is required. We can logically predict that if the heat input stays constant, as well as the process flow rate, when the inlet temperature falls by 10°C, the outlet temperature from the heating vessel will correspondingly decrease by 10° to 70°C. The energy balance on the heating vessel is

$$\begin{pmatrix} \text{Rate of enthalpy} \\ \text{into heating} \\ \text{vessel} \\ \text{with stream 3} \end{pmatrix} - \begin{pmatrix} \text{Rate of enthalpy} \\ \text{leaving heating} \\ \text{vessel} \\ \text{with exit stream} \end{pmatrix} + \begin{pmatrix} \text{Rate of energy} \\ \text{input to} \\ \text{heating vessel} \\ \text{from heater} \end{pmatrix} = \begin{pmatrix} \text{Rate of} \\ \text{accumulation} \\ \text{of enthalpy in} \\ \text{heating vessel} \end{pmatrix}$$

$$\rho v_3 C_p (T_3 - T_{\text{ref}}) \quad - \quad \rho v_3 C_p (T - T_{\text{ref}}) \quad + \quad Q \quad = \quad \frac{d}{dt}(\rho V C_p (T - T_{\text{ref}}))$$

Simplifying yields

$$\underbrace{\left(\frac{V}{v_3}\right)}_{\tau} \frac{dT}{dt} + T = T_3 + \frac{1}{\rho v_3 C_p} Q \tag{2.3}$$

Substituting values for the scenario we are considering gives

$$(5 \text{ min}) \frac{dT}{dt} + T = 35°C + \frac{1}{\left(1000 \, \frac{g}{L}\right)\left(30 \, \frac{L}{\min}\right)\left(1 \, \frac{cal}{g \cdot °C}\right)}\left(1.05 \times 10^6 \, \frac{cal}{\min}\right) = 70°C$$

$$T(0) = 80°C$$

Separating and integrating, we have

$$5\frac{dT}{dt} = 70 - T$$

$$\int_{80}^{T} \frac{dT}{70 - T} = \int_{0}^{t} \frac{dt}{5}$$

$$-\ln\left(\frac{70 - T}{-10}\right) = \frac{t}{5}$$

$$T = 70 + 10e^{-t/5} \tag{2.4}$$

A plot of the outlet temperature from the heating vessel as a function of time is shown in Fig. 2–7.

Outlet Temperature from Heating Vessel as a Function of Time
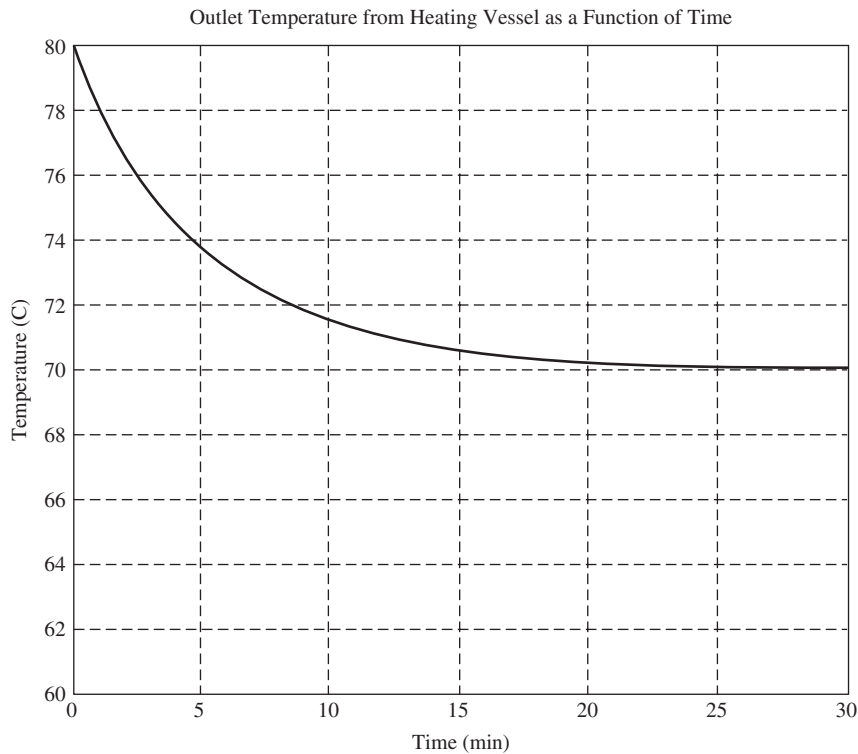
FIGURE 2–7
Outlet temperature transient due to the disturbance.

Notice the shape of the temperature response is the same as the shape of the concentration response that we saw previously. By appropriate modeling of the process, we can predict how the system will respond to changes in the operating conditions. Our ability to model the process will be extremely valuable as we design controllers to automatically control the process variables at their desired settings.

2.2   MATHEMATICAL TOOLS FOR MODELING

As we just saw in our analysis of the chemical mixer, the unsteady-state material and energy balance models that we wrote required us to solve differential equations to obtain the concentration and temperature versus time behavior for the process. This will be a common occurrence for us as we continue our studies of process dynamics and control. It would be beneficial to review some additional tools available to us for solving our process models. In Sec. 2.1, we solved the equations by separation and integration. A couple of other useful tools for solving such models are Laplace transforms and MATLAB/Simulink. In the next several sections, we will review the use of these additional tools for solving our model differential equations.

Definition of the Laplace Transform

The Laplace transform of a function $f(t)$ is *defined* to be $F(s)$ according to the equation

$$F(s) = \int_0^\infty f(t)e^{-st}\,dt \tag{2.5}$$

We often abbreviate this to

$$F(s) = L\{f(t)\}$$

where the operator L is defined by Eq. (2.5).

Example 2.1.   Find the Laplace transform of the function

$$f(t) = 1$$

According to Eq. (2.5),

$$F(s) = \int_0^\infty (1)e^{-st}\,dt = -\frac{e^{-st}}{s}\bigg|_{t=0}^{t=\infty} = \frac{1}{s}$$

Thus,

$$L\{1\} = \frac{1}{s}$$

There are several facts worth noting at this point:

1.   The Laplace transform $F(s)$ contains no information about the behavior of $f(t)$ for $t, 0$. This is not a limitation for control system study because t will represent the

time variable and we will be interested in the behavior of systems only for positive time. In fact, the variables and systems are usually defined so that $f(t)$ 0 for t , 0. The time we designate as $t = 0$ is arbitrary. We shall generally define $t = 0$ as the time when the process is disturbed from steady state (i.e., when an input is changed). Our usual starting point will be a steady-state system or process, and we will be interested in examining what happens when the system is disturbed. This will become clearer as we study specific examples.

2. Since the Laplace transform is defined in Eq. (2.5) by an improper integral, it will not exist for every function $f(t)$. A rigorous definition of the class of functions possessing Laplace transforms is beyond the scope of this book, but readers will note that every function of interest to us *does* satisfy the requirements for possession of a transform [see Churchill (1972)].

3. The Laplace transform is linear. In mathematical notation, this means

$$L\{af_1(t) + bf_2(t)\} = aL\{f_1(t)\} + bL\{f_2(t)\}$$

where a and b are constants and f_1 and f_2 are two functions of t.

Proof. Using the definition, we have

$$L\{af_1(t) + bf_2(t)\} = \int_0^\infty [af_1(t) + bf_2(t)]e^{-st}\,dt$$

$$= a\int_0^\infty f_1(t)e^{-st}\,dt + b\int_0^\infty f_2(t)e^{-st}\,dt$$

$$= aL\{f_1(t)\} + bL\{f_2(t)\}$$

4. The Laplace transform operator transforms a function of the variable t to a function of the variable s. The t variable is eliminated by the integration.

Transforms of Simple Functions

We now proceed to derive the transforms of some simple and useful functions. We shall see these common functions repeatedly during our future studies.

1. **The *step function* is**

$$f(t) = \begin{cases} 0 & t < 0 \\ 1 & t > 0 \end{cases}$$

This important function is known as the unit-step function and will henceforth be denoted by $u(t)$. From Example 2.1, it is clear that

$$L\{u(t)\} = \frac{1}{s}$$

As expected, the behavior of the function for $t < 0$ has no effect on its Laplace transform. Note that as a consequence of linearity, the transform of any constant A, that is, $f(t) = Au(t)$, is just $F(s) = A/s$. Notice in the chemical mixing example

that we just discussed that the inlet concentration and temperature are described by a step function initiated at time zero (3 P.M. in the example).

2. **The *exponential function* is**

$$f(t) = \begin{cases} 0 & t < 0 \\ e^{-at} & t > 0 \end{cases} = u(t)e^{-at}$$

where $u(t)$ is the unit-step function. Again proceeding according to definition, we have

$$L\left\{u(t)e^{-at}\right\} = \int_0^\infty e^{-(s+a)t}\, dt = -\frac{1}{s+a}e^{-(s+a)t}\bigg|_0^\infty = \frac{1}{s+a}$$

provided that $s + a > 0$, that is, $s > -a$. In this case, the convergence of the integral depends on a suitable choice of s. In case s is a complex number, it may be shown that this condition becomes

$$\mathrm{Re}(s) > -a$$

For problems of interest to us it will always be possible to choose s so that these conditions are satisfied, and the reader uninterested in mathematical niceties can ignore this point.

3. **The *ramp function* is**

$$f(t) = \begin{cases} 0 & t < 0 \\ t & t > 0 \end{cases} = tu(t)$$

$$L\{tu(t)\} = \int_0^\infty te^{-st}\, dt$$

Integration by parts yields

$$L\{tu(t)\} = -e^{-st}\left(\frac{t}{s} + \frac{1}{s^2}\right)\bigg|_0^\infty = \frac{1}{s^2}$$

4. **The *sine function* is**

$$f(t) = \begin{cases} 0 & t < 0 \\ \sin kt & t > 0 \end{cases} = u(t)\sin kt$$

$$L\{u(t)\sin kt\} = \int_0^\infty \sin kt\, e^{-st}\, dt$$

Integrating by parts, we have

$$L\{u(t)\sin kt\} = \frac{-e^{-st}}{s^2 + k^2}(s\,\sin kt + k\,\cos kt)\bigg|_0^\infty$$

$$= \frac{k}{s^2 + k^2}$$
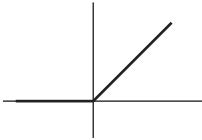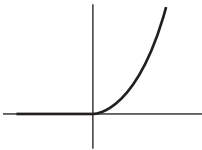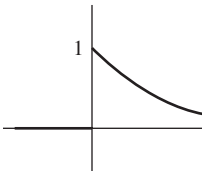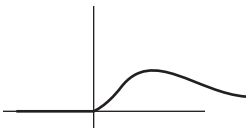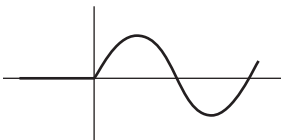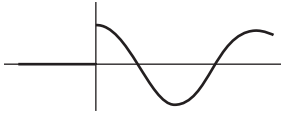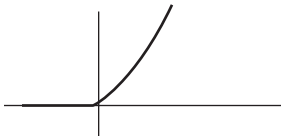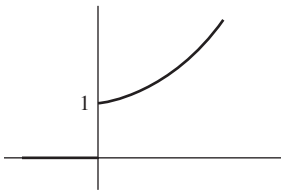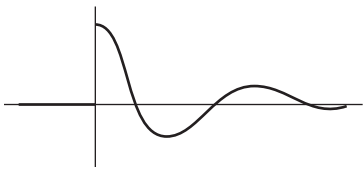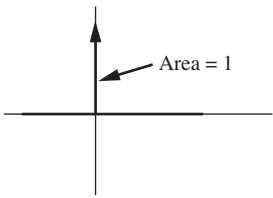
TABLE 2.1

Function	Graph	Transform
$u(t)$		$\dfrac{1}{s}$
$tu(t)$		$\dfrac{1}{s^2}$
$t^n u(t)$		$\dfrac{n!}{s^{n+1}}$
$e^{-at}u(t)$		$\dfrac{1}{s+a}$
$t^n e^{-at}u(t)$		$\dfrac{n!}{(s+a)^{n+1}}$
$\sin kt\, u(t)$		$\dfrac{k}{s^2+k^2}$

TABLE 2.1 *(Continued)*

Function	Graph	Transform
$\cos kt\, u(t)$		$\dfrac{s}{s^2 + k^2}$
$\sinh kt\, u(t)$		$\dfrac{k}{s^2 - k^2}$
$\cosh kt\, u(t)$	1	$\dfrac{s}{s^2 - k^2}$
$e^{-at}\sin kt\, u(t)$		$\dfrac{k}{(s + a)^2 + k^2}$
$e^{-at}\cos kt\, u(t)$		$\dfrac{s + a}{(s + a)^2 + k^2}$
$\delta(t)$, unit impulse	Area = 1	1

In a like manner, the transforms of other simple functions may be derived. Table 2.1 is a summary of transforms that will be of use to us. Those which have not been derived here can be easily established by direct integration, except for the transform of $\delta(t)$, which will be discussed in detail in the Appendix at the end of Chap. 3.

Using MATLAB for Symbolic Processing—Laplace Transforms

MATLAB is capable of symbolic processing. To prepare MATLAB for symbolic operations, some variable names will be declared symbolic (rather than numeric) using the `syms` command.

```
syms a x y z t k s
```

We can also define *u* as the Heaviside function (the unit step):

```
u=sym('Heaviside(t)')
u=
Heaviside(t)
```

Now we can determine the transform of the simple functions we have just discussed:
The step function:

```
laplace(u)
ans=
1/s
```

The exponential function:

```
laplace(exp(-a*t))
ans=
1/(s+a)
```

The ramp function:

```
laplace(t)
ans=
1/s^2
```

The sine function:

```
laplace(sin(k*t))
ans=
k/(s^2+k^2)
```

Transforms of Derivatives

At this point, the reader may wonder what has been gained by introduction of the Laplace transform. The transform merely changes a function of *t* into a function of *s.* The functions of *s* look no simpler than those of *t* and, as in the case of $A \rightarrow A/s,$ may actually be more complex. In the next few paragraphs, the motivation will become clear. It will be shown that the Laplace transform has the remarkable property of transforming

the operation of differentiation with respect to t to that of multiplication by s. Thus, we claim that

$$L\left\{\frac{df(t)}{dt}\right\} = sF(s) - f(0) \qquad (2.6)$$

where

$$F(s) = L\{f(t)\}$$

and $f(0)$ is $f(t)$ evaluated at $t = 0$. If $f(t)$ is discontinuous at $t = 0$, $f(0)$ should be evaluated at $t = 0^+$, that is, just to the right of the origin. Since we will seldom want to differentiate functions that are discontinuous at the origin, this detail is not of great importance. However, the reader is cautioned to watch carefully for situations in which such discontinuities occur.

Proof

$$L\left\{\frac{df(t)}{dt}\right\} = \int_0^\infty \frac{df}{dt} e^{-st}\, dt$$

To integrate this by parts, let

$$u = e^{-st} \qquad dv = \frac{df}{dt}\, dt$$

Then

$$du = -se^{-st}\, dt \qquad v = f(t)$$

Since

$$\int u\, dv = uv - \int v\, du$$

we have

$$\int_0^\infty \frac{df}{dt} e^{-st}\, dt = f(t)e^{-st}\Big|_0^\infty + s\int_0^\infty f(t)e^{-st}\, dt = -f(0) + sF(s)$$

The salient feature of this transformation is that whereas the function of t was to be differentiated with respect to t, the corresponding function of s is merely multiplied by s. We shall find this feature to be extremely useful in the solution of differential equations.

To find the transform of the second derivative we make use of the transform of the first derivative twice, as follows:

$$L\left\{\frac{d^2 f}{dt^2}\right\} = L\left\{\frac{d}{dt}\left(\frac{df}{dt}\right)\right\} = sL\left\{\frac{df}{dt}\right\} - \frac{df(t)}{dt}\bigg|_{t=0}$$

$$= s[sF(s) - f(0)] - f'(0)$$

$$= s^2 F(s) - sf(0) - f'(0)$$

where we have abbreviated

$$\frac{df(t)}{dt}\bigg|_{t=0} = f'(0)$$

In a similar manner, the reader can easily establish by induction that repeated application of Eq. (2.6) leads to

$$L\left\{\frac{d^n f}{dt^n}\right\} = s^n F(s) - s^{n-1} f(0) - s^{n-2} f^{(1)}(0) - \cdots - s f^{(n-2)}(0) - f^{(n-1)}(0)$$

where $f^i(0)$ indicates the ith derivative of $f(t)$ with respect to t, evaluated for $t = 0$.

Thus, the Laplace transform may be seen to change the operation of differentiation of the function to that of multiplication of the transform by s, the number of multiplications corresponding to the number of differentiations. In addition, some polynomial terms involving the initial values of $f(t)$ and its first $n - 1$ derivatives are involved. In later applications we usually define our variables so that these polynomial terms will vanish. Hence, they are of secondary concern here.

Example 2.2.   Find the Laplace transform of the function $x(t)$ that satisfies the differential equation and initial conditions

$$\frac{d^3 x}{dt^3} + 4\frac{d^2 x}{dt^2} + 5\frac{dx}{dt} + 2x = 2$$

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

It is permissible mathematically to take the Laplace transforms of both sides of a differential equation and equate them, since equality of functions implies equality of their transforms. Doing this, we obtain

$$s^3 x(s) - s^2 x(0) - sx'(0) - x''(0) + 4[s^2 x(s) - sx(0) - x'(0)]$$

$$+ 5[sx(s) - x(0)] + 2x(s) = \frac{2}{s}$$

where $x(s) = L\{x(t)\}$. Use has been made of the linearity property and of the fact that only positive values of t are of interest. Inserting the initial conditions and solving for $x(s)$, we have

$$x(s) = \frac{2}{s\left(s^3 + 4s^2 + 5s + 2\right)} \tag{2.7}$$

This is the required answer, the Laplace transform of $x(t)$.

2.3 SOLUTION OF ORDINARY DIFFERENTIAL EQUATIONS (ODEs)

There are two important points to note regarding this last example. First, application of the transformation resulted in an equation that was solved for the unknown function by *purely algebraic means.* Second, and most important, if the function $x(t)$, which has the Laplace transform $2/s(s^3 + 4s^2 + 5s + 2)$, were known, we would have the solution to the differential equation and initial conditions. This suggests a procedure for solving differential equations that is analogous to that of using logarithms to multiply or divide. To use logarithms, one transforms the pertinent numbers to their logarithms and then adds or subtracts, which is much easier than multiplying or dividing. The result of the addition or subtraction is the logarithm of the desired answer. The answer is found by reference to a table to find the number having this logarithm.

In the Laplace transform method for solution of differential equations, the functions are converted to their transforms, and the resulting equations are *algebraically* solved for the unknown function. This is much easier than solving a differential equation. However, at the last step the analogy to logarithms is not complete. We obviously cannot hope to construct a table containing the Laplace transform of every function $f(t)$ that possesses a transform. Instead, we will develop methods for expressing complicated transforms, such as $x(s)$ in Example 2.2, in terms of simple transforms that can be found in Table 2.1. For example, it is easily verified that the solution to the differential equation and initial conditions of Example 2.2 is

$$x(t) = 1 - 2te^{-t} - e^{-2t} \tag{2.8}$$

The Laplace transform of x, using Eq. (2.8) and Table 2.1, is

$$x(s) = \frac{1}{s} - 2\frac{1}{(s + 1)^2} - \frac{1}{s + 2} \tag{2.9}$$

Equation (2.7) is actually the result of placing Eq. (2.9) over a common denominator. Although it is difficult to find $x(t)$ from Eq. (2.7), Eq. (2.9) may be easily inverted to Eq. (2.8) by using Table 2.1. Therefore, what is required is a method for expanding the common-denominator form of Eq. (2.7) to the separated form of Eq. (2.9). This method is provided by the technique of partial fractions, which is developed in Chap. 3.

Using MATLAB for Symbolic Processing—Inversion of Laplace Transforms

Remember that we have previously declared some variables symbolic (a, k, x, y, z, t and s). Let's have MATLAB invert Eq. (2.3) for us and determine $x(t)$.

```
x=ilaplace(2/s/(s^3+4*s^2+5*s+2))
x=
1-exp(-2*t)-2*t*exp(-t)
```

which is the same as Eq. (2.8).

Using MATLAB for Symbolic Processing—Solution of Differential Equations

MATLAB can solve differential equations symbolically using the DSOLVE command. The derivatives are represented as Dx (first derivative), D2x (second derivative), etc.

```
x=dsolve('D3x+4*D2x+5*Dx+2*x=2','x(0)=0','Dx(0)=0','D2x(0)=0')
x=
1-exp(-2*t)-2*t*exp(-t)
```

which, again, is the same as Eq. (2.8).

Now, let's return to our chemical mixing scenario from earlier in the chapter.

Example 2.3.   Transform the differential equations resulting from the mass and energy balances for the chemical mixer to determine the transform of the exit concentration and temperature.

$$\tau \frac{dC_a}{dt} + C_a = C_{a3} \qquad \text{mass balance}$$

$$\tau \frac{dT}{dt} + T = T_3 + \left(\frac{1}{\rho v_3 C_p}\right) Q \qquad \text{energy balance}$$

For the mass balance,

$$\tau[sC_a(s) - C_a(0)] + C_a(s) = C_{a3}(s)$$

Rearranging and solving for $C_a(s)$, we have

$$C_a(s) = \frac{C_{a3}(s) + \tau C_a(0)}{\tau s + 1}$$

After the disturbance, C_{a3} has a constant value of 2 g/L. Therefore, $C_{a3}(s) = 2/s$. Also, from the process description, we know that $C_a(0) = 3$ g/L [this is the initial concentration of A in the tank at time 0 (3 P.M.)] and that τ, the time constant, is 5 min. Substituting these values into the expression for C_a yields

$$C_a(s) = \frac{2/s + 5(3)}{5s + 1} = \frac{2}{s(5s + 1)} + \frac{15}{5s + 1} \qquad (2.10)$$

For the energy balance,

$$\tau[sT(s) - T(0)] + T(s) = T_3(s) + \frac{1}{\rho v_3 C_p} Q(s)$$

Rearranging and solving for $T(s)$, we find

$$T(s) = \frac{T_3(s) + \left[1/(\rho v_3 C_p)\right]Q(s) + \tau T(0)}{\tau s + 1}$$

After the disturbance, T_3 is constant at 35°C, so $T_3(s) = 35/s$. The initial temperature in the tank $T(0)$, at 3 P.M. is 80°C. The heater input Q is constant at 1.05×10^6 cal/min. The time constant τ is 5 min. Substituting these values into the expression for $T(s)$ gives

$$T(s) = \frac{\dfrac{35}{s} + \left(\dfrac{1}{\left(1000\,\dfrac{g}{L}\right)\left(30\,\dfrac{L}{\text{min}}\right)\left(1\,\dfrac{\text{cal}}{g\cdot°\text{C}}\right)}\right)\dfrac{\left(1.05 \times 10^6\,\dfrac{\text{cal}}{\text{min}}\right)}{s} + 5(80)}{5s + 1}$$

Simplifying yields

$$T(s) = \frac{70/s + 5(80)}{5s + 1} = \frac{70}{s(5s + 1)} + \frac{400}{5s + 1} \tag{2.11}$$

If we can invert these expressions for $C_a(s)$, Eq. (2.10), and $T(s)$, Eq. (2.11), we will obtain the time domain solutions for the exiting concentration and temperature. We will address this topic in Chap. 3.

To summarize, we have reviewed a procedure using Laplace transforms for solving *linear, ordinary, differential equations (ODEs) with constant coefficients.*
The procedure is as follows:

1. Take the Laplace transform of both sides of the equation. The initial conditions are incorporated at this step in the transforms of the derivatives.
2. Solve the resulting equation for the Laplace transform of the unknown function algebraically.
3. Find the function of t that has the Laplace transform obtained in step 2. This function satisfies the differential equation and initial conditions and hence is the desired solution. This third step is frequently the most difficult or tedious step and will be developed further in Chap. 3. It is called inversion of the transform. Although there are other techniques available for inversion, the one that we will develop and make consistent use of is that of partial fraction expansion.

A simple example will serve to illustrate steps 1 and 2 (we'll save step 3 until Chap. 3).

Example 2.4. Solve

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

$$x(0) = 2$$

We number our steps according to the discussion in the preceding paragraphs:

1. $\underbrace{[sx(s) - 2]}_{sx(s) - x(0)} + 3x(s) = 0$

2. $x(s) = \dfrac{2}{s + 3} = 2\dfrac{1}{s + 3}$

3. $x(t) = 2e^{-3t}$

SUMMARY

In this chapter we discussed the importance of process modeling and worked through a chemical mixing example that led to two differential equations that described the process (one from the mass balance and one from the energy balance). We solved those relatively simple equations by separating and integrating. We also discussed using Laplace transforms for solving differential equations and presented a table of common transforms. We concluded by demonstrating the use of MATLAB for symbolically solving differential equations. In Chap. 3 we will discuss the method of partial fractions for inverting the solutions we obtained by using transforms to the time domain.

PROBLEMS

2.1. Transform the following:

   (a) $\sin(2t + \dfrac{\pi}{4})$

   (b) $e^{-t}\cos 2t$

   (c) Use the formula for the Laplace transform of a derivative to find $L\{\sinh(kt)\}$ if you are given that $L\{\cosh(kt)\} = s/(s^2 - k^2)$.

2.2. Invert the following transforms.

   (a) $\dfrac{3}{s}$

   (b) $\dfrac{3}{s + 2}$

   (c) $\dfrac{3}{(s + 2)^2}$

   (d) $\dfrac{3}{s^3}$

   (e) $\dfrac{\frac{1}{2}}{s^2 + 9}$

   (f) $\dfrac{3}{s^2 + 4s + 8}$

   (g) $\dfrac{s + 4}{s^2 + 4s + 8}$

   (h) $\dfrac{1}{(s + 2)^2}$

2.3. Find $x(s)$ for the following differential equations.

(a) $\dfrac{d^2x}{dt^2} + 4\dfrac{dx}{dt} + 3x = u(t)$ $x(0) = x'(0) = 0$

(b) $\dfrac{d^2x}{dt^2} + 2\dfrac{dx}{dt} + x = u(t)$ $x(0) = x'(0) = 1$

(c) $\dfrac{d^2x}{dt^2} + 2\dfrac{dx}{dt} + x = u(t)$ $x(0) = x'(0) = 0$

2.4. Solve Prob. 2.1 using the MATLAB `laplace` command.

2.5. Solve Prob. 2.2 using the MATLAB `ilaplace` command.

2.6. Solve Prob. 2.3 using the MATLAB `dsolve` command, and then use `ezplot` to graph the solutions.

2.7. Use the MATLAB `dsolve` command to solve the differential equations that we developed for the mass and energy balances for the chemical mixing scenario, and then use `ezplot` to graph the solutions. Compare the results with those presented in the text.

2.8. Use the MATLAB `ilaplace` command to invert Eqs. (2.10) and (2.11), and then use `ezplot` to graph the solutions. Compare the results with those presented in the text.

2.9. Rework the chemical mixing scenario if at 3 P.M. the operator mistakenly increases the flow rate of stream 1 to 20 L/min while stream 2 and the heater input remain unchanged.

CHAPTER
2

CAPSULE SUMMARY

Definition of the Laplace transform: $\quad L\{f(t)\} = f(s) = \int_0^\infty f(t)e^{-st}\,dt$

Linearity: $\quad L\{af_1(t) + bf_2(t)\} = aL\{f_1(t)\} + bL\{f_2(t)\}$

Transform of first derivative: $\quad L\left\{\dfrac{df(t)}{dt}\right\} = sf(s) - f(0)$

Transform of *nth* derivative:

$$L\left\{\frac{d^n f}{dt^n}\right\} = s^n f(s) - s^{n-1}f(0) - s^{n-2}f^{(1)}(0) - \mathrm{n} - sf^{(n-2)}(0) - f^{(n-1)}(0)$$

Transforms of some simple functions: See Table 2.1.

The following procedure uses Laplace transforms for solving *linear, ordinary, differential equations (ODEs) with constant coefficients:*

1. Take the Laplace transform of both sides of the equation. The initial conditions are incorporated at this step in the transforms of the derivatives.
2. Solve the resulting equation for the Laplace transform of the unknown function algebraically.
3. Find the function of *t* that has the Laplace transform obtained in step 2. This function satisfies the differential equation and initial conditions and hence is the desired solution. This third step is frequently the most difficult or tedious step. We will make consistent use of partial fraction expansions to accomplish this (see Chap. 3).

Useful MATLAB Commands:
`syms`—declares variables to be symbolic
`laplace`—takes the Laplace transform of a symbolic expression
`ilaplace`—inverts a symbolic Laplace transform expression
`dsolve`—solves a differential equation symbolically

CHAPTER
3

INVERSION BY PARTIAL FRACTIONS

Our study of the application of Laplace transforms to linear differential equations with constant coefficients has enabled us to rapidly establish the Laplace transform of the solution. We now wish to develop methods for inverting the transforms to obtain the solution in the time domain. In the first part of this chapter we give a series of examples that illustrate the partial fraction technique. After a generalization of these techniques, we proceed to a discussion of the qualitative information that can be obtained from the transform of the solution without inverting it.

The equations to be solved are all of the general form

$$a_n \frac{d^n x}{dt^n} + a_{n-1} \frac{d^{n-1} x}{dt^{n-1}} + \cdots + a_1 \frac{dx}{dt} + a_0 x = r(t)$$

The unknown function of time is $x(t)$, and $a_n, a_{n-1}, \ldots, a_1, a_0$ are constants. The given function $r(t)$ is called the *forcing function*. In addition, for all problems of interest in control system analysis, the initial conditions are given. In other words, values of $x, dx/dt, \ldots, d^{n-1} x/dt^{n-1}$ are specified at time 0. The problem is to determine $x(t)$ for all $t \geq 0$.

3.1 PARTIAL FRACTIONS

In the series of examples that follow, the technique of partial fraction inversion for solution of this class of differential equations is presented.

Example 3.1.   Solve

$$\frac{dx}{dt} + x = 1$$

$$x(0) = 0$$

Application of the Laplace transform yields

$$sx(s) + x(s) = \frac{1}{s}$$

or $$\qquad\qquad x(s) = \frac{1}{s(s + 1)}$$

The theory of partial fractions enables us to write this as

$$x(s) = \frac{1}{s(s + 1)} = \frac{A}{s} + \frac{B}{s + 1} \qquad\qquad (3.1)$$

where A and B are constants. Hence, from Table 2.1, it follows that

$$x(t) = A + Be^{-t} \qquad\qquad (3.2)$$

Therefore, if A and B were known, we would have the solution. The conditions on A and B are that they must be chosen to make Eq. (3.1) an identity in s.

   To determine A, multiply both sides of Eq. (3.1) by s.

$$\frac{1}{s + 1} = A + \frac{Bs}{s + 1} \qquad\qquad (3.3)$$

Since this must hold for all s, it must hold for $s = 0$. Putting $s = 0$ in Eq. (3.3) yields

$$A = 1$$

To find B, multiply both sides of Eq. (3.1) by $s + 1$.

$$\frac{1}{s} = \frac{A}{s}(s + 1) + B \qquad\qquad (3.4)$$

Since this must hold for all s, it must hold for $s = -1$. This yields

$$B = -1$$

This procedure for determining the coefficients is called the *Heaviside expansion.* There is an easy way to visualize the Heaviside procedure and quickly determine the coefficients of the partial fraction expansion (A and B in this case). Considering Eq. (3.1), we can determine A, the numerator of the $1/s$ factor, by ignoring (or "covering up") this term in the denominator of $x(s)$ and letting all the remaining s's equal the value of s that makes the "covered up" term equal to zero. The other coefficients are found in a similar manner.

For example, to solve for *A,* we "cover up" the *s* factor and let all the other *s* values equal 0.

$$A = \frac{1}{\cancel{s}(\cancel{s} + 1)} = 1$$

Similarly for *B,* we cover up the *s* + 1 term and let the other *s* values equal −1, so

$$B = \frac{1}{\underset{-1}{\cancel{s}} \; \cancel{(s + 1)}} = \frac{1}{-1} = -1$$

Cross-multiplication (as well as the quick visualization method) works for distinct roots (non-repeated factors in the denominator) and in a limited way for repeated roots. We will discuss the case of repeated roots shortly.

Now that we've found *A* and *B,* we have

$$x(s) = \frac{1}{s(s + 1)} = \frac{1}{s} - \frac{1}{s + 1} \tag{3.5}$$

and therefore,

$$x(t) = 1 - e^{-t} \tag{3.6}$$

Equation (3.5) may be checked by putting the right side over a common denominator, and Eq. (3.6) by substitution into the original differential equation and initial condition.

Example 3.2. Chemical mixing scenario revisited. In Chap. 2 we solved the chemical mixing scenario problem to the point where we had obtained the transformed solution to the material and energy balances. The transformed solutions, Eqs. (2.10) and (2.11), are repeated in this example for convenience.

$$C_a(s) = \frac{2}{s(5s + 1)} + \frac{15}{5s + 1} \tag{2.10}$$

We can now invert this expression for the concentration in the tank to the time domain.

Considering the first term on the left-hand side, we can separate it into partial fractions by using the same method that was employed in Example 3.1.

$$\frac{2}{s(5s + 1)} = \frac{\frac{2}{5}}{s\left(s + \frac{1}{5}\right)} = \frac{A}{s} + \frac{B}{s + \frac{1}{5}} = \frac{2}{s} + \frac{-2}{s + \frac{1}{5}}$$

Equation (2.10) may now be written as

$$C_a(s) = \frac{2}{s(5s + 1)} + \frac{15}{5s + 1} = \frac{2}{s} - \frac{2}{s + \frac{1}{5}} + \frac{3}{s + \frac{1}{5}} = \frac{2}{s} + \frac{1}{s + \frac{1}{5}}$$

We can now readily invert this expression to the time domain

$$C_a(t) = 2 + e^{-t/5}$$

This is the same solution that we previously obtained by separation and integration of the original mass balance differential equation in the time domain, which is plotted in Fig. 2–3.

Similarly, we can obtain the time domain solution for the temperature in the mixing vessel by inverting Eq. (2.11).

$$T(s) = \frac{70/s + 5(80)}{5s + 1} = \frac{70}{s(5s + 1)} + \frac{400}{5s + 1} \qquad (2.11)$$

Separating the right-hand side by using partial fractions, we get

$$T(s) = \frac{70}{s(5s + 1)} + \frac{400}{5s + 1} = \frac{70}{s} + \frac{-70}{s + \frac{1}{5}} + \frac{80}{s + \frac{1}{5}} = \frac{70}{s} + \frac{10}{s + \frac{1}{5}}$$

$$T(t) = 70 + 10e^{-t/5}$$

This is the same solution that we previously obtained by separation and integration of the original energy balance differential equation in the time domain which is plotted in Fig. 2–7.

Example 3.3. Solve

$$\frac{d^3x}{dt^3} + 2\frac{d^2x}{dt^2} - \frac{dx}{dt} - 2x = 4 + e^{2t}$$

$$x(0) = 1 \quad x'(0) = 0 \quad x''(0) = -1$$

Taking the Laplace transform of both sides yields

$$\left[s^3x(s) - s^2 + 1\right] + 2\left[s^2x(s) - s\right] - \left[sx(s) - 1\right] - 2x(s) = \frac{4}{s} + \frac{1}{s - 2}$$

Solving algebraically for $x(s)$, we find

$$x(s) = \frac{s^4 - 6s^2 + 9s - 8}{s(s - 2)(s^3 + 2s^2 - s - 2)} \qquad (3.7)$$

The cubic in the denominator may be factored, and $x(s)$ is expanded in partial fractions.

$$x(s) = \frac{s^4 - 6s^2 + 9s - 8}{s(s-2)(s+1)(s+2)(s-1)} = \frac{A}{s} + \frac{B}{s-2} + \frac{C}{s+1} + \frac{D}{s+2} + \frac{E}{s-1}$$

(3.8)

To find A, multiply both sides of Eq. (3.8) by s and then set $s = 0$; the result is

$$A = \frac{-8}{(-2)(1)(2)(-1)} = -2$$

The other constants are determined in the same way. The procedure and results are summarized in the following table.

To determine	Multiply Eq. (3.8) by	and set s to	Result
B	$s - 2$	2	$B = \frac{1}{12}$
C	$s + 1$	-1	$C = 1\frac{1}{3}$
D	$s + 2$	-2	$D = -\frac{17}{12}$
E	$s - 1$	1	$E = \frac{2}{3}$

Accordingly, the solution to the problem is

$$x(t) = -2 + \frac{1}{12}e^{2t} + \frac{11}{3}e^{-t} - \frac{17}{12}e^{-2t} + \frac{2}{3}e^{t}$$

A comparison between this method and the classical method, as applied to Example 3.2, may be profitable. In the classical method for solution of differential equations, we first write down the characteristic function of the homogeneous equation:

$$s^3 + 2s^2 - s - 2 = 0$$

This must be factored, as was also required in the Laplace transform method, to obtain the roots -1, -2, and $+1$. Thus, the complementary solution is

$$x_c(t) = C_1 e^{-t} + C_2 e^{-2t} + C_3 e^{t}$$

Furthermore, by inspection of the forcing function, we know that the particular solution has the form

$$x_p(t) = A + B e^{2t}$$

The constants A and B are determined by substitution into the differential equation and, as expected, are found to be -2 and $\frac{1}{12}$, respectively. Then

$$x(t) = -2 + \frac{1}{12}e^{2t} + C_1 e^{-t} + C_2 e^{-2t} + C_3 e^{t}$$

Using MATLAB for Symbolic Processing—Partial Fractions

Remember that we have previously declared some variables symbolic ($a, k, x, y, z, t,$ and s). MATLAB does not have a built-in function for performing partial fractions. However, we can force MATLAB to do the work for us by taking advantage of two other MATLAB functions, **diff** and **int.** We have MATLAB integrate $x(s)$, which it does internally by using partial fractions, and then immediately differentiate the resulting expression. The result will be the partial fraction expansion of $x(s)$.

Let's have MATLAB find the partial fraction expansion represented by Eq. (3.8).

```
x=(s^4-6*s^2+9*s-8)/s/(s-2)/(s+1)/(s+2)/(s-1)

x=
(s^4-6*s^2+9*s-8)/s/(s-2)/(s+1)/(s+2)/(s-1)

diff(int(x))

ans=
-2/s+1/12/(s-2)+11/3/(s+1)-17/12/(s+2)+2/3/(s-1)
pretty(ans)
```

$$-2 \,/\, s + 1 \,/\, 12 \,\frac{1}{s-2} + 11 \,/\, 3 \,\frac{1}{s+1} - \frac{17}{12}\,\frac{1}{s+2} + 2 \,/\, 3 \,\frac{1}{s-1}$$

Thus, MATLAB arrives at the same result as we did by hand.

---

and constants C_1, C_2, and C_3 are determined by the three initial conditions. The Laplace transform method has systematized the evaluation of these constants, avoiding the solution of three simultaneous equations. Four points are worth noting:

1. In both methods, one must find the roots of the characteristic equation. The roots give rise to terms in the solution *whose form is independent of the forcing function.* These terms make up the *complementary solution.*
2. The forcing function gives rise to terms in the solution *whose form depends on the form of the forcing function and is independent of the left side of the equation.* These terms comprise the *particular solution.*
3. The only interaction between these sets of terms, i.e., between the right side and left side of the differential equation, occurs in the evaluation of the constants involved.
4. The only effect of the initial conditions is in the evaluation of the constants. This is so because the initial conditions affect only the numerator of $x(s)$, as may be seen from the solution of this example.

In the three examples we have discussed, the denominator of $x(s)$ factored into real factors only. In the next example, we consider the complications that arise when the denominator of $x(s)$ has complex factors.

Using MATLAB for Symbolic Processing—Solving ODEs

MATLAB can symbolically solve ODEs. It uses the DSOLVE command for this purpose. We illustrate the use of this command with Example 3.2.

The problem consisted of

$$\frac{d^3x}{dt^3} + 2\frac{d^2x}{dt^2} - \frac{dx}{dt} - 2x = 4 + e^{2t}$$

$$x(0) = 1 \quad x'(0) = 0 \quad x''(0) = -1$$

The **DSOLVE** command is straighforward for solving this equation.

```
dsolve('D3x+2*D2x-Dx-2*x=4+exp(2*t)','x(0)=1','Dx(0)=0','D2x(0)=-1')
```

```
ans=
-1/12*exp(2*t)*(24*exp(-2*t)-1)+2/3*exp(t)-17/12*exp(-2*t)+11/3*exp(-t)
```

expand(ans)   This command multiplies out the expression to make it easier to compare with our original answer.

```
ans=
-2+1/12*exp(t)^2+2/3*exp(t)-17/12/exp(t)^2+11/3/exp(t)
```

which is the same result we obtained by hand: $x(t) = -2 + \frac{1}{12}e^{2t} + \frac{11}{3}e^{-t} - \frac{17}{12}e^{-2t} + \frac{2}{3}e^t$

We can verify this result with MATLAB by inverting the partial fraction expansion we obtained with MATLAB previously.

```
ilaplace(-2/s+1/12/(s-2)+11/3/(s+1)-17/12/(s+2)+2/3/(s-1))
```

```
ans=
-2+1/12*exp(2*t)+11/3*exp(-t)-17/12*exp(-2*t)+2/3*exp(t)
```

The result is the same!

Example 3.4.   Inversion of a transform that has complex roots in the denominator.   Solve

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

$$x(0) = 0 \qquad x'(0) = 0$$

Application of the Laplace transform yields

$$x(s) = \frac{2}{s\left(s^2 + 2s + 2\right)}$$

The quadratic term in the denominator may be factored by use of the quadratic formula. The roots are found to be $-1 - j$ and $-1 + j$. If we use these complex

roots in the partial fraction expansion, the algebra can get quite tedious. We present a method to obtain the partial fraction expansion for the case of complex roots, without resorting to the use of complex algebra.

Avoiding the use of complex algebra with a quadratic term. If we choose not to factor the quadratic term, we can use an alternate form of the partial fraction expansion.

$$x(s) = \frac{2}{s(s^2 + 2s + 2)} = \frac{A}{s} + \frac{Bs + C}{s^2 + 2s + 2}$$

Note that the second term of the expansion has the unfactored quadratic in the denominator. The numerator of each term in the expansion is a polynomial in s of one less degree than the denominator, hence the $Bs + C$ in the numerator (a first-order numerator with a second-order denominator). As before, we can determine A.

$$A = \frac{2}{0 + 2(0) + 2} = 1$$

So we now have

$$x(s) = \frac{2}{s(s^2 + 2s + 2)} = \frac{1}{s} + \frac{Bs + C}{s^2 + 2s + 2}$$

Clearing the denominator on the left-hand side, we obtain

$$2 = s^2 + 2s + 2 + Bs^2 + Cs$$

Collecting like terms, we get

$$(B + 1)s^2 + (2 + C)s + 2 = 2$$

We now match coefficients of like terms on the left and right sides of the equation.

$$s^2: \quad B + 1 = 0 \qquad B = -1$$
$$s: \quad 2 + C = 0 \qquad C = -2$$

Thus,

$$x(s) = \frac{1}{s} - \frac{s + 2}{s^2 + 2s + 2} \tag{3.9}$$

To invert the second term, we complete the square in the denominator to get a familiar transform. Remember that for a perfect square, the quadratic must have

the form $s^2 + \alpha s + (\alpha/2)^2 = (s + \alpha/2)^2$, where the constant is one-half the middle coefficient squared. The second term on the right-hand side becomes

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

where we've added and subtracted one-half the middle coefficent squared, $(2/2)^2 = 1$, so the denominator remains unchanged. The transform of the solution is now

$$x(s) = \frac{1}{s} - \frac{s + 2}{(s + 1)^2 + 1} \tag{3.10}$$

One last modification of the second term is required before inversion. A term of this type will lead to a sine term and a cosine term in the solution. From Table 2.1, we see that

$$L\{e^{-at}\sin(kt)\} = \frac{k}{(s + a)^2 + k^2} \tag{3.11a}$$

$$L\{e^{-at}\cos(kt)\} = \frac{s + a}{(s + a)^2 + k^2} \tag{3.11b}$$

Note that everywhere s appears in these forms, it appears as the quantity $s + a$. Thus, comparing these transforms with Eq. (3.10), we see that we need an $s + 1$ term in the numerator, to go with the $s + 1$ in the denominator. So we regroup as

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

We can easily invert these terms to obtain the solution to the differential equation.

$$x(t) = 1 - e^{-t}(\cos t + \sin t)$$

We now summarize the steps in this method for inverting quadratic terms with complex roots while avoiding the use of complex algebra.

Step 1. Form the partial expansion term for the quadratic with a first-order term in s in the numerator.

Step 2. Determine the numerators of the other terms in the expansion, using the Heaviside expansion.

Step 3. Cross-multiply the equation for $x(s)$ by the denominator of $x(s)$, and equate coefficients of like terms to determine the constants in the numerator of the quadratic term.

Step 4. Complete the square for the quadratic term.

Step 5. Regroup the terms in the numerator, such that if the quadratic is now $(s + a)^2$, everywhere else that s appears, it appears as $s + a$.

Step 6. Invert the resulting two terms to a sine and a cosine term (probably multiplied by an exponential).

In the next example, an exceptional case is considered; the denominator of $x(s)$ has *repeated roots*. The procedure in this case will vary slightly from that of the previous cases.

Example 3.5.   Inversion of a transform with repeated roots. Solve

$$\frac{d^3x}{dt^3} + \frac{3d^2x}{dt^2} + \frac{3dx}{dt} + x = 1$$

$$x(0) = x'(0) = x''(0) = 0$$

Application of the Laplace transform yields

$$x(s) = \frac{1}{s\left(s^3 + 3s^2 + 3s + 1\right)}$$

Factoring and expanding in partial fractions, we find

$$x(s) = \frac{1}{s(s+1)^3} = \frac{A}{s} + \frac{B}{(s+1)^3} + \frac{C}{(s+1)^2} + \frac{D}{s+1} \qquad (3.12)$$

As in the previous cases, to determine A, multiply both sides by s and then set s to zero. This yields

$$A = 1$$

Multiplication of both sides of Eq. (3.12) by $(s+1)^3$ results in

$$\frac{1}{s} = \frac{A(s+1)^3}{s} + B + C(s+1) + D(s+1)^2 \qquad (3.13)$$

Setting $s = -1$ in Eq. (3.13) gives

$$B = -1$$

Having found A and B, we introduce these values into Eq. (3.12) and place the right side of the equation over a common denominator; the result is

$$\frac{1}{s(s+1)^3} = \frac{(s+1)^3 - s + Cs(s+1) + Ds(s+1)^2}{s(s+1)^3} \qquad (3.14)$$

Expanding the numerator of the right side gives

$$\frac{1}{s(s+1)^3} = \frac{(1+D)s^3 + (3+C+2D)s^2 + (2+C+D)s + 1}{s(s+1)^3} \qquad (3.15)$$

We now equate the numerators on each side to get

$$1 = (1+D)s^3 + (3+C+2D)s^2 + (2+C+D)s + 1$$

Equating the coefficients of like powers of s gives

$$1 + D = 0$$
$$3 + C + 2D = 0$$
$$2 + C + D = 0$$

Solving these equations gives $C = -1$ and $D = -1$.
    The final result is then

$$x(s) = \frac{1}{s} - \frac{1}{(s + 1)^3} - \frac{1}{(s + 1)^2} - \frac{1}{s + 1} \tag{3.16}$$

By referring to Table 2.1, this can be inverted to

$$x(t) = 1 - e^{-t}\left(\frac{t^2}{2} + t + 1\right) \tag{3.17}$$

The reader should verify that Eq. (3.16) placed over a common denominator results in the original form

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

and that Eq. (3.17) satisfies the differential equation and initial conditions.

The result of Example 3.5 may be generalized. The appearance of the factor $(s + a)^n$ in the denominator of $x(s)$ leads to n terms in the partial fraction expansion

$$\frac{C_1}{(s + a)^n}, \frac{C_2}{(s + a)^{n-1}}, \ldots, \frac{C_n}{s + a}$$

The constant C_1 can be determined as usual by multiplying the expansion by $(s + a)^n$ and setting $s = -a$. The other constants are determined by the method shown in Example 3.5. These terms, according to Table 2.1, lead to the following expression as the inverse transform:

$$\left[\frac{C_1}{(n - 1)!}t^{n-1} + \frac{C_2}{(n - 2)!}t^{n-2} + \cdots + C_{n-1}t + C_n\right]e^{-at} \tag{3.18}$$

It is interesting to recall that in the classical method for solving these equations, one treats repeated roots of the characteristic equation by postulating the form of Eq. (3.18) and selecting the constants to fit the initial conditions.

3.2   QUALITATIVE NATURE OF SOLUTIONS

If we are interested *only in the form* of the solution $x(t)$, which is often the case in our work, *this information may be obtained directly from the roots of the denominator of* $x(s)$. As an illustration of this "qualitative" approach to differential equations, consider Example 3.4 in which

$$x(s) \; = \; \frac{2}{s\left(s^2 + 2s + 2\right)} \; = \; \frac{A}{s} \; + \; \frac{B}{s + 1 + j} \; + \; \frac{C}{s + 1 - j}$$

is the transformed solution of

$$\frac{d^2 x}{dt^2} + \frac{2' \, dx}{dt} + 2x \; = \; 2 \qquad x(0) \; = \; x'(0) \; = \; 0$$

It is evident by inspection of the partial fraction expansion, *without* evaluation of the constants, that the s in the denominator of $x(s)$ will give rise to a constant in $x(t)$. Also, since the roots of the quadratic term are $-1 \pm j$, it is known that $x(t)$ must contain terms of the form $e^{-t}(C_1 \cos t + C_2 \sin t)$ as we saw previously. This may be sufficient information for our purposes. Alternatively, we may be interested in the behavior of $x(t)$ as $t \to \infty$. It is clear that the terms involving sin and cos vanish because of the factor e^{-t}. Therefore, $x(t)$ ultimately approaches the constant, which by inspection must be unity.

The qualitative nature of the solution $x(t)$ can be related to the location of the roots of the denominator of $x(s)$ in the complex plane. These roots are the roots of the characteristic equation and the roots of the denominator of the transformed forcing function. Consider Fig. 3–1, a drawing of the complex plane, in which several typical
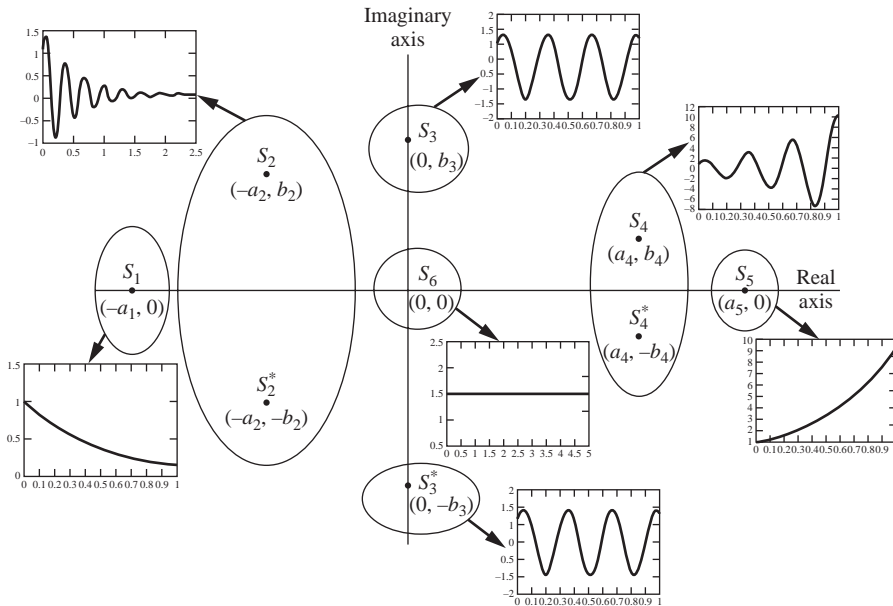
FIGURE 3–1
Location of typical roots of characteristic equation.

TABLE 3.1
Nature of terms in the solution $x(t)$ based on roots in the denominator of $X(s)$ from Fig. 3–1

Roots in denominator of $X(s)$	Terms in $x(t)$ for $t > 0$
s_1	$C_1 e^{-a_1 t}$
s_2, s_2^*	$e^{-a_2 t}(C_1 \cos b_2 t + C_2 \sin b_2 t)$
s_3, s_3^*	$C_1 \cos b_3 t + C_2 \sin b_3 t$
s_4, s_4^*	$e^{a_4 t}(C_1 \cos b_4 t + C_2 \sin b_4 t)$
s_5	$C_1 e^{a_5 t}$
s_6	C_1

roots are located and labeled with their coordinates. Table 3.1 gives the form of the terms in the equation for $x(t)$, corresponding to these roots. Note that all constants a_1, $a_2, \ldots, b_1, b_2, \ldots$ are taken as positive. The constants C_1 and C_2 are arbitrary and can be determined by the partial fraction expansion techniques. As discussed above, this determination is often not necessary for our work.

If any of these roots are repeated, the term given in Table 3.1 is multiplied by a power series in t

$$K_1 + K_2 t + K_3 t^2 + \cdots + K_r t^{r-1}$$

where r is the number of repetitions of the root and the constants $K_1, K_2, \ldots, K_r$ can be evaluated by partial fraction expansion.

It is thus evident that the imaginary axis divides the root locations into distinct areas, with regard to the behavior of the corresponding terms in $x(t)$ as t becomes large. Terms corresponding to roots to the left of the imaginary axis vanish exponentially in time, while those corresponding to roots to the right of the imaginary axis increase exponentially in time. Terms corresponding to roots at the origin behave as power series in time, a constant being considered as a degenerate power series. Terms corresponding to roots located elsewhere on the imaginary axis oscillate with constant amplitude in time unless they are multiple roots, in which case the amplitude of oscillation increases as a power series in time. Much use will be made of this information in later sections of the text.

SUMMARY

The reader now has available the basic tools for the use of Laplace transforms to solve differential equations. In addition, it is now possible to obtain considerable information about the qualitative nature of the solution with a minimum of labor. It should be pointed out that it is always necessary to factor the denominator of $x(s)$ to obtain any information about $x(t)$. If this denominator is a polynomial of order 3 or more, this may be far from a trivial problem.

The appendix to this chapter is a grouping of several Laplace transform theorems that will find later application. In addition, a discussion of the impulse function $\delta(t)$ is presented there. Unavoidably, this appendix is rather dry. It may be desirable for the reader to skip directly to Chap. 4, where our control studies begin. At each point where a theorem of App. 3A is applied, reference to the appropriate section of the appendix can be made.

PROBLEMS

3.1. Solve the following by using Laplace transforms.

(a) $\dfrac{d^2x}{dt^2} + \dfrac{dx}{dt} + x = 1 \qquad x(0) = x'(0) = 0$

(b) $\dfrac{d^2x}{dt^2} + 2\dfrac{dx}{dt} + x = 1 \qquad x(0) = x'(0) = 0$

(c) $\dfrac{d^2x}{dt^2} + 3\dfrac{dx}{dt} + x = 1 \qquad x(0) = x'(0) = 0$

Sketch the behavior of these solutions on a single graph. What is the effect of the coefficient of dx/dt?

3.2. Solve the following differential equations by Laplace transforms.

(a) $\dfrac{d^4x}{dt^4} + \dfrac{d^3x}{dt^3} = \cos t \qquad x(0) = x'(0) = x'''(0) = 0 \qquad x''(0) = 1$

(b) $\dfrac{d^2q}{dt^2} + \dfrac{dq}{dt} = t^2 + 2t \qquad q(0) = 4 \qquad q'(0) = -2$

3.3. Invert the following transforms.

(a) $\dfrac{3s}{\left(s^2 + 1\right)\left(s^2 + 4\right)}$

(b) $\dfrac{1}{s\left(s^2 - 2s + 5\right)}$

(c) $\dfrac{3s^3 - s^2 - 3s + 2}{s^2(s - 1)^2}$

3.4. Expand the following functions by partial fraction expansion. Do *not* evaluate coefficients or invert expressions.

(a) $X(s) = \dfrac{2}{(s + 1)\left(s^2 + 1\right)^2(s + 3)}$

(b) $X(s) = \dfrac{1}{s^3(s + 1)(s + 2)(s + 3)^3}$

(c) $X(s) = \dfrac{1}{(s + 1)(s + 2)(s + 3)(s + 4)}$

3.5. (*a*) Invert: $x(s) = 1/[s(s + 1)(0.5s + 1)]$

(*b*) Solve: $dx/dt + 2x = 2$ $x(0) = 0$

3.6. Obtain $y(t)$ for

(*a*) $y(s) = \dfrac{s + 1}{s^2 + 2s + 5}$

(*b*) $y(s) = \dfrac{s^2 + 2s}{s^4}$

(*c*) $y(s) = \dfrac{2s}{(s - 1)^3}$

3.7. (*a*) Invert the following function.

$$y(s) = 1/\left(s^2 + 1\right)^2$$

(*b*) Plot y versus t from 0 to 3π.

3.8. Determine $f(t)$ for $f(s) = 1/[s^2(s + 1)]$.

3.9. Solve the following differential equations.

(*a*) $\dfrac{d^2x}{dt^2} + 4\dfrac{dx}{dt} + 3x = u(t)$ $x(0) = x'(0) = 0$

(*b*) $\dfrac{d^2x}{dt^2} + 2\dfrac{dx}{dt} + x = u(t)$ $x(0) = x'(0) = 1$

(*c*) $2\dfrac{d^2x}{dt^2} + 2\dfrac{dx}{dt} + x = u(t)$ $x(0) = x'(0) = 0$

3.10. Use the trigonometric identities below to express the solution to Prob. 3.9*c* in terms of sine only. (*Note:* A sine and a cosine wave with the same frequency can be expressed as a single sine wave of the same frequency. The resulting sine wave will have a different amplitude and be phase-shifted from the original waves. This result will be important when we discuss frequency response in Chap. 15.)

$$a_1 \cos b + a_2 \sin b = a_3 \sin(b + \Phi)$$
$$\sin(A + B) = \sin A \cos B + \sin B \cos A$$

3.11. Find $f(t)$ if $F(s)$ is

(*a*) $\dfrac{1}{(s + 1)^3(s + 2)}$

(*b*) $\dfrac{s + 1}{s^2 + 2s + 5}$

(*c*) $\dfrac{s^2 - s - 6}{s^3 - 2s^2 - s + 2}$

(*d*) $\dfrac{s + 1}{s^2(s + 2)}$

(e) $\dfrac{1}{s(As + 1)(Bs + 1)}$

(f) $\dfrac{s + 1}{s(2s + 1)}$

(g) $\dfrac{s + 1}{s^2 + 3s + 1}$

(h) $\dfrac{s + 1}{s^2(2s + 1)}$

3.12. Find the solution to the following set of equations.

$$\left.\begin{array}{l} \dfrac{dx_1}{dt} = 2x_1 + 3x_2 + 1 \\[2mm] \dfrac{dx_2}{dt} = 2x_1 + x_2 + e^t \end{array}\right\} \qquad x_1(0) = x_2(0) = 0$$

Hint: Transformed equations can be manipulated algebraically to solve for each unknown (i.e., two equations in two unknowns) and then inverted separately.

3.13. Use MATLAB.
   (a) Find the partial fraction expansions for Prob. 3.11.
   (b) Invert the transforms in Prob. 3.11, using the **ILAPLACE** command.
   (c) Graph the solutions to Prob. 3.11 (skip Prob. 3.11e).

3.14. Use the MATLAB **DSOLVE** command to solve Prob. 3.12.

3.15. Use the MATLAB **DSOLVE** command to solve Prob. 3.9.

3.16. (a) Solve the differential equations in Prob. 2.3, using partial fractions.
   (b) Use the MATLAB **DSOLVE** command to solve the ODEs in Prob. 2.3.

CAPSULE SUMMARY

Any proper fraction may be resolved into a number of partial fractions subject to the following rules.

1. Factors such as $(as + b)$ in the denominator $F(s)/\cdots (as + b) \cdots$ will produce a term of type $A/(as + b)$, where A is a nonzero constant, in the expansion.
2. If there are repeated factors in the denominator, such as $F(s)/(as + b)^n$, they will produce n terms in the partial fraction expansion.

$$\underbrace{\frac{F(s)}{(as + b)^n}}_{\text{root repeated } n \text{ times}} = \underbrace{\frac{A_1}{as + b} + \frac{A_2}{(as + b)^2} + \cdots + \frac{A_n}{(as + b)^n}}_{\text{there are } n \text{ partial fractions in expansion}}$$

3. Quadratic or polynomial factors that you do not choose to factor yield

$$\frac{F(s)}{\cdots (as^2 + bs + c) \cdots} = \underbrace{\frac{As + B}{as^2 + bs + c}}_{\substack{\text{numerator is polynomial of one} \\ \text{less degree than denominator}}} + \cdots$$

4. If you have repeated quadratics,

$$\underbrace{\frac{F(s)}{\cdots (as^2 + bs + c)^n \cdots}}_{\text{yields} \rightarrow} =$$

$$\underbrace{\frac{A_1 s + B_1}{as^2 + bs + c} + \frac{A_2 s + B_2}{(as^2 + bs + c)^2} + \cdots + \frac{A_n s + B_n}{(as^2 + bs + c)^n} + \cdots}_{n \text{ terms in expansion}}$$

Useful MATLAB commands:
dsolve—used to symbolically solve ODEs
ilaplace—used to invert Laplace transforms
diff(int(...))—used to "force" MATLAB to perform a partial fraction expansion

CHAPTER APPENDIX
3

FURTHER PROPERTIES OF TRANSFORMS AND PARTIAL FRACTIONS

This appendix is a collection of theorems and results relative to the Laplace transformation. The theorems are selected because of their applicability to problems in control theory. Other theorems and properties of the Laplace transformation are available in standard texts on elementary differential equations. In later chapters, the theorems presented here will be used as needed.

3A.1   FINAL-VALUE THEOREM

If $f(s)$ is the Laplace transform of $f(t)$, then

$$\lim_{t \to \infty}[f(t)] = \lim_{s \to 0}[sf(s)]$$

provided that $sf(s)$ does not become infinite for any value of s satisfying $\mathrm{Re}(s) \geq 0$. If this condition does not hold, $f(t)$ does not approach a limit as $t \to \infty$. In the practical application of this theorem, the limit of $f(t)$ that is found by use of the theorem is correct only if $f(t)$ is bounded as t approaches infinity.

Proof.
From the Laplace transform of a derivative, we have

$$\int_0^\infty \frac{df}{dt} e^{-st}\, dt = sf(s) - f(0)$$

Hence,

$$\lim_{s \to 0} \int_0^\infty \frac{df}{dt} e^{-st} \, dt = \lim_{s \to 0} [sf(s)] - f(0)$$

It can be shown that the order of the integration and limit operation on the left side of this equation can be interchanged if the conditions of the theorem hold. Doing this gives

$$\int_0^\infty \frac{df}{dt} \, dt = \lim_{s \to 0} [sf(s)] - f(0)$$

Evaluating the integral gives

$$\lim_{t \to \infty} [f(t)] - f(0) = \lim_{s \to 0} [sf(s)] - f(0)$$

which immediately yields the desired result.

Example 3A.1.   Find the final value of the function $x(t)$ for which the Laplace transform is

$$x(s) = \frac{1}{s\left(s^3 + 3s^2 + 3s + 1\right)}$$

Direct application of the final-value theorem yields

$$\lim_{t \to \infty} [x(t)] = \lim_{s \to 0} \frac{1}{s^3 + 3s^2 + 3s + 1} = 1$$

As a check, note that this transform was inverted in Example 3.6 to give

$$x(t) = 1 - e^{-t}\left(\frac{t^2}{2} + t + 1\right)$$

which approaches unity as t approaches infinity. Note that since the denominator of $sx(s)$ can be factored to $(s + 1)^3$, the conditions of the theorem are satisfied; that is, $(s + 1)^3 \neq 0$ unless $s = -1$.

Example 3A.2.   Find the final value of the function $x(t)$ for which the Laplace transform is

$$x(s) = \frac{s^4 - 6s^2 + 9s - 8}{s(s - 2)\left(s^3 + 2s^2 - s - 2\right)}$$

In this case, the function $sx(s)$ can be written

$$sx(s) = \frac{s^4 - 6s^2 + 9s - 8}{(s + 1)(s + 2)(s - 1)(s - 2)}$$

Since this becomes infinite for $s = 1$ and $s = 2$, the conditions of the theorem are not satisfied. Note that we inverted this transform in Example 3.3, where it was found that

$$x(t) = -2 + \frac{1}{12}e^{2t} + \frac{11}{3}e^{-t} - \frac{17}{12}e^{-2t} + \tfrac{2}{3}e^{t}$$

This function continues to grow exponentially with t and, as expected, does not approach a limit. Mathematically we can understand why this is so, by referring to Fig. 3–1. A term in the denominator (the characteristic equation) such as $s - 1$ yields a root in the right half-plane (to the right of the imaginary axis in Fig. 3–1). Referring to Table 3.1, we see that the solution will involve a term of the form $C_1 e^{t}$ which will grow without bound and not approach a limit as $t \to \infty$. Thus, if any roots of the denominator have positive real parts, the final-value theorem does not apply.

The proof of the next theorem closely parallels the proof of the last one and is left as an exercise for the reader.

3A.2 INITIAL-VALUE THEOREM

$$\lim_{x \to 0}[f(t)] = \lim_{s \to \infty}[sf(s)]$$

The conditions on this theorem are not so stringent as those for the previous one because for functions of interest to us the order of integration and limiting process need not be interchanged to establish the result.

Example 3A.3. Find the initial value $x(0)$ of the function that has the transform

$$x(s) = \frac{s^4 - 6s^2 + 9s - 8}{s(s - 2)\left(s^3 + 2s^2 - s - 2\right)}$$

The function $sx(s)$ is written in the form

$$sx(s) = \frac{s^4 - 6s^2 + 9s - 8}{s^4 - 5s^2 + 4}$$

Dividing the numerator and denominator by s^4, we get

$$sx(s) = \frac{1 - 6/s^2 + 9/s^3 - 8/s^4}{1 - 5/s^2 + 4/s^4}$$

which clearly goes to unity as s becomes infinite. Hence

$$x(0) = 1$$

which again checks Example 3.3.

3A.3   TRANSLATION OF TRANSFORM

If $L\{f(t)\} = f(s)$, then

$$L\{e^{-at} f(t)\} = f(s + a)$$

In other words, the variable in the transform s is translated by a. We can transform $f(t)$ just as if the e^{-at} were not present, and then replace all the s's by $s + a$.

Proof

$$L\{e^{-at} f(t)\} = \int_0^\infty f(t) e^{-at} e^{-st}\, dt = \int_0^\infty f(t) e^{-(s+a)t}\, dt = f(s + a)$$

Example 3A.4.   Find $L\{e^{-at} \cos kt\}$. Since

$$L\{\cos kt\} = \frac{s}{s^2 + k^2}$$

then by the previous theorem,

$$L\{e^{-at} \cos kt\} = \frac{s + a}{(s + a)^2 + k^2}$$

which matches the result shown in Table 2.1.

A primary use for this theorem is in the inversion of transforms. For example, by using this theorem the transform

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

can be immediately inverted to

$$x(t) = te^{-at}$$

In obtaining this result, we made use of the following transform pair from Table 2.1:

$$L\{t\} = \frac{1}{s^2}$$

3A.4   TRANSLATION OF FUNCTION

If $L\{f(t)\} = f(s)$, then

$$L\{f(t - t_0)\} = e^{-st_0} f(s)$$

provided that

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

(which will always be true for functions we use).

Before we prove this theorem, it may be desirable to clarify the relationship between $f(t - t_0)$ and $f(t)$. This is done for an arbitrary function $f(t)$ in Fig. 3A–1, where it can be seen that $f(t - t_0)$ is simply translated horizontally from $f(t)$ through a distance t_0.

Proof.

$$L\{f(t - t_0)\} = \int_0^\infty f(t - t_0)e^{-st}\, dt$$

$$= e^{-st_0} \int_{-t_0}^\infty f(t - t_0)e^{-s(t - t_0)}\, d(t - t_0)$$

But since $f(t) = 0$ for $t < 0$, the lower limit of this integral may be replaced by zero. Since $t - t_0$ is now the dummy variable of integration, the integral may be recognized as the Laplace transform of $f(t)$; thus, the theorem is proved.

This result is also useful in inverting transforms. It follows that if $f(t)$ is the inverse transform of $f(s)$, then the inverse transform of

$$e^{-st_0} f(s)$$

is the function

$$g(t) = \begin{cases} 0 & t < t_0 \\ f(t-t_0) & t > t_0 \end{cases}$$

In other words, $g(t) = f(t - t_0)u(t - t_0)$.
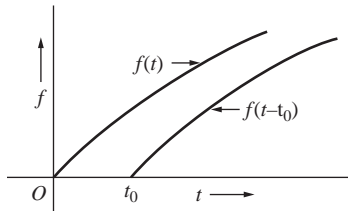
FIGURE 3A–1
Illustration of $f(t - t_0)$ as related to $f(t)$.

This property of transforms is very useful when pure time delays are present in a system. For example, consider a case where we have a stream exiting a chemical reactor, and the sensor that measures the exiting concentration lies downstream at a distance of 20 ft. If the exiting concentration changes at time $t = 0$, the sensor will not "see" the change until the fluid can travel the 20 ft down the pipeline to the sensor. The concentration signal at the sensor is delayed, and the delay time is due to the transportation lag in the pipeline. If the signal at the exit of the reactor is $C(t)$, then the signal at the sensor will be $C(t - t_0)u(t - t_0)$, where t_0 is the delay time. We will discuss this phenomenon further in Chap. 7.

Example 3A.5.   Find the Laplace transform of

$$f(t) = \begin{cases} 0 & t < 0 \\ \dfrac{1}{h} & 0 < t < h \\ 0 & t > h \end{cases}$$

FIGURE 3A–2
Pulse function of
Example 3A.5.

This function is pictured in Fig. 3A–2. It is clear that *f(t)* may be represented by the difference of two functions

$$f(t) = \frac{1}{h}[u(t) - u(t - h)]$$

where $u(t - h)$ is the unit-step function translated h units to the right. We may now use the linearity of the transform and the previous theorem to write immediately
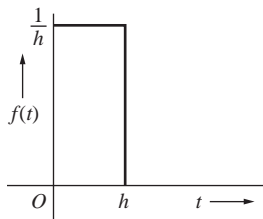
$$f(s) = \frac{1}{h}\frac{1 - e^{-hs}}{s}$$

This result is of considerable value in establishing the transform of the unit-impulse function, as will be described in the next section.

3A.5   TRANSFORM OF THE UNIT-IMPULSE FUNCTION

Consider again the function of Example 3A.5. If we allow h to shrink to zero, we obtain a new function which is zero everywhere except at the origin, where it is infinite. However, it is important to note that the area under this function always remains equal to unity. We call this new function $\delta(t)$, and the fact that its area is unity means that

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

The graph of $\delta(t)$ appears as a line of infinite height at the origin, as indicated in Table 2.1. The function $\delta(t)$ is called the unit-impulse function or, alternatively, the delta function.

It is mentioned here that, in the strict mathematical sense of a limit, the function *f(t)* does not possess a limit as h goes to zero. Hence, the function $\delta(t)$ does not fit the strict mathematical definition of a function. To assign a mathematically precise meaning to the unit-impulse function requires use of the theory of distributions, which is clearly beyond the scope of this text. However, for our work in automatic control, we will be able to obtain useful results by formal manipulation of the delta function, and hence we ignore these mathematical difficulties.

We have derived in Example 3A.5 the Laplace transform of *f(t)* as

$$L\{f(t)\} = \frac{1 - e^{-hs}}{hs}$$

Formally, then, the Laplace transform of $\delta(t)$ can be obtained by letting h go to 0 in $L\{f(t)\}$. Applying L'Hôpital's rule, we find

$$L\{\delta(t)\} = \lim_{h \to 0} \frac{1 - e^{-hs}}{hs} = \lim_{h \to 0} \frac{se^{-hs}}{s} = 1 \tag{3A.1}$$

This "verifies" the entry in Table 2.1.

It is interesting to note that since we rewrote $f(t)$ in Example 3A.5 as

$$f(t) = \frac{1}{h}[u(t) - u(t - h)]$$

then $\delta(t)$ can be written as

$$\delta(t) = \lim_{h \to 0} \frac{u(t) - u(t - h)}{h}$$

In this form, the delta function appears as the derivative of the unit-step function. The reader may find it interesting to ponder this statement in relation to the graphs of $\delta(t)$ and $u(t)$ and in relation to the integral of $\delta(t)$ discussed previously.

The unit-impulse function finds use as an idealized disturbance in control systems analysis and design. For example if one were to add water to a tank very quickly, the disturbance could be modeled as an impulse whose magnitude would be equal to the volume of the addition. If the contents of a 55-gal drum were quickly dumped into a storage tank, the addition could be modeled using an impulse function.

3A.6 TRANSFORM OF AN INTEGRAL

If $L\{f(t)\} = f(s)$, then

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

This important theorem is closely related to the theorem on differentiation. Since the operations of differentiation and integration are inverses of each other when applied to the time functions, i.e.,

$$\frac{d}{dt}\int_0^t f(t)\,dt = \int_0^t \frac{df}{dt}\,dt = f(t) \tag{3A.2}$$

it is to be expected that these operations when applied to the transforms will also be inverses. Thus assuming the theorem to be valid, Eq. (3A.2) in the transformed variable s becomes

$$s\frac{f(s)}{s} = \frac{1}{s}sf(s) = f(s)$$

In other words, multiplication of $f(s)$ by s corresponds to differentiation of $f(t)$ with respect to t, and division of $f(s)$ by s corresponds to integration of $f(t)$ with respect to t.

The proof follows from a straightforward integration by parts.

$$f(s) = \int_0^\infty f(t)e^{-st}\,dt$$

Let

$$u = e^{-st} \qquad dv = f(t)\,dt$$

Then

$$du = -se^{-st}\,dt \qquad v = \int_0^t f(t)\,dt$$

Hence,

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

Since $f(t)$ must satisfy the requirements for possession of a transform, it can be shown that the first term on the right, when evaluated at the upper limit of ∞, vanishes because of the factor e^{-st}. Furthermore, the lower limit clearly vanishes, and hence there is no contribution from the first term. The second term may be recognized as $sL\{\int_0^t f(t)\,dt\}$, and the theorem follows immediately.

Example 3A.6.   Solve the following equation for $x(t)$.

$$\frac{dx}{dt} = \int_0^t x(t)\,dt - t$$
$$x(0) = 3$$

Taking the Laplace transform of both sides and making use of the previous theorem yield

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

Solving for $x(s)$ gives

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

This may be expanded into partial fractions according to the usual procedure to give

$$x(s) = \frac{1}{s} + \frac{1}{s+1} + \frac{1}{s-1}$$

Hence,

$$x(t) = 1 + e^{-t} + e^t$$

The reader should verify that this function satisfies the original equation.

Example 3A.7.   We mentioned above that the unit impulse is the derivative of the unit-step function, and therefore the unit step must be the integral of the unit-impulse function. Use the theorem for the transform of an integral to determine the transform of the unit-step function if we know that $L\{\delta(t)\} = 1$.

Since we know that $u(t) = \int_0^\infty \delta(t)\,dt$, then

$$L\{u(t)\} = L\left\{\int_0^\infty \delta(t)\,dt\right\} = \frac{1}{s}L\{\delta(t)\} = \frac{1}{s}$$

Similarly, since we know that $du(t)/dt = \delta(t)$, then

$$L\{\delta(t)\} = L\left\{\frac{du(t)}{dt}\right\} = sL\{u(t)\} = s \cdot \frac{1}{s} = 1$$

3A.7   CUSTOM INPUTS

We can produce "custom" input signals by appropriately constructing them using standard input signals. These custom inputs are frequently useful when we analyze a process disturbance.

Example 3A.8.   Determine $f(t)$ and $f(s)$ for the input signal in Fig. 3A–3. This signal represents the step change in temperature of a furnace at hour 1, and then after holding for 2 h at that temperature, the slow ramping down to the original temperature over an hour-long period.

We can consider this input signal to be constructed from several individual "pieces," as shown in Fig. 3A–4. Thus,
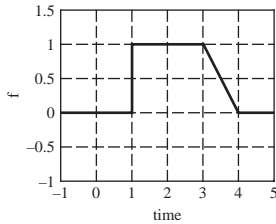
$$f(t) = u(t-1) - (t-3)u(t-3) + (t-4)u(t-4)$$

FIGURE 3A–3
Custom input signal.

A physical example where a custom input would be useful occurs if we decide to change the temperature set point of a process up by 5°F and then 10 min later we step the set
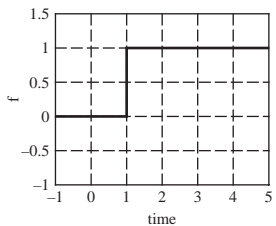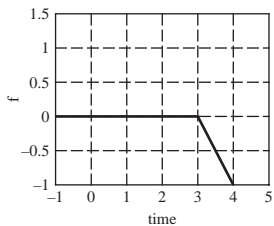
Start with this piece . . .   Add this piece . . .   Then add this piece.

FIGURE 3A–4
Constructing the custom input signal by combining standard inputs.

point back down to its original value. We've introduced a square wave "pulse" that we can construct from two unit steps.

3A.8   GENERAL DISCUSSION OF PARTIAL FRACTIONS ON A QUADRATIC TERM

In Chap. 3 we discussed how to express a quadratic term in the denominator of a fraction using partial fractions without resorting to complex algebra if we had complex roots. For completeness, we present the method using complex algebra.

Consider the general expression involving a quadratic term

$$x(s) = \frac{F(s)}{s^2 + \alpha s + \beta} \tag{3A.3}$$

where $F(s)$ is some function of s (say, $1/s$). Expanding the terms on the right side gives

$$x(s) = F_1(s) + \frac{Bs + C}{s^2 + \alpha s + \beta} \tag{3A.4}$$

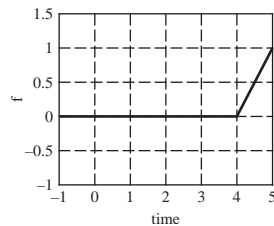where $F_1(s)$ represents other terms in the partial fraction expansion. First solve for B and C algebraically by placing the right side over a common denominator and equating the coefficients of like powers of s. The next step is to express the quadratic term in the form

$$s^2 + \alpha s + \beta = (s + a)^2 + k^2$$

The terms a and k can be found by solving for the roots of $s^2 + \alpha s + \beta = 0$ by the quadratic formula to give $s_1 = -a + jk$ and $s_2 = -a - jk$. The quadratic term can now be written

$$s^2 + \alpha s + \beta = (s - s_1)(s - s_2) = (s + a - jk)(s + a + jk) = (s + a)^2 + k^2$$

Equation (3A.4) now becomes

$$x(s) = F_1(s) + \frac{Bs + C}{(s + a)^2 + k^2} \tag{3A.5}$$

The numerator of the quadratic term is written to correspond to the transform pairs given by Eqs. (3.11*a*) and (3.11*b*).

$$Bs + C = B\left(s + a + \frac{C/B - a}{k}k\right) = B(s + a) + \frac{C - aB}{k}k$$

Equation (3A.5) becomes

$$x(s) = F_1(s) + B\frac{s + a}{(s + a)^2 + k^2} + \frac{C - aB}{k}\frac{k}{(s + a)^2 + k^2}$$

Applying the transform pairs of Eqs. (3.11*a*) and (3.11*b*) to the quadratic terms on the right gives

$$x(t) = F_1(t) + Be^{-at}\cos kt + \left(\frac{C - aB}{k}\right)e^{-at}\sin kt \qquad (3A.6)$$

where $F_1(t)$ is the result of inverting $F_1(s)$; B and C are coefficients of polynomial $Bs + C$ in numerator of quadratic term; and a and k correspond to the roots of the quadratic, roots $= -a \pm kj$. Let's use this generalized approach to solve a problem that we're already familiar with, Example 3.4.

Recall from Eq. (3.9) that

$$x(s) = \frac{A}{s} + \frac{Bs + C}{s^2 + 2s + 2} = \frac{1}{s} + \frac{-s - 2}{s^2 + 2s + 2}$$

The roots of the quadratic in the denominator are

$$\text{Roots} = \frac{-2 \pm \sqrt{4 - 8}}{2} = -1 \pm j$$

Summarizing the constants required for Eq. (3A.6), we have

A	B	C	a	k	$F_1(s)$
1	-1	-2	1	1	$1/s$

Substituting these quantities into Eq. (3A.6), we obtain the solution

$$x(t) = 1 + (-1)e^{-t}\cos t + \frac{-2 - (1)(-1)}{1}e^{-t}\sin t = 1 - e^{-t}(\cos t + \sin t)$$

which is the same as our previous result.

We now apply this method to another example.

Example 3A.9.   Solve

$$x(s) = \frac{1}{s\left(s^2 - 2s + 5\right)} = \frac{A}{s} + \frac{Bs + C}{s^2 - 2s + 5}$$

Applying the quadratic equation to the quadratic term gives

$$\text{Roots} = \frac{2 \pm \sqrt{4 - 20}}{2} = 1 \pm 2j$$

Thus, we find that $a = -1$ and $k = 2$. Solving for A, B, and C gives $A = \frac{1}{5}$, $B = -\frac{1}{5}$, and $C = \frac{2}{5}$. Introducing these values into the expression for $x(s)$ and applying Eq. (3A.6) give

$$x(t) = \frac{1}{5} - \frac{1}{5}e^t\cos 2t + \frac{1}{10}e^t\sin 2t$$

3A.9   USING COMPLEX ALGEBRA
FOR A QUADRATIC TERM

Reworking Example 3.4 using the complex roots of the quadratic, we can arrive at the partial fraction expansion

$$x(s) = \frac{2}{s(s + 1 + j)(s + 1 - j)} = \frac{A}{s} + \frac{B}{s + 1 + j} + \frac{C}{s + 1 - j} \quad (3A.7)$$

where A, B, and C are constants to be evaluated, so that this relation is an identity in s. The presence of complex factors does not alter the Heaviside procedure at all. However, the computations are more tedious.

To obtain A, multiply Eq. (3A.7) by s and set $s = 0$:

$$A = \frac{2}{(1 + j)(1 - j)} = 1$$

To obtain B, multiply Eq. (3A.7) by $s + 1 + j$ and set $s = -1 - j$:

$$B = \frac{2}{(-1 - j)(-2j)} = \frac{-1 - j}{2}$$

To obtain C, multiply Eq. (3A.7) by $s + 1 - j$ and set $s = -1 + j$:

$$C = \frac{2}{(-1 + j)(2j)} = \frac{-1 + j}{2}$$

Therefore,

$$x(s) = \frac{1}{s} + \frac{-1 - j}{2} \frac{1}{s + 1 + j} + \frac{-1 + j}{2} \frac{1}{s + 1 - j}$$

This is the desired result. To invert $x(s)$, we may now use the fact that $1/(s + a)$ is the transform of e^{-t}. The fact that a is complex does not invalidate this result, as can be seen by returning to the derivation of the transform of e^{-at}. The result is

$$x(t) = 1 + \frac{-1 - j}{2} e^{-(1 + j)t} + \frac{-1 + j}{2} e^{-(1 - j)t}$$

By using the identity

$$e^{(a + jb)t} = e^{at}(\cos bt + j\sin bt)$$

this can be converted to

$$x(t) = 1 - e^{-t}(\cos t + \sin t)$$

The details of this conversion are recommended as an exercise for the reader.

A more general discussion of this case will promote understanding. It was seen in Example 3.4 that the complex conjugate roots of the denominator of $x(s)$ gave rise to a pair of complex terms in the partial fraction expansion. The constants in these terms, B

and C, proved to be complex conjugates $(-1 - j)/2$ and $(-1 + j)/2$. When these terms were combined through a trigonometric identity, it was found that the complex terms canceled, leaving a real result for $x(t)$. Of course, it is necessary that $x(t)$ be real, since the original differential equation and initial conditions are real.

This information may be utilized as follows: The general case of complex conjugate roots arises in the form

$$x(s) = \frac{F(s)}{(s + k_1 + jk_2)(s + k_1 - jk_2)} \tag{3A.8}$$

where $F(s)$ is some real function of s.

For instance, in Example 3A.9 we had

$$F(s) = \frac{2}{s} \qquad k_1 = 1 \qquad k_2 = 1$$

Expanding Eq. (3A.8) in partial fractions gives

$$\frac{F(s)}{(s + k_1 + jk_2)(s + k_1 - jk_2)} = F_1(s) + \left(\frac{a_1 + jb_1}{s + k_1 + jk_2} + \frac{a_2 + jb_2}{s + k_1 - jk_2} \right) \tag{3A.9}$$

where a_1, a_2, b_1, and b_2 are the constants to be evaluated in the partial fraction expansion and $F_1(s)$ is a series of fractions arising from $F(s)$.

Again, in Example 3A.9,

$$a_1 = -\frac{1}{2} \qquad a_2 = -\frac{1}{2} \qquad b_1 = -\frac{1}{2} \qquad b_2 = \frac{1}{2} \qquad F_1(s) = \frac{1}{s}$$

Now, since the left side of Eq. (3A.9) is real for all real s, the right side must also be real for all real s. Since two complex numbers will add to form a real number if they are complex conjugates, the right side will be real *for all real s* if and only if the two terms are complex conjugates. Since the denominators of the terms are conjugates, this means that the numerators must also be conjugates, or

$$a_2 = a_1$$
$$b_2 = -b_1$$

This is exactly the result obtained in the specific case of Example 3.4. With this information, Eq. (3A.9) becomes

$$\frac{F(s)}{(s + k_1 + jk_2)(s + k_1 - jk_2)} = F_1(s) + \left(\frac{a_1 + jb_1}{s + k_1 + jk_2} + \frac{a_1 - jb_1}{s + k_1 - jk_2} \right) \tag{3A.10}$$

Hence, it has been established that terms in the inverse transform arising from the complex conjugate roots may be written in the form

$$(a_1 + jb_1)e^{(-k_1 - jk_2)t} + (a_1 - jb_1)e^{(-k_1 + jk_2)t}$$

Again, by using the identity

$$e^{(C_1 + jC_2)t} = e^{C_1 t}(\cos C_2 t + j \sin C_2 t)$$

this reduces to

$$2e^{-k_1 t}(a_1 \cos k_2 t + b_1 \sin k_2 t) \tag{3A.11}$$

Let us now rework Example 3A.9, using Eq. (3A.11). We return to the point at which we arrived, by our usual techniques, with the conclusion that

$$B = \frac{-1 - j}{2}$$

Comparison of Eqs. (3A.7) and (3A.10) and the result for B show that we have two possible ways to assign a_1, b_1, k_1, and k_2 so that we match the form of Eq. (3A.10). They are

$$
\begin{array}{cc}
a_1 = -\frac{1}{2} & a_1 = -\frac{1}{2} \\
b_1 = -\frac{1}{2} & b_1 = \frac{1}{2}
\end{array}
$$

or

$$
\begin{array}{cc}
k_1 = 1 & k_1 = 1 \\
k_2 = 1 & k_2 = -1
\end{array}
$$

The first way corresponds to matching the term involving B with the first term of the conjugates of Eq. (3A.10), and the second to matching it with the second term. *In either case,* substitution of these constants into Eq. (3A.11) yields

$$-e^{-t}(\cos t + \sin t)$$

which is, as we have discovered, the correct term in $x(t)$.

What this means is that one can proceed directly from the evaluation of one of the partial fraction constants, in this case B, to the complete term in the inverse transform, in this case $-e^{-t}(\cos t + \sin t)$. It is not necessary to perform all the algebra, since it has been done in the general case to arrive at Eq. (3A.11).

Another example will serve to emphasize the application of this technique.

Example 3A.10.   Solve

$$\frac{d^2 x}{dt^2} + 4x = 2e^{-1} \qquad x(0) = x'(0) = 0$$

The Laplace transform method yields

$$x(s) = \frac{2}{\left(s^2 + 4\right)(s + 1)}$$

Factoring and expanding into partial fractions give

$$\frac{2}{(s + 1)(s + 2j)(s - 2j)} = \frac{A}{s + 1} + \frac{B}{s + 2j} + \frac{C}{s - 2j} \qquad \text{(3A.12)}$$

Multiplying Eq. (3A.12) by $s + 1$ and setting $s = -1$ yields

$$A = \frac{2}{(-1 + 2j)(-1 - 2j)} = \frac{2}{5}$$

Multiplying Eq. (3A.12) by $s + 2j$ and setting $s = -2j$ yields

$$B = \frac{2}{(-2j + 1)(-4j)} = \frac{-2 + j}{10}$$

Matching the term

$$\frac{(-2 + j)/10}{s + 2j}$$

with the first term of the conjugates of Eq. (3A.10) requires that

$$a_1 = -\frac{2}{10} = -\frac{1}{5} \qquad b_1 = \frac{1}{10} \qquad k_1 = 0 \qquad k_2 = 2$$

Substituting in Eq. (3A.11) results in

$$-\frac{2}{5}\cos 2t + \frac{1}{5}\sin 2t$$

Hence the complete answer is

$$x(t) = \frac{2}{5}e^{-t} - \frac{2}{5}\cos 2t + \frac{1}{5}\sin 2t$$

Readers should verify that this answer satisfies the differential equation and initial conditions. In addition, they should show that it can also be obtained by matching the term with the second term of the conjugates of Eq. (3A.10) or by determining C instead of B.

SUMMARY

In this appendix, we have presented and discussed several properties of Laplace transforms. As we continue our studies with first-order systems, these properties will prove quite useful in our understanding and analysis of the process dynamics.

PROBLEMS

3A.1. If a forcing function $f(t)$ has the Laplace transform

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

graph the function $f(t)$.

3A.2. Solve the following equation for $y(t)$.

$$-2\int_0^t y(\tau)\,d\tau = \frac{dy}{dt} + 3y \qquad y(0) = 1$$

3A.3. Express the function given in Fig. P3A–3 in the t domain and the s domain.
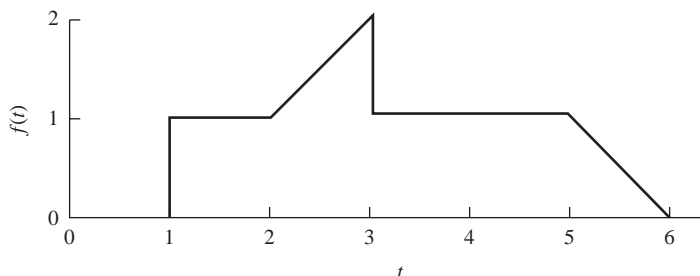

FIGURE P3A–3

3A.4. Sketch the following functions.
    (a) $f(t) = u(t) - 2u(t-1) + u(t-3)$
    (b) $f(t) = 3tu(t) - 3u(t-1) - u(t-2)$

3A.5. The function $f(t)$ has the Laplace transform

$$f(s) = (1 - 2e^{-s} + e^{-2s})/s^2$$

Obtain the function $f(t)$ and graph $f(t)$.

3A.6. Determine $f(t)$ at $t = 1.5$ and at $t = 3$ for the function

$$f(t) = 0.5u(t) - 0.5u(t-1) + (t-3)u(t-2)$$

3A.7. For the following ODE: $\dfrac{1}{2}\dfrac{dx}{dt} + x = f(t) \qquad x(0) = 0$
    (a) Use $f(t)$ from Prob. 3A.1 and determine $x(t)$. Plot $f(t)$ and $x(t)$ on the same graph.
    (b) Use $f(t)$ from Prob. 3A.3 and determine $x(t)$. Plot $f(t)$ and $x(t)$ on the same graph.
    (c) Use $f(t)$ from Prob. 3A.4a and determine $x(t)$. Plot $f(t)$ and $x(t)$ on the same graph.
    (d) Use $f(t)$ from Prob. 3A.4b and determine $x(t)$. Plot $f(t)$ and $x(t)$ on the same graph.
    (e) Use $f(t)$ from Prob. 3A.5 and determine $x(t)$. Plot $f(t)$ and $x(t)$ on the same graph.
    (f) Use $f(t)$ from Prob. 3A.6 and determine $x(t)$. Plot $f(t)$ and $x(t)$ on the same graph.

3A.8. Use the final-value theorem to predict the final values for the ODEs in Prob. 2.3. Verify your answers, using the time domain solutions obtained in Prob. 3.16.

3A.9. Find and sketch the solution to the following differential equations, using Laplace transforms.
    (a) $y' + y = \delta(t) \qquad y(0) = 0$
    (b) $y' + y = \delta(t-1) \qquad y(0) = 0$
    (c) $y' + y = u(t) \qquad y(0) = 0$
    (d) $y' + y = u(t-1) \qquad y(0) = 0$

3A.10. For the following transforms, find $\lim_{t \to \infty} f(t)$.

(a) $f(s) = \dfrac{1}{s(s + 1)^2}$

(b) $f(s) = \dfrac{1}{s(s - 1)^2}$

3A.11. Use the formula for the Laplace transform of an integral to find $L\{\sinh (kt)\}$ if you are given that $L\{\cosh (kt)\} = s/(s^2 - k^2)$.

3A.12. Transform the function $f(t) = 5e^{-2t} t \sin (3t)$.

APPENDIX
3A

CAPSULE SUMMARY

Final-value theorem: $\lim_{t\to\infty}[f(t)] = \lim_{s\to 0}[sf(s)]$

[Be careful, this does not apply if the denominator of $f(s)$ has roots with positive real parts, such as $s - 1$.]

Initial-value theorem: $\lim_{t\to 0}[f(t)] = \lim_{s\to\infty}[sf(s)]$

Translation of a transform: If $L\{f(t)\} = f(s)$, then $L\{e^{-at} f(t)\} = f(s + a)$. If the function of interest is multiplied by e^{-at}, we can transform $f(t)$ just as if the e^{-at} were not present, and then replace all the s's by $s + a$.

Translation of a function (time delay): If $L\{f(t)\} = f(s)$, then $L\{f(t - t_0)\} = e^{-st_0} f(s)$, provided that $f(t) = 0$ for $t < 0$ (which will always be true for functions we use). Alternatively, if $g(s) = e^{-st_0} f(s)$, then

$$g(t) = \begin{cases} 0 & t < t_0 \\ f(t - t_0) & t > t_0 \end{cases}$$

or, more compactly, $g(t) = f(t - t_0)u(t - t_0)$.

Transform of the unit-impulse function:

$$L\{\delta(t)\} = \lim_{h\to 0} \frac{1 - e^{-hs}}{hs} = \lim_{h\to 0} \frac{se^{-hs}}{s} = 1$$

Relationship between unit step and unit impulse: $\displaystyle\int_0^\infty \delta(t)\,dt = u(t)$

and $\dfrac{du(t)}{dt} = \delta(t)$

Transform of an integral and a derivative: Multiplication of $f(s)$ by s corresponds to differentiation of $f(t)$ with respect to t, and division of $f(s)$ by s corresponds to integration of $f(t)$ with respect to t.

$$L\left\{\frac{df(t)}{dt}\right\} = sf(s) \qquad \text{provided } f(0) = 0$$

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

When inverting an expansion containing a quadratic factor $x(s) = F_1(s) + Bs + C/(s^2 + \alpha s + \beta)$, we obtain

$$x(t) = F_1(t) + Be^{-at}\cos kt + \left(\frac{C - aB}{k}\right)e^{-at}\sin kt$$

where $F_1(t)$ is the result of inverting $F_1(s)$, B and C are the coefficients of the polynomial $Bs + C$, in the numerator of the quadratic term, and a and k correspond to the roots of the quadratic, roots $= -a \pm kj$.
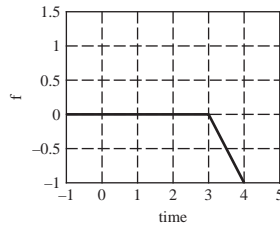
"Custom" input signals can be constructed using standard input signals.
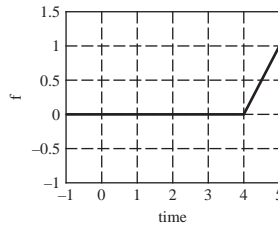

"Custom" signal
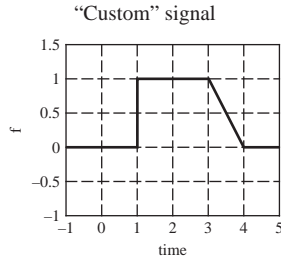
We can consider this input signal to be constructed from several individual "pieces" as follows:
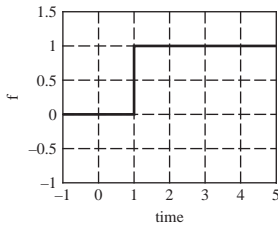
Start with this piece . . . step.   Add this piece . . . ramp.   Then add this piece . . . ramp.

PART
II

LINEAR OPEN-LOOP SYSTEMS

RESPONSE OF FIRST-ORDER SYSTEMS

Before we discuss a complete control system, it is necessary to become familiar with the responses of some of the simple, basic systems that often are the building blocks of a control system. This chapter and the three that follow describe in detail the behavior of several basic systems and show that a great variety of physical systems can be represented by a combination of these basic systems. Some of the terms and conventions that have become well established in the field of automatic control will also be introduced.

By the end of this part of the book, systems for which a transient must be calculated will be of high order and require calculations that are time-consuming if done by hand. Several software packages exist for streamlining this effort. We will use MATLAB as a tool throughout the book to demonstrate the applications of such software.

4.1 TRANSFER FUNCTION

MERCURY THERMOMETER. We develop the *transfer function* for a *first-order system* by considering the unsteady-state behavior of an ordinary mercury-in-glass thermometer. A cross-sectional view of the bulb is shown in Fig. 4–1*a*.

Consider the thermometer to be located in a flowing stream of fluid for which the temperature *x* varies with time. Our problem is to calculate the *response* or the time variation of the thermometer reading *y* for a particular change in *x*. *(In order that the result of the analysis of the thermometer be general and therefore applicable to other first-order systems, the symbols x and y have been selected to represent surrounding temperature and thermometer reading, respectively.)*

The following assumptions will be used in this analysis:

1.   All the resistance to heat transfer resides in the film surrounding the bulb (i.e., the resistance offered by the glass and mercury is neglected).
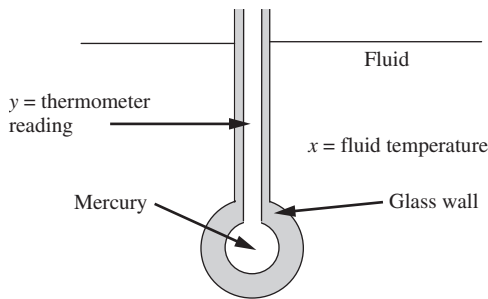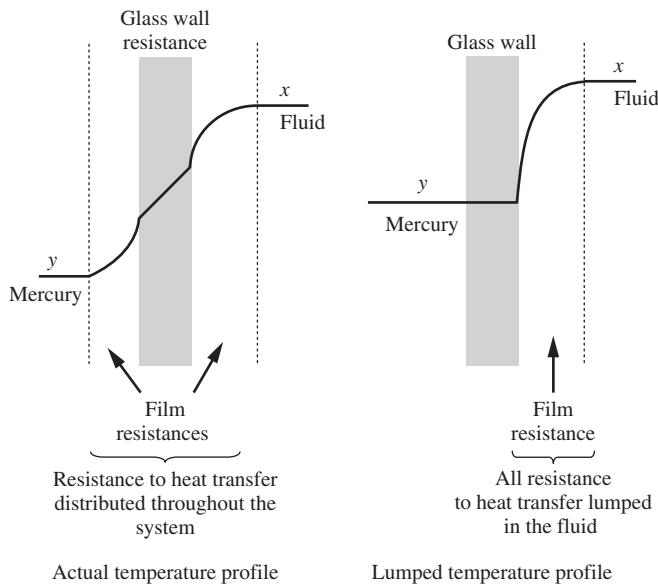
FIGURE 4–1a
Cross-sectional view of themometer.

FIGURE 4–1b
Temperature profiles in themometer.

2.   All the thermal capacity is in the mercury. Furthermore, at any instant the mercury assumes a uniform temperature throughout.

(Making these first two assumptions is often referred to as the lumping of parameters because all the resistance is "lumped" into one location and all the capacitance into another. As shown in the analysis, these assumptions make it possible to represent the dynamics of the system by an ordinary differential equation. If such assumptions were not made, the analysis would lead to a partial differential equation, and the representation would be referred to as a distributed-parameter *system. In Chap. 20, distributed-parameter systems will be considered in detail. See the difference between the actual temperature and lumped temperature profiles in Fig. 4–1b.)*

3. The glass wall containing the mercury does not expand or contract during the transient response. (In an actual thermometer, the expansion of the wall has an additional effect on the response of the thermometer reading. The glass initially expands and the cavity containing the mercury grows, resulting in a mercury reading that initially falls. Once the mercury warms and expands, the reading increases. This is an example of an inverse response. Inverse responses will be discussed in greater detail later.)

It is assumed that the thermometer is initially at steady state. This means that, before time 0, there is no change in temperature with time. At time 0, the thermometer will be subjected to some change in the surrounding temperature $x(t)$.

By applying the unsteady-state energy balance

$$\left(\text{Input rate}\right) - \left(\text{Output rate}\right) = \left(\text{Rate of accumulation}\right)$$

we get the result

$$hA(x - y) - 0 = mC\frac{dy}{dt} \tag{4.1}$$

where A = surface area of bulb for heat transfer, ft^2
C = heat capacity of mercury, $\text{Btu/(lb}_m \cdot {}^\circ\text{F)}$
m = mass of mercury in bulb, lb_m
t = time, h
h = film coefficient of heat transfer, $\text{Btu/(ft}^2 \cdot \text{h} \cdot {}^\circ\text{F)}$

For illustrative purposes, typical engineering units have been used.

Equation (4.1) states that the rate of flow of heat through the film resistance surrounding the bulb causes the internal energy of the mercury to increase at the same rate. The increase in internal energy is manifested by a change in temperature and a corresponding expansion of mercury, which causes the mercury column, or "reading" of the thermometer, to rise.

The coefficient h will depend on the flow rate and properties of the surrounding fluid and the dimensions of the bulb. We will assume that h is constant for a particular installation of the thermometer.

Our analysis has resulted in Eq. (4.1), which is a first-order differential equation. Before we solve this equation by means of the Laplace transform, *deviation variables* will be introduced into Eq. (4.1). The reason for these new variables will soon become apparent. Prior to the change in x, the thermometer is at steady state and the derivative dy/dt is zero. For the steady-state condition, Eq. (4.1) may be written

$$hA(x_s - y_s) = 0 \qquad t < 0 \tag{4.2}$$

The subscript s is used to indicate that the variable is the steady-state value. Equation (4.2) simply states that $y_s = x_s$, or the thermometer reads the true, bath temperature. Subtracting Eq. (4.2) from Eq. (4.1) gives

$$hA\left[(x - x_s) - (y - y_s)\right] = mC\frac{d(y - y_s)}{dt} \tag{4.3}$$

Notice that $d(y - y_s)/dt = dy/dt$ because y_s is a constant.

If we define the deviation variables to be the differences between the variables and their steady-state values

$$X = x - x_s$$
$$Y = y - y_s$$

then Eq. (4.3) becomes

$$hA(X - Y) = mC\frac{dY}{dt} \tag{4.4}$$

If we let $mC/hA = \tau$, Eq. (4.4) becomes

$$X - Y = \tau\frac{dY}{dt} \tag{4.5}$$

The parameter τ is called the *time constant* of the system and has the units of time. From above, we have

$$\tau = \frac{mC}{hA}[=]\frac{(\text{lb}_\text{m})\left(\dfrac{\text{Btu}}{\text{lb}_\text{m} \cdot {}^\circ\text{F}}\right)}{\left(\dfrac{\text{Btu}}{\text{ft}^2 \cdot \text{h} \cdot {}^\circ\text{F}}\right)(\text{ft}^2)}[=]\text{h}$$

Remember, in Eq. (4.5), X is the input to the system (the bath temperature) and Y is the output from the system (the indicated thermometer temperature).

Taking the Laplace transform of Eq. (4.5) gives

$$X(s) - Y(s) = \tau sY(s) - Y(0) = \tau sY(s) \tag{4.6}$$

The Laplace transform of the differential equation results in an equation that is free of initial conditions because the initial values of X and Y are zero. Since we start from steady state, $Y(0)$ must be zero,

$$Y(0) = y(0) - y_s = y_s - y_s = 0$$

And $X(0)$ is zero for the same reason. In control system engineering, we are primarily concerned with the deviations of system variables from their steady-state values. The use of deviation variables is, therefore, natural as well as convenient.

Rearranging Eq. (4.6) as a ratio of $Y(s)$ to $X(s)$ gives

$$\frac{Y(s)}{X(s)} = \frac{1}{\tau s + 1} = \frac{\text{output}}{\text{input}} \tag{4.7}$$

The expression on the right side of Eq. (4.7) is called the *transfer function* of the system. It is the ratio of the Laplace transform of the deviation in thermometer reading *(output)* to the Laplace transform of the deviation in the surrounding temperature *(input)*. In examining other physical systems, we usually attempt to obtain a transfer function.

Any physical system for which the relation between Laplace transforms of input and output deviation variables is of the form given by Eq. (4.7) is called a *first-order*

system. Synonyms for first-order systems are first-order lag and single exponential stage. The naming of all these terms is motivated by the fact that Eq. (4.7) results from a first-order, linear differential equation, Eq. (4.5). In Chap. 5 we discuss a number of other physical systems that are first-order.

To summarize the procedure for determining the transfer function for a process:

Step 1. Write the appropriate balance equations (usually mass or energy balances for a chemical process).
Step 2. Linearize terms if necessary (details on this step are given in Chap. 5).
Step 3. Place balance equations in deviation variable form.
Step 4. Laplace-transform the linear balance equations.
Step 5. Solve the resulting transformed equations for the transfer function, the output divided by the input.

This procedure is a very useful summary for developing the transfer function for a process.

Standard Form for First-Order Transfer Functions

The general form for a first-order system is

$$\tau \frac{dy}{dt} + y = K_p x(t) \tag{4.8}$$

where y is the output variable and $x(t)$ is the input forcing function. The initial conditions are

$$y(0) = y_s = K_p x(0) = K_p x_s$$

Introducing deviation variables gives

$$X = x - x_s$$
$$Y = y - y_s$$

Eq. (4.8) becomes

$$\tau \frac{dY}{dt} + Y = K_p X(t)$$
$$Y(0) = 0 \tag{4.9}$$

Transforming Eq. (4.9), we obtain

$$\tau s Y(s) + Y(s) = K_p X(s)$$

and rearranging, we obtain the standard first-order transfer function

$$\frac{Y(s)}{X(s)} = \frac{K_p}{\tau s + 1} \tag{4.10}$$

The important characteristics of the standard form are as follows:

- The denominator must be of the form $\tau s + 1$.
- The coefficient of the s term in the denominator is the system time constant τ.
- The numerator is the steady-state gain K_p.

Example 4.1.   Place the following transfer function in standard first-order form, and identify the time constant and the steady state gain.

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

Rearranging to standard form, we get

$$\frac{Y(s)}{X(s)} = \frac{6}{3s + 1}$$

Thus, the time constant is 3, and the steady-state gain is 6.

The physical significance of the steady-state gain becomes clear if we let $X(s) = 1/s$, the unit-step function. Then $Y(s)$ is given by

$$Y(s) = \frac{6}{s(3s + 1)}$$

The ultimate value of $Y(t)$ is

$$\lim_{s \to 0}[sY(s)] = \lim_{s \to 0}\left[\frac{6}{3s + 1}\right] = 6 = K_p$$

Thus the steady-state gain K_p is the steady-state value that the system attains after being disturbed by a unit-step input. It can be obtained by setting $s = 0$ in the transfer function.

PROPERTIES OF TRANSFER FUNCTIONS. In general, a transfer function relates two variables in a physical process; one of these is the cause (forcing function or input variable), and the other is the effect (response or output variable). In terms of the example of the mercury thermometer, the surrounding temperature is the cause or input, whereas the thermometer reading is the effect or output. We may write

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

where $G(s)$ = symbol for transfer function
$X(s)$ = transform of forcing function or input, in deviation form
$Y(s)$ = transform of response or output, in deviation form

The transfer function completely describes the dynamic characteristics of the system. If we select a particular input variation $X(t)$ for which the transform is $X(s)$, the response of the system is simply

$$Y(s) = G(s)X(s) \tag{4.11}$$

By taking the inverse of $Y(s)$, we get $Y(t)$, the response of the system.

   The transfer function results from a linear differential equation; therefore, the principle of superposition is applicable. This means that the transformed response of a system with transfer function $G(s)$ to a forcing function

$$X(s) = a_1 X_1(s) + a_2 X_2(s)$$

where X_1 and X_2 are particular forcing functions and a_1 and a_2 are constants, is

$$
\begin{aligned}
Y(s) &= G(s)X(s) \\
&= a_1 G(s)X_1(s) + a_2 G(s)X_2(s) \\
&= a_1 Y_1(s) + a_2 Y_2(s)
\end{aligned}
$$

where $Y_1(s)$ and $Y_2(s)$ are the responses to X_1 and X_2 alone, respectively. For example, the response of the mercury thermometer to a sudden change in surrounding temperature of 10°F is simply twice the response to a sudden change of 5°F in surrounding temperature.

   The functional relationship contained in a transfer function is often expressed by a *block diagram* representation, as shown in Fig. 4–2.
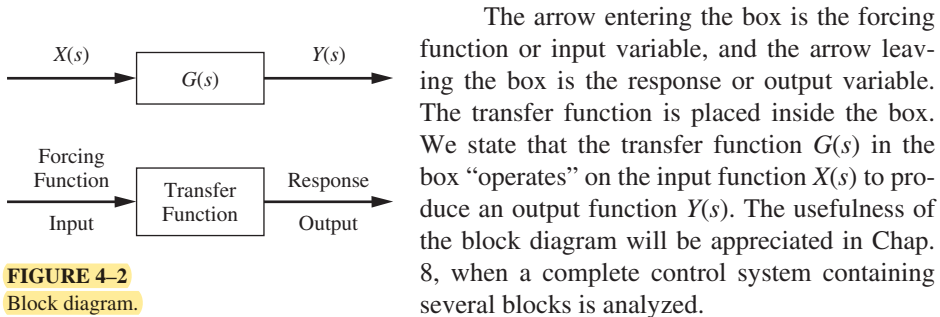
   The arrow entering the box is the forcing function or input variable, and the arrow leaving the box is the response or output variable. The transfer function is placed inside the box. We state that the transfer function $G(s)$ in the box "operates" on the input function $X(s)$ to produce an output function $Y(s)$. The usefulness of the block diagram will be appreciated in Chap. 8, when a complete control system containing several blocks is analyzed.

FIGURE 4–2
Block diagram.

4.2 TRANSIENT RESPONSE

Now that the transfer function of a first-order system has been established, we can easily obtain its transient response to *any* forcing function. Since this type of system occurs so frequently in practice, it is worthwhile to study its response to several common forcing functions: step, impulse, ramp, and sinusoidal. These forcing functions have been found to be very useful in theoretical and experimental aspects of process control. They will be used extensively in our studies, so let's explore each before we study the transient response of the first-order system to these forcing functions.

4.3 FORCING FUNCTIONS

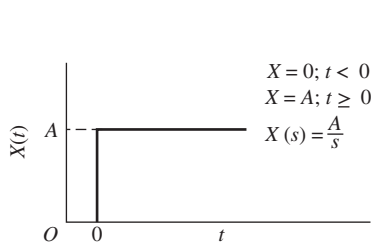**STEP FUNCTION.** Mathematically, the step function of magnitude A can be expressed as

$$X(t) = Au(t)$$

where $u(t)$ is the unit-step function defined in Chap. 2. A graphical representation is shown in Fig. 4–3.

The transform of this function is $X(s) = A/s$. A step function can be approximated very closely in practice. For example, a step change in flow rate can be obtained by the sudden opening of a valve.

$X = 0; t < 0$
$X = A; t \geq 0$
$X(s) = \frac{A}{s}$

FIGURE 4–3
Step input.

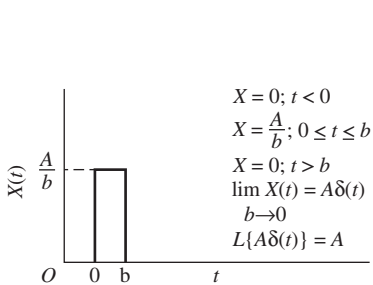**IMPULSE FUNCTION.** Mathematically, the impulse function of magnitude A is defined as

$$X(t) = A\delta(t)$$

where $\delta(t)$ is the unit-impulse function defined and discussed in App. 3A. A graphical representation of this function, before the limit is taken, is shown in Fig. 4–4.

The true impulse function, obtained by letting $b \to 0$ in Fig. 4–4, has a Laplace transform of A. It is used more frequently as a mathematical aid than as an actual input to a physical system. For some systems it is difficult even to approximate an impulse forcing function. For this reason the representation of Fig. 4–4 is valuable, since this form can usually be approximated physically by application and removal of a step function. If the time duration b is sufficiently small, we will see in Chap. 5 that the forcing function of Fig. 4–4 gives a response that closely resembles the response to a true impulse. In this sense, we often justify the use of A as the Laplace transform of the physically realizable forcing function of Fig. 4–4.
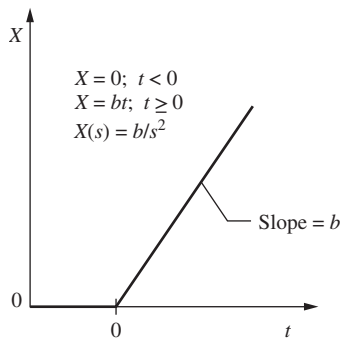
$X = 0; t < 0$
$X = \frac{A}{b}; 0 \leq t \leq b$
$X = 0; t > b$
$\lim_{b \to 0} X(t) = A\delta(t)$
$L\{A\delta(t)\} = A$

FIGURE 4–4
Impulse function.

RAMP FUNCTION. This function increases linearly with time and is described by the equations

$$X = 0 \qquad t < 0$$
$$X = bt \qquad t \geq 0$$

The ramp is shown graphically in Fig. 4–5. The transform of the ramp forcing function is $X(s) = b/s^2$.

We might, for example, desire to ramp up the temperature of an oven by 10°F/min. This would be an example of a ramp function.

$X = 0; \ t < 0$
$X = bt; \ t \geq 0$
$X(s) = b/s^2$
Slope = b

FIGURE 4–5
Ramp function.

SINUSOIDAL INPUT. This function is represented mathematically by the equations

$$X = 0 \qquad t < 0$$
$$X = A\sin \omega t \qquad t \geq 0$$

$$X = 0;\ t < 0$$
$$X = A \sin \omega t;\ t \geq 0$$
$$X(s) = \frac{A\omega}{s^2 + \omega^2}$$
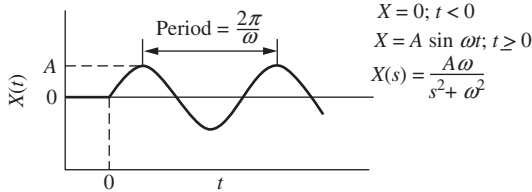
FIGURE 4–6
Sinusoidal input.

where A is the amplitude and ω is the radian frequency. The radian frequency ω is related to the frequency f in cycles per unit time by $\omega = 2\pi f$. Figure 4–6 shows the graphical representation of this function. The transform is $X(s) = A\omega/(s^2 + \omega^2)$. This forcing function forms the basis of an important branch of control theory known as *frequency response.* Historically, a large segment of the development of control theory was based on frequency-response methods, which will be presented in Chaps. 15 and 16. Physically, it is more difficult to obtain a sinusoidal forcing function in most process variables than to obtain a step function.

This completes the discussion of some of the common forcing functions. We now devote our attention to the transient response of the first-order system to each of the forcing functions just discussed.

4.4 STEP RESPONSE

If a step change of magnitude A is introduced into a first-order system, the transform of $X(t)$ is

$$X(s) = \frac{A}{s} \tag{4.12}$$

The transfer function, which is given by Eq. (4.7), is

$$\frac{Y(s)}{X(s)} = \frac{1}{\tau s + 1} \tag{4.7}$$

Combining Eqs. (4.7) and (4.12) gives

$$Y(s) = \frac{A}{s}\frac{1}{\tau s + 1} \tag{4.13}$$

This can be expanded by partial fractions to give

$$Y(s) = \frac{A/\tau}{s(s + 1/\tau)} = \frac{C_1}{s} + \frac{C_2}{s + 1/\tau} \tag{4.14}$$

Solving for the constants C_1 and C_2 by the techniques covered in Chap. 3 gives $C_1 = A$ and $C_2 = -A$. Inserting these constants into Eq. (4.14) and taking the inverse transform give the time response for Y:

$$Y(t) = 0 \qquad\qquad\qquad t < 0$$
$$Y(t) = A\left(1 - e^{-t/\tau}\right) \qquad t \geq 0$$
(4.15)

Hereafter, for the sake of brevity, it will be understood that, as in Eq. (4.15), the response is zero before $t = 0$. Equation (4.15) is plotted in Fig. 4–7 in terms of the dimensionless quantities $Y(t)/A$ and t/τ. (Note that if we refer to the standard form for a first-order system, Eq. (4.10), $K_p = A$ in this case.)

Having obtained the step response, Eq. (4.15), from a purely mathematical approach, we should consider whether the result seems to be correct from physical principles. Immediately after the thermometer is placed in the new environment, the temperature difference between the mercury in the bulb and the bath temperature is at its maximum value. With our simple lumped-parameter model, we should expect the flow of heat to commence immediately, with the result that the mercury temperature rises, causing a corresponding rise in the column of mercury. As the mercury temperature rises, the driving force causing heat to flow into the mercury will diminish, with the result that the mercury temperature changes at a slower rate as time proceeds. We see that this description of the response based on physical grounds does agree with the response given by Eq. (4–15) and shown graphically in Fig. 4–7.

Several features of this response are worth remembering:

1. The value of $Y(t)$ reaches 63.2 percent of its ultimate value when the time elapsed is equal to one time constant τ. When the time elapsed is 2τ, 3τ, and 4τ, the percent response is 86.5, 95, and 98, respectively. From these facts, one can consider the response essentially completed in three to four time constants.

2. One can show from Eq. (4.15) that the slope of the response curve at the origin in Fig. 4–7 is 1. This means that if the initial rate of change of $Y(t)$ were maintained, the response would be complete in one time constant. (See the dotted line in Fig. 4–7.)
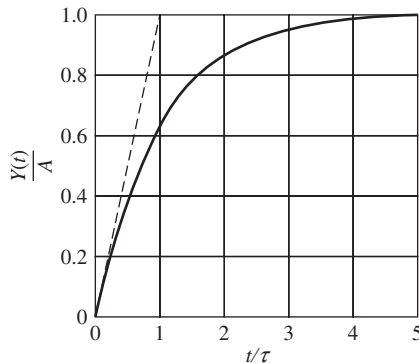
3. A consequence of the principle of superposition is that the response to a step input of any magnitude A may be obtained directly from Fig. 4–7 by multiplying the ordinate by A. Figure 4–7 actually gives the response to a unit-step function input, from which all other step responses are derived by superposition.

These results for the step response of a first-order system will now be applied to the following example.

FIGURE 4–7
Response of a first-order system to a step input.

Example 4.2. A thermometer having a time constant of 0.1 min is at a steady-state

temperature of 90°F. At time $t = 0$, the thermometer is placed in a temperature bath maintained at 100°F. Determine the time needed for the thermometer to read 98°F.

   (Note: The time constant given in this problem applies to the thermometer when it is located in the temperature bath. The time constant for the thermometer in air will be considerably different from that given because of the lower heat-transfer coefficient in air.)

   In terms of symbols used in this chapter, we have

$$\tau = 0.1 \text{ min} \qquad x_s = 90°\text{F} \qquad A = 10°\text{F}$$

The ultimate thermometer reading will, of course, be 100°F, and the ultimate value of the deviation variable $Y(\infty)$ is 10°F. When the thermometer reads 98°F, $Y(t) = 8°\text{F}$.

   Substituting into Eq. (4.12) the appropriate values of Y, A, and τ gives

$$8 = 10\left(1 - e^{-t/0.1}\right)$$

Solving this equation for t yields

$$t = 0.161 \text{ min}$$

The same result can also be obtained by referring to Fig. 4–7, where it is seen that $Y/A = 0.8$ at $t/\tau = 1.6$.

Using MATLAB to Obtain the Response of a First-Order System to a Step Function

The transform of the response is $10/s(0.1s + 1)$. We can simulate that in MATLAB by defining a system using the numerator and denominator of the response:

```
num=[10];
```

```
den=[0.1 1];
```

We then use the **step** function in MATLAB to obtain the response (Fig. 4–8).

```
step(num,den)
```

To obtain numerical values for the plot, we use the **tf** (transfer function command).

```
sys=tf(num,den)
```

```
 Transfer function:

    10
  _____
0.1 s + 1
```

```
[temp,t]=step(sys);    % Assigns the variables, temp and t, to the response.
 data=[t,temp]   % Concatenates time and temperature into one matrix and displays them.
```

FIGURE 4–8
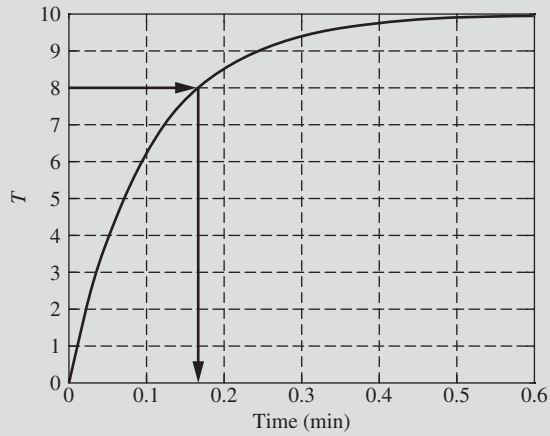Step response of thermometer in Example 4.1.

```
data =
        0                0
    0.0055           0.5372
    0.0110           1.0455
    0.0166           1.5265
    0.0221           1.9817
    0.0276           2.4124
    0.0331           2.8200
    0.0387           3.2057
    0.0442           3.5707
    0.0497           3.9161
    ......           ......
    0.1380           7.4851
    0.1436           7.6202
    0.1491           7.7481
    0.1546           7.8690
    0.1601           7.9835     temp=8.0 at approximately t = 0.16 min
    0.1656           8.0918
    0.1712           8.1943
    0.1767           8.2913
    ......           ......
    0.5908           9.9728
    0.5963           9.9743
```

Using Simulink to Obtain the Response of a First-Order System to a Step Function

We constructed the block diagram for the system using Simulink (Fig. 4–9). The simulation was run for 0.6 min, and the Scope output is shown in Fig. 4–10. The data were also exported to the MATLAB workspace and graphed in Fig. 4–11 using the **plot** command.
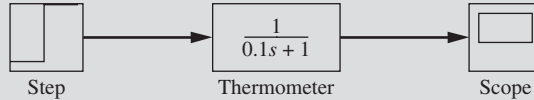
[plot (ScopeData.time, ScopeData.signals.values)]

FIGURE 4–9
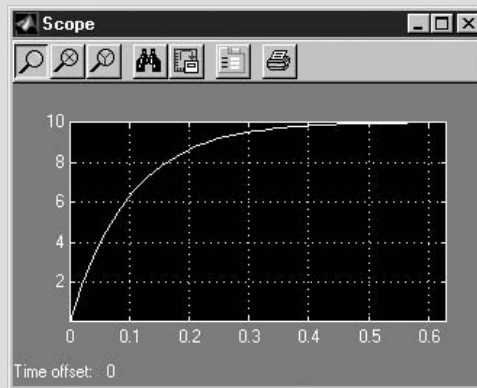Simulink block diagram for thermometer.

FIGURE 4–10
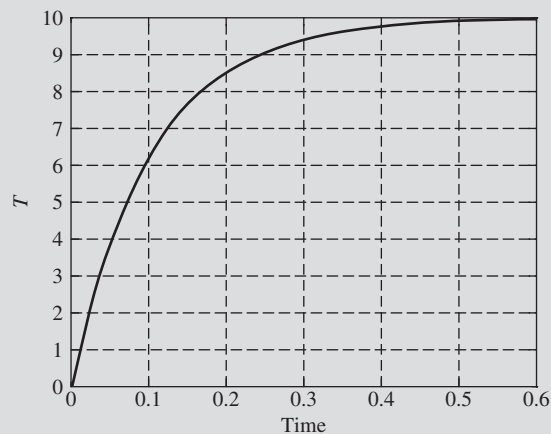Thermometer response to step input from Simulink scope.

FIGURE 4–11
Thermometer response to step input using MATLAB **plot** command.

Note that the Simulink results are the same as we obtained previously by hand calculation and with MATLAB.

The speed of the response of a first-order system is determined by the time constant for the system. Consider the following first-order system disturbed by a step input (Fig. 4–12).

The response of a first-order system for several values of τ is shown in Fig. 4–13.

It can be seen that as τ increases, it takes longer for the system to respond to the step disturbance.
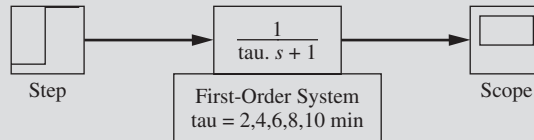

FIGURE 4–12
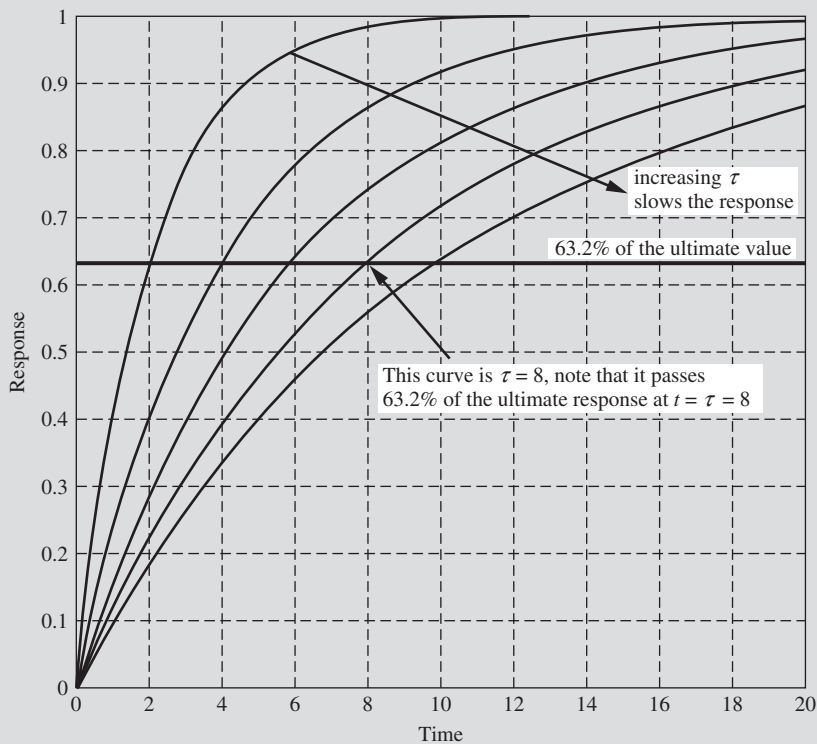Simulink model for examining the effect of τ on the step response.

FIGURE 4–13
Effect of τ on the step response of a first-order system.

4.5 IMPULSE RESPONSE

The impulse response of a first-order system will now be developed. Anticipating the use of superposition, we consider a unit impulse for which the Laplace transform is

$$X(s) = 1 \tag{4.16}$$

Combining this with the transfer function for a first-order system, which is given by Eq. (4.7), results in

$$Y(s) = \frac{1}{\tau s + 1} \tag{4.17}$$

This may be rearranged to

$$Y(s) = \frac{1/\tau}{s + 1/\tau} \tag{4.18}$$

The inverse of $Y(s)$ can be found directly from the table of transforms and can be written in the form

$$\tau Y(t) = e^{-t/\tau} \tag{4.19}$$

A plot of this response is shown in Fig. 4–14 in terms of the variables t/τ and $\tau Y(t)$. The response to an impulse of magnitude A is obtained, as usual, by multiplying $\tau Y(t)$ from Fig. 4–14 by A/τ.

Notice that the response rises immediately to 1.0 and then decays exponentially. Such an abrupt rise is, of course, physically impossible, but as we will see in Chap. 5, it is approached by the response to a finite pulse of narrow width, such as that of Fig. 4–4.
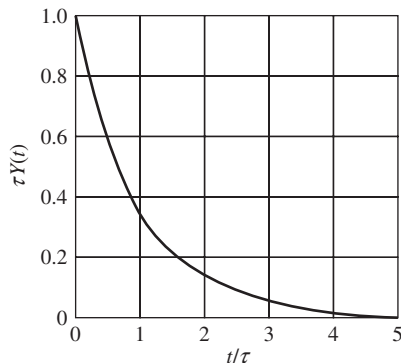

FIGURE 4–14
Unit-impulse response of a first-order system.

Using MATLAB to Generate the Impulse Response to a First-Order System

```
num=[1];
den=[1 1];
sys=tf(num,den)

 Transfer function:
   1
 ─────
 s + 1
```

`[x,y]=impulse(sys);`   % Assigns the variables x and y to the response.
`data=[x,y]`   % Concatenates x and y into one matrix and displays them.

```
data =
         0         1.0000
    0.0552         0.9463
    0.1104         0.8954
    0.1656         0.8473
    0.2209         0.8018
    0.2761         0.7588
    0.3313         0.7180
    0.3865         0.6794
    0.4417         0.6429
    0.4969         0.6084
    ......         ......
    5.0245         0.0066
    5.0797         0.0062
    5.1350         0.0059
    5.1902         0.0056
    5.2454         0.0053
    5.3006         0.0050
    5.3558         0.0047
    5.4110         0.0045
    5.4662         0.0042
    5.5215         0.0040
    5.5767         0.0038
    5.6319         0.0036
    5.6871         0.0034
    5.7423         0.0032
    5.7975         0.0030
    5.8527         0.0029
    5.9080         0.0027
    5.9632         0.0026
```

```
plot=[x,y]    % The result of this command is shown in Fig. 4–15.
```
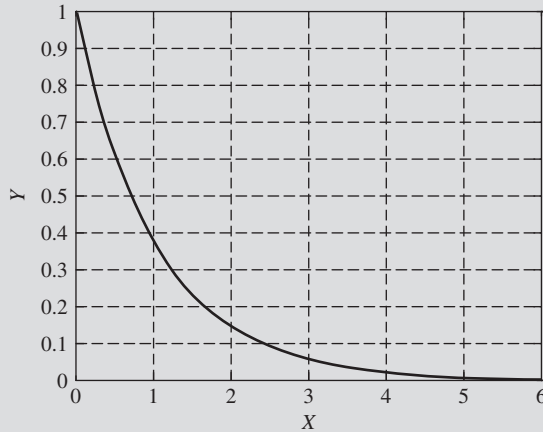

FIGURE 4–15
Impulse response of a first-order system using MATLAB.

4.6 RAMP RESPONSE

For a ramp input of $x(t) = bt$, where $X(s) = b/s^2$, the output is

$$Y(s) = \frac{b}{s^2(\tau s + 1)}$$

Rearranging and using partial fractions yield.

$$Y(s) = \frac{b}{s^2(\tau s + 1)} = \frac{b/\tau}{s^2(s + 1/\tau)} = \frac{b}{s^2} - \frac{b\tau}{s} + \frac{b\tau}{s^2 + 1/\tau}$$

$$Y(t) = bt - b\tau\left(1 - e^{-t/\tau}\right) = b(t - \tau) + b\tau e^{-t/\tau}$$

A plot of this response is shown in Fig. 4–16.

4.7 SINUSOIDAL RESPONSE

To investigate the response of a first-order system to a sinusoidal forcing function, the example of the mercury thermometer will be considered again. Consider a thermometer to be in equilibrium with a temperature bath at temperature x_s. At some time $t = 0$, the bath temperature begins to vary according to the relationship

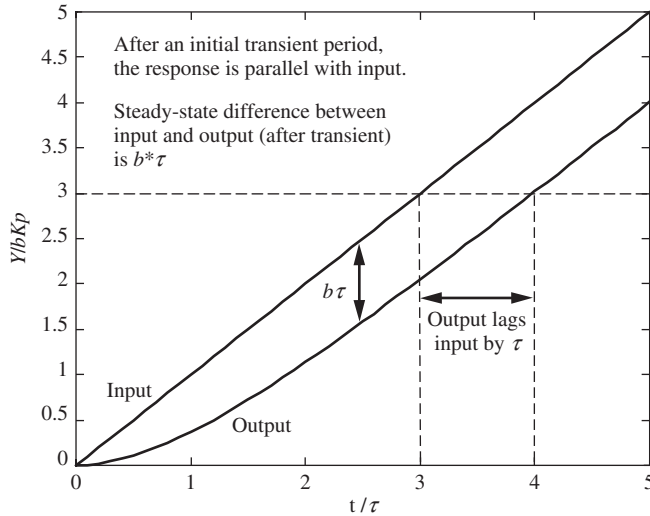$$x = x_s + A \sin \omega t \qquad t > 0 \tag{4.20}$$

FIGURE 4–16
Response of a first-order system to a ramp input.

where x = temperature of bath
          x_s = temperature of bath before sinusoidal disturbance is applied
          A = amplitude of variation in temperature
          ω = radian frequency, rad/time

In anticipation of a simple result, we introduce a deviation variable X which is defined as

$$X = x - x_s \tag{4.21}$$

Using this new variable in Eq. (4.20) gives

$$X = A \sin \omega t \tag{4.22}$$

By referring to a table of transforms, the transform of Eq. (4.22) is

$$X(s) = \frac{A\omega}{s^2 + \omega^2} \tag{4.23}$$

Combining Eqs. (4.7) and (4.23) to eliminate $X(s)$ yields

$$Y(s) = \frac{A\omega}{s^2 + \omega^2} \frac{1/\tau}{s + 1/\tau} \tag{4.24}$$

This equation can be solved for $Y(t)$ by means of a partial fraction expansion, as described in Chap. 3. The result is

$$Y(t) = \frac{A\omega\tau e^{-t/\tau}}{\tau^2\omega^2 + 1} - \frac{A\omega\tau}{\tau^2\omega^2 + 1}\cos \omega t + \frac{A}{\tau^2\omega^2 + 1}\sin \omega t \tag{4.25}$$

Equation (4.25) can be written in another form by using the trigonometric identity

$$p \cos B + q \sin B = r \sin (B + \theta) \qquad (4.26)$$

where

$$r = \sqrt{p^2 + q^2} \qquad \tan \theta = \frac{p}{q}$$

Applying the identity of Eq. (4.26) to Eq. (4.25) gives

$$Y(t) = \frac{A\omega\tau}{\tau^2\omega^2 + 1} e^{-t/\tau} + \frac{A}{\sqrt{\tau^2\omega^2 + 1}} \sin(\omega t + \phi) \qquad (4.27)$$

where

$$\phi = \tan^{-1}(-\omega\tau)$$

As $t \to \infty$, the first term on the right side of Eq. (4.27) vanishes and leaves only the ultimate periodic solution, which is sometimes called the steady-state solution

$$Y(t)\big|_s = \frac{A}{\sqrt{\tau^2\omega^2 + 1}} \sin(wt + \phi) \qquad (4.28)$$

By comparing Eq. (4.25) for the input forcing function with Eq. (4.28) for the ultimate periodic response, we see that

1. The output is a sine wave with a frequency ω equal to that of the input signal.
2. The ratio of output amplitude to input amplitude is $1/\sqrt{\tau^2\omega^2 + 1}$. This ratio is always smaller than 1. We often state this by saying that the signal is *attenuated.*
3. The output lags behind the input by an angle ϕ. It is clear that lag occurs, for the sign of ϕ is always negative.[*]

---

[*]By convention, the output sinusoid lags the input sinusoid if ϕ in Eq. (4.28) is negative. In terms of a recording of input and output, this means that the input peak occurs before the output peak. If ϕ is positive in Eq. (4.28), the system exhibits phase *lead,* or the output leads the input. In this book we always use the term *phase angle* (ϕ) and interpret whether there is lag or lead by the convention

$$\phi < 0 \qquad \text{phase lag}$$
$$\phi > 0 \qquad \text{phase lead}$$

For a particular system for which the time constant τ is a fixed quantity, it is seen from Eq. (4.28) that the attenuation of amplitude and the phase angle ϕ depend only on the frequency ω. The attenuation and phase lag increase with frequency, but the phase lag can never exceed 90° and approaches this value asymptotically.

Sinusoidal Response of a First-Order System Using MATLAB

For a first-order transfer function $1/(\tau s + 1)$, determine the response to an input function $x = \sin(4t)$. Plot the input and output on the same set of axes, and indicate the transient portion as well as the steady-state portion of the response.

```
num=[1];   % Set up the transfer function as before…
den=[1 1];
sys=tf(num,den)

 Transfer function:
   1
 ─────
 s + 1
```

```
t=0:0.1:10;              % Sets up a time vector to be used for the sine wave input.
u=sin(4*t);             % Defines the sine wave input function.
z=lsim(sys,u,t);        % Invokes the linear simulator within MATLAB and assigns
                             the output to z.
[plot(t,z,t,u)]         % Plots the input and output on the same axes.
[hold on]               % Holds the axes for further graphs.
[w=0.2353*exp(-t)+0.2425;]   % Transient envelope.
[plot(t,w)]             % Plots the transient envelope.
[q=0.2425;]             % Peak height for the ultimate periodic response.
[plot(t,q)]             % Plots the steady-state peak height
```
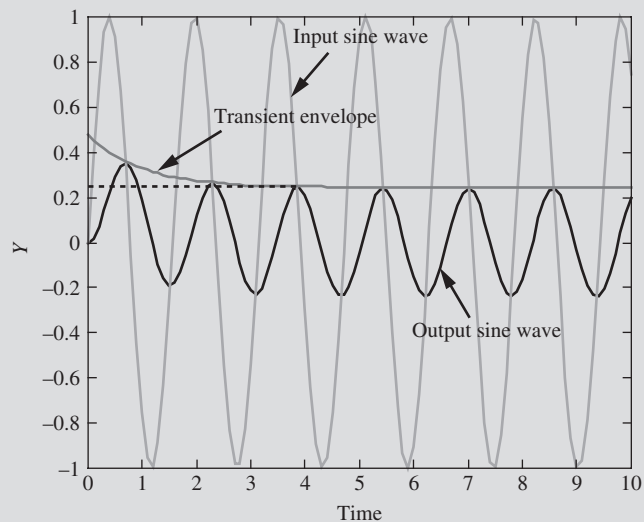
The resulting MATLAB graph is shown in Fig. 4–17.

FIGURE 4–17
Response of a first-order system to a sine wave using MATLAB.

Note that the system response has pretty much settled down to the steady periodic output wave after approximately 4 min (or 4 time constants).
   By using Eq. (4.27), the analytic solution for the response is

$$Y = 0.2353e^{-t} + 0.2425\sin(4t - 1.326 \text{ rad})$$

Note that the phase angle is -1.326 rad ($-75.96°$), and the response is nearly peaking as the input is zero and vice versa.

The sinusoidal response is interpreted in terms of the mercury thermometer by the following example.

   Example 4.3.   A mercury thermometer having a time constant of 0.1 min is placed in a temperature bath at 100°F and allowed to come to equilibrium with the bath. At time $t = 0$, the temperature of the bath begins to vary sinusoidally about its average temperature of 100°F with an amplitude of 2°F. If the frequency of oscillation is $10/\pi$ cycles/min, plot the ultimate response of the thermometer reading as a function of time. What is the phase lag?
      In terms of the symbols used in this chapter

$$\tau = 0.1 \text{ min}$$

$$x_s = 100°\text{F}$$

$$A = 2°\text{F}$$

$$f = \frac{10}{\pi} \text{ cycles/min}$$

$$\omega = 2\pi f = 2\pi \frac{10}{\pi} = 20 \text{ rad/min}$$

From Eq. (4.28), the amplitude of the response and the phase angle are calculated; thus

$$\frac{A}{\sqrt{\tau^2\omega^2 + 1}} = \frac{2}{\sqrt{4 + 1}} = 0.896°\text{F}$$

$$\phi = -\tan^{-1}2 = -63.5° = -1.11 \text{ rad}$$

or   Phase lag $= 63.5°$

The response of the thermometer is therefore

$$Y(t) = 0.896 \sin (20t - 1.11)$$

or $y(t) = 100 + 0.896 \sin (20t - 1.11)$

To obtain the lag in terms of time rather than angle, we proceed as follows: A frequency of $10/\pi$ cycles/min means that a complete cycle (peak to peak) occurs in $(10/\pi)^{-1}$ min. Since one cycle is equivalent to 360° and the lag is 63.5°, the time corresponding to this lag is

$$\text{Lag time} = \frac{63.5}{360} \times (\text{time for 1 cycle})$$

or

$$\text{Lag time} = \left(\frac{63.5}{360}\right)\left(\frac{\pi}{10}\right) = 0.0555 \text{ min}$$

thus,

$$y(t) = 100 + 0.896 \sin[20(t - 0.0555 \text{ min})]$$

In general, the lag in units of time is given by

$$\text{Lag time} = \frac{|\phi|}{360f}$$

when ϕ is expressed in degrees.

The response of the thermometer reading and the variation in bath temperature are shown in Fig. 4–18. Note that the response shown in this figure holds only after sufficient time has elapsed for the nonperiodic term of Eq. (4.27) to become negligible. For all practical purposes this term becomes negligible after a time equal to about 3τ. If the response were desired beginning from the time the bath temperature begins to oscillate, it would be necessary to plot the complete response as given by Eq. (4.27).
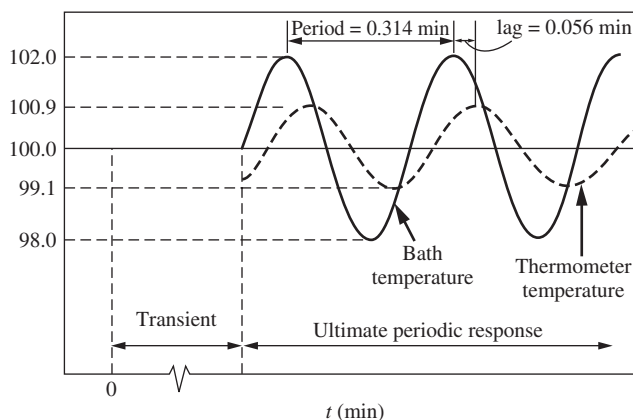

FIGURE 4–18
Response of a thermometer in Example 4.3.

SUMMARY

In this chapter several basic concepts and definitions of control theory have been introduced. These include input variable, output variable, deviation variable, transfer function, response, time constant, first-order system, block diagram, attenuation, and phase lag. Each of these ideas arose naturally in the study of the dynamics of the first-order system, which was the basic subject matter of the chapter. As might be expected, the concepts will find frequent use in succeeding chapters.

In addition to introducing new concepts, we have listed the response of the first-order system to forcing functions of major interest. This information on the dynamic behavior of the first-order system will be of significant value in the remainder of our studies.

PROBLEMS

4.1. A thermometer having a time constant of 0.2 min is placed in a temperature bath, and after the thermometer comes to equilibrium with the bath, the temperature of the bath is increased linearly with time at a rate of 1°/min. Find the difference between the indicated temperature and the bath temperature.

(*a*) 0.1 min after the change in temperature begins

(*b*) 1.0 min after the change in temperature begins

(*c*) What is the maximum deviation between indicated temperature and bath temperature, and when does it occur?

(*d*) Plot the forcing function and response on the same graph. After a long enough time, by how many minutes does the response lag the input?

4.2. A mercury thermometer bulb is $\frac{1}{2}$ in long by $\frac{1}{8}$-in diameter. The glass envelope is very thin. Calculate the time constant in water flowing at 10 ft/s at a temperature of 100°F. In your solution, give a summary that includes

(*a*) Assumptions used

(*b*) Source of data

(*c*) Results

4.3. Given: a system with the transfer function $Y(s)/X(s) = (T_1 s + 1)/(T_2 s + 1)$. Find $Y(t)$ if $X(t)$ is a unit-step function. If $T_1/T_2 = 5$, sketch $Y(t)$ versus t/T_2. Show the numerical values of minimum, maximum, and ultimate values that may occur during the transient. Check these by using the initial-value and final-value theorems of App. 3A.

4.4. A thermometer having first-order dynamics with a time constant of 1 min is placed in a temperature bath at 100°F. After the thermometer reaches steady state, it is suddenly placed in a bath at 110°F at $t = 0$ and left there for 1 min, after which it is immediately returned to the bath at 100°F.

(*a*) Draw a sketch showing the variation of the thermometer reading with time.

(*b*) Calculate the thermometer reading at $t = 0.5$ min and at $t = 2.0$ min.

4.5. Repeat Prob. 4.4 if the thermometer is in the 110°F bath for only 10 s.

4.6. A mercury thermometer, which has been on a table for some time, is registering the room temperature, 75°F. Suddenly, it is placed in a 400°F oil bath. The following data are obtained for the response of the thermometer.

Give two independent estimates of the thermometer time constant.

Time, s	Thermometer reading, °F
0	75
1	107
2.5	140
5	205
8	244
10	282
15	328
30	385

4.7. Rewrite the sinusoidal response of a first-order system [Eq. (4.27)] in terms of a cosine wave. Reexpress the forcing function [Eq. (4.22)] as a cosine wave, and compute the phase difference between input and output cosine waves.

4.8. The mercury thermometer of Prob. 4.6 is again allowed to come to equilibrium in the room air at 75°F. Then it is placed in the 400°F oil bath for a length of time less than 1 s and quickly removed from the bath and reexposed to the 75°F ambient conditions. It may be estimated that the heat-transfer coefficient to the thermometer in air is one-fifth that in the oil bath. If 10 s after the thermometer is removed from the bath it reads 98°F, estimate the length of time that the thermometer was in the bath.

4.9. A thermometer having a time constant of 1 min is initially at 50°C. It is immersed in a bath maintained at 100°C at $t = 0$. Determine the temperature reading at $t = 1.2$ min.

4.10. In Prob. 4.9, if at $t = 1.5$ min the thermometer is removed from the bath and put in a bath at 75°C, determine the maximum temperature indicated by the thermometer. What will be the indicated temperature at $t = 20$ min?

4.11. A process of unknown transfer function is subjected to a unit-impulse input. The output of the process is measured accurately and is found to be represented by the function $y(t) = te^{-t}$. Determine the unit-step response of this process.

4.12. The temperature of an oven being heated using a pulsed resistance heater varies as

$$T = 120 + 5\cos(25t + 30°)$$

where t is the time in seconds. The temperature of the oven is being measured with a thermocouple having a time constant of 5 s.

(*a*) What are the maximum and minimum temperatures indicated by the thermocouple?
(*b*) What is the maximum difference between the actual temperature and the indicated temperature?
(*c*) What is the time lag between the actual temperature and the indicated temperature?

4.13. The temperature of a experimental heated enclosure is being ramped up from 80 to 450°F at the rate of 20°F/min. A thermocouple, embedded in a thermowell for protection, is being used to monitor the oven temperature. The thermocouple has a time constant of 6 s.

(*a*) At $t = 10$ min, what is the difference between the actual temperature and the temperature indicated by the thermocouple? What is it at 60 min?
(*b*) When the thermocouple indicates 450°F, the heater will begin to modulate and maintain the temperature at the desired 450°F. What is the actual oven temperature when the thermocouple first indicates 450°F?
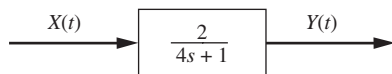

$X(t)$ → $\dfrac{2}{4s + 1}$ → $Y(t)$

FIGURE P4–14

4.14. For the transfer function in Fig. P4–14, the response $Y(t)$ is sinusoidal. The amplitude of the output wave is 0.6 and it lags behind the input by 1.5 min. Find $X(t)$. Note: the time constant in the transfer function is in minutes.

4.15. The graph in Fig. P4–15 is the response of a suspected first-order process to an impulse function of magnitude 3. Determine the transfer function $G(s)$ of the unknown process.
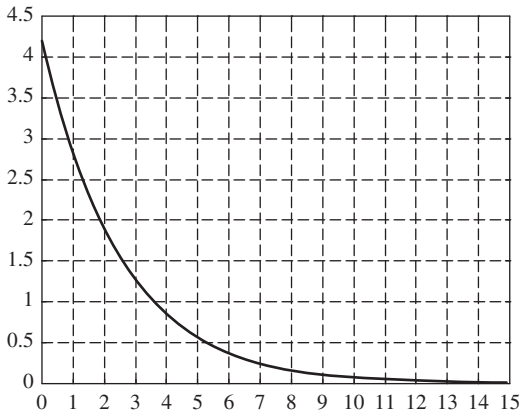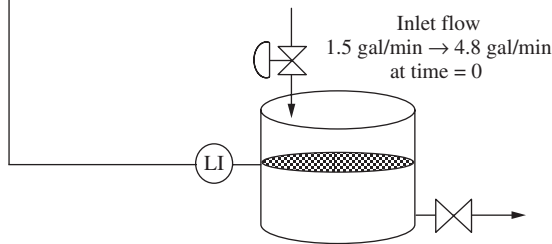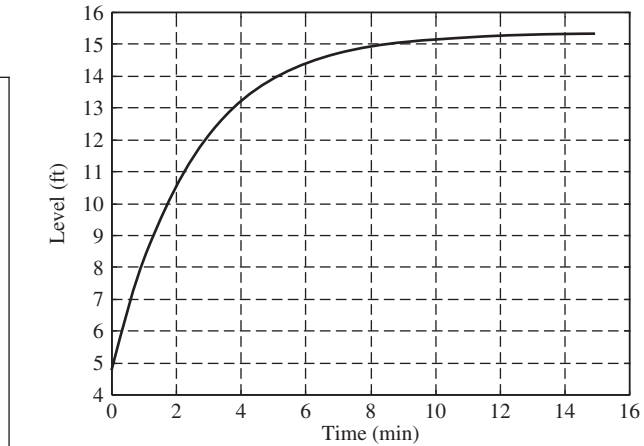

FIGURE P4–15

4.16. The level in a tank responds as a first-order system with changes in the inlet flow. Given the following level versus time data that were gathered (Fig. P4–16) after the inlet flow was

Time (min)	Level (ft)
0	4.8
0.138	5.3673
0.2761	5.9041
0.4141	6.412
0.5521	6.8927
0.6902	7.3475
0.8282	7.7779
0.9663	8.1852
1.1043	8.5706
1.2423	8.9354
1.3804	9.2805
1.5184	9.6071
1.6564	9.9161
1.7945	10.2085
1.9325	10.4853
2.0705	10.7471
2.2086	10.9949
2.3466	11.2294
2.4847	11.4513
2.6227	11.6612
2.7607	11.8599
....	....
14.3558	15.3261
14.4938	15.328
14.6319	15.3297
14.7699	15.3313

Inlet flow
1.5 gal/min → 4.8 gal/min
at time = 0

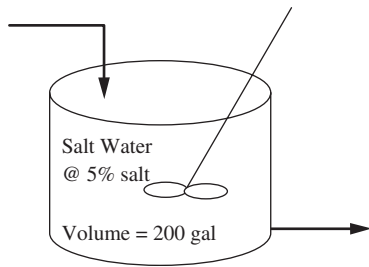*Note*: LI = level indicator

FIGURE P4–16

FIGURE P4–17

increased quickly from 1.5 to 4.8 gal/min, determine the transfer function that relates the height in the tank to the inlet flow. Be sure to use deviation variables and include units on the steady-state gain and the time constant.

4.17. A simple mixing process follows first-order behavior. A 200-gal mixing tank process, initially at steady state, is shown in Fig. P4–17. At time $t = 0$, the inlet flow is switched from 5% salt to freshwater. What does the inlet flow rate need to be to reduce the exit concentration to less than 0.5% in 30 min?

4.18. Joe, the maintenance man, dumps the contents of a 55-gal drum of water into the tank process shown below.
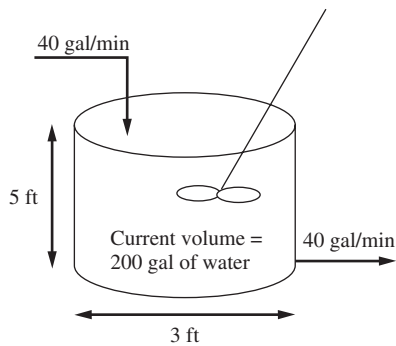

FIGURE P4–18

    (*a*) Will the tank overflow?
    (*b*) Plot the height as $f(t)$, starting at $t = 0$, the time of the dump.
    (*c*) Plot the output flow as $f(t)$, starting at $t = 0$, the time of the dump.
        NOTE: The output flow is proportional to the height of fluid in the tank.

CAPSULE SUMMARY

Standard form for first-order system transfer function:

$$G(s) = \frac{Y(s)}{X(s)} = \frac{K_p}{\tau s + 1}$$

where K_p is the steady-state gain and τ is the time constant (having units of time). Note the 1 in the denominator when the transfer function is in standard form.

Deviation variables: The difference between the process system variables and their steady-state values. When transfer functions are used, deviation values are always used. The convenience and utility of deviation variables lie in the fact that their initial values are most often zero.

$$X = x - x_s \qquad Y = y - y_s$$

Procedure for determining the transfer function for a process:

Step 1. Write the appropriate balance equations (usually mass or energy balances for a chemical process).

Step 2. Linearize terms if necessary (details on this step are given in Chap. 5).

Step 3. Place balance equations in deviation variable form.

Step 4. Laplace-transform the linear balance equations.

Step 5. Solve the resulting transformed equations for the transfer function, the output divided by the input.

Block diagrams: Graphically depict the relationship between the input variable, the transfer function, and the output variables $Y(s) = X(s)G(s)$. We always use transformed deviation variables with block diagrams.
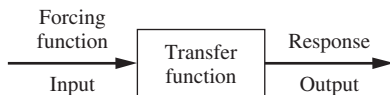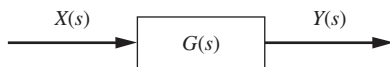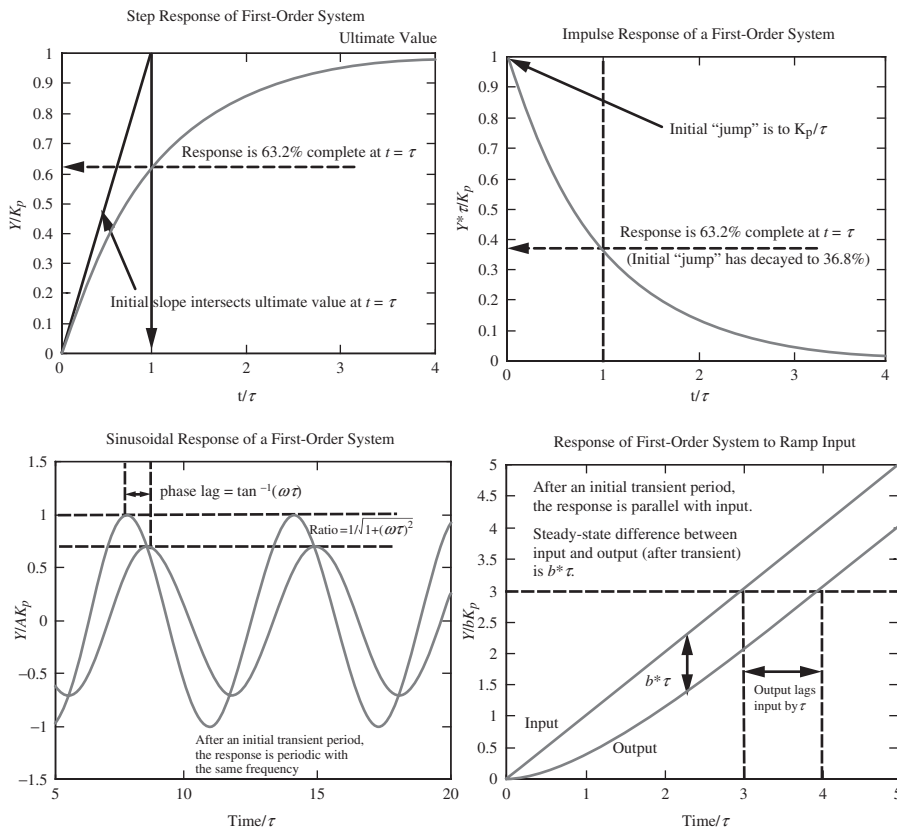

Standard responses of first-order systems to common inputs:

$$G(s) = \frac{Y(s)}{X(s)} = \frac{K_p}{\tau s + 1}$$

	Input		Output	
	$X(t)$	$X(s)$	$Y(s)$	$Y(t)$
Step	$u(t)$	$\dfrac{1}{s}$	$\dfrac{K_p}{s(\tau s + 1)}$	$K_p(1 - e^{-t/\tau})$
Impulse	$\delta(t)$	1	$\dfrac{K_p}{\tau s + 1}$	$\dfrac{K_p}{\tau} e^{-t/\tau}$
Ramp	$btu(t)$	$\dfrac{b}{s^2}$	$\dfrac{bK_p}{s^2(\tau s + 1)}$	$K_p[bt - b\tau(1 - e^{-t/\tau})]$
Sinusoid	$u(t)\,A\sin(\omega t)$	$\dfrac{A\omega}{s^2 + \omega^2}$	$\dfrac{A\omega K_p}{(s^2 + \omega^2)(\tau s + 1)}$	$\dfrac{AK_p\omega\tau}{1 + (\omega\tau)^2}e^{-t/\tau} + \dfrac{AK_p}{\sqrt{1 + (\omega\tau)^2}}\sin\left[\omega t + \tan^{-1}(-\omega\tau)\right]$

Key features of standard responses of first-order systems to common inputs

Step Response of First-Order System

Impulse Response of a First-Order System

Sinusoidal Response of a First-Order System

Response of First-Order System to Ramp Input

PHYSICAL EXAMPLES
OF FIRST-ORDER SYSTEMS

In the first part of this chapter, we will consider several physical systems that can be represented by a first-order transfer function. In the second part, a method for approximating the dynamic response of a nonlinear system by a linear response will be presented. This approximation is called linearization.

5.1 EXAMPLES OF FIRST-ORDER SYSTEMS

Liquid Level

Consider the system shown in Fig. 5–1, which consists of a tank of uniform cross-sectional area A to which is attached a flow resistance R such as a valve, a pipe, or a weir. Assume that q_o, the volumetric flow rate (volume/time) through the resistance, is related to the head h by the linear relationship
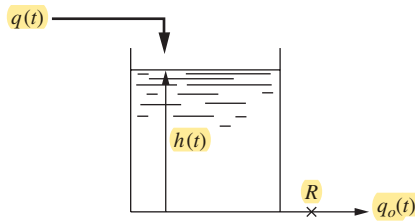
$$q_o = \frac{h}{R} \tag{5.1}$$

A resistance that has this linear relationship between flow and head is referred to as a *linear resistance.* (A pipe is a linear resistance if the flow is in the laminar range. A specially contoured weir, called a Sutro weir, produces a linear head-flow relationship. Turbulent flow through pipes and valves is generally proportional to $\sqrt{h}$. Flow through weirs having simple geometric shapes can be expressed as Kh^n, where K and n are positive constants. For example, the flow through a rectangular weir is proportional to $h^{3/2}$.)

A time-varying volumetric flow q of liquid of constant density ρ enters the tank. Determine the transfer function that relates head to flow.

We can analyze this system by writing a transient mass balance around the tank:

$$\begin{pmatrix} \text{Rate of} \\ \text{mass flow in} \end{pmatrix} - \begin{pmatrix} \text{Rate of} \\ \text{mass flow out} \end{pmatrix} = \begin{pmatrix} \text{Rate of accumulation} \\ \text{of mass in tank} \end{pmatrix}$$

$q(t)$

$h(t)$

R

$q_o(t)$

In terms of the variables used in this analysis, the mass balance becomes

$$\rho q(t) - \rho q_o(t) = \frac{d(\rho A h)}{dt}$$

$$q(t) - q_o(t) = A\frac{dh}{dt} \tag{5.2}$$

FIGURE 5–1
Liquid-level system.

Combining Eqs. (5.1) and (5.2) to eliminate $q_o(t)$ gives the following linear differential equation:

$$q - \frac{h}{R} = A\frac{dh}{dt} \tag{5.3}$$

We will introduce deviation variables into the analysis before proceeding to the transfer function. Initially, the process is operating at steady state, which means that $dh/dt = 0$ and we can write Eq. (5.3) as

$$q_s - \frac{h_s}{R} = 0 \tag{5.4}$$

where the subscript s has been used to indicate the steady-state value of the variable.
Subtracting Eq. (5.4) from Eq. (5.3) gives

$$q - q_s = \frac{1}{R}(h - h_s) + A\frac{d(h - h_s)}{dt} \tag{5.5}$$

If we define the deviation variables as

$$Q = q - q_s$$
$$H = h - h_s$$

then Eq. (5.5) can be written

$$Q = \frac{1}{R}H + A\frac{dH}{dt} \tag{5.6}$$

Taking the transform of Eq. (5.6) gives

$$Q(s) = \frac{1}{R}H(s) + AsH(s) \tag{5.7}$$

Notice that $H(0)$ is zero, and therefore the transform of dH/dt is simply $sH(s)$.
Equation (5.7) can be rearranged into the standard form of the first-order lag to give

$$\frac{H(s)}{Q(s)} = \frac{R}{\tau s + 1} \tag{5.8}$$

where $\tau = AR$.

In comparing the transfer function of the tank given by Eq. (5.8) with the transfer function for the thermometer given by Eq. (4.7), we see that Eq. (5.8) contains the factor R. The term R is simply the conversion factor that relates $h(t)$ to $q(t)$ when the system is at steady state. As we saw in Chap. 4, this value is the steady-state gain. We can again verify the physical significance of this value (as we did in Chap. 4) by applying the final-value theorem of App. 3A to the determination of the steady-state value of H when the flow rate $Q(t)$ changes according to a unit-step change; thus

$$Q(t) = u(t)$$

where $u(t)$ is the symbol for the unit-step change. The transform of $Q(t)$ is

$$Q(s) = \frac{1}{s}$$

Combining this forcing function with Eq. (5.8) gives

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

Applying the final-value theorem, proved in App. 3A, to $H(s)$ gives

$$H(t)|_{t\to\infty} = \lim_{s\to 0} \left[sH(s) \right] = \lim_{s\to 0} \frac{R}{\tau s + 1} = R$$

This shows that the ultimate change in $H(t)$ for a unit change in $Q(t)$ is simply R.

If the transfer function relating the inlet flow $q(t)$ to the outlet flow is desired, note that we have from Eq. (5.1)

$$q_{o_s} = \frac{h_s}{R} \tag{5.9}$$

Subtracting Eq. (5.9) from Eq. (5.1) and using the deviation variable $Q_o = q_o - q_{os}$ give

$$Q_o = \frac{H}{R} \tag{5.10}$$

Taking the transform of Eq. (5.10) gives

$$Q_o(s) = \frac{H(s)}{R} \tag{5.11}$$

Combining Eqs. (5.11) and (5.8) to eliminate $H(s)$ gives

$$\frac{Q_o(s)}{Q(s)} = \frac{1}{\tau s + 1} \tag{5.12}$$

Notice that the steady-state gain for this transfer function is dimensionless, which is to be expected because the input variable $q(t)$ and the output variable $q_o(t)$ have the same units (volume/time).

The possibility of approximating an impulse forcing function in the flow rate to the liquid-level system is quite real. Recall that the unit-impulse function is defined as a

pulse of unit area as the duration of the pulse approaches zero, and the impulse function can be approximated by suddenly increasing the flow to a large value for a very short time; that is, we may pour very quickly a volume of liquid into the tank. The nature of the impulse response for a liquid-level system will be described by the following example.

Example 5.1. A tank having a time constant of 1 min and a resistance of $\frac{1}{9}$ ft/cfm is operating at steady state with an inlet flow of 10 ft^3/min (or cfm). At time $t = 0$, the flow is suddenly increased to 100 ft^3/min for 0.1 min by adding an additional 9 ft^3 of water to the tank uniformly over a period of 0.1 min. (See Fig. 5–2a for this input disturbance.) Plot the response in tank level and compare with the impulse response.

Before proceeding with the details of the computation, we should observe that as the time interval over which the 9 ft^3 of water is added to the tank is shortened, the input approaches an impulse function having a magnitude of 9.

From the data given in this example, the transfer function of the process is

$$\frac{H(s)}{Q(s)} = \frac{\frac{1}{9}}{s + 1}$$

The input may be expressed as the difference in step functions, as was done in Example 3A.5.

$$Q(t) = 90[u(t) - u(t - 0.1)]$$

The transform of this is

$$Q(s) = \frac{90}{s}\left(1 - e^{-0.1s}\right)$$

Combining this and the transfer function of the process, we obtain

$$H(s) = 10\left[\frac{1}{s(s + 1)} - \frac{e^{-0.1s}}{s(s + 1)}\right] \tag{5.13}$$

The first term in Eq. (5.13) can be inverted as shown in Eq. (4.15) to give $10(1 - e^{-t})$. The second term, which includes $e^{-0.1s}$, must be inverted by use of the theorem on translation of functions given in App. 3A. According to this theorem, the inverse of $e^{-st_0} f(s)$ is $f(t - t_0)u(t - t_0)$ with $u(t - t_0) = 0$ for $t - t_0 < 0$ or $t < t_0$. The inverse of the second term in Eq. (5.13) is thus

$$L^{-1}\left\{\frac{e^{-0.1s}}{s(s + 1)}\right\} = 0 \qquad \text{for } t < 0.1$$

$$= 10\left(1 - e^{-(t - 0.1)}\right) \qquad \text{for } t > 0.1$$

or

$$10\left(1 - e^{-(t - 0.1)}\right)u(t - 0.1)$$

The complete solution to this problem, which is the inverse of Eq. (5.13), is

$$H(t) = 10\left(1 - e^{-t}\right)u(t) - 10\left(1 - e^{-(t-0.1)}\right)u(t - 0.1) \qquad (5.14)$$

which is equivalent to

$$H(t) = 10\left(1 - e^{-t}\right) \qquad\qquad\qquad t < 0.1$$

$$H(t) = 10\left[\left(1 - e^{-t}\right) - \left(1 - e^{-(t-0.1)}\right)\right] \qquad t > 0.1$$

Simplifying this expression for $H(t)$ for $t > 0.1$ gives

$$H(t) = 1.052e^{-t} \qquad t > 0.1$$

From Eq. (4.19), the response of the system to an impulse of magnitude 9 is given by

$$H(t)\big|_{\text{impulse}} = (9)\left(\tfrac{1}{9}e^{-t}\right) = e^{-t}$$

In Fig. 5–2, the pulse response of the liquid-level system and the ideal impulse response are shown for comparison. Notice that the level rises very rapidly during the 0.1 min that additional flow is entering the tank; the level then decays exponentially and follows very closely the ideal impulse response.
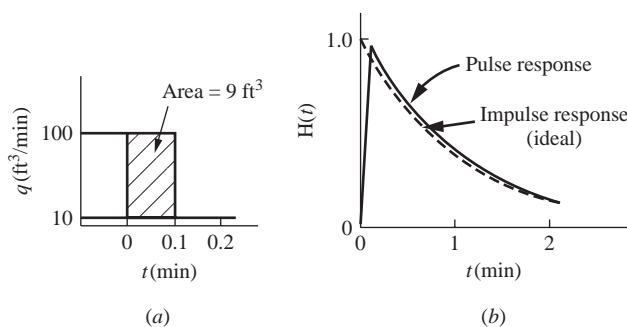

FIGURE 5–2
Approximation of an impulse function in a liquid-level system
(Example 5.1). (*a*) Pulse input; (*b*) response of tank level.

The responses to step and sinusoidal forcing functions are the same for the liquid-level system as for the mercury thermometer of Chap. 4. Hence, they need not be rederived. This is the advantage of characterizing all first-order systems by the same transfer function.

Liquid-Level Process with Constant-Flow Outlet

An example of a transfer function that often arises in control systems may be developed by considering the liquid-level system shown in Fig. 5–3. The resistance shown in
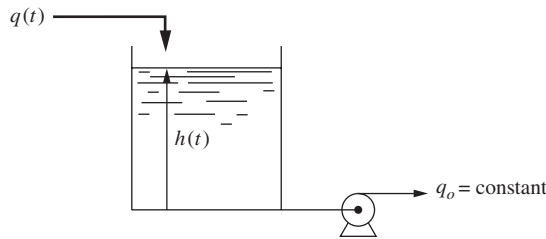
FIGURE 5–3
Liquid-level system with constant outlet flow.

Fig. 5–1 is replaced by a constant-flow pump. The same assumptions of constant cross-sectional area and constant density that were used before also apply here.

For this system, Eq. (5.2) still applies, but $q_o(t)$ is now a constant; thus

$$q(t) - q_o = A\frac{dh}{dt} \tag{5.15}$$

At steady state, Eq. (5.15) becomes

$$q_s - q_o = 0 \tag{5.16}$$

Subtracting Eq. (5.16) from Eq. (5.15) and introducing the deviation variables $Q = q - q_s$ and $H = h - h_s$ give

$$Q = A\frac{dH}{dt} \tag{5.17}$$

Taking the Laplace transform of each side of Eq. (5.17) and solving for H/Q give

$$\frac{H(s)}{Q(s)} = \frac{1}{As} \tag{5.18}$$

Notice that the transfer function $1/As$ in Eq. (5.18) is equivalent to integration. (Recall from App. 3A that multiplying the transform by s corresponds to differentiation of the function in the time domain, while dividing by s corresponds to integration in the time domain.) Therefore, the solution of Eq. (5.18) is

$$h(t) = h_s + \frac{1}{A}\int_0^t Q(t)\,dt \tag{5.19}$$

Clearly, if we increase the inlet flow to the tank, the level will increase because the outlet flow remains constant. The excess volumetric flow rate into the tank accumulates, and the level rises. For instance, if a step change $Q(t) = u(t)$ were applied to the system shown in Fig. 5–3 the result would be

$$h(t) = h_s + t/A \tag{5.20}$$

The step response given by Eq. (5.20) is a ramp function that grows without limit. Such a system that grows without limit for a sustained change in input is said to have

nonregulation. Systems that have a limited change in output for a sustained change in input are said to have *regulation.* An example of a system having regulation is the step response of a first-order system, such as that shown in Fig. 5–1. If the inlet flow to the process shown in Fig. 5–1 is increased, the level will rise until the outlet flow becomes equal to the inlet flow, and then the level stops changing. This process is said to be self-regulating.

The transfer function for the liquid-level system with constant outlet flow given by Eq. (5.18) can be considered as a special case of Eq. (5.8) as $R \rightarrow \infty$.

$$\lim_{R \to \infty} \left(\frac{R}{ARs + 1} \right) = \frac{1}{As}$$

The next example of a first-order system is a mixing process.

Mixing Process

Consider the mixing process shown in Fig. 5–4 in which a stream of solution containing dissolved salt flows at a constant volumetric flow rate q into a tank of constant holdup volume V. The concentration of the salt in the entering stream x (mass of salt/volume) varies with time. It is desired to determine the transfer function relating the outlet concentration y to the inlet concentration x.
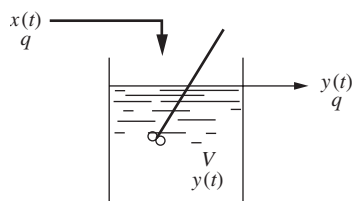

If we assume the density of the solution to be constant, the flow rate in must equal the flow rate out, since the holdup volume is fixed. We may analyze this system by writing a transient mass balance for the salt; thus

FIGURE 5–4
Mixing process.

$$\begin{pmatrix} \text{Flow rate of} \\ \text{salt in} \end{pmatrix} - \begin{pmatrix} \text{Flow rate of} \\ \text{salt out} \end{pmatrix} = \begin{pmatrix} \text{Rate of accumulation} \\ \text{of salt in tank} \end{pmatrix}$$

Expressing this mass balance in terms of symbols gives

$$qx - qy = \frac{d(Vy)}{dt} = V\frac{dy}{dt} \tag{5.21}$$

We will again introduce deviation variables as we have in the previous examples. At steady state, Eq. (5.21) may be written

$$qx_s - qy_s = 0 \tag{5.22}$$

Subtracting Eq. (5.22) from Eq. (5.21) and introducing the deviation variables

$$X = x - x_s$$
$$Y = y - y_s$$

give

$$qX - qY = V\frac{dY}{dt}$$

Taking the Laplace transform of this expression and rearranging the result give

$$\frac{Y(s)}{X(s)} = \frac{1}{\tau s + 1} \tag{5.23}$$

where $\tau = V/q$.

This mixing process is, therefore, another first-order process for which the dynamics are now well known.

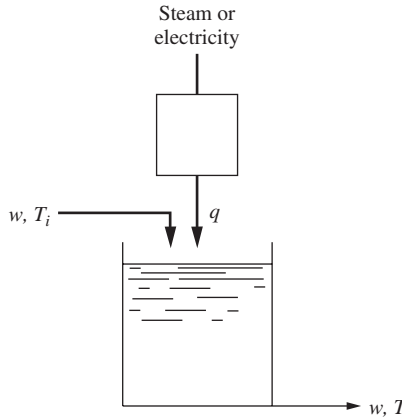Our last example of a first-order system is a heating process.

Heating Process

Consider the heating process shown in Fig. 5–5. A stream at temperature T_i is fed to the tank. Heat is added to the tank by means of an electric heater. The tank is well mixed, and the temperature of the exiting stream is T. The flow rate to the tank is constant at w lb/h.

A transient energy balance on the tank yields

Steam or electricity

w, T_i q

w, T

FIGURE 5–5
Heating process.

$$\begin{pmatrix} \text{Rate of} \\ \text{energy flow} \\ \text{into tank} \end{pmatrix} - \begin{pmatrix} \text{Rate of} \\ \text{energy flow} \\ \text{out of tank} \end{pmatrix} + \begin{pmatrix} \text{Rate of} \\ \text{energy flow in} \\ \text{from heater} \end{pmatrix} = \begin{pmatrix} \text{Rate of} \\ \text{accumulation of} \\ \text{energy in tank} \end{pmatrix}$$

Converting this energy balance to symbols results in

$$wC(T_i - T_{\text{ref}}) - wC(T - T_{\text{ref}}) + q = \rho VC \frac{d(T - T_{\text{ref}})}{dt} = \rho VC \frac{dT}{dt} \tag{5.24}$$

where T_{ref} is the reference temperature and C is the heat capacity of the fluid. At steady state, dT/dt is zero, and Eq. (5.24) can be written

$$wC(T_{is} - T_s) + q_s = 0 \tag{5.25}$$

where the subscript s has been used to indicate steady state. Subtracting Eq. (5.25) from Eq. (5.24) gives

$$wC(T_i - T_{is}) - wC(T - T_s) + q - q_s = \rho VC \frac{d(T - T_s)}{dt} \tag{5.26}$$

If we assume that T_i is constant (and so $T_i = T_{is}$) and introduce the deviation variables

$$T' = T - T_s$$
$$Q = q - q_s$$

Eq. (5.26) becomes

$$-wCT' + Q = \rho VC\frac{dT'}{dt} \tag{5.27}$$

Taking Laplace transforms of Eq. (5.27) gives

$$-wCT'(s) + Q(s) = \rho VCsT'(s) \tag{5.28}$$

Rearranging Eq. (5.28) produces the following first-order transfer function relating $T'(s)$ and $Q(s)$:

$$\frac{T'(s)}{Q(s)} = \frac{1/wC}{(\rho V/w)s + 1} = \frac{K}{\tau s + 1} \tag{5.29}$$

Thus, this process exhibits first-order dynamics as the tank temperature T responds to changes in the heat input to the tank.

Example 5.2.   Consider the mixed tank heater shown in Fig. 5–6. Develop a transfer function relating the tank outlet temperature to changes in the inlet temperature. Determine the response of the outlet temperature of the tank to a step change in the inlet temperature from 60 to 70°C. Before we proceed, intuitively what would we expect to happen? If the inlet temperature rises by 10°C, we expect the outlet temperature to eventually rise by 10°C if nothing else changes. Let's see what modeling the process will tell us.

From Eq. (5.26) we can write the following simplified balance, realizing that $q = q_s$:

$$wC(T_i - T_{is}) - wC(T - T_s) = \rho VC\frac{d(T - T_s)}{dt}$$

In terms of deviation variables, this becomes

$$wCT_i' - wCT' = \rho VC\frac{dT'}{dt}$$

Transforming, we get

$$wCT_i'(s) - wCT'(s) = \rho VCsT'(s)$$

and finally, after rearranging,

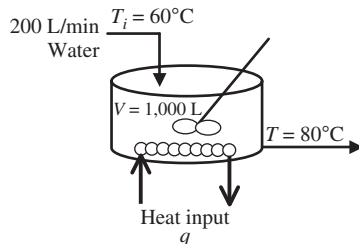$$\frac{T'(s)}{T_i'(s)} = \frac{1}{(\rho V/w)s + 1} = \frac{1}{\tau s + 1}$$

200 L/min $T_i = 60°C$
Water

$V = 1,000$ L

$T = 80°C$

Heat input
q

FIGURE 5–6
Mixed tank heater.

Substituting in numerical values for the variables, we obtain the actual transfer function for this mixed tank heater.

$$\tau = \frac{\rho V}{w} = \frac{V}{w/\rho} = \frac{V}{v} = \frac{\text{tank volume}}{\text{volumetric flow rate}} = \frac{1,000 \text{ L}}{200 \text{ L/min}} = 5 \text{ min}$$

$$\frac{T'(s)}{T_i'(s)} = \frac{1}{5s + 1}$$

If the inlet temperature is stepped from 60 to 70°C, $T_i'(t) = 70 - 60 = 10$ and $T_i'(s) = 10/s$. Thus,

$$T'(s) = \frac{10}{s}\left(\frac{1}{5s + 1}\right)$$

Inverting to the time domain, we obtain the expression for $T'(t)$

$$T'(t) = 10\left(1 - e^{-t/5}\right)$$

and finally, we obtain the expression for $T(t)$, the actual tank outlet temperature.

$$T(t) = T_s + T'(t) = 80 + 10\left(1 - e^{-t/5}\right)$$

A plot of the outlet temperature (in deviation variables) is shown in the Fig. 5–7a. The actual outlet temperature is shown in Fig. 5–7b. Note that for the uncontrolled mixing tank, a step change of 10°C in the inlet temperature
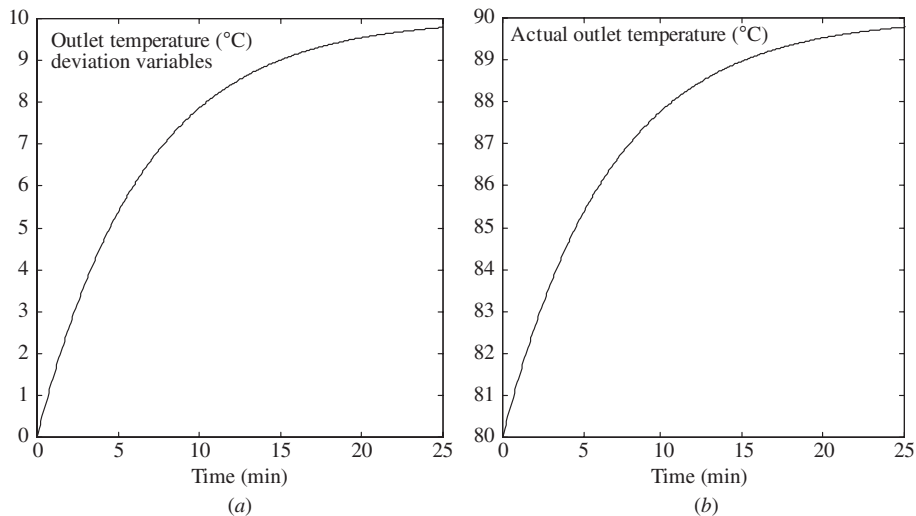

FIGURE 5–7
(a) Tank outlet temperature (deviation variable); (b) actual tank outlet temperature.

ultimately produces a 10°C change in the outlet temperature, just as we predicted intuitively before we began our modeling. This result is just what we expected.

The three examples presented in this section are intended to show that the dynamic characteristics of many physical systems can be represented by a first-order transfer function. In the remainder of the book, more examples of first-order systems will appear as we discuss a variety of control systems.

In summarizing the previous examples of first-order systems, the time constant for each has been expressed in terms of system parameters; thus

$$\tau = \frac{mC}{hA} \quad \text{for thermometer, Eq. (4.5)}$$

$$\tau = AR \quad \text{for liquid-level process, Eq. (5.8)}$$

$$\tau = \frac{V}{q} \quad \text{for mixing process, Eq. (5.23)}$$

$$\tau = \frac{\rho V}{w} \quad \text{for heating process, Eq. (5.29)}$$

5.2   LINEARIZATION

Thus far, all the examples of physical systems, including the liquid-level system of Fig. 5–1, have been linear. Actually, most physical systems of practical importance are nonlinear.

Characterization of a dynamic system by a transfer function can be done only for linear systems (those described by linear differential equations). The convenience of using transfer functions for dynamic analysis, which we have already seen in applications, provides significant motivation for approximating nonlinear systems by linear ones. A very important technique for such approximation is illustrated by the following discussion of the liquid-level system of Fig. 5–1.

We now assume that the flow out of the tank follows a square root relationship

$$q_o = Ch^{1/2} \tag{5.30}$$

where C is a constant.

For a liquid of constant density and a tank of uniform cross-sectional area A, a material balance around the tank gives

$$q(t) - q_o(t) = A\frac{dh}{dt} \tag{5.31}$$

Combining Eqs. (5.30) and (5.31) gives the nonlinear differential equation

$$q - Ch^{1/2} = A\frac{dh}{dt} \tag{5.32}$$

At this point, we cannot proceed as before and take the Laplace transform. This is due to the presence of the nonlinear term $h^{1/2}$, for which there is no simple transform. This difficulty can be circumvented by linearizing the nonlinear term.

By means of a Taylor series expansion, the function $q_o(h)$ may be expanded around the steady-state value h_s; thus

$$q_o = q_o(h_s) + q'_o(h_s)(h - h_s) + \frac{q''_o(h_s)(h - h_s)^2}{2} + \cdots$$

where $q'_o(h_s)$ is the first derivative of q_o evaluated at h_s, $q''_o(h_s)$ is the second derivative, etc. If we keep only the linear term, the result is

$$q_o \cong q_o(h_s) + q'_o(h_s)(h - h_s) \tag{5.33}$$

Taking the derivative of q_o with respect to h in Eq. (5.30) and evaluating the derivative at $h = h_s$ give

$$q'_o(h_s) = \tfrac{1}{2} C h_s^{-1/2}$$

Introducing this into Eq. (5.33) gives

$$q_o = q_{o_s} + \frac{1}{R_1}(h - h_s) \tag{5.34}$$

where $q_{o_s} = q_o(h_s)$ and $1/R_1 = \tfrac{1}{2} C h_s^{-1/2}$.

Substituting Eq. (5.34) into Eq. (5.31) gives

$$q - q_{o_s} - \frac{h - h_s}{R_1} = A \frac{dh}{dt} \tag{5.35}$$

At steady state the flow entering the tank equals the flow leaving the tank; thus

$$q_s = q_{o_s} \tag{5.36}$$

Introducing this last equation into Eq. (5.35) gives

$$A \frac{dh}{dt} + \frac{h - h_s}{R_1} = q - q_s \tag{5.37}$$

Introducing deviation variables $Q = q - q_s$ and $H = h - h_s$ into Eq. (5.37) and transforming give

$$\frac{H(s)}{Q(s)} = \frac{R_1}{\tau s + 1} \tag{5.38}$$

where

$$R_1 = \frac{2h_s^{1/2}}{C} \qquad \tau = R_1 A$$

We see that a transfer function is obtained that is identical in form with that of the linear system, Eq. (5.8). However, in this case, the resistance R_1 depends on the steady-state conditions around which the process operates. Graphically, the resistance R_1 is the reciprocal of the slope of the tangent line passing through the point (q_{os}, h_s), as shown in
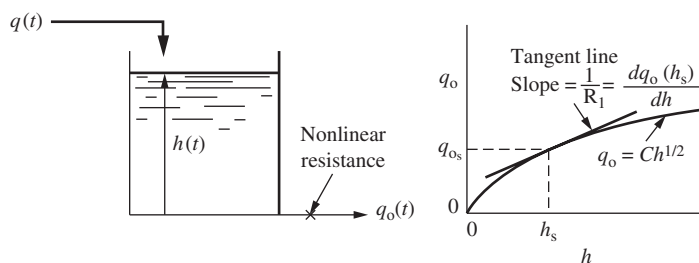
FIGURE 5–8
Liquid-level system with nonlinear resistance.

Fig. 5–8. Furthermore, the linear approximation given by Eq. (5.35) is the equation of the tangent line itself. From the graphical representation, it should be clear that the linear approximation improves as the deviation in h becomes smaller. If one does not have an analytic expression such as $h^{1/2}$ for the nonlinear function, but only a graph of the function, the technique can still be applied by representing the function by the tangent line passing through the point of operation.

Whether or not the linearized result is a valid representation depends on the operation of the system. If the level is being maintained by a controller, at or close to a fixed level h_s, then by the very nature of the control imposed on the system, deviations in level should be small (for good control) and the linearized equation is adequate. On the other hand, if the level should change over a wide range, the linear approximation may be very poor and the system may deviate significantly from the prediction of the linear transfer function. In such cases, it may be necessary to use the more difficult methods of nonlinear analysis, some of which are discussed in Chaps. 24 and 25. We shall extend the discussion of linearization to more complex systems in Chap. 20.

Using MATLAB to Compare Nonlinear (Exact) Solutions and Linearized Solutions

For the tank draining models of Eqs. (5.32) and (5.38) we have the following systems:

Nonlinear model
$$q - Ch^{1/2} = A\frac{dh}{dt} \tag{5.32}$$

Linearized model
$$\frac{H(s)}{Q(s)} = \frac{R_1}{\tau s + 1} \tag{5.38}$$

where
$$R_1 = \frac{2h_s^{1/2}}{C}$$
$$\tau = R_1 A$$

Consider the case where $A = 3$ ft^2 and the steady-state height is 4 ft when the inlet flow is 16 cfm. Compare the linearized and nonlinear (exact) solutions for the transient response of the tank height to a step change in feed flow from 16 to 20 cfm.

Solution:

From Eq. (5.30),

$$q_{os} = Ch_s^{1/2}$$

$$16 \text{ cfm} = C(4 \text{ ft})^{1/2}$$

thus,

$$C = 8 \text{ cfm}/\text{ft}^{1/2}$$

and

$$R_1 = \frac{2h_s^{1/2}}{C} = \frac{2\sqrt{4 \text{ ft}}}{8 \text{ cfm/ft}^{1/2}} = 0.5\frac{\text{ft}}{\text{cfm}}$$

$$\tau = R_1 A = \left(0.5\frac{\text{ft}}{\text{cfm}}\right)(3 \text{ ft}^2) = 1.5 \text{ min}$$

Substituting the numerical values into the nonlinear model, Eq. (5.32), yields

$$20 - 8\sqrt{h} = 3\frac{dh}{dt} \qquad h(0) = 4 \text{ ft}$$

The MATLAB m-file necessary to simulate this equation is shown below.

```
% FILENAME is level.m
  function  hprime=level(t,h)
  hprime=(20—8*sqrt(h))/3;
```

This file calculates the derivative dh/dt at any given t and h. We call the m-file using the numerical differential equation solver ODE45.

```
>> [t,h] = ode45(@level  ,   [0,10]  ,   [4]  );
```

Matrices that you want the answers returned into.	MATLAB routine to numerically solve ODE	m-file level.m contains the model	time span... $t_{initial}$ to t_{final}	initial condition $h(0)$ must correspond to $t_{initial}$

The linearized model, with numerical values substituted in, is

$$\frac{H(s)}{Q(s)} = \frac{0.5}{1.5s + 1}$$

where

$$H = h - 4 \qquad Q = q - 16 \qquad \text{deviation variables}$$

$$H(s) = \underbrace{\left(\frac{4}{s}\right)}_{\substack{\text{step change} \\ Q \text{ by 4 cfm}}} \left(\frac{0.5}{1.5s + 1}\right) = \frac{2}{s\,(1.5s + 1)}$$

Inverting gives

$$H(t) = 2\left(1 - e^{-t/1.5}\right) \qquad \text{linearized solution}$$

$$h(t) = 4 + 2\left(1 - e^{-t/1.5}\right)$$

Entering this equation into MATLAB yields

```
>>   hlin = 4 + 2*(1 − exp (−t/1.5));
```

Plot the linearized and nonlinear solutions on the same axes using MATLAB (see Fig. 5–9).

```
>>plot(t,h,t,hlin);
>> xlabel('Time');
>> ylabel('h');
>> title('Comparison of Nonlinear and Linearized Solutions');
```
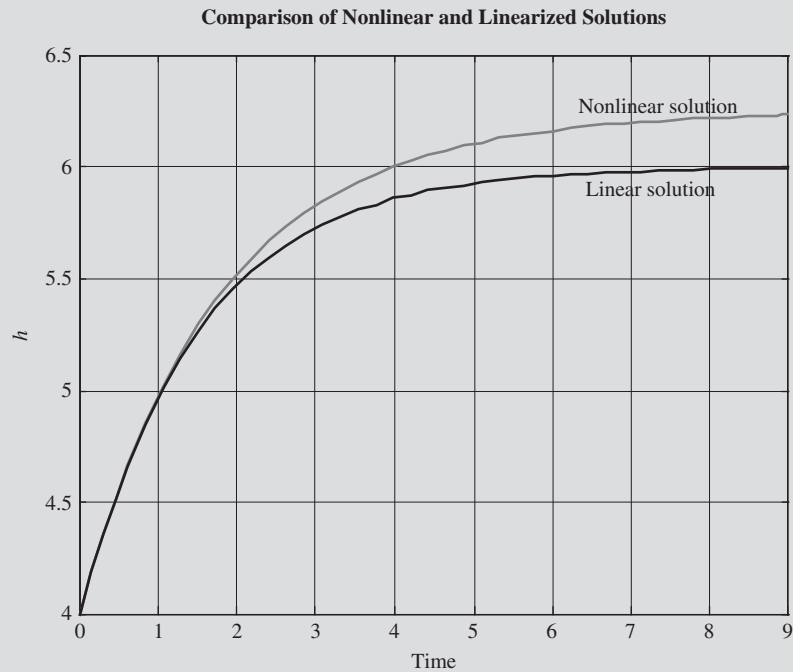

FIGURE 5–9
Comparison of nonlinear and linearized solutions for tank draining.

In general, the linearization of a nonlinear function is accomplished using a Taylor series expansion truncated to include only the linear terms. Thus for a single-variable function

$$f(x) = f(x_s) + \frac{df}{dx}\bigg|_{x_s} (x - x_s) + \text{(higher-order terms)} \tag{5.39}$$

For functions of two variables, we have

$$f(x,y) = f(x_s,y_s) + \frac{\partial f}{\partial x}\bigg|_{(x_s,y_s)} (x - x_s) + \frac{\partial f}{\partial y}\bigg|_{(x_s,y_s)} (y - y_s)$$
$$+ \text{(higher-order terms)} \tag{5.40}$$

Consider the differential equation describing the dynamics of a system

$$\frac{dy}{dt} + \underbrace{f(y)}_{\substack{\text{nonlinear} \\ \text{term}}} = x(t) \tag{5.41}$$

Linearizing the nonlinear term gives

$$\frac{dy}{dt} + \underbrace{f(y_s) + \frac{\partial f}{\partial y}\bigg|_{y_s} (y - y_s)}_{\text{linearized approximation}} = x(t) \tag{5.42}$$

Writing this equation again for the steady-state case gives

$$\frac{dy_s}{dt} + f(y_s) + \frac{\partial f}{\partial y}\bigg|_{y_s} (y_s - y_s) = x_s \tag{5.43}$$

Subtracting the steady-state case in Eq. (5.43) from Eq. (5.42), we can convert the original differential equation to deviation variables:

$$\frac{d(y - y_s)}{dt} + \frac{\partial f}{\partial y}\bigg|_{y_s} (y - y_s) = x - x_s$$

$$\frac{dY}{dt} + \frac{\partial f}{\partial y}\bigg|_{y_s} Y = X$$

where $X = x - x_s$ and $Y = y - y_s$. Note that the $f(y_s)$ term is eliminated in the process of forming deviation variables, and we are left with only linear terms in the equation which is now amenable to solution using Laplace transforms.

SUMMARY

In this chapter, we demonstrated several physical examples of first-order systems. Transfer functions were developed for those physical systems and placed into the standard form for first-order systems. We will see more examples of first-order systems as we discuss control systems in later chapters.

We have also characterized, in an approximate sense, a nonlinear system by a linear transfer function. In general, this technique may be applied to any nonlinearity that can be expressed in a Taylor series (or, equivalently, has a unique slope at the operating point). Since this includes most nonlinearities arising in process control, we have ample justification for studying linear systems in considerable detail.

PROBLEMS

5.1. Derive the transfer function $H(s)/Q(s)$ for the liquid-level system of Fig. P5–1 when

(a) The tank level operates about the steady-state value of $h_s = 1$ ft

FIGURE P5–1

FIGURE P5–3

FIGURE P5–4

(b) The tank level operates about the steady-state value of $h_s = 3$ ft
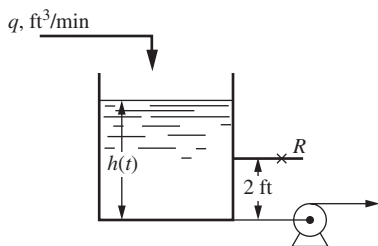
The pump removes water at a constant rate of 10 cfm (cubic feet per minute); this rate is independent of head. The cross-sectional area of the tank is 1.0 ft^2, and the resistance R is 0.5 ft/cfm.

5.2. A liquid-level system, such as the one shown in Fig. 5–1, has a cross-sectional area of 3.0 ft^2. The valve characteristics are
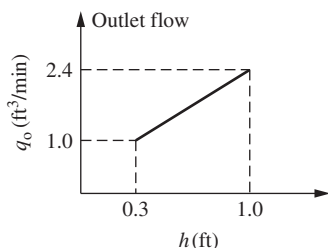
$$q = 8\sqrt{h}$$

where q = flow rate, cfm, and h = level above the valve, ft. Calculate the time constant for this system if the average operating level above the valve is
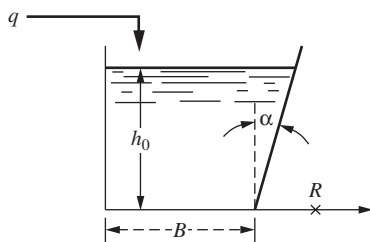
(a) 3 ft

(b) 9 ft

5.3. A tank having a cross-sectional area of 2 ft^2 is operating at steady state with an inlet flow rate of 2.0 cfm. The flow-head characteristics are shown in Fig. P5–3.

(a) Find the transfer function $H(s)/Q(s)$.

(b) If the flow to the tank increases from 2.0 to 2.2 cfm according to a step change, calculate the level h two minutes after the change occurs.

5.4. Develop a formula for finding the time constant of the liquid-level system shown in Fig. P5–4 when the average operating level is h_0. The resistance R is linear. The tank has three vertical walls and one that slopes at an angle α from the vertical as shown. The distance separating the parallel walls is 1.

C_i, F

Volume $\bar{V}$

C_o, F
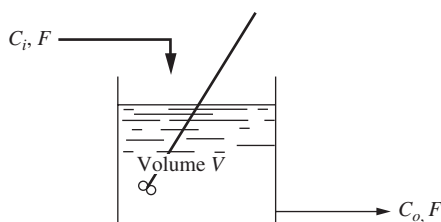
FIGURE P5–5

5.5. Consider the stirred-tank reactor shown in Fig. P5–5.

The reaction occurring is

$$A \rightarrow B$$

and it proceeds at a rate

$$r = kC_o$$

where r = (moles A reacting)/(volume)(time)

   k = reaction rate constant

   $C_o(t)$ = concentration of A in reactor at any time t (mol A/volume)

   V = volume of mixture in reactor

Further, let

$$F = \text{constant feed rate, volume/time}$$

$$C_i(t) = \text{concentration of } A \text{ in feed stream, moles/volume}$$

Assuming constant density and constant volume V, derive the transfer function relating the concentration in the reactor to the feed-stream concentration. Prepare a block diagram for the reactor. Sketch the response of the reactor to a unit-step change in C_i.

5.6. A thermocouple junction of area A, mass m, heat capacity C, and emissivity e is located in a furnace that normally is at $T_{i_s}°C$. At these temperatures convective and conductive heat transfer to the junction is negligible compared with radiative heat transfer. Determine the linearized transfer function between the furnace temperature T_i and the junction temperature T_0. For the case

$$m = 0.1 \text{ g}$$

$$C = 0.12 \text{ cal/(g} \cdot °C)$$

$$e = 0.7$$

$$A = 0.1 \text{ cm}^2$$

$$T_{i_s} = 1100°C$$

plot the response of the thermocouple to a 10°C step change in furnace temperature. Compare this with the true response obtained by integration of the differential equation.

5.7. A liquid-level system has the following properties:

Tank dimensions: 10 ft high by 5-ft diameter

Steady-state operating characteristics:

Inflow, gal/h	Steady-state level, ft
0	0
5,000	0.7
10,000	1.1
15,000	2.3
20,000	3.9
25,000	6.3
30,000	8.8

(a) Plot the level response of the tank under the following circumstances: The inlet flow rate is held at 300 gal/min for 1 h and then suddenly raised to 400 gal/min.

(b) How accurate is the steady-state level calculated from the dynamic response in part (a) when compared with the value given by the table above?

(c) The tank is now connected in series with a second tank that has identical operating characteristics, but which has dimensions 8 ft high by 4-ft diameter. Plot the response of the original tank (which is upstream of the new tank) to the change described in part (a) when the connection is such that the tanks are (1) interacting and (2) noninteracting. (See Chap. 6.)

5.8. A mixing process may be described as follows: A stream with solute concentration C_i (pounds/volume) is fed to a perfectly stirred tank at a constant flow rate of q (volume/time). The perfectly mixed product is withdrawn from the tank, also at the flow rate q at the same concentration as the material in the tank C_0. The total volume of solution in the tank is constant at V. Density may be considered to be independent of concentration.
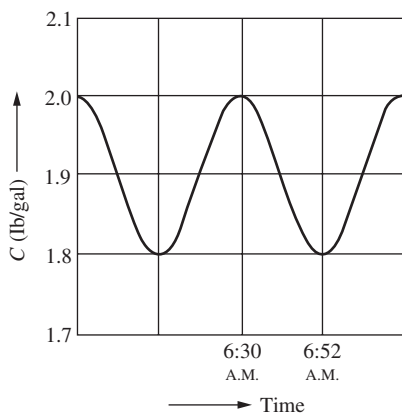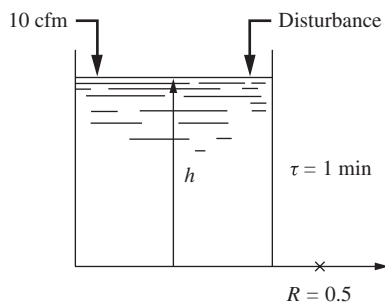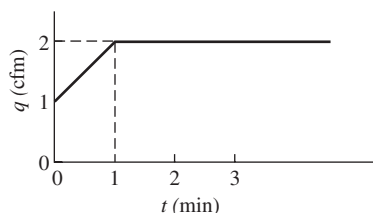
A trace of the tank concentration versus time appears as shown in Fig. P5–8.

(a) Plot on this same figure your best guess of the *quantitative* behavior of the inlet concentration versus time. Be sure to label the graph with *quantitative* information regarding times and magnitudes and any other data that will demonstrate your understanding of the situation.

(b) Write an equation for C_i as a function of time.

Data: tank dimensions: 8 ft high by 5-ft diameter
Tank volume V: 700 gal
Flow rate q: 100 gal/min
Average density: 70 lb/ft^3

FIGURE P5–8

FIGURE P5–9

FIGURE P5–10

5.9. The liquid-level process shown in Fig. P5–9 is operating at steady state when the following disturbance occurs: At time $t = 0$, 1 ft^3 water is added suddenly (unit impulse) to the tank; at $t = 1$ min, 2 ft^3 of water is added suddenly to the tank. Sketch the response of the level in the tank versus time, and determine the level at $t = 0.5$, 1, and 1.5 min.

5.10. A tank having a cross-sectional area of 2 ft^2 and a linear resistance of $R = 1$ ft/cfm is operating at steady state with a flow rate of 1 cfm. At time 0, the flow varies as shown in Fig. P5–10.

(a) Determine $Q(t)$ and $Q(s)$ by combining simple functions. Note that Q is the deviation in flow rate.

(b) Obtain an expression for $H(t)$ where H is the deviation in level.
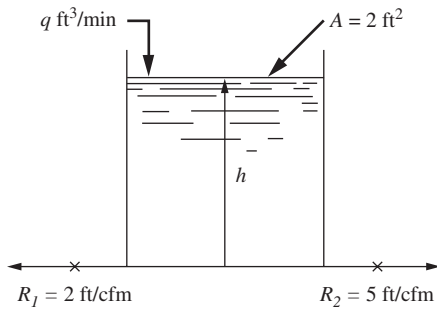
(c) Determine $H(t)$ at $t = 2$ and $t = \infty$.

FIGURE P5–12

5.11. Determine $y(t = 5)$ if
$Y(s) = e^{-3s}/s(7s + 1)$.

5.12. Derive the transfer function H/Q for the liquid-level system shown in Fig. P5–12. The resistances are linear; H and Q are deviation variables. Show clearly how you derived the transfer function. You are expected to give numerical values in the transfer function.

5.13. The liquid-level system shown in Fig. P5–13 is initially at steady state with the inlet flow rate at 1 cfm. At time 0, 1 ft^3 of water is suddenly added to the tank; at $t = 1$, 1 ft^3 is added; etc. In other words, a train of unit impulses is applied to the tank at intervals of 1 min. Ultimately the output wave train becomes periodic as shown in the sketch. Determine the maximum and minimum values of this output.
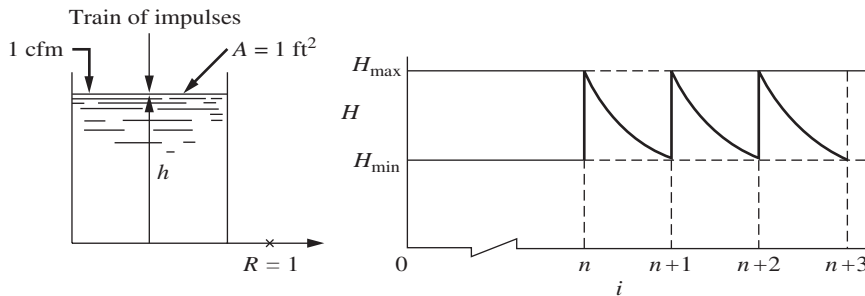

FIGURE P5–13

5.14. The two-tank mixing process shown in Fig. P5–14 contains a recirculation loop that transfers solution from tank 2 to tank 1 at a flow rate of αq_o.
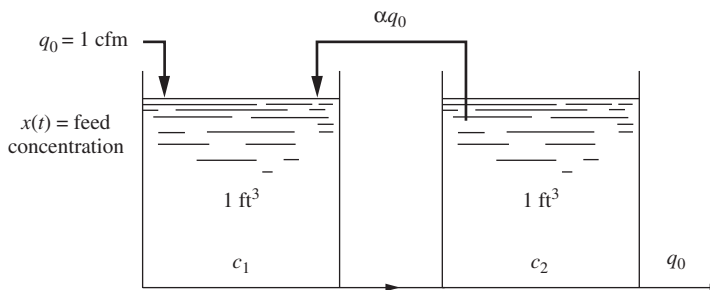

FIGURE P5–14

(a) Develop a transfer function that relates the concentration c_2 in tank 2 to the concentration x in the feed, that is, $C_2(s)/X(s)$ where C_2 and X are deviation variables. For convenience, assume that the initial concentrations are $x = c_1 = c_2 = 0$.

(b) If a unit-step change in x occurs, determine the time needed for c_2 to reach 60 percent of its ultimate value for the cases where $\alpha = 0$, 1, and ∞.

(c) Sketch the response for $\alpha = \infty$.

Assume that each tank has a constant holdup volume of 1 ft³. Neglect transportation lag in the line connecting the tanks and the recirculation line. Try to answer parts (b) and (c) by intuition.

5.15. Dye for our new line of blue jeans is being blended in a mixing tank. The desired color of blue is produced using a concentration of 1500 ppm blue dye, with a minimum acceptable concentration of 1400 ppm. At 9 A.M. today the dye injector plugged, and the dye flow was interrupted for 10 min, until we realized the problem and unclogged the nozzle. For how many minutes was the flow leaving the mixer off-specification (< 1400 ppm)? How many gallons of off-spec dye were made? See Fig. P5–15.
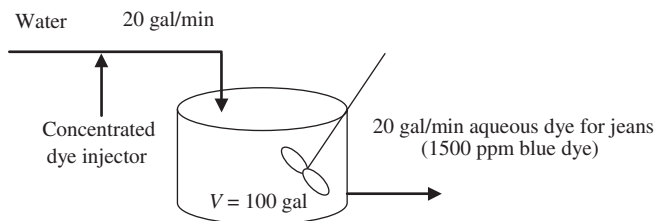

FIGURE P5–15

5.16. For the reactor (CSTR) shown in Fig. P5–16, determine the transfer function that relates the exit concentration from the reactor to changes in the feed concentration. If we instantaneously double the feed concentration from 1 to 2 mol/L, what is the new exiting concentration 1 min later? What is the new steady-state reactor concentration?

The rate constant is

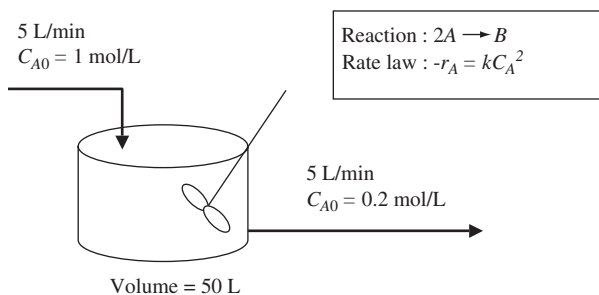$$k = \frac{2}{(\text{mol/L})(\text{min})}$$

FIGURE P5–16

The reaction rate law is $-r_A = kC_A{}^2$, where r_A is the production rate of A in moles per liter per minute.

5.17. The Antoine equation for the vapor pressure of a liquid at a given temperature is given by

$$P^* = e^{A - B/(T + C)}$$

The constants for benzene are

$$A = 15.9008$$
$$B = 2788.51°C$$
$$C = 220.80°C$$

for the vapor pressure in millimeters of mercury (mmHg). Linearize the equation about a temperature of 40°C.

Compare the actual vapor pressure (from the Antoine equation) at 45 and 60°C with the vapor pressure calculated from the linearized equation. What is the percent difference in each case? Comment on the suitability of the linearized equation.

5.18. Find the transfer function that relates the height in the vessel (Fig. P5–18) to changes in the inlet flow rate.
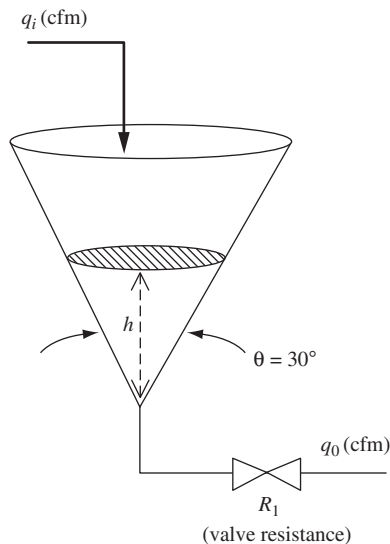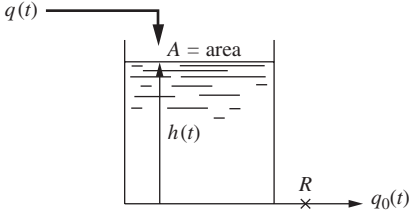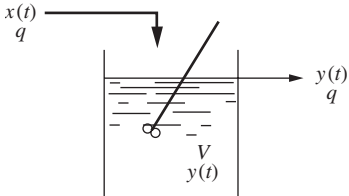

q_i (cfm)

h

$\theta = 30°$

q_0 (cfm)

R_1
(valve resistance)
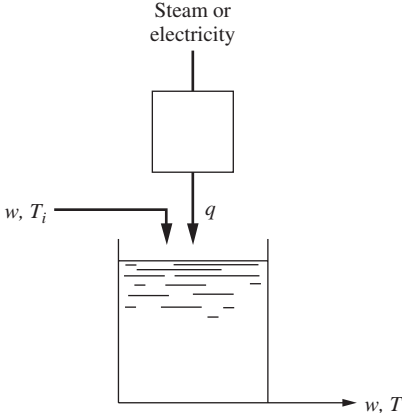
FIGURE P5–18
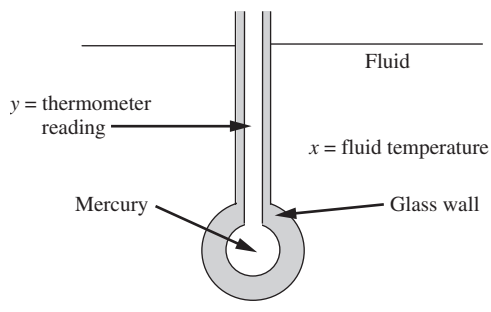
CAPSULE SUMMARY

Here are some physical examples of first-order systems:

System		Transfer function

Figure 5–1 Liquid level system

$$\frac{H(s)}{Q(s)} = \frac{R}{ARs + 1}$$

Figure 5–4 Mixing process

$$\frac{Y(s)}{X(s)} = \frac{1}{(V/q)s + 1}$$

Figure 5–5 Heating process

$$\frac{T'(s)}{Q(s)} = \frac{1/(wC)}{(\rho V/w)s + 1}$$

(continued)

System		Transfer function
Fluid y = thermometer reading x = fluid temperature Mercury — Glass wall	Figure 4–1a Thermometer	$\dfrac{Y(s)}{X(s)} = \dfrac{1}{(mc/hA)s + 1}$

TAYLOR SERIES EXPANSIONS FOR LINEARIZING NONLINEAR TERMS

Functions of a single variable: $f(x) = f(x_s) + \left.\dfrac{df}{dx}\right|_{x_s} (x - x_s)$

Functions of two variables:

$$f(x, y) = f(x_s, y_s) + \left.\frac{\partial f}{\partial x}\right|_{(x_s, y_s)} (x - x_s) + \left.\frac{\partial f}{\partial y}\right|_{(x_s, y_s)} (y - y_s)$$

# RESPONSE OF FIRST-ORDER SYSTEMS IN SERIES

6.1 INTRODUCTORY REMARKS

Very often, a physical system can be represented by several first-order processes connected in series. To illustrate this type of system, consider the liquid-level systems shown in Fig. 6–1 in which two tanks are arranged so that the outlet flow from the first tank is the inlet flow to the second tank.

Two possible piping arrangements are shown in Fig. 6–1. In Fig. 6–1a the outlet flow from tank 1 discharges directly into the atmosphere before spilling into tank 2, and the flow through R_1 depends only on h_1. The variation in h_2 in tank 2 does not affect the transient response occurring in tank 1. This type of system is referred to as a *noninteracting* system. In contrast to this, the system shown in Fig. 6–1b is said to be *interacting* because the flow through R_1 now depends on the difference between h_1 and h_2. We will consider first the noninteracting system of Fig. 6–1a.

6.2 NONINTERACTING SYSTEM

As in the previous liquid-level example, we shall assume the liquid to be of constant density, the tanks to have uniform cross-sectional area, and the flow resistances to be linear. Our problem is to find a transfer function that relates h_2 to q, that is, $H_2(s)/Q(s)$. The approach will be to obtain a transfer function for each tank, $Q_1(s)/Q(s)$ and $H_2(s)/Q_1(s)$, by writing a transient mass balance around each tank; these transfer functions will then be combined to eliminate the intermediate flow $Q_1(s)$ and produce the desired transfer function. A balance on tank 1 gives
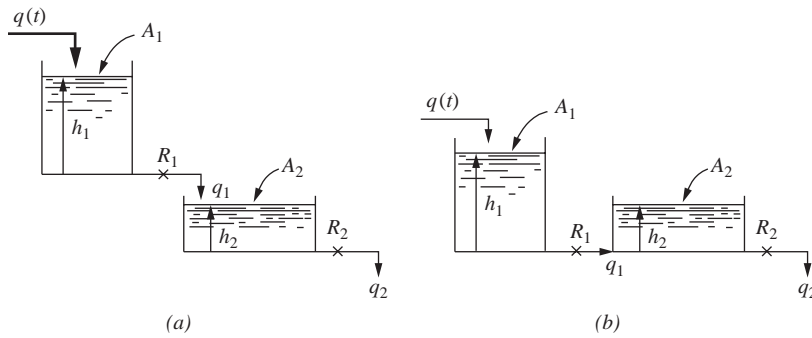
$$q - q_1 = A_1 \frac{dh_1}{dt} \tag{6.1}$$

Two-tank liquid-level system: *(a)* Noninteracting; *(b)* interacting.

A balance on tank 2 gives

$$q_1 - q_2 = A_2 \frac{dh_2}{dt} \tag{6.2}$$

The flow-head relationships for the two linear resistances are given by the expressions

$$q_1 = \frac{h_1}{R_1} \tag{6.3}$$

$$q_2 = \frac{h_2}{R_2} \tag{6.4}$$

Combining Eqs. (6.1) and (6.3) in exactly the same manner as was done in Chap. 5 and introducing deviation variables give the transfer function for tank 1

$$\frac{Q_1(s)}{Q(s)} = \frac{1}{\tau_1 s + 1} \tag{6.5}$$

where $Q_1 = q_1 - q_{1s}$, $Q = q - q_s$, and $\tau_1 = R_1 A_1$.

In the same manner, we can combine Eqs. (6.2) and (6.4) to obtain the transfer function for tank 2

$$\frac{H_2(s)}{Q_1(s)} = \frac{R_2}{\tau_2 s + 1} \tag{6.6}$$

where $H_2 = h_2 - h_{2s}$ and $\tau_2 = R_2 A_2$.

Having the transfer function for each tank, we can obtain the overall transfer function $H_2(s)/Q(s)$ by multiplying Eqs. (6.5) and (6.6) to eliminate $Q_1(s)$:

$$\frac{H_2(s)}{Q(s)} = \frac{1}{\tau_1 s + 1} \frac{R_2}{\tau_2 s + 1} \tag{6.7}$$

Notice that the overall transfer function of Eq. (6.7) is the product of two first-order transfer functions, each of which is the transfer function of a single tank operating independently

of the other. In the case of the interacting system of Fig. 6–1b, the overall transfer function *cannot* be found by simply multiplying the separate transfer functions; this will become apparent when the interacting system is analyzed later.

Example 6.1.   Two noninteracting tanks are connected in series as shown in Fig. 6–1a. The time constants are $\tau_2 = 1$ and $\tau_1 = 0.5$; $R_2 = 1$. Sketch the response of the level in tank 2 if a unit-step change is made in the inlet flow rate to tank 1.

The transfer function for this system is found directly from Eq. (6.7); thus

$$\frac{H_2(s)}{Q(s)} = \frac{R_2}{(\tau_1 s + 1)(\tau_2 s + 1)} \qquad (6.8)$$

For a unit-step change in Q, we obtain

$$H_2(s) = \frac{1}{s} \frac{R_2}{(\tau_1 s + 1)(\tau_2 s + 1)} \qquad (6.9)$$

Inversion by means of partial fraction expansion gives

$$H_2(t) = R_2 \left[1 - \frac{\tau_1 \tau_2}{\tau_1 - \tau_2} \left(\frac{1}{\tau_2} e^{-t/\tau_1} - \frac{1}{\tau_1} e^{-t/\tau_2} \right) \right] \qquad (6.10)$$

Substituting in the values of τ_1, τ_2, and R_2 gives

$$H_2(t) = 1 - \left(2e^{-t} - e^{-2t} \right) \qquad (6.11)$$

A plot of this response is shown in Fig. 6–2. Notice that the response is S-shaped and the slope dH_2/dt at the origin is zero. If the change in flow rate were introduced into the second tank, the response would be first-order and is shown for comparison in Fig. 6–2 by the dotted curve.
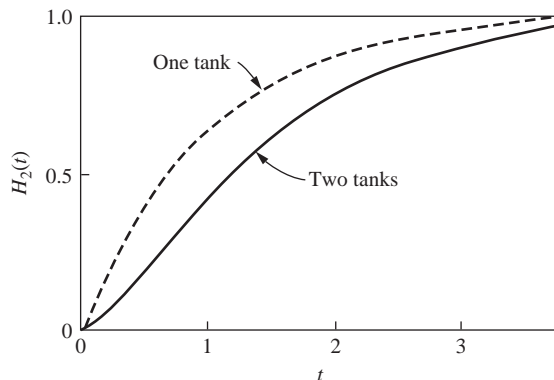

FIGURE 6–2
Transient response of liquid-level system (Example 6.1).

MATLAB/Simulink Simulation of the Transient Response in Fig. 6–2

It's quite easy to verify the result in Fig. 6–2 using Simulink (Fig. 6–3). The model is shown below at the left, and the output from the model is shown in the graph.
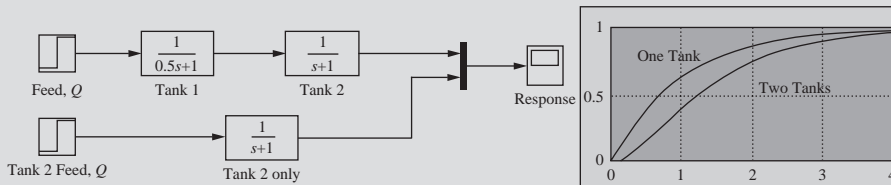

FIGURE 6–3
Simulink simulation of one tank and two tanks in series.

From Example 6.1, notice that the step response of a system consisting of two first-order systems is S-shaped and that the response changes very slowly just after introduction of the step input. This sluggishness or delay is sometimes called *transfer lag* and is always present when two or more first-order systems are connected in series. For a single first-order system, there is no transfer lag; i.e., the response begins immediately after the step change is applied, and the rate of change of the response (slope of response curve) is maximal at $t = 0$.

Generalization for Several Noninteracting Systems in Series

We have observed that the overall transfer function for two noninteracting first-order systems connected in series is simply the product of the individual transfer functions. We may now generalize this concept by considering n noninteracting first-order systems as represented by the block diagram of Fig. 6–4.
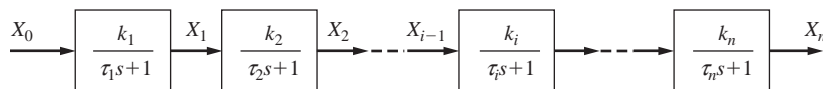

FIGURE 6–4
Noninteracting first-order systems.

The block diagram is equivalent to the relationships

$$\frac{X_1(s)}{X_0(s)} = \frac{k_1}{\tau_1 s + 1}$$

$$\frac{X_2(s)}{X_1(s)} = \frac{k_2}{\tau_2 s + 1}$$

etc.

$$\frac{X_n(s)}{X_{n-1}(s)} = \frac{k_n}{\tau_n s + 1}$$

To obtain the overall transfer function, we simply multiply the individual transfer functions; thus

$$\frac{X_n(s)}{X_0(s)} = \prod_{i=1}^{n} \frac{k_i}{\tau_i s + 1} \qquad (6.12)$$

To show how the transfer lag is increased as the number of stages increases, Fig. 6–5 gives the unit-step response curves for several systems containing one or more first-order stages in series.

FIGURE 6–5
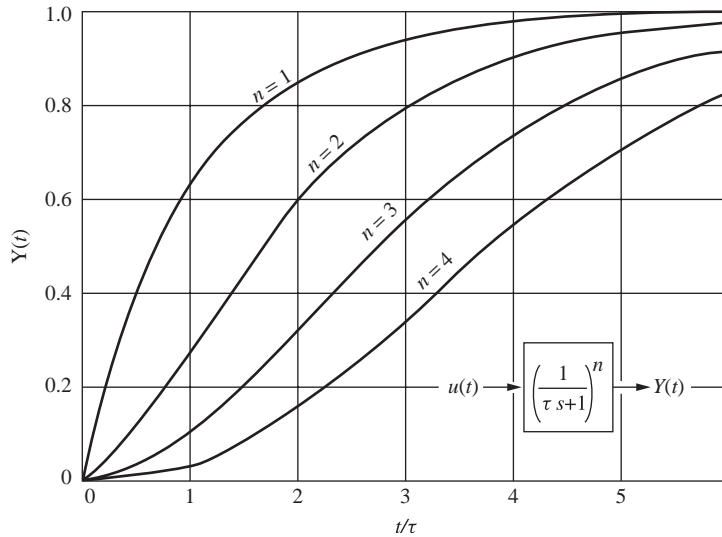Step response of noninteracting first-order systems in series.

MATLAB/Simulink Simulation of Noninteracting First-Order Systems in Fig. 6–5

Let's reproduce the response in Fig. 6–5 for four tanks in series using Simulink (Fig. 6–6). Note how simple the model is and that the result is identical to Fig. 6–5.
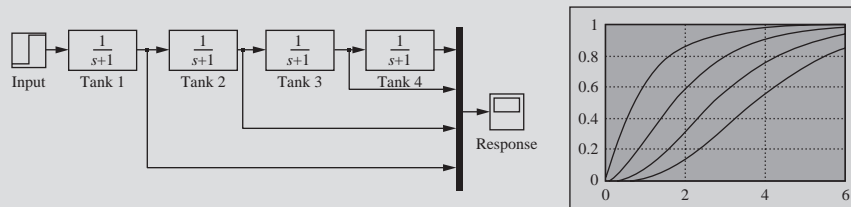

FIGURE 6–6
Simulink diagram for noninteracting first-order systems in series.

6.3   INTERACTING SYSTEM

To illustrate an interacting system, we will derive the transfer function for the system shown in Fig. 6–1*b*. The analysis is started by writing mass balances on the tanks as was done for the noninteracting case. The balances on tanks 1 and 2 are the same as before and are given by Eqs. (6.1) and (6.2).

$$\text{Tank 1} \qquad q - q_1 = A_1 \frac{dh_1}{dt} \qquad\qquad (6.1)$$

$$\text{Tank 2} \qquad q_1 - q_2 = A_2 \frac{dh_2}{dt} \qquad\qquad (6.2)$$

However, the flow-head relationship for R_1 is now

$$q_1 = \frac{1}{R_1}(h_1 - h_2) \qquad\qquad (6.13)$$

The flow-head relationship for R_2 is the same as before [Eq. (6.4)].

$$q_2 = \frac{h_2}{R_2} \qquad\qquad (6.4)$$

A simple way to combine Eqs. (6.1), (6.2), (6.4), and (6.13) is to first express them in terms of deviation variables, transform the resulting equations, and then combine the transformed equations to eliminate the unwanted variables.

At steady state, Eqs. (6.1) and (6.2) can be written

$$q_s - q_{1s} = 0 \qquad\qquad (6.14)$$

$$q_{1s} - q_{2s} = 0 \qquad\qquad (6.15)$$

Subtracting Eq. (6.14) from Eq. (6.1) and Eq. (6.15) from Eq. (6.2) and introducing deviation variables give

$$\text{Tank 1} \qquad Q - Q_1 = A_1 \frac{dH_1}{dt} \qquad\qquad (6.16)$$

$$\text{Tank 2} \qquad Q_1 - Q_2 = A_2 \frac{dH_2}{dt} \qquad\qquad (6.17)$$

Expressing Eqs. (6.13) and (6.4) in terms of deviation variables gives

$$\text{Valve 1} \qquad Q_1 = \frac{H_1 - H_2}{R_1} \qquad\qquad (6.18)$$

$$\text{Valve 2} \qquad Q_2 = \frac{H_2}{R_2} \qquad\qquad (6.19)$$

Transforming Eqs. (6.16) through (6.19) gives

$$\text{Tank 1} \qquad Q(s) - Q_1(s) = A_1 s H_1(s) \qquad\qquad (6.20)$$

$$\text{Tank 2} \qquad Q_1(s) - Q_2(s) = A_2 s H_2(s) \qquad\qquad (6.21)$$

$$\text{Valve 1} \qquad R_1 Q_1(s) = H_1(s) - H_2(s) \qquad\qquad (6.22)$$

$$\text{Valve 2} \qquad R_2 Q_2(s) = H_2(s) \qquad\qquad (6.23)$$

The analysis has produced four algebraic equations containing five unknowns: Q, Q_1, Q_2, H_1, and H_2. These equations may be combined to eliminate Q_1, Q_2, and H_1 and to arrive at the desired transfer function:

$$\frac{H_2(s)}{Q(s)} = \frac{R_2}{\tau_1 \tau_2 s^2 + (\tau_1 + \tau_2 + A_1 R_2)s + 1} \qquad\qquad (6.24)$$

Notice that the product of the transfer functions for the tanks operating separately, Eqs. (6.5) and (6.6), does not produce the correct result for the interacting system. The difference between the transfer function for the noninteracting system, Eq. (6.7), and that for the interacting system, Eq. (6.24), is the presence of the cross-product term $A_1 R_2$ in the coefficient of s.

The term *interacting* is often referred to as *loading*. The second tank of Fig. 6–1b is said to *load* the first tank.

To understand the effect of interaction on the transient response of a system, consider a two-tank system for which the time constants are equal ($\tau_1 = \tau_2 = \tau$). If the tanks are noninteracting, the transfer function relating inlet flow to outlet flow is

$$\frac{Q_2(s)}{Q(s)} = \left(\frac{1}{\tau s + 1}\right)^2 \qquad\qquad (6.25)$$

The unit-step response for this transfer function can be obtained by the usual procedure to give

$$Q_2(t) = 1 - e^{-t/\tau} - \frac{t}{\tau} e^{-t/\tau} \qquad\qquad (6.26)$$

If the tanks are interacting, the overall transfer function, according to Eq. (6.24), is (assuming further that $A_1 = A_2$)

$$\frac{Q_2(s)}{Q(s)} = \frac{1}{\tau^2 s^2 + 3\tau s + 1} \qquad\qquad (6.27)$$

By application of the quadratic formula, the denominator of this transfer function can be written as

$$\frac{Q_2(s)}{Q(s)} = \frac{1}{(0.38\tau s + 1)(2.62\tau s + 1)} \qquad\qquad (6.28)$$

For this example, we see that the effect of interaction has been to change the effective time constants of the interacting system. One time constant has become considerably larger and the other smaller than the time constant τ of either tank in the noninteracting system. The response of $Q_2(t)$ to a unit-step change in $Q(t)$ for the interacting case [Eq. (6.28)] is

$$Q_2(t) = 1 + 0.17e^{-t/0.38\tau} - 1.17e^{-t/2.62\tau} \qquad (6.29)$$

In Fig. 6–7, the unit-step responses [Eqs. (6.26) and (6.29)] for the two cases are plotted to show the effect of interaction. From this figure, it can be seen that interaction slows up the response. This result can be understood on physical grounds in the following way: If the same size step change is introduced into the two systems of Fig. 6–1, the flow from tank 1 (q_1) for the noninteracting case will not be reduced by the increase in level in tank 2. However, for the interacting case, the flow q_1 will be reduced by the buildup of level in tank 2. At any time t_1 following the introduction of the step input, q_1 for the interacting case will be less than for the noninteracting case with the result that h_2 (or q_2) will increase at a slower rate.
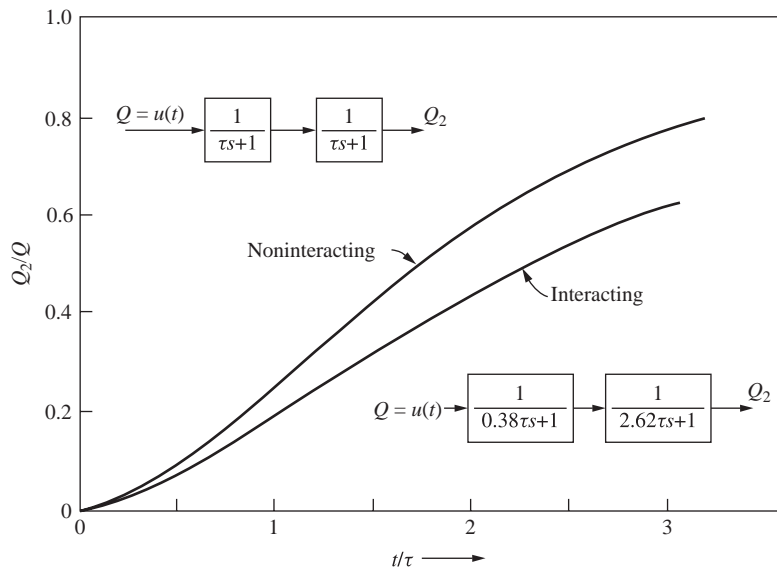

FIGURE 6–7
Effect of interaction on step response of two-tank system.

In general, the effect of interaction on a system containing two first-order lags is to change the ratio of effective time constants in the interacting system. In terms of the transient response, this means that the interacting system is more sluggish than the noninteracting system.

SUMMARY

In this chapter we discussed the response of first-order systems in series. We observed that the nature of the response is dependent upon whether the first-order systems in series form a noninteracting or an interacting system. We used MATLAB to visualize and analyze the response of these two different types of systems and to study their behavior.

This chapter concludes our specific discussion of first-order systems. We will make continued use of the material developed here in the succeeding chapters.

PROBLEMS

6.1. Determine the transfer function $H(s)/Q(s)$ for the liquid-level system shown in Fig. P6–1. Resistances R_1 and R_2 are linear. The flow rate from tank 3 is maintained constant at b by means of a pump; i.e., the flow rate from tank 3 is independent of head h. The tanks are noninteracting.
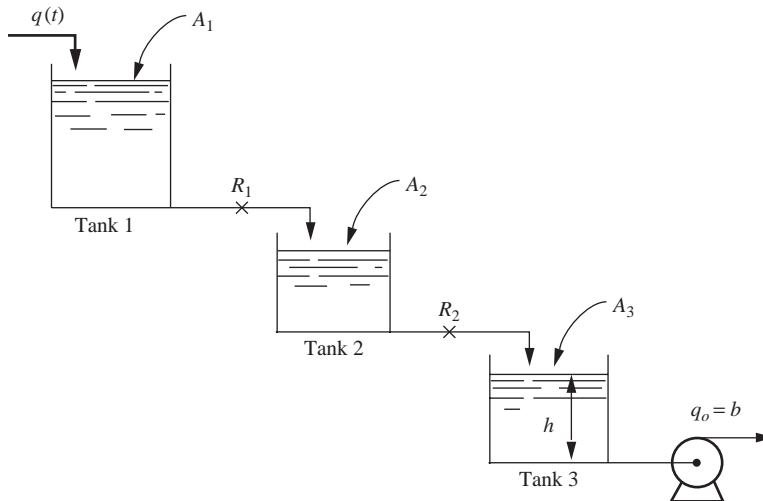

FIGURE P6–1

6.2. The mercury thermometer in Chap. 4 was considered to have all its resistance in the convective film surrounding the bulb and all its capacitance in the mercury. A more detailed analysis would consider both the convective resistance surrounding the bulb and that between the bulb and the mercury. In addition, the capacitance of the glass bulb would be included. Let

A_i = inside area of bulb, for heat transfer to mercury
A_o = outside area of bulb, for heat transfer from surrounding fluid
m = mass of mercury in bulb
m_b = mass of glass bulb

C = heat capacity of mercury
C_b = heat capacity of glass bulb
h_i = convective coefficient between bulb and mercury
h_o = convective coefficient between bulb and surrounding fluid
T = temperature of mercury
T_b = temperature of glass bulb
T_f = temperature of surrounding fluid

Determine the transfer function between T_f and T. What is the effect of the bulb resistance and capacitance on the thermometer response? Note that the inclusion of the bulb results in a pair of interacting systems, which give an overall transfer function somewhat different from that of Eq. (6.24).

6.3. There are N storage tanks of volume V arranged so that when water is fed into the first tank, an equal volume of liquid overflows from the first tank into the second tank, and so on. Each tank initially contains component A at some concentration C_o and is equipped with a perfect stirrer. At time 0, a stream of zero concentration is fed into the first tank at a volumetric rate q. Find the resulting concentration in each tank as a function of time.

6.4. (*a*) Find the transfer functions H_2/Q and H_3/Q for the three-tank system shown in Fig. P6–4 where H_2, H_3, and Q are deviation variables. Tank 1 and tank 2 are interacting.

   (*b*) For a unit-step change in q (that is, $Q = 1/s$), determine $H_3(0)$ and $H_3(\infty)$, and sketch $H_3(t)$ versus t.
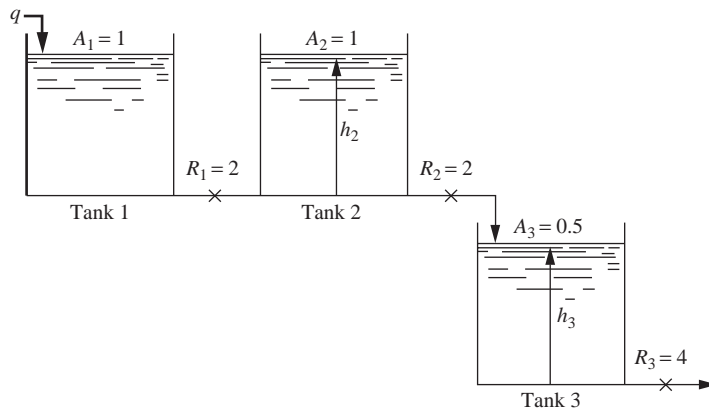

FIGURE P6–4

6.5. Three identical tanks are operated in series in a noninteracting fashion as shown in Fig. P6–5. For each tank, $R = 1$ and $\tau = 1$. The deviation in flow rate to the first tank is an impulse function of magnitude 2.

(*a*) Determine an expression for $H(s)$ where H is the deviation in level in the third tank.

(*b*) Sketch the response $H(t)$.
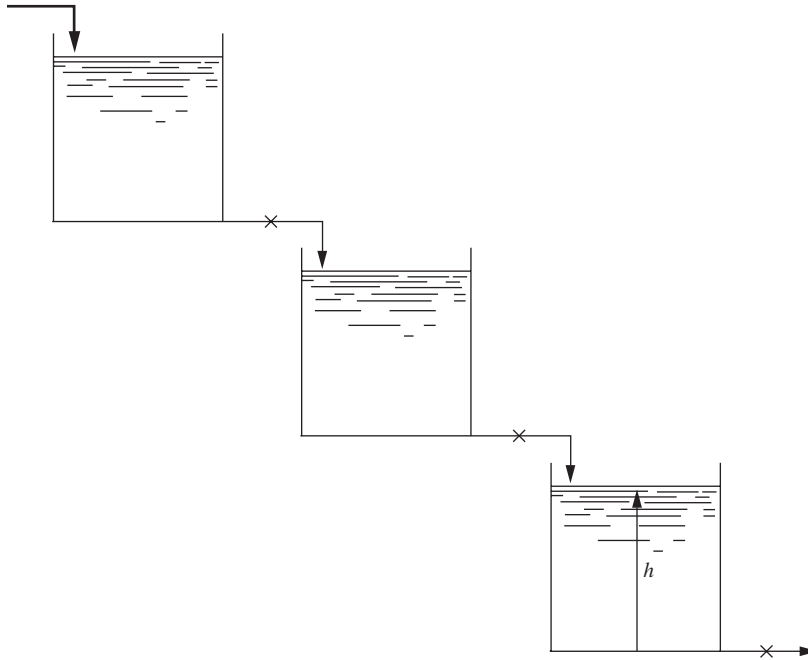
(*c*) Obtain an expression for $H(t)$.

FIGURE P6–5

6.6. In the two-tank mixing process shown in Fig. P6–6, x varies from 0 lb salt/ft^3 to 1 lb salt/ft^3 according to a step function. At what time does the salt concentration in tank 2 reach 0.6 lb salt/ft^3? The holdup volume of each tank is 6 ft^3.
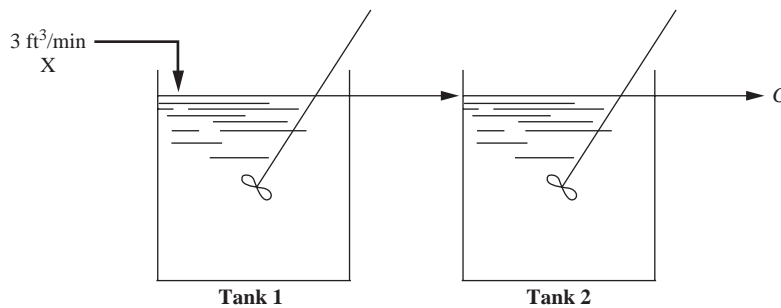
FIGURE P6–6

6.7. Starting from first principles, derive the transfer functions $H_1(s)/Q(s)$ and $H_2(s)/Q(s)$ for the liquid-level system shown in Fig. P6–7. The resistances are linear and $R_1 = R_2 = 1$. Note that two streams are flowing from tank 1, one of which flows into tank 2. You are expected to give numerical values of the parameters in the transfer functions and to show clearly how you derived the transfer functions.
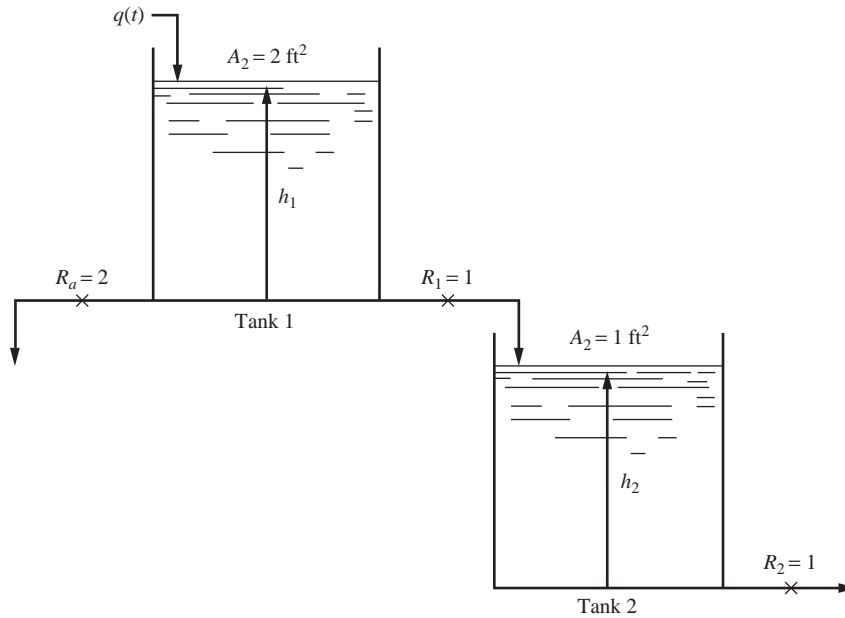
FIGURE P6–7

CAPSULE SUMMARY

Noninteracting systems For two systems in series, if the output from system 1 is not affected by the output from system 2, the systems are said to be *noninteracting*.

$$\frac{X_n(s)}{X_0(s)} = \prod_{i=1}^{n} G_i$$
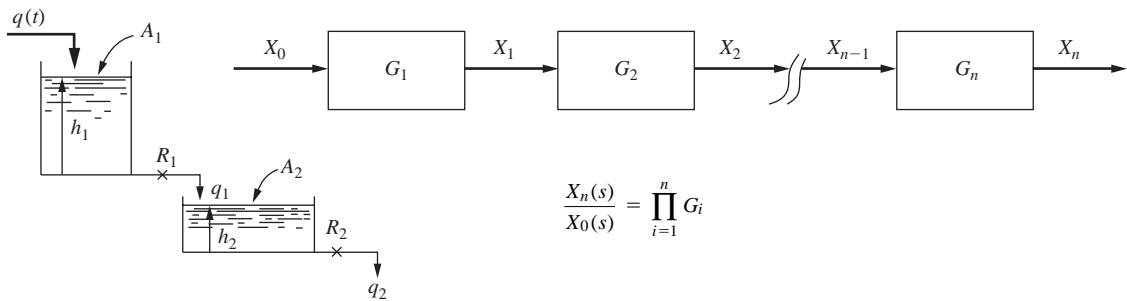
$$\frac{H_2(s)}{Q(s)} = \frac{1}{\tau_1 s + 1} \frac{R_2}{\tau_2 s + 1}$$

Interacting systems The output from system 1 *is* affected by the output from system 2. The overall transfer function for the process is *not* merely the product of the transfer functions in series.

$$\frac{X_n(s)}{X_0(s)} \neq \prod_{i=1}^{n} G_i$$
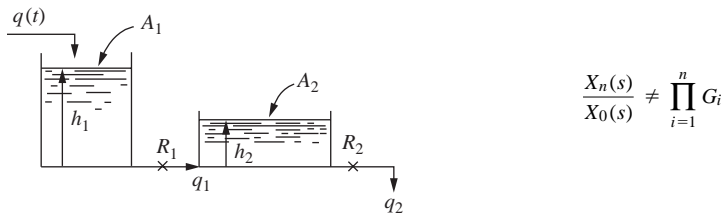
For the interacting two-tank system, the transfer function is

$$\frac{H_2(s)}{Q(s)} = \frac{R_2}{\tau_1\tau_2 s^2 + (\tau_1 + \tau_2 + A_1 R_2)s + 1}$$

Note the presence of the cross-product term in the denominator.

This term has the effect of slowing down the response of the process.

HIGHER-ORDER SYSTEMS: SECOND-ORDER AND TRANSPORTATION LAG

7.1 SECOND-ORDER SYSTEM

Transfer Function

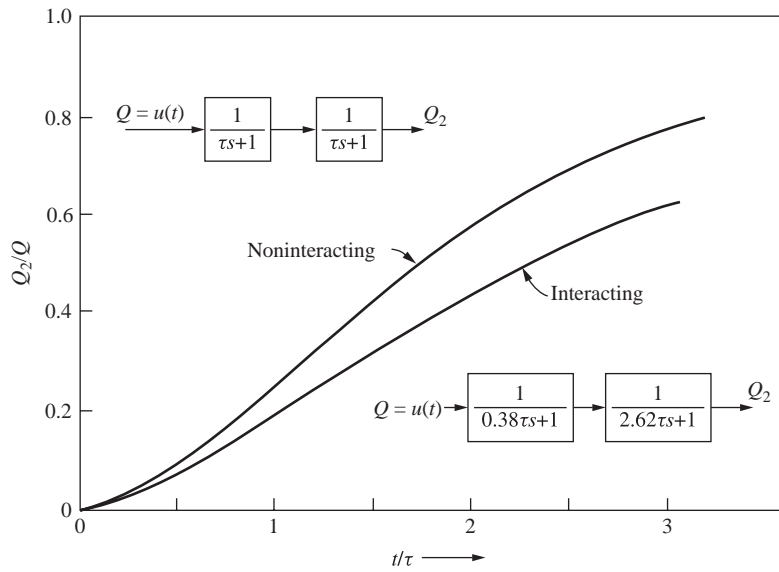This section introduces a basic system called a *second-order system* or a *quadratic lag*. Second-order systems are described by a second-order differential equation that relates the ouput variable y to the input variable x (the forcing function) with time as the independent variable.

$$A \frac{d^2 y}{dt^2} + B \frac{dy}{dt} + Cy = x(t) \tag{7.1}$$

A second-order system can arise from two first-order systems in series, as we saw in Chap. 6. Some systems are inherently second-order, and they do not result from a series combination of two first-order systems. Inherently second-order systems are not extremely common in chemical engineering applications. Most second-order systems that we encounter will result from the addition of a controller to a first-order process. Let's examine an inherently second-order system and develop some terminology that will be useful in our analysis of the control of chemical processes.

Consider a simple manometer as shown in Fig. 7–1. The pressure on both legs of the manometer is initially the same. The length of the fluid column in the manometer is L. At time $t = 0$, a pressure difference is imposed across the legs of the manometer. Assuming the resulting flow in the manometer to be laminar and the steady-state friction law for drag force in laminar flow to apply at each instant, we will determine the transfer function between the applied pressure difference ΔP and the manometer reading h. If we perform a momentum balance on the fluid in the manometer, we arrive at the following terms:

(Sum of forces causing fluid to move) $=$ (Rate of change of momentum of fluid)   (7.2)

FIGURE 7–1
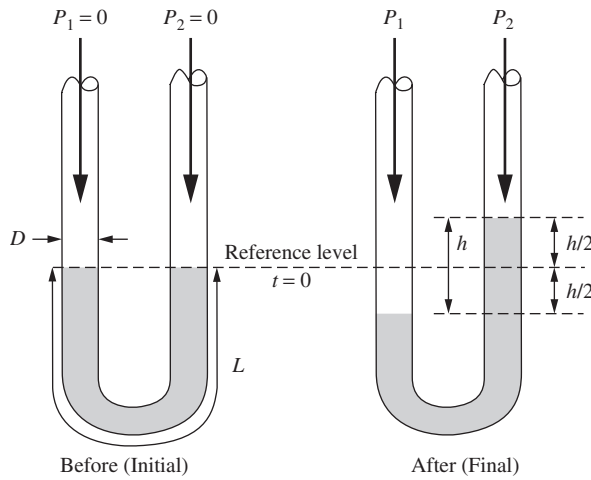Manometer.

where

$$\begin{pmatrix} \text{Sum of forces} \\ \text{causing fluid to move} \end{pmatrix} = \begin{pmatrix} \text{Unbalanced pressure forces} \\ \text{causing motion} \end{pmatrix} - \begin{pmatrix} \text{Frictional forces} \\ \text{opposing motion} \end{pmatrix}$$

$$\begin{pmatrix} \text{Unbalanced pressure forces} \\ \text{causing motion} \end{pmatrix} = (P_1 - P_2)\frac{\pi D^2}{4} - \rho g h \frac{\pi D^2}{4}$$

$$\begin{pmatrix} \text{Frictional forces} \\ \text{opposing motion} \end{pmatrix} = \begin{pmatrix} \text{Skin friction} \\ \text{at wall} \end{pmatrix} = \begin{pmatrix} \text{Shear stress} \\ \text{at wall} \end{pmatrix} \times \begin{pmatrix} \text{Area in contact} \\ \text{with wall} \end{pmatrix}$$

$$\begin{pmatrix} \text{Frictional forces} \\ \text{opposing motion} \end{pmatrix} = \tau_{\text{Wall}}(\pi D L) = \frac{8\mu\overline{V}}{D}(\pi D L) = \left(\frac{8\mu}{D}\right)\left(\frac{1}{2}\frac{dh}{dt}\right)(\pi D L)$$

The term for the skin friction at the wall is obtained from the Hagen-Poiseuille relationship for laminar flow (McCabe, Smith and Harriott, 2004). Note that $\overline{V}$ is the average velocity of the fluid in the tube, which is also the velocity of the interface, which is equal to $\frac{1}{2}\,dh/dt$ (see Fig. 7–2).
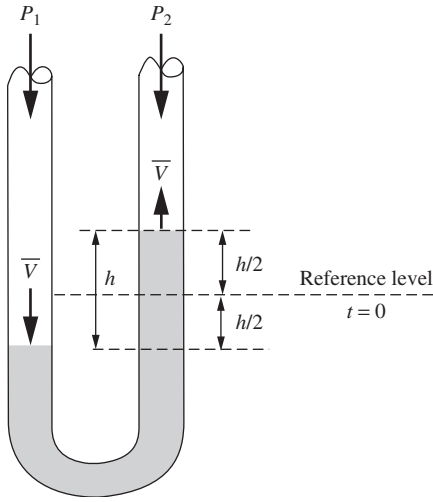
FIGURE 7–2
Average velocity of the fluid in the manometer.

The rate of change of momentum of the fluid [the right side of Eq. (7.2)] may be expressed as

$$\text{(Rate of change of momentum)} = \frac{d}{dt}(\text{mass} \times \text{velocity} \times \text{momentum correction factor})$$

$$= \left(\rho \frac{\pi D^2}{4} L\right)(\beta)\frac{d\overline{V}}{dt}$$

$$= \left(\rho \frac{\pi D^2}{4} L\right)(\beta)\left(\frac{1}{2}\frac{d^2h}{dt^2}\right)$$

The momentum correction factor β accounts for the fact that the fluid has a parabolic velocity profile in the tube, and the momentum must be expressed as $\beta m \overline{V}$ for laminar flow (see McCabe, Smith and Harriott, 2004). The value of β for laminar flow is 4/3. Substituting the appropriate terms into Eq. (7.2) produces the desired force balance equation for the manometer.

$$\left(\rho \frac{\pi D^2}{4} L\right)\left(\frac{4}{3}\right)\left(\frac{1}{2}\frac{d^2h}{dt^2}\right) = (P_1 - P_2)\frac{\pi D^2}{4} - \rho g h\frac{\pi D^2}{4} - \left(\frac{8\mu}{D}\right)\left(\frac{1}{2}\frac{dh}{dt}\right)(\pi D L \quad (7.3)$$

Rearranging Eq. (7.3), we obtain

$$\left(\rho \frac{\pi D^2}{4} L\right)\left(\frac{4}{3}\right)\left(\frac{1}{2}\frac{d^2h}{dt^2}\right) + \left(\frac{8\mu}{D}\right)\left(\frac{1}{2}\frac{dh}{dt}\right)(\pi D L) + \rho g h\frac{\pi D^2}{4} = (P_1 - P_2)\frac{\pi D^2}{4}$$

and finally, dividing both sides by $\rho g(\pi D^2/4)$, we arrive at the standard form for a second-order system.

$$\frac{2L}{3g}\frac{d^2h}{dt^2} + \frac{16\mu L}{\rho D^2 g}\frac{dh}{dt} + h = \frac{P_1 - P_2}{\rho g} = \frac{\Delta P}{\rho g} \tag{7.4}$$

(A more detailed version of the analysis of the manometer can be found in Bird et al., 1960). Note that as with first-order systems, standard form has a coefficient of 1 on the dependent variable term, h in this case. Second-order systems are described by a second-order differential equation. We may rewrite this Eq. (7.4) in general terms as

$$\tau^2 \frac{d^2Y}{dt^2} + 2\zeta\tau \frac{dY}{dt} + Y = X(t) \tag{7.5}$$

where

$$\tau^2 = \frac{2L}{3g} \tag{7.6}$$

$$2\zeta\tau = \frac{16\mu L}{\rho D^2 g} \tag{7.7}$$

$$X(t) = \frac{\Delta P}{\rho g} \quad \text{and} \quad Y = h \tag{7.8}$$

Solving for τ and ζ from Eqs. (7.6) and (7.7) gives

$$\tau = \sqrt{\frac{2L}{3g}} \quad \text{s} \tag{7.9}$$

$$\zeta = \frac{8\mu}{\rho D^2}\sqrt{\frac{3L}{2g}} \quad \text{dimensionless} \tag{7.10}$$

By definition, both τ and ζ must be positive. The reason for introducing τ and ζ in the particular form shown in Eq. (7.5) will become clear when we discuss the solution of Eq. (7.5) for particular forcing functions $X(t)$.

Equation (7.5) is written in a standard form that is widely used in control theory. If the fluid column is motionless ($dY/dt = 0$) and located at its rest position ($Y = 0$) before the forcing function is applied, the Laplace transform of Eq. (7.4) becomes

$$\tau^2 s^2 Y(s) + 2\zeta\tau s Y(s) + Y(s) = X(s) \tag{7.11}$$

From this, the transfer function follows:

$$\frac{Y(s)}{X(s)} = \frac{1}{\tau^2 s^2 + 2\zeta\tau s + 1} \tag{7.12}$$

The transfer function given by Eq. (7.12) is written in standard form, and we will show later that other physical systems can be represented by a transfer function having the

denominator of $\tau^2 s^2 + 2\zeta\tau s + 1$. All such systems are defined as second-order. Note that it requires two parameters, τ and ζ, to characterize the dynamics of a second-order system in contrast to only one parameter for a first-order system. We now discuss the response of a second-order system to some of the common forcing functions, namely, step, impulse, and sinusoidal.

Step Response

If the forcing function is a unit-step function, we have

$$X(s) = \frac{1}{s} \tag{7.13}$$

In terms of the manometer shown in Fig. 7–1, this is equivalent to suddenly applying a pressure difference [such that $X(t) = \Delta P/\rho g = 1$] across the legs of the manometer at time $t = 0$.

Superposition will enable us to determine easily the response to a step function of any other magnitude.

Combining Eq. (7.13) with the transfer function of Eq. (7.12) gives

$$Y(s) = \frac{1}{s} \frac{1}{\tau^2 s^2 + 2\zeta\tau s + 1} \tag{7.14}$$

The quadratic term in this equation may be factored into two linear terms that contain the roots

$$s_a = -\frac{\zeta}{\tau} + \frac{\sqrt{\zeta^2 - 1}}{\tau} \tag{7.15}$$

$$s_b = -\frac{\zeta}{\tau} + \frac{\sqrt{\zeta^2 - 1}}{\tau} \tag{7.16}$$

Equation (7.14) can now be written

$$Y(s) = \frac{1/\tau^2}{s(s - s_a)(s - s_b)} \tag{7.17}$$

The response of the system $Y(t)$ can be found by inverting Eq. (7.17). The roots s_a and s_b will be real or complex depending on value of the parameter ζ. The nature of the roots will, in turn, affect the form of $Y(t)$. The problem may be divided into the three cases shown in Table 7.1. Each case will now be discussed.

TABLE 7–1

Step response of a second-order system

Case	ζ	Nature of roots	Description of response
I	< 1	Complex	Underdamped or oscillatory
II	= 1	Real and equal	Critically damped
III	> 1	Real	Overdamped or nonoscillatory

CASE I STEP RESPONSE FOR $\zeta < 1$. For this case, the inversion of Eq. (7.17) yields the result

$$Y(t) = 1 - \frac{1}{\sqrt{1 - \zeta^2}} e^{-\zeta t / \tau} \sin\left(\sqrt{1 - \zeta^2}\, \frac{t}{\tau} + \tan^{-1} \frac{\sqrt{1 - \zeta^2}}{\zeta}\right) \tag{7.18}$$

To derive Eq. (7.18), use is made of the techniques of Chap. 3. Since $\zeta < 1$, Eqs. (7.15) to (7.17) indicate a pair of complex conjugate roots in the left half-plane and a root at the origin. In terms of the symbols of Fig. 3–1, the complex roots correspond to s_2 and s_2^* and the root at the origin to s_6.

The reader should realize that in Eq. (7.18), the argument of the sine function is in radians, as is the value of the inverse tangent term.

By referring to Table 3.1, we see that $Y(t)$ has the form

$$Y(t) = C_1 + e^{-\zeta t / \tau}\left(C_2 \cos \sqrt{1 - \zeta^2}\, \frac{t}{\tau} + C_3 \sin \sqrt{1 - \zeta^2}\, \frac{t}{\tau}\right) \tag{7.19}$$

The constants C_1, C_2, and C_3 are found by partial fractions. The resulting equation is then put in the form of Eq. (7.18) by applying the trigonometric identity used in Chap. 4, Eq. (4.26). The details are left as an exercise for the reader. It is evident from Eq. (7.18) that $Y(t) \to 1$ as $t \to \infty$.

The nature of the response can be understood most clearly by plotting the solution to Eq. (7.17) as shown in Fig. 7–3, where $Y(t)$ is plotted against the dimensionless variable t/τ for several values of ζ, including those above unity, which will be considered in the next section. Note that for $\zeta < 1$ all the response curves are oscillatory in nature and become less oscillatory as ζ is increased. The slope at the origin in Fig. 7–3 is zero for all values of ζ. The response of a second-order system for $\zeta < 1$ is said to be *underdamped*.

What is the physical significance of an underdamped response? Using the manometer as an example, if we step-change the pressure difference across an underdamped manometer, the liquid levels in the two legs will oscillate before stabilizing. The oscillations are characteristic of an underdamped response.

CASE II STEP RESPONSE FOR $\zeta = 1$. For this case, the response is given by the expression

$$Y(t) = 1 - \left(1 + \frac{t}{\tau}\right) e^{-t / \tau} \tag{7.20}$$

This is derived as follows: Equations (7.15) and (7.16) show that the roots s_1 and s_2 are real and equal. By referring to Fig. 3–1 and Table 3.1, it is seen that Eq. (7.20) is the correct form. The constants are obtained, as usual, by partial fractions.

The response, which is plotted in Fig. 7–3, is nonoscillatory. This condition, $\zeta = 1$, is called *critical damping* and allows the most rapid approach of the response to $Y = 1$ without oscillation.
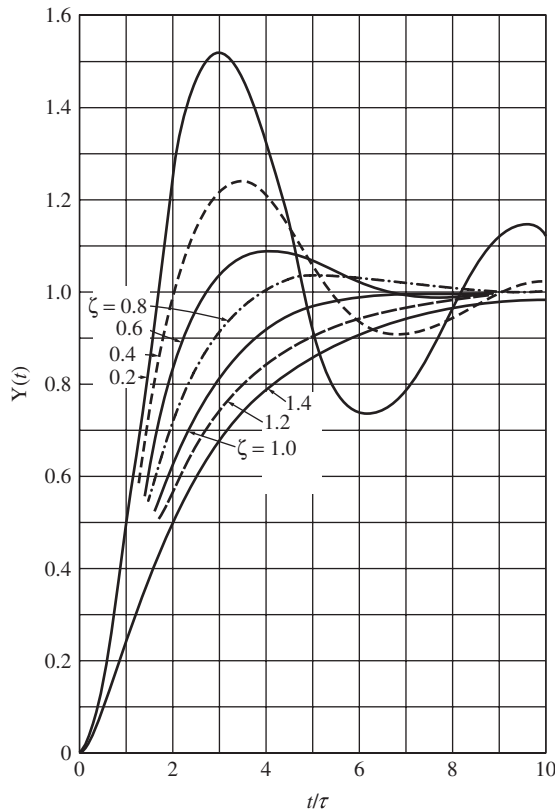
FIGURE 7–3
Response of a second-order system to a unit-step forcing function.

CASE III STEP RESPONSE FOR $\zeta > 1$. For this case, the inversion of Eq. (7.17) gives the result

$$Y(t) = 1 - e^{-\zeta t/\tau}\left(\cosh\sqrt{\zeta^2 - 1}\,\frac{t}{\tau} + \frac{\zeta}{\sqrt{\zeta^2 - 1}}\,\sinh\sqrt{\zeta^2 - 1}\,\frac{t}{\tau}\right) \quad (7.21)$$

where the hyperbolic functions are defined as

$$\sinh a = \frac{e^a - e^{-a}}{2}$$

$$\cosh a = \frac{e^a + e^{-a}}{2}$$

The procedure for obtaining Eq. (7.21) is parallel to that used in the previous cases.

The response has been plotted in Fig. 7–3 for several values of ζ. Notice that the response is nonoscillatory and becomes more "sluggish" as ζ increases. This is known as an *overdamped* response. As in previous cases, all curves eventually approach the line $Y = 1$.

Actually, the response for $\zeta > 1$ is not new. We saw it previously in the discussion of the step response of a system containing two first-order systems in series, for which the transfer function is

$$\frac{Y(s)}{X(s)} = \frac{1}{(\tau_1 s + 1)(\tau_2 s + 1)} \tag{7.22}$$

This is true for $\zeta > 1$ because the roots s_1 and s_2 are real, and the denominator of Eq. (7.12) may be factored into two real linear factors. Therefore, Eq. (7.12) is equivalent to Eq. (7.22) in this case. By comparing the linear factors of the denominator of Eq. (7.12) with those of Eq. (7.22), it follows that

$$\tau_1 = \left(\zeta + \sqrt{\zeta^2 - 1}\right)\tau \tag{7.23}$$

$$\tau_2 = \left(\zeta - \sqrt{\zeta^2 - 1}\right)\tau \tag{7.24}$$

Note that if $\tau_1 = \tau_2$, then $\tau = \tau_1 = \tau_2$ and $\zeta = 1$. The reader should verify these results.

Using MATLAB/Simulink to Determine the Step Response of the Manometer

Consider a manometer as illustrated in Fig. 7–1. The manometer is being used to determine the pressure difference between two instrument taps on an air line. The working fluid in the manometer is water. Determine the response of the manometer to a step change in pressure across the legs of the manometer.

Data
$L = 200$ cm
$g = 980$ cm/s^2
$\mu = 1$ cP $= 0.01$ g/(cm·s)
$\rho = 1.0$ g/cm^3 } for the working fluid, water

$$\frac{\Delta P}{\rho g} = \begin{cases} 0 & \text{for } t < 0 \\ 10 \text{ cm} & \text{for } t \geq 0 \end{cases}$$

$D = 0.11$ cm, 0.21 cm, 0.31 cm (Three Cases)

Solution. From Eq. (7.4), we have the governing differential equation for the manometer:

$$\frac{2L}{3g}\frac{d^2h}{dt^2} + \frac{16\mu L}{\rho D^2 g}\frac{dh}{dt} + h = \frac{P_1 - P_2}{\rho g} = \frac{\Delta P}{\rho g}$$

In terms of transformed deviation variables, this becomes

$$\tau^2 s^2 Y(s) + 2\zeta\tau s Y(s) + Y(s) = X(s)$$

where $\qquad Y = h - h_s \qquad$ and $\qquad X = \left(\dfrac{\Delta P}{\rho g} - \dfrac{\Delta P}{\rho g}\right)\Big|_s$

$$\tau = \sqrt{\dfrac{3L}{2g}} \qquad \text{and} \qquad \zeta = \dfrac{8\mu}{\rho D^2}\sqrt{\dfrac{3L}{2g}}$$

Let's calculate the time constant for the manometer.

$$\tau = \sqrt{\dfrac{2L}{3g}} = \sqrt{\dfrac{2(200 \text{ cm})}{3(980 \text{ cm}^2/\text{s})}} = 0.369 \text{ s}$$

and the damping coefficient for the three different tube diameters

$$\zeta = \dfrac{8\mu}{\rho D^2}\sqrt{\dfrac{3L}{2g}} = \dfrac{8[0.01 \text{ g}/(\text{cm} \cdot \text{s})]}{(1.0 \text{ g/cm}^3)(D^2)}\sqrt{\dfrac{3(200 \text{ cm})}{2(980 \text{ cm/s}^2)}} = \dfrac{0.0443}{D^2}$$

Diameter (cm)	ζ
0.11	3.66
0.21	1.00
0.31	0.46

Clearly we have one underdamped system ($\zeta < 1$), one critically damped system ($\zeta = 1$), and one overdamped system ($\zeta > 1$). One method of obtaining the responses is to substitute the values of τ and ζ into Eqs. (7.18), (7.20), and (7.21) and plot the resulting equations, realizing that the forcing function is 10 times a unit step. Another way to obtain the responses is to use MATLAB and Simulink to obtain the response of the transfer function Y/X to the forcing function input X.

$$\dfrac{Y(s)}{X(s)} = \dfrac{1}{\tau^2 s^2 + 2\zeta\tau s + 1} \qquad \text{and} \qquad X = \dfrac{10}{s}$$

The three necessary transfer functions are as follows:

Diameter (cm)	τ	ζ	τ^2	$2\zeta\tau$	Transfer function
0.11	0.369	3.66	0.136	2.70	$\dfrac{1}{0.136s^2 + 2.70s + 1}$
0.21	0.369	1.00	0.136	0.738	$\dfrac{1}{0.136s^2 + 0.738s + 1}$
0.31	0.369	0.46	0.136	0.340	$\dfrac{1}{0.136s^2 + 0.340s + 1}$

(continued)

The Simulink model for simulating the transfer functions is shown in Fig. 7–4, and the response is shown in Fig. 7–5.
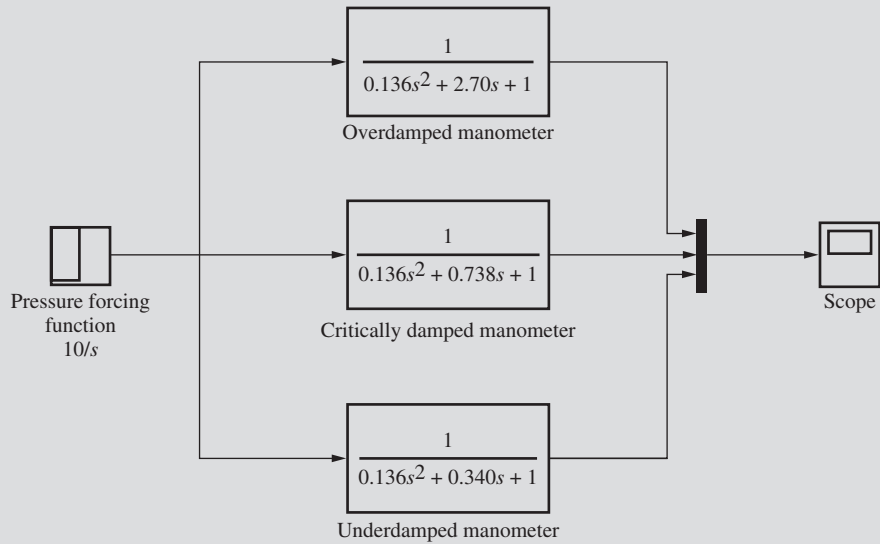

FIGURE 7–4
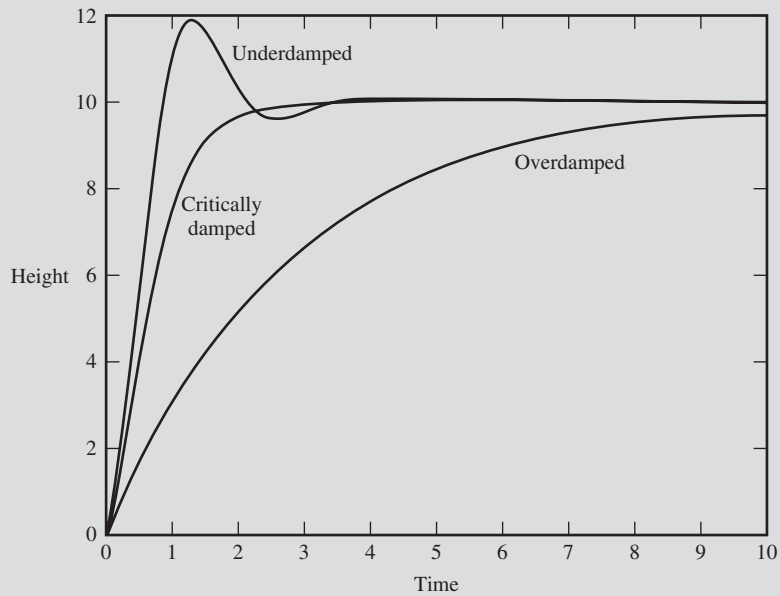Simulink diagram for manometer simulation.

FIGURE 7–5
Manometer response to step input.

Substituting the values for τ and ζ into Eqs. (7.18), (7.20), and (7.21), we get

$$Y(t) = 10\left[1 - \frac{1}{\sqrt{1 - \zeta^2}}e^{-\zeta t/\tau}\sin\left(\sqrt{1 - \zeta^2}\,\frac{t}{\tau} + \tan^{-1}\frac{\sqrt{1 - \zeta^2}}{\zeta}\right)\right] \quad \text{underdamped manometer}$$

$$= 10\left[1 - 1.13e^{-1.25t}\sin\left(2.41t + \underbrace{\tan^{-1} 1.93}_{1.09 \text{ rad}}\right)\right]$$

$$Y(t) = 10\left[1 - \left(1 + \frac{t}{\tau}\right)e^{-t/\tau}\right] = 10\left[1 - \left(1 + \frac{t}{0.369}\right)e^{-t/0.369}\right] \quad \text{critically damped manometer}$$

$$Y(t) = 10\left[1 - e^{-\zeta t/\tau}\left(\cosh\sqrt{\zeta^2 - 1}\,\frac{t}{\tau} + \frac{\zeta}{\sqrt{\zeta^2 - 1}}\sinh\sqrt{\zeta^2 - 1}\,\frac{t}{\tau}\right)\right] \quad \text{overdamped manometer}$$

$$= 10\left\{1 - e^{-9.92t}\left[\cosh(9.54t) + 1.04\sinh(9.54t)\right]\right\}$$

Plotting these responses gives the same results as the Simulink model.

On a practical note, notice that τ increases with the total length of the fluid column and that ζ increases with the viscosity of the fluid. If the damping coefficient ζ is small ($<< 1.0$), the response of the manometer to a change in pressure can be very oscillatory, and it becomes difficult to obtain accurate readings of the pressure. To dampen the oscillations, it is common practice to place a restriction on the bend of the tube. This increases the drag force of the fluid and is equivalent to increasing μ in the equation for ζ. Such a restriction (a partially open valve) is called a snubber.
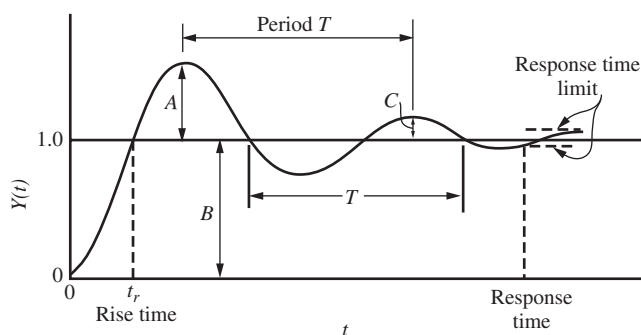
Terms Used to Describe an Underdamped System

Of these three cases, the underdamped response occurs most frequently in control systems. Hence a number of terms are used to describe the underdamped response quantitatively. Equations for some of these terms are listed below for future reference. In general, the terms depend on ζ and/or τ. All these equations can be derived from the time response as given by Eq. (7.18); however, the mathematical derivations are left to the reader as exercises.

1. *Overshoot.* Overshoot is a measure of how much the response exceeds the ultimate value following a step change and is expressed as the ratio A/B in Fig. 7–6.
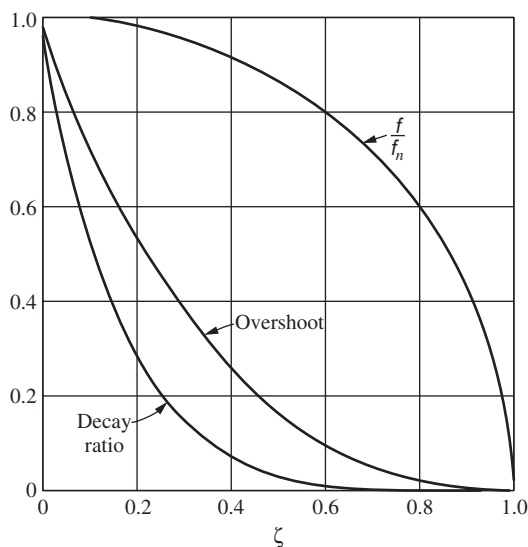   The overshoot for a unit step is related to ζ by the expression

$$\text{Overshoot} = \exp\left(\frac{-\pi\zeta}{\sqrt{1 - \zeta^2}}\right) \quad\quad (7.25)$$

This relation is plotted in Fig. 7–7. The overshoot increases for decreasing ζ.

FIGURE 7–6

Terms used to describe an underdamped second-order response.

FIGURE 7–7

Characteristics of a step response of an underdamped second-order system.

Why are we concerned about overshoot? Perhaps the temperature in our chemical reactor cannot be allowed to exceed a specified temperature to protect the catalyst from deactivation, or if it's a level control system, we don't want the tank to overflow. If we know these physical limitations, we can determine allowable values of ζ and choose our control system parameters to be sure to stay within those limits.

2. *Decay ratio.* The decay ratio is defined as the ratio of the sizes of successive peaks and is given by C/A in Fig. 7–6. The decay ratio is related to ζ by the expression

$$\text{Decay ratio} = \exp\left(\frac{-2\pi\zeta}{\sqrt{1-\zeta^2}}\right) = (\text{overshoot})^2 \qquad (7.26)$$

which is plotted in Fig. 7–7. Notice that larger ζ means greater damping, hence greater decay.

3. *Rise time.* This is the time required for the response to first reach its ultimate value and is labeled t_r in Fig. 7–6. The reader can verify from Fig. 7–3 that t_r increases with increasing ζ.
4. *Response time.* This is the time required for the response to come within ± 5 percent of its ultimate value and remain there. The response time is indicated in Fig. 7–6. The limits ± 5 percent are arbitrary, and other limits can be used for defining a response time.
5. *Period of oscillation.* From Eq. (7.18), the radian frequency (radians/time) is the coefficient of t in the sine term; thus,

$$\text{radian frequency } \omega = \frac{\sqrt{1 - \zeta^2}}{\tau} \tag{7.27}$$

Since the radian frequency ω is related to the cyclical frequency f by $\omega = 2\pi f$, it follows that

$$f = \frac{1}{T} = \frac{1}{2\pi} \frac{\sqrt{1 - \zeta^2}}{\tau} \tag{7.28}$$

where T is the period of oscillation (time/cycle). In terms of Fig. 7–6, T is the time elapsed between peaks. It is also the time elapsed between alternate crossings of the line $Y = 1$.

6. *Natural period of oscillation.* If the damping is eliminated [$B = 0$ in Eq. (7.1), or $\zeta = 0$], the system oscillates continuously without attenuation in amplitude. Under these "natural" or undamped conditions, the radian frequency is $1/\tau$, as shown by Eq. (7.27) when $\zeta = 0$. This frequency is referred to as the *natural frequency ω_n*:

$$\omega_n = \frac{1}{\tau} \tag{7.29}$$

The corresponding natural cyclical frequency f_n and period T_n are related by the expression

$$f_n = \frac{1}{T_n} = \frac{1}{2\pi\tau} \tag{7.30}$$

Thus, τ has the significance of the undamped period.
From Eqs. (7.28) and (7.30), the natural frequency is related to the actual frequency by the expression

$$\frac{f}{f_n} = \sqrt{1 - \zeta^2}$$

which is plotted in Fig. 7–7. Notice that for $\zeta < 0.5$ the natural frequency is nearly the same as the actual frequency.

In summary, it is evident that ζ is a measure of the degree of damping, or the oscillatory character, and τ is a measure of the period, or speed, of the response of a second-order system.

Impulse Response

If a unit impulse $\delta(t)$ is applied to the second-order system, then from Eqs. (7.12) and (3A.1) the transform of the response is

$$Y(s) = \frac{1}{\tau^2 s^2 + 2\zeta\tau s + 1} \tag{7.31}$$

As in the case of the step input, the nature of the response to a unit impulse will depend on whether the roots of the denominator of Eq. (7.31) are real or complex. The problem is again divided into the three cases shown in Table 7.1, and each is discussed below.

CASE I IMPULSE RESPONSE FOR $\zeta < 1$.   The inversion of Eq. (7.31) for $\zeta < 1$ yields the result

$$Y(t) = \frac{1}{\tau} \frac{1}{\sqrt{1 - \zeta^2}} e^{-\zeta t/\tau} \sin\sqrt{1 - \zeta^2}\, \frac{t}{\tau} \tag{7.32}$$

which is plotted in Fig. 7–8. The slope at the origin in Fig. 7–8 is 1.0 for all values of ζ.

A simple way to obtain Eq. (7.32) from the step response of Eq. (7.18) is to take the derivative of Eq. (7.18) with respect to t (remember from App. 3A that the derivative of the unit-step function is the impulse function). Comparison of Eqs. (7.14) and (7.31) shows that

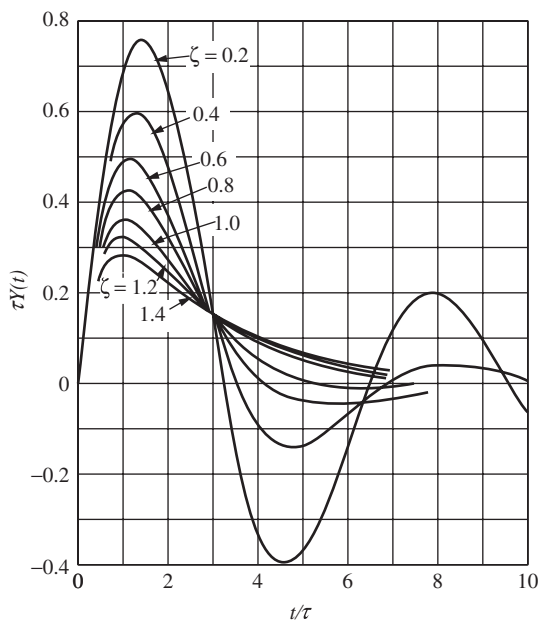$$Y(s)\big|_{\text{impulse}} = sY(s)\big|_{\text{step}} \tag{7.33}$$

FIGURE 7–8
Response of a second-order system to a unit-impulse forcing function.

The presence of s on the right side of Eq. (7.33) implies differentiation with respect to t in the time response. In other words, the inverse transform of Eq. (7.31) is

$$Y(t)|_{\text{impulse}} = \frac{d}{dt}\left(Y(t)|_{\text{step}}\right) \tag{7.34}$$

Application of Eq. (7.34) to Eq. (7.18) yields Eq. (7.32). This principle also yields the results for the next two cases.

CASE II IMPULSE RESPONSE FOR $\zeta = 1$. For the critically damped case, the response is given by

$$Y(t) = \frac{1}{\tau^2} t e^{-t/\tau} \tag{7.35}$$

which is plotted in Fig. 7–8.

CASE III IMPULSE RESPONSE FOR $\zeta > 1$. For the overdamped case, the response is given by

$$Y(t) = \frac{1}{\tau}\frac{1}{\sqrt{\zeta^2 - 1}}\, e^{-\zeta t/\tau}\, \sinh\sqrt{\zeta^2 - 1}\,\frac{t}{\tau} \tag{7.36}$$

which is also plotted in Fig. 7–8.

To summarize, the impulse–response curves of Fig. 7–8 show the same general behavior as the step response curves of Fig. 7–3. However, the impulse response always returns to zero. Terms such as *decay ratio, period of oscillation,* etc., may also be used to describe the impulse response. Many control systems exhibit transient responses such as those of Fig. 7–8.

Sinusoidal Response

If the forcing function applied to the second-order system is sinusoidal

$$X(t) = A \sin \omega t$$

then it follows from Eqs. (7.12) and (4.23) that

$$Y(s) = \frac{A\omega}{\left(s^2 + \omega^2\right)\left(\tau^2 s^2 + 2\zeta\tau s + 1\right)} \tag{7.37}$$

The inversion of Eq. (7.37) may be accomplished by first factoring the two quadratic terms to give

$$Y(s) = \frac{A\omega/\tau^2}{(s - j\omega)(s + j\omega)(s - s_a)(s - s_b)} \tag{7.38}$$

Here s_a and s_b are the roots of the denominator of the transfer function and are given by Eqs. (7.15) and (7.16). For the case of an underdamped system ($\zeta < 1$), the roots of the denominator of Eq. (7.38) are a pair of pure imaginary roots ($Ij\omega, -j\omega$) contributed by the forcing function and a pair of complex roots $\left(-\zeta/\tau + j\sqrt{1 - \zeta^2}/\tau, \ -\zeta/\tau - j\sqrt{1 - \zeta^2}/\tau\right.$

We may write the form of the response $Y(t)$ by referring to Fig. 3–1 and Table 3.1; thus

$$Y(t) = C_1 \cos \omega t + C_2 \sin \omega t + e^{-\zeta t/\tau}\left(C_3 \cos \sqrt{1 - \zeta^2}\,\frac{t}{\tau} + C_4 \sin \sqrt{1 - \zeta^2}\,\frac{t}{\tau}\right)$$

$$(7.39)$$

The constants are evaluated by partial fractions. Notice in Eq. (7.39) that as $t \to \infty$, only the first two terms do not become zero. These remaining terms are the ultimate periodic solution; thus

$$Y(t)\big|_{t \to \infty} = C_1 \cos \omega t + C_2 \sin \omega t \qquad (7.40)$$

The reader should verify that Eq. (7.40) is also true for $\zeta \geq 1$. From this little effort, we see already that the response of the second-order system to a sinusoidal driving function is ultimately sinusoidal and has the same frequency as the driving function. If the constants C_1 and C_2 are evaluated, we get from Eqs. (4.26) and (7.40)

$$Y(t) = \frac{A}{\sqrt{\left[1 - (\omega t)^2\right]^2 + (2\zeta\omega\tau)^2}}\sin(\omega t + \phi) \qquad (7.41)$$

where

$$\phi = -\tan^{-1}\frac{2\zeta\omega\tau}{1 - (\omega\tau)^2}$$

By comparing Eq. (7.41) with the forcing function

$$X(t) = A \sin \omega t$$

it is seen that

1. The ratio of the output amplitude to the input amplitude is

$$\text{Amplitude ratio} = \frac{\text{output amplitude}}{\text{input amplitude}} = \frac{1}{\sqrt{\left[1 - (\omega\tau)^2\right]^2 + (2\zeta\omega\tau)^2}}$$

It will be shown in Chap. 15 that this may be greater or less than 1, depending upon the values of ζ and $\omega\tau$. This is in direct contrast to the sinusoidal response of the *first-order* system, where the ratio of the output amplitude to the input amplitude is always *less than* 1.

2. The output lags the input by phase angle $|\phi|$.

$$\phi = \text{phase angle} = -\tan^{-1}\frac{2\zeta\omega\tau}{1 - (\omega\tau)^2}$$

It can be seen from Eq. (7.41), and will be shown in Chap. 15, that $|\phi|$ approaches 180° asymptotically as ω increases. The phase lag of the first-order system, on the other hand, can never exceed 90°. Discussion of other characteristics of the sinusoidal response will be deferred until Chap. 15.

We now have at our disposal considerable information about the dynamic behavior of the second-order system. It happens that many control systems that are not truly second-order exhibit step responses very similar to those of Fig. 7–3. Such systems are often characterized by second-order equations for approximate mathematical analysis. Hence, the second-order system is quite important in control theory, and frequent use will be made of the material in this chapter.

7.2   TRANSPORTATION LAG

A phenomenon that is often present in flow systems is the *transportation lag.* Synonyms for this term are *dead time* and *distance velocity lag.* As an example, consider the system shown in Fig. 7–9, in which a liquid flows through an insulated tube of uniform cross-sectional area A and length L at a constant volumetric flow rate q. The density ρ and the heat capacity C are constant. The tube wall has negligible heat capacity, and the velocity profile is flat (plug flow).
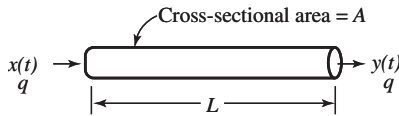
The temperature x of the entering fluid varies with time, and it is desired to find the response of the outlet temperature $y(t)$ in terms of a transfer function.

FIGURE 7–9
System with transportation lag.

As usual, it is assumed that the system is initially at steady state; for this system, it is obvious that the inlet temperature equals the outlet temperature; i.e.,

$$x_s = y_s \tag{7.42}$$

If a step change were made in $x(t)$ at $t = 0$, the change would not be detected at the end of the tube until τ s later, where τ is the time required for the entering fluid to pass through the tube. This simple step response is shown in Fig. 7–10.

If the variation in $x(t)$ were some arbitrary function, as shown in Fig. 7–10, the response $y(t)$ at the end of the pipe would be identical with $x(t)$ but again delayed by τ
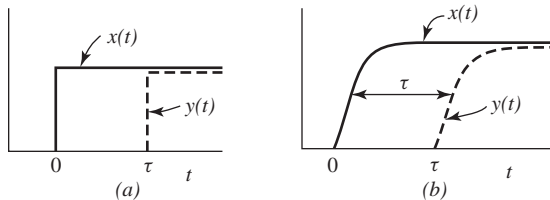

FIGURE 7–10
Response of transportation lag to various inputs.

units of time. The transportation lag parameter τ is simply the time needed for a particle of fluid to flow from the entrance of the tube to the exit, and it can be calculated from the expression

$$\tau = \frac{\text{volume of tube}}{\text{volumetric flow rate}}$$

or
$$\tau = \frac{AL}{q} \qquad (7.43)$$

It can be seen from Fig. 7–10 that the relationship between $y(t)$ and $x(t)$ is

$$y(t) = x(t - \tau) \qquad (7.44)$$

Subtracting Eq. (7.42) from Eq. (7.44) and introducing the deviation variables $X = x - x_s$ and $Y = y - y_s$ give

$$Y(t) = X(t - \tau) \qquad (7.45)$$

If the Laplace transform of $X(t)$ is $X(s)$, then the Laplace transform of $X(t - \tau)$ is $e^{-s\tau}X(s)$. This result follows from the theorem on translation of a function, which was discussed in App. 3A. Equation (7.45) becomes

$$Y(s) = e^{-s\tau}X(s)$$

or
$$\frac{Y(s)}{X(s)} = e^{-s\tau} \qquad (7.46)$$

Therefore, the transfer function of a transportation lag is $e^{-s\tau}$.

The transportation lag is quite common in the chemical process industries where a fluid is transported through a pipe. We shall see in a later chapter that the presence of a transportation lag in a control system can make it much more difficult to control. In general, such lags should be avoided if possible by placing equipment close together. They can seldom be entirely eliminated.

APPROXIMATION OF TRANSPORT LAG. The transport lag is quite different from the other transfer functions (first-order, second-order, etc.) that we have discussed in that it is not a rational function (i.e., a ratio of polynomials.) As shown in Chap. 13, a system containing a transport lag cannot be analyzed for stability by the Routh test. The transport lag can also be difficult to simulate by computer. For these reasons, several approximations of transport lag that are useful in control calculations are presented here.

One approach to approximating the transport lag is to write $e^{-\tau s}$ as $1/e^{\tau s}$ and to express the denominator as a Taylor series; the result is

$$e^{-\tau s} = \frac{1}{e^{\tau s}} = \frac{1}{1 + \tau s + \tau^2 s^2 / 2 + \tau^3 s^3 / 3! + \cdots}$$

Keeping only the first two terms in the denominator gives

$$e^{-\tau s} \cong \frac{1}{1 + \tau s} \qquad (7.47)$$

This approximation, which is simply a first-order lag, is a crude approximation of a transport lag. An improvement can be made by expressing the transport lag as

$$e^{-\tau s} = \frac{e^{-\tau s/2}}{e^{\tau s/2}}$$

Expanding numerator and denominator in a Taylor series and keeping only terms of first-order give

$$e^{-\tau s} \cong \frac{1 - \tau s /2}{1 + \tau s /2} \qquad \text{first-order Padé} \qquad (7.48)$$

This expression is also known as a *first-order Padé* approximation.

Another well-known approximation for a transport lag is the second-order Padé approximation:

$$e^{-\tau s} \cong \frac{1 - \tau s/2 + \tau^2 s^2/12}{1 + \tau s/2 + \tau^2 s^2/12} \qquad \text{second-order Padé} \qquad (7.49)$$

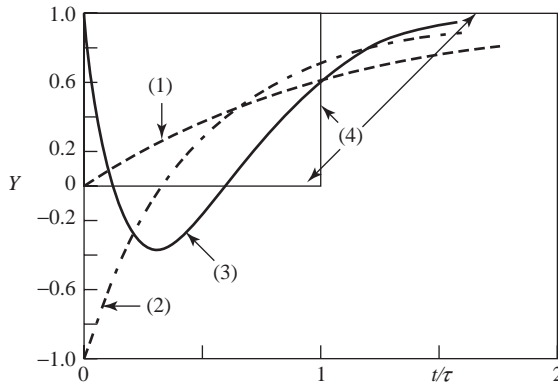Equation (7.48) is not merely the ratio of two Taylor series; it has been optimized to give a better approximation.

FIGURE 7–11

Step response to approximation of the transport lag $e^{-\tau s}$:

(1) $\dfrac{1}{\tau s + 1}$; (2) first-order Padé; (3) second-order Padé ; (4) $e^{-\tau s}$.

The step responses of the three approximations of transport lag presented here are shown in Fig. 7–11. The step response of $e^{-\tau s}$ is also shown for comparison. Notice that the response for the first-order Padé approximation drops to -1 before rising exponentially toward $+1$. The response for the second-order Padé approximation jumps to $+1$ and then descends to below 0 before returning gradually back to $+1$.

Although none of the approximations for $e^{-\tau s}$ is very accurate, the approximation for $e^{-\tau s}$ is more useful when it is multiplied by several first-order or second-order transfer functions. In this case, the other transfer functions filter out the high-frequency content of the signals passing through the transport lag, with the result that the transport lag approximation, when combined with other transfer functions, provides a satisfactory result in many cases. The accuracy of a transport lag can be evaluated most clearly in terms of frequency response, a topic covered later in this book.

SUMMARY

After studying the material in this chapter, we now have at our disposal considerable information about the dynamic behavior of the second-order systems and transportation lags. We noted that even though many control systems are not truly second-order, they frequently exhibit step responses very similar to what we have observed in this chapter. Such systems are often characterized by second-order equations for an approximate mathematical analysis. Therefore, the second-order system is quite important in control theory, and we will make use of this material often in future chapters.

PROBLEMS

7.1. A step change of magnitude 4 is introduced into a system having the transfer function

$$\frac{Y(s)}{X(s)} = \frac{10}{s^2 + 1.6s + 4}$$

Determine

(*a*) Percent overshoot

(*b*) Rise time

(*c*) Maximum value of $Y(t)$

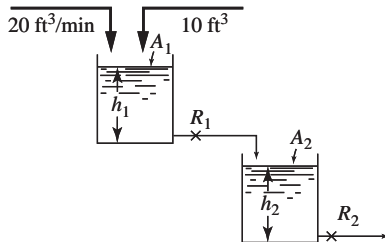(*d*) Ultimate value of $Y(t)$

(*e*) Period of oscillation

FIGURE P7–2

7.2. The two-tank system shown in Fig. P7–2 is operating at steady state. At time $t = 0$, $10\ \text{ft}^3$ of water is quickly added to the first tank. Using appropriate figures and equations in the text, determine the maximum deviation in level (feet) in both tanks from the ultimate steady-state values and the time at which each maximum occurs. Data:

$$A_1 = A_2 = 10\ \text{ft}^2$$
$$R_1 = 0.1\ \text{ft/cfm}$$
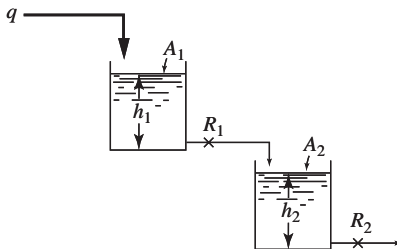$$R_2 = 0.35\ \text{ft/cfm}$$

FIGURE P7–3

7.3. The two-tank liquid-level system shown in Fig. P7–3 is operating at steady state when a step change is made in the flow rate to tank 1. The transient response is critically damped, and it takes 1.0 min for the change in level of the second tank to reach 50 percent of the total change.

If the ratio of the cross-sectional areas of the tanks is $A_1/A_2 = 2$, calculate the ratio R_1/R_2. Calculate the time constant for each tank. How long does it take for the change in level of the first tank to reach 90 percent of the total change?

7.4. Use Simulink to solve Prob. 7.3.

7.5. Design a mercury manometer that will measure pressures up to 2 atm absolute and will give responses that are slightly underdamped (that is, $\zeta = 0.7$).

7.6. Verify Eqs. (7.18), (7.20), and (7.21).

7.7. Verify Eqs. (7.25) and (7.26).

7.8. Verify Eq. (7.41).

7.9. If a second-order system is overdamped, it is more difficult to determine the parameters ζ and τ experimentally. One method for determining the parameters from a step response has been suggested by R. C. Oldenbourg and H. Sartorius (*The Dynamics of Automatic Controls.* ASME, p. 78, 1948), as described below.

(*a*) Show that the unit-step response for the overdamped case may be written in the form

$$S(t) = 1 - \frac{r_1 e^{r_2 t} - r_2 e^{r_1 t}}{r_1 - r_2}$$

where r_1 and r_2 are the (real and negative) roots of

$$\tau^2 s^2 + 2\zeta\tau s + 1 = 0$$

(*b*) Show that $S(t)$ has an inflection point at

$$t_i = \frac{\ln(r_2 / r_1)}{r_1 / r_2}$$

(*c*) Show that the slope of the step response at the inflection point

$$\left.\frac{dS(t)}{dt}\right|_{t=t_i} = S'(t_i)$$

has the value

$$S'(t_i) = -r_1 e^{r_1 t_i} = -r_2 e^{r_2 t_i}$$

$$= -r_1\left(\frac{r_2}{r_1}\right)^{r_1/(r_1-r_2)}$$

(*d*) Show that the value of the step response at the inflection point is

$$S(t_i) = 1 + \frac{r_1 + r_2}{r_1 r_2}S'(t_i)$$

and that hence

$$\frac{1 - S(t_i)}{S'(t_i)} = \frac{1}{r_1} - \frac{1}{r_2}$$

(*e*) On a typical sketch of a unit-step response, show distances equal to

$$\frac{1}{S'(t_i)} \quad \text{and} \quad \frac{1 - S(t_i)}{S'(t_i)}$$

and hence present two simultaneous equations resulting from a graphical method for determination of r_1 and r_2.

(*f*) Relate ζ and τ to r_1 and r_2.

7.10. Determine $Y(0)$, $Y(0.6)$, and $Y(\infty)$ if

$$Y(s) = \frac{1}{s}\frac{25(s+1)}{s^2+2s+25}$$

7.11. In the liquid-level system shown in Fig. P7–11, the deviation in flow rate to the first tank is an impulse function of magnitude 5. The following data apply: $A_1 = 1$ ft^2, $A_2 = A_3 = 2$ ft^2, $R_1 = 1$ ft/cfm, and $R_2 = 1.5$ ft/cfm.
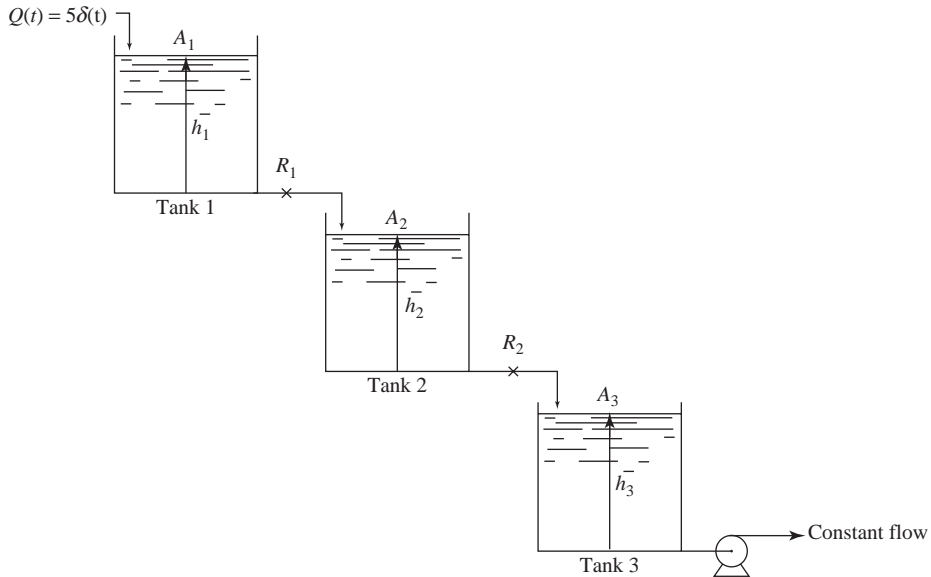

FIGURE P7–11

    (a) Determine expressions for $H_1(s)$, $H_2(s)$, and $H_3(s)$ where H_1, H_2, and H_3 are deviations in tank level for tanks 1, 2, and 3.

    (b) Sketch the responses of H_1 (t), H_2 (t), and H_3 (t). (You need show only the shape of the responses; do not plot.)

    (c) Determine H_1 (3.46), H_2 (3.46), and H_3 (3.46). For H_2 and H_3, use graphs in Chap. 7 of this text after first finding values of τ and ζ for an equivalent second-order system.

7.12. Sketch the response $Y(t)$ if $Y(s) = e^{-2s}/[s^2 + 1.2s + 1]$. Determine $Y(t)$ for $t = 0, 1, 5$, and ∞.

7.13. The two tanks shown in Fig. P7–13 are connected in an interacting fashion. The system is initially at steady state with $q = 10$ cfm. The following data apply to the tanks: $A_1 = 1$ ft^2, $A_2 = 1.25$ ft^2, $R_1 = 1$ ft/cfm, and $R_2 = 0.8$ ft/cfm.
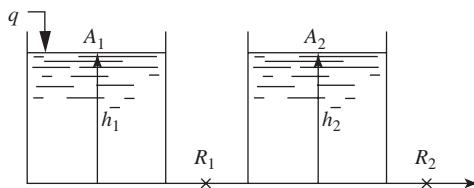

FIGURE P7–13

(a) If the flow changes from 10 to 11 cfm according to a step change, determine $H_2(s)$, i.e., the Laplace transform of $H_2(t)$, where H_2 is the deviation in h_2.

(b) Determine $H_2(1)$, $H_2(4)$, and $H_2(\infty)$.

(c) Determine the initial levels (actual levels) $h_1(0)$ and $h_2(0)$ in the tanks.

(d) Obtain an expression for $H_1(s)$ for the unit-step change described above.

7.14. From figures in this chapter, determine $Y(4)$ for the system response expressed by

$$y(s) = \frac{2}{s} \frac{2s + 4}{4s^2 + 0.8s + 1}$$

7.15. A step change of magnitude 3 is introduced into the transfer function

$$\frac{Y(s)}{X(s)} = \frac{10}{2s^2 + 0.3s + 0.5}$$

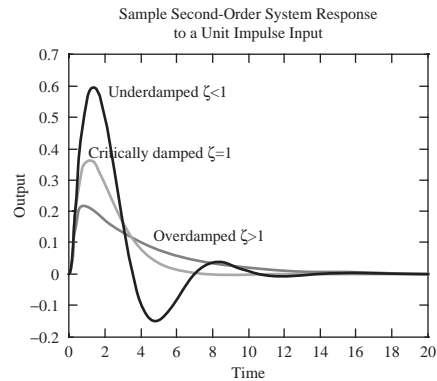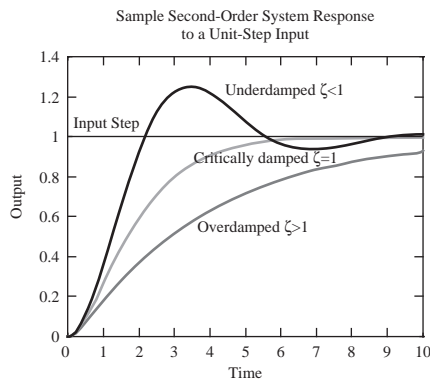Determine the overshoot and the frequency of oscillation.

CHAPTER
7

CAPSULE SUMMARY

STANDARD FORM FOR A SECOND-ORDER SYSTEM:

$$\tau^2 \frac{d^2Y}{dt^2} + 2\zeta\tau \frac{dY}{dt} + Y = X(t)$$

τ = Time constant

ζ = Damping coefficient (magnitude of this parameter determines the nature of the response)
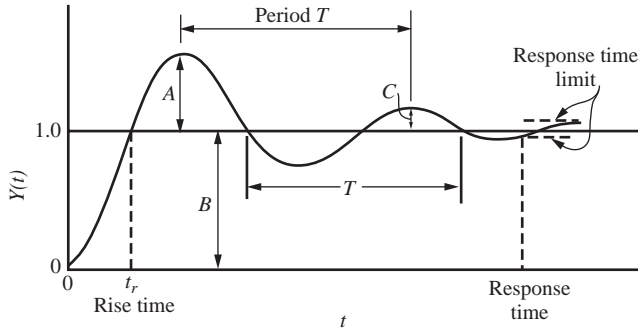

TERMS TO DESCRIBE AN UNDERDAMPED (OSCILLATORY) SECOND-ORDER RESPONSE
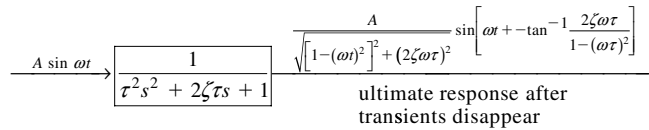
$$\text{Overshoot} = \exp\left(\frac{-\pi\zeta}{\sqrt{1-\zeta^2}}\right) = \frac{A}{B}$$

$$\text{Decay ratio} = \exp\left(\frac{-2\pi\zeta}{\sqrt{1-\zeta^2}}\right) = (\text{overshoot})^2 = \frac{C}{A}$$

$$\text{Radian frequency } \omega = \frac{\sqrt{1-\zeta^2}}{\tau} \qquad \omega_n = \text{natural frequency} = \frac{1}{\tau}$$

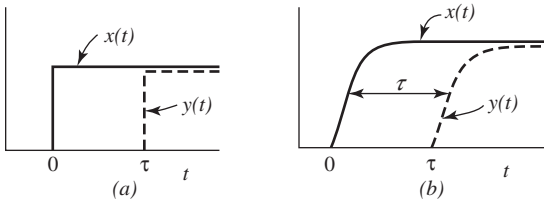## SINUSOIDAL RESPONSE OF A SECOND-ORDER SYSTEM

$$\text{Amplitude ratio} = \frac{\text{output amplitude}}{\text{input amplitude}} = \frac{1}{\sqrt{\left[1 - (\omega\tau)^2\right]^2 + (2\zeta\omega\tau)^2}} \qquad \phi = \text{phase angle} = -\tan^{-1}\frac{2\zeta\omega\tau}{1 - (\omega\tau)^2}$$

TRANSPORTATION LAG

$$Y(s) = e^{-s\tau}X(s)$$

LINEAR CLOSED-LOOP SYSTEMS

THE CONTROL SYSTEM

8.1 INTRODUCTION

In previous chapters, the dynamic behavior of several basic systems was examined. With this background, we can extend the discussion to a complete control system and introduce the fundamental concept of feedback. To work with a familiar system, the treatment will be based on a stirred-tank heater.

Figure 8–1 is a sketch of the apparatus. To orient the reader, the physical description of the process will be reviewed. A liquid stream at a temperature T_i enters an insulated, well-stirred tank at a constant flow rate w (mass/time). It is desired to maintain (or control) the temperature in the tank at T_R by means of the controller. If the measured tank temperature T_m differs from the desired temperature T_R, the controller senses the difference or *error* $\varepsilon = T_R - T_m$ and changes the heat input in such a way as to reduce the magnitude of ε. If the controller changes the heat input to the tank by an amount that is proportional to ε, we have *proportional* control.

In Fig. 8–1, it is indicated that the source of heat input q may be electricity or steam. If an electrical source were used, the final control element might be a variable transformer that is used to adjust current to a resistance heating element; if steam were used, the final control element would be a control valve that adjusts the flow of steam. In either case, the output signal from the controller should adjust q in such a way as to maintain control of the temperature in the tank.

8.2 COMPONENTS OF A CONTROL SYSTEM

The system shown in Fig. 8–1 may be divided into the following components:

1. Process (stirred-tank heater).
2. Measuring element (thermometer).
3. Controller.
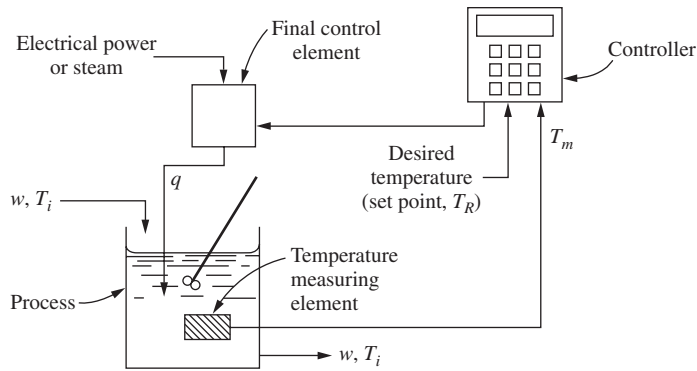4. Final control element (variable transformer or control valve).

FIGURE 8–1
Control system for a stirred-tank heater.

Each of these components can be readily identified as a separate physical item in the process. In general, these four components will constitute most of the control systems that we consider in this text; however, the reader should realize that more complex control systems exist in which more components are used. For example, some processes require a cascade control system in which two controllers and two measuring elements are used. A cascade system is discussed in Chap. 17.

8.3 BLOCK DIAGRAM

For computational purposes, it is convenient to represent the control system of Fig. 8–1 by means of the block diagram shown in Fig. 8–2. Such a diagram makes it much easier to visualize the relationships among the various signals. New terms, which appear in Fig. 8–2, are *set point* and *load*. The set point is a synonym for the desired value of the controlled variable. The load refers to a change in any variable that may cause the controlled variable of the process to change. In this example, the inlet temperature T_i is a load variable. Other possible loads for this system are changes in flow rate and heat loss from the tank. (These loads are not shown on the diagram.)
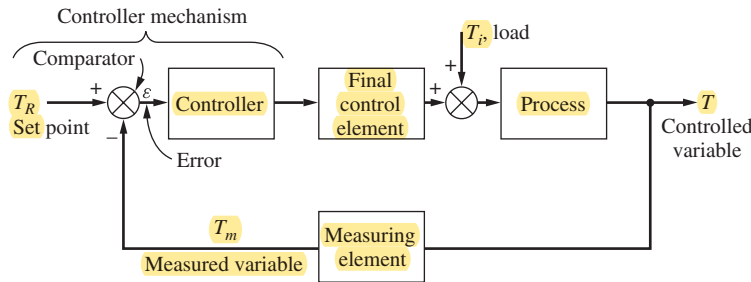

FIGURE 8–2
Block diagram of a simple control system.

The control system shown in Fig. 8–2 is called a *closed-loop* system or a feedback system because the measured value of the controlled variable is returned or "fed back" to a device called the *comparator*. In the comparator, the controlled variable is compared with the desired value or *set point*. If there is any difference between the measured variable and the set point, an error is generated. This error enters a *controller,* which in turn adjusts the *final control element* to return the controlled variable to the set point.

Negative Feedback Versus Positive Feedback

Several terms have been used that may need further clarification. The feedback principle, which is illustrated by Fig. 8–2, involves the use of the controlled variable T to maintain itself at a desired value T_R. The arrangement of the apparatus of Fig. 8–2 is often described as *negative feedback* to contrast with another arrangement called positive feedback. Negative feedback ensures that the difference between T_R and T_m is used to adjust the control element so that the tendency is to reduce the error. For example, assume that the system is at steady state and that $T = T_m = T_R$. If the load T_i should increase, T and T_m would start to increase, which would cause the error ε to become negative. With proportional control, the decrease in error would cause the controller and final control element to *decrease* the flow of heat to the system, with the result that the flow of heat would eventually be reduced to a value such that T approaches T_R. A verbal description of the operation of a feedback control system, such as the one just given, is admittedly inadequate, for this description necessarily is given as a sequence of events. Actually all the components operate simultaneously, and the only adequate description of what is occurring is a set of simultaneous differential equations. This more accurate description is the primary subject matter of the present and succeeding chapters.

If the signal to the comparator were obtained by adding T_R and T_m, we would have a *positive feedback* system, which is inherently unstable. To see that this is true, again assume that the system is at steady state and that $T = T_m = T_R$. If T_i were to increase, T and T_m would increase, which would cause the signal from the comparator (ε in Fig. 8–2) to increase, with the result that the heat to the system would increase. However, this action, which is just the opposite of that needed, would cause T to increase further. It should be clear that this situation would cause T to "run away" and control would not be achieved. For this reason, positive feedback would never be used intentionally in the system of Fig. 8–2. However, in more complex systems it may arise naturally. An example of this is discussed in Chap. 20.

Servo Problem Versus Regulator Problem

The control system of Fig. 8–2 can be considered from the point of view of its ability to handle either of two types of situations. In the first situation, which is called the servomechanism-type (or servo) problem, we assume that there is no change in load T_i and that we are interested in changing the bath temperature according to some prescribed function of time. For this problem, the set point T_R would be changed in accordance with the desired variation in bath temperature. If the variation is sufficiently

slow, the bath temperature may be expected to follow the variation in T_R very closely. There are occasions when a control system in the chemical industry will be operated in this manner. For example, one may be interested in varying the temperature of a reactor according to a prescribed time-temperature pattern. However, the majority of problems that may be described as the servo type come from fields other than the chemical industry. The tracking of missiles and aircraft and the automatic machining of intricate parts from a master pattern are well-known examples of the servo-type problem. The servo problem can be viewed as trying to follow a moving target (i.e., the changing set point). The other situation will be referred to as the regulator problem. In this case, the desired value T_R is to remain fixed, and the purpose of the control system is to maintain the controlled variable at T_R in spite of changes in load T_i. This problem is very common in the chemical industry, and a complicated industrial process will often have many self-contained control systems, each of which maintains a particular process variable at a desired value. These control systems are of the regulator type.

In considering control systems in the following chapters, we will frequently discuss the response of a linear control system to a change in set point (servo problem) separately from the response to a change in load (regulator problem). However, it should be realized that this is done only for convenience. The basic approach to obtaining the response of either type is essentially the same, and the two responses may be superimposed to obtain the response to any linear combination of set point and load changes.

8.4   DEVELOPMENT OF BLOCK DIAGRAM

Each block in Fig. 8–2 represents the functional relationship existing between the input and output of a particular component. In previous chapters, such input-output relations were developed in the form of transfer functions. In block diagram representations of control systems, the variables selected are *deviation variables,* and inside each block is placed the transfer function relating the input-output pair of variables. Finally, the blocks are combined to give the overall block diagram. This is the procedure to be followed in developing Fig. 8–2.

Process

Consider first the block for the process. This block will be seen to differ somewhat from those presented in previous chapters in that two input variables are present; however, the procedure for developing the transfer function remains the same.

An unsteady-state energy balance around the tank gives

$$q + wC(T_i - T_o) - wC(T - T_o) = \rho CV \frac{dT}{dt} \tag{8.1}$$

where T_o is the reference temperature.

[Note: In this analysis, it is assumed that the flow rate of heat q is instantaneously available and independent of the temperature in the tank. In some stirred-tank heaters, such as a jacketed kettle, q depends on both the temperature of the fluid in the jacket and the temperature of the fluid in the kettle. In this introductory chapter, systems (electrically heated tank or direct steam-heated tank) are selected for which this complication can be ignored. In Chap. 20, the analysis of a steam-jacketed kettle is given in which the effect of kettle temperature on q is taken into account.]

At steady state, dT/dt is zero, and Eq. (8.1) can be written

$$q_s + wC(T_{i_s} - T_o) - wC(T_s - T_o) = 0 \tag{8.2}$$

where the subscript s has been used to indicate steady state.

Subtracting Eq. (8.2) from Eq. (8.1) gives

$$q - q_s + wC[(T_i - T_{i_s}) - (T - T_s)] = \rho CV \frac{d(T - T_s)}{dt} \tag{8.3}$$

Notice that the reference temperature T_o cancels in the subtraction. If we introduce the deviation variables

$$T_i' = T_i - T_{i_s} \tag{8.4}$$

$$Q = q - q_s \tag{8.5}$$

$$T' = T - T_s \tag{8.6}$$

Eq. (8.3) becomes

$$Q + wC(T_i' - T') = \rho CV \frac{dT'}{dt} \tag{8.7}$$

Taking the Laplace transform of Eq. (8.7) gives

$$Q(s) + wC[T_i'(s) - T'(s)] = \rho CVsT'(s) \tag{8.8}$$

or

$$T'(s)\left(\frac{\rho V}{w}s + 1\right) = \frac{Q(s)}{wC} + T_i'(s) \tag{8.9}$$

This last expression can be written

Stirred heater transfer function $$T'(s) = \frac{1/wC}{\tau s + 1}Q(s) + \frac{1}{\tau s + 1}T_i'(s) \tag{8.10}$$

where

$$\tau = \frac{\rho V}{w}[=]\frac{kg}{kg/min}[=]min$$

The gain for $Q(t)$ is

$$\frac{1}{wC}[=]\frac{1}{\left(\dfrac{kg}{min}\right)[kJ/(kg \cdot ^\circ C)]}[=]\frac{^\circ C}{kJ/min}$$

Note that if we multiply this gain by a heat input with units of kilojoules per minute, we obtain a quantity with units of degrees Celsius, as we would expect. From Eq. (8.10) we can see that two independent quantities, the heater input and the inlet temperature, can

cause changes in the outlet temperature. The effects of these two quantities are additive (a consequence of linear equations and superposition).

If there is a change in $Q(t)$ only, then $T_i'(t) = 0$ and the transfer function relating T' to Q is

$$\frac{T'(s)}{Q(s)} = \frac{1/wC}{\tau s + 1} \qquad (8.11)$$

If there is a change in $T_i'(t)$ only, then $Q(t) = 0$ and the transfer function relating T' to T_i' is

$$\frac{T'(s)}{T_i'(s)} = \frac{1}{\tau s + 1} \qquad (8.12)$$

Equation (8.10) is represented by the block diagram shown in Fig. 8–3a. This diagram is simply an alternate way to express Eq. (8.10) in terms of the transfer functions of Eqs. (8.11) and (8.12). Superposition makes this representation possible. Notice that in Fig. 8–3 we have indicated summation of signals by the symbol shown in Fig. 8–4, which is called a *summing junction*. Subtraction can also be indicated with this symbol by placing a minus sign at the appropriate input. The summing junction was used previously as the symbol for the comparator of the controller (see Fig. 8–2). This symbol, which is standard in the control literature, may have several inputs but only one output.

A block diagram that is equivalent to Fig. 8–3a is shown in Fig. 8–3b. That this diagram is correct can be seen by rearranging Eq. (8.10); thus

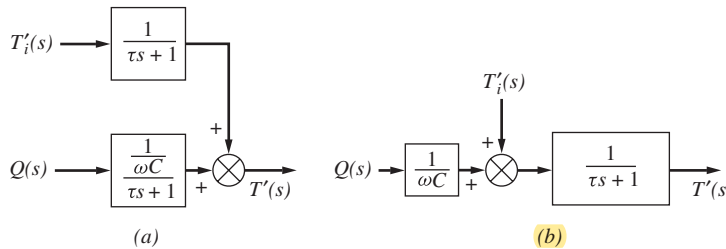$$T'(s) = \left[\frac{1}{wC} Q(s) + T_i'(s) \right] \frac{1}{\tau s + 1} \qquad (8.13)$$

(a)   *(b)*

FIGURE 8–3
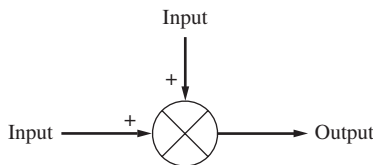Block diagram for process.

FIGURE 8–4
Summing junction.

In Fig. 8–3b, the input variables $(1/wC)Q(s)$ and $T_i'(s)$ are summed before being operated on by the transfer function $1/(\tau s + 1)$. Note that in all cases the units in the block diagram must be consistent. The quantities being combined at a summing junction must have the same units. The readers should convince themselves that this is indeed the
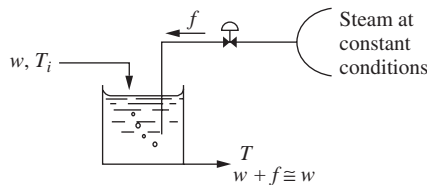
FIGURE 8–5
Supplying heat by steam.

case. We will see this again and again as we develop more complex block diagrams in future chapters. Note that the rearrangement in Eq. (8.13) is merely for mathematical convenience, and not what actually happens. Physically the two inputs (heat and inlet temperature) have independent effects on the outlet temperature.

The physical situation that exists for the control system (Fig. 8–1) if steam heating is used requires more careful analysis to show that Fig. 8–3 is an equivalent block diagram. Assume that a supply of steam at constant conditions is available for heating the tank. One method for introducing heat to the system is to let the steam flow through a control valve and discharge directly into the water in the tank, where it will condense completely and become part of the stream leaving the tank (see Fig. 8–5).

If the flow of steam, f (lb/time) is small compared with the inlet flow w, the total outlet flow is approximately equal to w. When the system is at steady state, the heat balance may be written

$$wC\left(T_{i_s} - T_o\right) - wC\left(T_s - T_o\right) + f_s\left(H_g - H_{l_s}\right) = 0 \qquad (8.14)$$

where T_o = reference temperature used to evaluate enthalpy of all streams entering and leaving tank
H_g = specific enthalpy of steam supplied, a constant
H_{l_s} = specific enthalpy of condensed steam flowing out at T_s, as part of total stream

The term H_{l_s} may be expressed in terms of heat capacity and temperature (assuming no phase change occurs between T_s and T_o).

$$H_{l_s} = C\left(T_s - T_o\right) \qquad (8.15)$$

From this, we see that if the steady-state temperature changes, H_{l_s} changes. In Eq. (8.14), $f_s(H_g - H_{l_s})$ is equivalent to the steady-state input q_s used previously, as can be seen by comparing Eq. (8.2) with Eq. (8.14).

Now consider an *unsteady-state* operation in which f is much less than w and the temperature T of the bath does not deviate significantly from the steady-state temperature T_s. For these conditions, we may write the unsteady-state balance approximately; thus

$$wC\left(T_i - T_o\right) - wC\left(T - T_o\right) + f\left(H_g - H_{l_s}\right) = \rho C V \frac{dT}{dt} \qquad (8.16)$$

In a practical situation for steam, H_g will be about 1000 Btu/lb$_m$. If the temperature of the bath T never deviates from T_s by more than 10°, the error in using the term $f(H_g - H_{ls})$ instead of $f(H_g - H_l)$ will be no more than 1 percent. Under these conditions, Eq. (8.16) represents the system closely, and by comparing Eq. (8.16) with Eq. (8.1), it is clear that

$$q = f(H_g - H_{ls}) \qquad (8.17)$$

Therefore, q is proportional to the flow of steam f, which may be varied by means of a control valve. It should be emphasized that the analysis presented here is only approximate. Both f and the deviation in T must be small. The smaller they become, the more closely Eq. (8.16) represents the actual physical system. An exact analysis of the problem leads to a differential equation with time-varying coefficients, and the transfer-function approach does not apply. The problem becomes considerably more difficult. A better approximation will be discussed in Chap. 20, where linearization techniques are used.

> **Example 8.1.   Stirred-tank heater model.** Let's revisit the stirred-tank heater introduced in Example 5.2 (Fig. 8–6).
>
> (a) Determine the response of the outlet temperature of the tank to a step change in the inlet temperature from 60° to 70°C.
> (b) Determine the response of the outlet temperature of the tank to a step increase in the heat input of 42 kW.
> (c) Determine the response of the outlet temperature of the tank to a simultaneous step change in the inlet temperature from 60° to 70°C and a step increase in the heat input of 42 kW.
>
> The energy balance for the stirred-tank heater is Eq. (8.10):
>
> $$T'(s) = \frac{1/wC}{\tau s + 1} Q(s) + \frac{1}{\tau s + 1} T_i'(s) \qquad (8.10)$$
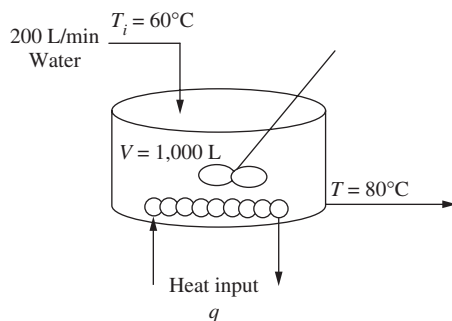

FIGURE 8–6
Stirred-tank heater revisited.

Substituting in numerical values for the variables, we obtain the actual transfer function for this stirred-tank heater.

$$\tau = \frac{\rho V}{w} = \frac{V}{w/\rho} = \frac{V}{v} = \frac{\text{tank volume}}{\text{volumetric flow rate}} = \frac{1{,}000 \text{ L}}{200 \text{ L/min}} = 5 \text{ min}$$

$$\frac{1}{wC} = \frac{1}{\left(\dfrac{200 \text{ L}}{\text{min}}\right)\left(\dfrac{1 \text{ kg}}{\text{L}}\right)\left(\dfrac{4.184 \text{ kJ}}{\text{kg}\cdot{}^\circ\text{C}}\right)\left(\dfrac{1 \text{ min}}{60 \text{ s}}\right)} = \frac{1}{14}\frac{{}^\circ\text{C}}{\text{kJ/s}} = \frac{{}^\circ\text{C}}{14 \text{ kW}}$$

$$T'(s) = \frac{1}{5s + 1}T_i'(s) + \frac{1}{5s + 1}\left(\frac{{}^\circ\text{C}}{14 \text{ kW}}\right)Q(s)$$

The block diagram for the tank is shown in Fig. 8–7.
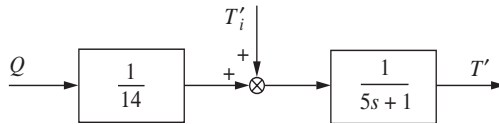

FIGURE 8–7
Block diagram for stirred-tank heater (tank and heater only).

Remember that T', T_i', and $Q(s)$ are deviation variables:

$$T' = T - 80$$
$$T_i' = T_i - 60$$
$$Q = q - q_s$$

The steady-state heat input q_s may be found from the steady-state energy balance, Eq. (8.2).

$$q_s + wC(T_{i_s} - T_o) - wC(T_s - T_o) = 0 \qquad (8.2)$$

$$q_s = wC(T_s - T_{i_s}) = \left(200 \frac{\text{kg}}{\text{min}}\right)\left(\frac{1 \text{ min}}{60 \text{ s}}\right)\left(4.184 \frac{\text{kJ}}{\text{kg}\cdot{}^\circ\text{C}}\right)(80{}^\circ\text{C} - 60{}^\circ\text{C}) = 280 \text{ kW}$$

Thus,

$$Q = q - q_s = q - 280 \text{ kW}$$

Solution

(a) If the inlet temperature is stepped from 60° to 70°C, then $T_i'(t) = 70 - 60 = 10$ and $T_i'(s) = 10/s$.
Note that $Q = 0$. Thus,

$$T'(s) = \frac{10}{s}\cdot\frac{1}{5s + 1}$$

Inverting to the time domain, we obtain the expression for $T'(t)$.
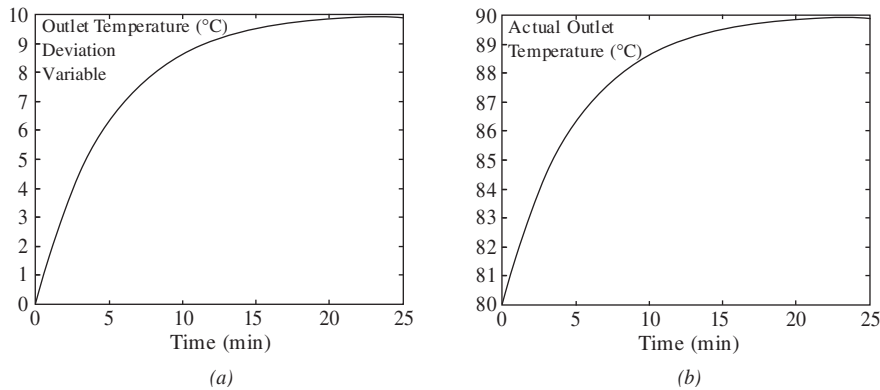
$$T'(t) = 10(1 - e^{-t/5})$$

FIGURE 8–8
Outlet temperature from stirred-tank heater. (*a*) Deviation variable; (*b*) actual temperature.

and finally, we obtain the expression for $T(t)$, the actual tank outlet temperature.

$$T(t) = T_s + T'(t) = 80 + 10(1 - e^{-t/5})$$

A plot of the outlet temperature (in deviation variables) is shown in Fig. 8–8*a*. The actual outlet temperature is shown in Fig. 8–8*b*. Note that for the uncontrolled mixing tank, a step change of 10°C in the inlet temperature ultimately produces a 10°C change in the outlet temperature. This result is just what we would expect after considering the physics of the situation.

(*b*) For a step increase in the heat input of 42 kW, $Q = 42/s$. Note that $T_i' = 0$ for this case. The expression for T' is given by

$$T'(s) = \frac{42}{s} \cdot \frac{1}{14} \frac{1}{5s + 1}$$

Inverting to the time domain, we obtain the expression for $T'(t)$.

$$T'(t) = 3(1 - e^{-t/5})$$

and finally, we obtain the expression for $T(t)$, the actual tank outlet temperature.

$$T(t) = T_s + T'(t) = 80 + 3(1 - e^{-t/5})$$

Qualitatively, this response is the same as for part (*a*); however, the ultimate temperature change due to the increased heat input is 3°C.

(*c*) If both changes occur simultaneously, the expression for $T'(t)$ is given by

$$T'(s) = \underbrace{\frac{42}{s} \cdot \frac{1}{14} \frac{1}{5s + 1}}_{\Delta T' \text{ due to change in heat input}} + \underbrace{\frac{10}{s} \cdot \frac{1}{5s + 1}}_{\Delta T' \text{ due to change in inlet temperature}}$$

Inverting, we obtain

$$T'(t) = 13(1 - e^{-t/5})$$

$$T(t) = T_s + T'(t) = 80 + 13(1 - e^{-t/5})$$

As we can see, because the system is linear, the principle of superposition applies and the effects are additive.

Measuring Element

The temperature measuring element, which senses the bath temperature T and transmits a signal T_m to the controller, may exhibit some dynamic lag. From the discussion of the mercury thermometer in Chap. 4, we observed this lag to be first-order. In this example, we will assume that the temperature measuring element is a first-order system, for which the transfer function is

$$\textbf{Measuring element transfer function}\quad \boxed{\frac{T_m'(s)}{T'(s)} = \frac{1}{\tau_m s + 1}} \qquad (8.18)$$

where the input-output variables T' and T_m' are deviation variables, defined as

$$T' = T - T_s$$

$$T_m' = T_m - T_{m_s}$$

$T'(s) \longrightarrow \boxed{\dfrac{1}{\tau_m s + 1}} \longrightarrow T_m'(s)$
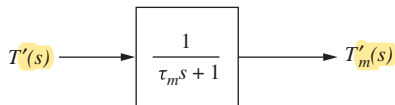
FIGURE 8–9
Block diagram of measuring element.

Note that when the control system is at steady state, $T_{m_s} = T_s$, which means that the temperature measuring element reads the true bath temperature. The transfer function for the measuring element may be represented by the block diagram shown in Fig. 8–9.

Thermocouples, commonly used temperature sensing devices in industry, have time constants on the order of 6 to 20 seconds (Riggs, 2007). The size of the time constant depends on the mass (size) of the thermocouple. Sometimes thermocouples are installed inside a protective sleeve called a thermowell that extends into the pipe or tank. The thermowell provides a protective barrier for the thermocouple against corrosion and abrasion, but can increase the effective time constant for the sensor.

Example 8.2. The temperature sensing element for the stirred-tank heater in Example 8.1 is a thermocouple. The manufacturer's specifications state that the thermocouple has a response time of 45 s (with the response time defined by the manufacturer as the time required for the thermocouple's reading to be 90 percent complete after a step change). Assuming that the thermocouple behaves as a first-order system, determine the transfer function for the temperature measuring element.

The model for the sensor is first-order. If $T'(s) = 1/s$, a unit-step change, then the sensor response is

$$T'_m(s) = \frac{1}{s}\left(\frac{1}{\tau_m s + 1}\right)$$

Inverting gives

$$T'_m(t) = (1 - e^{t/\tau_m})$$

Since the ultimate value of T'_m is 1, we know from the manufacturer's specifications that we can expect the response to be 90 percent complete at $t = 45$ s, which enables us to determine τ_m.

$$0.9 = (1 - e^{45 \text{ s}/\tau_m})$$

$$\frac{45 \text{ s}}{\tau_m} = 2.303$$

$$\tau_m = 19.5 \text{ s} = 0.33 \text{ min}$$

Therefore, the transfer function relating the actual temperature in the tank T' to the measured or indicated temperature T'_m is

$$\frac{T'_m(s)}{T'(s)} = \frac{1}{0.33s + 1}$$

The block diagram for the stirred-tank heater, including the thermocouple, is shown in Fig. 8–10.

Example 8.3.   For a step change in the inlet temperature to the stirred tank of 10°C (no change in heat input $Q = 0$), plot the actual tank temperature and the temperature indicated by the thermocouple in Example 8.2 as a function of time.

We can solve the equations by hand to obtain:

$$T'(t) = 10(1 - e^{-t/5}) \qquad \text{(from Example 8.1)}$$

and

$$T'_m(s) = \frac{10}{s(0.33s + 1)(5s + 1)} = \frac{0.71}{s + 3.03} - \frac{10.71}{s + 0.2} + \frac{10}{s}$$

Inverting to get $T'_m(t)$ yields

$$T'_m(t) = 10 + 0.71e^{-3.03t} - 10.71e^{-0.2t}$$

Plotting T' and T'_m, we obtain Fig. 8–11. From the graphs, it is clear that T'_m, lags behind T'.
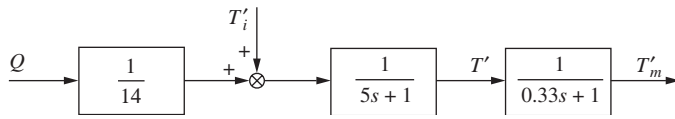
FIGURE 8–10
Block diagram for stirred-tank heater and measuring element.
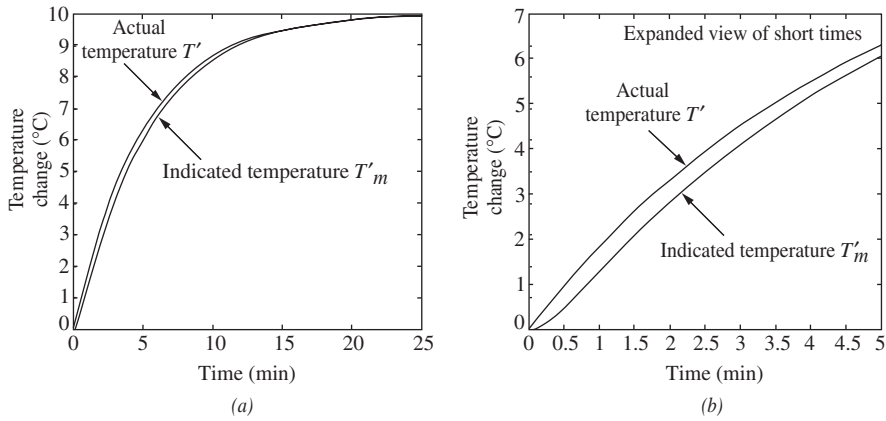

FIGURE 8–11
Comparison of temperature response for measured and actual temperatures in the stirred-tank heater. (*a*) Complete response; (*b*) expanded time scale for short times.
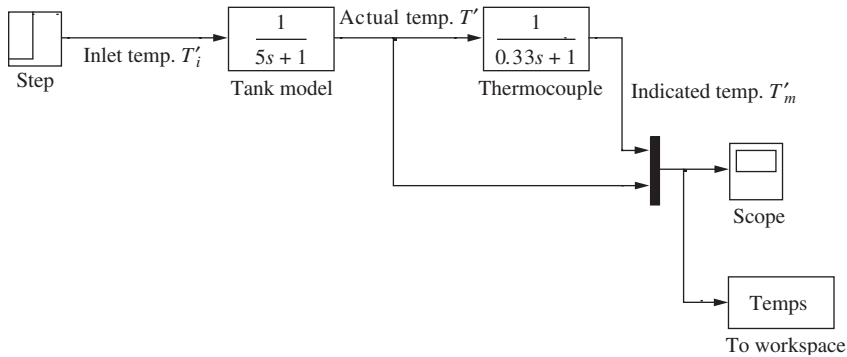

FIGURE 8–12
Simulink block diagram for Example 8.3.

We can also obtain these plots easily using Simulink (see Figs. 8–12 and 8–13). If we choose not to use the Scope, we can plot from the MATLAB workspace and obtain the same graph that we did previously (Fig. 8–14).

```
plot(temps.time,temps.signals.values)
```
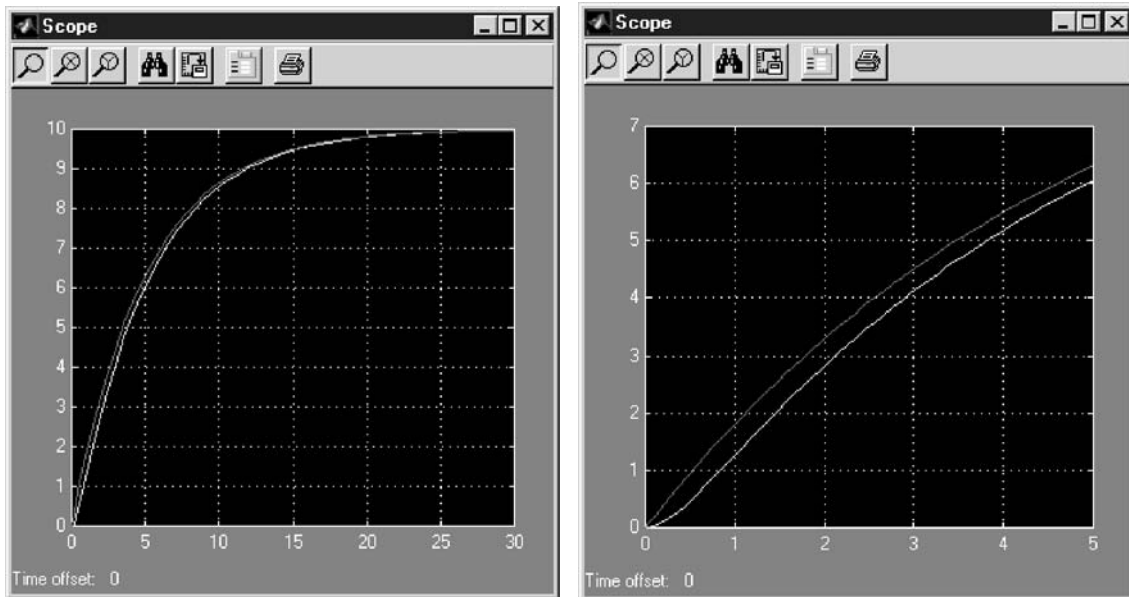

FIGURE 8–13
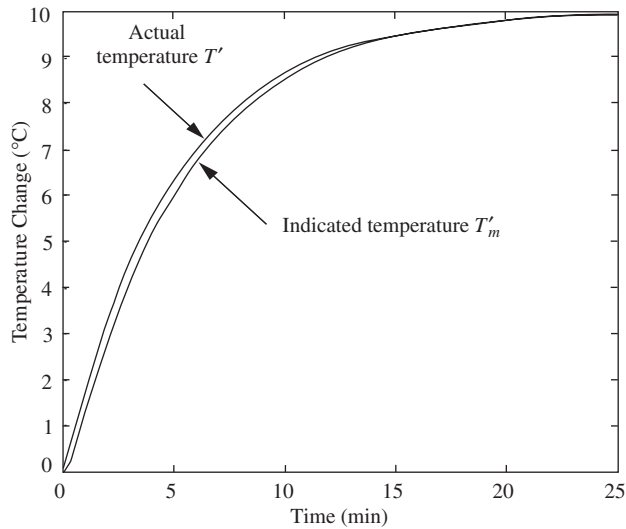Scope output from Simulink simulation of Example 8.3.

FIGURE 8–14
MATLAB plot of output for stirred-tank heater reponse.

Controller and Final Control Element

For convenience, the blocks representing the controller and the final control element are combined into one block. In this way, we need be concerned only with the overall response between the error in the temperature and the heat input to the tank. Also, it is assumed that the controller is a proportional controller. (In Chap. 9, the response of other controllers, which are commonly used in control systems, will be described.) The relationship for a proportional controller is

$$q = K_c \varepsilon + A \tag{8.19}$$

where

$$\varepsilon = T_R - T_M$$

T_R = set point temperature
K_c = proportional sensitivity or controller gain
A = heat input when $\varepsilon = 0$, also called the *bias* value (shortly we will show that $A = q_s$, the steady-state heat input)

At steady state, it is assumed that the set point, the process temperature, and the measured temperature are all equal to one another; thus

$$T_{R_s} = T_s = T_{m_s} \tag{8.20}$$

Let ε' be the deviation variable for error; thus

$$\varepsilon' = \varepsilon - \varepsilon_s \tag{8.21}$$

where $\varepsilon_s = T_{R_s} - T_{m_s}$.

Since $T_{R_s} = T_{m_s}$, $\varepsilon_s = 0$ and Eq. (8.21) becomes

$$\varepsilon' = \varepsilon - 0 = \varepsilon \tag{8.22}$$

This result shows that ε is itself a deviation variable.

Since at steady state there is no error, and A is the heat input for zero error, we see from Eq. (8.19) that

$$q_s = K_c \varepsilon_s + A = 0 + A = A$$

Note: The steady-state heat input q_s is the heat required to raise the steady-state inlet temperature from T_{i_s} to T_{R_s}, the desired steady-state set point temperature. The steady-state output from the controller/heater is termed the *bias value*. It is the output from the controller when the error is zero (i.e., steady state). It is simply calculated in this case from $q_s = wC(T_{R_s} - T_{i_s})$, which is easily deduced from Eq. (8.2).

Equation (8.19) may now be written in terms of q_s; thus

$$q = K_c \varepsilon + q_s$$

or

$$Q = K_c \varepsilon \tag{8.23}$$

where $Q = q - q_s$.

The transform of Eq. (8.23) is simply

Proportional controller transfer function $\boxed{Q(s) = K_c \varepsilon(s)}$ (8.24)

Note that ε, which is also equal to ε', may be expressed as

$$\varepsilon = T_R - T_{R_s} - (T_m - T_{m_s})$$ (8.25)

or

$$\varepsilon = T_R' - T_m'$$ (8.26)

Equation (8.25) follows from the definition of ε and the fact that $T_{R_s} = T_{m_s}$. Taking the transform of Eq. (8.26) gives

Error definition for stirred heater $\boxed{\varepsilon(s) = T_R'(s) - T_m'(s)}$ (8.27)

The transfer function for the proportional controller given by Eq. (8.24) and the definition of the error given by Eq. (8.27) may be expressed by the block diagram shown in Fig. 8–15.

   We have now completed the development of the separate blocks. If these are combined according to Fig. 8–2, we obtain the block diagram for the complete control system shown in Fig. 8–16. The reader should verify this figure. Note that when we construct a block diagram, it is indicating the flow of information around the control loop. If a line in the block diagram splits, the information in that line is merely being sent to two different places–there is not a process "stream" that is being divided in any way.
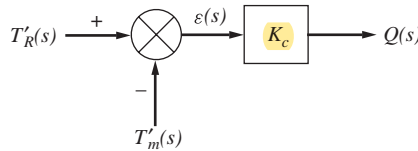

FIGURE 8–15
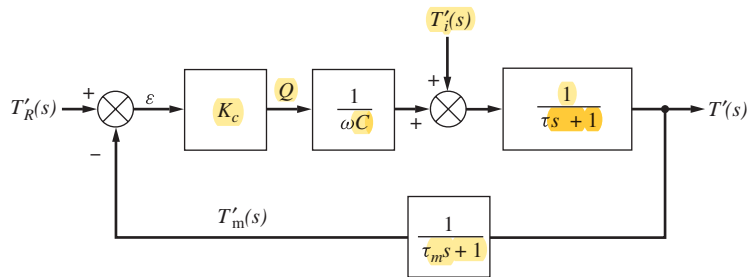Block diagram of proportional controller.

FIGURE 8–16
Block diagram of control system.

Example 8.4 Using the numerical values given in Examples 8.1 and 8.2 and a value of 20 for the proportional gain ($K_c = 20$), plot the response of the tank temperature to a change in the set point (or desired temperature) T_R' of 5°C.

At this point, we have not yet learned how to handle the block diagram algebra required to obtain this response by a hand calculation (this will be the topic of Chaps. 11 and 12). However, Simulink provides us with a very easy means of determining the response.

We merely construct the block diagram in Simulink exactly as written (see Fig. 8–17). The plot of the response of the tank temperature to a change in the set point is shown in Fig. 8–18.
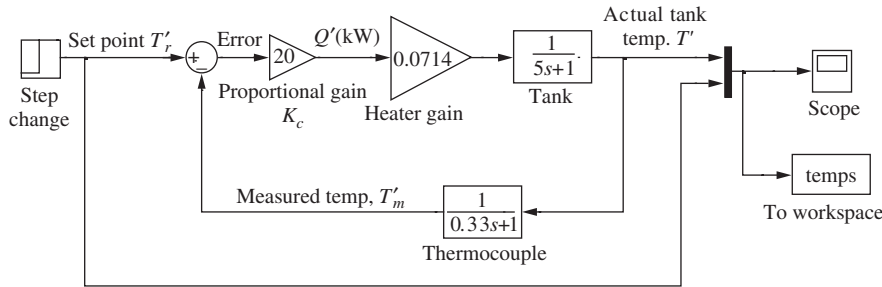

FIGURE 8–17
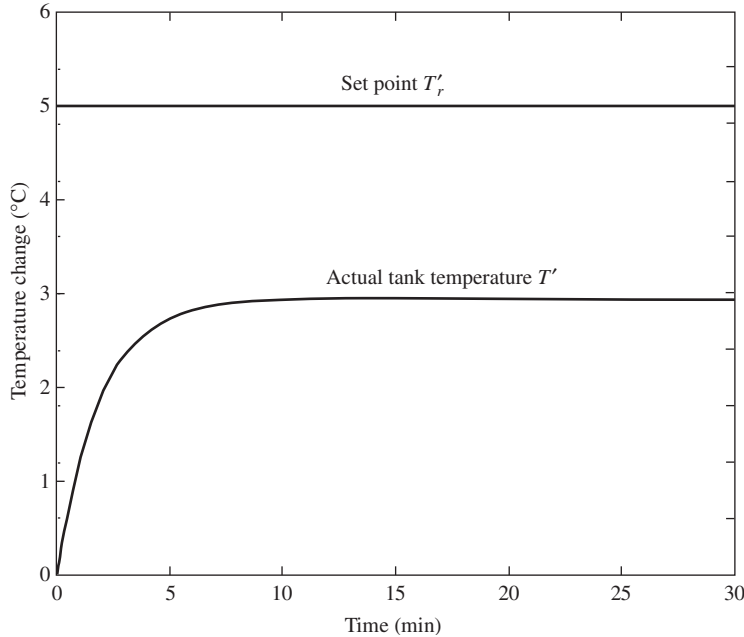Simulink model of stirred tank control system.

FIGURE 8–18
Response of stirred tank heater control system for proportional control only to a step change in the set point of 5°C ($K_c = 20$).

SUMMARY

It has been shown that a control system can be translated to a block diagram that includes the transfer functions of the various components. It should be emphasized that a block diagram is simply a systematic way of writing the simultaneous differential and algebraic equations that describe the dynamic behavior of the components. In the present case, these were Eqs. (8.10), (8.18), (8.24), and (8.27). The block diagram clarifies the relationships among the variables of these simultaneous equations. Another advantage of the block diagram representation is that it clearly shows the feedback relationship between measured variable and desired variable and how the difference in these two signals (the error ε) is used to maintain control. A set of equations generally does not clearly indicate the relationships shown by the block diagram.

In the next several chapters, tools will be developed that will enable us to reduce a block diagram such as the one in Fig. 8.16 to a single block that relates $T'(s)$ to T_i' or T_R'. We will then obtain the transient response of the control system shown in Fig. 8.16 to some specific changes in T_i' and T_R'. However, first we pause in Chap. 9 to look more carefully at the controller and control element blocks, which have been skimmed over in this chapter.

PROBLEMS

8.1. The two-tank heating process shown in Fig. P8–1 consists of two identical, well-stirred tanks in series. A flow of heat can enter tank 2. At time $t = 0$, the flow rate of heat to tank 2 suddenly increases according to a step function to 1,000 Btu/min, and the temperature of the inlet water T_i drops from 60 to 52°F according to a step function. These changes in heat flow and inlet water temperature occur simultaneously.

    (*a*) Develop a block diagram that relates the outlet temperature of tank 2 to the inlet temperature to tank 1 and the flow of heat to tank 2.

    (*b*) Obtain an expression for $T_2'(s)$ where T_2' is the deviation in the temperature of tank 2. This expression should contain numerical values of the parameters.

    (*c*) Determine $T_2(2)$ and $T_2(\infty)$.

    (*d*) Sketch the response $T_2'(t)$ versus t.

    Initially, $T_i = T_1 = T_2 = 60°F$ and $q = 0$. The following data apply:
    $w = 250$ lb/min
    Holdup volume of each tank = 5 ft^3
    Density of fluid = 50 lb/ft^3
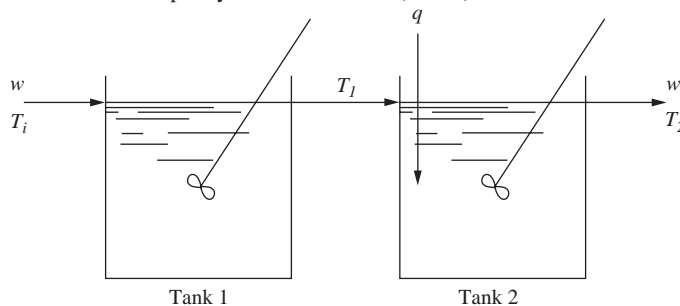   Heat capacity of fluid = 1 Btu/(lb · °F)

FIGURE P8–1

8.2. The two-tank heating process shown in Fig. P8–2 consists of two identical, well-stirred tanks in series. At steady state, $T_a = T_b = 60°F$. At time $t = 0$, the temperature of each stream entering the tanks changes according to a step function, that is, $T_a' = 10u(t)$ and $T_b' = 20u(t)$ where T_a' and T_b' are deviation variables.

(a) Develop the block diagram that relates T_2', the deviation in temperature in tank 2, to T_a' and T_b'.

(b) Obtain an expression for $T_2'(s)$.

(c) Determine $T_2(2)$.

The following data apply:

$w_1 = w_2 = 250$ lb/min

Holdup volume of each tank $= 10$ ft^3

Density of fluid $= 50$ lb/ft^3

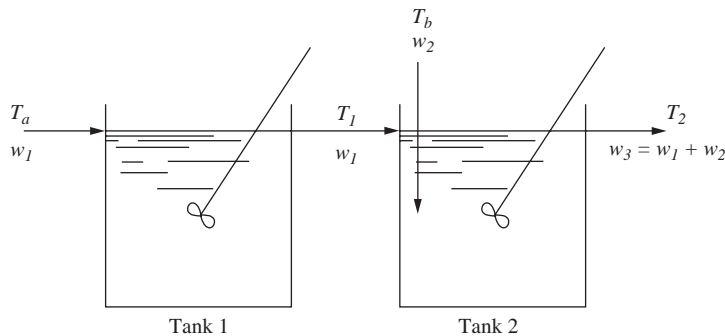Heat capacity of fluid $= 1$ Btu/(lb · °F)

Tank 1   Tank 2

FIGURE P8–2

8.3. The heat transfer equipment shown in Fig. P8–3 consists of two tanks, one nested inside the other. Heat is transferred by convection through the wall of the inner tank. The contents of each tank are well mixed. The following data and information apply:

1. The holdup volume of the inner tank is 1 ft^3. The holdup of the outer tank is 1 ft^3.
2. The cross-sectional area for heat transfer between the tanks is 1 ft^2.
3. The overall heat-transfer coefficient for the flow of heat between the tanks is 10 Btu/ (h · ft^2 · °F).
4. The heat capacity of fluid in each tank is 1 Btu/(lb · °F). The density of each fluid is 50 lb/ft^3.

Initially the temperatures of the feed stream to the outer tank and the contents of the outer tank are equal to 100°F. The contents of the inner tank are initially at 100°F. At time zero, the flow of heat to the inner tank Q is changed according to a step change from 0 to 500 Btu/h.

(a) Obtain an expression for the Laplace transform of the temperature of the inner tank $T(s)$.

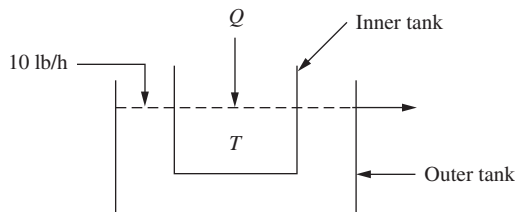(b) Invert $T(s)$ and obtain T for time $= 0, 5$ h, 10 h, and ∞.

FIGURE P8–3
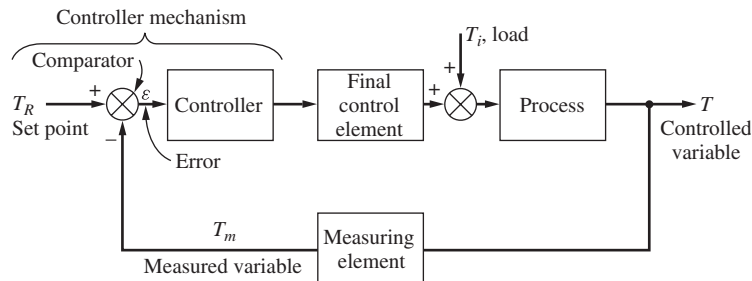
CHAPTER

8

CAPSULE SUMMARY

FIGURE 8–2
Block diagram of a simple control system.

Closed loop: The measured value of the controlled variable is fed back to the controller.

Controller: A device that outputs a signal to the process or final control element based on the magnitude of the error signal. A proportional controller outputs a signal proportional to the error.

Deviation variable: The difference between the actual value of a variable and its steady-state value. Block diagrams are *always* constructed using deviation variables.

Error: The difference between the value of the set point and that of the measured variable.

Final control element: A device that provides a modulated input to the process in response to a signal from the controller. For example, this may be a heater, a control valve, or a variety of other devices.

Load: The change in any process variable that can cause the controlled variable to change.

Measuring element: A sensor used to determine the value of the controlled variable and to send it to the comparator/controller. Examples include a thermocouple (temperature), a strain gage (pressure), a gas chromatograph (composition), and a pH electrode (acidity). These sensors typically have some dynamic behavior associated with them and can affect the design of the control system.

Negative feedback: The error is the difference between the set point and the measured variable (this is usually the desired configuration).

Positive feedback: The measured variable is added to the set point. (This is usually an undesirable situation, and frequently it leads to instability.)

Regulator problem: The goal of a control system for this type of problem is to enable the system to compensate for load changes and maintain the controlled variable at the set point.

Servo problem: The goal of a control system for this type of problem is to force the system to "track" the requested set point changes.

Set point: The desired value of the controlled variable.

CONTROLLERS AND FINAL CONTROL ELEMENTS

In Chap. 8, the block diagram representation of a simple control system (Fig. 8–2) was developed. This chapter will focus attention on the controller and final control element and will discuss the dynamic characteristics of some of these components that are in common use. As shown in Fig. 8–2, the input signal to the controller is the error, and the output signal of the controller is fed to the final control element. In many process control systems, this output signal is an air pressure, and the final control element is a pneumatic valve that opens and closes as the air pressure on the diaphragm changes.

For the mathematical analysis of control systems, it is sufficient to regard the controller as a simple computer. For example, a proportional controller may be thought of as a device that receives the error signal and puts out a signal proportional to it. Similarly, the final control element may be regarded as a device that produces corrective action on the process. The corrective action is regarded as mathematically related to the output signal from the controller. However, it is desirable to have some appreciation of the actual physical mechanisms used to accomplish this. For this reason, we begin this chapter with a physical description of a pneumatic control valve and a simplified description of a proportional controller.

Up to about 1960, most controllers were pneumatic. Although pneumatic controllers are still in use and function quite well in many installations, the controllers being installed today are electronic or computer-based instruments. For this reason, the proportional controller to be discussed in this chapter will be electronic or computer-based. The transfer functions that are presented in this chapter apply to either type of controller, and

the discussion is in no way restrictive. Other pneumatic devices, such as control valves, are found throughout chemical processing plants and are a very important part of chemical process control systems.

After the introductory discussion, transfer functions will be presented for simplified or idealized versions of the control valve and the conventional controllers. These transfer functions, for practical purposes, will adequately represent the dynamic behavior of control valves and controllers. Hence, they will be used in subsequent chapters for mathematical analysis and design of control systems.

9.1   MECHANISMS

Control Valve

The control valve shown in Fig. 9–1 contains a pneumatic device (valve motor) that moves the valve stem as the pressure on a spring-loaded diaphragm changes. The stem positions a plug in the orifice of the valve body. In the air-to-close valve, as the air pressure increases, the plug moves downward and restricts the flow of fluid through the valve. In the air-to-open valve, the valve opens and allows greater flow as the valve-top air pressure increases. The choice between air-to-open and air-to-close is usually made based on safety considerations. If the instrument air pressure fails, we would like the valve to fail in a safe position for the process. For example, if the control valve were on the cooling water inlet to a cooling jacket for an exothermic chemical reactor, we would want the valve to fail open so that we do not lose cooling water flow to the reactor. In such a situation, we would choose an air-to-close valve.

Valve motors are often constructed so that the valve stem position is proportional to the valve-top pressure. Most commercial valves move from fully open to fully closed as the valve-top pressure changes from 3 to 15 psig.

In general, the flow rate of fluid through the valve depends upon the upstream and downstream fluid pressures and the size of the opening through the valve. The plug and seat (or orifice) can be shaped so that various relationships between stem position and size of opening (hence, flow rate) are obtained. In our example, we assume for simplicity that at *steady state* the flow (for fixed upstream and downstream fluid pressures) is
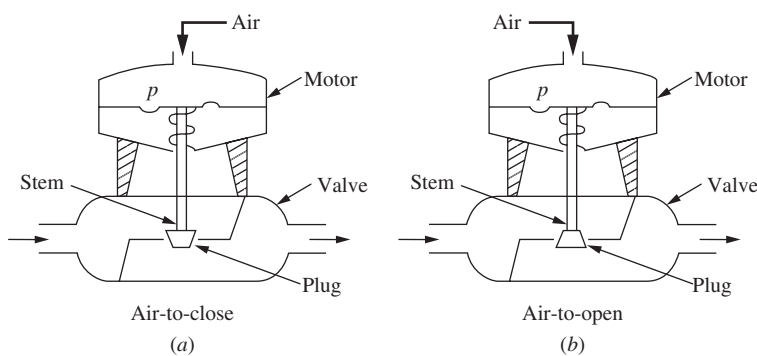

FIGURE 9–1

Pneumatic control valves. (*a*) Air to close; (*b*) air to open.

proportional to the valve-top pneumatic pressure. A valve having this relation is called a *linear valve.* A more complete discussion of control valves is presented in Chap. 19.

Controller

The control hardware required to control the temperature of a stream leaving a heat exchanger is shown in Fig. 9–2. This hardware consists of the following components listed here along with their respective conversions:

> Transducer (temperature-to-current)
> Computer/Controller (current-to-current)
> Converter (current-to-pressure)
> Control valve (pressure-to-flow rate)

Figure 9–2 shows that a thermocouple is used to measure the temperature; the signal from the thermocouple is sent to a transducer, which produces a current output in the range of 4 to 20 mA, which is a linear function of the input. The output of the transducer enters the controller where it is compared to the set point to produce an error signal. The computer/controller converts the error to an output signal in the range of 4 to 20 mA in accordance with the computer control algorithm. The only control algorithm we have considered so far has been proportional. Later in this chapter other control algorithms will be described. The output of the computer/controller enters the converter, which produces an output in the range of 3 to 15 psig, as a linear function of the input. Finally, the air pressure output of the converter is sent to the top of the control valve, which adjusts the flow of steam to the heat exchanger. We assume that the valve is linear and
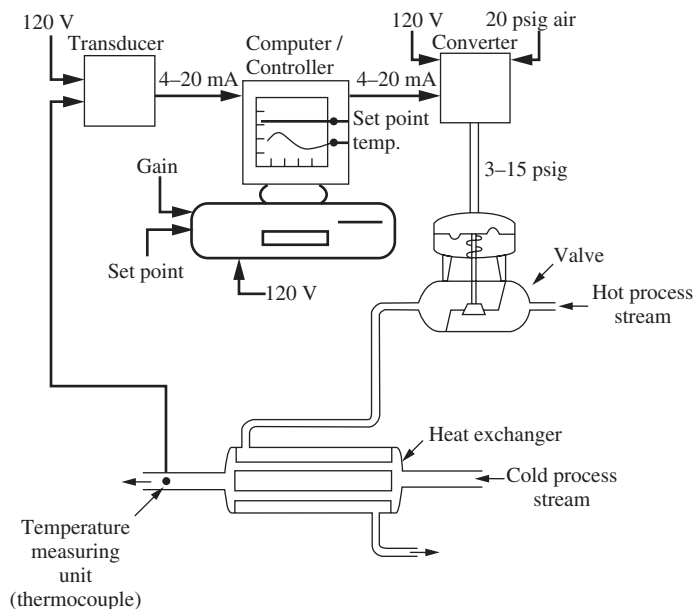

FIGURE 9–2
Schematic diagram of control system.

is the air-to-open type. The external power (120 V) needed for each component is also shown in Fig. 9–2. Electricity is needed for the transducer, computer/controller, and converter. A source of 20 psig air is also needed for the converter.

To see how the components interact with one another, consider the process to be operating at steady state with the outlet temperature equal to the set point. If the temperature of the cold process stream decreases, the following events occur: After some delay the thermocouple detects a decrease in the outlet temperature and produces a proportional change in the signal to the controller. As soon as the controller detects the drop in temperature, relative to the set point, the controller output increases according to proportional action. The increase in signal to the converter causes the output pressure from the converter to increase and to open the valve wider to admit a greater flow of the hot process stream. The increased flow of hot stream will eventually increase the output temperature and move it toward the set point. From this qualitative description, we see that the flow of signals from one component to the next is such that the outlet temperature of the heat exchanger should return toward the set point. An equivalent P&ID (piping and instrumentation diagram) for this control system is shown in Fig. 9–3 (for other P&ID symbols, see App. 9A). In a well-tuned control system, the response of the
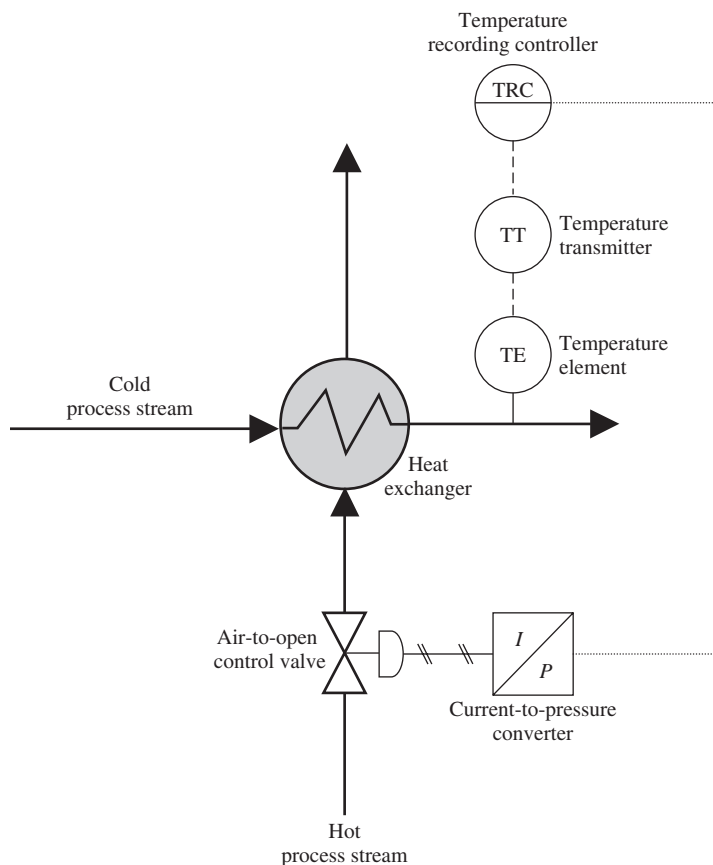

FIGURE 9–3
Piping and instrumentation diagram for control system of Fig. 9–2.

temperature will oscillate around the set point before coming to steady state. We will give considerable attention to the transient response of a control system in the remainder of this book. Further discussion will also be given on control valves in Chap. 19.

For convenience in describing various control laws (or algorithms) in the next part of this chapter, the transducer, controller, and converter will be lumped into one block, as shown in Fig. 9–4.

This concludes our brief introduction to valves and controllers. We now present transfer functions for such devices. These transfer functions, especially for controllers, are based on ideal devices that can be only approximated in practice. The degree of approximation is sufficiently good to warrant use of these transfer functions to describe the dynamic behavior of controller mechanisms for ordinary design purposes.
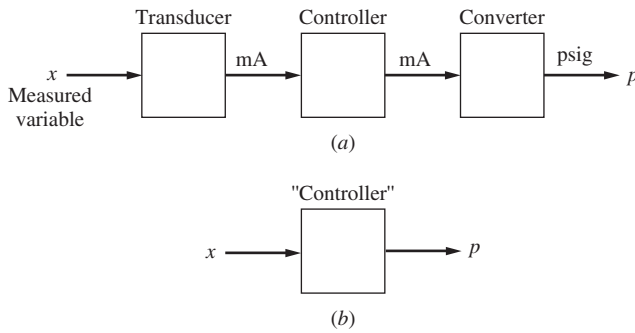

FIGURE 9–4
Equivalent block for transducer, controller, and converter.

9.2 IDEAL TRANSFER FUNCTIONS

Control Valve

A pneumatic valve always has some dynamic lag, which means that the stem position does not respond instantaneously to a change in the applied pressure from the controller. From experiments conducted on pneumatic valves, it has been found that the relationship between flow and valve-top pressure for a linear valve can often be represented by a first-order transfer function; thus

$$\boxed{\text{Control valve first-order transfer function} \qquad \frac{Q(s)}{P(s)} = \frac{K_v}{\tau_v s + 1}} \qquad (9.1)$$

where K_v is the steady-state gain, i.e., the constant of proportionality between the steady-state flow rate and the valve-top pressure, and τ_v is the time constant of the valve.

In many practical systems, the time constant of the valve is very small when compared with the time constants of other components of the control system, and the transfer function of the valve can be approximated by a constant.

$$\boxed{\text{Control valve (fast dynamics) transfer function} \qquad \frac{Q(s)}{P(s)} = K_v} \qquad (9.2)$$

Under these conditions, the valve is said to contribute negligible dynamic lag.

To justify the approximation of a fast valve by a transfer function, which is simply K_v, consider a first-order valve and a first-order process connected in series, as shown in Fig. 9–5.
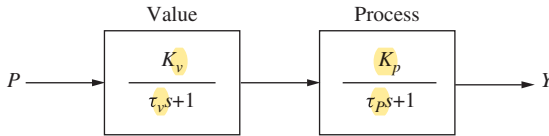

FIGURE 9–5
Block diagram for first-order valve and a first-order process.

According to the discussion of Chap. 6, if we assume no interaction (which is generally valid for this case), the relationship between the air pressure to the valve and the output from the process (perhaps a reactor temperature) is

$$\frac{Y(s)}{P(s)} = \frac{K_v K_P}{(\tau_v s + 1)(\tau_P s + 1)}$$

For a unit-step change in the valve-top pressure P,

$$Y = \frac{1}{s} \frac{K_v K_P}{(\tau_v s + 1)(\tau_P s + 1)}$$

the inverse of which is

$$Y(t) = K_v K_P \left[1 - \frac{\tau_v \tau_P}{\tau_v - \tau_P} \left(\frac{1}{\tau_P} e^{-t/\tau_v} - \frac{1}{\tau_v} e^{-t/\tau_P} \right) \right]$$

If $\tau_v \ll \tau_P$, this equation is approximately

$$Y(t) = K_v K_P \left(1 - e^{-t/\tau_P} \right)$$

The last expression is the unit-step response of the transfer function

$$\frac{Y(s)}{P(s)} = K_v \frac{K_P}{\tau_P s + 1}$$

so that the combination of process and valve is essentially first-order. This clearly demonstrates that when the time constant of the valve is much smaller than that of the process, the valve transfer function can be taken as K_v.

A typical pneumatic valve has a time constant of the order of 1 s. Many industrial processes behave as first-order systems or as a series of first-order systems having time constants that may range from a minute to an hour. For these systems we have shown that the lag of the valve is negligible, and we will make frequent use of this approximation.

Controllers

In this section, we present the transfer functions for the controllers frequently used in industrial processes. Because the transducer and the converter will be lumped together with the controller for simplicity, the result is that the input will be the measured variable x (e.g., temperature and fluid level) and the output will be a pneumatic signal p. (See Fig. 9–4.) Actually this form (x as input and p as output) applies to a pneumatic controller. For convenience, we will refer to the lumped components as the controller in the following discussion, even though the actual electronic controller is but one of the components.

PROPORTIONAL CONTROL. The simplest type of controller is the proportional controller. (The ON/OFF control is really the simplest, but it is a special case of the proportional controller as we'll see shortly.) Our goal is to reduce the error between the process output and the set point. The proportional controller, as we will see, can reduce the error, but cannot eliminate it. If we can accept some residual error, proportional control may be the proper choice for the situation.

The proportional controller has only one adjustable parameter, the controller gain. The proportional controller produces an output signal (pressure in the case of a pneumatic controller, current, or voltage for an electronic controller) that is proportional to the error ε. This action may be expressed as

$$\boxed{\begin{array}{l}\text{Proportional} \\ \text{controller}\end{array} \quad p = K_c\varepsilon + p_s} \qquad (9.3)$$

where p = output signal from controller, psig or mA
 K_c = proportional gain, or sensitivity
 ε = error = (set point) − (measured variable)
 p_s = a constant, the steady-state output from the controller [the bias value, see Eqs. (8.19) and (8.23)]

The error ε, which is the difference between the set point and the signal from the measuring element, may be in any suitable units. However, the units of the set point and the measured variable must be the same, since the error is the difference between these quantities.

In a controller having adjustable gain, the value of the gain K_c can be varied by entering it into the controller, usually by means of a keypad (or a knob on older equipment). The value of p_s is the value of the output signal when ε is zero, and in most controllers p_s can be adjusted to obtain the required output signal when the control system is at steady state and $\varepsilon = 0$.

To obtain the transfer function of Eq. (9.3), we first introduce the deviation variable

$$P = p - p_s$$

into Eq. (9.3). At time $t = 0$, we assume the error ε_s to be zero. Then ε is already a deviation variable. Equation (9.3) becomes

$$P(t) = K_c \varepsilon(t) \tag{9.4}$$

Taking the transform of Eq. (9.4) gives the transfer function of an ideal proportional controller.

$$\boxed{\text{Proportional controller transfer function} \qquad \frac{P(s)}{\varepsilon(s)} = K_c} \tag{9.5}$$

The actual behavior of a proportional controller is depicted in Fig. 9–6. The controller output will saturate (level out) at $p_{max} = 15$ psig or 20 mA at the upper end and at $p_{min} = 3$ psig or 4 mA at the lower end of the output. The ideal transfer function Eq. (9.5) does not predict this saturation phenomenon.
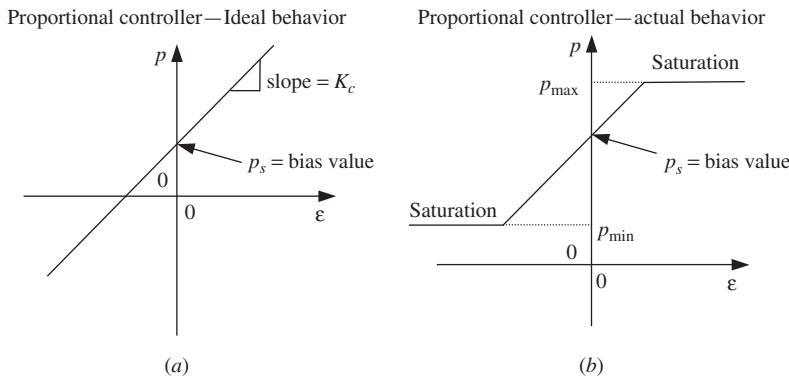
Proportional controller — Ideal behavior

Proportional controller — actual behavior

FIGURE 9–6
Proportional controller output as a function of error input to the controller. *(a)* Ideal behavior; *(b)* actual behavior.

The next example will help to clarify the concept of controller gain.

Example 9.1.   A pneumatic proportional controller is used in the process shown in Fig. 9–2 to control the cold stream outlet temperature within the range of 60 to 120°F. The controller gain is adjusted so that the output pressure goes from 3 psig (valve fully closed) to 15 psig (valve fully open) as the measured temperature goes from 71 to 75°F with the set point held constant. Find the controller gain K_c.

$$\text{Gain} = \frac{\Delta p}{\Delta \varepsilon} = \frac{15 \text{ psig} - 3 \text{ psig}}{75°\text{F} - 71°\text{F}} = 3 \text{ psi/°F}$$

Now assume that the gain of the controller is changed to 0.4 psi/°F. Find the error in temperature that will cause the control valve to go from fully closed to fully open.

$$\Delta T = \frac{\Delta p}{\text{gain}} = \frac{12 \text{ psi}}{0.4 \text{ psi/°F}} = 30°\text{F}$$

At this level of gain, the valve will be fully open if the error signal reaches 30°F. The gain K_c has the units of psi per unit of measured variable. [Regarding the units on controller gain, if the actual controller of Fig. 9–4 is considered, both the input and the output units are in milliamperes. In this case the gain will be dimensionless (i.e., mA/mA).]

ON/OFF CONTROL. A special case of proportional control is on/off control. If the gain K_c is made very high, the valve will move from one extreme position to the other if the process deviates only slightly from the set point. This very sensitive action is called on/off action because the valve is either fully open (on) or fully closed (off); i.e., the valve acts as a switch. This is a very simple controller and is exemplified by the thermostat used in a home-heating system. In practice, a dead band is inserted into the controller. With a dead band, the error reaches some finite positive value before the controller "turns on." Conversely, the error must fall to some finite negative value before the controller "turns off." This behavior is shown in Fig. 9–7. Expanding the width of the dead band makes the controller less sensitive to noise and prevents the phenomenon of *chattering,* where the controller will rapidly cycle on and off as the error fluctuates about zero. Chattering is detrimental to equipment performance.
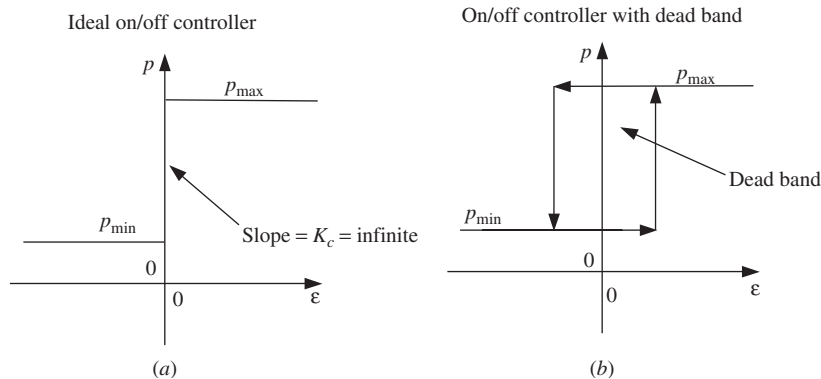

FIGURE 9–7
Output from on/off controller as a function of error input to the controller. *(a)* Ideal on/off controller; *(b)* on/off controller with dead band.

For various reasons, it is often desirable to add other modes of control to the basic proportional action. These modes, integral and derivative action, are discussed below with the objective of obtaining the ideal transfer functions of the expanded controllers. The reasons for introducing these modes will be discussed briefly at the end of this chapter and in greater detail in later chapters.

PROPORTIONAL-INTEGRAL (PI) CONTROL. If we cannot tolerate any residual error, we will have to introduce an additional control mode: integral control. If we add integral control to our proportional controller, we have what is termed PI, or proportional-integral

control. The integral mode ultimately drives the error to zero. This controller has two adjustable parameters for which we select values, the gain and the integral time. Thus it is a bit more complicated than a proportional controller, but in exchange for the additional complexity, we reap the advantage of no error at steady state.

PI control is described by the relationship

$$\boxed{\text{Proportional-integral controller} \qquad p = K_c\varepsilon + \frac{K_c}{\tau_I}\int_0^t \varepsilon\,dt + p_S} \qquad (9.6)$$

where K_c = proportional gain
τ_I = integral time, min
p_s = constant (the bias value)

In this case, we have added to the proportional action term $K_c\varepsilon$ another term that is proportional to the integral of the error. The values of K_c and τ_I are both adjustable.

To visualize the response of this controller, consider the response to a unit-step change in error, as shown in Fig. 9–8. This unit-step response is most directly obtained by inserting $\varepsilon = 1$ into Eq. (9.6), which yields

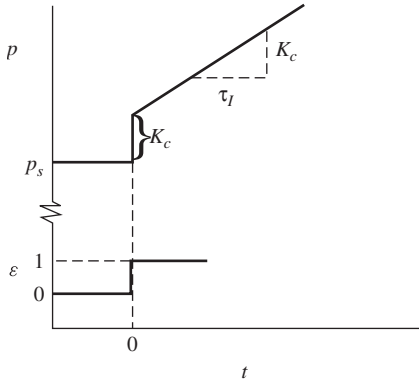$$p(t) = K_c + \frac{K_c}{\tau_I}t + p_S \qquad (9.7)$$

FIGURE 9–8
Response of a PI controller to a unit-step change in error.

Notice that p changes suddenly by an amount K_c and then changes linearly with time at a rate K_c/τ_I.

To obtain the transfer function of Eq. (9.6), we again introduce the deviation variable $P = p - p_s$ into Eq. (9.6) and then take the transform to obtain

$$\boxed{\text{Proportional-integral controller transfer function} \qquad \frac{P(s)}{\varepsilon(s)} = K_c\left(1 + \frac{1}{\tau_I s}\right)} \qquad (9.8)$$

Some manufacturers prefer to use the term *reset rate,* which is defined as the reciprocal of τ_I. The integral adjustment on a controller may be denoted by integral time or reset rate (carefully check the specific controller to be sure which value to enter). The calibration of the proportional and integral action is often checked by observing the jump and slope of a step response, as shown in Fig. 9–8.

PROPORTIONAL-DERIVATIVE (PD) CONTROL. Derivative control is another mode that can be added to our proportional or proportional-integral controllers. It acts upon the derivative of the error, so it is most active when the error is *changing* rapidly. It serves to reduce process oscillations.

This mode of control may be represented by

$$\text{Proportional-derivative controller} \qquad p = K_c\varepsilon + K_c\tau_D \frac{d\varepsilon}{dt} + p_s \qquad (9.9)$$

where K_c = proportional gain
  τ_D = derivative time, min
  p_s = constant (bias value)

In this case, we have added to the proportional term another term $K_c \tau_D \, d\varepsilon/dt$, which is proportional to the derivative of the error. The values of K_c and τ_D are both adjustable. Other terms that are used to describe the derivative action are *rate control* and *anticipatory control.*

The action of this controller can be visualized by considering the response to a linear change in error as shown in Fig. 9–9.
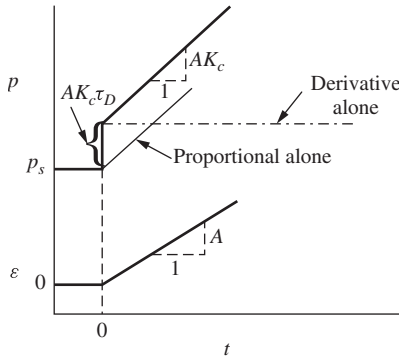

FIGURE 9–9
Response of a PD controller to a ramp input in error.

This response is obtained by introducing the linear function $\varepsilon(t) = At$ into Eq. (9.9) to obtain

$$p(t) = AK_ct + AK_c\tau_D + p_s$$

Notice that p changes suddenly by an amount $AK_c \, \tau_D$ as a result of the derivative action and then changes linearly at a rate AK_c. The effect of derivative action in this case is to

anticipate the linear change in error by adding output $AK_c \, \tau_D$ to the proportional action. The controller is taking preemptive action to counter the anticipated change in the error that it predicted from the slope of the error versus time curve.

To obtain the transfer function from Eq. (9.9), we introduce the deviation variable $P = p - p_s$ and then take the transform to obtain

$$
\boxed{\text{Proportional-derivative controller transfer function} \qquad \frac{P(s)}{\varepsilon(s)} = K_c \left(1 + \tau_D s\right)} \qquad (9.10)
$$

PROPORTIONAL-INTEGRAL-DERIVATIVE (PID) CONTROL. This mode of control is a combination of the previous modes and is given by the expression

$$
\boxed{\text{Proportional-integral-derivative controller} \qquad p = K_c \varepsilon + K_c \tau_D \frac{d\varepsilon}{dt} + \frac{K_c}{\tau_I} \int_0^t \varepsilon \, dt + p_s} \qquad (9.11)
$$

In this case, all three values K_c, τ_D, and τ_I can be adjusted in the controller. The transfer function for this controller can be obtained from the Laplace transform of Eq. (9.11); thus

$$
\boxed{\text{Proportional-integral-derivative controller transfer function} \qquad \frac{P(s)}{\varepsilon(s)} = K_c \left(1 + \tau_D s + \frac{1}{\tau_I s}\right)} \qquad (9.12)
$$

Derivative action is based on how rapidly the error is changing, not the magnitude of the error or how long the error has persisted. It is based on the slope of the error versus time curve at any instant in time. Therefore, a rapidly changing error signal will induce a large derivative response. "Noisy" error signals cause significant problems for derivative action because of the rapidly changing slope of the error caused by noise. Derivative control should be avoided in these situations unless the error signal can be filtered to remove the noise.

Motivation for Addition of Integral and Derivative Control Modes

Having introduced ideal transfer functions for integral and derivative modes of control, we now wish to indicate the practical motivation for use of these modes. The curves of Fig. 9–10 show the behavior of a typical feedback control system using different kinds of control when it is subjected to a permanent disturbance. This may be visualized in terms of the stirred-tank temperature control system of Chap. 8 after a step change in T_i. The value of the controlled variable is seen to rise at time zero owing to the disturbance. With no control, this variable continues to rise to a new steady-state value. With control, after some time the control system begins to take action to try to maintain the controlled variable close to the value that existed before the disturbance occurred.
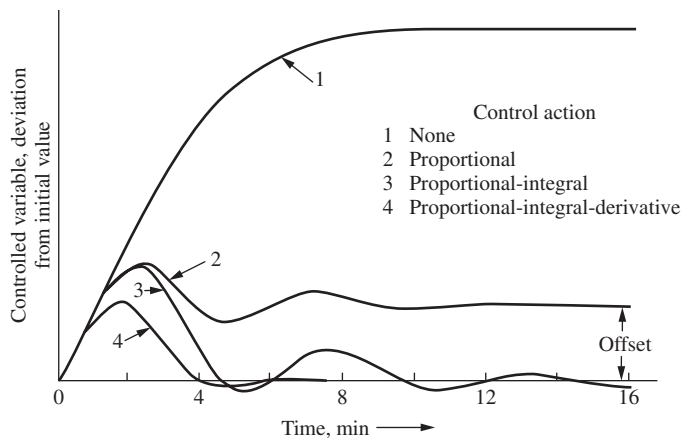
FIGURE 9–10
Response of a typical control system showing the effects of various modes of control.

With proportional action only, the control system is able to arrest the rise of the controlled variable and ultimately bring it to rest at a new steady-state value. The difference between this new steady-state value and the original value (the set point, in this case) is called *offset.* For the particular system shown, the offset is seen to be only about 20 percent of the ultimate change that would have been realized for this disturbance in the absence of control.

As shown by the PI curve, the addition of integral action eliminates the offset; the controlled variable ultimately returns to the original value. This advantage of integral action is balanced by the disadvantage of a more oscillatory behavior.

The addition of derivative action to the PI action gives a definite improvement in the response. The rise of the controlled variable is arrested more quickly, and it is returned rapidly to the original value with little or no oscillation. Discussion of the PD mode is deferred to a later chapter.

The selection among the control systems whose responses are shown in Fig. 9–10 depends on the particular application. If an offset of about 20 percent is tolerable, proportional action would likely be selected. If no offset were tolerable, integral action would be added. If excessive oscillations had to be eliminated, derivative action might be added. The addition of each mode means, as we will see in later chapters, more difficult controller adjustment. Our goal in forthcoming chapters will be to present the material that will enable the reader to develop curves such as those of Fig. 9–10 and thereby to design efficient, economic control systems.

SUMMARY

In this chapter we have presented a brief discussion of control valves and controllers. In addition, we presented ideal transfer functions to represent their dynamic behavior and some typical results of using these controllers.

The ideal transfer functions actually describe the action of many types of controllers, including pneumatic, electronic, computer-based, hydraulic, mechanical, and electrical systems. Hence, the mathematical analyses of control systems to be presented in later chapters, which are based upon first- and second-order systems, transportation lags, and ideal controllers, generalize to many branches of the control field. After studying this text on process control, the reader should be able to apply the knowledge to, e.g., problems in mechanical control systems. All that is required is a preliminary study of the physical nature of the systems involved.

PROBLEMS

9.1. A pneumatic PI temperature controller has an output pressure of 10 psig when the set point and process temperature coincide. The set point is suddenly increased by 10°F (i.e., a step change in error is introduced), and the following data are obtained:

Time, s	psig
0−	10
0+	8
20	7
60	5
90	3.5

Determine the actual gain (psig per degree Fahrenheit) and the integral time.

9.2. A unit-step change in error is introduced into a PID controller. If $K_c = 10$, $\tau_I = 1$, and $\tau_D = 0.5$, plot the response of the controller $P(t)$.

9.3. An ideal PD controller has the transfer function

$$\frac{P}{\varepsilon} = K_c(\tau_D s + 1)$$

An actual PD controller had the transfer function

$$\frac{P}{\varepsilon} = K_c \frac{\tau_D s + 1}{(\tau_D/\beta)s + 1}$$

where β is a large constant in an industrial controller.

If a unit-step change in error is introduced into a controller having the second transfer function, show that

$$P(t) = K_c\left(1 + Ae^{-\beta t/\tau_D}\right)$$

where A is a function of β which you are to determine. For $\beta = 5$ and $K_c = 0.5$, plot $P(t)$ versus t/τ_D. As $\beta \to \infty$, show that the unit-step response approaches that for the ideal controller.

9.4. A PID temperature controller is at steady state with an output pressure of 9 psig. The set point and process temperature are initially the same. At time $t = 0$, the set point is increased at the rate of 0.5°F/min. The motion of the set point is in the direction of *lower* temperatures. If the current settings are

$$K_c = 2 \text{ psig/}°F$$
$$\tau_i = 1.25 \text{ min}$$
$$\tau_D = 0.4 \text{ min}$$

plot the output pressure versus time.

9.5. The input ε to a PI controller is shown in Fig. P9–5. Plot the output of the controller if $K_c = 2$ and $\tau_I = 0.50$ min.
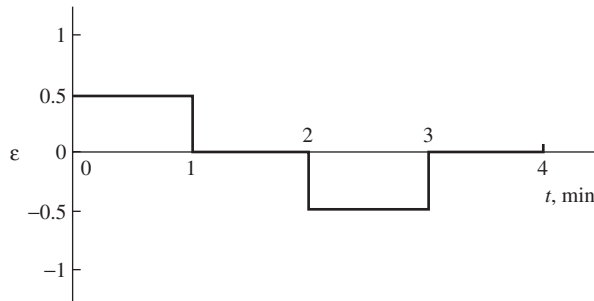

FIGURE P9–5

9.6. A PI controller has the transfer function

$$G_c = \frac{5s + 10}{s}$$

Determine the values of K_c and τ_I.

9.7. Dye for our new line of blue jeans is being blended in a mixing tank. The desired color of blue is produced using a concentration of 1500 ppm blue dye, with a minimum acceptable concentration of 1400 ppm. At 9 A.M. today the dye injector plugged, and the dye flow was interrupted for 10 min, until we realized the problem and unclogged the nozzle. see Fig. P9–7.
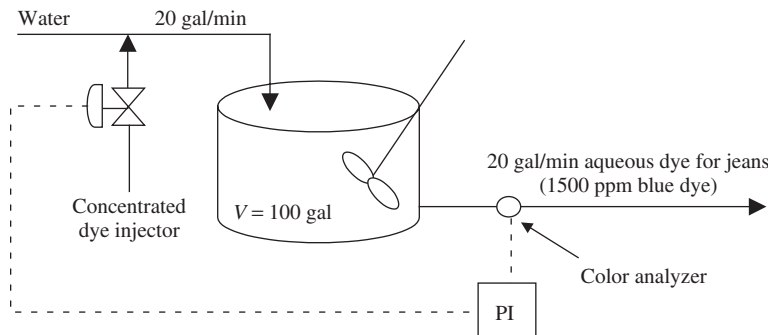

FIGURE P9–7

Plot the controller ouput from 9 A.M. to 9:10 A.M. The steady-state controller output (the bias value) is 8 psig. Does the controller output saturate (output range is 3 to 15 psig)? If so, at what time does it occur? The controller is a PI controller with $K_c = 0.001$ psig/ppm and $\tau_I = 1$ min.

CAPSULE SUMMARY

CONTROL VALVES

Two basic types of control valves are air-to-close and air-to-open. The air pressure (pneumatic) signal is usually 3 to 15 psig. The dynamics of the valves are adequately modeled as first-order systems. The time constant is on the order of 1 s.


Air-to-close (*a*)   Air-to-open (*b*)
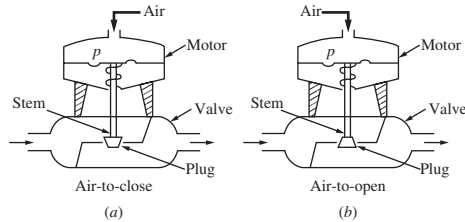
Control valuve first-order transfer function	$\dfrac{Q(s)}{P(s)} = \dfrac{K_v}{\tau_v s + 1}$	Control valve (fast dynamics) transfer function	$\dfrac{Q(s)}{P(s)} = K_v$

CONTROLLERS

Controller type	Time domain model	Transfer function
Proportional (P)	$p = K_c \varepsilon + p_s$	$\dfrac{P(s)}{\varepsilon(s)} = K_c$
Proportional-integral (PI)	$p = K_c \varepsilon + \dfrac{K_c}{\tau_I} \displaystyle\int_0^t \varepsilon \, dt + p_s$	$\dfrac{P(s)}{\varepsilon(s)} = K_c \left(1 + \dfrac{1}{\tau_I s} \right)$
Proportional-derivative (PD)	$p = K_c \varepsilon + K_c \tau_D \dfrac{d\varepsilon}{dt} + p_s$	$\dfrac{P(s)}{\varepsilon(s)} = K_c (1 + \tau_D s)$
Proportional-integral-derivative (PID)	$p = K_c \varepsilon + K_c \tau_D \dfrac{d\varepsilon}{dt} + \dfrac{K_c}{\tau_I} \displaystyle\int_0^t \varepsilon \, dt + p_s$	$\dfrac{P(s)}{\varepsilon(s)} = K_c \left(1 + \tau_D s + \dfrac{1}{\tau_I s} \right)$
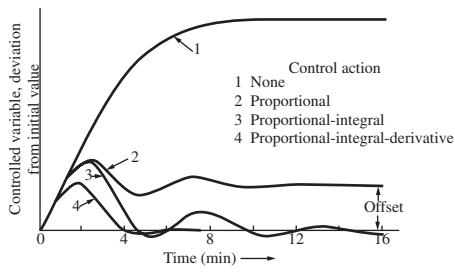
(Figure 9.10 Response of a typical control system showing the effects of various modes of control.)

CHAPTER 9 APPENDIX

PIPING AND INSTRUMENTATION DIAGRAM SYMBOLS

Piping and instrumentation diagrams (P&IDs) are used by the chemical process industry to document the control systems for their processes. The Instrument Society of America (ISA) produces a standards document [ISA-5.1-1984-(R1992)] "Instrument Symbols and Identification" that establishes a uniform means of designating instruments, control systems, and sensors used for measurement and control in a process. The standard is suitable for use in the chemical, petroleum, power generation, air conditioning, metal refining, and numerous other process industries. Table 9A.1 shows some common conventions used for identifying process instrumentation and control on process drawings. Figure 9A–1 shows an example of a level control loop for a process tank. The operation is as follows. The tank level is measured using a sensor, perhaps a differential pressure cell, indicated by the bubble containing LE on the diagram. The sensor is connected to a level transmitter, LT, that sends an electrical signal (4 to 20 mA) to a level indicating controller LIC. Level alarms high and low, LAH and LAL, "monitor" the signal from the level transmitter and indicate an alarm situation if necessary. Notice from the symbols (the line through the bubble) that LIC, LAH, and LAL are all located in the control
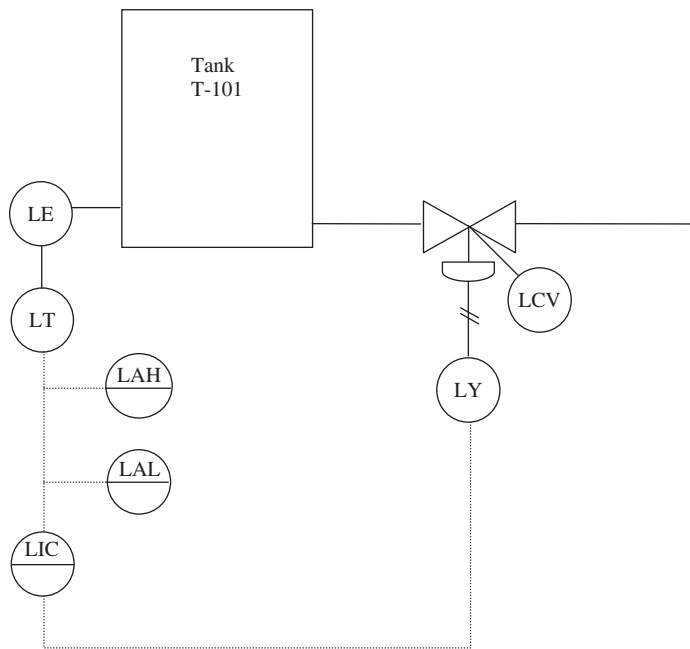

FIGURE 9A–1
Example of a level control loop using P&ID symbols.

room, while LE and LT are mounted in the field. The level indicating controller determines the necessary signal to send to the valve, based on the current level in the tank, the set point, and the selected control algorithm being used (P or PID, for example). LY computes the necessary control air pressure signal (3 to 15 psig) to send to the level control valve LCV to properly respond to the controller output signal from LIC.

TABLE 9A.1
Common symbols used on P&IDs [Instrument Society of America Standard, ISA-5.1-1984 (R1992)]

Instrument Line Symbols	
———————	Instrument supply or piping connection to process (capillary)
—⫽——⫽—	Pneumatic
....	Electrical
——○——○——	Software signal

Instrument Location and Identification	
(ABB)	Instrument located in plant
(ABB)	Instrument located on front of panel in control room
(ABB)	Instrument located on back of panel in control room.
[ABB]	Instrument accessible as part of a distributed control system

Common Identification Letters Used in Instrument Symbols		
	First Letter (A)	**Second or Third Letter (B)**
A	Analysis	Alarm
B	Burner, combustion	
C	User choice (sometimes conductivity)	Control
D	User choice (sometimes density or specific gravity)	
E	Voltage	Sensor element
F	Flow rate	
G	User choice	Glass, viewing device
H	Hand	High
I	Current (electrical)	Indicate
J	Power	
K	Time or time schedule	Control station
L	Level	Light or low
M	User choice (sometimes moisture or humidity)	Middle or intermediate
O		Orifice, restriction
P	Pressure, vacuum	Point
Q	Quantity	
R	Radioactivity or ratio	Record
S	Speed or frequency	Switch
T	Temperature	Transmit
V	Viscosity or vibration	Valve, damper or louver
W	Weight or force	Well
Y	Event	Relay, compute, convert
Z	Position	Drive

CHAPTER
10

BLOCK DIAGRAM OF A CHEMICAL-REACTOR CONTROL SYSTEM

To tie together the principles developed thus far and to illustrate further the procedure for reduction of a physical control system to a block diagram, we consider in this chapter the two-tank chemical-reactor control system of Fig. 10–1. This entire chapter serves as an example and may be omitted by the reader with no loss in continuity.
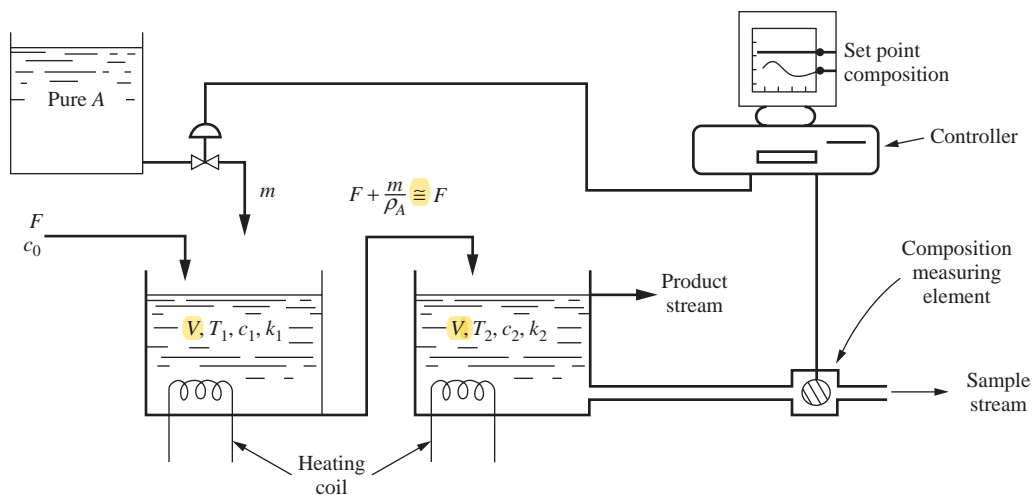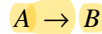

FIGURE 10–1
Control of a stirred-tank chemical reactor.

10.1   DESCRIPTION OF SYSTEM

A liquid stream enters tank 1 at a volumetric flow rate F cfm and contains reactant A at a concentration of c_0 mol A/ft^3. Reactant A decomposes in the tanks according to the irreversible chemical reaction

$$A \rightarrow B$$

The reaction is first-order and proceeds at a rate

$$r_A = -kc$$

where r_A = rate of formation of A, (mol A)/(ft$^3 \cdot$ time)

c = concentration of A, mol A/ft^3

k = reaction rate constant (a function of temperature), time^{-1} ($k_1 \equiv$ tank 1, $k_2 \equiv$ tank 2)

The reaction is to be carried out in a series of two continuous stirred-tank reactors. The tanks are maintained at different temperatures. The temperature in tank 2 is to be greater than the temperature in tank 1, with the result that k_2, the reaction rate constant in tank 2, is greater than that in tank 1, k_1. We will neglect any changes in physical properties due to chemical reaction.

The purpose of the control system is to maintain c_2, the concentration of A leaving tank 2, at some desired value in spite of variations in the inlet concentration c_0. This will be accomplished by adding a stream of pure A to tank 1 through a control valve. We wish to produce a block diagram for the process so that we can simulate its response to changes in inlet concentration.

10.2   REACTOR TRANSFER FUNCTIONS

We begin the analysis by making a material balance on A around tank 1; thus

$$\underbrace{V \frac{dc_1}{dt}}_{\substack{\text{accumulation of } A \\ \text{in tank 1}}} = \underbrace{Fc_0 + m}_{\substack{\text{flow of } A \text{ into} \\ \text{tank 1}}} - \underbrace{\left(F + \frac{m}{\rho_A}\right)c_1}_{\substack{\text{flow of } A \text{ out of} \\ \text{tank 1}}} - \underbrace{K_1 V c_1}_{\substack{\text{reaction rate} \\ \text{of } A \text{ in tank 1}}} \tag{10.1}$$

where m = molar flow rate of pure A through valve, (lb $\cdot$ mol/min)

ρ_A = density of pure A (lb $\cdot$ mol/ft^3)

V = holdup volume of tank, a constant (ft^3)

It is assumed that the volumetric flow of A through the valve m/ρ_A is much less than the inlet flow rate F, with the result that Eq. (10.1) can be written

$$V \frac{dc_1}{dt} + (F + k_1 V)c_1 = Fc_0 + m \tag{10.2}$$

This last equation may be written in the form

$$\frac{V}{F + k_1V} \frac{dc_1}{dt} + c_1 = \frac{F}{F + k_1V} c_0 + \frac{1}{F + k_1V} m$$

$$\tau_1 \frac{dc_1}{dt} + c_1 = \frac{1}{1 + k_1\tau} c_0 + \frac{1/F}{1 + k_1\tau} m \qquad (10.3)$$

where τ = residence time for each   tank $= \dfrac{V}{F}$, (time)

τ_1 = effective time constant for tank 1 $= \dfrac{V}{F + k_1V} = \dfrac{\tau}{1 + k_1\tau}$,   (time)

At steady state, $dc_1/dt = 0$, and Eq. (10.3) becomes

$$c_{1s} = \frac{1}{1 + k_1\tau} c_{0s} + \frac{1/F}{1 + k_1\tau} m_s \qquad (10.4)$$

where s refers to steady state.

Subtracting Eq. (10.4) from Eq. (10.3) and introducing the deviation variables

$$C_1 = c_1 - c_{1s}$$
$$C_0 = c_0 - c_{0s}$$
$$M = m - m_s$$

give

$$\tau_1 \frac{dC_1}{dt} + C_1 = \frac{1}{1 + k_1\tau} C_0 + \frac{1/F}{1 + k_1\tau} M \qquad (10.5)$$

Taking the transform of Eq. (10.5) yields the transfer function of the first reactor:

$$\boxed{\begin{array}{l} \text{Transfer} \\ \text{function for} \\ \text{tank 1} \end{array} \qquad C_1(s) = \frac{1/(1 + k_1\tau)}{\tau_1 s + 1}\left[C_0(s) + \frac{1}{F} M(s) \right]} \qquad (10.6)$$

A material balance on A around tank 2 gives

$$V \underbrace{\frac{dc_2}{dt}}_{\substack{\text{accumulation of } A \\ \text{in tank 2}}} = \underbrace{Fc_1}_{\substack{\text{flow of } A \\ \text{into tank 2}}} - \underbrace{Fc_2}_{\substack{\text{flow of } A \\ \text{out tank 2}}} - \underbrace{k_2Vc_2}_{\substack{\text{reaction rate} \\ \text{of } A \text{ in tank 2}}} \qquad (10.7)$$

Rearranging gives

$$V\frac{dc_2}{dt} + (F + k_2V)c_2 = Fc_1$$

$$\frac{V}{F + k_2V} \frac{dc_2}{dt} + c_2 = \frac{F}{F + k_2V} c_1 \qquad (10.7a)$$

As with tank 1, Eq. (10.7a) can be written in terms of deviation variables and arranged to give

$$\tau_2 \frac{dC_2}{dt} + C_2 = \frac{1}{1 + k_2 \tau} C_1 \tag{10.8}$$

where C_2 is the deviation variable for tank 2, $C_2 = c_2 - c_{2s}$, and τ_2 is the effective time constant for tank 2,

$$\tau_2 = \frac{V}{F + k_2 V} = \frac{V/F}{1 + k_2 (V/F)} = \frac{\tau}{1 + k_2 \tau}$$

Taking the transform of Eq. (10.8) gives the transfer function for the second reactor:

$$\boxed{\begin{array}{cc} \text{Transfer} \\ \text{function for} & C_2(s) = \dfrac{1/(1 + k_2 \tau)}{\tau_2 s + 1} C_1(s) \\ \text{tank 2} \end{array}} \tag{10.9}$$

To obtain some numerical results, we will assume the following data to apply to the system:

$\quad MW_A = 100$ lb/lb·mol A (molecular weight of A)

$\quad \rho_A = 0.8$ lb·mol/ft^3

$\quad c_{o_s} = 0.1$ lb·mol A/ft^3

$\quad F = 100$ cfm

$\quad m_s = 1.0$ lb·mol/min $\left(Note: \dfrac{m_s}{\rho_A} = \dfrac{1.0 \text{ lb·mol/min}}{0.8 \text{ lb·mol/ft}^3} = 1.25 \text{ cfm} \right)$

$\quad \left. \begin{array}{l} k_1 = \frac{1}{6} \text{ min}^{-1} \\ k_1 = \frac{2}{3} \text{ min}^{-1} \end{array} \right\} k_2 > k_1$ because $T_2 > T_1$

$\quad V = 300$ ft^3

$\quad \tau = \dfrac{V}{F} = \dfrac{300 \text{ ft}^3}{100 \text{ cfm}} = 3$ min

From Eq. (10.4), we can calculate the steady-state concentration of A in tank 1.

$$c_{1_s} = \frac{1}{1 + k_1 \tau} c_{0_s} + \frac{1/F}{1 + k_1 \tau} m_s = \frac{1}{\underbrace{1 + \left(\frac{1}{6}\right)(3)}_{\frac{1}{2}}} (0.1) + \frac{\frac{1}{100}}{1 + \frac{1}{2}} (1.0) = 0.0733 \frac{\text{lb·mol}}{\text{ft}^3}$$

We can calculate the steady-state concentration of A in tank 2 by using Eq. (10.7a) written at steady state (when $dc_2/dt = 0$).

$$c_{2_s} = \frac{1}{1 + k_2 \tau} c_{1_s} = \frac{1}{1 + \left(\frac{2}{3}\right)(3)} (0.0733) = 0.0244 \frac{\text{lb·mol}}{\text{ft}^3}$$

10.3   CONTROL VALVE

The air-to-open control valve selected for the process has the following characteristics. The flow of A through the valve varies linearly from 0 to 2 cfm as the valve-top pressure varies from 3 to 15 psig. The time constant τ_v of the valve is so small compared with the other time constants in the system that its dynamics can be neglected. From Eq. (10.2) the transfer function for a fast valve is

$$\frac{Q(s)}{P(s)} = K_v \qquad (10.2)$$

From this relationship, and remembering that Q and P are deviation variables, we can compute the valve gain using the given data

$$K_v = \frac{\Delta \text{Flow (cfm)}}{\Delta \text{Pressure (psi)}} = \frac{2 - 0}{15 - 3} = \frac{1 \text{ cfm}}{6 \text{ psi}}$$

Since $m_s/\rho_A = 1.25$ cfm, the normal operating pressure on the valve is

$$p_s = 3 + \frac{1.25}{2}(15 - 3) = 10.5 \text{ psig} \qquad (10.10)$$

Since our mole balances for the tanks are written in terms of molar flow rates instead of volumetric flow rates (cfm), we would like the valve equation to be written in terms of molar flow rates as well. Realizing that the volumetric flow rate (cfm) can be written as m/ρ_A, we can write the equation for the control valve as (see Fig. 10–2)

$$\frac{m/\rho_A - 1.25}{p - 10.5} = K_v$$

and finally,

$$m = \left[1.25 + K_v(p - 10.5)\right]\rho_A \qquad (10.11)$$

In terms of deviation variables, this can be written

$$M = K_v \rho_A P \qquad (10.12)$$

where

$$M = m - 1.25\rho_A = m - 1.0$$
$$P = p - 10.5$$

Taking the transform of Eq. (10.12) gives

$$\frac{M(s)}{P(s)} = K_v \rho_A = \left(\frac{1}{6}\frac{\text{ft}^3/\text{min}}{\text{psi}}\right)\left(\frac{0.8 \text{ lb·mol}}{\text{ft}^3}\right) = 0.133 \frac{\text{lb·mol/min}}{\text{psi}} \qquad (10.13)$$
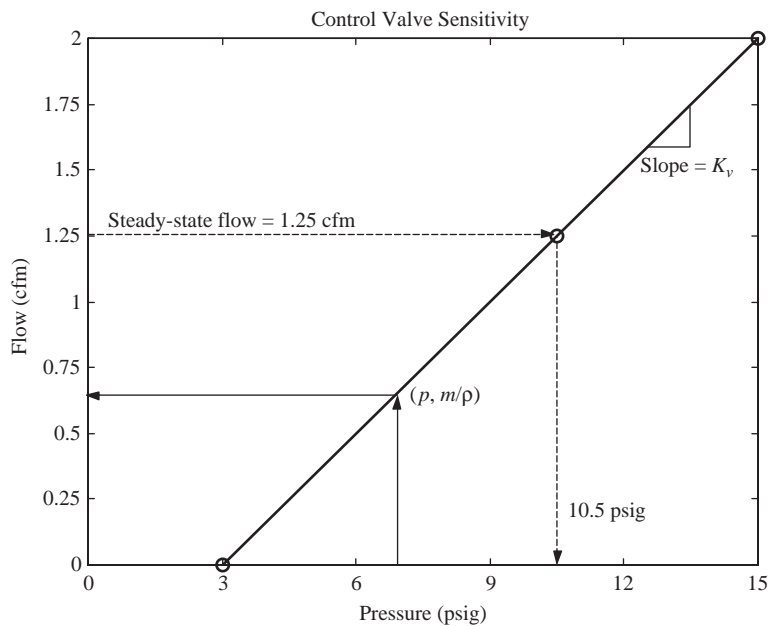
as the valve transfer function.

FIGURE 10–2
Control valve sensitivity.

10.4   MEASURING ELEMENT

For illustration, assume that the measuring element converts the concentration of A to an electronic signal. Specifically, the output of the measuring element varies from 4 to 20 mA as the concentration of A varies from 0.01 to 0.05 lb·mol A/ft^3. We will assume that the concentration measuring device is linear and has negligible lag. The sensitivity (or gain) of the measuring device is therefore

$$K_m = \frac{20 - 4}{0.05 - 0.01} = 400 \ \frac{\text{mA}}{\text{lb·mol/ft}^3}$$

Since c_{2_s} is 0.0244 lb·mol/ft^3, the normal signal from the measuring device is

$$\frac{0.0244 - 0.01}{0.05 - 0.01}(20 - 4) + 4.0 = 5.76 + 4.0 = 9.76 \text{ mA}$$

The equation for the measuring device is therefore

$$b = 9.76 + K_m(c_2 - 0.0244) = 9.76 \text{ mA} + 400 \ \frac{\text{mA}}{\text{lb·mol/ft}^3}\left(c_2 - 0.0244 \text{ lb·mol/ft}^3\right)$$

$$(10.14)$$

where b is the output current (milliamperes) from the measuring device. In terms of deviation variables, Eq. (10.14) becomes

$$B = K_m C_2 \qquad (10.15)$$

where $B = b - 9.76$ and $C_2 = c_2 - c_{2_s}$.

The transfer function for the measuring device is therefore

$$\frac{B(s)}{C_2(s)} = K_m \qquad (10.16)$$

A measuring device that changes the type of signal between its input and output is called a *transducer;* in the present case, the concentration signal is changed to a current signal. A thermocouple is another example of a transducer. It changes an input temperature signal to an output electrical signal (millivolts).

10.5   CONTROLLER

For convenience, we will assume the controller to have proportional action and produce a current output signal. The relation between the controller output signal (milliamperes) and the error (milliamperes) is

$$p = p_s + K_c(c_R - b) = p_s + K_c \varepsilon \qquad (10.17)$$

where c_R = desired current signal (or set point), mA
$\quad\quad K_c$ = controller gain, mA/mA
$\quad\quad \varepsilon$ = error = $c_R - b$, mA
In terms of deviation variables, Eq. (10.17) becomes

$$P = K_c \varepsilon \qquad (10.18)$$

The transform of this equation gives the transfer function of the controller

$$\frac{P(s)}{\varepsilon(s)} = K_c \qquad (10.19)$$

Assuming the set point and the signal from the measuring device to be the same when the system is at steady state under normal conditions, we have for the reference value of the set point

$$c_{R_s} = b_s = 9.76 \text{ mA}$$

The corresponding deviation variable for the set point is

$$C_R = c_R - c_{R_s}$$

10.6 CONTROLLER TRANSDUCER

The output from the controller is an electronic current signal ranging from 4 to 20 mA. The signal that the control valve requires is a pneumatic signal of 3 to 15 psig. A transducer is required to convert the current signal from the controller to the pneumatic signal required by the control valve. We will assume that the transducer is linear and has negligible lag. The transfer function for the transducer is merely a gain that is given by Eq. (10.20).

$$\frac{P_T(s)}{P(s)} = K_T = \frac{(15-3)\ \text{psig}}{(20-4)\ \text{mA}} = 0.75\ \frac{\text{psig}}{\text{mA}} \qquad (10.20)$$

10.7 TRANSPORTATION LAG

A portion of the liquid leaving tank 2 is continuously withdrawn through a sample line, containing a concentration measuring element, at a rate of 0.1 cfm. The measuring element must be remotely located from the process, because rigid ambient conditions must be maintained for accurate concentration measurements. The sample line has a length of 50 ft, and the cross-sectional area of the line is 0.001 ft^2.

The sample line can be represented by a transportation lag with parameter

$$\tau_d = \frac{\text{volume}}{\text{flow rate}} = \frac{(50)(0.001)}{0.1} = 0.5\ \text{min}$$

The transfer function for the sample line is, therefore,

$$e^{-\tau_d s} = e^{-0.5s}$$

10.8 BLOCK DIAGRAM

We have now completed the analysis of each component of the control system and have obtained a transfer function for each. These transfer functions can now be combined so that the overall system is represented by the block diagram in Fig. 10–3.
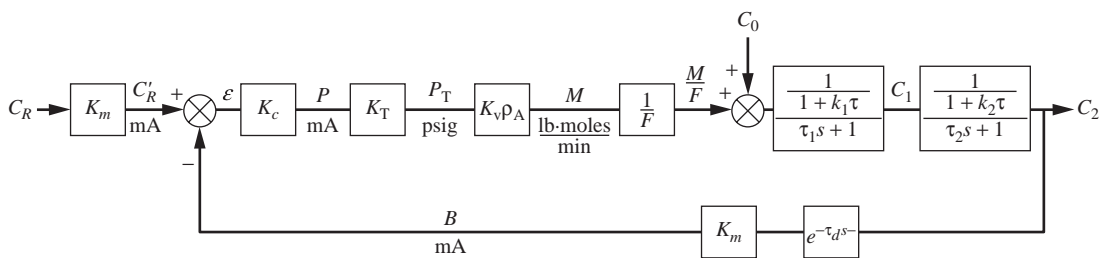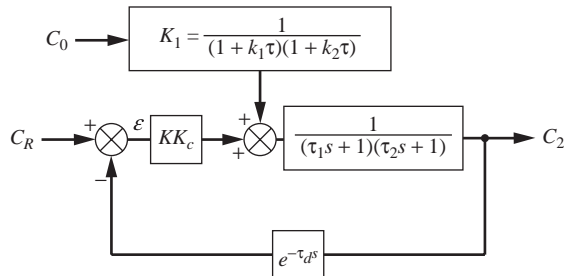

FIGURE 10–3
Block diagram for a chemical-reactor control system.

An equivalent diagram is shown in Fig. 10–4 in which some of the blocks have been combined.

$\tau_1 = 2$, $\tau_2 = 1$, $\tau_d = 0.5$, $K_1 = \dfrac{1}{4.5}$

Open-loop gain $= KK_c = \dfrac{K_m K_T K_v \rho_A}{F(1 + k_1\tau)(1 + k_2\tau)}$ $K_c = 0.09\, K_c$

FIGURE 10–4

Equivalent block diagram for a chemical-reactor control system (C_R is now in concentration units).

Numerical quantities for the parameters in the transfer functions are given in Fig. 10–4. It should be emphasized that the block diagram is written for deviation variables. The true steady-state values, which are not given by the diagram, must be obtained from the analysis of the problem.

The example analyzed in this chapter will be used later in discussion of control system design. The design problem will be to select a value of K_c that gives satisfactory control of the composition C_2 despite the rather long transportation lag involved in getting information to the controller. In addition, we will want to consider possible use of other modes of control for the system.

SUMMARY

We have now learned how to analyze a physical system and develop mathematical models for the various components of the system and then use those models to construct a block diagram of the model. We will make extensive use of these skills in upcoming chapters as we study control systems in greater depth.

PROBLEMS

10.1. In the process shown in Fig. P10–1, the concentration of salt leaving the second tank is controlled using a proportional controller by adding concentrated solution through a control valve. The following data apply:

(*a*) The controlled concentration is to be 0.1 lb salt/ft^3 solution. The inlet concentration c_i is always less than 0.1 lb/ft^3.

(*b*) The concentration of concentrated salt solution is 30 lb salt/ft^3 solution.

(c) Transducer: The output of the transducer varies linearly from 3 to 15 psig as the concentration varies from 0.05 to 0.15 lb/ft^3.

(d) Controller: The controller is a pneumatic, direct-acting, proportional controller.

(e) Control valve: As valve-top pressure varies from 3 to 15 psig, the flow through the control valve varies linearly from 0 to 0.005 cfm.

(f) It takes 30 s for the solution leaving the second tank to reach the transducer at the end of the pipe.

Draw a block diagram of the control system. Place in each block the appropriate transfer function. Calculate all the constants and give the units.
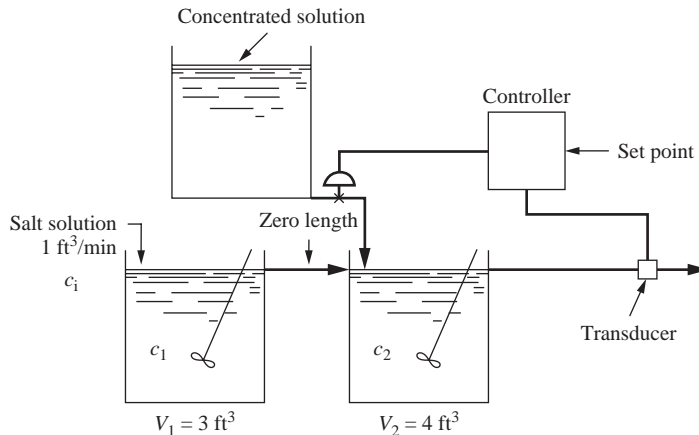

FIGURE P10–1

Use the process shown in Figs. 10–3 and 10–4 for Prob. 10.2 to 10.5.

10.2. Verify the values of τ_1 and τ_2.

10.3. Determine the steady-state value of the controller output p_s in milliamperes.

10.4. Use Simulink to simulate the open-loop response of the two chemical reactors to a step change in the feed concentration C_0 from 0.1 to 0.25 $lb \cdot mol\ A/ft^3$.

10.5. The open-loop process has an upset such that the flow rate to the process instantaneously rises to 120 cfm (from the original 100 cfm). How does the open-loop block diagram change? Plot the outlet concentration of A in both reactors as a function of time.

10.6. Two isothermal stirred-tank reactors (Fig. P10–6) are connected by a long pipe that acts as a pure time delay between the two tanks (no reaction takes place in the pipe). CSTR 1 is at a higher temperature than CSTR 2, but both temperatures remain constant. Assume constant throughputs and holdups (volumes) and a first-order, irreversible reaction taking place in each CSTR ($A \rightarrow B$). The flow rate through the system is 4 ft^3/min, and the delay time in the pipe is 30 s. The inlet concentration to CSTR 1 is initially at steady state at 1 $lb \cdot mol/ft^3$ and is increased at time 0 through a step change to 2 $lb \cdot mol/ft^3$.

(a) Draw the block diagram for the process, and be sure to include all necessary constants.

(b) Use Simulink to plot the exit concentration of A from each of the reactors.

(c) Use Simulink to plot the exit concentration of B from each of the reactors.

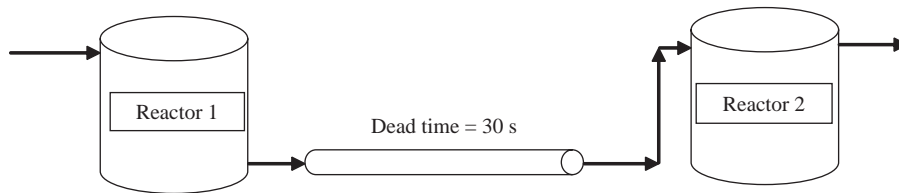| DATA		
	CSTR 1	CSTR 2
Rate constant (min^{-1})	0.3	0.15
Volume (ft^3)	25	15

FIGURE P10–6

CHAPTER
10

CAPSULE SUMMARY

A model for a two CSTRs in series was developed. The reaction that occurs in each vessel is $A \to B$. The rate of formation of A is given by $r_A = -kc$. A process schematic is shown below.
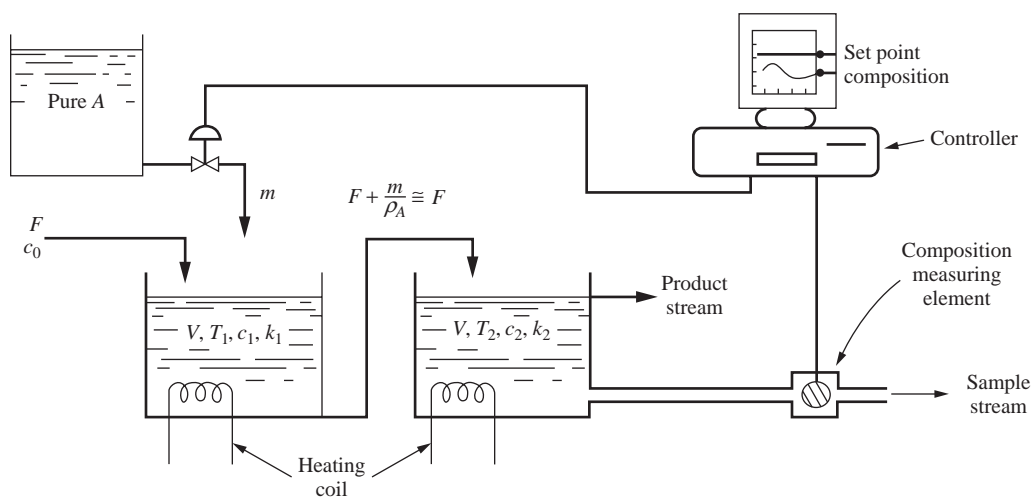

FIGURE 10–1
Control of a stirred-tank chemical reactor.

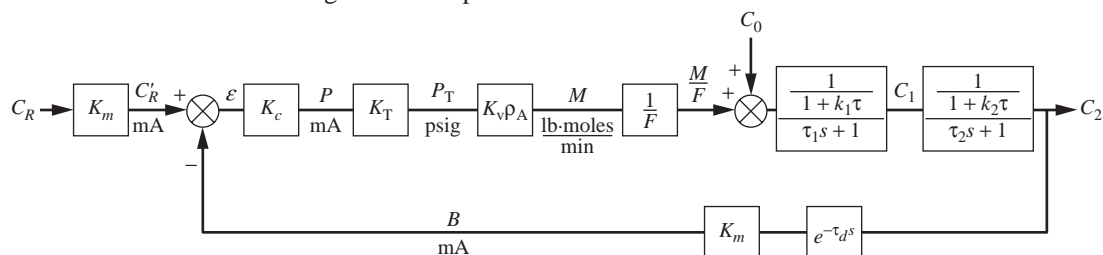A block diagram for the process is as follows:

FIGURE 10–3
Block diagram for a chemical-reactor control system.

Numerical values for the process are given below.

$$MW_A = 100 \text{ lb/lb} \cdot \text{mol } A$$

$$\rho_A = 0.8 \text{ lb} \cdot \text{mol/ft}^3$$

$$c_{os} = 0.1 \text{ lb} \cdot \text{mol } A/\text{ft}^3$$

$$F = 100 \text{ cfm}$$

$$m_s = 1.0 \text{ lb} \cdot \text{mol/min}$$

$$\frac{m_s}{\rho_A} = \frac{1.0 \text{ lb} \cdot \text{mol/min}}{0.8 \text{ lb} \cdot \text{mol/ft}^3} = 1.25 \text{ cfm}$$

$$\left.\begin{array}{l} k_1 = \frac{1}{6} \text{ min}^{-1} \\[2mm] k_2 = \frac{2}{3} \text{ min}^{-1} \end{array}\right\} \quad k_2 > k_1 \text{ because } T_2 > T_1$$

$$V = 300 \text{ ft}^3$$

$$\tau = \frac{V}{F} = \frac{300 \text{ ft}^3}{100 \text{ cfm}} = 3 \text{ min}, \ \tau_1 = 2 \text{ min}, \ \tau_2 = 1 \text{ min}$$

$$K_v = \frac{1}{6} \frac{\text{cfm}}{\text{psig}}$$

$$K_T = 0.75 \frac{\text{psig}}{\text{mA}}$$

$$K_m = 400 \frac{\text{mA}}{\text{lb} \cdot \text{mol/ft}^3}$$

$$\tau_d = 0.5 \text{ min}$$

$$c_{1s} = 0.01 \text{ lb} \cdot \text{mol/ft}^3$$

$$c_{2s} = 0.0244 \text{ lb} \cdot \text{mol/ft}^3$$

$$p_{T_s} = 10.5 \text{ psig}$$

$$m_s = 1.0 \text{ lb} \cdot \text{mol/min}$$

$$b_s = 9.76 \text{ mA}$$

$$p_s = 14 \text{ mA}$$

CHAPTER
11

CLOSED-LOOP TRANSFER FUNCTIONS

11.1 STANDARD BLOCK-DIAGRAM SYMBOLS

In Chap. 8, a block diagram was developed for the control of a stirred-tank heater (Fig. 8–2). In Fig. 11–1, the block diagram has been redrawn and incorporates some standard symbols for the variables and transfer functions, which are widely used in the control literature. These symbols are defined as follows:

R = set point or desired value

C = controlled variable

ε = error

B = variable produced by measuring element

M = manipulated variable

U = load variable or disturbance

G_c = transfer function of controller

G_1 = transfer function of final control element

G_2 = transfer function of process

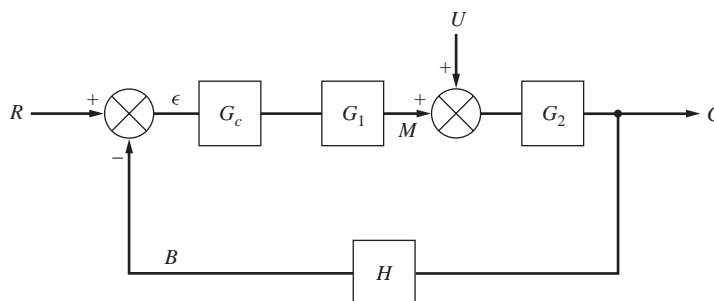H = transfer function of measuring element

FIGURE 11–1
Standard control system nomenclature.

In some cases, the blocks labeled G_c and G_1 will be lumped together into a single block. The series of blocks between the comparator and the controlled variable, which consist of G_c, G_1, and G_2, is referred to as the *forward path.* The block H between the controlled variable and the comparator is called the *feedback path.* The use of G for a transfer function in the forward path and H for one in the feedback path is a common convention.

The product *GH,* which is the product of all transfer functions ($G_cG_1G_2H$) in the loop, is called the *open-loop transfer function.* We call *GH* the open-loop transfer function because it relates the measured variable B to the set point R if the feedback loop (of Fig. 11–1) is disconnected (i.e., opened) from the comparator. The subject of this chapter is the closed-loop transfer function, which relates two variables when the loop of Fig. 11–1 is closed.

In more complex systems, the block diagram may contain several feedback paths and several loads. An example of a multiloop system, which is shown in Fig. 11–2, is cascade control. Several multiloop systems of industrial importance are presented in Chap. 17.
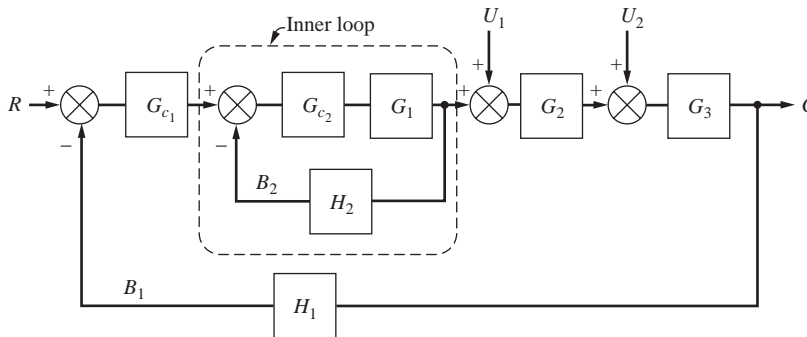

FIGURE 11–2
Block diagram for a multiloop, multiload system.

11.2   OVERALL TRANSFER FUNCTION FOR SINGLE-LOOP SYSTEMS

Once a control system has been described by a block diagram, such as the one shown in Fig. 11–1, the next step is to determine the transfer function relating C to R or C to U. We refer to these transfer functions as *overall* transfer functions because they apply to the entire system. These overall transfer functions are used to obtain considerable information about the control system, as will be demonstrated in the succeeding chapters. For the present it is sufficient to note that they are useful in determining the response of C to any change in R and U.

The response to a change in set point R, obtained by setting $U = 0$, represents the solution to the servo problem. The response to a change in load variable U, obtained by setting $R = 0$, is the solution to the regulator problem. A systematic approach for obtaining the overall transfer function for set point change and load change will now be presented.

Overall Transfer Function for Change in Set Point

For this case, $U = 0$ and Fig. 11–1 may be simplified or reduced as shown in Fig. 11–3. In this reduction, we have made use of a simple rule of block diagram reduction which states that a block diagram consisting of several transfer functions in series can be simplified to a single block containing a transfer function that is the product of the individual transfer functions.
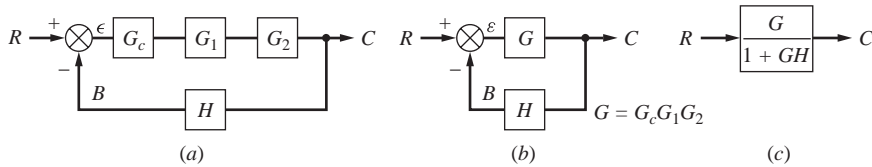

FIGURE 11–3
Block diagram reduction to obtain overall transfer function.

This rule can be proved by considering two noninteracting blocks in series as shown in Fig. 11–4. This block diagram is equivalent to the equations

$$\frac{Y}{X} = G_A \qquad \frac{Z}{Y} = G_B$$

Multiplying these equations gives

$$\frac{Y}{X}\frac{Z}{Y} = G_A G_B$$

which simplifies to

$$\frac{Z}{X} = G_A G_B$$
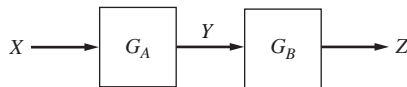
FIGURE 11–4
Two noninteracting blocks in series.

Thus, the intermediate variable Y has been eliminated, and we have shown the overall transfer function Z/X to be the product of the transfer functions $G_A G_B$. This proof for two blocks can be easily extended to any number of blocks to give the rule for the general case. This rule was developed in Chap. 7 for the specific case of several noninteracting, first-order systems in series.

With this simplification the following equations can be written directly from Fig. 11–3b.

$$C = G\varepsilon \tag{11.1}$$

$$B = HC \tag{11.2}$$

$$\varepsilon = R - B \tag{11.3}$$

Since there are four variables and three equations, we can solve the equations simultaneously for C in terms of R as follows:

$$C = G(R - B)$$
$$C = G(R - HC)$$
$$C = GR - GHC$$

or finally

$$\frac{C}{R} = \frac{G}{1 + GH} \tag{11.4}$$

This is the overall transfer function relating C to R and may be represented by an equivalent block diagram as shown in Fig. 11–3c.

Overall Transfer Function for Change in Load

In this case $R = 0$, and Fig. 11–1 is drawn as shown in Fig. 11–5a. From the diagram we can write the following equations:

$$C = G_2(U + M) \tag{11.5}$$

$$M = G_c G_1 \varepsilon \tag{11.6}$$

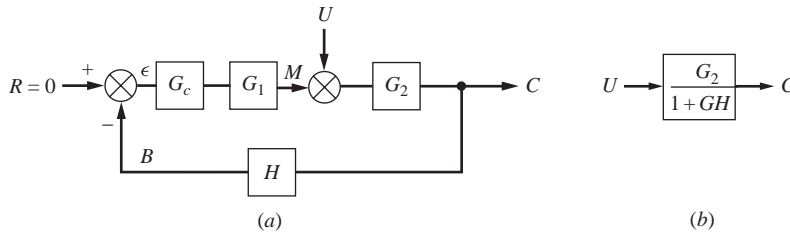$$\varepsilon = -B \tag{11.7}$$

$$B = HC \tag{11.8}$$

FIGURE 11–5
Block diagram for change in load.

Again the number of variables (C, U, M, B, ε) exceeds by 1 the number of equations, and we can solve for C in terms of U as follows:

$$C = G_2(U + G_c G_1 \varepsilon)$$
$$C = G_2 \big[U + G_c G_1(-HC) \big]$$

or finally

$$\frac{C}{U} = \frac{G_2}{1 + GH} \tag{11.9}$$

where $G = G_cG_1G_2$. Notice that the transfer functions for load change or set point change have denominators that are identical, $1 + GH$.

Another approach to finding the closed-loop transfer functions from the block diagram is a "brute-force" technique that involves "breaking the loop" and working your way across the block diagram.
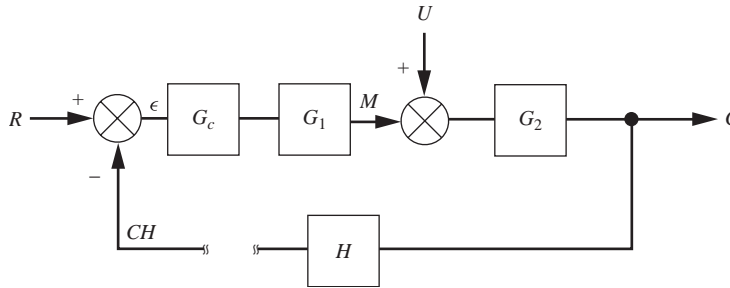

FIGURE 11–6
Modified block diagram for determining closed-loop transfer function.

Starting at the left edge of the modified block diagram in Fig. 11–6, we can work our way across the diagram and develop the following equation.

$$\big[(R - CH)G_cG_1 + U\big]G_2 = C$$

Solving this equation for C, we obtain (after a bit of algebra)

$$C = \frac{G_cG_1G_2}{1 + G_cG_1G_2H}R + \frac{G_2}{1 + G_cG_1G_2H}U = \underbrace{\frac{G}{1 + GH}}_{\substack{\text{closed loop} \\ \text{transfer function} \\ \text{relating } C \text{ and } R}}R + \underbrace{\frac{G_2}{1 + GH}}_{\substack{\text{closed loop} \\ \text{transfer function} \\ \text{relating } C \text{ and } U}}U$$

This result is the same as the individual results for C/R and C/U that we found previously.

The following simple rule serves to generalize the results for the single-loop feedback system shown in Fig. 11–1: The transfer function relating any pair of variables X, Y is obtained by the relationship

$$\frac{Y}{X} = \frac{\pi_{\text{forward}}}{1 + \pi_{\text{loop}}} \quad \text{negative feedback} \tag{11.10}$$

where π_{forward} = product of transfer functions in forward path between locations of X and Y

π_{loop} = product of all transfer functions in loop (i.e., in Fig. 11–1, $\pi_{\text{loop}} = G_cG_1G_2H$)

If this rule is applied to finding C/R in Fig. 11–1, we obtain

$$\frac{C}{R} = \frac{G_c G_1 G_2}{1 + G_c G_1 G_2 H} = \frac{G}{1 + GH}$$

which is the same as before. For positive feedback, the reader should show that the following result is obtained:

$$\frac{Y}{X} = \frac{\pi_{\text{forward}}}{1 - \pi_{\text{loop}}} \quad \text{positive feedback} \tag{11.11}$$

Example 11.1.   Determine the transfer functions C/R, C/U_1, and B/U_2 for the system shown in Fig. 11–7. Also determine an expression for C in terms of R and U_1 for the situation when both set point change and load change occur simultaneously.
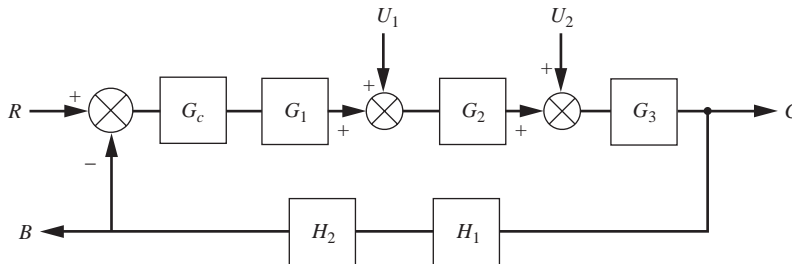
FIGURE 11–7
Block diagram for Example 11.1.

Using the rule given by Eq. (11.10), we obtain by inspection the results

$$\frac{C}{R} = \frac{G_c G_1 G_2 G_3}{1 + G} \tag{11.12}$$

$$\frac{C}{U_1} = \frac{G_2 G_3}{1 + G} \tag{11.13}$$

$$\frac{B}{U_2} = \frac{G_3 H_1 H_2}{1 + G} \tag{11.14}$$

where $G = G_c G_1 G_2 G_3 H_1 H_2$. The reader should check one or more of these results by the direct method of solution of simultaneous equations and the method described in Fig. 11–6.

For separate changes in R and U_1, we may obtain the response C from Eqs. (11.12) and (11.13); thus

$$C = \frac{G_c G_1 G_2 G_3}{1 + G} R \tag{11.15}$$

and

$$C = \frac{G_2 G_3}{1 + G} U_1 \tag{11.16}$$

If both R and U_1 occur simultaneously, the principle of superposition requires that the overall response be the sum of the individual responses; thus

$$C = \frac{G_c G_1 G_2 G_3}{1 + G} R + \frac{G_2 G_3}{1 + G} U_1 \tag{11.17}$$

11.3 OVERALL TRANSFER FUNCTION FOR MULTILOOP CONTROL SYSTEMS

To illustrate how one obtains the overall transfer function for a multiloop system, consider the next example in which the method used is to reduce the block diagram to a single-loop diagram by application of the rules summarized by Eqs. (11.10) and (11.11).

Example 11.2.   Determine the transfer function C/R for the system shown in Fig. 11–8. This block diagram represents a cascade control system, which will be discussed later.
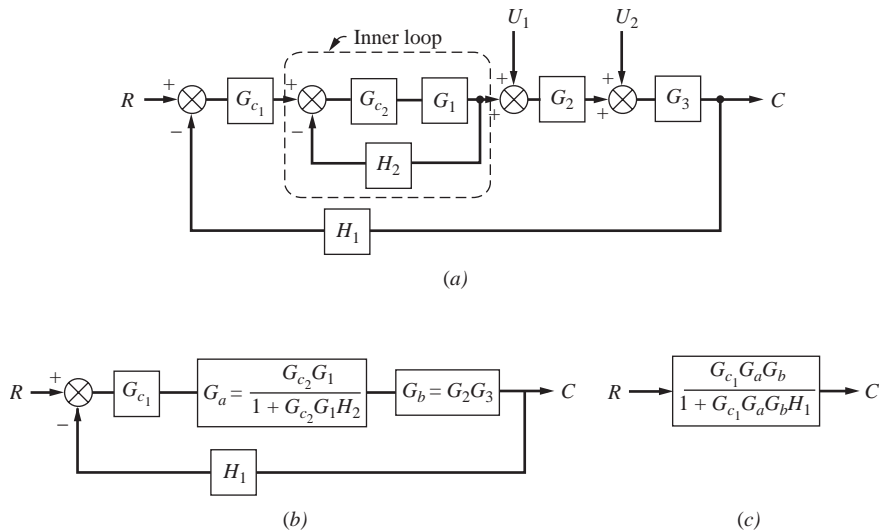

FIGURE 11–8
Block diagram reduction. (*a*) Original diagram; (*b*) first reduction; (*c*) final block diagram.

Obtaining the overall transfer function C/R for the system represented by Fig. 11–8*a* is straightforward if we first reduce the inner loop (or minor loop) involving G_{c_2}, G_1, and H_2 to a single block, as we have just done in the case of Fig. 11–1. For convenience, we may also combine G_2 and G_3 into a single block. These reductions are shown in Fig. 11–8*b*. Figure 11–8*b* is a single-loop block diagram that can be reduced to one block, as shown in Fig. 11–8*c*.

It should be clear without much detail that to find any other transfer function such as C/U_1 in Fig. 11–8*a,* we proceed in the same manner, i.e., first reduce the inner loop to a single-block equivalent.

SUMMARY

In this chapter, we have illustrated the procedure for reducing the block diagram of a control system to a single block that relates one input to one output variable. This procedure consists of writing, directly from the block diagram, a sufficient number of linear algebraic equations and solving them simultaneously for the transfer function of the desired pair of variables. For single-loop control systems, a simple rule was developed for finding the transfer function between any desired pair of input-output variables. This rule is also useful in reducing a multiloop system to a single-loop system.

It should be emphasized that regardless of the pair of variables selected, the denominator of the closed-loop transfer function will always contain the same term, $1 + G$, where G is the open-loop transfer function of the single-loop control system. In the succeeding chapters, frequent use will be made of the material in this chapter to determine the overall response of control systems.

PROBLEMS

11.1. Determine the transfer function $Y(s)/X(s)$ for the block diagrams shown in Fig. P11–1. Express the results in terms of G_a, G_b, and G_c.

(a)

(b)
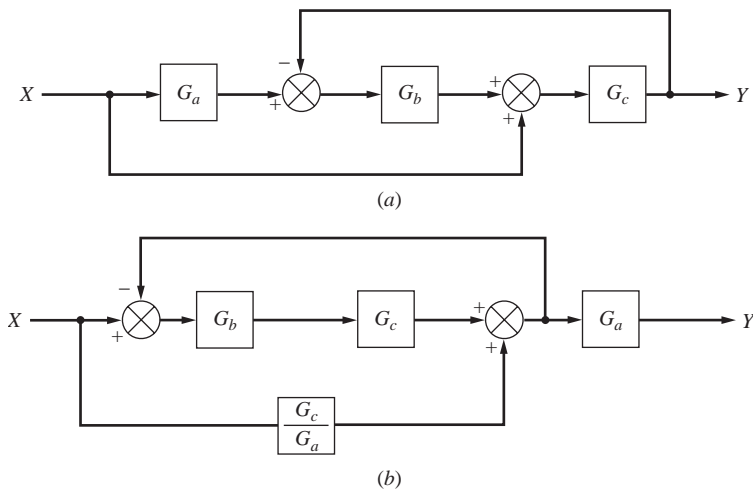
FIGURE P11–1

11.2. Find the transfer function $Y(s)/X(s)$ of the system shown in Fig. P11–2.
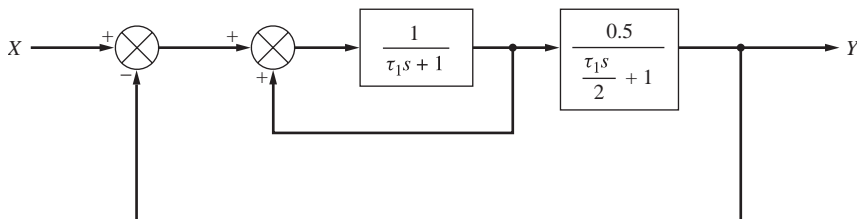
FIGURE P11–2

11.3.   For the control system shown in Fig. P11–3 determine the transfer function $C(s)/R(s)$.
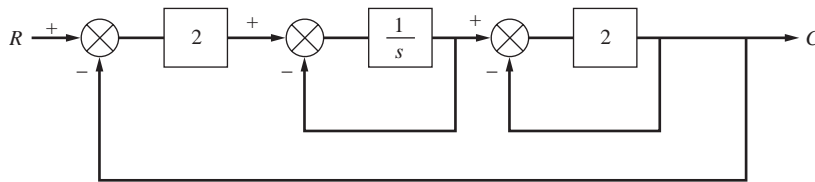
FIGURE P11–3

11.4.   Derive the transfer function Y/X for the control system shown in Fig. P11–4.
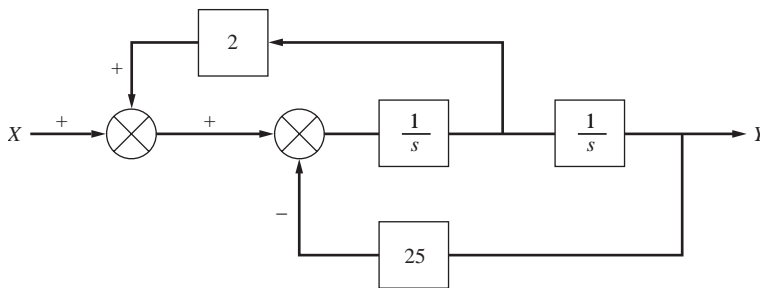
FIGURE P11–4

11.5.   Derive the transfer function T'/T'_R for the temperature control system shown in Fig. 8–16.

11.6.   Derive the transfer functions C_2/C_0 and C_2/C_R for the reactor control system shown in Fig. P10–3.

CHAPTER
11

CAPSULE SUMMARY

R = set point or desired value
C = controlled variable
ε = error
B = variable produced by measuring element
M = manipulated variable
U = load variable or disturbance
G_c = transfer function of controller
G_1 = transfer function of final control element
G_2 = transfer function of process
H = transfer function of measuring element
G = $G_c G_1 G_2$ (the product of the transfer functions in the forward path between R and C)
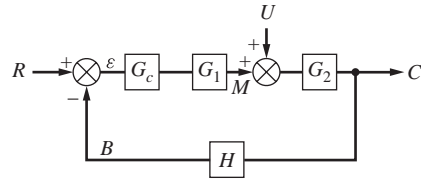

FIGURE 11.1
Standard control system nomenclature.

Closed-loop process   Process in which the feedback loop is connected to the comparator.
Closed-loop transfer function   Transfer functions relating two variables in the process when the feedback loop is connected to the comparator.
Feedback path   The path that connects the controlled variable and the comparator.
Forward path   The transfer functions that lie between two signals in the block diagram moving left to right as drawn in the block diagram above. The complete forward path consists of $G = G_c G_1 G_2$. The forward path between U and C is G_2 only.
Open-loop process   Process in which the feedback loop is disconnected from the comparator.
Open-loop transfer function   Product of all transfer functions in the loop relating B and R when the feedback loop is disconnected from the comparator.

Closed-loop formula
(negative feedback)
$$\frac{Y}{X} = \frac{\pi_{\text{forward}}}{1 + \pi_{\text{loop}}}$$

Closed-loop formula
(positive feedback)
$$\frac{Y}{X} = \frac{\pi_{\text{forward}}}{1 - \pi_{\text{loop}}}$$

π_{forward} = product of transfer functions in forward path between locations of X and Y
π_{loop} = product of all transfer functions in loop (i.e., from figure above, $\pi_{\text{loop}} = G_c G_1 G_2 H$)

CHAPTER
12

TRANSIENT RESPONSE OF SIMPLE CONTROL SYSTEMS

In this chapter the results of all the previous chapters will be applied to determining the transient response of a simple control system to changes in set point and load. Considerable use will be made of the results of Chaps. 4 through 7 (Part 2) because the overall transfer functions for the examples presented here reduce to first- and second-order systems.

Consider the control system for the heated, stirred tank that we previously discussed and is represented by Fig. 12–1. The reader may want to refer to Chap. 8 for a description of this control system.
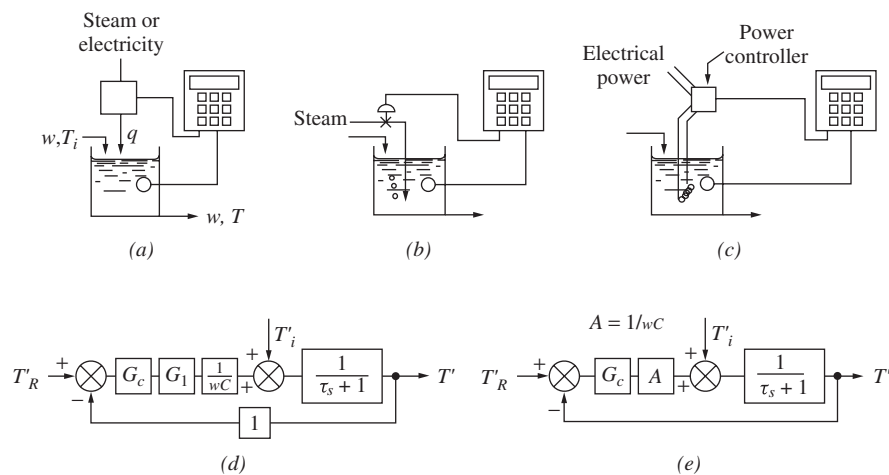

FIGURE 12–1
Block diagram of temperature control system.

228

In Fig. 12–1a, the sketch of the apparatus is drawn in such a way that the source of heat (electricity or steam) is not specified. To make this problem more realistic, we have shown in Fig. 12–1b that the source of heat is steam that is discharged directly into the water, and in Fig. 12–1c the source of heat is electrical. In the latter drawing, a device known as a power controller provides electric power to a resistance heater proportional to the signal from the controller.

The block diagram is shown in Fig. 12–1d. The block representing the process is taken directly from Fig. 8–3. To reduce the number of symbols, $1/wC$ has been replaced by A in Fig. 12–1e.

Throughout this chapter, we will assume that the valve does not have any dynamic lag, for which case the transfer function of the valve (G_1 in Fig. 12–1) will be taken as a constant K_v. This assumption was shown to be reasonable in Chap. 9. To simplify the discussion further, K_v has been taken as 1. (If K_v were other than 1, we may simply replace G_c by G_cK_v in the ensuing discussion.)

In the first part of the chapter, we will also assume that there is no dynamic lag in the measuring element ($\tau_m = 0$), so that it may be represented by a transfer function that is simply the constant 1. A bare thermocouple will have a response that is so fast that for all practical purposes it can be assumed to follow the slowly changing tank temperature without lag. When the feedback transfer function is unity, the system is called a *unity-feedback* system.

Introducing these assumptions leads to the simplified block diagram of Fig. 12–1e, for which we will obtain overall transfer functions for changes in set point and load when proportional control and proportional-integral control are used.

12.1   PROPORTIONAL CONTROL FOR SET POINT CHANGE (SERVO PROBLEM— SET POINT TRACKING)

The goal of the control system for this case is to force the system to "track" the desired set point changes. For proportional control, $G_c = K_c$. By using the methods developed in Chap. 11, the overall transfer function in Fig. 12–1 e is

$$\frac{T'}{T'_R} = \frac{K_cA/(\tau s + 1)}{1 + K_cA/(\tau s + 1)} = \frac{K_cA}{\tau s + (1 + K_cA)} \tag{12.1}$$

This may be rearranged in the form of a first-order lag to give

$$\frac{T'}{T'_R} = \frac{K_cA/(1 + K_cA)}{[\tau/(1 + K_cA)]s + 1} = \frac{A_1}{\tau_1 s + 1} \tag{12.2}$$

where $\tau_1 = \dfrac{\tau}{1 + K_cA}$

$$A_1 = \frac{K_cA}{1 + K_cA} = \frac{1}{1 + (1/K_cA)}$$

According to this result, the response of the tank temperature to a change in set point is first-order. The time constant for the closed-loop control system τ_1 is less than that of the stirred tank itself τ. This means that one of the effects of feedback control is to speed up the response. We may use the results of Chap. 4 to find the response to a variety of inputs.
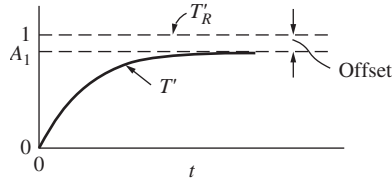

FIGURE 12–2

Unit-step response for set point change (P control).

The response of the system to a unit-step change in set point T_R' is shown in Fig. 12–2. (We have selected a unit change in set point for convenience; responses to steps of other magnitudes are obtained by superposition.) Remember, the goal of a control system is to force the system to track the set point. Thus, the desired ultimate value of T', which is $T'(\infty)$, is of course 1. For the case of a unit-step change in set point, T' approaches $A_1 = K_cA/(1 + K_cA)$, a fraction of unity. The desired change is, of course, 1. Thus, the ultimate value of the temperature $T'(\infty)$ does not match the desired change. This discrepancy is called *offset* and is defined as

$$\text{Offset} = T_R'(\infty) - T'(\infty) \tag{12.3}$$

The offset is actually the steady-state value of the error (for the case of unity feedback). In terms of the particular control system parameters

$$\text{Offset} = 1 - \frac{K_cA}{1 + K_cA} = \frac{1}{1 + K_cA} \tag{12.4}$$

This difference or discrepancy between set point and tank temperature at steady state is characteristic of proportional control. In some cases offset cannot be tolerated. However, notice from Eq. (12.4) that the offset decreases as K_c increases, and in theory the offset could be made as small as desired by increasing K_c to a sufficiently large value.

Proportional control :   As K_c ↑   Offset ↓

To give a full answer to the problem of eliminating offset by high controller gain requires a discussion of stability and the response of the system when other lags, which have been neglected, are included in the system. Both these subjects are to be covered later. For the present we simply say that whether or not proportional control is satisfactory depends on the amount of offset that can be tolerated, the speed of response of the system, and the amount of gain that can be provided by the controller without causing the system to go unstable.
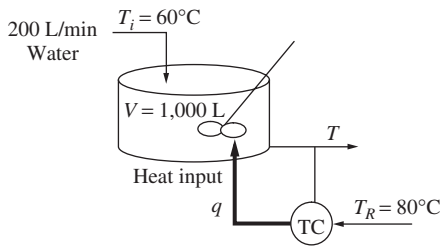
FIGURE 12–3
Stirred-tank heater control system.

Example 12.1.   Proportional control of a stirred tank heater for set point tracking. Consider the stirred-tank heater that we examined in Chap. 8 (Examples 8.1 to 8.3), shown in Fig. 12–3.

Determine the response of a closed-loop proportional control system for K_c = 5, 10, 20, and 100 for a set point change of 5°C. Determine the offset for each value of K_c.

As before, we define

$$\left.\begin{array}{l} T_i' \;=\; T_i - 60 \\[4pt] T_R' \;=\; T_R - 80 \\[4pt] q' \;=\; q - 280 \text{ kW} \end{array}\right\} \text{deviation variables}$$

$$\tau \;=\; \frac{1{,}000 \text{ L}}{200 \text{ L/min}} \;=\; 5 \text{ min} \qquad A \;=\; \frac{1}{wC} \;=\; \frac{1°C}{14 \text{ kW}} \;=\; \frac{0.0714°C}{\text{kW}}$$

If we assume that the thermocouple that senses the tank exiting temperature has fast dynamics compared to the tank, a Simulink block diagram for the process is shown in Fig. 12–4.

FIGURE 12–4
Simulink model for Example 12.1.

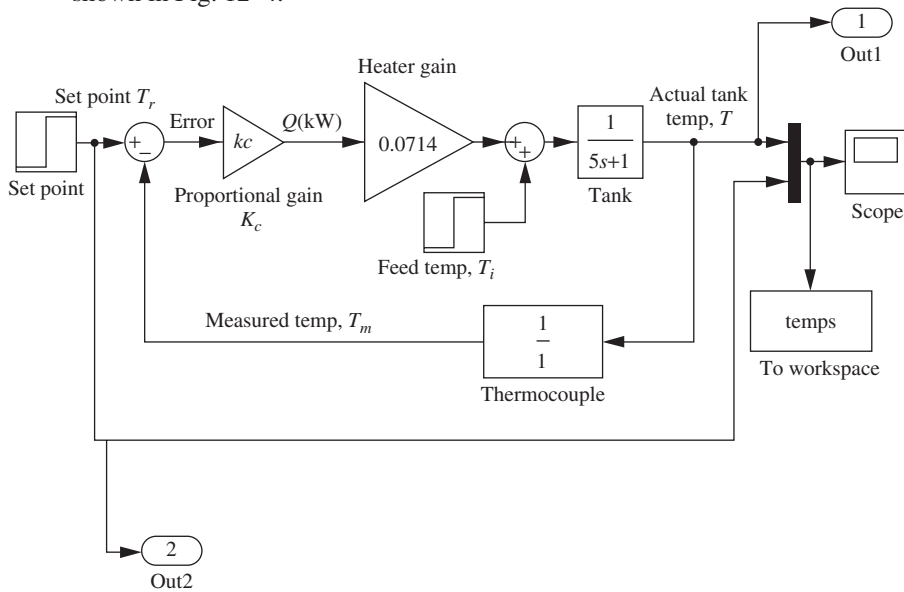We can program a MATLAB m-file to run this Simulink model (named: 'example12_1') for each value of K_c and then to plot the results on the same graph. The m-file that calls the model is shown below:

```
% M-file that calls Simulink Model 'Example12_1'
for i=1:4
  z=[5,10,20,100];
  kc=z(1,i);
% the variable y in the 'sim' statement is taken from the
outport in the model
  [t,x,y]=sim('example12_1',15);
  plot(t,y(:,1)
  hold on
% the last row in y contains the final values of the
setpoint and the tank temp
  [norow,nocol]=size(y);
  offset(i,1)=y(norow,2)-y(norow,1);
  kcplot(i,1)=kc;
end
grid
title('Temp vs time for Kc=5,10,20,100');
hold off
figure;
plot(kcplot,offset);
title('Offset vs Kc');
```

The results are shown in Figs. 12–5 and 12–6.

A hand calculation verifies the offset results obtained by MATLAB for the specified values of K_c.

K_c	**Offset** $= \dfrac{5}{1 + K_c A} = \dfrac{5}{1 + 0.0714 K_c}$
5	3.69
10	2.92
20	2.06
100	0.614

Note that as the proportional gain K_c increases, the process more closely approaches the desired set point (5°C in deviation variables, actually 85°C). For a gain of 100, the outlet temperature is 4.386°C (deviation variable; the actual outlet temperature is 84.386°C) with an offset of 0.614°C.

Deviation in Tank Temperature vs Time for $K_c = 5,10,20,100$
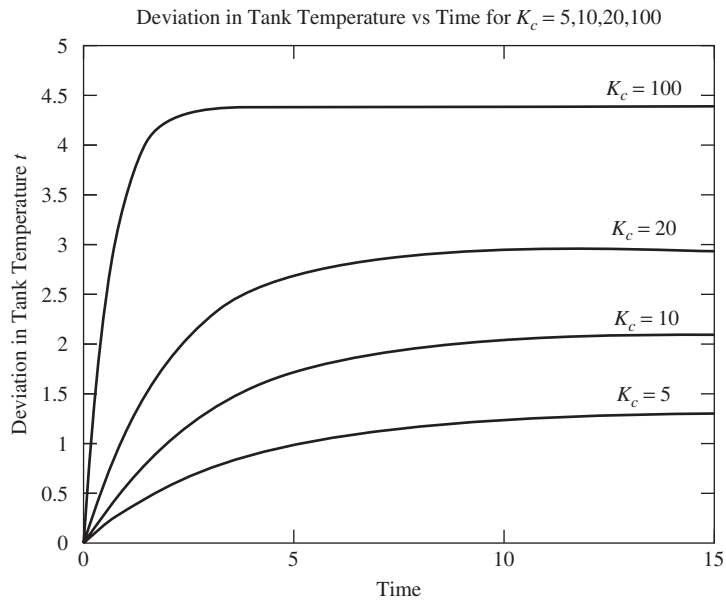
FIGURE 12–5
Response of stirred-tank temperature for proportional control—set point tracking.

Offset vs K_c for Set Point Tracking
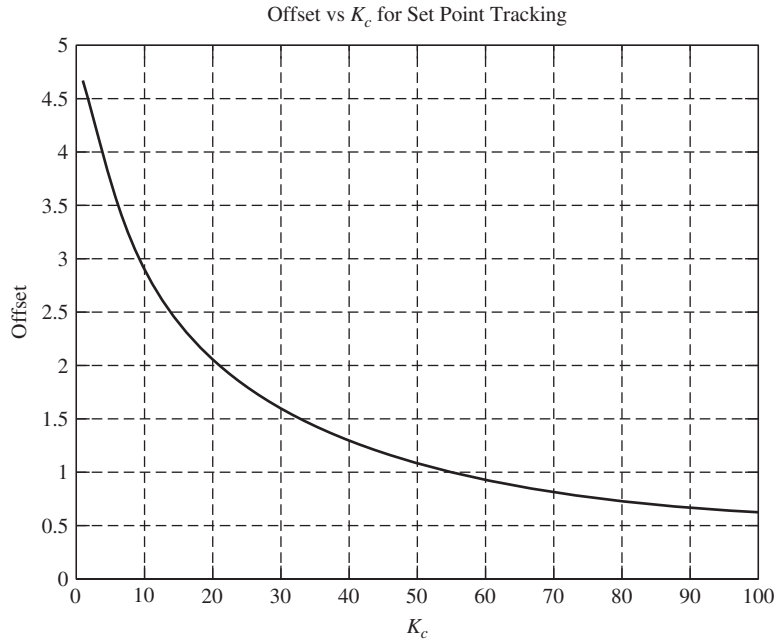
FIGURE 12–6
Effect of controller gain on offset for stirred-tank proportional control—set point tracking.

12.2 PROPORTIONAL CONTROL FOR LOAD CHANGE (REGULATOR PROBLEM— DISTURBANCE REJECTION)

The same control system shown in Fig. 12–1e is to be considered. This time the set point remains fixed; that is, $T'_R = 0$. We are interested in the response of the system to a change in the inlet stream temperature, i.e., to a load change. Remember that the goal of the control system in this case is to reject the effect of disturbances (changes in the inlet temperature T'_i for this process) and maintain the controlled variable, the tank temperature, at the set point. Since the set point has not changed from its steady-state value for this case ($T'_R = 0$), we want $T'(\infty) = 0$.

By using the methods of Chap. 11, the overall transfer function becomes

$$\frac{T'}{T'_i} = \frac{1/(\tau s + 1)}{1 + K_c A/(\tau s + 1)} = \frac{1}{\tau s + 1 + K_c A} \tag{12.5}$$

This may be arranged in the form of the first-order lag; thus

$$\frac{T'}{T'_i} = \frac{A_2}{\tau_1 s + 1} \tag{12.6}$$

where $A_2 = \dfrac{1}{1 + K_c A}$

$\tau_1 = \dfrac{\tau}{1 + K_c A}$

As for the case of set point change, we have an overall response that is first-order. The overall time constant τ_1 is the same as for set point changes. For a step change in inlet temperature, let us examine the response of the controlled process and how effectively the disturbance is rejected.

Thus for $T_i = 1/s$,

$$T' = \frac{1}{s}\frac{1/(1 + K_c A)}{\left[\tau/(1 + K_c A)\right]s + 1}$$

and the ultimate (steady-state) value of T' is

$$T'(\infty)\lim_{s \to 0}\ sT'(s) = \frac{1}{1 + K_c A}$$
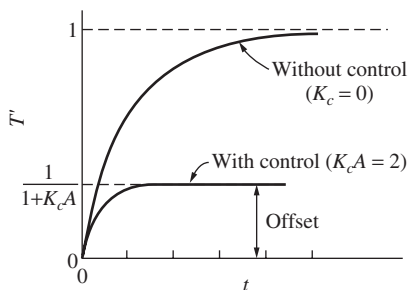
FIGURE 12–7
Unit-step response for load change (P control).

The response of the system to a unit-step change in inlet temperature T'_i is shown in Fig. 12–7.

It may be seen that T' approaches $1/(1 + K_c A)$. To demonstrate the benefit of control, we have shown the response of the tank temperature (open-loop response) to a unit-step change in inlet temperature if no control were present, that is, $K_c = 0$. In this case, the

major advantage of control lies in the reduction of offset. From Eq. (12.3), the offset becomes

$$\text{Offset} = T_R'(\infty) - T'(\infty) = 0 - \frac{1}{1 + K_c A}$$

$$\text{Offset} = -\frac{1}{1 + K_c A} \tag{12.7}$$

As for the case of a step change in set point, the absolute value of the offset is reduced as controller gain K_c is increased.

> **Example 12.2.   Proportional control of a stirred-tank heater for disturbance rejection.** Determine the response of a closed-loop proportional control system in Example 12.1 for $K_c = 5$, 10, 20, and 100 for a disturbance of 5°C (i.e., the feed temperature increases by 5°C from 60 to 65°C). Determine the offset for each value of K_c.
>
> We can use the same Simulink model for this analysis. We just have the step change enter the process as the disturbance and leave the set point set to zero.
>
> The results are as follows. Figure 12–8 shows the tank temperature (plotted as a deviation variable) versus time, and Fig. 12–9 shows the offset plotted as a function of K_c.
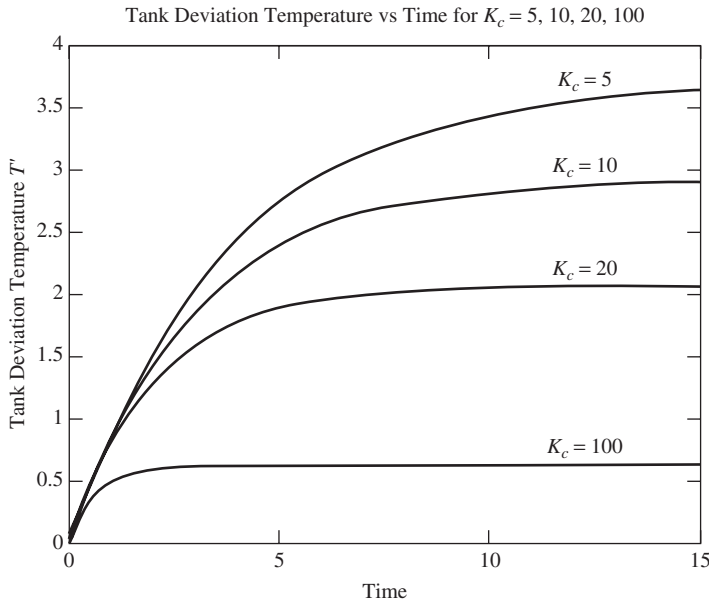

FIGURE 12–8
Response of stirred-tank temperature for proportional control—disturbance rejection.

Offset vs K_c for Disturbance Rejection
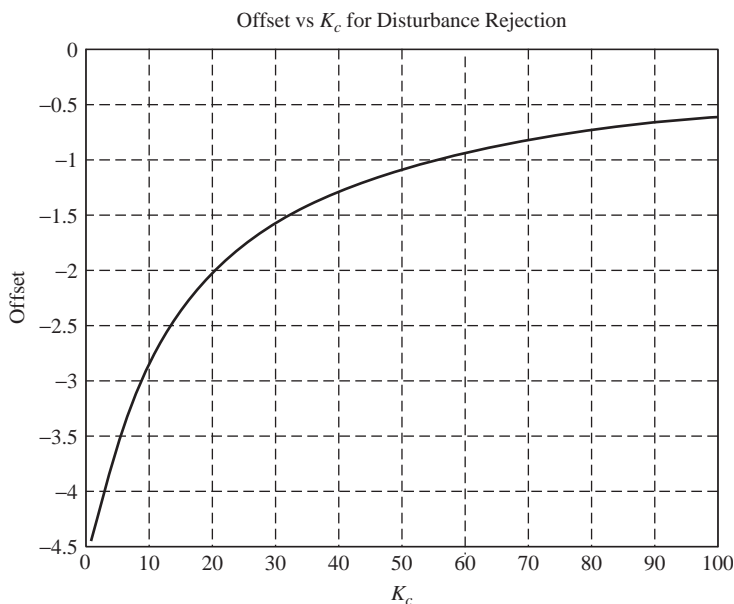
FIGURE 12–9

Effect of controller gain on offset for stirred-tank proportional control—disturbance rejection.

Note the results are qualitatively the same as for set point tracking. As K_c increases for proportional control, the system does a better job at rejecting the disturbance and bringing the tank temperature back to the original set point (0°C in deviation variables, or 80°C). The final temperature is closer to zero, and the offset also approaches zero.

12.3 PROPORTIONAL-INTEGRAL CONTROL FOR LOAD CHANGE

Again, for the case of a load change, the desired goal is the rejection of the disturbance, and the desired final value is zero (since the set point remains unchanged). For PI control, we replace G_c in Fig. 12–1e by $K_c (1 + 1/\tau_I s)$. The overall transfer function for load change is therefore

$$\frac{T'}{T_i'} = \frac{1/(\tau s + 1)}{1 + \dfrac{K_c (1 + 1/\tau_I s) A}{\tau s + 1}} \tag{12.8}$$

Rearranging this gives

$$\frac{T'}{T_i'} = \frac{\tau_I s}{(\tau s + 1)(\tau_I s) + K_c A(\tau_I s + 1)}$$

or

$$\frac{T'}{T_i'} = \frac{\tau_I s}{\tau \tau_I s^2 + (K_c A \tau_I + \tau_I)s + K_c A}$$

Since the denominator contains a quadratic expression, the transfer function may be written in the standard form of a second-order system to give

$$\frac{T'}{T'_i} = \frac{(\tau_I/K_cA)s}{(\tau\tau_I/K_cA)s^2 + \tau_I(1 + 1/K_cA)s + 1}$$

or

$$\frac{T'}{T'_i} = \frac{A_1s}{\tau_1^2 s^2 + 2\zeta\tau_1 s + 1} \tag{12.9}$$

where

$$A_1 = \frac{\tau_I}{K_cA}$$

$$\tau_1 = \sqrt{\frac{\tau\tau_I}{K_cA}}$$

$$\zeta = \frac{1}{2}\sqrt{\frac{\tau_I}{\tau}}\frac{1 + K_cA}{\sqrt{K_cA}}$$

For a unit-step change in load, $T'_i = 1/s$. Combining this with Eq. (12.9) gives

$$T' = \frac{A_1}{\tau_1^2 s^2 + 2\zeta\tau_1 s + 1} \tag{12.10}$$

Equation (12.10) shows that the response of the tank temperature to a step change in T'_i is equivalent to the response of a second-order system to an impulse function of magnitude A_1. Since we have studied the impulse response of a second-order system in Chap. 7, the solution to the present problem is already known. This justifies in part some of our previous work on transients. By using Eq. (7.32), the impulse response for this system may be written for $\zeta < 1$ as

$$T' = A_1\left(\frac{1}{\tau_1}\frac{1}{\sqrt{1 - \zeta^2}}e^{-\zeta t/\tau_1}\sin\sqrt{1 - \zeta^2}\,\frac{t}{\tau_1}\right) \tag{12.11}$$

Although the response of the system can be determined from Eq. (12.11) or Fig. 7–8, the effect of varying K_c and τ_I on the system response can be seen more clearly by plotting response curves, such as those shown in Fig. 12–10. From Fig. 12–10a, we see that an increase in K_c, for a fixed value of τ_I, improves the response by decreasing the maximum deviation and by making the response less oscillatory. The formula for ζ in Eq. (12.9) shows that ζ increases with K_c, which indicates that the response is less oscillatory. Figure 12–10b shows that, for a fixed value of K_c, a decrease in τ_I decreases
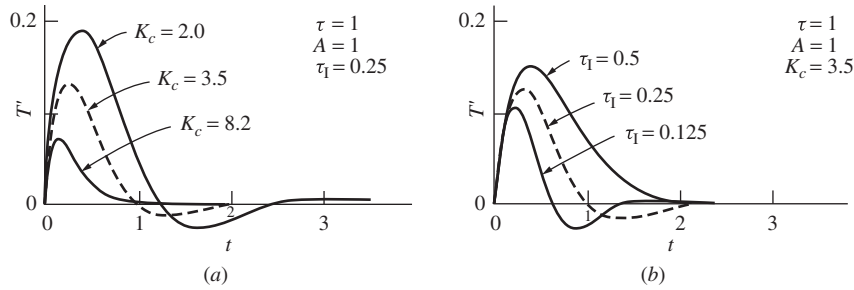
FIGURE 12–10
Unit-step response for load change—disturbance rejection (PI control).

the maximum deviation and period. However, a decrease in τ_I causes the response to become more oscillatory, which means that ζ decreases. This effect of τ_I on the oscillatory nature of the response is also given by the formula for ζ in Eq. (12.9).

Summary for Proportional-Integral Control—Response to Step Change in Load		
For fixed value of τ_I: As $K_c \uparrow$, $\zeta \uparrow$	Max. deviation $\downarrow$	Oscillations $\downarrow$
For fixed value of K_c: As $\tau_I \downarrow$, $\zeta \downarrow$	Max. deviation $\downarrow$	Oscillations $\uparrow$

For this case, the offset as defined by Eq. (12.3) is zero; thus

$$\text{Offset} = T'_R(\infty) - T'(\infty)$$
$$= 0 - 0 = 0$$

One of the most important advantages of PI control is the elimination of offset.

<div style="border:1px solid">

Key Point   **Proportional-Integral Control** $\rightarrow$ **No offset**

</div>

Example 12.3.   Proportional-integral control of a stirred tank heater for disturbance rejection.

(a) Determine the response of a closed-loop proportional control system in Example 12.1 for K_c = 5, 10, 20, and 100 for a disturbance of 5°C (i.e., the feed temperature increases by 5°C from 60 to 65°C). Use a value of 2 min for τ_I.

(b) Repeat part (a), using a constant value of 20 for K_c and values of 1, 2, 5, and 10 min for τ_I.

We can use nearly the same Simulink model for this analysis. We just change the gain block that we used for the proportional controller to a PID block (see Fig. 12–11). Within the PID block, the controller algorithm is given as $P + I/s + Ds$. Comparing this with our standard PID equation, we conclude that $P = K_c$, $I = K_c/\tau_I$, and $D = K_c\tau_D$. So that we can vary the model parameters with a MATLAB m-file, we insert the variable *kc* for P and *kc/taui* for I.
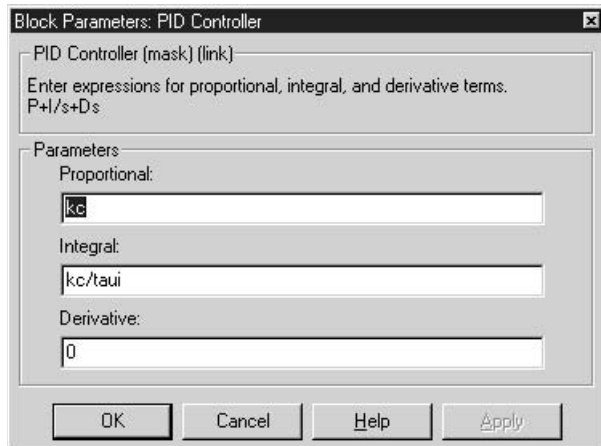
FIGURE 12–11
Simulink PID block parameters for Example 12.3.

For a disturbance change only, the step change enters the process as the distur-
bance, and we leave the set point input set to zero. The m-file that we use to drive
the model is shown below.

```
clear
% the variable h can be used to plot each line in a
  different color
h(1,:)='g- ';
h(2,:)='r-.';
h(3,:)='b: ';
h(4,:)='k--';
h(5,:)='k- ';
taui=2
for i=1:4
  z=[5,10,20,100];
  kc=z(1,i);
% the variable y in the 'sim' statement is taken from the
  outport in the model
 [t,x,y]=sim('example12_3',50);
  plot(t,y(:,1),h(i,:))
     hold on
end
grid
hold off
```

The result for a constant τ_I of 2 min is shown in Fig. 12–12.

FIGURE 12–12
Stirred-tank disturbance rejection using PI control, for $\tau_I = 2$ min.

The result for a constant value of $K_c = 20$ is shown in Fig. 12–13.

FIGURE 12–13
Stirred-tank disturbance rejection using PI control, for $K_c = 20$.

12.4   PROPORTIONAL-INTEGRAL CONTROL FOR SET POINT CHANGE

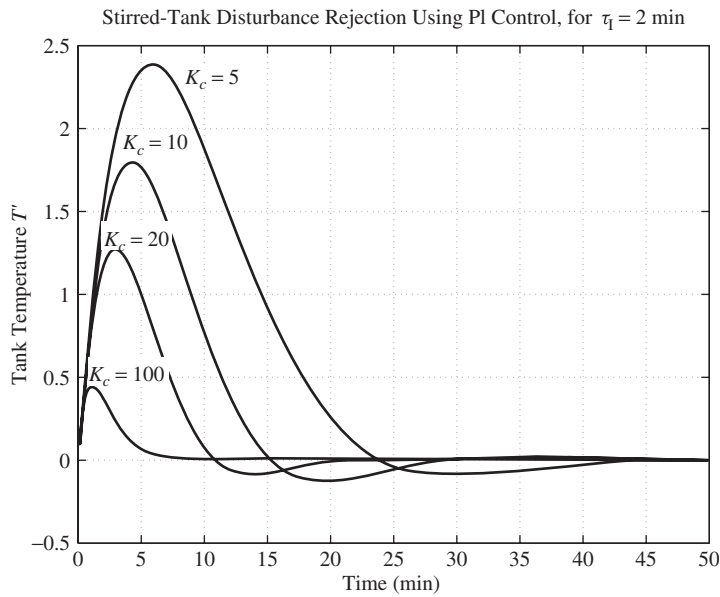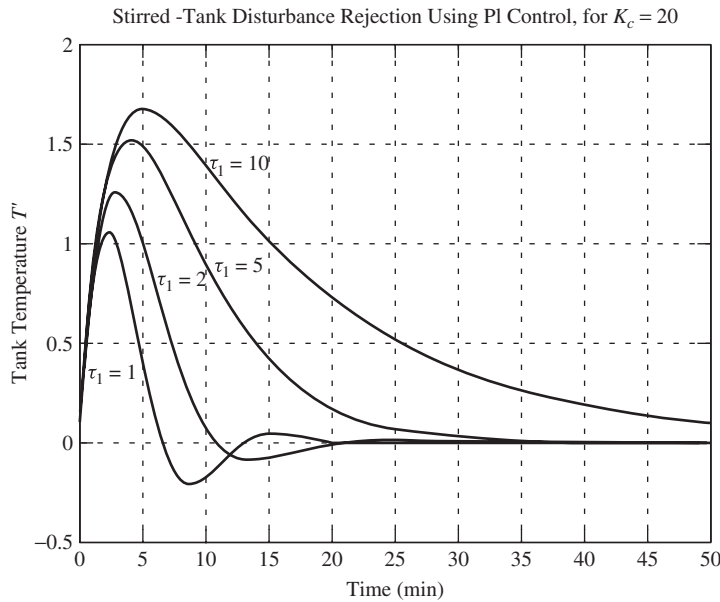Again, the controller transfer function is $K_c(1 + 1/\tau_I s)$, and we obtain from Fig. 12–1e the transfer function

$$\frac{T'}{T'_R} = \frac{AK_c(1 + 1/\tau_I s)[1/(\tau s + 1)]}{1 + AK_c(1 + 1/\tau_I s)[1/(\tau s + 1)]} \tag{12.12}$$

This equation may be reduced to the standard quadratic form to give

$$\frac{T'}{T'_R} = \frac{\tau_I s + 1}{\tau_1^2 s^2 + 2\zeta\tau_1 s + 1} \tag{12.13}$$

where τ_1 and ζ are the same functions of the parameters as in Eq. (12.9). Introducing a unit-step change ($T'_R = 1/s$) into Eq. (12.13) gives

$$T' = \frac{1}{s}\frac{\tau_I s + 1}{\tau_1^2 s^2 + 2\zeta\tau_1 s + 1} \tag{12.14}$$

To obtain the response of T' in the time domain, Eq. (12.14) is expanded into two terms:

$$T' = \frac{\tau_I}{\tau_1^2 s^2 + 2\zeta\tau_1 s + 1} + \frac{1}{s}\frac{1}{\tau_1^2 s^2 + 2\zeta\tau_1 s + 1} \tag{12.15}$$

The first term on the right is equivalent to the response of a second-order system to an impulse function of magnitude τ_I. The second term is the unit-step response of a second-order system. It is convenient to use Figs. 8–3 and 8–6 to obtain the response for Eq. (12.15). For $\zeta < 1$, an analytic expression for T' is

$$T' = \left(\frac{1}{\tau_1\sqrt{1 - \zeta^2}}e^{-\zeta t/\tau_1}\ \sin\sqrt{1 - \zeta^2}\ \frac{t}{\tau_1}\right) + 1$$
$$\hspace{3cm} \tag{12.16}$$
$$- \left[\frac{1}{\sqrt{1 - \zeta^2}}e^{-\zeta t/\tau_1}\ \sin\left(\sqrt{1 - \zeta^2}\ \frac{t}{\tau_1} + \tan^{-1}\frac{\sqrt{1 - \zeta^2}}{\zeta}\right)\right]$$

The last expression was obtained by combining Eqs. (7.18) and (7.32). A typical response for T' is shown in Fig. 12–14. The offset as defined by Eq. (12.3) is zero; thus

$$\text{Offset} = T'_R(\infty) - T'(\infty)$$
$$= 1 - 1 = 0$$

Again notice that the integral action in the controller has *eliminated the offset.*
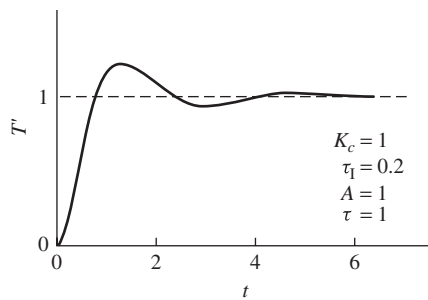
FIGURE 12–14
Unit-step response for set point change (PI control).

Example 12.4.   Proportional-integral control of a stirred-tank heater for set point tracking

(a) Determine the response of a closed-loop proportional control system in Example 12.1 for K_c = 5, 10, 20, and 100 for a set point change of 5°C. Use a value of 2 min for τ_I.

(b) Repeat part (a) using a constant value of 20 for K_c and values of 1, 2, 5, and 10 min for τ_I.

By proceeding as before using the Simulink model driven with a MATLAB m-file, these results are obtained, shown in Figs. 12–15 and 12–16.
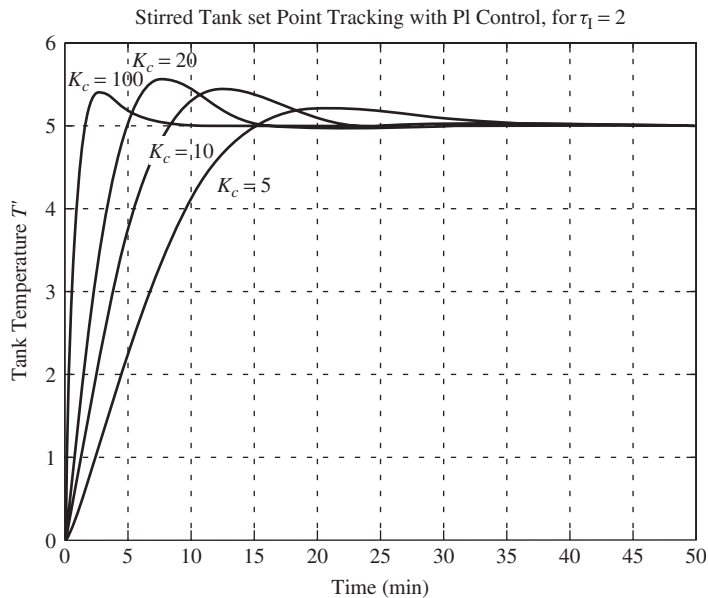
FIGURE 12–15
Stirred-tank set point tracking with PI control (τ_I = 2 min).

Stirred Tank Set Point Tracking with Pl Control for $K_c = 20$
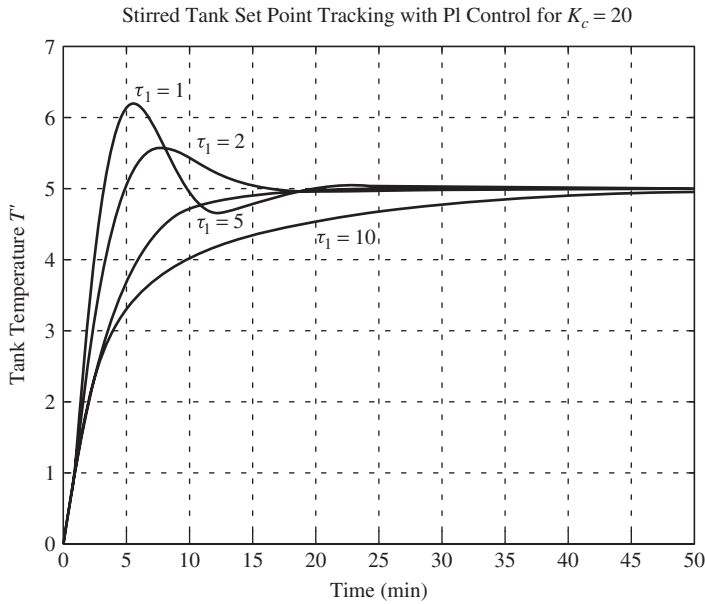
FIGURE 12–16
Stirred-tank set point tracking with PI control ($K_c = 20$).

Note that there is no offset with PI control. Eventually each system levels out at the desired set point of 5°C.

12.5 PROPORTIONAL CONTROL OF SYSTEM WITH MEASUREMENT LAG

In the previous examples the lag in the measuring element was assumed to be negligible (the feedback transfer function was taken as 1). We now consider the same control system, the stirred-tank heater of Fig. 12–1, with a first-order measuring element having a transfer function $1/(\tau_m s + 1)$.

The block diagram for the modified system is now shown in Fig. 12–17.
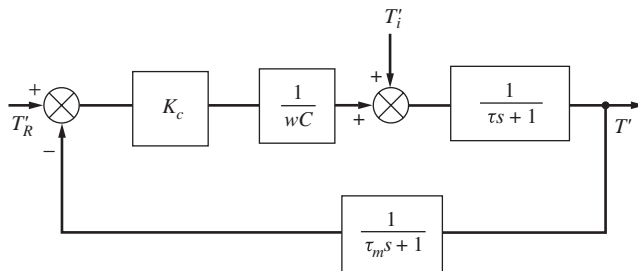
FIGURE 12–17
Control system with measurement lag.

By the usual procedure, the transfer function for set point changes may be written

$$\frac{T'}{T_R'} = \frac{A_1 (\tau_m s + 1)}{\tau_2^2 s^2 + 2\zeta_2 \tau_2 s + 1} \tag{12.17}$$

where

$$A_1 = \frac{K_c A}{1 + K_c A}$$

$$\tau_2 = \sqrt{\frac{\tau \tau_m}{1 + K_c A}}$$

$$\zeta_2 = \frac{\tau + \tau_m}{2\sqrt{\tau \tau_m}} \frac{1}{\sqrt{1 + K_c A}}$$

We will not obtain an expression for the transient response for this case, for it will be of the same form as Eq. (12.16). Adding the first-order measuring lag to the control system of Fig. 12–1 produces a second-order system even for proportional control. This means there will be an oscillatory response for an appropriate choice of the parameters τ, τ_m, K_c, and A. To understand the effect of gain K_c and measuring lag τ_m on the behavior of the system, response curves are shown in Fig. 12–18 for various combinations of K_c and τ_m for a fixed value of $\tau = 1$. In general, the response becomes more oscillatory, or less stable, as K_c or τ_m increases.

FIGURE 12–18
Effect of controller gain and measurement lag on system response for unit-step change in set point.

For a fixed value of $\tau_m = 1$, Fig. 12–18a shows that the offset is reduced as K_c increases; however, this improvement in steady-state performance is obtained at the

expense of a poorer transient response. As K_c increases, the overshoot becomes excessive, and the response becomes more oscillatory. In general, we will find that a control 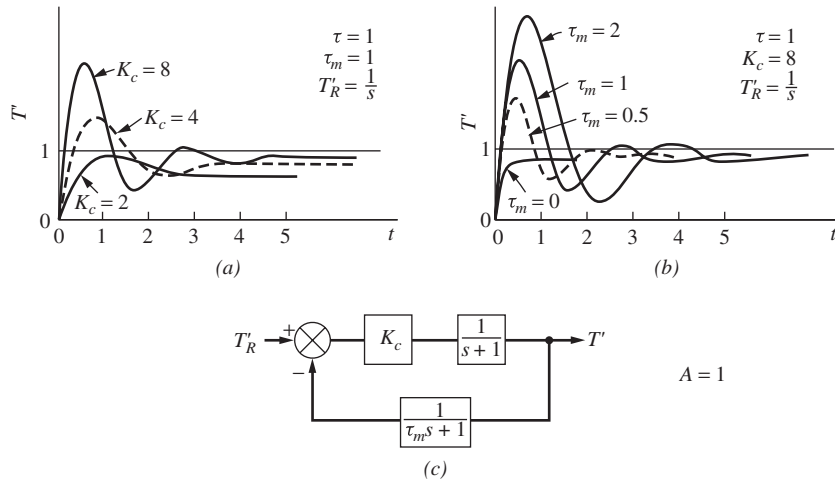system having proportional control will require a value of K_c that is based on a compromise between low offset and satisfactory transient response.

For a fixed value of controller gain ($K_c = 8$), Fig. 12–18b shows that an increase in measurement lag produces a poorer transient response in that the overshoot becomes greater and the response more oscillatory as τ_m increases. This behavior illustrates a general rule that the measuring element in a control system should respond quickly if satisfactory response is to be achieved.

Example 12.5   Effect of measurement lag on the proportional-integral control of a stirred-tank heater for set point tracking. Let's examine the effect of measurement lag on the PI control of the stirred-tank heater. We just concluded that as measurement lag increases, a proportional control only system becomes more oscillatory. What happens with PI control? The Simulink model is modified to include measurement lag (see Fig. 12–19). We'll fix K_c at 20 and $\tau_I = 2$, which gave us reasonably good results in Example 12.4.
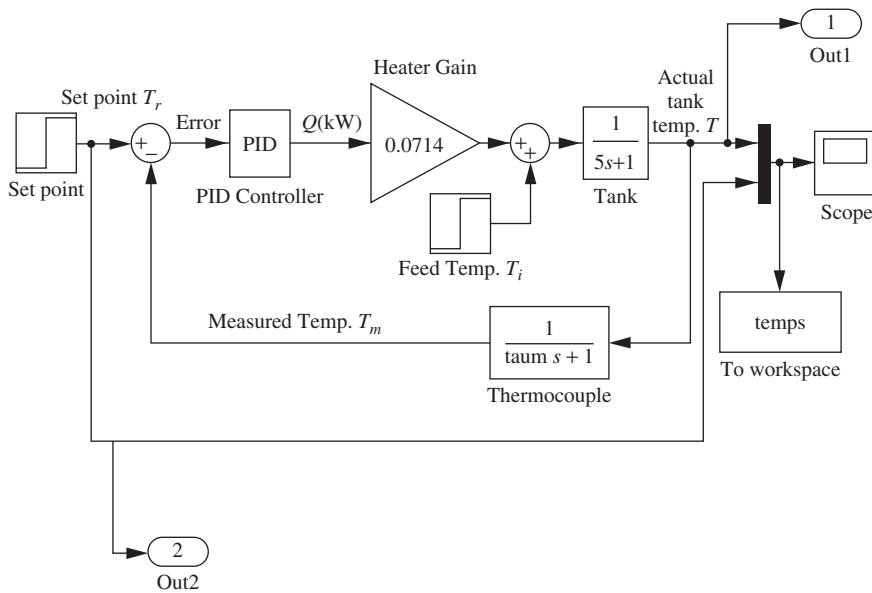

FIGURE 12–19
Simulink model for Example 12.5.

For values of $\tau_m = 0.33$, 1.0, 2.0, and 5.0, determine the nature of the response to a set point change of 5°C. Figure 12–20 shows the results. It is clear that as the measurement lag increases, the response becomes more oscillatory and less stable.
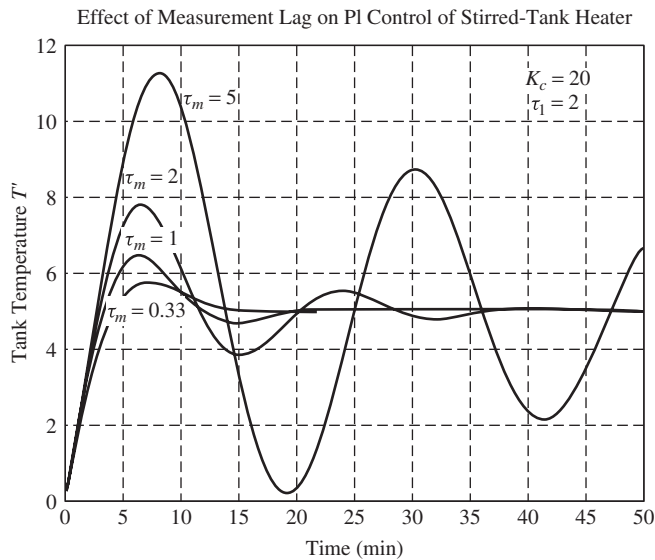
FIGURE 12–20
Effect of measurement lag on PI control of stirred-tank heater.

SUMMARY

In this chapter, we confined our attention to the response of simple control systems that were either first-order or second-order. This means that the transient response can be found by referring to Chaps. 4 and 7. However, if integral action were added to the controller in the system of Fig. 12–17, the overall transfer function would have a third-order polynomial in the denominator. Incorrect selection of controller parameters can lead to a response with increasing amplitude. These unstable responses can occur in all systems with third- or higher-order polynomials in the denominator of the overall transfer function. Inversion would require factoring a cubic, which is generally a difficult task by hand, but a routine task using a computer. Actually, systems with denominator polynomials of order greater than 2 are the rule rather than the exception. Hence, we will develop in forthcoming chapters convenient techniques for studying the response of higher-order control systems. These techniques will be of direct use in control system design.

Unstable responses can occur in all systems with third- or higher-order polynomials in the denominator of the overall transfer function. In Chap.13 we will present a concrete definition of stability and begin the development of methods for determining stability in control systems.

PROBLEMS

12.1. The set point of the control system shown in Fig. P12–1 is given a step change of 0.1 unit. Determine
   (*a*) The maximum value of C and the time at which it occurs
   (*b*) The offset
   (*c*) The period of oscillation

Draw a sketch of $C(t)$ as a function of time.
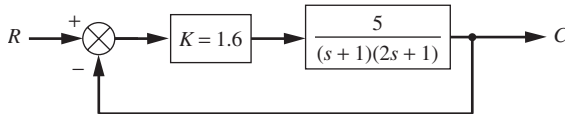
FIGURE P12–1

12.2. The control system shown in Fig. P12–2 contains a PID controller.
   (a) For the closed loop, develop formulas for the natural period of oscillation τ and the damping factor ζ in terms of the parameters K, τ_D, τ_I, and τ_1.
   For the following parts, $\tau_D = \tau_I = 1$ and $\tau_1 = 2$,
   (b) Calculate ζ when K is 0.5 and when K is 2.
   (c) Do ζ and τ approach limiting values as K increases, and if so, what are these values?
   (d) Determine the offset for a unit-step change in load if K is 2.
   (e) Sketch the response curve (C versus t) for a unit-step change in load when K is 0.5 and when K is 2.
   (f) In both cases of part (e) determine the maximum value of C and the time at which it occurs.
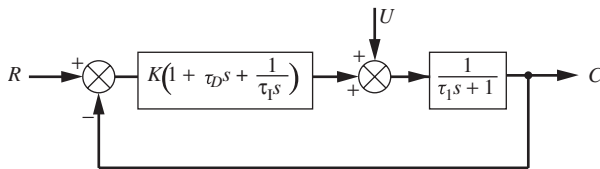
FIGURE P12–2

12.3. The location of a load change in a control loop may affect the system response. In the block diagram shown in Fig. P12–3, a unit-step change in load enters at either location 1 or location 2.
   (a) What is the frequency of the transient response when the load enters at location 1 and when the load enters at location 2?
   (b) What is the offset when the load enters at location 1 and when it enters at location 2?
   (c) Sketch the transient response to a step change in U_1 and to a step change in U_2.
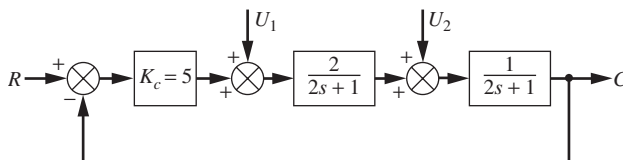
FIGURE P12–3

12.4. Consider the liquid-level control system shown in Fig. P12–4. The tanks are noninteracting. The following information is known:

- The resistances on the tanks are linear. These resistances were tested separately, and it was found that if the steady-state flow rate q cfm is plotted against steady-state tank level h ft, the slope of the line dq/dh is 2 ft²/min.
- The cross-sectional area of each tank is 2 ft².
- The control valve was tested separately, and it was found that a change of 1 psi in pressure to the valve produced a change in flow of 0.1 cfm.
- There is no dynamic lag in the valve or the measuring element.

(a) Draw a block diagram of this control system, and in each block give the transfer function, with numerical values of the parameters.
(b) Determine the controller gain K_c for a critically damped response.
(c) If the tanks were connected so that they were interacting, what is the value of K_c needed for critical damping?
(d) Using 1.5 times the value of K_c determined in part (c), determine the response of the level in tank 2 to a step change in set point of 1 in of level.
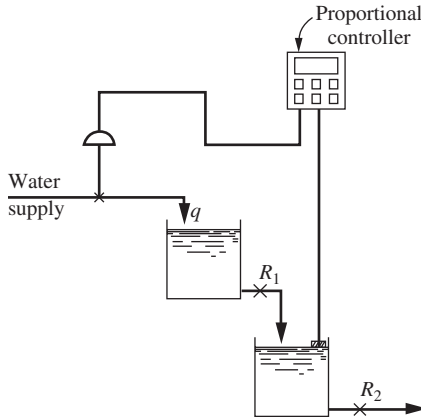
FIGURE P12–4

12.5. A PD controller is used in a control system having a first-order process and a measurement lag as shown in Fig. P12–5.

(a) Find expressions for ζ and τ for the closed-loop response.
(b) If $\tau_1 = 1$ min and $\tau_m = 10$ s, find K_c so that $\zeta = 0.7$ for the two cases (1) $\tau_D = 0$ s and (2) $\tau_D = 3$ s.
(c) Compare the offset and period realized for both cases, and comment on the advantage of adding the derivative mode.
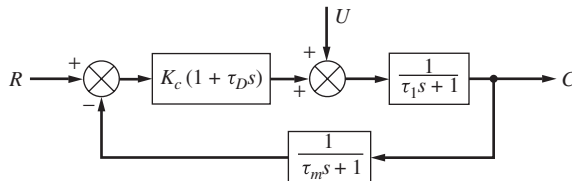
FIGURE P12–5

12.6. The thermal system shown in Fig. P12–6 is controlled by a PD controller. These data are given:

$$w = 250 \text{ lb/min}$$
$$\rho = 62.5 \text{ lb/ft}^3$$
$$V_1 = 4 \text{ ft}^3$$
$$V_2 = 5 \text{ ft}^3$$
$$V_3 = 6 \text{ ft}^3$$
$$C = 1 \text{ Btu/(lb·°F)}$$

A change of 1 psi from the controller changes the flow rate of heat q by 500 Btu/min. The temperature of the inlet stream may vary. There is no lag in the measuring element.

(a) Draw a block diagram of the control system with the appropriate transfer function in each block. Each transfer function should contain numerical values of the parameters.

(b) From the block diagram, determine the overall transfer function relating the temperature in tank 3 to a change in set point.

(c) Find the offset for a unit-step change in inlet temperature if the controller gain K_c is 3 psi/°F of temperature error and the derivative time is 0.5 min.
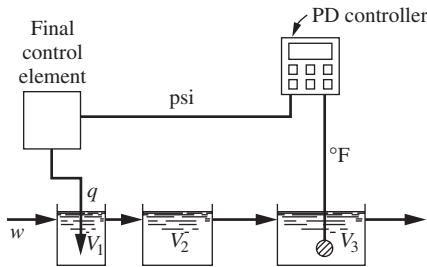
FIGURE P12–6

12.7. (a) For the control system shown in Fig. P12–7, obtain the closed-loop transfer function C/U.

(b) Find the value of K_c for which the closed-loop response has a ζ of 2.3.

(c) Find the offset for a unit-step change in U if $K_c = 4$.
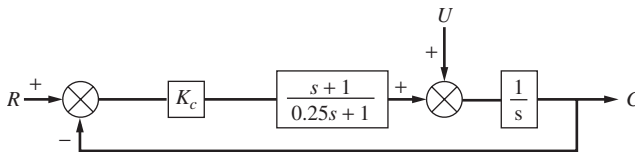
FIGURE P12–7

12.8. For the control system shown in Fig. P12–8, determine

(a) $C(s)/R(s)$

(b) $C(\infty)$

(c) Offset

(d) $C(0.5)$

(e) Whether the closed-loop response is oscillatory
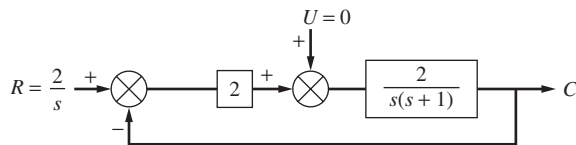
FIGURE P12–8

12.9. For the control system shown in Fig. P12–9, determine an expression for $C(t)$ if a unit-step change occurs in R. Sketch the response $C(t)$ and compute $C(2)$.
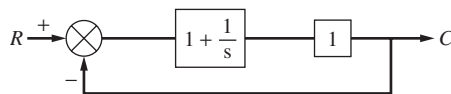
FIGURE P12–9

12.10. Compare the responses to a unit-step change in set point for the system shown in Fig. P12–10 for both negative feedback and positive feedback. Do this for K_c of 0.5 and 1.0. Compare these responses by sketching $C(t)$.
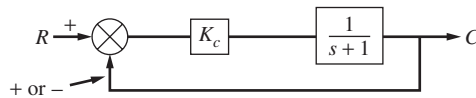
FIGURE P12–10

CAPSULE SUMMARY

Offset: The difference (at steady state) between the desired value of the controlled variable (set point) and the actual value of the controlled variable.

Measurement lag: The response time of the sensor used to measure the value of the controlled variable and send it to the controller. It usually appears in the feedback loop of the control system.

TRANSIENT RESPONSE OF A PROPORTIONAL-ONLY CONTROL SYSTEM

Key Point - With proportional-only control there will always be some offset and

$$\text{As} \quad K_C \uparrow \quad \text{Offset} \downarrow$$

TRANSIENT RESPONSE OF A PROPORTIONAL-INTEGRAL CONTROL SYSTEM

Key Point - With proportional-integral control there is no offset. However, the addition of the integral action in the controller can cause the system response to become more oscillatory and increase the maximum deviation (or overshoot) of the response.

PI Control = No offset, but may be oscillatory

TRANSIENT RESPONSE OF A SYSTEM WITH MEASUREMENT LAG

The larger the measurement lag in a control system, the poorer the control performance. As measurement lag increases, the response of the system will degrade and will become more oscillatory.

CHAPTER

13

STABILITY

13.1 CONCEPT OF STABILITY

In Chap. 12, the overall response of the control system was no higher than second-order. For these systems, the step response must resemble those of Fig. 4–7 or of Fig. 7–3. Hence, the system is inherently *stable*. In this chapter we consider the problem of stability in a control system (Fig. 13–1) only slightly more complicated than any studied previously. This system might represent proportional control of two stirred-tank heaters with measuring lag. In this discussion, only set point changes are to be considered. From the methods developed in Chap. 11 for determining the overall transfer function, we have from Fig. 13–1

$$\frac{C}{R} = \frac{G_1 G_2}{1 + G_1 G_2 H} \tag{13.1}$$
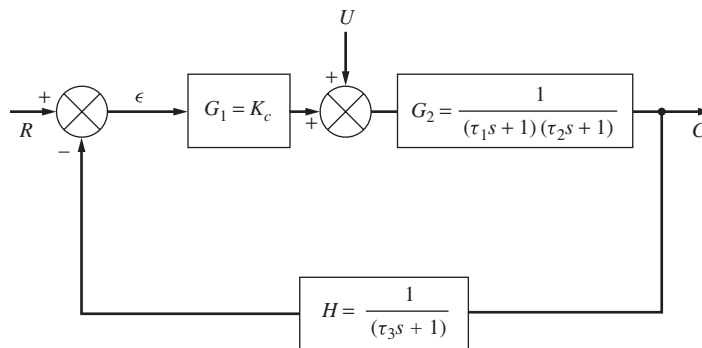
FIGURE 13–1
Third-order control system.

In terms of the particular transfer functions shown in Fig. 13–1, *C/R* becomes, after some rearrangement,

$$\frac{C}{R} = \frac{K_c(\tau_3 s + 1)}{(\tau_1 s + 1)(\tau_2 s + 1)(\tau_3 s + 1) + K_c} \tag{13.2}$$

The denominator of Eq. (13.2) is a third-order polynomial. For a unit-step change in *R*, the transform of the response is

$$C = \frac{1}{s}\frac{K_c(\tau_3 s + 1)}{(\tau_1 s + 1)(\tau_2 s + 1)(\tau_3 s + 1) + K_c} \tag{13.3}$$

To obtain the transient response *C(t)*, it is necessary to find the inverse of Eq. (13.3). This requires obtaining the roots of the denominator of Eq. (13.2), which is third-order. We can no longer find these roots as easily as we did for the second-order systems by use of the quadratic formula. However, in principle they can always be obtained by algebraic methods or through the use of computer software such as MATLAB.

It is apparent that the roots of the denominator depend upon the particular values of the time constants and K_c. These roots determine the nature of the transient response, according to the rules presented in Fig. 3–1 and Table 3.1. It is of interest to examine the nature of the response for the control system of Fig. 13–1 as K_c is varied, assuming the time constants τ_1, τ_2, and τ_3 to be fixed. To be specific, consider the step response for $\tau_1 = 1$, $\tau_2 = \frac{1}{2}$, and $\tau_3 = \frac{1}{3}$ for several values of K_c. Without going into the detailed calculations at this time, the results of inversion of Eq. (13.3) are shown as response curves in Fig. 13–2. From these response curves, it is seen that as K_c increases, the system response becomes more oscillatory. In fact, beyond a certain value of K_c, the successive amplitudes of the response grow rather than decay; this type of response is called *unstable*. Evidently, for some values of K_c, there is a pair of roots corresponding to s_4 and s_4^* of Fig. 3–1. As control system designers, we are clearly interested in being able to determine quickly the values of K_c that give unstable responses, such as that corresponding to $K_c = 12$ in Fig. 13–2.
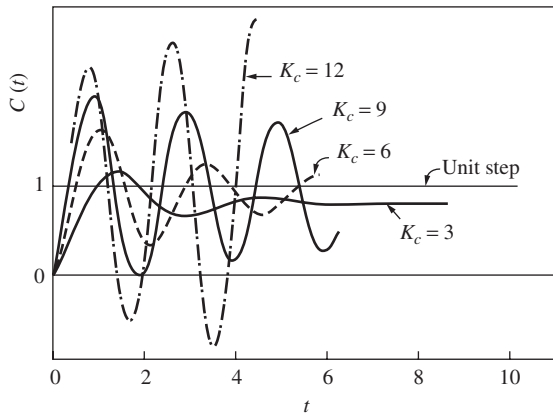
FIGURE 13–2
Response of control system of Fig. 13–1 for a unit-step change in set point.

If the order of Eq. (13.2) had been higher than 3, the hand calculations necessary to obtain Fig. 13–2 would have been even more difficult. In this chapter, the focus is on developing a clearer understanding of the concept of stability. In addition, we develop a quick test for detecting roots having positive real parts, such as s_4 and s_4^* in Fig. 3–1.

13.2   DEFINITION OF STABILITY (LINEAR SYSTEMS)

For our purposes, a *stable* system will be defined as one for which the output response is bounded for all bounded inputs. A system exhibiting an unbounded response to a bounded input is unstable. This definition, although somewhat loose, is adequate for most of the linear systems and simple inputs that we shall study.

STABLE SYSTEM → a *b*ounded *i*nput produces a *b*ounded *o*utput (BIBO)

A bounded input function is a function of time that always falls within certain bounds during the course of time. For example, the step function and sinusoidal function are bounded inputs. The function $f(t) = t$ is obviously unbounded.

Although the definition of an unstable system states that the output becomes unbounded, this is true only in the mathematical sense. An actual physical system always exhibits bounds or constraints. A linear mathematical model (set of linear differential equations describing the system) from which stability information is obtained is meaningful only over a certain range of variables. For example, a linear control valve gives a linear relation between flow and valve-top pressure only over the range of pressure (or flow) corresponding to values between which the valve is shut tight or wide open. When the valve is wide open, for example, further change in pressure to the diaphragm will not increase the flow. We often describe such a limitation by the term *saturation.* A physical system, when unstable, may not follow the response of its linear mathematical model beyond certain physical bounds but rather may saturate. However, the prediction of stability by the linear model is of utmost importance in a real control system since operation with the valve shut tight or wide open is clearly unsatisfactory control.

13.3   STABILITY CRITERION

The purpose of this section is to translate the stability definition into a simpler criterion, one that can be used to ascertain the stability of control systems of the form shown in Fig. 13–3.
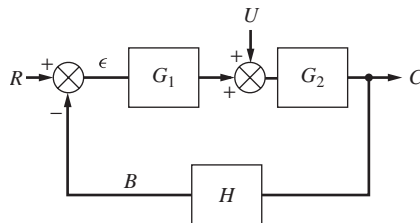
FIGURE 13–3
Basic single-loop control system.

CHARACTERISTIC EQUATION. From the block diagram of the control system (Fig. 13–3), we obtain by the methods of Chap. 12

$$C = \frac{G_1 G_2}{1 + G_1 G_2 H} R + \frac{G_2}{1 + G_1 G_2 H} U \tag{13.4}$$

To simplify the nomenclature, let $G = G_1 G_2 H$. We call G the *open-loop transfer function* because it relates the measured variable B to the set point R if the feedback loop of Fig. 13–3 is disconnected from the comparator (i.e., if the loop is opened). In terms of the open-loop transfer function G, Eq. (13.4) becomes

$$C = \frac{G_1 G_2}{1 + G} R + \frac{G_2}{1 + G} U \tag{13.5}$$

In principle, for given forcing functions $R(s)$ and $U(s)$, Eq. (13.5) may be inverted to give the control system response.

To determine under what conditions the system represented by Eq. (13.5) is stable, it is necessary to test the response to a bounded input. Suppose a unit-step change in set point is applied. Then

$$C(s) = \frac{G_1 G_2}{1 + G} \frac{1}{s} = \frac{G_1 G_2 F(s)}{s(s - r_1)(s - r_2)(s - r_n)} \tag{13.6}$$

where $r_1, r_2, \ldots, r_n$ are the n roots of the equation

$$1 + G(s) = 0 \tag{13.7}$$

and $F(s)$ is a function that arises in the rearrangement to the right-hand form of Eq. (13.6). Equation (13.7) is called the *characteristic equation* for the control system of Fig. 13–3. For example, for the control system of Fig. 13–1 the step response is

$$C(s) = \frac{G_1 G_2}{s(1 + G)}$$

$$= \frac{K_c / [(\tau_1 s + 1)(\tau_2 s + 1)]}{s[1 + K_c / (\tau_1 s + 1)(\tau_2 s + 1)(\tau_3 s + 1)]}$$

which may be rearranged to

$$C(s) = \frac{K_c(\tau_3 s + 1)}{s[\tau_1 \tau_2 \tau_3 s^3 + (\tau_1 \tau_2 + \tau_1 \tau_3 + \tau_2 \tau_3)s^2 + (\tau_1 + \tau_2 + \tau_3)s + (1 + K_c)]}$$

This is equivalent to

$$C(s) = \frac{K_c(\tau_3 s + 1)/\tau_1 \tau_2 \tau_3}{s(s - r_1)(s - r_2)(s - r_3)}$$

where r_1, r_2, and r_3 are the roots of the characteristic equation

$$\tau_1\tau_2\tau_3 s^3 + (\tau_1\tau_2 + \tau_1\tau_3 + \tau_2\tau_3)s^2 + (\tau_1 + \tau_2 + \tau_3)s + (1 + K_c) = 0 \quad (13.8)$$

Evidently, for this case the function $F(s)$ in Eq. (13.6) is

$$F(s) = \frac{(\tau_1 s + 1)(\tau_2 s + 1)(\tau_3 s + 1)}{\tau_1\tau_2\tau_3}$$

In Chap. 3, the qualitative nature of the inverse transforms of equations such as Eq. (13.6) was discussed. It was shown that (see Fig. 3–1 and Table 3.1) if there are any of the roots $r_1, r_2, \ldots, r_n$ in the right half of the complex plane, the response $C(t)$ will contain a term that grows exponentially in time and the system is unstable. If there are one or more roots of the characteristic equation at the origin, there is an s^m in the denominator of Eq. (13.6) (where $m \geq 2$) and the response is again unbounded, growing as a polynomial in time. This condition specifies m as greater than or equal to 2, not 1, because one of the s terms in the denominator is accounted for by the fact that the input is a unit step ($1/s$) in Eq. (13.6). (Note that a $1/s$ term will invert to a constant, while a $1/s^2$ term will invert to a term of the form $C_1 t$, which is unbounded.) Additionally, if there is a pair of conjugate roots of the characteristic equation on the imaginary axis, the contribution to the overall step response is a pure sinusoid, which is bounded. However, if the bounded input is taken as $\sin \omega t$, where ω is the imaginary part of the conjugate roots, the contribution to the overall response is a sinusoid with an amplitude that increases as a polynomial in time [the response will have a term of the form $C_1 t \sin(\omega t + \phi)$]. Thus, if a root lies on the imaginary axis, there is the potential for repeating the root of a bounded input (such as a step input or a sinusoid input), and the response will be unstable. Therefore, the right-half plane, including the imaginary axis, is the unstable region for location of roots of the characteristic equation. It is evident from Eq. (13.5) that precisely the same considerations apply to a change in the load U, as we just discussed for set point changes, since they have the same characteristic equation.

> **Therefore, the definition of *stability for linear systems* may be translated to the following criterion: A linear control system is unstable if any roots of its characteristic equation are on, or to the right of, the imaginary axis. Otherwise the system is stable.**

It is important to note that the characteristic equation of a control system, which determines its stability, is the same for set point or load changes. The stability depends only upon $G(s)$, the open-loop transfer function. Furthermore, although the rules derived above were based on a step input, they are applicable to any input. This is true, first, by the definition of stability and, second, because if there is a root of the characteristic equation in the right half-plane, it contributes an unbounded term in the response to any input. This follows from Eq. (13.5) after it is rearranged to the form of Eq. (13.6) for the particular input.
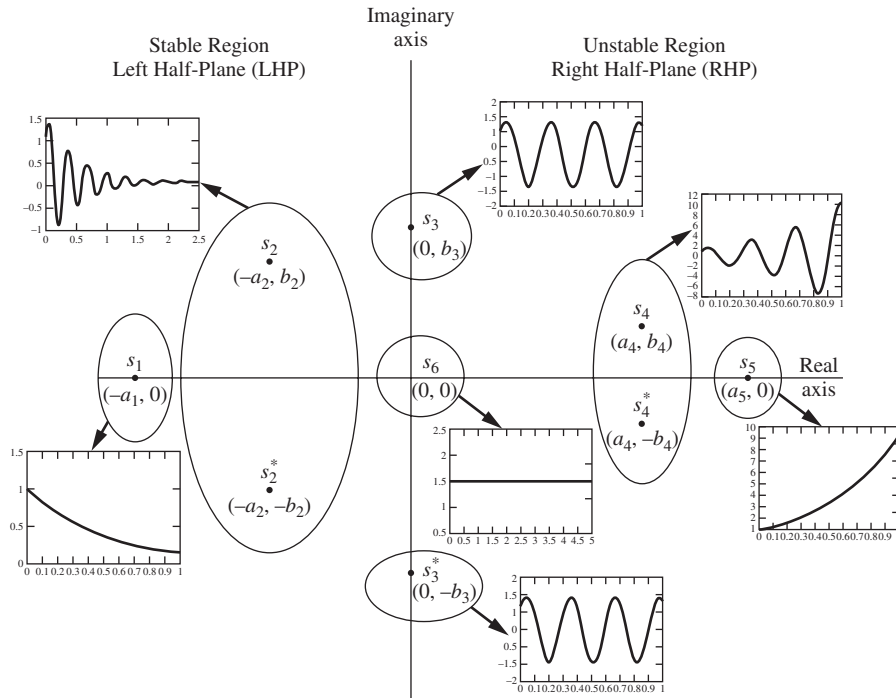
FIGURE 13–4
Stability of typical roots of the characteristic equation.

Therefore, the stability of a control system of the type shown in Fig. 13–3 is determined solely by its open-loop transfer function through the roots of the characteristic equation. This behavior is shown graphically in Fig. 13–4, which relates the nature of the response to the location of the roots of the characteristic equation.

Example 13.1.   In terms of Fig. 13–3, a control system has the transfer functions

$$G_1 = 10\frac{0.5s + 1}{s} \quad \text{(PI controller)}$$

$$G_2 = \frac{1}{2s + 1} \quad \text{(stirred tank)}$$

$$H = 1 \quad \text{(measuring element without lag)}$$

We have suggested a physical system by the components placed in parentheses. Find the characteristic equation and its roots, and determine whether the system is stable.

The first step is to write the open-loop transfer function

$$G = G_1 G_2 H = \frac{10(0.5s + 1)}{s(2s + 1)}$$

The characteristic equation is therefore

$$1 + \frac{10(0.5s + 1)}{s(2s + 1)} = 0$$

which is equivalent to

$$s^2 + 3s + 5 = 0$$

Solving by the quadratic formula gives

$$s = \frac{-3}{2} \pm \frac{\sqrt{9 - 20}}{2}$$

or

$$s_1 = \frac{-3}{2} + j\frac{\sqrt{11}}{2}$$

$$s_2 = \frac{-3}{2} - j\frac{\sqrt{11}}{2}$$

Since the real part of s_1 and s_2 is negative $(-3/2)$, the system is stable.

13.4   ROUTH TEST FOR STABILITY

The Routh test is a purely algebraic method for determining how many roots of the characteristic equation have positive real parts; from this it can also be determined whether the system is stable, for if there are no roots with positive real parts, the system is stable. The test is limited to systems that have polynomial characteristic equations. This means that it cannot be used to test the stability of a control system containing a transportation lag. The procedure for application of the Routh test is presented without proof. The proof is available elsewhere (Routh, 1905) and is mathematically beyond the scope of this text.

The procedure for examining the roots is to write the characteristic equation in the form

$$a_0 s^n + a_1 s^{n-1} + a_2 s^{n-2} + \cdots + a_n = 0 \tag{13.9}$$

where a_0 is positive. (If a_0 is originally negative, both sides are multiplied by -1.) In this form, it is *necessary* that all the coefficients

$$a_0, \ a_1, \ a_2, \ldots, \ a_{n-1}, \ a_n$$

be positive if all the roots are to lie in the left half-plane. If any coefficient is negative, the system is definitely unstable, and the Routh test is not needed to answer the question of stability. (However, in this case, the Routh test will tell us the number of roots in the right half-plane.) If all the coefficients are positive, the system may be stable or unstable. It is then necessary to apply the following procedure to determine stability.

Routh Array

Arrange the coefficients of Eq. (13.9) into the first two rows of the Routh array as follows:

Row				
1	a_0	a_2	a_4	a_6
2	a_1	a_3	a_5	a_7
3	b_1	b_2	b_3	
4	c_1	c_2	c_3	
5	d_1	d_2		
6	e_1	e_2		
7	f_1			
$n+1$	g_1			

The array has been filled in for $n = 7$ to simplify the discussion. For any other value of n, the array is prepared in the same manner. In general, there are $n + 1$ rows. For n even, the first row has one more element than the second row.

The elements in the remaining rows are found from the formulas

$$b_1 = \frac{a_1 a_2 - a_0 a_3}{a_1} \qquad b_2 = \frac{a_1 a_4 - a_0 a_5}{a_1} \quad \cdots$$

$$c_1 = \frac{b_1 a_3 - a_1 b_2}{b_1} \qquad c_2 = \frac{b_1 a_5 - a_1 b_3}{b_1} \quad \cdots$$

The elements for the other rows are found from formulas that correspond to those just given. The elements in any row are always derived from the elements of the two preceding rows. During the computation of the Routh array, any row can be divided by a positive constant without changing the results of the test. (The application of this rule often simplifies the arithmetic.)

Having obtained the Routh array, we can apply the following theorems to determine stability.

THEOREMS OF THE ROUTH TEST

Theorem 13.1.   The necessary and sufficient condition for all the roots of the characteristic equation [Eq. (13.9)] to have negative real parts (stable system) is that all elements of the first column of the Routh array (a_0, a_1, b_1, c_1, etc.) be positive and nonzero.

Theorem 13.2.   If some of the elements in the first column are negative, the number of roots with a positive real part (in the right half-plane) is equal to the number of sign changes in the first column.

Theorem 13.3.   If *one* pair of roots is on the imaginary axis, equidistant from the origin, and all other roots are in the left half-plane, then all the elements of the nth row will vanish and none of the elements of the preceding row will vanish. The location of the pair of imaginary roots can be found by solving the equation

$$Cs^2 + D = 0 \qquad\qquad (13.10)$$

where the coefficients C and D are the elements of the array in the $(n-1)$st row as read from left to right, respectively.

The algebraic method for determining stability is limited in its usefulness in that all we can learn from it is whether a system is stable. It does not give us any idea of the degree of stability or the roots of the characteristic equation.

Example 13.2.   Given the characteristic equation

$$s^4 + 3s^2 + 5s^2 + 4s + 2 = 0$$

determine the stability by the Routh criterion.

Since all the coefficients are positive, the system may be stable. To test this, form the following Routh array:

Row			
1	1	5	2
2	3	4	
3	11/3	6/3	
4	26/11	0	
5	2		

The elements in the array are found by applying the formulas presented in the rules; for example, b_1, which is the element in the first column, third row, is obtained by

$$b_1 = \frac{a_1 a_2 - a_0 a_3}{a_1}$$

or in terms of numerical values,

$$b_1 = \frac{(3)(5) - (1)(4)}{3} = \frac{15}{3} - \frac{4}{3} = \frac{11}{3}$$

Since there is no change in sign in the first column, there are no roots having positive real parts, and the system is stable.

Example 13.3.   (*a*) Using $\tau_1 = 1$, $\tau_2 = \frac{1}{2}$, and $\tau_3 = \frac{1}{3}$, determine the values of K_c for which the control system in Fig. 13–1 is stable. (*b*) For the value of K_c for which the system is on the threshold of instability, determine the roots of the characteristic equation with the help of Theorem 13.3.

Solution

(*a*) The characteristic equation $1 + G(s) = 0$ becomes

$$1 + \frac{K_c}{(s + 1)(s / 2 + 1)(s / 3 + 1)} = 0$$

Using MATLAB to Determine the Roots of a Polynomial

The MATLAB command `roots(C)` computes the roots of the polynomial whose coefficients are the elements of the vector C. If C has $n + 1$ elements, the polynomial is

$$C_1 x^n + C_2 x^{n-1} + C_3 x^{n-2} + \cdots + C_{n-1} x^2 + C_n x + C_{n+1}$$

For the characteristic equation of Example 13.2, $s^4 + 3s^2 + 5s^2 + 4s + 2 = 0$, the vector C would be

$$C = [1, 3, 5, 4, 2]$$

The MATLAB command and the resulting solution are shown below.
```
roots([1,3,5,4,2])

ans =
  -1.0000 + 1.0000i
  -1.0000 - 1.0000i
  -0.5000 + 0.8660i
  -0.5000 - 0.8660i
```

The real parts of all four roots are negative. Thus, the roots all lie in the left half-plane (LHP), and the system is stable. This is the same result that we obtained using the Routh test.

Rearrangement of this equation for use in the Routh test gives

$$s^3 + 6s^2 + 11s + 6(1 + K_c) = 0 \qquad (13.11)$$

The Routh array is

Row		
1	1	11
2	6	$6(1 + K_c)$
3	$10 - K_c$	
4	$6(1 + K_c)$	

Since the proportional sensitivity of the controller K_c is a positive quantity, we see that the fourth entry in the first column, $6(1 + K_c)$, is positive. According to Theorem 13.1, all the elements of the first column must be positive for stability; hence

$$10 - K_c > 0$$
$$K_c < 10$$

It is concluded that the system will be stable only if $K_c < 10$, which agrees with Fig. 13–2.

(b) At $K_c = 10$, the system is on the verge of instability, and the element in the nth (third) row of the array is zero. According to Theorem 13.3, the location of the imaginary roots is obtained by solving

$$Cs^2 + D = 0$$

where C and D are the elements in the $(n - 1)$st row. For this problem, with $K_c = 10$, we obtain

$$6s^2 + 66 = 0$$

$$s = \pm j\sqrt{11}$$

Therefore, two of the roots on the imaginary axis are located at $\sqrt{11}$ and $-\sqrt{11}$.
   The third root can be found by expressing Eq. (13.11) in factored form

$$(s - s_1)(s - s_2)(s - s_3) = 0 \qquad (13.12)$$

where s_1, s_2, and s_3 are the roots. Introducing the two imaginary roots ($s_1 = j\sqrt{11}$ and $s_2 = -j\sqrt{11}$) into Eq. (13.12) and multiplying out the terms give

$$s^3 - s_3 s^2 + 11s - 11s_3 = 0$$

Comparing this equation with Eq. (13.11), we see that $s_3 = -6$. The roots of the characteristic equation are therefore $s_1 = j\sqrt{11}$, $s_2 = -j\sqrt{11}$, and $s_3 = -6$.

Using MATLAB and $s^3 + 6s^2 + 11s + 6(1 + K_c) = 0$ with $K_c = 10$, we can use the **roots** command as follows.
roots([1,6,11,66])

```
ans =
  -6.0000
   0.0000 + 3.3166i
   0.0000 - 3.3166i
```
The result is the same, $\sqrt{11} = 3.3166$.

Example 13.4.   Determine the stability of the system shown in Fig. 13–1 for which a PI controller is used. Use $\tau_1 = 1$, $\tau_2 = \frac{1}{2}$, $\tau_3 = \frac{1}{3}$, $K_c = 5$, and $\tau_I = 0.25$.

Solution
Characteristic equation is

$$1 + \frac{(K_c \,/\, \tau_1 \tau_2 \tau_3)(\tau_I s + 1)}{\tau_I s [s + (1 \,/\, \tau_1)][s + (1 \,/\, \tau_2)][s + (1 \,/\, \tau_3)]} = 0$$

Using the parameters given above in this equation leads to

$$s^4 + 6s^3 + 11s^2 + 36s + 120 = 0$$

Notice that the order of the characteristic equation has increased from three to four as a result of adding integral action to the controller. The Routh array becomes

Row			
1	1	11	120
2	6	36	
3	5	120	
4	-108		
5	120		

Because there are two sign changes in the first column, we know from Theorem 13.2 of the Routh test that two roots have positive real parts. From Example 13.3 we know that for $K_c = 5$ the system is stable with proportional control. With integral action present, however, the system is unstable for $K_c = 5$.

Employing MATLAB to help us again yields

```
roots([1,6,11,36,120])

ans =
  -3.7953 + 0.8664i
  -3.7953 - 0.8664i
   0.7953 + 2.6992i
   0.7953 - 2.6992i
```

Note the two roots in the RHP (with the positive real parts of 0.7953). The system is unstable. This is the same result we obtained with the Routh test.

SUMMARY

A definition of stability for a control system has been presented and discussed. This definition was translated into a simple mathematical criterion relating stability to the location of roots of the characteristic equation. Briefly, it was found that a control system is stable if all the roots of its characteristic equation lie in the left half of the complex plane. The Routh criterion, a simple algebraic test for detecting roots of a polynomial lying in the right half of the complex plane, was presented and applied to control system stability analysis. This criterion suffers from two limitations: (1) It is applicable only to systems with polynomial characteristic equations, and (2) it gives no information about the actual location of the roots and, in particular, their proximity to the imaginary axis.

This latter point is quite important, as can be seen from Fig. 13–2 and the results of Example 13.3. The Routh criterion tells us only that for $K_c < 10$ the system is stable. However, from Fig. 13–2 it is clear that the value $K_c = 9$ produces a response that is undesirable because it has a response time that is too long. In other words, the controlled variable oscillates too long before returning to steady state. It will be shown later that this happens because for $K_c = 9$ there is a pair of roots close to the imaginary axis.

In Chap. 14 tools will be developed for obtaining more information about the actual location of the roots of the characteristic equation. This will enable us to predict the form of the curves of Fig. 13–2 for various values of K_c. The advantage of these tools is that they are graphical and are easy to apply compared with standard algebraic solutions of the characteristic equation.

There are two distinct approaches to this problem: root locus methods and frequency-response methods. The former are discussed in Chap. 14 and the latter in Chaps. 15 and 16. These groups of chapters are written in parallel, and the reader may study one or both groups in either order. As a guide to making this decision, here are some general comments concerning the two approaches.

Root locus methods allow rapid determination of the location of the roots of the characteristic equation as functions of parameters such as K_c of Fig. 13–1. However, they are difficult to apply to systems containing transportation lags. Also, they require a reasonably accurate knowledge of the theoretical process transfer function.

Frequency-response methods are an indirect solution to the location of the roots. They utilize the sinusoidal response of the open-loop transfer function to determine values of parameters such as K_c that keep these roots a "safe distance" from the right half-plane. The actual transient response for a given value of K_c can be only crudely approximated. However, frequency-response methods are easily applied to systems containing transportation lags and may be used with only experimental knowledge of the unsteady-state process behavior.

A mastery of control theory requires knowledge of both methods because they are complementary. However, the reader may choose to study only frequency response and still be adequately prepared for most of the material in the remainder of this book. The choice of studying only root locus will be more restrictive in terms of preparation for subsequent chapters. In addition, much of the literature on process dynamics relies heavily on frequency-response methods.

PROBLEMS

13.1. Write the characteristic equation and construct the Routh array for the control system shown in Fig. P13–1. Is the system stable for (*a*) $K_c = 9.5$, (*b*) $K_c = 11$, and (*c*) $K_c = 12$?
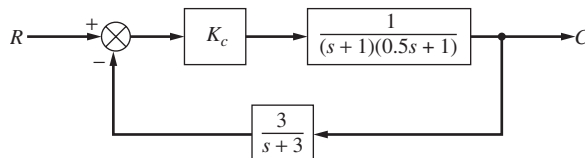

FIGURE P13–1

13.2. By means of the Routh test, determine the stability of the system shown in Fig. P13–2 when $K_c = 2$.
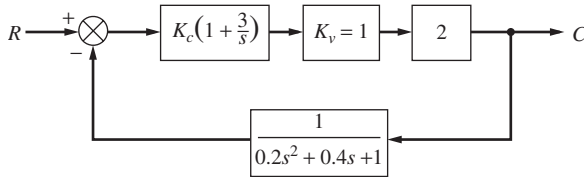
FIGURE P13–2

13.3. In the control system of Prob. 12.6, determine the value of gain (psi/°F) that just causes the system to be unstable if (a) $\tau_D = 0.25$ min and (b) $\tau_D = 0.5$ min.

13.4. Prove that if one or more of the coefficients $a_0, a_1, \ldots, a_n$ of the characteristic equation [Eq. (13.9)] is negative or zero, then there is necessarily an unstable root. *Hint:* First show that a_1/a_0 is minus the sum of all the roots, a_2/a_0 is plus the sum of all possible products of two roots, a_j/a_0 is $(-1)^j$ times the sum of all possible products of j roots, etc.

13.5. Prove that the converse statement of Prob. 13.4—that an unstable root implies that one or more of the coefficients will be negative or zero—is untrue for all $n > 2$. *Hint:* To prove that a statement is untrue, it is only necessary to demonstrate a single counterexample.

13.6. Deduce an extension of the Routh criterion that will detect the presence of roots with real parts greater than $-\sigma$ for any specified $\sigma > 0$.

13.7. Show that any complex number s satisfying $|s| < 1$ yields a value of

$$z = \frac{1 + s}{1 - s}$$

that satisfies

$$\text{Re}(z) > 0$$

(*Hint:* Let $s = x + jy$; $z = u + jv$. Rationalize the fraction, and equate real and imaginary parts of z and the rationalized fraction. Now consider what happens to the circle $x^2 + y^2 = 1$. To show that the *inside* of the circle goes over to the right half-plane, consider a convenient point inside the circle.)

On the basis of this transformation, deduce an extension of the Routh criterion that will determine whether the system has roots inside the unit circle. Why might this information be of interest? How can the transformation be modified to consider circles of other radii?
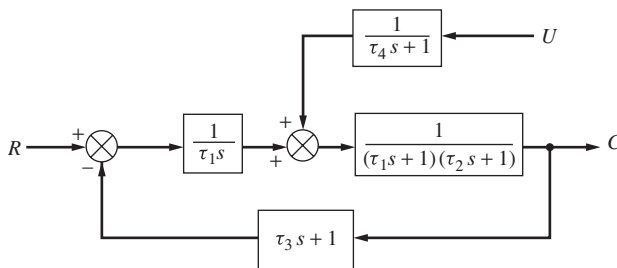

FIGURE P13–8

13.8. Given the control diagram shown in Fig. P13–8, deduce by means of the Routh criterion those values of τ_I for which the output C is stable for all inputs R and U.

13.9. In the control system shown in Fig. P13–9, find the value of K_c for which the system is on the verge of instability. The controller is replaced by a PD controller, for which the transfer function is $K_c\,(\tau_D s + 1)$. If $K_c = 10$, determine the range of τ_D for which the system is stable.
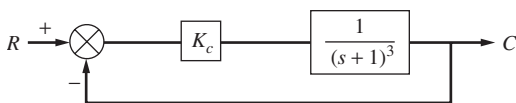
FIGURE P13–9

13.10. (*a*) Write the characteristic equation for the control system shown in Fig. P13–10.
   (*b*) Use the Routh test to determine if the system is stable for $K_c = 4$.
   (*c*) Determine the ultimate value of K_c above which the system is unstable.
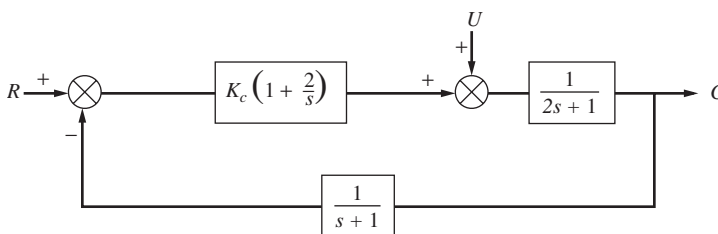
FIGURE P13–10

13.11. For the control system in Fig. P13–11, the characteristic equation is

$$s^4 + 4s^3 + 6s^2 + 4s + (1 + K) = 0$$

   (*a*) Determine the value of K above which the system is unstable.
   (b) Determine the value of K for which two of the roots are on the imaginary axis, and determine the values of these imaginary roots and the remaining two roots.

FIGURE P13–11

CAPSULE SUMMARY

STABLE SYSTEM $\rightarrow$
A *b*ounded *i*nput produces a *b*ounded *o*utput (BIBO).

Characteristic equation: $1 + G(s) = 0$, where $1 + G(s)$ is the denominator of the closed-loop transfer function for the process.

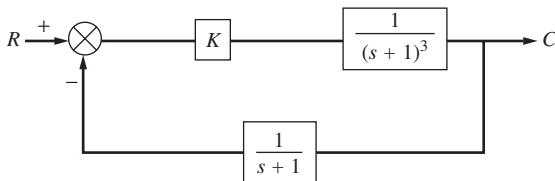**Routh stability test:** A test to determine the stability of a linear control system. To apply the Routh test, rewrite the characteristic equation as

$$a_0 s^n + a_1 s^{n-1} + a_2 s^{n-2} + \cdots + a_n = 0 \qquad (13.9)$$

FIGURE 13–4
Stability of typical roots of the characteristic equation.

where a_0 is positive. Form the following array by using the coefficients:

Row				
1	a_0	a_2	a_4	a_6
2	a_1	a_3	a_5	a_7
3	b_1	b_2	b_3	
4	c_1	c_2	c_3	
5	d_1	d_2		
6	e_1	e_2		
7	f_1			
$n+1$	g_1			

The other elements are found from the formulas

$$b_1 = \frac{a_1 a_2 - a_0 a_3}{a_1} \qquad b_2 = \frac{a_1 a_4 - a_0 a_5}{a_1} \quad \cdots$$

$$c_1 = \frac{b_1 a_3 - a_1 b_2}{b_1} \qquad c_2 = \frac{b_1 a_5 - a_1 b_3}{b_1} \quad \cdots$$

THEOREMS OF THE ROUTH TEST

Theorem 13.1. The necessary and sufficient condition for all the roots of the characteristic equation [Eq. (13.9)] to have negative real parts (stable system) is that all elements of the first column of the Routh array (a_0, a_1, b_1, c_1, etc.) be positive and nonzero.

Theorem 13.2. If some of the elements in the first column are negative, the number of roots with a positive real part (in the right half-plane) is equal to the number of sign changes in the first column.

Theorem 13.3. If *one* pair of roots is on the imaginary axis, equidistant from the origin, and all other roots are in the left half-plane, all the elements of the nth row will vanish and none of the elements of the preceding row will vanish. The location of the pair of imaginary roots can be found by solving the equation

$$Cs^2 + D = 0 \qquad\qquad (13.10)$$

where the coefficients C and D are the elements of the array in the $(n-1)$st row as read from left to right, respectively.

In Chap. 13 on stability, Routh's criterion was introduced to provide an algebraic method for determining the stability of a simple feedback control system (Fig. 13–3) from the characteristic equation of the system [Eq. (13.7)]. This criterion also yields the number of roots of the characteristic equation that are located in the right half of the complex plane. In this chapter, we develop a graphical method for finding the actual values of the roots of the characteristic equation, from which we can obtain the transient response of the system to an arbitrary forcing function.

14.1 CONCEPT OF ROOT LOCUS

In Chap. 13, the response of the simple feedback control system, shown again in Fig. 14–1, was given by the expression

$$C = \frac{G_1 G_2}{1 + G} R + \frac{G_2}{1 + G} U \tag{14.1}$$

where $G = G_1 G_2 H$. The factor in the denominator, $1 + G$, when set equal to zero, is called the characteristic equation of the closed-loop system. The roots of the characteristic equation determine the form (or character) of the response $C(t)$ to any particular forcing function $R(t)$ or $U(t)$.

The *root locus* method is a graphical procedure for finding the roots of $1 + G = 0$, as one of the parameters of G varies continuously. In our work, the parameter that will be varied is the gain (or sensitivity) K_c of the controller. We can illustrate the concept
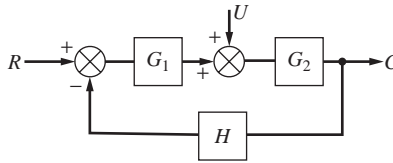
FIGURE 14–1
Simple feedback control system.

of a root locus diagram by considering the example presented in Fig. 13–1, which is represented by the block diagram of Fig. 14–1 with

$$G_1 = K_c$$

$$G_2 = \frac{1}{(\tau_1 s + 1)(\tau_2 s + 1)}$$

$$H = \frac{1}{\tau_3 s + 1}$$

For this case, the open-loop transfer function is

$$G = \frac{K_c}{(\tau_1 s + 1)(\tau_2 s + 1)(\tau_3 s + 1)}$$

which may be written in the alternate form

$$G(s) = \frac{K}{(s - p_1)(s - p_2)(s - p_3)} \tag{14.2}$$

where $\quad K = \dfrac{K_c}{\tau_1 \tau_2 \tau_3}$

$$p_1 = -\frac{1}{\tau_1} \quad p_2 = -\frac{1}{\tau_2} \quad p_3 = -\frac{1}{\tau_3}$$

The terms p_1, p_2, and p_3 are called the *poles* of the open-loop transfer function. A *pole* of $G(s)$ is any value of s for which $G(s)$ approaches infinity. For example, it is clear from Eq. (14.2) that if $s = p_1$, the denominator of Eq. (14.2) is zero and therefore $G(s)$ approaches infinity. Hence $p_1 = -1/\tau_1$ is a pole of $G(s)$.

The characteristic equation for the *closed-loop* system is

$$1 + \frac{K}{(s - p_1)(s - p_2)(s - p_3)} = 0$$

This expression may be written

$$(s - p_1)(s - p_2)(s - p_3) + K = 0 \tag{14.3}$$

Using the same numerical values for the poles that were used at the beginning of Chap. 13 ($-1, -2, -3$) gives

$$(s + 1)(s + 2)(s + 3) + K = 0 \qquad (14.4)$$

where $$K = 6K_c$$

Expanding the product of this equation gives

$$s^3 + 6s^2 + 11s + K + 6 = 0 \qquad (14.5)$$

which is third-order. For any particular value of controller gain K_c, we can obtain the roots of the characteristic equation [Eq. (14.5)]. For example, if $K_c = 4.41$ ($K = 26.5$), Eq. (14.5) becomes

$$s^3 + 6s^2 + 11s + 32.5 = 0$$

Solving this equation for the three roots gives

$$r_1 = -5.10$$
$$r_2 = -0.45 - j2.5$$
$$r_3 = -0.45 + j2.5$$

Note: MATLAB can easily be used to find the roots of a polynomial:

```
>> c=[1,6,11,32.5]; %placing the coefficients of the equation into a
vector
>> roots(c); %finding the roots
ans =
  -5.0931
  -0.4534 + 2.4851i
```

By selecting other values of *K,* other sets of roots are obtained, as shown in Table 14.1. If the roots are all real, the response will be nonoscillatory.

TABLE 14.1
Roots of the characteristic equation $(s + 1)(s + 2)(s + 3) + K = 0$

$K = 6 K_c$	r_1	r_2	r_3
0	-3	-2	-1
0.23	-3.10	-1.75	-1.15
0.39	-3.16	-1.42	-1.42
1.58	-3.45	$-1.28 - j0.75$	$-1.28 + j0.75$
6.6	-4.11	$-0.95 - j1.5$	$-0.95 + j1.5$
26.5	-5.10	$-0.45 - j2.5$	$-0.45 + j2.5$
60.0	-6.00	$0.0 - j3.32$	$0.0 + j3.32$
100.0	-6.72	$0.35 - j4$	$0.35 + j4$

For convenience, we may plot the roots r_1, r_2, and r_3 on the complex plane as K changes continuously. Such a plot is called a **root locus diagram** and is shown in Fig. 14–2. Notice that there are three loci or *branches* corresponding to the three roots and that they "emerge" or begin (for $K = 0$) at the poles of the open-loop transfer function $(-1, -2, -3)$. The direction of increasing K (remember, K is our controller gain, which we can adjust) is indicated on the diagram by an arrow. Also the values of K are marked on each locus. The root locus diagram for this system and others to follow is symmetric with respect to the real axis, and only the portion of the diagram in the upper half-plane need be drawn. This follows from the fact that the characteristic equation for a physical system contains coefficients that are real, and therefore complex roots of such an equation must appear in conjugate pairs.

The root locus diagram has the distinct advantage of giving at a glance the character of the response as the gain of the controller is continuously changed. The diagram of Fig. 14–2 reveals two critical values of K; one is at K_2 where two of the roots become equal, and the other is at K_3 where two of the roots are pure imaginary. It should be clear from the discussion in Chap. 13 that the nature of the response $C(t)$ will depend only on the roots r_1, r_2, r_3. Thus, if the roots are all real, which occurs for $K < 0.39$ in Fig. 14–2, the response will be nonoscillatory.

If two of the roots are complex and have negative real parts ($K_2 < K < K_3$), the response will include damped sinusoidal terms, which will produce an oscillatory response. If we adjust the controller gain such that $K > K_3$, two of the roots are complex and have positive real parts, and the response is a growing sinusoid. Some of these types of responses were shown in Fig. 13–2.
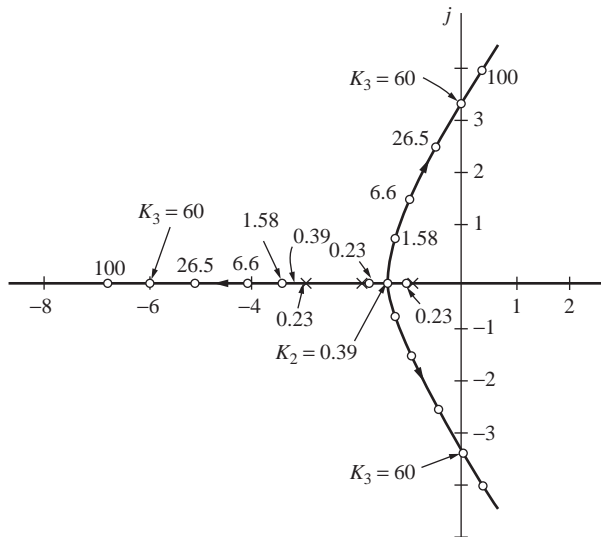

FIGURE 14–2
Root locus diagram for $(s + 1)(s + 2)(s + 3) + K = 0$.

MATLAB can also be used to quickly and easily construct a root locus diagram for the system. We enter the open-loop transfer function in the form of Eq. (14.2) with $K = 1$, and MATLAB generates the root locus diagram:

```
>> num = [1];
>> den = [1 6 11 6];
>> sys = tf(num,den)

Transfer function:
            1
---------------------------
s^3 + 6 s^2 + 11 s + 6
>> rlocus(sys)
```
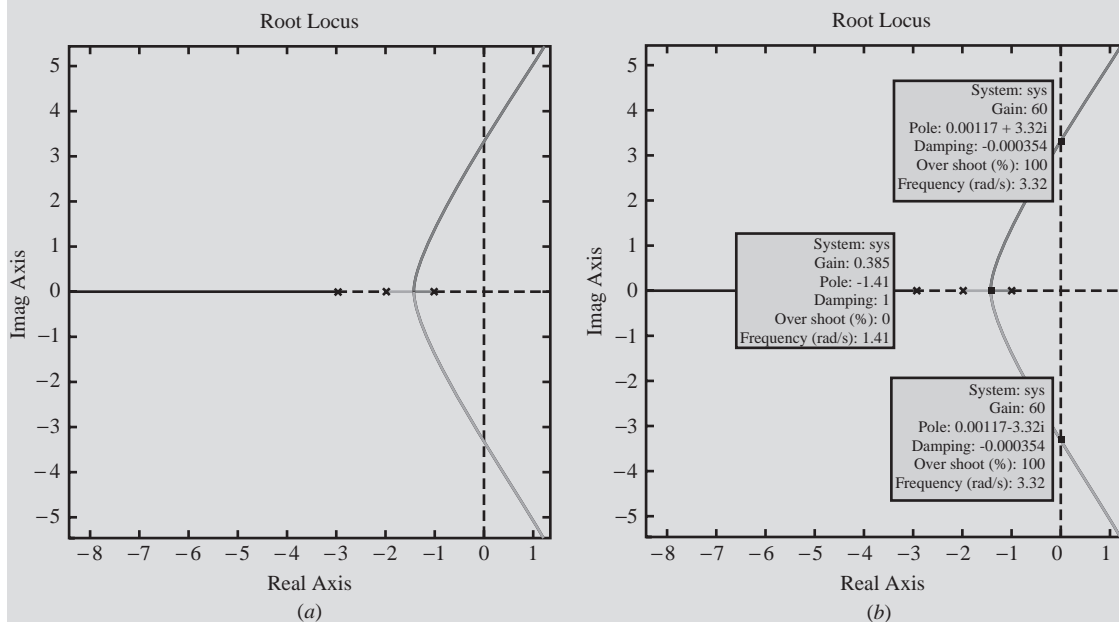

FIGURE 14–3

Root locus diagram for $(s + 1)(s + 2)(s + 3) + K = 0$ generated using MATLAB.

```
By clicking on one of the branches of the diagram in Fig. 14-3a with the
mouse, MATLAB displays information regarding that particular root
location, as shown in Fig. 14-3b.
```

Additionally MATLAB has a graphical tool (SISO tool) that can be used to plot and manipulate root locus diagrams. The command for this case is

```
sisotool(sys)
```

The root locus diagram pops up in an interactive window (see Fig. 14–4). The root locations (the heavy blocks) can be dragged along the branches, and the value of K is displayed as the Current Compensator in the upper left-hand corner of the window.
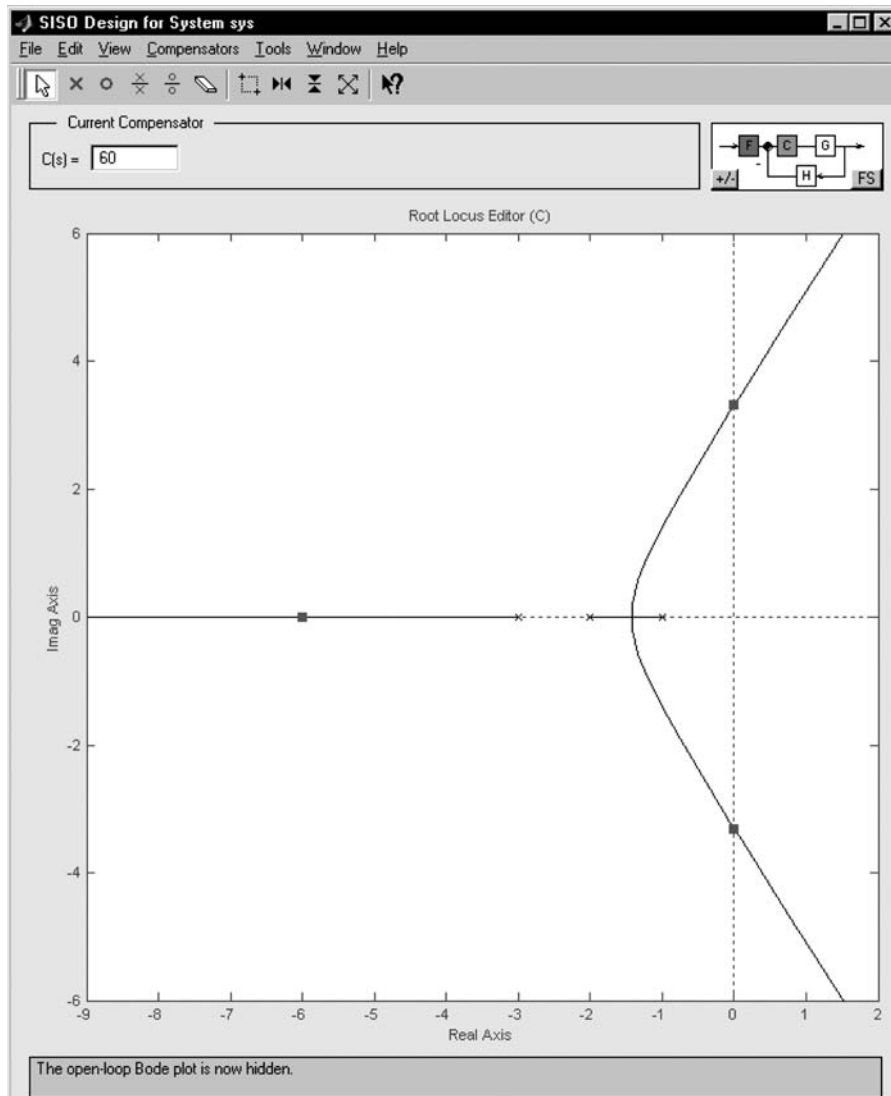

FIGURE 14–4
Root locus diagram generated using MATLAB SISO tool.

As another example of a root locus diagram, let the proportional controller be replaced with a PI controller, for which case G_1 in Fig. 14–1 is

$$G_1 = K_c\left(1 + \frac{1}{\tau_I s}\right)$$

For this case, the open-loop transfer function is

$$G(s) = \frac{K_c(\tau_I s + 1)}{\tau_I s(\tau_1 s + 1)(\tau_2 s + 1)(\tau_3 s + 1)}$$

which may be written in an alternate form

$$G(s) = \frac{K(s - z_1)}{s(s - p_1)(s - p_2)(s - p_3)} \tag{14.6}$$

where $K = \dfrac{K_c}{\tau_1 \tau_2 \tau_3}$

$z_1 = -\dfrac{1}{\tau_I}$

$p_1 = -\dfrac{1}{\tau_1}, \quad p_2 = -\dfrac{1}{\tau_2}, \quad p_1 = -\dfrac{1}{\tau_3}$

The term z_1 is called a *zero* of the open-loop transfer function. A zero of $G(s)$ is any value of s for which $G(s)$ approaches zero. By comparing Eq. (14.6) with Eq. (14.2), we see that the addition of integral action contributes to the open-loop transfer function one *zero* at z_1 and one additional pole at the origin.

The characteristic equation corresponding to Eq. (14.6) is

$$1 + \frac{K(s - z_1)}{s(s - p_1)(s - p_2)(s - p_3)} = 0 \tag{14.7}$$

This expression may be written

$$s(s - p_1)(s - p_2)(s - p_3) + K(s - z_1) = 0 \tag{14.8}$$

As a specific example of the root locus diagram corresponding to Eq. (14.8), let $\tau_1 = 1$, $\tau_2 = \frac{1}{2}$, $\tau_3 = \frac{1}{3}$, and $\tau_I = \frac{1}{4}$. These parameters are the same as those used in Example 13.4. The root locus diagram is shown in Fig. 14–5.

The MATLAB commands for constructing the root locus diagram for this system are

```
>> num1 = [1 4];
>> den1 = [1 6 11 6 0];
>> sys1 = tf(num1,den1)

Transfer function:
        s + 4
- - - - - - - - - - - - - - - -
s^4 + 6 s^3 + 11 s^2 + 6 s

>> rlocus(sys1)
```
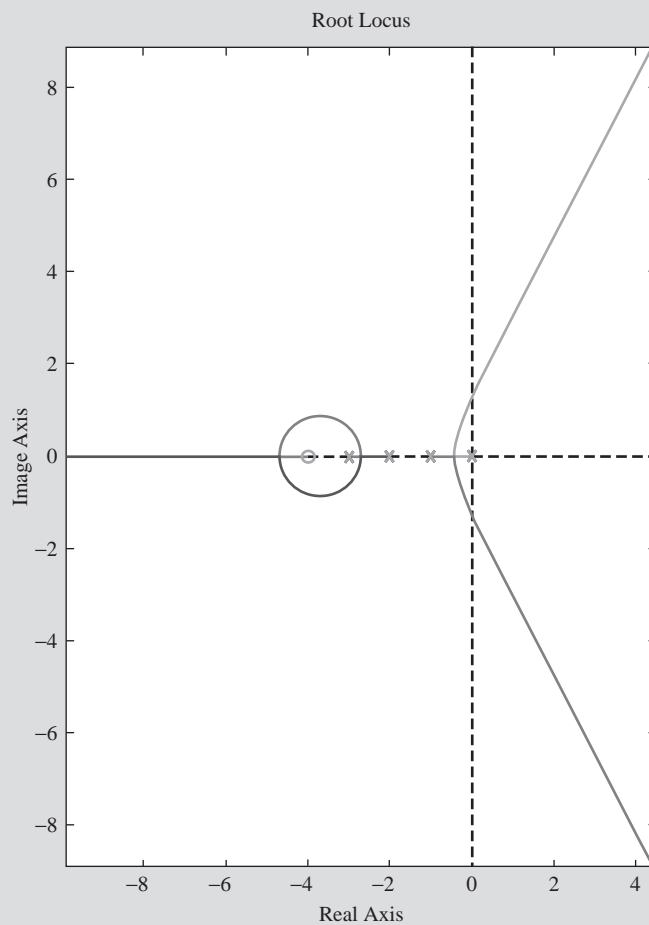

FIGURE 14–6
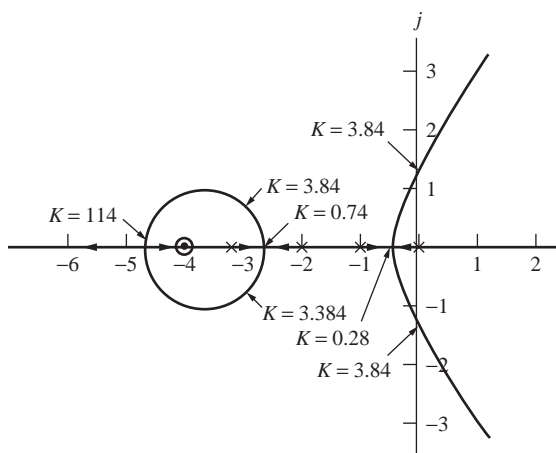Reproduction of root locus diagram of Fig. 14–5 using MATLAB.

FIGURE 14–5
Root locus diagram for $s(s + 1)(s + 2)(s + 3) + K(s + 4) = 0$; $K = 6K_c$.

Notice that for this case there are four loci corresponding to the four roots and that they emerge (at $K = 0$) from the open-loop poles $(0, -1, -2, -3)$. One of the loci moves toward the open-loop zero at -4 as K approaches infinity. The MATLAB box shows the commands for reproducing the root locus diagram in Fig. 14–5. The resulting MAT-LAB graph is shown in Fig. 14–6. The diagram in Fig. 14–5 should be compared with the one in Fig. 14–2 to see the effect of adding integral action to the control system. Notice that the value of $K = 3.84$, above which the roots move into the right half-plane, is lower than the corresponding value of $K = 60$ for proportional control. The effect of adding integral action has been to destabilize the system in terms of the amount of proportional action that can be used before instability occurs.

We can verify the points at which the loci cross the imaginary axis by using the Routh test (Theorem 13.3) of Chap. 13. The characteristic equation for proportional control from Eq. (14.4) is

$$(s + 1)(s + 2)(s + 3) + K = 0$$

or

$$s^3 + 6s^2 + 11s + K + 6 = 0$$

from which we can write the Routh array:

Row		
1	1	11
2	6	$K + 6$
3	b_1	

The theorem states that if one pair of roots is on the imaginary axis and all others are in the left half-plane, then all the elements of the nth row must be zero. From this we obtain for the element b_1

$$b_1 = \frac{(6)(11) - (K + 6)}{6} = 0$$

Solving for K gives

$$K = 60$$

as we found previously (Fig. 14–2). A root on the imaginary axis is expressed as simply ja. Substituting $s = ja$ and $K = 60$ into the polynomial gives

$$-ja^3 - 6a^2 + 11aj + 66 = 0$$
$$\left(66 - 6a^2\right) + \left(11a - a^3\right)j = 0$$

Equating the real part or the imaginary part to zero gives

$$a = \pm\sqrt{11} = \pm 3.32$$

Therefore the loci intersect the imaginary axis at $+3.32j$ and $-3.32j$ (which we also found previously).

Example 14.1   Consider the block diagram for the control system shown in Fig. 14–7. This system may represent a two-tank, liquid-level system having a PID controller and a first-order measuring lag. The open-loop transfer function is

$$G = K_c\frac{1 + 2s/3 + 1/3s}{(20s + 1)(10s + 1)(0.5s + 1)}$$

Rearranging this into the standard form gives

$$G = \frac{K(s - z_1)(s - z_2)}{s(s - p_1)(s - p_2)(s - p_3)}$$

where $K = K_c/150$
$z_1 = -0.5$
$z_2 = -1$
$p_1 = -0.05$
$p_2 = -0.1$
$p_3 = -2$

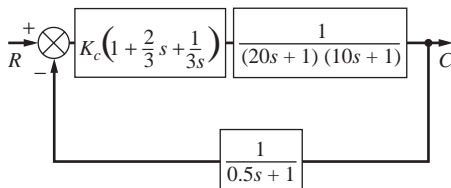**FIGURE 14–7**
Block diagram for Example 14.1.

In this case, there are four poles at 0, -0.05, -0.1, and -2, and two zeros at -0.5 and -1. These are plotted in Fig. 14–8. Note that the PID controller contributes the pole at the origin and the zeros at -0.5 and -1.
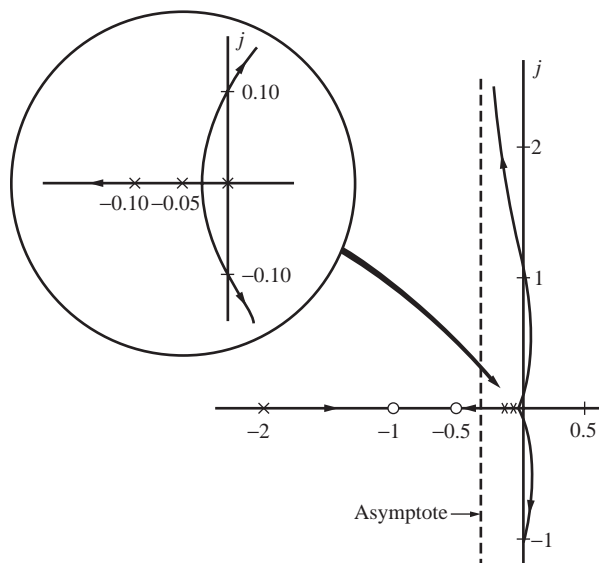

FIGURE 14–8
Root locus diagram for Example 14.1.

The two values of gain K which give a pair of roots of the characteristic equation that lie on the imaginary axis (and the corresponding roots) are

$$K = 0.004 \quad \text{or} \quad K_c = 0.6 \quad s = \pm j0.1$$
$$K = 2.4 \quad \quad \text{or} \quad K_c = 360 \quad s = \pm j1.1$$

From these results, we conclude that the system will oscillate with constant amplitude with a frequency $\omega = 0.1$ rad/time when $K_c = 0.6$. It will also oscillate at constant amplitude with $\omega = 1.1$ when $K_c = 360$. The system is unstable for $0.6 < K_c < 360$. The system is stable for $K_c < 0.6$ and for $K_c > 360$. The complete root locus diagram is sketched in Fig. 14–8.

SUMMARY

In this chapter, root locus diagrams have been presented and applied to a control system. MATLAB can be used to generate the complete root locus diagrams quickly and easily. It should be emphasized that the basic advantage of this method is the speed and ease with which the loci can be obtained. Once the roots are available, the response of the system to any forcing function can be obtained by the usual procedures of partial fractions and inversion given in Chap. 3.

PROBLEMS

14.1. Draw the root locus diagram for the system shown in Fig. P14–1 where
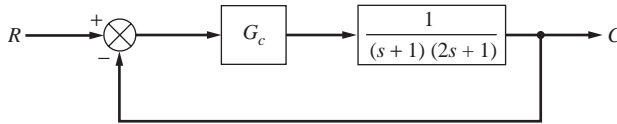
$G_c = K_c(1 + 0.5s + 1/s)$.

$R \longrightarrow \bigotimes \begin{matrix} + \\ - \end{matrix} \longrightarrow \boxed{G_c} \longrightarrow \boxed{\dfrac{1}{(s+1)(2s+1)}} \longrightarrow C$

FIGURE P14–1

14.2. Draw the root locus diagram for the system shown in Fig. P13–4 for (*a*) $\tau_I = 0.4$ min and (*b*) $\tau_I = 0.2$ min. (The proportional controller is replaced by a PI controller.) Determine the controller gain that just causes the system to become unstable. The values of parameters of the system are

$$K_v = \text{valve constant 0.070 cfm/psi}$$
$$K_m = \text{transducer constant } 6.74/(\text{ft of tank level})$$
$$R_2 = 0.55 \text{ ft level/cfm}$$
$$\tau_1 = \text{time constant of tank 1} = 2.0 \text{ min}$$
$$\tau_2 = \text{time constant of tank 2} = 0.5 \text{ min}$$

The controller gain K_c has the units of pounds per square inch (psi).

14.3. Construct the root locus diagram for the system shown in Fig. P13–2. If the system is unstable at higher values of K_c, find the roots on the imaginary axis and the corresponding value of K_c.

14.4. Construct the root loci for the following equations.

(*a*) $1 + \dfrac{K}{(s+1)(2s+1)} = 0$

(*b*) $1 + \dfrac{K}{s(s+1)(2s+1)} = 0$

(*c*) $1 + \dfrac{K(4s+1)}{s(s+1)(2s+1)} = 0$

(*d*) $1 + \dfrac{K(1.5s+1)}{s(s+1)(2s+1)} = 0$

(*e*) $1 + \dfrac{K(0.5s+1)}{s(s+1)(2s+1)} = 0$

On your diagrams you should locate quantitatively all poles and zeros. In addition show the parameter that is being varied along the locus and the direction in which the loci travel as this parameter is increased.

14.5. The control system is shown in Fig. P14–5. There are two cases: case 1: $\tau_D = \frac{2}{3}$ and case 2: $\tau_D = \frac{1}{9}$.
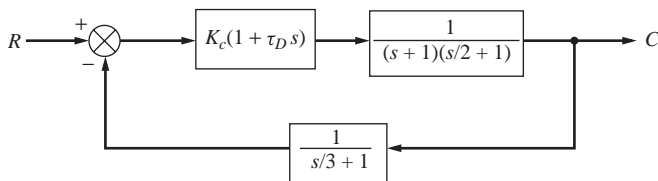

FIGURE P14–5

(a) Sketch the root locus diagram in each case.
(b) If the system can go unstable, find the value of Kc that just causes instability.
(c) Using Theorem 13.3 of the Routh test, find the locations (if any) at which the loci cross into the unstable region.

14.6. Draw the root locus diagram for the control system shown in Fig. P14–6.

(a) Determine the value of K_c needed to obtain a root of the characteristic equation of the closed-loop response which has an imaginary part 0.75.
(b) Using the value of K_c found in part (a), determine all the other roots of the characteristic equation from the root locus diagram.
(c) If a unit impulse is introduced into the set point, determine the response of the system C(t).
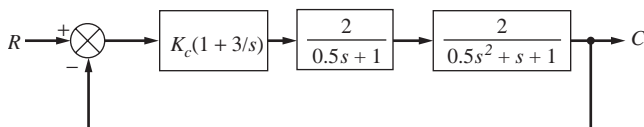
FIGURE P14–6

14.7. Plot the root locus diagram for the system shown in Fig. P14–7. We may consider this system to consist of a process having negligible lag; an underdamped, second-order measuring element; and a PD controller. This system may approximate the control of flow rate, in which case the block labeled K_p would represent a valve having no dynamic lag. The feedback element would represent a flow measuring device, such as a mercury manometer placed across an orifice plate. Mercury manometers are known to have underdamped, second-order dynamics. Plot the diagram for $\tau_D = \frac{1}{3}$.
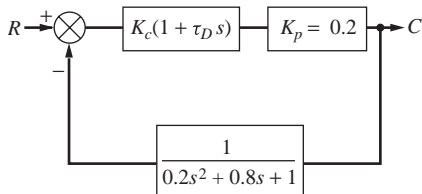
FIGURE P14–7

14.8. Draw the root locus diagram for the proportional control of a plant having the transfer function $2/(s + 1)^3$. Determine the roots on the imaginary axis and the corresponding value of K_c.

14.9. (*a*) Show how you would adopt the usual root locus method for variation in controller gain to the problem of obtaining the root locus diagram for variation in τ_D for the control system shown in Fig. P14–9 for $K_c = 2$.

   (b) Plot the root locus diagram for variation in τ_D with $K_c = 2$.

   (c) Determine the response of the system $C(t)$ for a unit-step change in R for $\tau_D = 0.5$ and $K_c = 2$. Sketch the response. What is the ultimate value of $C(t)$?
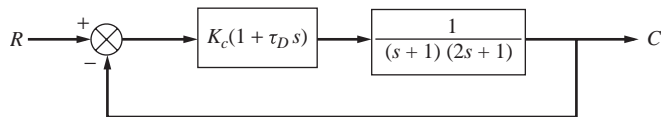
FIGURE P14–9

Hint: Rearrange the open-loop transfer function to be in the form

$$G(s) = \frac{\tau_D s}{s^2 + 1.5s + 1.5}$$

Then apply the usual root locus rules with τ_D taking the place of K_c.

CAPSULE SUMMARY
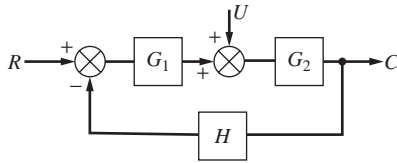
FIGURE 14–1
Simple feedback control system.

For this case, the open-loop transfer function is

$$G = G_1 G_2 H = \frac{K}{(\tau_1 s + 1)(\tau_2 s + 1)(\tau_3 s + 1)}$$

or, in the alternate form,

$$G(s) = \frac{K'}{(s - p_1)(s - p_2)(s - p_3)}$$

where $K' = \dfrac{K}{\tau_1 \tau_2 \tau_3}$

$$p_1 = -\frac{1}{\tau_1} \quad p_2 = -\frac{1}{\tau_2} \quad p_3 = -\frac{1}{\tau_3}$$

Here p_1, p_2, and p_3 are called the *poles* of the open-loop transfer function. A *pole* of $G(s)$ is any value of s for which $G(s)$ approaches infinity (i.e., a root of the denominator).

The characteristic equation for the *closed-loop* system is

$$1 + \frac{K'}{(s - p_1)(s - p_2)(s - p_3)} = 0$$

or

$$(s - p_1)(s - p_2)(s - p_3) + K' = 0$$

The root locus diagram is a plot of the location of the roots of this equation as K' is varied.

MATLAB commands for generating the root locus diagram are

```
% set up the open loop transfer function with  K' = 1
>> num = [1]; %coefficients of numerator polynomial
>> den = [1 6 11 6]; %coeffs. of denom. polynomial
>> sys = tf(num,den);
>> rlocus(sys)
```


Root Locus

(a)
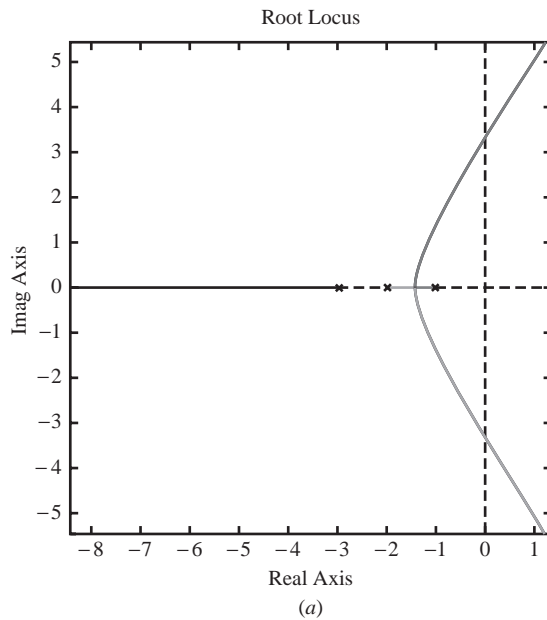
PART

IV

FREQUENCY
RESPONSE

CHAPTER
15

INTRODUCTION TO
FREQUENCY RESPONSE

Chapters 4 and 7 discussed briefly the response of first- and second-order systems to sinusoidal forcing functions. These frequency responses were derived by using the standard Laplace transform technique. In this chapter, a convenient graphical technique will be established for obtaining the frequency response of linear systems. The motivation for doing so will become apparent in Chap. 16, where it will be found that frequency response is a valuable tool in the analysis and design of control systems.

Many of the calculations in this chapter make use of complex numbers. The reader should review the two forms of complex numbers (rectangular and polar) and the basic operations used on complex numbers.

15.1 SUBSTITUTION RULE

A Fortunate Circumstance

Consider a simple first-order system with transfer function

$$G(s) = \frac{1}{\tau s + 1} \tag{15.1}$$

Substituting the quantity $j\omega$ for s in Eq. (15.1) gives

$$G(j\omega) = \frac{1}{j\omega\tau + 1}$$

We may convert this expression to polar form by multiplying numerator and denominator by the conjugate of $j\omega\tau + 1$; the result is

$$G(j\omega) = \frac{-j\omega\tau + 1}{(j\omega\tau + 1)(-j\omega\tau + 1)} = \frac{1}{1 + \omega^2\tau^2} - j\frac{\omega\tau}{1 + \omega^2\tau^2} \qquad (15.2)$$

To convert a complex number in rectangular form $(a + jb)$ to polar form $Re^{j\phi}$, where R = magnitude and ϕ = angle, one uses the relationships

$$R = \sqrt{a^2 + b^2} \quad \text{and} \quad \phi = \tan^{-1}\frac{b}{a}$$

For visualization of the polar form, see Fig. 15–1.
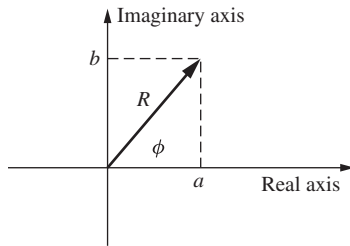
FIGURE 15–1
Complex number representations.

By fixing R and ϕ, we can define the complex number. Applying these relationships to Eq. (15.2) gives

$$G(j\omega) = \underbrace{\frac{1}{\sqrt{\omega^2\tau^2 + 1}}}_{R} \underbrace{\angle \tan^{-1}(-\omega\tau)}_{\phi} \qquad (15.3)$$

The quantities on the right side of Eq. (15.3) are familiar. Recall Eqs. (4.27) and (4.28)

$$Y(t) = \frac{A\omega\tau}{\tau^2\omega^2 + 1}e^{-t/\tau} + \frac{A}{\sqrt{\tau^2\omega^2 + 1}}\sin(\omega t + \phi) \qquad (4.27)$$

where

$$\phi = \tan^{-1}(-\omega\tau)$$

As $t \rightarrow \infty$, the first term on the right side of Eq. (4.27) vanishes and leaves only the ultimate periodic solution, which is sometimes called the steady-state solution and is shown in Fig. 15–2.

$$Y(t)\big|_s = \frac{A}{\sqrt{\tau^2\omega^2 + 1}}\sin(\omega t + \phi) \qquad (4.28)$$

So, after sufficient time elapses, the response of a first-order system to a sinusoidal input of frequency ω is also a sinusoid of frequency ω. Furthermore, from Eq. (4.28) the ratio of the amplitude of the response to that of the input is $1/\sqrt{\omega^2\tau^2 + 1}$, and the phase difference between output and input is $\tan^{-1}(-\omega\tau)$.

Hence, we have shown here that for the frequency response of a first-order system,

$$\text{Amplitude ratio (AR)} = \frac{\text{output amplitude}}{\text{input amplitude}} = \left|G(j\omega)\right| = R$$

$$\text{Phase angle} = \phi = \angle G(j\omega)$$

That is, to obtain the amplitude ratio AR and phase angle, one merely substitutes $j\omega$ for s in the transfer function and then finds the magnitude and argument (or angle) of the resulting complex number, respectively.
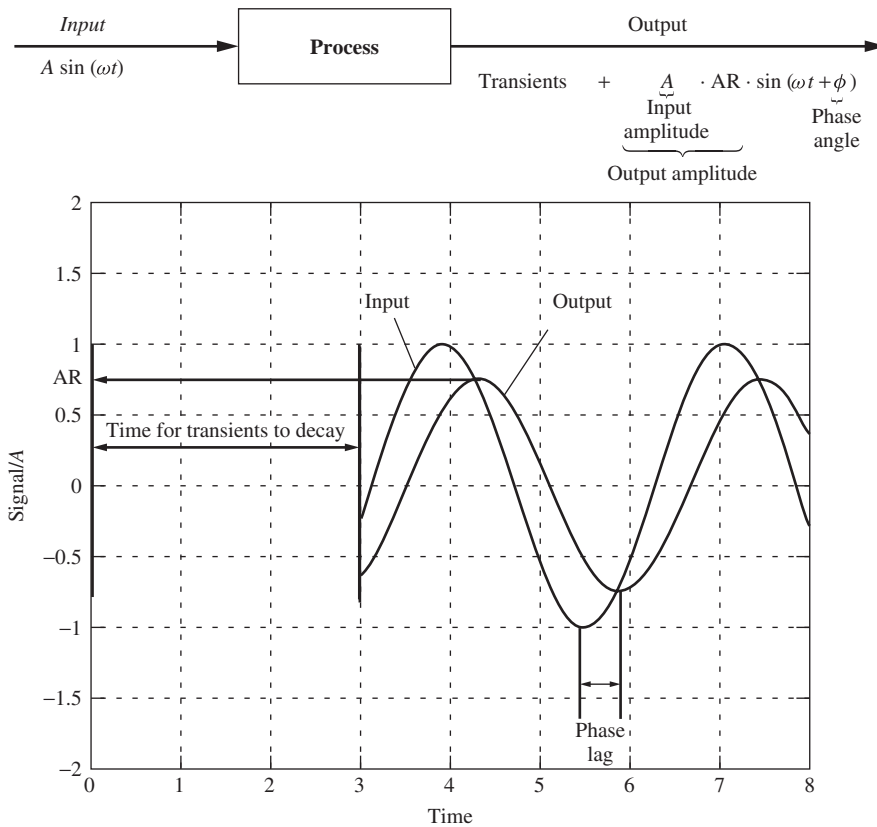

FIGURE 15–2
Characteristics of a steady-state sinusoidal response.

KEY FEATURES TO NOTE ABOUT THE FREQUENCY RESPONSE OF THE PROCESS

- After transients die out, the output is a sine wave.
- Input frequency = output frequency = ω.
- In general, the output is attenuated, that is, AR < 1.
- The output is shifted in time (it lags the input by the phase angle ϕ).
- Amplitude ratio and phase angle are both functions of frequency.

Example 15.1   Rework Example 4.2. The pertinent transfer function is

$$G(s) = \frac{1}{0.1s + 1}$$

The frequency of the bath-temperature variation is given as $10/\pi$ cycles/min which is equivalent to $\omega = (10 \ \text{cycles}/\pi \ \text{min})(2\pi \ \text{rad/cycle}) = 20 \ \text{rad/min}$. Hence, let

$$s = \omega j = 20 j$$

to obtain

$$G(20j) = \frac{1}{2j + 1}$$

In polar form, this is

$$G(20j) = \frac{1}{\sqrt{5}} \measuredangle - 1.11 \ \text{rad} = \frac{1}{\sqrt{5}} \measuredangle - 63.5°$$

which agrees with the previous result.

MATLAB Solution of Example 15.1 Using Simulink

The Simulink model of the process is shown in Fig. 15–3a. Once this Simulink model is run, we can use MATLAB to plot the process output and input as follows.

```
>> plot(temperature.time,temperature.signals.values)
```

The resulting output of this MATLAB command is shown in Fig. 15–3b.

   The tabular output can be examined to verify the actual peak locations so that we can check our hand-calculated results from Example 15.1.
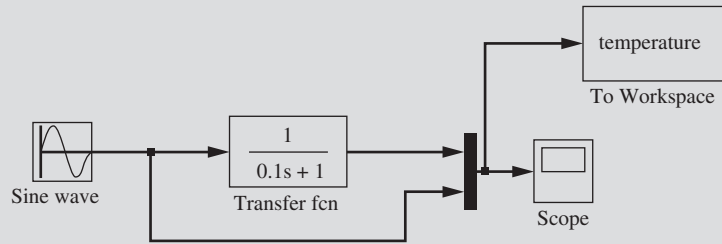
(*a*) Simulink model of first-order process for Example 15.1
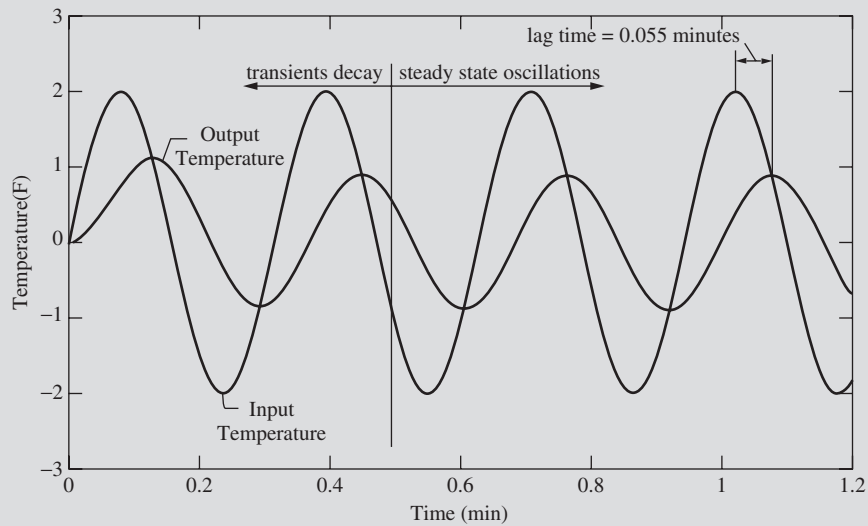
(*b*) MATLAB generated response for first-order system of Example 15.1.

FIGURE 15–3
Using MATLAB and Simulink to solve Example 15.1.

```
>> x=[temperature.time,temperature.signals.values]

x =
  [time] [output] [input]
      0        0        0
    ...      ...      ...      allow transients to decay...pick up data at
                              t=1 min, (from graph)
  1.0200   0.3837   1.9996← input peak
  1.0250   0.4624   1.9937
  1.0300   0.5365   1.9678
  1.0350   0.6053   1.9223
  1.0400   0.6680   1.8576
```

(continues)

```
1.0450  0.7241  1.7743
1.0500  0.7729  1.6733
1.0550  0.8140  1.5556
1.0600  0.8469  1.4223
1.0650  0.8714  1.2748
1.0700  0.8872  1.1146
1.0750  0.8941  0.9433← output peak
1.0800  0.8921  0.7625
1.0850  0.8812  0.5741
1.0900  0.8614  0.3800
1.0950  0.8331  0.1820
1.1000  0.7964 -0.0177
```

Using the MATLAB output to check our previous results, we get

$$\text{Amplitude ratio} = \frac{0.8941}{1.9996} = 0.447 \quad \text{compare to} \quad \frac{1}{\sqrt{5}} = 0.447 \text{ (check)}$$

$$\text{Phase lag} = \left(20\frac{\text{rad}}{\text{min}}\right)(0.055 \text{ min}) = 1.11 \text{ rad} = 63.6° \quad \text{(check)}$$

The procedure we have just discussed is general and can be used for other transfer functions with the important restriction that *it applies only to systems whose transfer functions yield stable responses.* A stable response is necessary so that the transients die out and we will be left with an oscillating, sinusoidal response. We prove the general result in the appendix at the end of this chapter for the reader who is interested.

Example 15.2   Find the frequency response of the system with the general second-order transfer function, and compare the results with those of Chap. 7. The transfer function is

$$\frac{1}{\tau^2 s^2 + 2\zeta\tau s + 1}$$

Putting $s = j\omega$ yields

$$\frac{1}{1 - \tau^2\omega^2 + j2\zeta\omega\tau}$$

which may be converted to polar form

$$\frac{1}{\sqrt{(1 - \omega^2\tau^2)^2 + (2\zeta\omega\tau)^2}} \angle \tan^{-1}\left(\frac{-2\zeta\omega\tau}{1 - \omega^2\tau^2}\right)$$

Hence,

$$\text{Amplitude ratio AR} = \frac{1}{\sqrt{\left[1 - (\omega\tau)^2\right]^2 + (2\zeta\omega\tau)^2}} \qquad (15.4)$$

$$\text{Phase angle} = \phi = -\tan^{-1}\frac{2\zeta\omega\tau}{1 - (\omega\tau)^2}$$

which agree with Eq. (7.41).

Example 15.3. Consider a second-order transfer function, with $\tau = 1$ and $\zeta = 0.8$, being disturbed with a sine wave input of $3\sin(0.5t)$ (Fig. 15–4). Determine the form of the response after the transients have decayed and steady-state oscillations are established.

$$3\sin(0.5t) \longrightarrow \boxed{\dfrac{1}{s^2 + 1.6s + 1}} \longrightarrow$$
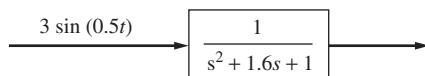
FIGURE 15–4
Block diagram for Example 15.3.

The steady-state oscillations will have the form $3(\text{AR})\sin(0.5t + \phi)$, which is an attenuated (smaller-amplitude) sine wave of the same frequency as the input and shifted by a phase angle ϕ. Thus, all we need to determine is AR and ϕ. We can calculate both from Eq. (15.4).

$$\text{AR} = \frac{1}{\sqrt{\left\{1 - [0.5(1)]^2\right\}^2 + [(2)(0.8)(0.5)(1)]^2}} = \frac{1}{\sqrt{0.5625 + 0.64}} = 0.91$$

$$\phi = -\tan^{-1}\left\{\frac{(2)(0.8)(0.5)(1)}{1 - [0.5(1)]^2}\right\} = -\tan^{-1}\left(\frac{0.8}{0.75}\right) = -0.818 \text{ rad} = -46.8^{\circ}$$

The form of the steady-state oscillations is therefore

$$3(\text{AR})\sin(0.5t + \phi) = 2.73\sin(0.5t - 0.818 \text{ rad})$$
$$= 2.73\sin\left[0.5(t - 1.636)\right]$$
$$= 2.73\sin(0.5t - 46.8^{\circ})$$

We can easily verify this response by using MATLAB and Simulink.
The Simulink model for this problem is shown in Fig. 15–5a.
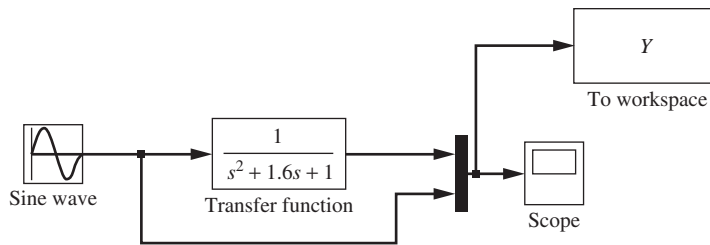
(*a*) Simulink model of first-order process for Example 15.3.
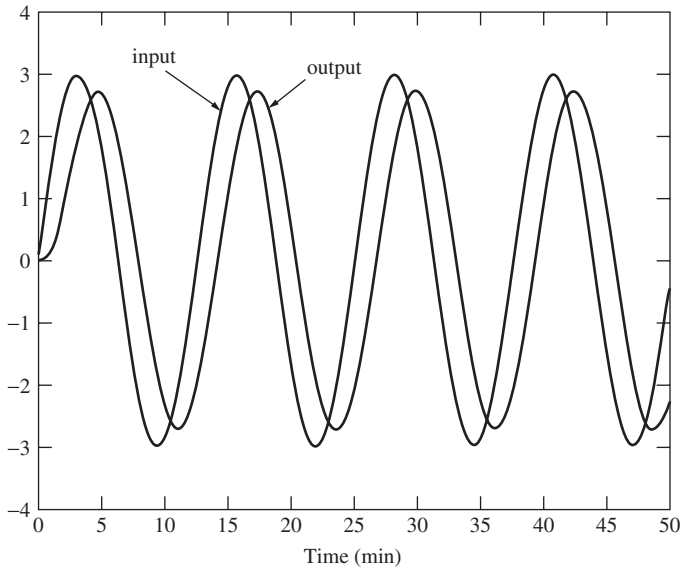
(*b*) MATLAB generated response for first-order system of Example 15.3.

FIGURE 15–5
Using MATLAB and Simulink to solve Example 15.3.

```
Once this Simulink model is run, we can use MATLAB to plot
the process output and input as follows:
>> plot(Y.time,Y.signals.values)
The resulting output of this MATLAB command is shown in
Fig. 15-5b.
>> x = [Y.time,Y.signals.values]
x =
     0          0          0
   0.1495     0.0008     0.2241
   0.2991     0.0059     0.4469
    ...        ...        ...      allow transients to decay
  26.6008    -0.2285     2.0096
  27.3739     0.8162     2.7011
  28.1273     1.7195     2.9919  ← input peak
```

```
28.7632    2.2986    2.9108
29.3446    2.6274    2.5806
29.9062    2.7358    2.0556 ← output peak
30.4614    2.6324    1.3781
31.0193    2.3254    0.5911
31.5891    1.8266   -0.2594
32.1830    1.1507   -1.1226
 ...        ...       ...
47.4619   -2.1802   -2.9573
48.0523   -2.5667   -2.6825
48.6172   -2.7289   -2.2019
49.1731   -2.6774   -1.5583
49.7295   -2.4201   -0.7944
50.0000   -2.2260   -0.3971
```

Transportation Lag

The response of a transportation lag is not described by a standard nth-order differential equation (that yields standard transfer functions). Rather, a transportation lag is described by the relation

$$Y(t) = X(t - \tau) \tag{15.5}$$

which states that the output Y lags the input X by an interval of time τ. If X is sinusoidal

$$X = A \sin \omega t$$

then from Eq. (15.5)

$$Y = A \sin \omega (t - \tau) = A \sin (\omega t - \omega \tau)$$

It is apparent that the AR is unity and the phase angle is $(-\omega\tau)$. To check the substitution rule of the previous section, recall that the transfer function for a transportation lag is given by

$$G(s) = \frac{Y(s)}{X(s)} = e^{-\tau s}$$

Putting $s = j\omega$ gives

$$G(j\omega) = e^{-j\omega\tau} = \left(\text{form of: } Re^{j\phi}, \ R = 1, \ \phi = -\omega\tau\right)$$

Then

$$\text{AR} = \left| e^{-j\omega\tau} \right| = 1 \tag{15.6}$$

$$\text{Phase angle} = \angle e^{-j\omega\tau} = -\omega\tau$$

and the validity of the rule is verified.

Example 15.4.   Consider a stirred-tank heater with a capacity of 15 gal. Water is entering and leaving the tank at the constant rate of 600 lb/min. The heated water that leaves the tank enters a well-insulated section of 6-in-ID pipe. Two feet from the tank, a thermocouple is placed in this line for recording the tank temperature, as shown in Fig. 15–6. The electrical heat input is held constant at 1000 kW.
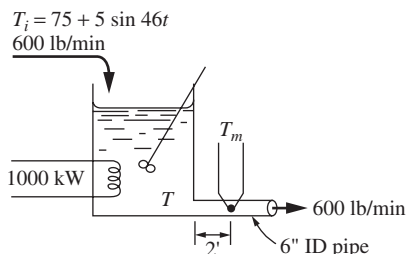
$T_i = 75 + 5 \sin 46t$
600 lb/min

FIGURE 15–6
Tank temperature system for Example 15.4.

If the inlet temperature is varied according to the relation

$$T_i = 75 + 5 \sin 46t$$

where T_i is in degrees Fahrenheit and t is in minutes, find the eventual behavior of the thermocouple reading T_m. Compare this with the behavior of the tank temperature T. It may be assumed that the thermocouple has a very small time constant and effectively measures the true fluid temperature at all times.

The problem is to find the frequency response of T_m to T_i. Deviation variables must be used. Define the deviation variable T_i' as

$$T_i' = T_i - 75 = 5 \sin 46t$$

To define a deviation variable for T_m, note that if T_i were held at 75°F, T_m would come to the steady state satisfying

$$q_s = wC\left(T_{m_s} - T_{i_s}\right)$$

This may be solved for T_{m_s}:

$$T_{m_s} = \frac{q_s}{wC} + T_{i_s} = \frac{(1000 \text{ kW})\left(1000 \dfrac{\text{W}}{\text{kW}}\right)\left(0.0569 \dfrac{\text{Btu/min}}{\text{W}}\right)}{\left(600 \dfrac{\text{lb}}{\text{min}}\right)\left(1.0 \dfrac{\text{Btu}}{\text{lb} \cdot {}^\circ\text{F}}\right)} + 75°\text{F} = 170°\text{F}$$

Hence, define a deviation variable T_m' as

$$T_m' = T_m - 170$$

Now, the overall system between T_i' and T_m' is made up of two components in series: the tank and the 2-ft section of pipe. The transfer function for the tank is

$$G_1(s) = \frac{1}{\tau_1 s + 1}$$

where, as we have seen before, τ_1 is given by

$$\tau_1 = \frac{\rho V}{w} = \frac{\left(60.8 \, \dfrac{\text{lb}}{\text{ft}^3}\right)(15 \text{ gal})}{\left(600 \, \dfrac{\text{lb}}{\text{min}}\right)\left(7.48 \, \dfrac{\text{gal}}{\text{ft}^3}\right)} = 0.203 \text{ min}$$

The transfer function of the 2-ft section of pipe, which corresponds to a transportation lag, is

$$G_2(s) = e^{-\tau_2 s}$$

where τ_2 is the length of time required for the fluid to traverse the length of pipe. This is

$$\tau_2 = \frac{L}{v} = \frac{2 \text{ ft}}{\dfrac{600 \text{ lb/min}}{\left(60.8 \text{ lb/ft}^3\right)\left(0.196 \text{ ft}^2\right)}} = 0.0397 \text{ min}$$

The factor 0.196 is the cross-sectional area of the pipe in square feet.

Since the two systems are in series, the overall transfer function between T_i' and T_m' is

$$\frac{T_m'}{T_i'} = \frac{e^{-\tau_2 s}}{\tau_1 s + 1} = \frac{e^{-0.0397 s}}{0.203 s + 1}$$

Notice that we must use the same time units throughout, minutes in this case. We are now ready to construct the block diagram for the process, which is shown in Fig. 15–7.
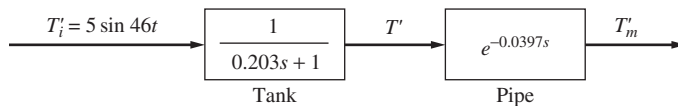
FIGURE 15–7
Block diagram for Example 15.4.

To find the AR and phase lag, we merely substitute $s = j\omega = 46j$ and find the magnitude and argument of the resulting complex number. However, note that we have previously derived the individual frequency responses for the first-order system and transportation lag. The overall transfer function is the product of the individual transfer functions; hence, its magnitude will be the product of the magnitudes and its argument the sum of the arguments of the individual transfer

functions. In general, if we have several transfer functions in series, the overall transfer function is

$$G(s) = G_1(s)G_2(s)\cdots G_n(s)$$

and substituting $s = j\omega$, we obtain

$$G(j\omega) = G_1(j\omega)G_2(j\omega)\cdots G_n(j\omega)$$

or, in polar form,

$$Re^{j\phi} = R_1 e^{j\phi_1} \cdot R_2 e^{j\phi_2} \cdots R_n e^{j\phi_n} = \underbrace{R_1\,R_2\cdots R_n}_{\substack{\text{product of individual} \\ \text{magnitudes}}}\; e^{\underbrace{j(\phi_1+\phi_2+\cdots+\phi_n)}_{\substack{\text{sum of individual} \\ \text{phase angles}}}}$$

then

$$\left|G(j\omega)\right| = \left|G_1(j\omega)\right|\left|G_2(j\omega)\right|\cdots\left|G_n(j\omega)\right|$$

$$\measuredangle G(j\omega) = \measuredangle G_1(j\omega) + \measuredangle G_2(j\omega) + \cdots + \measuredangle G_n(j\omega)$$

This rule makes it very convenient to find the frequency response of a number of systems in series.

Using Eq. (15.3) for the tank,

$$AR = \frac{1}{\sqrt{(46 \times 0.203)^2 + 1}} = \frac{1}{9.39} = 0.107$$

$$\text{Phase angle} = \tan^{-1}\left[(-46)(0.203)\right] = -84°$$

For the section of pipe, AR is unity, so that the overall AR is just 0.107. The phase lag due to the pipe may be obtained from Eq. (15.6) as

$$\text{Phase angle} = -\omega\tau_2 = -(46)(0.0397) = -1.82 \text{ rad} = -104°$$

The overall phase lag from T_i' to T_m' is the sum of the individual lags,

$$\measuredangle \frac{T_m'}{T_i'} = -84 - 104 = -188°$$

Hence

$$T_m = 170 + 5(0.107)\sin(46t - 188°)$$

For comparison, a plot of T_i', T_m', and T' is given in Fig. 15–8, where

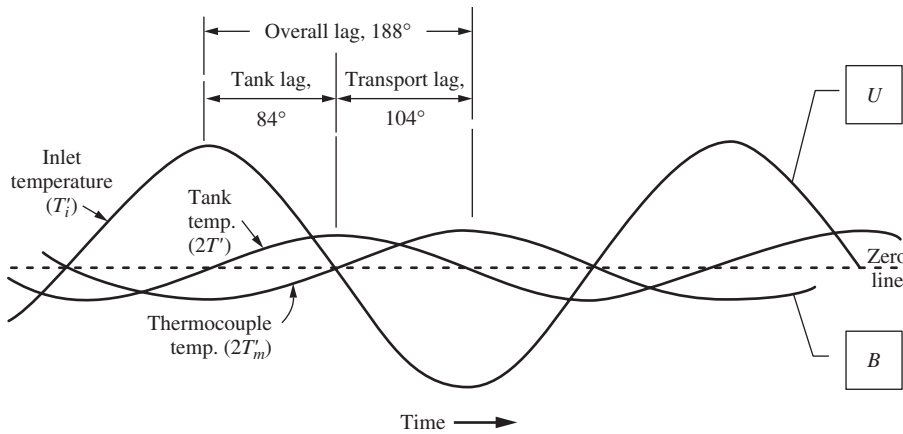$$T' = \text{tank temperature} - 170°F$$

FIGURE 15–8
Temperature variation in Example 15.4.

These temperatures are shown on a block diagram in Fig. 15–9.

It should be emphasized that this plot applies only after sufficient time has elapsed for the complementary solution to become negligible (i.e., the transients have decayed). This restriction applies to all the forthcoming work on frequency response. Also, note that, for convenience of scale (so that we easily can see them on the same axes), the tank and thermocouple temperatures have been plotted as $2T'$ and $2T'_m$, respectively.
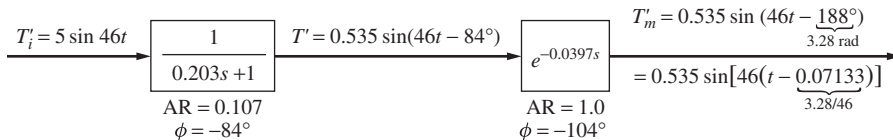
FIGURE 15–9
Response summary for Example 15.4.

A Control Problem

An interesting conclusion may be reached from a study of Fig. 15–8. Suppose that we are trying to control the tank temperature, using the deviation between the thermocouple reading and the set point as the error. A block diagram for proportional control might appear as in Fig. 15–10, where T'_i is replaced by U, T' by C, and T'_m by B to conform with our standard block diagram nomenclature. The variable R denotes the deviation of the set point from 170°F and is the desired value of the deviation C. The value of R is assumed to be zero in the following analysis (control at 170°F). The following arguments, while not rigorous, serve to give some insight regarding the application of frequency response to control system analysis.

FIGURE 15–10
Proportional control of heated, stirred tank.

If $R = 0$, then

$$\text{Error } \varepsilon = 0 - B = -B$$

$$\text{Heat input } Q = K_c \varepsilon = -K_c B$$

Thus, the heat being added to the tank is given in deviation variables as $-K_c B$. With reference to Fig. 15–8, which shows the response of the *uncontrolled* tank to a sinusoidal variation in U, it can be seen that the peaks of U (which is T'_i) an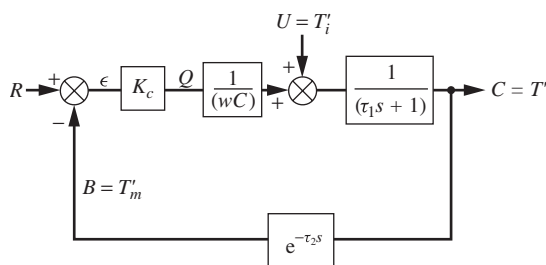d B (which is T'_m) are almost exactly opposite because the phase difference is 188°. This means that if the loop were closed, the control system would have a tendency to add *more* heat when the inlet temperature T_i was at its high peak, because B is then negative and $-K_c B$ becomes positive. (Recall that the set point R is held constant at zero.)

Conversely, when the inlet temperature is at a low point, the *tendency* will be for the control system to add less heat because B is positive. This is precisely opposite to the way the heat input should be controlled. Figure 15–11 clarifies this physical situation.

Therefore, the possibility of an unstable control system exists for this particular sinusoidal variation in frequency. Indeed, we shall demonstrate in Chap. 16 that *if K_c is taken too large,* the tank temperature will oscillate with increasing amplitude for *all* variations in U, and hence we have an unstable control system. The fact that such information may be obtained by study of the *frequency response* (i.e., the particular solution for a sinusoidal forcing function) justifies further study of this subject.

15.2   BODE DIAGRAMS

Thus far, it has been necessary to calculate AR and phase lag by direct substitution of $s = j\omega$ into the transfer function for the particular frequency of interest. It can be seen from Eqs. (15.3), (15.4), and (15.6) that the AR and phase lag are functions of frequency. There is a convenient graphical representation of their dependence on the frequency that largely eliminates direct calculation. *This is called a Bode diagram and consists of two graphs: logarithm of AR versus logarithm of frequency, and phase angle versus logarithm of frequency* (see Fig. 15–12).

The Bode diagram will be shown in Chap. 16 to be a convenient tool for analyzing control problems such as the one discussed in the preceding section. The remainder of this chapter is devoted to developing this tool and presenting Bode diagrams for the basic components of control loops.
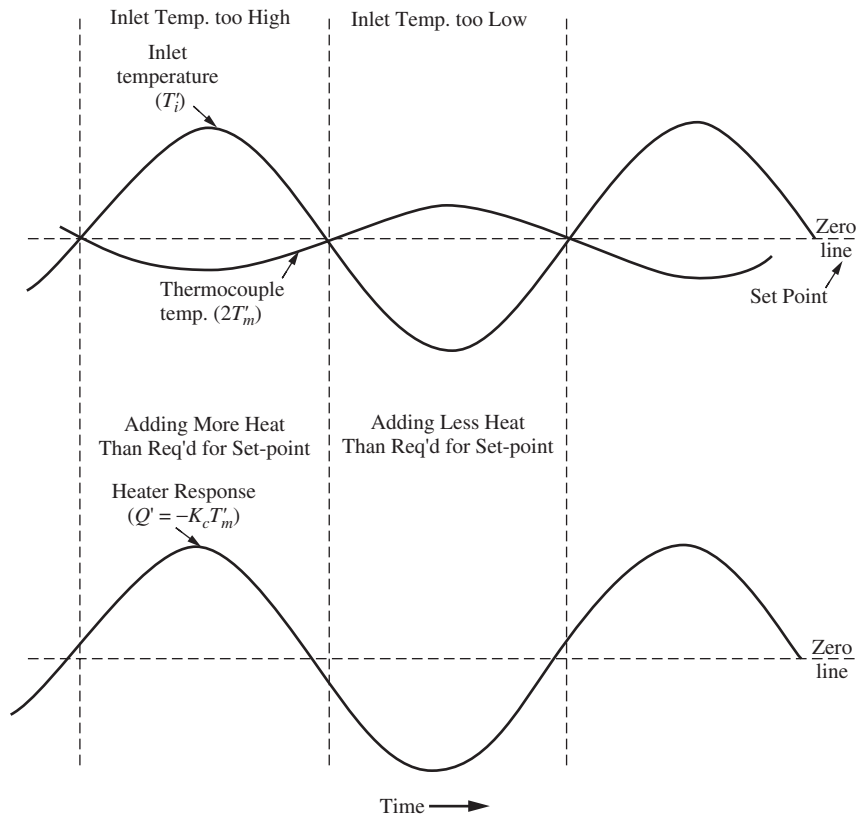
FIGURE 15–11
Illustration of possible unstable response of tank temperature system.
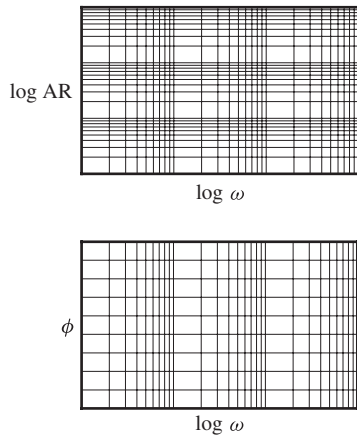
FIGURE 15–12
Bode diagram format.

First-Order System

As we saw earlier [(Eq. (15.3)], the AR and phase angle for the sinusoidal response of a first-order system are

$$AR = \frac{1}{\sqrt{\tau^2 \omega^2 + 1}} \tag{15.7}$$

$$\text{Phase angle} = \tan^{-1}(-\omega\tau) \tag{15.8}$$

It is convenient to regard these as functions of $\omega\tau$ for the purpose of generality. From Eq. (15.7)

$$\log AR = -\tfrac{1}{2}\log\left[(\omega\tau)^2 + 1\right] \tag{15.9}$$

The first part of the Bode diagram is a plot of Eq. (15.9). The true curve is shown as the solid line on the upper part of Fig. 15–13.
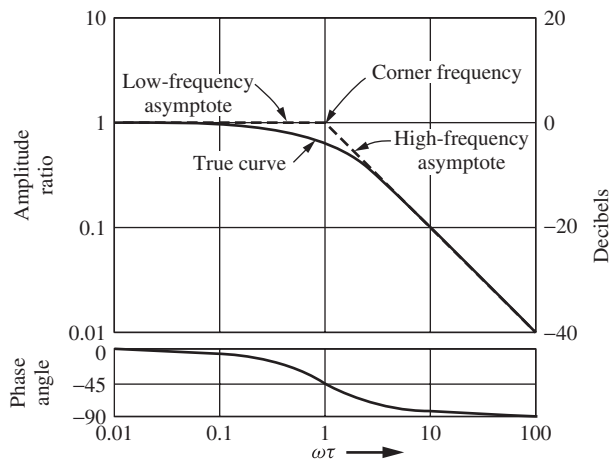
FIGURE 15–13
Bode diagram for first-order system.

Some asymptotic considerations can simplify the construction of this plot. As $\omega\tau \to 0$, Eq. (15.7) shows that $AR \to 1$. This is indicated by the low-frequency asymptote on Fig. 15–13. As $\omega\tau \to \infty$, Eq. (15.9) becomes asymptotic to

$$\log AR = -\log(\omega\tau)$$

which is a line of slope -1, passing through the point

$$\omega\tau = 1 \qquad AR = 1$$

This line is indicated as the high-frequency asymptote in Fig. 15–13. The frequency $\omega_c = 1/\tau$, where the two asymptotes intersect, is known as the *corner frequency;* it may

be shown that the deviation of the true AR curve from the asymptotes is a maximum at the corner frequency. Using $\omega_c = 1/\tau$ in Eq. (15.7) gives

$$\text{AR} = \frac{1}{\sqrt{2}} = 0.707$$

as the true value, whereas the intersection of the asymptotes occurs at AR = 1. Since this is the maximum deviation and is an error of less than 30 percent, for engineering purposes it is often sufficient to represent the curve entirely by the asymptotes. Alternately, the asymptotes and the value of 0.707 may be used to sketch the curve if greater accuracy is required. Computer packages (MATLAB, for example, see below) can also be used to generate the diagrams fairly easily.

In the lower half of Fig. 15–13, we have shown the phase curve as given by Eq. (15.8). Since

$$\phi = \tan^{-1}(-\omega\tau) = -\tan^{-1}(\omega\tau)$$

it is evident that ϕ approaches 0° at low frequencies and $-90°$ at high frequencies. This verifies the low- and high-frequency portions of the phase curve. At the corner frequency, $\omega_c = 1/\tau$,

$$\phi_c = -\tan^{-1}(\omega_c\tau) = -\tan^{-1} 1 = -45°$$

There are asymptotic approximations available for the phase curve, but they are not so accurate or so widely used as those for the AR. Instead, it is convenient to note that the curve is symmetric about $-45°$.

It should be stated that, in a great deal of the literature on control theory, amplitude ratios (or gains) are reported in decibels. The decibel (dB) is defined by

$$\text{Decibels} = 20 \log\left(\text{AR}\right)$$

Thus, an AR of unity corresponds to 0 dB, and an amplitude ratio of 0.1 corresponds to -20 dB. The value of the AR in decibels is given on the right-hand ordinate of Fig. 15–13.

Example 15.5.   Using MATLAB to Generate a Bode Diagram for a First-Order System
   To generate a Bode diagram, we must first form the transfer function in MATLAB.

```
>> num= [1];
>> den= [10 1];
>> sys= tf (num,den);%this statement generates a transfer
function assigned to sys
Transfer function:
    1
  _____

10 s + 1
>> bode(sys);%the "bode" command generates the Bode diagram
shown in Fig. 15-14
```
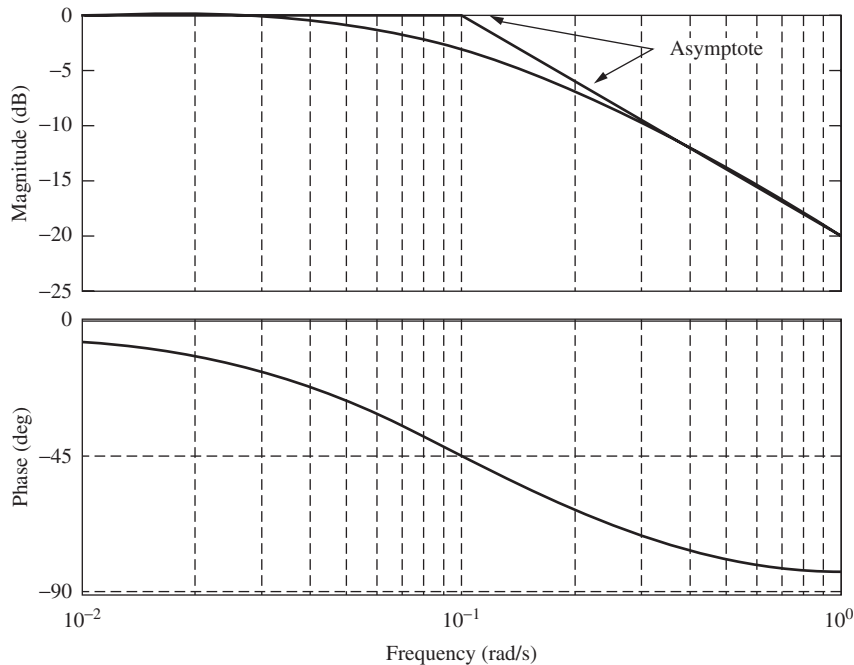
FIGURE 15–14
Bode diagram for first-order system in Example 15.5.

A drawback of this built-in Bode command is that the magnitude plot is always generated in decibels. An m-file that will generate a Bode plot without converting to decibels is shown below as function m-file named mybode.m.

mybode.m file

```
function mybode(sys);
[mag,phase,w]=bode(sys);
figure
subplot(2,1,1)
loglog(w,squeeze(mag))
grid
ylabel('amplitude')
xlabel('frequency(rad/time)')
subplot(2,1,2)
semilogx(w,squeeze(phase))
grid
ylabel('phase(deg)')
xlabel('frequency (rad/time)')
```

The result of using mybode on the current system is shown in Fig. 15–15.

Note that in both Figs. 15–14 and 15–15 the asymptotes were added after the magnitude graph was generated to illustrate the slopes of 0 and −1 for the low- and high-frequency asymptotes, respectively.
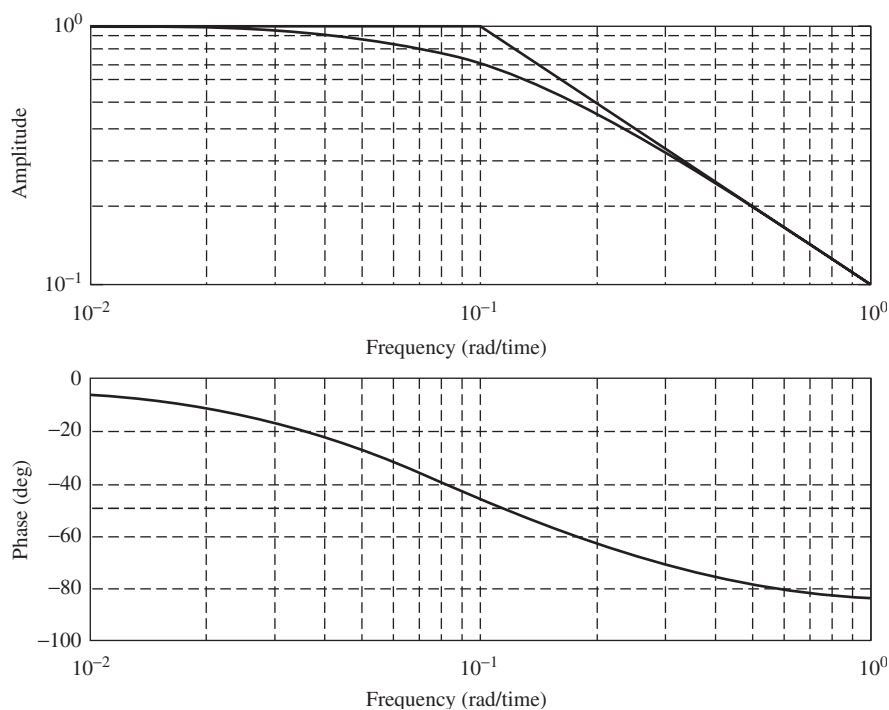
FIGURE 15–15
Bode diagram for Example 15.5 generated using `mybode` m-file and MATLAB.

First-Order Systems in Series

The advantages of the Bode plot become evident when we wish to plot the frequency response of systems in series. As shown in Example 15.4, the rules for multiplication of complex numbers indicate that the AR for two first-order systems in series is the product of the individual ARs:

$$\text{AR} = \frac{1}{\sqrt{\omega^2 \tau_1^2 + 1}\sqrt{\omega^2 \tau_2^2 + 1}} \tag{15.10}$$

Similarly, the phase angle is the sum of the individual phase angles

$$\phi = \tan^{-1}(-\omega\tau_1) + \tan^{-1}(-\omega\tau_2) \tag{15.11}$$

Since the AR is plotted on a logarithmic basis, multiplication of the ARs is accomplished by addition of logarithms on the Bode diagram (which, we shall see, is equivalent to adding the slopes of the asymptotes of the individual curves to get the asymptote of the overall curve on log-log coordinates). The phase angles are added directly. The procedure is best illustrated by an example.

Example 15.6. Plot the Bode diagram for the system whose overall transfer function is

$$\frac{1}{(s + 1)(s + 5)}$$

To put this in the form of two first-order systems in series, it is rewritten as

$$\frac{\frac{1}{5}}{(s + 1)(\frac{1}{5}s + 1)} \tag{15.12}$$

The time constants are $\tau_1 = 1$ and $\tau_2 = \frac{1}{5}$. The factor $\frac{1}{5}$ in the numerator corresponds to the steady-state gain.

From Eqs. (15.12) and (15.10)

$$AR_{overall} = \frac{\frac{1}{5}}{\sqrt{\omega^2 + 1}\sqrt{(\omega/5)^2 + 1}}$$

Hence,

$$\log AR_{overall} = \log\frac{1}{5} - \frac{1}{2}\log(\omega^2 + 1) - \frac{1}{2}\log\left[\left(\frac{\omega}{5}\right)^2 + 1\right]$$

or

$$\log AR_{overall} = \log\frac{1}{5} + \log(AR)_1 + \log(AR)_2 \tag{15.13}$$

where $(AR)_1$ and $(AR)_2$ are the ARs of the individual first-order systems, each with unity gain. Equation (15.13) shows that the overall AR is obtained, on logarithmic coordinates, by adding the individual ARs and a constant corresponding to the steady-state gain.

The individual ARs must be plotted as functions of log ω rather than log $(\omega\tau)$ because of the different time constants. This is easily done by shifting the curves of Fig. 15–13 to the right or left so that the corner frequency falls at $\omega = 1/\tau$. Thus, the individual curves of Fig. 15–16 are placed so that the corner frequencies fall at $\omega_{c_1} = 1$ and $\omega_{c_2} = 5$. These slopes of these curves are added to obtain the overall curve shown. Note that in this case the logarithms are negative and the addition is downward. To complete the AR curve, the factor log $\frac{1}{5}$ should be added to the overall curve. This would have the effect of shifting the entire curve down by a constant amount. Instead of doing this, the factor $\frac{1}{5}$ is incorporated by plotting the overall curve as $AR_{overall}/\left(\frac{1}{5}\right)$ instead of $AR_{overall}$. This procedure is usually more convenient.

Asymptotes have also been indicated on Fig. 15–16. The sum of the slopes of the individual asymptotes gives the slope of the overall asymptote, which is seen to be a good approximation to the overall curve. The overall asymptote has

a slope of 0 below $\omega = 1$, -1 for ω between 1 and 5, and -2 above $\omega = 5$. Its slope is obtained by simply adding the slopes of the individual asymptotes.
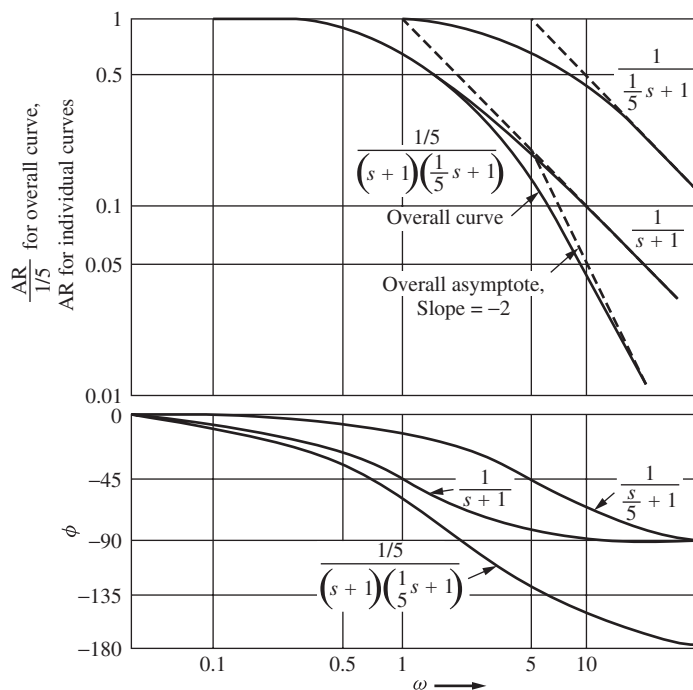
FIGURE 15–16
Bode diagram for $0.2/[(s + 1)(0.2s + 1)]$.

To obtain the phase angle, the individual phase angles are plotted and added according to Eq. (15.11). The factor $\frac{1}{5}$ has no effect on the phase angle, which approaches $-180°$ at high frequency.

Graphical Rules for Bode Diagrams

Before we proceed to a development of the Bode diagram for other systems, it is desirable to summarize the graphical rules that were utilized in Example 15.5.

Consider a number of systems in series. As shown in Examples 15.4 and 15.5, the overall AR is the product of the individual ARs, and the overall phase angle is the sum of the individual phase angles. Therefore,

$$\log(AR_{overall}) = \log(AR)_1 + \log(AR)_2 + \cdots + \log(AR)_n \quad (15.14)$$

and

$$\phi_{overall} = \phi_1 + \phi_2 + \cdots + \phi_n$$

where n is the total number of systems. Therefore, the following rules apply to the true curves or to the asymptotes on the Bode diagram:

1. The overall AR is obtained by adding the individual ARs. For this graphical addition, an individual AR that is above unity on the frequency response diagram is taken as positive; an AR that is below unity is taken as negative. To understand this, recall that the logarithm of a number greater than 1 is positive and the logarithm of a number less than 1 is negative. If we are plotting the asymptotes of the individual curves, the asymptote of the overall curve is obtained by adding the slopes of the individual curves.
2. The overall phase angle is obtained by addition of the individual phase angles.
3. The presence of a constant in the overall transfer function shifts the entire AR curve vertically by a constant amount and has no effect on the phase angle. It is usually more convenient to include a constant factor in the definition of the ordinate.

These rules will be of considerable value in later examples. Let us now proceed to develop Bode diagrams for other control system components.

Second-Order System

As shown in Example 15.2, the frequency response of a system with a second-order transfer function

$$G(s) = \frac{1}{\tau^2 s^2 + 2\zeta \tau s + 1}$$

is given by Eq. (15.4), repeated here for convenience,

$$\text{AR} = \frac{1}{\sqrt{\left(1 - \omega^2 \tau^2\right)^2 + (2\zeta\omega\tau)^2}} \qquad (15.4)$$

$$\text{Phase angle} = \tan^{-1}\frac{-2\zeta\omega\tau}{1 - (\omega\tau)^2}$$

If $\omega\tau$ is used as the abscissa for the general Bode diagram, it is clear that ζ will be a parameter. That is, there is a different curve for each value of ζ. These curves appear as in Fig. 15–17.

The calculation of phase angle as a function of ω from Eq. (15.4) requires careful attention. The calculation can be done most clearly with the aid of a plot of $\tan^{-1} x$ (or arctan x), as shown in Fig. 15–18.

As $\omega\tau$ goes from 0 to 1, we see from Eq. (15.4) that the argument of the arctan function goes from 0 to $-\infty$ and the phase angle goes from 0 to $-90°$ as shown by the branch from A to B in Fig. 15–18. As $\omega\tau$ crosses unity from a value less than unity to a value greater than unity, the sign of the argument of the arctan function in Eq. (15.4) shifts from negative to positive. To preserve continuity in angle as $\omega\tau$ crosses unity, the phase angle must go from -90 to $-180°$ as $\omega\tau$ goes from 1 to $+\infty$ and the branch of the arctan function goes from C to D (in Fig. 15–18).
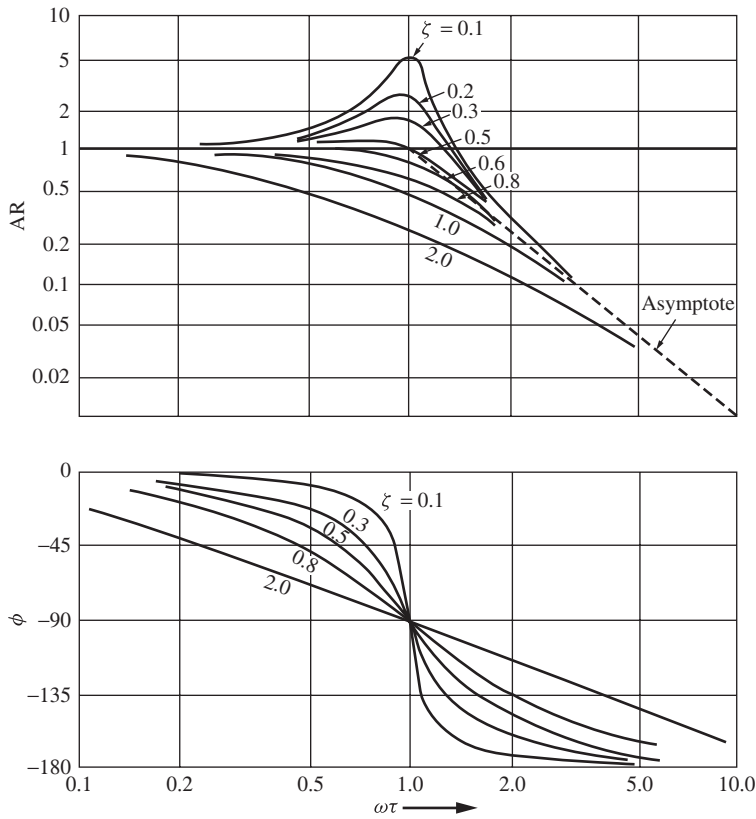
FIGURE 15–17
Block diagram for second-order system $\dfrac{1}{\tau^2 s^2 + 2\zeta\tau s + 1}$.

The arctan function available in calculators and computers normally covers the principal branches of the arctan function, shown as BAE in Fig. 15–18. For this reason, one must be very careful in calculating the phase angle with Eq. (15.4). If a calculator programmed for the principal branches of the arctan function is used and the argument is positive, one obtains the correct phase angle by subtracting 180° from the answer given by the calculator. Notice that for $\omega\tau = 1$, the phase angle is $-90°$, independent of ζ. This verifies that all phase curves intersect at $-90°$, as shown in Fig. 15–17.

We may now examine the amplitude curves obtained from Eq. (15.4). For $\omega\tau \ll 1$, the AR, or gain, approaches unity. For $\omega\tau \to \infty$, the only significant term in the denominator is $\sqrt{(\omega\tau)^4}$, and the AR becomes asymptotic to the line

$$AR = \frac{1}{(\omega\tau)^2}$$
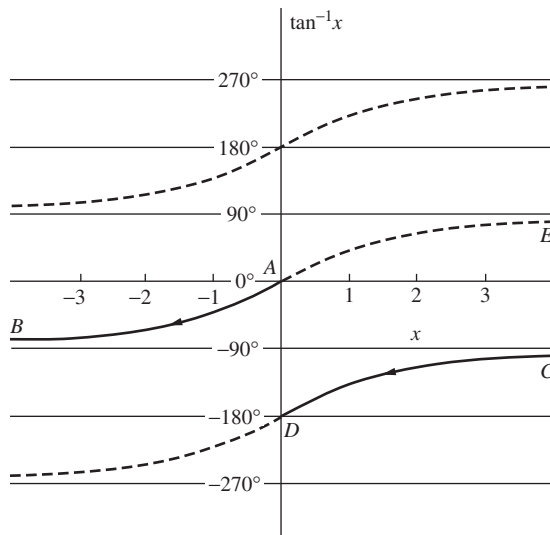
FIGURE 15–18
Use of plot of $\tan^{-1} x$ for computing phase angle of second-order system.

This asymptote has slope -2 and intersects the line AR $= 1$ at $\omega\tau = 1$. The asymptotic lines are indicated on Fig. 15–17. For $\zeta \geq 1$, we have shown that the second-order system is equivalent to two first-order systems in series. The fact that the AR for $\zeta \geq 1$ (as well as for $\zeta < 1$) attains a slope of -2 and phase of $-180°$ is, therefore, consistent.

Figure 15–17 also shows that, for $\zeta < 0.707$, the AR curves attain maxima in the vicinity of $\omega\tau = 1$. This can be checked by differentiating the expression for the AR with respect to $\omega\tau$ and setting the derivative to zero. The result is

$$(\omega\tau)_{max} = \sqrt{1 - 2\zeta^2} \qquad \zeta < 0.707 \tag{15.15}$$

for the value of $\omega\tau$ at which the maximum AR occurs. The value of the maximum AR, obtained by substituting $(\omega\tau)_{max}$ into Eq. (15.4), is

$$(AR)_{max} = \frac{1}{2\zeta\sqrt{1 - \zeta^2}} \qquad \zeta < 0.707$$

A plot of the maximum AR versus ζ is given in Fig. 15–19. The frequency at which the maximum AR is attained is called the resonant frequency and is obtained from Eq. (15.15),

$$\omega_r = \frac{1}{\tau}\sqrt{1 - 2\zeta^2} \tag{15.16}$$

FIGURE 15–19
Maximum AR versus damping for a second-order system.

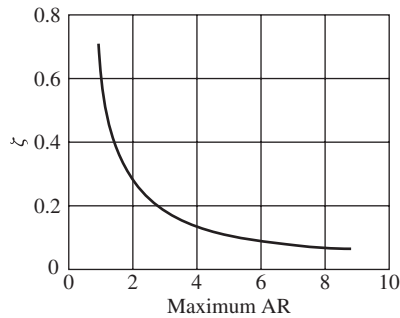The phenomenon of resonance is frequently observed in our everyday experience. A vase may vibrate when the stereo is playing a particular note. As a car decelerates, perceptible vibrations may occur at particular speeds. A suspension bridge oscillates violently when soldiers march across, stepping at a certain cadence.

It may be seen that AR values exceeding unity are attained by systems for which $\zeta < 0.707$. This is in sharp contrast to the first-order system, for which the AR is always less than unity.

The curves of Fig. 15–17 for $\zeta < 1$ are not simple to construct, particularly in the vicinity of the resonant frequency. Fortunately, almost all second-order control system components for which we will want to construct Bode diagrams have $\zeta > 1$. That is, they are composed of two first-order systems in series. Actually, the curves of Fig. 15–17 are presented primarily because they are useful in analyzing the *closed-loop frequency response* of many control systems.

Transportation Lag

As shown by Eq. (15.6), the frequency response for $G(s) = e^{-\tau s}$ is

$$\text{AR} = 1$$
$$\phi = -\omega\tau \quad \text{rad} \qquad \text{or} \qquad \phi = -57.2958\,\omega\tau \quad \text{deg}$$

In this expression, ω is in radians and 57.2958 is the number of degrees in 1 rad. There is no need to plot the AR since it is constant at 1.0. On logarithmic coordinates, the phase angle appears as in Fig. 15–20, where $\omega\tau$ is used as the abscissa to make the figure general. The transportation lag contributes a phase lag, which increases without bound as ω increases. Note that it is necessary to convert $\omega\tau$ from radians to degrees to prepare Fig. 15–20.

Proportional Controller

A proportional controller with transfer function K_c has amplitude ratio K_c and phase angle zero at all frequencies. No Bode diagram is necessary for this component.
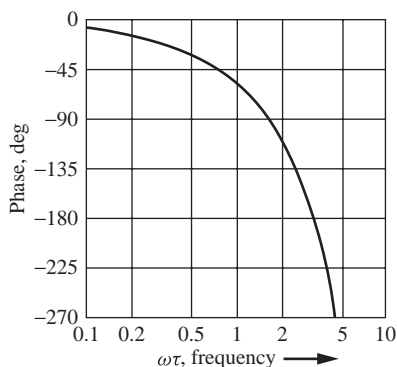
FIGURE 15–20
Phase characteristics of transportation lag.

Proportional-Integral Controller

This component has the ideal transfer function

$$G(s) = K_c\left(1 + \frac{1}{\tau_I s}\right)$$

Accordingly, the frequency response is given by

$$AR = |G(j\omega)| = K_c\left|1 + \frac{1}{\tau_I j\omega}\right| = K_c\sqrt{1 + \frac{1}{(\omega\tau_I)^2}}$$

$$Phase = \angle G(j\omega) = \angle\left(1 + \frac{1}{\tau_I j\omega}\right) = \tan^{-1}\left(-\frac{1}{\omega\tau_I}\right)$$

The Bode plot of Fig. 15–21 uses $\omega\tau_I$ as the abscissa. The constant factor K_c is included in the ordinate for convenience. Asymptotes with a corner frequency of $\omega_c = 1/\tau_I$ are indicated. The verification of Fig. 15–21 is recommended as an exercise for the reader.

Proportional-Derivative Controller

The transfer function is

$$G(s) = K_c(1 + \tau_D s)$$

The reader should show that this has amplitude and phase behavior that are just the inverse of the first-order system

$$\frac{1}{\tau s + 1}$$

Hence, the Bode plot is as shown in Fig. 15–22. The corner frequency is $\omega_c = 1/\tau_D$.

FIGURE 15–21
Bode diagram for PI controller.
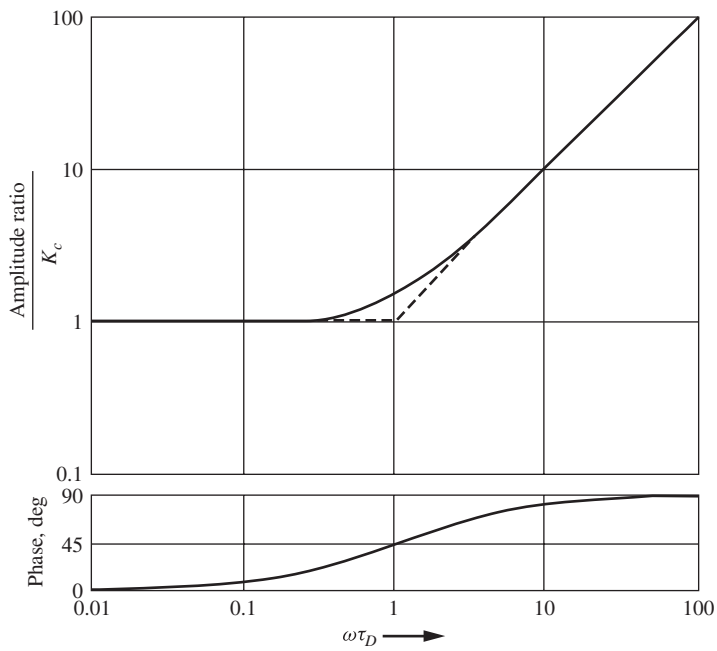
FIGURE 15–22
Bode diagram for PD controller.

This system is important because it introduces *phase lead.* Thus, it can be seen that using PD control for the tank temperature control system of Example 15.4 would decrease the phase lag at all frequencies. In particular, 180° of phase lag would not occur until a higher frequency. Thi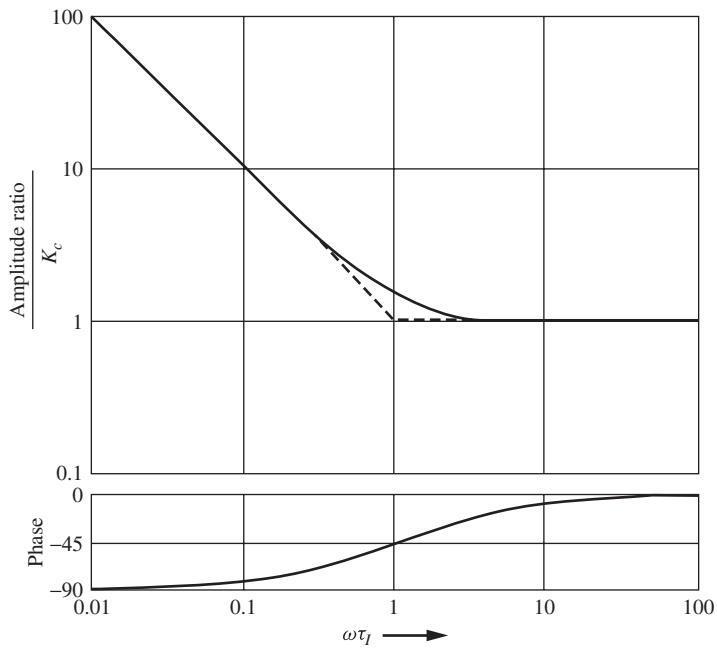s may exert a stabilizing influence on the control system. In Chap. 16 we look in detail at designing stabilizing controllers using Bode diagram analysis. It is appropriate to conclude this chapter with a summarizing example.

Example 15.7   Plot the Bode diagram for the open-loop transfer function of the control system of Fig. 15–23. This system might represent PD control of three tanks in series, with a transportation lag in the measuring element.
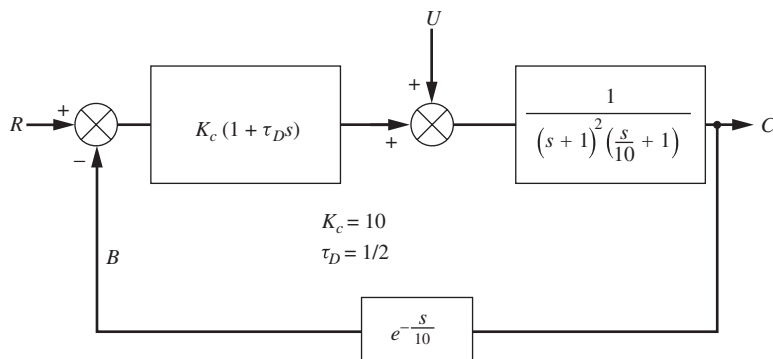
FIGURE 15–23
Block diagram of control system for Example 15.7.

The open-loop transfer function is

$$G(s) = \frac{10(0.5s + 1)e^{-s/10}}{(s + 1)^2(0.1s + 1)}$$

The individual components are plotted as dashed lines in Fig. 15–24. Only the asymptotes are used on the AR portion of the graph. Here it is easiest to plot the factor $(s + 1)^{-2}$ as a line of slope -2 through the corner frequency of 1. For the phase-angle graph, the factor $(s + 1)^{-1}$ is plotted and added in twice to form the overall curve. The overall curves are obtained by the graphical rules previously presented. For comparison, the overall curves obtained without derivative action (i.e., by not adding in the curves corresponding to $0.5s + 1$) are also shown. Note that on the asymptotic AR diagram, the slopes of the individual curves are added to obtain the slope of the overall curve.
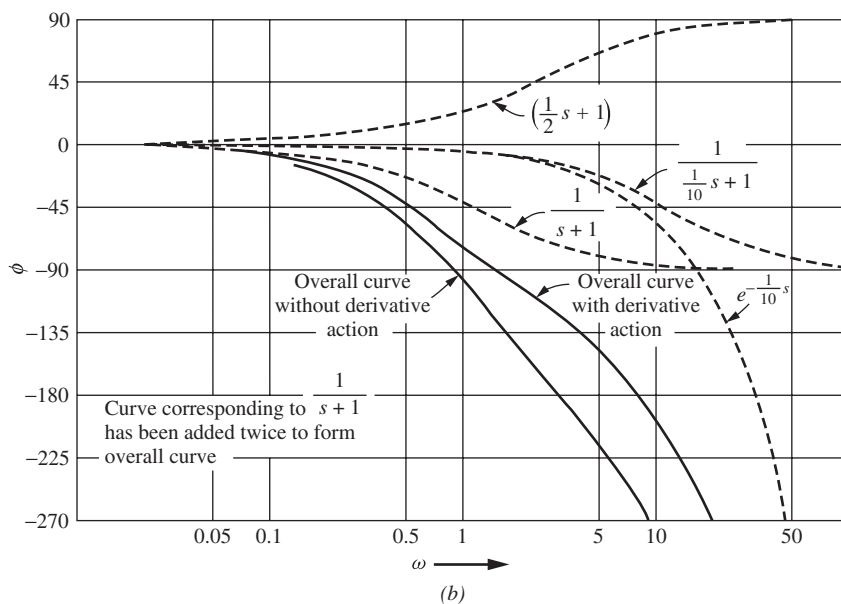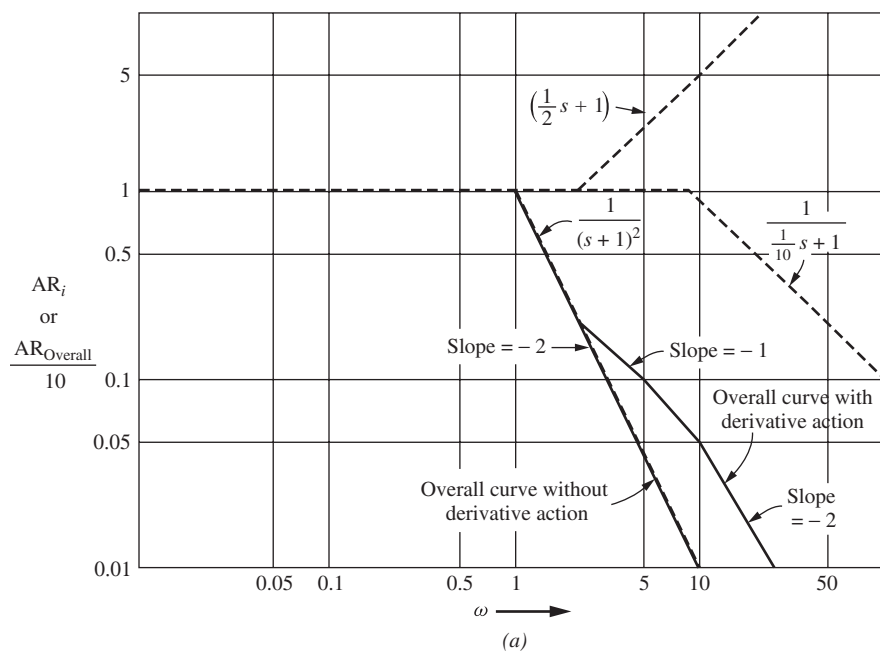
FIGURE 15–24
Block diagram for Example 15.7: (a) Amplitude ratio; (b) phase angle.

15.3 APPENDIX—GENERALIZATION OF SUBSTITUTION RULE

The substitution rule ($j\omega$ for s) is valid for transfer functions that yield stable systems. An nth-order linear system is characterized by an nth-order differential equation

$$a_n \frac{d^n Y}{dt^n} + a_{n-1} \frac{d^{n-1} Y}{dt^{n-1}} + \cdots + a_1 \frac{dY}{dt} + a_0 Y = X(t) \qquad (15.17)$$

where Y is the output variable and $X(t)$ is the forcing function or input variable. For specific cases of Eq. (15.17), refer to Eq. (4.5) for a first-order system and Eq. (7.5) for a second-order system. If $X(t)$ is sinusoidal

$$X(t) = A \sin \omega t$$

the solution of Eq. (15.17) will consist of a complementary solution and a particular solution of the form

$$Y_p(t) = C_1 \sin \omega t + C_2 \cos \omega t \qquad (15.18)$$

If the system is stable, the roots of the characteristic equation of Eq. (15.17) all lie to the left of the imaginary axis, and the complementary solution will vanish exponentially in time. Then Y_p is the quantity previously defined as the sinusoidal or *frequency response*. If the system is not stable, the complementary solution grows exponentially, and the term *frequency response* has no physical significance because $Y_p(t)$ is inconsequential.

The problem now is the evaluation of C_1 and C_2 in Eq. (15.18). Since we are interested in the amplitude and phase of $Y_P(t)$, Eq. (15.18) is rewritten as

$$Y_p = D_1 \sin (\omega t + D_2) \qquad (15.19)$$

as was done previously [compare to Eq. (4.25) and related equations].

It will be convenient to change $X(t)$ and $Y_p(t)$ from trigonometric to exponential form, using the identity

$$\sin \theta = \frac{e^{j\theta} - e^{-j\theta}}{2j}$$

Thus,

$$X(t) = \frac{A}{2j}\left(e^{j\omega t} - e^{-j\omega t}\right) \qquad (15.20)$$

and from Eq. (15.19)

$$Y_p(t) = \frac{D_1}{2j}\left(e^{j(\omega t + D_2)} - e^{-j(\omega t + D_2)}\right) \qquad (15.21)$$

Substitution of Eqs. (15.20) and (15.21) into Eq. (15.17) yields

$$
\frac{D_1 e^{j(\omega t + D_2)}}{2j} \left[a_n (j\omega)^n + a_{n-1}(j\omega)^{n-1} + \cdots + a_1(j\omega) + a_0 \right]
$$

$$
- \frac{D_1 e^{j(\omega t + D_2)}}{2j} \left[a_n (-j\omega)^n + a_{n-1}(-j\omega)^{n-1} + \cdots + a_1(-j\omega) + a_0 \right]
$$

$$
= \frac{A}{2j} \left(e^{j\omega t} - e^{-j\omega t} \right) \tag{15.22}
$$

The coefficients of $e^{j\omega t}$ on both sides of Eq. (15.22) must be equal. Hence,

$$
D_1 e^{jD_2} \left[a_n (j\omega)^n + a_{n-1}(j\omega)^{n-1} + \cdots + a_1(j\omega) + a_0 \right] = A \tag{15.23}
$$

Equation (15.23) will be satisfied if and only if

$$
\left| \frac{1}{a_n (j\omega)^2 + a_{n-1}(j\omega)^{n-1} + \cdots + a_1(j\omega) + a_0} \right| = \frac{D_1}{A} \tag{15.24}
$$

$$
\angle \frac{1}{a_n (j\omega)^n + a_{n-1}(j\omega)^{n-1} + \cdots + a_1(j\omega) + a_0} = D_2
$$

But D_1/A and D_2 are the AR and phase angle of the response, respectively, as may be seen from Eq. (15.19) and the forcing function. Furthermore, from Eq. (15.17) the transfer function relating X and Y is

$$
\frac{Y(s)}{X(s)} = \frac{1}{a_n s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0} \tag{15.25}
$$

Equations (15.24) and (15.25)[*] establish the general result.

[*] In writing this equation, it is assumed that X and Y have been written as deviation variables, so that initial conditions are zero.

SUMMARY

In this chapter the concept of frequency response was discussed. A graphical tool, the Bode diagram, was introduced to assist in the analysis of the frequency response characteristics of various system elements. Bode plots were constructed manually as well as with MATLAB. We will use this tool in Chap. 16 in the design of control systems.

PROBLEMS

15.1. For each of the following transfer functions, sketch the gain versus frequency, asymptotic Bode diagram. For each case, find the actual gain and phase angle at $\omega = 10$. *Note:* It is not necessary to use log-log paper; simply rule off decades on rectangular paper.

(a) $\dfrac{100}{(10s + 1)(s + 1)}$

(b) $\dfrac{10s}{(s + 1)(0.1s + 1)^2}$

(c) $\dfrac{s + 1}{(0.1s + 1)(10s + 1)}$

(d) $\dfrac{s - 1}{(0.1s + 1)(10s + 1)}$

(e) $(10s + 1)^2$

(f) $(10 + s)^2$

15.2. A temperature bath in which the temperature varies sinusoidally at various frequencies is used to measure the frequency response of a temperature-measuring element *B*. The apparatus is shown in Fig. P15.2. A standard thermocouple *A,* for which the time constant is 0.1 min for the arrangement shown in the sketch, is placed near the element to be measured. The response of each temperature-measuring element is recorded simultaneously on a two-channel recorder. The phase lag between the two chart records at different frequencies is shown in the table. From these data, show that it is reasonable to consider element *B* as a first-order process and calculate the time constant. Describe your method clearly.
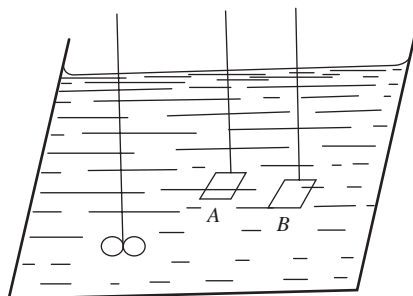
FIGURE P15–2

Frequency, cycles/min	Phase lag of B behind A, deg
0.1	7.1
0.2	12.9
0.4	21.8
0.8	28.2
1.0	29.8
1.5	26.0
2.0	23.6
3.0	18.0
4.0	14.2

15.3. Plot the asymptotic Bode diagram for the PID controller

$$G(s) = K_c\left(1 + \tau_D s + \frac{1}{\tau_I s}\right)$$

where $K_c = 10$, $\tau_I = 1$, and $\tau_D = 100$. Label corner frequencies and give slopes of asymptotes.

15.4. One way of experimentally measuring the frequency response is to plot the output sine wave versus the input sine wave. The results of such a plot look like Fig. P15–4. This is the sinusoidal *deviation* in output versus sinusoidal *deviation* in input and appears as an ellipse centered at the origin. Show how to obtain the AR and phase lag from this plot.
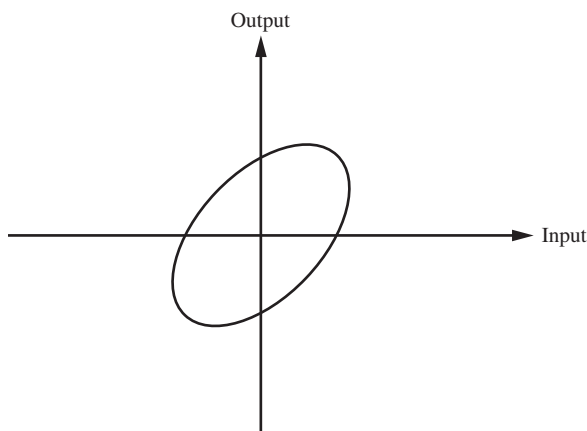
FIGURE P15–4

15.5. For the transfer function shown below, sketch carefully the gain versus frequency portion of the asymptotic plot of the Bode diagram. Determine the actual (exact) value of gain and phase angle at $\omega = 1$. Determine the phase angle as $\omega \to \infty$.

$$G(s) = \frac{2(0.1s + 1)}{s^2(10s + 1)}$$

Indicate very clearly the slopes of the asymptotic Bode diagram of $G(s)$.

15.6. (*a*) Plot accurately and neatly the Bode diagram for the process shown in Fig. P15–6, using log-log paper for gain versus frequency and semilog paper for phase versus frequency. Plot the frequency as radians per minute.

(*b*) Find the amplitude ratio and phase angle for Y/X at $\omega = 1$ rad/min and $\omega = 4$ rad/min.
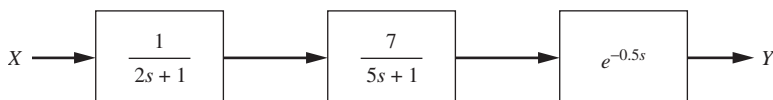
FIGURE P15–6

15.7. For the system shown in Fig. P15–7, determine accurately the phase angle in degrees between $Y(t)$ and $X(t)$ for $\omega = 0.5$. Determine the lag between the input wave and the output wave.

$$X = 2 \sin \omega t \longrightarrow \boxed{\dfrac{3}{9s^2 + 0.5s + 1}} \longrightarrow Y$$
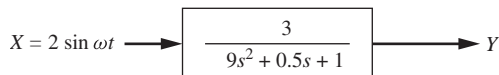
FIGURE P15–7

15.8. (*a*) For the transfer function given below, sketch carefully the asymptotic approximation of gain versus frequency. Show details such as slopes of asymptotes.

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

(*b*) Find the actual (exact) value of gain and phase angle for $\omega = 1$ and for $\omega = 2$.

15.9. Derive expressions for amplitude ratio and phase angle as functions of ω for the transfer function $G(s) = 1/(s^2 - 1)$.

15.10. The data given in the following table represent experimental, frequency response data for a process consisting of a first-order process and a transportation lag. Determine the time constant and the transportation lag parameter. Write the transfer function for the process, giving numerical values of the parameters.

Frequency, cycles/min	Gain	Phase angle, deg
0.01	1.0	0.0
0.02	1.0	−2.0
0.04	1.0	−6.0
0.06	1.0	−7.0
0.08	1.0	−8.5
0.10	1.0	−11.0
0.15	1.0	−17.0
0.20	1.0	−23.0
0.30	1.0	−36.0
0.40	0.98	−48.0
0.60	0.94	−73.0
0.80	0.88	−96.0
1.00	0.83	−122.0
1.50	0.71	−180.0
2.00	0.61	−239.0
4.00	0.37	—
6.00	0.26	—
8.00	0.20	—
10.00	0.16	—
20.00	0.080	—
40.00	0.041	—

CAPSULE SUMMARY

Substitution Rule: Substitute ($j\omega$ for s) in the transfer function to obtain a complex number. The magnitude of the resulting complex number is the amplitude ratio AR, while the angle of the complex number is the phase angle shift of the output response. This rule is valid for transfer functions that yield stable systems.

FIGURE 15–2

Characteristics of a steady-state sinusoidal response.

$$\text{Amplitude ratio AR } = \frac{\text{output amplitude}}{\text{input amplitude}} = \left|G(j\omega)\right| = R$$

$$\text{Phase angle } \phi = \angle G(j\omega)$$

Key Features to Note About the Frequency Response of the Process
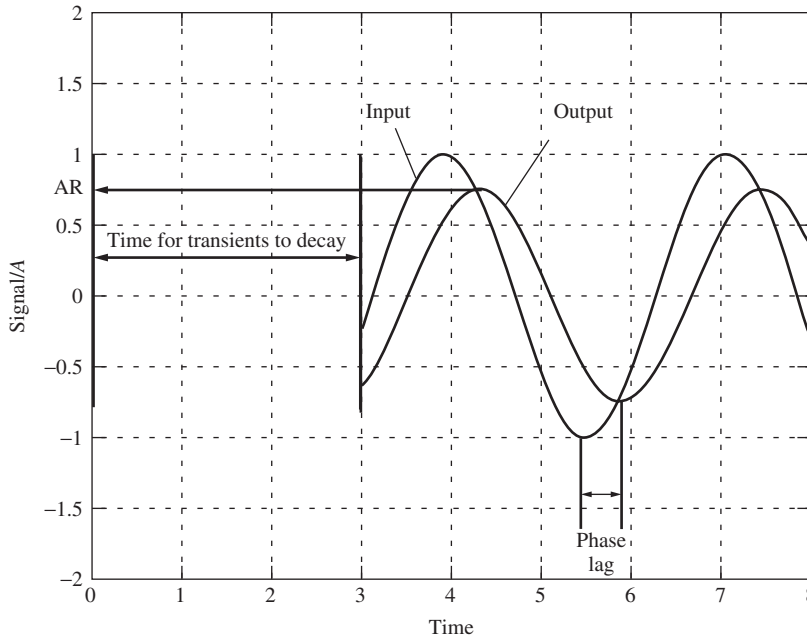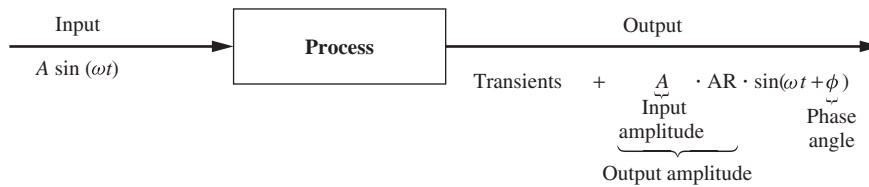
- After transients die out, the output is a sine wave.
- Input frequency = output frequency = ω.
- In general, the output is attenuated, that is, AR < 1.
- The output is shifted in time (it lags the input by the phase angle).
- Amplitude ratio and phase angle are both functions of frequency.

Bode diagrams are plots of the amplitude ratio and the phase angle of the steady response as functions of the frequency of the input sine wave.

Characteristics of Bode plots for some common transfer functions

Transfer function	Amplitude ratio AR	Phase angle (deg) ϕ	Corner frequency ω_c	Slope on Bode plot $\omega < \omega_c$	Slope on Bode plot $\omega > \omega_c$
$\dfrac{K}{1 + \tau s}$ First-order system	$K\dfrac{1}{\sqrt{1 + \omega^2 \tau^2}}$	$\tan^{-1}(-\omega\tau)$ $\phi \to -90°$ as $\omega \to \infty$	$\dfrac{1}{\tau}$	0	-1
$K(1 + \tau s)$ First-order lead	$K\sqrt{1 + \omega^2 \tau^2}$	$\tan^{-1}(\omega\tau)$ $\phi \to +90°$ as $\omega \to \infty$	$\dfrac{1}{\tau}$	0	1
$\dfrac{1}{\tau s}$ Integral	$\dfrac{1}{\omega\tau}$	-90	0	-1	-1
τs Derivative	$\omega\tau$	$+90$	0	1	1
$\dfrac{K}{\tau^2 s^2 + 2\zeta\tau s + 1}$ Second-order system	$K\dfrac{1}{\sqrt{\left(1 - \omega^2\tau^2\right)^2 + (2\zeta\omega\tau)^2}}$	$\tan^{-1}\left(\dfrac{-2\zeta\omega\tau}{1 - \omega^2\tau^2}\right)$ $\phi \to -180°$ as $\omega \to \infty$	$\dfrac{1}{\tau}$	0	-2
$\exp(-\tau s)$ Transportation lag	1	$-\omega\tau\left(\dfrac{180}{\pi}\right)$	0	—	0

CHAPTER
16

CONTROL SYSTEM DESIGN BY FREQUENCY RESPONSE

The purpose of this chapter is twofold. First, it is indicated that the stability of a control system can usually be determined from the Bode diagram of its open-loop transfer function. Then methods are presented for rational selection of controller parameters based on this Bode diagram. The material presented here is one of the more useful design aspects of the subject of frequency response.

16.1 TANK TEMPERATURE CONTROL SYSTEM

It was indicated in the discussion following Example 15.4 that the control system of Fig. 16–1 might offer stability problems because of excessive phase lag. To review, this system represents proportional control of tank temperature with a delay in the feedback loop. The factor $\frac{1}{600}$ is the process sensitivity $1/(wC)$, which gives the ultimate change in tank temperature per unit change in heat input Q.

$$\frac{1}{wC} = \frac{1}{\left(600\,\dfrac{\text{lb}}{\text{min}}\right)\left(1\,\dfrac{\text{Btu}}{\text{lb·°F}}\right)} = \frac{\text{°F}}{600\,\dfrac{\text{Btu}}{\text{min}}}$$

The proportional gain K_c, in Btu per minute per degree of temperature error, is to be specified by the designer.

The open-loop transfer function for this system is

$$G(s) = \frac{(K_c / 600)\,e^{-0.0396s}}{0.203s + 1} \tag{16.1}$$
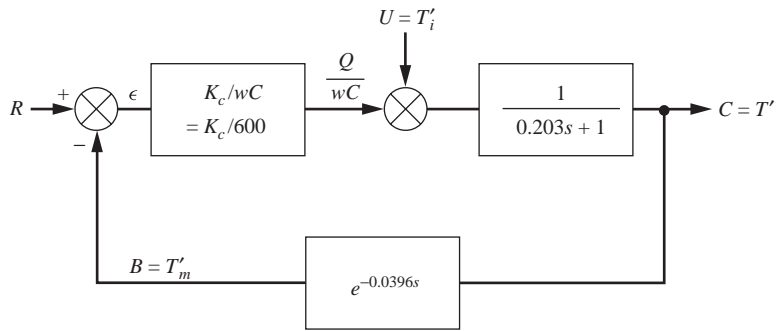
FIGURE 16–1
Control system for stirred-tank heater of Example 15.4.

The Bode diagram for $G(s)$ is plotted in Fig. 16–2. As usual, the constant factor $K_c/600$ is included in the definition of the ordinate for AR.
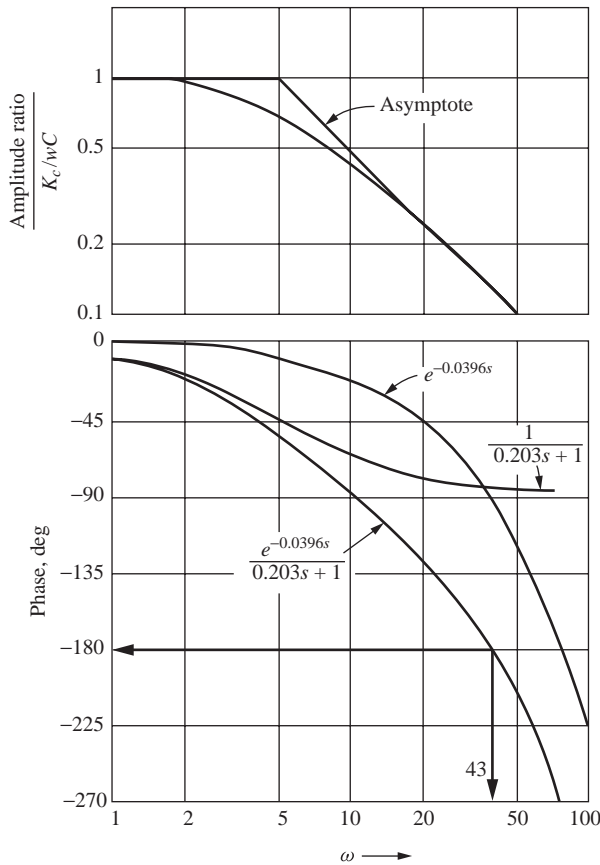
FIGURE 16–2
Bode diagram for open-loop transfer function of control system for stirred-tank heater:
$\left[(K_c/wC)e^{-\tau_2 s}\right]/(\tau_1 s + 1)$. Block diagram is shown in Fig. 16–1.

At the frequency of 43 rad/min, the phase lag is exactly 180° and $AR/(K_c/600) = 0.12$.

$$AR = \left(\frac{K_c}{600}\right)(0.12) = \frac{K_c}{600/0.12} = \frac{K_c}{5000}$$

Therefore, if a proportional gain of 5000 Btu/(min·°F) is used,

$$AR = \frac{5000}{5000} = 1$$

This is the AR between the signals ε and B. Note that it is dimensionless, as ε and B both have the units of temperature.

The control system is redrawn for $K_c = 5000$ in Fig. 16–3a, with the loop opened. That is, the feedback signal B is disconnected from the comparator. Imagine that a set point disturbance

$$R = \sin 43t$$

is applied to the opened loop. Then, since the open-loop $AR = 1$ and the phase lag $= 180°$,

$$B = \sin(43t - 180°) = -\sin 43t$$

Now imagine that, at some instant in time, R is set to zero and simultaneously the loop is closed. Figure 16–3b indicates that the closed loop continues to oscillate indefinitely. *This oscillation is theoretically sustained even though both R and U are zero.*

(a)

(b)
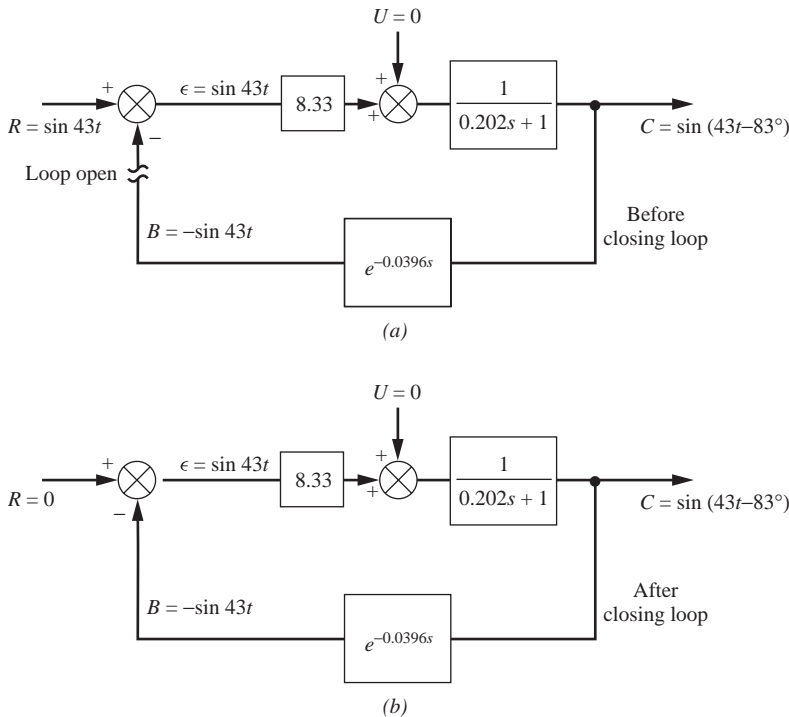
FIGURE 16–3
Sustained closed-loop oscillation.

Now suppose K_c is set to a slightly higher value and the same experiment repeated. This time, the signal ε is amplified slightly each time it passes around the loop. Thus, if K_c is set to 5001, after the first time around the loop the signal ε becomes (5001/5000) sin 43t. After the second time, it is $(5001/5000)^2$ sin 43t, etc. The phase angle is not affected by changing K_c. We thus conclude that, for $K_c > 5000$, the response is unbounded, since it oscillates with *increasing amplitude.*

By using the definition of stability presented in Chap. 13, it is concluded that the control system is unstable for $K_c > 5000$ because it exhibits an unbounded response to the bounded input described above. (The bounded input is zero in this case, for $U = R = 0$.) The condition $K_c > 5000$ corresponds to

$$\text{AR} > 1$$

for the open-loop transfer function, at the frequency 43 rad/min, where the open-loop phase lag is 180°.

This argument is not rigorous. We know the response B only if ε *remains constant in amplitude* because of the definition of frequency response. If, however, the change in K_c is very small, so that ε is amplified infinitesimally, then B will closely approximate the frequency response. While this does not *prove* anything, it shows that we are justified in suspecting instability and that closer investigation is warranted. A rigorous proof of stability requires application of the Nyquist stability criterion [see Coughanowr and Koppel (1965) or Kuo (1987)], which uses the theory of complex variables. For our purposes, it is sufficient to proceed with heuristic arguments.

16.2   THE BODE STABILITY CRITERION

It is tempting to generalize the results of the analysis of the tank temperature control system to the following rule. *A control system is unstable if the open-loop frequency response exhibits an AR exceeding unity at the frequency for which the phase lag is* 180°. This frequency is called the *crossover frequency.* The rule is called the *Bode stability criterion.*

Actually, since the discussion of Sec. 16.1 was based on heuristic arguments, this rule is not quite general. It applies readily to systems for which the gain and phase curves decrease continuously with frequency. However, if the phase curve appears as in Fig. 16–4, the more general Nyquist criterion must usually be used to determine stability. Other exceptions may occur. Fortunately, most process control systems can be analyzed with the simple Bode criterion, and it therefore finds wide application.

Application of the criterion requires nothing more than plotting the open-loop frequency response. This may be based on the theoretical transfer function, if it is available, as we have done for the tank temperature system. If the theoretical system dynamics are not known, the frequency response may be obtained experimentally. To do this, the open-loop system is disturbed with a sine wave input at several frequencies. At each frequency, records of the input and output waves are compared to establish the AR and phase lag. The results are plotted as a Bode diagram. This experimental technique is illustrated in greater detail in Chap. 18.

For the remainder of this chapter, we accept the Bode stability criterion as valid and use it to establish the control system design procedure.
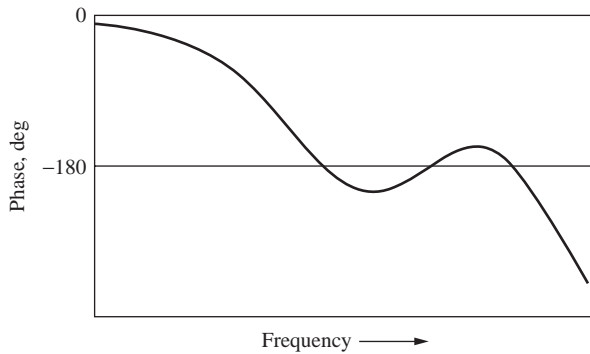
FIGURE 16–4
Phase behavior of a complex system for which the Bode criterion is not applicable.

16.3   GAIN AND PHASE MARGINS

Let us consider the general problem of selecting $G_c(s)$ for the system of Fig. 16–5. Suppose the open-loop frequency response, when a particular controller $G_c(s)$ is tried, is as shown in the Bode diagram of Fig. 16–6. The crossover frequency, at which the phase lag is 180°, is noted as ω_{co} on the Bode diagram. At this frequency, the AR is A. If A exceeds unity, we know from the Bode criterion that the system is unstable and that we have made a poor selection of $G_c(s)$. In Fig. 16–6 it is assumed that A is less than unity and therefore the system is stable.
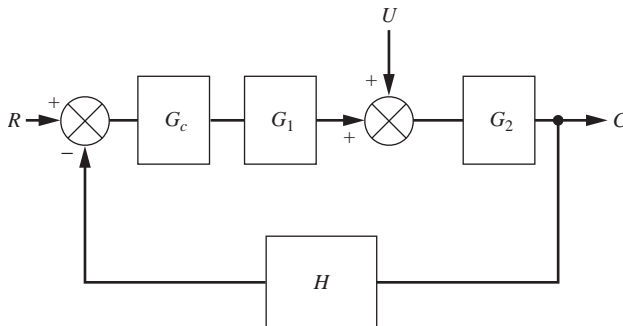
FIGURE 16–5
Block diagram for a general control system.

It is necessary to ascertain to what degree the system is stable. Intuitively, if A is only slightly less than unity, the system is "almost unstable" and may be expected to behave in a highly oscillatory manner even though it is theoretically stable. (Again, heuristic arguments are used. This statement is self-evident to the reader who has studied Chap. 14, where it is shown that the roots of the characteristic equation vary continuously with system parameters. Proof of the statement requires the Nyquist stability criterion.)
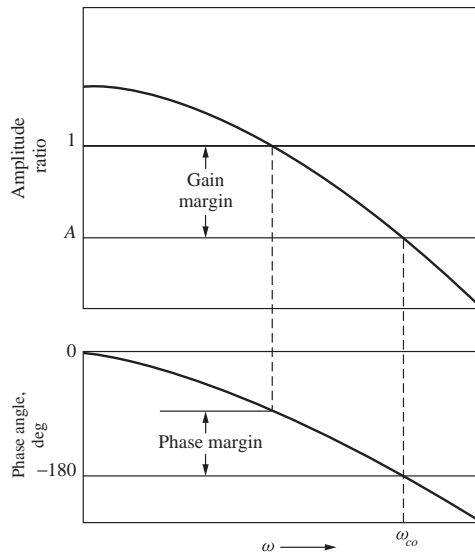
FIGURE 16–6
Open-loop Bode diagram for a typical control system.

Furthermore, the constant A is determined by the physical parameters of the system, such as time constants. These can be only estimated and may actually change slowly with time because of wear or corrosion. Hence, a design for which A is close to unity does not have an adequate safety factor.

To assign some quantitative measure to these considerations, the concept of gain margin (GM) is introduced. Using the nomenclature of Fig. 16–6,

$$\text{Gain margin} = \frac{1}{A} = \frac{1}{\text{AR}_{\phi=-180°}}$$

Typical specifications for design are that the gain margin should be greater than 1.7. This means that the AR at crossover could increase by a factor of 1.7 over the design value before the system became unstable.

$$\text{AR}_{\phi=-180°} = \frac{1}{\text{GM}} = \frac{1}{1.7} = 0.59$$

So, for GM = 1.7, AR = 0.59 at the crossover frequency. The design value of the gain margin is really a safety factor that maintains the AR a "safe distance" away from AR = 1 at ω_{co}. *As such, its value varies considerably with the application and designer.* A gain margin of unity or less indicates an unstable system.

Another margin frequently used for design is the phase margin. As indicated in Fig. 16–6, it is the difference between 180° and the phase lag at the frequency for which the gain is unity.

$$\text{Phase margin (PM)} = 180° - \phi_{\text{LAG}}\big|_{\text{AR}=1}$$

The phase margin therefore represents the additional amount of phase lag required to destabilize the system, just as the gain margin represents the additional gain for destabilization. *Typical design specifications are that the phase margin must be greater than 30°.* A negative phase margin indicates an unstable system.

Example 16.1.   Find a relation between relative stability (see below) and the phase margin for the control system of Fig. 16–7. A proportional controller is to be used.

This block diagram corresponds to the stirred-tank heater system, for which the block diagram has been given in Fig. 12–17. The particular set of constants is

$$\tau = \tau_m = 1$$

$$\frac{1}{wC} = 1$$

These are to be regarded as fixed, while the proportional gain K_c is to be varied to give a satisfactory phase margin.
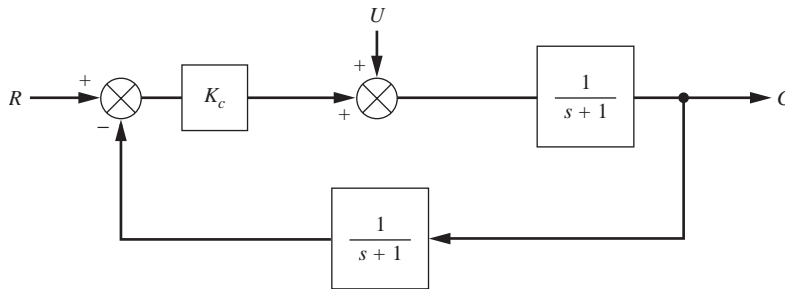
FIGURE 16–7
Block diagram for Example 16.1.

The *closed-loop* transfer function for this system is given by Eq. (12.17), rewritten for our particular case as

$$\frac{C}{R} = \frac{K_c}{1 + K_c} \frac{s + 1}{\tau_2^2 s^2 + 2\tau_2 \zeta_2 s + 1} \tag{16.2}$$

where
$$\tau_2 = \sqrt{\frac{1}{1 + K_c}} \qquad \zeta_2 = \sqrt{\frac{1}{1 + K_c}}$$

Since the closed-loop system is second-order, it can never be *unstable.* The shape of the response of the closed-loop system to a unit step in R must resemble the curves of Fig. 7–3. The meaning of *relative stability* is illustrated by Fig. 7–3. The lower ζ_2 is made, the more oscillatory and hence the "less stable" will be the response. Therefore, a relationship between phase margin and ζ_2 will give the relation between phase margin and relative stability.
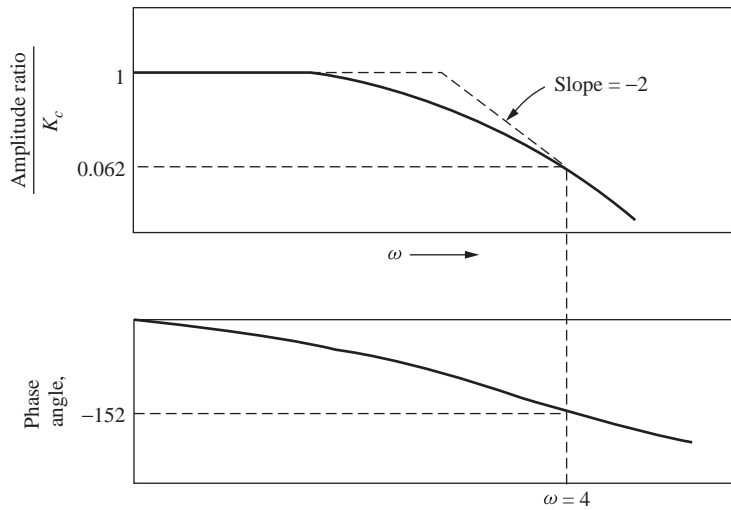
FIGURE 16–8
Open-loop Bode diagram for system of Example 16.1, $G = 1/(s + 1)^2$.

To find this relation the open-loop Bode diagram is prepared and is shown in Fig. 16–8. The simplest way to proceed from this diagram is as follows: Consider a typical frequency $\omega = 4$. If the open-loop gain were 1 at this frequency, then since the phase angle is $-152°$, the phase margin would be 28°. To make the open-loop gain 1 at $\omega = 4$, it is required that

$$K_c = \frac{1}{0.062} = 16.1$$

Then

$$\zeta_2 = \sqrt{\frac{1}{1 + K_c}} = 0.24$$

Hence, a point on the curve of ζ_2 versus phase margin is

$$\zeta_2 = 0.24 \qquad \text{phase margin} = 28°$$

Other points are calculated similarly at different frequencies, and the resulting curve is shown in Fig. 16–9. From this figure it is seen that ζ_2 decreases with decreasing phase margin and that if the phase margin is less than 30°, then ζ_2 is less than 0.26. From Fig. 7–3, it can be seen that the response of this system for $\zeta_2 < 0.26$ is highly oscillatory, hence relatively unstable, compared with a response for the system with phase margin 50° and $\zeta_2 = 0.4$.

For the particular system of Example 16.1, it was shown that the response became more oscillatory as the phase margin was decreased. This result generalizes to more complex systems. Thus, the phase margin is a useful design tool for application to systems of higher complexity, where the transient response cannot be easily determined and a plot such as Fig. 16–9 cannot be made. To repeat, the rule of thumb is that the phase margin must be greater than 30°.
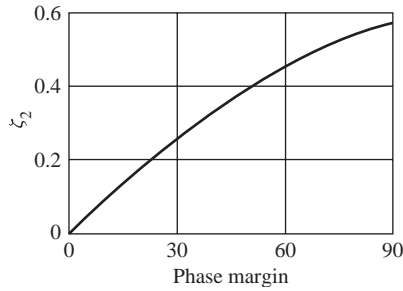
FIGURE 16–9
Damping versus phase margin for system of Fig. 16–7.

A similar statement can be made about the gain margin. As the gain margin is increased, the system response generally becomes less oscillatory, hence more stable. A control system designer will often try to make *both* the gain and phase margins equal to or greater than specified minimum values, typically 1.7 and 30°. Note that, for the case of Example 16.1, the gain margin is always infinite because the phase lag never quite reaches 180°. However, the phase margin requirement of 30° necessitates that $\zeta_2 > 0.26$, hence $K_c < 14$, which means that an offset of $\frac{1}{15}$ [see Eq. (16.2)] must be accepted. This illustrates the importance of considering both margins. The reader should refer to Fig. 16–6 to see that both margins exist simultaneously.

Example 16.2.   Specify the proportional gain K_c for the control system of Fig. 15–23, reproduced as Fig. 16–10.
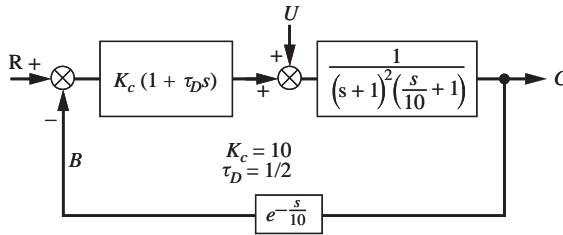
FIGURE 16.10
Block diagram of control system for Example 15.6.

The gain is to be specified for the two cases:

	Proportional gain	τ_D	Open-loop transfer function
Case 1 (PD control)	K_c	0.5 min	$G(s) = \dfrac{K_c\,(0.5s + 1)e^{-s/10}}{(s + 1)^2\,(0.1s + 1)}$
Case 2 (P control only)	K_c	0 min	$G(s) = \dfrac{K_c e^{-s/10}}{(s + 1)^2\,(0.1s + 1)}$

We'll use MATLAB to generate the exact Bode diagrams (we sketched the asymptotes in Chap. 15) for the two cases. We first need to generate the transfer function in MATLAB. For case 1, the MATLAB commands to produce the Bode diagram are as follows:

```
num=[0.5 1]  ;          % the coefficients of the polynomial
                        in the numerator of G(s);
den=conv([1 2 1],[0.1 1])    %conv multiplies the 2
                             polynomials in the denominator
                             of G(s);
den =
    0.1000    1.2000    2.1000    1.0000
% thus the resulting polynomial in the denominator is
```
$0.1s^3 + 1.2s^2 + 2.1s + 1$

```
sys=tf(num,den,'iodelay',0.1)


Transfer function:
                            0.5 s + 1
exp(-0.1*s)  *  ───────────────────────────────
                0.1 s^3 + 1.2 s^2 + 2.1 s + 1
% Now we can generate the Bode diagram(Fig.16-11) using
mybode (discussed in Chap. 15).
mybode(sys)
```
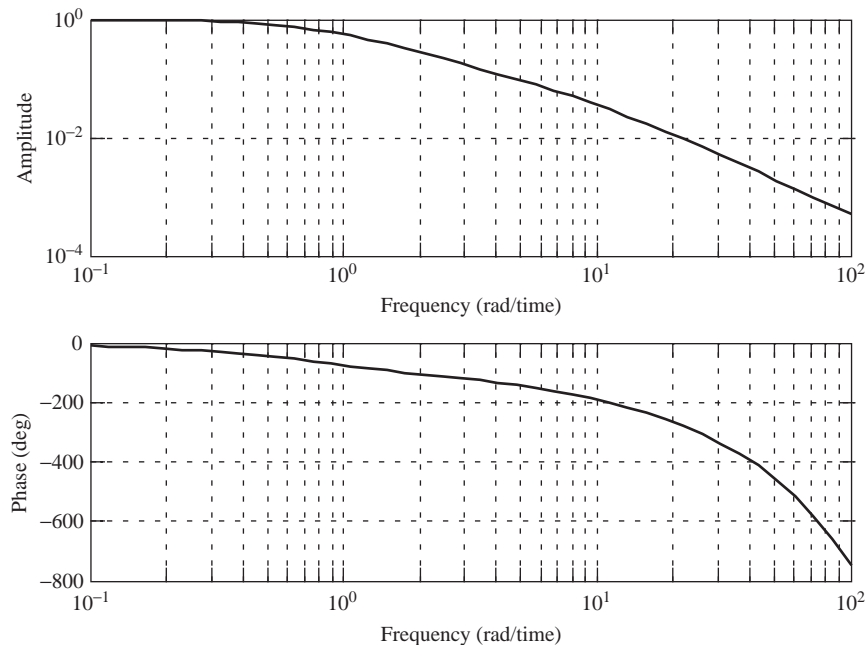

FIGURE 16–11
Bode diagram for Example 16.2, case 1 PD control.

We can examine the data used to generate Fig. 16–11 as follows.

Solution
Case 1. Consider first the gain margin. The crossover frequency for the curve with deriva-
tive action is 8.62 rad/min. At this frequency, the open-loop gain is 0.0445 if the value of
K_c is unity. (Including the factor of $\frac{1}{10}$ in the ordinate is actually equivalent to plotting
the case $K_c = 1$.) Therefore, according to the Bode criterion, the value of K_c necessary to
destabilize the loop is 1/0.0445, or 22.5. To achieve a gain margin of 1.7, K_c must be taken
as 22.5/1.7, or 13.2. To achieve proper phase margin, note that the frequency for which the
phase lag is 150° (phase margin is 30°) is 5.52 rad/min. At this frequency, a value for K_c of
1/0.0815, or 12.3, will cause the open-loop gain to be unity. Since this is lower than 13.2,
we use 12.3 as the design value of K_c. The resulting gain margin is then 1.83.
       The numerical data for the Bode diagram can be output to Excel and also calculated
in Excel by using the formulas for magnitude and phase angle. Goal Seek can also be used
in Excel to determine the frequency and magnitude at the phase angles of interest ($-180°$
and $-150°$). With Goal Seek, you have Excel vary the value of the frequency until the
phase angle equals $-180°$ or $-150°$. The Excel formulas to calculate the magnitude and
phase angle are:

```
Magnitude =(SQRT(frequency^2/4+1))/(frequency^2+1)/
           (SQRT((0.1*frequency)^2+1))
Phase Angle =(2*ATAN(-frequency)+ATAN(-0.1*frequency)
           +ATAN(0.5*frequency)+(-frequency/10))*180/3.14159
```

(The data exported from the MATLAB bode function are shown in the Fig. 16–12 for
comparison purposes.)

frequency	magnitude	magnitude	phase angle	phase angle
	matlab	Excel	matlab	Excel
0.1	0.9913	0.9913	−9.7047	−9.7047
0.1184	0.9878	0.9878	−11.4688	−11.4736
...	...	...	...	...
4.8187	0.097	0.0970	−142.4299	−142.4304
5.7029	0.0783	0.0783	−151.8055	−151.8054
6.7494	0.0627	0.0627	−162.339	−162.3387
7.988	0.0496	0.0496	−174.1712	−174.1712
9.4539	0.0388	0.0388	−187.4274	−187.4278
11.1887	0.03	0.0300	−202.2378	−202.2378
...	...	...	...	...
60.3237	0.0014	0.0014	−516.2162	−516.2168
71.3935	0.001	0.0010	−581.0807	−581.0813
84.4947	0.0007	0.0007	−657.369	−657.3697
100	0.0005	0.0005	−747.2471	−747.2477

Using Goal Seek ...

frequency	magnitude	phase angle
8.6223706	0.0445	−180.0000
5.5290187	0.0815	−150.0000

FIGURE 16–12
Use of Excel to calculate the magnitude and frequency for Bode diagrams.

Case 2. Proceeding exactly as in case 1 but using the curve for no derivative action, we found that $K_c = 6.7$ is needed for satisfactory gain margin and $K_c = 5.14$ for satisfactory phase margin. Hence K_c is taken as 5.14, and the resulting gain margin is 2.2.

The MATLAB code for generating this Bode diagram (Fig. 16–13) is

```
num=[1]

num =

     1

den=[0.1 1.2 2.1 1.0]

den =

    0.1000    1.2000    2.1000    1.0000

sys=tf(num,den,'iodelay',0.1)


Transfer function:
                                   1
exp(-0.1*s) *  ----------------------------------
                  0.1 s^3 + 1.2 s^2 + 2.1 s + 1

mybode(sys)
```
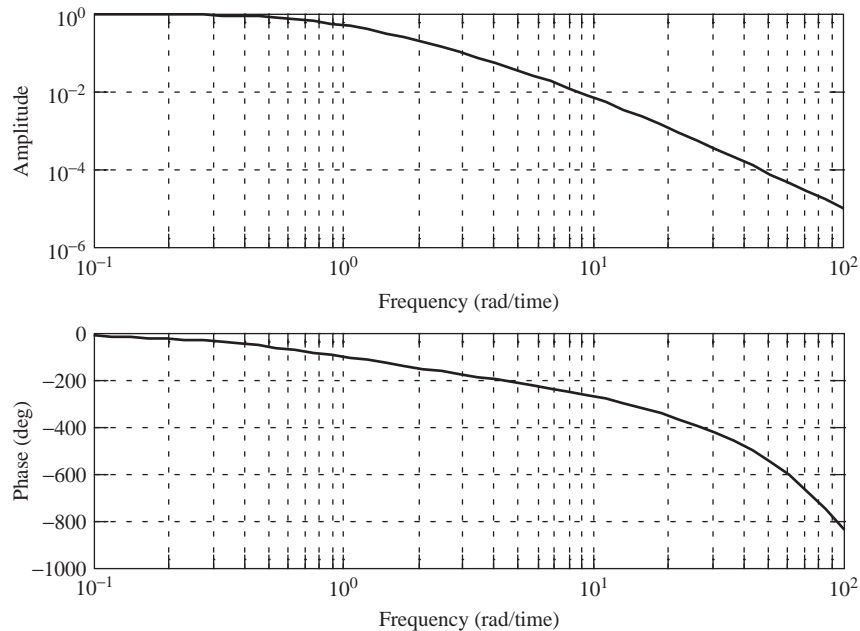

FIGURE 16–13
Bode diagram for Example 16.2, case 2, P control only.

To see the advantage of adding derivative control in this case, note from Fig. 16–10 that the final value of C for a unit step change in U is $1/(1 + K_c)$ for any value of τ_D. The addition of the derivative action allows an increase in the value of K_c from 5.14 to 12.3 while maintaining approximately the same relative stability in terms of gain and phase margins. This reduces the offset from 16 percent of the change in U to 7.5 percent of the change in U.

The reader is cautioned that the values of K_c selected in this way should be regarded as initial approximations to the actual values, which give "optimal" control of the system of Fig. 16–10. More will be said about this matter later in this chapter in conjunction with the two-tank chemical-reactor control system of Chap. 10.

Thus far, nothing has been said about upper limits on the gain and phase margins. Referring to Example 16.1 and Fig. 7–3, we see that if ζ_2 is too large, the response is sluggish. In fact, Fig. 7–3 suggests that for the system of Fig. 16–7 one should choose a value of ζ_2 low enough to give a short rise time without causing excessive response time and overshoot. In other words, one wants the most rapid response that has sufficient relative stability. The results of Example 16.1 generalize to many systems of higher complexity, in terms of margin. Hence, the designer frequently chooses the controller so that either the gain or phase margin is equal to its lowest acceptable value and the other margin is (probably) above its lowest acceptable value. This was the procedure followed in Example 16.2. In almost every situation, the designer faces this conflict between speed of response and degree of oscillation. In addition, if integral action is not used, the amount of the offset must be considered.

The concepts of gain and phase margin are useful in selecting K_c for proportional action. However, for additional modes of control such as PD, these concepts are difficult to apply in practice. Consider the selection of K_c and τ_D in Example 16.2. For a different value of τ_D the derivative contribution is shifted to the right or left on the Bode diagram of Fig. 15–24. This means that a different value of K_c will provide the proper margins. A typical design procedure is to select the value of τ_D for which the value of K_c resulting in a 30° phase margin is maximized. The motivation for this choice is that the offset will be minimized. However, the procedure is clearly trial and error. In the case of three-mode control, there are two parameters, τ_I and τ_D, which must be varied by trial to meet various design criteria. Fortunately, for this case and others there are simple rules for directly establishing values of the control parameters that usually give satisfactory gain and phase margins. These are the Ziegler-Nichols rules, which we develop next.

16.4   ZIEGLER-NICHOLS CONTROLLER SETTINGS

Consider selection of a controller G_c for the general control system of Fig. 16–5. We first plot the Bode diagram for the final control element, the process, and the measuring element in series $G_1G_2H(j\omega)$. *It should be emphasized that the controller is omitted from this plot.* Suppose the diagram appears as in Fig. 16–6. As noted on the figure, the crossover frequency for these three components in series is ω_{co}. At the crossover frequency, the overall amplitude ratio is A, as indicated. According to the Bode criterion, then, the gain of a proportional controller which would cause the system of Fig. 16–5 to be on the verge of instability is $1/A$. We define this quantity to be the ultimate gain K_u. Thus

$$K_u = \frac{1}{A} \tag{16.3}$$

TABLE 16.1
Ziegler-Nichols controller settings

Type of control	$G_c(S)$	K_c	τ_I	τ_D
Proportional	K_c	$0.5K_u$		
Proportional-integral (PI)	$K_c\left(1 + \dfrac{1}{\tau_I s}\right)$	$0.45K_u$	$\dfrac{P_u}{1.2}$	
Proportional-integral-derivative (PID)	$K_c\left(1 + \dfrac{1}{\tau_I s} + \tau_D s\right)$	$0.6K_u$	$\dfrac{P_u}{2}$	$\dfrac{P_u}{8}$

The ultimate period P_u is defined as the period of the sustained cycling that would occur if a proportional controller with gain K_u were used. From the discussion of Fig. 16–3, we know this to be

$$P_u = \frac{2\pi}{\omega_{co}}[=]\frac{\text{radians / cycle}}{\text{radians / sec}}[=]\frac{\text{sec}}{\text{cycle}} \tag{16.3a}$$

The factor of 2π appears, so P_u will be in units of time per cycle rather than time per radian. It should be emphasized that K_u and P_u are easily determined from the Bode diagram of Fig. 16–6.

The Ziegler-Nichols settings for controllers are determined directly from K_u and P_u according to the rules summarized in Table 16.1. Unfortunately, specifications of K_c and τ_D for PD control cannot be made using only K_u and P_u. In general, the values $0.6 K_u$ and $P_u/8$, which correspond to the limiting case of no integral action in a three-mode controller, are too conservative. That is, the resulting system will be too stable. There exist methods for this case which are in principle no more difficult to use than the Ziegler-Nichols rules. One of these is selection of τ_D for maximum K_c at 30° phase margin, which was discussed above. Another method, which utilizes the step response and avoids trial and error, is presented in Chap. 18.

The reasoning behind the Ziegler-Nichols selection of values of K_c is relatively clear. In the case of proportional control only, a gain margin of 2 is established. The addition of integral action introduces more phase lag at all frequencies (see Fig. 16–20); hence a lower value of K_c is required to maintain roughly the same gain margin. Adding derivative action introduces phase lead. Hence, more gain may be tolerated. This was demonstrated in Example 16.2. However, by and large the Ziegler-Nichols settings are based on experience with typical processes and should be regarded as first estimates.

Example 16.3. Using the Ziegler-Nichols rules, determine K_c and τ_I for the control system shown in Fig. 16–14.
For this problem, the computation will be done without plotting a Bode diagram; however, the reader may wish to do the problem with such a diagram. We first obtain the crossover frequency by applying the Bode stability criterion.

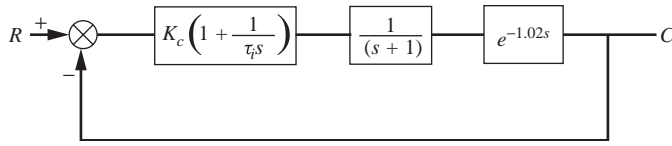$$-180° = -\tan^{-1}\omega - \left(\frac{180}{\pi}\right)(1.02)(\omega)$$

FIGURE 16–14
Block diagram for Example 16.3.

The value $180/\pi = 57.3$ converts radians to degrees. Solving this equation by trial and error gives for the crossover frequency $\omega_{co} = 2$ rad/min. The amplitude ratio AR at the crossover frequency for the open loop can be written as

$$\text{AR} = K_c \frac{1}{\sqrt{1 + \omega^2}}(1) = \frac{K_c}{2.24}$$

where we have used Eq. (15.7) for the first-order system and the fact that the amplitude ratio for a transport lag is 1. According to the Bode criterion, the AR is 1.0 at the crossover frequency when the system is on the verge of instability. Inserting AR $= 1$ into the above equation and solving for K_c gives $K_{cu} = 2.24$. From the Ziegler-Nichols rules of Table 16.1, we obtain

$$K_c = 0.45K_{cu} = (0.45)(2.24) = 1.01$$

and

$$\tau_I = \frac{P_u}{1.2} = \frac{2\pi / \omega_{co}}{1.2} = \frac{2\pi / 2}{1.2} = 2.62 \text{ min}$$

Example 16.4.   Using the Ziegler-Nichols rules, determine controller settings for various modes of control of the two-tank chemical-reactor system of Chap. 10. The block diagram is reproduced in Fig. 16–15.
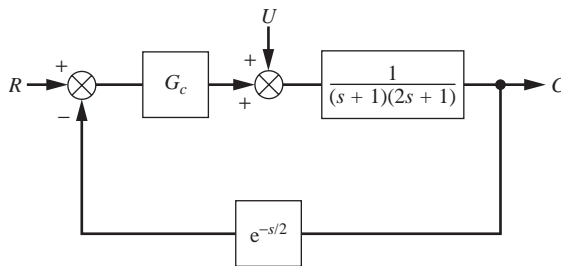

FIGURE 16–15
Block diagram for two-tank chemical-reactor system.

For convenience, the process gain K and the controller gain K_c are combined into an overall gain K_1. The equivalent controller transfer function is regarded as

$$G_c = K_1 \left(1 + \frac{1}{\tau_I s} + \tau_D s \right)$$

where K_1 (as well as τ_I and τ_D) is to be selected by the Ziegler-Nichols rules. The required value of K_c is then easily determined as

$$K_c = \frac{K_1}{K}$$

where $K = 0.09$ for the present case (see Chap. 10.)
The Bode diagram for the transfer function *without the controller*

$$\frac{e^{-0.5s}}{(s + 1)(2s + 1)}$$

is prepared by the usual procedures and is shown in Fig. 16–16.

From this figure, it is found that

$$\omega_{co} = 1.56 \text{ rad / min}$$

$$K_{1_u} = \frac{1}{0.145} = 6.9$$

$$P_u = \frac{2\pi}{1.56} = 4.0 \text{ min/cycle} \tag{16.4}$$

Hence, the Ziegler-Nichols control constants determined from Table 16.1 and Eq. (16.4) are given in Table 16.2.

A plot comparing the open-loop frequency responses *including the controller* for the three cases, using the controller constants of Table 16.2, is given in Fig. 16–17. This figure shows quite clearly the effect of the phase lead due to the derivative action. The resulting gain and phase margins are listed in Table 16.3. From this table it may be seen that the margins are adequate and generally conservative.

TABLE 16.2
Control constants for Example 16.4

Control	K_1	τ_I	τ_D
P	3.5		
PI	3.1	3.3	
PID	4.2	2.0	0.50
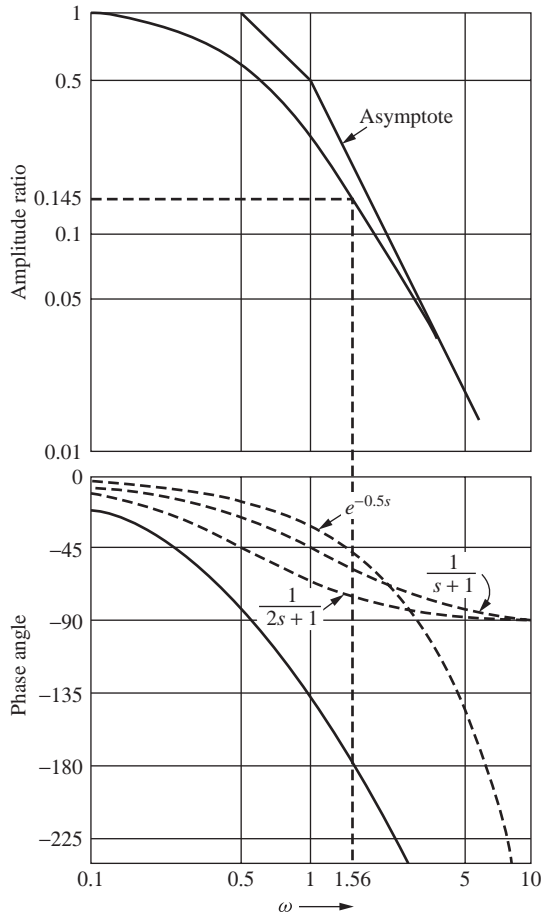
FIGURE 16–16
Bode diagram for $e^{-0.5s}/(s + 1)(2s + 1)$.

Note that to obtain the Bode diagram for systems including the PID controller, the controller transfer function is rewritten as

$$K_c\left(1 + \frac{1}{\tau_I s} + \tau_D s\right) = K_c \frac{\tau_D \tau_I s^2 + \tau_I s + 1}{\tau_I s} \tag{16.5}$$

TABLE 16.3
Gain and Phase margins for Example 16.4

Control	Gain margin	Phase margin
P	2.0	45°
PI	1.9	33°
PID	2.6	34°
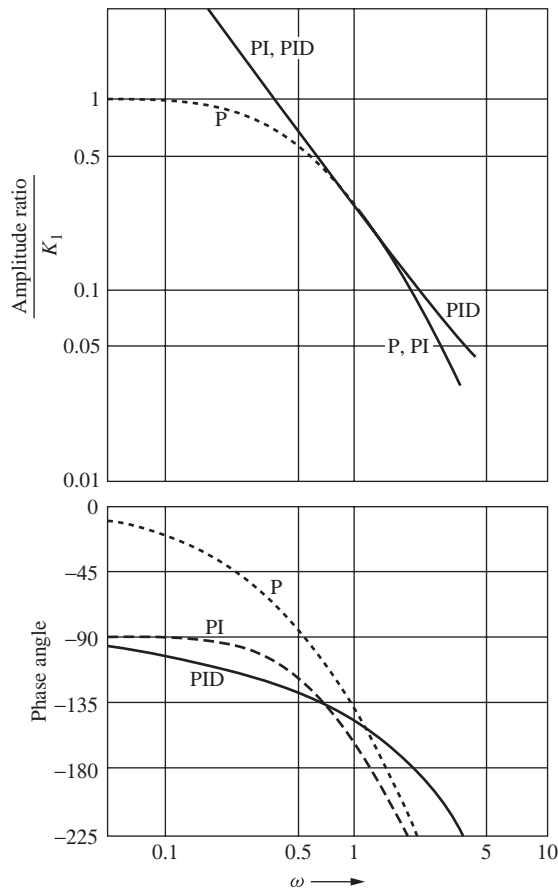
FIGURE 16–17
Open-loop Bode diagrams for various controllers with system of Fig. 16–15.

This is second-order in the numerator and has integral action in the denominator. In general, the numerator factors into first-order factors; hence it contributes two curves similar to that of Fig. 15–22 to the overall diagram. For the Ziegler-Nichols settings it is seen from Table 16.1 that $\tau_I = 4\tau_D$. Making this substitution into Eq. (16.5) gives

$$G_c = K_c \frac{4\tau_D^2 s^2 + 4\tau_D s + 1}{4\tau_D s} = \frac{K_c (2\tau_D s + 1)^2}{4\tau_D s} \tag{16.6}$$

and shows that the numerator is equivalent to two PD components in series. This AR is represented by a high-frequency asymptote of slope +2 passing through

the frequency $\omega = 1/2\tau_D$ and a low-frequency asymptote on the line AR = 1. It should be emphasized that these special considerations apply only to the Ziegler-Nichols settings. In the general case, the two time constants obtained by factoring the numerator of Eq. (16.5) will be different. The Bode plot of the denominator follows from

$$\frac{1}{\tau_I j\omega} = \left|\frac{1}{\omega\tau_I}\right| \measuredangle - 90°$$

The gain is a straight line of slope -1 passing through the point (AR = 1, ω = $1/\tau_I$). The phase lag is 90° at all frequencies. Plotting of the overall Bode diagram for the PID case to check the results of Fig. 16–17 is recommended as an exercise for the reader.

Transient Responses

For instructive purposes, the two-tank reactor system of Fig. 16–15 was simulated using MATLAB. Responses of $C(t)$ to a unit-step change in $R(t)$ are shown in Fig. 16–18. These responses were obtained using the Ziegler-Nichols controller settings determined in Example 16.4.

The responses to a step load change were also obtained using MATLAB. These are the curves of Fig. 9–9 that were discussed in Chap. 9 to illustrate the function of the various modes of control. A load change for this system corresponds to a change in the inlet concentration of reactant to tank 1 (refer to Fig. 10–1). As process control engineers, we would be more interested in controlling against this kind of disturbance than against a set point change because the set point or desired product concentration is likely to remain relatively fixed. In other words, this is a regulator problem and the curves of Fig. 9–9 are those we would use to determine the quality of control.
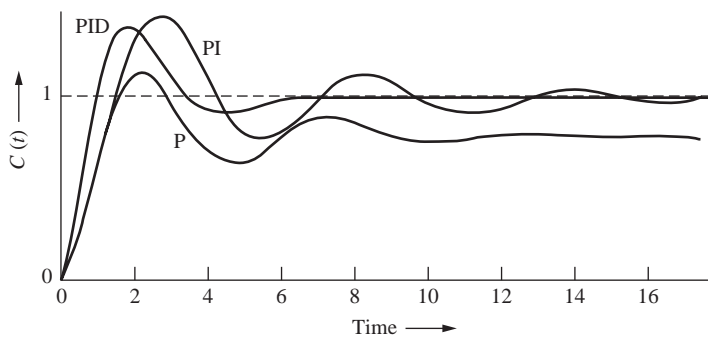

FIGURE 16–18
Closed-loop response to step change in set point for control system of Fig. 16–15 using various control modes.

However, the step change in set point is frequently used to test control systems despite the fact that the system will be primarily subject to load changes during actual operation. The reason for this is the existence of well-established terminology used

TABLE 16.4

Parameters for response of control system of Fig. 16–15 with Z-N settings

Control	Overshoot	Decay ratio	Rise time, min	Response time, min	Period of oscillation, min	Offset
P	0.49	0.26	1.3	10.4	5.0	0.21
PI	0.46	0.29	1.5	11.8	5.5	0
PID	0.42	0.05	0.9	4.9	5.0	0

to describe the step response of the underdamped second-order system. This terminology, which was presented in Chap. 7, is used to assign quantitative measure to responses that are not truly second-order, such as those of Fig. 16–18. Of course, the terminology can be applied only to responses that *resemble* damped sinusoids. Values of the various parameters determined for the responses of Fig 16–18 are summarized in Table 16.4. Offset, realized only with proportional control, is included for completeness.

It can be seen from Fig. 16–18 and Table 16.4 that addition of integral action eliminates offset at the expense of a more oscillatory response. When derivative action is also included, the response is much faster (lower rise time) and much less oscillatory (lower response time). The large overshoots realized in all three cases are characteristic of systems with relatively large time delays. In this case the controller is receiving information about the concentration in the second reactor that was true $\frac{1}{2}$ min ago. This is to be compared with the reactor time constants of 1 and 2 min. Hence, it is not surprising that the system overshoots before the controller can take sufficient action.

Figure 16–19 is presented for two purposes: (1) to illustrate that the Ziegler-Nichols controller settings should be regarded as first guesses rather than fixed values and (2) to show the effects of changing the various controller settings. These figures are transient responses to step changes in set point for the three-mode PID control. They show the effects of individually varying the three control parameters K_c, τ_I, and τ_D.

As an example of the use of these figures, suppose that it is decided that the maximum overshoot that can be tolerated is 25 percent. Figure 16–19a shows that overshoot may be reduced by decreasing K_c at the expense of a considerably more sluggish response. From Fig. 16–19b, we see that overshoot may be reduced by increasing τ_I (decreasing integral action) at a lesser expense in speed of response. Thus, for $\tau_I =$ 5 min, the overshoot is reduced to 20 percent without a serious sacrifice in speed. The overshoot cannot be significantly reduced by changing τ_D, as can be seen from Fig. 16–19c. However, the speed of response may be significantly increased by increasing the derivative action (sometimes at the expense of greater oscillation before the response has settled, as indicated by a higher decay ratio and a lower period). From this brief study of these figures, it may be concluded that, to decrease overshoot without seriously slowing the response, a combination of changes should be made. A possible
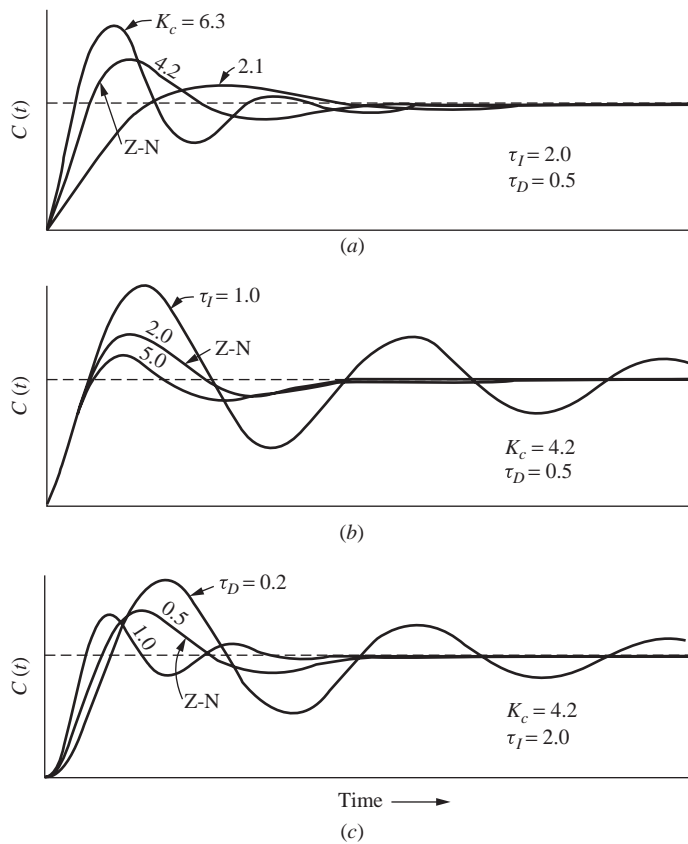
FIGURE 16–19
Effect of varying controller settings on system response. (Z-N indicates
response using Ziegler-Nichols settings.)

combination, which should be tried, is to reduce K_c slightly and to increase τ_I and τ_D
moderately. These changes would probably be tried on the actual reactor system when
it is put into operation. Such adjustments from the preliminary settings are usually made
by experienced control engineers, using trial procedures that are more art than science.
For this reason, we leave the problem of adjustment at this point.

SUMMARY

In this chapter we used frequency response tools, developed in Chap. 15, to design
control systems. We introduced the concepts of Bode stability criterion as well as gain
margin and phase margin for determining appropriate controller settings to obtain the
desired system response, while maintaining system stability. We also studied the use of
Ziegler-Nichols controller settings as initial estimates for controller tuning.

PROBLEMS

16.1. Calculate the value of gain K_c needed to produce continuous oscillations in the control system shown in Fig. P16–1 when

(*a*) *n* is 2

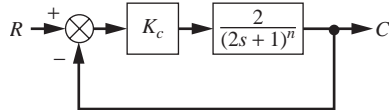(*b*) *n* is 3

Do not use a graph for this calculation.

$R \xrightarrow{+} \bigotimes \xrightarrow{} \boxed{K_c} \xrightarrow{} \boxed{\dfrac{2}{(2s+1)^n}} \xrightarrow{} C$

FIGURE P16–1

16.2. (*a*) Plot the asymptotic Bode diagram $|B/\varepsilon|$ versus ω for the control system shown in Fig. P16–2.

(*b*) The gain K_c is increased until the system oscillates continuously at a frequency of 3 rad/min. From this information, calculate the transportation lag parameter τ_d.
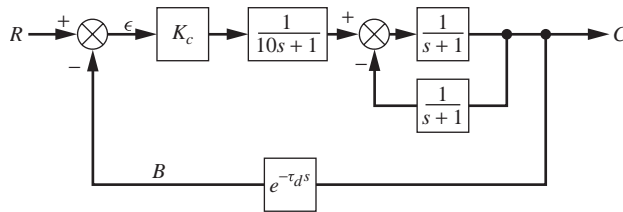

FIGURE P16–2

16.3. The frequency response for the block G_p in Fig. P16–3 is given in the following table:

f, cycles/min	Gain	Phase angle, deg
0.06	1.60	−68
0.08	1.40	−88
0.10	1.20	−105
0.15	0.84	−145
0.20	0.61	−177
0.30	0.35	−235
0.40	0.22	
0.60	0.11	
0.80	0.066	

Block G_p contains a distance velocity lag $e^{-\tau s}$ with $\tau = 1$ (this transfer function is included in the data given in the table).

(a) Find the value of K_c needed to produce a phase margin of 30° for the system if $\tau_I = 0.2$.

(b) Using the value of K_c found in part (a) and using $\tau_I = 0.2$, find the percentage change in the parameter τ to cause the system to oscillate continuously with constant amplitude.
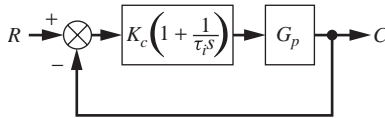

FIGURE P16–3

16.4. The system shown in Fig. P16–4 is controlled by a proportional controller. The concentration of salt in the solution leaving the tank is controlled by adding a concentrated solution through a control valve.
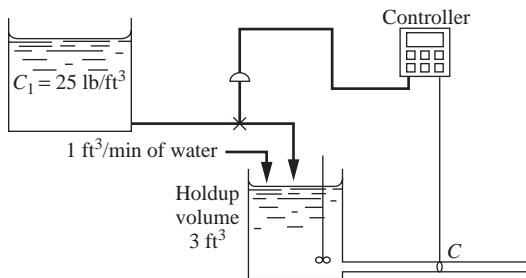

FIGURE P16–4

The following data apply:

1. Concentration of concentrated salt solution $C_1 = 25$ lb salt/ft^3 solution.
2. Controlled concentration $C = 0.1$ lb salt/ft^3 solution.
3. Control valve: The flow through the control valve varies from 0.002 to 0.0006 ft^3/min with a change of valve-top pressure from 3 to 15 psi. This relationship is linear.
4. Distance velocity lag: It takes 1 min for the solution leaving the tank to reach the concentration-measuring element at the end of the pipe.
5. Neglect lags in the valve.
    (a) Draw a block diagram of the control system. Place in each block the appropriate transfer function. Calculate all the constants and give the units.
    (b) Using a frequency-response diagram and the Ziegler-Nichols rules, determine the settings of the controller.
    (c) Using the controller settings of part (b), calculate the offset when the set point is changed by 0.02 unit of concentration.

16.5. The stirred-tank heater system shown in Fig. P16–5 is controlled by a PI controller. The following data apply:

Flow rate ω of liquid through the tanks: 250 lb/min

Holdup volume of each tank: 10 ft^3

Density of liquid: 50 lb/ft^3

Final control element: A change of 1 psi from the controller changes the heat input q by 100 Btu/min. The final control element is linear.
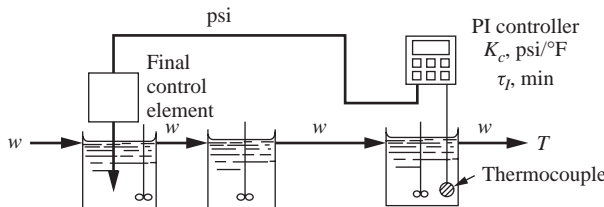

FIGURE P16.5

(a) Draw a block diagram of the control system. Show in detail such things as units and numerical values of the parameters.
(b) Determine the controller settings by the Ziegler-Nichols rules.
(c) If the control system is operated with *proportional mode only,* using the value of Kc found in part (b), determine the flow rate w at which the system will be on the verge of instability and oscillate continuously. What is the frequency of this oscillation?

16.6. The transfer function of a process and measurement element connected in series is given by

$$\frac{e^{-0.4s}}{(2s + 1)^2}$$

(a) Sketch the open-loop Bode diagram (gain and phase) for a control system involving this process and measurement lag.
(b) Specify the gain of a proportional controller to be used in this control system.

16.7. (a) For the control system shown in Fig. P16–7, determine the transfer function C/U.
(b) For $K_c = 2$ and $\tau_D = 1$, find $C(1.25)$ and the offset if $U(t) = u(t)$, a unit step.
(c) Sketch the open-loop Bode diagram for $K_c = 2$ and $\tau_D = 1$. For the upper part of the diagram (AR versus ω), show the asymptotic approximation. Include the transfer function for the controller in the open-loop Bode diagram.
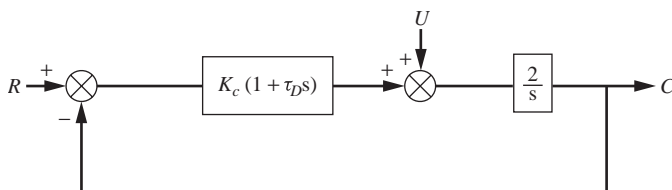(d) From the Bode diagram, what do you conclude about the stability of the closed-loop system?

FIGURE P16–7

16.8. The proportional controller of the temperature control system shown in Fig. P16–8 is properly tuned to give a good transient response for a standard set of operating conditions.

If changes are made in the operating conditions, the control system may become more or less stable. If the changes listed below are made *separately,* determine whether the system becomes more stable, less stable, or remains the same. Try to use the Bode stability criterion and sketches of frequency response graphs to solve this problem.

1. Controller gain increases.
2. Length of pipe between measuring element and tank increases.
3. Measuring element is inserted in tank.
4. Integral action is provided in controller.
5. A larger valve is used (i.e., one with a higher C_v value).

FIGURE P16–8
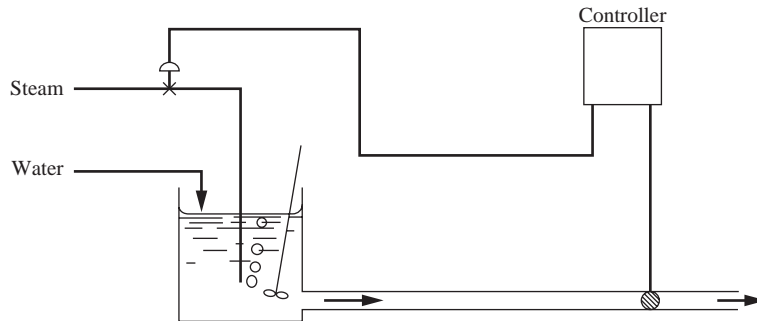
16.9. For each control system shown in Fig. P16–9, determine the characteristic equation of the closed-loop response and determine the value of K_c that will cause the system to be on the verge of instability (i.e., find the ultimate gain K_{cu}). If *possible,* use the Routh test. Note that the feedback element for system B is an approximation to e^{-2s}.
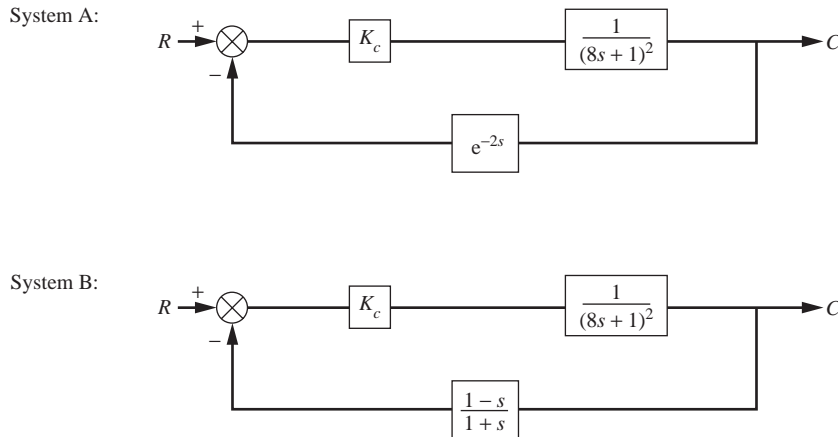

FIGURE P16–9

16.10. (*a*) For the system shown in Fig. P16–10 determine the value of K_c that will give 30° of phase margin.

(*b*) If a PI controller with $\tau_I = 2$ is used in place of the proportional controller, determine the value of K_c for 30° of phase margin.
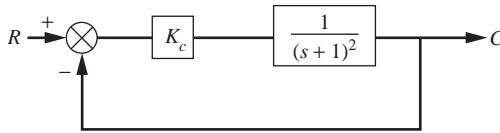
FIGURE P16–10

16.11. A stirred-tank heating process and its block diagram are shown in Fig. P16–11. The control system is tuned by the Ziegler-Nichols method, and the ultimate frequency ω_u is 2 rad/min.

(*a*) Determine the value of K_c by the Ziegler-Nichols method of tuning.

(*b*) What is the length of the pipe between the tank and the measuring element?

(*c*) What are the gain margin and the phase margin for the control system when K_c is set to the Ziegler-Nichols value found in part (*a*)?
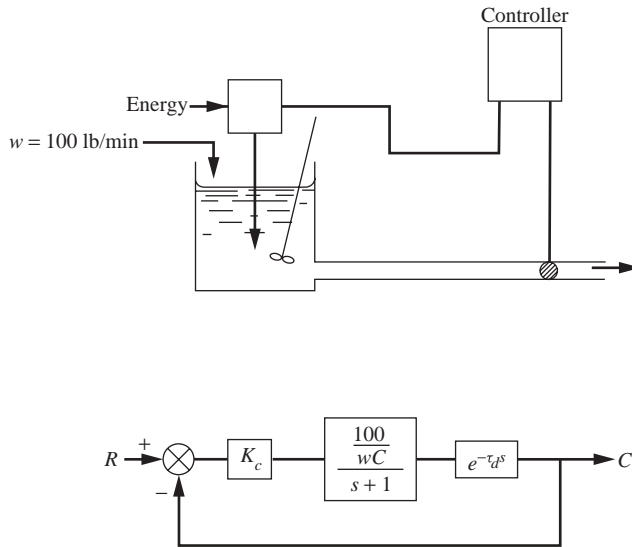
FIGURE P16–11

Data on process:

$$\text{Density of fluid, } \rho = 62 \text{ lb / ft}^3$$

$$\text{Heat capacity, } C \text{ of fluid} = 1.0 \text{ Btu / (lb·°F)}$$

$$\text{Inside diameter of pipe} = 2.0 \text{ in}$$

CAPSULE SUMMARY

THE BODE STABILITY CRITERION

A control system is unstable if the open-loop frequency response exhibits an AR exceeding unity at the frequency for which the phase lag is 180°. This frequency is called the *crossover frequency.* The rule is called the *Bode stability criterion.*

Gain Margin and Phase Margin

These quantities are defined to provide a safety margin in the design and selection of controllers.

Design rules of thumb: Set gain margin = 1.7 and phase margin = 30°. Choose the more conservative of the criteria (the one that yields the lower value of K_c). In general, the system becomes less oscillatory and more stable as GM and PM are increased. Note GM $\leq$ 0 and PM $\leq$ 0 are unstable.

$$\text{Gain margin GM} = \frac{1}{A} = \frac{1}{\text{AR}_{\phi=-180°}} \qquad \text{Phase margin (PM)} = 180° - \phi_{\text{LAG}}\big|_{\text{AR}=1}$$
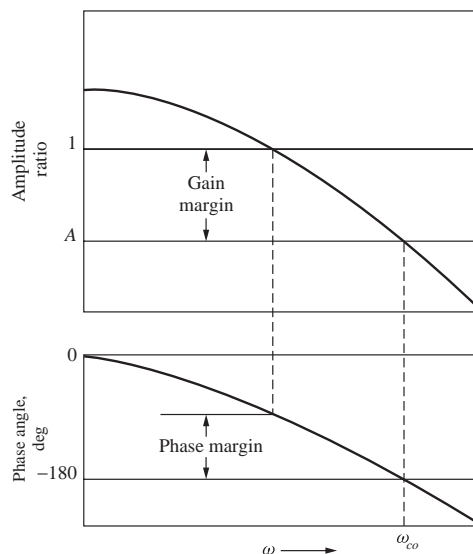
Suggested Initial Controller Tuning Parameters

Ziegler-Nichols controller settings

Type of control	$G_c(g)$	K_c	τ_1	τ_D
Proportional	K_c	$0.5K_u$		
Proportional-integral (PI)	$K_c\left(1 + \dfrac{1}{\tau_I s}\right)$	$0.45K_u$	$\dfrac{P_u}{1.2}$	
Proportional-integral-derivative (PID)	$K_c\left(1 + \dfrac{1}{\tau_I s} + \tau_D s\right)$	$0.6K_u$	$\dfrac{P_u}{2}$	$\dfrac{P_u}{8}$

PART
V

PROCESS APPLICATIONS

ADVANCED
CONTROL
STRATEGIES

Up to this point, the control systems considered have been single-loop systems involving one controller and one measuring element. In this chapter, several multiloop systems are described; these include cascade control, feedforward control, ratio control, Smith predictor control, and internal model control. The first three have found wide acceptance in industry. Smith predictor control has been known for about forty years, but it was considered impractical until the modern microprocessor-based controllers provided the simulation of transport lag. Internal model control, based on a rigorous mathematical foundation and an accurate model of the process, has been the subject of research for the past twenty years. The controller hardware and instrumentation for all these systems are readily available from manufacturers. Since this chapter is quite long, the reader may wish to select the type of advanced control strategy that is of particular interest. The descriptions of the five strategies are independent and need not be read in the order presented.

17.1   CASCADE CONTROL

To provide motivation for the study of cascade control, consider the single-loop control of a jacketed kettle as shown in Fig. 17–1a. The system consists of a kettle through which water, entering at temperature T_i, is heated to T_o by the flow of hot oil through a jacket surrounding the kettle. The temperature of the water in the kettle is measured and transmitted to the controller, which in turn adjusts the flow of hot oil through the jacket. This control system is satisfactory for controlling the kettle temperature; however, if the temperature of the oil supply should drop, the kettle temperature can undergo a large prolonged excursion from the set point before control is again established. The reason is that the controller does not take corrective action until the effect of the drop in oil

supply temperature has worked itself through the system of several resistances to reach the measuring element. To prevent the sluggish response of kettle temperature to a disturbance in oil supply temperature, the control system shown in Fig. 17–1*b* is proposed. In this system, which includes two controllers and two measuring elements, the output of the primary controller is used to adjust the set point of a secondary controller, which is used to control the jacket temperature. Under these conditions, the primary controller indirectly adjusts the jacket temperature. If the oil temperature should drop, the secondary control loop will act quickly to maintain the jacket temperature close to the value determined by the set point that is adjusted by the primary controller. This system shown in Fig. 17–1*b* is called a *cascade* control system. The primary controller is also referred to as the master controller, and the secondary controller is referred to as the slave controller.
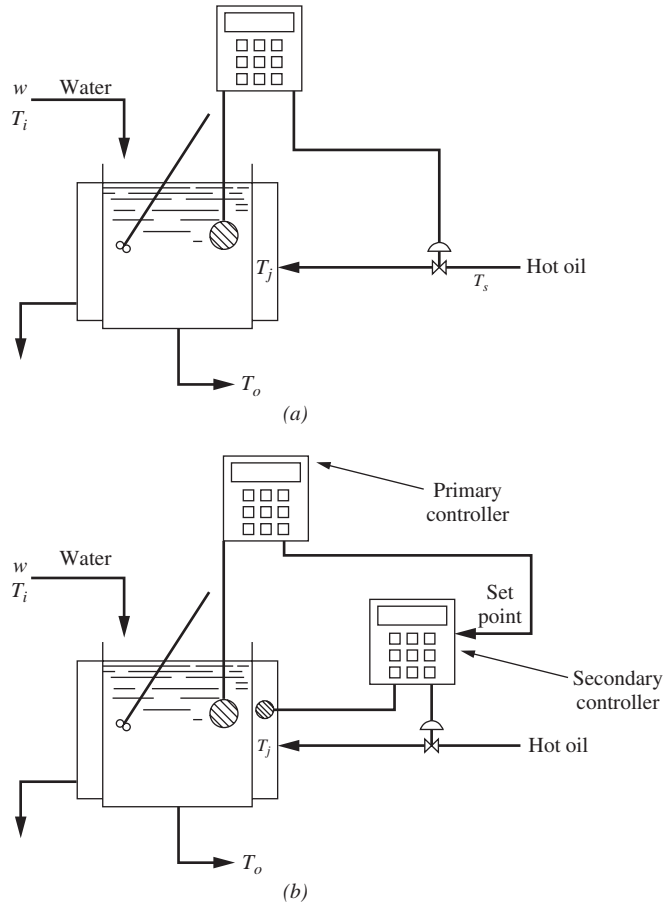

FIGURE 17–1
(*a*) Single-loop control of a jacketed kettle; (*b*) cascade control of a jacketed kettle.

A simplified block diagram of the single-loop system is shown in Fig. 17–2a. Figure 17–2b, which is a block diagram representation of the cascade control system, shows clearly that an inner loop has been added to the conventional control system.

(a)

(b)
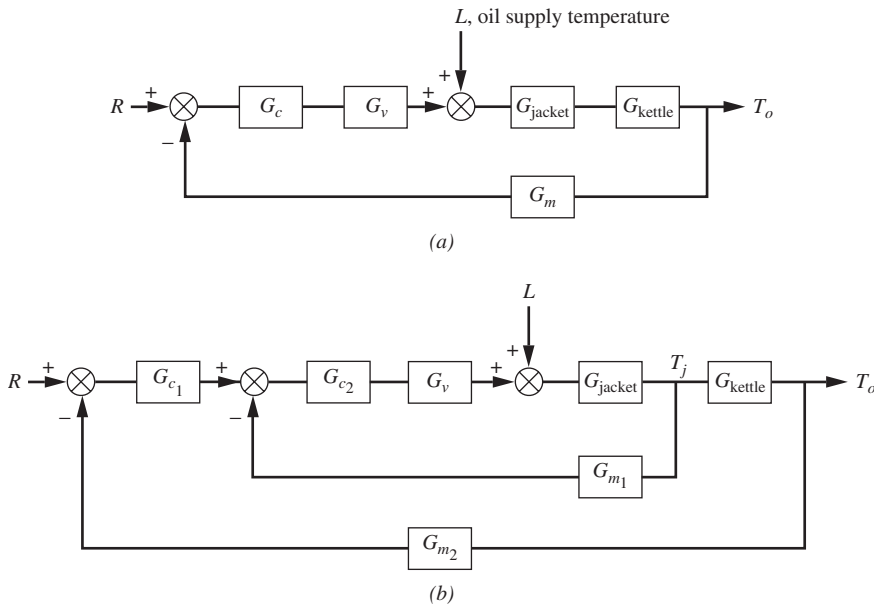
FIGURE 17–2
Block diagram: (a) single-loop conventional control; (b) cascade control.

Analysis of Cascade Control

To develop the closed-loop transfer functions for a cascade control system, consider the general block diagram shown in Fig. 17–3. In this diagram, the load disturbance U enters between two blocks of the plant, and the inner loop encloses this load disturbance.

To determine the transfer function $C/R,$ the inner loop is reduced to one block by the method shown in Chap. 11. The result is shown in Fig. 17–3b, and the block diagram of Fig. 17–3b can be used to give the result

$$\frac{C}{R} = \frac{G_{c_1} G_a G_3}{1 + G_{c_1} G_a G_3 H_1} \qquad (17.1)$$

where
$$G_a = \frac{G_{c_2} G_1 G_2}{1 + G_{c_2} G_1 G_2 H_2}$$

To obtain the transfer function relating output to load $C/U,$ the block diagram of Fig. 17–3a is rearranged by placing the transfer function $G_{c_2} G_1$ in the feedback paths of the primary and secondary loops; the new arrangement is shown in Fig. 17–4a. Since

FIGURE 17–3
Block diagram for cascade control for set point change.

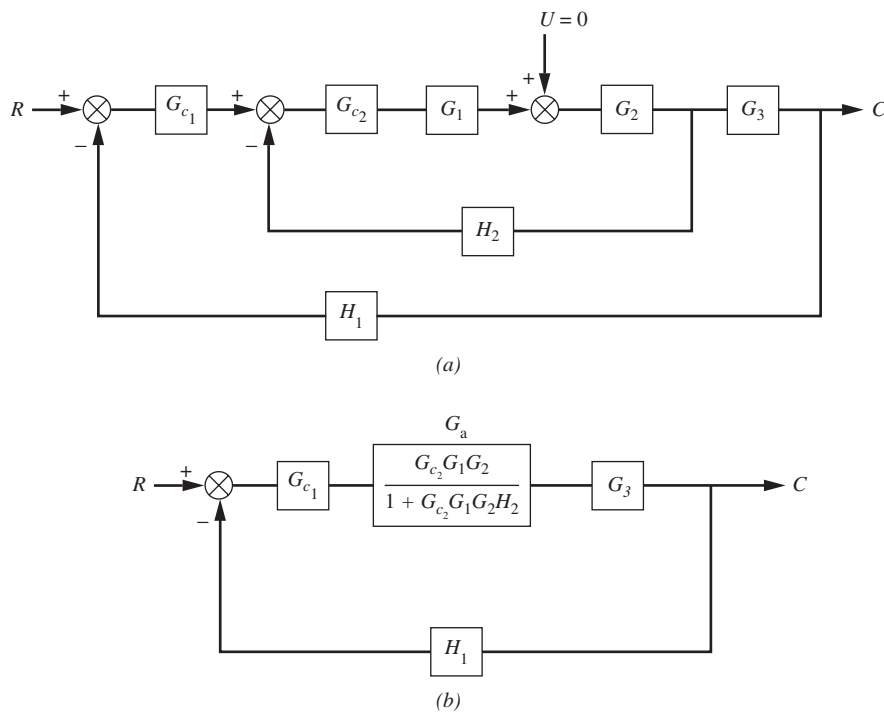$R = 0$ for the case under consideration, the block diagram can be redrawn as shown in Fig. 17–4b. This diagram, which has the same form as the one in Fig. 17–3a, can now be reduced to the form shown in Fig. 17–4c. Application of the rules of Chap. 11 to Fig. 17–4c finally gives

$$\frac{C}{U} = \frac{G_3}{G_1 G_{c_2}} \frac{G_a}{1 + G_a G_{c_1} H_1 G_3} \tag{17.2}$$

where G_a is the same as given in Eq. (17.1).

> **Example 17.1.**   To compare conventional control with cascade control, consider the conventional control system of Fig. 17–5a in which a third-order process is under PI control. A cascade version of this single-loop control system is shown in Fig. 17–5b in which an inner loop having proportional control encloses the load disturbance U.
>
> To obtain a response of the conventional control system for use in comparison with the response of the cascade system, the block diagram of Fig. 17–5a was simulated on a computer. The values of K_c and τ_I were chosen by trial and error to give the response to a step change in set point shown as curve I of Fig. 17–6; this response, which has a decay ratio of about $\frac{1}{4}$, was obtained with $K_c = 2.84$ and $\tau_I = 5$.
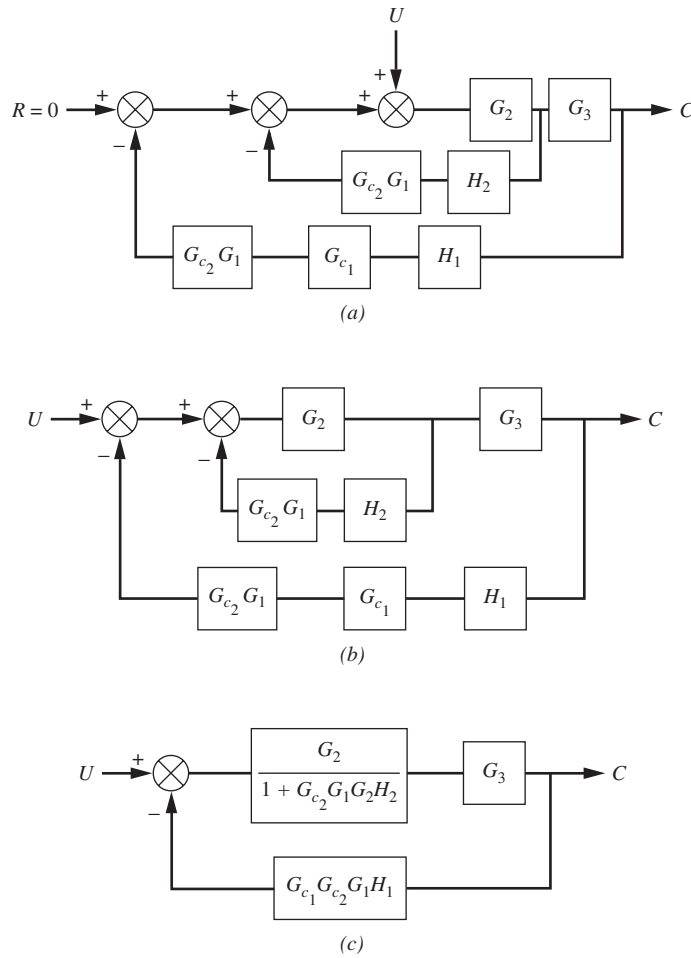
FIGURE 17–4
Block diagram for cascade control for load change.

The Ziegler-Nichols settings ($K_c = 3.65$ and $\tau_I = 3.0$) gave a set point response that was too oscillatory. Having obtained satisfactory controller settings ($K_c = 2.84$ and $\tau_I = 5.0$), we show the response of the system to a step change in U of 4 units as curve II of Fig. 17–7. The load response for no control (i.e., $K_c = 0$) is also shown as curve I for comparison.

The cascade control system of Fig. 17–5b was also simulated to obtain a load response. The controller gain K_{c_2} of the inner loop was chosen arbitrarily to be 10.0. This value was chosen to be high to obtain a fast-responding inner loop, a desirable situation for cascade control. Because of the introduction of the inner loop, the dynamics of the control system have changed, and it is necessary to tune the primary controller parameters for a good response to a step change in set point. By trial and error, primary controller settings of $K_{c_1} = 1.0$ and
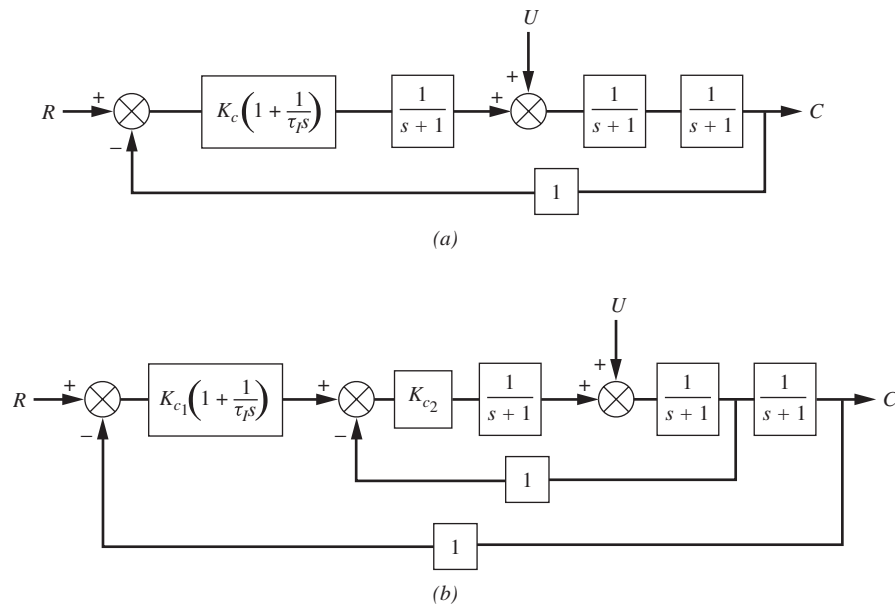
FIGURE 17–5
Block diagrams for Example 17.1: (*a*) single-loop conventional control; (*b*) cascade control.
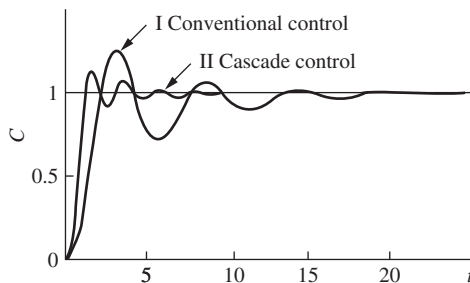
FIGURE 17–6
Responses to step change in set point for single-loop control and cascade control for Example 17.1.
I: Conventional control with $K_c = 2.84$ and $\tau_I = 5$; II: cascade control with $K_{c1} = 1.0$, $\tau_I = 0.63$, and $K_{c2} = 10$.

$\tau_I = 0.63$ were found that produced the response to a unit step in set point, shown as curve II in Fig. 17–6. The use of Ziegler-Nichols settings produced a less desirable response.

Using the controller parameters found from the step change in set point $(K_{c_1} = 1.0, \tau_I = 0.63)$, the response of the cascade system to a step change in load of 4 units was obtained and is shown as curve III of Fig. 17–7. As shown in Fig. 17–7, the load response for the cascade control system is far superior to the load response of the conventional control system. The maximum deviation of the cascade response has been reduced by a factor of about 4, and the frequency of oscillation has nearly doubled.

FIGURE 17–7
Responses to step change in set point for single-loop control and cascade control for Example 17.1. I: No control; II: conventional control with $K_c = 2.84$ and $\tau_I = 5$; III: cascade control with $K_{c1} = 1.0$, $\tau_I = 0.63$, and $K_{c2} = 10$.
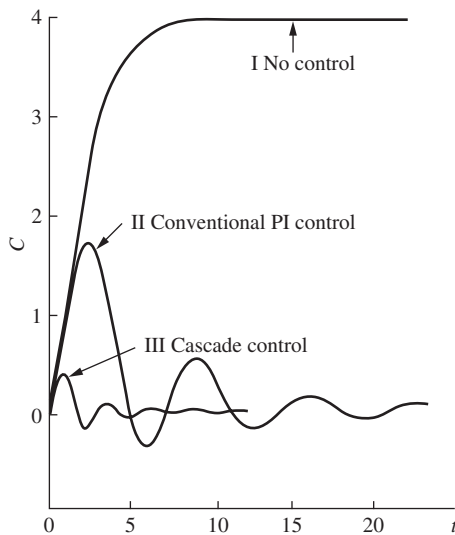
Generalizations

Cascade control is especially useful in reducing the effect of a load disturbance that moves through the control system slowly. The inner loop has the effect of reducing the lag in the outer loop, with the result that the cascade system responds more quickly with a higher frequency of oscillation. Example 17.2 will illustrate this effect of cascade control.

The choice of control action and tuning of the primary and secondary controllers for a cascade control system must be given careful consideration. The control action for the inner loop is often proportional with the gain set to a high value. The rationale for the use of proportional control rather than two- or three-mode control is that tuning is simplified and any offset associated with proportional control of the inner loop can be handled by the presence of integral action in the primary controller. The gain of the secondary controller should be set to a high value to give a tight inner loop that responds quickly to load disturbance; however, the gain should not be so high that the inner loop is unstable. Although the primary control loop can provide stable control even when the inner loop is unstable, it is considered unwise to have an unstable inner loop because the system will go unstable if the primary controller is placed in manual operation or if there is a break in the outer loop.

The action for the primary controller is generally PI or PID. The integral action is needed to reduce offset when sustained changes in load or set point occur. The problem of adjusting a primary controller is essentially the same as for a single-loop control system. Since the addition of the inner loop can change the dynamics of the outer loop significantly, the primary controller must be retuned when the inner loop is closed or when the secondary controller settings are changed.

Microprocessor-based controllers available today can implement cascade control very easily.

Example 17.2.   The claim is often made that cascade control gives a better response than conventional control because the lags in the outer loop are reduced. To illustrate this benefit, consider the conventional control and the cascade control of a third-order plant in Fig. 17–8*a* and *b*. The inner loop of the cascade system surrounds two of the first-order blocks in the plant. To simplify the discussion, the load disturbance is not shown since we are interested only in the closed-loop dynamics. The equivalent single-loop control system of the cascade system, shown in Fig. 17–8*c*, was obtained by the usual method for reducing a loop to a single block.

(a)

(b)

(c)
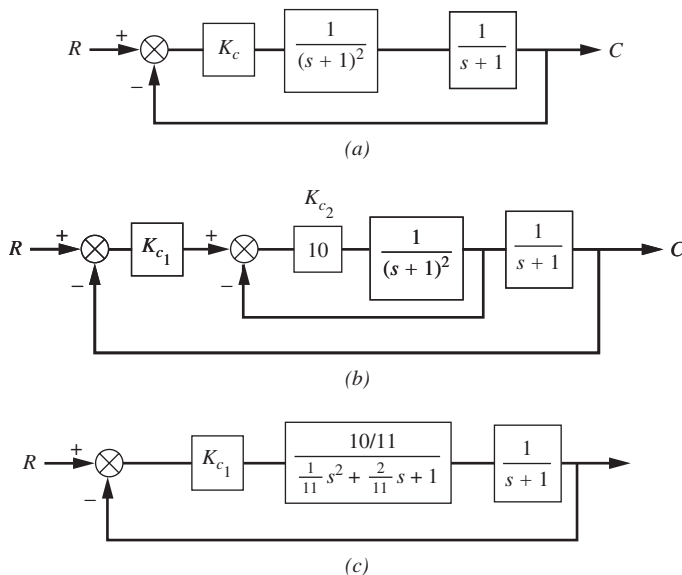
FIGURE 17–8
Block diagram for Example 17.2.

Comparing Fig. 17–8*a* with *c* shows that the use of cascade control has replaced a second-order critically damped system represented by the first two blocks of the plant [$1/(s + 1)^2$] with this underdamped second-order system

$$\frac{K}{\tau^2 s^2 + 2\zeta\tau s + 1}$$

where $K = \frac{10}{11}$

$$\tau = \sqrt{\frac{1}{11}}$$

$$\zeta = \sqrt{\frac{1}{11}}$$

This second-order underdamped system, for which τ and ζ are small, responds much faster than the critically damped second-order transfer function of the first two blocks of the open-loop system. Consequently, the cascade system will respond faster with a higher frequency of oscillation, as we have already seen in the simulated response of Fig. 17–6.

17.2   FEEDFORWARD CONTROL

If a particular load disturbance occurs frequently in a control process, the quality of control can often be improved by the addition of feedforward control. Consider the composition control system shown in Fig. 17–9a in which a concentrated stream of control reagent containing water and solute is used to control the concentration of the stream leaving a three-tank system. The stream to be processed passes through a pre-conditioning stirred tank where composition fluctuations are smoothed out before the outlet stream is mixed with control reagent. A three-tank system has been chosen for ease of computation in a numerical example that follows.

(a)

(b)
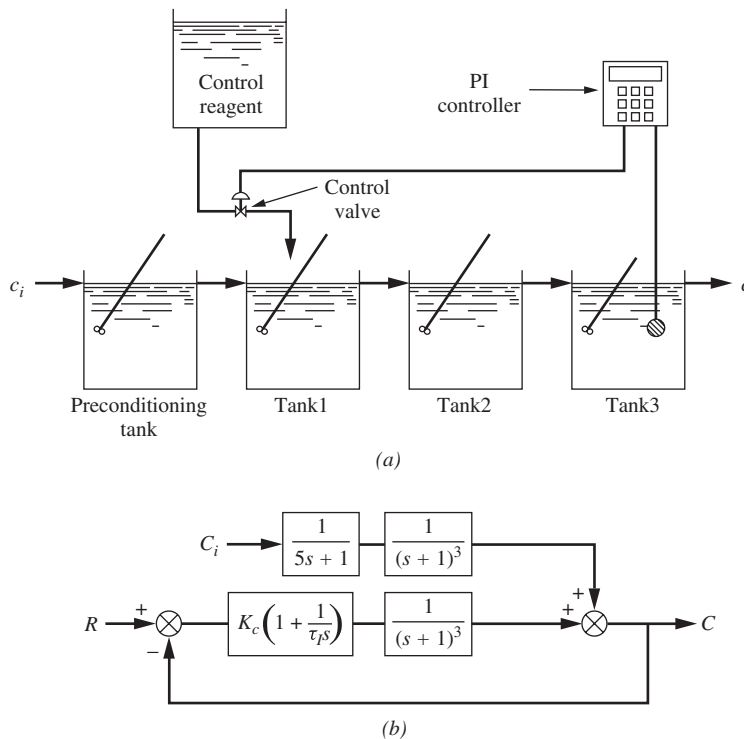
FIGURE 17–9
Composition control system: (*a*) physical process; (*b*) block diagram.

In the conventional feedback control system shown in Fig. 17–9a, the measurement of composition in the third tank is sent to a controller, which generates a signal that opens or closes the control valve, which in turn supplies concentrated reagent to the first tank. The block diagram corresponding to the control system of Fig. 17–9a is shown in Fig. 17–9b. [In Fig. 17–9a, concentration is denoted by c (lowercase letter). In the block diagram of the process in Fig. 17–9b, the symbol for concentration is denoted by C (capital letter) to denote a deviation variable. This use of symbols follows the procedure established in Chap. 4.]

To obtain some specific control system responses, numerical values of the time constants of the tanks have been chosen as shown in Fig. 17–9b. To study the response of this control system, the block diagram shown in Fig. 17–9b was simulated on a computer.

The values of K_c and τ_I were chosen by trial and error to give the response to a step change in set point shown in curve II of Fig. 17–10; this response, which has a decay ratio of about $\frac{1}{4}$, was obtained with $K_c = 2.84$ and $\tau_I = 5.0$. The Ziegler-Nichols settings ($K_c = 3.65$ and $\tau_I = 3.0$) give a set point response shown as curve I of Fig. 17–10, which is too oscillatory.
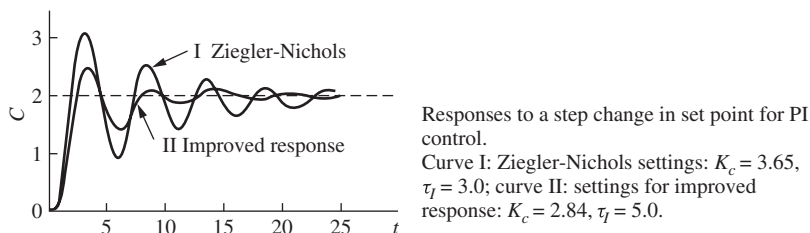

Responses to a step change in set point for PI control.
Curve I: Ziegler-Nichols settings: $K_c = 3.65$, $\tau_I = 3.0$; curve II: settings for improved response: $K_c = 2.84$, $\tau_I = 5.0$.

FIGURE 17–10

Having obtained satisfactory settings for the controller ($K_c = 2.84$, $\tau_I = 5.0$), we obtained the response of the system to a step change in C_i of 10 units, shown as curve I in Fig. 17–11. Note that the response is oscillatory and has a long tail. This response illustrates the fact that the feedback control system does not begin to respond until the load disturbance has worked its way through the forward loop and reaches the measuring element, with the result that the composition can move far from the set point during the transient.

Responses to a step change in load for feedforward-feedback control.
Curve I: PI control with $K_c = 2.84$, $\tau_I = 5.0$
Curve II: FF control with $K_c = 2.84$, $\tau_I = 5.0$, $G_f = -1/(5s + 1)$
Curve III: FF control with $K_c = 2.84$, $\tau_I = 5.0$, $G_f = -1$
Curve IV: FF control with $K_c = 2.84$, $\tau_I = 5.0$, $G_f = -0.5$
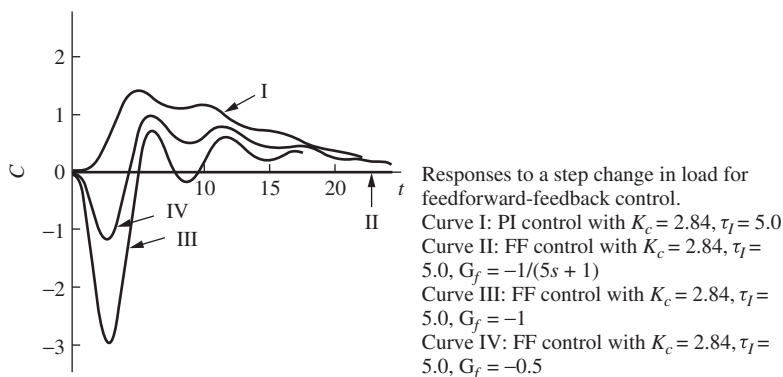
FIGURE 17–11

If the change in load disturbance C_i can be detected as soon as it occurs in the inlet stream, this information can be fed forward to a second controller that adjusts the control valve in such a way as to prevent any change in the outlet composition from the set point. A controller that uses information fed forward from the source of the load disturbance is called a *feedforward* controller. The block diagram that includes the feedforward controller G_f as well as the feedback controller G_c is shown in Fig. 17–12.

FIGURE 17–12
Control system with feedforward and feedback controllers.

Analysis of Feedforward Control

The response of C to changes in C_i and R can be written from Fig. 17–12 as follows:

$$C(s) = G_1(s)G_p(s)C_i(s) + G_f(s)G_p(s)C_i(s) + G_c(s)G_p(s)E(s) \quad (17.3)$$

where $E(s) = R(s) - C(s)$. To determine the transfer function of $G_f(s)$ that will prevent any change in the control variable C from its set point R, which is 0, we solve Eq. (17.3) for $G_f(s)$ with $C = 0$ and $R = 0$. The result is
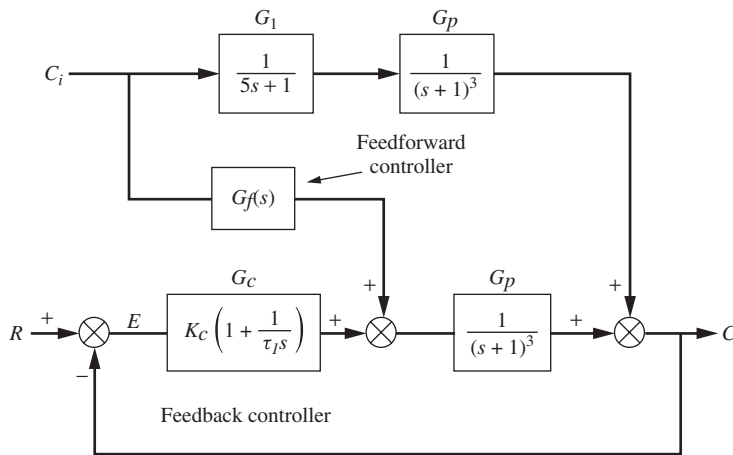
$$G_f(s) = -G_1(s) \quad (17.4)$$

For the example under consideration in Fig. 17–12,

$$G_f(s) = \frac{-1}{5s + 1} \quad (17.5)$$

This transfer function can be implemented easily with existing control hardware.

If the load response of the control system in Fig. 17–12, with $G_f(s)$ given by Eq. (17.5), were obtained for a step change in C_i, there would be no deviation of C from the set point (i.e., perfect control). This response is shown as curve II in Fig. 17–11, which of course is a horizontal line at $C = 0$.

Rather than use the $G_f(s)$ of Eq. (17.5) in the feedforward controller, one can try using only the constant term of $G_f(s)$, that is,

$$G_f(s) = -1$$

The response for $G_f = -1$ gives curve III in Fig. 17–11; this response has a very large undershoot before the feedback controller returns C to the set point. If we try using $G_f(s) = -0.5$, we obtain curve IV of Fig. 17–11; the undershoot is less in this case,

but the response is still unsatisfactory. As shown by curves III and IV, omitting the dynamic part of $G_f(s)$ can give very poor results. The success of using a feedforward controller depends on accurate knowledge of the process model, a luxury that may not be available in many applications.

Implementing Feedforward Transfer Functions

In applications of feedforward control, $G_f(s)$ may take the form of a lead expression, such as $G_f(s) = 1 + \tau_f s$. When this occurs, it is necessary to approximate $1 + \tau_f s$ by a lead-lag expression, such as

$$G_f(s) = \frac{1 + \tau_f s}{1 + \beta \tau_f s}$$

where $\beta \ll 1$ (see Prob. 10.3 for an example of this). To see how $G_f(s)$ takes the form of a lead expression, consider the load disturbance c_i of Fig. 17–9 to enter tank 2. Since no change in concentration occurs in the stream entering the preconditioning tank, we may eliminate it from the diagram for the case under consideration to obtain the diagram in Fig. 17–13.
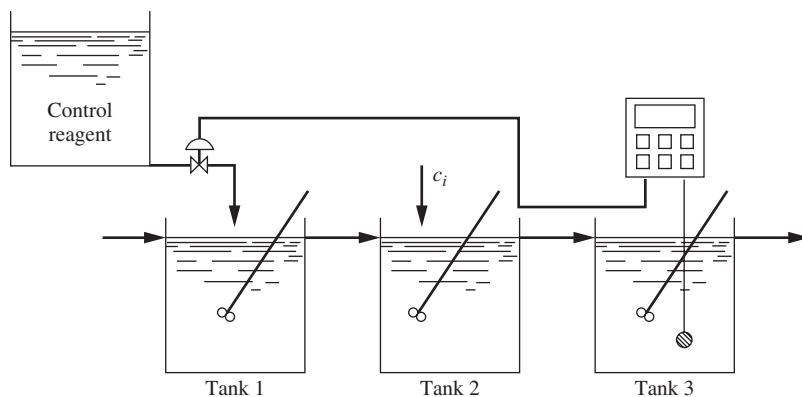
FIGURE 17–13
Composition control with disturbance to second tank.

Adding feedforward control and feedback control to the system in Fig. 17–13 gives the block diagram of Fig. 17–14. The diagram shown in Fig. 17–14 is the same as that in Fig. 17–12 with the exception that the disturbance C_i enters tank 2 instead of the preconditioning tank. As shown previously, the response of C to a change in C_i and R can be written directly from Fig. 17–14 as

$$C(s) = G_1(s)C_i(s) + G_f(s)G_p(s)C_i(s) + G_c(s)G_p(s)E(s) \qquad (17.6)$$

where $E(s) = R(s) - C(s)$.
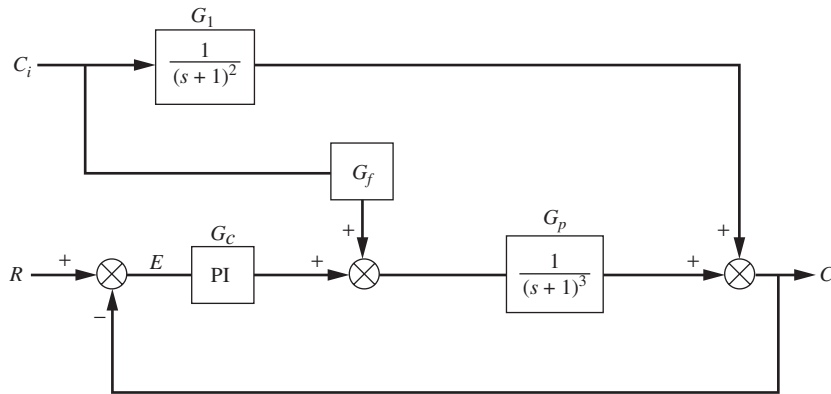
FIGURE 17–14
Feedforward-feedback control for system in Fig. 17–13.

In order for C not to change from the set point R, which is 0, we solve Eq. (17.6) for $G_f(s)$ with $C = 0$ and $R = 0$ to obtain

$$G_f(s) = -\frac{G_1(s)}{G_p(s)} \tag{17.7}$$

Introducing the expressions for $G_1(s)$ and $G_p(s)$ from Fig. 17–14 into Eq. (17.7) gives

$$G_f(s) = -(s + 1) \tag{17.8}$$

It is not practical to implement $-(s + 1)$. To see this, consider the response of $-(s + 1)$ to a step change as shown in Fig. 17–15.
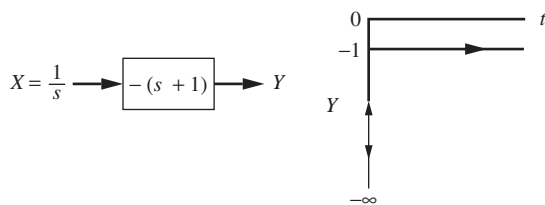
FIGURE 17–15
Step response for $-(s + 1)$.

There is no hardware that will produce an impulse as shown in Fig. 17–15; however, one can approximate $-(s + 1)$ by means of a lead-lag transfer function of the form

$$\frac{Y(s)}{X(s)} = -\frac{\tau_f s + 1}{\beta \tau_f s + 1} \tag{17.9}$$

where

$$\beta \ll 1$$

If we let $\beta = 0.1$ and $\tau_f = 1$ for the control system under consideration, we obtain as an approximation to Eq. (17.8)

$$G_f(s) = -\frac{s+1}{0.1s+1} \tag{17.10}$$

The response of this transfer function to a step input is shown in Fig. 17–16. The effect of this transfer function $-(s+1)/(0.1s+1)$ on the output of the feedforward controller for a step change in load is to give a sudden drop in flow followed by a fast exponential increase in the flow to a steady-state flow of -1. Note that for the parameters chosen for the transfer functions in Fig. 17–14, a unit increase in C_i must eventually be compensated by a unit decrease in the signal from the feedforward controller if there is to be no change in the process output. The sudden, initial drop in flow may be too abrupt for the control hardware, in which case the output would saturate. In practice, β can be increased (perhaps to 0.5) to reduce the magnitude of the initial drop.

$$-Y = (10-1)\,e^{-t/0.1} + 1$$
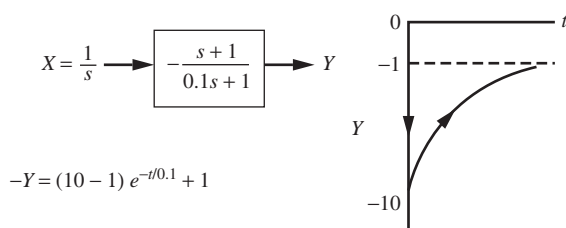
FIGURE 17–16
Step response for $-(s+1)/(0.1s+1)$.

The effect of using $G_f(s) = -(s+1)/(0.1s+1)$ with feedback control is shown in Fig. 17–17. The responses shown, which were obtained by simulation, are for a unit-step change in C_i. Curve I is for the case of feedback control only with $K_c = 2.84$ and $\tau_I = 5.0$. Curve II is for feedforward-feedback control using Eq. (17.10) for $G_f(s)$ and $K_c = 2.84$ and $\tau_I = 5.0$. One can see that the overshoot for the feedforward-feedback response has been reduced significantly.

FIGURE 17–17
Comparison of conventional feedback control with feedforward-feedback control for system in Fig. 17–14. Curve I: PI control with $K_c = 2.84$ and $\tau_I = 5$. Curve II : Feedforward-feedback control with $K_c = 2.84$, $\tau_I = 5$, and $G_f = -(s+1)/(0.1s+1)$.

Tuning Rules for Feedforward Feedback Control

In the practical application of feedforward control, one does not have a block diagram with transfer functions as shown in Figs. 17–12 and 17–14. For such a practical situation, one can still tune the feedforward controller by introducing a step change in the disturbance that enters the feedforward controller (C_i in Fig. 17–14) and then applying some tuning rules. The rules to be discussed here are from a training film on feedforward control produced by Foxboro Co. (1978).
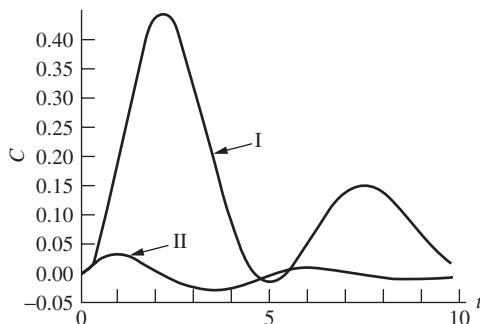
Feedforward Rules

In describing these rules, reference will be made to the general block diagram for a feedforward-feedback system shown in Fig. 17–18. It is assumed that $G_f(s)$ will be a lead-lag transfer function of the form

$$G_f(s) = \frac{K_f(T_1 s + 1)}{T_2 s + 1} \tag{17.11}$$

where K_f = steady-state gain of feedforward controller

       T_1, T_2 = time constants of dynamic part of feedforward controller
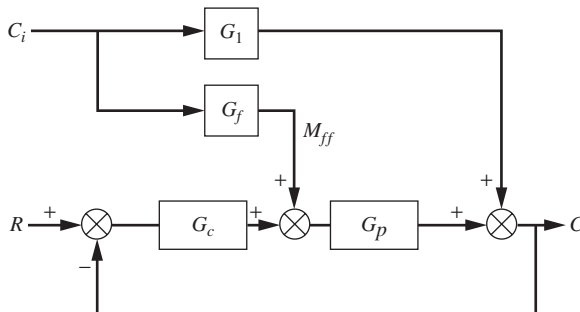
FIGURE 17–18
Feedforward-feedback control system.

Commercial microprocessor-based controllers provide this lead-lag transfer function.

The tuning rules listed below are explained with the help of Fig. 17–19. In that figure, a unit step is selected for the disturbance C_i, and K_f has been taken as -1. In practice, K_f will, of course, depend on the particular process being controlled.

1. Remove the control action in $G_c(s)$ by setting the controller to manual.
2. Set the feedforward controller to the computed steady-state gain (K_f) necessary to compensate ultimately for a step change in C_i. This means that the dynamic portion of $G_f(s)$ will be removed, and only the constant term (K_f) will remain.
3. Make a step change in C_i and observe the open-loop transient of C. The general shapes of the response to be expected are shown in Fig. 17–19.
4. If the response shown in Fig. 17–19a occurs, lead must predominate in $G_f(s)$ of Eq. (17.11) (i.e., $T_1 > T_2$). If the response of Fig. 17–19b occurs, lag must predominate in $G_f(s)$ (i.e., $T_1 < T_2$). The values of T_1 and T_2 in Eq. (17.11) are found by use of the information in Table 17.1. The value of K_f in Eq. (17.11) has been obtained in step 2.

FIGURE 17–19
Open-loop response to determine lead-lag time constants in feedforward tuning rules: (*a*) Lead must predominate in G_f; (*b*) lag must predominate in G_f.
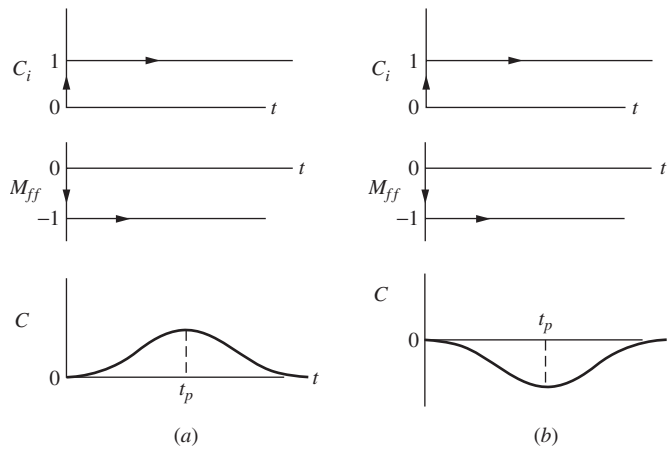
TABLE 17.1
Tuning parameters for feedforward control

Predominant mode	T_1	T_2
Lead	$1.5t_p$	$0.7t_p$
Lag	$0.7t_p$	$1.5t_p$

$$G_f(s) = K_f \frac{T_1 s + 1}{T_2 s + 1}$$

Example 17.3 will help clarify the use of these tuning rules.

Example 17.3.   *Use of feedforward tuning rules.* Apply the feedforward tuning rules to the system in Fig. 17–14. Since this example is concerned with the application of the tuning rules to a system for which a mathematical model is not generally available, the reader should assume that the transfer functions for $G_1(s)$ and $G_p(s)$ in Fig. 17–14 are unknown. The determination of $G_f(s)$ is to be obtained solely by information from open-loop transients.

We must first determine the steady-state gain K_f for the system of Fig. 17–14. If a step change in C_i is made, C will undergo a transient and eventually level out at a steady-state value. If the controller parameters are properly selected, the value of C at the end of the transient will be the same as it was before the transient occurred. By computation or experiment, one can determine the value of K_f needed to obtain no change in C. For the system in Fig. 17–14, one can see that K_f of Eq. (17.11) must be equal to -1.

We must now apply the feedforward tuning rules to obtain T_1 and T_2 in Eq. (17.11). After removing the feedback controller action $[G_c(s)]$, we have the equivalent diagram shown in Fig. 17–20. A unit-step change in C_i produces the transient for C shown as curve I in Fig. 17–22. Comparing the 2 shape of the transient with those of Fig. 17–19, we see that lead must predominate in $G_f(s)$. The peak value occurs at $t_p = 2$. Applying the rules in Table 17.1 gives

$$T_1 = 1.5t_p = 3$$

$$T_2 = 0.7t_p = 1.4$$

The feedforward controller transfer function is therefore

$$G_f(s) = \frac{-(3s + 1)}{1.4s + 1} \tag{17.12}$$

FIGURE 17–20
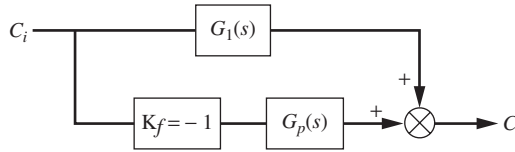Open-loop feedforward test to determine parameters for G_f.

It is of interest to show the response of C for feedforward only when the feedforward transfer function of Eq. (17.12) is used. MATLAB and Simulink can be used to simulate the result for a unit-step change in C_i (Fig. 17–21). It is shown as curve II in Fig. 17–22.
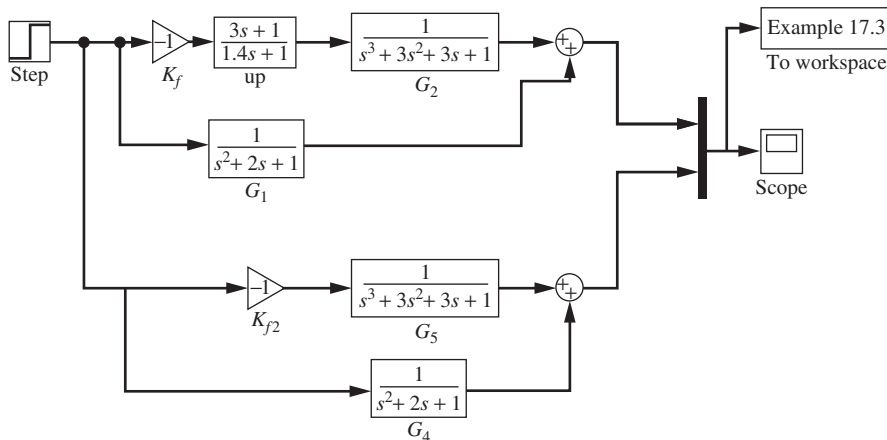
FIGURE 17–21
Simulink model for simulating Example 17.3.
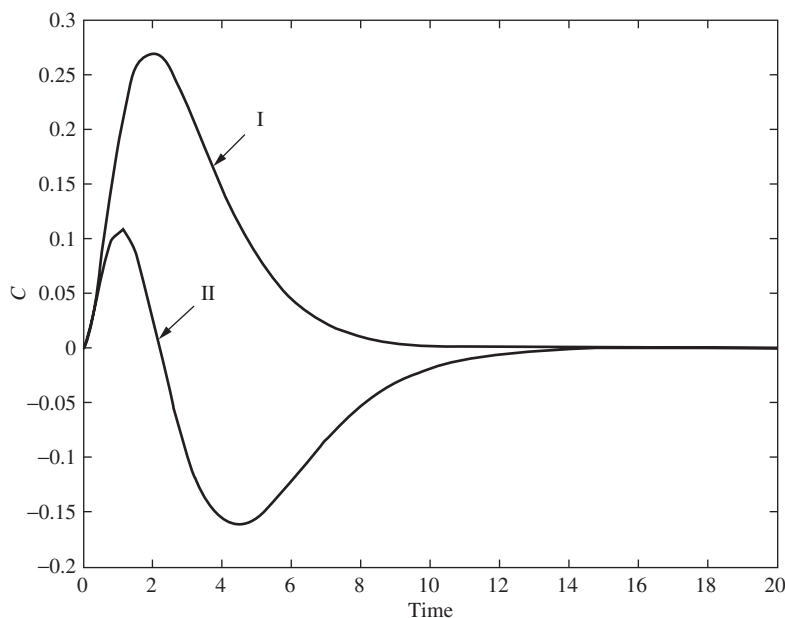
FIGURE 17–22
Open-loop response for step change in C_i for Example 17.3. Curve I: $G_f = -1$,
curve II: $G_f = -(3s + 1)/(1.4s + 1)$.

When the $G_f(s)$ of Eq. (17.12) is used and the controller parameters for $G_c(s)$ are
$K_c = 2.84$ and $\tau_I = 5.0$, the feedforward-feedback response to a unit-step change
in C_i is shown as curve II in Fig. 17–23. For comparison, the response for feed-
back control only is also shown in Fig. 17–23.
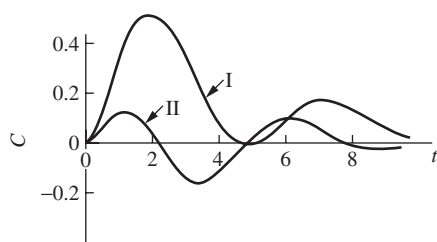
FIGURE 17–23
Comparison of conventional feedback control with feedforward-feedback control for Example 17.3.
Curve I: PI control with $K_c = 2.84$ and $\tau_I = 5$. Curve II: Feedforward-feedback control with
$K_c = 2.84$, $\tau_I = 5$, and $G_f = -(3s + 1)/(1.4s + 1)$.

17.3   RATIO CONTROL

An important control problem in the chemical industry is the combining of two or more
streams to provide a mixture having a desired ratio of components. Examples of this
mixing operation include the blending of reactants entering a chemical reactor or for the
injection of a fuel/air mixture into a furnace.

Figure 17–24*a* shows an example of a ratio control system. It depicts a control system for blending two liquid streams *A* and *B* to produce a mixed stream having the ratio K_r in units of mass *B*/mass *A*. Stream *A,* which is uncontrolled, is used to adjust the flow of stream *B* so that the desired ratio is maintained. The measured signal for stream *A* is multiplied by the desired ratio K_r to provide a signal that is the set point for the flow control loop for stream *B*. The parameter K_r can be adjusted to the desired value. Control hardware is available to perform the multiplication of two control signals.
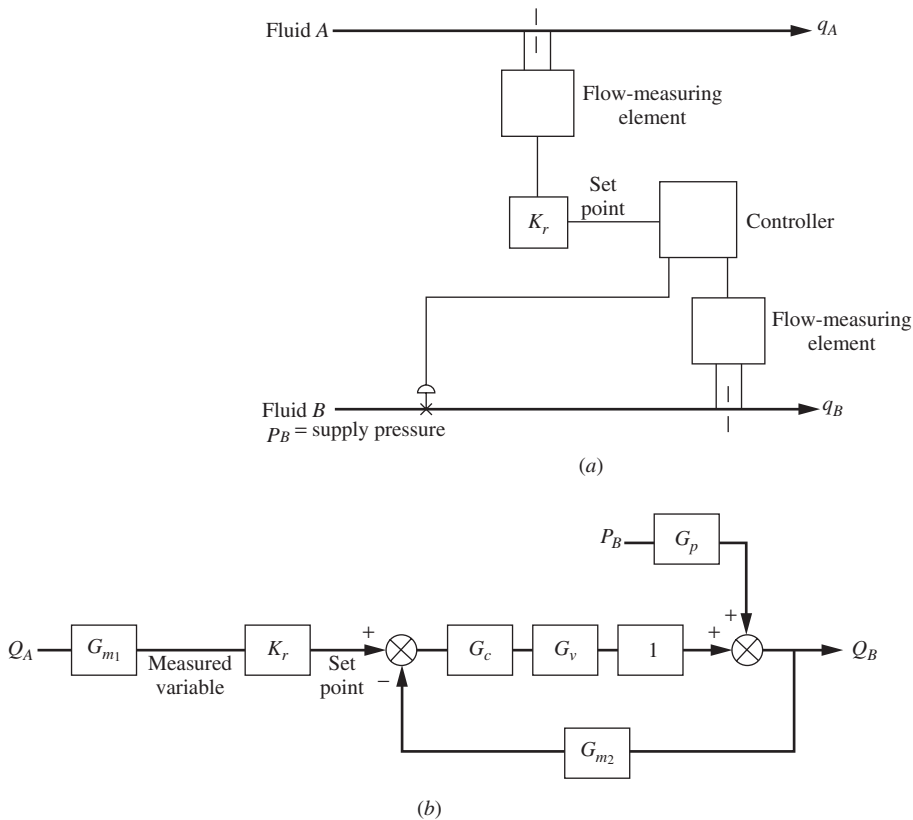
FIGURE 17–24
(*a*) Ratio control system; (*b*) block diagram for ratio control (set point $= G_{m_1} K_r Q_A$).

A block diagram of the ratio control system is shown in Fig. 17–24*b*. In a flow control loop, the dynamic elements consist of the controller, the flow-measuring element, and the control valve. For incompressible fluids, there is no lag between the change in valve position and the corresponding flow rate. For this reason, the transfer function between the valve and the measurement of flow rate is simply unity. The block diagram also shows a transfer function G_p that relates the flow rate of *B* to the supply pressure of *B*. A transfer function G_{m_1} is also shown that represents the dynamic lag of the flow-measuring element for stream *A*.

From the block diagram, the flow of B may be written

$$Q_B = \frac{G_{m_1} K_r G_c G_v}{1 + G_c G_v G_{m_2}} Q_A + \frac{G_p}{1 + G_c G_v G_{m_2}} P_B$$

The control action for a flow control system is usually PI. The integral action is needed to eliminate offset and thereby establish a precise ratio of the mixed streams of A and B. Derivative action is usually avoided in flow control because the signal from a flow-measuring element is inherently noisy. The presence of derivative action would amplify the noise and give poor control.

An example system from the LOOP-PRO software package (Cooper, 2005) is the control of the air/fuel ratio being fed to a process furnace. A description of the process from "Practical Process Control" that accompanies the software is as follows.

A furnace burns natural gas to heat a process liquid. The measured process variable is the temperature of the process liquid as it exits the furnace. To maintain temperature, controllers adjust the feed rate of combustion air and fuel to the firebox using a ratio control strategy. The flow rate of the process liquid acts as a load disturbance to the process. . . . For the furnace, the independent stream is the combustion air flow rate and the dependent stream is the fuel flow rate. Note that while air flow rate is considered the independent stream for ratio control, its flow rate is specified by the temperature controller on the process liquid exiting the furnace.
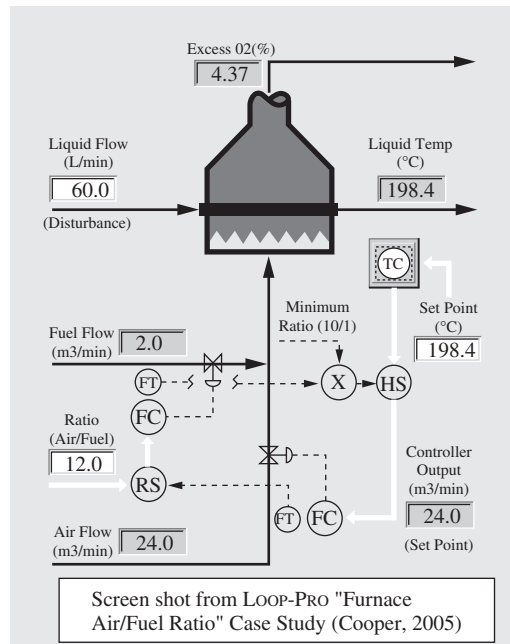
The process diagram is shown in Fig. 17–25.

Screen shot from LOOP-PRO "Furnace Air/Fuel Ratio" Case Study (Cooper, 2005)

FIGURE 17–25

This computer simulation by Cooper is an excellent means of exploring the behavior of a ratio control system. Additionally, he points out some interesting environmental and safety considerations that factor into the design of such a control system, such as

- Avoiding unburned hydrocarbons and carbon monoxide in the flue gases (too fuel-rich)
- Avoiding excess energy losses (too much air, a fuel-lean situation)
- Avoiding an explosive situation if the airflow fails (fail-safe considerations)

This interesting example provides some insights into the application of ratio control in a real-world situation.

17.4   DEAD-TIME COMPENSATION (SMITH PREDICTOR)

Processes that contain a large transport lag $[\exp(-\tau_D s)]$ can be difficult to control because a disturbance in set point or load does not reach the output of the process until τ_D units of time have elapsed. The control strategy to be described here, which is also known as dead-time compensation, attempts to reduce the deleterious effect of transport lag. Dead-time compensation, which is also referred to as a Smith predictor, was first described by O. J. M. Smith (1957).

Consider the single-loop control system of Fig. 17–26 in which the process transfer function $G_p(s)$ is to be modeled by

$$G_p(s) = G(s)e^{-\tau_D s} \tag{17.13}$$
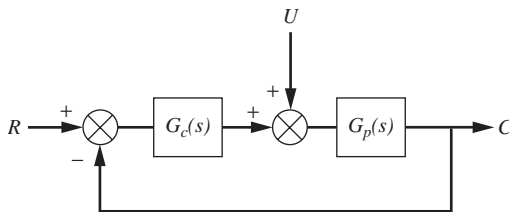
FIGURE 17–26
Control system.

The right side of Eq. (17.13) is the product of a transport lag $[\exp(-\tau_D s)]$ and a transfer function $G(s)$, which has minimum phase characteristics, such as $1/(\tau s + 1)$. For convenience in developing the dead-time compensation method, only a change in set point R will be considered. If a step change is made in R, the disturbance will not break through and appear at C until τ_D units of time elapse. Up to time τ_D, no control action occurs, with the result that the overall closed-loop response will be sluggish and

generally unsatisfactory. To overcome this difficulty, Smith suggested that $G_p(s)$ be modeled according to Eq. (17.13) and that additional feedback paths be inserted into Fig. 17–26 as shown in Fig. 17–27a.

If $G_p(s)$ is modeled exactly by Eq. (17.13), a close study of Fig. 17–27a shows that the signals entering comparator A will be identical; as a result, the signals cancel and cause the output of comparator A to be zero. The net effect is to completely eliminate the outer feedback path; this simplification is shown in Fig. 17–27b.
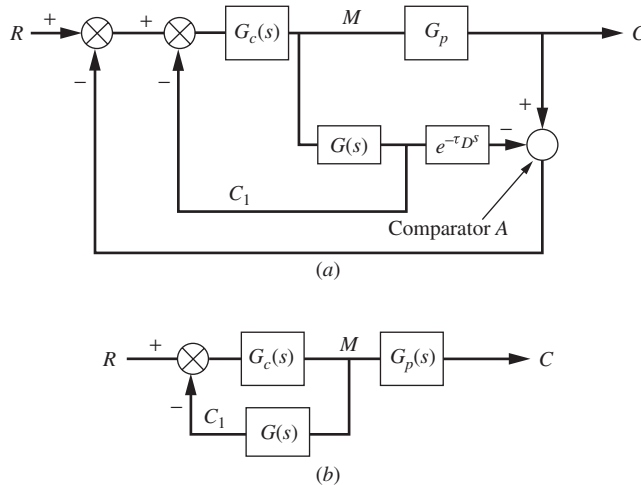
FIGURE 17–27
(a) Dead-time compensation (Smith predictor) block diagram; (b) equivalent diagram for part (a) when $G_p = G(s)e^{-\tau_D s}$.

The system of Fig. 17–27b is now much easier to control because the transport lag is not present in the loop. Of course, in the real system the transport lag is still present; we have eliminated it in a mathematical sense from the feedback path by the additional feedback paths of Fig. 17–27a and the assumption that the process transfer function $G_p(s)$ can be modeled exactly as shown in Eq. (17.13). To achieve the simplification suggested by Fig. 17–27b, we must now face reality and realize that the signal C_1 in Fig. 17–27b is not available to feed back. Only the signal C can be measured and fed back to the controller. In terms of controller hardware implementation, the diagram of Fig. 17–27a is redrawn in Fig. 17–28a to show which portion of the diagram will be implemented with controller hardware. Figure 17–28b, which is another way to represent Fig. 17–28a, is a form sometimes presented in the literature for dead-time compensation. The reader may legitimately ask whether hardware exists to actually implement what is shown within the dotted lines in Fig. 17–28. Until the appearance of microprocessor-based controllers, the answer was no. However, today many commercially available controllers provide dead-time compensation [$\exp(-\tau_D s)$] and $G(s)$ in the form of a first-order lag [$1/(\tau_s + 1)$].

The recommended procedure for applying dead-time compensation is as follows:

1. Model $G_p(s)$ by using a first-order plus dead-time (FOPDT) model

$$\frac{1}{\tau s + 1} e^{-\tau_D s}$$

In this step, we have chosen $G(s)$ of Fig. 17–28a to be first-order. Many processes in chemical engineering can be modeled by a first-order lag with dead time.

2. By means of appropriate hardware, implement the controller portion of Fig. 17–28a or b. If $G_p(s)$ can be exactly modeled by a first-order process with dead time, the response of the control system in Fig. 17–28 will be equivalent to the response obtained for the system in Fig. 17–27b in which the loop involves the control of a first-order process. In most practical situations, there will be some mismatch between $G_p(s)$ and its first-order with dead-time model. The greater the mismatch, the greater the deterioration in control response from the ideal situation of Fig. 17–27b. The application of the dead-time compensation technique and the effect of mismatch between $G_p(s)$ and its model will be illustrated in Example 17.4.

(a)

(b)
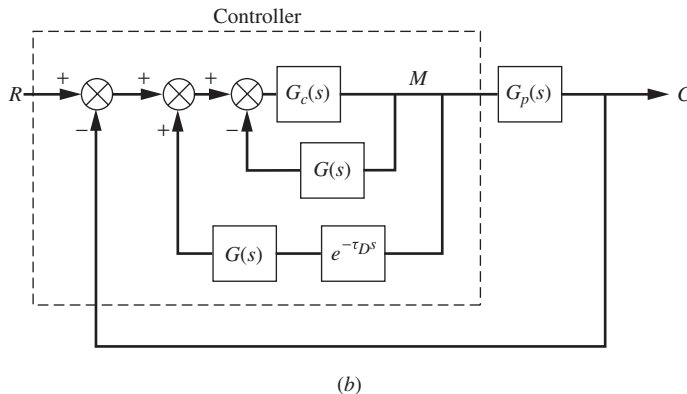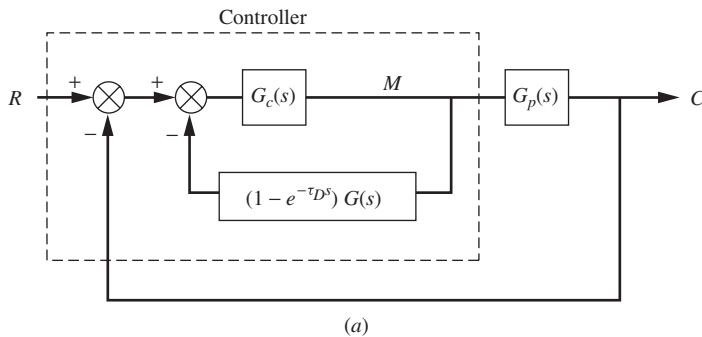
FIGURE 17–28
Hardware implementation of dead-time compensation.

Example 17.4. *Dead-time compensation.* Consider the control system shown in Fig. 17–29 in which the process is fourth-order; thus

$$G_p(s) = \left(\frac{1}{s+1}\right)^4$$

In a practical situation, we would not know the transfer function of the process. In this example, we have taken the process model to be fourth-order to provide a system sufficiently complex to show considerable transfer lag.
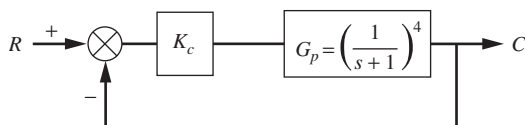
FIGURE 17–29
Control system for Example 17.4.

One can show for the system in Fig. 17–29 that the ultimate gain and the corresponding period are $K_{c_u} = 4.0$ and $P_u = 2\pi$. Using the Ziegler-Nichols rules, one gets $K_c = 2.0$. The response for a unit-step change in set point for $K_c = 2$ is shown in curve I of Fig. 17–32. Notice that the decay ratio is about $\frac{1}{4}$.

We now use the dead-time compensation method to control the process in Fig. 17–29. If we fit the step response of $(s+1)^{-4}$ to a first-order with dead-time model, we obtain

$$\frac{1.03}{2.62s+1}e^{-1.79s}$$

This model was obtained from a unit-step response using a least-squares fit procedure. We can now draw the diagram for the dead-time compensation system as shown in Fig. 17–30.
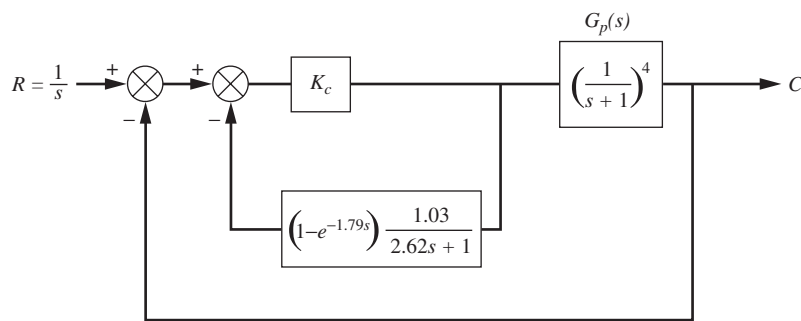
FIGURE 17–30
Dead-time compensation for Example 17.4.

The system shown in Fig. 17–30 was simulated by computer to compare the responses of the two control systems as shown in Fig. 17–32. The Simulink block diagram for the simulation is shown in Fig. 17–31.
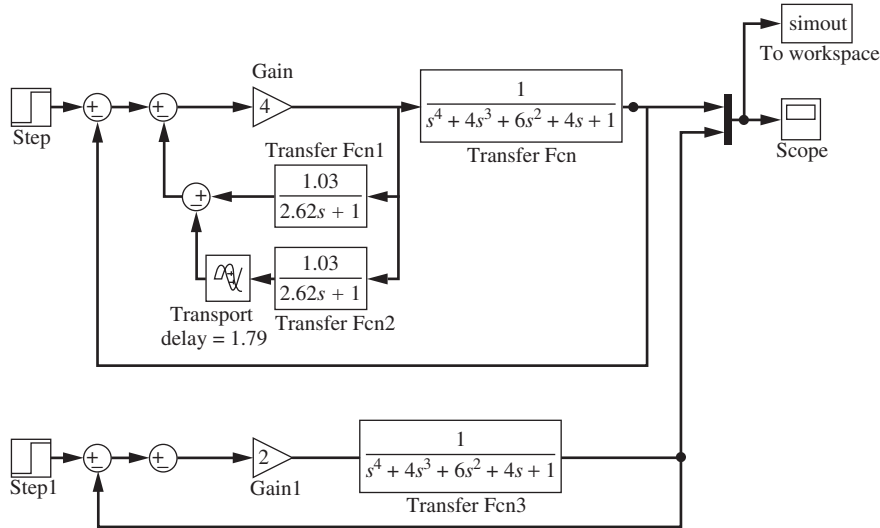
FIGURE 17–31
Simulink model for Example 17.4.

Using a K_c of 2.0 (the Ziegler-Nichols value) for the conventional control, we see from curve I that the response is quite oscillatory and has an offset of 0.333 as required for this value of gain. Using a K_c of 4.0 for the dead-time compensation, we see that the response is less oscillatory and the offset is 0.20. Note that if a K_c of 4.0 were applied to the conventional control system, the system would be on the verge of instability since a K_c of 4.0 is the ultimate gain.
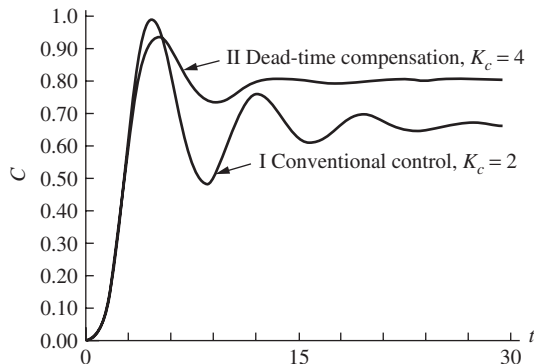
FIGURE 17–32
Comparison of response for conventional control with response for dead-time compensation for Example 17.4.

In conclusion, the dead-time compensation has permitted the use of a higher value of K_c, reduced the offset, and produced a less oscillatory response. The dead-time compensation response shown in Fig. 17–32 can be improved by adding integral action to the controller and tuning the controller parameters.

To successfully apply dead-time compensation to the control of a process, one must have an accurate model of the process, such as a first-order with dead-time model. The parameters in this model (τ and τ_D) can be considered as controller parameters along with the controller parameters of $G_c(s)$. For the case of dead-time compensation with proportional control in Example 17.4, we actually have three controller parameters: K_c, τ_D, and τ. If the process dynamics [$G_p(s)$] changes, all three parameters may need adjustment to achieve good control.

17.5   INTERNAL MODEL CONTROL

Internal model control (IMC), which is based on an accurate model of the process, leads to the design of a control system that is stable and robust. A robust control system is one that maintains satisfactory control in spite of changes in the dynamics of the process. In applying the IMC method of control system design, the following information must be specified:

- Process model
- Model uncertainty
- Type of input (step, ramp, etc.)
- Performance objective (integral square error, overshoot, etc.)

In many industrial applications for control systems, none of the above items is available, with the result that the system usually performs in a less than optimum manner. Determining the mathematical model and its uncertainty can be a difficult task. When the process is not sufficiently understood to obtain a mathematical model by applying fundamental principles, one must obtain a model experimentally. A discussion on the modeling of a process is presented in Chap. 18. The choice of a performance objective is subjective and often arbitrary. In the IMC method, the integral square error is implied.

A simple description of the IMC method will be presented here. The interested reader is advised to consult the book by Morari and Zafiriou (1989) for a full treatment of internal model control. The literature on IMC is difficult to understand without a good foundation in control theory and mathematics. A full treatment of IMC is beyond the scope of this text. It is hoped that the simple treatment given here will stimulate interest in this important area of process control.

Internal Model Control Structure

A block diagram of an IMC system is shown in Fig. 17–33a. Notice that the diagram is similar to the diagram for the Smith predictor method shown in Fig. 17–27a. In this diagram, G is the transfer function of the process and G_m is the model of the process. Although G and G_m are called the transfer functions of the process, they actually include

the valve and the process. The transfer function of the measuring element is taken as 1.0. The portion of the diagram that is implemented by the computer includes the IMC controller and the model; this portion is surrounded by the dotted boundary.

To compare the IMC structure of Fig. 17–33a with the conventional control structure, the diagram of Fig. 17–33a has been rearranged as shown in Fig. 17–33b. For convenience, the transfer function through which the load U passes has been omitted. We show only the output from the load block (U_1).
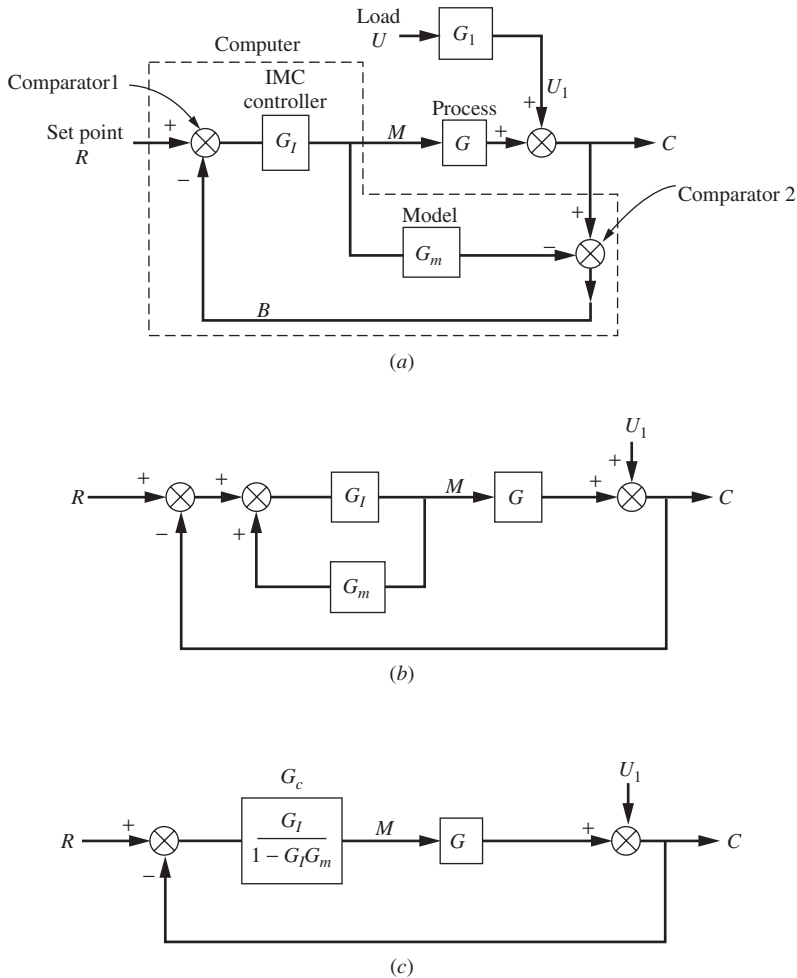

FIGURE 17–33
Internal model control structures: (a) basic structure; (b) alternate structure; (c) structure equivalent.

We may use the structure in Fig. 17–33b to relate the IMC controller to the conventional controller. Replacing the inner loop of Fig. 17–33b with a single block gives the structure shown in Fig. 17–33c. Since this structure is the conventional single-loop

control structure, we can identify the single controller block as G_c. After one designs the IMC controller (G_I) by the method to be described, one can determine the equivalent conventional controller G_c by the relation

$$G_c = \frac{G_I}{1 - G_I G_m} \tag{17.14}$$

For the structure shown in Fig. 17–33a, one can show that

$$C = U_1 + \frac{GG_I}{1 + G_I(G - G_m)}(R - U_1) \tag{17.15}$$

If the model exactly matches the process (i.e., $G_m = G$), the only signal entering comparator 1 in Fig. 17–33a is U_1. (The signals from G and G_m are equal and cancel each other in going through comparator 2.) Since U_1 is not the result of any processing by the transfer functions in the forward loop, U_1 is not a feedback signal but an independent signal that is equivalent to R in its effect on the output C. In fact, there is no feedback when $G = G_m$, and we have an open-loop system as shown in Fig. 17–34. In this case the stability of the control system depends only on G_I and G_m. If G_I and G_m are stable, the control system is stable.
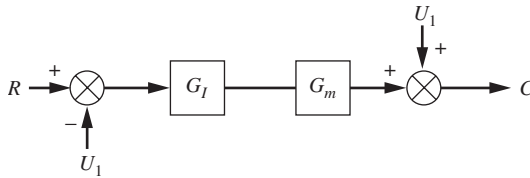
FIGURE 17–34
IMC structure when model matches the process ($G_m = G$).

Ideally, we should like to have C track R without lag when only a set point change occurs (i.e., $U_1 = 0$). For this to occur, we see from Fig. 17–34 or Eq. (17.15) that $G_I G = 1$, or since $G = G_m$, we may write $G_I G_m = 1$. Solving for G_I gives

$$G_I = \frac{1}{G_m} \tag{17.16}$$

Equation (17.16) simply states that the IMC controller should be the inverse of the transfer function of the process model. *Keep in mind that Eq. (17.16) is based on the assumption that the model exactly matches the process.*

For the case of only a change in load U_1 (i.e., $R = 0$), we should like to have the output C remain unchanged (i.e., $C = 0$). For this to occur, we see again from Fig. 17–34 or Eq. (17.15) that $G_I G_m = 1$; this leads to the same result as given by Eq. (17.16).

Even if there is no mismatch between the model and the process, the application of Eq. (17.16) will usually lead to a transfer function that cannot be implemented because it will be unstable, requires prediction, or requires pure differentiation. For example, if $G_m = 1/(\tau s + 1)$, the application of Eq. (17.16) gives

$$G_I = \tau s + 1$$

This result is equivalent to an ideal PD controller, which cannot be implemented because of the derivative term. If $G_m = e^{-\tau s}/(\tau_1 s + 1)$, we obtain

$$G_I = (\tau_1 s + 1)e^{\tau s}$$

The term $e^{\tau s}$, which represents prediction, cannot be implemented. If $G_m = (1 - s)/[(1 + s)(\tau s + 1)]$, then

$$G_I = \frac{(1 + s)(\tau s + 1)}{1 - s}$$

The term $1 - s$ in the denominator means that a pole is in the right half-plane, which leads to an unstable controller. With such difficulties of implementation of the internal model controller, one might ask if any practical result can be obtained. These difficulties can be overcome by application of the following simplified procedure.

Design of IMC Controllers

In using these rules, only a step change in disturbance is considered. The procedure for disturbances other than a step response is more complicated and beyond the scope of the limited discussion presented here.

1.   Separate the process model G_m into two terms

$$G_m = G_{m_a} G_{m_m} \tag{17.17}$$

where G_{m_a} is a transfer function of an all-pass filter. An all-pass filter is one for which $\left|G_{m_a}(j\omega)\right| = 1$ for all ω. Examples are $e^{-\tau_d s}$ and $(1 - s)/(1 + s)$. The G_{m_m} is a transfer function that has minimum phase characteristics. A system has nonminimum phase characteristics if its transfer function contains zeros in the right half-plane or transport lags, or both. Otherwise, a system has minimum phase characteristics. For a step change in disturbance ($R = 1/s$ or $U_1 = 1/s$), G_I is determined by

$$G_I = \frac{1}{G_{m_m}} \tag{17.18}$$

For a disturbance other than a step change, obtaining G_I is more complicated and the reader is referred to Morari and Zafiriou (1989).

The results of applying Eq. (17.18) will yield a transfer function that is stable and does not require prediction; however, it will have terms that cannot be implemented because they require pure differentiation (e.g., $\tau s + 1$).

2. To obtain a practical IMC controller, one multiplies G_I in step 1 by a transfer function of a filter $f(s)$. The simplest form recommended by Morari and Zafiriou is given by

$$f(s) = \frac{1}{(\lambda s + 1)^n} \tag{17.19}$$

where λ is a filter parameter and n is an integer. The practical IMC controller G_I can now be expressed as

$$G_I = \frac{f}{G_{m_m}} \tag{17.20}$$

The value of n is selected large enough to give a result for G_I that does not require pure differentiation. For the simple treatment of IMC design presented here, λ will be considered as a tunable parameter. In the full treatment of IMC given by Morari and Zafiriou, λ can be related to the model uncertainty. In practice, model uncertainty may not be available, in which case one is forced to treat λ as a tunable parameter.

3. If one wants to obtain the conventional controller transfer function G_c, use is made of Eq. (17.14), with G_I obtained from Eq. (17.20). For many simple process models, G_c turns out to be equivalent to a PID controller multiplied by a first-order transfer function; thus

$$G_c = K_c \left(1 + \tau_D s + \frac{1}{\tau_I s} \right) \left(\frac{1}{\tau_1 s + 1} \right) \tag{17.21}$$

where K_c, τ_D, τ_I, and τ_1 are functions of λ and the parameters in G_I and G_m. The examples that follow will illustrate the application of this simplified procedure for designing an IMC controller.

Example 17.5. *Internal model control.* Design an IMC controller for the process which is first-order:

$$G_m = \frac{K}{\tau s + 1}$$

For this case $G_{m_a} = 1$ and $G_{m_m} = K / (\tau s + 1)$. Applying Eq. (17.18) gives

$$G_I = \frac{1}{G_{m_m}} = \frac{\tau s + 1}{K}$$

To be able to implement this transfer function, let $f(s) = 1/(\lambda s + 1)$. The IMC controller becomes

$$G_I = \frac{1}{K} \frac{\tau s + 1}{\lambda s + 1}$$

This result is a lead-lag transfer function that can be easily implemented with computer-based controllers. We may now obtain G_c from Eq. (17.14).

$$G_c = \frac{G_I}{1 - G_I G_m}$$

Introducing the expressions for G_I and G_m into this equation gives

$$G_c = \frac{\dfrac{\tau s + 1}{K(\lambda s + 1)}}{1 - \dfrac{\tau s + 1}{K(\lambda s + 1)}\dfrac{K}{\tau s + 1}} = \frac{\tau s + 1}{K\lambda s} = \frac{\tau}{\lambda K}\left(1 + \frac{1}{\tau s}\right)$$

This result is in the form of a PI controller:

$$G_c = K_c\left(1 + \frac{1}{\tau_I s}\right) \qquad K_c = \frac{\tau}{\lambda K} \qquad \tau_I = \tau$$

Although this design procedure results in the equivalence of a PI controller, only one parameter (λ) must be used to tune the controller. This is a distinct advantage over the use of a conventional controller in which both K_c and τ_I must be tuned.

Example 17.6. *Internal model control.* Design an IMC controller for a process which is first-order with transport lag

$$G = K\frac{e^{-\tau_d s}}{\tau s + 1}$$

In the model of this process, use as an approximation to the transport lag a first-order Padé approximation [see Eq. (7.48)]; thus

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

The model becomes

$$G_m = K\frac{1 - (\tau_d / 2)s}{1 + (\tau_d / 2)s}\frac{1}{\tau s + 1}$$

For this model,

$$G_{m_a} = \frac{1 - (\tau_d / 2)s}{1 + (\tau_d / 2)s} \quad \text{(an all-pass filter)}$$

and

$$G_{m_m} = \frac{K}{\tau s + 1}$$

Following the same steps as used in Example 17.5, we obtain for the IMC controller

$$G_I = \frac{1}{K}\frac{\tau s + 1}{\lambda s + 1}$$

It is instructive to see the form G_c takes for this example. Applying Eq. (17.14) gives

$$G_c = \frac{G_I}{1 - G_I G_m} = \frac{\dfrac{\tau s + 1}{K(\lambda s + 1)}}{1 - \dfrac{\tau s + 1}{K(\lambda s + 1)} \dfrac{K[1 - (\tau_d / 2)s]}{[1 + (\tau_d / 2)s](\tau s + 1)}}$$

This may be reduced algebraically to the form given by Eq. (17.21) with

$$K_c = \frac{2\tau + \tau_d}{2(\lambda + \tau_d)}$$

$$\tau_I = \tau + \frac{\tau_d}{2}$$

$$\tau_D = \frac{\tau \tau_d}{2\tau + \tau_d}$$

$$\tau_1 = \frac{\lambda \tau_d}{2(\lambda + \tau_d)}$$

The response of this first-order with transport lag system for several values of λ and for $K = 1$, $\tau = 1$, and $\tau_d = 1$ is given in Fig. 17–35. The values of K_c, τ_I, τ_d, and τ_1 obtained from the above relations are shown in Table 17.2. Notice that once a model is accepted, the tuning of the modified IMC controller [Eq. (17.21)] depends only on the choice of λ. For the range of λ used, Fig. 17–35 shows that the step response is only slightly oscillatory for all values of λ, and the fastest response is for $\lambda = 0.5$. Also notice that λ affects only K_c and τ_1. This example shows that the design of a controller by the IMC method is a straightforward procedure and leads to a controller that requires the adjustment of only one parameter, λ.
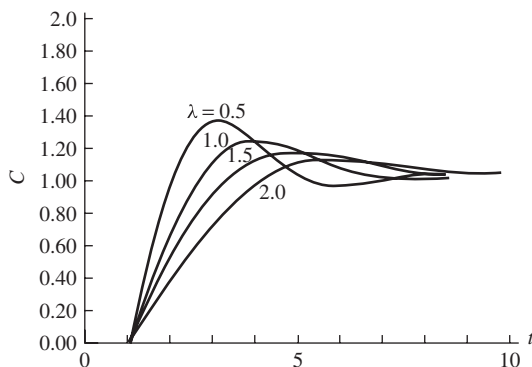

FIGURE 17–35
Response for IMC designed controller of Example 17.6.

TABLE 17.2
IMC derived controller settings for Example 17.6

λ	0.5	1.0	1.5	2.0
K_c	1.0	0.75	0.60	0.5
τ_I	1.5	1.5	1.5	1.5
τ_D	0.33	0.33	0.33	0.33
τ_1	0.167	0.25	0.30	0.33

It is instructive to compare the response for the IMC derived controller with the response for a PI controller using Ziegler-Nichols (Z-N) settings. The responses, which are given in Fig. 17–36, show that for this particular example the controller using Z-N settings produces a response with less overshoot and a higher frequency of oscillation than the controller designed by the IMC method.

These two examples show clearly how the parameters of the conventional controller G_c are related to the parameters of the model and the filter.

The treatment of internal model control presented here has been limited to single-input, single-output continuous systems for which the disturbance is a step change. Furthermore, we have not discussed the use of model uncertainty in selecting the filter parameters. Internal model control has been extended to sampled-data control systems and to multiple-input, multiple-output systems. IMC is an approach to the design of control systems that considers the process model as an essential part of the control system design. Computer-based controllers have the capability of implementing many of the control algorithms designed by the IMC method. There is no longer a need to be tied to the classical control algorithms.
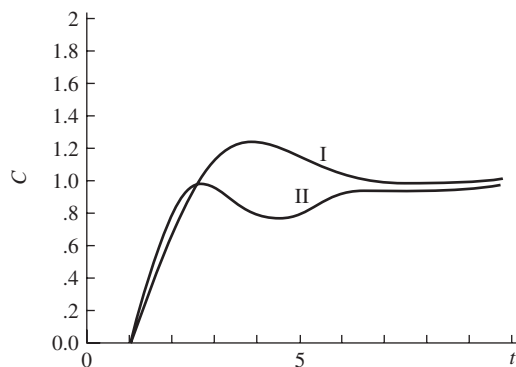

FIGURE 17–36
Comparison of response for IMC controller and conventional controller for Example 17.6:
Curve I: IMC derived controller with $\lambda = 1.0$; curve II: PI controller with Ziegler-Nichols settings
($K_c = 1.02$ and $\tau_I = 2.84$).

SUMMARY

In this chapter, we have examined five advanced control strategies. The first three on cascade control, feedforward-feedback control, and ratio control are advanced only in the sense that each strategy is more complex than the single-loop systems we have encountered up to this chapter. These three strategies are used extensively in industry, and computer-based controllers can implement them easily. The other strategies, on Smith prediction and internal model control (IMC), are less likely to be used in industry and are closely related in their block diagram structure. Three of the strategies—feedforward-feedback, Smith prediction, and IMC—are dependent on accurate models of the processes for their application.

Cascade control is especially useful in reducing the effect of a load disturbance that is located far from the control variable and that moves through the system slowly. The presence of the inner control loop reduces the lag in the outer loop, with the result that the cascade system responds more quickly to a load disturbance.

If a particular load disturbance occurs frequently, the quality of control can often be improved by applying feedforward control. Ideally the transfer function of the feedforward controller is obtained from knowledge of the model of the process. In cases where the feedforward controller transfer function requires prediction (for example, $\tau_f s + 1$), one must be satisfied with an approximation of the feedforward controller, which takes the form of a lead-lag transfer function. When a model of the process does not exist, the feedforward controller can be tuned after doing some open-loop step tests that relate the control variable to the load disturbance. To provide for load disturbances that cannot be measured or anticipated, feedforward control is always combined with feedback control in a practical situation.

Ratio control is widely used in industry in the blending of two component streams (A and B) to produce a mixed stream of desired composition (i.e., ratio of components). Ratio control is essentially a flow control problem in which the flow measurement of stream A (the wild stream) is used to compute the set point for the flow of stream B so that the desired ratio of components will be obtained.

The Smith predictor control scheme (dead-time compensation) was developed to improve the control of a system having a large transport lag. The method is based on a model of the process that is first-order with dead time. By introducing inner loops that contain elements of the transfer function of the model, the control system is transformed ideally to one without transport lag, a system that is much easier to control. This ideal situation occurs when the process and the model are in exact agreement. In reality, the success of the Smith predictor strategy depends on the degree of agreement between process and model.

Internal model control resembles the Smith predictor strategy in terms of the structure of the block diagram. To apply the IMC method, one must have an accurate model of the process, the model uncertainty, the type of disturbance (step, ramp, etc.), and the performance objective (integral of square error). The method, which is based on a rigorous mathematical foundation, leads to an IMC controller that is the best that can be designed in terms of the performance objective. The IMC structure can be reduced to a conventional control structure in which the conventional controller is related to the IMC controller and the parameters of the model. For many simple processes with simple disturbance (impulse, step, etc.), the equivalent conventional controller based on the IMC design method turns out to be the equivalent of a PID controller.

PROBLEMS

17.1. (a) Obtain G_f for the feedforward-feedback system shown in Fig. P17–1 so that C does not change when a disturbance in C_i occurs. Would there be any problem in implementing this G_f?

(b) If G_f is to be a lead-lag transfer function

$$\frac{T_1 s + 1}{T_2 s + 1}$$

determine T_1 and T_2 by the Foxboro rule. How do you determine whether lead or lag is to predominate? Use $t_p = 1.0$ in the Foxboro rule.

(c) When feedforward-feedback control is present, sketch the response $C(t)$ when $C_i = 1/s$ and when G_f from part (a) is used.

(d) Repeat part (c) when G_f from part (b) is used. Only a rough sketch that suggests the transient response is expected in this case.

(e) Determine $C(t)$ when $C_i = 1/s$, $G_f = -1$, and the feedback loop is broken at A-A. Obtain the numerical value of $C(t)$ at $t = 0.5$, 1.0, and 1.5.
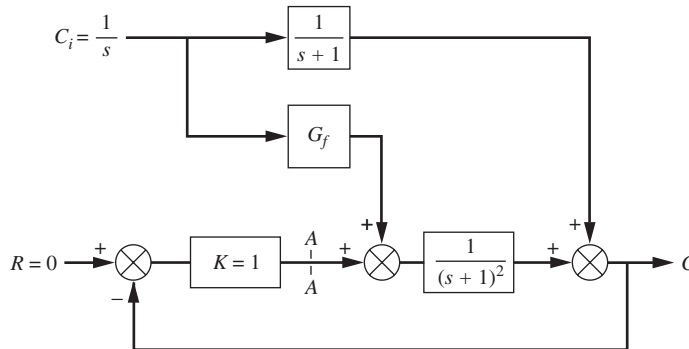
FIGURE P17–1

CHAPTER
17

CAPSULE SUMMARY

CASCADE CONTROL

Cascade control is used for improved disturbance rejection. It employs two controllers, a primary controller and a secondary controller. The primary controller provides the set point for the secondary controller.
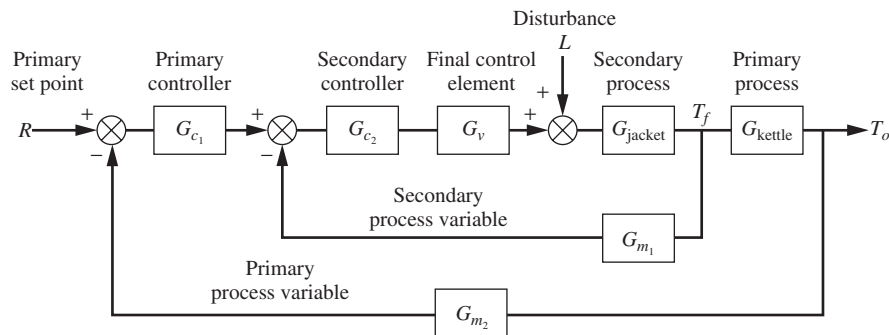
FEEDFORWARD CONTROL

Feedforward control is usually used in conjunction with feedback control to improve disturbance rejection. A model of the process is used to predict the effect of a disturbance on the process. This information instructs the controller on how to preemptively take action to negate the effect of the disturbance.
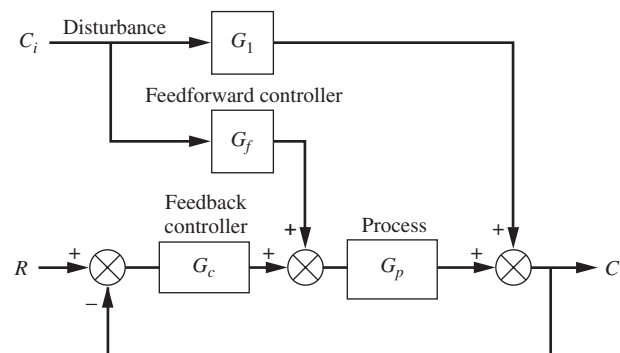
RATIO CONTROL

Ratio control is a control strategy used to control the blending of two components in a desired ratio.
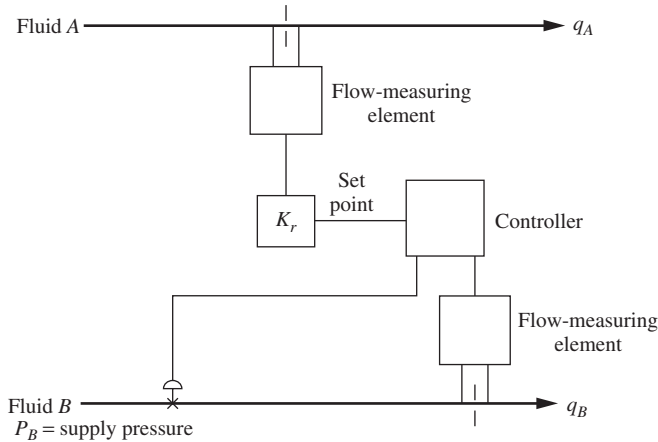
SMITH PREDICTOR CONTROL

Sometimes referred to as dead-time compensation, this control strategy is used to counteract the deleterious effects of dead time on process control. A model of the process is used to predict and negate the effect of dead time.

INTERNAL MODEL CONTROL

Internal model control resembles the Smith predictor strategy in terms of the structure of the block diagram. To apply the IMC method, one must have an accurate model of the process, the model uncertainty, the type of disturbance (step, ramp, etc.), and the performance objective (integral of square error). The method, which is based on a rigorous mathematical foundation, leads to an IMC controller that is the best that can be designed in terms of the performance objective. The IMC structure can be reduced to a conventional control structure in which the conventional controller is related to

the IMC controller and the parameters of the model. For many simple processes with simple disturbance (impulse, step, etc.), the equivalent conventional controller based on the IMC design method turns out to be the equivalent of a PID controller.

CONTROLLER TUNING AND PROCESS IDENTIFICATION

The selection of a controller type (P, PI, PID) and its parameters (K_c, τ_I, τ_D) is intimately related to the model of the process to be controlled. The adjustment of the controller parameters to achieve satisfactory control is called *tuning*. The selection of the controller parameters is essentially an optimization problem in which the designer of the control system attempts to satisfy some criterion of optimality, the result of which is often referred to as "good" control. The process of tuning can vary from a trial-and-error attempt to find suitable control parameters for good control to an elaborate optimization calculation based on a model of the process and a specific criterion of optimal control. In many applications, there is no model of the process, and the criterion for good control is only vaguely defined. A typical criterion for good control is that the response of the system to a step change in set point or load have minimum overshoot and a one-quarter decay ratio. Other criteria may include minimum rise time and minimum settling time.

In the first part of this chapter, some of the widely used tuning rules for continuous controllers will be presented. In the second part of the chapter, methods for determining the model of a process from experimental tests will be described. Determining the model of a process experimentally is referred to as *process identification.*

18.1 CONTROLLER TUNING

Before we present tuning rules, some discussion of the effect of each mode in a PID controller on the transient response of a controlled process will be instructive.

Selection of Controller Modes

Consider a typical loop as shown in Fig. 18–1 in which the process is second-order and the measuring element is a transport lag. (The transfer function of the valve is taken
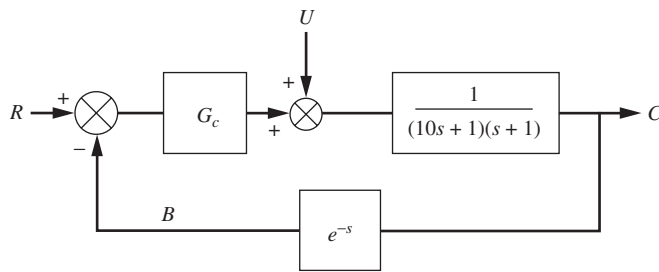
FIGURE 18–1
Typical control system used to study the effect of controller modes on load
responses shown in Fig. 18–2.

as 1.) Load responses for this process for four types of controllers (P, PD, PI, PID) are
shown in Fig. 18–2. For each response curve, the process was subjected to a unit-step
change in load ($U = 1/s$), and the controller parameters were selected by tuning rules to
be presented later. Regardless of the specific tuning rules used, the responses shown in
Fig. 18–2 are typical of well-tuned controllers for systems found in industry. The nature
of the response for each type controller will now be described. (The reader should also
refer to Figs. 9–10 and 16–18 to reinforce this discussion.)

FIGURE 18–2
Load response of a typical control system using
various modes of control (process shown in Fig. 18–1).

PROPORTIONAL CONTROL. As shown in Fig. 18–2, proportional control produces
an overshoot followed by an oscillatory response, which levels out at a value that does
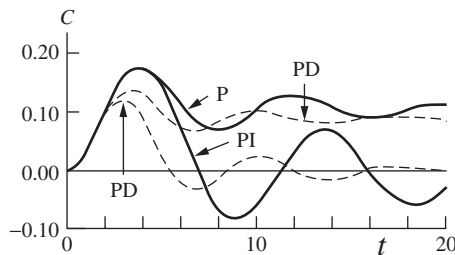not equal the set point; this ultimate displacement from the set point is the offset.

PROPORTIONAL-DERIVATIVE CONTROL. For this case the response exhibits a
smaller overshoot and a smaller period of oscillation compared to the response for pro-
portional control. The offset that still remains is less than that for proportional control.

PROPORTIONAL-INTEGRAL CONTROL. In this case, the response has about the same
overshoot as proportional control, but the period is larger; however, the response returns
to the set point (offset $= 0$) after a relatively long settling time. The most beneficial
influence of the integral action in the controller is the elimination of offset.

PROPORTIONAL-INTEGRAL-DERIVATIVE CONTROL. As one might expect, the use of PID control combines the beneficial features of PD and PI control. The response has lower overshoot and returns to the set point more quickly than the responses for the other types of controllers.

From the nature of the responses just described, we can make the following generalizations. Integral action, which is present in PI and PID controllers, eliminates offset. The addition of derivative action speeds up the response by contributing to the controller output a component of the signal that is proportional to the rate of change of the process variable.

For simple, low-order (first- or second-order) processes that can tolerate some offset, P or PD control is satisfactory. For processes that cannot tolerate offset and are of low order, PI control is required. For processes that are of high order (those with transport lag or many first-order lags in series), PID control is needed to prevent large overshoot and long settling time. There is probably little justification to select a P or PD controller for most processes. The PI controller is often the choice because it eliminates offset and requires only two parameter adjustments. Tuning a PID controller is more difficult because three parameters must be adjusted. The presence of derivative action can also cause the controller output to be very jittery if there is much noise in the signals. We now turn our attention to some of the criteria for good control that are used to judge whether a control system is well tuned.

Criteria for Good Control

Before we can be satisfied with the response of a control system for a choice of control parameters, we must have some concept of what we want as an ideal response. Most operators of processes know what they want in the form of a response to a change in set point or load. For example, a response that gives minimum overshoot and $\frac{1}{4}$ decay ratio is often considered as a satisfactory response. In many cases, tuning is done by trial and error until such a response is obtained. To compare different responses that use different sets of controller parameters, a criterion that reduces the entire response to a single number, or a *figure of merit,* is desirable. The figure of merit provides a means of "keeping score" for the different control parameters, and as we shall see, the low "score" generally wins. It is dangerous, however, to rely solely on the score to determine the best choice for the control parameters. The control system designer should examine the nature of the response in conjunction with the requirements for the process to determine the "best" choice of settings.

One criterion that is often used to evaluate a response of a control system is the integral of the square of the error (ISE) with respect to time. The definition of ISE is as follows:

Integral of the square of the error (ISE)

$$\text{ISE} \ = \ \int_0^\infty e^2 \, dt \tag{18.1}$$

where e is the usual error (i.e., set point – control variable). For a stable system for which there is no offset [i.e., $e(\infty) = 0$], Eq. (18.1) produces a single number as a figure of merit.

The objective of the designer is to obtain the minimum value of ISE by proper choice of control parameters. A response that has large errors and persists for a long time will produce a large ISE. For the cases of P and PD control, where offset occurs, the integral given by Eq. (18.1) does not converge. In these cases, one can use a modified integrand, which replaces the error $r(t) - c(t)$ by $c(\infty) - c(t)$. Since $c(\infty) - c(t)$ does approach zero as t goes to infinity, the integral will converge and serve as a figure of merit.

Two other criteria often used in process control are defined as follows:

Integral of the absolute value of error (IAE)

$$\text{IAE} = \int_0^\infty |e|\, dt \tag{18.2}$$

Integral of time-weighted absolute error (ITAE)

$$\text{ITAE} = \int_0^\infty |e|\; t\, dt \tag{18.3}$$

Each of the three figures of merit, given by Eqs. (18.1), (18.2), and (18.3), has different purposes. The ISE will penalize (i.e., increase the value of ISE) the response that has large errors, which usually occur at the beginning of a response, because the error is squared. The ITAE will penalize a response that has errors that persist for a long time. The IAE will be less severe in penalizing a response for large errors and treat all errors (large and small) in a uniform manner. The ISE figure of merit is often used in optimal control theory because it can be used more easily in mathematical operations (e.g., differentiation) than the figures of merit, which use the absolute value of error. In applying the tuning rules to be discussed in the next section, these figures of merit can be used in comparing responses that are obtained with different tuning rules.

> **Example 18.1.**   For the control system shown in Fig. 18–1, determine the ISE, ITAE, and IAE for a unit-step load disturbance. Let the controller be a PI controller with $K_c = 6$ and $\tau_I = 4$.
>
> Simulink can be used to easily simulate this process and calculate the figures of merit. The Simulink block diagram is shown in Fig. 18–3.
>
> The response generated using the model of Fig. 18–3 of the process for these controller settings is shown in Fig. 18–4.
>
> The figures of merit alone tell us nothing unless they are compared with figures for other controller settings. Remember, the goal is generally to obtain the lowest value. Once we have a Simulink model such as this prepared, it is relatively easy to check a variety of controller settings. We compare some other controller settings and their figures of merit shortly.

18.2   TUNING RULES

Ziegler-Nichols (Z-N) Rules

These rules were first proposed by Ziegler and Nichols (1942), who were engineers for a major control hardware company in the United States (Taylor Instrument Co.). Based
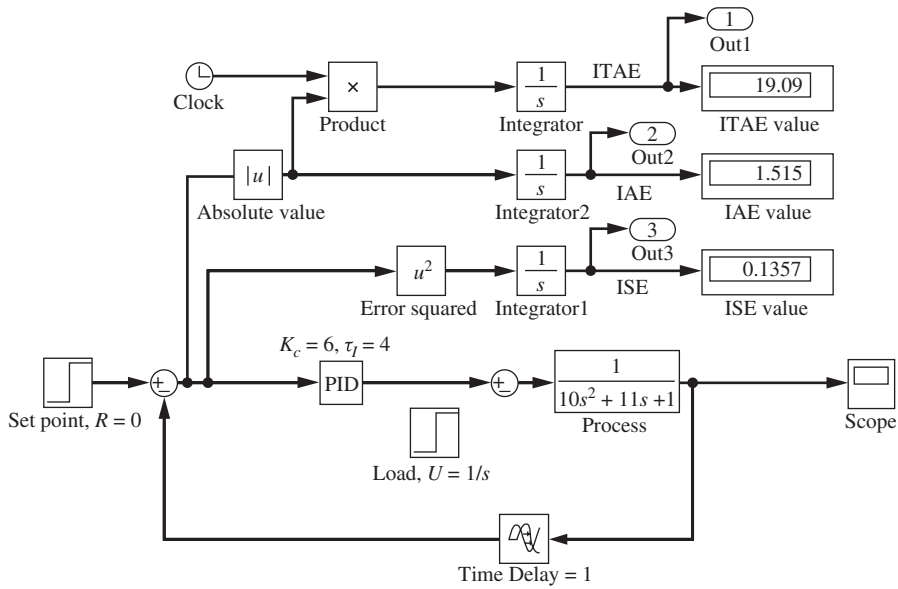
FIGURE 18–3
Simulink model for Example 18.1
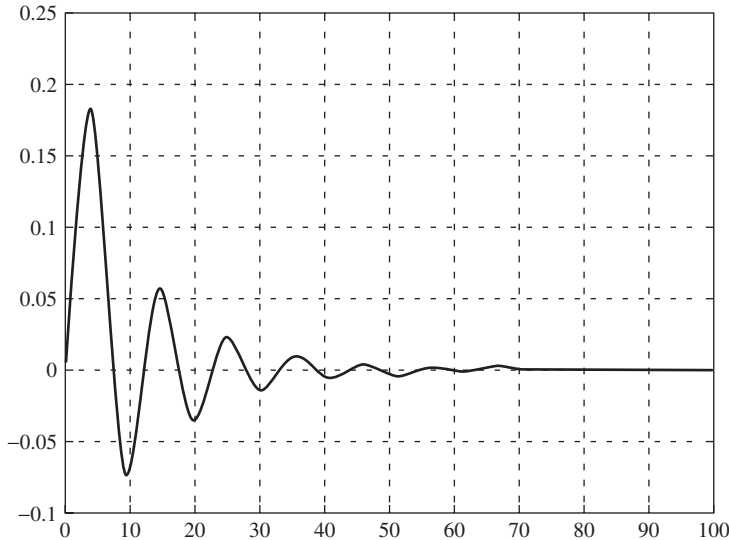

FIGURE 18–4
Process response for Example 18.1

on their experience with the transients from many types of processes, they developed a *closed-loop tuning method* still used today in one form or another. The method is described as a closed-loop method because the controller remains in the loop as an active controller in automatic mode. This closed-loop method will be contrasted with an

open-loop tuning method to be discussed later. We have already discussed the Ziegler-Nichols rules in Chap. 16 as a natural consequence of our study of frequency response. Ziegler and Nichols did not suggest that the ultimate gain K_{cu} and ultimate period P_u be computed from frequency response calculations based on the model of the process. They intended that K_{cu} and P_u be obtained from a closed-loop test of the actual process. When the rules were first proposed, frequency response methods and process models were not generally available to the control engineers. The rules are presented below and are in the form that one would use for actual application to a real process.

1. After the process reaches steady state at the normal level of operation, remove the integral and derivative modes of the controller, leaving only proportional control. On some PID controllers, this requires that the integral time τ_I be set to its maximum value and the derivative time τ_D to its minimum value. On computer-based controllers, the integral and derivative modes can be removed completely from the controller.

2. Select a value of proportional gain K_c, disturb the system, and observe the transient response. If the response decays, select a higher value of K_c and again observe the response of the system. Continue increasing the gain in small steps until the response first exhibits a sustained oscillation. The value of gain and the period of oscillation that correspond to the sustained oscillation are the ultimate gain K_{cu} and the ultimate period P_u.

Some very important precautions to take in applying this step of the tuning method are given in the next section.

3. From the values of K_{cu} and P_u found in step 2, use the Ziegler-Nichols rules given in Table 18.1 to determine controller settings (K_c, τ_I, τ_D). (This table is the same as Table 16.1.)

Although variations in the tuning rules given in Table 18.1 are used by industry, the same approach of using K_{cu} and P_u to obtain controller parameters is used. The Ziegler-Nichols rules generally provide conservative (and safe) controller settings. The Z-N settings should be considered as only approximate settings for satisfactory control. Fine-tuning of the controller settings is usually required to get an improved control response.

TABLE 18.1
Ziegler-Nichols controller settings

Type of control	$G_c(s)$	K_c	τ_I	τ_D
Proportional (P)	K_c	$0.5K_u$		
Proportional-integral (PI)	$K_c\left(1 + \dfrac{1}{\tau_I s}\right)$	$0.45K_u$	$\dfrac{P_u}{1.2}$	
Proportional-integral-derivative (PID)	$K_c\left(1 + \dfrac{1}{\tau_I s} + \tau_D s\right)$	$0.6K_u$	$\dfrac{P_u}{2}$	$\dfrac{P_u}{8}$

The experimental determination of K_{cu} and P_u described in step 2 can be replaced by a computation using frequency response methods if an accurate model of the process, valve, and measuring element is known. This type of calculation was done in Chap. 16.

PRECAUTIONS TO TAKE IN APPLYING THE Z-N METHOD. Some discussion is needed to avoid some pitfalls in applying step 2 of the Z-N method to obtain K_{cu} and P_u. These precautions are concerned with the type and size of the disturbance that induces the response and with the avoidance of using a limit cycle as the indication that the system is on the threshold of instability.

The simplest way to introduce a disturbance is to move the set point away from the control variable for a short time and then return the set point to its original value. This procedure, which is equivalent to introducing a pulse function in the error, causes the system to respond and yet stay within a narrow band surrounding the normal operating point of the process.

An alternate type of disturbance would be to introduce a small step change in set point. If step changes in set point are used to induce transients, the successive step changes should alternate around the normal operating point of the process. It is also important to make the disturbance as small as possible, especially as the gain of the controller is increased, so that the valve and other components do not exceed their physical limits.

When the valve moves to its limits during a closed-loop transient, we say that the valve saturates. Under these conditions, a sustained oscillation occurs, which is called a limit cycle. The limit cycle that is caused by saturation is a nonlinear phenomenon, which will be covered in Chap. 25 on nonlinear control. If a limit cycle occurs, the gain that produces it and the period of the cycle should not be used in the Ziegler-Nichols rules. Since the limit cycle will appear to the observer to be the same as a sustained oscillation when the system is on the verge of instability, the novice will often mistakenly use the information derived from the limit cycle (controller gain and period) to obtain controller settings. A simple way to know if one has a limit cycle is to observe the swing in pressure to the valve. If the limits of the valve (e.g., 3 to 15 psig) are reached repeatedly during the oscillatory response, one has a limit cycle and the controller gain and period should not be used to determine controller settings. It is for this reason step 2 states that K_c should be increased in small steps until the response first exhibits a sustained oscillation.

To appreciate the use of step 2 of the tuning method, one should have some laboratory experience in tuning a real process, or at least a computer simulation of a process. The experienced operator can develop some shortcuts to finding the ultimate gain and ultimate period.

Cohen and Coon (C-C) Rules

The next method of tuning to be discussed is an *open-loop method,* in which the control action is removed from the controller by placing it in manual mode and an open-loop transient is induced by a step change in the signal to the valve. This method was proposed by Cohen and Coon (1953) and is often used as an alternative to the Z-N method. Fig. 18–5 shows a typical control loop in which the control action is removed and the loop opened for the purpose of introducing a step change (*M/s*) to the valve. The step response is recorded at the output of the measuring element. The step change to the

valve is conveniently provided by the output from the controller, which is in manual mode. The response of the system (including the valve, process, and measuring element) is called the *process reaction curve;* a typical process reaction curve exhibits an S shape, as shown in Fig. 18–6. After we present the Cohen and Coon method of tuning, the basis for their recommendations will be discussed. The C-C method is summarized in the following steps:
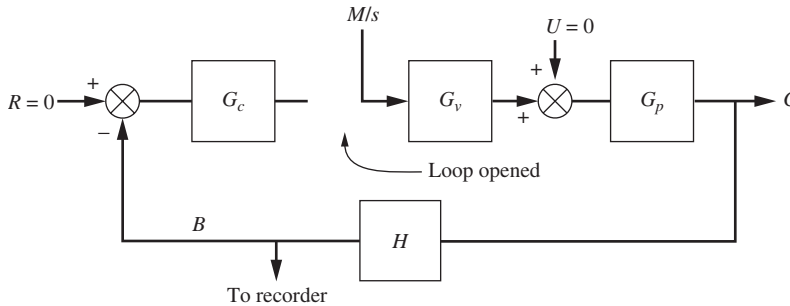

FIGURE 18–5
Block diagram of a control loop for measurement of a process reaction curve.

1. After the process reaches steady state at the normal level of operation, switch the controller to manual. In a modern controller, the controller output will remain at the same value after switching as it had before switching. (This is called "bumpless" transfer.)
2. With the controller in manual, introduce a small step change in the controller output that goes to the valve and record the transient, which is the process reaction curve (Fig. 18–6).
3. Draw a straight line tangent to the curve at the point of inflection, as shown in Fig. 18–6. The intersection of the tangent line with the time axis is the apparent transport lag T_d; the apparent first-order time constant T is obtained from

$$T = \frac{B_u}{S} \tag{18.4}$$

   where B_u is the ultimate value of B at large t and S is the slope of the tangent line. The steady-state gain that relates B to M in Fig. 18–5 is given by

$$K_P = \frac{B_u}{M} \tag{18.5}$$

4. Using the values of K_p, T, and T_d from step 3, the controller settings are found from the relations given in Table 18.2.

   Notice in Table 18.2 that all the controller settings are a function of the dimensionless group T_d/T, the ratio of the apparent transport lag to the apparent time constant. Also K_c is inversely proportional to K_p.
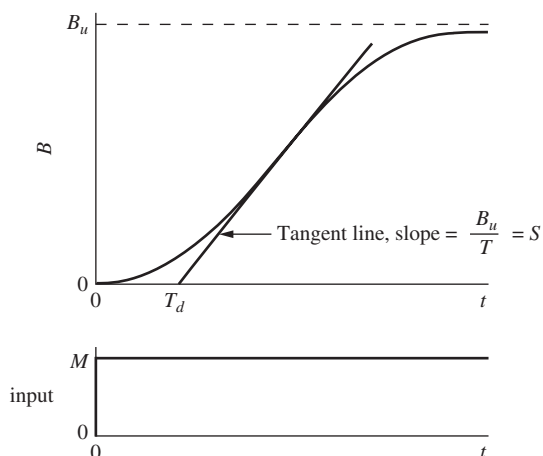
FIGURE 18–6
Typical process reaction curve showing graphical construction
to determine first-order with transport lag model.

TABLE 18.2
Cohen-Coon controller settings

Type of control	Parameter setting
Proportional (P)	$K_c = \dfrac{1}{K_p}\dfrac{T}{T_d}\left(1 + \dfrac{T_d}{3T}\right)$
Proportional-integral (PI)	$K_c = \dfrac{1}{K_p}\dfrac{T}{T_d}\left(\dfrac{9}{10} + \dfrac{T_d}{12T}\right)$
	$\tau_I = T_d\,\dfrac{30 + 3T_d/T}{9 + 20T_d/T}$
Proportional-derivative (PD)	$K_c = \dfrac{1}{K_p}\dfrac{T}{T_d}\left(\dfrac{5}{4} + \dfrac{T_d}{6T}\right)$
	$\tau_D = T_d\,\dfrac{6 - 2T_d/T}{22 + 3T_d/T}$
Proportional-integral-derivative (PID)	$K_c = \dfrac{1}{K_p}\dfrac{T}{T_d}\left(\dfrac{4}{3} + \dfrac{T_d}{4T}\right)$
	$\tau_I = T_d\,\dfrac{32 + 6T_d/T}{13 + 8T_d/T}$
	$\tau_D = T_d\,\dfrac{4}{11 + 2T_d/T}$

The rationale for the C-C tuning method begins with the representation of the S-shaped process reaction curve by a first-order with transport lag model; thus

$$G_p(s) = \frac{K_p e^{-T_d s}}{Ts + 1} \tag{18.6}$$

Using the system expressed by Eq. (18.6), Cohen and Coon obtained by theoretical means the controller settings given in Table 18.2. Their computations required that the response have $\frac{1}{4}$ decay ratio, minimum offset, minimum area under the load-response curve, and other favorable properties.

In applying the C-C tuning method, an important task is the graphical construction, shown in Fig. 18–6, which reduces the process reaction curve to first-order with the transport lag model given by Eq. (18.6). To understand the basis for the graphical procedure, consider the response of the transfer function of Eq. (18.6) to a step change in input; the resulting transient is shown in Fig. 18–7. After $t = T_d$, the response is a first-order response. The point of inflection of the curve in Fig. 18–7 occurs at $t = T_d$, and the slope of the tangent line at this point is related to the time constant by the relation

$$S = \frac{B_u}{T}$$

Solving for T gives the expression in Eq. (18.4). The response after $t = T_d$, shown in Fig. 18–7, was also presented in Fig. 4–7.

The attempt to model the process reaction curve by the method shown in Fig. 18–6 is crude and does not give a very good fit. Finding the point of inflection and drawing a tangent line at this point are quite difficult, especially if the data for the process reaction curve are not accurate and if they scatter. A better method for fitting the process reac-
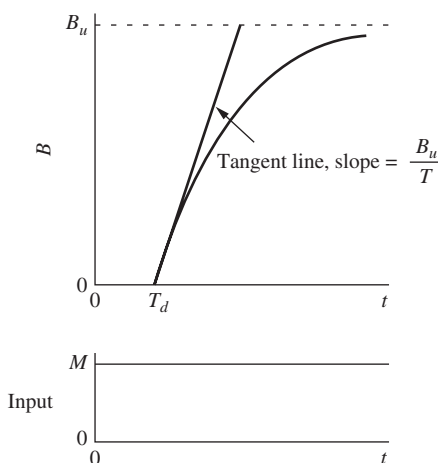

FIGURE 18–7
Step response for a first-order with transport lag model.

TABLE 18.3
Controller settings for the system of Fig. 18–1.

Control type	Parameter	Closed-loop method (Z-N method)	Open-loop method (C-C method)
P	K_c	6.4	8.1
PI	K_c	5.8	7.0
	τ_I	5.6	4.4
PD	K_c	11.4*	9.8
	τ_D	1.0*	0.43
PID	K_c	7.7	10.5
	τ_I	3.4	3.9
	τ_D	1.6	0.59

* Obtained by design for 30° phase margin and maximum K_c.

tion curve to a first-order with transport lag model is to perform a least-squares fit of the data. Some computer software, such as the LOOP-PRO package (see www.controlstation .com), provide an easy means of performing the fitting process. An example to be presented later will study the effect of the type of model-fitting procedure on the selection of controller parameters.

To illustrate the two methods of controller tuning just presented, the system shown in Fig. 18–1 was simulated using Simulink as shown above in Example 18.1. Table 18.3 gives the values of the controller parameters obtained by applying both tuning methods (Z-N and C-C). Fig. 18–8 shows the resulting transients. Since the Z-N method does
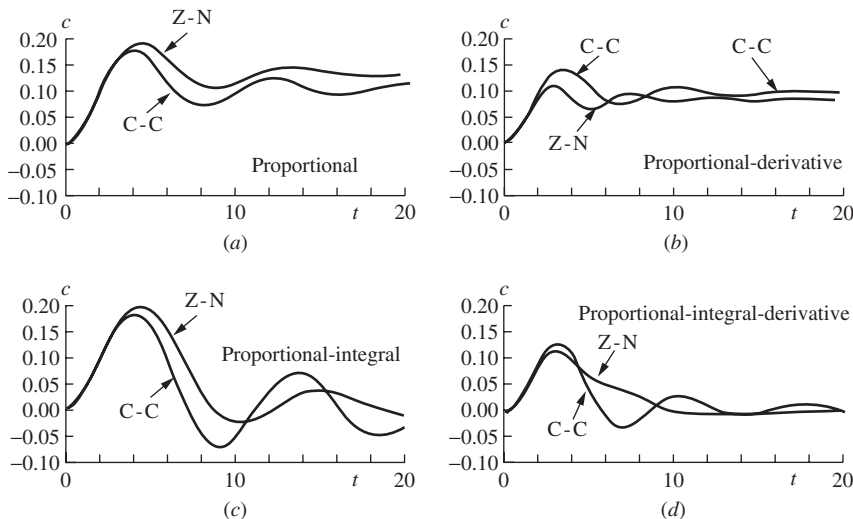

FIGURE 18–8
Comparison of load responses for the system of Fig. 18–1 using controller settings obtained by the Ziegler-Nichols (Z-N) method and the Cohen-Coon (C-C) method.

not give a rule for a PD controller, the settings listed for a PD controller under the Z-N heading of Table 18.3 were obtained by using a theoretical frequency response calculation in which the design was based on 30° phase margin and a maximum K_c. No general conclusions can be drawn about the relative merits of the two tuning methods from the results shown in Fig. 18–8, since these results apply to one specific example. About all that can be said is that for this specific example, both methods give reasonable first guesses of the control parameters.

Example 18.2. For the control system shown in Fig. 18–1, shown again in Fig. 18–9, determine the "best" PI controller settings as indicated by each of the figures of merit (ISE, ITAE, and IAE). Compare the responses of those settings with the ones obtained from Z-N and C-C tuning. We can use the same Simulink diagram that we used in Example 18.1. This time, however, it will be advantageous to write a MATLAB m-file to vary the controller parameters to minimize a particular figure of merit.
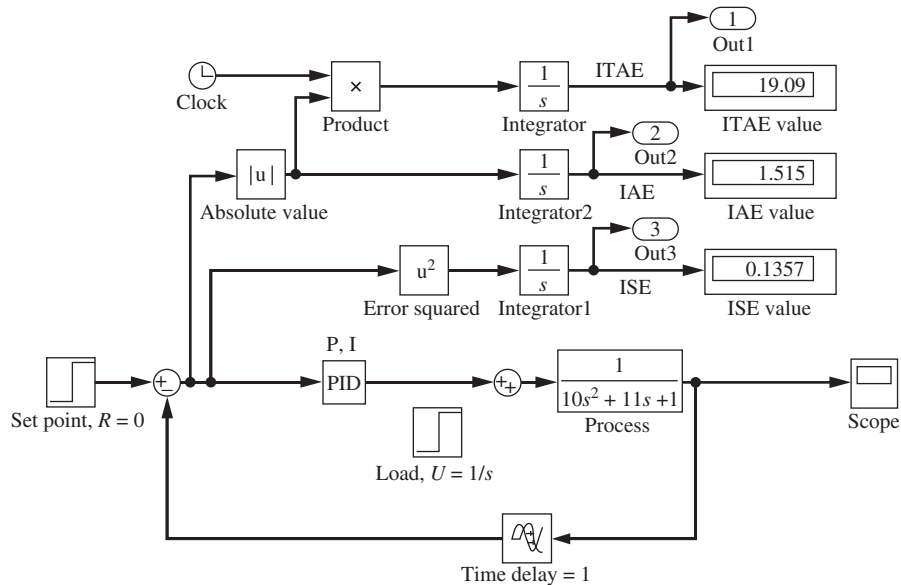
FIGURE 18–9
Simulink model for Example 18.2.

The Simulink block diagram is saved as *fig18_1.mdl*. The m-file that we use to call the Simulink diagram is called *merit_score.m*. We make use of the MATLAB optimization function FMINSEARCH to vary the controller parameters for us and to find the minimum figure of merit of interest. The comment statements in the m-file describe how it's used. An example of the MATLAB command that performs the optimization is

[zn,fval,exitflag]=FMINSEARCH('merit_score',[5.8 1],OPTIMSET ('MaxIter',100))

```
                          m-file: merit_score.m
 function merit_score=merit_score(Varg);
 % this m-file is for use with the FMINSEARCH function to vary P & I to
 % minimize the ITAE, IAE or ISE for the control system in Fig. 18.1:
 %
 % issue the FMINSEARCH command from the MATLAB workspace
 % be sure that the PID controller in simulink model, fig18_1.mdl,
 % has the P and I values set to the variables P & I
 % also be sure to declare P and I global in the MATLAB workspace
 % the 5.8 is the initial guess for P and 1 for I (these are the ZN values)
 %
 %[zn,fval,exitflag]= FMINSEARCH('merit_score',[5.8 1],options)
 %fval is the final value of the objective function
 %set the max iterations with:
 %options = OPTIMSET('MaxIter',100)
 %    If EXITFLAG is:
 %    1 then FMINSEARCH converged with a solution X.
 %    0 then the maximum number of iterations was reached.
 %
 %note that the ITAE is outport 1 on the simulink diagram, fig18_1
 %IAE is outport 2 and ISE is outport 3 on fig18_1
 %the column specified in the y variable is the outport. . .
 % y(:,1)=ITAE
 % y(:,2)=IAE
 % y(:,3)=ISE
 %
 global P I
 P=Varg(1);
 I=Varg(2);
 [t,x,y]=sim('fig18_1',50);
 merit_score=max(y(:,1));
```

The results of running this optimization to minimize each figure of merit for PI control are shown in Table 18.4 and Fig. 18–10.

TABLE 18.4
Comparison of controller settings for Example 18.2 using different tuning methods

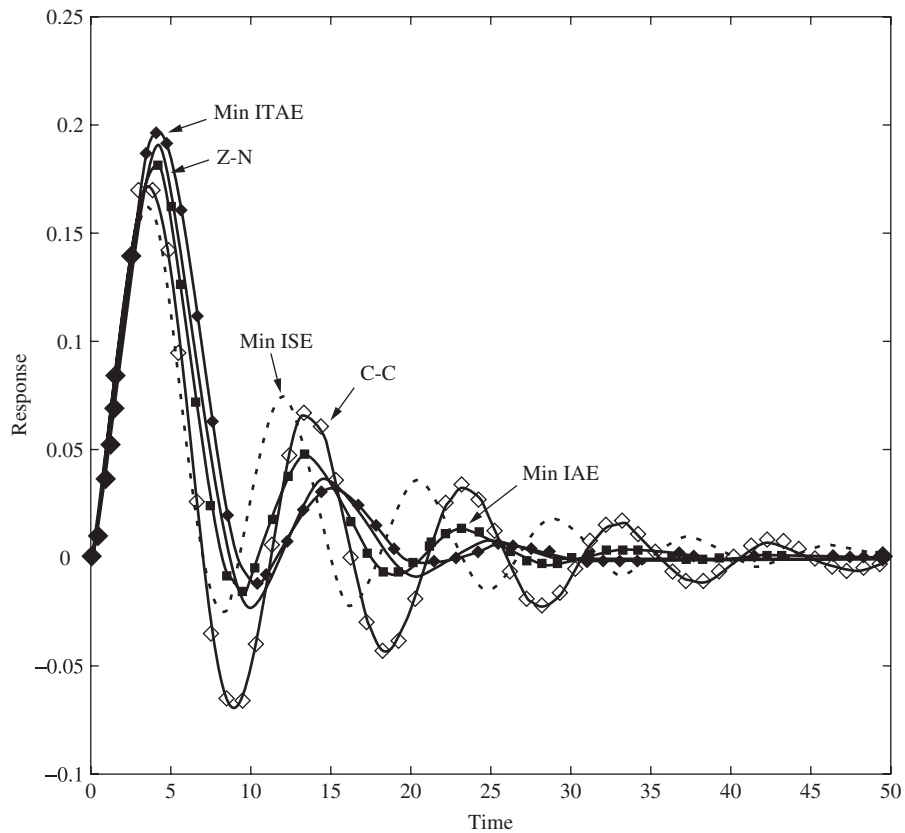| Parameter	Z-N method	C-C method	Minimum ITAE	Minimum IAE	Minimum ISE
K_c	5.8	7.0	5.5	6.6	8.7
τ_I	5.6	4.4	6.0	6.7	7.9

FIGURE 18–10
Plot of system response for Example 18.2 using different tuning methods.

The controller settings for all the settings are very similar. The C-C settings are the most oscillatory for this system. Z-N and minimum ITAE have the largest initial peak, but settle the most rapidly.

Let's look at some additional examples.

Example 18.3.   For the control system shown in Fig. 18–11, determine controller settings for a PI controller using the Z-N method and the C-C method. This problem will be instructive because the transfer function of the model is already in the form of first-order with transport lag, which is the form used by Cohen and Coon to derive their tuning rules.

C-C method.   Since the transfer function of the plant is in the form of Eq. (18.6), we obtain T and T_d immediately without having to draw a tangent line through
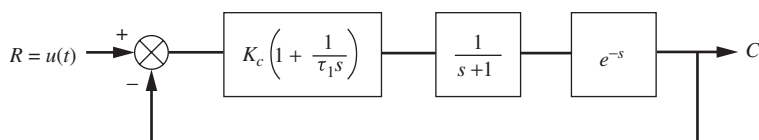
FIGURE 18–11
Process for Example 18.3.

the point of inflection; i.e., $T = 1$ and $T_d = 1$. We also observe from the block diagram that $K_p = 1$. Substituting these values into the appropriate equations of Table 18.2 gives

$$K_c = \frac{T}{K_p T_d}\left(0.9 + \frac{T_d}{12T}\right) = \frac{1}{1}\left(0.9 + \frac{1}{12}\right) = 0.983$$

and

$$\tau_I = T_d \frac{30 + 3\,T_d/T}{9 + 20T_d/T} = \frac{30 + 3}{9 + 20} = 1.14$$

Using these values for K_c and τ_I, the step response shown in Fig. 18–12 was obtained by simulation.
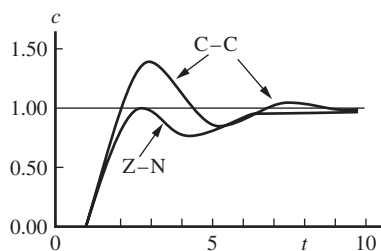

FIGURE 18–12
Response to unit-step change in set point for the system in Fig. 18–11 (Example 18.3).

Z-N method.   Application of the Bode criterion from Chap. 16 gives

$$\omega_{co} = 2.03 \qquad \text{or} \qquad P_u = \frac{2\pi}{\omega_{co}} = 3.09$$

$$K_{cu} = 2.26$$

The details for obtaining these results will not be given here since this type calculation was covered in depth in Chap. 16. Applying the Z-N rules for PI control from Table 18.1 gives

$$K_c = 0.45 K_{cu} = (0.45)(2.26) = 1.02$$

and

$$\tau_I = \frac{P_u}{1.2} = \frac{3.09}{1.2} = 2.58$$

The step response for these controller settings is shown in Fig. 18–12. The ISE value for each response was calculated out to a sufficiently long time (10 units of time) for the integral to converge; the results are as follows:

$$\text{C-C response : ISE} = 1.54 \quad \text{at} \quad t = 10$$
$$\text{Z-N response : ISE} = 1.49 \quad \text{at} \quad t = 10$$

Although the ISE values are nearly the same, the transient for the Z-N settings is better than the transient for the C-C settings. The Z-N transient has much less overshoot. The lesson to be learned from this example is that the comparison of two transients based on only one criterion (in this case, the ISE) may be misleading in the selection of the best transient. It is also important to judge the quality of a transient by its actual appearance. Note that for this example, in which there is a relatively large transport lag ($T_d = 1$), much of the contribution to the ISE occurs from $t = 0$ to $t = 1$, during which time the ISE reaches 1.0. This value of the ISE at $t = 1$ is the same, regardless of the tuning method used because the transport lag causes error to be constant from $t = 0$ to $t = 1$.

Example 18.4.   For the control system shown in Fig. 18–13, determine the controller settings for a PI controller using the Z-N method and the C-C method. In this problem, the process reaction curve must be modeled by the method shown in Fig. 18–6.
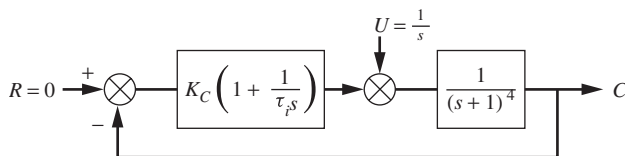

FIGURE 18–13
Process for Example 18.4.

C-C method.   Since the transfer function of the plant is given as $1/(s + 1)^4$, we can obtain the values of T_d and T for use in the C-C method analytically. A unit-step response for the plant transfer function is

$$c(t) = 1 - \left(\tfrac{1}{6}t^3 + \tfrac{1}{2}t^2 + t + 1 \right)e^{-t}$$

From this result one can readily obtain the first and second derivatives; thus

$$\dot{c}(t) = \frac{1}{6}t^3 e^{-t}$$

$$\ddot{c}(t) = \frac{1}{6}e^{-t}\left(3t^2 - t^3\right)$$

The location of the inflection point on the transient $c(t)$ is obtained by setting the second derivative to zero:

$$0 = \frac{1}{6}e^{-t}\left(3t^2 - t^3\right)$$

Solving for t gives as the root of interest in this problem $t = 3$. Knowing that the point of inflection occurs at $t = 3$, we can compute the slope of the tangent line through this point to be

$$S = \dot{c}(3) = \frac{1}{6}(3^3)e^{-3} = 0.224$$

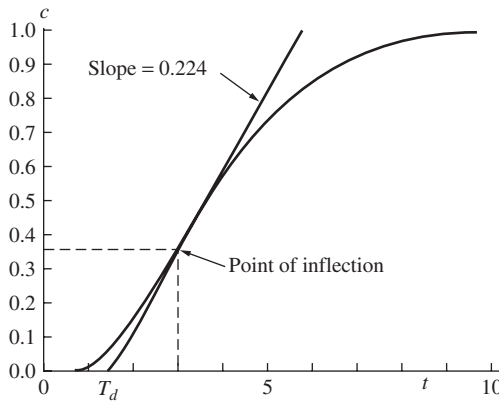We can now determine T_d as shown in Fig. 18–14.

FIGURE 18–14
Process reaction curve for Example 18.4.

From the expression for $c(t)$, we obtain the value of c at the inflection point to be $c(3) = 0.353$. The value of t where the tangent line intersects the t axis is obtained from the slope S; thus

$$\frac{0.353 - 0}{3 - T_d} = S = 0.224$$

Solving for T_d gives

$$T_d = 1.42$$

Solving for T from Eq. (18.4) gives

$$T = \frac{B_u}{S} = \frac{1}{0.224} = 4.46$$

Having found T_d and T, we can apply the appropriate equations from Table 18.2 to get

$$K_c = 2.91 \qquad \tau_I = 2.86$$

The transient for these settings that was obtained by simulation is shown as curve C-C1 in Fig. 18–15. To our surprise, it is unstable.

Z-N method. We apply the Z-N method for a PI controller, and we obtain the following results: $K_{cu} = 4$, $P_u = 2\pi$, $K_c = 1.8$, and $\tau_I = 5.23$.
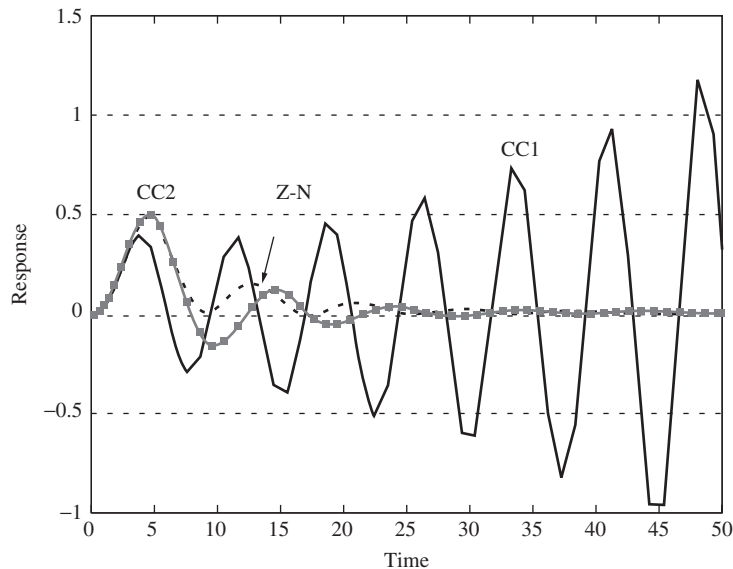
FIGURE 18–15
Comparison of transients produced by different tuning methods for Example 18.4 (shown in Fig. 18–13). Z-N: Ziegler-Nichols method; CC1: Cohen-Coon graphical method; CC2: Cohen-Coon method using least-squares fit of data.

The transient for this set of controller parameters is also shown in Fig. 18–15. We see that the response is stable and well damped.

The lesson learned in this example is that the application of a tuning method may not produce a satisfactory transient. Fine-tuning of these first guesses is usually needed.

Before we abandoned the C-C method for this example, the process reaction curve was fitted to a first-order plus dead-time (FOPDT) model by means

of a least-squares fitting procedure using Loop-Pro and Simulink software. The process was simulated in Simulink and disturbed with a unit step occurring at time $t = 1.0$. The response data were imported into Control Station and fit to a FOPDT model to produce the following results (see Fig. 18–16).
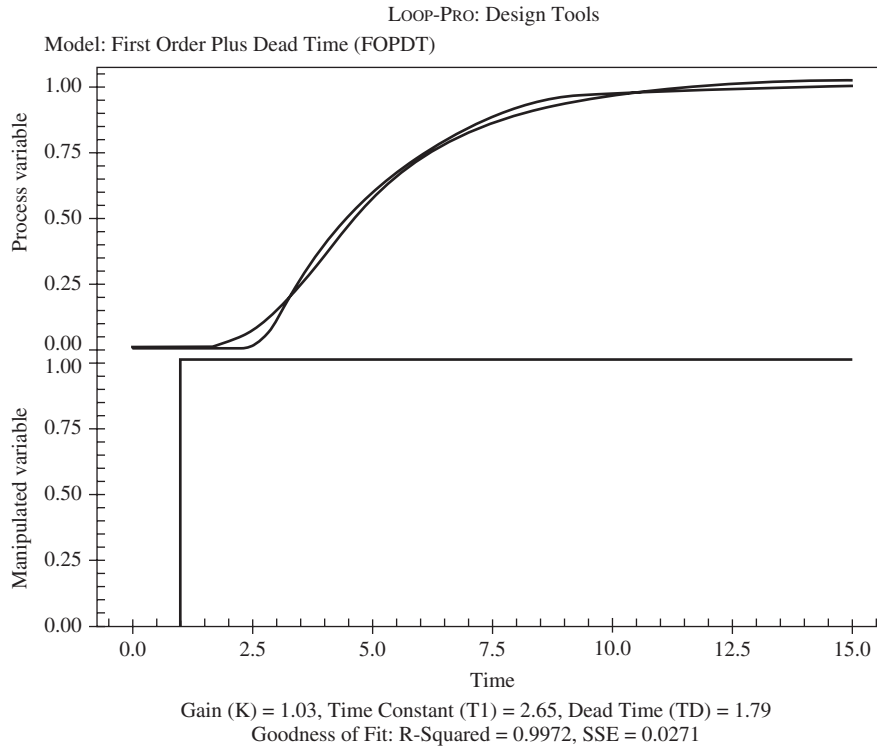
$$T_d = 1.79 \qquad T = 2.65 \qquad K_P = 1.03$$

Loop-Pro: Design Tools

Model: First Order Plus Dead Time (FOPDT)

Gain (K) = 1.03, Time Constant (T1) = 2.65, Dead Time (TD) = 1.79
Goodness of Fit: R-Squared = 0.9972, SSE = 0.0271

FIGURE 18–16
Loop-Pro fit of FOPDT model to process for Example 18.4.

Applying the C-C method for these FOPDT values gives

$$K_c = 1.374 \qquad \text{and} \qquad \tau_I = 2.548$$

Notice that the value of K_c is now considerably less than the value obtained from the fitting procedure shown in Fig. 18–14. This leads to the expectation that the response will now be stable. This expectation is fulfilled as shown by the transient labeled C-C2 in Fig. 18–15. Additionally, we can minimize ISE, ITAE, or IAE to find initial controller parameters as shown in Table 18.5. The response of the process to these controller settings is shown in Fig. 18–17.

The qualitative nature of these responses is similar to that for the Z-N settings and the CC2 settings shown in Fig. 18–15.

TABLE 18.5

Controller settings for Example 18.4 using various tuning methods

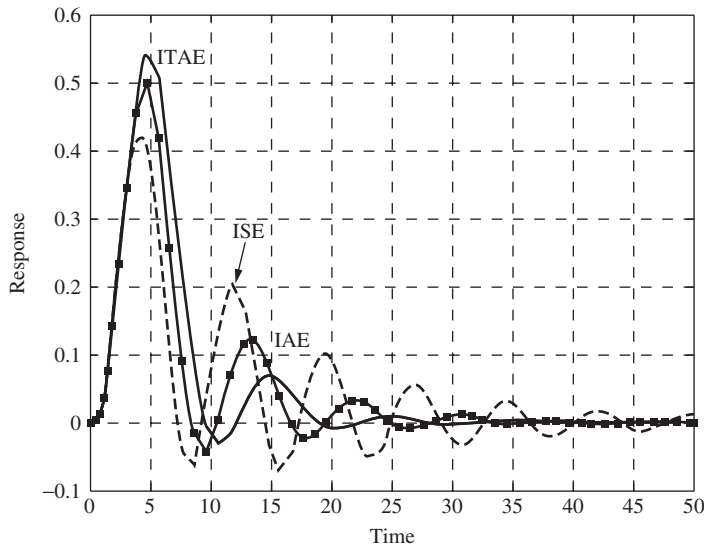| Parameter	Z-N method	C-C method	C-C using Loop-Pro FOPDT fit	Minimum ITAE	Minimum IAE	Minimum ISE
K_c	1.8	2.91	1.374	1.18	1.66	2.45
τ_I	5.23	2.86	2.548	3.28	4.16	5.51
ITAE	24.2	8431	30.7	19.37	21.64	41.33
IAE	2.93	116	3.34	2.96	2.79	3.311
ISE	0.819	284	0.910	1.06	0.816	0.67

FIGURE 18–17

Comparison of transients produced by minimizing the figures of merit for Example 18.4 (shown in Fig. 18–13), ITAE: minimum ITAE; ISE: minimum ISE; IAE: minimum IAE.

18.3   PROCESS IDENTIFICATION

Up to this point, the processes used in our control systems have been described by transfer functions that were derived by applying fundamental principles of physics and chemical engineering (e.g., Newton's law, material balance, heat transfer, fluid mechanics, reaction kinetics) to well-defined processes. In practice, many of the industrial processes to be controlled are too complex to be described by the application of fundamental principles. Either the task requires too much time and effort, or the fundamentals of the process are not understood. By means of experimental tests, one can identify the dynamical nature of such processes and from the results obtain a process model which is at least satisfactory for use in designing control systems. The experimental determination of the dynamic behavior of a process is called *process identification*.

The need for process models arises in many control applications, as we have seen in the use of tuning methods. Process models are also needed in developing feedforward control algorithms, self-tuning algorithms, and internal model control algorithms. Some of these advanced control strategies were discussed in Chap. 17.

Process identification provides several forms that are useful in process control; some of these forms are

Process reaction curve (obtained by step input)
Frequency response diagram (obtained by sinusoidal input)
Pulse response (obtained by pulse input)

We have already encountered the need for process identification in applying the tuning methods presented earlier in this chapter. In the case of the Z-N method, the procedure obtained one point on the open-loop frequency response diagram when the ultimate gain was found. (This point corresponds to a phase angle of $-180°$ and a process gain of $1/K_{cu}$ at the crossover frequency ω_{co}.) In the case of the C-C method, the process identification took the form of the process reaction curve.

Step Testing

As already described in the application of the C-C tuning method, a step change in the input to a process produces a response, which is called the process reaction curve. It is important that no disturbances other than the test step enter the system during the test; otherwise the transient will be corrupted by these uncontrolled disturbances and will be unsuitable for use in deriving a process model. Additionally, the step must be large enough to produce a change in the process that is much larger than any noise present in the process. (If the noise band is 0.5°F, the step should move the process at least 10 times this amount, or at least 5°F. This corresponds to a signal-to-noise ratio of 10.)

For many processes in the chemical industry, the process reaction curve is an S-shaped curve, as shown in Fig. 18–6. For systems that produce an S-shaped process reaction curve, a general model that can be fitted to the transient is a second-order plus dead time (SOPDT) model:

$$G_p(s) = \frac{K_p e^{-T_d s}}{(T_1 s + 1)(T_2 s + 1)} = \frac{Y(s)}{X(s)} \tag{18.7}$$

This model is an extension of the one used in the C-C tuning method, in which there was only one first-order term, the FOPDT model.

Computer software is available to fit a variety of models to process data using a least-squares technique. Loop-Pro is one of the easiest and most efficient. Cooper (2005) makes the following points regarding the successful use of computer software to fit process models to data for the purposes of controller tuning.

1.   The process must be at steady state before data collection begins.
2.   The first data point in the file must be the steady-state value.
3.   The test must be such that the dynamics caused by the test dominate the process noise.

4. There are no disturbance changes present during the test. The change in the process variable must be solely due to the changes introduced in the controller output.
5. Finally, does the model appear to visually fit the process data?

We will now work an example to demonstrate the use of such software to fit models to process data.

Example 18.5. Use LOOP-PRO software to obtain a process model for the two tanks in series process shown in Fig. 18–18. Fit a FOPDT model and a SOPDT model to the data.
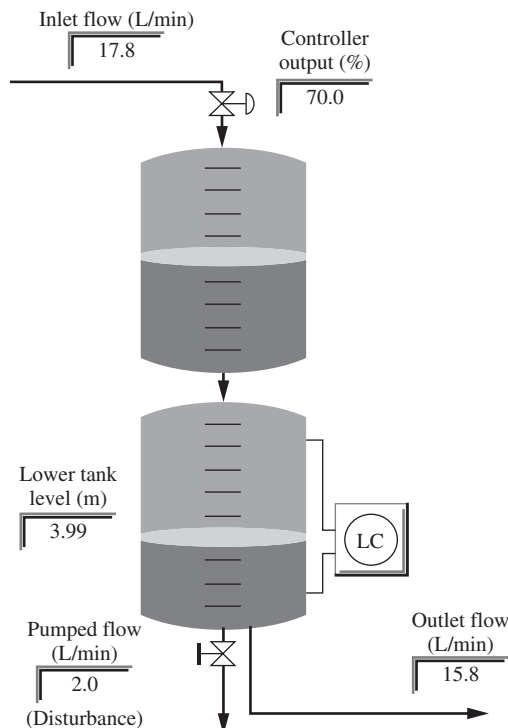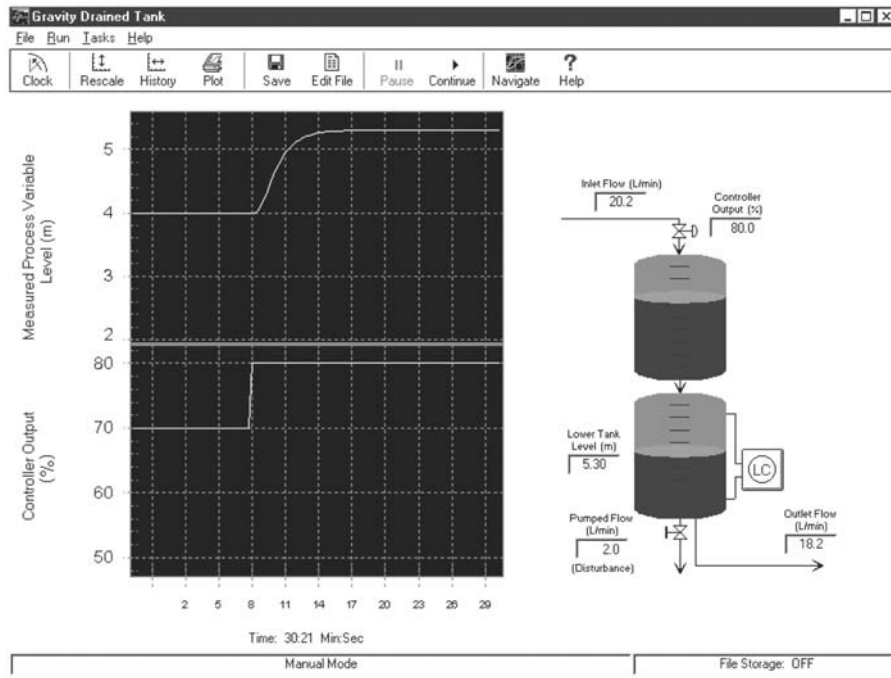

FIGURE 18–18
Screen shot from LOOP-PRO software for gravity-drained tanks.

The process is a pair of gravity-drained tanks. Flow enters the upper tank and drains into the lower tank. A level controller monitors the level in the lower tank and adjusts the flow into the upper tank to maintain the lower level at the desired set point. The process disturbance is a pumped flow (not dependent on tank level) from the lower tank.

We perform a step test on the inlet flow rate to the upper tank and record the data. In this test, the controller output was stepped from 70 to 80 percent, and the process response was recorded (Fig. 18–19).
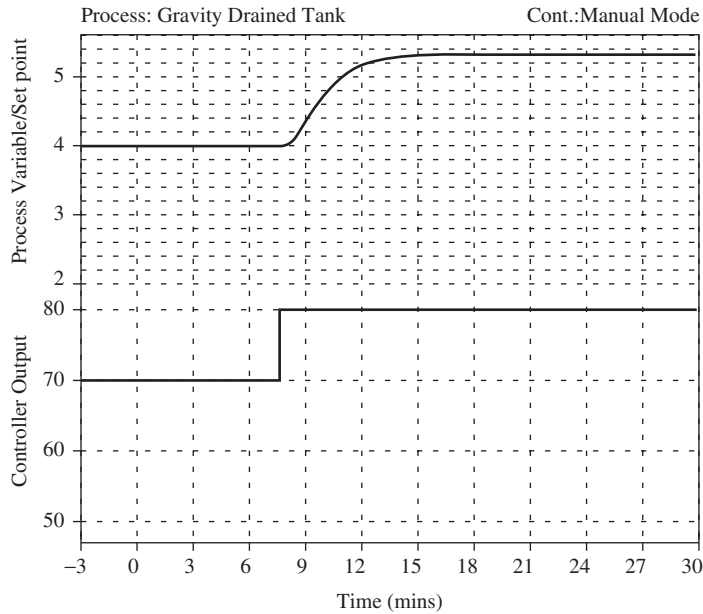
LOOP-PRO: Case Studies

FIGURE 18–19
LOOP-PRO gravity-drained tanks step test for Example 18.5.

Loop-Pro has the built-in capability to fit a FOPDT or SOPDT model (as well as some others) to the process data that we have generated. We use the Design Tools feature of the software and fit an FOPDT model to the data as shown in Fig. 18–20.
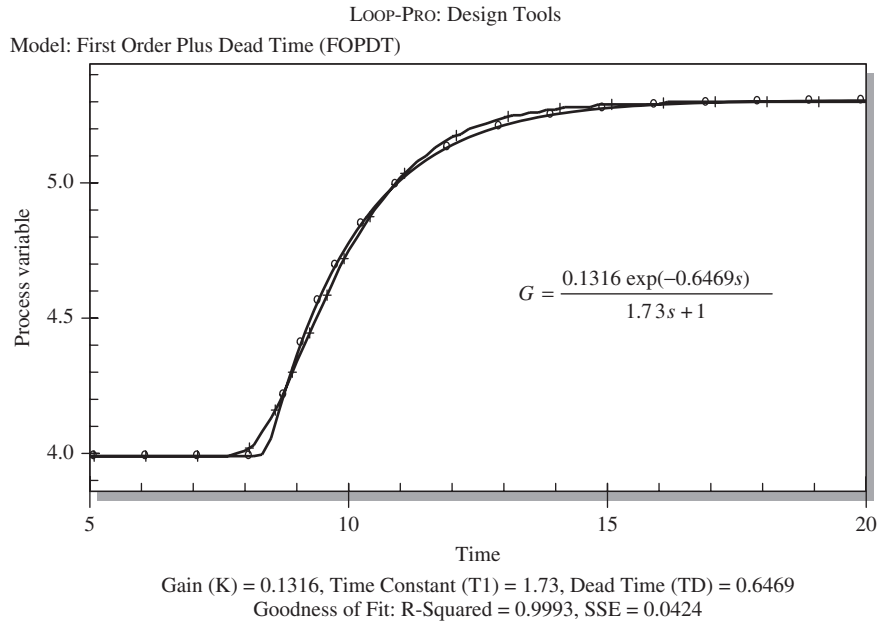
Loop-Pro: Design Tools

Model: First Order Plus Dead Time (FOPDT)

$$G = \frac{0.1316 \exp(-0.6469s)}{1.73s + 1}$$

Gain (K) = 0.1316, Time Constant (T1) = 1.73, Dead Time (TD) = 0.6469
Goodness of Fit: R-Squared = 0.9993, SSE = 0.0424

FIGURE 18–20
Loop-Pro design tools for data analysis. FOPDT fit of step test data for Example 18.5.

The FOPDT model fits the data very well. It captures the essential features of the response. The most noticeable difference is located at the time when the process variable begins to change. The process, being second-order (two tanks in series), rises as an S-shaped curve. The FOPDT, after the dead time, rises with a nonzero slope. There are two indications of the goodness of fit: $R^2 = 0.9993$ is very close to 1.00 (which would indicate a perfect fit). The sum of the squares of the errors, SSE, is defined as

$$\text{SSE} = \sum_{i=1}^{N} \left(\text{Measured Data}_i - \text{Model Data}_i \right)^2 \tag{18.8}$$

For this case SSE = 0.0424, which indicates a very close fit. We can also fit a SOPDT model to the process data as shown in Fig. 18–21.

The SOPDT model fits the data nearly perfectly ($R^2 = 1, \text{SSE} = 0.0019$), as we would expect, since the process is in fact a second-order process.

LOOP-PRO: Design Tools

Model: Second Order Plus Dead Time (SOPDT)

$K = 0.1311$, T1 = 1.04, T2 = 1.04, TD = 0.2208
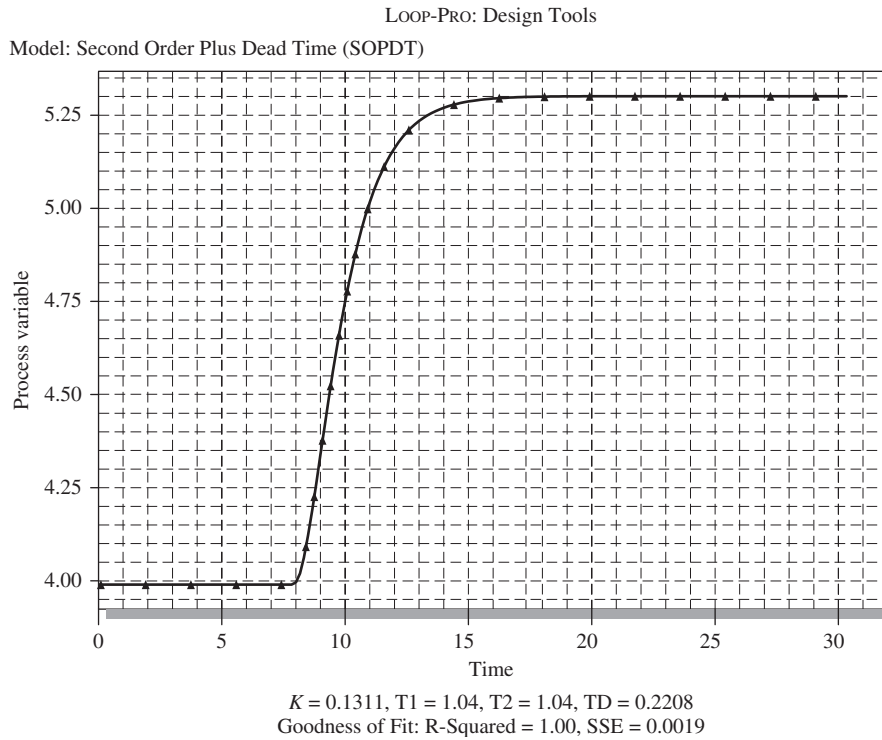Goodness of Fit: R-Squared = 1.00, SSE = 0.0019

FIGURE 18–21
LOOP-PRO design tools for data analysis. SOPDT fit of step-test data for Example 18.5.

Pulse and Doublet Testing

Pulse testing is similar to step testing; the only difference in the experimental procedure is that a pulse disturbance is used in place of a step disturbance. The pulse is introduced as a variation in valve top pressure as was done for step testing (see Fig. 18–5). In applying the pulse, the open-loop system is allowed to reach steady state, after which the valve top pressure is displaced from its steady-state value for a short time and then returned to its original value. The response is recorded at the output of the measuring element (*B* in Fig. 18–5). Usually the pulse shape is rectangular in experimental work, but other well-defined shapes are also used. An example of a pulse change in a process variable is shown in Fig. 18–22.
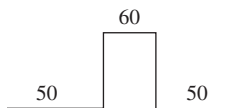

FIGURE 18–22
Sample pulse input.

This brief outline describing pulse testing may appear deceptively simple. In practice, the data on the response must be very accurate and noise-free for the method to succeed. This means that the recorder used to measure the response must be very sensitive. The selection of the pulse height and width is also critical. If the pulse height and width are too small, the disturbance to the system will be too small to produce a transient that can be measured accurately by the recorder. If the pulse height is too large, the system may be operating too far from the linear range of interest. The proper pulse height and width can be determined by some preliminary open-loop experiments. The pulse test is the least disruptive to plant operation among the process identification methods we have considered. The pulse disturbance does not cause the process output to depart far from its normal operating point.

A doublet test is two pulse tests performed in rapid succession and in opposite direction as shown in Fig. 18–23. (Cooper, 2005). For example, we might step the input to the process from 50 to 60 to 40 to 50. This is the equivalent of two pulse tests 50-60-50 and 50-40-50 performed back to back. The process variable does not need to come to steady state for either pulse. It merely needs to show a clear response to the first pulse.
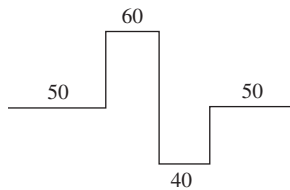

FIGURE 18–23
Sample doublet test input.

The advantages of the doublet test are that it starts from and returns to the design level of operation relatively quickly, it produces data above and below the design level, and from a process standpoint, it minimizes the amount of time that the process is away from the design level, and hence off-spec production (Cooper, 2005).

> **Example 18.6.**   Consider a process with an unknown transfer function such as that shown in Fig. 18–24. We wish to perform some process identification testing on it to determine a FOPDT model for use in controller tuning. Compare the results of pulse and doublet testing for this process.
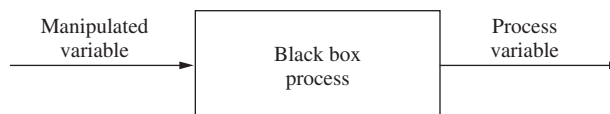

FIGURE 18–24
"Black-box" process for Example 18.6.

For calculation purposes we will use $G = 2/[(s + 1)(s + 2)(0.5s + 1)]$ for the transfer function of the black box.

Pulse test. We will disturb the process by using a pulse input for the manipulated variable, as shown in Fig. 18–22, and use LOOP-PRO to determine the parameters for a FOPDT model (see Fig. 18–25).
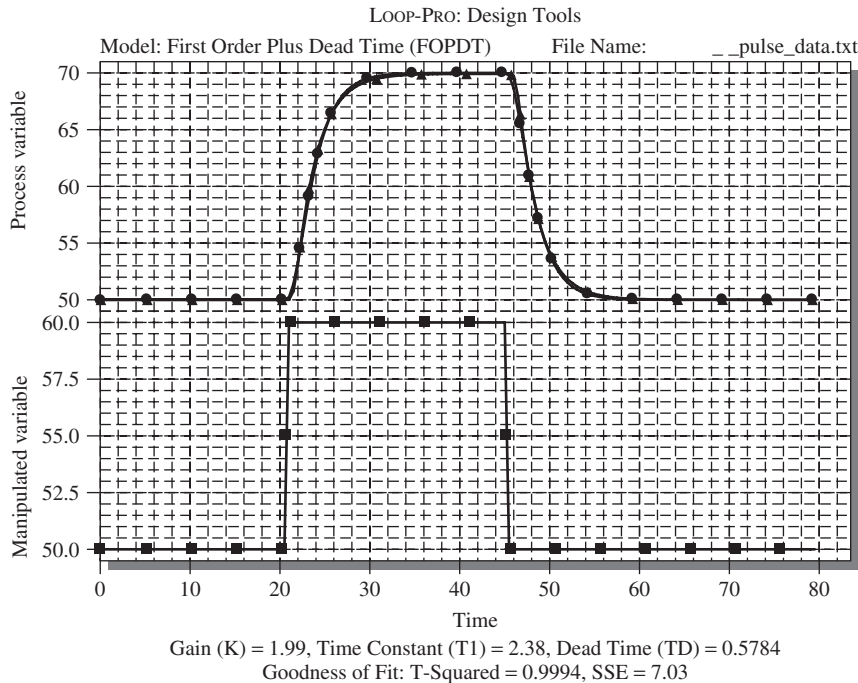

LOOP-PRO: Design Tools

Model: First Order Plus Dead Time (FOPDT)   File Name: _ _pulse_data.txt

Gain (K) = 1.99, Time Constant (T1) = 2.38, Dead Time (TD) = 0.5784
Goodness of Fit: T-Squared = 0.9994, SSE = 7.03

FIGURE 18–25
FOPDT fit of black-box process response to a pulse input (Example 18.6).

The resulting FOPDT model from a least-squares fit of the pulse data is

$$G = \frac{1.99 \exp(-0.5784s)}{2.38s + 1} \tag{18.9}$$

Doublet test. The doublet test on the black-box process is shown in Fig. 18–26. The FOPDT model from a least-squares fit of the doublet data is

$$G = \frac{2.09 \exp(-0.7458s)}{2.70s + 1} \tag{18.10}$$

If we use the model, Eq. (18.9), obtained from the pulse test, and perform a doublet test on it to compare with our actual black-box process response, we obtain Fig. 18–27. Notice that the pulse model deviates from the process response, mainly on the second half of the doublet. The SSE for the pulse model on the doublet
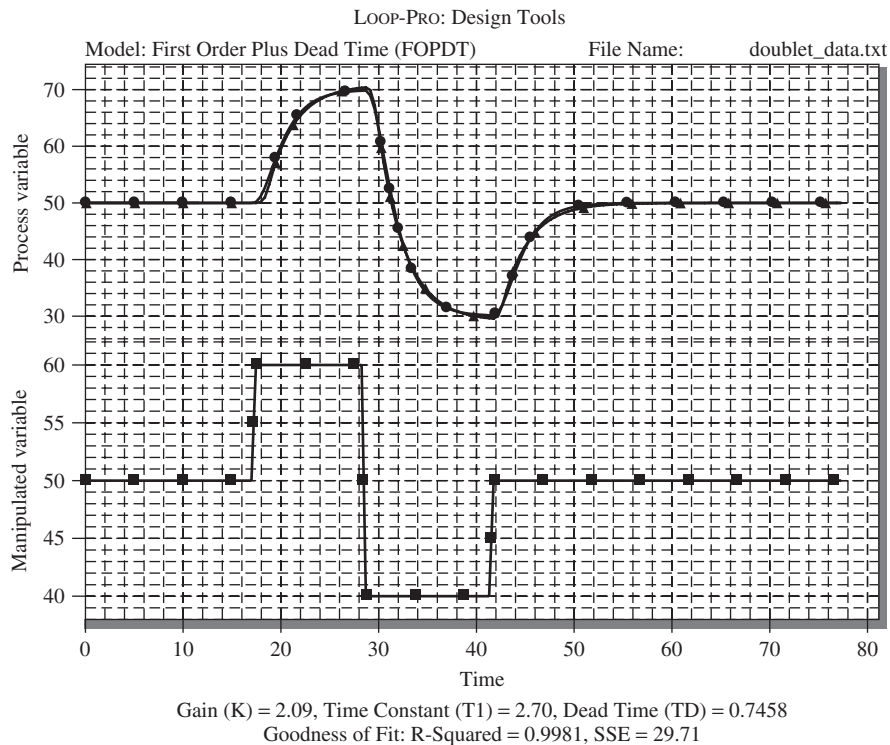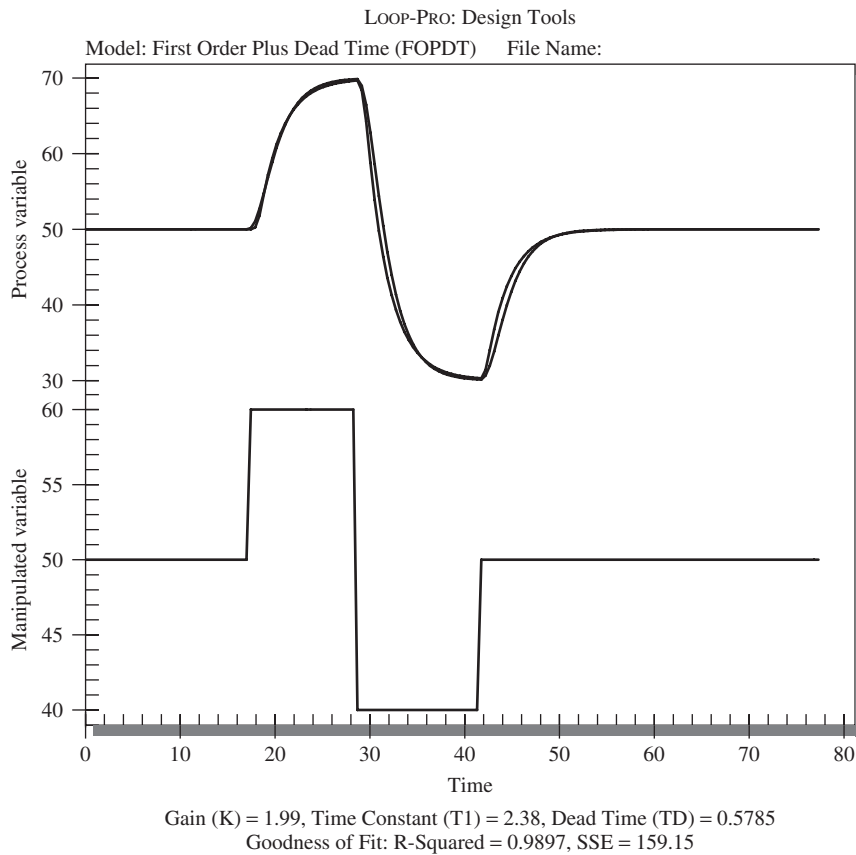
LOOP-PRO: Design Tools

Model: First Order Plus Dead Time (FOPDT)   File Name: doublet_data.txt

Gain (K) = 2.09, Time Constant (T1) = 2.70, Dead Time (TD) = 0.7458
Goodness of Fit: R-Squared = 0.9981, SSE = 29.71

FIGURE 18–26
FOPDT fit of black-box process response to a doublet input (Example 18.6).

test is 159.15 compared to 29.71 for the doublet model. Since we would like the model to capture process characteristics for deviations above and below the design operating level, we would probably be better off using the FOPDT model obtained from the doublet testing to determine our initial tuning parameters.

Frequency Testing

We have shown in the section on frequency response that a process having a transfer function $G(s)$ can be represented by a frequency response diagram (or Bode plot) by taking the magnitude and phase angle of $G(j\omega)$. This procedure can be reversed to obtain $G(s)$ from an experimentally determined frequency response diagram. The procedure requires that a device be available to produce a sinusoidal signal over a range of frequencies. We describe such a device as a *sine wave generator*. In frequency testing of an industrial process, a sinusoidal variation in pressure is applied to the top of the control valve so that the manipulated variable can be varied sinusoidally over a range of frequencies. The block diagram that applies during frequency testing is the same as that of Fig. 18–5 with the step input (*M/s*) replaced by a sinusoidal signal. The sine wave generator used to test electronic devices operates at frequencies that are too high for many slow-moving chemical processes. For frequency testing of chemical processes, special low-frequency generators must be built that can produce a sinusoidal variation

LOOP-PRO: Design Tools

Model: First Order Plus Dead Time (FOPDT)   File Name:

Gain (K) = 1.99, Time Constant (T1) = 2.38, Dead Time (TD) = 0.5785
Goodness of Fit: R-Squared = 0.9897, SSE = 159.15

FIGURE 18–27
A doublet test on the FOPDT model derived from a pulse input test (Example 18.6).

in pressure to a control valve. To preserve the sinusoidal signal in the flow of manipulated variable through the valve, the valve must be linear.

In the 1960s when frequency response methods were first introduced to chemical engineers as a means for process identification, several chemical and petroleum companies constructed mobile units containing low-frequency sine wave generators and recorders that could be moved to processing units in a plant for the purpose of frequency testing.

The great disadvantage of frequency testing is that it takes a long time to collect frequency response data over a range of frequencies that can be used to construct frequency response plots. The time is especially long for chemical processes, often having long time constants measured in minutes or even hours. The frequency test at a given frequency must last long enough to make sure that the transients have disappeared and only the ultimate periodic response is represented by the data. Frequency testing usually ties up plant equipment too long to be recommended as a means of process identification. Step testing, pulse testing, and doublet testing take much less time and can usually provide satisfactory process identification.

SUMMARY

In the practical application of process control, some methods for tuning and process identification are needed. The selection of controller modes depends on the process to be controlled. Proportional control is simple, but the response exhibits offset. The derivative action in PD control makes it possible to increase the controller gain with the result that the response has less offset and responds more quickly compared to proportional control. To eliminate offset, integral action must be present in the controller in the form of PI and PID control. PI control often causes the response to have large overshoot and a slow return to the set point especially for high-order processes. The presence of derivative action in a PID controller gives less overshoot and a faster return to the set point, compared to the response for PI control.

To compare the quality of control on a numerical basis, several criteria that integrate some function of the error with respect to time have been proposed. These include the integral of the square of the error (ISE), the integral of the absolute error (IAE), and the integral of the time-weighted absolute error (ITAE).

In the first part of this chapter two well-known tuning methods are presented: the Ziegler-Nichols (Z-N) method (a closed-loop method) and the Cohen-Coon (C-C) method (an open-loop method). These two methods were applied to several examples and the transients for each compared. The lesson to be learned through these examples is that the controller parameters obtained from a tuning rule should be considered as first guesses; fine online tuning is usually needed to get a satisfactory transient.

The Z-N and C-C methods actually require information about the process model. The Z-N method is based on the ultimate gain at the crossover frequency, which is equivalent to knowing one point on the open-loop frequency response diagram. The C-C method requires the use of an open-loop step response (process reaction curve).

In the advanced control strategies discussed in Chap. 17, a process model is often needed to apply the strategy. When a process model cannot be found by application of theoretical principles, one must obtain a model experimentally. The experimental approach to obtaining a model is called process identification. The three methods of process identification discussed in this chapter are step testing, frequency testing, and pulse testing. The frequency method is seldom used because of the time it takes to test a system over a wide range of frequencies. Step testing is easy to apply and ties the process up for only enough time to obtain one transient. Pulse testing is also simple to apply, but the analysis of the input-output data requires extensive calculations that must be done by a computer.

PROBLEMS

18.1. Calculate C-C and Z-N parameters for PI control for the process in Example 18.5, using the pulse FOPDT model and the doublet FOPDT model. Compare the results for a step change in load for a closed-loop PI control system. Use the ITAE figure of merit for comparison.

CAPSULE SUMMARY

Figures of merit for comparing control parameters (computed from process transients)

Figure of merit	Equation	Purpose
Integral of the square of the error (ISE)	$\text{ISE} = \int_0^\infty e^2 \, dt$	Penalizes large errors, particularly those that occur initially
Integral of the absolute value of error (IAE)	$\text{IAE} = \int_0^\infty |e| \, dt$	Treats all errors (large and small) the same
Integral of time-weighted absolute error (ITAE)	$\text{ITAE} = \int_0^\infty |e| \, t \, dt$	Penalizes errors that persist for a long time

First-order plus dead-time (FOPDT) model

$$G_p(s) = \frac{K_p e^{-T_d s}}{(T_1 s + 1)} = \frac{Y(s)}{X(s)}$$

Second-order plus dead-time (SOPDT) model

$$G_p(s) = \frac{K_p e^{-T_d s}}{(T_1 s + 1)(T_2 s + 1)} = \frac{Y(s)}{X(s)}$$

Ziegler-Nichols controller settings

Type of control	$G_c(s)$	K_c	τ_I	τ_D
Proportional (P)	K_c	$0.5K_u$		
Proportional-integral (PI)	$K_c\left(1 + \dfrac{1}{\tau_I s}\right)$	$0.45K_u$	$\dfrac{P_u}{1.2}$	
Proportional-integral-derivative (PID)	$K_c\left(1 + \dfrac{1}{\tau_I s} + \tau_D s\right)$	$0.6K_u$	$\dfrac{P_u}{2}$	$\dfrac{P_u}{8}$

Cohen-Coon controller settings

Type of control	Parameter setting
Proportional (P)	$K_c = \dfrac{1}{K_p} \dfrac{T}{T_d} \left(1 + \dfrac{T_d}{3T} \right)$
Proportional-integral (PI)	$K_c = \dfrac{1}{K_p} \dfrac{T}{T_d} \left(\dfrac{9}{10} + \dfrac{T_d}{12T} \right)$
	$\tau_I = T_d \dfrac{30 + 3T_d/T}{9 + 20T_d/T}$
Proportional-derivative (PD)	$K_c = \dfrac{1}{K_p} \dfrac{T}{T_d} \left(\dfrac{5}{4} + \dfrac{T_d}{6T} \right)$
	$\tau_D = T_d \dfrac{6 - 2T_d/T}{22 + 3T_d/T}$
Proportional-integral-derivative (PID)	$K_c = \dfrac{1}{K_p} \dfrac{T}{T_d} \left(\dfrac{4}{3} + \dfrac{T_d}{4T} \right)$
	$\tau_I = T_d \dfrac{32 + 6T_d/T}{13 + 8T_d/T}$
	$\tau_D = T_d \dfrac{4}{11 + 2T_d/T}$

One of the basic components of any control system is the final control element, which comes in a variety of forms depending on the specific control application. The most common type of final control element in chemical processing is the pneumatic control valve, which regulates the flow of fluids. Some other types include the variable-speed pump and the power controller (used in electrical heating).

Since the pneumatic control valve is so widely used in chemical processing, this chapter will be devoted to the description, selection, and sizing of control valves.

19.1   CONTROL VALVE CONSTRUCTION

The control valve is essentially a variable resistance to the flow of a fluid in which the resistance, and therefore the flow, can be changed by a signal from a process controller.

As shown in Fig. 19–1, the control valve consists of an actuator and a valve. The valve itself is divided into the body and the trim. The body consists of a housing for mounting the actuator and connections for attachment of the valve to a supply line and a delivery line. The trim, which is enclosed within the body, consists of a plug, a valve seat, and a valve stem. The actuator moves the valve stem as the pressure on a spring-loaded diaphragm changes. The stem moves a plug in a valve seat to change the resistance to flow through the valve. When a valve is supplied by the manufacturer, the actuator and the valve are attached to each other to form one unit.

For most actuators, the motion of the stem is proportional to the pressure applied on the diaphragm. In general, this type of actuator can be used for functions other than moving a valve stem. For example, it can be used to adjust dampers, variable-speed drives, rheostats, and other devices. As the pressure to the valve varies over its normal

range of operation (3 to 15 psig), the range of motion of the stem varies from a fraction of an inch to several inches depending on the size of the actuator. Manufacturers provide a range of actuators for various valve sizes.

The valves available vary over a wide range of sizes. The size is usually referred to by the size of the end connectors. For example, a 1-in valve would have connectors (threaded or flanged) to fit into a 1-in pipeline. In general, the larger the valve size, the larger the flow capacity of the valve.

For the control valve shown in Fig. 19–1, an increase in signal pressure above the diaphragm exerts a force on the diaphragm and back plate, which causes the stem to move down; this causes the cross-sectional area for flow between the plug and the seat to decrease, thereby reducing or throttling the flow. Such valve action as shown in Fig. 19–1 is called air-to-close (AC) action. The reverse action, air-to-open (AO), can be accomplished by designing the actuator so that pressure is applied to the underside of the diaphragm, for which case an increase in pressure to the valve raises the stem. An alternate method to reverse the valve action is to leave the actuator as shown in Fig. 19–1 and to invert the plug on the stem and place it under the valve seat. In general, selection of the type of valve (AO or AC) is made based on safety considerations. We would like the valve to fail in a safe position for the process in the event of a loss of air pressure. For example, if the control valve is controlling the inlet flow of cooling water to a cooling jacket on an exothermic chemical reactor, we would like the valve to fail in the open position so that we do not lose coolant flow. Thus, we would select an air-to-close (AC) valve.
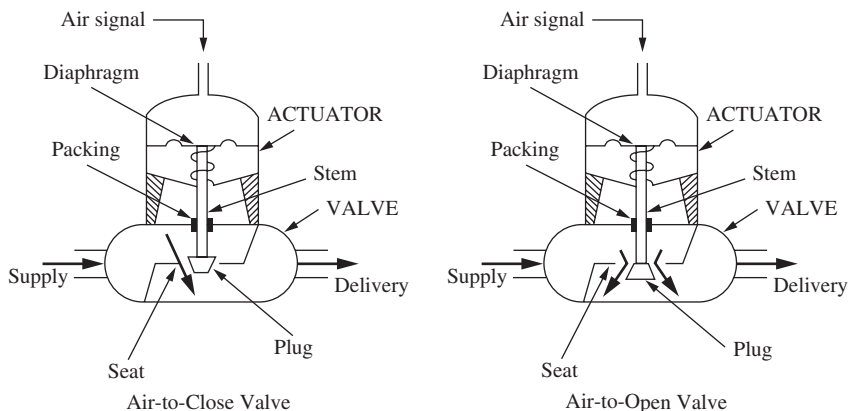

FIGURE 19–1
Single-seated control valves.

The valve shown in Fig. 19–1 is single-seated, meaning the valve contains one plug with one seating surface. For a single-seated valve, the plug must open against the full pressure drop across the valve. If the pressure drop is large, this means that a larger, more expensive actuator will be needed. To overcome this problem, valves are also constructed with double seating as shown in Fig. 19–2. In this type valve, two plugs are attached to the valve stem, and each one has a seat. The flow pattern through the

valve is designed so that the pressure drop across the seat at *A* tends to open the plug and the pressure drop across the seat at *B* tends to close the plug. This counterbalancing of forces on the plugs reduces the effort needed to open the valve with the result that a smaller, less expensive actuator is needed.
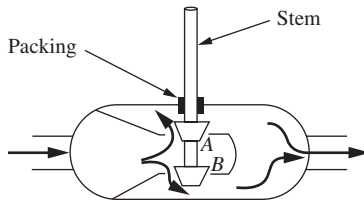

FIGURE 19–2
Double-seated control valve.

In a double-seated valve, it is difficult to have tight shutoff. If one plug has tight closure, there is usually a small gap between the other plug and its seat. For this reason, single-seated valves are recommended if the valve is required to be shut tight. In many processes, the valve is used for throttling flow and is never expected to operate near its shutoff position. For these conditions, the fact that the valve has a small leakage at shutoff position does not create a problem.

19.2   VALVE SIZING

To specify the size of a valve in terms of its capacity, the following equation is used:

$$q = C_v f(x) \sqrt{\frac{\Delta p_{\text{valve}}}{\text{sg}}} \tag{19.1}$$

where

q = flow rate, gal/min
$f(x)$ = fraction of maximum flow (= 1 for fully open)
x = fractional stem position (i.e., fraction open)
Δp_{valve} = pressure drop across the valve, psi
sg = specific gravity of fluid at stream temperature relative to water; for water sg = 1
C_v = factor associated with capacity of valve

Equation (19.1) applies to the flow of an incompressible, nonflashing fluid through the valve. Manufacturers rate the size of a valve in terms of the factor C_v. Sometimes C_v is defined as the flow (gal/min) of a fluid of unit specific gravity (water) through a fully open valve, across which a pressure drop of 1.0 psi exists. This verbal definition is, of course, obtained directly from Eq. (19.1) by letting $f(x) = 1$, $\Delta p_{\text{valve}} = 1$, and sg = 1. Equation (19.1) is based on the well-known Bernoulli equation for determining the pressure drop across valves and resistances. It is important to emphasize that C_v must be

determined from Eq. (19.1) using the units listed. Since so many valves in use are rated in terms of C_v, Eq. (19.1) is of practical importance; however, some industries now are defining a valve coefficient K_v defined by the equation

$$q = K_v \sqrt{\frac{\Delta p_{\text{valve}}}{\text{sg}}}$$

where q = flow rate, m^3/h
        Δp_{valve} = pressure drop across valve, kg_f/cm^2
            sg = specific gravity of fluid relative to water
The relation between K_v and C_v is

$$K_v = 0.856 C_v$$

For gases and steam, modified versions of Eq. (19.1) are used in which C_v is still used as a factor.

Manufacturers of valves provide brochures, nomographs, and computer programs for sizing valves for use with gases and steam.

In general, as the physical size of a valve body (i.e., size of pipe connectors) increases, the value of C_v increases. For a sliding stem and plug type of control valve, the value of C_v is roughly equal to the square of the pipe size multiplied by 10. Using this rule, a 3-in control valve should have a C_v of about 90, with units corresponding to those of Eq. (19.1). This implies the capacity of the fully open valve is 90 gal/min with a pressure drop of 1 psi.

> **Example 19.1**   A valve with a C_v rating of 4.0 is used to throttle the flow of glycerine for which sg = 1.26. Determine the maximum flow through the valve for a pressure drop of 100 psi.
>
> $$q = 4.0 \sqrt{\frac{100}{1.26}} = 35.6 \text{ gal/min}$$

The coefficient C_v varies with the design of the valve (shape, size, roughness) and the Reynolds number for the flow through the valve. This relationship is analogous to the relationship between the friction factor and roughness and the Reynolds number for flow through a pipe. For relatively nonviscous fluids, C_v in Eq. (19.1) can be taken as a constant for a valve of given size and type. The reason for this is that at high Reynolds numbers, the friction factor changes very little with flow rate. Except for very viscous fluids, the flow through a valve, which involves sudden contraction and expansion, is in the turbulent regime of fluid flow; turbulence in the valve exists even if the flow in the supply pipe is near the critical Reynolds number of 2100.

Consequently, for relatively nonviscous fluids, Eq. (19.1) is satisfactory for sizing a valve for any fluid. For the control of flow of very viscous fluids, such as tar or molasses, the value of C_v found from Eq. (19.1) must be multiplied by a correction factor that depends on viscosity, density, flow rate, and valve size (i.e., on the Reynolds number). Methods for determining the viscosity correction factor are provided by manufacturers for their valves. If one does not apply the correction factor for a very viscous fluid, the value of C_v will be too low and the valve will be undersized.

19.3   VALVE CHARACTERISTICS

The function of a control valve is to vary the flow of fluid through the valve by means of a change of pressure to the valve top. The relation between the flow through the valve and the valve stem position (or *lift*) is called the *valve characteristic,* which can be conveniently described by means of a graph as shown in Fig. 19–3 where three types of characteristics are illustrated.
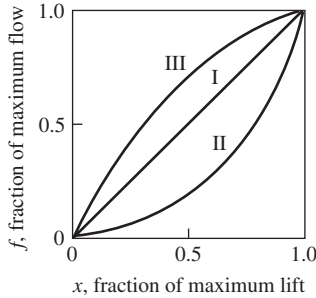

FIGURE 19–3
Inherent valve characteristics (pressure drop across valve is constant).
I: Linear, II: increasing sensitivity (e.g., equal-percentage valve), and III: decreasing sensitivity (e.g., square root valve).

In general, the flow through a control valve for a specific fluid at a given temperature can be expressed as

$$q = f_1(L, p_0, p_1) \qquad (19.2)$$

where q = volumetric flow rate
   L = valve stem position (or lift)
   p_0 = upstream pressure
   p_1 = downstream pressure

The inherent valve characteristic is determined for fixed values of p_0 and p_1, for which case Eq. (19.2) becomes

$$q = f_2(L) \qquad (19.3)$$

or, in other words, the flow is a function of the valve stem position.

For convenience, let

$$f = \frac{q}{q_{max}} \qquad \text{and} \qquad x = \frac{L}{L_{max}}$$

where q_{max} = maximum flow when valve is fully open (stem is at its maximum lift L_{max})
   x = fraction of maximum lift
   f = fraction of maximum flow

In general, f will be a function of x, which we will denote as $f(x)$. Equation (19.3) may be written as

$$f\left(\frac{L}{L_{max}}\right) = f(x) = \frac{q}{q_{max}} \qquad (19.4)$$

The types of valve characteristics can be defined in terms of the sensitivity of the valve, which is simply the fractional change in flow to the fractional change in stem position for fixed upstream and downstream pressures; mathematically, sensitivity may be written

$$\text{Sensitivity} = \frac{df}{dx}$$

In terms of valve characteristics, valves can be divided into three types: decreasing sensitivity (square root or quick-opening), linear, and increasing sensitivity (equal-percentage). These types are shown in Fig. 19–3 where the fractional flow $f(x)$ is plotted against fractional lift x. For the decreasing sensitivity type, the sensitivity (or slope) decreases with increasing flow. For the linear type, the sensitivity is constant and the characteristic curve is a straight line. For the increasing sensitivity type, the sensitivity increases with flow.

Valve characteristic curves, such as the ones shown in Fig. 19–3, can be obtained experimentally for any valve by measuring the flow through the valve as a function of lift (or valve-top pressure) under conditions of constant upstream and downstream pressures. Two types of valves that are widely used are the linear valve and the equal-percentage valve.

Linear Valves

The linear valve is one for which the sensitivity is constant and the relation between flow and lift is linear. For the linear valve, the mathematical relationship is

$$\frac{df}{dx} = \alpha \tag{19.5}$$

where α is a constant.

Assuming that the valve is shut tight when the lift is at lowest position, we have $f = 0$ at $x = 0$. For a single-seated valve that is not badly worn, the valve can be shut off for $x = 0$. Recall that the definitions of x and f require that $f = 1$ at $x = 1$. Integrating Eq. (19.5) and introducing the limits $f = 0$ at $x = 0$ and $f = 1$ at $x = 1$ give

$$\int_0^1 df = \int_0^1 \alpha\, dx$$

Integrating this equation and inserting limits give

$$\alpha = 1$$

For $\alpha = 1$, Eq. (19.5) can now be integrated to give

$$\boxed{f(x) = x \qquad \text{linear valve}} \tag{19.6}$$

Equal-Percentage Valves

For the equal-percentage valve, the defining equation is

$$\frac{df}{dx} = \beta f \tag{19.7}$$

where β is constant. Integration of this equation gives

$$\int_{f_0}^{f} \frac{df}{f} = \int_{0}^{x} \beta \, dx \qquad (19.8)$$

or

$$\ln \frac{f}{f_0} = \beta x \qquad (19.9)$$

where f_0 is the flow at $x = 0$. Rearranging gives

$$f = f_0 e^{\beta x} \qquad (19.10)$$

The term β can be expressed in terms of f_0 by inserting $f = 1$ at $x = 1$ into Eq. (19.9). The result is

$$\beta = \ln \frac{1}{f_0}$$

If we substitute for β in Eq. (19.10), we obtain

$$f = f_0 e^{x \ln(1/f_0)} = f_0 e^{\ln(1/f_0)^x} = f_0 e^{\ln(f_0)^{-x}} = f_0 f_0^{-x} = f_0^{1-x}$$

which is of the form

$$\boxed{f = \alpha^{x-1} \qquad \text{equal-percentage valve}} \qquad (19.11)$$

where $\alpha = 1/f_0$ is a constant.

Equation (19.9) shows that a plot of f versus x on semilog coordinates gives a straight line. A convenient way to determine if a valve is of the equal-percentage type is to plot the flow versus lift on semilog coordinates. The relation expressed by Eq. (19.9) is the basis for calling the valve characteristic logarithmic. The basis for calling the valve characteristic equal percentage can be seen by rearranging Eq. (19.7) into the form

$$\frac{df}{f} = \beta \, dx \qquad \text{or} \qquad \frac{\Delta f}{f} = \beta \, \Delta x$$

In this form it can be seen that an equal fractional (or percentage) change in flow $\Delta f/f$ occurs for a specified increment of change in stem position Δx, regardless of where the change in stem position occurs along the characteristic curve. In integrating Eq. (19.7), the flow was assumed to be f_0 at $x = 0$. Mathematically this is necessary, because f_0 cannot be taken as zero at $x = 0$ because the term on the left side of Eq. (19.9) becomes infinite. In practice, there may be some leakage (hence $f_0 \neq 0$) when the stem is at its lowest position for a double-seated valve or for a valve in which the plug and seat have become worn.

For some valves, especially large ones, the valve manufacturer intentionally allows some leakage at minimum lift ($x = 0$) to prevent binding and wearing of the plug and seat surfaces. For a valve that does shut tight and is also classified as an equal-percentage valve, the equal-percentage characteristic will not be followed when the valve is nearly shut. In practice, the control valve serves as a throttling valve and is not intended to be wide open or completely closed during normal operation.

To express the range over which an equal-percentage valve will follow the equal-percentage characteristic, the term *rangeability* is used. Rangeability is defined as the ratio of maximum flow to minimum controllable flow over which the valve characteristic is followed.

$$\text{Rangeability} = \frac{f(x)_{\text{max}}}{f(x)_{\text{min,controllable}}}$$

For example, if f_0 is 0.02, the rangeability is 50. It is not uncommon for a control valve to have a rangeability as high as 50.

In practice, the ideal characteristics for linear and equal-percentage valves are only approximated by commercially available valves. These discrepancies cause no difficulty because the inherent characteristics are changed considerably when the valve is installed in a line having resistance to flow, a situation that usually prevails in practice. The inherent valve characteristics are shown in Table 19.1. In the next section, the effect of line loss on the effective valve characteristic will be discussed.

TABLE 19.1
Inherent valve characteristics

Valve type	Sensitivity $\dfrac{df}{dx}$	Relationship
Linear	Constant	$f(x) = x$
Equal-percentage	Increasing	$f(x) = \alpha^{x-1}$
Square root or quick-opening	Decreasing	$f(x) = \sqrt{x}$

Effective Valve Characteristic

When a valve is placed in a line that offers resistance to flow, the inherent characteristic of the valve will be altered. The relation between flow and stem position (or valve-top pressure) for a valve installed in a process line is called the *effective valve characteristic.*
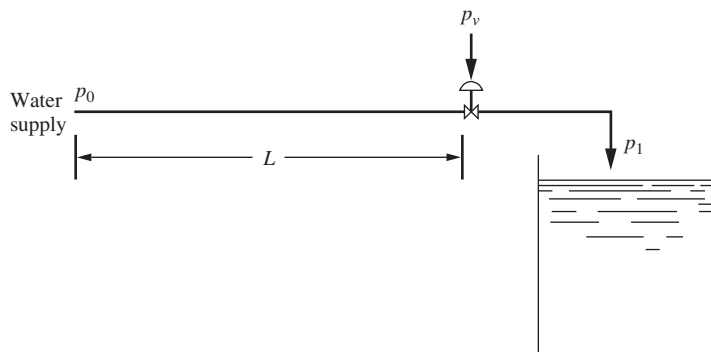

FIGURE 19–4
Control valve with supply line.

Consider a linear control valve attached to the end of a pipeline that delivers water to an open tank. A diagram of the system is shown in Fig. 19–4. If the pipe is of large diameter relative to the size of the control valve, the pressure drop in the line will be negligible and the full pressure drop $p_0 - p_1$ will be across the valve as the lift varies between 0 and 1 (from fully closed to fully open). In this case a plot of flow versus lift will give a linear relation as shown by curve I of Fig. 19–5. This curve is for the flow of water at 5°C through a control valve for which $C_v = 4.0$ and the overall pressure drop $p_0 - p_1$ is 100 psi. To show the effect of line loss, curve II is constructed for the same conditions as curve I, with the exception that 100 ft of 1.0-in (inside diameter) pipe is used to supply the valve.

Example 19.2 will give the detailed calculations used to obtain the results in Fig. 19–5. For 100 ft of pipe, the plot of flow versus lift gives curve II, shown in Fig. 19–5, in which the curve falls away or droops from the linear relation that holds for no line loss. Since line loss is proportional to the square of the velocity, the line loss is very small when the valve is nearly closed, for which case the total pressure drop is across the valve. For this reason, curves I and II in Fig. 19–5 are close together at low rates.

A rule often followed in industrial application of control valves is that the pressure drop across the wide-open valve should be greater than 25 percent of the pressure drop across the closed valve. A valve not selected according to this rule will lose its effectiveness to control at high flow rates.
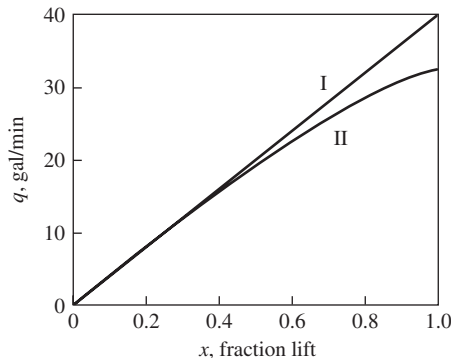

FIGURE 19–5
Effect of line loss on control valve characteristics from Example 19.2.
I: No pressure drop in supply line to valve, II: pressure drop present in supply line to valve.

Example 19.2.   Determine the flow versus lift relation for the linear control valve installed in the flow system of Fig. 19–4. The fluid is water at 5°C. The following data apply.

Pipe length	100 ft
Inside pipe diameter	1 in
Density of water	62.4 lb/ft^3
Viscosity of water	1.5 cP
C_v of control valve	4.0
Total pressure drop $p_0 - p_1 = \Delta p_{\text{total}}$	100 psi

If there is no line loss, as is the case for a large-diameter line, then $\Delta p_{total} = \Delta p_{valve}$, and the maximum flow can be calculated from Eq. (19.1):

$$q = C_v f(x) \sqrt{\frac{\Delta p_{valve}}{sg}} = 4.0(1.0)\sqrt{\frac{100}{1}} = 40.0 \text{ gal/min}$$

To determine the flow/lift relation for the case of line loss, we arbitrarily start the calculation with a flow of 30 gal/min. The pressure drop in the 100 ft-pipe can be calculated from the well-known expression from fluid mechanics

$$\Delta p_{pipe} = \frac{32 f\; L\rho Q^2}{144\pi^2 g_c D^5} \tag{19.12}$$

where Δp_{pipe} = pressure loss in line, psi

$$g_c = 32.174 \frac{\text{lb}_m \cdot \text{ft} / \text{s}^2}{\text{lb}_f}$$

ρ = density of fluid, lb_m/ft^3
f = Fanning friction factor, dimensionless
Q = volumetric flow through pipe, ft^3/s
L = pipe length, ft
D = inside pipe diameter, ft

The Fanning friction factor is a function of the Reynolds number and the pipe roughness. Equation (19.11) and a correlation for the Fanning friction factor can be found in the literature (Perry and Green, 1999). We now calculate the Reynolds number Re:

$$Re = \frac{Du\rho}{\mu}$$

Replacing the velocity u with $Q/[(\pi/4)D^2]$ gives

$$Re = \frac{4Q\rho}{\pi\mu D} \tag{19.13}$$

$$Q = \frac{30 \text{ gal/min}}{(60 \text{ s/min})(7.48 \text{ gal/ft}^3)} = 0.0668 \frac{\text{ft}^3}{\text{s}} = 240.6 \frac{\text{ft}^3}{\text{h}}$$

$$Re = \frac{(4)(240.6 \text{ ft}^3/\text{h})(62.4 \text{ lb/ft}^3)}{(\pi)\left(\frac{1}{12}\text{ ft}\right)(1.50 \text{ cP})\left[2.42 \frac{\text{lb}/(\text{ft}\cdot\text{h})}{\text{cP}}\right]} = 63,220$$

For this value of Reynolds number and for smooth pipe, the Fanning friction factor f is 0.005. Equation (19.12) may now be used to calculate the line loss:

$$\Delta p_{pipe} = \frac{(32)(0.005)(100)(62.4)(0.0668)^2}{(144)\left(\pi^2\right)(32.2)\left(\frac{1}{12}\right)^5} = 24.2 \text{ psi}$$

Therefore, $\Delta p_{valve} = 100 \text{ psi} - 24.2 \text{ psi} = 75.8 \text{ psi}$.

We next calculate the flow through the wide-open valve that gives a pressure drop of 75.8 psi:

$$q_{max} = C_v \underbrace{f(x)}_{\substack{x \text{ for linear,} \\ f(x)=x=1 \\ \text{for max. flow}}} \sqrt{\frac{\Delta p_{valve}}{sg}} = 4.0 \sqrt{\frac{75.8}{1}} = 34.8 \text{ gal / min}$$

Recall that we arbitrarily started this calculation with a flow rate of 30 gal/min. Now, since a flow of 34.8 gal/min through the wide-open valve produces a pressure drop of 75.8 psi, we conclude that the valve must be partially closed since we have $\Delta p_{valve} = 75.8$ psi with only 30 gal/min of flow. Since the valve is linear, we can calculate the lift x as follows:

$$\underbrace{f(x) = x}_{\text{Linear valve}} = \frac{q}{q_{max}} = \frac{30}{34.8} = 0.86$$

We can generalize this calculation to generate the data necessary to produce Fig. 19–5.

$$x = \frac{q}{q_{max}} = \frac{q}{C_v \sqrt{\Delta p_{valve} / sg}} = \frac{q}{4 \sqrt{\Delta p_{valve}}}$$

We also know that

$$\Delta p_{total} = 100 \text{ psi} = \Delta p_{pipe} + \Delta p_{valve}$$

We can observe that the pressure drop through the pipe is proportional to the square of the flow rate through the pipe [Eq. (19.12)]. Therefore, we can write

$$\frac{\left(\Delta p_{pipe}\right)_{old}}{\left(\Delta p_{pipe}\right)_{new}} = \frac{q_{old}^2}{q_{new}^2} = \left(\frac{q_{old}}{q_{new}}\right)^2$$

We have one data point that we have already calculated: $q = 30$ gal/min and $\Delta p_{valve} = 24.2$ psi, so

$$\frac{24.2 \text{ psi}}{\Delta p_{pipe}} = \left(\frac{30 \text{ gpm}}{q}\right)^2$$

and

$$\Delta p_{pipe} = \left(\frac{q}{30 \text{ gpm}}\right)^2 (24.2 \text{ psi})$$

From the expression for the total pressure drop, we can write

$$\Delta p_{valve} = \Delta p_{total} - \Delta p_{pipe} = 100 \text{ psi} - \left(\frac{q}{30 \text{ gpm}}\right)^2 (24.2 \text{ psi})$$

Finally, we can substitute this back into the expression for x.

$$x = \frac{q}{4\sqrt{\Delta p_{\text{valve}}}} = \frac{q}{4\sqrt{100 \text{ psi} - \left(\dfrac{q}{30 \text{ gpm}}\right)^2 (24.2 \text{ psi})}}$$

Using this equation, we can generate the effective characteristic curve (curve II) shown in Fig. 19–5. The data are summarized in Table 19.2.

TABLE 19.2
Effective characteristic for a linear valve with supply line loss

	Curve I pressure drop in valve only		Curve II pressure drop in valve and pipe	
Flow q, gal/min	Stem position x	Valve pressure drop, psi	Stem position x	Valve pressure drop, psi
0	0.00	100	0.00	100
10	0.25	100	0.25	97
20	0.50	100	0.53	89
30	0.75	100	0.86	76
30.5	0.76	100	0.88	75
31	0.78	100	0.90	74
31.5	0.79	100	0.92	73
32	0.80	100	0.94	72
32.5	0.81	100	0.96	72
33	0.83	100	0.98	71
33.5	0.84	100	1.00	70
34	0.85	100		
34.5	0.86	100		
35	0.88	100		
35.5	0.89	100		
36	0.90	100		
36.5	0.91	100		
37	0.93	100		
37.5	0.94	100		
38	0.95	100		
38.5	0.96	100		
39	0.98	100		
39.5	0.99	100		
40	1.00	100		

Example 19.3.   A control valve is to be installed in the flow system of Fig. 19–4. The valve is supplied by water at 5°C through 200 ft of pipe having an inside diameter of 1.0 in. The total pressure drop $p_0 - p_1$ is 100 psi. When the valve is wide open, the flow is to be 30 gal/min. Determine C_v for the valve. Plot the effective characteristic curve for the valve as flow versus lift. Do this problem for a linear valve and for an equal-percentage valve. The equal-percentage valve has an $\alpha = 33.3$.

Linear valve. To obtain the pressure drop in the line, use is made of Eqs. (19.12) and (19.13) as was done in Example 19.2. From Eq. (19.13), we obtain the Reynolds number as follows:

$$q = \frac{30}{(60)(7.48)} = 0.0668 \frac{ft^3}{s} = 240.6 \frac{ft^3}{h}$$

$$Re = \frac{4q\rho}{\pi\mu_D} = \frac{(4)\left(240.6 \ ft^3/h\right)\left(62.4 \ lb/ft^3\right)}{(\pi)\left(\frac{1}{12}ft\right)(1.50 \ cP)\left[2.42\frac{lb/(ft \cdot h)}{cP}\right]} = 63,220$$

From a correlation for the Fanning friction factor, we obtain $f = 0.005$. From Eq. (19.12), the line loss is calculated to be

$$\Delta p = \frac{(32)(0.005)(200)(62.4)(0.0668)^2}{(144)\left(\pi^2\right)(32.2)\left(\frac{1}{12}\right)^5} = 48.4 \ psi$$

$$\Delta p_{valve} = 100 - 48.4 = 51.6 \ psi$$

From knowledge of the maximum flow through the wide-open valve (30 gal/min) and Δp_{valve}, we calculate C_v from Eq. (19.1) as follows:

$$C_v = \frac{q_{max}}{\sqrt{\Delta p_{valve}/sg}} = \frac{30}{\sqrt{51.6}} = 4.18$$

Now that we have C_v, we can calculate the stem position x needed for various flow rates q.

As we showed in Example 19.2, the pressure drop through the pipe is proportional to the square of the flow rate. Since the total pressure drop is constant, we can calculate the pressure drop across the valve for any flow rate.

$$\Delta p_{valve} = \Delta p_{total} - \Delta p_{pipe} = 100 \ psi - \left(\frac{q}{30 \ gpm}\right)^2 (48.4 \ psi)$$

We can generalize the rest of the solution procedure for this problem as follows:

1. Pick a value for q (less than q_{max}, 30 gal/min in this case).

2. Find $\Delta p_{valve} = 100 \ psi - \left(\dfrac{q}{30 \ gpm}\right)^2 (48.4 \ psi)$

3. Find

$$q_{max} = C_v f(x)\sqrt{\frac{\Delta p_{valve}}{sg}} = (4.18)(1.0)\sqrt{\frac{\Delta p_{valve}}{1}} = 4.18\sqrt{100 \ psi - \left(\frac{q}{30 \ gpm}\right)^2 (48.4 \ psi)}.$$

4. Find $f(x) = \dfrac{q}{q_{max}} = \dfrac{q}{4.18\sqrt{100 - (q \ / \ 30)^2 (48.4)}}$.

5. Find $x = f(x)$ for a linear valve.

As a numerical example,

1. Choose $q = 20$ gal/min.

2. $\Delta p_{valve} = 100 \text{ psi} - \left(\dfrac{20}{30 \text{ gpm}}\right)^2 (48.4 \text{ psi}) = 78.5 \text{ psi}$

3. $q_{max} = 4.18\sqrt{78.5} = 37 \text{ gpm}$

4. $f(x) = \dfrac{20}{37} = 0.54$

5. $x = f(x) = 0.54$ for a linear valve

For other flow rates, one can repeat this procedure to obtain values of x. The results are shown in Table 19.3 and in Fig. 19–6. The latter also shows the inherent characteristic of the linear valve for comparison with the effective characteristic of the valve when line loss is present.

Equal-percentage valve. Calculation of the effective characteristic will now be made for an equal-percentage valve having the same C_v of 4.18 as calculated for the linear valve in the first part of this example. The procedure (steps 1 to 5) that we outlined above for the linear valve applies to this case as well, except that we must modify step 5 for an equal-percentage valve. The relationship between stem position and fraction of maximum flow is given by Eq. (19.11).

$$f(x) = \alpha^{x-1} \qquad\qquad (19.11)$$

Solving Eq. (19.11) for x, we get

$$x = 1 + \dfrac{\ln f(x)}{\ln \alpha}$$

For this example, $\alpha = 33.3$, and we modify step 5 from above to determine x as follows:

5. Find $x = 1 + \dfrac{\ln f(x)}{\ln \alpha} = 1 + \dfrac{\ln f(x)}{\ln 33.3} = 1 + \dfrac{\ln f(x)}{3.51}$

For a flow rate of 20 gal/min, we found $f(x) = 0.54$ above. For the equal-percentage valve, we can now find the stem position.

$$x = 1 + \dfrac{\ln(0.54)}{3.51} = 0.82$$

For other values of flow, corresponding values of x are calculated for the equal-percentage valve, and the results are shown in Table 19.3 and Fig. 19–6.

TABLE 19.3
Effective characteristics for a linear and an equal percentage valve (Example 19.3)

Q	P_{pipe}	P_{valve}	Q_{max}	x	x	x
(gal/min)	(psi)	(psi)	(gal/min)	Linear	Equal-percentage	Linear, inherent
2	0.22	99.78	41.76	0.05	0.13	0.07
4	0.86	99.14	41.62	0.10	0.33	0.13
6	1.94	98.06	41.39	0.14	0.45	0.20
8	3.44	96.56	41.07	0.19	0.53	0.27
10	5.38	94.62	40.66	0.25	0.60	0.33
12	7.74	92.26	40.15	0.30	0.66	0.40
14	10.54	89.46	39.54	0.35	0.70	0.47
16	13.77	86.23	38.82	0.41	0.75	0.53
18	17.42	82.58	37.98	0.47	0.79	0.60
20	21.51	78.49	37.03	0.54	0.82	0.67
22	26.03	73.97	35.95	0.61	0.86	0.73
24	30.98	69.02	34.73	0.69	0.89	0.80
26	36.35	63.65	33.35	0.78	0.93	0.87
28	42.16	57.84	31.79	0.88	0.96	0.93
30	48.40	51.60	30.03	1.00	1.00	1.00
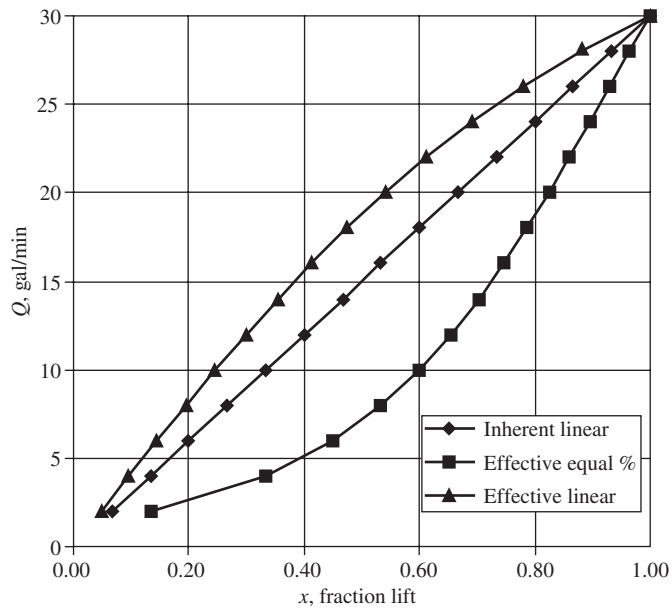
FIGURE 19–6
Comparison of effective characteristics for linear and equal-percentage
valves from Example 19.3.

BENEFIT OF AN EQUAL-PERCENTAGE VALVE. It is often stated in the control literature that the benefit derived from an equal-percentage valve arises from its inherent nonlinear characteristic that compensates for the line loss to give an effective valve characteristic that is nearly linear. A study of Fig. 19–6 shows that in this example an equal-percentage valve overcompensates for line loss and produces an effective characteristic that is not linear, but is bowed in the opposite direction to that of the effective characteristic of the linear valve. In summary, neither valve in this example produces an effective characteristic that is linear. One can show that as the line loss increases, the linear valve will depart more from the ideal linear relation and the equal-percentage valve will move more closely toward the linear relation.

In practice, a valve designated as linear will not give a linear characteristic exactly as defined in this chapter. To achieve a truly linear characteristic would require very careful design and precision machining of the valve plug and seat. The same comment can be made for an equal-percentage valve, as defined by Eq. (19.11). To know the effective characteristic of a valve, one must test it experimentally.

19.4   VALVE POSITIONER

The operation of an ideal air-to-open control valve is shown in Fig. 19–7a. Any given air pressure signal to the valve results in a unique stem position x. The friction in the packing and guiding surfaces of a control valve often causes a control valve to exhibit hysteresis, as shown in Fig. 19–7b. When the air pressure increases to the valve top, the stem position increases along the lower curve. When the air pressure decreases, the stem position decreases along the upper curve. At the moment the air pressure signal reverses, the stem position stays in the last position until the dead band H is exceeded, after which the pressure begins to decrease or increase along the paths shown by the arrows. If the valve is subjected to a slow periodic variation in pressure, a typical path taken by the stem position is shown by the closed curve *ABCDA* in Fig. 19–7b.
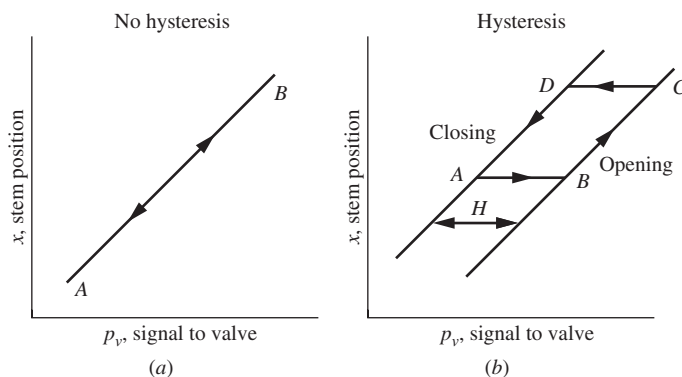
FIGURE 19–7
Control valve hysteresis.

The net result of this behavior is that the same air pressure signal to the valve top corresponds to two different stem positions, depending upon whether the valve is in the process of opening or closing.

The hysteresis described in the previous paragraph should be distinguished from the dynamic lag of a valve discussed in Chap. 9. The dynamic lag discussed in Chap. 9 is caused by the volume of space above the valve diaphragm, the resistance to flow of air to the valve top, and the inertia of the valve stem and plug; such a lag is expressed by a first-order or second-order transfer function. On the other hand, hysteresis, which is caused by the friction between the stem and the packing, is a nonlinear phenomenon and cannot be expressed by a transfer function. A valve can exhibit both dynamic lag and hysteresis.

The presence of hysteresis in the valve can cause the controlled signal to exhibit an oscillation or ripple called a *limit cycle.* Since this limit cycle is usually considered objectionable and contributes to the wear of the valve, a method is needed to eliminate it. Since the limit cycle is a nonlinear phenomenon related to the hysteresis, controller tuning is not a solution to the problem.

To reduce the deleterious effect of hysteresis and to also speed up the response of the valve, one can attach to the control valve a *positioner* that acts as a high-gain proportional controller that receives a set point signal from the primary controller and a measurement from the valve stem position. In this sense, the addition of a valve positioner introduces a form of cascade control, which was discussed in a previous chapter. A sketch of a control valve with a positioner attached is shown in Fig. 19–8. The positioner, bolted to the valve actuator, has an arm that is clamped to the valve stem to detect the stem position.

Notice that the valve positioner shown in Fig. 19–8 has the usual connections for a controller: a set point that calls for a desired stem position in the form of a signal from the primary controller p_c, a measurement in the form of stem position *x,* and a pneumatic output in the form of a pressure to the valve top p_v. Some positioners are now electronic microprocessor-based controllers, while others are still pneumatically based. Valve positioners are especially important for speeding up the valve motion and eliminating hysteresis and valve stem friction.
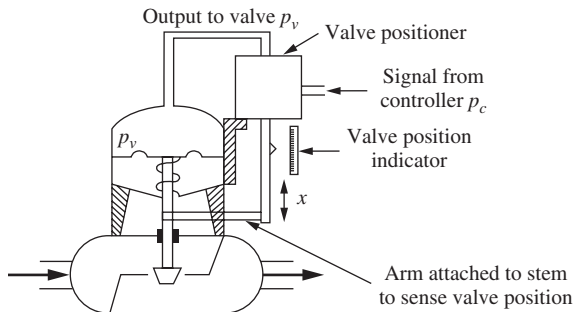
FIGURE 19–8
Control valve with positioner. (Compare with Fig. 19–1.)

SUMMARY

The control valve is a component of a control system often overlooked in a course on process control. In this chapter, the description, selection, and sizing of pneumatic control valves were presented. Valves may be of the pressure-to-close or the pressure-to-open type; the selection of the type is often related to safety considerations. If the air pressure fails, the valve should return to a position that ensures safe operating conditions for a process.

The flow capacity of a valve is based on an equation relating flow to the square root of the pressure drop across the valve; the proportionality constant C_v in this equation is a measure of the valve's capacity for flow—the larger C_v, the larger the flow.

Valves are classified according to their inherent flow characteristics such as linear or equal-percentage. A linear valve produces a flow (for constant pressure drop across the valve) that is proportional to the valve stem position, which in turn is proportional to the valve-top pressure.

The presence of a long, small-diameter line supplying a valve causes the pressure drop across the valve to decrease with the increase of flow, for a fixed overall pressure drop across the system. If the pressure drop in the line is excessive, the characteristic of the linear valve will become nonlinear and in terms of control theory, the steady-state gain K_v of the valve decreases with flow.

As a result of the change in valve gain, the controller in the loop must be readjusted for different flow rates to maintain the same degree of stability. To overcome this limitation of the linear valve, an equal-percentage (or logarithmic) valve is available for which the gain of the valve increases with flow rate. Such a valve compensates for the line loss and produces an effective characteristic that approaches a linear relation. The basis for the name *equal-percentage* (or *logarithmic*) is related to one form of the mathematical expression that describes the valve. In this form, an equal-percentage change in flow occurs for a specified change in stem position, regardless of the stem position.

To eliminate hysteresis, which can produce cycling and cause wear of the valve plug and seat, a valve positioner may be attached to a control valve. The positioner also speeds up the motion of the valve in response to a signal from the controller.

PROBLEMS

19.1. A linear control valve having a C_v of 0.1 is connected to a source of water. If the pressure drop across the valve is 400 psi and if the pneumatic pressure to the valve top is 12 psig, what is the flow rate through the valve? The valve goes from completely shut to completely open as the valve-top pressure varies from 3 to 15 psig.

19.2. (*a*) Under what conditions would an equal-percentage valve be used instead of a linear valve?

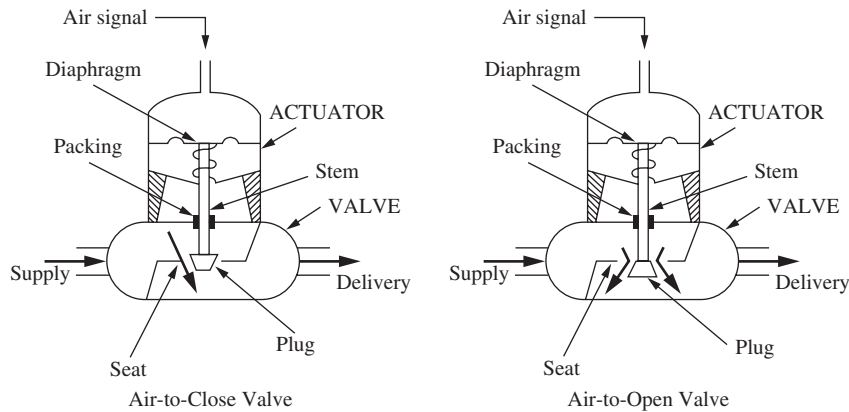(*b*) What are some reasons to use a valve positioner?

CAPSULE SUMMARY

FIGURE 19–1
Single-seated control valves.

Control valves can be air-to-open or air-to-close. The choice is usually made so that the valve fails in the safe position upon loss of air signal pressure.

$$\text{Valve sizing equation} \qquad q = C_v f(x) \sqrt{\frac{\Delta p_{\text{valve}}}{\text{sg}}} \qquad (19.1)$$

where
q = flow rate, gal/min
x = fractional stem position (i.e., fraction open)
Δp_{valve} = pressure drop across valve, psi
$f(x)$ = fraction of maximum flow (= 1 for fully open)
C_v = factor associated with capacity of valve
sg = specific gravity of fluid (water sg = 1)

Equation (19.1) applies to the flow of an incompressible, nonflashing fluid through a fully open valve. Factor C_v can be defined as the flow (gal/min) of a fluid of unit specific gravity (water) through a fully open valve, across which a pressure drop of 1.0 psi exists. Therefore, the bigger C_v, the larger the valve.

Valve type	Sensitivity $\dfrac{df}{dx}$	Relationship
Linear (I)	Constant	$f(x) = x$
Equal-percentage (II)	Increasing	$f(x) = \alpha^{x-1}$
Square root or quick-opening (III)	Decreasing	$f(x) = \sqrt{x}$

A **valve positioner** is a device that can be attached to a valve that drives the valve to the desired position, in spite of friction or hysteresis.
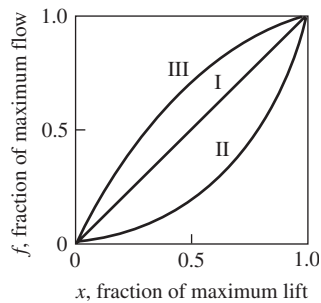
FIGURE 19–3
Inherent valve characteristics (pressure drop across valve is constant).
I: Linear, II: increasing sensitivity (e.g., equal-percentage valve), and III: decreasing sensitivity (e.g., square root valve).

THEORETICAL ANALYSIS OF COMPLEX PROCESSES

To investigate theoretically the control of a process, it is necessary first to know the dynamic character of the process that is being controlled. In previous chapters, the processes have been very simple for the purpose of illustrating control theory. Many physical processes are extremely complicated, and it requires considerable effort to construct a mathematical model that will adequately simulate the dynamics of the actual system. In this chapter, we analyze several complex systems to indicate some of the types of problems that can be encountered. In these examples, the technique of linearization, first presented in Chap. 5, is applied to a function of several variables. One example leads to a multiloop control system. In Sec. 20.3 distributed-parameter systems are discussed.

20.1 CONTROL OF A STEAM-JACKETED KETTLE

The dynamic response and control of the steam-jacketed kettle shown in Fig. 20–1 are to be considered. The system consists of a kettle through which water flows at a variable rate w lb/time. The entering water is at temperature T_i, which may vary with time. The kettle water, which is well agitated, is heated by steam condensing in the jacket at temperature T_v and pressure p_v. The temperature of the water in the kettle is measured and transmitted to the controller. The output signal from the controller is used to change the stem position of the valve, which adjusts the flow of steam to the jacket. The major problem in this example is to determine the dynamic characteristics of the kettle. The kettle is actually a nonlinear system, and to obtain a linear model, a number of simplifying assumptions are needed.
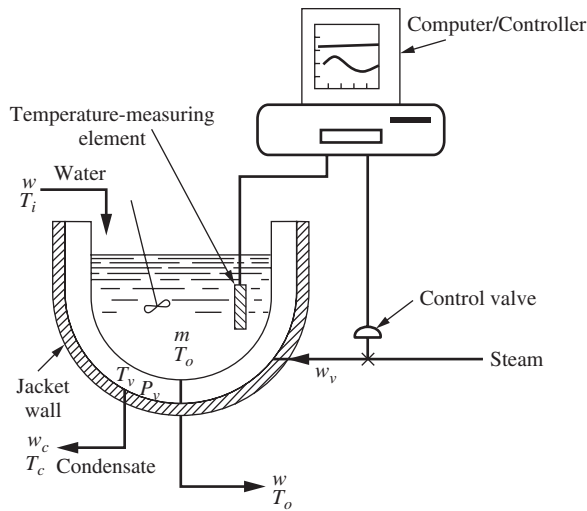
FIGURE 20–1
Control of a steam-jacketed kettle.

Analysis of Kettle

The following assumptions are made for the kettle:

1. The heat loss to the atmosphere is negligible.
2. The holdup volume of water in the kettle is constant.
3. The thermal capacity of the kettle wall, which separates steam from water, is negligible compared with that of the water in the kettle.
4. The thermal capacity of the outer jacket wall, adjacent to the surroundings, is finite, and the temperature of this jacket wall is uniform and equal to the steam temperature at any instant.
5. The kettle water is sufficiently agitated to result in a uniform temperature.
6. The flow of heat from the steam to the water in the kettle is described by the expression

$$q = U(T_v - T_o)$$

where q = flow rate of heat, Btu/(h·ft^2)
U = overall heat-transfer coefficient, Btu/(h·ft^2·°F)
T_v = steam temperature, °F
T_o = water temperature, °F

The overall heat-transfer coefficient U is constant.

7. The heat capacities of water and the metal wall are constant.
8. The density of water is constant.
9. The steam in the jacket is saturated.

The assumptions listed here are more or less arbitrary. For a specific kettle operating under a particular set of conditions, some of these assumptions may require modification.

The approach to this problem is to make an energy balance on the water side and another energy balance on the steam side. To aid the development of the transfer functions, a schematic diagram of the kettle is shown in Fig. 20–2. The symbols used throughout this analysis are defined as follows:

T_i = temperature of inlet water, °F
T_o = temperature of outlet water, °F
T_v = temperature of jacket steam, °F
T_c = temperature of condensate, °F
w = flow rate of inlet water, lb/time
w_v = flow rate of steam, lb/time
w_c = flow rate of condensate from kettle, lb/time
m = mass of water in kettle, lb
m_1 = mass of jacket wall, lb
V = volume of jacket steam space, ft^3
C = heat capacity of water, Btu/(lb·°F)
C_1 = heat capacity of metal in jacket wall, Btu/(lb·°F)
A = cross-sectional area for heat exchange, ft^2
t = time
H_v = specific enthalpy of steam entering, Btu/lb
H_c = specific enthalpy of condensate leaving, Btu/lb
U_v = specific internal energy of steam in jacket, Btu/lb
ρ_v = density of steam in jacket, lb/ft^3
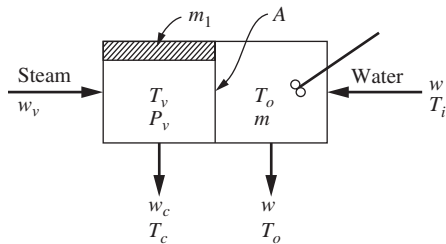
FIGURE 20–2
Schematic diagram of kettle.

An energy balance on the water side gives

$$wC(T_i - T_o) + UA(T_v - T_o) = mC\frac{dT_o}{dt} \qquad (20.1)$$

In Eq. (20.1), the terms $C, U, A,$ and m are constants. The first term in Eq. (20.1) is nonlinear, since it contains the product of flow rate and temperature, that is, wT_i and wT_o.

To obtain a transfer function from Eq. (20.1), these nonlinear terms must be linearized. Before continuing the analysis, we digress briefly to discuss the general problem of linearization of a function of several variables.

Recall from Chap. 5 [Eq. (5.39)] that a function of two variables $z(x, y)$ can be expanded around an operating point x_s, y_s by means of a Taylor series expansion:

$$z = z(x_s, y_s) + \left.\frac{\partial z}{\partial x}\right|_{x_s, y_s} (x - x_s) + \left.\frac{\partial z}{\partial y}\right|_{x_s, y_s} (y - y_s)$$
$$+ \text{higher-order terms in } x - x_s \text{ and } y - y_s \tag{20.2}$$

The subscript s stands for steady state.

In control problems, the operating point (x_s, y_s), around which the expansion is to be made, is selected at steady-state values of the variables before any disturbance occurs. Linearization of the function z consists of retaining only the linear terms, on the basis that the deviations (e.g., $x - x_s$) will be small. Thus,

$$z \cong z_s + z_{x_s}(x - x_s) + z_{y_s}(y - y_s) \tag{20.3}$$

where z_{x_s} and z_{y_s} are the partial derivatives in Eq. (20.2). If z is a function of three or more variables, the linearized form is the same as that of Eq. (20.3) with an additional term for each variable.

The linearization expressed by Eq. (20.3) may be applied to the terms wT_i and wT_o in Eq. (20.1) to obtain

$$wT_i = w_s T_{i_s} + w_s(T_i - T_{i_s}) + T_{i_s}(w - w_s) \tag{20.4}$$

and

$$wT_o = w_s T_{o_s} + w_s(T_o - T_{o_s}) + T_{o_s}(w - w_s) \tag{20.5}$$

Notice that for these cases the nonlinear terms are wT_i and wT_o. The first partial derivatives, evaluated at the operating point, are

$$\left.\frac{\partial(wT_{i_s})}{\partial w}\right|_{w_s, T_{i_s}} = T_{i_s}$$

$$\left.\frac{\partial(wT_{i_s})}{\partial w}\right|_{w_s, T_i} = w_s$$

and so on.

Introducing Eqs. (20.4) and (20.5) into Eq. (20.1) gives the following linearized equation:

$$\left[(T_{i_s} - T_{o_s})(w - w_s) + w_s(T_i - T_o)\right]C + UA(T_v - T_o) = mC\frac{dT_o}{dt} \tag{20.6}$$

At steady state, $dT_o/dt = 0$, and Eq. (20.1) can be written

$$w_s C (T_{i_s} - T_{o_s}) + UA(T_{v_s} - T_{o_s}) = 0 \tag{20.7}$$

Subtracting Eq. (20.7) from Eq. (20.6), introducing the deviation variables

$$T_i' = T_i - T_{i_s}$$
$$T_o' = T_o - T_{o_s}$$
$$T_v' = T_v - T_{v_s}$$
$$W' = w - w_s$$

and rearranging give the result

$$C\left[(T_{i_s} - T_{o_s})W + w_s \left(T_i' - T_o'\right)\right] + UA\left(T_v' - T_o'\right) = mC\frac{dT_o'}{dt} \tag{20.8}$$

Taking the transform of Eq. (20.8) and solving for $T_o'(s)$ give

$$T_o'(s) = \frac{K_1}{\tau_w s + 1} T_i'(s) + \frac{K_2}{\tau_w s + 1} T_v'(s) - \frac{K_3}{\tau_w s + 1} W'(s) \tag{20.9}$$

where

$$K_1 = \frac{w_s C}{UA + w_s C}$$

$$K_2 = \frac{UA}{UA + w_s C}$$

$$K_3 = \frac{C(T_{o_s} - T_{i_s})}{UA + w_s C}$$

$$\tau_w = \frac{mC}{UA + w_s C}$$

From Eq. (20.9), we see that the response of T_o' to T_i', T_v', or W' is first-order with a time constant τ_w. The steady-state gains (Ks) in Eq. (20.9) are all positive.

The following energy balance can be written for the steam side of the kettle:

$$\underbrace{w_v H_v}_{\text{in by flow}} - \underbrace{w_c H_c}_{\text{out by flow}} = \underbrace{UA(T_v - T_o)}_{\text{out by convection}} + \underbrace{\frac{Vd\left(\rho_v U_v\right)}{dt}}_{\text{accumulation in steam}}$$
$$+ \underbrace{m_1 C_1 \frac{dT_v}{dt}}_{\text{accumulation in kettle wall}} \tag{20.10}$$

Notice that we have made use of assumption 4 in writing the last term of Eq. (20.10), which implies that the metal in the outer jacket wall is always at the steam temperature.

A mass balance on the steam side of the kettle yields

$$w_v - w_c = V \frac{d\rho_v}{dt} \tag{20.11}$$

Combining Eqs. (20.10) and (20.11) to eliminate w_c gives

$$
\begin{aligned}
w_v (H_v - H_c) = (U_v - H_c)V \frac{d\rho_v}{dt} + m_1 C_1 \frac{dT_v}{dt} + UA(T_v - T_o) \\
+ V\rho_v \frac{dU_v}{dt}
\end{aligned}
\tag{20.12}
$$

The variables ρ_v, U_v, H_v, and H_c are functions of the steam and condensate temperatures and can be approximated by expansion in Taylor series and linearization as follows:

$$
\begin{aligned}
\rho_v &= \rho_{v_s} + \alpha (T_v - T_{v_s}) \\
U_v &= U_{v_s} + \phi (T_v - T_{v_s}) \\
H_v &= H_{v_s} + \gamma (T_v - T_{v_s}) \\
H_c &= H_{c_s} + \sigma (T_c - T_{c_s})
\end{aligned}
\tag{20.13}
$$

where

$$\alpha = \left. \frac{d\rho_v}{dT_v} \right|_s$$

$$\phi = \left. \frac{dU_v}{dT_v} \right|_s$$

$$\gamma = \left. \frac{dH_v}{dT_v} \right|_s$$

$$\sigma = \left. \frac{dH_c}{dT_c} \right|_s$$

The parameters α, ϕ, γ, and σ in these relationships can be obtained from the steam tables once the operating point is selected. For example, if the operating point is at 212°F and the deviation in steam temperature is 10°F, we obtain the following estimate of γ from the steam tables:

$$T_{v_s} = 212°\text{F}$$

$$H_{v_s} = 1150.4 \text{ Btu /lb}$$

$$H_v = 1154.1 \quad \text{at } T_v = 222°\text{F}$$

$$H_v = 1146.6 \quad \text{at } T_v = 202°\text{F}$$

$$\gamma \approx \frac{1154.1 - 1146.6}{222 - 202} = 0.375$$

$$H_v = 1150.4 + 0.375(T_v - 212)$$

In a similar manner, the properties of saturated steam can be used to evaluate α, ϕ, and σ.

Introducing the relationships of Eq. (20.13) into Eq. (20.12) and assuming the condensate temperature T_c to be the same as the steam temperature T_v give the following result:

$$\left[H_{v_s} - H_{c_s} + (\gamma - \sigma)(T_v - T_{v_s}) \right] w_v$$
$$= \left[(U_{v_s} - H_{c_s}) + (2\phi - \sigma)(T_v - T_{v_s}) + \frac{\phi}{\alpha} \rho_{v_s} + \frac{m_1 C_1}{\alpha V} \right] \qquad (20.14)$$
$$\alpha V \frac{dT_v}{dt} UA(T_v - T_o)$$

Some of the terms in Eq. (20.14) can be neglected. The term

$$(\gamma - \sigma)(T_v - T_{v_s})$$

can be dropped because it is negligible compared with $H_{v_s} - H_{c_s}$. For example, for steam at atmospheric pressure, a change of 10°F gives a value of $(\gamma - \sigma)(T_v - T_{v_s})$ of about 7 Btu/lb while $H_{v_s} - H_{c_s}$ is 970 Btu/lb. Similarly, the term $(2\phi - \sigma)(T_v - T_{v_s})$ can be neglected. For example, this term is about -4 Btu/lb for a change in steam temperature of 10°F for steam at about 1 atm pressure; the term $U_{v_s} - H_{c_s}$ is 897 Btu/lb under these conditions. Also, the term $\phi \rho_{v_s} / \alpha$ is about 15 Btu/lb and can be neglected. Discarding these terms, writing the remaining terms in deviation variables, and transforming yield

$$T_v'(s) = \frac{1}{\tau_v s + 1} T_o'(s) + \frac{K_5}{\tau_v s + 1} W_v'(s) \qquad (20.15)$$

where $T_v' = T_v - T_{vs}$

$W_v' = W_v - W_{v_s}$

$K_5 = \dfrac{H_{v_s} - H_{c_s}}{UA}$

$\tau_v = \dfrac{(U_{v_s} - H_{c_s})\alpha V + m_1 C_1}{UA}$

From Eq. (20.15), we see that the steam temperature T_v' depends on the steam flow rate W_v' and the water temperature T_o'. The combination of Eqs. (20.9) and (20.15) gives the dynamic response of the water temperature to changes in water flow rate, inlet water temperature, and steam flow rate. These equations are represented by a portion of the block diagram of Fig. 20–4. Before completing the analysis of the control system, we must consider the effect of valve stem position on the steam flow rate.

Analysis of Valve

The flow of steam through the valve depends on three variables: steam supply pressure, steam pressure in the jacket, and the valve stem position, which we assume to be proportional to the pneumatic valve-top pressure p. For simplicity, assume the steam supply pressure to be constant with the result that the steam flow rate is a function of only the two remaining variables; thus

$$w_v = f(p, p_v) \tag{20.16}$$

Because of the assumption that the steam in the jacket is always saturated, we know that p_v is a function of T_v; thus

$$p_v = g(T_v) \tag{20.17}$$

This functional relation can be obtained from the saturated steam tables. Equations (20.16) and (20.17) can be combined to give

$$w_v = f[p, g(T_v)] = f_1(p, T_v)$$

The function $f_1(p, T_v)$ is in general nonlinear, and if an analytic expression is available, the function can be linearized as described previously. The flow of steam through a control valve can often be represented by the relationship

$$w_v = A_0 C_v \sqrt{p_s - p_v} \tag{20.18}$$

where p_s = supply pressure of steam
$\quad\quad\quad p_v$ = pressure downstream of valve
$\quad\quad\quad A_0$ = cross-sectional area for flow of steam through valve
$\quad\quad\quad C_v$ = constant of valve

For a linear valve, A_0 is proportional to stem position, and the stem position is proportional to the valve-top pressure p; under these conditions, Eq. (20.18) takes the form

$$w_v = C'_v p \sqrt{p_s - p_v} \tag{20.19}$$

For this example, however, we assume that an analytic expression is not available. The linearized form of $f_1(p, T_v)$ can be obtained by making some experimental tests on the valve. If the valve-top pressure is fixed at its steady-state (or average) value and w_v is measured for several values of T_v (or p_v), a curve such as the one shown in Fig. 20–3a can be obtained. If the steam temperature T_v (or p_v) is held constant and the flow rate is measured at several values of valve-top pressure, a curve such as that shown in Fig. 20–3b can be obtained. These two curves can now be used to evaluate the partial derivatives in the linear expansion of $f_1(p, T_v)$ as we now demonstrate.

Expanding w_v about the operating point p_s, T_{v_s} and retaining only the linear terms give

$$w_v = w_{v_s} + \left.\frac{\partial w_v}{\partial p}\right|_{p_s,T_{v_s}} (p - p_s) + \left.\frac{\partial w_v}{\partial T_v}\right|_{p_s,T_{v_s}} (T_v - T_{v_s})$$

This equation can be written in the form

$$W_v = K_v P - \frac{1}{R_v} T_v' \qquad (20.20)$$

where $W_v' = w_v - w_{v_s}$

$\qquad P = p - p_s$

$\qquad T_v' = T_v - T_{v_s}$

$\qquad K_v = \left.\dfrac{\partial w_v}{\partial p}\right|_{p_s,T_{v_s}}$

$\qquad \dfrac{1}{R_v} = -\left.\dfrac{\partial w_v}{\partial T_v}\right|_{p_s,T_{v_s}}$

The coefficients K_v and $-1/R_v$ in Eq. (20.20) are the slopes of the curves of Fig. 20–3 at the operating point p_s, T_{v_s}. This follows from the definition of a partial derivative. Notice that $1/R_v$ has been defined as the negative of the slope so that R_v is a positive quantity. The experimental approach described here for obtaining a linear form for the flow characteristics of a valve is always possible in principle. However, it must be emphasized that the linear form is useful only for small deviations from the operating point. If the operating point is changed considerably, the coefficients K_v and $1/R_v$ must be reevaluated. Notice that, in writing Eq. (20.20), we have assumed the valve to have no dynamic lag between p and stem position. This assumption is valid for a system having large time constants, such as a steam-jacketed kettle, as was demonstrated in Chap. 9.
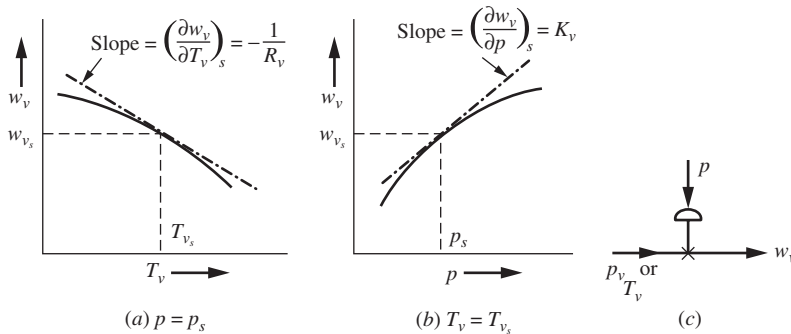
FIGURE 20–3
Linearization of valve characteristics from experimental tests.

Block Diagram of Control System

We have now completed the analysis of the kettle and valve. A block diagram of the control system, based on Eqs. (20.9), (20.15), and (20.20), is shown in Fig. 20–4.
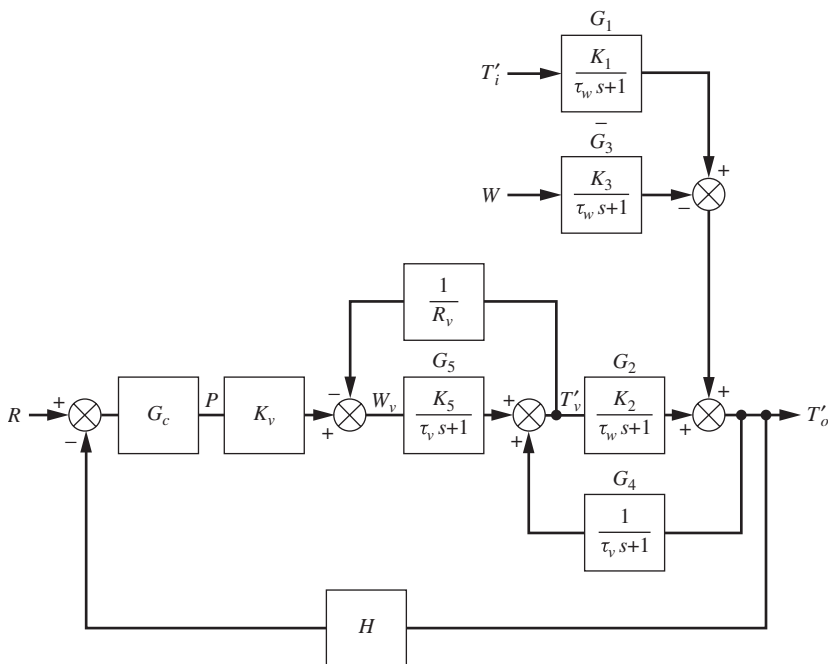
FIGURE 20–4
Block diagram for control of steam-jacketed kettle.

The controller action is not specified but merely denoted by G_c in the block diagram. Also, the feedback element is denoted as H. From Fig. 20–4, we see that the steam-jacketed kettle is a multiloop control system. Furthermore, the loops overlap. The block diagram can be used to obtain the overall transfer function between any two variables by applying the methods of Chap. 12. After considerable algebraic manipulation, the following result is obtained:

$$T'_o = \frac{G_c G_2 G_5 K_v}{D(s)} R + \frac{G_1 \left(1 + G_5 / R_v\right)}{D(s)} T'_i - \frac{G_3 \left(1 + G_5 / R_v\right)}{D(s)} W \quad (20.21)$$

where $D(s) = 1 + G_5/R_v + G_c G_2 G_5 K_v H - G_2 G_4$. The terms $G_1, G_2, G_3, G_4, G_5, G_c,$ and H are defined in Fig. 20–4. For example, if $G_c = K_c$ and $H = 1$, one obtains from Eq. (20.21) the transfer function

$$\frac{T'_o}{R} = \frac{K}{\tau^2 s^2 + 2\zeta \tau s + 1} \quad (20.22)$$

where
$$K = \frac{K_c K_v K_2 K_5}{D_1}$$

$$\tau^2 = \frac{\tau_v \tau_w}{D_1}$$

$$2\zeta\tau = \frac{\tau_v + \tau_w + K_5 \tau_w / R_v}{D_1}$$

$$D_1 = 1 + \frac{K_5}{R_v} + K_c K_v K_2 K_5 - K_2$$

It is seen that the response of the control system is second-order when proportional control is used and the measuring element does not have dynamic lag. Notice that the parameters K, τ^2, and $2\zeta\tau$ in Eq. (20.22) are positive. This follows from the fact that the parameters K_c, K_v, K_2, K_5, R_v, τ_v, and τ_w are all positive and that $K_2 < 1$. When a block diagram of a control system becomes very complicated, such as the one in this example, it is convenient to simulate the control system with a software package such as Simulink.

20.2   DYNAMIC RESPONSE OF A GAS ABSORBER

Another example of a complex system is the plate absorber shown in Fig. 20–5. The reader who has not studied gas absorption may find this subject presented in any textbook on chemical engineering unit operations; for example, see McCabe, Smith, and Harriott (2004).
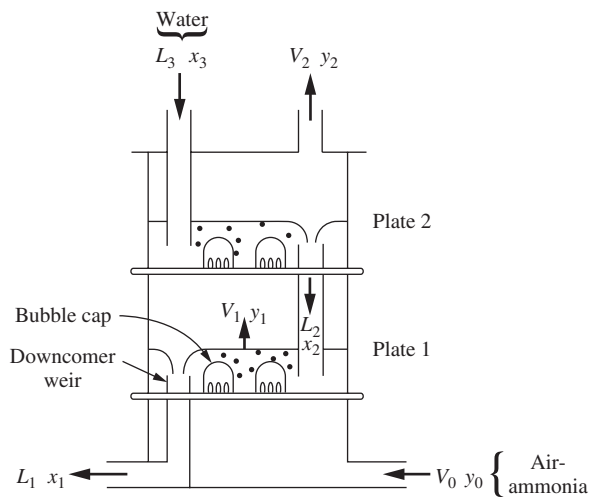
FIGURE 20–5
Bubble-cap absorber.

In this process, air containing a soluble gas such as ammonia is contacted with fresh water in a two-plate column in order to remove part of the ammonia from the gas. The action of gas bubbling through the liquid causes thorough mixing of the two phases on each plate. During the mixing process, ammonia diffuses from the bubbles into the liquid. In an industrial operation, many plates may be used; however, for simplicity, we consider only two plates in this example, since the basic principles are unaffected by the number of plates.

Our problem is to analyze the system for its dynamic response. In other words, we want to know how the concentrations of liquid and gas change as a result of change in inlet composition of the gas stream to be treated. We will consider the inlet flow of the liquid phase to be constant.

Throughout the analysis, the following symbols are used:

L_n = flow of liquid leaving nth plate, mol/min
V_n = flow of gas leaving nth plate, mol/min
x_n = concentration of liquid leaving nth plate, mole fraction NH_3
y_n = concentration of gas leaving nth plate, mole fraction NH_3
H_n = holdup (or storage) of liquid on nth plate, mole

To avoid too many complicating details, the following assumptions will be used:

1. The temperature and total pressure throughout the column are uniform and do not vary with changes in flow rate.
2. The entering gas stream is dilute (say 5 mol % NH_3) with the consequence that we can neglect the decrease in total molar flow rate of gas as ammonia is removed. Likewise, we can assume that the molar flow rate of liquid does not increase as ammonia is added.
3. The plate efficiency is 100 percent, which means that the vapor and liquid streams leaving a plate are in equilibrium. Such a plate is called an *ideal* equilibrium stage.
4. The equilibrium relationship is linear and is given by the expression

$$y_n = mx_n^* + b \qquad (20.23)$$

where m and b are constants that depend on the temperature and total pressure of the system and x_n^* is the concentration of liquid in equilibrium with gas of concentration y_n. For an ideal plate

$$x_n = x_n^*$$

(If the efficiency of the plate is not 100 percent, we can introduce an individual tray efficiency of the Murphree type, defined as

$$E_n = \frac{x_n - x_{n+1}}{x_n^* - x_{n+1}}$$

where x_n^* is the concentration of the liquid in equilibrium with gas of composition y_n. Notice that for an ideal plate $E_n = 1$ and $x_n = x_n^*$. In general, the efficiency

of a plate depends on the design of the plate, the properties of the gas and liquid streams, and the flow rates. We could include efficiency in our mathematical model; however, to do so would greatly increase the complexity of the problem. To account properly for the variation in efficiency with flow rates would require empirical relationships for a specific plate design.

In this list of assumptions, the one which is most likely to be invalid for a practical process is that the plate is an ideal equilibrium stage.)

5. The holdup of liquid H_n on each plate is constant and independent of flow rate. Furthermore, the holdup is the same for each plate, that is, $H_1 = H_2 = H$.

6. The inlet liquid flow is constant. This assumption, in addition to assumptions 5 and 2, leads us to the conclusion that the liquid flows throughout the column are constant and equal, that is, $L_3 = L_2 = L_1 = L$.

7. The holdup of gas between plates is negligible. As a consequence of this assumption and assumption 2, the flow rate of gas from each plate is the same and equal to the entering gas flow rate; that is,

$$V_0 \ = \ V_1 \ = \ V_2 \ = \ V$$

Analysis

We begin the analysis of this process by writing an ammonia balance around each plate. A mass balance on ammonia around plate 1 gives

$$H \frac{dx_1}{dt} \ = \ Lx_2 \ + \ Vy_0 \ - \ Lx_1 \ - \ Vy_1 \tag{20.24}$$

This last equation states that the accumulation of NH_3 on plate 1 is equal to the flow of NH_3 into the plate minus the flow of NH_3 out of the plate. Notice that V, L, and H do not have subscripts because of assumptions 5 through 7.

A mass balance on ammonia around plate 2 gives

$$H \frac{dx_2}{dt} \ = \ Vy_1 \ - \ Lx_2 \ - \ Vy_2 \tag{20.25}$$

The last equation does not contain a term Lx_3, since we have assumed that $x_3 = 0$ (pure water).

For an ideal plate $x_n = x_n^*$, and the equilibrium relation of Eq. (20.23) becomes

$$y_n \ = \ mx_n \ + \ b$$

Substituting the equilibrium relationship into Eqs. (20.24) and (20.25) gives

$$H \frac{dx_1}{dt} \ = \ Lx_2 \ - \ Lx_1 \ + \ Vm(x_0 \ - \ x_1)$$

and
$$H \frac{dx_2}{dt} \ = \ Vm(x_1 \ - \ x_2) \ - \ Lx_2$$

where $x_0 = (y_0 - b)/m$ is the composition of liquid that would be in equilibrium with the entering gas of composition y_0. Solving these last two equations for the derivatives gives

$$\frac{dx_1}{dt} = \frac{L}{H}(x_2 - x_1) + \frac{Vm}{H}(x_0 - x_1) \tag{20.26}$$

$$\frac{dx_2}{dt} = \frac{Vm}{H}(x_1 - x_2) - \frac{L}{H}x_2 \tag{20.27}$$

The analysis has resulted in two first-order differential equations. The forcing functions in this process, which must be specified as functions of t, are the inlet gas concentration $x_0 = (y_0 - b)/m$ and the inlet liquid flow rate $L_3 = L$. We can now solve for $x_1(t)$ and $x_2(t)$. Equations (20.26) and (20.27) can be written

$$\frac{dx_1}{dt} = -ax_1 + bx_2 + cx_0 \tag{20.28}$$

$$\frac{dx_2}{dt} = cx_1 - ax_2 \tag{20.29}$$

where

$$a = \frac{L}{H} + \frac{Vm}{H}$$

$$b = \frac{L}{H}$$

$$c = \frac{Vm}{H}$$

At steady state, $dx_1/dt = dx_2/dt = 0$, and Eqs. (20.30) and (20.31) can be written

$$0 = -ax_{1_s} + bx_{2_s} + cx_{0_s} \tag{20.30}$$

$$0 = cx_{1_s} - ax_{2_s} \tag{20.31}$$

Subtracting these steady-state equations from Eqs. (20.28) and (20.29) and introducing the deviation variables $X_1 = x_1 - x_{1_s}$, $X_2 = x_2 - x_{2_s}$, and $X_0 = x_0 - x_{0_s}$ give

$$\frac{dX_1}{dt} = -aX_1 + bX_2 + cX_0 \tag{20.32}$$

$$\frac{dX_2}{dt} = cX_1 - aX_2 \tag{20.33}$$

Notice that $X_0 = Y_0/m$ because

$$X_0 = x_0 - x_{0_s}$$

$$X_0 = \frac{y_0 - b}{m} - \frac{y_{0_s} - b}{m} = \frac{y_0 - y_{0_s}}{m} = \frac{Y_0}{m}$$

Equations (20.32) and (20.33) can be transformed to give

$$sX_1 = -aX_1 + bX_2 + cX_0$$

$$sX_2 = cX_1 - aX_2$$

We now have two algebraic equations and three unknowns (X_1, X_2, and X_0). Solving this pair of equations to eliminate X_1 and replacing X_2 by Y_2/m and X_0 by Y_0/m give the transfer function

$$\frac{Y_2(s)}{Y_0(s)} = \frac{c^2 / (a^2 - bc)}{\left[1 / (a^2 - bc)\right]s^2 + \left[2a / (a^2 - bc)\right]s + 1} \tag{20.34}$$

This result shows that the response of outlet gas concentration to a change in inlet gas concentration is second-order. Equation (20.34) is of the standard second-order form $K/(\tau^2 s^2 + 2\zeta\tau s + 1)$ with the parameters

$$\tau^2 = \frac{1}{a^2 - bc} \qquad \text{and} \qquad 2\zeta\tau = \frac{2a}{a^2 - bc}$$

Solving these two equations to eliminate τ gives

$$\zeta = \frac{1}{\sqrt{1 - bc / a^2}}$$

Writing a and b in terms of the original system parameters (L, H, V, m) gives

$$\zeta = \left[1 - \frac{(L / H)(Vm / H)}{(L / H + Vm / H)^2}\right]^{-1/2}$$

Simplifying this expression gives

$$\zeta = \left[1 - \frac{Vm / L}{(1 + Vm / L)^2}\right]^{-1/2}$$

Since $Vm/L > 0$, we see that $\zeta > 1$, meaning that the response is overdamped. If the analysis is repeated for a gas absorber containing n plates, it will be found that the response between inlet gas concentration and outlet gas concentration is nth-order.

20.3   DISTRIBUTED-PARAMETER SYSTEMS

Heat Conduction into a Solid

In Chap. 4, the analysis of the mercury thermometer was based on a "lumped para-meter" model. At that time, reference was made to a distributed-parameter model of the thermometer. To illustrate the difference between a lumped-parameter system and a distributed-parameter system, consider a slab of solid conducting material of infinite thickness, as shown in Fig. 20–6. Let the input to this system be the temperature at the left face ($x = 0$), which is some arbitrary function of time. The output will be the temperature at the position $x = L$. For convenience, we may consider this system to rep-resent the response of a bare thermocouple embedded in a thick wall, as the surface of the wall experiences a variation in temperature. The conductivity k, heat capacity C, and density ρ of the conducting material are constant, independent of temperature. Initially ($t < 0$), the slab is at a uniform steady-state temperature. Therefore in deviation vari-ables, which will be used henceforth, the initial temperature is zero. The cross-sectional area of the slab is A.
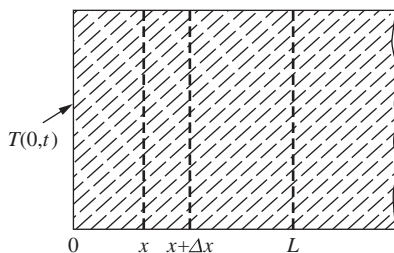
FIGURE 20–6
Heat conduction in a solid.

ANALYSIS. In this problem the temperature in the slab is a function of position and time and is indicated by $T(x,t)$. The temperature at the surface is indicated by $T(0,t)$, and that at $x = L$ by $T(L,t)$. To derive a differential equation that describes the heat conduc-tion in the slab, we first write an energy balance over a differential length Δx of the slab. This energy balance can be written

$$\left\{\begin{array}{c}\text{Flow of heat}\\\text{into left face}\\\text{by conduction}\end{array}\right\} - \left\{\begin{array}{c}\text{Flow of heat out}\\\text{of right face}\\\text{by conduction}\end{array}\right\} = \left\{\begin{array}{c}\text{Rate of accumulation}\\\text{of internal energy in}\\\text{volume element}\end{array}\right\} \qquad (20.35)$$

The flow of heat by conduction follows Fourier's law:

$$q = -k\frac{\partial T}{\partial x} \qquad (20.36)$$

where
$$q = \text{heat flux by conduction}$$
$$\partial T/\partial x = \text{temperature gradient}$$
$$k = \text{thermal conductivity}$$

Applying Eq. (20.36) to Eq. (20.35) gives

$$-Ak\frac{\partial T}{\partial x}\bigg|_x - \left(-Ak\frac{\partial T}{\partial x}\bigg|_{x+\Delta x}\right) = \frac{\partial}{\partial t}\left[C\rho A\,\Delta x\,(T-T_r)\right] \tag{20.37}$$

where T_r is the reference temperature used to evaluate internal energy. If we divide both sides by the volume of the element, and rearrange, we obtain

$$\frac{Ak\dfrac{\partial T}{\partial x}\bigg|_{x+\Delta x} - Ak\dfrac{\partial T}{\partial x}\bigg|_x}{A\,\Delta x} = \frac{k\dfrac{\partial T}{\partial x}\bigg|_{x+\Delta x} - k\dfrac{\partial T}{\partial x}\bigg|_x}{\Delta x} = C\rho\frac{\partial}{\partial t}\left[(T-T_r)\right] \tag{20.38}$$

Now, if we let the volume of the element shrink to zero, we obtain the fundamental equation describing conduction in a solid

$$k\frac{\partial^2 T}{\partial x^2} = \rho C\frac{\partial T}{\partial t}$$

This is often written as

$$\alpha\frac{\partial^2 T}{\partial x^2} = \frac{\partial T}{\partial t} \tag{20.39}$$

where $\alpha = k/(\rho C)$ is defined as the thermal diffusivity.

Several points are worth mentioning at this time. In this analysis, we have allowed the capacity for storing heat (ρCA per unit length of x) and the resistance to heat conduction [$1/(kA)$ per unit length of x] to be "spread out" or distributed uniformly throughout the medium. This distribution of capacitance and resistance is the basis for the term *distributed parameter*. The analysis has also led to a partial differential equation, which in general is more difficult to solve than the ordinary differential equation that results from a lumped-parameter model.

TRANSFER FUNCTION. We are now in a position to derive a transfer function from Eq. (20.39). First notice that since T is a function of both time t and position x, a transfer function may be written for an arbitrary value of x. In this problem, the temperature is to be observed at $x = L$; hence the transfer function will relate $T(L,t)$ to the temperature at the left surface $T(0,t)$ which is taken as the forcing function.

Equation (20.39) will be solved by the method of Laplace transforms. Taking the Laplace transform of both sides of Eq. (20.39) with respect to t gives

$$\alpha\int_0^\infty \frac{\partial^2 T}{\partial x^2}(x,t)e^{-st}\,dt = \int_0^\infty \frac{\partial T}{\partial t}(x,t)e^{-st}\,dt \tag{20.40}$$

Consider first the integral on the left side of Eq. (20.40). Interchanging the order of integration and differentiation results in

$$\int_0^\infty \frac{\partial^2 T}{\partial x^2}(x,t)e^{-st}\,dt = \frac{\partial^2}{\partial x^2}\int_0^\infty T(x,t)e^{-st}\,dt = \frac{d^2\overline{T}(x,s)}{dx^2} \tag{20.41}$$

where $\overline{T}(x,s)$ is the Laplace transform of $T(x,t)$. This interchange is allowed for most functions of engineering interest (Churchill,1972). In this chapter the overbar is often used to indicate the Laplace transform of a function of two variables. Note that the presence of x has no effect on the second integral of Eq. (20.41) because the integration is with respect to t. Also note that the derivative on the right side of Eq. (20.41) is taken as an ordinary derivative because $T(x,s)$ will later be seen to be a function of only one independent variable x and a parameter s. Next consider the integral on the right side of Eq. (20.40). Again, the presence of x has no effect on the integration with respect to t, and the rule for the transform of a derivative may be applied directly to yield

$$\int_0^\infty \frac{\partial T}{\partial t}(x,t)e^{-st}\,dt = s\overline{T}(x,s) - T(x,0) \tag{20.42}$$

where $T(x,0)$ is the initial temperature distribution in the solid. Introducing the results of the transformation into Eq. (20.40) gives

$$\alpha\frac{d^2\overline{T}(x,s)}{dx^2} = s\overline{T}(x,s) - T(x,0) \tag{20.43}$$

The partial differential equation has now been reduced to an ordinary differential equation, which can usually be solved without difficulty. It should be clear that s in Eq. (20.43) is merely a parameter, with the result that this equation is an *ordinary* second-order differential equation in the independent variable x. This follows because there are no derivatives with respect to s in Eq. (20.43). Since we have taken $T(x,0) = 0$ for the example under consideration, Eq. (20.43) becomes

$$\frac{d^2\overline{T}}{dx^2} - \frac{s}{\alpha}\overline{T} = 0 \tag{20.44}$$

Equation (20.44) is a linear differential equation and can be solved to give

$$\overline{T} = A_1 e^{-\sqrt{(s/\alpha)}x} + A_2 e^{\sqrt{(s/\alpha)}x} \tag{20.45}$$

The arbitrary coefficients A_1 and A_2 may be evaluated as follows: In order that $\overline{T}$ may be finite as $x \to \infty$, it is necessary that $A_2 = 0$. Equation (20.45) then becomes

$$\overline{T} = A_1 e^{-\sqrt{(s/\alpha)}x} \tag{20.45a}$$

The transformed forcing function at $x = 0$ is $\overline{T}(0,s)$, which can be substituted into Eq. (20.45a) to determine A_1; then

$$\overline{T}(0,s) = A_1 e^0$$

or
$$A_1 = \bar{T}(0,s)$$

Substituting A_1 into Eq. (20.45a) gives

$$\frac{\bar{T}(x,s)}{\bar{T}(0,s)} = e^{-\sqrt{(s/\alpha)}x} \tag{20.46}$$

By specifying a particular value of x, say $x = L$, the transfer function is

$$\frac{\bar{T}(L,s)}{\bar{T}(0,s)} = e^{-\sqrt{(s/\alpha)}L} \tag{20.47}$$

STEP RESPONSE. To illustrate the use of this transfer function, consider a forcing function that is the unit-step function; thus

$$T(0,t) = u(t)$$

for which case $\bar{T}(0,s) = 1/s$. Substituting this into Eq. (20.47) gives

$$\bar{T}(L,s) = \frac{1}{s}e^{-\sqrt{(s/\alpha)}L} \tag{20.48}$$

To obtain the response in the time domain, we must invert Eq. (20.48). A table of transforms gives the following transform pair:

$$L\left\{\frac{1}{s}e^{-\sqrt{(s/\alpha)}x}\right\} = \text{erfc}\,\frac{x}{\sqrt{4\alpha t}} \tag{20.49}$$

where erfc x is the complementary error function of x defined as

$$\text{erfc}\,x = 1 - \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}\,du$$

This function is tabulated in many textbooks and mathematical tables.
   Using this transform pair, Eq. (20.48) becomes

$$T(L,t) = \text{erfc}\,\frac{L}{\sqrt{4\alpha t}} = \text{erfc}\left[\frac{1}{2}\left(\frac{\alpha t}{L^2}\right)^{-1/2}\right] \tag{20.50}$$

A plot of T versus the dimensionless group $\alpha t/L^2$ is shown in Fig. 20–7.
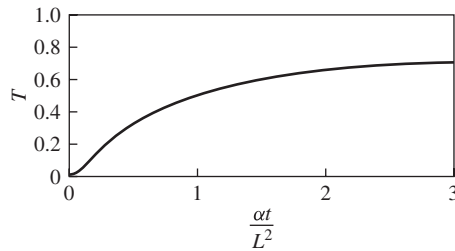
FIGURE 20–7
Response of temperature in the interior of a solid to
a unit-step change in temperature at the surface.

SINUSOIDAL RESPONSE. It is instructive to consider the response in temperature at
$x = L$ for the case where the forcing function is a sinusoidal variation; thus

$$T(0,t) = A \sin \omega t$$

Using the substitution rule of Chap. 15, in which s is replaced by $j\omega$, Eq. (20.47)
becomes

$$\frac{\bar{T}(L, j\omega)}{\bar{T}(0, j\omega)} = e^{-\sqrt{(jw/\alpha)}L} \qquad (20.51)$$

To obtain the AR and phase angle requires that the magnitude and argument of the right
side of Eq. (20.51) be evaluated. This can be done as follows: First write j in polar form
(see Fig. 20–8); thus

$$j = e^{(\pi/2)j} = \underbrace{1}_{\text{magnitude}} e^{\overbrace{(\pi/2)}^{\text{phase angle}} j}$$

from which we get

$$\sqrt{j} = \left(e^{j(\pi/2)}\right)^{1/2} = \pm e^{j\pi/4} = \pm \frac{1}{\sqrt{2}}(1 + j)$$
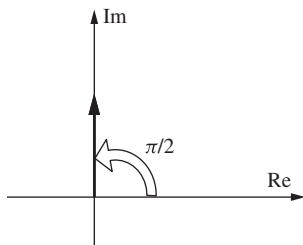
FIGURE 20–8
Polar representation of j.

Substituting the positive form of $\sqrt{j}$ into Eq. (20.51) gives

$$\frac{\bar{T}(L, j\omega)}{\bar{T}(0, j\omega)} = e^{-\sqrt{\omega/2\alpha}\, L} e^{-j\sqrt{\omega/2\alpha}\, L}$$

[Notice that the substitution of $-(1+j)/\sqrt{2}$ into Eq. (20.51) leads to a result in which the AR is greater than 1 and the phase angle leads. This is contrary to the response of the physical system and is not considered as a useful solution.]

From this form, we can write by inspection

$$\text{AR} = \left| \frac{\bar{T}(L, j\omega)}{\bar{T}(0, j\omega)} \right| = e^{-\sqrt{\omega/2\alpha}\, L} \tag{20.52}$$

$$\text{Phase angle} = \measuredangle \frac{\bar{T}(L, j\omega)}{\bar{T}(0, j\omega)} = -\sqrt{\frac{\omega}{2\alpha}} L \qquad \text{rad} \tag{20.53}$$

From these results, it is seen that the AR approaches zero as $\omega \to \infty$ and the phase angle decreases without limit as $\omega \to \infty$. Such a system is said to have *nonminimum* phase lag characteristics. With the exception of the transport lag, all the systems that have been considered up to now have given a limited value of phase angle as $\omega \to \infty$. These are called minimum phase systems and always occur for lumped-parameter systems. The nonminimum phase behavior is typical of distributed-parameter systems.

Transport Lag as a Distributed-Parameter System

We can demonstrate that the transport lag (distance-velocity lag) is, in fact, a distributed-parameter system as follows: Consider the flow of an incompressible fluid through an insulated pipe of uniform cross-sectional area A and length L, as shown in Fig. 20–9a. The fluid flows at velocity v, and the velocity profile is flat.
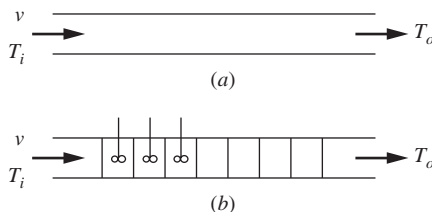
FIGURE 20–9
Obtaining the transfer function of a transport lag
from a lumped-parameter model.

We know from Chap. 7 that the transfer function relating outlet temperature T_o to the inlet temperature T_i is

$$\frac{T_o(s)}{T_i(s)} = e^{-(L/v)s}$$

Let the pipe be divided into n zones as shown in Fig. 20–9b. If each zone of length L/n is considered to be a well-stirred tank, then the pipe is equivalent to n noninteracting first-order systems in series, each having a time constant

$$\tau = \frac{L}{n}\frac{1}{v}$$

This expression for τ is equivalent to that appearing in Eq. (8.10). Since the transfer function for flow through a tank was developed in Chap. 8, the analysis will not be repeated here. (Note that taking each zone to be a well-stirred tank is called lumping of parameters.) The overall transfer function for this lumped-parameter model is therefore

$$\frac{T_o(s)}{T_i(s)} = \left(\frac{1}{\tau s + 1}\right)^n = \left\{\frac{1}{[(L/v)/n]s + 1}\right\}^n$$

To "distribute" the parameters, we let the size of the individual lumps go to zero by letting $n \to \infty$.

$$\frac{T_o(s)}{T_i(s)} = \lim_{n \to \infty}\left\{\frac{1}{[(L/v)/n]s + 1}\right\}^n$$

The thermal capacitance is now distributed over the tube length. It can be shown by use of calculus that the limit is

$$e^{-(L/v)s}$$

which is the transfer function derived previously. This demonstration should provide some initial insight into the relationship between a distributed-parameter system and a lumped-parameter system and indicates that a transport lag is a distributed system.

Heat Exchanger

As our last example of a distributed-parameter system, we consider the double-pipe heat exchanger shown in Fig. 20–10. [The analysis presented here essentially follows that of W. C. Cohen and E. F. Johnson (1956). These authors also present the experimental results of frequency response tests on a double-pipe, steam-to-water heat exchanger.] The fluid that flows through the inner pipe at constant velocity v is heated by steam condensing outside the pipe. The temperature of the fluid entering the pipe and the steam temperature vary according to some arbitrary functions of time. The steam temperature varies with time, but not with position in the exchanger. The metal wall separating steam from fluid is assumed to have negligible thermal capacity for the purpose of the analysis. The heat transfer from the steam to the fluid depends on the heat-transfer coefficient on the steam side h_o and the convective transfer coefficient on the water side h_i.

The resistance of the metal wall is neglected. The goal of the analysis will be to find transfer functions relating the exiting fluid temperature $T(L,t)$ to the entering fluid temperature $T(0,t)$ and the steam temperature $T_v(t)$.
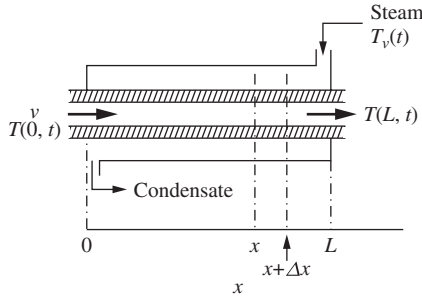
FIGURE 20–10
Double-pipe heat exchanger.

The following symbols will be used in this analysis:

$$
\begin{aligned}
T(x,t) &= \text{fluid temperature} \\
T_v(t) &= \text{steam temperature} \\
T_r &= \text{reference temperature for evaluating enthalpy} \\
v &= \text{fluid velocity} \\
\rho &= \text{density of fluid} \\
C &= \text{heat capacity of fluid} \\
A_i &= \text{cross-sectional area for flow inside pipe} \\
D_i &= \text{inside diameter of inner pipe} \\
D_o &= \text{outside diameter of inner pipe} \\
h_i &= \text{convective heat-transfer coefficient inside pipe} \\
h_o &= \text{heat-transfer coefficient for condensing steam} \\
U_i &= \text{overall heat-transfer coefficient based on inside area}
\end{aligned}
$$

$$
= \frac{1}{1/h_i + (1/h_o)(D_i/D_o)}
$$

ANALYSIS. We begin the analysis by writing a differential energy balance for the fluid inside the pipe over the volume element of length Δx (see Fig. 20–10). This balance can be stated as

$$
\left\{ \begin{array}{c} \text{Flow of} \\ \text{enthalpy in} \end{array} \right\} - \left\{ \begin{array}{c} \text{Flow of} \\ \text{enthalpy out} \end{array} \right\} + \left\{ \begin{array}{c} \text{Heat transferred} \\ \text{from steam} \\ \text{to liquid} \end{array} \right\}
$$

$$
= \left\{ \begin{array}{c} \text{Rate of accumulation} \\ \text{of internal energy} \end{array} \right\}
$$

(20.54)

The terms in this balance can be evaluated as follows:

Flow of enthalpy in at $x = vA_i \rho C(T - T_r)$

Flow of enthalpy out at $x + \Delta x = vA_i \rho C \left[\left(T + \dfrac{\partial T}{\partial x} \Delta x \right) - T_r \right]$

Heat transfer through film $= \pi D_i U_i \Delta x(T_v - T)$

Accumulation of internal energy $= \dfrac{\partial}{\partial t} \left[A_i \rho \, \Delta x \, C(T - T_r) \right]$

Introducing these terms into Eq. (20.54) gives, after simplification,

$$\frac{\partial T}{\partial t} = -v\frac{\partial T}{\partial x} + \frac{1}{\tau}(T_v - T) \tag{20.55}$$

where
$$\frac{1}{\tau} = \frac{\pi D_i U_i}{A_i \rho C}$$

We now have the differential equation that describes the dynamics of the system. As in previous problems, the dependent variables will be transformed to deviation variables. At steady state, the time derivative in Eq. (20.55) is zero, and it follows that

$$0 = -v\frac{dT_s}{dx} + \frac{1}{\tau}(T_{v_s} - T_s) \tag{20.56}$$

where the subscript s is used to denote the steady-state value. Note that to determine the steady-state temperature profile requires the solution of Eq. (20.56). We can rearrange this equation to

$$\frac{dT_s}{dx} + \frac{1}{\tau v}T_s = \frac{1}{\tau v}T_{v_s} u(t) \tag{20.57}$$

Transforming gives

$$\left(sT_s(s) - T_{s_{x=0}} \right) + \frac{1}{\tau v}T_s = \frac{1}{\tau v}T_{v_s}\left(\frac{1}{s} \right) \tag{20.58}$$

Now, let T_{s_0} represent the steady-state inlet (at $x = 0$) temperature of the water $T_{s_{x=0}}$ and rearrange:

$$T_s(s) = \frac{T_{v_s}/\tau v}{s(s + 1/\tau v)} + \frac{T_{s_0}}{s + 1/\tau v}$$

The steady-state temperature profile as a function of x may now be obtained by inverting

$$T_s(x) = T_{v_s} + (T_{s_0} - T_{v_s})\exp\left(-\frac{x}{v\tau}\right) \tag{20.59}$$

All equations for T' to be derived below should be recognized as deviations from this expression. Subtracting Eq. (20.56) from Eq. (20.55) and introducing deviation variables give

$$\frac{\partial T'}{\partial t} = -v\frac{\partial T'}{\partial x} + \frac{1}{\tau}\left(T'_v - T'\right) \tag{20.60}$$

where $T' = T - T_s$ and $T'_v = T_v - T_{v_s}$. Assuming that the exchanger is initially at steady state, so that $T_s(x,0) = T_s$ and $T_v(0) = T_{v_s}$, Eq. (20.60) may be transformed with respect to t to yield

$$sT'(x,s) = -v\frac{dT'(x,s)}{dx} + \frac{1}{\tau}\left[T'_v(s) - T'(x,s)\right]$$
$$\frac{dT'(x,s)}{dx} + \frac{s + 1/\tau}{v}T'(x,s) = \frac{1}{\tau v}T'_v(s) \tag{20.61}$$

Equation (20.61) is an ordinary first-order differential equation with the boundary condition $T'(x,s) = T'(0,s)$ at $x = 0$. We can transform Eq. (20.61) with respect to x (using p as the transform variable) to yield

$$pT'(p,s) - T'(0,s) + \frac{s + 1/\tau}{v}T'(p,s) = \frac{1/\tau v}{p}T'_v(s)$$

Rearranging gives

$$T'(p,s) = \frac{1/\tau v}{p[p + (s + 1/\tau)/v]}T'_v(s) + \frac{1}{p + [(s + 1/\tau)/v]}T'(0,s)$$

This expression is of the form

$$T'(p,s) = \frac{A}{p(p + B)}T'_v(s) + \frac{1}{p + B}T'(0,s) \tag{20.62}$$

where A and B are constants defined as $A = 1/\tau v$ and $B = (s + 1/\tau)/v$.

Equation (20.62) can now easily be inverted back to the (x,s) domain to yield

$$T'(x,s) = \frac{1}{\tau s + 1}\left(1 - e^{-[(\tau s + 1)/\tau v]x}\right)T'_v(s) + e^{-[(\tau s + 1)/\tau v]x}T'(0,s) \tag{20.63}$$

Let's look at the expression for the liquid outlet temperature from the exchanger, the quantity of interest in this example. The transform of the liquid outlet temperature is $T'(L,s)$. If we substitute $x = L$ into Eq. (20.63), we get

$$T'(L,s) = \frac{1}{\tau s + 1}\left(1 - e^{-[(\tau s + 1)/\tau v]L}\right)T'_v(s) + e^{-[(\tau s + 1)/\tau v]L}T'(0,s) \quad (20.64)$$

If we examine the exponent of the exponential term

$$-\left(\frac{\tau s + 1}{\tau v}\right)L = -\left(\frac{L}{v}s + \frac{L}{v\,\tau}\right) = -\left(\tau_d s + \frac{\tau_d}{\tau}\right)$$

we see a familiar term appear: τ_d is the transport lag on the liquid side of the exchanger. We can rewrite Eq. (20.64) as

$$T'(L,s) = \frac{1}{\tau s + 1}\left(1 - e^{-s\,\tau_d}\underbrace{e^{-\tau_d/\tau}}_{\text{constant}}\right)T'_v(s) + e^{-s\,\tau_d}\underbrace{e^{-\tau_d/\tau}}_{\text{constant}}T'(0,s)$$

$$\underbrace{T'(L,s)}_{\substack{\text{outlet temperature}\\(\text{output})}} = \underbrace{\left(\frac{1}{\tau s + 1} - \frac{Ke^{-s\,\tau_d}}{\tau s + 1}\right)}_{\substack{\text{transfer function relating}\\\text{outlet temperature to changes}\\\text{in steam temperature}}} \underbrace{T'_v(s)}_{\substack{\text{steam temperature}\\(\text{input})}} + \underbrace{\left(Ke^{-s\,\tau_d}\right)}_{\substack{\text{transfer function}\\\text{relating}\\\text{outlet temperature}\\\text{to changes in}\\\text{inlet temperature}}} \underbrace{T'(0,s)}_{\substack{\text{inlet temperature}\\(\text{disturbance})}}$$

$$(20.65)$$

where $K = e^{-\tau_d/\tau} = $ constant.

If the steam temperature is held constant $T'_v(s) = 0$, the transfer function relating the outlet temperature to the inlet temperature is

$$\frac{T'(L,s)}{T'(0,s)} = Ke^{-s\tau_d} \quad (20.66)$$

If we set $1/\tau = 0$ (which corresponds to the case of no heat transfer $U_0 = 0$), then $K = e^{-\tau_d/\tau} = e^0 = 1$. The response is simply that of a transport lag. This is the physical situation for which case the wall separating cold fluid from hot fluid acts as a perfect insulator. We saw in Chap. 7 that this situation is represented by a transport lag.

If the inlet temperature of the fluid entering the heat exchanger does not vary, the transfer function relating the exit fluid temperature to the steam temperature is

$$\frac{T'(L,s)}{T'_v(s)} = \frac{1}{\tau s + 1} - \frac{Ke^{-s\tau_d}}{\tau s + 1} \quad (20.67)$$

The response in the temperature of the fluid leaving the exchanger can be found for any forcing function, $T(0,t)$ or $T_v(t)$, by introducing the corresponding transforms into Eq. (20.65).

Example 20.1   Consider the heat exchanger shown in Fig. 20–10. The water is flowing through a 40-ft-long, $\frac{3}{4}$-in 14 BWG tube at a rate of 2 gal/min. The overall heat-transfer coefficient for the exchanger is $U_i = 100$ Btu/(ft$^2 \cdot$h$\cdot$°F). The steady-state inlet water temperature is 70°F, and the steady-state steam temperature is 212°F.

(a) Determine the steady-state temperature profile of the tube-side liquid as a function of distance from the inlet of the exchanger.
(b) For a 10°F step increase in the inlet temperature (steam temperature remains constant), determine the response of the exit temperature from the exchanger.
(c) For a 10°F step increase in the steam temperature (inlet temperature remains constant), determine the response of the exit temperature from the exchanger.

Solution

Data:

$D_o = 0.75$ in

$D_i = 0.584$ in

$$A_i = \text{flow area} = \frac{\pi D_i^2}{4} = \frac{\pi(0.584/12)^2}{4} = 0.00186 \text{ ft}^2$$

$$v = \left(2\frac{\text{gal}}{\text{min}}\right)\left(\frac{1 \text{ ft}^3}{7.48 \text{ gal}}\right)\left(\frac{1}{0.00186 \text{ ft}^2}\right)\left(\frac{1 \text{ min}}{60 \text{ s}}\right) = 2.4 \frac{\text{ft}}{\text{s}}$$

$$\frac{1}{\tau} = \frac{\pi D_i U_i}{A_i \rho C} = \frac{\pi\left[(0.584/12)\text{ft}\right]\left[100 \text{ Btu}/\left(\text{ft}^2 \cdot \text{h} \cdot °\text{F}\right)\right]}{\left(0.00186 \text{ ft}^2\right)\left(62.4 \text{ lb}/\text{ft}^3\right)\left[1 \text{ Btu}/(\text{lb} \cdot °\text{F})\right]} = 131.7 \text{ h}^{-1}$$

$$\tau = 0.00759 \text{ h} = 27.3 \text{ s}$$

(a) Making use of Eq. (20.59) gives

$$T_s(x) = T_{v_s} + (T_{s_0} - T_{v_s})\exp\left(-\frac{x}{v\tau}\right)$$

$$T_s(x) = 212 + (70 - 212)\exp\left[-\frac{x}{(2.4 \text{ ft}/\text{s})(27.3 \text{ s})}\right] = 212 - 142\exp\left(-\frac{x}{65.52}\right)$$

$$(20.59)$$

If we plot this result from $x = 0$ to $x = L = 40$ ft, we get the plot in Fig. 20–11.

Steady-State Temperature Profile
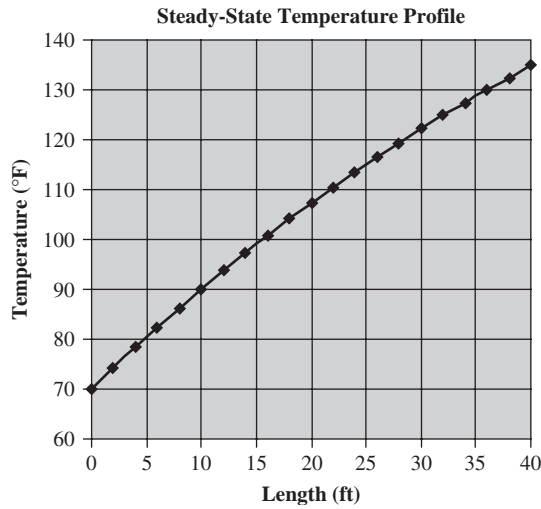
FIGURE 20–11
Steady-state temperature profile down length of heat exchanger tube
for Example 20.1.

This plot shows the temperature profile in the tube along the length of the exchanger. The initial steady-state temperature at $x = 40$ ft, $T_s(40)$, is 134.9°F. This is easily obtained by substituting $x = 40$ into the above expression.

(*b*) Using Eq. (20.66), we have

$$\frac{T'(L,s)}{T'(0,s)} = Ke^{-s\tau_d} \tag{20.66}$$

$$T'(L,s) = Ke^{-s\tau_d}T'(0,s) = e^{-\tau_d/\tau}e^{-s\tau_d}T'(0,s)$$

$$\tau_d = \frac{L}{v} = \frac{40 \text{ ft}}{2.4 \text{ ft}\!/\!_s} = 16.7 \text{ s}$$

$$T'(0,s) = \frac{10}{s} = \text{step function of magnitude 10}$$

$$T'(L,s) = \left(e^{-16.7 \text{ s}/27.3 \text{ s}}\right)\left(e^{-s(27.3 \text{ s})}\right)\left(\frac{10}{s}\right) = 0.542e^{-27.3\,s}\left(\frac{10}{s}\right) = \frac{5.42e^{-27.3\,s}}{s}$$

$$T'(L,s) = \left(e^{-16.7 \text{ s}/27.3 \text{ s}}\right)\left(e^{-s(16.7 \text{ s})}\right)\left(\frac{10}{s}\right) = 0.542e^{-16.7\,s}\left(\frac{10}{s}\right) = \frac{5.42e^{-16.7\,s}}{s}$$

and finally,

$$T'(L,t) = 5.42u(t - 16.7) = \begin{cases} 0 & t < 16.7 \text{ s} \\ 5.42 & t \geq 16.7 \text{ s} \end{cases}$$

So the response to a step change in the inlet temperature is delayed 16.7 s from the inlet change. The output increases by only 5.42°F (to 140.32°F) in response to a 10°F change in inlet temperature (the steady-state gain is 0.542).

(*c*) The applicable equation is Eq. (20.67).

$$\frac{T'(L,s)}{T'_v(s)} = \frac{1}{\tau s + 1} - \frac{Ke^{-s\tau_d}}{\tau s + 1}$$

$$T'(L,s) = \left(\frac{1}{\tau s + 1} - \frac{Ke^{-s\tau_d}}{\tau s + 1}\right)T'_v(s) = \left(\frac{1}{\tau s + 1} - \frac{Ke^{-s\tau_d}}{\tau s + 1}\right)\frac{10}{s}$$

$$T'(L,t) = 10\left(1 - e^{-t/\tau}\right)u(t) - 10K\left(1 - e^{-(t-\tau_d)/\tau}\right)u(t - \tau_d)$$

$$= 10\left(1 - e^{-t/27.3}\right)u(t) - 10(0.542)\left(1 - e^{-(t-16.7)/27.3}\right)u(t - 16.7)$$

$$T'(L,t) = \begin{cases} 0 & t < 0 \\ 10\left(1 - e^{-t/27.3}\right) & 0 \le t < 16.7 \\ 10\left(1 - e^{-t/27.3}\right) - 5.42\left(1 - e^{-(t-16.7)/27.3}\right) & t \ge 16.7 \end{cases}$$

A plot of the output response $T(L,t) = 134.9°F + T'\,(L,t)$ is shown in Fig. 20–12.
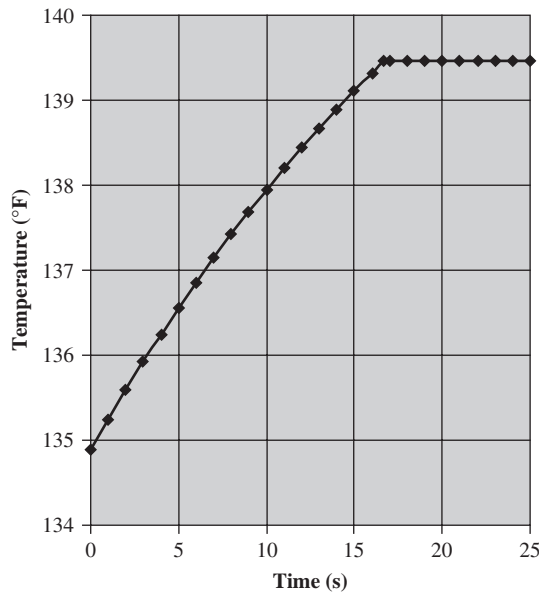
FIGURE 20–12
Plot of exit temperature from the heat exchanger as a function of time $T(L,t)$.

After the transport lag time (16.7 s) the outlet temperature from the exchanger remains constant.

SUMMARY

In this chapter, we analyzed several complex systems, using material and energy balances to predict their dynamic responses to disturbances. The process is often called *modeling,* and the resulting set of equations is referred to as the *mathematical model* of the system. In general, the model is based on the physics and chemistry of the system.

As systems are analyzed in greater detail and with fewer assumptions, the models that describe them become more complex, although more accurate. To predict the response of the system from the model requires that equations of the model be solved for some specific input disturbance. The only practical way to solve a complex model is to use computer simulation. The computer response will resemble that of the physical system if the model is accurate.

PROBLEMS

20.1 For the heat exchanger described in Example 20.1, determine the output response for a 10°F change in both steam temperature and inlet temperature.

20.2 What is the physical reason for the outlet temperature increasing less than 10°F for a step increase in the inlet temperature of 10°F? (That is, why doesn't the 10°F change propagate through the exchanger and ultimately appear at the outlet?)

CAPSULE SUMMARY

In this chapter, several complex systems have been analyzed mathematically. The result of each analysis was a set of equations (algebraic and/or differential) that presumably describe the dynamic response of the system to one or more disturbances. The process of obtaining the set of equations is often called *modeling,* and the set of equations is referred to as the *mathematical model* of the system. In general, the model is based on the physics and chemistry of the system. For example, in the analysis of a heat exchanger, one may write that the heat flux through a wall is equal to a convective transfer coefficient times a temperature driving force.

For a process not well understood, there is little chance that an accurate model can be obtained from the theoretical approach used here. For such systems, a direct dynamic test can be made. To do this, a known disturbance such as a pulse, step, or sinusoidal input is applied, and the response is recorded. This approach was discussed in Chap. 18. On the other hand, a model based on a theoretical analysis is extremely valuable, for it means that the system is well understood and that the effect of changes in system design and operation can be predicted.

The analysis of a steam-jacketed kettle provided an example of a nonlinear system containing nonlinear functions of several variables. The problem was handled by linearizing these functions about an operating point and ultimately obtaining a block diagram of the system from which the transfer function of the control system could be obtained. Although this approach is relatively straightforward, the resulting linear model can only be used over a narrow range of variables.

The analysis of the gas absorber gave some insight into the dynamic character of a typical multistage process that is widely used in the chemical process industries. A linear analysis of an *n*-plate column leads to *n* ordinary differential equations, which combine to give an overdamped *n*th-order response. Nonlinearities may be present in this system in such forms as a product of flow and concentration or a nonlinear equilibrium relationship. When the change of plate efficiency with flow is considered, the model of a gas absorber becomes even more complex. Most of the design techniques developed for multistage operations (e.g., gas absorption and distillation) have applied to steady-state operation. The dynamic analysis of such processes calls for dynamic parameters that are usually unavailable. For example, the liquid flow dynamics of trays used in distillation towers are relatively unknown.

The discussion of distributed-parameter systems further illustrated the complexities that can arise in physical systems. The distributed-parameter systems lead to partial differential equations, which may be very difficult to solve for most of the

forcing functions of practical interest. However, we saw that the response of distributed-parameter systems to sinusoidal forcing functions can be obtained directly by application of the substitution rule, in which s is replaced by $j\omega$. A distributed-parameter system features nonminimum phase lag characteristics. This is in sharp contrast to the lumped-parameter systems for which the phase angle approaches a limit at infinite frequency.

As systems are analyzed in greater detail and with fewer assumptions, the models that describe them become more complex, although more accurate. To predict the response of the system from the model requires that equations of the model be solved for some specific input disturbance. The only practical way to solve a complex model is to use computer simulation. The computer response will resemble that of the physical system if the model is accurate.

PART
VI

STATE-SPACE METHODS

CHAPTER
21

STATE-SPACE REPRESENTATION OF PHYSICAL SYSTEMS

21.1   INTRODUCTION

Up to this point, we have described dynamic physical systems by means of differential equations and transfer functions. Another method of description, which is widely used in all branches of control theory, is the state-space method. In fact, other disciplines of engineering (e.g., electrical engineering) introduce the state-space description before the transfer function description. The reader who plans to go beyond an introductory course in control or read from other engineering disciplines should be familiar with state-space methods. In the chapters of Part 6 of the book, the state-space method will be developed and compared with the transfer function method. It is much easier to start with the transfer function method and then develop the state-space method. The mathematical background needed for the transfer function approach involves differential equations and Laplace transforms. The additional mathematical background needed for the state-space method involves matrix algebra. Nearly all students today receive information on matrices in their mathematics courses. For those who are rusty in this topic, it is recommended that they review some of the fundamental matrix operations. A brief review of matrix algebra is given in App. 21A.

The transfer function approach is sufficient to calculate the response of linear control systems. The state-space approach is especially valuable in the field of optimal control of linear or nonlinear systems. The concepts developed in Part 6 of the book will be used in Part 7 on nonlinear control.

21.2   STATE VARIABLES

A linear physical system can be described mathematically by

- An nth-order differential equation
- A transfer function

- n first-order differential equations
- A matrix differential equation

So far, we have used the first two mathematical representations for describing physical systems. The third and fourth representations are referred to as state variable descriptions.

To illustrate these four methods of description, consider the familiar second-order process relating an output y to an input u. The four representations for this process are shown below.

1. **An nth-order differential equation** $(n = 2)$

$$\tau^2 \frac{d^2 y}{dt^2} + 2\zeta\tau \frac{dy}{dt} + y = u \tag{21.1}$$

2. **Transfer function.** The transfer function corresponding to Eq. (21.1) is

$$\frac{Y(s)}{U(s)} = \frac{1}{\tau^2 s^2 + 2\zeta\tau s + 1} \tag{21.2}$$

3. **n first-order differential equations** $(n = 2)$. Equation (21.1) can be expressed by two first-order differential equations. First, we need to make a variable change. Let's represent y as x_1 and $\dot{y}$ as x_2. Thus, $x_1 = y$ and $x_2 = \dot{y}$ (*Note:* $\dot{y} = dy/dt$ and $\ddot{y} = d^2 y/dt^2$.) Then,

$$\dot{x}_1 = \dot{y} = x_2 \tag{21.3}$$

and

$$\dot{x}_2 = \ddot{x}_1 = \ddot{y} \tag{21.4}$$

Solving Eq. (21.1) for $\ddot{y}$ yields

$$\ddot{y} = -\frac{1}{\tau^2} y - \frac{2\zeta}{\tau} \dot{y} + \frac{1}{\tau^2} u \tag{21.5}$$

Substituting in terms of x_1 and x_2, we obtain

$$\dot{x}_2 = -\frac{1}{\tau^2} x_1 - \frac{2\zeta}{\tau} x_2 + \frac{1}{\tau^2} u \tag{21.6a}$$

and

$$\dot{x}_1 = x_2 \tag{21.6b}$$

where x_1 and x_2 are called the state variables ($x_1 = y$ and $x_2 = \dot{y}$). Equations (21.6a) and (21.6b) are the system of two first-order differential equations that represent the second-order differential equation, Eq. (21.1).

We shall see later that other choices for x_1 and x_2 are possible; at this point, however, the reader is asked to accept Eqs. (21.6a) and (21.6b) as a valid description of the second-order system under consideration.

4. **Matrix differential equation.** Equations (21.6a) and (21.6b) can be written as one matrix differential equation as follows:

$$\dot{\mathbf{x}} \;=\; \mathbf{A}\mathbf{x} \,+\, \mathbf{b}u \tag{21.7}$$

or

$$\underbrace{\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix}}_{\dot{\mathbf{x}}} = \underbrace{\begin{bmatrix} 0 & 1 \\ \dfrac{-1}{\tau^2} & \dfrac{-2\zeta}{\tau} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{x}} + \underbrace{\begin{bmatrix} 0 \\ \dfrac{1}{\tau^2} \end{bmatrix}}_{\mathbf{b}} \underset{\text{scalar}}{\underset{\smile}{u}}$$

The representation given by Eqs. (21.6) and the representation given by Eq. (21.7) are exactly the same; Eq. (21.7) is in a more compact form. The state variables x_1 and x_2 are represented by the column vector **x.** The coefficients of the state variables on the right sides of Eqs. (21.6a) and (21.6b) are the elements of the matrix **A.** In this example, there is only one input or forcing term u, which is a scalar. Each term on the right side of Eq. (21.7) must be a vector containing two elements (i.e., a 2×1 matrix). For the expression given by Eq. (21.7) to agree with Eqs. (21.6a) and (21.6b), the coefficient of u must be a vector with the upper element zero. With some practice, the reader will be able to look at a matrix expression such as Eq. (21.7) and quickly see the equivalent set of differential equations.

The output y in representation 1 or 2 often represents a physical variable of interest, such as the temperature of a process or the position of a mechanical system. The alternate state variable representation given by Eq. (21.3) or Eq. (21.7) contains two state variables, one of which is y and the other of which is the derivative of y (or $\dot{y}$). In this case only y may be of interest to the control engineer; $\dot{y}$ is available, but may not be of interest since it cannot always be measured easily. (For example, there is no easy way to measure the rate of change of temperature if y represents temperature.)

State-Space Description

In general, a physical system can be described by state variables as follows

$$
\begin{aligned}
\dot{x}_1 &= f_1\left(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m\right) \\
\dot{x}_2 &= f_2\left(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m\right) \\
&\;\;\vdots \\
\dot{x}_n &= f_n\left(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m\right)
\end{aligned}
\tag{21.8}
$$

where x_1, x_2, . . ., x_n are n state variables and u_1, u_2, . . ., u_m are m inputs or forcing terms. The above set of equations may be written as a matrix expression as follows:

$$\dot{\mathbf{x}} \;=\; f\left(\mathbf{x}, \mathbf{u}\right)$$

If the system parameters vary with time, the vector f will contain explicit functions of time. An example for an element of f might be the expression on the right side of the following equation:

$$\dot{x}_1 = 2tx_1 + x_2 + u_1 + u_2$$

In this chapter, we shall be concerned with time-invariant systems for which $\dot{x}_i$ is a linear combination of state variables and the coefficients are constant. For the time-invariant case, we may write the general term $\dot{x}_i$ in Eq. (21.8) as follows:

$$\dot{x}_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n + b_{i1}u_1 + \cdots + b_{im}u_m \qquad (21.9)$$

for $i = 1, 2, 3, \ldots, n$. The equivalent matrix expression for Eq. (21.9) is

$$
\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_n \end{bmatrix} =
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\cdots & \cdots & \cdots & \cdots \\
a_{n1} & a_{n2} & \cdots & a_{nn}
\end{bmatrix}
\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} +
\begin{bmatrix}
b_{11} & b_{12} & \cdots & b_{1m} \\
b_{21} & b_{22} & \cdots & b_{2m} \\
\cdots & \cdots & \cdots & \cdots \\
b_{n1} & b_{n2} & \cdots & b_{nm}
\end{bmatrix}
\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix} \qquad (21.10)
$$

Writing this in the more compact matrix form, we have

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu} \qquad (21.11)$$

In this expression, there are m different inputs where $m \leq n$. Also, note that $\mathbf{u}$ in this expression is a vector quantity, not a scalar. The nature of the linear physical system expressed by Eq. (21.11) is completely stated by the matrices $\mathbf{A}$ and $\mathbf{B}$. For the time-invariant system, the elements of $\mathbf{A}$ and $\mathbf{B}$ are constants.

The outputs of interest to the control engineer may differ from the state variables x_i. The most general statement for relating the output to the state variables is

$$\mathbf{y} = \mathbf{Cx} \qquad (21.12)$$

where $\mathbf{y}$ is the vector of outputs $(y_1, y_2, \ldots, y_p)$ chosen by the control engineer for some practical reason. The matrix $\mathbf{C}$ is a $p \times n$ matrix containing constant elements. The way in which the matrix $\mathbf{C}$ is selected will be clarified in Example 21.1. In summary, the state-space description for a linear time-invariant system is given by Eqs. (21.11) and (21.12).

Example 21.1. For the two-tank noninteracting liquid-level system shown in Fig. 21–1, obtain the state-space description as expressed by Eqs. (21.11) and (21.12). The output y of interest is the level in tank 2. Notice that streams enter both tanks.

$$A_1 = 1 \qquad R_1 = 0.5$$
$$A_2 = 0.5 \qquad R_2 = \tfrac{2}{3}$$
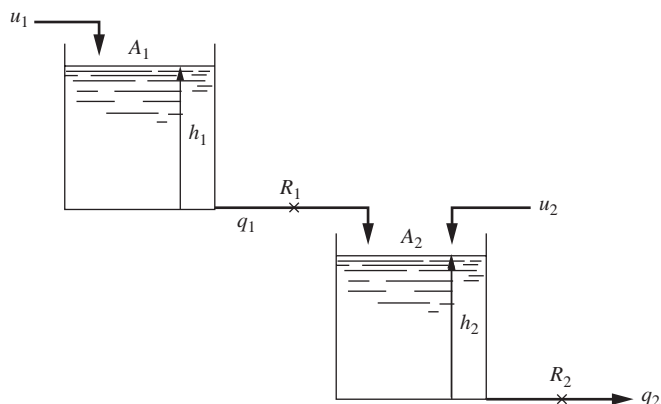
FIGURE 21–1
Two-tank liquid-level system.

For this example, let the state variables be the physical variables h_1 and h_2, which are the levels in tanks 1 and 2. These state variables are called physical variables because they can be easily measured or observed. (In another example, we shall consider a different set of state variables.)

For the liquid-level system shown in Fig. 21–1 we may write

$$A_1 \frac{dh_1}{dt} = u_1 - \frac{h_1}{R_1} \tag{21.13}$$

$$A_2 \frac{dh_2}{dt} = u_2 + \frac{h_1}{R_1} - \frac{h_2}{R_2} \tag{21.14}$$

or

$$\frac{dh_1}{dt} = -\frac{1}{R_1 A_1} h_1 + \frac{1}{A_1} u_1 \tag{21.15}$$

$$\frac{dh_2}{dt} = \frac{1}{R_1 A_2} h_1 - \frac{1}{A_2 R_2} h_2 + \frac{1}{A_2} u_2 \tag{21.16}$$

These equations can be written as

$$\dot{\mathbf{h}} = \mathbf{A}\mathbf{h} + \mathbf{B}\mathbf{u} \tag{21.17}$$

where

$$\underbrace{\begin{bmatrix} \dot{h}_1 \\ \dot{h}_2 \end{bmatrix}}_{\dot{\mathbf{h}}} = \underbrace{\begin{bmatrix} \dfrac{-1}{R_1 A_1} & 0 \\ \dfrac{1}{R_1 A_2} & \dfrac{-1}{R_2 A_2} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} h_1 \\ h_2 \end{bmatrix}}_{\mathbf{h}} + \underbrace{\begin{bmatrix} \dfrac{1}{A_1} & 0 \\ 0 & \dfrac{1}{A_2} \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}}_{\mathbf{u}}$$

If the output y is to be the level in tank 2 (h_2), we have

$$\mathbf{y} = \mathbf{Ch}$$

$$\underbrace{\begin{bmatrix} y_1 \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 0 & 1 \end{bmatrix}}_{\mathbf{C}} \underbrace{\begin{bmatrix} h_1 \\ h_2 \end{bmatrix}}_{\mathbf{h}}$$

and $\mathbf{y} = y_1 = h_2$. In this case $\mathbf{y}$ is a scalar (i.e., a 1×1 matrix).

The choice of output can be stated in many ways. Regardless of the choice, the output is related to the state variables by Eq. (21.12). To see how the matrix $\mathbf{C}$ depends on the choice of output, consider the following examples:

If $\mathbf{y}$ is to be a scalar that is the arithmetic average of the levels in the two tanks, one can show that

$$\underbrace{\begin{bmatrix} y_1 \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 0.5 & 0.5 \end{bmatrix}}_{\mathbf{C}} \underbrace{\begin{bmatrix} h_1 \\ h_2 \end{bmatrix}}_{\mathbf{h}}$$

If the output is to be h_1 and h_2, one can show that

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{C}} \underbrace{\begin{bmatrix} h_1 \\ h_2 \end{bmatrix}}_{\mathbf{h}}$$

Selection of State Variables

To the beginner, the selection of state variables may seem mysterious. The state variables of a system are the smallest set of variables that contain sufficient information to permit all future states to be determined. Although the number of state variables is fixed, the actual selection of these state variables is not unique. If possible, it is convenient to choose state variables that are directly related to physical variables which can be measured or observed (e.g., temperature, level, composition, position, velocity). For mechanical systems, transducers are available for measuring velocity; for this reason, velocity is considered a physical variable. On the other hand, since the measurement of rate of change of composition is not easily made, this variable is not usually considered a physical variable.

In the control literature, the types of state variables have been classified as follows.

1. *Physical variables.* State variables are called physical variables when they are readily measured and observed (level, temperature, composition, etc.). Physical variables were discussed at the beginning of this chapter and illustrated for a liquid-level system in Example 21.1 where $x_1 = h_1$ and $x_2 = h_2$.
2. *Phase variables.* State variables that are chosen to be the dependent variable and its successive derivatives are called phase variables. Phase variables were selected at the beginning of this chapter where $x_1 = y$ and $x_2 = \dot{y}$.

3. *Canonical variables.* If the state variables are selected to be canonical variables, the result is that the matrix **A** is diagonal. At this point, it is sufficient to say that canonical variables are selected as state variables for ease in matrix computation. In general, the canonical variables are not readily identified with physical variables.

In addition to the types of state variables listed above, any other legitimate set of variables can be selected. In Example 21.1, we used physical variables, namely, the levels in the tanks of the liquid-level system. In Examples 21.2 and 21.3, the method for selecting state variables will be shown.

Example 21.2.   For the two-tank liquid-level system of Example 21.1, shown in Fig. 21–1, obtain the state-space description as expressed by Eqs. (21.11) and (21.12) when phase variables are selected for the state variables. To simplify the problem, let $u_2 = 0$; i.e., there is only one input u_1.

For the system shown in Fig. 21–1, one can show that

$$\frac{H_2(s)}{U_1(s)} = \frac{R_2}{(\tau_1 s + 1)(\tau_2 s + 1)} \qquad (21.18)$$

where $\tau_1 = A_1 R_1$ and $\tau_2 = A_2 R_2$. Introducing the parameters in Fig. 21–1 into Eq. (21.18) gives

$$\frac{H_2(s)}{U_1(s)} = \frac{\frac{2}{3}}{\left(\frac{1}{2}s + 1\right)\left(\frac{1}{3}s + 1\right)} \qquad (21.19)$$

or

$$\frac{H_2(s)}{U_1(s)} = \frac{4}{(s + 2)(s + 3)} \qquad (21.20)$$

To obtain the differential equation corresponding to Eq. (21.20), we cross-multiply to obtain

$$(s + 2)(s + 3)H_2 = 4U_1$$

or

$$\left(s^2 + 5s + 6\right)H_2 = 4U_1$$

This may be expressed as the following differential equation:

$$\ddot{h}_2 + 5\dot{h}_2 + 6h_2 = 4u_1 \qquad (21.21)$$

Let the state variables be the following phase variables:

$$x_1 = h_2 \qquad (21.22)$$

$$x_2 = \dot{h}_2 \qquad (21.23)$$

We may now write

$$\dot{x}_1 = x_2\left[= \dot{h}_2\right] \qquad (21.24)$$

$$\dot{x}_2 = \ddot{h}_2 \qquad (21.25)$$

Equation (21.21) becomes

$$\dot{x}_2 + 5x_2 + 6x_1 = 4u_1 \tag{21.26}$$

The system can be described by Eqs. (21.24) and (21.26):

$$\dot{x}_1 = x_2 \tag{21.27a}$$
$$\dot{x}_2 = -6x_1 - 5x_2 + 4u_1 \tag{21.27b}$$

In terms of a matrix expression, Eqs. (21.27) may be written

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u_1$$

$$\underbrace{\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix}}_{\dot{\mathbf{x}}} = \underbrace{\begin{bmatrix} 0 & 1 \\ -6 & -5 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{x}} + \underbrace{\begin{bmatrix} 0 \\ 4 \end{bmatrix}}_{\mathbf{b}} \underbrace{u_1}_{\text{scalar}}$$

If the output **y** is to be the level in tank 2,

$$\underbrace{\begin{bmatrix} y \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{\mathbf{C}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{x}}$$

MATLAB Solution of Example 21.2

MATLAB has the capability to convert from a transfer function model to a state-space model of a system. From Eq. (21.20), the transfer function that relates the height in tank 2 to changes in the inlet flow to tank 1 is

$$\frac{H_2(s)}{U_1(s)} = \frac{4}{(s+2)(s+3)} = \frac{4}{s^2 + 5s + 6} \tag{21.20}$$

The MATLAB commands that create the state space representation are:

```
num=[4];

den=[1 5 6];

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

A =
      -5    -6
       1     0
B =
       1
       0
C =
       0     4
D =
       0
```

`tf2ss` is a "transfer function to state space" command that generates the matrices for the state-space model. The generalized model that MATLAB uses is

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu}$$

$$\mathbf{y} = \mathbf{Cx} + \mathbf{Du}$$

Note that the output **y** is some function of the states **x** and/or the inputs **u**. MATLAB returns a canonical form for the state-space representation that is equivalent (but not identical) to the one that we developed using phase variables. The output **y** that is returned from the **tf2ss** routine is a scalar that represents the dependent variable of the original transfer function, h2 in this case. Let's see if we can verify that the state-space model that MATLAB returns is a valid representation of the original system, Eq. (21.21). The MATLAB model is

$$\underbrace{\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix}}_{\dot{\mathbf{x}}} = \underbrace{\begin{bmatrix} -5 & -6 \\ 1 & 0 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{x}} + \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\mathbf{B}} u_1(t)$$

$$\underbrace{\begin{bmatrix} y \end{bmatrix}}_{\mathbf{y}} = h_2 = \underbrace{\begin{bmatrix} 0 & 4 \end{bmatrix}}_{\mathbf{C}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{x}} + \begin{bmatrix} 0 \end{bmatrix} u_1(t)$$

From the output equation,

$$h_2 = 4x_2$$
$$x_2 = 0.25h_2$$

and from the differential equations,

$$\dot{x}_1 = -5x_1 - 6x_2 + u_1(t)$$
$$\dot{x}_2 = x_1$$

and so we can write

$$x_1 = \dot{x}_2 = 0.25\dot{h}_2$$

and finally,

$$0.25\ddot{h}_2 = -5(0.25\dot{h}_2) - 6(0.25h_2) + u_1(t)$$
$$\ddot{h}_2 + 5\dot{h}_2 + 6h_2 = 4u_1(t)$$

which we recognize as Eq. (21.21), the original system. Although the state-space representation is not the same as the representation we derived using phase variables, it is a valid model for the system, and the output from a computer simulation of these models for h_2 would be identical.
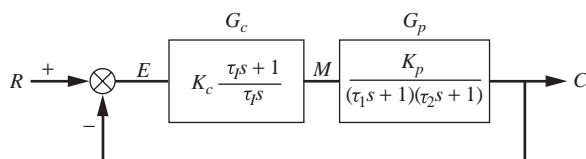
FIGURE 21–2
PI control system for Example 21.3.

Example 21.3. For the PI control system shown in Fig. 21–2, obtain a state-space representation in the form of Eq. (21.7); thus

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}r$$

where r is a scalar (the set point). Let

$$x_1 = c \tag{21.28}$$

$$x_2 = \dot{c} = \dot{x}_1 \tag{21.29}$$

With this choice of state variables, we have selected phase variables.

From Fig. 21–2, we may write

$$\frac{C(s)}{M(s)} = \frac{K_p}{(\tau_1 s + 1)(\tau_2 s + 1)} = \frac{K_p/\tau_1\tau_2}{(s + 1/\tau_1)(s + 1/\tau_2)}$$

or

$$\frac{C(s)}{M(s)} = \frac{A}{(s + a)(s + b)} \tag{21.30}$$

where $A = \dfrac{K_p}{\tau_1\tau_2}$

$a = \dfrac{1}{\tau_1}$

$b = \dfrac{1}{\tau_2}$

Cross-multiplying Eq. (21.30) gives

$$\left[s^2 + (a + b)s + ab \right] C(s) = AM(s)$$

or, in the time domain,

$$\ddot{c} + (a + b)\dot{c} + abc = Am \tag{21.31}$$

From Eqs. (21.28), (21.29), and (21.31) we obtain

$$\dot{x}_1 = x_2 \tag{21.32}$$

$$\dot{x}_2 = -abx_1 - (a + b)x_2 + Am \tag{21.33}$$

We must now obtain the state variables associated with the PI controller. From Fig. 21–2, we obtain

$$\frac{M(s)}{E(s)} = K_c \frac{\tau_I s + 1}{\tau_I s}$$

or $\qquad\qquad \tau_I s M(s) = K_c \tau_I s E(s) + K_c E(s)$

In terms of the time domain, this expression becomes

$$\dot{m} = K_c \dot{e} + \frac{K_c}{\tau_I} e \tag{21.34}$$

From the signals entering and leaving the comparator, we may write

$$e = r - c$$

or, since $x_1 = c,$ we may write

$$e = r - x_1 \tag{21.35a}$$

and $\qquad\qquad \dot{e} = \dot{r} - \dot{x}_1 \tag{21.35b}$

Combining Eqs. (21.34) and (21.35) gives

$$\dot{m} = K_c \left(\dot{r} - \underbrace{\dot{x}_1}_{x_2} \right) + \frac{K_c}{\tau_I}(r - x_1)$$

or $\qquad\qquad \dot{m} = K_c \dot{r} - K_c x_2 + \frac{K_c}{\tau_I} r - \frac{K_c}{\tau_I} x_1 \tag{21.36}$

   At this stage, we are faced with the difficulty of having a derivative term on the right side of Eq. (21.36). *In state-space representation, all variables on the right side must be state variables, not derivatives of state variables.* One way to handle the present difficulty is to define a new state variable x_3; let

$$x_3 = m - K_c r \tag{21.37}$$

or $\qquad\qquad \dot{x}_3 = \dot{m} - K_c \dot{r} \tag{21.38}$

Combining Eqs. (21.38) and (21.36) leads to

$$\dot{x}_3 = -\frac{K_c}{\tau_I}x_1 - K_c x_2 + \frac{K_c}{\tau_I}r \qquad (21.39)$$

or

$$\dot{x}_3 = -\alpha x_1 - K_c x_2 + \alpha r \qquad (21.40)$$

where $\alpha = K_c/\tau_I$. Summarizing the state variable equations given by Eqs. (21.32), (21.33), and (21.40) and using the definition of x_3 in Eq. (21.37) give

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = -abx_1 - (a + b)x_2 + Ax_3 + AK_c r$$
$$\dot{x}_3 = -\alpha x_1 - K_c x_2 + \alpha r$$

where

$$A = \frac{K_p}{\tau_1 \tau_2}$$

$$a = \frac{1}{\tau_1}$$

$$b = \frac{1}{\tau_2}$$

$$\alpha = \frac{K_c}{\tau_I}$$

Therefore, the reduced form $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}r$ is

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ -ab & -(a+b) & A \\ -\alpha & -K_c & 0 \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \underbrace{\begin{bmatrix} 0 \\ AK_c \\ \alpha \end{bmatrix}}_{\mathbf{b}} r(t)$$

If m is required as a function of t, it can always be found by solving Eq. (21.37) for m; thus

$$m = x_3 + K_c r$$

SUMMARY

State-space representation is an alternative to the transfer function representation of a physical system that we have used up to this point. A transfer function that relates an output variable to an input variable represents an nth-order differential equation. In the state-space representation, the nth-order differential equation is written as n first-order

differential equations in terms of n state variables. These n differential equations can also be written in a more compact form as a matrix differential equation

$$\dot{\mathbf{x}} \, = \, \mathbf{Ax} \, + \, \mathbf{Bu}$$

For an nth-order dynamic system, the number of state variables is fixed at n, but the selection of the variables is not unique. Of the many sets of state variables that one can choose, we discussed three sets that are useful in control theory, namely, physical variables, phase variables, and canonical variables. The state-space representation gives all the dynamic details of a system (e.g., the dependent variable and its successive derivatives for the case of phase variables, for example, h_1 and h_2 from Example (21.3). Whether or not this detail is needed depends on the problem being solved. We shall see the value of state-space representation in multivariable control and in nonlinear control in later chapters.

21A

ELEMENTARY MATRIX ALGEBRA

The purpose of this section is to provide, in a convenient location, a review of some of the elementary operations of matrix algebra for use in state-space methods. It is expected that the reader has had some course work in linear algebra discussing the concepts of a vector and a matrix and the operations performed on them.

VECTORS

An n-dimensional column vector is an ordered series of elements (numbers) $x_1, x_2, \ldots, x_n$ and is written as

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

Multiplication of a vector by a scalar $\lambda \mathbf{x}$ results in a vector for which each element is multiplied by λ.

MATRICES

A matrix is a rectangular array of elements (numbers) that takes the form

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

in which the elements are written a_{ij}. The subscript i refers to the ith row and j to the jth column.

Matrix $\mathbf{A}$ is called an $n \times m$ matrix where n is the number of rows and m is the number of columns. If $n = m$, the matrix is called a square matrix. If $m = 1$, the matrix is a column vector ($n \times 1$). If $n = 1$, the matrix is a row vector ($1 \times m$).

The transpose of a matrix $\mathbf{A}^T$ is a matrix for which the rows and columns of the matrix $\mathbf{A}$ are interchanged. If the diagonal elements (a_{ij} for all $i = j$) of a square matrix are 1 and all off-diagonal elements are 0, then the matrix is called a unit (or identity) matrix and is given the symbol $\mathbf{I}$. For example, a 2×2 unit matrix is given by $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

If $\mathbf{A} = \mathbf{A}^T$ for a square matrix, the matrix $\mathbf{A}$ is said to be symmetric.

When two matrices are added (or subtracted), the corresponding elements are added (or subtracted); thus

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & & a_{1m} + b_{1m} \\ a_{21} + b_{21} & a_{22} + b_{22} & & a_{2m} + b_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & & a_{nm} + b_{nm} \end{bmatrix}$$

The product of two matrices $\mathbf{C} = \mathbf{AB}$ is a matrix whose elements are obtained by the expression

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj} \qquad \text{for } i = 1, \dots, n \text{ and } j = 1, \dots, p$$

where $\mathbf{A}$ is an $n \times m$ matrix and $\mathbf{B}$ is an $m \times p$ matrix. The matrix $\mathbf{C}$ is an $n \times p$ matrix. The number of columns in $\mathbf{A}$ must equal the number of rows in $\mathbf{B}$ to perform this operation.

INVERSE OF A MATRIX

The inverse of a matrix is related to the concept of division for numbers. It can only be found for square matrices. The inverse of a number x is written $1/x$ or x^{-1}. The product of a number x and its inverse is equal to unity. The inverse of a matrix $\mathbf{A}$ is written $\mathbf{A}^{-1}$, and the product of a matrix and its inverse is equal to the unit matrix; thus

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

The expression used for matrix inversion takes the form

$$\mathbf{A}^{-1} = \frac{\text{adj } \mathbf{A}}{|\mathbf{A}|} \tag{21A.1}$$

where $|\mathbf{A}|$ is the determinant of $\mathbf{A}$ and adj $\mathbf{A}$ is the adjoint of $\mathbf{A}$. These two terms will now be described.

The determinant of a matrix $|\mathbf{A}|$ is a scalar which is computed from the elements of the matrix as follows:

$$|\mathbf{A}| = a_{i1} A_{i1} + a_{i2} A_{i2} + \dots + a_{in} A_{in}$$

or
$$|\mathbf{A}| = \sum_{j=1}^{n} a_{ij} A_{ij} \qquad \text{for any } i \qquad\qquad (21A.2)$$

where A_{ij}, the cofactor of element a_{ij}, is computed as

$$A_{ij} = (-1)^{i+j} M_{ij}$$

The determinant M_{ij} is the minor of the element a_{ij} and is defined as follows. If the row and column containing the element a_{ij} are deleted from a square matrix $\mathbf{A}$, the determinant of the resulting matrix, which is an $(n-1) \times (n-1)$ matrix, is the minor M_{ij}. An alternate expression for the calculation of a determinant which uses the elements of a specific column and its cofactors is as follows:

$$|\mathbf{A}| = \sum_{i=1}^{n} a_{ij} A_{ij} \qquad \text{for any } j \qquad\qquad (21A.3)$$

The determinant of a matrix with two equal rows or columns is zero.

We now define the adjoint of a matrix. Let the matrix $\mathbf{B}$ be an $n \times n$ matrix whose elements b_{ij} are the cofactors A_{ji} of $\mathbf{A}$, i.e., the transpose of the cofactor matrix. Matrix $\mathbf{B}$ is the adjoint of $\mathbf{A}$; thus $\mathbf{B} = \text{adj } \mathbf{A} = $ transpose of cofactor matrix, or

$$adj\ A = \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{bmatrix}$$

Some useful properties of the inverse are

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

$$\left(\mathbf{A}^{-1}\right)^{\mathrm{T}} = \left(\mathbf{A}^{\mathrm{T}}\right)^{-1}$$

$$\left(\mathbf{A}^{-1}\right)^{-1} = \mathbf{A}$$

The derivations of relationships presented here, as well as other properties of matrices, can be found in textbooks on linear algebra (see Anton, 1984).

EXAMPLES

1. Evaluate the determinant of $\mathbf{A}$ for the following matrix.

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 5 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}$$

For this problem, we use Eq. (21A.2) with $i = 1$ (i.e., use row 1).

$$|\mathbf{A}| = 2\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} - 3\begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix} + 5\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$$
$$= 2[(0)(0) - (1)(1)] - 3[(1)(0) - (1)(2)] + 5[(1)(1) - (0)(2)]$$
$$= 2(-1) - 3(-2) + 5(1) = 9$$

MATLAB Solution:

```
A=[2,3,5;1,0,1;2,1,0]

A =
     2     3     5
     1     0     1
     2     1     0
det(A)

ans =
     9
```

2. Find the inverse of the matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$$

$$\mathbf{A}^{-1} = \frac{\text{adj } \mathbf{A}}{|\mathbf{A}|}$$

The determinant of **A** is

$$|\mathbf{A}| = (2)(4) - (3)(1) = 5$$

The matrix of minors is

$$\begin{bmatrix} 4 & 1 \\ 3 & 2 \end{bmatrix}$$

The cofactor matrix is

$$\begin{bmatrix} 4 & -1 \\ -3 & 2 \end{bmatrix}$$

The adjoint of the matrix, which is the transpose of the cofactor matrix, is

$$\text{adj } \mathbf{A} = \begin{bmatrix} 4 & -3 \\ -1 & 2 \end{bmatrix}$$

therefore $\mathbf{A}^{-1} = \dfrac{1}{5}\begin{bmatrix} 4 & -3 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ -\frac{1}{5} & \frac{2}{5} \end{bmatrix}$

MATLAB Solution

```
A= [2,3;1,4]

A =
       2       3
       1       4
A^-1

ans =
     0.8000     -0.6000
    -0.2000      0.4000
```

3.   Obtain the inverse of the matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}$$

One can show that

$$|\mathbf{A}| = 18$$

The cofactor matrix is

$$\begin{bmatrix} 1 & 7 & -5 \\ -5 & 1 & 7 \\ 7 & -5 & 1 \end{bmatrix}$$

The adjoint matrix is

$$\begin{bmatrix} 1 & -5 & 7 \\ 7 & 1 & -5 \\ -5 & 7 & 1 \end{bmatrix}$$

$$\mathbf{A}^{-1} = \frac{1}{18}\begin{bmatrix} 1 & -5 & 7 \\ 7 & 1 & -5 \\ -5 & 7 & 1 \end{bmatrix}$$

MATLAB Solution

```
A=[2,3,1;1,2,3;3,1,2]
A =
     2      3      1
     1      2      3
     3      1      2
A^-1
ans =
    0.0556    -0.2778     0.3889
    0.3889     0.0556    -0.2778
   -0.2778     0.3889     0.0556 which is the same as we determined by
hand... if we multiply each element by 18, the result will be obvious.
ans*18
ans =
    1.0000    -5.0000     7.0000
    7.0000     1.0000    -5.0000
   -5.0000     7.0000     1.0000
```

PROBLEMS

21.1. In the liquid-level process shown in Fig. P21–1, the three tanks are interacting. The process
may be described by

$$\dot{x} = Ax + Bu$$

where $\quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad$ and $\quad u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$
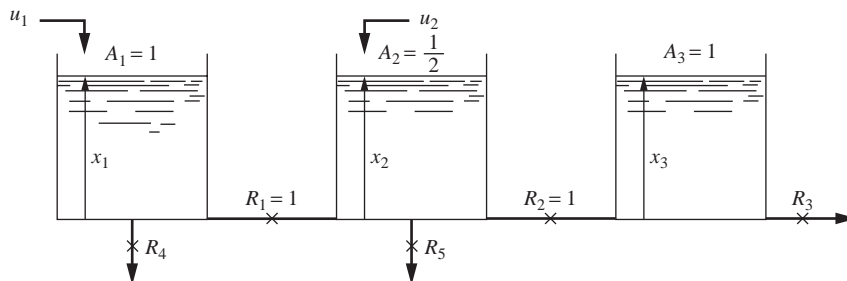
FIGURE P21–1

(*a*) If

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 0 \\ 2 & -3 & 2 \\ 0 & 1 & -3 \end{bmatrix}$$

determine values of R_3, R_4, and R_5. If one of these values of R is negative, what is your interpretation?

(*b*) Determine **B**.

21.2. For the system shown in Fig. P21–2, find **A** and **b** in

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

The tanks are interacting. The following data apply:

$$A_1 = 1, \qquad A_2 = \tfrac{1}{2}, \qquad R_1 = \tfrac{1}{2}, \qquad R_2 = 2, \qquad R_3 = 1$$
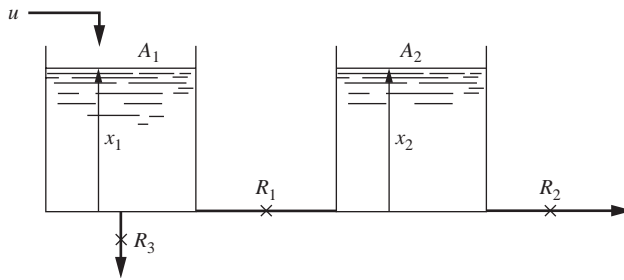
FIGURE P21–2

CAPSULE SUMMARY

A linear physical system can be described mathematically in several ways. For a second-order differential equation $\tau^2\, d^2y/dt^2 + 2\zeta\,\tau\, dy/dt + y = u,$ we may represent it by

- **A transfer function.**

$$\frac{Y(s)}{U(s)} = \frac{1}{\tau^2 s^2 + 2\zeta\tau s + 1}$$

- **Two first-order differential equations**

$$x_1 = y \quad \text{and} \quad x_2 = \dot{y}$$
$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = -\frac{1}{\tau^2}x_1 - \frac{2\zeta}{\tau}x_2 + \frac{1}{\tau^2}u$$

- **A matrix differential equation.**

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

$$\underbrace{\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix}}_{\dot{\mathbf{x}}} = \underbrace{\begin{bmatrix} 0 & 1 \\ -\dfrac{1}{\tau^2} & -\dfrac{2\zeta}{\tau} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{x}} + \underbrace{\begin{bmatrix} 0 \\ \dfrac{1}{\tau^2} \end{bmatrix}}_{\mathbf{b}} \underset{\text{scalar}}{u}$$

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

The x_i are the **state variables,** and the y_i are the **outputs.** In control literature, there are generally three types of state variables used:

1. *Physical variables* (readily measured and observed, i.e., level, temperature, composition, etc.)
2. *Phase variables* (the dependent variable and its successive derivatives)
3. *Canonical variables* (state variables selected so that matrix **A** is diagonal, chosen for ease in matrix computation)

SOME MATLAB COMMANDS

tf2ss—converts a transfer function model to a state-space model. The state variables are canonical.
det(A)—finds the determinant of square matrix **A**
A^-1 or **inv(A)**—finds the inverse of square matrix **A**

CHAPTER
22

TRANSFER
FUNCTION MATRIX

In Chap. 21, we have seen that a linear dynamic system can be expressed in terms of the following equations

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu}$$

(22.1)

$$\mathbf{y} = \mathbf{Cx}$$

(22.2)

where $\mathbf{x}$ = column vector of n state variables $(x_1, x_2, \ldots, x_n)$
  $\mathbf{u}$ = column vector of m inputs or forcing terms $(u_1, u_2, \ldots, u_m)$
  $\mathbf{y}$ = column vector of p outputs $(y_1, y_2, \ldots, y_p)$
  $\mathbf{A}$ = $n \times n$ matrix of coefficients
  $\mathbf{B}$ = $n \times m$ matrix of coefficients
  $\mathbf{C}$ = $p \times n$ matrix of coefficients
One of the objectives of this chapter is to show how one solves Eqs. (22.1) and (22.2) in a systematic manner.

Before discussing the solution of the matrix differential equation of Eq. (22.1), we consider the scalar differential equation

$$\frac{dx}{dt} = Ax + Bu$$

(22.3)

In this equation all the terms are scalars. If we multiply both sides of Eq. (22.3) by e^{-At}, we obtain, after rearrangement,

$$e^{-At}\, dx - Axe^{-At}\, dt = Be^{-At}u(t)\, dt$$

We can recognize that the left-hand side is an exact differential, and we can rewrite this expression as

$$d\left(xe^{-At}\right) = Be^{-At}u(t)\,dt$$

Integrating both sides from $t = 0$ to $t = t$, we obtain

$$\int_0^t d\left(xe^{-At}\right) = \int_{\tau=0}^{\tau=t} Be^{-A\tau}u(\tau)\,d\tau$$

$$xe^{-At} - x(0) = \int_{\tau=0}^{\tau=t} Be^{-A\tau}u(\tau)\,d\tau$$

$$x = e^{At}x(0) + e^{At}\int_{\tau=0}^{\tau=t} Be^{-A\tau}u(\tau)\,d\tau = e^{At}x(0) + \int_{\tau=0}^{\tau=t} Be^{A(t-\tau)}u(\tau)\,d\tau$$

The solution to Eq. (22.3) can thus be written as the sum of the complementary function and the particular integral as follows:

$$x(t) = \underbrace{e^{At}x(0)}_{\substack{\text{particular}\\\text{solution}}} + \underbrace{\int_0^t e^{A(t-\tau)}Bu(\tau)\,d\tau}_{\substack{\text{complementary}\\\text{solution}}} \tag{22.4}$$

Equation (22.4) is a well-known result that has been derived in many books on the solution of ordinary differential equations.

22.1   TRANSITION MATRIX

Let us now turn our attention to the solution of the matrix differential equation

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \tag{22.5}$$

This is Eq. (22.1) for the case of no inputs (i.e., $\mathbf{u} = 0$). The initial conditions for Eq. (22.5) may be expressed as $\mathbf{x}(0)$. One can show that the solution to Eq. (22.5) with initial conditions $\mathbf{x}(0)$ is given by

$$\mathbf{x}(t) = \left\{\mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2}{2!}t^2 + \cdots + \frac{\mathbf{A}^k}{k!}t^k\right\}\mathbf{x}(0) \tag{22.6}$$

The infinite series of matrix terms within the braces is given the symbol $e^{\mathbf{A}t}$. This symbol is chosen to recall that the infinite series of the scalar term e^{at} is

$$1 + at + \frac{a^2}{2!}t^2 + \cdots + \frac{a^k}{k!}t^k$$

Using the symbol $e^{\mathbf{A}t}$, we may write Eq. (22.6) as

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0) \tag{22.7}$$

The symbol $e^{\mathbf{A}t}$ is an $n \times n$ matrix in which each element contains a power series of t. The solution to Eq. (22.1) can be shown to be

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)\,d\tau \tag{22.8}$$

Notice that Eq. (22.8) resembles Eq. (22.4), which is the solution for the scalar differential equation. Since $e^{\mathbf{A}t}$ is awkward and perhaps misleading as to its nature, $e^{\mathbf{A}t}$ is sometimes replaced by $\phi(t)$; thus

$$\boldsymbol{\phi}(t) = e^{\mathbf{A}t} \qquad \text{(transition matrix)} \tag{22.9}$$

Either of the terms $\boldsymbol{\phi}(t)$ and $e^{\mathbf{A}t}$ can be used for the transition matrix. In this book, we shall use $e^{\mathbf{A}t}$.

Example 22.1.   Solution of a matrix differential equation. Solve the following matrix differential equation

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

where $u(t)$ is a unit-step function and

$$\mathbf{x}(0) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

One can show that

$$e^{\mathbf{A}t} = \begin{bmatrix} e^{-t} & e^{-t} - e^{-2t} \\ 0 & e^{-2t} \end{bmatrix}$$

In the next section, the method used to obtain the elements of this matrix will be developed. Applying Eq. (22.8) gives

$$\mathbf{x}(t) = \begin{bmatrix} e^{-t} & e^{-t} - e^{-2t} \\ 0 & e^{-2t} \end{bmatrix}\begin{bmatrix} -1 \\ 0 \end{bmatrix} + \int_0^t \begin{bmatrix} e^{-(t-\tau)} & e^{-(t-\tau)} - e^{-2(t-\tau)} \\ 0 & e^{-2(t-\tau)} \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix}d\tau$$

or $\qquad \mathbf{x}(t) = \begin{bmatrix} -e^{-t} \\ 0 \end{bmatrix} + \begin{bmatrix} e^{-(t-\tau)} - 0.5e^{-2(t-\tau)} \\ 0.5e^{-2(t-\tau)} \end{bmatrix}\Bigg|_{\tau=0}^{\tau=t}$

or $\qquad \mathbf{x}(t) = \begin{bmatrix} 0.5 - 2e^{-t} + 0.5e^{-2t} \\ 0.5 - 0.5e^{-2t} \end{bmatrix}$

MATLAB Solution of Matrix Differential Equation

Let's resolve Example 22.1 using MATLAB.

A=[-1,1;0,-2]

```
A =
    -1     1
     0    -2
```

syms t   %Declare *t* to be a symbolic variable.

B=expm(A*t)   %Calculate the transition matrix.

```
B =
[exp(-t), exp(-t)-exp(-2*t)]
[0,           exp(-2*t)]
```

C=B*[-1;0] %This is the "particular" solution, the first term on the right-hand side.

```
C =
[ -exp(-t)]
[        0]
```

syms tau   %Declare tau to be symbolic.

D=expm(A*(t-tau))   %Preparing the integrand
```
D =
[exp(-t+tau), -exp(-2*t+2*tau)+exp(-t+tau)]
[0,                  exp(-2*t+2*tau)]
```

E=D*[0;1]   %Completing the integrand

```
E =
[-exp(-2*t+2*tau)+exp(-t+tau)]
[exp(-2*t+2*tau)]
```

F=int(E,tau,0,t)   %Integrating the expression over tau, from tau=0 to tau=*t*

```
F =
[1/2+1/2*exp(-2*t)-exp(-t)]
[1/2-1/2*exp(-2*t)]   %The "complementary" solution
```

G=F+C

```
G =
[1/2+1/2*exp(-2*t)-2*exp(-t)]
[1/2-1/2*exp(-2*t)]   %The complete solution . . . this checks with our hand calculation above.
```

Determining $e^{\mathbf{A}t}$

One method for determining the elements of the transition matrix $e^{\mathbf{A}t}$ is to use Laplace transforms. Consider the matrix differential equation of Eq. (22.1)

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu}$$

If we take the Laplace transform of each side, we obtain

$$s\mathbf{X}(s) - \mathbf{x}(0) = \mathbf{AX}(s) + \mathbf{BU}(s)$$

or

$$s\mathbf{X}(s) - \mathbf{AX}(s) = \mathbf{x}(0) + \mathbf{BU}(s)$$

Solving for $\mathbf{X}(s)$ gives

$$(s\mathbf{I} - \mathbf{A})\mathbf{X}(s) = \mathbf{x}(0) + \mathbf{BU}(s) \tag{22.10}$$

To obtain an expression for $\mathbf{X}(s)$, premultiply both sides of Eq. (22.10) by $(s\mathbf{I} - \mathbf{A})^{-1}$; thus

$$(s\mathbf{I} - \mathbf{A})^{-1}(s\mathbf{I} - \mathbf{A})\mathbf{X}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}(0) + (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{BU}(s)$$

This equation becomes

$$\mathbf{X}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}(0) + (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{BU}(s) \tag{22.11}$$

To obtain $\mathbf{x}(t)$ from Eq. (22.11), we take the inverse transform; thus

$$\mathbf{x}(t) = L^{-1}\left\{(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}(0)\right\} + L^{-1}\left\{(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{BU}(s)\right\} \tag{22.12}$$

By comparing Eqs. (22.8) and (22.12), we see that

$$e^{\mathbf{A}t} = L^{-1}\left\{(s\mathbf{I} - \mathbf{A})^{-1}\right\} \tag{22.13}$$

and

$$\int_0^t e^{\mathbf{A}t - \tau}\mathbf{Bu}(\tau)\,d\tau = L^{-1}\left\{(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{BU}(s)\right\} \tag{22.14}$$

22.2   TRANSFER FUNCTION MATRIX

When $\mathbf{x}(0) = 0$, a case frequently used in control applications, we obtain from Eq. (22.11)

$$\mathbf{X}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{BU}(s) \tag{22.15}$$

This may be written

$$\mathbf{X}(s) = \mathbf{G}(s)\mathbf{U}(s) \tag{22.16}$$

where

$$\mathbf{G}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\,(\text{transfer function matrix}) \tag{22.17}$$

The term $\mathbf{G}(s)$ is called the *transfer function matrix* and serves the same purpose as the transfer function for the scalar case; namely, it relates a set of state variables $\mathbf{X}(s)$ to a set of inputs $\mathbf{U}(s)$.

If we prefer to relate the output to the input as expressed by Eq. (22.2), we may proceed as follows. Taking the Laplace transform of both sides of Eq. (22.2) gives

$$\mathbf{Y}(s) = \mathbf{C}\mathbf{X}(s) \tag{22.18}$$

Combining Eqs. (22.15) and (22.18) gives

$$\mathbf{Y}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{U}(s)$$

We may now write

$$\mathbf{Y}(s) = \mathbf{G}_1(s)\mathbf{U}(s) \tag{22.19}$$

where $$\mathbf{G}_1(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} \tag{22.20}$$

The term $\mathbf{G}_1(s)$ in Eq. (22.20) is also a transfer function matrix that relates the output vector $\mathbf{Y}$ to the input vector $\mathbf{U}.$

Example 22.2. Determine the transfer function matrix for the two-tank liquid-level system shown in Fig. 22–1. Given: $A_1 = 1, A_2 = 0.5, R_1 = 0.5, R_2 = \frac{2}{3}.$
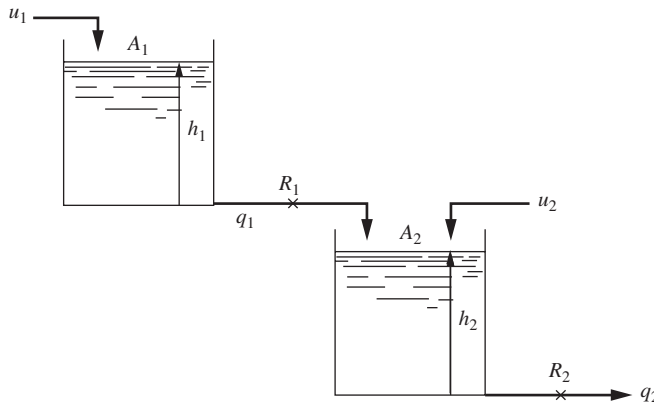
FIGURE 22–1
Two-tank liquid-level system.

As developed in Example 21.1 [Eq. (21.17)], this system is described by

$$\dot{\mathbf{h}} = \mathbf{Ah} + \mathbf{Bu}$$

where

$$\mathbf{A} = \begin{bmatrix} -2 & 0 \\ 4 & -3 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

From the definition of the transfer function matrix of Eq. (22.17), we write

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

The inverse of $s\mathbf{I} - \mathbf{A}$ is obtained as follows (see App. 21A for details on the inversion of a matrix):

$$(s\mathbf{I} - \mathbf{A})^{-1} = \frac{\mathrm{adj}(s\mathbf{I} - \mathbf{A})}{|s\mathbf{I} - \mathbf{A}|}$$

$$s\mathbf{I} - \mathbf{A} = \begin{bmatrix} s + 2 & 0 \\ -4 & s + 3 \end{bmatrix}$$

Cofactor of $\qquad s\mathbf{I} - \mathbf{A} = \begin{bmatrix} s + 3 & 4 \\ 0 & s + 2 \end{bmatrix}$

We can now find the adjoint:

$$\mathrm{adj}(s\mathbf{I} - \mathbf{A}) = \begin{bmatrix} s + 3 & 0 \\ 4 & s + 2 \end{bmatrix}$$

The determinant of $s\mathbf{I} - \mathbf{A}$ is

$$\begin{vmatrix} s + 2 & 0 \\ -4 & s + 3 \end{vmatrix} = (s + 2)(s + 3)$$

We can now determine the inverse of $s\mathbf{I} - \mathbf{A}.$

$$(s\mathbf{I} - \mathbf{A})^{-1} = \frac{\begin{bmatrix} s + 3 & 0 \\ 4 & s + 2 \end{bmatrix}}{(s + 2)(s + 3)} \tag{22.21}$$

$$\mathbf{G}(s) = \frac{\begin{bmatrix} s + 3 & 0 \\ 4 & s + 2 \end{bmatrix}}{(s + 2)(s + 3)} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} = \frac{\begin{bmatrix} s + 3 & 0 \\ 4 & 2(s + 2) \end{bmatrix}}{(s + 2)(s + 3)}$$

Simplifying this expression gives

$$\mathbf{G}(s) = \begin{bmatrix} \dfrac{1}{s + 2} & 0 \\ \dfrac{4}{(s + 2)(s + 3)} & \dfrac{2}{s + 3} \end{bmatrix}$$

From Eq. (22.16) we write

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

therefore

$$\begin{bmatrix} H_1(s) \\ H_2(s) \end{bmatrix} = \begin{bmatrix} \dfrac{1}{s+2} & 0 \\ \dfrac{4}{(s+2)(s+3)} & \dfrac{2}{s+3} \end{bmatrix} \begin{bmatrix} U_1(s) \\ U_2(s) \end{bmatrix} \qquad (22.22)$$

From Eq. (22.22), we obtain

$$H_1(s) = \frac{1}{s+2} U_1(s)$$

and $\qquad H_2(s) = \dfrac{4}{(s+2)(s+3)} U_1(s) + \dfrac{2}{s+3} U_2(s)$

For given inputs, the above equations may be inverted to obtain $h_1(t)$ and $h_2(t)$. For the case of $U_1(s) = 1/s$ and $U_2(s) = 0$, we get

$$H_1(s) = \frac{1}{s(s+2)} = \frac{0.5}{s(0.5s+1)}$$

and $\qquad H_2(s) = \dfrac{4}{s(s+2)(s+3)}$

Inversion of $H_1(s)$ and $H_2(s)$ gives

$$h_1(t) = 0.5\left(1 - e^{-2t}\right)$$

$$h_2(t) = \frac{2}{3} + \frac{4}{3}e^{-3t} - 2e^{-2t} = \frac{2}{3}\left(1 + 2e^{-3t} - 3e^{-2t}\right)$$

The results given above can be obtained, of course, by the methods presented earlier in this book.

The transition matrix can be obtained by applying Eq. (22.13) to Eq. (22.21):

$$e^{\mathbf{A}t} = L^{-1}\left\{\begin{bmatrix} \dfrac{1}{s+2} & 0 \\ \dfrac{4}{(s+2)(s+3)} & \dfrac{1}{s+3} \end{bmatrix}\right\}$$

Inverting each term in the matrix gives

$$e^{\mathbf{A}t} = \begin{bmatrix} e^{-2t} & 0 \\ 4\left(e^{-2t} - e^{-3t}\right) & e^{-3t} \end{bmatrix}$$

This matrix can be used in Eq. (22.8) to calculate $h_1(t)$ and $h_2(t)$. The result will be the same as obtained by inversion of Eq. (22.22).

MATLAB Solution of Example 22.2

```
A = [-2,0;4,-3]

A =
     -2    0
      4   -3

B = [1,0;0,2]

B =
      1    0
      0    2
syms s t
I = eye(2) %Set up the identity matrix

I =
      1    0
      0    1
E = s*I-A

E =
[s+2,   0]
[-4, s+3]

F = E^ -1

F =
[1/(s+2),                0]
[4/(s+2)/(s+3), 1/(s+3)]

G = F*B

G =
[1/(s+2),                0]
[4/(s+2)/(s+3), 2/(s+3)]
```
This is the same expression that we obtained by hand for $\mathbf{G}(s)$.

Let's check the transition matrix.

```
H = ilaplace(F)

H =
[ exp(-2*t),                        0]
[ 4*exp(-2*t)-4*exp(-3*t), exp(-3*t)]        %This is exp(At), the transition matrix
```

This matches the value we arrived at above for the transition matrix.

```
syms tau
K = [exp(-2*(t-tau)),0; 4*exp(-2*(t-tau))-4*exp(-3*(t-tau)),
exp(-3*(t-tau))]

K =
[ exp(-2*t+2*tau),                                    0]
[ 4*exp(-2*t+2*tau)-4*exp(-3*t+3*tau), exp(-3*t+3*tau)]   %This is exp(A(t-τ))


M = K*B

M =
[exp(-2*t+2*tau),                                     0]
[ 4*exp(-2*t+2*tau)-4*exp(-3*t+3*tau), 2*exp(-3*t+3*tau)]


N = M*[1;0]             %This is M*u, remember u₁ = 1 and u₂ = 0

N =
[exp(-2*t+2*tau)]
[4*exp(-2*t+2*tau)-4*exp(-3*t+3*tau)]


P = int(N,tau,0,t)    %The complementary solution.

P =
[1/2-1/2*exp(-2*t)]
[2/3-2*exp(-2*t)+4/3*exp(-3*t)]    % The first element is H1 and the second is H2.
```

This result matches the earlier result we obtained for H_1 and H_2.

Simulink Modeling of Example 22.2

Simulink can model a state-space system. Using the model developed in Example 22.2,

$$\begin{bmatrix} \dot{h}_1 \\ \dot{h}_2 \end{bmatrix} = \underbrace{\begin{bmatrix} -2 & 0 \\ 4 & -3 \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}}_{\mathbf{B}} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

If we desire to examine the response in tank 2, we want the output y to be equal to h_2. Therefore,

$$y = \underbrace{\begin{bmatrix} 0 & 1 \end{bmatrix}}_{\mathbf{C}} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & 0 \end{bmatrix}}_{\mathbf{D}} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \text{output}$$

The Simulink block diagram for the model is shown in Fig. 22–2.
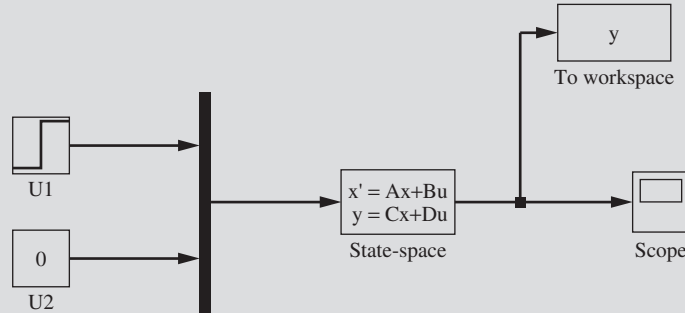
FIGURE 22–2
Simulink model for Example 22.2.

The simulation is quite easy to set up. Double-clicking on the state-space block pops up a menu that must be filled in using the values for the matrices **A, B, C,** and **D** (see Fig. 22–3). The completed menu box is shown below. The input signals u_1 and u_2 are multiplexed and fed into the state-space block. The upper signal is u_1, a unit step, and the lower signal is u_2, which is zero. The scope output is shown in Fig. 22–4.
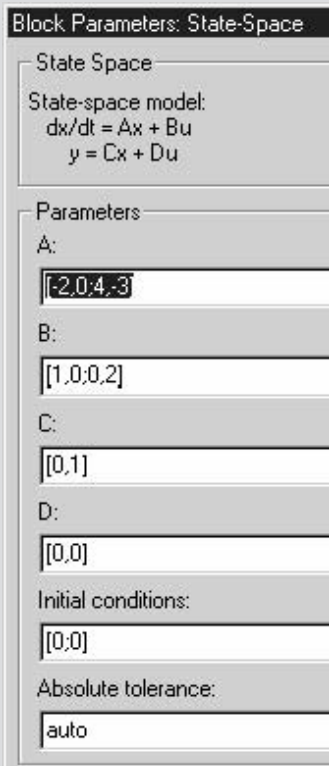
FIGURE 22–3
Simulink state-space menu block.
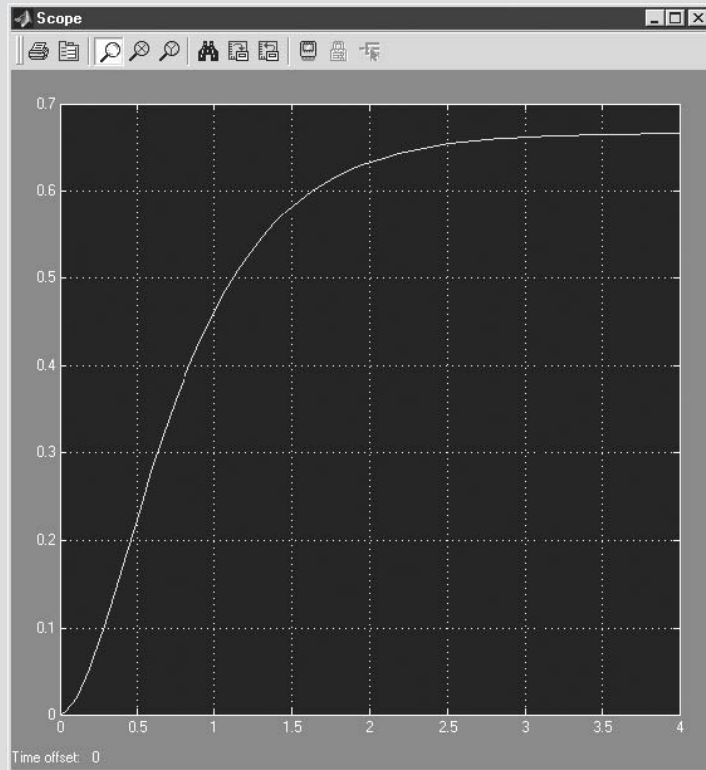
FIGURE 22–4
Simulink model output, Example 22.2.

The output from the model is also sent back to the MATLAB workspace, as *y,* where we can plot it in comparison to the analytical expression for h_2.

$$h_2(t) = \frac{2}{3} + \frac{4}{3}e^{-3t} - 2e^{-2t} = \frac{2}{3}\left(1 + 2e^{-3t} - 3e^{-2t}\right)$$

The MATLAB commands required to generate a comparison plot are

```
h2 = y.signals.values;
t = y.time;
h2analytical = (2/3)+(4/3)*exp(-3*t)-2*exp(-2*t);
plot(t,h2,t,h2analytical,'*')
```


Comparison of h_2 calculated with Simulink and the analytical expression
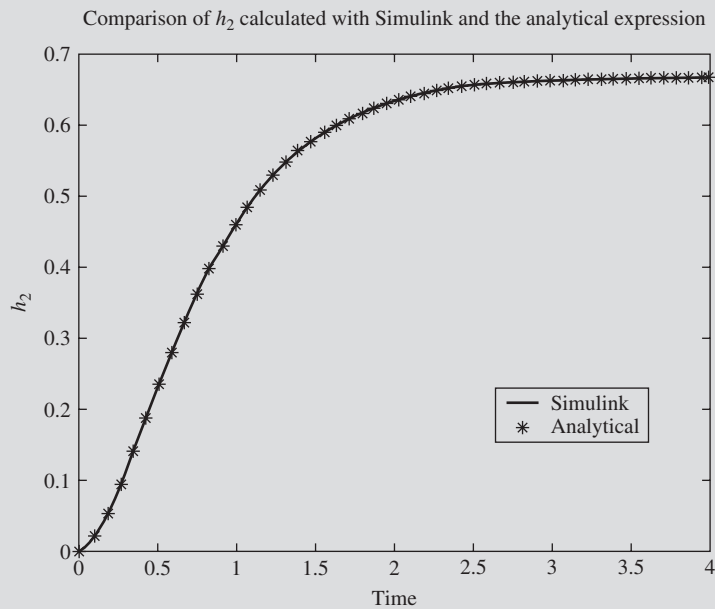
FIGURE 22–5
Comparison of h_2 calculated using Simulink with analytical solution.

From Fig. 22–5, we note that the solutions are identical.

SUMMARY

The matrix differential equation

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu}$$

used to describe a control system by the state-space method can be solved for the vector of state variables $\mathbf{x}$ by use of the transfer function matrix. It consists of a matrix of transfer functions that relate the state variables to the inputs. The transfer function matrix serves the same purpose in a multiple-input multiple-output system as the transfer function does for a single-input single-output system. The transfer function matrix is obtained from the matrix differential equation by application of Laplace transforms.

PROBLEMS

22.1. Determine $\mathbf{x}(t)$ for the system

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu}$$

where

$$e^{\mathbf{A}t} = \begin{bmatrix} e^{-5t} & -e^{-2t} + e^{-5t} \\ 0 & e^{-2t} \end{bmatrix}$$

$$\mathbf{x}(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad \mathbf{u}(t) = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 0 & -4 \end{bmatrix}$$

22.2 (*a*) Solve Prob. 21.1, using the techniques in this chapter.
   (*b*) Solve Prob. 21.1, using MATLAB/Simulink.

22.3 (*a*) Solve Prob. 21.2, using the techniques in this chapter.
   (*b*) Solve Prob. 21.2, using MATLAB/Simulink.

CAPSULE SUMMARY

For the general state-space model,

$$\dot{\mathbf{x}} \;=\; \mathbf{Ax} + \mathbf{Bu} \qquad \mathbf{y} \;=\; \mathbf{Cx} + \mathbf{Du}$$

The solution was shown to be

$$\mathbf{x}(t) \;=\; e^{\mathbf{A}t}\mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{Bu}(\tau)\,d\tau$$

where

$$e^{\mathbf{A}t} \;=\; L^{-1}\Big\{(s\mathbf{I} - \mathbf{A})^{-1}\Big\} \;\equiv\; \text{transition matrix}$$

and thus,

$$\int_0^t e^{\mathbf{A}t - \tau}\mathbf{Bu}(\tau)\,d\tau \;=\; L^{-1}\Big\{(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{BU}(s)\Big\}$$

If $\mathbf{x}(0) = 0$ (as is normally the case with deviation variables),

$$\mathbf{X}(s) \;=\; (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{BU}(s)$$

or

$$\mathbf{X}(s) \;=\; \mathbf{G}(s)\mathbf{U}(s)$$

where

$$\mathbf{G}(s) \;=\; (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} \;\equiv\; \text{transfer function matrix (relates } \mathbf{X} \text{ to } \mathbf{U})$$
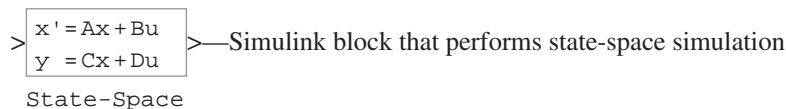
$$\mathbf{G}_1(s) \;=\; \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} \;\equiv\; \text{transfer function matrix (relates } \mathbf{Y} \text{ to } \mathbf{U})$$

USEFUL MATLAB COMMANDS

`expm(A)`—matrix exponential command.
`int`—performs a symbolic integration.
`A^ -1` — inverts matrix **A.**

> ┌─────────────┐
> │ x' = Ax + Bu │ >—Simulink block that performs state-space simulation
> │ y = Cx + Du │
> └─────────────┘
> `State-Space`

CHAPTER
23

MULTIVARIABLE CONTROL

U p to this point, the fundamentals of process dynamics and control have been illustrated by single-input single-output (SISO) systems. The processes encountered in the real world are usually multiple-input multiple-output (MIMO) systems. To explore these concepts, consider the interacting, two-tank liquid-level system in Fig. 23–1 where there is one input, the flow to tank 1 (m_1), and one output, the level in tank 2 (h_2). In this figure, h_2 is related to m_1 by a second-order transfer function. From the point of view of a SISO system, the relation between h_2 and m_1 may be represented by the block diagram in Fig. 23–1b. One may place a feedback control system around the open-loop system of Fig. 23–1b to maintain control of H_2.

(a)

(b)
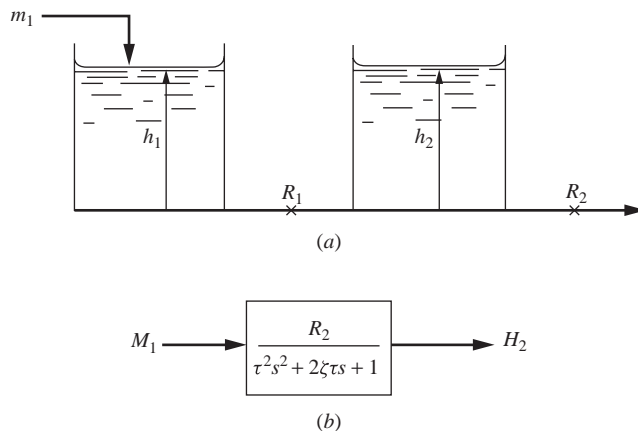
FIGURE 23–1
Single-input single-output (SISO) system: (a) two-tank interacting level system and (b) block diagram for SISO system.

Now consider the same process of Fig. 23–1 in which there are two inputs (m_1 and m_2) and two outputs (h_1 and h_2). This system is shown in Fig. 23–2a. A change in m_1 alone will affect both outputs (h_1 and h_2). A change in m_2 alone will also change both outputs. (Remember that this is an interacting process for which the level in tank 1 is affected by the level in tank 2.) The interaction between inputs and outputs can be seen more clearly by the block diagram of Fig. 23–2b. In this diagram, the transfer functions show how the change in one of the inputs affects both of the outputs. For example, if a change occurs in only M_1, the responses of H_1 and H_2 are

$$H_1(s) = G_{11}(s)M_1(s)$$
$$H_2(s) = G_{21}(s)M_1(s)$$

The transfer functions in Fig. 23–2b will be worked out for a specific set of process parameters in Example 23.1. (If the tanks were noninteracting, $G_{12} = 0$, with the result that a change in flow to tank 2 would not affect H_1.) If both H_1 and H_2 are to be controlled, a single control loop will not be sufficient; in this case two control loops are needed. The addition of control loops to the interacting system will be considered in Sec. 23.1.

(a)

(b)
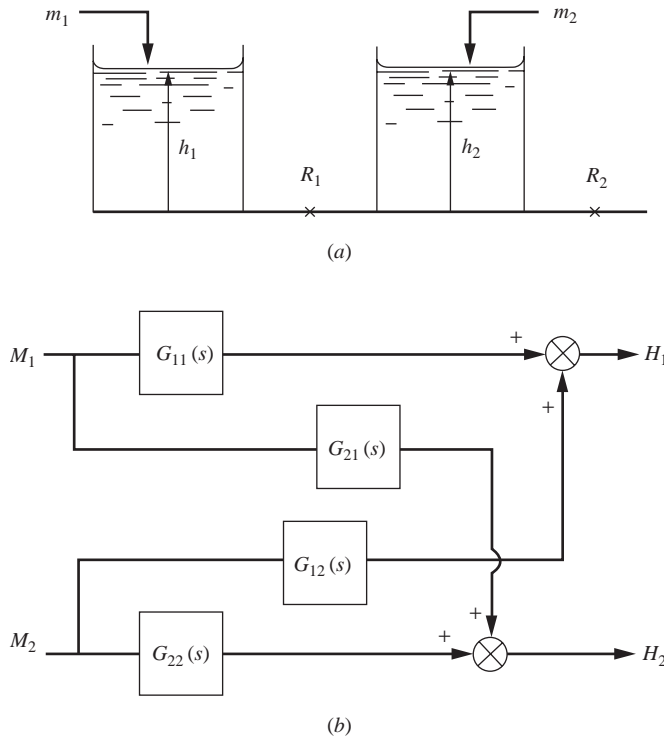
FIGURE 23–2
Multiple-input multiple-output (MIMO) system: (a) level process and (b) block diagram.

23.1 CONTROL OF INTERACTING SYSTEMS

The problem of controlling the outputs of a MIMO system will be discussed by means of a 2×2 system shown in Fig. 23–3. The problem can be extended to the case of more than two pairs of inputs and outputs by the same procedure described here. The control objective is to control C_1 and C_2 independently, in spite of changes in M_1 and M_2 or other load variables not shown. Two control loops are added to the diagram of Fig. 23–3 as shown in Fig. 23–4. Each loop has a block for the controller, the valve, and the measuring element. In principle, the multiloop control system of Fig. 23–4 will maintain control of C_1 and C_2. However, because of the interaction present in the system, a change in R_1 will also cause C_2 to vary because a disturbance enters the lower loop through the transfer function G_{21}. Because of interaction, both outputs (C_1 and C_2) will change if a change is made in either input alone. If G_{21} and G_{12} provide weak interaction, the two-controller scheme of Fig. 23–4 will give satisfactory control. In the extreme, if $G_{12} = G_{21} = 0$, we have no interaction and the two control loops are isolated from each other.
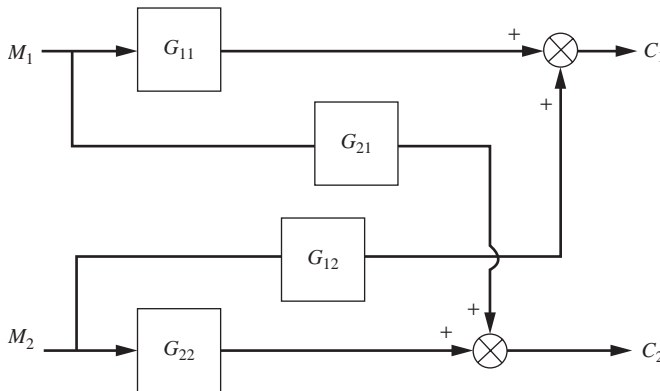
FIGURE 23–3
MIMO system for two
pairs of inputs and outputs
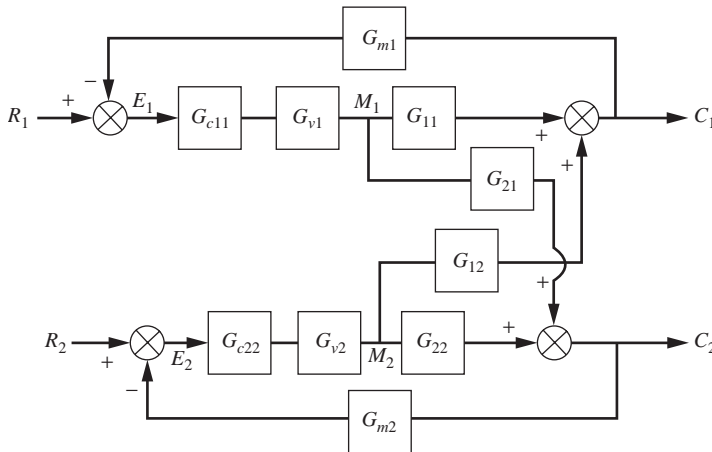
FIGURE 23–4
Multiloop control system with two controllers.

To completely eliminate the interaction between outputs and set points, two more controllers (cross-controllers) are added to the diagram of Fig. 23–4 to give the diagram shown in Fig. 23–5. In principle, these cross-controllers can be designed to eliminate interaction. The following analysis, which is expressed in matrix form, will lead to the method of design for cross-controllers that will eliminate interaction.

Response of Multiloop Control System

From Fig. 23–5, we may write by direct observation the following relationships in the form of the matrix expression

$$\mathbf{C} = \mathbf{G}_p\mathbf{M} \tag{23.1}$$

where $\mathbf{G}_p = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}$ $\mathbf{C} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}$ $\mathbf{M} = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix}$

We also may write from Fig. 23–5

$$M_1 = G_{v1}G_{c11}E_1 + G_{v1}G_{c12}E_2 \tag{23.2}$$

$$M_2 = G_{v2}G_{c21}E_1 + G_{v2}G_{c22}E_2 \tag{23.3}$$

where G_{v1} and G_{v2} are the transfer functions for the valves. Equations (23.2) and (23.3) may be written in matrix form as

$$\mathbf{M} = \mathbf{G}_v\mathbf{G}_c\mathbf{E} \tag{23.4}$$

where $\mathbf{G}_v = \begin{bmatrix} G_{v1} & 0 \\ 0 & G_{v2} \end{bmatrix}$ (valve matrix) $\mathbf{G}_c = \begin{bmatrix} G_{c11} & G_{c12} \\ G_{c21} & G_{c22} \end{bmatrix}$ (controller matrix)

$$\mathbf{E} = \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}$$
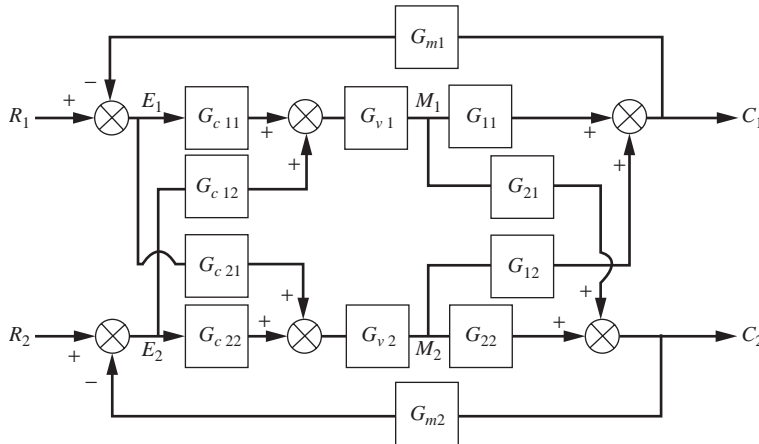
FIGURE 23–5
Multiloop control system with two primary controllers and two cross-controllers.

From Fig. 23–5, we write directly

$$E_1 = R_1 - G_{m1}C_1 \tag{23.5}$$

$$E_2 = R_2 - G_{m2}C_2 \tag{23.6}$$

where E_1 and E_2 are the error signals from the comparators. Equations (23.5) and (23.6) can be written in matrix form

$$\mathbf{E} = \mathbf{R} - \mathbf{G}_m\mathbf{C} \tag{23.7}$$

where $\mathbf{G}_m = \begin{bmatrix} G_{m1} & 0 \\ 0 & G_{m2} \end{bmatrix}$ (measuring element matrix)

$$\mathbf{R} = \begin{bmatrix} R_1 \\ R_2 \end{bmatrix}$$

From Eqs. (23.1) and (23.4), we obtain

$$\mathbf{C} = \mathbf{G}_p\mathbf{G}_v\mathbf{G}_c\mathbf{E} \tag{23.8}$$

If we let $\mathbf{G}_o = \mathbf{G}_p\,\mathbf{G}_v\,\mathbf{G}_c$, Eq. (23.8) becomes

$$\mathbf{C} = \mathbf{G}_o\mathbf{E} \tag{23.9}$$

Combining Eqs. (23.7) and (23.9) gives

$$\mathbf{C} = \mathbf{G}_o\mathbf{R} - \mathbf{G}_o\mathbf{G}_m\mathbf{C} \tag{23.10}$$

We may now solve Eq. (23.10) for $\mathbf{C}$ to obtain

$$\mathbf{C} = [\mathbf{I} + \mathbf{G}_o\mathbf{G}_m]^{-1}\mathbf{G}_o\mathbf{R} \tag{23.11}$$

Notice that the closed-loop behavior expressed by this matrix equation is analogous to the closed-loop response of a SISO system, which may be written

$$C(s) = \frac{G_o(s)}{1 + G_o(s)G_m(s)}R(s) \tag{23.12}$$

The matrix term $(\mathbf{I} + \mathbf{G}_o\,\mathbf{G}_m)^{-1}$ is equivalent to the scalar term $1/[1 + G_o(s)G_m(s)]$.
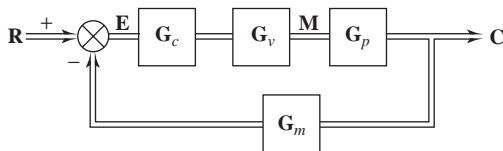
FIGURE 23–6
Block diagram for MIMO control system in terms of matrix blocks.
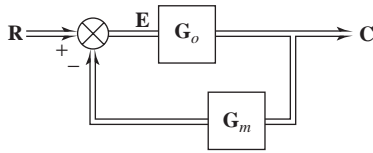
FIGURE 23–7
Reduced diagram for MIMO control
system where $\mathbf{G}_o = \mathbf{G}_p \mathbf{G}_v \mathbf{G}_c$.

A block diagram equivalent to the diagram for the MIMO control system in Fig. 23–5 is shown in Fig. 23–6. In this diagram, the blocks are filled with the matrices in Eqs. (23.1), (23.4), and (23.7). The double line indicates that more than one variable is being transmitted. Each block contains a matrix of transfer functions that relates an output vector to an input vector. The diagram can be simplified by multiplying the three matrices in the forward loop together and calling the result $\mathbf{G}_o$, as was done to obtain Eq. (23.9). The simplified diagram is shown in Fig. 23–7.

Noninteracting Control

In order for no interaction to occur between $\mathbf{C}$ and $\mathbf{R}$ in Fig. 23–5 (i.e., R_1 affects only C_1 and R_2 affects only C_2), the off-diagonal elements of $[\,\mathbf{I} + \mathbf{G}_o \mathbf{G}_m\,]^{-1} \mathbf{G}_o$ in Eq. (23.11) must be zero. Since $\mathbf{I}$ and $\mathbf{G}_m$ are diagonal, $[\,\mathbf{I} + \mathbf{G}_o \mathbf{G}_m\,]^{-1} \mathbf{G}_o$ will be diagonal if $\mathbf{G}_o$ is diagonal. Multiplication of the matrices in the expression for $\mathbf{G}_o$ is now shown:

$$\mathbf{G}_o = \mathbf{G}_p \mathbf{G}_v \mathbf{G}_c$$

$$\mathbf{G}_o = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \begin{bmatrix} G_{v1} & 0 \\ 0 & G_{v2} \end{bmatrix} \begin{bmatrix} G_{c11} & G_{c12} \\ G_{c21} & G_{c22} \end{bmatrix}$$

The result of multiplying these matrices gives

$$\mathbf{G}_o = \begin{bmatrix} G_{11}G_{v1}G_{c11} + G_{12}G_{v2}G_{c21} & G_{11}G_{v1}G_{c12} + G_{12}G_{v2}G_{c22} \\ G_{21}G_{v1}G_{c11} + G_{22}G_{v2}G_{c21} & G_{21}G_{v1}G_{c12} + G_{22}G_{v2}G_{c22} \end{bmatrix} \quad (23.13)$$

Setting the off-diagonal elements to zero and solving for G_{c12} and G_{c21} give

$$G_{c12} = -\frac{G_{12}G_{v2}G_{c22}}{G_{11}G_{v1}} \quad (23.14)$$

$$G_{c21} = -\frac{G_{21}G_{v1}G_{c11}}{G_{22}G_{v2}} \quad (23.15)$$

Example 23.1 will give some experience with the computations involved in applying the theory developed so far in this chapter.

Example 23.1.   For the two-tank, interacting liquid-level system shown in Fig. 23–8, develop the block diagram for a MIMO system corresponding to Fig. 23–3.

Material balances around tank 1 and tank 2 give the following differential equations:

$$A_1 \dot{c}_1 = m_1 - \frac{c_1 - c_2}{R_1} - \frac{c_1}{R_3} \quad (23.16)$$

$$A_2 \dot{c}_2 = m_2 + \frac{c_1 - c_2}{R_1} - \frac{c_2}{R_2} \quad (23.17)$$
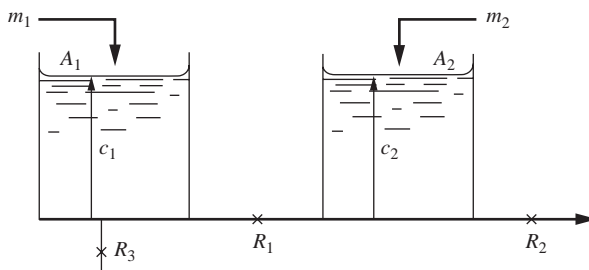
FIGURE 23–8
Process for Example 23.1: $A_1 = 1, A_2 = \frac{1}{2}, R_1 = \frac{1}{2}, R_2 = 2, R_3 = 1$.

Introducing the parameters given in Fig. 23–8 into Eqs. (23.16) and (23.17) gives

$$\dot{c}_1 = m_1 - 3c_1 + 2c_2 \tag{23.18}$$

$$\dot{c}_2 = 2m_2 + 4c_1 - 5c_2 \tag{23.19}$$

These equations may be written in matrix form as

$$\dot{\mathbf{c}} = \mathbf{Ac} + \mathbf{Bm}$$

where

$$\mathbf{A} = \begin{bmatrix} -3 & 2 \\ 4 & -5 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

We use Eq. (22.15) to obtain

$$\mathbf{C}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{BM}(s) \tag{23.20}$$

Writing Eq. (23.20) in the form of Eq. (23.1) gives

$$\mathbf{C} = \mathbf{G}_p\mathbf{M}$$

where $\mathbf{G}_p = (s\,\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}$.

After several steps involving the inversion of $s\,\mathbf{I} - \mathbf{A}$ and multiplying the result of inversion by $\mathbf{B}$, we get

$$G_p = \frac{\begin{bmatrix} s+5 & 4 \\ 4 & 2(s+3) \end{bmatrix}}{(s+1)(s+7)} \tag{23.21}$$

The block diagram can now be drawn as shown in Fig. 23–3 with

$$G_{11} = \frac{s+5}{(s+1)(s+7)} \qquad G_{12} = \frac{4}{(s+1)(s+7)}$$

$$G_{21} = \frac{4}{(s+1)(s+7)} \qquad G_{22} = \frac{2(s+3)}{(s+1)(s+7)}$$

Notice that the diagonal elements of $\mathbf{G}_p(s)$ are of the form

$$\frac{\alpha(s + \beta)}{(s + 1)(s + 7)}$$

(a)

(b)
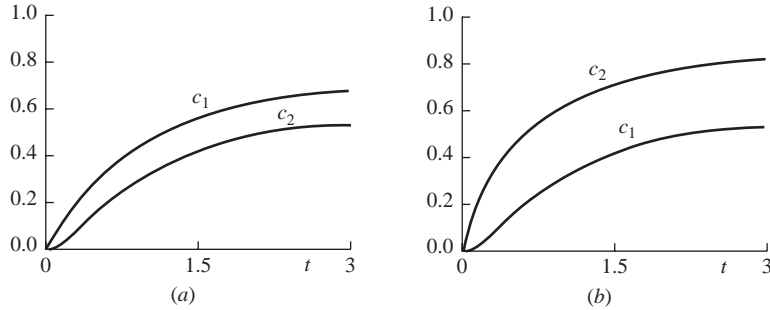
FIGURE 23–9
Open-loop response for Example 23.1. (a) $M_1 = \dfrac{1}{s}$, $M_2 = 0$ (step change to tank 1) and
(b) $M_1 = 0$, $M_2 = \dfrac{1}{s}$ (step change to tank 2).

These elements, which relate c_1 to m_1 and c_2 to m_2, will produce a second-order response to a step change in input that has a finite slope at the origin because of the numerator term $s + \beta$. In contrast, the off-diagonal elements have second-order transfer functions without numerator dynamics, for which case the step response will be second-order with zero slope at the origin. The responses of c_1 and c_2 for unit-step changes in m_1 and m_2 taken separately are shown in Fig. 23–9.

Example 23.2.   For the two-tank liquid-level system of Example 23.1, determine the controller transfer function matrix $\mathbf{G}_c$ needed to eliminate interaction. The primary controllers are to be proportional; i.e., $G_{c11} = K_1$ and $G_{c22} = K_2$. The diagram of the control system is shown in Fig. 23–10.
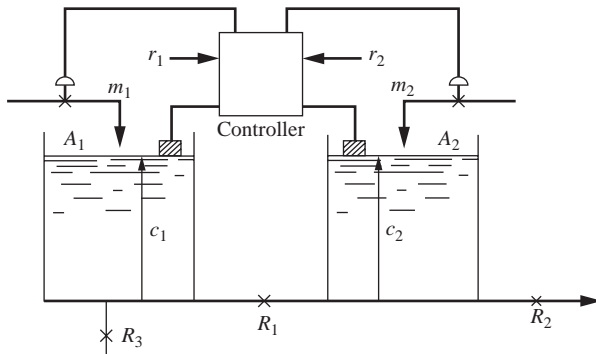
FIGURE 23–10
Process for Example 23.2. $A_1 = 1, A_2 = \frac{1}{2}, R_1 = \frac{1}{2}, R_2 = 2, R_3 = 1, G_{c11} = K_1, G_{c22} = K_2.$

The block labeled Controller contains the four transfer functions that are the elements of $\mathbf{G}_c$. In this problem, $\mathbf{G}_v$ is a unit diagonal matrix, i.e., $G_{v1} = G_{v2} = 1$.

From Eqs. (23.14) and (23.15) we obtain

$$G_{c12} = -\frac{G_{12}G_{c22}}{G_{11}} = -\frac{4}{(s+1)(s+7)}K_2\frac{(s+1)(s+7)}{s+5}$$

or
$$G_{c12} = \frac{-4K_2}{s+5} \tag{23.22}$$

$$G_{c21} = -\frac{G_{21}G_{c11}}{G_{22}} = -\frac{4}{(s+1)(s+7)}K_1\frac{(s+1)(s+7)}{2(s+3)}$$

or
$$G_{c21} = \frac{-2K_1}{s+3} \tag{23.23}$$

Having found the transfer functions for the cross-controllers, we can now determine the nature of the uncoupled response of c_1 to a change in r_1 and of c_2 to a change in r_2.

Inserting $G_{v1} = G_{v2} = 1$ and the expressions for G_{c12} and G_{c21} from Eqs. (23.14) and (23.15) into Eq. (23.13) gives for $\mathbf{G}_o$

$$\mathbf{G}_o = \begin{bmatrix} G_{11}G_{c11} + G_{12}G_{c21} & 0 \\ 0 & G_{21}G_{c12} + G_{22}G_{c22} \end{bmatrix} \tag{23.24}$$

Inserting the appropriate elements of the $\mathbf{G}_p$ matrix [Eq. (23.21)] and the $\mathbf{G}_c$ matrix in Eq. (23.24) gives, after considerable simplification,

$$\mathbf{G}_o = \begin{bmatrix} \dfrac{K_1}{s+3} & 0 \\ 0 & \dfrac{2K_2}{s+5} \end{bmatrix} \qquad \text{decoupled system} \tag{23.25}$$

The block diagram for this decoupled MIMO system is shown in Fig. 23–11.
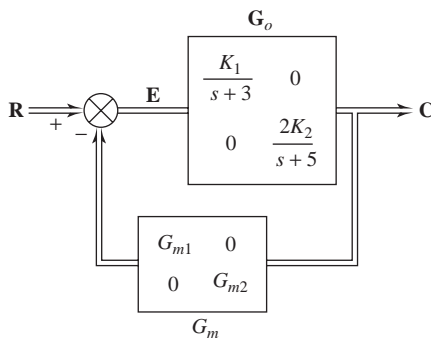
FIGURE 23–11
Block diagram for decoupled system in Example 23.2.

Assuming that the measurement matrix $\mathbf{G}_m$ is a unit diagonal matrix, the diagram in Fig. 23–11 can be simplified to the unity feedback diagram of Fig. 23–12.
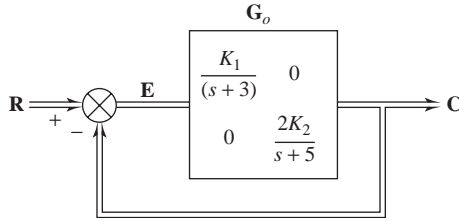
FIGURE 23–12
Simplified block diagram for Example 23.2.

From Fig. 23–12, we may write directly

$$\mathbf{C} = \mathbf{G}_o\mathbf{E} \qquad \mathbf{E} = \mathbf{R} - \mathbf{C}$$

therefore, $\mathbf{C} = \mathbf{G}_o\mathbf{R} - \mathbf{G}_o\mathbf{C}$

or
$$\begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} G_{o11} & 0 \\ 0 & G_{o22} \end{bmatrix}\begin{bmatrix} R_1 \\ R_2 \end{bmatrix} - \begin{bmatrix} G_{o11} & 0 \\ 0 & G_{o22} \end{bmatrix}\begin{bmatrix} C_1 \\ C_2 \end{bmatrix}$$

From this expression, we may write

$$C_1 = G_{o11}R_1 - G_{o11}C_1$$
$$C_2 = G_{o22}R_2 - G_{o22}C_2$$

Solving for $C_1(s)$ gives

$$C_1(s) = \frac{G_{o11}}{1 + G_{o11}} R_1(s)$$

Inserting G_{o11} from Eq. (23.25) gives

$$C_1(s) = \frac{K_1/(s+3)}{1 + K_1/(s+3)} R_1(s) \tag{23.26}$$

In a similar way, one can show that

$$C_2(s) = \frac{2K_2/(s+5)}{1 + 2K_2/(s+5)} R_2(s) \tag{23.27}$$

The result shows that the cross-controllers of Eqs. (23.22) and (23.23) give two separate noninteracting control loops, as shown in Fig. 23–13.

The response of the control system of Fig. 23–10 is shown in Fig. 23–14 for a unit-step change in R_1. In Fig. 23–14a, no cross-controllers are present in the matrix $\mathbf{G}_c$. In Fig. 23–14b, cross-controllers having the transfer functions given by Eqs. (23.22) and (23.23) are present. As expected, for the case of no cross-controllers, one sees from Fig. 23–14a that a request for a unit-step change in r_1 causes both c_1 and c_2 to change. For the case where cross-controllers are present, we see from Fig. 23–14b that a change in r_1 does not affect c_2 as demanded by a decoupled system.
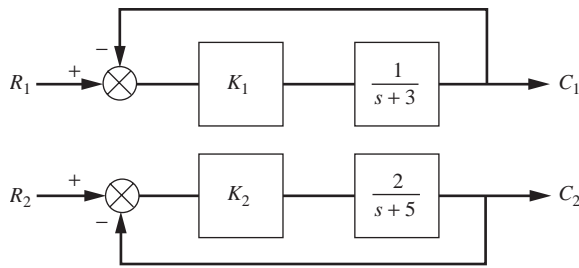
FIGURE 23–13
Decoupled control system for Example 23.2 where primary
controllers are proportional.
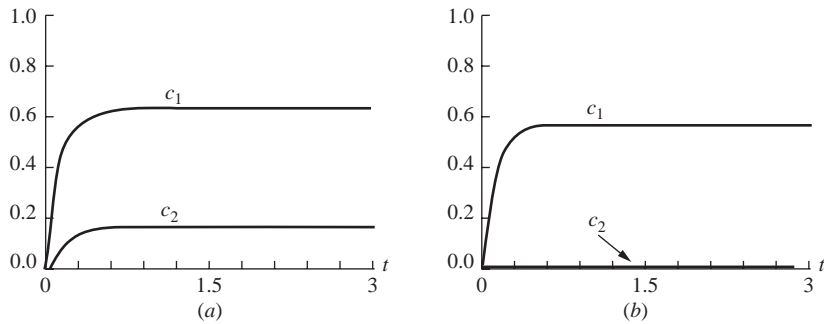
FIGURE 23–14
Response for control system in Example 23.2 for $R_1 = 1/s$, $R_2 = 0$, $G_{c11} = K_1 = 4$, $G_{c22} = K_2 = 4$.
(*a*) No cross-controllers; (*b*) cross-controllers present.

Example 23.3. Use MATLAB and Simulink to simulate the system shown in
Example 23.2. Examine the system behavior for values of K_c (both K_1 and K_2)
= 2, 4, 6, 8, and 10.

The Simulink diagram for the process is shown in Fig. 23–5.
The PID controller block for each loop is a proportional controller only, $G_{c11} = K_1$
and $G_{c22} = K_2$.

We use a MATLAB m-file to call this model, using the following command:
```
>>ex23_3driver
```

The MATLAB m-file that calls the Simulink model is
```
% The name of this file is ex23_3driver.m
% The Simulink file for the model is example23_3
% The variable h is used to plot each line in different
  colors
h(1,:)='g-';
h(2,:)='r-';
h(3,:)='b-';
```
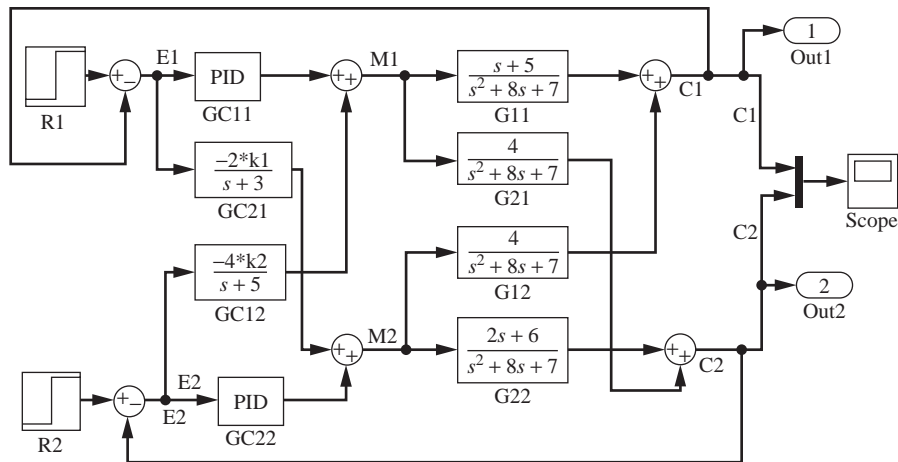
FIGURE 23–15
Simulink model for simulation of Example 23.2.

```
h(4,:)='k-';
h(5,:)='g--';
h(6,:)='r--';
for i=1:5
   K1=2*i;
   K2=2*i;
% The variable y in the 'sim' statement is taken from the
  outport in the model
  [t,x,y] = sim('example23_3',3);
  plot(t,y(:,1),h(i,:))
  hold on
  plot(t,y(:,2),h(i+1,:))
end
grid
title('c_1 and c_2 vs time for K_c=2,4,6,8,10');
grid
hold off
```

The output from the m-file is the desired plot shown in Fig. 23–16. (Note that
the text labels for the lines were added afterward using MATLAB plot edit-
ing commands by double-clicking the graph and filling in the appropriate menu
boxes.)

The MATLAB output shows that as we increase the controller gain,
the height in tank 1 more nearly tracks the set point change (as we might
expect), and the height in tank 2 is unaffected because of the decoupling cross-
controllers.
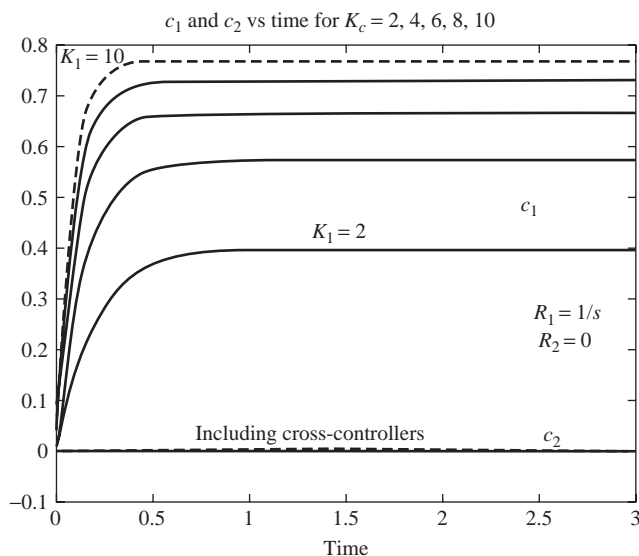
FIGURE 23–16
Solution of Example 23.2 using MATLAB/Simulink.

To completely avoid the offset associated with proportional control, we can use PI controllers for the primary controllers for the decoupled system. To study the effect of PI controllers for the decoupled system, let

$$G_{c11} = K_1\left(1 + \frac{1}{s}\right) \qquad \text{and} \qquad G_{c22} = K_2\left(1 + \frac{1}{s}\right)$$

Note that $\tau_I = 1$ for both controllers. For this case, the cross-controller transfer functions may be obtained from Eqs. (23.14) and (23.15); the results are

$$G_{c12} = \frac{-4K_2(s + 1)}{s(s + 5)} \qquad \text{and} \qquad G_{c21} = \frac{-2K_1(s + 1)}{s(s + 3)}$$

A simulation using these four controller transfer functions with $K_1 = K_2 = 4$ is shown in Fig. 23–17. From the transient response, we see that c_1 moves toward the set point of 1.0 and that c_2 does not change, as is expected for a decoupled system.
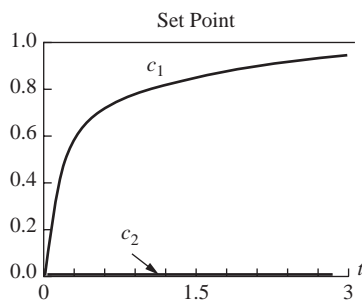
FIGURE 23–17
Response of decoupled control system in Example 23.2 for PI primary controllers: $G_{c11} = G_{c22} = 4(1 + 1/s)$, $R_1 = 1/s$, $R_2 = 0$.

23.2   STABILITY OF MULTIVARIABLE SYSTEMS

Determining the stability for a multivariable control system, such as the one in Fig. 23–4 or Fig. 23–5, can be much more complicated than for a SISO system. The transfer function for the closed-loop response of a MIMO system is given by Eq. (23.11):

$$\mathbf{C} = [\mathbf{I} + \mathbf{G}_o\mathbf{G}_m]^{-1}\mathbf{G}_o\mathbf{R}$$

To invert this expression, we write

$$\mathbf{C} = \frac{\text{adj } [\mathbf{I} + \mathbf{G}_o\mathbf{G}_m]\mathbf{G}_o\mathbf{R}}{|\mathbf{I} + \mathbf{G}_o\mathbf{G}_m|} \tag{23.28}$$

The numerator of this expression is an $n \times n$ matrix; the denominator is a nth-order polynomial. To simplify the following argument, let the matrix in Eq. (23.28) be 2×2. Let the elements of the numerator, after expansion, be written as follows:

$$\text{adj}[\mathbf{I} + \mathbf{G}_o\mathbf{G}_m]\mathbf{G}_o\mathbf{R} = \begin{bmatrix} \beta_{11}(s) & \beta_{12}(s) \\ \beta_{21}(s) & \beta_{22}(s) \end{bmatrix} \tag{23.29}$$

Let the elements of $\mathbf{G}_o$ $\mathbf{G}_m$ be written as follows:

$$\mathbf{G}_o\mathbf{G}_m = \begin{bmatrix} \alpha_{11}(s) & \alpha_{12}(s) \\ \alpha_{21}(s) & \alpha_{22}(s) \end{bmatrix} \tag{23.30}$$

Expansion of the determinant in Eq. (23.28), using Eq. (23.30), is

$$|\mathbf{I} + \mathbf{G}_o\mathbf{G}_m| = \begin{vmatrix} 1 + \alpha_{11}(s) & \alpha_{12}(s) \\ \alpha_{21}(s) & 1 + \alpha_{22}(s) \end{vmatrix}$$

or

$$|\mathbf{I} + \mathbf{G}_o\mathbf{G}_m| = [1 + \alpha_{11}(s)][1 + \alpha_{22}(s)] - \alpha_{12}(s)\alpha_{21}(s) \tag{23.31}$$

Equation (23.31) is a polynomial expression, for which the order will depend on the order of the transfer functions in $\mathbf{G}_o$ and $\mathbf{G}_m$. Equation (23.28) can now be written in terms of the expansions shown in Eqs. (23.29) and (23.31) as follows:

$$\mathbf{C} = \begin{bmatrix} \dfrac{\beta_{11}(s)}{|\mathbf{I} + \mathbf{G}_o\mathbf{G}_m|} & \dfrac{\beta_{12}(s)}{|\mathbf{I} + \mathbf{G}_o\mathbf{G}_m|} \\ \dfrac{\beta_{21}(s)}{|\mathbf{I} + \mathbf{G}_o\mathbf{G}_m|} & \dfrac{\beta_{22}(s)}{|\mathbf{I} + \mathbf{G}_o\mathbf{G}_m|} \end{bmatrix}$$

Since each term contains the polynomial $|\mathbf{I} + \mathbf{G}_o$ $\mathbf{G}_m|$ in the denominator, the stability of the multivariable system will depend on the roots of the polynomial equation

$$|\mathbf{I} + \mathbf{G}_o\mathbf{G}_m| = 0 \qquad \text{characteristic equation} \tag{23.32}$$

Equation (23.32) is the characteristic equation of the multivariable system. Although Eq. (23.32) has been derived here for the case where $\mathbf{G}_o \, \mathbf{G}_m$ is a 2×2 matrix, one can show that Eq. (23.32) applies to the general MIMO system of Fig. 23–7 in which $\mathbf{G}_o \, \mathbf{G}_m$ is a matrix of any size ($n \times n$). If the roots of the characteristic equation are in the left half of the complex plane, we know that the system is stable. One method to be used for examining the stability of a multivariable system is to apply the Routh test to the characteristic equation of Eq. (23.32). In practice, the characteristic equation can be of high order for a simple 2×2 multivariable control system. Example 23.4 illustrates the determination of stability for a multivariable control system.

Example 23.4.   For the control system of Example 23.2, which is shown in Fig. 23–10, determine stability for the case where $G_{c11} = K_1$, $G_{c22} = K_2$, and there are no cross-controllers present (i.e., $G_{c12} = G_{c21} = 0$); also let $\mathbf{G}_m$ and $\mathbf{G}_v$ be unit matrices. From Example 23.1, we have for the elements of $\mathbf{G}_p$

$$G_{11} = \frac{s + 5}{(s + 1)(s + 7)} \qquad G_{12} = \frac{4}{(s + 1)(s + 7)}$$

$$G_{21} = \frac{4}{(s + 1)(s + 7)} \qquad G_{22} = \frac{2(s + 3)}{(s + 1)(s + 7)}$$

Since $\mathbf{G}_v = \mathbf{I}$, $\mathbf{G}_o = \mathbf{G}_p \, \mathbf{G}_c$. Since $\mathbf{G}_m = \mathbf{I}$, the characteristic equation of Eq. (23.32) can now be written as

$$\left| \mathbf{I} + \mathbf{G}_p \mathbf{G}_c \right| = 0 \tag{23.33}$$

Introducing the elements of the matrices $\mathbf{G}_p$ and $\mathbf{G}_c$ into Eq. (23.33) gives, after expansion of the determinant,

$$\left[(s + 1)(s + 7) + K_1(s + 5) \right]\left[(s + 1)(s + 7) + 2K_2(s + 3) \right] - 16K_1K_2 = 0$$

For given values of K_1 and K_2, this expression can be expanded into a fourth-order polynomial equation of the form

$$s^4 + \alpha s^3 + \beta s^2 + \gamma s + \Delta = 0 \tag{23.34}$$

where α, β, γ, and Δ will include the gains K_1 and K_2.

The Routh test can be applied to Eq. (23.34) to determine whether the system is stable. From this simple example, the reader can appreciate the algebraic tedium that may be needed to determine the stability of a multivariable system.

One way to express the stability of this system is to plot the stability boundaries on a graph of K_1 versus K_2. The region within the boundaries gives the combinations of values of K_1 and K_2 for which the system is stable. Since the details of stability boundaries are beyond the scope of this chapter, the reader may consult Seborg, Edgar, and Mellichamp (2004) for examples of stability boundaries for multivariable systems.

SUMMARY

Most of the systems encountered are multiple-input multiple-output (MIMO) systems. Such systems have several inputs and several outputs that are often interacting, meaning that a disturbance at any input causes a response in some of or all the outputs. This interaction in a MIMO system makes control and stability analysis of the system very complicated compared to that of a single-input single-output (SISO) system. A convenient way to describe a MIMO system is by means of a block diagram in which each block contains a matrix of transfer functions that relates an input vector to an output vector.

It is often desirable to have a control system decoupled so that certain outputs can be controlled independently of other outputs. A systematic procedure was described for decoupling a control system by including cross-controllers along with the principal controllers. This approach to decoupling requires an accurate model of the system; the number of controllers (principal controllers and cross-controllers) increases rapidly with the number of inputs and outputs. A system represented by two inputs and two outputs requires as many as four controllers; a system of three inputs and three outputs requires as many as nine controllers; and so on.

The characteristic equation for a multivariable control system, from which one can determine stability by examining its roots, can be of high order for a relatively simple system. Expressing stability boundaries in terms of controller parameters becomes complex because of the large number of controller parameters that can be adjusted.

PROBLEMS

23.1. For the liquid-level system shown in Fig. P23–1 determine the cross-controller transfer functions that will decouple the system. Fill in each block of the diagram shown in Fig. 23–5 with a transfer function obtained from an analysis of the control system. The transfer function for each feedback measuring element is unity. The following data apply:

$$A_1 = 1 \quad A_2 = 0.5 \quad Res_1 = 0.5 \quad Res_2 = \tfrac{2}{3} \quad G_{c11} = K_1 \quad G_{c22} = K_2$$
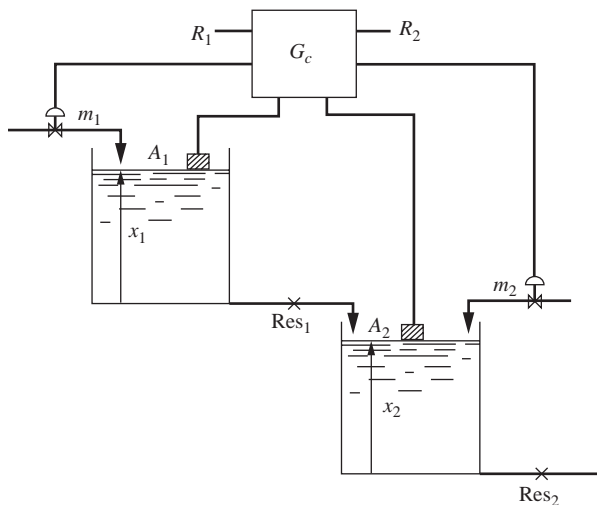
FIGURE P23–1

The resistance on the outlet of a tank has been denoted by Res to avoid confusion with the symbol for set point (R).

23.2. *(a)* For the interacting liquid-level system shown in Fig. P23–2, draw *very neatly* a block diagram that corresponds to Fig. 23–4. Each block should contain a transfer function obtained from an analysis of the liquid-level system. There are no cross-controllers in this system. The transfer function for each feedback element is unity. The following data apply:

$$A_1 = 1 \qquad A_2 = \tfrac{1}{2} \qquad \text{Res}_1 = \tfrac{1}{2} \qquad \text{Res}_2 = 2 \qquad \text{Res}_3 = 1$$
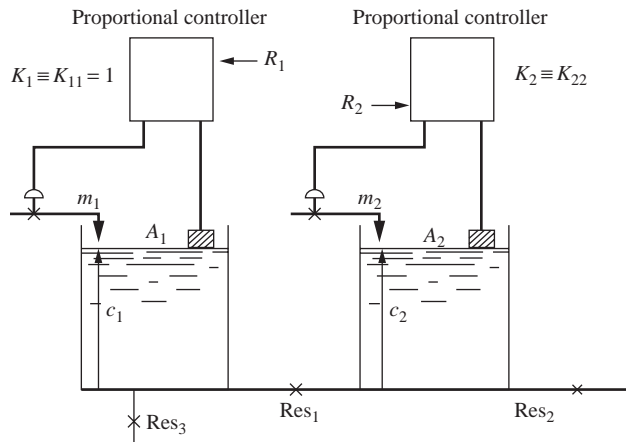
FIGURE P23–2

(b) Obtain the characteristic equation of this system in the form

$$s^n + \alpha s^{n-1} + \beta s^{n-2} + \cdots = 0$$

Obtain expressions for α, β, etc., in terms of K_2 ($K_1 = 1$).

(c) How would you determine stability limits for this interacting control system?

23.3. Modify the Simulink model from Example 23.3 to simulate a PI controller for the process and verify Fig. 23–17.

CAPSULE SUMMARY

SISO = single-input single-output system

MIMO = multiple-input multiple-output system. Most real-life systems are MIMO.

In MIMO systems, the variables can interact, making control difficult. A change in one variable by a controller attempting to drive the process variable to its desired set point can cause unwanted disturbances in other process variables.

Decoupling process variable interactions is possible by introducing appropriate cross-controllers.

MATLAB and Simulink are useful for simulating MIMO systems.

PART
VII

NONLINEAR CONTROL

CHAPTER
24

EXAMPLES OF NONLINEAR SYSTEMS

In previous chapters, we have confined our attention to the behavior of linear systems or to the analysis of linearized equations representative of nonlinear systems in the vicinity of the steady-state condition. While much useful information can be obtained from such analysis, it frequently is desirable or necessary to consider nonlinearities in control system design.

No real physical system is truly linear, particularly over a wide range of operating variables. Hence, to be complete, a control system design should allow for the possibility of a large deviation from steady-state behavior and resulting nonlinear behavior. The purpose of Chaps. 24 and 25 is to introduce some of the tools that can be used for this purpose and to indicate some of the complications that arise when nonlinear systems are considered.

24.1 DEFINITION OF A NONLINEAR SYSTEM

A nonlinear system is one for which the principle of superposition does not apply. Thus, by superposition, the response of a linear system to the sum of two inputs is the same as the sum of the responses to the individual inputs. This behavior, which allows us to characterize completely a linear system by a transfer function, is not true of nonlinear systems.

As an example, consider a liquid-level system. If the outflow is proportional to the square root of the tank level, superposition does not hold and the system is nonlinear. If the tank will always operate near the steady-state condition, the square root behavior may be adequately represented by a straight line and superposition applied, as we have done before. On the other hand, if the tank level were to fall to one-half the steady-state value, we would no longer expect the transfer function derived on the linearized basis to apply. The analysis becomes more complicated, as we shall see in our introduction to the study of nonlinear systems.

24.2 THE PHASE PLANE

The analysis of nonlinear dynamic systems may often be conceptually simplified by changing to a coordinate system known as *phase space.* In this coordinate system, time no longer appears explicitly, it being replaced by some other property of the system. For example, consider the flight of a rocket to the moon. In a grossly oversimplified manner, we may describe this motion by a plot of the distance of the rocket from the moon versus time. If all goes well, we would like such a plot to resemble Fig. 24–1. Note the initial acceleration during launch and the final deceleration at landing. We may, however, also represent this motion by a plot of rocket velocity versus distance from the moon. This plot is shown in Fig. 24–2, where velocity is defined as *d* (distance from moon)/*dt.* Figure 24–2 is called a *phase diagram* of the rocket motion. Time now appears merely as a parameter along the curve of the rocket motion. It has been replaced as a coordinate by the rocket velocity. Although in the present example Fig. 24–2 may not be of significant advantage over Fig. 24–1, we shall find phase diagrams very helpful in the analysis of certain nonlinear control systems.
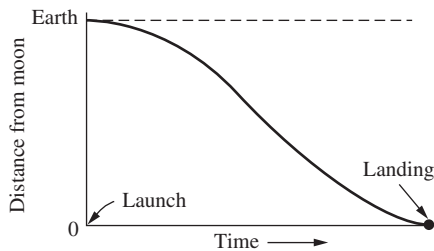
FIGURE 24–1
Distance-time plot for moon rocket.
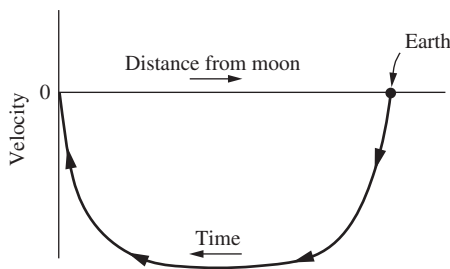
FIGURE 24–2
Velocity-distance plot for moon rocket.

To begin our study of phase diagrams, we convert the second-order system studied previously in Chap. 7 to the phase plane.

24.3   PHASE-PLANE ANALYSIS OF DAMPED OSCILLATOR

The differential equation describing the motion of the system of Fig. 7–1 in response to a unit-step function is

$$\tau^2 \frac{d^2Y}{dt^2} + 2\zeta\tau \frac{dY}{dt} + Y = 1 \tag{24.1}$$

Equation (24.1) has previously been solved to yield the motion in the form of $Y(t)$ versus t as shown in Fig. 7–3. For phase analysis, however, we want the motion in terms of velocity versus position, $\dot{Y}$ versus Y, where the dot notation is used to indicate differentiation with respect to t. Hence, we rewrite Eq. (24.1) as

$$\frac{dY}{dt} = \dot{Y}$$

$$\frac{d\dot{Y}}{dt} = \frac{-Y - 2\zeta\tau\dot{Y} + 1}{\tau^2} \tag{24.2}$$

It is sometimes convenient in phase-plane analysis to write the variables in terms of deviation about the *final condition*. In this case, the system will ultimately come to rest at $Y = 1$ (because both time derivatives $\dfrac{dy}{dt}$ and $\dfrac{d^2y}{dt^2}$ equal 0 at steady state). Hence we define

$$X = Y - 1$$

$$\dot{X} = \dot{Y}$$

Then Eq. (24.2) becomes

$$\frac{dX}{dt} = \dot{X}$$

$$\frac{d\dot{X}}{dt} = \frac{-X - 2\zeta\tau\dot{X}}{\tau^2} \tag{24.3}$$

These are now viewed as two simultaneous, first-order differential equations in the variables X and $\dot{X}$ (which, we recall, are phase variables).

To solve Eqs. (24.3), we may use the methods presented in Chaps. 21 and 22. For this purpose, let $X_1 = X$ and $X_2 = \dot{X}$. Equations (24.3) may be written in the form

$$\begin{bmatrix} \dot{X}_1 \\ \dot{X}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \dfrac{-1}{\tau^2} & \dfrac{-2\zeta}{\tau} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

or
$$\dot{\mathbf{X}} = \mathbf{AX} \tag{24.4}$$

where $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ \dfrac{-1}{\tau^2} & \dfrac{-2\zeta}{\tau} \end{bmatrix}$

Equation (24.4) is in the standard form of a matrix differential equation [Eq. (21.7)]. Notice that the term $\mathbf{b}u$ of Eq. (21.7) is not present because no forcing term is present in Eqs. (24.3). Equation (24.4) may be solved by use of Eq. (22.7):

$$\mathbf{X}(t) = e^{\mathbf{A}t}\mathbf{X}(0) \tag{22.7}$$

where

$$e^{\mathbf{A}t} = L^{-1}\left\{(s\mathbf{I} - \mathbf{A})^{-1}\right\} \tag{22.13}$$

Following the usual steps required to solve these equations gives the result

$$X_1 = X = C_1 e^{s_1 t} + C_2 e^{s_2 t}$$

$$X_2 = \dot{X} = s_1 C_1 e^{s_1 t} + s_2 C_2 e^{s_2 t} \tag{24.5}$$

where

$$C_1 = \frac{s_2 X_o - \dot{X}_o}{s_2 - s_1}$$

$$C_2 = \frac{\dot{X}_o - s_1 X_o}{s_2 - s_1}$$

and X_0 and $\dot{X}_0$ are the initial conditions; thus $X_0 = X(0)$ and $\dot{X}_0 = \dot{X}(0)$. The terms s_1 and s_2 are the roots of the characteristic equation

$$|s\mathbf{I} - \mathbf{A}| = 0 \tag{24.6}$$

Expanding this equation gives

$$\tau^2 s^2 + 2\zeta\tau s + 1 = 0$$

This quadratic equation has two roots:

$$s_{1,2} = \frac{-\zeta \pm \sqrt{\zeta^2 - 1}}{\tau}$$

If we take s_2 as the root with the positive sign

$$s_2 = \frac{-\zeta + \sqrt{\zeta^2 - 1}}{\tau}$$

the constants take the form

$$C_1 = \frac{\tau}{2\sqrt{\zeta^2 - 1}}\left(s_2 X_0 - \dot{X}_0\right)$$

$$C_2 = \frac{\tau}{2\sqrt{\zeta^2 - 1}}\left(\dot{X}_0 - s_1 X_0\right) \tag{24.7}$$

Equations (24.5) and (24.7) together give $X(t)$ and $\dot{X}(t)$ for all possible initial conditions X_0 and $\dot{X}_0$. For a given set of initial conditions, we compute C_1 and C_2 from Eq. (24.7), and then each value of t in Eq. (24.5) yields a pair of values for X and $\dot{X}$.

These may be plotted as a point on an $\dot{X}$-X diagram (i.e., a phase plane). The locus of these points as t varies from zero to infinity will be a curve in the $\dot{X}X$ plane. As an example, consider the case $X_0 = -1$, $\dot{X}_0 = 0$, $\zeta < 1$. The solution is already known to us in the form of X versus t (Chap. 7) and is replotted in Fig. 24–3 for convenience, together with a plot of $\dot{X}$ versus t. If these curves are replotted as $\dot{X}$ versus X, with t as a parameter, the result is as shown in Fig. 24–4. The reader should carefully compare Figs. 24–3 and 24–4 to be satisfied that they are indeed equivalent. The relationship between the two may be expressed by the statement that Fig. 24–3 is a parametric representation of Fig. 24–4. Having only the curve X versus t of Fig. 24–3, one can construct Fig. 24–4.
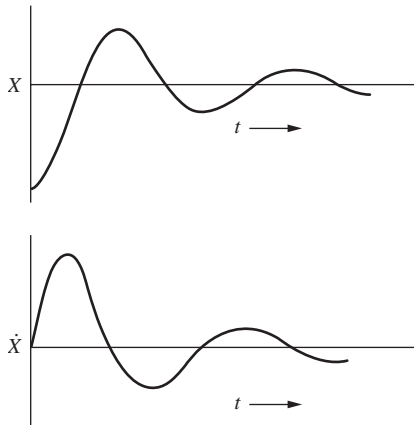
FIGURE 24–3
Typical motion of second-order system.
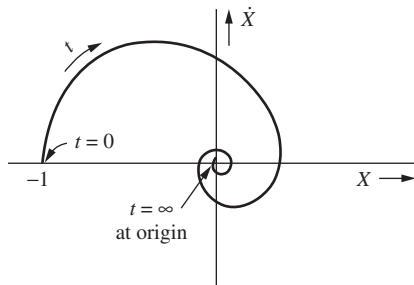
FIGURE 24–4
Phase plane corresponding to motion of Fig. 24–3.

To explore the phase diagram concept further, note that division of the second of Eqs. (24.3) by the first yields

$$\frac{d\dot{X}}{dX} = \frac{-X - 2\zeta\tau\dot{X}}{\tau^2\dot{X}} \tag{24.8}$$

in which the variable t has been eliminated. Equation (24.8) may be recognized as a homogeneous first-order differential equation. Hence, the substitution $\dot{X} = VX$ yields

$$\frac{X\,dV}{dX} = \frac{-1 - 2\zeta\tau V}{\tau^2 V} - V = \frac{-\left(1 - 2\zeta\tau V + \tau^2 V^2\right)}{\tau^2 V}$$

an equation which is separable in X and V. This can then be easily solved for V in terms of X. Finally, replacing $V = \dot{X}/X$ gives the solution for $\dot{X}$ versus X, or the equation for the curve of Fig. 24–4. The algebraic details of this rather tedious process are omitted. The point of the discussion is to emphasize further the equivalence between the description of the motion as X versus t or $\dot{X}$ versus X.

A convenient feature of the phase diagram is that several motions, corresponding to different initial conditions, can be readily plotted on the same diagram. Thus, if we add to Fig. 24–4 a curve for the motion under the initial condition $X_0 = 1$, $\dot{X}_0 = 0$, we obtain Fig. 24–5. This new trajectory represents the motion of the system after it is stretched 2 units and released from rest. (This follows from the definition $X = Y - 1$.) Furthermore, we have also interpolated in Fig. 24–5 to obtain the motion corresponding to $X_0 = 0$, $\dot{X}_0 = 1$, As we shall see later, this interpolation is justified. Hence, it is evident that the phase diagram gives us the "big picture" of the motion of the underdamped U-tube manometer or a spring-mass-damper system. No matter where the system starts, it spirals to the condition $X_0 = 1, \dot{X}_0 = 0$, the steady-state position. This spiral motion in the phase plane corresponds to the oscillatory nature of the X versus t curve of Fig. 24–3.
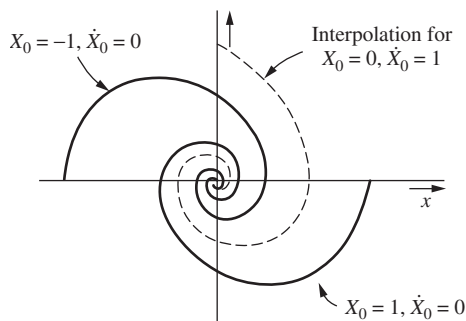
FIGURE 24–5
Interpolation on the phase plane.

Before we begin a more detailed study of the mechanics of phase analysis, it may be worthwhile to see how situations amenable to such analysis arise naturally in the physical world.

The Damped Oscillator

A classical example from mechanics is the damped oscillator, shown in Fig. 24–6. A block of mass W resting on a horizontal, frictionless table is attached to a linear spring. A viscous damper (dashpot) is also attached to the block. Assume that the system is free to oscillate horizontally under the influence of a forcing function $F(t)$. The origin of the coordinate system is taken as the right edge of the block when the spring is in the relaxed or unstretched condition. At time zero, the block is assumed to be at rest at this origin. In effect, this assumption makes the displacement variable $Y(t)$ a deviation variable. Also, the assumption that the block is initially at rest permits derivation of the second-order transfer function in its standard form. An initial velocity has the same effect as a forcing function. Hence, this assumption is in no way restrictive.
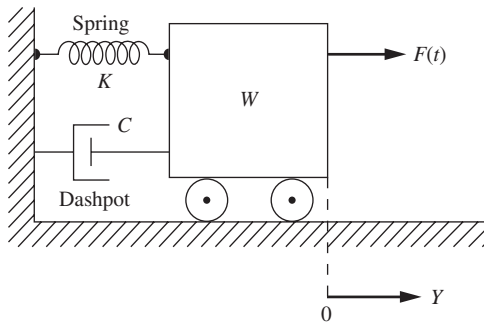
FIGURE 24–6
The damped oscillator.

Positive directions for force and displacement are indicated by the arrows in Fig. 24–6.

Consider the block at some instant when it is to the right of $Y = 0$ and when it is moving toward the right (positive direction). Under these conditions, the position Y and the velocity dY/dt are both positive. At this particular instant, the following forces are acting on the block:

1. The force exerted by the spring (toward the left) of $-KY$ where K is a positive constant, called Hooke's constant
2. The viscous friction force (acting to the left) of $-C\,dY/dt$, where C is a positive constant called the damping coefficient
3. The external force $F(t)$ (acting toward the right)

Newton's law of motion, which states that the sum of all forces acting on the mass is equal to the rate of change of momentum (mass $\times$ acceleration), takes the form

$$\frac{W}{g_c}\frac{d^2Y}{dt^2} = -KY - C\frac{dY}{dt} + F(t) \tag{24.9}$$

Rearrangement gives

$$\frac{W}{g_c}\frac{d^2Y}{dt^2} + C\frac{dY}{dt} + KY = F(t) \tag{24.10}$$

where W = mass of block, lb_m
g_c = 32.2 $\text{lb}_m \cdot \text{ft}/(\text{lb}_f \cdot \text{s}^2)$
C = viscous damping coefficient, $\text{lb}_f/(\text{ft/s})$
K = Hooke's constant, lb_f/ft
$F(t)$ = driving force, a function of time, lb_f

Dividing Eq. (24.10) by K gives

$$\frac{W}{g_c K}\frac{d^2Y}{dt^2} + \frac{C}{K}\frac{dY}{dt} + Y = \frac{F(t)}{K} \tag{24.11}$$

For convenience, this is written as

$$\tau^2 \frac{d^2Y}{dt^2} + 2\zeta\tau\frac{dY}{dt} + Y = X(t) \tag{24.12}$$

where

$$\tau^2 = \frac{W}{g_c K} \tag{24.13}$$

$$2\zeta\tau = \frac{C}{K} \tag{24.14}$$

$$X(t) = \frac{F(t)}{K} \tag{24.15}$$

Solving for τ and ζ from Eqs. (24.13) and (24.14) gives

$$\tau = \sqrt{\frac{W}{g_c K}} \quad \text{s} \tag{24.16}$$

$$\zeta = \sqrt{\frac{g_c C^2}{4WK}} \quad \text{dimensionless} \tag{24.17}$$

If the block is motionless ($dY/dt = 0$) and located at its rest position ($Y = 0$) before the forcing function is applied, the Laplace transform of Eq. (24.12) becomes

$$\tau^2 s^2 Y(s) + 2\zeta\tau s Y(s) + Y(s) = X(s) \tag{24.18}$$

Example 24.1.   Consider the motion of the damped oscillator in Fig. 24–6. Write the matrix differential equations that characterize the system. The following constants apply, and $Y(0) = 0$ and $\dot{Y}(0) = 0$.

$$\frac{W}{g_c} = 1\frac{\text{lb}_f}{\text{ft / s}^2} \qquad C = 2\frac{\text{lb}_f}{\text{ft / s}}$$

$$K = 5\frac{\text{lb}_f}{\text{ft}} \qquad F(t) = 10 \text{ lb}_f$$

From Eqs. (24.15) through Eq. (24.17), we can calculate

$$X(t) = \frac{10}{5} = 2 \text{ ft}$$

$$\tau = \sqrt{\frac{1}{5}}\text{ s}$$

$$\zeta = \sqrt{\frac{2^2}{(4)(5)}} = \sqrt{\frac{1}{5}}$$

The governing differential equation Eq. (24.12) can now be written as

$$\frac{1}{5}\frac{d^2Y}{dt^2} + \frac{2}{5}\frac{dY}{dt} + Y = X(t)$$

Defining the following phase variables

$$Y = Y_1$$

$$\frac{dy}{dt} = \dot{Y}_1 = Y_2$$

the differential equation becomes

$$\dot{Y}_2 = \left(\frac{-1}{\tau^2}\right)Y_1 - \frac{2\zeta\tau}{\tau^2}Y_2 + \frac{1}{\tau^2}X$$

Summarizing these equations in matrix form yields

$$\begin{bmatrix} \dot{Y}_1 \\ \dot{Y}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{-1}{\tau^2} & -\frac{2\zeta}{\tau} \end{bmatrix}\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{\tau^2} \end{bmatrix}X(t) = \begin{bmatrix} 0 & 1 \\ -5 & -2 \end{bmatrix}\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 5 \end{bmatrix}X(t) \quad (24.19)$$

Example 24.2.   Use MATLAB to solve the matrix differential equations developed in Example 24.1 for the damped oscillator, and plot the phase plane.

   We can use the MATLAB differential equation-solving routine ODE23 to solve the system of first-order differential equations represented by Eq. (24.19). The following m-file represents the model.

```
function ydot=ex24_2(t,y)
%function file for damped oscillator
%position is the first variable and velocity is the second
A=[0,1; -5, -2];
B=[0;5];
x=2;
ydot=A*y+B*x;
```

The corresponding MATLAB commands to solve the model are

```
>> [t,y] = ode45(@ex24_2,[0,5],[0,0]);
>> plot(t,y)
>> figure
>> plot(y(:,1),y(:,2))
```

The solution is returned in the *y* matrix. The first column is the displacement Y_1, and the second column is the velocity Y_2. Thus, a plot of *y* versus *t* yields the typical time plots, while the plot of the first column, `y(:,1)`, versus the second column, `y(:,2)`, is the phase-plane plot, with time as a parameter along the curve. (See Fig. 24–7)
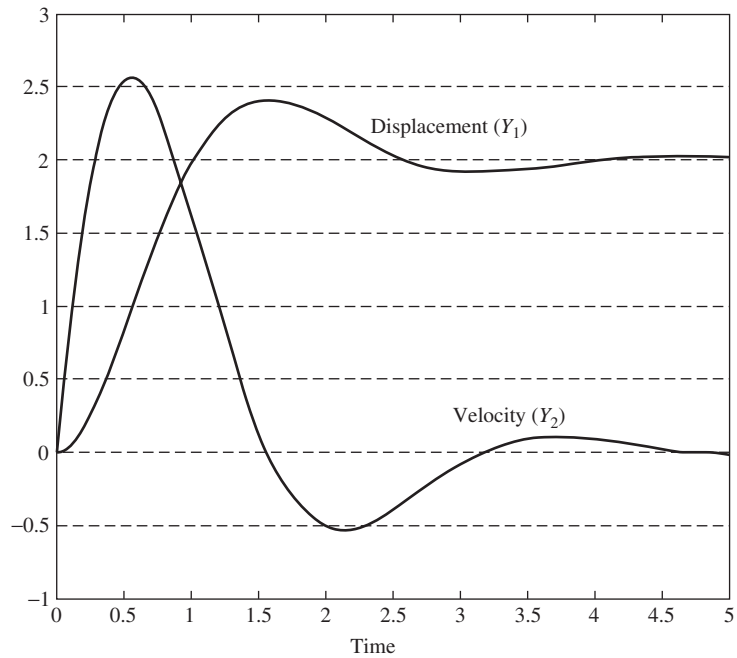
(a) Velocity and displacement for the damped oscillator in Example 24.2.
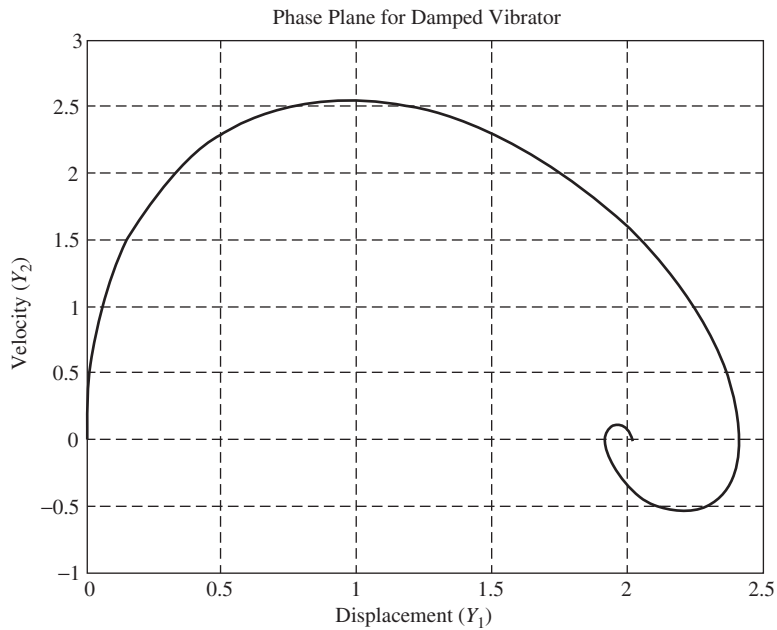
(b) Phase plane for the damped oscillator in Example 24.2.

FIGURE 24–7
Plot of Solution to Example 24.2.

Notice that the phase plane plot starts at (0,0) and spirals into the steady state (2,0) at $t = \infty$. It's also clear from the original differential equation [Eq. (24.18)] that $Y = 2$ at steady state.

24.4 MOTION OF A PENDULUM

Consider the pendulum of Fig. 24–8. As the pendulum is moving in the direction shown, there are two forces acting to oppose its motion.
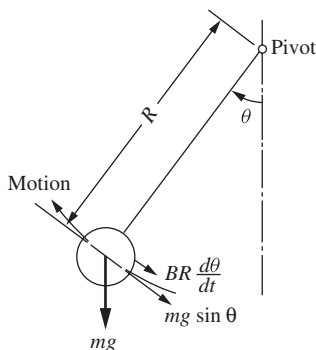
FIGURE 24–8
Forces acting on a pendulum.

These forces, which act tangentially to the circle of motion, are (1) the gravitational force $mg \sin \theta$ and (2) the friction in the pivot, which we suppose to be proportional to the tangential velocity of the mass $BR(d\theta/dt)$, where B is the proportionality constant. We shall assume the air resistance to be negligible and the rod to be of negligible mass. Application of Newton's second law gives

$$-mR\frac{d^2\theta}{dt^2} = mg \sin\theta + BR\frac{d\theta}{dt}$$

Rearrangement leads to

$$\frac{d^2\theta}{dt^2} + D\frac{d\theta}{dt} + \omega_n^2 \sin\theta = 0 \qquad (24.20)$$

where $D = \dfrac{B}{m} \qquad \omega_n^2 = \dfrac{g}{R}$

This equation resembles the equation for the motion of the damped oscillator system. However, the presence of the term involving $\sin\theta$ makes the equation nonlinear.

Equation (24.20) has the following form in phase coordinates:

$$\frac{d\theta}{dt} = \dot{\theta}$$

$$\frac{d\dot{\theta}}{dt} = -\omega_n^2 \sin\theta - D\dot{\theta} \qquad (24.21)$$

and a phase diagram would be a plot of angular velocity $\dot{\theta}$ versus position θ. At this point, we can gain some insight by simple analysis of Eq. (24.21) without actually obtaining a solution.

Referring for the moment to the spring-mass-damper system of Fig. 24–3, we saw that the system ceased to oscillate when the point $dX/dt = d\dot{X}/dt = 0$ was reached. That is, all curves stopped at the origin of Fig. 24–5. Since neither X nor $\dot{X}$ is changing with time, the motion ceases. Further examination of Eqs. (24.3) shows that $X = \dot{X} = 0$ is the *only* point at which both dX/dt and $d\dot{X}/dt$ are zero. Thus, we see that the mass will come to rest *only* when the situation of zero displacement (remember that in this case X is a deviation variable defined relative to the final steady-state displacement) and zero velocity is reached.

Now we can perform a similar analysis on Eqs. (24.21). We are asking the following question: At what point or points in the phase plane ($\dot{\theta}$ versus θ diagram) do both $d\theta/dt$ and $d\dot{\theta}/dt$ become zero? From the first of these equations, we see that this can happen only when $\dot{\theta} = 0$.

The reader should not be lulled into a false sense of security at this point. It would be wise to disregard the fact that $d\theta/dt$ and $\dot{\theta}$ are, in fact, the same quantity; $\dot{\theta}$ should be thought of as a coordinate in the phase plane, and $d\theta/dt$ as the rate of change with time of the other coordinate. The virtue of making this distinction will become clear in the next example, a chemical reactor.

Thus, the analysis leads to the conclusion that the motion will cease when the pendulum comes to rest in either of the positions shown in Fig. 24–9. In addition, it is clear from Eqs. (24.21) that if the pendulum stops at any other point, the motion continues. Of course, this analysis agrees with our physical intuition. However, we expect to find a distinction between the stability characteristics of the two equilibrium points, since the position at π is likely to be hard to attain and maintain. This distinction will be explored in greater detail in Chap. 25.
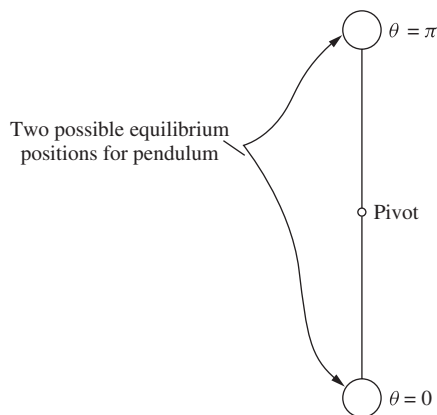
FIGURE 24–9
Equilibrium positions for pendulum.

By using this result in the second equation, it can be seen that it is also necessary that

$$\sin\theta = 0 \tag{24.22}$$

Equation (24.22) is satisfied at any of the points

$$\theta = n\pi$$

where n is a positive or negative integer or zero. However, from a physical standpoint, we can really distinguish between only two of these points, which we take as $\theta = 0$ and $\theta = \pi$. Thus, the positions $\theta = 0, 2\pi, 4\pi, -2\pi$, etc., all look the same; i.e., the pendulum is hanging straight down. Similarly, the points $\theta = \pi, 3\pi$, etc., all correspond to the pendulum standing straight up.

Example 24.3.   Use MATLAB to plot the displacement and velocity of a pendulum as a function of time as well as the phase-plane plot, using the following numerical values:

$$R = 1 \text{ m} \qquad m = 1 \text{ kg}$$
$$B = 2 \frac{\text{kg}}{\text{s}} \qquad g = 9.8 \frac{\text{m}}{\text{s}^2}$$
$$\theta(0) = \frac{\pi}{4} \qquad \frac{d\theta}{dt} = 0$$

Using these values, Eq. (24.21) becomes

$$\boxed{\theta = y_1}$$
$$\frac{d\theta}{dt} = \boxed{\frac{dy_1}{dt} = y_2}$$
$$\frac{d^2\theta}{dt^2} = \boxed{\frac{dy_2}{dt} = -9.8 \sin y_1 - 2y_2}$$

The MATLAB m-file for the model is

```
function ydot = ex24_3(t,y)
%function file for pendulum
%position is the first variable and velocity is the second
ydot(1,1) = y(2);
ydot(2,1) = -9.8*sin(y(1)) -2*y(2);
```

The MATLAB commands to execute the file and plot the necessary graphs are

```
>>[t,y] = ode45(@ex24_3,[0,10],[pi/4,0]);
>> plot(t,y)
>> figure
>> plot(y(:,1),y(:,2))
```

Note that the phase plane starts at the initial condition of $\theta = \pi/4, \dot{\theta} = 0$ and spirals into the steady state of $\theta = 0, \dot{\theta} = 0$ as we would expect (see Fig. 24–10).
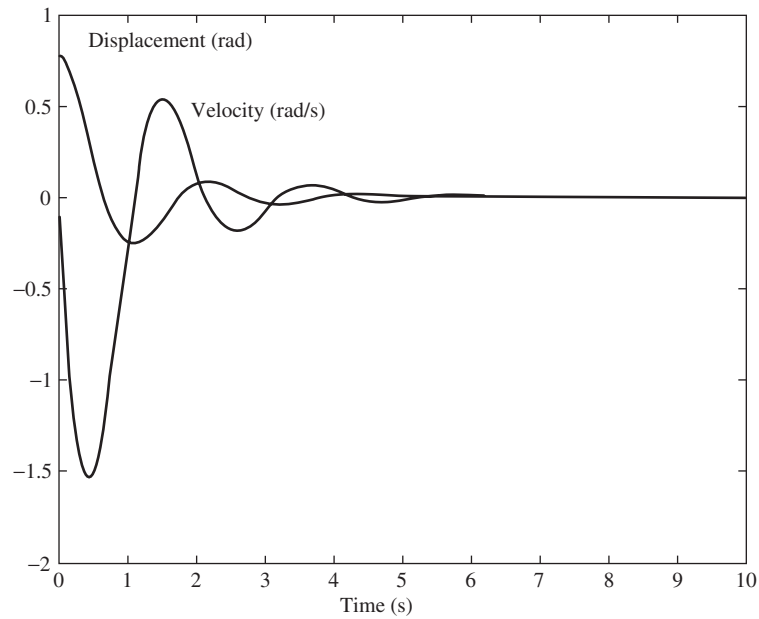
(*a*) Displacement and velocity of the pendulum in Example 24.3.
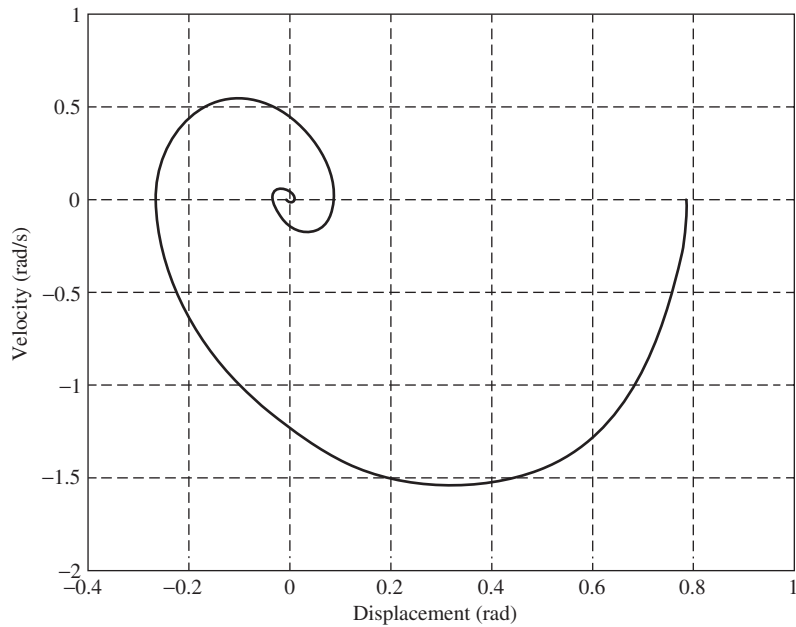
(*b*) Phase plane for the pendulum in Example 24.3.

FIGURE 24–10
Plot of the solution for Example 24.3.

24.5 A CHEMICAL REACTOR*
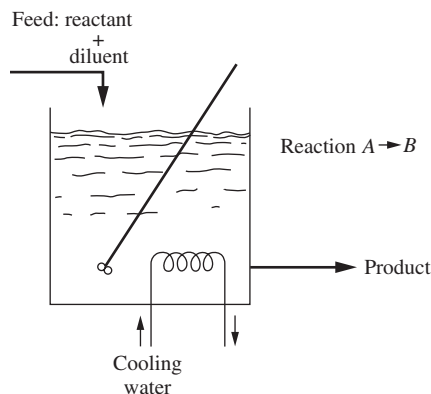
FIGURE 24–11
Schematic of exothermic chemical reactor.

Consider the stirred-tank chemical reactor of Fig. 24–11. The contents of the reactor are assumed to be perfectly mixed, and the reaction taking place is

$$A \rightarrow B \qquad (24.23)$$

which occurs at a rate

$$R_A = kC_Ae^{-E/RT} \qquad (24.24)$$

where R_A = moles of A decomposing per hour per cubic foot of reacting mixture
$\quad k$ = reaction velocity constant, h^{-1}
$\quad C_A$ = concentration of A in reacting mixture, mol/ft^3
$\quad E$ = activation energy, a constant, Btu/mol
$\quad R$ = universal gas law constant
$\quad T$ = absolute temperature of reacting mixture

The reaction is exothermic; ΔH Btu of heat is generated for each mole of A that reacts. Hence, to control the reactor, cooling water is supplied to a cooling coil. The actual reactor temperature is compared with a set point, and the rate of cooling water flow is adjusted accordingly. To indicate this control mathematically, we write that $Q(T)$ Btu/h of heat is removed through the cooling coil. In Chap. 25 we make a more detailed analysis of the dynamic behavior of the reactor. For the present preliminary analysis, it is not necessary to look carefully at $Q(T)$, and hence it is merely assumed that as T rises, more heat is removed in the coil. Let

$$x_{A_0} = \text{mole fraction of } A \text{ in feed stream}$$
$$x_{B_0} = \text{mole fraction of } B \text{ in feed stream}$$

---

* This example is based on the work of R. Aris and N. R. Amundson (1958).

Then $1 - x_{A_0} - x_{B_0}$ is the fraction of inerts in the feed stream. A mole balance on A

$$
\left(\begin{array}{c} \text{Molar flow rate} \\ \text{of } A \text{ in with feed} \end{array} \right) - \left(\begin{array}{c} \text{Molar flow rate} \\ \text{of } A \text{ out with product} \end{array} \right) - \left(\begin{array}{c} \text{Molar rate of} \\ A \text{ reacting} \end{array} \right)
$$

$$
= \left(\begin{array}{c} \text{Molar rate of accumulation} \\ \text{of } A \text{ in reactor} \end{array} \right)
$$

takes the form

$$
F\rho x_{A_0} - F\rho x_A - k\rho V e^{-E/RT} x_A = \rho V \frac{dx_A}{dt} \tag{24.25}
$$

where F = feed rate, ft^3/h

$\quad x_A$ = mole fraction of A in reactor

$\quad \rho$ = molar density of reacting mixture, mol/ft^3

$\quad V$ = volume of reacting mixture, ft^3

To arrive at Eq. (24.25), we have used Eq. (24.24) and made the following assumptions:

1. The density of the reacting mixture is constant, unaffected by the conversion of A to B.
2. The feed and product rates F are equal and constant.
3. Together, 1 and 2 imply that V, the volume of reacting mixture, is constant.
4. Perfect mixing occurs, so that x_A is the same in the reactor and product stream.

A similar mole balance may be derived for substance B. However, Eq. (24.23) shows that 1 mol of B appears for every mole of A destroyed. Hence

$$
x_B - x_{B_0} = x_{A_0} - x_A \tag{24.26}
$$

Equation (24.26) permits us to circumvent the mass balance for x_B, since knowing x_A, we can calculate x_B directly.

The energy balance on the reactor

(Sensible heat in feed) $-$ (Sensible heat in product)

$+$ (Heat generated by reaction) $-$ (Heat removed in cooling coil)

$=$ (Energy accumulating in reactor)

can be written as

$$
F\rho C_p (T_0 - T) + k\rho V (\Delta H) e^{-E/RT} x_A - Q(T) = \rho V C_p \frac{dT}{dt} \tag{24.27}
$$

where T_0 = temperature of feed stream

$\quad T$ = temperature in reactor (and product, since reactor is well mixed)

$\quad C_p$ = specific heat of reacting mixture

In writing Eq. (24.27), it is assumed that

1. The specific heat of the reacting mixture is constant, unaffected by the conversion of A to B.
2. The perfect mixing means that the temperatures of the reacting mixture and product stream are the same.
3. The heat of reaction ΔH is constant, independent of temperature and composition.

We remark here that these assumptions, as well as those made in Eq. (24.25), may be relaxed without affecting the conceptual aspects of the phase analysis. They are made only to keep the example as uncluttered as possible, without being trivial.

Equations (24.25) and (24.27) may be rearranged to the system

$$
\begin{aligned}
\frac{dx_A}{dt} &= \frac{F}{V}(x_{A_0} - x_A) - ke^{-E/RT}x_A \\
\frac{dT}{dt} &= \frac{F}{V}(T_0 - T) + \frac{k(\Delta H)}{C_p}e^{-E/RT}x_A - \frac{Q(T)}{\rho V C_p}
\end{aligned}
\tag{24.28}
$$

As a typical application of this system of equations, we might consider starting up the reactor, initially filled with a mixture at composition $x_A(0)$ and temperature $T(0)$. Suppose the feed rate, feed composition, feed temperature, and flow rate of cooling water are held constant and the reactor is operated in this manner until steady state is reached. To describe the transient behavior of the chemical reactor, one can solve Eqs. (24.28) by integrating them numerically, using a typical stepwise procedure such as the Euler or Runge-Kutta method. In MATLAB, ODE45, a differential equation solver, could be used. This will result in functions $x_A(t)$ and $T(t)$ for values of t from zero to some value (if one exists) at which, for practical purposes, $x_A(t)$ and $T(t)$ cease to change with t.

Alternatively, we may consider a phase-plane analysis of Eqs. (24.28) and seek solutions in the form of x_A versus T curves. Note that division of the first of Eqs. (24.28) by the second gives

$$
\frac{dx_A}{dT} = \frac{(F/V)(x_{A_0} - x_A) - ke^{-E/RT}x_A}{(F/V)(T_0 - T) + \left[k(\Delta H)/C_p\right]e^{-E/RT}x_A - Q(T)/\rho V C_p}
\tag{24.29}
$$

The parameter t has been eliminated in Eq. (24.29), which is simply a differential equation relating x_A and T. As we shall see in Chap. 25, this phase-plane analysis of the chemical reactor offers significant advantages over the ordinary analysis.

In the chemical reactor, we no longer have the special relationship among the phase variables that we had in both previous cases. For both the spring and pendulum problems, we more or less artificially changed a second-order differential equation to two first-order equations by introducing the phase variable $\dot{X}$ (or $\dot{\theta}$). This phase variable was directly related to the other phase variable X (or θ) by the equation

$$
\dot{X} = \frac{dX}{dt}
$$

For the chemical reactor, there is no such simple relation between x_A and T.

We can study the steady-state solutions to Eqs. (24.28) without solving the equations, much as was done in the case of the damped pendulum of the previous example. As before, we note that steady state requires that x_A and T simultaneously cease to change with time,

$$\frac{dx_A}{dt} = \frac{dT}{dt} = 0$$

From Eqs. (24.28), this implies that

$$\frac{F}{V}(x_{A_0} - x_{A_s}) - ke^{-E/RT_s}x_{A_s} = 0$$

$$\frac{F}{V}(T_0 - T_s) - \frac{k(\Delta H)}{C_p}e^{-E/RT_s}x_{A_s} - \frac{Q(T_s)}{\rho V C_p} = 0 \tag{24.30}$$

where x_{A_s} and T_s are the steady-state values of x_A and T.

The first of Eqs. (24.19) can be solved for x_{A_s}, yielding

$$x_{A_s} = x_{A_0}\frac{1}{1 + (kV / F)e^{-E/RT_s}} \tag{24.31}$$

Substitution of (24.31) into the second of Eqs. (24.30) yields

$$\frac{k(\Delta H)x_{A_0}/C_p}{e^{E/RT_s} + kV/F} = \frac{Q(T_s)}{\rho V C_p} + \frac{F}{V}(T_s - T_0) \tag{24.32}$$

Equation (24.32) is implicit in T_s, the steady-state temperature. In physical terms, it expresses an equality between the heat generated by the reaction and the heat removed in the cooling coil and product stream. To emphasize this, we have arranged it so that the left side is the heat generation and the right side is the heat removal.

Solution of Eq. (24.32) for T_s requires numerical values for the various parameters. Without going into this much detail at present, we may obtain some qualitative information. To do this, we sketch the right and left sides of this equation as functions of T_s. A typical shape for the left side is given by the sigmoidal curve of Fig. 24–12. (See Aris and Amundsen, 1958, p.121.) The unusual curvature, of course, is caused by the e^{E/RT_s} term in the denominator. To plot the right side, we must know $Q(T)$. While we have avoided specifying the form of $Q(T)$, we know it increases with T. If there were no control action, i.e., if the flow rate of cooling water were maintained constant regardless of T, then $Q(T)$ would increase almost linearly with T. This is so because at constant water rate, the heat transfer in the coil is approximately proportional to the difference between T and the mean temperature of the cooling water. This latter temperature would not vary as rapidly as T at practical flow rates. However, since we expect to have control action, we know that the cooling water flow rate will be increased with increasing T. Therefore, $Q(T)$ may be expected to increase faster than linearly with T, which means that the right side of Eq. (24.32) increases faster than linearly. Several typical curves of this right side are shown in Fig. 24–12.

A solution of Eq. (24.32) requires that the graphs of the right and left sides intersect. As shown in Fig. 24–12, there may be one, two, or three such intersections, depending on the relative locations of the heat generation (left side) and heat removal (right side). This means that there may be one, two, or three possible steady states for the reactor.

Different possible locations
of heat removal
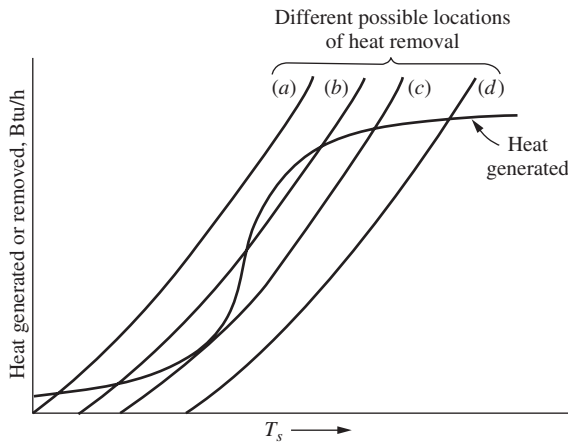
FIGURE 24–12
Steady-state generation and removal
functions for exothermic chemical
reactor

As we shall see in Chap. 25, the steady state actually attained by the reactor depends on initial conditions $x_A(0)$ and $T(0)$. The steady-state temperature T_s is then the temperature at the pertinent intersection, and the steady-state composition can be determined from Eq. (24.31). We shall also see that some of the steady states are unstable. In fact, the low-temperature steady state for curve (c) of Fig. 24.12, occurring as a point of tangency, is to be regarded with suspicion. Practically speaking, a perfect tangency would not occur. Minor variations in operating conditions (i.e., noise), which occur continually in actual process operation, may shift the curve (c) slightly to the left or right, resulting in two or zero low-temperature intersections, respectively.

SUMMARY

In this chapter, we have introduced the concept of a phase analysis and some of its basic elements. We have seen how physical situations give rise naturally to phase solutions. Furthermore, we have had our first look at true nonlinear behavior. In so doing, we have come to at least one interesting conclusion: a nonlinear motion or control system response may have more than one steady-state solution. This was true for the chemical reactor and for the pendulum. In contrast, the linear motions and control system responses we studied in previous chapters had only one steady-state solution. In Chap. 25, we shall discover still more differences that render nonlinear analysis more difficult than linear analysis.

PROBLEMS

24.1 Use MATLAB to analyze the damped oscillator shown in Fig. 24–6, using the following numerical values. Plot the displacement and velocity as a function of time as well as the phase plane for the sytem.

$$\frac{W}{g_c} = 1 \ \frac{\text{lb}_f}{\text{ft} / \text{s}^2} \qquad C = 2 \ \frac{\text{lb}_f}{\text{ft} / \text{s}} \qquad K = 2 \frac{\text{lb}_f}{\text{ft}}$$

(a) $F(t) = 6 \ \text{lb}_f, Y(0) = 0$, and $\dot{Y}(0) = 0$

(b) $F(t) = 6 \ \text{lb}_f, Y(0) = 2$, and $\dot{Y}(0) = 0$

(c) $F(t) = 6 \ \text{lb}_f, Y(0) = 0$, and $\dot{Y}(0) = 2$

24.2. Plot the phase planes for Prob. 24.1, parts (a) to (c), on the same set of axes.

CAPSULE SUMMARY

In this chapter, we have introduced the concept of a phase analysis and some of its basic elements. We have seen how physical situations give rise naturally to phase solutions. Furthermore, we have had our first look at true nonlinear behavior. In so doing, we have come to at least one interesting conclusion: a nonlinear motion or control system response may have more than one steady-state solution. This was true for the chemical reactor and for the pendulum. In contrast, the linear motions and control system responses we studied in previous chapters had only one steady-state solution. In Chap. 25, we shall discover still more differences that render nonlinear analysis more difficult than linear analysis.

Typical displacement versus time plot
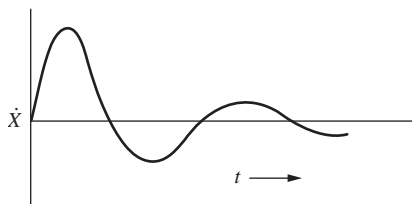
Typical velocity versus time plot
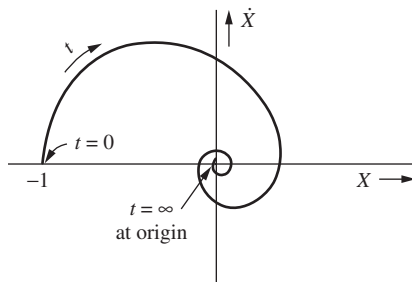
Phase-plane plot with time as a parameter along the length of the curve

CHAPTER
25

EXAMPLES
OF PHASE-PLANE
ANALYSIS

The advantages of the phase analysis introduced in Chap. 24 can be more fully appreciated after some acquaintance with the tools available for such analysis. To give a detailed exposition of all, or even most, of the aspects of this subject is not intended. Instead, this chapter strives to indicate its flavor and to stimulate further study.

25.1 PHASE SPACE

In Chap. 24, we considered three examples for which the dynamic response can be described by two state variables. For the cases of the damped oscillator and the pendulum, the state variables were phase variables in which the dependent variable and its derivative (X, $\dot{X}$ or θ, $\dot{\theta}$) were chosen as the state variables. For the exothermic chemical reactor, the state variables selected were temperature and composition (T, x_A); these variables, which arose naturally in the analysis of the chemical reactor, were called physical variables in Chap. 21.

In general, an nth-order dynamic system can be described by n state variables. The state variables ($x_1, x_2, \ldots, x_n$) can be located in a coordinate system called phase space. Each value of t, say t_1, defines a point in this space: $x_1(t_1)$, $x_2(t_1)$, $\ldots$, $x_n(t_1)$. The solution curve is a locus of these points for all values of t. It is called a *trajectory* and connects successive states of the system. For the damped oscillator presented in Chap. 24, the coordinate system was a plane with an axis for each state variable; we shall refer to this coordinate system as a phase plane. Figure 24–5 is a typical phase-plane representation of a dynamic system. When the physical system is third-order, the coordinate system consists of three axes, one for each state variable. Of course, systems of fourth or higher order require treatment in space that is of too many dimensions to be visualized. The graphic aspects of phase-space representation are advantageous primarily in the case of two dimensions (the phase plane) and to a limited extent for three

dimensions. The bulk of practical use of phase-space analysis has been made in the two-dimensional autonomous (time-invariant) case:

$$\frac{dx_1}{dt} = f_1(x_1, x_2)$$
$$\frac{dx_2}{dt} = f_2(x_1, x_2)$$
(25.1)

For this reason, we largely confine our attention in the remainder of this study to systems that may be written in the form of Eqs. (25.1). As we have seen, there is no loss in *conceptual generality,* but we cannot expect the *graphical aspects* of the material we shall develop to generalize to higher-dimension phase space. The solution of the system (25.1) may be presented as a family of trajectories in the $x_2 x_1$ plane. If we are given the initial conditions

$$x_1(t_0) = x_{10}$$
$$x_2(t_0) = x_{20}$$

the initial state of the system is the point (x_{10}, x_{20}) in the $x_2 x_1$ plane, and the trajectory may be traced from this point.

By dividing the second of Eqs. (25.1) by the first, we obtain

$$\frac{dx_2}{dx_1} = \frac{f_2(x_1, x_2)}{f_1(x_1, x_2)}$$
(25.2)

Now dx_2/dx_1 is merely the slope of a trajectory, since a trajectory is a plot of x_2 versus x_1 for the system. Hence, at each point in the phase plane (x_1, x_2), Eq. (25.2) yields a unique value for the slope of a trajectory through the point, namely, $f_2(x_1, x_2)/f_1(x_1, x_2)$. This last statement should be amended to exclude any point (x_1, x_2) at which $f_1(x_1, x_2)$ and $f_2(x_1, x_2)$ are *both* zero. These important points are called *critical points* and will be examined in greater detail below. Since the slope of the trajectory at a point, say (x_1, x_2), is unique by Eq. (25.2), it is clear that trajectories cannot intersect except at a critical point, where the slope is indeterminate.

Using Simulink/MATLAB to Plot a Phase Plane

Let us illustrate the use of Simulink/MATLAB in phase-plane analysis with an example.

Example 25.1.   Find the trajectory of the system

$$\frac{dx_1}{dt} = x_2$$
$$\frac{dx_2}{dt} = -5x_1 - 2x_2$$
(25.3)

which passes through the point
$$x_1 = 1 \qquad x_2 = 0$$

The Simulink diagram for simulating this system is shown in Fig. 25–1a.
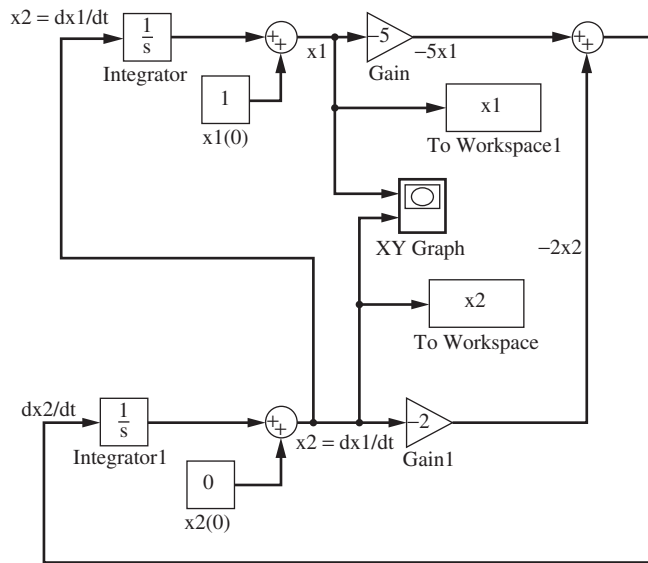
(*a*) Simulink diagram for Example 25.1.
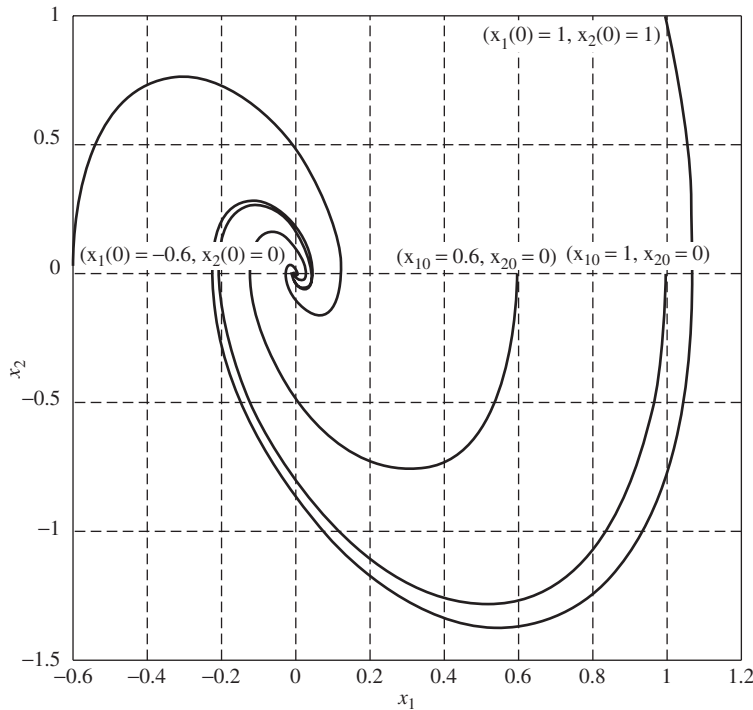
(*b*) Phase plane for Example 25.1 from Simulink.

FIGURE 25–1
Simulink solution for Example 25.1.

Another trajectory, starting from the point $(-0.6, 0)$, is shown on Fig. 25–1b. This serves to emphasize that interpolation is possible on the phase plane, and many trajectories representing various initial conditions are easily visualized or sketched.

Analysis of Critical Points

In the situations of most interest to us, Eq. (25.2) will represent the behavior of a (non-linear) control system, as in Eq. (24.29). Therefore, we shall be interested in maintaining the system at or near a steady state. Since, from Eq. (25.1), a steady-state point is defined by

$$f_1(x_1, x_2) = f_2(x_1, x_2) = 0$$

it is clear that the steady states are critical points. At the critical points, the slope of the trajectory is undefined; hence, many trajectories may intersect at these points. In Fig. 25–1 the origin is a critical point. It can be seen from the figure that, in this case, all trajectories spiral into the origin. Hence, this particular system is such that, no matter what the initial state (i.e., for any disturbance which is applied), the system returns to steady state at the critical point.

The critical point of Fig. 25–1 is called a *focus,* because the trajectories spiral into it. This spiral motion of the trajectories corresponds to the oscillatory approach of the system to steady state. The oscillatory motion occurs because the system of Eqs. (25.3) is underdamped, as indicated by the characteristic equation

$$|s\mathbf{I} - \mathbf{A}| = 0$$

or
$$\left| \underbrace{\begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix}}_{sI} - \underbrace{\begin{bmatrix} 0 & 1 \\ -5 & -2 \end{bmatrix}}_{A} \right| = \begin{vmatrix} s & -1 \\ 5 & s+2 \end{vmatrix} = s^2 + 2s + 5 = 0 = \tfrac{1}{5}s^2 + \tfrac{2}{5}s + 1$$

When put into standard form, this characteristic equation has parameters

$$\tau^2 = \tfrac{1}{5} \qquad 2\zeta\tau = \tfrac{2}{5}$$
$$\tau = \tfrac{1}{\sqrt{5}} \qquad \zeta = \tfrac{1}{\sqrt{5}}$$

Since $\zeta < 1$, the system is underdamped.

An overdamped system, such as that generated by the system

$$\frac{dx_1}{dt} = x_2 \qquad \frac{dx_2}{dt} = -5x_1 - 6x_2$$

having characteristic equation

$$s^2 + 6s + 5 = 0$$

so that

$$\tau = \frac{1}{\sqrt{5}} \qquad \zeta = \frac{3}{\sqrt{5}}$$

TABLE 25.1
Classification of critical points

Type of critical point	Characteristic equation	Pertinent values of ζ	Nature of roots	Sign of roots
Stable node	$\tau^2 s^2 + 2\zeta\tau s + 1 = 0$	$\zeta > 1$	Real	Both $-$
Stable focus	$\tau^2 s^2 + 2\zeta\tau s + 1 = 0$	$0 < \zeta < 1$	Complex	Real parts both $-$
Unstable focus	$\tau^2 s^2 + 2\zeta\tau s + 1 = 0$	$-1 < \zeta < 0$	Complex	Real parts both $+$
Unstable node	$\tau^2 s^2 + 2\zeta\tau s + 1 = 0$	$\zeta < -1$	Real	Both $+$
Saddle point	$\tau^2 s^2 + 2\zeta\tau s - 1 = 0$	All	Real	One $+$, one $-$

has a critical point such as that of Fig. 25–2a. Here the trajectories enter the critical point directly, without oscillation. This type of critical point is called a node. For comparison, a typical focus is sketched in Fig. 25–2b. In fact, other types of behavior may be exhibited by critical points of a second-order system, depending on the nature of the roots of the characteristic equation. These are summarized for *linear* systems in Table 25.1 and sketched in Fig. 25–2. The distinction between stable and unstable nodes or foci is
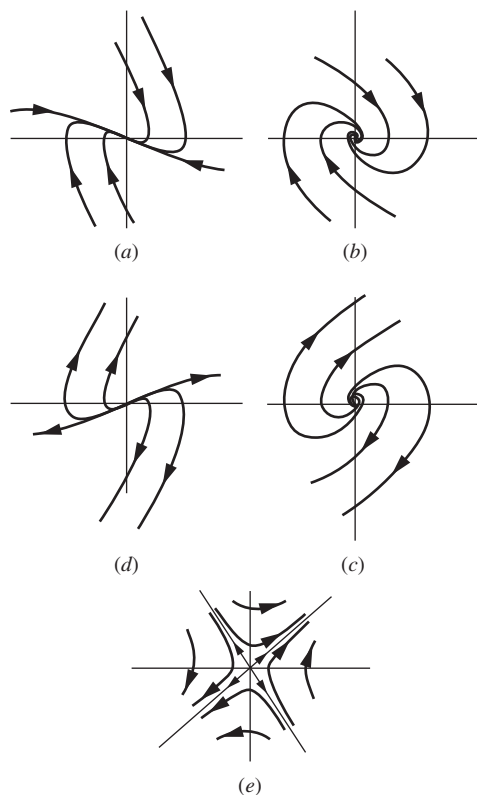
FIGURE 25–2
Second-order critical points: (*a*) stable node, (*b*) stable focus, (*c*) unstable focus, (*d*) unstable node, (*e*) saddle point.

made to indicate that the trajectories move toward the stable type of critical point and away from the unstable point. The saddle point arises when the roots of the characteristic equation are real and have opposite sign. In this case there are only two trajectories that enter the critical point, and after entering, the trajectories may leave the critical point (permanently) on either of two other trajectories. No other trajectory can enter the critical point, although some approach it very closely.

This categorization of critical points according to the particular linear system is often of value in the analysis of nonlinear systems. The reason for this is that, in a sufficiently small vicinity of a critical point, a nonlinear system behaves approximately linearly. Thus, the system of Eq. (24.21) for the pendulum is nonlinear. It has two physically distinguishable steady states, corresponding to the pendulum pointing up or down. The nonlinear term $\sin \theta$ may be linearized around each steady state. Near the steady state at $\theta = 0$,

$$\sin \theta \approx \sin \theta|_{\theta=0} + \frac{d(\sin \theta)}{d\theta}\bigg|_{\theta=0} (\theta - 0) + \text{higher-order term}$$

$$\sin \theta \approx 0 + \underbrace{\cos \theta|_{\theta=0}}_{1} (\theta - 0) + \text{higher-order terms}$$

$$\sin \theta \approx \theta$$

and near the steady state at $\theta = \pi$, a Taylor series yields

$$\sin \theta = -(\theta - \pi)$$

Therefore, near $\theta = 0$, Eqs. (24.21) are closely approximated by the linear equations

$$\frac{d\theta}{dt} = \dot{\theta}$$

$$\frac{d\dot{\theta}}{dt} = -\omega_n^2 \theta - D\dot{\theta}$$

(25.4)

and near $\theta = \pi$, by

$$\frac{dx}{dt} = \dot{x}$$

$$\frac{d\dot{x}}{dt} = \omega_n^2 x - D\dot{x}$$

(25.5)

where $x = \theta - \pi$. These linearized versions of Eqs. (24.10) can be easily solved to determine the nature of the *linear approximations* to the critical points. Thus, the characteristic equation for Eqs. (25.4) is

$$s^2 + Ds + \omega_n^2 = 0$$

(25.6)

while that for Eqs. (25.5) is

$$s^2 + Ds - \omega_n^2 = 0$$

(25.7)

As shown in Table 25.1, Eq. (25.6) yields a stable critical point, which may be a node or focus depending on the degree of damping. (Note that as the damping is increased, the behavior changes from focus to node, or from oscillatory to nonoscillatory.) On the other hand, Eq. (25.7) indicates a saddle point for the motion near $\theta = \pi$.

These conclusions apply strictly only to the linearized phase equations, Eqs. (25.4) and (25.5). To compare them with the behavior of the true system of Eqs. (24.10), the actual phase diagram is sketched for a lightly damped case in Fig. 25–3. For simplicity, this diagram is extended beyond the range $0 \le \theta \le 2\pi$ even though this is the only region of physical significance. Actually, the section for $0 \le \theta \le 2\pi$ should be cut out and rolled into a cylinder so that the lines corresponding to $\theta = 0$ and $\theta = 2\pi$ coincide. This phase cylinder would more realistically represent the motion of the pendulum. As seen from Fig. 25–3, the point at $\theta = \pi$ is, indeed, a saddle point and the point $\theta = 0$ (or 2π) is a stable focus. If the system were more heavily damped, this latter point would be a stable node.
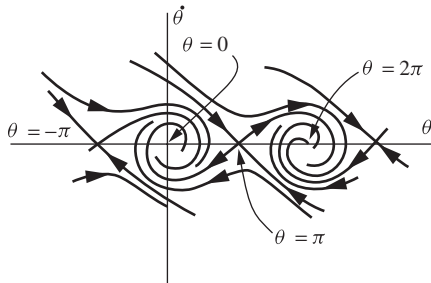
FIGURE 25–3
Phase portrait of lightly damped pendulum.

A greater understanding of the saddle point may now be obtained by analyzing the $\theta = \pi$ point in terms of what we know to be the physical behavior of the pendulum at this point. That is, the point may be approached from either of two directions. When the pendulum is at the point, an infinitesimal disturbance will cause it to fall in either of two directions. Other trajectories narrowly miss this point, indicating that just the right initial velocity must be imparted to the pendulum at a given initial point to cause it to stop in the $\theta = \pi$ position.

In summary, it can be concluded that in this case the linearized equations give valuable, accurate information about the behavior of the nonlinear system in the vicinity of the critical points. Because the linearized equations are more easily solved, it is always desirable to be able to relate the behavior of the actual system to the behavior of the linearized solutions in the vicinity of the operating point. In fact, in our previous work on control systems, we have assumed for nonlinear systems that design of a stable control system based on the linearized equations was adequate to ensure stable operation of the actual system. The basis for this assumption is given by the following theorem of Liapunov (see Letov, 1961).

Let the nonlinear equations of a motion be linearized by expansion in deviation variables around a particular critical point. If the linearized solution for the deviation variables is stable, the actual motion will be stable in some vicinity of

the critical point. If the linearized solution is neutrally stable (i.e., its characteristic equation has roots on the imaginary axis), no statement can be made about the actual motion. If the linearized solution is unstable, then the actual motion will be unstable.

It is necessary to define what is meant by stability and instability of the actual nonlinear motion in the vicinity of the critical point. Although stability in nonlinear systems is a complex subject, for our purposes it will suffice to state that a stable nonlinear motion in the vicinity of a critical point is one for which all phase-plane trajectories in this vicinity travel toward and end at the critical point. An unstable motion is one for which trajectories move away from the critical point. This would mean that while theoretically the state of the system may remain at the critical point indefinitely, any slight disturbance causes the unstable system to move away from the critical point. These conclusions agree with our physical understanding of the pendulum motion, since the steady condition at $\theta = \pi$ is easily destroyed.

It is because of Liapunov's theorem that linear control theory is so successful in control system design. One really hopes to control the system so that it remains permanently in the vicinity of a particular point (i.e., a steady state). However, when serious upsets occur in an automatically controlled plant, moving it far from steady state, it is often necessary to return the plant to manual control until conditions are again close to steady state. This is so because the controllers are designed for satisfactory operation in the linear range only. One of the great drawbacks of linear control theory is the fact that stability of the linearized equations guarantees stability of the nonlinear system only in *some* vicinity of the particular critical point. No information about the size of this vicinity or about the behavior outside this vicinity is obtained. If the linear vicinity is extremely small, then unknown to the designer who has used linear methods, almost any plant disturbance of practical size may result in control system failure. An example of this behavior will be given later.

Limit Cycles

The first major difference between linear and nonlinear motions is the possible existence of more than one critical point in the latter type. The second is the possible existence of limit cycles.

A *limit cycle* is defined as a periodic oscillation whose amplitude and frequency depend only on the properties of the system and not on the initial state of the system (provided the initial state lies in a certain nontrivial region of the phase space). In the phase plane, *stable* limit cycles are recognized as closed curves which are approached asymptotically by all nearby trajectories. *Unstable* limit cycles are closed curves from which all nearby trajectories diverge. An example of a stable limit cycle is the "steady-state" behavior of a home heating system when controlled by a thermostat. A periodic oscillation in house temperature is always reached, and the amplitude and frequency of the oscillation are independent of the temperature that existed in the house at the time that the furnace was started. Unstable limit cycles can never be realized physically for any system by definition. However, as will be seen later, they divide the phase plane into regions of totally different dynamic behavior and hence are of considerable importance.

It is important to distinguish between limit cycles and other closed curves that may occur. The linear system

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

has phase-space solution

$$x^2 + \tau^2 (\dot{x})^2 = C^2 \tag{25.8}$$

where $\dot{x} = dx/dt$ and the constant C depends on initial conditions. Equation (25.8) defines a family of concentric ellipses in the phase plane. However, *these are not limit cycles,* because the closed curve which is followed by the system depends on the initial state of the system through the constant C. In the next section, we study some limit cycles occurring in typical control systems.

OTHER ASPECTS. We have presented only those aspects of phase-plane analysis that will be of use in the examples to follow. This can be considered only as a brief introduction to the subject, and the interested reader is referred to the references already cited for more information. Among the important subjects that have been omitted are graphical methods for determination of time along a trajectory, various aspects of phase-plane topology, and the mathematical aspects of stability.

25.2   EXAMPLES OF PHASE-PLANE ANALYSIS

In this section, we consider two different examples of the use of the phase plane to analyze nonlinear control systems. The first is a simple on/off control system for a stirred-tank heater. The second is the chemical reactor of Chap. 24. In both cases, the systems are second-order and autonomous, so that they are ideal situations for use of the phase plane.

On/Off Control of Stirred-tank Heater

The use of on/off control offers significant economic advantages over proportional control or other more sophisticated modes of control. The control mechanism is simply a relay that turns on or off depending on the value of the measured variable. The disadvantage is usually that the quality of control is inferior to that realized with proportional control.

Consider the stirred-tank heater of Fig. 25–4. Water is being heated to a controlled temperature by mixing with steam. It is assumed for the analysis that the cold water input rate is constant. Heated water overflows into an outlet pipe at the top of the tank, so that no accumulation of mass occurs in the tank. Most of the steam is added, at a fixed flow rate, from the main steam supply. However, this amount of steam is set at a value somewhat less than the amount required to heat the cold water to the desired temperature. An additional amount of steam may be added whenever the solenoid valve is opened. When this additional steam is admitted, the sum of the two steam inputs is enough to heat the water to a temperature somewhat in excess of the desired
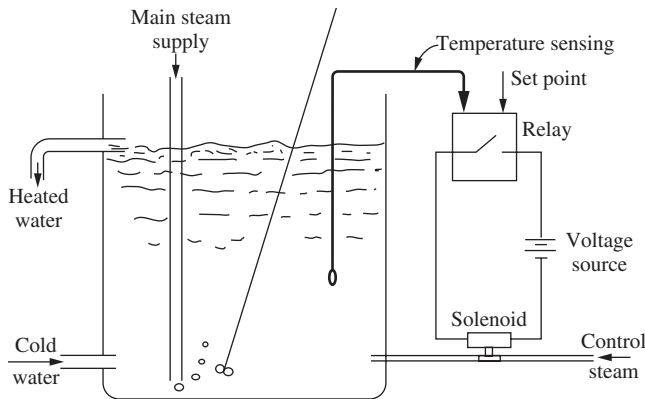
FIGURE 25–4
On/off control of a stirred-tank heater.

temperature. A temperature-measuring device such as a thermocouple or RTD trans-
mits the tank temperature to the relay. When this temperature is below the set point, the
relay closes, which opens the solenoid valve, thus admitting more steam. Eventually,
the additional steam will result in the temperature exceeding the set point, the relay
will open, the valve will close, cutting off the additional steam, and the temperature
will fall again.

It is apparent that an oscillating control will be achieved. In fact, from the discus-
sion in the previous section, we recognize that a limit cycle will occur. We consider
now a numerical example of this type of control system.

Water at 40°F, at a rate of 100 lb/min, is to be heated to 150°F. The main steam
supply is to be set so that it will heat this much water to 125°F, while additional steam,
through the controlled solenoid valve, is available to heat the water another 50°F. This
means that the steady-state temperatures with the solenoid closed and open, 125 to
175°F, are equally spaced about the set point. Heat losses to the surroundings are negli-
gible. The volume of the tank is 1.6 ft^3. The relay control system has a thermocouple for
measurement of temperature. This measuring system has a time constant of 30 s. The
solenoid valve is very rapid in response.

We first analyze this system considering the relay to behave ideally. This means
that it opens precisely at the instant the temperature exceeds the set point and closes
similarly. Later, we shall correct this to conform more closely to the behavior of
actual relays.

If the tank is perfectly stirred, it is a first-order system with a time constant of

$$\tau = \frac{\rho V}{\omega} = \frac{(62)(1.6)}{100} = 1.0 \, \text{min}$$

and its transfer function relating changes in the steam input rate to temperature is

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

where 10°F/lb/min of steam flow is the change in steady-state temperature per unit change in steady-state steam flow. The necessary main and controlled steam rates are (using 1000 Btu/lb for latent heat)

$$Q_{main} = \frac{(125 - 40)(100)}{1000} = 8.5 \text{ lb/min}$$

$$Q_{controlled} = \frac{(175 - 125)(100)}{1000} = 5.0 \text{ lb/min}$$

The amount of steam that would be necessary to maintain the water at a steady-state temperature of 150°F is

$$Q_s = \frac{(150 - 40)(100)}{1000} = 11.0 \text{ lb/min}$$

Hence, in terms of deviation variables, the controller output may be taken as $\pm$ 2.5 lb/min of steam.

A block diagram may now be constructed for this system, as shown in Fig. 25–5. This diagram uses deviations from 150°F as temperature variables, so the set point is taken as zero. The action of the relay is symbolized by the input-output relations, indicating that +2.5 lb/min of steam is admitted when the error is positive and −2.5 lb/min when the error is negative, again in deviation variables.
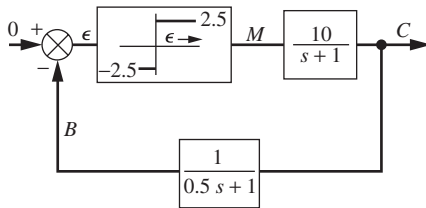
FIGURE 25–5
Block diagram for system of Fig. 25–4.

It is convenient to use a dimensionless version of Fig. 25–5. This is provided in Fig. 25–6, where the changes

$$M' = \frac{M}{2.5} \qquad \varepsilon' = \frac{\varepsilon}{25} \qquad C' = \frac{C}{25} \qquad B' = \frac{B}{25}$$

have been made.
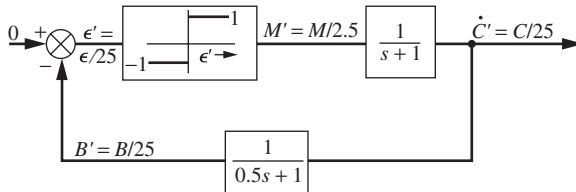
FIGURE 25–6
Dimensionless block diagram for system of Fig. 25–4.

The usual methods of linear control theory are not applicable to the block diagram of Fig. 25.6. The relay does not obey the principle of superposition in its input-output relation. It is necessary to revert to the differential equations describing the control loop. These are

$$M' = \frac{dC'}{dt} + C' \tag{25.9}$$

$$C' = \frac{1}{2}\frac{dB'}{dt} + B' \tag{25.10}$$

$$\varepsilon' = -B' \tag{25.11}$$

In addition we have

$$M' = \begin{cases} 1 & \varepsilon' > 0 \\ -1 & \varepsilon' < 0 \end{cases} \tag{25.12}$$

Combination of Eqs. (25.9) to (25.12) yields

$$\frac{1}{2}\frac{d^2\varepsilon'}{dt^2} + \frac{3}{2}\frac{d\varepsilon'}{dt} + \varepsilon' = \begin{cases} -1 & \varepsilon' > 0 \\ 1 & \varepsilon' < 0 \end{cases} \tag{25.13}$$

Equation (25.13) can be rewritten in phase notation as

$$\frac{d\varepsilon'}{dt} = \dot{\varepsilon}'$$

$$\frac{d\dot{\varepsilon}'}{dt} = \begin{cases} -\left(3\dot{\varepsilon}' + 2\varepsilon' + 2\right) & \varepsilon' > 0 \\ -\left(3\dot{\varepsilon}' + 2\varepsilon' - 2\right) & \varepsilon' < 0 \end{cases} \tag{25.14}$$

Figure 25–7 is a plot of the phase plane for this system. Note that, in this figure, the ε' scale has been expanded by a factor of 10 to magnify the behavior near the origin.

It can be seen from Fig. 25–7 that the trajectory approaches the origin. The final state is a limit cycle of zero amplitude and infinite frequency about the origin. In other words, the relay alternately opens and closes at very high frequency, a condition known as *chattering*.

Physically, this condition will never be realized because the dynamics of the sole-noid valve and the relay itself would become important. Instead, the final condition will be a limit cycle of high, rather than infinite, frequency and low, rather than zero, amplitude.

However, the basic idealization which has led us to this suspect conclusion is in the behavior of the relay. True relays have input-output characteristics more similar to those shown in Fig. 25–8. There is a dead band around the set point, of width $2\varepsilon'_0$, over which the relay is insensitive to changes in the error signal. Anyone who has made fine adjustments in the setting of a home thermostat has observed this behavior.
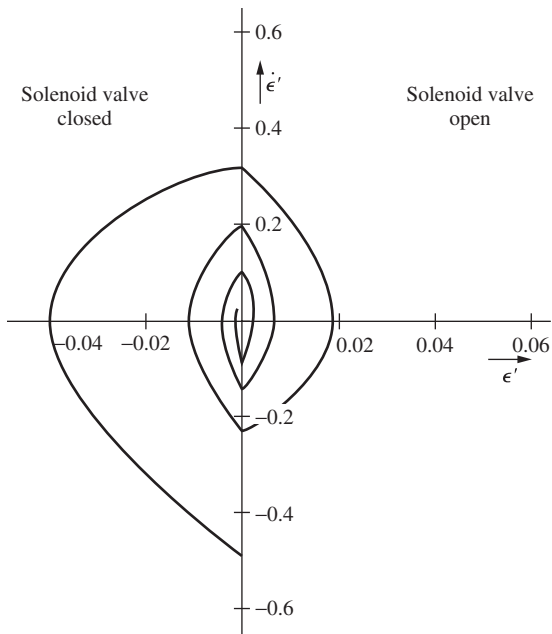
FIGURE 25–7
Phase-plane trajectory for on-off control system of Fig. 25–4.
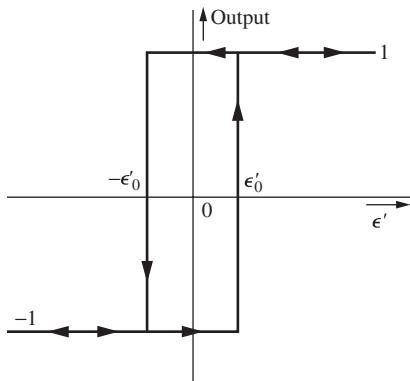
FIGURE 25–8
Characteristics of true relay with dead zone.

Consider as an example the case for $\varepsilon'_0 = 0.01$. The effect of this dead zone is to change the dividing line between on and off to that shown in Fig. 25–9. The new dividing line has the equation

$$\varepsilon' = \begin{cases} 0.01 & \dot{\varepsilon}' > 0 \\ -0.01 & \dot{\varepsilon}' < 0 \end{cases}$$
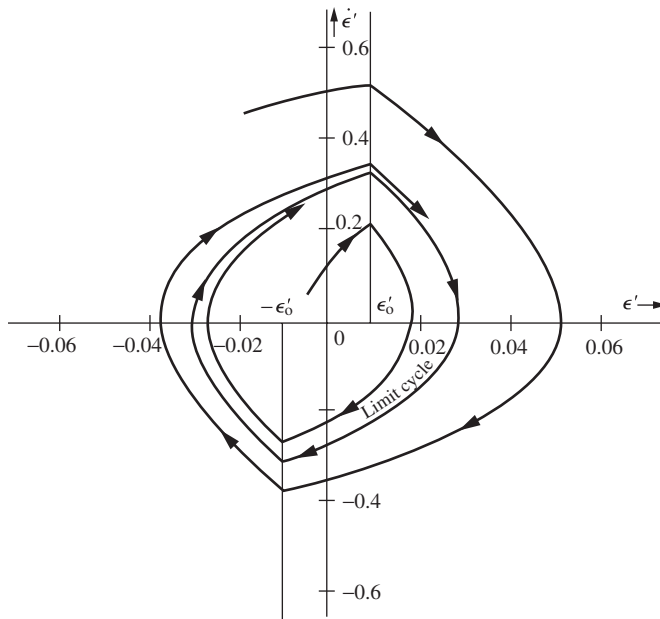
FIGURE 25–9
Phase plane for system of Fig. 25–4 using relay with characteristics of Fig. 25–8.

Now, as shown in Fig. 25–9, all trajectories approach a limit cycle, for which the error amplitude is approximately 0.03. The frequency is finite and is obtained by computing the time around the limit cycle. Although we have not presented here the graphical methods for determining this time, it can always be calculated by noting from the first of Eqs. (25.14) that

$$t = \int dt = \int \frac{d\varepsilon'}{\dot{\varepsilon}'} \tag{25.15}$$

Thus, time around the limit cycle can be computed by graphical evaluation of the integral in Eq. (25.15). The only difficulty is near the ε' axis, where $\dot{\varepsilon}'$ goes to zero. To circumvent this, we may use the second of Eqs. (25.14)

$$t = -\int \frac{d\dot{\varepsilon}'}{3\dot{\varepsilon}' + 2\varepsilon' \pm 2}$$

over a small segment of the trajectory as it crosses the ε' axis. The result of this graphical calculation is $\omega = 9.2$ rad/min.

The frequency thus computed for the error signal is, for obvious physical reasons, the same as the frequency of the controlled signal C'. However, the amplitude of C', which is of more direct interest, is not the same as the amplitude of ε'. It may be found in this case by noting from Eqs. (25.10) and (25.11) that

$$C' \approx -\tfrac{1}{2}\dot{\varepsilon}' - \varepsilon'$$

It is therefore clear from Fig. 25–9 that C' attains a maximum value near the switching points where

$$C' \approx \pm 0.17$$

Reverting to the original variables, it follows that the water temperature will oscillate with an amplitude of

$$(0.17)(25) = 4.25°F$$

The effect of a small dead zone, $2\varepsilon_0 = 2(0.01)(25) = 0.5°F$, is thus quite significant.

In practice, the width of this dead zone is usually an adjustable design parameter. This width is always chosen as a compromise. The wider it is made, the lower will be the limit-cycle frequency, thus saving excessive switching or chatter. However, the limit-cycle amplitude increases with dead-zone width, decreasing the quality of control.

Example 25.2.   Use of Simulink to generate a phase-plane diagram for a relay with dead zone. Use MATLAB and Simulink to simulate the control system shown in Fig. 25–6 containing a relay with characteristics shown in Fig. 25–8.

The Simulink block diagram for the system is shown in Fig. 25–10.
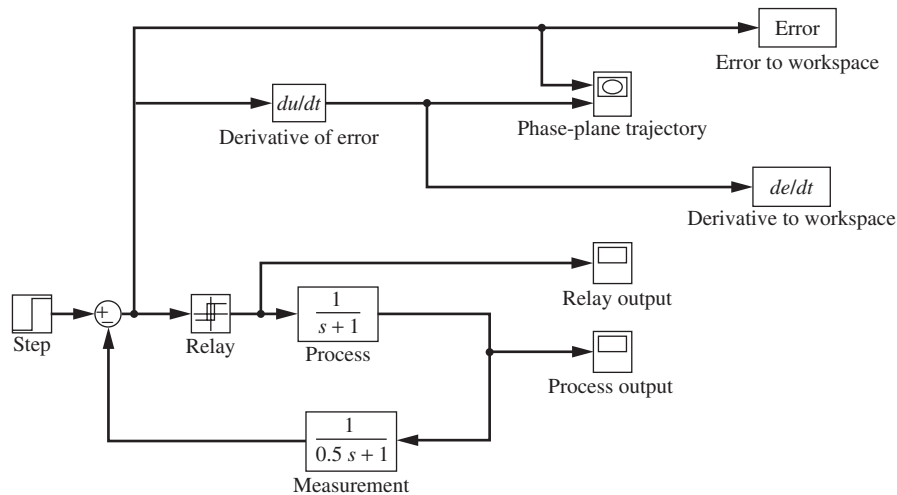
FIGURE 25–10
Simulink block diagram for process in Fig. 25–6.

The phase-plane trajectory for Example 25.2 is shown in Fig. 25–11.

We can also determine the cycle time by looking at the relay output as shown in Fig. 25–12.
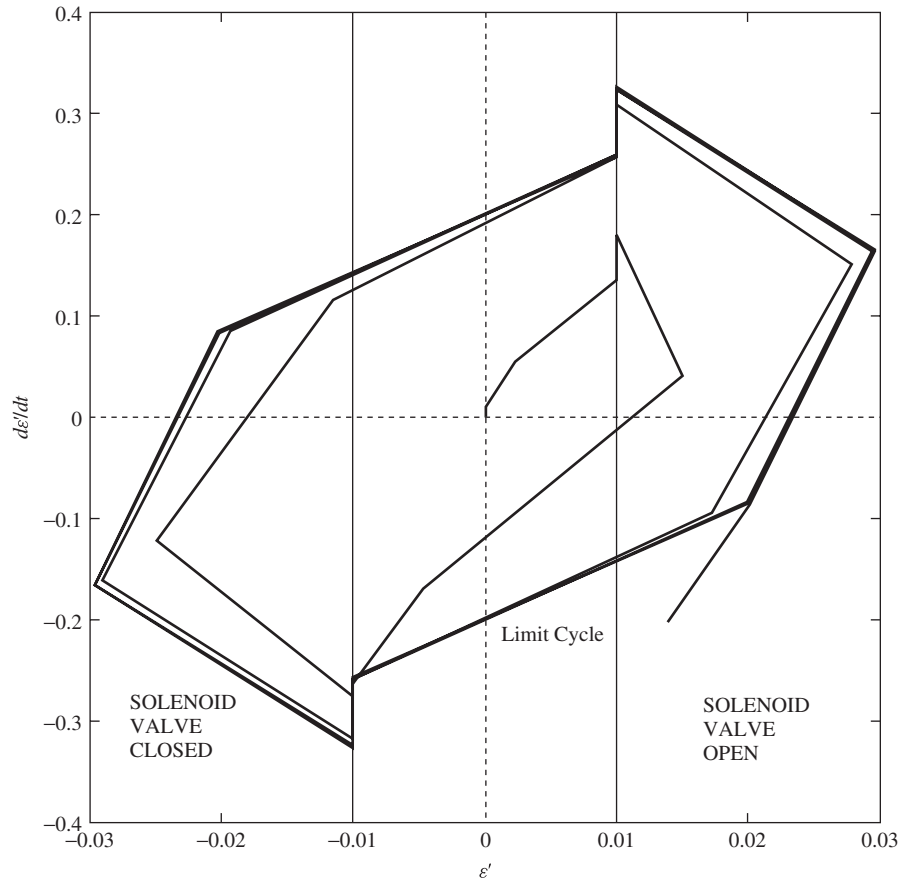
FIGURE 25–11
Phase plane generated using Simulink for Example 25.2.
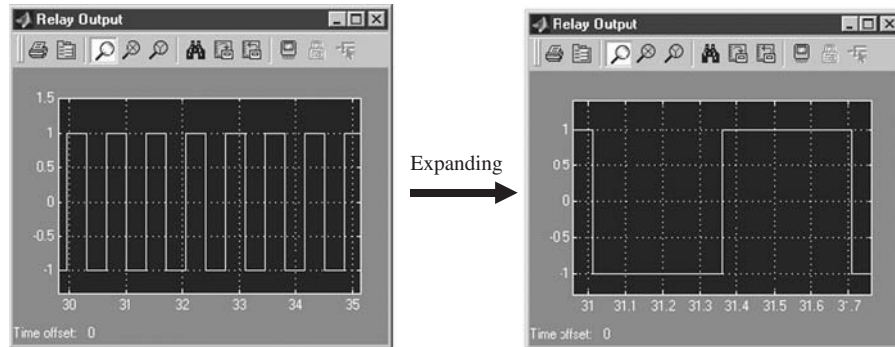
FIGURE 25–12
Relay ouput from Simulink model.

The relay cycle time from the graph is 0.7 min/cycle. The frequency is

$$\omega = \left[\frac{2\pi \text{ rad}}{\text{cycle}}\right]\left[\frac{1 \text{ cycle}}{0.7 \text{ min}}\right] = 9.0\frac{\text{rad}}{\text{min}}$$

This is the same result that we found earlier. We can also determine the maximum and minimum values of C' from the Simulink process output, as shown in Fig. 25–3.
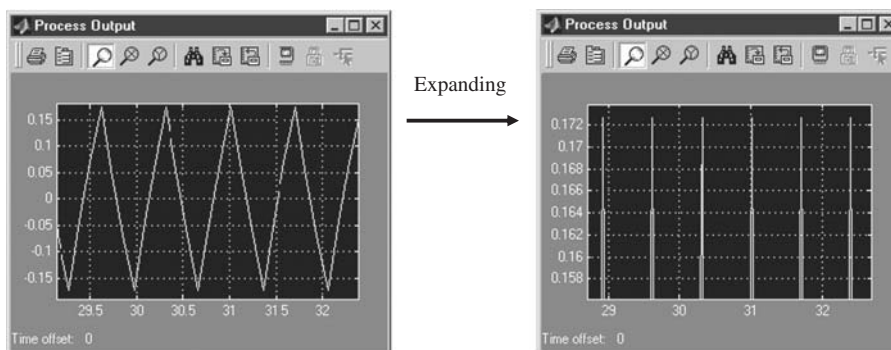
FIGURE 25–13
Minimum and maximum values of C' from Simulink.

From the output, we can see that $C' \approx \pm 0.173$, which is also the same as our earlier result. Simulink provides us with an easy way to study the behavior of these types of systems.

Exothermic Chemical Reactor

We now wish to consider the phase-plane behavior of the chemical reactor of Chap. 24. This study is based on the paper by Aris and Amundson (1958). For convenience, the dynamic equations are reproduced here:

$$\frac{dx_A}{dt} = \frac{F}{V}\left(x_{A_0} - x_A\right) - ke^{-E/RT}x_A$$

$$\frac{dT}{dt} = \frac{F}{V}\left(T_0 - T\right) + \frac{k\left(\Delta H\right)e^{-E/RT}}{C_p}x_A - \frac{Q\left(T\right)}{\rho V C_p} \qquad (24.28)$$

Defining the dimensionless variables

$$\tau = \frac{Ft}{V} \qquad y = \frac{x_A}{x_{A_0}} \qquad \theta = \frac{C_P T}{x_{A_0}\left(\Delta H\right)} \qquad \theta_0 = \frac{C_P T_0}{x_{A_0}\left(\Delta H\right)}$$

these equations become

$$\frac{dy}{d\tau} = 1 - y - r(y,\theta)$$

$$\frac{d\theta}{d\tau} = \theta_0 - \theta + r(y,\theta) - q(\theta)$$

(25.16)

where

$$r(y,\theta) = \frac{kVy}{F}e^{[-EC_P/Rx_{A_0}(\Delta H)\theta]}$$

$$q(\theta) = \frac{Q(T)}{F\rho x_{A_0}(\Delta H)}$$

As a control heat removal function $q(\theta)$, Aris and Amundson chose the form

$$q(\theta) = U(\theta - \theta_c)\left[1 + K_c(\theta - \theta_s)\right]$$

(25.17)

where θ_c is the dimensionless mean temperature of water in the cooling coil. This indicates that the heat removal is always proportional to the difference between the reactor temperature and mean cooling water temperature. In addition, the term in brackets indicates that proportional control on the cooling water flow rate is present. The flow rate is increased by an amount proportional to the difference between the actual reactor temperature θ and the desired steady-state temperature θ_s. This increase in cooling water flow rate is assumed for convenience to cause an approximately proportional increase in heat removal. The constant U is a dimensionless analog of U_0A, the overall heat-transfer rate.

As a specific numerical example, Aris and Amundson selected the following values for constants:

$$\frac{kV}{F} = e^{25}$$

$$\frac{EC_p}{Rx_{A_0}(\Delta H)} = 50$$

$$\theta_s = 2$$

$$\theta_0 = \theta_c = 1.75$$

$$U = 1$$

Under these conditions, Eqs. (25.16) become

$$\frac{dy}{d\tau} = 1 - y - ye^{50(1/2 - 1/\theta)}$$

$$\frac{d\theta}{d\tau} = 1.75 - \theta + ye^{50(1/2 - 1/\theta)} - (\theta - 1.75)\left[1 + K_c(\theta - 2)\right]$$

(25.18)

It can be seen that there is a critical point of Eqs. (25.18) at

$$y = \tfrac{1}{2} = y_s \qquad \theta = 2 = \theta_s$$

and this is the location at which control is desired. This point has the correct steady-state temperature and a 50 percent conversion of reactant. In addition, there may be two

more critical points of Eq. (25.18) depending on the proportional control constant K_c, as will be discussed below.

Since we are primarily interested in control about θ_s, we make use of Liapunov's theorem on local stability, presented earlier. Linearizing Eq. (25.18) in deviation variables $\theta - \theta_s$ and $y - y_s$ by using Taylor's series yields

$$\frac{d(y - y_s)}{d\tau} = -2(y - y_s) - 6.25(\theta - \theta_s)$$

$$\frac{d(\theta - \theta_s)}{d\tau} = y - y_s + \left(4.25 - \frac{K_c}{4}\right)(\theta - \theta_s)$$

(25.19)

where $y_s = \frac{1}{2}$. As we have seen before, the solution to this linear system is

$$y - y_s = c_1 e^{s_1 t} + c_2 e^{s_2 t}$$

$$\theta - \theta_s = c_3 e^{s_1 t} + c_4 e^{s_2 t}$$

where, in this case, s_1 and s_2 are the roots of [see Eq. (24.6) and the steps following this equation]

$$s^2 + \frac{K_c - 9}{4} s + \frac{2K_c - 9}{4} = 0$$

(25.20)

According to the Routh criteria, all coefficients in this characteristic equation must be positive in order that the real parts of the roots s_1 and s_2 be negative. Hence, we can see immediately from Eq. (25.20) that to achieve a stable node or focus, it is necessary that $K_c > 9$.
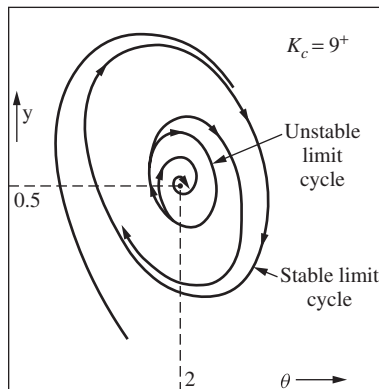
FIGURE 25–14
Stable and unstable limit cycles in an exothermic chemical reactor.

However, Aris and Amundson obtained the phase plane for (among other values) a value of K_c slightly greater than 9. This was accomplished by numerical solution of Eqs. (25.18). It was found that in the vicinity of the steady-state point, the situation is as depicted in Fig. 25–14. There are two limit cycles surrounding the stable focus critical point. The inner limit cycle is unstable, and the outer limit cycle is stable, according to the definitions given earlier. It may be seen that any disturbance (or initial condition) which moves the system no further from the critical point than the unstable limit cycle can be controlled. That is, the control system will eventually bring the system back to steady state. However, once the system is forced outside this limit cycle, it will eventually spiral out to the stable limit cycle. Control cannot be restored, and the reactor temperature and

concentration oscillate continuously. This example illustrates very well the limitations of linear control theory. All that the linear investigation could reveal is that, for $K_c > 9$, the system will be stable in some vicinity of the control point. The phase-plane analysis shows that for K_c slightly greater than 9, this vicinity is inside the unstable limit cycle of Fig. 25–14. If K_c is increased further, the two limit cycles disappear and good control can be achieved. This example points out the importance of unstable limit cycles. Although a physical system can never follow an unstable limit cycle, the limit cycle divides the phase plane into distinct dynamic regions for the physical system.

Other values of K_c were analyzed by Aris and Amundson. For low values of K_c, there are two other critical points besides the control point. For example, for $K_c = 0.8$, there are critical points at

$$y = 0.95 \qquad \theta = 1.77$$

and

$$y = 0.15 \qquad \theta = 2.15$$

Linear analysis shows that both these are stable, but for $K_c < 9$ the control point ($y = 0.5$, $\theta = 2$) is not. Phase-plane analysis shows that if the reactor is started at high temperatures, it will come to steady state at the high-temperature critical point, and vice versa. Starting the reactor at the desired control point will be of no avail, as it will leave and go to one of the other steady-state points, depending on the direction of the initial disturbance. For high values of K_c, there is only one critical point, which is at the control point. Phase-plane analysis shows that K_c must exceed approximately 30 before rapid return to steady state at the desired control point, following all disturbances, is achieved. Some phase-plane portraits for this system that were obtained by means of a computer are shown in Figs. 25–15 to 25–17.
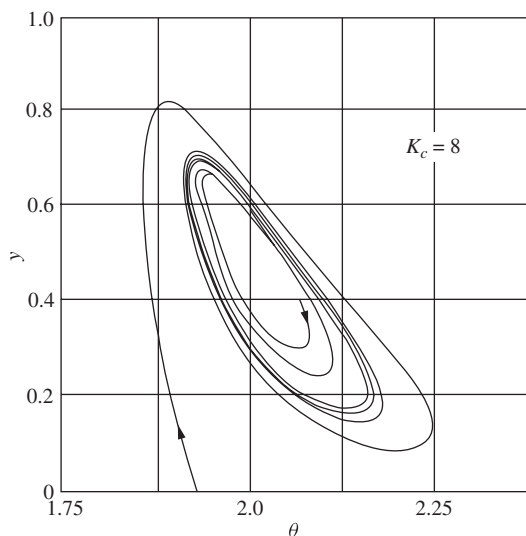
FIGURE 25–15
Phase-plane portrait of the control of a chemical reactor (limit cycle forms).
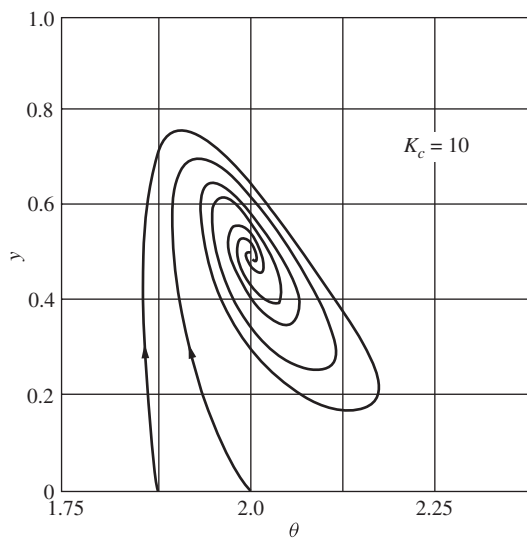
FIGURE 25–16
Phase-plane portrait of the control of a chemical reactor (no limit cycle forms).
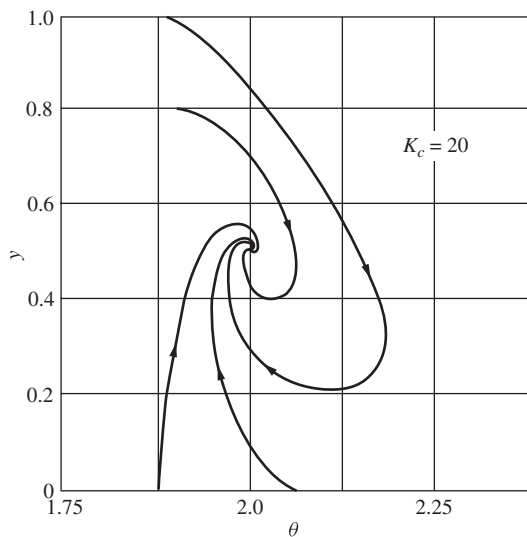
FIGURE 25–17
Phase-plane portrait of the control of a chemical reactor (no limit cycle forms).

This discussion is only a rather brief introduction to the extensive work by Aris and Amundson. The reader is strongly urged to consult the original paper for a more comprehensive treatment of the problem.

SUMMARY

We have seen that phase-plane analysis can be used for two typical nonlinear control problems. The results of this analysis give extensive information about the control system behavior.

PROBLEMS

25.1. For the system shown in Fig. P25–1, use Simulink to plot the trajectory from the initial point $e = 2$ and $\dot{e} = 0$.
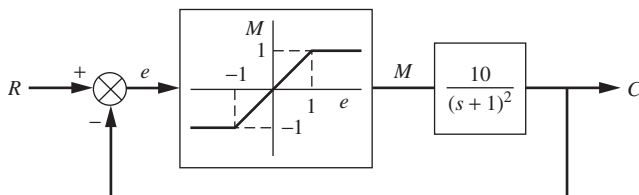
FIGURE P25–1

25.2. For the system shown in Fig. P25–2, plot the phase-plane trajectory. Use $\tau = 1$, $\zeta = \frac{1}{2}$, $K = 1$, $b = 0.25$. Let $R = 0$.
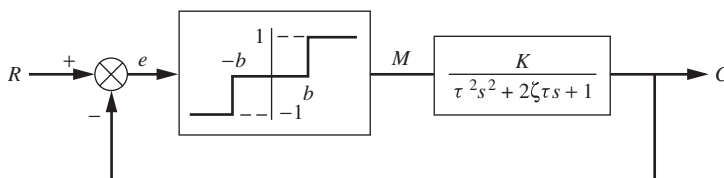
FIGURE P25–2

25.3. Consider the phase-plane equations

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

$$\dot{x}_2 = -x_1 + \frac{1}{10}x_2 - x_1^2 - \frac{10}{3}x\left(\frac{3}{2}\right)$$

(*a*) Determine the type of critical point at $x_1 = -1$, $x_2 = 0$.
(*b*) If there are any other critical points, find them.

25.4. The system shown in Fig. P25–4 is to be controlled by an ideal on/off relay.

(*a*) From the block diagram, write the differential equations for phase-plane description of the physical system in the form

$$\dot{x}_1 = f(x_1, x_2) \qquad \dot{x}_2 = f(x_1, x_2)$$

where $x_1 = c$ and $x_2 = \dot{c}$.
(*b*) Plot the trajectory that starts at $x_1 = 2$, $x_2 = 0$.
(*c*) Determine the values of x_1 and x_2 where the first switch occurs.
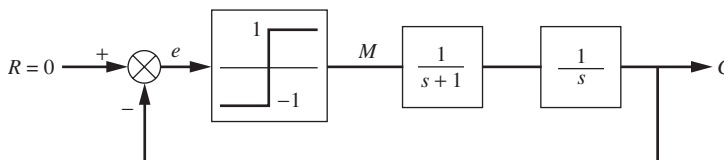
FIGURE P25–4

25.5. For the control system shown in Fig. P25–5 determine the frequency and amplitude of the limit cycle if one exists.
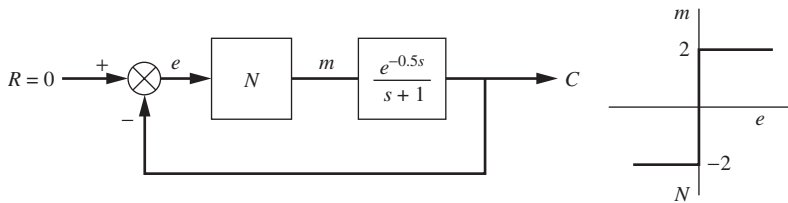
FIGURE P25–5

25.6. For the system shown in Fig. P25–6, $\tau = 1$, $\zeta = \frac{1}{2}$, $K = 1$, $M = 1$, $b = 0.25$, and $R = 0$.
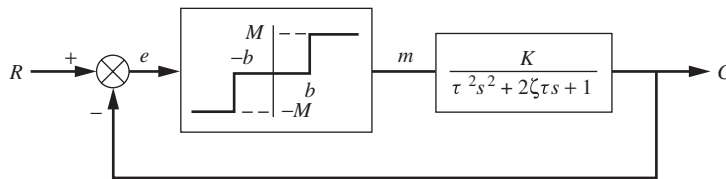
FIGURE P25–6

(*a*) For $K = 2$, does a limit cycle exist? If so, describe it.

(*b*) If a transport lag e^{-s} is introduced in the feedback loop, determine if a limit cycle exists for $K = 2$.

25.7. The stirred-tank system shown in Fig. P25–7 produces an aqueous solution of salt by use of a solenoid valve that switches from one reagent tank to the other as described below. The reagent tanks contain concentrated solutions of salt. When the measured concentration is above the set point, the control reagent of lower concentration enters the mixing tank at a constant flow rate of 0.01 L/min. When the measured concentration is less than the set point, the control reagent of higher concentration enters the mixing tank at a constant rate of 0.01 L/min. The holdup volume of the tank is 2 L, the transport lag between the tank and measuring element is 1.2 min, and the set point is 2 g salt/L.
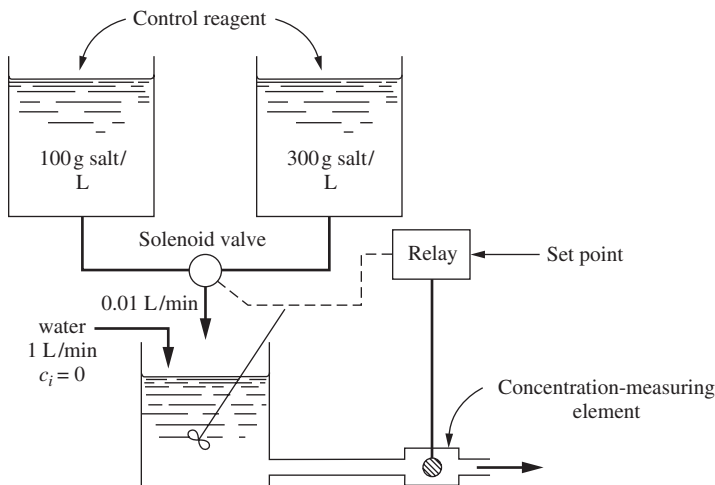
FIGURE P25–7

(*a*) Obtain a block diagram, in terms of deviation variables, for this control system.

(*b*) Determine the characteristics of the limit cycle (frequency and amplitude), if one exists.

25.8. For the control system shown in Fig. P25–8, determine if a limit cycle exists for $K = 1, 2,$ and 3. If a limit cycle exists, describe it in terms of amplitude and frequency.
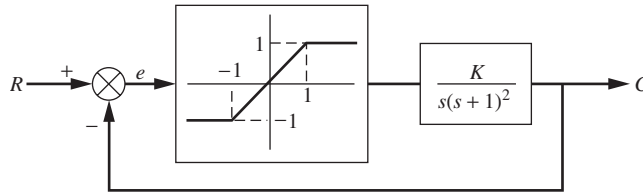
FIGURE P25–8

CAPSULE SUMMARY

MATLAB and Simulink can be used to plot phase planes:
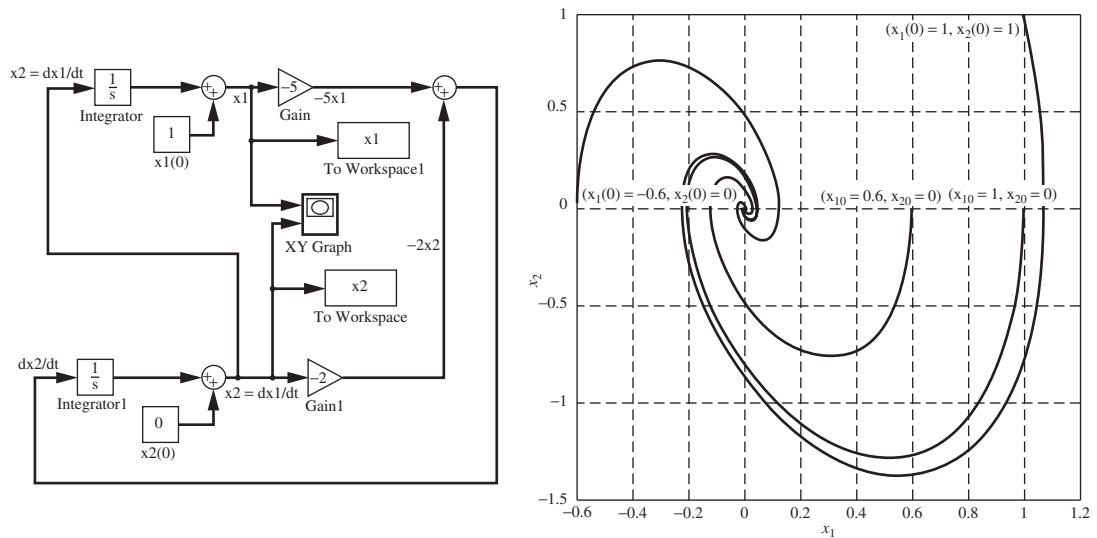
Limit cycle. A *limit cycle* is defined as a periodic oscillation whose amplitude and frequency depend only on the properties of the system and not on the initial state of the system (provided the initial state lies in a certain nontrivial region of the phase space). In the phase plane, *stable* limit cycles are recognized as closed curves that are approached asymptotically by all nearby trajectories. *Unstable* limit cycles are closed curves from which all nearby trajectories diverge. An example of a stable limit cycle is the "steady-state" behavior of a home heating system when controlled by a thermostat, an on/off controller.

Phase-Plane Trajectory for On/Off Control System

True Relay with Dead Zone
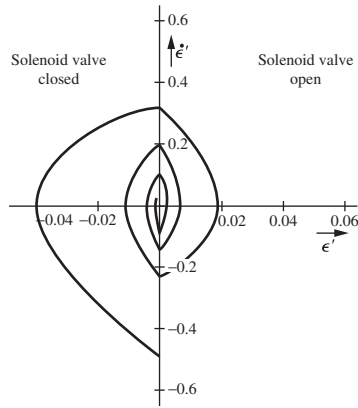
Phase-Plane Trajectory for Relay Control System
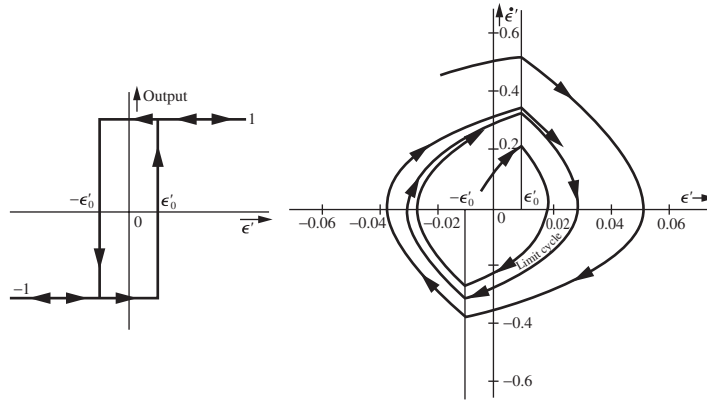
PART
VIII

COMPUTERS IN
PROCESS CONTROL

MICROPROCESSOR-BASED CONTROLLERS AND DISTRIBUTED CONTROL

In this chapter, some of the highlights of modern industrial microprocessor-based controllers and distributed control systems are presented. A microprocessor-based controller is essentially a computer programmed to perform the function of a process controller. For our purpose, the term *microprocessor* is synonymous with *computer,* and we could refer to a microprocessor-based controller as a computer-based controller. The number of features of these modern controllers is far too great to cover in one chapter. The best way for the reader to acquire some experience with modern controllers is through laboratory and plant use and by attending some of the short courses offered by the major suppliers of the equipment.

26.1   HISTORICAL BACKGROUND

During the past fifty years, tremendous development has occurred in process control hardware. The three phases of development are pneumatic control, electronic control, and microprocessor-based control. During the 1940s, the predominant controller was pneumatic, meaning that signals to and from the controller and within the controller mechanism were air-pressure signals that usually varied from 3 to 15 psig. The development of the high-gain operational electronic amplifier during World War II led to the development of the electronic controller and also the analog computer. The electronic controller mimicked the control functions of the pneumatic controller. It also provided some improvements, such as accurate and reproducible control parameter settings and reduction in size of the instruments. In contrast, the pneumatic controller required frequent calibration of the knobs used to set the various controller parameters (K_c, τ_I, τ_D). The pneumatic controller had interaction among the control modes and had inherent

lags that became significant at high-frequency operation. There were frequent debates over the pros and cons of pneumatic and electronic controllers. For example, the pneumatic controller was rugged, simple to install, and required little maintenance. Only a source of air pressure was needed to operate the controller. There was initially great concern about the possibility of explosions with the use of electronic controllers, so the instrument cases for these controllers were purged with steady streams of air when used in plants producing flammable substances. The maintenance of electronic controllers also required highly trained technicians.

In the 1960s, the chemical industry made its first attempt at computer process control. These control systems used large mainframe computers, for which the control programs had to be written from scratch. The first attempts at computer control were met with mixed reactions. In the 1970s, there appeared on the market the first generation of digital control hardware, which was based on the advances in microprocessor-based technology. This equipment was user-friendly, and all the software accompanied the hardware. The operator did not face the problem of writing computer code to implement the control functions; it was only necessary to learn the instructions needed to configure (set up) the controllers.

26.2   HARDWARE COMPONENTS

The hardware requirements for pneumatic, electronic, and microprocessor-based controls are shown in Fig. 26–1. In this figure, all the components are obtained from a manufacturer of control equipment; several of the components are common to the three systems. In Fig. 26–1a, all the signals are pneumatic (3 to 15 psig). The energy needed to operate these pneumatic components is a source of clean, dry air at a pressure of about 20 psig. The pressure can vary from 20 psig by about $\pm$ 10 percent without adversely affecting the operation of the instruments.

The electronic system shown in Fig. 26–1b requires both electrical and pneumatic power to operate the components. A transducer or converter is needed between the controller and the valve to convert current (4 to 20 mA) to pressure (3 to 15 psig). This is often referred to as an I/P (I = current, P = pressure) converter.

The components for a microprocessor-based system are shown in Fig. 26–1c. In this case, the control algorithm is software on the computer. The operator interacts with the control system using the computer. The computer can perform many more functions than just the implementation of the control algorithm, as will be discussed later. The chart recorder used on the older pneumatic and electronic systems has been replaced by a monitor on which the transients are shown.

Typical controllers are now capable of processing both analog and digital signals. An analog signal is the type that represents a continuous variable that varies over a range of values. A digital signal is a binary signal that can be represented by two states (on, off, or logic 1, logic 0, etc.). Examples of analog signals are the measurement from a temperature transmitter or the signal sent to a valve. Examples of digital signals are the output to a motor, which causes it to be on or off, or the output to an alarm light causing it to be on or off. The focus of this book has been on analog signals that are applicable to continuous control systems. However, there is an important area of control called *batch control* which frequently must deal with digital (on/off) signals. Batch control, as the name suggests, is the control of processes that are done in a batch
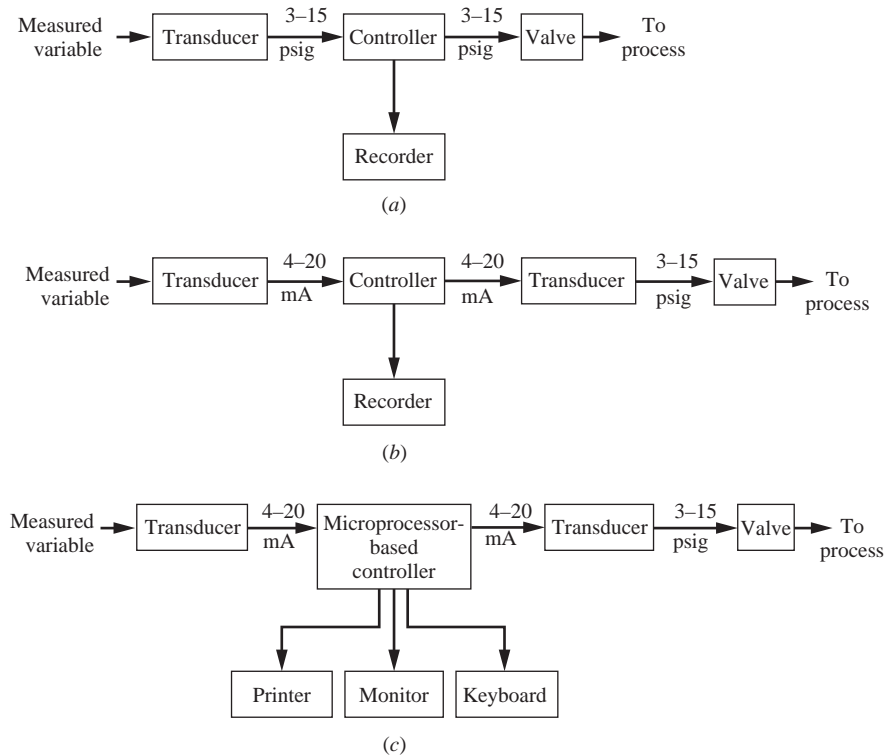
FIGURE 26–1
Controller components for (*a*) pneumatic control, (*b*) electronic control, (*c*) computer or
microprocessor-based control.

operation. Many examples of batch processing occur in the pharmaceutical industry
where small amounts of products of high unit cost are produced.

26.3   TASKS OF A MICROPROCESSOR-BASED CONTROLLER

The primary task of a microprocessor-based controller is implementation of a control
algorithm; however, the presence of a computer makes it possible to perform a number
of other peripheral tasks that are useful in process control and monitoring. Some of
these tasks provided in a modern control system are to

Implement classical and advanced control algorithms
Provide static and dynamic displays on the monitor
Provide process and diagnostic alarms
Provide mathematical functions
Provide data acquisition and storage (archiving)

The software to support all these tasks is supplied by the manufacturer of the control
equipment. We now look briefly at the nature of each task.

Implementation of Control Algorithms

The portion of the software that covers this task is organized into large numbers of blocks that can be connected to solve a specific control problem. A partial listing of the blocks typically provided is as follows:

>   Analog input
>   Analog output
>   Conventional control algorithms (P, PI, PD, PID)
>   Linearization
>   Lead-lag
>   Dead time
>   Self-tuning

There are many other blocks that have been omitted from this list because of the limitation of space in this chapter. Also a number of blocks process digital (or logic) signals (on/off) such as comparators, selectors, or timers, which are needed in batch control and automatic plant start-up and shutdown.

ANALOG INPUT BLOCK. The analog input block is an analog-to-digital device that converts a continuous signal from a transducer, which is in the form of a current or voltage, to a digital signal that can be used in the microprocessor.

ANALOG OUTPUT BLOCK. The analog output block reverses the operation of the analog input block by converting a digital signal, which has been computed in the microprocessor, to a voltage or a current that can be sent out to a transducer in the process in the field. Sometimes this block is called a field output block.

CONTROL BLOCK. The control block is a block for which many parameters can be specified. The manufacturer does not give any information on the method of implementing the control algorithms. The computer does not "watch" the process continuously, but rather it "samples" the necessary process signals at discrete time intervals. The sampling period T is one parameter that generally cannot be adjusted in a commercial controller; it is fixed by the developer of the software. Typical values of T in commercial controllers vary from 0.1 to 0.25 s. A controller operating with such a small T can be considered as a continuous controller for many chemical processes with large time constants. Parameters that can be selected are the controller parameters (K_c, τ_I, τ_D), limits on set point and controller output, and others. The familiar PID control algorithm from Chap. 9 is a continuous form of the equation

$$p = K_c\varepsilon + K_c\tau_D \frac{d\varepsilon}{dt} + \frac{K_c}{\tau_I}\int_0^t \varepsilon\, dt + p_s \qquad (9.11)$$

To implement the PID algorithm in the computer (digital) control system, we must discretize the equation, realizing that we are sampling the signals at fixed instants in time. The discrete version of Eq. (9.11) is

$$p_n = K_c \left[\varepsilon_n + \frac{T}{\tau_I} \sum_{i=1}^{n} \varepsilon_i + \frac{\tau_D}{T}(\varepsilon_n - \varepsilon_{n-1}) \right] + p_s \qquad (26.1)$$

Equation (26.1) can be interpreted as describing the output from the controller at the nth time interval (at $t = nT$). This form of the PID equation is called the *position* form of the equation. It specifies the exact value of the controller output at any time t (for example, the exact valve position if we are controlling a valve). Another form of this equation is called the *velocity* form of the equation. It specifies the change in the controller output from one time interval to the next. The velocity form of the PID equation is

$$\Delta p_n = K_c \left[(\varepsilon_n - \varepsilon_{n-1}) + \frac{T}{\tau_I} \varepsilon_n + \frac{\tau_D}{T}(\varepsilon_n - 2\varepsilon_{n-1} + \varepsilon_{n-2}) \right]$$

$$p_n = p_{n-1} + \Delta p_n \qquad (26.2)$$

Either the form of Eq. (26.1) or that of (26.2) can be used for the PID controller. Each is used in commercially available control systems.

LINEARIZATION BLOCK. The linearization block is used to "straighten out" a non-linear relation. The most common example of the need for this block is in processing a signal from an orifice plate used to measure flow. The signal (pressure) across an orifice plate is proportional to the square of the flow. To obtain a linear relation between flow rate and signal, the signal is sent through a linearization block, which has been configured to extract the square root of the input signal. The linearization block can also be configured to linearize any nonlinear relation that can be plotted on a coordinate system. This aspect of the linearization block can be useful for linearizing the input-output relation to a valve that is nonlinear in behavior. In Chap. 19, an equal-percentage valve was proposed as a device to linearize the relation between flow and valve-top pressure when line loss was large.

LEAD-LAG BLOCK. The lead-lag block simulates the lead-lag transfer function $K(T_1s + 1)/(T_2s + 1)$. The parameters K, T_1, and T_2 can be selected over a wide range of values. If one needs a first-order lag, T_1 can be set to zero. We have seen the need for the lead-lag block in feedforward control in Chap. 17.

DEAD-TIME BLOCK. The dead-time block simulates dead time (or transport lag) $e^{-\tau_d s}$. For this block, τ_d can be selected over a wide range of values. We have seen the need for this block in the Smith predictor control algorithm of Chap. 17.

Figure 26–2 shows a simple flow example using some of the blocks just described. The blocks are connected by computer code at a keyboard during the configuration of the control system. This connection of blocks is called *softwiring* since it is done through the software. The actual connection between the flow transmitter and the analog input block in the controller, which is made with wires, is called *hardwiring*.
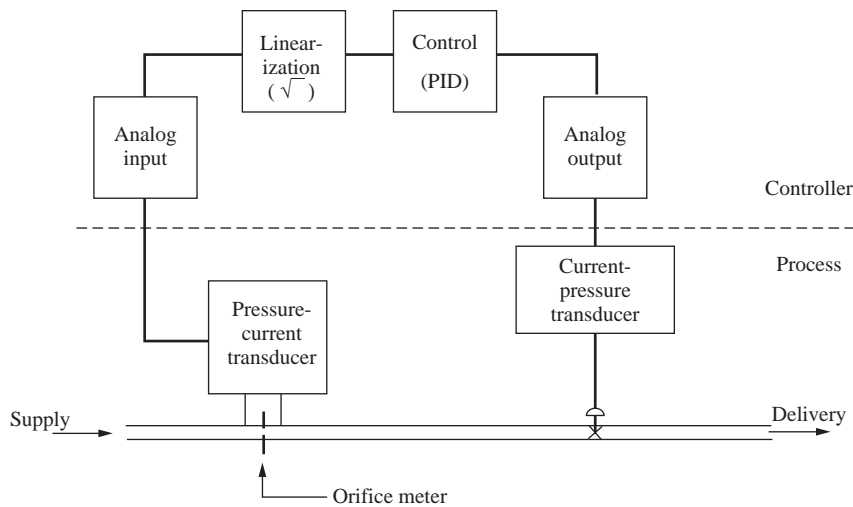
FIGURE 26–2
Example of the use of control blocks to control a flow process.

SELF-TUNE BLOCK. For years, one of the goals of control engineers has been to develop a device that would automatically tune a controller, online, while the process is operating. Until recently, this goal was reached for some special cases by the application of adaptive control theory, a branch of control that is beyond the scope of this book. In the mid-1980s a commercial device became available that uses the normal transients occurring in a controlled process (caused by set point and load upsets) to update the control parameters of a PID controller. This device is called a self-tuner and is one of the blocks available in the microprocessor-based controller of several hardware manufacturers. When the self-tuner is first applied to a process for which no process identification has been performed, the self-tuner is placed in the pretune phase, during which time the process is subjected to a pulse while it is operated open-loop. The introduction of the pulse and the analysis of the transient are done automatically by the self-tuner. The outcome of the pretune phase of operation is the selection of controller parameters. A conceivable approach to the development of the pretune phase of a self-tuner is to monitor an open-loop step response and apply a tuning method similar to the Cohen-Coon tuning method of Chap. 18. After the pretune phase, the control system is returned to closed-loop and the self-tuner continues to monitor the transients and make changes in controller parameters when needed. The self-tuning that occurs during closed-loop operation is based on the characteristics of the transients, such as decay ratio, overshoot, and period of oscillation. The self-tuning algorithm, being proprietary information, is described in only a general manner in the reference manual that accompanies the control equipment. Since many industrial processes are poorly tuned, the general-purpose self-tuner represents an impressive achievement in the application for computer control technology.

Displays

Software has made the strip chart recorder almost unnecessary. Transients (or trends) produced in a control system can now be displayed on a monitor screen dynamically.

As time progresses, the values of selected variables are displayed as a function of time. The segment of time shown on the screen can be selected to be a few minutes to a few hours to show dynamic detail or long-term trends. Transients that occurred in the past can be stored and displayed again.

Many process operators are more comfortable with control instruments that have a faceplate which shows bar graphs or pointers indicating set point, control variable, and output to the valve. In the older instruments, those indicators were obtained by use of mechanical motion or other means. The software for computer-based control can be used to obtain a dynamic display on the screen that mimics the faceplates of older, more traditional instruments.

Alarms

A part of the control software is devoted to detecting and reporting a problem in the form of an alarm. The alarm takes the form of a visual signal (flashing light), an audible signal (beeping horn), or the actuation of a switch. Examples of the use of switch closures include turning on or off a pump motor or opening or shutting a valve. The alarms are classified as *process alarms* and *diagnostic alarms.* The diagnostic alarm detects a malfunction in the control equipment or the loss of communication. For example, if a wire connecting the output of a temperature transmitter to an analog input block breaks, a diagnostic alarm will go off, indicating that the signal to the analog input block is out of range. The manufacturer of the control equipment provides all the software for detecting the problems that trigger diagnostic alarms.

The engineer who configures the control blocks selects the variables that are to trigger process alarms and specifies the alarm limits and the type of annunciation (flashing light, beeping horn, etc.) The alarms can be assigned a priority rating. Those variables in a process that are most critical are given the highest priority; less critical variables are given a lower priority.

Mathematical Functions

Control software provides basic mathematical functions such as summation, subtraction, multiplication, and accumulation (i.e., integration). These functions can be used along with other blocks in the design of a control system. A simple example of these functions is the calculation of mass flow rate of a gas from measurements of velocity, pressure, and temperature. These three measurements are combined according to the following relationship, which is based on the ideal gas law:

$$w = \frac{vAPM}{RT}$$

where w = mass flow rate, mass/time
v = velocity
P = pressure
T = absolute temperature
M = molecular weight of gas
R = gas constant
A = cross-sectional area for flow

The signal from the math block that represents *w* can then be sent as the control variable to a control block that controls the mass flow of gas.

Data Acquisition and Storage

Long-term storage of the transients can be obtained easily with a computer. This task is referred to as *archiving*. The automatic storage of critical process-control variables on disk or tape can be retrieved later to explain process operating difficulties. The computer can also be used to automatically record or log the type and location of an alarm, the time of a process alarm, the time of acknowledgment of an alarm, and the time it was cleared by operator intervention. This information is useful to supervisors in detecting violation of safety regulations or diagnosing process malfunctions.

26.4   SPECIAL FEATURES OF MICROPROCESSOR-BASED CONTROLLERS

In addition to the tasks just described, there are three special features available in modern microprocessor-based controllers that deserve attention. These are limiting, tracking, and anti-reset windup. Each will be discussed separately.

Limiting

In configuring a control system from basic control blocks, one can select lower and upper limits on controller output and set point. These limits are narrower than the limits inherently present in the hardware. Limits are often placed on a controller output for safety reasons or to protect equipment. For example, if one knows the flow rate of a liquid that causes a tank to overflow, one can set the limit on the output of a controller at a value less than the value that causes overflow. The limits on the controller output are active when the controller is in either automatic or manual mode. An example of a limit on set point is the selection of an upper limit on pressure for a steam-heated sterilizer to prevent damage to the equipment.

Tracking

A very useful feature of a microprocessor-based controller is tracking. Although tracking is not needed to successfully control a system, its presence is of great convenience to the process operator. Two examples of tracking are set point tracking and controller output tracking.

Set point tracking is useful when a controller is transferred from manual to automatic. When a process is started up for the first time, a common procedure is to bring the process on-stream in manual mode. In this case the operator adjusts the output of the controller (which goes to the valve) until the process variable comes to a desired steady state. When the tracking feature is not present in the controller, the set point must be manually adjusted until it equals the process variable before the controller is transferred to

automatic; the process then continues running in a smooth manner. If the operator adjusts the set point to the process variable after switching to automatic, there may be a temporary disturbance in the process variable. The expression for the disturbance is called a *bump*. With set point tracking, the operator does not need to think about adjusting the set point to the process variable, because it is done automatically. In other words, set point tracking provides "bumpless" transfer when switching from manual to automatic.

A second example of tracking can be seen in its use for transferring a cascade system from manual to automatic. (The reader should be familiar with the information on cascade control provided in Chap. 17 to understand this example.) To explain the use of tracking in cascade control, reference to Figs. 17–1*b* and 17–2*b* will be made. In starting up this system, the primary controller is placed in a standby condition, and the secondary controller is placed in manual mode. The means for accomplishing this is built into the software of the controller. With the secondary controller in manual mode, its output is adjusted until the temperature of the tank contents T_0 is at the desired value. Then, with the control system at steady state and T_0 at the desired value, the system is transferred to cascade mode by placing both controllers in automatic. Since the output of the primary controller adjusts the set point of the secondary controller, it is necessary to have the output of the primary controller equal to the jacket temperature T_j when the system is transferred to cascade mode. This goal can be achieved by having the output of the primary controller track the jacket temperature while the secondary controller is used in manual mode to adjust the tank temperature to the desired value. For this example, the set point of the primary controller can also automatically track the tank temperature T_0 before the transfer to cascade mode occurs. In this cascade control example, we have seen tracking used for both the set point and the controller output.

Anti-Reset Windup

A troublesome problem with a controller having integral action (PI or PID) is the possible occurrence of reset windup. When the error to a controller remains large for a long time, the integral action of the controller builds up a large value of output which often approaches the saturation value of the controller output. This accumulation of output is called *reset windup*. When the process variable returns to the set point, the output of the controller does not immediately return to a value that will hold the process variable at the set point because the controller output has built up (or has been wound up) and must be reduced by the presence of error of opposite sign over some duration of time. Thus the transient for the control variable exhibits a large overshoot that can persist while the output signal is being reduced through integral action being applied to the error of reversed sign.

Reset windup typically occurs during the start-up of a process. To gain some insight into the cause of reset windup, consider the start-up of the liquid-level process shown in Fig. 26–3 in which the level in the third tank is to be controlled by a PI controller. The valve is linear and saturates at 0 and at 0.5, as shown in Fig. 26–3*b*. Upon start-up with the PI controller in automatic mode, the tanks are empty, and the error $R - C$ is large and positive. The action of the controller on this error will result in a large output M due to proportional action and a rising contribution to M due to the integral action. The output of the controller will be at its saturation value, which is typically about 10 percent above the top of the 4- to 20-mA scale (i.e., 22 mA).
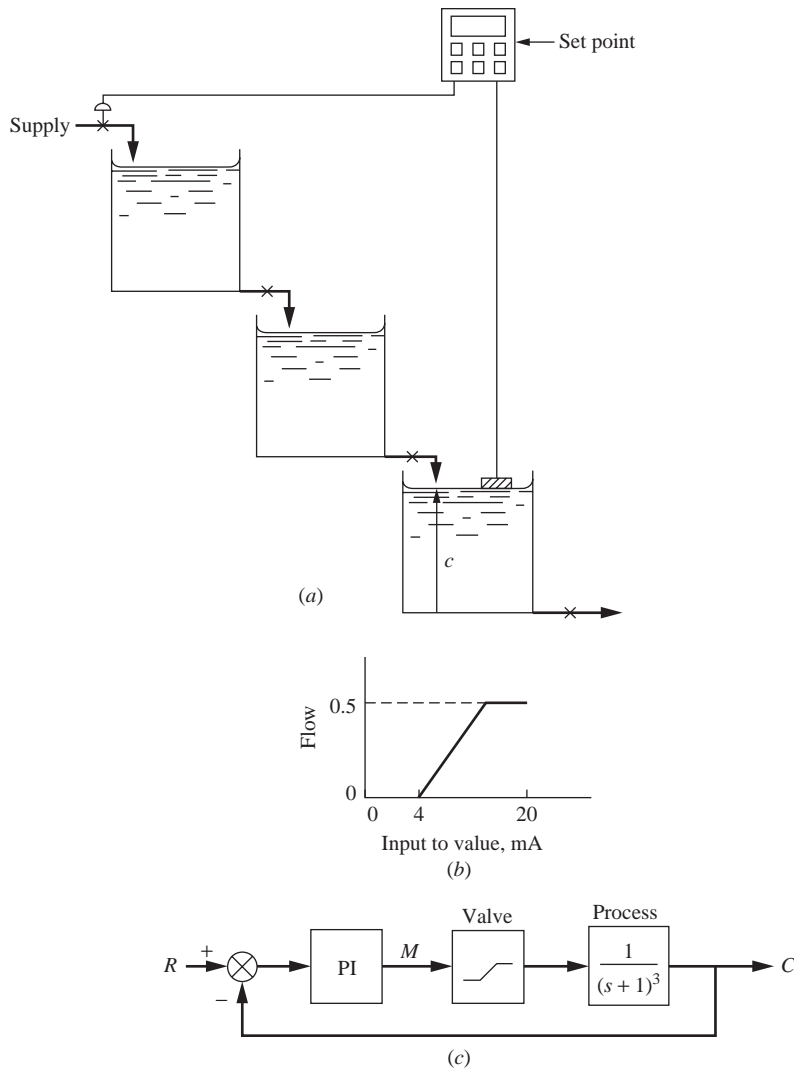
FIGURE 26–3

Plant start-up illustrating reset windup (tanks are initially empty): (*a*) process, (*b*) linear valve
with saturation limits, and (*c*) block diagram of process.

The large saturated value of M will in turn cause the valve to reach its saturation
value, which has been taken as 0.5. During the initial phase of the operation, the tanks are
being filled at the maximum rate of flow provided by the upper limit of the control valve.
During this filling stage of operation, the controller is not exercising any control since
the valve is at its limit. As the level rises toward the set point, the large error that existed
at start-up gradually diminishes toward zero. If only proportional action were present
in the controller, the output of the controller would return quickly to a mid-scale value;
however, because of the integral action, the controller output remains high, at its satura-
tion value, long after the process variable first reaches the set point. To reduce the output

M, the integral action must be applied to negative error so that the integration can lower the output to mid-scale. This negative error occurs as a result of the tank level remaining *above* the set point for some time after the tank level reaches the set point. Other causes of reset windup and some methods to prevent it are discussed by Shinskey (1979).

The control system shown in Fig. 26–3*c* was simulated for a start-up transient with the tanks initially empty; the transient is shown as curve I in Fig. 26–4. The large overshoot in tank level after the level reaches the set point is clearly illustrated. Now that the problem of reset windup has been described, we focus our attention on how to reduce or eliminate it. The development that follows on the use of external feedback to eliminate reset windup is based on the work of Shunta and Klein (1979).
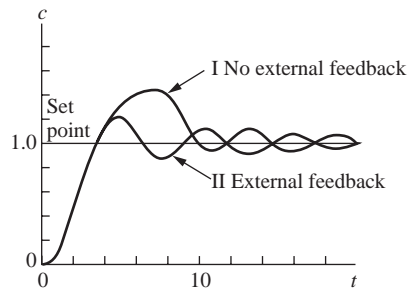
FIGURE 26–4
Start-up transients for system in Fig. 26–3 with and without external feedback.

A feature of microprocessor-based controllers is the availability of external feedback in the configuration of a PI or PID controller. The block diagram of a PI controller with external feedback is shown in Fig. 26–5.
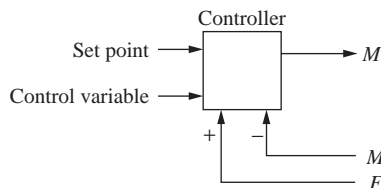
FIGURE 26–5
Controller with external feedback for use in anti-reset windup.

The output of this controller is given by

$$M(t) = K_c e(t) + \frac{K_c}{\tau_I}\int_0^t e(t)\,dt + \frac{1}{\tau_I}\int_0^t [F(t) - M(t)]\,dt \qquad (26.3)$$

where

$$M(t) = \text{controller output}$$

$$e(t) = \text{error} = \text{set point} - \text{control variable}$$

$$F(t) = \text{external feedback signal}$$

If the Laplace transform of both sides of Eq. (26.1) is taken, the result is

$$M(s) = K_c e(s) + \frac{K_c e(s)}{s\tau_I} + \frac{1}{s\tau_I}\big[F(s) - M(s)\big] \tag{26.4}$$

If the feedback signal is the controller output $F(s) = M(s)$, Eq. (26.4) becomes the usual transfer function for a PI controller:

$$M(s) = K_c\left(1 + \frac{1}{\tau_I s}\right)e(s) \tag{26.5}$$

The feedback signal $F(t)$ can be any signal available to the microprocessor-based controller. When $F(t)$ is not equal to $M(t)$, Eq. (26.4) can be solved for $M(s)$ to give

$$M(s) = K_c e(s) + \frac{F(s)}{\tau_I s + 1} \tag{26.6}$$

A controller following this equation provides a signal consisting of proportional action plus first-order tracking of $F(t)$. If $F(t)$ in Eq. (26.3) is taken as the output of the valve (or the output signal of the current-to-pressure transducer that goes to the valve) in our example in Fig. 26–3c, we have the basis for eliminating reset windup. During the filling stage of the tank, the feedback signal $F(t)$ will be constant at the saturation value of the valve output. When the tank level reaches the set point, the error will be zero and the only contribution from the controller output will be the tracked signal represented by the second term on the right side of Eq. (26.6). This value will be less than would be the case if external feedback were not employed. The overall result is that the controller output is less with the external feedback at the time the level first equals the set point, and the overshoot is reduced. The transient using external feedback is also shown in Fig. 26–4 as curve II. Notice that the overshoot is less when external feedback is used. To emphasize the benefit of external feedback for eliminating reset windup, no limits were placed on the output of the controller in the simulation of Fig. 26–3. In practice, there are physical limits on the controller output, and when this is the case, the reduction of overshoot with the use of external feedback may not be so pronounced as shown in Fig. 26–4.

26.5   DISTRIBUTED CONTROL

So far we have been concerned in this chapter with the operation of a single controller. Such a controller is referred to as a *stand-alone controller* because it is not communicating with other controllers, but only with the one control loop of which it is a part. Present-day microcomputer-based control systems have the capability of communicating with other controllers through a network, which is called *distributed control*. Figure 26–6 shows one version of the communication linkages that are usually present in a distributed control system. Each manufacturer of distributed control systems has a different way of organizing them.

A distributed control system is intended to be used for a large processing facility that involves as many as 50 to 100 loops. Some examples of these types of installations are a refinery, a brewery, and a power plant. In Fig. 26–6, the modules of control equipment that communicate with each other are

Control processor (CP)
Applications processor (AP)
Workstation (WS)
Field bus module (FBM)

The first three of these modules communicate with one another through a nodebus or "data highway," as it has been called. The field bus modules serve as devices that interface with transducers and valves in the process. Distributed control systems thus involve small modular controllers mounted in the field, process operator PC workstations in the control room or distributed throughout the plant, and a data highway (perhaps an Ethernet network) connecting the components.

The control processor contains the blocks described earlier (analog input, analog output, control, linearization, etc.) that are connected by softwiring to provide the control algorithm required for each loop. Communication between the control processor and the process (a distance away) in the field takes place in the field bus module.
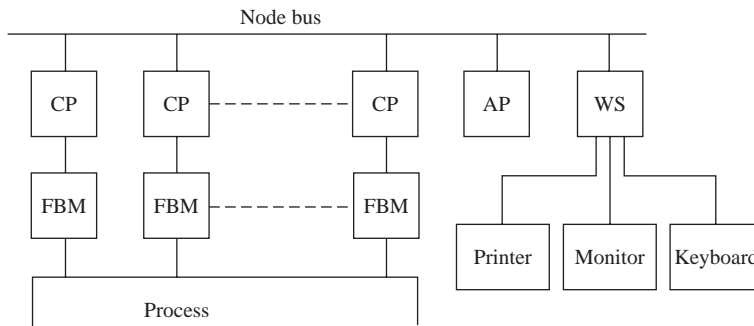
FIGURE 26–6
Typical connections in a distributed control system: CP: central processor, AP: applications processor, WS: workstation, FBM: field bus module.

One type of field bus module provides a set of analog inputs and a set of analog outputs that send to and receive from the field continuous signals (4 to 20 mA). Another type of module sends to and receives from the field digital signals that often take the form of switch-contact closures.

The application processor is a computer (which runs software) for performing the many tasks described earlier and for managing the communication among modules.

The workstation module is connected to a keyboard, a mouse, a monitor, and a printer for use by process operators to interact with the system. At the workstation, the process operator can call up on the screen various displays, change set points and controller parameters, switch from automatic to manual, acknowledge alarms, and perform other tasks needed to operate a control system consisting of many loops.

There are currently a number of types of digital networks (or buses) for plant control systems. In order of increasing complexity, they are sensor buses (used for simple devices and manufacturing control), device buses (which include Profibus and DeviceNet), and field buses (which include Profibus and Foundation Fieldbus) which are used for process control and diagnostics. These types of buses or networks are supported by numerous hardware vendors. "Smart" instrumentation is becoming more and more popular as equipment and control systems are replaced. Smart instrumentation has the capability to provide the control system with additional information and function that "nonsmart" devices do not, such as device ID verification, device diagnostics and status information, secondary process and device variables, remote device configuration and setup. According to the ARC Advisory Group of Dedham, Massachusetts (Fig. 26–7), worldwide about 40 percent of the installed devices are now smart devices D'Aquino and Greene, (May 2003). The rest are pneumatic and electronic.

40–45 million Devices Installed
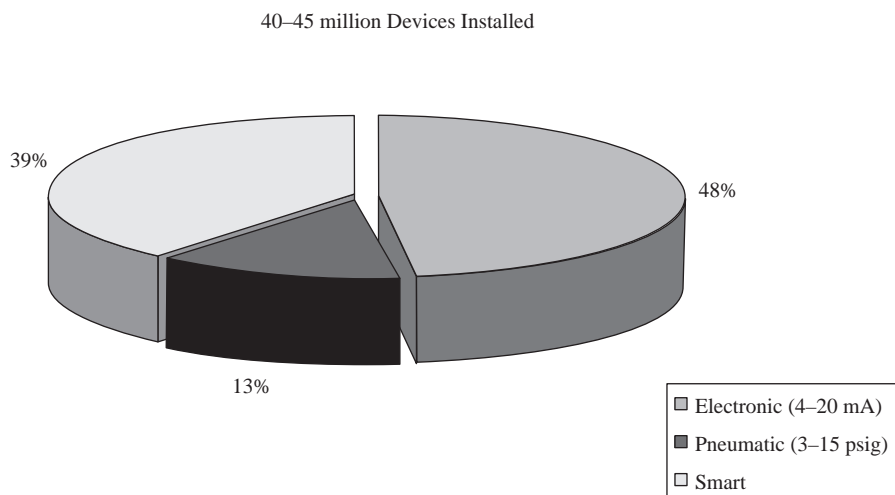
FIGURE 26–7
Nature of installed process instrumentation worldwide according to the ARC Advisory Group of Dedham, Massachusetts.

The trend is certainly moving toward the installation of computer control systems and smart instrumentation devices. Some of the challenges to be overcome include lack of compatibility with existing plant control systems and wiring, lack of a standard digital protocol, and initial cost. As these systems mature and the obstacles are overcome, we can look forward to exciting advances in process monitoring and control systems' hardware and software. The net result will be better controlled processes, and better access to process data for monitoring, diagnostics, and reporting capabilities.

SUMMARY

The computer has greatly changed the nature of industrial process control equipment. The microprocessor has become the heart of control instruments, and software has provided many functions besides the basic control algorithm. When the pneumatic controller

was the predominant type, one purchased a controller with very specific attributes (e.g., mode of control, type of measured variable, chart speed). The smart instruments and controllers available today contain not only the conventional control algorithms, but also many other functions such as simulation of basic transfer functions (e.g., lead-lag and transport lag), display building, mathematical functions, process and diagnostic alarms, and data acquisition. They also provide logic functions (comparators, timers, counters, etc.) for use in batch control and plant start-up and shutdown, as well as self-tuning algorithms.

In this chapter, some of the features of modern controllers were discussed (e.g., limiting, tracking, and anti-reset windup). Any controller having integral action can cause reset windup under certain conditions when the error persists for a long time. The result of such a phenomenon is a transient that has large overshoot. Manufacturers of control instruments now offer several methods for reducing reset windup; the one presented in this chapter was use of external feedback.

Before computer control appeared, most process loops were served by individual controllers with signals to and from these controllers being collected on a large panel board in a special control room. To obtain communication between the control room and the controllers required much wiring and piping (for pneumatic systems). Today, microprocessor-based control systems have the capability of communicating with other control instruments through networks, called distributed control. A distributed control system can control an entire plant and involve as many as one hundred or more control loops. Since each manufacturer has a different way of organizing a distributed control system, the practicing engineer must obtain the details of a particular system from the manufacturer. Most manufacturers offer a variety of short courses for technicians and engineers on the installation and use of their hardware and software.

CHAPTER
26

CAPSULE SUMMARY

DISTRIBUTED CONTROL SYSTEM SCHEMATIC
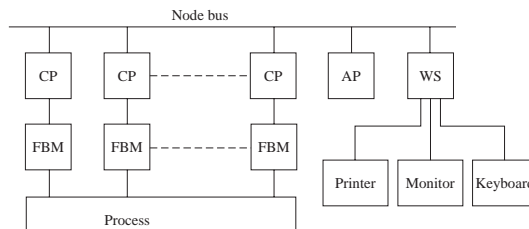
FIGURE 26-6
Typical connections in a distributed control system: CP: central processor, AP: applications processor, WS: workstation, FBM: field bus module.

SMART INSTRUMENTS

Smart instruments have the capability to provide the control system with additional information and function that "nonsmart" devices do not, such as device ID verification, device diagnostics and status information, secondary process and device variables, remote device configuration and setup.

FORMS OF PID ALGORITHM FOR DIGITAL IMPLEMENTATION

Standard continuous form

$$p = K_c \varepsilon + K_c \tau_D \frac{d\varepsilon}{dt} + \frac{K_c}{\tau_I} \int_0^t \varepsilon \, dt + p_s \qquad (9.11)$$

Position form of PID algorithm

$$p_n = K_c \left[\varepsilon_n + \frac{T}{\tau_I} \sum_{i=1}^n \varepsilon_i + \frac{\tau_D}{T} (\varepsilon_n - \varepsilon_{n-1}) \right] + p_s \qquad (26.1)$$

Velocity form of PID algorithm

$$\Delta p_n = K_c \left[(\varepsilon_n - \varepsilon_{n-1}) + \frac{T}{\tau_I} \varepsilon_n + \frac{\tau_D}{T} (\varepsilon_n - 2\varepsilon_{n-1} + \varepsilon_{n-2}) \right]$$
$$p_n = p_{n-1} + \Delta p_n \qquad (26.2)$$

Either the form of Eq. (26.1) or that of Eq. (26.2) can be used for the PID controller. Each is used in commercially available control systems..

BIBLIOGRAPHY

Aris, R., and N. R. Amundson (1958). *Chem. Eng. Sci., 7,* 121–155.

Anton, H. (1984). *Elementary Linear Algebra,* 4th ed., New York: Wiley.

Bequette, B. W. (2003). *Process Control: Modeling, Design and Simulation,* Upper Saddle River, N. J.: Prentice-Hall.

Bird, R. B., W. E. Stewart, and E. N. Lightfoot (1960). *Transport Phenomena,* New York: Wiley.

Carslaw, H. S., and J. C. Jaeger (1959). *Conduction of Heat in Solids,* 2d ed., Oxford at the Clarendon Press.

Chau, P. C. (2002). *Process Control: A First Course with MATLAB,* Cambridge UK: Cambridge University Press.

Churchill, R. V. (1972). *Operational Mathematics.* 3d ed., New York: McGraw-Hill.

Cohen, G. H., and G. A. Coon (1953). *Trans. ASME, 75,* 827.

Cohen, W. C., and E. F. Johnson (1956). *IEC, 48,* 1031–34.

Cooper, Douglas J. (2005). *Practical Process Control* using Loop-Pro Software, Control Station, Inc.

Coughanowr, D. R. (1991). *Process Systems Analysis and Control,* 2d ed., New York: McGraw-Hill.

Coughanowr, D. R., and L. B. Koppel (1965). *Process Systems Analysis and Control,* New York: McGraw-Hill.

D'Aquino R. and R. Greene, "*Forays in Smart Instrumentation*", Chem. Eng. Progress, 99, 8.

Doyle, F. J., III (2000). *Process Control Modules: A Software Laboratory for Control Design,* Upper Saddle River, N.J.: Prentice-Hall.

Evans, W. R. (1948). "Graphical Analysis of Control Systems," *Trans. AIEE, 67,* 547–551.

Evans, W. R. (1954). *Control-System Dynamics,* New York: McGraw-Hill.

Foxboro Co. (1978). "Principles of Feedforward Control." Audiovisual tape, No. B0150ME (1 hr, 23 min; 3 parts), Foxboro, Mass.

Kuo, B. C. (1987). *Automatic Control Systems,* 5th ed., Englewood Cliffs, NJ : Prentice Hall.

Letov, A. M. (1961). *Stability in Nonlinear Control Systems,* Princeton, N.J.: Princeton University Press.

Lopez, A. M., P. W. Murill, and C. L. Smith (1967). "Controller Tuning Relationships Based on Integral Performance Criteria.", Instrumentation Technology, 14, No.11, 57.

Luyben, W. L., and M. L. Luyben (1997). *Essentials of Process Control,* New York: McGraw-Hill.

Marlin, T. E. (2000). *Process Control: Designing Processes and Control Systems for Dynamic Performance,* 2d ed., Boston: McGraw-Hill.

Marlin, T. E. (2000), *Process Control: Designing Processes and Control Systems for Dynamic Performance,* 2d ed., New York: McGraw-Hill.

McCabe, W. L., J. C. Smith, and P. Harriott (2004). *Unit Operations of Chemical Engineering,* 7th ed., New York: McGraw-Hill.

Mickley, H. S., T. K. Sherwood, and C. E. Reed (1957). *Applied Mathematics in Chemical Engineering,* 2d ed., New York: McGraw-Hill.

Morari, M., and E. Zafiriou (1989). *Robust Process Control,* Englewood Cliffs, N.J.: Prentice-Hall.

Nobbe, L. B. (1961). "Transient Response of a Bubble-Cap Plate Absorber." M.S. thesis, Purdue University, Indiana.

Ogunnaike, B. A., and W. H. Ray (1994). *Process Dynamics, Modeling and Control,* New York: Oxford University Press.

Oldenbourg, R. C., and H. Sartorius (1948). "The Dynamics of Automatic Control," *Trans. ASME,* Vol. 77, p. 78.

Perry, R. H., and D. W. Green (2007). *Chemical Engineers' Handbook,* 8th ed., New York: McGraw-Hill.

Philips, C. L., and R. D. Harbor (2000). *Feedback Control Systems,* 4th ed., Englewood Cliffs, N.J.: Prentice-Hall.

Riggs, James B., and M. Nazmul Karim (2007). *Chemical and Bio-Process Control,* 3d ed., Lubbock, Tex.: Ferret Publishing.

Routh, E. J. (1905). *Dynamics of a System of Rigid Bodies.* Part II. London: Macmillan & Co., Ltd.

Seborg, D. E., T. F. Edgar, and D. A. Mellichamp (2004). *Process Dynamics and Control,* 2d ed., New York: Wiley.

Shinskey, F. G. (1979). *Process Control Systems,* 2d ed., New York: McGraw-Hill.

Shunta, J. P. and W. F. Klein (1979). "*Microcomputer Digital Control—What it ought to do,*" ISA Trans., 18, No. 1, 63.

Smith, C. A., and A. B. Corripio (2006). *Principles and Practice of Automatic Process Control,* 3d ed., New York: Wiley.

Smith, O. J. M. (1957). "Closer Control of Loops with Dead Time," *Chem. Eng. Prog., 53,* 217.

Stephanopoulos, G. (1984). *Chemical Process Control. An Introduction to Theory and Practice,* Englewood Cliffs, N.J.: Prentice-Hall.

Thaler, G. J., and M. P. Pastel (1962). *Analysis and Design of Nonlinear Feedback Control Systems,* New York: McGraw-Hill.

Wilts, C. H. (1960). *Principles of Feedback Control,* Reading, Mass: Addison-Wesley.

Ziegler, J. G., and N. B. Nichols (1942). "Optimum Settings for Automatic Controllers," *Trans. ASME, 64,* 759.

INDEX

Useful Functions and Laplace Transforms

TABLE 2.1

Function	Graph	Transform
$u(t)$		$\dfrac{1}{s}$
$tu(t)$		$\dfrac{1}{s^2}$
$t^n u(t)$		$\dfrac{n!}{s^{n+1}}$
$e^{-at}u(t)$		$\dfrac{1}{s+a}$
$t^n e^{-at}u(t)$		$\dfrac{n!}{s+a^{n+1}}$
$\sin kt\, u(t)$		$\dfrac{k}{s^2+k^2}$

TABLE 2.1 *(Continued)*

Function	Graph	Transform
$\cos kt\, u(t)$		$\dfrac{s}{s^2 + k^2}$
$\sinh kt\, u(t)$		$\dfrac{k}{s^2 - k^2}$
$\cosh kt\, u(t)$	1	$\dfrac{s}{s^2 - k^2}$
$e^{-at}\sin kt\, u(t)$		$\dfrac{k}{(s+a)^2 + k^2}$
$e^{-at}\cos kt\, u(t)$		$\dfrac{s+a}{(s+a)^2 + k^2}$
$\delta(t)$, unit impulse	Area = 1	1

Key Features of Standard Responses of First Order Systems to Common Inputs

Step Response of First Order System

Ultimate Value

Response is 63.2% complete at t=τ

Initial Slope intersects Ultimate Value at t=τ

Y/Kp

t/τ

Impulse Response of a First Order System

Initial "Jump" is to Kp/τ

Response is 63.2% complete at t=τ
(Initial "Jump" has decayed to 36.8%)

$Y^*\tau/Kp$

t/τ

Sinusoidal Response of a First Order System

phase lag = arctan($-w\tau$)

ratio=$1/\sqrt{1+(w\tau)^2}$

After an initial transient period, the response is periodic with the same frequency

Y/AKp

t/τ

Response of First Order System to Ramp Input

After an initial transient period, the response is parallel with input.

Steady state difference between input and output (after transient) is bτ

bτ

output lags input by τ
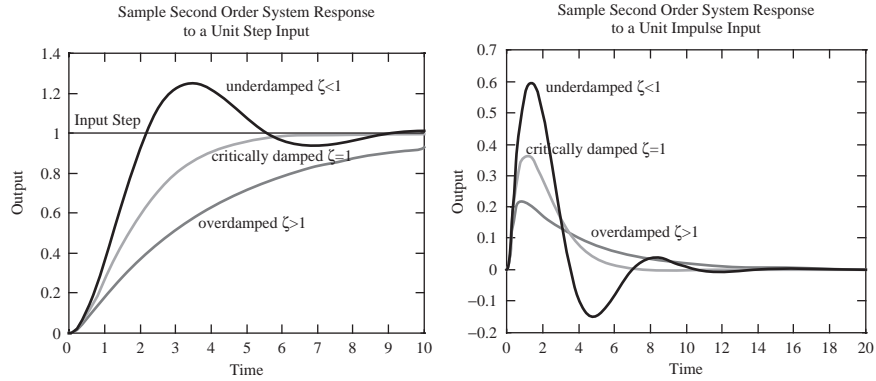
input

output

Y/bKp

t/τ

Standard Form for a Second Order System:

$$\tau^2 \frac{d^2Y}{dt^2} + 2\zeta\tau\frac{dY}{dt} + Y = X(t)$$

τ = Time Constant

ζ = Damping Coefficient, the magnitude of this parameter determines the nature of the response

Sample Second Order System Response to a Unit Step Input

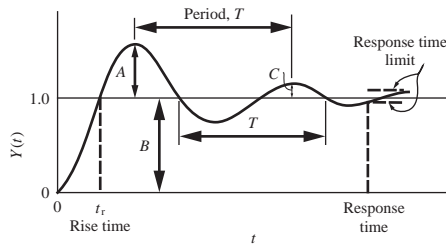Sample Second Order System Response to a Unit Impulse Input

Terms to Describe an Underdamped (Oscillatory) Second Order Response

$$\text{Overshoot} = \exp\left(-\pi\zeta/\sqrt{1-\zeta^2}\right) = \frac{A}{B}$$

$$\text{Decay ratio} = \exp\left(-2\pi\zeta/\sqrt{1-\zeta^2}\right) = (\text{overshoot})^2 = \frac{C}{A}$$

$$\omega, \text{ radian frequency} = \frac{\sqrt{1-\zeta^2}}{\tau}$$

$$\omega_n = \text{ natural frequency} = \frac{1}{\tau}$$

Process Systems Analysis and Control, Third Edition retains the clarity of presentation for which this book is well known. It is an ideal teaching and learning tool for a semester-long undergraduate chemical engineering course in process dynamics and control. It avoids the encyclopedic approach of many other texts on this topic. Computer examples using MATLAB® and Simulink® have been introduced throughout the book to supplement and enhance standard hand-solved examples. These packages allow the easy construction of block diagrams and quick analysis of control concepts to enable the student to explore "what-if" type problems that would be much more difficult and time consuming by hand. New homework problems have been added to each chapter. The new problems are a mixture of hand-solutions and computational-exercises. One-page capsule summaries have been added to the end of each chapter to help students review and study the most important concepts in each chapter.

Key Features:
- Short, clearly written chapters aimed at the undergraduate chemical engineer.
- Restructuring of the initial chapters to reduce the impression that students frequently have regarding control classes…that this is just another mathematics course disguised as an engineering course
- Computer applications in MATLAB®, Simulink® and Excel® have been introduced throughout the book.
  - to reduce the tedium of solving problems so that students may concentrate more on the concepts of dynamics and control and not get bogged down in the mathematical complexities of each problem
  - to give students the tools to be able to ask (and more easily answer) "what if…?" type of questions
  - to allow students to explore more difficult problems than would otherwise be possible in the time available for the course material
- One page capsule summaries have been added to the end of each chapter.

The Solutions to the End-of-Chapter Problems are available to Instructors at the text's website: **www.mhhe.com/coughanowr-leblanc**

Electronic Textbook Options

This text is offered through CourseSmart for both instructors and students. CourseSmart is an online browser where students can purchase access to this and other McGraw-Hill textbooks in a digital format. Through their browser, students can access the complete text online for one year at almost half the cost of a traditional text. Purchasing the eTextbook also allows students to take advantage of CourseSmart's web tools for learning, which include full text search, notes and highlighting, and email tools for sharing notes between classmates. To learn more about CourseSmart options, contact your sales representative or visit **www.CourseSmart.com.**

The McGraw·Hill Companies

McGraw Hill **Higher Education**