SCHAUM'S OUTLINE OF

THEORY AND PROBLEMS

OF

FEEDBACK and CONTROL SYSTEMS

Second Edition

CONTINUOUS (ANALOG) AND DISCRETE (DIGITAL)

JOSEPH J. DiSTEFANO, III, Ph.D.
Departments of Computer Science and Medicine
University of California, Los Angeles

ALLEN R. STUBBERUD, Ph.D.
Department of Electrical and Computer Engineering
University of California, Irvine

IVAN J. WILLIAMS, Ph.D.
Space and Technology Group, TRW, Inc.

SCHAUM'S OUTLINE SERIES

McGRAW-HILL

New York San Francisco Washington, D.C. Auckland Bogotá Caracas Lisbon London Madrid Mexico City Milan Montreal New Delhi San Juan Singapore Sydney Tokyo Toronto

JOSEPH J. DiSTEFANO, III received his M.S. in Control Systems and Ph.D. in Biocybernetics from the University of California, Los Angeles (UCLA) in 1966. He is currently Professor of Computer Science and Medicine, Director of the Biocybernetics Research Laboratory, and Chair of the Cybernetics Interdepartmental Program at UCLA. He is also on the Editorial boards of **Annals of Biomedical Engineering** and **Optimal Control Applications and Methods**, and is Editor and Founder of the **Modeling Methodology Forum** in the **American Journals of Physiology**. He is author of more than 100 research articles and books and is actively involved in systems modeling theory and software development as well as experimental laboratory research in physiology.

ALLEN R. STUBBERUD was awarded a B.S. degree from the University of Idaho, and the M.S. and Ph.D. degrees from the University of California, Los Angeles (UCLA). He is presently Professor of Electrical and Computer Engineering at the University of California, Irvine. Dr. Stubberud is the author of over 100 articles, and books and belongs to a number of professional and technical organizations, including the American Institute of Aeronautics and Astronautics (AIAA). He is a fellow of the Institute of Electrical and Electronics Engineers (IEEE), and the American Association for the Advancement of Science (AAAS).

IVAN J. WILLIAMS was awarded B.S., M.S., and Ph.D. degrees by the University of California at Berkeley. He has instructed courses in control systems engineering at the University of California, Los Angeles (UCLA), and is presently a project manager at the Space and Technology Group of TRW, Inc.

Schaum's Outline of Theory and Problems of
FEEDBACK AND CONTROL SYSTEMS

6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 BAW BAW 9 9

ISBN 0-07-017052-5 (Formerly published under ISBN 0-07-017047-9).

Sponsoring Editor: John Aliano
Production Supervisor: Louise Karam
Editing Supervisors: Meg Tobin, Maureen Walker

Library of Congress Cataloging-in-Publication Data

DiStefano, Joseph J.
Schaum's outline of theory and problems of feedback and control systems/Joseph J. DiStefano, Allen R. Stubberud, Ivan J. Williams. —2nd ed.
p. cm. — (Schaum's outline series)
ISBN 0-07-017047-9
1. Feedback control systems. 2. Control theory. I. Stubberud, Allen R. II. Williams, Ivan J. III. Title. IV. Title: Outline of theory and problems of feedback and control systems.
TJ216.5.D57 1990
629.8'3—dc20 89-14585

McGraw-Hill
A Division of The McGraw-Hill Companies

Preface

Feedback processes abound in nature and, over the last few decades, the word feedback, like *computer*, has found its way into our language far more pervasively than most others of technological origin. The conceptual framework for the theory of feedback and that of the discipline in which it is embedded—control systems engineering–have developed only since World War II. When our first edition was published, in 1967, the subject of linear continuous-time (or *analog*) control systems had already attained a high level of maturity, and it was (and remains) often designated *classical control* by the *conoscienti*. This was also the early development period for the digital computer and discrete-time data control processes and applications, during which courses and books in "sampled-data" control systems became more prevalent. Computer-controlled and *digital* control systems are now the terminology of choice for control systems that include digital computers or microprocessors.

In this second edition, as in the first, we present a concise, yet quite comprehensive, treatment of the fundamentals of feedback and control system theory and applications, for engineers, physical, biological and behavioral scientists, economists, mathematicians and students of these disciplines. Knowledge of basic calculus, and some physics are the only prerequisites. The necessary mathematical tools beyond calculus, and the physical and nonphysical principles and models used in applications, are developed throughout the text and in the numerous solved problems.

We have modernized the material in several significant ways in this new edition. We have first of all included discrete-time (digital) data signals, elements and control systems throughout the book, primarily in conjunction with treatments of their continuous-time (analog) counterparts, rather than in separate chapters or sections. In contrast, these subjects have for the most part been maintained pedagogically distinct in most other textbooks. Wherever possible, we have integrated these subjects, at the introductory level, in a *unified* exposition of continuous-time and discrete-time control system concepts. The emphasis remains on continuous-time and linear control systems, particularly in the solved problems, but we believe our approach takes much of the mystique out of the methodologic differences between the analog and digital control system worlds. In addition, we have updated and modernized the nomenclature, introduced state variable representations (models) and used them in a strengthened chapter introducing nonlinear control systems, as well as in a substantially modernized chapter introducing advanced control systems concepts. We have also solved numerous analog and digital control system analysis and design problems using special purpose computer software, illustrating the power and facility of these new tools.

The book is designed for use as a text in a formal course, as a supplement to other textbooks, as a reference or as a self-study manual. The quite comprehensive index and highly structured format should facilitate use by any type of readership. Each new topic is introduced either by section or by chapter, and each chapter concludes with numerous solved problems consisting of extensions and proofs of the theory, and applications from various fields.

Los Angeles, Irvine and
Redondo Beach, California
March, 1990

JOSEPH J. DISTEFANO, III
ALLEN R. STUBBERUD
IVAN J. WILLIAMS

Contents

Chapter 1 **INTRODUCTION** **1**
1.1 Control Systems: What They Are 1
1.2 Examples of Control Systems 2
1.3 Open-Loop and Closed-Loop Control Systems 3
1.4 Feedback 4
1.5 Characteristics of Feedback 4
1.6 Analog and Digital Control Systems 4
1.7 The Control Systems Engineering Problem 6
1.8 Control System Models or Representations 6

Chapter 2 **CONTROL SYSTEMS TERMINOLOGY** **15**
2.1 Block Diagrams: Fundamentals 15
2.2 Block Diagrams of Continuous (Analog) Feedback Control Systems 16
2.3 Terminology of the Closed-Loop Block Diagram 17
2.4 Block Diagrams of Discrete-Time (Sampled-Data, Digital) Components, Control Systems, and Computer-Controlled Systems 18
2.5 Supplementary Terminology 20
2.6 Servomechanisms 22
2.7 Regulators 23

Chapter 3 **DIFFERENTIAL EQUATIONS, DIFFERENCE EQUATIONS, AND LINEAR SYSTEMS** **39**
3.1 System Equations 39
3.2 Differential Equations and Difference Equations 39
3.3 Partial and Ordinary Differential Equations 40
3.4 Time Variability and Time Invariance 40
3.5 Linear and Nonlinear Differential and Difference Equations 41
3.6 The Differential Operator D and the Characteristic Equation 41
3.7 Linear Independence and Fundamental Sets 42
3.8 Solution of Linear Constant-Coefficient Ordinary Differential Equations 44
3.9 The Free Response 44
3.10 The Forced Response 45
3.11 The Total Response 46
3.12 The Steady State and Transient Responses 46
3.13 Singularity Functions: Steps, Ramps, and Impulses 47
3.14 Second-Order Systems 48
3.15 State Variable Representation of Systems Described by Linear Differential Equations 49
3.16 Solution of Linear Constant-Coefficient Difference Equations 51
3.17 State Variable Representation of Systems Described by Linear Difference Equations 54
3.18 Linearity and Superposition 56
3.19 Causality and Physically Realizable Systems 57

Chapter 4 **THE LAPLACE TRANSFORM AND THE z-TRANSFORM** **74**
4.1 Introduction 74
4.2 The Laplace Transform 74
4.3 The Inverse Laplace Transform 75
4.4 Some Properties of the Laplace Transform and Its Inverse 75
4.5 Short Table of Laplace Transforms 78
4.6 Application of Laplace Transforms to the Solution of Linear Constant-Coefficient Differential Equations 79
4.7 Partial Fraction Expansions 83
4.8 Inverse Laplace Transforms Using Partial Fraction Expansions 85
4.9 The z-Transform 86
4.10 Determining Roots of Polynomials 93
4.11 Complex Plane: Pole-Zero Maps 95
4.12 Graphical Evaluation of Residues 96
4.13 Second-Order Systems 98

Chapter 5 **STABILITY** **114**
5.1 Stability Definitions 114
5.2 Characteristic Root Locations for Continuous Systems 114
5.3 Routh Stability Criterion 115
5.4 Hurwitz Stability Criterion 116
5.5 Continued Fraction Stability Criterion 117
5.6 Stability Criteria for Discrete-Time Systems 117

Chapter 6 **TRANSFER FUNCTIONS** **128**
6.1 Definition of a Continuous System Transfer Function 128
6.2 Properties of a Continuous System Transfer Function 129
6.3 Transfer Functions of Continuous Control System Compensators and Controllers 129
6.4 Continuous System Time Response 130
6.5 Continuous System Frequency Response 130
6.6 Discrete-Time System Transfer Functions, Compensators and Time Responses 132
6.7 Discrete-Time System Frequency Response 133
6.8 Combining Continuous-Time and Discrete-Time Elements 134

Chapter 7 **BLOCK DIAGRAM ALGEBRA AND TRANSFER FUNCTIONS OF SYSTEMS** **154**
7.1 Introduction 154
7.2 Review of Fundamentals 154
7.3 Blocks in Cascade 155
7.4 Canonical Form of a Feedback Control System 156
7.5 Block Diagram Transformation Theorems 156
7.6 Unity Feedback Systems 158
7.7 Superposition of Multiple Inputs 159
7.8 Reduction of Complicated Block Diagrams 160

Chapter 8 **SIGNAL FLOW GRAPHS** **179**
8.1 Introduction 179
8.2 Fundamentals of Signal Flow Graphs 179

8.3 Signal Flow Graph Algebra 180
8.4 Definitions 181
8.5 Construction of Signal Flow Graphs 182
8.6 The General Input-Output Gain Formula 184
8.7 Transfer Function Computation of Cascaded Components 186
8.8 Block Diagram Reduction Using Signal Flow Graphs and the General Input-Output Gain Formula 187

Chapter 9 **SYSTEM SENSITIVITY MEASURES AND CLASSIFICATION OF FEEDBACK SYSTEMS** **208**
9.1 Introduction 208
9.2 Sensitivity of Transfer Functions and Frequency Response Functions to System Parameters 208
9.3 Output Sensitivity to Parameters for Differential and Difference Equation Models 213
9.4 Classification of Continuous Feedback Systems by Type 214
9.5 Position Error Constants for Continuous Unity Feedback Systems 215
9.6 Velocity Error Constants for Continuous Unity Feedback Systems 216
9.7 Acceleration Error Constants for Continuous Unity Feedback Systems 217
9.8 Error Constants for Discrete Unity Feedback Systems 217
9.9 Summary Table for Continuous and Discrete-Time Unity Feedback Systems 217
9.10 Error Constants for More General Systems 218

Chapter 10 **ANALYSIS AND DESIGN OF FEEDBACK CONTROL SYSTEMS: OBJECTIVES AND METHODS** **230**
10.1 Introduction 230
10.2 Objectives of Analysis 230
10.3 Methods of Analysis 230
10.4 Design Objectives 231
10.5 System Compensation 235
10.6 Design Methods 236
10.7 The w-Transform for Discrete-Time Systems Analysis and Design Using Continuous System Methods 236
10.8 Algebraic Design of Digital Systems, Including Deadbeat Systems 238

Chapter 11 **NYQUIST ANALYSIS** **246**
11.1 Introduction 246
11.2 Plotting Complex Functions of a Complex Variable 246
11.3 Definitions 247
11.4 Properties of the Mapping $P(s)$ or $P(z)$ 249
11.5 Polar Plots 250
11.6 Properties of Polar Plots 252
11.7 The Nyquist Path 253
11.8 The Nyquist Stability Plot 256
11.9 Nyquist Stability Plots of Practical Feedback Control Systems 256
11.10 The Nyquist Stability Criterion 260
11.11 Relative Stability 262
11.12 M- and N-Circles 263

Chapter 12 NYQUIST DESIGN 299
12.1 Design Philosophy 299
12.2 Gain Factor Compensation 299
12.3 Gain Factor Compensation Using M-Circles 301
12.4 Lead Compensation 302
12.5 Lag Compensation 304
12.6 Lag-Lead Compensation 306
12.7 Other Compensation Schemes and Combinations of Compensators 308

Chapter 13 ROOT-LOCUS ANALYSIS 319
13.1 Introduction 319
13.2 Variation of Closed-Loop System Poles: The Root-Locus 319
13.3 Angle and Magnitude Criteria 320
13.4 Number of Loci 321
13.5 Real Axis Loci 321
13.6 Asymptotes 322
13.7 Breakaway Points 322
13.8 Departure and Arrival Angles 323
13.9 Construction of the Root-Locus 324
13.10 The Closed-Loop Transfer Function and the Time-Domain Response 326
13.11 Gain and Phase Margins from the Root-Locus 328
13.12 Damping Ratio from the Root-Locus for Continuous Systems 329

Chapter 14 ROOT-LOCUS DESIGN 343
14.1 The Design Problem 343
14.2 Cancellation Compensation 344
14.3 Phase Compensation: Lead and Lag Networks 344
14.4 Magnitude Compensation and Combinations of Compensators 345
14.5 Dominant Pole-Zero Approximations 348
14.6 Point Design 352
14.7 Feedback Compensation 353

Chapter 15 BODE ANALYSIS 364
15.1 Introduction 364
15.2 Logarithmic Scales and Bode Plots 364
15.3 The Bode Form and the Bode Gain for Continuous-Time Systems 365
15.4 Bode Plots of Simple Continuous-Time Frequency Response Functions and Their Asymptotic Approximations 365
15.5 Construction of Bode Plots for Continuous-Time Systems 371
15.6 Bode Plots of Discrete-Time Frequency Response Functions 373
15.7 Relative Stability 375
15.8 Closed-Loop Frequency Response 376
15.9 Bode Analysis of Discrete-Time Systems Using the w-Transform 377

Chapter 16 BODE DESIGN 387
16.1 Design Philosophy 387
16.2 Gain Factor Compensation 387
16.3 Lead Compensation for Continuous-Time Systems 388
16.4 Lag Compensation for Continuous-Time Systems 392
16.5 Lag-Lead Compensation for Continuous-Time Systems 393
16.6 Bode Design of Discrete-Time Systems 395

CONTENTS

Chapter 17 NICHOLS CHART ANALYSIS 411
17.1 Introduction 411
17.2 db Magnitude-Phase Angle Plots 411
17.3 Construction of db Magnitude-Phase Angle Plots 411
17.4 Relative Stability 416
17.5 The Nichols Chart 417
17.6 Closed-Loop Frequency Response Functions 419

Chapter 18 NICHOLS CHART DESIGN 433
18.1 Design Philosophy 433
18.2 Gain Factor Compensation 433
18.3 Gain Factor Compensation Using Constant Amplitude Curves 434
18.4 Lead Compensation for Continuous-Time Systems 435
18.5 Lag Compensation for Continuous-Time Systems 438
18.6 Lag-Lead Compensation 440
18.7 Nichols Chart Design of Discrete-Time Systems 443

Chapter 19 INTRODUCTION TO NONLINEAR CONTROL SYSTEMS 453
19.1 Introduction 453
19.2 Linearized and Piecewise-Linear Approximations of Nonlinear Systems 454
19.3 Phase Plane Methods 458
19.4 Lyapunov's Stability Criterion 463
19.5 Frequency Response Methods 466

Chapter 20 INTRODUCTION TO ADVANCED TOPICS IN CONTROL SYSTEMS ANALYSIS AND DESIGN 480
20.1 Introduction 480
20.2 Controllability and Observability 480
20.3 Time-Domain Design of Feedback Systems (State Feedback) 481
20.4 Control Systems with Random Inputs 483
20.5 Optimal Control Systems 484
20.6 Adaptive Control Systems 485

APPENDIX A 486
Some Laplace Transform Pairs Useful for Control Systems Analysis

APPENDIX B 488
Some z-Transform Pairs Useful for Control Systems Analysis

REFERENCES AND BIBLIOGRAPHY 489

APPENDIX C 491
SAMPLE Screens from the Companion *Interactive Outline*

INDEX 507

Chapter 1

Introduction

1.1 CONTROL SYSTEMS: WHAT THEY ARE

In modern usage the word *system* has many meanings. So let us begin by defining what we mean when we use this word in this book, first abstractly then slightly more specifically in relation to scientific literature.

Definition 1.1a: A **system** is an arrangement, set, or collection of things connected or related in such a manner as to form an entirety or whole.

Definition 1.1b: A **system** is an arrangement of physical components connected or related in such a manner as to form and/or act as an entire unit.

The word **control** is usually taken to mean *regulate*, *direct*, or *command*. Combining the above definitions, we have

Definition 1.2: A **control system** is an arrangement of physical components connected or related in such a manner as to command, direct, or regulate itself or another system.

In the most abstract sense it is possible to consider every physical object a control system. Everything alters its environment in some manner, if not actively then passively—like a mirror *directing* a beam of light shining on it at some acute angle. The mirror (Fig. 1-1) may be considered an elementary control system, controlling the beam of light according to the simple equation "the angle of reflection α equals the angle of incidence α."

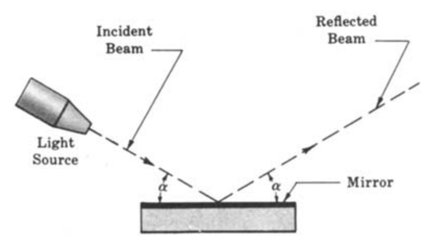


Fig. 1-1

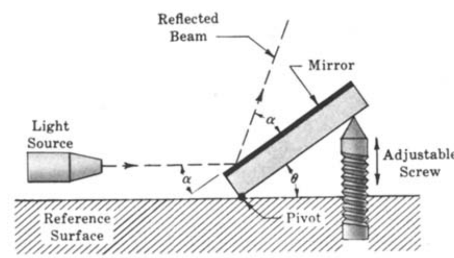


Fig. 1-2

In engineering and science we usually restrict the meaning of control systems to apply to those systems whose major function is to *dynamically* or *actively* command, direct, or regulate. The system shown in Fig. 1-2, consisting of a mirror pivoted at one end and adjusted up and down with a screw at the other end, is properly termed a *control system*. The angle of reflected light is regulated by means of the screw.

It is important to note, however, that control systems of interest for analysis or design purposes include not only those manufactured by humans, but those that normally exist in nature, and control systems with both manufactured and natural components.

1.2 EXAMPLES OF CONTROL SYSTEMS

Control systems abound in our environment. But before exemplifying this, we define two terms: *input* and *output*, which help in identifying, delineating, or defining a control system.

Definition 1.3: The **input** is the stimulus, excitation or command applied *to* a control system, typically from an external energy source, usually in order to produce a specified response *from* the control system.

Definition 1.4: The **output** is the actual response obtained from a control system. It may or may not be equal to the specified response implied by the input.

Inputs and outputs can have many different forms. Inputs, for example, may be physical variables, or more abstract quantities such as *reference*, *setpoint*, or *desired* values for the output of the control system.

The purpose of the control system usually identifies or defines the output and input. If the output and input are given, it is possible to identify, delineate, or define the nature of the system components.

Control systems may have more than one input or output. Often all inputs and outputs are well defined by the system description. But sometimes they are not. For example, an atmospheric electrical storm may intermittently interfere with radio reception, producing an unwanted output from a loudspeaker in the form of static. This "noise" output is part of the total output as defined above, but for the purpose of simply identifying a system, spurious inputs producing undesirable outputs are not normally considered as inputs and outputs in the system description. However, it is usually necessary to carefully consider these extra inputs and outputs when the system is examined in detail.

The terms input and output also may be used in the description of any type of system, whether or not it is a control system, and a control system may be part of a larger system, in which case it is called a **subsystem** or **control subsystem**, and its inputs and outputs may then be internal variables of the larger system.

EXAMPLE 1.1. An *electric switch* is a manufactured control system, controlling the flow of electricity. By definition, the apparatus or person flipping the switch is not a part of this control system.

Flipping the switch on or off may be considered as the input. That is, the input can be in one of two states, on or off. The output is the flow or nonflow (two states) of electricity.

The electric switch is one of the most rudimentary control systems.

EXAMPLE 1.2. A *thermostatically controlled heater or furnace automatically regulating the temperature of a room or enclosure* is a control system. The input to this system is a reference temperature, usually specified by appropriately setting a thermostat. The output is the actual temperature of the room or enclosure.

When the thermostat detects that the output is less than the input, the furnace provides heat until the temperature of the enclosure becomes equal to the reference input. Then the furnace is automatically turned off. When the temperature falls somewhat below the reference temperature, the furnace is turned on again.

EXAMPLE 1.3. The seemingly simple act of *pointing at an object with a finger* requires a biological control system consisting chiefly of the eyes, the arm, hand and finger, and the brain. The input is the precise direction of the object (moving or not) with respect to some reference, and the output is the actual pointed direction with respect to the same reference.

EXAMPLE 1.4. A part of the human temperature control system is the *perspiration system*. When the temperature of the air exterior to the skin becomes too high the sweat glands secrete heavily, inducing cooling of the skin by evaporation. Secretions are reduced when the desired cooling effect is achieved, or when the air temperature falls sufficiently.

The input to this system may be "normal" or comfortable skin temperature, a "setpoint," or the air temperature, a physical variable. The output is the actual skin temperature.

EXAMPLE 1.5. The control system consisting of *a person driving an automobile* has components which are clearly both manufactured and biological. The driver wants to keep the automobile in the appropriate lane of the roadway. He or she accomplishes this by constantly watching the direction of the automobile with respect to the direction of the road. In this case, the direction or heading of the road, represented by the painted guide line or lines on either side of the lane may be considered as the input. The heading of the automobile is the output of the system. The driver controls this output by constantly measuring it with his or her eyes and brain, and correcting it with his or her hands on the steering wheel. The major components of this control system are the driver's hands, eyes and brain, and the vehicle.

1.3 OPEN-LOOP AND CLOSED-LOOP CONTROL SYSTEMS

Control systems are classified into two general categories: *open-loop* and *closed-loop* systems. The distinction is determined by the **control action**, that quantity responsible for activating the system to produce the output.

The term *control action* is classical in the control systems literature, but the word *action* in this expression does not always *directly* imply change, motion, or activity. For example, the control action in a system designed to have an object hit a target is usually the *distance* between the object and the target. Distance, as such, is not an action, but action (motion) is implied here, because the goal of such a control system is to reduce this distance to zero.

Definition 1.5: An **open-loop** control system is one in which the control action is independent of the output.

Definition 1.6: A **closed-loop** control system is one in which the control action is somehow dependent on the output.

Two outstanding features of open-loop control systems are:

1. Their ability to perform accurately is determined by their calibration. To **calibrate** means to establish or reestablish the input-output relation to obtain a desired system accuracy.
2. They are not usually troubled with problems of *instability*, a concept to be subsequently discussed in detail.

Closed-loop control systems are more commonly called *feedback* control systems, and are considered in more detail beginning in the next section.

To classify a control system as open-loop or closed-loop, we must distinguish clearly the components of the system from components that interact with but are not part of the system. For example, the driver in Example 1.5 was defined as part of that control system, but a human operator may or may not be a component of a system.

EXAMPLE 1.6. Most *automatic toasters* are open-loop systems because they are controlled by a timer. The time required to make "good toast" must be estimated by the user, who is not part of the system. Control over the quality of toast (the output) is removed once the time, which is both the input and the control action, has been set. The time is typically set by means of a calibrated dial or switch.

EXAMPLE 1.7. An *autopilot mechanism and the airplane it controls* is a closed-loop (feedback) control system. Its purpose is to maintain a specified airplane heading, despite atmospheric changes. It performs this task by continuously measuring the actual airplane heading, and automatically adjusting the airplane control surfaces (rudder, ailerons, etc.) so as to bring the actual airplane heading into correspondence with the specified heading. The human pilot or operator who presets the autopilot is not part of the control system.

1.4 FEEDBACK

Feedback is that characteristic of closed-loop control systems which distinguishes them from open-loop systems.

Definition 1.7: **Feedback** is that property of a closed-loop system which permits the output (or some other controlled variable) to be compared with the input to the system (or an input to some other internally situated component or subsystem) so that the appropriate control action may be formed as some function of the output and input.

More generally, feedback is said to exist in a system when a *closed* sequence of cause-and-effect relations exists between system variables.

EXAMPLE 1.8. The concept of feedback is clearly illustrated by the autopilot mechanism of Example 1.7. The input is the specified heading, which may be set on a dial or other instrument of the airplane control panel, and the output is the actual heading, as determined by automatic navigation instruments. A comparison device continuously monitors the input and output. When the two are in correspondence, control action is not required. When a difference exists between the input and output, the comparison device delivers a control action signal to the controller, the autopilot mechanism. The controller provides the appropriate signals to the control surfaces of the airplane to reduce the input-output difference. Feedback may be effected by mechanical or electrical connections from the navigation instruments, measuring the heading, to the comparison device. In practice, the comparison device may be integrated within the autopilot mechanism.

1.5 CHARACTERISTICS OF FEEDBACK

The presence of feedback typically imparts the following properties to a system.

1. Increased accuracy. For example, the ability to faithfully reproduce the input. This property is illustrated throughout the text.
2. Tendency toward oscillation or instability. This all-important characteristic is considered in detail in Chapters 5 and 9 through 19.
3. Reduced sensitivity of the ratio of output to input to variations in system parameters and other characteristics (Chapter 9).
4. Reduced effects of nonlinearities (Chapters 3 and 19).
5. Reduced effects of external disturbances or noise (Chapters 7, 9, and 10).
6. Increased bandwidth. The **bandwidth** of a system is a frequency response measure of how well the system responds to (or filters) variations (or frequencies) in the input signal (Chapters 6, 10, 12, and 15 through 18).

1.6 ANALOG AND DIGITAL CONTROL SYSTEMS

The signals in a control system, for example, the input and the output waveforms, are typically functions of some independent variable, usually time, denoted t.

Definition 1.8: A signal dependent on a continuum of values of the independent variable t is called a **continuous-time** signal or, more generally, a **continuous-data** signal or (less frequently) an **analog** signal.

Definition 1.9: A signal defined at, or of interest at, only discrete (distinct) instants of the independent variable t (upon which it depends) is called a **discrete-time**, a **discrete-data**, a **sampled-data**, or a **digital** signal.

We remark that *digital* is a somewhat more specialized term, particularly in other contexts. We use it as a synonym here because it is the convention in the control systems literature.

EXAMPLE 1.9. The continuous, sinusoidally varying voltage $v(t)$ or alternating current $i(t)$ available from an ordinary household electrical receptable is a continuous-time (analog) signal, because it is defined at *each and every instant* of time t electrical power is available from that outlet.

EXAMPLE 1.10. If a lamp is connected to the receptacle in Example 1.9, and it is switched on and then immediately off every minute, the light from the lamp is a discrete-time signal, on only for an instant every minute.

EXAMPLE 1.11. The mean temperature T in a room at precisely 8 A.M. (08 hours) each day is a discrete-time signal. This signal may be denoted in several ways, depending on the application; for example $T(8)$ for the temperature at 8 o'clock—rather than another time; $T(1), T(2), \dots$ for the temperature at 8 o'clock on day 1, day 2, etc., or, equivalently, using a subscript notation, T_1, T_2, etc. Note that these discrete-time signals are *sampled* values of a continuous-time signal, the mean temperature of the room at all times, denoted $T(t)$.

EXAMPLE 1.12. The signals inside digital computers and microprocessors are inherently discrete-time, or discrete-data, or digital (or digitally coded) signals. At their most basic level, they are typically in the form of sequences of voltages, currents, light intensities, or other physical variables, at either of two constant levels, for example, ± 15 V; light-on, light-off; etc. These *binary signals* are usually represented in alphanumeric form (numbers, letters, or other characters) at the inputs and outputs of such digital devices. On the other hand, the signals of analog computers and other analog devices are continuous-time.

Control systems can be classified according to the types of signals they process: continuous-time (analog), discrete-time (digital), or a combination of both (hybrid).

Definition 1.10: **Continuous-time control systems**, also called **continuous-data control systems**, or **analog control systems**, contain or process only continuous-time (analog) signals and components.

Definition 1.11: **Discrete-time control systems**, also called **discrete-data control systems**, or **sampled-data control systems**, have discrete-time signals or components at one or more points in the system.

We note that discrete-time control systems can have continuous-time as well as discrete-time signals; that is, they can be hybrid. The distinguishing factor is that a discrete-time or digital control system *must* include at least one discrete-data signal. Also, digital control systems, particularly of sampled-data type, often have both open-loop and closed-loop modes of operation.

EXAMPLE 1.13. A target tracking and following system, such as the one described in Example 1.3 (tracking and pointing at an object with a finger), is usually considered an analog or continuous-time control system, because the distance between the "tracker" (finger) and the target is a continuous function of time, and the objective of such a control system is to *continuously* follow the target. The system consisting of a person driving an automobile (Example 1.5) falls in the same category. Strictly speaking, however, tracking systems, both natural and manufactured, can have digital signals or components. For example, control signals from the brain are often treated as "pulsed" or discrete-time data in more detailed models which include the brain, and digital computers or microprocessors have replaced many analog components in vehicle control systems and tracking mechanisms.

EXAMPLE 1.14. A closer look at the thermostatically controlled heating system of Example 1.2 indicates that it is actually a sampled-data control system, with both digital and analog components and signals. If the desired room temperature is, say, 68°F (22°C) on the thermostat and the room temperature falls below, say, 66°F, the thermostat switching system closes the circuit to the furnace (an analog device), turning it on until the temperature of the room reaches, say, 70°F. Then the switching system automatically turns the furnace off until the room temperature again falls below 66°F. This control system is actually operating open-loop between switching instants, when the thermostat turns the furnace on or off, but overall operation is considered closed-loop. The thermostat receives a continuous-time signal at its input, the actual room temperature, and it delivers a discrete-time (binary) switching signal at its output, turning the furnace on or off. Actual room temperature thus varies continuously between 66° and 70°F, and *mean* temperature is controlled at about 68°F, the *setpoint* of the thermostat.

The terms discrete-time and discrete-data, sampled-data, and continuous-time and continuous-data are often abbreviated as *discrete*, *sampled*, and *continuous* in the remainder of the book, wherever the meaning is unambiguous. *Digital* or *analog* is also used in place of discrete (sampled) or continuous where appropriate and when the meaning is clear from the context.

1.7 THE CONTROL SYSTEMS ENGINEERING PROBLEM

Control systems engineering consists of *analysis* and *design* of control systems configurations.

Analysis is the investigation of the properties of an existing system. The **design** problem is the choice and arrangement of system components to perform a specific task.

Two methods exist for design:

1. Design by analysis
2. Design by synthesis

Design by analysis is accomplished by modifying the characteristics of an existing or standard system configuration, and **design by synthesis** by defining the form of the system directly from its specifications.

1.8 CONTROL SYSTEM MODELS OR REPRESENTATIONS

To solve a control systems problem, we must put the specifications or description of the system configuration and its components into a form amenable to analysis or design.

Three basic representations (models) of components and systems are used extensively in the study of control systems:

1. Mathematical models, in the form of differential equations, difference equations, and/or other mathematical relations, for example, Laplace- and z-transforms
2. Block diagrams
3. Signal flow graphs

Mathematical models of control systems are developed in Chapters 3 and 4. Block diagrams and signal flow graphs are shorthand, graphical representations of either the schematic diagram of a system, or the set of mathematical equations characterizing its parts. Block diagrams are considered in detail in Chapters 2 and 7, and signal flow graphs in Chapter 8.

Mathematical models are needed when quantitative relationships are required, for example, to represent the detailed behavior of the output of a feedback system to a given input. Development of mathematical models is usually based on principles from the physical, biological, social, or information sciences, depending on the control system application area, and the complexity of such models varies widely. One class of models, commonly called *linear systems*, has found very broad application in control system science. Techniques for solving linear system models are well established and documented in the literature of applied mathematics and engineering, and the major focus of this book is linear feedback control systems, their analysis and their design. Continuous-time (continuous, analog) systems are emphasized, but discrete-time (discrete, digital) systems techniques are also developed throughout the text, in a unifying but not exhaustive manner. Techniques for analysis and design of *nonlinear* control systems are the subject of Chapter 19, by way of introduction to this more complex subject.

In order to communicate with as many readers as possible, the material in this book is developed from basic principles in the sciences and applied mathematics, and specific applications in various engineering and other disciplines are presented in the examples and in the solved problems at the end of each chapter.

Solved Problems

INPUT AND OUTPUT

1.1. Identify the input and output for the pivoted, adjustable mirror of Fig. 1-2.

The input is the angle of inclination of the mirror θ, varied by turning the screw. The output is the angular position of the reflected beam $\theta + \alpha$ from the reference surface.

1.2. Identify a possible input and a possible output for a rotational generator of electricity.

The input may be the rotational speed of the prime mover (e.g., a steam turbine), in revolutions per minute. Assuming the generator has no load attached to its output terminals, the output may be the induced voltage at the output terminals.

Alternatively, the input can be expressed as angular momentum of the prime mover shaft, and the output in units of electrical power (watts) with a load attached to the generator.

1.3. Identify the input and output for an automatic washing machine.

Many washing machines operate in the following manner. After the clothes have been put into the machine, the soap or detergent, bleach, and water are entered in the proper amounts. The wash and spin cycle-time is then set on a timer and the washer is energized. When the cycle is completed, the machine shuts itself off.

If the proper amounts of detergent, bleach, and water, and the appropriate temperature of the water are predetermined or specified by the machine manufacturer, or automatically entered by the machine itself, then the input is the time (in minutes) for the wash and spin cycle. The timer is usually set by a human operator.

The output of a washing machine is more difficult to identify. Let us define *clean* as the absence of foreign substances from the items to be washed. Then we can identify the output as the percentage of cleanliness. At the start of a cycle the output is less than 100%, and at the end of a cycle the output is ideally equal to 100% (*clean* clothes are not always obtained).

For most coin-operated machines the cycle-time is preset, and the machine begins operating when the coin is entered. In this case, the percentage of cleanliness can be controlled by adjusting the amounts of detergent, bleach, water, and the temperature of the water. We may consider all of these quantities as inputs.

Other combinations of inputs and outputs are also possible.

1.4. Identify the organ-system components, and the input and output, and describe the operation of the biological control system consisting of a human being reaching for an object.

The basic components of this intentionally oversimplified control system description are the brain, arm and hand, and eyes.

The brain sends the required nervous system signal to the arm and hand to reach for the object. This signal is amplified in the muscles of the arm and hand, which serve as power actuators for the system. The eyes are employed as a sensing device, continuously "feeding back" the position of the hand to the brain.

Hand position is the output for the system. The input is object position.

The objective of the control system is to reduce the distance between hand position and object position to zero. Figure 1-3 is a schematic diagram. The dashed lines and arrows represent the direction of information flow.

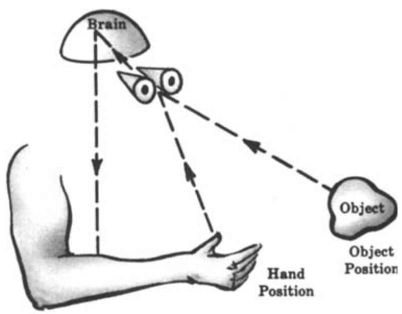


Fig. 1-3

OPEN-LOOP AND CLOSED-LOOP SYSTEMS

1.5. Explain how a closed-loop automatic washing machine might operate.

Assume all quantities described as possible inputs in Problem 1.3, namely cycle-time, water volume, water temperature, amount of detergent, and amount of bleach, can be adjusted by devices such as valves and heaters.

A closed-loop automatic washer might continuously or periodically measure the percentage of cleanliness (output) of the items being washing, adjust the input quantities accordingly, and turn itself off when 100% cleanliness has been achieved.

1.6. How are the following open-loop systems calibrated: (*a*) automatic washing machine, (*b*) automatic toaster, (*c*) voltmeter?

(*a*) Automatic washing machines are calibrated by estimating any combination of the following input quantities: (1) amount of detergent, (2) amount of bleach or other additives, (3) amount of water, (4) temperature of the water, (5) cycle-time.

On some washing machines one or more of these inputs is (are) predetermined. The remaining quantities must be estimated by the user and depend upon factors such as degree of hardness of the water, type of detergent, and type or strength of the bleach or other additives. Once this calibration has been determined for a specific type of wash (e.g., all white clothes, very dirty clothes), it does not normally have to be redetermined during the lifetime of the machine. If the machine breaks down and replacement parts are installed, recalibration may be necessary.

(*b*) Although the timer dial for most automatic toasters is calibrated by the manufacturer (e.g., light-medium-dark), the amount of heat produced by the heating element may vary over a wide range. In addition, the efficiency of the heating element normally deteriorates with age. Hence the amount of time required for "good toast" must be estimated by the user, and this setting usually must be periodically readjusted. At first, the toast is usually too light or too dark. After several successively different estimates, the required toasting time for a desired quality of toast is obtained.

(*c*) In general, a voltmeter is calibrated by comparing it with a known-voltage standard source, and appropriately marking the reading scale at specified intervals.

1.7. Identify the control action in the systems of Problems 1.1, 1.2, and 1.4.

Mathcad

For the mirror system of Problem 1.1 the control action is equal to the input, that is, the angle of inclination of the mirror θ. For the generator of Problem 1.2 the control action is equal to the input, the rotational speed or angular momentum of the prime mover shaft. The control action of the human reaching system of Problem 1.4 is equal to the distance between hand and object position.

1.8. Which of the control systems in Problems 1.1, 1.2, and 1.4 are open-loop? Closed-loop?

Mathcad

Since the control action is equal to the input for the systems of Problems 1.1 and 1.2, no feedback exists and the systems are open-loop. The human reaching system of Problem 1.4 is closed-loop because the control action is dependent upon the output, hand position.

1.9. Identify the control action in Examples 1.1 through 1.5.

The control action for the electric switch of Example 1.1 is equal to the input, the on or off command. The control action for the heating system of Example 1.2 is equal to the difference between the reference and actual room temperatures. For the finger pointing system of Example 1.3, the control action is equal to the difference between the actual and pointed direction of the object. The perspiration system of Example 1.4 has its control action equal to the difference between the "normal" and actual skin surface temperature. The difference between the direction of the road and the heading of the automobile is the control action for the human driver and automobile system of Example 1.5.

1.10. Which of the control systems in Examples 1.1 through 1.5 are open-loop? Closed-loop?

The electric switch of Example 1.1 is open-loop because the control action is equal to the input, and therefore independent of the output. For the remaining Examples 1.2 through 1.5 the control action is clearly a function of the output. Hence they are closed-loop systems.

FEEDBACK

1.11. Consider the voltage divider network of Fig. 1-4. The output is v_2 and the input is v_1.

Mathcad

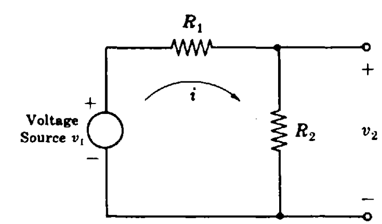


Fig. 1-4

(*a*) Write an equation for v_2 as a function of v_1, R_1, and R_2. That is, write an equation for v_2 which yields an open-loop system.

(*b*) Write an equation for v_2 in closed-loop form, that is, v_2 as a function of v_1, v_2, R_1, and R_2.

This problem illustrates how a passive network can be characterized as either an open-loop or a closed-loop system.

(*a*) From Ohm's law and Kirchhoff's voltage and current laws we have

$$v_2 = R_2 i \qquad i = \frac{v_1}{R_1 + R_2}$$

Therefore

$$v_2 = \left(\frac{R_2}{R_1 + R_2}\right) v_1 = f(v_1, R_1, R_2)$$

(*b*) Writing the current i in a slightly different form, we have $i = (v_1 - v_2)/R_1$. Hence

$$v_2 = R_2\left(\frac{v_1 - v_2}{R_1}\right) = \left(\frac{R_2}{R_1}\right) v_1 - \left(\frac{R_2}{R_1}\right) v_2 = f(v_1, v_2, R_1, R_2)$$

1.12. Explain how the classical economic concept known as the Law of Supply and Demand can be interpreted as a feedback control system. Choose the market price (selling price) of a particular item as the output of the system, and assume the objective of the system is to maintain price stability.

The Law can be stated in the following manner. The market *demand* for the item decreases as its price increases. The market *supply* usually increases as its price increases. The Law of Supply and Demand says that a stable market price is achieved if and only if the supply is equal to the demand.

The manner in which the price is regulated by the supply and the demand can be described with feedback control concepts. Let us choose the following four basic elements for our system: the Supplier, the Demander, the Pricer, and the Market where the item is bought and sold. (In reality, these elements generally represent very complicated processes.)

The input to our idealized economic system is *price stability* the "desired" output. A more convenient way to describe this input is *zero price fluctuation*. The output is the actual market price.

The system operates as follows: The Pricer receives a command (zero) for price stability. It estimates a price for the Market transaction with the help of information from its memory or records of past transactions. This price causes the Supplier to produce or supply a certain number of items, and the Demander to demand a number of items. The difference between the supply and the demand is the control action for this system. If the control action is nonzero, that is, if the supply is not equal to the demand, the Pricer initiates a change in the market price in a direction which makes the supply eventually equal to the demand. Hence both the Supplier and the Demander may be considered the feedback, since they determine the control action.

MISCELLANEOUS PROBLEMS

1.13. (*a*) Explain the operation of ordinary traffic signals which control automobile traffic at roadway intersections. (*b*) Why are they open-loop control systems? (*c*) How can traffic be controlled more efficiently? (*d*) Why is the system of (*c*) closed-loop?

(*a*) Traffic lights control the flow of traffic by successively confronting the traffic in a particular direction (e.g., north-south) with a red (stop) and then a green (go) light. When one direction has the green signal, the cross traffic in the other direction (east-west) has the red. Most traffic signal red and green light intervals are predetermined by a calibrated timing mechanism.

(*b*) Control systems operated by preset timing mechanisms are open-loop. The control action is equal to the input, the red and green intervals.

(*c*) Besides preventing collisions, it is a function of traffic signals to generally control the *volume* of traffic. For the open-loop system described above, the volume of traffic does not influence the preset red and green timing intervals. In order to make traffic flow more smoothly, the green light timing interval must be made longer than the red in the direction containing the greater traffic volume. Often a traffic officer performs this task.

The ideal system would automatically measure the volume of traffic in all directions, using appropriate sensing devices, compare them, and use the difference to control the red and green time intervals, an ideal task for a computer.

(*d*) The system of (*c*) is closed-loop because the control action (the difference between the volume of traffic in each direction) is a function of the output (actual traffic volume flowing past the intersection in each direction).

1.14. (*a*) Describe, in a simplified way, the components and variables of the biological control system involved in walking in a prescribed direction. (*b*) Why is walking a closed-loop operation? (*c*) Under what conditions would the human walking apparatus become an open-loop system? A sampled-data system? Assume the person has normal vision.

(*a*) The major components involved in walking are the brain, eyes, and legs and feet. The input may be chosen as the desired walk direction, and the output the actual walk direction. The control action is determined by the eyes, which detect the difference between the input and output and send this information to the brain. The brain commands the legs and feet to walk in the prescribed direction.

(*b*) Walking is a closed-loop operation because the control action is a function of the output.

(*c*) If the eyes are closed, the feedback loop is broken and the system becomes open-loop. If the eyes are opened and closed periodically, the system becomes a sampled-data one, and walking is usually more accurately controlled than with the eyes always closed.

1.15. Devise a control system to fill a container with water after it is emptied through a stopcock at the bottom. The system must automatically shut off the water when the container is filled.

The simplified schematic diagram (Fig. 1-5) illustrates the principle of the ordinary toilet tank filling system.

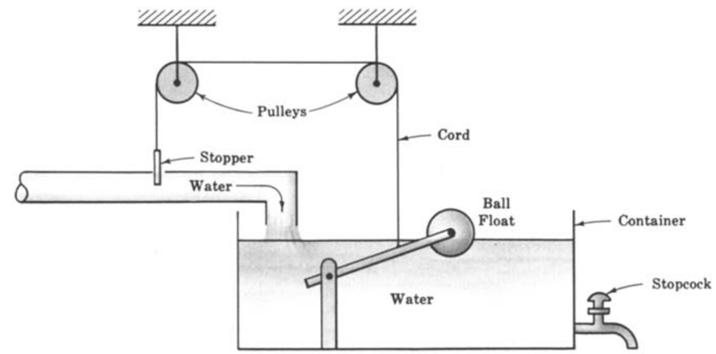


Fig. 1-5

The ball floats on the water. As the ball gets closer to the top of the container, the stopper decreases the flow of water. When the container becomes full, the stopper shuts off the flow of water.

1.16. Devise a simple control system which automatically turns on a room lamp at dusk, and turns it off in daylight.

A simple system that accomplishes this task is shown in Fig. 1-6.

At dusk, the photocell, which functions as a light-sensitive switch, closes the lamp circuit, thereby lighting the room. The lamp stays lighted until daylight, at which time the photocell detects the bright outdoor light and opens the lamp circuit.

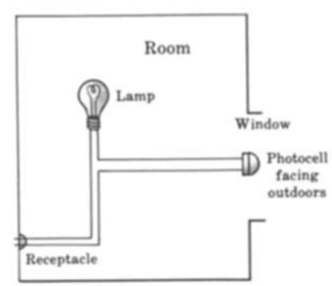


Fig. 1-6

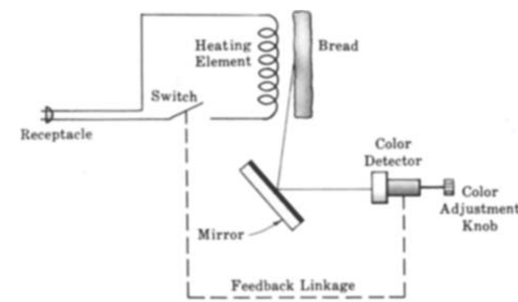


Fig. 1-7

1.17. Devise a closed-loop automatic toaster.

Assume each heating element supplies the same amount of heat to both sides of the bread, and toast quality can be determined by its color. A simplified schematic diagram of one possible way to apply the feedback principle to a toaster is shown in Fig. 1-7. Only one side of the toaster is illustrated.

The toaster is initially calibrated for a desired toast quality by means of the color adjustment knob. This setting never needs readjustment unless the toast quality criterion changes. When the switch is closed, the bread is toasted until the color detector "sees" the desired color. Then the switch is automatically opened by means of the feedback linkage, which may be electrical or mechanical.

1.18. Is the voltage divider network in Problem 1.11 an analog or digital device? Also, are the input and output analog or digital signals?

It is clearly an analog device, as are all electrical networks consisting only of passive elements such as resistors, capacitors, and inductors. The voltage source v_1 is considered an external input to this network. If it produces a continuous signal, for example, from a battery or alternating power source, the output is a continuous or analog signal. However, if the voltage source v_1 is a discrete-time or digital signal, then so is the output $v_2 = v_1 R_2/(R_1 + R_2)$. Also, if a switch were included in the circuit, in series with an analog voltage source, intermittent opening and closing of the switch would generate a sampled waveform of the voltage source v_1 and therefore a sampled or discrete-time output from this analog network.

1.19. Is the system that controls the total cash value of a bank account a continuous or a discrete-time system? Why? Assume a deposit is made only once, and no withdrawals are made.

If the bank pays no interest and extracts no fees for maintaining the account (like putting your money "under the mattress"), the system controlling the total cash value of the account can be considered continuous, because the value is always the same. Most banks, however, pay interest periodically, for example, daily, monthly, or yearly, and the value of the account therefore changes periodically, *at discrete times*. In this case, the system controlling the cash value of the account is a *discrete system*. Assuming no withdrawals, the interest is added to the principle each time the account earns interest, called *compounding*, and the account value continues to grow without bound (the "greatest invention of mankind," a comment attributed to Einstein).

1.20. What *type* of control system, open-loop or closed-loop, continuous or discrete, is used by an ordinary stock market investor, whose objective is to profit from his or her investment.

Stock market investors typically follow the progress of their stocks, for example, their prices, periodically. They might check the bid and ask prices daily, with their broker or the daily newspaper, or more or less often, depending upon individual circumstances. In any case, they periodically *sample* the pricing signals and therefore the system is sampled-data, or discrete-time. However, stock prices normally rise and fall between sampling times and therefore the system operates open-loop during these periods. The feedback loop is closed only when the investor makes his or her periodic observations and acts upon the information received, which may be to buy, sell, or do nothing. Thus overall control is closed-loop. The measurement (sampling) process could, of course, be handled more efficiently using a computer, which also can be programed to make decisions based on the information it receives. In this case the control system remains discrete-time, but not only because there is a digital computer in the control loop. Bid and ask prices do not change continuously but are inherently discrete-time signals.

Supplementary Problems

1.21. Identify the input and output for an automatic temperature-regulating oven.

1.22. Identify the input and output for an automatic refrigerator.

1.23. Identify an input and an output for an electric automatic coffeemaker. Is this system open-loop or closed-loop?

1.24. Devise a control system to automatically raise and lower a lift-bridge to permit ships to pass. No continuous human operator is permissible. The system must function entirely automatically.

1.25. Explain the operation and identify the pertinent quantities and components of an automatic, radar-controlled antiaircraft gun. Assume that no operator is required except to initially put the system into an operational mode.

1.26. How can the electrical network of Fig. 1-8 be given a *feedback* control system interpretation? Is this system analog or digital?

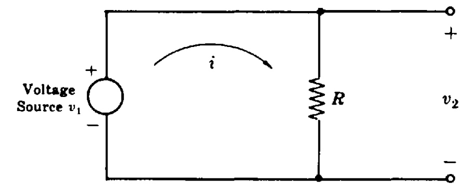


Fig. 1-8

1.27. Devise a control system for positioning the rudder of a ship from a control room located far from the rudder. The objective of the control system is to steer the ship in a desired heading.

1.28. What inputs in addition to the command for a desired heading would you expect to find acting on the system of Problem 1.27?

1.29. Can the application of "laissez faire capitalism" to an economic system be interpreted as a feedback control system? Why? How about "socialism" in its purest form? Why?

1.30. Does the operation of a stock exchange, for example, buying and selling equities, fit the model of the Law of Supply and Demand described in Problem 1.12? How?

1.31. Does a purely socialistic economic system fit the model of the Law of Supply and Demand described in Problem 1.12? Why (or why not)?

1.32. Which control systems in Problems 1.1 through 1.4 and 1.12 through 1.17 are digital or sampled-data and which are continuous or analog? Define the continuous signals and the discrete signals in each system.

1.33. Explain why economic control systems based on data obtained from typical accounting procedures are sampled-data control systems? Are they open-loop or closed-loop?

1.34. Is a rotating antenna radar system, which normally receives range and directional data once each revolution, an analog or a digital system?

1.35. What type of control system is involved in the treatment of a patient by a doctor, based on data obtained from laboratory analysis of a sample of the patient's blood?

Answers to Some Supplementary Problems

1.21. The input is the reference temperature. The output is the actual oven temperature.

1.22. The input is the reference temperature. The output is the actual refrigerator temperature.

1.23. One possible input for the automatic electric coffeemaker is the amount of coffee used. In addition, most coffeemakers have a dial which can be set for weak, medium, or strong coffee. This setting usually regulates a timing mechanism. The brewing time is therefore another possible input. The output of any coffeemaker can be chosen as coffee strength. The coffeemakers described above are open-loop.

Chapter 2

Control Systems Terminology

2.1 BLOCK DIAGRAMS: FUNDAMENTALS

A **block diagram** is a shorthand, pictorial representation of the cause-and-effect relationship between the input and output of a physical system. It provides a convenient and useful method for characterizing the functional relationships among the various components of a control system. System *components* are alternatively called *elements* of the system. The simplest form of the block diagram is the single *block*, with one input and one output, as shown in Fig. 2-1.

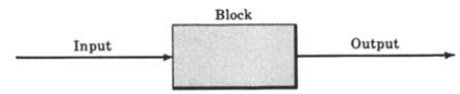


Fig. 2-1

The interior of the rectangle representing the block usually contains a description of or the name of the element, or the symbol for the mathematical operation to be performed on the input to yield the output. The *arrows* represent the direction of information or signal flow.

EXAMPLE 2.1

Fig. 2-2

The operations of addition and subtraction have a special representation. The block becomes a small circle, called a **summing point**, with the appropriate plus or minus sign associated with the arrows entering the circle. The output is the algebraic sum of the inputs. Any number of inputs may enter a summing point.

EXAMPLE 2.2

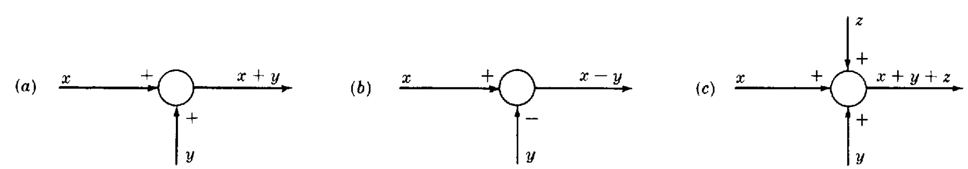


Fig. 2-3

Some authors put a cross in the circle: (Fig. 2-4)

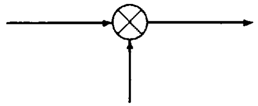


Fig. 2-4

This notation is avoided here because it is sometimes confused with the multiplication operation.

In order to have the same signal or variable be an input to more than one block or summing point, a **takeoff point** is used. This permits the signal to proceed unaltered along several different paths to several destinations.

EXAMPLE 2.3

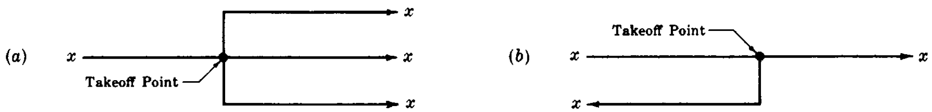


Fig. 2-5

2.2 BLOCK DIAGRAMS OF CONTINUOUS (ANALOG) FEEDBACK CONTROL SYSTEMS

The blocks representing the various components of a control system are connected in a fashion which characterizes their functional relationships within the system. The basic configuration of a simple closed-loop (feedback) control system with a single input and a single output (abbreviated SISO) is illustrated in Fig. 2-6 for a system with continuous signals only.

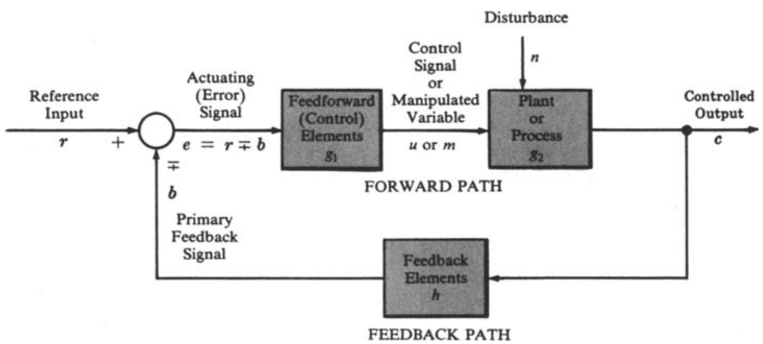


Fig. 2-6

We emphasize that the arrows of the closed loop, connecting one block with another, represent the direction of flow of *control* energy or information, which is not usually the main source of energy for the system. For example, the major source of energy for the thermostatically controlled furnace of Example

1.2 is often chemical, from burning fuel oil, coal, or gas. But this energy source would not appear in the closed control loop of the system.

2.3 TERMINOLOGY OF THE CLOSED-LOOP BLOCK DIAGRAM

It is important that the terms used in the closed-loop block diagram be clearly understood.

Lowercase letters are used to represent the input and output variables of each element as well as the symbols for the blocks g_1, g_2, and h. These quantities represent functions of time, unless otherwise specified.

EXAMPLE 2.4. $r = r(t)$

In subsequent chapters, we use capital letters to denote Laplace transformed or z-transformed quantities, as functions of the complex variable s, or z, respectively, or Fourier transformed quantities (frequency functions), as functions of the pure imaginary variable $j\omega$. Functions of s or z are often abbreviated to the capital letter appearing alone. Frequency functions are never abbreviated.

EXAMPLE 2.5. $R(s)$ may be abbreviated as R, or $F(z)$ as F. $R(j\omega)$ is never abbreviated.

The letters r, c, e, etc., were chosen to preserve the generic nature of the block diagram. This convention is now classical.

Definition 2.1: The **plant** (or **process**, or **controlled system**) g_2 is the system, subsystem, process, or object controlled by the feedback control system.

Definition 2.2: The **controlled output** c is the output variable of the plant, under the control of the feedback control system.

Definition 2.3: The **forward path** is the transmission path from the summing point to the controlled output c.

Definition 2.4: The **feedforward (control) elements** g_1 are the components of the forward path that generate the control signal u or m applied to the plant. *Note:* Feedforward elements typically include controller(s), compensator(s) (or equalization elements), and/or amplifiers.

Definition 2.5: The **control signal** u (or **manipulated variable** m) is the output signal of the feedforward elements g_1 applied as input to the plant g_2.

Definition 2.6: The **feedback path** is the transmission path from the controlled output c back to the summing point.

Definition 2.7: The **feedback elements** h establish the functional relationship between the controlled output c and the primary feedback signal b. *Note:* Feedback elements typically include sensors of the controlled output c, compensators, and/or controller elements.

Definition 2.8: The **reference input** r is an external signal applied to the feedback control system, usually at the first summing point, in order to command a specified action of the plant. It usually represents ideal (or desired) plant output behavior.

Definition 2.9: The **primary feedback signal** b is a function of the controlled output c, algebraically summed with the reference input r to obtain the actuating (error) signal e, that is, $r \pm b = e$. *Note:* An *open-loop* system has no primary feedback signal.

Definition 2.10: The **actuating** (or **error**) **signal** is the reference input signal r plus or minus the primary feedback signal b. The *control action* is generated by the actuating (error) signal in a feedback control system (see Definitions 1.5 and 1.6). *Note:* In an *open-loop* system, which has no feedback, the actuating signal is equal to r.

Definition 2.11: **Negative feedback** means the summing point is a subtractor, that is, $e = r - b$. **Positive feedback** means the summing point is an adder, that is, $e = r + b$.

2.4 BLOCK DIAGRAMS OF DISCRETE-TIME (SAMPLED-DATA, DIGITAL) COMPONENTS, CONTROL SYSTEMS, AND COMPUTER-CONTROLLED SYSTEMS

A *discrete-time (sampled-data or digital) control system* was defined in Definition 1.11 as one having discrete-time signals or components at one or more points in the system. We introduce several common discrete-time system components first, and then illustrate some of the ways they are interconnected in digital control systems. We remind the reader here that *discrete-time* is often abbreviated as *discrete* in this book, and *continuous-time* as *continuous*, wherever the meaning is unambiguous.

EXAMPLE 2.6. A digital computer or microprocessor is a discrete-time (discrete or digital) device, a common component in digital control systems. The internal and external signals of a digital computer are typically discrete-time or digitally coded.

EXAMPLE 2.7. A discrete system component (or components) with discrete-time input $u(t_k)$ and discrete-time output $y(t_k)$ signals, where t_k are discrete instants of time, $k = 1, 2, \ldots,$ etc., may be represented by a block diagram, as shown in Fig. 2-7.

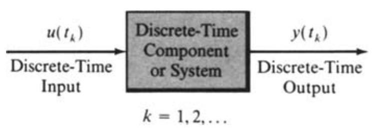


Fig. 2-7

Many digital control systems contain both continuous and discrete components. One or more devices known as *samplers*, and others known as *holds*, are usually included in such systems.

Definition 2.12: A **sampler** is a device that converts a continuous-time signal, say $u(t)$, into a discrete-time signal, denoted $u^*(t)$, consisting of a sequence of values of the signal at the instants $t_1, t_2, \ldots,$ that is, $u(t_1), u(t_2), \ldots,$ etc.

Ideal samplers are usually represented schematically by a switch, as shown in Fig. 2-8, where the switch is normally open except at the instants t_1, t_2, etc., when it is closed for an instant. The switch also may be represented as enclosed in a block, as shown in Fig. 2-9.

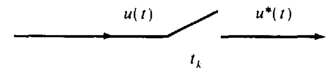


Fig. 2-8

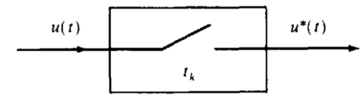


Fig. 2-9

EXAMPLE 2.8. The input signal of an ideal sampler and a few samples of the output signal are illustrated in Fig. 2-10. This type of signal is often called a *sampled-data signal.*

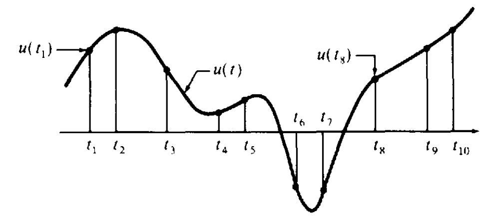


Fig. 2-10

Discrete-data signals $u(t_k)$ are often written more simply with the index k as the only argument, that is, $u(k)$, and the sequence $u(t_1), u(t_2), \ldots,$ etc., becomes $u(1), u(2), \ldots,$ etc. This notation is introduced in Chapter 3. Although sampling rates are in general nonuniform, as in Example 2.8, uniform sampling is the rule in this book, that is, $t_{k+1} - t_k \equiv T$ for all k.

Definition 2.13: A **hold**, or **data hold**, device is one that converts the discrete-time output of a sampler into a particular kind of continuous-time or analog signal.

EXAMPLE 2.9. A **zero-order hold** (or **simple hold**) is one that maintains (i.e., holds) the value of $u(t_k)$ constant until the next sampling time t_{k+1}, as shown in Fig. 2-11. Note that the output $y_{H0}(t)$ of the zero-order hold is continuous, except at the sampling times. This type of signal is called a **piecewise-continuous** signal.

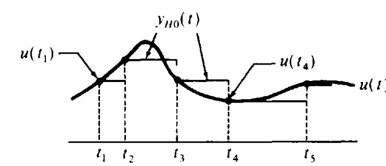


Fig. 2-11

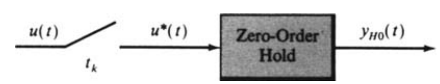


Fig. 2-12

Definition 2.14: An **analog-to-digital (A/D) converter is** a device that converts an analog or continuous signal into a discrete or digital signal.

Definition 2.15: A **digital-to-analog (D/A) converter is** a device that converts a discrete or digital signal into a continuous-time or analog signal.

EXAMPLE 2.10. The sampler in Example 2.8 (Figs. 2-9 and 2-10) is an A/D converter.

EXAMPLE 2.11. The zero-order hold in Example 2.9 (Figs. 2-11 and 2-12) is a D/A converter.

Samplers and zero-order holds are commonly used A/D and D/A converters, but they are not the only types available. Some D/A converters, in particular, are more complex.

EXAMPLE 2.12. Digital computers or microprocessors are often used to control continuous plants or processes. A/D and D/A converters are typically required in such applications, to convert signals from the plant to digital signals, and to convert the digital signal from the computer into a control signal for the analog plant. The joint operation of these elements is usually synchronized by a clock and the resulting controller is sometimes called a *digital filter*, as illustrated in Fig. 2-13.

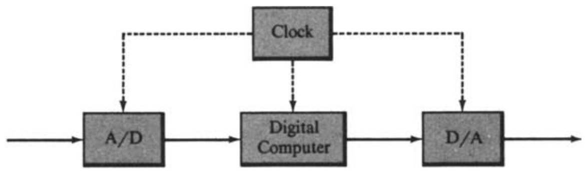


Fig. 2-13

Definition 2.16: A **computer-controlled system** includes a computer as the primary control element.

The most common computer-controlled systems have digital computers controlling analog or continuous processes. In this case, A/D and D/A converters are needed, as illustrated in Fig. 2-14.

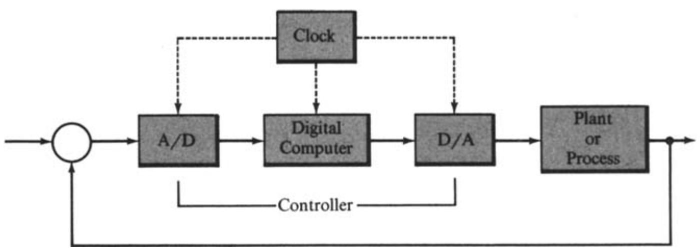


Fig. 2-14

The clock may be omitted from the diagram, as it synchronizes but is not an explicit part of signal flow in the control loop. Also, the summing junction and reference input are sometimes omitted from the diagram, because they may be implemented in the computer.

2.5 SUPPLEMENTARY TERMINOLOGY

Several other terms require definition and illustration at this time. Others are presented in subsequent chapters, as needed.

Definition 2.17: A **transducer** is a device that converts one energy form into another.

For example, one of the most common transducers in control systems applications is the *potentiometer*, which converts mechanical position into an electrical voltage (Fig. 2-15).

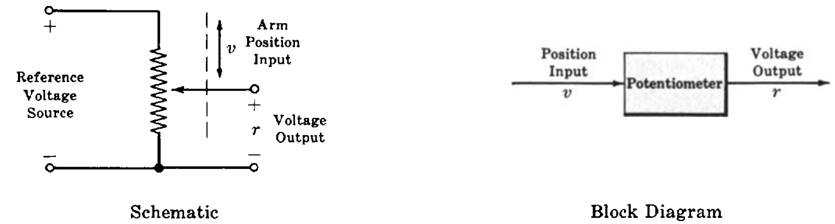


Fig. 2-15

Definition 2.18: The **command** v is an input signal, usually equal to the reference input r. But when the energy form of the command v is not the same as that of the primary feedback b, a transducer is required between the command v and the reference input r as shown in Fig. 2-16(a).

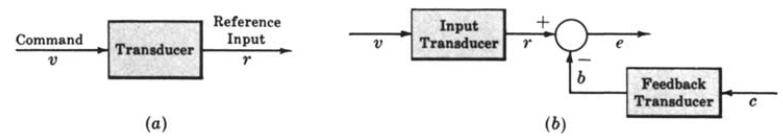


Fig. 2-16

Definition 2.19: When the feedback element consists of a transducer, and a transducer is required at the input, that part of the control system illustrated in Fig. 2-16(b) is called the **error detector**.

Definition 2.20: A **stimulus**, or **test input**, is any externally (exogenously) introduced input signal affecting the controlled output c. *Note:* The reference input r is an example of a stimulus, but it is not the only kind of stimulus.

Definition 2.21: A **disturbance** n (or **noise input**) is an undesired stimulus or input signal affecting the value of the controlled output c. It may enter the plant with u or m, as shown in the block diagram of Fig. 2-6, or at the first summing point, or via another intermediate point.

Definition 2.22: The **time response** of a system, subsystem, or element is the output as a function of time, usually following application of a prescribed input under specified operating conditions.

Definition 2.23: A **multivariable system** is one with more than one input (**multiinput, MI-**), more than one output (**multioutput, -MO**), or both (**multiinput-multioutput, MIMO**).

Definition 2.24: The term **controller** in a feedback control system is often associated with the elements of the forward path, between the actuating (error) signal e and the control variable u. But it also sometimes includes the summing point, the feedback elements, or both, and some authors use the term controller and compensator synonymously. The context should eliminate ambiguity.

The following five definitions are examples of **control laws**, or **control algorithms**.

Definition 2.25: An **on-off controller (two-position, binary controller)** has only two possible values at its output u, depending on the input e to the controller.

EXAMPLE 2.13. A binary controller may have an output $u = +1$ when the error signal is positive, that is, $e > 0$, and $u = -1$ when $e \leq 0$.

Definition 2.26: A **proportional (*P*) controller** has an output u proportional to its input e, that is, $u = K_P e$, where K_P is a proportionality constant.

Definition 2.27: A **derivative (*D*) controller** has an output proportional to the *derivative* of its input e, that is, $u = K_D\, de/dt$, where K_D is a proportionality constant.

Definition 2.28: An **integral (*I*) controller has an output *u* proportional to the integral** of its input e, that is, $u = K_I \int e(t)\, dt$, where K_I is a proportionality constant.

Definition 2.29: **PD, PI, DI, and PID controllers** are combinations of proportional (P), derivative (D), and integral (I) controllers.

EXAMPLE 2.14. The output u of a PD controller has the form:

$$u_{\mathrm{PD}} = K_P e + K_D \frac{de}{dt}$$

The output of a PID controller has the form:

$$u_{\mathrm{PID}} = K_P e + K_D \frac{de}{dt} + K_I \int e(t)\, dt$$

2.6 SERVOMECHANISMS

The specialized feedback control system called a *servomechanism* deserves special attention, due to its prevalence in industrial applications and control systems literature.

Definition 2.30: A **servomechanism** is a power-amplifying feedback control system in which the controlled variable c is mechanical position, or a time derivative of position such as velocity or acceleration.

EXAMPLE 2.15. An *automobile power-steering apparatus* is a servomechanism. The command input is the angular position of the steering wheel. A small rotational torque applied to the steering wheel is amplified hydraulically, resulting in a force adequate to modify the output, the angular position of the front wheels. The block diagram of such a system may be represented by Fig. 2-17. Negative feedback is necessary in order to return the control valve to the neutral position, reducing the torque from the hydraulic amplifier to zero when the desired wheel position has been achieved.

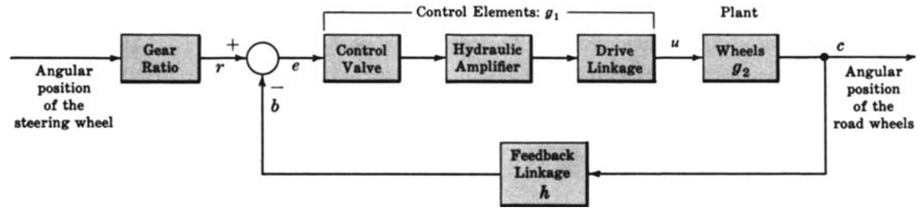


Fig. 2-17

2.7 REGULATORS

Definition 2.31: A **regulator** or **regulating system** is a feedback control system in which the reference input or command is constant for long periods of time, often for the entire time interval during which the system is operational. Such an input is often called a **setpoint**.

A regulator differs from a servomechanism in that the primary function of a regulator is usually to maintain a constant controlled output, while that of a servomechanism is most often to cause the output of the system to follow a varying input.

Solved Problems

BLOCK DIAGRAMS

2.1. Consider the following equations in which $x_1, x_2, \ldots, x_n$ are variables, and $a_1, a_2, \ldots, a_n$ are general coefficients or mathematical operators:

$$(a)\quad x_3 = a_1 x_1 + a_2 x_2 - 5$$

$$(b)\quad x_n = a_1 x_1 + a_2 x_2 + \cdots + a_{n-1} x_{n-1}$$

Draw a block diagram for each equation, identifying all blocks, inputs, and outputs.

(a) In the form the equation is written, x_3 is the output., The terms on the right-hand side of the equation are combined at a summing point, as shown in Fig. 2-18.

The $a_1 x_1$ term is represented by a single block, with x_1 as its input and $a_1 x_1$ as its output. Therefore the coefficient a_1 is put inside the block, as shown in Fig. 2-19. a_1 may represent any mathematical operation. For example, if a_1 were a constant, the block operation would be "multiply the input x_1 by the constant a_1." It is usually clear from the description or context of a problem what is meant by the symbol, operator, or description inside the block.

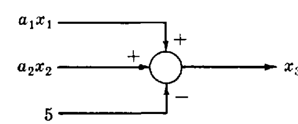


Fig. 2-18

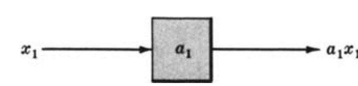


Fig. 2-19

The a_2x_2 term is represented in the same manner.

The block diagram for the entire equation is therefore shown in Fig. 2.20.

(*b*) Following the same line of reasoning as in part (*a*), the block diagram for

$$x_n = a_1x_1 + a_2x_2 + \cdots + a_{n-1}x_{n-1}$$

is shown in Fig. 2-21.

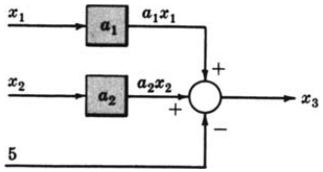


Fig. 2-20

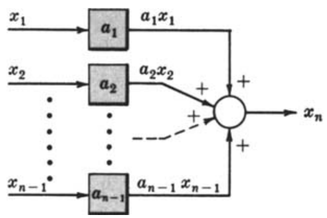
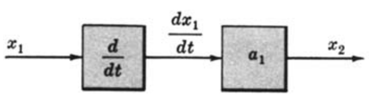


Fig. 2-21

2.2. Draw block diagrams for each of the following equations:

(*a*) $x_2 = a_1\left(\dfrac{dx_1}{dt}\right)$ (*b*) $x_3 = \dfrac{d^2x_2}{dt^2} + \dfrac{dx_1}{dt} - x_1$ (*c*) $x_4 = \int x_3\,dt$

(*a*) Two operations are specified by this equation, a_1 and differentiation d/dt. Therefore the block diagram contains two blocks, as shown in Fig. 2-22. Note the order of the blocks.



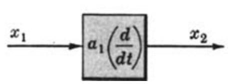


Fig. 2-23

Now, if a_1 were a constant, the a_1 block could be combined with the d/dt block, as shown in Fig. 2-23, since no confusion about the order of the blocks would result. But, if a_1 were an unknown operator, the reversal of blocks d/dt and a_1 would not necessarily result in an output equal to x_2, as shown in Fig. 2-24.

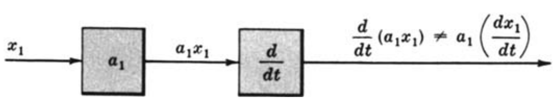


Fig. 2-24

(*b*) The + and − operations indicate the need for a summing point. The differentiation operation can be treated as in part (*a*), or by combining two first derivative operations into one second derivative operator block, giving two different block diagrams for the equation for x_3, as shown in Fig. 2-25.

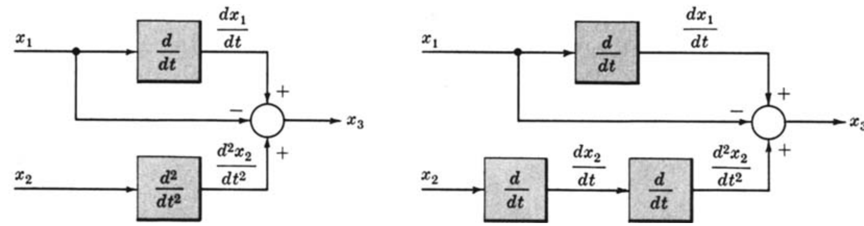


Fig. 2-25

(*c*) The integration operation can be represented in block diagram form as Fig. 2-26.

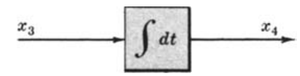


Fig. 2-26

2.3. Draw a block diagram for the pivoted, adjustable mirror mechanism of Section 1.1 with the output identified as in Problem 1.1. Assume that each 360° rotation of the screw raises or lowers the mirror k degrees. Identify all the signals and components of the control system in the diagram.

The schematic diagram of the system is repeated in Fig. 2-27 for convenience.

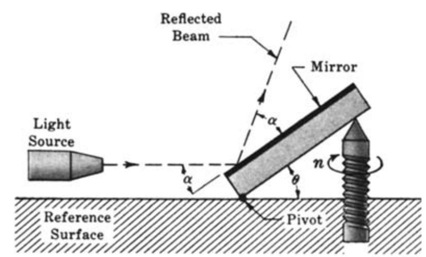


Fig. 2-27

Whereas the input was defined as θ in Problem 1.1, the specifications for this problem imply an input equal to the number of rotations of the screw. Let n be the number of rotations of the screw such that $n = 0$ when $\theta = 0°$. Therefore n and θ can be related by a block described by the constant k, since $\theta = kn$, as shown in Fig. 2-28.

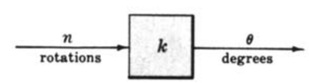


Fig. 2-28

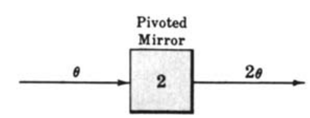


Fig. 2-29

The output of the system was determined in Problem 1.1 as $\theta + \alpha$. But since the light source is directed parallel to the reference surface, then $\alpha = \theta$. Therefore the output is equal to 2θ, and the mirror can be represented by a constant equal to 2 in a block, as shown in Fig. 2-29.

The complete open-loop block diagram is given by Fig. 2-30. For this simple example we also note that the output 2θ is equal to $2kn$ rotations of the screw. This yields the simpler block diagram of Fig. 2-31.

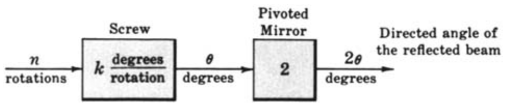


Fig. 2-30

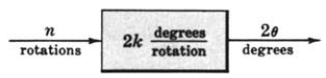


Fig. 2-31

2.4. Draw an open-loop and a closed-loop block diagram for the voltage divider network of Problem 1.11.

The open-loop equation was determined in Problem 1.11 as $v_2 = (R_2/(R_1 + R_2))v_1$, where v_1 is the input and v_2 is the output. Therefore the block is represented by $R_2/(R_1 + R_2)$ (Fig. 2-32), and clearly the operation is multiplication.

The closed-loop equation is

$$v_2 = \left(\frac{R_2}{R_1}\right)v_1 - \left(\frac{R_2}{R_1}\right)v_2 = \left(\frac{R_2}{R_1}\right)(v_1 - v_2)$$

The actuating signal is $v_1 - v_2$. The closed-loop negative feedback block diagram is easily constructed with the only block represented by R_2/R_1, as shown in Fig. 2-33.

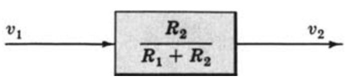


Fig. 2-32

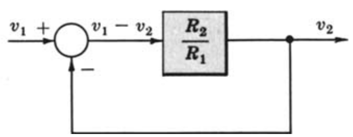


Fig. 2-33

2.5. Draw a block diagram for the electric switch of Example 1.1 (see Problems 1.9 and 1.10).

Both the input and output are binary (two-state) variables. The switch is represented by a block, and the electrical power source the switch controls is not part of the control system. One possible open-loop block diagram is given by Fig. 2-34.

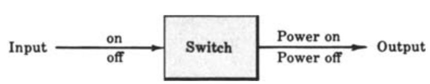


Fig. 2-34

For example, suppose the power source is an electrical current source. Then the block diagram for the switch might take the form of Fig. 2-35, where (again) the current source is not part of the control system, the input to the switch block is shown as a mechanical linkage to a simple "knife" switch, and the output is a nonzero current only when the switch is closed (on). Otherwise it is zero (off).

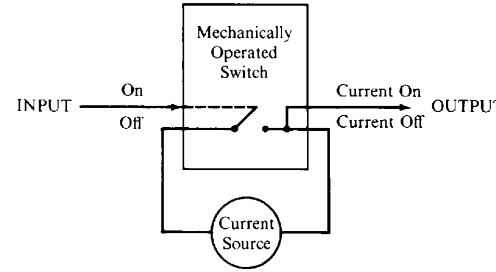


Fig. 2-35

2.6. Draw simple block diagrams for the control systems in Examples 1.2 through 1.5.

From Problem 1.10 we note that these systems are closed-loop, and from Problem 1.9 the actuating signal (control action) for the system in each example is equal to the input minus the output. Therefore negative feedback exists in each system.

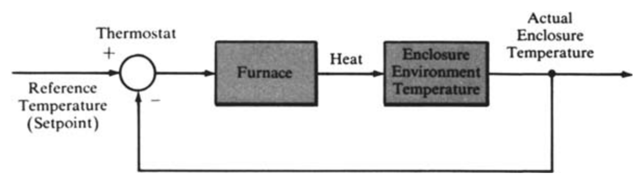
For the thermostatically controlled furnace of Example 1.2, the thermostat can be chosen as the summing point, since this is the device that determines whether or not the furnace is turned on. The enclosure environment (outside) temperature may be treated as a **noise** input acting directly on the enclosure.

The eyes may be represented by a summing point in both the human pointing system of Example 1.3 and the driver-automobile system of Example 1.5. The eyes perform the function of monitoring the input and output.

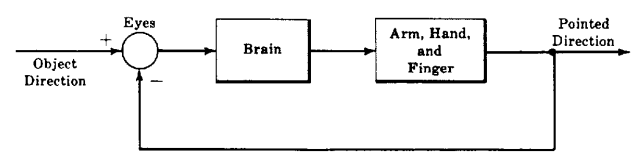
For the perspiration system of Example 1.4, the summing point is not so easily defined. For the sake of simplicity let us call it the nervous system.

The block diagrams are easily constructed as shown below from the information given above and the list of components, inputs, and outputs given in the examples.

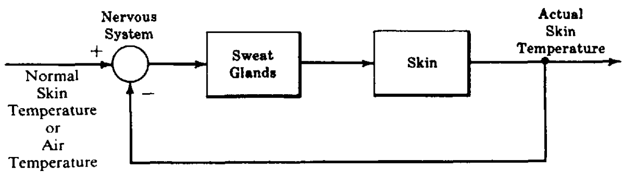
The arrows between components in the block diagrams of the biological systems in Examples 1.3 through 1.5 represent electrical, chemical, or mechanical signals controlled by the central nervous system.



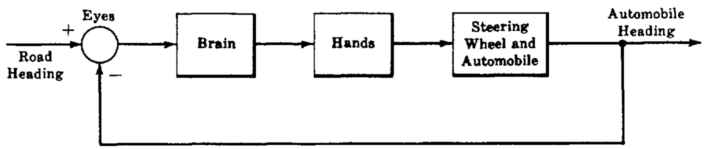
Example 1.2



Example 1.3



Example 1.4



Example 1.5

BLOCK DIAGRAMS OF FEEDBACK CONTROL SYSTEMS

2.7. Draw a block diagram for the water-filling system described in Problem 1.15. Which component or components comprise the plant? The controller? The feedback?

The container is the plant because the water level *of the container* is being controlled (see Definition 2.1). The stopper valve may be chosen as the control element; and the ball-float, cord, and associated linkage as the feedback elements. The block diagram is given in Fig. 2-36.

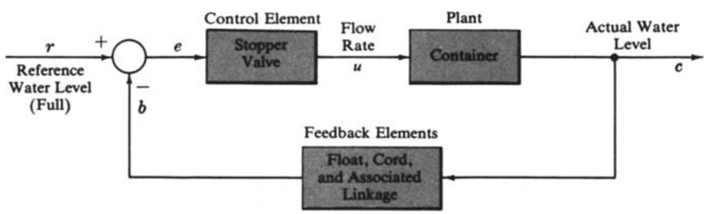


Fig. 2-36

The feedback is negative because the water flow rate to the container must decrease as the water level rises in the container.

2.8. Draw a simple block diagram for the feedback control system of Examples 1.7 and 1.8, the airplane with an autopilot.

The plant for this system is the airplane, including its control surfaces and navigational instruments. The controller is the autopilot mechanism, and the summing point is the comparison device. The feedback linkage may be simply represented by an arrow from the output to the summing point, as this linkage is not well defined in Example 1.8.

The autopilot provides control signals to operate the control surfaces (rudder, flaps, etc.). These signals may be denoted $u_1, u_2, \ldots$.

The simplest block diagram for this feedback system is given in Fig. 2-37.

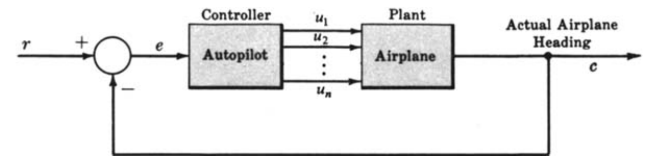


Fig. 2-37

SERVOMECHANISMS

2.9. Draw a schematic and a block diagram from the following description of a *position servomechanism* whose function is to open and close a water valve.

At the input of the system there is a rotating-type potentiometer connected across a battery voltage source. Its movable (third) terminal is calibrated in terms of angular position (in radians). This output terminal is electrically connected to one terminal of a voltage amplifier called a *servoamplifier*. The servoamplifier supplies enough output power to operate an electric motor called a *servomotor*. The servomotor is mechanically linked with the water valve in a manner which permits the valve to be opened or closed by the motor.

Assume the loading effect of the valve on the motor is negligible; that is, it does not "resist" the motor. A 360° rotation of the motor shaft completely opens the valve. In addition, the movable terminal of a second potentiometer connected in parallel at its fixed terminals with the input potentiometer is mechanically connected to the motor shaft. It is electrically connected to the remaining input terminal of the servoamplifier. The potentiometer ratios are set so that they are equal when the valve is closed.

When a command is given to open the valve, the servomotor rotates in the appropriate direction. As the valve opens, the second potentiometer, called the *feedback potentiometer*, rotates in the same direction as the input potentiometer. It stops when the potentiometer ratios are again equal.

A schematic diagram (Fig. 2-38) is easily drawn from the preceding description. Mechanical connections are shown as dashed lines.

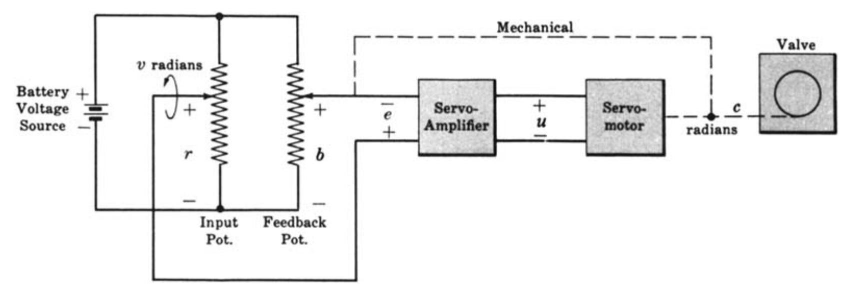


Fig. 2-38

The block diagram for this system (Fig. 2-39) is easily drawn from the schematic diagram.

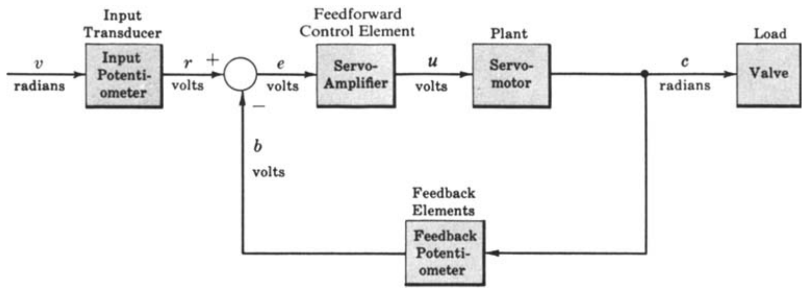


Fig. 2-39

2.10. Draw a block diagram for the elementary speed control system (velocity servomechanism) given in Fig. 2-40.

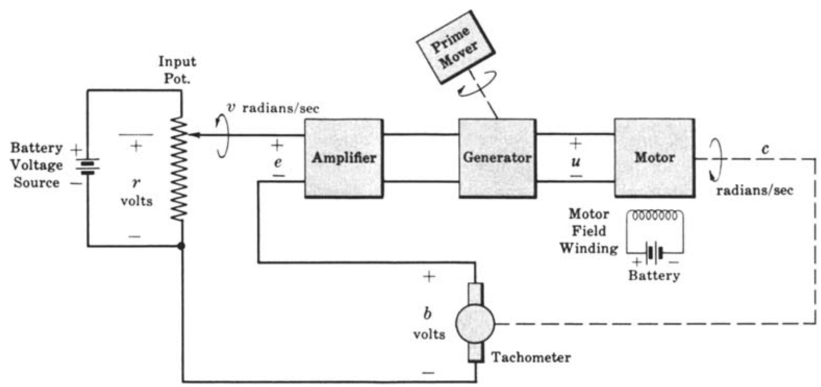


Fig. 2-40

The potentiometer is a rotating-type, calibrated in radians per seconds, and the prime-mover speed, motor field winding, and input potentiometer currents are constant functions of time. No load is attached to the motor shaft.

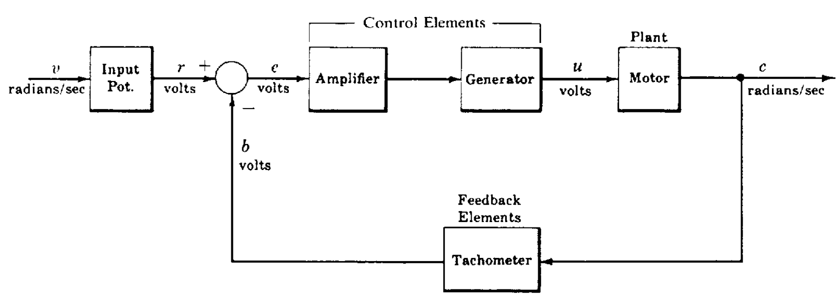


Fig. 2-41

The battery voltage sources for both the input potentiometer and motor field winding, and the prime-mover source for the generator are not part of the control loop of this servomechanism. The output of each of these sources is a constant function of time, and can be accounted for in the mathematical description of the input potentiometer, generator, and motor, respectively. Therefore the block diagram for this system is given in Fig. 2-41.

MISCELLANEOUS PROBLEMS

2.11. Draw a block diagram for the photocell light switch system described in Problem 1.16. The light intensity in the room must be maintained at a level greater than or equal to a prespecified level.

One way of describing this system is with two inputs, one input chosen as minimum reference room-light intensity r_1, and the second as room sunlight intensity r_2. The output c is actual room-light intensity.

The room is the plant. The manipulated variable (control signal) is the amount of light supplied to the room from both the lamp and the sun. The photocell and the lamp are the control elements because they control room-light intensity. Assume the minimum reference room-light intensity r_1 is equal to the intensity of room-light supplied by the lighted lamp alone. A block diagram for this system is given in Fig. 2-42.

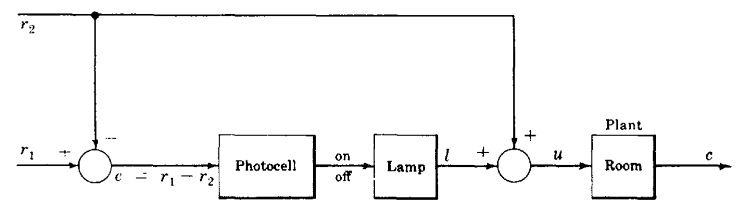


Fig. 2-42

The system is clearly open-loop. The actuating signal e is independent of the output c, and is equal to the difference between the two inputs: $r_1 - r_2$. When $e \leq 0$, $l = 0$ (the light is off). When $e > 0$, $l = r_1$ (the light is on).

2.12. Draw a block diagram for the closed-loop traffic signal system described in Problem 1.13.

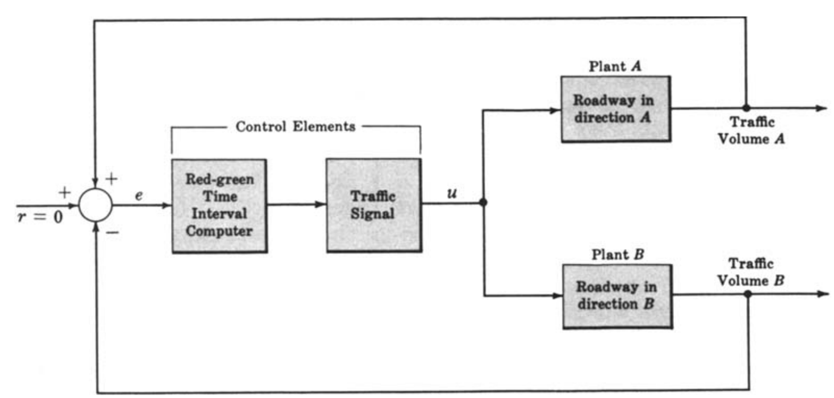


Fig. 2-43

This system has two outputs, the volume of traffic passing the intersection in one direction (the A direction), and the volume passing the intersection in the other direction (the B direction). The input is the command for equal traffic volumes in directions A and B; that is, the input is zero volume difference.

Suppose we call the mechanism for computing the appropriate red and green timing intervals the Red-Green Time Interval Computer. This device, in addition to the traffic signal, makes up the control elements. The plants are the roadway in direction A and the roadway in direction B. The block diagram of this traffic *regulator* is given in Fig. 2-43.

2.13. Draw a block diagram illustrating the economic Law of Supply and Demand, as described in Problem 1.12.

The block diagram is given by Fig. 2-44.

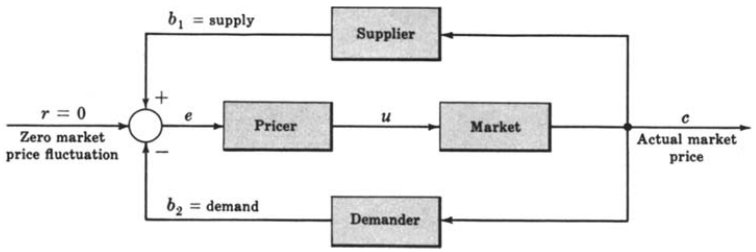


Fig. 2-44

2.14. The following very simplified model of the biological mechanism regulating human arterial blood pressure is an example of a feedback control system.

A well-regulated pressure must be maintained in the blood vessels (arteries, arterioles, and capillaries) supplying the tissues, so that blood flow is adequately maintained. This pressure is usually measured in the aorta (an artery) and is called the *blood pressure* p. It is not constant and normally has a range of 70–130 mm of mercury (mm Hg) in adults. Let us assume that p is equal to 100 mm Hg (on the average) in a normal individual.

A fundamental model of circulatory physiology is the following equation for arterial blood pressure:

$$p = Q\rho$$

where Q is the *cardiac output*, or the volume flow rate of blood from the heart to the aorta, and ρ is the *peripheral resistance* offered to blood flow by the arterioles. Under normal conditions, ρ is approximately inversely proportional to the fourth power of the diameter d of the vessels (arterioles).

Now d is believed to be controlled by the *vasomotor center* (VMC) of the brain, with increased activity of the VMC decreasing d, and vice versa. Although several factors affect VMC activity, the *baroreceptor cells* of the *arterial sinus* are believed to be the most important. Baroreceptor activity *inhibits* the VMC, and therefore functions in a negative feedback mode. According to this theory, if ρ increases, the baroreceptors send signals along the vagus and glossopharyngeal nerves to the VMC, decreasing its activity. This results in an increase in arteriole diameter d, a decrease in peripheral resistance ρ, and (assuming constant cardiac output Q) a corresponding drop in blood pressure p. This feedback network probably regulates, at least in part, blood pressure in the aorta.

Draw a block diagram of this feedback control system, identifying all signals and components.

Let the aorta be the plant, represented by Q (cardiac output); the VMC and arterioles may be chosen as the controller; the baroreceptors are the feedback elements. The input p_0 is the average normal (reference) blood pressure, 100 mm Hg. The output p is the actual blood pressure. Since $\rho = k(1/d)^4$, where k is a proportionality constant, the arterioles can be represented in the block by $k(\cdot)^4$. The block diagram is given in Fig. 2-45.

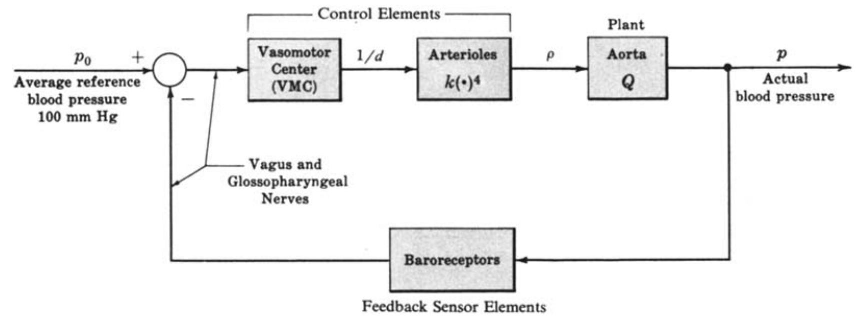


Fig. 2-45

2.15. The *thyroid* gland, an endocrine (internally secreting) gland located in the neck in the human, secretes *thyroxine* into the bloodstream. The bloodstream is the signal transmission system for most of the endocrine glands, just as conductive wires are the transmission system for the flow of electrical current, or pipes and tubes may be the transmission system for hydrodynamic fluid flow.

Like most human physiological processes, the production of thyroxine by the thyroid gland is automatically controlled. The amount of thyroxine in the bloodstream is regulated in part by a hormone secreted by the *anterior pituitary*, an endocrine gland suspended from the base of the brain. This "control" hormone is appropriately called *thyroid stimulating hormone* (TSH). In a simplified view of this control system, when the level of thyroxine in the circulatory system is higher than that required by the organism, TSH secretion is inhibited (reduced), causing a reduction in the activity of the thyroid. Hence less thyroxine is released by the thyroid.

Draw a block diagram of the simplified system described, identifying all components and signals.

Let the plant be the thyroid gland, with the controlled variable the level of thyroxine in the bloodstream. The pituitary gland is the controller, and the manipulated variable is the amount of TSH it secretes. The block diagram is given in Fig. 2-46.

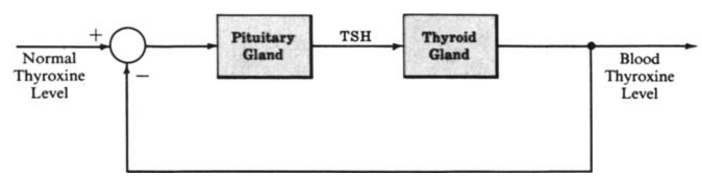


Fig. 2-46

We reemphasize that this is a very simplified view of this biological control system, as was that in the previous problem.

2.16. What type of **controller** is included in the more realistic thermostatically controlled heating system described in Example 1.14?

The thermostat-furnace controller has a binary output: furnace (full) on, or furnace off. Therefore it is an on-off controller. But it is not as simple as the sign-sensing binary controller of Example 2.13. The thermostat switch turns the furnace on when room temperature falls to 2° below its setpoint of 68°F (22°C), and turns it off when it rises to 2° above its setpoint.

Graphically, the characteristic curve of such a controller has the form given in Fig. 2-47.

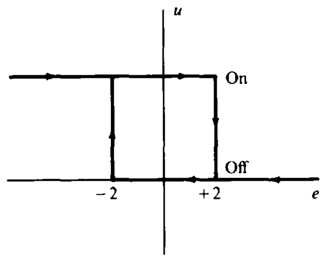


Fig. 2-47

This is called a **hysteresis** characteristic curve, because its output has a "memory"; that is, the switching points depend on whether the input e is rising or falling when the controller switches states from on to off, or off to on.

2.17. Sketch the error, control, and controlled output signals as functions of time and discuss how the on-off controller of Problem 2.16 maintains the average room temperature specified by the setpoint (68°F) of the thermostat?

The signals $e(t)$, $u(t)$, and $c(t)$ typically have the form shown in Fig. 2-48, assuming the temperature was colder than 66°F at the start.

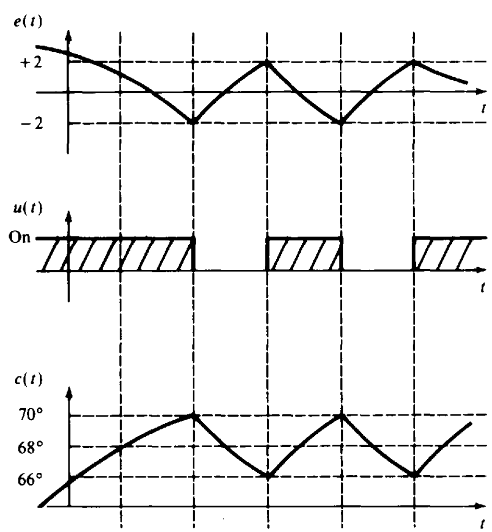


Fig. 2-48

The room temperature $c(t)$ is constantly varying. In each switching interval of the controller, it rises at an approximately constant rate, from 66° to 70°, or falls at an approximately constant rate, from 70° to 66°. The average temperature of the room is the mean value of this function $c(t)$, which is approximately 68°F.

2.18. What major advantage does a computer-controlled system have over an analog system?

The controller (control law) in a computer-controlled system is typically implemented by means of software, rather than hardware. Therefore the class of control laws that can be implemented conveniently is substantially increased.

Supplementary Problems

2.19. The schematic diagram of a semiconductor voltage amplifier called an *emitter follower* is given in Fig. 2-49.
An equivalent circuit for this amplifier is shown in Fig. 2-50, where r_p is the internal resistance of, and μ is a parameter of the particular semiconductor. Draw both an open-loop and a closed-loop block diagram for this circuit with an input v_{in} and an output v_{out}.

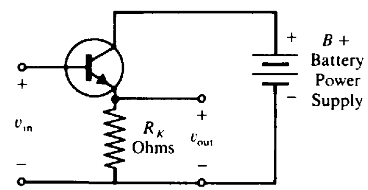


Fig. 2-49

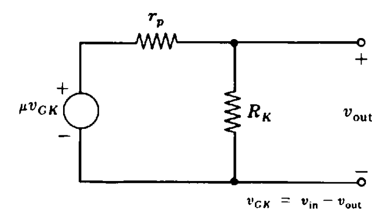


Fig. 2-50

2.20. Draw a block diagram for the human walking system of Problem 1.14.

2.21. Draw a block diagram for the human reaching system described in Problem 1.4.

2.22. Draw a block diagram for the automatic temperature-regulated oven of Problem 1.21.

2.23. Draw a block diagram for the closed-loop automatic toaster of Problem 1.17.

2.24. State the common dimensional units for the input and output of the following transducers: (a) accelerometer, (b) generator of electricity, (c) thermistor (temperature-sensitive resistor), (d) thermocouple.

2.25. Which systems in Problems 2.1 through 2.8 and 2.11 through 2.21 are servomechanisms?

2.26. The endocrine gland known as the *adrenal cortex* is located on top of each kidney (two parts). It secretes several hormones, one of which is *cortisol*. Cortisol plays an important part in regulating the metabolism of carbohydrates, proteins, and fats, particularly in times of stress. Cortisol production is controlled by adrenocorticotrophic hormone (ACTH) from the anterior pituitary gland. High blood cortisol inhibits ACTH production. Draw a block diagram of this simplified feedback control system.

2.27. Draw block diagrams for each of the following elements, first with voltage v as input and current i as output, and then vice versa: (a) resistance R, (b) capacitance C, (c) inductance L.

2.28. Draw block diagrams for each of the following mechanical systems, where force is the input and position the output: (a) a dashpot, (b) a spring, (c) a mass, (d) a mass, spring, and dashpot connected in series and fastened at one end (mass position is the output).

2.29. Draw a block diagram of a (a) parallel, (b) series R-L-C network.

2.30. Which systems described in the problems of this chapter are regulators?

2.31. What type of sampled-data system described in this chapter might be used in implementing a device or algorithm for approximating the integral of a continuous function $u(t)$, using the well-known rectangular rule, or rectangular integration technique?

2.32. Draw a simple block diagram of a computer-controlled system in which a digital computer is used to control an analog plant or process, with the summing point and reference input implemented in software in the computer.

2.33. What type of controller is the stopper valve of the water-filling system of Problem 2.7?

2.34. What types of controllers are included in: (a) each of the servomechanisms of Problems 2.9 and 2.10, (b) the traffic regulator of Problem 2.12?

Answers to Supplementary Problems

2.19. The equivalent circuit for the emitter follower has the same form as the voltage divider network of Problem 1.11. Therefore the open-loop equation for the output is

$$v_{\text{out}} = \frac{\mu R_K}{r_p + R_K}(v_{\text{in}} - v_{\text{out}}) = \left(\frac{\mu R_K}{r_p + (1+\mu) R_K}\right) v_{\text{in}}$$

and the open-loop block diagram is given in fig. 2-51.

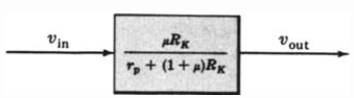


Fig. 2-51

The closed-loop output equation is simply

$$v_{\text{out}} = \frac{\mu R_K}{r_p + R_K}(v_{\text{in}} - v_{\text{out}})$$

and the closed-loop block diagram is given in Fig. 2-52.

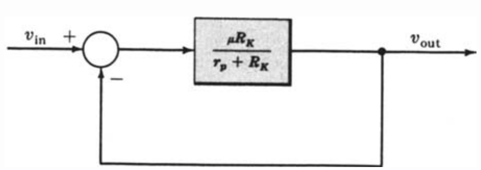
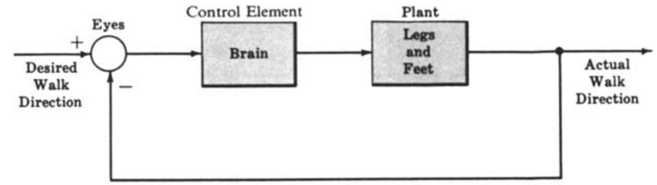
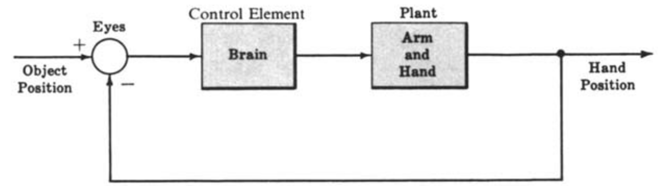


Fig. 2-52

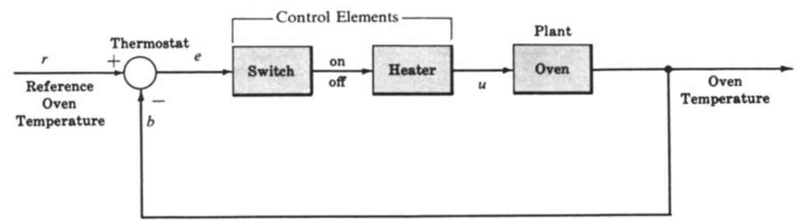
2.20.



2.21.

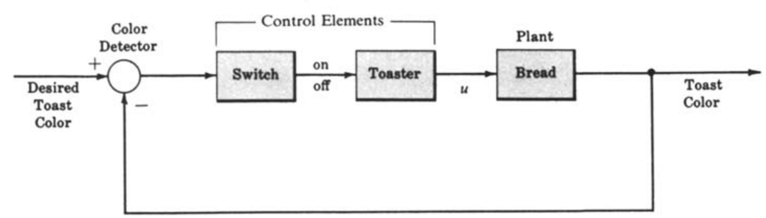


2.22.



When $e > 0$ $(r > b)$, the switch turns the heater on. When $e \leq 0$, the heater is turned off.

2.23.



2.24. (a) The input to an accelerometer is acceleration. The output is displacement of a mass, voltage, or another quantity proportional to acceleration.

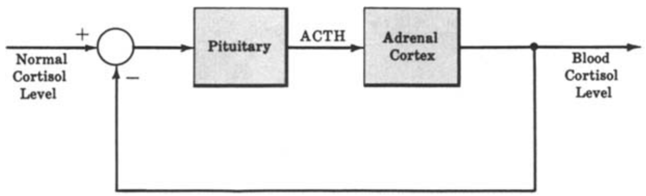
(b) See Problem 1.2.

(c) The input to a thermistor is temperature. The output is an electrical quantity measured in ohms, volts, or amperes.

(d) The input to a thermocouple is a temperature difference. The output is a voltage.

2.25. The following problems describe servomechanisms: Examples 1.3 and 1.5 in Problem 2.6, and Problems 2.7, 2.8, 2.17, and 2.21.

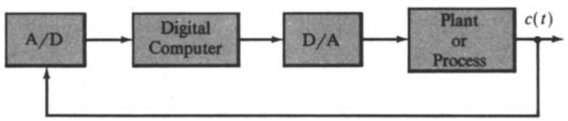
2.26.



2.30. The systems of Examples 1.2 and 1.4 in Problem 2.6, and the systems of Problems 2.7, 2.8, 2.12, 2.13, 2.14, 2.15, 2.22, 2.23, and 2.26 are regulators.

2.31. The sampler and zero-order hold device of Example 2.9 performs part of the process required for rectangular integration. For this simplest numerical integration algorithm, the "area under the curve" (i.e., the integral) is approximated by small rectangles of height $u(t_k)$ and width $t_{k+1} - t_k$. This result could be obtained by first multiplying the output of the hold device $u^*(t)$ by the width of the interval $t_{k+1} - t_k$, when $u^*(t)$ is on the interval between t_k and t_{k+1}. The sum of these products is the desired result.

2.32.



2.33. If the stopper valve is a simple one of the type that can be only fully open or fully closed, it is an *on-off controller*. But if it is that type that closes gradually as the tank fills, it is a *proportional controller*.

Chapter 3

Differential Equations, Difference Equations, and Linear Systems

3.1 SYSTEM EQUATIONS

A property common to all basic laws of physics is that certain fundamental quantities can be defined by numerical values. The physical laws define relationships between these fundamental quantities and are usually represented by equations.

EXAMPLE 3.1. The scalar version of Newton's second law states that, if a force of magnitude f is applied to a mass of M units, the acceleration a of the mass is related to f by the equation $f = Ma$.

EXAMPLE 3.2. Ohm's law states that, if a voltage of magnitude v is applied across a resistor of R units, the current i through the resistor is related to v by the equation $v = Ri$.

Many nonphysical laws can also be represented by equations.

EXAMPLE 3.3. The compound interest law states that, if an amount $P(0)$ is deposited for n equal periods of time at an interest rate I for each time period, the amount will grow to a value of $P(n) = P(0)(1 + I)^n$.

3.2 DIFFERENTIAL EQUATIONS AND DIFFERENCE EQUATIONS

Two classes of equations with broad application in the description of systems are differential equations and difference equations.

Definition 3.1: A **differential equation** is any algebraic or transcendental equality which involves either differentials or derivatives.

Differential equations are useful for relating rates of change of variables and other parameters.

EXAMPLE 3.4. Newton's second law (Example 3.1) can be written alternatively as a relationship between force f, mass M, and the rate of change of the velocity v of the mass with respect to time t, that is, $f = M(dv/dt)$.

EXAMPLE 3.5. Ohm's law (Example 3.2) can be written alternatively as a relationship between voltage v, resistance R, and the time rate of passage of charge through the resistor, that is, $v = R(dq/dt)$.

EXAMPLE 3.6. The diffusion equation in one dimension describes the relationship between the time rate of change of a quantity T in an object (e.g., heat concentration in an iron bar) and the positional rate of change of T: $\partial T/\partial x = k(\partial T/\partial t)$, where k is a proportionality constant, x is a position variable, and t is time.

Definition 3.2: A **difference equation** is an algebraic or transcendental equality which involves more than one value of the dependent variable(s) corresponding to more than one value of at least one of the independent variable(s). The dependent variables do not involve either differentials or derivatives.

Difference equations are useful for relating the evolution of variables (or parameters) from one discrete instant of time (or other independent variable) to another.

EXAMPLE 3.7. The compound interest law of Example 3.3 can be written alternatively as a difference equation relationship between $P(k)$, the amount of money after k periods of time, and $P(k+1)$, the amount of money after $k+1$ periods of time, that is, $P(k+1) = (1+I)P(k)$.

3.3 PARTIAL AND ORDINARY DIFFERENTIAL EQUATIONS

Definition 3.3: A **partial differential equation** is an equality involving one or more dependent and two or more independent variables, together with partial derivatives of the dependent with respect to the independent variables.

Definition 3.4: An **ordinary (total) differential equation** is an equality involving one or more dependent variables, one independent variable, and one or more derivatives of the dependent variables with respect to the independent variable.

EXAMPLE 3.8. The diffusion equation $\partial T/\partial x = k(\partial T/\partial t)$ is a partial differential equation. $T = T(x,t)$ is the dependent variable, which represents the concentration of some quantity at some position and some time in the object. The independent variable x defines the position in the object, and the independent variable t defines the time.

EXAMPLE 3.9. Newton's second law (Example 3.4) is an ordinary differential equation: $f = M(dv/dt)$. The velocity $v = v(t)$ and the force $f = f(t)$ are dependent variables, and the time t is the independent variable.

EXAMPLE 3.10. Ohm's law (Example 3.5) is an ordinary differential equation: $v = R(dq/dt)$. The charge $q = q(t)$ and the voltage $v = v(t)$ are dependent variables, and the time t is the independent variable.

EXAMPLE 3.11. A differential equation of the form:

$$a_n \frac{d^n y}{dt^n} + a_{n-1}\frac{d^{n-1}y}{dt^{n-1}} + \cdots + a_1 \frac{dy}{dt} + a_0 y = u(t)$$

or, more compactly,

$$\sum_{i=0}^{n} a_i \frac{d^i y(t)}{dt^i} = u(t) \tag{3.1}$$

where $a_0, a_1, \ldots, a_n$ are constants, is an ordinary differential equation. $y(t)$ and $u(t)$ are dependent variables, and t is the independent variable.

3.4 TIME VARIABILITY AND TIME INVARIANCE

In the remainder of this chapter, *time* is the only independent variable, unless otherwise specified. This variable is normally designated t, except that in difference equations the discrete variable k is often used, as an abbreviation for the time instant t_k (see Example 1.11 and Section 2.5); that is, $y(k)$ is used instead of $y(t_k)$, etc.

A **term** of a differential or difference equation consists of products and/or quotients of explicit functions of the independent variable, the dependent variables, and, for differential equations, derivatives of the dependent variables.

In the definitions of this and the next section, the term *equation* refers to either a differential equation or a difference equation.

Definition 3.5: A **time-variable equation** is an equation in which one or more terms depend *explicitly* on the independent variable time.

Definition 3.6: A **time-invariant equation** is an equation in which none of the terms depends *explicitly* on the independent variable time.

EXAMPLE 3.12. The difference equation $ky(k+2) + y(k) = u(k)$, where u and y are dependent variables, is time-variable because the term $ky(k+2)$ depends explicitly on the coefficient k, which represents the time t_k.

EXAMPLE 3.13. Any differential equation of the form:

$$\sum_{i=0}^{n} a_i \frac{d^i y}{dt^i} = \sum_{i=0}^{m} b_i \frac{d^i u}{dt^i} \tag{3.2}$$

where the coefficients $a_0, a_1, \ldots, a_n, b_0, b_1, \ldots, b_m$ are constants, is *time-invariant*. The equation depends *implicitly* on t, via the dependent variables u and y and their derivatives.

3.5 LINEAR AND NONLINEAR DIFFERENTIAL AND DIFFERENCE EQUATIONS

Definition 3.7: A **linear term** is one which is first degree in the dependent variables and their derivatives.

Definition 3.8: A **linear equation** is an equation consisting of a sum of linear terms. All others are **nonlinear equations**.

If any term of a differential equation contains higher powers, products, or transcendental functions of the dependent variables, it is nonlinear. Such terms include $(dy/dt)^3$, $u(dy/dt)$, and $\sin u$, respectively. For example, $(5/\cos t)(d^2y/dt^2)$ is a term of first degree in the dependent variable y, and $2uy^3(dy/dt)$ is a term of fifth degree in the dependent variables u and y.

EXAMPLE 3.14. The ordinary differential equations $(dy/dt)^2 + y = 0$ and $d^2y/dt^2 + \cos y = 0$ are nonlinear because $(dy/dt)^2$ is second degree in the first equation, and $\cos y$ in the second equation is *not* first degree, which is true of all transcendental functions.

EXAMPLE 3.15. The difference equation $y(k+2) + u(k+1)y(k+1) + y(k) = u(k)$, in which u and y are dependent variables, is a nonlinear difference equation because $u(k+1)y(k+1)$ is second degree in u and y. This type of nonlinear equation is sometimes called *bilinear* in u and y.

EXAMPLE 3.16. Any difference equation

$$\sum_{i=0}^{n} a_i(k)\, y(k+i) = \sum_{i=0}^{n} b_i(k)\, u(k+i) \tag{3.3}$$

in which the coefficients $a_i(k)$ and $b_i(k)$ depend only upon the independent variable k, is a linear difference equation.

EXAMPLE 3.17. Any ordinary differential equation

$$\sum_{i=0}^{n} a_i(t) \frac{d^i y}{dt^i} = \sum_{i=0}^{m} b_i(t) \frac{d^i u}{dt^i} \tag{3.4}$$

where the coefficients $a_i(t)$ and $b_i(t)$ depend only upon the independent variable t, is a linear differential equation.

3.6 THE DIFFERENTIAL OPERATOR D AND THE CHARACTERISTIC EQUATION

Consider the nth-order linear constant-coefficient differential equation

$$\frac{d^n y}{dt^n} + a_{n-1}\frac{d^{n-1}y}{dt^{n-1}} + \cdots + a_1 \frac{dy}{dt} + a_0 y = u \tag{3.5}$$

It is convenient to define a **differential operator**

$$D \equiv \frac{d}{dt}$$

and more generally an **nth-order differential operator**

$$D^n \equiv \frac{d^n}{dt^n}$$

The differential equation can now be written as

$$D^n y + a_{n-1}D^{n-1}y + \cdots + a_1 Dy + a_0 y = u$$

or

$$\left(D^n + a_{n-1}D^{n-1} + \cdots + a_1 D + a_0\right)y = u$$

Definition 3.9: The polynomial in D:

$$D^n + a_{n-1}D^{n-1} + \cdots + a_1 D + a_0 \tag{3.6}$$

is called the **characteristic polynomial**.

Definition 3.10: The equation

$$D^n + a_{n-1}D^{n-1} + \cdots + a_1 D + a_0 = 0 \tag{3.7}$$

is called the **characteristic equation**.

The fundamental theorem of algebra states that the characteristic equation has exactly n solutions $D = D_1, D = D_2, \ldots, D = D_n$. These n solutions (also called **roots**) are not necessarily distinct.

EXAMPLE 3.18. Consider the differential equation

$$\frac{d^2y}{dt^2} + 3\frac{dy}{dt} + 2y = u$$

The characteristic polynomial is $D^2 + 3D + 2$. The characteristic equation is $D^2 + 3D + 2 = 0$, which has the two distinct roots: $D = -1$ and $D = -2$.

3.7 LINEAR INDEPENDENCE AND FUNDAMENTAL SETS

Definition 3.11: A set of n functions of time $f_1(t), f_2(t), \ldots, f_n(t)$ is called **linearly independent** if the only set of constants $c_1, c_2, \ldots, c_n$ for which

$$c_1 f_1(t) + c_2 f_2(t) + \cdots + c_n f_n(t) = 0$$

for all t are the constants $c_1 = c_2 = \cdots = c_n = 0$.

EXAMPLE 3.19. The functions t and t^2 are linearly independent functions since

$$c_1 t + c_2 t^2 = t(c_1 + c_2 t) = 0$$

implies that $c_1/c_2 = -t$. There are *no constants* that satisfy this relationship.

A *homogeneous* nth-order linear differential equation of the form:

$$\sum_{i=0}^{n} a_i \frac{d^i y}{dt^i} = 0$$

has at least one set of n linearly independent solutions.

Definition 3.12: Any set of n *linearly independent* solutions of a homogeneous nth-order linear differential equation is called a **fundamental set**.

There is no unique fundamental set. From a given fundamental set other fundamental sets can be generated by the following technique. Suppose that $y_1(t), y_2(t), \ldots, y_n(t)$ is a fundamental set for an nth-order linear differential equation. Then a set of n functions $z_1(t), z_2(t), \ldots, z_n(t)$ can be formed:

$$z_1(t) = \sum_{i=1}^{n} a_{1i} y_i(t),\ z_2(t) = \sum_{i=1}^{n} a_{2i} y_i(t), \ldots, z_n(t) = \sum_{i=1}^{n} a_{ni} y_i(t) \tag{3.8}$$

where the a_{ji} are a set of n^2 constants. Each $z_i(t)$ is a solution of the differential equation. This set of n solutions is a *fundamental set* if the determinant

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \neq 0$$

EXAMPLE 3.20. The equation for simple harmonic motion, $d^2y/dt^2 + \omega^2 y = 0$, has as a fundamental set

$$y_1 = \sin \omega t \qquad y_2 = \cos \omega t$$

A second fundamental set is*

$$z_1 = \cos \omega t + j \sin \omega t = e^{j\omega t} \qquad z_2 = \cos \omega t - j \sin \omega t = e^{-j\omega t}$$

Distinct Roots

If the characteristic equation

$$\sum_{i=0}^{n} a_i D^i = 0$$

has distinct roots $D_1, D_2, \ldots, D_n$, then a fundamental set for the homogeneous equation

$$\sum_{i=0}^{n} a_i \frac{d^i y}{dt^i} = 0$$

is the set of functions $y_1 = e^{D_1 t}, y_2 = e^{D_2 t}, \ldots, y_n = e^{D_n t}$.

EXAMPLE 3.21. The differential equation

$$\frac{d^2y}{dt^2} + 3\frac{dy}{dt} + 2y = 0$$

has the characteristic equation $D^2 + 3D + 2 = 0$ whose roots are $D = D_1 = -1$ and $D = D_2 = -2$. A fundamental set for this equation is $y_1 = e^{-t}$ and $y_2 = e^{-2t}$.

Repeated Roots

If the characteristic equation has repeated roots, then for each root D_i of multiplicity n_i (i.e., n_i roots equal to D_i) there are n_i elements of the fundamental set $e^{D_i t}, te^{D_i t}, \ldots, t^{n_i - 1}e^{D_i t}$.

EXAMPLE 3.22. The equation

$$\frac{d^2y}{dt^2} + 2\frac{dy}{dt} + y = 0$$

*The *complex exponential function* e^w, where $w = u + jv$ for real u and v, and $j = \sqrt{-1}$, is defined in complex variable theory by $e^w \equiv e^u(\cos v + j \sin v)$. Therefore $e^{\pm j\omega t} = \cos \omega t \pm j \sin \omega t$.

with characteristic equation $D^2 + 2D + 1 = 0$, has the repeated root $D = -1$, and a fundamental set consisting of e^{-t} and te^{-t}.

3.8 SOLUTION OF LINEAR CONSTANT-COEFFICIENT ORDINARY DIFFERENTIAL EQUATIONS

Consider the class of differential equations of the form:

$$\sum_{i=0}^{n} a_i \frac{d^i y}{dt^i} = \sum_{i=0}^{m} b_i \frac{d^i u}{dt^i} \tag{3.9}$$

where the coefficients a_i and b_i are constant, $u = u(t)$ *(the input) is a known time function*, and $y = y(t)$ *(the output) is the unknown solution of the equation.* If this equation describes a physical system, then generally $m \le n$, and n is called the **order** of the differential equation. To completely specify the problem so that *a unique solution* $y(t)$ can be obtained, two additional items must be specified: (1) the interval of time over which a solution is desired and (2) a set of n *initial conditions* for $y(t)$ and its first $n-1$ derivatives. The time interval for the class of problems considered is defined by $0 \le t < +\infty$. This interval is used in the remainder of this book unless otherwise specified. The set of initial conditions is

$$y(0), \left.\frac{dy}{dt}\right|_{t=0}, \ldots, \left.\frac{d^{n-1}y}{dt^{n-1}}\right|_{t=0} \tag{3.10}$$

A problem defined over this interval and with these initial conditions is called an **initial value problem**.

The solution of a differential equation of this class can be divided into two parts, a *free response* and a *forced response*. The sum of these two responses constitutes the *total response*, or solution $y(t)$, of the equation.

3.9 THE FREE RESPONSE

The **free response** of a differential equation is the solution of the differential equation when the input $u(t)$ is identically zero.

If the input $u(t)$ is identically zero, then the differential equation has the form:

$$\sum_{i=0}^{n} a_i \frac{d^i y}{dt^i} = 0 \tag{3.11}$$

The solution $y(t)$ of such an equation depends only on the n initial conditions in Equation (3.10).

EXAMPLE 3.23. The solution of the homogeneous first-order differential equation $dy/dt + y = 0$ with initial condition $y(0) = c$, is $y(t) = ce^{-t}$. This can be verified by direct substitution. ce^{-t} is the free response of any differential equation of the form $dy/dt + y = u$ with the initial condition $y(0) = c$.

The *free response* of a differential equation can always be written as a linear combination of the elements of a *fundamental set*. That is, if $y_1(t), y_2(t), \ldots, y_n(t)$ is a fundamental set, then *any* free response $y_a(t)$ of the differential equation can be represented as

$$y_a(t) = \sum_{i=1}^{n} c_i y_i(t) \tag{3.12}$$

where the *constants* c_i are determined in terms of the initial conditions

$$y(0), \left.\frac{dy}{dt}\right|_{t=0}, \ldots, \left.\frac{d^{n-1}y}{dt^{n-1}}\right|_{t=0}$$

from the set of n algebraic equations

$$y(0) = \sum_{i=1}^{n} c_i y_i(0), \left.\frac{dy}{dt}\right|_{t=0} = \sum_{i=1}^{n} c_i \left.\frac{dy_i}{dt}\right|_{t=0}, \ldots, \left.\frac{d^{n-1}y}{dt^{n-1}}\right|_{t=0} = \sum_{i=1}^{n} c_i \left.\frac{d^{n-1}y_i}{dt^{n-1}}\right|_{t=0} \tag{3.13}$$

The linear independence of the $y_i(t)$ guarantees that a solution to these equations can be obtained for $c_1, c_2, \ldots, c_n$.

EXAMPLE 3.24. The free response $y_a(t)$ of the differential equation

$$\frac{d^2y}{dt^2} + 3\frac{dy}{dt} + 2y = u$$

with initial conditions $y(0) = 0$, $(dy/dt)|_{t=0} = 1$ is determined by letting

$$y_a(t) = c_1 e^{-t} + c_2 e^{-2t}$$

where c_1 and c_2 are unknown coefficients and e^{-t} and e^{-2t} are a fundamental set for the equation (Example 3.21). Since $y_a(t)$ must satisfy the initial conditions, that is,

$$y_a(0) = y(0) = 0 = c_1 + c_2 \qquad \left.\frac{dy_a(t)}{dt}\right|_{t=0} = \left.\frac{dy}{dt}\right|_{t=0} = 1 = -c_1 - 2c_2$$

then $c_1 = 1$ and $c_2 = -1$. The free response is therefore given by $y_a(t) = e^{-t} - e^{-2t}$.

3.10 THE FORCED RESPONSE

The **forced response** $y_b(t)$ of a differential equation is the solution of the differential equation when all the initial conditions

$$y(0), \left.\frac{dy}{dt}\right|_{t=0}, \ldots, \left.\frac{d^{n-1}y}{dt^{n-1}}\right|_{t=0}$$

are identically zero.

The implication of this definition is that the forced response depends only on the input $u(t)$. The *forced response* for a linear constant-coefficient ordinary differential equation can be written in terms of a *convolution integral* (see Example 3.38):

$$y_b(t) = \int_0^t w(t-\tau)\left[\sum_{i=0}^{m} b_i \frac{d^i u(\tau)}{d\tau^i}\right] d\tau \tag{3.14}$$

where $w(t-\tau)$ is the *weighting function (or kernel) of the differential equation*. This form of the convolution integral assumes that the weighting function describes a *causal* system (see Definition 3.22). This assumption is maintained below.

The weighting function of a linear constant-coefficient ordinary differential equation can be written as

$$\begin{aligned} w(t) &= \sum_{i=1}^{n} c_i y_i(t) & t \ge 0 \\ &= 0 & t < 0 \end{aligned} \tag{3.15}$$

where $c_1, \ldots, c_n$ are constants and the set of functions $y_1(t), y_2(t), \ldots, y_n(t)$ is a fundamental set of the differential equation. It should be noted that $w(t)$ *is a free response of the differential equation* and therefore requires n initial conditions for complete specification. These conditions fix the values of the constants $c_1, c_2, \ldots, c_n$. The initial conditions which all weighting functions of linear differential equations must satisfy are

$$w(0) = 0, \left.\frac{dw}{dt}\right|_{t=0} = 0, \ldots, \left.\frac{d^{n-2}w}{dt^{n-2}}\right|_{t=0} = 0, \left.\frac{d^{n-1}w}{dt^{n-1}}\right|_{t=0} = 1 \tag{3.16}$$

EXAMPLE 3.25. The weighting function of the differential equation

$$\frac{d^2y}{dt^2}+3\frac{dy}{dt}+2y=u$$

is a linear combination of e^{-t} and e^{-2t} (a fundamental set of the equation). That is,

$$w(t)=c_1e^{-t}+c_2e^{-2t}$$

c_1 and c_2 are determined from the two algebraic equations

$$w(0)=0=c_1+c_2 \qquad \left.\frac{dw}{dt}\right|_{t=0}=1=-c_1-2c_2$$

The solution is $c_1=1$, $c_2=-1$, and the weighting function is $w(t)=e^{-t}-e^{-2t}$.

EXAMPLE 3.26. For the differential equation of Example 3.25, if $u(t)=1$, then the forced response $y_b(t)$ of the equation is

$$y_b(t)=\int_0^t w(t-\tau)u(\tau)\,d\tau=\int_0^t\left[e^{-(t-\tau)}-e^{-2(t-\tau)}\right]d\tau$$

$$=e^{-t}\int_0^t e^{\tau}\,d\tau-e^{-2t}\int_0^t e^{2\tau}\,d\tau=\frac{1}{2}(1-2e^{-t}+e^{-2t})$$

3.11 THE TOTAL RESPONSE

The **total response** of a linear constant-coefficient differential equation is the sum of the *free response* and the *forced response*.

EXAMPLE 3.27. The total response $y(t)$ of the differential equation

$$\frac{d^2y}{dt^2}+3\frac{dy}{dt}+2y=1$$

with initial conditions $y(0)=0$ and $(dy/dt)|_{t=0}=1$ is the sum of the free response $y_a(t)$ determined in Example 3.24 and the forced response $y_b(t)$ determined in Example 3.26. Thus

$$y(t)=y_a(t)+y_b(t)=(e^{-t}-e^{-2t})+\frac{1}{2}(1-2e^{-t}+e^{-2t})=\frac{1}{2}(1-e^{-2t})$$

3.12 THE STEADY STATE AND TRANSIENT RESPONSES

The *steady state response* and *transient response* are another pair of quantities whose sum is equal to the total response. These terms are often used for specifying control system performance. They are defined as follows.

Definition 3.13: The **steady state response** is that part of the total response which *does not* approach zero as time approaches infinity.

Definition 3.14: The **transient response** is that part of the total response which approaches zero as time approaches infinity.

EXAMPLE 3.28. The total response for the differential equation in Example 3.27 was determined as $y=\frac{1}{2}-\frac{1}{2}e^{-2t}$. Clearly, the steady state response is given by $y_{ss}=\frac{1}{2}$. Since $\lim_{t\to\infty}[-\frac{1}{2}e^{-2t}]=0$, the transient response is $y_T=-\frac{1}{2}e^{-2t}$.

3.13 SINGULARITY FUNCTIONS: STEPS, RAMPS, AND IMPULSES

In the study of control systems and the equations which describe them, a particular family of functions called *singularity functions* is used extensively. Each member of this family is related to the others by one or more integrations or differentiations. The three most widely used singularity functions are the *unit step*, the *unit impulse*, and the *unit ramp*.

Definition 3.15: A **unit step function** $\mathbf{1}(t-t_0)$ is defined by

$$\mathbf{1}(t-t_0)=\begin{cases}1 & \text{for } t>t_0\\ 0 & \text{for } t\le t_0\end{cases} \tag{3.17}$$

The unit step function is illustrated in Fig. 3-1.

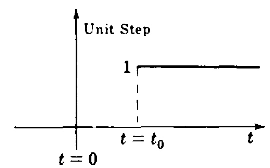


Fig. 3-1

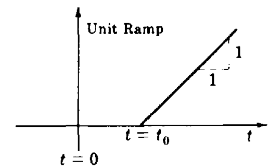


Fig. 3-2

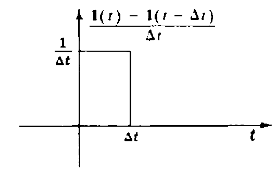


Fig. 3-3

Definition 3.16: A **unit ramp function** is the integral of a unit step function

$$\int_{-\infty}^{t}\mathbf{1}(\tau-t_0)\,d\tau=\begin{cases}t-t_0 & \text{for } t>t_0\\ 0 & \text{for } t\le t_0\end{cases} \tag{3.18}$$

The unit ramp function is illustrated in Fig. 3-2.

Definition 3.17: A **unit impulse function** $\delta(t)$ may be defined by

$$\delta(t)=\lim_{\substack{\Delta t\to 0\\ \Delta t>0}}\left[\frac{\mathbf{1}(t)-\mathbf{1}(t-\Delta t)}{\Delta t}\right] \tag{3.19*}$$

where $\mathbf{1}(t)$ is the unit step function.

The pair $\left\{\begin{matrix}\Delta t\to 0\\ \Delta t>0\end{matrix}\right\}$ may be abbreviated by $\Delta t\to 0^+$, meaning that Δt approaches zero *from the right*. The quotient in brackets represents a rectangle of height $1/\Delta t$ and width Δt as shown in Fig. 3-3. The limiting process produces a function whose height approaches infinity and width approaches zero. The area under the curve is equal to 1 for all values of Δt. That is,

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

The unit impulse function has the following very important property:

Screening Property: The integral of the product of a unit impulse function $\delta(t-t_0)$ and a function $f(t)$, continuous at $t=t_0$ over an interval which includes t_0, is equal to the function

*In a formal sense, Equation (*3.19*) defines the *one-sided derivative* of the unit step function. But neither the limit nor the derivative exist in the ordinary mathematical sense. However, Definition 3.17 is satisfactory for the purposes of this book, and many others.

$f(t)$ evaluated at t_0, that is,

$$\int_{-\infty}^{\infty} f(t)\,\delta(t-t_0)\,dt = f(t_0) \tag{3.20}$$

Definition 3.18: The **unit impulse response** of a system is the output $y(t)$ of the system when the input $u(t)=\delta(t)$ and all initial conditions are zero.

EXAMPLE 3.29. If the input-output relationship of a linear system is given by the convolution integral

$$y(t) = \int_0^t w(t-\tau)\,u(\tau)\,d\tau$$

then the unit impulse response $y_\delta(t)$ of the system is

$$y_\delta(t) = \int_0^t w(t-\tau)\,\delta(\tau)\,d\tau = \int_{-\infty}^{\infty} w(t-\tau)\,\delta(\tau)\,d\tau = w(t) \tag{3.21}$$

since $w(t-\tau)=0$ for $\tau>t$, $\delta(\tau)=0$ for $\tau<0$, and the screening property of the unit impulse has been used to evaluate the integral.

Definition 3.19: The **unit step response** is the output $y(t)$ when the input $u(t)=\mathbf{1}(t)$ and all initial conditions are zero.

Definition 3.20: The **unit ramp response** is the output $y(t)$ when the input $u(t)=t$ for $t>0$, $u(t)=0$ for $t\le 0$, and all initial conditions are zero.

3.14 SECOND-ORDER SYSTEMS

In the study of control systems, linear constant-coefficient second-order differential equations of the form:

$$\frac{d^2y}{dt^2} + 2\zeta\omega_n\frac{dy}{dt} + \omega_n^2 y = \omega_n^2 u \tag{3.22}$$

are important because higher-order systems can often be approximated by second-order systems. The constant ζ is called the **damping ratio**, and the constant ω_n is called the **undamped natural frequency** of the system. The forced response of this equation for inputs u belonging to the class of singularity functions is of particular interest. That is, the *forced response* to a unit impulse, unit step, or unit ramp is the same as the *unit impulse response*, *unit step response*, or *unit ramp response* of a system represented by this equation.

Assuming that $0\le\zeta\le 1$, the characteristic equation for Equation (*3.22*) is

$$D^2 + 2\zeta\omega_n D + \omega_n^2 = \left(D+\zeta\omega_n - j\omega_n\sqrt{1-\zeta^2}\right)\left(D+\zeta\omega_n + j\omega_n\sqrt{1-\zeta^2}\right) = 0$$

Hence the roots are

$$D_1 = -\zeta\omega_n + j\omega_n\sqrt{1-\zeta^2} \equiv -\alpha + j\omega_d \qquad D_2 = -\zeta\omega_n - j\omega_n\sqrt{1-\zeta^2} \equiv -\alpha - j\omega_d$$

where $\alpha\equiv\zeta\omega_n$ is called the **damping coefficient**, and $\omega_d\equiv\omega_n\sqrt{1-\zeta^2}$ is called the **damped natural frequency**. α is the inverse of the **time constant** τ of the system, that is, $\tau=1/\alpha$.

The weighting function of Equation (*3.22*) is $w(t)=(1/\omega_d)e^{-\alpha t}\sin\omega_d t$. The unit step response is given by

$$y_1(t) = \int_0^t w(t-\tau)\,\omega_n^2\,d\tau = 1 - \frac{\omega_n e^{-\alpha t}}{\omega_d}\sin(\omega_d t + \phi) \tag{3.23}$$

where $\phi\equiv\tan^{-1}(\omega_d/\alpha)$.

Figure 3-4 is a parametric representation of the unit step response. Note that the abscissa of this family of curves is normalized time $\omega_n t$, and the parameter defining each curve is the damping ratio ζ.

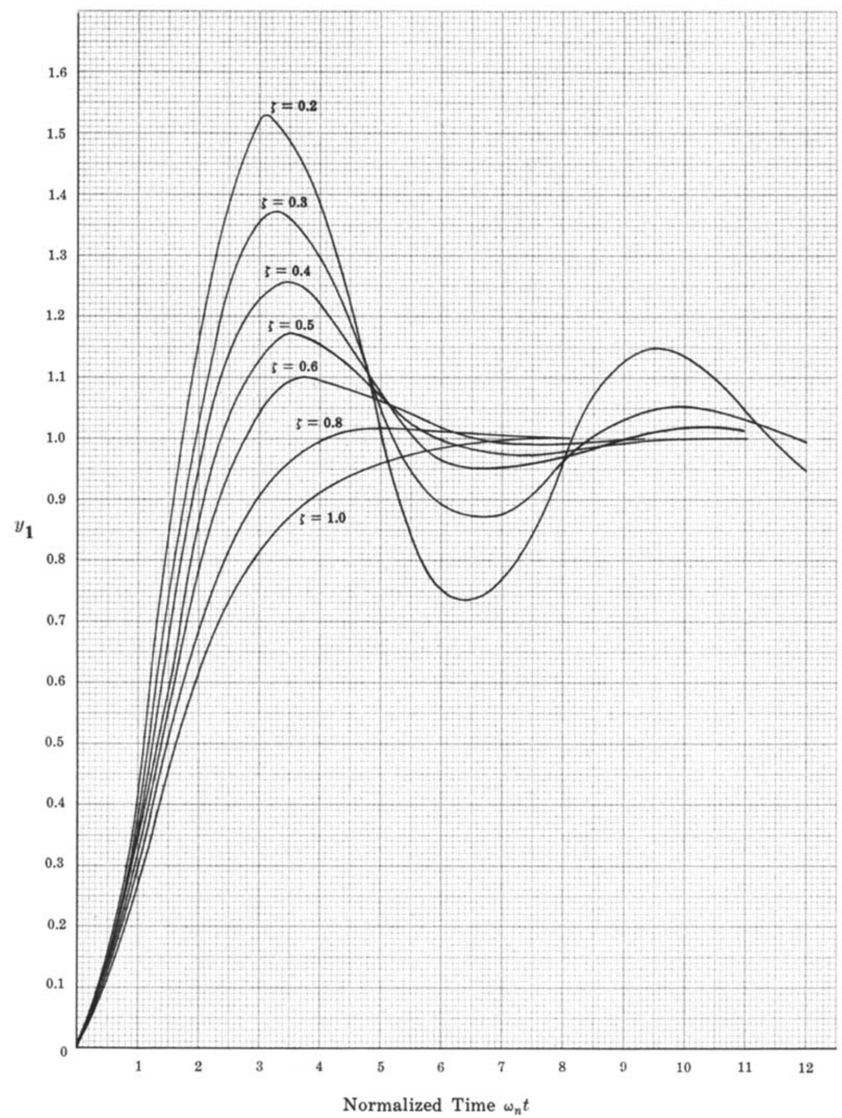


Fig. 3-4

3.15 STATE VARIABLE REPRESENTATION OF SYSTEMS DESCRIBED BY LINEAR DIFFERENTIAL EQUATIONS

In some problems of feedback and control, it is more convenient to describe a system by a set of first-order differential equations rather than by one or more nth-order differential equations. One reason is that quite general and powerful results from vector-matrix algebra can then be easily applied in deriving solutions for the differential equations.

EXAMPLE 3.30. Consider the differential equation form of Newton's second law, $f = M(d^2x/dt^2)$. It is clear from the meanings of velocity v and acceleration a that this second-order equation can be replaced by two first-order equations, $v = dx/dt$ and $f = M(dv/dt)$.

There are numerous ways to transform nth-order differential equations into n first-order equations. One of these is quite prevalent in the literature, and straightforward, and we introduce only this transformation here, to illustrate the approach. Consider the nth-order, single-input linear constant-coefficient differential equation

$$\sum_{i=0}^{n} a_i \frac{d^i y}{dt^i} = u$$

This equation can always be replaced by the following n first-order differential equations:

$$\begin{aligned}
\frac{dx_1}{dt} &= x_2 \\
\frac{dx_2}{dt} &= x_3 \\
&\ \vdots \\
\frac{dx_{n-1}}{dt} &= x_n \\
\frac{dx_n}{dt} &= -\frac{1}{a_n}\left[\sum_{i=0}^{n-1} a_i x_{i+1}\right] + \frac{1}{a_n}u
\end{aligned} \tag{3.24a}$$

where we have chosen $x_1 \equiv y$. Using *vector-matrix* notation, this set of equations can be written as

$$\begin{bmatrix} \dfrac{dx_1}{dt} \\ \dfrac{dx_2}{dt} \\ \vdots \\ \dfrac{dx_n}{dt} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\dfrac{a_0}{a_n} & -\dfrac{a_1}{a_n} & -\dfrac{a_2}{a_n} & \cdots & -\dfrac{a_{n-1}}{a_n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \dfrac{1}{a_n} \end{bmatrix} u \tag{3.24b}$$

or, more compactly, as

$$\frac{d\mathbf{x}}{dt} = A\mathbf{x} + \mathbf{b}u \tag{3.24c}$$

In Equation (*3.24c*) $\mathbf{x} \equiv \mathbf{x}(t)$ is called the **state vector**, with n time functions $x_1(t), x_2(t), \ldots, x_n(t)$ as its elements, called the **state variables** of the system. The scalar input of the system is $u(t)$.

More generally, *multiinput-multioutput* (*MIMO*) *systems* described by one or more linear constant-coefficient differential equations can be represented by a vector-matrix differential equation of the form:

$$\begin{bmatrix} \dfrac{dx_1}{dt} \\ \dfrac{dx_2}{dt} \\ \vdots \\ \dfrac{dx_n}{dt} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1r} \\ b_{21} & b_{22} & \cdots & b_{2r} \\ \vdots & & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nr} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_r \end{bmatrix} \tag{3.25a}$$

or, more compactly, as

$$\frac{d\mathbf{x}}{dt} = A\mathbf{x} + B\mathbf{u} \tag{3.25b}$$

In Equation (*3.25b*) $\mathbf{x}$ is defined as in Equation (*3.24c*), A is the $n \times n$ matrix of constants a_{ij}, and B is the $n \times r$ matrix of constants b_{ij}, each given in Equation (*3.25a*), and $\mathbf{u}$ is an r-vector of input functions.

The Transition Matrix

The matrix equation

$$\frac{d\Phi}{dt} = A\Phi$$

where Φ is an $n \times n$ matrix of time functions, called the **transition matrix of the differential equation** (*3.24c*) or (*3.25b*), has a special role in the solution of vector-matrix differential equations like Equation (*3.25b*). If I is the $n \times n$ *identity* or *unit* matrix, and $\Phi(0) = I$ is the initial condition of this homogeneous equation, the transition matrix has the special solution: $\Phi(t) = e^{At}$. In this case e^{At} is an $n \times n$ matrix function defined by the infinite series:

$$e^{At} = I + At + \frac{A^2t^2}{2!} + \frac{A^3t^3}{3!} + \cdots$$

Φ also has the property, called the *transition property*, that for all t_1, t_2, and t_3: $\Phi(t_1 - t_2)\Phi(t_2 - t_3) = \Phi(t_1 - t_3)$.

To solve the differential equation (*3.24*) or (*3.25*), the time interval of interest must be specified, for example, $0 \le t < +\infty$, and an initial condition vector $\mathbf{x}(0)$ is also needed. In this case, the general solution of Equation (3.25) is

$$\mathbf{x}(t) = e^{At}\mathbf{x}(0) + \int_0^t e^{A(t-\tau)}B\mathbf{u}(\tau)\,d\tau \tag{3.26}$$

The initial condition $\mathbf{x}(0)$ is sometimes referred to as the **state of the system at time t** $= 0$. From Equation (*3.26*) we see that knowledge of $\mathbf{x}(0)$, and the input $\mathbf{u}(t)$ on the interval $0 \le t < +\infty$, are adequate to completely determine the state variables for all time $t \ge 0$. Actually, knowledge of the state of the system at *any* time t', $0 < t' < +\infty$, and knowledge of the input $u(t)$, $t' \le t < +\infty$, are adequate to completely define the state vector $\mathbf{x}(t)$ at all subsequent times $t \ge t'$.

3.16 SOLUTION OF LINEAR CONSTANT-COEFFICIENT DIFFERENCE EQUATIONS

Consider the class of difference equations

$$\sum_{i=0}^{n} a_i y(k+i) = \sum_{i=0}^{m} b_i u(k+i) \tag{3.27}$$

where k is the integer-valued discrete-time variable, the coefficients a_i and b_i are constant, a_0 and a_n are nonzero, the input $u(k)$ is a known time sequence, and the output $y(k)$ is the unknown sequence solution of the equation. Since $y(k+n)$ is an explicit function of $y(k), y(k+1), \ldots, y(k+n-1)$, then n is the **order of the difference equation**. To obtain a unique solution for $y(k)$, two additional items must be specified, the time sequence over which a solution is desired, and a set of n initial conditions for $y(k)$. The time sequence for the class of problems treated in this book is the set of nonnegative integers, that is, $k = 0, 1, 2, \ldots$. The set of initial conditions is

$$y(0), y(1), \ldots, y(n-1) \tag{3.28}$$

A problem defined over this time sequence and with these initial conditions is called an **initial value problem**.

Consider the nth-order linear constant-coefficient difference equation

$$y(k+n)+a_{n-1}y(k+n-1)+\cdots+a_1y(k+1)+a_0y(k)=u(k) \tag{3.29}$$

It is convenient to define a **shift operator** Z by the equation

$$Z[y(k)]\equiv y(k+1)$$

By repeated application of this operation, we obtain

$$Z^n[y(k)]=Z[Z[\dots Z[y(k)]\dots]]=y(k+n)$$

Similarly, a **unity operator** I is defined by

$$I[y(k)]=y(k)$$

and $Z^0\equiv I$. The operator Z has the following important algebraic properties:

1. For constant c, $Z[cy(k)]=cZ[y(k)]$
2. $Z^m[y(k)+x(k)]=Z^m[y(k)]+Z^m[x(k)]$

The difference equation can thus be written as

$$Z^n[y(k)]+a_{n-1}Z^{n-1}[y(k)]+\cdots+a_1Z[y(k)]+a_0y(k)=u(k)$$

or

$$(Z^n+a_{n-1}Z^{n-1}+\cdots+a_1Z+a_0)[y(k)]=u(k)$$

The equation

$$Z^n+a_{n-1}Z^{n-1}+\cdots+a_1Z+a_0=0 \tag{3.30}$$

is called the **characteristic equation** of the difference equation, and, by the fundamental theorem of algebra, it has exactly n solutions: $Z=Z_1, Z=Z_2,\dots,Z=Z_n$.

EXAMPLE 3.31. Consider the difference equation

$$y(k+2)+\frac{5}{6}y(k+1)+\frac{1}{6}y(k)=u(k)$$

The characteristic equation is $Z^2+\frac{5}{6}Z+\frac{1}{6}=0$ with two solutions, $Z=-\frac{1}{2}$ and $Z=-\frac{1}{3}$.

A homogeneous nth-order linear difference equation has at least one set of n linearly independent solutions. Any such set is called a **fundamental set**. As with differential equations, fundamental sets are not unique.

If the characteristic equation has distinct roots $Z_1, Z_2,\dots,Z_n$, a fundamental set for the homogeneous equation

$$\sum_{i=0}^{n}a_iy(k+i)=0 \tag{3.31}$$

is the set of functions $Z_1^k, Z_2^k,\dots,Z_n^k$.

EXAMPLE 3.32. The difference equation

$$y(k+2)+\frac{5}{6}y(k+1)+\frac{1}{6}y(k)=0$$

has the characteristic equation $Z^2+\frac{5}{6}Z+\frac{1}{6}=0$, with roots $Z=Z_1=-\frac{1}{2}$ and $Z=Z_2=-\frac{1}{3}$. A fundamental set of this equation is $y_1(k)=(-\frac{1}{2})^k$ and $y_2(k)=(-\frac{1}{3})^k$.

If the characteristic equation has repeated roots, then for each root Z_i of multiplicity n_i, there are n_i elements of the fundamental set $Z_i^k, kZ_i^k,\dots,k^{n_i-2}Z_i^k, k^{n_i-1}Z_i^k$.

EXAMPLE 3.33. The equation $y(k+2)+y(k+1)+\frac{1}{4}y(k)=0$ with the repeated root $Z=-\frac{1}{2}$ has a fundamental set consisting of $(-\frac{1}{2})^k$ and $k(-\frac{1}{2})^k$.

The free response of a difference equation of the form of Equation (*3.27*) is the solution when the input sequence is identically zero. The equation then has the form of Equation (*3.31*) and its solution

depends only on the n initial conditions (*3.28*). If $y_1(k), y_2(k),\dots,y_n(k)$ is a fundamental set, then any free response of the difference equation (*3.27*) can be represented as

$$y_\alpha(k)=\sum_{i=1}^{n}c_iy_i(k)$$

where the constants c_i are determined in terms of the initial conditions $y_i(0)$ from the set of n algebraic equations:

$$\begin{aligned} y(0)&=\sum_{i=1}^{n}c_iy_i(0)\\ y(1)&=\sum_{i=1}^{n}c_iy_i(1)\\ &\vdots\\ y(n-1)&=\sum_{i=1}^{n}c_iy_i(n-1) \end{aligned} \tag{3.32}$$

The linear independence of the $y_i(k)$ guarantees a solution for $c_1, c_2,\dots,c_n$.

EXAMPLE 3.34. The free response of the difference equation $y(k+2)+\frac{5}{6}y(k+1)+\frac{1}{6}y(k)=u(k)$ with initial conditions $y(0)=0$ and $y(1)=1$ is determined by letting

$$y_\alpha(k)=c_1\left(-\frac{1}{2}\right)^k+c_2\left(-\frac{1}{3}\right)^k$$

where c_1 and c_2 are unknown coefficients and $(-\frac{1}{2})^k$ and $(-\frac{1}{3})^k$ are a fundamental set for the equation (Example 3.32). Since $y_\alpha(k)$ must satisfy the initial conditions, that is,

$$y_\alpha(0)=y(0)=0=c_1+c_2$$

$$y_\alpha(1)=y(1)=1=-\frac{1}{2}c_1-\frac{1}{3}c_2$$

then $c_1=-6$ and $c_2=6$. The free response is therefore given by $y_\alpha(k)=-6(-\frac{1}{2})^k+6(-\frac{1}{3})^k$.

The forced response $y_b(k)$ of a difference equation is its solution when all initial conditions $y(0), y(1),\dots,y(n-1)$ are zero. It can be written in terms of a *convolution sum*:

$$y_b(k)=\sum_{j=0}^{k-1}w(k-j)\left[\sum_{i=0}^{m}b_iu(j+i)\right]\qquad k=0,1,\dots,n \tag{3.33}$$

where $w(k-j)$ is the **weighting sequence of the difference equation**. Note that $y_b(0)=0$ by definition of the forced response, and $w(k-j)=0$ for $k<j$ (see Section 3.19). If $u(j)\equiv\delta(j)=1$ for $j=0$, and $\delta(j)=0$ for $j\neq 0$, the special input called the **Kronecker delta sequence**, then the forced response $y_b(k)\equiv y_\delta(k)$ is called the **Kronecker delta response**.

The weighting sequence of a linear constant-coefficient difference equation can be written as

$$w(k-l)=\sum_{j=1}^{n}\frac{M_j(l)}{a_nM(l)}y_j(k) \tag{3.34}$$

where $y_1(k), y_2(k),\dots,y_n(k)$ is a fundamental set of the difference equation, $M(l)$ is the **determinant**:

$$M(l)=\begin{vmatrix} y_1(l+1) & y_2(l+1) & \cdots & y_n(l+1)\\ y_1(l+2) & y_2(l+2) & \cdots & y_n(l+2)\\ \vdots & \vdots & \ddots & \vdots\\ y_1(l+n) & y_2(l+n) & \cdots & y_n(l+n) \end{vmatrix}$$

and $M_j(l)$ is the **cofactor** of the last element in the jth column of $M(l)$.

EXAMPLE 3.35. Consider the difference equation $y(k+2)+\frac{5}{6}y(k+1)+\frac{1}{6}y(k)=u(k)$. The weighting sequence is given by

$$w(k-l)=\frac{M_1(l)}{M(l)}y_1(k)+\frac{M_2(l)}{M(l)}y_2(k)$$

where $y_1(k)=(-\frac{1}{2})^k$, $y_2(k)=(-\frac{1}{3})^k$, $M_1(l)=-(-\frac{1}{3})^{l+1}$, $M_2(l)=(-\frac{1}{2})^{l+1}$, and

$$M(l)=\begin{vmatrix}\left(-\frac{1}{2}\right)^{l+1} & \left(-\frac{1}{3}\right)^{l+1}\\ \left(-\frac{1}{2}\right)^{l+2} & \left(-\frac{1}{3}\right)^{l+2}\end{vmatrix}=\frac{1}{36}\left(-\frac{1}{2}\right)^{l}\left(-\frac{1}{3}\right)^{l}$$

Therefore

$$w(k-l)=12\left(-\frac{1}{2}\right)^{k-l}-18\left(-\frac{1}{3}\right)^{k-l}$$

As for continuous systems, the **total response** of a difference equation is the sum of the free and forced responses of the equation. The **transient response** of a difference equation is that part of the total response which approaches zero as time approaches infinity. That part of the total response which does not approach zero is called the **steady state response**.

3.17 STATE VARIABLE REPRESENTATION OF SYSTEMS DESCRIBED BY LINEAR DIFFERENCE EQUATIONS

As with differential equations in Section 3.15, it is often useful to describe a system by a set of first-order difference equations, rather than by one or more nth-order difference equations.

EXAMPLE 3.36. The second-order difference equation

$$y(k+2)+\frac{5}{6}y(k+1)+\frac{1}{6}y(k)=u(k)$$

can be written as the two first-order equations:

$$x_1(k+1)=x_2(k)$$
$$x_2(k+1)=-\frac{5}{6}x_2(k)-\frac{1}{6}x_1(k)+u(k)$$

where we have chosen $x_1(k)\equiv y(k)$.

Consider the nth-order, single-input, linear constant-coefficient difference equation

$$\sum_{i=0}^{n}a_iy(k+i)=u(k)$$

This equation can always be replaced by the following n first-order difference equations:

$$\begin{aligned}x_1(k+1)&=x_2(k)\\ x_2(k+1)&=x_3(k)\\ &\vdots\\ x_{n-1}(k+1)&=x_n(k)\\ x_n(k+1)&=-\frac{1}{a_n}\left[\sum_{i=0}^{n-1}a_ix_{i+1}(k)\right]+\frac{1}{a_n}u(k)\end{aligned}\tag{3.35a}$$

where we have chosen $x_1(k)\equiv y(k)$. Using vector-matrix notation, this set of equations can be written

as the *vector-matrix* difference equation

$$\begin{bmatrix}x_1(k+1)\\ x_2(k+1)\\ \vdots\\ x_n(k+1)\end{bmatrix}=\begin{bmatrix}0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & & 0\\ \cdot & \cdot & \cdot & & 0\\ \cdot & \cdot & & & 1\\ -a_0/a_n & -a_1/a_n & \cdots & & -a_{n-1}/a_n\end{bmatrix}\begin{bmatrix}x_1(k)\\ x_2(k)\\ \vdots\\ x_n(k)\end{bmatrix}+\begin{bmatrix}0\\ 0\\ \vdots\\ 0\\ 1/a_n\end{bmatrix}u \tag{3.35b}$$

or, more compactly, as

$$\mathbf{x}(k+1)=A\mathbf{x}(k)+\mathbf{b}u \tag{3.35c}$$

In these equations, $\mathbf{x}(k)$ is an n-vector element of a time sequence called the **state vector**, made up of scalar elements $x_1(k), x_2(k),\ldots,x_n(k)$ called the **state variables** of the system at time k.

In general, *multiinput-multioutput* (MIMO) systems described by one or more linear constant-coefficient difference equations can be represented by

$$\mathbf{x}(k+1)=A\mathbf{x}(k)+B\mathbf{u}(k) \tag{3.36}$$

where $\mathbf{x}(k)$ is the state vector of the system, as above, A is an $n\times n$ matrix of constants a_{ij}, and B is an $n\times r$ matrix of constants b_{ij}, each defined as in Equation (*3.25a*), and $\mathbf{u}(k)$ is an r-vector element of a (multiple) input sequence. Given a time sequence of interest $k=0,1,2,\ldots,$ and an initial condition vector $\mathbf{x}(0)$, the solution of Equation (*3.36*) can be written as

$$\mathbf{x}(k)=A^k\mathbf{x}(0)+\sum_{j=0}^{k-1}A^{k-1-j}B\mathbf{u}(j) \tag{3.37}$$

Note that Equation (*3.37*) has a form similar to Equation (3.26). In general, however, A^k need not have the properties of a transition matrix of a differential equation. But there is one very important case when A^k does have such properties, that is, where A^k *is* a transition matrix. This case provides the basis for *discretization* of differential equations, as illustrated next.

Discretization of Differential Equations

Consider a *differential* system described by Equation (*3.26*). Suppose it is only necessary to have knowledge of the state variables at periodic time instants $t=0,T,2T,\ldots,kT,\ldots,$. In this case, the following *sequence* of state vectors can be written as

$$\mathbf{x}(T)=e^{AT}\mathbf{x}(0)+\int_0^T e^{A(T-\tau)}B\mathbf{u}(\tau)\,d\tau$$

$$\mathbf{x}(2T)=e^{AT}\mathbf{x}(T)+e^{AT}\int_T^{2T}e^{A(T-\tau)}B\mathbf{u}(\tau)\,d\tau$$

$$\vdots$$

$$\mathbf{x}(kT)=e^{AT}\mathbf{x}((k-1)T)+e^{A(k-1)T}\int_{(k-1)T}^{kT}e^{A(T-\tau)}B\mathbf{u}(\tau)\,d\tau$$

If we suppress the parameter T, use the abbreviation $\mathbf{x}(k)\equiv\mathbf{x}(kT)$, and define a new *input sequence* by

$$\mathbf{u}'(k)=e^{AkT}\int_{kT}^{(k+1)T}e^{A(T-\tau)}B\mathbf{u}(\tau)\,d\tau$$

then the set of solution equations above can be replaced by the single *vector-matrix difference equation*

$$\mathbf{x}(k+1)=e^{AT}\mathbf{x}(k)+\mathbf{u}'(k) \tag{3.38}$$

Note that $A'\equiv e^{AT}$ is a transition matrix in Equation (*3.38*).

3.18 LINEARITY AND SUPERPOSITION

The concept of linearity has been presented in Definition 3.8 as a property of differential and difference equations. In this section, linearity is discussed as a property of *general systems*, with one independent variable, time t. In Chapters 1 and 2, the concepts of system, input, and output were defined. The following definition of linearity is based on these earlier definitions.

Definition 3.21: If all initial conditions in the system are zero, that is, if the system is completely at rest, then the system is a **linear system** if it has the following property:

(*a*) If an input $u_1(t)$ produces an output $y_1(t)$, and

(*b*) an input $u_2(t)$ produces an output $y_2(t)$,

(*c*) then input $c_1u_1(t)+c_2u_2(t)$ produces an output $c_1y_1(t)+c_2y_2(t)$ for all pairs of inputs $u_1(t)$ and $u_2(t)$ and all pairs of constants c_1 and c_2.

Linear systems can often be represented by linear differential or difference equations.

EXAMPLE 3.37. A system is *linear* if its input-output relationship can be described by the linear differential equation

$$\sum_{i=0}^{n} a_i(t)\frac{d^iy}{dt^i}=\sum_{i=0}^{m} b_i(t)\frac{d^iu}{dt^i} \tag{3.39}$$

where $y=y(t)$ is the system output and $u=u(t)$ is the system input.

EXAMPLE 3.38. A system is linear if its input-output relationship can be described by the **convolution integral**

$$y(t)=\int_{-\infty}^{\infty} w(t,\tau)u(\tau)\,d\tau \tag{3.40}$$

where $w(t,\tau)$ is the **weighting function**, which embodies the internal physical properties of the system, $y(t)$ is the output, and $u(t)$ is the input.

The relationship between the systems of Examples 3.37 and 3.38 is discussed in Section 3.10.
The concept of linearity is often expressed by the *principle of superposition*.

Principle of Superposition: The response $y(t)$ of a linear system due to several inputs $u_1(t), u_2(t), \ldots, u_n(t)$ acting simultaneously is equal to the sum of the responses of each input acting alone, when all initial conditions in the system are zero. That is, if $y_i(t)$ is the response due to the input $u_i(t)$, then

$$y(t)=\sum_{i=1}^{n} y_i(t)$$

EXAMPLE 3.39. A linear system is described by the linear algebraic equation

$$y(t)=2u_1(t)+u_2(t)$$

where $u_1(t)=t$ and $u_2(t)=t^2$ are inputs, and $y(t)$ is the output. When $u_1(t)=t$ and $u_2(t)=0$, then $y(t)=y_1(t)=2t$. When $u_1(t)=0$ and $u_2(t)=t^2$, then $y(t)=y_2(t)=t^2$. The total output resulting from $u_1(t)=t$ and $u_2(t)=t^2$ is then equal to

$$y(t)=y_1(t)+y_2(t)=2t+t^2$$

The principle of superposition follows directly from the definition of linearity (Definition 3.21). Any system which satisfies the principle of superposition is linear.

3.19 CAUSALITY AND PHYSICALLY REALIZABLE SYSTEMS

The properties of a physical system restrict the form of its output. This restriction is embodied in the concept of *causality*.

Definition 3.22: A system in which time is the independent variable is called **causal** if the output depends only on the present and past values of the input. That is, if $y(t)$ is the output, then $y(t)$ depends only on the input $u(\tau)$ for values of $\tau\le t$.

The implication of Definition 3.22 is that a *causal* system is one which cannot anticipate what its future input will be. Accordingly, causal systems are sometimes called **physically realizable** systems. An important consequence of causality (physical realizability) is that the weighting function $w(t,\tau)$ of a causal linear continuous system is identically zero for $\tau>t$; that is, future values of the input are weighted zero. For causal discrete systems, the weighting sequence $w(k-j)\equiv 0$ for $j>k$.

Solved Problems

SYSTEM EQUATIONS

3.1. Faraday's law states that the voltage v induced between the terminals of an inductor is equal to the time rate of change of flux linkages. (A flux linkage is defined as one line of magnetic flux linking one turn of the winding of the inductor.) Suppose it is experimentally determined that the number of flux linkages λ is related to the current i in the inductor as shown in Fig. 3-5. The curve is approximately a straight line for $-I_0\le i\le I_0$. Determine a differential equation, valid for $-I_0\le i\le I_0$, which relates the induced voltage v and current i.

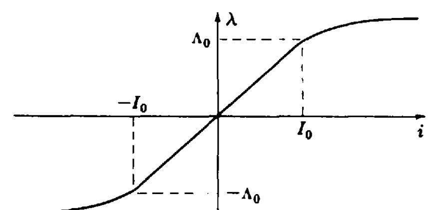


Fig. 3-5

Faraday's law can be written as $v=d\lambda/dt$. It is seen from the graph that

$$\lambda=\left(\frac{\Lambda_0}{I_0}\right)i=Li \qquad -I_0\le i\le I_0$$

where $L\equiv\Lambda_0/I_0$ is called the *inductance* of the inductor. The equation relating v and i is obtained by substituting Li for λ:

$$v=\frac{d\lambda}{dt}=\frac{d}{dt}(Li)=L\frac{di}{dt} \qquad \text{where} \quad -I_0\le i\le I_0$$

3.2. Determine a differential equation relating the voltage $v(t)$ and the current $i(t)$ for $t \geq 0$ for the electrical network given in Fig. 3-6. Assume the capacitor is uncharged at $t = 0$, the current i is zero at $t = 0$, and the switch S closes at $t = 0$.

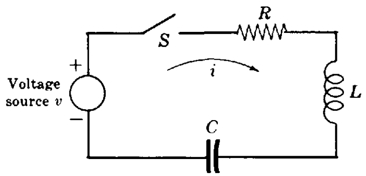


Fig. 3-6

By Kirchhoff's voltage law, the applied voltage $v(t)$ is equal to the sum of the voltage drops v_R, v_L, and v_C across the resistor R, the inductor L, and the capacitor C, respectively. Thus

$$v = v_R + v_L + v_C = Ri + L\frac{di}{dt} + \frac{1}{C}\int_0^t i(\tau)\, d\tau$$

To eliminate the integral, both sides of the equation are differentiated with respect to time, resulting in the desired differential equation:

$$L\frac{d^2 i}{dt^2} + R\frac{di}{dt} + \frac{i}{C} = \frac{dv}{dt}$$

3.3. Kepler's first two laws of planetary motion state that:

1. The orbit of a planet is an ellipse with the sun at a focus of the ellipse.
2. The radius vector drawn from the sun to a planet sweeps over equal areas in equal times.

Find a pair of differential equations that describes the motion of a planet about the sun, using Kepler's first two laws.

From Kepler's first law, the motion of a planet satisfies the equation of an ellipse:

$$r = \frac{p}{1 + e\cos\theta}$$

where r and θ are defined in Fig. 3-7, and $p \equiv b^2/a = a(1 - e^2)$.

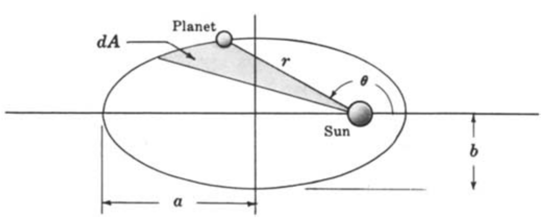


Fig. 3-7

In an infinitesimal time dt the angle θ increases by an amount $d\theta$. The area swept out by r over the period dt is therefore equal to $dA = \frac{1}{2}r^2\, d\theta$. The rate at which the area is swept out by r is constant

(Kepler's second law). Hence

$$\frac{dA}{dt} = \frac{1}{2}r^2\frac{d\theta}{dt} = \text{constant} \qquad \text{or} \qquad r^2\frac{d\theta}{dt} = k$$

The first differential equation is obtained by differentiating this result with respect to time:

$$2r\frac{dr}{dt}\frac{d\theta}{dt} + r^2\frac{d^2\theta}{dt^2} = 0 \qquad \text{or} \qquad 2\frac{dr}{dt}\frac{d\theta}{dt} + r\frac{d^2\theta}{dt^2} = 0$$

The second equation is obtained by differentiating the equation of the ellipse:

$$\frac{dr}{dt} = \left[\frac{pe\sin\theta}{(1 + e\cos\theta)^2}\right]\frac{d\theta}{dt}$$

Using the results that $d\theta/dt = k/r^2$ and $(1 + e\cos\theta) = p/r$, dr/dt can be rewritten as

$$\frac{dr}{dt} = \frac{ek}{p}\sin\theta$$

Differentiating again and replacing $r^2(d\theta/dt)$ with k yields

$$\frac{d^2 r}{dt^2} = \left(\frac{e}{p}\right)\left(\frac{k^2}{r^2}\right)\cos\theta$$

But $\cos\theta = (1/e)[p/r - 1]$. Hence

$$\frac{d^2 r}{dt^2} = \frac{k^2}{pr^2}\left[\frac{p}{r} - 1\right] = \frac{k^2}{r^3} - \frac{k^2}{pr^2}$$

Substituting $r(d\theta/dt)^2$ for k^2/r^3, we obtain the required second differential equation:

$$\frac{d^2 r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 + \frac{k^2}{pr^2} = 0 \qquad \text{or} \qquad \frac{d^2 r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 = -\frac{k^2}{pr^2}$$

3.4. A mathematical model for a feature of nervous system organization called *lateral inhibition* has been produced as a result of the work of several authors [2, 3, 4]. Lateral inhibitory phenomena can be simply described as inhibitory electrical interaction among laterally spaced, neighboring neurons (nerve cells). Each neuron in this model has a response c, measured by the frequency of discharge of pulses in its axon (the connection "cable" or "wire"). The response is determined by an excitation r supplied by an external stimulus, and is diminished by whatever inhibitory influences are acting on the neurons as a result of the activity of neighboring neurons. In a system of n neurons, the steady state response of the kth neuron is given by

$$c_k = r_k - \sum_{i=1}^{n} a_{k-i}c_i$$

where the constant a_{k-i} is the inhibitory coefficient of the action of neuron i on k. It depends only on the separation of the kth and ith neurons, and can be interpreted as a *spatial weighting function*. In addition, $a_m = a_{-m}$ (symmetrical spatial interaction).

(*a*) If the effect of neuron i on k is not immediately felt, but exhibits a small time lag Δt, how should this model be modified?

(*b*) If the input $r_k(t)$ is determined only by the output c_k, Δt seconds prior to t $[r_k(t) = c_k(t - \Delta t)]$, determine an approximate differential equation for the system of part (*a*).

(*a*) The equation becomes

$$c_k(t) = r_k(t) - \sum_{i=1}^{n} a_{k-i}c_i(t - \Delta t)$$

(*b*) Substituting $c_k(t-\Delta t)$ for $r_k(t)$,

$$c_k(t) - c_k(t-\Delta t) = -\sum_{i=1}^{n} a_{k-i}c_i(t-\Delta t)$$

Dividing both sides by Δt,

$$\frac{c_k(t) - c_k(t-\Delta t)}{\Delta t} = -\sum_{i=1}^{n}\left(\frac{a_{k-i}}{\Delta t}\right)c_i(t-\Delta t)$$

The left-hand side is approximately equal to dc_k/dt for small Δt. If we additionally assume that $c_i(t-\Delta t) \cong c_i(t)$ for small Δt, then we get the approximate differential equation

$$\frac{dc_k}{dt} + \sum_{i=1}^{n}\left(\frac{a_{k-i}}{\Delta t}\right)c_i(t) = 0$$

3.5. Determine a mathematical equation describing the sampled-data output of the ideal sampler described in Definition 2.12 and Example 2.8.

A convenient representation of the output of an ideal sampler is based on an extension of the concept of the unit impulse function $\delta(t)$ into an **impulse train**, defined for $t \geq 0$ as the function

$$m_{IT}(t) = \delta(t) + \delta(t-t_1) + \delta(t-t_2) + \cdots = \sum_{k=0}^{\infty}\delta(t-t_k)$$

where $t_0 = 0$ and $t_{k+1} > t_k$. The sampled signal $u^*(t)$ is then given by

$$u^*(t) = u(t)m_{IT}(t) = u(t)\sum_{k=0}^{\infty}\delta(t-t_k)$$

The utility of this representation is developed beginning in Chapter 4, following the introduction of transform methods.

3.6. Show how the simple R-C network given in Fig. 3-8 can be used to approximate the sample and (zero-order) hold function described in Example 2.9.

This system element operates as follows. When the sampling switch S is closed, the capacitor C is charged through the resistor R, and the voltage across C approaches the input $u(t)$. When S is opened, the capacitor cannot release its charge, because the current (charge) has nowhere to dissipate, so it *holds* its voltage until the next time S is closed. If we describe the opening and closing of the switch by the simple

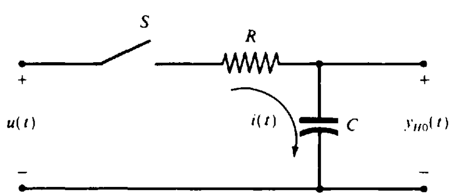


Fig. 3-8

function:

$$m_S(t) = \begin{cases} 1 & \text{if } S \text{ is closed} \\ 0 & \text{if } S \text{ is open} \end{cases}$$

we can say the current through R and C is *modulated* by $m_S(t)$. In these terms, we can write

$$i(t) = m_S(t)\left(\frac{u(t) - y_{H0}(t)}{R}\right)$$

and, since $i = C\,dy_{H0}/dt$, the differential equation for this circuit is

$$\frac{dy_{H0}}{dt} = \left(\frac{u - y_{H0}}{RC}\right)m_S(t)$$

We note that this is a *time-varying* differential equation, due to the multiplicative function $m_S(t)$ on the right-hand side. Also, as RC becomes smaller, that is, $1/RC$ becomes larger, dy_{H0}/dt becomes larger and the capacitor charges faster. Thus a smaller RC in this circuit creates a better approximation of the sample and hold function.

3.7. If the sampler in the previous problem is ideal, and the sampling rate is uniform, with period T, what is the differential equation?

The ideal sampler impulse train modulating function $m_{IT}(t)$ was defined in Problem 3.5. Thus the differential equation of the sample and hold becomes

$$\frac{dy_{H0}}{dt} = \left(\frac{u - y_{H0}}{RC}\right)\sum_{k=0}^{\infty}\delta(t-kT)$$

In this idealization, impulses replace current pulses.

CLASSIFICATIONS OF DIFFERENTIAL EQUATIONS

3.8. Classify the following differential equations according to whether they are ordinary or partial. Indicate the dependent and independent variables.

(*a*) $\dfrac{dx}{dt} + \dfrac{dy}{dt} + x + y = 0 \qquad x = x(t) \qquad y = y(t)$

(*b*) $\dfrac{\partial f}{\partial x} + \dfrac{\partial f}{\partial y} + x + y = 0 \qquad f = f(x, y)$

(*c*) $\dfrac{d}{dt}\left[\dfrac{\partial f}{\partial x}\right] = 0 \qquad f = x^2 + \dfrac{dx}{dt}$

(*d*) $\dfrac{df}{dx} = x \qquad f = y^2(x) + \dfrac{dy}{dx}$

(*a*) Ordinary; independent variable t; dependent variables x and y.

(*b*) Partial; independent variables x and y; dependent variable f.

(*c*) Since $\partial f/\partial x = 2x$, then $(d/dt)[\partial f/\partial x] = 2(dx/dt) = 0$, which is an ordinary differential equation; independent variable t; dependent variable x.

(*d*) $df/dx = 2y(dy/dx) + d^2y/dx^2 = x$, which is an ordinary differential equation; independent variable x; dependent variable y.

3.9. Classify the following linear differential equations according to whether they are time-variable or time-invariant. Indicate any time-variable terms.

(*a*) $\dfrac{d^2y}{dt^2} + 2y = 0 \qquad$ (*c*) $\left(\dfrac{1}{t+1}\right)\dfrac{d^2y}{dt^2} + \left(\dfrac{1}{t+1}\right)y = 0$

(*b*) $\dfrac{d}{dt}(t^2y) = 0 \qquad$ (*d*) $\dfrac{d^2y}{dt^2} + (\cos t)y = 0$

(*a*) Time-invariant.

(*b*) $(d/dt)(t^2y) = 2ty + t^2(dy/dt) = 0$. Dividing through by t, $t(dy/dt) + 2y = 0$ which is time-variable. The time-variable term is $t(dy/dt)$.

(*c*) Multiplying through by $t+1$, we obtain $d^2y/dt^2 + y = 0$ which is time-invariant.

(*d*) Time-variable. The time-variable term is $(\cos t)y$.

3.10. Classify the following differential equations according to whether they are linear or nonlinear. Indicate the dependent and independent variables and any nonlinear terms.

$(a)\ t\dfrac{dy}{dt}+y=0 \quad y=y(t) \qquad (d)\ (\cos t)\dfrac{d^2y}{dt^2}+(\sin 2t)y=0 \quad y=y(t)$

$(b)\ y\dfrac{dy}{dt}+y=0 \quad y=y(t) \qquad (e)\ (\cos y)\dfrac{d^2y}{dt^2}+\sin 2y=0 \quad y=y(t)$

$(c)\ \dfrac{dy}{dt}+y^2=0 \quad y=y(t) \qquad (f)\ (\cos x)\dfrac{d^2y}{dt^2}+\sin 2x=0 \quad y=y(t),\ x=x(t)$

(a) Linear; independent variable t; dependent variable y.
(b) Nonlinear; independent variable t; dependent variable y; nonlinear term $y(dy/dt)$.
(c) Nonlinear; independent variable t; dependent variable y; nonlinear term y^2.
(d) Linear; independent variable t; dependent variable y.
(e) Nonlinear; independent variable t; dependent variable y; nonlinear terms $(\cos y)\, d^2y/dt^2$ and $\sin 2y$.
(f) Nonlinear; independent variable t; dependent variables x and y; nonlinear terms $(\cos x)\, d^2y/dt^2$ and $\sin 2x$.

3.11. Why are all transcendental functions *not* of first degree?

Transcendental functions, such as the logarithmic, trigonometric, and hyperbolic functions and their corresponding inverses, are not of first degree because they are either defined by or can be written as infinite series. Hence their degree is in general equal to *infinity*. For example,

$$\sin x=\sum_{n=1}^{\infty}(-1)^{n-1}\frac{x^{2n-1}}{(2n-1)!}=x-\frac{x^3}{3!}+\frac{x^5}{5!}-\cdots$$

where the first term is first degree, the second is third degree, and so on.

THE CHARACTERISTIC EQUATION

3.12. Find the characteristic polynomial and characteristic equation for each system:

$$(a)\ \frac{d^4y}{dt^4}+9\frac{d^2y}{dt^2}+7y=u \qquad (b)\ \frac{d^4y}{dt^4}+9\frac{d^2y}{dt^2}+7y=\sin u$$

(a) Putting $D^n \equiv d^n/dt^n$ for $n=2$ and $n=4$, the characteristic polynomial is D^4+9D^2+7; and the characteristic equation is $D^4+9D^2+7=0$.

(b) Although the equation given in part (b) is nonlinear by Definition 3.8 (the term $\sin u$ is not first degree in u), we can treat it as a linear equation if we arbitrarily put $\sin u = x$, and treat x as a second dependent variable representing the input. In this case, part (b) has the same answer as part (a).

3.13. Determine the solution of the characteristic equation of the preceding problem.

Let $D^2 \equiv E$. Then $D^4=E^2$, and the characteristic equation becomes quadratic:

$$E^2+9E+7=0 \qquad E=-\frac{9\pm\sqrt{53}}{2} \qquad \text{and} \qquad D=\pm\sqrt{\frac{-9\pm\sqrt{53}}{2}}$$

LINEAR INDEPENDENCE AND FUNDAMENTAL SETS

3.14. Show that a sufficient condition for a set of n functions $f_1, f_2, \ldots, f_n$ to be linearly independent is that the determinant

$$\begin{vmatrix} f_1 & f_2 & \cdots & f_n \\ \dfrac{df_1}{dt} & \dfrac{df_2}{dt} & \cdots & \dfrac{df_n}{dt} \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{d^{n-1}f_1}{dt^{n-1}} & \dfrac{d^{n-1}f_2}{dt^{n-1}} & \cdots & \dfrac{d^{n-1}f_n}{dt^{n-1}} \end{vmatrix}$$

be nonzero. This determinant is called the **Wronskian** of the functions $f_1, f_2, \ldots, f_n$.

Assuming the f_i are differentiable at least $n-1$ times, let $n-1$ derivatives of

$$c_1f_1+c_2f_2+\cdots+c_nf_n=0$$

be formed as follows, where the c_i are unknown constants:

$$c_1\frac{df_1}{dt}+c_2\frac{df_2}{dt}+\cdots+c_n\frac{df_n}{dt}=0$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$c_1\frac{d^{n-1}f_1}{dt^{n-1}}+c_2\frac{d^{n-1}f_2}{dt^{n-1}}+\cdots+c_n\frac{d^{n-1}f_n}{dt^{n-1}}=0$$

These equations may be considered as n simultaneous linear homogeneous equations in the n unknown constants $c_1, c_2, \ldots, c_n$, with coefficients given by the elements of the Wronskian. It is well known that these equations have a nonzero solution for $c_1, c_2, \ldots, c_n$ (i.e., not all c_i are equal to zero) if and only if the determinant of the coefficients (the Wronskian) is equal to zero. Hence if the Wronskian is nonzero, then the only solution for $c_1, c_2, \ldots, c_n$ is the degenerate one, $c_1=c_2=\cdots=c_n=0$. Clearly, this is equivalent to saying that if the Wronskian is nonzero the functions $f_1, f_2, \ldots, f_n$ are linearly independent, since the only solution to $c_1f_1+c_2f_2+\cdots+c_nf_n=0$ is then $c_1=c_2=c_3=\cdots=c_n=0$. Hence a sufficient condition for the linear independence of $f_1, f_2, \ldots, f_n$ is that the Wronskian be nonzero. This condition is not *necessary*; that is, there exist sets of linearly independent functions for which the Wronskian *is* zero.

3.15. Show that the function $1, t, t^2$ are linearly independent.

The Wronskian of these three functions (see Problem 3.14) is

$$\begin{vmatrix} 1 & t & t^2 \\ 0 & 1 & 2t \\ 0 & 0 & 2 \end{vmatrix}=2$$

Since the Wronskian is nonzero, the functions are linearly independent.

3.16. Determine a fundamental set for the differential equations:

$$(a)\ \frac{d^3y}{dt^3}+5\frac{d^2y}{dt^2}+8\frac{dy}{dt}+4y=u \qquad (b)\ \frac{d^3y}{dt^3}+4\frac{d^2y}{dt^2}+6\frac{dy}{dt}+4y=u$$

(a) The characteristic polynomial is D^3+5D^2+8D+4, which can be written in factored form as $(D+2)(D+2)(D+1)$. Corresponding to the root $D_1=-1$ there is a solution e^{-t}, and corresponding to the repeated root $D_2 = D_3 = -2$ are the two solutions e^{-2t} and te^{-2t}. The three solutions constitute a fundamental set.

(*b*) The characteristic polynomial is $D^3 + 4D^2 + 6D + 4$, which can be written in factored form as $(D+1+j)(D+1-j)(D+2)$.

A fundamental set is then $e^{(-1-j)t}$, $e^{(-1+j)t}$, and e^{-2t}.

3.17. For the differential equations of Problem 3.16, find fundamental sets different from those found in Problem 3.16.

(*a*) Choose any 3×3 nonzero determinant, say

$$\begin{vmatrix} 1 & 2 & -1 \\ -3 & 2 & 0 \\ 1 & 3 & -2 \end{vmatrix} = -5$$

Using the elements of the first row as coefficients a_{1i} for the fundamental set $e^{-t}, e^{-2t}, te^{-2t}$ found in Problem 3.16, form

$$z_1 = e^{-t} + 2e^{-2t} - te^{-2t}$$

Using the second row, form

$$z_2 = -3e^{-t} + 2e^{-2t}$$

From the third row, form

$$z_3 = e^{-t} + 3e^{-2t} - 2te^{-2t}$$

The functions z_1, z_2, and z_3 constitute a fundamental set.

(*b*) For this equation generate the second fundamental set by letting

$$z_1 = e^{-2t}$$

$$z_2 = \frac{1}{2}e^{(-1+j)t} + \frac{1}{2}e^{(-1-j)t} = e^{-t}\left(\frac{e^{-jt}+e^{jt}}{2}\right)$$

$$= e^{-t}\left(\frac{\cos t - j\sin t + \cos t + j\sin t}{2}\right) = e^{-t}\cos t$$

$$z_3 = \frac{1}{2j}e^{(-1+j)t} - \frac{1}{2j}e^{(-1-j)t} = e^{-t}\left(\frac{e^{t}-e^{-jt}}{2j}\right)$$

$$= e^{-t}\left(\frac{\cos t + j\sin t - \cos t + j\sin t}{2j}\right) = e^{-t}\sin t$$

The coefficient determinant in this case is

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2j} & -\frac{1}{2j} \end{vmatrix} = -\frac{1}{2j}$$

SOLUTION OF LINEAR CONSTANT-COEFFICIENT ORDINARY DIFFERENTIAL EQUATIONS

3.18. Show that any free response $y_a(t) = \sum_{k=1}^{n} c_k y_k(t)$ satisfies $\sum_{i=0}^{n} a_i(d^i y/dt^i) = 0$.

By the definition of a fundamental set, $y_k(t)$, $k = 1, 2, \ldots, n$, satisfies $\sum_{i=0}^{n} a_i(d^i y_k/dt^i) = 0$. Substituting $\sum_{k=1}^{n} c_k y_k(t)$ into this differential equation yields

$$\sum_{i=0}^{n} a_i \frac{d^i}{dt^i}\left[\sum_{k=1}^{n} c_k y_k(t)\right] = \sum_{i=0}^{n}\sum_{k=1}^{n} a_i \frac{d^i}{dt^i}(c_k y_k(t)) = \sum_{k=1}^{n} c_k\left[\sum_{i=0}^{n} a_i \frac{d^i y_k(t)}{dt^i}\right] = 0$$

The last equality is obtained because the term in the brackets is zero for all k.

3.19. Show that the forced response given by Equation (3.14)

$$y_b(t) = \int_0^t w(t-\tau)\left[\sum_{i=0}^{m} b_i \frac{d^i u(\tau)}{d\tau^i}\right] d\tau$$

satisfies the differential equation

$$\sum_{i=0}^{n} a_i \frac{d^i y}{dt^i} = \sum_{i=0}^{m} b_i \frac{d^i u}{dt^i}$$

For simplification, let $r(t) \equiv \sum_{i=0}^{m} b_i(d^i u/dt^i)$. Then $y_b(t) = \int_0^t w(t-\tau) r(\tau)\, d\tau$ and

$$\frac{dy_b}{dt} = \int_0^t \frac{\partial w(t-\tau)}{\partial t} r(\tau)\, d\tau + w(t-\tau) r(\tau)\Big|_{\tau=t} = \int_0^t \frac{\partial w(t-\tau)}{\partial t} r(\tau)\, d\tau + 0 \cdot r(t)$$

Similarly,

$$\frac{d^2 y_b}{dt^2} = \int_0^t \frac{\partial^2 w(t-\tau)}{\partial t^2} r(\tau)\, d\tau, \ldots, \frac{d^{n-1} y_b}{dt^{n-1}} = \int_0^t \frac{\partial^{n-1} w(t-\tau)}{\partial t^{n-1}} r(\tau)\, d\tau$$

since, by Equation (*3.16*),

$$\frac{\partial^i w(t-\tau)}{\partial t^i}\Bigg|_{\tau=t} = \frac{d^i w(t)}{dt^i}\Bigg|_{t=0} = 0 \qquad \text{for } i = 0, 1, 2, \ldots, n-2$$

The nth derivative is

$$\frac{d^n y_b}{dt^n} = \int_0^t \frac{\partial^n w(t-\tau)}{\partial t^n} r(\tau)\, d\tau + \frac{\partial^{n-1} w(t-\tau)}{\partial t^{n-1}}\Bigg|_{\tau=t} \cdot r(t) = \int_0^t \frac{\partial^n w(t-\tau)}{\partial t^n} r(\tau)\, d\tau + r(t)$$

since, by Equation (*3.16*),

$$\frac{\partial^{n-1} w(t-\tau)}{\partial t^{n-1}}\Bigg|_{\tau=t} = \frac{d^{n-1} w(t)}{dt^{n-1}}\Bigg|_{t=0} = 1$$

The summation of the n derivatives is

$$\sum_{i=0}^{n} a_i \frac{d^i y_b}{dt^i} = \int_0^t \left[\sum_{i=0}^{n} a_i \frac{\partial^i w(t-\tau)}{\partial t^i}\right] r(\tau)\, d\tau + r(t)$$

Finally, making the change of variables $t - \tau = \theta$ in the bracketed term yields

$$\sum_{i=0}^{n} a_i \frac{\partial^i w(\theta)}{\partial \theta^i} = \sum_{i=0}^{n} a_i \frac{d^i w(\theta)}{d\theta^i} = 0$$

because $w(\theta)$ is a free response (see Section 3.10 and Problem 3.18). Hence

$$\sum_{i=0}^{n} a_i \frac{d^i y_b}{dt^i} = r(t) \equiv \sum_{i=0}^{m} b_i \frac{d^i u}{dt^i}$$

3.20. Find the free response of the differential equation

$$\frac{d^3y}{dt^3} + 4\frac{d^2y}{dt^2} + 6\frac{dy}{dt} + 4y = u$$

with initial conditions $y(0) = 1$, $(dy/dt)|_{t=0} = 0$, and $(d^2y/dt^2)|_{t=0} = -1$.

From the results of Problems 3.16 and 3.17, a fundamental set for this equation is e^{-2t}, $e^{-t}\cos t$, $e^{-t}\sin t$. Hence the free response can be written as

$$y_\alpha(t) = c_1e^{-2t} + c_2e^{-t}\cos t + c_3e^{-t}\sin t$$

The initial conditions provide the following set of algebraic equations for c_1, c_2, c_3:

$$y_\alpha(0) = c_1 + c_2 = 1 \qquad \left.\frac{dy_\alpha}{dt}\right|_{t=0} = -2c_1 - c_2 + c_3 = 0 \qquad \left.\frac{d^2y_\alpha}{dt^2}\right|_{t=0} = 4c_1 - 2c_3 = -1$$

from which $c_1 = \frac{1}{2}$, $c_2 = \frac{1}{2}$, $c_3 = \frac{3}{2}$. Therefore the free response is

$$y_\alpha(t) = \frac{1}{2}e^{-2t} + \frac{1}{2}e^{-t}\cos t + \frac{3}{2}e^{-t}\sin t$$

3.21. Find the weighting function of the differential equation

$$\frac{d^2y}{dt^2} + 4\frac{dy}{dt} + 4y = 3\frac{du}{dt} + 2u$$

The characteristic equation is $D^2 + 4D + 4 = (D+2)^2 = 0$ with the repeated root $D = -2$. A fundamental set is therefore given by e^{-2t}, te^{-2t}, and the weighting function has the form

$$w(t) = c_1e^{-2t} + c_2te^{-2t}$$

with the initial conditions

$$w(0) = \left[c_1e^{-2t} + c_2te^{-2t}\right]\Big|_{t=0} = c_1 = 0 \qquad \left.\frac{dw}{dt}\right|_{t=0} = \left[-2c_1e^{-2t} + c_2e^{-2t} - 2c_2te^{-2t}\right]\Big|_{t=0} = c_2 = 1$$

Thus $w(t) = te^{-2t}$.

3.22. Find the forced response of the differential equation (Problem 3.21):

$$\frac{d^2y}{dt^2} + 4\frac{dy}{dt} + 4y = 3\frac{du}{dt} + 2u$$

where $u(t) = e^{-3t}$, $t \geq 0$.

The forced response is given by Equation (*3.14*) as

$$y_b(t) = \int_0^t w(t-\tau)\left[3\frac{du}{d\tau} + 2u\right]d\tau = 3\int_0^t w(t-\tau)\frac{du}{d\tau}\,d\tau + 2\int_0^t w(t-\tau)\,u\,d\tau$$

Integrating the first integral by parts,

$$\int_0^t w(t-\tau)\frac{du}{d\tau}\,d\tau = w(t-\tau)u(\tau)\big|_0^t - \int_0^t \frac{\partial w(t-\tau)}{\partial\tau}u\,d\tau$$

$$= w(0)u(t) - w(t)u(0) - \int_0^t \frac{\partial w(t-\tau)}{\partial\tau}u\,d\tau$$

But $w(0) = 0$; hence the forced response can be written as

$$y_b(t) = \int_0^t\left[-3\frac{\partial w(t-\tau)}{\partial\tau} + 2w(t-\tau)\right]u(\tau)\,d\tau - 3w(t)u(0)$$

From Problem 3.21, $w(t-\tau) = (t-\tau)e^{-2(t-\tau)}$; hence

$$\left[-3\frac{\partial w(t-\tau)}{\partial\tau} + 2w(t-\tau)\right] = 3e^{-2(t-\tau)} - 4(t-\tau)e^{-2(t-\tau)}$$

and the forced response is

$$y_b(t) = 3e^{-2t}\int_0^t e^{2\tau}e^{-3\tau}\,d\tau - 4te^{-2t}\int_0^t e^{2\tau}e^{-3\tau}\,d\tau + 4e^{-2t}\int_0^t \tau e^{2\tau}e^{-3\tau}\,d\tau - 3te^{-2t}$$

$$= 7[e^{-2t} - e^{-3t} - te^{-2t}]$$

3.23. Find the output y of a system described by the differential equation

$$\frac{d^2y}{dt^2} + 3\frac{dy}{dt} + 2y = 1 + t$$

with initial conditions $y(0) = 0$ and $(dy/dt)|_{t=0} = 1$.

Let $u_1 \equiv 1$, $u_2 \equiv t$. The response y due to u_1 alone was determined in Example 3.27 as $y_1 = \frac{1}{2}(1 - e^{-2t})$. The free response y_α for the differential equation was found in Example 3.24 to be $y_\alpha = e^{-t} - e^{-2t}$. The forced response due to u_2 is given by Equation (*3.14*). Using the weighting function determined in Example 3.25, the forced response due to u_2 is

$$y_2 = \int_0^t w(t-\tau)u_2(\tau)\,d\tau = \int_0^t\left[e^{-(t-\tau)} - e^{-2(t-\tau)}\right]\tau\,d\tau$$

$$= e^{-t}\int_0^t \tau e^{\tau}\,d\tau - e^{-2t}\int_0^t \tau e^{2\tau}\,d\tau = \frac{1}{4}[4e^{-t} - e^{-2t} + 2t - 3]$$

Thus the forced response is

$$y_b = y_1 + y_2 = \frac{1}{4}[4e^{-t} - 3e^{-2t} + 2t - 1]$$

and the total response is

$$y = y_\alpha + y_b = \frac{1}{4}[8e^{-t} - 7e^{-2t} + 2t - 1]$$

3.24. Find the transient and steady state responses of a system described by the differential equation

$$\frac{d^2y}{dt^2} + 3\frac{dy}{dt} + 2y = 1 + t$$

with the initial conditions $y(0) = 0$ and $(dy/dt)|_{t=0} = 1$.

The total response for this equation was determined in Problem 3.23 as

$$y = \frac{1}{4}[8e^{-t} - 7e^{-2t} + 2t - 1]$$

Since $\lim_{t\to\infty}[\frac{1}{4}(8e^{-t} - 7e^{-2t})] = 0$, the transient response is $y_T = \frac{1}{4}(8e^{-t} - 7e^{-2t})$. The steady state response is $y_{ss} = \frac{1}{4}(2t - 1)$.

SINGULARITY FUNCTIONS

3.25. Evaluate: (*a*) $\int_5^8 t^2\delta(t-6)\,dt$, (*b*) $\int_0^4 \sin t\,\delta(t-7)\,dt$.

(*a*) Using the screening property of the unit impulse function, $\int_5^8 t^2\delta(t-6)\,dt = t^2|_{t=6} = 36$.

(*b*) Since the interval of integration $0 \leq t \leq 4$ does not include the position of the unit impulse function $t = 7$, then $\int_0^4 \sin t\,\delta(t-7)\,dt = 0$.

3.26. Show that the unit step response $y_1(t)$ of a causal linear system described by the convolution integral

$$y(t)=\int_0^t w(t-\tau)u(\tau)\,d\tau$$

is related to the unit impulse response $y_\delta(t)$ by the equation $y_1(t)=\int_0^t y_\delta(\tau)\,d\tau$.

The unit step response is given by $y_1(t)=\int_0^t w(t-\tau)u(\tau)d\tau$, where $\mathbf{1}(t)$ is a unit step function. In Example 3.29 it was shown that $y_\delta(t)=w(t)$. Hence

$$y_1(t)=\int_0^t y_\delta(t-\tau)u(\tau)\,d\tau=\int_0^t y_\delta(t-\tau)\,d\tau$$

Now make the change of variable $\theta=t-\tau$. Then $d\tau=-d\theta$, $\tau=0$ implies $\theta=t$, $\tau=t$ implies $\theta=0$, and the integral becomes

$$y_1(t)=-\int_t^0 y_\delta(\theta)\,d\theta=\int_0^t y_\delta(\theta)\,d\theta$$

3.27. Show that the unit ramp response $y_r(t)$ of a causal linear system described by the convolution integral (see Problem 3.26) is related to the unit impulse response $y_\delta(t)$ and the unit step response $y_1(t)$ by the equation

$$y_r(t)=\int_0^t y_1(\tau')\,d\tau'=\int_0^t\int_0^{\tau'} y_\delta(\theta)\,d\theta\,d\tau'$$

Proceeding as in Problem 3.26 with $w(t-\tau)=y_\delta(t-\tau)$ and τ changed to $t-\tau'$, we get

$$y_r(t)=\int_0^t y_\delta(t-\tau)\tau\,d\tau=\int_0^t (t-\tau')y_\delta(\tau')\,d\tau'=\int_0^t t y_\delta(\tau')\,d\tau'-\int_0^t \tau' y_\delta(\tau')\,d\tau'$$

From Problem 3.26, the first term can be written as $t\int_0^t y_\delta(\tau')\,d\tau'=ty_1(t)$. The second term can be integrated by parts, yielding

$$\int_0^t \tau' y_\delta(\tau')\,d\tau'=\tau' y_1(\tau')\Big|_0^t-\int_0^t y_1(\tau')\,d\tau'$$

where $dy_1(\tau')=y_\delta(\tau')\,d\tau'$. Therefore

$$y_r(t)=ty_1(t)-ty_1(t)+\int_0^t y_1(\tau')\,d\tau'=\int_0^t y_1(\tau')\,d\tau'$$

Again using the result of Problem 3.26, we obtain the required equation.

SECOND-ORDER SYSTEMS

3.28. Show that the weighting function of the second-order differential equation

$$\frac{d^2y}{dt^2}+2\zeta\omega_n\frac{dy}{dt}+\omega_n^2 y=\omega_n^2 u$$

is given by $w(t)=(1/\omega_d)e^{-\alpha t}\sin\omega_d t$, where $\alpha\equiv\zeta\omega_n$, $\omega_d\equiv\omega_n\sqrt{1-\zeta^2}$, $0\le\zeta\le 1$.

The characteristic equation

$$D^2+2\zeta\omega_n D+\omega_n^2=0$$

has the roots

$$D_1=-\zeta\omega_n+j\omega_n\sqrt{1-\zeta^2}=-\alpha+j\omega_d$$

$$D_2=-\zeta\omega_n-j\omega_n\sqrt{1-\zeta^2}=-\alpha-j\omega_d$$

One fundamental set is $y_1=e^{-\alpha t}e^{j\omega_d t}$, $y_2=e^{-\alpha t}e^{-j\omega_d t}$; and the weighting function can be written as

$$w(t)=c_1e^{-\alpha t}e^{-j\omega_d t}+c_2e^{-\alpha t}e^{j\omega_d t}$$

where c_1 and c_2 are, as yet, unknown coefficients. $w(t)$ can be rewritten as

$$\begin{aligned}w(t)&=e^{-\alpha t}[c_1\cos\omega_d t-jc_1\sin\omega_d t+c_2\cos\omega_d t+jc_2\sin\omega_d t]\\&=(c_1+c_2)e^{-\alpha t}\cos\omega_d t+j(c_2-c_1)e^{-\alpha t}\sin\omega_d t\\&=Ae^{-\alpha t}\cos\omega_d t+Be^{-\alpha t}\sin\omega_d t\end{aligned}$$

where $A\equiv c_1+c_2$ and $B\equiv j(c_2-c_1)$ are unknown coefficients determined from the initial conditions given by Equation (*3.16*). That is,

$$w(0)=\left[Ae^{-\alpha t}\cos\omega_d t+Be^{-\alpha t}\sin\omega_d t\right]\Big|_{t=0}=A=0$$

and

$$\left.\frac{dw}{dt}\right|_{t=0}=Be^{-\alpha t}[\omega_d\cos\omega_d t-\alpha\sin\omega_d t]\Big|_{t=0}=B\omega_d=1$$

Hence

$$w(t)=\frac{1}{\omega_d}e^{-\alpha t}\sin\omega_d t$$

3.29. Determine the damping ratio ζ, undamped natural frequency ω_n, damped natural frequency ω_d, damping coefficient α, and time constant τ for the following second-order system:

$$2\frac{d^2y}{dt^2}+4\frac{dy}{dt}+8y=8u$$

Dividing both sides of the equation by 2, $d^2y/dt^2+2(dy/dt)+4y=4u$. Comparing the coefficients of this equation with those of Equation (*3.22*), we obtain $2\zeta\omega_n=2$ and $\omega_n^2=4$ with the solutions $\omega_n=2$ and $\zeta=\frac{1}{2}=0.5$. Now $\omega_d=\omega_n\sqrt{1-\zeta^2}=\sqrt{3}$, $\alpha=\zeta\omega_n=1$, and $\tau=1/\alpha=1$.

3.30. The **overshoot** of a second-order system in response to a unit step input is the difference between the maximum value attained by the output and the steady state solution. Determine the overshoot for the system of Problem 3.29 using the normalized family of curves given in Section 3.14.

Since the damping ratio of this system is $\zeta=0.5$, the normalized curve corresponding to $\zeta=0.5$ is used. This curve has its maximum value (peak) at $\omega_n t=3.4$. From Problem 3.29, $\omega_n=2$; hence the time t_p at which the peak occurs is $t_p=3.4/\omega_n=3.4/2=1.7$ sec. The value attained at this time is 1.17, and the overshoot is $1.17-1.00=0.17$.

STATE VARIABLE REPRESENTATION OF SYSTEMS DESCRIBED BY LINEAR DIFFERENTIAL AND DIFFERENCE EQUATIONS

3.31. Put the differential equation

$$\frac{d^2y}{dt^2}=u$$

with initial conditions $y(0)=1$ and $(dy/dt)|_{t=0}=-1$, into state variable form. Then develop a solution for the resulting vector-matrix equation in the form of Equation (*3.26*) and, from this specify the free response and the forced response. Also, for $u(t)=1$, specify the transient and steady state responses.

Letting $x_1\equiv y$ and $dx_1/dt=x_2$, the state variable representation is $dx_1/dt=x_2$ with $x_1(0)=1$, and $dx_2/dt=u$ with $x_2(0)=-1$. The matrices A and B in the general equation form (*3.25*) are

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

Since $A^k=0$ for $k\ge 2$, the transition matrix is

$$e^{At}=I+At=\begin{bmatrix}1&t\\0&1\end{bmatrix}$$

and the solution of the state variable equation can be written as

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} + \int_0^t \begin{bmatrix} 1 & (t-\tau) \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ u(\tau) \end{bmatrix} d\tau$$

or, after multiplying the matrices in each term,

$$x_1(t) = 1 - t + \int_0^t (t-\tau)u(\tau)\,d\tau$$

$$x_2(t) = -1 + \int_0^t u(\tau)\,d\tau$$

The *free responses* are

$$x_{1\alpha}(t) = 1 - t$$
$$x_{2\alpha}(t) = -1$$

and the *forced responses* are

$$x_{1b}(t) = \int_0^t (t-\tau)u(\tau)\,d\tau$$

$$x_{2b}(t) = \int_0^t u(\tau)\,d\tau$$

For $u(t) = 1$, $x_1(t) = 1 - t + t^2/2$ and $x_2(t) = -1 + t$. The *transient responses* are $x_{1T}(t) = 0$ and $x_{2T}(t) = 0$ and the *steady state responses* are $x_{1ss}(t) = 1 - t + t^2/2$ and $y_{2ss}(t) = -1 + t$.

3.32. Show that the weighting sequence of the difference equation (*3.29*) has the form of Equation (*3.34*).

The technique used to solve this problem is called *variation of parameters*. It is assumed that the forced response of Equation (*3.29*) has the form:

$$y_b(k) = \sum_{j=1}^{n} c_j(k)y_j(k)$$

where $y_1(k),\ldots,y_n(k)$ is a fundamental set of solutions and $c_1(k),\ldots,c_n(k)$ is a set of unknown time-variable parameters to be determined. Since $y_b(0) = 0$ for any forced response of a difference equation, then $c_1(0) = 0,\ldots,c_n(0) = 0$. The parameter $c_j(k+1)$ is written as $c_j(k+1) = c_j(k) + \Delta c_j(k)$. Thus

$$y_b(k+1) = \sum_{j=1}^{n} c_j(k)y_j(k+1) + \left[\sum_{j=1}^{n} \Delta c_j(k)y_j(k+1)\right]$$

The increments $\Delta c_1(k),\ldots,\Delta c_n(k)$ are chosen such that the term in the brackets is zero. This process is then repeated for $y_b(k+2)$ so that

$$y_b(k+2) = \sum_{j=1}^{n} c_j(k)y_j(k+2) + \left[\sum_{j=1}^{n} \Delta c_j(k)y_j(k+2)\right]$$

Again the bracketed term is made zero by choice of the increments $\Delta c_1(k),\ldots,\Delta c_n(k)$. Similar expressions are generated for $y_b(k+3), y_b(k+4),\ldots,y_b(k+n-1)$. Finally,

$$y_b(k+n) = \sum_{j=1}^{n} c_j(k)y_j(k+n) + \left[\sum_{j=1}^{n} \Delta c_j(k)y_j(k+n)\right]$$

In this last expression, the bracketed term is not set to zero. Now the summation in Equation (*3.29*) is

$$\sum_{i=0}^{n} a_i y_b(k+i) = \sum_{j=1}^{n} c_j(k)\sum_{i=0}^{n} a_i y_j(k+i) + a_n\sum_{j=1}^{n} \Delta c_j(k)y_j(k+n) = u(k)$$

Since each element of the fundamental set is a free response, then

$$\sum_{i=0}^{n} a_i y_j(k+i) = 0$$

for each j. A set of n linear algebraic equations in n unknowns has thus been generated:

$$\sum_{j=1}^{n} \Delta c_j(k)y_j(k+1) = 0$$

$$\sum_{j=1}^{n} \Delta c_j(k)y_j(k+2) = 0$$

$$\vdots$$

$$\sum_{j=1}^{n} \Delta c_j(k)y_j(k+n) = \frac{u(k)}{a_n}$$

Now $\Delta c_j(k)$ can be written as

$$\Delta c_j(k) = \frac{M_j(k)}{M(k)}\frac{u(k)}{a_n}$$

where $M(k)$ is the determinant

$$M(k) = \begin{vmatrix} y_1(k+1) & y_2(k+1) & \cdots & y_n(k+1) \\ y_1(k+2) & y_2(k+2) & \cdots & y_n(k+2) \\ \vdots & \vdots & \ddots & \vdots \\ y_1(k+n) & y_2(k+n) & \cdots & y_n(k+n) \end{vmatrix}$$

$M_j(k)$ is the cofactor of the last element in the jth column of this determinant. The parameters $c_1(k),\ldots,c_n(k)$ are thus given by

$$c_j(k) = \sum_{l=0}^{k-1} \Delta c_j(l) = \sum_{l=0}^{k-1} \frac{M_j(l)}{M(l)}\frac{u(l)}{a_n}$$

The forced response then becomes

$$y_b(k) = \sum_{j=1}^{n}\sum_{l=0}^{k-1} \frac{M_j(l)}{M(l)}\frac{u(l)}{a_n}y_j(k)$$
$$= \sum_{l=0}^{k-1}\left[\sum_{j=1}^{n} \frac{M_j(l)}{a_n M(l)}y_j(k)\right]u(l)$$

This last equation is in the form of a convolution sum with weighting sequence

$$w(k-l) = \sum_{j=1}^{n} \frac{M_j(l)}{a_n M(l)}y_j(k)$$

LINEARITY AND SUPERPOSITION

3.33. Using the definition of linearity, Definition 3.21, show that any differential equation of the form:

$$\sum_{i=0}^{n} a_i(t)\frac{d^i y}{dt^i} = u$$

where y is the output and u is the input, is linear.

Let u_1 and u_2 be two arbitrary inputs, and let y_1 and y_2 be the corresponding outputs. Then, with all initial conditions equal to zero,

$$\sum_{i=0}^{n} a_i(t)\frac{d^i y_1}{dt^i} = u_1 \quad \text{and} \quad \sum_{i=0}^{n} a_i(t)\frac{d^i y_2}{dt^i} = u_2$$

Now form

$$c_1 u_1 + c_2 u_2 = c_1\left[\sum_{i=0}^{n} a_i(t)\frac{d^i y_1}{dt^i}\right] + c_2\left[\sum_{i=0}^{n} a_i(t)\frac{d^i y_2}{dt^i}\right]$$
$$= \sum_{i=0}^{n} a_i(t)\frac{d^i(c_1 y_1)}{dt^i} + \sum_{i=0}^{n} a_i(t)\frac{d^i(c_2 y_2)}{dt^i}$$
$$= \sum_{i=0}^{n} a_i(t)\frac{d^i}{dt^i}(c_1 y_1 + c_2 y)$$

Since this equation holds for all c_1 and c_2, the equation is linear.

3.34. Show that a system described by the convolution integral

$$y(t) = \int_{-\infty}^{\infty} w(t,\tau)u(\tau)\,d\tau$$

is linear, y is the output and u the input.

Let u_1 and u_2 be two arbitrary inputs and let

$$y_1 = \int_{-\infty}^{\infty} w(t,\tau)u_1(\tau)\,d\tau \qquad y_2 = \int_{-\infty}^{\infty} w(t,\tau)u_2(\tau)\,d\tau$$

Now let $c_1 u_1 + c_2 u_2$ be a third input and form

$$\int_{-\infty}^{\infty} w(t,\tau)[c_1 u_1(\tau) + c_2 u_2(\tau)]\,d\tau = c_1\int_{-\infty}^{\infty} w(t,\tau)u_1(\tau)\,d\tau + c_2\int_{-\infty}^{\infty} w(t,\tau)u_2(\tau)\,d\tau$$
$$= c_1 y_1 + c_2 y_2$$

Since this relationship holds for all c_1 and c_2, the convolution integral is a linear operation (or transformation).

3.35. Use the Principle of Superposition to determine the output y of Fig. 3-9.

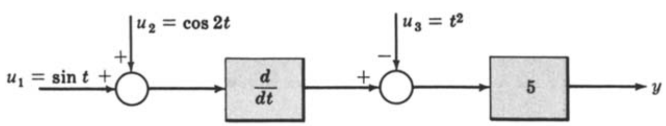


Fig. 3-9

For $u_2 = u_3 = 0$, $y_1 = 5(d/dt)(\sin t) = 5\cos t$. For $u_1 = u_3 = 0$, $y_2 = 5(d/dt)(\cos 2t) = -10\sin 2t$. For $u_1 = u_2 = 0$, $y_3 = -5t^2$. Therefore

$$y = y_1 + y_2 + y_3 = 5(\cos t - 2\sin 2t - t^2)$$

3.36. A linear system is described by the weighting function

$$w(t,\tau) = e^{-|t-\tau|} \qquad \text{for all} \quad t,\tau$$

Suppose the system is excited by an input

$$u(t) = t \qquad \text{for all} \quad t$$

Find the output $y(t)$.

The output is given by the convolution integral (Example 3.38):

$$y(t) = \int_{-\infty}^{\infty} e^{-|t-\tau|}\tau\,d\tau = \int_{-\infty}^{t} e^{-(t-\tau)}\tau\,d\tau + \int_{t}^{\infty} e^{(t-\tau)}\tau\,d\tau$$
$$= e^{-t}\int_{-\infty}^{t} e^{\tau}\tau\,d\tau + e^{t}\int_{t}^{\infty} e^{-\tau}\tau\,d\tau$$
$$= e^{-t}\left[e^{\tau}(\tau-1)|_{-\infty}^{t}\right] + e^{t}\left[e^{-\tau}(-\tau-1)|_{t}^{\infty}\right] = 2t$$

CAUSALITY

3.37. Two systems are defined by the relationships between their inputs and outputs as follows:

System 1: The input is $u(t)$ and at the same instant of time the output is $y(t) = u(t+T)$, $T > 0$.
System 2: The input is $u(t)$ and at the same instant of time the output is $y(t) = u(t-T)$, $T > 0$.

Are either of these systems causal?

In System 1, the output depends only on the input T seconds in the future. Thus it is not causal. An operation of this type is called **prediction**.

In System 2, the output depends only on the input T seconds in the past. Thus it is causal. An operation of this type is called a **time delay**.

Supplementary Problems

3.38. Which of the following terms are first degree in the dependent variable $y = y(t)$? (a) $t^2 y$, (b) $\tan y$, (c) $\cos t$, (d) e^{-y}, (e) te^{-t}.

3.39. Show that a system defined by the equation $y = mu + b$, where y is the output, u is the input, and m and b are nonzero constants, is nonlinear according to Definition 3.21.

3.40. Show that any differential equation of the form

$$\sum_{i=0}^{n} a_i(t)\frac{d^i y}{dt^i} = \sum_{i=0}^{m} b_i(t)\frac{d^i u}{dt^i}$$

satisfies Definition 3.21. (See Example 3.37 and Problem 3.33).

3.41. Show that the functions $\cos t$ and $\sin t$ are linearly independent.

3.42. Show that the functions $\sin nt$ and $\sin kt$, where n and k are integers, are linearly independent if $n \neq k$.

3.43. Show that the functions t and t^2 constitute a fundamental set for the differential equation

$$t^2\frac{d^2 y}{dt^2} - 2t\frac{dy}{dt} + 2y = 0$$

3.44. Find a fundamental set for

$$\frac{d^3 y}{dt^3} + 6\frac{d^2 y}{dt^2} + 21\frac{dy}{dt} + 26y = u$$

Chapter 4

The Laplace Transform and the *z*-Transform

4.1 INTRODUCTION

Several techniques used in solving engineering problems are based on the replacement of functions of a real variable (usually time or distance) by certain frequency-dependent representations, or by functions of a complex variable dependent upon frequency. A typical example is the use of Fourier series to solve certain electrical problems. One such problem consists of finding the current in some part of a linear electrical network in which the input voltage is a periodic or repeating waveform. The periodic voltage may be replaced by its Fourier series representation, and the current produced by each term of the series can then be determined. The total current is the sum of the individual currents (superposition). This technique often results in a substantial savings in computational effort.

Two very important transformation techniques for linear control system analysis are presented in this chapter: the *Laplace transform* and the *z-transform*. The Laplace transform relates time functions to frequency-dependent functions of a complex variable. The z-transform relates time sequences to a different, but related, type of frequency-dependent function. Applications of these mathematical transformations to solving linear constant-coefficient differential and difference equations are also discussed here. Together these methods provide the basis for the analysis and design techniques developed in subsequent chapters.

4.2 THE LAPLACE TRANSFORM

The Laplace transform is defined in the following manner:

Definition 4.1: Let $f(t)$ be a real function of a real variable t defined for $t > 0$. Then

$$\mathscr{L}[f(t)] \equiv F(s) \equiv \lim_{\substack{T\to\infty \\ \epsilon\to 0}} \int_{\epsilon}^{T} f(t)e^{-st}\,dt = \int_{0^+}^{\infty} f(t)e^{-st}\,dt \qquad 0<\epsilon<T$$

is called the **Laplace transform** of $f(t)$. s is a complex variable defined by $s \equiv \sigma + j\omega$, where σ and ω are real variables* and $j=\sqrt{-1}$.

Note that the lower limit on the integral is $t = \epsilon > 0$. This definition of the lower limit is sometimes useful in dealing with functions that are discontinuous at $t = 0$. When *explicit* use is made of this limit, it will be abbreviated $t = \lim_{\epsilon\to 0}\epsilon \equiv 0^+$, as shown above in the integral on the right.

The real variable t always denotes *time*.

Definition 4.2: If $f(t)$ is defined and single-valued for $t > 0$ and $F(\sigma)$ is absolutely convergent for some real number σ_0, that is,

$$\int_{0^+}^{\infty} |f(t)|e^{-\sigma_0 t}\,dt = \lim_{\substack{T\to\infty \\ \epsilon\to 0}} \int_{\epsilon}^{T} |f(t)|e^{-\sigma_0 t}\,dt < +\infty \qquad 0<\epsilon<T$$

then $f(t)$ is **Laplace transformable** for $\mathrm{Re}(s) > \sigma_0$.

*The real part σ of a complex variable s is often written as $\mathrm{Re}(s)$ (the real part of s) and the imaginary part ω as $\mathrm{Im}(s)$ (the imaginary part of s). Parentheses are placed around s only when there is a possibility of confusion.

EXAMPLE 4.1. The function e^{-t} is Laplace transformable since

$$\int_{0^+}^{\infty} |e^{-t}|e^{-\sigma_0 t}\,dt = \int_{0^+}^{\infty} e^{-(1+\sigma_0)t}\,dt = \frac{1}{-(1+\sigma_0)}e^{-(1+\sigma_0)t}\Big|_{0^+}^{\infty} = \frac{1}{1+\sigma_0} < +\infty$$

if $1+\sigma_0 > 0$ or $\sigma_0 > -1$.

EXAMPLE 4.2. The Laplace transform of e^{-t} is

$$\mathscr{L}[e^{-t}] = \int_{0^+}^{\infty} e^{-t}e^{-st}\,dt = \frac{-1}{(s+1)}e^{-(s+1)t}\Big|_{0^+}^{\infty} = \frac{1}{s+1} \qquad \text{for } \mathrm{Re}(s) > -1$$

4.3 THE INVERSE LAPLACE TRANSFORM

The Laplace transform transforms a problem from the real variable time domain into the complex variable s-domain. After a solution of the transformed problem has been obtained in terms of s, it is necessary to "invert" this transform to obtain the time domain solution. The transformation from the s-domain into the t-domain is called the *inverse Laplace transform*.

Definition 4.3: Let $F(s)$ be the Laplace transform of a function $f(t)$, $t > 0$. The contour integral

$$\mathscr{L}^{-1}[F(s)] \equiv f(t) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F(s)e^{st}\,ds$$

where $j=\sqrt{-1}$ and $c > \sigma_0$ (σ_0 as given in Definition 4.2), is called the **inverse Laplace transform** of $F(s)$.

It is seldom necessary in practice to perform the contour integration defined in Definition 4.3. For applications of the Laplace transform in this book, it is never necessary. A simple technique for evaluating the inverse transform for most control system problems is presented in Section 4.8.

4.4 SOME PROPERTIES OF THE LAPLACE TRANSFORM AND ITS INVERSE

The Laplace transform and its inverse have several important properties which can be used advantageously in the solution of linear constant-coefficient differential equations. They are:

1. The Laplace transform is a *linear transformation* between functions defined in the t-domain and functions defined in the s-domain. That is, if $F_1(s)$ and $F_2(s)$ are the Laplace transforms of $f_1(t)$ and $f_2(t)$, respectively, then $a_1F_1(s)+a_2F_2(s)$ is the Laplace transform of $a_1f_1(t)+a_2f_2(t)$, where a_1 and a_2 are arbitrary constants.
2. The inverse Laplace transform is a *linear transformation* between functions defined in the s-domain and functions defined in the t-domain. That is, if $f_1(t)$ and $f_2(t)$ are the inverse Laplace transforms of $F_1(s)$ and $F_2(s)$, respectively, then $b_1f_1(t)+b_2f_2(t)$ is the inverse Laplace transform of $b_1F_1(s)+b_2F_2(s)$, where b_1 and b_2 are arbitrary constants.
3. The Laplace transform of the *derivative* df/dt of a function $f(t)$ whose Laplace transform is $F(s)$ is

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

where $f(0^+)$ is the initial value of $f(t)$, evaluated as the one-sided limit of $f(t)$ as t approaches zero from positive values.

4. The Laplace transform of the *integral* $\int_0^t f(\tau)\,d\tau$ of a function $f(t)$ whose Laplace transform is $F(s)$ is

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

5. The initial value $f(0^+)$ of the function $f(t)$ whose Laplace transform is $F(s)$ is

$$f(0^+) = \lim_{t\to 0} f(t) = \lim_{s\to\infty} sF(s) \qquad t>0$$

This relation is called the *Initial Value Theorem.*

6. The final value $f(\infty)$ of the function $f(t)$ whose Laplace transform is $F(s)$ is

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

if $\lim_{t\to\infty} f(t)$ exists. This relation is called the *Final Value Theorem.*

7. The Laplace transform of a function $f(t/a)$ (*Time Scaling*) is

$$\mathscr{L}\left[f\left(\frac{t}{a}\right)\right] = aF(as)$$

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

8. The inverse Laplace transform of the function $F(s/a)$ (*Frequency Scaling*) is

$$\mathscr{L}^{-1}\left[F\left(\frac{s}{a}\right)\right] = af(at)$$

where $\mathscr{L}^{-1}[F(s)] = f(t)$.

9. The Laplace transform of the function $f(t-T)$ (*Time Delay*), where $T>0$ and $f(t-T)=0$ for $t \le T$, is

$$\mathscr{L}[f(t-T)] = e^{-sT}F(s)$$

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

10. The Laplace transform of the function $e^{-at}f(t)$ is given by

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

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

11. The Laplace transform of the *product of two functions* $f_1(t)$ and $f_2(t)$ is given by the *complex convolution integral*

$$\mathscr{L}[f_1(t)f_2(t)] = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F_1(\omega)F_2(s-\omega)\,d\omega$$

where $F_1(s) = \mathscr{L}[f_1(t)]$, $F_2(s) = \mathscr{L}[f_2(t)]$.

12. The inverse Laplace transform of the *product of the two transforms* $F_1(s)$ and $F_2(s)$ is given by the *convolution integrals*

$$\mathscr{L}^{-1}[F_1(s)F_2(s)] = \int_{0^+}^t f_1(\tau)f_2(t-\tau)\,d\tau = \int_{0^+}^t f_2(\tau)f_1(t-\tau)\,d\tau$$

where $\mathscr{L}^{-1}[F_1(s)] = f_1(t)$, $\mathscr{L}^{-1}[F_2(s)] = f_2(t)$.

EXAMPLE 4.3. The Laplace transforms of the functions e^{-t} and e^{-2t} are $\mathscr{L}[e^{-t}] = 1/(s+1)$, $\mathscr{L}[e^{-2t}] = 1/(s+2)$. Then, by Property 1,

$$\mathscr{L}[3e^{-t} - e^{-2t}] = 3\mathscr{L}[e^{-t}] - \mathscr{L}[e^{-2t}] = \frac{3}{s+1} - \frac{1}{s+2} = \frac{2s+5}{s^2+3s+2}$$

EXAMPLE 4.4. The inverse Laplace transforms of the functions $1/(s+1)$ and $1/(s+3)$ are

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

Then, by Property 2,

$$\mathscr{L}^{-1}\left[\frac{2}{s+1} - \frac{4}{s+3}\right] = 2\mathscr{L}^{-1}\left[\frac{1}{s+1}\right] - 4\mathscr{L}^{-1}\left[\frac{1}{s+3}\right] = 2e^{-t} - 4e^{-3t}$$

EXAMPLE 4.5. The Laplace transform of $(d/dt)(e^{-t})$ can be determined by application of Property 3. Since $\mathscr{L}[e^{-t}] = 1/(s+1)$ and $\lim_{t\to 0} e^{-t} = 1$, then

$$\mathscr{L}\left[\frac{d}{dt}(e^{-t})\right] = s\left(\frac{1}{s+1}\right) - 1 = \frac{-1}{s+1}$$

EXAMPLE 4.6. The Laplace transform of $\int_0^t e^{-\tau}\,d\tau$ can be determined by application of Property 4. Since $\mathscr{L}[e^{-t}] = 1/(s+1)$, then

$$\mathscr{L}\left[\int_0^t e^{-\tau}\,d\tau\right] = \frac{1}{s}\left(\frac{1}{s+1}\right) = \frac{1}{s(s+1)}$$

EXAMPLE 4.7. The Laplace transform of e^{-3t} is $\mathscr{L}[e^{-3t}] = 1/(s+3)$. The initial value of e^{-3t} can be determined by the Initial Value Theorem as

$$\lim_{t\to 0} e^{-3t} = \lim_{s\to\infty} s\left(\frac{1}{s+3}\right) = 1$$

EXAMPLE 4.8. The Laplace transform of the function $(1-e^{-t})$ is $1/s(s+1)$. The final value of this function can be determined from the Final Value Theorem as

$$\lim_{t\to\infty}(1-e^{-t}) = \lim_{s\to 0}\frac{s}{s(s+1)} = 1$$

EXAMPLE 4.9. The Laplace transform of e^{-t} is $1/(s+1)$. The Laplace transform of e^{-3t} can be determined by application of Property 7 (Time Scaling), where $a = \frac{1}{3}$:

$$\mathscr{L}[e^{-3t}] = \frac{1}{3}\left[\frac{1}{(\frac{1}{3}s+1)}\right] = \frac{1}{s+3}$$

EXAMPLE 4.10. The inverse transform of $1/(s+1)$ is e^{-t}. The inverse transform of $1/(\frac{1}{3}s+1)$ can be determined by application of Property 8 (Frequency Scaling):

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

EXAMPLE 4.11. The Laplace transform of the function e^{-t} is $1/(s+1)$. The Laplace transform of the function defined as

$$f(t) = \begin{cases} e^{-(t-2)} & t>2 \\ 0 & t \le 2 \end{cases}$$

can be determined by Property 9, with $T = 2$:

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

EXAMPLE 4.12. The Laplace transform of $\cos t$ is $s/(s^2+1)$. The Laplace transform of $e^{-2t}\cos t$ can be determined from Property 10 with $a = 2$:

$$\mathscr{L}[e^{-2t}\cos t] = \frac{s+2}{(s+2)^2+1} = \frac{s+2}{s^2+4s+5}$$

EXAMPLE 4.13. The Laplace transform of the product $e^{-2t}\cos t$ can be determined by application of Property 11 (Complex Convolution). That is, since $\mathscr{L}[e^{-2t}] = 1/(s+2)$ and $\mathscr{L}[\cos t] = s/(s^2+1)$, then

$$\mathscr{L}[e^{-2t}\cos t] = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty}\left(\frac{\omega}{\omega^2+1}\right)\left(\frac{1}{s-\omega+2}\right)d\omega = \frac{s+2}{s^2+4s+5}$$

The details of this contour integration are not carried out here because they are too complicated (see, e.g., Reference [1]) and unnecessary. The Laplace transform of $e^{-2t}\cos t$ was very simply determined in Example 4.12 using Property 10. There are, however, many instances in more advanced treatments of automatic control in which complex convolution can be used effectively.

EXAMPLE 4.14. The inverse Laplace transform of the function $F(s) = s/(s+1)(s^2+1)$ can be determined by application of Property 12. Since $\mathscr{L}^{-1}[1/(s+1)] = e^{-t}$ and $\mathscr{L}^{-1}[s/(s^2+1)] = \cos t$, then

$$\mathscr{L}^{-1}\left[\left(\frac{1}{s+1}\right)\left(\frac{s}{s^2+1}\right)\right] = \int_{0^+}^{t} e^{-(t-\tau)}\cos\tau\, d\tau = e^{-t}\int_{0^+}^{t} e^{\tau}\cos\tau\, d\tau = \tfrac{1}{2}(\cos t + \sin t - e^{-t})$$

4.5 SHORT TABLE OF LAPLACE TRANSFORMS

Table 4.1 is a short table of Laplace transforms. It is not complete, but when used in conjunction with the properties of the Laplace transform described in Section 4.4 and the partial fraction expansion techniques described in Section 4.7, it is adequate to handle all of the problems in this book. A more complete table of Laplace transform pairs is found in Appendix A.

TABLE 4.1

Time Function		Laplace Transform
Unit Impulse	$\delta(t)$	1
Unit Step	$\mathbf{1}(t)$	$\frac{1}{s}$
Unit Ramp	t	$\frac{1}{s^2}$
Polynomial	t^n	$\frac{n!}{s^{n+1}}$
Exponential	e^{-at}	$\frac{1}{s+a}$
Sine Wave	$\sin\omega t$	$\frac{\omega}{s^2+\omega^2}$
Cosine Wave	$\cos\omega t$	$\frac{s}{s^2+\omega^2}$
Damped Sine Wave	$e^{-at}\sin\omega t$	$\frac{\omega}{(s+a)^2+\omega^2}$
Damped Cosine Wave	$e^{-at}\cos\omega t$	$\frac{s+a}{(s+a)^2+\omega^2}$

Table 4.1 can be used to find both Laplace transforms and inverse Laplace transforms. To find the Laplace transform of a time function which can be represented by some combination of the elementary functions given in Table 4.1, the appropriate transforms are chosen from the table and are combined using the properties in Section 4.4.

EXAMPLE 4.15. The Laplace transform of the function $f(t) = e^{-4t} + \sin(t-2) + t^2e^{-2t}$ is determined as follows. The Laplace transforms of e^{-4t}, $\sin t$, and t^2 are given in the table as

$$\mathscr{L}[e^{-4t}] = \frac{1}{s+4} \qquad \mathscr{L}[\sin t] = \frac{1}{s^2+1} \qquad \mathscr{L}[t^2] = \frac{2}{s^3}$$

Application of Properties 9 and 10, respectively, yields

$$\mathscr{L}[\sin(t-2)] = \frac{e^{-2s}}{s^2+1} \qquad \mathscr{L}[t^2e^{-2t}] = \frac{2}{(s+2)^3}$$

Then Property 1 (Linearity) gives

$$\mathscr{L}[f(t)] = \frac{1}{s+4} + \frac{e^{-2s}}{s^2+1} + \frac{2}{(s+2)^3}$$

To find the inverse of the transform of a combination of those in Table 4.1, the corresponding time functions (inverse transforms) are determined from the table and combined appropriately using the properties in Section 4.4.

EXAMPLE 4.16. The inverse Laplace transform of $F(s) = [(s+2)/s^2+4]\cdot e^{-s}$ can be determined as follows. $F(s)$ is first rewritten as

$$F(s) = \frac{se^{-s}}{s^2+4} + \frac{2e^{-s}}{s^2+4}$$

Now

$$\mathscr{L}^{-1}\left[\frac{s}{s^2+4}\right] = \cos 2t \qquad \mathscr{L}^{-1}\left[\frac{2}{s^2+4}\right] = \sin 2t$$

Application of Property 9 for $t > 1$ yields

$$\mathscr{L}^{-1}\left[\frac{se^{-s}}{s^2+4}\right] = \cos 2(t-1) \qquad \mathscr{L}^{-1}\left[\frac{2e^{-s}}{s^2+4}\right] = \sin 2(t-1)$$

Then Property 2 (Linearity) gives

$$\begin{aligned}\mathscr{L}^{-1}[F(s)] &= \cos 2(t-1) + \sin 2(t-1) && t > 1\\ &= 0 && t \le 1\end{aligned}$$

4.6 APPLICATION OF LAPLACE TRANSFORMS TO THE SOLUTION OF LINEAR CONSTANT-COEFFICIENT DIFFERENTIAL EQUATIONS

The application of Laplace transforms to the solution of linear constant-coefficient differential equations is of major importance in linear control system problems. Two classes of equations of general interest are treated in this section. The first of these has the form:

$$\sum_{i=0}^{n} a_i \frac{d^iy}{dt^i} = u \qquad (4.1)$$

where y is the output, u is the input, the coefficients $a_0, a_1, \ldots, a_{n-1}$, are constants, and $a_n = 1$. The initial conditions for this equation are written as

$$\left.\frac{d^ky}{dt^k}\right|_{t=0^+} \equiv y_0^k \qquad k = 0, 1, \ldots, n-1$$

where y_0^k are constants. The Laplace transform of Equation (4.1) is given by

$$\sum_{i=0}^{n}\left[a_i\left(s^iY(s) - \sum_{k=0}^{i-1} s^{i-1-k}y_0^k\right)\right] = U(s) \qquad (4.2)$$

and the transform of the output is

$$Y(s)=\frac{U(s)}{\sum_{i=0}^{n} a_i s^i}+\frac{\sum_{i=0}^{n}\sum_{k=0}^{i-1} a_i s^{i-1-k} y_0^k}{\sum_{i=0}^{n} a_i s^i} \tag{4.3}$$

Note that the right side of Equation (*4.3*) is the sum of two terms: a term dependent only on the input transform, and a term dependent only on the initial conditions. In addition, note that the denominator of both terms in Equation (*4.3*), that is,

$$\sum_{i=0}^{n} a_i s^i = s^n + a_{n-1}s^{n-1} + \cdots + a_i s + a_0$$

is the *characteristic polynomial* of Equation (*4.1*) (see Section 3.6).

The time solution $y(t)$ of Equation (*4.1*) is the inverse Laplace transform of $Y(s)$, that is,

$$y(t)=\mathscr{L}^{-1}\left[\frac{U(s)}{\sum_{i=0}^{n} a_i s^i}\right]+\mathscr{L}^{-1}\left[\frac{\sum_{i=0}^{n}\sum_{k=0}^{i-1} a_i s^{i-1-k} y_0^k}{\sum_{i=0}^{n} a_i s^i}\right] \tag{4.4}$$

The first term on the right is the *forced response* and the second term is the *free response* of the system represented by Equation (*4.1*).

Direct substitution into Equations (*4.2*), (*4.3*), and (*4.4*) yields the transform of the differential equation, the solution transform $Y(s)$, or the time solution $y(t)$, respectively. But it is often easier to directly apply the properties of Section 4.4 to determine these quantities, especially when the order of the differential equation is low.

EXAMPLE 4.17. The Laplace transform of the differential equation

$$\frac{d^2y}{dt^2}+3\frac{dy}{dt}+2y=\mathbf{1}(t)=\text{unit step}$$

with initial conditions $y(0^+)=-1$ and $(dy/dt)|_{t=0^+}=2$ can be written directly from Equation (*4.2*) by first identifying n, a_i, and y_0^k: $n=2$, $y_0^0=-1$, $y_0^1=2$, $a_0=2$, $a_1=3$, $a_2=1$. Substitution of these values into Equation (*4.2*) yields

$$2Y+3(sY+1)+1(s^2Y+s-2)=\frac{1}{s} \qquad \text{or} \qquad (s^2+3s+2)Y=\frac{-(s^2+s-1)}{s}$$

It should be noted that when $i=0$ in Equation (*4.2*), the summation interior to the brackets is, by definition,

$$\left.\sum_{k=0}^{i-1}\right|_{i=0}=\sum_{k=0}^{k=-1}=0$$

The Laplace transform of the differential equation can also be determined in the following manner. The transform of d^2y/dt^2 is given by

$$\mathscr{L}\left[\frac{d^2y}{dt^2}\right]=s^2Y(s)-sy(0^+)-\left.\frac{dy}{dt}\right|_{t=0^+}$$

This equation is a direct consequence of Property 3, Section 4.4 (see Problem 4.17). With this information the transform of the differential equation can be determined by applying Property 1 (Linearity) of Section 4.4; that is,

$$\mathscr{L}\left[\frac{d^2y}{dt^2}+3\frac{dy}{dt}+2y\right]=\mathscr{L}\left[\frac{d^2y}{dt^2}\right]+\mathscr{L}\left[3\frac{dy}{dt}\right]+\mathscr{L}[2y]=(s^2+3s+2)Y+s+1=\mathscr{L}[\mathbf{1}(t)]=\frac{1}{s}$$

The output transform $Y(s)$ is determined by rearranging the previous equation and is

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

The output time solution $y(t)$ is the inverse transform of $Y(s)$. A method for determining the inverse transform of functions like $Y(s)$ above is presented in Sections 4.7 and 4.8.

Now consider constant-coefficient equations of the form:

$$\sum_{i=0}^{n} a_i\frac{d^iy}{dt^i}=\sum_{i=0}^{m} b_i\frac{d^iu}{dt^i} \tag{4.5}$$

where y is the output, u is the input, $a_n=1$, and $m\le n$. The Laplace transform of Equation (*4.5*) is given by

$$\sum_{i=0}^{n}\left[a_i\left(s^iY(s)-\sum_{k=0}^{i-1}s^{i-1-k}y_0^k\right)\right]=\sum_{i=0}^{m}\left[b_i\left(s^iU(s)-\sum_{k=0}^{i-1}s^{i-1-k}u_0^k\right)\right] \tag{4.6}$$

where $u_0^k=(d^ku/dt^k)|_{t=0^+}$. The output transform $Y(s)$ is

$$Y(s)=\left[\frac{\sum_{i=0}^{m} b_i s^i}{\sum_{i=0}^{n} a_i s^i}\right]U(s)-\frac{\sum_{i=0}^{m}\sum_{k=0}^{i-1} b_i s^{i-1-k}u_0^k}{\sum_{i=0}^{n} a_i s^i}+\frac{\sum_{i=0}^{n}\sum_{k=0}^{i-1} a_i s^{i-1-k}y_0^k}{\sum_{i=0}^{n} a_i s^i} \tag{4.7}$$

The time solution $y(t)$ is the inverse Laplace transform of $Y(s)$:

$$y(t)=\mathscr{L}^{-1}\left[\frac{\sum_{i=0}^{m} b_i s^i}{\sum_{i=0}^{n} a_i s^i}U(s)-\frac{\sum_{i=0}^{m}\sum_{k=0}^{i-1} b_i s^{i-1-k}u_0^k}{\sum_{i=0}^{n} a_i s^i}\right]+\mathscr{L}^{-1}\left[\frac{\sum_{i=0}^{n}\sum_{k=0}^{i-1} a_i s^{i-1-k}y_0^k}{\sum_{i=0}^{n} a_i s^i}\right] \tag{4.8}$$

The first term on the right is the *forced response*, and the second term is the *free response* of a system represented by Equation (*4.5*).

Note that the Laplace transform $Y(s)$ of the output $y(t)$ consists of ratios of polynomials in the complex variable s. Such ratios are generally called **rational (algebraic) functions**. If all initial conditions in Eq. (*4.8*) are zero and $U(s)=1$, (*4.8*) gives the *unit impulse response*. The denominator of each term in (*4.8*) is the *characteristic polynomial* of the system.

For problems in which initial conditions are not specified on $y(t)$ but on some other parameter of the system (such as the initial voltage across a capacitor not appearing at the output), y_0^k, $k=0,1,\ldots,n-1$, must be derived using the available information. For systems represented in the form of Equation (*4.5*), that is, including derivative terms in u, computation of y_0^k will also depend on u_0^k. Problem 4.38 illustrates these points.

The restriction $n\ge m$ in Equation (*4.5*) is based on the fact that real systems have a *smoothing* effect on their input. By a smoothing effect, it is meant that variations in the input are made less pronounced (at least no more pronounced) by the action of the system on the input. Since a differentiator generates the slope of a time function, it accentuates the variations of the function. An integrator, on the other hand, sums the area under the curve of a time function over an interval of time and thus averages (smooths) the variations of the function.

In Equation (*4.5*), the output y is related to the input u by an operation which includes m differentiations and n integrations of the input. Hence, in order that there be a smoothing effect (at least no accentuation of the variations) between the input and the output, there must be more (at least as many) integrations than differentiations; that is, $n\ge m$.

EXAMPLE 4.18. A certain system is described by the differential equation

$$\frac{d^2y}{dt^2} = \frac{du}{dt}, \qquad y(0^+) = \left.\frac{dy}{dt}\right|_{t=0^+} = 0$$

where the input u is graphed in Fig. 4-1. The corresponding functions du/dt and

$$y(t) = \int_{0^+}^{t} \int_{0^+}^{\theta} \frac{du}{d\alpha}\, d\alpha\, d\theta = \int_{0^+}^{t} u(\theta)\, d\theta$$

are also shown. Note from these graphs that differentiation of u accentuates the variations in u while integration smooths them.

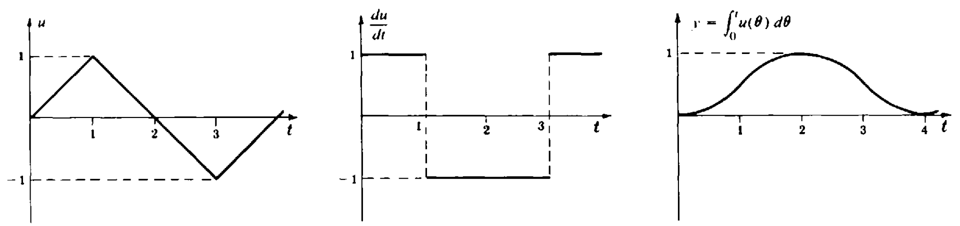


Fig. 4-1

EXAMPLE 4.19. Consider a system described by the differential equation

$$\frac{d^2y}{dt^2} + 3\frac{dy}{dt} + 2y = \frac{du}{dt} + 3u$$

with initial conditions $y_0^0 = 1$, $y_0^1 = 0$. If the input is given by $u(t) = e^{-4t}$, then the Laplace transform of the output $y(t)$ can be obtained by direct application of Equation (4.7) by first identifying m, n, a_i, b_i and u_0^0: $n = 2$, $a_0 = 2$, $a_1 = 3$, $a_2 = 1$, $m = 1$, $u_0^0 = \lim_{t \to 0} e^{-4t} = 1$, $b_0 = 3$, $b_1 = 1$. Substitution of these values into Equation (4.7) yields

$$Y(s) = \left(\frac{s+3}{s^2+3s+2}\right)\left(\frac{1}{s+4}\right) + \frac{s+3}{s^2+3s+2} - \frac{1}{s^2+3s+2}$$

This transform can also be obtained by direct application of Properties 1 and 3 of Section 4.4 to the differential equation, as was done in Example 4.17.

The linear constant-coefficient vector-matrix differential equations discussed in Section 3.15 also can be solved by Laplace transform techniques, as illustrated in the following example.

EXAMPLE 4.20. Consider the vector-matrix differential equation of Problem 3.31:

$$\frac{d\mathbf{x}}{dt} = A\mathbf{x} + \mathbf{b}u$$

where

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \qquad A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad \mathbf{x}(0) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

and with $u = \mathbf{1}(t)$, the unit step function. The Laplace transform of the vector-matrix form of this equation is

$$s\mathbf{X}(s) - \mathbf{x}(0) = A\mathbf{X}(s) + \frac{1}{s}\mathbf{b}$$

where $\mathbf{X}(s)$ is the vector Laplace transform whose components are the Laplace transforms of the components of $\mathbf{x}(t)$. This can be rewritten as

$$[sI - A]\mathbf{X}(s) = \mathbf{x}(0) + \frac{1}{s}\mathbf{b}$$

where I is the *identity* or *unit* matrix. The Laplace transform of the solution vector $\mathbf{x}(t)$ can thus be written as

$$\mathbf{X}(s) = [sI - A]^{-1}\mathbf{x}(0) + \frac{1}{s}[sI - A]^{-1}\mathbf{b}$$

where $[\cdot]^{-1}$ represents the *inverse* of the matrix. Since

$$sI - A = \begin{bmatrix} s & -1 \\ 0 & s \end{bmatrix}$$

then

$$[sI - A]^{-1} = \frac{1}{s^2}\begin{bmatrix} s & 1 \\ 0 & s \end{bmatrix}$$

Substituting for $[sI - A]^{-1}$, $\mathbf{x}(0)$, and $\mathbf{b}$ gives

$$\mathbf{X}(s) = \begin{bmatrix} \dfrac{s-1}{s^2} \\ -\dfrac{1}{s} \end{bmatrix} + \begin{bmatrix} \dfrac{1}{s^3} \\ \dfrac{1}{s^2} \end{bmatrix}$$

where the first term is the Laplace transform of the *free response*, and the second term is the Laplace transform of the *forced response*. Using Table 4.1, the Laplace transform of these vectors can be inverted term by term, providing the solution vector:

$$\mathbf{x}(t) = \begin{bmatrix} \mathbf{1}(t) - t + t^2/2 \\ -\mathbf{1}(t) + t \end{bmatrix}$$

4.7 PARTIAL FRACTION EXPANSIONS

In Section 4.6 it was shown that the Laplace transforms encountered in the solution of linear constant-coefficient differential equations are rational functions of s (i.e., ratios of polynomials in s). In this section an important representation of rational functions, the partial fraction expansion, is presented. It will be shown in the next section that this representation greatly simplifies the inversion of the Laplace transform of a rational function.

Consider the rational function

$$F(s) = \frac{\sum_{i=0}^{m} b_i s^i}{\sum_{i=0}^{n} a_i s^i} \tag{4.9}$$

where $a_n = 1$ and $n \geq m$. By the fundamental theorem of algebra, the denominator polynomial equation

$$\sum_{i=0}^{n} a_i s^i = 0$$

has n roots. Some of these roots may be repeated.

EXAMPLE 4.21. The polynomial $s^3 + 5s^2 + 8s + 4$ has three roots: -2, -2, -1. -2 is a repeated root.

Suppose the denominator polynomial equation above has n_1 roots equal to $-p_1$, n_2 roots equal to $-p_2, \ldots, n_r$ roots equal to $-p_r$, where $\sum_{i=1}^{r} n_i = n$. Then

$$\sum_{i=0}^{n} a_i s^i = \prod_{i=1}^{r} (s + p_i)^{n_i}$$

The rational function $F(s)$ can then be written as

$$F(s)=\frac{\sum_{i=0}^{m} b_i s^i}{\prod_{i=1}^{r}(s+p_i)^{n_i}}$$

The **partial fraction expansion** representation of the rational function $F(s)$ is

$$F(s)=b_n+\sum_{i=1}^{r}\sum_{k=1}^{n_i}\frac{c_{ik}}{(s+p_i)^k} \tag{4.10a}$$

where $b_n=0$ unless $m=n$. The coefficients c_{ik} are given by

$$c_{ik}=\frac{1}{(n_i-k)!}\frac{d^{n_i-k}}{ds^{n_i-k}}\left[(s+p_i)^{n_i}F(s)\right]\Bigg|_{s=-p_i} \tag{4.10b}$$

The particular coefficients c_{i1}, $i=1,2,\dots,r$, are called the **residues** of $F(s)$ at $-p_i$, $i=1,2,\dots,r$. If none of the roots are repeated, then

$$F(s)=b_n+\sum_{i=1}^{n}\frac{c_{i1}}{s+p_i} \tag{4.11a}$$

where

$$c_{i1}=(s+p_i)F(s)|_{s=-p_i} \tag{4.11b}$$

EXAMPLE 4.22. Consider the rational function

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

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

$$F(s)=b_2+\frac{c_{11}}{s+1}+\frac{c_{21}}{s+2}$$

The numerator coefficient of s^2 is $b_2=1$. The coefficients c_{11} and c_{21} are determined from Equation (*4.11b*) as

$$c_{11}=(s+1)F(s)|_{s=-1}=\frac{s^2+2s+2}{s+2}\Bigg|_{s=-1}=1$$

$$c_{21}=(s+2)F(s)|_{s=-2}=\frac{s^2+2s+2}{s+1}\Bigg|_{s=-2}=-2$$

Hence

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

EXAMPLE 4.23. Consider the rational function

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

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

$$F(s)=b_3+\frac{c_{11}}{s+1}+\frac{c_{12}}{(s+1)^2}+\frac{c_{21}}{s+2}$$

The coefficients $b_3, c_{11}, c_{12}, c_{21}$ are given by

$$b_3=0$$

$$c_{11}=\frac{d}{ds}(s+1)^2F(s)\Bigg|_{s=-1}=\frac{d}{ds}\frac{1}{s+2}\Bigg|_{s=-1}=-1$$

$$c_{12}=(s+1)^2F(s)|_{s=-1}=\frac{1}{s+2}\Bigg|_{s=-1}=1$$

$$c_{21}=(s+2)F(s)|_{s=-2}=1$$

Thus

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

4.8 INVERSE TRANSFORMS USING PARTIAL FRACTION EXPANSIONS

In Section 4.6 it was shown that the solution to a linear constant-coefficient ordinary differential equation can be determined by finding the inverse Laplace transform of a rational function. The general form of this operation can be written using Equation (*4.10*) as

$$\mathscr{L}^{-1}\left[\frac{\sum_{i=0}^{m} b_i s^i}{\sum_{i=0}^{n} a_i s^i}\right]=\mathscr{L}^{-1}\left[b_n+\sum_{i=0}^{r}\sum_{k=1}^{n_i}\frac{c_{ik}}{(s+p_i)^k}\right]=b_n\delta(t)+\sum_{i=1}^{r}\sum_{k=1}^{n_i}\frac{c_{ik}}{(k-1)!}t^{k-1}e^{-p_it} \tag{4.12}$$

where $\delta(t)$ is the unit impulse function and $b_n=0$ unless $m=n$. We remark that the rightmost term in Equation (*4.12*) is the general form of the *unit impulse response* for Equation (*4.5*).

EXAMPLE 4.24. The inverse Laplace transform of the function

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

is given by

$$\mathscr{L}^{-1}\left[\frac{s^2+2s+2}{(s+1)(s+2)}\right]=\mathscr{L}^{-1}\left[1+\frac{1}{s+1}-\frac{2}{s+2}\right]=\mathscr{L}^{-1}[1]+\mathscr{L}^{-1}\left[\frac{1}{s+1}\right]-\mathscr{L}^{-1}\left[\frac{2}{s+2}\right]=\delta(t)+e^{-t}-2e^{-2t}$$

which is the unit impulse response for the differential equation:

$$\frac{d^2y}{dt^2}+3\frac{dy}{dt}+2y=\frac{d^2u}{dt^2}+2\frac{du}{dt}+2u$$

EXAMPLE 4.25. The inverse Laplace transform of the function

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

is given by

$$\mathscr{L}^{-1}\left[\frac{1}{(s+1)^2(s+2)}\right]=\mathscr{L}^{-1}\left[-\frac{1}{s+1}+\frac{1}{(s+1)^2}+\frac{1}{s+2}\right]$$

$$=-\mathscr{L}^{-1}\left[\frac{1}{s+1}\right]+\mathscr{L}^{-1}\left[\frac{1}{(s+1)^2}\right]+\mathscr{L}^{-1}\left[\frac{1}{s+2}\right]=-e^{-t}+te^{-t}+e^{-2t}$$

4.9 THE z-TRANSFORM

The z-transform is used to describe signals and components in discrete-time control systems. It is defined as follows:

Definition 4.4: Let $\{f(k)\}$ denote a real-valued sequence $f(0), f(1), f(2), \ldots$, or equivalently, $f(k)$ for $k = 0, 1, 2, \ldots$. Then

$$\mathcal{Z}\{f(k)\} \equiv F(z) = \sum_{k=0}^{\infty} f(k) z^{-k}$$

is called the **z-transform** of $\{f(k)\}$. z is a *complex variable defined* by $z \equiv \mu + j\nu$, where μ and ν are real variables and $j = \sqrt{-1}$.

Remark 1: The kth term of the series in this definition is always the kth element of the sequence being z-transformed times z^{-k}.

Remark 2: Often $\{f(k)\}$ is defined over equally spaced times: $0, T, 2T, \ldots, kT, \ldots$, where T is a fixed time interval. The resulting sequence is thus sometimes written as $\{f(kT)\}$, or $f(kT)$, $k = 0, 1, 2, \ldots$, and $\mathcal{Z}\{f(kT)\} = \sum_{k=0}^{\infty} f(kT) z^{-k}$, but the dependence on T is usually suppressed. We use the variable arguments k and kT interchangeably for time sequences, when there is no ambiguity.

Remark 3: The z-transform is defined differently by some authors, as the transformation $z \equiv e^{sT}$, which amounts to a simple exponential change of variables between the complex variable $z = \mu + j\nu$ and the complex variable $s = \sigma + j\omega$ in the Laplace transform domain, where T is the sampling period of the discrete-time system. This definition implies a sequence $\{f(k)\}$, or $\{f(kT)\}$, obtained by ideal sampling (sometimes called *impulse sampling*) of a continuous signal $f(t)$ at uniformly spaced times kT, $k = 1, 2, \ldots$. Then $s = \ln z/T$, and our definition above, that is, $F(z) = \sum_{k=0}^{\infty} f(kT) z^{-k}$, follows directly from the result of Problem 4.39. Additional relationships between continuous and discrete-time systems, particularly for systems with both types of elements, are developed further beginning in Chapter 6.

EXAMPLE 4.26. The series $F(z) = 1 + z^{-1} + z^{-2} + \cdots + z^{-k} + \cdots$, is the z-transform of the sequence $f(k) = 1$, $k = 0, 1, 2, \ldots$.

If the rate of increase in the terms of the sequence $\{f(k)\}$ is no greater than that of some geometric series as k approaches infinity, then $\{f(k)\}$ is said to be of **exponential order**. In this case, there exists a real number r such that

$$F(z) = \sum_{k=0}^{\infty} f(k) z^{-k}$$

converges for $|z| > r$. r is called the **radius of convergence** of the series. If r is finite, the sequence $\{f(k)\}$ is called **z-transformable**.

EXAMPLE 4.27. The series in Example 4.26 is convergent for $|z| > 1$ and can be written in closed form as the function

$$F(z) = \frac{1}{1 - z^{-1}} \qquad \text{for} \quad |z| > 1$$

If $F(z)$ exists for $|z| > r$, then the integral and derivative of $F(z)$ can be evaluated by operating term by term on the defining series. In addition, if

$$F_1(z) = \sum_{k=0}^{\infty} f_1(k) z^{-k} \qquad \text{for} \quad |z| > r_1$$

and

$$F_2(z) = \sum_{k=0}^{\infty} f_2(k) z^{-k} \qquad \text{for} \quad |z| > r_2$$

then

$$F_1(z) F_2(z) = \sum_{k=0}^{\infty} \left(\sum_{i=0}^{k} f_1(k-i) f_2(i) \right) z^{-k} = \sum_{k=0}^{\infty} \left(\sum_{i=0}^{k} f_2(k-i) f_1(i) \right) z^{-k}$$

The term $\sum_{i=0}^{k} f_1(k-i) f_2(i)$ is called the **convolution sum** of the sequences $\{f_1(k)\}$ and $\{f_2(k)\}$, where the radius of convergence is the larger of the two radii of convergence of $F_1(z)$ and $F_2(z)$.

EXAMPLE 4.28. The derivative of the series in Example 4.26 is

$$\frac{dF}{dz} = -z^{-2} - 2z^{-3} - \cdots - kz^{-(k+1)} - \cdots$$

The indefinite integral is

$$\int F(z)\, dz = z + \ln z - z^{-1} + \cdots$$

EXAMPLE 4.29. The z-transform of the sequence $f_2(k) = 2^k$, $k = 0, 1, 2, \ldots$, is

$$F_2(z) = 1 + 2z^{-1} + 4z^{-2} + \cdots$$

for $|z| > 2$. Let $F_1(z)$ be the z-transform in Example 4.26. Then

$$F_1(z) F_2(z) = \sum_{k=0}^{\infty} \left(\sum_{i=0}^{k} 1^{k-i} 2^i \right) z^{-k} = \sum_{k=0}^{\infty} (2^{k+1} - 1) z^{-k} \qquad \text{for} \quad |z| > 2$$

The z-transform of the sequence $f(k) = A^k$, $k = 0, 1, 2, \ldots$, where A is any finite complex number, is

$$\mathcal{Z}\{A^k\} = 1 + Az^{-1} + A^2 z^{-2} + \cdots$$
$$= \frac{1}{1 - Az^{-1}} = \frac{z}{z - A}$$

where the radius of convergence $r = |A|$. By suitable choice of A, the most common types of sequences can be defined and their z-transforms generated from this relationship.

EXAMPLE 4.30. For $A = e^{\alpha T}$, the sequence $\{A^k\}$ is the sampled exponential $1, e^{\alpha T}, e^{2\alpha T}, \ldots$, and the z-transform of this sequence is

$$\mathcal{Z}\{e^{\alpha kT}\} = \frac{1}{1 - e^{\alpha T} z^{-1}}$$

with radius of convergence $r = |e^{\alpha T}|$.

The z-transform has an inverse very similar to that of the Laplace transform.

Definition 4.5: Let C be a circle centered at the origin of the z-plane and with radius greater than the radius of convergence of the z-transform $F(z)$. Then

$$\mathcal{Z}^{-1}[F(z)] \equiv \{f(k)\} = \frac{1}{2\pi j} \int_c F(z) z^{k-1}\, dz$$

is the **inverse of the z-transform** $F(z)$.

In practice, it is seldom necessary to perform the contour integration in Definition 4.5. For applications of z-transforms in this book, it is never necessary. The properties and techniques in the remainder of this section are adequate to evaluate the inverse transform for most discrete-time control system problems.

Following are some additional **properties of the z-transform and its inverse** which can be used advantageously in discrete-time control system problems.

1. The z-transform and its inverse are *linear transformations* between the time domain and the z-domain. Therefore, if $\{f_1(k)\}$ and $F_1(z)$ are a transform pair and if $\{f_2(k)\}$ and $F_2(z)$ are a

transform pair, then $\{a_1 f_1(k) + a_2 f_2(k)\}$ and $a_1 F_1(z) + a_2 F_2(z)$ are a transform pair for any a_1 and a_2.

2. If $F(z)$ is the z-transform of the sequence $f(0), f(1), f(2), \ldots$, then
$$z^n F(z) - z^n f(0) - z^{n-1} f(1) - \cdots - z f(n-1)$$
is the z-transform of the sequence $f(n), f(n+1), f(n+2), \ldots$, for $n > 1$. Note that the kth element of this sequence is $f(n+k)$.

3. The initial term $f(0)$ of the sequence $\{f(k)\}$ whose z-transform is $F(z)$ is
$$f(0) = \lim_{z \to \infty} (1 - z^{-1}) F(z) = F(\infty)$$
This relation is called the **Initial Value Theorem**

4. Let the sequence $\{f(k)\}$ have the z-transform $F(z)$, with radius of convergence ≤ 1. Then the final value $f(\infty)$ of the sequence is given by
$$f(\infty) = \lim_{z \to 1} (1 - z^{-1}) F(z)$$
if the limit exists. This relation is called the **Final Value Theorem.**

5. The inverse z-transform of the function $F(z/a)$ (**Frequency Scaling**) is
$$\mathcal{Z}^{-1}\left[F\left(\frac{z}{a}\right)\right] = a^k f(k) \qquad k = 0, 1, 2, \ldots$$
where $\mathcal{Z}^{-1}[F(z)] = \{f(k)\}$.

6. If $F(z)$ is the z-transform of the sequence $f(0), f(1), f(2), \ldots$, then $z^{-1}F(z)$ is the z-transform of the time-shifted sequence $f(-1), f(0), f(1), \ldots$, where $f(-1) \equiv 0$. This relationship is called the **Shift Theorem.**

EXAMPLE 4.31. The z-transforms of the sequences $\{(\frac{1}{3})^k\}$ and $\{(\frac{1}{2})^k\}$ are $\mathcal{Z}\{(\frac{1}{3})^k\} = z/(z - \frac{1}{3})$, and $\mathcal{Z}\{(\frac{1}{2})^k\} = z/(z - \frac{1}{2})$. Then, by Property 1,
$$\mathcal{Z}\left\{3\left(\frac{1}{2}\right)^k - \left(\frac{1}{3}\right)^k\right\} = \frac{3z}{z - \frac{1}{2}} - \frac{z}{z - \frac{1}{3}}$$
$$= \frac{2z^2 - \dfrac{z}{2}}{z^2 - \dfrac{5z}{6} + \dfrac{1}{6}}$$

EXAMPLE 4.32. The inverse z-transforms of the functions $z/(z + \frac{1}{2})$ and $z/(z - \frac{1}{4})$ are
$$\mathcal{Z}^{-1}\left[\frac{z}{z + \frac{1}{2}}\right] = \left\{\left(-\frac{1}{2}\right)^k\right\}, \qquad \mathcal{Z}^{-1}\left[\frac{z}{z - \frac{1}{4}}\right] = \left\{\left(\frac{1}{4}\right)^k\right\}$$
Then, by Property 1,
$$\mathcal{Z}^{-1}\left[2\frac{z}{z + \frac{1}{2}} - 4\frac{z}{z - \frac{1}{4}}\right] = 2\mathcal{Z}^{-1}\left[\frac{z}{z + \frac{1}{2}}\right] - 4\mathcal{Z}^{-1}\left[\frac{z}{z - \frac{1}{4}}\right] = \left\{2\left(-\frac{1}{2}\right)^k - 4\left(\frac{1}{4}\right)^k\right\}$$

EXAMPLE 4.33. The z-transform of the sequence $1, \frac{1}{2}, \frac{1}{4}, \ldots, (\frac{1}{2})^k, \ldots$ is $z/(z - \frac{1}{2})$. Then, by Property 2, the z-transform of the sequence $\frac{1}{4}, \frac{1}{8}, \ldots, (\frac{1}{2})^{k+2}, \ldots$ is
$$z^2\left(\frac{z}{z - \frac{1}{2}}\right) - z^2 - \frac{z}{2} = \frac{1}{4}\frac{z}{z - \frac{1}{2}}$$

EXAMPLE 4.34. The z-transform of $\{(\frac{1}{4})^k\}$ is $z/(z - \frac{1}{4})$. The initial value of $\{(\frac{1}{4})^k\}$ can be determined by the Initial Value Theorem as
$$\lim_{k \to 0} \{(\tfrac{1}{4})^k\} = \lim_{z \to \infty} (1 - z^{-1})\left(\frac{z}{z - \frac{1}{4}}\right) = 1$$

EXAMPLE 4.35. The z-transform of the sequence $\{1 - (\frac{1}{4})^k\}$ is $\frac{3}{4}z/(z^2 - \frac{5z}{4} + \frac{1}{4})$. The final value of this sequence can be determined from the Final Value Theorem as
$$\lim_{k \to \infty}\left\{1 - \left(\frac{1}{4}\right)^k\right\} = \lim_{z \to 1}(1 - z^{-1})\left(\frac{\frac{3}{4}z}{z^2 - \dfrac{5z}{4} + \frac{1}{4}}\right) = 1$$

EXAMPLE 4.36. The inverse z-transform of $z/(z - \frac{1}{4})$ is $\{(\frac{1}{4})^k\}$. The inverse transform of $(\frac{z}{2})/(\frac{z}{2} - \frac{1}{4})$ is $\{2^k(\frac{1}{4})^k\} = \{(\frac{1}{2})^k\}$.

For the types of control problems considered in this book, the resulting z-transforms are rational algebraic functions of z, as illustrated below, and there are two practical methods for inverting them. The first is a numerical technique, generating a power series expansion by long division.

Suppose the z-transform has the form:
$$F(z) = \frac{b_n z^n + b_{n-1} z^{n-1} + \cdots + b_1 z + b_0}{a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0}$$
It is easily rewritten in powers of z^{-1} as
$$F(z) = \frac{b_n + b_{n-1} z^{-1} + \cdots + b_0 z^{-n}}{a_n + a_{n-1} z^{-1} + \cdots + a_0 z^{-n}}$$
by multiplying each term by z^{-n}. Then, by long division, the denominator is divided into the numerator, yielding a polynomial in z^{-1} of the form:
$$F(z) = \frac{b_n}{a_n} + \frac{1}{a_n}\left(b_{n-1} - \frac{b_n a_{n-1}}{a_n}\right) z^{-1} + \cdots$$

EXAMPLE 4.37. The z-transform $z/(z - \frac{1}{2})$ is rewritten as $1/(1 - z^{-1}/2)$ which, by long division, has the form:
$$\frac{1}{1 - z^{-1}/2} = 1 + \left(\frac{1}{2}\right) z^{-1} + \left(\frac{1}{2}\right)^2 z^{-2} + \cdots$$

For the second inversion method, $F(z)$ is first expanded into a special partial fraction form and each term is inverted using the properties previously discussed.

Table 4.2 is a short table of z-transform pairs. When used in conjunction with the properties of the z-transform described earlier, and the partial fraction expansion techniques described in Section 4.7, it

Table 4.2

kth Term of the Time Sequence	z-Transform
1 at k, 0 elsewhere (Kronecker delta sequence)	z^{-k}
1 (unit step sequence)	$\dfrac{z}{z-1}$
k (unit ramp sequence)	$\dfrac{z}{(z-1)^2}$
A^k (for complex numbers A)	$\dfrac{z}{z-A}$
kA^k	$\dfrac{Az}{(z-A)^2}$
$\dfrac{(k+1)(k+2)\cdots(k+n-1)}{(n-1)!}A^k$	$\dfrac{z^n}{(z-A)^n}$

is adequate to handle all the problems in this book. A more complete table of z-transform pairs is given in Appendix B.

The final transform pair in Table 4.2 can be used to generate many other useful transforms by proper choice of A and use of Property 1.

The following examples illustrate how z-transforms can be inverted using the partial fraction expansion method.

EXAMPLE 4.38. To invert the z-transform $F(z)=1/(z+1)(z+2)$, we form the partial fraction expansion of $F(z)/z$:

$$\frac{F(z)}{z}=\frac{1}{z(z+1)(z+2)}=\frac{\frac{1}{2}}{z}+\frac{-1}{z+1}+\frac{\frac{1}{2}}{z+2}$$

Then

$$F(z)=\frac{1}{2}-\frac{z}{z+1}+\frac{1}{2}\frac{z}{z+2}$$

which can be inverted term by term as

$$f(0)=0$$

$$f(k)=-(-1)^k+\frac{1}{2}(-2)^k \qquad \text{for all} \quad k\geq 1$$

EXAMPLE 4.39. To invert $F(z)=1/(z+1)^2(z+2)$, we take the partial fraction expansion of $F(z)/z$:

$$\frac{F(z)}{z}=\frac{\frac{1}{2}}{z}+\frac{0}{z+1}+\frac{-1}{(z+1)^2}+\frac{-\frac{1}{2}}{z+2}$$

Then

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

$$f(k)=-k(-1)^k-\frac{1}{2}(-2)^k \qquad \text{for all} \quad k\geq 1 \text{ and } f(0)=0$$

EXAMPLE 4.40. Using the last transform pair in Table 4.2, the z-transform of the sequence $\{k^2/2\}$ can be generated by noting the following transform pairs:

$$\left\{\frac{(k+1)(k+2)}{2!}\right\}\leftrightarrow\frac{z^3}{(z-1)^3}$$

$$\{k\}\leftrightarrow\frac{z}{(z-1)^2}$$

$$\{1\}\leftrightarrow\frac{z}{z-1}$$

Since

$$\frac{(k+1)(k+2)}{2!}=\frac{k^2}{2}+\frac{3}{2}k+1$$

then, by Property 1,

$$\mathcal{Z}\left\{\frac{k^2}{2}\right\}=\frac{z^3}{(z-1)^3}-\frac{3}{2}\frac{z}{(z-1)^2}-\frac{z}{z-1}=\frac{z(z+1)/2}{(z-1)^3}$$

Linear nth-order constant-coefficient difference equations can be solved using z-transform methods by a procedure virtually identical to that used to solve differential equations by Laplace transform methods. This is illustrated step by step in the following example.

EXAMPLE 4.41. The difference equation

$$x(k+2)+\frac{5}{6}x(k+1)+\frac{1}{6}x(k)=1$$

with initial conditions $x(0)=0$ and $x(1)=1$ is z-transformed by applying Properties 1 and 2. By Property 1 (Linearity):

$$\mathcal{Z}\left\{x(k+2)+\frac{5}{6}x(k+1)+\frac{1}{6}x(k)\right\}=\mathcal{Z}\{x(k+2)\}+\frac{5}{6}\mathcal{Z}\{x(k+1)\}+\frac{1}{6}\mathcal{Z}\{x(k)\}=\mathcal{Z}\{1\}$$

By Property 2, if $\mathcal{Z}[x(k)]\equiv X(z)$, then

$$\mathcal{Z}\{x(k+1)\}=zX(z)-zx(0)=zX(z)$$

$$\mathcal{Z}\{x(k+2)\}=z^2X(z)-z^2x(0)-zx(1)=z^2X(z)-z$$

From Table 4.2, the z-transform of the unit step sequence is

$$\mathcal{Z}\{1\}=\frac{z}{z-1}$$

Direct substitution of these expressions into the transformed equation then gives

$$\left(z^2+\frac{5}{6}z+\frac{1}{6}\right)X(z)-z=\frac{z}{z-1}$$

Thus the z-transform $X(z)$ of the solution sequence $x(k)$ is

$$X(z)=\frac{z}{z^2+\frac{5}{6}z+\frac{1}{6}}+\frac{z}{(z-1)\left(z^2+\frac{5}{6}z+\frac{1}{6}\right)}=X_a(z)+X_b(z)$$

Note that the first term $X_a(z)$ results from the initial conditions and the second term $X_b(z)$ results from the input sequence. Thus the inverse of the first term is the *free response*, and the inverse of the second term is the *forced response*. The first term can be inverted by forming the partial fraction expansion

$$\frac{X_a(z)}{z}=\frac{1}{z^2+\frac{5}{6}z+\frac{1}{6}}=-\frac{6}{z+\frac{1}{2}}+\frac{6}{z+\frac{1}{3}}$$

From this,

$$X_a(z)=-6\frac{z}{z+\frac{1}{2}}+6\frac{z}{z+\frac{1}{3}}$$

and from Table 4.2, the inverse of $X_a(z)$ (the free response) is

$$x_a(k)=-6\left(-\frac{1}{2}\right)^k+6\left(-\frac{1}{3}\right)^k \qquad k=0,1,2,\ldots$$

Similarly, to find the forced response, the following partial fraction expansion is formed:

$$\frac{X_b(z)}{z}=\frac{1}{(z-1)\left(z+\frac{1}{2}\right)\left(z+\frac{1}{3}\right)}=\frac{\frac{1}{2}}{z-1}+\frac{4}{z+\frac{1}{2}}+\frac{-\frac{9}{2}}{z+\frac{1}{3}}$$

Thus

$$X_b(z)=\frac{\frac{1}{2}z}{z-1}+\frac{4z}{z+\frac{1}{2}}-\frac{\frac{9}{2}z}{z+\frac{1}{3}}$$

Then, from Table 4.2, the inverse of $X_b(z)$ (the forced response) is

$$x_b(k)=\frac{1}{2}+4\left(-\frac{1}{2}\right)^k-\frac{9}{2}\left(-\frac{1}{3}\right)^k \qquad k=0,1,2,\ldots$$

The total response $x(k)$ is

$$x(k) = x_a(k) + x_b(k) = \frac{1}{2} - 2\left(-\frac{1}{2}\right)^k + \frac{3}{2}\left(-\frac{1}{3}\right)^k \qquad k = 0,1,2,\ldots$$

Linear constant-coefficient vector-matrix difference equations presented in Section 3.17 can also be solved by z-transform techniques, as illustrated in the following example.

EXAMPLE 4.42. Consider the difference equation of Example 4.41 written in state variable form (see Example 3.36):

$$x_1(k+1) = x_2(k)$$

$$x_2(k+1) = -\frac{5}{6}x_2(k) - \frac{1}{6}x_1(k) + 1$$

with initial conditions $x_1(0) = 0$ and $x_2(0) = 1$. In vector-matrix form, these two equations are written as

$$\mathbf{x}(k+1) = A\mathbf{x}(k) + \mathbf{b}u(k)$$

where

$$A = \begin{bmatrix} 0 & 1 \\ -\frac{1}{6} & -\frac{5}{6} \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad \mathbf{x}(k) = \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} \qquad \mathbf{x}(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$u(k) = 1$. The z-transform of the vector-matrix form of the equation is

$$z\mathbf{X}(z) - z\mathbf{x}(0) = A\mathbf{X}(z) + \frac{z}{z-1}\mathbf{b}$$

where $\mathbf{X}(z)$ is a vector-valued z-transform whose components are the z-transforms of the corresponding components of the state vector $\mathbf{x}(k)$. This transformed equation can be rewritten as

$$(zI - A)\mathbf{X}(z) = z\mathbf{x}(0) + \frac{z}{z-1}\mathbf{b}$$

where I is the identity or unit matrix. The z-transform of the solution vector $\mathbf{x}(k)$ is

$$\mathbf{X}(z) = z(zI - A)^{-1}\mathbf{x}(0) + \frac{z}{z-1}(zI - A)^{-1}\mathbf{b}$$

where $(\cdot)^{-1}$ represents the *inverse* of the matrix. Since

$$zI - A = \begin{bmatrix} z & -1 \\ \frac{1}{6} & z + \frac{5}{6} \end{bmatrix}$$

then

$$(zI - A)^{-1} = \frac{1}{z^2 + \frac{5}{6}z + \frac{1}{6}}\begin{bmatrix} z + \frac{5}{6} & 1 \\ -\frac{1}{6} & z \end{bmatrix}$$

Substituting for $(zI - A)^{-1}$, $\mathbf{x}(0)$, and $\mathbf{b}$ yields

$$\mathbf{X}(z) = \begin{bmatrix} \dfrac{z}{z^2 + \frac{5}{6}z + \frac{1}{6}} \\[2ex] \dfrac{z^2}{z^2 + \frac{5}{6}z + \frac{1}{6}} \end{bmatrix} + \begin{bmatrix} \dfrac{z}{(z-1)\left(z^2 + \frac{5}{6}z + \frac{1}{6}\right)} \\[2ex] \dfrac{z^2}{(z-1)\left(z^2 + \frac{5}{6}z + \frac{1}{6}\right)} \end{bmatrix}$$

where the first term is the z-transform of the free response and the second of the forced response. Using the partial fraction expansion method and Table 4.2, the inverse of this z-transform is

$$\mathbf{x}(k) = \begin{bmatrix} \frac{1}{2} - 2\left(-\frac{1}{2}\right)^k + \frac{3}{2}\left(-\frac{1}{3}\right)^k \\ \frac{1}{2} + \left(-\frac{1}{2}\right)^k - \frac{1}{2}\left(-\frac{1}{3}\right)^k \end{bmatrix} \qquad k = 0,1,2,\ldots$$

4.10 DETERMINING ROOTS OF POLYNOMIALS

The results of Sections 4.7, 4.8, and 4.9 indicate that finding the solution of linear constant-coefficient differential and difference equations by transform techniques generally requires the determination of the roots of polynomial equations of the form:

$$Q_n(s) = \sum_{i=0}^{n} a_i s^i = 0$$

where $a_n = 1$, $a_0, a_1, \ldots, a_{n-1}$, are real constants and s is replaced by z for z-transform polynomials.

The roots of a second-order polynomial equation $s^2 + a_1 s + a_0 = 0$ can be obtained directly from the quadratic formula and are given by

$$s_1 = \frac{-a_1 + \sqrt{a_1^2 - 4a_0}}{2} \qquad s_2 = \frac{-a_1 - \sqrt{a_1^2 - 4a_0}}{2}$$

But for higher-order polynomials such analytical expressions do not, in general, exist. The expressions that do exist are very complicated. Fortunately, numerical techniques exist for determining these roots.

To aid in the use of these numerical techniques, the following general properties of $Q_n(s)$ are given:

1. If a repeated root of multiplicity n_i is counted as n_i roots, then $Q_n(s) = 0$ has exactly n roots (Fundamental theorem of algebra).
2. If $Q_n(s)$ is divided by the factor $s + p$ until a constant remainder is obtained, the remainder is $Q_n(-p)$.
3. $s + p$ is a factor of $Q_n(s)$ if and only if $Q_n(-p) = 0$ [$-p$ is a root of $Q_n(s) = 0$].
4. If $\sigma + j\omega$ (σ, ω real) is a root of $Q_n(s) = 0$, then $\sigma - j\omega$ is also a root of $Q_n(s) = 0$.
5. If n is odd, $Q_n(s) = 0$ has at least one real root.
6. The number of positive real roots of $Q_n(s) = 0$ cannot exceed the number of variations in sign of the coefficients in the polynomial $Q_n(s)$, and the number of negative roots cannot exceed the number of variations in sign of the coefficients of $Q_n(-s)$ (Descartes' rule of signs).

Of the techniques available for iteratively determining the roots of a polynomial equation (or equivalently the factors of the polynomial), some can determine only real roots and others both real and complex roots. Both types are presented below.

Horner's Method

This method can be used to determine the *real roots* of the polynomial equation $Q_n(s) = 0$. The steps to be followed are:

1. Evaluate $Q_n(s)$ for real integer values of s, $s = 0, \pm 1, \pm 2, \ldots$, until for two consecutive integer values such as k_0 and $k_0 + 1$, $Q_n(k_0)$ and $Q_n(k_0 + 1)$ have opposite signs. A real root then lies between k_0 and $k_0 + 1$. Assume this root is positive without loss of generality. A first approximation of the root is taken to be k_0. Corrections to this approximation are obtained in the remaining steps.
2. Determine a sequence of polynomials $Q_n^l(s)$ using the recursive relationship

$$Q_n^{l+1}(s) = Q_n^l\left(\frac{k_l}{10^l} + s\right) = \sum_{i=0}^{n} a_i^{l+1} s^i \qquad l = 0,1,2,\ldots \tag{4.13}$$

   where $Q_n^0(s) = Q_n(s)$, and the values k_l, $l = 1, 2, \ldots$, are generated in Step 3.
3. Determine the integer k_l at each iteration by evaluating $Q_n^l(s)$ for real values of s given by $s = k/10^l$, $k = 0, 1, 2, \ldots, 9$. For two consecutive values of k, say k_l and k_{l+1}, the values $Q_n(k_l/10^l)$ and $Q_n(k_{l+1}/10^l)$ have opposite signs.

4. Repeat until the desired accuracy of the root has been achieved. The approximation of the real root after the Nth iteration is given by

$$s_N = \sum_{l=0}^{N} \frac{k_l}{10^l} \tag{4.14}$$

Each iteration increases the accuracy of the approximation by one decimal place.

Newton's Method

This method can determine *real roots* of the polynomial equation $Q_n(s) = 0$. The steps to be followed are:

1. Obtain a first approximation s_0 of a root by making an "educated" guess, or by a technique such as the one in Step 1 of Horner's method.
2. Generate a sequence of improved approximations until the desired accuracy is achieved by the recursive relationship

$$s_{l+1} = s_l - \left. \frac{Q_n(s)}{\dfrac{d}{ds}[Q_n(s)]} \right|_{s=s_l}$$

which can be rewritten as

$$s_{l+1} = \frac{\displaystyle\sum_{i=0}^{n} (i-1) a_i s_l^i}{\displaystyle\sum_{i=1}^{n} i a_i s_l^{i-1}} \tag{4.15}$$

where $l = 0, 1, 2, \ldots$.

This method does not provide a measure of the accuracy of the approximation. Indeed, there is no guarantee that the approximations converge to the correct value.

Lin-Bairstow Method

This method can determine both *real and complex roots* of the polynomial equation $Q_n(s) = 0$. More exactly, this method determines quadratic factors of $Q_n(s)$ from which two roots can be determined by the quadratic formula. The roots can, of course, be either real or complex. The steps to be followed are:

1. Obtain a first approximation of a quadratic factor

$$s^2 + \alpha_1 s + \alpha_0$$

of $Q_n(s) = \sum_{i=0}^{n} a_i s^i$ by some method, perhaps an "educated" guess. Corrections to this approximation are obtained in the remaining steps.
2. Generate a set of constants $b_{n-2}, b_{n-3}, \ldots, b_0, b_{-1}, b_{-2}$ from the recursive relationship

$$b_{i-2} = a_i - \alpha_1 b_{i-1} - \alpha_0 b_i$$

where $b_n = b_{n-1} = 0$, and $i = n, n-1, \ldots, 1, 0$.
3. Generate a set of constants $c_{n-2}, c_{n-3}, \ldots, c_1, c_0$ from the recursive relationship

$$c_{i-1} = b_{i-1} - \alpha_1 c_i - \alpha_0 c_{i+1}$$

where $c_n = c_{n-1} = 0$, and $i = n, n-1, \ldots, 1$.

4. Solve the two simultaneous equations

$$c_0 \Delta\alpha_1 + c_1 \Delta\alpha_0 = b_{-1}$$
$$(-\alpha_1 c_0 - \alpha_0 c_1)\Delta\alpha_1 + c_0 \Delta\alpha_0 = b_{-2}$$

for $\Delta\alpha_1$ and $\Delta\alpha_0$. The new approximation of the quadratic factor is

$$s^2 + (\alpha_1 + \Delta\alpha_1)s + (\alpha_0 + \Delta\alpha_0)$$

5. Repeat Steps 1 through 4 for the quadratic factor obtained in Step 4, until successive approximations are sufficiently close.

This method does not provide a measure of the accuracy of the approximation. Indeed, there is no guarantee that the approximations converge to the correct value.

Root-Locus Method

This method can be used to determine both real and complex roots of the polynomial equation $Q_n(s) = 0$. The technique is discussed in Chapter 13.

4.11 COMPLEX PLANE: POLE-ZERO MAPS

The rational functions $F(s)$ for continuous systems can be rewritten as

$$F(s) = \frac{b_m \displaystyle\sum_{i=0}^{m} (b_i/b_m) s^i}{\displaystyle\sum_{i=0}^{n} a_i s^i} = \frac{b_m \displaystyle\prod_{i=1}^{m} (s+z_i)}{\displaystyle\prod_{i=0}^{n} (s+p_i)}$$

where the terms $s + z_i$ are factors of the numerator polynomial and the terms $s + p_i$ are factors of the denominator polynomial, with $a_n \equiv 1$. If s is replaced by z, $F(z)$ represents a system function for discrete-time systems.

Definition 4.6: Those values of the complex variable s for which $|F(s)|$ [absolute value of $F(s)$] is zero are called the **zeros** of $F(s)$.

Definition 4.7: Those values of the complex variable s for which $|F(s)|$ is infinite are called the **poles** of $F(s)$.

EXAMPLE 4.43. Let $F(s)$ be given by

$$F(s) = \frac{2s^2 - 2s - 4}{s^3 + 5s^2 + 8s + 6}$$

which can be rewritten as

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

$F(s)$ has *finite zeros* at $s = -1$ and $s = 2$, and a *zero* at $s = \infty$. $F(s)$ has *finite poles* at $s = -3$, $s = -1 - j$, and $s = -1 + j$.

Poles and **zeros** are complex numbers determined by two real variables, one representing the real part and the other the imaginary part of the complex number. A pole or zero can therefore be represented as a point in rectangular coordinates. The *abscissa* of this point represents the real part and the *ordinate* the imaginary part. In the s-plane, the abscissa is called the **σ-axis** and the ordinate the **jω-axis**. In the z-plane, the abscissa is called the **μ-axis** and the ordinate the **jν-axis**. The planes defined by these coordinate systems are generally called the **complex plane** (s-plane or z-plane). That half of the complex plane in which $\text{Re}(s) < 0$ or $\text{Re}(z) < 0$ is called the **left half of the s-plane** or **z-plane** (LHP), and that half in which $\text{Re}(s) > 0$ or $\text{Re}(z) > 0$ is called the **right half of the s-plane** or **z-plane** (RHP). That portion of the z-plane in which $|z| < 1$ is called (the interior of) the **unit circle** in the z-plane.

The location of a pole in the complex plane is denoted symbolically by a cross ($\times$), and the location of a zero by a small circle ($\circ$). The s-plane including the locations of the finite poles and zeros of $F(s)$ is called the **pole-zero map** of $F(s)$. A similar comment holds for the z-plane.

EXAMPLE 4.44. The rational function

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

has finite poles $s = -3$, $s = -1 - j$, and $s = -1 + j$, and finite zeros $s = -1$ and $s = 2$. The pole-zero map of $F(s)$ is shown in Fig. 4-2.

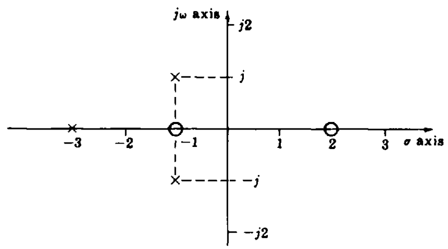


Fig. 4-2

4.12 GRAPHICAL EVALUATION OF RESIDUES*

Let $F(s)$ be a rational function written in its factored form:

$$F(s) = \frac{b_m \prod_{i=1}^{m} (s + z_i)}{\prod_{i=1}^{n} (s + p_i)}$$

Since $F(s)$ is a complex function, it can be written in *polar form* as

$$F(s) = |F(s)|e^{j\phi} = |F(s)|\underline{/\phi}$$

where $|F(s)|$ is the absolute value of $F(s)$ and $\phi \equiv \arg F(s) = \tan^{-1}[\text{Im } F(s)/\text{Re } F(s)]$.

$F(s)$ can further be written in terms of the polar forms of the factors $s + z_i$ and $s + p_i$ as

$$F(s) = \frac{b_m \prod_{i=1}^{m} |s + z_i|}{\prod_{i=1}^{n} |s + p_i|} \underline{\Bigg/ \left[\sum_{i=1}^{m} \phi_{iz} - \sum_{i=1}^{n} \phi_{ip} \right]}$$

where $s + z_i = |s + z_i|\underline{/\phi_{iz}}$ and $s + p_i = |s + p_i|\underline{/\phi_{ip}}$.

*While s is used to represent the complex variable in this section, it is not intended to represent the Laplace variable only but rather to be a general complex variable and the discussion is applicable to both the Laplace and z-transforms.

Each complex number s, z_i, p_i, $s + z_i$, and $s + p_i$ can be represented by a vector in the s-plane. If p is a general complex number, then the vector representing p has magnitude $|p|$ and direction defined by the angle

$$\phi = \tan^{-1}\left[\frac{\text{Im } p}{\text{Re } p}\right]$$

measured counterclockwise from the positive σ-axis.

A typical pole $-p_i$ and zero $-z_i$ are shown in Fig. 4-3, along with a general complex variable s. The *sum vectors* $s + z_i$ and $s + p_i$ are also shown. Note that the vector $s + z_i$ is a vector which starts at the zero $-z_i$ and terminates at s, and $s + p_i$ starts at the pole $-p_i$ and terminates at s.

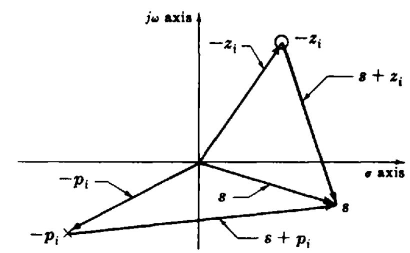


Fig. 4-3

For distinct poles of the rational function $F(s)$, the *residue* $c_{k1} \equiv c_k$ of the pole $-p_k$ is given by

$$c_k = (s + p_k)F(s)\big|_{s=-p_k} = \left. \frac{b_m (s + p_k) \prod_{i=1}^{m} (s + z_i)}{\prod_{i=1}^{n} (s + p_i)} \right|_{s=-p_k}$$

These residues can be determined by the following graphical procedure:

1. Plot the pole-zero map of $(s + p_k)F(s)$.
2. Draw vectors on this map starting at the poles and zeros of $(s + p_k)F(s)$, and terminating at $-p_k$. Measure the magnitude (in the scale of the pole-zero map) of these vectors and the angles of the vectors measured from the positive real axis in the counterclockwise direction.
3. Obtain the magnitude $|c_k|$ of the residue c_k as the product of b_m and the magnitudes of the vectors from the zeros to $-p_k$, divided by the product of the magnitudes of the vectors from the poles to $-p_k$.
4. Determine the angle ϕ_k of the residue c_k as the sum of the angles of the vectors from the zeros to $-p_k$, minus the sum of the angles of the vectors from the poles to $-p_k$. This is true for positive b_m. If b_m is negative, then add 180° to this angle.

The residue c_k is given in polar form by

$$c_k = |c_k|e^{j\phi_k} = |c_k|\underline{/\phi_k}$$

or in rectangular form by

$$c_k = |c_k|\cos\phi_k + j|c_k|\sin\phi_k$$

This graphical technique is not directly applicable for evaluating residues of multiple poles.

4.13 SECOND-ORDER SYSTEMS

As indicated in Section 3.14, many control systems can be described or approximated by the general second-order differential equation

$$\frac{d^2y}{dt^2} + 2\zeta\omega_n\frac{dy}{dt} + \omega_n^2 y = \omega_n^2 u$$

The positive coefficient ω_n is called the **undamped natural frequency** and the coefficient ζ is the **damping ratio** of the system.

The Laplace transform of $y(t)$, when the initial conditions are zero, is

$$Y(s) = \left[\frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}\right]U(s)$$

where $U(s) = \mathscr{L}[u(t)]$. The poles of the function $Y(s)/U(s) = \omega_n^2/(s^2 + 2\zeta\omega_n s + \omega_n^2)$ are

$$s = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$$

Note that:

1. If $\zeta > 1$, both poles are negative and real.
2. If $\zeta = 1$, the poles are equal, negative, and real ($s = -\omega_n$).
3. If $0 < \zeta < 1$, the poles are complex conjugates with negative real parts ($s = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2}$).
4. If $\zeta = 0$, the poles are imaginary and complex conjugate ($s = \pm j\omega_n$).
5. If $\zeta < 0$, the poles are in the right half of the s-plane (RHP).

Of particular interest in this book is Case 3, representing an **underdamped second-order system**. The poles are complex conjugates with negative real parts and are located at

$$s = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2}$$

or at

$$s = -\alpha \pm j\omega_d$$

where $1/\alpha \equiv 1/\zeta\omega_n$ is called the **time constant** of the system and $\omega_d \equiv \omega_n\sqrt{1-\zeta^2}$ is called the **damped natural frequency** of the system. For fixed ω_n, Fig. 4-4 shows the locus of these poles as a function of ζ, $0 < \zeta < 1$. The locus is a semicircle of radius ω_n. The angle θ is related to the damping ratio by $\theta = \cos^{-1}\zeta$.

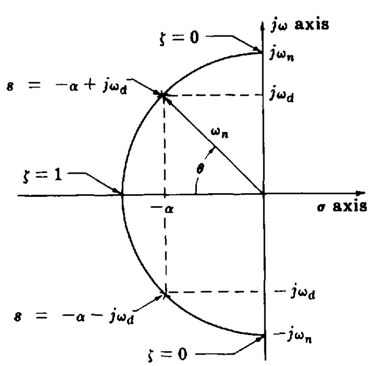


Fig. 4-4

A similar description for second-order systems described by difference equations does not exist in such a simple and useful form.

Solved Problems

LAPLACE TRANSFORMS FROM THE DEFINITION

4.1. Show that the unit step function $\mathbf{1}(t)$ is Laplace transformable and determine its Laplace transform.

Direct substitution into the equation of Definition 4.2 yields

$$\int_{0^+}^{\infty} |\mathbf{1}(t)|e^{-\sigma_0 t}\,dt = \int_{0^+}^{\infty} e^{-\sigma_0 t}\,dt = -\frac{1}{\sigma_0}e^{-\sigma_0 t}\Big|_{0^+}^{\infty} = \frac{1}{\sigma_0} < +\infty$$

for $\sigma_0 > 0$. The Laplace transform is given by Definition 4.1:

$$\mathscr{L}[\mathbf{1}(t)] = \int_{0^+}^{\infty}\mathbf{1}(t)e^{-st}\,dt = -\frac{1}{s}e^{-st}\Big|_{0^+}^{\infty} = \frac{1}{s} \qquad \text{for} \quad \operatorname{Re} s > 0$$

4.2. Show that the unit ramp function t is Laplace transformable and determine its Laplace transform.

Direct substitution into the equation of Definition 4.2 yields

$$\int_{0^+}^{\infty} |t|e^{-\sigma_0 t}\,dt = \frac{e^{-\sigma_0 t}}{\sigma_0^2}(-\sigma_0 t - 1)\Big|_{0^+}^{\infty} = \frac{1}{\sigma_0^2} < +\infty$$

for $\sigma_0 > 0$. The Laplace transform is given by Definition 4.1:

$$\mathscr{L}[t] = \int_{0^+}^{\infty} te^{-st}\,dt = \frac{e^{-st}}{s^2}(-st - 1)\Big|_{0^+}^{\infty} = \frac{1}{s^2} \qquad \text{for} \quad \operatorname{Re} s > 0$$

4.3. Show that the sine function $\sin t$ is Laplace transformable and determine its Laplace transform.

The integral $\int_{0^+}^{\infty}|\sin t|e^{-\sigma_0 t}\,dt$ can be evaluated by writing the integral over the positive half cycles of $\sin t$ as

$$\int_{n\pi}^{(n+1)\pi}\sin t\, e^{-\sigma_0 t}\,dt = \frac{e^{-\sigma_0 n\pi}}{\sigma_0^2 + 1}\left[e^{-\sigma_0\pi} + 1\right]$$

for n even, and over negative half-cycles of $\sin t$ as

$$-\int_{n\pi}^{(n+1)\pi}\sin t\, e^{-\sigma_0 t}\,dt = \frac{e^{-\sigma_0 n\pi}}{\sigma_0^2 + 1}\left[e^{-\sigma_0\pi} + 1\right]$$

for n odd. Then

$$\int_{0^+}^{\infty}|\sin t|e^{-\sigma_0 t}\,dt = \frac{e^{-\sigma_0\pi} + 1}{\sigma_0^2 + 1}\sum_{n=0}^{\infty}e^{-\sigma_0 n\pi}$$

The summation converges for $e^{-\sigma_0\pi} < 1$ or $\sigma_0 > 0$ and can be written in closed form as

$$\sum_{n=0}^{\infty}e^{-\sigma_0 n\pi} = \frac{1}{1 - e^{-\sigma_0\pi}}$$

Then

$$\int_{0^+}^{\infty}|\sin t|e^{-\sigma_0 t}\,dt = \left[\frac{1 + e^{-\sigma_0\pi}}{1 - e^{-\sigma_0\pi}}\right]\left(\frac{1}{\sigma_0^2 + 1}\right) < +\infty \qquad \text{for} \quad \sigma_0 > 0$$

Finally, $\mathscr{L}[\sin t] = \int_{0^+}^{\infty}\sin t\, e^{-st}\,dt = \dfrac{e^{-st}(-s\sin t - \cos t)}{s^2 + 1}\Big|_{0^+}^{\infty} = \dfrac{1}{s^2 + 1} \qquad \text{for} \quad \operatorname{Re} s > 0$

4.4. Show that the Laplace transform of the unit impulse function is given by $\mathscr{L}[\delta(t)] = 1$.

Direct substitution of Equation (3.19) into the equation of Definition 4.1 yields

$$\int_{0^+}^{\infty} \delta(t) e^{-st}\,dt = \int_{0^+}^{\infty} \lim_{\Delta t \to 0}\left[\frac{\mathbf{1}(t) - \mathbf{1}(t-\Delta t)}{\Delta t}\right] e^{-st}\,dt$$

$$= \lim_{\Delta t \to 0}\left[\int_{0^+}^{\infty} \frac{\mathbf{1}(t)}{\Delta t} e^{-st}\,dt - \int_{0^+}^{\infty} \frac{\mathbf{1}(t-\Delta t)}{\Delta t} e^{-st}\,dt\right] = \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left[\frac{1}{s} - \frac{e^{-\Delta t s}}{s}\right]$$

where the Laplace transform of $\mathbf{1}(t)$ is $1/s$, as shown in Problem 4.1, and the second term is obtained using Property 9. Now

$$e^{-\Delta t s} = 1 - \Delta t s + \frac{(\Delta t s)^2}{2!} - \frac{(\Delta t s)^3}{3!} + \cdots$$

(see Reference [1]). Thus

$$\mathscr{L}[\delta(t)] = \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left[\frac{1}{s} - \frac{e^{-\Delta t s}}{s}\right] = \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left[\Delta t - \frac{(\Delta t)^2 s}{2!} + \frac{(\Delta t)^3 s^2}{3!} - \cdots\right] = 1$$

PROPERTIES OF THE LAPLACE TRANSFORM AND ITS INVERSE

4.5. Show that $\mathscr{L}[a_1 f_1(t) + a_2 f_2(t)] = a_1 F_1(s) + a_2 F_2(s)$, where $F_1(s) = \mathscr{L}[f_1(t)]$ and $F_2(s) = \mathscr{L}[f_2(t)]$ (Property 1).

By definition,

$$\mathscr{L}[a_1 f_1(t) + a_2 f_2(t)] = \int_{0^+}^{\infty} [a_1 f_1(t) + a_2 f_2(t)] e^{-st}\,dt$$

$$= \int_{0^+}^{\infty} a_1 f_1(t) e^{-st}\,dt + \int_{0^+}^{\infty} a_2 f_2(t) e^{-st}\,dt$$

$$= a_1 \int_{0^+}^{\infty} f_1(t) e^{-st}\,dt + a_2 \int_{0^+}^{\infty} f_2(t) e^{-st}\,dt$$

$$= a_1 \mathscr{L}[f_1(t)] + a_2 \mathscr{L}[f_2(t)] = a_1 F_1(s) + a_2 F_2(s)$$

4.6. Show that $\mathscr{L}^{-1}[a_1 F_1(s) + a_2 F_2(s)] = a_1 f_1(t) + a_2 f_2(t)$, where $\mathscr{L}^{-1}[F_1(s)] = f_1(t)$ and $\mathscr{L}^{-1}[F_2(s)] = f_2(t)$ (Property 2).

By definition,

$$\mathscr{L}^{-1}[a_1 F_1(s) + a_2 F_2(s)] = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} [a_1 F_1(s) + a_2 F_2(s)] e^{st}\,ds$$

$$= \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} a_1 F_1(s) e^{st}\,ds + \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} a_2 F_2(s) e^{st}\,ds$$

$$= a_1\left[\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F_1(s) e^{st}\,ds\right] + a_2\left[\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F_2(s) e^{st}\,ds\right]$$

$$= a_1 \mathscr{L}^{-1}[F_1(s)] + a_2 \mathscr{L}^{-1}[F_2(s)] = a_1 f_1(t) + a_2 f_2(t)$$

4.7. Show that the Laplace transform of the derivative df/dt of a function $f(t)$ is given by $\mathscr{L}[df/dt] = sF(s) - f(0^+)$, where $F(s) = \mathscr{L}[f(t)]$ (Property 3).

By definition,

$$\mathscr{L}\left[\frac{df}{dt}\right] = \lim_{\substack{T \to \infty \\ \epsilon \to 0}} \int_{\epsilon}^{T} \frac{df}{dt} e^{-st}\,dt$$

Integrating by parts,

$$\lim_{\substack{T \to \infty \\ \epsilon \to 0}} \int_{\epsilon}^{T} \frac{df}{dt} e^{-st}\,dt = \lim_{\substack{T \to \infty \\ \epsilon \to 0}} \left[f(t) e^{-st}\Big|_{\epsilon}^{T} + s\int_{\epsilon}^{T} f(t) e^{-st}\,dt\right] = -f(0^+) + sF(s)$$

where $\lim_{\epsilon \to 0} f(\epsilon) = f(0^+)$.

4.8. Show that

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

where $F(s) = \mathscr{L}[f(t)]$ (Property 4).

By definition and a change in the order of integrations, we have

$$\mathscr{L}\left[\int_0^t f(\tau)\,d\tau\right] = \int_{0^+}^{\infty}\int_0^t f(\tau)\,d\tau\, e^{-st}\,dt = \int_{0^+}^{\infty} f(\tau)\int_{\tau}^{\infty} e^{-st}\,dt\,d\tau$$

$$= \int_{0^+}^{\infty} f(\tau)\left[-\frac{1}{s}e^{-st}\Big|_{\tau}^{\infty}\right] d\tau = \int_{0^+}^{\infty} f(\tau)\frac{e^{-s\tau}}{s}\,d\tau = \frac{F(s)}{s}$$

4.9. Show that $f(0^+) \equiv \lim_{t \to 0} f(t) = \lim_{s \to \infty} sF(s)$, where $F(s) = \mathscr{L}[f(t)]$ (Property 5).

From Problem 4.7,

$$\mathscr{L}\left[\frac{df}{dt}\right] = sF(s) - f(0^+) = \lim_{\substack{T \to \infty \\ \epsilon \to 0}} \int_{\epsilon}^{T} \frac{df}{dt} e^{-st}\,dt$$

Now let $s \to \infty$, that is,

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

Since the limiting processes can be interchanged, we have

$$\lim_{s \to \infty}\left[\lim_{\substack{T \to \infty \\ \epsilon \to 0}} \int_{\epsilon}^{T} \frac{df}{dt} e^{-st}\,dt\right] = \lim_{\substack{T \to \infty \\ \epsilon \to 0}} \int_{\epsilon}^{T} \frac{df}{dt}\left(\lim_{s \to \infty} e^{-st}\right) dt$$

But $\lim_{s \to \infty} e^{-st} = 0$. Hence the right side of the equation is zero and $\lim_{s \to \infty} sF(s) = f(0^+)$.

4.10. Show that if $\lim_{t \to \infty} f(t)$ exists then $f(\infty) \equiv \lim_{t \to \infty} f(t) = \lim_{s \to 0} sF(s)$, where $F(s) = \mathscr{L}[f(t)]$ (Property 6).

From Problem 4.7,

$$\mathscr{L}\left[\frac{df}{dt}\right] = sF(s) - f(0^+) = \lim_{\substack{T \to \infty \\ \epsilon \to 0}} \int_{\epsilon}^{T} \frac{df}{dt} e^{-st}\,dt$$

Now let $s \to 0$, that is,

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

Since the limiting processes can be interchanged, we have

$$\lim_{s \to 0}\left[\lim_{\substack{T \to \infty \\ \epsilon \to 0}} \int_{\epsilon}^{T} \frac{df}{dt} e^{-st}\,dt\right] = \lim_{\substack{T \to \infty \\ \epsilon \to 0}} \int_{\epsilon}^{T} \frac{df}{dt}\left(\lim_{s \to 0} e^{-st}\right) dt = \lim_{\substack{T \to \infty \\ \epsilon \to 0}} \int_{\epsilon}^{T} \frac{df}{dt}\,dt = f(\infty) - f(0^+)$$

Adding $f(0^+)$ to both sides of the last equation yields $\lim_{s \to 0} sF(s) = f(\infty)$ if $f(\infty) = \lim_{t \to \infty} f(t)$ exists.

4.11. Show that $\mathscr{L}[f(t/a)] = aF(as)$, where $F(s) = \mathscr{L}[f(t)]$ (Property 7).

By definition, $\mathscr{L}[f(t/a)] = \int_{0^+}^{\infty} f(t/a)e^{-st}\,dt$. Making the change of variable $\tau = t/a$,

$$\mathscr{L}\left[f\left(\frac{t}{a}\right)\right] = a\int_{0^+}^{\infty} f(\tau)\,e^{-(as)\tau}\,d\tau = aF(as)$$

4.12. Show that $\mathscr{L}^{-1}[F(s/a)] = af(at)$, where $f(t) = \mathscr{L}^{-1}[F(s)]$ (Property 8).

By definition,

$$\mathscr{L}^{-1}\left[F\left(\frac{s}{a}\right)\right] = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F\left(\frac{s}{a}\right)e^{st}\,ds$$

Making the change of variable $\omega = s/a$,

$$\mathscr{L}^{-1}\left[F\left(\frac{s}{a}\right)\right] = \frac{a}{2\pi j}\int_{c-j\infty}^{c+j\infty} F(\omega)\,e^{\omega(at)}\,d\omega = af(at)$$

4.13. Show that $\mathscr{L}[f(t-T)] = e^{-sT}F(s)$, where $f(t-T) = 0$ for $t \le T$ and $F(s) = \mathscr{L}[f(t)]$ (Property 9).

By definition,

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

Making the change of variable $\theta = t - T$,

$$\mathscr{L}[f(t-T)] = \int_{0^+}^{\infty} f(\theta)\,e^{-s\theta}e^{-sT}\,d\theta = e^{-sT}F(s)$$

4.14. Show that $\mathscr{L}[e^{-at}f(t)] = F(s+a)$, where $F(s) = \mathscr{L}[f(t)]$ (Property 10).

By definition,

$$\mathscr{L}\left[e^{-at}f(t)\right] = \int_{0^+}^{\infty} e^{-at}f(t)\,e^{-st}\,dt = \int_{0^+}^{\infty} f(t)\,e^{-(s+a)t}\,dt = F(s+a)$$

4.15. Show that

$$\mathscr{L}[f_1(t)f_2(t)] = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F_1(\omega)\,F_2(s-\omega)\,d\omega$$

where $F_1(s) = \mathscr{L}[f_1(t)]$ and $F_2(s) = \mathscr{L}[f_2(t)]$ (Property 11).

By definition,

$$\mathscr{L}[f_1(t)f_2(t)] = \int_{0^+}^{\infty} f_1(t)f_2(t)\,e^{-st}\,dt$$

But

$$f_1(t) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F_1(\omega)\,e^{\omega t}\,d\omega$$

Hence

$$\mathscr{L}[f_1(t)f_2(t)] = \frac{1}{2\pi j}\int_{0^+}^{\infty}\int_{c-j\infty}^{c+j\infty} F_1(\omega)\,e^{\omega t}\,d\omega\, f_2(t)\,e^{-st}\,dt$$

Interchanging the order of integrations yields

$$\mathscr{L}[f_1(t)f_2(t)] = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F_1(\omega)\int_{0^+}^{\infty} f_2(t)\,e^{-(s-\omega)t}\,dt\,d\omega$$

Since $\int_{0^+}^{\infty} f_2(t)e^{-(s-\omega)t}\,dt = F_2(s-\omega)$,

$$\mathscr{L}[f_1(t)f_2(t)] = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F_1(\omega)\,F_2(s-\omega)\,d\omega$$

4.16. Show that

$$\mathscr{L}^{-1}[F_1(s)F_2(s)] = \int_{0^+}^{t} f_1(\tau)f_2(t-\tau)\,d\tau$$

where $f_1(t) = \mathscr{L}^{-1}[F_1(s)]$ and $f_2(t) = \mathscr{L}^{-1}[F_2(s)]$ (Property 12).

By definition,

$$\mathscr{L}^{-1}[F_1(s)F_2(s)] = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F_1(s)\,F_2(s)\,e^{st}\,ds$$

But $F_1(s) = \int_{0^+}^{\infty} f_1(\tau)e^{-s\tau}\,d\tau$. Hence

$$\mathscr{L}^{-1}[F_1(s)F_2(s)] = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty}\int_{0^+}^{\infty} f_1(\tau)\,e^{-s\tau}\,d\tau\,F_2(s)\,e^{st}\,ds$$

Interchanging the order of integrations yields

$$\mathscr{L}^{-1}[F_1(s)F_2(s)] = \int_{0^+}^{\infty}\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F_2(s)\,e^{s(t-\tau)}\,ds\,f_1(\tau)\,d\tau$$

Since

$$\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F_2(s)\,e^{s(t-\tau)}\,ds = f_2(t-\tau)$$

then

$$\mathscr{L}^{-1}[F_1(s)F_2(s)] = \int_{0^+}^{\infty} f_1(\tau)f_2(t-\tau)\,d\tau = \int_{0^+}^{t} f_1(\tau)f_2(t-\tau)\,d\tau$$

where the second equality is true since $f_2(t-\tau) = 0$ for $\tau \ge t$.

4.17. Show that

$$\mathscr{L}\left[\frac{d^iy}{dt^i}\right] = s^iY(s) - \sum_{k=0}^{i-1} s^{i-1-k}y_0^k$$

for $i > 0$, where $Y(s) = \mathscr{L}[y(t)]$ and $y_0^k = (d^ky/dt^k)|_{t=0^+}$

This result can be shown by mathematical induction. For $i = 1$,

$$\mathscr{L}\left[\frac{dy}{dt}\right] = sY(s) - y(0^+) = sY(s) - y_0^0$$

as shown in Problem 4.7. Now assume the result holds for $i = n-1$, that is,

$$\mathscr{L}\left[\frac{d^{n-1}y}{dt^{n-1}}\right] = s^{n-1}Y(s) - \sum_{k=0}^{n-2} s^{n-2-k}y_0^k$$

Then $\mathscr{L}[d^ny/dt^n]$ can be written as

$$\begin{aligned}
\mathscr{L}\left[\frac{d^ny}{dt^n}\right] &= \mathscr{L}\left[\frac{d}{dt}\left(\frac{d^{n-1}y}{dt^{n-1}}\right)\right] = s\mathscr{L}\left[\frac{d^{n-1}y}{dt^{n-1}}\right] - \left.\frac{d^{n-1}y}{dt^{n-1}}\right|_{t=0^+} \\
&= s\left(s^{n-1}Y(s) - \sum_{k=0}^{n-2} s^{n-2-k}y_0^k\right) - y_0^{n-1} = s^nY(s) - \sum_{k=0}^{n-1} s^{n-1-k}y_0^k
\end{aligned}$$

For the special case $n = 2$, we have $\mathscr{L}[d^2y/dt^2] = s^2Y(s) - sy_0^0 - y_0^1$.

LAPLACE TRANSFORMS AND THEIR INVERSE FROM THE TABLE OF TRANSFORM PAIRS

4.18. Find the Laplace transform of $f(t) = 2e^{-t}\cos 10t - t^4 + 6e^{-(t-10)}$ for $t > 0$.

From the table of transform pairs,

$$\mathscr{L}[e^{-t}\cos 10t] = \frac{s+1}{(s+1)^2 + 10^2} \qquad \mathscr{L}[t^4] = \frac{4!}{s^5} \qquad \mathscr{L}[e^{-t}] = \frac{1}{s+1}$$

Using Property 9, $\mathscr{L}[e^{-(t-10)}] = e^{-10s}/(s+1)$. Using Property 1,

$$\mathscr{L}[f(t)] = 2\mathscr{L}[e^{-t}\cos 10t] - \mathscr{L}[t^4] + 6\mathscr{L}[e^{-(t-10)}] = \frac{2(s+1)}{s^2+2s+101} - \frac{24}{s^5} + \frac{6e^{-10s}}{s+1}$$

4.19. Find the inverse Laplace transform of

$$F(s) = \frac{2e^{-0.5s}}{s^2 - 6s + 13} - \frac{s-1}{s^2 - 2s + 2}$$

for $t > 0$.

$$\frac{2}{s^2 - 6s + 13} = \frac{2}{(s-3)^2 + 2^2} \qquad \frac{s-1}{s^2 - 2s + 2} = \frac{s-1}{(s-1)^2 + 1}$$

The inverse transforms are determined directly from Table 4.1 as

$$\mathscr{L}^{-1}\left[\frac{1}{(s-3)^2 + 2^2}\right] = e^{3t}\sin 2t \qquad \mathscr{L}^{-1}\left[\frac{s-1}{(s-1)^2 + 1}\right] = e^t \cos t$$

Using Property 9, *then* Property 2, results in

$$f(t) = \begin{cases} -e^t \cos t & 0 < t \le 0.5 \\ e^{3(t-0.5)}\sin 2(t-0.5) - e^t\cos t & t > 0.5 \end{cases}$$

LAPLACE TRANSFORMS OF LINEAR CONSTANT-COEFFICIENT DIFFERENTIAL EQUATIONS

4.20. Determine the output transform $Y(s)$ for the differential equation

$$\frac{d^3y}{dt^3} + 3\frac{d^2y}{dt^2} - \frac{dy}{dt} + 6y = \frac{d^2u}{dt^2} - u$$

where y = output, u = input, and initial conditions are

$$y(0^+) = \left.\frac{dy}{dt}\right|_{t=0^+} = 0 \qquad \left.\frac{d^2y}{dt^2}\right|_{t=0^+} = 1$$

Using Property 3 or the result of Problem 4.17, the Laplace transforms of the terms of the equation are given as

$$\mathscr{L}\left[\frac{d^3y}{dt^3}\right] = s^3Y(s) - s^2y(0^+) - s\left.\frac{dy}{dt}\right|_{t=0^+} - \left.\frac{d^2y}{dt^2}\right|_{t=0^+} = s^3Y(s) - 1$$

$$\mathscr{L}\left[\frac{d^2y}{dt^2}\right] = s^2Y(s) - sy(0^+) - \left.\frac{dy}{dt}\right|_{t=0^+} = s^2Y(s)$$

$$\mathscr{L}\left[\frac{dy}{dt}\right] = sY(s) - y(0^+) = sY(s) \qquad \mathscr{L}\left[\frac{d^2u}{dt^2}\right] = s^2U(s) - su(0^+) - \left.\frac{du}{dt}\right|_{t=0^+}$$

where $Y(s) = \mathscr{L}[y(t)]$ and $U(s) = \mathscr{L}[u(t)]$. The Laplace transform of the given equation can now be

written as

$$\mathscr{L}\left[\frac{d^3y}{dt^3}\right] + 3\mathscr{L}\left[\frac{d^2y}{dt^2}\right] - \mathscr{L}\left[\frac{dy}{dt}\right] + 6\mathscr{L}[y]$$

$$= s^3Y(s) - 1 + 3s^2Y(s) - sY(s) + 6Y(s)$$

$$= \mathscr{L}\left[\frac{d^2u}{dt^2}\right] - \mathscr{L}[u] = s^2U(s) - su(0^+) - \left.\frac{du}{dt}\right|_{t=0^+} - U(s)$$

Solving for $Y(s)$, we obtain

$$Y(s) = \frac{(s^2-1)U(s)}{s^3 + 3s^2 - s + 6} - \frac{su(0^+) + \left.\frac{du}{dt}\right|_{t=0^+}}{s^3 + 3s^2 - s + 6} + \frac{1}{s^3 + 3s^2 - s + 6}$$

4.21. What part of the solution of Problem 4.20 is the transform of the free response? The forced response?

The transform of the free response $Y_\alpha(s)$ is that part of the output transform $Y(s)$ which does not depend on the input $u(t)$, its derivatives or its transform; that is,

$$Y_\alpha(s) = \frac{1}{s^3 + 3s^2 - s + 6}$$

The transform of the forced response $Y_b(s)$ is that part of $Y(s)$ which depends on $u(t)$, its derivative and its transform; that is,

$$Y_b(s) = \frac{(s^2-1)U(s)}{s^2 + 3s^2 - s + 6} - \frac{su(0^+) + \left.\frac{du}{dt}\right|_{t=0^+}}{s^3 + 3s^2 - s + 6}$$

4.22. What is the characteristic polynomial for the differential equation of Problems 4.20 and 4.21?

The characteristic polynomial is the denominator polynomial which is common to the transforms of the free and forced responses (see Problem 4.21), that is, the polynomial $s^3 + 3s^2 - s + 6$.

4.23. Determine the output transform $Y(s)$ of the system of Problem 4.20 for an input $u(t) = 5\sin t$.

From Table 4.1, $U(s) \equiv \mathscr{L}[u(t)] = \mathscr{L}[5\sin t] = 5/(s^2+1)$.
The initial values of $u(t)$ and du/dt are $u(0^+) = \lim_{t\to 0} 5\sin t = 0$, $(du/dt)|_{t=0^+} = \lim_{t\to 0} 5\cos t = 5$.
Substituting these values into the output transform $Y(s)$ given in Problem 4.20,

$$Y(s) = \frac{s^2 - 9}{(s^3 + 3s^2 - s + 6)(s^2 + 1)}$$

PARTIAL FRACTION EXPANSIONS

4.24. A rational function $F(s)$ can be represented by

$$F(s) = \frac{\sum_{i=0}^{n} b_i s^i}{\prod_{i=1}^{r}(s + p_i)^{n_i}} = b_n + \sum_{i=1}^{r}\sum_{k=1}^{n_i}\frac{c_{ik}}{(s+p_i)^k} \tag{4.10a}$$

where the second form is the partial fraction expansion of $F(s)$. Show that the constants c_{ik} are given by

$$c_{ik} = \frac{1}{(n_i - k)!}\frac{d^{n_i-k}}{ds^{n_i-k}}\left[(s+p_i)^{n_i}F(s)\right]\Big|_{s=-p_i} \tag{4.10b}$$

Let $(s+p_j)$ be the factor of interest and form

$$(s+p_j)^{n_j}F(s) = (s+p_j)^{n_j}b_n + \sum_{i=1}^{r}\sum_{k=1}^{n_i}\frac{(s+p_j)^{n_j}c_{ik}}{(s+p_i)^k}$$

This can be rewritten as

$$(s+p_j)^{n_j}F(s) = (s+p_j)^{n_j}b_n + \sum_{i=1}^{j-1}\sum_{k=1}^{n_i}\frac{(s+p_j)^{n_j}c_{ik}}{(s+p_i)^k}$$

$$+\sum_{i=j+1}^{r}\sum_{k=1}^{n_i}\frac{(s+p_j)^{n_j}c_{ik}}{(s+p_i)^k} + \sum_{k=1}^{n_j}(s+p_j)^{n_j-k}c_{jk}$$

Now form $$\frac{d^{n_j-l}}{ds^{n_j-l}}\left[(s+p_j)^{n_j}F(s)\right]\bigg|_{s=-p_j}$$

Note that the first three terms on the right-hand side of $(s+p_j)^{n_j}F(s)$ will have a factor $s+p_j$ in the numerator even after being differentiated $n_j - l$ times $(l = 1,2,\ldots,n_j)$ and thus these three terms become zero when evaluated at $s=-p_j$. Therefore

$$\frac{d^{n_j-l}}{ds^{n_j-l}}\left[(s+p_j)^{n_j}F(s)\right]\bigg|_{s=-p_j} = \frac{d^{n_j-l}}{ds^{n_j-l}}\left[\sum_{k=1}^{n_j}(s+p_j)^{n_j-k}c_{jk}\right]\bigg|_{s=-p_j}$$

$$=\sum_{k=1}^{l}(n_j-k)(n_j-k-1)\cdots(l-k+1)(s+p_j)^{(-k+l)}c_{jk}\bigg|_{s=-p_j}$$

Except for that term in the summation for which $k=l$, all the other terms are zero since they contain factors $s+p_j$. Then

$$\frac{d^{n_j-l}}{ds^{n_j-l}}\left[(s+p_j)^{n_j}F(s)\right]\bigg|_{s=-p_j} = (n_j-l)(n_j-l-1)\cdots(1)c_{jl}$$

or $$c_{jl} = \frac{1}{(n_j-l)!}\frac{d^{n_j-l}}{ds^{n_j-l}}\left[(s+p_j)^{n_j}F(s)\right]\bigg|_{s=-p_j}$$

4.25. Expand $Y(s)$ of Example 4.17 in a partial fraction expansion.

$Y(s)$ can be rewritten with the denominator polynomial in factored form as

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

The partial fraction expansion of $Y(s)$ is [see Equation (*4.11*)]

$$Y(s) = b_3 + \frac{c_{11}}{s} + \frac{c_{21}}{s+1} + \frac{c_{31}}{s+2}$$

where $b_3 = 0$,

$$c_{11} = \frac{-(s^2+s-1)}{(s+1)(s+2)}\bigg|_{s=0} = \frac{1}{2} \qquad c_{21} = \frac{-(s^2+s-1)}{s(s+2)}\bigg|_{s=-1} = -1 \qquad c_{31} = \frac{-(s^2+s-1)}{s(s+1)}\bigg|_{s=-2} = -\frac{1}{2}$$

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

4.26. Expand $Y(s)$ of Example 4.19 in a partial fraction expansion.

$Y(s)$ can be rewritten with the denominator polynomial in factored form as

$$Y(s) = \frac{s^2+9s+19}{(s+1)(s+2)(s+4)}$$

The partial fraction expansion of $Y(s)$ is [see Equation (*4.11*)]

$$Y(s) = b_3 + \frac{c_{11}}{s+1} + \frac{c_{21}}{s+2} + \frac{c_{31}}{s+4}$$

where $b_3 = 0$,

$$c_{11} = \frac{s^2+9s+19}{(s+2)(s+4)}\bigg|_{s=-1} = \frac{11}{3} \qquad c_{21} = \frac{s^2+9s+19}{(s+1)(s+4)}\bigg|_{s=-2} = -\frac{5}{2}$$

$$c_{31} = \frac{s^2+9s+19}{(s+1)(s+2)}\bigg|_{s=-4} = -\frac{1}{6}$$

Thus $$Y(s) = \frac{11}{3(s+1)} - \frac{5}{2(s+2)} - \frac{1}{6(s+4)}$$

INVERSE LAPLACE TRANSFORMS USING PARTIAL FRACTION EXPANSIONS

4.27. Determine $y(t)$ for the system of Example 4.17.

Mathcad

From the result of Problem 4.25, the transform of $y(t)$ can be written as

$$\mathscr{L}[y(t)] \equiv Y(s) = \frac{1}{2s} - \frac{1}{s+1} - \frac{1}{2(s+2)}$$

Therefore

$$y(t) = \frac{1}{2}\mathscr{L}^{-1}\left[\frac{1}{s}\right] - \mathscr{L}^{-1}\left[\frac{1}{s+1}\right] - \frac{1}{2}\mathscr{L}^{-1}\left[\frac{1}{s+2}\right] = \frac{1}{2}\left[1 - 2e^{-t} - e^{-2t}\right] \qquad t>0$$

4.28. Determine $y(t)$ for the system of Example 4.19.

From the result of Problem 4.26, the transform of $y(t)$ can be written as

$$\mathscr{L}[y(t)] = Y(s) = \frac{11}{3(s+1)} - \frac{5}{2(s+2)} - \frac{1}{6(s+4)}$$

Therefore $$y(t) = \frac{11}{3}e^{-t} - \frac{5}{2}e^{-2t} - \frac{1}{6}e^{-4t}$$

ROOTS OF POLYNOMIALS

4.29. Find an approximation of a real root of the polynomial equation

$$Q_3(s) = s^3 - 3s^2 + 4s - 5 = 0$$

to an accuracy of three significant figures using *Horner's method*.

By Descartes' rule of signs, $Q_3(s)$ has three variations in the signs of its coefficients (1 to -3, -3 to 4, and 4 to -5). Thus there may be three positive real roots. $Q_3(-s) = -s^3 - 3s^2 - 4s - 5$ has no sign changes; therefore $Q_3(s)$ has no negative real roots and only real values of s greater than zero need be considered.

Step 1—We have $Q_3(0) = -5$, $Q_3(1) = -3$, $Q_3(2) = -1$, $Q_3(3) = 7$. Therefore $k_0 = 2$ and the first approximation is $s_0 = k_0 = 2$.

Step 2—Determine $Q_3^1(s)$ as

$$Q_3^1(s) = Q_3^0(2+s) = (2+s)^3 - 3(2+s)^2 + 4(2+s) - 5 = s^3 + 3s^2 + 4s - 1$$

Step 3—$Q_3^1(0) = -1$, $Q_3^1(\frac{1}{10}) = -0.569$, $Q_3^1(\frac{2}{10}) = -0.072$, $Q_3^1(\frac{3}{10}) = 0.497$. Hence $k_1 = 0.2$ and $s_1 = k_0 + k_1 = 2.2$.

Now repeat Step 2 to determine $Q_3^2(s)$:

$$Q_3^2(s) = Q_3^1(0.2+s) = (0.2+s)^3 + 3(0.2+s)^2 + 4(0.2+s) - 1 = s^3 + 3.6s^2 + 5.32s - 0.072$$

Repeating Step 3: $Q_3^2(0) = -0.072$, $Q_3^2(1/100) = -0.018$, $Q_3^2(2/100) = 0.036$. Hence $k_2 = 0.01$ and $s_2 = k_0 + k_1 + k_2 = 2.21$ which is an approximation of the root accurate to three significant figures.

4.30. Find an approximation of a real root of the polynomial equation given in Problem 4.29 using *Newton's method*. Perform four iterations and compare the result with the solution of Problem 4.29.

The sequence of approximations is defined by letting $n = 3$, $a_3 = 1$, $a_2 = -3$, $a_1 = 4$, and $a_0 = -5$ in the recursive relationship of Newton's method [Equation (*4.15*)]. The result is

$$s_{l+1} = \frac{2s_l^3 - 3s_l^2 + 5}{3s_l^2 - 6s_l + 4} \qquad l = 0, 1, 2, \ldots$$

Let the first guess be $s_0 = 0$. Then

$$s_1 = \frac{5}{4} = 1.25 \qquad s_3 = \frac{2(3.55)^3 - 3(3.55)^2 + 5}{3(3.55)^2 - 6(3.55) + 4} = 2.76$$

$$s_2 = \frac{2(1.25)^3 - 3(1.25)^2 + 5}{3(1.25)^2 - 6(1.25) + 4} = 3.55 \qquad s_4 = \frac{2(2.76)^3 - 3(2.76)^2 + 5}{3(2.76)^2 - 6(2.76) + 4} = 2.35$$

The next iteration yields $s_5 = 2.22$ and the sequence is converging.

4.31. Find an approximation of a quadratic factor of the polynomial

$$Q_3(s) = s^3 - 3s^2 + 4s - 5$$

of Problems 4.29 and 4.30, using the *Lin-Bairstow method*. Perform two iterations.

Step 1—Choose as a first approximation the factor $s^2 - s + 2$.
The constants needed in Step 2 are $\alpha_1 = -1$, $\alpha_0 = 2$, $n = 3$, $a_3 = 1$, $a_2 = -3$, $a_1 = 4$, $a_0 = -5$.
Step 2—From the recursive relationship

$$b_{i-2} = a_i - \alpha_1 b_{i-1} - \alpha_0 b_i$$

$i = n, n-1, \ldots, 1, 0$, the following constants are formed:

$$b_1 = a_3 = 1 \qquad b_0 = a_2 + b_1 = -2$$

$$b_{-1} = a_1 + b_0 - 2b_1 = 0 \qquad b_{-2} = a_0 + b_{-1} - 2b_0 = -1$$

Step 3—From the recursive relationship

$$c_{i-1} = b_{i-1} - \alpha_1 c_i - \alpha_0 c_{i+1}$$

$i = n, n-1, \ldots, 1$, the following constants are determined:

$$c_1 = b_1 = 1 \qquad c_0 = b_0 + c_1 = -1$$

Step 4—The simultaneous equations

$$c_0\,\Delta\alpha_1 + c_1\,\Delta\alpha_0 = b_{-1}$$

$$(-\alpha_1 c_0 - \alpha_0 c_1)\,\Delta\alpha_1 + c_0\,\Delta\alpha_0 = b_{-2}$$

can now be written as

$$-\Delta\alpha_1 + \Delta\alpha_0 = 0$$

$$-3\Delta\alpha_1 - \Delta\alpha_0 = -1$$

whose solution is $\Delta\alpha_1 = \frac{1}{4}$, $\Delta\alpha_0 = \frac{1}{4}$, and the new approximation of the quadratic factor is

$$s^2 - 0.75s + 2.25$$

If Steps 1 through 4 are repeated for $\alpha_1 = -0.75$, $\alpha_0 = 2.25$, the second iteration produces

$$s^2 - 0.7861s + 2.2583$$

POLE-ZERO MAPS

4.32. Determine all of the poles and zeros of $F(s) = (s^2 - 16)/(s^5 - 7s^4 - 30s^3)$.

The finite poles of $F(s)$ are the roots of the denominator polynomial equation

$$s^5 - 7s^4 - 30s^3 = s^3(s+3)(s-10) = 0$$

Therefore $s = 0$, $s = -3$, and $s = 10$ are the finite poles of $F(s)$. $s = 0$ is a triple root of the equation and is called a **triple pole** of $F(s)$. These are the only values of s for which $|F(s)|$ is infinite and are all the poles of $F(s)$. The finite zeros of $F(s)$ are the roots of the numerator polynomial equation

$$s^2 - 16 = (s-4)(s+4) = 0$$

Therefore $s = 4$ and $s = -4$ are the *finite zeros* of $F(s)$. As $|s| \to \infty$, $F(s) \cong 1/s^3 \to 0$. Then $F(s)$ has a triple zero at $s = \infty$.

4.33. Draw a pole-zero map for the function of Problem 4.32.

From the solution of Problem 4.32, $F(s)$ has *finite zeros* at $s = 4$ and $s = -4$, and *finite poles* at $s = 0$ (a triple pole), $s = -3$ and $s = 10$. The pole-zero map is shown in Fig. 4-5.

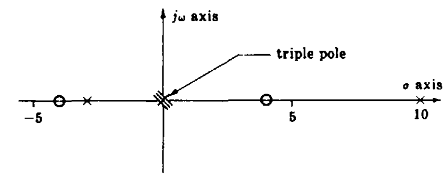


Fig. 4-5

4.34. Using the graphical technique, evaluate the residues of the function

$$F(s) = \frac{20}{(s+10)(s+1+j)(s+1-j)}$$

The pole-zero map of $F(s)$ is shown in Fig. 4-6.

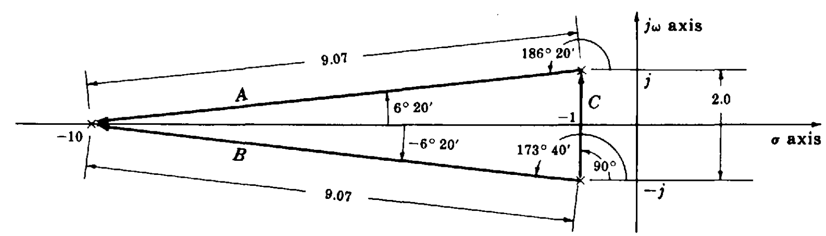


Fig. 4-6

Included in this pole-zero map are the vector displacements between the poles. For example, A is the vector displacement of the pole $s = -10$ relative to the pole $s = -1 + j$. Clearly then, $-A$ is the vector displacement of the pole $s = -1 + j$ relative to the pole $s = -10$.

The magnitude of the residue at the pole $s = -10$ is

$$|c_1| = \frac{20}{|A|\,|B|} = \frac{20}{(9.07)(9.07)} = 0.243$$

The angle ϕ_1 of the residue at $s = -10$ is the negative of the sum of the angles of A and B, that is, $\phi_1 = -[186°20' + 173°40'] = -360°$. Hence $c_1 = 0.243$.

The magnitude of the residue at the pole $s=-1+j$ is

$$|c_2| = \frac{20}{|-A||C|} = \frac{20}{(9.07)(2)} = 1.102$$

The angle ϕ_2 of the residue at the pole $s=-1+j$ is the negative of the sums of the angles of $-A$ and C: $\phi_2 = -[6°20' + 90°] = -96°20'$. Hence $c_2 = 1.102\angle -96°20' = -0.128 - j1.095$.

The magnitude of the residue at the pole $s=-1-j$ is

$$|c_3| = \frac{20}{|-B||-C|} = \frac{20}{(9.07)(2)} = 1.102$$

The angle ϕ_3 of the residue at the pole $s=-1-j$ is the negative of the sum of the angles of $-B$ and $-C$: $\phi_3 = -[-90° - 6°20'] = 96°20'$. Hence $c_3 = 1.102\angle 96°20' = -0.128 + j1.095$.

Note that the residues c_2 and c_3 of the complex conjugate poles are also complex conjugates. This is always true for the residues of complex conjugate poles.

SECOND-ORDER SYSTEMS

4.35. Determine (*a*) the undamped natural frequency ω_n, (*b*) the damping ratio ζ, (*c*) the time constant τ, (*d*) the damped natural frequency ω_d, (*e*) characteristic equation for the second-order system given by

Mathcad

$$\frac{d^2y}{dt^2} + 5\frac{dy}{dt} + 9y = 9u$$

Comparing this equation with the definitions of Section 4.13, we have

(*a*) $\omega_n^2 = 9$ or $\omega_n = 3$ rad/sec (*c*) $\tau = \dfrac{1}{\zeta\omega_n} = \dfrac{2}{5}$ sec (*e*) $s^2 + 5s + 9 = 0$

(*b*) $2\zeta\omega_n = 5$ or $\zeta = \dfrac{5}{2\omega_n} = \dfrac{5}{6}$ (*d*) $\omega_d = \omega_n\sqrt{1-\zeta^2} = 1.66$ rad/sec

4.36. How and why can the following system be approximated by a second-order system?

Mathcad

$$\frac{d^3y}{dt^3} + 12\frac{d^2y}{dt^2} + 22\frac{dy}{dt} + 20y = 20u$$

When the initial conditions on $y(t)$ and its derivatives are zero, the output transform is

$$\mathscr{L}[y(t)] \equiv Y(s) = \frac{20}{s^3 + 12s^2 + 22s + 20}U(s)$$

where $U(s) = \mathscr{L}[u(t)]$. This can be rewritten as

$$Y(s) = \frac{10}{41}\left(\frac{1}{s+10} - \frac{s}{s^2+2s+2}\right)U(s) + \frac{80}{41}\left(\frac{U(s)}{s^2+2s+2}\right)$$

The constant factor $\frac{80}{41}$ of the second term is 8 times the constant factor $\frac{10}{41}$ of the first term. The output $y(t)$ will then be dominated by the time function

$$\frac{80}{41}\mathscr{L}^{-1}\left[\frac{U(s)}{s^2+2s+2}\right]$$

The output transform $Y(s)$ can then be approximated by this second term; that is,

$$Y(s) \simeq \frac{80}{41}\left(\frac{U(s)}{s^2+2s+2}\right) \simeq \left(\frac{2}{s^2+2s+2}\right)U(s)$$

The second-order approximation is $d^2y/dt^2 + 2(dy/dt) + 2y = 2u$.

4.37. In Chapter 6 it will be shown that the output $y(t)$ of a time-invariant linear causal system with all initial conditions equal to zero is related to the input $u(t)$ in the Laplace transform domain

by the equation $Y(s) = P(s)U(s)$, where $P(s)$ is called the *transfer function* of the system. Show that $p(t)$, the inverse Laplace transform of $P(s)$, is equal to the *weighting function* $w(t)$ of a system described by the constant-coefficient differential equation

$$\sum_{i=0}^{n} a_i \frac{d^iy}{dt^i} = u$$

The forced response for a system described by the above equation is given by Equation (*3.15*), with all $b_i = 0$ except $b_0 = 1$:

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

and $w(t-\tau)$ is the weighting function of the differential equation.

The inverse Laplace transform of $Y(s) = P(s)U(s)$ is easily determined from the convolution integral of Property 12 as

$$y(t) = \mathscr{L}^{-1}[Y(s)] = \mathscr{L}^{-1}[P(s)U(s)] = \int_{0^+}^{t} p(t-\tau)u(\tau)\,d\tau$$

Hence
$$\int_{0^+}^{t} w(t-\tau)u(\tau)\,d\tau = \int_{0^+}^{t} p(t-\tau)u(\tau)\,d\tau \quad \text{or} \quad w(t) = p(t)$$

MISCELLANEOUS PROBLEMS

4.38. For the *R-C* network in Fig. 4-7:

(*a*) Find a differential equation which relates the output voltage y and the input voltage u.

(*b*) Let the initial voltage across the capacitor C be $v_{c0} = 1$ volt with the polarity shown, and let $u = 2e^{-t}$. Using the Laplace transform technique, find y.

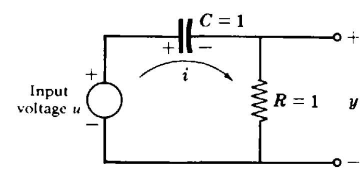


Fig. 4-7

(*a*) From Kirchhoff's voltage law

$$u = v_{c0} + \frac{1}{C}\int_0^t i\,dt + Ri = v_{c0} + \int_0^t i\,dt + i$$

But $y = Ri = i$. Therefore $u = v_{c0} + \int_0^t y\,dt + y$. Differentiating both sides of this integral equation yields the differential equation $\dot{y} + y = \dot{u}$.

(*b*) The Laplace transform of the differential equation found in part (*a*) is

$$sY(s) - y(0^+) + Y(s) = sU(s) - u(0^+)$$

where $U(s) = \mathscr{L}[2e^{-t}] = 2/(s+1)$ and $u(0^+) = \lim_{t\to 0} 2e^{-t} = 2$. To find $y(0^+)$, limits are taken on both sides of the original voltage equation:

$$u(0^+) = \lim_{t\to 0} u(t) = \lim_{t\to 0}\left[v_{c0} + \int_0^t y\,dt + y(t)\right] = v_{c0} + y(0^+)$$

Hence $y(0^+) = u(0^+) - v_{c0} = 2 - 1 = 1$. The transform of $y(t)$ is then

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

Finally,

$$y(t)=\mathscr{L}^{-1}\left[-\frac{2}{(s+1)^2}\right]+\mathscr{L}^{-1}\left[\frac{1}{s+1}\right]=-2te^{-t}+e^{-t}$$

4.39. Determine the Laplace transform of the output of the ideal sampler described in Problem 3.5.

From Definition 4.1 and Equation (*3.20*), the screening property of the unit impulse, we have

$$\begin{aligned}U^*(s)&=\int_{0^+}^{\infty}e^{-st}U^*(t)\,dt=\int_{0^+}^{\infty}e^{-st}\sum_{k=0}^{\infty}u(t)\,\delta(t-kT)\,dt\\&=\sum_{k=0}^{\infty}\int_{0^+}^{\infty}e^{-st}u(t)\,\delta(t-kT)\,dt=\sum_{k=0}^{\infty}e^{-skT}u(kT)\end{aligned}$$

4.40. Compare the result of Problem 4.39 with the z-transform of the sampled signal $u(kT)$, $k=0,1,2,\ldots$.

By definition the z-transform of the sampled signal is

$$U(z)=\sum_{k=0}^{\infty}u(kT)z^{-k}$$

This result could have been obtained directly by substituting $z=e^{sT}$ in the result of Problem 4.39.

4.41. Prove the Shift Theorem (Property 6, Section 4.9).

By definition,

$$\mathcal{Z}\{f(k)\}\equiv F(z)\equiv\sum_{k=0}^{\infty}f(k)z^{-k}$$

If we define a new, shifted sequence by $g(0)\equiv f(-1)=0$ and $g(k)\equiv f(k-1)$, $k=1,2,\ldots,$ then

$$\mathcal{Z}\{g(k)\}\equiv\sum_{k=0}^{\infty}g(k)z^{-k}\equiv\sum_{j=0}^{\infty}g(j)z^{-j}=\sum_{j=0}^{\infty}f(j-1)z^{-j}$$

(see Remark 1 following Definition 4.4). Now let k be redefined as $k\equiv j-1$ in the last equation. Then

$$\begin{aligned}\mathcal{Z}\{f(k-1)\}&=\sum_{k=-1}^{\infty}f(k)z^{-k-1}=z^{-1}\sum_{k=-1}^{\infty}f(k)z^{-k}\\&=z^{-1}f(-1)z^{+1}+z^{-1}\sum_{k=0}^{\infty}f(k)z^{-k}\\&=z^0\cdot 0+z^{-1}\sum_{k=0}^{\infty}f(k)z^{-k}=z^{-1}F(z)\end{aligned}$$

Note that repeated application of this result gives

$$Z[f(k-j)]=z^{-j}F(z)$$

Supplementary Problems

4.42. Show that $\mathscr{L}[-tf(t)]=dF(s)/ds$, where $F(s)=\mathscr{L}[f(t)]$.

4.43. Using the convolution integral find the inverse transform of $1/s(s+2)$.

4.44. Determine the final value of the function $f(t)$ whose Laplace transform is

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

4.45. Determine the initial value of the function $f(t)$ whose Laplace transform is

$$F(s)=\frac{4s}{s^3+2s^2+9s+6}$$

4.46. Find the partial fraction expansion of the function $F(s)=10/(s+4)(s+2)^3$.

4.47. Find the inverse Laplace transform $f(t)$ of the function $F(s)=10/(s+4)(s+2)^3$.

4.48. Solve Problem 3.24 using the Laplace transform technique.

4.49. Using the Laplace transform technique, find the forced response of the differential equation

$$\frac{d^2y}{dt^2}+4\frac{dy}{dt}+4y=3\frac{du}{dt}+2u$$

where $u(t)=e^{-3t}$, $t>0$. Compare this solution with that obtained in Problem 3.26.

4.50. Using the Laplace transform technique, find the transient and steady state responses of the system described by the differential equation $d^2y/dt^2+3(dy/dt)+2y=1$ with initial conditions $y(0^+)$ and $(dy/dt)|_{t=0^+}=1$.

4.51. Using the Laplace transform technique, find the unit impulse response of the system described by the differential equation $d^3y/dt^3+dy/dt=u$.

Answers to Some Supplementary Problems

4.43. $\frac{1}{2}[1-e^{-2t}]$

4.44. $\frac{2}{75}$

4.45. 0

4.46. $F(s)=\dfrac{5}{(s+2)^3}-\dfrac{5}{2(s+2)^2}+\dfrac{5}{4(s+2)}-\dfrac{5}{4(s+4)}$

4.47. $f(t)=\dfrac{5t^2e^{-2t}}{2}-\dfrac{5te^{-2t}}{2}+\dfrac{5e^{-2t}}{4}-\dfrac{5e^{-4t}}{4}$

4.49. $y_b(t)=7e^{-2t}-7e^{-3t}-7te^{-2t}$

4.50. Transient response $=2e^{-t}-\frac{3}{2}e^{-2t}$. Steady state response $=\frac{1}{2}$.

4.51. $y_\delta(t)=1-\cos t$

Chapter 5

Stability

5.1 STABILITY DEFINITIONS

The stability of a continuous or discrete-time system is determined by its response to inputs or disturbances. Intuitively, a stable system is one that remains at rest unless excited by an external source and returns to rest if all excitations are removed. Stability can be precisely defined in terms of the impulse response $y_\delta(t)$ of a continuous system, or Kronecker delta response $y_\delta(k)$ of a discrete-time system (see Sections 3.13 and 3.16), as follows:

***Definition 5.1a*:** A continuous system (discrete-time system) is **stable** if its impulse response $y_\delta(t)$ (Kronecker delta response $y_\delta(k)$) approaches zero as time approaches infinity.

Alternatively, the definition of a stable system can be based upon the response of the system to **bounded inputs**, that is, inputs whose magnitudes are less than some finite value for all time.

***Definition 5.1b*:** A continuous or discrete-time system is **stable** if every bounded input produces a bounded output.

Consideration of the *degree* of stability of a system often provides valuable information about its behavior. That is, if it is stable, how close is it to being unstable? This is the concept of **relative stability**. Usually, relative stability is expressed in terms of some allowable variation of a particular system parameter, over which the system remains stable. More precise definitions of relative stability indicators are presented in later chapters. Stability of nonlinear systems is treated in Chapter 19.

5.2 CHARACTERISTIC ROOT LOCATIONS FOR CONTINUOUS SYSTEMS

A major result of Chapters 3 and 4 is that the impulse response of a linear time-invariant continuous system is a sum of exponential time functions whose exponents are the roots of the system characteristic equation (see Equation 4.12). *A necessary and sufficient condition for the system to be stable is that the real parts of the roots of the characteristic equation have negative real parts.* This ensures that the impulse response will decay exponentially with time.

If the system has some roots with real parts equal to zero, but none with positive real parts, the system is said to be **marginally stable**. In this instance, the impulse response does not decay to zero, although it is bounded, but certain other inputs will produce unbounded outputs. Therefore marginally stable systems are unstable.

EXAMPLE 5.1. The system described by the Laplace transformed differential equation,

$$(s^2+1)Y(s)=U(s)$$

has the characteristic equation

$$s^2+1=0$$

This equation has the two roots $\pm j$. Since these roots have zero real parts, the system is not stable. It is, however, marginally stable since the equation has no roots with positive real parts. In response to most inputs or disturbances, the system oscillates with a bounded output. However, if the input is $u=\sin t$, the output will contain a term of the form: $y=t\cos t$, which is unbounded.

5.3 ROUTH STABILITY CRITERION

The Routh criterion is a method for determining continuous system stability, for systems with an nth-order characteristic equation of the form:

$$a_n s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0 = 0$$

The criterion is applied using a **Routh table** defined as follows:

$$\begin{array}{c|cccc} s^n & a_n & a_{n-2} & a_{n-4} & \cdots \\ s^{n-1} & a_{n-1} & a_{n-3} & a_{n-5} & \cdots \\ \cdot & b_1 & b_2 & b_3 & \cdots \\ \cdot & c_1 & c_2 & c_3 & \cdots \\ \cdot & \cdots & \cdots & \cdots & \cdots \end{array}$$

where $a_n, a_{n-1}, \ldots, a_0$ are the coefficients of the characteristic equation and

$$b_1 \equiv \frac{a_{n-1}a_{n-2}-a_n a_{n-3}}{a_{n-1}} \qquad b_2 \equiv \frac{a_{n-1}a_{n-4}-a_n a_{n-5}}{a_{n-1}} \qquad \text{etc.}$$

$$c_1 \equiv \frac{b_1 a_{n-3}-a_{n-1}b_2}{b_1} \qquad c_2 \equiv \frac{b_1 a_{n-5}-a_{n-1}b_3}{b_1} \qquad \text{etc.}$$

The table is continued horizontally and vertically until only zeros are obtained. Any row can be multiplied by a positive constant before the next row is computed without disturbing the properties of the table.

The Routh Criterion: *All the roots of the characteristic equation have negative real parts if and only if the elements of the first column of the Routh table have the same sign. Otherwise, the number of roots with positive real parts is equal to the number of changes of sign.*

EXAMPLE 5.2.

$$s^3+6s^2+12s+8=0$$

$$\begin{array}{c|ccc} s^3 & 1 & 12 & 0 \\ s^2 & 6 & 8 & 0 \\ s^1 & \frac{64}{6} & 0 & \\ s^0 & 8 & & \end{array}$$

Since there are no changes of sign in the first column of the table, all the roots of the equation have negative real parts.

Often it is desirable to determine a range of values of a particular system parameter for which the system is stable. This can be accomplished by writing the inequalities that ensure that there is no change of sign in the first column of the Routh table for the system. These inequalities then specify the range of allowable values of the parameter.

EXAMPLE 5.3.

$$s^3+3s^2+3s+1+K=0$$

$$\begin{array}{c|ccc} s^3 & 1 & 3 & 0 \\ s^2 & 3 & 1+K & 0 \\ s^1 & \dfrac{8-K}{3} & 0 & \\ s^0 & 1+K & & \end{array}$$

For no sign changes in the first column, it is necessary that the conditions $8-K>0$, $1+K>0$ be satisfied. Thus the characteristic equation has roots with negative real parts if $-1<K<8$, the simultaneous solution of these two inequalities.

A row of zeros for the s^1 row of the Routh table indicates that the polynomial has a pair of roots which satisfy the **auxiliary equation** formed as follows:

$$As^2 + B = 0$$

where A and B are the first and second elements of the s^2 row.

To continue the table, the zeros in the s^1 row are replaced with the coefficients of the derivative of the auxiliary equation. The derivative of the auxiliary equation is

$$2As + 0 = 0$$

The coefficients $2A$ and 0 are then entered into the s^1 row and the table is continued as described above.

EXAMPLE 5.4. In the previous example, the s^1 row is zero if $K = 8$. In this case, the auxiliary equation is $3s^2 + 9 = 0$. Therefore two of the roots of the characteristic equation are $s = \pm j\sqrt{3}$.

5.4 HURWITZ STABILITY CRITERION

The Hurwitz criterion is another method for determining whether all the roots of the characteristic equation of a continuous system have negative real parts. This criterion is applied using determinants formed from the coefficients of the characteristic equation. It is assumed that the first coefficient, a_n, is positive. The determinants Δ_i, $i = 1, 2, \ldots, n-1$, are formed as the principal minor determinants of the determinant

$$\Delta_n = \begin{vmatrix} a_{n-1} & a_{n-3} & \cdots & \begin{bmatrix} a_0 & \text{if } n \text{ odd} \\ a_1 & \text{if } n \text{ even} \end{bmatrix} & 0 & \cdots & 0 \\ a_n & a_{n-2} & \cdots & \begin{bmatrix} a_1 & \text{if } n \text{ odd} \\ a_0 & \text{if } n \text{ even} \end{bmatrix} & 0 & \cdots & 0 \\ 0 & a_{n-1} & a_{n-3} & \cdots & \cdots & \cdots & 0 \\ 0 & a_n & a_{n-2} & \cdots & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & a_0 \end{vmatrix}$$

The determinants are thus formed as follows:

$$\Delta_1 = a_{n-1}$$

$$\Delta_2 = \begin{vmatrix} a_{n-1} & a_{n-3} \\ a_n & a_{n-2} \end{vmatrix} = a_{n-1}a_{n-2} - a_n a_{n-3}$$

$$\Delta_3 = \begin{vmatrix} a_{n-1} & a_{n-3} & a_{n-5} \\ a_n & a_{n-2} & a_{n-4} \\ 0 & a_{n-1} & a_{n-3} \end{vmatrix} = a_{n-1}a_{n-2}a_{n-3} + a_n a_{n-1} a_{n-5} - a_n a_{n-3}^2 - a_{n-4} a_{n-1}^2$$

and so on up to Δ_{n-1}.

Hurwitz Criterion: *All the roots of the characteristic equation have negative real parts if and only if* $\Delta_i > 0$, $i = 1, 2, \ldots, n$.

EXAMPLE 5.5. For $n = 3$,

$$\Delta_3 = \begin{vmatrix} a_2 & a_0 & 0 \\ a_3 & a_1 & 0 \\ 0 & a_2 & a_0 \end{vmatrix} = a_2 a_1 a_0 - a_0^2 a_3, \qquad \Delta_2 = \begin{vmatrix} a_2 & a_0 \\ a_3 & a_1 \end{vmatrix} = a_2 a_1 - a_0 a_3, \qquad \Delta_1 = a_2$$

Thus all the roots of the characteristic equation have negative real parts if

$$a_2 > 0 \qquad a_2 a_1 - a_0 a_3 > 0 \qquad a_2 a_1 a_0 - a_0^2 a_3 > 0$$

5.5 CONTINUED FRACTION STABILITY CRITERION

This criterion is applied to the characteristic equation of a continuous system by forming a continued fraction from the odd and even portions of the equation, in the following manner. Let

$$Q(s) \equiv a_n s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0$$

$$Q_1(s) \equiv a_n s^n + a_{n-2}s^{n-2} + \cdots$$

$$Q_2(s) \equiv a_{n-1}s^{n-1} + a_{n-3}s^{n-3} + \cdots$$

Form the fraction Q_1/Q_2, and then divide the denominator into the numerator and invert the remainder, to form a continued fraction as follows:

$$\frac{Q_1(s)}{Q_2(s)} = \frac{a_n s}{a_{n-1}} + \frac{\left(a_{n-2} - \dfrac{a_n a_{n-3}}{a_{n-1}}\right)s^{n-2} + \left(a_{n-4} - \dfrac{a_n a_{n-5}}{a_{n-1}}\right)s^{n-4} + \cdots}{Q_2}$$

$$= h_1 s + \cfrac{1}{h_2 s + \cfrac{1}{h_3 s + \cfrac{1}{h_4 s + \cfrac{1}{\ddots \; \cfrac{1}{h_n s}}}}}$$

If $h_1, h_2, \ldots, h_n$ are all positive, then all the roots of $Q(s) = 0$ have negative real parts.

EXAMPLE 5.6.

$$Q(s) = s^3 + 6s^2 + 12s + 8$$

$$\frac{Q_1(s)}{Q_2(s)} = \frac{s^3 + 12s}{6s^2 + 8} = \frac{1}{6}s + \frac{\frac{32}{3}s}{6s^2 + 8}$$

$$= \frac{1}{6}s + \cfrac{1}{\frac{9}{16}s + \cfrac{1}{\frac{4}{3}s}}$$

Since all the coefficients of s in the continued fraction are positive, that is, $h_1 = \frac{1}{6}$, $h_2 = \frac{9}{16}$, and $h_3 = \frac{4}{3}$, all the roots of the polynomial equation $Q(s) = 0$ have negative real parts.

5.6 STABILITY CRITERIA FOR DISCRETE-TIME SYSTEMS

The stability of discrete systems is determined by the roots of the discrete system characteristic equation

$$Q(z) = a_n z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0 = 0 \tag{5.1}$$

However, in this case the stability region is defined by the *unit circle* $|z| = 1$ in the z-plane. **A necessary and sufficient condition for system stability is that all the roots of the characteristic equation have a magnitude less than one**, that is, be within the *unit circle*. This ensures that the Kronecker delta response decays with time.

A stability criterion for discrete systems similar to the Routh criterion is called the **Jury test**. For this test, the coefficients of the characteristic equation are first arranged in the *Jury array*:

$$
\begin{array}{r|lllllll}
\text{row} & & & & & & & \\
1 & a_0 & a_1 & a_2 & & \cdots & a_{n-1} & a_n \\
2 & a_n & a_{n-1} & a_{n-2} & & \cdots & a_1 & a_0 \\
3 & b_0 & b_1 & b_2 & & \cdots & b_{n-1} & \\
4 & b_{n-1} & b_{n-2} & b_{n-3} & & \cdots & b_0 & \\
5 & c_0 & c_1 & c_2 & \cdots & c_{n-2} & & \\
6 & c_{n-2} & c_{n-3} & c_{n-4} & \cdots & c_0 & & \\
\vdots & \vdots & \vdots & \vdots & & & & \\
2n-5 & r_0 & r_1 & r_2 & r_3 & & & \\
2n-4 & r_3 & r_2 & r_1 & r_0 & & & \\
2n-3 & s_0 & s_1 & s_2 & & & &
\end{array}
$$

where

$$b_k = \begin{vmatrix} a_0 & a_{n-k} \\ a_n & a_k \end{vmatrix} \qquad c_k = \begin{vmatrix} b_0 & b_{n-1-k} \\ b_{n-1} & b_k \end{vmatrix}$$

$$s_0 = \begin{vmatrix} r_0 & r_3 \\ r_3 & r_0 \end{vmatrix} \qquad s_1 = \begin{vmatrix} r_0 & r_2 \\ r_3 & r_1 \end{vmatrix} \qquad s_2 = \begin{vmatrix} r_0 & r_1 \\ r_3 & r_2 \end{vmatrix}$$

The first two rows are written using the characteristic equation coefficients and the next two rows are computed using the determinant relationships shown above. The process is continued with each succeeding pair of rows having one less column than the previous pair until row $2n-3$ is computed, which only has three entries. The array is then terminated.

Jury Test: *Necessary and sufficient conditions for the roots of* $Q(z)=0$ *to have magnitudes less than one are:*

$$
\begin{aligned}
&Q(1) > 0 \\
&Q(-1) \quad \begin{cases} >0 & \text{for } n \text{ even} \\ <0 & \text{for } n \text{ odd} \end{cases} \\
&|a_0| < a_n \\
&|b_0| > |b_{n-1}| \\
&|c_0| > |c_{n-2}| \\
&\quad\vdots \\
&|r_0| > |r_3| \\
&|s_0| > |s_2|
\end{aligned}
$$

Note that if the $Q(1)$ or $Q(-1)$ conditions above are not satisfied, the system is unstable and it is not necessary to construct the array.

EXAMPLE 5.7. For $Q(z) = 3z^4 + 2z^3 + z^2 + z + 1 = 0$ (n even),

$$Q(1) = 3+2+1+1+1 = 8 > 0$$

$$Q(-1) = 3-2+1-1+1 = 2 > 0$$

Thus the Jury array must be completed as

row					
1	1	1	1	2	3
2	3	2	1	1	1
3	−8	−5	−2	−1	
4	−1	−2	−5	−8	
5	63	38	11		

The remaining test condition constraints are therefore

$$|a_0| = 1 < 3 = a_n$$

$$|b_0| = |-8| > |-1| = |b_{n-1}|$$

$$|c_0| = 63 > 11 = |c_{n-2}|$$

Since all the constraints of the Jury test are satisfied, all the roots of the characteristic equation are within the unit circle and the system is stable.

The *w*-Transform

The stability of a linear discrete-time system expressed in the z-domain also can be determined using the s-plane methods developed for continuous systems (e.g., Routh, Hurwitz). The following *bilinear transformation* of the complex variable z into the new complex variable w given by the equivalent expressions:

$$z = \frac{1+w}{1-w} \tag{5.2}$$

$$w = \frac{z-1}{z+1} \tag{5.3}$$

transforms the interior of the unit circle in the z-plane onto the left half of the w-plane. Therefore the stability of a discrete-time system with characteristic polynomial $Q(z)$ can be determined by examining the locations of the roots of

$$Q(w) = Q(z)\big|_{z=(1+w)/(1-w)} = 0$$

in the w-plane, treating w like s and using s-plane techniques to establish stability properties. This transformation is developed more extensively in Chapter 10 and is also used in subsequent frequency domain analysis and design chapters.

EXAMPLE 5.8. The polynomial equation

$$27z^3 + 27z^2 + 9z + 1 = 0$$

is the characteristic equation of a discrete-time system. To test for roots outside the unit circle $|z|=1$, which would signify instability, we set

$$z = \frac{1+w}{1-w}$$

which, after some algebraic manipulation, leads to a new characteristic equation in w:

$$w^3 + 6w^2 + 12w + 8 = 0$$

This equation was found to have roots only in the left half of the complex plane in Example 5.2. Therefore the original discrete-time system is stable.

Solved Problems

STABILITY DEFINITIONS

5.1. The impulse responses of several linear continuous systems are given below. For each case determine if the impulse response represents a stable or an unstable system.

Mathcad

(a) $h(t)=e^{-t}$, (b) $h(t)=te^{-t}$, (c) $h(t)=1$, (d) $h(t)=e^{-t}\sin 3t$, (e) $h(t)=\sin\omega t$.

If the impulse response decays to zero as time approaches infinity, the system is stable. As can be seen in Fig. 5-1, the impulse responses (a), (b), and (d) decay to zero as time approaches infinity and therefore

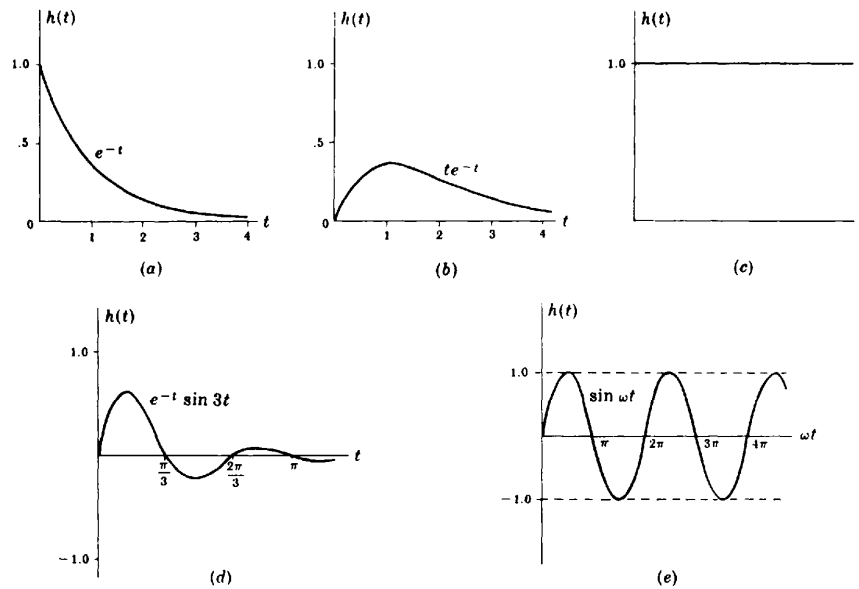


Fig. 5-1

represent *stable* systems. Since the impulse responses (c) and (e) do not approach zero, they represent *unstable* systems.

5.2. If a step function is applied at the input of a continuous system and the output remains below a certain level for all time, is the system stable?

The system is not necessarily stable since the output must be bounded for every bounded input. A bounded output to one specific bounded input does not ensure stability.

5.3. If a step function is applied at the input of a continuous system and the output is of the form $y = t$, is the system stable or unstable?

The system is unstable since a bounded input produced an unbounded output.

CHARACTERISTIC ROOT LOCATIONS FOR CONTINUOUS SYSTEMS

5.4. The roots of the characteristic equations of several systems are given below. Determine in each case if the set of roots represents stable, marginally stable, or unstable systems.

(a) $-1, -2$	(d) $-1+j, -1-j$	(g) $-6, -4, 7$
(b) $-1, +1$	(e) $-2+j, -2-j, 2j, -2j$	(h) $-2+3j, -2-3j, -2$
(c) $-3, -2, 0$	(f) $2, -1, -3$	(i) $-j, j, -1, 1$

The sets of roots (a), (d), and (h) represent stable systems since all the roots have negative real parts. The sets of roots (c) and (e) represent marginally stable systems since all the roots have nonpositive real parts, that is, zero or negative. The sets (b), (f), (g), and (i) represent unstable systems since each has at least one root with a positive real part.

5.5. A system has poles at -1 and -5 and zeros at 1 and -2. Is the system stable?

The system is stable since the poles are the roots of the system characteristic equation (Chapter 3) which have negative real parts. The fact that the system has a zero with a positive real part does not affect its stability.

5.6. Determine if the system with the following characteristic equation is stable:

$$(s+1)(s+2)(s-3)=0.$$

This characteristic equation has the roots -1, -2, and 3 and therefore represents an unstable system since there is a positive real root.

5.7. The differential equation of an integrator may be written as follows: $dy/dt = u$. Determine if an integrator is stable.

The characteristic equation of this system is $s = 0$. Since the root does not have a negative real part, an integrator is not stable. Since it has no roots with positive real parts, an integrator is marginally stable.

5.8. Determine a bounded input which will produce an unbounded output from an integrator.

The input $u = 1$ will produce the output $y = t$, which is unbounded.

ROUTH STABILITY CRITERION

5.9. Determine if the following characteristic equation represents a stable system:

$$s^3 + 4s^2 + 8s + 12 = 0$$

The Routh table for this system is

$$\begin{array}{c|cc} s^3 & 1 & 8 \\ s^2 & 4 & 12 \\ s^1 & 5 & 0 \\ s^0 & 12 & \end{array}$$

Since there are no changes of sign in the first column, all the roots of the characteristic equation have negative real parts and the system is stable.

5.10. Determine if the following characteristic equation has any roots with positive real parts:

$$s^4 + s^3 - s - 1 = 0$$

Note that the coefficient of the s^2 term is zero. The Routh table for this equation is

$$\begin{array}{r|ccc} s^4 & 1 & 0 & -1 \\ s^3 & 1 & -1 & 0 \\ s^2 & 1 & -1 & \\ s^1 & 0 & 0 & \\ \text{new } s^1 & 2 & 0 & \\ s^0 & -1 & & \end{array}$$

The presence of the zeros in the s^1 row indicates that the characteristic equation has two roots which satisfy the auxiliary equation formed from the s^2 row as follows: $s^2 - 1 = 0$. The roots of this equation are $+1$ and -1.

The new s^1 row was formed using the coefficients from the derivative of the auxiliary equation: $2s - 0 = 0$. Since there is one change of sign, the characteristic equation has one root with a positive real part, the one at $+1$ determined from the auxiliary equation.

5.11. The characteristic equation of a given system is

$$s^4 + 6s^3 + 11s^2 + 6s + K = 0$$

What restrictions must be placed upon the parameter K in order to ensure that the system is stable?

The Routh table for this system is

$$\begin{array}{c|ccc}
s^4 & 1 & 11 & K \\
s^3 & 6 & 6 & 0 \\
s^2 & 10 & K & 0 \\
s^1 & \dfrac{60-6K}{10} & 0 & \\
s^0 & K & &
\end{array}$$

For the system to be stable, $60 - 6K > 0$, or $K < 10$, and $K > 0$. Thus $0 < K < 10$.

5.12. Construct a Routh table and determine the number of roots with positive real parts for the equation

$$2s^3 + 4s^2 + 4s + 12 = 0$$

The Routh table for this equation is given below. Here the s^2 row was divided by 4 before the s^1 row was computed. The s^1 row was then divided by 2 before the s^0 row was computed.

$$\begin{array}{c|cc}
s^3 & 2 & 4 \\
s^2 & 1 & 3 \\
s^1 & -1 & 0 \\
s^0 & 3 &
\end{array}$$

Since there are two changes of sign in the first column of the Routh table, the equation above has two roots with positive real parts.

HURWITZ STABILITY CRITERION

5.13. Determine if the characteristic equation below represents a stable or an unstable system.

$$s^3 + 8s^2 + 14s + 24 = 0$$

The Hurwitz determinants for this system are

$$\Delta_3 = \begin{vmatrix} 8 & 24 & 0 \\ 1 & 14 & 0 \\ 0 & 8 & 24 \end{vmatrix} = 2112 \qquad \Delta_2 = \begin{vmatrix} 8 & 24 \\ 1 & 14 \end{vmatrix} = 88 \qquad \Delta_1 = 8$$

Since each determinant is positive, the system is stable. Note that the general formulation of Example 5.5 could have been used to check the stability in this case by substituting the appropriate values for the coefficients a_0, a_1, a_2, and a_3.

5.14. For what range of values of K is the system with the following characteristic equation stable?

$$s^2 + Ks + 2K - 1 = 0$$

The Hurwitz determinants for this system are

$$\Delta_2 = \begin{vmatrix} K & 0 \\ 1 & 2K-1 \end{vmatrix} = 2K^2 - K = K(2K-1) \qquad \Delta_1 = K$$

In order for these determinants to be positive, it is necessary that $K > 0$ and $2K - 1 > 0$. Thus the system is stable if $K > \frac{1}{2}$.

5.15. A system is designed to give satisfactory performance when a particular amplifier gain $K = 2$. Determine how much K can vary before the system becomes unstable if the characteristic equation is

$$s^3 + (4+K)s^2 + 6s + 16 + 8K = 0$$

Substituting the coefficients of the given equation into the general Hurwitz conditions of Example 5.5 results in the following requirements for stability:

$$4 + K > 0 \qquad (4+K)6 - (16+8K) > 0 \qquad (4+K)(6)(16+8K) - (16+8K)^2 > 0$$

Assuming the amplifier gain K cannot be negative, the first condition is satisfied. The second and third conditions are satisfied if K is less than 4. Hence with an amplifier gain design value of 2, the system could tolerate an increase in gain of a factor of 2 before it would become unstable. The gain could also drop to zero without causing instability.

5.16. Determine the Hurwitz conditions for stability of the following general fourth-order characteristic equation, assuming a_4 is positive.

$$a_4s^4 + a_3s^3 + a_2s^2 + a_1s + a_0 = 0$$

The Hurwitz determinants are

$$\Delta_4 = \begin{vmatrix} a_3 & a_1 & 0 & 0 \\ a_4 & a_2 & a_0 & 0 \\ 0 & a_3 & a_1 & 0 \\ 0 & a_4 & a_2 & a_0 \end{vmatrix} = a_3\left(a_2a_1a_0 - a_3a_0^2\right) - a_1^2a_0a_4$$

$$\Delta_3 = \begin{vmatrix} a_3 & a_1 & 0 \\ a_4 & a_2 & a_0 \\ 0 & a_3 & a_1 \end{vmatrix} = a_3a_2a_1 - a_0a_3^2 - a_4a_1^2$$

$$\Delta_2 = \begin{vmatrix} a_3 & a_1 \\ a_4 & a_2 \end{vmatrix} = a_3a_2 - a_4a_1$$

$$\Delta_1 = a_3$$

The conditions for stability are then

$$a_3 > 0 \qquad a_3a_2 - a_4a_1 > 0 \qquad a_3a_2a_1 - a_0a_3^2 - a_4a_1^2 > 0 \qquad a_3\left(a_2a_1a_0 - a_3a_0^2\right) - a_1^2a_0a_4 > 0$$

5.17. Is the system with the following characteristic equation stable?

$$s^4 + 3s^3 + 6s^2 + 9s + 12 = 0$$

Substituting the appropriate values for the coefficients in the general conditions of Problem 5.16, we have

$$3 > 0 \qquad 18 - 9 > 0 \qquad 162 - 108 - 81 \ngtr 0 \qquad 3(648 - 432) - 972 \ngtr 0$$

Since the last two conditions are not satisfied, the system is unstable.

CONTINUED FRACTION STABILITY CRITERION

5.18. Repeat Problem 5.9 using the continued fraction stability criterion.

The polynomial $Q(s) = s^3 + 4s^2 + 8s + 12$ is divided into the two parts:

$$Q_1(s) = s^3 + 8s \qquad Q_2(s) = 4s^2 + 12$$

The continued fraction for $Q_1(s)/Q_2(s)$ is

$$\frac{Q_1(s)}{Q_2(s)} = \frac{s^3+8s}{4s^2+12} = \frac{1}{4}s + \frac{5s}{4s^2+12} = \frac{1}{4}s + \cfrac{1}{\frac{4}{5}s + \cfrac{1}{\frac{5}{12}s}}$$

Since all the coefficients of s are positive, the polynomial has all its roots in the left half-plane and the system with the characteristic equation $Q(s)=0$ is stable.

5.19. Determine bounds upon the parameter K for which a system with the following characteristic equation is stable:

$$s^3 + 14s^2 + 56s + K = 0$$

$$\frac{Q_1(s)}{Q_2(s)} = \frac{s^3+56s}{14s^2+K} = \frac{1}{14}s + \frac{(56-K/14)s}{14s^2+K} = \frac{1}{14}s + \cfrac{1}{\left[\frac{14}{56-K/14}\right]s + \cfrac{1}{\left[\frac{56-K/14}{K}\right]s}}$$

For the system to be stable, the following conditions must be satisfied: $56 - K/14 > 0$ and $K > 0$, that is, $0 < K < 784$.

5.20. Derive conditions for all the roots of a general third-order polynomial to have negative real parts.

For $Q(s) = a_3s^3 + a_2s^2 + a_1s + a_0$,

$$\frac{Q_1(s)}{Q_2(s)} = \frac{a_3s^3+a_1s}{a_2s^2+a_0} = \frac{a_3}{a_2}s + \frac{[a_1 - a_3a_0/a_2]s}{a_2s^2+a_0} = \frac{a_3}{a_2}s + \cfrac{1}{\left[\frac{a_2}{a_1-a_3a_0/a_2}\right]s + \cfrac{1}{\left[\frac{a_1-a_3a_0/a_2}{a_0}\right]s}}$$

The conditions for all the roots of $Q(s)$ to have negative real parts are then

$$\frac{a_3}{a_2} > 0 \qquad \frac{a_2}{a_1 - a_3a_0/a_2} > 0 \qquad \frac{a_1 - a_3a_0/a_2}{a_0} > 0$$

Thus if a_3 is positive, the required conditions are $a_2, a_1, a_0 > 0$ and $a_1a_2 - a_3a_0 > 0$. Note that if a_3 is not positive, $Q(s)$ should be multiplied by -1 before checking the above conditions.

5.21. Is the system with the following characteristic equation stable?

$$s^4 + 4s^3 + 8s^2 + 16s + 32 = 0$$

$$\frac{Q_1(s)}{Q_2(s)} = \frac{s^4+8s^2+32}{4s^3+16s} = \frac{1}{4}s + \frac{4s^2+32}{4s^3+16s}$$

$$= \frac{1}{4}s + \cfrac{1}{s + \cfrac{-16s}{4s^2+32}} = \frac{1}{4}s + \cfrac{1}{s + \cfrac{1}{-\frac{1}{4}s + \cfrac{1}{-\frac{1}{2}s}}}$$

Since the coefficients of s are not all positive, the system is unstable.

DISCRETE-TIME SYSTEMS

5.22. Is the system with the following characteristic equation stable?

$$Q(z) = z^4 + 2z^3 + 3z^2 + z + 1 = 0$$

Applying the Jury test, with $n = 4$ (even),

$$Q(1) = 1 + 2 + 3 + 1 + 1 = 8 > 0$$
$$Q(-1) = 1 - 2 + 3 - 1 + 1 = 2 > 0$$

The Jury array must be constructed, as follows:

row					
1	1	1	3	2	1
2	1	2	3	1	1
3	0	−1	0	1	
4	1	0	−1	0	
5	−1	1	0		

The Jury test constraints are

$$|a_0| = 1 \not< 1 = a_n$$
$$|b_0| = 0 \not> 1 = |b_{n-1}|$$
$$|c_0| = |-1| > 0 = |c_{n-2}|$$

Since all the constraints are not satisfied, the system is unstable.

5.23. Is the system with the following characteristic equation stable?

$$Q(z) = 2z^4 + 2z^3 + 3z^2 + z + 1 = 0$$

Applying the Jury test, with $n = 4$ (even),

$$Q(1) = 2 + 2 + 3 + 1 + 1 = 9 > 0$$
$$Q(-1) = 2 - 2 + 3 - 1 + 1 = 3 > 0$$

The Jury array must be constructed, as follows:

row					
1	1	1	3	2	2
2	2	2	3	1	1
3	3	3	2	0	
4	0	2	3	3	
5	9	7	0		

The test constraints are

$$|a_0| = 1 < 2 = a_n$$
$$|b_0| = 3 > 0 = |b_{n-1}|$$
$$|c_0| = 9 > 0 = |c_{n-2}|$$

Since all the constraints are satisfied, the system is stable.

5.24. Is the system with the following characteristic equation stable?

$$Q(z) = z^5 + 3z^4 + 3z^3 + 3z^2 + 2z + 1 = 0$$

Applying the Jury test, with $n = 5$ (odd),

$$Q(1) = 1 + 3 + 3 + 3 + 2 + 1 = 13 > 0$$
$$Q(-1) = -1 + 3 - 3 + 3 - 2 + 1 = 1 > 0$$

Since n is odd, $Q(-1)$ must be less than zero for the system to be stable. Therefore the system is unstable.

MISCELLANEOUS PROBLEMS

5.25. If a zero appears in the first column of the Routh table, is the system necessarily unstable?

Strictly speaking, a zero in the first column must be interpreted as having no sign, that is, neither positive nor negative. Consequently, all the elements of the first column cannot have the same sign if one of them is zero, and the system is unstable. In some cases, a zero in the first column indicates the presence of two roots of equal magnitude but opposite sign (see Problem 5.10). In other cases, it indicates the presence of one or more roots with zero real parts. Thus a characteristic equation having one or more roots with zero real parts and no roots with positive real parts will produce a Routh table in which all the elements of the first column do not have the same sign and do not have any sign changes.

5.26. Prove that a continuous system is unstable if any coefficients of the characteristic equation are zero.

The characteristic equation may be written in the form

$$(s-s_1)(s-s_2)(s-s_3)\cdots(s-s_n)=0$$

where $s_1, s_2, \ldots, s_n$ are the roots of the equation. If this equation is multiplied out, n new equations can be obtained relating the roots and the coefficients of the characteristic equation in the usual form. Thus

$$a_n s^n + a_{n-1}s^{n-1} + \cdots + a_0 = 0 \qquad \text{or} \qquad s^n + \frac{a_{n-1}}{a_n}s^{n-1} + \cdots + \frac{a_0}{a_n} = 0$$

and the relations are

$$\frac{a_{n-1}}{a_n} = -\sum_{i=1}^{n} s_i, \quad \frac{a_{n-2}}{a_n} = \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} s_i s_j, \quad \frac{a_{n-3}}{a_n} = -\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\ i\neq j\neq k}}^{n} s_i s_j s_k, \ldots, \frac{a_0}{a_n} = (-1)^n s_1 s_2 \cdots s_n$$

The coefficients $a_{n-1}, a_{n-2}, \ldots, a_0$ all have the same sign as a_n and are nonzero if all the roots $s_1, s_2, \ldots, s_n$ have negative real parts. The only way any one of the coefficients can be zero is for one or more of the roots to have zero or positive real parts. In either case, the system would be unstable.

5.27. Prove that a continuous system is unstable if all the coefficients of the characteristic equation do not have the same sign.

From the relations presented in Problem 5.26, it can be seen that the coefficients $a_{n-1}, a_{n-2}, \ldots, a_0$ have the same sign as a_n if all the roots $s_1, s_2, \ldots, s_n$ have negative real parts. The only way any of these coefficients may differ in sign from a_n is for one or more of the roots to have a positive real part. Thus the system is necessarily unstable if all the coefficients do not have the same sign. Note that a system is *not* necessarily stable if all the coefficients do have the same sign.

5.28. Can the continuous system stability criteria presented in this chapter be applied to continuous systems which contain time delays?

No they cannot be directly applied because systems which contain time delays do not have characteristic equations of the required form, that is, finite polynomials in s. For example, the following characteristic equation represents a system which contains a time delay:

$$s^2 + s + e^{-sT} = 0$$

Strictly speaking, this equation has an infinite number of roots. However, in some cases an approximation may be employed for e^{-sT} to give useful, although not entirely accurate, information concerning system stability. To illustrate, let e^{-sT} in the equation above be replaced by the first two terms of its Taylor series. The equation then becomes

$$s^2 + s + 1 - sT = 0 \qquad \text{or} \qquad s^2 + (1-T)s + 1 = 0$$

One of the stability criteria of this chapter may then be applied to this approximation of the characteristic equation.

5.29. Determine an approximate upper limit on the time delay in order that the system discussed in the solution of Problem 5.28 be stable.

Employing the approximate equation $s^2 + (1-T)s + 1 = 0$, the Hurwitz determinants are $\Delta_1 = \Delta_2 = 1 - T$. Hence for the system to be stable, the time delay T must be less than 1.

Supplementary Problems

5.30. For each characteristic polynomial, determine if it represents a stable or an unstable system.

Mathcad

(*a*) $2s^4 + 8s^3 + 10s^2 + 10s + 20$ (*c*) $s^5 + 6s^4 + 10s^2 + 5s + 24$ (*e*) $s^4 + 8s^3 + 24s^2 + 32s + 16$

(*b*) $s^3 + 7s^2 + 7s + 46$ (*d*) $s^3 - 2s^2 + 4s + 6$ (*f*) $s^6 + 4s^4 + 8s^2 + 16$

5.31. For what values of K does the polynomial $s^3 + (4+K)s^2 + 6s + 12$ have roots with negative real parts?

5.32. How many roots with positive real parts does each polynomial have?

(*a*) $s^3 + s^2 - s + 1$ (*b*) $s^4 + 2s^3 + 2s^2 + 2s + 1$ (*c*) $s^3 + s^2 - 2$ (*d*) $s^4 - s^2 - 2s + 2$

(*e*) $s^3 + s^2 + s + 6$

5.33. For what positive value of K does the polynomial $s^4 + 8s^3 + 24s^2 + 32s + K$ have roots with zero real parts? What are these roots?

Answers to Supplementary Problems

5.30. (*b*) and (*e*) represent stable systems; (*a*), (*c*), (*d*), and (*f*) represent unstable systems.

5.31. $K > -2$

5.32. (*a*) 2, (*b*) 0, (*c*) 1, (*d*) 2, (*e*) 2

5.33. $K = 80$; $s = \pm j2$

Chapter 6

Transfer Functions

6.1 DEFINITION OF A CONTINUOUS SYSTEM TRANSFER FUNCTION

As shown in Chapters 3 and 4, the response of a time-invariant linear system can be separated into two parts: the forced response and the free response. This is true for both continuous and discrete systems. We consider continuous transfer functions first, and for single-input, single-output systems only. Equation (*4.8*) clearly illustrates this division for the most general constant-coefficient, linear, ordinary differential equation. The forced response includes terms due to initial values u_0^k of the input, and the free response depends only on initial conditions y_0^k on the output. If terms due to *all* initial values, that is, u_0^k *and* y_0^k, are lumped together, Equation (*4.8*) can be written as

$$y(t)=\mathscr{L}^{-1}\left[\left(\sum_{i=0}^{m} b_i s^i \bigg/ \sum_{i=0}^{n} a_i s^i\right)U(s)+\left(\text{terms due to } all \text{ initial values } u_0^k,\ y_0^k\right)\right]$$

or, in transform notation, as

$$Y(s)=\left(\sum_{i=0}^{m} b_i s^i \bigg/ \sum_{i=0}^{n} a_i s^i\right)U(s)+\left(\text{terms due to } all \text{ initial values } u_0^k,\ y_0^k\right)$$

The **transfer function** $P(s)$ of a continuous system is defined as that factor in the equation for $Y(s)$ multiplying the transform of the input $U(s)$. For the system described above, the transfer function is

$$P(s)=\sum_{i=0}^{m} b_i s^i \bigg/ \sum_{i=0}^{n} a_i s^i=\frac{b_m s^m+b_{m-1}s^{m-1}+\cdots+b_0}{a_n s^n+a_{n-1}s^{n-1}+\cdots+a_0}$$

the denominator is the characteristic polynomial, and the transform of the response may be rewritten as

$$Y(s)=P(s)U(s)+\left(\text{terms due to } all \text{ initial values } u_0^k,\ y_0^k\right)$$

If the quantity (terms due to *all* initial values u_0^k, y_0^k) is zero, the Laplace transform of the output $Y(s)$ in response to an input $U(s)$ is given by

$$Y(s)=P(s)U(s)$$

If the system is at rest prior to application of the input, that is, $d^k y/dt^k=0$, $k=0,1,\dots,n-1$, for $t<0$, then

$$\left(\text{terms due to } all \text{ initial values } u_0^k,\ y_0^k\right)=0$$

and the output as a function of time $y(t)$ is simply the inverse transform of $P(s)U(s)$.

It is emphasized that not all transfer functions are rational algebraic expressions. For example, the transfer function of a continuous system including time delays contains terms of the form e^{-sT} (e.g., Problem 5.28). The transfer function of an element representing a pure time delay is $P(s)=e^{-sT}$, where T is the time delay in units of time.

Since the formation of the output transform $Y(s)$ is purely an algebraic multiplication of $P(s)$ and $U(s)$ when (terms due to *all* initial values u_0^k, y_0^k) $=0$, the multiplication is commutative; that is,

$$Y(s)=U(s)P(s)=P(s)U(s) \tag{6.1}$$

6.2 PROPERTIES OF A CONTINUOUS SYSTEM TRANSFER FUNCTION

The transfer function of a continuous system has several useful properties:

1. It is the Laplace transform of its impulse response $y_\delta(t)$, $t\ge 0$. That is, if the input to a system with transfer function $P(s)$ is an impulse and all initial values are zero the transform of the output is $P(s)$.
2. The system transfer function can be determined from the system differential equation by taking the Laplace transform and ignoring all terms arising from initial values. The transfer function $P(s)$ is then given by
$$P(s)=\frac{Y(s)}{U(s)}$$
3. The system differential equation can be obtained from the transfer function by replacing the s variable with the differential operator D defined by $D\equiv d/dt$.
4. The stability of a time-invariant linear system can be determined from the characteristic equation (see Chapter 5). The denominator of the system transfer function is the *characteristic polynomial*. Consequently, for continuous systems, if all the roots of the denominator have negative real parts, the system is stable.
5. The roots of the denominator are the system poles and the roots of the numerator are the system zeros (see Chapter 4). The system transfer function can then be specified to within a constant by specifying the system poles and zeros. This constant, usually denoted by K, is the system **gain factor**. As was described in Chapter 4, Section 4.11, the system poles and zeros can be represented schematically by a pole-zero map in the s-plane.
6. If the system transfer function has no poles or zeros with positive real parts, the system is a **minimum phase** system.

EXAMPLE 6.1. Consider the system with the differential equation $dy/dt+2y=du/dt+u$.

The Laplace transform version of this equation with all initial values set equal to zero is $(s+2)Y(s)=(s+1)U(s)$.

The system transfer function is thus given by $P(s)=Y(s)/U(s)=(s+1)/(s+2)$.

EXAMPLE 6.2. Given $P(s)=(2s+1)/(s^2+s+1)$, the system differential equation is

$$y=\left[\frac{2D+1}{D^2+D+1}\right]u \qquad\text{or}\qquad D^2y+Dy+y=2Du+u \qquad\text{or}\qquad \frac{d^2y}{dt^2}+\frac{dy}{dt}+y=2\frac{du}{dt}+u$$

EXAMPLE 6.3. The transfer function $P(s)=K(s+a)/(s+b)(s+c)$ can be specified by giving the zero location $-a$, the pole locations $-b$ and $-c$, and the gain factor K.

6.3 TRANSFER FUNCTIONS OF CONTINUOUS CONTROL SYSTEM COMPENSATORS AND CONTROLLERS

The transfer functions of four common control system components are presented below. Typical mechanizations of three of these transfer functions, using R-C networks, are presented in the solved problems.

EXAMPLE 6.4. The general transfer function of a **continuous system lead compensator** is

$$P_{\text{Lead}}(s)=\frac{s+a}{s+b} \qquad b>a \tag{6.2}$$

This compensator has a zero at $s=-a$ and a pole at $s=-b$.

EXAMPLE 6.5. The general transfer function of a **continuous system lag compensator** is

$$P_{\mathrm{Lag}}(s)=\frac{a(s+b)}{b(s+a)}\qquad b>a \tag{6.3}$$

However, in this case the zero is at $s=-b$ and the pole is at $s=-a$. The gain factor a/b is included because of the way it is usually mechanized (Problem 6.13).

EXAMPLE 6.6. The general transfer function of a **continuous system lag-lead compensator** is

$$P_{\mathrm{LL}}(s)=\frac{(s+a_1)(s+b_2)}{(s+b_1)(s+a_2)}\qquad b_1>a_1,\ b_2>a_2 \tag{6.4}$$

This compensator has two zeros and two poles. For mechanization considerations, the restriction $a_1b_2=b_1a_2$ is usually imposed (Problem 6.14).

EXAMPLE 6.7. The transfer function of the **PID controller** of Example 2.14 is

$$P_{\mathrm{PID}}(s)\equiv\frac{U_{\mathrm{PID}}(s)}{E(s)}=K_P+K_Ds+\frac{K_I}{s}=\frac{K_Ds^2+K_Ps+K_I}{s} \tag{6.5}$$

This controller has two zeros and one pole. It is similar to the lag-lead compensator of the previous example except that the smallest pole is at the origin (an integrator) and it does not have the second pole. It is typically mechanized in an analog or digital computer.

6.4 CONTINUOUS SYSTEM TIME RESPONSE

The Laplace transform of the response of a continuous system to a specific input is given by

$$Y(s)=P(s)U(s)$$

when all initial conditions are zero. The inverse transform $y(t)=\mathscr{L}^{-1}[P(s)U(s)]$ is then the time response and $y(t)$ may be determined by finding the poles of $P(s)U(s)$ and evaluating the residues at these poles (when there are no multiple poles). Therefore $y(t)$ depends on both the poles and zeros of the transfer function and the poles and zeros of the input.

The residues can be determined graphically from a *pole-zero map* of $Y(s)$, constructed from the pole-zero map of $P(s)$ by simply adding the poles and zeros of $U(s)$. Graphical evaluation of the residues may then be performed as described in Chapter 4, Section 4.12.

6.5 CONTINUOUS SYSTEM FREQUENCY RESPONSE

The steady state response of a continuous system to sinusoidal inputs can be determined from the system transfer function. For the special case of a step function input of amplitude A, often called a **d.c. input**, the Laplace transform of the system output is given by

$$Y(s)=P(s)\frac{A}{s}$$

If the system is stable, the steady state response is a step function of amplitude $AP(0)$, since this is the residue at the input pole. The amplitude of the input signal is thus multiplied by $P(0)$ to determine the amplitude of the output. $P(0)$ is therefore the **d.c. gain** of the system.

Note that for an unstable system such as an integrator ($P(s)=1/s$), a steady state response does not always exist. If the input to an integrator is a step function, the output is a ramp, which is unbounded (see Problems 5.7 and 5.8). For this reason, integrators are sometimes said to have infinite d.c. gain.

The steady state response of a stable system to an input $u=A\sin\omega t$ is given by

$$y_{ss}=A|P(j\omega)|\sin(\omega t+\phi)$$

where $|P(j\omega)|$ = magnitude of $P(j\omega)$, $\phi=\arg P(j\omega)$, and the complex number $P(j\omega)$ is determined

from $P(s)$ by replacing s by $j\omega$ (see Problem 6.20). The system output has the same frequency as the input and can be obtained by multiplying the magnitude of the input by $|P(j\omega)|$ and shifting the phase angle of the input by $\arg P(j\omega)$. The magnitude $|P(j\omega)|$ and angle $\arg P(j\omega)$ for all ω together define **the system frequency response**. The magnitude $|P(j\omega)|$ is the *gain* of the system for sinusoidal inputs with frequency ω.

The system frequency response can be determined graphically in the s-plane from a pole-zero map of $P(s)$ in the same manner as the graphical calculation of residues. In this instance, however, the magnitude and phase angle of $P(s)$ are computed at a point on the $j\omega$ axis by measuring the magnitudes and angles of the vectors drawn from the poles and zeros of $P(s)$ to the point on the $j\omega$ axis.

EXAMPLE 6.8. Consider the system with the transfer function

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

Referring to Fig. 6-1, the magnitude and angle of $P(j\omega)$ for $\omega=1$ are computed in the s-plane as follows. The magnitude of $P(j1)$ is

$$|P(j1)|=\frac{1}{\sqrt{5}\cdot\sqrt{2}}=0.316$$

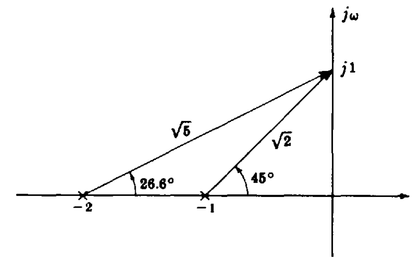


Fig. 6-1

and the angle is

$$\arg P(j1)=-26.6^\circ-45^\circ=-71.6^\circ$$

EXAMPLE 6.9. The system frequency response is usually represented by two graphs (see Fig. 6-2): one of $|P(j\omega)|$ as a function of ω and one of $\arg P(j\omega)$ as a function of ω. For the transfer function of Example 6.8, $P(s)=1/(s+1)(s+2)$, these graphs are easily determined by plotting the values of $|P(j\omega)|$ and $\arg P(j\omega)$ for several values of ω as shown below.

ω	0	0.5	1.0	2.0	4.0	8.0
$\lvert P(j\omega)\rvert$	0.5	0.433	0.316	0.158	0.054	0.015
$\arg P(j\omega)$	0	-40.6°	-71.6°	-108.5°	-139.4°	-158.9°

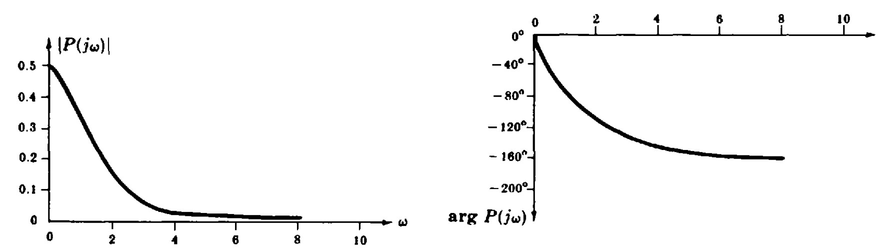


Fig. 6-2

6.6 DISCRETE-TIME SYSTEM TRANSFER FUNCTIONS, COMPENSATORS AND TIME RESPONSES

The transfer function $P(z)$ for a discrete-time system is defined as that factor in the equation for the transform of the output $Y(z)$ that multiplies the transform of the input $U(z)$. If all terms due to initial conditions are zero, then the system response to an input $U(z)$ is given by: $Y(z) = P(z)U(z)$ in the z-domain, and $\{y(k)\} = \mathcal{Z}^{-1}[P(z)U(z)]$ in the time-domain.

The transfer function of a discrete-time system has the following properties:

1. $P(z)$ is the z-transform of its Kronecker delta response $y_\delta(k)$, $k = 0, 1, \ldots$.
2. The system difference equation can be obtained from $P(z)$ by replacing the z variable with the shift operator Z defined for any integers k and n by
$$Z^n[y(k)] = y(k+n) \tag{6.6}$$
3. The denominator of $P(z)$ is the system *characteristic polynomial.* Consequently, if all the roots of the denominator are within the unit circle of the z-plane, the system is stable.
4. The roots of the denominator are system poles and the roots of the numerator are the system zeros. $P(z)$ can be specified by specifying the system poles and zeros and the gain factor K:
$$P(z) = \frac{K(z+z_1)(z+z_2)\cdots(z+z_m)}{(z+p_1)(z+p_2)\cdots(z+p_n)} \tag{6.7}$$
The system poles and zeros can be represented schematically by a pole-zero map in the z-plane. The pole-zero map of the output response can be constructed from the pole-zero map of $P(z)$ by including the poles and zeros of the input $U(z)$.
5. The order of the denominator polynomial of the transfer function of a causal (physically realizable) discrete-time system must be greater than or equal to the order of the numerator polynomial.
6. The steady state response of a discrete-time system to a unit step input is called the **d.c. gain** and is given by the Final Value Theorem (Section 4.9):
$$\lim_{k\to\infty} y(k) = \lim_{z\to 1}\left[\frac{z-1}{z}P(z)\frac{z}{z-1}\right] = P(1) \tag{6.8}$$

EXAMPLE 6.10. Consider a discrete-time system characterized by the difference equation
$$y(k+2) + 1.1y(k+1) + 0.3y(k) = u(k+2) + 0.2u(k+1)$$
The z-transform version of this equation with all initial conditions set equal to zero is
$$(z^2 + 1.1z + 0.3)Y(z) = (z^2 + 0.2z)U(z)$$
The system transfer function is given by
$$P(z) = \frac{z(z+0.2)}{z^2+1.1z+0.3} = \frac{z(z+0.2)}{(z+0.5)(z+0.6)}$$
This system has a zero at -0.2 and two poles, at -0.5 and -0.6. Since the poles are inside the unit circle, the system is stable. The d.c. gain is
$$P(1) = \frac{1(1.2)}{(1.5)(1.6)} = 0.5$$

EXAMPLE 6.11. The general transfer function of a **digital lead compensator** is
$$P_{\text{Lead}}(z) = \frac{K_{\text{Lead}}(z-z_c)}{z-p_c} \qquad z_c > p_c \tag{6.9}$$
This compensator has a zero at $z = z_c$ and a pole at $z = p_c$. Its steady state gain is
$$P_{\text{Lead}}(1) = \frac{K_{\text{Lead}}(1-z_c)}{1-p_c} \tag{6.10}$$

The gain factor K_{Lead} is included in the transfer function to adjust its gain at a given ω to a desired value. In Problem 12.13, for example, K_{Lead} is chosen to render the steady state gain of P_{Lead} (at $\omega = 0$) equal to that of its analog counterpart.

EXAMPLE 6.12. The general transfer function of a **digital lag compensator** is
$$P_{\text{Lag}}(z) = \frac{(1-p_c)(z-z_c)}{(1-z_c)(z-p_c)} \qquad z_c < p_c \tag{6.11}$$
This compensator has a zero at $z = z_c$ and a pole at $z = p_c$. The gain factor $(1-p_c)/(1-z_c)$ is included so that the low frequency or steady state gain $P_{\text{Lag}}(1) = 1$, analogous to the continuous-time lag compensator.

EXAMPLE 6.13. Digital lag and lead compensators can be designed directly from s-domain specifications by using the transform between the s- and z-domains defined by $z = e^{sT}$. That is, the poles and zeros of
$$P_{\text{Lead}}(s) = \frac{s+a}{s+b} \qquad \text{and} \qquad P_{\text{Lag}} = \frac{a(s+b)}{b(s+a)}$$
can be mapped according to $z = e^{sT}$. For the lead compensator, the zero at $s = -a$ maps into the zero at $z = z_c \equiv e^{-aT}$, and the pole at $s = -b$ maps into the pole at $z = p_c \equiv e^{-bT}$. This gives
$$P'_{\text{Lead}}(z) = \frac{z-e^{-aT}}{z-e^{-bT}} \tag{6.12}$$
Similarly,
$$P'_{\text{Lag}}(z) = \left(\frac{1-e^{-aT}}{1-e^{-bT}}\right)\left(\frac{z-e^{-bT}}{z-e^{-aT}}\right) \tag{6.13}$$
Note that $P'_{\text{Lag}}(1) = 1$.

This transformation is only one of many possible for digital lead and lag compensators, or any type of compensators for that matter. Another variant of the lead compensator is illustrated in Problems 12.13 through 12.15.

An example of how Equation (6.13) can be used in applications is given in Example 12.7.

6.7 DISCRETE-TIME SYSTEM FREQUENCY RESPONSE

The steady state response to an input sequence $\{u(k) = A\sin\omega kT\}$ of a stable discrete-time system with transfer function $P(z)$ is given by
$$y_{ss} = A|P(e^{j\omega T})|\sin(\omega kT + \phi) \qquad k = 0, 1, 2, \ldots \tag{6.14}$$
where $|P(e^{j\omega T})|$ is the magnitude of $P(e^{j\omega T})$, $\phi = \arg P(e^{j\omega T})$, and the complex function $P(e^{j\omega T})$ is determined from $P(z)$ by replacing z by $e^{j\omega T}$ (see Problem 6.40). The system output is a sequence of samples of a sinusoid with the same frequency as the input sinusoid. The output sequence is obtained by multiplying the magnitude A of the input by $|P(e^{j\omega T})|$ and shifting the phase angle of the input by $\arg P(e^{j\omega T})$. The magnitude $|P(e^{j\omega T})|$ and phase angle $\arg P(e^{j\omega T})$, for all ω, together define the **discrete-time system frequency response function.** The magnitude $|P(e^{j\omega T})|$ is the **gain** of the system for sinusoidal inputs with angular frequency ω.

A discrete-time system frequency response function can be determined in the z-plane from a pole-zero map of $P(z)$ in the same manner as the graphical calculation of residues (Section 4.12). In this instance, however, the magnitude and phase angle are computed on the $e^{j\omega T}$ circle (the unit circle), by measuring the magnitude and angle of the vectors drawn from the poles and zeros of P to the point on the unit circle. Since $P(e^{j\omega T})$ is periodic in ω, with period $2\pi/T$, the frequency response function need only be determined over the angular frequency range $-\pi/T \le \omega \le \pi/T$. Also, since the magnitude function is an even function of ω, and the phase angle is an odd function of ω, actual computations need only be performed over half this angular frequency range, that is, $0 \le \omega \le \pi/T$.

6.8 COMBINING CONTINUOUS-TIME AND DISCRETE-TIME ELEMENTS

Thus far the z-transform has been used mainly to describe systems and elements which operate on and produce only discrete-time signals, and the Laplace transform has been used only for continuous-time systems and elements, with continuous-time input and output signals. However, many control systems include both types of elements. Some of the important relationships between the z-transform and the Laplace transform are developed here, to facilitate analysis and design of mixed (continuous/discrete) systems.

Discrete-time signals arise either from the sampling of continuous-time signals, or as the output of inherently discrete-time system components, such as digital computers. If a continuous-time signal $y(t)$ with Laplace transform $Y(s)$ is sampled uniformly, with period T, the resulting sequence of samples $y(kT)$, $k = 0, 1, 2, \ldots$, can be written as

$$y(kT) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} Y(s)e^{skT}\,ds \qquad k = 0, 1, 2, \ldots$$

where $c > \sigma_0$ (see Definition 4.3). The z-transform of this sequence is $Y^*(z) = \sum_{k=0}^{\infty} y(kT)z^{-k}$ (Definition 4.4) which, as shown in Problem 6.41, can be written as

$$Y^*(z) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} Y(s)\left(\frac{1}{1-e^{sT}z^{-1}}\right)ds \tag{6.15}$$

for the region of convergence $|z| > e^{cT}$. This relationship between the Laplace transform and the z-transform can be evaluated by application of Cauchy's integral law [1]. However, in practice, it is usually not necessary to use this complex analysis approach.

The continuous-time function $y(t) = \mathscr{L}^{-1}[Y(s)]$ can be determined from $Y(s)$ and a table of Laplace transforms, and the time variable t is then replaced by kT, providing the kth element of the desired sequence:

$$y(kT) = \mathscr{L}^{-1}[Y(s)]\big|_{t=kT}$$

Then the z-transform of the sequence $y(kT)$, $k = 0, 1, 2, \ldots$, is generated by referring to a table of z-transforms, which yields the desired result:

$$Y^*(z) = \mathscr{Z}\{y(kT)\} = \mathscr{Z}\left\{\mathscr{L}^{-1}[Y(s)]\big|_{t=kT}\right\} \tag{6.16}$$

Thus, in Equation (6.16), the symbolic operations $\mathscr{L}^{-1}$ and $\mathscr{Z}$ represent straightforward table lookups, and $|_{t=kT}$ generates the sequence to be z-transformed.

A common combination of discrete-time and continuous-time elements and signals is shown in Fig. 6-3.

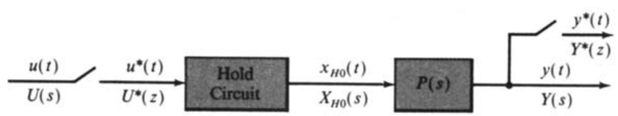


Fig. 6-3

If the hold circuit is a *zero-order hold*, then as shown in Problem 6.42, the discrete-time transfer function from $U^*(z)$ to $Y^*(z)$ is given by

$$\frac{Y^*(z)}{U^*(z)} = (1-z^{-1})\mathscr{Z}\left\{\mathscr{L}^{-1}\frac{P(s)}{s}\bigg|_{t=kT}\right\} \tag{6.17}$$

In practice, the sampler at the output, generating $y^*(t)$ in Fig. 6-3, may not exist. However, it is sometimes convenient to assume one exists at that point, for purposes of analysis (see, e.g., Problem 10.13). When this is done, the sampler is often called a **fictitious sampler**.

If both the input and output of a system like the one shown in Fig. 6-3 are continuous-time signals, and the input is subsequently sampled, then Equation (6.17) generates a discrete-time transfer function

which relates the input at the sampling times $T, 2T, \ldots$ to the output at the same sampling times. However, this discrete-time system transfer function does *not* relate input and output signals at times τ *between* sampling times, that is, for $kT < \tau < (k+1)T$, $k = 0, 1, 2, \ldots$.

EXAMPLE 6.14. In Fig. 6-3, if the hold circuit is a zero-order hold and $P(s) = 1/(s+1)$, then from Equation (6.17), the discrete-time transfer function of the mixed-element subsystem is

$$\begin{aligned}
\frac{Y^*(z)}{U^*(z)} &= (1-z^{-1})\mathscr{Z}\left\{\mathscr{L}^{-1}\left(\frac{1}{s(s+1)}\right)\bigg|_{t=kT}\right\}\\
&= (1-z^{-1})\mathscr{Z}\left\{\mathscr{L}^{-1}\left(\frac{1}{s}-\frac{1}{s+1}\right)\bigg|_{t=kT}\right\}\\
&= (1-z^{-1})\mathscr{Z}\left\{(\mathbf{1}(t)-e^{-t})\big|_{t=kT}\right\}\\
&= (1-z^{-1})\mathscr{Z}\left\{\mathbf{1}(kT)-e^{-kT}\right\}\\
&= (1-z^{-1})\left[\mathscr{Z}\{\mathbf{1}(kT)\}-\mathscr{Z}\{e^{-kT}\}\right]\\
&= (1-z^{-1})\left[\frac{1}{1-z^{-1}}-\frac{1}{1-e^{-T}z^{-1}}\right]\\
&= \left(\frac{z-1}{z}\right)\left(\frac{z}{z-1}\right)\left[\frac{1-e^{-T}}{z-e^{-T}}\right]\\
&= \frac{1-e^{-T}}{z-e^{-T}}
\end{aligned}$$

Solved Problems

TRANSFER FUNCTION DEFINITIONS

6.1. What is the transfer function of a system whose input and output are related by the following differential equation?

$$\frac{d^2y}{dt^2} + 3\frac{dy}{dt} + 2y = u + \frac{du}{dt}$$

Taking the Laplace transform of this equation, ignoring terms due to initial conditions, we obtain

$$s^2Y(s) + 3sY(s) + 2Y(s) = U(s) + sU(s)$$

This equation can be written as

$$Y(s) = \left[\frac{s+1}{s^2+3s+2}\right]U(s)$$

The transfer function of this system is therefore given by

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

6.2. A particular system containing a time delay has the differential equation $(d/dt)y(t) + y(t) = u(t-T)$. Find the transfer function of this system.

The Laplace transform of the differential equation, ignoring terms due to initial conditions, is $sY(s) + Y(s) = e^{-sT}U(s)$. $Y(s)$ and $U(s)$ are related by the following function of s, which is the system transfer function

$$P(s) = \frac{Y(s)}{U(s)} = \frac{e^{-sT}}{s+1}$$

6.3. The position y of a moving object of constant mass M is related to the total force f applied to the object by the differential equation $M(d^2y/dt^2) = f$. Determine the transfer function relating the position to the applied force.

Taking the Laplace transform of the differential equation, we obtain $Ms^2Y(s) = F(s)$. The transfer function relating $Y(s)$ to $F(s)$ is therefore $P(s) = Y(s)/F(s) = 1/Ms^2$.

6.4. A motor connected to a load with inertia J and viscous friction B produces a torque proportional to the input current i. If the differential equation for the motor and load is $J(d^2\theta/dt^2) + B(d\theta/dt) = Ki$, determine the transfer function between the input current i and the shaft position θ.

The Laplace transform version of the differential equation is $(Js^2 + Bs)\Theta(s) = KI(s)$, and the required transfer function is $P(s) = \Theta(s)/I(s) = K/s(Js + B)$.

PROPERTIES OF TRANSFER FUNCTIONS

6.5. An impulse is applied at the input of a continuous system and the output is observed to be the time function e^{-2t}. Find the transfer function of this system.

The transfer function is $P(s) = Y(s)/U(s)$ and $U(s) = 1$ for $u(t) = \delta(t)$. Therefore

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

6.6. The impulse response of a certain continuous system is the sinusoidal signal $\sin t$. Determine the system transfer function and differential equation.

The system transfer function is the Laplace transform of its impulse response, $P(s) = 1/(s^2+1)$. Then $P(D) = y/u = 1/(D^2+1)$, $D^2y + y = u$ or $d^2y/dt^2 + y = u$.

6.7. The step response of a given system is $y = 1 - \frac{7}{3}e^{-t} + \frac{3}{2}e^{-2t} - \frac{1}{6}e^{-4t}$. What is the transfer function of this system?

Since the derivative of a step is an impulse (see Definition 3.17), the impulse response for this system is $p(t) = dy/dt = \frac{7}{3}e^{-t} - 3e^{-2t} + \frac{2}{3}e^{-4t}$.

The Laplace transform of $p(t)$ is the desired transfer function. Thus

$$P(s) = \frac{\frac{7}{3}}{s+1} + \frac{-3}{s+2} + \frac{\frac{2}{3}}{s+4} = \frac{s+8}{(s+1)(s+2)(s+4)}$$

Note that an alternative solution would be to compute the Laplace transform of y and then multiply by s to determine $P(s)$, since a multiplication by s in the s-domain is equivalent to differentiation in the time domain.

6.8. Determine if the transfer function $P(s) = (2s+1)/(s^2+s+1)$ represents a stable or an unstable system.

The characteristic equation of the system is obtained by setting the denominator polynomial to zero, that is, $s^2 + s + 1 = 0$. The characteristic equation may then be tested using one of the stability criteria described in Chapter 5. The Routh table for this system is given by

$$\begin{array}{c|cc} s^2 & 1 & 1 \\ s^1 & 1 & \\ s^0 & 1 & \end{array}$$

Since there are no sign changes in the first column, the system is stable.

6.9. Does the transfer function $P(s) = (s+4)/(s+1)(s+2)(s-1)$ represent a stable or an unstable system?

The stability of the system is determined by the roots of the denominator polynomial, that is, the *poles* of the system. Here the denominator is in factored form and the poles are located at $s = -1, -2, +1$. Since there is one pole with a positive real part, the system is unstable.

6.10. What is the transfer function of a system with a gain factor of 2 and a pole-zero map in the s-plane as shown in Fig. 6-4?

The transfer function has a zero at -1 and poles at -2 and the origin. Hence the transfer function is $P(s) = 2(s+1)/s(s+2)$.

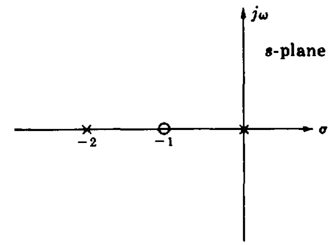


Fig. 6-4

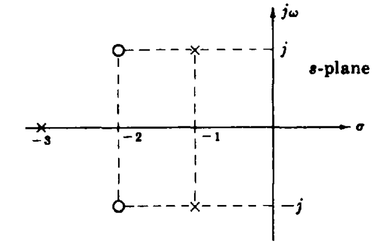


Fig. 6-5

6.11. Determine the transfer function of a system with a gain factor of 3 and the pole-zero map shown in Fig. 6-5

The transfer function has zeros at $-2 \pm j$ and poles at -3 and at $-1 \pm j$. The transfer function is therefore $P(s) = 3(s+2+j)(s+2-j)/(s+3)(s+1+j)(s+1-j)$.

TRANSFER FUNCTIONS OF CONTINUOUS CONTROL SYSTEM COMPONENTS

6.12. An R-C network mechanization of a lead compensator is shown in Fig. 6-6. Find its transfer function.

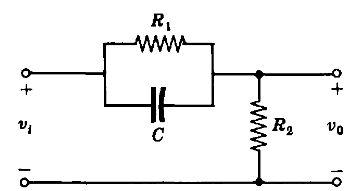


Fig. 6-6

Assuming the circuit is not loaded, that is, no current flows through the output terminals, Kirchhoff's current law for the output node yields

$$C\frac{d}{dt}(v_i - v_0) + \frac{1}{R_1}(v_i - v_0) = \frac{1}{R_2}v_0$$

The Laplace transform of this equation (with zero initial conditions) is

$$Cs[V_i(s) - V_0(s)] + \frac{1}{R_1}[V_i(s) - V_0(s)] = \frac{1}{R_2}V_0(s)$$

The transfer function is

$$P_{\text{Lead}} = \frac{V_0(s)}{V_i(s)} = \frac{Cs + 1/R_1}{Cs + 1/R_1 + 1/R_2} = \frac{s+a}{s+b}$$

where $a = 1/R_1C$ and $b = 1/R_1C + 1/R_2C$.

6.13. Determine the transfer function of the *R-C* network mechanization of the lag compensator shown in Fig. 6-7.

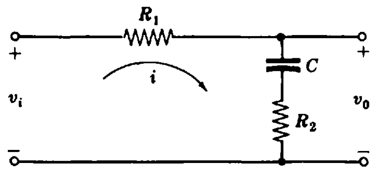


Fig. 6-7

Kirchhoff's voltage law for the loop yields the equation

$$iR_1 + \frac{1}{C}\int_0^t i\,dt + iR_2 = v_i$$

whose Laplace transform is

$$\left(R_1 + R_2 + \frac{1}{Cs}\right)I(s) = V_i(s)$$

The output voltage is given by

$$V_0(s) = \left(R_2 + \frac{1}{Cs}\right)I(s)$$

The transfer function of the lag network is therefore

$$P_{\text{Lag}} = \frac{V_0(s)}{V_i(s)} = \frac{R_2 + 1/Cs}{R_1 + R_2 + 1/Cs} = \frac{a(s+b)}{b(s+a)} \qquad \text{where } a = \frac{1}{(R_1+R_2)C} \qquad b = \frac{1}{R_2C}$$

6.14. Derive the transfer function of the *R-C* network mechanization of the lag-lead compensator shown in Fig. 6-8.

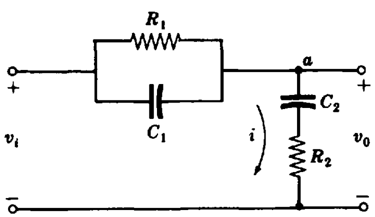


Fig. 6-8

Equating currents at the output node a yields

$$\frac{1}{R_1}(v_i - v_0) + C_1\frac{d}{dt}(v_i - v_0) = i$$

The voltage v_0 and the current i are related by

$$\frac{1}{C_2}\int_0^t i\,dt + iR_2 = v_0$$

Taking the Laplace transform of these two equations (with zero initial conditions) and eliminating $I(s)$ results in the equation

$$\left(\frac{1}{R_1} + C_1 s\right)[V_i(s) - V_0(s)] = \frac{V_0(s)}{1/sC_2 + R_2}$$

The transfer function of the network is therefore

$$P_{\text{LL}} = \frac{V_0(s)}{V_i(s)} = \frac{\left(s + \frac{1}{R_1C_1}\right)\left(s + \frac{1}{R_2C_2}\right)}{s^2 + \left(\frac{1}{R_2C_2} + \frac{1}{R_2C_1} + \frac{1}{R_1C_1}\right)s + \frac{1}{R_1C_1R_2C_2}} = \frac{(s+a_1)(s+b_2)}{(s+b_1)(s+a_2)}$$

where

$$a_1 = \frac{1}{R_1C_1} \qquad b_1a_2 = a_1b_2 \qquad b_1 + a_2 = a_1 + b_2 + \frac{1}{R_2C_1} \qquad b_2 = \frac{1}{R_2C_2}$$

6.15. Find the transfer function of the *simple* lag network shown in Fig. 6-9.

This network is a special case of the lag compensation network of Problem 6.13 with R_2 set equal to zero. Hence the transfer function is given by

$$P(s) = \frac{V_0(s)}{V_i(s)} = \frac{1/Cs}{R + 1/Cs} = \frac{1/RC}{s + 1/RC}$$

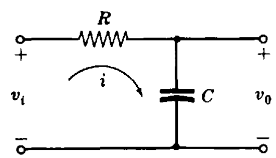


Fig. 6-9

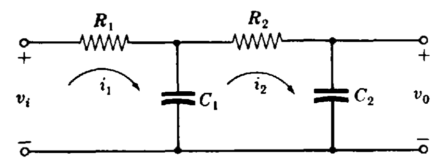


Fig. 6-10

6.16. Determine the transfer function of two simple lag networks connected in series as shown in Fig. 6-10.

The two loop equations are

$$R_1i_1 + \frac{1}{C_1}\int_0^t (i_1 - i_2)\,dt = v_i$$

$$R_2i_2 + \frac{1}{C_2}\int_0^t i_2\,dt + \frac{1}{C_1}\int_0^t (i_2 - i_1)\,dt = 0$$

Using the Laplace transformation and solving the two loop equations for $I_2(s)$, we obtain

$$I_2(s) = \frac{C_2 s V_i(s)}{R_1R_2C_1C_2s^2 + (R_1C_1 + R_1C_2 + R_2C_2)s + 1}$$

The output voltage is given by $v_0 = (1/C_2)\int_0^t i_2\,dt$. Thus

$$\frac{V_0(s)}{V_i(s)} = \frac{1}{R_1R_2C_1C_2s^2 + (R_1C_1 + R_1C_2 + R_2C_2)s + 1}$$

CONTINUOUS SYSTEM TIME RESPONSE

6.17. What is the unit step response of a continuous system whose transfer function has a zero at -1, a pole at -2, and a gain factor of 2?

The Laplace transform of the output is given by $Y(s) = P(s)U(s)$. Here

$$P(s) = \frac{2(s+1)}{s+2} \qquad U(s) = \frac{1}{s} \qquad Y(s) = \frac{2(s+1)}{s(s+2)} = \frac{1}{s} + \frac{1}{s+2}$$

Evaluating the inverse transform of the partial fraction expansion of $Y(s)$ gives $y(t) = 1 + e^{-2t}$.

6.18. Graphically evaluate the unit step response of a continuous system whose transfer function is given by

$$P(s) = \frac{(s+2)}{(s+0.5)(s+4)}$$

The pole-zero map of the output is obtained by adding the poles and zeros of the input to the pole-zero map of the transfer function. The output pole-zero map therefore has poles at 0, -0.5, and -4 and a zero at -2 as shown in Fig. 6-11.

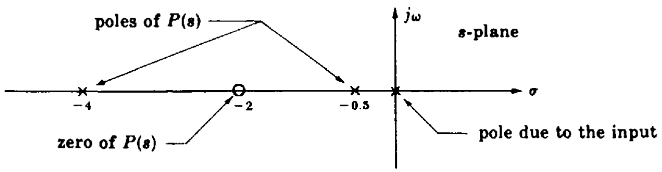


Fig. 6-11

The residue for the pole at the origin is

$$|R_1| = \frac{2}{0.5(4)} = 1 \qquad \arg R_1 = 0°$$

For the pole at -0.5,

$$|R_2| = \frac{1.5}{0.5(3.5)} = 0.857 \qquad \arg R_2 = -180°$$

For the pole at -4,

$$|R_3| = \frac{2}{4(3.5)} = 0.143 \qquad \arg R_3 = -180°$$

The time response is therefore $y(t) = R_1 + R_2 e^{-0.5t} + R_3 e^{-4t} = 1 - 0.857e^{-0.5t} - 0.143e^{-4t}$.

6.19. Evaluate the unit step response of the system of Problem 6.11.

The Laplace transform of the system output is

$$Y(s) = P(s)U(s) = \frac{3(s+2+j)(s+2-j)}{s(s+3)(s+1+j)(s+1-j)}$$

Expanding $Y(s)$ into partial fractions yields

$$Y(s) = \frac{R_1}{s} + \frac{R_2}{s+3} + \frac{R_3}{s+1+j} + \frac{R_4}{s+1-j}$$

where

$$R_1 = \frac{3(2+j)(2-j)}{3(1+j)(1-j)} = \frac{5}{2} \qquad R_3 = \frac{3(1)(1-2j)}{(-1-j)(2-j)(-2j)} = \frac{-3}{20}(7+j)$$

$$R_2 = \frac{3(-1+j)(-1-j)}{-3(-2+j)(-2-j)} = \frac{-2}{5} \qquad R_4 = \frac{3(1+2j)(1)}{(2+j)(-1+j)(2j)} = \frac{-3}{20}(7-j)$$

Evaluating the inverse Laplace transform,

$$y = \frac{5}{2} - \frac{2}{5}e^{-3t} - \frac{3\sqrt{2}}{4}e^{-t}\left[e^{-j(t+\theta)} + e^{j(t+\theta)}\right] = \frac{5}{2} - \frac{2}{5}e^{-3t} - \frac{3\sqrt{2}}{2}e^{-t}\cos(t+\theta)$$

where $\theta = -\tan^{-1}[\frac{1}{7}] = -8.13°$.

CONTINUOUS SYSTEM FREQUENCY RESPONSE

6.20. Prove that the steady state output of a stable system with transfer function $P(s)$ and input $u = A\sin\omega t$ is given by

$$y_{ss} = A|P(j\omega)|\sin(\omega t + \phi) \qquad \text{where} \quad \phi = \arg P(j\omega)$$

The Laplace transform of the output is $Y(s) = P(s)U(s) = P(s)[A\omega/(s^2+\omega^2)]$.

When this transform is expanded into partial fractions, there will be terms due to the poles of $P(s)$ and two terms due to the poles of the input ($s = \pm j\omega$). Since the system is stable, all time functions resulting from the poles of $P(s)$ decay to zero as time approaches infinity. Thus the steady state output contains only the time functions resulting from the terms in the partial fraction expansion due to the poles of the input. The Laplace transform of the steady state output is therefore

$$Y_{ss}(s) = \frac{AP(j\omega)}{2j(s-j\omega)} + \frac{AP(-j\omega)}{-2j(s+j\omega)}$$

The inverse transform of this equation is

$$y_{ss} = A|P(j\omega)|\left[\frac{e^{j\phi}e^{j\omega t} - e^{-j\phi}e^{-j\omega t}}{2j}\right] = A|P(j\omega)|\sin(\omega t + \phi) \qquad \text{where} \quad \phi = \arg P(j\omega)$$

6.21. Find the *d.c. gain* of each of the systems represented by the following transfer functions:

$$(a)\quad P(s) = \frac{1}{s+1} \qquad (b)\quad P(s) = \frac{10}{(s+1)(s+2)} \qquad (c)\quad P(s) = \frac{(s+8)}{(s+2)(s+4)}$$

The d.c. gain is given by $P(0)$. Then (a) $P(0) = 1$, (b) $P(0) = 5$, (c) $P(0) = 1$.

6.22. Evaluate the gain and phase shift of $P(s) = 2/(s+2)$ for $\omega = 1$, 2, and 10.

The gain of $P(s)$ is given by $|P(j\omega)| = 2/\sqrt{\omega^2+4}$. For $\omega = 1$, $|P(j1)| = 2/\sqrt{5} = 0.894$; for $\omega = 2$, $|P(j2)| = 2/\sqrt{8} = 0.707$; for $\omega = 10$, $|P(j10)| = 2/\sqrt{104} = 0.196$.

The phase shift of the transfer function is the phase angle of $P(j\omega)$, $\arg P(j\omega) = -\tan^{-1}\omega/2$. For $\omega = 1$, $\arg P(j1) = -\tan^{-1}\frac{1}{2} = -26.6°$; for $\omega = 2$, $\arg P(j2) = -\tan^{-1}1 = -45°$; for $\omega = 10$, $\arg P(j10) = -\tan^{-1}5 = -78.7°$.

6.23. Sketch the graphs of $|P(j\omega)|$ and $\arg P(j\omega)$ as a function of frequency for the transfer function of Problem 6.22.

In addition to the values calculated in Problem 6.22 for $|P(j\omega)|$ and $\arg P(j\omega)$, the values for $\omega = 0$ will also be useful: $|P(j0)| = 2/2 = 1$, $\arg P(j0) = -\tan^{-1} 0 = 0$.

As ω becomes large, $|P(j\omega)|$ asymptotically approaches zero while $\arg P(j\omega)$ asymptotically approaches $-90°$. The graphs representing the frequency response of $P(s)$ are shown in Fig. 6-12.

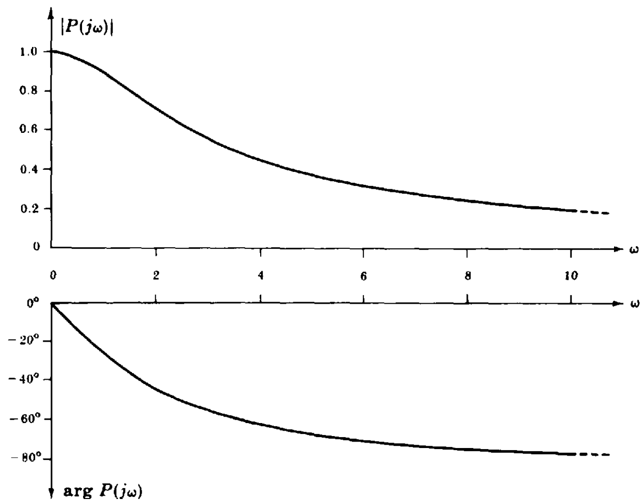


Fig. 6-12

DISCRETE-TIME SYSTEM TRANSFER FUNCTIONS AND TIME RESPONSES

6.24. The Kronecker delta response of a discrete-time system is given by $y_\delta(k) = 1$ for all $k \geq 0$. What is its transfer function?

The transfer function is the z-transform of the Kronecker delta response, as given in Example 4.26:

$$P(z) = 1 + z^{-1} + z^{-2} + z^{-3} + \cdots$$

To determine a pole-zero representation of $P(z)$, note that

$$zP(z) - z = P(z)$$

or

$$(z-1)P(z) = z$$

so that

$$P(z) = \frac{z}{z-1}$$

Alternatively, note that the Kronecker delta response is the unit step sequence, which has the z-transform

$$P(z) = \frac{z}{z-1}$$

(see Table 4.2).

6.25. The Kronecker delta response of a particular discrete system is given by $y_\delta(k) = (0.5)^k$ for $k \geq 0$. What is its transfer function?

The form of the Kronecker delta response indicates the presence of a single pole at 0.5. The Kronecker delta response of a system with a single pole and no zero has no output at $k = 0$. That is,

$$\frac{1}{z-0.5} = z^{-1} + 0.5z^{-2} + 0.25z^{-3} + \cdots + (0.5)^{n-1}z^{-n} + \cdots$$

Consequently, the transfer function must have a zero in the numerator to advance the output sequence one sample interval. That is,

$$P(z) = \frac{z}{z-0.5}$$

6.26. What is the difference equation for a system whose transfer function is

$$P(z) = \frac{z-0.1}{z^2+0.3z+0.2}$$

Replacing z^n with Z^n, we get

$$P(Z) = \frac{Z-0.1}{Z^2+0.3Z+0.2}$$

Then

$$y(k) = P(Z)u(k) = \frac{(Z-0.1)u(k)}{Z^2+0.3Z+0.2} = \frac{u(k+1)-0.1u(k)}{Z^2+0.3Z+0.2}$$

and, by cross multiplying,

$$y(k+2) + 0.3y(k+1) + 0.2y(k) = u(k+1) - 0.1u(k)$$

6.27. What is the transfer function of a discrete system with a gain factor of 2, zeros at 0.2 and −0.5, and poles at 0.5, 0.6, and −0.4? Is it stable?

Mathcad

The transfer function is

$$P(z) = \frac{2(z-0.2)(z+0.5)}{(z-0.5)(z-0.6)(z+0.4)}$$

Since all the system poles are inside the unit circle, the system is stable.

MISCELLANEOUS PROBLEMS

6.28. A *d.c. (direct current) motor* is shown schematically in Fig. 6-13. L and R represent the inductance and resistance of the motor armature circuit, and the voltage v_b represents the generated back e.m.f. (electromotive force) which is proportional to the shaft velocity $d\theta/dt$. The torque T generated by the motor is proportional to the armature current i. The inertia J represents the combined inertia of the motor armature and the load, and B is the total viscous friction acting on the output shaft. Determine the transfer function between the input voltage V and the angular position Θ of the output shaft.

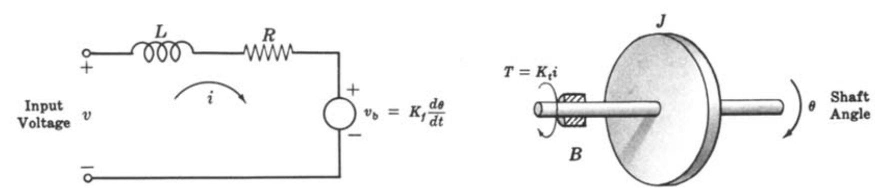


Fig. 6-13

The differential equations of the motor armature circuit and the inertial load are

$$Ri + L\frac{di}{dt} = v - K_f\frac{d\theta}{dt} \quad \text{and} \quad K_t i = J\frac{d^2\theta}{dt^2} + B\frac{d\theta}{dt}$$

Taking the Laplace transform of each equation, ignoring initial conditions,

$$(R+sL)I = V - K_f s\Theta \quad \text{and} \quad K_t I = (Js^2+Bs)\Theta$$

Solving these equations simultaneously for the transfer function between V and Θ, we have

$$\frac{\Theta}{V} = \frac{K_t}{(Js^2+Bs)(Ls+R)+K_tK_fs} = \frac{K_t/JL}{s\left[s^2+(B/J+R/L)s+BR/JL+K_tK_f/JL\right]}$$

6.29. The back e.m.f generated by the armature circuit of a d.c. machine is proportional to the angular velocity of its shaft, as noted in the problem above. This principle is utilized in the *d.c. tachometer* shown schematically in Fig. 6-14, where v_b is the voltage generated by the armature, L is the armature inductance, R_a is the armature resistance, and v_0 is the output voltage. If K_f is the proportionality constant between v_b and shaft velocity $d\theta/dt$, that is, $v_b = K_f(d\theta/dt)$, determine the transfer function between the shaft position Θ and the output voltage V_0. The output load is represented by a resistance R_L and $R_L + R_a \equiv R$.

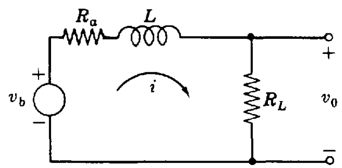


Fig. 6-14

The Laplace transformed equation representing the tachometer is $I(R+sL) = K_f s\Theta$. The output voltage is given by

$$V_0 = IR_L = \frac{R_L K_f s\Theta}{R+sL}$$

The transfer function of the d.c. tachometer is then

$$\frac{V_0}{\Theta} = \frac{R_L K_f}{L}\left(\frac{s}{s+R/L}\right)$$

6.30. A simple mechanical *accelerometer* is shown in Fig. 6-15. The position y of the mass M with respect to the accelerometer case is proportional to the acceleration of the case. What is the transfer function between the input acceleration A ($a = d^2x/dt^2$) and the output Y?

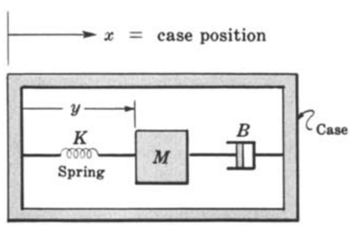


Fig. 6-15

Equating the sum of the forces acting on the mass M to its inertial acceleration, we obtain

$$-B\frac{dy}{dt} - Ky = M\frac{d^2}{dt^2}(y-x)$$

or

$$M\frac{d^2y}{dt^2} + B\frac{dy}{dt} + Ky = M\frac{d^2x}{dt^2} = Ma$$

where a is the input acceleration. The zero initial condition transformed equation is

$$(Ms^2 + Bs + K)Y = MA$$

The transfer function of the accelerometer is therefore

$$\frac{Y}{A} = \frac{1}{s^2 + (B/M)s + K/M}$$

6.31. A differential equation describing the dynamic operation of the *one-degree-of-freedom gyroscope* shown in Fig. 6-16 is

$$J\frac{d^2\theta}{dt^2} + B\frac{d\theta}{dt} + K\theta = H\omega$$

where ω is the angular velocity of the gyroscope about the input axis, θ is the angular position of the spin axis—the measured output of the gyroscope, H is angular momentum stored in the spinning wheel, J is the inertia of the wheel about the output axis, B is the viscous friction coefficient about the output axis, and K is the spring constant of the restraining spring attached to the spin axis.

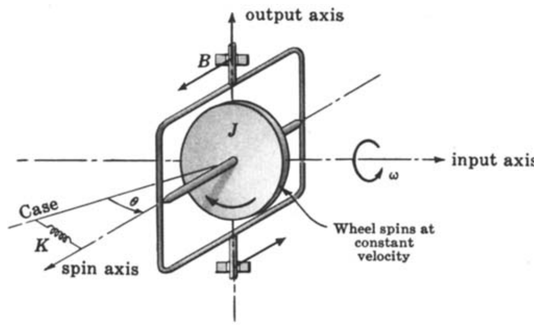


Fig. 6-16

(*a*) Determine the transfer function relating the Laplace transforms of ω and θ, and show that the steady state output is proportional to the magnitude of a constant rate input. This type of gyroscope is called a *rate gyro*.

(*b*) Determine the transfer function between ω and θ with the restraining spring removed ($K = 0$). Since here the output is proportional to the integral of the input rate, this type of gyroscope is called an *integrating gyro*.

(*a*) The zero initial condition transform of the gyroscope differential equation is

$$(Js^2 + Bs + K)\Theta = H\Omega$$

where Θ and Ω are the Laplace transforms of θ and ω, respectively. The transfer function relating Θ and Ω is therefore

$$\frac{\Theta}{\Omega} = \frac{H}{(Js^2 + Bs + K)}$$

For a constant or d.c. rate input ω_K, the magnitude of the steady state output θ_{ss} can be obtained by multiplying the input by the d.c. gain of the transfer function, which in this case is H/K. Thus the steady state output is proportional to the magnitude of the rate input, that is, $\theta_{ss} = (H/K)\omega_K$.

(*b*) Setting K equal to zero in the transfer function of (*a*) yields $\Theta/\Omega = H/s(Js+B)$. This transfer function now has a pole at the origin, so that an integration is obtained between the input Ω and the output Θ. The output is thus proportional to the integral of the input rate or, equivalently, the input angle.

6.32. A differential equation approximating the rotational dynamics of a rigid vehicle moving in the atmosphere is

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

where θ is the vehicle attitude angle, J is its inertia, N is the normal-force coefficient, L is the distance from the center of gravity to the center of pressure, and T is any applied torque (see Fig. 6-17). Determine the transfer function between an applied torque and the vehicle attitude angle.

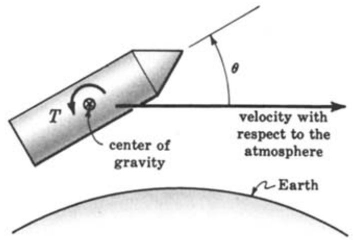


Fig. 6-17

The zero initial condition, transformed system differential equation is

$$(Js^2 - NL)\Theta = T$$

The desired transfer function is

$$\frac{\Theta}{T} = \frac{1}{Js^2 - NL} = \frac{1/J}{s^2 - NL/J}$$

Note that if NL is positive (center of pressure forward of the vehicle center of gravity), the system is *unstable* because there is a pole in the right half-plane at $s = \sqrt{NL/J}$. If NL is negative, the poles are imaginary and the system is *oscillatory* (marginally stable). However, aerodynamic damping terms not included in the differential equation are actually present and perform the function of damping out any oscillations.

6.33. Pressure receptors called *baroreceptors* measure changes in arterial blood pressure, as outlined in Problem 2.14. They are shown as a block in the feedback path of the block diagram determined in the solution of that problem. The frequency $b(t)$ at which signals (action potentials) move along the vagus and glossopharyngeal nerves from the baroreceptors to the vasomotor center (VMC) in the brain is proportional to arterial blood pressure p plus the time rate of change of blood pressure. Determine the form of the transfer function for the baroreceptors.

From the description given above, the equation for b is

$$b = k_1 p + k_2 \frac{dp}{dt}$$

where k_1 and k_2 are constants, and p is blood pressure. [p should not be confused here with the notation $p(t)$, the inverse Laplace transform of $P(s)$ introduced in this chapter as a general representation for a transfer function.] The Laplace transform of the above equation, with zero initial conditions, is

$$B = k_1 P + k_2 sP = P(k_1 + k_2 s)$$

The transfer function of the baroreceptors is therefore $B/P = k_1 + k_2 s$. We again remind the reader that P represents the transform of arterial blood pressure in this problem.

6.34. Consider the transfer function C_k/R_k for the biological system described in Problem 3.4(*a*) by the equations

$$c_k(t) = r_k(t) - \sum_{i=1}^{n} a_{k-i} c_i(t - \Delta t)$$

for $k = 1, 2, \ldots, n$. Explain how C_k/R_k may be computed.

Taking the Laplace transform of the above equations, ignoring initial conditions, yields the following set of equations:

$$C_k = R_k - \sum_{i=1}^{n} a_{k-i} C_i e^{-s\Delta t}$$

for $k = 1, 2, \ldots, n$. If all n equations were written down, we would have n equations in n unknowns (C_k for $k = 1, 2, \ldots, n$). The general solution for any C_k in terms of the inputs R_k can then be determined using the standard techniques for solving simultaneous equations. Let D represent the determinant of the coefficient matrix:

$$D \equiv \begin{vmatrix} 1 + a_0 e^{-s\Delta t} & a_{-1} e^{-s\Delta t} & \cdots & a_{1-n} e^{-s\Delta t} \\ a_1 e^{-s\Delta t} & 1 + a_0 e^{-s\Delta t} & \cdots & a_{2-n} e^{-s\Delta t} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n-1} e^{-s\Delta t} & \cdots & a_1 e^{-s\Delta t} & 1 + a_0 e^{-s\Delta t} \end{vmatrix}$$

Then in general,

$$C_k = \frac{D_k}{D}$$

where D_k is the determinant of the coefficient matrix with the kth column replaced by

$$\begin{matrix} R_1 \\ R_2 \\ \vdots \\ R_n \end{matrix}$$

The transfer function C_k/R_k is then determined by setting all the inputs except R_k equal to zero, computing C_k from the formula above, and dividing C_k by R_k.

6.35. Can you determine the s-domain transfer function of the ideal sampler described in Problems 3.5 and 4.39? Why?

No. From the results of Problem 4.39, the output transform $U(s)$ of the ideal sampler is

$$U^*(s) = \sum_{k=0}^{\infty} e^{-skT} u(kT)$$

It is not possible to factor out the transform $U(s)$ of the input signal $u(t)$ applied to the sampler, because the sampler is not a time-invariant system element. Therefore it cannot be described by an ordinary transfer function.

6.36. Based on the developments of the sampler and zero-order hold function given in Problems 3.5, 3.6, 3.7, and 4.39, design an idealization of the zero-order hold transfer function.

In Problem 3.7, impulses in $m_{1T}(t)$ replaced the current pulses modulated by $m_s(t)$ in Problem 3.6. Then, by the screening property of the unit impulse, Equation (*3.20*), the integral of each impulse is the value of $u(t)$ at the sampling instant kT, $k = 0, 1, \ldots$, etc. Therefore it is logical to replace the capacitor (and resistor) in the approximate hold circuit of Problem 3.6 by an integrator, which has the Laplace transform $1/s$. To complete the design, the output of the hold must be equal to u at each sampling time, not $u - y_{H0}$; therefore we need a function that automatically resets the integrator to zero after each sampling period. The transfer function of such a device is given by the "pulse" transfer function:

$$P_{H0}(s) = \frac{1}{s}(1 - e^{-sT})$$

Then we can write the transform of the output of the ideal hold device as

$$Y_{H0}(s) = P_{H0}(s) U^*(s) = \frac{1}{s}(1 - e^{-sT}) \sum_{k=0}^{\infty} e^{-sT} u(kT)$$

6.37. Can you determine the s-domain transfer function of the ideal sampler and ideal zero-order hold combination of the previous problem? Why?

No. It is not possible to factor out the transform $U(s)$ of $u(t)$ applied to the sampler. Again, the sampler is not a time-invariant device.

6.38. The simple lag circuit of Fig. 6-3, with a switch S in the input line, was described in Problem 3.6 as an approximate sample and zero-order hold device, and idealized in Problem 6.36. Why is this the case, and under what circumstances?

The transfer function of the simple lag was shown in Problem 6-15 to be

$$P(s) = \frac{1/RC}{s + 1/RC}$$

If $RC \ll 1$, $P(s)$ can be approximated as $P(s) \approx 1$, and the capacitor ideally holds the output constant until the next sample time.

6.39. Show that for a rational function $P(z)$ to be the transfer function of a *causal* discrete-time system, the order of its denominator polynomial must be equal to or greater than the order of its numerator polynomial (Property 6, Section 6.6).

In Section 3.16 we saw that a discrete-time system is causal if its weighting sequence $w(k) = 0$ for $k < 0$. Let $P(z)$, the system transfer function, have the form:

$$P(z) = \frac{b_m z^m + b_{m-1} z^{m-1} + \cdots + b_1 z + b_0}{a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0}$$

where $a_n \neq 0$ and $b_m \neq 0$. The weighting sequence $w(k)$ can be generated by inverting $P(z)$, using the long division technique of Section 4.9.

We first divide the numerator and denominator of $P(z)$ by z^m, thus forming:

$$P(z) = \frac{b_m + b_{m-1} z^{-1} + \cdots + b_0 z^{-m}}{a_n z^{n-m} + a_{n-1} z^{n-m-1} + \cdots + a_0 z^{-m}}$$

Dividing the denominator of $P(z)$ into its numerator then gives

$$P(z) = \left(\frac{b_m}{a_n}\right) z^{m-n} + \left(b_{m-1} - \frac{b_m a_{n-1}}{a_n}\right) z^{m-n-1} + \cdots$$

The coefficient of z^{-k} in this expansion of $P(z)$ is $w(k)$, and we see that $w(k) = 0$ for $k < n - m$ and

$$w(n-m) = \frac{b_m}{a_n} \neq 0$$

For causality, $w(k) = 0$ for $k < 0$, therefore $n - m \geq 0$ and $n \geq m$.

6.40. Show that the steady state response of a stable discrete-time system to an input sequence $u(k) = A \sin \omega kT$, $k = 0, 1, 2, \ldots$, is given by

$$y_{ss} = A\left|P(e^{j\omega T})\right| \sin(\omega kT + \phi) \qquad k = 0, 1, 2, \ldots \tag{6.14}$$

where $P(z)$ is the system transfer function.

Since the system is linear, if this result is true for $A = 1$, then it is true for arbitrary values of A. To simplify the arguments, an input $u'(k) = e^{j\omega kT}$, $k = 0, 1, 2, \ldots$, is used. By noting that

$$u'(k) = e^{j\omega kT} = \cos \omega kT + j \sin \omega kT$$

the response of the system to $\{u'(k)\}$ is a complex combination of the responses to $\{\cos \omega kT\}$ and $\{\sin \omega kT\}$, where the imaginary part is the response to $\{\sin \omega kT\}$. From Table 4.2 the z-transform of $\{e^{j\omega kT}\}$ is

$$\frac{z}{z - e^{j\omega T}}$$

Thus the z-transform of the system output $Y'(z)$ is

$$Y'(z) = P(z)\frac{z}{z - e^{j\omega T}}$$

To invert $Y'(z)$, we form the partial fraction expansion of

$$\frac{Y'(z)}{z} = P(z)\frac{1}{z - e^{j\omega T}}$$

This expansion consists of terms due to the poles of $P(z)$ and a term due to the pole at $z = e^{j\omega T}$. Therefore

$$Y'(z) = z\left[\sum \text{terms due to poles of } P(z) + \frac{P(e^{j\omega T})}{z - e^{j\omega T}}\right]$$

and

$$\{y'(k)\} = \mathcal{Z}^{-1}\left[z\sum \text{terms due to poles of } P(z)\right] + \left\{P(e^{j\omega T}) e^{j\omega kT}\right\}$$

Since the system is stable, the first term vanishes as k becomes large and

$$\begin{aligned} y_{ss} &= P(e^{j\omega T}) e^{j\omega kT} = \left|P(e^{j\omega T})\right| e^{j(\omega kT + \phi)} \\ &= \left|P(e^{j\omega T})\right| [\cos(\omega kT + \phi) + j\sin(\omega kT + \phi)] \qquad k = 0, 1, 2, \ldots \end{aligned}$$

where $\phi = \arg P(e^{j\omega T})$. The steady state response to the input $\sin \omega kT$ is the imaginary part of y_{ss}, or

$$y_{ss} = \left|P(e^{j\omega T})\right| \sin(\omega kT + \phi) \qquad k = 0, 1, 2, \ldots$$

6.41. Show that, if a continuous-time function $y(t)$ with Laplace transform $Y(s)$ is sampled uniformly with period T, the z-transform of the resulting sequence of samples $Y^*(z)$ is related to $Y(s)$ by Equation (*6.15*).

From Definition 4.3:

$$y(t) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} Y(s) e^{st}\, ds$$

where $c > \sigma_0$. Uniformly sampling $y(t)$ generates the samples $y(kT)$, $k = 0, 1, 2, \ldots$. Therefore

$$y(kT) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} Y(s) e^{skT}\, ds \qquad k = 0, 1, 2, \ldots$$

The z-transform of this sequence is

$$Y^*(z) = \sum_{k=0}^{\infty} y(kT) z^{-k} = \sum_{k=0}^{\infty} \frac{z^{-k}}{2\pi j}\int_{c-j\infty}^{c+j\infty} Y(s) e^{skT}\, ds$$

and after interchanging summation and integration,

$$Y^*(z) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} Y(s) \sum_{k=0}^{\infty} e^{skT} z^{-k}\, ds$$

Now

$$\sum_{k=0}^{\infty} e^{skT} z^{-k} = \sum_{k=0}^{\infty} \left(e^{sT} z^{-1}\right)^k$$

is a geometric series, which converges if $|e^{sT} z^{-1}| < 1$. In this case,

$$\sum_{k=0}^{\infty} \left(e^{sT} z^{-1}\right)^k = \frac{1}{1 - e^{sT} z^{-1}}$$

The inequality $|e^{sT} z^{-1}| < 1$ implies that $|z| > |e^{sT}|$. On the integration contour, $|e^{sT}| = |e^{(c+j\omega)T}| = e^{cT}$

Thus the series converges for $|z| > e^{cT}$. Therefore

$$Y^*(z) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} Y(s)\frac{1}{1-e^{sT}z^{-1}}\,ds$$

for $|z| > e^{cT}$, which is Equation (*6.15*).

6.42. Show that if the hold circuit in Fig. 6-3 is a zero-order hold, the discrete-time transfer function is given by Equation (*6.17*).

Let $p(t) = \mathscr{L}^{-1}[P(s)]$. Then, using the convolution integral (Definition 3.23), the output of $P(s)$ can be written as

$$y(t) = \int_0^t p(t-\tau)x_{H0}(\tau)\,d\tau$$

Since $x_{H0}(t)$ is the output of a zero-order hold, it is constant over each sampling interval. Thus $y(t)$ can be written as

$$y(t) = \int_0^T p(t-\tau)x(0)\,d\tau + \int_T^{2T} p(t-\tau)x(1)\,d\tau + \cdots$$

$$+\int_{(j-2)T}^{(j-1)T} p(t-\tau)x[(j-2)T]\,d\tau + \int_{(j-1)T}^{t} p(t-\tau)x[(j-1)T]\,d\tau$$

where $(j-1)T \le t \le jT$. Now

$$y(jT) = \sum_{i=0}^{j-1}\left(\int_{iT}^{(i+1)T} p(jT-\tau)\,d\tau\right)x(iT)$$

By letting $\theta \equiv jT-\tau$, the integral can be rewritten as

$$\int_{iT}^{(i+1)T} p(jT-\tau)\,d\tau = \int_{(j-i-1)T}^{(j-i)T} p(\theta)\,d\theta$$

where $i = 0,1,2,3,\ldots, j-1$. Now, defining $h(t) \equiv \int_0^t p(\theta)\,d\theta$ and $k = j-1$ or $j = k+1$ yields

$$\int_{(j-i-1)T}^{(j-i)T} p(\theta)\,d\theta = \int_0^{(j-i)T} p(\theta)\,d\theta - \int_0^{(j-i-1)T} p(\theta)\,d\theta = \int_0^{(k-i+1)T} p(\theta)\,d\theta - \int_0^{(k-i)T} p(\theta)\,d\theta$$

$$= h[(k-i+1)T] - h[(k-i)T]$$

Therefore we can write

$$y[(k+1)T] = \sum_{i=0}^{k} h[(k-i+1)T]x(iT) - \sum_{i=0}^{k} h[(k-i)T]x(iT)$$

Using the relationship between the convolution sum and the product of z-transforms in Section 4.9, the Shift Theorem (Property 6, Section 4.9), and the definition of the z-transform, the z-transform of the last equation is

$$zY^*(z) = zH^*(z)X^*(z) - H^*(z)X^*(z)$$

where $Y^*(z)$ is the z-transform of the sequence $y(kT)$, $k = 0,1,2,\ldots$, $H^*(z)$ is the z-transform of $\int_0^{kT} p(\theta)\,d\theta$, $k = 0,1,2,\ldots$, and $X^*(z)$ is the z-transform of $x(kT)$, $k = 0,1,2,\ldots$. Rearranging terms yields

$$\frac{Y^*(z)}{X^*(z)} = (1-z^{-1})H^*(z)$$

Then, since $h(t) = \int_0^t p(\theta)\,d\theta$, $\mathscr{L}[h(t)] = P(s)/s$ and

$$\frac{Y^*(z)}{X^*(z)} = (1-z^{-1})\mathcal{Z}\left\{\mathscr{L}^{-1}\left(\frac{P(s)}{s}\right)\Bigg|_{t=kT}\right\}$$

6.43. Compare the solution in Problem 6.42 with that in Problem 6.37. What is fundamentally different about Problem 6.42, thereby permitting the use of linear frequency domain methods on this problem?

The presence of a sampler at the output of $P(s)$ permits the use of z-domain transfer functions for the combination of the sampler, zero-order hold, and $P(s)$.

Supplementary Problems

6.44. Determine the transfer function of the *R-C* network shown in Fig. 6-18

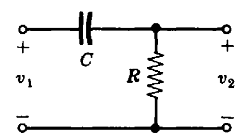


Fig. 6-18

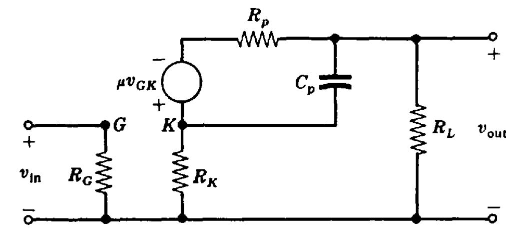


Fig. 6-19

6.45. An equivalent circuit of an electronic amplifier is shown in Fig. 6-19. What is its transfer function?

6.46. Find the transfer function of a system having the impulse response $p(t) = e^{-t}(1 - \sin t)$.

6.47. A sinusoidal input $x = 2\sin 2t$ is applied to a system with the transfer function $P(s) = 2/s(s+2)$. Determine the steady state output y_{ss}.

6.48. Find the step response of a system having the transfer function $P(s) = 4/(s^2-1)(s^2+1)$.

6.49. Determine which of the following transfer functions represent stable systems and which represent unstable systems:

(*a*) $P(s) = \dfrac{(s-1)}{(s+2)(s^2+4)}$ (*c*) $P(s) = \dfrac{(s+2)(s-2)}{(s+1)(s-1)(s+4)}$

(*b*) $P(s) = \dfrac{(s-1)}{(s+2)(s+4)}$ (*d*) $P(s) = \dfrac{6}{(s^2+s+1)(s+1)^2}$

(*e*) $P(s) = \dfrac{5(s+10)}{(s+5)(s^2-s+10)}$

6.50. Use the Final Value Theorem (Chapter 4) to show that the steady state value of the output of a stable system in response to a unit step input is equal to the d.c. gain of the system.

6.51. Determine the transfer function of two of the networks shown in Problem 6.44 connected in cascade (series).

6.52. Examine the literature for the transfer functions of two- and three-degree-of-freedom gyros and compare them with the one-degree-of-freedom gyro of Problem 6.31.

6.53. Determine the ramp response of a system having the transfer function $P(s) = (s+1)/(s+2)$.

6.54. Show that if a system described by

$$\sum_{i=0}^{n} a_i \frac{d^i y}{dt^i} = \sum_{i=0}^{m} b_i \frac{d^i u}{dt^i}$$

for $m \le n$ is at rest prior to application of the input, that is, $d^k y/dt^k = 0$, $k = 0, 1, \ldots, n-1$, for $t < 0$, then (terms due to *all* initial values u_0^k, y_0^k) $= 0$.
(*Hint:* Integrate the differential equation n times from $0^- \equiv \lim_{\epsilon \to 0,\, \epsilon < 0} \epsilon$ to t, and then let $t \to 0^+$.)

6.55. Determine the frequency response of the ideal zero-order hold (ZOH) device, with transfer function given in Problem 6.36, and sketch the gain and phase characteristics.

6.56. A zero-order hold was defined in Definition 2.13 and Example 2.9. A **first-order hold** maintains the **slope** of the function defined by the last two values of the sampler output, until the next sample time. Determine the discrete-time transfer function from $U^*(z)$ to $Y^*(z)$ for the subsystem in Fig. 6-3, with a first-order hold element.

Answers to Supplementary Problems

6.44. $\dfrac{V_2}{V_1} = \dfrac{s}{s + 1/RC}$

6.45. $\dfrac{V_{out}}{V_{in}} = \dfrac{-\mu R_L}{(R_k + R_L) R_p C_p s + (\mu + 1) R_k + R_p + R_L}$

6.46. $P(s) = \dfrac{s^2 + s + 1}{(s+1)(s^2 + 2s + 2)}$

6.47. $y_{ss} = 0.707 \sin(2t - 135°)$

6.48. $y = -4 + e^{-t} + e^{t} + 2\cos t$

6.49. (b) and (d) represent stable systems; (a), (c), and (e) represent unstable systems.

6.51. $\dfrac{V_2}{V_1} = \dfrac{s^2}{s^2 + (3/RC)s + 1/R^2C^2}$

6.53. $y = \frac{1}{4} - \frac{1}{4}e^{-2t} + \frac{1}{2}t$

6.55. $P(j\omega) = \left[\dfrac{T \sin(\omega T/2)}{\omega T/2}\right] e^{-j\omega T/2}$

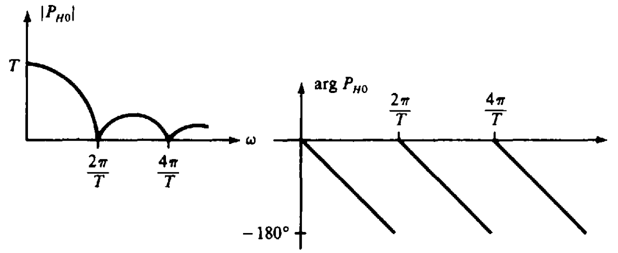


Fig. P6-55

6.56. $\dfrac{Y^*(z)}{U^*(z)} = (1 - z^{-1})^2 \mathcal{Z}\left\{ \mathcal{L}^{-1}\left(\dfrac{G(s)}{s} + \dfrac{1}{T}\dfrac{G(s)}{s^2} \right)\Bigg|_{t=kT} \right\}$

Chapter 7

Block Diagram Algebra and Transfer Functions of Systems

7.1 INTRODUCTION

It is pointed out in Chapters 1 and 2 that the block diagram is a shorthand, graphical representation of a physical system, illustrating the functional relationships among its components. This latter feature permits evaluation of the contributions of the individual elements to the overall performance of the system.

In this chapter we first investigate these relationships in more detail, utilizing the frequency domain and transfer function concepts developed in preceding chapters. Then we develop methods for reducing complicated block diagrams to manageable forms so that they may be used to predict the overall performance of a system.

7.2 REVIEW OF FUNDAMENTALS

In general, a block diagram consists of a specific configuration of four types of elements: blocks, summing points, takeoff points, and arrows representing unidirectional signal flow:

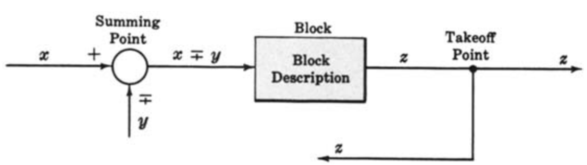


Fig. 7-1

The meaning of each element should be clear from Fig. 7-1.

Time-domain quantities are represented by lowercase letters.

EXAMPLE 7.1. $r = r(t)$ for continuous signals, and $r(t_k)$ or $r(k)$, $k = 1, 2, \ldots$, for discrete-time signals.

Capital letters in this chapter are used for Laplace transforms, or z-transforms. The argument s or z is often suppressed, to simplify the notation, if the context is clear, or if the results presented are the same for both Laplace (continuous-time system) and z-(discrete-time system)transfer function domains.

EXAMPLE 7.2. $R = R(s)$ or $R = R(z)$.

The basic feedback control system configuration presented in Chapter 2 is reproduced in Fig. 7-2, with all quantities in abbreviated transform notation.

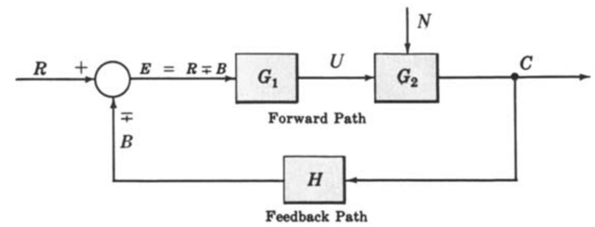


Fig. 7-2

The quantities G_1, G_2, and H are the transfer functions of the components in the blocks. They may be either Laplace or z-transform transfer functions.

EXAMPLE 7.3. $G_1 = U/E$ or $U = G_1 E$.

It is important to note that these results apply *either* to Laplace transform *or* to z-transform transfer functions, but not necessarily to *mixed* continuous/discrete block diagrams that include *samplers*. Samplers are linear devices, but they are not time-invariant. Therefore they cannot be characterized by an ordinary s-domain transfer function, as defined in Chapter 6. See Problem 7.38 for some exceptions, and Section 6.8 for a more extensive discussion of mixed continuous/discrete systems.

7.3 BLOCKS IN CASCADE

Any finite number of blocks in series may be algebraically combined by multiplication of transfer functions. That is, n components or blocks with transfer functions $G_1, G_2, \ldots, G_n$ connected in cascade are equivalent to a single element G with a transfer function given by

$$G = G_1 \cdot G_2 \cdot G_3 \cdots G_n = \prod_{i=1}^{n} G_i \tag{7.1}$$

The symbol for multiplication "·" is omitted when no confusion results.

EXAMPLE 7.4.

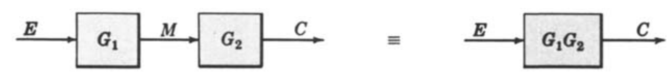


Fig. 7-3

Multiplication of transfer functions is *commutative*; that is,

$$G_i G_j = G_j G_i \tag{7.2}$$

for any i or j.

EXAMPLE 7.5.

Fig. 7-4

Loading effects (interaction of one transfer function upon its neighbor) must be accounted for in the derivation of the individual transfer functions before blocks can be cascaded. (See Problem 7.4.)

7.4 CANONICAL FORM OF A FEEDBACK CONTROL SYSTEM

The two blocks in the forward path of the feedback system of Fig. 7-2 may be combined. Letting $G \equiv G_1G_2$, the resulting configuration is called the **canonical form** of a feedback control system. G and H are not necessarily unique for a particular system.

The following definitions refer to Fig. 7-5.

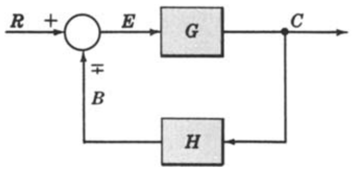


Fig. 7-5

Definition 7.1: $G \equiv$ direct transfer function $\equiv$ forward transfer function

Definition 7.2: $H \equiv$ feedback transfer function

Definition 7.3: $GH \equiv$ loop transfer function $\equiv$ open-loop transfer function

Definition 7.4: $C/R \equiv$ closed-loop transfer function $\equiv$ control ratio

Definition 7.5: $E/R \equiv$ actuating signal ratio $\equiv$ error ratio

Definition 7.6: $B/R \equiv$ primary feedback ratio

In the following equations, the $-$ sign refers to a *positive* feedback system, and the $+$ sign refers to a *negative* feedback system:

$$\frac{C}{R} = \frac{G}{1 \pm GH} \tag{7.3}$$

$$\frac{E}{R} = \frac{1}{1 \pm GH} \tag{7.4}$$

$$\frac{B}{R} = \frac{GH}{1 \pm GH} \tag{7.5}$$

The denominator of C/R determines the *characteristic equation* of the system, which is usually determined from $1 \pm GH = 0$ or, equivalently,

$$D_{GH} \pm N_{GH} = 0 \tag{7.6}$$

where D_{GH} is the denominator and N_{GH} is the numerator of GH, unless a pole of G cancels a zero of H (see Problem 7.9). Relations (*7.1*) through (*7.6*) are valid for both continuous (s-domain) and discrete (z-domain) systems.

7.5 BLOCK DIAGRAM TRANSFORMATION THEOREMS

Block diagrams of complicated control systems may be simplified using easily derivable transformations. The first important transformation, combining blocks in cascade, has already been presented in Section 7.3. It is repeated for completeness in the chart illustrating the transformation theorems (Fig. 7-6). The letter P is used to represent any transfer function, and W, X, Y, Z denote any transformed signals.

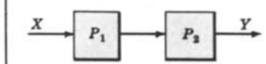
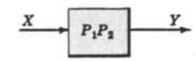
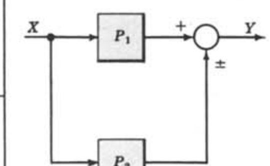
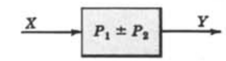
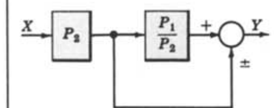
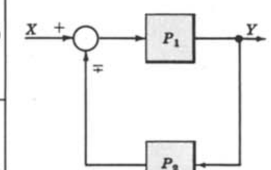
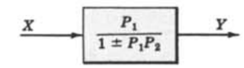
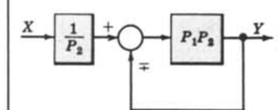
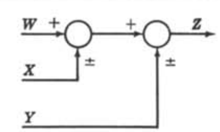
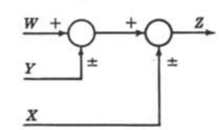
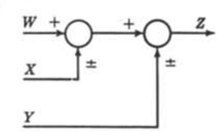
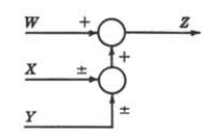
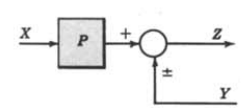
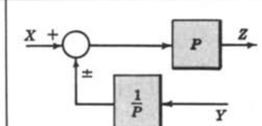
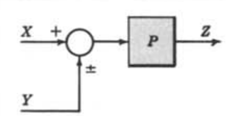
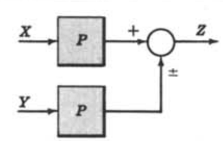
	Transformation	Equation	Block Diagram	Equivalent Block Diagram
1	Combining Blocks in Cascade	$Y = (P_1P_2)X$		
2	Combining Blocks in Parallel; or Eliminating a Forward Loop	$Y = P_1X \pm P_2X$		
3	Removing a Block from a Forward Path	$Y = P_1X \pm P_2X$		
4	Eliminating a Feedback Loop	$Y = P_1(X \mp P_2Y)$		
5	Removing a Block from a Feedback Loop	$Y = P_1(X \mp P_2Y)$		
6a	Rearranging Summing Points	$Z = W \pm X \pm Y$		
6b	Rearranging Summing Points	$Z = W \pm X \pm Y$		
7	Moving a Summing Point Ahead of a Block	$Z = PX \pm Y$		
8	Moving a Summing Point Beyond a Block	$Z = P[X \pm Y]$		

Fig. 7-6

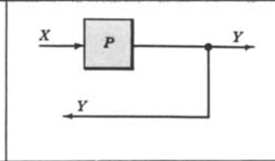
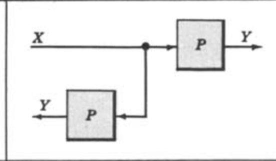
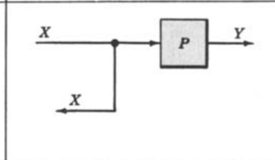
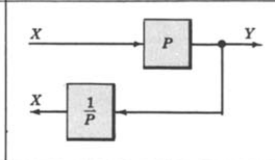
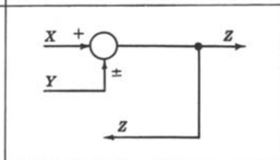
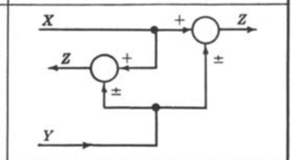
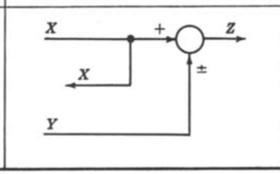
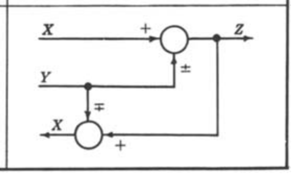
	Transformation	Equation	Block Diagram	Equivalent Block Diagram
9	Moving a Takeoff Point Ahead of a Block	$Y = PX$		
10	Moving a Takeoff Point Beyond a Block	$Y = PX$		
11	Moving a Takeoff Point Ahead of a Summing Point	$Z = X \pm Y$		
12	Moving a Takeoff Point Beyond a Summing Point	$Z = X \pm Y$		

Fig. 7-6 *Continued*

7.6 UNITY FEEDBACK SYSTEMS

Definition 7.7: A **unity feedback system** is one in which the primary feedback b is identically equal to the controlled output c.

EXAMPLE 7.6. $H = 1$ for a linear, unity feedback system (Fig. 7-7).

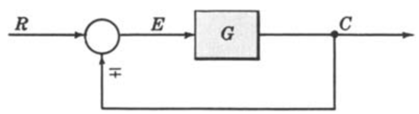


Fig. 7-7

Any feedback system with only linear time-invariant elements can be put into the form of a unity feedback system by using Transformation 5.

EXAMPLE 7.7.

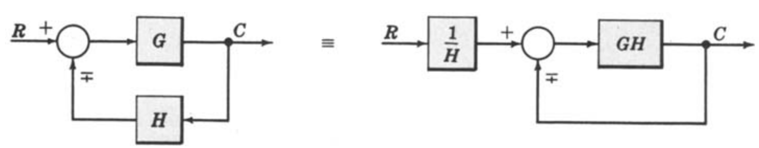


Fig. 7-8

The characteristic equation for the unity feedback system, determined from $1 \pm G = 0$, is

$$D_G \pm N_G = 0 \tag{7.7}$$

where D_G is the denominator and N_G the numerator of G.

7.7 SUPERPOSITION OF MULTIPLE INPUTS

Sometimes it is necessary to evaluate system performance when several inputs are simultaneously applied at different points of the system.

When multiple inputs are present in a *linear* system, each is treated independently of the others. The output due to all stimuli acting together is found in the following manner. We assume zero initial conditions, as we seek the system response only to inputs.

Step 1: Set all inputs except one equal to zero.

Step 2: Transform the block diagram to canonical form, using the transformations of Section 7.5.

Step 3: Calculate the response due to the chosen input acting alone.

Step 4: Repeat Steps 1 to 3 for each of the remaining inputs.

Step 5: Algebraically add all of the responses (outputs) determined in Steps 1 to 4. This sum is the total output of the system with all inputs acting simultaneously.

We reemphasize here that the above superposition process is dependent on the system being linear.

EXAMPLE 7.8. We determine the output C due to inputs U and R for Fig. 7-9.

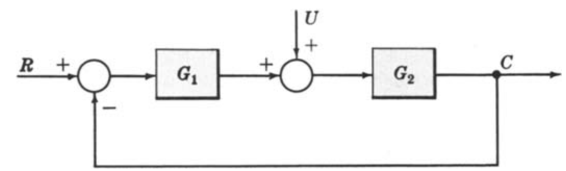
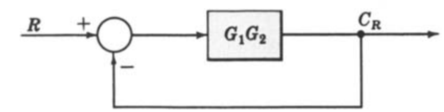
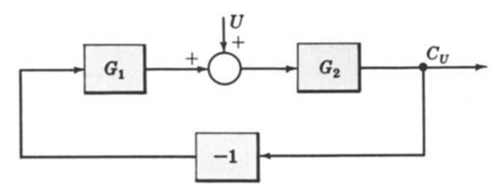


Fig. 7-9

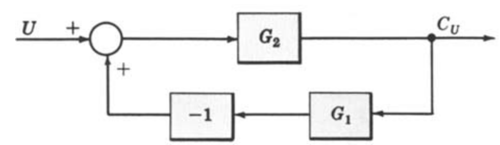
Step 1: Put $U \equiv 0$.
Step 2: The system reduces to



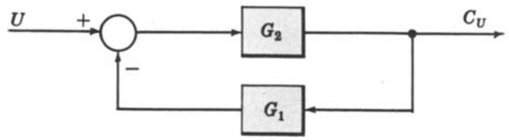
Step 3: By Equation (7.3), the output C_R due to input R is $C_R = [G_1G_2/(1 + G_1G_2)]R$.
Step 4a: Put $R = 0$.
Step 4b: Put -1 into a block, representing the negative feedback effect:



Rearrange the block diagram:



Let the -1 block be absorbed into the summing point:



Step 4c: By Equation (7.3), the output C_U due to input U is $C_U = [G_2/(1 + G_1G_2)]U$.

Step 5: The total output is

$$C = C_R + C_U = \left[\frac{G_1G_2}{1+G_1G_2}\right]R + \left[\frac{G_2}{1+G_1G_2}\right]U = \left[\frac{G_2}{1+G_1G_2}\right][G_1R + U]$$

7.8 REDUCTION OF COMPLICATED BLOCK DIAGRAMS

The block diagram of a practical feedback control system is often quite complicated. It may include several feedback or feedforward loops, and multiple inputs. By means of systematic block diagram reduction, every multiple loop linear feedback system may be reduced to canonical form. The techniques developed in the preceding paragraphs provide the necessary tools.

The following general steps may be used as a basic approach in the reduction of complicated block diagrams. Each step refers to specific transformations listed in Fig. 7-6.

Step 1: Combine all cascade blocks using Transformation 1.

Step 2: Combine all parallel blocks using Transformation 2.

Step 3: Eliminate all minor feedback loops using Transformation 4.

Step 4: Shift summing points to the left and takeoff points to the right of the major loop, using Transformations 7, 10, and 12.

Step 5: Repeat Steps 1 to 4 until the canonical form has been achieved for a particular input.

Step 6: Repeat Steps 1 to 5 for each input, as required.

Transformations 3, 5, 6, 8, 9, and 11 are sometimes useful, and experience with the reduction technique will determine their application.

EXAMPLE 7.9. Let us reduce the block diagram (Fig. 7-10) to canonical form.

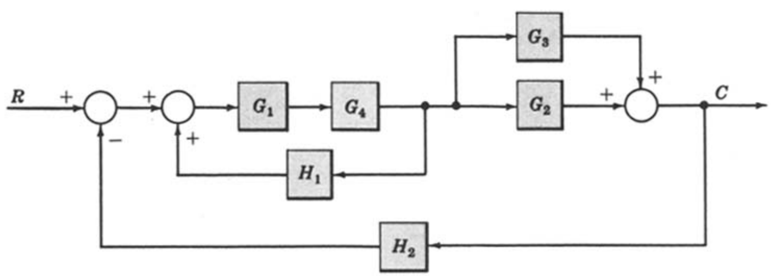


Fig. 7-10

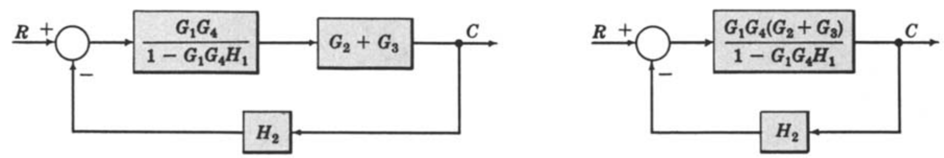
Step 1:

Step 2:

Step 3:

Step 4: Does not apply.

Step 5:



Step 6: Does not apply.

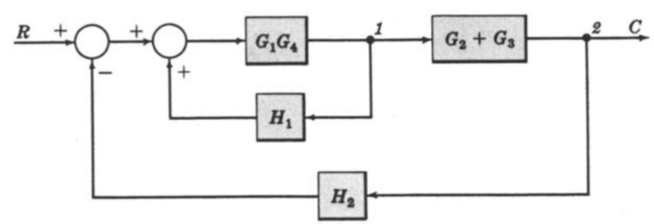
An occasional requirement of block diagram reduction is the isolation of a particular block in a feedback or feedforward loop. This may be desirable to more easily examine the effect of a particular block on the overall system.

Isolation of a block generally may be accomplished by applying the same reduction steps to the system, but usually in a different order. Also, the block to be isolated cannot be combined with any others.

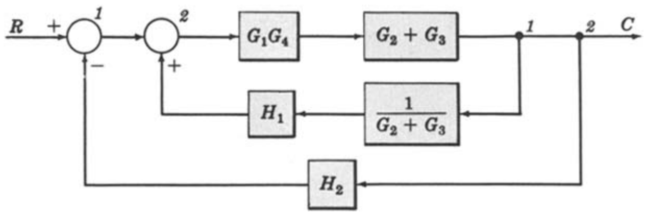
Rearranging Summing Points (Transformation 6) and Transformations 8, 9, and 11 are especially useful for isolating blocks.

EXAMPLE 7.10. Let us reduce the block diagram of Example 7.9, isolating block H_1.

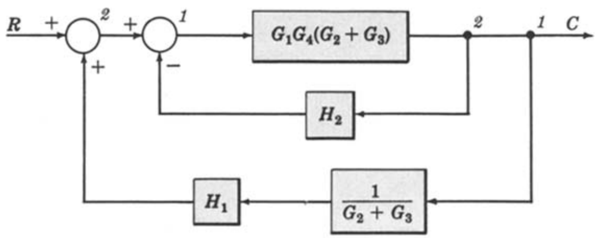
Steps 1 and 2:



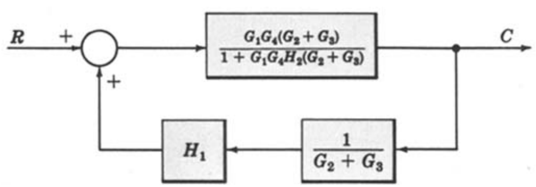
We do not apply Step 3 at this time, but go directly to Step 4, moving takeoff point *1* beyond block $G_2 + G_3$:



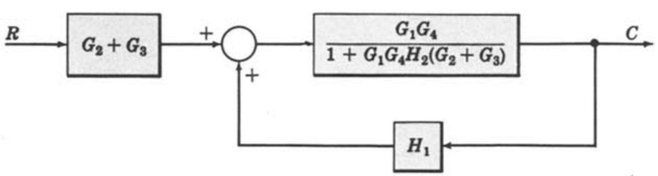
We may now rearrange summing points *1* and *2* and combine the cascade blocks in the forward loop using Transformation 6, then Transformation 1:



Step 3:



Finally, we apply Transformation 5 to remove $1/(G_2 + G_3)$ from the feedback loop:



Note that the same result could have been obtained after applying Step 2 by moving takeoff point *2 ahead* of $G_2 + G_3$, instead of takeoff point *1 beyond* $G_2 + G_3$. Block $G_2 + G_3$ has the same effect on the control ratio C/R whether it directly follows R or directly precedes C.

Solved Problems

BLOCKS IN CASCADE

7.1. Prove Equation (*7.1*) for blocks in cascade.

The block diagram for n transfer functions $G_1, G_2, \ldots, G_n$ in cascade is given in Fig. 7-11.

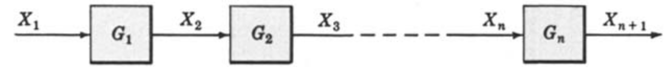


Fig. 7-11

The output transform for any block is equal to the input transform multiplied by the transfer function (see Section 6.1). Therefore $X_2 = X_1G_1$, $X_3 = X_2G_2, \ldots, X_n = X_{n-1}G_{n-1}$, $X_{n+1} = X_nG_n$. Combining these equations, we have

$$X_{n+1} = X_nG_n = X_{n-1}G_{n-1}G_n = \cdots = X_1G_1G_2 \cdots G_{n-1}G_n$$

Dividing both sides by X_1, we obtain $X_{n+1}/X_1 = G_1G_2 \cdots G_{n-1}G_n$.

7.2. Prove the commutativity of blocks in cascade, Equation (*7.2*).

Consider two blocks in cascade (Fig. 7-12):

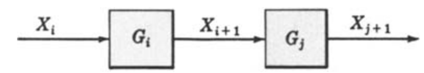


Fig. 7-12

From Equation (*6.1*) we have $X_{i+1} = X_iG_i = G_iX_i$ and $X_{j+1} = X_{i+1}G_j = G_jX_{i+1}$. Therefore $X_{j+1} = (X_iG_i)G_j = X_iG_iG_j$. Dividing both sides by X_i, $X_{j+1}/X_i = G_iG_j$.

Also, $X_{j+1} = G_j(G_iX_i) = G_jG_iX_i$. Dividing again by X_i, $X_{j+1}/X_i = G_jG_i$. Thus $G_iG_j = G_jG_i$.

This result is extended by mathematical induction to any finite number of transfer functions (blocks) in cascade.

7.3. Find X_n/X_1 for each of the systems in Fig. 7-13.

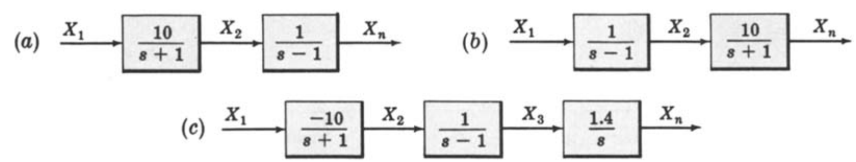


Fig. 7-13

(*a*) One way to work this problem is to first write X_2 in terms of X_1:

$$X_2 = \left(\frac{10}{s+1}\right)X_1$$

Then write X_n in terms of X_2:

$$X_n = \left(\frac{1}{s-1}\right)X_2 = \left(\frac{1}{s-1}\right)\left(\frac{10}{s+1}\right)X_1$$

Multiplying out and dividing both sides by X_1, we have $X_n/X_1 = 10/(s^2 - 1)$.

A shorter method is as follows. We know from Equation (*7.1*) that two blocks can be reduced to one by simply multiplying their transfer functions. Also, the transfer function of a single block is its output-to-input transform. Hence

$$\frac{X_n}{X_1} = \left(\frac{1}{s-1}\right)\left(\frac{10}{s+1}\right) = \frac{10}{s^2-1}$$

(*b*) This system has the same transfer function determined in part (*a*) because multiplication of transfer functions is commutative.

(*c*) By Equation (*7.1*), we have

$$\frac{X_n}{X_1} = \left(\frac{-10}{s+1}\right)\left(\frac{1}{s-1}\right)\left(\frac{1.4}{s}\right) = \frac{-14}{s(s^2-1)}$$

7.4. The transfer function of Fig. 7-14a is $\omega_0/(s+\omega_0)$, where $\omega_0 = 1/RC$. Is the transfer function of Fig. 7-14b equal to $\omega_0^2/(s+\omega_0)^2$? Why?

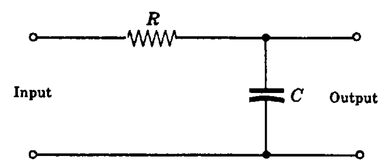


Fig. 7-14a

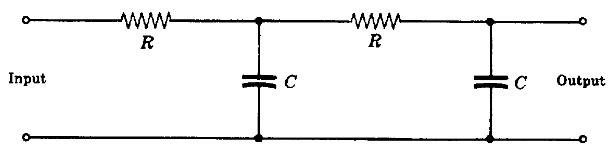


Fig. 7-14b

No. If two networks are connected in series (Fig. 7-15) the second loads the first by drawing current from it. Therefore Equation (*7.1*) cannot be directly applied to the combined system. The correct transfer function for the connected networks is $\omega_0^2/(s^2+3\omega_0 s+\omega_0^2)$ (see Problem 6.16), and this is *not* equal to $(\omega_0/(s+\omega_0))^2$.

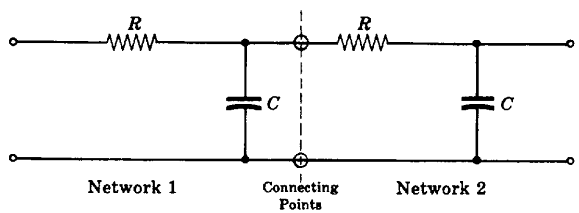


Fig. 7-15

CANONICAL FEEDBACK CONTROL SYSTEMS

7.5. Prove Equation (*7.3*), $C/R = G/(1 \pm GH)$.

The equations describing the canonical feedback system are taken directly from Fig. 7-16. They are given by $E = R \mp B$, $B = HC$, and $C = GE$. Substituting one into the other, we have

$$C = G(R \mp B) = G(R \mp HC)$$
$$= GR \mp GHC = GR + (\mp GHC)$$

Subtracting $(\mp GHC)$ from both sides, we obtain $C \pm GHC = GR$ or $C/R = G/(1 \pm GH)$.

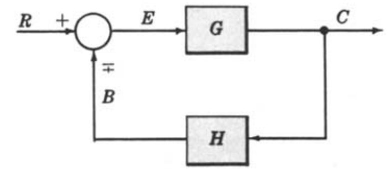


Fig. 7-16

7.6. Prove Equation (*7.4*), $E/R = 1/(1 \pm GH)$.

From the preceding problem, we have $E = R \mp B$, $B = HC$, and $C = GE$. Then $E = R \mp HC = R \mp HGE$, $E \pm GHE = R$, and $E/R = 1/(1 \pm GH)$.

7.7. Prove Equation (*7.5*), $B/R = GH/(1 \pm GH)$.

From $E = R \mp B$, $B = HC$, and $C = GE$, we obtain $B = HGE = HG(R \mp B) = GHR \mp GHB$. Then $B \pm GHB = GHR$, $B = GHR/(1 \pm GH)$, and $B/R = GH/(1 \pm GH)$.

7.8. Prove Equation (*7.6*), $D_{GH} \pm N_{GH} = 0$.

The characteristic equation is usually obtained by setting $1 \pm GH = 0$. (See Problem 7.9 for an exception.) Putting $GH \equiv N_{GH}/D_{GH}$, we obtain $D_{GH} \pm N_{GH} = 0$.

7.9. Determine (*a*) the loop transfer function, (*b*) the control ratio, (*c*) the error ratio, (*d*) the primary feedback ratio, (*e*) the characteristic equation, for the feedback control system in which K_1 and K_2 are constants (Fig. 7-17).

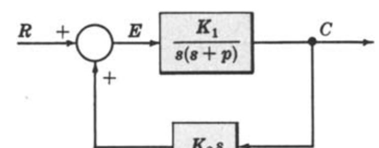


Fig. 7-17

(*a*) The loop transfer function is equal to GH.

Hence
$$GH = \left[\frac{K_1}{s(s+p)}\right] K_2 s = \frac{K_1 K_2}{s+p}$$

(*b*) The control ratio, or closed-loop transfer function, is given by Equation (*7.3*) (with a minus sign for positive feedback):

$$\frac{C}{R} = \frac{G}{1-GH} = \frac{K_1}{s(s+p-K_1K_2)}$$

(*c*) The error ratio, or actuating signal ratio, is given by Equation (*7.4*):

$$\frac{E}{R} = \frac{1}{1-GH} = \frac{1}{1-K_1K_2/(s+p)} = \frac{s+p}{s+p-K_1K_2}$$

(*d*) The primary feedback ratio is given by Equation (*7.5*):

$$\frac{B}{R} = \frac{GH}{1-GH} = \frac{K_1K_2}{s+p-K_1K_2}$$

(*e*) The characteristic equation is given by the denominator of C/R above, $s(s+p-K_1K_2) = 0$. In this case, $1 - GH = s + p - K_1K_2 = 0$, which is *not* the characteristic equation, because the pole s of G cancels the zero s of H.

BLOCK DIAGRAM TRANSFORMATIONS

7.10. Prove the equivalence of the block diagrams for Transformation 2 (Section 7.5).

The equation in the second column, $Y = P_1 X \pm P_2 X$, governs the construction of the block diagram in the third column, as shown. Rewrite this equation as $Y = (P_1 \pm P_2)X$. The equivalent block diagram in the last column is clearly the representation of this form of the equation (Fig. 7-18)

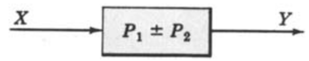


Fig. 7-18

7.11. Repeat Problem 7.10 for Transformation 3.

Rewrite $Y = P_1 X \pm P_2 X$ as $Y = (P_1/P_2)P_2 X \pm P_2 X$. The block diagram for this form of the equation is clearly given in Fig. 7-19.

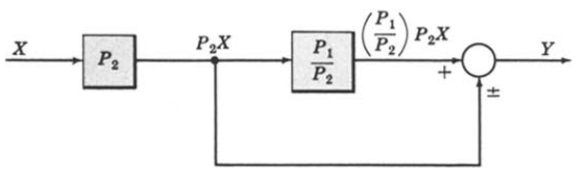


Fig. 7-19

7.12. Repeat Problem 7.10 for Transformation 5.

We have $Y = P_1[X \mp P_2 Y] = P_1 P_2[(1/P_2)X \mp Y]$. The block diagram for the latter form is given in Fig. 7-20.

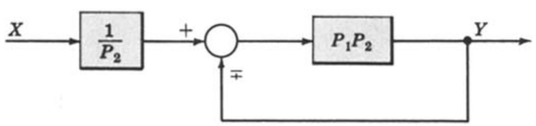


Fig. 7-20

7.13. Repeat Problem 7.10 for Transformation 7.

We have $Z = PX \pm Y = P[X \pm (1/P)Y]$, which yields the block diagram given in Fig. 7-21.

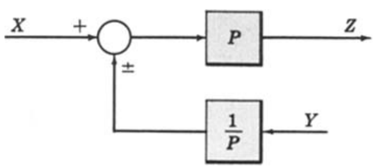


Fig. 7-21

7.14. Repeat Problem 7.10 for Transformation 8.

We have $Z = P(X \pm Y) = PX \pm PY$, whose block diagram is clearly given in Fig. 7-22.

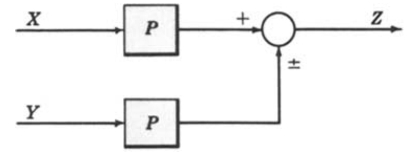


Fig. 7-22

UNITY FEEDBACK SYSTEMS

7.15. Reduce the block diagram given in Fig. 7-23 to unity feedback form and find the system characteristic equation.

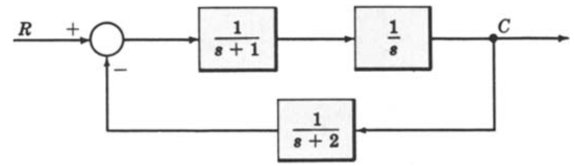


Fig. 7-23

Combining the blocks in the forward path, we obtain Fig. 7-24.

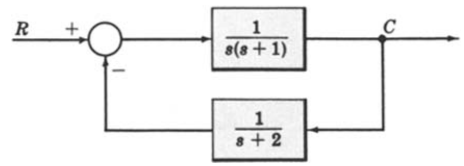


Fig. 7-24

Applying Transformation 5, we have Fig. 7-25.

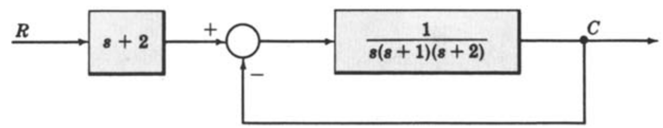


Fig. 7-25

By Equation (7.7), the characteristic equation for this system is $s(s+1)(s+2)+1=0$ or $s^3+3s^2+2s+1=0$.

MULTIPLE INPUTS AND OUTPUTS

7.16. Determine the output C due to U_1, U_2, and R for Fig. 7-26.

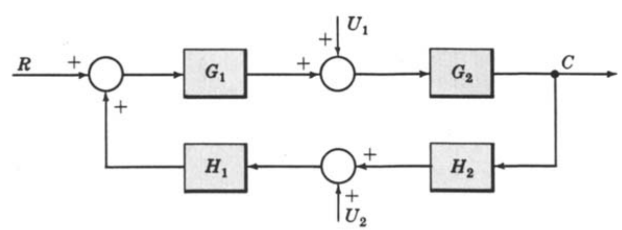


Fig. 7-26

Let $U_1 = U_2 = 0$. After combining the cascaded blocks, we obtain Fig. 7-27, where C_R is the output due to R acting alone. Applying Equation (*7.3*) to this system, $C_R = [G_1G_2/(1 - G_1G_2H_1H_2)]R$.

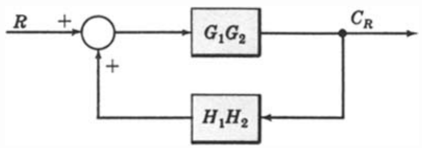


Fig. 7-27

Now let $R = U_2 = 0$. The block diagram is now given in Fig. 7-28, where C_1 is the response due to U_1 acting alone. Rearranging the blocks, we have Fig. 7-29. From Equation (*7.3*), we get $C_1 = [G_2/(1 - G_1G_2H_1H_2)]U_1$.

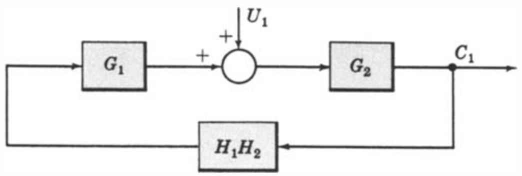


Fig. 7-28

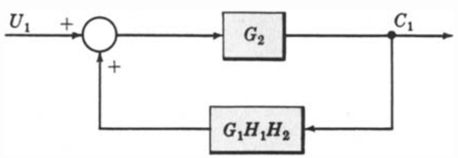


Fig. 7-29

Finally, let $R = U_1 = 0$. The block diagram is given in Fig. 7-30, where C_2 is the response due to U_2 acting alone. Rearranging the blocks, we get Fig. 7-31. Hence $C_2 = [G_1G_2H_1/(1 - G_1G_2H_1H_2)]U_2$.

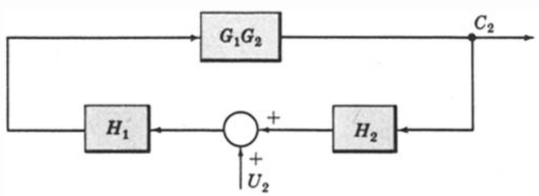


Fig. 7-30

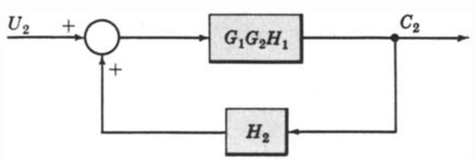


Fig. 7-31

By superposition, the total output is

$$C = C_R + C_1 + C_2 = \frac{G_1G_2R + G_2U_1 + G_1G_2H_1U_2}{1 - G_1G_2H_1H_2}$$

7.17. Figure 7-32 is an example of a multiinput-multioutput system. Determine C_1 and C_2 due to R_1 and R_2.

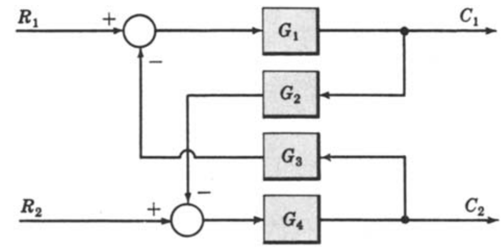


Fig. 7-32

First put the block diagram in the form of Fig. 7-33, ignoring the output C_2.

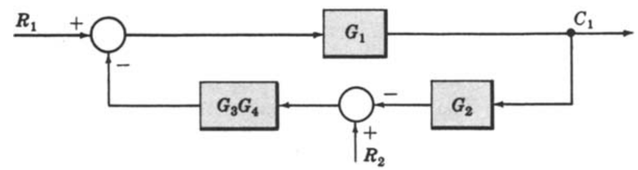


Fig. 7-33

Letting $R_2 = 0$ and combining the summing points, we get Fig. 7-34.

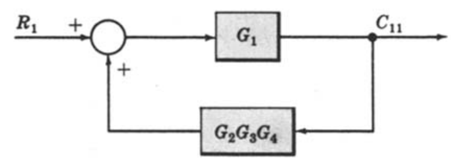


Fig. 7-34

Hence C_{11}, the output at C_1 due to R_1 alone, is $C_{11} = G_1R_1/(1 - G_1G_2G_3G_4)$. For $R_1 = 0$, we have Fig. 7-35.

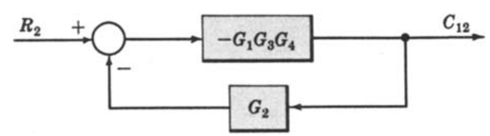


Fig. 7-35

Hence $C_{12} = -G_1G_3G_4R_2/(1 - G_1G_2G_3G_4)$ is the output at C_1 due to R_2 alone. Thus $C_1 = C_{11} + C_{12} = (G_1R_1 - G_1G_3G_4R_2)/(1 - G_1G_2G_3G_4)$.

Now we reduce the original block diagram, ignoring output C_1. First we obtain Fig. 7-36.

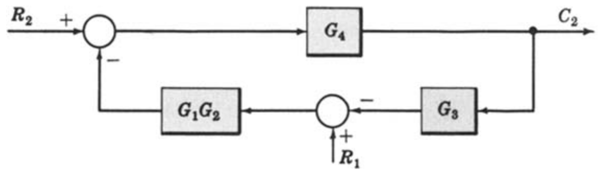


Fig. 7-36

Then we obtain the block diagram given in Fig. 7-37. Hence $C_{22} = G_4R_2/(1 - G_1G_2G_3G_4)$. Next, letting $R_2 = 0$, we obtain Fig. 7-38. Hence $C_{21} = -G_1G_2G_4R_1/(1 - G_1G_2G_3G_4)$. Finally, $C_2 = C_{22} + C_{21} = (G_4R_2 - G_1G_2G_4R_1)/(1 - G_1G_2G_3G_4)$.

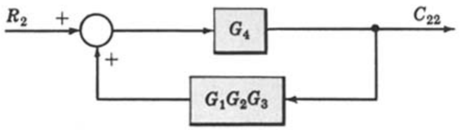


Fig. 7-37

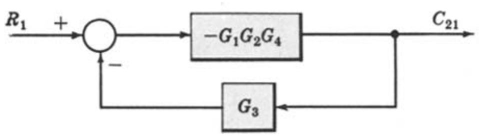


Fig. 7-38

BLOCK DIAGRAM REDUCTION

7.18. Reduce the block diagram given in Fig. 7-39 to canonical form, and find the output transform C. K is a constant.

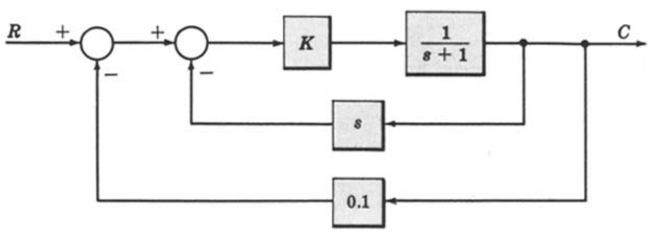


Fig. 7-39

First we combine the cascade blocks of the forward path and apply Transformation 4 to the innermost feedback loop to obtain Fig. 7-40.

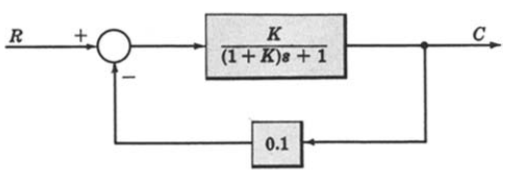


Fig. 7-40

Equation (*7.3*) or the reapplication of Transformation 4 yields $C = KR/[(1 + K)s + (1 + 0.1K)]$.

7.19. Reduce the block diagram of Fig. 7-39 to canonical form, isolating block K in the forward loop.

Mathcad

By Transformation 9 we can move the takeoff point ahead of the $1/(s + 1)$ block (Fig. 7-41):

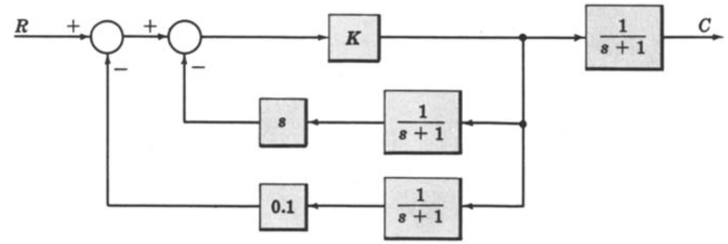


Fig. 7-41

Applying Transformations 1 and 6*b*, we get Fig. 7-42.

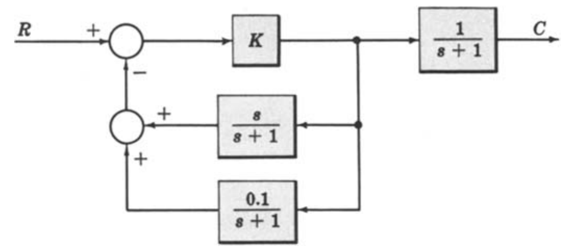


Fig. 7-42

Now we can apply Transformation 2 to the feedback loops, resulting in the final form given in Fig. 7-43.

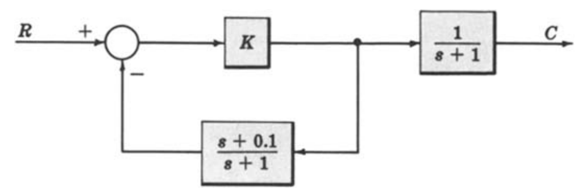


Fig. 7-43

7.20. Reduce the block diagram given in Fig. 7-44 to open-loop form.

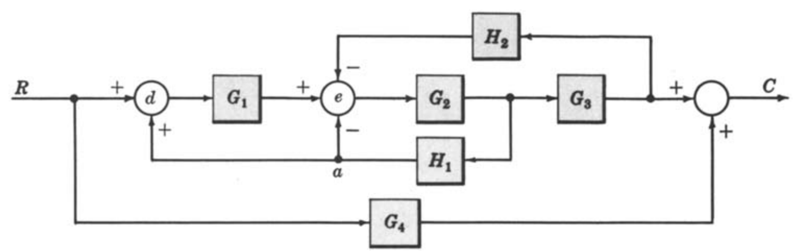


Fig. 7-44

First, moving the leftmost summing point beyond G_1 (Transformation 8), we obtain Fig. 7-45.

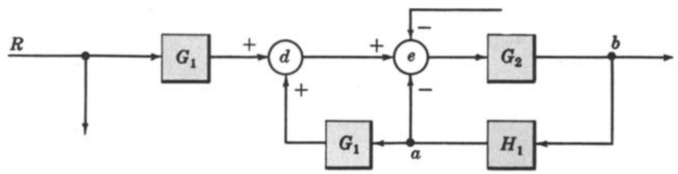


Fig. 7-45

Next, moving takeoff point a beyond G_1, we get Fig. 7-46.

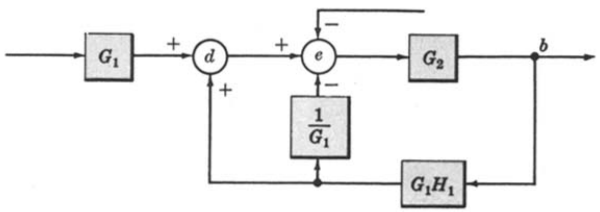


Fig. 7-46

Now, using Transformation 6*b*, and then Transformation 2, to combine the two lower feedback loops (from G_1H_1) entering d and e, we obtain Fig. 7-47.

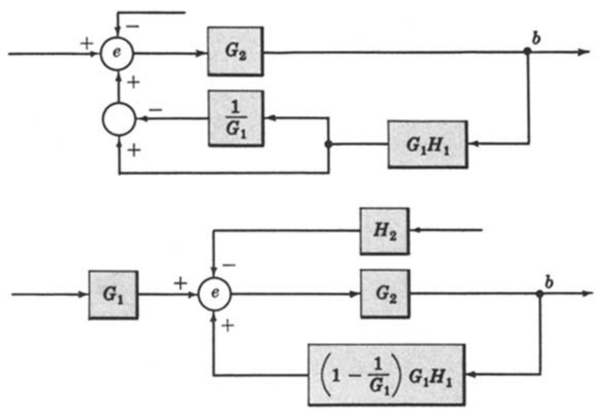
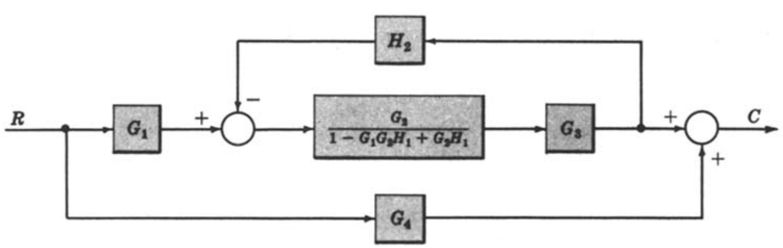
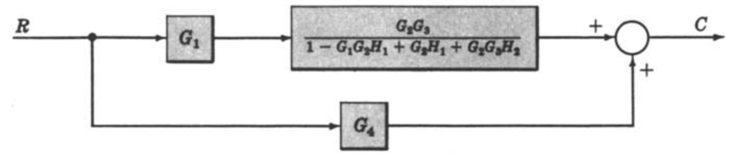


Fig. 7-47

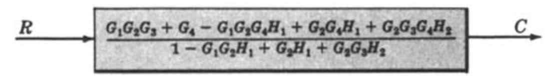
Applying Transformation 4 to this inner loop, the system becomes



Again, applying Transformation 4 to the remaining feedback loop yields



Finally, Transformation 1 and 2 give the open-loop block diagram:



MISCELLANEOUS PROBLEMS

7.21. Show that simple block diagram Transformation 1 of Section 7.5 (combining blocks in cascade) is not valid if the first block is (or includes) a *sampler*.

The output transform $U^*(s)$ of an ideal sampler was determined in Problem 4.39 as

$$U^*(s) = \sum_{k=0}^{\infty} e^{-skT}u(kT)$$

Taking $U^*(s)$ as the input of block P_2 of Transformation 1 of the table, the output transform $Y(s)$ of block P_2 is

$$Y(s) = P_2(s)U^*(s) = P_2(s)\sum_{k=0}^{\infty} e^{-skT}u(kT)$$

Clearly, the input transform $X(s) = U(s)$ cannot be factored from the right-hand side of $Y(s)$, that is, $Y(s) \neq F(s)U(s)$. The same problem occurs if P_1 includes other elements, as well as a sampler.

7.22. Why is the characteristic equation invariant under block diagram transformation?

Block diagram transformations are determined by *rearranging* the input-output equations of one or more of the subsystems that make up the total system. Therefore the final transformed system is governed by the same equations, probably arranged in a different manner than those for the original system.

Now, the characteristic equation is determined from the denominator of the overall system transfer function set equal to zero. Factoring or other rearrangement of the numerator and denominator of the system transfer function clearly does not change it, nor does it alter the denominator set equal to zero.

7.23. Prove that the transfer function represented by C/R in Equation (*7.3*) can be approximated by $\pm 1/H$ when $|G|$ or $|GH|$ are very large.

Dividing the numerator and denominator of $G/(1 \pm GH)$ by G, we get $1\Big/\left(\frac{1}{G} \pm H\right)$. Then

$$\lim_{|G|\to\infty}\left[\frac{C}{R}\right] = \lim_{|G|\to\infty}\left[\frac{1}{\frac{1}{G} \pm H}\right] = \pm\frac{1}{H}$$

Dividing by GH and taking the limit, we obtain

$$\lim_{|GH|\to\infty}\left[\frac{C}{R}\right] = \lim_{|GH|\to\infty}\left[\frac{\frac{1}{H}}{\frac{1}{GH} \pm 1}\right] = \pm\frac{1}{H}$$

7.24. Assume that the characteristics of G change radically or unpredictably during system operation. Using the results of the previous problem, show how the system should be designed so that the output C can always be predicted reasonably well.

In problem 7.23 we found that

$$\lim_{|GH|\to\infty}\left[\frac{C}{R}\right] = \pm\frac{1}{H}$$

Thus $C \to \pm R/H$ as $|GH| \to \infty$, or C is independent of G for large $|GH|$. Hence the system should be designed so that $|GH| \gg 1$.

7.25. Determine the transfer function of the system in Fig. 7-48. Then let $H_1 = 1/G_1$ and $H_2 = 1/G_2$.

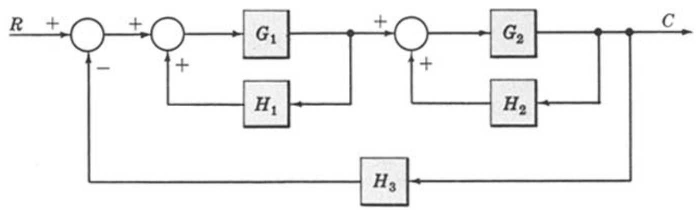


Fig. 7-48

Reducing the inner loops, we have Fig. 7-49.

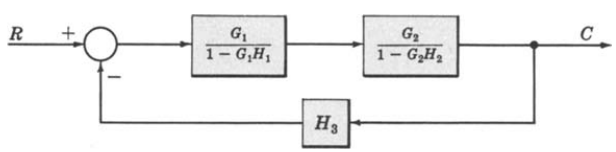


Fig. 7-49

Applying Transformation 4 again, we obtain Fig. 7-50.

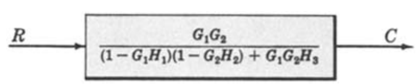


Fig. 7-50

Now put $H_1 = 1/G_1$ and $H_2 = 1/G_2$. This yields

$$\frac{C}{R} = \frac{G_1G_2}{(1-1)(1-1) + G_1G_2H_3} = \frac{1}{H_3}$$

7.26. Show that Fig. 7-51 is valid.

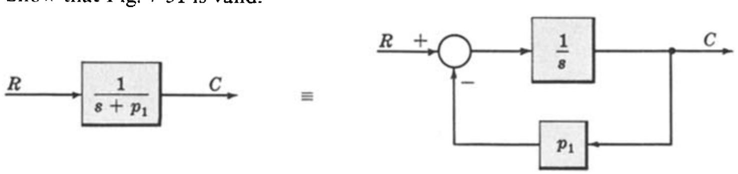


Fig. 7-51

From the open-loop diagram, we have $C = R/(s + p_1)$. Rearranging, $(s + p_1)C = R$ and $C = (1/s)(R - p_1C)$. The closed-loop diagram follows from this equation.

7.27. Prove Fig. 7-52.

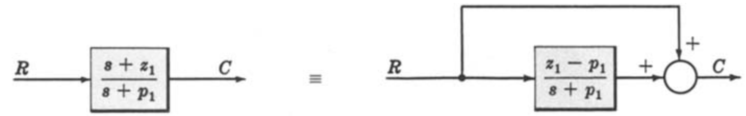


Fig. 7-52

This problem illustrates how a finite zero may be removed from a block.

From the forward-loop diagram, $C = R + (z_1 - p_1)R/(s + p_1)$. Rearranging,

$$C = \left(1 + \frac{z_1 - p_1}{s + p_1}\right)R = \left(\frac{s + p_1 + z_1 - p_1}{s + p_1}\right)R = \left(\frac{s + z_1}{s + p_1}\right)R$$

This mathematical equivalence clearly proves the equivalence of the block diagrams.

7.28. Assume that linear approximations in the form of transfer functions are available for each block of the Supply and Demand System of Problem 2.13, and that the system can be represented by Fig. 7-53.

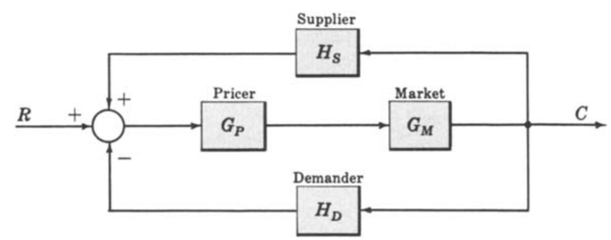


Fig. 7-53

Determine the overall transfer function of the system.

Block diagram Transformation 4, applied twice to this system, gives Fig. 7-54.

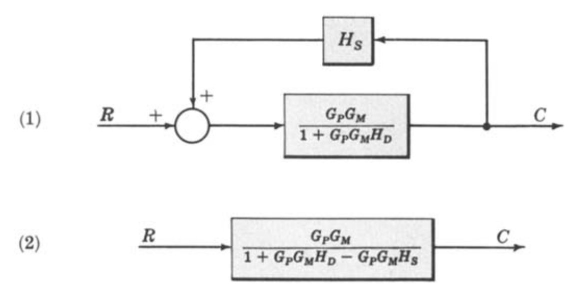


Fig. 7-54

Hence the transfer function for the linearized Supply and Demand model is: $\dfrac{G_PG_M}{1 + G_PG_M(H_D - H_S)}$.

Supplementary Problems

7.29. Determine C/R for each system in Fig. 7-55.

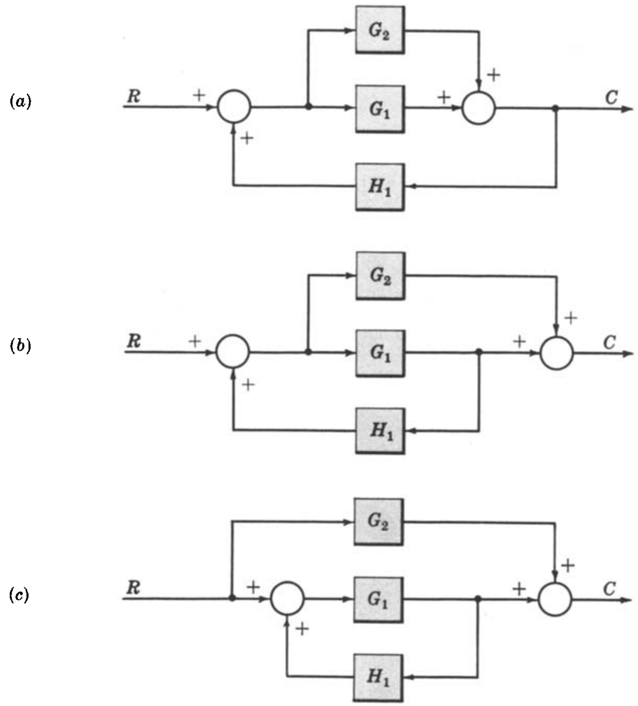


Fig. 7-55

7.30. Consider the blood pressure regulator described in Problem 2.14. Assume the vasomotor center (VMC) can be described by a linear transfer function $G_{11}(s)$, and the baroreceptors by the transfer function $k_1 s + k_2$ (see Problem 6.33). Transform the block diagram into its simplest, unity feedback form.

7.31. Reduce Fig. 7-56 to canonical form.

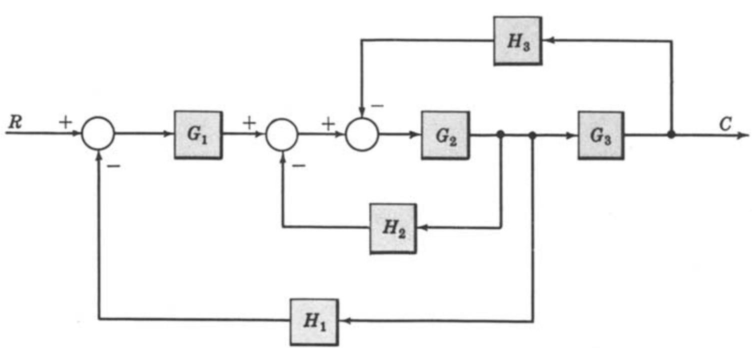


Fig. 7-56

7.32. Determine C for the system represented by Fig. 7-57.

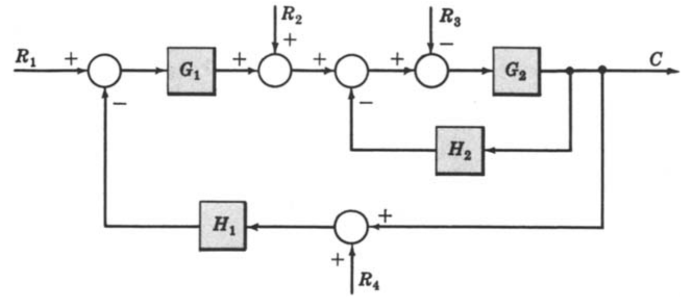


Fig. 7-57

7.33. Give an example of two feedback systems in canonical form having identical control ratios C/R but different G and H components.

7.34. Determine C/R_2 for the system given in Fig. 7-58.

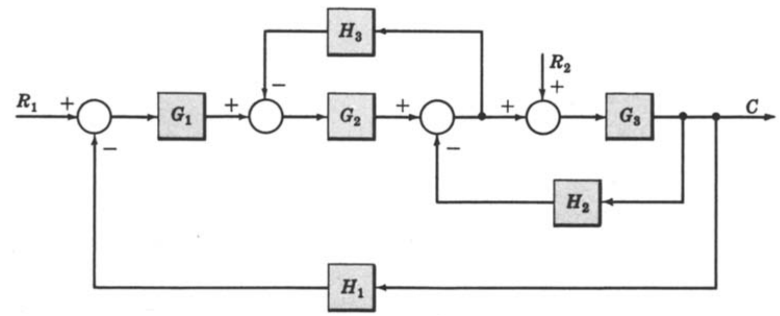


Fig. 7-58

7.35. Determine the complete output C, with both inputs R_1 and R_2 acting simultaneously, for the system given in the preceding problem.

7.36. Determine C/R for the system represented by Fig. 7-59.

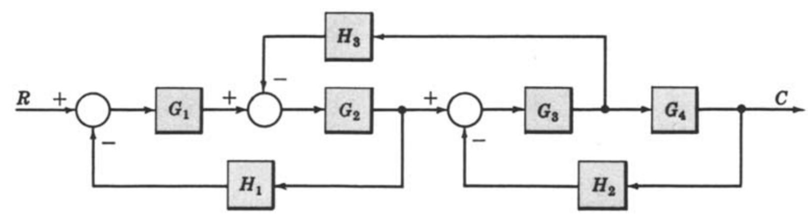


Fig. 7-59

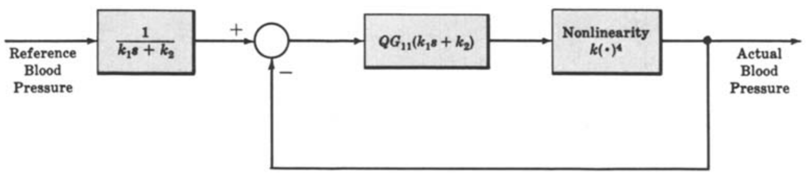
7.37. Determine the characteristic equation for each of the systems of Problems (a) 7.32, (b) 7.35, (c) 7.36.

7.38. What block diagram transformation rules in the table of Section 7.5 permit the inclusion of a sampler?

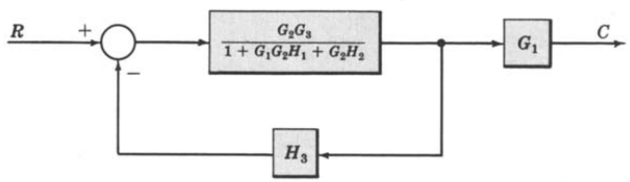
Answers to Supplementary Problems

7.29. See Problem 8.15.

7.30.



7.31.



7.32. $C = \dfrac{G_1G_2R_1 + G_2R_2 - G_2R_3 - G_1G_2H_1R_4}{1 + G_2H_2 + G_1G_2H_1}$

7.34. $\dfrac{C}{R_2} = \dfrac{G_3(1 + G_2H_3)}{1 + G_3H_2 + G_2H_3 + G_1G_2G_3H_1}$

7.35. $C = \dfrac{G_1G_2G_3R_1 + G_3(1 + G_2H_3)R_2}{1 + G_3H_2 + G_2H_3 + G_1G_2G_3H_1}$

7.36. $\dfrac{C}{R} = \dfrac{G_1G_2G_3G_4}{(1 + G_1G_2H_1)(1 + G_3G_4H_2) + G_2G_3H_3}$

7.37. (*a*) $1 + G_2H_2 + G_1G_2H_1 = 0$

(*b*) $1 + G_3H_2 + G_2H_3 + G_1G_2G_3H_1 = 0$

(*c*) $(1 + G_1G_2H_1)(1 + G_3G_4H_2) + G_2G_3H_3 = 0.$

7.38. The results of Problem 7.21 indicate that any transformation that involves any *product* of two or more transforms is not valid if a sampler is included. But all those that simply involve the sum or difference of signals are valid, that is, Transformations 6, 11, and 12. Each represents a simple rearrangement of signals as a linear-sum, and addition is a commutative operation, even for sampled signals, that is $Z = X \pm Y = Y \pm X$.

Chapter 8

Signal Flow Graphs

8.1 INTRODUCTION

The most extensively used graphical representation of a feedback control system is the block diagram, presented in Chapters 2 and 7. In this chapter we consider another model, the signal flow graph.

A **signal flow graph** is a pictorial representation of the simultaneous equations describing a system. It graphically displays the transmission of signals through the system, as does the block diagram. But it is easier to draw and therefore easier to manipulate than the block diagram.

The properties of signal flow graphs are presented in the next few sections. The remainder of the chapter treats applications.

8.2 FUNDAMENTALS OF SIGNAL FLOW GRAPHS

Let us first consider the simple equation

$$X_i = A_{ij}X_j \tag{8.1}$$

The variables X_i and X_j can be functions of time, complex frequency, or any other quantity. They may even be constants, which are "variables" in the mathematical sense.

For signal flow graphs, A_{ij} is a mathematical operator mapping X_j into X_i, and is called the **transmission function**. For example, A_{ij} may be a constant, in which case X_i is a constant times X_j in Equation (*8.1*); if X_i and X_j are functions of s or z, A_{ij} may be a transfer function $A_{ij}(s)$ or $A_{ij}(z)$.

The signal flow graph for Equation (*8.1*) is given in Fig. 8-1. This is the simplest form of a signal flow graph. Note that the variables X_i and X_j are represented by a small dot called a **node**, and the transmission function A_{ij} is represented by a line with an arrow, called a **branch**.

Fig. 8-1

Every variable in a signal flow graph is designated by a node, and every transmission function by a branch. Branches are always unidirectional. The arrow denotes the direction of signal flow.

EXAMPLE 8.1. Ohm's law states that $E = RI$, where E is a voltage, I a current, and R a resistance. The signal flow graph for this equation is given in Fig. 8-2.

Fig. 8-2

8.3 SIGNAL FLOW GRAPH ALGEBRA

1. The Addition Rule

The value of the variable designated by a node is equal to the sum of all signals entering the node. In other words, the equation

$$X_i = \sum_{j=1}^{n} A_{ij} X_j$$

is represented by Fig. 8-3.

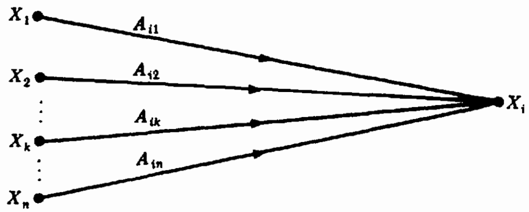


Fig. 8-3

EXAMPLE 8.2. The signal flow graph for the equation of a line in rectangular coordinates, $Y = mX + b$, is given in Fig. 8-4. Since b, the Y-axis intercept, is a constant it may represent a node (variable) or a transmission function.

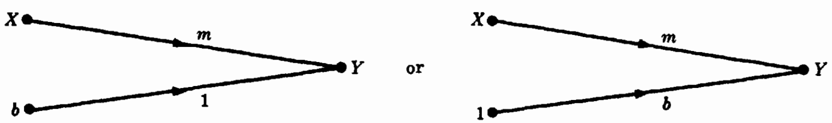


Fig. 8-4

2. The Transmission Rule

The value of the variable designated by a node is transmitted on every branch leaving that node. In other words, the equation

$$X_i = A_{ik} X_k \qquad i = 1, 2, \ldots, n, \; k \text{ fixed}$$

is represented by Fig. 8-5.

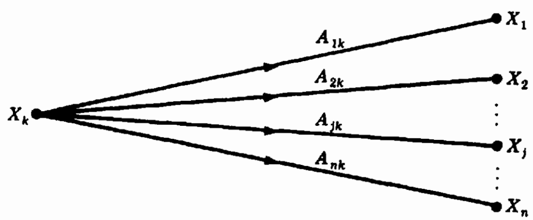


Fig. 8-5

EXAMPLE 8.3. The signal flow graph of the simultaneous equations $Y = 3X$, $Z = -4X$ is given in Fig. 8-6.

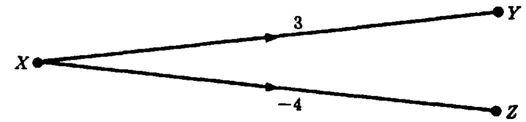


Fig. 8-6

3. The Multiplication Rule

A cascaded (series) connection of $n - 1$ branches with transmission functions $A_{21}, A_{32}, A_{43}, \ldots, A_{n(n-1)}$ can be replaced by a single branch with a new transmission function equal to the product of the old ones. That is,

$$X_n = A_{21} \cdot A_{32} \cdot A_{43} \cdots A_{n(n-1)} \cdot X_1$$

The signal flow graph equivalence is represented by Fig. 8-7.

Fig. 8-7

EXAMPLE 8.4. The signal flow graph of the simultaneous equations $Y = 10X$, $Z = -20Y$ is given in Fig. 8-8.

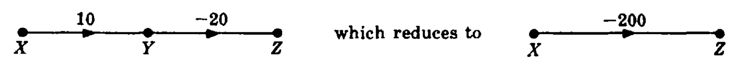


Fig. 8-8

8.4 DEFINITIONS

The following terminology is frequently used in signal flow graph theory. The examples associated with each definition refer to Fig. 8-9.

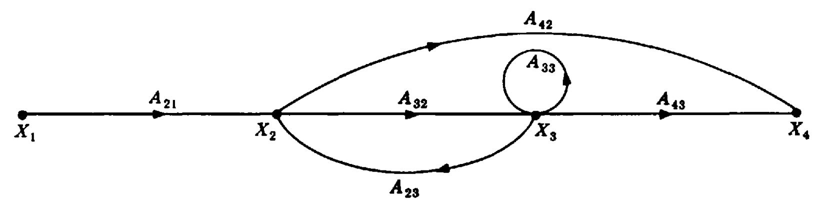


Fig. 8-9

Definition 8.1: A **path** is a continuous, unidirectional succession of branches along which no node is passed more than once. For example, X_1 to X_2 to X_3 to X_4, X_2 to X_3 and back to X_2, and X_1 to X_2 to X_4 are paths.

Definition 8.2: An **input node** or **source** is a node with only outgoing branches. For example, X_1 is an input node.

Definition 8.3: An **output node** or **sink** is a node with only incoming branches. For example, X_4 is an output node.

Definition 8.4: A **forward path** is a path from the input node to the output node. For example, X_1 to X_2 to X_3 to X_4, and X_1 to X_2 to X_4 are forward paths.

Definition 8.5: A **feedback path** or **feedback loop** is a path which originates and terminates on the same node. For example, X_2 to X_3 and back to X_2 is a feedback path.

Definition 8.6: A **self-loop** is a feedback loop consisting of a single branch. For example, A_{33} is a self-loop.

Definition 8.7: The **gain** of a branch is the transmission function of that branch when the transmission function is a multiplicative operator. For example, A_{33} is the gain of the self-loop if A_{33} is a constant or transfer function.

Definition 8.8: The **path gain** is the product of the branch gains encountered in traversing a path. For example, the path gain of the forward path from X_1 to X_2 to X_3 to X_4 is $A_{21}A_{32}A_{43}$.

Definition 8.9: The **loop gain** is the product of the branch gains of the loop. For example, the loop gain of the feedback loop from X_2 to X_3 and back to X_2 is $A_{32}A_{23}$.

Very often, a variable in a system is a function of the output variable. The canonical feedback system is an obvious example. In this case, if the signal flow graph were to be drawn directly from the equations, the "output node" would require an outgoing branch, contrary to the definition. This problem may be remedied by adding a branch with a transmission function of unity entering a "dummy" node. For example, the two graphs in Fig. 8-10 are equivalent, and Y_4 is an output node. Note that $Y_4 = Y_3$.

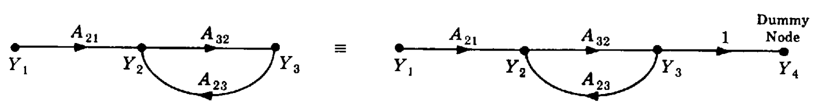


Fig. 8-10

8.5 CONSTRUCTION OF SIGNAL FLOW GRAPHS

The signal flow graph of a linear feedback control system whose components are specified by noninteracting transfer functions can be constructed by direct reference to the block diagram of the system. Each variable of the block diagram becomes a node and each block becomes a branch.

EXAMPLE 8.5. The block diagram of the canonical feedback control system is given in Fig. 8-11.

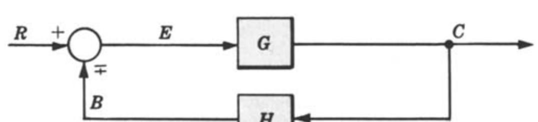


Fig. 8-11

The signal flow graph is easily constructed from Fig. 8-12. Note that the − or + sign of the summing point is associated with H.

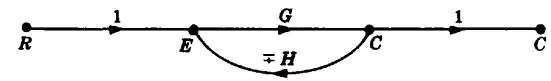


Fig. 8-12

The signal flow graph of a system described by a set of simultaneous equations can be constructed in the following general manner.

1. Write the system equations in the form

$$X_1 = A_{11}X_1 + A_{12}X_2 + \cdots + A_{1n}X_n$$

$$X_2 = A_{21}X_1 + A_{22}X_2 + \cdots + A_{2n}X_n$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$X_m = A_{m1}X_1 + A_{m2}X_2 + \cdots + A_{mn}X_n$$

An equation for X_1 is not required if X_1 is an input node.

2. Arrange the m or n (whichever is larger) nodes from left to right. The nodes may be rearranged if the required loops later appear too cumbersome.
3. Connect the nodes by the appropriate branches A_{11}, A_{12}, etc.
4. If the desired output node has outgoing branches, add a dummy node and a unity gain branch.
5. Rearrange the nodes and/or loops in the graph to achieve maximum pictorial clarity.

EXAMPLE 8.6. Let us construct a signal flow graph for the simple resistance network given in Fig. 8-13. There are five variables, v_1, v_2, v_3, i_1, and i_2. v_1 is known. We can write four independent equations from Kirchhoff's voltage and current laws. Proceeding from left to right in the schematic, we have

$$i_1 = \left(\frac{1}{R_1}\right)v_1 - \left(\frac{1}{R_1}\right)v_2 \qquad v_2 = R_3 i_1 - R_3 i_2 \qquad i_2 = \left(\frac{1}{R_2}\right)v_2 - \left(\frac{1}{R_2}\right)v_3 \qquad v_3 = R_4 i_2$$

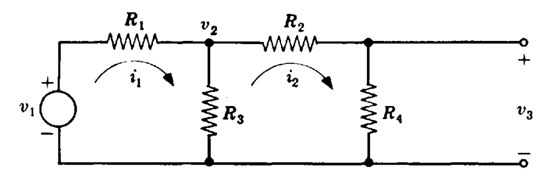


Fig. 8-13

Laying out the five nodes in the same order with v_1 as an input node, and connecting the nodes with the appropriate branches, we get Fig. 8-14. If we wish to consider v_3 as an output node, we must add a unity gain

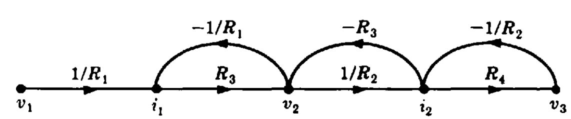


Fig. 8-14

branch and another node, yielding Fig. 8-15. No rearrangement of the nodes is necessary. We have one forward path and three feedback loops clearly in evidence.

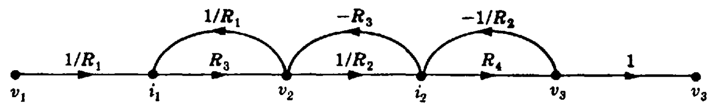


Fig. 8-15

Note that signal flow graph representations of equations are not unique. For example, the addition of a unity gain branch followed by a dummy node changes the graph, but not the equations it represents.

8.6 THE GENERAL INPUT-OUTPUT GAIN FORMULA

We found in Chapter 7 that we can reduce complicated block diagrams to canonical form, from which the control ratio is easily written as

$$\frac{C}{R} = \frac{G}{1 \pm GH}$$

It is possible to simplify signal flow graphs in a manner similar to that of block diagram reduction. But it is also possible, and much less time-consuming, to write down the input-output relationship *by inspection* from the original signal flow graph. This can be accomplished using the formula presented below. This formula can also be applied directly to block diagrams, but the signal flow graph representation is easier to read—especially when the block diagram is very complicated.

Let us denote the ratio of the input variable to the output variable by T. For linear feedback control systems, $T = C/R$. For the general signal flow graph presented in preceding paragraphs $T = X_n/X_1$, where X_n is the output and X_1 is the input.

The general formula for any signal flow graph is

$$T = \frac{\sum_i P_i \Delta_i}{\Delta} \tag{8.2}$$

where P_i = the ith forward path gain

P_{jk} = jth possible product of k nontouching loop gains

$$\Delta = 1 - (-1)^{k+1} \sum_k \sum_j P_{jk}$$

$$= 1 - \sum_j P_{j1} + \sum_j P_{j2} - \sum_j P_{j3} + \cdots$$

= 1 − (sum of all loop gains) + (sum of all gain products of two nontouching loops) − (sum of all gain products of three nontouching loops) + ⋯

$\Delta_i = \Delta$ evaluated with all loops touching P_i eliminated

Two loops, paths, or a loop and a path are said to be **nontouching** if they have no nodes in common.

Δ is called the **signal flow graph determinant** or **characteristic function**, since $\Delta = 0$ is the system characteristic equation.

The application of Equation (*8.2*) is considerably more straightforward than it appears. The following examples illustrate this point.

EXAMPLE 8.7. Let us first apply Equation (*8.2*) to the signal flow graph of the canonical feedback system (Fig. 8-16).

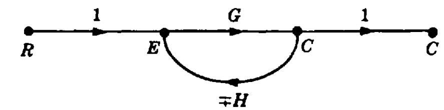


Fig. 8-16

There is only one forward path; hence

$$P_1 = G$$
$$P_2 = P_3 = \cdots = 0$$

There is only one (feedback) loop. Hence

$$P_{11} = \mp GH$$
$$P_{jk} = 0 \qquad j \neq 1 \qquad k \neq 1$$

Then

$$\Delta = 1 - P_{11} = 1 \pm GH \qquad \text{and} \qquad \Delta_1 = 1 - 0 = 1$$

Finally,

$$T = \frac{C}{R} = \frac{P_1 \Delta_1}{\Delta} = \frac{G}{1 \pm GH}$$

EXAMPLE 8.8. The signal flow graph of the resistance network of Example 8.6 is shown in Fig. 8-17. Let us apply Equation (*8.2*) to this graph and determine the voltage gain $T = v_3/v_1$ of the resistance network.

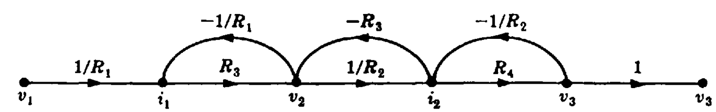


Fig. 8-17

There is one forward path (Fig. 8-18). Hence the forward path gain is

$$P_1 = \frac{R_3 R_4}{R_1 R_2}$$

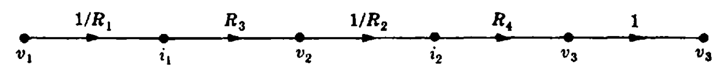


Fig. 8-18

There are three feedback loops (Fig. 8-19). Hence the loop gains are

$$P_{11} = -\frac{R_3}{R_1} \qquad P_{21} = -\frac{R_3}{R_2} \qquad P_{31} = -\frac{R_4}{R_2}$$

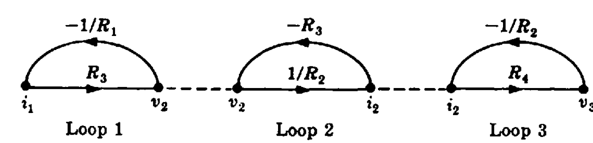


Fig. 8-19

There are two nontouching loops, loops 1 and 3. Hence

$$P_{12} = \text{gain product of the only two nontouching loops} = P_{11} \cdot P_{31} = \frac{R_3 R_4}{R_1 R_2}$$

There are no three loops that do not touch. Therefore

$$\Delta = 1 - (P_{11} + P_{21} + P_{31}) + P_{12} = 1 + \frac{R_3}{R_1} + \frac{R_3}{R_2} + \frac{R_4}{R_2} + \frac{R_3 R_4}{R_1 R_2}$$
$$= \frac{R_1 R_2 + R_1 R_3 + R_1 R_4 + R_2 R_3 + R_3 R_4}{R_1 R_2}$$

Since all loops touch the forward path, $\Delta_1 = 1$. Finally,

$$\frac{v_3}{v_1} = \frac{P_1 \Delta_1}{\Delta} = \frac{R_3 R_4}{R_1 R_2 + R_1 R_3 + R_1 R_4 + R_2 R_3 + R_3 R_4}$$

8.7 TRANSFER FUNCTION COMPUTATION OF CASCADED COMPONENTS

Loading effects of interacting components require little special attention using signal flow graphs. Simply combine the graphs of the components at their normal joining points (output node of one to the input node of another), account for loading by adding new loops at the joined nodes, and compute the overall gain using Equation (*8.2*). This procedure is best illustrated by example.

EXAMPLE 8.9. Assume that two identical resistance networks are to be cascaded and used as the control elements in the forward loop of a control system. The networks are simple voltage dividers of the form given in Fig. 8-20.

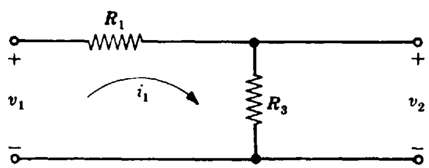


Fig. 8-20

Two independent equations for this network are

$$i_1 = \left(\frac{1}{R_1}\right)v_1 - \left(\frac{1}{R_1}\right)v_2 \qquad \text{and} \qquad v_2 = R_3 i_1$$

The signal flow graph is easily drawn (Fig. 8-21). The gain of this network is, by inspection, equal to

$$\frac{v_2}{v_1} = \frac{R_3}{R_1 + R_3}$$

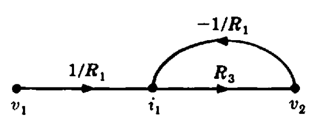


Fig. 8-21

If we were to ignore loading, the overall gain of two cascaded networks would simply be determined by multiplying the individual gains:

$$\left(\frac{v_2}{v_1}\right)^2 = \frac{R_3^2}{R_1^2 + R_3^2 + 2R_1 R_3}$$

This answer is incorrect. We prove this in the following manner. When the two identical networks are cascaded, we note that the result is equivalent to the network of Example 8.6, with $R_2 = R_1$ and $R_4 = R_3$ (Fig. 8-22).

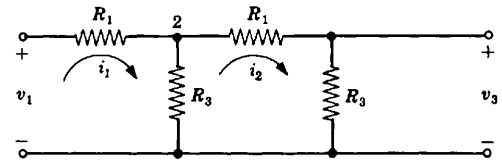


Fig. 8-22

The signal flow graph of this network was also determined in Example 8.6 (Fig. 8-23).

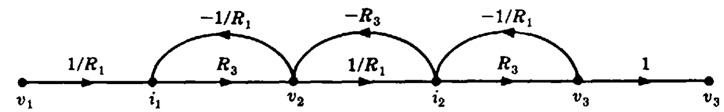


Fig. 8-23

We observe that the feedback branch $-R_3$ in Fig. 8-23 does not appear in the signal flow graph of the cascaded signal flow graphs of the individual networks connected from node v_2 to v_1' (Fig. 8-24). This means that, as a result of connecting the two networks, the second one loads the first, changing the equation for v_2 from

$$v_2 = R_3 i_1 \qquad \text{to} \qquad v_2 = R_3 i_1 - R_3 i_2$$

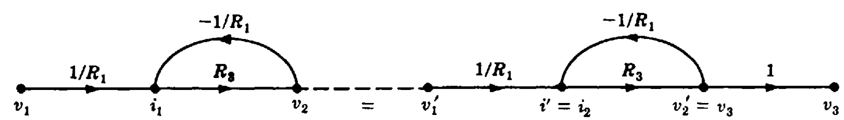


Fig. 8-24

This result could also have been obtained by directly writing the equations for the combined networks. In this case, only the equation for v_2 would have changed form.

The gain of the combined networks was determined in Example 8.8 as

$$\frac{v_3}{v_1} = \frac{R_3^2}{R_1^2 + R_3^2 + 3R_1 R_3}$$

when R_2 is set equal to R_1 and R_4 is set equal to R_3. We observe that

$$\left(\frac{v_2}{v_1}\right)^2 = \frac{R_3^2}{R_1^2 + R_3^2 + 2R_1 R_3} \neq \frac{v_3}{v_1}$$

It is good general practice to calculate the gain of cascaded networks directly from the *combined* signal flow graph. Most practical control system components load each other when connected in series.

8.8 BLOCK DIAGRAM REDUCTION USING SIGNAL FLOW GRAPHS AND THE GENERAL INPUT-OUTPUT GAIN FORMULA

Often, the easiest way to determine the control ratio of a complicated block diagram is to translate the block diagram into a signal flow graph and apply Equation (*8.2*). Takeoff points and summing points must be separated by a unity gain branch in the signal flow graph when using Equation (*8.2*).

If the elements G and H of a canonical feedback representation are desired, Equation (*8.2*) also provides this information. The direct transfer function is

$$G = \sum_i P_i \Delta_i \tag{8.3}$$

The loop transfer function is

$$GH = \Delta - 1 \tag{8.4}$$

Equations (*8.3*) and (*8.4*) are solved simultaneously for G and H, and the canonical feedback control system is drawn from the result.

EXAMPLE 8.10. Let us determine the control ratio C/R and the canonical block diagram of the feedback control system of Example 7.9 (Fig. 8-25).

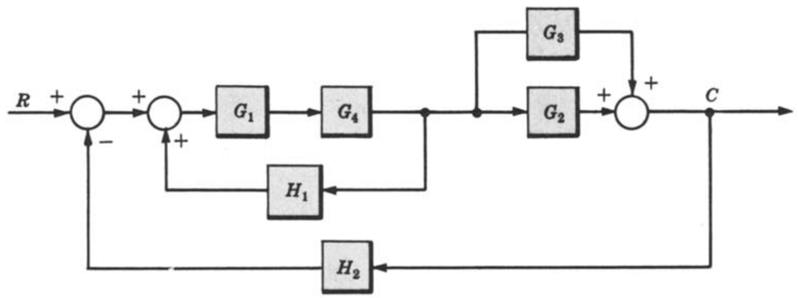


Fig. 8-25

The signal flow graph is given in Fig. 8-26. There are two forward paths:

$$P_1 = G_1G_2G_4 \qquad P_2 = G_1G_3G_4$$

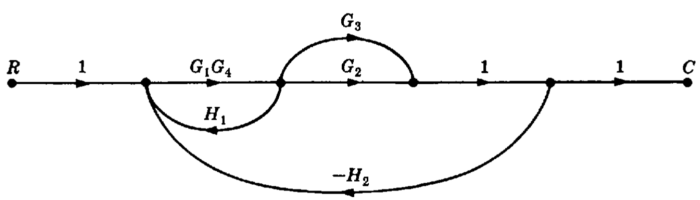


Fig. 8-26

There are three feedback loops:

$$P_{11} = G_1G_4H_1 \qquad P_{21} = -G_1G_2G_4H_2 \qquad P_{31} = -G_1G_3G_4H_2$$

There are no nontouching loops, and all loops touch both forward paths; then

$$\Delta_1 = 1 \qquad \Delta_2 = 1$$

Therefore the control ratio is

$$T = \frac{C}{R} = \frac{P_1\Delta_1 + P_2\Delta_2}{\Delta} = \frac{G_1G_2G_4 + G_1G_3G_4}{1 - G_1G_4H_1 + G_1G_2G_4H_2 + G_1G_3G_4H_2}$$

$$= \frac{G_1G_4(G_2 + G_3)}{1 - G_1G_4H_1 + G_1G_2G_4H_2 + G_1G_3G_4H_2}$$

From Equations (*8.3*) and (*8.4*), we have

$$G = G_1G_4(G_2 + G_3) \qquad \text{and} \qquad GH = G_1G_4(G_3H_2 + G_2H_2 - H_1)$$

Therefore

$$H = \frac{GH}{G} = \frac{(G_2 + G_3)H_2 - H_1}{G_2 + G_3}$$

The canonical block diagram is therefore given in Fig. 8-27.

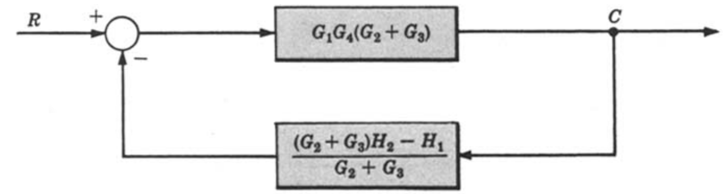


Fig. 8-27

The negative summing point sign for the feedback loop is a result of using a positive sign in the GH formula above. If this is not obvious, refer to Equation (*7.3*) and its explanation in Section 7.4.

The block diagram above may be put into the final form of Examples 7.9 or 7.10 by using the transformation theorems of Section 7.5.

Solved Problems

SIGNAL FLOW GRAPH ALGEBRA AND DEFINITIONS

8.1. Simplify the signal flow graphs given in Fig. 8-28.

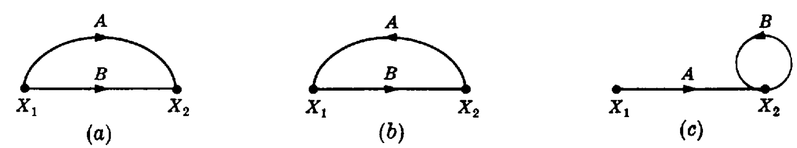
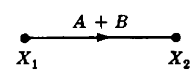
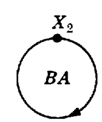


Fig. 8-28

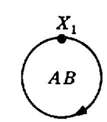
(*a*) Clearly, $X_2 = AX_1 + BX_1 = (A + B)X_1$. Therefore we have



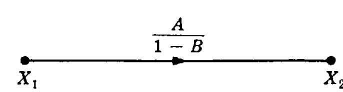
(*b*) We have $X_2 = BX_1$ and $X_1 = AX_2$. Hence $X_2 = BAX_2$, or $X_1 = ABX_1$, yielding



or



(*c*) If A and B are multiplicative operators (e.g., constants or transfer functions), we have $X_2 = AX_1 + BX_2 = (A/(1 - B))X_1$. Hence the signal flow graph becomes



8.2. Draw signal flow graphs for the block diagrams in Problem 7.3 and reduce them by the multiplication rule (Fig. 8-29).

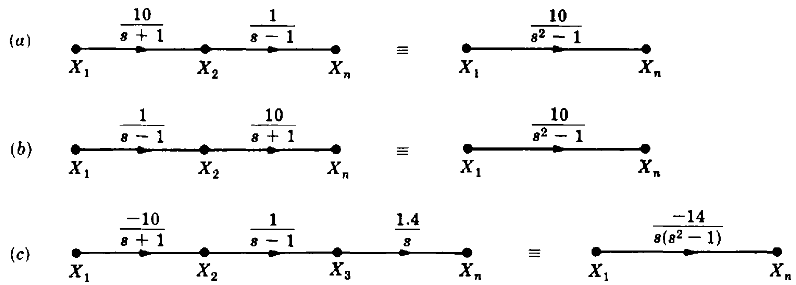


Fig. 8-29

8.3. Consider the signal flow graph in Fig. 8-30.

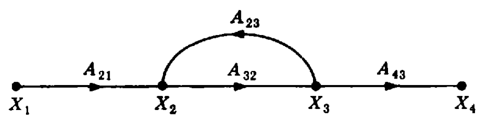


Fig. 8-30

(*a*) Draw the signal flow graph for the system equivalent to that graphed in Fig. 8-30, but in which X_3 becomes kX_3 (k constant) and X_1, X_2, and X_4 remain the same.

(*b*) Repeat part (*a*) for the case in which X_2 and X_3 become k_2X_2 and k_3X_3, and X_1 and X_4 remain the same (k_2 and k_3 are constants).

This problem illustrates the fundamentals of a technique that can be used for *scaling* variables.

(*a*) For the system to remain the same when a node variable is multiplied by a constant, all signals entering the node must be multiplied by the same constant, and all signals leaving the node divided by that constant. Since X_1, X_2, and X_4 must remain the same, the *branches* are modified (Fig. 8-31).

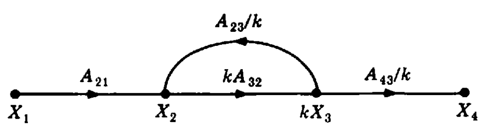


Fig. 8-31

(*b*) Substitute k_2X_2 for X_2, and k_3X_3 for X_3 (Fig. 8-32)

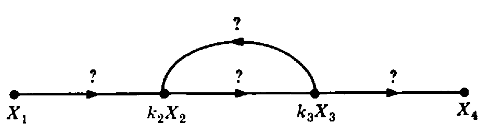


Fig. 8-32

It is clear from the graph that A_{21} becomes k_2A_{21}, A_{32} becomes $(k_3/k_2)A_{32}$, A_{23} becomes $(k_2/k_3)A_{23}$, and A_{43} becomes $(1/k_3)A_{43}$ (Fig. 8-33).

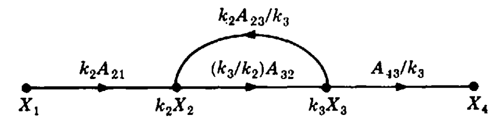


Fig. 8-33

8.4. Consider the signal flow graph given in Fig. 8-34.

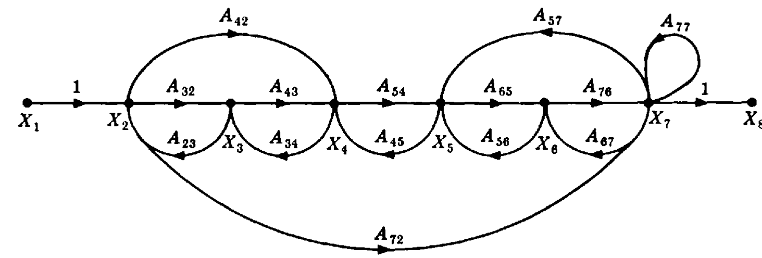


Fig. 8-34

Identify the (*a*) input node, (*b*) output node, (*c*) forward paths, (*d*) feedback paths, (*e*) self-loop. Determine the (*f*) loop gains of the feedback loops, (*g*) path gains of the forward paths.

(*a*) X_1

(*b*) X_8

(*c*) X_1 to X_2 to X_3 to X_4 to X_5 to X_6 to X_7 to X_8
X_1 to X_2 to X_7 to X_8
X_1 to X_2 to X_4 to X_5 to X_6 to X_7 to X_8

(*d*) X_2 to X_3 to X_2; X_3 to X_4 to X_3; X_4 to X_5 to X_4; X_2 to X_4 to X_3 to X_2;
X_2 to X_7 to X_5 to X_4 to X_3 to X_2; X_5 to X_6 to X_5; X_6 to X_7 to X_6;
X_5 to X_6 to X_7 to X_5; X_7 to X_7; X_2 to X_7 to X_6 to X_5 to X_4 to X_3 to X_2

(*e*) X_7 to X_7

(*f*) $A_{32}A_{23}$; $A_{43}A_{34}$; $A_{54}A_{45}$; $A_{65}A_{56}$; $A_{76}A_{67}$; $A_{65}A_{76}A_{57}$; A_{77}; $A_{42}A_{34}A_{23}$;
$A_{72}A_{57}A_{45}A_{34}A_{23}$; $A_{72}A_{67}A_{56}A_{45}A_{34}A_{23}$

(*g*) $A_{32}A_{43}A_{54}A_{65}A_{76}$; A_{72}; $A_{42}A_{54}A_{65}A_{76}$

SIGNAL FLOW GRAPH CONSTRUCTION

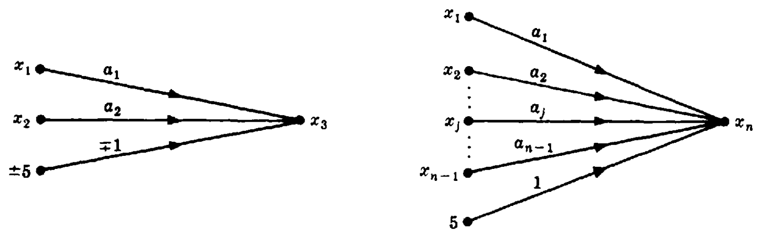
8.5. Consider the following equations in which $x_1, x_2, \ldots, x_n$ are variables and $a_1, a_2, \ldots, a_n$ are coefficients or mathematical operators:

(*a*) $x_3 = a_1x_1 + a_2x_2 \mp 5$ (*b*) $x_n = \sum_{k=1}^{n-1} a_k x_k + 5$

What are the minimum number of nodes and the minimum number of branches required to construct the signal flow graphs of these equations? Draw the graphs.

(*a*) There are four variables in this equation: x_1 x_2, x_3, and ± 5. Therefore a minimum of four nodes are required. There are three coefficients or transmission functions on the right-hand side of the equation:

a_1, a_2, and ∓ 1. Hence a minimum of three branches are required. A minimal signal flow graph is shown in Fig. 8-35(a).



(a) (b)

Fig. 8-35

(b) There are $n+1$ variables: $x_1, x_2, \ldots, x_n$, and 5; and there are n coefficients: $a_1, a_2, \ldots, a_{n-1}$, and 1. Therefore a minimal signal flow graph is shown in Fig. 8-35(b).

8.6. Draw signal flow graphs for

$$(a)\quad x_2 = a_1\left(\frac{dx_1}{dt}\right) \qquad (b)\quad x_3 = \frac{d^2x_2}{dt^2} + \frac{dx_1}{dt} - x_1 \qquad (c)\quad x_4 = \int x_3\,dt$$

(a) The operations called for in this equation are a_1 and d/dt. Let the equation be written as $x_2 = a_1 \cdot (d/dt)(x_1)$. Since there are two operations, we may define a new variable dx_1/dt and use it as an intermediate node. The signal flow graph is given in Fig. 8-36.

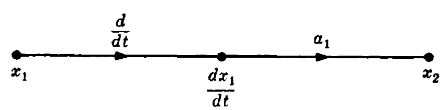


Fig. 8-36

(b) Similarly, $x_3 = (d^2/dt^2)(x_2) + (d/dt)(x_1) - x_1$. Therefore we obtain Fig. 8-37

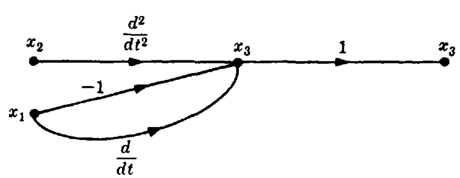


Fig. 8-37

(c) The operation is integration. Let the operator be denoted by $\int dt$. The signal flow graph is given in Fig. 8-38.

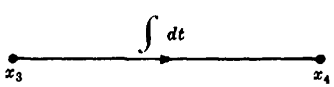


Fig. 8-38

8.7. Construct the signal flow graph for the following set of simultaneous equations:

$$x_2 = A_{21}x_1 + A_{23}x_3 \qquad x_3 = A_{31}x_1 + A_{32}x_2 + A_{33}x_3 \qquad x_4 = A_{42}x_2 + A_{43}x_3$$

There are four variables: $x_1, \ldots, x_4$. Hence four nodes are required. Arranging them from left to right and connecting them with the appropriate branches, we obtain Fig. 8-39.

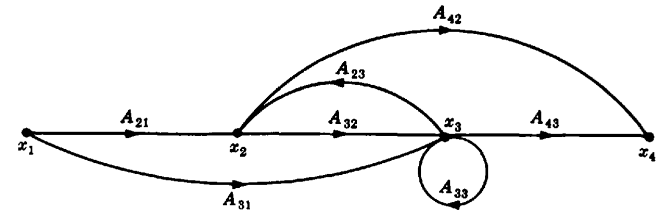


Fig. 8-39

A neater way to arrange this graph is shown in Fig. 8-40.

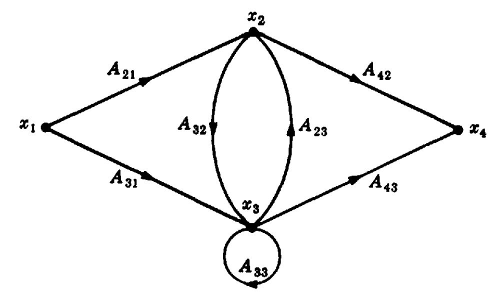


Fig. 8-40

8.8. Draw a signal flow graph for the resistance network shown in Fig. 8-41 in which $v_2(0) = v_3(0) = 0$. v_2 is the voltage across C_1.

Mathcad

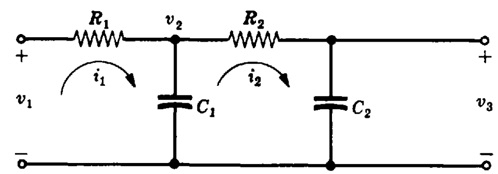


Fig. 8-41

The five variables are v_1, v_2, v_3, i_1, and i_2; and v_1 is the input. The four independent equations derived from Kirchhoff's voltage and current laws are

$$i_1 = \left(\frac{1}{R_1}\right)v_1 - \left(\frac{1}{R_1}\right)v_2 \qquad v_2 = \frac{1}{C_1}\int_0^t i_1\,dt - \frac{1}{C_1}\int_0^t i_2\,dt$$

$$i_2 = \left(\frac{1}{R_2}\right)v_2 - \left(\frac{1}{R_2}\right)v_3 \qquad v_3 = \frac{1}{C_2}\int_0^t i_2\,dt$$

The signal flow graph can be drawn directly from these equations (Fig. 8-42).

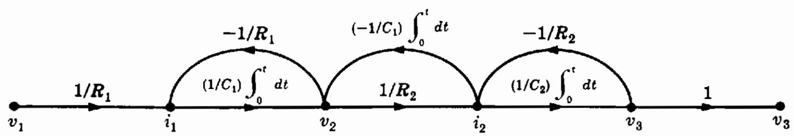


Fig. 8-42

In Laplace transform notation, the signal flow graph is given in Fig. 8-43.

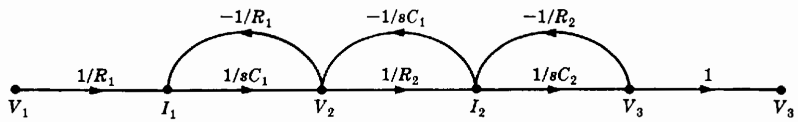


Fig. 8-43

THE GENERAL INPUT-OUTPUT GAIN FORMULA

8.9. The transformed equations for the mechanical system given in Fig. 8-44 are

$$\text{(i)} \qquad F + k_1 X_2 = \left(M_1 s^2 + f_1 s + k_1\right) X_1$$

$$\text{(ii)} \qquad k_1 X_1 = \left(M_2 s^2 + f_2 s + k_1 + k_2\right) X_2$$

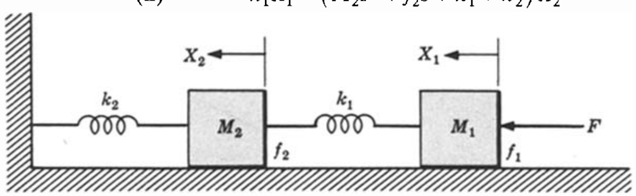


Fig. 8-44

where F is force, M is mass, k is spring constant, f is friction, and X is displacement. Determine X_2/F using Equation (*8.2*).

There are three variables: X_1, X_2, and F. Therefore we need three nodes. In order to draw the signal flow graph, divide Equation (i) by A and Equation (ii) by B, where $A \equiv M_1 s^2 + f_1 s + k_1$, and $B \equiv M_2 s^2 + f_2 s + k_1 + k_2$:

$$\text{(iii)} \qquad \left(\frac{1}{A}\right) F + \left(\frac{k_1}{A}\right) X_2 = X_1$$

$$\text{(iv)} \qquad \left(\frac{k_1}{B}\right) X_1 = X_2$$

Therefore the signal flow graph is given in Fig. 8-45.

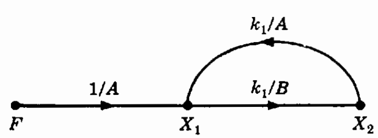


Fig. 8-45

The forward path gain is $P_1 = k_1/AB$. The feedback loop gain is $P_{11} = k_1^2/AB$. then $\Delta = 1 - P_{11} = (AB - k_1^2)/AB$ and $\Delta_1 = 1$. Finally,

$$\frac{X_2}{F} = \frac{P_1\Delta_1}{\Delta} = \frac{k_1}{AB - k_1^2} = \frac{k_1}{\left(M_1 s^2 + f_1 s + k_1\right)\left(M_2 s^2 + f_2 s + k_1 + k_2\right) - k_1^2}$$

8.10. Determine the transfer function for the block diagram in Problem 7.20 by signal flow graph techniques.

The signal flow graph, Fig. 8-46, is drawn directly from Fig. 7-44. There are two forward paths. The path gains are $P_1 = G_1G_2G_3$ and $P_2 = G_4$. The three feedback loop gains are $P_{11} = -G_2H_1$, $P_{21} = G_1G_2H_1$, and $P_{31} = -G_2G_3H_2$. No loops are nontouching. Hence $\Delta = 1 - (P_{11} + P_{21} + P_{31})$. Also, $\Delta_1 = 1$; and since no loops touch the nodes of P_2, $\Delta_2 = \Delta$. Thus

$$T = \frac{P_1\Delta_1 + P_2\Delta_2}{\Delta} = \frac{G_1G_2G_3 + G_4 + G_2G_4H_1 - G_1G_2G_4H_1 + G_2G_3G_4H_2}{1 + G_2H_1 - G_1G_2H_1 + G_2G_3H_2}$$

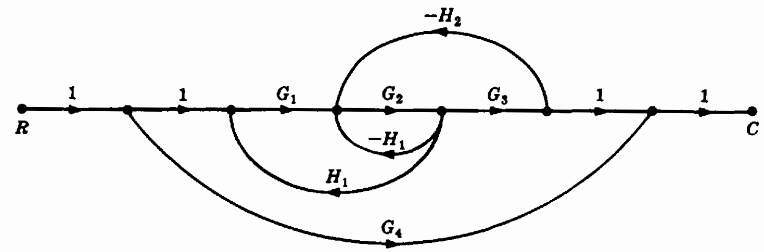


Fig. 8-46

8.11. Determine the transfer function V_3/V_1 from the signal flow graph of Problem 8.8.

Mathcad

The single forward path gain is $1/(s^2R_1R_2C_1C_2)$. The loop gains of the three feedback loops are $P_{11} = -1/(sR_1C_1)$, $P_{21} = -1/(sR_2C_1)$, and $P_{31} = -1/(sR_2C_2)$. The gain product of the only two nontouching loops is $P_{12} = P_{11} \cdot P_{31} = 1/(s^2R_1R_2C_1C_2)$. Hence

$$\Delta = 1 - (P_{11} + P_{21} + P_{31}) + P_{12} = \frac{s^2R_1R_2^2C_1^2C_2 + s\left(R_2^2C_1C_2 + R_1R_2C_1C_2 + R_1R_2C_1^2\right) + R_2C_1}{s^2R_1R_2^2C_1^2C_2}$$

Since all loops touch the forward path, $\Delta_1 = 1$. Finally,

$$\frac{V_3}{V_1} = \frac{P_1\Delta_1}{\Delta} = \frac{1}{s^2R_1R_2C_1C_2 + s(R_2C_2 + R_1C_2 + R_1C_1) + 1}$$

8.12. Solve Problem 7.16 with signal flow graph techniques.

The signal flow graph is drawn directly from Fig. 7-26, as shown in Fig. 8-47:

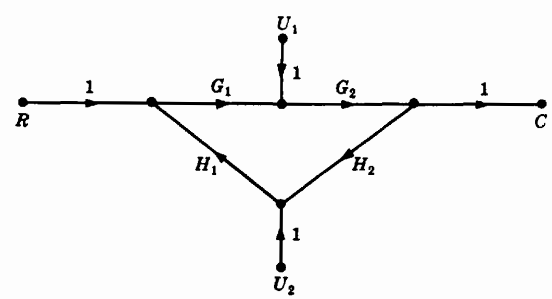


Fig. 8-47

With $U_1 = U_2 = 0$, we have Fig. 8-48. Then $P_1 = G_1G_2$ and $P_{11} = G_1G_2H_1H_2$. Hence $\Delta = 1 - P_{11} = 1 - G_1G_2H_1H_2$, $\Delta_1 = 1$, and

$$C_R = TR = \frac{P_1\Delta_1 R}{\Delta} = \frac{G_1G_2R}{1 - G_1G_2H_1H_2}$$

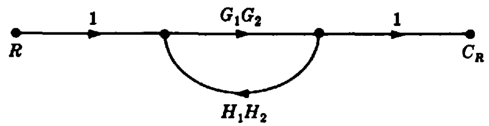


Fig. 8-48

Now put $U_2 = R = 0$ (Fig. 8-49).

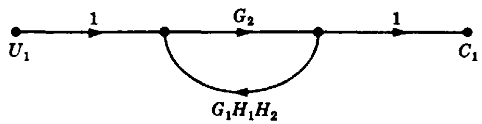


Fig. 8-49

Then $P_1 = G_2$, $P_{11} = G_1G_2H_1H_2$, $\Delta = 1 - G_1G_2H_1H_2$, $\Delta_1 = 1$, and

$$C_1 = TU_1 = \frac{G_2U_1}{1 - G_1G_2H_1H_2}$$

Now put $R = U_1 = 0$ (Fig. 8-50).

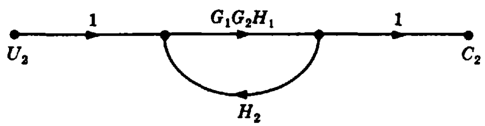


Fig. 8-50

Then $P_1 = G_1G_2H_1$, $P_{11} = G_1G_2H_1H_2$, $\Delta = 1 - G_1G_2H_1H_2$, $\Delta_1 = 1$, and

$$C_2 = TU_2 = \frac{P_1\Delta_1U_2}{\Delta} = \frac{G_1G_2H_1U_2}{1 - G_1G_2H_1H_2}$$

Finally, we have

$$C = C_R + C_1 + C_2 = \frac{G_1G_2R + G_2U_1 + G_1G_2H_1U_2}{1 - G_1G_2H_1H_2}$$

TRANSFER FUNCTION COMPUTATION OF CASCADED COMPONENTS

8.13. Determine the transfer function for two of the networks in cascade shown in Fig. 8-51.

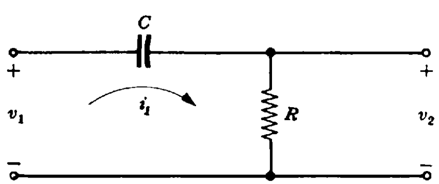


Fig. 8-51

In Laplace transform notation the network becomes Fig. 8-52.

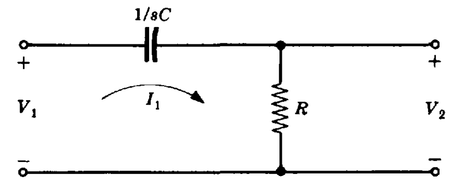


Fig. 8-52

By Kirchhoff's laws, we have $I_1 = sCV_1 - sCV_2$ and $V_2 = RI_1$. The signal flow graph is given in Fig. 8.53.

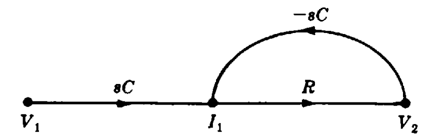


Fig. 8-53

For two networks in cascade (Fig. 8-54) the V_2 equation is also dependent on I_2: $V_2 = RI_1 - RI_2$. Hence two networks are joined at node 2 (Fig. 8-55) and a feedback loop $(-RI_2)$ is added between I_2 and V_2 (Fig. 8-56).

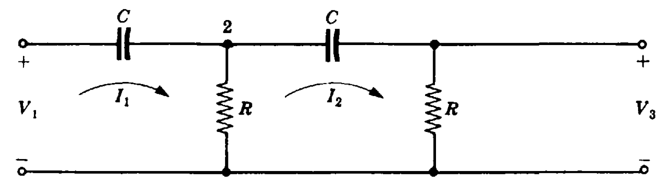


Fig. 8-54

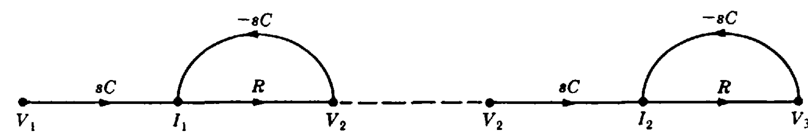


Fig. 8-55

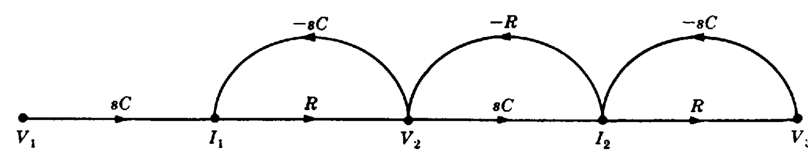


Fig. 8-56

Then $P_1 = s^2R^2C^2$, $P_{11} = P_{31} = -sRC$, $P_{12} = P_{11} \cdot P_{31} = s^2R^2C^2$, $\Delta = 1 - (P_{11} + P_{21} + P_{31}) + P_{12} = 1 + 3sRC + s^2R^2C^2$, $\Delta_1 = 1$, and

$$T = \frac{P_1\Delta_1}{\Delta} = \frac{s^2}{s^2 + (3/RC)s + 1/(RC)^2}$$

8.14. Two resistance networks in the form of that in Example 8.6 are to be used for control elements in the forward path of a control system. They are to be cascaded and shall have identical respective component values as shown in Fig. 8-57. Find v_5/v_1 using Equation (*8.2*).

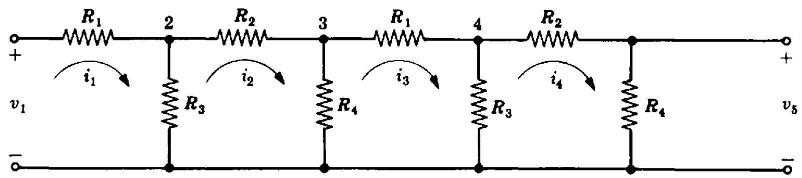


Fig. 8-57

There are nine variables: v_1, v_2, v_3, v_4, v_5, i_1, i_2, i_3, and i_4. Eight independent equations are

$$i_1=\left(\frac{1}{R_1}\right)v_1-\left(\frac{1}{R_1}\right)v_2 \qquad i_3=\left(\frac{1}{R_1}\right)v_3-\left(\frac{1}{R_1}\right)v_4$$

$$v_2=R_3i_1-R_3i_2 \qquad v_4=R_3i_3-R_3i_4$$

$$i_2=\left(\frac{1}{R_2}\right)v_2-\left(\frac{1}{R_2}\right)v_3 \qquad i_4=\left(\frac{1}{R_2}\right)v_4-\left(\frac{1}{R_2}\right)v_5$$

$$v_3=R_4i_2-R_4i_3 \qquad v_5=R_4i_4$$

Only the equation for v_3 is different from those of the single network of Example 8.6; it has an extra term, $(-R_4i_3)$. Therefore the signal flow diagram for each network alone (Example 8.6) may be joined at node v_3, and an extra branch of gain $-R_4$ drawn from i_3 to v_3. The resulting signal flow graph for the double network is given in Fig. 8-58.

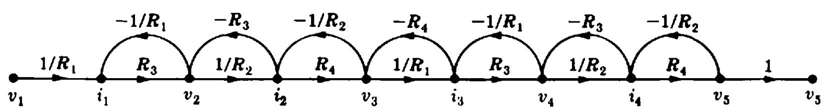


Fig. 8-58

The voltage gain $T=v_5/v_1$ is calculated from Equation (*8.2*) as follows. One forward path yields $P_1=(R_3R_4/R_1R_2)^2$. The gains of the seven feedback loops are $P_{11}=-R_3/R_1=P_{51}$, $P_{21}=-R_3/R_2=P_{61}$, $P_{31}=-R_4/R_2=P_{71}$, and $P_{41}=-R_4/R_1$.

There are 15 gain products of two nontouching loops. From left to right, we have

$$P_{12}=\frac{R_3R_4}{R_1R_2} \qquad P_{42}=\frac{R_3^2}{R_1R_2} \qquad P_{72}=\frac{R_3^2}{R_1R_2} \qquad P_{10,2}=\frac{R_3R_4}{R_1R_2} \qquad P_{13,2}=\frac{R_3R_4}{R_1R_2}$$

$$P_{22}=\frac{R_3R_4}{R_1^2} \qquad P_{52}=\frac{R_3R_4}{R_1R_2} \qquad P_{82}=\left(\frac{R_3}{R_2}\right)^2 \qquad P_{11,2}=\frac{R_3R_4}{R_2^2} \qquad P_{14,2}=\frac{R_4^2}{R_1R_2}$$

$$P_{32}=\left(\frac{R_3}{R_1}\right)^2 \qquad P_{62}=\frac{R_3R_4}{R_1R_2} \qquad P_{92}=\frac{R_3R_4}{R_2^2} \qquad P_{12,2}=\left(\frac{R_4}{R_2}\right)^2 \qquad P_{15,2}=\frac{R_3R_4}{R_1R_2}$$

There are 10 gain products of three nontouching loops. From left to right, we have

$$P_{13}=\frac{R_3^2R_4}{R_1^2R_2} \qquad P_{33}=-\frac{R_3R_4^2}{R_1R_2^2} \qquad P_{63}=-\frac{R_3^2R_4}{R_1^2R_2} \qquad P_{83}=-\frac{R_3R_4^2}{R_1R_2^2} \qquad P_{53}=-\frac{R_3R_4^2}{R_1^2R_2}$$

$$P_{23}=-\frac{R_3^2R_4}{R_1R_2^2} \qquad P_{43}=-\frac{R_3^2R_4}{R_1^2R_2} \qquad P_{73}=-\frac{R_3^2R_4}{R_1R_2^2} \qquad P_{93}=-\frac{R_3^2R_4}{R_1R_2^2} \qquad P_{10,3}=-\frac{R_3R_4^2}{R_1R_2^2}$$

There is one gain product of four nontouching loops: $P_{14}=P_{11}P_{31}P_{51}P_{71}=(R_3R_4/R_1R_2)^2$. Therefore the determinant is

$$\Delta=1-\sum_{j=1}^{7}P_{j1}+\sum_{j=1}^{15}P_{j2}-\sum_{j=1}^{10}P_{j3}+P_{14}$$

$$=1+\frac{R_1R_3+R_1R_4+R_2R_3+R_2R_4+6R_3R_4+2R_3^2+R_4^2}{R_1R_2}+\frac{R_3R_4+R_3^2}{R_1^2}+\frac{R_3^2+R_4^2+R_3R_4}{R_2^2}$$

Since all loops touch the forward path, $\Delta_1=1$ and

$$T=\frac{P_1\Delta_1}{\Delta}=\frac{(R_3R_4)^2}{(R_1R_2)^2+R_1^2(R_2R_3+R_2R_4+R_3R_4+R_3^2+R_4^2)+R_2^2(R_3^2+R_1R_3+R_1R_4+R_3R_4)+2R_1R_2R_3^2+R_1R_2R_4^2+6R_1R_2R_3R_4}$$

BLOCK DIAGRAM REDUCTION

8.15. Determine C/R for each system shown in Fig. 8-59 using Equation (*8.2*).

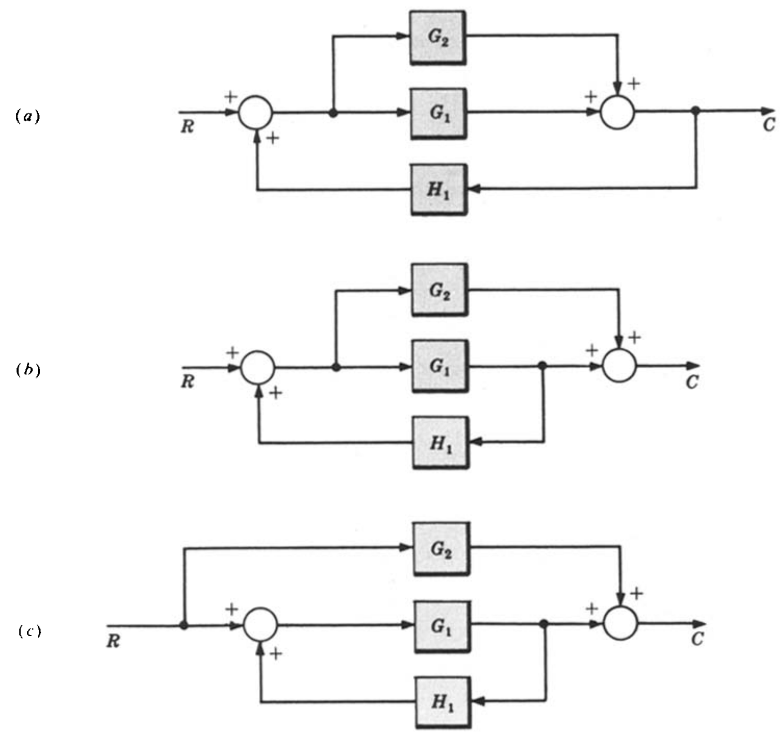


Fig. 8-59

(*a*) The signal flow graph is given in Fig. 8-60. The two forward path gains are $P_1=G_1$, $P_2=G_2$. The two feedback loop gains are $P_{11}=G_1H_1$, $P_{21}=G_2H_1$. Then

$$\Delta=1-(P_{11}+P_{21})=1-G_1H_1-G_2H_1$$

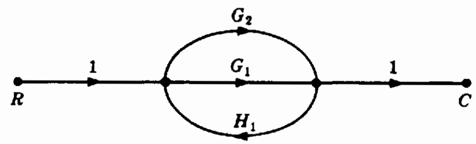


Fig. 8-60

Now, $\Delta_1 = 1$ and $\Delta_2 = 1$ because both paths touch the feedback loops at both interior nodes. Hence

$$\frac{C}{R} = \frac{P_1\Delta_1 + P_2\Delta_2}{\Delta} = \frac{G_1 + G_2}{1 - G_1H_1 - G_2H_1}$$

(*b*) The signal flow graph is given in Fig. 8-61. Again, we have $P_1 = G_1$ and $P_2 = G_2$. But now there is only one feedback loop, and $P_{11} = G_1H_1$; then $\Delta = 1 - G_1H_1$. The forward path through G_1 clearly touches the feedback loop at nodes a and b; thus $\Delta_1 = 1$. The forward path through G_2 touches the feedback loop at node a; then $\Delta_2 = 1$. Hence

$$\frac{C}{R} = \frac{P_1\Delta_1 + P_2\Delta_2}{\Delta} = \frac{G_1 + G_2}{1 - G_1H_1}$$

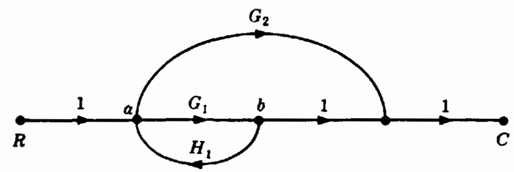


Fig. 8-61

(*c*) The signal flow graph is given in Fig. 8-62. Again, we have $P_1 = G_1$, $P_2 = G_2$, $P_{11} = G_1H_1$, $\Delta = 1 - G_1H_1$, and $\Delta_1 = 1$. But the feedback path *does not* touch the forward path through G_2 at *any* node. Therefore $\Delta_2 = \Delta = 1 - G_1H_1$ and

$$\frac{C}{R} = \frac{P_1\Delta_1 + P_2\Delta_2}{\Delta} = \frac{G_1 + G_2(1 - G_1H_1)}{1 - G_1H_1}$$

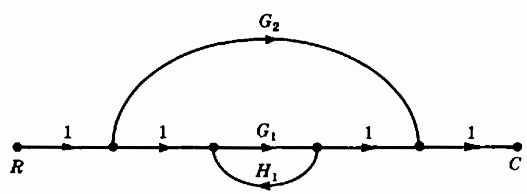


Fig. 8-62

This problem illustrates the importance of separating summing points and takeoff points with a branch of unity gain when applying Equation (*8.2*).

8.16. Find the transfer function C/R for the system shown in Fig. 8-63 in which K is a constant.

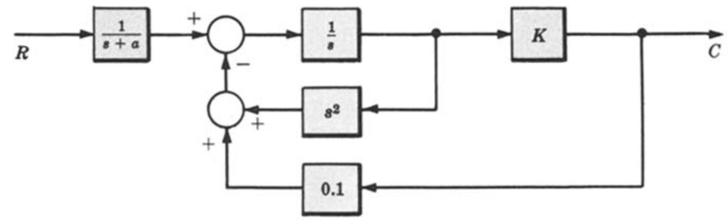


Fig. 8-63

The signal flow graph is given in Fig. 8-64. The only forward path gain is

$$P_1 = \left(\frac{1}{s+a}\right)\cdot\left(\frac{1}{s}\right)K = \frac{K}{s(s+a)}$$

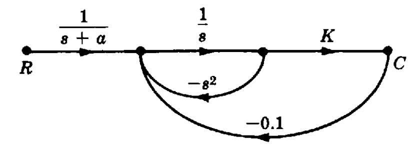


Fig. 8-64

The two feedback loop gains are $P_{11} = (1/s)\cdot(-s^2) = -s$ and $P_{21} = -0.1K/s$. There are no nontouching loops. Hence

$$\Delta = 1 - (P_{11} + P_{21}) = \frac{s^2 + s - 0.1K}{s} \qquad \Delta_1 = 1 \qquad \frac{C}{R} = \frac{P_1\Delta_1}{\Delta} = \frac{K}{(s+a)(s^2 + s + 0.1K)}$$

8.17. Solve Problem 7.18 using signal flow graph techniques.

The signal flow graph is given in Fig. 8-65.

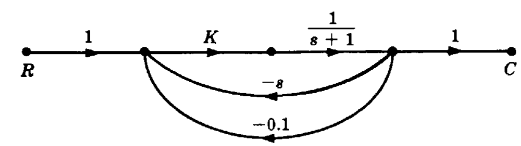


Fig. 8-65

Applying the multiplication and addition rules, we obtain Fig. 8-66. Now

$$P_1 = \frac{K}{s+1} \qquad P_{11} = -\frac{K(s+0.1)}{s+1} \qquad \Delta = 1 + \frac{K(s+0.1)}{s+1} \qquad \Delta_1 = 1,$$

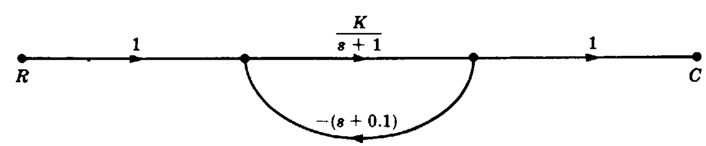


Fig. 8-66

and

$$C = TR = \frac{P_1\Delta_1 R}{\Delta} = \frac{KR}{(1+K)s + 1 + 0.1K}$$

8.18. Find C/R for the control system given in Fig. 8-67.

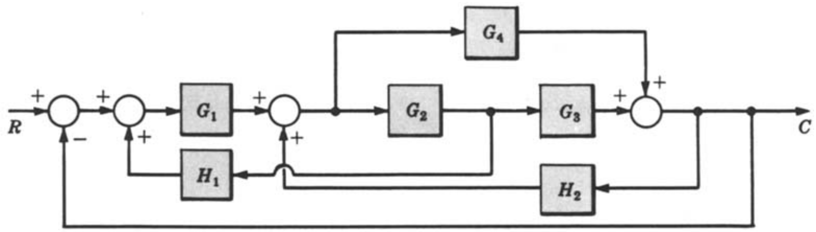


Fig. 8-67

The signal flow graph is given in Fig. 8.68. The two forward path gains are $P_1 = G_1G_2G_3$ and $P_2 = G_1G_4$. The five feedback loop gains are $P_{11} = G_1G_2H_1$, $P_{21} = G_2G_3H_2$, $P_{31} = -G_1G_2G_3$, $P_{41} = G_4H_2$, and $P_{51} = -G_1G_4$. Hence

$$\Delta = 1 - (P_{11} + P_{21} + P_{31} + P_{41} + P_{51}) = 1 + G_1G_2G_3 - G_1G_2H_1 - G_2G_3H_2 - G_4H_2 + G_1G_4$$

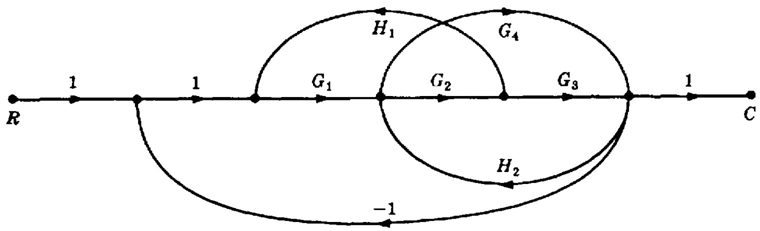


Fig. 8-68

and $\Delta_1 = \Delta_2 = 1$. Finally,

$$\frac{C}{R} = \frac{P_1\Delta_1 + P_2\Delta_2}{\Delta} = \frac{G_1G_2G_3 + G_1G_4}{1 + G_1G_2G_3 - G_1G_2H_1 - G_2G_3H_2 - G_4H_2 + G_1G_4}$$

8.19. Determine C/R for the system given in Fig. 8-69. *Then* put $G_3 = G_1G_2H_2$.

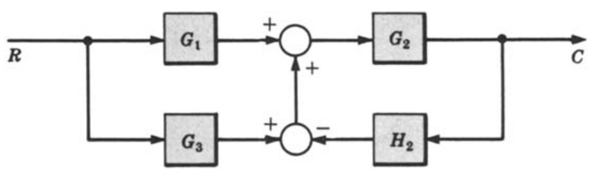


Fig. 8-69

The signal flow graph is given in Fig. 8-70. We have $P_1 = G_1G_2$, $P_2 = G_2G_3$, $P_{11} = -G_2H_2$, $\Delta = 1 + G_2H_2$, $\Delta_1 = \Delta_2 = 1$, and

$$\frac{C}{R} = \frac{P_1\Delta_1 + P_2\Delta_2}{\Delta} = \frac{G_2(G_1 + G_3)}{1 + G_2H_2}$$

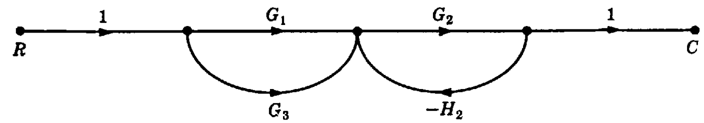


Fig. 8-70

Putting $G_3 = G_1G_2H_2$, we obtain $C/R = G_1G_2$ and the system transfer function becomes open-loop.

8.20. Determine the elements for a canonical feedback system for the system of Problem 8.10.

From Problem 8.10, $P_1 = G_1G_2G_3$, $P_2 = G_4$, $\Delta = 1 + G_2H_1 - G_1G_2H_1 + G_2G_3H_2$, $\Delta_1 = 1$, and $\Delta_2 = \Delta$. From Equation (*8.3*) we have

$$G = \sum_{i=1}^{2} P_i\Delta_i = G_1G_2G_s + G_4 + G_2G_4H_1 - G_1G_2G_4H_1 + G_2G_3G_4H_2$$

and from Equation (*8.4*) we obtain

$$H = \frac{\Delta - 1}{G} = \frac{G_2H_1 - G_1G_2H_1 + G_2G_3H_2}{G_1G_2G_3 + G_4 + G_2G_4H_1 - G_1G_2G_4H_1 + G_2G_3G_4H_2}$$

Supplementary Problems

8.21. Find C/R for Fig. 8-71, using Equation (*8.2*).

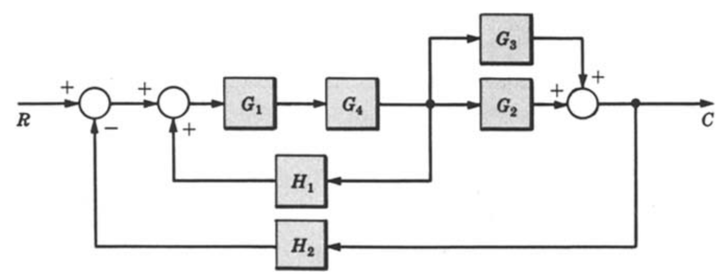


Fig. 8-71

8.22. Determine a set of canonical feedback system transfer functions for the preceding problem, using Equations (*8.3*) and (*8.4*).

8.23. Scale the signal flow graph in Fig. 8-72 so that X_3 becomes $X_3/2$ (see Problem 8.3).

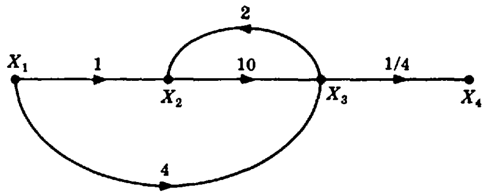


Fig. 8-72

8.24. Draw a signal flow graph for several nodes of the lateral inhibition system described in Problem 3.4 by the equation

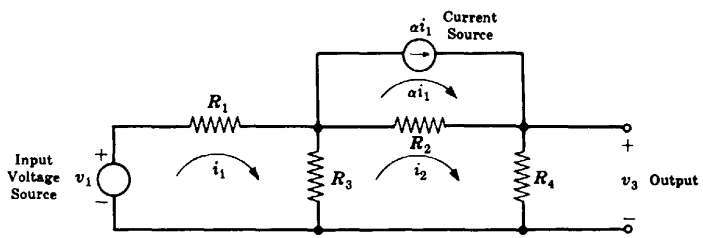
$$c_k = r_k - \sum_{i=1}^{n} a_{k-i} c_i$$

8.25. Draw a signal flow graph for the system presented in Problem 7.31.

8.26. Draw a signal flow graph for the system presented in Problem 7.32.

8.27. Determine C/R_4 from Equation (*8.2*) for the signal flow graph drawn in Problem 8.26.

8.28. Draw a signal flow graph for the electrical network in Fig. 8-73.



α = constant

Fig. 8-73

8.29. Determine V_3/V_1 from Equation (*8.2*) for the network of Problem 8.28.

8.30. Determine the elements for a canonical feedback system for the network of Problem 8.28, using Equations (*8.3*) and (*8.4*).

8.31. Draw the signal flow graph for the analog computer circuit in Fig 8-74.

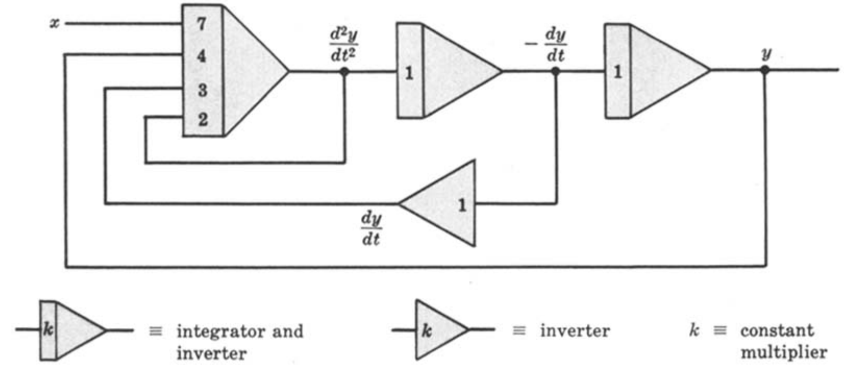


Fig. 8-74

8.32. Scale the analog computer circuit of Problem 8.31 so that y becomes $10y$, dy/dt becomes $20(dy/dt)$, and d^2y/dt^2 becomes $5(d^2y/dt^2)$.

Answers to Supplementary Problems

8.21. $P_1 = G_1G_2G_4$; $P_2 = G_1G_3G_4$, $P_{11} = G_1G_4H_1$, $P_{21} = -G_1G_2G_4H_2$, $P_{31} = -G_1G_3G_4H_2$, $\Delta = 1 - G_1G_4H_1 + G_1G_2G_4H_2 + G_1G_3G_4H_2$, and $\Delta_1 = \Delta_2 = 1$. Therefore

$$\frac{C}{R} = \frac{P_1\Delta_1 + P_2\Delta_2}{\Delta} = \frac{G_1G_4(G_2+G_3)}{1 - G_1G_4[H_1 - H_2(G_2+G_3)]}$$

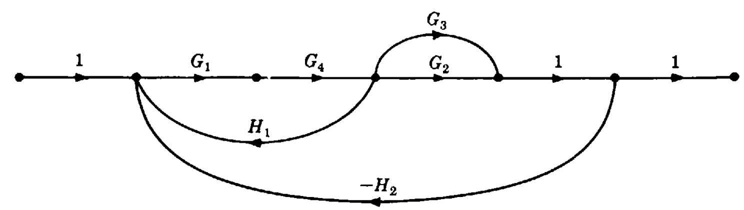


Fig. 8-75

8.22. $G = P_1\Delta_1 + P_2\Delta_2 = G_1G_4(G_2+G_3)$ $\qquad H = \dfrac{\Delta - 1}{G} = H_2 - \dfrac{H_1}{G_2+G_3}$

8.23.

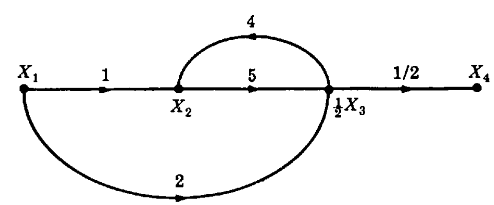


Fig. 8-76

8.24.

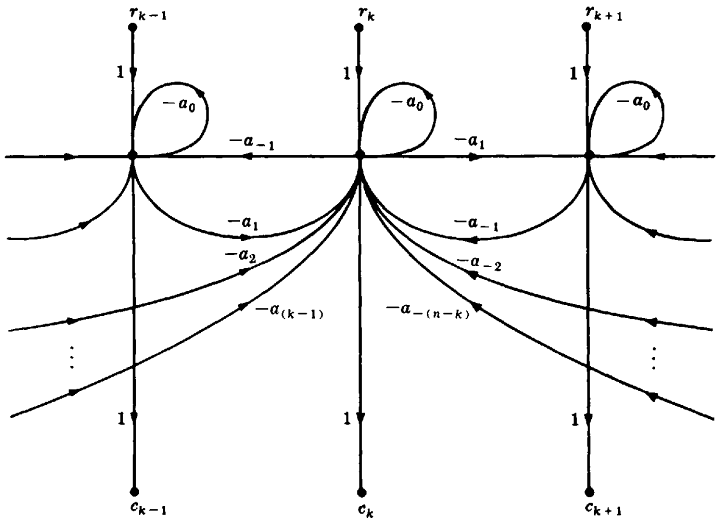


Fig. 8-77

8.25.

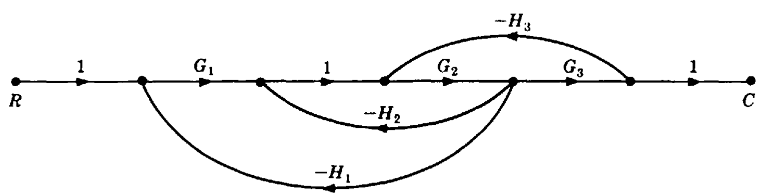


Fig. 8-78

8.26.

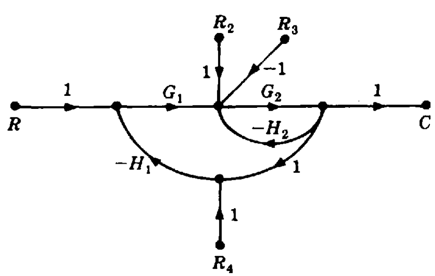


Fig. 8-79

8.27. $\dfrac{C}{R_4} = \dfrac{-G_1G_2H_1}{1 + G_2H_2 + G_1G_2H_1}$

8.28.

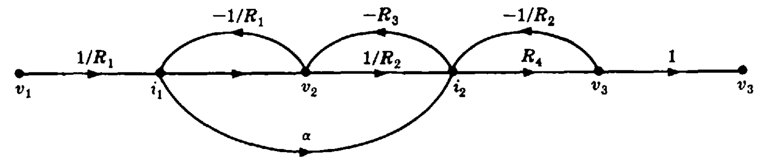


Fig. 8-80

8.29. $\dfrac{V_3}{V_1} = \dfrac{R_3R_4 + \alpha R_2R_4}{R_1R_2 + R_1R_3 + R_1R_4 + R_2R_3 + R_3R_4 - \alpha R_2R_3}$

8.30. $G = R_4(R_3 + \alpha R_2)$

$$H = \frac{R_1(R_2 + R_3 + R_4) + R_3R_4 + R_2R_3(1-\alpha)}{R_4(R_3 + \alpha R_2)}$$

8.31.

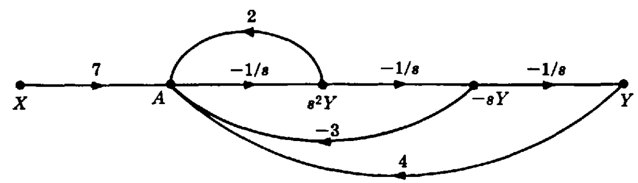


Fig. 8-81

8.32.

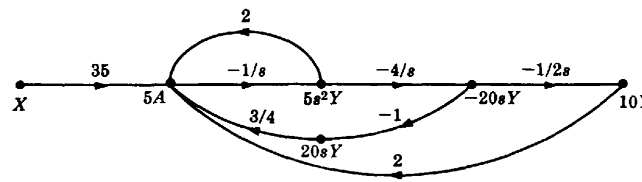


Fig. 8-82

Chapter 9

System Sensitivity Measures and Classification of Feedback Systems

9.1 INTRODUCTION

In earlier chapters the concepts of feedback and feedback systems have been emphasized. Since a system with a given transfer function can be synthesized in either an open-loop or a closed-loop configuration, a closed-loop (feedback) configuration must have some desirable properties which an open-loop configuration does not have.

In this chapter some of the properties of feedback and feedback systems are further discussed, and quantitative measures of the effectiveness of feedback are developed in terms of the concepts of *sensitivity* and *error constants*.

9.2 SENSITIVITY OF TRANSFER FUNCTIONS AND FREQUENCY RESPONSE FUNCTIONS TO SYSTEM PARAMETERS

An early step in the analysis or design of a control system is the generation of models for the various elements in the system. If the system is linear and time-invariant, two useful mathematical models for these elements are the *transfer function* and the *frequency response function* (see Chapter 6).

The transfer function is fixed when its parameters are specified, and the values given to these parameters are called **nominal values**. They are rarely, if ever, known exactly, so nominal values are actually approximations to true parameter values. The corresponding transfer function is called the **nominal transfer function**. The accuracy of the model then depends in part, on how closely these nominal parameter values approximate the real system parameters they represent, and also how much these parameters deviate from nominal values during the course of system operation. The **sensitivity** of a system to its parameters is a measure of how much the system transfer function differs from its nominal when each of its parameters differs from its nominal value.

System sensitivity can also be defined and analyzed in terms of the frequency response function. The frequency response function of a continuous system can be determined directly from the transfer function of the system, if it is known, by replacing the complex variable s in the transfer function by $j\omega$. For discrete-time systems, the frequency response function is obtained by replacing z by $e^{j\omega T}$. Thus the frequency response function is defined by the same parameters as those of the transfer function, and its accuracy is determined by the accuracy of these parameters. The frequency response function can alternatively be defined by graphs of its magnitude and phase- angle, both plotted as a function of the real frequency ω. These graphs are often determined experimentally, and in many cases cannot be defined by a finite number of parameters. Hence an infinite number of values of amplitude and phase angle (values for all frequencies) define the frequency response function. The **sensitivity** of the system is in this case a measure of the amount by which its frequency response function differs from its nominal when the frequency response function of an element of the system differs from its nominal value.

Consider the mathematical model $T(k)$ (transfer function or frequency response function) of a linear time-invariant system, written in polar form as

$$T(k)=|T(k)|e^{j\phi_T} \tag{9.1}$$

where k is a parameter upon which $T(k)$ depends. Usually both $|T(k)|$ and ϕ_T depend on k, and k is a real or complex parameter of the system.

Definition 9.1: For the mathematical model $T(k)$, with k regarded as the only parameter, the **sensitivity of $T(k)$ with respect to the parameter k** is defined by

$$S_k^{T(k)}\equiv\frac{d\ln T(k)}{d\ln k}=\frac{dT(k)/T(k)}{dk/k}=\frac{dT(k)}{dk}\,\frac{k}{T(k)} \tag{9.2}$$

In some treatments of this subject, $S_k^{T(k)}$ is called the **relative sensitivity**, or **normalized** sensitivity, because it represents the variation dT relative to the nominal T, for a variation dk relative to the nominal k. $S_k^{T(k)}$ is also sometimes called the *Bode sensitivity.*

Definition 9.2: The **sensitivity of the magnitude of $T(k)$ with respect to the parameter k** is defined by

$$S_k^{|T(k)|}\equiv\frac{d\ln|T(k)|}{d\ln k}=\frac{d|T(k)|/|T(k)|}{dk/k}=\frac{d|T(k)|}{dk}\,\frac{k}{|T(k)|} \tag{9.3}$$

Definition 9.3: The **sensitivity of the phase angle ϕ_T of $T(k)$ with respect to the parameter k** is defined by

$$S_k^{\phi_T}\equiv\frac{d\ln\phi_T}{d\ln k}=\frac{d\phi_T/\phi_T}{dk/k}=\frac{d\phi_T}{dk}\,\frac{k}{\phi_T} \tag{9.4}$$

The sensitivities of $T(k)=|T(k)|e^{j\phi_T}$, the magnitude $|T(k)|$, and the phase angle ϕ_T with respect to the parameter k are related by the expression

$$S_k^{T(k)}=S_k^{|T(k)|}+j\phi_T S_k^{\phi_T} \tag{9.5}$$

Note that, in general, $S_k^{|T(k)|}$ and $S_k^{\phi_T}$ are complex numbers. In the special but very important case where k is real, then both $S_k^{|T(k)|}$ and $S_k^{\phi_T}$ are real. When $S_k^{T(k)}=0$, $T(k)$ is **insensitive** to k.

EXAMPLE 9.1. Consider the frequency response function

$$T(\mu)=e^{-j\omega\mu}$$

where $\mu\equiv k$. The magnitude of $T(\mu)$ is $|T(\mu)|=1$, and the phase angle of $T(\mu)$ is $\phi_T=-\omega\mu$.

The sensitivity of $T(\mu)$ with respect to the parameter μ is

$$S_\mu^{T(\mu)}=\frac{d(e^{-j\omega\mu})}{d_\mu}\,\frac{\mu}{e^{-j\omega\mu}}=-j\omega\mu$$

The sensitivity of the magnitude of $T(\mu)$ with respect to the parameter μ is

$$S_\mu^{|T(\mu)|}=\frac{d|T(\mu)|}{\mu}\,\frac{\mu}{|T(\mu)|}=0$$

The sensitivity of the phase angle of $T(\mu)$ with respect to the parameter μ is

$$S_\mu^{\phi_T}=\frac{d\phi_T}{d\mu}\,\frac{\mu}{\phi_T}=-\omega\cdot\frac{\mu}{-\omega\mu}=1$$

Note that

$$S_\mu^{|T(\mu)|}+j\phi_T S_\mu^{\phi_T}=-j\omega\mu=S_\mu^{T(\mu)}$$

The following development is in terms of transfer functions. However, everything is applicable to frequency response functions (for continuous systems) by simply replacing s in all equations by $j\omega$, or $z=e^{j\omega T}$ for discrete systems.

A special but very important class of system transfer functions has the form:

$$T=\frac{A_1+kA_2}{A_3+kA_4} \tag{9.6}$$

where k is a parameter and A_1, A_2, A_3, and A_4 are polynomials in s (or z). This type of dependence between a parameter k and a transfer function T is general enough to include many of the systems considered in this book.

For a transfer function with the form of Equation (9.6), the *sensitivity of T with respect to the parameter k* is given by

$$S_k^T \equiv \frac{dT}{dk}\cdot\frac{k}{T} = \frac{k(A_2A_3 - A_1A_4)}{(A_3 + kA_4)(A_1 + kA_2)} \tag{9.7}$$

In general, S_k^T is a function of the complex variable s (or z).

EXAMPLE 9.2. The transfer function of the discrete-time system given in Fig. 9-1 is

$$T \equiv \frac{C}{R} = \frac{K}{z^3 + (a+b)z^2 + abz + K}$$

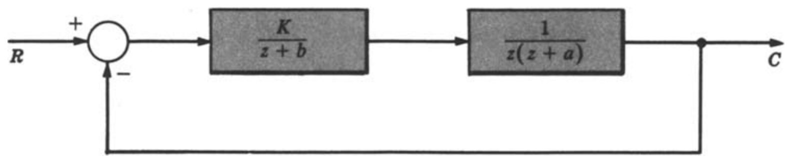


Fig. 9-1

If K is the parameter of interest ($k \equiv K$), we group terms in T as follows:

$$T = \frac{K}{\left[z^3 + (a+b)z^2 + abz\right] + K}$$

Comparing T with Equation (9.6), we see that

$$A_1 = 0 \qquad A_2 = 1 \qquad A_3 = z^3 + (a+b)z^2 + abz \qquad A_4 = 1$$

If a is the parameter of interest ($k \equiv a$), T can be rewritten as

$$T = \frac{K}{\left[z^3 + bz^2 + K\right] + a\left[z^2 + bz\right]}$$

Comparing this expression with Equation (9.6) we see that

$$A_1 = K \qquad A_2 = 0 \qquad A_3 = z^3 + bz^2 + K \qquad A_4 = z^2 + bz$$

If b is the parameter of interest ($k \equiv b$), T can be rewritten as

$$T = \frac{K}{\left[z^3 + az^2 + K\right] + b\left[z^2 + az\right]}$$

Again comparing this expression with Equation (9.6), we see that

$$A_1 = K \qquad A_2 = 0 \qquad A_3 = z^3 + az^2 + K \qquad A_4 = z^2 + az$$

EXAMPLE 9.3. For the lead network shown in Fig. 9-2 the transfer function is

$$T \equiv \frac{E_0}{E_i} = \frac{1 + RCs}{2 + RCs}$$

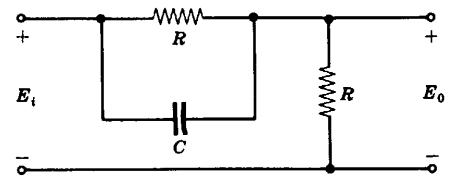


Fig. 9-2

If C (capacitance) is the parameter of interest, we write $T = (1 + C(Rs))/[2 + C(Rs)]$. Comparing this expression with Equation (9.6), we see that $A_1 = 1$, $A_2 = Rs$, $A_3 = 2$, $A_4 = Rs$.

EXAMPLE 9.4. For the system of Example 9.2 *the sensitivity of T with respect to K* is

$$S_K^T = \frac{K\left[z^3 + (a+b)z^2 + abz\right]}{K\left[z^3 + (a+b)z^2 + abz + K\right]} = \frac{1}{1 + \dfrac{K}{z^3 + (a+b)z^2 + abz}}$$

The sensitivity of T with respect to the parameter a is

$$S_a^T = \frac{-aK(z^2 + bz)}{K\left[z^3 + bz^2 + K + a(z^2 + bz)\right]} = \frac{-1}{1 + \dfrac{z^3 + bz^2 + K}{a(z^2 + bz)}}$$

The sensitivity of T with respect to the parameter b is

$$S_b^T = \frac{-bK(z^2 + az)}{K\left[z^3 + az^2 + K + b(z^2 + az)\right]} = \frac{-1}{1 + \dfrac{z^3 + az^2 + K}{b(z^2 + az)}}$$

EXAMPLE 9.5. For the lead network of Fig. 9-2 the sensitivity of T with respect to the capacitance C is

$$S_C^T = \frac{C(2Rs - Rs)}{(2 + RCs)(1 + RCs)} = \frac{RCs}{(2 + RCs)(1 + RCs)} = \frac{1}{(1 + 2/RCs)(1 + 1/RCs)}$$

EXAMPLE 9.6. The open-loop and closed-loop systems given in Fig. 9-3 have the same plant and the same overall system transfer function for $K = 2$.

$$\left(\frac{C}{R}\right)_1 = \frac{K}{s^2 + 4s + 5}$$

$$\left(\frac{C}{R}\right)_2 = \frac{K}{s^2 + 4s + 3 + K}$$

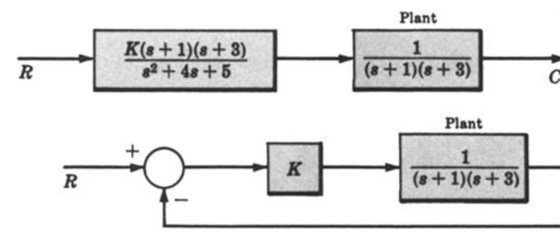


Fig. 9-3

Although these systems are precisely equivalent for $K = 2$, their properties differ significantly for small (and large) deviations of K from $K = 2$. The transfer function of the first system is

$$T_1 \equiv \left(\frac{C}{R}\right) = \frac{K}{s^2 + 4s + 5}$$

Comparing this expression with Equation (9.6) gives $A_1 = 0$, $A_2 = 1$, $A_3 = s^2 + 4s + 5$, $A_4 = 0$. Substituting these values into Equation (9.7), we obtain

$$S_K^{T_1} = \frac{K(s^2+4s+5)}{(s^2+4s+5)K} = 1$$

for all K.

The transfer function of the second system is

$$T_2 \equiv \left(\frac{C}{R}\right)_2 = \frac{K}{s^2+4s+3+K}$$

Comparing this expression with Equation (9.6) yields $A_1 = 0$, $A_2 = 1$, $A_3 = s^2+4s+3$, $A_4 = 1$. Substituting these values into Equation (9.7), we obtain

$$S_K^{T_2} = \frac{K(s^2+4s+3)}{(s^2+4s+3+K)(K)} = \frac{1}{1+K/(s^2+4s+3)}$$

For $K=2$, $S_K^{T_2} = 1/[1+2/(s^2+s+3)]$.

Note that the sensitivity of the open-loop system T_1 is fixed at 1 for all values of gain K. On the other hand, the closed-loop sensitivity is a function of K and the complex variable s. Thus $S_K^{T_2}$ may be adjusted in a design problem by varying K or maintaining the frequencies of the input function within an appropriate range.

For $\omega < \sqrt{3}$ rad/sec, the sensitivity of the closed-loop system is

$$S_K^{T_2} \simeq \frac{1}{1+\frac{2}{3}} = \frac{3}{5} = 0.6$$

Thus the feedback system is 40% less sensitive than the open-loop system for low frequencies. For high frequencies, the sensitivity of the closed-loop system approaches 1, the same as that of the open-loop system.

EXAMPLE 9.7. Suppose G is a frequency response function, either $G(j\omega)$ for a continuous system, or $G(e^{j\omega T})$ for a discrete-time system. The frequency response function for the unity feedback system (continuous or discrete-time) given in Fig. 9-4 is related to the forward-loop frequency response function G by

$$\frac{C}{R} = \left|\frac{C}{R}\right| e^{j\phi_{C/R}} = \frac{G}{1+G} = \frac{|G|e^{j\phi_G}}{1+|G|e^{j\phi_G}}$$

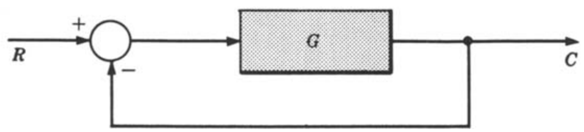


Fig. 9-4

where $\phi_{C/R}$ is the phase angle of C/R and ϕ_G is the phase angle of G. *The sensitivity of C/R with respect to $|G|$* is given by

$$S_{|G|}^{C/R} = \frac{d(C/R)}{d|G|}\cdot\frac{|G|}{C/R} = \frac{e^{j\phi_G}}{(1+|G|e^{j\phi_G})^2}\cdot\frac{|G|}{\dfrac{|G|e^{j\phi_G}}{1+|G|e^{j\phi_G}}}$$

$$= \frac{1}{1+|G|e^{j\phi_G}} = \frac{1}{1+G} \qquad (9.8)$$

Note that for large $|G|$ the sensitivity of C/R to $|G|$ is relatively small.

EXAMPLE 9.8. Suppose the system of Example 9.7 is continuous, that $\omega = 1$, and for some given $G(j\omega)$, $G(j1) = 1+j$. Then $|G(j\omega)| = \sqrt{2}$, $\phi_G = \pi/4$ rad, $(C/R)(j\omega) = \frac{3}{5} + j\frac{1}{5}$, $|(C/R)(j\omega)| = \sqrt{10}/5$, and $\phi_{C/R} = 0.3215$ rad.

Using the result of the previous example, the *sensitivity* of $(C/R)(j\omega)$ *with respect to* $|G(j\omega)|$ is

$$S_{|G(j\omega)|}^{(C/R)(j\omega)} = \frac{1}{2+j} = \frac{2}{5} - j\frac{1}{5}$$

Then from Equation (9.5) we have

$$S_{|G(j\omega)|}^{|(C/R)(j\omega)|} = \frac{2}{5} = 0.4 \qquad \phi_{C/R} S_{|G(j\omega)|}^{\phi_{C/R}} = -\frac{1}{5} \qquad S_{|G(j\omega)|}^{\phi_{C/R}} = -\frac{1}{5(0.3215)} = -0.622$$

These real values of sensitivity mean that a 10% change in $|G(j\omega)|$ will produce a 4% change in $|(C/R)(j\omega)|$ and a -6.22% change in $\phi_{C/R}$.

A qualitative attribute of a system related to its sensitivity is its *robustness*. A system is said to be **robust** when its operation is insensitive to parameter variations. Robustness may be characterized in terms of the sensitivity of its transfer or frequency response function, or of a set of performance indices to system parameters.

9.3 OUTPUT SENSITIVITY TO PARAMETERS FOR DIFFERENTIAL AND DIFFERENCE EQUATION MODELS

The concept of sensitivity is also applicable to system models expressed in the time domain. **The sensitivity of the model output y to any parameter p** is given by

$$S_p^{y(t)} \equiv S_p^y = \frac{d(\ln y)}{d(\ln p)} = \frac{dy/y}{dp/p} = \frac{dy}{dp}\frac{p}{y}$$

Since the model is defined in the time domain, the sensitivity is usually found by solving for the output $y(t)$ in the time domain. The derivative dy/dp is sometimes called the **output sensitivity coefficient**, which is generally a function of time, as is the sensitivity S_p^y.

EXAMPLE 9.9. We determine the sensitivity of the output $y(t) = x(t)$ to the parameter a for the differential system $\dot{x} = ax + u$. The sensitivity is

$$S_a^y = \frac{dy}{da}\frac{a}{y} = \frac{dx}{da}\frac{a}{x}$$

To determine S_a^y, consider the *time derivative* of dx/da, and interchange the order of differentiation, that is,

$$\frac{d}{dt}\left(\frac{dx}{da}\right) = \frac{d}{da}\left(\frac{dx}{dt}\right) = \frac{d}{da}(ax+u)$$

Now define a new variable $v \equiv dx/da$. Then

$$\dot{v} = \frac{d}{da}(ax+u) = a\frac{dx}{da} + 1\cdot x = av + x$$

The sensitivity function S_a^y can then be found by first solving the system differential equation for $x(t)$, because $x(t)$ is the forcing function in the differential equation for $v(t)$ above. The required solutions were developed in Section 3.15 as

$$x(t) = e^{at}x(0) + \int_0^t e^{a(t-\tau)}u(\tau)\,d\tau$$

and

$$v(t) = \int_0^t e^{a(t-\tau)}x(\tau)\,d\tau$$

because $v(0) = 0$. The time-varying output sensitivity is computed from these two functions as

$$S_a^y = \frac{dx}{da}\frac{a}{x} = \frac{av(t)}{x(t)}$$

EXAMPLE 9.10. For the discrete system defined by

$$x(k+1) = ax(k) + u(k)$$
$$y(k) = cx(k)$$

we determine the sensitivity of the output y to the parameter a as follows. Let

$$v(k) \equiv \frac{\partial x(k)}{\partial a}$$

Then

$$v(k+1) = \frac{\partial x(k+1)}{\partial a} = \frac{\partial}{\partial a}[ax(k) + u(k)]$$
$$= x(k) + a\frac{\partial x(k)}{\partial a} = av(k) + x(k)$$

and

$$\frac{\partial y(k)}{\partial a} = \frac{\partial cx(k)}{\partial a} = c\frac{\partial x(k)}{\partial a} = cv(k)$$

Thus, to determine S_a^y, we first solve the two discrete equations:

$$x(k+1) = ax(k) + u(k)$$
$$v(k+1) = av(k) + x(k)$$

(e.g., see Section 3.17). Then

$$S_a^y = \frac{\partial y(k)}{\partial a} \cdot \frac{a}{y(k)} = \frac{av(k)}{x(k)}$$

9.4 CLASSIFICATION OF CONTINUOUS FEEDBACK SYSTEMS BY TYPE

Consider the class of canonical feedback systems defined by Fig. 9-5. For continuous systems, the open-loop transfer function may be written as

$$GH = \frac{K\prod_{i=1}^{m}(s+z_i)}{\prod_{i=1}^{n}(s+p_i)}$$

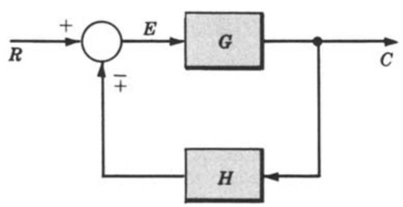


Fig. 9-5

where K is a constant, $m \le n$, and $-z_i$ and $-p_i$ are the finite zeros and poles, respectively, of GH. If there are a zeros and b poles at the origin, then

$$GH = \frac{Ks^a\prod_{i=1}^{m-a}(s+z_i)}{s^b\prod_{i=1}^{n-b}(s+p_i)}$$

In the remainder of this chapter, only systems for which $b \ge a$ are considered, and $l \equiv b - a$.

Definition 9.4: A canonical feedback system whose open-loop transfer function can be written in the form:

$$GH = \frac{K\prod_{i=1}^{m-a}(s+z_i)}{s^l\prod_{i=1}^{n-a-l}(s+p_i)} \equiv \frac{KB_1(s)}{s^lB_2(s)} \tag{9.9}$$

where $l \ge 0$ and $-z_i$ and $-p_i$ are the nonzero finite zeros and poles of GH, respectively, is called a **type l system**.

EXAMPLE 9.11. The system defined by Fig. 9-6 is a *type 2 system*.

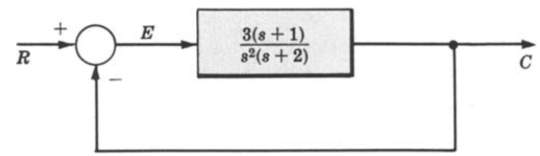


Fig. 9-6

EXAMPLE 9.12. The system defined by Fig. 9-7 is a *type 1 system*.

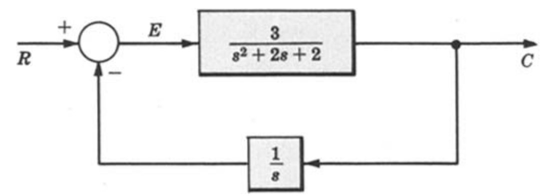


Fig. 9-7

EXAMPLE 9.13. The system defined by Fig. 9-8 is a *type 0 system*.

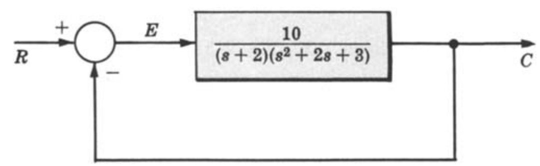


Fig. 9-8

9.5 POSITION ERROR CONSTANTS FOR CONTINUOUS UNITY FEEDBACK SYSTEMS

One criterion of the effectiveness of feedback in a *stable type l unity feedback system* is the *position (step) error constant*. It is a measure of the steady state error between the input and output when the input is a unit step function, that is, the difference between the input and output when the system is in steady state and the input is a step.

Definition 9.5: The **position error constant** K_p of a type l unity feedback system is defined as

$$K_p \equiv \lim_{s\to 0} G(s) = \lim_{s\to 0} \frac{KB_1(s)}{s^l B_2(s)} = \begin{cases} \dfrac{KB_1(0)}{B_2(0)} & \text{for } l=0 \\ \infty & \text{for } l>0 \end{cases} \tag{9.10}$$

The steady state error of a stable type l unity feedback system when the input is a unit step function $[e(\infty) = 1 - c(\infty)]$ is related to the position error constant by

$$e(\infty) = \lim_{t\to\infty} e(t) = \frac{1}{1+K_p} \tag{9.11}$$

EXAMPLE 9.14. The position error constant for a type 0 system is finite. That is,

$$|K_p| = \left|\frac{KB_1(0)}{B_2(0)}\right| < \infty$$

The steady state error for a type 0 system is nonzero and finite.

EXAMPLE 9.15. The position error constant for a type 1 system is

$$K_p = \lim_{s\to 0} \frac{KB_1(0)}{sB_2(0)} = \infty$$

Therefore the steady state error is $e(\infty) = 1/(1+K_p) = 0$.

EXAMPLE 9.16. The position error constant for a type 2 system is

$$K_p = \lim_{s\to 0} \frac{KB_1(s)}{s^2 B_2(s)} = \infty$$

Therefore the steady state error is $e(\infty) = 1/(1+K_p) = 0$.

9.6 VELOCITY ERROR CONSTANTS FOR CONTINUOUS UNITY FEEDBACK SYSTEMS

Another criterion of the effectiveness of feedback in a *stable type l unity feedback system* is the *velocity (ramp) error constant*. It is a measure of the steady state error between the input and output of the system when the input is a unit ramp function.

Definition 9.6: The **velocity error constant** K_v of a stable type l unity feedback system is defined as

$$K_v \equiv \lim_{s\to 0} sG(s) = \lim_{s\to 0} \frac{KB_1(s)}{s^{l-1} B_2(s)} = \begin{cases} 0 & \text{for } l=0 \\ \dfrac{KB_1(0)}{B_2(0)} & \text{for } l=1 \\ \infty & \text{for } l>1 \end{cases} \tag{9.12}$$

The steady state error of a stable type l unity feedback system when the input is a unit ramp function is related to the velocity error constant by

$$e(\infty) = \lim_{t\to\infty} e(t) = \frac{1}{K_v} \tag{9.13}$$

EXAMPLE 9.17. The velocity error constant for a type 0 system is $K_v = 0$. Hence the steady state error is infinite.

EXAMPLE 9.18. The velocity error constant for a type 1 system, $K_v = KB_1(0)/B_2(0)$, is finite. Therefore the steady state error is nonzero and finite.

EXAMPLE 9.19. The velocity error constant for a type 2 system is infinite. Therefore the steady state error is zero.

9.7 ACCELERATION ERROR CONSTANTS FOR CONTINUOUS UNITY FEEDBACK SYSTEMS

A third criterion of the effectiveness of feedback in a *stable type l unity feedback* system is the *acceleration (parabolic) error constant*. It is a measure of the steady state error of the system when the input is a unit parabolic function; that is, $r = t^2/2$ and $R = 1/s^3$.

Definition 9.7: The **acceleration error constant** K_a of a stable type l unity feedback system is defined as

$$K_a \equiv \lim_{s\to 0} s^2 G(s) = \lim_{s\to 0} \frac{KB_1(s)}{s^{l-2} B_2(s)} = \begin{cases} 0 & \text{for } l=0,1 \\ \dfrac{KB_1(0)}{B_2(0)} & \text{for } l=2 \\ \infty & \text{for } l>2 \end{cases} \tag{9.14}$$

The steady state error of a stable type l unity feedback system when the input is a unit parabolic function is related to the acceleration error constant by

$$e(\infty) = \lim_{t\to\infty} e(t) = \frac{1}{K_a} \tag{9.15}$$

EXAMPLE 9.20. The acceleration error constant for a type 0 system is $K_a = 0$. Hence the steady state error is infinite.

EXAMPLE 9.21. The acceleration error constant for a type 1 system is $K_a = 0$. Hence the steady state error is infinite.

EXAMPLE 9.22. The acceleration error constant for a type 2 system, $K_a = KB_1(0)/B_2(0)$, is finite. Hence the steady state error is nonzero and finite.

9.8 ERROR CONSTANTS FOR DISCRETE UNITY FEEDBACK SYSTEMS

The open-loop transfer function for a type l discrete system can be written as

$$GH = \frac{K(z+z_1)\cdots(z+z_m)}{(z-1)^l(z+p_1)\cdots(z+p_n)} = \frac{KB_1(z)}{(z-1)^l B_2(z)}$$

where $l \geq 0$ and $-z_i$ and $-p_i$ are the nonunity zeros and poles of GH in the z-plane.

All the results developed for continuous unity feedback systems in Sections 9.5 through 9.7 are the same for discrete systems with this open-loop transfer function.

9.9 SUMMARY TABLE FOR CONTINUOUS AND DISCRETE-TIME UNITY FEEDBACK SYSTEMS

In Table 9.1 the error constants are given in terms of α, where $\alpha = 0$ for continuous systems, and $\alpha = 1$ for discrete-time systems. For continuous systems $T = 1$ in the steady state error.

TABLE 9.1

Input	Unit Step		Unit Ramp		Unit Parabola	
System Type	K_p	Steady State Error	K_v	Steady State Error	K_a	Steady State Error
Type 0	$\dfrac{KB_1(\alpha)}{B_2(\alpha)}$	$\dfrac{1}{1+K_p}$	0	∞	0	∞
Type 1	∞	0	$\dfrac{KB_1(\alpha)}{B_2(\alpha)}$	$\dfrac{T}{K_v}$	0	∞
Type 2	∞	0	∞	0	$\dfrac{KB_1(\alpha)}{B_2(\alpha)}$	$\dfrac{T^2}{K_a}$

9.10 ERROR CONSTANTS FOR MORE GENERAL SYSTEMS

The results of Sections 9.5 through 9.9 are only applicable to stable unity feedback linear systems. They can be readily extended, however, to more general stable linear systems. In Fig. 9-9, T_d represents the transfer function of a desired (ideal) system, and C/R represents the transfer function of the actual system (an approximation of T_d). R is the input to both systems, and E is the difference (the error) between the desired output and the actual output. For this more general system, three error constants are defined below and are related to the steady state error.

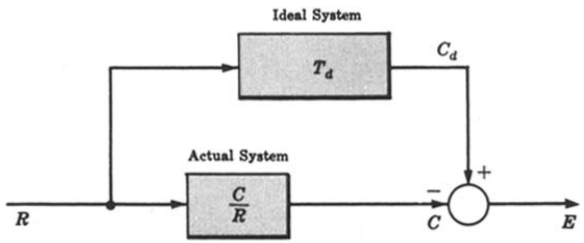


Fig. 9-9

Definition 9.8: The **step error constant** K_s is defined for continuous systems as

$$K_s \equiv \frac{1}{\lim\limits_{s\to 0}\left[T_d - \dfrac{C}{R}\right]} \tag{9.16}$$

The steady state error for the general system when the input is a unit step function is related to K_s by

$$e(\infty) = \lim_{t\to\infty} e(t) = \frac{1}{K_s} \tag{9.17}$$

Definition 9.9: The **ramp error constant** K_r is defined for continuous systems as

$$K_r \equiv \frac{1}{\lim\limits_{s\to 0}\dfrac{1}{s}\left[T_d - \dfrac{C}{R}\right]} \tag{9.18}$$

The steady state error for the general system when the input is a unit ramp function is related to K_r by

$$e(\infty) = \lim_{t\to\infty} e(t) = \frac{1}{K_r} \tag{9.19}$$

Definition 9.10: The **parabolic error constant** K_{pa} is defined for continuous systems as

$$K_{pa} \equiv \frac{1}{\lim\limits_{s\to 0}\dfrac{1}{s^2}\left[T_d - \dfrac{C}{R}\right]} \tag{9.20}$$

The steady state error for the general system when the input is a unit parabolic function is related to K_{pa} by

$$e(\infty) = \lim_{t\to\infty} e(t) = \frac{1}{K_{pa}} \tag{9.21}$$

EXAMPLE 9.23. The nonunity feedback system given in Fig. 9-10 has the transfer function $C/R = 2/(s^2+2s+4)$. If the desired transfer function which C/R approximates is $T_d = \frac{1}{2}$, then

$$T_d - \frac{C}{R} = \frac{s(s+2)}{2(s^2+2s+4)}$$

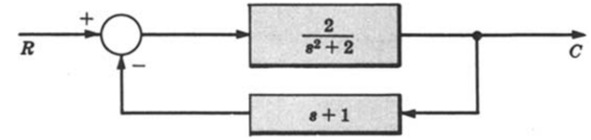


Fig. 9-10

Therefore

$$K_s = \frac{1}{\lim\limits_{s\to 0}\left[\dfrac{s(s+2)}{2(s^2+2s+4)}\right]} = \infty \qquad K_r = \frac{1}{\lim\limits_{s\to 0}\dfrac{1}{s}\left[\dfrac{s(s+2)}{2(s^2+2s+4)}\right]} = 4$$

$$K_{pa} = \frac{1}{\lim\limits_{s\to 0}\dfrac{1}{s^2}\left[\dfrac{s(s+2)}{2(s^2+2s+4)}\right]} = 0$$

EXAMPLE 9.24. For the system of Example 9.23 the steady state errors due to a unit step input, a unit ramp input, and a unit parabolic input can be found using the results of that example. For a unit step input, $e(\infty) = 1/K_s = 0$. For a unit ramp input, $e(\infty) = 1/K_r = \frac{1}{4}$. For a unit parabolic input, $e(\infty) = 1/K_{pa} = \infty$.

To establish relationships between the general error constants K_s, K_r, and K_{pa} and the error constants K_p, K_v, and K_a for unity feedback systems, we let the actual system be a continuous unity feedback system and let the desired system have a unity transfer function. That is, we let

$$T_d = 1 \qquad \text{and} \qquad \frac{C}{R} = \frac{G}{1+G}$$

Therefore

$$K_s = \frac{1}{\lim\limits_{s\to 0}\left[\dfrac{1}{1+G}\right]} = 1 + \lim_{s\to 0} G(s) = 1 + K_p \tag{9.22}$$

$$K_r = \frac{1}{\lim\limits_{s\to 0}\left[\dfrac{1}{s}\left(\dfrac{1}{1+G}\right)\right]} = \lim_{s\to 0} sG(s) = K_v \tag{9.23}$$

$$K_{pa} = \frac{1}{\lim\limits_{s\to 0}\left[\dfrac{1}{s^2}\left(\dfrac{1}{1+G}\right)\right]} = \lim_{s\to 0} s^2G(s) = K_a \tag{9.24}$$

Solved Problems

SYSTEM CONFIGURATIONS

9.1. A given plant has the transfer function G_2. A system is desired which includes G_2 as the output element and has a transfer function C/R. Show that, if no constraints (such as stability) are placed on the compensating elements, then such a system can be synthesized as either an open-loop or a unity feedback system.

If the system can be synthesized as an open-loop system, then it will have the configuration given in Fig. 9-11, where G_1 is an unknown compensating element. The system transfer function is $C/R = G_1'G_2$, from which $G_1' = (C/R)/G_2$. This value for G_1' permits synthesis of C/R as an open-loop system.

Fig. 9-11

If the system can be synthesized as a unity feedback system, then it will have the configuration given in Fig. 9-12.

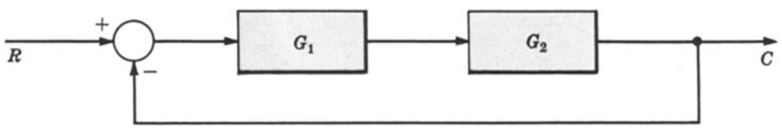


Fig. 9-12

The system transfer function is $C/R = G_1G_2/(1+G_1G_2)$ from which

$$G_1 = \frac{1}{G_2}\left(\frac{C/R}{1-C/R}\right)$$

This value for G_1 permits synthesis of C/R as a unity feedback system.

9.2. Using the results of Problem 9.1, show how the system transfer function $C/R = 2/(s^2+s+2)$ which includes as its output element the plant $G_2 = 1/s(s+1)$ can be synthesized as (*a*) an open-loop system, (*b*) a unity feedback system.

(*a*) For the open-loop system,

$$G_1' = \frac{C/R}{G_2} = \frac{2s(s+1)}{s^2+s+2}$$

and the system block diagram is given in Fig. 9-13.

Fig. 9-13

(*b*) For the unity feedback system,

$$G_1 = \frac{1}{G_2}\left(\frac{C/R}{1-C/R}\right) = s(s+1)\left[\frac{2/(s^2+s+2)}{(s^2+s+2-2)/(s^2+s+2)}\right] = 2$$

and the system block diagram is given in Fig. 9-14.

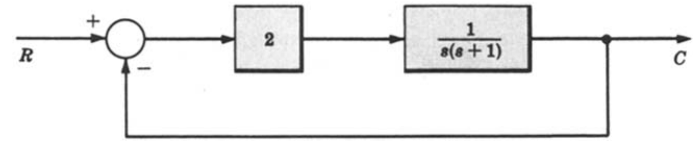


Fig. 9-14

TRANSFER FUNCTION SENSITIVITY

9.3. The two systems given in Fig. 9-15 have the same transfer function when $K_1 = K_2 = 100$.

$$T_1 = \left(\frac{C}{R}\right)_1\Bigg|_{\substack{K_1=100\\K_2=100}} = \frac{K_1K_2}{1+0.0099K_1K_2} = 100$$

$$T_2 = \left(\frac{C}{R}\right)_2\Bigg|_{\substack{K_1=100\\K_2=100}} = \left(\frac{K_1}{1+0.09K_1}\right)\left(\frac{K_2}{1+0.09K_2}\right) = 100$$

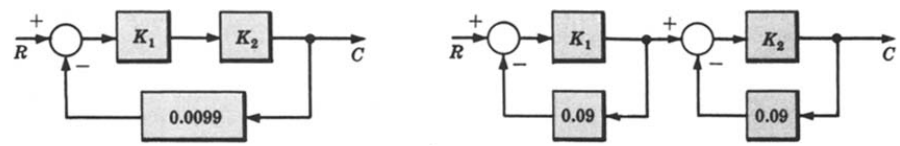


Fig. 9-15

Compare the sensitivities of these two systems with respect to parameter K_1 for nominal values $K_1 = K_2 = 100$.

For the first system, $T_1 = K_1K_2/[1+K_1(0.0099K_2)]$. Comparing this expression with Equation (*9.6*) yields $A_1 = 0$, $A_2 = K_2$, $A_3 = 1$, $A_4 = 0.0099K_2$. Substituting these values into Equation (*9.7*), we obtain

$$S_{K_1}^{T_1} = \frac{K_1K_2}{(1+0.0099K_1K_2)(K_1K_2)} = \frac{1}{1+0.0099K_1K_2} = 0.01 \qquad \text{for} \quad K_1 = K_2 = 100$$

For the second system,

$$T_2 = \left(\frac{K_1}{1+0.09K_1}\right)\left(\frac{K_2}{1+0.09K_2}\right) = \frac{K_1K_2}{1+0.09K_1+0.09K_2+0.0081K_1K_2}$$

Comparing this expression with Equation (*9.6*) yields $A_1 = 0$, $A_2 = K_2$, $A_3 = 1+0.09K_2$, $A_4 = 0.09 + 0.0081K_2$. Substituting these values into Equation (*9.7*), we have

$$S_{K_1}^{T_2} = \frac{K_1K_2(1+0.09K_2)}{(1+0.09K_1)(1+0.09K_2)(K_1K_2)} = \frac{1}{1+0.09K_1} = 0.1 \qquad \text{for} \quad K_1 = K_2 = 100$$

A 10% variation in K_1 will approximately produce a 0.1% variation in T_1 and a 1% variation in T_2. Thus the second system T_2 is 10 times more sensitive to variations in K_1 than is the first system T_1.

9.4. The closed-loop system given in Fig. 9-16 is defined in terms of the frequency response function of the feedforward element $G(j\omega)$.

$$\frac{C}{R}(j\omega) = \frac{G(j\omega)}{1+G(j\omega)}$$

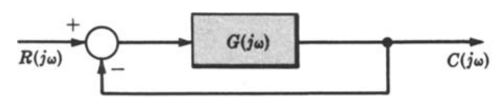


Fig. 9-16

Suppose that $G(j\omega) = 1/(j\omega + 1)$. In Chapter 15 it is shown that the frequency response functions $1/(j\omega + 1)$ can be approximated by the straight line graphs of magnitude and phase of $G(j\omega)$ given in Fig. 9-17.

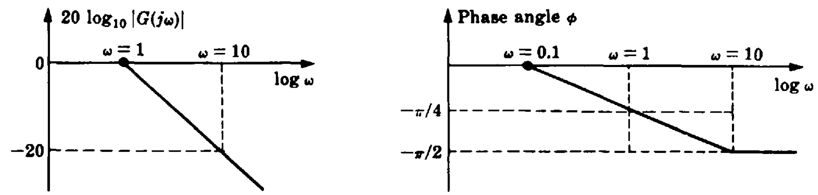


Fig. 9-17

At $\omega = 1$ the true values of $20\log_{10}|G(j\omega)|$ and ϕ are -3 and $-\pi/4$, respectively. For $\omega = 1$, find:

(*a*) The sensitivity of $|(C/R)(j\omega)|$ with respect to $|G(j\omega)|$.

(*b*) Using the result of part (*a*), determine an approximate value for the error in $|(C/R)(j\omega)|$ caused by using the straight-line approximations for $1/(j\omega + 1)$.

(*a*) Using Equation (*9.8*) the sensitivity of $(C/R)(j\omega)$ with respect to $|G(j\omega)|$ is given by

$$S_{|G(j\omega)|}^{(C/R)(j\omega)} = \frac{1}{1 + G(j\omega)} = \frac{1}{2 + j\omega} = \frac{2 - j\omega}{4 + \omega^2}$$

Since $|G(j\omega)|$ is real,

$$S_{|G(j\omega)|}^{|(C/R)(j\omega)|} = \mathrm{Re}\, S_{|G(j\omega)|}^{(C/R)(j\omega)} = \frac{2}{4 + \omega^2}$$

For $\omega = 1$, $S_{|G(j\omega)|}^{|(C/R)(j\omega)|} = 0.4$.

(*b*) For $\omega = 1$, the exact value of $|G(j\omega)|$ is $|G(j\omega)| = 1/\sqrt{2} = 0.707$. The approximate value taken from the graph is $|G(j\omega)| = 1$. Then the percentage error in the approximation is $100(1 - 0.707)/0.707 = 41.4\%$. The approximate percentage error in $|(C/R)(j\omega)|$ is $41.4\, S_{|G(j\omega)|}^{|(C/R)(j\omega)|} = 16.6\%$.

9.5. Show that the sensitivities of $T(k) = |T(k)|e^{j\phi_T}$, the magnitude $|T(k)|$, and the phase angle ϕ_T with respect to parameter k are related by

$$S_k^{T(k)} = S_k^{|T(k)|} + j\phi_T \cdot S_k^{\phi_T} \qquad [\text{Equation } (9.5)]$$

Using Equation (*9.2*),

$$S_k^{T(k)} = \frac{d\ln T(k)}{d\ln k} = \frac{d\ln\left[|T(k)|e^{j\phi_T}\right]}{d\ln k} = \frac{d\left[\ln|T(k)| + j\phi_T\right]}{d\ln k}$$

$$= \frac{d\ln|T(k)|}{d\ln k} + j\frac{d\phi_T}{d\ln k} = \frac{d\ln|T(k)|}{d\ln k} + j\phi_T\frac{d\ln\phi_T}{d\ln k} = S_k^{|T(k)|} + j\phi_T S_k^{\phi_T}$$

Note that if k is real, then $S_k^{|T(k)|}$ and $S_k^{\phi_T}$ are both real, and

$$S_k^{|T(k)|} = \mathrm{Re}\, S_k^{T(k)} \qquad \phi_T S_k^{\phi_T} = \mathrm{Im}\, S_k^{T(k)}$$

9.6. Show that the sensitivity of the transfer function $T = (A_1 + kA_2)/(A_3 + kA_4)$ with respect to the parameter k is given by $S_k^T = k(A_2A_3 - A_1A_4)/(A_3 + kA_4)(A_1 + kA_2)$.

By definition, the sensitivity of T with respect to the parameter k is

$$S_k^T = \frac{d\ln T}{d\ln k} = \frac{dT}{dk} \cdot \frac{k}{T}$$

Now

$$\frac{dT}{dk} = \frac{A_2(A_3 + kA_4) - A_4(A_1 + kA_2)}{(A_3 + kA_4)^2} = \frac{A_2A_3 - A_1A_4}{(A_3 + kA_4)^2}$$

Thus

$$S_k^T = \frac{A_2A_3 - A_1A_4}{(A_3 + kA_4)^2} \cdot \frac{k(A_3 + kA_4)}{A_1 + kA_2} = \frac{k(A_2A_3 - A_1A_4)}{(A_3 + kA_4)(A_1 + kA_2)}$$

9.7. Consider the system of Example 9.6 with the addition of a load disturbance and a noise input as shown in Fig. 9-18. Show that the feedback controller improves the output sensitivity to the noise input and the load disturbance.

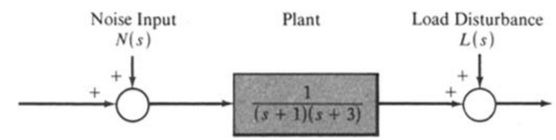


Fig. 9-18

For the open-loop system, the output due to the noise input and load disturbance is

$$C(s) = L(s) + \frac{1}{(s+1)(s+3)}N(s)$$

independent of the action of open-loop controller. For the closed-loop system,

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

For low frequencies the closed-loop system attenuates both the load disturbance and the noise input, compared to the open-loop system. In particular, the closed-loop system has steady state or d.c. gain:

$$C(0) = \frac{3}{5}L(0) + \frac{1}{5}N(0)$$

while the open-loop system has

$$C(0) = L(0) + \frac{1}{3}N(0)$$

At high frequencies these gains are approximately equal.

SYSTEM OUTPUT SENSITIVITY IN THE TIME DOMAIN

9.8. For the system defined by

$$\dot{\mathbf{x}} = A(\mathbf{p})\mathbf{x} + B(\mathbf{p})\mathbf{u}$$
$$\mathbf{y} = C(\mathbf{p})\mathbf{x}$$

show that the matrix of output sensitivities

$$\left[\frac{\partial y_i}{\partial p_j}\right]$$

is determined by solution of the differential equations

$$\dot{\mathbf{x}} = A\mathbf{x} + \mathbf{u} \tag{9.25}$$

$$\dot{V} = AV + \frac{\partial A}{\partial \mathbf{p}}\mathbf{x} + \frac{\partial B}{\partial \mathbf{p}}\mathbf{u} \tag{9.26}$$

with
$$\left[\frac{\partial y_i}{\partial p_j}\right] = CV + \frac{\partial C}{\partial \mathbf{p}}\mathbf{x} \tag{9.27}$$

where
$$V \equiv [v_{ij}] \equiv \frac{\partial \mathbf{x}}{\partial \mathbf{p}} \equiv \left[\frac{\partial x_i}{\partial p_j}\right]$$

that is, V is the matrix of sensitivity functions. The derivative of the sensitivity function v_{ij} is given by

$$\dot{v}_{ij} = \frac{d}{dt}\left(\frac{\partial x_i}{\partial p_j}\right)$$

Assuming the state variables have continuous derivatives, we can interchange the order of total and partial differentiation, so that

$$\dot{v}_{ij} = \frac{\partial}{\partial p_j}\left(\frac{dx_i}{dt}\right)$$

In matrix form,

$$\dot{V} = \frac{\partial \dot{\mathbf{x}}}{\partial \mathbf{p}} = \frac{\partial}{\partial \mathbf{p}}[A\mathbf{x} + B\mathbf{u}] = \frac{\partial A}{\partial \mathbf{p}}\mathbf{x} + A\frac{\partial \mathbf{x}}{\partial \mathbf{p}} + \frac{\partial B}{\partial \mathbf{p}}\mathbf{u}$$

Since $V = \partial \mathbf{x}/\partial \mathbf{p}$, we have

$$\dot{V} = AV + \frac{\partial A}{\partial \mathbf{p}}\mathbf{x} + \frac{\partial B}{\partial \mathbf{p}}\mathbf{u}$$

Then

$$\frac{\partial \mathbf{y}}{\partial \mathbf{p}} = \frac{\partial C\mathbf{x}}{\partial \mathbf{p}} = \frac{\partial C}{\partial \mathbf{p}}\mathbf{x} + C\frac{\partial \mathbf{x}}{\partial \mathbf{p}} = CV + \frac{\partial C}{\partial \mathbf{p}}\mathbf{x}$$

Note that, in the above equations, the partial derivative of a matrix with respect to the vector $\mathbf{p}$ is understood to generate a series of matrices, each one of which, when multiplied by $\mathbf{x}$, generates a column in the resulting matrix. That is, $(\partial A/\partial \mathbf{p})\mathbf{x}$ is a matrix with jth column $(\partial A/\partial p_j)\mathbf{x}$. This is easily verified by writing out all the scalar equations explicitly and differentiating term by term.

SYSTEMS CLASSIFICATION BY TYPE

9.9. The canonical feedback system is represented by Fig. 9-19.

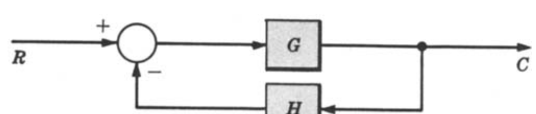


Fig. 9-19

Classify this system according to type if

(a) $G = \dfrac{1}{s}$ $\quad H = 1$

(b) $G = \dfrac{5}{s(s+3)}$ $\quad H = \dfrac{s+1}{s+2}$

(c) $G = \dfrac{2}{s^2+2s+5}$ $\quad H = s+5$

(d) $G = \dfrac{24}{(2s+1)(4s+1)}$ $\quad H = \dfrac{4}{4s(3s+1)}$

(e) $G = \dfrac{4}{s(s+3)}$ $\quad H = \dfrac{1}{s}$

(a) $GH = \dfrac{1}{s}$; *type 1*

(b) $GH = \dfrac{5(s+1)}{s(s+2)(s+3)}$; *type 1*

(c) $GH = \dfrac{2(s+5)}{s^2+2s+5}$; *type 0*

(d) $GH = \dfrac{96}{4s(2s+1)(3s+1)(4s+1)} = \dfrac{1}{s(s+\frac{1}{2})(s+\frac{1}{3})(s+\frac{1}{4})}$; *type 1*

(e) $GH = \dfrac{4}{s^2(s+3)}$; *type 2*

9.10. Classify the system given in Fig. 9-20 by type.

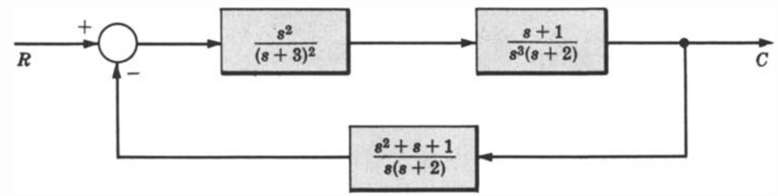


Fig. 9-20

The open-loop transfer function of this system is

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

Therefore it is a *type 2* system.

ERROR CONSTANTS AND STEADY STATE ERRORS

9.11. Show that the steady state error $e(\infty)$ of a stable type l unity feedback system when the input is a unit step function is related to the position error constant by

$$e(\infty) = \lim_{t\to\infty} e(t) = \frac{1}{1+K_p}$$

The error ratio (Definition 7.5) for a unity negative feedback system is given by Equation *(7.4)* with $H = 1$, that is, $E/R = 1/(1+G)$. For $R = 1/s$, $E = (1/s)(1/(1+G))$. From the Final Value Theorem, we obtain

$$e(\infty) = \lim_{s\to 0} sE(s) = \lim_{s\to 0}\left(\frac{s}{s[1+G(s)]}\right) = \frac{1}{1+\lim_{s\to 0}G(s)} = \frac{1}{1+K_p}$$

where we have used the definition $K_p \equiv \lim_{s\to 0} G(s)$.

9.12. Show that the steady state error $e(\infty)$ of a stable type l unity feedback system with a unit ramp function input is related to the velocity error constant by $e(\infty) = \lim_{t\to\infty} e(t) = 1/K_v$.

We have $E/R = 1/(1+G)$, and $E = (1/s^2)(1/(1+G))$ for $R = 1/s^2$. Since $G = KB_1(s)/s^l B_2(s)$ by Definition 9.4,

$$E = \frac{1}{s^2}\left[\frac{s^l B_2(s)}{s^l B_2(s) + KB_1(s)}\right]$$

For $l > 0$, we have

$$sE(s) = \frac{B_2(s)}{sB_2(s) + KB_1(s)/s^{l-1}}$$

where $l - 1 \geq 0$. Now we can use the Final Value Theorem, as was done in the previous problem, because the condition for the application of this theorem is satisfied. That is, for $l > 0$ we have

$$e(\infty) = \lim_{s\to 0} sE(s) = \begin{cases} 0 & \text{for } l > 1 \\ \dfrac{B_2(0)}{KB_l(0)} & \text{for } l = 1 \end{cases}$$

$B_1(0)$ and $B_2(0)$ are nonzero and finite by Definition 9.4; hence the limit exists (i.e., it is finite).

We cannot evoke the Final Value Theorem for the case $l = 0$ because

$$sE(s)|_{l=0} = \frac{1}{s}\left[\frac{B_2(s)}{B_2(s) + KB_1(s)}\right]$$

and the limit as $s \to 0$ of the quantity on the right does not exist. However, we may use the following argument for $l = 0$. Since the system is stable, $B_2(s) + KB_1(s) = 0$ has roots only in the left half-plane. Therefore E can be written with its denominator in the general factored form:

$$E = \frac{B_2(s)}{s^2 \prod_{i=1}^{r}(s+p_i)^{n_i}}$$

where $\mathrm{Re}(p_i) > 0$ and $\sum_{i=1}^{r} n_i = n - a$ (see Definition 9.4), that is, some roots may be repeated. Expanding E into partial fractions [Equation (*4.10a*)], we obtain

$$E = \frac{c_{20}}{s^2} + \frac{c_{10}}{s} + \sum_{i=1}^{r}\sum_{k=1}^{n_i} \frac{c_{ik}}{(s+p_i)^k}$$

where b_n in Equation (*4.10a*) is zero because the degree of the denominator is greater than that of the numerator ($m < n$). Inverting $E(s)$ (Section 4.8), we get

$$e(t) = c_{20}t + c_{10} + \sum_{i=1}^{r}\sum_{k=1}^{n_i} \frac{c_{ik}}{(k-1)!} t^{k-1} e^{-p_i t}$$

Since $\mathrm{Re}(p_i) > 0$ and c_{20} and c_{10} are finite nonzero constants (E is a rational algebraic expression), then

$$e(\infty) = \lim_{t\to\infty} e(t) = \lim_{t\to\infty}(c_{20}t) + c_{10} = \infty$$

Collecting results, we have

$$e(\infty) = \begin{cases} \infty & \text{for } l = 0 \\ \dfrac{B_2(0)}{KB_1(0)} & \text{for } l = 1 \\ 0 & \text{for } l > 1 \end{cases}$$

Equivalently,

$$\frac{1}{e(\infty)} = \begin{cases} 0 & \text{for } l = 0 \\ \dfrac{KB_1(0)}{B_2(0)} & \text{for } l = 1 \\ \infty & \text{for } l > 1 \end{cases}$$

These three values for $1/e(\infty)$ define K_v; thus

$$e(\infty) = \frac{1}{K_v}$$

9.13. For Fig. 9-21 find the position, velocity, and acceleration error constants.

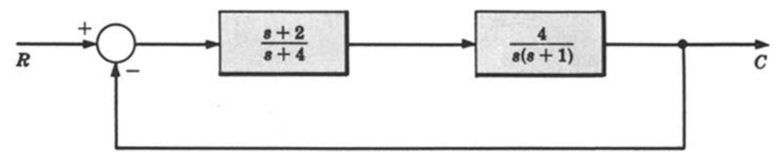


Fig. 9-21

Position error constant:

$$K_p = \lim_{s\to 0} G(s) = \lim_{s\to 0} \frac{4(s+2)}{s(s+1)(s+4)} = \infty$$

Velocity error constant:

$$K_v = \lim_{s\to 0} sG(s) = \lim_{s\to 0} \frac{4(s+2)}{(s+1)(s+4)} = 2$$

Acceleration error constant:

$$K_a = \lim_{s\to 0} s^2 G(s) = \lim_{s\to 0} \frac{4s(s+2)}{(s+1)(s+4)} = 0$$

9.14. For the system in Problem 9.13, find the steady state error for (a) a unit step input, (b) a unit ramp input, (c) a unit parabolic input.

(a) The steady state error for a unit step input is given by $e(\infty) = 1/(1+K_p)$. Using the result of Problem 9.13 yields $e(\infty) = 1/(1+\infty) = 0$.

(b) The steady state error for a unit ramp input is given by $e(\infty) = 1/K_v$. Again using the result of Problem 9.13, we get $e(\infty) = \frac{1}{2}$.

(c) The steady state error for a unit parabolic input is given by $e(\infty) = 1/K_a$. Then $e(\infty) = 1/0 = \infty$.

9.15. Figure 9-22 approximately represents a differentiator. Its transfer function is $C/R = Ks/[s(\tau s+1) + K]$. Note that $\lim_{\tau\to 0\, K\to\infty} C/R = s$, that is, C/R is a pure differentiator in the limit. Find the step, ramp, and parabolic error constants for this system, where the ideal system T_d is assumed to be a differentiator.

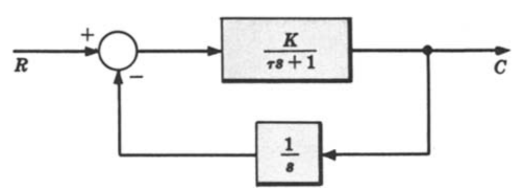


Fig. 9-22

Using the notation of Section 9.10, $T_d = s$ and $T_d - C/R = s^2(\tau s+1)/[s(\tau s+1)+K]$. Applying Definitions 9.8, 9.9, and 9.10 yields

$$K_s = \frac{1}{\lim_{s\to 0}\left[T_d - \frac{C}{R}\right]} = \frac{1}{\lim_{s\to 0}\left[\frac{s^2(\tau s+1)}{s(\tau s+1)+K}\right]} = \infty$$

$$K_r = \frac{1}{\lim_{s\to 0}\frac{1}{s}\left[T_d - \frac{C}{R}\right]} = \frac{1}{\lim_{s\to 0}\left[\frac{s(\tau s+1)}{s(\tau s+1)+K}\right]} = \infty$$

$$K_{pa} = \frac{1}{\lim_{s\to 0}\frac{1}{s^2}\left[T_d - \frac{C}{R}\right]} = \frac{1}{\lim_{s\to 0}\left[\frac{\tau s+1}{s(\tau s+1)+K}\right]} = K$$

9.16. Find the steady state value of the difference (error) between the outputs of a pure differentiator and the approximate differentiator of the previous problem for (a) a unit step input, (b) a unit ramp input, (c) a unit parabolic input.

From Problem 9.15, $K_s = \infty$, $K_r = \infty$, and $K_{pa} = K$.

(a) The steady state error for a unit step input is $e(\infty) = 1/K_s = 0$.

(b) The steady state error for a unit ramp input is $e(\infty) = 1/K_r = 0$.

(c) The steady state error for a unit parabolic input is $e(\infty) = 1/K_{pa} = 1/K$.

9.17. Given the stable type 2 unity feedback system shown in Fig. 9-23, find (a) the position, velocity, and acceleration error constants, (b) the steady state error when the input is $R = \frac{3}{s} - \frac{1}{s^2} + \frac{1}{2s^3}$.

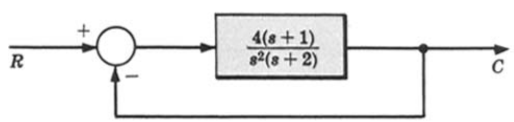


Fig. 9-23

(a) Using the last row of Table 9.1 (*type 2* systems), the error constants are $K_p = \infty$, $K_v = \infty$, $K_a = (4)(1)/2 = 2$.

(b) The steady state errors for unit step, unit ramp, and unit parabolic inputs are obtained from the same row of the table and are given by: $e_1(\infty) = 0$ for a unit step; $e_2(\infty) = 0$ for a unit ramp; $e_3(\infty) = \frac{1}{2}$ for a unit parabola.

Since the system is linear, the errors can be superimposed. Thus the steady state error when the input is $R = \frac{3}{s} - \frac{1}{s^2} + \frac{1}{2s^3}$ is given by $e(\infty) = 3e_1(\infty) - e_2(\infty) + \frac{1}{2}e_3(\infty) = \frac{1}{4}$.

Supplementary Problems

9.18. Prove the validity of Equation (*9.17*). (*Hint:* See Problems 9.11 and 9.12.)

9.19. Prove the validity of Equation (*9.19*). (*Hint:* See Problems 9.11 and 9.12.)

9.20. Prove the validity of Equation (*9.21*). (*Hint:* See Problems 9.11 and 9.12.)

9.21. Determine the sensitivity of the system in Problem 7.9, to variations in each of the parameters K_1, K_2 and p individually.

9.22. Generate an expression, in terms of the sensitivities determined in Problem 9.21, which relates the total variation in the transfer function of the system in Problem 7.9 to variations in K_1, K_2, and p.

9.23. Show that the steady state error $e(\infty)$ of a stable type l unity feedback system with a unit parabolic input is related to the acceleration error constant by $e(\infty) = \lim_{t\to\infty} e(t) = 1/K_a$. (*Hint:* See Problem 9.12.)

9.24. Verify Equations (*9.26*) and (*9.27*) by performing all differentiations on the full set of scalar simultaneous differential equations making up Equation (*9.25*).

Answers to Some Supplementary Problems

9.21. $S_{K_1}^{C/R} = \dfrac{s+p}{s+p-K_1K_2}$ $\quad S_{K_2}^{C/R} = \dfrac{K_1K_2}{s+p-K_1K_2}$ $\quad S_p^{C/R} = \dfrac{-p}{s+p-K_1K_2}$

9.22. $\Delta\dfrac{C}{R} = \dfrac{(s+p)\Delta K_1 + (K_1K_2)\Delta K_2 - p\Delta p}{s+p-K_1K_2}$