إلى قارئ هذا الكتاب ، تحية طيبة وبعد ...

لقد أصبحنا نعيش في عالم يعج بالأبحاث والكتب والمعلومات، وأصبح العلم معياراً حقيقياً لتفاضل الأمم والدول والمؤسسات والأشخاص على حدٍّ سواء، وقد أمسى بدوره حلاً شبه وحيدٍ لأكثر مشاكل العالم حدة وخطورة، فالبيئة تبحث عن حلول، وصحة الإنسان تبحث عن حلول، والموارد التي تشكل حاجة أساسية للإنسان تبحث عن حلول كذلك، والطاقة والغذاء والماء جميعها تحديات يقف العلم في وجهها الآن ويحاول أن يجد الحلول لها. فأين نحن من هذا العلم ؟ وأين هو منا؟

نسعى في موقع عالم الإلكترون www.4electron.com لأن نوفر بين أيدي كل من حمل على عاتقه مسيرة درب تملؤه التحديات ما نستطيع من أدوات تساعده في هذا الدرب، من مواضيع علمية، ومراجع أجنبية بأحدث إصداراتها، وساحات لتبادل الآراء والأفكار العلمية والمرتبطة بحياتنا الهندسية، وشروحٍ لأهم برمجيات الحاسب التي تتداخل مع تطبيقات الحياة الأكاديمية والعملية، ولكننا نتوقع في نفس الوقت أن نجد بين الطلاب والمهندسين والباحثين من يسعى مثلنا لتحقيق النفع والفائدة للجميع، ويحلم أن يكون عضواً في مجتمعٍ يساهم بتحقيق بيئة خصبة للمواهب والإبداعات والتألق، فهل تحلم بذلك ؟

حاول أن تساهم بفكرة، بومضة من خواطر تفكيرك العلمي، بفائدة رأيتها في إحدى المواضيع العلمية، بجانب مضيء لمحته خلف ثنايا مفهوم هندسي ما. تأكد بأنك ستلتمس الفائدة في كل خطوة تخطوها، وترى غيرك يخطوها معك ...

أخي القارئ، نرجو أن يكون هذا الكتاب مقدمة لمشاركتك في عالمنا العلمي التعاوني، وسيكون موقعكم عالم الإلكترون ww.4electron.com بكل الإمكانيات المتوفرة لديه جاهزاً على الدوام لأن يحقق البيئة والواقع الذي يبحث عنه كل باحث أو طالب في علوم الهندسة، ويسعى فيه للإفادة كل ساعٍ ، فأهلاً وسهلاً بكم .

مع تحيات إدارة الموقع وفريق عمله

www.4electron.com

LINEAR CONTROL SYSTEM ANALYSIS AND DESIGN WITH MATLAB

Fifth Edition, Revised and Expanded

John J. D'Azzo and Constantine H. Houpis
Air Force Institute of Technology
Wright-Patterson Air Force Base, Ohio, U.S.A.

Stuart N. Sheldon
U.S. Nuclear Regulatory Commission
Lisle, Illinois, U.S.A.

MARCEL DEKKER, INC. NEW YORK • BASEL

The fourth edition was published as *Linear Control System Analysis and Design: Conventional and Modern*, by John J. D'Azzo and Constantine H. Houpis (McGraw-Hill, 1995).

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress.

ISBN: 0-8247-4038-6

This book is printed on acid-free paper.

Headquarters
Marcel Dekker, Inc., 270 Madison Avenue, New York, NY 10016, U.S.A.
tel: 212-696-9000; fax: 212-685-4540

Distribution and Customer Service
Marcel Dekker, Inc., Cimarron Road, Monticello, New York 12701, U.S.A.
tel: 800-228-1160; fax: 845-796-1772

Eastern Hemisphere Distribution
Marcel Dekker AG, Hutgasse 4, Postfach 812, CH-4001 Basel, Switzerland
tel: 41-61-260-6300; fax: 41-61-260-6333

World Wide Web
http://www.dekker.com

The publisher offers discounts on this book when ordered in bulk quantities. For more information, write to Special Sales/Professional Marketing at the headquarters address above.

Current printing (last digit):

10 9 8 7 6 5 4 3 2 1

PRINTED IN UNITED STATES OF AMERICA

CONTROL ENGINEERING

A Series of Reference Books and Textbooks

1. Nonlinear Control of Electric Machinery, *Darren M. Dawson, Jun Hu, and Timothy C. Burg*
2. Computational Intelligence in Control Engineering, *Robert E. King*
3. Quantitative Feedback Theory: Fundamentals and Applications, *Constantine H. Houpis and Steven J. Rasmussen*
4. Self-Learning Control of Finite Markov Chains, *A. S. Poznyak, K. Najim, and E. Gómez-Ramírez*
5. Robust Control and Filtering for Time-Delay Systems, *Magdi S. Mahmoud*
6. Classical Feedback Control: With MATLAB, *Boris J. Lurie and Paul J. Enright*
7. Optimal Control of Singularly Perturbed Linear Systems and Applications: High-Accuracy Techniques, *Zoran Gajić and Myo-Taeg Lim*
8. Engineering System Dynamics: A Unified Graph-Centered Approach, *Forbes T. Brown*
9. Advanced Process Identification and Control, *Enso Ikonen and Kaddour Najim*
10. Modern Control Engineering, *P. N. Paraskevopoulos*
11. Sliding Mode Control in Engineering, *edited by Wilfrid Perruquetti and Jean Pierre Barbot*
12. Actuator Saturation Control, *edited by Vikram Kapila and Karolos M. Grigoriadis*
13. Nonlinear Control Systems, *Zoran Vukić, Ljubomir Kuljača, Dali Đonlagić, Sejid Tešnjak*

14. Linear Control System Analysis and Design with MATLAB: Fifth Edition, Revised and Expanded, *John J. D'Azzo, Constantine H. Houpis, and Stuart N. Sheldon*

Additional Volumes in Preparation

Robot Manipulator Control: Theory and Practice, Second Edition, Revised and Expanded, *Frank L. Lewis, Darren M. Dawson, and Chaouki T. Abdallah*

Robust Control System Design: Advanced State Space Techniques, Second Edition, Revised and Expanded, *Chia-Chi Tsui*

Series Introduction

Many textbooks have been written on control engineering, describing new techniques for controlling systems, or new and better ways of mathematically formulating existing methods to solve the ever-increasing complex problems faced by practicing engineers. However, few of these books fully address the applications aspects of control engineering. It is the intention of this new series to redress this situation.

The series will stress applications issues, and not just the mathematics of control engineering. It will provide texts that present not only both new and well-established techniques, but also detailed examples of the application of these methods to the solution of real-world problems. The authors will be drawn from both the academic world and the relevant applications sectors.

There are already many exciting examples of the application of control techniques in the established fields of electrical, mechanical (including aerospace), and chemical engineering. We have only to look around in today's highly automated society to see the use of advanced robotics techniques in the manufacturing industries; the use of automated control and navigation systems in air and surface transport systems; the increasing use of intelligent control systems in the many artifacts available to the domestic consumer market; and the reliable supply of water, gas, and electrical power to the domestic consumer and to industry. However, there are currently many challenging problems that could benefit from wider exposure to the applicability of control methodologies, and the systematic systems-oriented basis inherent in the application of control techniques.

This series presents books that draw on expertise from both the academic world and the applications domains, and will be useful not only as academically recommended course texts but also as handbooks for practitioners in many applications domains. *Linear Control System Analysis and Design with MATLAB* is another outstanding entry in Dekker's Control Engineering series.

Neil Munro

Preface

The countless technological advances of the twentieth century require that future engineering education emphasize *bridging the gap* between theory and the real world. This edition has been prepared with particular attention to the needs of undergraduates, especially those who seek a solid foundation in control theory as well as an ability to bridge the gap between control theory and its real-world applications. To help the reader achieve this goal, computer-aided design accuracy checks (CADAC) are used throughout the text to encourage good habits of computer literacy. Each CADAC uses fundamental concepts to ensure the viability of a computer solution.

This edition has been enhanced as a solid undergraduate and first-year graduate text; it emphasizes applying control theory fundamentals to both analog and sampled-data single-input single-output (SISO) feedback control systems. At the same time, the coverage of digital control systems is greatly expanded. Extensive reference is made to computer-aided design (CAD) packages to simplify the design process. The result is a comprehensive presentation of control theory and design—one that has been thoroughly class-tested, ensuring its value for classroom and self-study use.

This book features extensive use of explanations, diagrams, calculations, tables, and symbols. Such mathematical rigor is necessary for design applications and advanced control work. A solid foundation is built on concepts of modern control theory as well as those elements of conventional control theory that are relevant in analysis and design of control systems. The presentation of various techniques helps the reader understand what A. T. Fuller has called "the enigmatic control system." To provide a coherent development of the subject, we eschew formal proofs and lemmas, instead using an organization that draws the perceptive student steadily and surely to the demanding theory of multivariable control systems. Design examples are included throughout each chapter to reinforce the student's understanding of the material. A student who has reached this point is fully equipped to undertake the challenges of more advanced control theories, as presented in advanced control theory textbooks.

Chapter 2 sets forth the appropriate differential equations to describe the performance of physical systems, networks, and devices. The block diagram, the transfer function, and the state space (the essential concept of modern control theory) are also introduced. The approach used for the state space is the simultaneous derivation of the state-vector differential equation with the SISO differential equation for a chosen physical system. The chapter also shows how to derive the mathematical description of a physical system using LaGrange equations.

Chapter 3 presents the classical method of solving differential equations. Once the state-variable equation has been introduced, a careful explanation of its solution is provided. The relationship of the transfer function to the state equation of the system is presented in Chapter 14. The importance of the state transition matrix is described, and the state transition equation is derived. The idea of eigenvalues is explained next; this theory is used with the Cayley–Hamilton and Sylvester theorems to evaluate the state transition matrix.

The early part of Chapter 4 presents a comprehensive description of Laplace transform methods and pole-zero maps. Some further aspects of matrix algebra are introduced as background for solving the state equation using Laplace transforms. Finally, the evaluation of transfer matrices is clearly explained.

Chapter 5 begins with system representation by the conventional block-diagram approach. This is followed by a discussion of simulation diagrams and the determination of the state transition equation using signal flow graphs. The chapter also explains how to derive parallel state diagrams from system transfer functions, establishing the advantages of having the state equation in uncoupled form.

Chapter 6 introduces basic feedback system characteristics. This includes the relationship between system type and the ability of the system to follow or track polynomial inputs.

Chapter 7 presents the details of the root-locus method. Chapters 8 and 9 describe the frequency-response method using both log and polar plots. These chapters address the following topics: the Nyquist stability criterion; the correlation between the s-plane, frequency domain, and time domain; and gain setting to achieve a desired output response peak value while tracking polynomial command inputs. Chapters 10 and 11 describe the methods for improving system performance, including examples of the techniques for applying cascade and feedback compensators. Both the root-locus and frequency-response methods of designing compensators are covered.

Chapter 12 develops the concept of modeling a desired control ratio with figures of merit to satisfy system performance specifications. The system inputs generally fall into two categories: (1) desired input that the system output is to track (a tracking system) and (2) an external disturbance input for which the system output is to be minimal (a disturbance-rejection system). For both types of systems, the desired control ratio is synthesized by the proper placement of its poles and inclusion of zeros, if required. Chapter 12 also introduces the Guillemin-Truxal design procedure, which is used for designing a tracking control system and a design procedure emphasizing disturbance rejection.

Chapter 13 explains how to achieve desired system characteristics using complete state-variable feedback. Two important concepts of modern control theory—controllability and observability—are treated in a simple and straightforward manner.

Chapter 14 presents the sensitivity concepts of Bode, as used in variation of system parameters. Other tools include the method of using feedback transfer functions to form estimates of inaccessible states for use in state feedback, and a technique for linearizing a nonlinear system about its equilibrium points.

Chapter 15 presents the fundamentals of sampled data (S-D) control systems. Chapter 16 describes the design of digital control systems, demonstrating, for example, the effectiveness of digital compensation. The concept of a pseudo-continuous-time (PCT) model of a digital system permits the use of continuous-time methods for the design of digital control systems.

The text has been prepared so that it can be used for self-study by engineers in various areas of practice (electrical, aeronautical, mechanical, etc.). To make it valuable to all engineers, we use various examples of feedback control systems and unify the treatment of physical control systems by using mathematical and block-diagram models common to all.

There are many computer-aided design (CAD) packages (e.g., MATLAB® [see App. C], Simulink, and TOTAL-PC) available to help students and practicing engineers analyze, design, and simulate control systems. The use of MATLAB is emphasized throughout the book, and many MATLAB m-files are presented as examples.

We thank the students who have used this book in its previous editions and the instructors who have reviewed this edition for their helpful comments and recommendations. We thank especially Dr. R. E. Fontana, Professor Emeritus of Electrical Engineering, Air Force Institute of Technology, for the encouragement he provided for the previous editions. This edition is dedicated to the memory of Dr. T. J. Higgins, Professor Emeritus of Electrical Engineering, University of Wisconsin, for his thorough review of the earlier manuscripts.

We also express our appreciation to Professor Emeritus Donald McLean of the University of Southampton, England, formerly a visiting professor at the Air Force Institute of Technology. Our association with him has been an enlightening and refreshing experience. The personal relationship with him has been a source of inspiration and deep respect.

John J. D'Azzo
Constantine H. Houpis
Stuart N. Sheldon

Contents

1

Introduction

1.1 INTRODUCTION

The technological explosion of the twentieth century, which was accelerated by the advent of computers and control systems, has resulted in tremendous advances in the field of science. Thus, automatic control systems and computers permeate life in all advanced societies today. These systems and computers have acted and are acting as catalysts in promoting progress and development, propelling society into the twenty-first century. Technological developments have made possible high-speed bullet trains; exotic vehicles capable of exploration of other planets and outer space; the establishment of the Alpha space station; safe, comfortable, and efficient automobiles; sophisticated civilian and military [manual and uninhabited (see Fig. 1.1)] aircraft; efficient robotic assembly lines; and efficient environmentally friendly pollution controls for factories. The successful operation of all of these systems depends on the proper functioning of the large number of control systems used in such ventures.

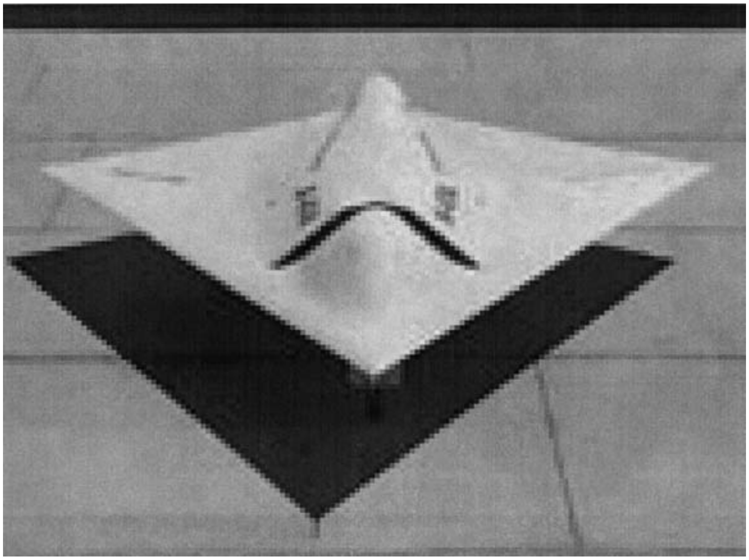


FIGURE 1.1 An unmanned aircraft.

1.2 INTRODUCTION TO CONTROL SYSTEMS

Classical Examples

The toaster in Fig. 1.2*a* can be set for the desired darkness of the toasted bread. The setting of the "darkness" knob, or timer, represents the input quantity, and the degree of darkness and crispness of the toast produced is the output quantity. If the degree of darkness is not satisfactory, because of the condition of the bread or some similar reason, this condition can in no way automatically alter the length of time that heat is applied. Since the output quantity has no influence on the input quantity, there is no *feedback* in this system. The heater portion of the toaster represents the dynamic part of the overall system, and the timer unit is the *reference selector.*

The dc shunt motor of Fig. 1.2*b* is another example. For a given value of field current, a required value of voltage is applied to the armature to produce the desired value of motor speed. In this case the motor is the dynamic part of the system, the applied armature voltage is the input quantity, and the speed of the shaft is the output quantity. A variation of the speed from the desired value, due to a change of mechanical load on the shaft, can in no way cause a change in the value of the applied armature voltage to maintain the desired speed. Therefore, the output quantity has no influence on the input quantity.

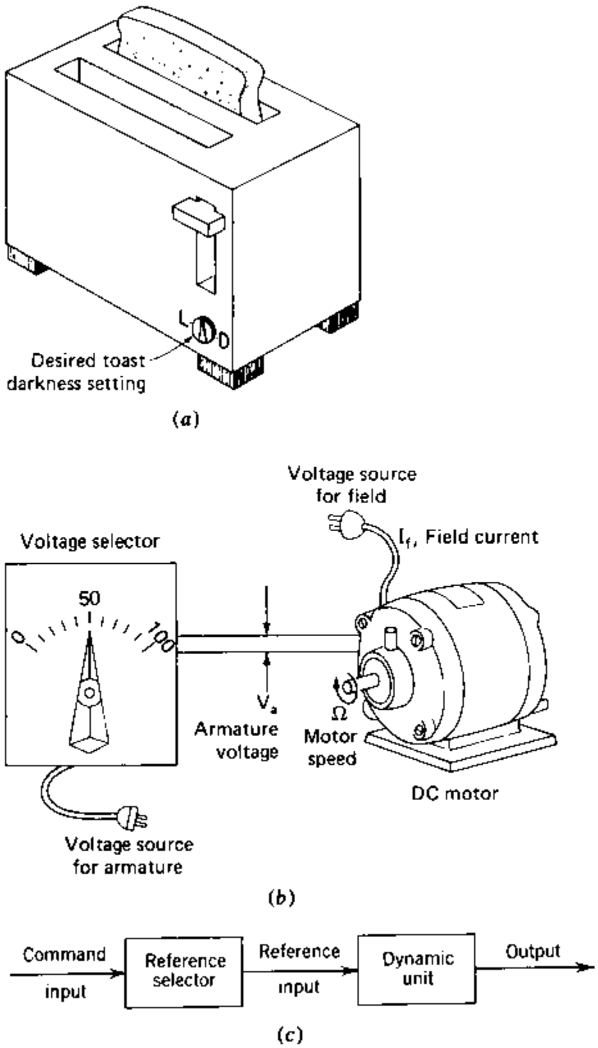


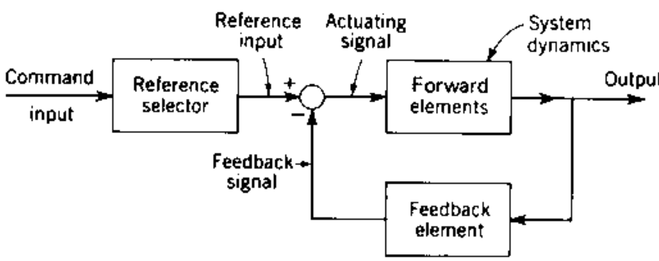
FIGURE 1.2 Open-loop control systems: (*a*) automatic toaster; (*b*) electric motor; (*c*) functional block diagram.

Systems in which the output quantity has no effect upon the input quantity are called *open-loop control systems.* The examples just cited are represented symbolically by a functional block diagram, as shown in Fig. 1.2*c.* In this figure, (1) the desired darkness of the toast or the desired speed of the motor is the command input, (2) the selection of the value of time on the toaster timer or the value of voltage applied to the motor armature is represented by the reference-selector block, and (3) the output of this block is identified as the reference input. The reference input is applied to the dynamic unit that performs the desired control function, and the output of this block is the desired output.

A person could be assigned the task of sensing the actual value of the output and comparing it with the command input. If the output does not have the desired value, the person can alter the reference-selector position to achieve this value. Introducing the person provides a means through which the output is *fed back* and is compared with the input. Any necessary change is then made in order to cause the output to equal the desired value. The *feedback* action therefore controls the input to the dynamic unit. Systems in which the output has a direct effect upon the input quantity are called *closed-loop control systems.*

To improve the performance of the closed-loop system so that the output quantity is as close as possible to the desired quantity, the person can be replaced by a mechanical, electrical, or other form of a comparison unit. The functional block diagram of a *single-input single-output* (SISO) closed-loop control system is illustrated in Fig. 1.3. Comparison between the reference input and the feedback signals results in an *actuating signal* that is the difference between these two quantities. The actuating signal acts to maintain the output at the desired value. This system is called a *closed-loop control system*. The designation *closed-loop* implies the action resulting from the comparison between the output and input quantities in order to maintain the output at the desired value. Thus, the output is controlled in order to achieve the desired value.

Examples of closed-loop control systems are illustrated in Figs. 1.4 and 1.5. In a home heating system the desired room temperature (command input)



loop control system.

FIGURE 1.3 Functional block diagram of a closed-loop system.

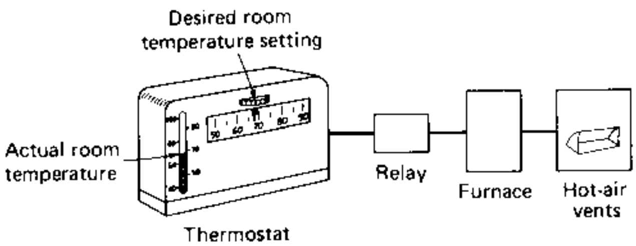


FIGURE 1.4 Home heating control system.

is set on the thermostat in Fig. 1.4. (reference selector). A bimetallic coil in the thermostat is affected by both the actual room temperature (output) and the reference-selector setting. If the room temperature is lower than the desired temperature, the coil strip alters its shape and causes a mercury switch to operate a relay, which turns on the furnace to produce heat in the room. When the room temperature [1] reaches the desired temperature, the shape of the coil strip is again altered so that the mercury switch opens. This deactivates the relay and in turn shuts off the furnace. In this example, the bimetallic coil performs the function of a comparator since the output (room temperature) is fed back directly to the comparator. The switch, relay, and furnace are the dynamic elements of this closed-loop control system.

A closed-loop control system of great importance to all multistory buildings is the automatic elevator of Fig. 1.5. A person in the elevator presses the button corresponding to the desired floor. This produces an actuating signal that indicates the desired floor and turns on the motor that raises or lowers the elevator. As the elevator approaches the desired floor, the actuating signal decreases in value and, with the proper switching sequences, the elevator stops at the desired floor and the actuating signal is reset to zero. The closed-loop control system for the express elevator in the Sears Tower building in Chicago is designed so that it ascends or descends the 103 floors in just under 1 min with maximum passenger comfort.

Modern Examples

The examples in this section represent complex closed-loop control systems that are at the forefront of the application of control theory to the control system challenges of the twenty-first century.

The ultimate objective in robotic arm control research [2]* is to provide human arm emulation. Payload invariance is a necessary component of

*References are indicated by numbers in brackets and are found at the end of the chapter.

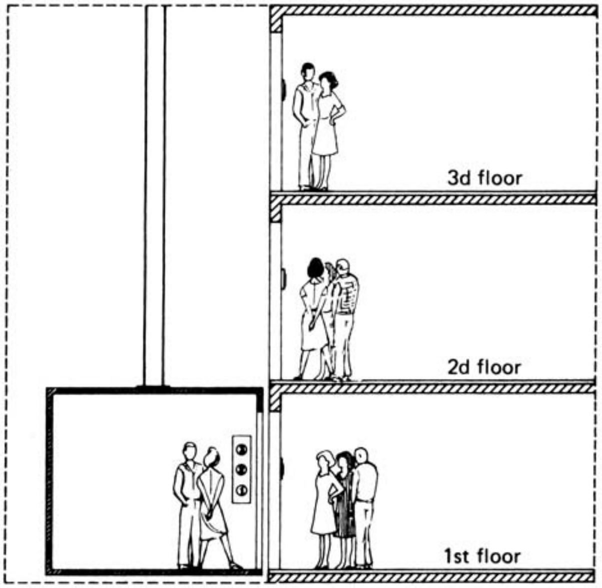


FIGURE 1.5 Automatic elevator.

human arm emulation. Model-based controllers require accurate knowledge of payload and drive system dynamics to provide good high-speed tracking accuracy. A robust multivariable control system design technique is required which solves the payload and dynamics uncertainty. Thus, the model-based quantitative feedback theory (QFT) design technique [3] is applied which results in controllers that are implemented by a series of simple backwards difference equations. QFT high-speed tracking accuracy was experimentally evaluated on the first three links of the PUMA-500 of Fig. 1.6. This robust design technique increased tracking accuracy by up to a factor of 4 over the model-based controller performance baseline. The QFT tracking performance is robust for both unmodeled drive system dynamics and payload uncertainty. The nonheuristic nature of the QFT design and tuning should allow application to a wide range of manipulators.

The interest in improving the fuel efficiency of automobiles has spurred the improvement of the idle speed control for the automotive fuel-injected engine [4,5]. The following is the abstract from the paper entitled "Robust Controller Design and Experimental Verification of I.C. Engine Speed Control" by G.K. Hamilton and M.A. Franchek, School of Mechanical Engineering, Purdue University [4].

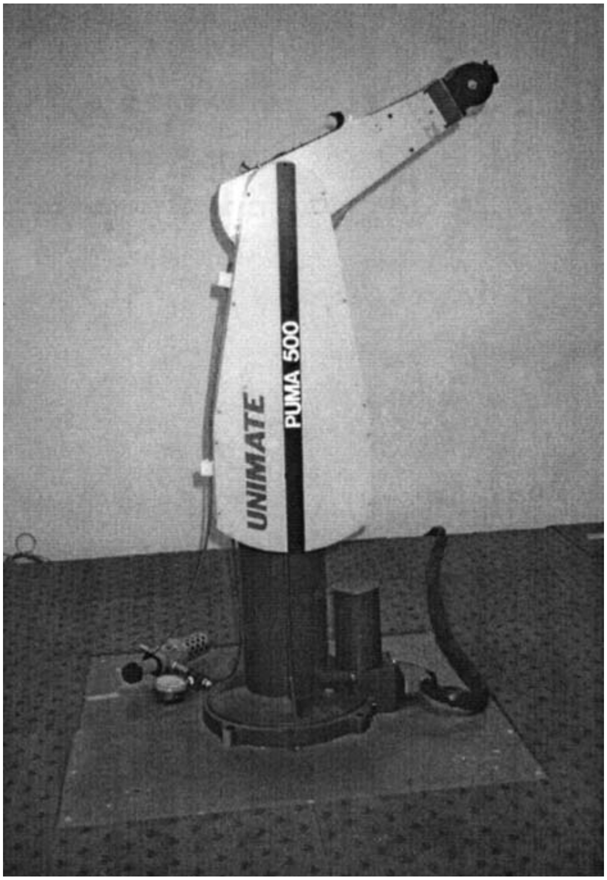


FIGURE 1.6 Robot arm (From Ref. 2).

Presented in this paper is the robust idle speed control of a Ford 4.6L V-8 fuel injected engine. The goal of this investigation is to design a robust feedback controller that maintains the idle speed within a *150 rpm* tolerance of about *600 rpm* despite a *20 Nm* step torque disturbance delivered by the power steering pump. The controlled input is the by-pass air valve which is subjected to an output saturation constraint. Issues complicating the controller design include the nonlinear nature of the engine dynamics, the induction-to-power delay of the manifold filling dynamics, and the saturation constraint of the by-pass air valve. An experimental verification of the proposed controller, utilizing the nonlinear plant, is included.

The desired performance has been demonstrated on the laboratory test setup shown in Figure 1.7*a*. The authors show in their paper that they met all the design objectives and have achieved excellent results.

Shown in Figure 1.7*b* is the testing and simulation setup of a mass air flow (MAF) sensor diagnostics for adaptive fueling control of internal combustion engines performed at the Purdue Engine Research Facility/Engine Control Technology, Purdue University, by Professor M.A. Franchek and his associates [6]. An information synthesis solution is attractive for diagnostics since the algorithm automatically calibrates itself, reduces the number of false detections, and compresses a large amount of engine health information into the model coefficients. There are three primary parts to information synthesis diagnostics. First, an IS model is used to predict the MAF sensor output based on the engine operating condition. The inputs to this IS model include the throttle position sensor (TPS) and the engine speed sensor information. The second part concerns an adaptation process that is used to reduce the errors between the IS model output and the actual MAF sensor output. Finally, the adapted model coefficients are used to diagnose the sensor as well as identify the source for changes in the sensor characteristics. This proposed solution is experimentally tested and validated on a Ford 4.6 L

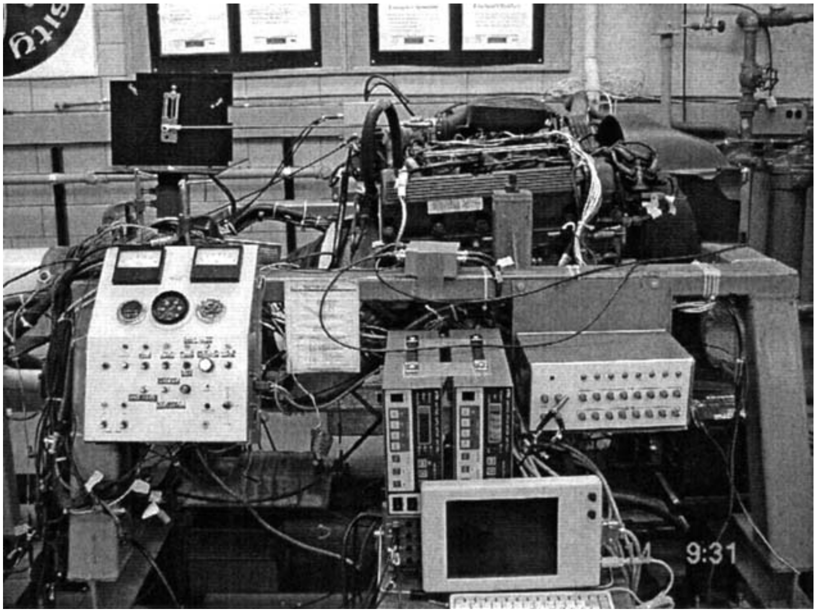


FIGURE 1.7*a* Fuel injection engine.

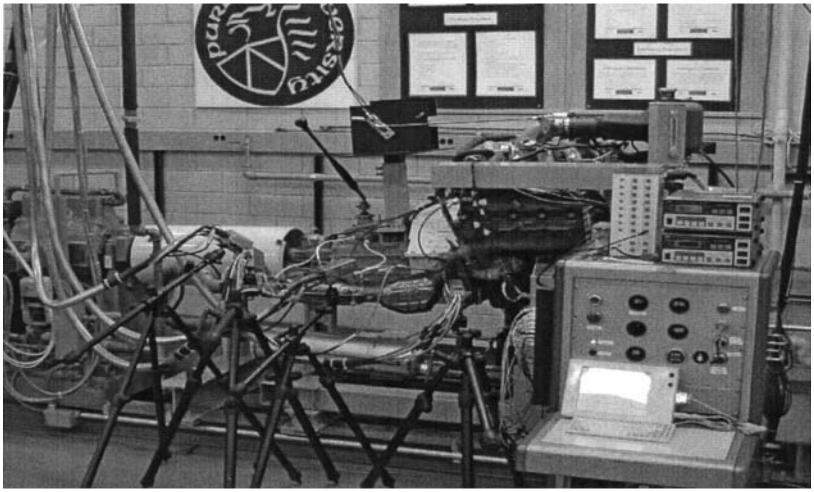


FIGURE 1.7*b* Testing and simulation setup of a mass air flow sensor diagnostics for internal combustion engines.

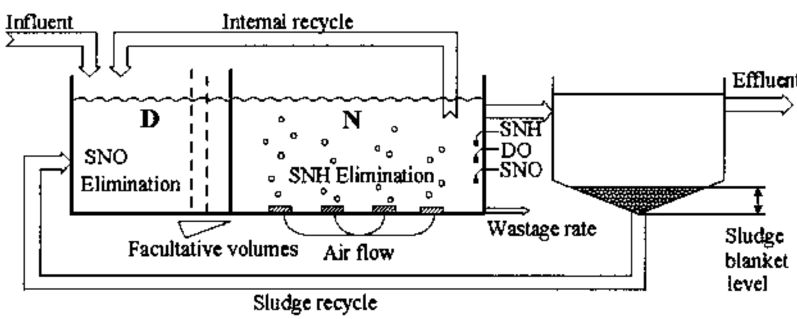


FIGURE 1.8 Wastewater treatment plant.

V-8 fuel injected engine. The specific MAF sensor faults to be identified include sensor bias and a leak in the intake manifold.

One of the most important objectives of a wastewater treatment plant (WWTP) [7], shown in Fig. 1.8, is to protect the water environment from negative effects produced by residual water, controlling the maximum concentration of pernicious substances. A computer simulation of the QFT-designed WWTP-compensated control system met the desired performance specifications. The control system design resulted in an improved performance of the plant because the concentration levels obtained are nearer to those required by environmental law, and a notable reduction in the running costs is produced. Thus, the operation of the plant is notably more efficient. The controller developed is also suitable for low-cost microcomputer implementation.

Design methods for analog SISO control systems shown in Fig. 1.3 are covered in Chaps. 6 to 16. Some systems require a precision in their performance that cannot be achieved by the structure of Fig. 1.3. Also, systems exist for which there are multiple inputs and/or multiple outputs. They are discussed in References 3 and 8. The design methods for such systems are often based on a representation of the system in terms of *state variables.* For example, position, velocity, and acceleration may represent the state variables of a position control system. The definition of state variables and their use in representing systems are contained in Chaps. 2, 3, and 5. The use of state-variable methods for the design of control systems is presented in Chaps. 13 and 14. The design methods presented in Chaps. 7 to 16 require knowledge of a fixed mathematical model of the system that is being controlled. The parameters of some systems change because of the range of conditions under which they operate. The quantitative feedback theory is a design technique for nonlinear plants that contain structured parametric uncertainty [3]. Using QFT, the parameter variations and performance specifications are included at the onset of the design process. The use of a digital computer to assist the engineer in the design process is emphasized throughout this book, and an available computer-aided design (CAD) package is given in Appendix C.

The design of the robust flight control system (FCS) for the VISTA F-16 of Fig. 1.9*b* was accomplished by an Air Force Institute of Technology

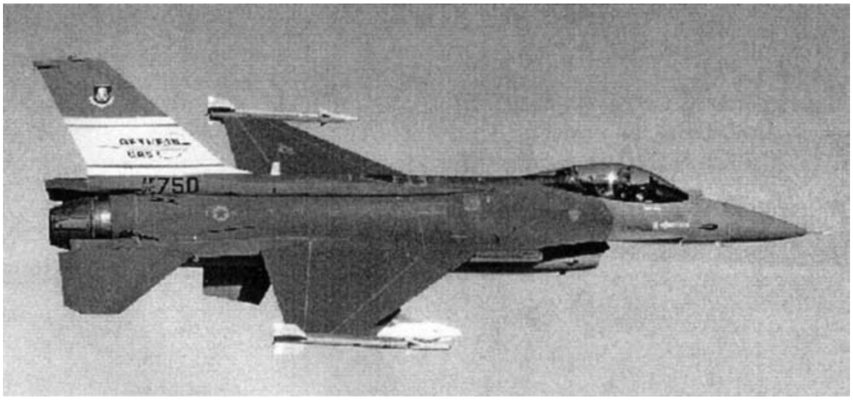


FIGURE 1.9 VISTA F-16.

student who is an F-16 pilot [9]. He was able to utilize his real-world knowledge of the aircraft and its handling qualities to achieve the desired robust FCS. Traditionally, flight control engineers have taken a conservative, brute force approach to designing a full envelope FCS for an aircraft. First, many design points, which for this design were points representing airspeed vs. altitude, within and along the border of the flight envelope plot were selected. Second, individual compensator designs were accomplished for each of these points. Third, smooth transitions between these compensators must be engineered. Making the transitions imperceptible to the pilot is very difficult and time consuming because each airspeed-altitude design point can be approached from an infinite number of initial conditions. Obviously, if the number of the design points can be reduced, thus reducing the number of transitions required, the design process can be made more efficient, and the resulting FCS less complex.

A way to reduce the number of necessary design points is to apply a robust control design technique to the problem. A compensator synthesized using robust control principles should be able to handle large parts of, if not the whole, flight envelope. Unfortunately, many previous attempts at applying robust control design algorithms to practical, real-world problems have been dismal failures [9]. Although the problem is well posed, the failure is due to the fact that the resulting compensator is impractical to implement. Either the compensator is of too high order, or its gain is too large to accommodate real-world nonlinearities. Also, any sensor noise present is accentuated by this gain. The typical reason for these poor results is that the robust design is synthesized in the essentially noiseless world of the digital computer, and then validated on the digital computer through the use of small signal, linear simulation.

A robust control design technique that overcomes the aforementioned pitfalls is the QFT design technique. Although a QFT design effort could very easily result in a compensator of high order and of high gain, it does give the designer complete control over the gain and the order of the compensator; hence, QFT is not constrained to produce an impractical compensator. In addition, if a decision is made to decrease or limit the order or gain of a compensator, the performance trade-offs due to this action can be clearly seen by the designer.

In summary, although excellent FCSs have been designed for aircraft using traditional design methods, the synthesis of those FCSs has been a costly, time-consuming endeavor. Thus, limiting robustness in FCS design results in a convoluted, complex, full envelope design. QFT offers the ability to incorporate enough robustness to simplify the design process and the resulting FCS, but not so much robustness that the resulting FCS is impractical to implement due to violation of physical limitations imposed by the "real-world" (i.e., actuator saturation or sensor noise amplification). Also, QFT has the feature of utilizing the control system designer's knowledge of the real-world characteristics of the plant, etc. during the ongoing design process in maximizing the ability to achieve the desired robust system performance. A simulation [10], involving the nonlinear plant was performed on the Lamars Simulator [11] by the FCS designer—an F-16 pilot. The excellent performance in these simulations demonstrated the viability of a QFT design approach in producing flight-worthy aircraft control systems. It illustrated the benefits of designing flight control systems with the QFT robust control system design technique in contrast to the brute force approach of optimizing a flight control system for performance in expected configurations and then scheduling the gains.

1.3 DEFINITIONS

From the preceding discussion the following definitions are evolved, based in part on the standards of the IEEE [1], and are used in this text.

System. A combination of components that act together to perform a function not possible with any of the individual parts. The word *system* as used herein is interpreted to include physical, biological, organizational, and other entities, and combinations thereof, which can be represented through a common mathematical symbolism. The formal name *systems engineering* can also be assigned to this definition of the word *system*. Thus, the study of feedback control systems is essentially a study of an important aspect of systems engineering and its application.

Command input. The motivating input signal to the system, which is independent of the output of the system and exercises complete control over it (if the system is completely controllable).

Reference selector (*reference input element*). The unit that establishes the value of the reference input. The reference selector is calibrated in terms of the desired value of the system output.

Reference input. The reference signal produced by the reference selector, i.e., the command expressed in a form directly usable by the system. It is the actual signal input to the control system.

Disturbance input. An external disturbance input signal to the system that has an unwanted effect on the system output.

Forward element (*system dynamics*). The unit that reacts to an actuating signal to produce a desired output. This unit does the work of controlling the output and thus may be a power amplifier.

Output (*controlled variable*). The quantity that must be maintained at a prescribed value, i.e., following the command input without responding the disturbance inputs.

Open-loop control system. A system in which the output has no effect upon the input signal.

Feedback element. The unit that provides the means for feeding back the output quantity, or a function of the output, in order to compare it with the reference input.

Actuating signal. The signal that is the difference between the reference input and the feedback signal. It is the input to the control unit that causes the output to have the desired value.

Closed-loop control system. A system in which the output has an effect upon the input quantity in such a manner as to maintain the desired output value.

The fundamental difference between the open- and closed-loop systems is the *feedback action*, which may be continuous or discontinuous. In one form of discontinuous control the input and output quantities are periodically sampled and discontinuous. Continuous control implies that the output is continuously fed back and compared with the reference input compared; i.e., the control action is discontinuous in time. This is commonly called a *digital*, *discrete-data* or *sampled-data* feedback control system. A discrete-data control system may incorporate a digital computer that improves the performance achievable by the system. In another form of discontinuous control system the actuating signal must reach a prescribed value before the system dynamics reacts to it; i.e., the control action is discontinuous in amplitude rather than in time. This type of discontinuous control system is commonly called an *on-off* or *relay* feedback control system. Both forms may be present in a system. In this text continuous control systems are considered in detail since they lend themselves readily to a basic understanding of feedback control systems. The fundamentals of sampled-data (S-D) control systems are given in Chap. 15. Digital control systems are introduced in Chap. 16.

With the above introductory material, it is proper to state a definition [1] of a feedback control system: "A control system that operates to achieve prescribed relationships between selected system variables by comparing functions of these variables and using the comparison to effect control." The following definitions are also used.

Servomechanism (often abbreviated as *servo*). The term is often used to refer to a mechanical system in which the steady-state error is zero for a constant input signal. Sometimes, by generalization, it is used to refer to any feedback control system.

Regulator. This term is used to refer to systems in which there is a constant steady-state output for a constant signal. The name is derived from the early speed and voltage controls, called speed and voltage regulators.

1.4 HISTORICAL BACKGROUND [12]

The action of steering an automobile to maintain a prescribed direction of movement satisfies the definition of a feedback control system. In Fig. 1.10, the prescribed direction is the reference input. The driver's eyes perform the function of comparing the actual direction of movement with the prescribed direction, the desired output. The eyes transmit a signal to the brain, which

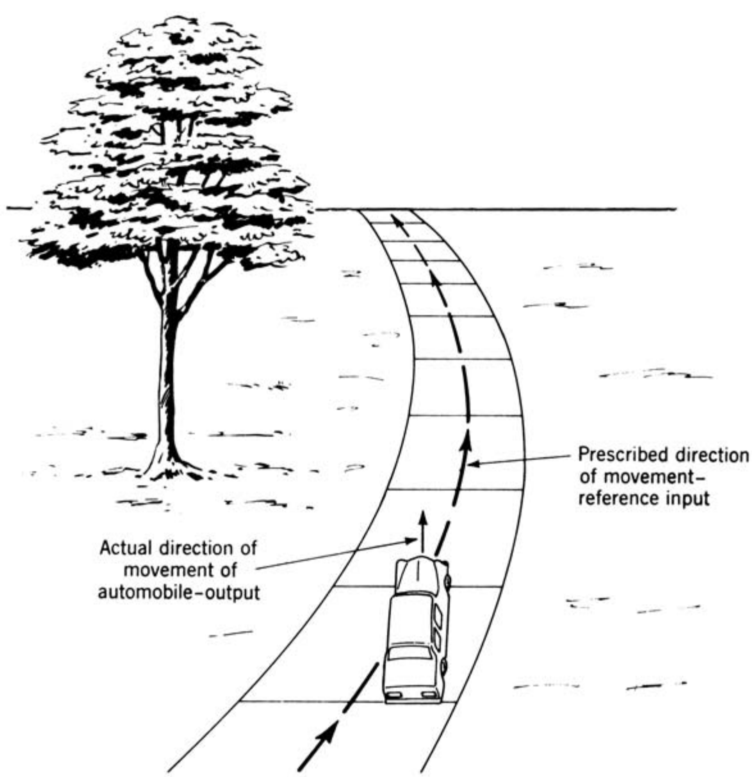


FIGURE 1.10 A pictorial demonstration of an automobile as a feedback control system.

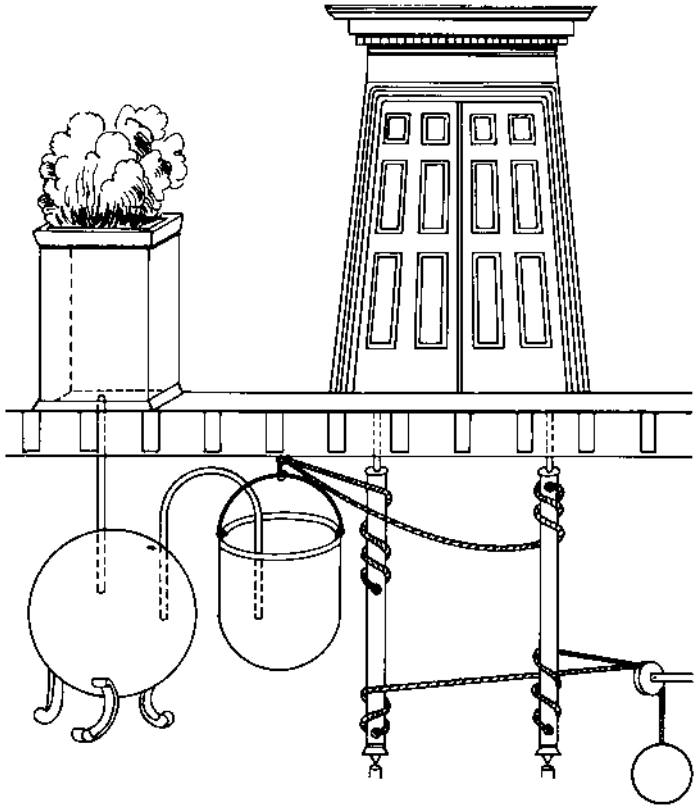


FIGURE 1.11 Hero's device for opening temple doors.

interprets this signal and transmits a signal to the arms to turn the steering wheel, adjusting the actual direction of movement to bring it in line with the desired direction. Thus, steering an automobile constitutes a feedback control system.

One of the earliest open-loop control systems was Hero's device for opening the doors of a temple. The command input to the system (see Fig. 1.11) was lighting a fire upon the altar. The expanding hot air under the fire drove the water from the container into the bucket. As the bucket became heavier, it descended and turned the door spindles by means of ropes, causing the counterweight to rise. The door could be closed by dousing the fire. As the air in the container cooled and the pressure was thereby reduced, the water from the bucket siphoned back into the storage container. Thus, the bucket became lighter and the counterweight, being heavier, moved down, thereby closing the door. This occurs as long as the bucket is higher than the container. The device was probably actuated when the ruler and his entourage started to ascend the temple steps. The system for opening the door was not visible or known to the masses. Thus, it created an air of mystery and demonstrated the power of the Olympian gods.

James Watt's flyball governor for controlling speed, developed in 1788, can be considered the first widely used automatic feedback control system. Maxwell, in 1868, made an analytic study of the stability of the flyball governor. This was followed by a more detailed solution of the stability of a third-order flyball governor in 1876 by the Russian engineer Wischnegradsky [13]. Minorsky made one of the earlier deliberate applications of nonlinear elements in closed-loop systems in his study of automatic ship steering about 1922 [14].

A significant date in the history of automatic feedback control systems is 1934, when Hazen's paper "Theory of Servomechanisms" was published in the *Journal of the Franklin Institute*, marking the beginning of the very intense interest in this new field. It was in this paper that the word *servomechanism* originated, from the words *servant* (or slave) and *mechanism*. Black's important paper on feedback amplifiers appeared [15] in the same year. After World War II, control theory was studied intensively and applications have proliferated. Many books and thousands of articles and technical papers have been written, and the application of control systems in the industrial and military fields has been extensive. This rapid growth of feedback control systems was accelerated by the equally rapid development and widespread use of computers.

An early military application of a feedback control system is the antiaircraft radar tracking control system shown in Fig. 1.12. The radar antenna detects the position and velocity of the target airplane, and the computer takes this information and determines the correct firing angle for the gun. This angle includes the necessary lead angle so that the shell reaches the projected position at the same time as the airplane. The output signal of the computer, which is a function of the firing angle, is fed into an amplifier that provides power for the drive motor. The motor then aims the gun at the necessary firing angle. A feedback signal proportional to the gun position ensures correct alignment with the position determined by the computer. Since the gun must be positioned both horizontally and vertically, this system has two drive motors, which are parts of two coordinated feedback loops.

The advent of the nuclear reactor was a milestone in the advancement of science and technology. For proper operation the power level of the reactor must be maintained at a desired value or must vary in a prescribed manner. This must be accomplished automatically with minimum human supervision. Figure 1.13 is a simplified block diagram of a feedback control system for controlling the power output level of a reactor. If the power output level differs

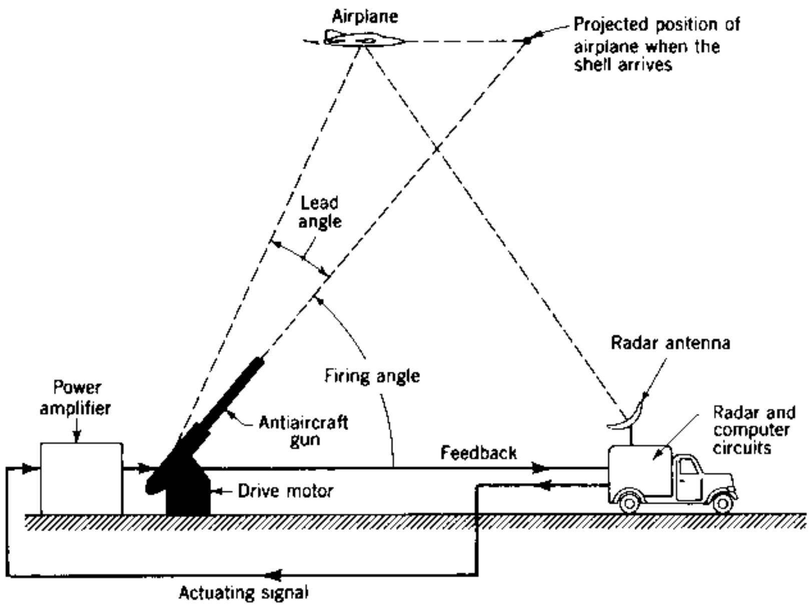


FIGURE 1.12 Antiaircraft radar-tracking control systems.

from the reference input value, the actuating signal produces a signal at the output of the control elements. This, in turn, moves the regulating rod in the proper direction to achieve the desired power level of the nuclear reactor. The position of the regulating rod determines the rate of nuclear fission and therefore the total power generated. This output nuclear power can be converted into steam power, for example, which is then used for generating electric energy.

The control theory developed through the late 1950s may be categorized as *conventional* control theory and is effectively applied to many control-design problems, especially to SISO systems. Since then, control theory has been developed for the design of more complicated systems and for *multiple-input multiple-output* (MIMO) systems. Space travel has become possible only

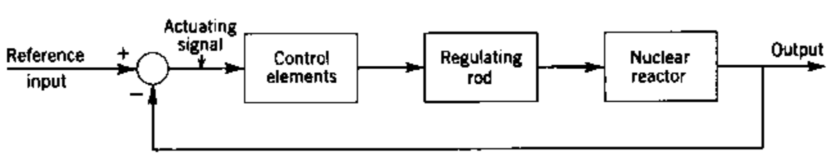


FIGURE 1.13 A feedback system for controlling the power level of a nuclear reactor.

because of the advent of modern control theory. Areas such as trajectory optimization and minimum-time and/or minimum-fuel problems, which are very important in space travel, can be readily handled by multivariable control theory. The introduction of microprocessors as control elements, i.e., performing control functions in contrast to being used solely as computational tools, has had an enormous impact on the design of feedback control systems which achieve desired control-system specifications.

The development of control concepts in the engineering field has been extended to the realm of human and biomedical engineering. The basic concept of feedback control is used extensively in the field of business management. The field of medicine is also one to which the principles of control systems and systems engineering are being applied extensively. Thus, standards of optimum performance are established in all areas of endeavor: the actual performance is compared with the desired standard, and any difference between the two is used to bring them into closer agreement.

1.5 DIGITAL CONTROL DEVELOPMENT [16]

The advances of the twentieth century have expedited the decrease in cost of digital hardware; thus economical digital control implementation is enabling the tremendous advances that will be made in the twenty-first century. Applications include process control, automatic aircraft stabilization and control, guidance and control of aerospace vehicles, aerospace vehicle management systems (VMS), uninhabited (unmmaned) aerospace vehicles such as the Global Hawk, and robotics. The development of digital control systems is illustrated by the following example of a digital flight control system.

Numerous changes have been made in aircraft flight control systems. Initially, flight control systems were purely mechanical, which was ideal for smaller, slow-speed, low-performance aircraft because they were easy to maintain. However, more control-surface force is required in modern high-performance airplanes. Thus, during the twentieth century a hydraulic power boost system was added to the mechanical control. This modification maintained the direct mechanical linkage between the pilot and the control surface. As aircraft became larger, faster, and heavier, and had increased performance, they became harder to control because the pilot could not provide the necessary power to directly operate the control surfaces. Thus, the entire effort of moving the control surface had to be provided by the actuator. A stability augmentation system (SAS) was added to the hydraulic boosted mechanical regulator system to make the aircraft flyable under all flight configurations. Motion sensors were used to detect aircraft perturbations and to provide electric signals to a SAS computer, which, in turn, calculated the proper amount of servo actuator force required. When

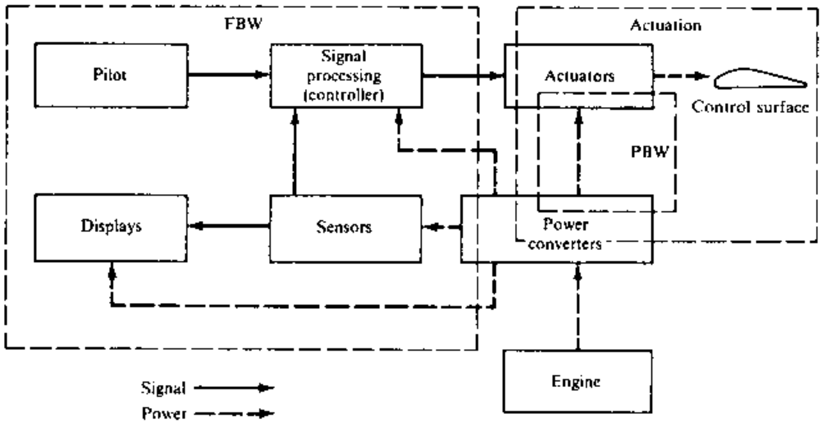


FIGURE 1.14 Fly-by-wire (FBW) and power-by-wire (PBW) control systems. (Control Systems Development Branch, Flight Dynamics Laboratory, Wright-Patterson AFB, Ohio.)

higher-authority SAS was required, both series- and parallel-pitch axis, dampers were installed. This so-called command augmentation system (CAS), which is no longer utilized, allowed greater flexibility in control because the parallel damper could provide full-authority travel without restricting the pilot's stick movements. Although planes were originally designed to be statically stable, longitudinally unstable aircraft are more agile. In these aircraft the flight control system provided the required stability.

The next step in the evolution of flight control systems was the use of a fly-by-wire (FBW) control system shown in Fig. 1.14. In this design, all pilot commands are transmitted to the control-surface actuators through electric wires. Thus, all mechanical linkages from the pilot's control stick to the servo actuators are removed from the aircraft. The FBW system provided the advantages of reduced weight, improved survivability, and decreased maintenance. However, the pilot is required to believe in and accept the increased survivability that is provided by using redundancy throughout the entire flight control system.

Originally the flight control computers were analog (such as the F-16 aircraft computers), but these have been replaced by digital computers. In addition, the controller consists of a digital computer which accepts the pilot commands and signals from the sensors (position and rate gyros) and accelerometers, and sends commands to the actuators. This is now referred to as a *digital flight control system* (DFCS). For twenty-first century aerospace vehicles the use of hydraulics has essentially been eliminated in favor of an all-electric system incorporating the use of digital computers. No longer do

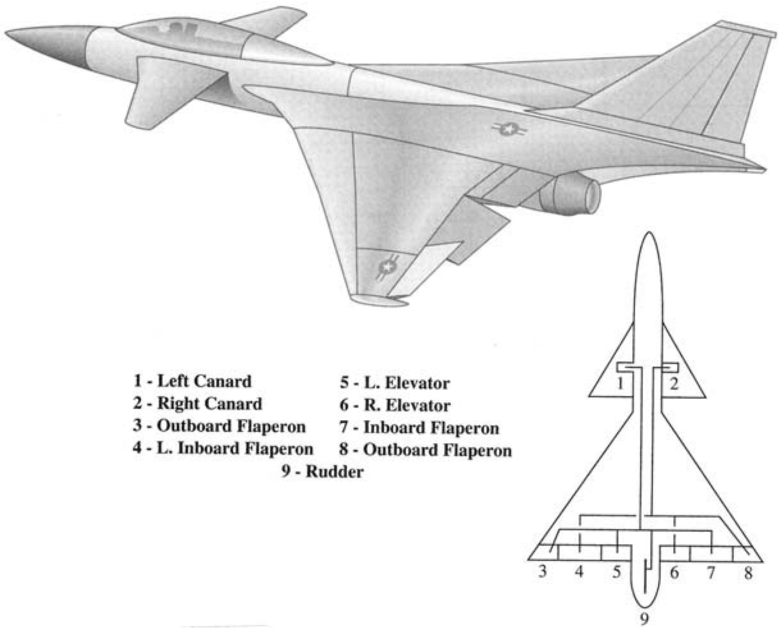


FIGURE 1.15 An aircraft designed for aero-redundancy for reconfiguration to maintain desired flying qualities.

they simply control the flight control system; they now command and control utilities and other aircraft subsystems. The flight control system has now become a more inclusive VMS.

The improved airborne digital processors have further reduced the cost, weight, and maintenance of modern aircraft. Other advantages associated with twenty-first century digital equipment include greater accuracy, increased modification flexibility through the use of software changes, improved in-flight reconfiguration techniques that compensate for physical damage and equipment failures, and more reliable preflight and postflight maintenance testing. An example of a modern high-performance aircraft is shown in Fig. 1.15. This aircraft has a quad-redundant three-axis fly-by-wire flight control system. It also includes a digital built-in task computer that runs through all the preflight tests to make sure that all equipment is functioning properly.

1.6 MATHEMATICAL BACKGROUND

The early studies of control systems were based on the solution of differential equations by classical methods. Other than for simple systems, the analysis in this approach is tedious and does not readily indicate what changes should be made to improve system performance. Use of the Laplace transform simplifies this analysis somewhat. Nyquist's paper [17] published in 1932 dealt with the application of steady-state frequency-response techniques to feedback amplifier design. This work was extended by Black [15] and Bode [18]. Hall [19] and Harris [20] applied frequency-response analysis in the study of feedback control systems, which furthered the development of control theory as a whole.

Another advance occurred in 1948, when Evans [21] presented his root-locus theory. This theory affords a graphical study of the stability properties of a system as a function of loop gain and permits the graphical evaluation of both the time and the frequency response. Laplace transform theory and network theory are joined in the root-locus calculation. In the conventional control-theory portion of this text the reader learns to appreciate the simplicity and value of the root locus technique.

The Laplace transform and the principles of linear algebra are used in the application of modern control theory to system analysis and design. The nth-order differential equation describing the system can be converted into a set of n first-order differential equations expressed in terms of the state variables. These equations can be written in matrix notation for simpler mathematical manipulation. The matrix equations lend themselves very well to computer computation. This characteristic has enabled modern control theory to solve many problems, such as nonlinear and optimization problems, which could not be solved by conventional control theory.

Mathematical models are used in the linear analysis presented in this text. Once a physical system has been described by a set of mathematical equations, they are manipulated to achieve an appropriate mathematical format. When this has been done, the subsequent method of analysis is independent of the nature of the physical system; i.e., it does not matter whether the system is electrical, mechanical, etc. This technique helps the designer to spot similarities based upon previous experience.

The reader should recognize that no single design method is intended to be used to the exclusion of the others. Depending upon the known factors and the simplicity or complexity of a control-system problem, a designer may use one method exclusively or a combination of methods. With experience in the design of feedback control systems comes the ability to use the advantages of each method.

The modern control theory presented in this text provides great potential for shaping the system output response to meet desired performance standards. Additional state-of-the-art design techniques for MIMO controls systems are presented that bring together many of the fundamentals presented earlier in the text. These chapters present the concepts of designing a robust control system in which the plant parameters may vary over specified ranges during the entire operating regime.

A control engineer must be proficient in the use of available comprehensive CAD programs similar to MATLAB (see Appendix C) or TOTAL-PC [8] (see Appendix D), which are control-system computer-aided-design programs. Many CAD packages for personal computers (PCs) and mainframe computers are available commercially. The use of a CAD package enhances the designer's control-system design proficiency, since it minimizes and expedites the tedious and repetitive calculations involved in the design of a satisfactory control system. To understand and use a computer-aided analysis and design package, one must first achieve a conceptual understanding of the theory and processes involved in the analysis and synthesis of control systems. Once the conceptual understanding is achieved by direct calculations, the reader is urged to use all available computer aids. For complicated design problems, engineers must write their own digital-computer program that is especially geared to help achieve a satisfactory system performance.

1.7 THE ENGINEERING CONTROL PROBLEM

In general, a control problem can be divided into the following steps:

1. A set of performance specifications is established.
2. The performance specifications establish the control problem.
3. A set of linear differential equations that describe the physical system is formulated or a system identification technique is applied in order to obtain the plant model transfer functions.
4. A control-theory design approach, aided by available computer-aided-design (CAD) packages or specially written computer programs, involves the following:
   (*a*) The performance of the basic (original or uncompensated) system is determined by application of one of the available methods of analysis (or a combination of them).
   (*b*) If the performance of the original system does not meet the required specifications, a control design method is selected that will improve the system's response.
   (*c*) For plants having structured parameter uncertainty, the quantitative feedback theory (QFT) [3] design technique may be used. Parametric uncertainty is present when parameters of the plant to be controlled vary during its operation, as explained in Ref. 3.
5. A simulation of the designed nonlinear system is performed.
6. The actual system is implemented and tested.

Design of the system to obtain the desired performance is the control problem. The necessary basic equipment is then assembled into a system to perform the desired control function. Although most systems are nonlinear, in many cases the nonlinearity is small enough to be neglected, or the limits of operation are small enough to allow a linear analysis to be used. This textbook considers only linear systems.

A basic system has the minimum amount of equipment necessary to accomplish the control function. After a control system is synthesized to achieve the desired performance, final adjustments can be made in a simulation, or on the actual system, to take into account the nonlinearities that were neglected. A computer is generally used in the design, depending upon the complexity of the system. The essential aspects of the control system design process are illustrated in Fig. 1.16. *Note:* The development of this figure is based upon the application of the QFT [3,8] design technique. A similar figure may be developed for other design techniques. The intent of (Fig. 1.16) is to give the reader an overview of what is involved in achieving a successful and practical control system design. The aspects of this figure that present the factors that help in *bridging the gap* between theory and the real world are addressed in the next paragraph. While accomplishing a practical control system design, the designer must keep in mind that the goal of the design process, besides achieving a satisfactory theoretical robust design, is to implement a control system that meets the functional requirements. In other words, during the design process one must keep the constraints of the real world in mind. For instance, in performing the simulations, one must be able to interpret the results obtained, based upon a knowledge of what can be reasonably expected of the plant that is being controlled. For example, in performing a time simulation of an aircraft's transient response to a pilot's maneuvering command to the flight control system, the *simulation run time* may need to be only 5s since by that time a pilot would have instituted a new command signal. If within this 5s window the performance specifications are satisfied, then it will be deemed that a successful design has been achieved. However, if the performance varies even more dramatically, rate saturation of the output effectors will significantly affect the achievement of the functional requirements. Linear and nonlinear simulations are very helpful in early evaluation of the controlled system, but if the system is to operate in the real word, hardware-in-the-loop and system tests must be performed to check for unmodeled effects not taken into account during the design and implementation phases. In order to be a successful control system designer, an individual must be fully cognizant of the role corresponding to each aspect illustrated in Fig. 1.16.

Bridging the gap, as illustrated in Fig. 1.16, is enhanced by the transparency of the metrics depicted by the oval items in the interior of the QFT

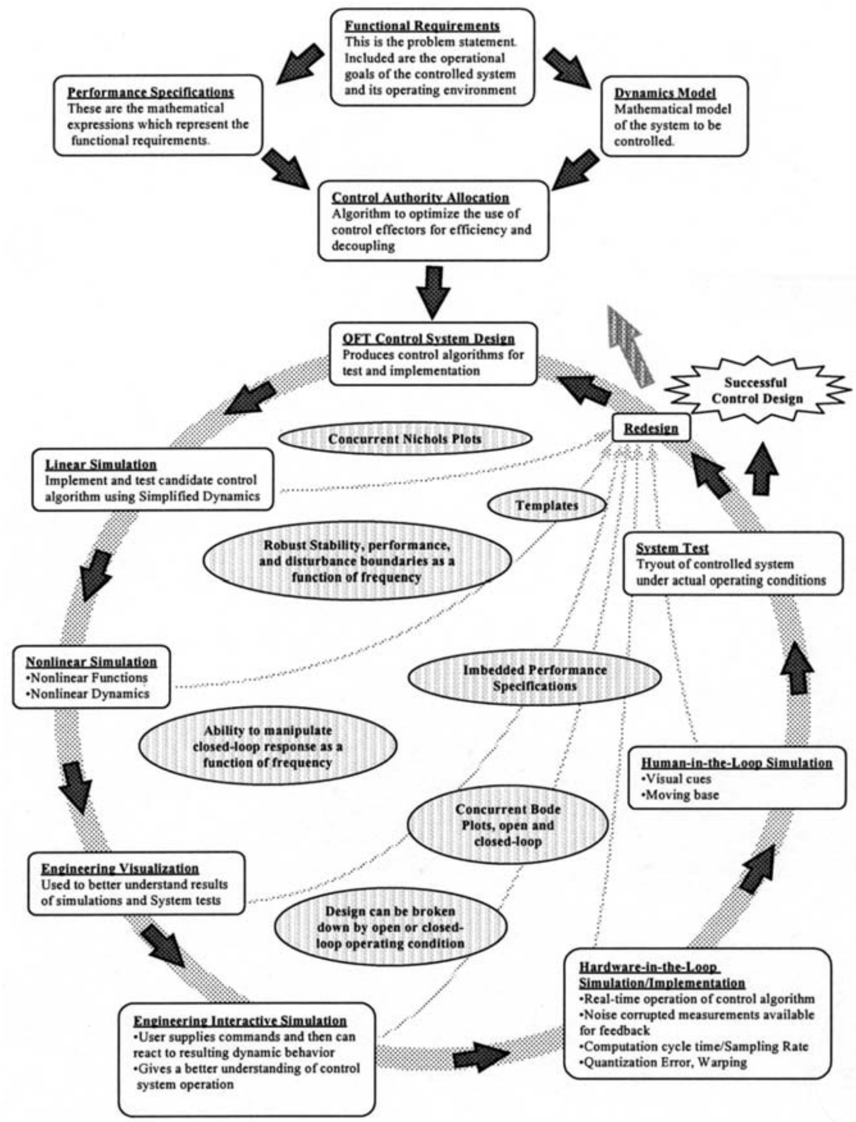


FIGURE 1.16 A control system design process: bridging the gap.

design process. A key element of QFT is embedding the performance specifications at the onset of the design process. This establishes design goals that enhance and expedite the achievement of a successful design. Another important element is the creation of templates at various frequencies.

The sizes of the templates indicate whether or not a robust design is achievable. If a robust design is not achievable, then the templates can be used as a metric in the reformation of the control design problem. Another element of the QFT design process is the ability to concurrently analyze frequency responses of the J linear-time invariant (LTI) plants that represent the nonlinear dynamical system throughout its operating environment. This gives the designer insight into the behavior of the system. The designer can use this insight for such things as picking out key frequencies to use during the design process, as an indicator of potential problems such as nonminimum phase behavior, and as a tool to compare the nonlinear system with the desired performance boundaries. The next element of QFT consists of the design boundaries. During the actual loop shaping process, the designer uses boundaries plotted on the Nichols chart. These boundaries are only guidelines and the designer can exercise engineering judgment to determine if all the boundaries are critical or if some of the boundaries are not important. For example, based on knowledge of the real world system, the designer may determine that meeting performance boundaries below a certain frequency is not important, but it is important to meet the disturbance rejection boundaries below that frequency. Once the initial design has been accomplished, all of the J loop transmission functions can be plotted on a Nichols chart to analyze the results of applying the designed compensator (controller) to the nonlinear system. This gives the designer a first look at any areas of the design that may present problems during simulation and implementation. The last two elements of the QFT design process that help *bridging the gap* is the relationship of the controlled system's behavior to the frequency domain design and the operating condition. These relationships enable the designer to better analyze simulation or system test results for problems in the control design. To obtain a successful control design, the controlled system must meet all of the requirements during simulation and system test. If the controlled system fails any of the simulation or system tests, then, using the design elements of QFT, the designer can trace that failure back through the design process and make necessary adjustments to the design. QFT provides many metrics that provide the link between the control design process and real world implementation; this is the *transparency of QFT.*

1.8 COMPUTER LITERACY

In the mid 1960s, the first practical CAD package, FREQR, for frequency domain analysis was developed at the Air Force Institute of Technology (AFIT). This was followed by the development of two other AFIT CAD packages: PARTL for partial-fraction expansion of transfer functions and for obtaining a time response, and ROOTL for obtaining root-locus data and plots. These CAD packages became the basis for the practical control system design CAD package called *TOTAL*, the forerunner of TOTAL-PC [8], which was developed in 1978 at AFIT. TOTAL became the catalyst, along with other control CAD packages developed by other individuals, for the development of the current highly developed commercial control system design CAD packages that are now readily available. One of these CAD packages is *MATLAB*, which has become a valuable tool for a control engineer and is illustrated in this text. The detailed contents of the TOTAL-PC CAD package are described in Appendix D. The program is contained in the disk which is included with this book.

Becoming proficient (computer literate) in the use of these CAD packages (tools) is essential for a control systems engineer. It is also essential to develop procedures for checking the CAD results at each stage of the analysis and design. This is necessary in order to verify that these CAD tools have generated results which are consistent with theory. Whenever in doubt, concerning the operation of a specific CAD tool, apply the CAD tool to a simple problem whose known analytical solution can be readily compared to the computer generated output.

1.9 OUTLINE OF TEXT

The text is essentially divided into three parts. The first part, consisting of Chapters 2 through 4, provides the mathematical foundation for modeling physical systems and obtaining time solutions using classical or Laplace transform methods. The second part consists of Chapters 5 through 9 that provide the fundamentals of conventional control theory and state-variable concepts. The remaining portion of the text represents material that is usually covered in the first or second undergraduate course in control theory and control system design.

The first few chapters deal with the mathematics and physical system modeling that underlie the analysis of control systems. Once the technique of writing the system equations (and, in turn, their Laplace transforms) that describe the performance of a dynamic system has been mastered, the ideas of block and simulation diagrams and transfer functions are developed. When physical systems are described in terms of block diagrams and transfer functions, they exhibit basic servo characteristics. These characteristics are described and discussed. The concept of state is introduced, and the system equations are developed in the standard matrix format. The necessary linear algebra required to manipulate the matrix equations is included. A presentation of the various methods of analysis is next presented that can be used in the study of feedback control systems. SISO systems are used initially to facilitate an understanding of the synthesis methods. These methods rely on root-locus and steady-state frequency-response analysis. If the initial design does not meet the desired specifications, then improvement of the basic system by using compensators in presented. These compensators can be designed by analyzing and synthesizing a desired open-loop transfer function or by synthesizing a desired closed-loop transfer function that produces the desired overall system performance. Chapter 12 is devoted to modeling a desired closed-loop transfer function for tracking a desired input or for rejecting (not responding to) disturbance inputs.

The next portion of the text deals with SISO system design using state feedback. Topics such as controllability and observability, pole placement via state-variable feedback, and parameter sensitivity are presented. The fundamentals of sampled-data (S-D) control-system analysis are presented in Chap. 15. The design of cascade and feedback compensators (controllers) for improving the performance of S-D control systems is presented in Chap. 16. Chapter 16 introduces the concept of the representation of, and the design of digital control systems by a psuedo-continuous-time (PCT) control system.

Appendix A gives a table of Laplace transform pairs. Fundamentals of basic matrix algebra are presented in Appendix B. A description of the MATLAB SIMULINK computer-aided-design packages, that are useful to a feedback control system engineer, is presented in Appendix C.

In closing this introductory chapter it is important to stress that feedback control engineers are essentially "system engineers," i.e., people whose primary concern is with the design and synthesis of an overall system. To an extent depending on their own background and experience, they rely on, and work closely with, engineers in the various recognized branches of engineering to furnish them with the transfer functions and/or system equations of various portions of a control system and the special characteristics of the physical plant being controlled. For example, the biomedical control engineer works closely with the medical profession in modeling biological functions, man-machine interface problems, etc.

The following design policy includes factors that are worthy of consideration in the control system design problem:

1. Use proven design methods.
2. Select the system design that has the minimum complexity.
3. Use minimum specifications or requirements that yield a satisfactory system response. Compare the cost with the performance and select the fully justified system implementation.
4. Perform a complete and adequate simulation and testing of the system.

REFERENCES

1. *IEEE Standard Dictionary of Electrical and Electronics Terms*, Wiley-Interscience, New York, 1972.
2. Bossert, D. E.: "Design of Pseudo-Continuous-Time Quantitative Feedback Theory Robot Controllers," MS thesis, AFIT/GE/ENG/89D-2, Graduate School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB, Ohio, December 1989.
3. Houpis, C. H., and S. J. Rasmussen: *Quantitative Feedback Theory Fundamentals and Applications*, Marcel Dekker, New York, 1999.
4. Hamilton, G. K., and M. A. Franchel: "Robust Controller Design and Experimental Verification of I.C. Engine Speed Control," *Int. J. of Robust and Nonlinear Control*, vol. **7**, pp. 609–627, 1997.
5. Jayasuriya, S., and M. A. Franchel: "A QFT Type Design Methodology for a Parallel Plant Structure and Its Application in Idle Speed Control," *Int. J. Control*, vol. **60**, pp. 653–670, 1994.
6. Buehler, P. J., M. A. Franchek, and I. Makki: "Mass Air Flow Sensor Diagnostics for Adaptive Fueling Control of Internal Combustion Engines," Proceedings of the American Control Conference, Anchorage, Alaska, May 8–10, 2002.
7. Ostolaza, J. X., and M. Garcia-Sanz: "Control of an Activated Sludge Wastewater Treatment Plant with Nitrification-Denitrification Configuration Using QFT Technique," Symposium on Quantitative Feedback Theory and Other Frequency Domain Methods and Applications Proceedings, University of Strathclyde, Glasgow, Scotland, August 1997.
8. D'Azzo, J. J., and C. H. Houpis: *Linear Control System Analysis and Design: Conventional and Modern*, 4th ed., McGraw-Hill, New York, 1995.
9. Phillips, S., M. Pachter, and C. H. Houpis: "A QFT Subsonic Envelope Flight Control System Design," National Aerospace Electronics Conference (NAECON), Dayton, Ohio, May 1995.
10. Sheldon, S. N., and C. Osmor: "Piloted Simulation of An F-16 Flight Control System Designed Using Quantitative Feedback Theory," Symposium on Qantitative Feedback Theory and Other Frequency Domain Methods and Applications Proceedings, University of Strathclyde, Glasgow, Scotland, August 1997
11. Sheldon, S. N., and C. Osmon: "Piloted Simulation of An F-16 Flight Control System Designed Using Quantitative Feedback Theory," Symposium on Quantitative Feedback theory and Other Frequency Domain Methods and Applications Proceedings, University of Strathclyde, Glasgow, Scotland, August 1997.
12. Mayr, O.: *Origins of Feedback Control*, M.I.T. Press, Cambridge, Mass., 1971.
13. Trinks, W.: *Governors and the Governing of Prime Movers*, Van Nostrand, Princeton, N.J., 1919.
14. Minorsky, N.: "Directional Stability and Automatically Steered Bodies," *J. Am. Soc. Nav. Eng.*, vol. **34**, p. 280, 1922.
15. Black, H. S.: "Stabilized Feedback Amplifiers," *Bell Syst. Tech. J.*, 1934.
16. Houpis, C. H., and G. B. Lamont: *Digital Control Systems Theory, Hardware, Software*, 2nd ed., McGraw-Hill, New York, 1992.

17. Nyquist, H.: "Regeneration Theory," *Bell Syst. Tech. J.*, 1932.
18. Bode, H. W.: *Network Analysis and Feedback Amplifier Design*, Van Nostrand, Princeton, N.J., 1945.
19. Hall, A. C.: "Application of Circuit Theory to the Design of Servomechanisms," *J. Franklin Inst.*, 1946.
20. Harris, H.: The Frequency Response of Automatic Control System, "*Trans. AIEE*, vol. **65**, pp. 539–1546, 1946.
21. Evans, W. R.: "Graphical Analysis of Control Systems," *Trans. AIEE*, vol. **67**, pt. II, pp. 547–1551, 1948.
22. Allan, R.: "Busy Robots Spur Productivity," *IEEE Spectrum*, vol. **16**, pp. 31–136, 1979.
23. Dorf, R. C.: *Encyclopedia of Robotics*, Wiley-Interscience, New York, 1988.
24. Vidyasagar, M.: "System Theory and Robotics," *IEEE Control Systems*, pp. 16–117, April 1987.

2

Writing System Equations

2.1 INTRODUCTION

An accurate mathematical model that describes a system completely must be determined in order to analyze a dynamic system. The derivation of this model is based upon the fact that the dynamic system can be completely described by known differential equations (Chap. 2) or by experimental test data (Sec. 8.1). The ability to analyze the system and determine its performance depends on how well the characteristics can be expressed mathematically. Techniques for solving linear differential equations with constant coefficients are presented in Chaps. 3 and 4. However, the solution of a time-varying or nonlinear equation often requires a numerical, graphic, or computer procedure [1]. The systems considered in this chapter are described completely by a set of linear *constant-coefficient* differential equations. Such systems are said to be *linear time-invariant* (LTI) [2]; i.e., the relationship between the system input and output is independent of time. Since the system does not change with time, the output is independent of the time at which the input is applied. Linear methods are used because there is extensive and elegant mathematics for solving linear equations. For many systems, there are regions of operation for which the linear representation works very well.

This chapter presents methods for writing the differential and state equations for a variety of electrical, mechanical, thermal, and hydraulic systems [3]. This step is the first that must be mastered by the control-systems engineer. The basic physical laws are given for each system, and the associated parameters are defined. Examples are included to show the application of the basic laws to physical equipment. The result is a differential equation, or a set of differential equations, that describes the system. The equations derived are limited to linear systems or to systems that can be represented by linear equations over their useful operating range. The important concepts of system and of state variables are also introduced. The system equations, expressed in terms of state variables, are called *state equations.*

The analytical tools of linear algebra are presented in Appendix B. They are introduced in the text as they are needed in the development of the system differential equations. They are also introduced in a logical format in the development of state equations. This facilitates the solution of the differential and state equations that are covered in Chap. 3. This is followed by the use of Laplace transforms to facilitate solution of differential and state equations in Chap. 4. The linear algebra format also facilitates the use of computer-aided techniques for solving the differential and state equations as presented in Appendix C.

The analysis of behavior of the system equations is enhanced by using the block diagram representation of the system. Complete drawings showing all the detailed parts are frequently too congested to show the specific functions that are performed by each system component. It is common to use a block diagram, in which each function is represented by a block in order to simplify the picture of the complete system. Each block is labeled with the name of the component, and the blocks are appropriately interconnected by line segments. This type of diagram removes excess detail from the picture and shows the functional operation of the system. The use of a block diagram provides a simple means by which the functional relationship of the various components can be shown and reveals the operation of the system more readily than does observation of the physical system itself. The simple functional block diagram shows clearly that apparently different physical systems can be analyzed by the same techniques. Since a block diagram is involved not with the physical characteristics of the system but only with the functional relationship between various points in the system, it can reveal the similarity between apparently unrelated physical systems.

A block diagram [4] represents the flow of information and the functions performed by each component in the system. A further step taken to increase the information supplied by the block diagram is to label the input quantity into each block and the output quantity from each block. Arrows are used to show the direction of the flow of information. The block represents the function or dynamic characteristics of the component and is represented by a transfer function (Sec. 2.4). The complete block diagram shows how the functional components are connected and the mathematical equations that determine the response of each component. Examples of block diagrams are shown throughout this chapter.

In general, variables that are functions of time are represented by lowercase letters. These are sometimes indicated by the form $x(t)$, but more often this is written just as x. There are some exceptions, because of established convention, in the use of certain symbols.

To simplify the writing of differential equations, the D operator notation is used [5]. The symbols D and $1/D$ are defined by

$$Dy \equiv \frac{dy(t)}{dt} \qquad D^2 y \equiv \frac{d^2 y(t)}{dt^2} \tag{2.1}$$

$$D^{-1} y \equiv \frac{1}{D} y \equiv \int_0^t y(\tau)\, d\tau + \int_{-\infty}^0 y(\tau)\, d\tau = \int_0^t y(\tau)\, d\tau + Y_0 \tag{2.2}$$

where Y_0 represents the value of the integral at time $t = 0$, that is, the *initial value* of the integral.

2.2 ELECTRIC CIRCUITS AND COMPONENTS [6]

The equations for an electric circuit obey Kirchhoff's laws, which state the following:

1. The algebraic sum of the potential differences around a closed path equals zero. This can be restated as follows: In traversing any closed loop, the sum of the voltage rises equals the sum of the voltage drops.
2. The algebraic sum of the currents entering (or leaving) a node is equal to zero. In other words, the sum of the currents entering the junction equals the sum of the currents leaving the junction.

The voltage sources are usually alternating-current (ac) or direct-current (dc) generators. The usual dc voltage source is shown as a battery. The voltage drops appear across the three basic electrical elements: resistors, inductors, and capacitors. These elements have constant component values.

The voltage drop across a resistor is give by Ohm's law, which states that the voltage drop across a resistor is equal to the product of the current through the resistor and its resistance. Resistors absorb energy from the system. Symbolically, this voltage is written as

$$v_R = Ri \tag{2.3}$$

The voltage drop across an inductor is given by Faraday's law, which is written

$$v_L = L\frac{di}{dt} \equiv L\, Di \tag{2.4}$$

TABLE 2.1 Electrical Symbols and Units

Symbol	Quantity	Units
e or v	Voltage	Volts
i	Current	Amperes
L	Inductance	Henrys
C	Capacitance	Farads
R	Resistance	Ohms

This equation states that the voltage drop across an inductor is equal to the product of the inductance and the time rate of increase of current. A positive-valued derivative Di implies an increasing current, and thus a positive voltage drop; a negative-valued derivative implies a decreasing current, and thus a negative voltage drop.

The positively directed voltage drop across a capacitor is defined as the ratio of the magnitude of the positive electric charge on its positive plate to the value of its capacitance. Its direction is from the positive plate to the negative plate. The charge on a capacitor plate is equal to the time integral of the current entering the plate from the initial instant to the arbitrary time t, plus the initial value of the charge. The capacitor voltage is written in the form

$$v_C = \frac{q}{C} = \frac{1}{C}\int_0^t i\,d\tau + \frac{Q_0}{C} \equiv \frac{i}{CD} \tag{2.5}$$

The mks units for these electrical quantities in the practical system are given in Table 2.1.

Series Resistor-Inductor Circuit

The voltage source e in Fig. 2.1 is a function of time. When the switch is closed, setting the voltage rise equal to the sum of the voltage drops produces

$$\begin{aligned} v_R + v_L &= e \\ Ri + L\frac{di}{dt} &= Ri + L\,Di = e \end{aligned} \tag{2.6}$$

When the applied voltage e is known, the equation can be solved for the current i, as shown in Chap. 3. This equation can also be written in terms of the voltage v_L across the inductor in the following manner. The voltage across the inductor is $v_L = L\,Di$. Therefore, the current through the inductor is

$$i = \frac{1}{LD}v_L$$

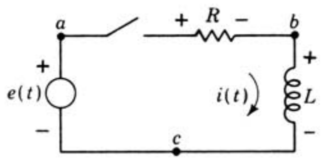


FIGURE 2.1 Series resistor-inductor circuit.

Substituting this value into the original equation gives

$$\frac{R}{LD}v_L + v_L = e \tag{2.7}$$

The node method is also convenient for writing the system equations directly in terms of the voltages. The junctions of any two elements are called *nodes.* This circuit has three nodes, labeled a, b, and c (see Fig. 2.1). One node is used as a reference point; in this circuit node c is used as the reference. The voltages at all the other nodes are considered with respect to the reference node. Thus v_{ac} is the voltage drop from node a to node c, and v_{bc} is the voltage drop from node b to the reference node c. For simplicity, these voltages are written just as v_a and v_b.

The source voltage $v_a = e$ is known; therefore, there is only one unknown voltage, v_b, and only one node equation is necessary. Kirchhoff's second law, that the algebraic sum of the currents entering (or leaving) a node must equal zero, is applied to node b. The current from node b to node a, through the resistor R, is $(v_b - v_a)/R$. The current from node b to node c through the inductor L is $(1/LD)v_b$. The sum of these currents must equal zero:

$$\frac{v_b - v_a}{R} + \frac{1}{LD}v_b = 0 \tag{2.8}$$

Rearranging terms gives

$$\left(\frac{1}{R} + \frac{1}{LD}\right)v_b - \frac{1}{R}v_a = 0 \tag{2.9}$$

Except for the use of different symbols for the voltages, this is the same as Eq. (2.7). Note that the node method requires writing only one equation.

Series Resistor-Inductor-Capacitor Circuit

For the series RLC circuit shown in Fig. 2.2, the applied voltage is equal to the sum of the voltage drops when the switch is closed:

$$v_L + v_R + v_C = e \qquad L\,Di + Ri + \frac{1}{CD}i = e \tag{2.10}$$

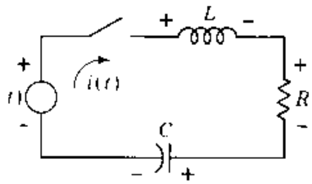


FIGURE 2.2 Series *RLC* circuit.

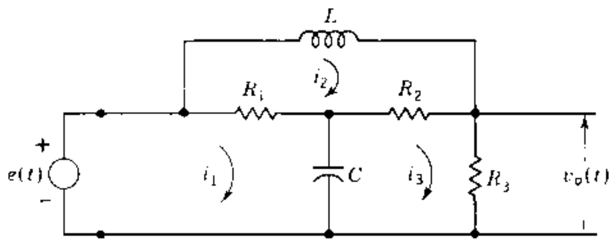


FIGURE 2.3 Multiloop network.

The circuit equation can be written in terms of the voltage drop across any circuit element. For example, in terms of the voltage across the resistor, $v_R = Ri$, Eqs. (2.10) become

$$\frac{L}{R}Dv_R + v_R + \frac{1}{RCD}v_R = e \tag{2.11}$$

Multiloop Electric Circuits

Multiloop electric circuits (see Fig. 2.3) can be solved by either loop or nodal equations. The following example illustrates both methods. The problem is to solve for the output voltage v_o.

LOOP METHOD. A loop current is drawn in each closed loop (usually in a clockwise direction); then Kirchhoff's voltage equation is written for each loop:

$$\left(R_1 + \frac{1}{CD}\right)i_1 - R_1 i_2 - \frac{1}{CD}i_3 = e \tag{2.12}$$

$$-R_1 i_1 + (R_1 + R_2 + LD)i_2 - R_2 i_3 = 0 \tag{2.13}$$

$$-\frac{1}{CD}i_1 - R_2 i_2 + \left(R_2 + R_3 + \frac{1}{CD}\right)i_3 = 0 \tag{2.14}$$

The output voltage is

$$v_o = R_3 i_3 \tag{2.15}$$

These four equations must be solved simultaneously to obtain $v_o(t)$ in terms of the input voltage $e(t)$ and the circuit parameters.

NODE METHOD. The junctions, or nodes, are labeled by letters in Fig. 2.4. Kirchhoff's current equations are written for each node in terms of the node voltages, where node d is taken as reference. The voltage v_{bd} is the voltage of node b with reference to node d. For simplicity, the voltage v_{bd} is written just as v_b. Since there is one known node voltage $v_a = e$ and two unknown voltages v_b and v_o, only two equations are required:

$$\text{For node } b: \qquad i_1 + i_2 + i_3 = 0 \tag{2.16}$$

$$\text{For node } c: \qquad -i_3 + i_4 + i_5 = 0 \tag{2.17}$$

In terms of the node voltages, these equations are

$$\frac{v_b - v_a}{R_1} + CDv_b + \frac{v_b - v_o}{R_2} = 0 \tag{2.18}$$

$$\frac{v_o - v_b}{R_2} + \frac{v_o}{R_3} + \frac{1}{LD}(v_o - e) = 0 \tag{2.19}$$

Rearranging the terms in order to systematize the form of the equations gives

$$\left(\frac{1}{R_1} + CD + \frac{1}{R_2}\right)v_b - \frac{1}{R_2}v_o = \frac{1}{R_1}e \tag{2.20}$$

$$-\frac{1}{R_2}v_b + \left(\frac{1}{R_2} + \frac{1}{R_3} + \frac{1}{LD}\right)v_o = \frac{1}{LD}e \tag{2.21}$$

For this example, only two nodal equations are needed to solve for the potential at node c. An additional equation must be used if the current in R_3 is required. With the loop method, three equations must be solved simultaneously to obtain the current in any branch; an additional equation must be used if the voltage across R_3 is required. The method that requires the solution of the fewest equations should be used. This varies with the circuit.

The rules for writing the node equations are summarized as follows:

1. The number of equations required is equal to the number of unknown node voltages.
2. An equation is written for each node.
3. Each equation includes the following:
   (*a*) The node voltage multiplied by the sum of all the admittances that are connected to this node. This term is positive.

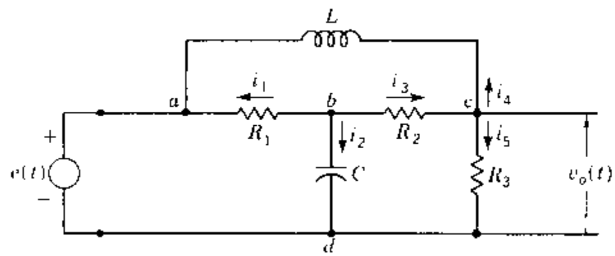


FIGURE 2.4 Multinode network.

(*b*) The node voltage at the other end of each branch multiplied by the admittance connected between the two nodes. This term is negative.

The reader should learn to apply these rules so that Eqs. (2.20) and (2.21) can be written directly from the circuit of Fig. 2.4.

2.3 STATE CONCEPTS

Basic matrix properties are used to introduce the concept of state and the method of writing and solving the state equations. The state of a system (henceforth referred to only as state) is defined by Kalman [7] as follows:

STATE. The *state* of a system is a mathematical structure containing a set of n variables $x_1(t), x_2(t), \ldots, x_i(t), \ldots, x_n(t)$, called the *state variables*, such that *the initial values* $x_i(t_0)$ of this set and the system inputs $u_j(t)$ are sufficient to describe uniquely the system's future response of $t \geq t_0$. A minimum set of state variables is required to represent the system accurately. The m inputs, $u_1(t), u_2(t), \ldots, u_j(t), \ldots, u_m(t)$, are deterministic; i.e., they have specific values for all values of time $t \geq t_0$.

Generally the initial starting time t_0 is taken to be zero. The state variables need not be physically observable and measurable quantities; they may be purely mathematical quantities. The following additional definitions apply:

STATE VECTOR. The set of state variables $x_i(t)$ represents the elements or components of the n-dimensional state vector $\mathbf{x}(t)$; that is,

$$\mathbf{x}(t) \equiv \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \equiv \mathbf{x} \tag{2.22}$$

The order of the system characteristic equation is n, and the state equation representation of the system consists of n first-order differential equations. When all the inputs $u_j(t)$ to a given system are specified for $t > t_0$, the resulting state vector uniquely determines the system behavior for any $t > t_0$.

STATE SPACE. State space is defined as the n-dimensional space in which the components of the state vector represent its coordinate axes.

STATE TRAJECTORY. State trajectory is defined as the path produced in the state space by the state vector $\mathbf{x}(t)$ as it changes with the passage of time. State space and state trajectory in the two-dimensional case are referred to as the *phase plane* and *phase trajectory*, respectively [1].

The first step in applying these definitions to a physical system is the selection of the system variables that are to represent the state of the system. Note that there is no unique way of making this selection. The three common representations for expressing the system state are the *physical*, *phase*, and *canonical state variables*. The physical state-variable representation is introduced in this chapter. The other two representations are introduced in later chapters.

The selection of the state variables for the physical-variable method is based upon the energy-storage elements of the system. Table 2.2 lists some common energy-storage elements that exist in physical systems and the corresponding energy equations. The physical variable in the energy equation for each energy-storage element *can* be selected as a state variable of the system. Only independent physical variables are chosen to be state variables. *Independent state variables* are those state variables that cannot be expressed in terms of the remaining assigned state variables. In some systems it may be necessary to identify more state variables than just the energy-storage variables. This situation is illustrated in some of the following examples, where velocity is a state variable. When position, the integral of this state variable, is of interest, it must also be assigned as a state variable.

Example 1: Series *RL* Circuit (Fig. 2.1). Only one energy-storage element, the inductor, is present in this circuit; thus there is only one state variable. From Table 2.2, the state variable is $x_1 = i$. The equation desired is one that contains the first derivative of the state variable. Letting $u = e$, the loop equation, Eq. (2.6), can be rewritten as

$$Rx_1 + L\dot{x}_1 = u$$

$$\dot{x}_1 = -\frac{R}{L}x_1 + \frac{1}{L}u \tag{2.23}$$

The letter u is the standard notation for the input forcing function and is called the *control variable*. Equation (2.23) is the standard form of the *state* *equation* of the system. There is only one state equation because this is a first-order system, $n = 1$.

Example 2: Series *RLC* Circuit (Fig. 2.2). This circuit contains two energy-storage elements, the inductor and capacitor. From Table 2.2, the two assigned state variables are identified as $x_1 = v_c$ (the voltage across the capacitor) and $x_2 = i$ (the current in the inductor). Thus two state equations are required.

A loop or branch equation is written to obtain an equation containing the derivative of the current in the inductor. A node equation is written to obtain an equation containing the derivative of the capacitor voltage. The number of loop equations that must be written is equal to the number of state variables representing currents in inductors. The number of equations involving node voltages that must be written is equal to the number of state variables representing voltages across capacitors. These are usually, but not always, node equations. These loop and node equations are written in terms of the inductor *branch* currents and capacitor branch voltages. From these equations, it is necessary to determine which of the assigned physical variables are independent. When a variable of interest is not an energy physical variable, then this variable is identified as an *augmented state variable.*

TABLE 2.2 Energy-Storage Elements

Element	Energy	Physical variable
Capacitor C	$\frac{Cv^2}{2}$	Voltage v
Inductor L	$\frac{Li^2}{2}$	Current i
Mass M	$\frac{Mv^2}{2}$	Translational velocity v
Moment of inertia J	$\frac{J\omega^2}{2}$	Rotational velocity ω
Spring K	$\frac{Kx^2}{2}$	Displacement x
Fluid compressibility $\frac{V}{K_B}$	$\frac{VP_L^2}{2K_B}$	Pressure P_L
Fluid capacitor $C = \rho A$	$\frac{\rho A h^2}{2}$	Height h
Thermal capacitor C	$\frac{C\theta^2}{2}$	Temperature θ

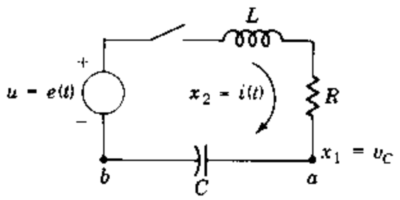


FIGURE 2.5 Series *RLC* circuit.

Figure 2.2 is redrawn in Fig. 2.5 with node b as the reference node. The node equation for node a and the loop equation are, respectively,

$$C\dot{x}_1 = x_2 \tag{2.24}$$

and

$$L\dot{x}_2 + Rx_2 + x_1 = u \tag{2.25}$$

Rearranging terms to the standard state equation format yields

$$\dot{x}_1 = \frac{1}{C}x_2 \tag{2.26}$$

$$\dot{x}_2 = -\frac{1}{L}x_1 - \frac{R}{L}x_2 + \frac{1}{L}u \tag{2.27}$$

Equations (2.26) and (2.27) represent the state equations of the system containing two independent state variables. Note that they are *first-order* linear differential equations and are $n = 2$ in number. They are the minimum number of state equations required to represent the system's future performance.

The following definition is based upon these two examples.

STATE EQUATION. The state equations of a system are a set of n first-order differential equations, where n is the number of independent states.

The state equations represented by Eqs. (2.26) and (2.27) are expressed in matrix notation as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{C} \\ -\frac{1}{L} & -\frac{R}{L} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{L} \end{bmatrix} [u] \tag{2.28}$$

It can be expressed in a more compact form as

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}\mathbf{u} \tag{2.29}$$

where

$$\dot{\mathbf{x}} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} \qquad \text{an } n \times 1 \text{ column vector} \tag{2.30}$$

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 0 & \dfrac{1}{C} \\ -\dfrac{1}{L} & -\dfrac{R}{L} \end{bmatrix}$$

$$\text{an } n \times n \text{ plant coefficient matrix} \tag{2.31}$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \text{an } n \times 1 \text{ state vector} \tag{2.32}$$

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0 \\ \dfrac{1}{L} \end{bmatrix} \qquad \text{an } n \times 1 \text{ control matrix} \tag{2.33}$$

and, in this case, $\mathbf{u} = [u]$ is a one-dimensional control vector. In Eq. (2.29), matrices **A** and **x** are conformable.

If the output quantity $y(t)$ for the circuit of Fig. 2.2 is the voltage across the capacitor v_C, then

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

Thus the matrix *system output* equation for this example is

$$\mathbf{y}(t) = \mathbf{c}^T\mathbf{x} + \mathbf{du} = [1 \quad 0]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + [0][u] \tag{2.34}$$

where

$\mathbf{c}^T = [1 \quad 0]$ a 1×2 row vector

$\mathbf{y} = [y]$ a one-dimensional output vector

$\mathbf{d} = 0$

Equations (2.29) and (2.34) are for a single-input single-output (SISO) system. For a multiple-input multiple-output (MIMO) system, with m inputs and l outputs, these equations become

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu} \tag{2.35}$$

$$\mathbf{y} = \mathbf{Cx} + \mathbf{Du} \tag{2.36}$$

where

$\mathbf{A} = n \times n$ plant matrix

$\mathbf{B} = n \times m$ control matrix

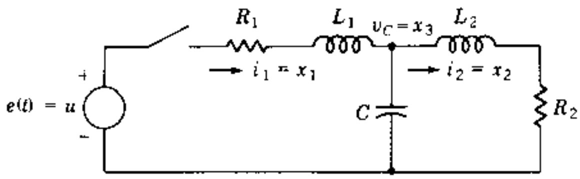


FIGURE 2.6 An electric circuit.

$\mathbf{C} = l \times n$ output matrix

$\mathbf{D} = l \times m$ feedforward matrix

$\mathbf{u} = m$-dimensional control vector

$\mathbf{y} = l$-dimensional output vector

Example 3. Obtain the state equation for the circuit of Fig. 2.6, where i_2 is considered to be the output of this system. The assigned state variables are $x_1 = i_1$, $x_2 = i_2$, and $x_3 = v_C$. Thus, two loop and one node equations are written

$$R_1 x_1 + L_1 \dot{x}_1 + x_3 = u \tag{2.37}$$

$$-x_3 + L_2 \dot{x}_2 + R_2 x_2 = 0 \tag{2.38}$$

$$-x_1 + x_2 + C\dot{x}_3 = 0 \tag{2.39}$$

The three state variables are independent, and the system state and output equations are

$$\dot{\mathbf{x}} = \begin{bmatrix} -\dfrac{R_1}{L_1} & 0 & -\dfrac{1}{L_1} \\ 0 & -\dfrac{R_2}{L_2} & \dfrac{1}{L_2} \\ \dfrac{1}{C} & -\dfrac{1}{C} & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} \dfrac{1}{L_1} \\ 0 \\ 0 \end{bmatrix} \mathbf{u} \tag{2.40}$$

$$\mathbf{y} = [0 \quad 1 \quad 0]\mathbf{x} \tag{2.41}$$

Example 4. Obtain the state equations for the circuit of Fig. 2.7. The output is the voltage v_1. The input or control variable is a current source $i(t)$. The assigned state variables are i_1, i_2, i_3, v_1, and v_2. Three loop equations and two node equations are written:

$$v_1 = L_1\, Di_1 \tag{2.42}$$

$$v_2 = L_2\, Di_2 + v_1 \tag{2.43}$$

$$v_2 = L_3\, Di_3 \tag{2.44}$$

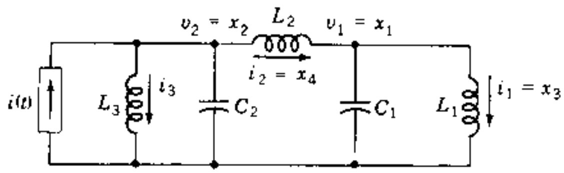


FIGURE 2.7 Electric circuit.

$$i_2 = C_1\, Dv_1 + i_1 \tag{2.45}$$

$$i = i_3 + C_2\, Dv_2 + i_2 \tag{2.46}$$

Substituting from Eqs. (2.42) and (2.44) into Eq. (2.43), or writing the loop equation through L_1, L_2, and L_3 and then integrating (multiplying by $1/D$), gives

$$L_3 i_3 = L_2 i_2 + L_1 i_1 + K \tag{2.47}$$

where K is a function of the initial conditions. This equation reveals that one inductor current is dependent upon the other two inductor currents. Thus, this circuit has only *four* independent physical state variables, two inductor currents and two capacitor voltages. The four independent state variables are designated as $x_1 = v_1$, $x_2 = v_2$, $x_3 = i_1$, and $x_4 = i_2$, and the control variable is $u = i$. Three state equations are obtainable from Eqs. (2.42), (2.43), and (2.45). The fourth equation is obtained by eliminating the dependent current i_3 from Eqs. (2.46) and (2.47). The result in matrix form is

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 0 & -\dfrac{1}{C_1} & \dfrac{1}{C_1} \\ 0 & 0 & -\dfrac{L_1}{L_3 C_2} & -\dfrac{L_2 + L_3}{L_3 C_2} \\ \dfrac{1}{L_1} & 0 & 0 & 0 \\ -\dfrac{1}{L_2} & \dfrac{1}{L_2} & 0 & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ \dfrac{1}{C_2} \\ 0 \\ 0 \end{bmatrix} = \mathbf{u} \tag{2.48}$$

$$\mathbf{y} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \mathbf{x} \tag{2.49}$$

The dependence of i_3 on i_1 and i_2 as shown by Eq. (2.47) may not be readily observed. In that case the matrix state equation for this example would be written with five state variables.

The examples in this section are fairly straightforward. In general it is necessary to write more than just the number of state equations, because other system variables appear in them. These equations are solved

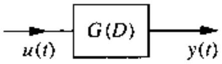


FIGURE 2.8 Block diagram representation of Fig. 2.2.

simultaneously to eliminate all internal variables in the circuit except for the state variables. This procedure is necessary in some of the problems for this chapter. For more complex circuits it is possible to introduce more generalized and systematized linear graphs for obtaining the state equations. They are not included in this book.

2.4 TRANSFER FUNCTION AND BLOCK DIAGRAM

A quantity that plays a very important role in control theory is the system *transfer function*, defined as follows:

TRANSFER FUNCTION. If the system differential equation is linear, the ratio of the output variable to the input variable, where the variables are expressed as functions of the D operator, is called the transfer function.

Consider the system output $v_C = y$ in the RLC circuit of Fig. 2.2. Substituting $i = CDv_C$ into Eq. (2.10) yields

$$(LCD^2 + RCD + 1)v_C(t) = e(t) \tag{2.50}$$

The system transfer function is

$$G(D) = \frac{y(t)}{u(t)} = \frac{v_C(t)}{e(t)} = \frac{1}{LCD^2 + RCD + 1} \tag{2.51}$$

The notation $G(D)$ is used to denote a transfer function when it is expressed in terms of the D operator. It may also be written simply as G.

The *block diagram* representation of this system (Fig. 2.8) represents the mathematical operation $G(D)u(t) = y(t)$; that is, the transfer function times the input is equal to the output of the block. The resulting equation is the differential equation of the system.

2.5 MECHANICAL TRANSLATION SYSTEMS [1,8,9]

Mechanical systems obey Newton's law that the sum of the forces equals zero; that is, the sum of the applied forces must be equal to the sum of the reactive forces. The three qualities characterizing elements in a mechanical translation* system are mass, elastance, and damping. The following analysis

** Translation* means motion in a straight line.

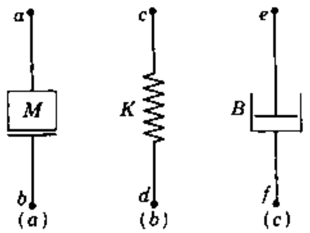


FIGURE 2.9 Network elements of mechanical translation systems.

includes only linear functions. Static friction, Coulomb friction, and other nonlinear friction terms are not included. Basic elements entailing these qualities are represented as network elements [10], and a mechanical network is drawn for each mechanical system to facilitate writing the differential equations.

The mass M is the inertial element. A force applied to a mass produces an acceleration of the mass. The reaction force f_M is equal to the product of mass and acceleration and is opposite in direction to the applied force. In terms of displacement x, velocity v, and acceleration a, the force equation is

$$f_M = Ma = M\,Dv = M\,D^2 x \tag{2.52}$$

The network representation of mass is shown in Fig. 2.9*a*. One terminal, a, has the motion of the mass; and the other terminal, b, is considered to have the motion of the reference. The reaction force f_M is a function of time and acts "through" M.

The elastance, or stiffness, K provides a restoring force as represented by a spring. Thus, if stretched, the string tries to contract; if compressed, it tries to expand to its normal length. The reaction force f_K on each end of the spring is the same and is equal to the product of the stiffness K and the amount of deformation of the spring. The network representation of a spring is shown in Fig. 2.9*b*. The displacement of each end of the spring is measured from the original or equilibrium position. End c has a position x_c, and end d has a position x_d, measured from the respective equilibrium positions. The force equation, in accordance with Hooke's law, is

$$f_K = K(x_c - x_d) \tag{2.53}$$

If the end d is stationary, then $x_d = 0$ and the preceding equation reduces to

$$f_K = Kx_c \tag{2.54}$$

The plot f_K vs. x_c for a real spring is not usually a straight line, because the spring characteristic is nonlinear. However, over a limited region of operation, the linear approximation, i.e., a constant value for K, gives satisfactory results.

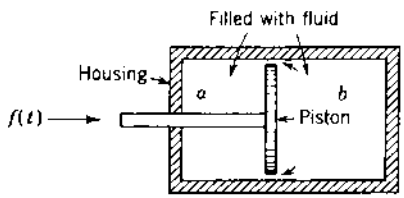


FIGURE 2.10 Dashpot construction.

The damping, or viscous, friction B characterizes the element that absorbs energy. The damping force is proportional to the difference in velocity of two bodies. The assumption that the viscous friction is linear simplifies the solution of the dynamic equation. The network representation of damping action is a dashpot, as shown in Fig. 2.9*c*. Damping may either be intentional or occur unintentionally and is present because of physical construction. The reaction damping force f_B is approximated by the product of damping B and the relative velocity of the two ends of the dashpot. The direction of this force, given by Eq. (2.55), depends on the relative magnitudes and directions of the velocities Dx_e and Dx_f:

$$f_B = B(v_e - v_f) = B(Dx_e - Dx_f) \tag{2.55}$$

Damping may be added to a system by use of a dashpot. The basic operation of a dashpot, in which the housing is filled with an incompressible fluid, is shown in Fig. 2.10. If a force f is applied to the shaft, the piston presses against the fluid, increasing the pressure on side b and decreasing the pressure on side a. As a result, the fluid flows around the piston from side b to side a. If necessary, a small hole can be drilled through the piston to provide a positive path for the flow of fluid. The damping force required to move the piston inside the housing is given by Eq. (2.55), where the damping B depends on the dimensions and the fluid used.

Before the differential equations of a complete system can be written, the mechanical network must first be drawn. This is done by connecting the terminals of elements that have the same displacement. Then the force equation is written for each node or position by equating the sum of the forces at each position to zero. The equations are similar to the node equations in an electric circuit, with force analogous to current, velocity analogous to voltage, and the mechanical elements with their appropriate operators analogous to admittance. The reference position in all of the following examples should be taken from the static equilibrium positions. The force of gravity therefore does not appear in the system equations. The U.S. customary and metric systems of units are shown in Table 2.3.

TABLE 2.3 Mechanical Translation Symbols and Units

Symbol	Quantity	U.S. customary units	Metric units
f	Force	Pounds	Newtons
x	Distance	Feet	Meters
v	Velocity	Feet/second	Meters/second
a	Acceleration	Feet/second^2	Meters/second^2
M†	Mass	Slugs $= \dfrac{\text{pound-seconds}^2}{\text{foot}}$	Kilograms
K	Stiffness coefficient	Pounds/foot	Newtons/meter
B	Damping coefficient	Pounds/(foot/second)	Newtons/(meter/second)

†Mass M in the U.S. customary system above has the dimensions of slugs. Sometimes it is given in units of pounds. If so, then in order to use the consistent set of units above, the mass must be expressed in slugs by using the conversion factor 1 slug = 32.2 lb.

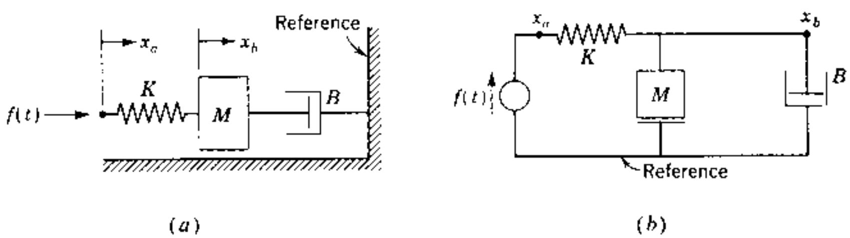


FIGURE 2.11 (*a*) Simple mass-spring-damper mechanical system; (*b*) corresponding mechanical network.

Simple Mechanical Translation System

The system shown in Fig. 2.11 is initially at rest. The end of the spring and the mass have positions denoted as the reference positions, and any displacements from these reference positions are labeled x_a and x_b, respectively. A force f applied at the end of the spring must be balanced by a compression of the spring. The same force is also transmitted through the spring and acts at point x_b.

To draw the mechanical network, the first step is to locate the points x_a and x_b and the reference. The network elements are then connected between these points. For example, one end of the spring has the position x_a, and the other end has the position x_b. Therefore, the spring is connected between these points. The complete mechanical network is drawn in Fig. 2.11*b*.

The displacements x_a and x_b are nodes of the circuit. At each node the sum of the forces must add to zero. Accordingly, the equations may be written for nodes a and b as

$$f = f_K = K(x_a - x_b) \tag{2.56}$$

$$f_K = f_M + f_B = MD^2 x_b + BDx_b \tag{2.57}$$

These two equations can be solved for the two displacements x_a and x_b and their respective velocities $v_a = Dx_a$ and $v_b = Dx_b$.

It is possible to obtain one equation relating x_a to f, x_b to x_a, or x_b to f by combining Eqs. (2.57) and (2.58):

$$K(MD^2 + BD)x_a = (MD^2 + BD + K)f \tag{2.58}$$

$$(MD^2 + BD + K)x_b = Kx_a \tag{2.59}$$

$$(MD^2 + BD)x_b = f \tag{2.60}$$

The solution of Eq. (2.59) shows the motion x_b resulting from a given motion x_a. Also, the solution of Eqs. (2.58) and (2.60) show the motions x_a and x_b, respectively, resulting from a given force f. From each of these three equations the following transfer functions are obtained:

$$G_1 = \frac{x_a}{f} = \frac{MD^2 + BD + K}{K(MD^2 + BD)} \tag{2.61}$$

$$G_2 = \frac{x_b}{x_a} = \frac{K}{MD^2 + BD + K} \tag{2.62}$$

$$G = \frac{x_b}{f} = \frac{1}{MD^2 + BD} \tag{2.63}$$

Note that the last equation is equal to the product of the first two, i.e.,

$$G = G_1 G_2 = \frac{x_a}{f}\frac{x_b}{x_a} = \frac{x_b}{f} \tag{2.64}$$

The block diagram representing the mathematical operation of Eq. (2.64) is shown in Fig. 2.12. Figure 2.12a is a detailed representation that indicates all variables in the system. The two block G_1 and G_2 are said to be in *cascade*. Figure 2.12b, called the *overall block diagram* representation, shows only the input f and the output x_b, where x_b is considered the output variable of the system of Fig. 2.11.

The multiplication of transfer functions, as in Eq. (2.64), is valid as long as there is no coupling or loading between the two blocks in Fig. 2.12a.

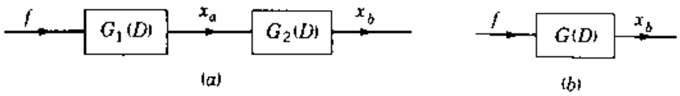


FIGURE 2.12 Block diagram representation of Fig. 2.11: (a) detailed; (b) overall.

The signal x_a is unaffected by the presence of the block having the transfer function G_2; thus the multiplication is valid. When electric circuits are coupled, the transfer functions may not be independent unless they are isolated by an electronic amplifier with a very high input impedance.

Example 1. Determine the state equations for Eq. (2.57). Equation (2.57) involves only one enery-storage element, the mass M, whose energy variable is v_b. The output quantity in this system is the position $y = x_b$. Since this quantity is not one of the physical or energy-related state variables, it is necessary to increase the number of state variables to 2; i.e., $x_b = x_1$ is an augmented state variable. Equation (2.57) is of second order, which confirms that two state variables are required. Note that the spring constant does not appear in this equation since the force f is transmitted through the spring to the mass. With $x_1 = x_b, x_2 = v_b = \dot{x}_1, u = f$, and $y = x_1$, the state and output equations are

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 \\ 0 & -\dfrac{B}{M} \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ \dfrac{1}{M} \end{bmatrix} \mathbf{u} = \mathbf{A}\mathbf{x} + \mathbf{b}\mathbf{u} \qquad \mathbf{y} = [1 \quad 0]\ \mathbf{x} = \mathbf{c}^T \mathbf{x} \tag{2.65}$$

Example 2. Determine the state equations for Eq. (2.59). Equation (2.59) involves two energy-storage elements, K and M, whose energy variables are x_b and v_b, respectively. Note that the spring constant K does appear in this equation since x_a is the input u. Therefore, a state variable must be associated with this energy element. Thus, the assigned state variables are x_b and $v_b = Dx_b$, which are independent state variables of this system. Let $x_1 = x_b = y$ and $x_2 = v_b = \dot{x}_1$. Converting Eq. (2.59) into state-variable form yields the two equations

$$\dot{x}_1 = x_2 \qquad \dot{x}_2 = -\frac{K}{M}x_1 - \frac{B}{M}x_2 + \frac{K}{M}u \tag{2.66}$$

They can be put into the matrix form

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}\mathbf{u} \qquad \mathbf{y} = \mathbf{c}^T\mathbf{x} \tag{2.67}$$

Multiple-Element Mechanical Translation System

A force $f(t)$ is applied to the mass M_1 of Fig. 2.13a. The sliding friction between the masses M_1 and M_2 and the surface is indicated by the viscous friction coefficients B_1 and B_2. The system equations can be written in terms of the two displacements x_a and x_b. The mechanical network is drawn by connecting the terminals of the elements that have the same displacement (see Fig. 2.13b). Since the sum of the forces at each node must add to zero, the equations are

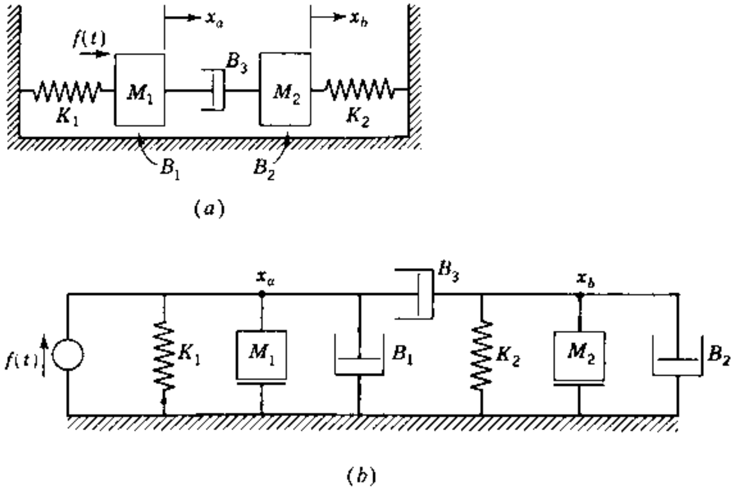


FIGURE 2.13 (*a*) Multiple-element mechanical system: (*b*) corresponding mechanical network.

written according to the rules for node equations:

$$\text{For node } a: \quad (M_1D^2 + B_1D + B_3D + K_1)x_a - (B_3D)x_b = f \tag{2.68}$$

$$\text{For node } b: \quad -(B_3D)x_a + (M_2D^2 + B_2D + B_3D + K_2)x_b = 0 \tag{2.69}$$

A definite pattern to these equations can be detected. Observe that K_1, M_1, B_1, and B_3 are connected to node a and that Eq. (2.68), for node a, contains all four of these terms as coefficients of x_a. Notice that element B_3 is also connected to node b and that the term $-B_3$ appears as a coefficient of x_b. When this pattern is used, Eq. (2.69) can be written directly. Thus, since K_2, M_2, B_2, and B_3 are connected to node b, they appear as coefficients of x_b. Element B_3 is also connected to node a, and $-B_3$ appears as the coefficient of x_a. Each term in the equation must be a force.

The node equations for a mechanical system follow directly from the mechanical network of Fig. 2.13*b*. They are similar in form to the node equations for an electric circuit and follow the same rules.

The block diagram representing the system in Fig. 2.13 is also given by Fig. 2.12, where again the system output is x_b. The transfer functions $G_1 = x_a/f$ and $G_2 = x_b/x_a$ are obtained by solving the equations for this system.

Example 3. Obtain the state equations for the system of Fig. 2.13, where x_b is the system output. There are four energy-storage elements, thus the four assigned state variables are x_a, x_b, Dx_a and Dx_b. Analyzing Eqs. (2.68) and

(2.69) shows that this is a fourth-order system. Let

$$x_1 = x_b \quad \text{for spring } K_2 \qquad x_2 = \dot{x}_1 = v_b \quad \text{for mass } M_2$$
$$x_3 = x_a \quad \text{for spring } K_1 \qquad x_4 = \dot{x}_3 = v_a \quad \text{for mass } M_1$$
$$u = f \qquad y = x_b = x_1$$

The four state variables are independent, i.e., no one state variable can be expressed in terms of the remaining state variables. The system equations are

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}\mathbf{u} \qquad \mathbf{y} = \mathbf{c}^T\mathbf{x} \tag{2.70}$$

where,

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -\dfrac{K_2}{M_2} & -\dfrac{B_2+B_3}{M_2} & 0 & \dfrac{B_3}{M_2} \\ 0 & 0 & 0 & 1 \\ 0 & \dfrac{B_3}{M_1} & -\dfrac{K_1}{M_1} & -\dfrac{B_1+B_3}{M_1} \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \dfrac{1}{M_1} \end{bmatrix}$$

$$c_T = [1 \quad 0 \quad 0 \quad 0]$$

2.6 ANALOGOUS CIRCUITS

Analogous circuits represent systems for which the differential equations have the same form. The corresponding variables and parameters in two circuits represented by equations of the same form are called *analogs.* An electric circuit can be drawn that looks like the mechanical circuit and is represented by node equations that have the same mathematical form as the mechanical equations. The analogs are listed in Table 2.4. In this table the force f and the current i are analogs and are classified as "through" variables. There is a

TABLE 2.4 Electrical and Mechanical Analogs

Mechanical translation element		Electrical element	
Symbol	Quantity	Symbol	Quantity
f	Force	i	Current
$v = Dx$	Velocity	e or v	Voltage
M	Mass	C	Capacitance
K	Stiffness coefficient	$1/L$	Reciprocal inductance
B	Damping coefficient	$G = 1/R$	Conductance

physical similarity between the two, because a force indicator must be placed in series with the system. Also, the velocity "across" a mechanical element is analogous to voltage across an electrical element. Again, there is a physical similarity, because a measuring instrument must be placed across the system in both cases. A voltmeter must be placed across a circuit to measure voltage; it must have a point of reference. A velocity indicator must also have a point of reference. Nodes in the mechanical network are analogous to nodes in the electric network.

2.7 MECHANICAL ROTATIONAL SYSTEMS

The equations characterizing rotational systems are similar to those for translation systems. Writing torque equations parallels the writing of force equations, where the displacement, velocity, and acceleration terms are now angular quantities. The applied torque is equal to the sum of the reaction torques. The three elements in a rotational system are inertia, the spring, and the dashpot. The mechanical-network representation of these elements is shown in Fig. 2.14.

The torque applied to a body having a moment of inertia J produces an angular acceleration. The reaction torque T_J is opposite to the direction of the applied torque and is equal to the product of moment of inertia and acceleration. In terms of angular acceleration α, angular velocity ω, or angular displacement θ, the torque equation is

$$T_J = J\alpha = J\,D\omega = J\,D^2\theta \tag{2.71}$$

When a torque is applied to a spring, the spring is twisted by an angel θ. The applied torque is transmitted through the spring and appears at the other end. The reaction spring torque T_K that is produced is equal to the product of the *stiffness*, or *elastance*, K of the spring and the angle of twist. By denoting the positions of the two ends of the spring, measured from the neutral position, as θ_c and θ_d, the reaction torque is

$$T_K = K(\theta_c - \theta_d) \tag{2.72}$$

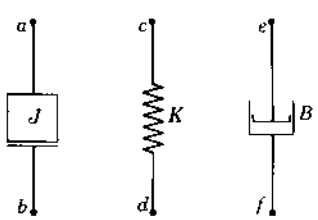


FIGURE 2.14 Network elements of mechanical rotational systems.

Damping occurs whenever a body moves through a fluid, which may be a liquid or a gas such as air. To produce motion of the body, a torque must be applied to overcome the reaction damping torque. The damping is represented as a dashpot with a *viscous friction coefficient B*. The damping torque T_B is equal to the product of damping B and the relative angular velocity of the ends of the dashpot. The reaction torque of a damper is

$$T_B = B(\omega_e - \omega_f) = B(D\theta_e - D\theta_f) \tag{2.73}$$

Writing the differential equations is simplified by first drawing the mechanical network for the system. This is done by first designating nodes that correspond to each angular displacement. Then each element is connected between the nodes that correspond to the two motions at each end of that element. The inertia elements are connected from the reference node to the node representing its position. The spring and dashpot elements are connected to the two nodes that represent the position of each end of the element. Then the torque equation is written for each node by equating the sum of the torques at each node to zero. These equations are similar to those for mechanical translation and are analogous to those for electric circuits. The units for these elements are given in Table 2.5.

An electrical analog can be obtained for a mechanical rotational system in the manner described in Sec. 2.7. The changes to Table 2.4 are due only to the fact that rotational quantities are involved. Thus torque is the analog of current, and the moment of inertia is the analog of capacitance.

TABLE 2.5 Mechanical Rotational Symbols and Units

Symbol	Quantity	U.S. customary units	Metric units
T	Torque	Pound-feet	Newton-meters
θ	Angle	Radians	Radians
ω	Angular velocity	Radians/second	Radians/second
α	Angular acceleration	Radians/second^2	Radians/second^2
J^*	Moment of inertia	Slug-feet^2 (or pound-feet-seconds^2)	Kilogram-meters^2
K	Stiffness coefficient	Pound-feet/radian	Newton-meters/radian
B	Damping coefficient	Pound-feet/(radian/second)	Newton-meters/radians/second)

*The moment of inertia J has the dimensions of mass-distance^2, which have the units slug feet^2 in the U.S. customary system. Sometimes the units are given as pound-feet^2. To use the consistent set of units above, the moment of inertia must be expressed in slug-feet^2 by using the conversion factor 1 slug = 32.2 lb.

Simple Mechanical Rotational System

The system shown in Fig. 2.15 has a mass, with a moment of inertia J, immersed in a fluid. A torque T is applied to the mass. The wire produces a reactive torque proportional to its stiffness K and to the angle of twist. The fins moving through the fluid having a damping B, which requires a torque proportional to the rate at which they are moving. The mechanical network is drawn in Fig. 2.15b. There is one node having a displacement θ; therefore, only one equation is necessary:

$$J\,D^2\theta + B\,D\theta + K\theta = T(t) \tag{2.74}$$

Multiple-Element Mechanical Rotational System

The system represented by Fig. 2.16a has two disks that have damping between them and also between each of them and the frame. The mechanical

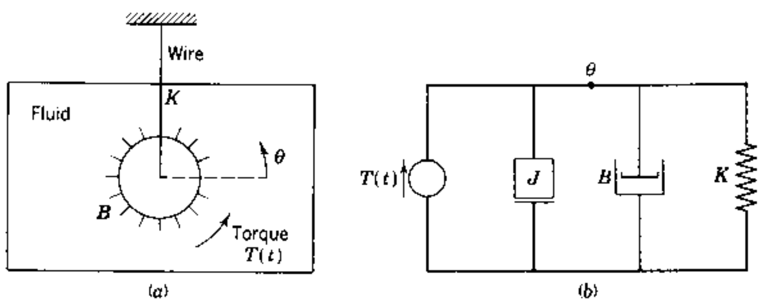


FIGURE 2.15 (a) Simple rotational system; (b) mechanical network.

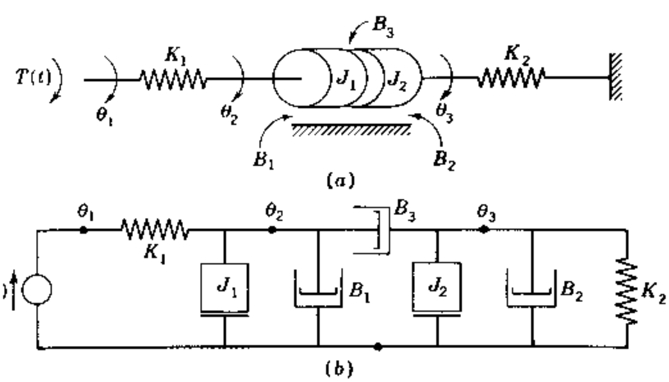


FIGURE 2.16 (a) Rotational system; (b) corresponding mechanical network.

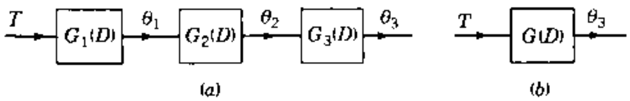


FIGURE 2.17 Detailed and overall block diagram representations of Fig. 2.16.

network is drawn in Fig. 2.16*b* by first identifying the three angular displacements θ_1, θ_2, and θ_3. Then the elements are connected to the proper nodes. The equations are written directly in systematized form, since the torques at each node must add to zero, as follows:

$$\text{Node 1:}\quad K_1\theta_1 - K_1\theta_2 = T(t) \tag{2.75}$$

$$\text{Node 2:}\quad -K_1\theta_1 + \left[J_1D^2 + (B_1 + B_3)D + K_1\right]\theta_2 - (B_3D)\theta_3 = 0 \tag{2.76}$$

$$\text{Node 3:}\quad -(B_3D)\theta_2 + \left[J_2D^2 + (B_2 + B_3)D + K_2\right]\theta_3 = 0 \tag{2.77}$$

These three equations can be solved simultaneously for θ_1, θ_2, and θ_3 as a function of the applied torque. If θ_3 is the system output, these equations can be solved for the following four transfer functions:

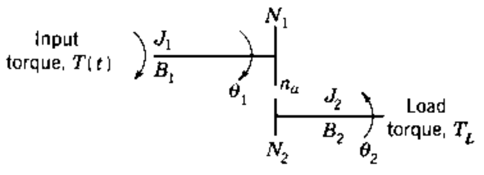
$$G_1 = \frac{\theta_1}{T} \qquad G_2 = \frac{\theta_2}{\theta_1} \qquad G_3 = \frac{\theta_3}{\theta_2} \qquad G = G_1G_2G_3 = \frac{\theta_1}{T}\frac{\theta_2}{\theta_1}\frac{\theta_3}{\theta_2} = \frac{\theta_3}{T} \tag{2.78}$$

The detailed and overall block diagram representations of Fig. 2.16 are shown in Fig. 2.17. The overall transfer function G given by Eqs. (2.78) is the product of the transfer functions, which are said to be *in cascade*. This product of transfer functions applies in general to any number of elements in cascade when there is no loading between the blocks.

Example. Determine the state variables for the system of Fig. 2.16, where θ_3 is the system output y and torque T is the system input u. The spring K_1 is effectively not in the system because the torque is transmitted through the spring. Since there are three energy-storage elements J_1, J_2, and K_2, the assigned state variables are $D\theta_2$, $D\theta_3$, and θ_3. These are independent state variables. Therefore, let $x_1 = \theta_3$, $x_2 = D\theta_3$, and $x_3 = D\theta_2$. If the system input is $u = \theta_1$, then the spring K_1 is included, resulting in four state variables (see Prob. 2.17).

2.8 EFFECTIVE MOMENT OF INERTIA AND DAMPING OF A GEAR TRAIN

When the load is coupled to a drive motor through a gear train, the moment of inertia and damping relative to the motor are important. Since the shaft length



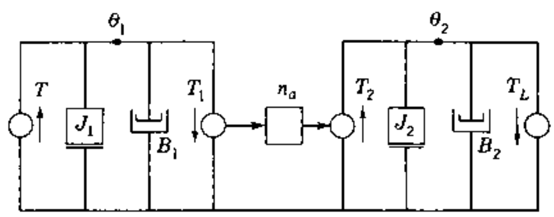


FIGURE 2.18 (*a*) Representation of a gear train; (*b*) corresponding mechanical network.

between gears is very short, the stiffness may be considered infinite. A simple representation of a gear train is shown in Fig. 2.18*a*. The following definitions are used:

$N =$ number of teeth on each gear
$\omega = D\theta =$ velocity of each gear
$n_a =$ ratio of (speed of driving shaft)/(speed of driven shaft)
$\theta =$ angular position

Typically, gearing may not mesh perfectly, and the teeth of one gear may not fill the full space between the gear teeth of the matching gear. As a consequence, there can be times when the drive gear is not in contact with the other gear. This nonlinear characteristic is called *dead zone*. When the dead zone is very small, it is neglected for the linearized gearing model.

The mechanical network for the gear train is shown in Fig. 2.18*b*. At each gear pair, two torques are produced. For example, a restraining torque T_1 is the load on gear 1 produced by the rest of the gear train. A driving torque T_2 is also produced on gear 2. T_2 is the torque transmitted to gear 2 to drive the rest of the gear train. These torques are inversely proportional to the ratio of the speeds of the respective gears. The block labeled n_a between T_1 and T_2 is used to show the relationship between them; that is, $T_2 = n_a T_1$. A transformer may be considered as the electrical analog of the gear train, with angular velocity analogous to voltage and torque analogous to current.

The equations describing the system are

$$J_1 D^2\theta_1 + B_1 D\theta_1 + T_1 = T \qquad J_2 D^2\theta_2 + B_2 D\theta_2 + T_L = T_2$$
$$n_a = \frac{\omega_1}{\omega_2} = \frac{\theta_1}{\theta_2} = \frac{N_2}{N_1} \qquad \theta_2 = \frac{\theta_1}{n_a} \qquad T_2 = n_a T_1 \tag{2.79}$$

The equations can be combined to produce

$$J_1 D^2\theta_1 + B_1 D\theta_1 + \frac{1}{n_a}\left(J_2 D^2\theta_2 + B_2 D\theta_2 + T_L\right) = T \tag{2.80}$$

This equation can be expressed in terms of the torque and the input position θ_1 only:

$$\left(J_1 + \frac{J_2}{n_a^2}\right)D^2\theta_1 + \left(B_a + \frac{B_2}{n_a^2}\right)D\theta_1 + \frac{T_L}{n_a} = T \tag{2.81}$$

Equation (2.81) represents the system performance as a function of a single dependent variable. An equivalent system is one having an equivalent moment of inertia and damping equal to

$$J_{\text{eq}} = J_1 + \frac{J_2}{n_a^2} \tag{2.82}$$

$$B_{\text{eq}} = B_1 + \frac{B_2}{n_a^2} \tag{2.83}$$

If the solution for θ_2 is wanted, the equation can be altered by the substitution of $\theta_1 = n_a\theta_2$. This system can be generalized for any number of gear stages. When the gear-reduction ratio is large, the load moment of inertia may contribute a negligible value to the equivalent moment of inertia. When a motor rotor is geared to a long shaft that provides translational motion, then the elastance or stiffness K of the shaft must be taken into account in deriving the system's differential equations (see Prob. 2.24).

2.9 THERMAL SYSTEMS [11]

A limited number of thermal systems can be represented by linear differential equations. The basic requirement is that the temperature of a body be considered uniform. This approximation is valid when the body is small. Also, when the region consists of a body of air or liquid, the temperature can be considered uniform if there is perfect mixing of the fluid. The necessary condition of equilibrium requires that the heat added to the system equal the heat stored plus the heat carried away. This requirement can also be expressed in terms of rate of heat flow.

The symbols shown in Table 2.6 are used for thermal systems. A thermal-system network is drawn for each system in which thermal capacitance and

TABLE 2.6 Thermal Symbols and Units

Symbol	Quantity	U.S. customary units	Metric units
q	Rate of heat flow	Btu/minute	Joules/second
M	Mass	Pounds	Kilograms
S	Specific heat	Btu/(pounds)(°F)	Joules/(kilogram)(°C)
C	Thermal capacitance $C = MS$	Btu/°F	Joules/°C
K	Thermal conductance	Btu/(minute)(°F)	Joules/(second)(°C)
R	Thermal resistance	Degrees/ (Btu/minute)	Degrees/ (joule/second)
θ	Temperature	°F	°C
h	Heat energy	Btu	Joules

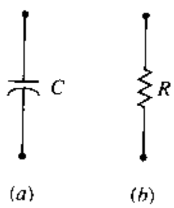


FIGURE 2.19 Network elements of thermal systems.

thermal resistance are represented by network elements. The differential equations can then be written form the thermal network.

The additional heat stored in a body whose temperature is raised from θ_1 to θ_2 is given by

$$h = \frac{q}{D} = C(\theta_2 - \theta_1) \tag{2.84}$$

In terms of rate of heat flow, this equation can be written as

$$q = CD(\theta_2 - \theta_1) \tag{2.85}$$

The thermal capacitance determines the amount of heat stored in a body. It is analogous to the electric capacitance of a capacitor in an electric circuit, which determines the amount of charge stored. The network representation of thermal capacitance is shown in Fig. 2.19a.

Rate of heat flow through a body in terms of the two boundary temperatures θ_3 and θ_4 is

$$q = \frac{\theta_3 - \theta_4}{R} \tag{2.86}$$

The thermal resistance determines the rate of heat flow through the body. This is analogous to the resistance of a resistor in an electric circuit, which determines the current flow. The network representation of thermal resistance is shown in Fig. 2.19*b*. In the thermal network the temperature is analogous to potential.

Simple Mercury Thermometer

Consider a thin glass-walled thermometer filled with mercury that has stabilized at a temperature θ_1. It is plunged into a bath of temperature θ_0 at $t = 0$. In its simplest form, the thermometer can be considered to have a capacitance C that stores heat and a resistance R that limits the heat flow. The temperature at the center of the mercury is θ_m. The flow of heat into the thermometer is

$$q = \frac{\theta_0 - \theta_m}{R} \tag{2.87}$$

The heat entering the thermometer is stored in the thermal capacitance and is given by

$$h = \frac{q}{D} = C(\theta_m - \theta_1) \tag{2.88}$$

These equations can be combined to form

$$\frac{\theta_0 - \theta_m}{RD} = C(\theta_m - \theta_1) \tag{2.89}$$

Differentiating Eq. (2.89) and rearranging terms gives

$$RC\,D\theta_m + \theta_m = \theta_0 \tag{2.90}$$

The thermal network is drawn in Fig. 2.20. The node equation for this circuit, with the temperature considered as a voltage, gives Eq. (2.90) directly. From this equation the transfer function $G = \theta_m/\theta_0$ may be obtained. Since there is one energy-storage element, the independent state variable is $x_1 = \theta_m$ and the input is $u = \theta_0$. Thus the state equation is

$$\dot{x}_1 = -\frac{1}{RC}x_1 + \frac{1}{RC}u \tag{2.91}$$

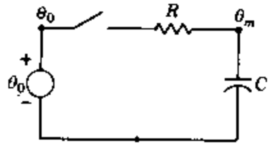


FIGURE 2.20 Simple network representation of a thermometer.

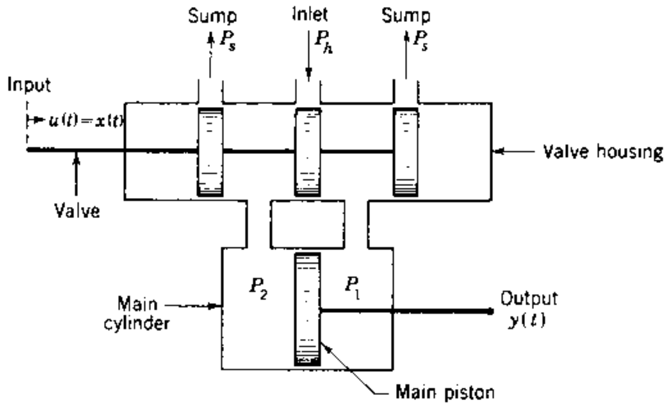


FIGURE 2.21 Hydraulic actuator.

A more exact model of a mercury thermometer is presented in Ref. 11. A simple transfer system for a water tank heater is also presented in this reference.

2.10 HYDRAULIC LINEAR ACTUATOR

The valve-controlled hydraulic actuator is used in many applications as a force amplifier. Very little force is required to position the valve, but a large output force is controlled. The hydraulic unit is relatively small, which makes its use very attractive. Figure 2.21 shows a simple hydraulic actuator in which motion of the valve regulates the flow of oil to either side of the main cylinder. An input motion x of a few thousandths of an inch results in a large change of oil flow. The resulting difference in pressure on the main piston causes motion of the output shaft. The oil flowing in is supplied by a source that maintains a constant high pressure P_h, and the oil on the opposite side of the piston flows into the drain (sump) at low pressure P_s. The load-induced pressure P_L is the difference between the pressures on each side of the main piston:

$$P_L = P_1 - P_2 \tag{2.92}$$

The flow of fluid through an inlet orifice is given by [12]

$$q = ca\sqrt{2g\frac{\Delta p}{w}} \tag{2.93}$$

where

c = orifice coefficient
a = orifice area
Δp = pressure drop across orifice

$w =$ specific weight of fluid
$g =$ gravitational acceleration constant
$q =$ rate of flow of fluid

Simplified Analysis

As a first-order approximation, it can be assumed that the orifice coefficient and the pressure drop across the orifice are constant and independent of valve position. Also, the orifice are a is proportional to the valve displacement x. Equation (2.93), which gives the rate of flow of hydraulic fluid through the valve, can be rewritten in linearized form as

$$q = C_x x \tag{2.94}$$

where x is the displacement of the valve. The displacement of the main piston is directly proportional to the volume of fluid that enters the main cylinder, or, equivalently, the rate of flow of fluid is proportional to the rate of displacement of the output. When the compressibility of the fluid and the leakage around the valve and main piston are neglected, the equation of motion of the main piston is

$$q = C_b D y \tag{2.95}$$

Combining the two equations gives

$$Dy = \frac{C_x}{C_b} x = C_1 x \tag{2.96}$$

This analysis is essentially correct when the load reaction is small.

More Complete Analysis

When the load reaction is not negligible, a load pressure P_L must be produced across the piston. A more complete analysis takes into account the pressure drop across the orifice, the leakage of oil around the piston, and the compressibility of the oil. The pressure drop Δp across the orifice is a function of the source pressure P_h and the load pressure P_L. Since P_h is assumed constant, the flow equation for q is a function of valve displacement x and load pressure P_L:

$$q = f(x, P_L) \tag{2.97}$$

The differential rate of flow dq, expressed in terms of partial derivatives, is

$$dq = \frac{\partial q}{\partial x} dx + \frac{\partial q}{\partial P_L} dP_L \tag{2.98}$$

If q, x, and P_L are measured from zero values as reference points, and if the partial derivatives are constant at the values they have at zero, the integration of Eq. (2.98) gives

$$q = \left(\frac{\partial q}{\partial x}\right)_0 x + \left(\frac{\partial q}{\partial P_L}\right)_0 P_L \tag{2.99}$$

By defining

$$C_x \equiv \left(\frac{\partial p}{\partial x}\right)_0 \qquad \text{and} \qquad C_p \equiv \left(\frac{-\partial q}{\partial P_L}\right)_0$$

the flow equation for fluid entering the main cylinder can be written as

$$q = C_x x - C_p P_L \tag{2.100}$$

Both C_x and C_p have positive values. A comparison of Eq. (2.100) with Eq. (2.94) shows that the load pressure reduces the flow into the main cylinder. The flow of fluid into the cylinder must satisfy the continuity conditions of equilibrium. This flow is equal to the sum of the components:

$$q = q_0 + q_l + q_c \tag{2.101}$$

where

$q_0 =$ incompressible component (causes motion of piston)
$q_l =$ leakage component
$q_c =$ compressible component

The component q_0, which produces a motion y of the main piston, is

$$q_0 = C_b\, Dy \tag{2.102}$$

The compressible component is derived in terms of the bulk modulus of elasticity of the fluid, which is defined as the ratio of incremental stress to incremental strain.

Thus

$$K_B = \frac{\Delta P_L}{\Delta V / V} \tag{2.103}$$

Solving for ΔV and dividing both sides of the equation by Δt gives

$$\frac{\Delta V}{\Delta t} = \frac{V}{K_B}\frac{\Delta P_L}{\Delta t} \tag{2.104}$$

Taking the limit as Δt approaches zero and letting $q_c = dV/dt$ gives

$$q_c = \frac{V}{K_B} DP_L \tag{2.105}$$

where V is the effective volume of fluid under compression and K_B is the bulk modulus of the hydraulic oil. The volume V at the middle position of the piston stroke is often used in order to linearize the differential equation.

The leakage components is

$$q_l = LP_L \tag{2.106}$$

where L is the leakage coefficient of the whole system.

Combining these equations gives

$$q = C_x x - C_p P_L = C_b\ Dy + \frac{V}{K_B} DP_L + LP_L \tag{2.107}$$

and rearranging terms gives

$$C_b\ Dy + \frac{V}{K_B} DP_L + (L + C_p)P_L = C_x x \tag{2.108}$$

The force developed by the main piston is

$$F = n_F AP_L = CP_L \tag{2.109}$$

where n_F is the force conversion efficiency of the unit and A is the area of the main actuator piston.

An example of a specific type of load, consisting of a mass and a dashpot, is shown in Fig. 2.22. The equation for this system is obtained by equating the force produced by the piston, which is given by Eq. (2.109), to the reactive load forces:

$$F = M\,D^2 y + B\,Dy = CP_L \tag{2.110}$$

Substituting the value of P_L from Eq. (2.110) into Eq. (2.108) gives the equation relating the input motion x to the response y:

$$\frac{MV}{CK_B} D^3 y + \left[\frac{BV}{CK_B} + \frac{M}{C}(L + C_p)\right] D^2 y + \left[C_b + \frac{B}{C}(L + C_p)\right] Dy = C_x x \tag{2.111}$$

The preceding analysis is based on perturbations about the reference set of values $x = 0$, $q = 0$, $P_L = 0$. For the entire range of motion x of the valve, the quantities $\partial q/\partial x$ and $-\partial q/\partial P_L$ can be determined experimentally. Although

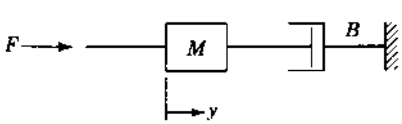


FIGURE 2.22 Load on a hydraulic piston.

they are not constant at values equal to the values C_x and C_p at the zero reference point, average values can be assumed in order to simulate the system by linear equations. For conservative design the volume V is determined for the main piston at the midpoint.

To write the state equation for the hydraulic actuator and load of Figs. 2.21 and 2.22, the energy-related variables must be determined. The mass M has the output velocity Dy as an energy-storage variable. The compressible component q_c also represents an energy-storage element in a hydraulic system. The compression of a fluid produces stored energy, just as in the compression of a spring. The equation for hydraulic energy is

$$E(t) = \int_0^t P(\tau)q(\tau)\,d\tau \tag{2.112}$$

where $P(t)$ is the pressure and $q(t)$ is the rate of flow of fluid. The energy storage in a compressed fluid is obtained in terms of the bulk modulus of elasticity K_B. Combining Eq. (2.105) with Eq. (2.112) for a constant volume yields

$$E_c(P_L) = \int_0^{P_L} \frac{V}{K_B} P_L\, dP_L = \frac{V}{2K_B} P_L^2 \tag{2.113}$$

The stored energy in a compressed fluid is proportional to the pressure P_L squared; thus P_L may be used as a physical state variable.

Since the output quantity in this system is the position y, it is necessary to increase the state variables to three. Further evidence of the need for three state variables is the fact that Eq. (2.111) is a third-order equation. Therefore, in this example, let $x_1 = y$, $x_2 = Dy = \dot{x}_1$, $x_3 = P_L$, and $u = x$. Then, from Eqs. (2.108) and (2.110), the state and output equations are

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -\dfrac{B}{M} & \dfrac{C}{M} \\ 0 & -\dfrac{C_b K_B}{V} & -\dfrac{K_B(L + C_p)}{V} \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ 0 \\ \dfrac{K_B C_X}{V} \end{bmatrix} \mathbf{u} \tag{2.114}$$

$$\mathbf{y} = [y] = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \mathbf{x} = [x_1] \tag{2.115}$$

The effect of augmenting the state variables by adding the piston displacement $x_1 = y$ is to produce a singular system; that is, $|\mathbf{A}| = 0$. This property does not appear if a spring is added to the load, as shown in Prob. 2.19. In that case $x_1 = y$ is an independent state variable.

A detailed derivation of a positive-displacement rotational hydraulic transmission is presented in Ref. 1.

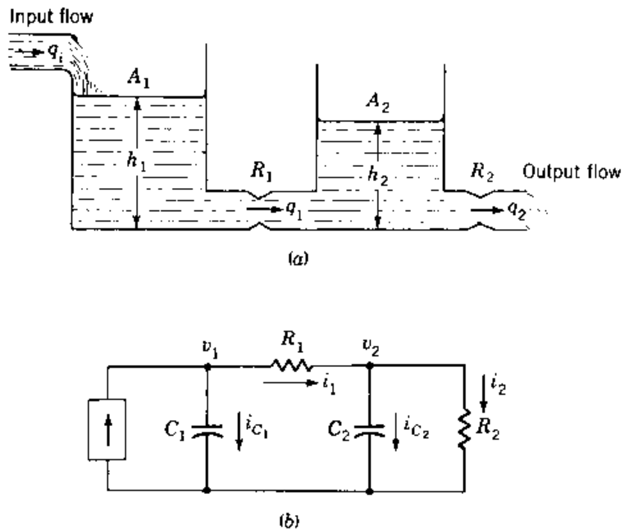


FIGURE 2.23 Liquid-level system and its equivalent electrical analog.

2.11 LIQUID-LEVEL SYSTEM [12,13]

Figure 2.23*a* represents a two-tank liquid-level control system. Definitions of the system parameters are

$q_i, q_1, q_2 =$ rates of flow of fluid $\quad h_1, h_2 =$ heights of fluid level
$R_1, R_2 =$ flow resistance $\quad A_1, A_2 =$ cross-sectional tank areas

The following basic linear relationships hold for this system:

$$q = \frac{h}{R} = \text{rate of flow through orifice} \tag{2.116}$$

$$\begin{aligned} q_n &= (\text{tank input rate of flow}) - (\text{tank output rate of flow}) \\ &= \text{net tank rate of flow} = A\,Dh \end{aligned} \tag{2.117}$$

Applying Eq. (2.117) to tanks 1 and 2 yields, respectively,

$$q_{n_1} = A_1\ Dh_1 = q_i - q_1 = q_i - \frac{h_1 - h_2}{R_1} \tag{2.118}$$

$$q_{n_2} = A_2\ Dh_2 = q_1 - q_2 = \frac{h_1 - h_2}{R_1} - \frac{h_2}{R_2} \tag{2.119}$$

These equations can be solved simultaneously to obtain the transfer functions h_1/q_i and h_2/q_i.

TABLE 2.7 Hydraulic and Electrical Analogs

Hydraulic element		Electrical element	
Symbol	Quantity	Symbol	Quantity
q_i	Input flow rate	i_i	Current source
h	Height	v_C	Capacitor voltage
A	Tank area	C	Capacitance
R	Flow resistance	R	Resistance

The energy stored in each tank represents potential energy, which is equal to $\rho A h^2/2$, where ρ is the fluid density coefficient. Since there are two tanks, the system has two energy-storage elements, whose energy-storage variables are h_1 and h_2. Letting $x_1 = h_1$, $x_2 = h_2$, and $u = q_i$ in Eqs. (2.118) and (2.119) reveals that x_1 and x_2 are independent state variables. Thus, the state equation is

$$\dot{\mathbf{x}} = \begin{bmatrix} -\dfrac{1}{R_1 A_1} & \dfrac{1}{R_1 A_1} \\ \dfrac{1}{R_1 A_2} & -\dfrac{1}{R_1 A_2} - \dfrac{1}{R_2 A_2} \end{bmatrix} \mathbf{x} + \begin{bmatrix} \dfrac{1}{A_1} \\ 0 \end{bmatrix} \mathbf{u} \tag{2.120}$$

The levels of the two tanks are the outputs of the system. Letting $y_1 = x_1 = h_1$ and $y_2 = x_2 = h_2$ yields

$$y = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x} \tag{2.121}$$

The potential energy of a tank can be represented as a capacitor whose stored potential energy is $C v_c^2/2$; thus, the electrical analog of h is v_C. As a consequence, an analysis of Eqs. (2.118) and (2.119) yields the analogs between the hydraulic and electrical quantities listed in Table 2.7. The analogous electrical equations are

$$C_1\, Dv_1 = i_i - \frac{v_1 - v_2}{R_1} \tag{2.122}$$

$$C_2\, Dv_2 = \frac{v_1 - v_2}{R_1} - \frac{v_2}{R_2} \tag{2.123}$$

These two equations yield the analogous electric circuit of Fig. 2.23*b*.

2.12 ROTATING POWER AMPLIFIERS [14,15]

A dc generator can be used as a power amplifier, in which the power required to excite the field circuit is lower than the power output rating of the armature

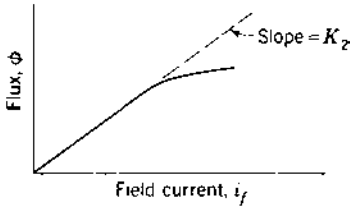


FIGURE 2.24 Magnetization curve.

circuit. The voltage e_g induced in the armature circuit is directly proportional to the product of the magnetic flux ϕ set up by the field and the speed of rotation ω of the armature. Neglecting hysteresis, this is expressed by

$$e_g = K_1 \phi \omega \tag{2.124}$$

The flux is a function of the field current and the type of iron used in the field. A typical magnetization curve showing flux as the function of field current is given in Fig. 2.24. Up to saturation the relation is approximately linear, and the flux is directly proportional to field current:

$$\phi = K_2 i_f \tag{2.125}$$

Combining these equations gives

$$e_g = K_1 K_2 \omega i_f \tag{2.126}$$

When the generator is used as a power amplifier, the armature is driven at constant speed, and this equation becomes

$$e_g = K_g i_f \tag{2.127}$$

A generator is represented schematically in Fig. 2.25, in which L_f and R_f and L_g and R_g are the inductance and resistance of the field and armature circuits, respectively. The equations for the generator are

$$e_f = (L_f\ D + R_f) i_f \qquad e_g = K_g i_f \qquad e_t = e_g - (L_g\ D + R_g) i_a \tag{2.128}$$

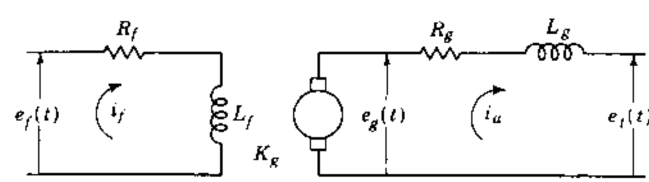


FIGURE 2.25 Schematic diagram of a generator.

The armature current depends on the load connected to the generator terminals. Combining the first two equations gives

$$(L_f\ D + R_f)e_g = K_g e_f \tag{2.129}$$

Equation (2.129) relates the generated voltage e_g to the input field voltage e_f. The electrical power output $p_o = e_t i_a$ is much larger than the power input $p_i = e_f i_f$. Thus, the ratio p_o/p_i represents the electrical power gain. Remember, however, that mechanical power drives the generator armature. Therefore, mechanical power is being converted into electrical power output.

2.13 DC SERVOMOTOR

The development torque of a dc motor is proportional to the magnitude of the flux due to the field current i_f and the armature current i_m. For any given motor the only two adjustable quantities are the flux and the armature current. The developed torque can therefore be expressed as

$$T(t) = K_3 \phi i_m \tag{2.130}$$

To avoid confusion with t for time, the capital letter T is used to indicate torque. It may denote either a constant or a function that varies with time. There are two modes of operation of a servomotor. For the armature control mode, the field current is held constant and an adjustable voltage is applied to the armature. In the field control mode, the armature current is held constant and an adjustable voltage is applied to the field. The armature control method of operation is considered in detail. The reader can refer to Ref. 1 for the field control method of speed control.

A constant field current is obtained by separately exciting the field from a fixed dc source. The flux is produced by the field current and is, therefore, essentially constant. Thus the torque is proportional only to the armature current, and Eq. (2.130) becomes

$$T(t) = K_t i_m \tag{2.131}$$

When the motor armature is rotating, a voltage e_m is induced that is proportional to the product of flux and speed. Because the polarity of this voltage opposes the applied voltage e_a, it is commonly called the *back emf*. Since the flux is held constant, the induced voltage e_m is directly proportional to the speed ω_m:

$$e_m = K_1 \phi \omega_m = K_b \omega_m = K_b\ D\theta_m \tag{2.132}$$

The torque constant K_t and the generator constant K_b are the same for mks units. Control of the motor speed is obtained by adjusting the voltage applied to the armature. Its polarity determines the direction of the armature

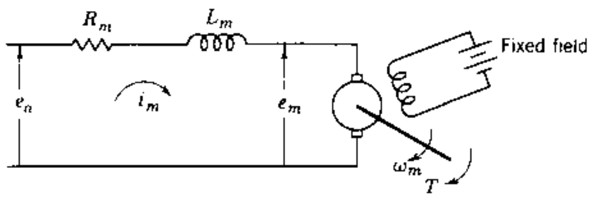


FIGURE 2.26 Circuit diagram of a dc motor.

current and, therefore, the direction of the torque generated. This, in turn, determines the direction of rotation of the motor. A circuit diagram of the armature-controlled dc motor is shown in Fig. 2.26. The armature inductance and resistance are labeled L_m and R_m. The voltage equation of the armature circuit is

$$L_m Di_m + R_m i_m + e_m = e_a \tag{2.133}$$

The current in the armature produces the required torque according to Eq. (2.131). The required torque depends on the load connected to the motor shaft. If the load consists only of a moment of inertia and damper (friction), as shown in Fig. 2.27 the torque equation can be written:

$$J\, D\omega_m + B\omega_m = T(t) \tag{2.134}$$

The required armature current i_m can be obtained by equating the generated torque of Eq. (2.131) to the required load torque of Eq. (2.134). Inserting this current, and the back emf from Eq. (2.132) into Eq. (2.133) produces the system equation in terms of the motor velocity ω_m:

$$\frac{L_m J}{K_t} D^2\omega_m + \frac{L_m B + R_m J}{K_t} D\omega_m + \left(\frac{R_m B}{K_t} + K_b\right)\omega_m = e_a \tag{2.135}$$

This equation can also be written in terms of motor position θ_m:

$$\frac{L_m J}{K_t} D^3\theta_m + \frac{L_m B + R_m J}{K_t} D^2\theta_m + \frac{R_m B + K_b K_t}{K_t} D\theta_m = e_a \tag{2.136}$$

There are two energy-storage elements, J and L_m, for the system represented by Figs. 2.26 and 2.27. Designating the independent state variables

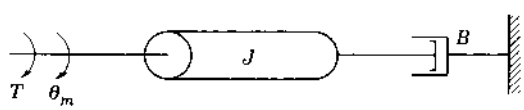


FIGURE 2.27 Inertia and friction as a motor load.

as $x_1 = \omega_m$ and $x_2 = i_m$ and the input as $u = e_a$ yields the state equation

$$\dot{\mathbf{x}} = \begin{bmatrix} -\dfrac{B}{J} & \dfrac{K_t}{J} \\ -\dfrac{K_b}{L_m} & -\dfrac{R_m}{L_m} \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ \dfrac{1}{L_m} \end{bmatrix} \mathbf{u} \tag{2.137}$$

If motor position θ_m is the output, another differential equation is required; that is, $\theta_m = \omega$. If the solution of θ_m is required, then the system is of third order. A third state variable, $x_3 = \theta_m$, must therefore be added.

The transfer functions ω_m/e_a and θ_m/e_a can be obtained from Eqs. (2.135) and (2.136). The armature inductance is small and can usually be neglected, $L_m \approx 0$. Equation (2.136) is thus reduced to a second-order equation. The corresponding transfer function has the form

$$G = \frac{\theta_m}{e_a} = \frac{K_M}{D(T_m D + 1)} \tag{2.138}$$

2.14 AC SERVOMOTOR [16]

An ac servomotor is basically a two-phase induction motor that has two stator field coils placed 90 electrical degrees apart, as shown in Fig. 2.28*a*. In a two-phase motor the ac voltage e and e_c are equal in magnitude and separated by a phase angle of 90°. A two-phase induction motor runs at a speed slightly below the synchronous speed and is essentially a constant-speed motor. The synchronous speed n_s is determined by the number of poles P produced by the stator windings and the frequency f of the voltage applied to the stator windlings: $n_s = 120f/P$ revolutions per minute.

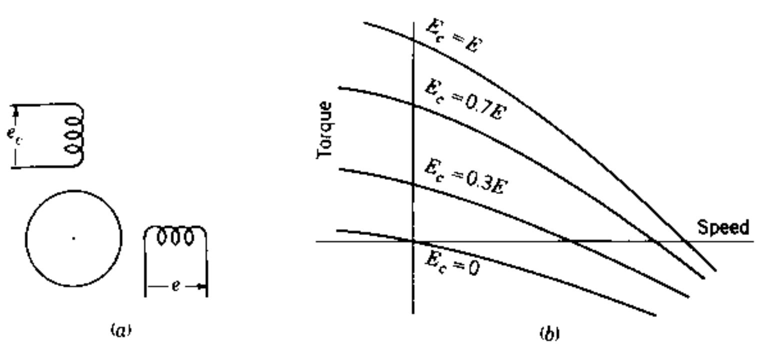


FIGURE 2.28 (*a*) Schematic diagram of a two-phase induction motor; (*b*) servomotor characteristics.

When the unit is used as a servomotor, the speed must be proportional to an input voltage. The two-phase motor can be used as a servomotor by applying an ac voltage e of fixed amplitude to one of the motor windings. When the other voltage e_c is varied, the torque and speed are a function of this voltage. Figure 2.28b shows a set of torque-speed curves for various control voltages.

Note that the curve for zero control-field voltage goes through the origin and that the slope is negative. This means that when the control-field voltage becomes zero, the motor develops a decelerating torque, causing it to stop. The curves show a large torque at zero speed. This is a requirement for a servomotor in order to provide rapid acceleration. It is accomplished in an induction motor by building the rotor with a high resistance.

The torque-speed curves are not straight lines. Therefore, a linear differential equation cannot be used to represent the exact motor characteristics. Sufficient accuracy may be obtained by approximating the characteristics by straight lines. The following analysis is based on this approximation.

The torque generated is a function of both the speed ω and the control-field voltage E_c. In terms of partial derivatives, the torque equation is approximated by effecting a double Taylor series expansion of $T(E_c, \omega)$ about the origin and keeping only the linear terms:

$$\left.\frac{\partial T}{\partial E_c}\right|_{\text{origin}} E_c + \left.\frac{\partial T}{\partial \omega}\right|_{\text{origin}} \omega = T(E_c, \omega) \tag{2.139}$$

If the torque-speed motor curves are approximated by parallel straight lines, the partial derivative coefficients of Eq. (2.139) are constants that can be evaluated from the graph. Let

$$\frac{\partial T}{\partial E_c} = K_c \qquad \text{and} \qquad \frac{\partial T}{\partial \omega} = K_\omega \tag{2.140}$$

For a load consisting of a moment of inertia and damping, the load torque required is

$$T_L = J\,D\omega + B\omega \tag{2.141}$$

Since the generated and load torques must be equal, Eqs. (2.139) and (2.141) are equated:

$$K_c E_c + K_\omega \omega = J\,D\omega + B\omega$$

Rearranging terms gives

$$J\,D\omega + (B - K_\omega)\omega = K_c E_c \tag{2.142}$$

In terms of position θ, this equation can be written as

$$J\,D^2\theta + (B - K_\omega)\,D\theta = K_c E_c \tag{2.143}$$

In order for the system to be stable (see Chap. 3) the coefficient $(B - K_\omega)$ must be positive. Observation of the motor characteristics shows that $K_\omega = \partial T/\partial\omega$ is negative; therefore, the stability requirement is satisfied.

Analyzing Eq. (2.142) reveals that this system has only one enery-storage element J. Thus the state equation, where $x_1 = \omega$ and $u = E_c$, is

$$\dot{x}_1 = -\frac{B - K_\omega}{J}x_1 + \frac{K_c}{J}u \tag{2.144}$$

2.15 LAGRANGE'S EQUATION

Previous sections show the application of Kirchhoff's laws for writing the differential equations of electric networks and the application of Newton's laws for writing the equations of motion of mechanical systems. These laws can be applied, depending on the complexity of the system, with relative ease. In many instances systems contain combinations of electric and mechanical components. The use of Lagrange's equation provides a systematic unified approach for handling a broad class of physical systems, no matter how complex their structure [8].

Lagrange's equation is given by

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_n}\right) - \frac{\partial T}{\partial q_n} + \frac{\partial D}{\partial \dot{q}_n} + \frac{\partial V}{\partial q_n} = Q_n \qquad n = 1, 2, 3, \ldots \tag{2.145}$$

where

T = total kinetic energy of system
D = dissipation function of system
V = total potential energy of system
Q_n = generalized applied force at the coordinate n
q_n = generalized coordinate
$\dot{q}_n = dq_n/dt$ (generalized velocity)

and $n = 1, 2, 3, \ldots$ denotes the number of independent coordinates or degrees of freedom that exist in the system. The total kinetic energy T includes all energy terms, regardless of whether they are electrical or mechanical. The dissipation function D represents one-half the rate at which energy is dissipated as heat; dissipation is produced by friction in mechanical systems and by resistance in electric circuits. The total potential energy stored in the system is designated by V. The forcing functions applied to a system are designated by Q_n; they take the form of externally applied forces or torques in mechanical systems and appear as voltage or current sources in electric circuits. These quantities are illustrated for electrical and mechanical components in the following examples. Kinetic energy is associated with the generalized velocity and, for an inductor and mass, is given by $T_L = L\dot{q}^2/2 = Li^2/2$ and $T_M = M\dot{x}^2/2 = Mv^2/2$, respectively. Potential energy is associated with the generalized position and, for a capacitor and spring, is given by $V_C = q^2/2C$ and $V_K = Kx^2/2$, respectively. The dissipation function is always a function of the generalized velocity and, for an electrical resistor and mechanical friction, is given by $D_R = R\dot{q}^2/2 = Ri^2/2$ and $D_B = B\dot{x}^2/2 = Bv^2/2$, respectively. The generalized applied force for an electric circuit is the electromotive force of a generator, thus $Q_q = E$. For the Earth's gravitational force applied to a mass, the generalized applied force is $Q_f = Mg$. Similar relationships can be developed for other physical systems. The application of Lagrange's equation is illustrated by the following example.

Example: Electromechanical System with Capacitive Coupling. Figure 2.29 shows a system in which mechanical motion is converted into electric energy. This represents the action that takes place in a capacitor microphone. Plate *a* of the capacitor is fastened rigidly to the frame. Sound waves impinge upon and exert a force on plate *b* of mass *M*, which is suspended from the frame by a spring *K* and which has damping *B*. The output voltage that appears across the resistor *R* is intended to reproduce electrically the sound-wave patterns which strike the plate *b*.

At equilibrium, with no external force extended on plate *b*, there is a charge q_0 on the capacitor. This produces a force of attraction between the plates so that the spring is stretched by an amount x_1 and the space between the plates is x_0. When sound waves exert a force on plate *b* there will be a

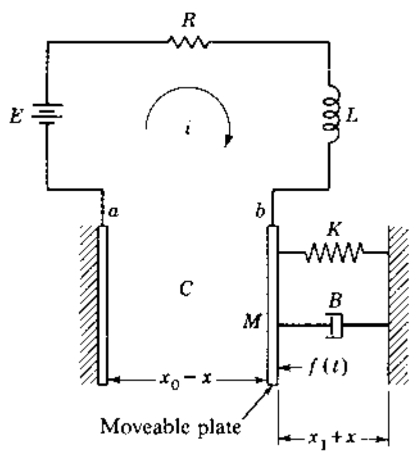


FIGURE 2.29 Electromechanical system with capacitive coupling.

resulting motion x that is measured from the equilibrium position. The distance between the plates will then be $x_0 - x$, and the charge on the plates will be $q_0 + q$.

The capacitance is approximated by

$$C = \frac{\varepsilon A}{x_0 - x} \qquad \text{and} \qquad C_0 = \frac{\varepsilon A}{x_0}$$

where ε is the dielectric constant for air and A is the area of the plate.

The energy expressions for this system are

$$\begin{aligned} T &= \tfrac{1}{2} L\dot{q}^2 + \tfrac{1}{2} M\dot{x}^2 \\ D &= \tfrac{1}{2} R\dot{q}^2 + \tfrac{1}{2} B\dot{x}^2 \\ V &= \frac{1}{2C}(q_0 + q)^2 + \tfrac{1}{2} K(x_1 + x)^2 \\ &= \frac{1}{2\varepsilon A}(x_0 - x)(q_0 + q)^2 + \tfrac{1}{2} K(x_1 + x)^2 \end{aligned}$$

The method is simple and direct. It is merely necessary to include all the energy terms, whether electrical or mechanical. The electromechanical coupling in this example appears in the potential energy. Here, the presence of charge on the plates of the capacitor exerts a force on the mechanical system. Also, motion of the mechanical system produces an equivalent emf in the electric circuit.

The two degrees of freedom are the displacement x of plate b and the charge q on the capacitor. Applying Lagrange's equation twice gives

$$M\ddot{x} + B\dot{x} - \frac{1}{2\varepsilon A}(q_0 + q)^2 + K(x_1 + x) = f(t) \tag{2.146}$$

$$L\ddot{q} + R\dot{q} + \frac{1}{\varepsilon A}(x_0 - x)(q_0 - q) = E \tag{2.147}$$

These equations are nonlinear. However, a good linear approximation can be obtained, because x and q are very small quantities and, therefore, the x^2, q^2, and xq terms can be neglected. This gives

$$\begin{aligned} &(q_0 + q)^2 \approx q_0^2 + 2q_0 q \\ &(x_0 - x)(q_0 - q) \approx x_0 q_0 - q_0 x + x_0 q \end{aligned}$$

With these approximations the system equations become

$$M\ddot{x} + Kx_1 + Kx - \frac{q_0^2}{2\varepsilon A} - \frac{2q_0 q}{2\varepsilon A} + B\dot{x} = f(t) \tag{2.148}$$

$$L\ddot{q} + \frac{x_0 q_0}{\varepsilon A} - \frac{q_0 x}{\varepsilon A} + \frac{x_0 q}{\varepsilon A} + R\dot{q} = E \tag{2.149}$$

From Eq. (2.148), by setting $f(t) = 0$ and taking steady-state conditions, the result is

$$Kx_l - \frac{q_0^2}{2\varepsilon A} = 0$$

This simply equates the force on the spring and the force due to the charges at the equilibrium condition. Similarly, in Eq. (2.149) at equilibrium

$$\frac{x_0 q_0}{\varepsilon A} = \frac{q_0}{C_0} = E$$

Therefore, the two system equations can be written in linearized form as

$$M\ddot{x} + B\dot{x} + Kx - \frac{q_0}{\varepsilon A} q = f(t) \tag{2.150}$$

$$L\ddot{q} + R\dot{q} + \frac{q}{C_0} - \frac{q_0}{\varepsilon A} x = 0 \tag{2.151}$$

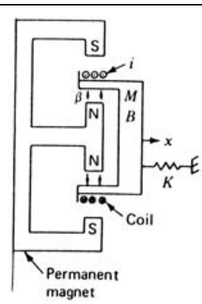
These equations show that $q_0/\varepsilon A$ is the coupling factor between the electrical and mechanical portions of the system.

Another form of electromechanical coupling exists when current in a coil produces a force that is exerted on a mechanical system and, simultaneously, motion of a mass induces an emf in an electric circuit. The electromagnetic coupling is represented in Table 2.8. In that case the kinetic energy includes a term

$$T = l\beta N_c xi = Uxi \tag{2.152}$$

The energy for this system, which is shown in Table 2.8, may also be considered potential energy. Since Eq. (2.145) contains the terms $-\partial T/\partial q_n$ and $\partial V/\partial q_n$, identification of the coupling energy as potential energy V would change the sign on the corresponding term in the differential equation.

Table 2.8 Identification of Energy Functions for Electromechanical Elements

Definition	Element and symbol	Kinetic energy T
x = relative motion i = current in coil $U = l\beta N_c$ (electromechanical coupling constant) β = flux density produced by permanent magnet l = length of coil N_c = number of turns in coil		Uxi

www.4electron.com

The main advantage of Lagrange's equation is the use of a single systematic procedure, eliminating the need to consider separately Kirchhoff's laws for the electrical aspects and Newton's law for the mechanical aspects of the system in formulating the statements of equilibrium. Once this procedure is mastered, the differential equations that describe the system are readily obtained.

2.16 SUMMARY

The examples in this chapter cover many of the basic elements of control systems. In order to write the differential and state equations, the basic laws governing performance are first stated for electrical, mechanical, thermal, and hydraulic systems. These basic laws are then applied to specific devices, and their differential equations of performance are obtained. The basic state, transfer function, and block diagram concepts are introduced. Lagrange's equation has been introduced to provide a systematized method for writing the differential equations of electrical, mechanical, and electromechanical systems. This chapter constitutes a reference for the reader who wants to review the fundamental concepts involved in writing differential equations of performance. Basic matrix fundamentals are presented in Appendix B.

REFERENCES

1. D'Azzo, J. J., and C. H. Houpis: *Linear Control System Analysis & Design: Conventional and Modern*, 3rd ed., McGraw-Hill, New York, 1988.
2. Rugh, W. J.: *Linear System Theory*, Prentice Hall, Englewood Cliffs. NJ. 1993.
3. Blackburn, J. F., ed.: *Components Handbook*, McGraw-Hill, New York, 1948.
4. Stout, T.M.: "A Block Diagram Approach to Network Analysis," *Trans. AIEE*, vol. **71**, pp. 255–260. 1952.
5. Kreyszig, E.: *Advanced Engineering Mathematics*, 7th ed., John Wiley, New York, 1993.
6. Nilson, J.W.: *Electric Circuits*, 2nd ed., Addison-Wesley, Reading, Mass., 1980.
7. Kalman, R. E.: "Mathematical Description of Linear Dynamical Systems." *J. Soc. Ind. Appl. Math., Ser. A. Control*, vol. **1**, no. 2, pp. 152–192, 1963.
8. Ogar, G. W., and J. J. D'Azzo: "A Unified Procedure for Deriving the Differential Equations of Electrical and Mechanical Systems," *IRE Trans. Educ.*, vol. E-5, no.1, pp. 18–26, March 1962.
9. Davis, S. A., and B. K. Ledgerwood: *Electromechanical Components for Servomechanisms*, McGraw-Hill, New York, 1961.
10. Gardner, M. F., and J. L. Barnes: *Transients in Linear Systems*. Wiley, New York, 1942, chap. 2.
11. Hornfeck, A. J.: "Response Characteristics of Thermometer Elements," *Trans. ASME,* vol. 71. pp. 121–132, 1949.

12. *Flow Meters: Their Theory and Application*, American Society of Mechanical Engineers, New York, 1937.
13. Takahashi, Y., M. J. Rabins, and D.M. Auslander: *Control and Dynamic Systems*, Addison-Wesley, Reading, Mass., 1970.
14. Saunders, R. M.: "The Dynamo Electric Amplifier: Class A Operation," *Trans. AIEE.* vol. 68, pp. 1368–1373, 1949.
15. Litman, B.: "An Analysis of Rotating Amplifiers," *Trans. AIEE*. vol. 68, pt. II, pp. 1111–1117, 1949.
16. Hopkin, A. M.: "Transient Response of Small Two-Phase Servomotors," *Trans. AIEE*, vol. 70, pp. 881–886, 1951.
17. Cochin, I.: *Analysis and Design of Dynamic Systems*, Harper & Row, New York, 1980.

3

Solution of Differential Equations

3.1 INTRODUCTION

The general solution of a linear differential equation [1,2] is the sum of two components, the particular integral and the complementary function. Often the particular integral is the steady-state component of the solution of the differential equation; the complementary function, which is the solution of the corresponding homogeneous equation, is then the transient component of the solution. Often the steady-state component of the response has the same form as the driving function. In this book the particular integral is called the steady-state solution even when it is not periodic. The form of the transient component of the response depends only on the roots of the characteristic equation. For nonlinear differential equations the form of the response also depends on the initial or boundary conditions. The instantaneous value of the transient component depends on the boundary conditions, the roots of the characteristic equation, and the instantaneous value of the steady-state component.

This chapter covers methods of determining the steady-state and the transient components of the solution. These components are first determined separately and then added to form the complete solution. Analysis of the transient solution should develop in the student a feel for the solution to be expected.

The method of solution is next applied to the matrix state equation. A general format is obtained for the complementary solution in terms of the state transition matrix (STM). Then the complete solution is obtained as a function of the STM and the input forcing function.

3.2 STANDARD INPUTS TO CONTROL SYSTEMS

For some control systems the input has a specific form that may be represented either by an analytical expression or as a specific curve. An example of the latter is the pattern used in a machining operation, where the cutting tool is required to follow the path indicated by the pattern outline. For other control systems the input may be random in shape. In this case it cannot be expressed analytically and is not repetitive. An example is the camera platform used in a photographic airplane. The airplane flies at a fixed altitude and speed, and the camera takes a series of pictures of the terrain below it, which are then fitted together to form one large picture of the area. This task requires that the camera platform remain level regardless of the motion of the airplane. Since the attitude of the airplane varies with wind gusts and depends on the stability of the airplane itself, the input to the camera platform is obviously a random function.

It is important to have a basis of comparison for various systems. One way of doing this is by comparing the response with a standardized input. The input or inputs used as a basis of comparison must be determined from the required response of the system and the actual form of its input. The following standard inputs, with unit amplitude, are often used in checking the response of a system:

1. Sinusoidal function $r = \cos \omega t$
2. Power series function $r = a_0 + a_1 t + a_2(t^2/2) + \cdots$
3. Step function $r = u_{-1}(t)$
4. Ramp (step velocity) function $r = u_{-2}(t) = tu_{-1}(t)$
5. Parabolic (step acceleration) function $r = u_{-3}(t) = (t^2/2)u_{-1}(t)$
6. Impulse function $r = u_0(t)$

Functions 3 to 6 are called *singularity functions* (Fig. 3.1). The singularity functions can be obtained from one another by successive differentiation or integration. For example, the derivative of the parabolic function is the ramp function, the derivative of the ramp function is the step function, and the derivative of the step function is the impulse function.

For each of these inputs a complete solution of the differential equation is determined in this chapter. First, generalized methods are developed to determine the steady-state output $c(t)_{ss}$ for each type of input.

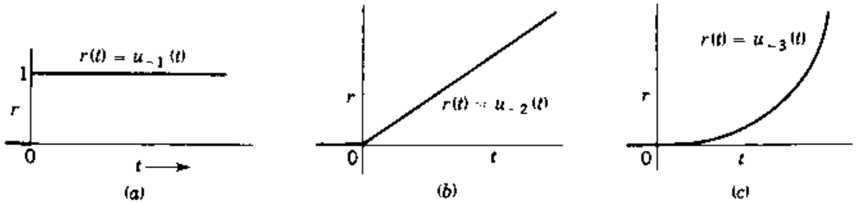


FIGURE 3.1 Singularity functions: (*a*) step function, $u_{-1}(t)$; (*b*) ramp function, $u_{-2}(t)$; (*c*) parabolic function, $u_{-3}(t)$.

These methods are applicable to linear differential equations of any order. Next, the method of evaluating the transient component of the response $c(t)_t$ is determined, and the form of the transient component of the response is shown to depend on the characteristic equation. Addition of the steady-state component and the transient component gives the complete solution, that is, $c(t) = c(t)_{ss} + c(t)_t$. The coefficients of the transient terms are determined by the instantaneous value of the steady-state component, the roots of the characteristic equation, and the initial conditions. Several examples are used to illustrate these principles. The solution of the differential equation with a pulse input is postponed until the next chapter.

3.3 STEADY-STATE RESPONSE: SINUSOIDAL INPUT

The input quantity r is assumed to be a sinusoidal function of the form

$$r(t) = R\cos(\omega t + \alpha) \tag{3.1}$$

The general integrodifferential equation to be solved is of the form

$$\begin{aligned} &A_v D^v c + A_{v-1} D^{v-1} c + \cdots + A_0 D^0 c + A_{-1} D^{-1} c \\ &\quad + \cdots + A_{-w} D^{-w} c = r \end{aligned} \tag{3.2}$$

The steady-state solution can be obtained directly by use of Euler's identity,

$$e^{j\omega t} = \cos\omega t + j\sin\omega t$$

The input can then be written

$$\begin{aligned} r = R\cos(\omega t + \alpha) &= \text{ real part of } \left(\mathrm{R}e^{j(\omega t+\alpha)}\right) \\ &= \mathrm{Re}\left(\mathrm{R}e^{j(\omega t+\alpha)}\right) = \mathrm{Re}(\mathrm{R}e^{j\alpha}e^{j\omega t}) = \mathrm{Re}(\mathbf{R}e^{j\omega t}) \end{aligned} \tag{3.3}$$

For simplicity, the phrase *real part of* or its symbolic equivalent Re is often omitted, but is must be remembered that the real part is intended.

The quantity $\mathbf{R} = Re^{j\alpha}$ is the *phasor* representation of the input; i.e., it has both a magnitude R and an angle α. The magnitude R represents the maximum value of the input quantity $r(t)$. For simplicity the angle $\alpha = 0^\circ$ usually is chosen for $\mathbf{R}$. The input r from Eq. (3.3) is inserted in Eq. (3.2). Then, for the expression to be an equality, the response c must be of the form

$$c(t)_{ss} = C\cos(\omega t + \phi) = \mathrm{Re}\,(Ce^{j\phi}e^{j\omega t}) = \mathrm{Re}\,(\mathbf{C}e^{j\omega t}) \tag{3.4}$$

where $\mathbf{C} = Ce^{j\phi}$ is a *phasor* quantity having the magnitude C and the angle ϕ. The nth derivative of c_{ss} with respect to time is

$$D^n c(t)_{ss} = \mathrm{Re}\left[(j\omega)^n \mathbf{C}e^{j\omega t}\right] \tag{3.5}$$

Inserting c_{ss} and its derivatives from Eqs. (3.4) and (3.5) into Eq. (3.2) gives

$$\begin{aligned} &\mathrm{Re}\,(A_v(j\omega)^v \mathbf{C}e^{j\omega t} + A_{v-1}(j\omega)^{v-1}\mathbf{C}e^{j\omega t} + \cdots + A_{-w}(j\omega)^{-w}\mathbf{C}e^{j\omega t}] \\ &\quad = \mathrm{Re}\,(\mathbf{R}e^{j\omega t}) \end{aligned} \tag{3.6}$$

Canceling $e^{j\omega t}$ from both sides of the equation and then solving for $\mathbf{C}$ gives

$$\mathbf{C} = \frac{\mathbf{R}}{A_v(j\omega)^v + A_{v-1}(j\omega)^{v-1} + \cdots + A_0 + A_{-1}(j\omega)^{-1} + \cdots + A_{-w}(j\omega)^{-w}} \tag{3.7}$$

where $\mathbf{C}$ is the phasor representation of the output; i.e., it has a magnitude C and an angle ϕ. Since the values of C and ϕ are functions of the frequency ω, the phasor output is written as $\mathbf{C}(j\omega)$ to show this relationship. Similarly, $\mathbf{R}(j\omega)$ denotes the fact that the input is sinusoidal and may be a function of frequency.

Comparing Eqs. (3.2) and (3.6), it can be seen that one equation can be determined easily from the other. Substituting $j\omega$ for D, $\mathbf{C}(j\omega)$ for c, and $\mathbf{R}(j\omega)$ for r in Eq. (3.2) results in Eq. (3.6). The reverse is also true and is independent of the order of the equation. It should be realized that this is simply a *rule of thumb* that yields the desired expression.

The time response can be obtained directly from the phasor response. The output is

$$c(t)_{ss} = \mathrm{Re}\,(\mathbf{C}e^{j\omega t}) = |\mathbf{C}|\cos(\omega t + \phi) \tag{3.8}$$

As an example, consider the rotational hydraulic transmission described in eq. (2.122), Sec. 2.11, of Ref. 10, which has an input x and an output angular position θ_m. The equation that relates input to output in terms of system parameters is repeated here:

$$\frac{VJ}{K_B C}D^3\theta_m + \frac{LJ}{C}D^2\theta_m + d_m D\theta_m = d_p\omega_p x \tag{3.9}$$

When the input is the sinusoid, $x = X \sin \omega t$, the corresponding phasor equation can be obtained from Eq. (3.9) by replacing $x(t)$ by $\mathbf{X}(j\omega)$, θ_m by $\mathbf{\Theta}_m(j\omega)$, and D by $j\omega$. The ratio of phasor output to phasor input is termed the *frequency transfer function*, often designated by $\mathbf{G}(j\omega)$. This ratio, in terms of the steady-state sinusoidal phasors, is

$$\mathbf{G}(j\omega) = \frac{\mathbf{\Theta}_m(j\omega)}{\mathbf{X}(j\omega)} = \frac{d_p\omega_p/d_m}{j\omega\left[(VJ/K_B C\, d_m)(j\omega)^2 + (LJ/C\, d_m)(j\omega) + 1\right]} \tag{3.10}$$

3.4 STEADY-STATE RESPONSE: POLYNOMIAL INPUT

A development of the steady-state solution of a differential equation with a general polynomial input is presented first. Particular cases of the power series are then covered in detail.

The general differential equation (3.2) is repeated here:

$$A_v D^v c + A_{v-1} D^{v-1} c + \cdots + A_0 c + A_{-1} D^{-1} c + \cdots + A_{-w} D^{-w} c = r \tag{3.11}$$

The polynomial input is of the form

$$r(t) = R_0 + R_1 t + \frac{R_2 t^2}{2!} + \cdots + \frac{R_k t^k}{k!} \tag{3.12}$$

where the highest-order term in the input is $R_k t^k/k!$. For $t < 0$ the value of $r(t)$ is zero. The object is to find the steady-state or particular solution of the dependent variable c. The method used is to assume a polynomial solution of the form

$$c(t)_{ss} = b_0 + b_1 t + \frac{b_2 t^2}{2!} + \cdots + \frac{b_q t^q}{q!} \tag{3.13}$$

where determination of the value of q is given as follows:

The assumed solution is then substituted into the differential equation. The coefficients $b_0, b_1, b_2, \ldots$ of the polynomial solution are evaluated by equating the coefficients of like powers of t on both sides of the equation. Inserting $r(t)$ from Eq. (3.12) into Eq. (3.11), the highest power of t on the right side of Eq. (3.11) is k; therefore t^k must also appear on the left side of this equation. The highest power of t on the left side of the equation is produced by the lowest-order derivative term $D^{-w}c$ and is equal to q plus the order of the lowest derivative. Therefore, the value of the highest-order exponent of t that appears in the assumed solution of Eq. (3.13) is

$$q = k - w \qquad q \geq 0 \tag{3.14}$$

where k is the highest exponent appearing in the input and w is the lowest-order exponent of the differential operator D appearing in the general differential equation (3.11). Equation (3.14) is valid only for positive values of q. For each of the following examples the response is a polynomial, because the input is of that form. However, the highest power in the response may not be the same as that of the input. When the lowest-order derivative is zero, that is, $w = 0$ in Eq. (3.11), then $q = k$. When integral terms are present, the highest order of t in $c(t)$ is smaller than k.

Step-Function Input

A convenient input $r(t)$ to a system is an abrupt change represented by a unit step function, as shown in Fig. 3.1*a*. This type of input cannot always be put into a system since it may take a definite length of time to make the change in input, but it represents a good mathematical input for checking system response.

The servomotor described in Sec. 2.13 is used as an example. The response of motor velocity ω_m in terms of the voltage e_a applied to the armature, as given by Eq. (2.135), is of the form

$$A_2 D^2 x + A_1 Dx + A_0 x = r \tag{3.15}$$

The unit step function $r = u_{-1}(t)$ is a polynomial in which the highest exponent of t is $k = 0$. When the input is a unit step function, the method of solution for a polynomial can therefore be used, that is, the steady-state response is also a polynomial. The lowest-order derivative in Eq. (3.15) is $w = 0$; therefore, $q = 0$ and the steady-state response has only one term of the form

$$x_{ss} = b_0 \tag{3.16}$$

The derivatives are $Dx_{ss} = 0$ and $D^2 x_{ss} = 0$. Inserting these values into Eq. (3.15) yields

$$x(t)_{ss} = b_0 = \frac{1}{A_0} \tag{3.17}$$

Ramp-Function Input (Step Function of Velocity)

The ramp function is a fixed rate of change of a variable as a function of time. This input is shown in Fig. 3.1*b*. The input and its derivative are expressed mathematically as

$$r(t) = u_{-2}(t) = tu_{-1}(t) \qquad Dr = u_{-1}(t)$$

A ramp input is a polynomial input where the highest power of t is $k = 1$. When the input $r(t)$ in Eq. (3.15) is the ramp function $u_{-2}(t)$, the highest power of t in the polynomial output is $q = k = 1$. The output is therefore

$$x(t)_{ss} = b_0 + b_1 t \tag{3.18}$$

The derivatives are $Dx_{ss} = b_1$ and $D^2 x_{ss} = 0$. Inserting these values into Eq. (3.15), and then equating coefficients of t raised to the same power, yields

$$\begin{aligned} t^0: &\qquad A_1 b_1 + A_0 b_0 = 0 &\qquad b_0 &= -\frac{A_1 b_1}{A_0} = -\frac{A_1}{A_0^2} \\ t^1: &\qquad A_0 b_1 = 1 &\qquad b_1 &= \frac{1}{A_0} \end{aligned}$$

Thus, the steady-state solution is

$$x(t)_{ss} = -\frac{A_1}{A_0^2} + \frac{1}{A_0} t \tag{3.19}$$

Parabolic-Function Input (Step Function of Acceleration)

A parabolic function has a constant second derivative, which means that the first derivative is a ramp. The input $r(t)$ is expressed mathematically as

$$r = u_{-3}(t) = \frac{t^2}{2} u_{-1}(t) \qquad Dr = u_{-2}(t) = t u_{-1}(t) \qquad D^2 r = u_{-1}(t)$$

A parabolic input is a polynomial input where the highest power of t is $k = 2$. A steady-state solution is found in the conventional manner for a polynomial input. Consider Eq. (3.15) with a parabolic input. The order of the lowest derivative in the system equation is $w = 0$. The value of q is therefore equal to 2, and the steady-state response is of the form

$$x(t)_{ss} = b_0 + b_1 t + \frac{b_2 t^2}{2} \tag{3.20}$$

The derivatives of x are $Dx_{ss} = b_1 + b_2 t$ and $D^2 x_{ss} = b_2$. Inserting these values into Eq. (3.15), followed by equating the coefficients of t raised to the same power, yields $b_2 = 1/A_0$, $b_1 = -A_1/A_0^2$, and $b_0 = A_1^2/A_0^3 - A_2/A_0^2$.

3.5 TRANSIENT RESPONSE: CLASSICAL METHOD

The classical method of solving for the complementary function or transient response of a differential equation requires, first, the writing of the homogeneous equation. The general differential equation has the form

$$b_v D^v c + b_{v-1} D^{v-1} c + \cdots + b_0 D^0 c + b_{-1} D^{-1} c + \cdots + b_{-w} D^{-w} c = r \tag{3.21}$$

where r is the forcing function and c is the response. The homogeneous equation is formed by letting the right side of the differential equation equal zero:

$$b_v D^v c_t + b_{v-1} D^{v-1} c_t + \cdots + b_0 c_t + b_{-1} D^{-1} c_t + \cdots + b_{-w} D^{-w} c_t = 0 \tag{3.22}$$

where c_t is the transient component of the general solution.

The general expression for the transient response, which is the solution of the homogeneous equation, is obtained by assuming a solution of the form

$$c_t = A_m e^{mt} \tag{3.23}$$

where m is a constant yet to be determined. Substituting this value of c_t into Eq. (3.22) and factoring $A_m e^{mt}$ from all terms gives

$$A_m e^{mt}(b_v m^v + b_{v-1} m^{v-1} + \cdots + b_0 + \cdots + b_{-w} m^{-w}) = 0 \tag{3.24}$$

Equation (3.24) must be satisfied for $A_m e^{mt}$ to be a solution. Since e^{mt} cannot be zero for all values of time t, it is necessary that

$$Q(m) = b_v m^v + b_{v-1} m^{v-1} + \cdots + b_0 + \cdots + b_{-w} m^{-w} = 0 \tag{3.25}$$

This equation is purely algebraic and is termed the *characteristic equation*. The roots of the characteristic equation, Eq. (3.25), can be readily obtained by using the MATLAB CAD program. There are $v + w$ roots, or *eigenvalues*, of the characteristic equation; therefore, the complete transient solution contains the same number of terms of the form $A_m e^{mt}$ if all the roots are simple. Thus the transient component, when there are no multiple roots, is

$$c_t = A_1 e^{m_1 t} + A_2 e^{m_2 t} + \cdots + A_k e^{m_k t} + \cdots + A_{v+w} e^{m_{v+w} t} \tag{3.26}$$

where each $e^{m_k t}$ is described as a *mode* of the system. If there is a root m_q of multiplicity p, the transient includes corresponding terms of the form

$$A_{q1} e^{m_q t} + A_{q2} t e^{m_q t} + \cdots + A_{qp} t^{p-1} e^{m_q t} \tag{3.27}$$

Instead of using the detailed procedure just outlined, the characteristic equation of Eq. (3.25) is usually obtained directly from the homogeneous equation of Eq. (3.22) by substituting m for Dc_t, m^2 for $D^2 c_t$, etc.

Since the coefficients of the transient solution must be determined from the initial conditions, there must be $v + w$ known initial conditions. These conditions are values of the variable c and of its derivatives that are known at specific times. The $v + w$ initial conditions are used to set up $v + w$ simultaneous equations of c and its derivatives. The value of c includes both the steady-state and transient components. Since determination of the coefficients includes consideration of the steady-state component, the input affects the value of the coefficient of each exponential term.

Complex Roots

If all values of m_k are real, the transient terms can be evaluated as indicated above. Frequently, some values of m_k are complex. When this happens, they always occur in pairs that are complex conjugates and are of the form

$$m_k = \sigma + j\omega_d \qquad m_{k+1} = \sigma - j\omega_d \tag{3.28}$$

where σ is called the *damping coefficient* and ω_d is called the *damped natural frequency*. The transient terms corresponding to these values of m are

$$A_k e^{(\sigma+j\omega_d)t} + A_{k+1} e^{(\sigma-j\omega_d)t} \tag{3.29}$$

These terms are combined to a more useful form by factoring the term $e^{\sigma t}$:

$$e^{\sigma t}(A_k e^{j\omega_d t} + A_{k+1} e^{-j\omega_d t}) \tag{3.30}$$

By using the Euler identity $e^{\pm j\omega_d t} = \cos \omega_d t \pm j \sin \omega_d t$ and then combining terms, expression (3.30) can be put into the form

$$e^{\sigma t}(B_1 \cos \omega_d t + B_2 \sin \omega_d t) \tag{3.31}$$

This can be converted into the form

$$Ae^{\sigma t} \sin(\omega_d t + \phi) \tag{3.32}$$

where $A = \sqrt{B_1^2 + B_2^2}$ and $\phi = \tan^{-1}(B_1/B_2)$. This form is very convenient for plotting the transient response. The student must learn to use this equation directly without deriving it each time. Often the constants in the transient term are evaluated more readily from the initial conditions by using the form of Eq. (3.31), and then the expression is converted to the more useful form of Eq. (3.32).

This transient term is called an *exponentially damped sinusoid*; it consists of a sine wave of frequency ω_d whose magnitude is $Ae^{\sigma t}$; that is, it is decreasing exponentially with time if σ is negative. It has the form shown in Fig. 3.2, in which the curves $\pm Ae^{\sigma t}$ constitute the *envelope*. The plot of the time solution always remains between the two branches of the envelope.

For the complex roots given in Eq. (3.28), where σ is negative, the transient decays with time and eventually dies out. A negative σ represents a stable system. When σ is positive, the transient increases with time and will destroy the equipment unless otherwise restrained. A positive σ represents the undesirable case of an unstable system. Control systems must be designed so that they are always stable.

Damping Ratio ζ and Undamped Natural Frequency ω_n

When the characteristic equation has a pair of complex-conjugate roots, it has a quadratic factor of the form $b_2 m^2 + b_1 m + b_0$. The roots of this

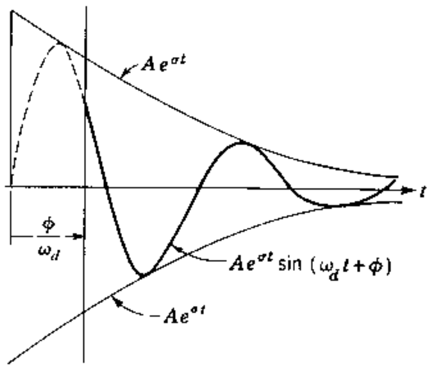


FIGURE 3.2 Sketch of an exponentially damped sinusoid.

factor are

$$m_{1,2} = -\frac{b_1}{2b_2} \pm j\sqrt{\frac{4b_2b_0 - b_1^2}{4b_2^2}} = \sigma \pm j\omega_d \tag{3.33}$$

The real part σ is recognized as the exponent of e, and ω_d is the frequency of the oscillatory portion of the component stemming from this pair of roots, as given by expression (3.32).

The quantity b_1 represents the effective damping constant of the system. If the numerator under the square root in Eq. (3.33) is zero, then b_1 has the value $2\sqrt{b_2b_0}$ and the two roots $m_{1,2}$ are equal. This represents the critical value of the damping constant and is written $b_1' = 2\sqrt{b_2b_0}$. The *damping ratio* ζ is defined as the ratio of the actual damping constant to the critical value of the damping constant:

$$\zeta = \frac{\text{actual damping constant}}{\text{critical damping constant}} = \frac{b_1}{b_1'} = \frac{b_1}{2\sqrt{b_2b_0}} \tag{3.34}$$

When ζ is positive and less than unity, the roots are complex and the transient is a damped sinusoid of the form of expression (3.32). When ζ is less than unity, the response is said to be *underdamped*. When ζ is greater than unity, the roots are real and the response is *overdamped*; i.e., the transient solution consists of two exponential terms with real exponents. For $\zeta > 0$ the transients decay with time and the response $c(t)$ approaches the steady state value. For $\zeta < 0$ the transient increases with time and the response is unstable.

The undamped natural frequency ω_n is defined as the frequency of the sustained oscillation of the transient if the damping is zero:

$$\omega_n = \sqrt{\frac{b_0}{b_2}} \tag{3.35}$$

The case of zero damping constant, $b_1 = 0$, means that the transient response does not die out; it is a sine wave of constant amplitude.

The quadratic factors are frequently written in terms of the damping ratio and the undamped natural frequency. Underdamped systems are generally analyzed in terms of these two parameters. After factoring b_0, the quadratic factor of the characteristic equation is

$$\frac{b_2}{b_0}m^2 + \frac{b_1}{b_0}m + 1 = \frac{1}{\omega_n^2}m^2 + \frac{2\zeta}{\omega_n}m + 1 \tag{3.36}$$

When it is multiplied through by ω_n^2, the quadratic appears in the form

$$m^2 + 2\zeta\omega_n m + \omega_n^2 \tag{3.37}$$

The two forms given by Eqs. (3.36) and (3.37) are called the *standard forms* of the quadratic factor, and the corresponding roots are

$$m_{1,2} = \sigma \pm j\omega_d = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2} \tag{3.38}$$

The transient response of Eq. (3.32) for the underdamped case, written in terms of ζ and ω_n, is

$$Ae^{\sigma t}\sin(\omega_d t + \phi) = Ae^{-\zeta\omega_n t}\sin(\omega_n\sqrt{1-\zeta^2}\,t + \phi) \tag{3.39}$$

From this expression the effect on the transient of the terms ζ and ω_n can readily be seen. The larger the product $\zeta\omega_n$, the faster the transient will decay. These terms also affect the damped natural frequency of oscillation of the transient, $\omega_d = \omega_n\sqrt{1-\zeta^2}$, which varies directly as the undamped natural frequency and decreases with an increase in damping ratio.

3.6 DEFINITION OF TIME CONSTANT

The transient terms have the exponential form Ae^{mt}. When $m = -a$ is real and negative, the plot of Ae^{-at} has the form shown in Fig. 3.3. The value of time that makes the exponent of e equal to -1 is called the *time constant T*. Thus

$$-aT = -1 \qquad \text{and} \qquad T = \frac{1}{a} \tag{3.40}$$

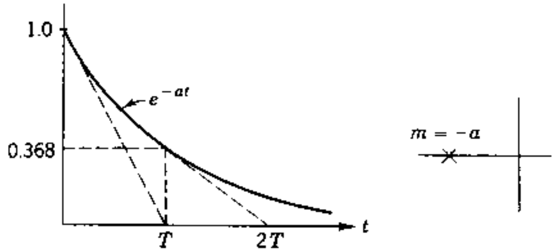


FIGURE 3.3 Plot of the exponential e^{-at} and the root location.

In a duration of time equal to one time constant the exponential e^{-at} decreases from the value 1 to the value 0.368. Geometrically, the tangent drawn to the curve Ae^{-at} at $t=0$ intersects the time axis at the value of time equal to the time constant T.

When $m = \sigma \pm j\omega_d$ is a complex quantity, the transient has the form $Ae^{\sigma t}\sin(\omega_d t + \phi)$. A plot of this function is shown in Fig. 3.2. In the case of the damped sinusoid, the time constant is defined in terms of the parameter σ that characterizes the envelope $Ae^{-\sigma t}$. Thus the time constant T is equal to

$$T = \frac{1}{|\sigma|} \tag{3.41}$$

In terms of the damping ratio and the undamped natural frequency, the time constant is $T = 1/\zeta\omega_n$. Therefore, the larger the product $\zeta\omega_n$, the greater the instantaneous rate of decay of the transient.

3.7 EXAMPLE: SECOND-ORDER SYSTEM—MECHANICAL

The simple translational mechanical system of Sec. 2.5 is used as an example and is shown in Fig. 3.4. Equation (2.59) relates the displacement x_b to x_a:

$$M\,D^2x_b + B\,Dx_b + Kx_b = Kx_a \tag{3.42}$$

Consider the system to be originally at rest. Then the function x_a moves 1 unit at time $t=0$; that is, the input is a unit step function $x_a(t) = u_{-1}(t)$. The problem is to find the motion $x_b(t)$.

The displacement of the mass is given by

$$x_b(t) = x_{b,\mathrm{ss}} + x_{b,t}$$

where $x_{b,\mathrm{ss}}$ is the steady-state solution and $x_{b,t}$ is the transient solution.

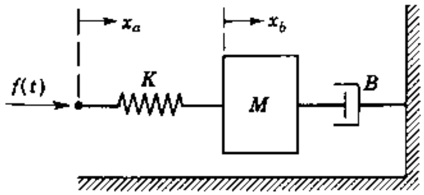


FIGURE 3.4 Simple mechanical system.

The steady-state solution is found first by using the method of Sec. 3.4. In this example it may be easier to consider the following:

1. Since the input x_a is a constant, the response x_b must reach a fixed steady-state position.
2. When x_b reaches a constant value, the velocity and acceleration become zero.

By putting $D^2x_b = Dx_b = 0$ into Eq. (3.42), the final or steady-state value of x_b is $x_{b,\text{ss}} = x_a$. After the steady-state solution has been found, the transient solution is determined. The characteristic equation is

$$Mm^2 + Bm + K = M\left(m^2 + \frac{B}{M}m + \frac{K}{M}\right) = 0$$

Putting this in terms of ζ and ω_n gives

$$m^2 + 2\zeta\omega_n m + \omega_n^2 = 0 \tag{3.43}$$

for which the roots are $m_{1,2} = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$. The transient solution depends on whether the damping ratio ζ is (1) greater than unity, (2) equal to unity, or (3) smaller than unity. For ζ greater than unity the roots are real and have the values $m_1 = -a$ and $m_2 = -b$, and the transient response is

$$x_{b,t} = A_1 e^{-at} + A_2 e^{-bt} \tag{3.44}$$

For ζ equal to unity, the roots are real and equal; that is, $m_1 = m_2 = -\zeta\omega_n$. Since there are multiple roots, the transient response is

$$x_{b,t} = A_1 e^{-\zeta\omega_n t} + A_2 t e^{-\zeta\omega_n t} \tag{3.45}$$

For ζ less than unity, the roots are complex,

$$m_{1,2} = -\zeta\omega_n \pm j\omega_n\sqrt{1 - \zeta^2}$$

and the transient solution, as outlined in Sec. 3.5, is

$$x_{b,t} = Ae^{-\zeta\omega_n t} \sin\left(\omega_n\sqrt{1 - \zeta^2}\,t + \phi\right) \tag{3.46}$$

The complete solution is the sum of the steady-state and transient solutions. For the underdamped case, $\zeta < 1$, the complete solution of Eq. (3.42) with a unit step input is

$$x_b(t) = 1 + Ae^{-\zeta\omega_n t}\sin(\omega_n\sqrt{1-\zeta^2}\,t + \phi) \tag{3.47}$$

The two constants A and ϕ must next be determined from the initial conditions. In this example the system was initially at rest; therefore, $x_b(0) = 0$. The energy stored in a mass is $W = \frac{1}{2}Mv^2$. From the principle of conservation of energy, the *velocity of a system with mass cannot change instantaneously*; thus $Dx_b(0) = 0$. Two equations are necessary, one for $x_b(t)$ and one for $Dx_b(t)$. Differentiating Eq. (3.47) yields

$$\begin{aligned} Dx_b(t) = {} & -\zeta\omega_n A e^{-\zeta\omega_n t}\sin(\omega_n\sqrt{1-\zeta^2}\,t + \phi) \\ & + \omega_n\sqrt{1-\zeta^2}\,Ae^{-\zeta\omega_n t}\cos(\omega_n\sqrt{1-\zeta^2}\,t + \phi) \end{aligned} \tag{3.48}$$

Inserting the initial conditions $x_b(0) = 0$, $Dx_b(0) = 0$, and $t = 0$ into Eqs. (3.47) and (3.48) yields the two simultaneous equations

$$0 = 1 + A\sin\phi \qquad 0 = -\zeta\omega_n A\ \sin\ \phi + \omega_n\sqrt{1-\zeta^2}\,A\ \cos\ \phi$$

These equations are then solved for A and ϕ:

$$A = \frac{-1}{\sqrt{1-\zeta^2}} \qquad \phi = \tan^{-1}\frac{\sqrt{1-\zeta^2}}{\zeta} = \cos^{-1}\zeta$$

where the units of ϕ are radians. Thus the complete solution is

$$x_b(t) = 1 - \frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}}\sin(\omega_n\sqrt{1-\zeta^2}\,t + \cos^{-1}\zeta) \tag{3.49}$$

When the complete solution has been obtained, it should be checked to see that it satisfies the known conditions. For example, putting $t = 0$ into Eq. (3.49) gives $x_b(0) = 0$; therefore, the solution checks. In a like manner, the constants can be evaluated for the other two cases of damping. Equation (3.49) shows that the steady-state value $x_{b,\mathrm{ss}}$ is equal to x_a. Thus, the output *tracks* the input.

3.8 EXAMPLE: SECOND-ORDER SYSTEM—ELECTRICAL

The electric circuit of Fig. 3.5 is used to illustrate further the determination of initial conditions. The circuit, as shown, is in the steady state. At time $t = 0$ the switch is closed. The problem is to solve for the current $i_2(t)$ through the inductor for $t > 0$.

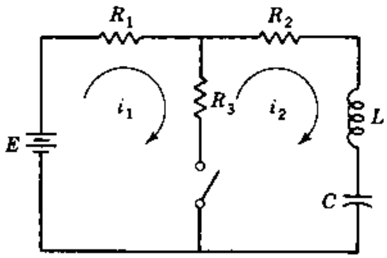


FIGURE 3.5 Electric circuit: $E = 10\,\text{V}$, $R_1 = 10\,\Omega$, $R_2 = 15\,\Omega$, $R_3 = 10\,\Omega$, $L = 1\,\text{H}$, $C = 0.01\,\text{F}$.

Two loop equations are written for this circuit:

$$10 = 20i_1 - 10i_2 \tag{3.50}$$

$$0 = -10i_1 + \left(D + 25 + \frac{100}{D}\right)i_2 \tag{3.51}$$

Eliminating i_1 from these equations yields

$$10 = \left(2D + 40 + \frac{200}{D}\right)i_2$$

Differentiating this equation yields

$$(2D^2 + 40D + 200)i_2 = 0 \tag{3.52}$$

The steady-state solution is found by using the method of Sec. 3.4. Since the input of Eq. (3.52) is a step function of value zero, the steady-state output is

$$i_{2,\text{ss}} = 0 \tag{3.53}$$

This can also be deduced from an inspection of the circuit. When a branch contains a capacitor, the steady-state current is always zero for a dc source.

Next the transient solution is determined. The characteristic equation is $m^2 + 20m + 100 = 0$, for which the roots are $m_{1,2} = -10$. Thus the circuit is critically damped, and the current through the inductor can be expressed by

$$i_2(t) = i_2(t)_t = A_1e^{-10t} + A_2te^{-10t} \tag{3.54}$$

Two equations and two initial conditions are necessary to evaluate the two constants A_1 and A_2. Equation (3.54) and its derivative,

$$Di_2(t) = -10A_1e^{-10t} + A_2(1 - 10t)e^{-10t} \tag{3.55}$$

are utilized in conjunction with the initial conditions $i_2(0^+)$ and $Di_2(0^+)$. In this example the currents just before the switch is closed are

$i_1(0^-) = i_2(0^-) = 0$. The energy stored in the magnetic field of a single inductor is $W = \frac{1}{2}Li^2$. Since this energy cannot change instantly, the current through an inductor cannot change instantly either. Therefore, $i_2(0^+) = i_2(0^-) = 0$. $Di_2(t)$ is found from the original circuit equation, Eq. (3.51), which is rewritten as

$$Di_2(t) = 10i_1(t) - 25i_2(t) - v_c(t) \tag{3.56}$$

To determine $Di_2(0^+)$, it is necessary first to determine $i_1(0^+)$, $i_2(0^+)$, and $v_c(0^+)$. Since $i_2(0^+) = 0$, Eq. (3.50) yields $i_1(0^+) = 0.5$ A. Since the energy $W = \frac{1}{2}Cv^2$ stored in a capacitor cannot change instantly, the voltage across the capacitor cannot change instantly either. The steady-state value of capacitor voltage for $t < 0$ is $v_c(0^-) = 10$ V; thus

$$v_c(0^-) = v_c(0^+) = 10\text{V} \tag{3.57}$$

Inserting these values into Eq. (3.56) yields $Di_2(0^+) = -5$. Substituting the initial conditions into Eqs. (3.54) and (3.55) results in $A_1 = 0$ and $A_2 = -5$. Therefore, the current through the inductor for $t \geq 0$ is

$$i_2(t) = -5te^{-10t} \tag{3.58}$$

3.9 SECOND-ORDER TRANSIENTS [2]

The response to a unit step-function input is usually used as a means of evaluating the response of a system. The example of Sec. 3.7 is used as an illustrative simple second-order system. The differential equation given by Eq. (3.42) can be expressed in the form

$$\frac{D^2 c}{\omega_n^2} + \frac{2\zeta}{\omega_n} Dc + c = r$$

This is defined as a *simple* second-order equation because there are no derivatives of r on the right side of the equation. The underdamped response ($\zeta < 1$) to a unit step input, subject to zero initial conditions, is derived in Sec. 3.7 and is given by

$$c(t) = 1 - \frac{e^{-\zeta\omega_n t}}{\sqrt{1 - \zeta^2}} \sin\left(\omega_n \sqrt{1 - \zeta^2}\, t + \cos^{-1} \zeta\right) \tag{3.59}$$

A family of curves representing this equation is shown in Fig. 3.6, where the abscissa is the dimensionless variable $\omega_n t$. The curves are thus a function only of the damping ratio ζ.

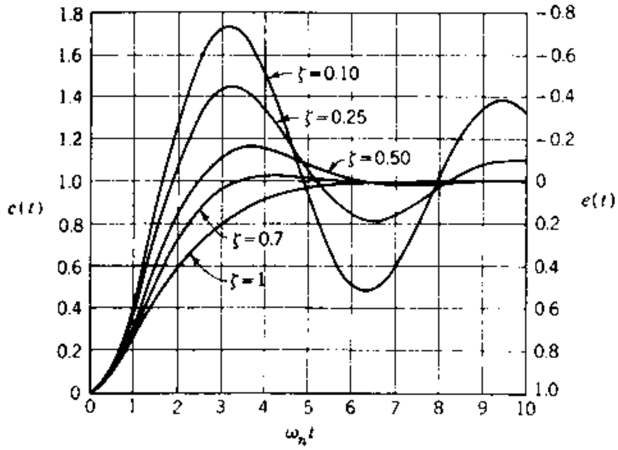


FIGURE 3.6 Simple second-order transients.

These curves show that the amount of overshoot depends on the damping ratio ζ. For the overdamped and critically damped case, $\zeta \geq 1$, there is no overshoot. For the underdamped case, $\zeta < 1$, the system oscillates around the final value. The oscillations decrease with time, and the system response approaches the final value. The peak overshoot for the underdamped system is the first overshoot. The time at which the peak overshoot occurs, t_p, can be found by differentiating $c(t)$ from Eq. (3.59) with respect to time and setting this derivative equal to zero:

$$\begin{aligned}\frac{dc}{dt} &= \frac{\zeta\omega_n e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}} \sin(\omega_n\sqrt{1-\zeta^2}t + \cos^{-1}\zeta) \\ &\quad - \omega_n e^{-\zeta\omega_n t}\cos(\omega_n\sqrt{1-\zeta^2}t + \cos^{-1}\zeta) \\ &= 0\end{aligned}$$

This derivative is zero at $\omega_n\sqrt{1-\zeta^2}t = 0, \pi, 2\pi, \ldots$. The peak overshoot occurs at the first value after zero, provided there are zero initial conditions; therefore,

$$t_p = \frac{\pi}{\omega_n\sqrt{1-\zeta^2}} \tag{3.60}$$

Inserting this value of time in Eq. (3.59) gives the normalized peak overshoot with a step input as $M_p = c(t_p)/R_0$:

$$M_p = c_p = 1 + \exp\left(-\frac{\zeta\pi}{\sqrt{1-\zeta^2}}\right) \tag{3.61}$$

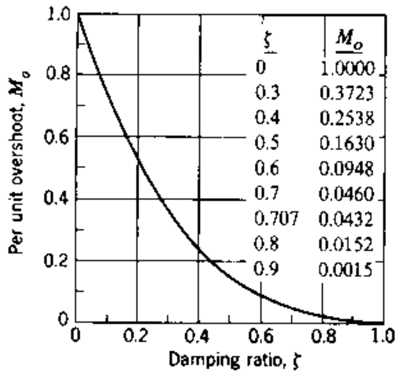


FIGURE 3.7 Peak overshoot vs. damping ratios for a simple second-order equation $D^2c/\omega_n^2 + (2\varsigma/\omega_n)Dc + c = r(t)$.

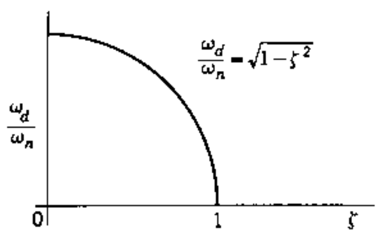


FIGURE 3.8 Frequency of oscillation vs. damping ratio.

The per unit overshoot M_o as a function of damping ratio is shown in Fig. 3.7, where

$$M_o = \frac{c_p - c_{ss}}{c_{ss}} \tag{3.62}$$

The variation of the frequency of oscillation of the transient with variation of damping ratio is also of interest. In order to represent this variation by one curve, the quantity ω_d/ω_n is plotted against ζ in Fig. 3.8. If the scales of ordinate and abscissa are equal, the curve is an arc of a circle. Note that this curve has been plotted for $\zeta \leq 1$. Values of damped natural frequency for $\zeta > 1$ are mathematical only, not physical.

The error in the system is the input minus the output; thus the error equation is

$$e = r - c = \frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}} \sin\left(\omega_n\sqrt{1-\zeta^2}\,t + \cos^{-1}\zeta\right) \tag{3.63}$$

The variation of error with time is sometimes plotted. These curves can be obtained from Fig. 3.6 by realizing that the curves start at $e(0) = +1$ and have the final value $e(\infty) = 0$.

Response Characteristics

The transient-response curves for the second-order system show a number of significant characteristics. The overdamped system is slow-acting and does not oscillate about the final position. For some applications the absence of oscillations may be necessary. For example, an elevator cannot be allowed to oscillate at each stop. But for systems where a fast response is necessary, the slow response of an overdamped system cannot be tolerated.

The underdamped system reaches the final value faster than the overdamped system, but the response oscillates about this final value. If this oscillation can be tolerated, the underdamped system is faster acting. The amount of permissible overshoot determines the desirable value of the damping ratio. For example, a damping ratio $\zeta = 0.4$ has an overshoot of 25.4 percent, and a damping ratio $\zeta = 0.8$ has an overshoot of 1.52 percent.

The settling time is the time required for the oscillations to decrease to a specified absolute percentage of the final value and thereafter to remain less than this value. Errors of 2 or 5 percent are common values used to determine settling time. For second-order systems the value of the transient component at any time is equal to or less than the exponential $e^{-\zeta\omega_n t}$. The value of this term is given in Table 3.1 for several values of t expressed in a number of time constants T.

The settling time for a 2 percent error criterion is approximately 4 time constants; for a 5 percent error criterion, it is 3 time constants. The percent error criterion used must be determined from the response desired for the system. The time for the envelope of the transient

Table 3.1 Exponential Values

t	$e^{-\zeta\omega_n t}$	Error, %
$1T$	0.368	36.8
$2T$	0.135	13.5
$3T$	0.050	5.0
$4T$	0.018	1.8
$5T$	0.007	0.7

to die out is

$$T_s = \frac{\text{number of time constants}}{\zeta\omega_n} \tag{3.64}$$

Since ζ must be determined and adjusted for the permissible overshoot, the undamped natural frequency determines the settling time. When $c(t)_{ss} = 0$, then t_s may be based upon ± 2 percent of $c(t_p)$.

3.10 TIME-RESPONSE SPECIFICATIONS [3]

The desired performance characteristics of a tracking system of any order may be specified in terms of the transient response to a unit step-function input. The performance of a system may be evaluated in terms of the following quantities, as shown in Fig. 3.9.

1. Peak overshoot c_p, which is the magnitude of the largest overshoot, often occurring at the first overshoot. This may also be expressed in percent of the final value.
2. Time to maximum overshoot t_p is the time to reach the maximum overshoot.
3. Time to first zero error t_0, which is the time required to reach the final value the first time. It is often referred to as *duplicating time.*
4. Settling time t_s, the time required for the output response to first reach and thereafter remain within a prescribed percentage of the final value. This percentage must be specified in the individual case. Common values used for settling time are 2 and 5 percent. As commonly used, the 2 or 5 percent is applied to the envelope that yields T_s. The actual t_s may be smaller than T_s.
5. Rise time t_r, defined as the time for the response, on its initial rise, to go from 0.1 to 0.9 times the steady-state value.

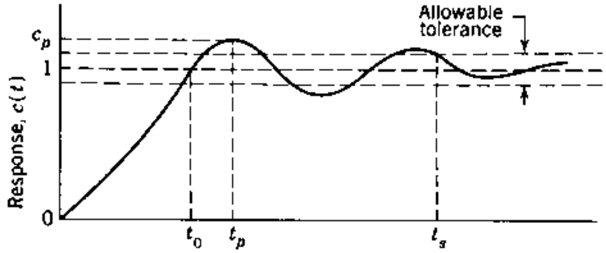


FIGURE 3.9 Typical underdamped response to a step input.

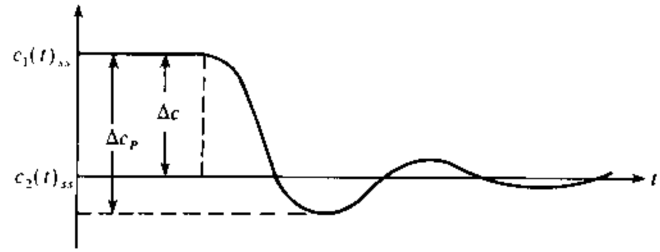


FIGURE 3.10 Underdamped transient.

6. Frequency of oscillation ω_d of the transient depends on the physical characteristics of the system.

If the system has an initial constant output $c_1(t)_{ss}$ and then the input is changed to obtain a new constant value $c_2(t)_{ss}$, the transient response for an underdamped system is illustrated in Fig. 3.10. The per unit overshoot may then be redefined from the expression in Eq. (3.62) to the new value

$$M_o = \frac{\text{maximum overshoot}}{\text{signal transition}} = \frac{|\Delta c_p| - |\Delta c|}{|\Delta c|} \tag{3.65}$$

The time response differs for each set of initial conditions. Therefore, to compare the time responses of various systems it is necessary to start with standard initial conditions. The most practical standard is to start with the system at rest. Then the response characteristics, such as maximum overshoot and settling time, can be compared meaningfully.

For some systems these specifications are also applied for a ramp input. In such cases the plot of error with time is used with the definitions. For systems subject to shock inputs the response due to an impulse is used as a criterion of performance.

3.11 CAD ACCURACY CHECKS (CADAC)

A proficient control system engineer is an individual who has mastered control theory and its application in analysis and design. The competent engineer must become *computer literate* in the use of CAD packages like MATLAB [4] or TOTAL-PC. The individual must fully comprehend the importance of the following factors:

1. It is essential to have a firm understanding of the CAD algorithm being used. This leads to a judgment about the reasonableness of the CAD program results.

2. CAD input data must be entered accurately. Remember: *garbage in–garbage out*.
3. No computer has infinite accuracy. Therefore, round-off must be considered.
4. CAD accuracy checks (CADAC) help to validate the accuracy of the CAD output data. *The CADAC are necessary, but not necessarily sufficient, conditions for ensuring the accuracy of the CAD output data*.

The figures of merit (FOM) equations for M_p, t_p, and T_s can serve as CADAC. They apply not only for simple second-order systems but also for systems in which the dominant roots (a pair of complex roots or a real root) of the characteristic equation dictate the time response characteristics. These CADAC are important for the design methods presented in later chapters.

3.12 STATE-VARIABLE EQUATIONS [5–8]

The differential equations and the corresponding state equations of various physical systems are derived in Chap. 2. The state variables selected in Chap. 2 are restricted to the energy-storage variables. In Chap. 5 different formulations of the state variables are presented. In large-scale systems, with many inputs and outputs, the state-variable approach can have distinct advantages over conventional methods, especially when digital computers are used to obtain the solutions. Although this text is restricted to linear time-invariant (LTI) systems, the state-variable approach is applicable to nonlinear and to time-varying systems. In these cases a computer is a practical method for obtaining the solution. A feature of the state-variable method is that it decomposes a complex system into a set of smaller systems that can be normalized to have a minimum interaction and that can be solved individually. Also, it provides a unified approach that is used extensively in modern control theory.

The block diagram of Fig. 3.11 represents a system S that has m inputs, l outputs, and n state variables. The coefficients in the equations

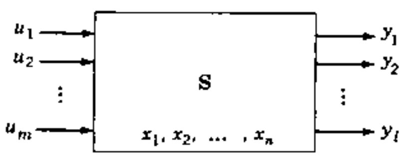


FIGURE 3.11 General system representation.

representing an LTI system are constants. The matrix state and output equations are then

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \tag{3.66}$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) \tag{3.67}$$

The variables $\mathbf{x}(t)$, $\mathbf{u}(t)$, and $\mathbf{y}(t)$ are column vectors, and $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are matrices having constant elements. Equation (3.66) is solved first for the state vector $\mathbf{x}(t)$. This result is then used in Eq. (3.67) to determine the output $\mathbf{y}(t)$.

The homogeneous state equation, with the input $\mathbf{u}(t) = 0$, is

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \tag{3.68}$$

where $\mathbf{A}$ is a constant $n \times n$ matrix and $\mathbf{x}$ is an $n \times 1$ column vector.

For the scalar first-order equation $\dot{x} = ax$, the solution, in terms of the initial conditions at time $t = 0$, is

$$x(t) = e^{at}x(0) \tag{3.69}$$

For any other initial condition, at time $t = t_0$, the solution is

$$x(t) = e^{a(t-t_0)}x(t_0) \tag{3.70}$$

Comparing the scalar and the state equations shows the solution of Eq. (3.68) to be analogous to the solution given by Eq. (3.70); it is

$$\mathbf{x}(t) = \exp[\mathbf{A}(t - t_0)]\mathbf{x}(t_0) \tag{3.71}$$

The exponential function of a scalar that appears in Eq. (3.69) can be expressed as the infinite series

$$e^{at} = \exp[at] = 1 + \frac{at}{1!} + \frac{(at)^2}{2!} + \frac{(at)^3}{3!} + \ldots + \frac{(at)^k}{k!} + \ldots \tag{3.72}$$

The analogous exponential function of a square matrix $\mathbf{A}$ that appears in Eq. (3.71), with $t_0 = 0$, is

$$e^{\mathbf{A}t} = \exp[\mathbf{A}t] = \mathbf{I} + \frac{\mathbf{A}t}{1!} + \frac{(\mathbf{A}t)^2}{2!} + \frac{(\mathbf{A}t)^3}{3!} + \ldots + \frac{(\mathbf{A}t)^k}{k!} + \ldots \tag{3.73}$$

Thus $\exp[\mathbf{A}t]$ is a square matrix of the same order as $\mathbf{A}$. A more useful form for the infinite series of Eq. (3.73) is the closed form developed in Sec. 3.14. The closed form is called the *state transition matrix* (STM) or the *fundamental matrix* of the system and is denoted by

$$\boldsymbol{\Phi}(t) = e^{\mathbf{A}t} = \exp[\mathbf{A}t] \tag{3.74}$$

The term STM is descriptive of the unforced or natural response and is the expression preferred by engineers. The STM has the following properties [7,9].

1. $\mathbf{\Phi}(t_2 - t_1)\mathbf{\Phi}(t_1 - t_0) = \mathbf{\Phi}(t_2 - t_0)$ for any t_0, t_1, t_2 (3.75)
2. $\mathbf{\Phi}(t)\mathbf{\Phi}(t)\cdots\mathbf{\Phi}(t) = \mathbf{\Phi}^q(t) = \mathbf{\Phi}(qt)$ $q =$ positive integer (3.76)
3. $\mathbf{\Phi}^{-1}(t) = \mathbf{\Phi}(-t)$ (3.77)
4. $\mathbf{\Phi}(0) = \mathbf{I}$ unity matrix (3.78)
5. $\mathbf{\Phi}(t)$ is nonsingular for all finite values of t (3.79)

3.13 CHARACTERISTIC VALUES

Consider a system of equations represented by

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \tag{3.80}$$

The signals $\dot{\mathbf{x}}$ and $\mathbf{x}$ are column vectors, and $\mathbf{A}$ is a square matrix of order n. One case for which a solution of this equation exists is if $\mathbf{x}$ and $\dot{\mathbf{x}}$ have the same direction in the state space but differ only in magnitude by a scalar proportionality factor λ. The solution must therefore have the form $\dot{\mathbf{x}} = \lambda\mathbf{x}$. Inserting this into Eq. (3.80) and rearranging terms yields

$$[\lambda\mathbf{I} - \mathbf{A}]\mathbf{x} = \mathbf{0}$$

This equation has a nontrivial solution only if $\mathbf{x}$ is not zero. It is therefore required that the matrix $[\lambda\mathbf{I} - \mathbf{A}]$ not have full rank; therefore, the determinant of the coefficients of $\mathbf{x}$ must be zero:

$$Q(\lambda) \equiv |\lambda\mathbf{I} - \mathbf{A}| = 0 \tag{3.81}$$

When $\mathbf{A}$ is of order n, the resulting polynomial equation is the *characteristic equation*

$$Q(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = 0 \tag{3.82}$$

The roots λ_i of the characteristic equation are called the *characteristic values* or *eigenvalues* of $\mathbf{A}$. The roots may be distinct (simple) or repeated with a multiplicity p. Also, a root may be real or complex. Complex roots must appear in conjugate pairs, and the set of roots containing the complex-conjugate roots is said to be *self-conjugate*. The polynomial $Q(\lambda)$ may be written in factored form as

$$Q(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n) \tag{3.83}$$

The product of *eigenvalues* of a matrix **A** is equal to its determinant, that is, $\lambda_1\lambda_2 \cdots \lambda_n = |\mathbf{A}|$. Also, the sum of the *eigenvalues* is equal to the sum of the elements on the main diagonal (the trace) of **A**, that is,

$$\sum_{l}^{n} \lambda_i = \sum_{i=1}^{n} a_{ii} \equiv \text{trace } \mathbf{A}$$

3.14 EVALUATING THE STATE TRANSITION MATRIX

There are several methods for evaluating the STM $\boldsymbol{\Phi}(t) = \exp[\mathbf{A}t]$ in closed form for a given matrix **A**. The method illustrated here is based on the Cayley-Hamilton theorem. The Laplace transform method is covered in the next chapter, and a state transformation method is covered in Chap. 5. Consider a general polynomial of the form

$$N(\lambda) = \lambda^m + C_{m-1}\lambda^{m-1} + \cdots + C_1\lambda + C_0 \tag{3.84a}$$

When the polynomial $N(\lambda)$ is divided by the characteristic polynomial $Q(\lambda)$, the result is

$$\frac{N(\lambda)}{Q(\lambda)} = F(\lambda) + \frac{R(\lambda)}{Q(\lambda)}$$

or

$$N(\lambda) = F(\lambda)Q(\lambda) + R(\lambda) \tag{3.84b}$$

The function $R(\lambda)$ is the remainder, and it is a polynomial whose maximum order is $n-1$, or 1 less than the order of $Q(\lambda)$. For $\lambda = \lambda_i$ the value $Q(\lambda_i) = 0$; thus

$$N(\lambda_i) = R(\lambda_i) \tag{3.85}$$

The matrix polynomial corresponding to Eq. (3.84*a*), using **A** as the variable, is

$$\mathbf{N}(\mathbf{A}) = \mathbf{A}^m + C_{m-1}\mathbf{A}^{m-1} + \cdots + C_1\mathbf{A} + C_0\mathbf{I} \tag{3.86}$$

Since the characteristic equation $Q(\lambda) = 0$ has n roots, there are n equations $Q(\lambda_1) = 0$, $Q(\lambda_2) = 0, \ldots, Q(\lambda_n) = 0$. The analogous matrix equation is

$$\mathbf{Q}(\mathbf{A}) = \mathbf{A}^n + a_{n-1}\mathbf{A}^{n-1} + \cdots + a_1\mathbf{A} + a_0\mathbf{I} = \mathbf{0}$$

where **0** indicates a null matrix of the same order as **Q**(**A**). This equation implies the Cayley-Hamilton theorem, which is sometimes expressed as "every square matrix **A** satisfies its own characteristic equation." The matrix polynomial corresponding to Eqs. (3.84*b*) and (3.85) is therefore

$$\mathbf{N}(\mathbf{A}) = \mathbf{F}(\mathbf{A})\mathbf{Q}(\mathbf{A}) + \mathbf{R}(\mathbf{A}) = \mathbf{R}(\mathbf{A}) \tag{3.87}$$

Equations (3.85) and (3.87) are valid when $N(\lambda)$ is a polynomial of any order (or even an infinite series) as long as it is analytic. The exponential function $N(\lambda) = e^{\lambda t}$ is an analytic function that can be represented by an infinite series, as shown in Eq. (3.72). Since this function converges in the region of analyticity, it can be expressed in closed form by a polynomial in λ of degree $n-1$. Thus, for each eigenvalue, from Eq. (3.85),

$$e^{\lambda_i t} = R(\lambda_i) = \alpha_0(t) + \alpha_1(t)\lambda_i + \cdots + \alpha_k(t)\lambda_i^k + \cdots + \alpha_{n-1}(t)\lambda_i^{n-1} \tag{3.88}$$

Inserting the n distinct roots λ_i into Eq. (3.88) yields n equations that can be solved simultaneously for the coefficients α_k. These coefficients may be inserted into the corresponding matrix equation:

$$\begin{aligned} e^{\mathbf{A}t} &= \mathbf{N}(\mathbf{A}) = \mathbf{R}(\mathbf{A}) \\ &= \alpha_0(t)\mathbf{I} + \alpha_1(t)\mathbf{A} + \cdots + \alpha_k(t)\mathbf{A}^k + \cdots + \alpha_{n-1}(t)\mathbf{A}^{n-1} \end{aligned} \tag{3.89}$$

For the STM this yields

$$\boldsymbol{\Phi}(t) = \exp[\mathbf{A}t] = \sum_{k=0}^{n-1} \alpha_k(t)\mathbf{A}^k \tag{3.90}$$

For the case of multiple roots, refer to a text on linear algebra [5,8,9] for the evaluation of $\boldsymbol{\Phi}(t)$.

Example. Find $\exp[\mathbf{A}t]$ when

$$\mathbf{A} = \begin{bmatrix} 0 & 6 \\ -1 & -5 \end{bmatrix}$$

The characteristic equation is

$$Q(\lambda) = |\lambda\mathbf{I} - \mathbf{A}| = \begin{vmatrix} \lambda & -6 \\ 1 & \lambda + 5 \end{vmatrix} = \lambda^2 + 5\lambda + 6 = 0$$

The roots of this equation are $\lambda_1 = -2$ and $\lambda_2 = -3$. Since $\mathbf{A}$ is a second-order matrix, the remainder polynomial given by Eq. (3.88) is of first order:

$$N(\lambda_i) = R(\lambda_i) = \alpha_0 + \alpha_1\lambda_i \tag{3.91}$$

Substituting $N(\lambda_i) = e^{\lambda_i t}$ and the two roots λ_1 and λ_2 into Eq. (3.91) yields

$$e^{-2t} = \alpha_0 - 2\alpha_1 \qquad \text{and} \qquad e^{-3t} = \alpha_0 - 3\alpha_1$$

Solving these equations for the coefficients α_0 and α_1 yields

$$\alpha_0 = 3e^{-2t} - 2e^{-3t} \qquad \text{and} \qquad \alpha_1 = e^{-2t} - e^{-3t}$$

The STM obtained by using Eq. (3.90) is

$$\Phi(t) = \exp[\mathbf{A}t] = \alpha_0\mathbf{I} + \alpha_1\mathbf{A} + \begin{bmatrix} \alpha_0 & 0 \\ 0 & \alpha_0 \end{bmatrix} + \begin{bmatrix} 0 & 6\alpha_1 \\ -\alpha_1 & -5\alpha_1 \end{bmatrix}$$

$$= \begin{bmatrix} 3e^{-2t} - 2e^{-3t} & 6e^{-2t} - 6e^{-3t} \\ -e^{-2t} + e^{-3t} & -2e^{-2t} + 3e^{-3t} \end{bmatrix} \tag{3.92}$$

Another procedure for evaluating the STM, $\Phi(t)$, is by use of the Sylvester expansion. The format in this method makes it especially useful with a digital computer. The method is presented here without proof for the case when **A** is a square matrix with n distinct eigenvalues. The polynomial **N(A)** can be written

$$e^{\mathbf{A}t} = \mathbf{N}(\mathbf{A}) = \sum_{i=1}^{n} N(\lambda_i)\mathbf{Z}_i(\lambda) \tag{3.93}$$

$$\mathbf{Z}_i(\lambda) = \frac{\sum_{j=1, j\neq 1}^{n} (\mathbf{A} - \lambda_j\mathbf{I})}{\sum_{j=1, j\neq i}^{n} (\lambda_i - \lambda_j)} \tag{3.94}$$

For the previous example there are two eigenvalues $\lambda_1 = -2$ and $\lambda_2 = -3$. The two matrices $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are evaluated from Eq. (3.94):

$$\mathbf{Z}_1 = \frac{\mathbf{A} - \lambda_2\mathbf{I}}{\lambda_1 - \lambda_2} = \frac{\begin{bmatrix} 0+3 & 6 \\ -1 & -5+3 \end{bmatrix}}{-2-(-3)} = \begin{bmatrix} 3 & 6 \\ -1 & -2 \end{bmatrix}$$

$$\mathbf{Z}_2 = \frac{\mathbf{A} - \lambda_1\mathbf{I}}{\lambda_2 - \lambda_1} = \frac{\begin{bmatrix} 0+2 & 6 \\ -1 & -5+2 \end{bmatrix}}{-3-(-2)} = \begin{bmatrix} -2 & -6 \\ 1 & 3 \end{bmatrix}$$

Using these values in Eq. (3.93), where $N(\lambda_i) = e^{\lambda_i t}$, yields

$$\Phi(t) = \exp[\mathbf{A}t] = e^{\lambda_1 t}\mathbf{Z}_1 + e^{\lambda_2 t}\mathbf{Z}_2$$

Equation (3.92) contains the value of $\Phi(t)$.

The matrices $\mathbf{Z}_i(\lambda)$ are called the *constituent matrices* of **A** and have the following properties:

1. Orthogonality:
$$\mathbf{Z}_i\mathbf{Z}_j = \mathbf{0} \qquad i \neq j \tag{3.95}$$
2. The sum of the complete set is equal to the unit matrix:
$$\sum_{i=1}^{n} \mathbf{Z}_i = \mathbf{I} \tag{3.96}$$
3. Idempotency:
$$\mathbf{Z}_i^r = \mathbf{Z}_i \tag{3.97}$$
where $r > 0$ is any positive integer.

Equation (3.93) can be conveniently applied to evaluate other functions of **A**. For example, to obtain $\mathbf{A}^r$ use the form

$$\mathbf{A}^r = \lambda_1^r \mathbf{Z}_1 + \lambda_2^r \mathbf{Z}_2 + \cdots + \lambda_n^r \mathbf{Z}_n = \sum_{i=1}^{n} \lambda_i^r \mathbf{Z}_i \tag{3.98}$$

The operations involved in computing Eq. (3.98) consist primarily of addition, which is simpler than multiplying **A** by itself r times.

3.15 COMPLETE SOLUTION OF THE STATE EQUATION [10]

When an input $\mathbf{u}(t)$ is present, the complete solution for $\mathbf{x}(t)$ is obtained from Eq. (3.66). The starting point for obtaining the solution is the following equality, obtained by applying the rule for the derivative of the product of two matrices (see Appendix B):

$$\frac{d}{dt}\left[e^{-\mathbf{A}t}\mathbf{x}(t)\right] = e^{-\mathbf{A}t}[\dot{\mathbf{x}}(t) - \mathbf{A}\mathbf{x}(t)] \tag{3.99}$$

Utilizing Eq. (3.66), $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$, in the right side of Eq. (3.99) yields

$$\frac{d}{dt}\left[e^{-\mathbf{A}t}\mathbf{x}(t)\right] = e^{-\mathbf{A}t}\mathbf{B}\mathbf{u}(t)$$

Integrating this equation between 0 and t gives

$$e^{-\mathbf{A}t}\mathbf{x}(t) - \mathbf{x}(0) = \int_0^t e^{-\mathbf{A}\tau}\mathbf{B}\mathbf{u}(\tau)\, d\tau \tag{3.100}$$

Multiplying by $e^{\mathbf{A}t}$ on the left and rearranging terms produces

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)\, d\tau = \mathbf{\Phi}(t)\mathbf{x}(0) + \int_0^t \mathbf{\Phi}(t-\tau)\mathbf{B}\mathbf{u}(\tau)\, d\tau \tag{3.101}$$

Using the STM and generalizing for initial conditions at time $t = t_0$ gives the solution to the state-variable equation with an input $\mathbf{u}(t)$ as

$$\mathbf{x}(t) = \mathbf{\Phi}(t-t_0)\mathbf{x}(t_0) + \int_{t_0}^t \mathbf{\Phi}(t-\tau)\mathbf{B}\mathbf{u}(\tau)\, d\tau \qquad t > t_0 \tag{3.102}$$

This equation is called the *state transition equation*; i.e., it describes the change of state relative to the initial conditions $\mathbf{x}(t_0)$ and the input $\mathbf{u}(t)$. It may be more convenient to change the variables in the integral of Eq. (3.101) by letting $\beta = t - \tau$, which gives

$$\mathbf{x}(t) = \mathbf{\Phi}(t)\,\mathbf{x}(0) + \int_0^t \mathbf{\Phi}(\beta)\mathbf{B}\mathbf{u}(t-\beta)\, d\beta \tag{3.103}$$

where $\boldsymbol{\Phi}(t) = e^{At}$ is the *state transition matrix*. Note that $\boldsymbol{\Phi}(\mathbf{s})\mathbf{x}(0)$ is the zero-input response and $\boldsymbol{\Phi}(\mathbf{s})\mathbf{BU}(\mathbf{s})$ is the zero-state response.

In modern terminology the state solution $\mathbf{x}(t)$ as given by Eqs. (3.101) to (3.103) is described by

$$\mathbf{x}(t) = \mathbf{x}_{zi}(t) + \mathbf{x}_{zs}(t) \tag{3.104}$$

where $\mathbf{x}_{zi}(t)$ is called the *zero-input* response, that is, $\mathbf{u}(t) = \mathbf{0}$, and $\mathbf{x}_{zs}(t)$ is called the *zero-state* response, that is, $\mathbf{x}(t_0) = \mathbf{0}$.

Example. Solve the following state equation for a unit step-function scalar input, $u(t) = u_{-1}(t)$:

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 6 \\ -1 & -5 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}u(t)$$

The STM for this equation is given by Eq. (3.92). Thus, the total time solution is obtained by substituting this value of $\boldsymbol{\Phi}(t)$ into Eq. (3.103). Since $u(t-\beta)$ has the value of unity for $0 < \beta < t$, Eq. (3.103) yields

$$\begin{aligned}
\mathbf{x}(t) &= \boldsymbol{\Phi}(t)\mathbf{x}(0) + \int_0^t \begin{bmatrix} 6e^{-2\beta} - 6e^{-3\beta} \\ -2e^{-2\beta} + 3e^{-3\beta} \end{bmatrix} d\beta \\
&= \begin{bmatrix} 3e^{-2t} - 2e^{-3t} & 6e^{-2t} - 6e^{-3t} \\ -e^{-2t} + e^{-3t} & -2e^{-2t} + 3e^{-3t} \end{bmatrix}\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} \\
&\quad + \begin{bmatrix} 1 - 3e^{-2t} + 2e^{-3t} \\ e^{-2t} - e^{-3t} \end{bmatrix}
\end{aligned} \tag{3.105}$$

The integral of a matrix is the integral of each element in the matrix with respect to the indicated scalar variable. This property is used in evaluating the particular integral in Eq. (3.105). Inserting the initial conditions produces the final solution for the state transition equation. The Laplace transform method presented in Chap. 4 is a more direct method for finding this solution. An additional method is shown in Chap. 5.

3.16 SUMMARY

This chapter establishes the manner of solving differential equations. The steady-state and transient solutions are determined separately and then added to form the total response. Then the constants of the transient component of the solution are evaluated to satisfy the initial conditions. The ability to anticipate the form of the response is very important. The solution of differential equations has been extended to include solution of the matrix state and output equations. This method is also applicable to multiple-input multiple-output, time-varying, and nonlinear systems. The matrix formulation of these systems lends itself to digital-computer solutions. Solution of differential and state equations is facilitated by use of the MATLAB computer aided Design method. Use of MATLAB is described in Appendix C, and a number of examples are contained throughout this text. Obtaining the complete solution by means of the Laplace transform is covered in the next chapter.

REFERENCES

1. Kreyszig, E.: *Advanced Engineering Mathematics*, 7th ed., John Wiley & Sons, New York, 1993.
2. Kinariwala, B.K., F.F. Kuo, and N.K. Tsao: *Linear Circuits and Computation*, John Wiley & Sons, New York, 1973.
3. D'Azzo, J. J., and C. H. Houpis: *Feedback Control System Analysis and Synthesis*, 2nd ed., McGraw-Hill, New York, 1966.
4. Saadat, H.: *Computational Aids in Control Systems Using MATLAB*™, McGraw-Hill, New York, 1993.
5. Rugh, W.J.: *Linear System Theory*, Prentice Hall, Englewood Cliffs, N.J., 1993.
6. Kailath, T.: *Linear Systems*, Prentice-Hall, Englewood Cliffs, N.J., 1980.
7. Ward, J.R., and R.D. Strum: *State Variable Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1970.
8. Wiberg, D.M.: *State Space and Linear Systems*, Schaum's Outline Series, McGraw-Hill, New York, 1971.
9. De Russo, P.M., R.J. Roy and C.M. Close: *State Variables for Engineers*, Krieger, Malabar, Fla., 1990.
10. D'Azzo, J.J., and C.H. Houpis: *Linear Control System Analysis and Design*, 3rd Ed., McGraw-Hill, New York, 1988.

4

Laplace Transform

4.1 INTRODUCTION

The Laplace transform method is used extensively [1–3] to facilitate and systematize the solution of ordinary constant-coefficient differential equations. The advantages of this modern transform method for the analysis of linear-time-invariant (LTI) systems are the following:

1. It includes the boundary or initial conditions.
2. The work involved in the solution is simple algebra.
3. The work is systematized.
4. The use of a table of transforms reduces the labor required.
5. Discontinuous inputs can be treated.
6. The transient and steady-state components of the solution are obtained simultaneously.

The disadvantage of transform methods is that if they are used mechanically, without knowledge of the actual theory involved, they sometimes yield erroneous results. Also, a particular equation can sometimes be solved more simply and with less work by the classical method. Although an understanding of the Laplace transform method is essential, it must be emphasized that the solutions of differential equations are readily obtained by use of CAD packages such as MATLAB and TOTAL-PC [4] (see Appendixes C and D). Laplace transforms are also applied to the solution of system equations that are in matrix state-variable format. The method for using the state and output equations to obtain the system transfer function is presented.

4.2 DEFINITION OF THE LAPLACE TRANSFORM

The direct Laplace transformation of a function of time $f(t)$ is given by

$$\mathscr{L}[f(t)] = \int_0^\infty f(t)e^{-st}\,dt = F(s) \tag{4.1}$$

where $\mathscr{L}[f(t)]$ is a shorthand notation for the Laplace integral. Evaluation of the integral results in a function $F(s)$ that has s as the parameter. This parameter s is a complex quantity of the form $\sigma + j\omega$. Since the limits of integration are zero and infinity, it is immaterial what value $f(t)$ has for negative or zero time.

There are limitations on the functions $f(t)$ that are Laplace-transformable. Basically, the requirement is that the Laplace integral converge, which means that this integral has a definite functional value. To meet this requirement [3] the function $f(t)$ must be (1) piecewise continuous over every finite interval $0 \le t_1 \le t \le t_2$ and (2) of exponential order. A function is piecewise continuous in a finite interval if that interval can be divided into a finite number of subintervals, over each of which the function is continuous and at the ends of each of which $f(t)$ possesses finite right- and left-hand limits. A function $f(t)$ is of exponential order if there exists a constant a such that the product $e^{-at}|f(t)|$ is bounded for all values of t greater than some finite value T. This imposes the restriction that σ, the real part of s, must be greater than a lower bound σ_a for which the product $e^{-\sigma_a t}|f(t)|$ is of exponential order. A linear differential equation with constant coefficients and with a finite number of terms is Laplace transformable if the driving function is Laplace transformable.

All cases covered in this book are Laplace transformable. The basic purpose in using the Laplace transform is to obtain a method of solving differential equations that involves only simple algebraic operations in conjunction with a table of transforms.

4.3 DERIVATION OF LAPLACE TRANSFORMS OF SIMPLE FUNCTIONS

A number of examples are presented to show the derivation of the Laplace transform of several time functions. A list of common transform pairs is given in Appendix A.

Step Function $u_{-1}(t)$

The Laplace transform of the unit step function $u_{-1}(t)$ (see Fig. 3.1*a*) is

$$\mathscr{L}[u_{-1}(t)] = \int_0^\infty u_{-1}(t)e^{-st}\,dt = U_{-1}(s) \tag{4.2}$$

Since $u_{-1}(t)$ has the value 1 over the limits of integration,

$$U_{-1}(s) = \int_0^\infty e^{-st}\,dt = -\left.\frac{e^{-st}}{s}\right|_0^\infty = \frac{1}{s} \qquad \text{if } \sigma > 0 \tag{4.3}$$

The step function is undefined at $t=0$, but this is immaterial, for the integral is defined by a limit process

$$\int_0^\infty f(t)e^{-st}\,dt = \lim_{\substack{T\to\infty \\ \varepsilon\to 0}} \int_\varepsilon^T f(t)e^{-st}\,dt \tag{4.4}$$

and the explicit value at $t=0$ does not affect the value of the integral. The value of the integral obtained by taking limits is implied in each case but is not written out explicitly.

Decaying Exponential $e^{-\alpha t}$

The exponent α is a positive real number.

$$\begin{aligned}\mathscr{L}[e^{-at}] &= \int_0^\infty e^{-\alpha t}e^{-st}\,dt \int_0^\infty e^{-(s+\alpha)t}\,dt \\ &= -\left.\frac{e^{-(s+\alpha)t}}{s+\alpha}\right|_0^\infty = \frac{1}{s+\alpha} \qquad \sigma > -\alpha\end{aligned} \tag{4.5}$$

Sinusoid $\cos\omega t$

Here ω is a positive real number.

$$\mathscr{L}[\cos\omega t] = \int_0^\infty \cos\omega t\, e^{-st}\,dt \tag{4.6}$$

Expressing $\cos\omega t$ in exponential form gives

$$\cos\omega t = \frac{e^{j\omega t} + e^{-j\omega t}}{2}$$

Then

$$\begin{aligned}\mathscr{L}[\cos\omega t] &= \frac{1}{2}\left(\int_0^\infty e^{(j\omega-s)t}\,dt + \int_0^\infty e^{(-j\omega-s)t}\,dt\right) = \frac{1}{2}\left[\frac{e^{(j\omega-s)t}}{j\omega-s} + \frac{e^{(-j\omega-s)t}}{-j\omega-s}\right]_0^\infty \\ &= \frac{1}{2}\left(-\frac{1}{j\omega-s} - \frac{1}{-j\omega-s}\right) = \frac{s}{s^2+\omega^2} \qquad \sigma>0\end{aligned} \tag{4.7}$$

Ramp Function $u_{-2}(t) = tu_{-1}(t)$

$$\mathscr{L}[t] = \int_0^\infty te^{-st}\,dt \qquad \sigma > 0 \tag{4.8}$$

This expression is integrated by parts by using

$$\int_a^b u\,dv = uv\Big|_a^b - \int_a^b v\,du$$

Let $u = t$ and $dv = e^{-st}\,dt$. Then $du = dt$ and $v = -e^{-st}/s$. Thus

$$\int_0^\infty te^{-st}\,dt = -\frac{te^{-st}}{s}\Big|_0^\infty - \int_0^\infty \left(-\frac{e^{-st}}{s}\right) dt$$

$$= 0 - \frac{e^{-st}}{s^2}\Big|_0^\infty = \frac{1}{s^2} \qquad \sigma > 0 \tag{4.9}$$

4.4 LAPLACE TRANSFORM THEOREMS

Several theorems that are useful in applying the Laplace transform are presented in this section. In general, they are helpful in evaluating transforms.

Theorem 1: Linearity. If a is a constant or is independent of s and t, and if $f(t)$ is transformable, then

$$\mathscr{L}[af(t)] = a\mathscr{L}[f(t)] = aF(s) \tag{4.10}$$

Theorem 2: Superposition. If $f_1(t)$ and $f_2(t)$ are both Laplace-transformable, the principle of superposition applies:

$$\mathscr{L}[f_1(t) \pm f_2(t)] = \mathscr{L}[f_1(t)] \pm \mathscr{L}[f_2(t)] = F_1(s) \pm F_2(s) \tag{4.11}$$

Theorem 3: Translation in time. If the Laplace transform of $f(t)$ is $F(s)$ and a is a positive real number, the Laplace transform of the translated function $f(t-a)u_{-1}(t-a)$ is

$$\mathscr{L}[f(t-a)u_{-1}(t-a)] = e^{-as}F(s) \tag{4.12}$$

Translation in the positive t direction in the real domain becomes multiplication by the exponential e^{-as} in the s domain.

Theorem 4: Complex differentiation. If the Laplace transform of $f(t)$ is $F(s)$, then

$$\mathscr{L}[tf(t)] = -\frac{d}{ds}F(s) \tag{4.13}$$

Multiplication by time in the real domain entails differentiation with respect to s in the s domain.

Example 1. Using $\mathscr{L}[\cos \omega t]$ from Eq. (4.7),

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

Example 2. Using $\mathscr{L}[e^{-\alpha t}]$ from Eq. (4.5),

$$\mathscr{L}[te^{-\alpha t}] = -\frac{d}{ds}\mathscr{L}[e^{-\alpha t}] = -\frac{d}{ds}\left(\frac{1}{s+\alpha}\right) = \frac{1}{(s+\alpha)^2}$$

Theorem 5: Translation in the s Domain. If the Laplace transform of $f(t)$ is $F(s)$ and a is either real or complex, then

$$\mathscr{L}[e^{at} f(t)] = F(s - a) \tag{4.14}$$

Multiplication of e^{at} in the real domain becomes translation in the s domain.

Example 3. Starting with $\mathscr{L}[\sin \omega t] = \omega/(s^2 + \omega^2)$ and applying Theorem 5 gives

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

Theorem 6: Real Differentiation. If the Laplace transform of $f(t)$ is $F(s)$, and if the first derivative of $f(t)$ with respect to time $Df(t)$ is transformable, then

$$\mathscr{L}[Df(t)] = sF(s) - f(0^+) \tag{4.15}$$

The term $f(0^+)$ is the value of the right-hand limit of the function $f(t)$ as the origin $t = 0$ is approached from the right side (thus through positive values of time). This includes functions, such as the step function, that may be undefined at $t = 0$. For simplicity, the plus sign following the zero is usually omitted, although its presence is implied.

The transform of the second derivative $D^2 f(t)$ is

$$\mathscr{L}[D^2 f(t)] = s^2 F(s) - sf(0) - Df(0) \tag{4.16}$$

where $Df(0)$ is the value of the limit of the derivative of $f(t)$ as the origin $t = 0$, is approached from the right side.

The transform of the nth derivative $D^n f(t)$ is

$$\mathscr{L}[D^n f(t)] = s^n F(s) - s^{n-1} f(0) - s^{n-2} Df(0) - \cdots - s\, D^{n-2} f(0) \\ - D^{n-1} f(0) \tag{4.17}$$

Note that the transform includes the initial conditions, whereas in the classical method of solution the initial conditions are introduced separately to evaluate the coefficients of the solution of the differential equation. When all initial conditions are zero, the Laplace transform of the nth derivative of $f(t)$ is simply $s^n F(s)$.

Theorem 7: Real Integration. If the Laplace transform of $f(t)$ is $F(s)$, its integral

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

is transformable and the value of its transform is

$$\mathscr{L}[D^{-1} f(t)] = \frac{F(s)}{s} + \frac{D^{-1} f(0^+)}{s} \tag{4.18}$$

The term $D^{-1} f(0^+)$ is the constant of integration and is equal to the value of the integral as the origin is approached from the positive or right side. The plus sign is omitted in the remainder of this text.

The transform of the double integral $D^{-2} f(t)$ is

$$\mathscr{L}\left[D^{-2} f(t) \right] = \frac{F(s)}{s^2} + \frac{D^{-1} f(0)}{s^2} + \frac{D^{-2} f(0)}{s} \tag{4.19}$$

The transform of the nth-order integral $D^{-n} f(t)$ is

$$\mathscr{L}[D^{-n} f(t)] = \frac{F(s)}{s^n} + \frac{D^{-1} f(0)}{s^n} + \cdots + \frac{D^{-n} f(0)}{s} \tag{4.20}$$

Theorem 8: Final Value. If $f(t)$ and $Df(t)$ are Laplace transformable, if the Laplace transform of $f(t)$ is $F(s)$, and if the limit $f(t)$ as $t \to \infty$ exists, then

$$\lim_{s \to 0} sF(s) = \lim_{t \to \infty} f(t) \tag{4.21}$$

This theorem states that the behavior of $f(t)$ in the neighborhood of $t = \infty$ is related to the behavior of $sF(s)$ in the neighborhood of $s = 0$. If $sF(s)$ has poles [values of s for which $|sF(s)|$ becomes infinite] on the imaginary axis (excluding the origin) or in the right-half s plane, there is no finite final value of $f(t)$ and the theorem cannot be used. If $f(t)$ is sinusoidal, the theorem is invalid, since $\mathscr{L}[\sin \omega t]$ has poles at $s = \pm j\omega$ and $\lim_{t\to\infty} \sin \omega t$ does not exist. However, for poles of $sF(s)$ at the origin, $s = 0$, this theorem gives the final value of $f(\infty) = \infty$. This correctly describes the behavior of $f(t)$ as $t \to \infty$.

Theorem 9: Initial Value. If the function $f(t)$ and its first derivative are Laplace transformable, if the Laplace transform of $f(t)$ is $F(s)$, and if $\lim_{s\to\infty} sF(s)$ exists, then

$$\lim_{s\to\infty} sF(s) = \lim_{t\to 0} f(t) \tag{4.22}$$

This theorem states that the behavior of $f(t)$ in the neighborhood of $t = 0$ is related to the behavior of $sF(s)$ in the neighborhood of $|s| = \infty$. There are no limitations on the locations of the poles of $sF(s)$.

Theorem 10: Complex Integration. If the Laplace transform of $f(t)$ is $F(s)$ and if $f(t)/t$ has a limit as $t \to 0^+$, then

$$\mathscr{L}\left[\frac{f(t)}{t}\right] = \int_0^\infty F(s)\, ds \tag{4.23}$$

This theorem states that division by the variable in the real domain entails integration with respect to s in the s domain.

4.5 CAD ACCURACY CHECKS: CADAC

The Laplace transform theorems 8 and 9 for the initial and final values are valuable CADAC. These theorems should be used, when appropriate, to assist in the validation of the computer solutions of control system problems. Additional CADAC are presented throughout this text.

4.6 APPLICATION OF THE LAPLACE TRANSFORM TO DIFFERENTIAL EQUATIONS

The Laplace transform is now applied to the solution of the differential equation for the simple mechanical system that is solved by the classical method in Sec. 3.7. The differential equation of the system is repeated here:

$$MD^2x_2 + BDx_2 + Kx_2 = Kx_1 \tag{4.24}$$

The position $x_1(t)$ undergoes a unit step displacement. This is the input and is called the *driving function*. The unknown quantity for which the equation is to be solved is the output displacement $x_2(t)$, called the *response function*.

The Laplace transform of Eq. (4.24) is

$$\mathscr{L}[Kx_1] = \mathscr{L}[MD^2x_2 + BDx_2 + Kx_2] \tag{4.25}$$

The transform of each term is

$$\mathscr{L}[Kx_1] = KX_1(s)$$

$$\mathscr{L}[Kx_2] = KX_2(s)$$

$$\mathscr{L}[BDx_2] = B[sX_2(s) - x_2(0)]$$

$$\mathscr{L}[MD^2x_2] = M[s^2X_2(s) - sx_2(0) - Dx_2(0)].$$

Substituting these terms into Eq. (4.25) and collecting terms gives

$$KX_1(s) = (Ms^2 + Bs + K)X_2(s) - [Msx_2(0) + MDx_2(0) + Bx_2(0)] \tag{4.26}$$

Equation (4.26) is the transform equation and shows how the initial conditions—the initial position $x_2(0)$ and the initial velocity $Dx_2(0)$—are incorporated into the equation. The function $X_1(s)$ is called the *driving transform*; the function $X_2(s)$ is called the *response transform*. The coefficient of $X_2(s)$, which is $Ms^2 + Bs + K$, is called the *characteristic function*. The equation formed by setting the characteristic function equal to zero is called the *characteristic equation* of the system. Solving for $X_2(s)$ gives

$$X_2(s) = \frac{K}{Ms^2 + Bs + K}X_1(s) + \frac{Msx_2(0) + Bx_2(0) + MDx_2(0)}{Ms^2 + Bs + K} \tag{4.27}$$

The coefficient of $X_1(s)$ is defined as the *system transfer function*. The second term on the right side of the equation is called the *initial condition component*. Combining the terms of Eq. (4.27) yields

$$X_2(s) = \frac{KX_1(s) + Msx_2(0) + Bx_2(0) + MDx_2(0)}{Ms^2 + Bs + K} \tag{4.28}$$

Finding the function $x_2(t)$ whose transform is given by Eq. (4.28) is symbolized by the inverse transform operator $\mathscr{L}^{-1}$; thus

$$\begin{aligned} x_2(t) &= \mathscr{L}^{-1}[X_2(s)] \\ &= \mathscr{L}^{-1}\left[\frac{KX_1(s) + Msx_2(0) + Bx_2(0) + MDx_2(0)}{Ms^2 + Bs + K}\right] \end{aligned} \tag{4.29}$$

The function $x_2(t)$ can be found in the table of Appendix A after inserting numerical values into Eq. (4.29). For this example the system is initially at rest. From the principle of conservation of energy the initial conditions are $x_2(0) = 0$ and $Dx_2(0) = 0$. Assume, as in Sec. 3.7, that the damping ratio ζ is less than unity. Since $x_1(t)$ is a step function, the time response function is

$$x_2(t) = \mathscr{L}^{-1}\left[\frac{K/M}{s(s^2 + Bs/M + K/M)}\right] = \mathscr{L}^{-1}\left[\frac{\omega_n^2}{s(s^2 + 2\zeta\omega_n s + \omega_n^2)}\right] \tag{4.30}$$

where $\omega_n = \sqrt{K/M}$ and $\zeta = B/2\sqrt{KM}$. Reference to transform pair 27*a* in Appendix A provides the solution directly as

$$x_2(t) = 1 - \frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}} \sin\left(\omega_n\sqrt{1-\zeta^2}\,t + \cos^{-1}\zeta\right)$$

4.7 INVERSE TRANSFORMATION

The application of Laplace transforms to a differential equation yields an algebraic equation. From the algebraic equation the transform of the response function is readily found. To complete the solution the inverse transform must be found. In some cases the inverse-transform operation

$$f(t) = \mathscr{L}^{-1}[F(s)] = \frac{1}{j2\pi}\int_{\sigma-j\infty}^{\sigma+j\infty} F(s)e^{st}\,ds \tag{4.31}$$

can be performed by direct reference to transform tables or by use of a digital computer program (see Appendix C). The linearity and translation theorems are useful in extending the tables. When the response transform cannot be found in the tables, the general procedure is to express $F(s)$ as the sum of partial fractions with constant coefficients. The partial fractions have a first-order or quadratic factor in the denominator and are readily found in the table of transforms. The complete inverse transform is the sum of the inverse transforms of each fraction. This procedure is illustrated next.

The response transform $F(s)$ can be expressed, in general, as the ratio of two polynomials $P(s)$ and $Q(s)$. Consider that these polynomials are of degree w and n, respectively, and are arranged in descending order of the powers of the variables s; thus,

$$F(s) = \frac{P(s)}{Q(s)} = \frac{a_w s^w + a_{w-1}s^{w-1} + \cdots + a_1 s + a_0}{s^n + b_{n-1}s^{n-1} + \cdots + b_1 s + b_0} \tag{4.32}$$

The a's and b's are real constants, and the coefficient of the highest power of s in the denominator has been made equal to unity. Only those $F(s)$ that are proper fractions are considered, i.e., those in which n is greater than w.*

*If $n = w$, first divide $P(s)$ by $Q(s)$ to obtain $F(s) = a_w + P_1(s)/Q(s) = a_w + F_1(s)$. Then express $F_1(s)$ as the sum of partial fractions with constant coefficients.

The first step is to factor $Q(s)$ into first-order and quadratic factors with real coefficients:

$$F(s) = \frac{P(s)}{Q(s)} = \frac{P(s)}{(s - s_1)(s - s_2)\cdots(s - s_k)\cdots(s - s_n)} \tag{4.33}$$

The values $s_1, s_2, \ldots, s_n$ in the finite plane that make the denominator equal to zero are called the *zeros* of the denominator. These values of s, which may be either real or complex, also make $|F(s)|$ infinite, and so they are called *poles* of $F(s)$. Therefore, the values $s_1, s_2, \ldots, s_n$ are referred to as zeros of the denominator or poles of the complete function in the finite plane, i.e., there are n poles of $F(s)$. Methods of factoring polynomials exist in the literature. Digital-computer programs are available to perform this operation (see Appendix C) [4].

The transform $F(s)$ can be expressed as a series of fractions. If the poles are simple (nonrepeated), the number of fractions is equal to n, the number of poles of $F(s)$. In such case the function $F(s)$ can be expressed as

$$F(s) = \frac{P(s)}{Q(s)} = \frac{A_1}{s - s_1} + \frac{A_2}{s - s_2} + \cdots + \frac{A_k}{s - s_k} + \cdots + \frac{A_n}{s - s_n} \tag{4.34}$$

The procedure is to evaluate the constants $A_1, A_2, \ldots, A_n$ corresponding to the poles $s_1, s_2, \ldots, s_n$. The coefficients $A_1, A_2, \ldots$ are termed the *residues*† of $F(s)$ at the corresponding poles. Cases of repeated factors and complex factors are treated separately. Several ways of evaluating the constants are shown in the following section.

4.8 HEAVISIDE PARTIAL-FRACTION EXPANSION THEOREMS

The technique of partial-fraction expansion is set up to take care of all cases systematically. There are four classes of problems, depending on the denominator $Q(s)$. Each of these cases is illustrated separately.

Case 1 $F(s)$ has first-order real poles.
Case 2 $F(s)$ has repeated first-order real poles.
Case 3 $F(s)$ has a pair of complex-conjugate poles (a quadratic factor in the denominator).
Case 4 $F(s)$ has repeated pairs of complex-conjugate poles (a repeated quadratic factor in the denominator).

†More generally the residue is the coefficient of the $(s - s_i)^{-1}$ term in the Laurent expansion of $F(s)$ about $s = s_i$.

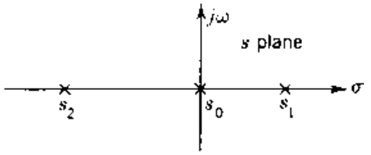


FIGURE 4.1 Location of real poles in the s plane.

Case 1: First-Order Real Poles

The locations of three real poles of $F(s)$ in the s plane are shown in Fig. 4.1. The poles may be positive, zero, or negative, and they lie on the real axis in the s plane. In this example, s_1 is positive, s_0 is zero, and s_2 is negative. For the poles shown in Fig. 4.1 the transform $F(s)$ and its partial fractions are

$$F(s) = \frac{P(s)}{Q(s)} = \frac{P(s)}{s(s - s_1)(s - s_2)} = \frac{A_0}{s} + \frac{A_1}{s - s_1} + \frac{A_2}{s - s_2} \tag{4.35}$$

There are as many fractions as there are factors in the denominator of $F(s)$. Since $s_0 = 0$, the factor $s - s_0$ is written simply as s. The inverse transform of $F(s)$ is

$$f(t) = A_0 + A_1 e^{s_1 t} + A_2 e^{s_2 t} \tag{4.36}$$

The pole s_1 is positive; therefore, the term $A_1 e^{s_1 t}$ is an increasing exponential and the system is unstable. The pole s_2 is negative, and the term $A_2 e^{s_2 t}$ is a decaying exponential with a final value of zero. Therefore, for a system to be stable, all real poles that contribute to the complementary solution must be in the left half of the s plane.

To evaluate a typical coefficient A_k, multiply both sides of Eq. (4.34) by the factor $s - s_k$. The result is

$$\begin{aligned}(s - s_k)F(s) &= (s - s_k)\frac{P(s)}{Q(s)} \\ &= A_1 \frac{s - s_k}{s - s_1} + A_2 \frac{s - s_k}{s - s_2} + \cdots + A_k + \cdots + A_n \frac{s - s_k}{s - s_n}\end{aligned} \tag{4.37}$$

The multiplying factor $s - s_k$ on the left side of the equation and the same factor of $Q(s)$ should be canceled. By letting $s = s_k$, each term on the right side of the equation is zero except A_k. Thus, a general rule for evaluating the constants for single-order real poles is

$$A_k = \left[(s - s_k)\frac{P(s)}{Q(s)}\right]_{s=s_k} = \left[\frac{P(s)}{Q'(s)}\right]_{s=s_k} \tag{4.38}$$

where $Q'(s_k) = [dQ(s)/ds]_{s=s_k} = [Q(s)/(s - s_k)]_{s=s_k}$. The coefficients A_k are the *residues* of $F(s)$ at the corresponding poles. For the case of

$$F(s) = \frac{s+2}{s(s+1)(s+3)} = \frac{A_0}{s} + \frac{A_1}{s+1} + \frac{A_2}{s+3}$$

the constants are

$$A_0 = [sF(s)]_{s=0} = \left[\frac{s+2}{(s+1)(s+3)}\right]_{s=0} = \frac{2}{3}$$

$$A_1 = [(s+1)F(s)]_{s=-1} = \left[\frac{s+2}{s(s+3)}\right]_{s=-1} = -\frac{1}{2}$$

$$A_2 = [(s+3)F(s)]_{s=-3} = \left[\frac{s+2}{s(s+1)}\right]_{s=-3} = -\frac{1}{6}$$

The solution as a function of time is

$$f(t) = \frac{2}{3} - \frac{1}{2}e^{-t} - \frac{1}{6}e^{-3t} \tag{4.39}$$

Case 2: Multiple-Order Real Poles

The position of real poles of $F(s)$, some of which are repeated, is shown in Fig. 4.2. The notation $]_r$ indicates a pole of order r. All real poles lie on the real axis of the s plane. For the poles shown in Fig. 4.2 the transform $F(s)$ and its partial fractions are

$$\begin{aligned} F(s) &= \frac{P(s)}{Q(s)} = \frac{P(s)}{(s-s_1)^3(s-s_2)} \\ &= \frac{A_{13}}{(s-s_1)^3} + \frac{A_{12}}{(s-s_1)^2} + \frac{A_{11}}{s-s_1} + \frac{A_2}{s-s_2} \end{aligned} \tag{4.40}$$

The order of $Q(s)$ in this case is 4, and there are four fractions. Note that the multiple pole s_1, which is of order 3, has resulted in three fractions on the right

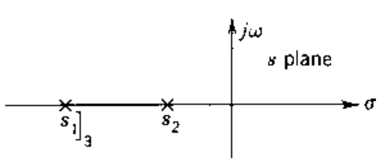


FIGURE 4.2 Location of real poles in the s plane.

side of Eq. (4.40). To designate the constants in the partial fractions, a single subscript is used for a first-order pole. For multiple-order poles a double-subscript notation is used. The first subscript designates the pole, and the second subscript designates the order of the pole in the partial fraction. The constants associated with first-order denominators in the partial-fraction expansion are called residues: therefore, only the constants A_{11} and A_2 are residues of Eq. (4.40).

The inverse transform of $F(s)$ (see transform pairs 7 and 8 in Appendix A) is

$$f(t) = A_{13}\frac{t^2}{2}e^{s_1 t} + A_{12}te^{s_1 t} + A_{11}e^{s_1 t} + A_2 e^{s_2 t} \tag{4.41}$$

For the general transform with repeated real roots,

$$\begin{aligned} F(s) &= \frac{P(s)}{Q(s)} = \frac{P(s)}{(s-s_q)^r(s-s_1)\cdots} \\ &= \frac{A_{qr}}{(s-s_q)^r} + \frac{A_{q(r-1)}}{(s-s_q)^{r-1}} + \cdots + \frac{A_{q(r-k)}}{(s-s_q)^{r-k}} + \cdots \\ &\quad + \frac{A_{q1}}{s-s_q} + \frac{A_1}{s-s_1} + \cdots \end{aligned} \tag{4.42}$$

The constant A_{qr} can be evaluated by simply multiplying both sides of Eq. (4.42) by $(s-s_q)^r$, giving

$$\begin{aligned} (s-s_q)^r F(s) &= \frac{(s-s_q)^r P(s)}{Q(s)} = \frac{P(s)}{(s-s_1)\cdots} \\ &= A_{qr} + A_{q(r-1)}(s-s_q) + \cdots \\ &\quad + A_{q1}(s-s_q)^{r-1} + A_1\frac{(s-s_q)^r}{s-s_1} + \cdots \end{aligned} \tag{4.43}$$

Note that the factor $(s-s_q)^r$ must be divided out of the left side of the equation.

For $s = s_q$, all terms on the right side of the equation are zero except A_{qr}; therefore,

$$A_{qr} = \left[(s-s_q)^r \frac{P(s)}{Q(s)}\right]_{s=s_q} \tag{4.44}$$

Evaluation of $A_{q(r-1)}$ cannot be performed in a similar manner. Multiplying both sides of Eq. (4.42) by $(s-s_q)^{r-1}$ and letting $s = s_q$ would result in both sides being infinite, which leaves $A_{q(r-1)}$ indeterminate. If the term A_{qr} were eliminated from Eq. (4.43), $A_{q(r-1)}$ could be evaluated. This can be done by differentiating Eq. (4.43) with respect to s:

$$\frac{d}{ds}\left[(s-s_q)^r \frac{P(s)}{Q(s)}\right] = A_{q(r-1)} + 2A_{q(r-2)}(s-s_q) + \cdots \tag{4.45}$$

Letting $s = s_q$ gives

$$A_{q(r-1)} = \left\{ \frac{d}{ds} \left[(s - s_q)^r \frac{P(s)}{Q(s)} \right] \right\}_{s=s_q} \tag{4.46}$$

Repeating the differentiation gives the coefficient $A_{q(r-2)}$ as

$$A_{q(r-2)} = \left\{ \frac{1}{2} \frac{d^2}{ds^2} \left[(s - s_q)^r \frac{P(s)}{Q(s)} \right] \right\}_{s=s_q} \tag{4.47}$$

This process can be repeated until each constant is determined. A general formula for finding these coefficients associated with the repeated real pole of order r is

$$A_{q(r-k)} = \left\{ \frac{1}{k!} \frac{d^k}{ds^k} \left[(s - s_q)^r \frac{P(s)}{Q(s)} \right] \right\}_{s=s_q} \tag{4.48}$$

For the case of

$$F(s) = \frac{1}{(s+2)^3(s+3)} = \frac{A_{13}}{(s+2)^3} + \frac{A_{12}}{(s+2)^2} + \frac{A_{11}}{s+2} + \frac{A_2}{s+3} \tag{4.49}$$

the constants are

$$A_{13} = [(s+2)^3 F(s)]_{s=-2} = 1 \qquad A_{12} = \left\{ \frac{d}{ds} [(s+2)^3 F(s)] \right\}_{s=-2} = -1$$

$$A_{11} = \left\{ \frac{d^2}{2ds^2} [(s+2)^3 F(s)] \right\}_{s=-2} = 1 \qquad A_2 = [(s+3)F(s)]_{s=-3} = -1$$

and the solution as a function of time is

$$f(t) = \frac{t^2}{2} e^{-2t} - te^{-2t} + e^{-2t} - e^{-3t} \tag{4.50}$$

A recursive algorithm for computing the partial fraction expansion of rational functions having any number of multiple poles is readily programmed on a digital computer and is given in Ref. 5.

Case 3: Complex-Conjugate Poles

The position of complex poles of $F(s)$ in the s plane is shown in Fig. 4.3. Complex poles always are present in complex-conjugate pairs; their real part may be either positive or negative. For the poles shown in Fig. 4.3 the

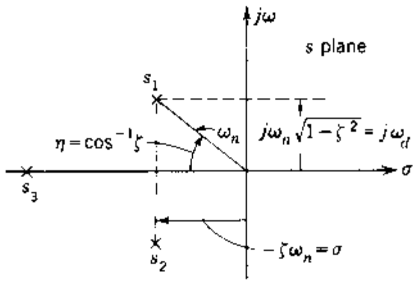


FIGURE 4.3 Location of complex-conjugate poles in the *s* plane.

transform $F(s)$ and its partial fractions are

$$\begin{aligned}F(s) &= \frac{P(s)}{Q(s)} = \frac{P(s)}{(s^2 + 2\zeta\omega_n s + \omega_n^2)(s - s_3)} \\ &= \frac{A_1}{s - s_1} + \frac{A_2}{s - s_2} + \frac{A_3}{s - s_3} \\ &= \frac{A_1}{s + \zeta\omega_n - j\omega_n\sqrt{1-\zeta^2}} + \frac{A_2}{s + \zeta\omega_n + j\omega_n\sqrt{1-\zeta^2}} + \frac{A_3}{s - s_3} \qquad (4.51)\end{aligned}$$

The inverse transform of $F(s)$ is

$$\begin{aligned}f(t) &= A_1 \exp\left[\left(-\zeta\omega_n + j\omega_n\sqrt{1-\zeta^2}\right)t\right] \\ &\quad + A_2 \exp\left[\left(-\zeta\omega_n - j\omega_n\sqrt{1-\zeta^2}\right)t\right] + A_3 e^{s_3 t} \qquad (4.52)\end{aligned}$$

Since the poles s_1 and s_2 are complex conjugates, and since $f(t)$ is a real quantity, the coefficients A_1 and A_2 must also be complex conjugates. Equation (4.52) can be written with the first two terms combined to a more useful damped sinusoidal form:

$$\begin{aligned}f(t) &= 2|A_1|e^{-\zeta\omega_n t}\sin\left(\omega_n\sqrt{1-\zeta^2}\,t + \phi\right) + A_3 e^{s_3 t} \\ &= 2|A_1|e^{\sigma t}\sin(\omega_d t + \phi) + A_3 e^{s_3 t} \qquad (4.53)\end{aligned}$$

where $\phi =$ angle of $A_1 + 90^\circ$, $\omega_d = \omega_n\sqrt{1-\zeta^2}$, and $\sigma = -\zeta\omega_n$.

The values of A_1 and A_3, as found in the manner shown previously, are

$$A_1 = [(s - s_1)F(s)]_{s=s_1} \qquad \text{and} \qquad A_3 = [(s - s_3)F(s)]_{s=s_3}$$

Since s_1 is complex, the constant A_1 is also complex. *Remember that A_1 is associated with the complex pole with the positive imaginary part.*

In Fig. 4.3 the complex poles have a negative real part, $\sigma = -\zeta\omega_n$, where the damping ratio ζ is positive. For this case the corresponding transient response is known as a damped sinusoid and is shown in Fig. 3.2. Its final value is zero. The angle η shown in Fig. 4.3 is measured from the negative real axis and is related to the damping ratio by

$$\cos\,\eta = \zeta$$

If the complex pole has a positive real part, the time response increases exponentially with time and the system in unstable. If the complex roots are in the right half of the s plane, the damping ratio ζ is negative. The angle η for this case is measured from the positive real axis and is given by

$$\cos\,\eta = |\zeta|$$

For the case of

$$F(s) = \frac{10}{(s^2+6s+25)(s+2)} = \frac{A_1}{s+3-j4} + \frac{A_2}{s+3+j4} + \frac{A_3}{s+2} \tag{4.54}$$

the constants are

$$\begin{aligned} A_1 &= \left[(s+3-j4)\frac{10}{(s^2+6s+25)(s+2)}\right]_{s=-3+j4} \\ &= \left[\frac{10}{(s+3+j4)(s+2)}\right]_{s=-3+j4} = 0.303\angle{-194^\circ}\,. \\ A_3 &= \left[(s+2)\frac{10}{(s^2+6s+25)(s+2)}\right]_{s=-2} = \left(\frac{10}{s^2+6s+25}\right)_{s=-2} = 0.59 \end{aligned}$$

Using Eq. (4.53), the solution is

$$f(t) = 0.606e^{-3t}\sin(4t - 104^\circ) + 0.59e^{-2t} \tag{4.55}$$

The function $F(s)$ of Eq. (4.54) appears in Appendix A as transform pair 29. With the notation of Appendix A, the phase angle in the damped sinusoidal term is

$$\phi = \tan^{-1}\frac{b}{c-a} = \tan^{-1}\frac{4}{2-3} = \tan^{-1}\frac{4}{-1} = 104^\circ \tag{4.56}$$

It is important to note that $\tan^{-1}[4/(-1)] \neq \tan^{-1}(-4/1)$. To get the correct value for the angle ϕ, it is useful to draw a sketch in order to avoid ambiguity and ensure that ϕ is evaluated correctly.

IMAGINARY POLES. The position of imaginary poles of $F(s)$ in the s plane is shown in Fig. 4.4. As the real part of the poles is zero, the poles lie on the imaginary axis. This situation is a special case of complex poles, i.e., the damping ratio $\zeta = 0$. For the poles shown in Fig. 4.4 the transform $F(s)$ and its partial fractions are

$$F(s) = \frac{P(s)}{Q(s)} = \frac{P(s)}{(s^2+\omega_n^2)(s-s_3)} = \frac{A_1}{s-s_1} + \frac{A_2}{s-s_2} + \frac{A_3}{s-s_3} \tag{4.57}$$

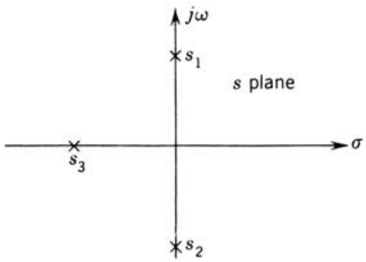


FIGURE 4.4 Poles of $F(s)$ containing imaginary conjugate poles in the s plane.

The quadratic can be factored in terms of the poles s_1 and s_2; thus,

$$s^2 + \omega_n^2 = (s - j\omega_n)(s + j\omega_n) = (s - s_1)(s - s_2)$$

The inverse transform of Eq. (4.57) is

$$f(t) = A_1 e^{j\omega_n t} + A_2 e^{-j\omega_n t} + A_3 e^{s_3 t} \tag{4.58}$$

As $f(t)$ is a real quantity, the coefficients A_1 and A_2 are complex conjugates. The terms in Eq. (4.58) can be combined into the more useful form

$$f(t) = 2|A_1| \sin(\omega_n t + \phi) + A_3 e^{s_3 t} \tag{4.59}$$

Note that since there is no damping term multiplying the sinusoid, this term represents a steady-state value. The angle ϕ is given by

$$\phi = \text{angle of } A_1 + 90^\circ$$

The values of A_1 and A_3, found in the conventional manner, are

$$A_1 = [(s - s_1)F(s)]_{s=s_1} \qquad A_3 = [(s - s_3)F(s)]_{s=s_3}$$

For the case where

$$F(s) = \frac{100}{(s^2 + 25)(s + 2)} = \frac{A_1}{s - j5} + \frac{A_2}{s + j5} + \frac{A_3}{s + 2} \tag{4.60}$$

the values of the coefficients are

$$A_1 = [(s - j5)F(s)]_{s=j5} = \left[\frac{100}{(s + j5)(s + 2)}\right]_{s=j5} = 1.86\angle{-158.2^\circ}\,.$$

$$A_3 = [(s + 2)F(s)]_{s=-2} = \left(\frac{100}{s^2 + 25}\right)_{s=-2} = 3.45$$

The solution is

$$f(t) = 3.72 \sin(5t - 68.2^\circ) + 3.45e^{-2t} \tag{4.61}$$

Case 4: Multiple-Order Complex Poles

Multiple-order complex-conjugate poles are rare. They can be treated in much the same fashion as repeated real poles.

4.9 MATLAB PARTIAL-FRACTION EXAMPLE

The MATLAB *RESIDUE* command is used to obtain the partial-fraction expansion of Eq. (4.54).

MATLAB Partial-Fraction m-file

```
%Set up transfer function
%
%   Define numerator constant of F(s)
k=10;
%   Define poles of F(s)
p(1)=-3+j*4; p(2)=-3-j*4; p(3)=-2;
%Define numerator and denominator
num=k;
den=poly(p)
den=
1    8    37   50
%
%   Use the RESIDUE command to call for residues:
[r1, p1, k1=residue (num, den);
%   List the residues (which are complex)
r1
r1=
-0.2941+0.0735i
-0.2941-0.0735i
0.5882
%   List the poles to check data entry
p1
p1=
-3.0000+4.0000i
-3.0000-4.0000i
-2.0000
%   List the angle of the residues in degrees,
angle (r1)*180/pi
ans=
165.9638
-165.9638
0
%   List the magnitudes of the residues
abs (r1)
```

```
ans =
0.3032
0.3032
0.5882
%
%  To show numerical stability, form the transfer function from
%  the residues (CADAC)(accuracy check)
%  CADAC (Accuracy check)
[num1, den1] = residue (r1, p1, k1);
printsys (num1, den1)
num/den =

1.1102e - 016s^2 + 4.4409e - 016s + 10
----------------------------------
        s^3 + 8s^2 + 37s + 50

%
%   Note the false roots in the numerator
%   Form the Poles
roots (den1)
ans =
-3.0000 + 4.0000i
-3.0000 - 4.0000i
-2.0000
%
%   Plot a three second response to an impulse
%    Define a time vector [initial value: incremental value:
%    final value]
time = [0: .02: 3];
%    Use impulse response command
impulse (num, den, time)
%    Print the plot to compare it with the plot in Fig. 4.5
print -dtiff Sec.4.9. tif
```

The PARTFRAC Command

The m-file name *partfrac.m* calculates time responses and is included on the accompanying CDROM.

```
Partfrac (num, den)
```

MATLAB produces the following output, which corresponds to Eq. (4.55).

```
The partial fraction expansion of
Num/den =

        10
-------------------
 s^3 + 8s^2 + 37s + 50
```

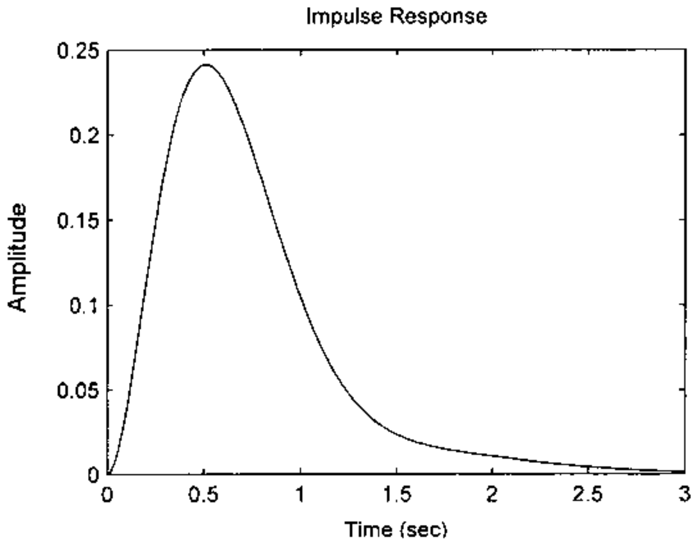


FIGURE 4.5 Response of Eq. (4.55) from MATLAB.

```
Representing the time response for a unit impulse input is
y(t) =

0.60634 exp( − 3t) cos(4t + 165.9638deg) + 0.58824 exp( − 2t)
```

4.10 PARTIAL-FRACTION SHORTCUTS

Some shortcuts can be used to simplify the partial-fraction expansion procedures. They are most useful for transform functions that have multiple poles or complex poles. With multiple poles, the evaluation of the constants by the process of repeated differentiation given by Eq. (4.48) can be very tedious, particularly when a factor containing s appears in the numerator of $F(s)$. The modified procedure is to evaluate only those coefficients for real poles that can be obtained readily without differentiation. The corresponding partial functions are then subtracted from the original function. The resultant function is easier to work with to get the additional partial fractions than the original function $F(s)$.

Example 1. For the function $F(s)$ of Eq. (4.49), which contains a pole $s_1 = -2$ of multiplicity 3, the coefficient $A_{13} = 1$ is readily found. Subtracting the associated fraction from $F(s)$ and putting the result over a common denominator yields.

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

Note that the remainder $F_x(s)$ contains the pole $s_1 = -2$ with a multiplicity of only 2. Thus the coefficient A_{12} is now easily obtained from $F_x(s)$ without differentiation. The procedure is repeated to obtain all the coefficients associated with the multiple pole.

Example 2. Consider the function

$$F(s) = \frac{20}{(s^2+6s+25)(s+1)} = \frac{As+B}{s^2+6s+25} + \frac{C}{s+1} \tag{4.62}$$

which has one real pole and a pair of complex poles. Instead of partial fractions for each pole, leave the quadratic factor in one fraction. Note that the numerator of the fraction containing the quadratic is a polynomial in s and is one degree lower than the denominator. Residue C is $C = [(s+1)F(s)]_{s=-1} = 1$. Subtracting this fraction from $F(s)$ gives

$$\frac{20}{(s^2+6s+25)(s+1)} - \frac{1}{s+1} = \frac{-s^2-6s-5}{(s^2+6s+25)(s+1)}$$

Since the pole at $s = -1$ has been removed, the numerator of the remainder must be divisible by $s+1$. The simplified function must be equal to the remaining partial fraction; i.e.,

$$-\frac{s+5}{s^2+6s+25} = \frac{As+B}{s^2+6s+25} = \frac{-(s+5)}{(s+3)^2+4^2}$$

The inverse transform for this fraction can be obtained directly from Appendix A, pair 26. The complete inverse transform of $F(s)$ is

$$f(t) = e^{-t} - 1.12e^{-3t}\sin(4t + 63.4^\circ) \tag{4.63}$$

Leaving the complete quadratic in one fraction has resulted in real constants for the numerator. The chances for error are probably less than evaluating the residues, which are complex numbers, at the complex poles.

A rule presented by Hazony and Riley [6] is very useful for evaluating the coefficients of partial-fraction expansions. For a normalized‡ ratio of polynomials the rule is expressed as follows:

1. If the denominator is 1 degree higher than the numerator, the sum of the residues is 1.

‡A normalized polynomial is a monic polynomial, that is, the coefficient of the term containing the highest degree is equal to unity.

2. If the denominator is 2 or more degrees higher than the numerator, the sum of the residues is 0.

Equation (4.32) with a_w factored from the numerator is a normalized ratio of polynomials. These rules are applied to the ratio $F(s)/a_w$. It should be noted that *residue* refers only to the coefficients of terms in a partial-fraction expansion with first-degree denominators. Coefficients of terms with higher-degree denominators are referred to only as coefficients. These rules can be used to simplify the work involved in evaluating the coefficients of partial-fraction expansions, particularly when the original function has a multiple-order pole. For example, only A_{11} and A_2 in Eq. (4.49) are residues, and therefore $A_{11} + A_2 = 0$. Since $A_2 = -1$, the value of $A_{11} = 1$ is obtained directly.

Digital-computer programs (see Appendixes C and D) are readily available for evaluating the partial-fraction coefficients and obtaining a tabulation and plot of $f(t)$ [4].

4.11 GRAPHICAL INTERPRETATION OF PARTIAL-FRACTION COEFFICIENTS [7]

The preceding sections describe the analytical evaluation of the partial-function coefficients. These constants are directly related to the pole-zero pattern of the function $F(s)$ and can be determined graphically, whether the poles and zeros are real or in complex-conjugate pairs. As long as $P(s)$ and $Q(s)$ are in factored form, the coefficients can be determined graphically by inspection. Rewriting Eq. (4.32) with the numerator and denominator in factored form and with $a_w = K$ gives

$$F(s) = \frac{P(s)}{Q(s)} = \frac{K(s - z_1)(s - z_2)\cdots(s - z_m)\cdots(s - z_w)}{(s - p_1)(s - p_2)\cdots(s - p_k)\cdots(s - p_n)}$$

$$= K\frac{\prod_{m=1}^{w}(s - z_m)}{\prod_{k=1}^{n}(s - p_k)} \tag{4.64}$$

The zeros of this function, $s = z_m$, are those values of s for which the function is zero; that is, $F(z_m) = 0$. Zeros are indicated by a small circle on the s plane. The poles of this function, $s = p_k$, are those values of s for which the function is infinite; that is, $|F(p_k)| = \infty$. Poles are indicated by a small $\times$ on the s plane. Poles are also known as singularities[§] of the function. When $F(s)$ has only simple poles (first-order poles), it can be expanded into partial fractions of

[§]A singularity of a function $F(s)$ is a point where $F(s)$ does not have a derivative.

the form

$$F(s) = \frac{A_1}{s - p_1} + \frac{A_2}{s - p_2} + \cdots + \frac{A_k}{s - p_k} + \cdots + \frac{A_n}{s - p_n} \tag{4.65}$$

The coefficients are obtained from Eq. (4.38) and are given by

$$A_k = [(s - p_k)F(s)]_{s=p_k} \tag{4.66}$$

The first coefficient is

$$A_1 = \frac{K(p_1 - z_1)(p_1 - z_2)\cdots(p_1 - z_w)}{(p_1 - p_2)(p_1 - p_3)\cdots(p_1 - p_n)} \tag{4.67}$$

Figure 4.6 shows the poles and zeros of a function $F(s)$. The quantity $s = p_1$ is drawn as an arrow from the origin of the s plane to the pole p_1. This arrow is called a *directed line segment*. It has a magnitude equal to $|p_1|$ and an angle of 180°. Similarly, the directed line segment $s = z_1$ is drawn as an arrow from the origin to the zero z_1 and has a corresponding magnitude and angle. A directed line segment is also drawn from the zero z_1 to the pole p_1. By the rules of vector addition, this directed line segment is equal to $p_1 - z_1$; in polar form it has a magnitude equal to its length and an angle ψ, as shown. By referring to Eq. (4.67), it is seen that $p_1 - z_1$ appears in the numerator of A_1. Whereas this quantity can be evaluated analytically, it can also be measured from the pole-zero diagram. In a similar fashion, each factor in Eq. (4.67) can be obtained graphically; the directed line segments are drawn *from* the zeros and each of the other poles *to* the pole p_1. The angle from a zero is indicated by the symbol ψ, and the angle from a pole is indicated by the symbol θ.

The general rule for evaluating the coefficients in the partial-fraction expansion is quite simple. The value of A_k is the product of K and the directed distances from each zero to the pole p_k divided by the product of the directed distances from each of the other poles to the pole p_k. Each of these

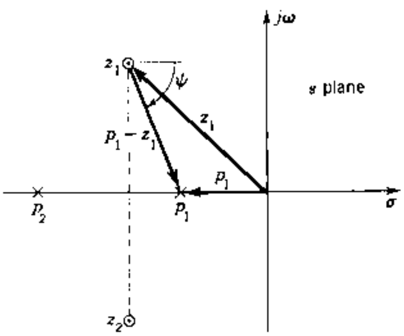


FIGURE 4.6 Directed line segments drawn on a pole-zero diagram of a function $F(s)$.

directed distances is characterized by a magnitude and an angle. This statement can be written in equation form as

$$A_k = K\frac{\text{product of directed distances from each zero to pole } p_k}{\text{product of directed distances from all other poles to pole } p_k}$$

$$K = \frac{\prod_{m=1}^{w} |p_k - z_m| \;\angle \sum_{m=1}^{w} \psi(p_k - z_m)}{\prod_{c=1, c\neq k}^{n} |p_k - p_c| \;\angle \sum_{c=1}^{n} \theta(p_k - p_c)} \tag{4.68}$$

Equations (4.67) and (4.68) readily reveals that as the pole p_k comes closer to the zero z_{m}, the corresponding coefficient A_k becomes smaller (approaches a zero value). This characteristic is very important in analyzing system response characteristics. For real poles the values of A_{k} must be real but can be either positive or negative. A_k is positive if the total number of real poles and real zeros to the right of p_k is even, and it is negative if the total number of real poles and real zeros to the right of p_k is odd. For complex poles the values of A_k are complex. If $F(s)$ has no finite zeros, the numerator of Eq. (4.68) is equal to K. A repeated factor of the form $(s + a)^r$ in either the numerator or the denominator is included r times in Eq. (4.68), where the values of $|s + a|$ and $\angle s + a$ are obtained graphically. The application of the graphical technique is illustrated by the following example:

Example. A function $F(s)$ with complex poles (step function response):

$$F(s) = \frac{K(s+\alpha)}{s\left[(s+a)^2+b^2\right]} = \frac{K(s+\alpha)}{s(s+a-jb)(s+a+jb)}$$

$$= \frac{K(s-z_1)}{s(s-p_1)(s-p_2)} = \frac{A_0}{s} + \frac{A_1}{s+a-jb} + \frac{A_2}{s+a+jb} \tag{4.69}$$

The coefficient A_0, for the pole $p_0 = 0$, is obtained by use of the directed line segments shown in Fig. 4.7*a*:

$$A_0 = \frac{K(\alpha)}{(a-jb)(a+jb)} = \frac{K\alpha}{\sqrt{a^2+b^2}\,e^{j\theta_1}\sqrt{a^2+b^2}\,e^{j\theta_2}} = \frac{K\alpha}{a^2+b^2} \tag{4.70}$$

Note that the angles θ_1 and θ_2 are equal in magnitude but opposite in sign. The coefficient for a real pole is therefore always a real number.

The coefficient A_1, for the pole $p_1 = -a + jb$, is obtained by use of the directed line segments shown in Fig. 4.7*b*:

$$A_1 = \frac{K[(\alpha - a) + jb]}{(-a+jb)(j2b)} = \frac{K\sqrt{(\alpha-a)^2+b^2}\;e^{j\psi}}{\sqrt{a^2+b^2}\,e^{j\theta}2be^{j\pi/2}}$$

$$= \frac{K}{2b}\sqrt{\frac{(\alpha-a)^2+b^2}{a^2+b^2}}\;e^{j(\psi-\theta-\pi/2)} \tag{4.71}$$

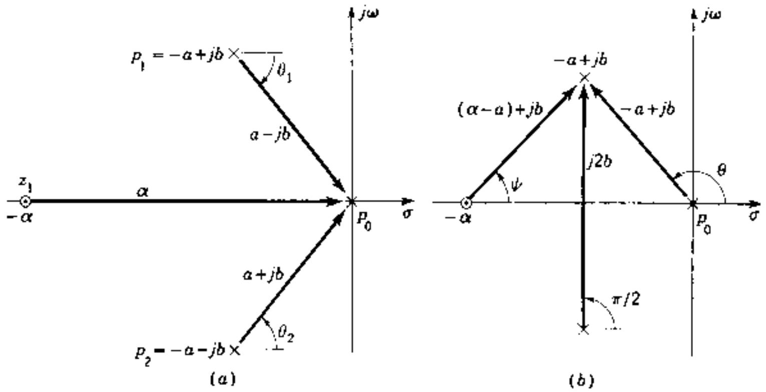


FIGURE 4.7 Directed line segments for evaluating (*a*) A_0 of Eq. (4.70) and (*b*) A_1 of Eq. (4.71).

where

$$\psi = \tan^{-1} \frac{b}{\alpha - a} \qquad \text{and} \qquad \theta = \tan^{-1} \frac{b}{-a}$$

Since the constants A_1 and A_2 are complex conjugates, it is not necessary to evaluate A_2. It should be noted that zeros in $F(s)$ affect both the magnitude and the angle of the coefficients in the time function.

The response as a function of time is therefore

$$f(t) = A_0 + 2|A_1| e^{-at} \sin\left(bt + \angle A_1 + 90^\circ\right)$$

$$= \frac{K\alpha}{a^2 + b^2} + \frac{K}{b} \sqrt{\frac{(\alpha - a)^2 + b^2}{a^2 + b^2}}\, e^{-at} \sin(bt + \phi) \tag{4.72}$$

where

$$\phi = \psi - \theta = \tan^{-1} \frac{b}{\alpha - a} - \tan^{-1} \frac{b}{-a} = \text{angle of } A_1 + 90^\circ \tag{4.73}$$

This agrees with transform pair 28 in Appendix A. A_0 and A_1 can be evaluated by measuring the directed line segments directly on the pole-zero diagram. It is interesting to note that, for a given ζ, as the magnitude $\omega_n = \sqrt{a^2 + b^2}$ becomes larger, the coefficients A_0 and A_1 approach zero as revealed by an analysis of Eqs. (4.70) and (4.71), respectively. (The limiting value is $\lim_{\omega_n \to \infty}[A_1] = K/2b$.) This is an important characteristic when it is desired that $f(t)$ be very small. The application of this characteristic is used in disturbance rejection (Chap. 12). Also see Prob. 4.17 for further illustration.

4.12 FREQUENCY RESPONSE FROM THE POLE-ZERO DIAGRAM

The frequency response of a system is described as the steady-state response with a sine-wave forcing function for all values of frequency. This information is often presented graphically, using two curves. One curve shows the ratio of output amplitude to input amplitude M and the other curve shows the phase angle of the output α, where both are plotted as a function of frequency, often on a logarithmic scale.

Consider the input to a system as sinusoidal and given by

$$x_1(t) = X_1 \sin \omega t \tag{4.74}$$

and the output is

$$x_2(t) = X_2 \sin(\omega t + \alpha) \tag{4.75}$$

The magnitude of the frequency response for the function given in Eq. (4.64) is

$$M = \left|\frac{\mathrm{X}_2(j\omega)}{\mathrm{X}_1(j\omega)}\right| = \left|\frac{\mathrm{P}(j\omega)}{\mathrm{Q}(j\omega)}\right| = \left|\frac{K(j\omega - z_1)(j\omega - z_2)\cdots}{(j\omega - p_1)(j\omega - p_2)\cdots}\right| \tag{4.76}$$

and the angle is

$$\begin{aligned} \alpha &= \angle \mathrm{P}(j\omega) - \angle Q(j\omega) \\ &= \angle K + \angle j\omega - z_1 + \angle j\omega - z_2 + \cdots - \angle j\omega - p_1 - \angle j\omega - p_2 - \cdots \end{aligned} \tag{4.77}$$

Figure 4.8 shows a pole-zero diagram and the directed line segments for evaluating M and α corresponding to a frequency ω_1. As ω increases, each of the directed line segments changes in both magnitude and angle. Several characteristics can be noted by observing the change in the magnitude and angle of each directed line segment as the frequency ω increases from 0 to ∞. The magnitude and angle of the frequency response can be obtained graphically from the pole-zero diagram. The magnitude M and the angle α are a composite of all the effects of all the poles and zeros. In particular, for a system function that has all poles and zeros in the left half of the s plane, the following characteristics of the frequency response are noted:

1. At $\omega = 0$ the magnitude is a finite value, and the angle is 0°.
2. If there are more poles than zeros, then as $\omega \to \infty$, the magnitude of M approaches zero and the angle is -90° times the difference between the number of poles and zeros.

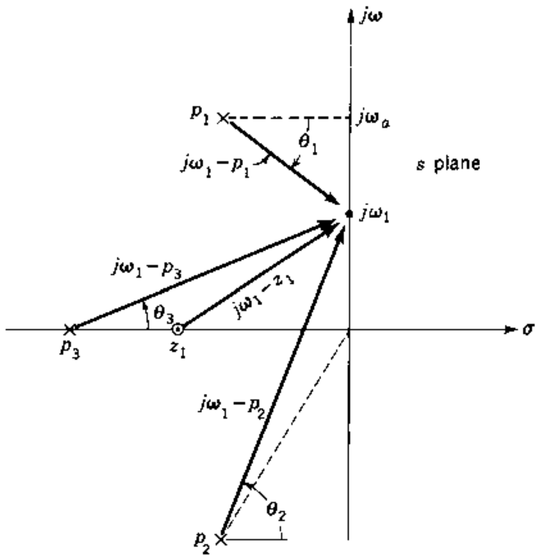


FIGURE 4.8 Pole-zero diagram of Eq. (4.76) showing the directed line segments for evaluating the frequency response $M\angle\alpha$.

3. The magnitude can have a peak value M_m only if there are complex poles fairly close to the imaginary axis. It is shown later that the damping ratio ζ of the complex poles must be less than 0.707 to obtain a peak value for M. Of course, the presence of zeros could counteract the effect of the complex poles, and so it is possible that no peak would be present even if ζ of the poles were less than 0.707.

Typical frequency-response characteristics are illustrated for two common functions. As a first example,

$$\frac{X_2(s)}{X_1(s)} = \frac{P(s)}{Q(s)} = \frac{a}{s+a} \tag{4.78}$$

The frequency response shown in Fig. 4.9*b* and *c* starts with a magnitude of unity and an angle of 0°. As ω increases, the magnitude decreases and approaches zero while the angle approaches $-90°$. At $\omega = a$ the magnitude is 0.707 and the angle is $-45°$. The frequency $\omega = a$ is called the *corner frequency* ω_{cf} or the break frequency. Figure 4.9 shows the pole location, the frequency-response magnitude M, and the frequency-response angle α as a function of ω.

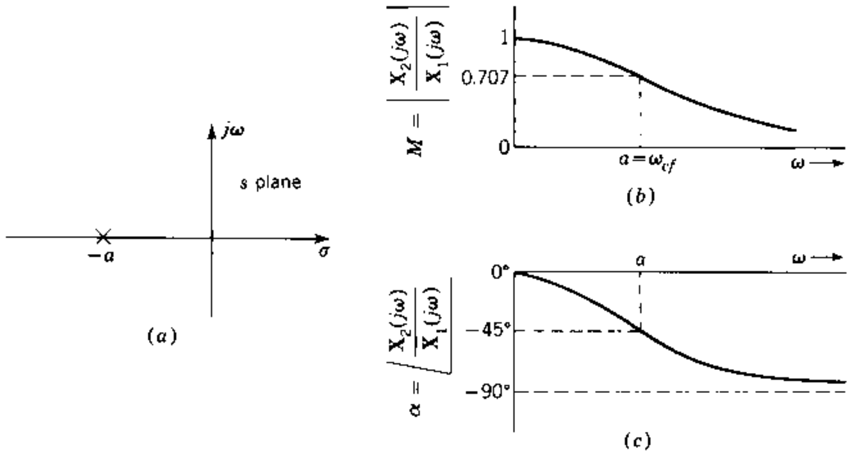


FIGURE 4.9 (*a*) Pole location of Eq. (4.78); (b) *M* vs. ω; (c) α vs. ω.

The second example is a system transform with a pair of conjugate poles:

$$\frac{X_2(s)}{X_1(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} = \frac{\omega_n^2}{(s - p_1)(s - p_2)}$$

$$= \frac{\omega_n^2}{\left(s + \zeta\omega_n - j\omega_n\sqrt{1-\zeta^2}\right)\left(s + \zeta\omega_n + j\omega_n\sqrt{1-\zeta^2}\right)} \tag{4.79}$$

The location of the poles of Eq. (4.79) and the directed line segments for $s = j\omega$ are shown in Fig. 4.10*a*. At $\omega = 0$ the values are $M = 1$ and $\alpha = 0^\circ$. As ω increases, the magnitude M first increases, because $j\omega - p_1$ is decreasing faster than $j\omega - p_2$ is increasing. By differentiating the magnitude of $|\mathbf{M}(j\omega)|$, obtained from Eq. (4.79), with respect to ω, it can be shown (see Sec. 9.3) that the maximum value M_m and the frequency ω_m at which it occurs are given by

$$M_m = \frac{1}{2\zeta\sqrt{1-\zeta^2}} \tag{4.80}$$

$$\omega_m = \omega_n\sqrt{1 - 2\zeta^2} \tag{4.81}$$

A circle drawn on the *s* plane in Fig. 4.10*a*, with the poles p_1 and p_2 as the end-points of the diameter, intersects the imaginary axis at the value ω_m given by Eq. (4.81). Both Eq. (4.81) and the geometrical construction of the circle show that the *M* curve has a maximum value, other than at $\omega = 0$, only if $\zeta < 0.707$. The angle α becomes more negative as ω increases. At $\omega = \omega_n$ the angle is equal to -90°. This is the corner frequency for a quadratic factor

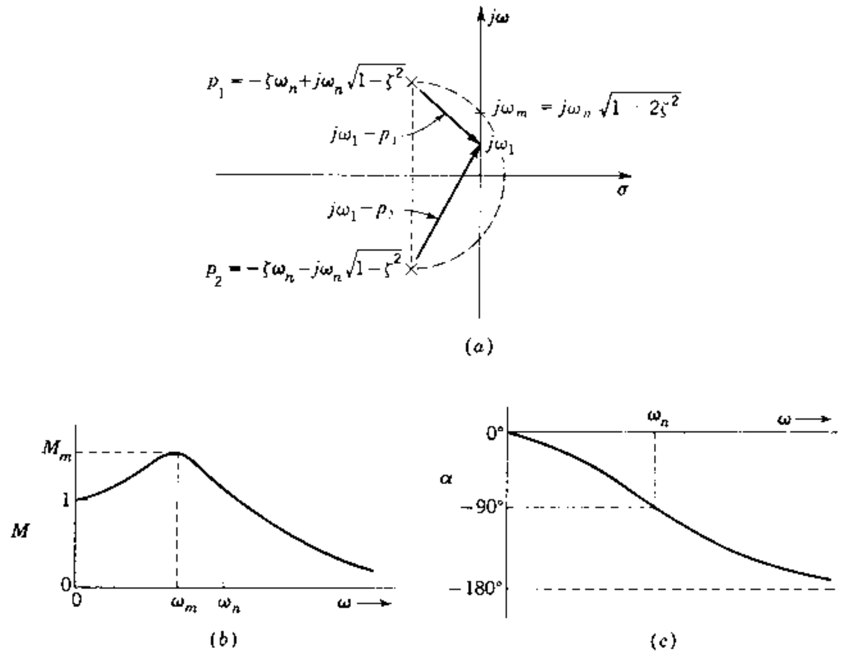


FIGURE 4.10 (*a*) Pole-zero diagram, < 0.707; (*b*) magnitude of frequency response; (*c*) angle of frequency response.

with complex roots. As $\omega \to \infty$, the angle α approaches -180°. The magnitude and angle of the frequency response for $\zeta < 0.707$ are shown in Fig. 4.10. The M and α curves are continuous for all linear systems. The M vs. ω plot of Fig. 4.10*b* describes the frequency response characteristics for the undamped simple second-order system of Eq. (4.79). In later chapters this characteristic is used as the standard reference for analyzing and synthesizing higher-order systems.

4.13 LOCATION OF POLES AND STABILITY

The stability and the corresponding response of a system can be determined from the location of the poles of the response transform $F(s)$ in the s plane. The possible positions of the poles are shown in Fig. 4.11, and the responses are given in Table 4.1. These poles are the roots of the characteristic equation.

Poles of the response transform at the origin or on the imaginary axis that are not contributed by the forcing function result in a continuous output. These outputs are undesirable in a control system. Poles in the

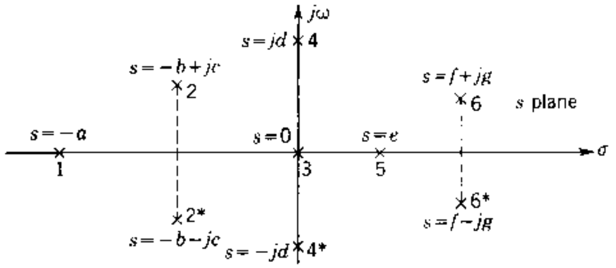


FIGURE 4.11 Location of poles in the s plane. (Numbers are used to identify the poles.)

right-half s plane result in transient terms that increase with time. Such performance characterizes an unstable system; therefore, poles in the right-half s plane are not desirable, in general.

4.14 LAPLACE TRANSFORM OF THE IMPULSE FUNCTION

Figure 4.12 shows a rectangular pulse of amplitude $1/a$ and of duration a. The analytical expression for this pulse is

$$f(t) = \frac{u_{-1}(t) - u_{-1}(t-a)}{a} \tag{4.82}$$

and its Laplace transform is

$$F(s) = \frac{1 - e^{-as}}{as} \tag{4.83}$$

where

$$e^{-as} = 1 - as - \frac{(as)^2}{2!} - \frac{(as)^3}{3!} - \cdots \tag{4.84}$$

Table 4.1 Relation of Response to Location of Poles

Position of pole	Form of response	Characteristics
1	Ae^{-at}	Damped exponential
2–2*	$Ae^{-bt}\sin(ct + \phi)$	Exponentially damped sinusoid
3	A:	Constant
4–4*	$A\sin(dt + \phi)$	Constant-sinusoid
5	Ae^{et}	Increasing exponential (unstable)
6–6*	$Ae^{ft}\sin(gt + \phi)$	Exponentially increasing sinusoid (unstable)

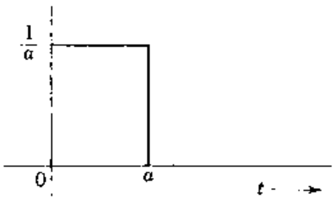


FIGURE 4.12 Rectangular pulse of unit area.

and

$$\lim_{a\to 0}\left(\frac{1-e^{-as}}{as}\right)=\lim_{a\to 0}\left(1+\frac{as}{2!}+\frac{(as)^2}{3!}+\cdots\right)=1 \tag{4.85}$$

If a is decreased, the amplitude increases and the duration of the pulse decreases but the area under the pulse, and thus its strength, remains unity. The limit of $f(t)$ as $a \to 0$ is termed a unit impulse and is designated by $\delta(t) = u_0(t)$. The Laplace transform of the unit impulse, as evaluated by use of L'Hospital's theorem, is $\Delta(s) = 1$.

When any function has a jump discontinuity, the derivative of the function has an impulse at the discontinuity. Some systems are subjected to shock inputs. For example, an airplane in flight may be jolted by a gust of wind of short duration. When a gun is fired, it has a reaction force of large magnitude and very short duration, and so does a steel ball bouncing off a steel plate. When the duration of a disturbance or input to a system is very short compared with the natural periods of the system, the response can often be well approximated by considering the input to be an impulse of proper strength. An impulse of infinite magnitude and zero duration does not occur in nature; but if the pulse duration is much smaller than the time constants of the system, a representation of the input by an impulse is a good approximation. The shape of the impulse is unimportant as long as the strength of the equivalent impulse is equal to the strength of the actual pulse.

Since the Laplace transform of an impulse is defined [8], the approximate response of a system to a pulse is often found more easily by this method than by the classical method. Consider the *RLC* circuit shows in Fig. 4.13 when the input voltage is an impulse. The problem is to find the voltage e_c across the capacitor as a function of time. The differential equation relating e_c to the input voltage e and the impedances is written by the node method:

$$\left(CD+\frac{1}{R+LD}\right)e_c-\frac{1}{R+LD}e=0 \tag{4.86}$$

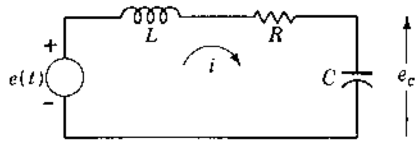


FIGURE 4.13 *RLC* series circuit.

Rationalizing the equation yields

$$(LCD^2 + RCD + 1)e_c = e \tag{4.87}$$

Taking the Laplace transform of this equation gives

$$(s^2LC + sRC + 1)E_c(s) - sLCe_c(0) - LCDe_c(0) - RCe_c(0) = E(s) \tag{4.88}$$

Consider an initial condition $e_c(0^-) = 0$. The voltage across the capacitor is related to the current in the series circuit by the expression $De_c = i/C$. Since the current cannot change instantaneously through an inductor, $De_c(0^-) = 0$. These initial conditions that exist at $t = 0^-$ are used in the Laplace-transformed equation when the forcing function is an impulse. The solution for the current shows a discontinuous step change at $t = 0$.

The input voltage e is an impulse; therefore, $E(s) = 1$. Solving for $E_c(s)$ gives

$$E_c(s) = \frac{1/LC}{s^2 + (R/L)s + 1/LC} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{4.89}$$

Depending on the relative sizes of the parameters, the circuit may be overdamped or underdamped. In the overdamped case the voltage $e_c(t)$ rises to a peak value and then decays to zero. In the underdamped case the voltage oscillates with a frequency $\omega_d = \omega_n\sqrt{1 - \zeta^2}$ and eventually decays to zero. The form of the voltage $e_c(t)$ for several values of damping is shown in Sec. 4.15.

The stability of a system is revealed by the impulse response. With an impulse input the response transform contains only poles and zeros contributed by the system parameters. The impulse response is therefore the inverse transform of the system transfer function and is called the *weighting function*:

$$g(t) = \mathscr{L}^{-1}[G(s)] \tag{4.90}$$

The impulse response provides one method of evaluating the system function; although the method is not simple, it is possible to take the impulse response, which is a function of time, and convert it into the transfer function $G(s)$, which is a function of s.

The system response $g(t)$ to an impulse input $\delta(t)$ can be used to determine the response $c(t)$ to any input $r(t)$ [8]. The response $c(t)$ is determined by the use of the convolution or superposition integral

$$c(t) = \int_0^t r(\tau)g(t-\tau)\,d\tau \tag{4.91}$$

This method is normally used for inputs that are not sinusoidal or a power series.

4.15 SECOND-ORDER SYSTEM WITH IMPULSE EXCITATION

The response of a second-order system to a step-function input is shown in Fig. 3.6. The impulse function is the derivative of the step function; therefore, the response to an impulse is the derivative of the response to a step function. Figure 4.14 shows the response for several values of damping ratio ζ for the second-order function of Eq. (4.89).

The first zero of the impulse response occurs at t_p, which is the time at which the maximum value of the step-function response occurs.

For the underdamped case the impulse response is

$$f(t) = \frac{\omega_n}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t} \sin \omega_n \sqrt{1-\zeta^2}\; t \tag{4.92}$$

By setting the derivative of $f(t)$ with respect to time equal to zero, the maximum overshoot can be shown to occur at

$$t_m = \frac{\cos^{-1}\zeta}{\omega_n\sqrt{1-\zeta^2}} \tag{4.93}$$

The maximum value of $f(t)$ is

$$f(t_m) = \omega_n \exp\left(-\frac{\zeta\cos^{-1}\zeta}{\sqrt{1-\zeta^2}}\right) \tag{4.94}$$

The impulse response $g(t)$ can be measured after applying an input pulse that has a short duration compared with the largest system time constant. Then it can be used in a deconvolution technique to determine the transfer function $G(s)$. It can also be used to determine the disturbance-rejection characteristics of control systems.

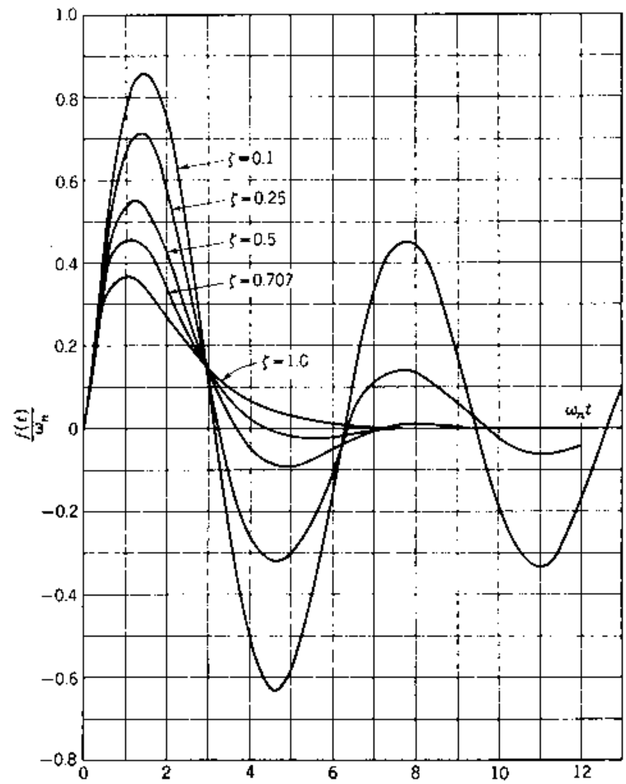


FIGURE 4.14 Response to an impulse for $F(s) = \omega_n^2/(s^2 + 2\omega_n s + \omega_n^2)$.

4.16 SOLUTION OF STATE EQUATION [9,10]

The homogeneous state equation is

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \tag{4.95}$$

The solution of this equation can be obtained by taking the Laplace transform

$$s\mathbf{X}(s) - \mathbf{x}(0) = \mathbf{A}\mathbf{X}(s) \tag{4.96}$$

Grouping of the terms containing $\mathbf{X}(s)$ yields

$$[s\mathbf{I} - \mathbf{A}]\mathbf{X}(s) = \mathbf{x}(0) \tag{4.97}$$

The unit matrix $\mathbf{I}$ is introduced so that all terms in the equations are proper matrices. Premultiplying both sides of the equation by $[s\mathbf{I} - \mathbf{A}]^{-1}$ yields

$$\mathbf{X}(s) = [s\mathbf{I} - \mathbf{A}]^{-1}\mathbf{x}(0) \tag{4.98}$$

The inverse Laplace transform of Eq. (4.98) gives the time response

$$\mathbf{x}(t) = \mathscr{L}^{-1}\left\{[s\mathbf{I} - \mathbf{A}]^{-1}\right\}\mathbf{x}(0) \tag{4.99}$$

Comparing this solution with Eq. (3.70) yields the following expression for the state transition matrix:

$$\boldsymbol{\Phi}(t) = \mathscr{L}^{-1}[s\mathbf{I} - \mathbf{A}]^{-1} \tag{4.100}$$

The *resolvent matrix* is designated by $\boldsymbol{\Phi}(s)$ and is defined by*

$$\boldsymbol{\Phi}(s) = [s\mathbf{I} - \mathbf{A}]^{-1} \tag{4.101}$$

Example 1. Solve for $\boldsymbol{\Phi}(t)$ (see the example of Sec. 3.14) for

$$\mathbf{A} = \begin{bmatrix} 0 & 6 \\ -1 & -5 \end{bmatrix}$$

Using Eq. (4.101) gives

$$\boldsymbol{\Phi}(s) = \left[\begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix} - \begin{bmatrix} 0 & 6 \\ -1 & -5 \end{bmatrix}\right]^{-1} = \begin{bmatrix} s & -6 \\ 1 & s+5 \end{bmatrix}^{-1}$$

Using Eq. (4.97) gives

$$\Phi(s) = \frac{1}{s^2 + 5s + 6}\begin{bmatrix} s+5 & 6 \\ -1 & s \end{bmatrix} = \begin{bmatrix} \dfrac{s+5}{(s+2)(s+3)} & \dfrac{6}{(s+2)(s+3)} \\ \dfrac{-1}{(s+2)(s+3)} & \dfrac{s}{(s+2)(s+3)} \end{bmatrix} \tag{4.102}$$

Using transform pairs 12 to 14 in Appendix A yields the inverse transform

$$\Phi(t) = \begin{bmatrix} 3e^{-2t} - 2e^{-3t} & 6e^{-2t} - 6e^{-3t} \\ -e^{-2t} + e^{-3t} & -2e^{-2t} + 3e^{-3t} \end{bmatrix}$$

This is the same as the value obtained in Eq. (3.92) and is an alternate method of obtaining the same result.

The state equation with an input present is

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \tag{4.103}$$

Solving by means of the Laplace transform yields

$$\mathbf{X}(s) = [s\mathbf{I} - \mathbf{A}]^{-1}\mathbf{x}(0) + [s\mathbf{I} - \mathbf{A}]^{-1}\mathbf{B}\mathbf{U}(s) = \boldsymbol{\Phi}(s)\mathbf{x}(0) + \boldsymbol{\Phi}(s)\mathbf{B}\mathbf{U}(s) \tag{4.104}$$

* When $\mathbf{A}$ is a companion matrix, a direct algorithm for evaluating $[s\mathbf{I} - \mathbf{A}]^{-1}$ and $\boldsymbol{\Phi}(t)$ is presented in Ref. 11.

Note that $\Phi(s)x(0)$ is the zero-input response and $\Phi(s)\mathbf{BU}(s)$ is the zero-state response (See Sec. 3.15.)

Example 2. Solve for the state equation in the example of Sec. 3.14 for a step function input. The function $\Phi(s)$ is given by Eq. (4.102) in Example 1 and $U(s) = 1/s$. Inserting these values into Eq. (4.104) gives

$$\mathbf{X}(s) = \Phi(s)\mathbf{x}(0) + \begin{bmatrix} \dfrac{6}{s(s+2)(s+3)} \\ \dfrac{1}{(s+2)(s+3)} \end{bmatrix} \tag{4.105}$$

Applying the transform pairs 10 and 12 of Appendix A yields the same result for $\mathbf{x}(t)$ as in Eq. (3.105). The Laplace transform method is a direct method of solving the state equation.

The possibility exists in the elements of $\Phi(s)$ that a numerator factor of the form $s + a$ will cancel a similar term in the denominator. In that case the transient mode e^{-at} does not appear in the corresponding element of $\Phi(t)$. This has a very important theoretical significance since it affects the *controllability and observability* of a control system. These properties are described in Sec. 13.2. Also, as described in Chap. 7, the ability to design a tracking control system requires that the system be observable and controllable.

4.17 EVALUATION OF THE TRANSFER-FUNCTION MATRIX

The transfer-function matrix of a system with multiple inputs and outputs is evaluated from the state and output equations

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu} \qquad \mathbf{y} = \mathbf{Cx} + \mathbf{Du}$$

Taking the Laplace transform of these equations and solving for the output $\mathbf{Y}(s)$ in terms of the input $\mathbf{U}(s)$ yields

$$\mathbf{Y}(s) = C\Phi(s)\mathbf{x}(0) + C\Phi(s)\mathbf{BU}(s) + \mathbf{DU}(s) \tag{4.106}$$

Since the transfer-function relationship is $\mathbf{Y}(s) = \mathbf{G}(s)\mathbf{U}(s)$ and is defined for zero initial conditions, the system transfer-function matrix is given by analogy as

$$\mathbf{G}(s) = \mathbf{C}\Phi(s)\mathbf{B} + \mathbf{D} \tag{4.107}$$

For multiple-input multiple-output system $\mathbf{G}(s)$ is a matrix in which the elements are the transfer functions between each output and each input of the system.

Example. Determine the transfer functions and draw a block diagram for the two-input two-output system represented by

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1 & 1 \\ 0 & -2 \end{bmatrix}\mathbf{u} \qquad \mathbf{y} = \begin{bmatrix} 0 & -2 \\ 1 & 0 \end{bmatrix}\mathbf{x}$$

The characteristic polynomial is

$$|s\mathbf{I} - \mathbf{A}| = (s+1)(s+2)$$

From Eq. (4.101), the resolvent matrix is

$$\mathbf{\Phi}(s) = \frac{1}{(s+1)(s+2)}\begin{bmatrix} s+3 & 1 \\ -2 & s \end{bmatrix}$$

Then, from Eq. (4.107)

$$\mathbf{G}(s) = \mathbf{C}\mathbf{\Phi}(s)\mathbf{B} = \begin{bmatrix} \dfrac{4}{(s+1)(s+2)} & \dfrac{4}{s+2} \\ \dfrac{s+3}{(s+1)(s+2)} & \dfrac{1}{s+2} \end{bmatrix} = \begin{bmatrix} G_{11}(s) & G_{12}(s) \\ G_{21}(s) & G_{22}(s) \end{bmatrix} \tag{4.108}$$

The block diagram for the system is shown in Fig. 4.15.

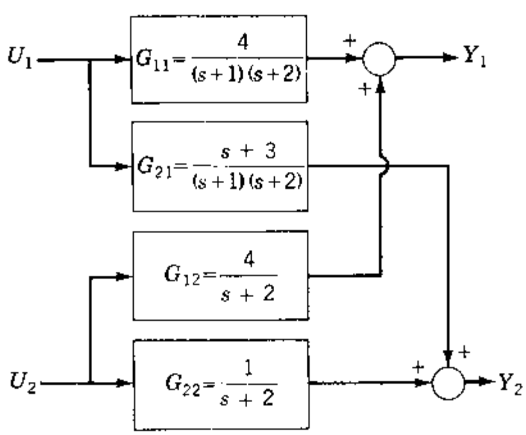


FIGURE 4.15 Block diagram for Eq. (4.108)

4.18 MATLAB m-FILE FOR MIMO SYSTEMS

The following commands illustrate how MIMO systems are manipulated with MATLAB.

```
%
% Define the plant, control input, output, and feedforward
  matrices
% A=[0 1; -2 -3]; B=[1 1; 0 -2]; C=[0 -2; 1 0]; D=zeros(2,2);
%
% Form the state space system model ``statesys''
%
statesys=ss(A, B, C, D)
a=

      x1  x2
 x1    0   1
 x2   -2  -3

b=

      u1  u2
 x1    1   1
 x2    0  -2

c=

      x1  x2
 y1    0  -2
 y2    1   0

d=

      u1  u2
 y1    0   0
 y2    0   0

Continuous-time model.
%
%  Transform from state space model to transfer function model
``tfsys''
%
%
tfsys=tf (statesys)
Transfer function from input 1 to output ...

            4
#1 : ----------
     s^2+3s+2

         s+3
#2 : ----------
     s^2+3s+2
```

www.4electron.com

```
Transfer function from input 2 to output ...

       4s + 4
#1 :----------
    s^2 + 3s + 2

        s + 1
#2 :----------
    s^2 + 3s + 2
%
%  Poles and zeros that cancel are removed with the minreal
%  command
%  Minreal is short for minimum realization
%
tfsys = minreal(tfsys)
Transfer function from input 1 to output ...

           4
#1 :  ----------
      s^2 + 3s + 2

         s + 3
#2 :  ----------
      s^2 + 3s + 2

Transfer function from input 2 to output ...

        4
#1 :  ----
      s + 2

        1
#2 :  ----
      s + 2

%  To get factored form use ''zpk'' which stands for zero, pole,
%  gain
%
factoredsys = zpk(tfsys)
Zero/pole/gain from input 1 to output ...

             4
#1 :  -----------
      (s + 2) (s + 1)

         (s + 3)
#2 :  -----------
      (s + 2) (s + 1)

Zero/pole/gain from input 2 to output ...

         4
#1 :  -----
      (s + 2)

         1
#2 :  -----
      (s + 2)
```

```
%  Individual elements of the model are addressed (output,
%  input)
%
tfsys(2, 1)   % The transfer function to output 2 from input 1
Transfer function:

     s + 3
  ---------
  s^2 + 3s + 2

%
%  Subelements use a pointer notation with {output, input}
%  addressing
%  See 1tiprops for more information
%
tfsys.num{2, 1}    % Numerator of tfsys (2, 1)
ans =
      0   1    3
tfsys.den{2, 1}    % Denominator of tfsys (2, 1)
ans =
      1   3    2
factoredsys.z{2, 1}    %    zeroes of factoredsys (2. 1)
ans =
      -3
factoredsys.p{2, 1}    % Poles of factoredsys (2, 1)
ans =
      -2
      -1
```

4.19 SUMMARY

This chapter discusses the important characteristics and the use of the Laplace transform, which is employed extensively with differential equations because it systematizes their solution. Also it is used extensively in feedback-system synthesis. The pole-zero pattern has been introduced to represent a system function. The pole-zero pattern is significant because it determines the amplitudes of all the time-response terms. The frequency response has also been shown to be a function of the pole-zero pattern. The solution of linear time-invariant (LTI) state equations can be obtained by means of the Laplace transform. The procedure is readily adapted for the use of a digital computer [4], which is advantageous for multiple-input multiple-output (MIMO) systems.

Later chapters cover feedback-system analysis and synthesis by three methods. The first of these is the root-locus method, which locates the poles and zeros of the system in the *s* plane. Knowing the poles and zeros permits an exact determination of the time response. The second method is based on the frequency response. Since the frequency response is a function of the pole-zero pattern, the two methods are complementary and give equivalent information in different forms. The root-locus and frequency-response methods are primarily applicable to SISO systems and rely on use of transfer functions to represent the system. They can be adapted for the synthesis of MIMO systems. The state-feedback method is readily used for both SISO and MIMO systems. The necessary linear algebra operations for state-feedback synthesis are presented in App. B and in this chapter in preparation for their use in later chapters.

System stability requires that all roots of the characteristic equation be located in the left half of the s plane. They can be identified on the pole-zero diagram and are the poles of the overall transfer function. The transfer function can be obtained from the overall differential equation relating an input to an output. The transfer function can also be obtained from the state-equation formulation, as shown in this chapter.

REFERENCES

1. Churchill, R.V.: *Operational Mathematics*, 3rd ed., McGraw-Hill, New York, 1972.
2. Thomson, W.T.: *Laplace Transformation*, 2nd ed., Prentice-Hall, Englewood Cliffs, N.J., 1996.
3. Kreyszig, E.: *Advanced Engineering Mathematics*, 7th ed., John Wiley & Sons, New York 1993.
4. *TOTAL-PC* by Dr. R. Ewing, Graduate School of Engineering. Air Force Institute of Technology, AFIT/ENG, Bldg. 642, 2950 P St., Wright-Patterson Air Force Base, Ohio 45433-7765.
5. Bongiorno, J.J., Jr.: "A Recursive Algorithm for Computing the Partial Fraction Expansion of Rational Functions Having Multiple Poles," *IEEE Trans. Autom. Control*, vol. Ac-29, pp. 650–652, 1984.
6. Hazony, D., and I. Riley: "Simplified Technique for Evaluating Residues in the Presence of High Order Poles,: paper presented at the Western Electric Show and Convention, August 1959.
7. Aseltine, J.A.: *Transform Method in Linear System Analysis*, McGraw-Hill, New York, 1958.
8. Nilson, J.W.: *Electric Circuits*, 2nd ed., Addison-Wesley, Reading, Mass., 1980.
9. Ward, J.R., and R.D. Strum: *State Variable Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1970.
10. Strang, G.: *Linear Algebra and Its Applications*, 3rd ed., Harcourt, Brace, Jovanovich, New York, 1988.
11. Taylor, F.J.: "A Novel Inversion of (sI-A)," *Int. J. Systems Science*, vol. 5, no. 2, pp. 153–160, Feb. 1974.

5

System Representation

5.1 INTRODUCTION

This chapter introduces the basic principles of system representation. From the concepts introduced in earlier chapters, a number of systems are represented in block-diagram form. Feedback is included in these systems in order to achieve the desired performance. Also, the standard symbols and definitions are presented. These are extensively used in the technical literature on control systems and form a common basis for understanding and clarity. While block diagrams simplify the representation of functional relationships within a system, the use of signal flow graphs (SFG) provides further simplification for larger systems that contain intercoupled relationships. Simulation diagrams are presented as a means of representing a system for which the overall differential equation is known. Then the inclusion of initial condition in SFG leads to the state-diagram representation. To provide flexibility in the method used by the designer for system analysis and design, the system representation may be in either the transfer-function or the state-equation form. The procedures for converting from one representation to the other are presented in Chap. 4 and in this chapter. When the state-variable format is used, the mathematical equations may be written so that the states are uncoupled. This simplifies the equations, leading to an A matrix that is diagonal. The techniques for transforming the state vector in order to achieve different representations are presented.

5.2 BLOCK DIAGRAMS

The representation of physical components by blocks is shown in Chap. 2. For each block the transfer function provides the dynamical mathematical relationship between the input and output quantities. Also, Chap. 1 and Fig. 1.3 describe the concept of feedback, which is used to achieve a better response of a control system to a command input. This section presents several examples of control systems and their representation by block diagrams. The blocks represent the functions performed rather than the components of the system.

Example 1: A Temperature Control System. An industrial process temperature control system is shown in Fig. 5.1*a*. The requirement is to control the temperature θ in the tank. The voltage r, obtained from a potentiometer, is calibrated in terms of the desired temperature θ_{comm}. This voltage represents the input quantity to the feedback control system. The actual temperature θ, the output quantity, is measured by means of a thermocouple immersed in the tank. The voltage e_{th} produced in the thermocouple is proportional to θ. The voltage e_{th} is amplified to produce the voltage b, which is the feedback quantity. The voltage $e = r - b$ is the *actuating signal* and is amplified by the

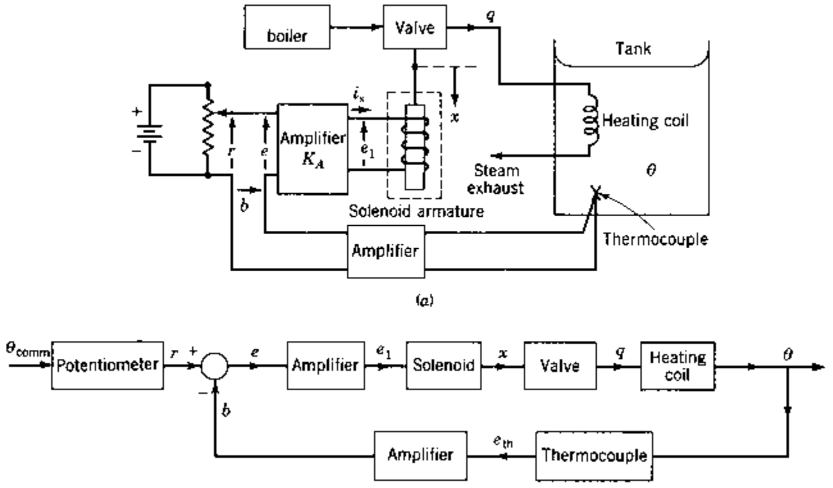


FIGURE 5.1 An industrial process temperature control system.

www.4electron.com

amplifier having the gain K_A to produce the solenoid voltage e_1. The current i_s in the solenoid, which results from applying e_1 produce a proportional force f_s that acts on the solenoid armature and valve to control the valve position x. The valve position in turn controls the flow of hot steam q from the boiler into the heating coil in the tank. The resulting temperature θ of the tank is directly proportional to the steam flow with a time delay that depends on the specific heat of the fluid and the mixing rate. The block diagram representation for this system in Fig. 5.1*b* shows the functions of each unit and the signal flow through the system.

To show the operation of the system, consider that an increase in the tank temperature is required. The voltage r is increased to a value that represents the value of the desired temperature. This change in r causes an increase in the actuating signal e. This increase in e, through its effect on the solenoid and valve, causes an increase in the amount of hot steam flowing through the heating coil in the tank. The temperature θ therefore increases in proportion to the steam flow. When the output temperature rises to a value essentially equal to the desired temperature, the feedback voltage b is equal to the reference input r ($b \approx r$). The flow of steam through the heating coil is stabilized at a steady-state value but maintains θ at the desired value.

Example 2: Command Guidance Interceptor System. A more complex system is the command guidance system shown in Fig. 5.2, which directs the flight of a missile in space in order to intercept a moving target. The target may be an enemy bomber whose aim is to drop bombs at some position.

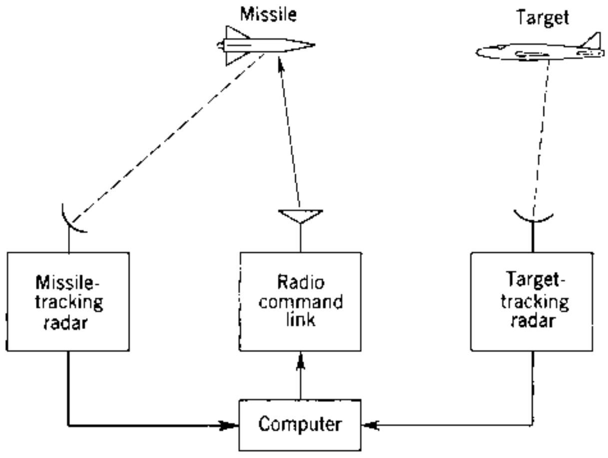


FIGURE 5.2 Command guidance interceptor system.

The defense uses the missile with the objective of intercepting and destroying the bomber before it launches its bombs. The target-tracking radar is used first for detecting and then for tracking the target. It supplies information on target range and angle and their rates of change (time derivatives). This information is continuously fed into the computer, which calculates a predicted course for the target. The missile-tracking radar supplies similar information that is used by the computer to determine its flight path. The computer compares the two flight paths and determines the necessary change in missile flight path to produce a collision course. The necessary flight path changes are supplied to the radio command link, which transmits this information to the missile. This electrical information containing corrections in flight path is used by a control system in the missile. The missile control system converts the error signals into mechanical displacements of the missile airframe control surfaces by means of actuators. The missile responds to the positions of the aerodynamic control surfaces to follow the prescribed flight path, which is intended to produce a collision with the target. Monitoring of the target is continuous so that changes in the missile course can be corrected up to the point of impact. The block diagram of Fig. 5.3 depicts the functions of this command guidance system. Many individual functions are performed within each block. Some of the components of the missile control system are shown within the block representing the missile.

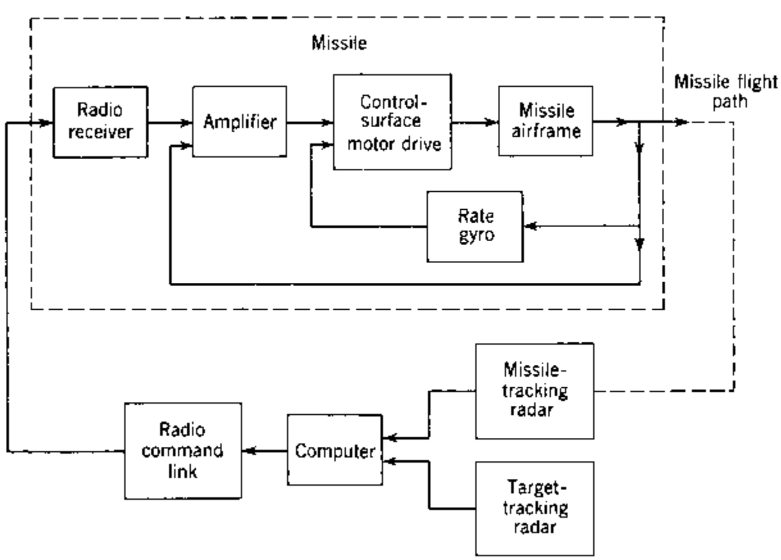


FIGURE 5.3 Block diagram of a generalized command guidance interceptor system.

Example 3: Aircraft Control System [1]. The feedback control system used to keep an airplane on a predetermined course or heading is necessary for the navigation of commercial airliners. Despite poor weather conditions and lack of visibility, the airplane must maintain a specified heading and altitude in order to reach its destination safely. In addition, in spite of rough air, the trip must be made as smooth and comfortable as possible for the passengers and crew. The problem is considerably complicated by the fact that the airplane has six degrees of freedom. This fact makes control more difficult than the control of a ship, whose motion is limited to the surface of the water. A *flight controller* is the feedback control system used to control the aircraft motion.

Two typical signals to the system are the correct flight path, which is set by the pilot, and the level position (attitude) of the airplane. The ultimately controlled variable is the actual course and position of the airplane. The output of the control system, the controlled variable, is the aircraft heading. In conventional aircraft three primary control surfaces are used to control the physical three-dimensional attitude of the airplane, the elevators, rudder, and ailerons. The axes used for an airplane and the motions produced by the control surface are shown in Fig. 5.4.

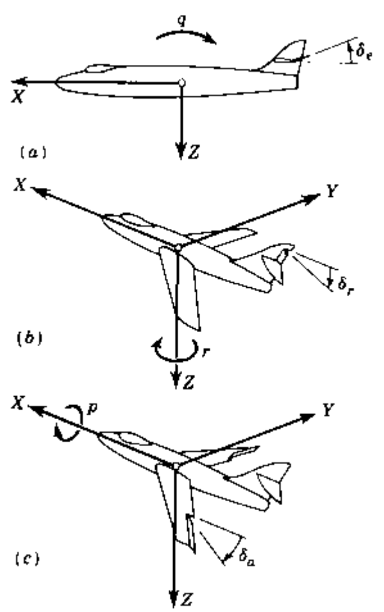


FIGURE 5.4 Airplane control surfaces: (*a*) elevator deflection produces pitching velocity *q*; (*b*) rudder deflection produces yawing velocity *r*; (*c*) aileron deflection produces rolling velocity *p*.

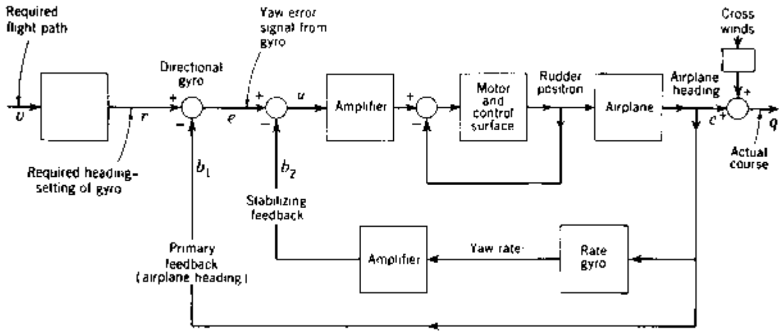


FIGURE 5.5 Airplane directional-control system.

The directional gyroscope is used as the error-measuring device. Two gyros must be used to provide control of both heading and attitude (level position) of the airplane. The error that appears in the gyro as an angular displacement between the rotor and case is translated into a voltage by various methods, including the use of transducers such as potentiometers, synchros, transformers, or microsyns.

Additional stabilization for the aircraft can be provided in the control system by rate feedback. In other words, in addition to the primary feedback, which is the position of the airplane, another signal proportional to the angular rate of rotation of the airplane around the vertical axis is fed back in order to achieve a stable response. A "rate" gyro is used to supply this signal. This additional stabilization may be absolutely necessary for some of the newer high-speed aircraft.

A block diagram of the aircraft control system (Fig. 5.5) illustrates control of the airplane heading by controlling the rudder position. In this system the heading that is controlled is the direction the airplane would travel in still air. The pilot corrects this heading, depending on the crosswinds, so that the actual course of the airplane coincides with the desired path. Another control included in the complete airplane control system controls the ailerons and elevators to keep the airplane in level flight.

5.3 DETERMINATION OF THE OVERALL TRANSFER FUNCTION

The block diagram of a control system with negative feedback can often be simplified to the form shown in Fig. 5.6 where the standard symbols and definitions used in feedback systems are indicated. In this feedback control system the output is the controlled variable C. This output is measured by

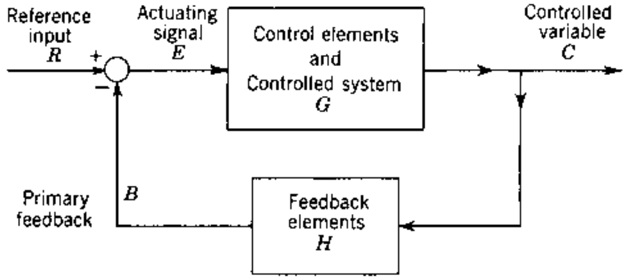


FIGURE 5.6 Block diagram of a feedback system.

a feedback element H to produce the primary feedback signal B, which is then compared with the reference input R. The difference E, between the reference input R and the feedback signal B, is the input to the controlled system G and is referred to as the *actuating signal*. For unity feedback systems where $H = 1$, the actuating signal is equal to the error signal, which is the difference $R - C$. The transfer functions of the forward and feedback components of the system are G and H, respectively.

In using the block diagram to represent a linear feedback control system where the transfer functions of the components are known, the letter symbol is capitalized, indicating that it is a transformed quantity; i.e., it is a function of the operator D, the complex parameter s, or the frequency parameter $j\omega$. This representation applies for the transfer function, where G is used to represent $G(D)$, $G(s)$, or $\mathbf{G}(j\omega)$. It also applies to all variable quantities, such as C, which represents $C(D)$, $C(s)$, or $\mathbf{C}(j\omega)$. Lowercase symbols are used, as in Fig. 5.5, to represent any function in the time domain. For example, the symbol c represents $c(t)$.

The important characteristic of such a system is the *overall transfer function*, which is the ratio of the transform of the controlled variable C to the transform of the reference input R. This ratio may be expressed in operational, Laplace transform, or frequency (phasor) form. The overall transfer function is also referred to as the *control ratio*. The terms are used interchangeably in this text.

The equations describing this system in terms of the transform variable are

$$C(s) = G(s)E(s) \tag{5.1}$$

$$B(s) = H(s)C(s) \tag{5.2}$$

$$E(s) = R(s) - B(s) \tag{5.3}$$

Combining these equations produces the *control ratio*, or *overall transfer function*,

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)H(s)} \tag{5.4}$$

The characteristic equation of the closed-loop system is obtained from the denominator of the control ratio:

$$1 + G(s)H(s) = 0 \tag{5.5}$$

The stability and response of the closed-loop system, as determined by analysis of the characteristic equation, are discussed more fully in later chapters.

For simplified systems where the feedback is unity, that is, $H(s) = 1$, the actuating signal, given by Eq. (5.3), is now the error present in the system, i.e., the *reference input minus the controlled variable, expressed by*

$$E(s) = R(s) - C(s) \tag{5.6}$$

The control ratio with unity feedback is

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)} \tag{5.7}$$

The *open-loop transfer function* is defined as the ratio of the output of the feedback path $B(s)$ to the actuating signal $E(s)$ for any given feedback loop. In terms of Fig. 5.6, the open-loop transfer function is

$$\frac{B(s)}{E(s)} = G(s)H(s) \tag{5.8}$$

The *forward transfer function* is defined as the ratio of the controlled variable $C(s)$ to the actuating signal $E(s)$. For the system shown in Fig. 5.5 the forward transfer function is

$$\frac{C(s)}{E(s)} = G(s) \tag{5.9}$$

In the case of unity feedback, where $H(s) = 1$, the open-loop and the forward transfer functions are the same. The forward transfer function $G(s)$ may be made up not only of elements in cascade but may also contain internal, or *minor*, feedback loops. The algebra of combining these internal feedback loops is similar to that used previously. An example of a controlled system with an internal feedback loop is shown in Fig. 5.5.

It is often useful to express the actuating signal E in terms of the input R. Solving from Eqs. (5.1) to (5.3) gives

$$\frac{E(s)}{R(s)} = \frac{1}{1 + G(s)H(s)} \tag{5.10}$$

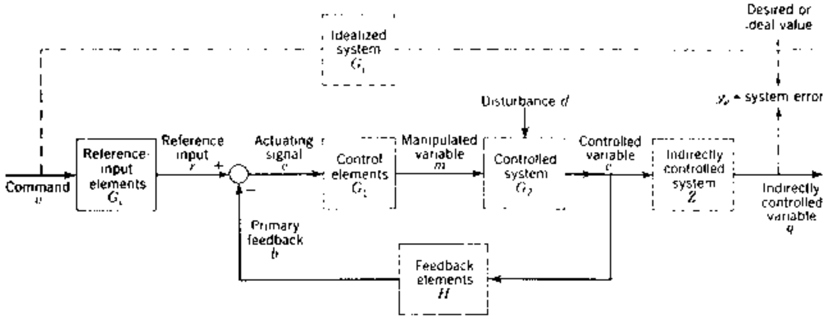


FIGURE 5.7 Block diagram of feedback control system containing all basic elements.

The concept of system error y_e is defined as the ideal or desired system value minus the actual system output. The ideal value establishes the desired performance of the system. For unity-feedback systems the actuating signal is an actual measure of the error and is directly proportional to the system error (see Fig. 5.7).

Example: Overall Transfer Function

Computation of several calculations using MATLAB are illustrated below. Assume that the transfer functions for Fig. 5.6 are:

$$G(s) = \frac{5}{s+5} \qquad H(s) = \frac{1}{s}$$

Determine the overall control ratio, C(s)/R(s). MATLAB provides the "feedback" function to calculate the overall (closed-loop) transfer function as shown by the following m-file.

```
%
% The feedback command calculates the closed-loop transfer
% function for a given forward transfer function G(s)
% and feedback transfer function H(s)
%
% Type ''help feedback'' for more information
% Example 1:
% Define a forward transfer function G(s) = 5/s + 5
% using the tf command

% G = tf([5],[1 5]) % System = tf (numerator, denominator)

Transfer function:

  5
-----
s + 5
```

www.4electron.com

```
%  Example 2:
%  Define a feedback transfer function H(s) = 1/s
%  using the tf command
H = tf([1],[1 0])  %  System = tf (numerator, denominator)

Transfer function:

1
-
s


%  Example 3:
%  For a non-unity, negative feedback system the
%  closed loop transfer function is
%
cltf = feedback (G,H,-1)

Transfer function:

     5s
----------
s^2 + 5s + 5


%  For positive feedback use ''feedback (G,H,1)''

     5s
----------
s^2 + 5s + 5


%  Example 4:
%  The forward transfer function is calculated by multiplying
%  G(s) and H(s)
%
G * H

Transfer function:

    5
-------
 s^2 + 5s
```

5.4 STANDARD BLOCK DIAGRAM TERMINOLOGY [2]

Figure 5.7 shows a block diagram representation of a feedback control system containing the basic elements. Figure 5.8 shows the block diagram and symbols of a more complicated system with multiple paths and inputs. Numerical subscripts are used to distinguish between blocks of similar functions in the

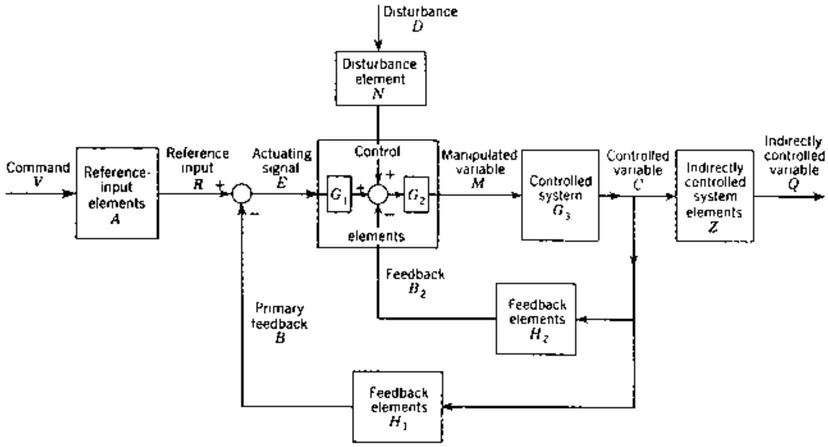


FIGURE 5.8 Block diagram of representative feedback control system showing multiple feedback loops. (All uppercase letters denote transformation.)

circuit. For example, in Fig. 5.7 the control elements are designated by G_1 and the controlled system by G_2. In Fig. 5.8 the control elements are divided into two blocks, G_1 and G_2, to aid in representing relations between parts of the system. Also, in this figure the primary feedback is represented by H_1. Additional minor feedback loops might be designated by H_2, H_3, and so forth.

The idealized system represented by the block enclosed with dashed lines in Fig. 5.7 can be understood to show the relation between the basic input to the system and the performance of the system in terms of the desired output. This would be the system agreed upon to establish the ideal value for the output of the system. In systems where the command is actually the desired value or ideal value, the idealized system would be represented by unity. The arrows and block associated with the idealized system, the ideal value, and the system error are shown in dashed lines on the block diagram because they do not exist physically in any feedback control system. For any specific problem, it represents the system (conceived in the mind of the designer) that gives the best approach, when considered as a perfect system, to the desired output or ideal value.

Definitions: Variables in the System

Command v is the input that is established by some means external to, and independent of, the feedback control system.

Reference input r is derived from the command and is the actual signal input to the system.

Controlled variable c is the quantity that is directly measured and controlled. It is the output of the controlled system.

Primary feedback b is a signal that is a function of the controlled variable and that is compared with the reference input to obtain the actuating signal.

Actuating signal e is obtained from a comparison measuring device and is the reference input minus the primary feedback. This signal, usually at a low energy level, is the input to the control elements that produce the manipulated variable.

Manipulated variable m is the quantity obtained from the control elements that is applied to the controlled system. The manipulated variable is generally at a higher energy level than the actuating signal and may also be modified in form.

Indirectly controlled variable q is the output quantity that is related through the indirectly controlled system to the controlled variable. It is outside the closed loop and is not directly measured for control.

Ultimately controlled variable is a general term that refers to the indirectly controlled variable. In the absence of the indirectly controlled variable, it refers to the controlled variable.

Ideal value i is the value of the ultimately controlled variable that would result from an idealized system operating with the same command as the actual system.

System error y_e is the ideal value minus the value of the ultimately controlled variable.

Disturbance d is the unwanted signal that tends to affect the controlled variable. The disturbance may be introduced into the system at many places.

Definitions: System Components

Reference input elements G_v produce a signal r proportional to the command.

Control elements G_c produce the manipulated variable m from the actuating signal.

Controlled system G is the device that is to be controlled. This is frequently a high-power element.

Feedback element H produces the primary feedback b from the controlled variable. This is generally a proportionality device but may also modify the characteristics of the controlled variable.

Indirectly controlled system Z relates the indirectly controlled variable q to the controlled quantity c. This component is outside the feedback loop.

Idealized system G_i is one whose performance is agreed upon to define the relationship between the ideal value and the command. This is often called the *model* or *desired* system.

Disturbance element N denotes the functional relationship between the variable representing the disturbance and its effect on the control system.

5.5 POSITION CONTROL SYSTEM

Figure 5.9 shows a simplified block diagram of an angular position control system. The reference selector and the sensor, which produce the reference input $R = \theta_R$ and the controlled output position $C = \theta_o$, respectively, consist of rotational potentiometers. The combination of these units represents a rotational comparison unit that generates the actuating signal E for the position control system, as shown in Fig. 5.10*a*, where K_θ, in volts per radian, is the potentiometer sensitivity constant. The symbolic comparator for this system is shown in Fig. 5.10*b*.

The transfer function of the motor-generator control is obtained by writing the equations for the schematic diagram shown in Fig. 5.11*a*. This figure shows a dc motor that has a constant field excitation and drives an

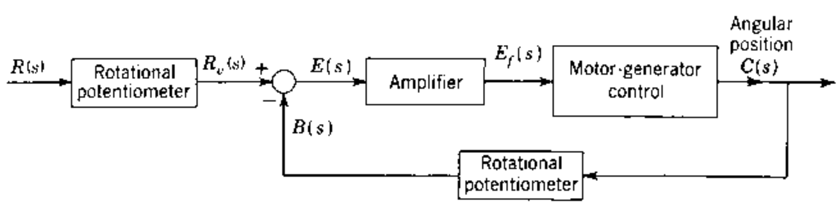


FIGURE 5.9 Position control system.

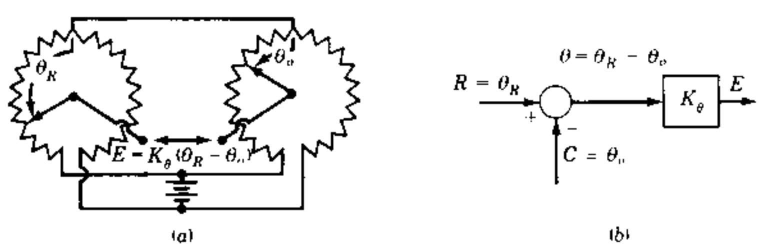
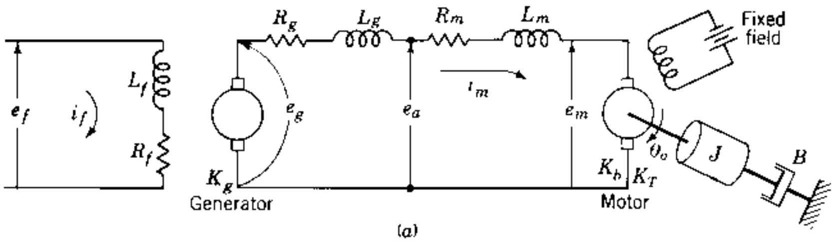


FIGURE 5.10 (*a*) Rotational position comparison; (*b*) its block diagram representation.



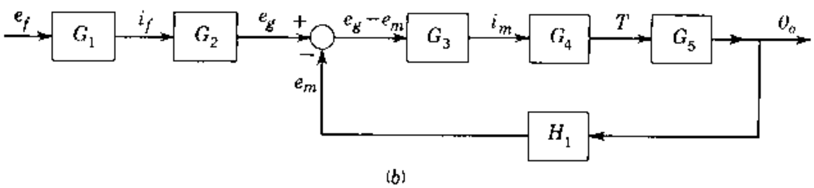


FIGURE 5.11 (*a*) Motor-generator control; (*b*) block diagram.

inertia and friction load. The armature voltage for the motor is furnished by the generator, which is driven at constant speed by a prime mover. The generator voltage e_g is determined by the voltage e_f applied to the generator field. The generator is acting as a power amplifier for the signal voltage e_f. The equations for this system are as follows:

$$e_f = L_f\, Di_f + R_f i_f \tag{5.11}$$

$$e_g = K_g i_f \tag{5.12}$$

$$e_g - e_m = (L_g + L_m)\, Di_m + (R_g + R_m) i_m \tag{5.13}$$

$$e_m = K_b\, D\theta_o \tag{5.14}$$

$$T = K_T i_m = J\, D^2\theta_o + B\, D\theta_o \tag{5.15}$$

The block diagram drawn in Fig. 5.11*b* is based on these equations. Starting with the input quantity e_f, Eq. (5.11) shows that a current i_f is produced. Therefore, a block is drawn with e_f as the input and i_f as the output. Equation (5.12) shows that a voltage e_g is generated as a function of the current i_f. Therefore, a second block is drawn with the current i_f as the input and e_g as the output. Equation (5.13) relates the current i_m in the motor to the difference of two voltages, $e_g - e_m$. To obtain this difference a summation point is introduced. The quantity e_g from the previous block enters this summation point. To obtain the quantity $e_g - e_m$ there must be added the quantity e_m, entering summation point with a minus sign. Up to this point the manner in which e_m is obtained has not yet been determined. The output of this summation point is used as the input to the next block, from which the current i_m is the output. Similarly, the block with current i_m as the input and the generated torque T as the output and the block with the torque input and the resultant motor position as the output are drawn. There must be no loose ends in the complete diagram; i.e., every dependent variable must be connected through a block or blocks into the system. Therefore, e_m is obtained from Eq. (5.14), and a block representing this relationship is drawn with θ_o as the input and e_m as the output. By using this procedure, the block diagram is completed and the functional relationships in the system are described. Note that the generator and the motor are no longer separately distinguishable. Also, the input and the output of each block are not in the same units, as both electrical and mechanical quantities are included.

The transfer functions of each block, as determined in terms of the pertinent Laplace transforms, are as follows:

$$G_1(s) = \frac{I_f(s)}{E_f(s)} = \frac{1/R_f}{1 + (L_f/R_f)s} = \frac{1/R_f}{1 + T_f s} \tag{5.16}$$

$$G_2(s) = \frac{E_g(s)}{I_f(s)} = K_g \tag{5.17}$$

$$G_3(s) = \frac{I_m(s)}{E_g(s) - E_m(s)} = \frac{1/(R_g + R_m)}{1 + [(L_g + L_m)/(R_g + R_m)]s}$$
$$= \frac{1/R_{gm}}{1 + (L_{gm}/R_{gm})s} = \frac{1/R_{gm}}{1 + T_{gm}s} \tag{5.18}$$

$$G_4(s) = \frac{T(s)}{I_m(s)} = K_T \tag{5.19}$$

$$G_5(s) = \frac{\Theta_o(s)}{T(s)} = \frac{1/B}{s[1 + (J/B)s]} = \frac{1/B}{s(1 + T_n s)} \tag{5.20}$$

$$H_1(s) = \frac{E_m(s)}{\Theta_o(s)} = K_b s \tag{5.21}$$

The block diagram can be simplified, as shown in Fig. 5.12, by combining the blocks in cascade. The block diagram is further simplified, as shown in

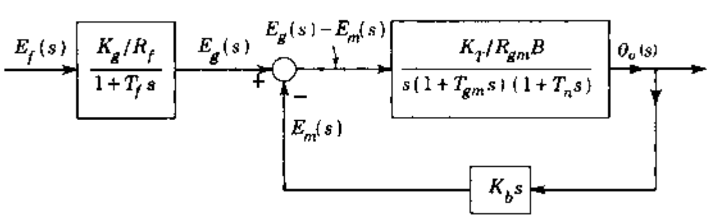


FIGURE 5.12 Simplified block diagram.

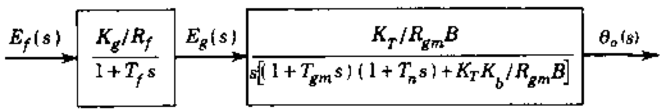


FIGURE 5.13 Reduced block diagram.

Fig. 5.13, by evaluating an equivalent block from $E_g(s)$ to $\Theta_o(s)$, using the principle of Eq. (5.4). The final simplification results in Fig. 5.14. The overall transfer function $G_x(s)$ is

$$G_x(s) = \frac{\Theta_o(s)}{E_f(s)} = \frac{K_g K_T/R_f B R_{gm}}{s(1+T_f s)[(1+K_T K_b/BR_{gm}) + (T_{gm}+T_n)s + T_{gm}T_n s^2]} \tag{5.22}$$

This expression is the exact transfer function for the entire motor and generator combination. Certain approximations, if valid, may be made to simplify this expression. The first approximation is that the inductance of the generator and motor armatures is very small, therefore, $T_{gm} \approx 0$. With this approximation, the transfer function reduces to

$$\frac{\Theta_o(s)}{E_f(s)} = G_x(s) \approx \frac{K_x}{s(1+T_f s)(1+T_m s)} \tag{5.23}$$

where

$$K_x = \frac{K_g K_T}{R_f(BR_{gm} + K_T K_b)} \tag{5.24}$$

$$T_f = \frac{L_f}{R_f} \tag{5.25}$$

$$T_m = \frac{JR_{gm}}{BR_{gm} + K_T K_b} \tag{5.26}$$

If the frictional effect of the load is very small, the approximation can be made that $B \approx 0$. With this additional approximation, the transfer function has the same form as Eq. (5.23), but the constants are now

$$K_x = \frac{K_g}{R_f K_b} \tag{5.27}$$

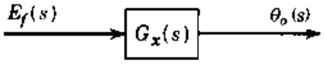


FIGURE 5.14 Simplified block diagram for Fig. 5.11.

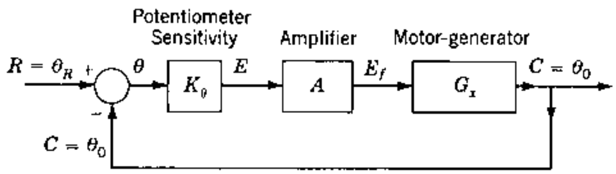


FIGURE 5.15 Equivalent representation of Fig. 5.9.

$$T_f = \frac{L_f}{R_f} \tag{5.28}$$

$$T_m = \frac{JR_{gm}}{K_T K_b} \tag{5.29}$$

For simple components, as in this case, an overall transfer function can often be derived more easily by combining the original system equations. This can be done without the intermediate steps shown in this example. However, the purpose of this example is to show the representation of dynamic components by individual blocks and the combination of blocks.

A new block diagram representing the position-control system of Fig. 5.9 is drawn in Fig. 5.15. Since the transfer function of the amplifier of Fig. 5.9 is $E_f(s)/E(s) = A$, the forward transfer function of the position-control system of Fig. 5.15 is

$$G(s) = \frac{\Theta_o(s)}{\Theta(s)} = \frac{AK_x K_\theta}{s(1 + T_f s)(1 + T_m s)} \tag{5.30}$$

The overall transfer function (control ratio) for this system, with $H(s) = 1$, is

$$\frac{C(s)}{R(s)} = \frac{\Theta_o(s)}{\Theta_R(s)} = \frac{G(s)}{1 + G(s)H(s)} = \frac{AK_x K_\theta}{s(1 + T_f s)(1 + T_m s) + AK_x K_\theta} \tag{5.31}$$

5.6 SIMULATION DIAGRAMS [3,4]

The simulation used to represent the dynamic equations of a system may show the actual physical variables that appear in the system, or it may show variables that are used purely for mathematical convenience. In either case the overall response of the system is the same. The simulation diagram is similar to the diagram used to represent the system on an analog computer. The basic elements used are ideal integrators, ideal amplifiers, and ideal summers, shown in Fig. 5.16. Additional elements such as multipliers and dividers may be used for nonlinear systems.

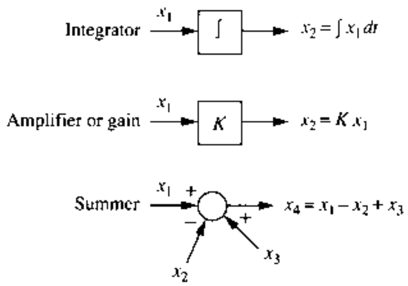


FIGURE 5.16 Elements used in a simulation diagram.

One of the methods used to obtain a simulation diagram includes the following steps:

1. Start with differential equation.
2. On the left side of the equation put the highest-order derivative of the *dependent* variable. A first-order or higher-order derivative of the input may appear in the equation. In this case the highest-order derivative of the input is also placed on the left side of the equation. All other terms are put on the right side.
3. Start the diagram by assuming that the signal, represented by the terms on the left side of the equation, is available. Then integrate it as many times as needed to obtain all the lower-order derivatives. It may be necessary to add a summer in the simulation diagram to obtain the dependent variable explicitly.
4. Complete the diagram by feeding back the approximate outputs of the integrators to a summer to generate the original signal of step 2. Include the input function if it is required.

Example. Draw the simulation diagram for the series *RLC* circuit of Fig. 2.2 in which the output is the voltage across the capacitor.

Step 1. When $y = v_c$ and $u = e$ are used, Eq. (2.10) becomes

$$LC\ddot{y} + RC\dot{y} + y = u \tag{5.32}$$

Step 2. Rearrange terms to the form

$$\begin{aligned}\ddot{y} &= \frac{1}{LC}u - \frac{R}{L}\dot{y} - \frac{1}{LC}y \\ &= bu - a\dot{y} - by \end{aligned} \tag{5.33}$$

where $a = R/L$ and $b = 1/LC$.

Step 3. The signal $\ddot{y}$ is integrated twice, as shown in Fig. 5.17*a*.

Step 4. The block or simulation diagram is completed as shown in Fig. 5.17*b* in order to satisfy Eq. (5.33).

The state variables are often selected as the outputs of the integrators in the simulation diagram. In this case they are $y = x_1$ and $\dot{y} = x_2 = \dot{x}_1$. These values and $\ddot{y} = \dot{x}_2$ are shown in Fig. 5.17*b*. The state and output equations are therefore

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\dfrac{1}{LC} & -\dfrac{R}{L} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \dfrac{1}{LC} \end{bmatrix} u = \mathbf{A}\mathbf{x} + \mathbf{B}u \tag{5.34}$$

$$\mathbf{y} = [1 \quad 0] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \mathbf{0}u = \mathbf{C}\mathbf{x} + \mathbf{D}u \tag{5.35}$$

It is common in a physical system for the **D** matrix to be zero, as in this example. These equations are different from Eqs. (2.26) and (2.27), yet they represent the same system. This difference illustrates that state variables are not unique. When the state variables are the dependent variable and the derivatives of the dependent variable, as in this example, they are called *phase variables.* The phase variables are applicable to differential equations of any order and can be used without drawing the simulation diagram, as shown below.

CASE 1. The general differential equation that contains no derivatives of the input is

$$D^n y + a_{n-1} D^{n-1} y + \cdots + a_1 D_y + a_0 y = u \tag{5.36}$$

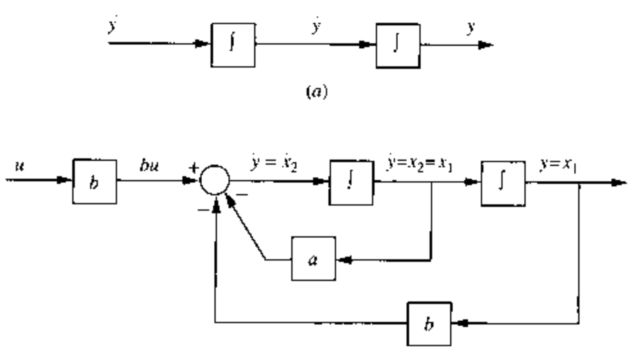


FIGURE 5.17 Simulation diagram for Eq. (5.32).

The state variables are selected as the *phase variables*, which are defined by $x_1 = y$, $x_2 = Dx_1 = \dot{x}_1 = Dy$, $x_3 = \dot{x}_2 = D^2y, \ldots, x_n = \dot{x}_{n-1} = D^{n-1}y$. Thus, $D^n y = \dot{x}_n$. When these state variables are used, Eq. (5.36) becomes

$$\dot{x}_n + a_{n-1}x_n + a_{n-2}x_{n-1} + \cdots + a_1x_2 + a_0x_1 = u \tag{5.37}$$

The resulting state and output equations using phase variables are

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \\ \vdots \\ \\ \dot{x}_{n-1} \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} 0 & 1 & & & & & \\ & 0 & 1 & & & & \mathbf{0} \\ & & 0 & 1 & & & \\ & & & 0 & \ddots & & \\ & & & & \ddots & & \\ & \mathbf{0} & & & & 0 & 1 \\ -a_0 & -a_1 & -a_2 & -a_3 & \cdots & -a_{n-2} & -a_{n-1} \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ 0 \\ \\ \vdots \\ \\ 0 \\ 1 \end{bmatrix} \mathbf{u}$$

$$\dot{\mathbf{x}} = \mathbf{A}_c\mathbf{x} + \mathbf{b}_c\mathbf{u} \tag{5.38}$$

$$\mathbf{y} = [x_1] = \begin{bmatrix} 1 & 0 & 0 & \cdots \end{bmatrix} \mathbf{x} = \mathbf{c}_c^T \mathbf{x} \tag{5.39}$$

These equations can be written directly from the original differential equation (5.36). The plant matrix $\mathbf{A}_c$ contains the number 1 in the *superdiagonal* and the negative of the coefficients of the original differential equation in the *n*th row. In this simple form the matrix $\mathbf{A}_c$ is called the *companion matrix*. Also, the **B** matrix takes on the form shown in Eq. (5.38) and is indicated by $\mathbf{b}_c$.

CASE 2. When derivatives of $u(t)$ appear in the differential equation, the phase variables may be used as the state variables and the state equation given in Eq. (5.38) still applies; however, the output equation is no longer given by Eq. (5.39). This is shown by considering the differential equation

$$\begin{aligned}(D^n + a_{n-1}D^{n-1} + \cdots + a_1 D + a_0)y \\ = (c_w D^w + c_{w-1}D^{w-1} + \cdots + c_1 D + c_0)u \qquad w \le n\end{aligned} \tag{5.40}$$

The output y is specified as

$$y = (c_w D^w + c_{w-1}D^{w-1} + \cdots + c_1 D + c_0)x_1 \tag{5.41}$$

In terms of the state variables this output equation becomes

$$y = c_w\dot{x}_w + c_{w-1}x_w + \cdots + c_1x_2 + c_0x_1 \tag{5.42}$$

Substituting Eq. (5.41) into Eq. (5.40) and canceling common terms on both sides of the equation yields

$$(D^n + a_{n-1}D^{n-1} + \cdots a_1 D + a_0)x_1 = u \tag{5.43}$$

This equation has the same form as Eq. (5.36). Thus, in terms of phase variables, Eq. (5.43) can be represented in the form $\dot{\mathbf{x}} = \mathbf{A}_c\mathbf{x} + \mathbf{b}_c u$, where the matrices $\mathbf{A}_c$ and $\mathbf{b}_c$ are the same as in Eq. (5.38). The case for two values of w are:

1. For $w = n$, the output equation (5.42) becomes

$$y = c_n\dot{x}_n + c_{n-1}x_n + \cdots + c_1x_2 + c_0x_1$$

Using Eq. (5.37) to eliminate $\dot{x}_n$ yields

$$\begin{aligned} \mathbf{y} &= \begin{bmatrix} (c_0 - a_0c_n) & (c_1 - a_1c_n) & \cdots & (c_{n-1} - a_{n-1}c_n) \end{bmatrix}\mathbf{x} + [c_n]\mathbf{u} \\ &= \mathbf{c}_c^T\mathbf{x} + \mathbf{d}_c\mathbf{u} \end{aligned} \tag{5.44}$$

2. For $w < n$, $\dot{x}_w = x_{w+1}$, and $c_{w+i} = 0$ for $i = 1, 2, \ldots, n - w$. In this case the output $\mathbf{y}$ in Eq. (5.44) reduces to

$$\mathbf{y} = \begin{bmatrix} c_0 & c_1 & \cdots & c_w & 0 & \cdots & 0 \end{bmatrix}\mathbf{x} \tag{5.45}$$

Figure 5.18 shows the simulation diagram that represents the system of Eq. (5.40) in terms of the state equation, Eq. (5.38), and the output equation

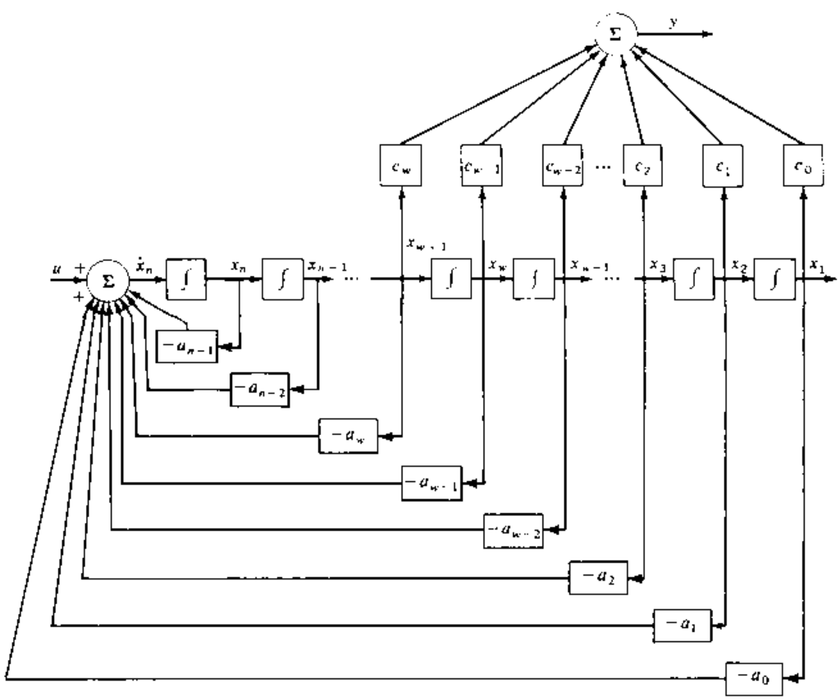


FIGURE 5.18 Simulation diagram representing the system of Eq. (5.40) in terms of Eqs. (5.38) and (5.45).

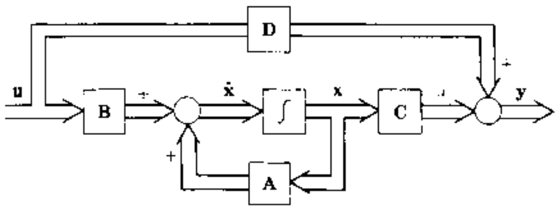


FIGURE 5.19 General matrix block diagram representing the state and output equations.

Eq. (5.45). It has the desirable feature that differentiation of the input signal $u(t)$ is not required to satisfy the differential equation. The simulation diagram has the advantage of requiring only two summers, regardless of the order of the system, one at the input and one at the output. This representation of a differential equation is considered again in Sec. 5.8. Differentiators are avoided in analog- and digital-computer simulations, because they accentuate any noise present in the signals.

Different sets of state variables may be selected to represent a system. The selection of the state variables determines the **A**, **B**, **C**, and **D** matrices of Eqs. (2.35) and (2.36). A general matrix block diagram representing the state and the output equations is shown in Fig. 5.19. The matrix **D** is called the *feedforward matrix*. The analysis and design of multiple-input multiple-output (MIMO) systems are covered in detail in Refs. 5 to 7.

5.7 SIGNAL FLOW GRAPHS [8,9]

The block diagram is a useful tool for simplifying the representation of a system. The block diagrams of Figs. 5.9 and 5.15 have only one feedback loop and may be categorized as simple block diagrams. Figure 5.5 has three feedback loops; thus it is not a simple system. When intercoupling exists between feedback loops, and when a system has more than one input and one output, the control system and block diagram are more complex. Having the block diagram simplifies the analysis of complex system. Such an analysis can be even further simplified by using a *signal flow graph* (SFG), which looks like a simplified block diagram.

An SFG is a diagram that represents a set of simultaneous equations. It consists of a *graph* in which *nodes* are connected by directed *branches.* The nodes represent each of the system variables. A branch connected between two nodes acts as a one-way signal multiplier: the direction of signal flow is indicated by an arrow placed on the branch, and the multiplication

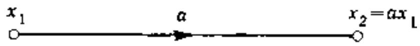


FIGURE 5.20 Signal flow graph for $x_2 = ax_1$.

factor (transmittance or transfer function) is indicated by a letter placed near the arrow. Thus, in Fig. 5.20, the branch transmits the signal x_1 from left to right and multiplies it by the quantity a in the process. The quantity a is the transmittance, or transfer function. It may also be indicated by $a = t_{12}$, where the subscripts show that the signal flow is from node 1 to node 2.

Flow-Graph Definitions

A *node* performs two functions:

1. *Addition* of the signals on all incoming branches
2. *Transmission* of the total node signal (the sum of all incoming signals) to all outgoing branches

These functions are illustrated in the graph of Fig. 5.21, which represents the equations

$$w = au + bv \qquad x = cw \qquad y = dw \tag{5.46}$$

Three types of nodes are of particular interest:

Source nodes (*independent nodes*). These represent independent variables and have only outgoing branches. In Fig. 5.21, nodes u and v are source nodes.

Sink nodes (*dependent nodes*). These represent dependent variables and have only incoming branches. In Fig. 5.21, nodes x and y are sink nodes.

Mixed nodes (*general nodes*). These have both incoming and outgoing branches. In Fig. 5.21, node w is a mixed node. A mixed node may be treated as a sink node by adding an outgoing branch of unity transmittance, as shown in Fig. 5.22, for the equation $x = au + bv$ and $w = cx = cau + cbv$.

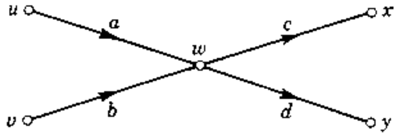


FIGURE 5.21 Signal flow graph for Eqs. (5.46).

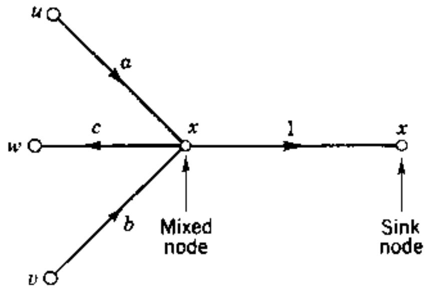


FIGURE 5.22 Mixed and sink nodes for a variable.

A *path* is any connected sequence of branches whose arrows are in the same direction.

A *forward path* between two nodes is one that follows the arrows of successive branches and in which a node appears only once. In Fig. 5.21 the path *uwx* is a forward path between the nodes *u* and *x*.

Flow-Graph Algebra

The following rules are useful for simplifying a signal flow graph:

Series paths (*cascade nodes*). Series paths can be combined into a single path by multiplying the transmittances as shown in Fig. 5.23*a*.

Path gain. The product of the transmittances in a series path.

Parallel paths. Parallel paths can be combined by adding the transmittances as shown in Fig. 5.23*b*.

Node absorption. A node representing a variable other than a source or sink can be eliminated as shown in Fig. 5.23*c*.

Feedback loop. A closed path that starts at a node and ends at the same node.

Loop gain. The product of the transmittances of a feedback loop.

The equations for the feedback system of Fig. 5.6 are as follows:

$$C = GE \tag{5.47}$$

$$B = HC \tag{5.48}$$

$$E = R - B \tag{5.49}$$

Note that an equation is written for each dependent variable. The corresponding SFG is shown in Fig. 5.24*a*. The nodes *B* and *E* can be eliminated in turn to produce Fig. 5.24*b* and Fig. 5.24*c*, respectively. Figure 5.24*c* has a *self-loop* of value $-GH$. The final simplification is to eliminate the self-loop to produce the

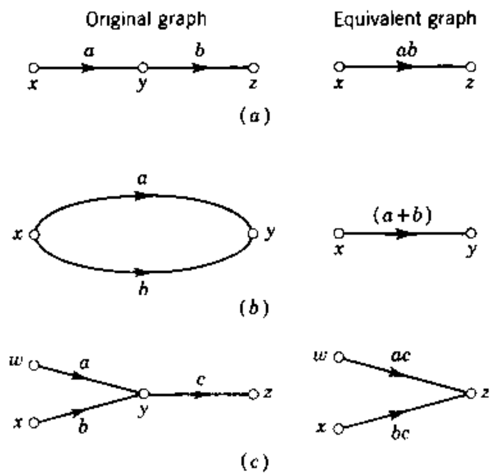


FIGURE 5.23 Flow-graph simplifications.

overall transmittance from the input R to the output C. This is obtained by summing signals at node C in Fig. 5.24c, yielding $C = GR - GHC$. Solving for C produces Fig. 5.24d.

General Flow-Graph Analysis

If all the source nodes are brought to the left and all the sink nodes are brought to the right, the SFG for an arbitrarily complex system can be represented by Fig. 5.25a. The effect of the *internal* nodes can be factored out by ordinary algebraic processes to yield the equivalent graph represented

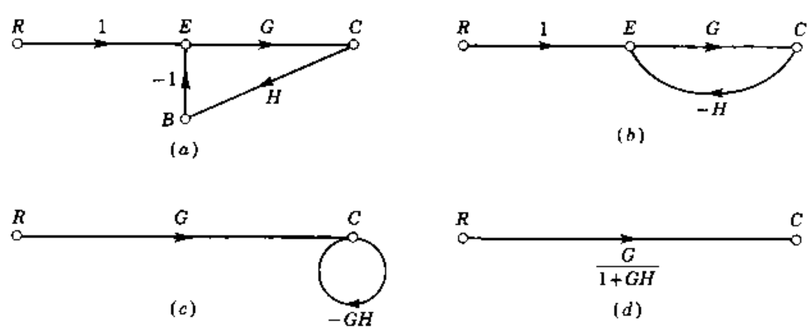


FIGURE 5.24 Successive reduction of the flow graph for the feedback system of Fig. 5.6.

by Fig. 5.25*b*. This simplified graph is represented by

$$y_1 = T_a x_1 + T_d x_2 \tag{5.50}$$

$$y_2 = T_b x_1 + T_e x_2 \tag{5.51}$$

$$y_3 = T_c x_1 + T_f x_2 \tag{5.52}$$

The T's, called overall graph transmittances, are the overall transmittances from a specified source node to a specified dependent node. For linear systems the principle of superposition can be used to "solve" the graph. That is, the sources can be considered one at a time. Then the output signal is equal to the sum of the contributions produced by each input. The overall transmittance can be found by the ordinary processes of linear algebra, i.e., by the solution of the set of simultaneous equations representing the system. However, the same results can be obtained directly from the SFG. The fact that they can produce answers to large sets of linear equations by *inspection* gives the SFGs their power and usefulness.

The Mason Gain Rule

The overall transmittance can be obtained from the Mason gain formula [8]. The formula and definitions are followed by an example to show its application. The overall transmittance T is given by

$$T = \frac{\sum T_n \Delta_n}{\Delta} \tag{5.53}$$

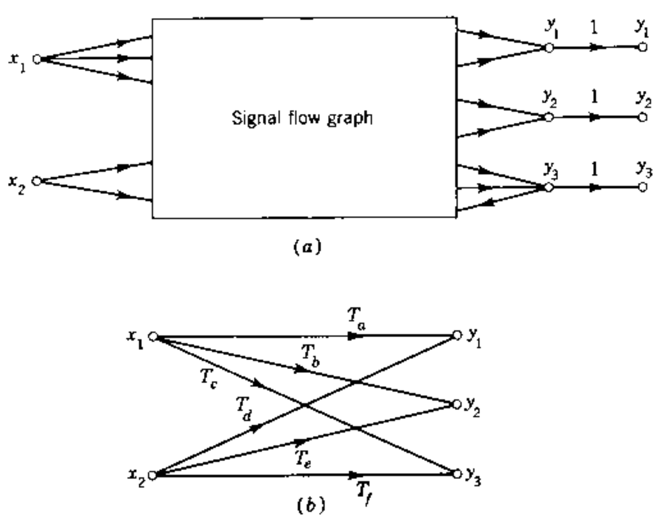


FIGURE 5.25 Equivalent signal flow graphs.

www.4electron.com

where T_n is the transmittance of each forward path between a source and a sink node and Δ is the graph determinant, found from

$$\Delta = 1 - \sum L_1 + \sum L_2 - \sum L_3 + \cdots \tag{5.54}$$

In this equation L_1 is the transmittance of each closed path, and $\sum L_1$ is the sum of the transmittances of all closed paths in the graph. L_2 is the produce of the transmittances of two *nontouching* loops; loops are nontouching if they do not have any common nodes. $\sum L_2$ is the sum of the products of transmittances in all possible combinations of nontouching loops taken two at a time. L_3 is the product of the transmittances of three nontouching loops. $\sum L_3$ is the sum of the products of transmittances in all possible combinations of nontouching loops taken three at a time.

In Eq. (5.53), Δ_n is the cofactor of T_n. It is the determinant of the remaining subgraph when the forward path that produces T_n is removed. Thus, Δ_n does not include any loops that touch the forward path in question. Δ_n is equal to unity when the forward path touches all the loops in the graph or when the graph contains no loops. Δ_n has the same form as Eq. (5.54).

Example. Figure 5.26*a* shows a block diagram. Since $E_1 = M_1 - B_1 - B_3 = G_1 E - H_1 M_2 = H_3 C$, it is not necessary to show M_l, $-B_l$, and B_3 explicitly. The SFG is shown in Fig. 5.26*b*. Since this is a fairly complex system, the resulting equation is expected to be complex. However, the application of Mason's rule produces the resulting overall transmittance in a systematic manner. This system has four loops, whose transmittances are $-G_2H_1$, $-G_5H_2$, $-G_1G_2G_3G_5$, and $-G_2G_3G_5H_3$. Therefore

$$\sum L_t = -G_2H_1 - G_5H_2 - G_1G_2G_3G_5 - G_2G_3G_5H_3 \tag{5.55}$$

Only two loops are nontouching; therefore,

$$\sum L_2 = (-G_2H_1)(-G_5H_2) \tag{5.56}$$

Although there are four loops, there is no set of three loops that are nontouching; therefore,

$$\sum L_3 = 0 \tag{5.57}$$

The system determinant can therefore be obtained from Eq. (5.54).

There is only one forward path between R and C. The corresponding forward transmittance is $G_1G_2G_3G_5$. If this path, with its corresponding nodes, is removed from the graph, the remaining subgraph has no loops. The cofactor Δ_n is therefore equal to unity. The complete overall transmittance from R to C, obtained from Eq. (5.53), is

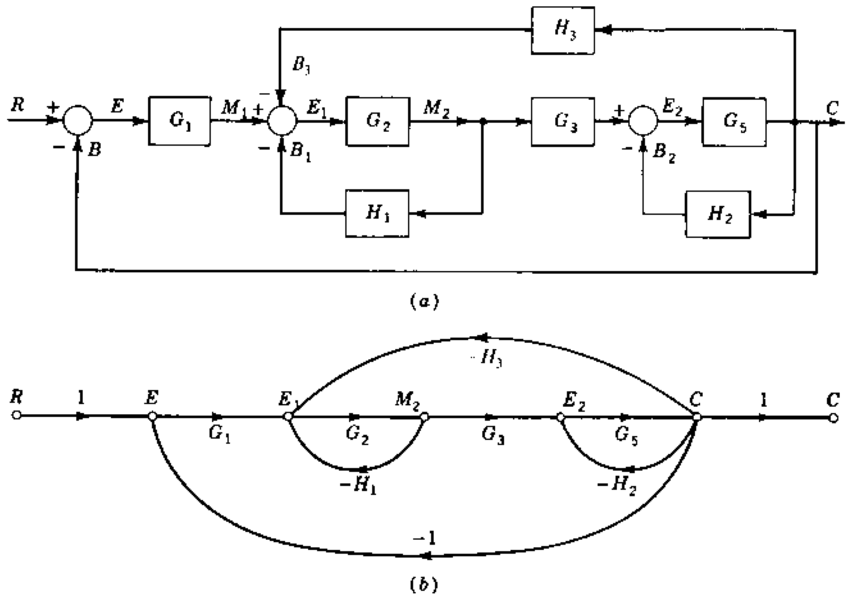


FIGURE 5.26 Block diagram and its signal flow graph.

$$T = \frac{G_1G_2G_3G_5}{1 + G_2H_1 + G_5H_2 + G_1G_2G_3G_5 + G_2G_3G_5H_3 + G_2G_5H_1H_2} \tag{5.58}$$

The complete expression for T is obtained by inspection from the SFG. The SFG is simpler than solving the five simultaneous equations that represent this system.

5.8 STATE TRANSITION SIGNAL FLOW GRAPH [10]

The state transition SFG or, more simply, the *state diagram*, is a simulation diagram for a system of equations and includes the initial conditions of the states. Since the state diagram in the Laplace domain satisfies the rules of Mason's SFG, it can be used to obtain the transfer function of the system and the transition equation. As described in Sec. 5.6, the basic elements used in a simulation diagram are a gain, a summer, and an integrator. The signal-flow representation in the Laplace domain for an integrator is obtained as follows:

$$\dot{x}_1(t) = x_2(t)$$

$$X_1(s) = \frac{X_2(s)}{s} + \frac{x_1(t_0)}{s} \tag{5.59}$$

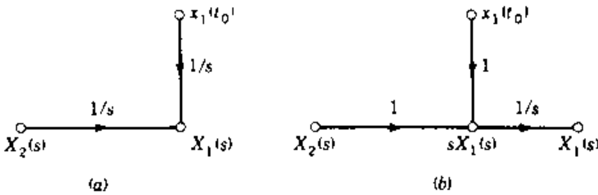


FIGURE 5.27 Representations of an integrator in the Laplace domain in a signal flow graph.

Equation (5.59) may be represented either by Fig. 5.27*a* or Fig. 5.27*b*.

A differential equation that contains no derivatives of the input, as given by Eq. (5.36), is repeated here:

$$D^n y + a_{n-1} D^{n-1} y + \cdots + a_1 Dy + a_0 y = u \tag{5.60}$$

In terms of the phase variables, with $x_1 = y$ and $\dot{x}_i = x_{i+1}$, this equation becomes

$$\dot{x}_n + a_{n-1} x_n + \cdots + a_1 x_2 + a_0 x_1 = u \tag{5.60a}$$

The state diagram is drawn in Fig. 5.28 with phase variables. It is obtained by first drawing the number of integrator branches $1/s$ equal to the order of the differential equation. The outputs of the integrators are designated as the state variables, and the initial conditions of the n states are included as inputs in accordance with Fig. 5.27*a*. Taking the Laplace transform of Eq. (5.60*a*) yields, after dividing by s.

$$X_n(s) = \frac{1}{s}[-a_0 X_1(s) - a_1 X_2(s) - \cdots - a_{n-1} X_n(s) + U(s)] + \frac{1}{s} x_n(t_0)$$

The node $X_n(s)$ in Fig. 5.28 satisfies this equation. The signal at the unlabeled node between $U(s)$ and $X_n(s)$ is equal to $sX_n(s) - x_n(t_0)$.

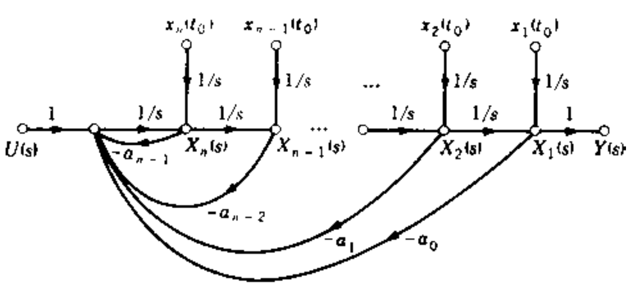


FIGURE 5.28 State diagram for Eq. (5.60).

The overall transfer function $Y(s)/U(s)$ is defined with all initial values of the states equal to zero. Using this condition and applying the Mason gain formula given by Eq. (5.53) yields

$$G(s) = \frac{Y(s)}{U(s)} = \frac{s^{-n}}{\Delta(s)} = \frac{s^{-n}}{1 + a_{n-1}s^{-1} + \cdots + a_1 s^{-(n-1)} + a_0 s^{-n}}$$
$$= \frac{1}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0} \tag{5.61}$$

The state transition equation of the system is obtained from the state diagram by applying the Mason gain formula and considering each initial condition as a source. This is illustrated by the following example.

Example 1. Equation (5.32) can be expressed as

$$\ddot{y} + \frac{R}{L}\dot{y} + \frac{1}{LC}y = \frac{1}{LC}u$$

(a) Draw the state diagram. (*b*) Determine the state transition equation.

Solution. (*a*) The state Diagram, Fig. 5.29, includes two integrators, because this is a second-order equation. The state variables are selected as the phase variables that are the outputs of the integrators, that is, $x_1 = y$ and $x_2 = \dot{x}_1$.

(*b*) The state transition equations are obtained by applying the Mason gain formula with the three inputs u, $x_1(t_0)$, and $x_2(t_0)$:

$$X_1(s) = \frac{s^{-1}(1 + s^{-1}R/L)}{\Delta(s)}x_1(t_0) + \frac{s^{-2}}{\Delta(s)}x_2(t_0) + \frac{s^{-2}/LC}{\Delta(s)}U(s)$$

$$X_2(s) = \frac{s^{-2}/LC}{\Delta(s)}x_1(t_0) + \frac{s^{-1}}{\Delta(s)}x_2(t_0) + \frac{s^{-1}/LC}{\Delta(s)}U(s)$$

$$\Delta(s) = 1 + \frac{s^{-1}R}{L} + \frac{s^{-2}}{LC}$$

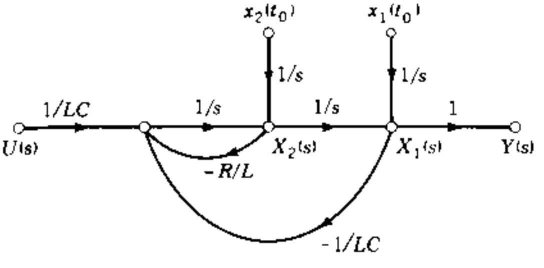


FIGURE 5.29 State diagram for Example 1.

After simplification, these equations become

$$\mathbf{X}(s) = \begin{bmatrix} X_1(s) \\ X_2(s) \end{bmatrix}$$

$$= \frac{1}{s^2 + (R/L)s + 1/LC} \left\{ \begin{bmatrix} s + \dfrac{R}{L} & 1 \\ \dfrac{-1}{LC} & s \end{bmatrix} \begin{bmatrix} x_1(t_0) \\ x_2(t_0) \end{bmatrix} + \begin{bmatrix} \dfrac{1}{LC} \\ \dfrac{s}{LC} \end{bmatrix} U(s) \right\}$$

This equation is of the form given by Eq. (4.104), i.e.,

$$\mathbf{X}(s) = \mathbf{\Phi}(s)\mathbf{x}(t_0) + \mathbf{\Phi}(s)\mathbf{B}\mathbf{U}(s)$$

Thus the resolvent matrix $\mathbf{\Phi}(s)$ is readily identified and, by use of the inverse Laplace transform, yields the state transition matrix $\mathbf{\Phi}(t)$. The elements of the resolvent matrix can be obtained directly from the SFG with $\mathbf{U}(s) = 0$ and only one $x_j(t_0)$ considered at a time, i.e., all other initial conditions set to zero. Each element of $\mathbf{\Phi}(s)$ is

$$\phi_{ij}(s) = \frac{X_i(s)}{x_j(t_0)}$$

For example,

$$\phi_{11}(s) = \frac{X_1(s)}{x_1(t_0)} = \frac{s^{-1}(1 + s^{-1}R/L)}{\Delta(s)} = \frac{s + R/L}{s^2 \Delta(s)}$$

The complete state transition equation $\mathbf{x}(t)$ is obtainable through use of the state diagram. Therefore, $\mathbf{\Phi}$(s) is obtained without performing the inverse operation $[s\mathbf{I} - \mathbf{A}]^{-1}$. Since $\mathbf{x}(t)$ represents phase variables, the system output is $y(t) = x_1(t)$.

Example 2. A differential equation containing derivatives of the input, given by Eq. (5.40), is repeated here.

$$\begin{aligned}(D^n + a_{n-1}D^{n-1} + \cdots + a_1 D + a_0)y \\ = (c_w D^w + c_{w-1}D^{w-1} + \cdots + c_1 D + c_0)u \qquad w \le n\end{aligned} \tag{5.62}$$

Phase variables can be specified as the state variables, provided the output is identified by Eq. (5.41) as

$$y = (c_w D^w + c_{w-1}D^{w-1} + \cdots + c_1 D + c_0)x_1 \tag{5.63}$$

Equation (5.62) then reduces to

$$(D^n + a_{n-1}D^{n-1} + \cdots + a_1 D + a_0)x_1 = u \tag{5.64}$$

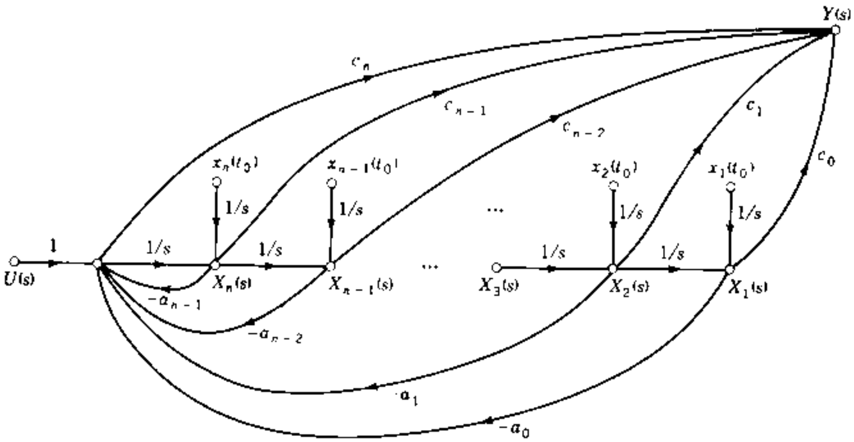


FIGURE 5.30 State diagram for Eq. (5.62) using phase variables, for $w = n$.

The resulting state diagram for Eqs. (5.63) and (5.64) is shown in Fig. 5.30 for $w = n$. The state equations obtained from the state diagram are given by Eq. (5.38). The output equation in matrix form is readily obtained from Fig. 5.30 in terms of the transformed variables $X(s)$ and the input $U(s)$ as

$$\begin{aligned}\mathbf{Y}(s) &= [(c_0 - a_0 c_n) \quad (c_1 - a_1 c_n) \quad \cdots \\ &\qquad (c_{n-2} - a_{n-2} c_n)\ (c_{n-1} - a_{n-1} c_n)]\mathbf{X}(s) + c_n \mathbf{U}(s) \\ &= \mathbf{c}_c^T \mathbf{X}(s) + d_c \mathbf{U}(s)\end{aligned} \tag{5.65}$$

Equation (5.65) is expressed in terms of the state $\mathbf{X}(s)$ and the input $\mathbf{U}(s)$. To obtain Eq. (5.65) from Fig. 5.30, delete all the branches having a transmittance $1/s$ and all the initial conditions in Fig. 5.30. The Mason gain formula is then used to obtain the output in terms of the state variables and the input. Note: there are two forward paths from each state variable to the output when $w = n$. If $w < n$, Eq. (5.65) becomes

$$\mathbf{Y}(s) = [c_0 \quad c_1 \quad \cdots \quad c_w \quad 0 \quad \cdots \quad 0]\mathbf{X}(s) \tag{5.66}$$

5.9 PARALLEL STATE DIAGRAMS FROM TRANSFER FUNCTIONS

A single-input single-output (SISO) system represented by the differential equation (5.62) may be represented by an overall transfer function of the form

$$\frac{Y(s)}{U(s)} = G(s) = \frac{c_w s^w + c_{w-1} s^{w-1} + \cdots + c_1 s + c_0}{s^n + a_{n-1} s^{n-1} + \cdots + a_1 s + a_0} \qquad w \le n \tag{5.67}$$

An alternate method of determining a simulation diagram and a set of state variables is to factor the denominator and to express $G(s)$ in partial fractions. When there are no repeated roots and $w = n$, the form is

$$G(s) = \frac{c_n s^n + c_{n-1} s^{n-1} + \cdots + c_1 s + c_0}{(s-\lambda_1)(s-\lambda_2)\cdots(s-\lambda_n)}$$
$$= c_n + \frac{f_1}{s-\lambda_1} + \frac{f_2}{s-\lambda_2} + \cdots + \frac{f_n}{s-\lambda_n} = c_n + \sum_{i=1}^{n} G_i(s)$$

where each

$$G_i(s) = \frac{f_i Z_i(s)}{U(s)} = \frac{f_i}{s-\lambda_i}$$

The symbols z_i and their Laplace transforms $Z_i(s)$ are used for the state variables in order to distinguish the form of the associated diagonal matrix. The output $Y(s)$ produced by the input $U(s)$ is therefore

$$Y(s) = c_n U(s) + \frac{f_1 U(s)}{s-\lambda_1} + \frac{f_2 U(s)}{s-\lambda_2} + \cdots$$
$$= c_n U(s) + f_1 Z_1(s) + f_2 Z_2(s) + \cdots \tag{5.68}$$

The state variables $Z_i(s)$ are selected to satisfy this equation. Each fraction represents a first-order differential equation of the form

$$\dot{z}_i - \lambda_i z_i = u \tag{5.69}$$

This expression can be simulated by an integrator with a feedback path of gain equal to λ_i, followed by a gain f_i. Therefore, the complete simulation diagram is drawn in Fig. 5.31. The reader may add the initial conditions of the states, $z_i(t_0)$, to Fig. 5.31. The $z_i(t_0)$ are the inputs to additional branches having transmittances of value $1/s$ and terminating at the nodes $Z_i(s)$. The term $c_n U(s)$ in Eq. (5.68) is satisfied in Fig. 5.31 by the feedforward path of gain c_n. This term appears only when the numerator and denominator of $G(s)$ have the same degree. The output of each integrator is defined as a state variable. The state equations are therefore of the form

$$\dot{\mathbf{z}} = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix} \mathbf{z} + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \mathbf{u} = \mathbf{\Lambda z} + \mathbf{b}_n \mathbf{u} \tag{5.70}$$

$$\mathbf{y} = [\,f_1 \quad f_2 \quad \cdots \quad f_n]\mathbf{z} + c_n \mathbf{u} = c_n^T \mathbf{z} + \mathbf{d}_n \mathbf{u} \tag{5.71}$$

An important feature of these equations is that the **A** matrix appears in diagonal or *normal or canonical form*. This diagonal form is indicated by

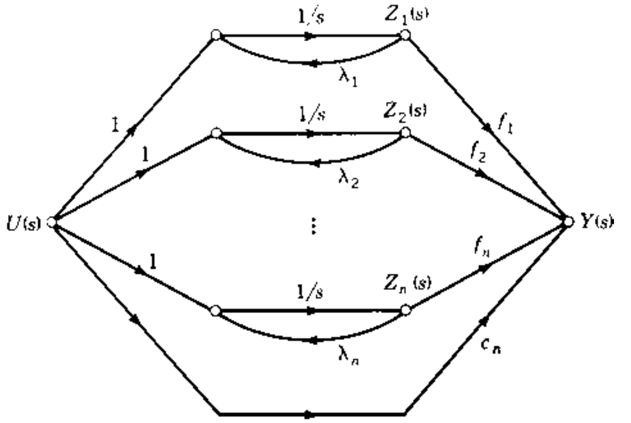


FIGURE 5.31 Simulation of Eq. (5.67) by parallel decomposition for $w = n$ (Jordan diagram for distinct roots).

$\mathbf{\Lambda}$ (or $\mathbf{A}^*$), and the corresponding state variables are often called *canonical variables.* The elements of the **B** vector (indicated by $\mathbf{b}_n$) are all unity, and the elements of the **C** row matrix (indicated by $[f_1 \quad f_2 \quad \dots \quad f_n] = \mathbf{c}_n^T$) are the coefficients of the partial fractions in Eq. (5.68). The state diagram of Fig. 5.31 and the state equation of Eq. (5.70) represent the Jordan form for the system with distinct eigenvalues. The state and output equations for a multiple-input multiple-output (MIMO) system can be expressed as

$$\dot{\mathbf{z}} = \mathbf{\Lambda}\mathbf{z} + \mathbf{B}_n\mathbf{u} \tag{5.72}$$

$$\mathbf{y} = \mathbf{C}_n\mathbf{z} + \mathbf{D}_n\mathbf{u} \tag{5.73}$$

The diagonal matrix $\mathbf{A} = \mathbf{\Lambda}$ means that each state equation is *uncoupled;* i.e., each state z_i can be solved independently of the other states. This fact simplifies the procedure for finding the state transition matrix $\mathbf{\Phi}(t)$. This form is also useful for studying the observability and controllability of a system, as discussed in Chap. 13. When the state equations are expressed in normal form, the state variables are often denoted by z_i. Transfer functions with repeated roots are not encountered very frequently and are not considered in this text [3].

Example. For the given transfer function $G(s)$, draw the parallel state diagram and determine the state equation. A partial-fraction expansion is performed on $G(s)$, and the state diagram is drawn in Fig. 5.32. Then the state and output equations are written.

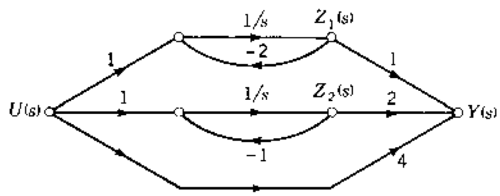


FIGURE 5.32 Simulation diagram for example of Sec. 5.9.

$$G(s) = \frac{4s^2 + 15s + 13}{s^2 + 3s + 2} = 4 + \frac{1}{s+2} + \frac{2}{s+1}$$

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

$$\mathbf{y}(t) = [1 \quad 2]\, \mathbf{z}(t) + 4\mathbf{u}(t)$$

5.10 DIAGONALIZING THE A MATRIX [11,12]

In Sec. 5.9 the method of partial-fraction expansion is shown to lead to the desirable normal form of the state equation, in which the A matrix is diagonal. The partial-fraction method is not convenient for multiple-input multiple-output systems or when the system equations are already given in state form. Therefore, this section presents a more general method for converting the state equation by means of a linear *similarity transformation*. Since the state variables are not unique, the intention is to transform the state vector **x** to a new state vector **z** by means of a constant, square, nonsingular transformation matrix **T** so that

$$\mathbf{x} = \mathbf{T}\mathbf{z} \tag{5.74}$$

Since **T** is a constant matrix, the differentiation of this equation yields

$$\dot{\mathbf{x}} = \mathbf{T}\dot{\mathbf{z}} \tag{5.75}$$

Substituting these values into the state equation $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$ produces

$$\mathbf{T}\dot{\mathbf{z}} = \mathbf{A}\mathbf{T}\mathbf{z} + \mathbf{B}\mathbf{u} \tag{5.76}$$

Premultiplying by $\mathbf{T}^{-1}$ gives

$$\dot{\mathbf{z}} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}\mathbf{z} + \mathbf{T}^{-1}\mathbf{B}\mathbf{u} \tag{5.77}$$

The corresponding output equation is

$$\mathbf{y} = \mathbf{C}\mathbf{T}\mathbf{z} + \mathbf{D}\mathbf{u} \tag{5.78}$$

It is easily shown that the eigenvalues are the same for both the original and the transformed equations. From Eq. (5.77) the new characteristic equation is

$$\left|\lambda\mathbf{I} - \mathbf{T}^{-1}\mathbf{A}\mathbf{T}\right| = 0 \tag{5.79}$$

In this equation the unit matrix can be replaced by $\mathbf{I} = \mathbf{T}^{-1}\mathbf{T}$. Then, after prefactoring $\mathbf{T}^{-1}$ and postfactoring $\mathbf{T}$, the characteristic equation becomes

$$\left|\mathbf{T}^{-1}(\lambda\mathbf{I} - \mathbf{A})\mathbf{T}\right| = 0 \tag{5.80}$$

Using the property (Appendix B) that the determinant of the product of matrices is equal to the product of the determinants of the individual matrices, Eq. (5.80) is equal to

$$|\mathbf{T}^{-1}| \cdot |\lambda\mathbf{I} - \mathbf{A}| \cdot |\mathbf{T}| = |\mathbf{T}^{-1}| \cdot |\mathbf{T}| \cdot |\lambda\mathbf{I} - \mathbf{A}| = 0$$

Since $|\mathbf{T}^{-1}| \cdot |\mathbf{T}| = |\mathbf{I}| = 1$, the characteristic equation of the transformed equation as given by Eq. (5.79) is equal to the characteristic equation of the original system, i.e.,

$$|\lambda\mathbf{I} - \mathbf{A}| = 0 \tag{5.81}$$

Therefore, it is concluded that the eigenvalues are invariant in a linear transformation given by Eq. (5.74).

The matrix $\mathbf{T}$ is called the *modal* matrix when it is selected so that $\mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ is diagonal, i.e.,

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix} \tag{5.82}$$

This supposes that the eigenvalues are distinct. With this transformation the system equations are

$$\dot{\mathbf{z}} = \boldsymbol{\Lambda}\mathbf{z} + \mathbf{B}'\mathbf{u} \tag{5.83}$$

$$\mathbf{y} = \mathbf{C}'\mathbf{z} + \mathbf{D}'\mathbf{u} \tag{5.84}$$

where

$$\boldsymbol{\Lambda} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T} \qquad \mathbf{B}' = \mathbf{T}^{-1}\mathbf{B} \qquad \mathbf{C}' = \mathbf{C}\mathbf{T} \qquad \mathbf{D}' = \mathbf{D}$$

When $\mathbf{T}$ is selected so that $\mathbf{B}' = \mathbf{T}^{-1}\mathbf{B}$ contains only unit elements, then it is denoted as $\mathbf{B}_n$ as in Eq. (5.70) and in method 3 (see following). When $\mathbf{x} = \mathbf{T}\mathbf{z}$ is used to transform the equations into any other state-variable representation

$\mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \mathbf{A}'$, then $\mathbf{T}$ is just called a **transformation matrix**. The case where $\mathbf{A}'$ is the companion matrix $\mathbf{A}_c$ is covered in Sec. 5.13.

Four methods are presented for obtaining the modal matrix $\mathbf{T}$ for the case where the matrix $\mathbf{A}$ has distinct eigenvalues. When $\mathbf{A}$ has multiple eigenvalues, this matrix can be transformed into diagonal form only if the number of independent eigenvectors is equal to the multiplicity of the eigenvalues. In such cases, the columns of $\mathbf{T}$ may be determined by methods of 2, 3, and 4.

Method 1: Matrix A in Companion Form

When there are distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ for the matrix $\mathbf{A}$ and it is in companion form ($\mathbf{A} = \mathbf{A}_c$), the Vandermonde matrix, which is easily obtained, is the modal matrix. The Vandermonde matrix is defined by

$$\mathbf{T} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \lambda_1^2 & \lambda_2^2 & \cdots & \lambda_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \cdots & \lambda_n^{n-1} \end{bmatrix} \tag{5.85}$$

Example 1. Transform the state variables in the following equation in order to uncouple the states:

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -24 & -26 & -9 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}\mathbf{u} \tag{5.86}$$

$$\mathbf{y} = \begin{bmatrix} 3 & 3 & 1 \end{bmatrix}\mathbf{x} \tag{5.87}$$

The characteristic equation $|\lambda\mathbf{I} - \mathbf{A}_c|$ yields the root, $\lambda_1 = -2$, $\lambda_2 = -3$, and $\lambda_3 = -4$. Since eigenvalues are distinct and the $\mathbf{A}$ matrix is in companion form, the Vandermonde matrix can be used as the modal matrix. Using Eq. (5.85) yields

$$\mathbf{T} = \begin{bmatrix} 1 & 1 & 1 \\ -2 & -3 & -4 \\ 4 & 9 & 16 \end{bmatrix}$$

A matrix inversion routine (using a programmable calculator or computer) yields

$$\mathbf{T}^{-1} = \frac{1}{2}\begin{bmatrix} 12 & 7 & 1 \\ -16 & -12 & -2 \\ 6 & 5 & 1 \end{bmatrix}$$

To show that this procedure does uncouple the states, the matrix system equations, in terms of the new state vector $\mathbf{z}$, are obtained from Eqs. (5.77) and (5.78) as

$$\dot{\mathbf{z}} = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -4 \end{bmatrix}\mathbf{z} + \begin{bmatrix} 7 \\ -10 \\ 4 \end{bmatrix}\mathbf{u} \tag{5.88}$$

$$\mathbf{y} = \begin{bmatrix} 1 & 3 & 7 \end{bmatrix}\mathbf{z} \tag{5.89}$$

This confirms that

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \mathbf{\Lambda} \tag{5.90}$$

Method 2: Adjoint Method

The second method does not require the $\mathbf{A}$ matrix to be in companion form. Premultiplying Eq. (5.82) by $\mathbf{T}$, where the elements of the modal matrix are defined by $\mathbf{T} = [v_{ij}]$, yields

$$\mathbf{A}\mathbf{T} = \mathbf{T}\mathbf{\Lambda} \tag{5.91}$$

The right side of this equation can be expressed in the form

$$\mathbf{T} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \mathbf{0} \\ & \lambda_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \lambda_n \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 v_{11} & \lambda_2 v_{12} & \cdots & \lambda_n v_{1n} \\ \lambda_1 v_{21} & \lambda_2 v_{22} & \cdots & \lambda_n v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 v_{n1} & \lambda_2 v_{n2} & \cdots & \lambda_n v_{nn} \end{bmatrix} = \begin{bmatrix} \lambda_1 \mathbf{v}_1 & \lambda_2 \mathbf{v}_2 & \cdots & \lambda_n \mathbf{v}_n \end{bmatrix}$$

The column vectors $\mathbf{v}_i$ of the modal matrix

$$\mathbf{T} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} \tag{5.92}$$

are called the *eigenvectors*: In terms of the $\mathbf{v}_i$, Eq. (5.91) is written as

$$\begin{bmatrix} \mathbf{A}\mathbf{v}_1 & \mathbf{A}\mathbf{v}_2 & \cdots & \mathbf{A}\mathbf{v}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n \end{bmatrix} \tag{5.93}$$

Equating columns on the left and right sides of Eq. (5.93) yields

$$\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i \tag{5.94}$$

This equation can be put in the form

$$[\lambda_i\mathbf{I} - \mathbf{A}]\mathbf{v}_i = \mathbf{0} \tag{5.95}$$

Equation (5.95) identifies that the eigenvectors* $\mathbf{v}_i$ are in the *null space* of the matrix $[\lambda_i\mathbf{I} - \mathbf{A}]$. Using the property $\mathbf{M}\ \mathrm{adj}\ \mathbf{M} = |\mathbf{M}|\mathbf{I}$ [see Eq. (B.25)] and letting $\mathbf{M} = \lambda_i\mathbf{I} - \mathbf{A}$ yields $[\lambda_i\mathbf{I} - \mathbf{A}]\ \mathrm{adj}\ [\lambda_i\mathbf{I} - \mathbf{A}] = |\lambda_i\mathbf{I} - \mathbf{A}|\mathbf{I}$. Since λ_i is an eigenvalue, the characteristic equation is $|\lambda_i\mathbf{I} - \mathbf{A}| = 0$. Therefore, this equation becomes

$$[\lambda_i\mathbf{I} - \mathbf{A}]\ \mathrm{adj}[\lambda_i\mathbf{I} - \mathbf{A}] = \mathbf{0} \tag{5.96}$$

A comparison of Eqs. (5.95) and (5.96) shows that the eigenvector

$$\mathbf{v}_i \text{ is proportional to any nonzero column of } \mathrm{adj}[\lambda_i\mathbf{I} - \mathbf{A}] \tag{5.97}$$

Since Eq. (5.77) contains $\mathbf{T}^{-1}$, a necessary condition is that $\mathbf{T}$ must be nonsingular. This is satisfied if the $\mathbf{v}_i$ are linearly independent. When the eigenvalues are distinct, Eq. (5.95) is a homogeneous equation whose rank is $n-1$, and each value of λ_i yields only one eigenvector $\mathbf{v}_i$. When the eigenvalue λ_i has multiplicity r, the *rank deficiency* α of Eq. (5.95) is $1 \le \alpha \le r$. Rank deficiency denotes that the rank of Eq. (5.95) is $n - \alpha$. Then the linearly independent eigenvectors $\mathbf{v}_i$ that satisfy Eq. (5.95) are α in number. If $\alpha < r$, then the matrix $\mathbf{A}$ cannot be diagonalized.

Example 2. Use $\mathrm{adj}[\lambda\mathbf{I} - \mathbf{A}]$ to obtain the modal matrix for the equations

$$\dot{\mathbf{x}} = \begin{bmatrix} -9 & 1 & 0 \\ -26 & 0 & 1 \\ -24 & 0 & 0 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 2 \\ 5 \\ 0 \end{bmatrix}u \tag{5.98}$$

$$Y = \begin{bmatrix} 1 & 2 & -1 \end{bmatrix}\mathbf{x} \tag{5.99}$$

The characteristic equation $|\lambda\mathbf{I} - \mathbf{A}| = 0$ yields the roots $\lambda_1 = -2$, $\lambda_2 = -3$, and $\lambda_3 = -4$. These eigenvalues are distinct, but $\mathbf{A}$ is not in companion form. Therefore, the Vandermode matrix is not the modal matrix. Using the adjoint method 2 gives

*These are sometimes called the right eigenvectors because $[\lambda_i\mathrm{I} - \mathrm{A}]$ is multiplied on the right by v_i in Eq. (5.95).

$$\mathrm{adj}[\lambda\mathbf{I}-\mathbf{A}] = \mathrm{adj}\begin{bmatrix} \lambda+9 & -1 & 0 \\ 26 & \lambda & -1 \\ 24 & 0 & \lambda \end{bmatrix}$$

$$= \begin{bmatrix} \lambda^2 & \lambda & 1 \\ -26\lambda-24 & \lambda^2+9\lambda & \lambda+9 \\ -24\lambda & -24 & \lambda^2+9\lambda+26 \end{bmatrix}$$

For $\lambda_1 = -2$: $\quad \mathrm{adj}[-2\mathbf{I}-\mathbf{A}] = \begin{bmatrix} 4 & -2 & 1 \\ 28 & -14 & 7 \\ 48 & -24 & 12 \end{bmatrix} \quad \mathbf{v}_1 = \begin{bmatrix} 1 \\ 7 \\ 12 \end{bmatrix}$

For $\lambda_2 = -3$: $\quad \mathrm{adj}[-3\mathbf{I}-\mathbf{A}] = \begin{bmatrix} 9 & -3 & 1 \\ 54 & -18 & 6 \\ 72 & -24 & 8 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 6 \\ 8 \end{bmatrix}$

For $\lambda_3 = -4$: $\quad \mathrm{adj}[-4\mathbf{I}-\mathbf{A}] = \begin{bmatrix} 16 & -4 & 1 \\ 80 & -20 & 5 \\ 96 & -24 & 6 \end{bmatrix} \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 5 \\ 6 \end{bmatrix}$

For each λ_i the columns of adj $[\lambda_i\mathbf{I}-\mathbf{A}]$ are linearly related; i.e., they are proportional. The $\mathbf{v}_i$ may be multiplied by a constant and are selected to contain the smallest integers; often the leading term is reduced to 1. In practice, it is necessary to calculate only one column of the adjoint matrix. The modal matrix is

$$\mathbf{T} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 7 & 6 & 5 \\ 12 & 8 & 6 \end{bmatrix}$$

The inverse is

$$\mathbf{T}^{-1} = -\frac{1}{2}\begin{bmatrix} -4 & 2 & -1 \\ 18 & -6 & 2 \\ -16 & 4 & -1 \end{bmatrix}$$

The matrix equations, in terms of the new state vector $\mathbf{z}$, are

$$\dot{\mathbf{z}} = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -4 \end{bmatrix}\mathbf{z} + \begin{bmatrix} -1 \\ -3 \\ 6 \end{bmatrix}\mathbf{u} = \mathbf{\Lambda}\mathbf{z} + \mathbf{B}'\mathbf{u} \tag{5.100}$$

$$\mathbf{y} = \begin{bmatrix} 3 & 5 & 5 \end{bmatrix} \mathbf{z} = \mathbf{C}'\mathbf{z} \tag{5.101}$$

Once each eigenvalue is assigned the designation λ_1, λ_2, and λ_3, they appear in this order along the diagonal of $\mathbf{\Lambda}$. Note that the systems of Examples 1 and 2 have the same eigenvalues, but different modal matrices are required to uncouple the states.

Method 3: Simultaneous Equation Method

An alternate method for evaluating the n elements of each $\mathbf{v}_i$ is to form a set of n equations from the matrix equation, Eq. (5.94). This is illustrated by the following example.

Example 3. Rework Example 2 using Eq. (5.94). The matrices formed by using the **A** matrix of Eq. (5.98) are

$$\begin{bmatrix} -9 & 1 & 0 \\ -26 & 0 & 1 \\ -24 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_{1i} \\ v_{2i} \\ v_{3i} \end{bmatrix} = \lambda_i \begin{bmatrix} v_{1i} \\ v_{2i} \\ v_{3i} \end{bmatrix} \tag{5.102}$$

Performing the multiplication yields

$$\begin{bmatrix} -9v_{1i} + v_{2i} \\ -26v_{1i} + v_{3i} \\ -24v_{1i} \end{bmatrix} = \begin{bmatrix} \lambda_i v_{1i} \\ \lambda_i v_{2i} \\ \lambda_i v_{3i} \end{bmatrix} \tag{5.103}$$

Each value of λ_i is inserted in this matrix, and the corresponding elements are equated to form three equations. For $\lambda_1 = -2$, the equations are

$$\begin{aligned} -9v_{11} + v_{21} &= -2v_{11} \\ -26v_{11} + v_{31} &= -2v_{21} \\ -24v_{11} &= -2v_{31} \end{aligned} \tag{5.104}$$

Only two of these equations are independent. This fact can be demonstrated by inserting v_{21} from the first equation and v_{31} from the third equation into the second equation. The result is the identity $v_{11} = v_{11}$. This merely confirms that $|\lambda_i \mathbf{I} = \mathbf{A}|$ is of rank $n - 1 = 2$. Since there are three elements (v_{1i}, v_{2i}, and v_{3i}) in the eigenvector $\mathbf{v}_i$ and only two independent equations are obtained from Eq. (5.102), the procedure is to arbitrarily set one of the elements equal to unity. Thus, after letting $v_{11} = 1$, these equations yield

$$\mathbf{v}_1 = \begin{bmatrix} v_{11} \\ v_{21} \\ v_{31} \end{bmatrix} = \begin{bmatrix} v_{11} \\ 7v_{11} \\ 12v_{11} \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \\ 12 \end{bmatrix}$$

This result is the same as obtained in Example 2. The procedure is repeated to evaluate $\mathbf{v}_2$ and $\mathbf{v}_3$.

The need to arbitrarily select one of the elements in each column vector v_i can be eliminated by specifying that the desired matrix $\mathbf{B}' = \mathbf{B}_n$. In Sec. 5.9 the partial fraction expansion of the transfer function leads to a column matrix $\mathbf{b}_n$ containing all 1s, that is,

$$\mathbf{b}_n = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T \tag{5.105}$$

To achieve this form for $\mathbf{b}_n$ the restriction is imposed that $\mathbf{b} = \mathbf{T}\mathbf{b}_n$. This produces n equations that are added to the n^2 equations formed by equating the elements of $\mathbf{AT} = \mathbf{T}\boldsymbol{\Lambda}$. Only n^2 of these equations are independent, and they can be solved simultaneously for the n^2 elements of $\mathbf{T}$. The restriction that $\mathbf{b} = \mathbf{T}\mathbf{b}_n$ can be satisfied only if the system is controllable (see Sec. 13.2).

Method 4: Reid's Method [13]

The eigenvectors† $\mathbf{v}_i$ that satisfy the equation

$$[\lambda_i \mathbf{I} - \mathbf{A}]\mathbf{v}_i = \mathbf{0} \tag{5.106}$$

are said to lie in the *null* space of the matrix $[\lambda_i \mathbf{I} - \mathbf{A}]$. These eigenvectors can be computed by the following procedure [13]:

1. Use elementary row operation (see Appendix B) to transform $[\lambda_i \mathbf{I} - \mathbf{A}]$ into HNF.
2. Rearrange the rows of this matrix, if necessary, so that the leading unit elements of each row appear on the principal diagonal.
3. Identify the column of this matrix that contains a zero on the principal diagonal. After replacing the zero on the principal diagonal by -1, this column vector is identified as the basis vector for the eigenvector space.

Example 4. Rework Example 2 using method 4.

$$[\lambda_i \mathbf{I} - \mathbf{A}] = \begin{bmatrix} \lambda_i + 9 & -1 & 0 \\ 26 & \lambda_i & -1 \\ 24 & 0 & \lambda_i \end{bmatrix} \tag{5.107}$$

†The eigenvector v_i are not unique, because they can be formed by any combination of vectors in the null space and can be multiplied by any nonzero constant.

For $\lambda_i = -2$ this matrix and its HNF normal form are

$$\begin{bmatrix} 7 & -1 & 0 \\ 26 & -2 & -1 \\ 24 & 0 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -\frac{1}{12} \\ 0 & 1 & -\frac{7}{12} \\ 0 & 0 & 0 \end{bmatrix}$$

where $\sim$ indicates the result of elementary operations. The eigenvector $\mathbf{v}_1$ lies in the space defined by the third column after inserting -1 on the principal diagonal. This is written as

$$\mathbf{v}_1 \in \text{span}\left\{\begin{bmatrix} -\frac{1}{12} \\ -\frac{7}{12} \\ -1 \end{bmatrix}\right\} \tag{5.108}$$

Similarly, for $\lambda_2 = -3$,

$$\begin{bmatrix} 6 & -1 & 0 \\ 26 & -3 & -1 \\ 24 & 0 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -\frac{1}{8} \\ 0 & 1 & \frac{6}{8} \\ 0 & 0 & 0 \end{bmatrix}$$

The eigenvector $\mathbf{v}_2$ lies in the space defined by the third column after inserting -1 on the principal diagonal, i.e.,

$$\mathbf{v}_2 \in \text{span}\left\{\begin{bmatrix} -\frac{1}{8} \\ -\frac{6}{8} \\ -1 \end{bmatrix}\right\} \tag{5.109}$$

Also, for $\lambda_3 = -4$,

$$\begin{bmatrix} 5 & -1 & 0 \\ 26 & -4 & -1 \\ 24 & 0 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -\frac{1}{6} \\ 0 & 1 & -\frac{5}{6} \\ 0 & 0 & 0 \end{bmatrix}$$

The eigenvector $\mathbf{v}_3$ lies in the space defined by the third column after inserting -1 on the principal diagonal, i.e.,

$$\mathbf{v}_3 \in \text{span}\left\{\begin{bmatrix} -\frac{1}{6} \\ -\frac{5}{6} \\ -1 \end{bmatrix}\right\} \tag{5.110}$$

The eigenvectors are selected from the one-dimensional spaces indicated in Eqs. (5.108) to (5.110). These vectors are linearly independent. For convenience the basic vectors can be multiplied by a negative constant such that the elements are positive integers. Therefore, one selection for the modal matrix is

$$\mathbf{T} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 7 & 6 & 5 \\ 12 & 8 & 6 \end{bmatrix} \tag{5.111}$$

This coincides with the results of method 2.

Method 5: Eigenvector Method

1. Form the matrix

$$\mathbf{S}(\lambda_i) = \left[\frac{\lambda_i \mathbf{I} - \mathbf{A}}{\mathbf{I}}\right] \tag{5.112}$$

2. Use *column* operations in order to achieve one or more columns in the form

$$\left[\begin{array}{cc} 0 & \cdots \\ 0 & \cdots \\ 0 & \cdots \\ \hline x & \cdots \\ y & \cdots \\ z & \cdots \end{array}\right] \tag{5.113}$$

3. Then the eigenvector associated with the eigenvalue λ_i is

$$\mathbf{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Example. Refer to Example 4, Eq. (5.107). With $\lambda_1 = -2$, the matrix $\mathbf{S}(-2)$ is

$$\mathbf{S}(-2) = \left[\begin{array}{ccc} 7 & -1 & 0 \\ 26 & -2 & -1 \\ 24 & 0 & -2 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] \overset{(a)}{\sim} \left[\begin{array}{ccc} 0 & -1 & 0 \\ 12 & -2 & -1 \\ 24 & 0 & -2 \\ \hline 1 & 0 & 0 \\ 7 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] \overset{(b)}{\sim} \left[\begin{array}{ccc} 0 & -1 & 0 \\ 0 & -2 & -1 \\ 0 & 0 & -2 \\ \hline \begin{bmatrix} 1 \\ 7 \\ 12 \end{bmatrix} \mathbf{v}_1 & \begin{matrix} 0 \\ 1 \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 1 \end{matrix} \end{array}\right] \tag{5.114}$$

a. Multiply column 2 by 7 and add it to column 1. This makes the first element in the first column equal to zero.

b. Multiply column 3 by 12 and add it to column 1. This makes the second element equal to zero.

c. Continue the process until the form shown is achieved.

d. Note that the eigenvector is

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 7 \\ 12 \end{bmatrix} \tag{5.115}$$

With $\lambda_2 = -3$, the matrix $\mathbf{S}(-3)$ is

$$\mathbf{S}(-3) = \left[\begin{array}{ccc} 6 & -1 & 0 \\ 26 & -3 & -1 \\ 24 & 0 & -3 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{ccc} 0 & -1 & 0 \\ 8 & -3 & -1 \\ 24 & 0 & -3 \\ \hline 1 & 0 & 0 \\ 6 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{c} \begin{matrix} 0 & \quad -1 & \quad 0 \\ 0 & \quad -3 & \quad -1 \\ 0 & \quad 0 & \quad -3 \end{matrix} \\ \hline \begin{bmatrix} 1 \\ 6 \\ 8 \end{bmatrix} \mathbf{v}_2 \quad \begin{matrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{matrix} \end{array}\right] \tag{5.116}$$

With $\lambda_3 = -4$, the matrix $\mathbf{S}(-4)$ is

$$\mathbf{S}(-4) = \left[\begin{array}{ccc} 5 & -1 & 0 \\ 26 & -4 & -1 \\ 24 & 0 & -4 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{ccc} 0 & -1 & 0 \\ 6 & -4 & -1 \\ 24 & 0 & -4 \\ \hline 1 & 0 & 0 \\ 5 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{c} \begin{matrix} 0 & \quad -1 & \quad 0 \\ 0 & \quad -4 & \quad -1 \\ 0 & \quad 0 & \quad -4 \end{matrix} \\ \hline \begin{bmatrix} 1 \\ 5 \\ 8 \end{bmatrix} \mathbf{v}_3 \quad \begin{matrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{matrix} \end{array}\right] \tag{5.117}$$

An alternate method is to leave $\mathbf{S}(\lambda)$ in terms of λ, that is

$$\mathbf{S}(\lambda) = \left[\frac{\lambda\mathbf{I} - \mathbf{A}}{\mathbf{I}}\right] = \left[\begin{array}{ccc} \lambda + 9 & -1 & 0 \\ 26 & \lambda & -1 \\ 24 & 0 & \lambda \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right]$$

Using column operations:

1. Multiply column 2 by $\lambda + 9$ and add to column 1.

$$\left[\begin{array}{ccc} 0 & -1 & 0 \\ \lambda^2+9\lambda+26 & \lambda & -1 \\ 24 & 0 & \lambda \\ \hline 1 & 0 & 0 \\ \lambda+9 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] \tag{5.118}$$

2. Multiply column 3 by $\lambda^2+9\lambda+26$ and add to column 1.

$$\left[\begin{array}{ccc} 0 & -1 & 0 \\ 0 & \lambda & -1 \\ \lambda^3+9\lambda^2+26\lambda+24 & 0 & \lambda \\ \hline \begin{bmatrix} 1 \\ \lambda+9 \\ \lambda^2+9\lambda+26 \end{bmatrix} & \begin{matrix} 0 \\ 1 \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 1 \end{matrix} \end{array}\right]$$

Note that $\lambda^3+9\lambda^2+26\lambda+24=(\lambda+2)(\lambda+3)(\lambda+4)$.

Method 6: Using MATLAB

MATLAB's *eig* function is designed to return the eigenvalues and eigenvectors of a matrix A, such that if

```
[T,D] = eig(A)
```

then

```
A * T = T * D
```

Using the A matrix from Example 2, the MATLAB solution is

```
[T,D] = eig (A)

T =

 -0.1270  0.0995  0.0718
 -0.6350  0.5970  0.5026
 -0.7620  0.7960  0.8615

D =

 -4.0000        0        0
       0  -3.0000        0
       0        0  -2.0000
```

The vectors of T are unit length versions of the eigenvectors in Eq. (5.111). The following command verifies that this transformation is correct:

```
lambda = inv (T)*A*T
```

When the matrix $\mathbf{A}$ has *multiple eigenvalues*, the number of independent eigenvectors associated with the eigenvalue may or may not be equal to the multiplicity of the eigenvalue. This property can be determined by using method 4. The number of columns of the modified HNF of $[\lambda_i \mathbf{I} - \mathbf{A}]$ containing zeros on the principal diagonal may be equal to r, the multiplicity of λ_i. When this occurs, the independent eigenvectors are used to form the modal matrix $\mathbf{T}$, with the result that $\mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \boldsymbol{\Lambda}$ is a diagonal matrix.

5.11 USE OF STATE TRANSFORMATION FOR THE STATE EQUATION SOLUTION

The usefulness of the transformation of a matrix to diagonal form is now illustrated in solving the state equation. In Sec. 3.15 the state equation and its solution are given, respectively, by

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \tag{5.119}$$

and

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)\,d\tau \tag{5.120}$$

The evaluation of the state transition matrix $e^{\mathbf{A}t}$ is computed in Sec. 3.14 by using the Cayley-Hamilton theorem. This procedure is simplified by first expressing the matrix $\mathbf{A}$ in terms of the diagonalized matrix $\boldsymbol{\Lambda}$ and the modal matrix $\mathbf{T}$. Thus, from the development in Sec. 5.10,

$$\mathbf{A} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^{-1} \tag{5.121}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix given in Eq. (5.82) and $\mathbf{T}$ is the modal matrix given in Eq. (5.92). Using the series representation of $e^{\mathbf{A}t}$ given in Eq. (3.73), and then using Eq. (5.121), yields

$$\begin{aligned} e^{\mathbf{A}t} &= \mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2 t^2}{2!} + \frac{\mathbf{A}^3 t^3}{3!} + \cdots \\ &= \mathbf{I} + (\mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^{-1})t + \frac{(\mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^{-1})^2 t^2}{2!} + \frac{(\mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^{-1})^3 t^3}{3!} + \cdots \\ &= \mathbf{T}\left(\mathbf{I} + \boldsymbol{\Lambda}t + \frac{\boldsymbol{\Lambda}^2 t^2}{2!} + \frac{\boldsymbol{\Lambda}^3 t^3}{3!} + \cdots\right)\mathbf{T}^{-1} = \mathbf{T}e^{\boldsymbol{\Lambda}t}\mathbf{T}^{-1} \end{aligned} \tag{5.122}$$

Since Λ is a diagonal matrix, the matrix $e^{\Lambda t}$ is also diagonal and has the form

$$e^{\Lambda t} = \begin{bmatrix} e^{\lambda_1 t} & & & 0 \\ & e^{\lambda_2 t} & & \\ & & \ddots & \\ 0 & & & e^{\lambda_n t} \end{bmatrix} \tag{5.123}$$

The work of finding the modal matrix $\mathbf{T}$ is therefore balanced by the ease of obtaining $e^{\Lambda t}$. This is illustrated in the following example.

Example. Evaluate $e^{\mathbf{A}t}$ for the example of Sec. 3.14. The plant matrix $\mathbf{A}$ is

$$\mathbf{A} = \begin{bmatrix} 0 & 6 \\ -1 & -5 \end{bmatrix}$$

and the eigenvalues are $\lambda_1 = -2$ and $\lambda_2 = -3$. Since $\mathbf{A}$ is not in companion form, the modal matrix is evaluated using method 2, 3, 4, or 5 of Sec. 5.10:

$$\mathbf{T} = [\mathbf{v}_1 \quad \mathbf{v}_2] = \begin{bmatrix} 3 & 2 \\ -1 & -1 \end{bmatrix}$$

Now, using Eqs. (5.122) and (5.123) yields

$$\begin{aligned} e^{\mathbf{A}t} &= \mathbf{T}e^{\Lambda t}\mathbf{T}^{-1} = \mathbf{T}\begin{bmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{bmatrix}\mathbf{T}^{-1} \\ &= \begin{bmatrix} 3 & 2 \\ -1 & -1 \end{bmatrix}\begin{bmatrix} e^{-2t} & 0 \\ 0 & e^{-3t} \end{bmatrix}\begin{bmatrix} 1 & 2 \\ -1 & -3 \end{bmatrix} \\ &= \begin{bmatrix} 3e^{-2t} - 2e^{-3t} & 6e^{-2t} - 6e^{-3t} \\ -e^{-2t} + e^{-3t} & -2e^{-2t} + 3e^{-3t} \end{bmatrix} \end{aligned}$$

This result is the same as that given in Eq. (3.92).

5.12 TRANSFORMING A MATRIX WITH COMPLEX EIGENVALUES

The transformation of variables described in Sec. 5.10 simplifies the system representation mathematically by removing any coupling between the system modes. Thus, diagonalizing the $\mathbf{A}$ matrix produces the decoupled form of system representation shown in Fig. 5.31. Each state equation corresponding to the normal form of the state equation, Eq. (5.83), has the form

$$\dot{z}_i(t) = \lambda_i z_i(t) + b_i u \tag{5.124}$$

Since only one state variable z_i appears in this equation, it can be solved independently of all the other states.

When the eigenvalues are complex, the matrix **T** that produces the diagonal matrix **Λ** contains complex numbers. This property is illustrated for the differential equation

$$\ddot{x} - 2\sigma\dot{x} + (\sigma^2 + \omega_d^2)x = u(t) \tag{5.125}$$

where $\omega_n^2 = \sigma^2 + \omega_d^2$ and the eigenvalues are $\lambda_{1,2} = \sigma \pm j\omega_d$. Using the phase variables $x_1 = x$ and $x_2 = \dot{x}_1$, the state and output equations are

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 \\ -\omega_n^2 & 2\sigma \end{bmatrix}\mathbf{x} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\mathbf{u} \tag{5.126}$$

$$\mathbf{y} = \begin{bmatrix} 1 & 0 \end{bmatrix}\mathbf{x} \tag{5.127}$$

The modes can be uncoupled, but doing so is undesirable, as shown next. Since the eigenvalues are distinct and **A** is a companion matrix, the modal matrix **T** is the Vandermonde matrix

$$\mathbf{T} = \begin{bmatrix} 1 & 1 \\ \sigma + j\omega_d & \sigma - j\omega_d \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} \tag{5.128}$$

When the transformation $\mathbf{x} = \mathbf{Tz}$ is used [14], the normal form of the state and output equations obtained from Eqs. (5.83) and (5.84) are

$$\dot{\mathbf{z}} = \begin{bmatrix} \sigma + j\omega_d & 0 \\ 0 & \sigma - j\omega_d \end{bmatrix}\mathbf{z} + \begin{bmatrix} \dfrac{1}{j2\omega_d} \\ -\dfrac{1}{j2\omega_d} \end{bmatrix}\mathbf{u} \tag{5.129}$$

$$\mathbf{y} = \begin{bmatrix} 1 & 1 \end{bmatrix}\mathbf{z} \tag{5.130}$$

The simulation diagram based on these equations is shown in Fig. 5.33. The parameters in this diagram are complex quantities, which increases the

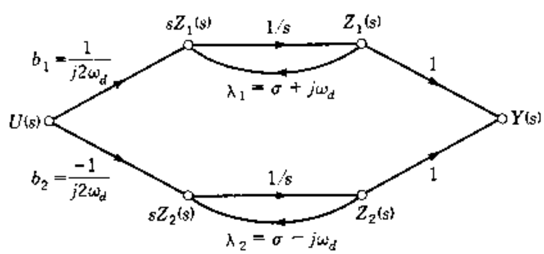


FIGURE 5.33 Simulation diagram for Eqs. (5.129) and (5.130).

difficulty when obtaining the mathematical solution. Thus, it may be desirable to perform another transformation in order to obtain a simulation diagram that contains only real quantities. The pair of complex-conjugate eigenvalues λ_1 and λ_2 jointly contribute to one transient mode that has the form of a damped sinusoid. The additional transformation is $\mathbf{z} = \mathbf{Q}\mathbf{w}$, where

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{2} & -\frac{j}{2} \\ \frac{1}{2} & \frac{j}{2} \end{bmatrix} \tag{5.131}$$

The new state and output equations are

$$\dot{\mathbf{w}} = \mathbf{Q}^{-1}\boldsymbol{\Lambda}\mathbf{Q}\mathbf{w} + \mathbf{Q}^{-1}\mathbf{T}^{-1}\mathbf{B}\mathbf{u} = \boldsymbol{\Lambda}_m\mathbf{w} + \mathbf{B}_m\mathbf{u} \tag{5.132}$$

$$\mathbf{y} = \mathbf{C}\mathbf{T}\mathbf{Q}\mathbf{w} + \mathbf{D}\mathbf{u} = \mathbf{C}_m\mathbf{w} + \mathbf{D}_m\mathbf{u} \tag{5.133}$$

For this example these equations [5]* are in the modified canonical form

$$\dot{\mathbf{w}} = \begin{bmatrix} \sigma & \omega_d \\ -\omega_d & \sigma \end{bmatrix}\mathbf{w} + \begin{bmatrix} 0 \\ 1/\omega_d \end{bmatrix}\mathbf{u} \tag{5.134}$$

$$\mathbf{y} = \begin{bmatrix} 1 & 0 \end{bmatrix}\mathbf{w} \tag{5.135}$$

The simulation diagram for these equations is shown in Fig. 5.34. Note that only real quantities appear. Also, the two modes have not been isolated, which is an advantage for complex-conjugate roots.

The effect of the two transformations $\mathbf{x} = \mathbf{T}\mathbf{z}$ and $\mathbf{z} = \mathbf{Q}\mathbf{w}$ can be considered as one transformation given by

$$\mathbf{x} = \mathbf{T}\mathbf{Q}\mathbf{w} = \begin{bmatrix} 1 & 0 \\ \sigma & \omega_d \end{bmatrix}\mathbf{w} = \mathbf{T}_m\mathbf{w} \tag{5.136}$$

Comparison of the modified modal matrix $\mathbf{T}_m$ in Eq. (5.136) with the original matrix in Eq. (5.128) shows that

$$\mathbf{T}_m = \begin{bmatrix} \mathrm{Re}(\mathbf{v}_1) & \mathrm{Im}(\mathbf{v}_1) \end{bmatrix} \tag{5.137}$$

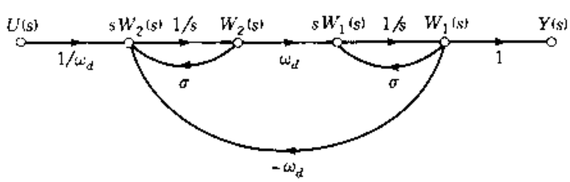


FIGURE 5.34 Simulation diagram for Eqs. (5.134) and (5.135).

---

*Although the trajectories in the state space are not drawn here, the result of this transformation produces trajectories that are symmetrical with respect to the w axes.

Therefore, $\mathbf{T}_m$ can be obtained directly from $\mathbf{T}$ without using the $\mathbf{Q}$ transformation.

In the general case containing both complex and real eigenvalues the modified block diagonal matrix $\mathbf{\Lambda}_m$ and the resulting matrices $\mathbf{B}_m$ and $\mathbf{C}_m$ have the form

$$\mathbf{\Lambda}_m = \left[\begin{array}{cc:cc:ccc} \sigma_1 & \omega_1 & & & & & \mathbf{0} \\ -\omega_1 & \sigma_1 & & & & & \\ \hdashline & & \sigma_2 & \omega_2 & & & \\ & & -\omega_2 & \sigma_2 & & & \\ \hdashline & & & & \lambda & & \\ & & & & & \ddots & \\ & \mathbf{0} & & & & & \lambda_n \end{array}\right] \qquad \mathbf{B}_m = \mathbf{T}_m^{-1}\mathbf{B} \tag{5.138}$$

$$\mathbf{C}_m = \mathbf{C}\mathbf{T}_m \tag{5.139}$$

With this modified matrix $\mathbf{\Lambda}_m$ the two oscillating modes produced by the two sets of complex eigenvalues $\sigma_1 \pm j\omega_1$ and $\sigma_2 \pm j\omega_2$ are uncoupled. The column vectors contained in $\mathbf{T}$ for the conjugate eigenvalues are also conjugates. Thus

$$\mathbf{T} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_1^* & \mathbf{v}_3 & \mathbf{v}_3^* & \mathbf{v}_5 & \cdots & \mathbf{v}_n \end{bmatrix} \tag{5.140}$$

The modified modal matrix $\mathbf{T}_m$ can then be obtained directly by using

$$T_m = \begin{bmatrix} \mathrm{Re}\,(\mathbf{v})_1 & \mathrm{Im}\,(\mathbf{v}_1) & \mathrm{Re}\,(\mathbf{v}_3) & \mathrm{Im}\,(\mathbf{v}_3) & \mathbf{v}_5 & \cdots & \mathbf{v}_n \end{bmatrix} \tag{5.141}$$

5.13 TRANSFORMING AN A MATRIX INTO COMPANION FORM

The synthesis of feedback control systems and the analysis of their response characteristics is often considerably facilitated when the plant and control matrices in the matrix state equation have particular forms. The transformation from a physical or a general state variable $\mathbf{x}_p$ to the phase variable $\mathbf{x}$ is presented in this section [3,14]. In this case the state equation in terms of the general state variable is

$$\dot{\mathbf{x}}_p = \mathbf{A}_p\mathbf{x}_p + \mathbf{b}_p\mathbf{u} \tag{5.142}$$

and the transformed state equation in terms of the phase variables is

$$\dot{\mathbf{x}} = \mathbf{A}_c\mathbf{x} + \mathbf{b}_c\mathbf{u} \tag{5.143}$$

where the companion matrix $\mathbf{A}_c$ and the vector $\mathbf{b}_c$ have the respective forms

$$\mathbf{A}_c = \begin{bmatrix} 0 & 1 & & & & 0 \\ & 0 & 1 & & & \\ & & 0 & \ddots & & \\ & & & \ddots & & \\ 0 & & & \cdots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-2} & -a_{n-1} \end{bmatrix} \qquad b_c = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \tag{5.144}$$

The state equation given by Eq. (5.143) is often described in the literature as being in the *control canonical form*. The general state and phase variables are related by

$$\mathbf{x}_p = \mathbf{T}\mathbf{x} \tag{5.145}$$

so that

$$\mathbf{A}_c = \mathbf{T}^{-1}\mathbf{A}_p\mathbf{T} \qquad \mathbf{b}_c = \mathbf{T}^{-1}\mathbf{b}_p \tag{5.146}$$

Since the characteristic polynomial is invariant under a similarity transformation, it can be obtained from $\mathbf{Q}(\lambda) = |\lambda\mathbf{I} - \mathbf{A}_p|$ or by the method of Sec. 4.16 and has the form

$$\mathbf{Q}(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 \tag{5.147}$$

Thus, by using the coefficients of Eq. (5.147), the companion matrix can readily be formed, placing the negatives of the coefficients along the bottom row of $\mathbf{A}_c$.

The controllability matrix [see Eq. (13.10)] for the phase-variable state equation is

$$\begin{aligned} \mathbf{M}_{cc} &= \begin{bmatrix} \mathbf{b}_c & \mathbf{A}_c\mathbf{b}_c & \cdots & \mathbf{A}_c^{n-1}\mathbf{b}_c \end{bmatrix} \\ &= \begin{bmatrix} & \mathbf{0} & & & & 1 \\ & & & & \iddots & \\ & & & 0 & & \\ & & 0 & 1 & \cdots & \\ & 0 & 1 & -a_{n-1} & \cdots & \\ 0 & 1 & -a_{n-1} & a_{n-1}^2 - a_{n-2} & \cdots & \\ 1 & -a_{n-1} & a_{n-1}^2 - a_{n-2} & -a_{n-1}^3 + 2a_{n-1}a_{n-2} - a_{n-3} & \cdots & \end{bmatrix} \end{aligned} \tag{5.148}$$

The matrix $\mathbf{M}_{cc}$* in Eq. (5.148) can be formed since the coefficients of the characteristic polynomial $\mathbf{Q}(\lambda)$ are known. The matrix $\mathbf{M}_{cc}$ expressed in terms of the transformed matrices of Eq. (5.146) is

$$\begin{aligned}\mathbf{M}_{cc} &= \begin{bmatrix} \mathbf{T}^{-1}\mathbf{b}_p & \mathbf{T}^{-1}\mathbf{A}_p\mathbf{b}_p & \mathbf{T}^{-1}\mathbf{A}_p^2\mathbf{b}_p & \cdots & \mathbf{T}^{-1}\mathbf{A}_p^{n-1}\mathbf{b}_p \end{bmatrix} \\ &= \mathbf{T}^{-1}\begin{bmatrix} \mathbf{b}_p & \mathbf{A}_p\mathbf{b}_p & \mathbf{A}_p^2\mathbf{b}_p & \cdots & \mathbf{A}_p^{n-1}\mathbf{b}_p \end{bmatrix} = \mathbf{T}^{-1}\mathbf{M}_{cp} \end{aligned} \tag{5.149}$$

The matrix $\mathbf{M}_{c_p}$ is obtained by using $\mathbf{b}_p$ and $\mathbf{A}_p$ of the general state-variable equation (5.142). Equation (5.149) yields

$$\mathbf{T} = \mathbf{M}_{cp}\mathbf{M}_{cc}^{-1} \tag{5.150}$$

An alternative procedure is to determine $\mathbf{T}^{-1}$ as follows:

1. Form the augmented matrix $[\mathbf{M}_{cp} \quad \mathbf{I}]$.
2. Perform elementary row operations (see Sec. 4.16) until this matrix has the form $[\mathbf{M}_{cc} \quad \mathbf{T}^{-1}]$.
3. The right half of this matrix identifies the required transformation matrix.

Example. A general state-variable equation contains the matrices

$$\mathbf{A}_p = \begin{bmatrix} 1 & 6 & -3 \\ -1 & -1 & 1 \\ -2 & 2 & 0 \end{bmatrix} \qquad \mathbf{b}_p = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \tag{5.151}$$

for which the characteristic polynomial is evaluated in Sec. 4.16 as

$$\mathbf{Q}(\lambda) = \lambda^3 - 3\lambda + 2 = (\lambda - 1)^2(\lambda + 2) \tag{5.152}$$

The matrix $\mathbf{M}_{cc}$ obtained in accordance with Eq. (5.148) is

$$\mathbf{M}_{cc} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 3 \end{bmatrix} \tag{5.153}$$

The augmented matrix obtained in accordance with step 1 of the procedure is

$$\begin{bmatrix} \mathbf{M}_{cp} & \mathbf{I} \end{bmatrix} = \left[\begin{array}{ccc:ccc} 1 & 4 & -2 & 1 & 0 & 0 \\ 1 & -1 & -3 & 0 & 1 & 0 \\ 1 & 0 & -10 & 0 & 0 & 1 \end{array}\right] \tag{5.154}$$

*The matrix $\mathbf{M}_{cc}$ has *striped* form, that is, the elements constituting the secondary diagonal (from the lower left to the upper right) are all equal. The same is true of all elements in the lines parallel to it.

Performing row operations until this matrix has the form $[\mathbf{M}_{cc} \quad \mathbf{T}^{-1}]$ yields

$$\begin{bmatrix} 0 & 0 & 1 & \vdots & 1/36 & 4/36 & -5/36 \\ 0 & 1 & 0 & \vdots & 7/36 & -8/36 & 1/36 \\ 1 & 0 & 3 & \vdots & 13/36 & 52/36 & -29/36 \end{bmatrix} = \begin{bmatrix} \mathbf{M}_{cc} & \mathbf{T}^{-1} \end{bmatrix} \tag{5.155}$$

The matrix $\mathbf{T}^{-1}$ from Eq. (5.155) yields the transformation matrix

$$\mathbf{T} = \begin{bmatrix} -5 & 4 & 1 \\ -6 & -1 & 1 \\ -13 & 0 & 1 \end{bmatrix} \tag{5.156}$$

which produces the desired forms of the matrices

$$\mathbf{A}_c = \mathbf{T}^{-1}\mathbf{A}_p\mathbf{T} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & 3 & 0 \end{bmatrix} \tag{5.157}$$

and

$$\mathbf{b}_c = \mathbf{T}^{-1}\mathbf{b}_p \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{5.158}$$

The transformation from the general state variable $\mathbf{x}_p$ to the phase variable $\mathbf{x}$ can also be accomplished by first obtaining the transfer function $G(s) = \mathbf{c}_p^T[s\mathbf{I} - \mathbf{A}_p]^{-1}\mathbf{b}_p$. Then the state and output equations can be obtained directly by using Eqs. (5.38) and (5.44), respectively. The advantage of the method presented in this section is that it avoids having to obtain the inverse of the *polynomial matrix* $[s\mathbf{I} - \mathbf{A}_p]^{-1}$.

5.14 USING MATLAB TO OBTAIN THE COMPANION A MATRIX

The example in Sec. 5.13 can be solved for the correct transformation matrix **T** by using the MATLAB CAD package. Note the MATLAB built-in function *canon* assumes a different definition of companion form. The MATLAB default companion form may be called right companion form as the coefficients of the characteristic equation are along the right column. Using the *canon* function on the matrices from Eq. (5.151) produces the following results:

```
[Ac, Bc, Cc, Dc, T] = canon (A, B, C, D, 'companion')
  Ac =

    -0.0000  0.0000  -2.0000
     1.0000       0   3.0000
    -0.0000  1.0000  -0.0000

  Bc =

    1
    0
    0

  Cc =

    3  3  -15

  Dc =
    0
  T =

    0.2778   1.1111  -0.3889
    0.1944  -0.2222   0.0278
    0.0278   0.1111  -0.1389
```

Alternatively, a MATLAB user defined function can be written to calculate the bottom companion form described in Sec. 5.13. User defined functions are m-files with predefined inputs and outputs (See App. C). The following user defined function *botcomp* is based upon the logic of the *canon* function:

```
function [Ac, Bc, Cc, Dc, T] = botcomp (A,B,C,D)
%
%  BOTCOMP Compute a bottom companion form canonical state-space
%  realization.
%
%  [Ac,Bc,Cc,Dc,T] = botcomp (A,B,C,D) computes the canonical
%  matrices for a bottom companion form realization and
%  returns the state transformation matrix T relating the
%  new state vector z to the old state vector x by z = Tx.
%
%  This algorithm uses the method from D'Azzo and Houpis
%  Sec. 5.13.
%
%  Stuart N. Sheldon
%  Revision: 0
%
```

```
Mcp=ctrb (A, B);            %  Form the controllability matrix
                            %  (physical form)
                            %  to use the transformation
                            %  Eq (5.141)
Ac=(Mcp\A*Mcp);             %  Use Mcp to generate the Companion
                            %  form State Matrix
Bc=zeros (length (A), 1);   %  Create an input matrix in
                            %  companion form
Bc (length (A), 1)=1;
Mcc=ctrb (Ac, Bc);          %  Compute the companion form
                            %  controllability matrix
T=Mcp* inv (Mcc);           %  Compute a transformation matrix
                            %  using Eq.(5.143)
Ac=inv (T) *A*T;            %  Return the companion form state
                            %  matrix
Bc=inv (T) *B;              %  Return the companion form
                            %  control input matrix
Cc=C*T;                     %  Return the companion form output
                            %  matrix
%  end botcomp
```

Using the *botcomp* function on the matrices from Eq. (5.151) produces the following results:

```
[Ac, Bc, Cc, Dc, T]=botcomp (A, B, C, D)
Ac =

    0.0000  1.0000  -0.0000
         0  0.0000   1.0000
   -2.0000  3.0000        0

Bc =

   0
   0
   1

Cc =

   -24.0000  3.0000  3.0000

Dc =
  0
```

```
T =

    -5.0000    4.0000  1.0000
    -6.0000   -1.0000  1.0000
   -13.0000   -0.0000  1.0000
```

Note that these results agree with Eqs. (5.156) and (5.157).

5.15 SUMMARY

The foundation has been completed in Chaps. 2 to 5 for analyzing the performance of feedback control systems. The analysis of such systems and the adjustments of parameters to achieve the desired performance are covered in the following chapters. In order to provide clarity in understanding the principles and to avoid ambiguity, the systems considered in Chaps. 6 to 16 are restricted primarily to single-input, single-output (SISO) systems. The procedures can be extended to multiple-input multiple-output (MIMO) systems, and some synthesis techniques for multivariable systems are presented in Refs. 5 and 6. The block diagram representation containing a single feedback loop is the basic unit used to develop the various techniques of analysis.

Since state variables are not unique, there are several ways of selecting them. One method is the use of physical variables, as in Chap. 2. The selection of the state variables as phase or canonical variables is demonstrated in this chapter through the medium of the simulation diagram. This diagram is similar to an analog-computer diagram and uses integrators, amplifiers, and summers. The outputs of the integrators in the simulation diagram are defined as the state variables.

Preparation for overall system analysis has been accomplished by writing the conventional differential and state equations in Chap. 2, obtaining the solution of these equations in Chap. 3, introducing the Laplace transform and MATLAB to facilitate the design procedure in Chap. 4, and incorporating this into the representation of a complete system in Chap. 5. Matrix methods are also presented in App. B in preparation for system synthesis via the use of modern control theory.

REFERENCES

1. Etkin, B.; *Dynamics of Atmospheric Flight*, John Wiley and Sons, New York, 1972.
2. *IEEE Standard Dictionary of Electrical and Electronics Terms*, Wiley-Interscience, New York, 1972.
3. DeRusso, P. M., R. J. Roy, and C. M. Close: *State Variables for Engineers*, Krieger, Malabar, Fla., 1990.

4. Wiberg, D. M.: *State Space and Linear System*, Schaum's Outline Series, McGraw-Hill, New York, 1971.
5. D'Azzo, J. J., and C. H. Houpis: *Linear Control System Analysis and Design: Conventional and Modern*, 4th ed., McGraw-Hill, New York, 1995.
6. Houpis, C. H., and S. J. Rasmussen: *Quantitative Feedback Theory: Fundamentals and Applications*, Marcel Dekker, New York, 1999.
7. Kailath, T.: *Linear Systems*, Prentice Hall, Englewood Cliffs, N.J., 1980.
8. Mason, S. J.: "Feedback Theory: Further Properties of Signal Flow Graphs," *Proc. IRE*, vol. 44, no. 7, pp. 920–926, July 1956.
9. Mason, S. J., and H. J. Zimmerman: *Electronic Circuits, Signals, and Systems*, The M.I.T. Press, Cambridge, Mass., 1960.
10. Kuo, B. C.: *Linear Networks and Systems*, McGraw-Hill, New York, 1967.
11. Trimmer, J. D.: *Response of Physical Systems*, John Wiley and Sons, New York, 1950, p. 17.
12. Langholz, G., and S. Frankenthal: "Reduction to Normal Form of a State Equation in the Presence of Input Derivatives," *Intern. J. Systems Science*, vol. 5, no. 7, pp. 705–706, July 1974.
13. Reid, J. G.: *Linear System Fundamentals: Continuous and Discrete, Classic and Modern*, McGraw-Hill, New York, 1983.
14. Crossley, T. R., and B. Porter: "Inversion of Complex Matrices," *Electronics Letters*, vol. 6, no. 4, pp. 90–91, February 1970.
15. Rugh, W. J.: *Linear System Theory*, 2nd ed., Prentice-Hall, Englewood Cliffs, N.J., 1993.

6

Control-System Characteristics

6.1 INTRODUCTION

Open- and closed-loop transfer functions have certain basic characteristics that permit transient and steady-state analyses of the feedback-controlled system. Five factors of prime importance in feedback-control systems are *stability, the existence and magnitude of the steady-state error*, *controllability, observability*, *and parameter sensitivity.* The stability characteristic of a linear time-invariant system is determined from the system's characteristic equation. Routh's stability criterion provides a means for determining stability without evaluating the roots of this equation. The steady-state characteristics are obtainable from the open-loop transfer function for unity-feedback systems (or equivalent unity-feedback systems), yielding figures of merit and a ready means for classifying systems. The first two properties are developed in this chapter. The remaining properties are presented in Chaps. 12 through 14. Computer-aided-design (CAD) programs like MATLAB [1] (Appendix C) and TOTAL-PC (Appendix D) are available to assist in the determination of these system characteristics.

6.2 ROUTH'S STABILITY CRITERION [2–5]

The response transform $X_2(s)$ has the general form given by Eq. (4.32), which is repeated here in slightly modified form. $X_1(s)$ is the driving transform.

$$\begin{aligned} X_2(s) &= \frac{P(s)}{Q(s)} X_1(s) \\ &= \frac{P(s)X_1(s)}{b_n s^n + b_{n-1}s^{n-1} + b_{n-2}s^{n-2} + \cdots + b_1 s + b_0} \end{aligned} \tag{6.1}$$

Sections 4.6, 4.7, and 4.10 describe the methods used to evaluate the inverse transform $\mathscr{L}^{-1}[F(s)] = f(t)$. However, before the inverse transformation can be performed, the characteristic polynomial $Q(s)$ must be factored. CAD programs are readily available for obtaining the roots of a polynomial [1,6]. Section 4.13 shows that stability of the response $x_2(t)$ requires that all zeros of $Q(s)$ have negative real parts. Since it is usually not necessary to find the exact solution when the response is unstable, a simple procedure to determine the existence of zeros with positive real parts is needed. If such zeros of $Q(s)$ with positive real parts are found, the system is unstable and must be modified. Routh's criterion is a simple method of determining the number of zeros with positive real parts without actually solving for the zeros of $Q(s)$. Note that zeros of $Q(s)$ are poles of $X_2(s)$. The characteristic equation is

$$Q(s) = b_n s^n + b_{n-1}s^{n-1} + b_{s-2}s^{n-2} + \cdots + b_1 s + b_0 = 0 \tag{6.2}$$

If the b_0 term is zero, divide by s to obtain the equation in the form of Eq. (6.2). The b's are real coefficients, and all powers of s from s^n to s^0 must be present in the characteristic equation. A necessary but not sufficient condition for stable roots is that all the coefficients in Eq. (6.2) must be positive. If any coefficients other than b_0 are zero, or if all the coefficients do not have the same sign, then there are pure imaginary roots or roots with positive real parts and the system is unstable. In that case it is unnecessary to continue if only stability or instability is to be determined. When all the coefficients are present and positive, the system may or may not be stable because there still may be roots on the imaginary axis or in the right-half s plane. Routh's criterion is mainly used to determine stability. In special situations it may be necessary to determine the actual number of roots in the right-half s plane. For these situations the procedure described in this section can be used.

The coefficients of the characteristic equation are arranged in the pattern shown in the first two rows of the following Routhian array. These coefficients are then used to evaluate the rest of the constants to complete the array.

$$\begin{array}{l|lllll}
s^n & b_n & b_{n-2} & b_{n-4} & b_{n-6} & \cdots \\
s^{n-1} & b_{n-1} & b_{n-3} & b_{n-5} & b_{n-7} & \cdots \\
s^{n-2} & c_1 & c_2 & c_3 & \cdots & \\
s^{n-3} & d_1 & d_2 & \cdots & & \\
\cdots & \cdots & \cdots & & & \\
s^1 & j_1 & & & & \\
s^0 & k_1 & & & &
\end{array}$$

The constants $c_1, c_2, c_3,$ etc., in the third row are evaluated as follows:

$$c_1 = \frac{b_{n-1}b_{n-2} - b_n b_{n-3}}{b_{n-1}} \tag{6.3}$$

$$c_2 = \frac{b_{n-1}b_{n-4} - b_n b_{n-5}}{b_{n-1}} \tag{6.4}$$

$$c_3 = \frac{b_{n-1}b_{n-6} - b_n b_{n-7}}{b_{n-1}} \tag{6.5}$$

This pattern is continued until the rest of the c's are all equal to zero. Then the d row is formed by using the s^{n-1} and s^{n-2} rows. The constants are

$$d_1 = \frac{c_1 b_{n-3} - b_{n-1}c_2}{c_1} \tag{6.6}$$

$$d_2 = \frac{c_1 b_{n-5} - b_{n-1}c_3}{c_1} \tag{6.7}$$

$$d_3 = \frac{c_1 b_{n-7} - b_{n-1}c_4}{c_1} \tag{6.8}$$

This process is continued until no more d terms are present. The rest of the rows are formed in this way down to the s^0 row. The complete array is triangular, ending with the s^0 row. Notice that the s^1 and s^0 rows contain only one term each. Once the array has been found, *Routh's criterion states that the number of roots of the characteristic equation with positive real parts is equal to the number of changes of sign of the coefficients in the first column.* Therefore, the system is stable if all terms in the first column have the same sign.*

The following example illustrates this criterion:

$$Q(s) = s^5 + s^4 + 10s^3 + 72s^2 + 152s + 240 \tag{6.9}$$

*Reference 7 shows that a system is unstable if there is a negative element in any position in any row. Thus, it is not necessary to complete the Routhian array if any negative number is encountered.

The Routhian array is formed by using the procedure described above:

$$\begin{array}{c|lll}
s^5 & 1 & 10 & 152 \\
s^4 & 1 & 72 & 240 \\
s^3 & -62 & -88 & \\
s^2 & 70.6 & 240 & \\
s^1 & 122.6 & & \\
s^0 & 240 & &
\end{array}$$

In the first column there are two changes of sign, from 1 to –62 and from – 62 to 70.6; therefore, $Q(s)$ has two roots in the right-half s plane (RHP). Note that this criterion gives the number of roots with positive real parts but does not tell the values of the roots. If Eq. (6.9) is factored, the roots are $s_1 = -3$, $s_{2,3} = -1 \pm j\sqrt{3}$, and $s_{4,5} = +2 \pm j4$. This calculation confirms that there are two roots with positive real parts. The Routh criterion does not distinguish between real and complex roots.

Theorem 1: Division of a Row. The coefficients of any row may be multiplied or divided by a positive number without changing the signs of the first column. The labor of evaluating the coefficients in Routh's array can be reduced by multiplying or dividing any row by a constant. This may result, for example, in reducing the size of the coefficients and therefore simplifying the evaluation of the remaining coefficients.

The following example illustrates this theorem:

$$Q(s) = s^6 + 3s^5 + 2s^4 + 9s^3 + 5s^2 + 12s + 20 \tag{6.10}$$

The Routhian array is

$$\begin{array}{c|llll}
s^6 & 1 & 2 & 5 & 20 \\
s^5 & \cancel{3} & \cancel{9} & \cancel{12} & \\
 & 1 & 3 & 4 & \text{(after dividing by 3)} \\
s^4 & -1 & 1 & 20 & \\
s^3 & \cancel{4} & \cancel{24} & & \\
 & 1 & 6 & & \text{(after dividing by 4)} \\
s^2 & 7 & 20 & & \\
s^1 & 22 & & & \text{(after multiplying by 7)} \\
s^0 & 20 & & &
\end{array}$$

Notice that the size of the numbers has been reduced by dividing the s^5 row by 3 and the s^3 row by 4. The result is unchanged; i.e., there are two changes of signs in the first column and, therefore, there are two roots with positive real parts.

Theorem 2: Zero Coefficient in the First Column. When the first term in a row is zero but not all the other terms are zero, the following methods can be used:

1. Substitute $s = 1/x$ in the original equation; then solve for the roots of x with positive real parts. The number of roots x with positive real parts will be the same as the number of s roots with positive real parts.
2. Multiply the original polynomial by the factor $(s+1)$, which introduces an additional negative root. Then form the Routhian array for the new polynomial.

Both of these methods are illustrated in the following example:

$$Q(s) = s^4 + s^3 + 2s^2 + 2s + 5 \tag{6.11}$$

The Routhian array is

$$\begin{array}{c|ccc} s^4 & 1 & 2 & 5 \\ s^3 & 1 & 2 & \\ s^2 & 0 & 5 & \end{array}$$

The zero in the first column prevents completion of the array. The following methods overcome this problem.

METHOD 1. Letting $s = 1/x$ and rearranging the polynomial gives

$$Q(x) = 5x^4 + 2x^3 + 2x^2 + x + 1 \tag{6.12}$$

The new Routhian array is

$$\begin{array}{c|ccc} x^4 & 5 & 2 & 1 \\ x^3 & 2 & 1 & \\ x^2 & -1 & 2 & \\ x^1 & 5 & & \\ x^0 & 2 & & \end{array}$$

There are two changes of sign in the first column. Therefore, there are two roots of x in the RHP. The number of roots s with positive real parts is also two. This method does not work when the coefficients of $Q(s)$ and of $Q(x)$ are identical.

METHOD 2.

$$Q_1(s) = Q(s)(s+1) = s^5 + 2s^4 + 3s^3 + 4s^2 + 7s + 5$$

$$\begin{array}{c|ccc} s^5 & 1 & 3 & 7 \\ s^4 & 2 & 4 & 5 \\ s^3 & 2 & 9 & \\ s^2 & -10 & 10 & \\ s^1 & 11 & & \\ s^0 & 10 & & \end{array}$$

The same result is obtained by both methods. There are two changes of sign in the first column, so there are two zeros of $Q(s)$ with positive real parts. An additional method is described in Ref. 6.

Theorem 3: A Zero Row. When all the coefficients of one row are zero, the procedure is as follows:

1. The auxiliary equation can be formed from the preceding row, as shown below.
2. The Routhian array can be completed by replacing the all-zero row with the coefficients obtained by differentiating the auxiliary equation.
3. The roots of the auxiliary equation are also roots of the original equation. These roots occur in pairs and are the negative of each other. Therefore, these roots may be imaginary (complex conjugates) or real (one positive and one negative), may lie in quadruplets (two pairs of complex-conjugate roots), etc.

Consider the system that has the characteristic equation

$$Q(s) = s^4 + 2s^3 + 11s^2 + 18s + 18 = 0 \tag{6.13}$$

The Routhian array is

$$\begin{array}{c|ccl} s^4 & 1 & 11 & 18 \\ s^3 & \cancel{2} & \cancel{18} & \\ & 1 & 9 & \text{(after dividing by 2)} \\ s^2 & \cancel{2} & \cancel{18} & \\ & 1 & 9 & \\ s^1 & 0 & & \text{(after dividing by 2)} \end{array}$$

The presence of a zero row (the s^1 row) indicates that there are roots that are the negatives of each other. The next step is to form the *auxiliary equation* from the preceding row, which is the s^2 row. The highest power of s is s^2, and only even powers of s appear. Therefore, the auxililary equation is

$$s^2 + 9 = 0 \tag{6.14}$$

The roots of this equation are

$$s = \pm j3$$

These are also roots of the original equation. The presence of imaginary roots indicates that the output includes a sinusoidally oscillating component.

To complete the Routhian array, the auxiliary equation is differentiated and is

$$2s + 0 = 0 \tag{6.15}$$

The coefficients of this equation are inserted in the s^1 row, and the array is then completed:

$$\begin{array}{c|c} s^1 & 2 \\ s^0 & 9 \end{array}$$

Since there are no changes of sign in the first column, there are no roots with positive real parts.

In feedback systems, covered in detail in the following chapters, the ratio of the output to the input does not have an explicitly factored denominator. An example of such a function is

$$\frac{X_2(s)}{X_1(s)} = \frac{P(s)}{Q(s)} = \frac{K(s+2)}{s(s+5)(s^2+2s+5) + K(s+2)} \tag{6.16}$$

The value of K is an adjustable parameter in the system and may be positive or negative. The value of K determines the location of the poles and therefore the stability of the system. It is important to know the range of values of K for which the system is stable. This information must be obtained from the characteristic equation, which is

$$Q(s) = s^4 + 7s^3 + 15s^2 + (25+K)s + 2K = 0 \tag{6.17}$$

The coefficients must all be positive in order for the zeros of $Q(s)$ to lie in the left half of the s plane (LHP), but this condition is not sufficient for stability. The Routhian array permits evaluation of precise boundaries for K:

$$\begin{array}{c|ccc} s^4 & 1 & 15 & 2K \\ s^3 & 7 & 25+K & \\ s^2 & 80-K & 14K & \\ s^1 & \dfrac{(80-K)(25+K)-98K}{80-K} & & \\ s^0 & 14K & & \end{array}$$

The term $80-K$ for the s^2 row imposes the restriction $K<80$, and the s^0 row requires $K>0$. The numerator of the first term in the s^1 row is equal to $-K^2-43K+2000$, and this function must be positive for a stable system. By use of the quadratic formula the zeros of this function are $K=28.1$ and $K=-71.1$, and the numerator of the s^1 row is positive between these values. The combined restrictions on K for stability of the system are therefore $0<K<28.1$. For the value $K=28.1$ the characteristic equation has imaginary roots that can be evaluated by applying Theorem 3 to form the auxiliary equation. Also, for $K=0$, it can be seen from Eq. (6.16) that there is no output. The methods for selecting the "best" value of K in the range between 0 and 28.1 are contained in later chapters. It is important to note that the Routh criterion provides useful but restricted information. Another method of determining stability is Hurwitz's criterion [2,8], which establishes the necessary conditions in terms of the system determinants.

6.3 MATHEMATICAL AND PHYSICAL FORMS

In various systems the controlled variable, labeled C, shown in Fig. 6.1 may have the physical form of position, speed, temperature, rate of change of temperature, voltage, rate of flow, pressure, etc. Once the blocks in the diagram are related to transfer functions, it is immaterial to the analysis of the system what the physical form of the controlled variable may be. Generally, the important quantities are the controlled quantity c, its rate of change Dc, and its second derivative D^2c, that is, the first several derivatives of c, including the zeroth derivative. For any specific control system each of these "mathematical" functions has a definite "physical" meaning. For example, if the controlled variable c is position, then Dc is velocity and D^2c is acceleration. As a second example, if the controlled variable c is velocity, then Dc is acceleration and D^2c is the rate of change of acceleration.

Often the input signal to a system has an irregular form, such as that shown in Fig. 6.2, that cannot be expressed by any simple equation. This prevents a straight-forward analysis of system response. It is noted, though, that the signal form shown in Fig. 6.2 may be considered to be composed of

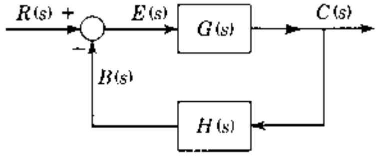


FIGURE 6.1 Simple feedback system.

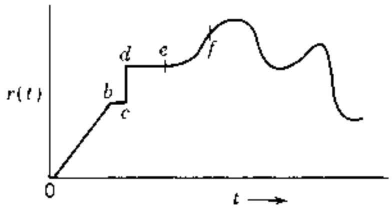


FIGURE 6.2 Input signal to a system.

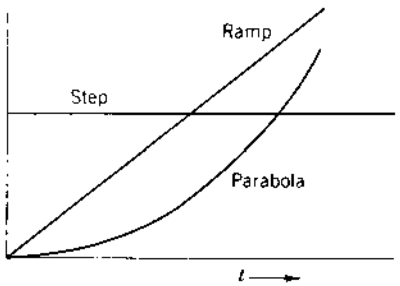


FIGURE 6.3 Graphical forms of step, ramp, and parabolic input functions.

three basic forms of known types of input signals, i.e., a step in the region *cde*, a ramp in the region 0*b*, and a parabola in the region *ef*. Thus, if the given linear system is analyzed separately for each of these types of input signals, there is then established a fair measure of performance with the irregular input. For example, consider the control system for a missile that receives its command input from an interceptor guidance system (see Fig. 5.3). This input changes slowly compared with the control-system dynamics. Thus, step and ramp inputs adequately approximate most command inputs. Therefore, the three standard inputs not only approximate most command inputs but also provide a means for comparing the performance of different systems.

Consider that the system shown in Fig. 6.1 is a position-control system. Feedback control systems are often analyzed on the basis of a unit step input signal (see Fig. 6.3). This system can first be analyzed on the basis that the unit step input signal $r(t)$ represents position. Since this gives only a limited idea of how the system responds to the actual input signal, the system can then be analyzed on the basis that the unit step signal represents a constant velocity $Dr(t) = u_{-1}(t)$. This in reality gives an input position signal of the form of a ramp (Fig. 6.3) and thus a closer idea of how the system responds to the actual input signal. In the same manner the unit step input signal can represent a constant acceleration, $D^2r(t) = u_{-1}(t)$, to obtain the system's performance to a parabolic position input signal. The curves shown in Fig. 6.3 then represent acceleration, velocity, and position.

6.4 FEEDBACK SYSTEM TYPES

The simple closed-loop feedback system, with unity feedback, shown in Fig. 6.4, may be called a *tracker* since the output $c(t)$ is expected to track or follow the input $r(t)$. The open-loop transfer function for this system is $G(s) = C(s)/E(s)$, which is determined by the components of the actual control system. Generally $G(s)$ has one of the following mathematical forms:

$$G(s) = \frac{K_0(1 + T_1 s)(1 + T_2 s)\cdots}{(1 + T_a s)(1 + T_b s)\cdots} \tag{6.18}$$

$$G(s) = \frac{K_1(1 + T_1 s)(1 + T_2 s)\cdots}{s(1 + T_a s)(1 + T_b s)\cdots} \tag{6.19}$$

$$G(s) = \frac{K_2(1 + T_1 s)(1 + T_2 s)\cdots}{s^2(1 + T_a s)(1 + T_b s)\cdots} \tag{6.20}$$

Note that the constant term in each factor is equal to unity. The preceding equations are expressed in a more generalized manner by defining the standard form of the transfer function as

$$G(s) = \frac{K_m(1 + b_1 s + b_2 s^2 + \cdots + b_w s^w)}{s^m(1 + a_1 s + a_2 s^2 + \cdots + a_u s^u)} = K_m G'(s) \tag{6.21}$$

where

$$\begin{aligned}
a_1, a_2, \ldots &= \text{constant coefficients} \\
b_1, b_2, \ldots &= \text{constant coefficients} \\
K_m &= \text{gain constant of the transfer function } G(s) \\
m &= 0, 1, 2, \ldots \text{ denotes the transfer function type} \\
G'(s) &= \text{forward transfer function with unity gain}
\end{aligned}$$

The degree of the denominator is $n = m + u$. For a unity-feedback system, E and C have the same units. Therefore, K_0 is nondimensional, K_1 has the units of seconds^{-1}, K_2 has the units of seconds^{-2}.

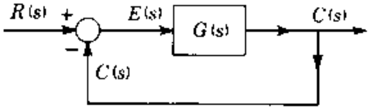


FIGURE 6.4 Unity-feedback control system.

In order to analyze each control system, a "type" designation is introduced. The designation is based upon the value of the exponent m of s in Eq. (6.21). Thus, when $m = 0$, the system represented by this equation is called a Type 0 system; when $m = 1$, it is called a Type 1 system; when $m = 2$, it is called a Type 2 system; etc. [See Prob. 6.20 for a Type -1 ($m = -1$) system.] Once a physical system has been expressed mathematically, the analysis is independent of the nature of the physical system. It is immaterial whether the system is electrical, mechanical, hydraulic, thermal, or a combination of these. The most common feedback control systems have Type 0, 1, or 2 open-loop transfer functions, as shown in Eqs. (6.18) to (6.20). It is important to analyze each type thoroughly and to relate it as closely as possible to its transient and steady-state solution.

The various types exhibit the following steady-state properties:

Type 0: A constant actuating signal results in a constant value for the controlled variable.

Type 1: A constant actuating signal results in a constant rate of change (constant velocity) of the controlled variable.

Type 2: A constant actuating signal results in a constant second derivative (constant acceleration) of the controlled variable.

Type 3: A constant actuating signal results in a constant rate of change of acceleration of the controlled variable.

These classifications lend themselves to definition in terms of the differential equations of the system and to identification in terms of the forward transfer function. For all classifications the degree of the denominator of $G(s)H(s)$ usually is equal to or greater than the degree of the numerator because of the physical nature of feedback control systems. That is, in every physical system there are energy-storage and dissipative elements such that there can be no instantaneous transfer of energy from the input to the output. However, exceptions do occur.

6.5 ANALYSIS OF SYSTEM TYPES

The properties presented in the preceding section are now examined in detail for each type of $G(s)$ appearing in stable unity-feedback tracking systems. First, remember the following theorems:

Final-value theorem:

$$\lim_{t \to \infty} f(t) = \lim_{s \to 0} sF(s) \tag{6.22}$$

Differential theorem:

$$\mathscr{L}[D^m c(t)] = s^m C(s) \qquad \text{when all initial conditions are zero} \tag{6.23}$$

Also, when the input is a polynomial, the steady-state output of a stable closed-loop system that has *unity feedback* $[H(s) = 1]$ has the same form as the input. Therefore, if the input is a ramp function, the steady-state output must also include a ramp function plus a constant, and so forth.

From the preceding section, it is seen that the forward transfer function defines the system type and is generally in the factored form

$$G(s) = \frac{C(s)}{E(s)} = \frac{K_m(1 + T_1 s)(1 + T_2 s)\cdots}{s^m(1 + T_a s)(1 + T_b s)(1 + T_c s)\cdots} \tag{6.24}$$

Solving this equation for $E(s)$ yields

$$E(s) = \frac{(1 + T_a s)(1 + T_b s)(1 + T_c s)\cdots}{K_m(1 + T_1 s)(1 + T_2 s)\cdots} s^m C(s) \tag{6.25}$$

Thus, applying the final-value theorem gives

$$\begin{aligned} e(t)_{ss} &= \lim_{s\to 0}[sE(s)] = \lim_{s\to 0}\left[\frac{s(1 + T_a s)(1 + T_b s)(1 + T_c s)\cdots}{K_m(1 + T_1 s)(1 + T_2 s)\cdots} s^m C(s)\right] \\ &= \lim_{s\to 0}\frac{s[s^m C(s)]}{K_m} \end{aligned} \tag{6.26}$$

Thus, applying the final-value theorem to Eq. (6 .23) gives

$$\lim_{s\to 0} s[s^m C(s)] = D^m c(t)_{ss} \tag{6.27}$$

Therefore, Eq. (6.26) can be written as

$$e(t)_{ss} = \frac{D^m c(t)_{ss}}{K_m} \tag{6.28}$$

or

$$K_m e(t)_{ss} = D^m c(t)_{ss} \tag{6.29}$$

This equation relates a derivative of the output to the error; it is most useful for the case where $D^m c(t)_{ss} = \text{constant}$. Then $e(t)_{ss}$ must also equal a constant, that is, $e(t)_{ss} = E_0$, and Eq. (6.29) may be expressed as

$$K_m E_0 = D^m c(t)_{ss} = \text{constant} = C_m \tag{6.30}$$

Note that $C(s)$ has the form

$$C(s) = \frac{G(s)}{1 + G(s)} R(s)$$
$$= \frac{K_m[(1 + T_1 s)(1 + T_2 s) \cdots]}{s^m(1 + T_a s)(1 + T_b s) \cdots + K_m(1 + T_1 s)(1 + T_2 s) \cdots} R(s) \tag{6.31}$$

The expression for $E(s)$ in terms of the input $R(s)$ is obtained as follows:

$$E(s) = \frac{C(s)}{G(s)} = \frac{1}{G(s)} \frac{G(s)R(s)}{1 + G(s)} = \frac{R(s)}{1 + G(s)} \tag{6.32}$$

With $G(s)$ given by Eq. (6.24), ther expression for $E(s)$ is

$$E(s) = \frac{s^m(1 + T_a s)(1 + T_b s) \cdots R(s)}{s^m(1 + T_a s)(1 + T_b s) \cdots + K_m(1 + T_1 s)(1 + T_2 s) \cdots} \tag{6.33}$$

Applying the final-value theorem to Eq. (6.33) yields

$$e(t)_{ss} = \lim_{s \to 0} s \left[\frac{s^m(1 + T_a s)(1 + T_b s) \cdots R(s)}{s^m(1 + T_a s)(1 + T_b s) \cdots + K_m(1 + T_1 s)(1 + T_2 s) \cdots} \right] \tag{6.34}$$

Equation (6.34) relates the steady-state error to the input; it is now analyzed for various system types and for step, ramp, and parabolic inputs.

Case 1: $m = 0$ (Type 0 System)

STEP INPUT $r(t) = R_0 u_{-1}(t)$, $R(s) = R_0/s$. From Eq. (6.34):

$$e(t)_{ss} = \frac{R_0}{1 + K_0} = \text{constant} = E_0 \neq 0 \tag{6.35}$$

Applying the final-value theorem to Eq. (6.31) yields:

$$c(t)_{ss} = \frac{K_0}{1 + K_0} R_0 \tag{6.36}$$

From Eq. (6.35) it is seen that a Type 0 system with a constant input produces a constant value of the output and a constant actuating signal. [The same results can be obtained by applying Eq. (6.30).] Thus in a Type 0 system a fixed error E_0 is required to produce a desired constant output C_0; that is, $K_0 E_0 = c(t)_{ss} = \text{constant} = C_0$. For steady-state conditions:

$$e(t)_{ss} = r(t)_{ss} - c(t)_{ss} = R_0 - C_0 = E_0 \tag{6.37}$$

Differentiating the preceding equation yields $Dr(t)_{ss} = Dc(\mathrm{t})_{ss} = 0$. Figure 6.5 illustrates these results.

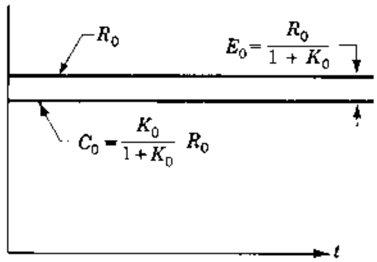


FIGURE 6.5 Steady-state response of a Type 0 system with a step input.

RAMP INPUT $r(t) = R_1 t u_{-1}(t)$, $R(s) = R_1/s^2$. From Eq. (6.34), $e(t)_{ss} = \infty$. Also, by the use of the Heaviside partial-fraction expansion, the particular solution of $e(t)$ obtained from Eq. (6.33) contains the term $[R_1/(1+K_0)]t$. Therefore, the conclusion is that a Type 0 system with a ramp-function input produces a ramp output with a smaller slope; thus, there is an error that increases with time and approaches a value of infinity, meaning that a Type 0 system cannot follow a ramp input.

Similarly, a Type 0 system cannot follow the parabolic input

$$r(t) = \frac{R_2 t^2}{2} u_{-1(t)} \tag{6.38}$$

since $e(t)_{ss} = r(t)_{ss} - c(t)_{ss}$ approaches a value of infinity.

Case 2: $m = 1$ (Type 1 System)

STEP INPUT $R(s) = R_0/s$. From Eq. (6.34), $e(t)_{ss} = 0$. Therefore, a Type 1 system with a constant input produces a steady-state constant output of value identical with the input. That is, for a Type 1 system there is zero steady-state error between the output and input for a step input, i.e.,

$$e(t)_{ss} = r(t)_{ss} - c(t)_{ss} = 0 \tag{6.39}$$

The preceding analysis is in agreement with Eq. (6.30). That is, for a step input, the steady-state output must also be a constant $c(t) = C_0 u_{-1}(t)$ so that

$$Dc(t)_{ss} = 0 = K_1 E_0 \tag{6.40}$$

or

$$E_0 = 0 \tag{6.41}$$

RAMP INPUT $R(s) = R_1/s^2$. From Eq. (6.34)

$$e(t)_{ss} = \frac{R_1}{K_1} = \text{constant} = E_0 \neq 0 \tag{6.42}$$

From Eq. (6.42) it can be seen that a Type 1 system with a ramp input produces a constant actuating signal. That is, in a Type 1 system a fixed error E_0 is required to produce a ramp output. This result can also be obtained from Eq. (6.30), that is, $K_1 E_0 = Dc\,(t)_{ss} = \text{constant} = C_1$. For steady-state conditions:

$$e(t)_{ss} = r(t)_{ss} - c(t)_{ss} = E_0 \tag{6.43}$$

For the ramp input

$$r(t) = R_1 t u_{-1}(t) \tag{6.44}$$

the particular solution for the output has the form of a power series:

$$c(t)_{ss} = C_0 + C_1 t \tag{6.45}$$

Substituting Eqs. (6.44) and (6.45) into Eq. (6.43) yields

$$E_0 = R_1 t - C_0 - C_1 t \tag{6.46}$$

This result can occur only with

$$R_1 = C_1 \tag{6.47}$$

This result signifies that the slope of the ramp input and the ramp output are equal. Of course, this condition is necessary if the difference between input and output is a constant. Figure 6.6 illustrates these results. The delay of the steady-state output ramp shown in Fig. 6.6 is equal to

$$\text{Delay} = \frac{e(t)_{ss}}{Dr(t)} = \frac{1}{K_1} \tag{6.48}$$

PARABOLIC INPUT. $r(t) = (R_2 t^2/2)u_{-1}(t)$, $R(s) = R_2/s^3$. The particular solution of $e(t)$ obtained from Eq. (6.33) contains the term

$$e(t) = \frac{R_2 t}{K_1} \tag{6.49}$$

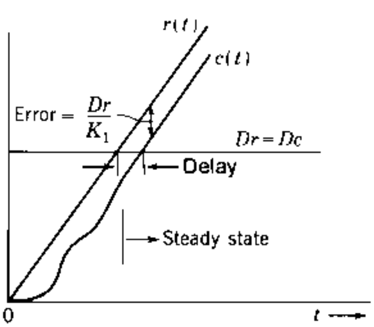


FIGURE 6.6 Steady-state response of a Type 1 system with a ramp input.

Therefore, a Type 1 system with a parabolic input produces a parabolic output, but with an error that increases with time and approaches a value of infinity. This limit means that a Type 1 system cannot follow a parabolic input.

Case 3: $m = 2$ (Type 2 System)

STEP INPUT $R(s) = R_0/s$. Performing a corresponding analysis in the same manner as for Cases 1 and 2 reveals that a Type 2 system can follow a step input with zero steady-state error, i.e., $c(t)_{ss} = r(t) = R_0 u_{-1}(t)$ and $e(t)_{ss} = 0$.

RAMP INPUT $R(s) = R_1/s^2$. Performing a corresponding analysis in the same manner as for Cases 1 and 2 reveals that a Type 2 system can follow a ramp input with zero steady-state error, i.e., $c(t)_{ss} = r(t) = R_1 t$ and $e(t)_{ss} = 0$.

PARABOLIC INPUT $R(s) = R_2/s^3$. From Eq. (6.34),

$$e(t)_{ss} = \frac{R_2}{K_2} = \text{constant} = E_0 \neq 0 \tag{6.50}$$

From Eq. (6.50) it can be seen that a Type 2 system with a parabolic input produces a parabolic output with a constant actuating signal. That is, in a Type 2 system a fixed error E_0 is required to produce a parabolic output. This result is further confirmed by applying Eq. (6.30), i.e.,

$$K_2 E_0 = D^2 c(t)_{ss} = \text{constant} = C_2 \tag{6.51}$$

Thus, for steady-state conditions

$$e(t)_{ss} = r(t)_{ss} - c(t)_{ss} = E_0 \tag{6.52}$$

For a parabolic input given by

$$r(t) = \frac{R_2 t^2}{2} u_{-1}(t) \tag{6.53}$$

the particular solution of the output must be given by the power series

$$c(t)_{ss} = \frac{C_2}{2} t^2 + C_1 t + C_0 \tag{6.54}$$

Substituting Eqs. (6.53) and (6.54) into Eq. (6.52) yields

$$E_0 = \frac{R_2 t^2}{2} - \frac{C_2 t^2}{2} - C_1 t - C_0 \tag{6.55}$$

Differentiating Eq. (6.55) twice results in $C_1 = 0$ and $R_2 = C_2$. Therefore, $E_0 = -C_0$ and the input and output curves have the same shape but are displaced by a constant error. Figure 6.7 illustrates the results obtained above.

The results determined in this section verify the properties stated in the previous section for the system types. The steady-state response characteristics

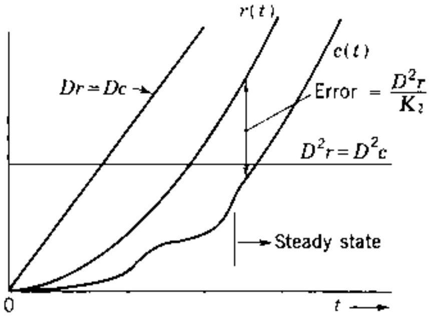


FIGURE 6.7 Steady-state response of a Type 2 system with a parabolic input.

for Types 0, 1, and 2 unity-feedback systems are given in Table 6.1 and apply only to stable systems.

6.6 EXAMPLE: TYPE 2 SYSTEM

A Type 2 system is one in which the second derivative of the output is maintained constant by a constant actuating signal. In Fig. 6.8 is shown a positioning system that is a Type 2 system, where

K_b = motor back-emf constant, V/(rad/s)
K_x = potentiometer constant, V/rad
K_T = motor torque constant, lb.ft/A
A = integrator amplifier gain, s^{-1}
K_g = generator constant, V/A
$T_m = J R_{gm}/(K_b K_T + R_{gm} B)$ = motor mechanical constant, s
T_f = generator field constant, s
$K_M = K_T/(BR_{gm} + K_T K_b)$ = overall motor constant, rad/V.s

The motor-generator shown in Fig. 6.8 is a power amplifier and can be represented by a combination of the circuits shown in Figs. 2.25 and 2.26. Assuming that the inductances of the generator and motor armatures, described in Secs. 2.12 and 2.13, are negligible, the transfer function is given by Eq. (2.135). Figure 6.9 is the block diagram representation of the position control system shown in Fig. 6.8. Since there are two integrating actions in the forward path, this is a Type 2 system. The forward transfer function is

$$G(s) = \frac{\theta_0(s)}{E(s)} = \frac{K_2}{s^2(1 + T_f s)(1 + T_m s)} \tag{6.56}$$

where $K_2 = K_x A K_g K_M / R_f$ has the units of seconds^{-2}.

TABLE 6.1 Steady-State Response Characteristics for Stable Unity-Feedback Systems

System type m	$r(t)_{ss}$	$c(t)_{ss}$	$e(t)_{ss}$	$e(\infty)$	Derivatives
0	$R_0u_{-1}(t)$	$\frac{K_0}{1+K_0}R_0$	$\frac{R_0}{1+K_0}$	$\frac{R_0}{1+K_0}$	$Dr = Dc = 0$
	$R_1tu_{-1}(t)$	$\frac{K_0R_1}{1+K_0}t + C_0$	$\frac{R_1}{1+K_0}t - C_0$	∞	$Dr \neq Dc$
	$\frac{R_2t^2}{2}u_{-1}(t)$	$\frac{K_0R_2}{2(1+K_0)}t^2 + C_1t + C_0$	$\frac{R_2}{2(1+K_0)}t^2 - C_1t - C_0$	∞	$Dr \neq Dc$
1	$R_0u_{-1}(t)$	R_0	0	0	$Dr = Dc = 0$
	$R_1tu_{-1}(t)$	$R_1t - \frac{R_1}{K_1}$	$\frac{R_1}{K_1}$	$\frac{R_1}{K_1}$	$Dr = Dc = R_1$
	$\frac{R_2t^2}{2}u_{-1}(t)$	$\frac{R_2t^2}{2} - \frac{R_2}{K_1}t + C_0$	$\frac{R_2}{K_1}t - C_0$	∞	$Dr \neq Dc$
2	$R_0u_{-1}(t)$	R_0	0	0	$Dr = Dc = 0$
	$R_1tu_{-1}(t)$	R_1t	0	0	$Dr = Dc = R_1$
	$\frac{R_2t^2}{2}u_{-1}(t)$	$\frac{R_2t^2}{2} - \frac{R_2}{K_2}$	$\frac{R_2}{K_2}$	$\frac{R_2}{K_2}$	$D^2r = D^2c = R_2$ $Dr = Dc = R_2t$

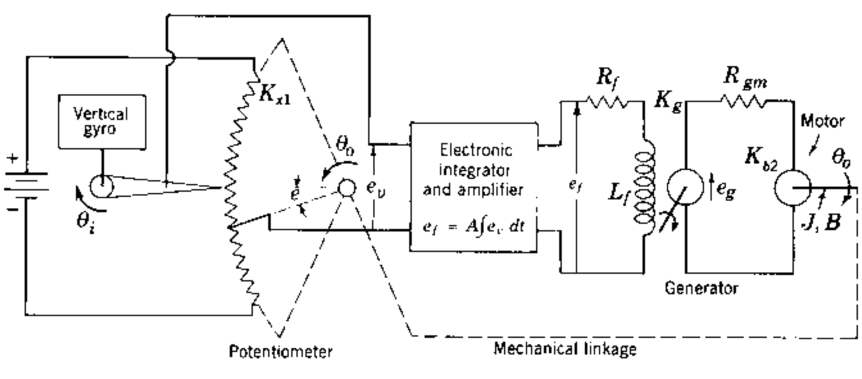


FIGURE 6.8 Position control of a space-vehicle camera (Type 2 system).

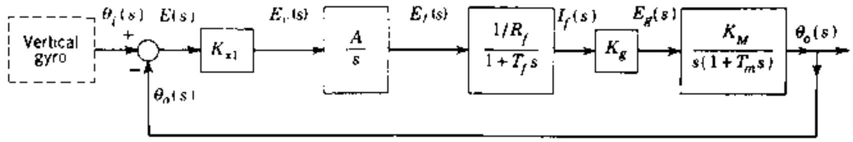


FIGURE 6.9 Block diagram representation of the system of Fig. 6.8.

This Type 2 control system is unstable. As shown in Chap. 10, an appropriate cascade compensator can be added to the forward path to produce a stable system. The new transfer function, with $T_1 > T_2$, has the form

$$G_0(s) = G_c(s)G(s) = \frac{K_2'(1 + T_1 s)}{s^2(1 + T_f s)(1 + T_m s)(1 + T_2 s)} \tag{6.57}$$

For a constant input $r(t) = R_0 u_{-1}(t)$ the steady-state value of $c(t)_{ss}$ is

$$c(t)_{ss} = \lim_{s \to 0}\left[s \frac{C(s)}{R(s)} \frac{R_0}{s}\right] = \lim_{s \to 0}\left[\frac{G_0(s)}{1 + G_0(s)} R_0\right] = R_0 \tag{6.58}$$

Thus

$$E_0 = r(t)_{ss} - c(t)_{ss} = R_0 - R_0 = 0 \tag{6.59}$$

As expected, a Type 2 system follows a step-function input with no steady-state error.

6.7 STEADY-STATE ERROR COEFFICIENTS [9]

The definition of system types is the first step toward establishing a set of standard characteristics that permit the engineer to obtain as much information as possible about a given system with a minimum amount of calculation. Also, these standard characteristics must point the direction in which a given system must be modified to meet a given set of performance specifications. An important characterisitic is the ability of a system to achieve the desired steady-state output with a minimum error. Thus, in this section there are defined *system error coefficients* that are a measure of a stable unity-feedback control system's steady-state accuracy for a given desired output that is relatively constant or slowly varying.

In Eq. (6.30) it is shown that a constant actuating signal exists when the derivative of the output is constant. This derivative is proportional to the actuating signal E_0 and to a constant K_m, which is the gain of the forward transfer function. The conventional names of these constants for the Type 0, 1, and 2, systems are *position, velocity,* and *acceleration error coefficients,* respectively. Since the conventional names were originally selected for

TABLE 6.2 Correspondence Between the Conventional and the Authors' Designation of Steady-State Error Coefficients

Symbol	Conventional designation of error coefficients	Authors' designation of error coefficients
K_p	Position	Step
K_v	Velocity	Ramp
K_a	Acceleration	Parabolic

application to position-control systems (servomechanisms), these names referred to the actual physical form of $c(t)$ or $r(t)$, which represent position, as well as to the mathematical form of $c(t)$, that is, c, Dc, and D^2c. These names are ambiguous when the analysis is extended to cover control of temperature, velocity, etc. In order to define general terms that are universally applicable, the authors have selected the terminology *step, ramp,* and *parabolic steady-state error coefficients.* Table 6.2 lists the conventional and the author's designation of the error coefficients.

The *error coefficients are independent of the system type. They apply to any system type and are defined for specific forms of the input, i.e., for a step, ramp, or parabolic input. These error coefficients are applicable only for stable unity-feedback systems.* The results are summarized in Table 6.3.

Steady-State Step Error Coefficient

The step error coefficient is defined as

Step error coefficient

$$= \frac{\text{steady-state value of output } c(t)_{ss}}{\text{steady-state actuating signal } e(t)_{ss}} = K_p \tag{6.60}$$

and applies only for a step input, $r(t) = R_0 u_{-1}(t)$. The steady-state value of the output is obtained by applying the final-value theorem to Eq. (6.31):

$$c(t)_{ss} = \lim_{s\to 0} sC(s) = \lim_{s\to 0}\left[\frac{sG(s)}{1+G(s)}\frac{R_0}{s}\right] = \lim_{s\to 0}\left[\frac{G(s)}{1+G(s)}R_0\right] \tag{6.61}$$

Similarly, from Eq. (6.32), for a unity-feedback system

$$e(t)_{ss} = \lim_{s\to 0}\left[s\frac{1}{1+G(s)}\frac{R_0}{s}\right] = \lim_{s\to 0}\left[\frac{1}{1+G(s)}R_0\right] \tag{6.62}$$

TABLE 6.3 Definitions of Steady-State Error Coefficients for Stable Unity-Feedback Systems

Error coefficient	Definition of error coefficient	Value of error coefficient	Form of input signal $r(t)$
Step (K_p)	$\dfrac{c(t)_{ss}}{e(t)_{ss}}$	$\lim_{s\to 0} G(s)$	$R_0 u_{-1}(t)$
Ramp (K_v)	$\dfrac{(Dc)_{ss}}{e(t)_{ss}}$	$\lim_{s\to 0} sG(s)$	$R_1 t u_{-1}(t)$
Parabolic (K_a)	$\dfrac{(D^2c)_{ss}}{e(t)_{ss}}$	$\lim_{s\to 0} s^2 G(s)$	$\dfrac{R_1 t^2}{2} u_{-1}(t)$

Substituting Eqs. (6.61) and (6.62) into Eq. (6.60) yields

$$\text{Step error coefficient} = \frac{\lim_{s\to 0}\left[\dfrac{G(s)}{1+G(s)} R_0\right]}{\lim_{s\to 0}\left[\dfrac{1}{1+G(s)} R_0\right]} \tag{6.63}$$

Since both the numerator and the denominator of Eq. (6.63) in the limit can never be zero or infinity simultaneously, where $K_m \neq 0$, the indeterminate forms 0/0 and ∞/∞ never occur. Thus, this equation reduces to

$$\text{Step error coefficient} = \lim_{s\to 0} G(s) = K_p \tag{6.64}$$

Therefore, applying Eq. (6.64) to each type system yields

$$K_p \begin{cases} \lim_{s\to 0} \dfrac{K_0(1+T_1 s)(1+T_2 s)\cdots}{(1+T_a s)(1+T_b s)(1+T_c s)\cdots} = K_0 & \text{Type 0 system} \quad (6.65) \\ \infty & \text{Type 1 system} \quad (6.66) \\ \infty & \text{Type 2 system} \quad (6.67). \end{cases}$$

Steady-State Ramp Error Coefficient

The ramp error coefficient is defined as

$$\text{Ramp error coefficient} = \frac{\text{steady-state value of derivative of output } Dc(t)_{ss}}{\text{steady-state actuating signal} e(t)_{ss}}$$

$$= K_v \tag{6.68}$$

and applies only for a ramp input, $r(t) = R_1 t u_{-1}(t)$. The first derivative of the output is given by

$$\mathscr{L}[Dc] = sC(s) = \frac{sG(s)}{1 + G(s)} R(s) \tag{6.69}$$

The steady-state value of the derivative of the output is obtained by using the final-value theorem:

$$Dc(t)_{ss} = \lim_{s\to 0} s[sC(s)] = \lim_{s\to 0}\left[\frac{s^2 G(s)}{1 + G(s)}\frac{R_1}{s^2}\right] = \lim_{s\to 0}\left[\frac{G(s)}{1 + G(s)} R_1\right] \tag{6.70}$$

Similarly, from Eq. (6.32), for a unity-feedback system

$$e(t)_{ss} = \lim_{s\to 0}\left[s\frac{1}{1 + G(s)}\frac{R_1}{s^2}\right] = \lim_{s\to 0}\left[\frac{1}{1 + G(s)}\frac{R_1}{s}\right] \tag{6.71}$$

Substituting Eqs. (6.70) and (6.71) into Eq. (6.68) yields

$$\text{Ramp error coefficient} = \frac{\lim_{s\to 0}\left[\dfrac{G(s)}{1 + G(s)} R_1\right]}{\lim_{s\to 0}\left[\dfrac{1}{1 + G(s)}(R_1/s)\right]} \tag{6.72}$$

Since this equation never has the indeterminate form 0/0 or ∞/∞, it can be simplified to

$$\text{Ramp error coefficient} = \lim_{s\to 0} sG(s) = K_v \tag{6.73}$$

Therefore, applying Eq. (6.73) to each type system yields

$$K_v = \begin{cases} \lim\limits_{s\to 0} \dfrac{sK_0(1 + T_1 s)(1 + T_2 s)\cdots}{(1 + T_a s)(1 + T_b s)(1 + T_c s)\cdots} = 0 & \text{Type 0 system} \quad (6.74) \\ K_1 & \text{Type 1 system} \quad (6.75) \\ \infty & \text{Type 2 system} \quad (6.76) \end{cases}$$

Steady-State Parabolic Error Coefficient

The parabolic error coefficient is defined as

$$\begin{aligned} &\text{Parabolic error coefficient} \\ &\quad = \frac{\text{steady-state value of second derivative of output } D^2 c(t)_{ss}}{\text{steady-state actuating signal } e(t)_{ss}} \\ &\quad = K_a \end{aligned} \tag{6.77}$$

and applies only for a parabolic input, $r(t) = (R_2t^2/2)u_{-1}(t)$. The second derivative of the output is given by

$$\mathscr{L}[D^2c] = s^2C(s) = \frac{s^2G(s)}{1+G(s)}R(s) \tag{6.78}$$

The steady-state value of the second derivative of the output is obtained by using the final-value theorem:

$$D^2c(t)_{ss} = \lim_{s\to 0} s[s^2C(s)] = \lim_{s\to 0}\left[\frac{s^3G(s)}{1+G(s)}\frac{R_2}{s^3}\right] = \lim_{s\to 0}\left[\frac{G(s)}{1+G(s)}R_2\right] \tag{6.79}$$

Similarly, from Eq. (6.32), for a unity-feedback system

$$e(t)_{ss} = \lim_{s\to 0}\left[s\frac{1}{1+G(s)}\frac{R_2}{s^3}\right] = \lim_{s\to 0}\left[\frac{1}{1+G(s)}\frac{R_2}{s^2}\right] \tag{6.80}$$

Substituting Eqs. (6.79) and (6.80) into Eq. (6.77) yields

$$\text{Parabolic error coefficient} = \frac{\lim_{s\to 0}\left[\dfrac{G(s)}{1+G(s)}R_2\right]}{\lim_{s\to 0}\left[\dfrac{1}{1+G(s)}(R_2/s^2)\right]} \tag{6.81}$$

Since this equation never has the indeterminate form 0/0 or ∞/∞, it can be simplified to

$$\text{Parabolic error coefficient} = \lim_{s\to 0} s^2G(s) = K_a \tag{6.82}$$

Therefore, applying Eq. (6.82) to each type system yields

$$K_a = \begin{cases} \lim_{s\to 0}\dfrac{s^2K_0(1+T_1s)(1+T_2s)\cdots}{(1+T_as)(1+T_bs)(1+T_cs)\cdots} = 0 & \text{Type 0 system} \quad (6.83)\\ 0 & \text{Type 1 system} \quad (6.84)\\ K_2 & \text{Type 2 system} \quad (6.85)\end{cases}$$

Based upon the steady-state error coefficient definitions and referring to the corresponding Figs. 6.5 through 6.7, the values of K_0, K_1, and K_2 can also be determined from the computer simulation of the control system. For example, from Fig. 6.6 and using Eq. (6.48), the gain is $K_1 = 1/\text{delay}$.

6.8 CAD ACCURACY CHECKS: CADAC

When appropriate, the steady-state error coefficient definitions can serve as CADAC. For example, before entering the $G(s)$ function into a CAD package, determine the values of K_m. This step is followed by obtaining the values of K_m from the expression of $G(s)$ on the computer screen or printed

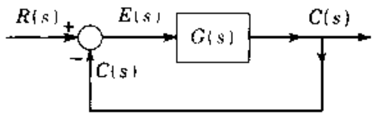


FIGURE 6.10 Unity-feedback system.

out by the CAD package. If the two sets of values agree, then the next required CAD manipulation is performed.

6.9 USE OF STEADY-STATE ERROR COEFFICIENTS

The use of steady-state error coefficients is discussed for stable unity-feedback systems, as illustrated in Fig. 6.10. Note that the steady-state error coefficients *can only be used* to find errors for unity-feedback systems that are *also* stable.

Type 1 System

For a Type 1 system with a step or ramp input considered at steady-state, the value of $(Dc)_{ss}$ is

$$Dc(t)_{ss} = K_1 E_0 \tag{6.86}$$

Thus, the larger, the value of K_1, the smaller the size of the actuating signal e necessary to maintain a constant rate of change of the output. From the standpoint of trying to maintain $c(t) = r(t)$ at all times, a larger K_1 results in a more sensitive system. In other words, a larger K_1 results in a greater speed of response of the system to a given actuating signal $e\,(t)$. Therefore, the gain K_1 is another standard characteristic of a system's performance. The maximum value K_1 is limited by stability considerations and is discussed in later chapters.

For the Type 1 system with a step input, the step error coefficient is equal to infinity, and the steady-state error is therefore zero. The steady-state output $c(t)_{ss}$ for a Type 1 system is equal to the input when $r(t) = $ constant.

Consider now a ramp input $r(t) = R_1 t u_{-1}(t)$ The steady-state value of $Dc(\mathrm{t})_{ss}$ is found by using the final-value theorem:

$$Dc(t)_{ss} = \lim_{s\to 0} s[sC(s)] = \lim_{s\to 0}\left[s\frac{sG(s)}{1+G(s)}R(s)\right] \tag{6.87}$$

where $R(\mathrm{s}) = R_1/\mathrm{s}^2$ and

$$G(s) = \frac{K_1(1+T_1 s)(1+T_2 s)\cdots(1+T_w s)}{s(1+T_a s)(1+T_b s)\cdots(1+T_u s)}$$

Inserting these values into Eq. (6.87) gives

$$Dc(t)_{ss} = \lim_{s\to 0}\left[s\frac{sK_1(1+T_1s)(1+T_2s)\cdots(1+T_ws)}{\begin{matrix} s(1+T_as)(1+T_bs)\cdots(1+T_us) \\ +K_1(1+T_1s)(1+T_2s)\cdots(1+T_ws)\end{matrix}}\frac{R_1}{s^2}\right] = R_1$$

Therefore,

$$(Dc)_{ss} = (Dr)_{ss} \tag{6.88}$$

The magnitude of the steady-state error is found by using the ramp error coefficient. From Eq. (6.73).

$$K_v = \lim_{s\to 0} sG(s) = K_1$$

From the definition of ramp error coefficient, the steady-state error is

$$e(t)_{ss} = \frac{Dc(t)_{ss}}{K_1} \tag{6.89}$$

Since $Dc(t)_{ss} = Dr(t)$,

$$e(t)_{ss} = \frac{Dr}{K_1} = \frac{R_1}{K_1} = E_0 \tag{6.90}$$

Therefore, a Type 1 system follows a ramp input with a constant error E_0. Figure 6.11 illustrates these conditions graphically.

Table of Steady-State Error Coefficients

Table 6.4 gives the values of the error coefficients for the Type 0,1, and 2 systems. These values are determined from Table 6.3. The reader should be able to make ready use of Table 6.4 for evaluating the appropriate error

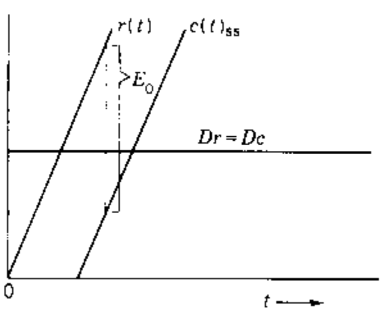


FIGURE 6.11 Steady-state response of a Type 1 system for $Dr =$ constant.

TABLE 6.4 Steady-State Error Coefficients for Stable Systems

System type	Step error coefficient K_p	Ramp error coefficient K_v	Parabolic error coefficient K_a
0	K_0	0	0
1	∞	K_1	0
2	∞	∞	K_2

coefficient. The error coefficient is used with the definitions given in Table 6.3 to evaluate the magnitude of the steady-state error.

Polynomial Input: t^{m+1}

Note that a Type m system can follow an input of the form t^{m-1} with zero steady-state error. It can follow an input t^m, but there is a constant steady-state error. It cannot follow an input t^{m+1} because the steady-state error approaches infinity. However, if the input is present only for a finite length of time, the error is also finite. Then the error can be evaluated by taking the inverse Laplace transform of Eq. (6.33) and inserting the value of time. The maximum permissible error limits the time ($0 < t < t_1$) that an input t^{m+1} can be applied to a control system.

6.10 NONUNITY-FEEDBACK SYSTEM

The nonunity-feedback system of Fig. 6.1 may be mathematically converted to an equivalent unity-feedback system from which "effective" system type and steady-state error coefficient can be determined. The control ratio for Fig. 6.1 is

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)H(s)} = \frac{N(s)}{D(s)} \tag{6.91}$$

and for the *equivalent* unity-feedback control system of the form of Fig. 6.4, the control ratio is

$$\frac{C(s)}{R(s)} = \frac{G_{eq}(s)}{1 + G_{eq}(s)} \tag{6.92}$$

Since the transfer functions $G(s)$ and $H(s)$ are known, equating Eqs. (6.91) and (6.92) yields

$$G_{eq}(s) = \frac{N(s)}{D(s) - N(s)} \tag{6.93}$$

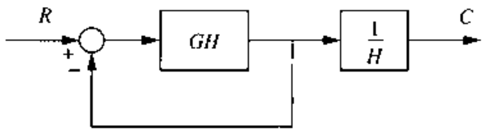


FIGURE 6.12 Equivalent nonunity feedback representation of Fig. 6.1.

When the nonunity-feedback system is stable, its steady-state performance characteristics can be determined based on Eq. (6.93).

Equation (6.91) can also be expressed in the form

$$\frac{C(s)}{R(s)} = \left[\frac{G(s)H(s)}{1 + G(s)H(s)}\right]\frac{1}{H(s)} \tag{6.94}$$

Thus, the nonunity feedback system of Fig. 6.1 can be represented by the block diagram of Fig. 6.12 in accordance with Eq. (6.94). This diagram shows a unity feedback system in cascade with $1/H(s)$. When H is a constant, Fig. 6.12 is an especially useful representation that permits the design to be performed using unity feedback methods.

6.11 SUMMARY

In this chapter the physical forms of the reference input and the controlled variable are related to the mathematical representations of these quantities. Since, in general, the forward transfer functions of most unity feedback control systems fall into three categories, they can be identified as Type 0, 1, and 2 systems. The associated definitions of the steady-state error coefficients readily provide information that is indicative of a stable system's steady-state performance. Thus, a start has been made in developing a set of standard characteristics. A nonunity-feedback system can also be represented by an equivalent unity-feedback system having the forward transfer function $G_{eq}(s)$ (see Sec. 6.9). Then the steady-state performance can be evaluated on the basis of the system type and the error coefficients determined from $G_{eq}(s)$.

REFERENCES

1. Etter, D. M. :*Engineering Problem Solving with MATLAB®*, Prentice-Hall, Englewood Cliffs, N.J., 1993.
2. Guillemin, E. A.: *The Mathematics of Circuit Analysis,* Wiley, New York, 1949.
3. Singh, V.: "Comments on the Routh-Hurwitz Criterion," *IEEE Trans. Autom. Control,* vol. AC-26, p.612, 1981.
4. Khatwani, K. J.: "On Routh-Hurwitz Criterion," *IEEE Trans. Autom. Control,* vol. AC-26, pp. 583–584, 1981.

5. Pillai, S. K.: "On the ε-Method of the Routh-Hurwitz Criterion," *IEEE Trans. Autom. Control,* vol. AC-26, p. 584, 1981.
6. Gantmacher, F. R.: *Applications of the Theory of Matrices,* Wiley-Interscience, New York, 1959.
7. Krishnamurthi, V.: "Implications of Routh Stability Criteria," *IEEE Trans. Autom. Control,* vol. AC-25, pp. 554–555, 1980.
8. Porter, Brian: *Stability Criteria for Linear Dynamica Systems,* Academic, New York, 1968.
9. Chestnut, H., and R. W. Mayer: *Servomechanisms and Regulating System Design,* 2nd ed., vol. 1, Wiley, New York, 1959.

7

Root Locus

7.1 INTRODUCTION

A designer uses the time response of a system to determine if his or her design of a control system meets specifications. An accurate solution of the system's performance can be obtained by deriving the differential equations for the control system and solving them. However, if the response does not meet specifications, it is not easy to determine from this solution just what physical parameters in the system should be changed to improve the response. A designer wishes to be able to predict a system's performance by an analysis that does not require the actual solution of the differential equations. Also, the designer would like this analysis to indicate readily the manner or method by which this system must be adjusted or compensated to produce the desired performance characteristics.

The first thing that a designer wants to know about a given system is whether or not it is stable. This can be determined by examining the roots obtained from the characteristic equation $1 + G(s)H(s) = 0$. By applying Routh's criterion to this equation, it is possible in short order to determine stability without solving for the roots. Yet this does not satisfy the designer because it does not indicate the degree of stability of the system, i.e., the amount of overshoot and the settling time of the controlled variable. Not only must the system be stable, but also the overshoot must be maintained within prescribed limits and transients must die out in a sufficiently short time. The graphical methods described in this text not only indicate whether a system is stable or unstable but, for a stable system, also show the degree of stability. A number of commercially available computer-aided-design (CAD) programs can be used to obtain the solution. MATLAB is used as the main CAD program in this text and its use is described in Appendix C. The CAD program TOTAL-PC is described in Appendix D and is especially useful for root locus and frequency response design. In addition to the time response, a designer can choose to analyze and interpret the steady-state sinusoidal response of the transfer function of the system to obtain an idea of the system's response. This method is based upon the interpretation of a Nyquist plot, discussed in Chaps. 8 and 9. The frequency-response approach yields enough information to indicate whether the system needs to be adjusted or compensated and how the system should be compensated.

This chapter deals with *the root-locus method* [1,2], which incorporates the more desirable features of both the classical and the frequency-response methods. *The root locus is a plot of the roots of the characteristic equation of the closed-loop system as a function of the gain of the open-loop transfer function.* With relatively small effort this graphical approach yields a clear indication of the effect of gain adjustment. The underlying principle is that the poles of $C(s)/R(s)$ (transient-response modes) are related to the zeros and poles of the open-loop transfer function $G(s)H(s)$ and also to the gain. This relationship is shown by Eq. (5.4), which is repeated here:

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)H(s)} \tag{7.1}$$

An important advantage of the root-locus method is that the roots of the characteristic equation of the system can be obtained directly. This yields a complete and accurate solution of the transient and steady-state response of the controlled variable. Also, an approximate solution can be obtained with a reduction of the work required. A person can readily obtain proficiency with the root-locus method. With the help of a CAD program, such as MATLAB or TOTAL-PC (see Appendixes C and D), it is possible to design a system with relative ease.

7.2 PLOTTING ROOTS OF A CHARACTERISTIC EQUATION

To give a better insight into root-locus plots, consider the position-control system shown in Fig. 7.1. The motor produces a torque $T(s)$ that is proportional to the actuating signal $E(s)$. The load consists of the combined motor and load inertia J and viscous friction B. The forward transfer function

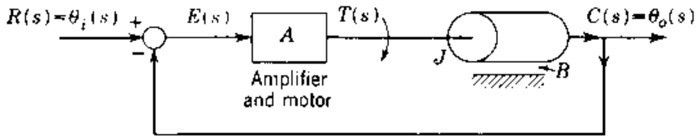


FIGURE 7.1 A position-control system.

(see Sec. 2.12) is

$$G(s) = \frac{\theta_o(s)}{E(s)} = \frac{A/J}{s(s + B/J)} = \frac{K}{s(s + a)} \tag{7.2}$$

where $K = A/J$ and $a = B/J$. Assume that $a = 2$. Thus

$$G(s) = \frac{C(s)}{E(s)} = \frac{K}{s(s + 2)} \tag{7.3}$$

When the transfer function is expressed with the coefficients of the highest powers of s in both the numerator and the denominator polynomials equal to unity, the value of K is defined as the *static loop sensitivity*. The control ratio (closed-loop transfer function), using Eq. (7.1), is

$$\frac{C(s)}{R(s)} = \frac{K}{s(s + 2) + K} = \frac{K}{s^2 + 2s + K} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{7.4}$$

where $\omega_n = \sqrt{K}$, $\zeta = 1/\sqrt{K}$, and K is considered to be adjustable from zero to an infinite value. The design problem is to determine the roots of the characteristic equation for all values of K and to plot these roots in the s plane. The roots of the characteristic equation, $s^2 + 2s + K = 0$, are given by

$$s_{1,2} = -1 \pm \sqrt{1 - K} \tag{7.5}$$

For $K = 0$, the roots are $s_1 = 0$ and $s_2 = -2$, which also are the poles of the open-loop transfer function given by Eq. (7.3). When $K = 1$, then $s_{1,2} = -1$. Thus, when $0 < K < 1$, the roots $s_{1,2}$ are real and lie on the negative real axis of the s plane between -2 and -1 and 0 to -1, respectively. For the case where $K > 1$, the roots are complex and are given by

$$s_{1,2} = \sigma \pm j\omega_d = -\zeta\omega_n \pm j\omega_n\sqrt{1 - \zeta^2} = -1 \pm j\sqrt{K - 1} \tag{7.6}$$

Note that the real part of all the roots is constant for all values of $K > 1$.

The roots of the characteristic equation are determined for a number of values of K (see Table 7.1) and are plotted in Fig. 7.2. Curves are drawn through these plotted points. On these curves, containing two branches, lie all possible roots of the characteristic equation for all values of K from zero to infinity.

TABLE 7.1 Location of Roots for the Characteristic Equation $s^2 + 2s + K = 0$

K	s_1	s_2
0	$-0 + j0$	$-2.0 - j0$
0.5	$-0.293 + j0$	$-1.707 - j0$
0.75	$-0.5 + j0$	$-1.5 - j0$
1.0	$-1.0 + j0$	$-1.0 - j0$
2.0	$-1.0 + j1.0$	$-1.0 - j10$
3.0	$-1.0 + j1.414$	$-1.0 - j1.414$
50.0	$-1.0 + j7.0$	$-1.0 - j7.0$

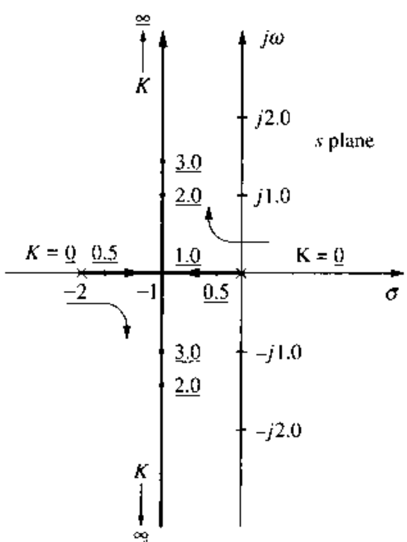


FIGURE 7.2 Plot of all roots of the characteristic equation $s^2 + 2s + K = 0$ for $0 \le K < \infty$. Values of K are underlined.

Note that each branch is calibrated with K as a parameter and the values of K at points on the locus are underlined; the arrows show the direction of increasing values of K. *These curves are defined as the root-locus plot of Eq.* (7.4). Once this plot is obtained, the roots that best fit the system performance specifications can be selected. Corresponding to the selected roots there is a required value of K that can be determined from the plot. When the roots have been selected, the time response can be obtained. Since this process of finding the root locus by calculating the roots for various values of K becomes tedious for characteristic equations of order higher than second, a simpler method of obtaining the root locus would be preferable. The graphical methods for determining the root-locus plot are the subject of this chapter. CAD methods are used to facilitate obtaining the root locus.

The value of K is normally considered to be positive. However, it is possible for K to be negative. For the example in this section, if the value of K is negative, Eq. (7.5) gives only real roots. Thus, the entire locus lies on the real axis, that is, $0 \le s_1 < +\infty$ and $-2 \ge s_2 > -\infty$ for $0 \ge K > -\infty$. For any negative value of K there is always a root in the right half of the s plane, and the system is unstable. Once the root locus has been obtained for a control system, it is possible to determine the variation in system performance with respect to a variation in sensitivity K. For the example of Fig. 7.1, the control ratio is written in terms of its roots, for $K > 1$, as

$$\frac{C(s)}{R(s)} = \frac{K}{(s - \sigma - j\omega_d)(s - \sigma + j\omega_d)} \tag{7.7}$$

Note, as defined in Fig. 4.3, that a root with a damping ratio ζ lies on a line making the angle $\eta = \cos^{-1}\zeta$ with the negative real axis. The damping ratio of several roots is indicated in Fig. 7.3. For an increase in the gain K of the system, analysis of the root locus reveals the following characteristics:

1. A decrease in the damping ratio ζ. This increases the overshoot of the time response.
2. An increase in the undamped natural frequency ω_n. The value of ω_n is the distance from the origin to the complex root.

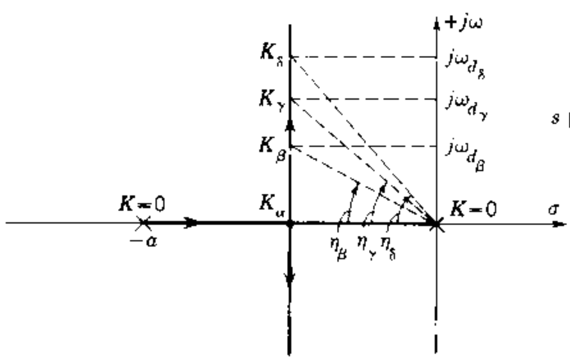


FIGURE 7.3 Root-locus plot of the position-control system of Fig. 7.1. $\eta = \cos^{-1}\zeta$.

3. An increase in the damped natural frequency ω_d. The value of ω_d is the imaginary component of the complex root and is equal to the frequency of the transient response.
4. No effect on the rate of decay σ; that is, it remains constant for all values of gain equal to or greater than K_α. For higher-order systems this is not the case.
5. The root locus is a vertical line for $K \geq K_\alpha$, and $\sigma = -\zeta\omega_n$ is constant. This means that no matter how much the gain is increased in a linear *simple* second-order system, the system can never become unstable. The time response of this system with a step-function input, for $\zeta < 1$, is of the form

$$c(t) = A_0 + A_1 \exp(-\zeta\omega_n t)\sin(\omega_d t + \phi)$$

The root locus of each control system can be analyzed in a similar manner to determine the variation in its time response that results from a variation in its loop sensitivity K.

7.3 QUALITATIVE ANALYSIS OF THE ROOT LOCUS

A zero is added to the simple second-order system of the preceding section so that the open-loop transfer function is

$$G(s) = \frac{K(s + 1/T_2)}{s(s + 1/T_1)} \tag{7.8}$$

The root locus of the control system having this transfer function is shown in Fig. 7.4*b*. Compare this root locus with that of the original system, shown in Fig. 7.4*a* whose open loop transfer function is:

$$G(s) = \frac{K}{s(s + 1/T_1)}$$

It is seen that, in Fig. 7.4*b*, the branches have been "pulled to the left," or farther from the imaginary axis. For values of static loop sensitivity greater than K_α, the roots are farther to the left than for the original system. Therefore, the transients will decay faster, yielding a *more stable* system.

If a pole, instead of a zero, is added to the original system, the resulting transfer function is

$$G(s) = \frac{K}{s(s + 1/T_1)(s + 1/T_3)} \tag{7.9}$$

Figure 7.4*c* shows the corresponding root locus of the closed-loop control system. Note that the addition of a pole has "pulled the locus to the right"

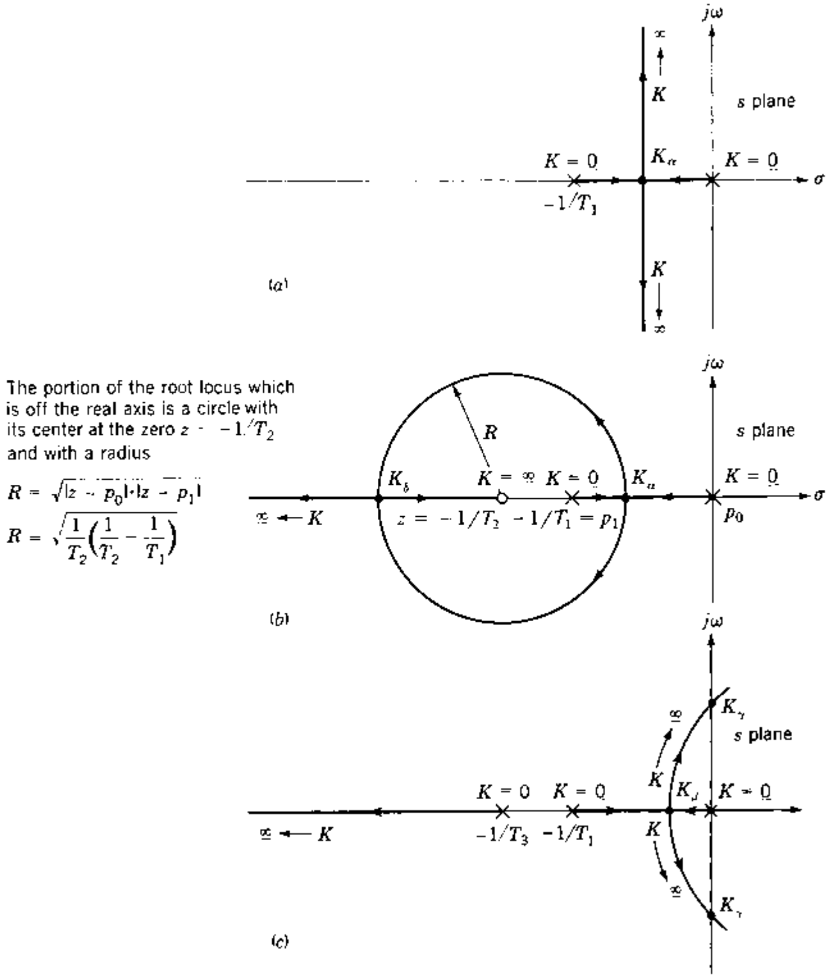


FIGURE 7.4 Various root-locus configurations for $H(s) = 1$: (*a*) root locus of basic transfer function; (*b*) root locus with additional zero; (*c*) root locus with additional pole.

so that two branches cross the imaginary axis. For values of static loop sensitivity greater than K_β, the roots are closer to the imaginary axis than for the original system. Therefore, the transients will decay more slowly, yielding a *less stable* system. Also, for values of $K > K_\gamma$, two of the three roots lie in the right half of the *s* plane, resulting in an unstable system. The addition of the pole has resulted in a less stable system, compared with the original second-order system. Thus, the following general conclusions can be drawn:

1. The addition of a zero to a system has the general effect of pulling the root locus to the left, tending to make it a more stable and faster-responding system (shorter settling time T_s).
2. The addition of a pole to a system has the effect of pulling the root locus to the right, tending to make it a less stable and slower-responding system.

Figure 7.5 illustrates the root-locus configurations for negative-feedback control systems for more complex transfer functions. Note that the third system contains a pole of $G(s)H(s)$ in the right half of the s plane (RHP). It represents the performance of an airplane with an autopilot in the longitudinal mode.

The root-locus method is a graphical technique for readily determining the location of all possible roots of a characteristic equation as the gain is varied from zero to infinity. Also, it can readily be determined how the locus should be altered in order to improve the system's performance, based upon the knowledge of the effect of the addition of poles or zeros.

7.4 PROCEDURE OUTLINE

The procedure to be followed in applying the root-locus method is outlined. This procedure is modified when a CAD program is used to obtain the root locus. Such a program can provide the desired data for the root locus in plotted or tabular form. The procedure outlined next and discussed in the following sections is intended to establish firmly for the reader the fundamentals of the root-locus method and to enhance the interpretation of the data obtained from the CAD program.

Step 1. Derive the open-loop transfer function $G(s)H(s)$ of the system.

Step 2. Factor the numerator and denominator of the transfer function into linear factors of the form $s + a$, where a may be real or complex.

Step 3. Plot the zeros and poles of the open-loop transfer function in the $s = \sigma + j\omega$ plane.

Step 4. The plotted zeros and poles of the open-loop function determine the roots of the characteristic equation of the closed-loop system $[1 + G(s)H(s) = 0]$. Use the geometrical shortcuts summarized in Sec. 7.8 or a CAD program to determine the locus that describes the roots of the closed-loop characteristic equation.

Step 5. Calibrate the locus in terms of the loop sensitivity K. If the gain of the open-loop system is predetermined, the locations of the exact

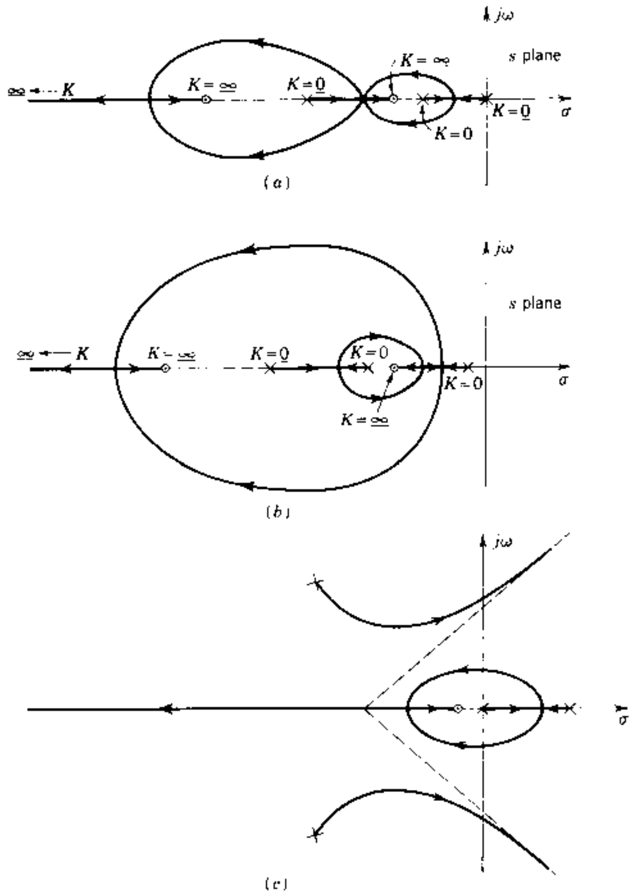


FIGURE 7.5 Various root-locus configurations:

$$(a) \quad G(s)H(s) = \frac{K(s + 1/T_2)(s + 1/T_4)}{s(s + 1/T_1)(s + 1/T_3)}$$

$$(b) \quad G(s)H(s) = \frac{K(s + 1/T_2)(s + 1/T_4)}{(s + 1/T_1)(s + 1/T_3)(s + 1/T_5)}$$

$$(c) \quad G(s)H(s) = \frac{K(s + 1/T_2)}{s(s - 1/T)(s^2 + 2\zeta\omega_n s + \omega_n^2)}$$

These figures are not drawn to scale. Several other root-locus shapes are possible for a given pole-zero arrangement, depending on the specific values of the poles and zeros. Variations of the possible root-locus plots for a given pole-zero arrangement are shown in Ref. 8.

roots of $1 + G(s)H(s)$ can be immediately identified. If the location of the roots is specified, the required value of K can be determined.

Step 6. Once the roots have been found in step 5, the system's time response can be calculated by taking the inverse Laplace transform of $C(s)$, either manually or by use of a CAD program.

Step 7. If the response does not meet the desired specifications, determine the shape that the root locus must have to meet these specifications.

Step 8. If gain adjustment alone does not yield the desired performance, synthesize the cascade compensator that must be inserted into the system. This process, called *compensation*, is described in later chapters.

7.5 OPEN-LOOP TRANSFER FUNCTION

In securing the open-loop transfer function, keep the terms in the factored form of $(s + a)$ or $(s^2 + 2\zeta\omega_n s + \omega_n^2)$. For unity feedback the open-loop function is equal to the forward transfer function $G(s)$. For nonunity feedback it also includes the transfer function of the feedback path. This open-loop transfer function is of the form

$$G(s)H(s) = \frac{K(s + a_1)\cdots(s + a_h)\cdots(s + a_w)}{s^m(s + b_1)(s + b_2)\cdots(s + b_c)\cdots(s + b_u)} \tag{7.10}$$

where a_h and b_c may be real or complex numbers and may lie in either the left-half plane (LHP) or right-half plane (RHP). The value of K may be either positive or negative. For example, consider

$$G(s)H(s) = \frac{K(s + a_1)}{s(s + b_1)(s + b_2)} \tag{7.11}$$

When the transfer function is in this form (with all the coefficients of s equal to unity), the K is defined as the *static loop sensitivity.* For this example, a zero of the open-loop transfer exists at $s = -a_1$ and the poles are at $s = 0$, $s = -b_1$, and $s = -b_2$. Now let the zeros and poles of $G(s)H(s)$ in Eq. (7.10) be denoted by the letters z and p, respectively, i.e.,

$$\begin{array}{llll} z_1 = -a_1 & z_2 = -a_2 & \cdots & z_w = -a_w \\ p_1 = -b_1 & p_2 = -b_2 & \cdots & p_u = -b_u \end{array}$$

Equation (7.10) can then be rewritten as

$$G(s)H(s) = \frac{K(s - z_1)\cdots(s - z_w)}{s^m(s - p_1)\cdots(s - p_u)} = \frac{K\prod_{h=1}^{w}(s - z_h)}{s^m\prod_{c=1}^{u}(s - p_c)} \tag{7.12}$$

where $\prod$ indicates a product of terms. The degree of the numerator is w and that of the denominator is $m + u = n$.

7.6 POLES OF THE CONTROL RATIO $C(s)/R(s)$

The underlying principle of the root-locus method is that the poles of the control ratio $C(s)/R(s)$ are related to the zeros and poles of the open-loop transfer function $G(s)H(s)$ and to the loop sensitivity K. This relationship can be shown as follows. Let

$$G(s) = \frac{N_1(s)}{D_1(s)} \tag{7.13}$$

and

$$H(s) = \frac{N_2(s)}{D_2(s)} \tag{7.14}$$

Then

$$G(s)H(s) = \frac{N_1 N_2}{D_1 D_2} \tag{7.15}$$

Thus,

$$\frac{C(s)}{R(s)} = M(s) = \frac{A(s)}{B(s)} = \frac{G(s)}{1 + G(s)H(s)} = \frac{N_1/D_1}{1 + N_1 N_2/D_1 D_2} \tag{7.16}$$

where

$$B(s) \equiv 1 + G(s)H(s) = 1 + \frac{N_1 N_2}{D_1 D_2} = \frac{D_1 D_2 + N_1 N_2}{D_1 D_2} \tag{7.17}$$

Rationalizing Eq. (7.16) gives

$$\frac{C(s)}{R(s)} = M(s) = \frac{P(s)}{Q(s)} = \frac{N_1 D_2}{D_1 D_2 + N_1 N_2} \tag{7.18}$$

From Eqs. (7.17) and (7.18) it is seen that the zeros of $B(s)$ are the poles of $M(s)$ and determine the form of the system's transient response. In terms of $G(s)H(s)$, given by Eq. (7.12), the degree of $B(s)$ is equal to $m + u$; therefore, $B(s)$ has $n = m + u$ finite zeros. As shown in Chap. 4 on Laplace transforms, the factors of $Q(s)$ produce transient components of $c(t)$ that fall into the categories shown in Table 7.2. The numerator $P(s)$ of Eq. (7.18) merely modifies the constant multipliers of these transient components. The roots of $B(s) = 0$, which is the characteristic equation of the system, must satisfy the equation

$$B(s) \equiv 1 + G(s)H(s) = 0 \tag{7.19}$$

These roots must also satisfy the equation:

$$G(s)H(s) = \frac{K(s - z_1)\cdots(s - z_w)}{s^m(s - p_1)\cdots(s - p_u)} = -1 \tag{7.20}$$

Table 7.2 Time Response Terms in $c(t)$

Denominator factor of $C(s)$	Corresponding inverse	Form
s	$u_{-1}(t)$	Step function
$s + (1/T)$	$e^{-t/T}$	Decaying exponential
$s^2 + 2\zeta\omega_n s + \omega_n^2$	$\exp(-\zeta\omega_n t)\sin(\omega_n\sqrt{1-\zeta^2}t + \phi)$ where $\zeta < 1$	Damped sinusoid

Thus, as the loop sensitivity K assumes values from zero to infinity, the transfer function $G(s)H(s)$ must always be equal to -1. The corresponding values of s that satisfy Eq. (7.20), for any value of K, are the poles of $M(s)$. The plots of these values of s are defined as the root locus of $M(s)$.

Conditions that determine the root locus for *positive values* of loop sensitivity are now determined. The general form of $G(s)H(s)$ for *any* value of s is

$$G(s)H(s) = Fe^{-j\beta} \tag{7.21}$$

The right side of Eq. (7.20), -1, can be written as

$$-1 = e^{j(1+2h)\pi} \qquad h = 0, \pm 1, \pm 2, \ldots$$

Equation (7.20) is satisfied *only* for those values of s for which

$$Fe^{-j\beta} = e^{j(1+2h)\pi}$$

where

$$F = |G(s)H(s)| = 1$$

and

$$-\beta = (1 + 2h)\pi \tag{7.22}$$

From the preceding it can be concluded that the magnitude of $G(s)H(s)$, a function of the complex variable s, must always be unity and its phase angle must be an odd multiple of π if the particular value of s is a zero of $B(s) = 1 + G(s)H(s)$. Consequently, the following two conditions are formalized for the root locus for all positive values of K from zero to infinity:

For $K > 0$

Magnitude condition: $\quad |G(s)H(s)| = 1 \qquad (7.23)$

Angle condition: $\quad \angle G(s)H(s) = (1 + 2h)180^\circ \quad$ for $\quad h = 0, \pm 1, \pm 2, \ldots \qquad (7.24)$

In a similar manner, the conditions for *negative* values of loop sensitivity ($-\infty < K < 0$) can be determined. [This corresponds to positive feedback, $e(t) = r(t) + b(t)$, and positive values of K.] The root locus must satisfy the following conditions.

For $K > 0$

$$\text{Magnitude condition:} \quad |G(s)H(s)| = 1 \tag{7.25}$$

$$\text{Angle condition:} \quad \angle G(s)H(s) = h360^\circ \quad \text{for} \quad h = 0, \pm 1, \pm 2, \ldots \tag{7.26}$$

Thus the root-locus method provides a plot of the variation of each of the poles of $C(s)/R(s)$ in the complex s plane as the loop sensitivity is varied from $K = 0$ to $K = \pm\infty$.

7.7 APPLICATION OF THE MAGNITUDE AND ANGLE CONDITIONS

Once the open-loop transfer function $G(s)H(s)$ has been determined and put into the proper form, the poles and zeros of this function are plotted in the $s = \sigma + j\omega$ plane. As an example, consider

$$G(s)H(s) = \frac{K(s + 1/T_1)^2}{s(s + 1/T_2)(s^2 + 2\zeta\omega_n s + \omega_n^2)} = \frac{K(s - z_1)^2}{s(s - p_1)(s - p_2)(s - p_3)} \tag{7.27}$$

For the quadratic factor $s^2 + 2\zeta\omega_n s + \omega_n^2$ with the damping ratio $\zeta < 1$, the complex-conjugate poles of Eq. (7.27) are

$$P_{2,3} = -\zeta\omega_n \pm j\omega_n\sqrt{1 - \zeta^2} = \sigma \pm j\omega_d$$

The plot of the poles and zeros of Eq. (7.27) is shown in Fig. 7.6. Remember that complex poles or zeros always occur in conjugate pairs, that σ is the damping constant, and that ω_d is the damped natural frequency of oscillation. A multiple pole or zero is indicated on the pole-zero diagram by $\times]_q$ or $\odot]_q$, where $q = 1, 2, 3, \ldots$ is the multiplicity of the pole or zero. The open-loop poles and zeros are plotted, and they are used in the graphical construction of the locus of the poles (the roots of the characteristic equation) of the closed-loop control ratio.

For any particular value (real or complex) of s the terms s, $s - p_1$, $s - p_2$, $s - z_1, \ldots$ are complex numbers designating *directed line segments*. For example, if $s = -4 + j4$ and $p_1 = -1$, then $s - p_1 = -3 + j4$ or $|s - p_1| = 5$, and

$$\phi_1 = \angle s - p_1 = 126.8^\circ \qquad \text{(see Fig. 7.6)} \tag{7.28}$$

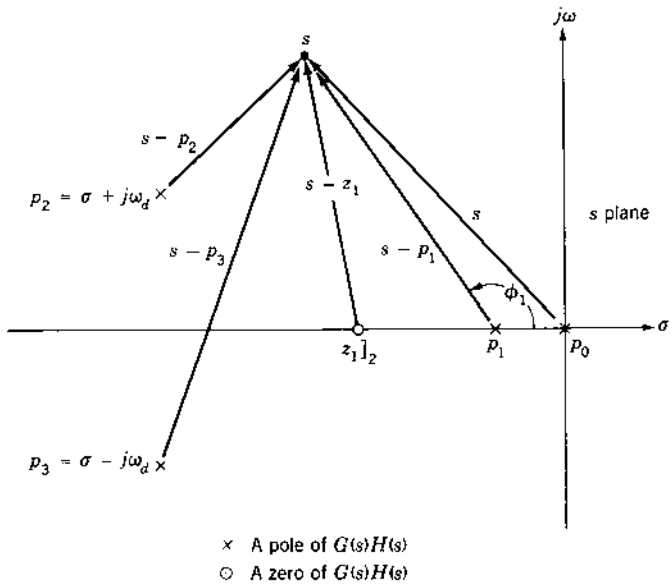


FIGURE 7.6 Pole-zero diagram for Eq. (7.27).

Recall that it is shown in Sec. 7.6 that the roots of the characteristic equation $1 + G(s)H(s) = 0$ are all the values of s that satisfy the two conditions:

$$|G(s)H(s)| = 1 \tag{7.29}$$

$$\underline{/G(s)H(s)} = \begin{cases} (1+2h)180^\circ & \text{for } K > 0 \\ h360^\circ & \text{for } K < 0 \end{cases} \qquad h = 0, \pm 1, \pm 2, \ldots \tag{7.30}$$

These are labeled as the magnitude and angle conditions, respectively. Therefore, applying these two conditions to the general Eq. (7.12) results in

$$\frac{|K| \cdot |s - z_1| \cdots |s - z_w|}{|s^m| \cdot |s - p_1| \cdot |s - p_2| \cdots |s - p_u|} = 1 \tag{7.31}$$

and

$$\begin{aligned} -\beta &= \underline{/s - z_1} + \cdots + \underline{/s - z_w} - m\angle s - \underline{/s - p_1} - \cdots - \underline{/s - P_u} \\ &= \begin{cases} (1+2h)180^\circ & \text{for } K > 0 \\ h360^\circ & \text{for } K < 0 \end{cases} \end{aligned} \tag{7.32}$$

Solving Eq. (7.31) for $|K|$ gives

$$|K| = \frac{|s^m| \cdot |s - p_1| \cdot |s - p_2| \cdots |s - p_u|}{|s - z_1| \cdots |s - z_w|} = \text{loop sensitivity} \tag{7.33}$$

Multiplying Eq. (7.32) by -1 gives

$$\beta = \sum(\text{angles of denominator terms}) - \sum(\text{angles of numerator terms})$$
$$= \begin{cases} (1+2h)180^\circ & \text{for } K > 0 \\ h360^\circ & \text{for } K < 0 \end{cases} \tag{7.34}$$

All angles are considered positive, measured in the counterclockwise (CCW) sense. This form is convenient since $G(s)H(s)$ usually has more poles than zeros. Equations (7.33) and (7.34) are in the form used in the graphical construction of the root locus. In other words, there are particular values of s for which $G(s)H(s)$ satisfies the angle condition. For these values of s, Eq. (7.33) is used to determine the corresponding magnitude K. Those values of s that satisfy both the angle and the magnitude conditions are the roots of the characteristic equation and are $n = m + u$ in number. Thus, corresponding to step 4 in Sec. 7.4, the locus of all possible roots is obtained by applying the angle condition. This root locus can be calibrated in terms of the loop sensitivity K by using the magnitude condition.

Example. Determine the locus of all the closed-loop poles of $C(s)/R(s)$, for

$$G(s)H(s) = \frac{K_0(1 + 0.25s)}{(1+s)(1+0.5s)(1+0.2s)} \tag{7.35}$$

$$G(s)H(s) = \frac{K(s+4)}{(s+1)(s+2)(s+5)} \qquad K = 2.5K_0 \tag{7.36}$$

Step 1. The poles and zeros are plotted in Fig. 7.7.

Step 2. In Fig. 7.7 the ϕ's are denominator angles and the ψ's are numerator angles to a search point. Also, the l's are the lengths of the directed segments stemming from the denominator factors, and (l)'s are the lengths of the directed segments stemming from the numerator factors. After plotting the poles and zeros of the open-loop transfer functions, arbitrarily choose a search point. To this

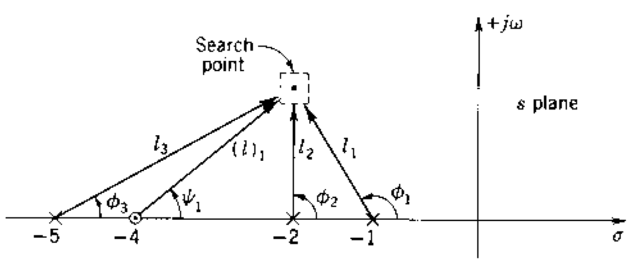


FIGURE 7.7 Construction of the root locus.

point, draw directed line segments from all the open-loop poles and zeros and label as indicated. For this search point to be a point on the root locus, the following angle condition must be true:

$$\beta = \phi_1 + \phi_2 + \phi_3 - \psi_1 = \begin{cases} (1+2h)180^\circ & \text{for } K > 0 \\ h360^\circ & \text{for } K < 0 \end{cases} \tag{7.37}$$

If this equation is not satisfied, select another search point until it is satisfied. Locate a sufficient number of points in the s plane that satisfy the angle condition. In the next section additional information is given that lessens and systematizes the work involved in this trial-and-error approach.

Step 3. Once the complete locus has been determined, the locus can be calibrated in terms of the loop sensitivity for any root s_1 as follows:

$$|K| = \frac{l_1 l_2 l_3}{(l)_1} \tag{7.38}$$

where $l_1 = |s_1 + 1|$; $l_2 = |s_1 + 2|$; $l_3 = |s_1 + 5|$; and $(l)_1 = |s_1 + 4|$. In other words, the values of l_1, l_2, l_3, and $(l)_1$ can be measured for a given point s_1 that satisfies the angle condition; thus, the value of $|K|$ for this point can be calculated. The appropriate sign must be given to the magnitude of K, compatible with the particular angle condition that is utilized to obtain the root locus. Note that since complex roots must occur in conjugate pairs, the locus is symmetrical about the real axis. Thus, the bottom half of the locus can be drawn once the locus above the real axis has been determined. The root locus for this system is shown in Fig. 7.8. Note that for negative values of K, the three branches of the root locus lie in the left half plane for $K > -2.5$.

$W(s)$ Plane

From Eqs. (7.36) and (7.20),

$$W(s) = u_x + jv_y = \frac{(s+1)(s+2)(s+5)}{s+4} = -K \tag{7.39}$$

The line $u_x = -K$ in the $W(s)$ plane maps into the curves indicated in Fig. 7.8. That is, for each value of u_x in the $W(s)$ plane there is a particular value or a set of values of s in the s plane.

7.8 GEOMETRICAL PROPERTIES (CONSTRUCTION RULES)

To facilitate the application of the root-locus method, the following rules are established for $K > 0$. These rules are based upon the interpretation of the

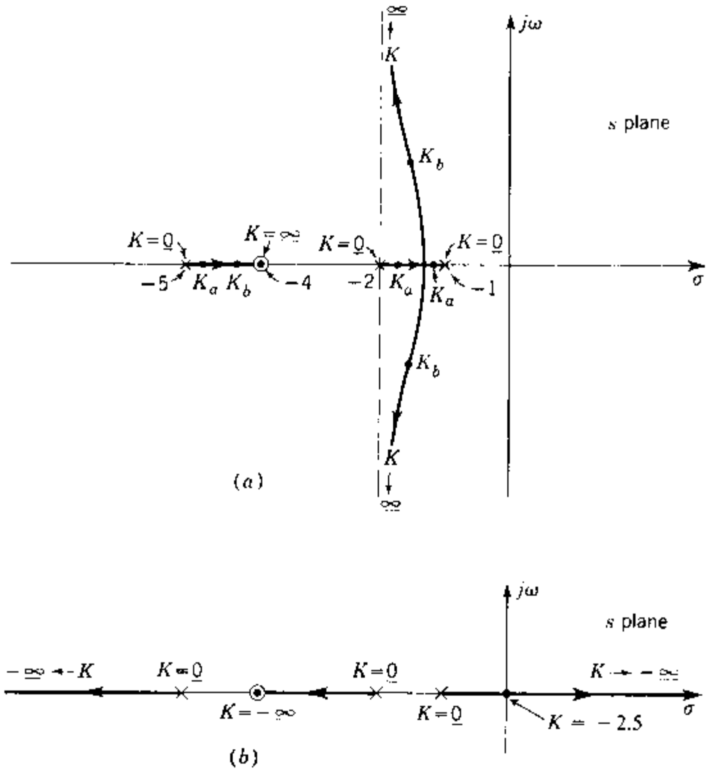


FIGURE 7.8 The complete root locus of Eq. (7.36): (*a*) for $K > 0$; (*b*) for $K < 0$.

angle condition and an analysis of the characteristic equation. These rules can be extended for the case where $K < 0$. The rules for both $K > 0$ and $K < 0$ are listed in Sec. 7.16 for easy reference. The rules presented aid in obtaining the root locus by expediting the plotting of the locus. The root locus can also be obtained by using the MATLAB CAD program. These rules provide *checkpoints* to ensure that the computer solution is correct. They also permit rapid sketching of the root locus, which provides a qualitative idea of achievable closed-loop system performance.

Rule 1: Number of Branches of the Locus

The characteristic equation $B(s) = 1 + G(s)H(s) = 0$ is of degree $n = m + u$; therefore, there are n roots. As the open-loop sensitivity K is varied from zero to infinity, each root traces a continuous curve. Since there are n roots, there are the same number of curves or branches in the complete root locus.

Since the degree of the polynomial $B(s)$ is determined by the poles of the open-loop transfer function, *the number of branches of the root locus is equal to the number of poles of the open-loop transfer function.*

Rule 2: Real-Axis Locus

In Fig. 7.9 are shown a number of open-loop poles and zeros. If the angle condition is applied to any search point such as s_1 on the real axis, the angular contribution of all the poles and zeros on the real axis to the left of this point is zero. The angular contribution of the complex-conjugate poles to this point is 360°. (This is also true for complex-conjugate zeros.) Finally, the poles and zeros on the real axis to the right of this point each contribute 180° (with the appropriate sign included). From Eq. (7.34) the angle of $G(s)H(s)$ to the point s_1 is given by

$$\phi_0 + \phi_1 + \phi_2 + \phi_3 + [(\phi_4)_{+j} + (\phi_4)_{-j}] - (\psi_1 + \psi_2) = (1 + 2h)180^\circ \tag{7.40}$$

or

$$180^\circ + 0^\circ + 0^\circ + 0^\circ + 360^\circ - 0^\circ - 0^\circ = (1 + 2h)180^\circ \tag{7.41}$$

Therefore, s_1 is a point on a branch of the locus. Similarly, it can be shown that the point s_2 is not a point on the locus. The poles and zeros to the left of a point s on the real axis and the 360° contributed by the complex-conjugate poles or zeros do not affect the odd-multiple-of-180° requirement. Thus, *if the total number of real poles and zeros to the right of a search point s on the real axis is odd, this point lies on the locus.* In Fig. 7.9 the root locus exists on the real axis from p_0 to p_1, z_1 to p_2, and p_3 to z_2.

All points on the real axis between z_1 and p_2 in Fig. 7.9 satisfy the angle condition and are therefore points on the root locus. However, there is

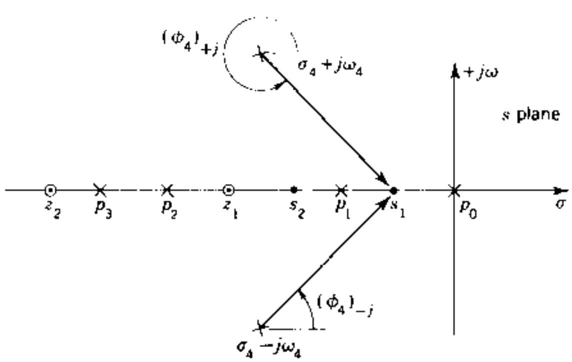


FIGURE 7.9 Determination of the real-axis locus.

no guarantee that this section of the real axis is part of just one branch. Figure 7.5*b* and Prob. 7.4 illustrate the situation where part of the real axis between a pole and a zero is divided into three sections that are parts of three different branches.

Rule 3: Locus End Points

The magnitude of the loop sensitivity that satisfies the magnitude condition is given by Eq. (7.33) and has the general form

$$|W(s)| = K = \frac{\prod_{c=1}^{n} |s - p_c|}{\prod_{h=1}^{w} |s - z_h|} \tag{7.42}$$

Since the numerator and denominator factors of Eq. (7.42) locate the poles and zeros, respectively, of the open-loop transfer function, the following conclusions can be drawn:

1. When $s = p_c$ (the open-loop poles), the loop sensitivity K is zero.
2. When $s = z_h$ (the open-loop zeros), the loop sensitivity K is infinite. When the numerator of Eq. (7.42) is of higher order than the denominator, then $s = \infty$ also makes K infinite, thus being equivalent in effect to a zero.

Thus, the locus starting points ($K = 0$) are at the open-loop poles and the locus ending points ($K = \infty$) are at the open-loop zeros (the point at infinity being considered as an equivalent zero of multiplicity equal to the quantity $n - w$).

Rule 4: Asymptotes of Locus as *s* Approaches Infinity

Plotting of the locus is greatly facilitated by evaluating the asymptotes approached by the various branches as s takes on large values. Taking the limit of $G(s)H(s)$ as s approaches infinity, based on Eqs. (7.12) and (7.20), yields

$$\lim_{s\to\infty} G(s)H(s) = \lim_{s\to\infty}\left[K\frac{\prod_{h=1}^{w}(s - z_h)}{\prod_{c=1}^{n}(s - p_c)}\right] = \lim_{s\to\infty}\frac{K}{s^{n-w}} = -1 \tag{7.43}$$

Remember that K in Eq. (7.43) is still a variable in the manner prescribed previously, thus allowing the magnitude condition to be met. Therefore, as $s \to \infty$,

$$-K = s^{n-w} \tag{7.44}$$

$$|-K| = |s^{n-w}| \qquad \text{Magnitude condition} \tag{7.45}$$

$$\angle -K = \angle s^{n-w} = (1 + 2h)180^\circ \qquad \text{Angle condition} \tag{7.46}$$

Rewriting Eq. (7.46) gives $(n - w)\angle s = (1 + 2h)180^\circ$ or

$$\gamma = \frac{(1 + 2h)180^\circ}{n - w} \qquad \text{as } s \to \infty \tag{7.47}$$

There are $n - w$ asymptotes of the root locus, and their angles are given by

$$\gamma = \frac{(1+2h)180^\circ}{[\text{number of poles of } G(s)H(s)] - [\text{number of zeros of } G(s)H(s)]} \tag{7.48}$$

Equation (7.48) reveals that, no matter what magnitude s may have, after a sufficiently large value has been reached, the argument (angle) of s on the root locus remains constant. For a search point that has a sufficiently large magnitude, the open-loop poles and zeros appear to it as if they had collapsed into a single point. Therefore, the branches are asymptotic to straight lines whose slopes and directions are given by Eq. (7.48) (see Fig. 7.10). These asymptotes usually do not go through the origin. The correct real-axis intercept of the asymptotes is obtained from Rule 5.

Rule 5: Real-Axis Intercept of the Asymptotes

The real-axis crossing σ_o of the asymptotes can be obtained by applying the theory of equations. The result is

$$\sigma_o = \frac{\sum_{c=1}^{n} \mathrm{Re}(p_c) - \sum_{h=1}^{w} \mathrm{Re}(z_h)}{n - w} \tag{7.49}$$

The asymptotes are not dividing lines, and a locus may cross its asymptote. It may be valuable to know from which side the root locus approaches its asymptote. Lorens and Titsworth [3] present a method for obtaining this information. The locus lies exactly along the asymptote if the pole-zero pattern is symmetric about the asymptote line extended through the point σ_o.

Rule 6: Breakaway Point on the Real Axis [4]

The branches of the root locus start at the open-loop poles where $K = 0$ and end at the finite open-loop zeros or at $s = \infty$. When the root locus has branches

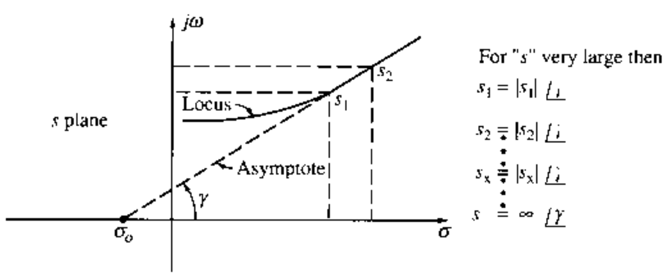


FIGURE 7.10 Asymptotic condition for large values of s.

on the real axis between two poles, there must be a point at which the two branches *breakaway* from the real axis and enter the complex region of the s plane in order to approach zeros or the point at infinity. (Examples are shown in Fig. 7.11*a*-3: between p_0 and p_1, and in Fig. 7.11*b*-2: between p_2 and p_3.) For two finite zeros (see Fig. 7.11*b*-1) or one finite zero and one at infinity (see Fig. 7.11*a*-1) the branches are coming from the complex region and enter the real axis.

In Fig. 7.11*a*-3 between two poles there is a point s_a for which the loop sensitivity K_z is greater than for points on either side of s_a on the real axis. In other words, since K starts with a value of zero at the poles and increases in value as the locus moves away from the poles, there is a point somewhere in between where the K's for the two branches simultaneously reach a maximum value. This point is called the *breakaway point.* Plots of K vs. σ utilizing Eq. (7.33) are shown in Fig. 7.11 for the portions of the root locus that exist

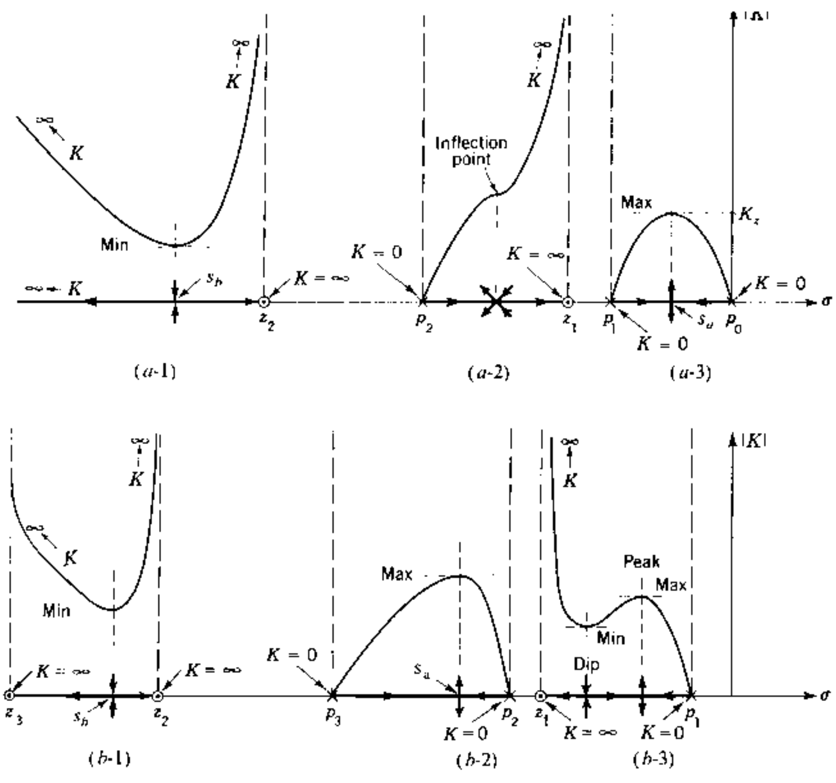


FIGURE 7.11 Plots of K and the corresponding real-axis locus for (*a*) Fig. 7.5*a* and (*b*) Fig. 7.5*b*.

on the real axis for $K > 0$. The point s_b for which the value of K is a minimum between two zeros is called the *break-in point*. The breakaway and break-in points can easily be calculated for an open-loop pole-zero combination for which the derivatives of $W(s) = -K$ is of the second order. As an example, if

$$G(s)H(s) = \frac{K}{s(s+1)(s+2)} \tag{7.50}$$

then

$$W(s) = s(s+1)(s+2) = -K \tag{7.51}$$

Multiplying the factors together gives

$$W(s) = s^3 + 3s^2 + 2s = -K \tag{7.52}$$

When $s^3 + 3s^2 + 2s$ is a minimum, $-K$ is a minimum and K is a maximum. Thus, by taking the derivative of this function and setting it equal to zero, the points can be determined:

$$\frac{dW(s)}{ds} = 3s^2 + 6s + 2 = 0 \tag{7.53}$$

or

$$s_{a,b} = -1 \pm 0.5743 = -0.4257, \ -1.5743$$

Since the breakaway points s_a for $K > 0$ must lie between $s = 0$ and $s = -1$ in order to satisfy the angle condition, the value is

$$s_a = -0.4257 \tag{7.54}$$

The other point, $s_b = -1.5743$, is the break-in point on the root locus for $K < 0$. Substituting $s_a = -0.4257$ into Eq. (7.52) gives the value of K at the breakaway point for $K > 0$ as

$$K = -[(-0.426)^3 + (3)(-0.426)^2 + (2)(-0.426)] = 0.385 \tag{7.55}$$

When the derivative of $W(s)$ is of higher order than 2, a digital-computer program can be used to calculate the roots of the numerator polynomial of $dW(s)/ds$; these roots locate the breakaway and break-in points. Note that it is possible to have both a breakaway and a break-in point between a pole and zero (finite or infinite) on the real axis, as shown in Figs. 7.5, 7.11*a*-2, and 7.11*b*-3. The plot of K vs. σ for a locus between a pole and zero falls into one of the following categories:

1. The plot clearly indicates a peak and a dip, as illustrated between p_1 and z_1 in Fig. 7.11*b*-3. The peak represents a "maximum" value of K that identifies a break-in point.

2. The plot contains an inflection point. This occurs when the break-away and break-in points coincide, as is the case between p_2 and z_1 in Fig. 7.11*a*-2.
3. The plot does not indicate a dip-and-peak combination or an inflection point. For this situation there are no break-in or break-away points.

Rule 7: Complex Pole (or Zero): Angle of Departure

The next geometrical shortcut is the rapid determination of the direction in which the locus leaves a complex pole or enters a complex zero. Although in Fig. 7.12 a complex pole is considered, the results also hold for a complex zero.

In Fig. 7.12*a*, an area about p_2 is chosen so that l_2 is very much smaller than l_0, l_1, l_3, and $(l)_1$. For illustrative purposes, this area has been enlarged many times in Fig. 7.12*b*. Under these conditions the angular contributions from all the other poles and zeros, except p_2, to a search point anywhere in this area are approximately constant. They can be considered to have values determined as if the search point were right at p_2. Applying the angle condition to this small area yields

$$\phi_0 + \phi_1 + \phi_2 + \phi_3 - \psi_1 = (1 + 2h)180^\circ \tag{7.56}$$

or the departure angle is

$$\phi_{2_D} = (1 + 2h)180^\circ - (\phi_0 + \phi_1 + 90^\circ - \psi_1)$$

In a similar manner the approach angle to a complex zero can be determined. For an open-loop transfer function having the pole-zero

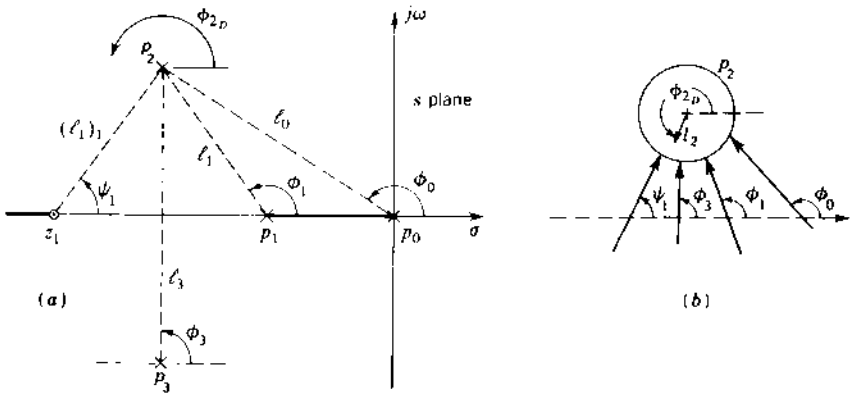


FIGURE 7.12 Angle condition in the vicinity of a complex pole.

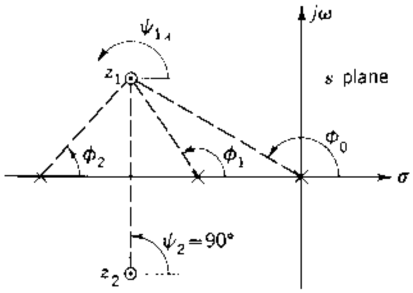


FIGURE 7.13 Angle condition in the vicinity of a complex zero.

arrangement shown in Fig. 7.13, the approach angle ψ_1 to the zero z_1 is given by

$$\psi_{1_A} = (\phi_0 + \phi_1 + \phi_2 - 90^\circ) - (1 + 2h)180^\circ \tag{7.57}$$

In other words, the direction of the locus as it leaves a pole or approaches a zero can be determined by adding up, according to the angle condition, all the angles of all vectors from all the other poles and zeros to the pole or zero in question. Subtracting this sum from $(1 + 2h)180^\circ$ gives the required direction.

Rule 8: Imaginary-Axis Crossing Point

In cases where the locus crosses the imaginary axis into the right-half s plane, the crossover point can usually be determined by Routh's method or by similar means. For example, if the closed-loop characteristic equation $D_1D_2 + N_1N_2 = 0$ is of the form

$$s^3 + bs^2 + cs + Kd = 0$$

the Routhian array is

$$\begin{array}{c|ll} s^3 & 1 & c \\ s^2 & b & Kd \\ s^1 & (bc - Kd)/b & \\ s^0 & Kd & \end{array}$$

An undamped oscillation may exist if the s^{-1} row in the array equals zero. For this condition the *auxiliary equation* obtained from the s^2 row is

$$bs^2 + Kd = 0 \tag{7.58}$$

and its roots are

$$s_{1,2} = \pm j\sqrt{\frac{Kd}{b}} = \pm j\omega_n \tag{7.59}$$

The loop sensitivity term K is determined by setting the s^1 row to zero:

$$K = \frac{bc}{d} \tag{7.60}$$

For $K > 0$, Eq. (7.59) gives the natural frequency of the undamped oscillation. This corresponds to the point on the imaginary axis where the locus crosses over into the right-half s plane. The imaginary axis divides the s plane into stable and unstable regions. Also, the value of K from Eq. (7.60) determines the value of the loop sensitivity at the crossover point. For values of $K < 0$ the term in the s^0 row is negative, thus characterizing an unstable response. The limiting values for a stable response are therefore

$$0 < K < \frac{bc}{d} \tag{7.61}$$

In like manner, the crossover point can be determined for higher-order characteristic equations. For these higher-order systems care must be exercised in *analyzing all terms in the first column that contain the term K* in order to obtain the correct range of values of gain for stability.

Rule 9: Intersection or Nonintersection of Root-Locus Branches [5]

The theory of complex variables yields the following properties:

1. A value of s that satisfies the angle condition of Eq. (7.34) is a point on the root locus. If $dW(s)/ds \neq 0$ at this point, there is one and only one branch of the root locus through the point.
2. If the first $y-1$ derivatives of $W(s)$ vanish at a given point on the root locus, there are y branches approaching and y branches leaving this point; thus, there are root-locus intersections at this point. The angle between two adjacent *approaching* branches is given by

   $$\lambda_y = \pm\frac{360^\circ}{y} \tag{7.62}$$

   Also, the angle between a branch *leaving* and an adjacent branch that is *approaching* the same point is given by

   $$\theta_y = \pm\frac{180^\circ}{y} \tag{7.63}$$

Figure 7.14 illustrates these angles at $s = -3$, with $\theta_y = 45^\circ$ and $\lambda_y = 90^\circ$.

Rule 10: Conservation of the Sum of the System Roots

The technique described by this rule aids in the determination of the general shape of the root locus. Consider the general open-loop transfer function

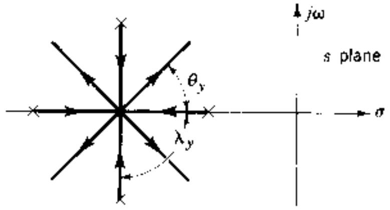


FIGURE 7.14 Root locus for $G(s)H(s) = \dfrac{K}{(s+2)(s+4)(s^2+6s+10)}$.

in the form

$$G(s)H(s) = \frac{K\prod_{h=1}^{w}(s-z_h)}{s^m\prod_{c=1}^{u}(s-p_c)} \tag{7.64}$$

Recalling that for physical systems $w \le n = u + m$, the denominator of $C(s)/R(s)$ can be written as

$$B(s) = 1 + G(s)H(s) = \frac{\prod_{j=1}^{n}(s-r_j)}{s^m\prod_{c=1}^{u}(s-p_c)} \tag{7.65}$$

where r_j are the roots described by the root locus.

By substituting from Eq. (7.64) into Eq. (7.65) and equating numerators on each side of the resulting equation, the result is

$$s^m\prod_{c=1}^{u}(s-p_c) + K\prod_{h=1}^{w}(s-z_h) = \prod_{j=1}^{n}(s-r_j) \tag{7.66}$$

Expanding both sides of this equation gives

$$\begin{aligned}&\left(s^n - \sum_{c=1}^{u}p_c s^{n-1} + \cdots\right) + K\left(s^w - \sum_{h=1}^{w}z_h s^{w-1} + \cdots\right)\\ &\quad = s^n - \sum_{j=1}^{n}r_j s^{n-1} + \cdots\end{aligned} \tag{7.67}$$

For those open-loop transfer functions in which $w \le n-2$, the following is obtained by equating the coefficients of s^{n-1} of Eq. (7.67):

$$\sum_{c=1}^{u}p_c = \sum_{j=1}^{n}r_j$$

Since m open-loop poles have values of zero, this equation can also be written as

$$\sum_{j=1}^{n}p_j = \sum_{j=1}^{n}r_j \tag{7.68}$$

where p_j now represents all the open-loop poles, including those at the origin, and r_j are the roots of the characteristic equation. This equation reveals that as the system gain is varied from zero to infinity, the sum of the system roots is constant. In other words, the sum of the system roots is conserved and is independent of K. When a system has several root-locus branches that go to infinity (as $K \to \infty$), the directions of the branches are such that the sum of the roots is constant. A branch going to the right therefore requires that there will be a branch going to the left. The root locus of Fig. 7.2 satisfies the conservancy law for the root locus. The sum of the roots is a constant for all values of K.

For a unity-feedback system, Rao [6] has shown that the closed-loop pole and zero locations of

$$\frac{C(s)}{R(s)} = \frac{K(s - z_1)\cdots(s - z_w)}{(s - p_1)\cdots(s - p_n)} \tag{7.69}$$

satisfy the relation

$$\sum_{i=1}^{w}(-z_i)^{-q} = \sum_{j=1}^{n}(-p_j)^{-q} \qquad q = 1, 2, \ldots, m - 1 \tag{7.70}$$

for $m > 1$, that is, for a Type 2 (or higher) system. This serves as a check on the accuracy of the root determination.

Rule 11: Determination of Roots on the Root Locus

After the root locus has been plotted, the specifications for system performance are used to determine the dominant roots. The *dominant branch* is the branch that is the closest to the imaginary axis. This root has the largest influence on the time response. When this branch yields complex dominant roots, the time response is oscillatory and the figures of merit (see Sec. 3.10) are the peak overshoot M_p, the time t_p at which the peak overshoot occurs, and the settling time t_s. An additional figure of merit, the gain K_m, significantly affects the steady-state error (see Chap. 6). All of these quantities can be used to select the *dominant roots*. Thus, the designer may use the damping ratio ζ, the undamped natural frequency ω_n, the damped natural frequency ω_d, the damping coefficient σ, or the gain K_m to select the dominant roots. When the dominant roots are selected, the required loop sensitivity can be determined by applying the magnitude condition, as shown in Eq. (7.33). The remaining roots on each of the other branches can be determined by any of the following methods:

Method 1. Determine the point on each branch of the locus that satisfies the same value of loop sensitivity as for the dominant roots.

Method 2. If all except one real or a complex pair of roots are known, either of the following procedures can be used.

Procedure 1. Divide the characteristic polynomial by the factors representing the known roots. The remainder gives the remaining roots.

Procedure 2. Equation (7.68) can be used to find some of the roots. A necessary condition is that the denominator of $G(s)H(s)$ be at least of degree 2 higher than the numerator. If all the roots except one real root are known, application of Eq. (7.68) yields directly the value of the real root. However, for complex roots of the form $r = \sigma \pm j\omega_d$ it yields only the value of the real component σ.

CAD programs are available that yield the loop sensitivity and the roots of the characteristic equation of the system. When the damping ratio is specified for the dominant roots, these roots determine the value of K and all the remaining roots.

7.9 CAD ACCURACY CHECKS (CADAC)

The first seven geometrical construction rules, which can be readily evaluated, should be used as *CADAC* to assist in the validation of the root locus data being obtained from a CAD package.

7.10 ROOT LOCUS EXAMPLE

Find $C(s)/R(s)$ with $\zeta = 0.5$ for the dominant roots (roots closest to the imaginary axis) for the feedback control system represented by

$$G(s) = \frac{K_1}{s(s^2/2600 + s/26 + 1)} \qquad \text{and} \qquad H(s) = \frac{1}{0.04s + 1}$$

Rearranging gives

$$G(s) = \frac{2600K_1}{s(s^2 + 100s + 2600)} = \frac{N_1}{D_1} \qquad \text{and} \qquad H(s) = \frac{25}{s + 25} = \frac{N_2}{D_2}$$

Thus,

$$G(s)H(s) = \frac{65{,}000K_1}{s(s+25)(s^2+100s+2600)} = \frac{K}{s^4 + 125s^3 + 5100s^2 + 65{,}000s}$$

where $K = 65{,}000K_1$.

1. The poles of $G(s)H(s)$ are plotted on the s plane in Fig. 7.15; the values of these poles are $s = 0, -25, -50 + j10, -50 - j10$.

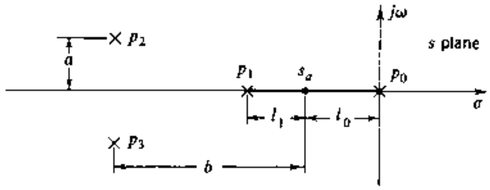


FIGURE 7.15 Location of the breakaway point.

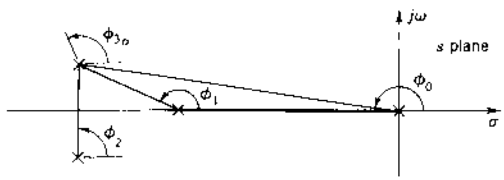


FIGURE 7.16 Determination of the departure angle.

The system is completely unstable for $K < 0$. Therefore, this example is solved only for the condition $K > 0$.

2. There are four branches of the root locus.
3. The locus exists on the real axis between 0 and -25.
4. The angles of the asymptotes are

$$\gamma = \frac{(1 + 2h)180^\circ}{4} = \pm 45^\circ,\ \pm 135^\circ$$

5. The real-axis intercept of the asymptotes is

$$\sigma_o = \frac{0 - 25 - 50 - 50}{4} = -31.25$$

6. The breakaway point s_a (see Fig. 7.15) on the real axis between 0 and -25 is found by solving $dW(s)/ds = 0$ [see Eq. (7.39)]:

$$-K = s^4 + 125s^3 + 5100s^2 + 65{,}000s$$
$$\frac{d(-K)}{ds} = 4s^3 + 375s^2 + 10{,}200s + 65{,}000 = 0$$
$$s_a = -9.15$$

7. The angle of departure ϕ_{3_D} (see Fig. 7.16) from the pole $-50 + j10$ is obtained from

$$\phi_0 + \phi_1 + \phi_2 + \phi_{3_D} = (1 + 2h)180^\circ$$
$$168.7^\circ + 158.2^\circ + 90^\circ + \phi_{3_D} = (1 + 2h)180^\circ$$
$$\phi_{3_D} = 123.1^\circ$$

Similarly, the angle of departure from the pole $-50 + j10$ is -123.1°.

8. The imaginary-axis intercepts are obtained from

$$\frac{C(s)}{R(s)} = \frac{2600K_1(s+25)}{s^4 + 125s^3 + 5100s^2 + 65{,}000s + 65{,}000K_1} \tag{7.71}$$

The Routhian array for the denominator of $C(s)/R(s)$, which is the characteristic polynomial, is (see Sec. 6.2, Theorem 1)

$$\begin{array}{l|lll} s^4 & 1 & 5100 & 65{,}000K_1 \\ s^3 & 1 & 520 \text{ (after division by 125)} & \\ s^2 & 1 & 14.2K_1 \text{ (after division by 4580)} & \\ s^1 & 520 - 14.2K_1 & & \\ s^0 & 14.2K_1 & & \end{array}$$

Pure imaginary roots exist when the s^1 row is zero. This occurs when $K_1 = 520/14.2 = 36.6$. The auxiliary equation is formed from the s^2 row:

$$s^2 + 14.2K_1 = 0$$

and the imaginary roots are

$$s = \pm j\sqrt{14.2K_1} = \pm j\sqrt{520} = \pm j\,22.8$$

9. Additional points on the root locus are found by locating points that satisfy the angle condition

$$\underline{/s} + \underline{/s+25} + \underline{/s+50-j10} + \underline{/s+50+j10} = (1+2m)180^\circ$$

The root locus is shown in Fig. 7.17.

10. The radial line for $\zeta = 0.5$ is drawn on the graph of Fig. 7.17 at the angle (see Fig. 4.3 for definition of η)

$$\eta = \cos^{-1} 0.5 = 60^\circ$$

The dominant roots obtained from the graph are

$$s_{1,2} = -6.6 \pm j11.4$$

11. The gain is obtained from the expression

$$K = 65{,}000K_1 = |s| \cdot |s+25| \cdot |s+50-j10| \cdot |s+50+j10|$$

Inserting the value $s_1 = -6.6 + j11.4$ into this equation yields

$$K = 65{,}000K_1 = 598{,}800$$
$$K_1 = 9.25$$

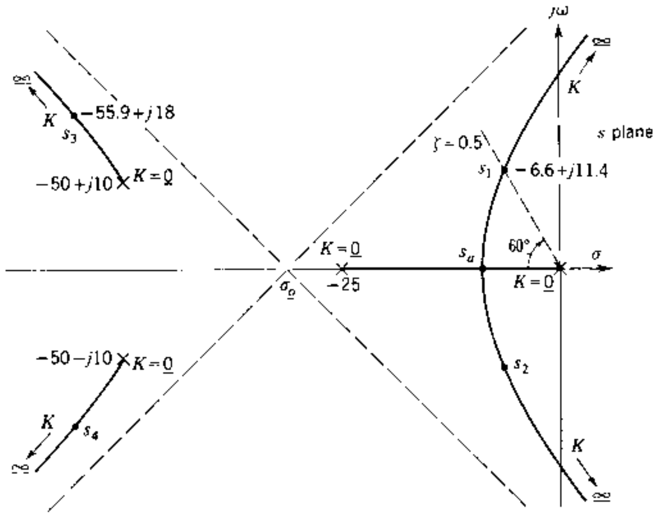


FIGURE 7.17 Root locus for $G(s)H(s) = \dfrac{65{,}000K_1}{s(s+25)(s^2+100s+2600)}$.

12. The other roots are evaluated to satisfy the magnitude condition $K = 598{,}800$. The remaining roots of the characteristic equation are

$$s_{3,4} = -55.9 \pm j18.0$$

The real part of the additional roots can also be determined by using the rule from Eq. (7.68):

$$\begin{aligned} &0 - 25 + (-50 + j10) + (-50 - j10) \\ &\quad = (-6.6 + j11.4) + (-6.6 - j11.4) + (\sigma + j\omega_d) + (\sigma - j\omega_d) \end{aligned}$$

This gives

$$\sigma = -55.9$$

By using this value, the roots can be determined from the root locus as $-55.9 \pm j18.0$.

13. The control ratio, using values of the roots obtained in steps 10 and 12, is

$$\frac{C(s)}{R(s)} = \frac{N_1 D_2}{\text{factors determined from root locus}}$$

$$
\begin{aligned}
&= \frac{1}{(s+6.6+j11.4)(s+6.6-j11.4)} \\
&\quad \times \frac{24{,}040(s+25)}{(s+55.9+j18)(s+55.9-j18)} \\
&= \frac{24{,}040(s+25)}{(s^2+13.2s+173.5)(s^2+111.8s+3450)}
\end{aligned}
$$

14. The response $c(t)$ for a unit step input is found from

$$
\begin{aligned}
C(s) &= \frac{24{,}040(s+25)}{s(s^2+13.2s+173.5)(s^2+111.8s+3450)} \\
&= \frac{A_0}{s} + \frac{A_1}{s+6.6-j11.4} + \frac{A_2}{s+6.6+j11.4} \\
&\quad + \frac{A_3}{s+55.9-j18} + \frac{A_4}{s+55.9+j18}
\end{aligned}
$$

The constants are

$$
A_0 = 1.0 \qquad A_1 = 0.604\,\angle{-201.7^\circ} \qquad A_3 = 0.14\,\angle{-63.9^\circ}
$$

Note that A_0 must be exactly 1, since $G(s)$ is Type 1 and the gain of $H(s)$ is unity. It can also be obtained from Eq. (7.71) for $C(s)/R(s)$ in step 8. Inserting $R(s) = 1/s$ and finding the final value gives $c(t)_{ss} = 1.0$. The response $c(t)$ is

$$
\begin{aligned}
c(t) &= 1 + 1.21e^{-6.6t}\sin(11.4t - 111.7^\circ) \\
&\quad + 0.28e^{-55.9t}\sin(18t + 26.1^\circ)
\end{aligned} \tag{7.72}
$$

A plot of $c(t)$ is shown in Figs. 7.19 and 7.26.

15. The solution of the root locus design problem by use of MATLAB is presented in Sec. 7.11.

7.11 EXAMPLE OF SECTION 7.10: MATLAB ROOT LOCUS

MATLAB provides a number of options for plotting the root locus [9,10]. A root locus plot is obtained using the *rlocus* command. The root locus version of *sisotool* described in Appendix C allows a user to design compensators and to adjust the gain on a root locus plot. To provide more efficient design, the *rloczeta* function was written to calculate the gain necessary to yield a specified closed-loop damping ratio ζ. The macro *fom* computes the closed-loop figures of merit. Both of these commands are illustrated below.

```
%  This example illustrates the rloczeta function which
%  calculates the compensator gain required to yield a specified
```

```
%  damping ratio zeta
%
%  Define a G, H and G*H for the example of section 7.10

G = tf([2600],[1  100 2600 01])  % System = tf(numerator, denominator)

Transfer function:

         2600
 --------------------
  s^3 + 100 s^2 + 2600s

H = tf ([25],  [1 25])

Transfer function:

   25
 ------
  s + 25
oltf = G * H                %   oltf = G(s)H(s)

Transfer function:

               6500
 ------------------------------
  s^4 + 125s^3 + 5100s^2 + 65000s

%
%  Extract the numerator and denominator polynomials from oltf
%  (see Sect 4.18)
GHnum=oltf.num{1}
GHnum =

        0       0        0          0     65000

GHden=oltf.den{1}
GHden =

        1    125    5100    65000          0

%
%  Select a desired damping ratio zeta and tolerance
zeta = 0.5;
tol = 1e-12;

%  The function
%  [k, r] = rloczeta (num,den,zeta,tol)
%  Calculates the gain, k, which yields dominant closed-loop
%  roots having a damping ratio given by zeta. Also returns
%  all of the closed-loop roots.
%
[k,r] = rloczeta (GHnum,GHden,zeta,tol)
```

```
K =
   9.2435
r =
   -55.9022 -18.0414i
   -55.9022 +18.0414i
   -6.5978 -11.4277i
   -6.5978 +11.4377i
%
%  Calculate the damping ratio for each closed loop pole
zetas = -real(r)./abs(r)      % ./ is element by element division
zetas =
   0.9517
   0.9517
   0.5000
   0.5000
%
%  Plot the root locus

rlocus (oltf)          % Alternatively rlocus (Ghnum,Ghden)
%
%  Hold the plot and add the closed loop roots (Fig 7.18)
```

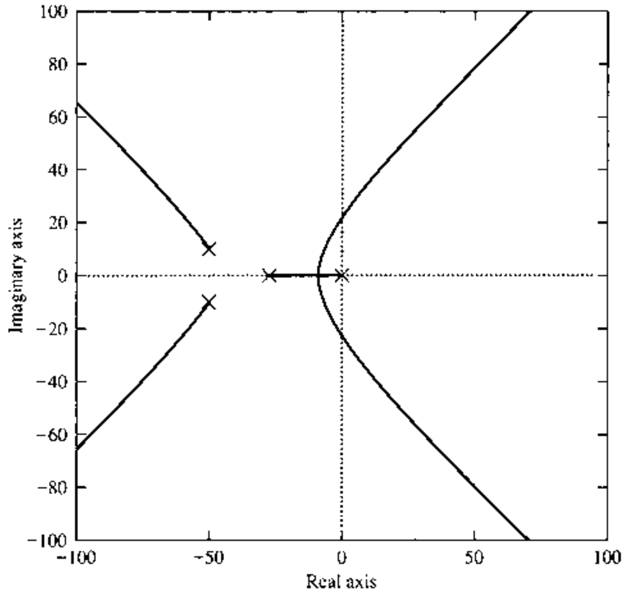


FIGURE 7.18 MATLAB root-locus plot corresponding to Fig. 7.17.

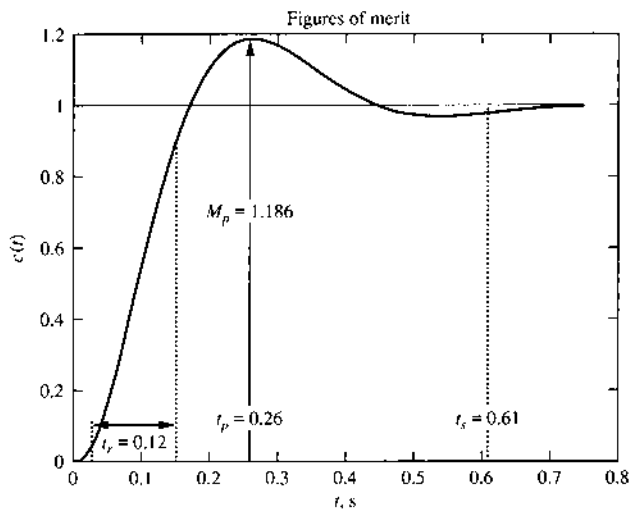


FIGURE 7.19 MATLAB time response plot corresponding to Eq. (7.72) and Fig. 7.26.

```
hold on
rlocus (GHnum,GHden,k)
%  Reshape the plot and make it square

axis([ -70 10 -40 40 ])
axis('square')

%  Form the closed-loop transfer function
cltf=feedback (k*G,H,-1)     % Remember to include the gain k
Transfer function:

          2.403e004s + 6.008e005
----------------------------------------
 s^4 + 125s^3 + 5100s^2 + 65000s + 6.008e005

%  Open a second figure window for the time response of Fig. 7.19
%  with the command:
figure (2)
%
%  The m-file fom.m was designed to compute the figures of merit
%  Its input arguments are the numerator and denominator
%  polynomials
%
fom(cltf.num{1},cltf.den{1})

Figures of merit for a unit step input
```

```
rise time       =  0.12
peak value      =  1.186
peak time       =  0.26
settling time   =  0.61
final value     =  1
```

7.12 ROOT LOCUS EXAMPLE WITH AN RH PLANE ZERO

Many aircraft flight control systems use an adjustable gain in the feedback path. Also, the forward transfer function may be nonminimum phase (nmp; see Sec. 8.11) and have a negative gain. This is illustrated by the control system shown in Fig. 7.20, where

$$G(s) = \frac{-2(s+6)(s-6)}{s(s+3)(s+4-j4)(s+4+j4)} \qquad H(s) = K_h > 0$$

Thus the open-loop transfer function is

$$G(s)H(s) = \frac{K(s+6)(s-6)}{s(s+3)(s+4-j4)(s+4+j4)} \tag{7.73}$$

where $K = -2K_h$. The root-locus characteristics are summarized as follows:

1. $G(s)H(s)$ has zeros located at $s = -6$ and $+6$. The zero at $s = 6$ means that this transfer function is nonminimum phase. The poles are located at $s = 0, -3, -4 + j4$ and $-4 - j4$.
2. There are four branches of the root locus.
3. For positive values of K, a branch of the root locus is located on the positive real axis from $s = 0$ to $s = 6$. Thus, there is a positive real root and the system is unstable. For $K < 0$ (the values of K_h are positive), the root locus must satisfy the 360° angle condition given in Eq. (7.26). The root locus is shown in Fig. 7.21. This plot is readily obtained using a CAD program.
4. The closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{-2(s+6)(s-6)}{s^4 + 11s^3 + (56+K)s^2 + 96s - 36K} \tag{7.74}$$

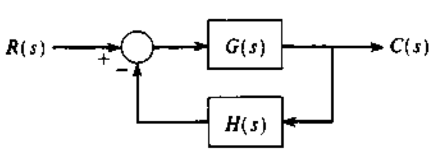


FIGURE 7.20 Nonunity-feedback control system.

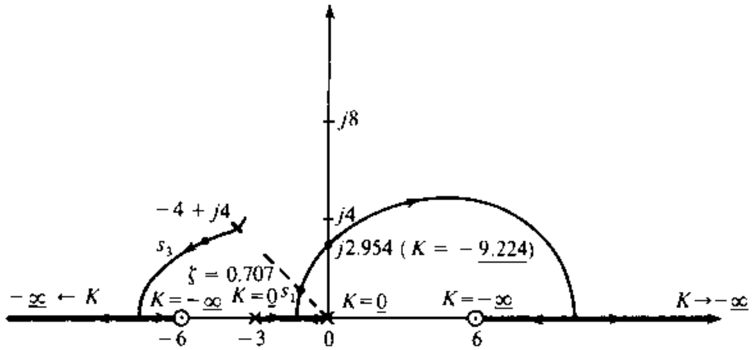


FIGURE 7.21 Root locus for Eq. (7.73) for negative values of K.

The Routhian array for the denominator polynomial shows that the imaginary axis crossing occurs at $s = \pm j2.954$ for $K = -9.224$. Thus, the closed-loop system is stable for $K > -9.224$.

5. Using $\zeta = 0.707$ for selecting the dominant poles yields $s_{1,2} = -1.1806 \pm j1.1810$ and $s_{3,4} = -4.3194 \pm j3.4347$, with $K_h = 1.18(K = -2.36)$. The closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{-2(s+6)(s-6)}{(s+1.1806 \pm j1.1810)(s+4.3194 \pm j3.4347)} \tag{7.75}$$

6. The output response for a unit-step input is

$$\begin{aligned} c(t) = 0.84782 - 1.6978e^{-1.1806t} \sin(1.181t + 29.33^\circ) \\ - 0.20346e^{-4.3194t} \sin(3.437t + 175.47^\circ) \end{aligned} \tag{7.76}$$

The step response for this system is plotted in Fig. 7.22. Note that the slight initial undershoot is due to the zero in the RHP. The figures of merit are $M_p = 0.8873$, $t_p = 2.89s$, $t_s = 3.85s$, and $c(\infty) = 0.8478$. Note that the forward transfer function $G(s)$ is Type 1, but the steady-state value of the output is not equal to the input because of the nonunity feedback. Also note that the numerator of $C(s)/R(s)$ is the same as the numerator of $G(s)$; thus, it is not affected by the value selected for K.

7.13 PERFORMANCE CHARACTERISTICS

As pointed out early in this chapter, the root-locus method incorporates the more desirable features of both the classical method and the steady-state sinusoidal phasor analysis. In the example of Sec. 7.10 a direct relationship

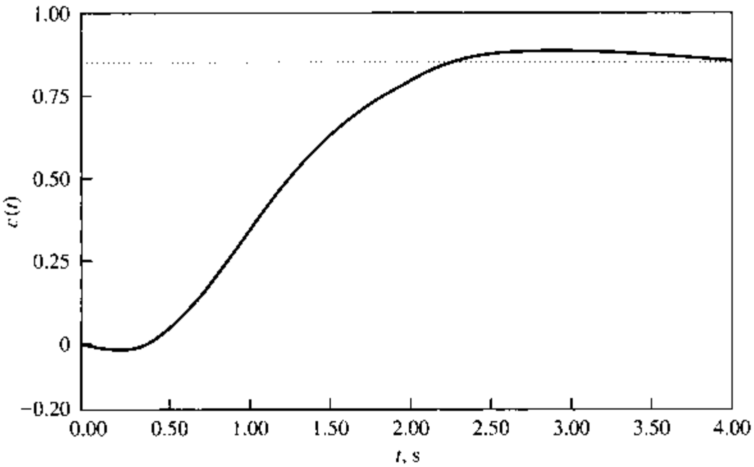


FIGURE 7.22 Step response for the closed-loop system given in Eq. (7.75).

is noted between the root locus and the time solution. This section is devoted to strengthening this relationship to enable the designer to synthesize and/or compensate a system.

General Introduction

In review, consider a simple second-order system whose control ratio is

$$\frac{C(s)}{R(s)} = \frac{K}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{7.77}$$

and whose transient component of the response to a step input is

$$c(t)_t = C_1 e^{s_1 t} + C_2 e^{s_2 t} \tag{7.78}$$

where, for $\zeta < 1$,

$$s_{1,2} = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2} = \sigma \pm j\omega_d \tag{7.79}$$

Thus

$$c(t)_t = Ae^{\sigma t}\sin(\omega_d t + \phi) \tag{7.80}$$

Consider now a plot of the roots in the s plane and their correlation with the transient solution in the time domain for a step input. Six cases are illustrated in Fig. 7.23 for different values of damping, showing both the locations of the roots and the corresponding transient plots. In the case of $\zeta = 0$, the roots lie on the $\pm j\omega$ axis and the response has sustained

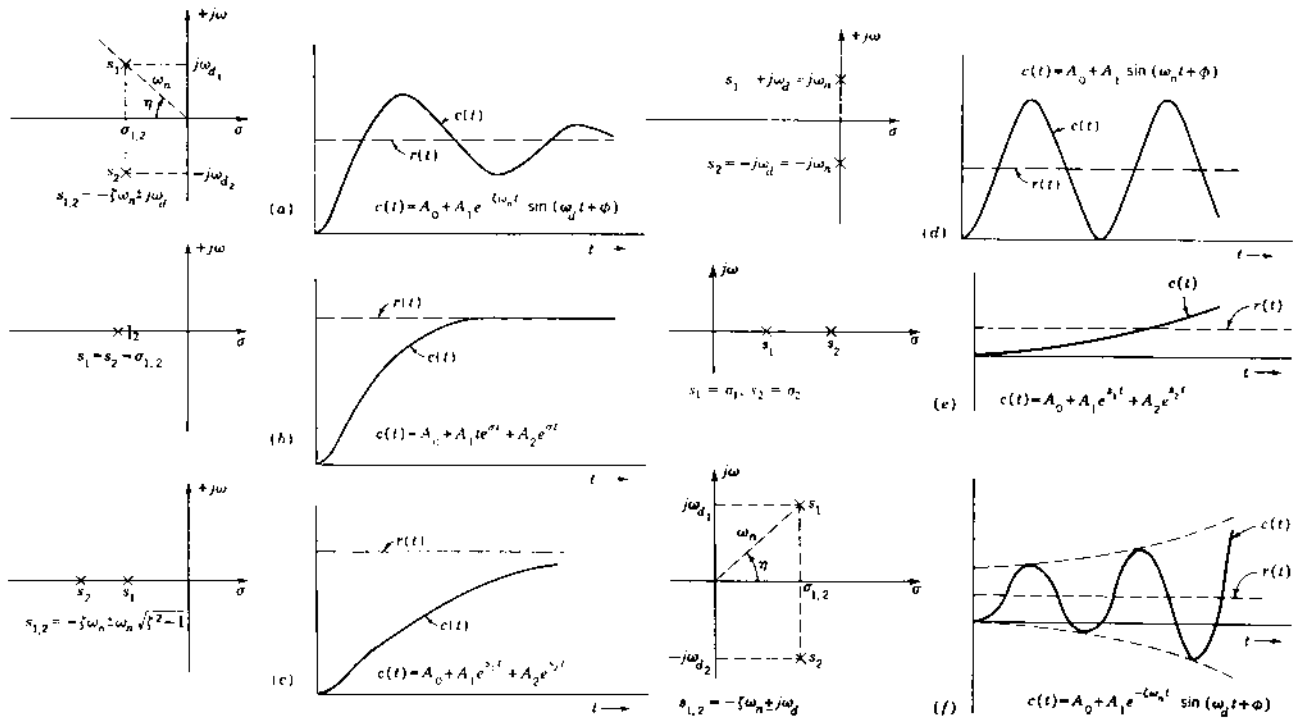


FIGURE 7.23 Plot of roots of a second-order characteristic equation in the *s* plane and their correlation to the transient solution in the time domain for step input: (*a*) underdamped, stable, $0 < \zeta < 1$; (*b*) critically damped, stable, $\zeta = 1$; (*c*) real roots, overdamped, stable, $\zeta > 1$; (*d*) undamped, sustained oscillations, $\zeta = 0$; (*e*) real roots, unstable, $\zeta < -1$; (*f*) underdamped, unstable, $0 > \zeta > -1$.

oscillations (Fig. 7.23*d*). Those portions of the root locus that yield roots in the RHP result in unstable operation (Fig. 7.23*e* and *f*). Thus, the desirable roots are on that portion of the root locus in the LHP. (As shown in Chap. 9, the $\pm j\omega$ axis of the s plane corresponds to the $-1+j0$ point of the Nyquist plot.)

Table 4.1 summarizes the information available from Fig. 7.23; i.e., it shows the correlation between the location of the closed-loop poles and the corresponding transient component of the response. Thus, the value of the root-locus method is that it is possible to determine all the forms of the transient component of the response that a control system may have.

Plot of Characteristic Roots for $0 < \zeta < 1$

The important desired roots lie in the region in which $0 < \zeta < 1$ (generally between 0.4 and 0.8). In Fig. 7.23*a*, the radius r from the origin to the root s_1 is

$$r = \sqrt{\omega_{d_1}^2 + \sigma_{1,2}^2} = \sqrt{\omega_n^2(1-\zeta^2) + \omega_n^2\zeta^2} = \omega_n \tag{7.81}$$

and

$$\cos\eta = \left|\frac{-\sigma_{1,2}}{r}\right| = \frac{\zeta\omega_n}{\omega_n} = \zeta \tag{7.82}$$

or

$$\eta = \cos^{-1}\zeta \tag{7.83}$$

Based on the preceding equations, the constant-parameter loci are drawn in Fig. 7.24. From Figs. 7.23 and 7.24 and Eqs. (7.81) to (7.83)

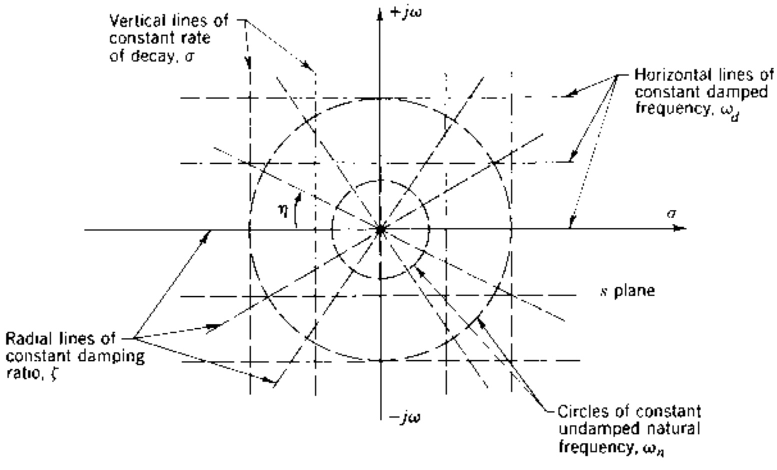


FIGURE 7.24 Constant-parameter curves on the s plane.

the following characteristics are summarized for the s plane:

1. Horizontal lines represent lines of constant damped natural frequency ω_d. The closer these lines are to the real axis, the lower the value of ω_d. For roots lying on the real axis ($\omega_d = 0$) there is no oscillation in the transient.
2. Vertical lines represent lines of constant damping or constant rate of decay of the transient. The closer these lines (or the characteristic root) are to the imaginary axis, the longer it takes for the transient response to die out.
3. Circles about the origin are circles of constant undamped natural frequency ω_n. Since $\sigma^2 + \omega_d^2 = \omega_n^2$, the locus of the roots of constant ω_n is a circle in the s plane; that is, $|s_1| = |s_2| = \omega_n$. The smaller the circles, the lower the value of ω_n.
4. Radial lines passing through the origin with the angle η are lines of constant damping ratio ζ. The angle η is measured clockwise from the negative real axis for positive ζ, as shown in Figs. 7.23*a* and 7.24.

Variation of Roots with ζ

Note in Fig. 7.25 the following characteristics:

For $\zeta > 1$, the roots $s_{1,2} = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$ are real.
For $\zeta = 1$, the roots $s_{1,2} = -\zeta\omega_n$ are real and equal.
For $\zeta < 1$, the roots $s_{1,2} = -\zeta\omega_n \pm j\omega_n\sqrt{1 - \zeta^2}$ are complex conjugates.
For $\zeta = 0$, the roots $s_{1,2} = \pm j\omega_n$ are imaginary.

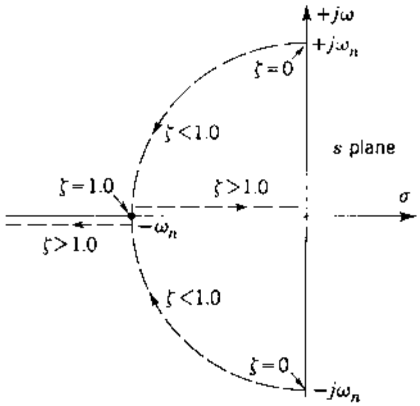


FIGURE 7.25 Variation of roots of a simple second-order system for a constant ω_n.

Higher-Order Systems

Consider the control ratio of a system of order n given by

$$\frac{C(s)}{R(s)} = \frac{P(s)}{s^n + a_{n-1}s^{n-1} + \cdots + a_0} \tag{7.84}$$

In a system having one or more sets of complex-conjugate roots in the characteristic equation, each quadratic factor is of the form

$$s^2 + 2\zeta\omega_n s + \omega_n^2 \tag{7.85}$$

and the roots are

$$s_{1,2} = \sigma \pm j\omega_d \tag{7.86}$$

The relationship developed for ω_n and ζ earlier in this section apply equally well for each complex-conjugate pair of roots of an nth-order system. The distinction is that the *dominant* ζ and ω_n apply for that pair of complex-conjugate roots that lies closest to the imaginary axis. These values of ζ and ω_n are dominant because the corresponding transient term has the longest settling time and the largest magnitude. Thus, the dominant values of ζ and ω_n are selected from the root locus for the desired response. Remember that in selecting the values of the dominant ζ and ω_n, the other roots are automatically set. Depending on the location of the other roots, they modify the solution obtained from the dominant roots.

In Example 1 of Sec. 7.10 the response $c(t)$ with a step input given by Eq. (7.72) is

$$c(t) = 1 + 1.21e^{-6.6t}\sin(11.4t - 111.7^\circ) + 0.28e^{-55.9t}\sin(18t + 26.1^\circ) \tag{7.87}$$

Note that the transient term due to the roots of $s = -55.9 \pm j18$ dies out in approximately one-tenth the time of the transient term due to the dominant roots $s = -6.6 \pm j11.4$. Therefore, the solution can be approximated by the simplified equation

$$c(t) \approx 1 + 1.21e^{-6.6t}\sin(11.4t - 111.7^\circ) \tag{7.88}$$

This expression is valid except for a short initial period of time while the other transient term dies out. Equation (7.88) suffices if an approximate solution is desired, i.e., for determining M_p, t_p, and T_s, since the neglected term does not appreciably affect these three quantities. In general, the designer can judge from the location of the roots which ones may be neglected. Plots of $c(t)$ corresponding to Eqs. (7.87) and (7.88) are shown in Fig. 7.26. Since there is little difference between the two curves, the system can be approximated as a second-order system. The significant characteristics of the time response are $M_p = 1.19$, $t_p = 0.26s$, $t_s = 0.61s$, and $\omega_d \approx 11.4$ rad/s.

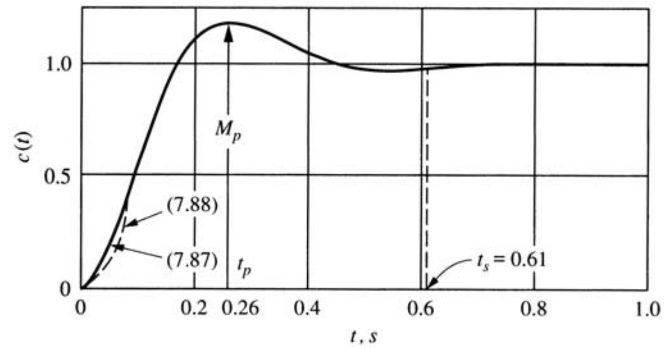


FIGURE 7.26 Plot $c(t)$ vs. t for Eqs. (7.87) and (7.88).

The transient response of any complex system, with the effect of all the roots taken into account, can often be considered to be the result of an *equivalent* second-order system. On this basis, it is possible to define effective (or equivalent) values of ζ and ω_n. Realize that in order to alter either effective quantity, the location of one or more of the roots must be altered.

7.14 TRANSPORT LAG [7]

Some elements in control systems are characterized by dead time or transport lag. This appears as a dead interval for which the output is delayed in response to an input. Figure 7.27*a* shows the block diagram of a transport-lag element. In Fig. 7.27*b* a typical resultant output $e_o(t) = e_i(t - \tau)u_{-1}(t - \tau)$ is shown, which has the same form as the input but is delayed by a time interval τ. Dead time is a nonlinear characteristic that can be represented precisely

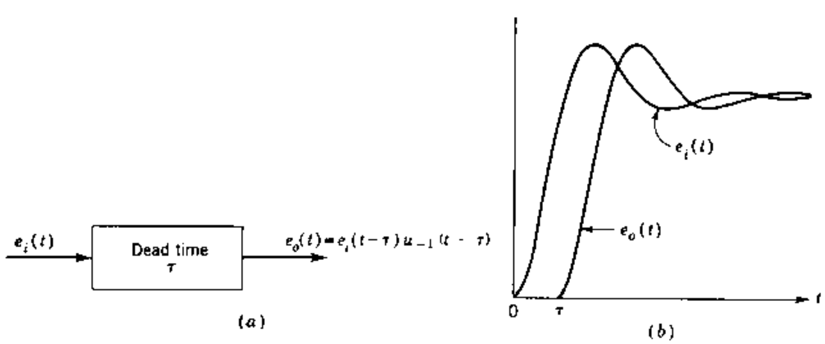


FIGURE 7.27 (*a*) Transport lag τ; (*b*) time characteristic.

in terms of the Laplace transform (see Sec. 4.4) as a transcendental function. Thus,

$$E_o(s) = e^{-\tau s} E_i(s) \tag{7.89}$$

The transfer function of the transport-lag element, where $s = \sigma + j\omega$, is

$$G_\tau(s) = \frac{E_o(s)}{E_i(s)} = e^{-\tau s} = e^{-\sigma\tau} \underline{/-\omega\tau} \tag{7.90}$$

It has a negative angle that is directly proportional to the frequency.

The root locus for a system containing a transport lag is illustrated by the following example [7].

$$G(s)H(s) = \frac{Ke^{-\tau s}}{s(s-p_1)} = \frac{Ke^{-\sigma\tau}e^{-j\omega\tau}}{s(s-p_1)} \tag{7.91}$$

For $K > 0$ the angle condition is

$$\underline{/s} + \underline{/s - p_1} + \underline{/\omega\tau} = (1+2h)180^\circ \tag{7.92}$$

The root locus for the system with transport lag is drawn in Fig. 7.28. The magnitude condition used to calibrate the root locus is

$$K = |s| \cdot |s - p_1| e^{\sigma\tau} \tag{7.93}$$

The same system without transport lag has only two branches (see Fig. 7.2) and is stable for all $K > 0$. The system with transport lag has an infinite number of branches and the asymptotes are

$$\omega = \frac{(1+2h)\pi}{\tau} \tag{7.94}$$

where $h = 0, \pm 1, \pm 2, \ldots$. The values of K at the points where the branches cross the imaginary axis are

$$K = \omega\sqrt{p_1^2 + \omega^2} \tag{7.95}$$

The two branches closest to the origin thus have the largest influence on system stability and are therefore the principal branches. The maximum value of loop sensitivity K given by Eq. (7.95) is therefore determined by the frequency ω_1 (shown in Fig. 7.28) that satisfies the angle condition $\omega_1\tau = \tan^{-1}|p_1|/\omega_1$.

7.15 SYNTHESIS

The root-locus method lends itself very readily to the synthesis problem because of the direct relationship between the frequency and the time

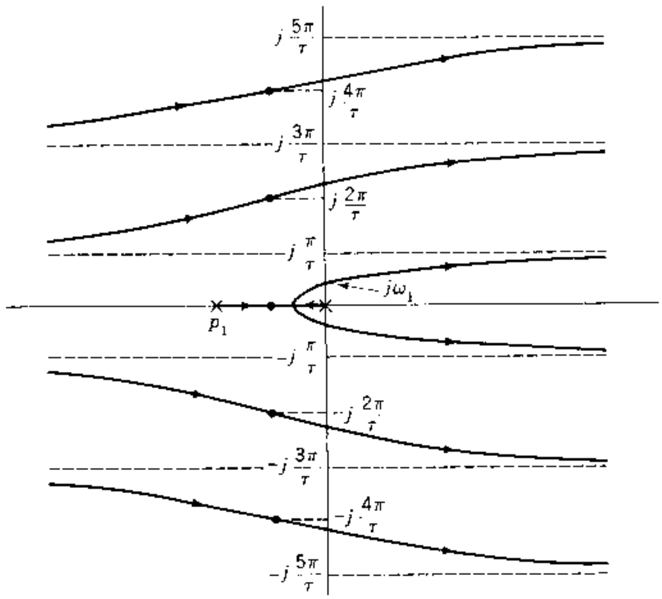


FIGURE 7.28 Root locus for $G(s)H(s) = \dfrac{Ke^{-\tau s}}{s(s - p_1)}$.

domains to the *s* domain. The desired response can be achieved by keeping in mind the following five points:

1. First plot the root locus; then the locations of the dominant closed-loop poles must be specified, based on the desired transient response.
2. Use these guidelines for selecting the roots:
   (*a*) Since the desired time response is often specified to have a peak value M_p between 1.0 and 1.4, the locus has at least one set of complex-conjugate poles. Often there is a pair of complex poles near the imaginary axis that dominates the time response of the system; if so, these poles are referred to as the *dominant-pole pair.*
   (*b*) To set the system gain, any of the following characteristics must be specified for the dominant-pole pair: the damping ratio ζ, the settling time T_s (for 2 percent: $T_s = 4/\zeta\omega_n$), the undamped natural frequency ω_n, or the damped natural frequency ω_d. As previously determined (see Fig. 7.24), each of these factors corresponds to either a line or a circle in the *s* domain.

(*c*) When the line or circle corresponding to the given information has been drawn, as stated in point 2*b*, its intersection with the locus determines the dominant-pole pair and fixes the value of the system gain. By applying the magnitude condition to this point, the value of the gain required can be determined. Setting the value of the gain to this value results in the transient-response terms, corresponding to the dominant-pole pair, having the desired characterizing form.

3. When the gain corresponding to the dominant roots has been determined, the remaining roots can be found by applying the magnitude condition to each branch of the locus or by using a CAD program. Once the value of the gain on the dominant branch is fixed, thus fixing the dominant poles, the locations of all remaining roots on the other branches are also fixed.
4. If the root locus does not yield the desired response, the locus must be altered by compensation to achieve the desired results. The subject of compensation in the s domain is discussed in later chapters.

7.16 SUMMARY OF ROOT-LOCUS CONSTRUCTION RULES FOR NEGATIVE FEEDBACK

RULE 1. The number of branches of the root locus is equal to the number of poles of the open-loop transfer function.

RULE 2. For *positive* values of K, the root locus exists on those portions of the real axis for which the total number of real poles and zeros to the right is an odd number. For *negative* values of K, the root locus exists on those portions of the real axis for which the total number of real poles and zeros to the right is an even number (including zero).

RULE 3. The root locus starts ($K = 0$) at the open-loop poles and terminates ($K = \pm\infty$) at the open-loop zeros or at infinity.

RULE 4. The angles of the asymptotes for those branches of the root locus that end at infinity are determined by

$$\gamma = \frac{(1+2h)180^\circ \text{ for } K > 0 \qquad \text{or} \qquad h360^\circ \text{ for } K < 0}{[\text{Number of poles of } G(s)H(s)] - [\text{Number of zeros of } G(s)H(s)]}$$

RULE 5. The real-axis intercept of the asymptotes is

$$\sigma_o = \frac{\sum_{c=1}^{n} \mathrm{Re}(p_c) - \sum_{h=1}^{w} \mathrm{Re}(z_h)}{n - w}$$

RULE 6. The breakaway point for the locus between two poles on the real axis (or the break-in point for the locus between two zeros on the real axis) can be determined by taking the derivative of the loop sensitivity K with respect to s. Equate this derivative to zero and find the roots of the resulting equation. The root that occurs between the poles (or the zeros) is the breakaway (or break-in) point.

RULE 7. For $K > 0$ the angle of departure from a complex pole is equal to 180° minus the sum of the angles from the other poles plus the sum of the angles from all the zeros. Any of these angles may be positive or negative. For $K < 0$ the departure angle is 180° from that obtained for $K > 0$.

For $K > 0$ the angle of approach to a complex zero is equal to the sum of the angles from all the poles minus the sum of the angles from the other zeros minus 180°. For $K < 0$ the approach angle is 180° from that obtained from $K > 0$.

RULE 8. The imaginary-axis crossing of the root locus can be determined by setting up the Routhian array from the closed-loop characteristic polynomial. Equate the s^1 row to zero and the auxiliary equation from the s^2 row. The roots of the auxiliary equation are the imaginary-axis crossover points.

RULE 9. The selection of the dominant roots of the characteristic equation is based on the specifications that give the required system performance; i.e., it is possible to evaluate $\sigma, \omega_d,$ and ζ from Eqs. (3.60), (3.61), and (3.64), which in turn determine the location of the desired dominant roots. The loop sensitivity for these roots is determined by applying the magnitude condition. The remaining roots are then determined to satisfy the same magnitude condition.

RULE 10. For those open-loop transfer functions for which $w \leq n - 2$, the sum of the closed-loop roots is equal to the sum of the open-loop poles. Thus, once the dominant roots have been located, the rule

$$\sum_{j=1}^{n} p_j = \sum_{j=1}^{n} r_j$$

can be used to find one real or two complex roots. Factoring known roots from the characteristic equation can simplify the work of finding the remaining roots.

A root-locus CAD program (see Appendix C) will produce an accurate calibrated root locus. This considerably simplifies the work required for the system design. By specifying ζ for the dominant roots or the gain K_m, a computer program can determine all the roots of the characteristic equation.

7.17 SUMMARY

In this chapter the root-locus method is developed to solve graphically for the roots of the characteristic equation. This method can be extended to solving for the roots of any polynomial. Any polynomial can be rearranged and put into the mathematical from of a ratio of factored polynomials that is equal to plus or minus unity, i.e.,

$$\frac{N(s)}{D(s)} = \pm 1$$

Once this form has been obtained, the procedures given in this chapter can be utilized to locate the roots of the polynomial.

The root locus permits an analysis of the performance of feedback systems and provides a basis for selecting the gain in order to best achieve the performance specifications. Since the closed-loop poles are obtained explicitly, the form of the time response is directly available. CAD programs are available for obtaining $c(t)$ vs. t. Widely used CAD packages for control systems simplify considerably the analysis and design process. They provide information including root-locus plots, roots of interest for a given gain, closed-loop transfer functions, and the time response with a specified input. If the performance specifications cannot be met, the root locus can be analyzed to determine the appropriate compensation to yield the desired results. This is covered in Chap. 10.

REFERENCES

1. Truxal, J. G.: *Automatic Feedback Control System Synthesis*, McGraw-Hill, New York, 1955.
2. Evans, W. R.: *Control-System Dynamics*, McGraw-Hill, New York, 1954.
3. Lorens, C. S., and R. C. Titsworth: "Properties of Root Locus Asymptotes," letter in *IRE Trans. Autom. Control*, vol. AC-5, pp. 71–72, January 1960.
4. Yeung, K. S.: "A Remark on the Use of Remec's Method of Finding Breakaway Points," *IEEE Trans. Autom. Control*, vol. AC-26, pp.940–941, 1981.
5. Wilts, C. H.: *Principles of Feedback Control*, Addison-Wesley, Reading, Mass., 1960.
6. Rao, S. N.: "A Relation between Closed Loop Pole Zeros Locations," *Int. J. Control*, vol. 24, p. 147, 1976.
7. Chang, C. S.: "Analytical Method for Obtaining the Root Locus with Positive and Negative Gain," *IEEE Trans. Autom. Control*, vol. AC-10, pp. 92–94, 1965.
8. Yeh, V. C. M.: "The Study of Transients in Linear Feedback Systems by Conformal Mapping and Root-Locus Method," *Trans. ASME*, vol. 76, pp. 349–361, 1954.
9. Saadat, H.: *Computational Aids in Control Systems Using MATLAB®*, McGraw-Hill, New York, 1993.
10. Leonard, N. E., and W. S. Levine: *Using MATLAB to Analyze and Design Control Systems*, 2nd Ed. Benjamin/Cummings, Meulo Park, Calif. 1995.

8

Frequency Response

8.1 INTRODUCTION

The frequency response [1] method of analysis and design is a powerful technique for the comprehensive study of a system by conventional methods. Performance requirements can be readily expressed in terms of the frequency response as an alternative to the s plane analysis using the root locus. Since noise, which is always present in any system, can result in poor overall performance, the frequency-response characteristics of a system permit evaluation of the effect of noise. The design of a passband for the system response may result in excluding the noise and therefore improving the system performance as long as the dynamic tracking performance specifications are met. The frequency response is also useful in situations for which the transfer functions of some or all of the components in a system are unknown. The frequency response can be determined experimentally for these situations, and an approximate expression for the transfer function can be obtained from the graphical plot of the experimental data. The frequency-response method is also a very powerful method for analyzing and designing a robust multi-input/multi-output (MIMO) [2] system with structured uncertain plant parameters. In this chapter two graphical representations of transfer functions are presented: the logarithmic plot and the polar plot. These plots are used to develop Nyquist's stability criterion [1,3–5] and closed-loop design procedures. The plots are also readily obtained by use of computer-aided design (CAD) packages like MATLAB or TOTAL-PC (see Ref. 6). The closed-loop feedback response $\mathbf{M}(j\omega)$ is obtained as a function of the open-loop transfer function $\mathbf{G}(j\omega)$. Design methods for adjusting the open-loop gain are developed and demonstrated. They are based on the polar plot of $\mathbf{G}(j\omega)$ and the Nicholas plot. Both methods achieve a peak value M_m and a resonant frequency ω_m of the closed-loop frequency response. A correlation between these frequency-response characteristics and the time response is developed.

Design of control systems by state variable technique is often based upon achieving an optimum performance according to a specified performance index PI; for example, minimizing the integral of squared error (ISE): $\mathrm{PI} = \int_0^\infty e(t)^2 dt$, where $e \equiv r - c$. Both frequency-response and root-locus methods are valuable complementary tools for many of the techniques of modern control theory.

8.2 CORRELATION OF THE SINUSOIDAL AND TIME RESPONSE [3]

As pointed out earlier, solving for $c(t)$ by the classical method is laborious and impractical for synthesis purposes, especially when the input is not a simple analytical function. The use of Laplace transform theory lessens the work involved and permits the engineer to synthesize and improve a system. The root-locus method illustrates this fact. The advantages of the graphical representations in the frequency domain of the transfer functions are developed in the following pages.

Once the frequency response of a system has been determined, the time response can be determined by inverting the corresponding Fourier transform. The behavior in the frequency domain for a given driving function $r(t)$ can be determined by the Fourier transform as

$$\mathbf{R}(j\omega) = \int_{-\infty}^{\infty} r(t)e^{-j\omega t} dt \tag{8.1}$$

For a given control system the frequency response of the controlled variable is

$$\mathbf{C}(j\omega) = \frac{\mathbf{G}(j\omega)}{1 + \mathbf{G}(j\omega)\mathbf{H}(j\omega)}\mathbf{R}(j\omega) \tag{8.2}$$

By use of the inverse Fourier transform, which is much used in practice, the controlled variable as a function of time is

$$c(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{C}(j\omega)e^{j\omega t} d\omega \tag{8.3}$$

If the design engineer cannot evaluate Eq. (8.3) by reference to a table of definite integrals, this equation can be evaluated by numerical or graphical integration. This is necessary if $\mathbf{C}(j\omega)$ is available only as a curve and cannot be simply expressed in analytical form, as is often the case. The procedure is described in several books [14]. In addition, methods have been developed based on the Fourier transform and a step input signal, relating $\mathbf{C}(j\omega)$ qualitatively to the time solution without actually taking the inverse Fourier transform. These methods permit the engineer to make an approximate determination of the system response through the interpretation of graphical plots in the frequency domain. This makes the design and improvement of feedback systems possible with a minimum effort and is emphasized in this chapter.

Section 4.12 shows that the frequency response is a function of the pole-zero pattern in the s plane. It is therefore related to the time response of the system. Two features of the frequency response are the maximum value M_m and the resonant frequency ω_m. Section 9.3 describes the qualitative relationship between the time response and the values M_m and ω_m. Since the location of the closed-loop poles can be determined from the root locus, there is a direct relationship between the root-locus and frequency-response methods.

8.3 FREQUENCY-RESPONSE CURVES

The frequency domain plots that have found great use in graphical analysis in the design of feedback control systems belong to two categories. The first category is the plot of the magnitude of the output-input ratio vs. frequency in rectangular coordinates, as illustrated in Sec. 4.12. In logarithmic coordinates these are known as *Bode plots.* Associated with this plot is a second plot of the corresponding phase angle vs. frequency. In the second category the output-input ratio may be plotted in polar coordinates with frequency as a parameter. There are two types of polar plots, direct and inverse. Polar plots are generally used only for the open-loop response and are commonly referred to as *Nyquist Plots* [8]. The plots can be obtained experimentally or by a CAD package (see Appendixes C and D). These programs assist in obtaining plots and the designs developed in this text. When a CAD program is not available, the Bode plots are easily obtained by a graphical procedure. The other plots can then be obtained from the Bode plots.

For a given sinusoidal input signal, the input and steady-state output are of the following forms:

$$r(t) = \mathbf{R} \sin \omega t \tag{8.4}$$

$$c(t) = \mathbf{C} \sin (\omega t + \alpha) \tag{8.5}$$

The closed-loop frequency response is given by

$$\frac{\mathbf{C}(j\omega)}{\mathbf{R}(j\omega)} = \frac{\mathbf{G}(j\omega)}{1 + \mathbf{G}(j\omega)\mathbf{H}(j\omega)} = M(\omega)\,\underline{/\alpha(\omega)} \tag{8.6}$$

For each value of frequency, Eq. (8.6) yields a phasor quantity whose magnitude is M and whose phase angle α is the angle between $C(j\omega)$ and $R(j\omega)$.

An ideal system may be defined as one where $\alpha = 0^\circ$ and $R(j\omega) = C(j\omega)$ for $0 < \omega < \infty$. (See curves 1 in Fig. 8.1.) However, this definition implies an instantaneous transfer of energy from the input to the output. Such a transfer cannot be achieved in practice since any physical system has some energy dissipation and some energy-storage elements. Curves 2 and 3 in Fig. 8.1 represent the frequency responses of practical control systems. The passband, or bandwidth, of the frequency response is defined as the range of frequencies from 0 to the frequency ω_b; where $M = 0.707$ of the value at $\omega = 0$. However, the frequency ω_m is more easily obtained than ω_b. The values M_m and ω_m are often used as figures of merit (FOM).

In any system the input signal may contain spurious noise signals in addition to the true signal input, or there may be sources of noise within the closed-loop system. This noise is generally in a band of frequencies above the dominant frequency band of the true signal. Thus, to reproduce the true signal and attenuate the noise, feedback control systems are designed to have a definite passband. In certain cases the noise frequency may exist in the same frequency band as the true signal. When this occurs, the problem of estimating

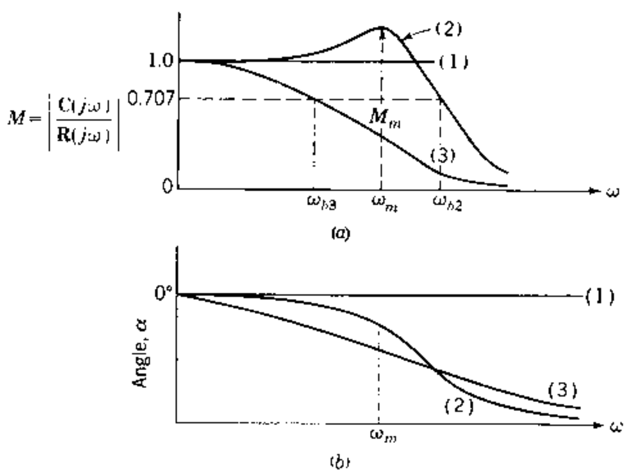


FIGURE 8.1 Frequency-response characteristics of $C(j\omega)/R(j\omega)$ in rectangular coordinates.

the desired singal is more complicated. Therefore, even if the ideal system were possible, it would not be desirable.

8.4 BODE PLOTS (LOGARITHMIC PLOTS)

The plotting of the frequency transfer function can be systematized and simplified by using logarithmic plots. The use of semilog paper eliminates the need to take logarithms of very many numbers and also expands the low-frequency range, which is of primary importance. The advantages of logarithmic plots are that (1) the mathematical operations of multiplication and division are transformed to addition and subtraction and (2) the work of obtaining the transfer function is largely graphical instead of analytical. The basic factors of the transfer function fall into three categories, and these can easily be plotted by means of straight-line asymptotic approximations.

In preliminary design studies the straight-line approximations are used to obtain approximate performance characteristics very quickly or to check values obtained from the computer. As the design becomes more firmly specified, the straight-line curves can be corrected for greater accuracy. From these logarithmic plots enough data in the frequency range of concern can readily be obtained to determine the corresponding polar plots.

Some basic definitions of logarithmic terms follow.

Logarithm. The logarithm of a complex number is itself a complex number. The abbreviation "log" is used to indicate the logarithm to the base 10:

$$\begin{aligned} \log\left|\mathbf{G}(j\omega)\right|e^{j\phi(\omega)} &= \log\left|\mathbf{G}(j\omega)\right| + \log\, e^{j\phi(\omega)} \\ &= \log\left|\mathbf{G}(j\omega)\right| + j0.434\phi(\omega) \end{aligned} \tag{8.7}$$

The real part is equal to the logarithm of the magnitude, log $|\mathbf{G}(j\omega)|$, and the imaginary part is proportional to the angle, $0.434\phi(\omega)$. In the rest of this book the factor 0.434 is omitted and only the angle $\phi(\omega)$ is used.

Decibel. In feedback-system work the unit commonly used for the logarithm of the magnitude is the *decibel* (dB). When logarithms of transfer functions are used, the input and output variables are not necessarily in the same units; e.g., the output may be speed in radians per second, and the input may be voltage in volts.

Log magnitude. The logarithm of the magnitude of a transfer functions $\mathbf{G}(j\omega)$ expressed in decibel is

$$20\,\log\left|\mathbf{G}(j\omega)\right| \quad \text{dB}$$

This quantity is called the *log magnitude*, abbreviated Lm. Thus

$$\mathrm{Lm}\, \mathbf{G}(j\omega) = 20 \log\left|\mathbf{G}(j\omega)\right| \qquad \mathrm{dB} \tag{8.8}$$

Since the transfer function is a function of frequency, the Lm is also a function of frequency.

Octave and Decade. Two units used to express frequency bands of frequency ratios are the octave and the decade. An octave is a frequency band from f_1 to f_2, where $f_2/f_1 = 2$. Thus, the frequency band from 1 to 2 Hz is 1 octave in width, and the frequency band from 17.4 to 34.8 Hz is also 1 octave in width. Note that 1 octave is not a fixed frequency bandwidth but depends on the frequency range being considered. The number of octaves in the frequency range from f_1 to f_2 is

$$\frac{\log(f_2/f_1)}{\log 2} = 3.32 \log \frac{f_2}{f_1} \qquad \text{octaves} \tag{8.9}$$

There is an increase of 1 decade from f_1 to f_2 when $f_2/f_1 = 10$. The frequency band from 1 to 10 Hz or from 2.5 to 25 Hz is 1 decade in width. The number of decades from f_1 to f_2 is given by

$$\log \frac{f_2}{f_1} \qquad \text{decades} \tag{8.10}$$

The dB values of some common numbers are given in Table 8.1. Note that the reciprocals of numbers differ only in sign. Thus, the dB value of 2 is +6 dB and the dB value of $\frac{1}{2}$ is −6 dB. Two properties are illustrated in Table 8.1.

Property 1. As a number doubles, the decibel value increases by 6 dB. The number 2.70 is twice as big as 1.35, and its decibel value is 6 dB more. The number 200 is twice as big as 100, and its decibel value is 6 dB greater.

Property 2. As a number increases by a factor of 10, the decibel value increases by 20 dB. The number 100 is 10 times as large as the number 10, and its decibel value is 20 dB more. The number 200 is 100 times larges than the number 2, and its decibel value is 40 dB greater.

TABLE 8.1 Decibel Values of Some Common Numbers

Number	Decibels	Number	Decibels
0.01	−40	2.0	6
0.1	−20	10.0	20
0.5	−6	100.0	40
1.0	0	200.0	46

8.5 GENERAL FREQUENCY-TRANSFER-FUNCTION RELATIONSHIPS

The frequency transfer function can be written in generalized form as the ratio of polynomials:

$$
\begin{aligned}
G(j\omega) &= \frac{K_m(1+j\omega T_1)(1+j\omega T_2)^r\cdots}{(j\omega)^m(1+j\omega T_a)[1+(2\zeta/\omega_n)j\omega+\left(1/\omega_n^2\right)(j\omega)^2]\cdots} \\
&= K_m\mathbf{G}'(j\omega)
\end{aligned}
\tag{8.11}
$$

where K_m is the gain constant. The logarithm of the transfer function is a complex quantity; the real portion is proportional to the log of the magnitude, and the complex portion is proportional to the angle. Two separate equations are written, one for the log magnitude and one for the angle, respectively:

$$
\begin{aligned}
\mathrm{Lm}\,\mathbf{G}(j\omega) &= \mathrm{Lm}\,K_m + \mathrm{Lm}\,(1+j\omega T_1) + r\,\mathrm{Lm}\,(1+j\omega T_2) + \cdots - m\,\mathrm{Lm}\,j\omega \\
&\quad - \mathrm{Lm}(1+j\omega T_a) - \mathrm{Lm}\left[1+\frac{2\zeta}{\omega_n}j\omega+\frac{1}{\omega_n^2}(j\omega)^2\right] - \cdots
\end{aligned}
\tag{8.12}
$$

$$
\begin{aligned}
\underline{/\mathbf{G}(j\omega)} &= \underline{/K_m} + \underline{/1+j\omega T_1} + r\,\underline{/1+j\omega T_2} + \cdots - m\,\underline{/j\omega} \\
&\quad - \underline{/1+j\omega T_a} - \underline{/1+\frac{2\zeta}{\omega_n}j\omega+\frac{1}{\omega_n^2}(j\omega)^2} - \cdots
\end{aligned}
\tag{8.13}
$$

The angle equation may be written as

$$
\begin{aligned}
\underline{/\mathbf{G}(j\omega)} &= \underline{/K_m} + \tan^{-1}\omega T_1 + r\tan^{-1}\omega T_2 + \cdots - m90^\circ \\
&\quad - \tan^{-1}\omega T_a - \tan^{-1}\frac{2\zeta\omega/\omega_n}{1-\omega^2/\omega_n^2} - \cdots
\end{aligned}
\tag{8.14}
$$

The gain K_m is a real number but may be positive or negative; therefore, its angle is correspondingly 0° or 180°. Unless otherwise indicated, a positive value of gain is assumed in this book. Both the log magnitude and the angle given by these equations are functions of frequency. When the log magnitude and the angle are plotted as functions of the log of frequency, the resulting pair of curves are referred to as the Bode plots or the *log magnitude diagram and phase diagram*. Equations (8.12) and (8.13) show that the resultant curves are obtained by the addition and subtractions of the corresponding individual terms in the transfer function equation. The two curves can be combined into a single curve of log magnitude vs. angle, with frequency ω as a parameter. This curve, called the Nichols chart or the *log magnitude–angle diagram*, or the corresponding Nyquist polar plot, is used for the quantitative design of the feedback system to meet specifications of required performance.

8.6 DRAWING THE BODE PLOTS

The properties of frequency- response plots are presented in this section, but the data for these plots usually are obtained from a CAD program. The generalized form of the transfer function as given by Eq. (8.11) shows that the numerator and denominator have four basic types of factors:

$$K_m \tag{8.15}$$

$$(j\omega)^{\pm m} \tag{8.16}$$

$$(1 + j\omega T)^{\pm r} \tag{8.17}$$

$$\left[1 + \frac{2\zeta}{\omega_n} j\omega + \frac{1}{\omega_n^2}(j\omega)^2\right]^{\pm p} \tag{8.18}$$

Each of these terms except K_m may appear raised to an integral power other than 1. The curves of Lm and angle vs. the log frequency can easily be drawn for each factor. Then these curves for each factor can be added together graphically to get the curves for the complete transfer function. The procedure can be further simplified by using asymptotic approximations to these curves, as shown in the following pages.

Constants

Since the constant K_m is frequency-invariant, the plot of

$$\text{Lm } K_m = 20 \log K_m \quad \text{dB}$$

is a horizontal straight line. The constant raises or lowers the Lm curve of the complete transfer function by a fixed amount. The angle, of course, is zero as long as K_m is positive.

$j\omega$ Factors

The factor $j\omega$ appearing in the denominator has a log magnitude

$$\text{Lm } (j\omega)^{-1} = 20 \log \left|(j\omega)^{-1}\right| = -20 \log \omega \tag{8.19}$$

When plotted against log ω, this curve is a straight line with a negative slope of 6 dB/octave or 20 dB/decade. Values of this function can be obtained from Table 8.1 for several values of ω. The angle is constant and equal to -90°.

When the factor $j\omega$ appears in the numerator, the Lm is

$$\text{Lm } (j\omega) = 20 \log|j\omega| = 20 \log \omega \tag{8.20}$$

This curve is a straight line with a positive slope of 6 dB/octave or 20 dB/decade. The angle is constant and equal to $+90^\circ$. Notice that the only difference between the curves for $j\omega$ and for $1/j\omega$ is a change in the sign of the slope of the Lm and a change in the sign of the angle. Both curves go through the point 0 dB at $\omega = 1$.

For the factor $(j\omega)^{\pm m}$ the Lm curve has a slope of $\pm 6m$ dB/octave or $\pm 20m$ dB/decade, and the angel is constant and equal to $\pm m\, 90^\circ$.

$1 + j\omega T$ Factors

The factor $1 + j\omega T$ appearing in the denominator has a log magnitude

$$\text{Lm } (1 + j\omega T)^{-1} = 20 \text{ log} \left|1 + j\omega T\right|^{-1} = -20 \text{ log} \sqrt{1 + \omega^2 T^2} \qquad (8.21)$$

For very small values of ω, that is $\omega T \ll 1$,

$$\text{Lm } (1 + j\omega T)^{-1} \approx \log 1 = 0 \text{ dB} \qquad (8.22)$$

Thus, the plot of the Lm at small frequencies is the 0-dB line. For every large values of ω, that is, $\omega T \gg 1$,

$$\text{Lm } (1 + j\omega T)^{-1} \approx 20 \log |j\omega T|^{-1} = -20 \log \omega T \qquad (8.23)$$

The value of Eq. (8.23) at $\omega = 1/T$ is 0. For values of $\omega > 1/T$ this function is a straight line with a negative slope of 6 dB/octave. Therefore, the asymptotes of the plot of Lm $(1 + j\omega T)^{-1}$ are two straight lines, one of zero slope below $\omega = 1/T$and one of -6 dB/octave slope above $\omega = 1/T$. These asymptotes are drawn in Fig. 8.2.

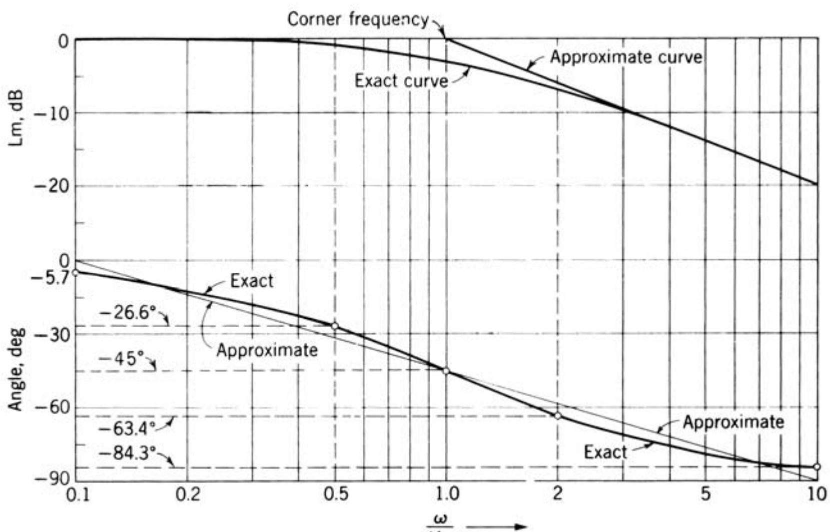


FIGURE 8.2 Log magnitude and phase diagram for $(1 + j\omega T)^{-1} = [1 + j(\omega/\omega_{cf})]^{-1}$.

Table 8.2 Values of Lm $(1+j\omega T)^{-1}$ for Several Frequencies

ω/ω_{cf}	Exact value, dB	Value of the asymptote, dB	Error, dB
0.1	−0.04	0	−0.04
0.25	−0.26	0	−0.26
0.5	−0.97	0	−0.97
0.76	−2.00	0	−2.00
1	−3.01	0	−3.01
1.31	−4.35	−2.35	−2.00
2	−6.99	−6.02	−0.97
4	−12.30	−12.04	−0.26
10	−20.04	−20.0	−0.04

The frequency at which the asymptotes to the log magnitude curve interesect is defined as the corner frequency ω_{cf}. The value $\omega_{cf} = 1/T$ is the corner frequency for the function $(1 + j\omega T)^{\pm r} = (1 + j\omega/\omega_{cf})^{\pm r}$.

The exact values of Lm $(1 + j\omega T)^{-1}$ are given in Table 8.2 for several frequencies in the range a decade above and below the corner frequency. The exact curve is also drawn in Fig. 8.2. The error, in dB, between the exact curve and the asymptotes is approximately as follows:

1. At the corner frequency: 3 dB
2. One octave above and below the corner frequency: 1 dB
3. Two octaves from the corner frequency: 0.26 dB

Frequently the preliminary design studies are made by using the asymptotes only. The correction to the straight-line approximation to yield the true Lm curve is shown in Fig. 8.3. For more exact studies the corrections are put in at the corner frequency and at 1 octave above and below the corner frequency, and the new curve is drawn with a French curve. For a more accurate curve a CAD program should be used.

The phase angle curve for this function is also plotted in Fig. 8.2. At zero frequency the angle is 0°; at the corner frequency $\omega = \omega_{cf}$ the angle is -45°; and at infinite frequency the angle is -90°. The angle curve is symmetrical about the corner-frequency value when plotted against $\log(\omega/\omega_{cf})$ or $\log \omega$. Since the abscissa of the curves in Fig. 8.2 is ω/ω_{cf}, the shapes of the angle and Lm curves are independent of the time constant T. Thus, when the curves are plotted with the abscissa in terms of ω, changing T just "slides" the Lm and the angle curves left or right so that the -3 dB and the -45°points occur at the frequency $\omega = \omega_{cf}$.

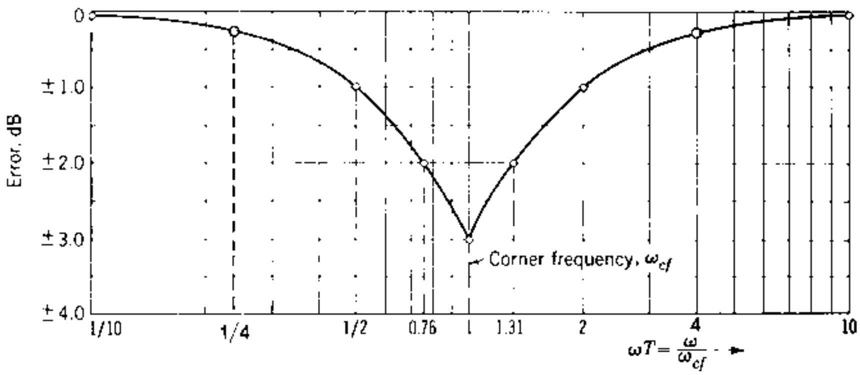


FIGURE 8.3 Log magnitude correction for $(1 + j\omega T)^{\pm 1}$

For preliminary design studies, straight-line approximations of the phase curve can be used. The approximation is a straight line drawn through the following three points:

ω/ω_{cf}	0.1	1.0	10
Angle	0°	−45°	−90°

The maximum error resulting from this approximation is about $\pm 6°$. For greater accuracy, a smooth curve is drawn through the points given in Table 8.3.

The factor $1 + j\omega T$ appearing in the numerator has the log magnitude

$$\text{Lm}\,(1 + j\omega T) = 20 \log \sqrt{1 + \omega^2 T^2}$$

This is the same function as its inverse $\text{Lm}\,(1 + j\omega T)^{-1}$ except that it is positive. The corner frequency is the same, and the angle varies from 0 to 90° as the frequency increases from zero to infinity. The Lm and angle curves for the function $1 + j\omega T$ are symmetrical about the abscissa to the curves for $(1 + j\omega T)^{-1}$.

TABLE 8.3 Angles of $(1 + j\omega/\omega_{cf})^{-1}$ for Key Frequency Points

ω/ω_{cf}	Angle, deg
0.1	−5.7
0.5	−26.6
1.0	−45.0
2.0	−63.4
10.0	−84.3

Quadratic Factors

Quadratic factors in the denominator of the transfer function have the form

$$\left[1+\frac{2\zeta}{\omega_{\rm n}}j\omega+\frac{1}{\omega_n^2}(j\omega)^2\right]^{-1} \tag{8.24}$$

For $\zeta > 1$ the quadratic can be factored into two first-order factors with real zeros that can be plotted in the manner shown previously. But for $\zeta < 1$ Eq. (8.24) contains conjugate-complex factors, and the entire quadratic is plotted without factoring:

$$\text{Lm}\left[1+\frac{2\zeta}{\omega_n}j\omega+ > \frac{1}{\omega_n^2}(j\omega)^2\right]^{-1} = -20\ \log\left[\left(1-\frac{\omega^2}{\omega_n^2}\right)^2+\left(\frac{2\zeta\omega}{\omega_n}\right)^2\right]^{1/2} \tag{8.25}$$

$$\text{Angle}\left[1+\frac{2\zeta}{\omega_n}j\omega+\frac{1}{\omega_n^2}(j\omega)^2\right]^{-1} = -\tan^{-1}\frac{2\zeta\omega/\omega_n}{1-\omega^2/\omega_n^2} \tag{8.26}$$

From Eq. (8.25) it is seen that for every small values of ω the low-frequency asymptote is represented by $\text{Lm} = 0$ dB. For every high values of frequency, the Lm is approximately

$$-20\ \log\frac{\omega^2}{\omega_n^2} = -40\log\frac{\omega}{\omega_n}$$

and the high-frequency asymptote has a slope of -40 dB/decade. The asymptotes cross at the corner frequency $\omega_{\rm cf} = \omega_{\rm n}$.

From Eq. (8.25) it is seen that a resonant condition exists in the vicinity of $\omega = \omega_n$, where the peak value of the Lm is greater than 0 dB. Therefore, there may be a substantial deviation of the Lm curve from the straight-line asymptotes, depending on the value of ζ. A family of curves for several values of $\zeta < 1$ is plotted in Fig. 8.4. For the appropriate ζ the curve can be drawn by selecting sufficient points from Fig. 8.4 or computed from Eq. (8.25).

The phase-angle curve for this function also varies with ζ. At zero frequency the angle is 0°, at the corner frequency the angle is -90°, and at infinite frequency the angle is -180°. A family of curves for various values of $\zeta < 1$ is plotted in Fig. 8.4. Enough values to draw the appropriate curve can be taken from Fig. 8.4 or computed from Eq. (8.26). When the quadratic factor appears in the numerator, the magnitude of the Lm and phase angle are the same as those in Fig. 8.4 except that they are changed in sign.

The $\text{Lm}[1 + j2\zeta\omega/\omega_n + (j\omega/\omega_n)^2]^{-1}$ with $\zeta < 0.707$ has a peak value. The magnitude of this peak value and the frequency at which it occurs are

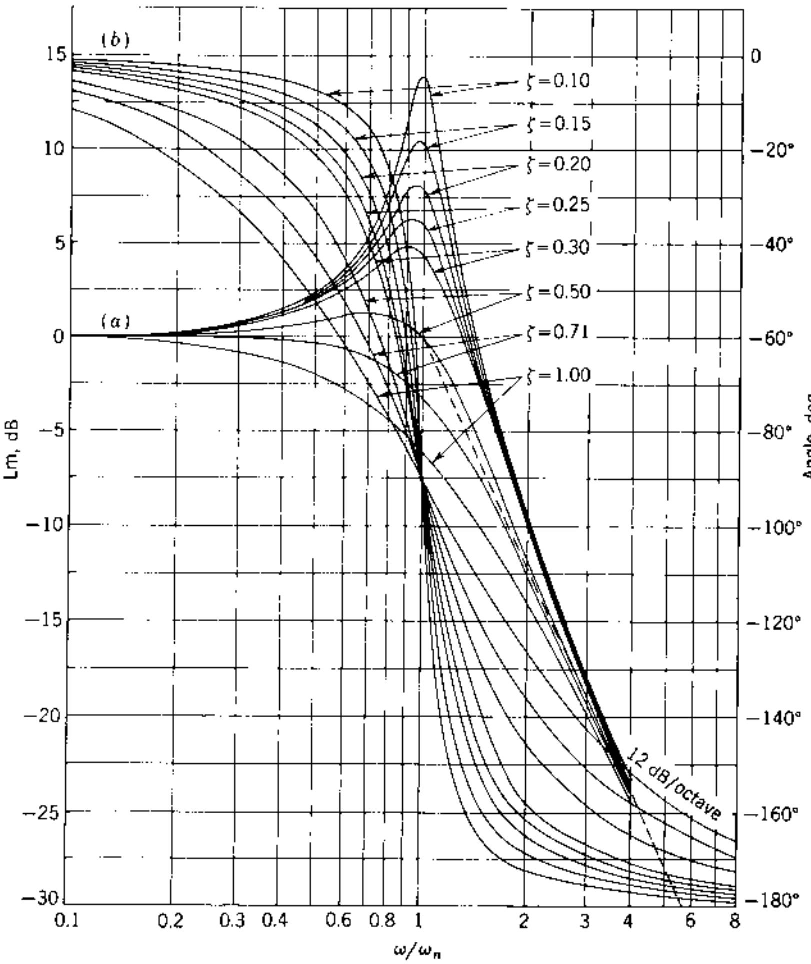


FIGURE 8.4 Log magnitude and phase diagram for $[1 + j2\zeta\omega/\omega_n + (j\omega/\omega_n)^2]^{-1}$.

important terms. These values, given earlier in Sec. 4.12 and derived in Sec. 9.3, are repeated here:

$$M_m = \frac{1}{2\zeta\sqrt{1-\zeta^2}} \tag{8.27}$$

$$\omega_m = \omega_n\sqrt{1-2\zeta^2} \tag{8.28}$$

Note that the peak value M_m depends only on the damping ratio ζ. Since Eq. (8.28) is meaningful only for real values of ω_m, the curve of M vs. ω has a peak value greater than unity only for $\zeta < 0.707$. The frequency at which the peak value occurs depends on both the damping ratio ζ and the undamped natural frequency ω_n. This information is used when adjusting a control system for good response characteristics. These characteristics are discussed in Chap. 9.

The discussion thus far has dealt with poles and zeros that are located in the left-half (LH) s plane. The Lm curves for poles and zeros lying in the right-half (RH) s plane are the same as those for poles and zeros located in the LH s plane. However, the angle curves are different. For example, the angle for the factor $(1 - j\omega T)$ varies from 0 to -90° as ω varies from zero to infinity. Also, if ζ is negative, the quadratic factor of expression (8.24) contains RH s plane poles or zeros. Its angle varies from -360° at $\omega = 0$ to -180° at $\omega = \infty$. This information can be obtained from the pole-zero diagram discussed in Sec. 4.12, with all angles measured in a *counterclockwise* (ccw) direction. Some CAD packages do not consistently use a ccw measurement direction and thus yield inaccurate angle values.

8.7 EXAMPLE OF DRAWING A BODE PLOT

Figure 8.5 is the block diagram of a feedback control system with unity feedback. In this section the Lm and phase diagram is drawn in a systematic manner for the open-loop transfer function of this system. The Lm curve is drawn both for the straight-line approximation and for the exact curve. Table 8.4 lists the pertinent characteristics for each factor. The Lm asymptotes and angle curves for each factor are shown in Figs. 8.6 and 8.7, respectively. They are added algebraically to obtain the composite curve. The composite Lm curve using straight-line approximations is drawn directly, as outlined below.

Step 1. At frequencies less than ω_1, the first corner frequency, only the factors Lm 4 and $\mathrm{Lm}(j\omega)^{-1}$ are effective. All the other factors have zero value. At ω_1, Lm $4 = 12$ dB and $\mathrm{Lm}(j\omega_1)^{-1} = 6$ dB; thus, at this frequency the composite curve has the value of 18 dB. Below ω_1 the composite curve has a slope of -20 dB/decade because of the $\mathrm{Lm}\,(j\omega)^{-1}$ term.

Step 2. Above ω_1, the factor $\mathrm{Lm}\,(1 + j2\omega)^{-1}$ has a slope of -20 dB/decade and must be added to the terms in step 1. When the slopes are added,

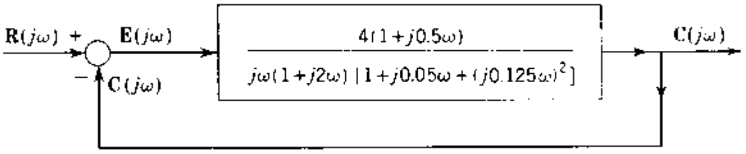


FIGURE 8.5 Block diagram of a control system with unity feedback.

TABLE 8.4 Characteristics of Log Magnitude and Angle Diagrams for Various Factors

Factor	Corner frequency, ω_{cf}	Log magnitude	Angle characteristics
4	None	Constant magnitude of $+12$ dB	Constant 0°
$(j\omega)^{-1}$	None	Constant slope of -20 dB/decade (0 dB at $\omega = 1$)	Constant $-90°$
$(1 + j2\omega)^{-1}$	$\omega_1 = 0.5$	0 slope below corner frequency; -20 dB/decade slope above corner frequency. (18 dB at ω_1)	Varies from 0 to $-90°$
$1 + j0.5\omega$	$\omega_2 = 2.0$	0 slope below corner frequency; $+20$ dB/decade slope above corner frequency. (-6 dB at ω_2)	Varies from 0 to $+90°$
$[1 + j0.05\omega + (j0.125\omega)^2]^{-1}$ $\zeta = 0.2$ $\omega_n = 8$	$\omega_3 = 8.0$	0 slope below corner frequency; -40 dB/decade slope above corner frequency. (-18 dB at ω_3)	Varies from 0 to $-180°$

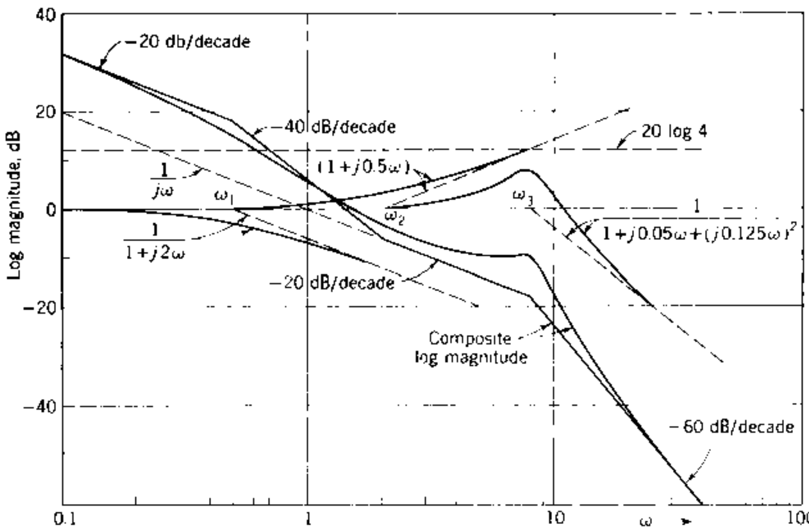


FIGURE 8.6 Log magnitude curve for

$$\mathbf{G}(j\omega) = \frac{4(1 + j0.5\omega)}{j\omega(1 + j2\omega) \times [1 + j0.05\omega + (j0.125\omega)^2]}.$$

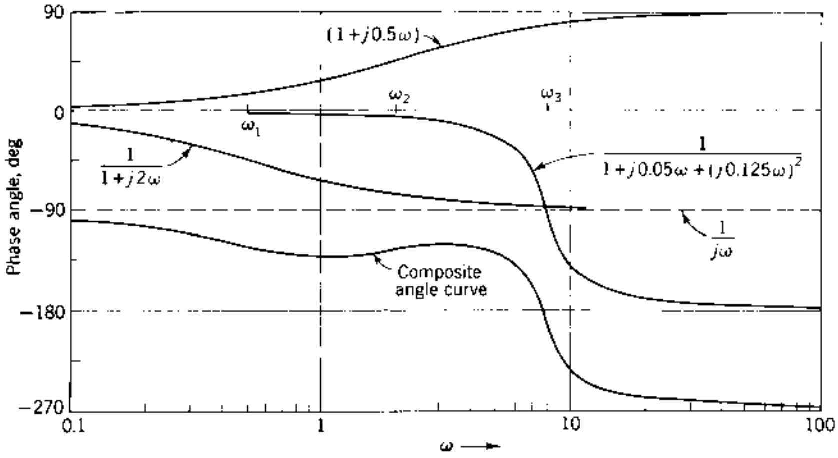


FIGURE 8.7 Phase-angle curve for

$$\mathbf{G}(j\omega) = \frac{4(1+j0.5\omega)}{j\omega(1+j2\omega)[1+j0.05\omega+(j0.125\omega)^2]}$$

the composite curve has a total slope of –40 dB/decade in the frequency band from ω_1 to ω_2. Since this bandwidth is 2 octaves, the value of the composite curve at ω_2 is –6 dB.

Step 3. Above ω_2, the numerator factor Lm $(1+j0.5\omega)$ is effective. This factor has a slope of $+20$ dB/decade above ω_2 and must be added to obtain the composite curve. The composite curve now has a total slope of –20 dB/decade in the frequency band from ω_2 to ω_3. The bandwidth from ω_2 to ω_3 is 2 octaves; therefore, the value of the composite curve at ω_3 is –18 dB.

Step 4. Above ω_3 the last term Lm $[1+j0.05\omega+(j0.125\omega)^2]^{-1}$ must be added. This factor has a slope of –40 dB/decade; therefore, the total slope of the composite curve above ω_3 is –60 dB/decade.

Step 5. Once the asymptotic plot of Lm G$(j\omega)$ has been drawn, the corrections can be added if desired. The corrections at each corner frequency and at an octave above and below the corner frequency are usually sufficient. For first-order terms the corrections are ± 3 dB at the corner frequencies and ± 1 dB at an octave above and below the corner frequency. The values for quadratic terms can be obtained from Fig. 8.4 since they are a function of ζ. If the correction curve for the quadratic factor is not available, the correction at the frequencies $\omega = \omega_n$ and $\omega = 0.707\omega_n$ can easily be calculated from Eq. (8.25). The correction at ω_m can also be obtained using Eqs. (8.27 and (8.28).

The corrected Lm curves for each factor and for the composite curve are shown in Fig. 8.6.

Determination of the phase-angle curve of G($j\omega$) can be simplified by using the following procedure.

1. For the $(j\omega)^{-m}$ term, draw a line at the angle of $(-m)90^\circ$.
2. For each $(1+j\omega T)\pm 1$ term, locate the angles from Table 8.4 at the corner frequency, an octave above and an octave below the corner frequency, and a decade above and below the corner frequency. Then draw a curve through these points for each $(1+j\omega T)\pm 1$ term.
3. For each $1+j2\zeta\omega/\omega_n+(j\omega/\omega_n)^2$ term:

   (a) Locate the $\pm 90^\circ$ point at the corner frequency $\omega=\omega_n$.

   (b) From Fig. 8.4, for the respective ζ, obtain a few points to draw the phase plot for each term with the aid of a French curve. The angle at $\omega = 0.707\omega_n$ may be sufficient and can be evaluated easily from Eq. (8.26) as $\tan^{-1} 2.828\zeta$.
4. Once the phase plot of each term of G($j\omega$) has been drawn, the composite phase plot of G($j\omega$) is determined by adding the individual phase curves.

   (a) Use the line respresenting the angle equal to

   $$\underline{/\lim_{\omega\to 0} G(j\omega)} = (-m)90^\circ$$

   as the base line, where m is the system type. Add or subtract the angles of each factor from this reference line.

   (b) At a particular frequency on the graph, measure the angle for each single-order and quadratic factor. Add and/or subtract them from the base line until all terms have been accounted for. The number of frequency points used is determined by the desired accuracy of the phase plots.

   (c) At $\omega=\infty$ the phase is 90° times the difference of the orders of the numerator and denominator $[-(n-\omega)90^\circ]$.

The characteristics of the log magnitude-angle diagram for this example are described in Sec. 8.18. Both the curves obtained from the asymptotes and the exact plots are shown in Fig. 8.6 and 8.7. The data for this diagram can be obtained from MATLAB.

8.8 GENERATION OF MATLAB BODE PLOTS

Accurate plots for Lm G($j\omega$) and $\underline{/G(j\omega)}$ can be obtained by using MATLAB (see Fig. 8.8). The MATLAB commands shown below are used to obtain Bode

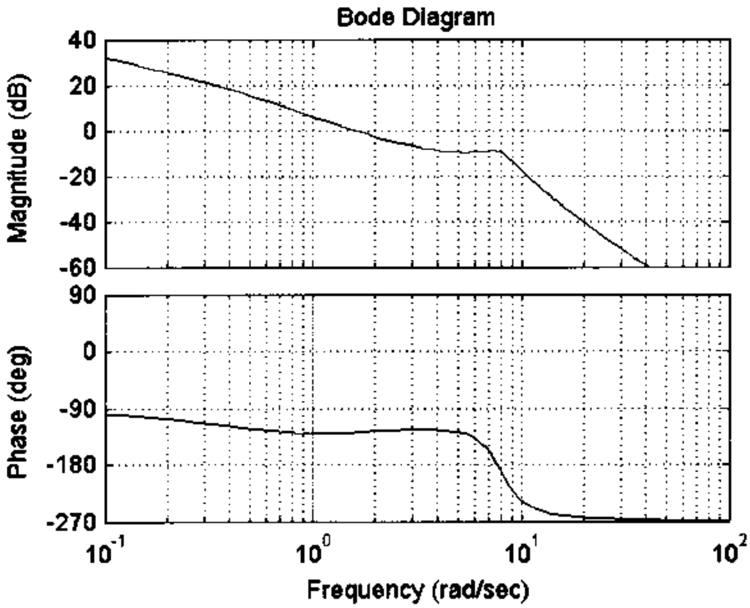


FIGURE 8.8 MATLAB Bode plots corresponding to Fig. 8.6 and Fig. 8.7.

plots for the transfer function.

$$G(s) = \frac{4(1 + j0.5\omega)}{j\omega(1 + j2\omega)[1 + j0.05\omega + (j0.125\omega)^2} \tag{8.29}$$

The figure format and layout is adjusted within the figure window. Further details of a frequency response m-file are shown in Sec. 9.9 and App. C.

```
num=4*[.5 1]
den=conv ([1 0], conv ([2 1], [.125^2 .05 1]))
bode (num, den)
```

8.9 SYSTEM TYPE AND GAIN AS RELATED TO LOG MAGNITUDE CURVES

The steady-state error of a closed-loop system depends on the system type and the gain. The system error coefficients are determined by these two characteristics, as noted in Chap. 6. For any given Lm curve the system type and gain can be determined. Also, with the transfer function given so that the system type and gain are known, they can expedite drawing the Lm curve. This is described for Type 0, 1, and 2 systems.

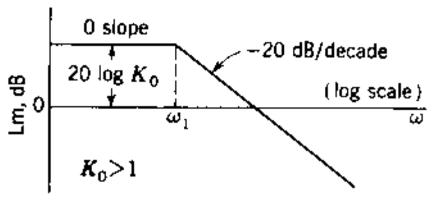


FIGURE 8.9 Log magnitude plot for $\mathbf{G}(j\omega) = K_0/(1 + j\omega T_a)$.

Type 0 Systems

A first-order Type 0 system has a transfer function of the form

$$\mathbf{G}(j\omega) = \frac{K_0}{1 + j\omega T_a}$$

At low frequencies, $\omega < 1/T_a$, Lm $\mathbf{G}(j\omega) \approx 20 \log K_0$, which is a constant. The slope of the Lm curve is zero below the corner frequency $\omega_1 = 1/T_a$ and –20 dB/decade above the corner frequency. The Lm curve is shown in Fig. 8.9.

For a Type 0 system the characteristics are as follows:

1. The slope at low frequencies is zero.
2. The magnitude at low frequencies is 20 log K_0.
3. The gain K_0 is the steady-state step error coefficient.

Type 1 System

A second-order Type 1 system has a transfer function of the form

$$\mathbf{G}(j\omega) = \frac{K_1}{j\omega(1 + j\omega T_a)}$$

At low frequencies, $\omega < 1/T_a$, Lm $\mathbf{G}(j\omega) \approx$ (Lm $(K_1/j\omega) =$ Lm $K_1 -$ Lm $j\omega$, which has a slope of –20 dB/decade. At $\omega = K_1$, Lm $(K_1/j\omega) = 0$. If the corner frequency $\omega_1 = 1/T_a$ is greater than K_1, the low-frequency portion of the curve of slope −20 dB/decade crosses the 0-dB axis at a value of $\omega_x = K_1$, as shown in Fig. 8.10*a*. If the corner frequency is less than K_1, the low-frequency portion of the curve of slope –20 dB/decade may be extended until it does cross the 0-dB axis. The value of the frequency at which the extension crosses the 0-dB axis is $\omega_x = K_1$. In other words, the plot Lm $(K_1/j\omega)$ crosses the 0-dB value at $\omega_x = K_1$, as illustrated in Fig. 8.10*b*.

At $\omega = 1$, Lm $j\omega = 0$; therefore, Lm $(K_1/j\omega)_{\omega=1} = 20 \log K_1$. For $T_a < 1$ this value is a point on the slope of –20 dB/decade. For $T_a > 1$ this value is a point on the extension of the initial slope, as shown in Fig. 8.10*b*. The frequency ω_x is smaller or larger than unity according as K_1 is smaller or larger than unity.

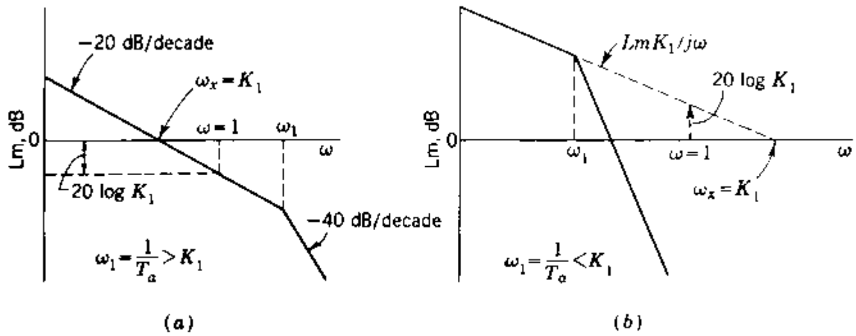


FIGURE 8.10 Log magnitude plot for $G(j\omega) = K1/j\omega(1 + j\omega T_a)$.

For a Type 1 system the characteristics are as follows:

1. The slope at low frequencies is –20 dB/decade.
2. The intercept of the low-frequency slope of –20 dB/decade (or its extension) with the 0-dB axis occurs at the frequency ω_x, where $\omega_x = K_1$.
3. The value of the low-frequency slope of –20 dB/decade (or its extension) at the frequency $\omega = 1$ is equal to 20 log K_1.
4. The gain K_1 is the steady-state ramp error coefficient.

Type 2 System

A third-order Type 2 system has a transfer function of the form

$$G(j\omega) = \frac{K_2}{(j\omega)^2(1 + j\omega T_a)}$$

At low frequencies, $\omega < 1/T_a$, $\text{Lm } G(j\omega) = \text{Lm}[K_2/(j\omega)^2] = \text{Lm}K_2 - \text{Lm}(j\omega)^2$, for which the slope is –40 dB/decade. At $\omega^2 = K_2$, $\text{Lm}\ [K_2/(j\omega)^2] = 0$; therefore, the intercept of the initial slope of –40 dB/decade (or its extension, if necessary) with the 0-dB axis occurs at the frequency ω_y, where $\omega_y^2 = K_2$.

At $\omega = 1$, $\text{Lm}\ (j\omega)^2 = 0$; therefore, $\text{Lm}\ [K_2/(j\omega)^2]_{\omega=1} = 20 \log K_2$. This point occurs on the initial slope or on its extension, according to whether $\omega_1 = 1/T_a$ is larger or smaller than $\sqrt{K_2}$. If $K_2 > 1$, the quantity $20 \log K_2$ is positive, and if $K_2 < 1$, the quantity $20 \log K_2$ is negative.

The Lm curve for a Type 2 transfer function is shown in Fig. 8.11. The determination of gain K_2 from the graph is shown.

For a Type 2 system the characteristics are as follows:

1. The slope at low-frequencies is –40 dB/decade.

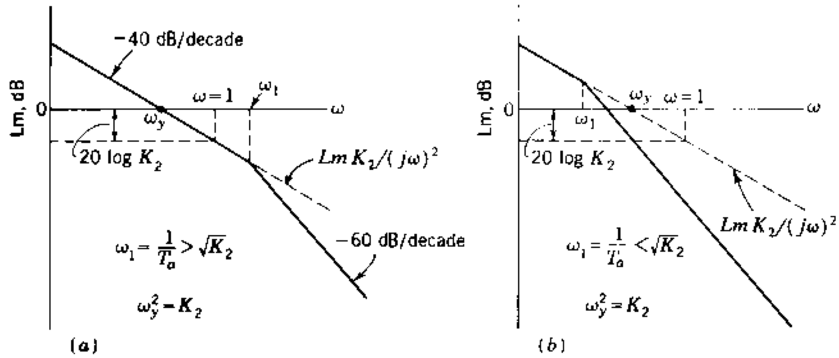


FIGURE 8.11 Log magnitude plot for $G(j\omega) = K_2/(j\omega)^2(1 + j\omega T_a)$.

2. The intercept of the low-frequency slope of –40 dB/decade (or its extension, if necessary) with the 0-dB axis occurs at a frequency ω_y, where $\omega_y^2 = K_2$.
3. The value on the low-frequency slope of –40 dB/decade (or its extension) at the frequency $\omega = 1$ is equal to $20 \log K_2$.
4. The gain K_2 is the steady-state parabolic error coefficient.

8.10 CAD ACCURACY CHECK (CADAC)

CADAC checks that apply for CAD package frequency domain analysis are the following:

1. Based on the given $G(j\omega)$, determine the angle of this function, assuming a minimum-phase stable plant, at $\omega = 0$ and at $\omega = \infty$ (see Sec. 8.7). if the plant has any RH plane poles and/or zeros, it is necessary to ascertain whether the CAD package being used correctly measures all angles in the ccw (or cw) direction as ω goes from 0 to ∞.
2. Based on the system type of $G(s)$, the Lm $G(j\omega)$ can be readily determined at $\omega = 0$ for a Type 0 plant. The value of ω_x for a Type 1 plant or ω_y for a Type 2 plant should be verified.

8.11 EXPERIMENTAL DETERMINATION OF TRANSFER FUNCTION [5,9]

The Lm and phase-angle diagram is of great value when the mathematical expression for the transfer function of a given system is not known. The magnitude and angle of the ratio of the output to the input can be obtained experimentally for a steady-state sinusoidal input signal at a number of frequencies. These data are used to obtain the exact Lm and angle diagram. Asymptotes are drawn on the exact Lm curve, using the fact that their slopes must be multiples of $\pm$ 20 dB/decade. From these asymptotes and the intersection the system type and the approximate time constants are determined. In this manner, the transfer function of the system can be synthesized.

Care must be exercised in determining whether any zeros of the transfer function are in the right half (RH) s plane. A system that has no open-loop zeros in the RH s plane is defined as a *minimum-phase system* [10,11] and all factors have the form $1 + Ts$ and/or $1 + As + Bs^2$. A system that has open-loop zeros in the RH s plane is a *nonminimum-phase system* [12,13]. The stability is determined by the location of the poles and does not affect the designation of minimum or nonminimum phase.

The angular variation for poles or zeros in the RH s plane is different from those in the LH plane [10]. For this situation, one or more terms in the transfer function have the form $1 - Ts$ and/or $1 \pm As \pm Bs^2$. As an example, consider the functions $1 + j\omega T$ and $1 - j\omega T$. The Lm plots of these functions are identical, but the angle diagram for the former goes from 0 to 90° whereas for the latter it goes from 0 to –90°. Therefore, care must be exercised in interpreting the angle plot to determine whether any factors of the transfer function lie in the RH s plane. Many practical systems are in the minimum-phase category.

Unstable plants must be handled with care. That is, first a stabilizing compensator must be added to form a stable closed-loop system. From the experimental data for the stable closed-loop system, the plant transfer function is determined by using the known compensator transfer function.

8.12 DIRECT POLAR PLOTS

The earlier sections of this chapter present a simple graphical method for plotting the characteristic curves of the transfer functions. These curves are the Lm and the angle of $G(j\omega)$ vs. ω, plotted on semilog graph paper. The reason for presenting this method first is that these curves can be constructed easily and rapidly. However, frequently the polar plot of the transfer function is desired. The magnitude and angle of $G(j\omega)$, for sufficient frequency points, are readily obtainable from the Lm $G(j\omega)$ and $\angle G(j\omega)$ vs. log ω curves. This approach often takes less time than calculating $|G(j\omega)|$ and $\phi(\omega)$ analytically for each frequency point desired, unless a CAD program is available (see Appendix C).

The data for drawing the polar plot of the frequency response can also be obtained from the pole-zero diagram, as described in Sec. 4.12. It is possible to visualize the complete shape of the frequency-response curve from the pole-zero diagram because the angular contribution of each pole and zero is readily apparent. The polar plot of $G(j\omega)$ is called the direct polar plot. The polar plot

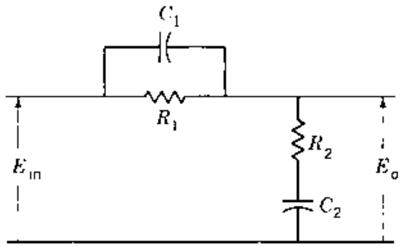


FIGURE 8.12 An RC circuit used as a lag-lead compensator.

of $[G(j\omega)]^{-1}$ is called the *inverse polar plot* [11], which is discussed briefly in Chap. 9.

Complex *RC* Network (Lag-Lead Compensator)

The circuit of Fig. 8.12 is used in later chapters as a compensator:

$$G(s) = \frac{E_o(s)}{E_{in}(s)} = \frac{1 + (T_1 + T_2)s + T_1 T_2 s^2}{1 + (T_1 + T_2 + T_{12})s + T_1 T_2 s^2} \tag{8.30}$$

Where the time constants are

$$T_1 = R_1 C_1 \qquad T_2 = R_2 C_2 \qquad T_{12} = R_1 C_2$$

As a function of frequency, the transfer function is

$$G(j\omega) = \frac{E_o(j\omega)}{E_{in}(j\omega)} = \frac{\left(1 - \omega^2 T_1 T_2\right) + j\omega(T_1 + T_2)}{\left(1 - \omega^2 T_1 T_2\right) + j\omega(T_1 + T_2 + T_{12})} \tag{8.31}$$

By the proper choice of the time constants, the circuit acts as a lag network in the lower-frequency range of 0 to ω_x and as a lead network in the higher-frequency range of ω_x to ∞. That is, the steady-state sinusoidal output E_o lags or leads the sinusoidal input E_{in}, according to whether ω is smaller or larger than ω_x. The polar plot of this transfer function is a circle with its center on the real axis and lying in the first and fourth quadrants. Figure 8.13 illustrates the polar plot in nondimensionalized form for a typical circuit. When T_2 and T_{12} in Eq. (8.31) are expressed in terms of T_1, the expression of the form given in the title of Fig. 8.13 results. Its properties are the following:

1. $\lim_{\omega \to 0} \mathbf{G}(j\omega T_1) \to 1\,\underline{/0^\circ}$
2. $\lim_{\omega \to \infty} \mathbf{G}(j\omega T_1) \to 1\,\underline{/0^\circ}$
3. At the frequency $\omega = \omega_x$, for which $\omega_x^2 T_1 T_2 = 1$, Eq. (8.31) becomes

$$\mathbf{G}(j\omega_x T_1) = \frac{T_1 + T_2}{T_1 + T_2 + T_{12}} = \left|\mathbf{G}(j\omega_x T_1)\right| \angle 0^0 \tag{8.32}$$

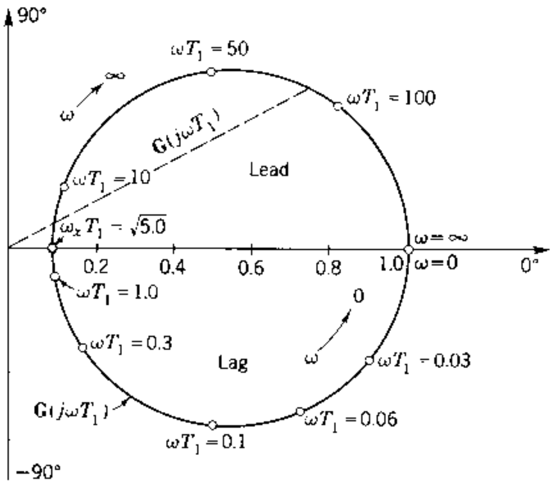


FIGURE 8.13 Polar plot of

$$G(j\omega T_1) = \frac{(1 + j\omega T_1)(1 + j0.2\omega T_1)}{(1 + j11.1\omega T_1)(1 + j0.0179\omega T_1)}$$

$T_2 = 0.2T_1,\ T_{12} = 10.0T_1.$

Note that Eq. (8.32) represents the minimum value of the transfer function in the whole frequency spectrum. From Fig. 8.13 it is seen that for frequencies below ω_x the transfer function has a negative or lag angle. For frequencies above ω_x it has a positive or lead angle. The applications and advantages of this circuit are discussed in later chapters on compensation.

Type 0 Feedback Control System

The field-controlled servomotor [6] illustrates a typical Type 0 device. It has the transfer function

$$\mathbf{G}(j\omega) = \frac{\mathbf{C}(j\omega)}{\mathbf{E}(j\omega)} = \frac{K_0}{\left(1 + j\omega T_f\right)(1 + j\omega T_m)} \tag{8.33}$$

Note from Eq. (8.33) that

$$\mathbf{G}(j\omega) \to \begin{cases} K_0 \angle 0^\circ & \omega \to 0^+ \\ 0 \angle -180^\circ & \omega \to \infty \end{cases}$$

Also, for each term in the denominator the angular contribution to $\mathbf{G}(j\omega)$, as ω goes from 0 to ∞, goes from 0 to -90°. Thus, the polar plot of this transfer function must start at $\mathbf{G}(j\omega) = K_0 \angle 0^\circ$ for $\omega = 0$ and proceeds first through

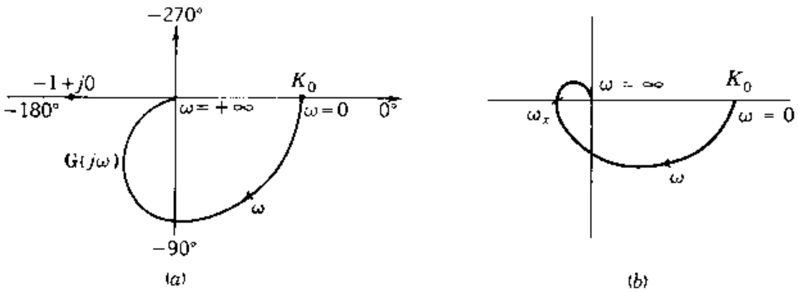


FIGURE 8.14 Polar plot for typical Type 0 transfer functions: (*a*) Eq. (8.33); (*b*) Eq. (8.34).

the fourth and then through the third quadrants to $\lim_{\omega\to\infty} \mathbf{G}(j\omega) = 0\,\underline{/-180^\circ}$ as the frequency approaches infinity, as shown in Fig. 8.14 (a). In other words, the angular variation of $\mathbf{G}(j\omega)$ is continuously decreasing, going in a clockwise (cw) direction from 0° to -180°. The exact shape of this plot is determined by the particular values of the time constants T_f and T_m.

If Eq. (8.33) had another term of the form $1 + j\omega T$ in the denominator, the transfer function would be

$$\mathbf{G}(j\omega) = \frac{K_0}{(1 + j\omega T_f)(1 + j\omega T_m)(1 + j\omega T)} \tag{8.34}$$

The point $\mathbf{G}(j\omega)|_{\omega=\infty}$ rotates cw by an additional 90°. In other words, when $\omega \to \infty$, $\mathbf{G}(j\omega) \to 0\,\underline{/-270^\circ}$. In this case the curve crosses the real axis as shown in Fig. 8.14(b) at a frequency ω_x for which the imaginary part of the transfer function is zero.

When a term of the form $1 + j\omega T$ appears in the numerator, the transfer function experiences an angular variation of 0 to 90° (a ccw rotation) as the frequency is varied from 0 to ∞. Thus, with increasing frequency, the angle of $\mathbf{G}(j\omega)$ may not change continuously in one direction. Also, the resultant polar plot may not be as smooth as the one shown in Fig. 8.14. For example, the transfer function

$$G(j\omega) = \frac{K_0(1 + j\omega T_1)^2}{(1 + j\omega T_2)(1 + j\omega T_3)(1 + j\omega T_4)^2} \tag{8.35}$$

whose polar plot, shown in Fig. 8.15, has a "dent" when the time constants T_2 and T_3 are greater than T_1, and T_1 is greater than T_4. From the angular contribution of each factor and from the preceding analysis, it can be surmised that the polar plot has the general shape shown in the figure. In the event that T_1 is smaller than all the other time constants, the polar plot is similar in shape to the one shown in Fig. 8.14a.

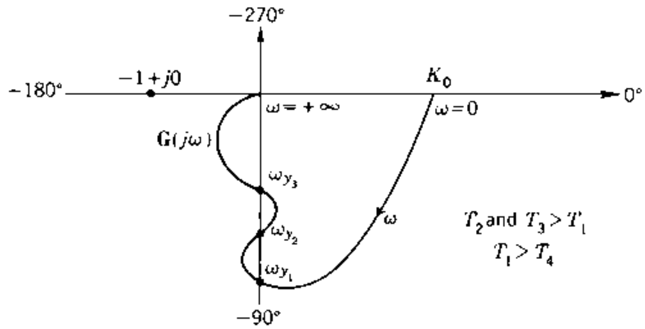


FIGURE 8.15 Polar plot for a more complex Type 0 transfer function [Eq. (8.35)].

In the same manner, a quadratic in either the numerator or the denominator of a transfer function results in an angular contribution of 0 to $\pm 180^\circ$, respectively, and the polar plot of $G(j\omega)$ is affected accordingly. It can be seen from the examples that the polar of a Type 0 system always starts at a value K_0 (step error coefficient) on the positive real axis for $\omega = 0$ and ends at zero magnitude (for $n > w$) and tangent to one of the major axes at $\omega = \infty$. The final angle is -90° times the order of the denominator minus the order of the numerator of $G(j\omega)$. The arrows on the plots of $G(j\omega)$ in Figs. 8.14 and 8.15 indicate the direction of increasing frequency.

Type 1 Feedback Control System

The transfer function of a typical Type 1 system is

$$G(j\omega) = \frac{C(j\omega)}{E(j\omega)} = \frac{K_1}{j\omega(1 + j\omega T_m)(1 + j\omega T_c)\left(1 + j\omega T_q\right)} \tag{8.36}$$

Note from Eq. (8.36) that

$$G(j\omega) \to \begin{cases} \infty\angle{-90^\circ} & \omega \to 0^+ \qquad (8.37) \\ 0\angle{-360^\circ} & \omega \to \infty \qquad (8.38) \end{cases}$$

Note that the $j\omega$ term in the denominator contributes the angle -90° to the total angle of $G(j\omega)$ for all frequencies. Thus, the basic difference between Eqs. (8.34) and (8.36) is the presence of the term $j\omega$ in the denominator of the latter equation. Since all the $1 + j\omega T$ terms of Eq. (8.36) appear in the denominator, its polar plot, as shown in Fig. 8.16, has no dents. From the remarks of this and previous sections, it can be seen that the angular variation of $G(j\omega)$ decreases continuously (ccw) in the same direction from -90 to -360° as ω increases from 0 to ∞. The presence of any frequency-dependent factor in the

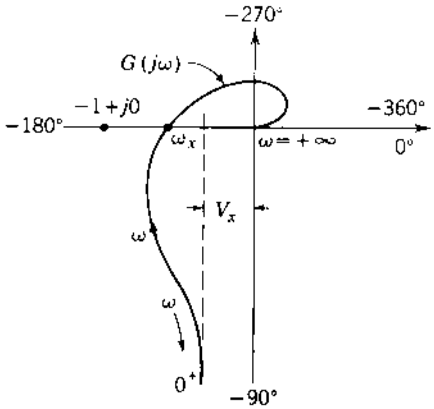


FIGURE 8.16 Polar plot for a typical Type 1 transfer function [Eq. (8.36)].

numerator has the same general effect on the polar plot as that described previously for the Type 0 system.

From Eq. (8.37) it is seen that the magnitude of the function G($j\omega$) approaches infinity as the value of ω approaches zero. There is a line parallel to the -90° axis but displaced to the left of the origin, to which G($j\omega$) tends asymptotically as $\omega \to 0$. The true asymptote is determined by finding the value of the real part of G($j\omega$) as ω approaches zero. Thus

$$V_x = \lim_{\omega \to 0} \mathrm{Re}\,[G(j\omega)] \tag{8.39}$$

or, for this particular transfer function,

$$V_x = -K_1(T_q + T_c + T_m) \tag{8.40}$$

Equation (8.40) shows that the magnitude of G($j\omega$) approaches infinity asymptotically to a vertical line whose real-axis intercept equals V_x, as illustrated in Fig. 8.16. Note that the value of V_x is a direct function of the ramp error coefficient K_1.

The frequency of the crossing point on the negative real axis of the G($j\omega$) function is that value of frequency ω_x for which the imaginary part of G($j\omega$) is equal to zero. Thus,

$$\mathrm{Im}\,[\mathrm{G}(j\omega_x)] = 0 \tag{8.41}$$

or, for this particular transfer function,

$$\omega_x = (T_c T_q + T_q T_m + T_m T_c)^{-1/2} \tag{8.42}$$

The real-axis crossing point is very important because it determines closed-loop stability, as described in later sections dealing with system stability.

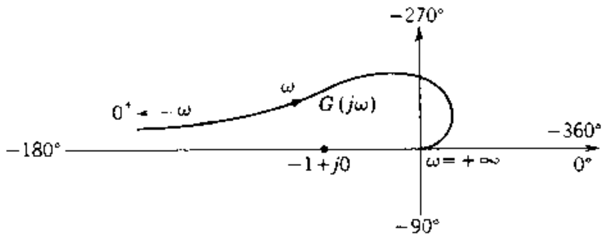


FIGURE 8.17 Polar plot of Eq. (8.43), a typical Type 2 transfer function resulting in an unstable feedback control system.

Type 2 Feedback Control System

The transfer function of the Type 2 system illustrated in Sec. 6.6 and represented by Eq. (6.56) is

$$\mathbf{G}(j\omega) = \frac{\mathbf{C}(j\omega)}{\mathbf{E}(j\omega)} = \frac{K_2}{(j\omega)^2\left(1 + j\omega T_f\right)(1 + j\omega T_m)} \tag{8.43}$$

Its properties are

$$\mathbf{G}(j\omega) \to \begin{cases} \infty \underline{/-180^\circ} & \omega \to 0^+ & (8.44) \\ 0 \underline{/-360^\circ} & \omega \to +\infty & (8.45) \end{cases}$$

The presence of the $(j\omega)^2$ term in the denominator contributes -180° to the total angle of $\mathbf{G}(j\omega)$ for all frequencies. For the transfer function of Eq. (8.43) the polar plot (Fig. 8.17) is a smooth curve whose angle $\phi(\omega)$ decreases continuously from -180 to -360°.

The introduction of an additional pole and a zero can alter the shape of the polar plot. Consider the transfer function

$$\mathbf{G}_0(j\omega) = \frac{K_2'(1 + j\omega T_1)}{(j\omega)^2\left(1 + j\omega T_f\right)(1 + j\omega T_m)(1 + j\omega T_2)} \tag{8.46}$$

where $T_1 > T_2$. The polar plot can be obtained from the pole-zero diagram shown in Fig. 8.18. At $s = j\omega = j0^+$ the angle of each factor is zero except for the double pole at the origin. The angle at $\omega = 0^+$ is therefore -180°, as given by Eq. (8.44), which is still applicable. As ω increases from zero, the angle of $j\omega + 1/T_1$ increases faster than the angles of the other poles. In fact, at low frequencies the angle due to the zero is larger than the sum of the angles due to the poles located to the left of the zero. This fact is shown qualitatively at the frequency ω_1 in Fig. 8.18. Therefore, the angle of $\mathbf{G}(j\omega)$ at low frequencies is greater that -180°. As the frequency increases to a value ω_x, the sum of the component angles of $\mathbf{G}(j\omega)$ is -180° and the

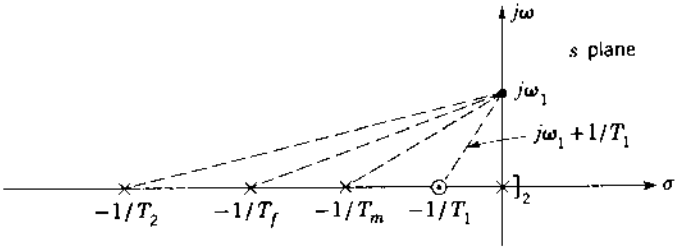


FIGURE 8.18 Pole-zero diagram for

$$G_o(s) = \frac{K_2' T_1}{T_f T_m T_2} \frac{s + 1/T_1}{s^2(s + 1/T_f)(s + 1/T_m)(s + 1/T_2)}.$$

polar plot crosses the real axis, as shown in Fig. 8.19. As ω increases further, the angle of $j\omega + 1/T_1$ shows only a small increase, but the angles from the poles increase rapidly. In the limit, as $\omega \to \infty$, the angles of $j\omega + 1/T_1$ and $j\omega + 1/T_2$ are equal and opposite in sign, so the angel of $G(j\omega)$ approaches -360°, as given by Eq. (8.45).

Fig. 8.19 shown the complete polar plot of $G(j\omega)$. A comparison of Figs. 8.17 and 8.19 shows that both curves approach -180° at $\omega = 0^+$, which is typical of Type 2 systems. As $\omega \to \infty$, the angle approaches -360° since both $G(j\omega)$ and $G_o(j\omega)$ have the same degree, $n\text{–}w = 4$. When the Nyquist stability criterion described in Secs. 8.14 and 8.15 is used, the feedback system containing $G(j\omega)$ can be shown to be unstable, whereas the closed-loop system containing $G_0(j\omega)$ can be stable.

It can be shown that as $\omega \to 0^+$, the polar plot of a Type 2 system is below the real axis if $\sum(T_{\text{numerator}}) - \sum(T_{\text{denominator}})$ is a positive value and is above the real axis if it is a negative value. Thus, for this example the necessary condition is $T_1 > (T_f + T_m + T_2)$.

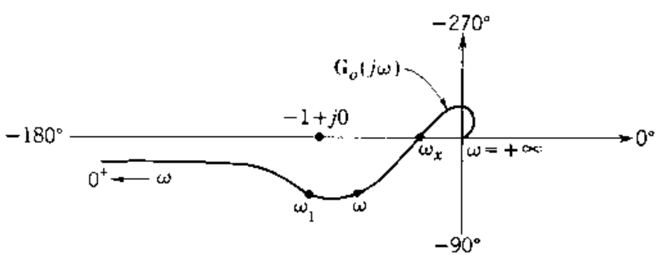


FIGURE 8.19 Polar plot for a typical Type 2 transfer function resulting in a stable system [Eq. (8.46)].

8.13 SUMMARY: DIRECT POLAR PLOTS

To obtain the direct polar plot of a system's forward transfer function, the following criteria are used to determine the key parts of the curve.

Step 1. The forward transfer function has the general form

$$\mathbf{G}(j\omega) = \frac{K_m(1 + j\omega T_a)(1 + j\omega T_b)\cdots(1 + j\omega T_w)}{(j\omega)^m(1 + j\omega T_1)(1 + j\omega T_2)\cdots(1 + j\omega T_u)} \tag{8.47}$$

For this transfer function the system type is equal to the value of m and determines the portion of the polar plot representing the $\lim_{\omega\to 0} \mathbf{G}(j\omega)$. The low-frequency polar plot characteristic (as $\omega \to 0$) of the different system types are summarized in Fig. 8.20. The angle at $\omega = 0$ is $m(-90^\circ)$. The arrow on the polar plots·indicates the direction of increasing frequency.

Step 2. The high-frequency end of the polar plot can be determined as follows:

$$\lim_{\omega\to+\infty} \mathbf{G}(j\omega) = 0 \,\underline{/(w - m - u)90^\circ} \tag{8.48}$$

Note that since the degree of the denominator Eq. (8.47) is always greater than the degree of the numerator, the high-frequency point ($\omega = \infty$) is approached (i.e., the angular condition) in the cw sense. The plot ends at the origin tangent to the axis determined by Eq. (8.48). Tangency may occur on either side of the axis.

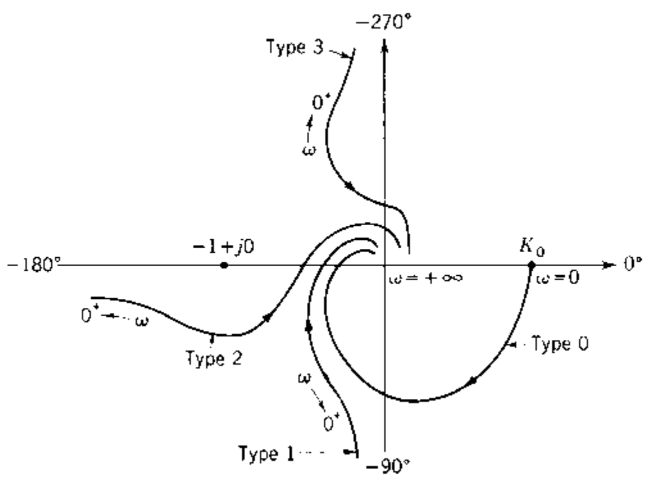


FIGURE 8.20 A summary of direct polar plots of different types of system.

www.4electron.com

Step 3. The asymptote that the low-frequency end approaches, for a Type 1 system, is determined by taking the limit as $\omega \to 0$ of the real part of the transfer function.

Step 4. The frequencies at the points of intersection of the polar plot with the negative real axis and the imaginary axis are determined, respectively, by setting

$$\text{Im } G(j\omega)] = 0 \tag{8.49}$$

$$\text{Re } G(j\omega)] = 0 \tag{8.50}$$

Step 5. If there are no frequency-dependent terms in the numerator of the transfer function, the curve is a smooth one in which the angle of $G(j\omega)$ continuously decreases as ω goes from 0 to ∞. With time constants in the numerator, and depending upon their values, the angle may not continuously vary in the same direction, thus creating "dents" in the polar plot.

Step 6. As is seen later in this chapter, it is important to know the exact shape of the polar plot of $G(j\omega)$ in the vicinity of the $-1 + j0$ point and the crossing point on the negative real axis.

8.14 NYQUIST'S STABILITY CRITERION

A system designer must be sure that the closer-loop system he or she designs is stable. The Nyquist stability criterion [3,4,8,10] provides a simple graphical procedure for determining closed-loop stability from the frequency-response curves of the open-loop transfer function $G(j\omega)H(j\omega)$. The application of this method in terms of the polar plot is covered in this section; appliction in terms of the log magnitude–angle (Nichols) diagram is covered in Sec. 8.17.

For a stable system the roots of the characteristic equation

$$B(s) = 1 + G(s)H(s) = 0 \tag{8.51}$$

cannot be permitted to lie in the RH s plane or on the $j\omega$ axis, as shown in Fig. 8.21. In terms of $G = N_1/D_1$ and $H = N_2/D_2$, Eq. (8.51) becomes

$$B(s) = 1 + \frac{N_1 N_2}{D_1 D_2} = \frac{D_1 D_2 + N_1 N_2}{D_1 D_2} \tag{8.52}$$

Note that the numerator and denominator of B(s) have the same degree, and the *poles of the open-loop transfer function* $G(s)H(s)$ *are the poles of* $B(s)$. The closed-loop transfer function of the system is

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)H(s)} = \frac{N_1 D_2}{D_1 D_2 + N_1 N_2} \tag{8.53}$$

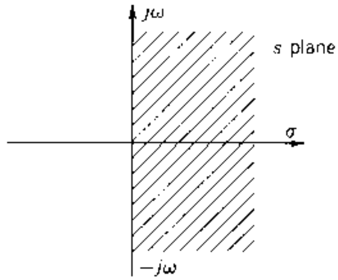


FIGURE 8.21 Prohibited region in the s plane.

The denominator of $C(s)/R(s)$ is the same as the numerator of $B(s)$. The condition for stability may therefore be restated as follows: For a stable system none of the zeros of $B(s)$ can lie in the RH s plane or on the imaginary axis. Nyquist's stability criterion relates the number of zeros and poles of $B(s)$ that lie in the RH s plane to the polar plot of $G(s)H(s)$.

Limitations

In this analysis the assumption is that all the control systems are inherently linear or that their limits of operation are confined to give a linear operation. This yields a set of linear differential equations that describe the dynamic performance of the systems. Because of the physical nature of feedback control systems, the degree of the denominator D_1D_2 is equal to or greater than the degree of the numerator N_1N_2 of the open-loop transfer function $G(s)H(s)$. Mathematically, this statement means that $\lim_{s\to\infty} G(s)H(s) \to 0$ or a constant. These two factors satisfy the necessary limitations to the generalized Nyquist stability criterion.

Mathematical Basis for Nyquist's Stability Criterion

A rigorous mathematical derivation of Nyquist's stability criterion involves complex-variable theory and is available in the literature [11,15]. Fortunately, the result of the deviation is simple and is readily applied. A qualitative approach is now presented for the special case that $B(s)$ is a rational fraction. The characteristic function $B(s)$ given by Eq. (8.52) is rationalized, factored, and then written in the form

$$B(s) = \frac{(s - Z_1)(s - Z_2)\cdots(s - Z_n)}{(s - p_1)(s - p_2)\cdots(s - p_n)} \tag{8.54}$$

where $Z_1, Z_2, \ldots, Z_n$ are the zeros and $p_1, p_2, \ldots, p_n$ are the poles. [Remember that z_i is used to denote a zero of $G(s)H(s)$.] The poles p_i are the same as the

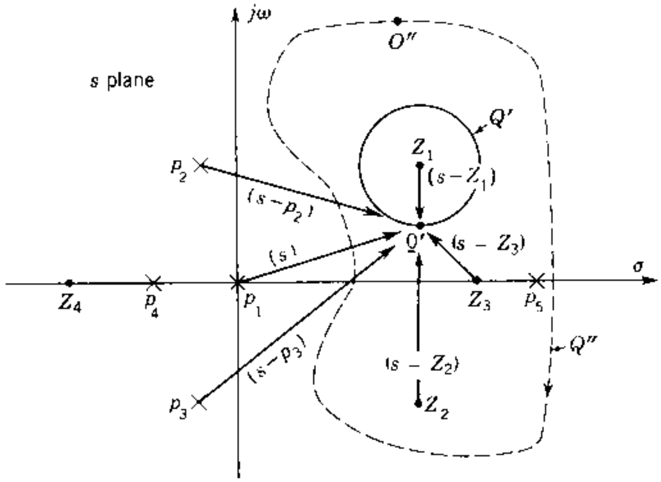


FIGURE 8.22 A plot of some poles and zeros of Eq. (8.54).

poles of the open-loop transfer function $G(s)H(s)$ and include the s term for which $p = 0$, if it is present.

In Fig. 8.22 some of the poles and zeros of a generalized function $B(s)$ are arbitrarily drawn on the s plane. Also, an arbitrary *closed contour* Q' is drawn that encloses the zero Z_1. To the point O' on Q', whose coordinates are $s = \sigma + j\omega$, are drawn directed line segments from all the poles and zeros. The lengths of these directed line segments are given by $s - Z_1, s - Z_2, s - p_1, s - p_2$, etc. Not all the directed segments from the poles and zeros are indicated in the figure, as they are not necessary to proceed with this development. *As the point O' is rotated cw once around the closed contour Q', the line of length $(s - Z_1)$* rotates through a net cw angle of 360°. *All the other directed* segments have rotated through a *net angle of* 0°. Thus, by referring to Eq. (8.54) it is seen that the cw rotation of 360° for the length $(s - Z_1)$ *must simultaneously be realized* by the function $B(s)$ for the enclosure of the zero Z_1 by the path Q'.

Consider now a larger closed contour Q'' that includes the zeros Z_1, Z_2, Z_3, and the pole p_5. *As a point O'' is rotated cw once around the closed curve Q'', each of the directed line segments from the enclosed pole and zeros rotates through a net cw angle of* 360°. Since the angular rotation of the pole is experienced by the characteristic function $B(s)$ in its denominator, the net angular rotation realized by Eq. (8.54) must be equal to the net angular rotations due to the pole p_5 minus the net angular rotations due to the zeros Z_1, Z_2, and Z_3. In other words, the net angular rotation experienced by $1 + G(s)H(s)$ is $360^\circ - (3)(360^\circ) = -720^\circ$. Therefore, for this case, the net number of rotations N experienced by $B(s) = 1 + G(s)H(s)$ for the cw movement of point

O' *once* about the closed contour Q' is equal to –2; that is

$$N = (\text{number of poles enclosed}) - (\text{number of zeros enclosed})$$
$$= 1 - 3 = -2$$

Where the minus sign denotes cw rotation. Note that if the contour Q' includes only the pole p_5, $B(s)$ experiences one ccw rotation as the point O'' is moved cw around the contour. Also, for *any closed path* that may be chosen, all the poles and zeros that lie outside the path each contribute a net angular rotation of 0° to $B(s)$ as a point is moved once around this contour.

Generalizing Nyquist's Stability Criterion

Consider now a closed contour Q such that the whole RH s plane is encircled (see Fig. 8.23), thus enclosing all zeros and poles of $B(s)$ that have positive real parts. The theory of complex variables used in the rigorous derivation requires that the contour Q must not pass *through any poles or zeros of* $B(s)$. When the results of the preceding discussion are applied to the contour Q, the following properties are noted:

1. The total number of cw rotations of $B(s)$ due to its zeros is equal to the total number of zeros Z_R in the RH s plane.
2. The total number of ccw rotations of $B(s)$ due to its poles is equal to the total number of poles P_R in the RH s plane.
3. The *net* number of rotations N of $B(s) = 1 + G(s)H(s)$ about the origin is equal to its total number of poles P_R minus its total number of zeros Z_R in the RH s plane. N may be positive (ccw), negative (cw), or zero.

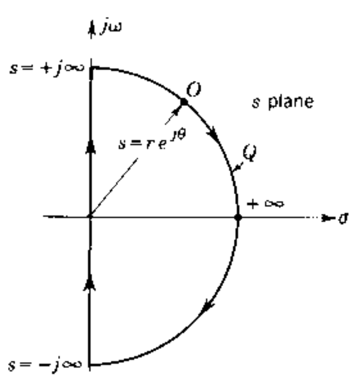


FIGURE 8.23 The contour that encloses the entire right-half s plane.

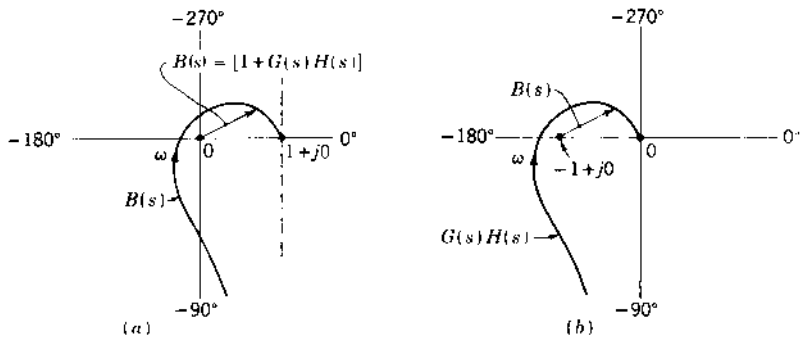


FIGURE 8.24 A change of reference for B(s).

The essence of these three conclusions can be represented by the equation

$$N = \frac{\text{change in phase of } [1+G(s)H(s)]}{2\pi} = P_R - Z_R \tag{8.55}$$

where ccw rotation is defined as being positive and cw rotation is negative. In order for the characteristic function $B(s)$ to realize a net rotation N, the directed line segment representing $B(s)$ (see Fig. 8.24a) must rotate about the origin $360N$ degrees or N complete revolutions. Solving for Z_R in Eq. (8.55) yields

$$Z_R = P_R - N \tag{8.56}$$

Since $B(s)$ can have no zeros Z_R in the RH s plane for a stable system, it is therefore concluded that, *for a stable system, the net number of rotations of $B(s)$ about the origin must be ccw and equal to the number of poles P_R that lie in the RH s plane*. In other words, if $B(s)$ experiences a net cw rotation, this indicates that $Z_R > P_R$, where $P_R \geq 0$, and thus the closed-loop system is unstable. If there are zero net rotations, then $Z_R = P_R$ and the system may or may not be stable, according to whether $P_R = 0$ or $P_R > 0$.

Obtaining a Plot of $B(s)$

Figure 8.24*a* and *b* shows a plot of $B(s)$ and a plot of $G(s)H(s)$, respectively. By moving the origin of Fig. 8.24*b* to the $-1 + j0$ point, the curve is now equal to $1 + G(s)H(s)$, which is $B(s)$. Since $G(s)H(s)$ is known, this function is plotted, and then the origin is moved to the -1 point to obtain $B(s)$.

In general, the open-loop transfer functions of many physical systems do not have any poles P_R in the RH s plane. In this case, $Z_R = N$. *Thus, for a stable system the net number of rotations about the $-1 + j0$ point must be zero when there are no poles of $G(s)H(s)$ in the RH s plane.*

In the event that the function $G(s)H(s)$ has some poles in the RH s plane and the denominator is not in factored form, the number P_R can be determined by applying Routh's criterion to D_1D_2. The Routhian array gives the number of roots in the RH s plane by the number of sign changes in the first column.

Analysis of Path Q

In applying Nyquist's criterion, the whole RH s plane must be encircled to ensure the inclusion of all poles or zeros of $B(s)$ in this portion of the plane. In Fig. 8.23 the entire RH s plane is enclosed by the closed path Q, which is composed of the following two segments:

1. One segment is the imaginary axis from $-j\infty$ to $+j\infty$.
2. The other segment is a semicircle of infinite radius that encircles the entire RH s plane.

The portion of the path along the imaginary axis is represented mathematically by $s = j\omega$. Thus, replacing s by $j\omega$ in Eq. (8.54), followed by letting ω take on all values from $-\infty$ to $+\infty$, gives the portion of the $B(s)$ plot corresponding to that portion of the closed contour Q that lies on the imaginary axis.

One of the requirements of the Nyquist criterion is that $\lim_{s\to\infty} G(s)H(s) \to 0$ or a constant. Thus $\lim_{s\to\infty} B(s) = \lim_{s\to\infty}[1 + G(s)\times H(s)] \to 1$ or 1 plus the constant. As a consequence, as the point O moves along the segment of the closed contour represented by the semicircle of infinite radius, the corresponding portion of the $B(s)$ plot is a fixed point. As a result, the movement of point O along only the imaginary axis from $-j\infty$ to $+j\infty$ results in the same net rotation of $B(s)$ as if the whole contour Q were considered. *In other words, all the rotation of $B(s)$ occurs while the point O goes from $-j\infty$ to $+j\infty$ along the imaginary axis.* More generally, this statement applies only to those transfer functions $G(s)H(s)$ that conform to the limitations stated earlier in this section.*

Effect of Poles at the Origin on the Rotation of $B(s)$

Some transfer functions $G(s)H(s)$ have an s^m in the denominator. Since no poles or zeros can lie on the contour Q, the contour shown in Fig. 8.23 must be modified. For these cases the manner in which the $\omega = 0^-$ and $\omega = 0^+$ portions of the plot are joined is now investigated. This can best be done by taking an example. Consider the transfer function with positive

---

*A transfer function that does not conform to these limitations and to which the immediately preceding italicized statement does not apply is $G(s)H(s) = s$.

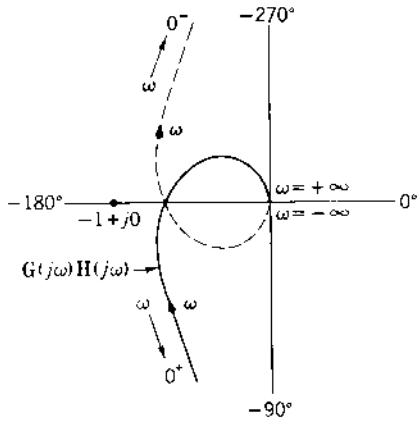


FIGURE 8.25 A plot of the transfer function $G(j\omega)H(j\omega)$ for Eq.(8.57).

values of T_1 and T_2:

$$G(s)H(s) = \frac{K_1}{s(1 + T_1 s)(1 + T_2 s)} \tag{8.57}$$

The direct polar plot of $\mathbf{G}(j\omega)\mathbf{H}(j\omega)$ of this function is obtained by substituting $s = j\omega$ into Eq. (8.57). Figure 8.25 represents the plot of $G(s)H(s)$ for values of s along the imaginary axis; that is, $j0^+ < j\omega < j\infty$ and $-j\infty < j\omega < j0^-$. The plot is drawn for both positive and negative frequency values. *The polar plot drawn for negative frequencies is the conjugate of the plot drawn for positive frequencies.* That is, the curve for negative frequencies is symmetrical to the curve for positive frequencies, with the real axis as the axis of symmetry.

The closed contour Q of Fig. 8.23, in the vicinity of $s = 0$, is modified as shown in Fig. 8.26a to avoid passing through the origin. In other words, the point O is moved along the negative imaginary axis from $s = -j\infty$ to a point where $s = -j\varepsilon = 0^- \angle -\pi/2$ becomes very small. Then the point O moves along a semicircular path of radius $s = \varepsilon e^{j\theta}$ in the RH s plane with a very small radius ε until it reaches the positive imaginary axis at $s = +j\varepsilon = j0^+ = 0^+ \angle \pi/2$. From here the point O proceeds along the positive imaginary axis to $s = +j\infty$. Then, letting the radius approach zero, $\varepsilon \to 0$, for the semicircle around the origin ensures the inclusion of all poles and zeros in the RH s plane. To complete the plot of $B(s)$ in Fig. 8.25, it is necessary to investigate the effect of moving point O on this semicircle around the origin.

For the semicircular portion of the path Q represented by $s = \varepsilon e^{j\theta}$, where $\varepsilon \to 0$ and $-\pi/2 \le \theta \le \pi/2$, Eq. (8.57) becomes

$$G(s)H(s) = \frac{K_1}{s} = \frac{K_1}{\varepsilon e^{j\theta}} = \frac{K_1}{\varepsilon} e^{-j\theta} = \frac{K_1}{\varepsilon} e^{j\psi} \tag{8.58}$$

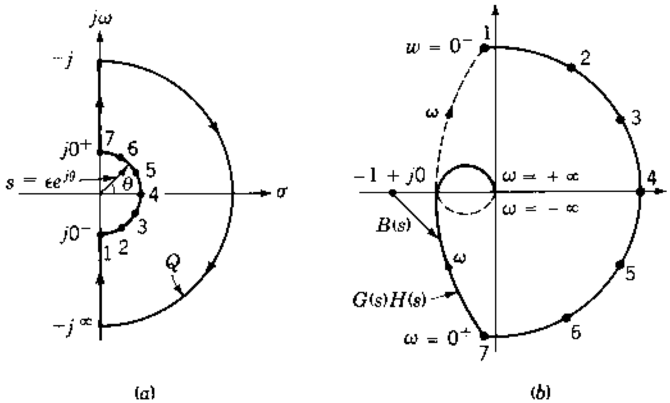


FIGURE 8.26 (*a*) The contour Q that encircles the right-half *s* plane, (*b*) complete plot for Eq.(8.57).

where $K_1/\varepsilon \to \infty$ as $\varepsilon \to \infty$, and $\psi = -\theta$ goes from $\pi/2$ to $-\pi/2$ as the directed segment *s* goes ccw from $\varepsilon \underline{/-\pi/2}$ to $\varepsilon \underline{/+\pi/2}$. Thus, in Fig. 8.25, the endpoints from $\omega \to 0^-$ and $\omega \to 0^+$ are joined by a semicircle of infinite radius in the first and fourth quadrants. Figure 8.25 illustrates this procedure and shows the completed contour of $G(s)H(s)$ as the point O moves along the modified contour Q in the *s* plane from point 1 to point 7. When the origin is moved to the $-1 + j0$ point, the curve becomes $B(s)$, as shown in Fig.8.26b.

The plot of $B(s)$ in Fig.8.26b does not encircle the $-1 + j0$ point; therefore, the encirclement N is zero. From Eq. (8.57) there are no poles within Q; that is, $P_R = 0$. Thus, when Eq. (8.56) is applied, $Z_R = 0$ and the closed-loop system is stable.

Transfer functions that have the term s^m in the denominator have the general form, as $\varepsilon \to 0$,

$$G(s)H(s) = \frac{K_m}{s^m} = \frac{K_m}{(\varepsilon^m)e^{jm\theta}} = \frac{K_m}{\varepsilon^m} e^{-jm\theta} = \frac{K_m}{\varepsilon^m} e^{jm\psi} \tag{8.59}$$

where $m = 1, 2, 3, 4, \ldots$ With the reasoning used in the preceding example, it is seen from Eq.(8.59) that, as *s* moves from 0^- to 0^+, the plot of $G(s)H(s)$ traces m cw semicircles of infinite radius about the origin. If $m = 2$, then, as θ goes from $-\pi/2$ to $\pi/2$ in the *s* plane with radius ε, $G(s)H(s)$ experiences a net rotation of $(2)(180^\circ) = 360^\circ$.

Since the polar plots are symmetrical about the real axis, it is only necessary to determine the shape of the plot of $G(s)H(s)$ for a range of values of $0 < \omega < +\infty$. The net rotation of the plot for the range of $-\infty < \omega < +\infty$ is twice that of the plot for the range of $0 < \omega < +\infty$.

When G($j\omega$)H($j\omega$) Passes Through the Point $-1+j0$

When the curve of G($j\omega$)H($j\omega$) passes through the $-1+j0$ point, the number of encirclements N is indeterminate. This situation corresponds to the condition where $B(s)$ has zeros on the imaginary axis. A necessary condition for applying the Nyquist criterion is that the path encircling the specified area must not pass through any poles or zeros of $B(s)$. When this condition is violated, the value for N becomes indeterminate and the Nyquist stability criterion cannot be applied. Simple imaginary zeros of $B(s)$ mean that the closed-loop system will have a continuous steady-state sinusoidal component in its output that is independent of the form of the input. Unless otherwise stated, this condition is considered unstable.

8.15 EXAMPLES OF NYQUIST'S CRITERION USING DIRECT POLAR PLOT

Several polar plots are illustrated in this section. These plots can be obtained with the aid of pole-zero diagrams, from calculations for a few key frequencies, or from a CAD program. The angular variation of each term of a $G(s)H(s)$ function, as the contour Q is traversed, is readily determined from its pole-zero diagram. Both stable and unstable open-loop transfer functions $G(s)H(s)$ are illustrated in the following examples for which all time constants T_i are positive.

Example 1: Type 0. The direct polar plot is shown in Fig. 8.27 for the following Type 0 open-loop transfer function:

$$G(s)H(s) = \frac{K_0}{(1 + T_1 s)(1 + T_2 s)} \tag{8.60}$$

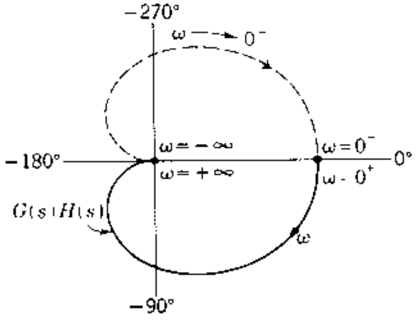


FIGURE 8.27 Plot of a typical Type 0 transfer function G(s)H(s) for $-\infty < \omega < \infty$.

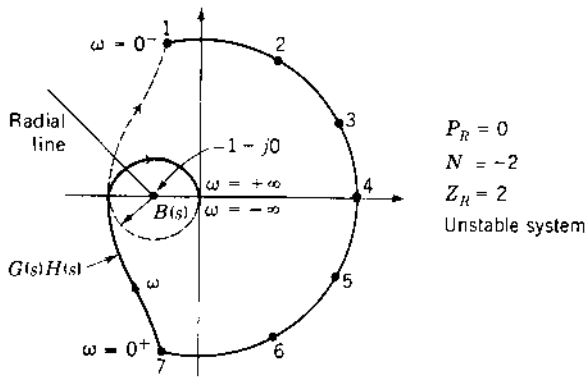


FIGURE 8.28 Polar plot of Fig. 8.26b with increased gain.

From this plot it is seen that $N = 0$, and Eq. (8.60) yields the value $P_R = 0$. Therefore, the value of Z_R is $Z_R = P_R - N = 0$ and the closed-loop system is stable. Thus, the Nyquist stability criterion indicates that, no matter how much the gain K_o is increased, this system is always stable.

Example 2: Type 1. The example in Fig. 8.26 for the transfer function of the form

$$G(s)H(s) = \frac{K_1}{s(1 + T_1 s)(1 + T_2 s)} \tag{8.61}$$

is shown in the preceding section to be stable. However, if the gain K_1 is increased, the system is made unstable, as seen from Fig. 8.28. Note that the rotation of $B(s)$, in the direction of increasing frequency, produces a net angular rotation of -720° or two complete cw rotations ($N = -2$), so that $Z_R = P_R - N = 2$. Thus, the closed-loop system is unstable.

The number of rotations N can be determined by drawing a radial line from the $-1 + j0$ point. Then, at each intersection of this radial line with the curve of $G(s)H(s)$, the direction of increasing frequency is noted. If the frequency along the curve of $G(s)H(s)$ is increasing as the curve crosses the radial line in a ccw direction, this crossing is positive. Similarly, if the frequency along the curve of $G(s)H(s)$ is increasing as the curve crosses the radial line in a cw direction, this crossing is negative. The sum of the crossings, including the sign, is equal to N. In Fig. 8.28 both crossings of the radial line are cw and negative. Thus, the sum of the crossings is $N = -2$.

Example 3: Type 2. Consider the transfer function

$$G(s)H(s) = \frac{K_2(1 + T_4 s)}{s^2(1 + T_1 s)(1 + T_2 s)(1 + T_3 s)} \tag{8.62}$$

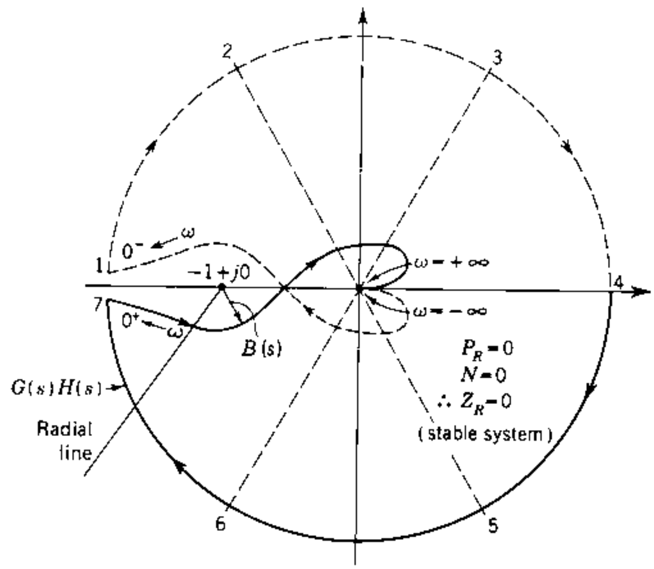


FIGURE 8.29 The complete polar plot of Eq.(8.62).

where $T_4 > T_1 + T_2 + T_3$. Figure 8.29 shows the mapping of $G(s)H(s)$ for the contour Q of the s plane. The word *mapping*, as used here, means that for a given point in the s plane there corresponds a given value of $G(s)H(s)$ or $B(s)$.

As pointed out in the preceding section, the presence of the s^2 term in the denominator of Eq. (8.62) results in a net rotation of 360° in the vicinity of $\omega = 0$, as shown in Fig. 8.29. For the complete contour of Fig. 8.26a the net rotation is zero; thus, since $P_R = 0$, the system is stable. The value of N can be determined by drawing the line radiating from the $-1 + j0$ point (see Fig. 8.29) and noting one cw and one ccw crossing; thus, $N = 0$. Like the previous example, this system can be made unstable by increasing the gain sufficiently for the $G(s)H(s)$ plot to cross the negative real axis to the left of the $-1 + j0$ point.

Example 4: Conditionally Stable System. In this instance,

$$G(s)H(s) = \frac{K_0(1 + T_1 s)^2}{(1 + T_2 s)(1 + T_3 s)(1 + T_4 s)(1 + T_5 s)^2} \tag{8.63}$$

where $T_5 < T_1 < T_2$, T_3, and T_4. The complete polar plot of Eq. (8.63) is illustrated in Fig. 8.30 for a particular value of gain. For the radial line drawn from the $-1 + j0$ point the number of rotations is readily determined as $N = 0$.

It is seen in this example that the system can be made unstable not only by increasing the gain but also by decreasing the gain. If the gain is increased

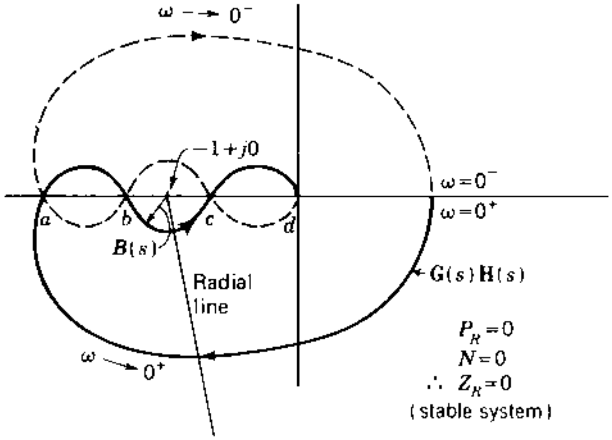


FIGURE 8.30 The complete polar plot of Eq. (8.63).

sufficiently for the $-1 + j0$ point to lie between the points c and d of the polar plot, the net cw rotation is equal to 2. Therefore, $Z_R = 2$ and the system is unstable. On the other hand, if the gain is decreased so that the $-1 + j0$ point lies between the points a and b of the polar plot, the net cw rotation is again 2 and the system is unstable. The gain can be further decreased so that the $-1 + j0$ point lies to the left of point a of the polar plot, resulting in a stable system. *This system is therefore conditionally stable.* A conditionally stable system is stable for a given range of values of gain but becomes unstable if the gain is either reduced or increased sufficiently. Such a system places a greater restriction on the stability and drift of amplifier gain and system characteristics. In addition, an effective gain reduction occurs in amplifiers that reach saturation with large input signals. This, in turn, may result in an unstable operation for the conditionally stable system.

Example 5: Open-Loop Unstable. For this case

$$G(s)H(s) = \frac{K_1(T_2 s + 1)}{s(T_1 s - 1)} \tag{8.64}$$

The complete polar plot for this equation is illustrated in Fig. 8.31 for a particular value of gain K_1. There is one cw intersection with the radial line drawn from the $-1 + j0$ point, and thus $N = -1$. Therefore, since $P_R = 1$, $Z_R = 2$, and the system is unstable. In this example the system is unstable for low values of gain $(0 < K_1 < K_{1x})$ for which the real-axis crossing b is to the right of the $-1 + j0$ point. For the range $0 < K_1 < K_{1x}$ the $-1 + j0$ point is located, as shown in Fig. 8.31, between the points a and b. However, for the range $K_{1x} < K_1 < \infty$ the $-1 + j0$ point lies between the points b and c, which

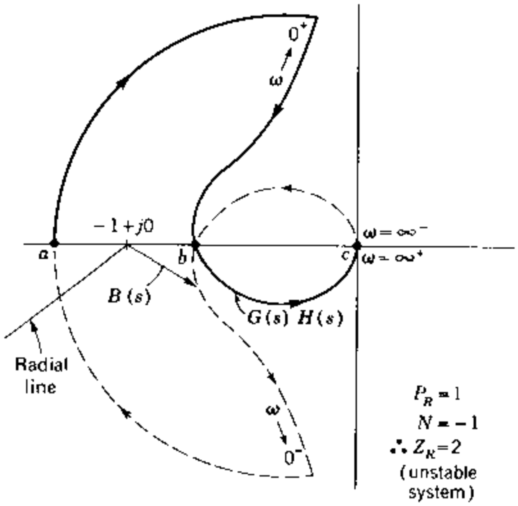


FIGURE 8.31 The complete polar plot of the open-loop unstable transfer function $G(s)H(s) = K_1(T_2s + 1)/s(T_1s - 1)$.

yields $N = +1$, thus resulting in $Z_R = P_R - N = 1 - N = 0$ and a stable closed-loop system.

8.16 NYQUIST'S STABILITY CRITERION APPLIED TO SYSTEM HAVING DEAD TIME

Transport lag (dead time), as described in Sec. 7.14, is represented by the transfer function $G_\tau(s) = e^{-\tau s}$. The frequency transfer function is

$$G_\tau(j\omega) = e^{-j\omega\tau} = 1\,\angle{-\omega\tau}\,. \tag{8.65}$$

It has a magnitude of unity and a negative angle whose magnitude increases directly in proportion to frequency. The polar plot of Eq. (8.65) is a unit circle that is traced indefinitely, as shown in Fig. 8.32. The Lm and phase-angle diagram shows a constant value of 0 dB and a phase angle that decreases with frequency.

An example can best illustrate the application of the Nyquist criterion to a control system having transport lag. Figure 8.33a illustrates the complete polar plot of a stable system having a specified value of gain and the transfer function

$$G_x(s)H(s) = \frac{K_1}{s(1 + T_1s)(1 + T_2s)} \tag{8.66}$$

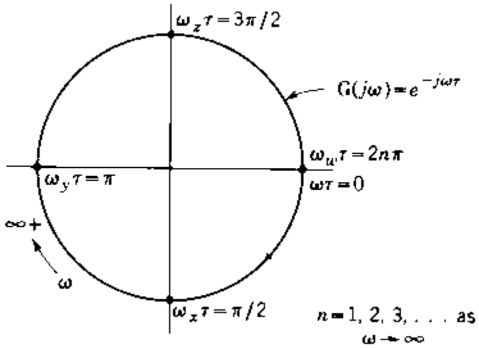


FIGURE 8.32 Polar plot characteristic for transport lag.

If dead time is added to this system, its transfer function becomes

$$G(s)H(s) = \frac{K_1 e^{-\tau s}}{s(1 + T_1 s)(1 + T_2 s)} \tag{8.67}$$

and the resulting complete polar plot is shown in Fig. 8.33*b*. When the contour Q is traversed and the polar plot characteristic of dead time, shown in Fig. 8.32, is included, the effects on the complete polar plot are as follows:

1. In traversing the imaginary axis of the contour Q between $0^+ < \omega < +\infty$, the polar plot of $\mathbf{G}(j\omega)\mathbf{H}(j\omega)$ in the third quadrant is shifted clockwise, closer to the $-1 + j0$ point. Thus, if the dead time is increased sufficiently, the $-1 + j0$ point will be enclosed by the polar plot and the system becomes unstable.
2. As $\omega \to +\infty$, the magnitude of the angle contributed by the transport lag increases indefinitely. This yields a spiraling curve as $|\mathbf{G}(j\omega)\mathbf{H}(j\omega)| \to 0$.

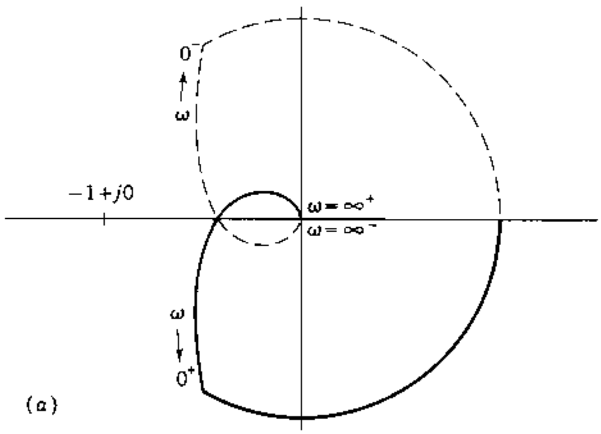
A transport lag therefore tends to make a system less stable.

8.17 DEFINITIONS OF PHASE MARGIN AND GAIN MARGIN AND THEIR RELATION TO STABILITY [16]

The stability and approximate degree of stability can be determined from the Lm and phase diagram. The stability characteristic is specified in terms of the following quantities:

Gain crossover. This is the point on the plot of the transfer function at which the magnitude of $\mathbf{G}(j\omega)$ is unity [Lm $\mathbf{G}(j\omega) = 0$ dB]. The frequency at gain crossover is called the phase-margin frequency ω_ϕ.

Phase margin angle. This is 180° plus the negative trigonometrically considered angle of the transfer function at the gain-crossover point. It is



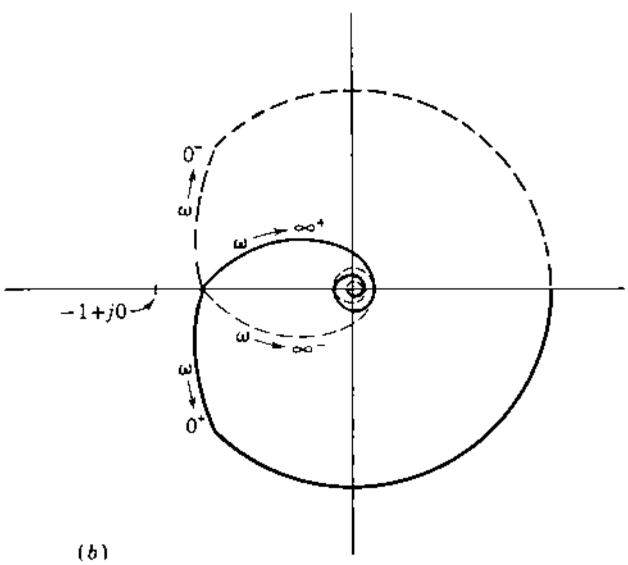


FIGURE 8.33 Complete polar plots of (a) Eq. (8.66) and (b) Eq. (8.67).

designated as the angle γ, which can be expressed as $\gamma = 180^\circ + \phi$, where $\angle G(j\omega_\phi) = \phi$ is negative.

Phase crossover. This is the point on the plot of the transfer function at which the phase angle is -180°. The frequency at which phase crossover occurs is called the gain-margin frequency ω_c.

Gain margin. The gain margin is the factor a by which the gain must be changed in order to produce instability. Expressed in terms of the transfer function at the frequency ω_c, it is

$$\left| G(j\omega_c)a = 1 \right| \tag{8.68}$$

On the polar plot of $G(j\omega)$ the value at ω_c is

$$\left| G(j\omega_c) \right| = \frac{1}{a} \tag{8.69}$$

In terms of the Lm, in decibels, this is

$$\text{Lm}\, a = -\text{Lm}\, G(j\omega_c) \tag{8.70}$$

which identifies the gain margin on the Lm diagram.

These quantities are illustrated in Fig. 8.34 on both the log and the polar curves. Note the algebraic sign associated with these two quantities as marked on the curves. Figure 8.34 *a* and *b* represents a stable system, and Fig. 8.34 *c* and *d* represents an unstable system.

The phase margin angle is the amount of phase shift at the frequency ω_ϕ that would just produce instability. This angle would make the polar plot go through the –1 point. The phase margin angle for minimum-phase (m.p.) systems must be positive for a stable system, whereas a negative phase margin angle means that the system is unstable.

It can be shown that the phase margin angle is related to the effective damping ratio ζ of the system. This topic is discussed qualitatively in the next chapter. Satisfactory response is usually obtained with a phase margin of 45° to 60°. As an individual gains experience and develops his or her own particular technique, a desirable value of γ to be used for a particular system becomes more evident. This guideline for system performance applies only to those systems where behaviour is that of an equivalent second-order system. The gain margin must be positive when expressed in decibels (greater than unity as a numeric) for a stable system. A negative gain margin means that the system is unstable. The damping ratio ζ of the system is also related to the gain margin. However, the phase margin gives a better estimate of damping ratio, and therefore of the transient overshoot of the system, than the gain margin.

The phase-margin frequency ω_ϕ, phase margin angle γ, crossover frequency ω_c, and the gain margin Lm a are also readily identified on the Nichols plot as shown in Fig. 8.35 and described in Sec. 8.19. Further information about the speed of response of the system can be obtained from the maximum value of the control ratio and the frequency at which this maximum occurs. The Lm–angle diagram or a CAD program can be used to obtain this data (see Chap. 9).

The relationship of stability and gain margin is modified for a conditionally stable system. This can be seen by reference to the polar plot of Fig. 8.29.

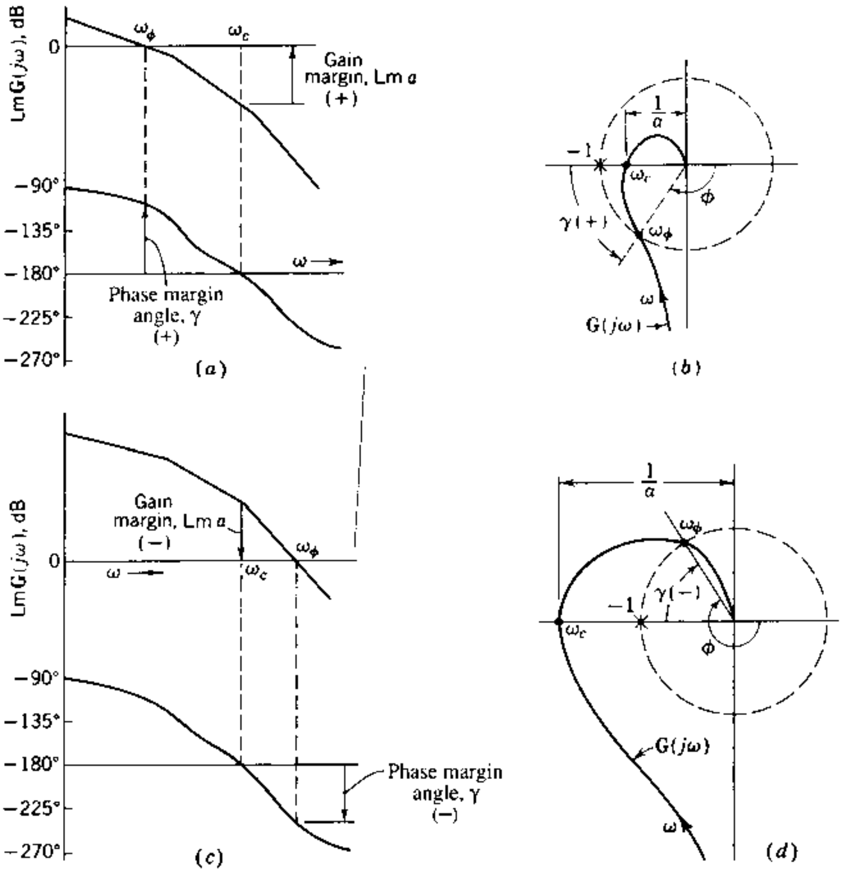


FIGURE 8.34 Log magnitude and phase diagram and polar plots of G($j\omega$), showing gain margin and phase margin angle: (a) and (b) stable; (c) and (d) unstable.

The system for the particular gain represented in Fig. 8.29 is stable. Instability can occur with an increase or a decrease in gain. Therefore, both "upper" and "lower" gain margins must be identified, corresponding to the upper crossover frequency ω_{cu} and the lower crossover frequency ω_{cl}. In Fig. 8.30 the upper gain margin is larger than unity and lower gain margin is less than unity.

8.18 STABILITY CHARACTERISTICS OF THE LOG MAGNITUDE AND PHASE DIAGRAM

The earlier sections of this chapter discuss how the asymptotes of the Lm vs. ω curve are related to each factor of the transfer function. For example, factors of the form $(1 + j\omega)^{-1}$ have a negative slope equal to –20 dB/decade at high frequencies. Also, the angle for this factor varies from 0° at low frequencies to -90° at high frequencies.

The total phase angle of the transfer function at any frequency is closely related to the slope of the Lm curve at that frequency. A slope of -20 dB/decade is related to an angle of -90°; a slope of –40 dB/decade is related to an angle of -180°; a slope of –60 dB/decade is related to an angle of -270°; etc. Changes of slope at higher and lower corner frequencies, around the particular frequency being considered, contribute to the total angle at that frequency. The farther away the changes of slope (corner frequency) are from the particular frequency, the less they contribute to the total angle at that frequency.

By observing the asymptotes of the Lm curve, it is possible to estimate the approximate value of the angle. With reference to Fig. 8.6 for the example of Sec. 8.7, the angle at $\omega = 4$ is now investigated. The slope of the curve at $\omega = 4$ is –20 dB/decade; therefore, the angle is near -90°. The slope changes at the corner frequencies $\omega = 2$ and $\omega = 8$, and so the slopes beyond these frequencies contribute to the total angle at $\omega = 4$. The actual angle, as read from Fig. 8.7, is -122°. The farther away the corner frequencies occur, the closer the angle is to -90°.

The stability of a minimum phase (m.p.) system requires that the phase margin angle be positive. For this to be true, the angle at the gain crossover [Lm $\mathbf{G}(j\omega) = 0$ dB] must be greater than -180°. This requirement places a limit on the slope of the Lm curve at the gain crossover. *The slope at the gain crossover should be more positive than –40* dB/*decade if the adjacent corner frequencies are not close.* A slope of –20 dB/decade is preferable. This value is derived from the consideration of a theorem by Bode. However, it can be seen qualitatively from the association of the slope the Lm curve with the value of the phase angle. This guide should be used to assist in system design.

The Lm and phase diagram reveals some pertinent information, just as the polar plots do. For example, the gain can be adjusted (which raises or lowers the Lm curve) to produce a phase margin angle in the desirable range of 45° to 60°. The shape of the low-frequency portion of the curve determines system type and therefore the degree of steady-state accuracy. The system type and the gain determine the error coefficients and therefore the steady-state error. The phase-margin frequency ω_ϕ gives a qualitative indication of the speed of response of a system. However, this is only a qualitative relationship. A more detailed analysis of this relationship is made in the next chapter.

8.19 STABILITY FROM THE NICHOLS PLOT (LOG MAGNITUDE—ANGLE DIAGRAM)

The Lm–angle diagram is drawn by picking for each frequency the values of Lm and angle from the Lm and phase diagram. The resultant curve has

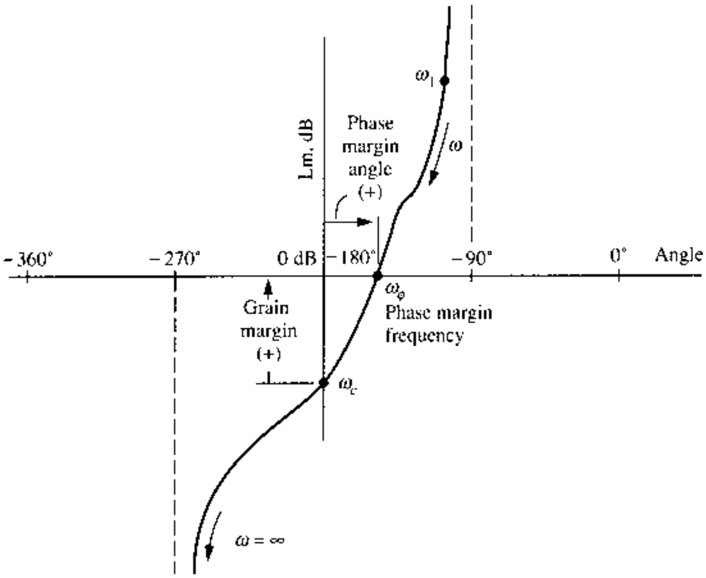


FIGURE 8.35 Example of a log magnitude–angle diagram for a minimum-phase open-loop transfer function.

frequency as a parameter. The curve for the example of Sec. 8.7, sketched in Fig. 8.35, shows a positive gain margin and phase margin angle; therefore, this represents a stable system. Changing the gain raises or lowers the curve without changing the angle characteristics. Increasing the gain raises the curve, thereby decreasing the gain margin and phase margin angle, with the result that the stability is decreased. Increasing the gain so that the curve has a positive Lm at -180° results in negative gain and phase margins; therefore, an unstable system results. Decreasing the gain lowers the curve and increases stability. However, a large gain is desired to reduce steady-state errors, as shown in Chap. 6.

The Lm–angle diagram for $G(s)H(s)$ can be drawn for all values of s on the contour Q of Fig. 8.26a. The resultant curve for minimum phase systems is a closed contour. Nyquist's criterion can be applied to this contour by determining the number of points (having the values 0 dB and odd multiples of 180°) enclosed by the curve of $G(s)H(s)$. This number is the value of N that is used in the equation $Z_R = N - P_R$ to determine the value of Z_R. As an example, consider a control system whose transfer function is given by

$$G(s) = \frac{K_1}{s(1 + T_s)} \tag{8.71}$$

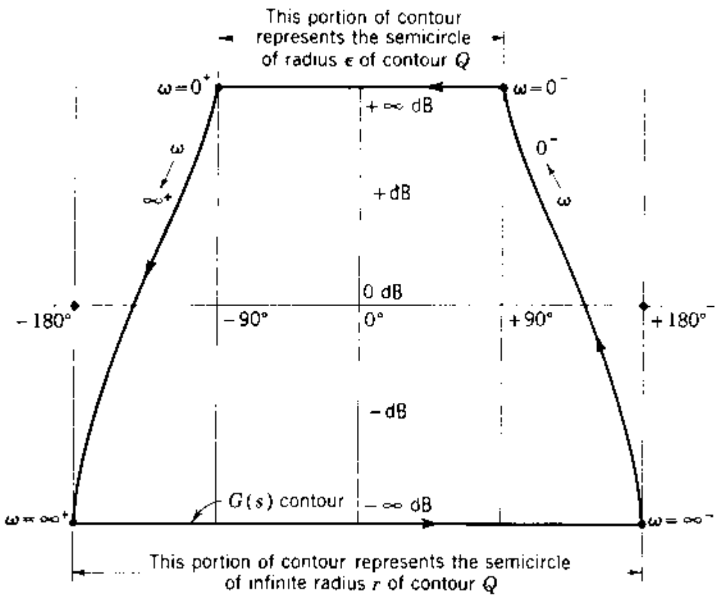


FIGURE 8.36 The log magnitude–angle contour for the minimum-phase system of Eq. (8.71).

Its Lm–angle diagram, for the contour Q, is shown in Fig. 8.36. From this figure it is seen that the value of N is zero and the system is stable. The Lm–angle contour for a nonminimum-phase system does not close; thus, it is more difficult to determine the value of N. For these cases the polar plot is easier to use to determine stability.

In the case of m.p. systems, it is not necessary to obtain the complete Lm–angle contour to determine stability. Only that portion of the contour is drawn representing $G(j\omega)$ for the range of values $0^+ < \omega < \infty$. The stability is then determined from the position of the curve of $G(j\omega)$ relative to the (0 dB, $-180°$) point. In other words, the curve is traced in the direction of increasing frequency, i.e., "walking" along the curve in the direction of increasing frequency. The system is stable if the (0 dB, $-180°$) point is to the right of the curve. This simplified rule of thumb is based on Nyquist's stability criterion for an m.p. system.

A conditionally stable system (as defined in Sec. 8.15, Example 4) is one in which the curve crosses the $-180°$ axis at more than one point. Figure 8.37 shows the transfer-function plot for such a system with two stable and two unstable regions. The gain determines whether the system is stable or unstable. The stable system represented by Fig. 8.37c has two values of gain

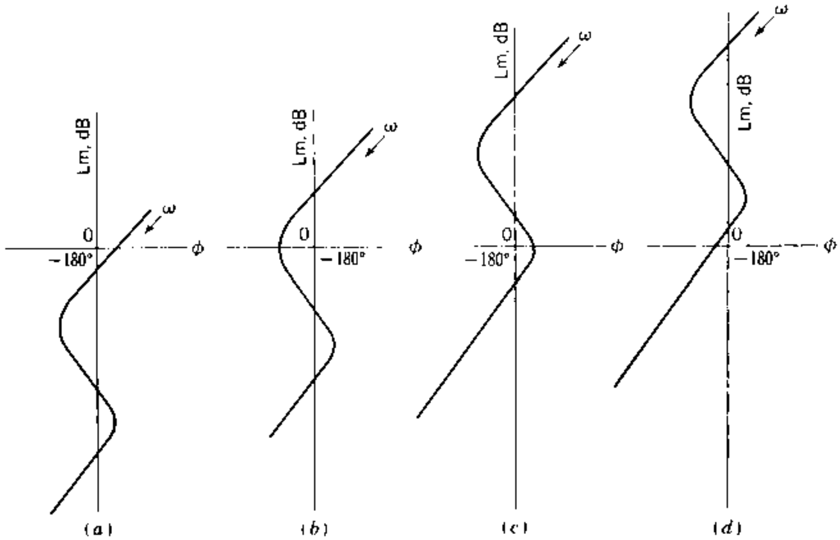


FIGURE 8.37 Log magnitude–angle diagram for a conditionally stable system: (*a*), (*c*) stable; (*b*), (*d*) unstable.

margin, since instability can occur with either an increase or a decrease in gain. Corresponding to the upper and lower crossover frequencies ω_{cu} and ω_{cl}, the gain margin Lm a_u and Lm a_l are positive and negative, respectively. Additional stability information can be obtained from the Lm–angle diagram (see Chapter 9).

8.20 SUMMARY

In this chapter different types of frequency-response plots are introduced. From the Lm and phase-angle diagrams the polar plot can be obtained rapidly and with ease. All of these plots indicate the type of system that is represented and the necessary gain adjustments that must be made to improve the response. How these adjustments are made is discussed in the following chapter. The methods presented for obtaining the Lm frequency-response plots stress graphical techniques. For greater accuracy a CAD program can be used to calculate these data (see Appendixes C and D).

The methods described in this chapter for obtaining frequency-response plots are based upon the condition that the transfer function of a given system is known. These plots can also be obtained from experimental data, which do not require the analytical expression for the transfer function of a system to be known. These experimental data can be used to synthesize an analytical expression for the transfer function using the Lm plot (see Sec. 8.10).

This chapter shows that the polar plot of the transfer function *G(s)H(s)*, in conjunction with Nyquist's stability criterion, gives a rapid means of determining whether a system is stable or unstable. The same information is obtained from the Lm and phase-angle diagrams and the Lm–angle diagram (Nichols plot); the phase margin angle and gain margin are also used as a means of indicating the degree of stability. In the next chapter it is shown that other key information related to the time domain is readily obtained from any of these plots. Thus, the designer can determine whether the given system is satisfactory.

Another useful application of the Nyquist criterion is the analysis of systems having the characteristic of dead time. The Nyquist criterion has one disadvantage in that open-loop poles or zeros on the $j\omega$ axis require special treatment. This treatment is not expounded since, in general most problems are free of poles or zerox on the $j\omega$ axis.

REFERENCES

1. Maccoll, L. A.: *Fundamental Theory of Servomechanisms*, Van Nostrand, Princeton, N.J., 1945.
2. Houpis, C. H., and S. J. Rasmussen: *Quantitative Feedback Theory Fundamentals and Applications*, Marcel Dekker, New York, 1999.
3. James, H. M., N. B. Nichols and R. S. Phillips: *Theory of Servomechanisms*, McGraw-Hill, New York, 1947.
4. Bode, H. W.: *Network Analysis and Feedback Amplifier Design*, Van Nostrand, Princeton, N.J., 1945, chap 8.
5. Sanathanan, C. K., and H. Tsukui: "Synthesis of Transfer Function from Frequency Response Date", *Intern. J. Systems Science*, Vol. 5, no. 1, pp. 41–54, January 1974.
6. D'Azzo, J. J., and C. H. Houpis: *Linear Control System Analysis and Design*: *Conventional and Modern*, 3rd ed., McGraw-Hill, New York, 1988.
7. D'Azzo, J. J. and C. H. Houpis: *Feedback Control System Analysis and Synthesis*, 2nd ed., McGraw-Hill, New York, 1966.
8. Nyquist, H.: "Regeneration Theory", *Bell Syst. Tech. J.*, Vol. 11, pp. 126–146, 1932.
9. Bruns R. A and R. M. Saunders: *Analysis of Feedback Control Systems*, McGraw-Hill, New York, 1955, chap. 14.
10. Balabanian, N., and W. R. LePage: "What Is a Minimum-Phase Network?', *Trans. AIEE*, Vol. 74, pt. II, pp. 785–788, January 1956.
11. Freudenberg, J. S. and D. P. Looze: "Right Half-Plane Poles and Zeros and Design Trade-Offs in Feedback Systems", *IEEE Trans. Automatic Control*, AC-30, pp. 555–565, 1985.
12. Chu, Y.: "correlation between frequency and transient responses of feedback control systems", *Trans. AIEE*, Vol. 72, pt. II, pp. 82–92, 1953.

13. Higgins, P. J. and C. M. Siegel: "determination of the Maximum Modulus, or the Specified Gain of a Servomechanism by Complex Variable Differentiation', *Trans. AIEE*, vol. 72, pt. II, pp. 467, January 1954.
14. Brown, G. S. and D. P. Campbell: *Principles of Servomechanisms*, Wiley, New York, 1948.
15. Rowland, J. R.: *Linear Control Systems*, Wiley, New York, 1986.
16. Chestnut, H. and R. W. Mayer: *Servomechanisms and Regulating System Design*. 2nd Ed., Vol. 1, Wiley, New York, 1959.

9

Closed-Loop Tracking Performance Based on the Frequency Response

9.1 INTRODUCTION

Chapter 8 is devoted to plotting the open-loop transfer function $G(j\omega)H(j\omega)$. The two types of plots that are emphasized are the direct polar plot (Nyquist plot) and the log magnitude–angle plot (Nichols plot). Also included in Chapter 8 is the determination of closed-loop stability in terms of the open-loop transfer function by use of the Nyquist stability criterion, which is illustrated in terms of both types of plot. The result of the stability study places bounds on the permitted range of values of gain.

This chapter develops a correlation between the frequency and the time responses of a system, leading to a method of gain setting in order to achieve a specified closed-loop frequency response [1]. The closed-loop frequency response is obtained as a function of the open-loop frequency response. Through the graphical representations in this chapter, the reader should obtain an appreciation of the responses that can be achieved by gain adjustment. CAD programs like TOTAL-PC and MATLAB (see Appendix C and D) can be used interactively to perform a number of designs rapidly. In addition, although the design is performed in the frequency domain, the closed-loop responses in the time domain are also obtained. Then a "best" design is selected by considering both the frequency and the time responses. CAD programs considerably reduce the computational time of the designer in achieving a system that meets the design specifications.

9.2 DIRECT POLAR PLOT

Consider a simple control system with unity feedback, as shown in Fig. 9.1. The performance of this system with a sinusoidal input $\mathbf{R}(j\omega)$ is represented by the following equations:

$$\mathbf{E}(j\omega) = \mathbf{R}(j\omega) - \mathbf{C}(j\omega) \tag{9.1}$$

$$\frac{\mathbf{C}(j\omega)}{\mathbf{E}(j\omega)} = \mathbf{G}(j\omega) = \left|\mathbf{G}(j\omega)\right| e^{j\phi} \tag{9.2}$$

$$\frac{\mathbf{C}(j\omega)}{\mathbf{R}(j\omega)} = \frac{\mathbf{G}(j\omega)}{1 + \mathbf{G}(j\omega)} = \mathbf{M}(j\omega) = M(\omega)e^{j\alpha(\omega)} \tag{9.3}$$

$$\frac{\mathbf{E}(j\omega)}{\mathbf{R}(j\omega)} = \frac{1}{1 + \mathbf{G}(j\omega)} \tag{9.4}$$

Figure 9.2 is the polar plot of $\mathbf{G}(j\omega)$ for this control system. The directed line segment drawn from the $-1 + j0$ point to any point on the $\mathbf{G}(j\omega)\mathbf{H}(j\omega)$

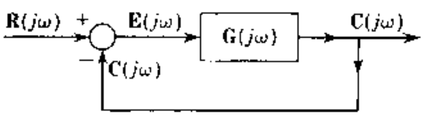


FIGURE 9.1 Block diagram of a simple control system.

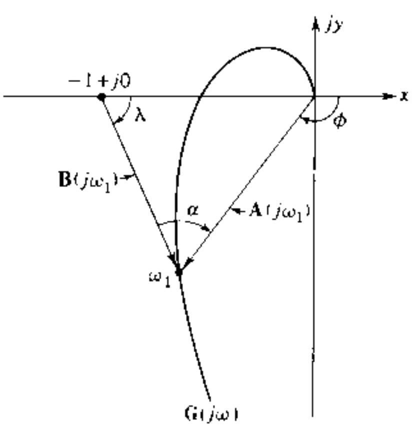


FIGURE 9.2 Polar plot of $\mathbf{G}(j\omega)$ for the control system of Figure 9.1.

curve represents the quantity $\mathbf{B}(j\omega) = 1 + \mathbf{G}(j\omega)\mathbf{H}(j\omega)$. For a unity-feedback system, $\mathbf{H}(j\omega) = 1$; therefore,

$$\mathbf{B}(j\omega) = |\mathbf{B}(j\omega)| e^{j\lambda(\omega)} = 1 + \mathbf{G}(j\omega) \tag{9.5}$$

The frequency control ratio $\mathbf{C}(j\omega)/\mathbf{R}(j\omega)$ is therefore equal to the ratio

$$\frac{\mathbf{C}(j\omega)}{\mathbf{R}(j\omega)} = \frac{\mathbf{A}(j\omega)}{\mathbf{B}(j\omega)} = \frac{|\mathbf{A}(j\omega)| e^{j\phi(\omega)}}{|\mathbf{B}(j\omega)| e^{j\lambda(\omega)}} = \frac{\mathbf{G}(j\omega)}{1 + \mathbf{G}(j\omega)} \tag{9.6}$$

$$\frac{\mathbf{C}(j\omega)}{\mathbf{R}(j\omega)} = \frac{|\mathbf{A}(j\omega)|}{|\mathbf{B}(j\omega)|} e^{j(\phi-\lambda)} = M(\omega) e^{j\alpha(\omega)} \tag{9.7}$$

Note that the angle $\alpha(\omega)$ can be determined directly from the construction shown in Fig. 9.2. Since the magnitude of the angle $\phi(\omega)$ is greater than the magnitude of the angle $\lambda(\omega)$, the value of the angle $\alpha(\omega)$ is negative. Remember that ccw rotation is taken as positive.

Combining Eqs. (9.4) and (9.5) gives

$$\frac{\mathbf{E}(j\omega)}{\mathbf{R}(j\omega)} = \frac{1}{1 + \mathbf{G}(j\omega)} = \frac{1}{|\mathbf{B}(j\omega)| e^{j\lambda}} \tag{9.8}$$

From Eq. (9.8) it is seen that the greater the distance from the $-1 + j0$ point to a point on the $\mathbf{G}(j\omega)$ locus, for a given frequency, the smaller the steady-state sinusoidal error for a particular sinusoidal input. Thus, the usefulness and importance of the polar plot of $\mathbf{G}(j\omega)$ have been enhanced.

9.3 DETERMINATION OF M_m AND ω_m FOR A SIMPLE SECOND-ORDER SYSTEM

The frequency at which the maximum value of $\mathbf{C}(j\omega)/\mathbf{R}(j\omega)$ occurs is referred to as the *resonant frequency* ω_m. The maximum value is labeled M_m. These two quantities are figures of merit of a system. As shown in later chapters, compensation to improve system performance is based upon a knowledge of ω_m and M_m.

For a *simple second-order system* a direct and simple relationship can be obtained for M_m and ω_m in terms of the system parameters. Consider the position-control system of Fig. 9.3, which contains the servomotor described by Eq. (2.138). The forward and closed-loop transfer functions are,

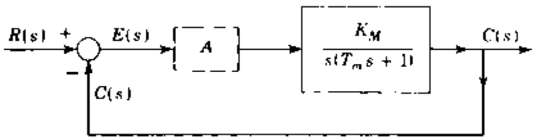


FIGURE 9.3 A position-control system.

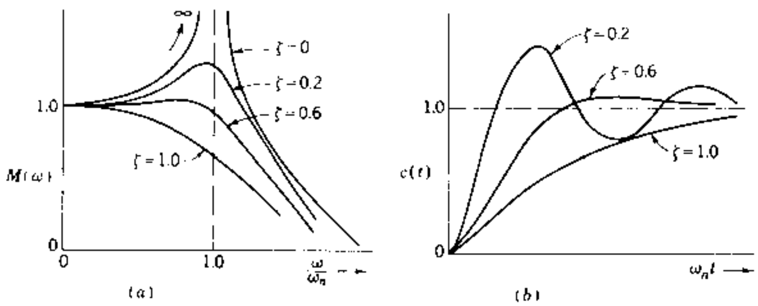


FIGURE 9.4 (*a*) Plots of M vs. ω/ω_n for a simple second-order system; (*b*) corresponding time plots for a step input.

respectively,

$$\frac{C(s)}{E(s)} = \frac{K_1}{s(T_m s + 1)} \tag{9.9}$$

$$\frac{C(s)}{R(s)} = \frac{1}{(T_m/K_1)s^2 + (1/K_1)s + 1} = \frac{1}{s^2/\omega_n^2 + (2\zeta/\omega_n)s + 1} \tag{9.10}$$

where $K_1 = AK_M$. The damping ratio and the undamped natural frequency for this system, as determined from Eq. (9.10), are $\zeta = 1/(2\sqrt{K_1 T_m})$ and $\omega_n = \sqrt{K_1/T_m}$. The control ratio as a function of frequency is therefore

$$\frac{\mathbf{C}(j\omega)}{\mathbf{R}(j\omega)} = \frac{1}{(1 - \omega^2/\omega_n^2) + j2\zeta(\omega/\omega_n)} = M(\omega)e^{j\alpha(\omega)} \tag{9.11}$$

Plots of $M(\omega)$ and $\alpha(\omega)$ vs. ω, for a particular value of ω_n, can be obtained for different values of ζ (see Fig. 8.4). In order to obtain a family of curves that are applicable to all simple second-order systems with different values of ω_n, they are plotted vs. ω/ω_n in Fig. 9.4*a*.

The inverse Laplace transform of $C(s)$, for a unit-step input, gives the time response, as derived in Sec. 3.9 for $\zeta < 1$, as

$$c(t) = 1 - \frac{1}{\sqrt{1 - \zeta^2}} e^{-\zeta\omega_n t} \sin\left(\omega_n\sqrt{1 - \zeta^2}\, t + \cos^{-1}\zeta\right) \tag{9.12}$$

The plot of $c(t)$ for several values of damping ratio ζ is drawn in Fig. 9.4*b*.

Next, consider the magnitude M^2, as derived from Eq. (9.11):

$$M^2 = \frac{1}{(1-\omega^2/\omega_n^2)^2 + 4\zeta^2(\omega^2/\omega_n^2)} \tag{9.13}$$

To find the maximum value of M and the frequency at which it occurs, Eq. (9.13) is differentiated with respect to frequency and set equal to zero:

$$\frac{dM^2}{d\omega} = -\frac{-4(1-\omega^2/\omega_n^2)(\omega/\omega_n^2) + 8\zeta^2(\omega/\omega_n^2)}{[(1-\omega^2/\omega_n^2)^2 + 4\zeta^2(\omega^2/\omega_n^2)]^2} = 0 \tag{9.14}$$

The frequency ω_m at which the value M exhibits a peak (see Fig. 9.5), as found from Eq. (9.14), is

$$\omega_m = \omega_n\sqrt{1-2\zeta^2} \tag{9.15}$$

This value of frequency is substituted into Eq. (9.13) to yield

$$M_m = \frac{1}{2\zeta\sqrt{1-\zeta^2}} \tag{9.16}$$

From these equations it is seen that the curve of M vs. ω has a peak value, other than at $\omega = 0$, only for $\zeta < 0.707$. Figure 9.6 is a plot of M_m vs. ζ for a simple

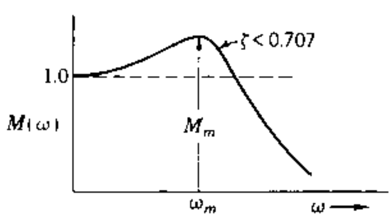


FIGURE 9.5 A closed-loop frequency-response curve indicating M_m and ω_m.

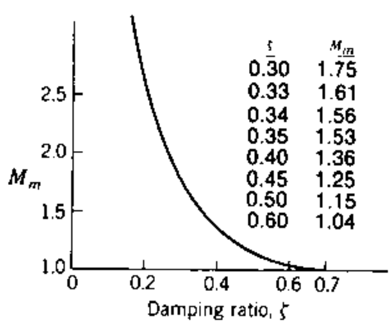


FIGURE 9.6 A plot of M_m vs. ζ for a simple second-order system.

second-order system. For values of $\zeta < 0.4$, M_m increases very rapidly in magnitude; the transient oscillatory response is therefore excessively large and might damage the physical equipment.

In Secs. 3.5 and 7.13 the damped natural frequency ω_d for the transient of the simple second-order system is shown to be

$$\omega_d = \omega_n\sqrt{1-\zeta^2} \tag{9.17}$$

Also, in Sec. 3.9 the expression for the peak value M_p for a unit-step input to this simple system is determined. It is repeated here:

$$M_p = \frac{c_p}{r} = 1 + \exp\left(-\frac{\zeta\pi}{\sqrt{1-\zeta^2}}\right) \tag{9.18}$$

Therefore, for this simple second-order system the following conclusions are obtained in correlating the frequency and time responses:

1. Inspection of Eq. (9.15) reveals that ω_m is a function of both ω_n and ζ. Thus, for a given value of ζ, the larger the value of ω_m, the larger ω_n, and the faster the transient time of response for this system.
2. Inspection of Eqs. (9.16) and (9.18) shows that both M_m and M_p are functions of ζ. The smaller ζ becomes, the larger in value M_m and M_p become. Thus, it can be concluded that the larger the value of M_m, the larger the value of M_p. For values of $\zeta < 0.4$ the correspondence between M_m and M_p is only qualitative for this simple case. In other words, for $\zeta = 0$ the time domain yields $M_p = 2$, whereas the frequency domain yields $M_m = \infty$. When $\zeta > 0.4$, there is a close correspondence between M_m and M_p. As an example, for ζ equal to 0.6, $M_m = 1.04$ and $M_p = 1.09$.
3. In Fig. 9.7 are shown polar plots for different damping ratios for the simple second-order system. Note that the shorter the distance between the $-1 + j0$ point and a particular $\mathbf{G}(j\omega)$ plot, as indicated by the dashed lines, the smaller the damping ratio. Thus, M_m is larger and consequently M_p is also larger.

From these characteristics a designer can obtain a good approximation of the time response of a simple second-order system by knowing only the M_m and ω_m of its frequency response.

The procedure of setting the derivative, with respect to ω, of $\mathbf{C}(j\omega)/\mathbf{R}(j\omega)$ to zero works very well with a simple system. But the differentiation of $\mathbf{C}(j\omega)/\mathbf{R}(j\omega)$ and solution for ω_m and M_m become tedious for more complex systems. Therefore, a graphic procedure is generally used, as shown in the following sections [2].

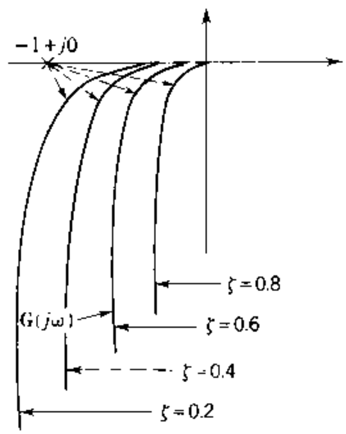


FIGURE 9.7 Polar plots of $G(j\omega)$ for different damping ratios for the system shown in Fig. 9.3.

9.4 CORRELATION OF SINUSOIDAL AND TIME RESPONSES [3]

Although the correlation in the preceding section is for a simple second-order system, it has been found by experience that M_m is also a function of the *effective* ζ and ω_n for higher-order systems. The effective ζ and ω_n of a higher-order system (see Sec.7.13) is dependent upon the combined effect of the ζ and ω_n of each second-order term and the values of the real roots of the characteristic equation of $C(s)/R(s)$. Thus, in order to alter the effective M_m, the location of the roots must be changed. Which ones should be altered depends on which are dominant in the time domain.

From the analysis for a simple second-order system, whenever the frequency response has the shape shown in Fig 9.5, the following correlation exists between the frequency and time responses for systems of any order:

1. The larger ω_m is made, the faster the time of response for the system.
2. The value of M_m gives a qualitative measure of M_p within the acceptable range of the effective damping ratio $0.4 < \zeta < 0.707$. In terms of M_m, the acceptable range is $1 < M_m < 1.4$.
3. The closer the $G(j\omega)$ curve comes to the $-1 + j0$ point, the larger the value of M_m.

The characteristics developed in Chap. 6 can be added to these three items; i.e., the larger K_p, K_v or K_a is made, the greater the steady-state accuracy for a step, a ramp, and a parabolic input, respectively. In terms of the polar plot, the farther the point $\mathbf{G}(j\omega)|_{\omega=0} = K_0$ for a Type 0 system is from the origin, the more accurate is the steady-state time response for a step input. For a Type 1 system, the farther the low-frequency asymptote (as $\omega \to 0$) is from the imaginary axis, the more accurate is the steady-state time response for a ramp input.

All the factors just mentioned are merely *guideposts* in the *frequency domain* to assist the designer in obtaining an *approximate* idea of the time response of a system. They serve as "stop-and-go signals" to indicate if the design is headed in the right direction in achieving the desired time response. If the desired performance specifications are not satisfactorily met, compensation techniques (see later chapters) must be used. After compensation the exact time response can be obtained by taking the inverse Laplace transform of $C(s)$. The approximate approach saves much valuable time. Exceptions to the preceding analysis do occur for higher-order systems. The digital computer is a valuable tool for obtaining the exact closed-loop frequency and time responses. Thus a direct correlation between them can easily be made.

9.5 CONSTANT $M(\omega)$ AND $\alpha(\omega)$ CONTOURS OF $C(j\omega)/R(j\omega)$ ON THE COMPLEX PLANE (DIRECT PLOT)

The open-loop transfer function and its polar plot for a given feedback control system have provided the following information so far:

1. The stability or instability of the system
2. If the system is stable, the degree of its stability
3. The system type
4. The degree of steady-state accuracy
5. A graphical method of determining $\mathbf{C}(j\omega)/\mathbf{R}(j\omega)$

These items provide a qualitative idea of the system's time response. The polar plot is useful since it permits the determination of M_m and ω_m. Also, it is used to determine the gain required to produce a specified value of M_m.

The contours of constant values of M drawn in the complex plane yield a rapid means of determining the values of M_m and ω_m and the value of gain required to achieve a desired value M_m. In conjunction with the contours of constant values of $\alpha(\omega)$, also drawn in the complex plane, the plot of $\mathbf{C}(j\omega)/\mathbf{R}(j\omega)$ can be obtained rapidly. The M and α contours are developed only for unity-feedback systems.

Equation of a Circle

The equation of a circle with its center at the point (a, b) and having radius r is

$$(x - a)^2 + (y - b)^2 = r^2 \tag{9.19}$$

The circle is displaced to the left of the origin for a negative value of a, and for negative values of b the circle is displaced below the origin. This equation of a circle in the xy plane is used later in this section to express contours of constant M and α.

$M(\omega)$ Contours

Figure 9.2 is the polar plot of a forward transfer function $\mathbf{G}(j\omega)$ of a *unity-feedback system*. From Eq. (9.3), the magnitude M of the control ratio is

$$M(\omega) = \frac{|\mathbf{A}(j\omega)|}{|\mathbf{B}(j\omega)|} = \frac{|\mathbf{G}(j\omega)|}{|1 + \mathbf{G}(j\omega)|} \tag{9.20}$$

The question at hand is: How many points are there in the complex plane for which the ratios of the magnitudes of the phasors $\mathbf{A}(j\omega)$ and $\mathbf{B}(j\omega)$ have the same value of $M(\omega)$? For example, by referring to Fig. 9.2. it can be seen that for the frequency $\omega = \omega_1$, M has a value M_a. It is desirable to determine all the other points in the complex plane for which

$$\frac{|\mathbf{A}(j\omega)|}{|\mathbf{B}(j\omega)|} = M_a$$

To derive the constant M locus, express the transfer function in rectangular coordinates. That is,

$$\mathbf{G}(j\omega) = x + jy \tag{9.21}$$

Substituting this equation into Eq. (9.20) gives

$$M = \frac{|x + jy|}{|1 + x + jy|} = \left[\frac{x^2 + y^2}{(1 + x)^2 + y^2}\right]^{1/2} \qquad \text{or} \qquad M^2 = \frac{x^2 + y^2}{(1 + x)^2 + y^2}$$

Rearranging the terms of this equation yields

$$\left(x + \frac{M^2}{M^2 - 1}\right)^2 + y^2 = \frac{M^2}{(M^2 - 1)^2} \tag{9.22}$$

By comparison with Eq. (9.19) it is seen that Eq. (9.22) is the equation of a circle with its center on the real axis with M as a parameter. The center is located at

$$x_0 = -\frac{M^2}{M^2 - 1} \tag{9.23}$$

$$y_0 = 0 \tag{9.24}$$

and the radius is

$$r_0 = \left|\frac{M}{M^2 - 1}\right| \tag{9.25}$$

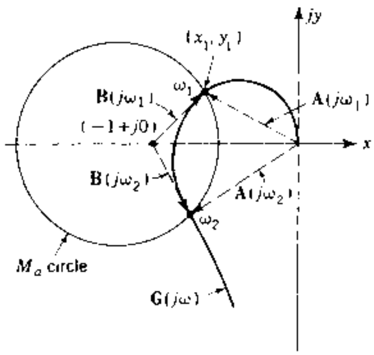


FIGURE 9.8 A plot of the transfer function $\mathbf{G}(j\omega)$ and the M_a circle.

Inserting a given value of $M = M_a$ into Eq. (9.22) results in a circle in the complex plane having a radius r_0 and its center at (x_0, y_0). This circle is called a constant M contour for $M = M_\mathrm{a}$.

The ratio of magnitudes of the phasors $\mathbf{A}(j\omega)$ and $\mathbf{B}(j\omega)$ drawn to any point on the M_a circle has the same value. As an example, refer to Fig. 9.8, in which the M_a circle and the $\mathbf{G}(j\omega)$ function are plotted. Since the circle intersects the $\mathbf{G}(j\omega)$ plot at the two frequencies ω_1 and ω_2, there are two frequencies for which

$$\frac{|\mathbf{A}(j\omega_1)|}{|\mathbf{B}(j\omega_1)|} = \frac{|\mathbf{A}(j\omega_2)|}{|\mathbf{B}(j\omega_2)|} = M_a$$

In other words, a given point (x_1, y_1) is simultaneously a point on a particular transfer function $\mathbf{G}(j\omega)$ and a point on the M circle passing through it.

The plot of Fig. 9.8 is redrawn in Fig. 9.9 with two more M circles added. In this figure the circle $M = M_b$ is just tangent to the $\mathbf{G}(j\omega)$ plot. There is only one point (x_3, y_3) for $\mathbf{G}(j\omega_3)$ in the complex plane for which the ratio $\mathbf{A}(j\omega)/\mathbf{B}(j\omega)$ is equal to M_b. Also, the M_c circle does not intersect and is not tangent to the $\mathbf{G}(j\omega)$ plot, indicating that there are no points in the plane that can simultaneously satisfy Eqs. (9.20) and (9.22) for $M = M_c$.

Figure 9.10 shows a family of circles in the complex plane for different values of M. In this figure, notice that the larger the value M, the smaller is its corresponding M circle. Thus, for the example shown in Fig. 9.9 the ratio $\mathbf{C}(j\omega)/\mathbf{R}(j\omega)$, for a unity-feedback control system, has a maximum value of M equal to $M_m = M_b$. A further inspection of Fig. 9.10 and Eq. (9.23) reveals the following:

1. For $M \to \infty$, which represents a condition of oscillation ($\zeta \to 0$), the center of the M circle $x_0 \to -1 + j0$ and the radius $r_0 \to 0$.

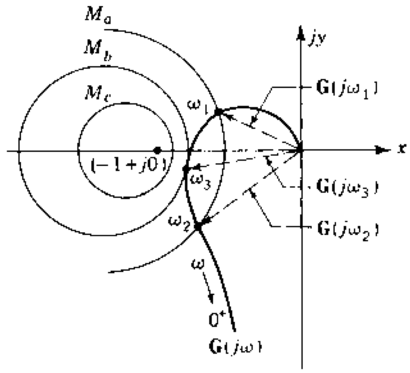


FIGURE 9.9 *M* contours and a $\mathbf{G}(j\omega)$ plot.

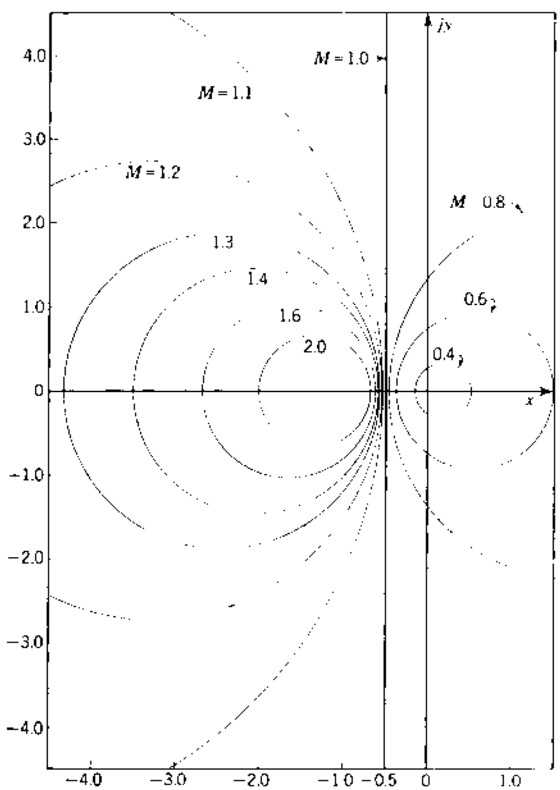


FIGURE 9.10 Constant *M* contours.

This agrees with the statement made previously that as the $\mathbf{G}(j\omega)$ plot comes closer to the $-1 + j0$ point, the system's effective ζ becomes smaller and the degree of its stability decreases.

2. For $M(\omega) = 1$, which represents the condition where $\mathbf{C}(j\omega) = \mathbf{R}(j\omega)$, $r_0 \to \infty$ and the M contour becomes a straight line perpendicular to the real axis at $x = -\frac{1}{2}$.
3. For $M \to 0$, the center of the M circle $x_0 \to 0$ and the radius $r_0 \to 0$.
4. For $M > 1$ the centers of the circles lie to the left of $x = -1 + j0$, and for $M < 1$ the centers of the circles lie to the right of $x = 0$. All centers are on the real axis.

$\alpha(\omega)$ Contours

The $\alpha(\omega)$ contours, representing constant values of phase angle $\alpha(\omega)$ for $\mathbf{C}(j\omega)/\mathbf{R}(j\omega)$, can also be determined in the same manner as the M contours. Substituting Eq. (9.21) into Eq. (9,3) yields.

$$M(\omega)e^{j\alpha(\omega)} = \frac{x + jy}{(1 + x) + jy} = \frac{\mathbf{A}(j\omega)}{\mathbf{B}(j\omega)} \tag{9.26}$$

The question at hand now is: How many points are there in the complex plane for which the ratio of the phasors $\mathbf{A}(j\omega)$ and $\mathbf{B}(j\omega)$ yields the same angle α? To answer this question, express the angle α obtained from Eq. (9.26) as follows:

$$\alpha = \tan^{-1}\frac{y}{x} - \tan^{-1}\frac{y}{1 + x}$$

$$= \tan^{-1}\frac{y/x - y/(1 + x)}{1 + (y/x)[y/(1 + x)]} = \tan^{-1}\frac{y}{x^2 + x + y^2}$$

$$\tan\alpha = \frac{y}{x^2 + x + y^2} = N \tag{9.27}$$

For a constant value of the angle α, $\tan\alpha = N$ is also constant. Rearranging Eq. (9.27) results in

$$\left(x + \frac{1}{2}\right)^2 + \left(y - \frac{1}{2N}\right)^2 = \frac{1}{4}\frac{N^2 + 1}{N^2} \tag{9.28}$$

By comparing with Eq. (9.19) it is seen that Eq. (9.28) is the equation of a circle with N as a parameter. It has its center at

$$x_q = -\frac{1}{2} \tag{9.29}$$

$$y_q = \frac{1}{2N} \tag{9.30}$$

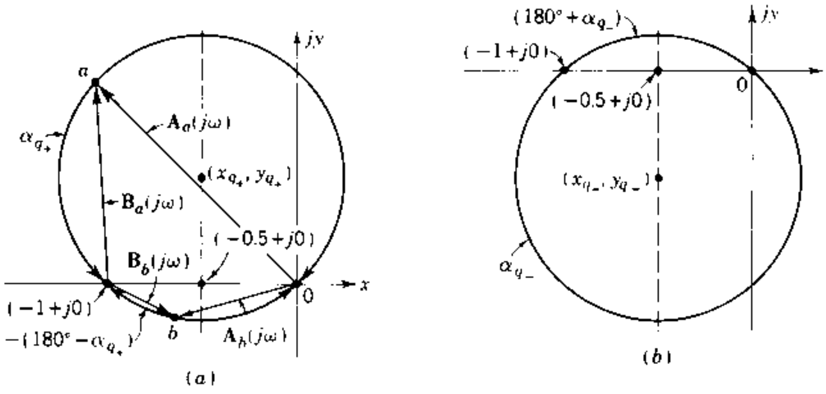


FIGURE 9.11 Arcs of constant α.

and has a radius of

$$r_q = \frac{1}{2}\left(\frac{N^2+1}{N^2}\right)^{1/2} \tag{9.31}$$

Inserting a given value of $N_q = \tan\alpha_q$ into Eq. (9.28) results in a circle (see Fig. 9.11) of radius r_q with its center at (x_q, y_q)

The tangent of angles in the first and third quadrant is positive. Therefore, the y_q coordinate given by Eq. (9.30) is the same for an angle in the first quadrant and for the negative of its supplement, which is in the third quadrant. As a result, the constant α contour is only an arc of the circle. As an example, the $\alpha = -310^\circ$ and -130° arcs are part of the same circle, as shown in Fig. 9.11*a*. Similarly, angles α in the second and fourth quadrants have the same value y_q if they are negative supplements of each other. The constant α contours for these angles are shown in Fig. 9.11*b*.

The ratio of the complex quantities $\mathbf{A}(j\omega)$ and $\mathbf{B}(j\omega)$, for all points on the α_q arc, yields the same phase angle α_q, that is,

$$\angle\mathbf{A}(j\omega) - \angle\mathbf{B}(j\omega) = \alpha_q \tag{9.32}$$

A good procedure, in order to avoid ambiguity in determining the angle α is to start at zero frequency, where $\alpha = 0$, and to proceed to higher frequencies, knowing that the angle is a continuous function.

Different values of α result in a family of circles in the complex plane with centers on the line represented by $(-\frac{1}{2}, y)$, as illustrated in Fig. 9.12.

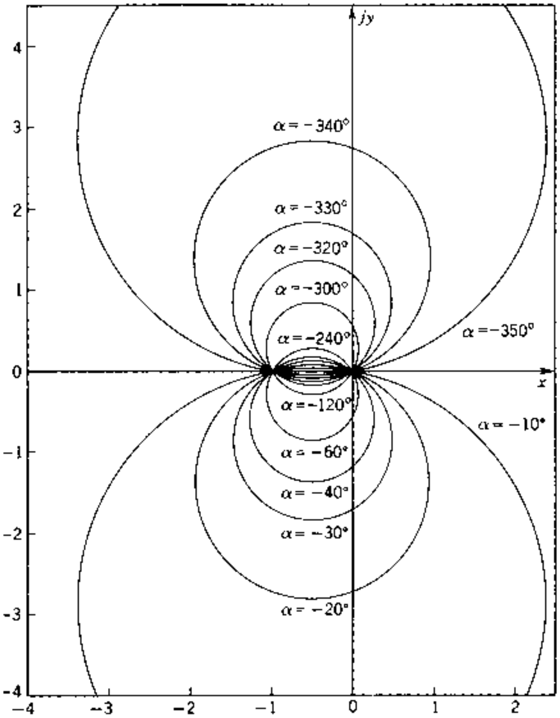


FIGURE 9.12 Constant α contours.

Tangents to the *M* Circles

The line drawn through the origin of the complex plane tangent to a given M circle plays an important part in setting the gain of $G(j\omega)$. Referring to Fig. 9.13 and recognizing that the line $bc = r_0$ is the radius and $ob = x_0$ is the distance to the center of the particular M circle, it becomes evident that

$$\sin\psi = \frac{bc}{ob} = \frac{M/(M^2-1)}{M^2/(M^2-1)} = \frac{1}{M} \tag{9.33}$$

This relationship is utilized in the section on gain adjustment. Also, the point a in Fig. 9.13 is the $-1 + j0$ point. This statement is proved from

$$(oc)^2 = (ob)^2 - (bc)^2 \qquad ac = oc \ \sin\psi \quad \text{and} \quad (oa)^2 = (oc)^2 - (ac)^2$$

Combining these equations yields

$$(oa)^2 = \frac{M^2-1}{M^2}\left[(ob)^2 - (bc)^2\right] \tag{9.34}$$

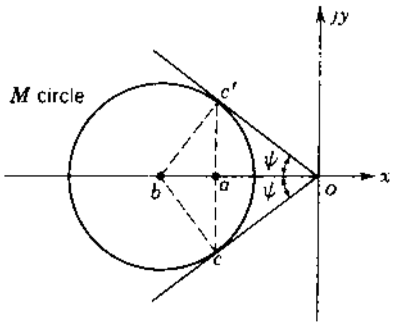


FIGURE 9.13 Determination of sin Ψ.

The values of the distances to the center ob and of the radius bc of the M circle [see Eqs. (9.23) and (9.25)] are then substituted into Eq. (9.34), which results in

$$oa = 1 \tag{9.35}$$

Some typical values necessary for constructing M and α contours are as follows:

$M_o = 1.1$	$x_o = -5.76$	$r_o = 5.24$	$\psi = 65.4^\circ$
$M_o = 1.3$	$x_o = -2.45$	$r_o = 1.88$	$\psi = 50.3^\circ$
$\alpha = -60^\circ$	$N = -1.73$	$r_q = 0.577$	$y_q = -0.289$
$\alpha = 30^\circ$	$N = 0.577$	$r_q = 1.000$	$y_q = 0.866$

9.6 CONSTANT 1/*M* AND α CONTOURS (UNITY FEEDBACK) IN THE INVERSE POLAR PLANE

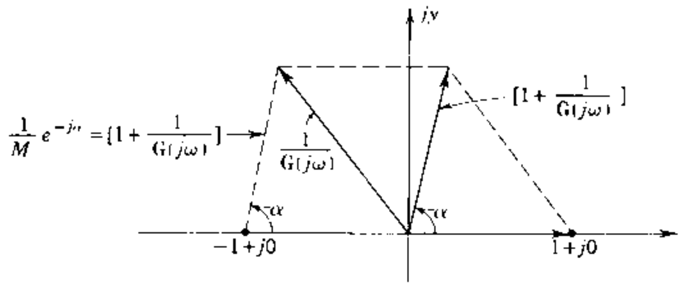
The closed-loop unity-feedback system can be represented by either of the following equations:

$$\frac{\mathbf{C}(j\omega)}{\mathbf{R}(j\omega)} = \frac{\mathbf{G}(j\omega)}{1 + \mathbf{G}(j\omega)} = Me^{j\alpha} \tag{9.36}$$

$$\frac{\mathbf{R}(j\omega)}{\mathbf{C}(j\omega)} = \frac{1}{\mathbf{G}(j\omega)} + 1 = \frac{1}{M}e^{-j\alpha} \tag{9.37}$$

Note that Eq. (9.37) is composed of two complex quantities that can be readily plotted in the complex plane, as shown in Fig. 9.14. Thus, $\mathbf{R}(j\omega)/\mathbf{C}(j\omega)$ can be obtained graphically by plotting the complex quantities $1/\mathbf{G}(j\omega)$ and 1.

Constant $1/M$ and α contours for the inverse plots are developed for unity-feedback systems. Figure 9.14 illustrates the quantities in Eq. (9.37) for



(9.37) with $\mathbf{H}(j\omega) = 1$.

FIGURE 9.14 Representation of the of the quantities in Eq. (9.37) with $\mathbf{H}(j\omega) = 1$.

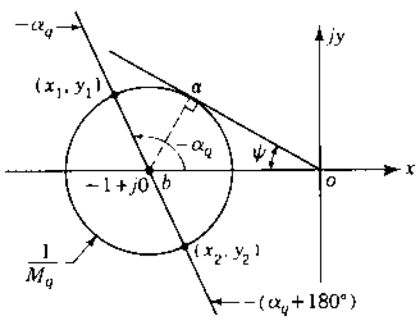


FIGURE 9.15 Contour for a particular magnitude $\mathbf{R}(j\omega)/\mathbf{C}(j\omega) = 1/M_q$.

a unity-feedback system, with $\mathbf{H}(j\omega) = 1$. Note that, because of the geometry of construction, the directed line segment drawn from the $-1 + j0$ point is also the quantity $1 + 1/\mathrm{G}(j\omega)$ or $(1/M)e^{-j\alpha}$. Thus, the inverse plot yields very readily the form and location of the contours of constant $1/M$ and α. In other words, in the inverse plane it can be seen that:

1. Contours of constant values of M are circles whose centers are at the $-1 + j0$ point, and the radii are equal to $1/M$.
2. Contours of constant values of $-\alpha$ are radial lines that pass through the $-1 + j0$ point.

Figure 9.15 shows the contour for a particular magnitude of $\mathbf{R}(j\omega)/\mathbf{C}(j\omega)$ and a tangent drawn to the M circle from the origin. In Fig. 9.15,

$$\sin\psi = \left|\frac{ab}{ob}\right| = \frac{1/M}{1} = \frac{1}{M}$$

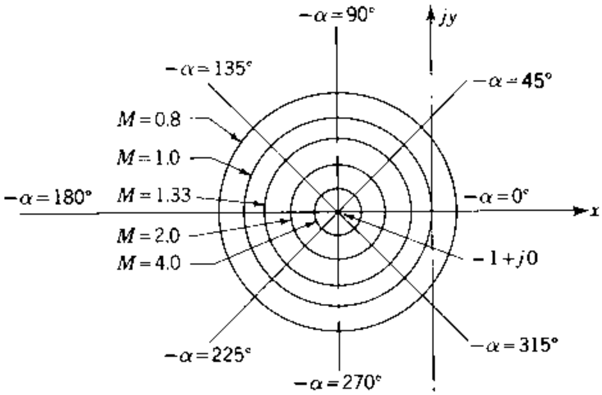


FIGURE 9.16 Typical M and α contours on the inverse polar plane for unity feedback.

which is the same as for the contours in the direct plane. The directed line segment from $-1 + j\,0$ to the point (x_1, y_1) has a value of $1/M_q$ and angle of $-\alpha_q$. If the polar plot of $1/\mathbf{G}(j\omega)$ passes through the (x_1, y_1) point, then $\mathbf{R}(j\omega_1)/\mathbf{C}(j\omega_1)$ has these values. The directed line segment from $-1 + j0$ to the point (x_2, y_2) has a magnitude of $1/M_q$ and an angle of $-\alpha q - 180°$. If the polar plot of $1/\mathbf{G}(j\omega)$ passes through the (x_2, y_2) point, then $|\mathbf{R}(j\omega_2)/\mathbf{C}(j\omega_2)|$ has these values.

Figure 9.16 illustrates families of typical M and α contours. The plots of constant M values shown in Fig. 9.16 are concentric circles having the same center at $-1 + j\,0$. They are used in Sec. 9.8 to develop the constant M contours in the log magnitude–angle diagram (Nichols chart).

9.7 GAIN ADJUSTMENT OF A UNITY-FEEDBACK SYSTEM FOR A DESIRED M_m: DIRECT POLAR PLOT

Gain adjustment is the first step in adjusting the system for the desired performance. The procedure for adjusting the gain is outlined in this section. Figure 9.17a shows $\mathbf{G}_x(j\omega)$ with its respective M_m circle in the complex plane. Since

$$\mathbf{G}_x(j\omega) = x + jy = K_x \mathbf{G}'_x(j\omega) = K_x(x' + jy') \tag{9.38}$$

it follows that

$$x' + jy' = \frac{x}{K_x} + j\frac{y}{K_x}$$

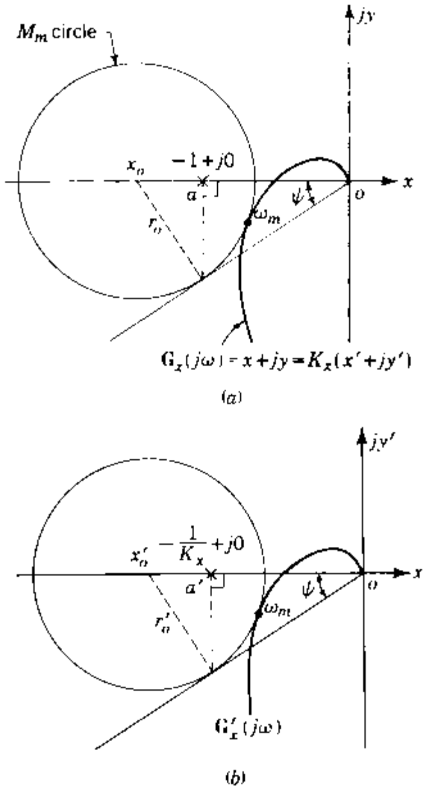


FIGURE 9.17 (*a*) Plot of $\mathbf{G}_x(j\omega)$ with the respective M_m circle. (*b*) Circle drawn tangent both to the plot of $\mathbf{G}'_x(j\omega)$ and to the line representing the angle $\Psi = \sin^{-1}(1/M_m)$.

where $\mathbf{G}'_x(j\omega) = \mathbf{g}_x(j\omega)/K_x$ is defined as the frequency-sensitive portion of $\mathbf{G}_x(j\omega)$ with unity gain. Note that changing the gain merely changes the amplitude and not the angle of the locus of points of $\mathbf{G}_x(j\omega)$. Thus, if in Fig. 9.17*a* a change of scale is made by dividing the, x, y coordinates by K_x so that the new coordinates are x', y', the following are true:

1. The $\mathbf{G}_x(j\omega)$ plot becomes the $\mathbf{G}'_x(j\omega)$ plot.
2. The M_m circle becomes simultaneously tangent to $\mathbf{G}'_x(j\omega)$ and to the line representing $\sin\Psi = 1/M_m$.
3. The $-1 + j0$ point becomes the $-1/K_x + j0$ point.
4. The radius r_0 becomes $r'_0 = r_0/K_x$.

In other words, if $\mathbf{G}'_x(j\omega)$ and $\mathbf{G}_x(j\omega)$ are drawn on separate sheets of graph paper (see Fig. 9.17b), and the two graphs are superimposed so that the axes coincide, the circles and the $\mathbf{G}_x(j\omega)$ and $\mathbf{G}'_x(j\omega)$ plots also coincide. Referring to Fig. 9.17*a* and 9.17*b*, note that $oa = -1$ and $oa' = -1/K_x$.

As a consequence, it is possible to determine the required gain to achieve a desired M_{m} for a given system by the following graphical procedure:

Step 1. If the original system has a transfer function

$$\mathbf{G}_x(j\omega) = \frac{K_x(1 + j\omega T_1)(1 + j\omega T_2)\cdots}{(j\omega)^m(1 + j\omega T_a)(1 + j\omega T_b)(1 + j\omega T_c)\cdots} = K_x\mathbf{G}'_x(j\omega) \tag{9.39}$$

with an original gain K_x, only the frequency-sensitive portion $\mathbf{G}'_x(j\omega)$ is plotted.

Step 2. Draw the straight line at the angle $\Psi = \sin^{-1}(1/M_{\mathrm{m}})$, measured from the negative real axis.

Step 3. By trial and error, find a circle whose center lies on the negative real axis and is simultaneously tangent to both the $\mathbf{G}'_x(j\omega)$ plot and the line drawn at the angle Ψ.

Step 4. Having found this circle, locate the point of tangency on the Ψ-angle line. Draw a line from the point of the tangency perpendicular to the real axis. Label the point where this line intersects the real axis as a'.

Step 5. For this circle to be an M circle representing M_m, the point a' must be the $-1 + j0$ point. Thus, the x', y' coordinates must be multiplied by a gain factor K_m in order to convert this plot into a plot of $\mathbf{G}(j\omega)$. From the graphical construction the value K_m is $1/oa'$.

Step 6. The original gain must be changed by a factor $A = K_m/K_x$.

Note that if $\mathbf{G}_x(j\omega)$ that includes a gain K_x is already plotted, it is possible to work directly with the plot of the function $\mathbf{G}_x(j\omega)$. Following the procedure just outlined results in the determination of the *additional* gain required to produce the specified M_m; that is, the additional gain is

$$A = \frac{K_m}{K_x} = \frac{1}{oa''}$$

Example. It is desired that the closed-loop system that has the open-loop transfer function.

$$\mathbf{G}_x(j\omega) = \frac{1.47}{j\omega(1 + j0.25\omega)(1 + j0.1\omega)}$$

have an $M_m = 1.3$. The problem is to determine the actual gain K_1 needed and the amount by which the original gain K_x must be changed to obtain this M_m.

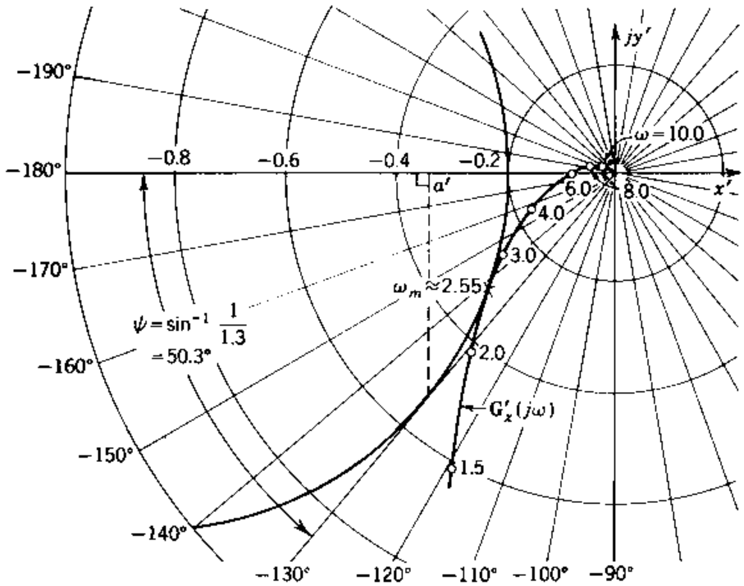


FIGURE 9.18 Gain adjustment using the procedure outlined in Sec. 9.7 for $\mathbf{G}'_x(j\omega) = 1/j\omega(1 + j0.25\omega)(1 + j0.1\omega)$.

The procedure previously outlined is applied to this problem. $\mathbf{G}'_x(j\omega)$ is plotted in Fig. 9.18 and results in $\omega_m = 2.55$ and

$$K_1 = \frac{1}{oa'} \approx \frac{1}{0.34} = 2.94\ \mathrm{s}^{-1}$$

The additional gain required is

$$A = \frac{K_1}{K_x} = \frac{2.94}{1.47} = 2.0$$

In other words, the original gain of $\mathbf{G}(j\omega)$ must be doubled to obtain M_m of 1.3 for $\mathbf{C}(j\omega)/\mathbf{R}(j\omega)$.

9.8 CONSTANT M AND α CURVES ON THE LOG MAGNITUDE—ANGLE DIAGRAM (NICHOLS CHART) [4]

As derived earlier in this chapter, the constant M curves on the direct and inverse polar plots are circles. The transformation of these curves to the log magnitude (Lm)–angle diagram is done more easily by starting from the inverse polar plot since all the M circles have the same center. This requires a change of sign of the Lm and angle obtained, since the transformation is

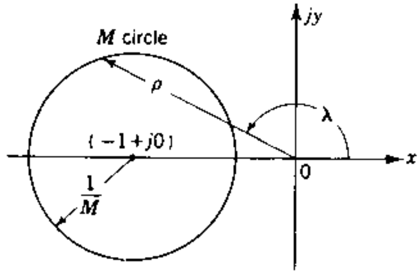


FIGURE 9.19 Constant M circle on the inverse polar plot.

from the inverse transfer function on the polar to the direct transfer function on the Lm plot.

A constant M circle is shown in Fig. 9.19 on the inverse polar plot. The magnitude ρ and angle λ drawn to any point on this circle are shown.

The equation for this M circle is

$$y^2 + (1+x)^2 = \frac{1}{M^2} \tag{9.40}$$

where

$$x = \rho \cos\lambda \qquad y = \rho \sin\lambda \tag{9.41}$$

Combining these equations produces

$$\rho^2 M^2 + 2\rho M^2 \cos\lambda + M^2 - 1 = 0 \tag{9.42}$$

Solving for ρ and λ yields

$$\rho = -\cos\lambda \pm \left(\cos^2\lambda - \frac{M^2-1}{M^2}\right)^{1/2} \tag{9.43}$$

$$\lambda = \cos^{-1}\frac{1 - M^2 - \rho^2 M^2}{2\rho M^2} \tag{9.44}$$

These equations are derived from the inverse polar plot $1/\mathrm{G}(j\omega)$. Since the Lm–angle diagram is drawn for the transfer function $\mathrm{G}(j\omega)$ and not for its reciprocal, a change in the equations must be made by substituting

$$r = \frac{1}{\rho} \qquad \phi = -\lambda$$

Since Lm $r = -$Lm ρ, it is only necessary to change the sign of Lm ρ. For any value of M a series of values of angles λ can be inserted in Eq. (9.43) to solve for ρ. This magnitude must be changed to decibels. Alternately, for any value of M a series of values of ρ can be inserted in Eq. (9.44) to solve for

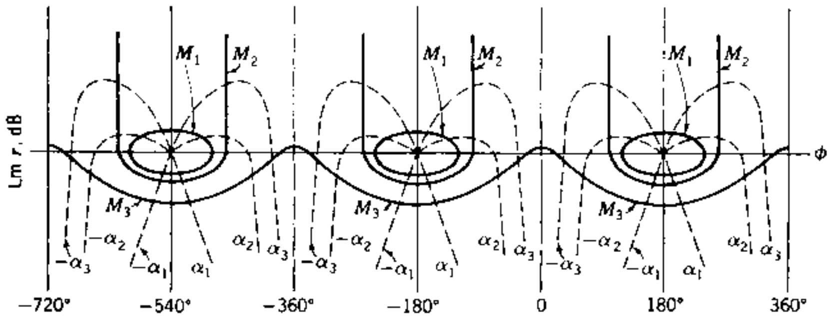


FIGURE 9.20 Constant M and α curves on the Nichols chart.

the corresponding angle λ. Therefore, the constant M curve on the Lm-angle diagram can be plotted by using either of these two equations.

In a similar fashion the constant α curves can be drawn on the Lm-angle diagram. The α curves on the inverse polar plot are semi-infinite straight lines terminating on the $-1 + j0$ point and are given by

$$\tan \alpha + x \tan \alpha + y = 0 \tag{9.45}$$

Combining Eq. (9.45) with Eq. (9.41) produces

$$\tan \alpha + \rho \cos \lambda \tan \alpha + \rho \sin \lambda = 0 \tag{9.46}$$

$$\tan \alpha = \frac{-\rho \sin \lambda}{1 + \rho \cos \lambda} = \frac{-y}{1 + x} \tag{9.47}$$

For constant values of α, a series of values of λ can be inserted in this equation to solve for ρ. The constant α semi-infinite curves can then be plotted on the Lm-angle diagram.

Constant M and α curves, as shown in Fig. 9.20, repeat for every 360°. Also, there is symmetry at every 180° interval. An expanded 300° section of the constant M and α graph is shown in Fig. 9.23. This graph is commonly referred to as the Nichols chart. Note that the $M = 1$(0 dB) curve is asymptotic to $\phi = -90°$ and $-270°$, and the curves for $M < \frac{1}{2}$ (−6 dB) are always negative. $M = \infty$ is the point at 0 dB, −180°, and the curves for $M > 1$ are closed curves inside the limits $\phi = -90°$ and $\phi = -270°$. The Nichols chart should be viewed as two charts. The first has the Cartesian coordinates of dB vs. phase angle ϕ that are used to plot the open-loop frequency response, with the frequencies noted along the plot. The second is the chart for loci of constant M and α for the closed-loop transfer function that is superimposed on the open-loop frequency-response plot. These loci for constant M and α for the Nichols chart are applied only for stable unity-feedback systems. A macro, utilizing Eqs. (9.44) and (9.47), can be written to generate a computer printout of the constant M and α contours on the Nichols chart.

9.9 GENERATION OF MATLAB BODE AND NYQUIST PLOTS

The MATLAB commands shown below are used to obtain Bode and Nyquist plots for the transfer function

$$\mathbf{G}(s) = \frac{K}{s(s+1)(s+5)} \tag{9.48}$$

In using these commands as a basis of other problems, it is important to pay attention to the selection of the appropriate scales for the plots. The smallest scale range for Bode plots should be selected based upon the 0-dB crossing of the Lm plot and the value of the gain-margin frequency. For the Nyquist plot it is important to select the scale so that the horizontal axis crossing points (-180° and/or $+180^\circ$) are clearly shown.

Bode Plot

```
%  Set up transfer function using open-loop poles and gain
p=[0 -1 -5];
num=5;
den=poly(p);
%
%  Use Bode command to obtain plots
%
%  In general, select dB scale of +20 to -40 dB, a minimum
%  angle scale, and a log frequency (w) scale that includes
%  the phase and gain margin frequencies
%
%  Get Bode plot of interest
%  Specify frequency range w space logarithmically from
%  10^-1 to 10^2
w=logspace (-1, 2,200);
%
%  Entering the command ''bode(num, den, w)'' will open a figure
%  window with the Bode plot on a default scale. The plot can be
%  edited using the menu buttons at the top of the window.
%  Alternatively, the plot format can be hard coded as follows:
%
%    Generate Bode plot data
[mag, phase]=bode(num, den, w);
%    Select top plot of two plots
subplot(2,1,1)
```

```
%    Plot magnitude data on semilog scale in dB
semilogx(w,20*log10(mag))
%    Set axis limits
axis([0.1,100,−40,20]);
%    Set title and labels
title('Example'),xlabel('omega(rad/sec)'),ylabel('dB'),grid
%    Select bottom plot of two plots
subplot (2,1,2)
%    Plot angle data on semilog scale
semilog x(w phase)
%    Set axis limits
axis([0.1,100,−270,0]);
%    Plot a line at −180
hold on
semilogx([0.1 100], [−180 −180])
hold off
%    Set title and labels
xlabel ('omega (rad/sec)'),ylabel ('Degrees'),grid
```

Nyquist Plot

```
%  Set up transfer function using open-loop poles and gain
p=[0 −1 −5];
num=5;
den=poly(p);
%
%  Use Nyquist command to obtain plots
%
%  In general, select a log frequency (w) scale that includes
%  the gain margin frequency
%
%  Get Nyquist plot of interest
%  Specify frequency range w
w=logspace(log10(0.5), log10 (20), 200);
%
%  Entering the command ''nyquist (num,den,w)'' will open a
%  figure window with the nyquist plot on a default scale.
%  The plot can be edited using the menu buttons at the top
%  of the window. Alternatively, the plot format can be hard
%  coded as follows:
%
nyquist (num,den,w);
hold on
num=50;
nyquist (num,den,w);
hold off
axis ([−2 0 −0.8 0.8]);
```

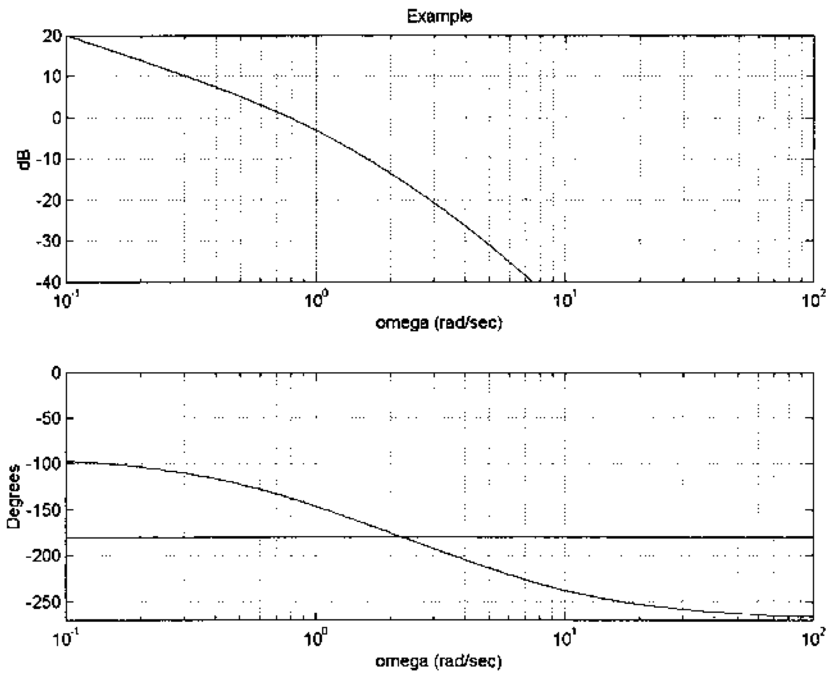


FIGURE 9.21 MATLAB Bode plots for $K = 5$.

The Bode plots for $K = 5$ and the Nyquist plots for $K = 5$ and 50 are shown in Figs. 9.21 and 9.22, respectively.

9.10 ADJUSTMENT OF GAIN BY USE OF THE LOG MAGNITUDE–ANGLE DIAGRAM (NICHOLS CHART)

The Lm–angle diagram for

$$\mathbf{G}_x(j\omega) = \frac{2.04(1 + j2\omega/3)}{j\omega(1 + j\omega)(1 + j0.2\omega)(1 + j0.2\omega/3)} \tag{9.49}$$

is drawn as the solid curve in Fig. 9.23 on graph paper that has the constant M and α contours. The $M = 1.12$ (1 dB) curve is tangent to the curve at $\omega_{m1} = 1.1$. These values are the maximum values of the control ratio M_m and the resonant frequency ω_m with the gain $K_x = 2.04$ given in Eq. (9.49). The designer specifies that the closed-loop system should be adjusted to produce an $M_m = 1.26$ (2 dB) by changing the gain. The dashed curve in Fig. 9.23 is obtained by raising the transfer function curve until it is tangent to the $M_m = 1.26$ (2 dB) curve. The resonant frequency is now equal to $\omega_{m2} = 2.09$. The curve has been raised by the amount Lm $A = 4.5$ dB, meaning that an additional gain of $A = 1.679$ must

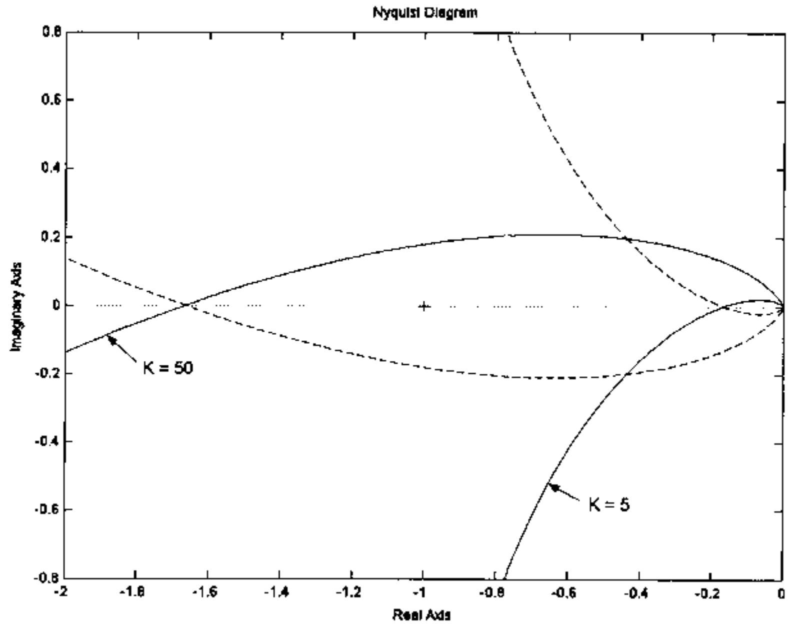


FIGURE 9.22 MATLAB Nyquist plots.

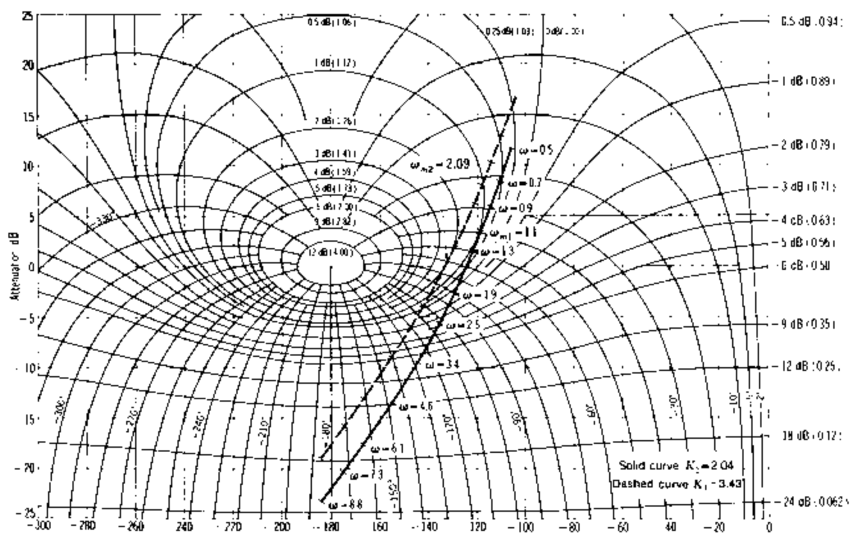


FIGURE 9.23 Log magnitude-angle diagram for Eq. (9.49).

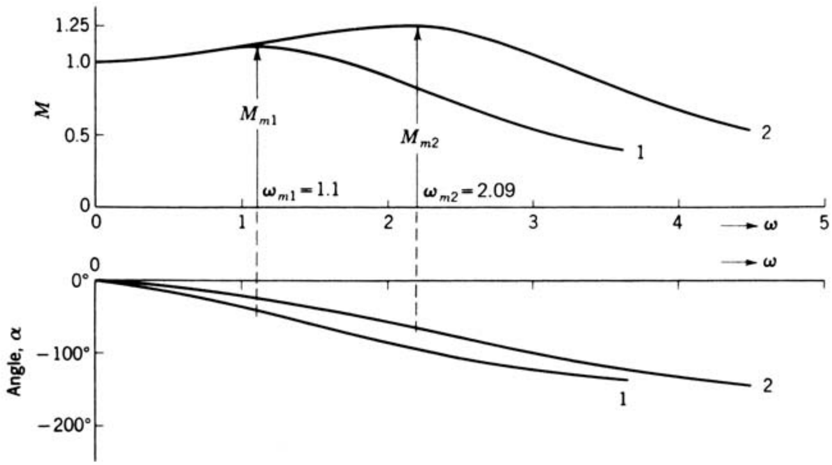


FIGURE 9.24 Control ratio vs. frequency obtained from the log magnitude-angle diagram of Fig. 9.23.

be added in cascade with $\mathbf{G}_x(j\omega)$. The gain is now equal to $K_1 = (1.679)(2.04) = 3.43$.

For any frequency point on the transfer function curve, the values of M and α can be read from the graph. For example, with M_m made equal to 1.26 in Fig. 9.23, the value of M at $\omega = 3.4$ is -0.5 dB and α is -110°. The closed loop frequency response, both Lm and angle, obtained from Fig. 9.23 is also plotted in Fig. 9.24 for both values of gain, $K_1 = 2.04$ and 3.43. The following MATLAB commands in **Sect9p10.m** illustrate the Nichols chart design process.

```
echo on
gxnum=2.04*[2/3 1];     % Create numerator
gxden=conv([1 1 0],conv([0.2 1],[0.2/3 1])); % Create denominator
Gx=tf(gxnum,gxden);% Create transfer function from num and den
zpk (Gx)                 % Show zero - pole - gain form
sisotool('nichols', Gx) % Open the SISO tool window
%
% The design is accomplished by setting a constraint of
% Closed-Loop Peak Gain<2 (value in dB) and moving the curve
% up to lie tangent to the constraint.
%
% See Figure C. 9     for a screen view of the SISO design tool
% in the 'nichols' mode. Axes and grid values are adjusted
% to obtain this view.
```

If the closed-loop response with the resultant gain adjustment has too low a resonant frequency for the desired performance, the next step is to compensate the system to improve the frequency response. Compensation is studied in detail in Chaps. 10 to 12.

9.11 CORRELATION OF POLE-ZERO DIAGRAM WITH FREQUENCY AND TIME RESPONSES

Whenever the closed-loop control ratio $M(j\omega)$ has the characteristic from shown in Fig. 9.5, the system may be approximated as a simple second-order system. This usually implies that the poles, other than the dominant complex pair, are either far to the left of the dominant complex poles or are close to zeros. When these conditions are not satisfied, the frequency response may have other shapes. This can be illustrated by considering the following three control ratios:

$$\frac{C(s)}{R(s)} = \frac{1}{s^2 + s + 1} \tag{9.50}$$

$$\frac{C(s)}{R(s)} = \frac{0.313(s + 0.8)}{(s + 0.25)(s^2 + 0.3s + 1)} \tag{9.51}$$

$$\frac{C(s)}{R(s)} = \frac{4}{(s^2 + s + 1)(s^2 + 0.4s + 4)} \tag{9.52}$$

The pole-zero diagram, the frequency response and the time response to a step input for each of these equations are shown in Fig. 9.25.

For Eq. (9.50) the following characteristics are noted from Fig. 9.25*a*:

1. The control ratio has only two complex dominant poles and no zeros.
2. The frequency-response curve has the following characteristics:
   (*a*) A single peak $M_m = 1.157$ at $\omega_m = 0.7$.
   (*b*) $1 < M < M_m$ in the frequency range $0 < \omega < 1$.
3. The time response has the typical waveform described in Chap. 3 for a simple second-order system. That is, the first maximum of $c(t)$ due to the oscillatory term is greater than $c(t)_{ss}$, and the $c(t)$ response after this maximum oscillates around the value of $c(t)_{ss}$.

For Eq. (9.51) the following characteristics are noted from Fig. 9.25*b*:

1. The control ratio has two complex poles and one real pole that are all dominant, and one real zero.
2. The frequency-response curve has the following characteristics:
   (*a*) A single peak, $M_m = 1.27$ at $\omega_m = 0.95$.
   (*b*) $M < 1$ in the frequency range $0 < \omega < \omega_x$.
   (*c*) The peak M_m occurs at $\omega_m = 0.95 > \omega_x$.

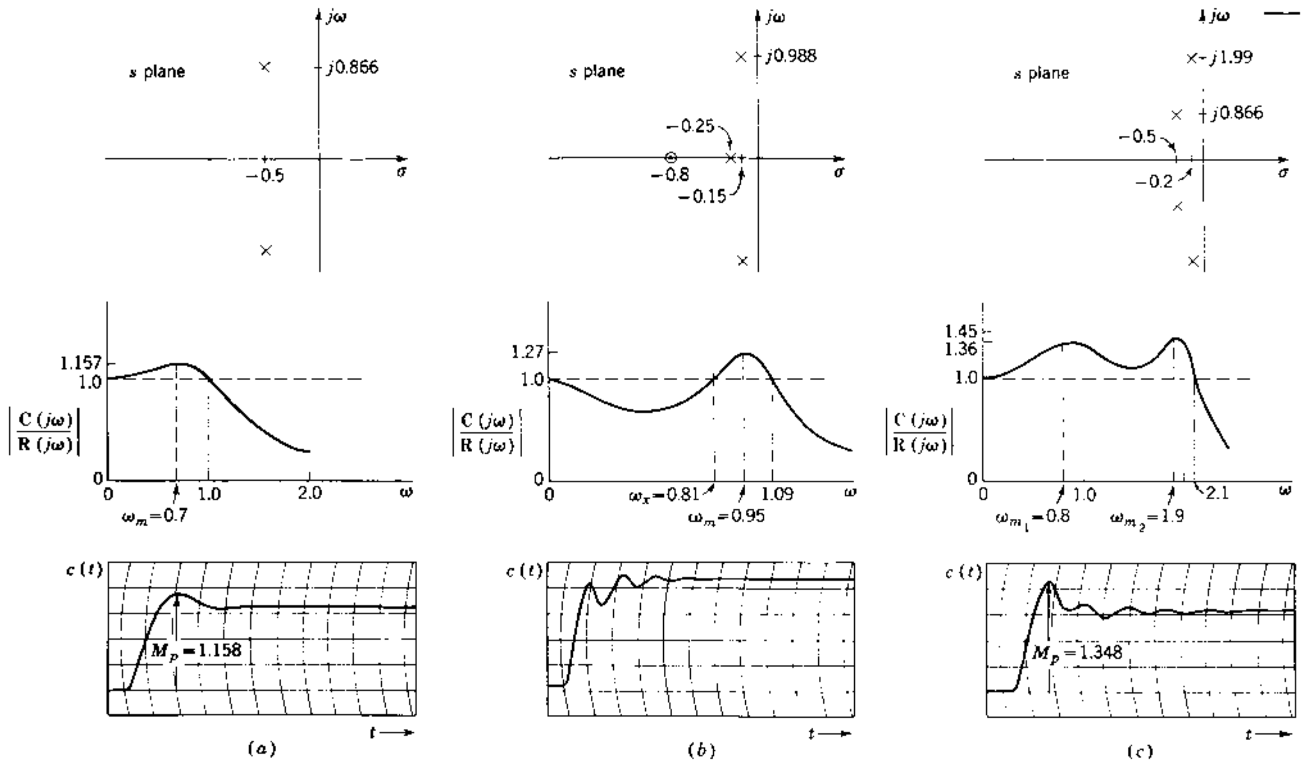


FIGURE 9.25 Comparison of frequency and time responses for three pole-zero patterns.

3. The time response does not have the conventional waveform. That is, the first maximum of $c(t)$ due to the oscillatory term is less than $c(t)_{ss}$ because of the transient term $A_3e^{-0.25t}$.

For Eq. (9.52) the following characteristics are noted from Fig. 9.25*c*:

1. The control ratio has four complex poles, all are dominant, and no zeros.
2. The frequency-response curve has the following characteristics:
   (*a*) There are two peaks, $M_{m1} = 1.36$ at $\omega_{m1} = 0.81$ and $M_{m2} = 1.45$ at $\omega_{m2} = 1.9$.
   (*b*) In the frequency range $0 < \omega < 2.1$, $1 < M < 1.45$.
   (*c*) The time response does not have the simple second-order waveform. That is, the first maximum of $c(t)$ in the oscillation is greater than $c(t)_{ss}$, and the oscillatory portion of $c(t)$ does not oscillate about a value of $c(t)_{ss}$. This time response can be predicted from the pole locations in the *s* plane and from the plot of M vs. ω (the two peaks).

Another example is the system represented by

$$M(s) = \frac{K}{(s - p_1)(s^2 + 2\zeta\omega_n s + \omega_n^2)} \tag{9.53}$$

When the real pole p_1 and real part of the complex poles are equal as shown in Fig. 9.26*a*, the corresponding frequency response is as shown in Fig. 9.26*b*. The magnitude at ω_m is less than unity. The corresponding time response to a step input is monotonic; i.e., there is no overshoot as shown in Fig. 9.26*c*. This time response may be considered as a critically damped response.

The examples discussed in this section show that the time-response waveform is closely related to the frequency response of the system. In other words, the time response waveform of the system can be predicted from the shape of the frequency response plot. This time-response correlation with the plot of $M = |\mathbf{C}(j\omega)/\mathbf{R}(j\omega)|$ vs. ω, when there is no distinct dominant complex pair of poles, is an important advantage of frequency-response analysis. Thus, as illustrated in this section, the frequency-response plot may be used as a guide in determining (or predicting) time-response characteristics.

9.12 SUMMARY

In summary, this chapter is devoted to the correlation between the frequency and time responses. The figures of merit M_m and ω_m, along with ω_0 and γ, are established as guideposts for evaluating a system's tracking performance. The addition of a pole to an open-loop transfer function produces a cw shift of the direct polar plot, which results in a larger value of M_m. The time response also

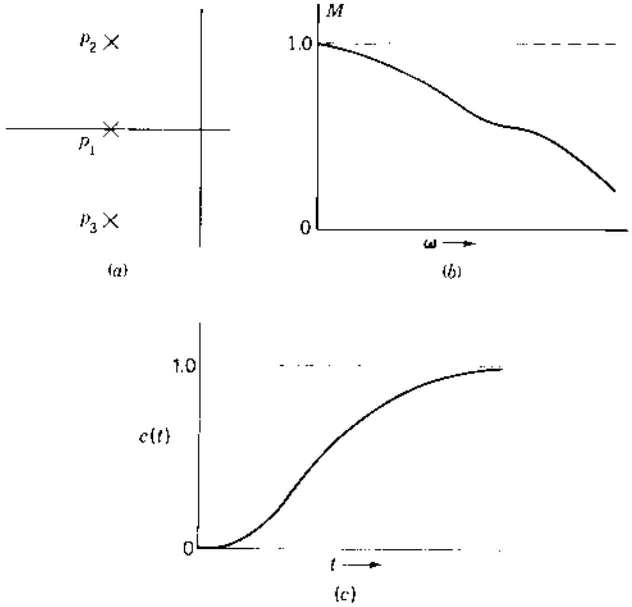


FIGURE 9.26 Form of the frequency and time responses for a particular pole pattern of Eq. (9.53).

suffers because ω_m becomes smaller. The reverse is true if a zero is added to the open-loop transfer function, This agrees with the root locus analysis, which shows that the addition of a pole or zero results in a less stable or more stable system, respectively. Thus, the qualitative correlation between the root locus and the frequency response is enhanced. The M and α contours are developed as an aid in obtaining the closed-loop frequency response and in adjusting the gain to obtain a desired M_m. The methods described for setting the gain for a desired M_m are based on the fact that generally the desired values of M_m are greater than 1 in order to obtain an underdamped response. When the gain adjustment does not yield a satisfactory value of ω_m, the system must be compensated in order to increase ω_m without changing the value M_m. Compensation procedures are covered in the following chapters.

The procedure used in the frequency-response method is summarized as follows:

Step 1. Derive the open-loop transfer function $G(s)H(s)$ of the system.
Step 2. Put the transfer function into the form $G(j\omega)H(j\omega)$.
Step 3. Arrange the various factors of the transfer function so that they are in the complex form, $j\omega$, $1 + j\omega T$, and $1 + aj\omega + b(j\omega)^2$.

Step 4. Plot the Lm and phase-angle diagram for $G(j\omega)H(j\omega)$. The graphical methods of Chap. 8 may be used, but a CAD program can provide more extensive data.

Step 5. Transfer the data from the plots in step 4 to either of the following; (*a*) Lm–angle diagram or (*b*) direct polar plot.

Step 6. Apply the Nyquist stability criterion and adjust the gain for the desired degree of stability M_m of the system. Then check the correlation to the time response for a step-input signal. This correlation reveals qualitative information about the time response.

Step 7. If the response does not meet the desired frequency and time response specifications, determine the shape that the plot must have to meet these specifications. Synthesize the compensator that must be inserted into the system. The procedure is developed in the following chapters.

A digital computer is used extensively in system design. Standard CAD programs like MATLAB and TOTAL-PC can be used to obtain the frequency response of both the open-loop and closed-loop transfer functions. The computer techniques are especially useful for systems containing frequency-sensitive feedback. The exact time response is then obtained to provide a correlation between the frequency and time responses. This chapter illustrates the use of the Nichols chart for analysing the tracking performances of unity-feedback systems. The Nichols chart can also be used for analysing the disturbance rejection performance of a nonunity-feedback system. This is discussed in detail in later chapters.

REFERENCES

1. Brown, G. S. and D. P. Campbell: *Principles of Servomechanisms*, Wiley, New York, 1948, Chaps. 6 and 8.
2. Higgins, T. J., and C. M. Siegel: Determination of the Maximum Modulus, or the Specified Gain of a Servomechanism by Complex Variable differentiation, *Trans. AIEE*, Vol. 72, pt II, pp. 467–469, January 1954.
3. Chu, Y.: Correlation between Frequency and Transient Responses of Feedback Control Systems, *Trans. AIEE*, Vol. 72, pt II, pp. 81–92, May 1953.
4. James. H. M., N. B. Nichols and R. S. Phillips: *Theory of Servomechanisms*, McGraw-Hill, New York, 1947, chap. 4.
5. D'Azzo, J, J, and C. H. Houpis: *Linear Control System Analysis and Design: Conventional and Modern*, 4th ed., McGraw-Hill, New York, 1995.

10

Root-Locus Compensation: Design

10.1 INTRODUCTION TO DESIGN

The preceding chapters deal with basic feedback control systems composed of the minimum equipment required to perform the control function and the necessary sensors and comparators to provide feedback. The designer first analyzes the performance of the basic system. This chapter presents methods of improving the design of a control system. To increase the reader's understanding of how poles and zeros affect the time response, a graphical means is presented for calculating the figures of merit of the system and the effect of additional significant nondominant poles. Typical pole-zero patterns are employed to demonstrate the correlation between the pole-zero diagram and the frequency and time responses. As a result of this analysis, the control system may be modified to achieve the desired time response. The next few chapters are devoted to achieving the necessary refinements.

Modifying a system to reshape its root locus in order to improve system performance is called *compensation* or stabilization. When the system is satisfactorily compensated: it is stable, has a satisfactory transient response, and has a large enough gain to ensure that the steady-state error does not exceed the specified maximum. Compensation devices may consist of electric networks or mechanical equipment containing levers, springs, dashpots, etc. The compensator (also called a filter) may be placed in cascade with the

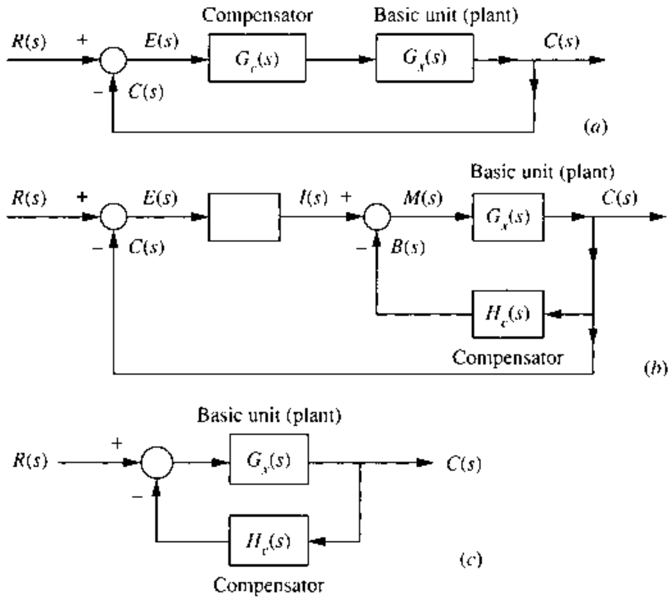


FIGURE 10.1 Block diagram showing the location of compensators: (*a*)cascade; (*b*) minor loop feedback; and (*c*) output feedback.

forward transfer function $G_x(s)$ (cascade or series compensation), as shown in Fig. 10.1*a*, or in the feedback path (feedback or parallel compensation), as shown in Fig. 10.1*b* and *c*. In Fig. 10.1*b* the compensation is introduced by means of a minor feedback loop. The selection of the location for inserting the compensator depends largely on the control system, the necessary physical modifications, and the results desired. The cascade compensator $G_c(s)$ is inserted at the low-energy point in the forward path so that power dissipation is very small. This also requires that G_x have a high input impedance. Isolation amplifiers may be necessary to avoid loading of or by the compensating network. The networks used for compensation are generally called lag, lead, and lag-lead compensation. The examples used for each of these compensators are not the only ones available but are intended primarily to show the methods of applying compensation. Feedback compensation is used in later design examples to improve the system's tracking of a desired input $r(t)$ or to improve the system's rejection of a disturbance input $d(t)$.

The following factors must be considered when making a choice between cascade and feedbeck compensation:

1. The design procedures for a cascade compensator $G_c(s)$ are more direct than those for a feedback compensator $H_c(s)$. Although the

design of feedback compensation is sometimes more laborious, it may be easier to implement.

2. The physical form of the control system, i.e., whether it is electrical, hydraulic, mechanical, etc., determines the ease of implementing a practical cascade or feedback (parallel) compensator. A microprocessor may also be used to implement the required compensation.
3. The economics, size, weight, and cost of components and amplifiers are important. In the forward path the signal goes from a low to a high energy level, whereas the reverse is true in the feedback loop. Thus, generally an amplifier may not be necessary in the feedback path. The cascade path generally requires an amplifier for gain and/or isolation. In applications such as aircraft and space-craft, minimum size and weight of equipment are essential.
4. Environmental conditions in which the feedback control system is to operate affect the accuracy and stability of the controlled quantity. For example, this is a critical problem in an airplane, which is subjected to rapid changes in altitude and temperature.
5. The problem of noise within a control system may determine the choice of compensator. This is accentuated in situations in which a greater amplifier gain is required with a forward compensator than by the use of feedback networks.
6. The time of response desired for a control system is a determining factor. Often a faster time of response can be achieved by the use of feedback compensation.
7. Some systems require 'tight-loop' stabilization to isolate the dynamics of one portion of a control system from other portions of the complete system. This can be accomplished by introducing an inner feedback loop around the portion to be isolated.
8. When $G_x(s)$ has a pair of dominant complex poles that yield the dominant poles of $C(s)/R(s)$, the simple first-order cascade compensators discussed in this chapter provide minimal improvement to the system's time-response characteristics. For this situation, as shown in Sec. 10.20, feedback compensation is more effective and is therefore more desirable.
9. The designer's experience and preferences influence the choice between a cascade and feedback compensator for achieving the desired performance.
10. The feedback compensator of Fig. 10.1*c* is used in Chap. 12 for disturbance rejection to minimize the response $c(t)$ to a disturbance input $d(t)$.

Meeting the system-performance specification often requires the inclusion of a properly designed compensator. The design process is normally performed by successively changing the compensator parameter(s) and comparing the resulting system performance. This process is expedited by the use of interactive CAD packages like MATLAB or TOTAL-PC. The availability of a graphics capability in a CAD program permits the designer to see the changes in system performance in real time. An interactive capability allows the designer to try a number of compensator designs quickly in order to select a design that best meets the system-performance requirements. The use of an interactive CAD package is standard practice for control system designers.

10.2 TRANSIENT RESPONSE: DOMINANT COMPLEX POLES [1]

The root-locus plot permits the designer to select the best poles for the control ratio. The criteria for determining which poles are best must come from the specifications of system performance and from practical considerations. For example, the presence of nonlinear characteristics like backlash, dead zone, and Coulomb friction can produce a steady-state error with a step input, even though the linear portion of the system is Type 1 or higher. The response obtained from the pole-zero locations of the control ratio does not show the presence of a steady-state error caused by the nonlinearities.* Frequently the system gain is adjusted so that there is a dominant pair of complex poles, and the response to a step input has the underdamped form shown in Fig. 10.2*a*. In this case the transient contributions from the other poles must be small. From Eq. (4.68), the necessary conditions for the time response to be dominated by

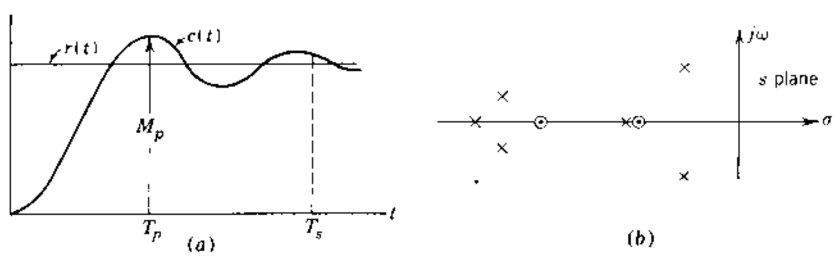


FIGURE 10.2 Transient response to a step input with dominant complex poles. (*b*) Pole-zero pattern of $C(s)/R(s)$ for the desired response.

*For a detailed study of the effect of nonlinear characteristics on system performance, the reader is referred to the literature [2,3].

one pair of complex poles require the pole-zero pattern of Fig. 10.2*b* and have the following characteristics:

1. The other poles must be far to the left of the dominant poles so that the transients due to these other poles are small in amplitude and die out rapidly.
2. Any other pole that is not far to the left of the dominant complex poles must be near a zero so that the magnitude of the transient term due to that pole is small.

With the system designed so that the response to a unit step input has the under-damped form shown in Fig. 10.2*a*, the following transient figures of merit (described in Secs. 3.9 and 3.10) are used to judge its performance:

1. M_p, peak overshoot, is the amplitude of the first overshoot.
2. T_p, peak time, is the time to reach the peak overshoot.
3. T_s, settling time, is the time for the response envelope first to reach and thereafter remain within 2 percent of the final value.
4. N is the number of oscillations in the response up to the settling time.

Consider the nonunity-feedback system shown in Fig. 10.3. using

$$G(s) = \frac{N_1}{D_1} = \frac{K_G \prod (s - z_k)}{\prod (s - p_g)} \tag{10.1}$$

$$H(s) = \frac{N_2}{D_2} = \frac{K_H \prod (s - z_j)}{\prod (s - p_i)} \tag{10.2}$$

$$G(s)H(s) = \frac{N_1 N_2}{D_1 D_2} = \frac{K_G K_H \prod_{h=1}^{w} (s - z_h)}{\prod_{c=1}^{n} (s - p_c)} \tag{10.3}$$

The product $K_G K_H = K$ is defined as the *static loop sensitivity.* For $G(s)H(s)$ the degree of the numerator is w and the degree of the denominator is n. These symbols are used throughout Chap. 6. The control ratio is

$$\frac{C(s)}{R(s)} = \frac{P(s)}{Q(s)} = \frac{N_1 D_2}{D_1 D_2 + N_1 N_2} = \frac{K_G \prod_{m=1}^{w'} (s - z_m)}{\prod_{k=1}^{n} (s - p_k)} \tag{10.4}$$

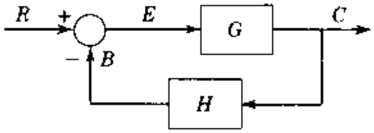


FIGURE 10.3 Feedback system.

Note that the constant K_G in Eq. (10.4) is not the same as the *static loop sensitivity* unless the system has unity feedback. The degree n of the denominator of $C(s)/R(s)$ is the same as for $G(s)H(s)$, regardless of whether the system had unity or nonunity feedback. *The degree w' of the numerator of $C(s)/R(s)$ is equal to the sum of the degrees of N_1 and D_2.* For a unity-feedback system the degree of the numerator is $w' = w$. For a nonunity-feedback system the w' zeros of Eq. (10.4) include both the zeros of $G(s)$, i.e., N_1 and the poles of $H(s)$, i.e., D_2, as shown in Eq. (10.4).

For a unit step input the output of Fig. 10.3 is

$$
\begin{aligned}
C(s) &= \frac{P(s)}{sQ(s)} = \frac{K_G \prod_{m=1}^{w'} (s - z_m)}{s \prod_{k=1}^{n} (s - p_k)} \\
&= \frac{A_0}{s} + \frac{A_1}{s - p_1} + \cdots + \frac{A_k}{s - p_k} + \cdots + \frac{A_n}{s - p_n}
\end{aligned}
\tag{10.5}
$$

The poles $s = 0$ that comes from $R(s)$ must be included as a pole of $C(s)$, as shown in Eq. (10.5). The values of the coefficients can be obtained graphically from the pole-zero diagram by the method described in Sec. 4.11 and Eq. (4.68) or by the more practical use of a CAD program. Assume that the system represented by Eq. (10.5) has a dominant complex pole $p_1 = \sigma \pm j\omega_d$. The complete time solution is

$$
\begin{aligned}
c(t) = \frac{P(0)}{Q(0)} &+ 2\left| \frac{K_G \prod_{m=1}^{w'} (p_1 - z_m)}{P_1 \prod_{k=2}^{n} (p_1 - p_k)} \right| e^{\sigma t} \\
&\times \cos\left[\omega_d t + \angle P(p_1) - \angle p_1 - \angle Q'(p_1) \right] \\
&+ \sum_{k=3}^{n} \left[\frac{P(p_k)}{p_k Q'(p_k)} \right] e^{p_k t}
\end{aligned}
\tag{10.6}
$$

where

$$
Q'(p_k) = \left[\frac{dQ(s)}{ds} \right]_{s=p_k} = \left[\frac{Q(s)}{s - p_k} \right]_{s=p_k}
\tag{10.7}
$$

By assuming the pole-zero pattern of Fig. 10.2, the last term of Eq. (10.6) may often be neglected. The time response is therefore approximated by

$$
\begin{aligned}
c(t) \approx \frac{P(0)}{Q(0)} &+ 2\left| \frac{K_G \prod_{m=1}^{w'} (p_1 - z_m)}{p_1 \prod_{k=2}^{n} (p_1 - p_k)} \right| e^{\sigma t} \\
&\times \cos\left[\omega_d t + \angle P(p_1) - \angle p_1 - \angle Q'(p_1) \right]
\end{aligned}
\tag{10.8}
$$

Note that, although the transient terms due to the other poles have been neglected, the effect of those poles on the amplitude and phase angle of the dominant transient has not been neglected.

The peak time T_p is obtained by setting the derivative with respect to time of Eq. (10.8) equal to zero. This gives

$$\begin{aligned} T_p &= \frac{1}{\omega_d}\left[\frac{\pi}{2} - \angle P(p_1) + \angle Q'(p_1)\right] \\ &= \frac{1}{\omega_d}\left\{\frac{\pi}{2} - [\text{sum of angles from zeros of } C(s)/R(s) \text{ to dominant pole } p_1]\right. \\ &\quad \left. + \begin{bmatrix} \text{sum of angles from all other poles of } C(s)/R(s) \text{ to} \\ \text{dominant pole } p_1\text{, including conjugate pole} \end{bmatrix}\right\} \end{aligned} \tag{10.9}$$

This value of T_p is more accurate than that given by Eq. (3.60) for a second-order system. Since the phase angle $\phi = \angle P(p_1) - \angle p_1 - \angle Q'(p_1)$ in Eq. (10.8) cannot have a value greater than 2π, the value T_p must occur within one "cycle" of the transient. Inserting this value of T_p into Eq. (10.8) gives the peak overshoot M_p. By using the value $\cos(\pi/2 - \angle p_1) = \omega_d/\omega_n$, the value M_p can be expressed as

$$M_p = \frac{P(0)}{Q(0)} + \frac{2\omega_d}{\omega_n^2}\left|\frac{K_G \prod_{m=1}^{w'}(p_1 - z_m)}{\prod_{k=2}^{n}(p_1 - p_k)}\right| e^{\sigma T_p} \tag{10.10}$$

The first term in Eq. (10.10) represents the final value, and the second term represents the overshoot M_o. For a unity-feedback system that is Type 1 or higher, there is zero steady-state error when the input is a step function. Therefore, the first term on the right side of Eq. (10.8) is equal to unity. Its magnitude can also be obtained by applying the final-value theorem to $C(s)$ given by Eq. (10.5). This equality is used to solve for K_G. Note that under these conditions $K_G = K$ and $w' = w$:

$$K_G = \frac{\prod_{k=1}^{n}(-p_k)}{\prod_{m=1}^{w}(-z_m)} \tag{10.11}$$

The value of the overshoot M_o can therefore be expressed as

$$M_o = \frac{2\omega_d}{\omega_n^2}\left|\frac{\prod_{k=1}^{n}(-p_k)\prod_{m=1}^{w}(p_1 - z_m)}{\prod_{m=1}^{w}(-z_m)\prod_{k=2}^{n}(p_1 - p_k)}\right| e^{\sigma T_p} \tag{10.12}$$

Some terms inside the brackets in Eq. (10.12) cancel the terms in front so that M_o can be expressed in words as

$$M_o = \left[\frac{\begin{array}{l}\text{product of distances from all poles}\\ \text{of } C(s)/R(s) \text{ to origin, excluding}\\ \text{distances of two dominant poles from}\\ \text{origin}\end{array}}{\begin{array}{l}\text{product of distances from all other poles}\\ \text{of } C(s)/R(s) \text{ to dominant pole } p_1,\\ \text{excluding distance between dominant}\\ \text{poles}\end{array}} \right] \times \left[\frac{\begin{array}{l}\text{product of distances}\\ \text{from all zeros of}\\ C(s)/R(s) \text{ to dominant}\\ \text{pole } p_1\end{array}}{\begin{array}{l}\text{product of distances}\\ \text{from all zeros of}\\ C(s)/R(s) \text{ to origin}\end{array}} \right] e^{\sigma T_p} \tag{10.13}$$

The denominator terms of Eq. (10.13) clearly reveal that as a zero of Eq. (10.4) is *located closer to the imaginary axis, the magnitude of M_o increases.* If there are no finite zeros of $C(s)/R(s)$, the factors in M_o involving zeros become unity. Equation (10.13) is valid only for a unity-feedback system that is Type 1 or higher. It may be sufficiently accurate for a Type 0 system that has a large value for K_0. The value of M_o can be calculated either from the right-hand term of Eq. (10.10) or from Eq. (10.13), whichever is more convenient. The effect on M_o of other poles, which cannot be neglected, is discussed in the next section.

The values of T_s and N can be approximated from the dominant roots. Section 3.9 shows that T_s is four time constants for 2 percent error:

$$T_s = \frac{4}{|\sigma|} = \frac{4}{\zeta\omega_n} \tag{10.14}$$

$$N = \frac{\text{settling time}}{\text{period}} = \frac{T_s}{2\pi/\omega_d} = \frac{2\omega_d}{\pi|\sigma|} = \frac{2}{\pi}\frac{\sqrt{1-\zeta^2}}{\zeta} \tag{10.15}$$

Equation (10.9) shows that zeros of the control ratio cause a decrease in the peak time T_p, whereas poles increase the peak time. Peak time can also be decreased by shifting zeros to the right or poles (other than the dominant poles) to the left. Equation (10.13) shows that the larger the value of T_p, the smaller the value of M_o because $e^{\sigma T_p}$ decreases. M_o can also be decreased by reducing the ratios $p_k/(p_k-p_1)$ and $(z_m-p_1)/z_m$. But this reduction can have an adverse effect on T_p. The conditions on pole-zero locations that lead to a small M_o may therefore produce a large T_p. Conversely, the conditions that lead to a small T_p may be obtained at the expense of a large M_o. Thus, a compromise is required in the peak time and peak overshoot that are attainable. Some improvements can be obtained by introducing additional poles and zeros into the system and locating them appropriately. This topic is covered in the following sections on compensation.

The approximate equations, given in this section, for the response of a system to a step-function input are based on the fundamental premise that there is a pair of dominant complex poles and that the effect of other poles is small. When this premise is satisfied, a higher-order system, $n > 2$, effectively acts like a simple second-order system. These approximate equations are presented primarily to enhance the reader's understanding of how poles and zeros affect the time response. A CAD program can be used to calculate and plot $c(t)$ and to yield precise values of M_p, t_p, and t_s.

10.3 ADDITIONAL SIGNIFICANT POLES [4]

When there are two dominant complex poles, the approximations developed in Sec. 10.2 give accurate results. However, there are cases where an additional pole of $C(s)/R(s)$ is significant. Figure 10.4*a* shows a pole-zero diagram which contains dominant complex poles and an additional real pole p_3. The control ratio, with $K = -\omega_n^2 p_3$, is given by

$$\frac{C(s)}{R(s)} = \frac{K}{(s^2 + 2\zeta\omega_n s + \omega_n^2)(s - p_3)} \tag{10.16}$$

With a unit-step input the time response is

$$c(t) = 1 + 2|A_1|e^{-\zeta\omega_n t} \sin\left(\omega_n\sqrt{1-\zeta^2 t} + \phi\right) + A_3 e^{p_3 t} \tag{10.17}$$

The transient term due to the real pole p_3 has the form $A_3 e^{p_3 t}$ where A_3, based upon Eq. (4.68) is *always negative*. Thus, the overshoot M_p is reduced, and settling time t_s may be increased or decreased. This effect is typical of an additional real pole. The magnitude A_3 depends on the location of p_3 relative to the complex poles. The further to the left the pole p_3 is located, the smaller the magnitude of A_3, therefore the smaller its effect on the total response. A pole that is six times as far to the left as the complex poles has negligible effect on the time response. The typical time response is shown in Fig. 10.4*a*. As the pole p_3 moves to the right, the magnitude of A_3 increases and the overshoot becomes smaller. As p_3 approaches but is still to the left of the complex poles, the first maximum in the time response is less than the final value.

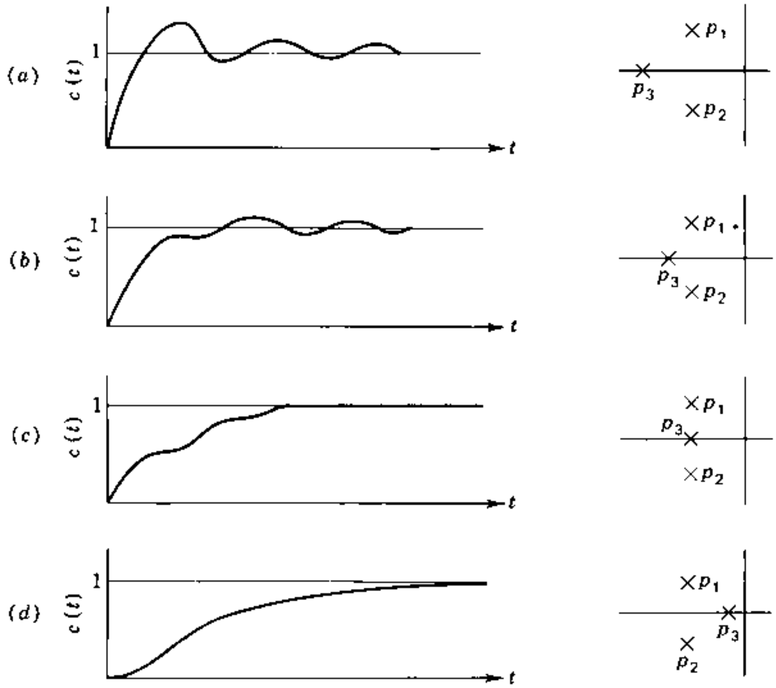


FIGURE 10.4 Typical time response as a function of the real-pole location.

The largest overshoot can occur at the second or a later maximum as shown in Fig. 10.4*b* (see also Fig. 9.25*b*).

When p_3 is located at the real-axis projection of the complex poles, the response is monotonic; i.e., there is no overshoot. This represents the critically damped situation, as shown in Fig. 10.4*c*. The complex poles may contribute a "ripple" to the time response, as shown in the figure. When p_3 is located to the right of the complex poles, it is the dominant pole and the response is overdamped.

When the real pole p_3 is to the left of the complex poles, the peak time T_p is approximately given by Eq. (10.9). Although the effect of the *real* pole is to increase the peak time, this change is small if the real pole is fairly far to the left. A first-order correction can be made to the peak overshoot M_o given by Eq. (10.13) by adding the value of $A_3 e^{p_3 T_p}$ to the peak overshoot due to the complex poles. The effect of the real pole on the actual settling time t_s can be estimated by calculating $A_3 \exp(p_3 T_s)$ and comparing it with the size and sign of the underdamped transient at time T_s obtained from Eq. (10.14). If both are negative at T_s, then the true settling time t_s is increased. If they have opposite

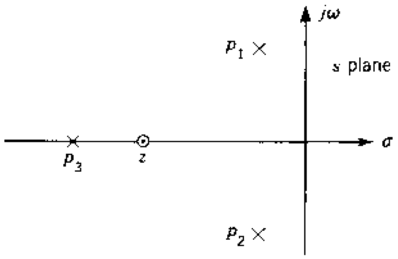


FIGURE 10.5 Pole-zero diagram of $C(s)/R(s)$ for Eq. (10.18).

signs, then the value $T_s = 4T$ based on the complex roots is a good approximation for t_s.

The presence of a real zero in addition to the real pole further modifies the transient response. A possible pole-zero diagram for which the control ratio is

$$\frac{C(s)}{R(s)} = \frac{K(s-z)}{(s^2 + 2\zeta\omega_n s + \omega_n^2)(s-p_3)} \tag{10.18}$$

is shown in Fig. 10.5. The complete time response to a unit step-function input still has the form given in Eq. (10.17). However, the sign of A_3, based upon Eq. (4.68), depends on the relative location of the real pole and the real zero. A_3 is negative if the zero is to the left of p_3, and it is positive if the zero is to the right of p_3. Also, the magnitude of A_3 is proportional to the distance from p_3 to z [see Eq. (4.68)]. Therefore, if the zero is close to the pole, A_3 is small and the contribution of this transient term is correspondingly small. Compared with the response for Eq. (10.16), as shown in Fig. 10.4, note the following:

1. If the zero z is to the left of the real pole p_3, the response is qualitatively the same as that for a system with only complex poles; but the peak overshoot is smaller. Depending on the ζ of the dominant complex poles, either the response of Fig. 10.4*a* or 10.4*b* may be obtained.
2. If the zero z is to the right of the real pole p_3, the peak overshoot is greater than that for a system with only complex poles.

These characteristics are evident in the example in Fig. 10.20.

When the real pole p_3 and the real zero z are close together, the value of peak time T_p obtained from Eq. (10.9) can be considered essentially correct. Actually T_p decreases if the zero is to the right of the real pole, and vice versa. A first-order correction to M_o from Eq. (10.13) can be obtained by adding the contribution of $A_3 e^{p_3 t}$ at the time T_p. The analysis of the effect on t_s is similar to that described previously for the case of just an additional real pole.

As a summary, many control systems having an underdamped response can be approximated by one having the following characteristics: (1) two complex poles; (2) two complex poles and one real pole; and (3) two complex poles, one real pole, and one real zero. For case 1 the relations T_p, M_o, T_s, and N developed in Sec. 10.2 give an accurate representation of the time response. For case 2 these approximate values can be corrected if the real pole is far enough to the left. Then the contribution of the additional transient term is small, and the total response remains essentially the sum of a constant and an underdamped sinusoid. For case 3 the approximate values can be corrected provided the zero is near the real pole so that the amplitude of the additional transient term is small. More exact calculation of the figures of merit can be obtained by plotting the exact response as a function of time by use of a CAD program.

10.4 ROOT-LOCUS DESIGN CONSIDERATIONS

The root-locus design technique presented in Chap. 7 stresses the desired transient figures of merit. Thus, the selection of the dominant complex root can be based on a desired damping ratio ζ_D (which determines the peak overshoot), the desired damped natural frequency ω_D (which is the frequency of oscillation of the transient), or the desired real part σ_d of the dominant root (which determines the settling time). Two additional factors must be included in the design. First, the presence of additional poles will modify the peak overshoot, as described in Sec. 10.3. Second, it may be necessary to specify the value of the gain K_m. For example, in a Type 1 system it mat be necessary to specify a minimum value of K_1 in order to limit the magnitude of the steady-state error e_{ss} with a ramp input, where $e(t)_{ss} = Dr(t)/K_1$. The following example incorporates these factors into a root-locus design.

Example. Consider a unity-feedback system where

$$\begin{aligned} G_x(s) &= \frac{K_x}{s(s^2 + 4.2s + 14.4)} \\ &= \frac{K_x}{s(s + 2.1 + j3.1607)(s + 2.1 - j3.1607)} \end{aligned} \tag{10.19}$$

and $K_1 = K_x/14.4$. For this system it is specified that the following figures of merit must be satisfied: $1 < M_p \leq 1.123$, $t_s \leq 3\,\text{s}$, $e(t)_{ss} = 0$, $t_p \leq 1.6\,\text{s}$, $K_1 \geq 1.5\,\text{s}^{-1}$.

First Design

With the assumption that the closed-loop system can be represented by an effective simple second-order model, Eqs. (3.60), (3.61), and (10.14) can be

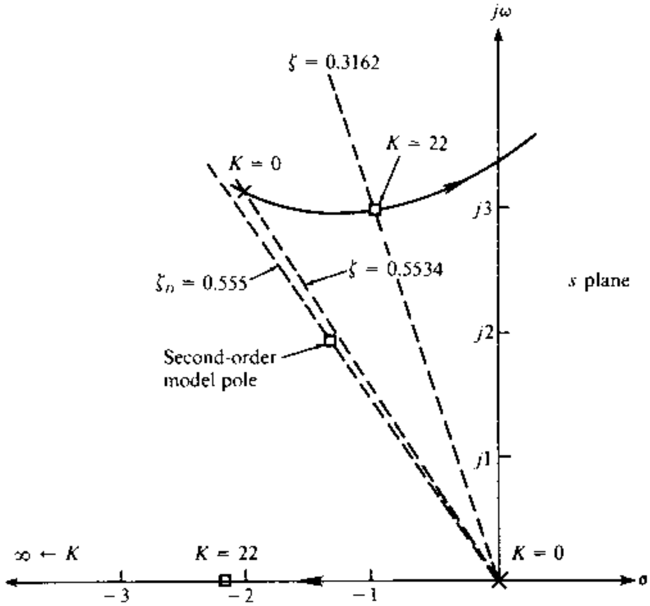


FIGURE 10.6 Root locus for Eq. (10.19).

used to determine the required values of ζ_D, ω_d, and ω_n:

$$M_p = 1.123 = 1 + \exp\frac{-\zeta\pi}{\sqrt{1-\zeta^2}} \qquad \rightarrow \zeta_D = 0.555$$

$$t_p = \frac{\pi}{\omega_d} \qquad \rightarrow \omega_d > 1.9635$$

$$T_s = \frac{4}{\zeta_D \omega_n} \qquad \rightarrow \omega_n > 2.4024$$

Using the values $\zeta_D = 0.555$, $\omega_n = 2.4024$, and $\omega_d = 1.9984$, the required dominant complex roots are $s_{1,2} = -1.3333 \pm j1.9984$. Figure 10.6 contains the root locus for $G_x(s)$ and the line representing $\zeta_D = 0.555$. It is evident that the required dominant roots are not achievable.

Second Design

The first design does not consider the fact that there is a third root s_3. The associated transient $A_3 e^{s_3 t}$ is negative (see Sec. 10.3), and it will reduce the peak overshoot of the underdamped transient associated with the complex roots $s_{1,2}$. Therefore, it is possible to select a smaller value of ζ for the complex roots, and the effect of the third root can be used to keep M_p within the specified limits. In order simultaneously to satisfy the specification for $K_1 > 1.5$, the design is therefore based on obtaining the complex roots $s_{1,2} = -1 \pm j3$ on the root locus. The corresponding sensitivity is $K = 14.4$ $K_1 = 22$, which yields $K_1 = 1.528$. The closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{22}{(s+1+j3)(s+1-j3)(s+2.2)} \tag{10.20}$$

The plot of $|C(j\omega)/R(j\omega)|$ vs. ω very closely approximates the frequency response for a simple second-order underdamped system shown in Fig. 9.25*a*. A unit-step input results in the following values: $M_p \approx 1.123$, $t_p \approx 1.51$ s, and $t_s \approx 2.95$ s. Therefore, the desired specifications have been achieved.

10.5 RESHAPING THE ROOT LOCUS

The root-locus plots described in Chap. 7 show the relationship between the gain of the system and the time response. Depending on the specifications established for the system, the gain that best achieves the desired performance is selected. The performance specifications may be based on the desired damping ratio, undamped natural frequency, time constant, or steady-state error. The root locus may show that the desired performance cannot be achieved just by adjustment of the gain. In fact, in some cases the system may be unstable for all values of gain. The control-systems engineer must then investigate the methods for reshaping the root locus to meet the performance specifications.

The purpose of reshaping the root locus generally falls into one of the following categories:

1. A given system is stable and its transient response is satisfactory, but its steady-state error is too large. Thus, the gain must be increased to reduce the steady-state error (see Chap. 6). This increase must be accomplished without appreciably reducing ths system stability.
2. A given system is stable, but its transient response is unsatisfactory. Thus, the root locus must be reshaped so that it is moved farther to the left, away from the imaginary axis.
3. A given system is stable, but both its transient response and its steady-state response are unsatisfactory. Thus, the locus must be moved to the left and the gain must be increased.
4. A given system is unstable for all values of gain. Thus, the root locus must be reshaped so that part of each branch falls in the left-half s plane, thereby making the system stable.

Compensation of a system by the introduction of poles and zeros is used to improve the operating performance. However, each additional compensator pole increases the number of roots of the closed-loop characteristic equation. If an underdamped response of the form shown in Fig. 10.2*a* is desired, the system gain must be adjusted so that there is a pair of dominant complex poles. This requires that any other pole be far to the left or near a zero so that its transient response has a small amplitude and therefore has a small effect on the total time response. The required pole-zero diagram is shown in Fig 10.2*b*. The approximate values of peak overshoot M_o, peak time T_p, settling time T_s, and the number of oscillations up to settling time N can be obtained from the pole-zero pattern as described in Sec. 10.2. The effect of compensator poles and zeros on these quantities can be evaluated rapidly by use of a CAD program. (see Appendixes C and D)

10.6 CAD ACCURACY CHECKS (CADAC)

The CADAC given in Secs. 3.11, 4.5, 6.8, 7.9, and 8.10 should be used, when applicable, in doing the problems associated with this chapter and the remaining chapters of this text. Each of these CADAC should be invoked, when applicable, at the appropriate point during the utilization of a CAD program.

10.7 IDEAL INTEGRAL CASCADE COMPENSATION (PI CONTROLLER)

When the transient response of a feedback control system is considered satisfactory but the steady-state error is too large, it is possible to eliminate the error by increasing the system type. This change must be accomplished *without appreciably changing the dominant roots of the characteristic equation*. The system type can be increased by operating on the actuating signal e to produce one that is proportional to both the magnitude and the integral of this signal. This proportional plus integral (PI) controller is shown in Fig. 10.7 where

$$E_1(s) = \left(1 + \frac{K_i}{s}\right)E(s) \tag{10.21}$$

$$G_c(s) = \frac{E_1(s)}{E(s)} = \frac{s + K_i}{s} \tag{10.22}$$

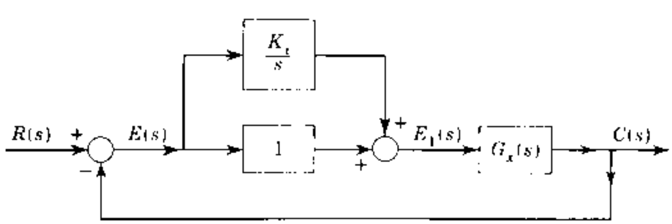


FIGURE 10.7 Ideal proportional plus integral (PI) control.

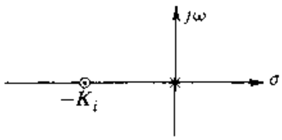


FIGURE 10.8 Location of the pole and zero of an ideal proportional plus integral compensator.

In this system the quantity $e_1(t)$ continues to increase as long as an error $e(t)$ is present. Eventually $e_1(t)$ becomes large enough to produce an output $c(t)$ equal to the input $r(t)$. The error $e(t)$ is then equal to zero. The constant K_i (generally very small) and the overall gain of the system must be selected to produce satisfactory roots of the characteristic equation. Since the system type has been increased, the corresponding error coefficient is equal to infinity. Provided that the new roots of the characteristic equation can be satisfactorily located, the transient response is still acceptable.

The locations of the pole and zero of this ideal proportional plus integral compensator are shown in Fig. 10.8. The pole alone would move the root locus to the right, thereby slowing the time response. The zero must be near the origin in order to minimize the increase in response time of the complete system. The required integrator can be achieved by an electronic circuit. Mechanically the integral signal can be obtained by an integrating gyroscope, which is used in vehicles such as aircraft, space vehicles, ships, and submarines where the improved performance justifies the cost. Frequently, however a passive electric network consisting of resistors and capacitors sufficiently approximates the proportional plus integral action, as illustrated in the next section.

10.8 CASCADE LAG COMPENSATION DESIGN USING PASSIVE ELEMENTS

Figure 10.9*a* shows a network that approximates a proportional plus integral output and is used as a lag compensator. Putting an amplifier of gain A in

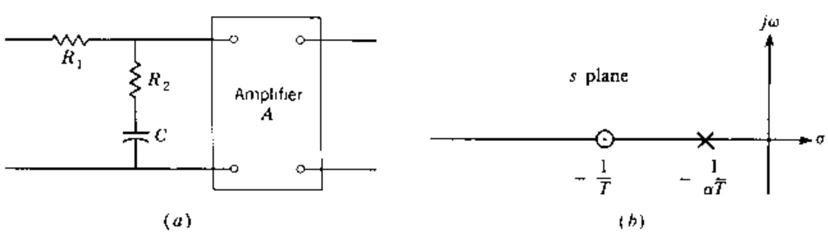


FIGURE 10.9 (*a*) Integral or lag compensator; (*b*) pole and zero location.

series with this network yields the transfer function

$$G_c(s) = A\frac{1 + Ts}{1 + \alpha Ts} = \frac{A}{\alpha}\frac{(s + 1/T)}{(s + 1/\alpha T)} = \frac{A}{\alpha}\frac{(s - z_c)}{(s - p_c)} \tag{10.23}$$

where $\alpha = (R_1 + R_2)/R_2 > 1$ and $T = R_2C$. The pole $p_c = -1/\alpha T$ is therefore to the right of the zero $z_c = -1/T$, as shown in Fig. 10.9*b*. The locations of the pole p_c and the zero z_c of $G_c(s)$ on the s plane can be made close to those of the ideal compensator. Besides furnishing the necessary gain, the amplifier also acts as an isolating unit to prevent any loading effects between the compensator and the original system.

Assume that the original forward transfer function is

$$G_x(s) = \frac{K\prod_{h=1}^{w}(s - z_h)}{\prod_{g=1}^{n}(s - p_g)} \tag{10.24}$$

when the closed-loop root s_1 is selected on the root locus for the original system, the loop sensitivity K is

$$K = \frac{\prod_{g=1}^{n}|s - p_g|}{\prod_{h=1}^{w}|s - z_h|} \tag{10.25}$$

With the addition of the lag compensator in cascade, the new forward transfer function is

$$G(s) = G_c(s)G_x(s) = \frac{AK}{\alpha}\frac{(s + 1/T)}{(s + 1/\alpha T)}\frac{\prod_{h=1}^{w}(s - z_h)}{\prod_{g=1}^{n}(s - p_g)} \tag{10.26}$$

When the desired roots of the charateristic equation are located on the root locus, the magnitude of the loop sensitivity for the new root s' becomes

$$K' = \frac{AK}{\alpha} = \frac{|s' + 1/\alpha T|}{|s' + 1/T|}\frac{\prod_{g=1}^{n}|s' - p_g|}{\prod_{h=1}^{w}|s' - z_h|} \tag{10.27}$$

As an example, a lag compensator is applied to a Type 0 system. To improve the steady-state accuracy, the error coefficient must be increased. The value of K_0 before and after the addition of the compensator is calculated by using the definition $K_0 = \lim_{s\to 0} G(s)$ from Eqs. (10.24) and (10.26), respectively,

$$K_0 = \frac{\prod_{h=1}^{w}(-z_h)}{\prod_{g=1}^{n}(-p_g)}K \tag{10.28}$$

$$K_0' = \frac{\prod_{h=1}^{w}(-z_h)}{\prod_{g=1}^{n}(-p_g)}\alpha K' \tag{10.29}$$

The following procedure is used to design the passive lag cascade compensator. First, the pole $p_c = -1/\alpha T$ and the zero $z_c = -1/T$ of the compensator are placed very close together. This step means that most of the original root locus remains practically unchanged. If the angle contributed by the compensator *at the original closed-loop dominant root* is less than 5°, the new locus is displaced only slightly. This 5° is only a guide and should not be applied arbitrarily. The new closed-loop dominant pole s' is therefore only slightly changed from the uncompensated value. This new value satisfies the restriction that the transient response must not change appreciably. As a result, the values $p'_c + 1/\alpha T$ and $z'_c + 1/T$ are almost equal, and the values K and K' in Eqs. (10.25) and (10.27) are approximately equal. The values K_0 and K'_0 in Eqs. (10.28) and (10.29) now differ only by the factor α so that $K'_0 \approx \alpha K_0$. The gain required to produce the new root s' therefore increases approximately by the factor α, which is the ratio of the compensator zero and pole. Summarizing, the necessary conditions on the compensator are that (1) the pole and zero must be close together and (2) the ratio α of the zero and pole must approximately equal the desired increase in gain. These requirements can be achieved by placing the compensator pole and zero very close to the origin. The size of α is limited by the physical parameters required in the network. A value $\alpha = 10$ is often used.

Although the preceding statements are based on a Type 0 system, the same conditions apply equally well for a Type 1 or higher system.

DESIGN EXAMPLE OF LAG COMPENSATION APPLIED TO A TYPE 1 SYSTEM. A control system with unity feedback has the forward transfer function

$$G_x(s) = \frac{K_1}{s(1+s)(1+0.2s)} = \frac{K}{s(s+1)(s+5)} \tag{10.30}$$

that yields the root locus shown in Fig. 10.10. For the basic control system a damping ratio $\zeta = 0.45$ yields the following pertinent data:

Dominant roots: $s_{1,2} = -0.404 \pm j0.802$

Static loop sensitivity:
$$\begin{aligned} K &= |s_1| \cdot |s_1 + 1| \cdot |s_1 + 5| \\ &= |-0.404 + j0.802| \cdot |0.596 + j0.802| \\ &\quad \cdot |4.596 + j0.802| = 4.188 \end{aligned}$$

Ramp error coefficient: $K_1 = \lim_{s\to 0} sG(s) = \dfrac{K}{5} = \dfrac{4.188}{5} = 0.838\ \mathrm{s}^{-1}$

Undamped natural frequency: $\omega_n = 0.898$ rad/s

Third root: $s_3 = -5.192$

The values of peak time T_p, peak overshoot M_o, and settling time T_s obtained from Eqs. (10.9), (10.13), and (10.14) are $T_p = 4.12$ s, $M_o = 0.202$, and

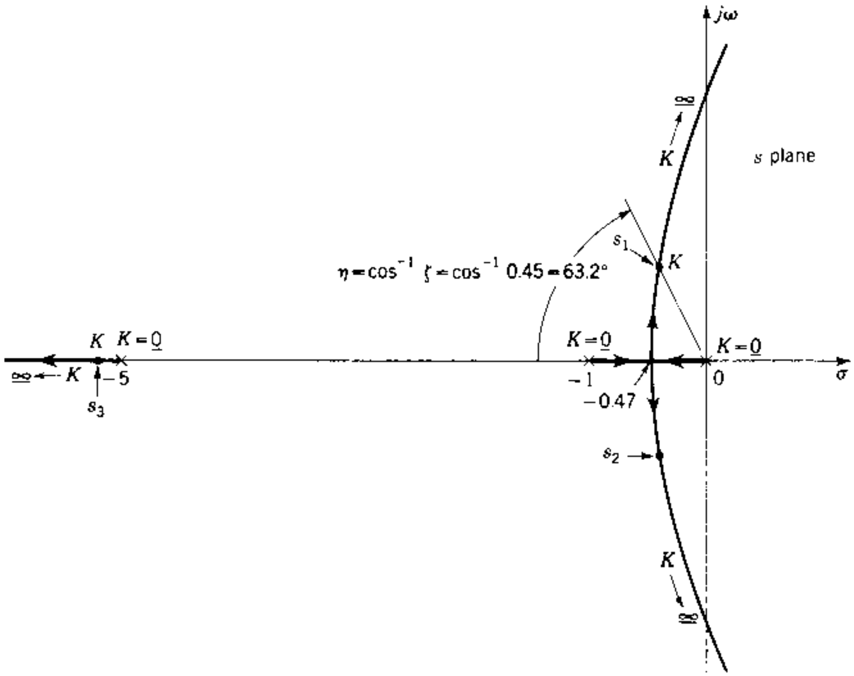


FIGURE 10.10 Root locus of $G_x(s) = K/s(s+1)(s+5)$.

$T_s = 9.9$ s. These are approximate values, whereas the computer program gives more accurate values (see Table 10.1) including $t_s = 9.48$ s.

To increase the gain, a lag compensator is put in cascade with the forward transfer function. With the criteria discussed in the preceding section and $\alpha = 10$, the compensator pole is located at $p_c = -0.005$ and the zero at $z_c = -0.05$. This pole and zero are achieved for the network of Fig. 10.9 with the values $R_1 = 18$ MΩ, $R_2 = 2$ MΩ, and $C = 10$ μF. The angle of the compensator at the original dominant roots is about 2.6° and is acceptable. Since the compensator pole and zero are very close together, they make only a small change in the new root locus in the vicinity of the original roots. The compensator transfer function is

$$G_c(s) = \frac{A}{\alpha} \frac{s + 1/T}{s + 1/\alpha T} = \frac{A}{10} \frac{s + 0.05}{s + 0.005} \tag{10.31}$$

Figure 10.11 (not to scale) shows the new root locus for $G(s) = G_x(s)G_c(s)$ as solid lines and the original locus as dashed lines. Note that for the damping ratio $\zeta = 0.45$ the new locus and the original locus are close together. The new dominant roots are $s_{1,2} = -0.384 \pm j0.763$; thus, the roots are essentially

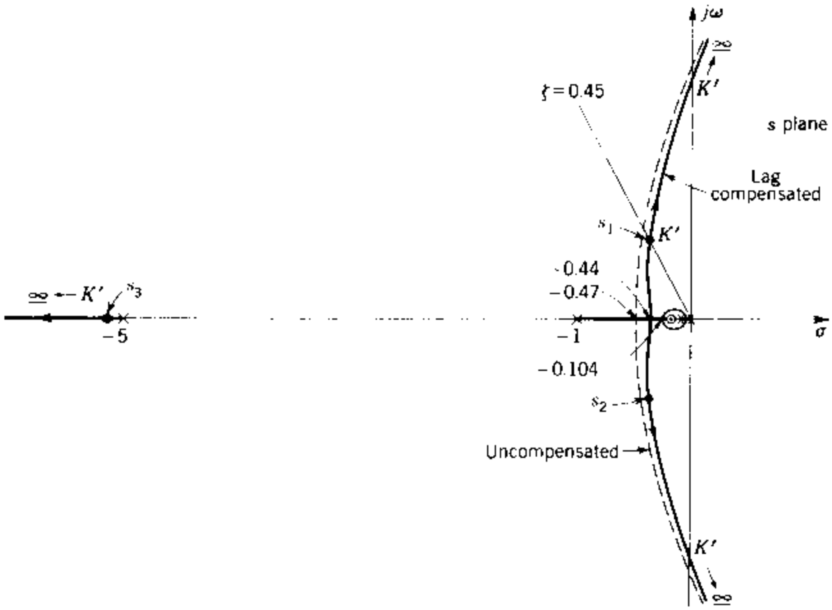


FIGURE 10.11 Root locus of $G(s) = K'(s+0.05)/s(s+1)(s+5)(s+0.005)$.

unchanged. Adjusting the gain of the compensated system to obtain roots with a damping ratio 0.45, the following results are obtained:

Dominant roots: $s_{1,2} = -0.384 \pm j0.763$

Loop sensitivity : $K' = \dfrac{|s| \cdot |s+1| \cdot |s+5| \cdot |s+0.005|}{|s+0.05|} = 4.01$

Ramp error coefficient: $K_1' = \dfrac{K'\alpha}{5} = \dfrac{(4.01)(10)}{5} = 8.02\,\text{s}^{-1}$

Increase in gain: $A = \dfrac{K_1'}{K_1} = \dfrac{8.02}{0.838} = 9.57$

Undamped natural frequency: $\omega_n = 0.854$ rad/s

Others roots: $s_3 = -5.183$ and $s_4 = -0.053$

The value of T_p for the compensated system is determined by use of Eq. (10.9), with the values shown in Fig. 10.12:

$$T_p = \frac{1}{\omega_d}\left[\frac{\pi}{2} - \psi_1 + (\phi_2 + \phi_3 + \phi_4)\right]$$
$$= \frac{1}{0.763}[90^\circ - 113.6^\circ + (113.4^\circ + 9.04^\circ + 90^\circ)]\frac{\pi\,\text{rad}}{180^\circ} = 4.32\,\text{s}$$

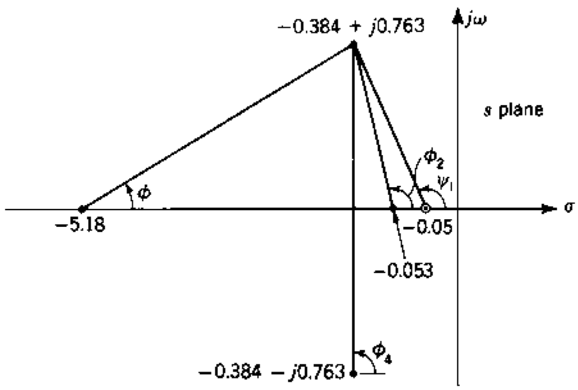


FIGURE 10.12 Angles used to evaluate T_p.

The value of M_o for the compensated system is determined by use of Eq. (10.13), with the values shown in Fig. 10.13:

$$M_o = \frac{l_1 l_2 (l)_1}{l_3 l_4 (l)_2} e^{\sigma T_p} = \frac{(5.18)(0.053)}{(4.76)(0.832)} \frac{0.834}{0.05} e^{(-0.384)(4.32)}$$
$$= 1.1563 e^{-1.659} = 0.22$$

The values obtained from a computer are $t_p = 4.30$ s and $M_o = 0.266$ and are listed in Table 10.1, Section 10.13, where a comparison is made of the performances achieved with several types of cascade compensators.

A comparison of the uncompensated system and the system with integral compensation shows that the ramp error coefficient K_1' has increased by a factor of 9.57 but the undamped natural frequency has been decreased slightly from 0.898 to 0.854 rad/s. That is, the steady-state error with a ramp input has decreased but the settling time has been increased from 9.9 to 22.0 s. Provided that the increased settling time and peak overshoot are acceptable, the system has been improved. To reduce this increase in settling time, it is necessary to include the constriant [5] that $|(s_4 - z_c)/s_4| \le -0.02$. The reader may demonstrate this by redesigning the lag compensator. Although the complete root locus is drawn in Fig. 10.11, in practice this is not necessary. Only the general shape is sketched and the particular needed points are accurately determined by using a CAD program.

10.9 IDEAL DERIVATIVE CASCADE COMPENSATION (PD CONTROLLER)

When the transient response of a feedback system must be improved, it is necessary to reshape the root locus so that it is moved farther to the left of the

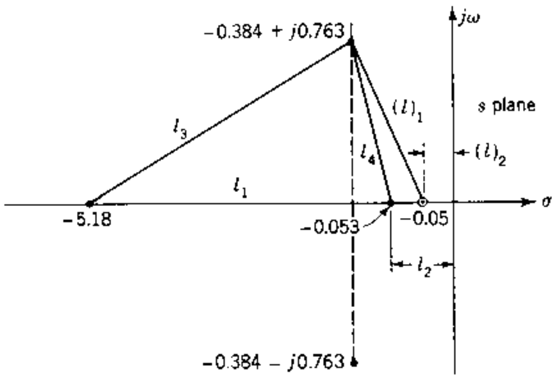


FIGURE 10.13 Lengths used to evaluate M_o.

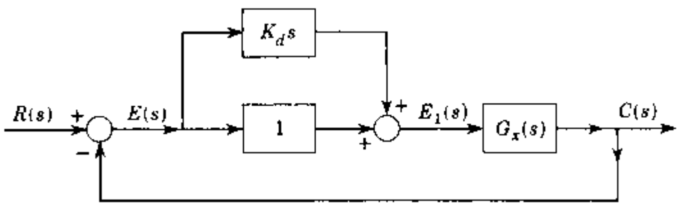


FIGURE 10.14 Ideal proportional plus derivative (PD) control.

imaginary axis. Section 7.3 shows that introducing an additional zero in the forward transfer function produces this effect. A zero can be produced by operating on the actuating signal to produce a signal that is proportional to both the magnitude and the derivative (rate of change) of the actuating signal. This proportional plus derivative (PD) controller is shown in Fig. 10.14, where

$$E_1(s) = (1 + K_d s)E(s) \tag{10.32}$$

Physically the effect can be described as introducing anticipation into the system. The system reacts not only to the magnitude of the error but also to its rate of change. If the error is changing rapidly, then $e_1(t)$ is large and the system responds faster. The net result is to speed up the response of the system. On the root locus the introduction of a zero has the effect of shifting the curves to the left, as shown in the following example.

Example. The root locus for a Type 2 system that is unstable for all values of loop sensitivity K is shown in Fig. 10.15a. Adding a zero to the system by means of the PD compensator $G_c(s)$ has the effect of moving the locus to the left, as

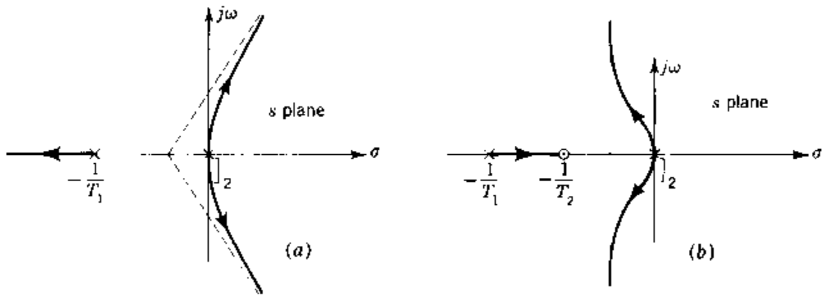


FIGURE 10.15 Root locus for (a) $G_x(s) = K/s^2(s + 1/T_1)$; (b) $G(s) = K(s + 1/T_2)/s^2(s + 1/T_1)$.

shown in Fig. 10.15*b*. The addition of a zero $s = -1/T_2$ between the origin and the pole at $-1/T_1$ stabilizes the system for all positive values of gain.

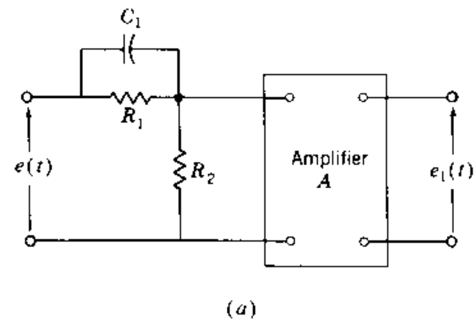
This example shows how the introduction of a zero in the forward transfer function by a proportional plus derivative compensator has modified and improved the feedback control system. However, the derivative action amplifies any spurious signal or noise that may be present in the actuating signal. This noise amplification may saturate electronic amplifiers so that the system does not operate properly. A simple passive network is used to approximate proportional plus derivative action as shown in the next section.

10.10 LEAD COMPENSATION DESIGN USING PASSIVE ELEMENTS

Figure 10.16*a* shows a derivative or lead compensator made up of electrical elements. To this network an amplifier of gain A is added in cascade. The transfer function is

$$G_c(s) = A\alpha\frac{1 + Ts}{1 + \alpha Ts} = A\frac{s + 1/T}{s + 1/\alpha T} = A\frac{s - z_c}{s - p_c} \tag{10.33}$$

where $\alpha = R_2/(R_1 + R_2) < 1$ and $T = R_1C$. This lead compensator introduces a zero at $z_c = -1/T$ and a pole at $p_c = -1/\alpha T$. Thus, $\alpha = z_c/p_c$. By making α sufficiently small, the location of the pole p_c is far to the left of the zero z_c, and therefore the pole has small effect on the dominant part of the root locus. Near the zero the net angle of the compensator is due predominantly to the zero. Figure 10.16*b* shows the location of the lead-compensator pole and zero and the angles contributed by each at a point s_1. It is also found that the gain of the compensated system is often increased. The maximum increases in gain and in the real part ($\zeta\omega_n$) of the dominant root of the characteristic equation do not coincide. The compensator zero



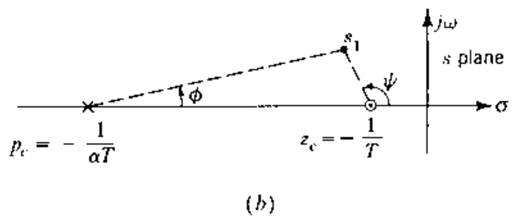


FIGURE 10.16 (*a*) Derivative or lead compensator. (*b*) Location of pole and zero of a lead compensator.

location must be determined by making several trials with the aid of a CAD program for the desired optimum performance.

The loop sensitivity is proportional to the ratio of $|p_c/z_c|$. Therefore, as α decreases, the loop sensitivity increases. The minimum value of α is limited by the size of the parameters needed in the network to obtain the minimum input impedance required. Note also from Eq. (10.33) that a small α requires a large value of additional gain A from the amplifier. The value $\alpha = 0.1$ is a common choice.

Design Example—Lead Compensation Applied to a Type 1 System

The same system used in Sec. 10.7 is now used with lead compensation. The locations of the pole and zero of the compensator are first selected by trial. At the conclusion of this section some rules are given to show the best location.

The forward transfer function of the original system is

$$G_x(s) = \frac{K_1}{s(1+s)(1+0.2s)} = \frac{K}{s(s+1)(s+5)} \tag{10.34}$$

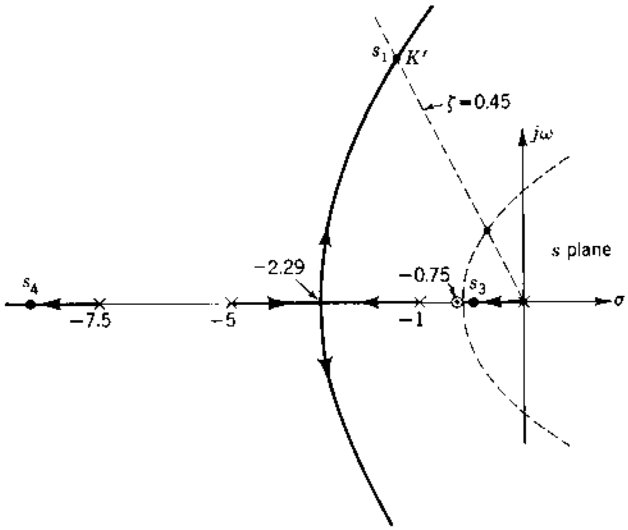


FIGURE 10.17 Root locus of $G(s) = AK(s + 0.75)/(s(s + 1)(s + 5)(s + 7.5))$.

The transfer function of the compensator, using a value $\alpha = 0.1$, is

$$G_c(s) = 0.1A\frac{1 + Ts}{1 + 0.1Ts} = A\frac{s + 1/T}{s + 1/0.1T} \tag{10.35}$$

The new forward transfer function is

$$G(s) = G_c(s)G_x(s) = \frac{AK(s + 1/T)}{s(s + 1)(s + 5)(s + 1/0.1T)} \tag{10.36}$$

Three selections for the position of the zero $s = -1/T$ are made. A comparison of the results shows the relative merits of each. The three locations of the zero are $s = -0.75$, $s = -1.00$, and $s = -1.50$. Figures 10.17 to 10.19 show the resultant root loci for these three cases. In each figure the dashed lines are the locus of the original uncompensated system.

The loop sensitivity is evaluated in each case from the expression

$$K' = AK = \frac{|s| \cdot |s + 1| \cdot |s + 5| \cdot |s + 1/\alpha T|}{|s + 1/T|} \tag{10.37}$$

and the ramp error coefficient is evaluated from the equation

$$K_1' = \lim_{s \to 0} sG(s) = \frac{K'\alpha}{5} \tag{10.38}$$

A summary of the results is given in Table 10.1, which shows that large increase in ω_n has been achieved by adding the lead compensators. Since the value of

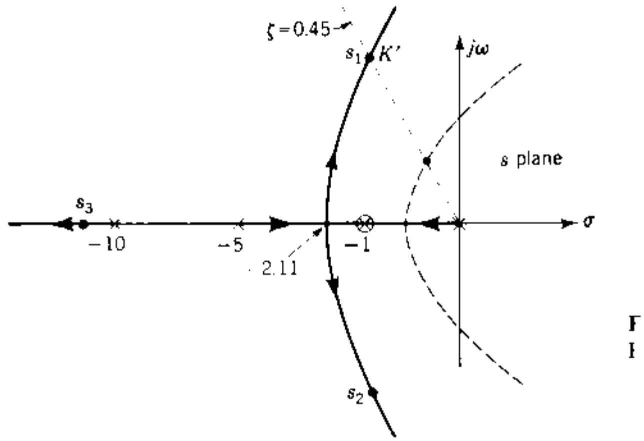


FIGURE 10.18 Root locus of $G(s) = AK(s+1)/(s(s+1)(s+5)(s+10))$.

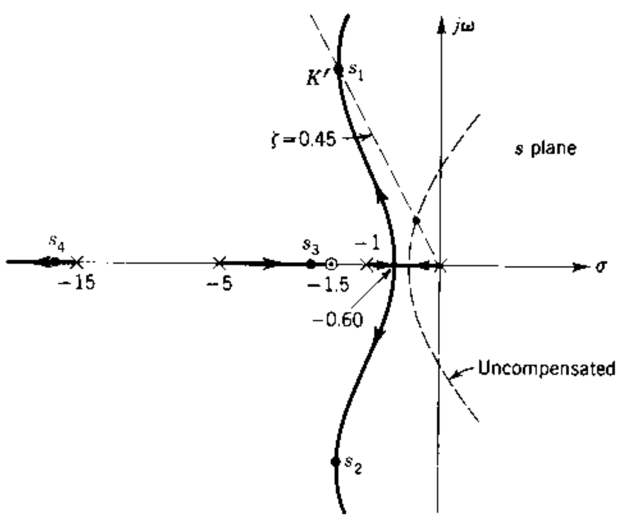


FIGURE 10.19 Root locus of $G(s) = AK(s+1.5)/(s(s+1)(s+5)(s+15))$.

ζ is held constant for all three cases, the system has a much faster response time. A comparison of the individual lead compensators shows the following.

Case A. Placing the zero of the compensator to the right of the plant pole $s = -1$ results in a root s_3 on the negative real axis close to the origin (see Fig. 10.17). From Eq. (4.68) it can be seen that the coefficient associated

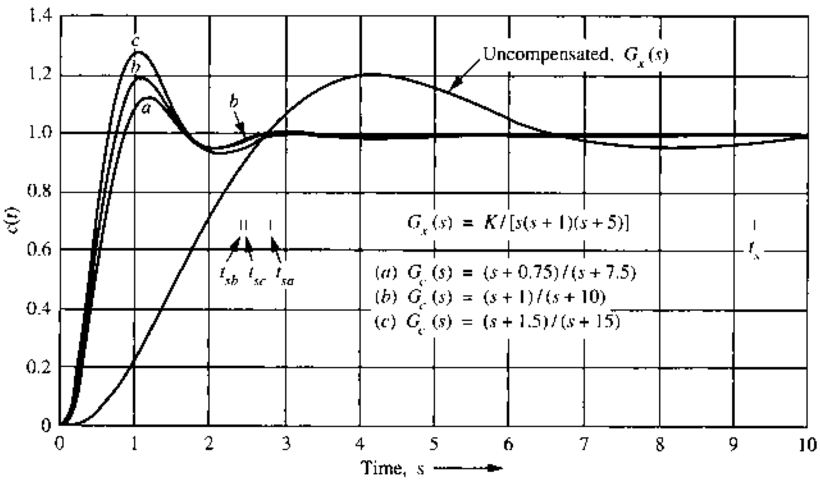


FIGURE 10.20 Time response with a step input for the uncompensated and the lead-compensated cases of Table 10.1.

with this transient term is negative. Therefore, it has the desirable effect of decreasing the peak overshoot and the undesirable characteristic of having a longer settling time. If the size of this transient term is small, (the magnitude is proportional to the distance from the real root to the zero), the effect on the settling time is small.

The decrease in M_o can be determined by calculating the contribution of the transient term $A_3 e^{s_3 t}$ due to the real root at time t_p, as discussed in Sec. 10.3. At the time $t = T_s$, the sign of the underdamped transient is negative and the transient due to the real root is also negative; therefore the actual settling time is increased. The exact settling time obtained from a complete solution of the time response is $t_{sa} = 2.78$ s and is shown as plot *a* in Fig. 10.20.

Case B. The best settling time occurs when the zero of the compensator cancels the pole $s = -1$ of the original transfer function (see Fig. 10.18). There is no real root near the imaginary axis in this case. The actual settling time is $t_{sb} = 2.45$ s as shown in plot *b* in Fig. 10.20. The order of the characteristic equation is unchanged.

Case C. The largest gain occurs when the zero of the compensator is located at the left of the pole $s = -1$ of the original transfer function (see Fig. 10.19). Since the transient due to the root $s_3 = -1.76$ is positive, it increases the peak overshoot of the response. The value of the transient due to the real root can be added to the peak value of the underdamped sinusoid to obtain the correct value of the peak overshoot. The real root may also affect the settling time.

In this case the real root decreases the peak time and increases the settling time. From plot c in Fig. 10.20 the actual values are $t_p = 1.05$ s and $t_{sc} = 2.50$ s.

From the root-locus plots it is evident that increasing the gain increases the peak overshoot and the settling time. The responses for each of the three lead compensators are shown in Fig. 10.20. The effect of the additional real root is to change the peak overshoot and the settling time. The designer must make a choice between the largest gain attainable, the smallest peak overshoot, and the shortest settling time.

Based on this example, the following guidelines are used to apply cascade lead compensators to a Type 1 or higher system:

1. If the zero $z_c = -1/T$ of the lead compensator is superimposed and cancels the largest real pole (excluding the pole at zero) of the original transfer function, a good improvement in the transient response is obtained.
2. If a larger gain is desired than that obtained by guideline 1, several trials should be made with the zero of the compensator moved to the left or right of the largest real pole of $G_x(s)$. The location of the zero that results in the desired gain and roots is selected.

For a Type 0 system it is often found that a better time response and a larger gain can be obtained by placing the compensator zero so that it cancels or is close to the second-largest real pole of the original transfer function.

10.11 GENERAL LEAD-COMPENSATOR DESIGN

Additional methods are available for the design of an appropriate lead compensator G_c that is placed in cascade with the open loop transfer function G_x, as shown in Fig. 10.1a. Figure 10.21 shows the original root locus of a control system. For a specified damping ratio ζ the dominant root of the uncompensated system is s_1. Also shown is s_2, which is the desired root of the system's characteristic equation. Selection of s_2 as a desired root is based on the performance required for the system. The design problem is to select a lead compensator that results in s_2 being a root. The first step is to find the sum of the angles at the point s_2 due to the poles and zeros of the original transfer function G_x. This angle is set equal to $180^\circ + \phi$. For s_2 to be on the new root locus, it is necessary for the lead compensator to contribute an angle $\phi_c = -\phi$ at this point. The total angle at s_2 is then 180°, and it is a point on the new root locus. A simple lead compensator represented by Eq. (10.33), with its zero to the right of its pole, can be used to provide the angle ϕ_c at the point s_2.

Method 1. Actually, there are many possible locations of the compensator pole and zero that will produce the necessary angle ϕ_c at the point s_2.

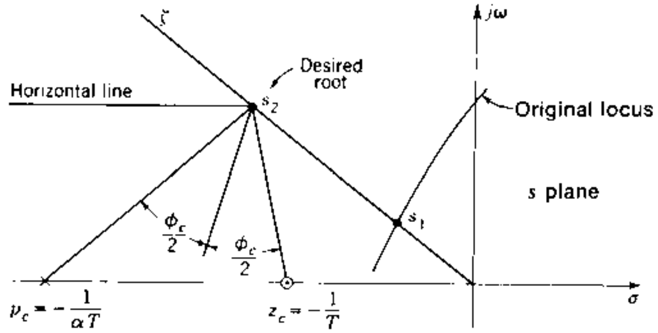


FIGURE 10.21 Graphical construction for locating the pole and zero of a simple lead compensator.

Cancellation of poles of the original open-loop transfer function by means of compensator zeros may simplify the root locus and thereby reduce the complexity of the problem. The compensator zero is simply placed over a real pole. Then the compensator pole is placed further to the left at a location that makes s_2 a point on the new root locus. The pole to be canceled depends on the system type. For a Type 1 system the largest real pole (excluding the pole at zero) should be canceled. For a Type 0 system the second-largest real pole should be canceled.

Method 2. The following construction is based on obtaining the maximum value of α, which is the ratio of the compensator zero and pole. The steps in locating the lead-compensator pole and zero (see Fig. 10.21) are these:

1. Locate the desired root s_2. Draw a line from this root to the origin and a horizontal line to the left of s_2.
2. Bisect the angle between the two lines drawn in step 1
3. Measure the angle $\phi_c/2$ on either side of the bisector drawn in step 2.
4. The intersections of these lines with the real axis locate the compensator pole p_c and zero z_c.

The construction just outlined is shown in Fig. 10.21. It is left for the reader to show that this construction results in the largest possible value of α. Since α is less than unity and appears as the gain of the compensator [see Eq. (10.33)], the largest value of α requires the smallest additional gain A.

Method 3. Changing the locations of the compensator pole and zero can produce a range of values for the gain while maintaining the desired root s_2. A specified procedure for achieving both the desired root s_2 and the system gain K_m is given in Refs. 6 and 7.

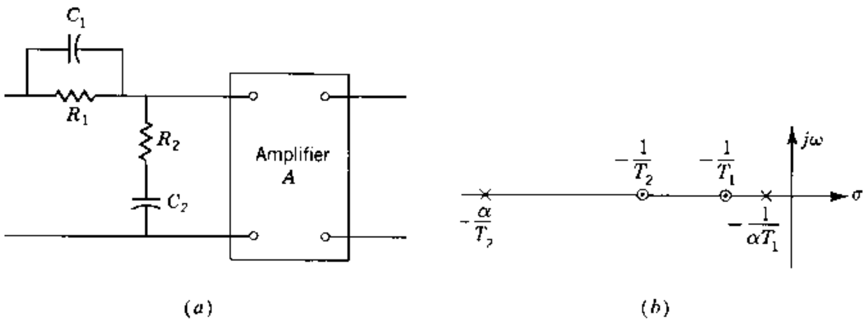


FIGURE 10.22 (*a*) Lag-lead compensator. (*b*) Pole and zero locations.

10.12 LAG-LEAD CASCADE COMPENSATION DESIGN

The preceding sections show that (1) the insertion of an integral or lag compensator in cascade results in a large increase in gain and a small reduction in the undamped natural frequency and (2) the insertion of a derivative or lead compensator in cascade results in a small increase in gain and large increase in the undamped natural frequency. By inserting both the lag and the lead compensators in cascade with the original transfer function, the advantages of both can be realized simultaneously; i.e., a large increase in gain and a large increase in the undamped natural frequency can be obtained. Instead of using two separate networks, it is possible to use one network that acts as both a lag and a lead compensator. Such a network, shown in Fig. 10.22 with an amplifier added in cascade, is called a lag-lead compensator. The transfer function of this compensator (see Sec. 8.12) is

$$G_c(s) = A\frac{(1 + T_1 s)(1 + T_2 s)}{(1 + \alpha T_1 s)[1 + (T2/\alpha)s]} = A\frac{(s + 1/T_1)(s + 1/T_2)}{(s + 1/\alpha T_1)(s + \alpha/T_2)} \qquad (10.39)$$

where $\alpha > 1$ and $T_1 > T_2$.

The fraction $(1 + T_1 s)/(1 + \alpha T_1 s)$ represents the lag compensator, and the fraction $(1 + T_2 s)/[1 + (T_2/\alpha)s]$ represents the lead compensator. The values T_1, T_2, and α are selected to achieve the desired improvement in system performance. The specific values of the compensator components are obtained from the relationships $T_1 = R_1 C_1, T_2 = R_2 C_2$, $\alpha T_1 + T_2/\alpha = R_1 C_1 + R_2 C_2 + R_1 C_2$. It may also be desirable to specify a minimum value of the input impedance over the passband frequency range.

The procedure for applying the lag-lead compensator is a combination of the procedures for the individual units.

1. For the integral or lag component:
   (*a*) The zero $s = -1/T_1$ and the pole $s = -1/\alpha T_1$ are selected close together, with α set to a large value such as $\alpha = 10$.
   (*b*) The pole and zero are located to the left of and close to the origin, resulting in an increased gain.
2. For the derivative or lead component, the zero $s = -1/T_2$ can be superimposed on a pole of the original system. Since α is already selected in step 1, the new pole is located at $s = -\alpha/T_2$. As a result the root locus moves to the left and therefore the undamped natural frequency increases.

The relative positions of the poles and zeros of the lag-lead compensators are shown in Fig. 10.22*b*.

Design Example—Lag-Lead Compensation Applied to a Type 1 System

The system described in Secs. 10.8 and 10.10 is now used with a lag-lead compensator. The forward transfer function is

$$G_x(s) = \frac{K_1}{s(1+s)(1+0.2s)} = \frac{K}{s(s+1)(s+5)} \tag{10.40}$$

The lag-lead compensator used is shown in Fig. 10.22, and the transfer function $G_c(s)$ is given by Eq. (10.39). The new complete forward transfer function is

$$G(s) = G_c(s)G_x(s) = \frac{AK(s+1/T_1)(s+1/T_2)}{s(s+1)(s+5)(s+1/\alpha T_1)(s+\alpha/T_2)} \tag{10.41}$$

The poles and zeros of the compensator are selected in accordance with the principles outlined prevoiusly in this section. They coincide with the values used for the integral and derivative compensators when applied individually in Secs. 10.8 and 10.10. With an $\alpha = 10$, the compensator poles and zeros are

$$z_{\text{lag}} = -\frac{1}{T_1} = -0.05 \qquad p_{\text{lag}} = -\frac{1}{\alpha T_1} = -0.005$$

$$z_{\text{lead}} = -\frac{1}{T_2} = -1 \qquad p_{\text{lead}} = -\frac{\alpha}{T_2} = -10$$

The forward transfer function becomes

$$G(s) = G_c(s)G_x(s) = \frac{K'(s+1)(s+0.05)}{s(s+1)(s+0.005)(s+5)(s+10)} \tag{10.42}$$

The root locus (not to scale) for this system is shown in Fig. 10.23.

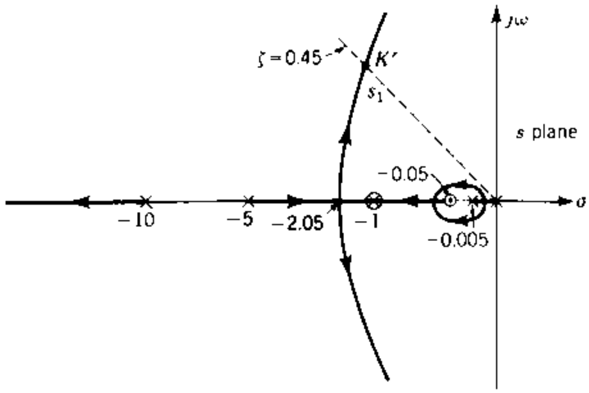


FIGURE 10.23 Root locus of $G(s) = K'(s + 0.05)/(s(s + 0.005)(s + 5)(s + 10))$.

For damping ratio $\zeta = 0.45$ the following pertinent data are obtained for the compensated system:

Dominant roots: $s_{1,2} = -1.571 \pm j3.119$

Loop sensitivity (evaluated at $s = s_1$): $K' = \dfrac{|s| \cdot |s + 0.005| \cdot |s + 5| \cdot |s + 10|}{|s + 0.05|} = 146.3$

Ramp error coefficient: $K_1' = \lim_{s \to 0} sG(s) = \dfrac{K'(0.05)}{(0.005)(5)(10)} = 29.3\text{s}^{-1}$

Undamped natural frequency: $\omega_n = 3.49$ rad/s

The additional roots are $s_3 = -11.81$ and $s_4 = -0.051$. These results, when compared with those of the uncompensated system, show a large increase in both the gain and the undamped natural frequency. They indicate that the advantages of the lag-lead compensator are equivalent to the combined improvement obtained by the lag and the lead compensators. Note the practicality in the choice of $\alpha = 10$, which is based on experience. The response characteristics are shown in Table 10.1.

10.13 COMPARISON OF CASCADE COMPENSATORS

Table 10.1 shows a comparison of the system response obtained when a lag, a lead, and a lag-lead compensator are added in cascade with the same basic feedback system. These results are for the examples of Secs. 10.8, 10.10, and 10.12 and are obtained by use of a CAD program. These results are typical of the changes introduced by the use of cascade compensators, which can be

Table 10.1 Comparison of Performance of Several Cascade Compensators for the System of Eq. (10.30) with $\zeta = 0.45$

Compensator	Dominant complex roots	Undamped natural frequency, ω_n	Other roots	Ramp error coefficients, K_1	t_p, s	M_o	t_s, s	Additional gain, required, A
1. Uncompensated	$-0.404 + j0.802$	0.898	-5.192	0.84	4.12	0.202	9.48	—
2. Lag: $\dfrac{s+0.05}{s+0.005}$	$-0.384 + j0.763$	0.854	-0.053 -5.183	8.02	4.30	0.266	22.0	9.57
3. Lead: $\dfrac{s+0.75}{s+7.5}$	$-1.522 + j3.021$	3.38	-0.689 -9.77	2.05	1.18	0.127	2.78	24.4
4. Lead: $\dfrac{s+1}{s+10}$	$-1.588 + j3.152$	3.53	-11.83	2.95	1.09	0.195	2.45	35.1
5. Lead: $\dfrac{s+1.5}{s+15}$	$-1.531 + j3.039$	3.40	-1.76 -16.18	4.40	1.05	0.280	2.50	52.5
6. Lag-lead: $\dfrac{(s+0.05)(s+1)}{(s+0.005)(s+10)}$	$-1.571 + j3.119$	3.49	-0.051 -11.81	29.26	1.10	0.213	3.33	34.9
7. PID: $\dfrac{7.194(s+1)(s+0.1)}{s}$	$-2.449 + j4.861$	5.443	-0.1017	∞	0.646	0.225	2.10	7.194

summarized as follows:

1. Lag compensator
   (*a*) Results in a large increase in gain K_m (by a factor almost equal to α), which means a much smaller steady-state error
   (*b*) Decreases ω_n and therefore has the disadvantage of producing an increase in the settling time
2. Lead compensator
   (*a*) Results in a moderate increase in gain K_m, thereby improving steady-state accuracy
   (*b*) Results in a large increase in ω_n and therefore significantly reduces the settling time
   (*c*) The transfer function of the lead compensator, using the passive network of Fig. 10.16*a*, contains the gain α [see Eq. (10.33)], which is less than unity. Therefore the additional gain A, which must be added, is larger than the increase in K_m for the system
3. Lag-lead compensator (essentially combines the desirable characteristics of the lag and the lead compensators)
   (*a*) Results in a large increase in gain K_m, which improves the steady-state response
   (*b*) Results in a large increase in ω_n, which improves the transient-response settling time

Figure 10.24 shows the frequency response for the original and compensated systems corresponding to Table 10.1. There is a close qualitative correlation between the time-response characteristics listed in Table 10.1 and the corresponding frequency-response curves shown in Fig. 10.24. The peak time response value M_p is directly proportional to the maximum value M_m, and the settling time t_s is a direct function of the passband frequency. Curves 3, 4, and 5 have a much wider passband than curves 1 and 2; therefore, these systems have a faster settling time.

For systems other than the one used as an example, these simple compensators may not produce the desired changes. *However, the basic principles described here are applicable to any system for which the open-loop poles closest to the imaginary axis are on the negative real axis.* For these systems the conventional compensators may be used, and good results are obtained.

When the complex open-loop poles are dominant, the conventional compensators produce only small improvements. A pair of complex poles near the imaginary axis often occurs in aircraft and in missile transfer functions. Effective compensation could remove the open-loop dominant complex poles and replace them with poles located farther to the left. This means that the compensator should have a transfer function of the

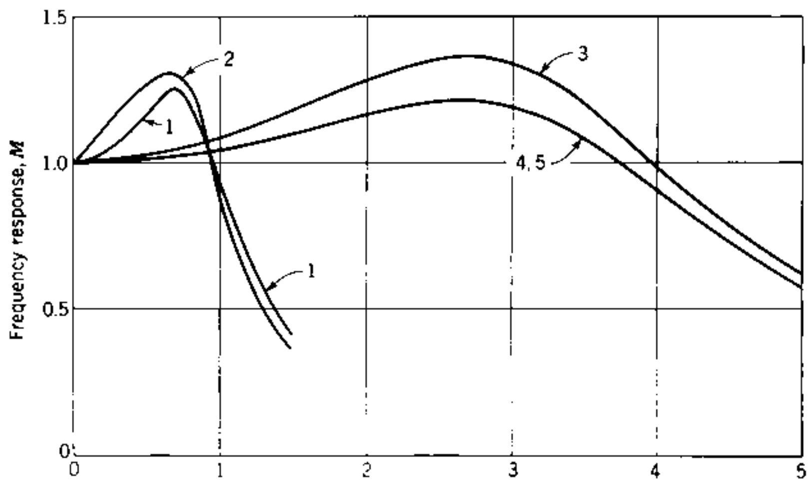


FIGURE 10.24 Frequency response for systems of Table 10.1: (1) original basic system; (2) lag-compensated; (3) lead-compensated, zero at −1.5; (4) lead-compensated, zero at −1.0; (5) lag-lead compensated.

form

$$G_c(s) = \frac{1 + Bs + As^2}{1 + Ds + Cs^2} \tag{10.43}$$

If exact cancellation is assumed, the zeros of $G_c(s)$ would coincide with the complex poles of the original transfer function. A representative root locus, when the compensator has real poles, is shown in Fig. 10.25. The solid lines are the new locus, and the dashed lines are the original locus. Thus, the transient response has been improved. However, the problem remaining to be solved is the synthesis of a network with the desired poles and zeros. There are references that cover network synthesis [8, 9]. An alternative design method for improving the performance of plants that have dominant complex poles is the use of rate feedback compensation (see Sec. 10.21). This method does not require the assumption of exact cancellation shown in Fig. 10.25.

10.14 PID CONTROLLER

The PI and PD controller of Secs. 10.7 and 10.9, respectively, can be combined into a single controller, i.e., a proportional plus integral plus derivative (PID) controller as shown in Fig. 10.26. The transfer function of this controller is

$$G_c(s) = \frac{E_1(s)}{E(s)} = K_d s + K_p + \frac{K_i}{s} = \frac{K_d(s - z_1)(s - z_2)}{s} \tag{10.44}$$

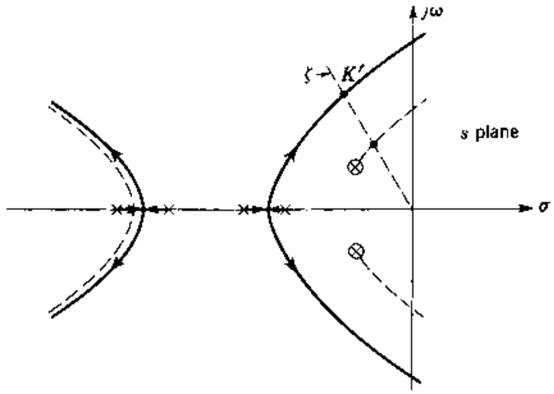


FIGURE 10.25 Root locus of $G_c(s)G_x(s) = K_0'/((1 + T_1 s)(1 + T_2 s)(1 + T_3 s) \times (1 + T_4 s))$.

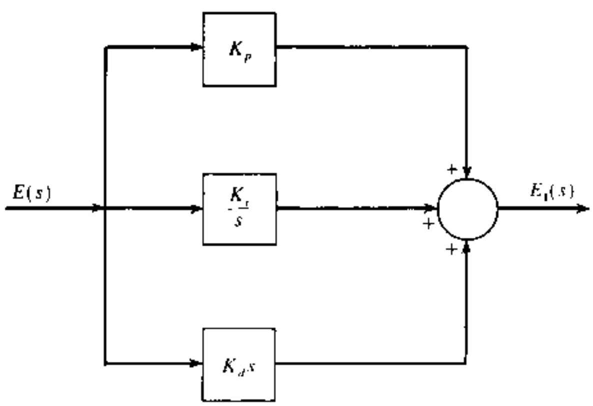


FIGURE 10.26 A PID controller.

where $(z_1 + z_2) = -K_p/K_d$ and $z_1 z_2 = K_i/K_d$. The compensator of Eq. (10.44) is placed in cascade with the basic system of Eq. (10.30), with $z_1 = -1$ and $z_2 = -0.1$, so the forward transfer function is

$$G_c(s)G_x(s) = \frac{KK_d(s + 0.1)}{s^2(s + 5)}$$

For a damping ratio $\zeta = 0.45$ the resulting control ratio is

$$\frac{C(s)}{R(s)} = \frac{30.13(s + 0.1)}{(s + 2.449 \pm j4.861)(s + 0.1017)} \tag{10.45}$$

The figures of merit for the PID-compensated system are $M_p = 1.225$, $t_p = 0.646$ s, $t_s = 2.10$ s, $K_v = \infty$, and $K_2 = 0.6026\, s^{-2}$. Comparing these results with those in Table 10.1 shows that the PID controller results in the best improvement in system performance. Also note that the use of PID controller increases the system type from Type 1 to Type 2. With the advances in the state of the art of microelectronics, satisfactory derivative action may be achieved to produce the PID controller. However, to avoid saturation of the differentiated signal e due to noise that is present in the system, it is necessary to add a cascade filter to attenuate the high frequencies.

10.15 INTRODUCTION TO FEEDBACK COMPENSATION

System performance may be improved by using either cascade or feedback compensation. The factors that must be considered when making that decision are listed in Sec. 10.1. The effectiveness of cascade compensation is illustrated in the preceding sections of this chapter. The use of feedback compensation is demonstrated in the remainder of this chapter. The location of a feedback compensator depends on the complexity of the basic system, the accessibility of insertion points, the form of the signal being fed back, the signal with which it is being compared, and the desired improvement. Figure 10.27 shows various forms of feedback-compensated systems. For each case there is a unity-feedback loop (major loop) in addition to the

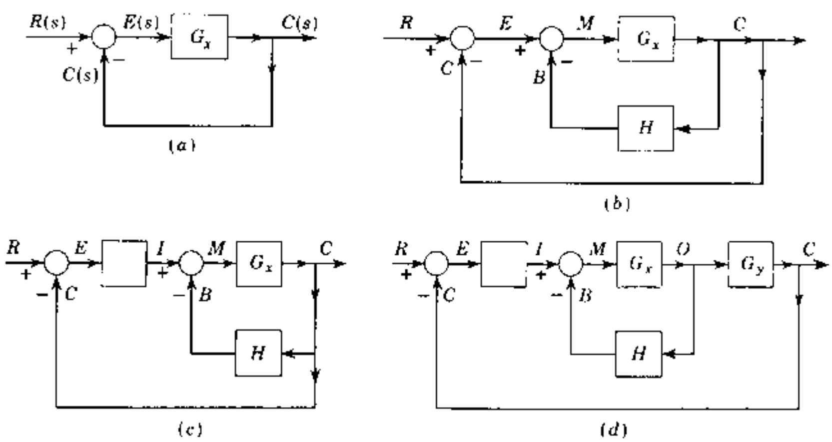


FIGURE 10.27 Various forms of feedback systems: (*a*) uncompensated system; (*b*) feedback-compensated system; (*c*) and (*d*) feedback-compensated systems with a cascade gain *A* or compensator G_c.

TABLE 10.2 Basic Forms of $H(s)$ to Be Used to Prevent Change of System Type

Basic form of $H(s)$	Type with which it can be used
K	0
Ks	0 and 1
Ks^2	0, 1, and 2
Ks^3	0, 1, 2, and 3

compensation loop (minor loop) in order to maintain a direct correspondence (tracking) between the output and input.

It is often necessary to maintain the basic system type when applying feddback compensation. Care must therefore be exercised in determining the number of differentiating terms that are permitted in the numerator of $H(s)$. Table 10.2 presents the basic forms that $H(s)$ can have so that the system type remains the same after the minor feedback-compensation loop is added. $H(s)$ may have additional factors in the numerator and denominator. It can be shown by evaluating $G(s) = C(s)/E(s)$ that, to maintain the system type, the degree of the factor s in the numerator of $H(s)$ must be equal to or higher than the type of the forward transfer function $G_x(s)$ shown in Fig. 10.27.

The use of a direct or unity-feedback loop on any non–Type 0 control element converts it into a Type 0 element. This effect is advantageous when going from a low-energy-level input to a high-energy-level output for a given control element and also when the forms of the input and output energy are different. When $G_x(s)$ in Fig. 10.28 is Type 1 or higher, it may represent either a single control element or a group of control elements whose transfer function is of the general form

$$G_x(s) = \frac{K_m(1 + T_1 s)\cdots}{s^m(1 + T_a s)(1 + T_b s)\cdots} \tag{10.46}$$

when $m \neq 0$. The overall ratio of output to input is given by

$$G(s) = \frac{O(s)}{I(s)} = \frac{G_x(s)}{1 + G_x(s)} \tag{10.47}$$

or

$$G(s) = \frac{K_m(1 + T_1 s)\cdots}{s^m(1 + T_a s)(1 + T_b s)\cdots + K_m(1 + T_1 s)\cdots} \tag{10.48}$$

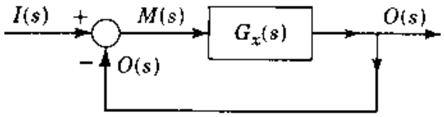


FIGURE 10.28 Transforming non-Type 0 control element to Type 0.

Equation (10.48) represents an equivalent forward transfer function between $O(s)$ and $I(s)$ of Fig. 10.28 and has the form of a Type 0 control element. Thus, the transformation of type has been effected. Therefore, for a step input, $O(t)_{ss} = \lim_{s\to 0}[sG(s)I(s)] = 1$. As a consequence, a given input signal produces an output signal of equal magnitude but at a higher energy level and/or with a different form of energy. The polar plots of $G_x(j\omega)$ and $O(j\omega)/I(j\omega)$ would show that for any given frequency the phase-margin angle is greater with feedback, thus indicating a more stable system (a smaller value of M_m).

10.16 FEEDBACK COMPENSATION: DESIGN PROCEDURES

Techniques are now described for applying feedback compensation to the root-locus method of system design [10]. The system to be investigated, see Fig. 10.29, has a minor feedback loop around the original forward transfer function $G_x(s)$. The minor-loop feedback transfer function $H(s)$ has the form described in the following pages. Either the gain portion of $G_x(s)$ or the cascade amplifier A is adjusted to fix the closed-loop system characteristics. For more complex systems a design approach based on the knowledge obtained for this simple system can be developed.

Two methods of attack can be used. The characteristic equation of the complete system may be obtained and the root locus plotted, by use of the *partition*[11] method, as a function of a gain that appears in the system. A more general approach is first to adjust the roots of the characteristic equation of the inner loop. These roots are then the poles of the forward transfer function

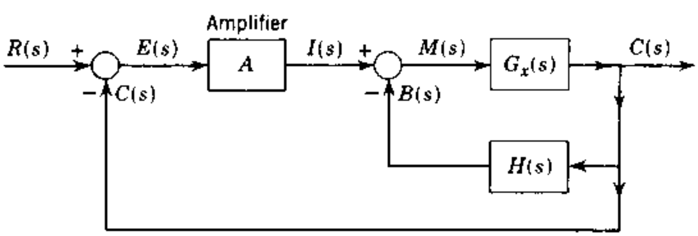


FIGURE 10.29 Block diagram for feedback compensation.

and are used to draw the root locus for the overall system. Both methods, using various types of feedback compensators, are illustrated in the succeeding sections. The techniques developed with cascade compensation should be kept in mind. Basically, the addition of poles and zeros changes the root locus. To improve the system time response the locus must be moved to the left, away from the imaginary axis.

10.17 SIMPLIFIED RATE FEEDBACK COMPENSATION: A DESIGN APPROACH

The feedback system shown in Fig. 10.29 provides the opportunity for varying three functions, $G_x(s)$, $H(s)$ and A. The larger the number of variables inserted into a system, the better the opportunity for achieving a specified performance. However, the design problem becomes much more complex. To develop a feel for the changes that can be introduced by feedback compensation, first a simplification is made by letting $A = 1$. The system of Fig. 10.29 then reduces to that shown in Fig. 10.30*a*. It can be further simplified as shown in Fig. 10.30*b*.

The first and simplest case of feedback compensation to be considered is the use of a rate transducer that has the transfer function $H(s) = K_t s$. The original forward transfer function is

$$G_x(s) = \frac{K}{s(s+1)(s+5)} \qquad H(s) = K_t s \tag{10.49}$$

The control ratio is

$$\frac{C(s)}{R(s)} = \frac{G_x(s)}{1 + G_x(s)H_1(s)} = \frac{G_x(s)}{1 + G_x(s)[1 + H(s)]} \tag{10.50}$$

The characteristic equation is

$$1 + G_x(s)[1 + H(s)] = 1 + \frac{K(1 + K_t s)}{s(s+1)(s+5)} = 0 \tag{10.51}$$

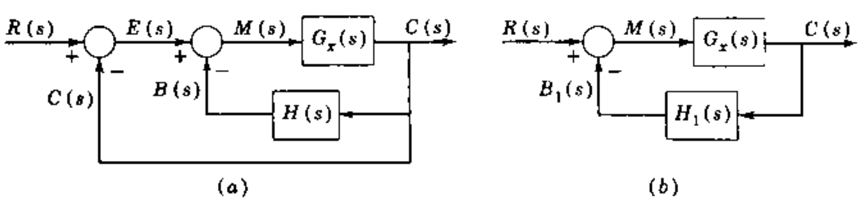


FIGURE 10.30 (*a*) Minor-loop compensation. (*b*) Simplified block diagram, $H_1(s) = 1 + H(s)$.

Equation (10.51) shows that the transducer plus the unity feedback have introduced the ideal proportional plus derivative control $(1 + K_t s)$ that can also be achieved by a cascade compensator. The results should therefore be qualitatively the same.

Partitioning the characteristic equation (10.51) to obtain the general format $F(s) = -1$, which is required for obtaining a root locus, yields

$$G_x(s)H_1(s) = \frac{KK_t(s + 1/K_t)}{s(s+1)(s+5)} = -1 \tag{10.52}$$

The root locus is drawn as a function of KK_t. Introduction of minor loop rate feedback results in the introduction of a zero [see Eq. 10.52]. It therefore has the effect of moving the locus to the left and improving the time response. *This zero is not the same as introducing a cascade zero because no zero appears in the rationalized equation of the control ratio*

$$\frac{C(s)}{R(s)} = \frac{K}{s(s+1)(s+5) + KK_t(s + 1/K_t)} \tag{10.53}$$

The rate constant K_t must be chosen before the root locus can be drawn for Eq. (10.52). *Care must be taken in selecting the value of K_t. For example, if $K_t = 1$* is used, cancellation of $(s+1)$ in the numerator and denominator of Eq. (10.52) yields

$$\frac{K}{s(s+5)} = -1 \tag{10.54}$$

and it is possible to get the impression that $C(s)/R(s)$ has only two poles. *This is not the case*, as can be seen from Eq. (10.53). Letting $K_t = 1$ provides the common factor $s+1$ in the characteristic equation so that it becomes

$$(s+1)[s(s+5) + K] = 0 \tag{10.55}$$

One closed-loop pole has therefore been made equal to $s = -1$, and the other two poles are obtained from the root locus of Eq. (10.54). If complex roots are selected from the root locus of Eq. (10.54), then the pole $s = -1$ is dominant and the response of the closed-loop system is overdamped (see Fig. 10.4*d*).

The proper design procedure is to select a number of trial locations for the zero $s = -1/K_t$, to tabulate or plot the system characteristics that result, and then to select the best combination. As an example, with $\zeta = 0.45$ as the criterion for the complex closed-loop poles and $K_t = 0.4$, the root locus is similar to that of Fig. 10.31 with $A = 1$. Table 10.3 gives a comparison of the original system with the rate-compensated system. From the root locus of Eq. (10.52) the roots are $s_{1,2} = -1 \pm j2$, and $s_3 = -4$, with $K = 20$. The gain

TABLE 10.3 Comparison of Performances

System	Dominant complex poles $\zeta = 0.45$	Other roots	Ramp error coefficient K_1, s^{-1}	t_p, s	M_o	t_s, s
Original	$-0.40 \pm j0.8$	-5.19	0.84	4.12	0.202	9.48
Cascade lead	$-1.59 \pm j3.15$	-11.83	2.95	1.09	0.195	2.45
Rate feedback	$-1.0 \pm j2.0$	-4.0	1.54	1.87	0.17	3.95

K_1 must be obtained from the forward transfer function:

$$K_1 = \lim_{s\to 0} sG(s) = \lim_{s\to 0} \frac{sG_x(s)}{1 + G_x(s)H(s)} = \lim_{s\to 0} \frac{sK}{s[(s+1)(s+5) + KK_t]}$$
$$= \frac{K}{5 + KK_t} = 1.54 \tag{10.56}$$

The results in Table 10.3 show that the effects of rate feedback are similar to those of a cascade lead compensator, but the gain K_t is not so large (see Table 10.1). The control ratio of the system with rate feedback, with $K = 20$ and $K_t = 0.4$, is

$$\frac{C(s)}{R(s)} = \frac{20}{(s + 1 - j2)(s + 1 + j2)(s + 4)} \tag{10.57}$$

The performance characteristics for various compensation methods are summarized in Tables 10.1 and 10.5.

10.18 DESIGN OF RATE FEEDBACK

The more general case of feedback compensation shown in Fig. 10.29, with $H(s) = K_t s$ and the amplifier gain A not equal to unity, is considered in this

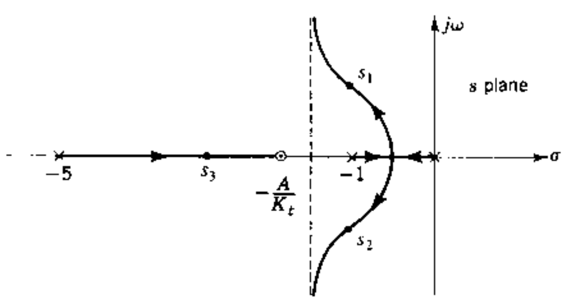
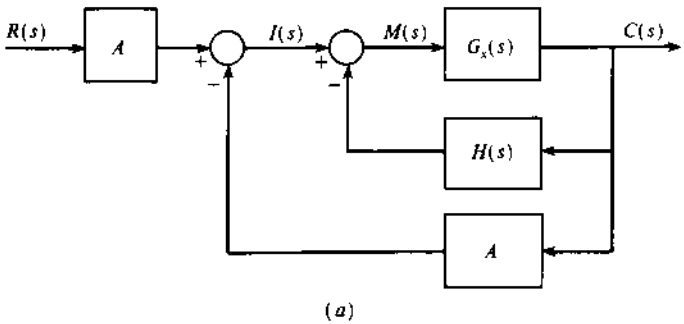


FIGURE 10.31 Root locus for Eq. (10.60).



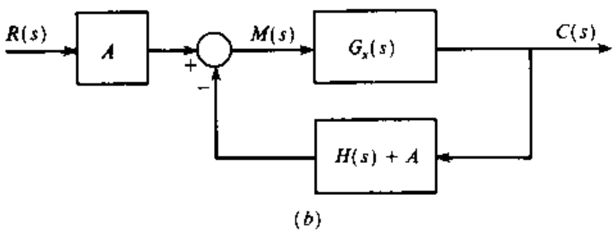


FIGURE 10.32 Block diagrams equivalent to Fig. 10.29.

section. Two methods of attacking this problem are presented. The block diagrams of Fig. 10.32*a* and *b* are mathematically equivalent to the block diagram of Fig. 10.29 since they all yield the same closed-loop transfer function:

$$\frac{C(s)}{R(s)} = \frac{AG_x(s)}{1 + G_x(s)H(s) + AG_x(s)} = \frac{AG_x(s)}{1 + G_x(s)[H(s) + A]} \tag{10.58}$$

METHOD 1. The characteristic equation obtained from the denominator of Eq. (10.58) for the plant given by Eq. (10.49) is

$$1 + G_x(s)[H(s) + A] = 1 + \frac{K(K_t s + A)}{s(s+1)(s+5)} = 0 \tag{10.59}$$

Partitioning the characteristic equation to obtain the general format $F(s) = -1$, which is required for the root locus, yields

$$\frac{KK_t(s + A/K_t)}{s(s+1)(s+5)} = -1 \tag{10.60}$$

Since the numerator and denominator are factored, the roots can be obtained by plotting a root locus. The angle condition of $(1 + 2h)180^\circ$ must be satisfied for the root locus of Eq. (10.60). It should be kept in mind, however, that Eq. (10.60) is *not* the forward transfer function of this system; *it is just an equation whose roots are the poles of the control ratio.* The three roots of this characteristic equation vary as K, K_t, and A are varied. Obviously, it is necessary to fix two of these quantities and to determine the roots as a function of the third. A sketch of the root locus as a function of K is shown in Fig. 10.31 for arbitrary values of A and K_t. Note that rate feedback results in an equivalent open-loop zero, so the system is stable for all values of gain for the value of A/K_t shown. This observation is typical of the stabilizing effect that can be achieved by derivative feedback and corresponds to the effect achieved with a proportional plus derivative (PD) compensator in cascade.

The angles of the asymptotes of two of the branches of Eq. (10.60) are $\pm 90^\circ$. For the selection A/K_t shown in Fig. 10.31, these asymptotes are in the left half of the s plane, and the system is stable for any value of K. However, if $A/K_t > 6$, the asymptotes are in the right half of the s plane and the system can become unstable if K is large enough. For the best transient response the root locus should be moved as far to the left as possible. This is determined by the location of the zero $s = -A/K_t$. To avoid having a dominant real root, the value of A/K_t is made greater than unity.

The forward transfer function obtained from Fig. 10.29 is

$$G(s) = \frac{C(s)}{E(s)} = \frac{AG_x(s)}{1 + G_x(s)H(s)} = \frac{AK}{s(s+1)(s+5) + KK_t s} \tag{10.61}$$

The ramp error coefficient obtained from Eq. (10.61) is

$$K_1' = \frac{AK}{5 + KK_t} \tag{10.62}$$

Selection of the zero ($s = -A/K_t$), the sensitivity KK_t for the root locus drawn for Eq. (10.60), and the ramp error coefficient of Eq. (10.62) are all interrelated. The designer therefore has some flexibility in assigning the values of A, K_t, and K. The values $A = 20$ and $K_t = 0.8$ result in a zero at $s = -2.5$, the same as the example in Sec. 10.17. For $\zeta = 0.45$, the control ratio is given by Eq. (10.57) and the performance is the same as that listed in Table 10.3. The use of the additional amplifier A does not change the performance; instead, it permits greater flexibility in the allocation of gains. For example, K can be decreased and both A and K_t increased by a corresponding amount with no change in performance.

METHOD 2. A more general design method is to start with the inner loop, adjusting the gain to select the poles. Then successive loops are adjusted, the poles for each loop being set to desired values. For the system of Eq. (10.49), the transfer function of the inner loop is

$$\frac{C(s)}{I(s)} = \frac{K}{s(s+1)(s+5) + KK_t s} = \frac{K}{s[(s+1)(s+5) + KK_t]}$$
$$= \frac{K}{s(s-p_1)(s-p_2)} \tag{10.63}$$

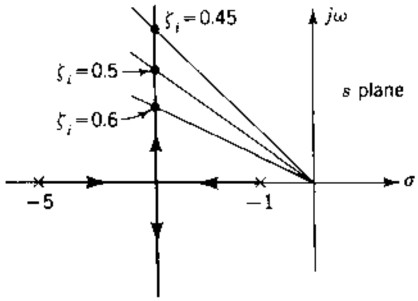


FIGURE 10.33 Poles of the inner loop for several values of ζ_i.

The equation for which the root locus is plotted to find the roots p_1 and p_2 is

$$(s+1)(s+5) + KK_t = 0$$

which can be written as

$$\frac{KK_t}{(s+1)(s+5)} = -1 \tag{10.64}$$

The locus is plotted in Fig. 10.33, where several values of the inner-loop damping ratio ζ_i for the roots are shown. Selection of a ζ_i permits evaluation of the product KK_t by use of Eq. (10.64). Upon assignment of a value to one of these quantities, the other can then be determined. The poles of the inner loop are the points on the locus that are determined by the ζ_i selected. They also are the poles of the open-loop transfer function for the overall system as represented in Eq. (10.63). It is therefore necessary that the damping ratio ζ_i of the inner loop have a higher value than the desired damping ratio of the overall system. The forward transfer function of the complete system becomes

$$G(s) = \frac{C(s)}{E(s)} = \frac{AK}{s(s^2 + 2\zeta_i\omega_{n,i}s + \omega_{n,i}^2)} \tag{10.65}$$

The root locus of the complete system is shown in Fig. 10.34 for several values of the inner-loop damping ratio ζ_i. These loci are plotted as a function of AK.

The control ratio has the form

$$\begin{aligned}\frac{C(s)}{R(s)} &= \frac{AK}{s(s^2 + 2\zeta_i\omega_{n,i}s + \omega_{n,i}^2) + AK} \\ &= \frac{AK}{(s+\alpha)(s^2 + 2\zeta\omega_n s + \omega_n^2)}\end{aligned} \tag{10.66}$$

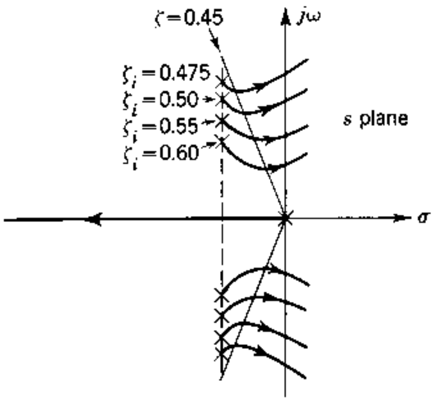


FIGURE 10.34 Root locus of complete system for several values of ζ_i.

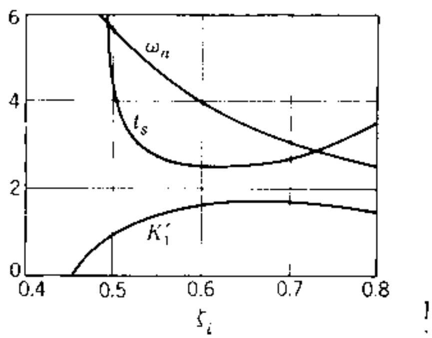


FIGURE 10.35 Variation of K_1',t_s, and ω_n with ζ_i.

In this problem the damping ratio ζ of the complex poles of the control ratio is used as the basis for design. Both the undamped natural frequency ω_n and the ramp error coefficient $K_1' = AK/(5 + KK_t)$ depend on the selction of the inner-loop damping ratio ζ_i. Several trials are necessary to show how these quantities and the remaining figures of merit vary with the selection of ζ_i. This information is shown graphically in Fig. 10.35. The system performance specifications must be used to determine a suitable design. The associated value of K_1' determines the required values of A, K and K_t. The results in Table 10.4 show that $\zeta_i = 0.6$ produces the shortest settling time $t_s = 2.19$, an overshoot $M_o = 0.01$ and $K_1' = 1.54$. Table 10.4 also shows for $\zeta < 0.6$ that a real root is most dominant, which yields an overdamped system with an increased value of t_s.

TABLE 10.4 System Response Characteristics Using Method 2

ζ_i	KK_t	AK	Roots	t_p, s	M_o	t_s, s	K_1'
0.475	34.91	19.33	$-2.74 \pm j5.48$ -0.52	†	0	7.64	0.484
0.5	31.01	30.53	$-2.5 \pm j4.98$ -0.93	†	0	4.34	0.847
0.55	24.76	39.0	$-2.1 \pm j4.22$ -1.75	†	0	2.36	1.31
0.6	20.01	38.63	$-1.82 \pm j3.62$ -2.35	1.40	0.01	2.19	1.54
0.7	13.37	30.51	$-1.38 \pm j2.74$ -3.24	1.52	0.12	2.88	1.66
0.8	9.06	22.14	$-1.03 \pm j2.14$ -3.83	1.77	0.20	3.75	1.57

†Overdamped response.

10.19 DESIGN: FEEDBACK OF SECOND DERIVATIVE OF OUTPUT

To improve the system performance further, a signal proportional to the second derivative of the output may be fed back. For a position system an accelerometer will generate such a signal. An approximation to this desired signal may be generated by modifying the output of a rate transducer with a high-pass filter. An example of this technique for a position-control system is shown in Fig. 10.36.

The transfer function of the inner feedback, which consists of a tachometer in cascade with a lead compensator, is

$$H(s) = \frac{B(s)}{C(s)} = K_t s\left(\frac{s}{s+1/RC}\right) = \frac{K_t s^2}{s+1/RC} \tag{10.67}$$

Method 1 is used for the application of this compensator. Using the denominator of Eq. (10.58), the root locus is drawn for $G_x(s)[H(s)+A] = -1$, which yields

$$\frac{KK_t(s^2 + As/K_t + A/K_tRC)}{s(s+1)(s+5)(s+1/RC)} = -1 \tag{10.68}$$

Since A/K_t and $1/RC$ must be selected in order to specify the zeros in Eq. (10.68) before the locus can be drawn, it is well to look at the expression for K_1' to determine the limits to be placed on these values. The forward transfer

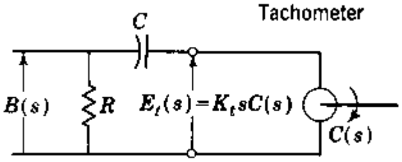


FIGURE 10.36 Tachometer and *RC* feedback network.

function is

$$G(s) = \frac{AG_x(s)}{1 + G_x(s)H(s)} = \frac{AK(s + 1/RC)}{s(s+1)(s+5)(s+1/RC) + KK_t s^2} \tag{10.69}$$

Since it is a Type 1 transfer function, the gain K_1' is equal to

$$K_1' = \lim_{s\to 0} sG(s) = \frac{AK}{5} \tag{10.70}$$

Since K_1' is independent of RC, the value of RC is chosen to produce the most desirable root locus. The two zeros of $s^2 + (A/K_t)s + A/K_tRC$ may be either real or complex and are equal to

$$s_{a,b} = -\frac{A}{2K_t} \pm \sqrt{\left(\frac{A}{2K_t}\right)^2 - \frac{A}{K_tRC}} \tag{10.71}$$

A representative sketch of the root locus is shown if Fig. 10.37 for an arbitrary selection of A, K_t, and RC. For Fig. 10.37 the values $A=1$, $K_t=0.344$, and $1/RC=0.262$ give zeros at $s_a=-0.29$ and $s_b=-2.62$. Note that the zero s_a and the pole $s=-1/RC$ have the properies of a cascade lag compensator; therefore, an increase in gain can be expected. The zero s_b results in an improved time response. By specifying a damping ratio $\zeta = 0.35$ for the dominant roots, the result is $K=52.267$, $K_1' = 10.45$, and $\omega_n = 3.75$; and the overall

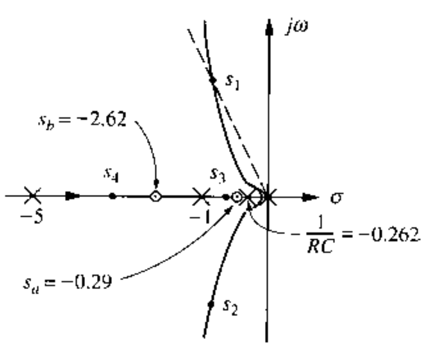


FIGURE 10.37 Root locus of Eq. (10.68).

TABLE 10.5 Comparison of Cascade and Feedback Compensation Using the Root Locus

System	K_t	ω_n	ζ	M_o	t_s, s
Uncompensated	0.84	0.89	0.45	0.202	9.48
Rate feedback	1.54	4.0	0.45	0.01	2.19
Second derivative of output	10.45	3.75	0.35	0.267	6.74
Cascade lag compensator	8.02	0.854	0.45	0.266	22.0
Cascade lead compensator	2.95	3.53	0.45	0.195	2.45
Cascade lag-lead compensator	29.26	3.49	0.45	0.213	3.33

control ratio is

$$\frac{C(s)}{R(s)} = \frac{52.267(s + 0.262)}{(s + 1.31 + j3.51)(s + 1.31 - j3.51)(s + 3.35)(s + 0.291)} \tag{10.72}$$

Note that a passive lead network in the feedback loop acts like a passive lag network in cascade and results in a large increase in K_1'. Several trials may be necessary to determine the best selection of the closed-loop poles and ramp error coefficient. A $\zeta = 0.35$ is chosen instead of $\zeta = 0.45$ to obtain a larger value for K_1'. Since the third pole $s = -3.34$ reduces the peak overshoot, a smaller value of ζ is justifiable; however, the peak overshoot and settling time are larger than the best results obtained with just rate feedback (see Table 10.5).

The use of an RC network in the feedback path may not always produce a desired increase in gain. Also, selecting the proper location of the zeros and the corresponding values of A, K_t, and RC in Eq. (10.68) may require several trials. Therefore, a possible alternative is to use only rate feedback to improve the time response, and then to add a cascade lag compensator to increase the gain. The corresponding design steps are essentially independent.

10.20 RESULTS OF FEEDBACK COMPENSATION DESIGN

The preceding sections show the application of feedback compensation by the root-locus method. Basically, rate feedback is similar to ideal derivative control in cascade. The addition of a high-pass RC filter in the feedback path permits a large increase in the error coefficient. For good results the feedback compensator $H(s)$ should have a zero, $s = 0$, of higher order than the type of the original transfer function $G_x(s)$. The method of adjusting the poles of the inner loop and then adjusting successively larger loops is applicable to all multiloop systems. This method is used extensively in the design of aircraft flight control systems [12]. The results of feedback

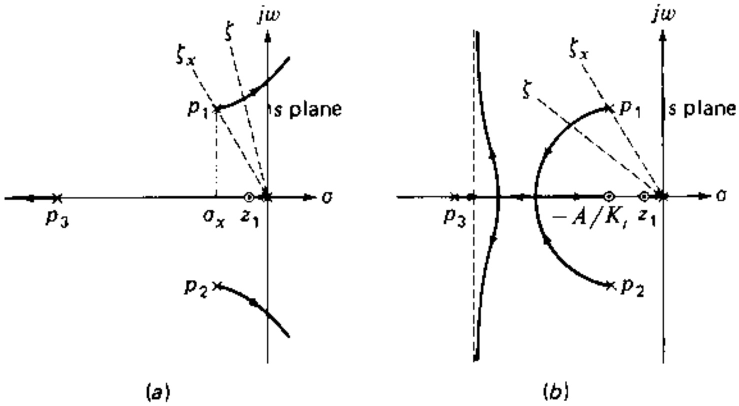


FIGURE 10.38 Root locus for the longitudinal motion of an aircraft: (*a*) uncompensated; (*b*) compensated (rate feedback).

compensation applied in the preceding sections are summarized in Table 10.5. Results of cascade compensation are also listed for comparison. This tabulation shows that comparable results are obtained by cascade and feedback compensation.

10.21 RATE FEEDBACK: PLANTS WITH DOMINANT COMPLEX POLES

The simple cascade compensators of Secs. 10.8 to 10.12 yield very little system performance improvement for plants having the pole-zero pattern of Fig. 10.38*a* where the plant has dominant complex poles. The following transfer function, representing the longitudinal motion of an aircraft [12], yields such a pattern:

$$G_x(s) = \frac{\theta(s)}{e_g(s)} = \frac{K_x(s - z_1)}{s(s - p_1)(s - p_2)(s - p_3)} \tag{10.73}$$

where θ is the pitch angle and e_g is the input to the elevator servo. With unity feedback the root locus of Fig. 10.38*a* shows, for $K_x > 0$ that only $\zeta < \zeta_x$ is achievable for the dominant roots. However, good improvement in system performance is obtained by using a minor loop with rate feedback $H(s) = K_t s$. Using method 1 of Sec. 10.18 for the system configuration of Fig. 10.29 yields

$$\frac{C(s)}{E(s)} = \frac{AK_x(s - z_1)}{s(s - p_1)(s - p_2)(s - p_3) + K_t K_x s(s - z_1)} \tag{10.74}$$

and

$$\frac{C(s)}{R(s)} = \frac{AK_x(s - z_1)}{s(s - p_1)(s - p_2)(s - p_3) + K_t K_x s(s - z_1) + AK_x(s - z_1)} \tag{10.75}$$

The characteristic equation formed from the denominator of Eq. (10.58), or from the denominator of Eq. (10.75), is partitioned to yield

$$\frac{K_t K_x (s - z_1)(s + A/K_t)}{s(s - p_1)(s - p_2)(s - p_3)} = -1 \tag{10.76}$$

A comparison of Eqs. (10.73) and (10.76) shows that rate feed back has added a zero to the root locus. One possible root locus, depending on the value of A/K_t, is shown in Fig. 10.38*b*. This illustrates that using rate feedback with $K_t > 0$ allows a good improvement in system performance, since it is possible to achieve $\zeta > \zeta_x$ and $\sigma < \sigma_x$ for the complex roots. The real root near the origin may appear to be dominant, but it is close to the zero z_1. Therefore, it may not make a large contribution to the transient response.

Note that the cascade compensator of Eq. (10.43), as applied in Fig. 10.25, shows exact cancellation of the pair of complex poles by the compensator zeros. In practice, exact cancellation cannot be achieved because the aircraft poles vary with different flight conditions. Thus, as illustrated by this section, using simple rate feedback is a more practical approach for achieving improved system performance.

10.22 SUMMARY

This chapter shows the applications of cascade and feedback compensation. The procedures for applying the compensators are given so that they can be used with any system. They involve a certain amount of trial and error and the exercise of judgment based on past experience. The improvements produced by each compensator are shown by application to a specific system. In cases where the conventional compensators do not produce the desired results, other methods of applying compensators are available in the literature.

It is important to emphasize that the conventional compensators cannot be applied indiscriminately. For example, the necessary improvements may not fit the four categories listed in Sec. 10.5, or the use of the conventional compensator may cause a deterioration of other performance characteristics. In such cases the designer must use some ingenuity in locating the compensator poles and zeros to achieve the desired results. The use of several compensating networks in cascade and/or in feedback may be required to achieve the desired root locations.

Realization of compensator pole-zero locations by means of passive networks requires an investigation of network synthesis techniques. Design procedures for specialized networks that produce two poles and two zeros are available [8, 9, 13].

It may occur to the reader that the methods of compensation in this and the following chapters can become laborious because some trial and error is involved. The use of an interactive CAD program to solve design problems minimizes the work involved. Using a computer means that an acceptable design of the compensator can be determined in a shorter time. It is still useful to obtain a few values of the solution, by an analytical or graphical method, to ensure that the problem has been properly set up and that the CAD program is operating correctly.

REFERENCES

1. Chu, Y.: "Synthesis of Feedback Control System by Phase-Angle Loci," *Trans. AIEE*, vol. 71. pt. II, pp. 330–339, November 1952.
2. Gibson, J. E., *Nonlinear automatic control*, McGraw-Hill, New York, 1963.
3. Thaler, G. J., and M. P. Pastel: *Analysis and Design of Nonlinear Feedback Control Systems*, McGraw-Hill, New York, 1960.
4. Elgerd, O. I., and W. C. Stephens: "Effect of Closed-Loop Transfer Function Pole and Zero Locations on the Transient Response of Linear Control Systems," *Trans. AIEE*, vol. 78, pt. II, pp. 121–127, May 1959.
5. Ali, W., and J. H. Burghart: *Effects of Lag Controllers on Transient Response*, IEE Proc. Part D, vol. 138, no. 2, pp. 119–122, March 1991.
6. Ross, E. R., T. C. Warren, and G. J. Thaler: "Design of Servo Compensation Based on the Root Locus Approach," *Trans. AIEE*, vol. 79, pt. II, pp. 272–277, September 1960.
7. Pena, L. Q.: "Designing Servocompensators Graphically," *Control Eng.*, vol 82, pp. 79–81, January 1964.
8. Lazear, T. J., and A. B. Rosenstein: "On Pole-Zero Synthesis and the General Twin-T," *Trans. IEEE*, vol. 83, pt. II, pp. 389–393, November 1964.
9. Mitra, S. K.: *Analysis and Synthesis of Linear Active Networks*, Wiley, New York, 1969.
10. Truxal, J.G.: *Automatic Feedback Control System Synthesis*, McGraw-Hill, New York, 1955.
11. D'Azzo, J. J., and C. H. Houpis: *Feedback Control System Analysis and Synthesis*. 2nd ed., McGraw-Hill, New York, 1966.
12. Blakelock, J. H.: *Automatic Control of Aircraft and Missiles*, 2nd ed. Wiley, New York, 1991.
13. Kuo, F. F.: *Network Analysis and Synthesis*, 2nd ed. Wiley, New York, 1966.

11

Frequency-Response Compensation Design

11.1 INTRODUCTION TO FEEDBACK COMPENSATION DESIGN

Improvements in the response of feedback tracking control systems and disturbance rejection can be achieved in the frequency domain. This chapter presents frequency-domain tracking techniques, and Chap. 12 presents a disturbance-rejection technique. The preceding chapter presents the factors used with the root-locus method for selecting compensators and the corresponding effects on the time response. The designer can place the poles and zeros of the proposed compensator on the s plane and determine the poles of the closed-loop tracking system, which in turn permits evaluation of the closed-loop time response. Similarly, new values of M_m, ω_m, and K_m can be determined by the frequency-response method, using either the polar plot or the log plot. However, the closed-loop poles are not explicitly determined. Furthermore, the correlation between the frequency-response parameters M_m and ω_m and the time response is only qualitative, as discussed in Secs. 9.3, 9.4, and 9.11. The presence of real roots near the complex dominant roots further changes the correlation between the frequency-response parameters and the time response, as discussed in Sec. 9.11.

This chapter presents the changes that can be made in the frequency-response characteristics by use of the three types of cascade compensators (lag, lead, and lag-lead) and by feedback compensation. They are only representative of the compensators that can be used. Design procedures used to obtain improvement of the system performance are described. A designer can extend or modify these design procedures, which are based on the presence of one pair of complex dominant roots for the closed-loop system, to those systems where there are other dominant roots besides the main complex pair [1]. Both the log plots and the polar plots are used in this chapter in applying the frequency-response compensation criteria to show that either type of plot can be used; the choice depends on individual preference.

The performance of a closed-loop system can be described in terms of M_m, ω_m, and the system error coefficient K_m. The value of M_m (or γ) essentially describes the damping ratio ζ and therefore the amount of overshoot in the transient response. For a specific value of M_m, the resonant frequency ω_m (or ω_ϕ) determines the undamped natural frequency ω_n, which in turn determines the response time of the system. The system error coefficient K_m is important because it determines the steady-state error with an appropriate standard input. The design procedure is usually based on selecting a value of M_m and using the methods described in Chap. 9 to find the corresponding values of ω_m and the required gain K_m. If the desired performance specifications are not met by the basic system after this is accomplished, compensation must be used. This alters the shape of the frequency-response plot in an attempt to meet the performance specifications. Also, for those systems that are unstable for all values of gain, it is mandatory that a stabilizing or compensating network be inserted in the system. The compensator may be placed in cascade or in a minor feedback loop. The reasons for reshaping the frequency-response plot generally fall into the following categories:

1. A given system is stable, and its M_m and ω_m (and therefore the transient response) are satisfactory, but its steady-state error is too large. The gain K_m must therefore be increased to reduce the steady-state error (see Chap. 6) without appreciably altering the values of M_m and ω_m. It is shown later that in this case the high-frequency portion of the frequency-response plot is satisfactory, but the low-frequency portion is not.
2. A given system is stable, but its transient settling time is unsatisfactory; i.e., the M_m is set to a satisfactory value, but the ω_m is too low. The gain may be satisfactory, or a small increase may be desirable. The high-frequency portion of the frequency-response plot must be altered in order to increase the value of ω_m.
3. A given system is stable and has a desired M_m, but both its transient response and its steady-state response are unsatisfactory. Therefore,

the values of both ω_m and K_m must be increased. The portion of the frequency plot in the vicinity of the $-1 + j0$ (or -180°, 0-dB) point must be altered to yield the desired ω_m, and the low-frequency portion must be changed to obtain the increase in gain desired.

4. A given system is unstable for all values of gain. The frequency-response plot must be altered in the vicinity of the $-1 + j0$ or (-180°, 0-dB) point to produce a stable system with a desired M_m and ω_m.

Thus, the objective of compensation is to reshape the frequency-response plot of the basic system by means of a compensator in order to achieve the performance specifications. Examples that demonstrate this objective are presented in the following sections. The design methods in this chapter and the effects of compensation are illustrated by the use of frequency-response plots in order to provide a means for understanding the principles involved. It is expected, however, that CAD programs such as MATLAB and TOTAL-PC will be used extensively in the design process (see Sec. 10.6). Such a program provides both the frequency response and the time response of the resultant system.

11.2 SELECTION OF A CASCADE COMPENSATOR

Consider the system shown in Fig. 11.1

$\mathbf{G}_x(j\omega)$ = forward transfer function of a basic feedback control system

$\mathbf{G}(j\omega) = \mathbf{G}_c(j\omega)\mathbf{G}_x(j\omega)$ = forward transfer function that results in the required stability and steady-state accuracy

Dividing $\mathbf{G}(j\omega)$ by $\mathbf{G}_x(j\omega)$ gives the necessary cascade compensator

$$\mathbf{G}_c(j\omega) = \frac{\mathbf{G}(j\omega)}{\mathbf{G}_x(j\omega)} \tag{11.1}$$

A physical electrical network for the compensator described in Eq. (11.1) can be synthesized by the techniques of Foster, Cauer, Brune, and Guillemin.

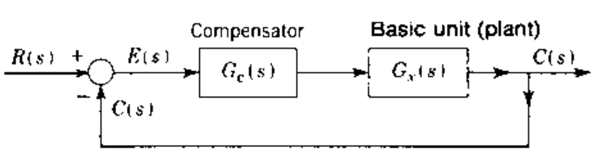


FIGURE 11.1 Cascade compensation.

In the preceding and present chapters the compensators are limited to relatively simple *RC* networks. Based on the design criteria presented, the parameters of these simple networks can be evaluated in a rather direct manner. The simple networks quite often provide the desired improvement in the control system's performance. When they do not, more complex networks must be synthesized [2,3]. Physically realizable networks are limited in their performance over the whole frequency spectrum; this makes it difficult to obtain the exact transfer-function characteristic given by Eq. (11.1). However, this is not a serious limitation, because the desired performance specifications are interpreted in terms of a comparatively small bandwidth of the frequency spectrum. A compensator can be constructed to have the desired magnitude and angle characteristics over this bandwidth. Minor loop feedback compensation can also be used to achieve the desired $\mathbf{G}(j\omega)$. This may yield a simpler network than the cascade compensator given by Eq. (11.1).

The characteristics of the lag and lead compensators and their effects upon a given polar plot are shown in Fig. 11.2*a* and *b*. The cascade lag compensator results in a clockwise (cw) rotation of $\mathbf{G}(j\omega)$ and can produce a large increase in gain K_m with a small decrease in the resonant frequency ω_m. The cascade lead compensator results in a counterclockwise (ccw) rotation of $\mathbf{G}(j\omega)$ and can produce a large increase in ω_m with a small increase in K_m. When the characteristics of both the lag and the lead networks are necessary to achieve the desired specifications, a lag-lead compensator can be used. The transfer function and polar plot of this lag-lead compensator are shown in Fig. 11.2*c*. The corresponding log plots for each of these compensators are included with the design procedures later in this chapter.

The choice of a compensator depends on the characteristics of the given (or basic) feedback control system and the desired specifications. When these networks are inserted in cascade, they must be of the (1) proportional plus integral and (2) proportional plus derivative types. The plot of the lag compensator shown in Fig. 11.2*a* approximates the ideal integral plus proportional characteristics shown in Fig. 11.3*a* at high frequencies. Also, the plot of the lead compensator shown in Fig. 11.2*b* approximates the ideal derivative plus proportional characteristics shown in Fig 11.3*b* at low frequencies. These are the frequency ranges where the proportional plus integral and proportional plus derivative controls are needed to provide the necessary compensation.

The equations of $\mathbf{G}_c(j\omega)$ in Figs. 11.2*a* and 11.2*b*, which represent the lag and the lead compensators used in this chapter, can be expressed as

$$\mathbf{G}_c(j\omega) = |\mathbf{G}_c(j\omega)| \underline{/\phi_c} \tag{11.2}$$

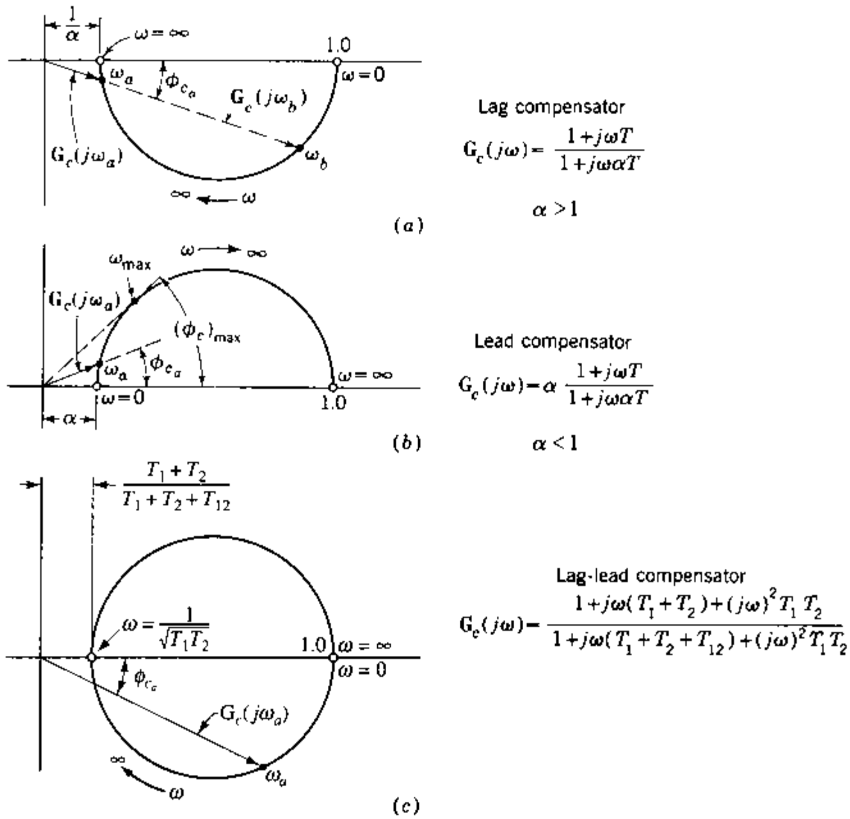


FIGURE 11.2 Polar plots: (*a*) lag compensator; (*b*) lead compensator; (*c*) lag-lead compensator.

Solving for the angle ϕ_c yields

$$\phi_c = \tan^{-1}\omega T - \tan^{-1}\omega\alpha T = \tan^{-1}\frac{\omega T - \omega\alpha T}{1+\omega^2\alpha T^2} \tag{11.3}$$

or

$$T^2 + \frac{\alpha - 1}{\omega\alpha\tan\phi_c}T + \frac{1}{\omega^2\alpha} = 0 \tag{11.4}$$

where $\alpha > 1$ for the lag compensator and $\alpha < 1$ for the lead compensator. In the design procedures shown in later sections, a frequency ω is specified at which a particular value of ϕ_c is desired. For the selected values of α, ϕ_c, and ω,

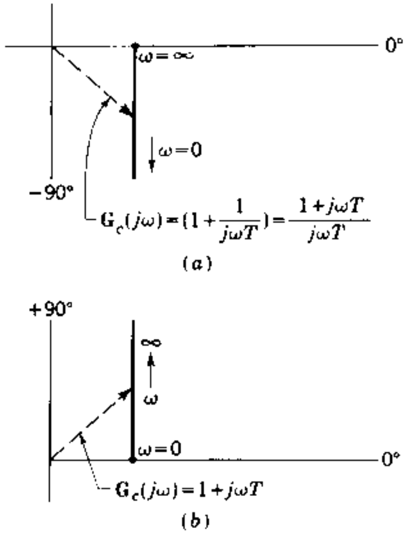


FIGURE 11.3 (*a*) Ideal integral plus proportional compensator. (*b*) Ideal derivative plus proportional compensator.

Eq. (11.4) is used to determine the value of T required for the compensator. For a lag compensator, a small value of ϕ_c is required, whereas a large value of ϕ_c is required for a lead compensator. For a lead compensator the value ω_{max}, at which the maximum phase shift occurs, can be determined for a given T and α by setting the derivative of ϕ_c with respect to the frequency ω equal to zero. Therefore, from Eq. (11.3) the maximum angle occurs at the frequency

$$\omega_{max} = \frac{1}{T\sqrt{\alpha}} \tag{11.5}$$

Inserting this value of frequency into Eq. (11.3) gives the maximum phase shift

$$(\phi_c)_{max} = \sin^{-1}\frac{1-\alpha}{1+\alpha} \tag{11.6}$$

Typical values for a lead compensator are

α	0	0.1	0.2	0.3	0.4	0.5
$(\phi_c)_{max}$	90°	54.9°	41.8°	32.6°	25.4°	19.5°

A knowledge of the maximum phase shift is useful in the application of lead compensators. The design procedures developed in the following sections utilize the characteristics of compensators discussed in this section.

11.3 CASCADE LAG COMPENSATOR

When M_m and ω_m, which determine the transient response of a feedback control system, are satisfactory but the steady-state error is too large, it is necessary to increase the gain without appreciably altering that portion of the given log magnitude (Lm)-angle plot in the vicinity of the $(-180^\circ, 0\text{-dB})$ point (category 1). This can be done by introducing an integral or lag compensator in cascade with the original forward transfer function as shown in Fig. 11.1. Figure 10.9 shows a representative lag network having the transfer function

$$\mathbf{G}_c(s) = A\frac{1+Ts}{1+\alpha Ts} \tag{11.7}$$

where $\alpha > 1$. As an example of its application, consider a basic Type 0 system. For an improvement in the steady-state accuracy, the system step error coefficient K_p must be increased. Its values before and after the addition of the compensator are

$$K_p = \lim_{s\to 0} \mathbf{G}_x(s) = K_0 \tag{11.8}$$

$$K_p' = \lim_{s\to 0} \mathbf{G}_c(s)\mathbf{G}_x(s) = AK_0 = K_0' \tag{11.9}$$

The compensator amplifier gain A must have a value that gives the desired increase in the value of the step error coefficient. Thus

$$A = \frac{K_0'}{K_0} \tag{11.10}$$

This increase in gain must be achieved while maintaining the same M_m and without appreciably changing the value of ω_m. However, the lag compensator has a negative angle, which moves the original Lm-angle plot to the left, or closer to the $(-180^\circ, 0\text{-dB})$ point. This has a destabilizing effect and reduces ω_m. To limit this decrease in ω_m, the lag compensator is designed so that it introduces a small angle, generally no more than -5°, at the original resonant frequency ω_{m1}. The value of -5° at $\omega' = \omega_{m1}$ is an empirical value, determined from practical experience in applying this type of network. A slight variation in method, which gives equally good results, is to select $\mathbf{G}_c(s)$ so that the magnitude of its angle is 5° or less at the original phase-margin frequency $\omega_{\phi 1}$.

The overall characteristics of lag compensators are best visualized from the log plot. The Lm and phase-angle equations for the compensator of Eq. (11.7) are

$$\mathrm{Lm}\mathbf{G}'_c(j\omega) = \mathrm{Lm}\ (1 + j\omega T) - \mathrm{Lm}\ (1 + j\omega\alpha T) \tag{11.11}$$

$$\angle \mathbf{G}_c(j\omega) = \angle 1 + j\omega T - \angle 1 + j\omega\alpha T \tag{11.12}$$

For various values of α there is a family of curves for the log magnitude and phase-angle diagram, as shown in Fig. 11.4. These curves show that an attenuation equal to Lm α is introduced above $\omega T = 1$. Assume that a system has the original forward transfer function $\mathbf{G}_x(j\omega)$ and that the gain has been adjusted for the desired phase margin γ. The Lm and phase-angle diagram of $\mathbf{G}_x(j\omega)$ are sketched in Fig. 11.5. The addition of the lag compensator $\mathbf{G}_c(j\omega)$ reduces the log magnitude by Lm α for the frequencies above $\omega = 1/T$. The value of T of the compensator is selected so that the attenuation Lm α occurs at the original phase-margin frequency $\omega_{\phi 1}$. Notice that the phase margin angle at $\omega_{\phi 1}$ has been reduced. To maintain the specified value of γ, the phase-margin frequency has been reduced to $\omega_{\phi 2}$. It is now possible to increase the gain of $\mathbf{G}_x(j\omega)\mathbf{G}'_c(j\omega)$ by selecting the value A so that the Lm curve will have the value 0 dB at the frequency $\omega_{\phi 2}$.

The effects of the lag compensator can now be analyzed. The reduction in log magnitude at the higher frequencies due to the compensator is desirable. This permits an increase in the gain to maintain the desired γ. Unfortunately,

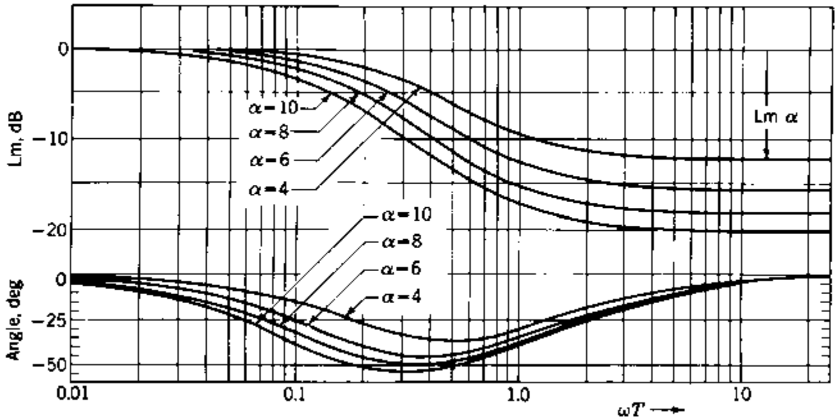


FIGURE 11.4 Log magnitude and phase-angle diagram for lag compensator $G'_c(j\omega) = \dfrac{1 + j\omega T}{1 + j\omega\alpha T}$.

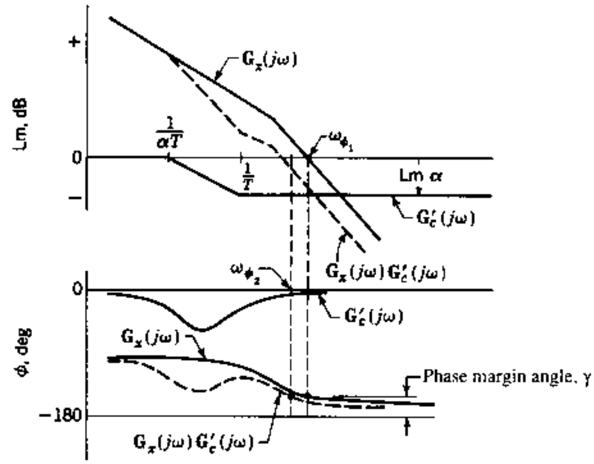


FIGURE 11.5 Original and lag-compensated log magnitude and phase-angle plots.

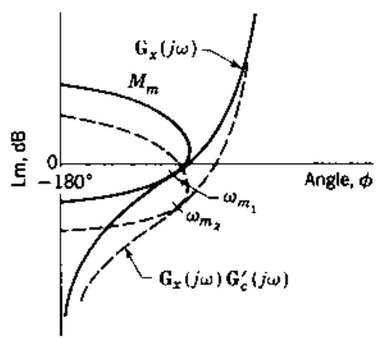


FIGURE 11.6 Original and lag-compensated Nichols plots.

the negative angle of the compensator lowers the angle curve. To maintain the desired γ, the phase-margin frequency is reduced. To keep this reduction small, the value of $\omega_{cf} = 1/T$ should be made as small as possible.

The Lm-angle diagram permits gain adjustment for the desired M_m. The curves for $\mathbf{G}_x(j\omega)$ and $\mathbf{G}_x(j\omega)\mathbf{G}_c'(j\omega)$ are shown in Fig. 11.6. The increase in gain to obtain the same M_m is the amount that the curve of $\mathbf{G}_x(j\omega)\mathbf{G}_c'(j\omega)$ must be raised in order to be tangent to the M_m curve. Because of the negative angle introduced by the lag compensator, the new resonant frequency ω_{m2} is smaller than the original resonant frequency ω_{m1}. Provided that the decrease in ω_m is small and the increase in gain is sufficient to meet the specifications, the system is considered satisfactorily compensated.

In summary, the lag compensator is basically a low-pass filter: the low frequencies are passed and the higher frequencies are attenuated. This attenuation characteristic of the lag compensator is useful and permits an increase in the gain. The negative phase-shift characteristic is detrimental to system performance but must be tolerated. Because the predominant and useful feature of the compensator is attenuation, a more appropriate name for it is *high-frequency attenuation compensator.*

11.4 DESIGN EXAMPLE: CASCADE LAG COMPENSATION

To give a basis for comparison of all methods of compensation, the examples of Chap. 10 are used in this chapter. For the unity-feedback Type 1 system of Sec. 10.8 the forward transfer function is

$$\mathbf{G}_x(s) = \frac{K_1}{s(1+s)(1+0.2s)} \tag{11.13}$$

By the methods of Secs. 9.7 and 9.10, the gain required for this system to have an M_m of 1.26 (2 dB) is $K_1 = 0.872\ \mathrm{s}^{-1}$, and the resulting resonant frequency is $\omega_{m1} = 0.715$ rad/s. With these values the transient response is considered to be staisfactory, but the steady-state error is too large. Putting a lag compensator in cascade with a basic system (category 1) permits an increase in gain and therefore reduces the steady-state error.

The cascade compensator to be used has the transfer function given by Eq. (11.7). Selection of an $\alpha = 10$ means that an increase in gain of almost 10 is desired. Based on the criterion discussed in the last section (a phase shift of -5° at the original ω_m), the compensator time constant can be determined. Note that in Fig. 11.4, $\phi_c = -5^\circ$ occurs for values of ωT approximately equal to 0.01 and 10. These two points correspond, respectively, to points ω_b and ω_a in Fig. 11.2*a*. The choice $\omega T = \omega_m T = 10$ or $\omega T = \omega_{\phi 1} T = 10$ (where $\omega_a = \omega_{\phi 1}$ or $\omega_a = \omega_{m1}$) is made because this provides the maximum attenuation at ω_{m1}. This ensures a maximum possible gain increase in $\mathbf{G}(j\omega)$ to achieve the desired M_m. Thus, for $\omega_{m1} = 0.72$, the value $T \approx 13.9$ is obtained. Note that the value of T can also be obtained from Eq. (11.4). For ease of calculation, the value of T is rounded off to 14.0, which does not appreciably affect the results.

In Figs. 11.7*a* and 11.7*b* the Lm and phase-angle diagrams for the compensator $\mathbf{G}'_c(j\omega)$ are shown, together with the curves for $\mathbf{G}'_x(j\omega)$. The sum of the curves is also shown and represents the function $\mathbf{G}'_c(j\omega)\mathbf{G}'_x(j\omega)$. The curves of $\mathbf{G}'_c(j\omega)\mathbf{G}'_x(j\omega)$ from Fig. 11.7 are used to draw the Lm–angle diagram of Fig. 11.8. By using this curve, the gain is adjusted to obtain an $M_m = 1.26$. The results are $K'_1 = 6.68$ and $\omega_m = 0.53$. Thus an improvement in K_1 has been obtained, but a penalty of a smaller ω_m results. An increase in

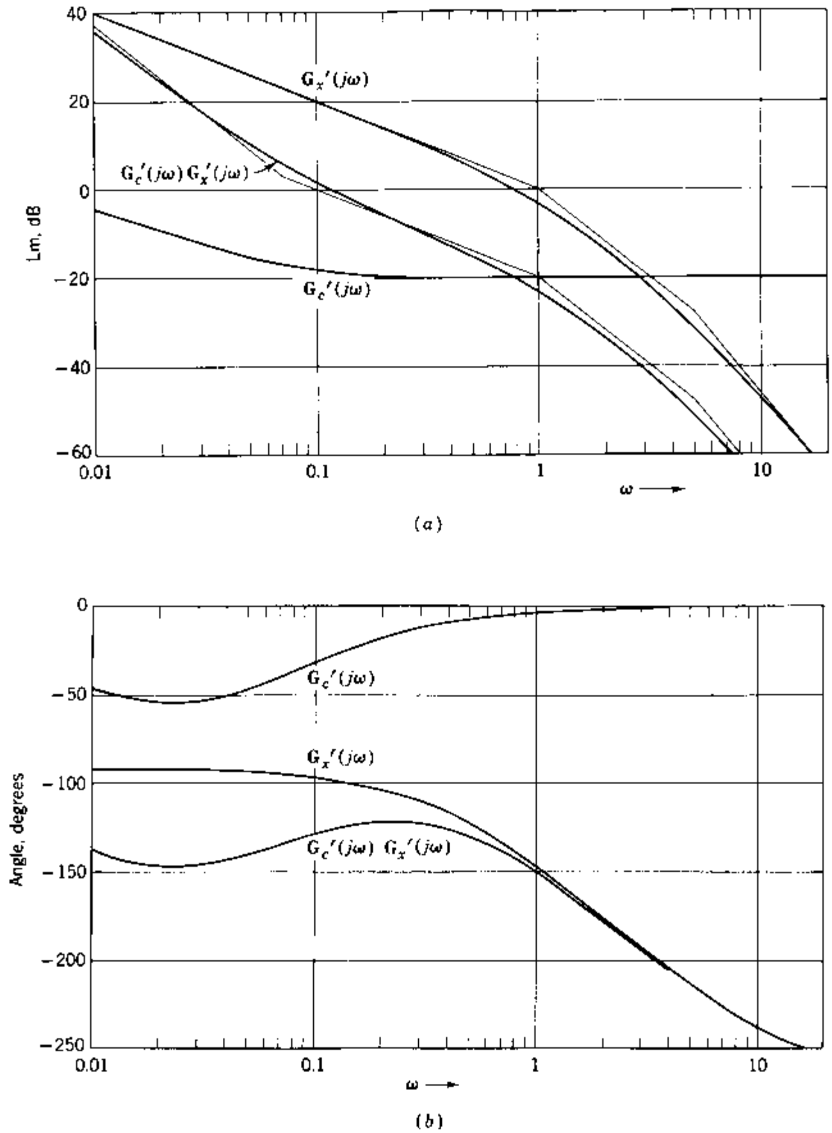


FIGURE 11.7 Log magnitude and phase-angle diagrams for

$$\mathbf{G}_x'(j\omega) = \frac{1}{j\omega(1 + j\omega)(1 + j0.2\omega)}$$

$$\mathbf{G}_c'(j\omega)\mathbf{G}_x'(j\omega) = \frac{1 + j14\omega}{j\omega(1 + j\omega)(1 + j0.2\omega)(1 + j140\omega)}.$$

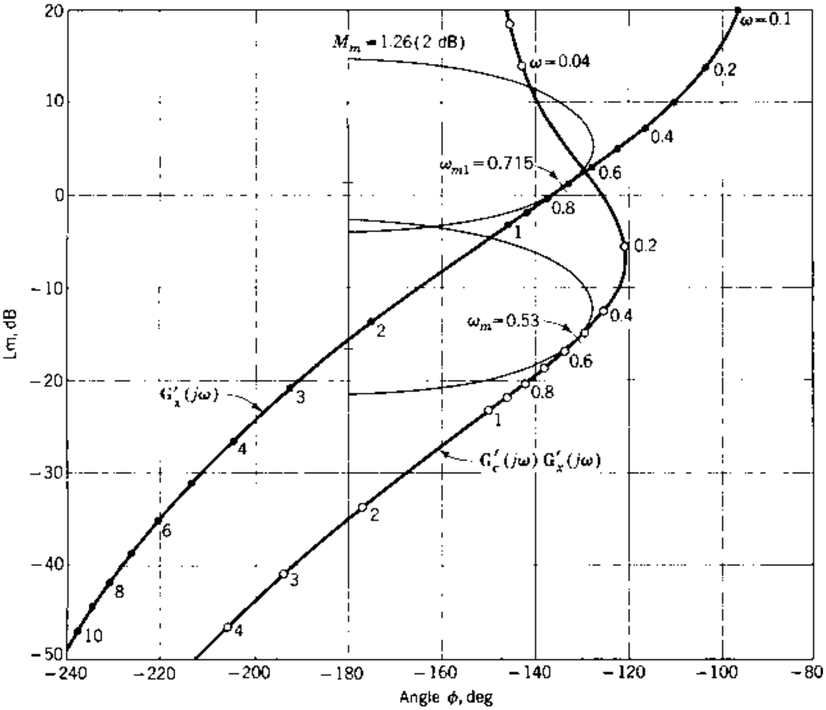


FIGURE 11.8 Lm-angle diagram: original and lag-compensated systems.

gain of approximately $A = 7.7$ results from the use of a lag compensator with $\alpha = 10$, which is typical of the gain increase permitted by a given α. This additional gain may be obtained from an amplifier already present in the basic system if it is not already operating at maximum gain. Note that the value of gain is somewhat less than the chosen value of α. Provided that the decrease in ω_m and the increase in gain are acceptable, the system has been suitably compensated.

The polar plots representing the basic and compensated systems shown in Fig. 11.9 also show the effects of compensation. Both curves are tangent to the circle representing $M_m = 1.26$. Note that there is essentially no change of the high-frequency portion of $\mathbf{G}(j\omega)$ compared with $\mathbf{G}_x(j\omega)$. However, in the low-frequency region there is a cw rotation of $\mathbf{G}(j\omega)$ due to the lag compensator.

Table 11.1 presents a comparison between the root-locus method and the frequency-plot method of applying a lag compensator to a basic control

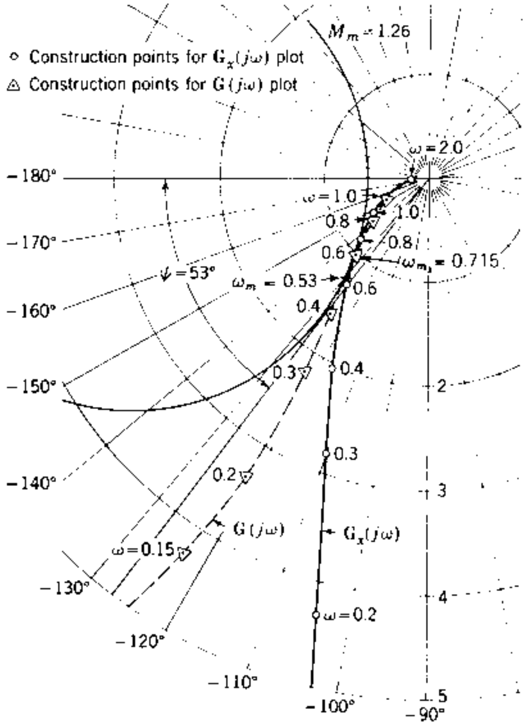


FIGURE 11.9 Polar plots of a Type 1 control system, with and without lag compensation:

$$\mathbf{G}_x(j\omega) = \frac{0.872}{j\omega(1 + j\omega)(1 + j0.2\omega)}$$

$$\mathbf{G}(j\omega) = \mathbf{G}_c(j\omega)\mathbf{G}_x(j\omega) = \frac{6.675(1 + j14\omega)}{j\omega(1 + j\omega)(1 + j0.2\omega)(1 + j140\omega)}.$$

system. Note the similarity of results obtained by these methods. The selection of the method depends upon individual preference. The compensators developed in the root-locus and frequency-plot methods are different, but achieve comparable results. To show this explicitly, Table 11.1 includes the compensator developed in the root-locus method as applied to the frequency-response method. In the frequency-response method, either the log or the polar plot may be used. An analytical technique for designing a lag or PI compensator, using both phase margin and a desired value of K_m, is presented in Ref. 4.

TABLE 11.1 Comparison Between Root-Locus and Frequency Methods of Applying a Cascade Lag Compensator to a Basic Control System

$G'_c(s)$	ω_{m1}	ω_m	ω_{n1}	ω_n	K_1	K'_1	M_m	M_p	t_p, s	t_s, s
		Root-locus method, ζ of dominant roots = 0.45								
$\dfrac{1+20s}{1+200s}$	—	0.66	0.898	0.854	0.840	8.02	1.31	1.266	4.30	22.0
	Frequency-response method, $M_m = 1.26$ (effective $\zeta = 0.45$)									
$\dfrac{1+14s}{1+140s}$	0.715	0.53	—	—	0.872	6.68	1.260	1.257	4.90	22.8
$\dfrac{1+20s}{1+200s}$	0.715	0.60	—	—	0.872	7.38	1.260	1.246	4.55	23.6

11.5 CASCADE LEAD COMPENSATOR

Figure 10.16*a* shows a lead compensator made up of passive elements that has the transfer function

$$\mathbf{G}'_c(j\omega) = \alpha \frac{1 + j\omega T}{1 + j\omega\alpha T} \qquad \alpha < 1 \tag{11.14a}$$

This equation is marked with a prime since it does not contain the gain A. The Lm and phase-angle equations for this compensator are

$$\text{Lm}\ \mathbf{G}'_c(j\omega) = \text{Lm}\,\alpha + \text{Lm}\,(1 + j\omega T) - \text{Lm}\,(1 + j\omega\alpha T) \tag{11.14b}$$

$$\angle \mathbf{G}'_c(j\omega) = \angle 1 + j\omega T - \angle 1 + j\omega\alpha T \tag{11.14c}$$

A family of curves for various values of α is shown in Fig. 11.10. It is seen from the shape of the curves that an attenuation equal to Lm α is introduced at frequencies below $\omega T = 1$. Thus, the lead network is basically a high-pass filter: the high frequencies are passed, and the low frequencies are attenuated. Also, an appreciable lead angle is introduced in the frequency range from $\omega = 1/T$ to $\omega = 1/\alpha T$ (from $\omega T = 1$ to $\omega T = 1/\alpha$). Because of its angle characteristic a lead network can be used to increase the bandwidth for a system that falls in category 2.

Application of the lead compensator can be based on adjusting the γ and ω_ϕ. Assume that a system has the original forward transfer function $\mathbf{G}_x(j\omega)$, and that the gain has been adjusted for the desired γ or for the desired M_m. The Lm and phase-angle diagram of $\mathbf{G}_x(j\omega)$ is sketched in Fig. 11.11. The purpose of the lead compensator is to increase ω_ϕ, and therefore to increase ω_m. The lead compensator introduces a positive angle over a relatively narrow bandwidth. By properly selecting the value of T,

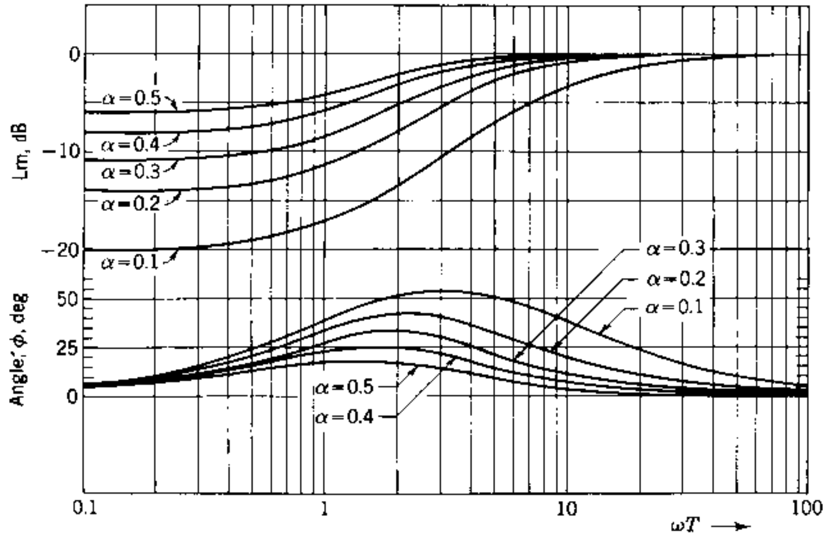


FIGURE 11.10 Log magnitude and phase-angle diagram for lead compensator $\mathbf{G}_c'(j\omega) = \alpha \dfrac{1 + j\omega T}{1 + j\omega\alpha T}$.

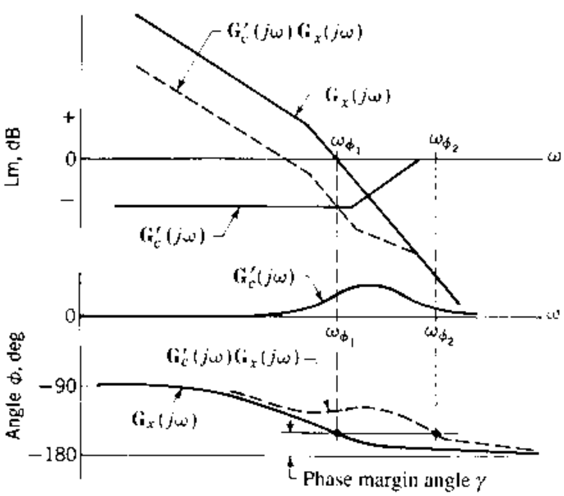


FIGURE 11.11 Original and lead-compensated log magnitude and phase-angle plots.

the phase-margin frequency can be increased from $\omega_{\phi 1}$ to $\omega_{\phi 2}$. Selection of T can be accomplished by physically placing the angle curve for the compensator on the same graph with the angle curve of the original system. The location of the compensator angle curve must produce the specified γ at the highest possible frequency; this location determines the value of the time constant T. The gain of $\mathbf{G}'_c(j\omega)\mathbf{G}_x(j\omega)$ must then be increased so that the Lm curve has the value 0 dB at the frequency $\omega_{\phi 2}$. For a given α, an analysis of the Lm curve of Fig. 11.10 shows that the farther to the right the compensator curves are placed, i.e., the smaller T is made, the larger the new gain of the system.

It is also possible to use the criterion derived in the previous chapter for the selection of T. This criterion is to select T, for Type 1 or higher systems, equal to or slightly smaller than the largest time constant of the original forward transfer function. This is the procedure used in the example of the next section. For a Type 0 system, the compensator time constant T is made equal to or slightly smaller than the second-largest time constant of the original system. Several locations of the compensator curves should be tested and the best results selected.

More accurate application of the lead compensator is based on adjusting the gain to obtain a desired M_m by use of the Lm-angle diagram. Figure 11.12 shows the original curve $\mathbf{G}_x(j\omega)$ and the new curve $\mathbf{G}'_c(j\omega)\mathbf{G}_x(j\omega)$. The increase in gain to obtain the same M_m is the amount that the curve $\mathbf{G}'_c(j\omega)\mathbf{G}_x(j\omega)$ must be raised to be tangent to the M_m curve. Because of the positive angle introduced by the lead compensator, the new resonant frequency ω_{m2} is larger than the original resonant frequency ω_{m1}. The increase in gain is not as large as that obtained by use of the lag compensator.

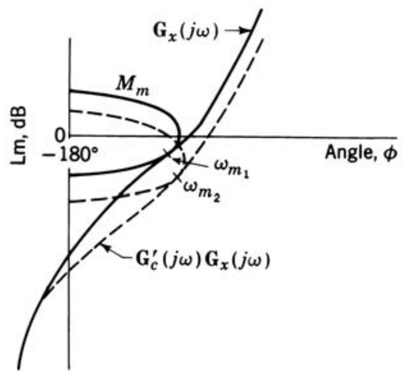


FIGURE 11.12 Original and lead-compensated log magnitude–angle curves.

11.6 DESIGN EXAMPLE: CASCADE LEAD COMPENSATION

For the unity-feedback Type 1 system of Sec. 10.8, the forward transfer function is

$$\mathbf{G}_x(s) = \frac{K_1}{s(1+s)(1+0.2s)} \tag{11.15}$$

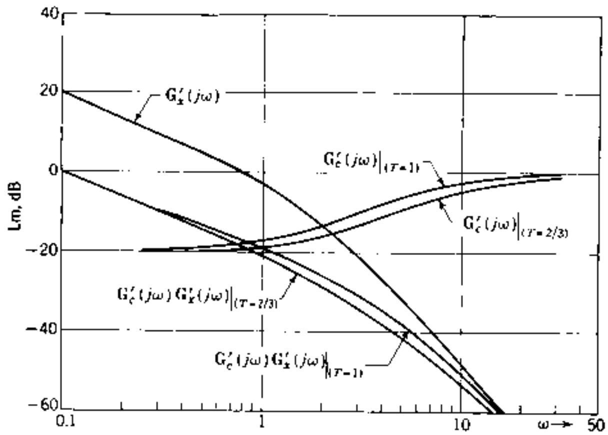
From Sec. 11.4 for $M_m = 1.26$ it is found that $K_1 = 0.872\ \mathrm{s}^{-1}$ and $\omega_{m1} = 0.715$ rad/s. Putting a lead compensator in cascade with the basic system increases the value of ω_m, and therefore improves the time response of the system. This situation falls in category 2.

The choice of values of α and T of the compensator must be such that the angle ϕ_c adds a sizeable positive phase shift at frequencies above ω_{m1} (or $\omega_{\phi 1}$) of the overall forward transfer function. The nominal value of $\alpha = 0.1$ is often used; thus only the determination of the value of T remains. Two values of T selected in Sec. 10.10 are used in this example so that a comparison can be made between the root-locus and frequency-response techniques. The results are used to establish the design criteria for lead compensators.

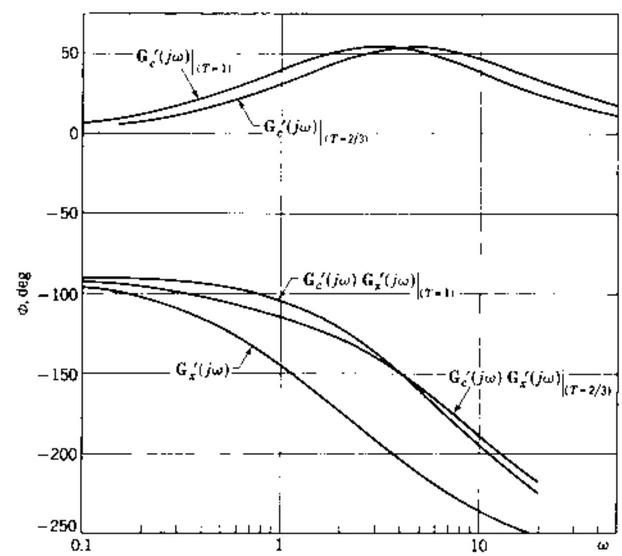
In Figs. 11.13*a* and 11.13*b* the Lm and phase-angle diagrams for the basic system $\mathbf{G}'_x(j\omega)$, the two compensators $\mathbf{G}'_c(j\omega)$, and the two composite transfer functions $\mathbf{G}_c(j\omega)\mathbf{G}'_x(j\omega)$ are drawn. Using the Lm and phase-angle curves of $\mathbf{G}'_c(j\omega)\mathbf{G}'_x(j\omega)$ from Fig. 11.13 the Lm-angle diagrams are drawn in Fig. 11.14. The gain is again adjusted to obtain M_m=1.26 (2 dB). The corresponding polar plots of the original system $\mathbf{G}_x(j\omega)$ and the compensated systems $\mathbf{G}(j\omega) = \mathbf{G}_c(j\omega)\mathbf{G}_x(j\omega)$ are shown in Fig. 11.15. Note that the value of ω_m is increased for both values of T.

Table 11.2 presents a comparison between the root-locus method and the frequency-response method of applying a lead compensator to a basic control system. The results show that a lead compensator increases the resonant frequency ω_m. However, a range of values of ω_m is possible, depending on the compensator time constant used. The larger the value of ω_m, the smaller the value of K_1. The characteristics selected must be based on the system specifications. The increase in gain is not as large as that obtained with a lag compensator. Note in the table the similarity of results obtained by the root-locus and frequency-response methods. Because of this similarity, the design rules of Sec. 10.10 can be interpreted in terms of the frequency-response method as follows.

RULE 1. The value of the time constant T in the numerator of $\mathbf{G}_c(j\omega)$ given by Eq. (11.14*a*) is made equal to the value of the largest time constant in the denominator of the original transfer function $\mathbf{G}_x(j\omega)$ for a Type 1 or higher system. This usually results in the largest increase in ω_m and ω_ϕ (best time response), and there is an increase in system gain K_m.



(*a*)



(*b*)

FIGURE 11.13 Log magnitude and phase-angle diagrams of

$$\mathbf{G}_x'(j\omega) = \frac{1}{j\omega(1+j\omega)(1+j0.2\omega)}$$

$$\mathbf{G}_c'(j\omega)\mathbf{G}_x'(j\omega) = \frac{0.1(1+j\omega T)}{j\omega(1+j\omega)(1+j0.2\omega)(1+j0.1\omega T)}.$$

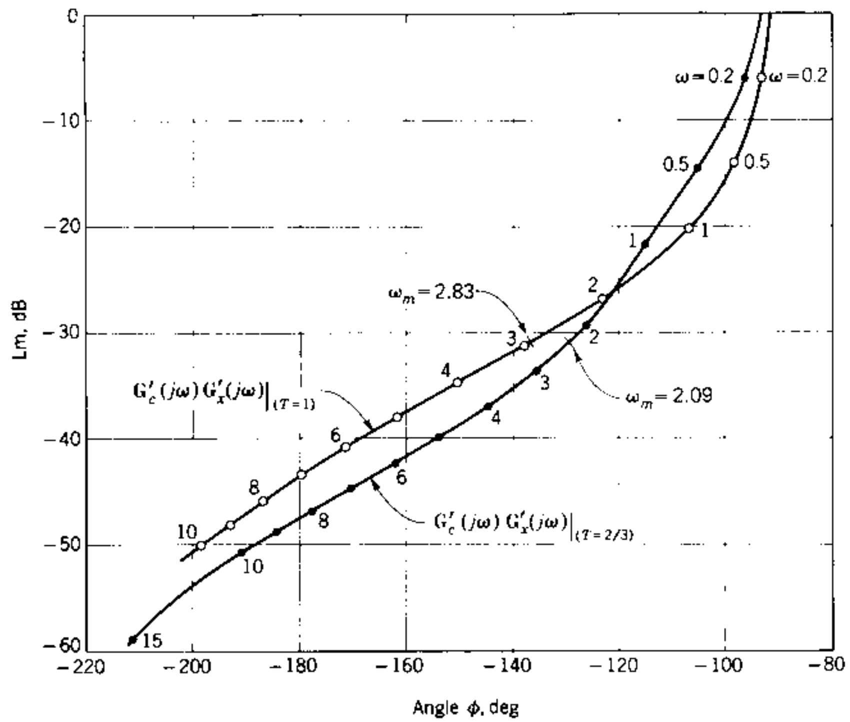


FIGURE 11.14 Log magnitude-angle diagrams with lead compensators from Figs. 11.13*a* and *b*.

RULE 2. The value of the time constant T in the numerator of $\mathbf{G}_c(j\omega)$ is made slightly smaller than the largest time constant in the denominator of the original transfer function for a Type 1 or higher system. This results in achieving the largest gain increase for the system with an appreciable increase in ω_m and ω_ϕ.

Rule 1 is the simplest to apply to the design of the lead compensator. Where both a maximum gain increase and a good improvement in the time of response are desired, Rule 2 is applicable. For a Type 0 system, Rules 1 and 2 are modified so that the lead-compensator time constant is selected either equal to or slightly smaller than the second-largest time constant of the original system.

Remember that the lag and lead networks are designed for the same basic system; thus from Tables 11.1 and 11.2 it is seen that both result in a gain increase. The distinction between the two types of compensation is that the lag network gives the largest increase in gain (with the best improvement in

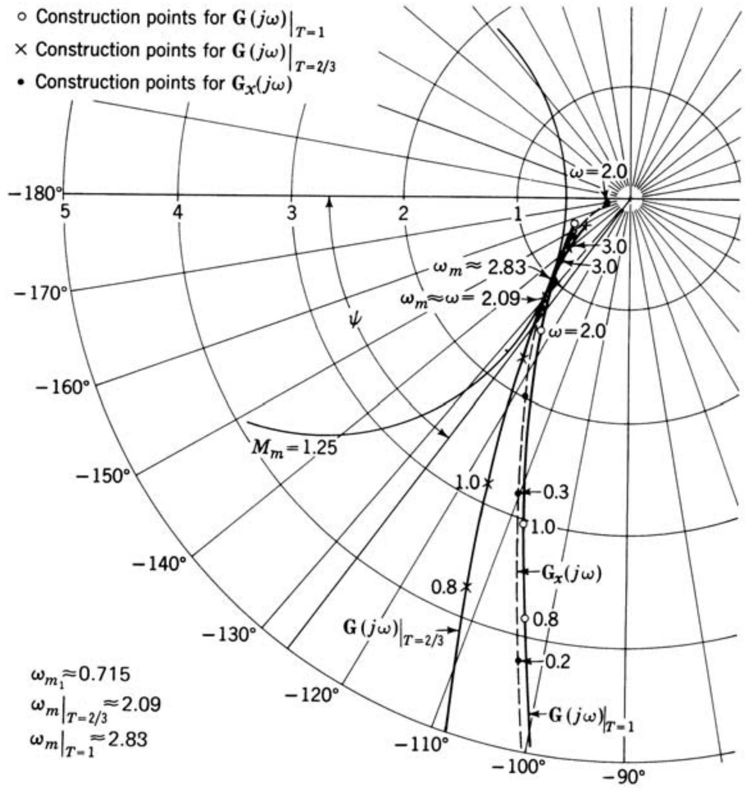


FIGURE 11.15 Polar plots of Type 1 control system with and without lead compensation:

$$\mathbf{G}_x(j\omega) = \frac{0.872}{j\omega(1 + j\omega)(1 + j0.2\omega)}$$

$$\mathbf{G}(j\omega)\big|_{T=2/3} = \frac{3.48(1 + j2\omega/3)}{j\omega(1 + j\omega)(1 + j0.2\omega)(1 + j0.2\omega/3)}$$

$$\mathbf{G}(j\omega)\big|_{T=1} = \frac{3.13}{j\omega(1 + j0.1\omega)(1 + j0.2\omega)}.$$

steady-state accuracy) at the expense of increasing the response time, whereas the lead network gives an appreciable improvement in the time of response and a small improvement in steady-state accuracy. The particular problem at hand dictates the type of compensation to be used. An analytical technique for designing a lead compensator is presented in Ref. 4, using γ as the basis of design.

TABLE 11.2 Comparison Between Root-Locus and Frequency-Response Methods of Applying a Lead Compensator to a Basic Control System

$\mathbf{G}_c'(s)$	ω_{m1}	ω_m	ω_{n1}	ω_n	K_1	K'	M_m	M_p	t_p	s
			Root-locus method, ζ of dominant roots $= 0.45$							
$0.1\dfrac{1+s}{1+0.1s}$	—	2.66	0.89	3.52	0.84	2.95	1.215	1.196	1.09	2.45
$0.1\dfrac{1+0.667s}{1+0.0667s}$	—	2.67	0.89	3.40	0.84	4.40	1.38	1.283	1.05	2.50
			Frequency-response method, $M_m = 1.26$ (effective $\zeta = 0.45$)							
$0.1\dfrac{1+s}{1+0.1s}$	0.715	2.83	—	—	0.872	3.13	1.26	1.215	1.05	2.40
$0.1\dfrac{1+0.667s}{1+0.0667s}$	0.715	2.09	—	—	0.872	3.48	1.26	1.225	1.23	2.73

11.7 CASCADE LAG-LEAD COMPENSATOR

The introduction of a lag compensator results in an increase in the gain, with a consequent reduction of the steady-state error in the system. Introducing a lead compensator results in an increase in the resonant frequency ω_m and a reduction in the system's settling time. If both the steady-state error and the settling time are to be reduced, a lag and a lead compensator must be used simultaneously. This improvement can be accomplished by inserting the individual lag and lead networks in cascade, but it is more effective to use a new network that has both the lag and the lead characteristics. Such a network, shown in Fig. 10.22*a*, is called a lag-lead compensator. Its transfer function is

$$\mathbf{G}_c(j\omega) = \frac{A(1+j\omega T_1)(1+j\omega T_2)}{(1+j\omega\alpha T_1)(1+j\omega T_2/\alpha)} \tag{11.16}$$

where $\alpha > 1$ and $T_1 > T_2$. The first half of this transfer function produces the lag effect, and the second half produces the lead effect. The Lm and phase-angle diagram for a representative lag-lead compensator is shown in Fig. 11.16, where the selection is made $T_1 = 5T_2$. A simple design method is to make T_2 equal to a time constant in the denominator of the original system. An alternative procedure is to locate the compensator curves on the log plots of the original system to produce the highest possible ω_ϕ. Figure 11.17 shows a sketch of the Lm and phase-angle diagram for the original forward transfer function $\mathbf{G}_x(j\omega)$, the compensator $\mathbf{G}_c'(j\omega)$, and the combination $\mathbf{G}_c(j\omega)\mathbf{G}_x'(j\omega)$. The phase-margin frequency is increased from $\omega_{\phi 1}$ to $\omega_{\phi 2}$,

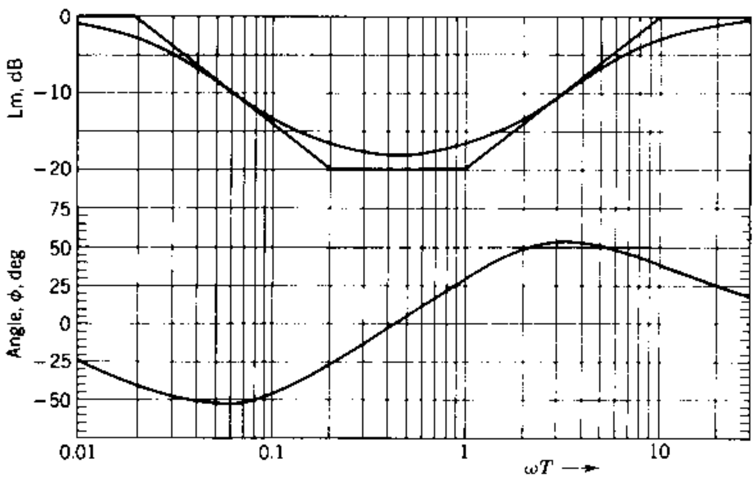


FIGURE 11.16 Log magnitude and phase-angle diagram for lag-lead compensator

$$\mathbf{G}_c'(j\omega) = \frac{(1 + j5\omega T_2)(1 + j\omega T_2)}{(1 + j50\omega T_2)(1 + j0.1\omega T_2)}.$$

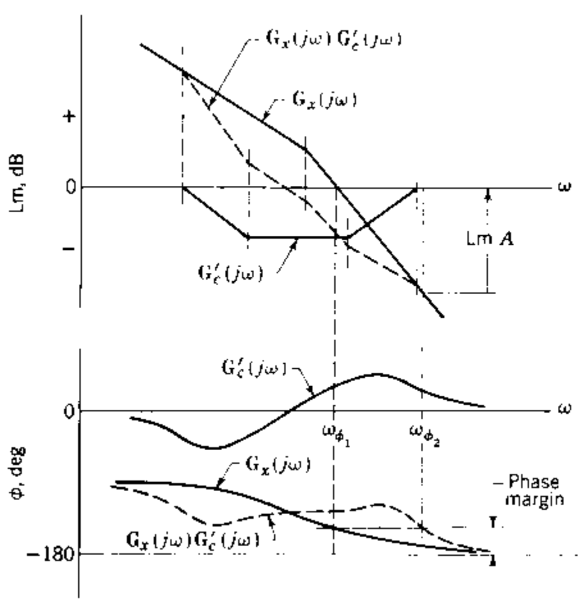


FIGURE 11.17 Original and lag-lead compensated log magnitude and phase angle plots.

and the additional gain is labeled Lm A. The new value of M_m can be obtained from the Lm-angle diagram in the usual manner.

Another method is simply the combination of the design procedures used for the lag and lead networks, respectively. Basically, the Lm and phase-angle curves of the lag compensator can be located as described in Sec. 11.3. The lag-compensator curves are given by Fig. 11.4 and are located so that the negative angle introduced at either the original $\omega_{\phi 1}$ or the original ω_{m1} is small—of the order of -5°. This permits determination of the time constant T_1. The Lm and phase-angle curves of the lead compensator can be located as described in Sec. 11.5. The lead-compensator curves are given by Fig. 11.10, except that the Lm curve should be raised to have a value of 0 dB at the low frequencies. This is necessary because the lag-lead compensator transfer function does not contain α as a factor. Either the time constant T_2 of the compensator can be made equal to the largest time constant of the original system (for a Type 1 or higher system), or the angle curve is located to produce the specified phase margin angle at the highest possible frequency. In the latter case, the location chosen for the angle curve permits determination of the time constant T_2. When using the lag-lead network of Eq. (11.16), the value of α used for the lag compensation must be the reciprocal of the α used for the lead compensation. The example given in the next section implements this second method because it is more flexible.

11.8 DESIGN EXAMPLE: CASCADE LAG-LEAD COMPENSATION

For the basic system treated in Secs. 11.4 and 11.6, the forward transfer function is

$$\mathbf{G}_x(j\omega) = \frac{K_1}{j\omega(1 + j\omega)(1 + j0.2\omega)} \tag{11.17}$$

For an $M_m = 1.26$, the result is $K_1 = 0.872\ \mathrm{s}^{-1}$ and $\omega_{m1} = 0.715$ rad/s. According to the second method described above, the lag-lead compensator is the combination of the individual lag and lead compensators designed in Secs. 11.4 and 11.6. Thus the compensator transfer function is

$$\mathbf{G}_c(j\omega) = A\frac{(1 + jT_1\omega)(1 + j\omega T_2)}{(1 + j\alpha T_1\omega)(1 + j0.1\omega T_2)} \quad \text{for } \alpha = 10 \tag{11.18}$$

The new forward transfer function with $T_1 = 14$ and $T_2 = 1$ is

$$\mathbf{G}(j\omega) = \mathbf{G}_c(j\omega)\mathbf{G}_x(j\omega) = \frac{AK_1(1 + j14\omega)}{j\omega(1 + j0.2\omega)(1 + j140\omega)(1 + j0.1\omega)} \tag{11.19}$$

Figures 11.18*a* and 11.18*b* show the Lm and phase-angle diagrams of the basic system $\mathbf{G}'_x(j\omega)$, the compensator $\mathbf{G}'_c(j\omega)$, and the composite transfer function $\mathbf{G}'_c(j\omega)\mathbf{G}'_x(j\omega)$. From these curves, the Lm-angle diagram $\mathbf{G}'_a(j\omega)$ is shown in Fig. 11.19. Upon adjusting for an $M_m = 1.26$, the results are $K'_1 = 29.93$ and $\omega_{ma} = 2.69$. When these values are compared with those of Tables 11.1 and 11.2, it is seen that lag-lead compensation results in a value of K_1 approximately equal to the product of the K_1's of the lag and lead-compensated systems. Also, the ω_m is only slightly less than that obtained with the lead-compensated system. Thus, a larger increase in gain is achieved than with lag compensation, and the value of ω_m is almost as large as the value obtained by using the lead-compensated system. One may therefore be inclined to use a lag-lead compensator exclusively. $\mathbf{G}_b$ in Fig. 11.19 is with $T_1 = 20$.

The resulting error coefficient (gain) due to the insertion of a lag-lead compensator is approximately equal to the product of

1. The original error coefficient K_m of the uncompensated system
2. The increase in K_m due to the insertion of a lag compensator
3. The increase in K_m due to the insertion of a lead compensator

Table 11.3 shows the results for both the root-locus and log-plot methods of applying a lag-lead compensator to a basic control system. The compensator designed from the log plots is different than the one designed by the root-locus method. To show that the difference in performance is small, Table 11.3 includes the results obtained with both compensators when applying the frequency-response method. Both plots are shown in Fig. 11.19. The plot using Eq. (10.42) with $T_1 = 20$ yields $\omega_{mb} = 2.74$, as shown in this figure.

11.9 FEEDBACK COMPENSATION DESIGN USING LOG PLOTS [1]

The previous sections demonstrate straightforward procedures for applying cascade compensation. When feedback compensation is applied to improve the tracking qualities of a control system, additional parameters must be determined. Besides the gain constant of the basic forward transfer function, the gain constant and frequency characteristics of the minor feedback must also be selected. Additional steps must therefore be introduced into the design procedure. The general effects of feedback compensation are demonstrated through application to the specific system that is used throughout this text. The use of a minor loop for feedback compensation is shown in Fig. 11.20. The transfer function $\mathbf{G}_x(j\omega)$ represents the basic system, $\mathbf{H}(j\omega)$ is the feedback compensator forming a minor loop around $\mathbf{G}_x(j\omega)$, and A is an

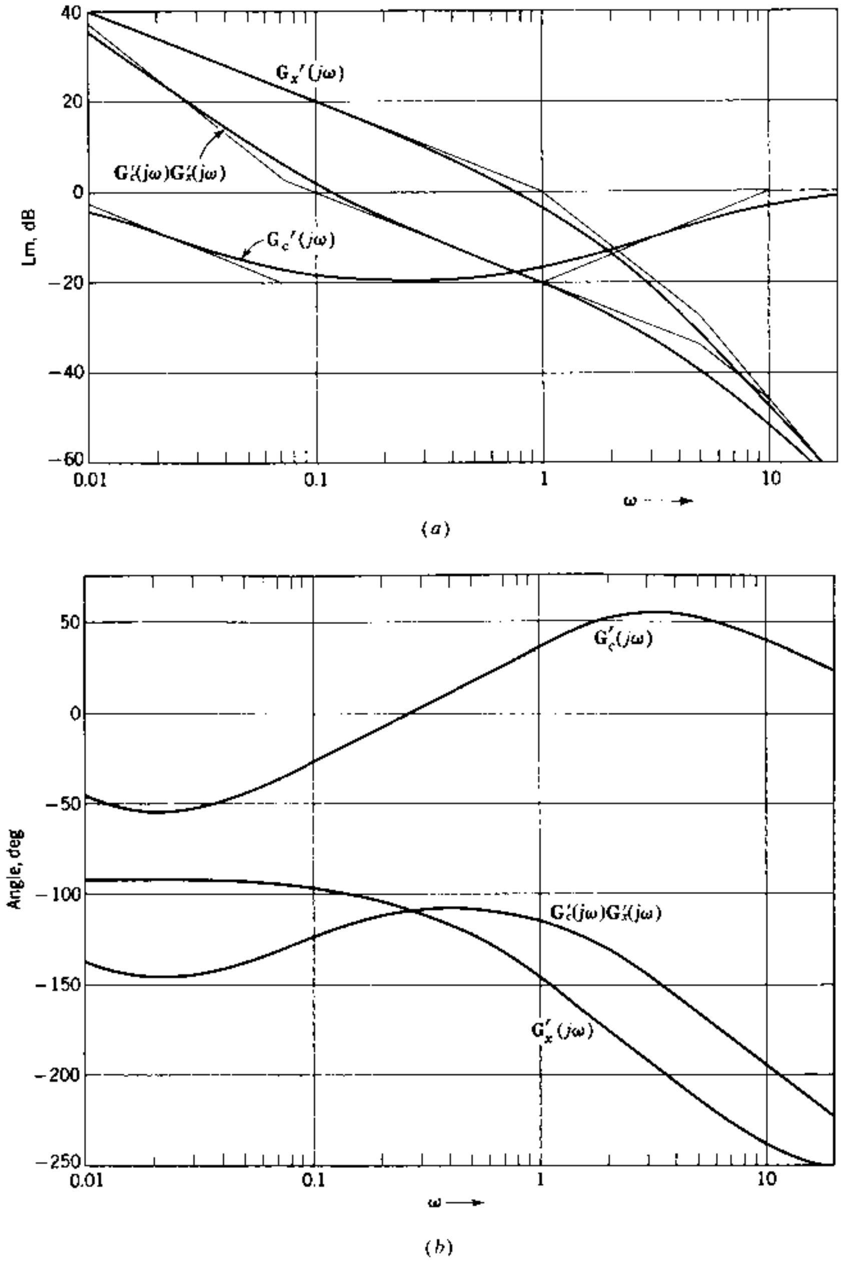


FIGURE 11.18 Log magnitude and phase-angle diagrams of Eqs. (11.17) to (11.19) with $T_1 = 14$.

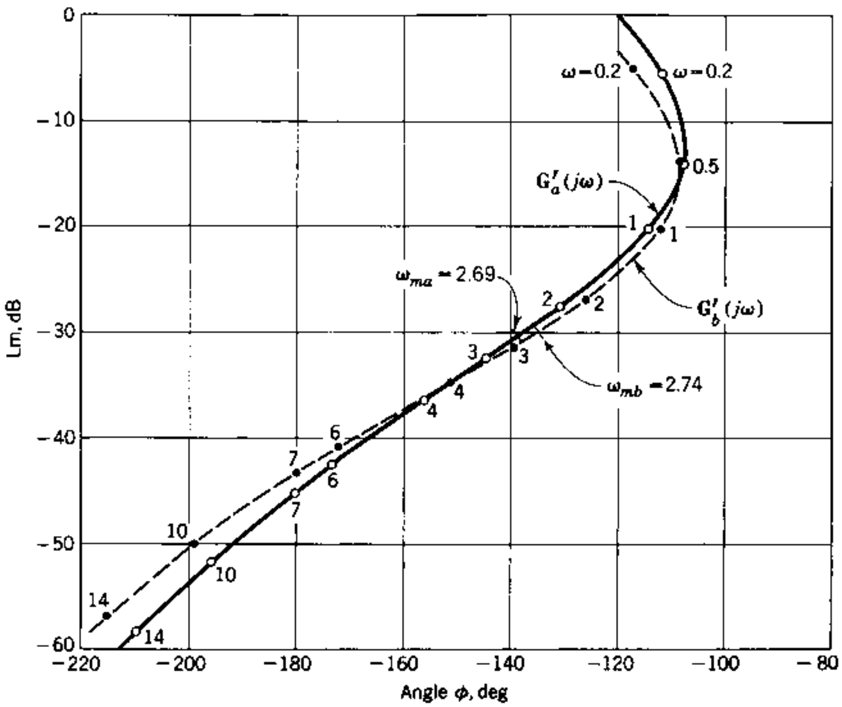


FIGURE 11.19 Log magnitude–angle diagram of $\mathbf{G}'(j\omega)$ from Eq. (11.19)—$\mathbf{G}_a'$ with $T_1 = 14$ and $\mathbf{G}_b^1$ with $T_1 = 20$.

TABLE 11.3 Comparison Between Root-Locus and Frequency-Response Methods of Applying a Lag-Lead Compensator to a Basic Control System

$G_c'(s)$	ω_{m1}	ω_m	ω_{n1}	ω_n	K_1	K_1'	M_m	M_p	t_p, s	t_s, s
Root locus method, ζ of dominant roots = 0.45										
$\frac{(1+s)(1+20s)}{(1+0.1s)(1+200s)}$	—	2.63	0.898	3.49	0.84	29.26	1.23	1.213	1.10	3.33
Frequency response method, $M_m = 1.26$ (effective $\zeta = 0.45$)										
$\frac{(1+s)(1+20s)}{(1+0.1s)(1+200s)}$	0.715	2.74	—	—	0.872	30.36	1.26	1.225	1.08	3.30
$\frac{(1+s)(1+14s)}{(1+0.1s)(1+140s)}$	0.715	2.69	—	—	0.872	29.93	1.26	1.229	1.09	3.55

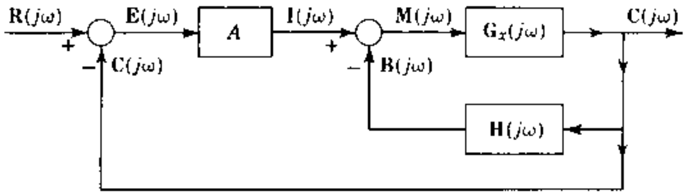


FIGURE 11.20 Block diagram for feedback compensation.

amplifier that is used to adjust the overall performance. The forward transfer function of this system is

$$\mathbf{G}_1(j\omega) = \frac{\mathbf{C}(j\omega)}{\mathbf{I}(j\omega)} = \frac{\mathbf{G}_x(j\omega)}{1 + \mathbf{G}_x(j\omega)\mathbf{H}(j\omega)} \tag{11.20}$$

In order to apply feedback compensation, new techniques must be developed. This is done by first using some approximations with the straight-line Lm curves and then developing an exact procedure. Consider the two cases when

$$\left|\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)\right| \ll 1 \qquad \text{and} \qquad \left|\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)\right| \gg 1$$

The forward transfer function of Eq. (11.20) can be approximated by

$$\mathbf{G}_1(j\omega) \approx \mathbf{G}_x(j\omega) \qquad \text{for } \left|\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)\right| \ll 1 \tag{11.21}$$

and

$$\mathbf{G}_1(j\omega) \approx \frac{1}{\mathbf{H}(j\omega)} \qquad \text{for } \left|\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)\right| \gg 1 \tag{11.22}$$

The condition when $|\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)| \approx 1$ is still undefined, in which case neither Eq. (11.21) nor Eq. (11.22) is applicable. In the approximate procedure this condition is neglected, and Eqs. (11.21) and (11.22) are used when $|\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)| < 1$ and $|\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)| > 1$, respectively. This approximation allows investigation of the qualitative results to be obtained. After these results are found to be satisfactory, the refinements for an exact solution are introduced.

An example illustrates the use of these approximations. Assume that $\mathbf{G}_x(j\omega)$ represents a motor having inertia and damping. The transfer function derived in Sec. 2.13 can be represented by

$$\mathbf{G}_x(j\omega) = \frac{K_M}{j\omega(1 + j\omega T_m)} \tag{11.23}$$

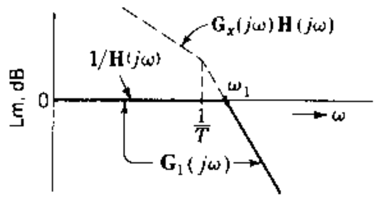


FIGURE 11.21 Log magnitude curves for $\mathbf{G}_x(j\omega)\mathbf{H}(j\omega) = K_1/j\omega(1 + j\omega T)$ and the approximate forward transfer function $\mathbf{G}_1(j\omega)$ for $\mathbf{H}(j\omega) = 1$.

Let the feedback $\mathbf{H}(j\omega) = 1\angle 0^\circ$. This problem is sufficiently simple to be solved exactly algebraically. However, use is made of the Lm curve and the approximate conditions. In Fig. 11.21 the Lm curve for $\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)$ is sketched.

From Fig. 11.21 it is seen that $|\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)| > 1$ for all frequencies below ω_1. Using the approximation of Eq. (11.22), $\mathbf{G}_1(j\omega)$ can be represented by $1/\mathbf{H}(j\omega)$ for frequencies below ω_1. Also, because $|\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)| < 1$ for all frequencies above ω_1, the approximate of Eq. (11.21) can be used. Therefore $\mathbf{G}_1(j\omega)$ can be represented, as shown, by the line of zero slope and 0 dB for frequencies up to ω_1 and the line of slope −12-dB/octave above ω_1. The approximate equation of $\mathbf{G}_1(j\omega)$ therefore has a quadratic in the denominator with $\omega_n = \omega_1$:

$$\mathbf{G}_1(j\omega) = \frac{1}{1 + 2\zeta j\omega/\omega_1 + (j\omega/\omega_1)^2} \tag{11.24}$$

Of course $\mathbf{G}(j\omega)$ can be obtained algebraically for this simple case from Eq. (11.20), with the result given as

$$\mathbf{G}_1(j\omega) = \frac{1}{1 + (1/K_M)j\omega + (j\omega)^2 T/K_M} \tag{11.25}$$

where $\omega_1 = \omega_n = (K_M/T)^{1/2}$ and $\zeta = 1/[2(K_M T)^{1/2}]$. Note that the approximate result is basically correct, but that some detailed information, in this case the value of ζ, is missing. The approximate angle curve can be drawn to correspond to the approximate Lm curve of $\mathbf{G}_1(j\omega)$ or to Eq. (11.24).

11.10 DESIGN EXAMPLE: FEEDBACK COMPENSATION (LOG PLOTS)

The system of Fig. 11.20 is investigated with the value of $\mathbf{G}_x(j\omega)$ given by

$$\mathbf{G}_x(j\omega) = \frac{K_x}{j\omega(1 + j\omega)(1 + j0.2\omega)} \tag{11.26}$$

The system having this transfer function has been used throughout this text for the various methods of compensation. The Lm and phase-angle diagram using the straight-line Lm curve is drawn in Fig. 11.22. A phase margin angle, γ of 45° can be obtained, which yields $\omega_{\phi 1} = 0.8$, provided the gain is changed by -2 dB.

The object in applying compensation is to increase both the ω_ϕ and the gain. In order to select a feedback compensator $\mathbf{H}(j\omega)$, consider the following facts:

1. The system type should be maintained. In accordance with the conditions outlined in Sec. 10.15 and the results given in Sec. 10.20 for the root-locus method, it is known that good improvement is achieved by using a feedback compensator $\mathbf{H}(j\omega)$ that has a zero, $s = 0$, of order equal to (or preferably higher than) the type of the original transfer function $\mathbf{G}_x(j\omega)$.
2. It is shown in Sec. 11.9 that the new forward transfer function can be approximated by

$$\mathbf{G}_1(j\omega) \approx \begin{cases} \mathbf{G}_x(j\omega) & \text{for } \left|\mathbf{G}_x(j\omega)\mathbf{H}(j\omega) < 1\right| \\ \dfrac{1}{\mathbf{H}(j\omega)} & \text{for } \left|\mathbf{G}_x(j\omega)\mathbf{H}(j\omega) > 1\right| \end{cases}$$

This information is used for the proper selection of $\mathbf{H}(j\omega)$.

In order to increase the values of ω_ϕ and the gain, a portion of the curves of $\mathbf{G}_x(j\omega)$ for a selected range of frequencies can be replaced by the curves

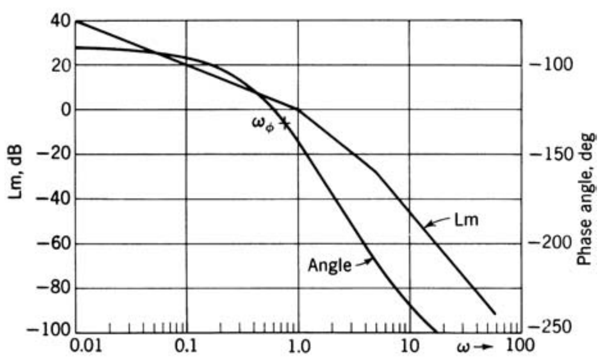


FIGURE 11.22 Log magnitude and phase-angle diagram of

$$\mathbf{G}'_x(j\omega) = \frac{1}{j\omega(1 + j\omega)(1 + j0.2\omega)}.$$

of $1/\mathbf{H}(j\omega)$. This requires that the value of $\mathbf{H}(j\omega)$ be selected such that $|\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)| > 1$ for that range of frequencies. A feedback unit using a tachometer and an RC derivative network is considered. The circuit is shown in Fig. 10.36, and the transfer function is

$$\mathbf{H}(j\omega T) = \frac{(K_t/T)(j\omega T)^2}{1 + j\omega T} = K_h \mathbf{H}'(j\omega T) \tag{11.27}$$

where $K_h = K_t/T$. The Lm and phase-angle diagram for $1/\mathbf{H}'(j\omega)$ is plotted in Fig. 11.23 with a nondimensionalized frequency scale, ωT. When this scale is converted to a dimensionalized scale ω, the curves of Fig. 11.23 shift to the left or to the right as T is increased or decreased. The angle curve $1/\mathbf{H}(j\omega)$ shows that the desired γ can be obtained at a frequency that is a function of the compensator time constant T. This shows promise for increasing γ, provided that the magnitude $|\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)|$ can be made greater than unity over the correct range of frequencies. Also, the larger K_t/T is made, the smaller the magnitude of $1/\mathbf{H}(j\omega)$ at the phase-margin frequency. This permits a large value of gain A and therefore a large error coefficient.

Section 8.18 describes the desirability of a slope of −20 dB/decade to produce a large γ. The design objective is therefore to produce a section of the $\mathbf{G}_1(j\omega)$ Lm curve with a slope of −20 dB/decade and with a smaller magnitude than (i.e., below) the Lm curve of $\mathbf{G}_x(j\omega)$. This new section of the $\mathbf{G}_1(j\omega)$ curve must occur at a higher frequency than the original $\omega_{\phi 1}$. To achieve compensation, the Lm curve $1/\mathbf{H}(j\omega)$ is placed over the $\mathbf{G}_x(j\omega)$ curve so that there are one to two points of intersection. There is no reason to restrict the value of K_x, but a practical procedure is to use the value of K_x that

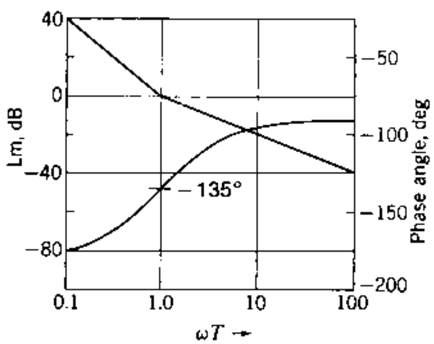


FIGURE 11.23 Log magnitude and phase-angle diagram of $\dfrac{1}{\mathbf{H}'(j\omega T)} = \dfrac{1 + j\omega T}{(j\omega T)^2}$.

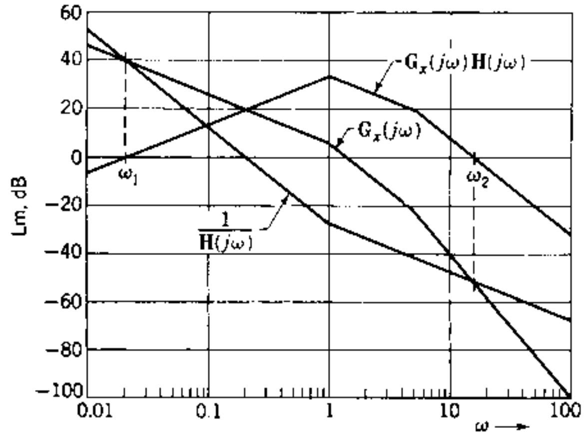


FIGURE 11.24 Log magnitude plots of $1/\mathbf{H}(j\omega)$, $\mathbf{G}_x(j\omega)$, and $\mathbf{G}_x(j\omega)\mathbf{G}(j\omega)$.

produces the desired M_m without compensation. One such arrangement is shown in Fig. 11.24, with $K_x = 2$, $T = 1$, and $K_t = 25$. Intersections between the two curves occur at $\omega_1 = 0.021$ and $\omega_2 = 16$. The Lm plot of $\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)$ is also drawn in Fig. 11.24 and shows that

$$\left|\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)\right| \begin{cases} < 1 & \text{for } \omega_2 < \omega < \omega_1 \\ > 1 & \text{for } \omega_2 > \omega > \omega_1 \end{cases}$$

Therefore $\mathbf{G}_1(j\omega)$ can be represented approximately by

$$\mathbf{G}_1(j\omega) \approx \begin{cases} \mathbf{G}_x(j\omega) & \text{for } \omega_2 < \omega < \omega_1 \\ 1/\mathbf{H}(j\omega) & \text{for } \omega_2 > \omega > \omega_1 \end{cases}$$

The composite curve has corner frequencies at $1/\omega_1$, $1/T$, and $1/\omega_2$ and can therefore be approximately represented, when T=1, by

$$\mathbf{G}_1(j\omega) = \frac{2(1 + j\omega/T)}{j\omega(1 + j\omega/\omega_1)(1 + j\omega/\omega_2)^2} = \frac{2(1 + j\omega)}{j\omega(1 + j47.5\omega)(1 + j0.0625\omega)^2} \tag{11.28}$$

The Lm and phase-angle diagram for Eq. (11.28) is shown in Fig. 11.25. Selecting $T = 1$, then $\gamma = 45^\circ$ can occur at $\omega = 1$. However, the new approximate forward transfer function $\mathbf{G}_1(j\omega)$ has a second-order pole at $\omega = \omega_2$, as shown in Eq. (11.28). Therefore, there is a higher frequency at which $\gamma = 45^\circ$ can be achieved, as shown in Fig. 11.25. The higher frequency $\omega_\phi = 4.5$ is chosen in this example, which yields a gain of Lm $A = 41$ dB. This shows a considerable improvement of the system performance.

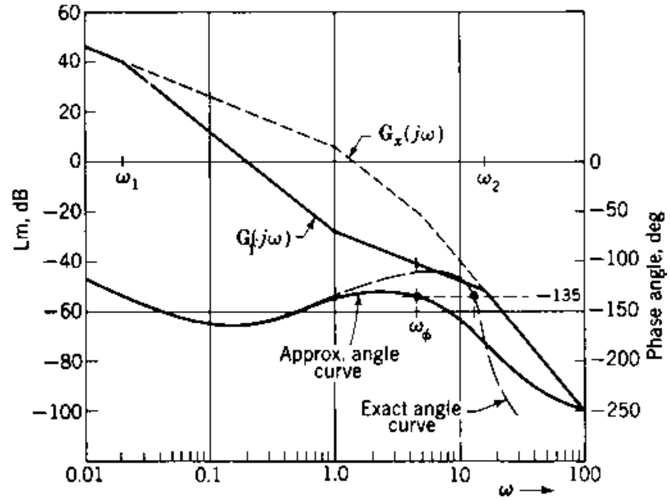


FIGURE 11.25 Log magnitude and phase-angle diagrams of

$$\mathbf{G}_1(j\omega) = \frac{2(1+j\omega)}{j\omega(1+j47.5\omega)(1+j0.0625\omega)^2}.$$

The angle curve of $1/\mathbf{H}'(j\omega T)$ in Fig. 11.23 shows that the desired values of γ and ω_ϕ are a function of the value of T. For a given value of γ the angle $\angle 1/\mathbf{H}'(j\omega T) = -180^\circ + \gamma$ is evaluated. The corresponding value of $(\omega T)_\gamma$ is then $\tan\gamma = \omega T$. Then the required value of T is $T = (\omega T)_\gamma/\omega_\phi$. It is desired that the frequency ω_2 at the upper intersection of $\mathbf{G}_x(j\omega)$ and $1/\mathbf{H}(j\omega)$ (see Figs. 11.24 and 11.25) be much larger than the new phase margin frequency ω_ϕ. This is achieved with $\omega_2 \geq 20\omega_\phi$, which satisfies the desired condition that $\angle \mathbf{G}_1(j\omega_\phi) = \angle 1/\mathbf{H}(j\omega_\phi)$. When this condition is satisfied, a precise determination of T is possible.

An exact curve of $\mathbf{G}_1(j\omega)$ can be obtained analytically. The functions $\mathbf{G}_x(j\omega)$ and $\mathbf{H}(j\omega)$ are

$$\mathbf{G}_x(s) = \frac{2}{s(1+s)(1+0.2s)} \qquad \mathbf{H}(s) = \frac{25s^2}{1+s} \tag{11.29}$$

The function $\mathbf{G}_1(j\omega)$ is obtained as follows:

$$\begin{aligned}\mathbf{G}_1(j\omega) &= \frac{\mathbf{G}_x(j\omega)}{1+\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)} = \frac{2(1+j\omega)}{j\omega(1+j\omega)^2(1+j0.2\omega)+50(j\omega)^2} \\ &= \frac{2(1+j\omega)}{j\omega(1+j52\omega)[1+0.027(j\omega)+0.00385(j\omega)^2]}\end{aligned} \tag{11.30}$$

A comparison of Eqs. (11.28) and (11.30) shows that there is little difference between the approximate and the exact equations. The main difference is that the approximate equation has the term $(1 + j0.0625\omega)^2$, which assumes a damping ratio $\zeta = 1$ for this quadratic factor. The actual damping ratio is $\zeta = 0.217$ in the correct quadratic factor $[1 + 0.027(j\omega) + 0.00385(j\omega)^2]$. The corner frequency is the same for both cases, $\omega_n = 16$. As a result, the exact angle curve given in Fig. 11.25 shows a higher value of $\omega_\phi(= 13)$ than that obtained for the approximate curve ($\omega_\phi = 4.5$). If the basis of design is the value of M_m, the exact curve should be used. The data for the exact curve are readily available from a computer (see Appendix C).

The adjustment of the amplifier gain A is based on either γ or M_m. An approximate value for A is the gain necessary to raise the curve of Lm $\mathbf{G}_1(j\omega)$ so that it crosses the 0-dB line at ω_ϕ. A more precise value for A is obtained by plotting $\mathbf{G}_1(j\omega)$ on the Lm–angle diagram and adjusting the system gain for a desired value of M_m. The exact Lm $\mathbf{G}_1(j\omega)$ curve should be used for best results.

The shape of the Nichols plot for $\mathbf{G}_1(j\omega)$ is shown in Fig. 11.26. When adjusting the gain for a desired M_m, the objectives are to achieve a large resonant frequency ω_m and a large gain constant K_1. The tangency of the M_m curve at $\omega_c = 15.3$ requires a gain $A = 95.5$ and yields $K_1 = 191$. However,

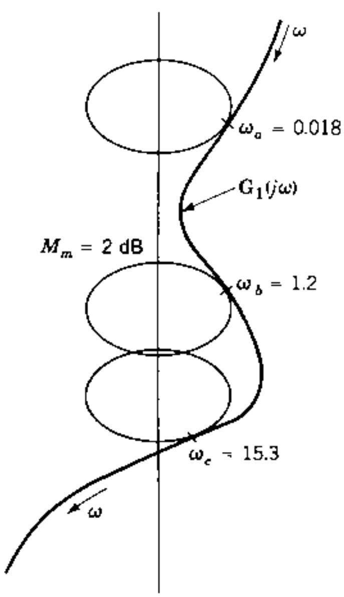


FIGURE 11.26 Nichols plot of $\mathbf{G}_1(j\omega)$.

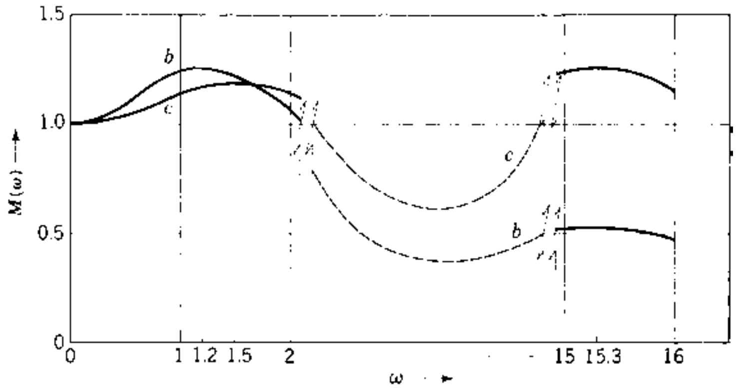


FIGURE 11.27 Closed-loop frequency response

the rapid change in phase angle in the vicinity of ω_c can produce unexpected results. The M vs. ω characteristic curve c, shown in Fig. 11.27, has two peaks. This is not the form desired and must be rejected. This example serves to call attention to the fact that one point, the peak M value, is not sufficient to determine the suitability of the system response.

Tangency at ω_b, in Fig. 11.26 requires $A = 57.5$ and yields $K_1 = 115$. Curve b in Fig. 11.27 has the form desired and is acceptable. Compared with the original system, the minor-loop feedback compensation has produced a modest increase in ω_m from 0.715 to 1.2. There is also a very large increase in K_1 from 0.872 to 115. Tangency at ω_a is not considered, as it would represent a degradation of the uncompensated performance.

This example qualitatively demonstrates that feedback compensation can be used to produce a section of the Lm curve of the forward transfer function $\mathbf{G}(j\omega)$ with a slope of -20 dB/decade. This section of the curve can be placed in the desired frequency range, with the result that the new ω_ϕ is larger than that of the original system. However, to obtain the correct quantitative results, the exact curves should be used instead of the straight-line approximations. Also, the use of the Lm–angle diagram as the last step in the design process, to adjust the gain for a specified M_m, gives a better indication of performance than the γ does.

The reader may wish to try other values of K_t and T in the feedback compensator in order to modify the system performance. Also, it is worth considering whether or not the RC network should be used as part of the feedback compensator. The results using the root-locus method (see Secs. 10.16 to 10.20) show that rate feedback alone does improve system performance. The use of second-derivative feedback produces only limited improvements.

TABLE 11.4 Basic Feedback Compensators

	$H(s)$	$1/H(s)$	Characteristic of $\angle 1/H(j\omega)$ for $0 \le \omega \le \infty$
1.	$K_t s$	$\dfrac{1}{K_t s}$	-90°
2.	$\dfrac{K_t s}{Ts+1}$	$\dfrac{Ts+1}{K_t s}$	-90° to 0°
3.	$\dfrac{K_t Ts^2}{Ts+1}$	$\dfrac{Ts+1}{K_t Ts^2}$	-180° to -90°

11.11 APPLICATION GUIDELINES: BASIC MINOR-LOOP FEEDBACK COMPENSATORS

Table 11.4 presents three basic feedback compensators and the corresponding angle characteristics of $1/\mathbf{H}(j\omega)$. One approach in designing these compensators is to determine the parameters of $H(s)$ that yield the desired value of the phase margin angle γ at a desired value of ω_ϕ.

For compensators 1 and 2 in Table 11.4, the desired values of $\gamma(<90^\circ)$ and ω_ϕ should occur in the vicinity of the intersection of the Lm $[1/\mathbf{H}(j\omega)]$ and the Lm $\mathbf{G}_x(j\omega)$ plots, assuming that the angle of $\mathbf{G}_x(j\omega)$ varies from its initial value to $-90^\circ i$ ($i = 2, 3, \ldots$) as the frequency increases from zero to infinity. Thus, to achieve an acceptable design of these compensators, an initial choice is made for the value of K_x or K_t and T. It may be necessary to lower the plot of Lm $[1/\mathbf{H}(j\omega)]$ to ensure that the intersection value of ω is in the proper range, such that the desired value of ω_ϕ is attainable. The intersection of Lm $[1/\mathbf{H}(j\omega)]$ with Lm $\mathbf{G}_x(j\omega)$ must occur on that part of Lm $G(j\omega)$ that has a slope that will allow a $\gamma < 90^\circ$ to be achievable in the vicinity of the intersection. By trial and error, the parameters of $1/\mathbf{H}(s)$ are varied to achieve an acceptable value of ω_ϕ. For practical control systems there is an upper limit to an acceptable value of ω_ϕ, based upon the desire to attenuate high-frequency system noise.

For compensator 3, based upon Table 11.4 and Sec. 11.10, there are usually two intersections of Lm $\mathbf{G}_x(j\omega)$ and Lm $(1/\mathbf{H}(j\omega))$. For the system of Fig. 11.20, the desired values of $\gamma(<90$ deg) and ω_ϕ for Lm $A\mathbf{G}_1$ can be made to occur on the -6 dB/octave portion of the plot of Lm $[1/\mathbf{H}(j\omega)]$. This requires that the intersections of the plots of Lm $[\mathbf{G}_x(j\omega)]$ and Lm $[1/\mathbf{H}(j\omega)]$ occur at least one decade away from the value of ω that yields the desired values of ω_ϕ and γ. With this separation, the forward transfer function $\mathbf{G}_1(j\omega_\phi) \approx 1/\mathbf{H}(j\omega_\phi)$.

The guidelines presented in this section for the three compensators of Table 11.4 are intended to provide some fundamental insight to the reader in developing effective minor loop feedback compensator design procedures.

11.12 SUMMARY

A basic unity-feedback control system is used throughout this and other chapters. It is therefore possible to compare the results obtained by adding each of the compensators to the system. The results obtained are consolidated in Table 11.5. The results achieved by using the frequency response are comparable to those obtained by using the root locus (see Chap. 10). The design methods that have been demonstrated use some trial-and error techniques. This design process can be expedited by use of an appropriate CAD package (see Appendixes C and D). The greater complexity of the feedback design procedures make reliance on CAD packages more valuable.

The following properties can be attributed to each compensator:

Cascade lag compensator. Results in an increase in the gain K_m and a small reduction in ω_m. This reduces steady-state error but increases transient settling time.

TABLE 11.5 Summary of Cascade and Feedback Compensation of a Basic Control System Using the Frequency-Response Method with $M_m = 1.26$ (Effective $\zeta = 0.45$)

Compensator	ω_m	K_1	Additional gain required	M_p	t_p, s	t_s, s
Uncompensated	0.715	0.872	—	1.203	4.11	9.9
Lag:						
$\dfrac{1+14s}{1+140s}$	0.53	6.68	7.66	1.257	4.90	22.8
$\dfrac{1+20s}{1+200s}$	0.60	7.38	8.46	1.246	4.55	23.6
Lead:						
$0.1\dfrac{1+s}{1+0.1s}$	2.83	3.13	35.9	1.215	1.05	2.40
$0.1\dfrac{1+2s/3}{1+0.2s/3}$	2.09	3.48	39.9	1.225	1.23	2.73
Lag-lead:						
$\dfrac{(1+s)(1+14s)}{(1+0.1s)(1+140s)}$	2.69	29.93	34.3	1.229	1.088	3.55
$\dfrac{(1+s)(1+20s)}{(1+0.1s)(1+200s)}$	2.74	30.36	35.0	1.225	1.077	3.30
Feedback:						
$H(s) = \dfrac{25s^2}{1+s}$	1.20	115	57.5	1.203	1.49	3.27

Cascade lead compensator. Results in an increase in ω_m, thus reducing the settling time. There may be a small increase in gain.

Cascade lag-lead compensator. Results in both an increase in gain K_m and in resonant frequency ω_m. This combines the improvements of the lag and the lead compensators.

Feedback compensator. Results in an increase in both the gain K_m and the resonant frequency ω_m. The basic improvement in ω_m is achieved by derivative feedback. This may be modified by adding a filter such as an *RC* network.

REFERENCES

1. Chestnut, H., and R.W. Mayer: *Servomechanisms and Regulating System Design*, 2nd ed., vol. 1, Wiley, New York, 1959, chaps. 10 and 12.
2. Kuo, F. F.: *Network Analysis and Synthesis*, 2nd ed., Wiley, New York, 1966.
3. Mitra, S. K.: *Analysis and Synthesis of Linear Active Networks*, Wiley, New York, 1969.
4. Philips, C. L.: "Analytical Bode Design Of Controllers," *IEEE Trans. Educ.*, vol. E-28, no. 1, pp. 43–44, 1985.

12

Control-Ratio Modeling

12.1 INTRODUCTION [1]

The conventional control-design techniques of the previous chapters determine closed-loop system performance based on the open-loop transfer function [$G(s)$ or $G(s)H(s)$] or the open-loop transfer function of an equivalent unity-feedback system. The frequency response or the root-locus technique yields a closed-loop response that is based on some of the closed-loop specifications. If necessary, compensation is then added in order to meet additional specifications. There are some techniques that first require the modeling of a desired control ratio that satisfies the desired figures of merit. The desired control ratio is then used to determine the necessary compensation. Formulating the desired control ratio is the main objective of this chapter. Figure 12.1 represents a multiple-input single-output (MISO) control system in which $r(t)$ is the input, $y(t)$ is the output required to track the input, $d(t)$ is an external disturbance input that must have minimal effect on the output, and $F(s)$ is a prefilter to assist in the tracking of $r(t)$ by $y(t)$. Thus, two types of control ratios need to be modeled: a tracking transfer function $M_T(s)$ and a disturbance transfer function $M_D(s)$. In Fig. 12.1 and the remainder of this text, except for Chaps 15 and 16, the output is denoted by $y(t)$ or $Y(s)$ instead of $c(t)$ or $C(s)$; this conforms with symbols typically found in the literature for this material. For single-input single output (SISO) systems, the

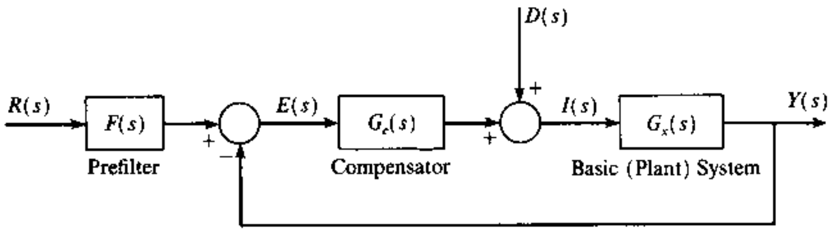


FIGURE 12.1 A MISO control system.

symbol $c(t)$ is most often used to represent the output in most basic texts and agrees with the presentation in the earlier parts of this book.

The control systems considered in the previous chapters represent systems for which $F(s) = 1$ and $d(t) = 0$; thus, they are SISO control systems. For these SISO systems the Guillemin-Truxal method, which requires the designer to specify a desired model $M_T(s)$, represents a method of designing the required compensator $G_c(s)$ of Fig. 12.1.

MISO control systems, in which both a desired input $r(t)$ and a disturbance input $d(t)$ are present, require the design of a cascade compensator $G_c(s)$ and an input prefilter $F(s)$, as shown in Fig. 12.1. The *quantitative feedback theory* (QFT) technique addresses the design of such systems, which have both a desired and a disturbance input. This design technique requires both types of control-ratio models, $M_T(s)$, and $M_D(s)$, in order to effect a satisfactory design. Some state-variable design techniques (Chapter 13) also require the specification of a model tracking control ratio. The use of proportional plus integral plus derivative (PID) controllers leads naturally to the rejection of disturbance inputs and the tracking of the command inputs.

This chapter is devoted to the discussion of tracking control-ratio models, the presentation of the Guillemin-Truxal design method, the presentation of a disturbance-rejection design method for a control system having $r(t) = 0$ and a disturbance input $d(t) \neq 0$, and the discussion of disturbance-rejection control-ratio models. CAD packages (see appendixes C and D) are available to assist the designer in synthesizing desired model control ratios (see Sec. 10.6).

12.2 MODELING A DESIRED TRACKING CONTROL RATIO

The pole-zero placement method of this section requires the modeling of a desired tracking control ratio $[Y(s)/R(s)]_T = M_T(s)$ in order to satisfy the desired tracking performance specifications for M_p, t_p, and T_s (or t_s) for a step input (transient characteristic) and the gain K_m (steady-state characteristic). The model must be consistent with the degrees of the numerator and denominator, w and n respectively, of the basic plant $G_x(s)$. Before proceeding, it is suggested that the reader review Secs. 3.9, 3.10, 9.11, 10.2, and 10.3.

CASE 1. The model that yields the desired figures of merit and is simplest to synthesize is one that represents an effective simple second-order control ratio $[Y(s)/R(s)]_T$, and is independent of the numerator and denominator degrees w and n of the plant. The simple second-order closed-loop model for this case is

$$\begin{aligned} M_{T1}(s) &= \frac{A_1}{s^2 + as + A_1} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \\ &= \frac{\omega_n^2}{(s - p_1)(s - p_2)} = \frac{G_{\text{eq}}(s)}{1 + G_{\text{eq}}(s)} \end{aligned} \tag{12.1}$$

where $A_1 = \omega_n^2$ ensures that, with a step input of amplitude R_0, the output tracks the input so that $y(t)_{ss} = R_0$. The control ratio of Eq. (12.1) corresponds to a unity-feedback system with an equivalent forward transfer function

$$G_{\text{eq}}(s) = \frac{A_1}{s(s + a)} \tag{12.2}$$

Note that although there are four possible specifications, M_p, t_p, T_s, and K_m, only two adjustable parameters, ζ and ω_n, are present in Eq. (12.1). When Eqs. (3.60), (3.61), and (3.64) are analyzed for peak time t_p, peak overshoot M_p, and setting time T_s, respectively, and the appropriate value for K_m is selected, it is evident that only two of these relationships can be used to obtain the two adjustable parameters that locate the two dominant poles $p_{1,2}$ of Eq. (12.1). Therefore, it may not be possible to satisfy all four specifications. For example, specifying the minimum value of the gain K_1 (where $K_1 = A_1/a$) yields the minimum acceptable value of ω_n. Also, the required value of ζ is determined from $M_p = 1 + \exp(-\pi\zeta/\sqrt{1 - \zeta^2})$. Having determined ω_n and ζ, the values of $t_p = \pi/\omega_d$ and $T_s = 4/\zeta\omega_n$ can be evaluated. If these values of t_p and T_s are within the acceptable range of values, the desired model has been achieved. If not, an alternative procedure is to determine ζ and ω_n by using the desired M_p and T_s, or M_p and $t_p = \pi/\omega_n\sqrt{1 - \zeta^2}$.

In order to satisfy the requirement that the specifications $1 < M_p \le \alpha_p$, $t_p \le t_{\alpha_p}$, and $t_s \le t_{\alpha_s}$ two of these values α_p, t_{α_p}, and t_{α_s} are used to determine the two unknowns ζ_α and ω_{n_α}. These values of ζ_α and ω_{n_α} yield the desired dominant poles $p_{1,2}$ shown in Fig. 12.2. Any dominant complex pole not lying in the shaded area satisfies these specifications. If an acceptable second-order $M_T(s)$, as given by Eq. (12.1), cannot be found that satisfies all the desired performance specifications, a higher-order control ratio must be used.

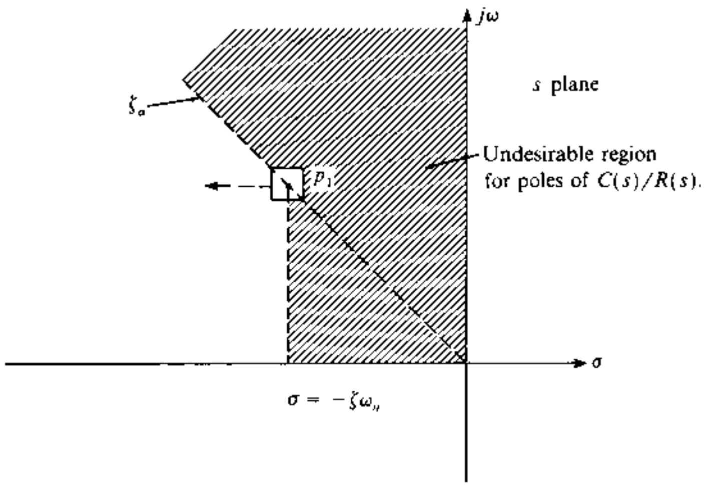


FIGURE 12.2 Location of the desired dominant pole.

CASE 2. It may be possible to satisfy the specifications by use of the following third-order model:

$$M_{T2}(s) = \frac{A}{s+d} M_{T1}(s) = \frac{A}{(s+d)(s^2 + 2\zeta\omega_n s + \omega_n^2)} \tag{12.3}$$

The dominant complex poles of Eq. (12.3) are determined in the same manner as for case 1; i.e., the values of ζ and ω_n are determined by using the specifications for M_p, t_p, T_s, and K_m. Because the model of Eq. (12.1) cannot satisfy all four specifications, the known effects of the real pole $s = -d$ in Eq. (12.3) can be used to try to achieve these specifications. With a CAD program (see appendixes C and D), various trial values of d are used to check whether or not the desired specifications can be achieved. Note that for each value of d, the corresponding value of A_2 must be equal to $A_2 = d\omega_n^2$ in order to satisfy the requirement that $e(t)_{ss} = 0$ for a step input. As an example, the specifications $M_p = 1.125$, $t_p = 1.5s$, and $T_s = 3.0s$ cannot be satisfied by the second-order control ratio of the form of Eq. (12.1), but they can be satisfied by

$$\frac{Y(s)}{R(s)} = \frac{22}{(s + 1 \pm j3)(s + 2.2)} \tag{12.4}$$

For a third-order all-pole open-loop plant, the model of Eq. (12.3) must be used. Then, if the model of Eq. (12.1) is satisfactory, the pole $-d$ is made a nondominant pole.

CASE 3. The third case deals with the third-order model [see Eq. (10.18)]

$$M_{T3}(s) = \frac{A(s+c)}{s+d} M_{T1}(s) = \frac{A_3(s+c)}{(s+d)(s^2 + 2\zeta\omega_n s + \omega_n^2)}$$
$$= \frac{A_3(s - z_1)}{(s-p_1)(s-p_2)(s-p_3)} \tag{12.5}$$

which has one zero. The zero can be a zero of the plant or a compensator zero that is inserted into the system to be used in conjunction with the pole $-d$ in order to satisfy the specifications. The procedure used in synthesizing Eq. (12.5) is based upon the nature of the zero; i.e., the zero (1) is not close to the origin or (2) it is close to the origin and is therefore a troublesome zero. (A stable system with a zero close to the origin may have a large transient overshoot for a step input.)

When the system has a step input, Fig. 12.3 represents the pole-zero diagram of $Y(s)$ for Eq. (12.5). Consider three possible locations of the real zero; to the right of, canceling, and to the left of the real pole $p_3 = -d$. When the zero cancels the pole p_3, the value of M_p is determined only by the complex-conjugate poles p_1 and p_2; when the zero is to the right of p_3, the value of M_p is larger (see Sec. 10.9); and when it is to the left, the value of M_p is smaller [see Eq. (4.68) and Sec. 10.9]. The amount of this increase or decrease is determined by how close the zero is to the pole p_3. Thus the optimum location of the real pole, to the right of the zero, minimizes the effect of the troublesome zero; i.e., it reduces the peak of the transient overshoot.

In order to achieve M_T, it may be necessary to insert additional equipment in cascade and/or in feedback with $G_x(s)$. Also, in achieving the desired model for a plant containing a zero, the design method may use (1) the

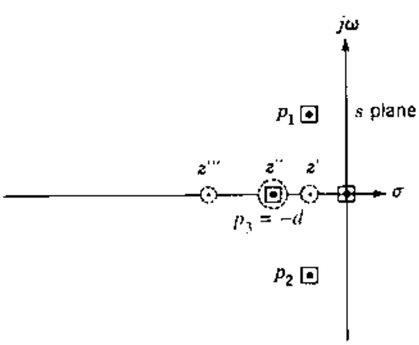


FIGURE 12.3 Pole-zero location of $Y(s)$ for Eq. (12.5), with a step input.

cancellation of the zero by a dominator pole, (2) the addition of a second zero, or (3) the replacement of the zero with one at a new location.

In synthesizing the desired control ratio, a system designer has a choice of one of the following approaches.

METHOD 1. Cancel some or all of the zeros of the control ratio by poles of the control ratio. The number of zeros to be canceled must be equal to or less than the excess of the number of poles n minus the number of desired dominant poles β. If necessary, cascade compensators of the form $G_c(s) = 1/(s - P_i)$ can be inserted into the open-loop system in order to increase the number of poles of $Y(s)/R(s)$ so that cancellation of zeros can be accomplished. The number of compensator poles γ required must increase the order of the system sufficiently to ensure that $\alpha \leq n + \gamma - \beta$. Theoretically it is possible to insert a compensator of this form just preceding the block that contains the unwanted zero to be canceled by the pole of this compensator. However, in many systems it is not physically possible to locate the compensator in this manner. Also, setting P_i equal to the value of the unwanted zero z_h of the plant results in an unobservable state. All the states must be controllable and observable in order to achieve an *optimal response*. Thus, to achieve an optimal response and to cancel the unwanted zero, choose P_i such that $P_i \neq z_h$.

METHOD 2. Relocate some or all of the zeros of $G_x(s)$ by using cascade compensators of the form $G_c(s) = (s - z_i)/(s - P_i)$. The P_i is the value of the zero of $G_x(s)$ that is to be canceled, and the zero z_i assumes the value of the desired zero location. The number γ of these compensators to be inserted equals the number of zeros α to be relocated. The values of z_i represent the desired zeros. The values of P_i are selected to meet the condition that α *poles of the control ratio* are used to cancel the α unwanted zeros.

METHOD 3. A combination of the above two approaches can be used; i.e., canceling some zeros by closed-loop poles and replacing some of them by the zeros of the cascade compensators $G_c(s)$.

The modeling technique discussed in this section requires that the poles and zeros of $Y(s)/R(s)$ be located in a pattern that achieves the desired time response. This is called the *pole-zero placement technique*. The number of poles minus zeros of $(Y/R)_T$ must equal or exceed the number of poles minus zeros of the plant transfer function. For a plant having w zeros, w of the closed-loop poles are either located "very close" to these zeros, in order to minimize the effect of these poles on $y(t)$, or suitably placed for the desired time response to be obtained. If there are β dominant poles, the remaining $n - \beta - w$ poles are located in the nondominant region of the s plane. Another method of selecting the locations of the poles and zeros of $M_T(s)$ is given in Ref. 3; it is based on a performance index that optimizes the closed-loop time responses. The specification of $Y(s)/R(s)$ is the basis for the Guillemin-Truxal method of Sec. 12.3 and for the state-feedback designs discussed in this and in the next chapter.

12.3 GUILLEMIN-TRUXAL DESIGN PROCEDURE [4]

The modeling technique discussed in this section requires that the poles and zeros of $Y(s)/R(s)$ be located in a pattern which achieves the desired time response. This is called the *pole-zero placement technique.* The number of poles minus zeros of $(Y/R)_T$ must equal or exceed the number of poles minus zeros of the plant transfer function. For a plant having w zeros, w of the closed-loop poles are either located "very close" to these zeros in order to minimize the effect of these poles on $y(t)$ or suitably placed for the desired time response to be obtained. If there are β dominant poles, the remaining $n - \beta - w$ poles are located in the nondominant region of the s plane. Another method of selecting the locations of the poles and zeros of $M_T(s)$ is given in Chap. 16 of Ref. 5; it is based on a performance index that optimizes the closed-loop time response. The specification of $Y(s)/R(s)$ is the basis for the Guillemin-Truxal method of this section and for the state-feedback designs discussed in the next chapter.

In contrast to designing the cascade compensator $G_c(s)$ of Fig. 12.4 based upon the analysis of the open-loop transfer function, the Guillemin-Truxal method is based upon designing $G_c(s)$ to yield a desired control ratio $[Y(s)/R(s)]_T = M_T(s)$. The control ratio for the unity-feedback cascade-compensated control system of Fig. 12.4 is

$$M_T(s) = \frac{Y(s)}{R(s)} = \frac{N(s)}{D(s)} = \frac{G_c(s)G_x(s)}{1 + G_c(s)G_x(s)} \tag{12.6}$$

where $N(s)$ and $D(s)$ represent the specified zeros and poles, respectively, of the desired control ratio. Solving this equation for the required cascade compensator $D(s)$ yields

$$G_c(s) = \frac{M_T(s)}{[1 - M_T(s)]G_x(s)} = \frac{N(s)}{[D(s) - N(s)]G_x(s)} \tag{12.7}$$

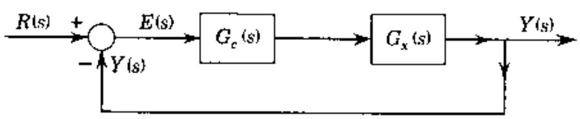


FIGURE 12.4 A cascade-compensated control system.

In other words, when $N(s)$ and $D(s)$ have been specified, the required transfer function for $D(s)$ is obtained. Network-synthesis techniques are then utilized to synthesize a network having the required transfer function of Eq. (12.7). Usually a passive network containing only resistors and capacitors is desired. Option 28 of the TOTAL-PC CAD package can be utilized to determine $G_c(s)$ [2]. Further details of the Guillemin-Truxal method are given in Ref. 6.

Example 1. The desired control ratio of a unity-feedback control system is

$$\frac{Y(s)}{R(s)} = \frac{N(s)}{D(s)} = \frac{210(s+1.5)}{(s+1.75)(s+16)(s+1.5\pm j3)} \tag{12.8}$$

The numerator constant has been selected to yield zero stedy-state error with a step input and the desired figures of merit are $M_p = 1.28$, $t_p = 1.065s$, and $t_s = 2.54s$.

$$G_x(s) = \frac{4}{s(s+1)(s+5)} \tag{12.9}$$

The required cascade compensator, determined by substituting Eqs. (12.8) and (12.9) into Eq. (12.7), is

$$\begin{aligned} G_c(s) &= \frac{52.5s(s+1)(s+1.5)(s+5)}{s^4 + 20.75s^3 + 92.6s^2 + 73.69s} \\ &= \frac{52.5s(s+1)(s+1.5)(s+5)}{s(s+1.02)(s+4.88)(s+14.86)} \approx \frac{52.5(s+1.5)}{s+14.86} \end{aligned} \tag{12.10}$$

The approximation made in $G_c(s)$ yields a simple, physically realizable, minimum-phase lead network with an $\alpha = 0.101$. Based on a root-locus analysis of $G_c(s)G_x(s)$, the ratios of the terms $(s+1)/(s+1.02)$ and $(s+5)/(s+4.88)s$ of $G_c(s)$ have a negligible effect on the location of the desired dominant poles of $Y(s)/R(s)$. With the simplified $G_c(s)$ the root-locus plot, using $G_c(s)G_x(s)$, is shown in Fig. 12.5, and the control ratio achieved is

$$\frac{Y(s)}{R(s)} = \frac{210(s+1.5)}{(s+1.755)(s+16.0)(s+1.52\pm j3.07)}$$

Because this equation is very close to that of Eq. (12.8), it is satisfactory. Note that the root-locus plot of $G_c(s)\,G_x(s) = -1$ has asymptotes of 180° and $\pm 60^\circ$; therefore the system becomes unstable for high gain.

The Guillemin-Truxal design method, as illustrated by this example, involves three steps:

1. Specifying the desired zeros, poles, and numerator constant of the desired closed-loop function $Y(s)/R(s)$ in the manner discussed in Sec. 12.2.

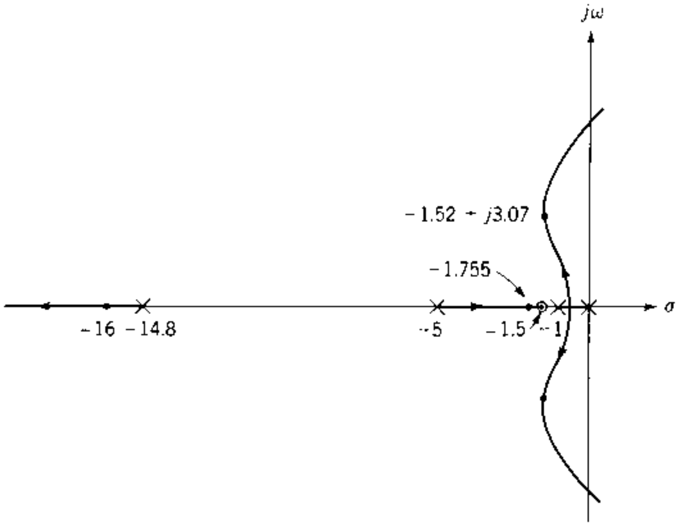


FIGURE 12.5 Root-locus plot for Example 1.

2. Solving for the required cascade compensator $G_c(s)$ from Eq. (12.7).
3. Synthesizing a physically realizable compensator $G_c(s)$, preferably a passive network.

Using practical cascade compensators, the number of poles minus zeros of the closed-loop system must be equal to or greater than the number of poles minus zeros of the basic open-loop plant. This must include the poles and zeros introduced by the cascade compensator. The practical aspects of system synthesis using the Guillemin-Truxal method impose the limitation that the poles of $G_x(s)$ must lie in the left-half s plane. Failure to exactly cancel poles in the right-half s plane would result in an unstable closed-loop system. The state-variable-feedback method, described in the next chapter, is a compensation method that achieves the desired $Y(s)/R(s)$ without this limitation.

12.4 INTRODUCTION TO DISTURBANCE REJECTION [6,7]

The previous chapters and Sec. 12.3 deal with designing a control system whose output follows (tracks) the input signal. The design approach for the disturbance-rejection problem, by contrast, is based upon the performance specification that the output of the control system of Fig. 12.6 must not be affected by a system disturbance input $d(t)$. In other words, the steady-state

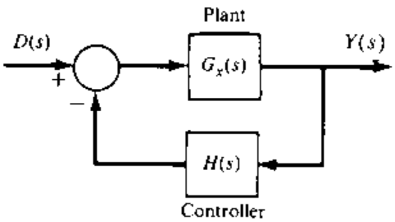


FIGURE 12.6 A simple disturbance-rejection control system.

output $y(t)_{ss} = 0$ for the input $d(t) \neq 0$. Consider the control ratio

$$\frac{Y(s)}{D(s)} = \frac{K(s^w + c_{w-1}s^{w-1} + \cdots + c_1 s + c_0)}{s^n + q_{n-1}s^{n-1} + \cdots + q_1 s + q_0} \tag{12.11}$$

Assuming a stable system and a step disturbance $D(s) = D_0/s$, the output is given by

$$Y(s) = \frac{K(s^w + c_{w-1}s^{w-1} + \cdots + c_1 s + c_0)}{s(s^n + q_{n-1}s^{n-1} + \cdots + q_1 s + q_0)} \tag{12.12}$$

The desired output $y(t)_{ss} = 0$ is achieved only if $c_0 = 0$. *This requires that the numerator of $Y(s)/D(s)$ have at least one zero at the origin.*

It is desired for this type of control problem that the disturbance have no effect on the steady-state output; also, the resulting transient must die out as fast as possible with a limit α_p on the maximum magnitude of the output.

The next three sections present a frequency-domain design technique for minimizing the effect of a disturbance input on the output of a control system. This technique is enhanced by simultaneously performing a qualitative root-locus analysis for a proposed feedback-compensator structure.

12.5 A SECOND-ORDER DISTURBANCE-REJECTION MODEL

A simple second-order continuous-time disturbance-rejection model [7] transfer function for the system of Fig. 12.6 is

$$M_D(s) = \frac{Y(s)}{D(s)} = \frac{G_x}{1 + G_x H} = \frac{K_x s}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{12.13}$$

where $K_x > 0$. For a unit-step disturbance $D(s) = 1/s$ and $y(t)_{ss} = 0$. Equation (12.13) yields (see Transform Pair 24*a* in Appendix A)

$$y(t) = \mathscr{L}^{-1}\left[\frac{K_x}{s^2 + 2\zeta\omega_n s + \omega_n^2}\right]$$
$$= \frac{K_x}{\omega_n\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t} \sin \omega_n\sqrt{1-\zeta^2}\, t \qquad (12.14)$$

where $a^2 + b^2 = \omega_n^2$, $b = \omega_d$, and $a = |\sigma| = \zeta\omega_n$. The time domain or frequency domain approaches can be used to determine appropriate model values of ζ and ω_n, while K_x is assumed fixed. The application of these methods is described below.

Time Domain

The time response of Eq. (12.14) is plotted vs. $\omega_n t$ in Fig. 4.14 for several values of $\zeta \leq 1$. By setting the derivative of $y(t)$ with respect to time equal to zero, it can be shown that the maximum overshoot occurs at the time

$$t_p = \frac{\cos^{-1}\zeta}{\omega_n\sqrt{1-\zeta^2}} \qquad (12.15)$$

and the maximum value of $y(t)$ is

$$y(t_p) = \frac{K_x}{\omega_n}\exp\left(-\frac{\zeta\cos^{-1}\zeta}{\sqrt{1-\zeta^2}}\right) \qquad (12.16)$$

Letting $K_x = v\omega_n^2$ (with $v > 0$), Eq. (12.16) is rearranged to the form

$$\frac{y(t_p)}{\omega_n} = v\exp\left(-\frac{\zeta\cos^{-1}\zeta}{\sqrt{1-\zeta^2}}\right) \qquad (12.17)$$

A plot of Eq. (12.17) vs. ζ is shown in Fig. 12.7.

Frequency Domain

The frequency response function of Eq. (12.13) is

$$\mathbf{M}_D(j\omega) = \frac{\mathbf{Y}(j\omega)}{\mathbf{D}(j\omega)} = \frac{(K_x/\omega_n^2)(j\omega)}{[1-(\omega/\omega_n)^2] + j(2\zeta\omega/\omega_n)} = \frac{K_x}{\omega_n^2}\mathbf{M}'_D(j\omega) \qquad (12.18)$$

A straight-line approximation of the Bode plot of Lm $\mathbf{M}_D(j\omega)$ vs. log ω (drawn on semilog paper) is shown in Fig. 12.7*b* for a selected value of ω_n. This plot reveals that

1. The closed-loop transfer function $\mathbf{M}_D(j\omega)$ must attenuate all frequencies present in the disturbance input $\mathbf{D}(j\omega)$. The plot of

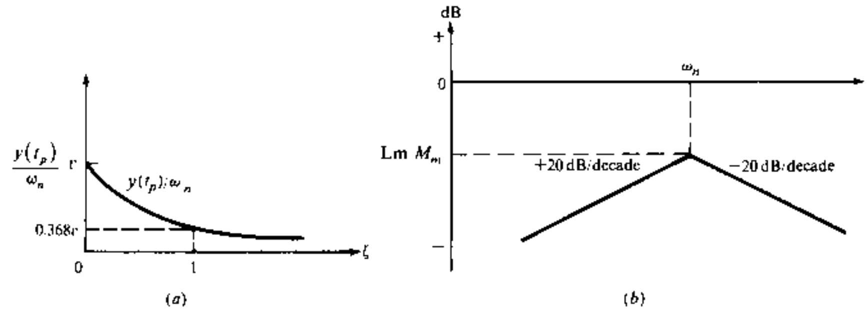


FIGURE 12.7 (*a*) Peak value of Eq. (12.14), using Eq. (12.17). (*b*) Bode plot characteristics of Eq. (12.18).

Lm $\mathbf{M}'_D(j\omega)$ vs. ω is raised or lowered by an amount equal to $\mathrm{Lm}(K_x/\omega_n^2)$. The value of (K_x/ω_n^2) therefore determines the maximum value M_m shown in Fig. 12.7*b*.

2. As ω_n increases, the peak value moves to the right to a higher frequency.

Selection of the desired disturbance model of Eq. (12.13) is based on both the time-response and the frequency-response characteristics. For example, selection of ω_n determines the bandwidth, as shown in Fig. 12.7*b*. Also, the maximum value of the permitted peak overshoot, as shown in Figs. 4.14 and 12.7*a*, is determined by the value of ζ. In addition, the peak time t_p and the settling time $t_s = 4/\zeta\omega_n$ of the transient may be used to determine the parameters in the model of Eq. (12.13). Some trial and error may be necessary to ensure that all specifications for the disturbance model are satisfied.

12.6 DISTURBANCE-REJECTION DESIGN PRINCIPLES FOR SISO SYSTEMS [7]

The desired disturbance model $M_D(s)$ of a system must be based on the specified time-domain and frequency-domain characteristics as described in Secs. 12.4 and 12.5. In order to obtain the desired characteristics of $M_D(s)$, it is necessary to design the required feedback controller $H(s)$ in Fig. 12.6. A first trial may be $H(s) = K_H/s$. If this choice of $H(s)$ does not produce the desired characteristics, then a higher-order feedback transfer function must be used. The order of the transfer functions in the system may also require a disturbance model of higher than second order.

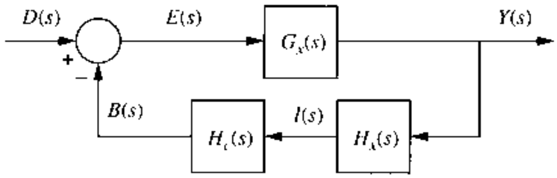


FIGURE 12.8 A control system with a disturbance input.

The approximations developed in Sec. 11.9 are now restated for the disturbance rejection system of Fig. 12.8.

$$\frac{\mathbf{Y}(j\omega)}{\mathbf{D}(j\omega)} = \frac{\mathbf{G}_x(j\omega)}{1 + \mathbf{G}_x(j\omega)\mathbf{H}(j\omega)} \tag{12.19}$$

where $\mathbf{H}(j\omega) = \mathbf{H}_x(j\omega)\mathbf{H}_c(j\omega)$. This control ratio can be approximated by

$$\frac{\mathbf{Y}(j\omega)}{\mathbf{D}(j\omega)} \approx \begin{cases} \mathbf{G}_x(j\omega) & \text{for} \quad |\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)| \ll 1 \qquad (12.20) \\ \dfrac{1}{\mathbf{H}(j\omega)} & \text{for} \quad |\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)| \gg 1 \qquad (12.21) \end{cases}$$

The condition when $|\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)| \approx 1$ is still undefined, in which case neither of these conditions is applicable, In the approximate procedure the straight-line Bode plots are used and this condition is neglected. Therefore Eqs. (12.20) and (12.21) are used when $|\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)| < 1$ and $|\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)| > 1$, respectively. This approximation, along with a root-locus sketch of $G_x(s)H(s) = -1$, allows investigation of the qualitative results to be obtained. After these results are found to be satisfactory, the refinements for an exact solution are introduced.

The design procedure using the approximations of Eqs. (12.20) and (12.21) follows.

Step 1. Using the straight-line approximations, plot Lm $\mathbf{G}_x(j\omega)$ vs. log ω.

Step 2. Assume that the feedback compensator $H_c(s)$ is put in cascade with a specified feedback unit $H_x(s)$, in order to yield the desired system performance. To use the straight-line approximations, with $\mathbf{H}(j\omega) = \mathbf{H}_x(j\omega)\mathbf{H}_c(j\omega)$, plot Lm $1/\mathbf{H}(j\omega)$ vs. ω. Note that:

1. The condition Lm $\mathbf{G}_x(j\omega)\mathbf{H}(j\omega) = 0$ may also be expressed as

$$\text{Lm}\,\mathbf{G}_x(j\omega) = -\text{Lm}\,\mathbf{H}(j\omega) = \text{Lm}\,1/\mathbf{H}(j\omega) \tag{12.22}$$

This condition is satisfied at the intersection of the plots of $\mathbf{G}_x(j\omega)$ and $1/\mathbf{H}(j\omega)$. The frequency at the intersections represents the boundary between the two regions given by Eqs. (12.20) and (12.21).

There must be at least one intersection between the plots in order for both approximations to be valid. Disturbance rejection over the frequency band of $0 \le \omega \le \omega_b$, the condition of Eq. (12.21), results in only one intersection.

2. For disturbance rejection it is desired to have $\mathrm{Lm}\,\mathbf{Y}(j\omega)/\mathbf{D}(j\omega) < \mathrm{Lm}\,\mathbf{M}_D$. Also, Eq. (12.21) must hold for $0 \le \omega \le \omega_b$ in order to yield the desired frequency-domain specifications. To satisfy these objectives it is usually necessary for the gain constant of $H(s)$ to be very high.

Step 3. Plot $\mathrm{Lm}\,\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)$ vs. ω by using the straight-line approximation, and determine the slope of the plot at the 0-dB crossover point. Achieving a stable system requires (see Sec. 8.17) that this slope be greater than –40 dB/decade if the adjacent corner frequencies are not close. Also, the loop transmission frequency ω_ϕ (the 0-dB crossover or phase-margin frequency) of $\mathrm{Lm}\,\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)$ must be within the allowable specified value in order to attenuate high-frequency loop noise. Placing zeros of $H_c(s)$ "too close" to the imaginary axis may increase the value of ω_ϕ. It may be necessary at this stage of the design procedure to modify $H_c(s)$ in order to satisfy the stability requirement and the frequency-domain specifications.

Step 4. Once a trial $H(s)$ is found that satisfies the frequency-domain specifications, sketch the root locus $[G_x(s)H(s) = -1]$ to ascertain that (*a*) at least a conditionally stable system exists, and (*b*) it is possible to satisfy the time-domain specifications. [Section 12.5 discusses the desired locations of the poles of $Y(s)$ that achieve the desired specifications (see also Prob. 12.7)]. If the sketch seems reasonable, (1) obtain the exact root-locus plots and select a value of the static loop sensitivity $K = K_G K_x K_c$ and (2) for this value of K, obtain a plot of $y(t)$ vs. t in order to obtain the figures of merit.

Step 5. If the results of step 4 do not yield satisfactory results, then use these results to select a new $H_c(s)$ and repeat steps 2 through 4.

Example 2. Consider the nonunity-feedback system of Fig. 12.8, where

$$G_x(s) = \frac{K_G}{s(s+1)} = \frac{1}{s(s+1)} \tag{12.23}$$

and

$$H_x(s) = \frac{K_x}{s+200} = \frac{200}{s+200} \tag{12.24}$$

Because a disturbance input is present, it is necessary to design the compensator $H_c(s) = K_c H_c'(s)$ so that the control system of Fig. 12.8 satisfies the

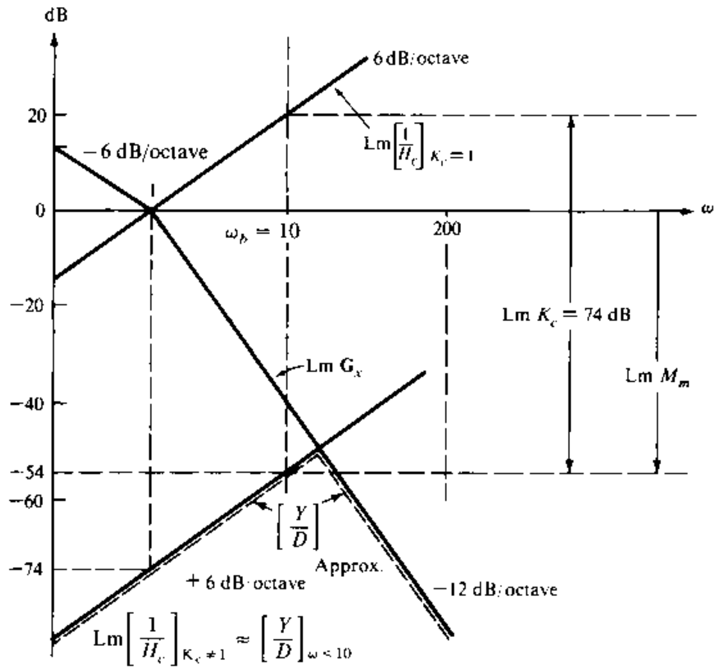


FIGURE 12.9 Log magnitude plots of Example 2.

following specifications for a unit-step disturbance input $d(t)$: $|y(t_p)| \leq 0.002$, $y(t)_{ss} = 0$, and $t_p \leq 0.2$ s. Using the empirical estimate that $M_p = |y(t_p)/d(t)| \approx M_m$ gives the specification $\mathrm{Lm}\, M_m \leq \mathrm{Lm}\, 0.002 = -54$ dB for the desired model $M_D(s)$. It is also estimated that the bandwidth of the disturbance transfer function model $\mathbf{Y}(j\omega)/\mathbf{D}(j\omega)$ is $0 \leq \omega \leq 10$ rad/s.

Trial Solution

Step 1. The plot of Lm $\mathbf{G}_x(j\omega)$ vs. ω is shown in Fig. 12.9.

Step 2. Because $G_x(s)$ does not have a zero at the origin and $Y(s)/D(s)$ does require one for disturbance rejection [so that $y(\infty)_{ss} = 0$], then $H(s)$ must have a pole at the origin. Thus the feedback transfer function must have the form

$$H(s) = H_x(s)H_c(s) = \frac{K_H(s + a_1)\cdots}{s(s + 200)(s + b_1)\cdots} \tag{12.25}$$

where the degree of the denominator is equal to or greater than the degree of the numerator. As a first trial, let

$$H_c(s) = \frac{K_c}{s} \tag{12.26}$$

Then $K_H = K_x K_c = 200K_c$. To determine the value of K_c required to satisfy $\mathrm{Lm}\,\mathbf{Y}(j\omega)/\mathbf{D}(j\omega) \le -54\,\mathrm{dB}$ for $0 \le \omega \le 10\,\mathrm{rad/s}$, initially plot the curve for $[\mathrm{Lm}\,1/\mathbf{H}_c(j\omega)]_{K_c=1}$. The value of K_c is then selected so that the plot of $\mathrm{Lm}\,1/\mathbf{H}_c(j\omega)$ is located such that it is on or below the -54 dB line for the frequency range $\omega \le \omega_b = 10$. Because the corner frequency of $\mathbf{H}(j\omega)$ at $\omega = 200$ lies outside the desired bandwidth, $\mathrm{Lm}\,1/H_c(j\omega) = \mathrm{Lm}\,1/H(j\omega)$ is the controlling factor in the frequency range $\omega \le 10$. Next, lower this plot, as shown in the figure, until the plot of $\mathrm{Lm}\,1/H(j\omega)$ yields the desired value of -54 dB at $\omega_b = 10\,\mathrm{rad/s}$. Equation (12.21), for this case, satisfies the requirement $|\mathbf{M}(j\omega)| \le M_m$ over the desired bandwidth and is below -54 dB. $\mathrm{Lm}\;200K_c = 74\,\mathrm{dB}$ is required to lower the plot of $\mathrm{Lm}\,1/\mathbf{H}(j\omega)$ to the required location. With this value of K_c the resulting first-trial feedback transfer function is

$$H(s) = \frac{5000}{s(s+200)} \tag{12.27}$$

Step 3. An analysis of

$$\mathbf{G}_x(j\omega)\mathbf{H}(j\omega) = \frac{K_c}{(j\omega)^2(1+j\omega)\left(1+j\dfrac{\omega}{200}\right)} \tag{12.28}$$

reveals that the plot of $\mathrm{Lm}\,[\mathbf{G}_x(j\omega)\mathbf{H}(j\omega)]$ vs. ω crosses the 0-dB axis with a minimum slope of -40 dB/decade regardless of the value of K_c. This crossover characteristic is indicative of an unstable system.

Step 4. This is confirmed by the root-locus sketch in Fig. 12.10, for

$$G_x(s)H(s) = \frac{K_H}{s^2(s+1)(s+200)} = -1 \tag{12.29}$$

which reveals a completely unstable system for the $H(s)$ of Eq. (12.27).

The next trial requires that at least one zero be added to Eq. (12.27) in order to make the system at least conditionally stable while maintaining the desired bandwidth. A final acceptable design of $H_c(s)$ that meets the system specifications requires a trial-and-error procedure. A solution to this design problem is given in Sec. 12.7.

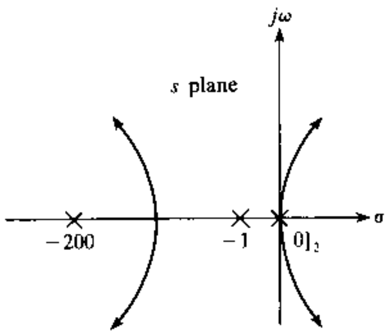


FIGURE 12.10 A root-locus sketch for Eq. (12.29)

Example 3. Consider the nonunity-feedback system of Fig. 12.8, where

$$G_x(s) = \frac{2}{s+2}$$

and

$$H(s) = H_x(s)H_c(s) = \frac{K_H}{s}$$

Determine the response $y(t)$ for a step disturbance input $d(t) = u_{-1}(t)$ as the gain K_H is increased.

The system output is

$$Y(s) = \frac{2}{s^2 + 2s + 2K_H} = \frac{2}{(s-p_1)(s-p_2)} = \frac{A_1}{(s-p_1)} + \frac{A_2}{(s-p_2)} \tag{12.30}$$

Thus

$$y(t) = A_1 e^{p_1 t} + A_2 e^{p_2 t} \tag{12.31}$$

where $p_{1,2} = -1 \pm j\sqrt{2K_H - 1}$. For $K_H > 0.5$,

$$A_1 = \frac{1}{j\sqrt{2K_H - 1}} \qquad A_2 = -A_1$$

The coefficients of the output transient, as $K_H \to \infty$, are

$$\lim_{K_H \to \infty} A_1 = \lim_{K_H \to \infty} A_2 = 0$$

For this example, as K_H is made larger, the coefficients A_1 and A_2 in Eq. (12.31) become smaller. Thus

$$\lim_{K_H \to \infty} y(t) = 0$$

which is the desired result for disturbance rejection. The output for a step disturbance input has two desired properties. First, the steady-state output is zero because the feedback contains an integrator. Second, the magnitude of the transient becomes smaller as the feedback gain K_H increases. (Note that the value of K_H is limited by practical considerations). A similar analysis can be made for a more complicated system (see Probs. 12.8 and 12.9). In some applications it may not be necessary to achieve $y(\infty)_{ss} = 0$.

12.7 DISTURBANCE-REJECTION DESIGN EXAMPLE

In Example 2, the feedback compensator $H_c(s) = K_c/s$ results in an unstable system. An analysis of the root locus in Fig. 12.10 indicates that $H_c(s)$ must contain at least one zero near the origin in order to pull the root-locus branches into the left-half s plane. In order to minimize the effect of an unwanted disturbance on $y(t)$, it is necessary to minimize the magnitude of all the coefficients of

$$y(t) = A_1 e^{p_1 t} + \cdots + A_n e^{p_n t} \tag{12.32}$$

Problem 4.18 shows that this requires the poles of $Y(s)/D(s)$ to be located as far to the left in the s plane as possible. Figure 12.11 graphically describes the result of the analysis made in Prob. 12.7 for the desired placement of the control-ratio poles. The root-locus branch which first crosses the imaginary axis determines the maximum value of gain for a stable system. The root selected on this

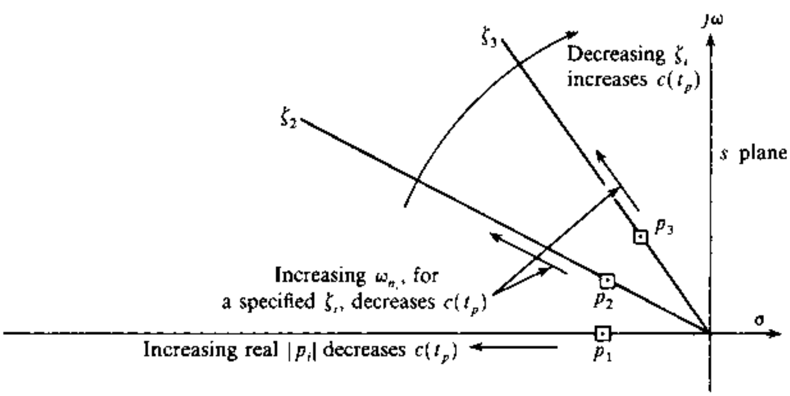


FIGURE 12.11 Effects on the peak overshoot to a disturbance input due to changes in the parameters ζ_i and ω_{n_i} of complex poles and the values of the real poles of the control ratio.

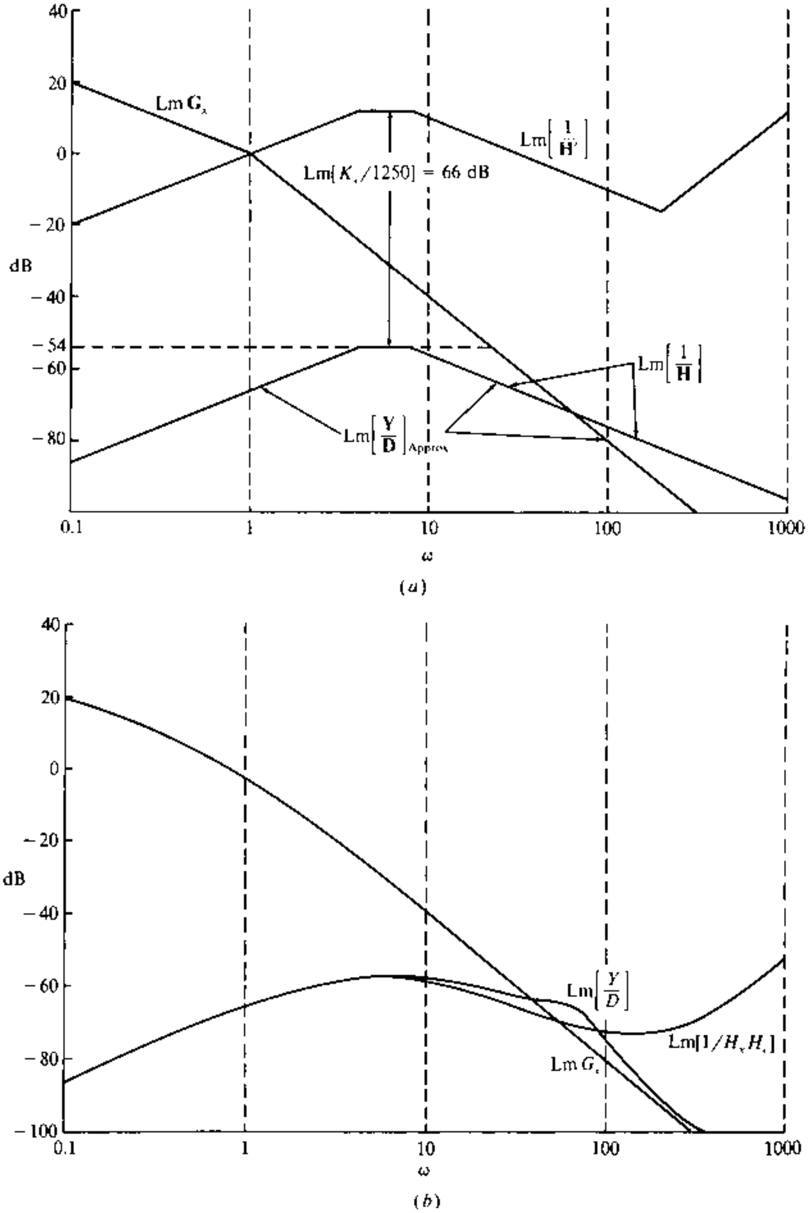


FIGURE 12.12 Bode plot for Example 4: (*a*) straight-line approximations; (*b*) exact poles.

branch should have $0.5 < \zeta < 0.7$, which has been determined empirically. When there is a choice of roots for a given ζ, the root selected should have the largest value of gain and ω_n. These properties are indicated in Fig. 12.11.

Example 4. The system of Example 2, Sec. 12.6, is now redesigned in order to achieve a stable system. Also, the root-locus branches must be pulled as far to the left as possible. Therefore, assume the following compensator, which contains two zeros:

$$H_c(s) = \frac{K_c(s+a_1)(s+a_2)}{s(s+b_1)(s+b_2)} \tag{12.33}$$

As a trial, let $a_1 = 4$ and $a_2 = 8$. The poles $-b_1$ and $-b_2$ must be chosen as far to the left as possible; therefore values $b_1 = b_2 = 200$ are selected. The plot of $\text{Lm}\, 1/\mathbf{H}_x(j\omega)\mathbf{H}_c(j\omega) \le -54\,\text{dB}$, where $|\mathbf{Y}(j\omega)/\mathbf{D}(j\omega)| \approx |1/\mathbf{H}_x(j\omega)\mathbf{H}_c(j\omega)|$ over the range $0 \le \omega \le \omega_b = 10\,\text{rad/s}$, is shown in Fig. 12.12. This requires that $200K_c \ge 50 \times 10^7$. The open-loop transfer function of Fig. 12.8 is now

$$\begin{aligned} G_x(s)H_x(s)H_c(s) &= \frac{200K_c(s+4)(s+8)}{s^2(s+1)(s+200)^3} \\ &= \frac{(K_c/1250)(0.25s+1)(0.125s+1)}{s^2(s+1)(0.005s+1)^3} \end{aligned} \tag{12.34}$$

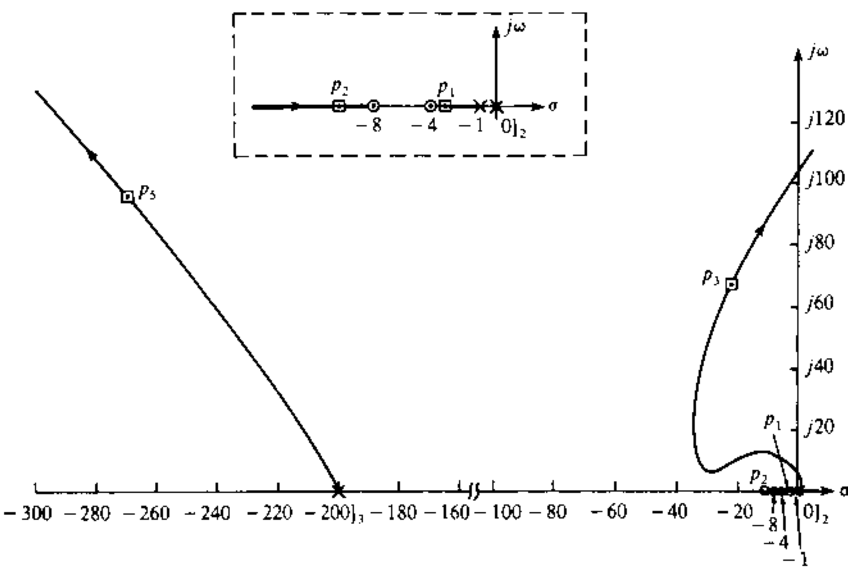


FIGURE 12.13 Sketch of root locus for Eq. (12.34). Not to scale.

The poles and zeros of Eq. (12.34) are plotted in Fig. 12.13. The value of static-loop sensitivity $200K_c = 5 \times 10^8$ yields the set of roots

$$p_1 = -3.8469 \qquad p_2 = -10.079$$
$$p_{3,4} = -23.219 \pm j66.914 \qquad p_{5,6} = -270.32 \pm j95.873$$

that are shown in Fig. 12.13. This choice of gain results in a satisfactory design, because $M_p = 0.001076 \leq 0.002$ (−54 dB), $t_p = 0.155$ s (<0.2 s), $y(t)_{ss} = 0$, and $0 \leq \omega \leq \omega_b = 10$ rad/s. Note that the gain can be reduced until either $M_p = 0.002$ with $t_p \leq 0.2$ s or $t_p = 0.2$ s with $M_p \leq 0.002$ is achieved. Although not specified in this problem, the phase-margin frequency $\omega_\phi = 56.2$ rad/s must also be taken into account in the design of a practical system.

12.8 DISTURBANCE-REJECTION MODELS

The multiple-input single-output (MISO) control system of Fig. 12.14 has three inputs, $r(t)$, $d_1(t)$, and $d_2(t)$, which represent a tracking and two disturbance inputs, respectively. The previous chapters are concerned with the tracking problem alone, with $d(t) = 0$. This section uses the pole-zero placement method for the development of a desired disturbance-rejection control-ratio model $(Y/R)_D = M_D$. The specification for disturbance rejection for the control system of Fig. 12.14, with a step disturbance input $d_2(t) = D_0 u_{-1}(t)$ and $r(t) = d_1(t) = 0$ is

$$|y(t)| \leq \alpha_p \qquad \text{for } t \geq t_x \tag{12.35}$$

as shown in Fig. 12.15. Note that the initial value of the output for this condition is $y(0) = D_0$. When $d_1(t) = D_0 u_{-1}(t)$ and $r(t) = d_2(t) = 0$, then the requirement for disturbance rejection is

$$|y(t_p)| \leq \alpha_p \tag{12.36}$$

as shown in Fig. 12.16, where $y(0) = 0$. The simplest control-ratio models for Figs. 12.15 and 12.16, respectively, are shown in Table 12.1.

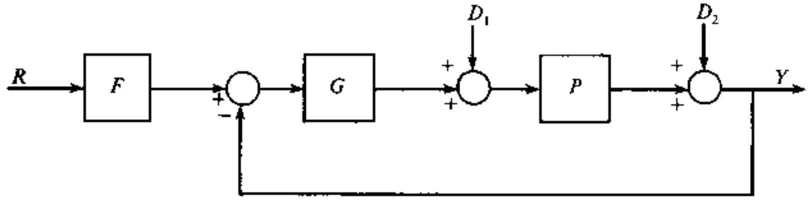


FIGURE 12.14 A multiple-input single-output (MISO) system.

TABLE 12.1 Disturbance Rejection Control Ratio Models

$r(t) = d_1(t) = 0$ and $d_2(t) = D_0 u_{-1}(t)$ (See Fig. 12.15)

$M_D = \dfrac{Y_1(s)}{D_2(s)} = \dfrac{s}{s+a}$	(12.37)	$Y_1(s) = \dfrac{D_0}{s+a}$		$y_1(t) = D_0 e^{-at}$	(12.39)
$M_D = \dfrac{Y_2(s)}{D_2(s)} = \dfrac{s(s+a)}{(s+a)^2 + b^2}$	(12.38)	$Y_2(s) = \dfrac{D_0(s+a)}{(s+a)^2 + b^2}$		$y_2(t) = D_0 e^{-at} \cos bt$	(12.40)

$r(t) = d_2(t) = 0$ and $d_1(t) = D_0 u_{-1}(t)$ (See Fig. 12.16)

$M_D = \dfrac{Y_3(s)}{D(s)} = \dfrac{K_D s}{(s+a)(s+b)}$	(12.41)	$Y_3(s) = \dfrac{K_D D_0}{(s+a)(s+b)}$		$Y_3(t) = \dfrac{K_D D_0}{b-a}(e^{-at} - e^{-bt})$	(12.43)
$M_D = \dfrac{Y_4(s)}{D_1(s)} = \dfrac{K_D s}{(s+a)^2 + b^2}$	(12.42)	$Y_4(s) = \dfrac{K_D D_0}{(s+a)^2 + b^2}$		$Y_4(t) = \dfrac{K_D D_0}{b} e^{-at} \sin bt$	(12.44)

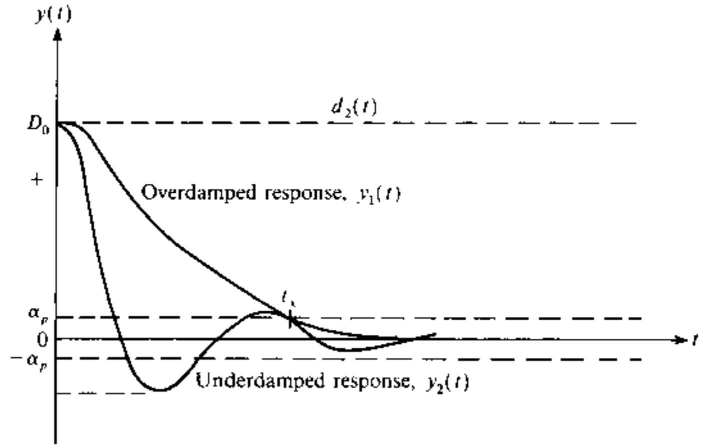


FIGURE 12.15 Disturbance-rejection time response for Fig. 12.14 with $r(t) = d_1(t) = 0$ and $d_2(t) = D_0 u_{-1}(t)$.

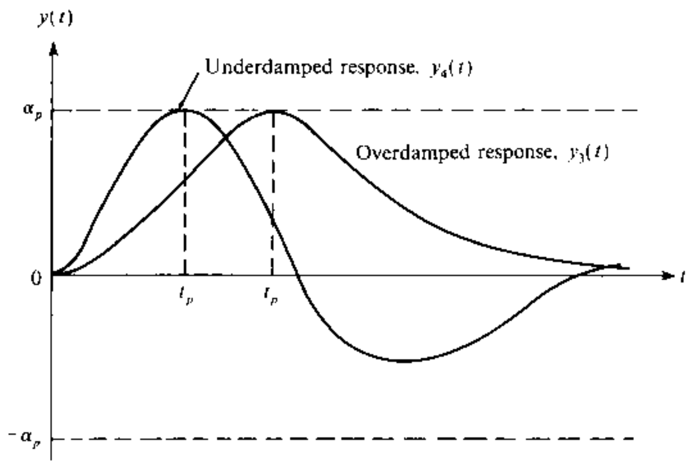


FIGURE 12.16 Disturbance-rejection time response for Fig. 12.14 with $r(t) = d_2(t) = 0$ and $d_1(t) = D_0 u_{-1}(t)$.

With the values of α_p, t_x, and D_0 specified, the value of a for the case represented by Eqs. (12.37) and (12.39) is given by

$$a = -\frac{1}{t_x}\ln\left(\frac{\alpha_p}{D_0}\right) \tag{12.45}$$

For the case represented by Eqs. (12.38) and (12.40), then

$$a = -\frac{1}{t_x}\ln\left(\frac{\alpha_p}{D_0 \cos bt_x}\right) \tag{12.46}$$

In Eqs. (12.38), the two quantities a and b must be determined. Since Eq. (12.46) shows that a is a function of b, a trial-and-error procedure can be used to determine suitable values. Thus, assume a value of b and determine the corresponding value of a. If the time-response characteristics are satisfactory, the model is completely determined. If not, try another value for b.

Consider the case where the settling time specification is given by

$$|y(t_s)| = \left|D_0 e^{-at_s} \cos bt_s\right| = 0.02\alpha_p = \alpha_s \tag{12.47}$$

that is, the output decays to 2 percent of the maximum value at $t = t_s$. With t_s specified, then Eq. (12.40) can be used to obtain the value of a, i.e.,

$$a = -\frac{1}{t_s}\ln\left(\frac{\alpha_s}{D_0 \cos bt_s}\right) \tag{12.48}$$

If all four quantities (t_p, t_s, α_p, and D_0) are specified, then the intersection of the plots of Eqs. (12.46) and (12.48), if one exists, yields the value of a that meets all the specifications.

For the case represented by Eqs. (12.41) and (12.43), assume that the values of α_p, α_s, t_p, t_s, and D_0 are specified. Then, from Eq. (12.43),

$$\alpha_p = K_\alpha\left(e^{-at_p} - e^{-bt_p}\right) > 0 \tag{12.49}$$

where $K_\alpha = K_D D_0/(b - a)$. Also from Eq. (12.43)

$$\left.\frac{dy(t)}{dt}\right|_{t=t_p} = K_\alpha\left(-ae^{-at_p} + be^{-bt_p}\right) = 0 \tag{12.50}$$

At $t = t_s$, Eq. (12.43) yields

$$\alpha_s = K_\alpha(e^{-at_s} - e^{-bt_s}) \tag{12.51}$$

From Eqs. (12.49) and (12.50) obtain, respectively

$$e^{-at_p} = \frac{\alpha_p}{K_\alpha} + e^{-bt_p} \tag{12.52}$$

and

$$\frac{b}{a} = e^{(b-a)t_p} \tag{12.53}$$

Also, from Eq. (12.51) with $K_\alpha > 0$, obtain

$$e^{-at_s} > e^{-bt_s} \rightarrow 1 > \frac{e^{-(b/a)at_s}}{e^{-at_s}} = e^{(1-(b/a))at_s} \tag{12.54}$$

Equation (12.54) is valid only when $b/a > 1$, or

$$b > a \tag{12.55}$$

Thus, assume a value for b/a, and for various values of b (or a), obtain values for a (or b) from Eqs. (12.51) and (12.53). Obtain a plot of a vs. b for each of these equations. If these plots intersect, then the values of a or b at the intersection result in a transfer function $M_D(s)$ that satisfies the desired performance specifications. If no intersection exists, assume another value for b/a and repeat the process.

If the desired performance specifications cannot be obtained by the simple models of Table 12.1, then poles and/or zeros must be added, by a trial-and-error procedure, until a model is synthesized that is satisfactory. For either case, it may be necessary to add nondominant poles and/or zeros to $M_D(s)$ in order that the degrees of its numerator and denominator polynomials are consistent with the degrees of the numerator and denominator polynomials of the control ratio of the actual control system.

12.9 SUMMARY

This chapter presents the concepts of synthesizing a desired control-ratio model for a system's output in order to either track a desired input or to reject a disturbance input. These models are based on the known characteristics of a second-order system. Once these simple models are obtained, nondominant poles and/or zeros can be added to ensure that the final model is compatible with the degree of the numerator and denominator ploynomials of the control ratio of the actual control system. The Guillemin-Truxal method is presented as one design method that requires the use of a tracking model. A disturbance-rejection frequency-domain design technique is presented. This technique is based upon the known correlation between the frequency domain, time domain, and the s plane.

REFERENCES

1. Voland, G.: *Control Systems Modeling and Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1986.
2. Houpis, C. H., and S. J. Rasmussen: *Quantitative Feedback Theory: Fundamentals and Application*. Marcel Dekker, New York, 1999.
3. Moore, B. C.: "On the Flexibility Offered by State Feedback in Multivariable Systems Beyond Closed-Loop Eigenvalue Assignment." *IEEE Trans. Autom. Control*. vol. AC-21, pp. 689–692.

4. Truxal, J. G.: *Automatic Feedback Control System Synthesis*, McGraw-Hill, New York, 1955.
5. D'Azzo, J. J., and C. H. Houpis: *Linear Control System Analysis and Design: Conventional and Modern*, 3rd ed., McGraw-Hill, New York, 1988.
6. D'Azzo, J. J., and C. H. Houpis: *Feedback Control System Analysis and Synthesis*, 2nd ed., McGraw Hill, New York, 1966.
7. Houpis, C. H., and G. B. Lamont: *Digital Control Systems: Theory, Hardware, Software*, 2nd ed., McGraw-Hill, New York, 1992.

13

Design: Closed-Loop Pole-Zero Assignment (State-Variable Feedback)

13.1 INTRODUCTION

Chapter 10 to 12 deal with the improvement of system performance in the time and frequency domains, respectively, by conventional design procedures. Such procedures are based on an analysis of the open-loop transfer function and yield an improved closed-loop response as a consequence of properly designed cascade and/or feedback compensators. The design method presented in this chapter is based upon achieving a desired or model control ratio $M_T(s)$, i.e., this is a pole-zero placement method. Section 12.2 describes the modeling of a desired closed-loop transfer function. This chapter introduces the concept of feeding back all the system states to achieve the desired improvement in system performance. The design concepts developed with conventional control-theory are also used to design state-variable feedback. The state-variable-feedback concept requires that all states be accessible in a physical system, but for most systems this requirement is not met; i.e., some of the states are inaccessible. Techniques for handling systems with inaccessible states are presented in Chap. 14. The state-variable-feedback design method presented in this chapter is based upon achieving a desired closed-loop state equation or a desired control ratio for a *single-input single-output* (SISO) system. Synthesis methods using matrix methods for *multiple-input multiple-output* (MIMO) systems are available in the literature [1–14]. Following the convention often used with the state-variable representation of systems, the output and actuating signals are represented in this chapter and Chap. 14 by $y(t)$ and $u(t)$ instead of $c(t)$ and $e(t)$, respectively.

The principal objective of the state-variable method in this chapter is the assignment of the closed-loop eigenvalues for SISO systems. This design process can be simplified for higher-order systems by first transforming the state equation into control canonical form. The procedure is described in Sec. 13.4. The analysis and synthesis involved in the problems associated with this chapter, like the previous chapters, relies on the use of the CAD packages like MATLAB, TOTAL, or ICECAP/PC (see Appendix B). The utilization of such programs expedites the design process for achieving a solution and eliminates the tedious hand calculations.

13.2 CONTROLLABILITY AND OBSERVABILITY [5–9]

Before introducing the state feedback design methods, the necessary conditions and tests for controllability and observability are presented. Recall that the state and output equations representing a system, as defined in Sec. 2.3, have the form

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \tag{13.1}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u} \tag{13.2}$$

where **x**, **y**, and **u** are vectors and **A**, **B**, **C**, and **D** are matrices having the dimensions:

$$\mathbf{x}: n \times 1 \quad \mathbf{y}: \ell \times 1 \quad \mathbf{u}: m \times 1 \quad \mathbf{A}: n \times n \quad \mathbf{B}: n \times m \quad \mathbf{C}: \ell \times n \quad \mathbf{D}: \ell \times m$$

An important objective of state variable control is the design of systems that have an optimum performance. The optimal control [1,2,8,9] is based on the optimization of some specific performance criterion. The achievement of such optimal linear control systems is governed by the controllability and observability properties of the system. For example, in order to be able to relocate or reassign the open-loop plant poles to more desirable closed-loop locations in the s-plane, it is necessary that the plant satisfy the controllability property. Further, these properties establish the conditions for complete equivalence between the state-variable and transfer-function representations. A study of these properties, presented below, provides a basis for consideration of the assignment of the closed-loop eigenstructure (eigenvalue and eigenvector spectra).

CONTROLLABILITY. A system is said to be completely *state-controllable* if, for any initial time t_0, each initial state $\mathbf{x}(t_0)$ can be transferred to any final state $\mathbf{x}(t_f)$ in a finite time, $t_f > t_0$, by means of an unconstrained control input vector $\mathbf{u}(t)$. An unconstrained control vector has no limit on the amplitudes of $\mathbf{u}(t)$. This definition implies that $\mathbf{u}(t)$ is able to affect each state variable in the state equation of Eq. (13.1). Controllability also means that each mode $\mathbf{v}_i e^{\lambda_i t}$ in $\mathbf{x}(t)$ is directly affected by the input $\mathbf{u}(t)$. The solution of the state equation (see Sec. 3.14) is

$$\mathbf{x}(t) = \mathbf{\Phi}(t - t_0)\mathbf{x}(t_0) + \int_{t_0}^{t} (t - \tau)\mathbf{B}\mathbf{u}(\tau)\, d\tau \tag{13.3}$$

OBSERVABILITY. A system is said to be completely *observable* if every initial state $\mathbf{x}(t_0)$ can be exactly determined from the measurements of the output $\mathbf{y}(t)$ over the finite interval of time $t_0 \le t \le t_f$. This definition implies that every state of $\mathbf{x}(t)$ affects the output $\mathbf{y}(t)$:

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) = \mathbf{C}\mathbf{\Phi}(t - t_0)\mathbf{x}(t_0) + \mathbf{C}\int_{t_0}^{t} \mathbf{\Phi}(t - \tau)\mathbf{B}\mathbf{u}(\tau)\, d\tau \tag{13.4}$$

where the initial state $\mathbf{x}(t_0)$ is the result of control inputs prior to t_0.

The concepts of controllability and observability [7] can be illustrated graphically by the block diagram of Fig. 13.1. By the proper selection of the state variable it is possible to divide a system into the four subdivisions shown in Fig. 13.1:

S_{co} = completely controllable and completely observable subsystem

S_o = completely observable but uncontrollable subsystem

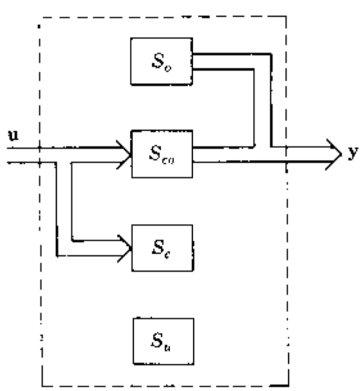


FIGURE 13.1 The four possible subdivisions of a system.

$S_c =$ completely controllable but unobservable subsystem

$S_u =$ uncontrollable and unobservable subsystem

This figure readily reveals that only the controllable and observable subsystem S_{co} satisfies the definition of a transfer-function matrix

$$\mathbf{G}(s)\mathbf{U}(s) = \mathbf{Y}(s) \tag{13.5}$$

If the entire system is completely controllable and completely observable, then the state-variable and transfer-function matrix representations of a system are equivalent and accurately represent the system. In that case the resulting transfer function does carry all the information characterizing the dynamic performance of the system. As mentioned in Chap. 2, there is no unique method of selecting the state variables to be used in representing a system, but the transfer-function matrix is completely and uniquely specified once the state-variable representation of the system is known. In the transfer-function approach, the state vector is suppressed; the transfer-function method is concerned only with the system's input-output characteristics. The state-variable method also includes a description of the system's internal behavior.

The determination of controllability and observability of a system by subdividing it into its four possible subdivisions is not always easy. A simple method for determining these system characteristics is developed by using Eq. (13.3) with $t_0 = 0$ and specifying that the final state vector $\mathbf{x}(t_f) = \mathbf{0}$. This yields

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

or

$$\mathbf{x}(0) = -\int_0^{t_f} e^{-\mathbf{A}\tau}\mathbf{B}\mathbf{u}(\tau)\,d\tau \tag{13.6}$$

The Cayley-Hamilton method presented in Sec. 3.14 shows that exp $(-\mathbf{A}\tau)$ can be expressed as a polynomial of order $n-1$. Therefore,

$$e^{-\mathbf{A}\tau} = \sum_{k=0}^{n-1} \alpha_k(\tau)\mathbf{A}^k \tag{13.7}$$

Inserting this value into Eq. (13.6) yields

$$\mathbf{x}(0) = -\sum_{k=0}^{n-1} \mathbf{A}^k\mathbf{B}\int_0^{tf} \alpha_k(\tau)\mathbf{u}(\tau)\,d\tau \tag{13.8}$$

With the input $\mathbf{u}(t)$ of dimension r, the integral in Eq. (13.8) can be evaluated, and the result is

$$\int_0^{t_f} \alpha_k(\tau)\mathbf{u}(\tau)\, d\tau = \beta_k$$

where β_k is a vector of dimension $n \times 1$. Equation (13.8) can now be expressed as

$$\mathbf{x}(0) = -\sum_{k=0}^{n-1} \mathbf{A}^k \mathbf{B} \beta_k = -\left[\mathbf{B} \;\vdots\; \mathbf{AB} \;\vdots\; \cdots \;\vdots\; \mathbf{A}^{n-1}\mathbf{B}\right] \begin{bmatrix} \beta_0 \\ --- \\ \beta_1 \\ --- \\ \vdots \\ --- \\ \beta_{n-1} \end{bmatrix} \tag{13.9}$$

According to the definition of complete controllability, each mode $\mathbf{v}_i e^{\lambda_i t}$ must be directly affected by the input $\mathbf{u}(t)$. This requires that the *controllability* matrix $\mathbf{M}_c$ have the property

$$\text{Rank}\, \mathbf{M}_c = \text{Rank}\left[\mathbf{B} \;\vdots\; \mathbf{AB} \;\vdots\; \cdots \;\vdots \mathbf{A}^{n-1}\mathbf{B}\right] = n \tag{13.10}$$

For a single-input system, the matrix $\mathbf{B}$ is the vector $\mathbf{b}$, and the matrix of Eq. (13.10) has dimension $n \times n$.

When the state equation of a SISO system is in control canonical form; that is, the plant matrix $\mathbf{A}$ is in companion form and the control vector $\mathbf{b} = [0 \quad \cdots \quad 0 \quad K_G]^T$, then the controllability matrix $\mathbf{M}_c$ has full rank n (see Sec. 5.13). Thus, the system is completely controllable and it is not necessary to perform a controllability test.

When the matrix pair $(\mathbf{A}, \mathbf{B})$ is not completely controllable, the uncontrollable modes can be determined by transforming the state equation so that the plant matrix is in diagonal form (see Sec. 5.10). This approach is described below. A simple method for determining controllability when the system has nonrepeated eigenvalues is to use the transformation $\mathbf{x}=\mathbf{Tz}$ to convert the $\mathbf{A}$ matrix into diagonal (canonical) form, as described in Sec. 5.10, so that the modes are decoupled. The resulting state equation is

$$\dot{\mathbf{z}} = \mathbf{T}^{-1}\mathbf{ATz} + \mathbf{T}^{-1}\mathbf{Bu} = \mathbf{\Lambda z} + \mathbf{B}'\mathbf{u} \tag{13.11}$$

Each transformed state z_i represents a mode and can be directly affected by the input $\mathbf{u}(t)$ only if $\mathbf{B}' = \mathbf{T}^{-1}\mathbf{B}$ has no zero row. Therefore, for nonrepeated eigenvalues, *a system is completely controllable if* $\mathbf{B}'$ has no zero row. For repeated eigenvalues see Refs. 5 and 7. $\mathbf{B}'$ is called the *mode-controllability* matrix. A zero row of $\mathbf{B}'$ indicates that the corresponding mode is uncontrollable. The eigenvalues associated with uncontrollable modes are called *input-decoupling zeros.*

In a similar manner, the condition for observability is derived from the homogeneous system equations

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \tag{13.12}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} \tag{13.13}$$

When Eq. (13.7) is used, the output vector $\mathbf{y}(t)$ can be expressed as

$$\mathbf{y}(t) = \mathbf{C}e^{\mathbf{A}t}\mathbf{x}(0) + \sum_{k=0}^{n-1} \alpha_k(t)\mathbf{C}\mathbf{A}^k\mathbf{x}(0) \tag{13.14}$$

For observability the output $\mathbf{y}(t)$ must be influenced by each state $\mathbf{x}_i$. This imposes restrictions on $\mathbf{C}\mathbf{A}^k$. It can be shown that the system is completely observable if the following *observability* matrix $\mathbf{M}_o$ has the property

$$\text{Rank } \mathbf{M}_o = \text{Rank}\left[\mathbf{C}^T \;\vdots\; \mathbf{A}^T\mathbf{C}^T \;\vdots\; (\mathbf{A}^T)^2\mathbf{C}^T \;\vdots\; \cdots \;\vdots (\mathbf{A}^T)^{(n-1)}\mathbf{C}^T)\right] = n \tag{13.15}$$

For a single-output system, the matrix $\mathbf{C}^T$ is a column matrix that can be represented by $\mathbf{c}$, and the matrix of Eq. (13.15) has dimension $n \times n$. When the matrix pair (**A**,**C**) is not completely observable, the unobservable modes can be determined by transforming the state and output equations so that the plant matrix is in diagonal form. This approach is described below.

A simple method of determining observability that can be used when the system has nonrepeated eigenvalues is to use the transformation $\mathbf{x} = \mathbf{T}\mathbf{z}$ in order to put the equations in canonical form; that is, the plant matrix is transformed to the diagonal form $\mathbf{\Lambda}$. The output equation is then

$$\mathbf{y} = \mathbf{C}\mathbf{x} = \mathbf{C}\mathbf{T}\mathbf{z} = \mathbf{C}'\mathbf{z} \tag{13.16}$$

If a column of $\mathbf{C}'$ has all zeros, then one mode is not coupled to *any* of the outputs and the system is unobservable. The *condition for observability, for the case of nonrepeated eigenvalues, is that* $\mathbf{C}'$ *have no zero columns.* $\mathbf{C}'$ is called the *mode-observability* matrix. The eigenvalues associated with an unobservable mode are called *output-decoupling zeros.*

Example. Figure 13.2 shows the block-diagram representation of a system whose state and output equations are

$$\dot{\mathbf{x}} = \begin{bmatrix} -2 & 0 \\ -1 & -1 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1 \\ 1 \end{bmatrix}u \tag{13.17}$$

$$\mathbf{y} = \begin{bmatrix} 0 & 1 \end{bmatrix}\mathbf{x} \tag{13.18}$$

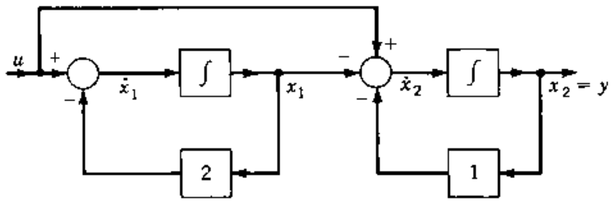


FIGURE 13.2 Simulation diagram for the system of Eq. (13.17).

Determine: (*a*) the eigenvalues from the state equation, (*b*) the transfer function $Y(s)/U(s)$, (*c*) whether the system is controllable and/or observable, (*d*) the transformation matrix T that transforms the general state-variable representation of Eqs. (13.17) and (13.18) into the canonical form representation, and (*e*) draw a simulation diagram based on the canonical representation.

(*a*) $$|s\mathbf{I} - \mathbf{A}| = \begin{vmatrix} s+2 & 0 \\ 1 & s+1 \end{vmatrix} = (s+2)(s+1)$$

The eigenvalues are $\lambda_1 = -2$ and $\lambda_2 = -1$. Thus,

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

(*b*) $$\mathbf{\Phi}(s) = [s\mathbf{I} - \mathbf{A}]^{-1} = \frac{\begin{bmatrix} s+1 & 0 \\ -1 & s+2 \end{bmatrix}}{(s+1)(s+2)}$$

$$\mathbf{G}(s) = \mathbf{c}^T\mathbf{\Phi}(s)\mathbf{b} = \left(\begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} s+1 & 0 \\ -1 & s+2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) \frac{1}{(s+1)(s+2)}$$

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

Note that there is a cancellation of the factor $s+1$, and only the mode with $\lambda_1 = -2$ is both controllable and observable. When a mode is either uncontrollable or unobservable, it does not appear in the transfer function.

(*c*) $$\text{Rank}\left[\mathbf{b} \;\vdots\; \mathbf{Ab}\right] = \text{Rank}\begin{bmatrix} 1 & -2 \\ 1 & -2 \end{bmatrix} = 1$$

The controllability matrix $\mathbf{M}_c = [\mathbf{b} \quad \mathbf{Ab}]$ is singular and has a rank deficiency of one; thus the system is *not* completely controllable and has one uncontrollable mode.

$$\text{Rank}\left[\mathbf{c} \;\vdots\; \mathbf{A}^T\mathbf{c}\right] = \text{Rank}\begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix} = 2$$

The observability matrix $\left[\mathbf{c} \;\vdots\; \mathbf{A}^T\mathbf{c}\right]$ is nonsingular; thus the system is completely observable. Note that the mode with $\lambda_2 = -1$, which is canceled in the determination of the transfer function, is observable but not controllable.
(*d*) Using Eq. (5.91) of Sec. 5.10, with $\mathbf{x} = \mathbf{Tz}$, gives

$$\mathbf{AT} = \mathbf{T\Lambda}$$

$$\begin{bmatrix} -2 & 0 \\ -1 & -1 \end{bmatrix}\begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}\begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix}$$

$$\begin{bmatrix} -2v_{11} & -2v_{12} \\ -v_{11} - v_{21} & -v_{12} - v_{22} \end{bmatrix} = \begin{bmatrix} -2v_{11} & -v_{12} \\ -2v_{21} & -v_{22} \end{bmatrix}$$

Equating elements on both sides of this equation yields

$$\begin{aligned} -2v_{11} &= -2v_{11} & \qquad -2v_{12} &= -v_{12} \\ -v_{11} - v_{21} &= -2v_{21} & \qquad -v_{12} - v_{22} &= -v_{22} \end{aligned}$$

Because there are only two independent equations, assume that $v_{11} = v_{22} = 1$. Then $v_{21} = 1$ and $v_{12} = 0$. Utilizing these values yields

$$\mathbf{b}' = \mathbf{T}^{-1}\mathbf{b} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\mathbf{c}'^T = \mathbf{c}^T\mathbf{T} = \begin{bmatrix} 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \end{bmatrix}$$

The canonical equations, for this system are therefore

$$\dot{\mathbf{z}} = \begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix}\mathbf{z} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}u$$

$$\mathbf{y} = \begin{bmatrix} 1 & 1 \end{bmatrix}\mathbf{z}$$

Because $\mathbf{b}'$ contains a zero row, the system is uncontrollable and $\lambda = -1$ is an input-decoupling zero. Because $\mathbf{c}'$ contains all nonzero elements, the system is completely observable. These results agree with the results obtained by the method used in part (*c*).
(*e*) The individual state and output equations are

$$\dot{z}_1 = -2z_1 + u \tag{13.19}$$

$$\dot{z}_2 = -z_2 \tag{13.20}$$

$$y = z_1 + z_2 \tag{13.21}$$

Equation (13.19) satisfies the requirement of controllability; for any t_0 the initial state $z_1(t_0)$ can be transferred to any final state $z_1(t_f)$ in a finite time $t_f \geq t_0$ by means of the unconstrained control vector u. Because u does not appear

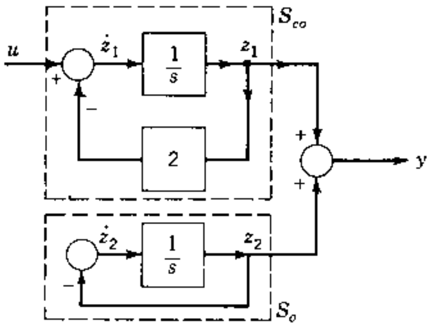


FIGURE 13.3 Simulation diagram for Eqs. (13.17) and (13.18).

in Eq. (13.20), it is impossible to control the state z_2; thus $\lambda_2 = -1$ is an input-decoupling zero and the system is uncontrollable. Equation (13.21) satisfies the definition of observability, i.e., both states appear in y and, therefore, $z_1(t_0)$ and $z_2(t_0)$ can be determined from the measurements of the output $y(t)$ over the finite interval of time $t_0 \le t \le t_f$.

The simulation diagram of Fig. 13.3 is synthesized from Eqs. (13.19) to (13.21). The advantage of the canonical form is that the modes $\lambda_1 = -2$ and $\lambda_2 = -1$ are completely decoupled. This feature permits the determination of the existence of some or all of the four subdivisions of Fig. 13.1 and the application of the definitions of controllability and observability.

Example: MATLAB Controllability and Observability

```
A = [-2  0; -1  -1];
B = [1;  1];
C = [0  1];

%
% (a) Determine the eigenvalues
eig(A)

ans =
    -1
    -2

%
% (b) Determine the transfer function
sys = ss(A, B, C, 0);
sys = tf(sys)
```

www.4electron.com

```
Transfer function:
```

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

```
sys = minreal (sys)

Transfer function :
```

$$\frac{1}{s+2}$$

```
%
% (c) Determine controllability and observability
%
% The “ctrb” function computes the controllability matrix
co = ctrb(A, B)
co =

    1  -2
    1  -2

rank(co)
ans =
     1
% The number of uncontrollable states
uncon = length(A) - rank (co)
uncon =
        1
%
% The “obsv” function computes the observability matrix
ob = obsv(A, C)
ob =
     0   1
    -1  -1
rank (ob)
ans =
    2
%
% (d) MATLAB’s version shows one controllable state [Ab, Bb, Cb,
Db, T] = CANON (A, B, C,0, ‘modal’)
Ab =
    -1   0
     0  -2
Bb =
         0
    1.4142
Cb =
    1.0000   0.7071
```

www.4electron.com

```
Db =
    0
T =
    -1.0000   1.0000
     1.4142        0
T*A* inv (T)
ans =
    -1   0
     0  -2
```

13.3 STATE FEEDBACK FOR SISO SYSTEMS

This section covers the design of a single-input single-output (SISO) system in which the set of closed-loop eigenvalues is assigned by state feedback. The plant state and output equations have the form

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \tag{13.22}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} \tag{13.23}$$

The state-feedback control law (see Fig. 13.4) is

$$\mathbf{u} = \mathbf{r} + \mathbf{K}\mathbf{x} \tag{13.24}$$

where $\mathbf{r}$ is the input vector and $\mathbf{K}$ is the state-feedback matrix having dimension $m \times n$. The closed-loop state equation obtained by combining Eqs. (13.22) and (13.24) yields

$$\dot{\mathbf{x}} = (\mathbf{A} + \mathbf{B}\mathbf{K})\mathbf{x} + \mathbf{B}\mathbf{r} \tag{13.25}$$

Thus the closed-loop system matrix produced by state feedback is

$$\mathbf{A}_{c1} = \mathbf{A} + \mathbf{B}\mathbf{K} \tag{13.26}$$

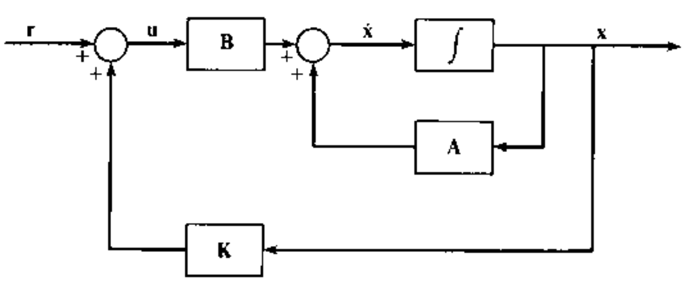


FIGURE 13.4 State feedback control system.

which results in the closed-loop characteristic equation

$$Q(\lambda) = \left|\lambda\mathbf{I} - \mathbf{A}_{c1}\right| = |\lambda\mathbf{I} - \mathbf{A} - \mathbf{BK}| = 0 \tag{13.27}$$

Equation (13.27) reveals that the closed-loop eigenvalues can be assigned by the proper selection of the state-feedback matrix **K**. A necessary and sufficient condition for the selection of **K** is that the plant be completely controllable [9] (see Sec. 13.2).

Example 1. A second-order system is represented by the state and output equations

$$\dot{\mathbf{x}} = \begin{bmatrix} -1 & 3 \\ 0 & -2 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1 \\ 1 \end{bmatrix}\mathbf{u} \tag{13.28}$$

$$\mathbf{y} = \begin{bmatrix} 1 & 0 \end{bmatrix}\mathbf{x} = \mathbf{x}_1 \tag{13.29}$$

The open-loop eigenvalue spectrum is $\sigma(\mathbf{A}) = \{-1, -2\}$. The system is completely controllable since, from Eq. (13.10),

$$\text{Rank}\,\mathbf{M}_c = \text{Rank}\begin{bmatrix} 1 & 2 \\ 1 & -2 \end{bmatrix} = 2 \tag{13.30}$$

The closed-loop eigenvalues are determined from Eq. (13.27) with $\mathbf{K} = \begin{bmatrix} k_1 & k_2 \end{bmatrix}$:

$$\begin{aligned} |\lambda\mathbf{I} - \mathbf{A} - \mathbf{BK}| &= \left|\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} - \begin{bmatrix} -1 & 3 \\ 0 & -2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix}\begin{bmatrix} k_1 & k_2 \end{bmatrix}\right| \\ &= \lambda^2 + (3 - k_1 - k_2)\lambda + (2 - 5k_1 - k_2) \end{aligned} \tag{13.31}$$

Assigning the closed-loop eigenvalues $\lambda_1 = -5$ and $\lambda_2 = -6$ requires the characteristic polynomial to be $\lambda^2 + 11\lambda + 30$. Equating the coefficients of this desired characteristic equation with those of Eq. (13.31) results in the values $k_1 = -5$ and $k_2 = -3$, or $\mathbf{K} = [-5 \quad -3]$. The resulting closed-loop state equation obtained by using Eq. (13.25) is

$$\dot{\mathbf{x}} = \begin{bmatrix} -6 & 0 \\ -5 & -5 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1 \\ 1 \end{bmatrix}\mathbf{r} \tag{13.32}$$

This state equation has the required eigenvalue spectrum $\sigma(\mathbf{A}_{c1}) = \{-5, -6\}$.

The following example illustrates the requirement of controllability for closed-loop eigenvalue assignment by state feedback.

Example 2. A cascade gain K is inserted in the plant of the example in Sec. 13.2, Eq. (13.17) so that its transfer function is

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

It is shown in that example that this plant is not controllable. Figure 13.5 shows a state-feedback system for this plant. The closed-loop control ratio can be obtained by use of the Mason gain rule [Eq. (5.53)], and is given by

$$\frac{Y(s)}{R(s)} = \frac{K(s+1)}{s^2 + (3 + Kk_1 + Kk_2)s + (2 + Kk_2 + Kk_1)} \tag{13.33}$$

In order for the system to follow a step input with no steady-state error, the requirement obtained by applying the final value theorem is that $K = 2 + Kk_2 + Kk_1$, or

$$K = \frac{2}{1 - k_1 - k_2} \tag{13.34}$$

Because this system is uncontrollable, as determined in Sec. 13.2, the pole located at $s_2 = -1$ cannot be altered by state feedback. The characteristic equation of Eq. (13.33) illustrates this property because it can be factored to the form

$$(s+1)(s + 2 + Kk_2 + Kk_1) = 0 \tag{13.35}$$

Therefore one eigenvalue, $s_2 = -2$, is altered by state feedback to the value $s_1 = -(2 + Kk_2 + Kk_1) = -K$, where K can be assigned as desired. Because the uncontrollable pole cancels the zero of Eq. (13.33), the control ratio becomes

$$\frac{Y(s)}{R(s)} = \frac{K(s+1)}{(s+1)(s+K)} = \frac{K}{s+K} \tag{13.36}$$

This example demonstrates that the uncontrollable pole $s_1 = -1$ cannot be changed by state feedback.

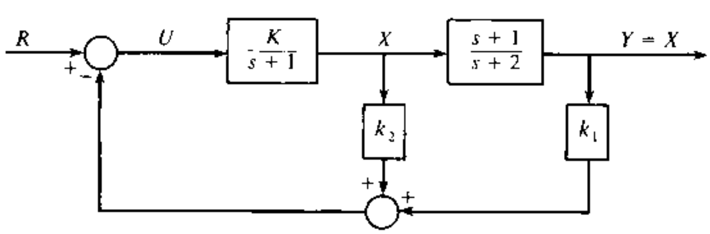


FIGURE 13.5 An uncontrollable state-feedback system.

When deriving the transfer function from the system differential equations, cancellation of a pole by a zero is not permitted in developing the state representation. Cancellation of a pole by a zero in the transfer function may hide some of the dynamics of the entire system. This eliminates information pertaining to S_o, S_c, and S_u; that is, the observability and controllability of the system are determined from the state equations, as illustrated in these examples.

13.4 STATE-FEEDBACK DESIGN FOR SISO SYSTEMS USING THE CONTROL CANONICAL (PHASE-VARIABLE) FORM

The examples of Sec. 13.3 are of second order, and therefore the algebraic solution for the **K** matrix is computationally simple and direct. For higher-order systems, the computations are simplified by first transforming the state equation into the control canonical form $\dot{\mathbf{x}}_c = \mathbf{A}_c\mathbf{x}_c + \mathbf{b}_c\mathbf{u}$. For this representation of the state equation, the plant matrix $\mathbf{A}_c$ is in companion form, and the control matrix b_c contains all zeros except for the unit element in the last row. This form leads directly to the determination of the required state-feedback matrix that is unique. The procedure is presented in this section.

The model of an open-loop SISO system can be represented by a physical state variable and an output equation

$$\dot{\mathbf{x}} = \mathbf{A}_p\mathbf{x}_p + \mathbf{b}_p\mathbf{u} \tag{13.37}$$

$$\mathbf{y} = \mathbf{c}_p^T\mathbf{x}_p \tag{13.38}$$

The actuating signal **u** is formed in accordance with the feedback control law

$$\mathbf{u} = \mathbf{r} + \mathbf{k}_p^T\mathbf{x}_p \tag{13.39}$$

where **r** is the reference input and

$$\mathbf{k}_p^T = \begin{bmatrix} k_{p1} & k_{p2} & \cdots & k_{pn} \end{bmatrix} \tag{13.40}$$

The closed-loop state equation obtained from Eqs. (13.37) and (13.39) is

$$\dot{\mathbf{x}}_p = \left[\mathbf{A}_p + \mathbf{b}_p\mathbf{k}_p^T\right]\mathbf{x}_p + \mathbf{b}_p\mathbf{r} = \mathbf{A}_{p_{\mathrm{cl}}}\mathbf{x}_p + \mathbf{b}_p\mathbf{r} \tag{13.41}$$

The design procedure that assigns the desired set of closed-loop eigenvalues, $\sigma(\mathbf{A}_{p_{\mathrm{cl}}}) = \{\lambda_i\}$, $i = 1, 2, \ldots, n$, is accomplished as follows:

1. Use the method of Sec. 5.13 to obtain the matrix **T**, which transforms the state equation (13.37), by the change of state variables $\mathbf{x}_p = \mathbf{T}\mathbf{x}_c$, into the *control canonical form* $\dot{\mathbf{x}}_c = \mathbf{A}_c\mathbf{x}_c + \mathbf{b}_c\mathbf{u}$, where A_c is in companion form and $\mathrm{b}_c = [0 \quad \cdots \quad 0 \quad 1]^T$.

2. Select the set of desired closed-loop eigenvalues and obtain the corresponding closed-loop characteristic polynomial

$$\begin{aligned} Q_{\mathrm{cl}}(\lambda) &= (\lambda - \lambda_1)(\lambda - \lambda_2) \quad \cdots \quad (\lambda - \lambda_n) \\ &= \lambda^n + \alpha_{n-1}\lambda^{n-1} + \quad \cdots \quad + \alpha_1\lambda + \alpha_0 \end{aligned} \tag{13.42}$$

3. Write the companion form of the desired closed-loop plant matrix

$$\mathbf{A}_{c_{\mathrm{cl}}} = \begin{bmatrix} 0 & 1 & & & & \\ & 0 & 1 & & & \\ & & 0 & \ddots & & \\ & & & \ddots & 1 & \\ & & & & 0 & 1 \\ -\alpha_0 & -\alpha_1 & & \cdots & -\alpha_{n-2} & -\alpha_{n-1} \end{bmatrix} \tag{13.43}$$

and $\mathbf{b}_c = [0 \quad \cdots \quad 0 \quad 1]$.

4. Use the feedback matrix

$$\mathbf{k}_c^T = \begin{bmatrix} k_{c1} & k_{c2} & \cdots & k_{cn} \end{bmatrix} \tag{13.44}$$

to form the closed-loop matrix

$$\mathbf{A}_{C_{c\mathrm{l}}} = \mathbf{A}_c + \mathbf{b}_c\mathbf{u} = \mathbf{A}_c + \mathbf{b}_c\mathbf{k}_c^T$$

$$= \begin{bmatrix} 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \hdashline 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -a_0 + k_{c1} & -a_1 + k_{c2} & \cdots & -a_{n-2}k_{c(n-1)} & -a_{n-1} + k_{cn} \end{bmatrix} \tag{13.45}$$

5. Equate the matrices of Eqs. (13.43) and (13.45), and solve for the elements of the matrix $\mathbf{k}_c^T$.
6. Compute the required physical feedback matrix using

$$\mathbf{k}_p^T = \mathbf{k}_c^T T^{-1} \tag{13.46}$$

Example. Using the physical system shown by Eq. (13.47), assign the set of eigenvalues $\{-1+j1, -1-j1, -4\}$.

$$\mathbf{A}_p = \begin{bmatrix} 1 & 6 & -3 \\ -1 & -1 & 1 \\ -2 & 2 & 0 \end{bmatrix} \qquad \mathbf{b}_p = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \tag{13.47}$$

This open-loop system is unstable because $\sigma(\mathbf{A}) = \{1, 1, -2\}$.

Step 1. As derived in the example of Sec. 5.13,

$$\mathbf{A}_c = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & 3 & 0 \end{bmatrix} \qquad \mathbf{b}_c = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{13.48}$$

and, for $x_p = \mathbf{T}x_c$,

$$\mathbf{T} = \begin{bmatrix} -5 & 4 & 1 \\ -6 & -1 & 1 \\ -13 & 0 & 1 \end{bmatrix} \tag{13.49}$$

Step 2. The required closed-loop characteristics polynomial is

$$Q_{c1}(\lambda) = (\lambda + 1 - j1)(\lambda + 1 + j1)(\lambda + 4) = \lambda^3 + 6\lambda^2 + 10\lambda + 8 \tag{13.50}$$

Step 3. The closed-loop plant matrix in companion form is

$$\mathbf{A}_{c_{c1}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -8 & -10 & -6 \end{bmatrix} \tag{13.51}$$

Step 4

$$\mathbf{A}_c + \mathbf{b}_c\mathbf{k}_c^T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 + k_{c1} & 3 + k_{c2} & k_{c3} \end{bmatrix} \tag{13.52}$$

Step 5. Equating the coefficients of $\mathbf{A}_{c_{c1}}$ and $\mathbf{A}_c + \mathbf{b}_c\mathbf{k}_c^T$ yields

$$\mathbf{k}_c^T = \begin{bmatrix} -6 & -13 & -6 \end{bmatrix} \tag{13.53}$$

Step 6. Using Eq. (13.46), the feedback matrix required for the physical system is

$$\mathbf{k}_p^T = \mathbf{k}_c^T\mathbf{T}^{-1} = 1/36\begin{bmatrix} -175 & -232 & 191 \end{bmatrix} \tag{13.54}$$

The reader should verify that the physical closed-loop matrix

$$\mathbf{A}_{p_{c1}} = \mathbf{A}_p + \mathbf{b}_p\mathbf{k}_p^T = \frac{1}{36}\begin{bmatrix} -139 & -16 & 83 \\ -211 & -268 & 227 \\ -247 & -160 & 191 \end{bmatrix} \tag{13.55}$$

has the required set of eigenvalues $\sigma(\mathbf{A}_{p_{c1}}) = \{-1+j1, -1-j1, -4\}$.

This design method provides a very straightforward procedure for computing the state feedback matrix $\mathbf{k}_p^T$. MATLAB contains procedures for assisting with this design.

13.5 STATE-VARIABLE FEEDBACK [10] (PHYSICAL VARIABLES)

The open-loop position-control system of Fig. 13.6 is used to illustrate the effects of state-variable feedback. The dynamic equations for this basic system are

$$e_a = Ae \qquad A > 0 \tag{13.56}$$

$$e_a - e_m = (R_m + L_m D)i_m \tag{13.57}$$

$$e_m = K_b \omega_m \tag{13.58}$$

$$T = K_T i_m = JD\omega_m + B\omega_m \tag{13.59}$$

$$\omega_m = D\theta_m \tag{13.60}$$

In order to obtain the maximum number of accessible (measurable) states, *physical variables are selected as the state variables.* The two variables i_m and ω_m and the desired output quantity θ_m are identified as the three state variables: $x_1 = \theta_m = y$, $x_2 = \omega_m = \dot{x}_1$, and $x_3 = i_m$, and the input $u = e$. Figure 13.6 shows that all three state variables are accessible for measurement. The sensors are selected to produce voltages that are proportional to the state variables and to have effect on the measured variables.

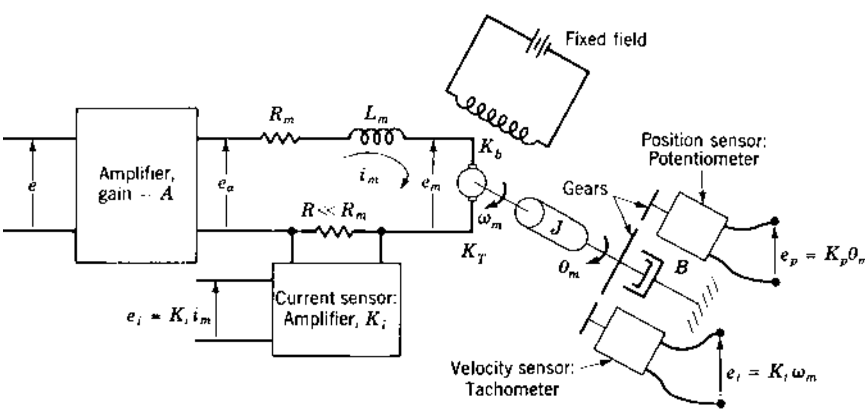


FIGURE 13.6 Open-loop position-control system and measurement of states.

Figure 13.7*a* is the block diagram of a closed-loop position-control system utilizing the system of Fig. 13.6. Each of the states is fed back through amplifiers whose gain values k_1, k_2, and k_3 are called *feedback coefficients.* These feedback coefficients include the sensor constants and any additional gain required by the system design to yield the required values of k_1, k_2, and k_3. The summation of the three feedback quantities, where $\mathbf{k}^T = \begin{bmatrix} k_1 & k_2 & k_3 \end{bmatrix}$ is the feedback matrix, is

$$\mathbf{k}^T\mathbf{X} = k_1X_1 + k_2X_2 + k_3X_3 \tag{13.61}$$

Using block-diagram manipulation techniques, Fig. 13.7*a* is simplified to the block diagram of Fig. 13.7*b*. A further reduction is achieved, as shown in Fig. 13.7*c*. Note that the forward transfer function $G(s) = X_1(s)/U(s)$ shown in this figure is identical with that obtained from Fig. 13.7*a* with zero-state feedback coefficients. Thus, state-variable feedback is *equivalent* to inserting a feedback compensator $H_{eq}(s)$ in the feedback loop of a conventional feedback control system, as shown in Fig. 13.7*d*. The reduction of 13.7*a* to that of Fig. 13.7*c* or Fig. 13.7*d* is the *H-equivalent reduction.* The system control ratio is therefore

$$\frac{Y(s)}{R(s)} = \frac{G(s)}{1 + G(s)H_{eq}(s)} \tag{13.62}$$

For Figs. 13.7*c* and 13.7*d* the following equations are obtained:

$$G(s)H_{eq}(s) = \frac{100A[k_3s^2 + (2k_2 + k_3)s + 2k_1]}{s(s^2 + 6s + 25)} \tag{13.63}$$

$$\frac{Y(s)}{R(s)} = \frac{200A}{s^3 + (6 + 100k_3A)s^2 + (25 + 200k_2A + 100k_3A)s + 200k_1A} \tag{13.64}$$

The equivalent feedback transfer function $H_{eq}(s)$ must be designed so that the system's desired figures of merit are satisfied. The following observations are noted from Eqs. (13.63) and (13.64) for this system in which $G(s)$ has no zeros; i.e., it is an all-pole plant.

1. In order to achieve zero steady-state error for a step input, $R(s) = R_0/s$, Eq. (13.64) must yield

   $$y(t)_{ss} = \lim_{s\to 0} sY(s) = \frac{R_0}{k_1} = R_0$$

   This requires that $k_1 = 1$.

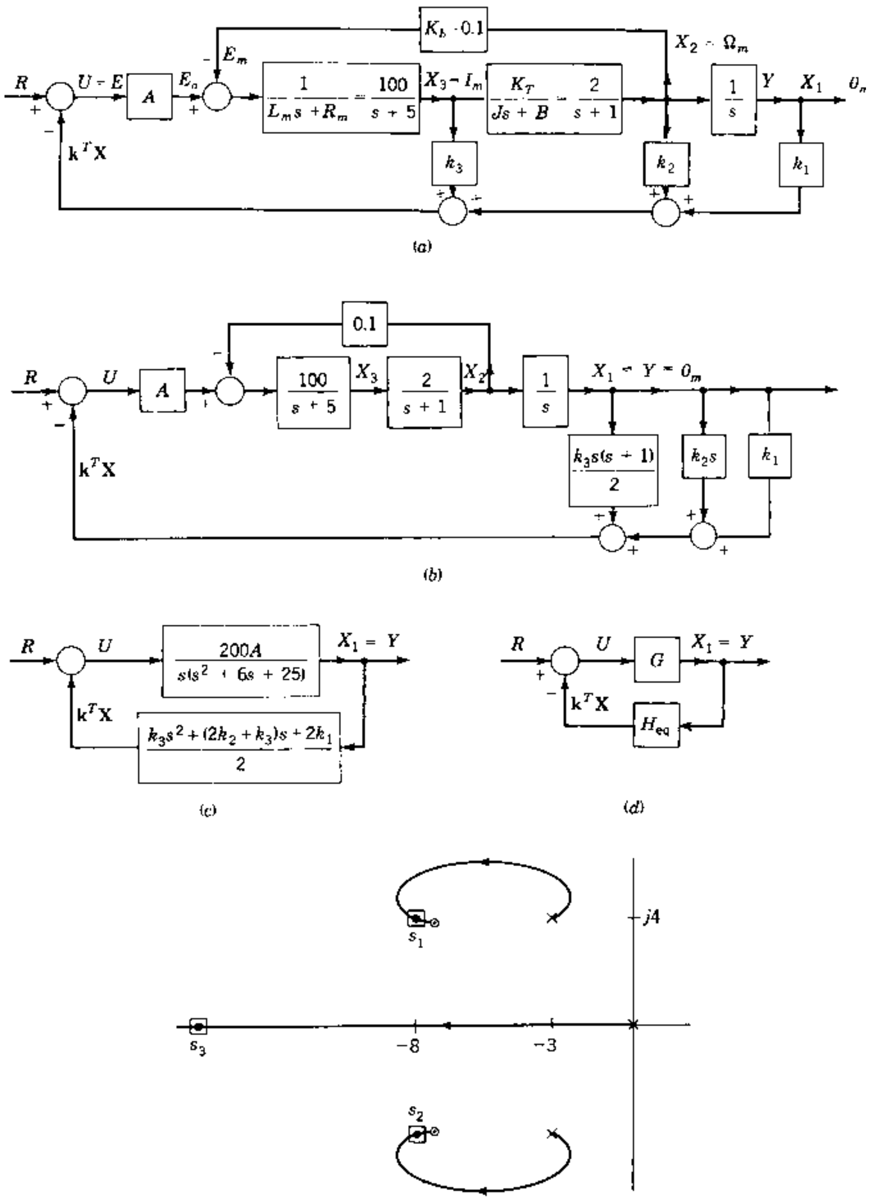


FIGURE 13.7 A position-control system with feedback, its H_{eq} representation, and the corresponding root locus.

2. The numerator of $H_{eq}(s)$ is of second order; i.e., the denominator of $G(s)$ is of degree $n = 3$ and the numerator of $H_{eq}(s)$ is of degree $n - 1 = 2$.
3. The poles of $G(s)H_{eq}(s)$ are the poles of $G(s)$.
4. The zeros of $G(s)H_{eq}(s)$ are the zeros of $H_{eq}(s)$.
5. The use of state feedback produces additional zeros in the open-loop transfer function $G(s)H_{eq}(s)$ without adding poles. This is in contrast to cascade compensation using passive networks. The zeros of $H_{eq}(s)$ are to be located to yield the desired response. It is known that the addition of zeros to the open-loop transfer function moves the root locus to the left; i.e., they make the system more stable and improve the time-response characteristics.
6. Since the root-locus plot of $G(s)H_{eq}(s) = -1$ has one asymptote, $\gamma = -180^\circ$ for $k_3 > 0$, it is possible to pick the locations of the zeros of $H_{eq}(s)$ to ensure a completely stable system for all positive values of gain. Figure 13.7*e* shows a possible selection of the zeros and the resulting root locus. Note that when the conventional cascade compensators of Chap. 10 are used, the angles of the asymptotes yield branches that go, in general, into the right-half (RH) s plane. The maximum gain is therefore limited in order to maintain system stability. This restriction is removed with state feedback.
7. The desired figures of merit, for an underdamped response, are satisfied by selecting the desired locations of the pole of $Y(s)/R(s)$ of Eq. (13.64) at $s_{1,2} = -a \pm jb$ and $s_3 = -c$. For these specified pole locations the known characteristic equation has the form

$$(s + a \pm jb)(s + c) = s^3 + d_2 s^2 + d_1 s + d_0 = 0 \tag{13.65}$$

   Equating the corresponding coefficients of the denominators of Eqs. (13.64) and (13.65) yields

$$d_2 = 6 + 100k_3 A \tag{13.66}$$

$$d_1 = 25 + 200k_2 A + 100k_3 A \tag{13.67}$$

$$d_0 = 200k_1 A = 200A \tag{13.68}$$

   These three equations with three unknowns can be solved for the required values of A, k_2, and k_3.

It should be noted that if the basic system $G(s)$ has any zeros, they become poles of $H_{eq}(s)$. Thus, the zeros of $G(s)$ cancel these poles of $H_{eq}(s)$ in $G(s)H_{eq}(s)$. Also, the zeros of $G(s)$ become the zeros of the control ratio. The reader is able to verify this property by replacing $2/(s+1)$ in Fig. 13.7*a* by

$2(s+\alpha)/(s+1)$ and performing an H-equivalent reduction. This property is also illustrated in the examples of Sec. 13.11.

13.6 GENERAL PROPERTIES OF STATE FEEDBACK (USING PHASE VARIABLES)

In order to obtain general state-variable-feedback properties that are applicable to any $G(s)$ transfer function, consider the minimum-phase transfer function

$$G(s) = \frac{Y(s)}{U(s)} = \frac{K_G(s^w + c_{w-1}s^{w-1} + \cdots + c_1 s + c_0)}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0} \tag{13.69}$$

where $K_G > 0$, $w < n$, and the coefficients c_i are all positive. The system characteristics and properties are most readily obtained by representing this system in terms of phase variables [see Secs. 5.6 and 5.8, Eqs (5.38) and (5.45), with $c_w = 1$ and u replaced by $K_G u$, and Sec. 13.4]:

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix}\mathbf{x} + \begin{bmatrix} 0 \\ 0 \\ \cdot \\ K_G \end{bmatrix} u \tag{13.70}$$

$$\mathrm{y} = \begin{bmatrix} c_0 & c_1 & \cdots & c_{w-1} & 1 & 0 & \cdots & 0 \end{bmatrix}\mathbf{x} = \mathbf{c}_c^T\mathbf{x} \tag{13.71}$$

From Fig. 13.7*d* and Eq. (13.71)

$$H_{\mathrm{eq}}(s) = \frac{\mathbf{k}_c^T\mathbf{X}(s)}{Y(s)} = \frac{\mathbf{k}_c^T\mathbf{X}(s)}{\mathbf{c}_c^T\mathbf{X}(s)} = \frac{k_1X_1(s) + k_2X_2(s) + \cdots + k_nX_n(s)}{c_0X_1(s) + c_1X_2(s) + \cdots + X_{w+1}} \tag{13.72}$$

If $w = n$, use Eq. (5.44) to obtain $y = Cx + Du$.

The state-feedback matrix is $\mathbf{k}_c^T$ because the state variables are phase variables. Using the phase-variable relationship $X_j(s) = s^{j-1}X_1(s)$, so that $\dot{X}_1 = X_2,\ \dot{X}_2 = X_3,\ \cdots,$ then Eq. (13.72) becomes

$$H_{\mathrm{eq}}(s) = \frac{k_n s^{n-1} + k_{n-1}s^{n-2} + \cdots + k_2 s + k_1}{s^w + c_{w-1}s^{w-1} + \cdots + c_1 s + c_0} \tag{13.73}$$

Thus, from Eqs. (13.69) and (13.73), after cancellation of common terms,

$$G(s)H_{\mathrm{eq}}(s) = \frac{K_G(k_n s^{n-1} + k_{n-1}s^{n-2} + \cdots + k_2 s + k_1)}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0} \tag{13.74}$$

Factoring k_n from the numerator polynomial gives an alternate form of this equation,

$$G(s)H_{\mathrm{eq}}(s) = \frac{K_G k_n(s^{n-1} + \cdots + \alpha_1 s + \alpha_0)}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0} \tag{13.75}$$

where the loop sensitivity is $K_G k_n = K$. Substituting Eqs. (13.69) and (13.74) into Eq. (13.62) yields

$$\frac{Y(s)}{R(s)} = \frac{K_G(s^w + c_{w-1}s^{w-1} + \cdots + c_0)}{s^n + (a_{n-1} + K_G k_n)s^{n-1} + \cdots + (a_0 + K_G k_1)} \tag{13.76}$$

An analysis of these equations yields the following conclusions:

1. The numerator of $H_{\mathrm{eq}}(s)$ is a polynomial of degree $n-1$ in s; that is, it has $n-1$ zeros. The values of the coefficients k_i of this numerator polynomial, and thus the $n-1$ zeros of $H_{\mathrm{eq}}(s)$, can be selected by the system designer.
2. The numerator of $G(s)$, except for K_G, is equal to the denominator of $H_{\mathrm{eq}}(s)$. Therefore, the open-loop transfer function $G(s)H_{\mathrm{eq}}(s)$ has the same n poles as $G(s)$ and has the $n-1$ zeros of $H_{\mathrm{eq}}(s)$ which are to be determined in the design process. [An exception to this occurs when physical variables are utilized and the transfer function $X_n(s)/U(s)$ contains a zero. In that case this zero does not appear as a pole of $H_{\mathrm{eq}}(s)$, and the numerator and denominator of $G(s)H_{\mathrm{eq}}(s)$ are both of degree n.]
3. The root-locus diagram, based on Eq. (13.75) for $k_n > 0$, reveals that $n-1$ branches terminate on the $n-1$ arbitrary zeros, which can be located anywhere in the s plane by properly selecting the values of the k_i. There is a single asymptote at $\gamma = -180^\circ$, and therefore one branch terminates on the negative real axis at $s = -\infty$. System stability is ensured for high values of K_G if all the zeros of $H_{\mathrm{eq}}(s)$ are located in the left-half (LH) s plane. For $k_n < 0$, a single asymptote exists at $\gamma = 0^\circ$; thus the system becomes unstable for high values of K_G.
4. State-variable feedback provides the ability to select the locations of the zeros of $G(s)H_{\mathrm{eq}}(s)$, as given by Eq. (13.75), in order to locate closed-loop system poles that yield the desired system performance. From Eq. (13.76) it is noted that the result of feedback of the states through constant gains k_i is to change the closed-loop poles and to leave the system zeros, if any, the same as those of $G(s)$. (This requires the plant to be controllable, as described in Sec. 13.2.) It may be desirable to specify that one or more of

the closed-loop poles cancel one or more unwanted zeros in $Y(s)/R(s)$.

5. In synthesizing the desired $Y(s)/R(s)$, the number of poles minus the number of zeros of $Y(s)/R(s)$ must equal the number of poles minus the number of zeros of $G(s)$.

A thorough steady-state error analysis of Eq. (13.76) is made in Secs. 13.7 and 13.8 for the standard step, ramp, and parabolic inputs. From the conclusions obtained in that section, some or all of the feedback gains k_1, k_2, and k_3 are specified.

Summarizing, the $H_{eq}(s)$ can be designed to yield the specified closed-loop transfer function in the manner described for the system of Fig. 13.7. A root-locus sketch of $G(s)H_{eq}(s) = \pm 1$ can be used as an aid in synthesizing the desired $Y(s)/R(s)$ function. From this sketch it is possible to determine the locations of the closed-loop poles, in relation to any system zeros, that will achieve the desired performance specifications. Note that although the zeros of $G(s)$ are *not* involved in obtaining the root locus for $G(s)H_{eq}(s)$, they do appear in $Y(s)/R(s)$ and thus affect the time function $y(t)$.

DESIGN PROCEDURE. The design procedure is summarized as follows:

Step 1. Draw the system state-variable block diagram when applicable.
Step 2. Assume that all state variable are accessible.
Step 3. Obtain $H_{eq}(s)$ and $Y(s)/R(s)$ in terms of the k_i's.
Step 4. Determine the value of k_1 required for zero steady-state error for a unit step input, $u_{-1}(t)$; that is, $y(t)_{ss} = \lim_{s\to 0} sY(s) = 1$.
Step 5. Synthesize the desired $[Y(s)/R(s)]_T$ from the desired performance specifications for $M_p, t_p, t_s,$ and steady-state error requirements (see Sec. 12.2). It may also be desirable to specify that one or more of the closed-loop poles should cancel unwanted zeros of $Y(s)/R(s)$.
Step 6. Set the closed-loop transfer functions obtained in steps 3 and 5 equal to each other. Equate the coefficients of like powers of s of the denominator polynomials and solve for the required values of the k_i. If phase variables are used in the design, a linear transformation must be made in order to convert the phase-variable-feedback coefficients required for the physical variables present in the control system.

If some of the states are not accessible, suitable cascade or minor-loop compensators can be determined using the known values of the k_i. This method is discussed in Sec. 14.6.

The procedure for designing *a state-variable-feedback system* is illustrated by means of examples using physical variables in Secs. 13.9

through 13.12. Before the examples are considered, however, the steady-state error properties are investigated in Secs. 13.7 and 13.8.

13.7 STATE-VARIABLE FEEDBACK: STEADY-STATE ERROR ANALYSIS

A steady-state error analysis of a stable control system is an important design consideration (see Chap. 6). The three standard inputs are used for a steady-state analysis of a state-variable-feedback control system. The phase-variable representation is used in the following developments, in which the necessary values of the feedback coefficients are determined. The results apply only when phase variables are used and do not apply when general state variables are used.

Step Input $r(t) = R_0 u_{-1}(t), R(s) = R_0/s$

For a step input, solving Eq. (13.76), which is based upon the phase-variable representation for $Y(s)$, and applying the final-value theorem yields

$$y(t)_{ss} = \frac{K_G c_0 R_0}{a_0 + K_G k_1} \tag{13.77}$$

The system error is defined by

$$e(t) \equiv r(t) - y(t) \tag{13.78}$$

For zero steady-state error, $e(t)_{ss} = 0$, the steady-state output must be equal to the input:

$$y(t)_{ss} = r(t) = R_0 = \frac{K_G c_0 R_0}{a_0 + K_G k_1} \tag{13.79}$$

where $c_0 \neq 0$. This requires that $K_G c_0 = a_0 + K_G k_1$, or

$$k_1 = c_0 - a_0/K_G \tag{13.80}$$

Thus, zero steady-state error with a step input can be achieved with state-variable feedback even if $G(s)$ is Type 0. For an all-pole plant $c_0 = 1$, and for a Type 1 or higher plant, $a_0 = 0$. In that case the required value is $k_1 = 1$. *Therefore, the feedback coefficient k_1 is determined by the plant parameters once zero steady-state error for a step input is specified.* This specification is maintained for all state-variable-feedback control-system designs throughout the remainder of this text. From Eq. (13.76), the system characteristic equation is

$$s^n + (a_{n-1} + K_G k_n)s^{n-1} + \cdots + (a_2 + K_G k_3)s^2 + (a_1 + K_G k_2)s \\ + (a_0 + K_G k_1) = 0 \tag{13.81}$$

where $a_1 + K_G k_2 = K_G c_1$ and $a_0 + K_G k_1 = K_G c_0$. A necessary condition for all closed-loop poles of $Y(s)/R(s)$ to be in the LH s plane is that all coefficients in Eq. (13.81) must be positive. Therefore, the constant term of the polynomial $a_0 + K_G k_1 = K_G c_0$ must be positive.

Ramp Input $r(t) = R_1 u_{-2}(t) = R_1 t u_{-1}(t), R(s) = R_1/s^2$

To perform the analysis for ramp and parabolic inputs, the error function $E(s)$ must first be determined. Solving for $Y(s)$ from Eq. (13.76) and evaluating $E(s) = R(s) - Y(s)$ yields

$$E(s) = R(s)\frac{\begin{matrix} s^n + (a_{n-1} + K_G k_n)s^{n-1} + \cdots + (a_0 + K_G k_1) \\ -K_G(s^w + c_{w-1}s^{w-1} + \cdots + c_1 s + c_0) \end{matrix}}{s^n + (a_{n-1} + K_G k_n)s^{n-1} + \cdots + (a_1 + K_G k_2)s + (a_0 + K_G k_1)} \tag{13.82}$$

Substituting Eq. (13.80), the requirement for zero steady-state error for a step input, into Eq. (13.82) produces a cancellation of the constant terms in the numerator. Thus,

$$E(s) = R(s)\frac{\begin{matrix} s^n + (a_{n-1} + K_G k_n)s^{n-1} + \cdots + (a_1 + K_G k_2)s \\ -K_G(s^w + c_{w-1}s^{w-1} + \cdots + c_1 s) \end{matrix}}{s^n + (a_{n-1} + K_G k_n)s^{n-1} + \cdots + (a_1 + K_G k_2)s + K_G c_0} \tag{13.83}$$

Not that the numerator has a factor of s.

For a ramp input, the steady-state error is the constant

$$c(t)_{ss} = \lim_{s \to 0} sE(s) = \frac{a_1 + K_G k_2 - K_G c_1}{K_G c_0} R_1 \tag{13.84}$$

In order to obtain $e(t)_{ss} = 0$, it is necessary that $a_1 + K_G k_2 - K_G c_1 = 0$. This is achieved when

$$k_2 = c_1 - a_1/K_G \tag{13.85}$$

For an all-pole plant, $c_0 = 1$ and $c_1 = 0$; thus $k_2 = -a_1/K_G$. In that case, the coefficient of s in the system characteristic equation (13.81) is zero, and the system is unstable. *Therefore, zero steady-state error for a ramp input cannot be achieved by state-variable feedback for an all-pole stable system.* If $a_1 + K_G k_2 > 0$, in order to obtain a positive coefficient of s in Eq. (13.81), the system may be stable but a finite steady-state error exists. The error in all-pole plant is

$$e(t)_{ss} = \left(k_2 + \frac{a_1}{K_G}\right) R_1 \tag{13.86}$$

When a plant contains at least one zero, so that $c_1 > 0$, it is possible to specify the value of k_2 given by Eq. (13.85) and to have a stable system. In that case the system has zero steady-state error with a ramp input. Substituting Eq. (13.85) into Eq. (13.81) results in the positive coefficient $K_G c_1$ for the first power of s. *Thus, a stable state-variable-feedback system having zero steady-state error with both a step and a ramp input can be achieved only when the control ratio has at least one zero.* Note that this occurs even though $G(s)$ may be Type 0 or 1.

Parabolic Input $r(t) = R_2 u_{-3}(t) = (R_2 t^2/2) u_{-1}(t), R(s) = R_2/s^3$

The all-pole and pole-zero plants are considered separately for a parabolic input. For the all-pole plant with Eq. (13.80) satisfied, Eq. (13.82) reduces to

$$E(s) = \frac{R_2}{s^3}\left[\frac{s^n + (a_{n-1} + K_G k_n)s^{n-1} + \cdots + (a_1 + K_G k_2)s}{s^n + (a_{n-1} + K_G k_n)s^{n-1} + \cdots + (a_1 + K_G k_2)s + K_G c_0}\right] \tag{13.87}$$

Because $e(t)_{ss} = \lim_{s \to 0} sE(s) = \infty$, *it is concluded that an all-pole plant cannot follow a parabolic input.* This is true even if Eq. (13.80) is not satisfied.

Under the condition this $e(t)_{ss} = 0$ for both step and ramp inputs to a pole-zero stable system, Eqs. (13.80) and (13.85) are substituted into Eq. (13.83) to yield, for a plant containing at least one zero,

$$E(s) = \frac{R_2}{s^3}\left[\frac{\begin{matrix} s^n + (a_{n-1} + K_G k_n)s^{n-1} + \cdots + (a_2 + K_G k_3)s^2 \\ -K_G(c_w s^w + c_{w-1}s^{w-1} + \cdots + c_2 s^2) \end{matrix}}{s^n + (a_{n-1} + K_G k_n)s^{n-1} + \cdots + K_G c_1 s + K_G c_0}\right] \tag{13.88}$$

Note that s^2 is the lowest power term in the numerator. Applying the final-value theorem to this equation yields

$$e(t)_{ss} = \lim_{s \to 0} sE(s) = \frac{a_2 + K_G k_3 - K_G c_2}{K_G c_0} R_2 \tag{13.89}$$

In order to achieve $e(t)_{ss} = 0$, it is necessary that $a_2 + K_G k_3 - K_G c_2 = 0$. Thus

$$k_3 = c_2 - a_2/K_G \tag{13.90}$$

With this value of k_3 the s^2 term in Eq. (13.81) has the positive coefficient $K_G c_2$ only when $w \geq 2$. *Thus, a stable system having zero steady-state error with a* *parabolic input can be achieved, even if* $G(s)$ *is Type* 0, *only for a pole-zero system with two or more zeros.*

In a system having only one zero ($w = 1$), the value $c_2 = 0$ and $k_3 = -a_2/K_G$ produces a zero coefficient for the s^2 term in the characteristic equation (13.81), resulting in an unstable system. *Therefore, zero steady-state error with a parabolic input cannot be achieved by a state-variable-feedback pole-zero stable system with one zero.* If $k_3 \neq -a_2/K_G$ and $a_2 + K_G k_3 > 0$ in order to maintain a positive coefficient of s^2 in Eq. (13.81), a finite error exists. The system can be stable, and the steady-state error is

$$e(t)_{ss} = \frac{R_2(a_2 + K_G k_3)}{K_G c_0} \tag{13.91}$$

When a performance specifications requires the closed-loop system to follow ramp and parabolic inputs with no steady-state error, and if the plant has no zeros, it is possible to add a cascade block that contains one or more zeros. The insertion point for this block must take into account the factors presented in Sec. 13.11.

13.8 USE OF STEADY-STATE ERROR COEFFICIENTS

For the example represented by Fig. 13.7*a* the feedback coefficient blocks are manipulated to yield the *H*-equivalent block diagram in Fig. 13.7*d*. Similarly a *G-equivalent* unity-feedback block diagram, as shown in Fig. 13.8, can be obtained. The overall transfer functions for Figs. 13.7*d* and 13.8 are, respectively,

$$M(s) = \frac{Y(s)}{R(s)} = \frac{G(s)}{1 + G(s)H_{eq}(s)} \tag{13.92}$$

$$M(s) = \frac{Y(s)}{R(s)} = \frac{G_{eq}(s)}{1 + G_{eq}(s)} \tag{13.93}$$

Solving for $G_{eq}(s)$ from Eq. (13.93) yields

$$G_{eq}(s) = \frac{M(s)}{1 - M(s)} \tag{13.94}$$

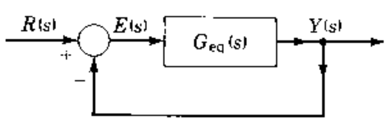


FIGURE 13.8 *G*-equivalent block diagram.

Using the control ratio $M(s)$ from Eq. (13.76) in Eq. (13.94) yields

$$G_{eq}(s) = \frac{K_G(s^w + \cdots + c_1 s + c_0)}{s^n + (a_{n-1} + K_G k_n)s^{n-1} + \cdots + (a_1 + K_G k_2)s + (a_0 + K_G k_1) - K_G(s^w + \cdots + c_1 s + c_0)} \tag{13.95}$$

This is a Type 0 transfer function and has the step-error coefficient $K_p = K_G c_0/(a_0 + K_G k_1 - K_G c_0)$. When the condition $k_1 = c_0 - a_0/K_G$ is satisfied so that $e(t)_{ss} = 0$ with a step input, Eq. (13.95) reduces to

$$G_{eq}(s) = \frac{K_G(s^w + \cdots + c_1 s + c_0)}{s^n + (a_{n-1} + K_G k_n)s^{n-1} + \cdots + (a_1 + K_G k_2)s - K_G(s^w + \cdots + c_1 s)} \tag{13.96}$$

Since this is a Type 1 transfer function, the step-error coefficient is $K_p = \lim_{s\to 0}[sG_{eq}(s)] = \infty$, and ramp error coefficient is

$$K_v = K_G c_0/(a_1 + K_G k_2 - K_G c_1). \tag{13.97}$$

Satisfying the conditions for $e(t)_{ss} = 0$ for both step and ramp inputs, as given by Eqs. (13.80) and (13.85), Eq. (13.96) reduces to

$$G_{eq}(s) = \frac{K_G(s^w + \cdots + c_1 s + c_0)}{s^n + (a_{n-1} + K_G k_n)s^{n-1} + \cdots + (a_2 + K_G k_3)s^2 - K_G(s^w + \cdots + c_2 s^2)} \tag{13.98}$$

Since this is a Type 2 transfer function, there is zero steady-state error with a ramp input. It follows a parabolic input with a steady-state error. The error coefficients are $K_p = \infty$, $K_v = \infty$, and $K_a = K_G c_0/(a_2 + K_G k_3 - K_G c_2)$.

For the case when $w \geq 2$ and the conditions for $e(t)_{ss} = 0$ for all three standard inputs are satisfied [Eqs. (13.80), (13.85), and (13.96)], Eq. (13.98) reduces to

$$G_{eq}(s) = \frac{K_G(s^w + \cdots + c_1 s + c_0)}{s^n + (a_{n-1} + K_G k_n)s^{n-1} + \cdots + (a_3 + K_G k_4)s^3 - K_G(s^w + \cdots + c_3 s^3)} \tag{13.99}$$

This is a Type 3 transfer function that has zero steady-state error with a parabolic input. The error coefficients are $K_p = K_v = K_a = \infty$. The results of the previous analysis are contained in Table 13.1. Table 13.2 summarizes the zero steady-state error requirements for a stable state-variable-feedback control system. Based on Tables 13.1 and 13.2, the specification on the type of input that the system must follow determines the number of zeros that must be included in $Y(s)/R(s)$. The number of zeros determines the ability of the system to follow a particular input with no steady-state error and to achieve a stable system.

TABLE 13.1 State-Variable-Feedback System: Steady-State Error Coefficents[†]

State-variable feedback system	Number of zeros	K_p	K_v	K_a
All-pole plant	$w=0$	∞	$\dfrac{K_G}{a_1+K_Gk_2}$	0
Pole-zero plant	$w=1$	∞	∞	$\dfrac{K_Gc_0}{a_2+K_Gk_3}$
	$w\geq 2$	∞	∞	∞

[†]For phase-variable representation.

TABLE 13.2 Stable State-Variable-Feedback System: Requirements for Zero Steady-State Error[†]

System	Number of zeros required	Input: Step	Input: Ramp	Input: Parabola	G_{eq} (s) type
All-pole plant	$w=0$	$k_1=1-\dfrac{a_0}{K_G}$	‡	§	1
Pole-zero plant	$w=1$	$k_1=c_0-\dfrac{a_0}{K_G}$	$k_2=c_1-\dfrac{a_1}{K_G}$	‡	2
	$w\geq 2$			$k_3=c_1-\dfrac{a_2}{K_G}$	3

[†]For phase-variable representation.
[‡]A steady-steady error exists.
§Cannot follow.

With state-variable feedback, any desired system type can be achieved for $G_{eq}(s)$, regardless of the type represented by $G(s)$. In order to achieve a Type m system, the requirement for $m>0$ is that

$$a_{i-1}+K_Gk_1-K_Gc_{i-1}=0 \qquad \text{for all } i=1,2,3,\ldots,m \tag{13.100}$$

Thus, a system in which the plant is not Type m can be made to behave as a Type m system by use of state feedback. This is achieved without adding pure integrators in cascade with the plant. *In order to achieve a Type m system it is seen from Table* 13.2 *that the control ratio must have a minimum of $m-1$ zeros.* This is an important characteristic of state-variable-feedback systems.

The requirement for $e(t)_{ss}=0$ for a ramp input can also be specified in terms of the locations of the zeros of the plant with respect to the desired locations of the poles of the control ratio (see Chap. 12). Expanding

$$\frac{E(s)}{R(s)} = \frac{1}{1 + G_{\text{eq}}(s)} \tag{13.101}$$

in a Maclaurin series in s yields

$$F(s) = \frac{E(s)}{R(s)} = e_0 + e_1 s + e_2 s^2 + \cdots + e_i s^i + \cdots \tag{13.102}$$

where the coefficients $e_0, e_1, e_2, \ldots$ are called the *generalized error coefficients* [11] and are given by

$$e_i = \frac{1}{i!}\left[\frac{d^i F(s)}{ds^i}\right]_{s=0} \tag{13.103}$$

From this equation, e_0 and e_1 are:

$$e_0 = \frac{1}{1 + K_p} \tag{13.104}$$

$$e_1 = \frac{1}{K_v} \tag{13.105}$$

Note that e_0 and e_1 are expressed in terms of the system error coefficients defined in Chap. 6. When the requirement of Eq. (13.80) is satisfied, Eq. (13.96) yields $K_p = \infty$. Solving for $E(s)$ from Eq. (13.102) and substituting the result into $Y(s) = R(s) - E(s)$ yields

$$\frac{Y(s)}{R(s)} = 1 - e_0 - e_1 s - e_2 s^2 - \cdots \tag{13.106}$$

The relation between K_v and the poles and zeros of $Y(s)/R(s)$ is readily determined if the control ratio is written in factored form:

$$\frac{Y(s)}{R(s)} = \frac{K_G(s - z_1)(s - z_2)\cdots(s - z_w)}{(s - p_1)(s - p_2)\cdots(s - p_n)} \tag{13.107}$$

The first derivative of $Y(s)/R(s)$ in Eq. (13.106) with respect to s, evaluated at $s = 0$, yields $-e_1$ which is equal to $-1/K_v$.

When it is noted that $[Y(s)/R(s)]_{s=0} = 1$ for a Type 1 or higher system with unity feedback, the following equation can be written [11]:

$$\frac{1}{K_v} = -\frac{\left[\frac{d}{ds}\left(\frac{Y}{R}\right)\right]_{s=0}}{\left[\frac{Y}{R}\right]_{s=0}} = -\left[\frac{d}{ds}\ln\left(\frac{Y(s)}{R(s)}\right)\right]_{s=0} \tag{13.108}$$

Substitution of Eq. (13.107) into Eq. (13.108) yields

$$\frac{1}{K_v} = -\left(\frac{1}{s-z_1} + \cdots + \frac{1}{s-z_w} - \frac{1}{s-p_1} - \cdots - \frac{1}{s-p_n}\right)_{s=0}$$

$$= \sum_{j=1}^{n} \frac{1}{p_j} - \sum_{j=1}^{w} \frac{1}{z_j} \tag{13.109}$$

Thus, the conditions for $e(t)_{ss} = 0$ for a ramp input, which occurs for $K_v = \infty$, that is, $1/K_v = 0$, is satisfied when

$$\sum_{j=1}^{n} \frac{1}{p_j} = \sum_{j=1}^{w} \frac{1}{z_j} \tag{13.110}$$

Therefore, in synthesizing the desired control ratio, its poles can be located not only to yield the desired transient response characteristics but also to yield the value $K_v = \infty$. In addition, it must contain at least one zero that is not canceled by a pole.

The values of k_1, k_2, and k_3 determined for $e(t)_{ss} = 0$ for the respective input from Eqs. (13.80), (13.85), and (13.90) are based upon the phase-variable representation of the system. These values may not be compatible with the desired roots of the characteristic equation. In that case it is necessary to leave at least one root unspecified so that these feedback gains may be used. If the resulting location of the unspecified root(s) is not satisfactory, other root locations may be tried in order to achieve a satisfactory performance. Once a system is designed that satisfies both the steady-state and the transient time responses, the feedback coefficients for the physical system can be calculated. This is accomplished by equating the corresponding coefficients of the characteristic equations representing, respectively, the physical- and the phase-variable systems. It is also possible to achieve the desired performance by working with the system represented entirely by physical variables, as illustrated by the examples in the following sections.

13.9 STATE-VARIABLE FEEDBACK: ALL-POLE PLANT

The design procedure presented in Sec. 13.6 is now applied to an all-pole plant.

Design Example. For the basic plant having the transfer function shown in Fig. 13.9*a*, state-variable feedback is used in order to achieve a closed-loop response with $M_p = 1.043$, $t_s = 5.65$ s, and zero steady-state error for a step input $r(t) = R_0 u_{-1}(t)$.

Steps 1–3. See Fig. 13.9*b* and *c*, and use Eq. (13.62) to obtain

$$\frac{Y(s)}{R(s)} = \frac{10A}{s^3 + (6 + 2Ak_3)s^2 + [5 + 10A(k_3 + k_2)]s + 10Ak_1} \tag{13.111}$$

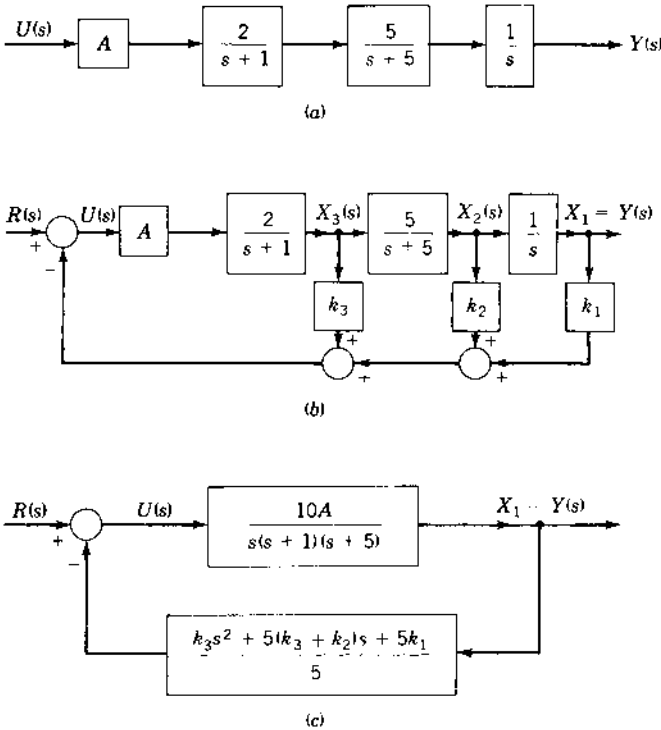


FIGURE 13.9 A state-variable-feedback system, all-pole plant: (*a*) plant; (*b*) state-variable feedback; (*c*) $H_{eq}(s)$.

Step 4. Inserting Eq. (13.111) into $y(t)_{ss} = \lim_{s\to 0} sY(s) = R_0$, for a step input, yields the requirement that $k_1 = 1$.

Step 5. Assuming that Eq. (13.111) has complex poles that are truly dominant, Eq. (3.61) is used to determine the damping ratio.

$$M_p = 1.043 = 1 + \exp\left(-\frac{\zeta\pi}{\sqrt{1-\zeta^2}}\right)$$

or

$$\zeta \approx 0.7079$$

From $T_s = 4/|\sigma|, |\sigma| = 4/5.65 = 0.708$. For these values of ζ and $|\sigma|$, the desired closed-loop complex-conjugate poles are $p_{1,2} = -0.708 \pm j0.7064$. Because the denominator of $G(s)$ is of third order, the control ratio also has a third-order denominator, as shown in Eq. (13.111).

To ensure the dominancy of the complex poles, the third pole is made significantly nondominant and is arbitrarily selected as $p_3 = -100$. Therefore, the desired closed-loop transfer function is

$$\frac{Y(s)}{R(s)} = \frac{10A}{(s+100)(s+0.708 \pm j0.7064)}$$
$$= \frac{10A}{s^3 + 101.4s^2 + 142.6s + 100} \tag{13.112}$$

Step 6. In order to achieve zero steady-state error, a step input requires the selection $k_1 = 1$ (see Sec. 13.5). Then, equating corresponding terms in the denominators of Eqs. (13.111) and (13.113) yields

$$\begin{aligned} 10Ak_1 &= 100 & A &= 10 \\ 6 + 2Ak_3 &= 101.4 & k_3 &= 4.77 \\ 5 + 100(k_3 + k_2) &= 142.7 & k_2 &= -3.393 \end{aligned}$$

Inserting Eq. (13.112) into Eq. (13.94) yields the G-equivalent form (see Fig. 13.8):

$$G_{eq}(s) = \frac{100}{s(s^2 + 101.4s + 142.7)} \tag{13.113}$$

Thus, $G_{eq}(s)$ is Type 1 and the ramp error coefficient is $K_1 = 0.701$.

An analysis of the root locus for the unity-feedback and the state-feedback systems, shown in Fig. 13.10*a* and *b*, respectively, reveals several characteristics.

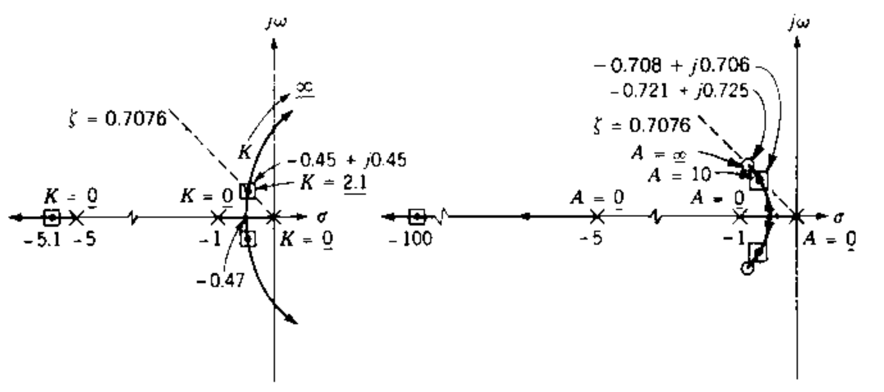


FIGURE 13.10 Root-locus diagrams for (*a*) basic unity-feedback and (*b*) state-variable control systems.

For large values of gain the unity-feedback system is unstable, even if cascade compensation is added. For any variation in gain the corresponding change in the dominant roots changes the system response. Thus, the response is sensitive to changes of gain.

The state-variable-feedback system is stable for all $A > 0$. Because the roots are very close to the zeros of $G(s)H_{eq}(s)$, any increase in gain from the value $A = 10$ produces negligible changes in the location of the roots. In other words, for high-forward-gain (hfg) operation, $A > 0$, the system's response is essentially unaffected by variations in A, provided the feedback coefficients remain fixed. Thus, for the example represented by Fig. 13.10*b*, when $K = 9.54A \geq 95.4$, the change in the values of the dominant roots is small and there is little effect on the time response for a step input. The insensitivity of the system response to the variation of the gain A is related to the system's sensitivity function, which is discussed in the next chapter. *This insensitivity is due to the fact that the two desired dominant roots are close to the two zeros of* $H_{eq}(s)$ *and the third root is nondominant. To achieve this insensitivity to gain variation there must be* β *dominant roots that are located close to* β *zeros of* $H_{eq}(s)$. *The remaining* $n - \beta$ *roots must be nondominant.* In this example $\beta = n - 1 = 2$, but in general $\beta \leq n - 1$.

Both the conventional and state-feedback systems can follow a step input with zero steady-state error. Also, they both follow a ramp input with a steady-state error. The state-variable-feedback system cannot be designed to produce zero steady-state error for a ramp input because $G(s)$ does not contain a zero. In order to satisfy this condition a cascade compensator that adds a zero to $G(s)$ can be utilized in the state-variable-feedback system.

In general, as discussed in Sec. 12.2, when selecting the poles and zeros of $[Y(s)/R(s)]_T$ to satisfy the desired specifications, one pair of dominant complex-conjugate poles $p_{1,2}$ is often chosen. The other $n - 2$ poles are located far to the left of the dominant pair or close to zeros in order to ensure the dominancy of $p_{1,2}$. It is also possible to have more than just a pair of dominant complex-conjugate poles to satisfy the desired specifications. For example, an additional pole of $Y(s)/R(s)$ can be located to the left of the dominant complex pair. When properly located, it can reduce the peak overshoot and may also reduce the settling time. Choosing only one nondominant root, that is, $n - 1$ dominant roots, results in lower values of K and the k_i than when a larger number of nondominant roots is chosen. *For state-variable feedback the value of the gain A is always taken as positive.*

13.10 PLANTS WITH COMPLEX POLES

This section illustrates how the state-variable representation of a plant containing complex poles is obtained. Because the system shown in

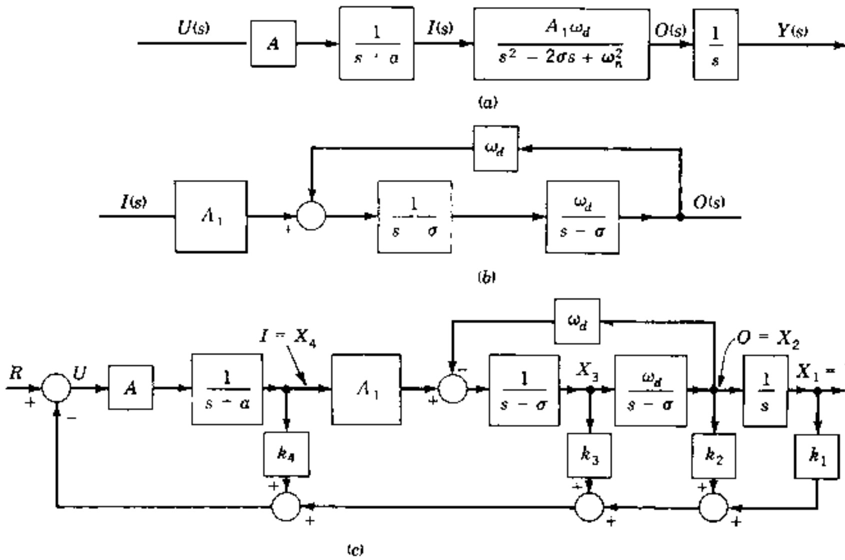


FIGURE 13.11 Representation of a plant containing complex poles.

Fig. 13.11*a* is of the fourth order, four state variables must be identified. The transfer function $O(s)/I(s)$ involves two states. If the detailed block-diagram representation of $O(s)/I(s)$ contains two explicit states, it should not be combined into the quadratic form; by combining the blocks, the identity of one of the states is lost. That situation is illustrated by the system of Fig. 13.6. As represented in Fig. 13.7*a*, the two states X_2 and X_3 are accessible, but in the combined mathematical form of Fig. 13.7*c* the physical variables are not identifiable. When one of the states is not initially identified, an equivalent model containing two states must be formed for the quadratic transfer function $O(s)/I(s)$ of Fig. 13.11*a*. This decomposition can be performed by the method of Sec. 5.12. The simulation diagram of Fig. 5.34 is shown in the block diagram of Fig. 13.11*b*. This decomposition is inserted into the complete control system that contains state-variable feedback, as shown in Fig. 13.11*c*. If the state X_3 is inaccessible, then, once the values of k_i are determined, the physical realization of the state feedback can be achieved by block-diagram manipulation. An alternate approach is to represent the transfer function $O(s)/I(s)$ in state-variable form, as shown in Fig. 13.12. This representation is particularly attractive when X_3 is recognizable as a physical variable. If it is not, then the pick-off for k_3 can be moved to the right by block-diagram manipulation.

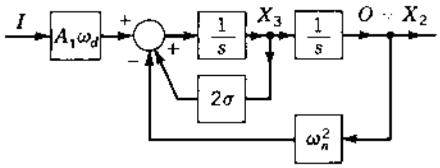


FIGURE 13.12 State-variable representation of $O(s)/I(s)$ in Fig. 13.11*a*.

13.11 COMPENSATOR CONTAINING A ZERO

When a cascade compensator of the form $G_c(s) = (s - z_i)/(s - p_i)$ is required, it can be constructed in either of the forms shown in Fig. 13.13*a* or *b*. The input signal is shown as AU because, in general, this is the most accessible place to insert the compensator. The differential equation for the two forms are

$$\text{For Fig.13.13}a\text{:} \quad \dot{x}_{i+n} = P_i x_{i+n} - z_i A u + A\dot{u} \tag{13.114}$$

$$\text{For Fig.13.13}b\text{:} \quad \dot{x}'_{i+n} = P_i x'_{i+n} + (P_i - z_i) A u \tag{13.115}$$

The presence of $\dot{u}$ in Eq. (13.114) does not permit describing the system with this compensator by the state equation $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u$, and thus x_{i+n} in Fig. 13.13*a* is not a state variable. The presence of a zero in the first block of the forward path can also occur in a system before a compensator is inserted.

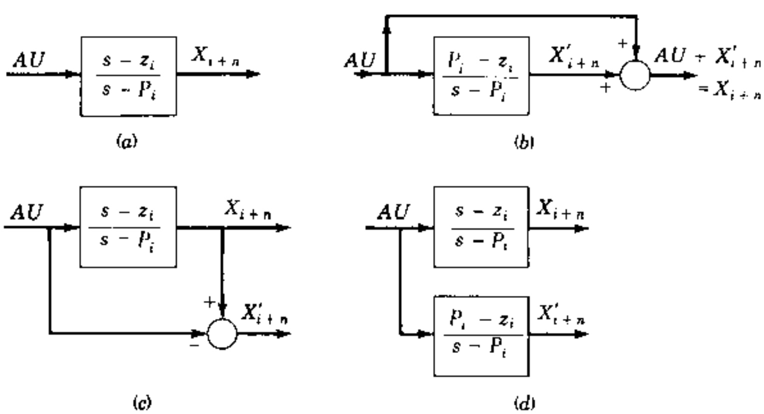


FIGURE 13.13 Cascade compensators: $G_c(s) = (s - z_i)/(s - P_i)$: (*a*) conventional; (*b*) feedforward form; (*c*) generating the state variable X'_{i+n}; (*d*) mathematical model for (*c*).

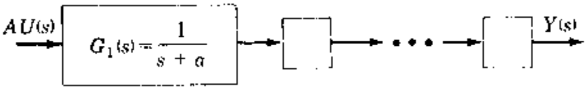


FIGURE 13.14 Forward path.

The design of such a system can be treated like that of a system containing the compensator of Fig. 13.13*a*. The consequence of the $\dot{u}$ term in Eq. (13.114) is that the zero of $G_c(s)$ does not appear as a pole of $H_{\text{eq}}(s)$. Thus the root-locus characteristics discussed in Sec. 13.6 must be modified accordingly. There is no difficulty in expressing the system in state-variable form when the function of Fig. 13.13*a* is not the first frequency-sensitive term in the forward path. When the compensator of Fig. 13.13*a* is inserted at any place to the right of the G_1 block in the forward path shown in Fig. 13.14, the system can be described by a state equation. Thus, the root-locus characteristics discussed in Sec. 13.6 also apply for this case.

A system containing the feedforward compensator in Fig. 13.13*b* can be described by a state equation, because $\dot{u}$ does not appear in Eq. (13.115). Thus, the poles of $H_{\text{eq}}(s)$ continue to be the zeros of $G(s)$. Therefore, the root-locus characteristics discussed in Sec. 13.6 also apply for this case. This is a suitable method of constructing the compensator $G_c(s) = (s - z_i)/(s - P_i)$ because it produces the accessible state variable X'_{i+n}. The implementation of $G_c(s)$ consists of the lag term $K_c/(s - P_i)$, where $K_c = P_i - z_i$, with the unity-gain feedforward path shown in Fig. 13.13*b*. The transfer function for the system of Fig. 13.13*b* is

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

An additional possibility is to generate a new state variable X'_{i+n} from the physical arrangement shown in Fig. 13.13*c*. This is equivalent to the form shown in Fig. 13.13*d* and is mathematically equivalent to Fig. 13.13*b*. Thus, X'_{i+n} is a valid state variable and may be used with a feedback coefficient k_{i+n}.

13.12 STATE-VARIABLE FEEDBACK: POLE-ZERO PLANT

Section 13.9 deals with a system having an all-pole plant; the control ratio therefore has no zeros. Locating the roots of the characteristic equation for such a system to yield the desired system response is relatively easy. This section deals with a system having a plant that contains both poles and zeros and has a control ratio of the form

$$\frac{Y(s)}{R(s)} = \frac{K_G(s - z_1)(s - z_2) \cdots (s - z_h) \cdots (s - z_w)}{(s - p_1)(s - p_2) \cdots (s - p_c) \cdots (s - p_n)} \tag{13.116}$$

that has the same zeros as the plant. When open-loop zeros are present, the synthesis of the overall characteristic equation becomes more complex than for the all-pole plant. As discussed in Sec. 13.8 and summarized in Table 13.2 the presence of a zero in the control ratio is necessary in order to achieve $e(t)_{ss} = 0$ for a ramp input. In synthesizing the desired control ratio, Eq. (13.110) can also be used to assist in achieving this condition while simultaneously satisfying the other desired figures of merit.

If the locations of the zeros of $G_x(s)$ are satisfactory, the design procedure is the same as for the all-pole case of Sec. 13.9. That is, the coefficients of the desired characteristic equation are equated to the corresponding coefficients of the numerator polynomial of $1 + G(s)H_{\text{eq}}(s)$ in order to determine the required K_G and k_i. It is possible to achieve low sensitivity to variation in gain by requiring that

1. $Y(s)/R(s)$ have β dominant poles and *at least one nondominat pole.*
2. The β dominant poles be close β zeros of $H_{\text{eq}}(s)$.

The later can be achieved by a judicious choice of the compensator pole(s) and/or the nondominant poles.

Three examples are used to illustrate state feedback design for plants that contain one zero. The procedure is essentially the same for systems having more than one zero. The design technique illustrated in these examples can be modified as desired.

Trial 1: Design Example 1. The control ratio for the system of Fig. 13.15 is

$$\frac{Y(s)}{R(s)} = \frac{A(s+2)}{s^3 + [6 + (k_3 + k_2)A]s^2 + [5 + (k_3 + 2k_2 + k_1)A]s + 2k_1 A} \tag{13.117}$$

A trial function for the desired control ratio for this system is specified as

$$\left.\frac{Y(s)}{R(s)}\right|_{\text{desired}} = \frac{2}{s^2 + 2s + 2} \tag{13.118}$$

which follows a step input with no steady-state error. Note that three dominant poles are being assigned: $p_1 = -1 + j1, p_2 = -1 - j1, p_3 = -2$. There is no nondominant pole in $Y(s)/R(s)$.

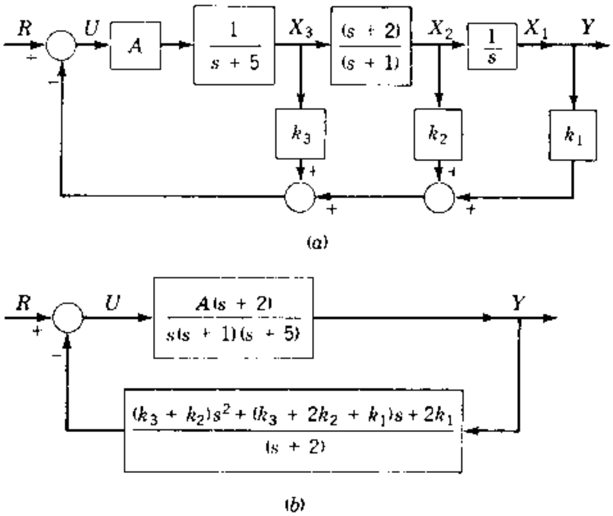


FIGURE 13.15 Closed-loop zero cancellation by state-variable-feedback: (*a*) detailed block diagram; (*b*) *H*-equivalent reduction (where $K_G = A$).

If Eq. (13.117) is to reduce to Eq. (13.119), then one pole of Eq. (13.117) must have the value $s = -2$, in order to cancel the zero, and $A = 2$. Equation (13.118) is modified to reflect this requirement.

$$\left.\frac{Y(s)}{R(s)}\right|_{\text{desired}} = \frac{2(s+2)}{(s^2+2s+2)(s+2)} = \frac{2(s+2)}{(s^3+4s^2+6s+4)} \tag{13.119}$$

Equating the corresponding coefficients of the denominators of Eqs. (13.117) and (13.119) yields $k_1 = 1$, $k_2 = 1/2$, and $k_3 = -3/2$. Thus

$$G(s)H_{\text{eq}}(s) = \frac{-A(s^2 - s/2 - 2)}{s(s+1)(s+5)} = -1 \tag{13.120}$$

whose zeros are $s_1 = -1.189$ and $s_2 = 1.686$. Canceling the minus sign shows that the 0° root locus is required, and it is plotted in Fig. 13.16. For $A = 2$, the roots are $p_{1,2} = -1 \pm j1$ and $p_3 = -2$, as specified by Eq. (13.119). Note that because Eq. (13.118) has no other poles in addition to the dominant poles, the necessary condition of at least one nondominant pole is not satisfied. As a result, two braches of the root locus enter the RH of the s plane, yielding a system that can become unstable for high gain. The response of this system to a unit step input has $M_p = 1.043$, $t_p = 3.1\,s$, and $t_s = 4.2\,s$. Because the roots are not near zeros, the system is highly sensitive to gain variation,

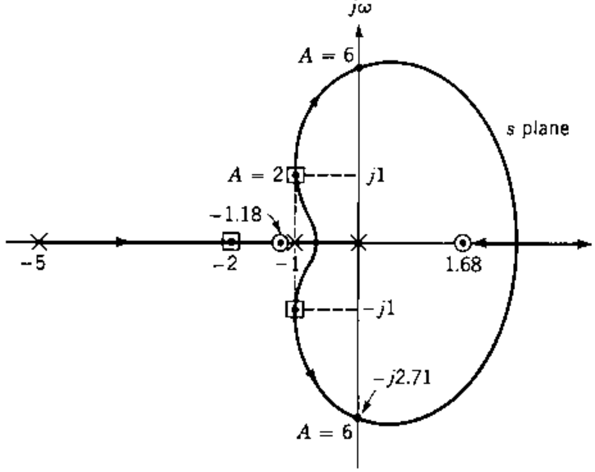


FIGURE 13.16 Root locus for Eq. (13.120).

as seen from the root locus in Fig. 13.16. Any increase in A results in a significant change in performance characteristics, and the system becomes unstable if A exceeds the value of 6. The ramp error coefficient obtained from $G_{eq}(s)$ is $K_1 = 1.0$.

Design Example 1 illustrates that insensitivity to gain variation and stability for all positive values of gain is not achieved. In order to satisfy these conditions, $[Y(s)/R(s)]_T$ must have β dominant poles and at least one nondominant pole. The next example illustrates this design concept.

Trial 2: Design Example 2. For the third-order system of Fig. 13.15, it is desired to cancel the control ratio zero at $s = -2$ and to reduce the system output sensitivity to variations in the gain A while effectively achieving the second-order response of Eq. (13.118). The time response is satisfied if $Y(s)/R(s)$ has three dominant poles $s_{1,2} = -1 \pm j1$ and $s_3 = -2$. However, in order to achieve insensitivity to gain variation, the transfer function $H_{eq}(s)$ must be increased to $n = 4$. Therefore, one pole must be added to $G(s)$ by using the cascade compensator $G_c(s) = 1/(s + a)$. Then $Y(s)/R(s)$ can have one nondominant pole p and the desired control ratio is given by

$$\left.\frac{Y(s)}{R(s)}\right|_{\text{desired}} = \frac{K_G(s+2)}{(s^2+2s+2)(s+2)(s-p)}. \tag{13.121}$$

The selections of the values of $K_G = A$ and p are not independent because the condition of $e(t)_{ss} = 0$ for a step input $r(t) = R_0 u_{-1}(t)$ is required.

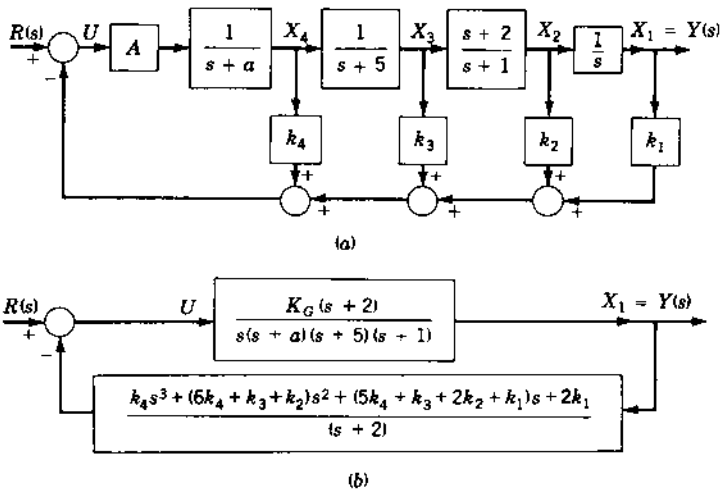


FIGURE 13.17 Closed-loop zero cancellation by state-variable feedback: (*a*) detailed block diagram; (*b*) *H*-equivalent reduction (where $K_G = A$).

Thus

$$y(t)_{ss} = \lim_{s \to 0} sY(s) = \frac{-K_G}{2p} R_0 = R_0 \tag{13.122}$$

or

$$K_G = -2p \tag{13.123}$$

As mentioned previously, the larger the value of the loop sensitivity for the root locus of $G(s)H_{eq}(s) = -1$, the closer the roots will be to the zeros of $H_{eq}(s)$. It is this "closeness" that reduces the system's output sensitivity to gain variations. Therefore, the value of K_G should be chosen as large as possible consistent with the capabilities of the system components. This large value of K_G may have to be limited to a value that does not accentuate the system noise to an unacceptable level. For this example, selecting a value of $K_G = 100$ ensures that p is a nondominant pole. This results in a value $p = -50$ so that Eq. (13.121) becomes

$$\left.\frac{Y(s)}{R(s)}\right|_{\text{desired}} = \frac{100(s+2)}{s^4 + 54s^3 + 206s^2 + 304s + 200)}. \tag{13.124}$$

The detailed state feedback and *H*-equivalent block diagrams for the modified control system are shown in Fig. 13.17*a* and *b*, respectively.

In order to maintain a controllable and observable system the values of a can be chosen to be any value *other than* the value of the zero term of $G(s)$. Thus, assume $a = 1$. With this value of a and with $K_G = 100$, the control ratio of the system of Fig. 13.17b is

$$\frac{Y(s)}{R(s)} = \frac{100(s+2)}{s^4 + (7 + 100k_4)s^3 + [11 + 100(6k_4 + k_3 + k_2)]s^2 + [5 + 100(5k_4 + k_3 + k_2 + k_1)]s + 200k_1} \tag{13.125}$$

For Eq. (13.125) to yield $e(t)_{ss} = 0$ with a step input requires $k_1 = 1$. Equating the corresponding coefficients of the denominators of Eqs. (13.124) and (13.125) and solving for the feedback coefficients yields $k_2 = 0.51$, $k_3 = -1.38$, and $k_4 = 0.47$. Thus, from Fig. 13.17b:

$$G(s)H_{eq}(s) = \frac{0.47K_G(s^3 + 4.149s^2 + 6.362s + 4.255)}{s^4 + 7s^3 + 11s^2 + 5s} \tag{13.126}$$

Verification. The root locus shown in Fig. 13.18 reveals that the three desired dominant roots are close to the three zeros of $H_{eq}(s)$, the fourth root is nondominant, and the root locus lies entirely in the left-half s plane. Thus, the full benefits of state-variable feedback have been achieved. In other words, a completely stable system with low sensitivity to parameter variation has been realized. The figures of merit of this system are $M_p = 1.043$, $t_p = 3.2$ s, $t_s = 4.2$ s, and $K_1 = 0.98\ \text{s}^{-1}$.

Trial 3: Design Example 3. A further improvement in the time response can be obtained by adding a pole and zero to the overall system function of

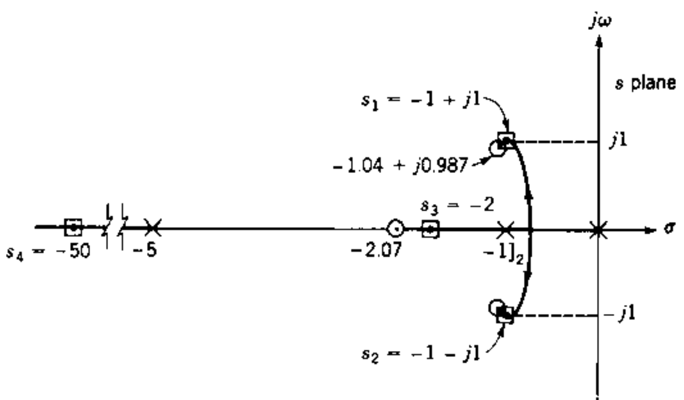


FIGURE 13.18 The root locus for Eq. (13.126).

Design Example 2. The control ratio to be achieved is

$$\frac{Y(s)}{R(s)}=\frac{K_G(s+1.4)(s+2)}{(s+1)(s^2+2s+2)(s+2)(s-p)}$$
$$=\frac{K_G(s+1.4)(s+2)}{s^5+(s-p)s^4+(10-5p)s^3+(10-10p)s^2+(4-10p)s-4p} \tag{13.127}$$

Analysis of Eq. (13.127) shows the following features:

1. Because $Y(s)/R(s)$ has four dominant poles, the nondominant pole p yields the desired property of insensitivity to gain variation.
2. The pole-zero combination $(s+1.4)/(s+1)$ is added to reduce the peak overshoot in the transient response.
3. The factor $s+2$ appears in the numerator because it is present in $G(s)$. Because it is not desired in $Y(s)/R(s)$, it is canceled by producing the same factor in the denominator.
4. In order to achieve zero steady-state error with a step input $r(t)=u_{-1}(t)$, the output is obtained by applying the final-value theorem to $Y(s)$. This yields $y(t)_{ss}=K_G\times 1.4\times 2/(-4p)=1$. Therefore, $p=-0.7K_G$. Assuming the large value of $K_G=100$ to be satisfactory results in $p=-70$, which is nondominant, as required.

The required form of the modified system is shown in Fig. 13.19. The cascade compensator $1/(s+1)$ is added to the system in order to achieve the degree of the denominator of Eq. (13.127). The cascade compensator $(s+1.4)/(s+3)$ is included in order to produce the desired numerator factor $s+1.4$. The denominator is arbitrarily chosen as $s+3$. This increases the degree of the denominator in order to permit canceling the factor $s+2$ in $Y(s)/R(s)$.

The forward transfer function obtained from Fig. 13.19 is

$$G(s)=\frac{100(s+1.4)(s+2)}{s(s+1)^2(s+3)(s+5)}=\frac{100(s^2+3.4s+2.8)}{s^5+10s^4+32s^3+38s^2+15s} \tag{13.128}$$

The value of $H_{eq}(s)$ is evaluated from Fig. 13.19 and is

$$H_{eq}(s)=\frac{\begin{array}{l}(k_5+k_4)s^4+(9k_5+7.4k_4+k_3+k_2)s^3\\ \quad+(23k_5+13.4k_4+2.4k_3+3.4k_2+k_1)s^2\\ \quad+(15k_5+7k_4+1.4k_3+2.8k_2+3.4k_1)s+2.8k_1\end{array}}{(s+1.4)(s+2)} \tag{13.129}$$

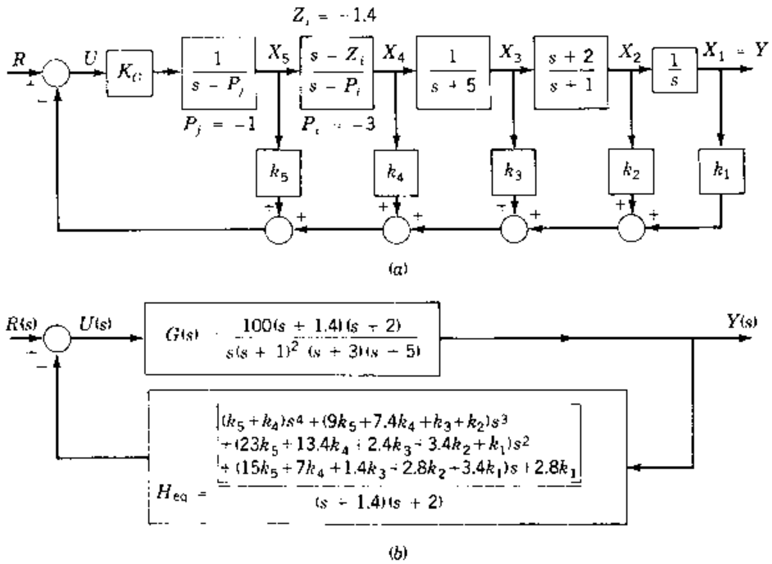


FIGURE 13.19 State feedback diagram for Design Example 3.

The desired overall system function is obtained from Eq. (13.127), with the values $K_G = 100$ and $p = -70$, as follows.

$$\frac{Y(s)}{R(s)} = \frac{100(s + 1.4)(s + 2)}{s^5 + 75s^4 + 360s^3 + 710s^2 + 704s + 280} \tag{13.130}$$

The system function obtained by using $G(s)$ and $H_{eq}(s)$ is

$$\frac{Y(s)}{R(s)} = \frac{100(s+1.4)(s+2)}{\begin{aligned} & s^5 + [10 + 100(k_5 + k_4)]s^4 \\ & + [32 + 100(9k_5 + 7.4k_4 + k_3 + k_2)]s^3 \\ & + [38 + 100(23k_5 + 13.4k_4 + 2.4k_3 + 3.4k_2 + k_1)]s^2 \\ & + [15 + 100(15k_5 + 7k_4 + 1.4k_3 + 2.8k_2 + 3.4k_1)]s + 100(2.8k_1) \end{aligned}} \tag{13.131}$$

Equating the coefficients in the denominators of Eqs. (13.130) and (13.131) yields $280k_1 = 280$, or $k_1 = 1$, and

$$\begin{aligned} 1500k_5 + \ \ 700k_4 + 140k_3 + 280k_2 &= 349 \\ 2300k_5 + 1340k_4 + 240k_3 + 340k_2 &= 572 \\ 900k_5 + \ \ 740k_4 + 100k_3 + 100k_2 &= 328 \\ 100k_5 + \ \ 100k_4 \qquad\qquad\qquad\quad &= 65 \end{aligned} \tag{13.132}$$

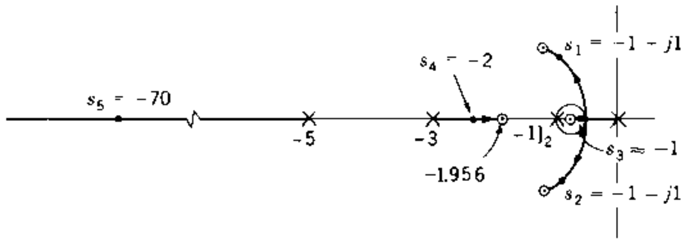


FIGURE 13.20 Root locus for Eq. (13.133). (Drawing is not to scale.)

Solving these four simultaneous equations gives $k_2 = 1.0$, $k_3 = -2.44167$, $k_4 = 0.70528$, and $k_5 = -0.05528$.

In order to demonstrate the desirable features of this system, the root locus is drawn in Fig. 13.20 for

$$G(s)H_{\text{eq}}(s) = \frac{100(0.65s^4 + 3.28s^3 + 6.72s^2 + 6.89s + 2.8)}{s(s+1)^2(s+3)(s+5)}$$
$$= \frac{65(s + 1.04532 \pm j1.05386)(s + 0.99972)(s + 1.95565)}{s(s+1)^2(s+3)(s+5)} \tag{13.133}$$

The root locus shows that the zero of $Y(s)/R(s)$ at $s = -2$ has been canceled by the pole at $s = -2$. Because the root $s_4 = -2$ is near a zero in Fig. 13.20, its value is insensitive to increases in gain. This ensures the desired cancellation. The dominant roots are all near zeros in Fig. 13.20, thus assuring that the desired control ratio $Y(s)/R(s)$ as given by Eq. (13.127) has been achieved. The system is stable for all values of gain K_G. Also, state-variable feedback has produced a system in which $Y(s)/R(s)$ is unaffected by increases in A above 100. The time response $y(t)$ to step input has a peak overshoot $M_p = 1.008$, a peak time $t_p = 3.92$, and a settling time $t_s = 2.94$. The ramp error coefficient evaluated for $G_{\text{eq}}(s)$ is $K_1 = 0.98$. These results show that the addition of the pole-zero combination $(s + 1.4)/(s + 1)$ to $Y(s)/R(s)$ has improved the performance. This system is insensitive to variation in gain.

The three design examples in this section are intended to provide a firm foundation in the procedures for

1. Synthesizing the desired system performance and the overall control ratio.
2. Analyzing the root locus of $G(s)H_{\text{eq}}(s) = -1$ for the desired system performance and verifying insensitivity to variations in gain.

3. Determining the values of k_i to yield the desired system performance.

These approaches are not all-inclusive, and variations may be developed [8, 10].

13.13 OBSERVERS [12–17]

The synthesis of state-feedback control presented in the earlier portions of this chapter assumes that all the states $\mathbf{x}$ are measurable or that they can be generated from the output. In many practical control systems it is physically or economically impractical to install all the transducers which would be necessary to measure all of the states. The ability to reconstruct the plant states from the output requires that all the states be observable. The necessary condition for complete observability is given by Eq. (13.15).

The purpose of this section is to present methods of reconstructing the states from the measured outputs by a dynamical system, which is called an observer. The reconstructed state vector $\hat{\mathbf{x}}$ may then be used to implement a state-feedback control law $\mathbf{u} = \mathbf{K}\hat{\mathbf{x}}$. A basic method of reconstructing the states is to simulate the state and output equations of the plant on a digital computer. Thus, consider a system represented by the state and output equations of Eqs. (13.12) and (13.13). These equations are simulated on a computer with the same input $\mathbf{u}$ that is applied to the actual physical system. The states of the simulated system and of the actual system will then be identical only if the initial conditions of the simulation and of the physical plant are the same. However, the physical plant may be subjected to unmeasurable disturbances, which cannot be applied to the simulation. Therefore, the difference between the actual plant output $\mathbf{y}$ and the simulation output $\hat{\mathbf{y}}$ is used as another input in the simulation equation. Thus, the observer state and output equations become

$$\dot{\hat{\mathbf{x}}} = \mathbf{A}\hat{\mathbf{x}} + \mathbf{B}\mathbf{u} + \mathbf{L}(\mathbf{y} - \hat{\mathbf{y}}) \tag{13.134}$$

and

$$\hat{\mathbf{y}} = \mathbf{C}\hat{\mathbf{x}} \tag{13.135}$$

where $\mathbf{L}$ is the $n \times l$ observer matrix.

A method for synthesizing the matrix $\mathbf{L}$ uses the *observer reconstruction error and its derivative* defined, respectively, by

$$\mathbf{e} \equiv \mathbf{x} - \hat{\mathbf{x}} \tag{13.136a}$$

$$\dot{\mathbf{e}} = \dot{\mathbf{x}} - \dot{\hat{\mathbf{x}}} \tag{13.136b}$$

Subtracting Eq. (13.134) from Eq. (13.22) and using Eqs. (13.23), (13.135), and (13.136) yields the observer-error state equation

$$\dot{\mathbf{e}} = (\mathbf{A} - \mathbf{LC})\mathbf{e} \tag{13.137}$$

By an appropriate choice of the observer matrix **L**, all of the eigenvalues of $\mathbf{A} - \mathbf{LC}$ can be assigned to the left-half plane, so that the steady-state value of $e(t)$ for *any* initial condition is zero:

$$\lim_{t\to 0} \mathbf{e}(t) = \mathbf{0} \tag{13.138}$$

Equation (13.137) indicates that the observer error equation has no input and is excited only by the initial condition, and thus the observer error is not determined by the system input. The steady-state value of the error is therefore equal to zero. The significance of this is that the observer states approach the plant states in the steady-state for any input to the plant. The eigenvalues of $[\mathbf{A} - \mathbf{LC}]$ are usually selected so that they are to the left of the eigenvalues of A. Thus, the observer states have an initial "start-up" error when the observer is turned on, but they rapidly approach the plant states.

The selection of the observer matrix **L** can be used to assign both the eigenvalues and the eigenvectors of the matrix $[\mathbf{A} - \mathbf{LC}]$. Only eigenvalue assignment is considered in this section. The method of Refs. 12 and 13 is used to assign the eigenvectors also. The representation of the physical plant represented by Eqs. (13.22) and (13.23) and the observer represented by Eqs. (13.134) and (13.135) are shown in Fig. 13.21.

Example 3. In the example of Sec. 13.3, the output is used to implement a state feedback control law. An observer is to be synthesized to reconstruct the inaccessible state x_2. The system is completely observable since, from Eq. (13.15),

$$\text{Rank}\,\mathbf{M}_o = \text{Rank}\begin{bmatrix} 1 & -1 \\ 0 & 3 \end{bmatrix} = 2 = n \tag{13.139}$$

The eignevalues $[\mathbf{A} - \mathbf{LC}]$ are determined by the characteristic polynomial

$$\begin{aligned} |\lambda\mathbf{I} - (\mathbf{A} - \mathbf{LC})| &= \left| \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} - \begin{bmatrix} -1 & 3 \\ 0 & -2 \end{bmatrix} + \begin{bmatrix} \ell_1 \\ \ell_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} \right| \\ &= \lambda^2 + (3 + l_1)\lambda + (2 + 2l_1 + 3l_2) = (\lambda + 9)(\lambda + 10) \end{aligned} \tag{13.140}$$

Assigning the eigenvalues as $\lambda_1 = -9$ and $\lambda_2 = -10$ requires that $l_1 = 16$ and $l_2 = \frac{56}{3}$. The effectiveness of the observer in reconstructing the state $x_2(t)$ is

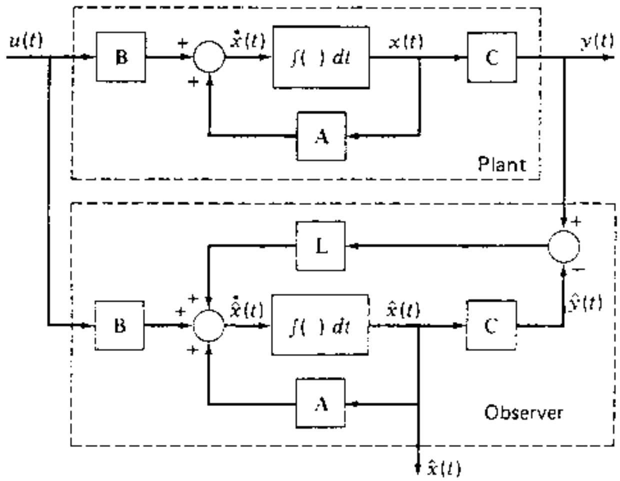


FIGURE 13.21 Plant with inaccessible states and an observer that generates estimates of the states.

shown by comparing the responses of the plant and the observer. For $u(t) = u_{-1}(t)$, $x_1(0) = 0$, $x_2(0) = 1$, $\hat{x}_1(0) = 0$, and $\hat{x}_2(0) = 0$, the initial state reconstruction error are $e_1(0) = 0$ and $e_2(0) = 1$. The responses obtained from Eqs. (13.22), (13.137), and (13.134), respectively, are:

$$x_2(t) = 0.5 + 0.5e^{-2t} \tag{13.141}$$

$$e_2(t) = 8e^{-9t} - 7e^{-10t} \tag{13.142}$$

$$\hat{x}_2(t) = 0.5 + 0.5e^{-2t} - 8e^{-9t} + 7e^{-10t} \tag{13.143}$$

These quantities are plotted in Fig. 13.22 and show that $\hat{x}_2(t)$ converges rapidly to the actual state $x_2(t)$.

The observer described by Eqs. (13.134) and (13.135) is called a full-order observer since it generates estimates of all the states. If some of the states are accessible, as in the example above in which the state $x_1 = y$ is measurable, the order of the observer may be reduced. The description of such reduced-order observers is contained in the literature [6].

13.14 CONTROL SYSTEMS CONTAINING OBSERVERS

This section covers the synthesis of control systems for plants with inaccessible states. In such cases an observer can be used to generate estimates of the states, as described in Sec. 13.13. Consider the case of a controllable and

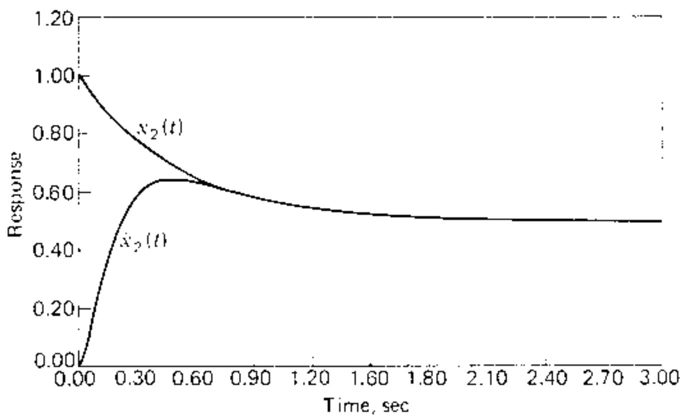


FIGURE 13.22 Plant and observer states, Eqs. (13.141) and (13.143).

observable system governed by the state and output equations of Eqs. (13.22) and (13.23). The observer governed by Eq. (13.134) produces the states $\hat{\mathbf{x}}$ which are estimates of the plant states $\mathbf{x}$. The reconstruction error between the plant states x and the observer states $\hat{\mathbf{x}}$ is given by Eq. (13.136*a*). The matrix $\mathbf{L}$ is selected, as shown in Sec. 13.13, so that the eigenvalues of $[\mathbf{A} - \mathbf{LC}]$ are far to the left of the eigenvalues of $\mathbf{A}$. The control law used to modify the performance of the plant is

$$\mathbf{u} = \mathbf{r} + \mathbf{K}\hat{\mathbf{x}} = \mathbf{r} + \mathbf{K}(\mathbf{x} - \mathbf{e}) \tag{13.144}$$

where $\mathbf{r}$ is the input. Then Eq. (13.22) becomes

$$\dot{\mathbf{x}} = (\mathbf{A} + \mathbf{BK})\mathbf{x} - \mathbf{BKe} + \mathbf{Br} \tag{13.145}$$

The composite closed-loop system consists of the interconnected plant and observer (see Fig. 13.21) and the controller of Eq. (13.144). This closed-loop system is represented by Eqs. (13.145) and (13.137), which are written in the composite matrix form

$$\begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{e}} \end{bmatrix} = \begin{bmatrix} \mathbf{A} + \mathbf{BK} & -\mathbf{BK} \\ \mathbf{0} & \mathbf{A} - \mathbf{LC} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{e} \end{bmatrix} + \begin{bmatrix} \mathbf{B} \\ \mathbf{0} \end{bmatrix} \mathbf{r} \tag{13.146}$$

The characteristic polynomial for the matrix of Eq. (13.146) is

$$Q(\lambda) = \left| \lambda\mathbf{I} - (\mathbf{A} + \mathbf{BK}) \right| \cdot \left| \lambda\mathbf{I} - (\mathbf{A} - \mathbf{LC}) \right| \tag{13.147}$$

Therefore, the eigenvalues of $[\mathbf{A} + \mathbf{BK}]$ and $[\mathbf{A} + \mathbf{LC}]$ can be assiagned independently by the selection of appropriate matrices $\mathbf{K}$ and $\mathbf{L}$. This is an important property which permits the observer and the controller to be designed separately.

Example 4. For the system in the example of Sec. 13.3, the eigenvalues of $[\mathbf{A}+\mathbf{BK}]$ are assigned as $\lambda_1=-5$ and $\lambda_2=-6$ by selecting $k_1=-5$ and $k_2=-3$, or $\mathbf{K}=\begin{bmatrix}-5 & 3\end{bmatrix}$. For $u(t)=u_{-1}(t), x_1(0)=0$, and $x_2(0)=1$, the state response of the system of Eq. (13.32) when the states are accessible is

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \frac{1}{6}-\frac{1}{6}e^{-6t} \\ \frac{1}{30}+\frac{9}{5}e^{-5t}-\frac{5}{6}e^{-6t} \end{bmatrix} \tag{13.148}$$

It is noted, since $y=x_1$, that the output does not track the input. When the states are not accessible and the observer of the example in Sec. 13.14 is used, the composite representation of Eq. (13.146) is

$$\begin{bmatrix} \dot{x}(t) \\ \dot{e}(t) \end{bmatrix} = \begin{bmatrix} -6 & 0 & 5 & 3 \\ -5 & -5 & 5 & 3 \\ 0 & 0 & -17 & 3 \\ 0 & 0 & \frac{-56}{3} & -2 \end{bmatrix} \begin{bmatrix} x(t) \\ e(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} r(t) \tag{13.149}$$

The solution of the plant states from Eq. (13.149) with the input $r(t)=u_{-1}(t), x_1(0), x_2(0)=1$, $e_1(0)=0$, and $e_2(0)=1$ is

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \frac{1}{6}+\frac{23}{6}e^{-6t}-13e^{-9t}+9e^{-10t} \\ \frac{1}{30}-\frac{42}{5}e^{-5t}+\frac{115}{6}e^{-6t}-26e^{-9t}+\frac{81}{5}e^{-10t} \end{bmatrix} \tag{13.150}$$

The difference between Eqs. (13.148) and (13.150) is due to the presence of the observer in the latter case. The responses are plotted in Fig. 13.23 and show that the states reach the same steady-state values.

The contribution of the observer to the plant-state response can be changed by assigning the eigenvectors as well as the eigenvalues. This can be accomplished by the method of Refs. 3 and 4.

13.15 SUMMARY

This chapter presents the controllability and observability properties of a system. These properties are important in the application of many control system design techniques. This is followed by the presentation of the design method using the state-variable representation of SISO systems. Assignment of the closed-loop eigenvalues for SISO systems is readily accomplished when the plant state equation is transformed into the control canonical form $\dot{\mathbf{x}}=\mathbf{A}_c\mathbf{x}+\mathbf{b}_c\mathbf{u}$ and the system is controllable. In this representation,

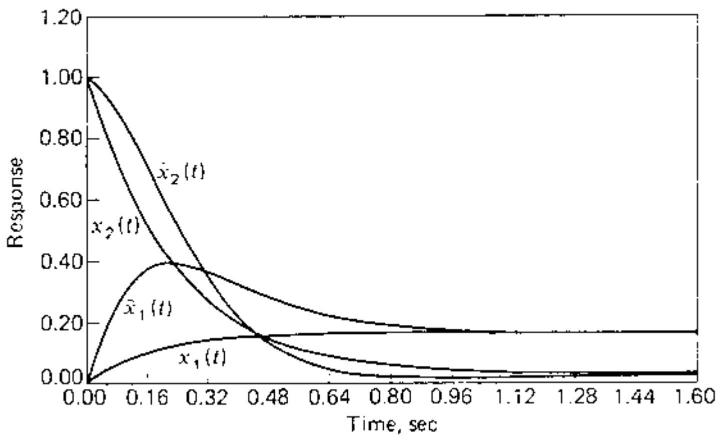


FIGURE 13.23 Plant and observer states, Eqs. (13.148) and (13.150).

the plant matrix $\mathbf{A}_c$ is in companion form and the control matrix $\mathbf{b}_c$ contains all zeros except for the unit element in the last row.

A model control ratio incorporating the desired system performance specifications is used in this chapter to design a compensated state-variable-feedback control system. The root-locus properties and characterization, the steady-state error analysis, and the use of steady-state error coefficients, as they pertain to state-variable-feedback systems (all-pole and pole-zero plants) are discussed in detail. The presence of at least one nondominant closed-loop pole is required in order to minimize the sensitivity of the output response to gain variations. A comparison of the sensitivity to gain variation is made of unity-feedback and state-variable-feedback systems.

The design of an observer to implement a state feedback design is illustrated when all the state variables are observable. The use of CAD programs expedite the achievement of an acceptable design.

REFERENCES

1. Kailath, T.: *Linear Systems*, Prentice-Hall, Englewood Cliffs, N.J., 1980.
2. Wolovich, W.A.: *Linear Multivariable Systems*, Springer-Verlag, New York, 1974.
3. Porter, B. and J. J. D'Azzo: "Closed-Loop Eigenstructure Assignment by State Feedback in Multivariable Systems," *Int. J. Control*, vol. 27, 487–492, 1978.
4. Porter, B. and J. J. D'Azzo: "Algorithm for Closed-Loop Eigenstructure Assignment by State Feedback In Multivariable Feedback," *Int. J. Control*, vol. 27, pp. 943–947, 1978.

5. Chen, C. T., and C. A. Desoer: "Proof of Controllability of Jordan Form State Equations," *Trans. IEEE*, vol. AC-13, pp. 195–196, 1968.
6. Rugh, W. J.: *Linear System Theory*, Prentice-Hall, Englewood Cliffs, N.J., 1993.
7. Gilbert, E. G.: "Controllability and Observability in Multi-Variable Control Systems," *J. Soc. Ind. Appl. Math., Ser. A., Control*, vol. 1, no. 2, pp. 128–151, 1963.
8. Anderson, B. D. O., and J. B. Moore: *Linear Optimal Control*, Prentice-Hall, Englewood Cliffs, N.J., 1971.
9. Kalman, R. E.: "When is a Linear System Optimal," *ASME J. Basic Eng.*, vol. 86, pp. 51–60, March, 1964.
10. Schultz, D. G., and J. L. Melsa: *State Functions and Linear Control Systems*, McGraw-Hill, New York, 1967.
11. James, H. M., N. B. Nichols, and R. S. Phillips: *Theory of Servomechanism*, McGraw-Hill, New York, 1947.
12. Luenberger, D. G.: "Observing the State of a Linear System", *IEEE Trans. Mil. Electron.*, vol. 8, pp. 74–80, 1964.
13. Luenberger, D. G.: "Observers for Multivariable Systems," *IEEE Trans. Autom. Control*, vol. AC-11, pp. 190–197, 1966.
14. Luenberger, D. G.: "An Introduction to Observers," *IEEE Trans. Autom. Control*, vol. AC-16, pp. 596–602, 1971.
15. Kwakernak, H., and R. Sivan: *Linear Optimal Control Systems*, Wiley-Interscience, New York, 1972.
16. Chen, C. T.: *Introduction to Linear Sytem Theory*, Holt, Rinehart and Winston, New York, 1970.
17. Barnett, S.: *Polynomial and Linear Control Systems*, Marcel Dekker, New York, 1983.

14

Parameter Sensitivity and State-Space Trajectories

14.1 INTRODUCTION

In Chap. 13 it is shown that an advantage of a state-feedback control system operating with high forward gain (hfg) is the insensitivity of the system output to gain variations in the forward path. The design method of Chap. 13 assumes that all states are accessible. This chapter investigates in depth the insensitive property of hfg operation of a state-feedback control system. This is followed by a treatment of inaccessible states. In order for the reader to develop a better feel for the kinds of transient responses of a system, this chapter includes an introduction to state-space trajectories. This includes the determination of the steady-state, or equilibrium, values of a system response. While linear time-invariant (LTI) systems have only one equilibrium value, a nonlinear system may have a number of equilibrium solutions. These can be determined from the state equations. They lead to the development of the Jacobian matrix, which is used to represent a nonlinear system by approximate linear equations in the region close to the singular points.

14.2 SENSITIVITY

The environmental conditions to which a control system is subjected affect the accuracy and stability of the system. The performance characteristics of most components are affected by their environment and by aging. Thus, any change in the component characteristics causes a change in the transfer function and therefore in the controlled quantity. The effect of a parameter change on system performance can be expressed in terms of a *sensitivity function*. This sensitivity function S_δ^M is a measure of the sensitivity of the system's response to a system parameter variation and is given by

$$S_\delta^M = \frac{d(\ln M)}{d(\ln \delta)} = \frac{d(\ln M)}{d\delta}\frac{d\delta}{d(\ln \delta)} \tag{14.1}$$

where ln = logarithm to base e, M = system's output response (or its control ratio), and δ = system parameter that varies. Now

$$\frac{d(\ln M)}{d\delta} = \frac{1}{M}\frac{dM}{d\delta} \qquad \text{and} \qquad \frac{d(\ln \delta)}{d\delta} = \frac{1}{\delta} \tag{14.2}$$

Accordingly, Eq. (14.1) can be written

$$\begin{aligned} S_\delta^M \Bigg|_{\substack{M=M_o \\ \delta=\delta_o}} &= \frac{dM/M_o}{d\delta/\delta_o} \\ &= \frac{\delta}{M}\left(\frac{dM}{d\delta}\right)\Bigg|_{\substack{M=M_o \\ \delta=\delta_o}} = \frac{\text{fractional change in output}}{\text{fractional change in system parameter}} \end{aligned} \tag{14.3}$$

where M_o and δ_o represent the nominal values of M and δ. When M is a function of more than one parameter, say $\delta_1, \delta_2, \ldots, \delta_k$, the corresponding formulas for the sensitivity entail partial derivatives. For a small change in δ from δ_o, M changes from M_o, and the sensitivity can be written as

$$S_\delta^M \Bigg|_{\substack{M=M_o \\ \delta=\delta_o}} \approx \frac{\Delta M/M_o}{\Delta\delta/\delta_o} \tag{14.4}$$

To illustrate the effect of changes in the transfer function, four cases are considered for which the input signal $r(t)$ and its transform $R(s)$ are fixed. Although the response $Y(s)$ is used in these four cases, the results are the same when $M(s)$ is the control ratio.

Case 1: Open-Loop System of Fig. 14.1*a*

The effect of a change in $G(s)$, for a fixed $r(t)$ and thus a fixed $R(s)$, can be determined by differentiating, with respect to $G(s)$, the output expression

$$Y_o(s) = R(s)G(s) \tag{14.5}$$

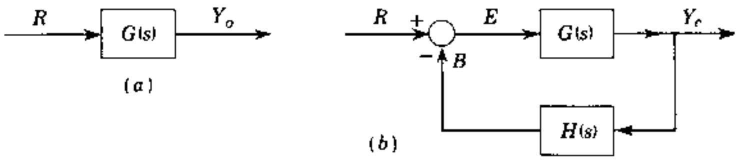


FIGURE 14.1 Control systems: (*a*) open loop; (*b*) closed loop.

giving

$$dY_o(s) = R(s)\,dG(s) \tag{14.6}$$

Combining these two equations gives

$$dY_o(s) = \frac{dG(s)}{G(s)}\,Y_o(s) \rightarrow S_{G(s)}^{Y(s)}(s) = \frac{dY_o(s)/Y_o(s)}{dG(s)/G(s)} = 1 \tag{14.7}$$

A change in the transfer function $G(s)$ therefore causes a proportional change in the transform of the output $Y_o(s)$. This requires that the performance specifications of $G(s)$ be such that any variation still results in the degree of accuracy within the prescribed limits. In Eq. (14.7) the varying function in the system is the transfer function $G(s)$.

Case 2: Closed-Loop Unity-Feedback System of Fig. 14.1*b* $[H(s) = 1]$

Proceeding in the same manner as for case 1, for $G(s)$ varying, leads to

$$Y_c(s) = R(s)\frac{G(s)}{1 + G(s)} \tag{14.8}$$

$$dY_c(s) = R(s)\frac{dG(s)}{[1 + G(s)]^2} \tag{14.9}$$

$$dY_c(s) = \frac{dG(s)}{G(s)[1 + G(s)]}\,Y_c(s) = \frac{1}{1 + G(s)}\,\frac{dG(s)}{G(s)}\,Y_c(s)$$

$$S_{G(s)}^{Y(s)} = \frac{dY_c(s)/Y_c(s)}{dG(s)/G(s)} = \frac{1}{1 + G(s)} \tag{14.10}$$

Comparing Eq. (14.10) with Eq. (14.7) readily reveals that the effect of changes of $G(s)$ upon the transform of the output of the closed-loop control is reduced by the factor $1/|1 + G(s)|$ compared to the open-loop control. *This is an important reason why feedback systems are used.*

Case 3: Closed-Loop Nonunity-Feedback System of Fig. 14.1*b* [Feedback Function $H(s)$ Fixed and $G(s)$ Variable]

Proceeding in the same manner as for case 1, for $G(s)$ varying, leads to

$$Y_c(s) = R(s)\frac{G(s)}{1 + G(s)H(s)} \tag{14.11}$$

$$dY_c(s) = R(s)\frac{dG(s)}{[1 + G(s)H(s)]^2} \tag{14.12}$$

$$dY_c(s) = \frac{dG(s)}{G(s)[1 + G(s)H(s)]}Y_c(s) = \frac{1}{1 + G(s)H(s)}\frac{dG(s)}{G(s)}Y_c(s)$$

$$S_{G(s)}^{Y(s)} = \frac{dY_c(s)/Y_c(s)}{dG(s)/G(s)} = \frac{1}{1 + G(s)H(s)} \tag{14.13}$$

Comparing Eqs. (14.7) and (14.13) shows that the closed-loop variation is reduced by the factor $1/|1 + G(s)H(s)|$. In comparing Eqs. (14.10) and (14.13), if the term $|1+G(s)H(s)|$ is larger than the term $|1 + G(s)|$, then there is an advantage to using a nonunity-feedback system. Further, $H(s)$ may be introduced both to provide an improvement in system performance and to reduce the effect of parameter variations within $G(s)$.

Case 4: Closed-Loop Nonunity-Feedback System of Fig. 14.1*b* [Feedback Function $H(s)$ Variable and $G(s)$ Fixed]

From Eq. (14.11),

$$dY_c(s) = R(s)\frac{-G(s)^2 dH(s)}{[1 + G(s)dH(s)]^2} \tag{14.14}$$

Multiplying and dividing Eq. (14.14) by $H(s)$ and also dividing by Eq. (14.11) results in

$$dY_c(s) = -\left[\frac{G(s)H(s)}{1 + G(s)H(s)}\frac{dH(s)}{H(s)}\right]Y_c(s) \approx -\frac{dH(s)}{H(s)}Y_c(s)$$

$$S_{H(s)}^{Y(s)} = \frac{dY_c(s)/Y_c(s)}{dH(s)/H(s)} \approx -1 \tag{14.15}$$

The approximation applies for those cases where $|G(s)H(s)| >> 1$. When Eq. (14.15) is compared with Eq. (14.7), it is seen that a variation in the feedback function has approximately a direct effect upon the output, the same as for the open-loop case. Thus, the components of $H(s)$ must be selected as precision fixed elements in order to maintain the desired degree of accuracy and stability in the transform $Y(s)$.

The two situations of cases 3 and 4 serve to point out the advantage of feedback compensation from the standpoint of parameter changes. Because the use of fixed feedback compensation minimizes the effect of variations in the components of $G(s)$, prime consideration can be given to obtaining the necessary power requirements in the forward loop rather than to accuracy and stability. $H(s)$ can be designed as a precision device so that the transform $Y(s)$, or the output $y(t)$, has the desired accuracy and stability. *In other words, by use of feedback compensation the performance of the system can be made to depend more on the feedback term than on the forward term.*

Applying the sensitivity equation (14.3) to each of the four cases, where $M(s) = Y(s)/R(s)$, where $\delta = G(s)$ (for cases 1, 2, and 3), and $\delta = H(s)$ (for case 4), yields the results shown in Table 14.1. This table reveals that S_δ^M *never exceeds a magnitude of* 1, and the smaller this value, the less sensitive the system is to a variation in the transfer function. For an increase in the variable function, a positive value of the sensitivity function means that the output increases from its nominal response. Similarly, a negative value of the sensitivity function means that the output decreases from its nominal response. The results presented in this table are based upon a functional analysis; i.e., the "variations" considered are in $G(s)$ and $H(s)$. The results are easily interpreted when $G(s)$ and $H(s)$ are real numbers. When they are not real numbers, and where δ represents the parameter that varies within $G(s)$ or $H(s)$, the interpretation can be made as a function of frequency.

The analysis in this section so far has considered variations in the transfer functions $G(s)$ and $H(s)$. Take next the case when $r(t)$ is sinusoidal. Then the input can be represented by the phasor $\mathbf{R}(j\omega)$ and the output by the phasor $\mathbf{Y}(j\omega)$. The system is now represented by the frequency transfer

TABLE 14.1 Sensitivity Functions

Case	System variable parameter	S_δ^M
1	$G(s)$	$\dfrac{dY_o/Y_o}{dG/G} = 1$
2	$G(s)$	$\dfrac{dY_c/Y_c}{dG/G} = \dfrac{1}{1+G}$
3	$G(s)$	$\dfrac{dY_c/Y_c}{dG/G} = \dfrac{1}{1+GH}$
4	$H(s)$	$\dfrac{dY_c/Y_c}{dH/H} \approx -1$

functions G($j\omega$) and H($j\omega$). All the formulas developed earlier in this section are the same in form, but the arguments are $j\omega$ instead of s. As parameters vary within G($j\omega$) and H($j\omega$), the magnitude of the sensitivity function can be plotted as a function of frequency. The magnitude $\left|S_\delta^M(j\omega)\right|$ does not have the limits of 0 to 1 given in Table 14.1, but can vary from 0 to any large magnitude. An example of sensitivity to parameter variation is investigated in detail in the following section. That analysis shows that the sensitivity function can be considerably reduced with appropriate feedback included in the system structure.

14.3 SENSITIVITY ANALYSIS [3,4]

An important aspect in the design of a control system is the insensitivity of the system outputs to items such as: sensor noise, parameter uncertainty, cross-coupling effects, and external system disturbances. The analysis in this section is based upon the control system shown in Fig. 14.2 where $T(s)=Y(s)/R(s)$ and where $F(s)$ represents a prefilter. The plant is described by $P(s)$ and may include some parameter uncertainties (see Sec. 14.4). In this system $G(s)$ represents a compensator. The prefilter and compensator are designed to minimize the effect of the parameter uncertainties. The goal of the design is to satisfy the desired figures of merit (FOM). In this text it is assumed that $F(s)=1$. The effect of these items on system performance can be expressed in terms of the *sensitivity function* which is defined by

$$S_\delta^T = \frac{\delta}{T}\left[\frac{\partial T}{\partial \delta}\right] \tag{14.16}$$

where δ represents the variable parameter in T. Figure 14.2 is used for the purpose of analyzing the sensitivity of a system to three of these items.

Using the linear superposition theorem, where

$$Y = Y_R + Y_C + Y_N$$

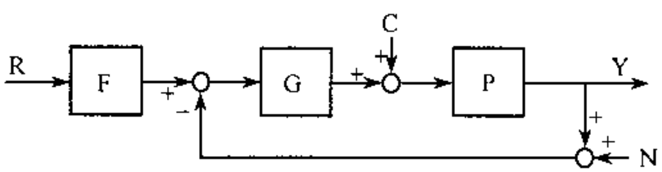


FIGURE 14.2 An example of system sensitivity analysis.

and

$$T_R = \frac{Y_R}{R} = \frac{FL}{1+L} \tag{14.17a}$$

$$T_N = \frac{Y_N}{N} = \frac{-L}{1+L} \tag{14.17b}$$

$$T_C = \frac{Y_C}{C} = \frac{P}{1+L} \tag{14.17c}$$

and $L = GP$ (the loop transmission function), the following transfer functions and sensitivity functions (where $\delta = P$) are obtained, respectively, as:

$$S_P^{T_R} = \frac{FG}{1+L} \tag{14.18a}$$

$$S_P^{T_N} = \frac{-G}{1+L} \tag{14.18b}$$

$$S_P^{T_C} = \frac{1}{1+L} \tag{14.18c}$$

Since sensitivity is a function of frequency, it is necessary to achieve a slope for Lm $\mathbf{L}_o(j\omega)$ that minimizes the effect on the system due to sensor noise. This is the most important case, since the "minimum BW" of Eq. (14.17*b*) tends to be greater than the BW of Eq. (14.17*a*), as illustrated in Fig. 14.3. Based on the magnitude characteristic of $\mathbf{L}_o$ for low- and high-frequency ranges, then:

For the ***low-frequency range***, where $|\mathbf{L}(j\omega)| >> 1$, from Eq. (14.18*b*),

$$S_P^{T_N} \approx \left|\frac{-1}{\mathbf{P}(j\omega)}\right| \tag{14.19}$$

For the ***high-frequency range***, where $|\mathbf{L}(j\omega)| << 1$, from Eq. (14.18*b*),

$$S_P^{T_N} \approx |-\mathbf{G}(j\omega)| = \left|\frac{-\mathbf{L}(j\omega)}{\mathbf{P}(j\omega)}\right| \tag{14.20}$$

The BW characteristics of the open-loop function $\mathbf{L}(j\omega)$, with respect to sensitivity, are illustrated in Fig. 14.4. As seen from this figure, the low-frequency sensitivity given by Eq. (14.19) is satisfactory but the high-frequency sensitivity given by Eq. (14.20) is unsatisfactory since it can present a serious noise rejection problem.

Based upon the analysis of Fig. 14.4, it is necessary to try to make the phase margin frequency ω_ϕ (the loop transmission BW), small enough in

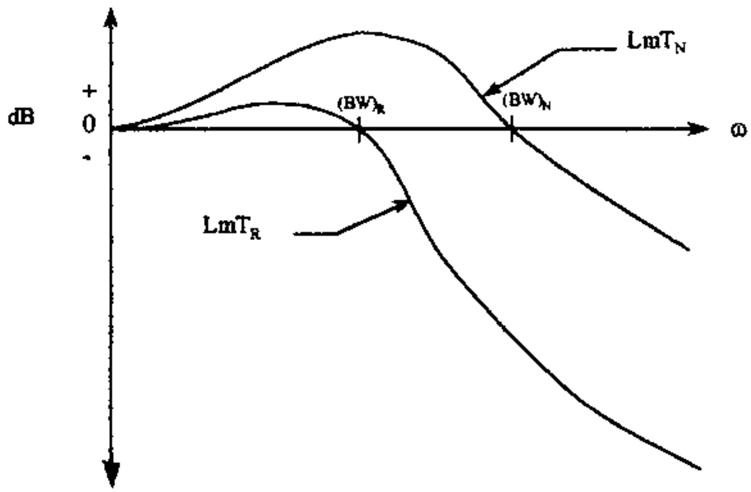


FIGURE 14.3 Frequency response characteristics for the system of Fig. 14.2.

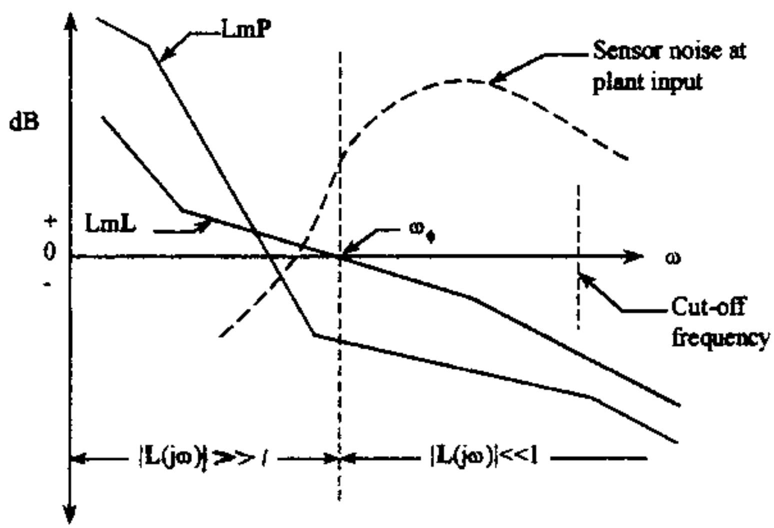


FIGURE 14.4 Bandwidth characteristics of Fig. 14.2.

order to minimize the sensor noise effect on the system's output. For most practical systems $n \geq w + 2$ and

$$\int_0^\infty \log\left[\mathbf{S}_P^T\right] d\omega = 0 \tag{14.21}$$

$$\left|\mathbf{S}_P^T N\right| < 1 \rightarrow \log\left[\mathbf{S}_P^T N\right] < 0 \tag{14.22a}$$

$$\left|\mathbf{S}_P^T N\right| > 1 \rightarrow \log\left[\mathbf{S}_P^T N\right] > 0 \tag{14.22b}$$

Thus, the designer must try to locate the condition of Eq. (14.22*b*) in the high-frequency range where the system performance is not of concern; i.e., the noise effect on the output is negligible.

The analysis for external disturbance effect (see Ref. 4) on the system output is identical to that for cross-coupling effects. For either case, low sensitivity is conducive to their rejection.

14.4 SENSITIVITY ANALYSIS [3,4] EXAMPLES

This section shows how the use of complete state-variable or output feedback minimizes the sensitivity of the output response to a parameter variation. This feature of state feedback is in contrast to conventional control-system design and is illustrated in the example of Sec. 13.9 for the case of system gain variation. In order to illustrate the advantage of state-variable feedback over the conventional method of achieving the desired system performance, a system sensitivity analysis is made for both designs. The basic plant $G_x(s)$ of Fig. 13.9*a* is used for this comparison. The conventional control system is shown in Fig. 14.5*a*. The control-system design can be implemented by the complete state-feedback configuration of Fig. 14.5*b* or by the output feedback representation of Fig. 14.5*c*, which uses H_{eq} as a fixed feedback compensator. The latter configuration is shown to be best for minimizing the sensitivity of the output response with a parameter variation that occurs in the forward path, between the state x_n and the output y. For both the state-feedback configuration and its output feedback equivalent, the coefficients k_i are determined for the nominal values of the parameters and are assumed to be invariant. The analysis is based upon determining the value of the passband frequency ω_b at $|\mathbf{M}_o(j\omega_b)| = M_o = 0.707$ for the nominal system values of the conventional and state-variable-feedback control systems. The system sensitivity function of Eq. (14.3), repeated here, is determined for each system of Fig. 14.5 and evaluated for the frequency range $0 \leq \omega \leq \omega_b$:

$$S_\delta^M(s)\Bigg|_{\substack{M=M_o\\ \delta=\delta_o}} = \left(\frac{\delta}{M}\frac{dM}{d\delta}\right)\Bigg|_{\substack{M=M_o\\ \delta=\delta_o}} \tag{14.23}$$

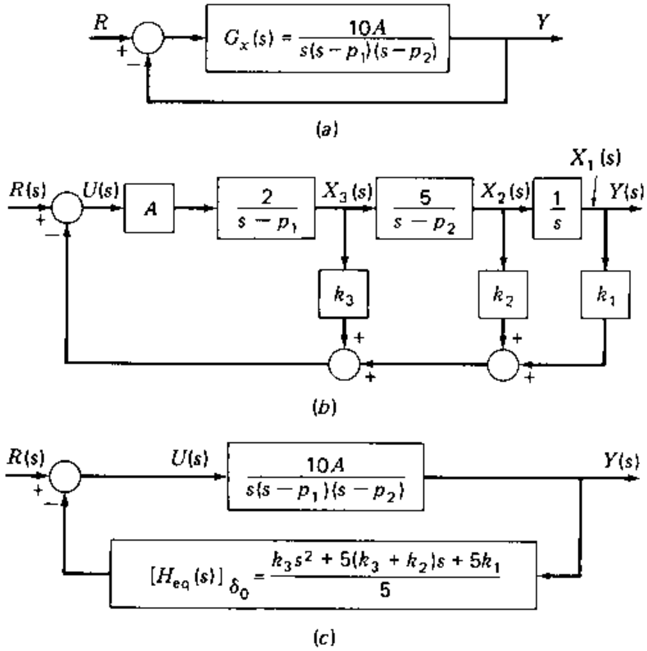


FIGURE 14.5 Control: (*a*) unity feedback; (*b*) state feedback; (*c*) state feedback, *H*-equivalent. The nominal values are $p_1 = -1$, $p_2 = -5$ and $A = 10$.

Note that M is the control ratio. The passband frequency is $\omega_b = 0.642$ for the system of Fig. 14.5*a* with nominal values, and $\omega_b = 1.0$ for the state-feedback system of Figs. 14.5*b* and 14.5*c* with nominal values.

The control ratios for each of the three cases to be analyzed are

Figure 14.5*a*:

$$M(s) = \frac{10A}{s^3 - (p_1 + p_2)s^2 + p_1p_2s + 10A} \tag{14.24}$$

Figure 14.5*b*:

$$M(s) = \frac{10A}{s^3 + (2Ak_3 - p_1 - p_2)s^2 + [p_1p_2 + 2A(5k_2 - k_3p_2)]s + 10Ak_1} \tag{14.25}$$

Figure 14.5*c*:

$$M(s) = \frac{10A}{s^3 + (2Ak_3 - p_1 - p_2)s^2 + [p_1p_2 + 10A(k_3 + k_2)]s + 10Ak_1} \tag{14.26}$$

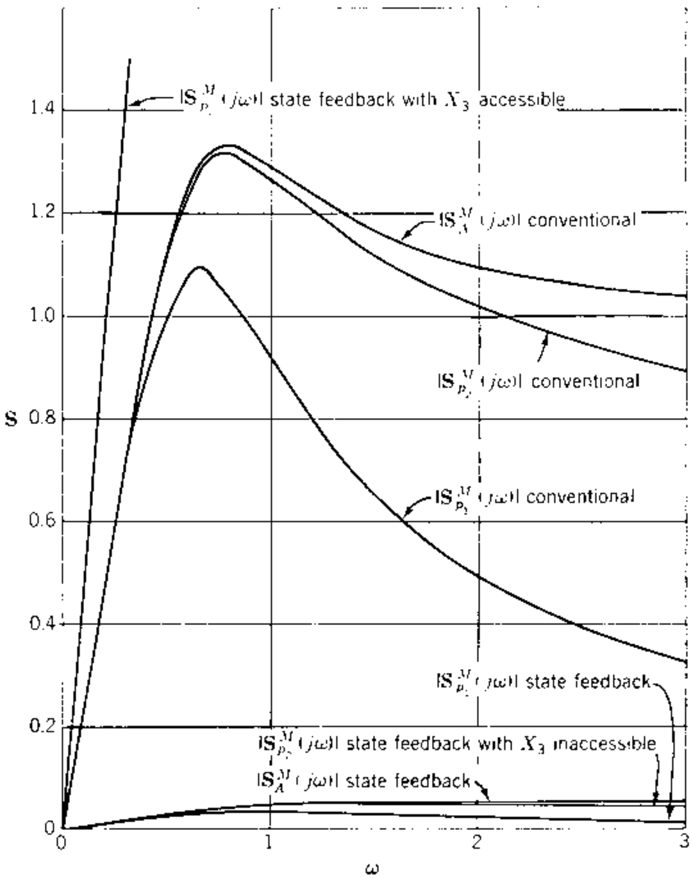


FIGURE 14.6 Sensitivity due to pole and gain variation.

The system sensitivity function for any parameter variation is readily obtained for each of the three configurations of Fig. 14.5 by using the respective control ratios of Eqs. (14.24) to (14.26). The plots of $|\mathbf{S}_{p_1}^M(j\omega)|$, $|\mathbf{S}_{p_2}^M(j\omega)|$, and $|\mathbf{S}_A^M(j\omega)|$ vs. ω, shown in Fig. 14.6, are drawn for the nominal values of $p_1 = -1$, $p_2 = -5$, and $A = 10$ for the state-variable-feedback system. For the conventional system, the value of loop sensitivity equal to 2.1 corresponds to $\zeta = 0.7076$, as shown on the root locus in Fig. 13.10*a*. The same damping ratio is used in the state-variable-feedback system developed in the example in Sec. 13.9. Table 14.2 summarizes the values obtained.

TABLE 14.2 System Sensitivity Analysis

Control system	Figure	$\lvert M(j\omega_b)\rvert$	$\%\lvert\Delta M(j\omega_b)\rvert$	$\lvert S_{p_1}^{M}(j\omega_b)\rvert$	$\lvert S_{p_2}^{M}(j\omega_b)\rvert$	$\lvert S_{A}^{M}(j\omega_b)\rvert$
Conventional system design						
1. Nominal plant ($A = 0.21$, $p_1 = -1$, $p_2 = -5$): $G(s) = \dfrac{2.1}{s(s+1)(s+5)}$	14.5a	0.707 ($\omega_b = 0.642$)	—	1.09	1.28	1.29
2. $P_1 = -2$: $G(s) = \dfrac{2.1}{s(s+2)(s+5)}$	14.5a	0.338	52.4	—	—	—
3. $P_2 = -10$: $G(s) = \dfrac{2.1}{s(s+1)(s+10)}$	14.5a	0.658	6.9	—	—	—
4. $A = 0.42$: $G(s) = \dfrac{4.2}{s(s+1)(s+5)}$	14.5a	1.230	74.0	—	—	—

State-variable feedback system design [*M* is given in Eqs. (13.111), (14.25), and (14.26), respectively]						
5. Nominal plant: $G(s)H_{eq}(s)=\dfrac{95.4(s^2+1.443s+1.0481)}{s(s+1)(s+5)}$	14.5*b*,*c*	0.707 ($\omega_b=1.0$)	—	0.036	0.05 for X_3 inaccessible 3.40 for X_3 accessible	0.051
6. $p_1=-2$: $G(s)H_{eq}(s)=\dfrac{95.4(s^2+1.443s+1.0481)}{s(s+2)(s+5)}$	14.5*b*,*c*	0.683	3.39	—	—	—
7*a*. $p_2=-10$, X_3 accessible: $G(s)H_{eq}(s)=\dfrac{95.4(s^2+6.443s+1.0481)}{s(s+1)(s+10)}$	14.5*b*	0.160	77.4	—	—	—
7*b*. $p_2=-10$, X_3 inaccessible: $G(s)H_{eq}(s)=\dfrac{95.4(s^2+1.443s+1.0481)}{s(s+1)(s+10)}$	14.5*c*	0.682	3.54	—	—	—
8. A = 20: $G(s)H_{eq}(s)=\dfrac{190.8(s^2+1.443s+1.0481)}{s(s+1)(s+5)}$	14.5*b*,*c*	0.717	1.41	—	—	—

Table 14.3 Time-Response Data

Control system	$M_p(t)$	t_p, s	t_s, s	$\%\Delta M_p(t) = \dfrac{\lvert M_p - M_{po}\rvert}{M_{po}} 100$
1	1.048	7.2	9.8	—
2	1.000	—	16.4	4.8
3	1.000	—	15.0	4.8
4	1.189	4.1	10.15	13.45
5	1.044	4.45	6	—
6	1.035	4.5	6^-	0.86
7*a*	1.000	—	24	4.4
7*b*	1.046	4.6	6.5	0.181
8	1.046	4.4	5.6^-	0.192

The sensitivity function for each system, at $\omega = \omega_b$, is determined for a 100 percent variation of the plant poles $p_1 = -1$, $p_2 = -5$, and of the forward gain A, respectively. The term $\%|\Delta M|$ is defined as

$$\%|\Delta M| = \frac{|\mathbf{M}(j\omega_b)| - |\mathbf{M}_o(j\omega_b)|}{|M_o(j\omega_b)|} 100 \tag{14.27}$$

where $|\mathbf{M}_o(j\omega_b)|$ is the value of the control ratio with nominal values of the plant parameters, and $|\mathbf{M}(j\omega_b)|$ is the value with one parameter changed to the extreme value of its possible variation. The time-response data for systems 1 to 8 are given in Table 14.3 for a unit step input. Note that the response for the state-variable-feedback system (5 to 8, except 7*a*) is essentially unaffected by parameter variations.

The sensitivity function S_δ^M represents a measure of the change to be expected in the system performance for a change in a system parameter. This measure is borne out by comparing Tables 14.2 and 14.3. That is, in the frequency-domain plots of Fig. 14.6, a value $|\mathbf{S}_A^M(j\omega)| = 1.29$ at $\omega = \omega_b = 0.642$ is indicated for the conventional system, compared with 0.051 for the state-variable-feedback system. Also, the percent change $\%|\Delta M(j\omega)|$ for a doubling of the forward gain is 74 percent for the conventional system and only 1.41 percent for the state-variable-feedback system. These are consistent with the changes in the peak overshoot in the time domain of $\%\Delta M_p(t) = 13.45$ and 0.192 percent, respectively, as shown in Table 14.3. Thus, the magnitude of $\%\Delta M(j\omega)$ and $\%\Delta M_p(t)$ are consistent with the relative magnitudes of $\%|\mathbf{S}_A^M(j\omega_b)|$. Similar results are obtained for variations of the pole p_1.

An interesting result occurs when the pole $p_2 = -5$ is subject to variation. If the state X_3 (see Fig. 14.5*b*) is accessible and is fed back directly through k_3, the sensitivity $|\mathbf{S}_{p_2}^M(j\omega)|$ is much larger than for the conventional system, having a value of 3.40 at $\omega_b = 1.0$. However, if the equivalent feedback is obtained from X_1 through the fixed transfer function $[H_{eq}(s)]_{\delta_0 = p_2}$, Fig. 14.5*c*, the sensitivity is considerably reduced to $|S_{p_2}^M(j1)| = 0.050$, compared with 1.28 for the conventional system. If the components of the feedback unit are selected so that the transfer function $[H_{eq}(s)]$ is invariant, then the time-response characteristic for output feedback through $[H_{eq}(s)]$ is essentially unchanged when p_2 doubles in value (see Table 14.3) compared with the conventional system.

In analyzing Fig. 14.6, the values of the sensitivities must be considered over the entire passband ($0 \le \omega \le \omega_b$). In this passband the state-variable-feedback system has a much lower value of sensitivity than the conventional system. The performance of the state-variable-feedback systems is essentially unaffected by any variation in A, p_1, or p_2 (with restrictions) when the feedback coefficients remain fixed, provided that the forward gain satisfies the condition $K \ge K_{\min}$ (see Probs. 14.2 and 14.3). *The low-sensitivity state-variable-feedback systems considered in this section satisfy the following conditions:* the system's characteristic equation must have (1) β dominant roots ($< n$) which are located close to zeros of $H_{eq}(s)$ and (2) at least one nondominant root.

One approach for determining the nominal value for the variable parameter δ for a state-variable-feedback design is to determine **k** for the minimum, midrange, and maximum value of δ. Then evaluate and plot $|S_\delta^M(j\omega)|$ for the range of $0 \le \omega \le \omega_b$ for each of these values of δ. The value of δ that yields the minimum area under the curve of $|S_\delta^M(j\omega)|$ vs. ω is chosen as the value for δ_o.

The analysis of this section reveals that for the state-variable-feedback system any parameter variation between the control U and the state variable X_n has minimal effect on system performance. In order to minimize the effect on system performance for a parameter variation that occurs between the states X_n and X_1, the feedback signals should be moved to the right of the block containing the variable parameter (see Sec. 14.5), even if all the states are accessible. The output feedback system that incorporates an invariant $H_{eq}(s)$ has a response that is more invariant to parameter variations than a conventional system.

14.5 PARAMETER SENSITIVITY EXAMPLES

Example 1. The $(Y/R)_T$, Eq. (13.118), for Design Example 1 of Chap. 13 contains no nondominant pole. As seen from the root locus of Fig. 13.16 for this example, the system is highly sensitive to gain variations. Using Eq. (13.117) in Eq. (14.3) with $\delta = A$ yields $|S_\delta^M(j\omega)| = 1.84$ at $\omega_b = 1.41$.

Example 2. The $(Y/R)_T$ of Design Example 1 is modified to contain one nondominant pole; see Eq. (13.121). Using Eq. (13.121), with $p = -50$ and $\delta = K_G$, in Eq. (14.3) yields $|S_\delta^M(j\omega_b)| = 0.064$ at $\omega_b = 1.41$. Thus, the presence of a

TABLE 14.4 Time-Response and Gain-Sensitivity Data

Example	$(Y/R)_D$	M_P	t_p, s	t_s, s	K_1, s^{-1}	$\lvert S_{K_G}^M(j\omega_b)\rvert$ $\omega_b = 1.41$
1	$\dfrac{2(s+2)}{(s^2+2s+2)(s+2)}$	1.043	3.1	4.2	1.0	1.84
2	$\dfrac{100(s+2)}{(s^2+2s+2)(s+2)(s+50)}$	1.043	3.2	4.2	0.98	0.064
3	$\dfrac{100(s+1.4)(s+2)}{(s^2+2s+2)(s+1)(s+2)(s+70)}$	1.008	3.92	2.94	0.98	0.087

nondominant pole drastically reduces $\left|S_\delta^M(j\omega_b)\right|$, which is indicative of the reduction of the sensitivity magnitude for the frequency range $0 \le \omega \le \omega_b$.

Example 3. The $(Y/R)_T$ of Design Example 2 is modified to improve the value of t_s; see Eq. (13.127). Using Eq. (13.127), with $p = -70$ and $\delta = K_G$, in Eq. (14.3) yields $\left|S_\delta^M(j\omega_b)\right| = 0.087$ at $\omega_b = 1.41$.

The data in Table 14.4 summarize the results of these three examples and show that the presence of at least one nondominant pole in $(Y/R)_T$ drastically reduces the system's sensitivity to gain variation while achieving the desired time-response characteristics.

14.6 INACCESSIBLE STATES [5]

In order to achieve the full benefit of a state-variable-feedback system, all the states must be accessible. Although, in general, this requirement may not be satisfied, the design procedures presented in this chapter are still valid. That is, on the basis that all states are accessible, the required value of the state-feedback gain vector **k** which achieves the desired $Y(s)/R(s)$ is computed. Then, for states that are not accessible, the corresponding k_i blocks are moved by block-diagram manipulation techniques to states that are accessible. As a result of these manipulations the full benefits of state-variable feedback may not be achieved (low sensitivity S_δ^M and a completely stable system), as described in Sec. 14.4. For the extreme case when the output $y = x_1$ is the only accessible state, the block-diagram manipulation to the G-equivalent (see Fig. 14.8c) reduces to the Guillemin-Truxal design. More sophisticated methods for reconstructing (estimating) the values of the missing states do exist (Luenberger observer theory). They permit the accessible and the reconstructed (inaccessible) states all to be fed back through the feedback coefficients, thus eliminating the need for block-diagram manipulations and maintaining the full benefits of the state-variable feedback-designed

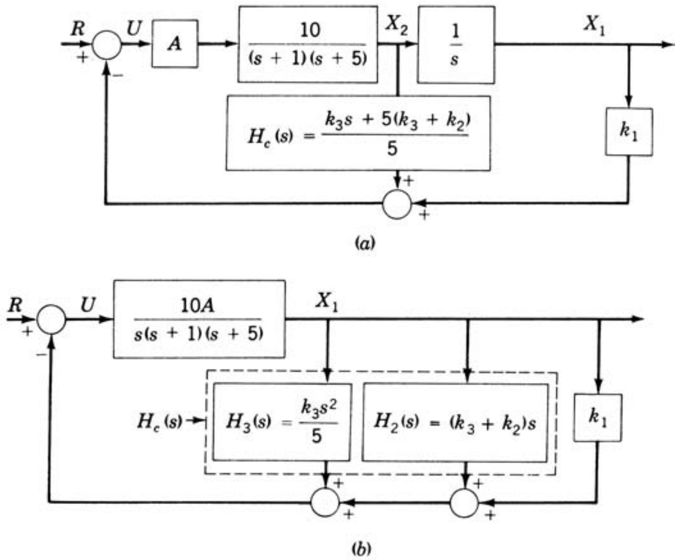


FIGURE 14.7 Minor-loop compensation (manipulation to the right) of the control system of Fig. 14.5*b* with $p_1 = -1$ and $p_2 = -2$.

system [5]. Even when all states are theoretically accessible, design and/or cost of instrumentation may make it prohibitive to measure some of the states.

Various block diagram manipulations of Fig. 14.5*b*, depending on which states are accessible, are shown in Figs. 14.7 and 14.8. The availability of practical state-variable sensors determines which physical implementation is most feasible. The active minor-loop compensation of Fig. 14.7*b* results in the requirement for first- and second-order derivative action in the minor loops; thus $H_c(s)$ is an improper function. It is possible to add poles to $H_c(s)$ that are far to the left in order to achieve a realizable $H_c(s)$.

The compensators required for each case represented in Figs. 14.7 and 14.8, using the values of $k_1 = 1$, $k_2 = -3.393$, $k_3 = 4.77$, and $A = 10$ determined in Sec. 13.9, are given in Table 14.5, which also gives the sensitivity function $|\mathbf{S}_{K_G}^M(j\omega_b)|$ for each case at $\omega_b = 1.0$ and $M = 0.707$. As noted in Table 14.5, the low sensitivity of the output response to gain variations is maintained as long as invariant feedback compensators $H_c(s)$ are used. In other words, the block-diagram manipulations that result in only *minor-loop* compensation, i.e., using only *feed-back* compensators $H_c(s)$, have the same $H_{eq}(s)$ as before the manipulations are performed. Therefore, the same stability and sensitivity characteristics are maintained for the manipulated as for the nonmanipulated systems. When the shifting is to the left, to yield a resulting cascade

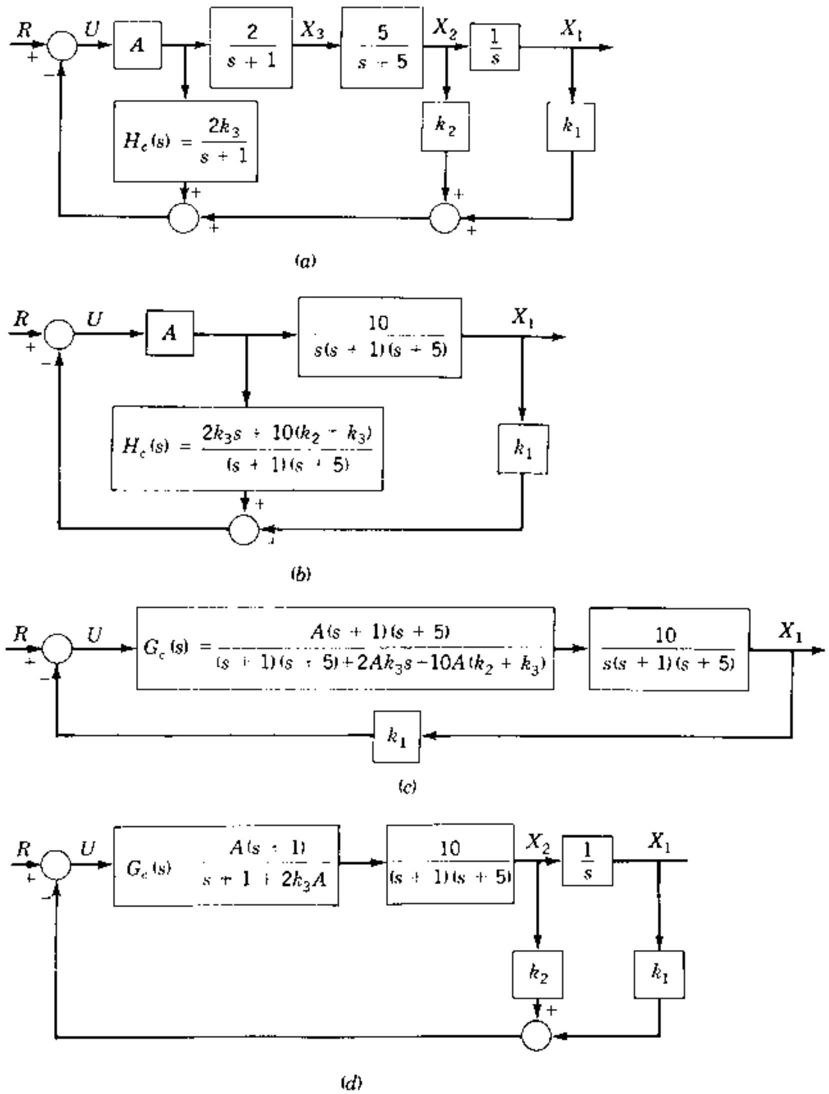


FIGURE 14.8 Manipulation of state feedback to the left for the control system of Fig. 14.5*b*.

Table 14.5 Compensators and Sensitivity Functions for Control Systems of Figs. 14.7 and 14.8

Figure	Compensator	$\lvert S_A^M(j\omega_b)\rvert$
14.7*a*	$H_c(s) = 0.954s + 1.377$	0.0508
14.7*b*	$H_c(s) = 0.954s^2 + 1.377s$	0.0508
14.8*a*	$H_c(s) = \dfrac{9.54}{s+1}$	0.0508
14.8*b*	$H_c(s) = \dfrac{9.54(s+1.443)}{(s+1)(s+5)}$	0.0508
14.8*c*	$G_c(s) = \dfrac{10(s+1)(s+5)}{(s+1.43)(s+99.98)}$	0.51
14.8*d*	$G_c(s) = \dfrac{10(s+1)}{s+96.4}$	0.51

compensator $G_c(s)$, the low sensitivity to parameter variation is not achieved. This sensitivity characteristic is borne out by analyzing the root locus of Figs. 13.10*b* and 14.9. Figure 13.10*b* is the root locus representing the block-diagram manipulations using only $H_c(s)$ networks. Figure 14.9 shows the root locus representing the block-diagram manipulations using cascade $G_c(s)$ networks. In Fig. 13.10*b*, an increase in the gain does not produce any appreciable change in the closed-loop system response. In contrast, any gain variation in the systems represented by Fig. 14.9 greatly affects the closed-loop system response.

Note that the manipulated system of Fig. 14.8*d*, using both cascade $G_c(s)$ and minor-loop $H_c(s) = k_2$ compensators, yields the same relative sensitivity as for the system of Fig. 14.8*c* which uses only cascade compensation. In general, for a system using both $G_c(s)$ and $H_c(s)$ compensators, it may be possible to achieve a sensitivity that is between that of a system using only minor-loop compensation and that of one using only cascade compensation.

The technique presented in this section for implementing a system having inaccessible states is rather simple and straightforward. It is very satisfactory for systems that are noise-free and time-invariant. For example, if in Fig. 14.5*b* the state X_3 is inaccessible and X_2 contains noise, it is best to shift the k_3 block to the left. This assumes that the amplifier output signal has minimal noise content. If all signals contain noise, the system performance is degraded from the desired performance. This technique is also satisfactory under parameter variations and/or if the poles and zeros of the transfer function $G_x(s)$ are not known exactly when operating under high forward gain.

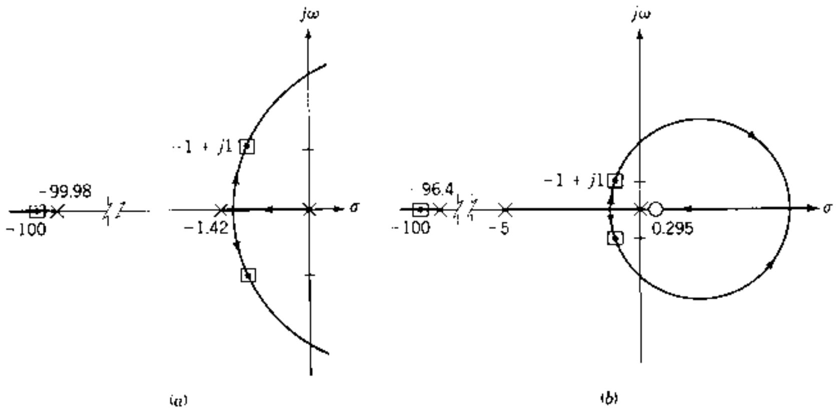


FIGURE 14.9 Root locus for (*a*) Fig. 14.8*c*, and (*b*) Fig. 14.8*d*.

14.7 STATE-SPACE TRAJECTORIES [6]

The definition of stability for a linear time-invariant (LTI) system is an easy concept to understand. The complete response to any input contains a particular solution that has the same form as the input and a complementary solution containing terms of the form $Ae^{\lambda_i t}$. When the eigenvalues λ_i have negative real parts, the transients associated with the complementary solution decay with time, and the response is called stable. When the roots have positive real parts, the transients increase without bound and the system is called unstable. It is necessary to extend the concept of stability to nonlinear systems. This section introduces the properties of state-space trajectories, which give a visual picture of transient and steady-state performance. The kinds of singular points and their associated stability are categorized for both linear and nonlinear systems.

The *state space* is defined as the n-dimensional space in which the components of the state vector represent its coordinate axes. The unforced response of a system having n state variables, released from any initial point $\mathbf{x}(t_0)$, traces a curve or trajectory in an n-dimensional state space. Time t is an implicit function along the trajectory. When the state variables are represented by phase variables, the state space may also be called a phase space. While it is impossible to visualize a space with more than three dimensions, the concept of a state space is nonetheless very useful in systems analysis. The behavior of second-order systems is conveniently viewed in the state space because it is two-dimensional and is in fact a state (or phase) plane. It is easy to obtain a graphical or geometrical representation and interpretation of second-order system behavior in the state plane. The concepts developed for the state plane can be extrapolated to the state space of higher-order systems. Examples of a number of second-order systems are presented in this section. These concepts are extended qualitatively to third-order systems. The symmetry that results in the state-plane trajectories is demonstrated for the cases when the A matrix is converted to normal (canonical) or modified normal (modified canonical) form. The second-order examples that are considered include the cases where the eigenvalues are (1) negative real and unequal (overdamped), (2) negative real and equal (critically damped), (3) imaginary (oscillatory), and (4) complex with negative real parts (underdamped). The system that is studied has no forcing function and is represented by the state equation

$$\dot{\mathbf{x}} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \mathbf{x} \tag{14.28}$$

Because there is no forcing function, the response results from energy stored in the system as represented by initial conditions.

Example 1. Overdamped response. Consider the following example with the given initial conditions:

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 3 \\ -1 & -4 \end{bmatrix} \mathbf{x} \qquad \mathbf{x}(0) = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \tag{14.29}$$

Because one state equation is $\dot{x}_1 = 3x_2$, then x_2 is a phase variable. Therefore, the resulting state plane is called a *phase plane.*

The state transition matrix $\Phi(t)$ is obtained by using the Sylvester expansion theorem (see Sec. 3.14) or any other convenient method. The system response is overdamped, with eigenvalues $\lambda_1 = -1$ and $\lambda_2 = -3$. For the given initial conditions the system response is

$$\mathbf{x}(t) = \mathbf{\Phi}(t)\mathbf{x}(0) = \begin{bmatrix} 3e^{-t} - 3e^{-3t} \\ -e^{-t} + 3e^{-3t} \end{bmatrix} \tag{14.30}$$

The trajectory represented by this equation can be plotted in the phase plane x_2 versus x_1. A family of trajectories in the phase plane for a number of initial conditions is drawn in Fig. 14.10. Such a family of trajectories is called a *phase portrait*. The arrows show the direction of the states along the trajectories for increasing time. Motion of the point on a trajectory is in the clockwise direction about the origin because $\dot{x}_1 = 3x_2$. Thus, when x_2 is positive, x_1 must be increasing in value, and when x_2 is negative, x_1 must be decreasing in value. The value of N in Fig. 14.10 is the slope of the trajectory at each point in the phase plane, i.e., $N = dx_2/dx_1$, as described in Eq. (14.32). The rate of change of x_1 is zero at the points where the trajectories cross the x_1 axis. Therefore, the trajectories always cross the x_1 axis in a perpendicular direction.

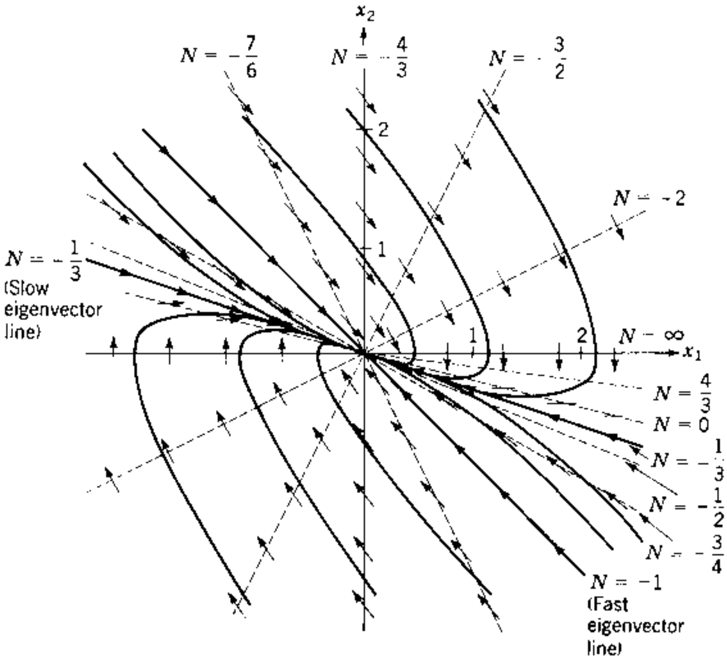


FIGURE 14.10 Phase-plane portrait for Eq. (14.30).

One method of drawing the phase trajectory is to insert values of time t into the solution $\mathbf{x}(t)$ and to plot the results. For the example above the solutions are $x_1 = 3e^{-t} - 3e^{-3t}$ and $x_2 = -e^{-t} + 3e^{-3t}$. Another approach is to eliminate t from these equations to obtain an analytical expression for the trajectory. This can be done by first solving for e^{-t} and e^{-3t}:

$$e^{-t} = \frac{x_1 + x_2}{2} \qquad \text{and} \qquad \mathrm{e}^{-3t} = \frac{x_1 + 3x_2}{6}$$

Raising e^{-t} to the third power, e^{-3t} to the first power, and equating the results yields

$$\left(\frac{x_1 + x_2}{2}\right)^3 = \left(\frac{x_1 + 3x_2}{6}\right)^1 \tag{14.31}$$

This equation is awkward to use and applies only for the specified initial conditions. Further, it is difficult to extend this procedure to nonlinear systems.

A practical graphical method for obtaining trajectories in the state plane is the *method of isoclines.* The previous example is used to illustrate this method. Taking the ratio of the state equations of this example yields

$$\frac{\dot{x}_2}{\dot{x}_1} = \frac{dx_2/dt}{dx_1/dt} = \frac{dx_2}{dx_1} = \frac{-x_1 - 4x_2}{3x_2} = N \tag{14.32}$$

This equation represents the slope of the trajectory passing through any point in the state plane. The integration of this equation traces the trajectory starting at any initial condition. The method of isoclines is a graphical technique for performing this integration. When the slope of the trajectory in Eq. (14.32) is selected at any specific value N, the result is the equation of a straight line having a slope m and represented by

$$x_2 = -\frac{1}{4+3N} x_1 = m x_1 \tag{14.33}$$

This equation describes a family of curves that are called *isoclines.* They have the property that all trajectories crossing a particular isocline have the same slope at the crossing points. For LTI second-order systems the isoclines are all straight lines. A trajectory can be sketched as shown in Fig. 14.10.

The system represented by Example 1 is stable, as indicated by the negative eigenvalues. Because there is no forcing function, the response is due to the energy initially stored in the system. As this energy is dissipated, the response approaches equilibrium at $x_1 = x_2 = 0$. All the trajectories approach and terminate at this point. The slope of the trajectory going through this equilibrium point is therefore indeterminate; that is, $N = dx_2/dx_1 = 0/0$, representing a *singularity* point. This indeterminacy can be used to locate the equilibrium point by letting the slope of the trajectory in the state plane be $N = dx_2/dx_1 = 0/0$. Applying this condition to Eq. (14.32) yields the two equations $x_2 = 0$ and $-x_1 - 4x_2 = 0$. This confirms that the equilibrium point $x_1 = x_2 = 0$. For the overdamped response this equilibrium point is called a *node.* The procedure can be extended to all systems of any order, even when they are nonlinear. This is done by solving the vector equation $\dot{\mathbf{x}} = 0$ for the equilibrium point(s). For an underdamped system the equilibrium point is called a focus.

14.8 LINEARIZATION (JACOBIAN MATRIX) [6,7]

The definition of stability for a linear time-invariant (LTI) system is an easy concept to understand. The complete response has two components; the particular and complementary solutions. The particular solution that has the same form as the input and the complementary solution contains terms of the form $Ae^{\lambda_i t}$. When the eigenvalues λ_i have negative real parts, the transients associated with the complementary solution decay with time, and the response is called stable. When the roots have positive real parts, the transients increase without bound and the system is called unstable. Some nonlinear systems design techniques are based upon obtaining a set of J LTI plants about specified operating points. A method for obtaining a set of LTI plants for a nonlinear system, about J operating (equilibrium) points, is presented in this section. This method permits the determination of stability condition in the neighborhood of each equilibrium point.

A linear system with no forcing function (an autonomous system) and with $|\mathbf{A}| \neq \mathbf{0}$ has only one equilibrium point, $\mathbf{x}_0$. All the examples in this text have equilibrium points at the origin $\mathbf{x}_0 = \mathbf{0}$. A nonlinear system, on the other hand, may have more than one equilibrium point. This is easily illustrated by considering the unity-feedback angular position-control system shown in Fig. 14.11. The feedback action is provided by synchros that generate the actuating signal $e = \sin(\theta_i - \theta_o)$. With no input, $\theta_i = 0$, the differential equation of the system is

$$\ddot{\theta}_o + a\dot{\theta}_o + K \sin\theta_o = 0 \tag{14.34}$$

This is obviously a nonlinear differential equation because of the term $\sin\theta_o$. With the phase variables $x_1 = \theta_o$ and $x_2 = \dot{x}_1 = \dot{\theta}_o$, the corresponding state equations are

$$\dot{x}_1 = x_2 \qquad \dot{x}_2 = -K \sin x_1 - a x_2 \tag{14.35}$$

The slope of the trajectories in the phase plane is obtained from

$$N = \frac{\dot{x}_2}{\dot{x}_1} = \frac{-K \sin x_1 - a x_2}{x_2} \tag{14.36}$$

In the general case the unforced nonlinear state equation is

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) \tag{14.37}$$

Because equilibrium points x_0 exist $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}_0) = \mathbf{0}$, the singularities are $x_2 = 0$ and $x_2 = k\pi$, where k is an integer. The system therefore has multiple equilibrium points.

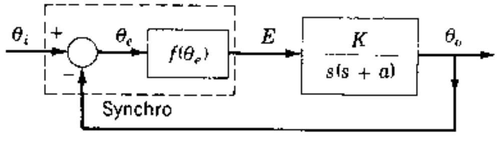


FIGURE 14.11 A nonlinear feedback control system.

In a small neighborhood about each of the equilibrium points, a nonlinear system behaves like a linear system. The states can therefore be written as $\mathbf{x} = \mathbf{x}_0 + \mathbf{x}^*$, where $\mathbf{x}^*$ represents the perturbation or *state deviation* from the equilibrium point $\mathbf{x}_0$. Each of the elements of $\mathbf{f}(\mathbf{x})$ can be expanded in a Taylor series about one of the equilibrium points $\mathbf{x}_0$. Assuming that $\mathbf{x}^*$ is restricted to a small neighborhood of the equilibrium point, the higher-order terms in the Taylor series may be neglected. Thus, the resulting linear variational state equation is

$$\dot{\mathbf{x}}^* = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}_{\mathbf{x}=\mathbf{x}_0} \mathbf{x}^* = \mathbf{J}_x \mathbf{x}^* \tag{14.38}$$

where f_i is the ith row of $\mathbf{f}(\mathbf{x})$ and $\mathbf{J}_x = \partial \mathbf{f}/\partial \mathbf{x}^T$ is called the *Jacobian matrix* and is evaluated at $\mathbf{x}_0$.

For the system of Eq. (14.35), the motion about the equilibrium point $x_1 = x_2 = 0$ is represented by

$$\dot{\mathbf{x}}^* = \begin{bmatrix} \dot{\mathbf{x}}_1^* \\ \dot{\mathbf{x}}_2^* \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -K & -a \end{bmatrix} \begin{bmatrix} \mathbf{x}_1^* \\ \mathbf{x}_2^* \end{bmatrix} = \mathbf{J}_x \mathbf{x}^* \tag{14.39}$$

For these linearized equations the eigenvalues are $\lambda_{1,2} = -a/2 \pm \sqrt{(a/2)^2 - K}$. This equilibrium point is stable, depending upon the magnitudes of a and K.

For motion about the equilibrium point $x_1 = \pi$, $x_2 = 0$, the state equations are

$$\dot{\mathbf{x}}^* = \begin{bmatrix} 0 & 1 \\ K & -a \end{bmatrix} \mathbf{x}^* \tag{14.40}$$

The eigenvalues of this $\mathbf{J}_x$ are $\lambda_{1,2} = -a/2 \pm \sqrt{(a/2)^2 + K}$. Thus, one eigenvalue is positive and the other is negative. Therefore, the equilibrium point is an unstable saddle point. The motion around the saddle point is considered unstable because every point on all trajectories, except on the two-separatrices, moves away from this equilibrium point. A phase-plane portrait for this system can be obtained by the method of isoclines. From Eq. (14.36), the isocline equation is

$$x_2 = \frac{-K \sin x_1}{N + a} \tag{14.41}$$

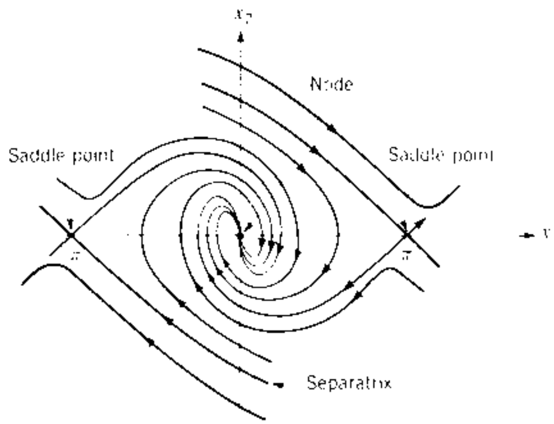


FIGURE 14.12 Phase portrait for Eq. (14.34).

The phase portrait is shown in Fig. 14.12. Note that the linearized equations are applicable only in the neighborhood of the singular points. Thus, they describe stability *in the small.*

When an input $\mathbf{u}$ is present, the nonlinear state equation is

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x},\mathbf{u}) \tag{14.42}$$

Its equilibrium point $\mathbf{x}_0$, when $\mathbf{u} = \mathbf{u}_0$ is a constant, is determined by letting $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}_0,\mathbf{u}_0) = \mathbf{0}$. Then the linearized variational state equation describing the variation from the equilibrium point is obtained by using only the linear terms from the Taylor series expansion. For variations in $\mathbf{u}$, appearing as $\mathbf{u}^*$ in the input $\mathbf{u} = \mathbf{u}_0 + \mathbf{u}^*$, the linearized state equation is

$$\mathbf{x}^* = \begin{bmatrix} \dfrac{\partial f_1(\mathbf{x},\mathbf{u})}{\partial x_1} & \dfrac{\partial f_1(\mathbf{x},\mathbf{u})}{\partial x_2} & \cdots & \dfrac{\partial f_1(\mathbf{x},\mathbf{u})}{\partial x_n} \\ \dfrac{\partial f_2(\mathbf{x},\mathbf{u})}{\partial x_1} & \dfrac{\partial f_2(\mathbf{x},\mathbf{u})}{\partial x_2} & \cdots & \dfrac{\partial f_2(\mathbf{x},\mathbf{u})}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial f_n(\mathbf{x},\mathbf{u})}{\partial x_1} & \dfrac{\partial f_n(\mathbf{x},\mathbf{u})}{\partial x_2} & \cdots & \dfrac{\partial f_n(\mathbf{x},\mathbf{u})}{\partial x_n} \end{bmatrix}_{\substack{\mathbf{x}=\mathbf{x}_0 \\ \mathbf{u}=\mathbf{u}_0}} \mathbf{x}^*$$

$$+ \begin{bmatrix} \dfrac{\partial f_1(\mathbf{x},\mathbf{u})}{\partial u_1} & \cdots & \dfrac{\partial f_1(\mathbf{x},\mathbf{u})}{\partial u_r} \\ \cdots\cdots\cdots & \cdots\cdots & \cdots\cdots \\ \dfrac{\partial f_n(\mathbf{x},\mathbf{u})}{\partial u_1} & \cdots & \dfrac{\partial f_n(\mathbf{x},\mathbf{u})}{\partial u_r} \end{bmatrix}_{\substack{\mathbf{x}=\mathbf{x}_0 \\ \mathbf{u}=\mathbf{u}_0}} \mathbf{u}^* = \mathbf{J}_x\mathbf{x}^* + \mathbf{J}_u\mathbf{u}^* \tag{14.43}$$

In analyzing the system performance in the vicinity of equilibrium, it is usually convenient to translate the origin of the state space to that point. This is done by inserting $\mathbf{x} = \mathbf{x}_0 + \mathbf{x}^*$ into the original equations. With the origin at $\mathbf{x}_0$, the variations from this point are described by $\mathbf{x}^*$.

The characteristic equation describing the motion about the equilibrium point can be obtained from the linearized state equations. For a second-order system it has the form

$$s^2 + ps + q = (s - \lambda_1)(s - \lambda_2) = 0 \tag{14.44}$$

The eigenvalues λ_1 and λ_2 determine whether the singular point produces an overdamped or an underdamped oscillatory response (a node, a focus, or a saddle). The eigenvalues also determine whether the transient response about the equilibrium point is stable or unstable.

14.9 SUMMARY

This chapter presents an analysis of the sensitivity of the system output to variations of system parameters. This is accomplished by defining a sensitivity function S_δ^M, where M is the control ratio and δ is the parameter that varies. A sensitivity analysis for a specific example system shows that a state-variable-feedback system has a much lower sensitivity than a conventional output feedback design. When there are inaccessible states, the states can be regenerated from the output, provided that the system is completely observable. For SISO systems this requires derivative action, accentuating noise that may be present in the output. This can be avoided by designing an observer [8–10]. The properties of state-space trajectories and singular points are introduced. Most of the examples used are second-order systems because they clearly and easily demonstrate the important characteristics, such as stability and the kinds of singular points. Both linear and nonlinear differential equations are considered. The Jacobian matrix is used to represent nonlinear systems by approximate linear equations in the vicinity of their singular points. The important property is demonstrated that a linear equation has only one singular equilibrium point, whereas a nonlinear equation may have more than one singular equilibrium point.

REFERENCES

1. Bode, H. W.: *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, 1945.
2. Houpis, C. H., and S. J. Rasmussen: *Quantitative Feedback Theory*, Marcel Dekker, New York, 1999.

3. "Sensitivity and Modal Response for Single-Loop and Multiloop Systems," *Flight Control Lab. ASD, AFSC Tech. Doc. Rep.* ASD-TDR-62-812, Wright-Patterson Air Force Base, Ohio, January 1963.
4. Cruz, J. B., Jr.: Feedback Systems, McGraw-Hill, New York, 1972.
5. D'Azzo, J. J., and C. H. Houpis: *Linear Control System Analysis & Design: Conventional and Modern*, McGraw-Hill, New York, 3rd ed., 1988.
6. Andronow, A. A., et al.: *Theory of Oscillations,* Addison-Wesley, Reading, Mass., 1966.
7. Csaki, F.: *Modern Control Theories*, Akademiai, Kiado, Budapest, 1972.
8. Luenberger, D. G.: "Observing the State of a Linear System," *IEEE Trans. Mil. Electron.*, vol. 8, pp. 74–80, 1964.
9. Luenberger, D. G.: "Observers for Multivariable Systems," *IEEE Trans. Autom. Control*, vol. AC-11, pp. 190–197, 1966.
10. Luenberger, D. G.: "An Introduction to Observer," *IEEE Trans. Autom. Control*, vol. AC-16, pp. 569–602, 1971.

15

Sampled-Data Control Systems

15.1 INTRODUCTION [1–5]

The methods presented in the previous chapters deal with the fundamental properties of continuous systems. However, digital computers are available not only for the design of control systems but also to perform the control function. Digital computers and microprocessors are used in many control systems. The size and cost advantages associated with the microcomputer make its use as a controller economical and practical. Using such a digital processor with a plant that operates in the continuous-time domain requires that its input signal be discrete, which requires the sampling of the signals used by the controller. Such sampling may be an inherent characteristic of the system. For example, a radar tracking system supplies information on an airplane's position and motion to a digital processor at discrete periods of time. This information is therefore available as a succession of data points. When there is no inherent sampling, an analog-to-digital (A/D) converter must be incorporated in a digital or sampled-data (S-D) control system. The sampling process can be performed at a constant rate or at a variable rate, or it may be random. The discussion in this chapter is based on a constant-rate sampling. The output of the controller must then be converted from discrete form into an analog signal by a digital-to-analog (D/A) converter. A functional block diagram of such a system is shown in Fig. 15.1, in which the signals e^* and u^*

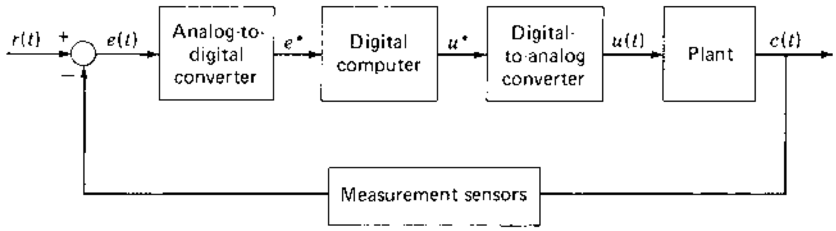


FIGURE 15.1 Control system incorporating a digital computer.

are in discrete form, u is piecewise continuous, and the remaining signals are continuous. The notation e^* and u^* denotes that these signals are sampled at specified time intervals and are therefore in discrete form. Systems that include a digital computer are known as *digital control systems.* The synthesis techniques for such systems are based on *representing the entire system* as an equivalent sampled or discrete or an equivalent *pseudo-continuous-time* (PCT) system.

There are various approaches that may be used in analyzing the stability and time-response characteristics of S-D systems. These approaches may be divided into two distinct categories: (1) *direct* (DIR) and (2) *digitization* (DIG) or *discrete* digital control analysis techniques. In the first (DIR) category the analysis is performed entirely in the discrete domain (z plane). In the second (DIG) category the analysis and synthesis of the sampled-data system is carried out entirely by transformation to the w' plane or, by use of the Padé approximation (see Ref. 1, App. C), entirely in the s plane. The use of the Padé approximation, along with the information provided by a Fourier analysis of the sampled signal, results in the modeling of a sampled-data control system by a PCT control system. The w' plane analysis is not covered in this text but the reader is referred to Ref. 1 where an overview of the DIR and DIG methods is presented.

The analysis and design of sampled-data control systems, as discussed in this chapter, is expedited by the use of CAD packages, such as MATLAB or TOTAL-PC (see Sec. 10.6 and Appendixes C and D).

15.2 SAMPLING

Sampling may occur at one or more places in a system. The sampling operation is represented in a block diagram by the symbol for a switch. Figure 15.2 shows a system with sampling of the actuating signal. Note that the output $c(t)$ is a continuous function of time. A fictitious sampler that produces the mathematical function $c^*(t)$ (see Sec. 15.6) is also shown. The sampling

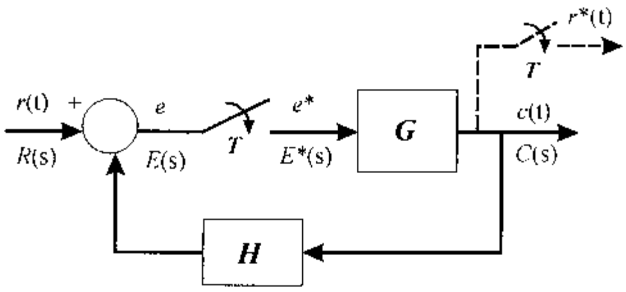


FIGURE 15.2 Block diagram of a sampled-data control system involving the sampling of the actuating signal.

process can be considered a modulation process in which a pulse train $p(t)$, with magnitude $1/\gamma$ and period T, multiplies a continuous-time function $f(t)$ and produces the sampled function $f_p^*(t)$. This is represented by

$$f_p^*(t) = p(t)f(t) \tag{15.1}$$

These quantities are shown in Fig. 15.3. A Fourier-series expansion of $p(t)$ is

$$p(t) = \frac{1}{\gamma}\sum_{n=-\infty}^{+\infty} C_n e^{jn\omega_s t} \tag{15.2}$$

where the sampling frequency is $\omega_s = 2\pi/T$ and the Fourier coefficients C_n are given by

$$C_n = \frac{1}{T}\int_0^T p(t)e^{-j\omega_s t}\,dt = \frac{1}{T}\frac{\sin(n\omega_s\gamma/2)}{n\omega_s\gamma/2}e^{-jn\omega_s\gamma/2} \tag{15.3}$$

The sampled function is therefore

$$f_p^*(t) = \sum_{n=-\infty}^{+\infty} C_n f(t)e^{jn\omega_s t} \tag{15.4}$$

If the Fourier transform of the continuous function $f(t)$ is $F(j\omega)$, the Fourier transform of the sampled function is [2,3,5]

$$F_p^*(j\omega) = \sum_{n=-\infty}^{+\infty} C_n F(j\omega + jn\omega_s) \tag{15.5}$$

A comparison of the Fourier spectra of the continuous and sampled functions is shown in Fig. 15.4. It is seen that the sampling process produces a fundamental spectrum similar in shape to that of the continuous function. It also

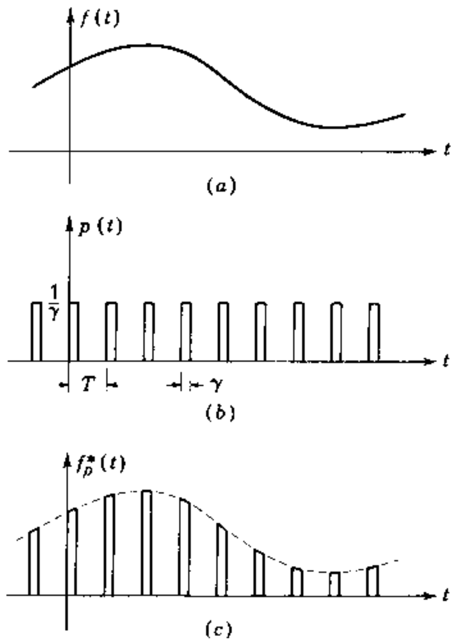


FIGURE 15.3 (*a*) Continuous function $f(t)$; (*b*) sampling train $p(t)$; (*c*) sampled function $f_p^*(t)$.

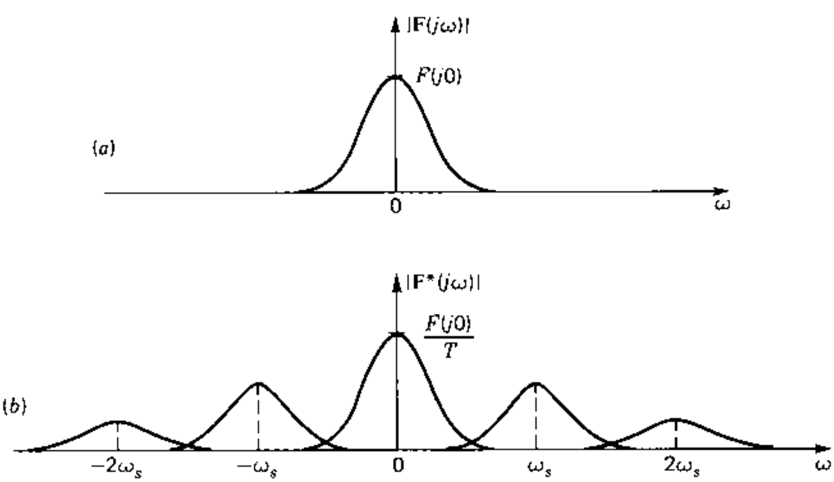


FIGURE 15.4 Frequency spectra for (*a*) a continuous function $f(t)$ and (*b*) a pulse-sampled function $f_p^*(t)$.

produces a succession of spurious complementary spectra that are shifted periodically by a frequency separation $n\omega_s$.

If the sampling frequency is sufficiently high, there is very little overlap between the fundamental and complementary frequency spectra. In that case a low-pass filter could extract the spectrum of the continuous input signal by attenuating the spurious higher-frequency spectra. The forward transfer function of a control system generally has a low-pass characteristic, so the system responds with more or less accuracy to the continuous signal.

15.3 IDEAL SAMPLING

To simplify the mathematical analysis of sampled-data control systems, the concept of an ideal sampler (impulse sampling) is introduced. If the duration of γ of the sampling pulse is much less than the sampling time T and much smaller than the smallest time constant of $e(t)$, then the pulse train $p(t)$ can be represented by an ideal sampler. Since the area of each pulse of $p(t)$ has unit value, the impulse train also has a magnitude of unity. The ideal sampler produces the impulse train $\delta_T(t)$, which is shown in Fig. 15.5 and is represented by

$$\delta_T(t) = \sum_{k=-\infty}^{+\infty} \delta(t - kT) \tag{15.6}$$

where $\delta(t - kT)$ is the unit impulse which occurs at $t = T$. The frequency spectrum of the function that is sampled by the ideal impulse train is shown in Fig. 15.6. It is similar to that shown in Fig. 15.4, except that the complementary spectra have the same amplitude as the fundamental spectrum [1]. Because the forward transfer function of a control system attenuates the higher frequencies, the overall system response is essentially the same with the idealized impulse sampling as with the actual pulse sampling. Note in Figs. 15.4 and 15.6 that sampling reduces the amplitude of the spectrum by the factor $1/T$. The use of impulse sampling simplifies the mathematical analysis of sampled systems and is therefore used extensively to represent the sampling process.

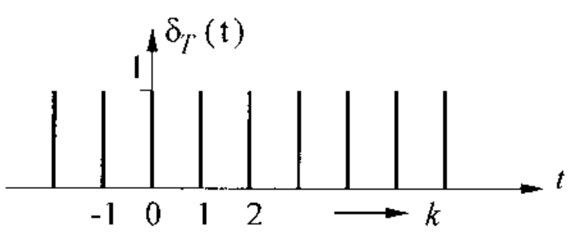


FIGURE 15.5 An impulse train.

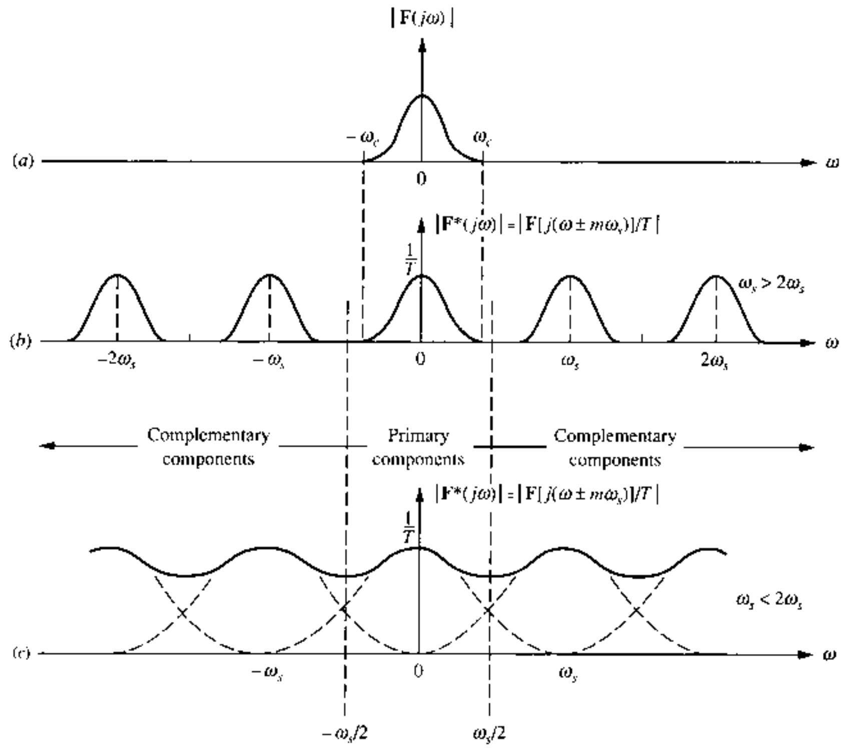


FIGURE 15.6 Frequency spectra for (*a*) a continuous function $f(t)$; (*b*) an impulse-sampled function $f^*(t)$, when $\omega_s > 2\omega_c$ [1]; (*c*) an impulse-sampled function $f^*(t)$, when $\omega < 2\omega_c$ [1].

When $f(t) = 0$ for $t < 0$, the impulse sequence is

$$f^*(t) = f(t)\,\delta_T(t) = \sum_{k=0}^{+\infty} f(t)\,\delta_T(t - kT) \tag{15.7}$$

and the Laplace transform is given by the infinite series

$$F^*(s) = \sum_{k=0}^{+\infty} f(kT)e^{-kTs} \tag{15.8}$$

where $f(kT)$ represents the function $f(t)$ at the sampling times kT. Because the expression $F^*(s)$ contains the term e^{Ts}, it is not an algebraic expression but a transcendental one. Therefore, a change of variable is made:

$$z \equiv e^{Ts} \tag{15.9}$$

where $s = (1/T)\ln z$. Equation (15.8) can now be written as

$$[F^*(s)]_{s=(1/T)\ln z} = F(z) = \sum_{k=0}^{+\infty} f(kt)z^{-k} \tag{15.10}$$

For functions $f(t)$ which have zero value for $t < 0$, the one-sided $\mathcal{Z}$ transform is the infinite series

$$F(z) = f(0) + f(T)z^{-1} + f(2T)z^{-2} + \cdots \tag{15.11}$$

If the Laplace transform of $f(t)$ is a rational function, it is possible to write $F^*(s)$ in closed form. When the degree of the denominator of $F(s)$ is at least 2 higher than the degree of the numerator, the closed form can be obtained from

$$F^*(s) = \sum_{\text{at poles of } F(p)} \left\{ \text{residues of } \left[F(p) \frac{1}{1 - e^{-(s-p)T}} \right] \right\} \tag{15.12}$$

where $F(p)$ is the Laplace transform of $f(t)$ with s replaced by p. The $\mathcal{Z}$ transform in closed form may be obtained from Eq. (15.12) and is written as [4]

$$F(z) = \hat{F}(z) + \beta = \sum_{\text{at poles of } F(p)} \left\{ \text{residues of} \left[F(p) \frac{1}{1 - e^{pT}z^{-1}} \right] \right\} + \beta \tag{15.13}$$

where

$$\beta = \lim_{s \to \infty} sF(x) - \lim_{z \to \infty} \hat{F}(z) \tag{15.14}$$

The value of β given by Eq. (15.14) ensures that the initial value $f(0)$ represented by $F(s)$ and $F(z)$ are identical. *$F(z)$ is called the $\mathcal{Z}$ transform of $f^*(t)$*. The starred and z forms of the *impulse response transfer function* are easily obtained for the ordinary Laplace transfer function.

Example 15.1. Consider the transfer function

$$G(s) = \frac{K}{s(s+a)} \tag{15.15}$$

The corresponding impulse transfer function in the s domain is

$$G^*(s) = \frac{Ke^{-sT}(1 - e^{-at})}{a(1 - e^{-sT})(1 - e^{-(s+a)T})} \tag{15.16}$$

and in the z domain,

$$G(z) = \frac{(1 - e^{-aT})Kz^{-1}}{a(1 - z^{-1})(1 - e^{-aT}z^{-1})} \tag{15.17}$$

The transformation of Eq. (15.16) into the z domain, as given by Eq. (15.17), results in the primary strip $(-j\omega_s/2 < j\omega < j\omega_s/2)$ and the infinite number of complementary strips $(\ldots - j5\omega_s/2 < -j\omega < -j3\omega_s/2,\ \ -j3\omega_s/2 < -j\omega < -j\omega_s/2$, and $j\omega_s/2 < j\omega < j3\omega_s/2,\ j3\omega_s/2 < j\omega < j5\omega_s/2, \ldots)$ in the s plane. The pole for $k = 0$ is said to lie in the *primary strip* in the s plane and the remaining poles, for $k \neq 0$, are said to lie in the *complementary strips* in the s plane. Note that they are uniformly spaced with respect to their imaginary parts with a separation $j\omega_s$. The poles of $G^*(s)$ in Eq. (15.16) are infinite in number. These poles exist at $s = jk\omega_s$ and $s = -a + jk\omega_s$ for all values of $-\infty < k < +\infty$. The complementary strips are transformed into the same (overlapping) portions in the z plane. Thus, by contrast, there are just two poles of Eq. (15.17), located at $z = 1$ and $z = e^{-aT}$. The root-locus method can therefore be applied easily in the z plane, whereas for sampled functions it is not very convenient to use in the s plane because of the infinite number of poles.

The transformation of the s plane into the z plane can be investigated by inserting $s = \sigma + j\omega_d$ into

$$z = e^{Ts} = e^{\sigma T} e^{j\omega_d T} = e^{\sigma T} e^{j2\pi\omega_d/\omega_s} \tag{15.18}$$

where $\omega_d = \omega_n\sqrt{1-\zeta^2}$.

1. Lines of constant σ in the s plane map into circles of radius equal to $e^{\sigma T}$ in the z plane as illustrated in Fig. 15.7 where $\sigma = -\zeta\omega_n$. In Fig. 15.8 are shown loci of constant ζ and loci of constant ω_n. Specifically, the segment of the imaginary axis in the s plane of width ω_s maps into the circle of unit radius [unit circle (UC)] in the z plane as shown in Fig. 15.17*b*; successive segments map into

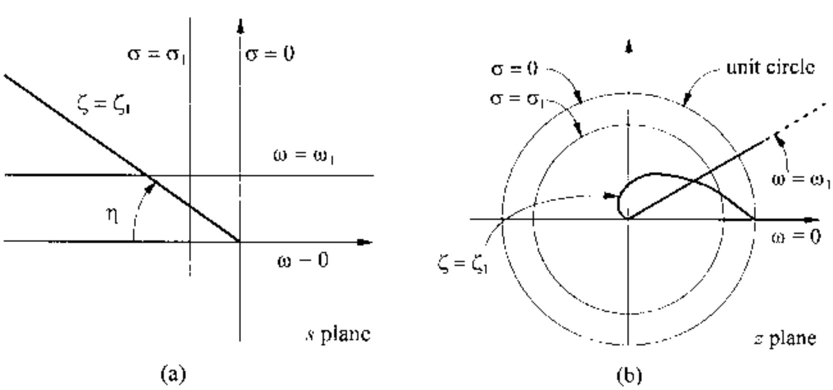


FIGURE 15.7 Transformation from the s plane to the z plane.

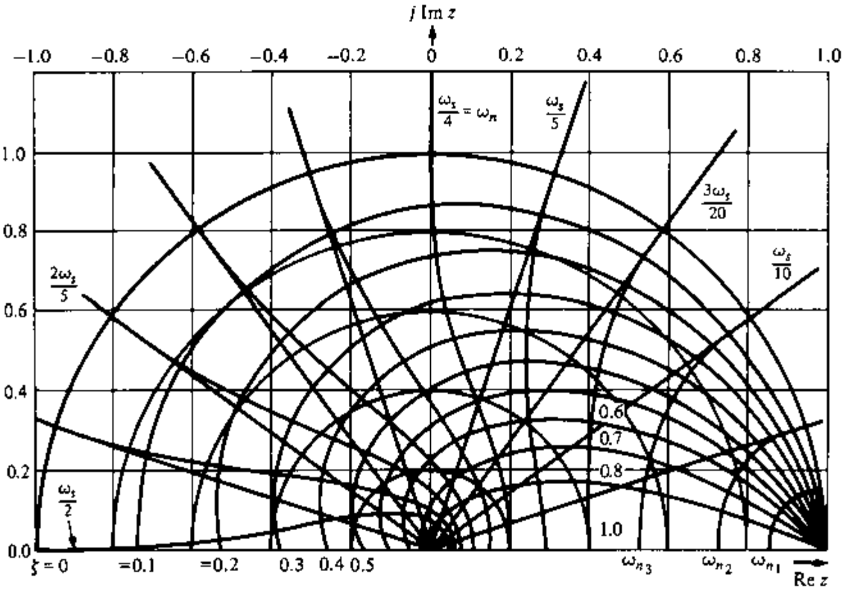


FIGURE 15.8 Plots of loci of constant ζ and loci of constant ω_n. (From Ref. 1, with permission of the McGraw-Hill Companies.)

overlapping circles. This fact shows that a *proper* consideration of sampled-data (S-D) systems in the z plane requires the use of a multiple-sheeted surface, i.e., a Riemann surface. But by virtue of the uniform repetition of the roots of the characteristic equation, the condition for stability is that all roots of the characteristic equation contained in the principal branch lie within the unit circle in the principal sheet of the z plane.

2. Lines of constant ω_d in the s plane map into radial rays drawn at the angle ωT in the z plane. The portion of the constant ω_d line in the left half of the s plane becomes the radial ray within the unit circle in the z plane. The negative part of real axis $-\infty < \sigma < 0$ in the s plane is mapped on the segment of the real axis defined by $0 < z \leq 1$.
3. The constant-damping-ratio ray in the s plane is defined by the equation

$$s = \sigma + j\omega_d = -\omega_d \cot\ \eta + j\omega_d$$

when $\eta = \cos^{-1}\zeta$. Therefore,

$$z = e^{sT} = e^{-\omega_d T \cot\eta + j\omega_d T} = e^{-\omega_d T \cot\eta} \angle \omega_d T \tag{15.19}$$

The corresponding map describes a logarithmic spiral in the z plane.

In summary, the strips in the left-half s plane ($\sigma < 0$) map into the region inside the UC in the z plane, and the strips in the right-half s plane ($\sigma > 0$) map into the region outside the UC in the z plane.

15.4 $\mathcal{Z}$-TRANSFORM THEOREMS

A short list of $\mathcal{Z}$ transforms is given in Table 15.1. Several simple properties of the one-sided $\mathcal{Z}$ transform permit the extension of the table and facilitate its use in solving difference equations. These properties are presented as theorems. They apply when the $\mathcal{Z}$ transform of $f^*(t)$, denoted by $\mathcal{Z}[f^*(t)]$, is $F(z)$ and the sampling time is T.

Theorem 1. Translation in Time (Time Shift). Shifting to the right (delay) yields

$$\mathcal{Z}[f^*(t - pT)] = z^{-p}F(z) \tag{15.20}$$

Shifting to the left (advance) yields

$$\mathcal{Z}[f^*(t + pT)] = z^{p}F(z) - \sum_{i=0}^{p-1} f(iT)z^{p-i} \tag{15.21}$$

Theorem 2. Final Value. If $F(z)$ converges for $|z| > 1$ and all poles of $(1 - z)F(z)$ are inside the unit circle, then

$$\lim_{k\to\infty} f(kT) = \lim_{z\to 1} [(1 - z^{-1})F(z)] \tag{15.22}$$

Theorem 3. Initial Value. If $\lim_{z\to\infty} F(z)$ exists, then

$$\lim_{k\to 0} f(kT) = \lim_{z\to\infty} F(z) \tag{15.23}$$

15.5 DIFFERENTIATION PROCESS [1]

An approximation to the continuous derivative $c(t) = \dot{r}(t)$ uses the first-backward difference or Euler's technique:

$$c(kT) = \frac{\{r(kT) - r[(k-1)T]\}}{T} \tag{15.24}$$

Using this numerical analysis approach to approximate the derivative results in an anticipatory solution requiring future inputs to be known in order to generate the current solution (noncausal situation). This future knowledge is usually difficult to generate in a real-time control system. These numerical

TABLE 15.1 Table of z Transforms

Continuous time function	Laplace transform	z transform of $f^*(t)$ $F(z)$	Discrete time function $f(kt)$ for $k \geq 0$
1. $\delta_0(t)$	1	1	$\delta_0(0)$
2. $\delta(t-kT)$ k is any integer	e^{-kTs}	z^{-1}	$\delta(t-kT)$
3. $u_{-1}(t)$	$\frac{1}{s}$	$\frac{z}{z-1}$	$\delta(t-kT)$
4. $tu_{-1}(t)$	$\frac{1}{s^2}$	$\frac{Tz}{(z-1)^2}$	$kT\delta(t-kT)$
5. $\frac{t^2}{2}u_{-1}(t)$	$\frac{1}{s^3}$	$\frac{T^2z(z+1)}{2(z-1)^3}$	$kT^2\delta(t-kT)$
6. e^{-at}	$\frac{1}{s+a}$	$\frac{z}{z-e^{-aT}}$	e^{-akT}
7. $\frac{e^{-bt}-e^{-at}}{a-b}$	$\frac{1}{(s+a)(s+b)}$	$\frac{1}{a-b}\left[\frac{z}{z-e^{-bT}}-\frac{z}{z-e^{-aT}}\right]$	$\frac{1}{a-b}(e^{-bkT}-e^{-akT})$
8. $u_{-1}(t)-e^{-at}$	$\frac{a}{s(s+a)}$	$a\frac{(1-e^{-aT})z}{(z-1)(z-e^{-aT})}$	$\delta(t-kT)e^{-akT}$
9. $t-\frac{1-e^{-aT}}{a}$	$\frac{a}{s^2(s+a)}$	$\frac{Tz}{(z-1)^2}-\frac{(1-e^{-aT})z}{a(z-1)(z-e^{-aT})}$	$kT-\frac{1-e^{-akT}}{a}$
10. $\sin at$	$\frac{a}{s^2+a^2}$	$\frac{z\sin aT}{z^2-2z\cos aT+1}$	$\sin akT$
11. $\cos at$	$\frac{s}{s^2+a^2}$	$\frac{z(z-\cos aT)}{z^2-2z\cos aT+1}$	$\cos akT$
12. $e^{-at}\sin bt$	$\frac{b}{(s+a)^2+b^2}$	$\frac{ze^{-aT}\sin bT}{z^2-2ze^{-aT}\cos bT+e^{-2aT}}$	$e^{-akT}\sin bkT$
13. $e^{-at}\cos bt$	$\frac{s+a}{(s+a)^2+b^2}$	$\frac{z^2-ze^{-aT}\cos bT}{z^2-2ze^{-aT}\cos bT+e^{-2aT}}$	$e^{-akT}\cos bkT$
14. te^{-at}	$\frac{1}{(s+a)^2}$	$\frac{Tze^{-aT}}{\left(z-e^{-aT}\right)^{-2}}$	$kT\, e^{-akT}$

analysis concepts are very important in realizing a digital control law or equation that will prove to be a linear difference equation.

15.5.1 First Derivative Approximation

The backward-difference discretization of the differentiation process is defined by

$$\dot{c}(t) \equiv Dc(t) \equiv \frac{dc(t)}{dt} \equiv \lim_{T\to 0}\frac{c(kT)-c[(k-1)T]}{T} \tag{15.25}$$

where $c(t)$ has the appropriate continuity characteristics and D is the derivative operator. This equation may be approximated by the first-backward difference as follows:

$$Dc(kT) = \left.\frac{dc(t)}{dt}\right|_{t=kT} \approx \frac{c(kT) - c[(k-1)T]}{T} = \frac{1}{T}\nabla c(kT) \tag{15.26}$$

15.5.2 Second Derivative Approximation

The second derivation backward-difference discretization is defined as

$$D \cdot Dc(t) \equiv D^2 c(t) \equiv \frac{d^2 c(t)}{dt^2} \equiv \lim_{T\to 0} \frac{Dc(kT) - Dc[(k-1)T]}{T} \tag{15.27}$$

which may be approximated by

$$D^2 c(kT) = \left.\frac{d^2 c(t)}{dt^2}\right|_{t=kT} \approx \frac{Dc(kT) - Dc[(k-1)T]}{T} \tag{15.28}$$

Note that the term $Dc(kT)$ in Eq. (15.28) is given by Eq. (15.26). The term $Dc[(k-1)T]$ is expressed in the format of Eq. (15.26) by replacing k by $(k-1)$ to yield

$$\frac{1}{T}\nabla c[(k-1)T] = \frac{c[(k-1)T] - c[(k-2)T]}{T} \tag{15.29}$$

Substituting Eqs. (15.26) and (15.29) into Eq. (15.28) yields the second-backward difference

$$\begin{aligned} D^2 c(kT) &\approx \frac{1}{T^2}\{\nabla c(kT) - \nabla c[(k-1)T]\} \\ &\approx \frac{1}{T^2}\{c(kT) - 2c[(k-1)T]\} \approx \frac{1}{T^2}\nabla^2 c(kT) \end{aligned} \tag{15.30}$$

15.5.3 rth Derivative Approximation

The approximation of the rth-backward difference of the rth derivative is

$$\begin{aligned} D^r c(kT) &\approx \frac{1}{T^r}\{\nabla^{r-1} c(kT) - \nabla^{r-1} c[(k-1)T] + c[(k-2)T]\} \\ &= \frac{1}{T^r}\nabla^r c(kT) \end{aligned} \tag{15.31}$$

where $r = 1, 2, 3, \ldots$ and $k = 0, 1, 2, \ldots$. For a given value of r, Eq. (15.31) can be expanded in a similar manner as is done for the second derivative backward difference approximation to obtain the expanded representation of $D^r c(kT)$.

15.6 SYNTHESIS IN THE z DOMAIN (DIRECT METHOD)

For the block diagram of Fig. 15.2, the system equations are

$$C(s) = G(s)E^*(s) \tag{15.32}$$

$$E(s) = R(s) - B(s) = R(s) - G(s)H(s)E^*(s) \tag{15.33}$$

The starred transform of Eq. (15.33) is

$$E^*(s) = R^*(s) - GH^*(s)E^*(s) \tag{15.34}$$

where $GH^*(s) \equiv [G(s)H(s)]$. Note that, in general, $GH^*(s) \neq G^*(s)H^*(s)$.

Solving for $E^*(s)$ from Eq. (15.34) and substituting into Eq. (15.32) yields

$$C(s) = \frac{G(s)R^*(s)}{1 + GH^*(s)} \tag{15.35}$$

The starred transform $C^*(s)$ obtained from Eq. (15.35) is

$$C^*(s) = \frac{G^*(s)R^*(s)}{1 + GH^*(s)} \tag{15.36}$$

The $\mathcal{Z}$ transform of this equation is obtained by replacing each starred transform by the corresponding function of z:

$$C(z) = \frac{G(z)R(z)}{1 + GH(z)} \tag{15.37}$$

Although the output $c(t)$ is continuous, the inverse of $C(z)$ in Eq. (15.37) yields only the set of values $c(kT)$, $k = 0, 1, 2, \ldots$, corresponding to the values at the sampling instants. Thus, $c(kT)$ is the set of impulses from an ideal fictitious sampler located at the output, as shown in Fig. 15.2, which operates in synchronism with the sampler of the actuating signal.

Example 15.2. Derive the expressions (1) $C^*(s)$, (2) $C(z)$, and (3) the control ratio $C(z)/R(z)$ if possible for the S-D control system of Fig. 15.9.

Part (1). To derive $C^*(s)$, complete the following two steps.

Step **1:** The expressions that relate all the external and internal variables shown in Fig. 15.9 are:

$$C(s) = G_2(s)E_1^*(s) \tag{a}$$

$$E_1(s) = G_1(s)E(s) \tag{b}$$

$$E(s) = R(s) - B(s) \tag{c}$$

$$B(s) = H_1(s)I^*(s) \tag{d}$$

$$I(s) = H_2(s)C(s) \tag{e}$$

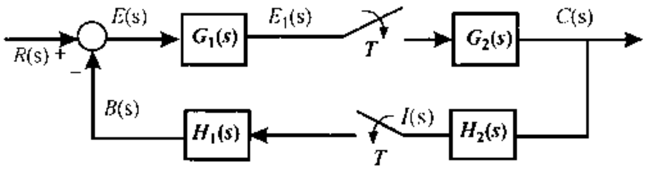


FIGURE 15.9 A nonunity feedback sampled-data control system. (From Ref. 1, with permission of the McGraw-Hill Companies.)

Step **2:** The objectives are (i) to manipulate these equations in such a manner that all the internal variables are eliminated; and (ii) to obtain a relationship between the input and output system variables only. Substituting equation (*a*) into equation (*e*) and equation (*c*) into equation (*b*) yields, respectively:

$$I(s) \quad = G_2(s)H_2(s)E_1^*(s) \tag{f}$$

$$E_1(s) = G_1(s)R(s) - G_1(s)B(s) \tag{g}$$

Equation (*d*) is substituted into equation (*g*) to yield:

$$E_1(s) = G_1(s)R(s) - G_1(s)H_1(s)I^*(s) \tag{h}$$

The starred transform of equations (*a*), (*f*), and (*h*) are, respectively:

$$C^*(s) = G_2^*(s)E_1^*(s) \tag{i}$$

$$I^*(s) = G_2H_2^*(s)E_1^*(s) \tag{j}$$

$$E_1^*(s) = G_1R^*(s) - G_1H_1^*(s)I^*(s) \tag{k}$$

Note that the input forcing function $r(t)$ is now incorporated as part of the starred transform $G_1R^*(s)$.
Substituting $I^*(s)$ from equation (*j*) into equation (*k*) and then solving for $E_1^*(s)$ yields:

$$E_1^*(s) = \frac{G_1R^*(s)}{1 + G_1H_1^*(s)G_2H_2^*(s)} \tag{l}$$

This equation is substituted into equation (*i*) to yield:

$$C^*(s) = \frac{G_2^*(s)G_1R^*(s)}{1 + G_1H_1^*(s)G_2H_2^*(s)} \tag{15.38}$$

Part (2). The $\mathcal{Z}$ transform of Eq. (15.38) is obtained by replacing each starred transform by the corresponding function of z.

$$C(z) = \frac{G_2(z)G_1R(z)}{1 + G_1H_1(z)G_2H_2(z)} \tag{15.39}$$

Part (3). Note that the input $R(z)$ is incorporated as part of $G_1R(z)$; thus this equation cannot be manipulated to obtain the desired control ratio. A *pseudo control ratio* can be obtained as follows:

$$\left[\frac{C(z)}{R(z)}\right]_p = \frac{1}{R(z)}[C(z)] = \frac{1}{R(z)}\left[\frac{G_2(z)G_1R(z)}{1 + G_1H_1(z)G_2H_2(z)}\right] \tag{15.40}$$

15.6.1 z-Plane Stability

The control ratio for the system of Fig. 15.2, obtained from Eq. (15.37), is

$$\frac{C(z)}{R(z)} = \frac{G(z)}{1 + GH(z)} = \frac{W(z)}{Q(z)} \tag{15.41}$$

The characteristic equation of a S-D control system is obtained by setting to zero the denominator of the control system's control ratio. For example, the characteristic equation for the closed-loop system as represented by Eq. (15.41) is given by

$$Q(z) = 1 + GH(z) = 0 \tag{15.42}$$

In general, for a closed-loop S-D control system, the characteristic equation has the form:

$$Q(z) \equiv 1 + P(z) = 0 \tag{15.43}$$

As mentioned previously, the nature and location of the roots of the characteristic equation (15.42) determine the stability and the dynamic behavior of the closed-loop system in a manner similar to that for analog systems. For a continuous-time system, the root locus for the closed-loop system is based on the characteristic equation $1 + G(s)H(s) = 0$ or, equivalently, $G(s)H(s) = \varepsilon^{-j(1+2h)\pi}$, where $h = 0, \pm 1, \pm 2, \ldots$. Similarly, a root-locus analysis can be made for a S-D system whose closed-loop characteristic equation is given by Eq. (15.42).

15.6.2 System Stability [5]

The fundamental concepts of stability are presented in the previous chapters. For the discrete model $c(kT + 1) = f[c(kT), kT]$, the system is *stable* if all solutions remain within a small distance of each other after a specific discrete value of time, k_nT. *Asymptotic stability* refers to the condition that all the solutions to the discrete model are stable and the distance between them goes to zero as $k \to \infty$. Asymptotic stability then implies stability. For a discrete time-invariant linear system model where $F(x,kT) \to F(z)$, asymptotic stability requires that all characteristic roots of the characteristic equation [or poles for $F(z)$] are within the unit circle (UC). Another stability relationship defined as *bounded-input bounded output* (BIBO) refers to the application of a bounded input generating a bounded output in a discrete time-invariant linear model. Asymptotic stability for this model implied BIBO stability.

To determine if a given discrete time-invariant linear model is stable (asymptotically stable), a number of techniques [1] are available: generation of characteristic values, Nyquist's method, Jury's stability test, root-locus, Bode diagram, and Lyapunov's second method. The characteristic values for a given z domain transfer function can be determined by using a computer-aided-design (CAD) package with the appropriate accuracy.

System stability is determined by the location of the roots of the system characteristic equation in the z domain. In Sec. 15.3 it is stated that the primary and complementary strips are transformed into the same (overlapping) portions of the z plane. That is, the strips in the left-half s plane ($\sigma < 0$) map into the region inside the UC in the z plane, and the strips in the right-half s plane ($\sigma > 0$) map into the region outside the UC in the z plane. Figure 15.10 illustrates the mapping of the primary strip into the inside of the unit circle in the z plane. Therefore, a sample-data control system is stable if the roots of the z domain characteristic equation lie inside the UC. In Chapter 16 it is shown that by transforming a sample-data control system to its corresponding PCT analog control system its stability can be determined in the same manner as for analog systems.

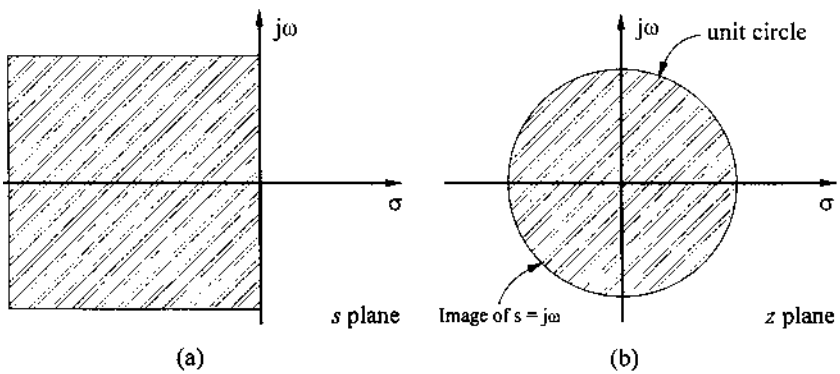


FIGURE 15.10 The mapping of the s plane into the z plane by means of $z = e^{Ts}$. (From Ref. 1, with permission of the McGraw-Hill Companies.)

15.6.3 System Analysis

Analysis of the performance of the sampled system, corresponding to the response given by Eq. (15.37), can be performed by the frequency-response or root-locus method. It should be kept in mind that the geometric boundary for stability in the z plane is the unit circle. As an example of the root-locus method, consider the sampled feedback system represented by Fig. 15.2 with $H(s) = 1$. Using $G(s)$ given by Eq. (15.15) with $a = 1$ and a sampling time $T = 1$, the equation for $G(z)$ is

$$G(z) = GH(z) = \frac{0.632Kz}{(z-1)(z-0.368)} \tag{15.44}$$

The characteristic equation, as obtained from Eq. (15.42), is $1 + GH(z) = 0$ or

$$GH(z) = -1 \tag{15.45}$$

The usual root-locus techniques can be used to obtain a plot of the roots of Eq. (15.45) as a function of the sensitivity K. The root locus is drawn in Fig. 15.11. The maximum value of K for stability is obtained from the magnitude condition which yields $K_{\max} = 2.73$. This occurs at the crossing of the root locus and the unit circle. The selection of the desired roots can be based on the damping ratio ζ desired. For the specified value of ζ, the spiral given by Eq. (15.19) and shown in Fig. 15.8, must be drawn. The intersection of this curve with the root locus determines the roots. Alternatively, it is possible to specify the setting time. This determines the value of σ in the s plane, so the circle of radius $e^{\sigma T}$ can be drawn in the z plane. The intersection of this circle and the root locus determines the roots. For the value $\zeta = 0.48$ the roots are

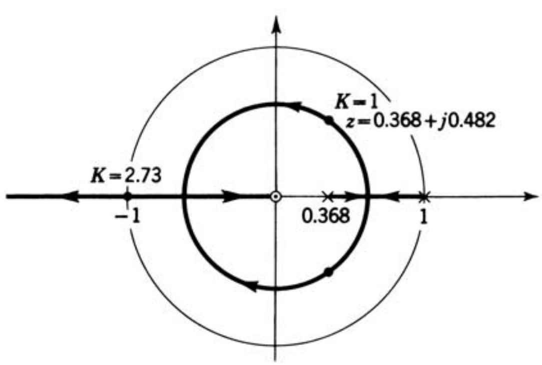


FIGURE 15.11 Root locus for Eqs. (15.44) and (15.45).

$z = 0.368 + j0.482$, shown on the root locus in Fig. 15.11, and the value of $K = 1$. The control ratio is therefore

$$\frac{C(z)}{R(z)} = \frac{0.632z}{z - 0.368 \pm j0.482} = \frac{0.632z}{z^2 - 0.736z + 0.368} \tag{15.46}$$

For a unit-step input the value of $R(z)$ is

$$R(z) = \frac{z}{z-1} \tag{15.47}$$

so that

$$C(z) = \frac{0.632z^2}{(z-1)(z^2 - 0.736z + 0.368)} \tag{15.48}$$

The expression for $C(z)$ can be expanded by dividing its denominator into its numerator to get a power series in z^{-1}.

$$C(z) = 0.632z^{-1} + 1.096z^{-2} + 1.205z^{-3} + 1.120z^{-4} + 1.014z^{-5} + 0.98z^{-6} + \cdots \tag{15.49}$$

The inverse transform of $C(z)$ is

$$c(kT) = 0\delta(t) + 0.632\delta(t - T) + 1.096\delta(t - 2T) + \cdots \tag{15.50}$$

Hence, comparing Eq. (15.49) with Eq. (15.11), the values of $c(kT)$ at the sampling instants are the coefficients of the terms in the series of Eq. (15.50) at the corresponding sampling instants. A plot of the values $c(kT)$ is shown in Fig. 15.12. The curve of $c(t)$ is drawn as a smooth curve through these plotted points.

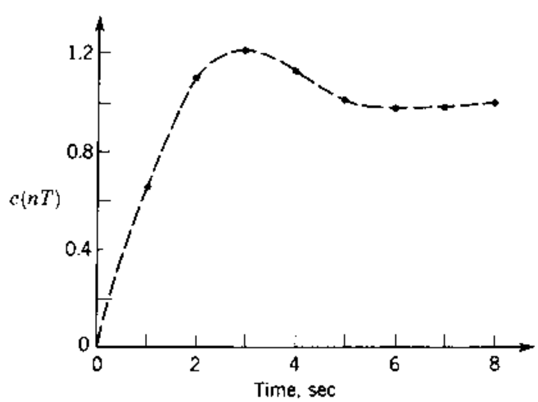


FIGURE 15.12 Plot of $c(nT)$ for Eq. (15.50).

This section presents several basic characteristics in the analysis of sampled-data (S-D) systems. Additional topics of importance include the reconstruction of the continuous signal from the sampled signal by use of *hold* circuits and the problem of compensation to improve performance. These topics are presented in the following sections.

15.7 THE INVERSE $\mathcal{Z}$ TRANSFORM

In the example of Sec. 15.6 the inverse of $C(z)$ is obtained by expanding this function into an infinite series in terms of z^{-k}. This is referred to as the *power-series method*. Then the coefficients of z^{-k} are the values of $c(kT)$. The division required to obtain $C(z)$ as an infinite series (open form) is easily performed on a digital computer. However, if an analytical expression in closed form for $c^*(t)$ is desired, $C(z)$ can be expanded into partial fractions which appear in Table 15.1. Since the $\mathcal{Z}$ transforms in Table 15.1 contain a zero at the origin in the numerator, the Heaviside partial fraction expansion is first performed on $C(z)/z$; i.e., functions of $C(z)$ that contain a zero at the origin must first be put into proper form.

Example 15.3. For $C(z)$ given by Eq. (15.48) with $T = 1$,

$$\frac{C(z)}{z} = \frac{A}{z-1} + \frac{Bz + C}{z^2 - 0.736z + 0.368} \tag{15.51}$$

The coefficients are evaluated by the usual Heaviside partial-fraction method and yield $A = 1$, $B = -1$, and $C = 0.368$. Inserting these values into Eq. (15.51) and using the form of entries 3 and 13 of Table 15.1, the response transform $C(z)$ is

$$C(z) = \frac{z}{z-1} - \frac{z(z - e^{-a}\cos b)}{z^2 - 2ze^{-a}\cos b + e^{-2a}} \tag{15.52}$$

where $a = 0.5$ and $b = 1$. Thus, from Table 15.1 the system output of Eq. (15.52) in closed form is

$$c(kT) = 1 - e^{-0.5k}\cos k \tag{15.53}$$

The open form introduces an error that is propagated in the division due to round off. This does not occur with the closed form. Remember that $C(z)$ represents the sampled values $c^*(t)$ of $c(t)$ at the sample instants kT. Thus, the inversion from the z domain to the time domain yields $c^*(t)$ or the values $c(kT)$. Thus, the inversion process yields values of $c(t)$ only at the sampling instants. In other words, $C(z)$ does not contain any information about the values in between the sampling instants. This is an inherent limitation of the $\mathcal{Z}$ transform method.

In order to obtain the inverse $\mathcal{Z}$ transform of functions of $C(z)$ that do not contain a zero at the origin, the reader is referred to Ref. 1.

15.8 ZERO-ORDER HOLD

The function of a zero-order hold (ZOH) is to reconstruct a piecewise-continuous signal from the sampled function $f^*(t)$. It holds the output amplitude constant at the value of the impulse for the duration of the sampling period T. Thus, it has the transfer function

$$G_{zo}(s) = \frac{1 - e^{-Ts}}{s} \tag{15.54}$$

The action of the ZOH is shown in Fig. 15.13. When the sampling time T is small or the signal is slowly varying, the output of the ZOH is frequently used in digital control systems. It is used to convert the discrete signal obtained from a digital computer into a piecewise-continuous signal that is the input to the plant. The ZOH is a low-pass filter that, together with the basic plant,

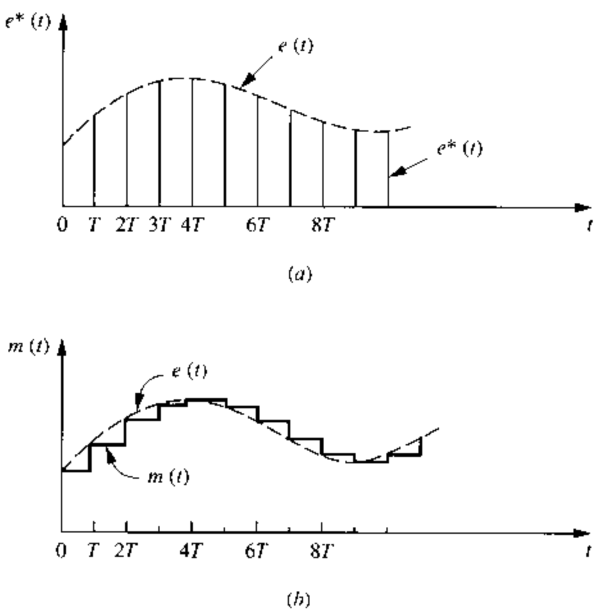


FIGURE 15.13 Input and output signals for a zero-order hold: (*a*) continuous input signal $e(t)$ and the sampled signal $e^*(t)$: (*b*) continuous signal $e(t)$ and the piecewise-constant output $m(t)$ of the zero-order hold.

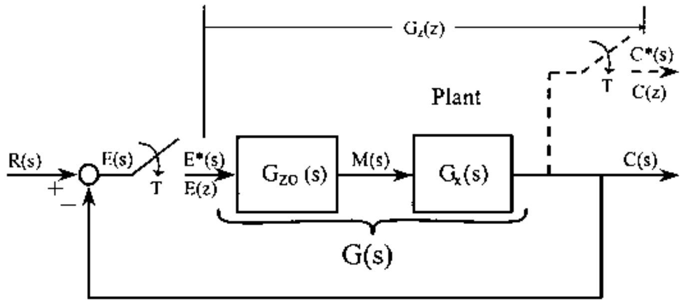


FIGURE 15.14 The uncompensated sampled-data (S-D) control system.

attenuates the complementary frequency spectra introduced by the sampling process. The ZOH does introduce a lag angle into the system, and therefore it can affect the stability and time-response characteristic of the system.

Example 15.4. Figure 15.14 shows a unity-feedback system in which a zero-order hold is placed after the sampler and before the plant $G_x(s)$, which is given by Eq. (15.14) where $a = 1$. The $\mathcal{Z}$ transform of the forward transfer function $G_z(z)$ is

$$G_z(z) = \frac{C(z)}{E(z)} = \mathcal{Z}[G(s)] = \mathcal{Z}[G_{zo}(s)G_x(s)] = \mathcal{Z}\left[\frac{1 - e^{-Ts}}{s}\frac{K}{s(s+1)}\right] \quad (15.55)$$

Because $\mathcal{Z}(1 - e^{-Ts}) = 1 - z^{-1}$, this term can be factored from Eq. (15.55) to give, for $T = 1$,

$$G_z(z) = (1 - z^{-1})\mathcal{Z}\left[\frac{K}{s^2(s+1)}\right] = \frac{K(ze^{-1} + 1 - 2e^{-1})}{(z-1)(z-e^{-1})} = \frac{0.368K(z+0.717)}{(z-1)(z-0.368)} \quad (15.56)$$

A comparison with the $\mathcal{Z}$ transform without the ZOH as given by Eq. (15.44) shows that the zero is moved from the origin to $z = -0.717$.

The root locus for the system of Fig. 15.14 is shown in Fig. 15.15 and may be compared with the root locus without the ZOH in Fig. 15.11. The maximum value of the gain K for stability is $K_{\max} = 0.8824$. With $\zeta = 0.707$, the closed-loop control ratio is

$$\frac{C(z)}{R(z)} = \frac{0.118(z + 0.717)}{(z - 0.625 + j0.249)(z - 0.625 - j0.249)} = \frac{0.118(z + 0.17)}{z^2 - 1.25z + 0.453} \quad (15.57)$$

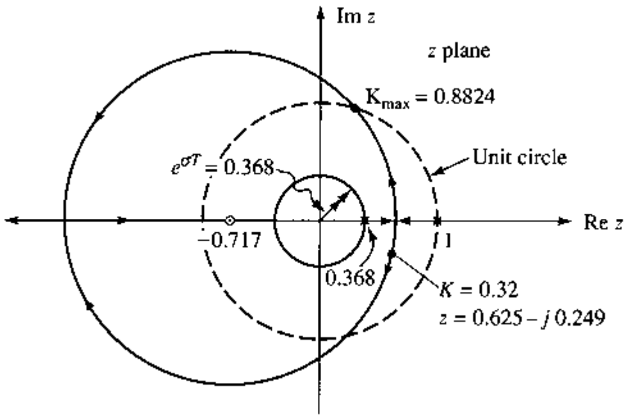


FIGURE 15.15 Root locus for the system of Fig. 15.14.

For this control ratio, $e^{\sigma T} = 0.673$ so that the corresponding real part of the closed-loop poles in the s domain is $\sigma = -0.396$ and the approximate setting time is $T_s = 4/0.396 = 10.1$ s.

This example illustrates that a second-order sampled system becomes unstable for large values of gain. This is in contrast to a continuous second-order system, which is stable for all positive values of gain. The advantage of a discrete control system is the greater flexibility of compensation that can be achieved with a digital compensator.

15.9 LIMITATIONS

Like any method of analysis, there are limitations to the use of the $\mathcal{Z}$-transform method. In spite of these limitations, however, this method is a very useful and convenient analytical tool because of its simplicity. Nevertheless, the following limitations must be kept in mind when applying and interpreting the results of the $\mathcal{Z}$-transform method:

1. The use of an ideal sampler in the analysis of a discrete-data system is based upon the model in which the strengths (area) of the impulses in the impulse train $c^*(t)$ are equal to the corresponding values of the input signal $c(t)$ at the sampling instants kT. For this mathematical model to be "close" to correct, it is necessary for the sampling duration (or pulse width) γ to be very small in comparison with the smallest time constant of the system. It must also be very much smaller than the sampling time T.

2. While the inverse $\mathcal{Z}$ transform $\mathcal{Z}^{-1}[C(z)]$ yields only $c(kT)$ and not a unique analytical function $c(t)$, the *modified* $\mathcal{Z}$-transform method $\mathcal{Z}_m$,[5] removes this limitation.
3. The impulse responses of rational $C(s)$ functions (*exclusive of a hold device*) experience a jump behavior at $t=0$ [at $t=kT$ for $c^*(t)$] unless these functions have at least two more poles than zeros. For these functions,

$$\lim_{t \to kT^+} c(t) = L_+ \neq \lim_{t \to kT^-} c(t) = L_-$$

i.e., the value of $c(kT)$, as t approaches the value of kT from the low side, is not identical to the value obtained when approaching kT from the high side. In other words, a discontinuity occurs in the value of $c(kT)$ at $t=kT$. Fortunately, the $C(z)$ encountered in the analysis of most practical control systems have at least two more poles than zeros ($n \geq w+2$), and this limitation does not apply for such systems.

15.10 STEADY-STATE ERROR ANALYSIS FOR STABLE SYSTEMS

The three important characteristics of a control system are (1) *stability*, (2) *steady-state performance*, and (3) *transient response*. The first item of interest in the analysis of a system is its stability characteristics. If there is a range of gain for which a system yields a stable performance, then the next item of interest is the system's steady-state error characteristics; i.e., can the system output $c(kT)$ follow a given input $r(kT)$ with zero or a small value of error $[e(kT) = r(kT) - c(kT)]$. If the first two items are satisfactory, a transient time-response analysis is then made. Section 15.6.2 and 15.6.3 deal with the first item. This section discusses the second characteristic, and the remaining sections deal, mainly, with the transient response.

Fortunately, the analysis of the steady-state error characteristics of a *unity-feedback S-D system* parallels the analysis for a unity-feedback continuous-time stable system based upon system types and upon specific forms of the system input function. For the unity-feedback S-D system of Fig. 15.14, the control ratio and the output and error signals expressed in the z domain are, respectively,

$$\frac{C(z)}{R(z)} = \frac{G_z(z)}{1 + G_z(z)} \tag{15.58}$$

$$C(z) = \frac{G_z(z)}{1 + G_z(z)} R(z) \tag{15.59}$$

and

$$E(z) = R(z) - C(z) = \frac{R(z)}{1 + G_z(z)} \tag{15.60}$$

Thus, *if the system is stable*, the final-value theorem can be applied to Eqs. (15.59) and (15.60) to obtain the steady-state or final value of the output and the error at the sampling instants: i.e.,

$$c^*(\infty) = \lim_{t\to\infty} c^*(t) = \lim_{z\to 1}\left[\frac{(1 - z^{-1})G_z(z)}{1 + G_z(z)} R(z)\right] \tag{15.61}$$

$$e^*(\infty) = \lim_{t\to\infty} e^*(t) = \lim_{z\to 1}\left[\frac{(1 - z^{-1})R(z)}{1 + G_z(z)}\right] \tag{15.62}$$

The steady-state error analysis for the stable nonunity feedback system of Fig. 15.16*a* may be analyzed by determining its equivalent z domain-stable unity-feedback system shown in Fig. 15.16*b*. The control ratios of the respective configurations of Fig. 15.16 are, respectively,

$$\frac{C(z)}{R(z)} = \frac{G_z(z)}{1 + G_{ZO}G_xH(z)} = \frac{N(z)}{D(z)} \tag{15.63}$$

$$\frac{C(z)}{R(z)} = \frac{G_{eq}(z)}{1 + G_{eq}(z)} \tag{15.64}$$

where $N(z)$ and $D(z)$ are the numerator and denominator polynomials, respectively, of Eq. (15.63). $G_{eq}(z)$ is determined by equating Eqs. (15.63) and (15.64) and manipulating to obtain

$$G_{eq}(z) = \frac{N(z)}{D(z) - N(z)} = \frac{N(z)}{D_G(z)} = \frac{N(z)}{(z - 1)^m D'_G(z)} \tag{15.65}$$

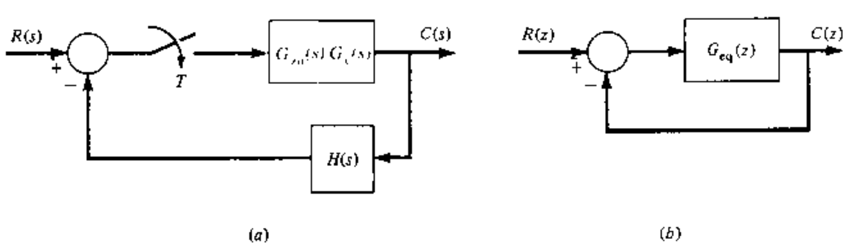


FIGURE 15.16 (*a*) A stable nonunity-feedback sampled-data system; (*b*) the equivalent *z* domain stable unity feedback system of Fig. 15.16*a*. (From Ref. 1, with permission of the McGraw-Hill Companies.)

where $D'_G(z)$ is the resultant polynomial remaining after the term $(z-1)^m$ is factored out of $D_G(z) = D(z) - N(z)$. A similar procedure can be utilized to determine $G_{eq}(z)$ for other stable nonunity-feedback S-D systems. Thus, for Fig. 15.16*b*, the steady or final value of the output and the error at the sampling instants, for this stable system, are given respectively as follows:

$$c^*(\infty) = \lim_{t\to\infty} c^*(t) = \lim_{z\to 1}\left[\frac{(1-z^{-1})G_{eq}(z)}{1+G_{eq}(z)}R(z)\right] \tag{15.66}$$

$$e^*(t) = \lim_{t\to\infty} e^*(t) = \lim_{z\to 1}\left[\frac{(1-z^{-1})R(z)}{1+G_{eq}(z)}\right] \tag{15.67}$$

15.10.1 Steady-State Error-Coefficients

The steady-state error coefficients have the same meaning and importance for sampled-data systems as for continuous time systems; i.e., how well the system output can follow a given type of input forcing function. *The following derivations of the error coefficients are independent of the system type. They apply to any system type and are defined for specific forms of the input. These error coefficients are useful only for a stable system and are defined in this text for a unity-feedback system.*

Step Input. $R(z) = R_o z/(z-1)$. The step error coefficient K_p is defined as

$$K_p \equiv \frac{c^*(\infty)}{e^*(\infty)} \tag{15.68}$$

Substituting from Eqs. (15.61) and (15.62) into Eq. (15.68) yields

$$K_p = \lim_{z\to 1} G(z) \tag{15.69}$$

which applies only for a step input, $r(t) = R_0 u_{-1}(t)$.

Ramp Input. $R(z) = R_1 z/(z-1)^2$. The ramp error coefficient K_v is defined as

$$K_v \equiv \frac{\text{steady-state value of derivative of output}}{e^*(\infty)} \tag{15.70}$$

Since Eq. (15.26) represents the derivative of $c(t)$ in the discrete-time domain, use of the translation and final-value theorems [Eqs. (15.20) and (15.22), respectively] permits Eq. (15.70) to be written as

$$K_v = \frac{\lim_{z\to 1}\left[\frac{(1-z^{-1})^2}{T}C(z)\right]}{e^*(\infty)} \tag{15.71}$$

Substituting from Eqs. (15.61) and (15.62) into Eq. (15.71) yields

$$K_v = \frac{1}{T} \lim_{z \to 1} \left[\frac{z-1}{z} G_z(z) \right] \qquad \mathrm{s}^{-1} \tag{15.72}$$

which applies only for a ramp input, $r(t) = R_1 t u_{-1}(t)$.

Parabolic Input. $R(z) = R_2 T^2 z(z+1)/2(z-1)^3$. The parabolic error coefficient K_a is defined as

$$K_a \equiv \frac{\text{steady-state value of second derivative of output}}{e^*(\infty)} \tag{15.73}$$

Since Eq. (15.30) represents the second derivative of the output in the discrete-time domain, use of the translation and final-value theorems permits Eq. (15.73) to be written as

$$K_a = \frac{\lim_{z \to 1} \left[\frac{(1-z^{-1})^3}{T^2} \right]}{e^*(\infty)} \tag{15.74}$$

Substituting from Eqs. (15.59) and (15.67) into Eq. (15.74) yields

$$K_a = \frac{1}{T^2} \lim_{z \to 1} \left[\frac{(z-1)^2}{z^2} G_z(z) \right] \qquad \mathrm{s}^{-2} \tag{15.75}$$

which applies only for a parabolic input, $r(t) = R_2\, t^2 u_{-1}(t)/2$.

15.10.2 Evaluation of Steady-State Error Coefficients

The forward transfer function of a sampled-data unity-feedback system in the z domain has the general form

$$G(z) = \frac{Kz^d (z-a_1)(z-a_2)\cdots(z-a_i)\cdots}{(z-1)^m (z-b_1)(z-b_2)\cdots(z-b_j)\cdots} \tag{15.76}$$

where a_i and b_j may be real or complex, d and m are positive integers, $m = 0, 1, 2, \ldots$, and *m represents the system type.* Note that the $(z-1)^m$ term in Eq. (15.76) corresponds to the s^m term in the denominator of the forward transfer function of a continuous-time Type m system. Substituting from Eq. (15.76) into Eqs. (15.69), (15.72), and (15.75), respectively, yields the following values of the

TABLE 15.2 Steady-State Error Coefficients for Stable Systems

System type	Step error Coefficient K_p	Ramp error Coefficient K_v	Parabolic error Coefficient K_a
0	K_0	0	0
1	∞	K_1	0
2	∞	∞	K_2

steady-state error coefficients for the various Type m systems:

$$K_p = \begin{cases} \lim_{z\to 1} \dfrac{Kz^d(z-a_1)(z-a_2)\cdots}{(z-b_1)(z-b_2)\cdots} = K_0 & \text{Type 0} \quad (15.77) \\ \infty & \text{Type 1} \quad (15.78) \\ \infty & \text{Type 2} \quad (15.79) \end{cases}$$

$$K_v = \begin{cases} \frac{1}{T}\lim_{z\to 1} \dfrac{Kz^d(z-1)(z-a_1)(z-a_2)\cdots}{z(z-b_1)(z-b_2)\cdots} = 0 & \text{Type 0} \quad (15.80) \\ \frac{1}{T}\lim_{z\to 1} \dfrac{Kz^d(z-a_1)(z-a_2)\cdots}{z(z-b_1)(z-b_2)\cdots} = K_1 & \text{Type 1} \quad (15.81) \\ \infty & \text{Type 2} \quad (15.82) \end{cases}$$

$$K_a = \begin{cases} \frac{1}{T^2}\lim_{z\to 1} \dfrac{Kz^d(z-1)^2(z-a_1)(z-a_2)\cdots}{z^2(z-b_1)(z-b_2)\cdots} = 0 & \text{Type 0} \quad (15.83) \\ \frac{1}{T^2}\lim_{z\to 1} \dfrac{Kz^d(z-1)(z-a_1)(z-a_2)\cdots}{z^2(z-b_1)(z-b_2)\cdots} = 0 & \text{Type 1} \quad (15.84) \\ \frac{1}{T^2}\lim_{z\to 1} \dfrac{Kz^d(z-a_1)(z-a_2)\cdots}{z^2(z-b_1)(z-b_2)\cdots} = K_2 & \text{Type 2} \quad (15.85) \end{cases}$$

In applying Eqs. (*15.77*) *to* (*15.85*) *it is required that the denominator of* $G(z)$ *be in factored form to ascertain if it contains any* $z-1$ *factor*(*s*). Table 15.2 summarizes the results of Eqs. (15.77) to (15.85).

15.10.3 Use of Steady-State Error Coefficients

The importance of the steady-state error coefficients is illustrated by means of an example.

Example 15.5. For the system of Fig. 15.14, consider $G_z(z)$ of the form of Eq. (15.76) with $m = 1$ (a Type 1 system). Determine the value of $e^*(\infty)$ for each of the three standard inputs (step, ramp, and parabolic), assuming that the system is stable.

For the definitions of Eqs. (15.68), (15.70), and (15.73) and from Table 15.2, the following results are obtained:

$$e^*(\infty) = \frac{c^*(\infty)}{K_p} = \frac{c^*(\infty)}{\infty} = 0 \tag{15.86}$$

$$e^*(\infty) = \frac{\text{steady-state value of derivative of output}}{K_1} = E_0 \tag{15.87}$$

$$e^*(\infty) = \frac{\text{steady-state value of second derivative of output}}{0} = \infty \tag{15.88}$$

Thus, a Type 1 sampled-data stable system (1) can follow a step input with zero steady-state error, (2) can follow a ramp input with a constant error E_0, and (3) cannot follow a parabolic input. Equation (15.87) indicates that the value of E_0, for a given value of R_1 (for the ramp input), may be made smaller by making K_1 larger. This assumes that the desired degree of stability and the desired transient performance are maintained while K_1 is increasing in value. A similar analysis can be made for Types 0 and 2 systems (see problems).

Example 15.6. For the system of Fig. 15.17, consider that $G_{eq}(z)$, given by Eq. (15.65), is of the form of Eq. (15.76) with $m = 2$ (an equivalent Type 2 system). Determine the steady-state error coefficients for this system.

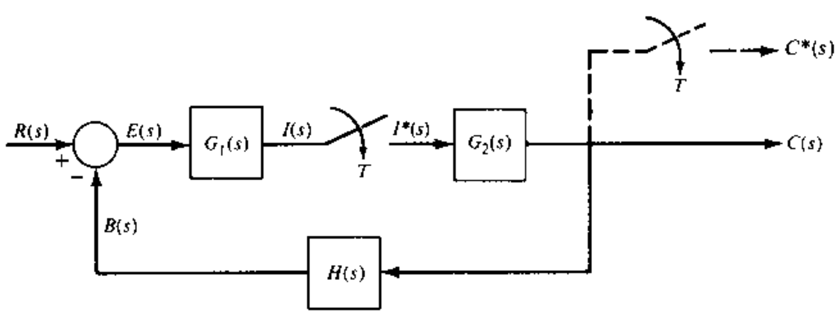


FIGURE 15.17 A nonunity-feedback sampled-data control system. (From Ref. 1, with permission of the McGraw-Hill Companies.)

Applying the definitions of Eqs. (15.69), (15.71), and (15.75) yields, respectively,

$$K_p = \infty \quad K_v = \infty \quad K_a = \frac{1}{T^2}\left[\frac{N(z)}{z^2 D'_G(z)}\right] = K_2$$

Example 15.7. The expression for the output of the control system shown in Fig. 15.17 is

$$C(z) = \frac{G_2(z)G_1(z)R(z)}{1 + G_2G_1H(z)} \tag{15.89}$$

For a specified input $r(t)$ (step, ramp, or parabolic), obtain $R(z) = \mathcal{Z}[r(t)]$, corresponding to a sampled $r(t)$, and then obtain the pseudo control ratio $[C(z)/R(z)]_p$; i.e., divide both sides of Eq. (15.89) by $R(z)$ to obtain

$$\left[\frac{C(z)}{R(z)}\right]_p = \frac{1}{R(z)}\left[\frac{G_2(z)G_1(z)R(z)}{1 + G_2G_1H(z)}\right] \tag{15.90}$$

By setting Eq. (15.90) equal to Eq. (15.64), it is then possible to obtain an expression for $G_{eq}(z)$ that represents the forward transfer function of an equivalent unity-feedback system represented by Fig. 15.16*b*, where $G_z(z) \equiv G_{eq}(z)$. This expression of $G_{eq}(z)$ can then be used to determine the "effective" Type m system that Fig. 15.17 represents and to solve for K_p, K_v, and K_a.

As a specific example, let

$$\left[\frac{C(z)}{R(z)}\right]_p = \frac{K(z^2 + az + b)}{z^3 + cz^2 + dz + e} = \frac{N(z)}{D(z)} \tag{15.91}$$

where $r(t) = u_{-1}(t)$, $K = 0.129066$, $a = 0.56726$, $b = -0.386904$, $c = -1.6442$, $d = 1.02099$, and $e = -0.224445$. For these values of the coefficients the system is stable. Setting Eq. (15.91) equal to Eq. (15.65) and solving for $G_{eq}(z)$ yields

$$G_{eq}(z) = \frac{K(z^2 + az + b)}{z^3 - 1.773266z^2 + 0.97776z - 0.174509} = \frac{N(z)}{D_G(z)}$$

Applying the final-value theorem to Eq. (15.91) yields $c(\infty) = 1$. Therefore, the nonunity feedback system of Fig. 15.17 effectively acts at least as a Type 1 system. Based upon Eq. (15.76) this implies that $D_G(z)$ contains at least one factor of the form $z - 1$. Dividing $D_G(z)$ by $z - 1$ yields $z^2 - 0.773266z + 0.17451$, which does not contain $z - 1$ as a factor. Thus, the nonunity-feedback system is a Type 1 system. The equation for $G_{eq}(z)$ is rewritten as follows:

$$G_{eq}(z) = \frac{K(z^2 + az + b)}{(z-1)(z^2 - 0.773266z + 0.17451)}$$

For the ramp input $r(t) = R_1 t u_{-1}(t)$ with $R_1 = 0.5$ and for $T = 0.1$, then applying Eqs. (15.72) and (15.87) yields, respectively,

$$K_1 = \frac{1}{T}\lim_{z\to 1}\left[\frac{z-1}{z} G_{eq}(z)\right] = \frac{K(1+a+b)}{T(0.401244)} \qquad \mathrm{s}^{-1}$$

and

$$e^*(\infty) = \frac{R_1}{K_1} = \frac{0.5}{3.796} = 0.1317$$

Thus, the system follows a ramp input with a steady-state error.

15.11 ROOT-LOCUS ANALYSIS FOR SAMPLED-DATA CONTROL SYSTEMS

The first thing that a designer wants to know about a given S-D system is whether or not it is stable. This can be determined by examining the roots obtained from the characteristic equation $1 + G(z)H(z) = 0$. Thus, the root-locus method is used to analyze the performance of a S-D control system in the same manner as for a continuous-time control system. For either type of system, the *root locus is a plot of the roots of the characteristic equation of the closed-loop system as a function of the gain constant*. This graphical approach yields a clear indication of gain-adjustment effects with relatively small effort compared with other methods. The underlying principle is that the poles of $C(z)/R(z)$ or $C(z)$ (transient-response modes) are related to the zeros and poles of the open-loop transfer function $G(z)H(z)$ and also the gain. An important advantage of the root-locus method is that the roots of the characteristic equation of the system can be obtained directly, which results in a complete and accurate solution of the transient and steady-state response of the controlled variable. Another important feature is that an approximate control solution can be obtained with a reduction of the required work. With the help of a CAD package, it is possible to synthesize a compensator, if one is required, with relative ease.

This section presents a detailed summary of the root-locus method. The first subsection details a procedure for obtaining the root locus, the next subsection defines the root-locus construction rules for negative feedback, and the last subsection contains examples of this method.

15.11.1 Procedure Outline

The procedure to be followed in applying the root-locus method is outlined in this subsection. This procedure is easier when a CAD program is used to obtain the root locus. Such a program can provide the desired data for the root locus in plotted or tabular form. This procedure, which is a modified version of the one applicable for the continuous-time systems, is summarized as follows

Step 1. Derive the open-loop transfer function $G(z)H(z)$ of the system.

Step 2. Factorize the numerator and denominator of the transfer function into linear factors of the form $z + a$, where a may be real or complex.

Step 3. Plot the zeros and poles of the open-loop transfer function in the z plane, where $z = \sigma_z + j\omega_z$.

Step 4. The plotted zeros and poles of the open-loop function determine the roots of the characteristic equation of the closed-loop system $[1 + G(z)H(z) = 0]$. By use of geometrical shortcuts or a digital-computer program, determine the locus that describes the roots of the closed-loop characteristic equation.

Step 5. Calibrate the locus in terms of the loop sensitivity K. If the gain of the open-loop system is predetermined, the location of the exact roots of $1 + G(z)H(z) = 0$ is immediately known. If the location of the roots (or ζ) is specified, the required value of K can be determined.

Step 6. Once the roots have been found in step 5, the system's time response can be calculated by taking the inverse $\mathcal{Z}$ transform, either manually or by use of a computer program.

Step 7. If the response does not meet the desired specifications, determine the shape that the root locus must have to meet these specifications.

Step 8. Synthesize the network that must be inserted into the system, if other than gain adjustment is required, to make the required modification on the original locus. This process, called *compensation*, is described in Chaps. 10 to 12.

15.11.2 Root-Locus Construction Rules for Negative Feedback

The system characteristic equation, Eq. (15.43), is rearranged as follows:

$$P(z) = -1 \tag{15.92}$$

Assume that $P(z)$ represents the open-loop function

$$P(z) = G(z)H(z) = \frac{K(z - z_1)\cdots(z - z_i)\cdots(z - z_w)}{(z - p_1)\cdots(z - p_c)\cdots(z - p_n)} \tag{15.93}$$

where z_i and p_c are the open-loop zeros and poles, respectively, n is the number of poles, w is the number of zeros, and K *is defined as the static loop sensitivity* (gain constant) when $P(z)$ is expressed in this format. Equation (15.92) falls into the mathematical format for root-locus analysis; i.e.,

$$\text{Magnitude condition:} |P(z)| = 1 \tag{15.94}$$

$$\text{Angle condition:} \quad -\beta = \begin{cases} (1 + 2h)180^\circ & \text{for } K > 0 \\ h360^\circ & \text{for } K < 0 \end{cases} \tag{15.95}$$

Thus, the construction rules for continuous-time systems, with minor modifications since the plot is in the z plane, are applicable for S-D systems and are summarized as follows:

Rule 1. The number of branches of the root locus is equal to the number of poles of the open-loop transfer function.

Rule 2. For *positive* values of K, the root exist on those portions of the real axis for which the sum of the poles and zeros to the right is an odd integer. For *negative* values of K, the root locus exists on those portions of the real axis for which the sum of the poles and zeros to the right is an even integer (including zero).

Rule 3. The root locus starts ($K = 0$) at the open-loop poles and terminates ($K = \pm\infty$) at the open-loop zeros or at infinity.

Rule 4. The angles of the asymptotes of the root locus that end at infinity are determined by

$$\gamma = \frac{(1 + 2h)180^\circ}{[\textit{no. of poles of } G(z)H(z)] - [\textit{no. of zeros of } G(z)H(z)]} \quad \text{for } k > 0 \tag{15.96}$$

and

$$\gamma = \frac{h360^\circ}{[\textit{no. of poles of } G(z)H(z)] - [\textit{no. of zeros of } G(z)H(z)]} \quad \text{for } k < 0 \tag{15.97}$$

Rule 5. The real-axis intercept of the asymptotes is

$$z_0 = \frac{\sum_{c=1}^{n} \text{Re}\, p_c - \sum_{i=1}^{w} \text{Re}\, z_i}{n - w} \tag{15.98}$$

Rule 6. The breakaway point for the locus between two poles on the real axis (or the break-in point for the locus between two zeros on the real axis) can be determined by taking the derivative of the loop sensitivity K with respect to z. Equate this derivative to zero and find the roots of the resulting equation. The root that occurs between the poles (or the zero) is the breakaway (or break-in) point.

Rule 7. For $K > 0$ the angle of departure from a complex pole is equal to 180° minus the sum of the angle from the other poles plus the sum of the angle from the zeros. Any of these angles may be positive or negative. For $K < 0$ the departure angle is 180° from that obtained for $K > 0$.

For $K > 0$ the angle of approach to a complex zero is equal to the sum of the angles from the poles minus the sum of the angles from the other zeros minus 180°. For $K < 0$ the approach angle is 180° from that obtained for $K > 0$.

Rule 8. The root loci are symmetrical about the real axis.

Rule 9. The static loop sensitivity calibration K of the root locus can be made by applying the magnitude condition given by Eq. (15.94) as follows:

$$K = \frac{|z - p_1| \cdot |z - p_2| \cdots |z - p_c| \cdots |z - p_n|}{|z - z_1| \cdot |z - z_2| \cdots |z - z_w|} \tag{15.99}$$

Rule 10. The selection of the dominant roots of the characteristic equation is based upon the specification that give the required system performance; i.e., it is possible to evaluate σ, ω_d, and ζ, from Eqs. (3.60), (3.61), and (3.64).

These values in turn are mapped into the z domain to determine the location of the desired dominant roots in the z plane. The loop sensitivity for these roots is determined by means of the magnitude condition. The remaining roots are then determined to satisfy the same magnitude condition.

A root-locus CAD digital-computer program produces an accurate calibrated root locus. This considerably simplifies the work required for the system design. By specifying ζ for the dominant roots or K, use of a computer program can yield all the roots of the characteristic equation.†

†A good engineering design rule as a first estimate for a calculation step size in a CAD program is $T/10$ in order to generate accurate results.

It should be remembered that the entire *unbounded left-hand s* plane is mapped into the unit circle (UC) in the z plane. The mappings of the poles and zeros in the left-half s plane into the z plane migrate toward the vicinity of the $1+j0$ point as $T \to 0$. Thus, in plotting the poles and zeros of $G(z) = \mathcal{Z}[G(s)]$, they approach the $1+j0$ point as $T \to 0$. For a "small-enough" value of T and an inappropriate plot scale, some or all of these poles and zeros "appear to lie on top of one another." Therefore, *caution* should be exercised in *selecting the scale* for the root-locus plot and the *degree of accuracy* that needs to be maintained for an accurate analysis and design of a S-D system.

15.11.3 Root-Locus Design Examples

Example 15.8. The second-order characteristic equation, for a given control ratio, $z^2 - 0.2Az + 0.1A = 0$ is partitioned to put it into format of Eq. (15.92) as follows:

$$z^2 = 0.2Az - 0.1A = -K(z - 0.5) \tag{15.100}$$

where $K = -0.2A$. Equation (15.100) is rearranged to yield

$$P(z) = \frac{K(z - 0.5)}{z^2} = -1 \tag{15.101}$$

that is of the mathematical format of Eq. (15.92). The poles and zero of Eq. (15.101) are plotted in the z plane as shown in Fig. 15.18. The construction rules applied to this example yield the following information.

Rule 1. Number of branches of the root locus is given by $n = 2$.
Rule 2. For $K > 0$ ($A < 0$) the real-axis locus exists between $z = 0.5$ and $z = -\infty$, and for $K < 0$ ($A > 0$) the real-axis locus exists between $z = 0.5$ and $z = +\infty$.

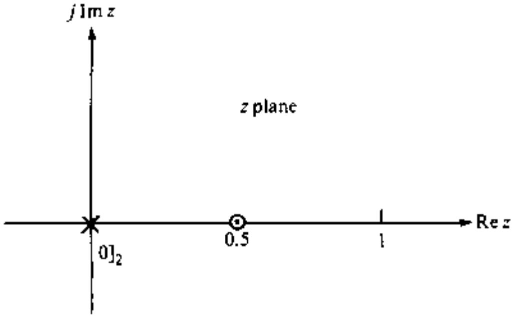


FIGURE 15.18 Poles and zero of Eq. (15.101). (From Ref. 1, with permission of the McGraw-Hill Companies.)

Rule 3. The root-locus branches start, with $K = 0$, at the poles of $P(z)$. One branch ends at the zero, $z = 0.5$, and one branch ends at infinity for $K = \pm\infty$.

Rule 4. Asymptotes

For $K > 0$:

$$\gamma = \frac{(1+2h)180^\circ}{2-1} = (1+2h)180^\circ \qquad \text{thus } \gamma = 180^\circ$$

For $K < 0$:

$$\gamma = \frac{h360^\circ}{2-1} = h360^\circ \qquad \text{thus } \gamma = 0^\circ (\text{or } 360^\circ)$$

Rule 5. Not applicable for the asymptotes determined by Rule 4.

Rule 6. For this example there is no breakaway point on the real axis for $K > 0$. For $K < 0$, Eq. (15.101) is rearranged to obtain the function

$$W(z) = \frac{z^2}{-z + 0.5} = K \tag{15.102}$$

Taking the derivative of this function and setting it equal to zero yields the break-in and breakaway points. Thus,

$$\frac{dW(z)}{dz} = \frac{z(z-1)}{(-z+0.5)^2} = 0$$

which yields $z_{1,2} = 0, 1$. Therefore, $z_1 = 0$ is the breakaway point and $z_2 = 1$ is the break-in point.

Rule 7. Not applicable for this example.

Rule 8. The root locus is symmetrical about the real axis.

Rule 9. The static loop sensitivity calibration of the root locus can be made by evaluating Eq. (15.102) for various values of z that lie on the root locus.

Rule 10. Not applicable for this example.

The root locus for Eq. (15.101) is shown in Fig. 15.19. The UC intersections of the root locus as determined by a CAD program occur for the static loop sensitivity values of $K = 2/3$ and $K = -2$.

Example 15.9. Given the unity-feedback sampled-data system shown in Fig. 15.14, where $G_x(s) = K_G/[s(s+2)]$, the objectives of this example are as follows: (*a*) determine $C(z)/R(z)$ in terms of K_G and T (i.e., the values of K_G and T are unspecified); (*b*) determine the root locus and the maximum value of K_G for a stable response with $T = 0.1$ s; (*c*) determine the steady-state error characteristics with various inputs for this system for those values of K_G and T that yield a stable system response; and (*d*) determine the roots of the

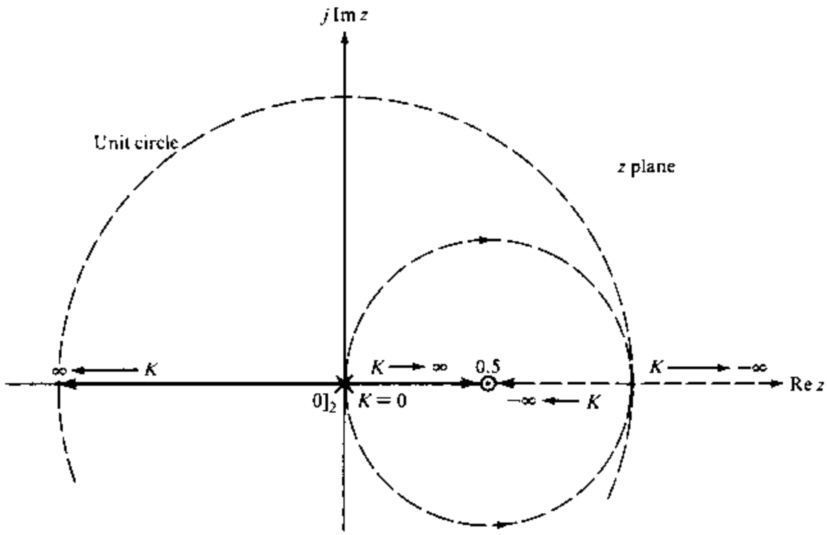


FIGURE 15.19 Root-locus plot of Example 15.8. (From Ref. 1, with permission of the McGraw-Hill Companies.)

characteristic equation for $\zeta = 0.6$, the corresponding time response $c(kT)$, and the figure of merit (FOM).

Solution.

a. The forward transfer function of the open-loop system is

$$G_z(s) = G_{zo}(s)G_x(s) = \frac{K_G(1 - e^{-sT})}{s^2(s+2)} = (1 - e^{-sT})\frac{K_G}{s^2(s+2)}$$
$$= G_e^*(s)\frac{K_G}{s^2(s+2)}$$

Thus, using entry 8 in Table 15.1 yields

$$G_z(z) = G_e(z)\mathcal{Z}\left[\frac{K_G}{s^2(s+2)}\right] = (1 - z^{-1})\mathcal{Z}\left[\frac{K_G}{s^2(s+2)}\right]$$
$$= \frac{K_G[T + 0.5e^{-2T} - 0.5)z + (0.5 - 0.5e^{-2T} - Te^{-2T})]}{2(z-1)(z - e^{-2T})} \tag{15.103}$$

Substituting Eq. (15.103) into

$$\frac{C(z)}{R(z)} = \frac{G_z(z)}{1 + G_z(z)} = \frac{N(z)}{D(z)}$$

yields

$$\frac{C(z)}{R(z)} = \frac{0.5K_G[(T-0.5+0.5e^{-2T})z+(0.5-0.5e^{-2T}-Te^{-2T})]}{z^2-[(1+e^{-2T})-0.5K_G(T-0.5+0.5e^{-2T})]z \\ +e^{-2T}+0.5K_G(0.5-0.5e^{-2T}-Te^{-2T})} \tag{15.104}$$

b. From Eq. (15.104), the characteristic equation is given by $Q(z)=1+G_z(z)=0$, which yields

$$G_z(z) = \frac{K\left(z+\frac{0.5-0.5e^{-2T}-Te^{-2T}}{T+0.5e^{-2T}-0.5}\right)}{(z-1)(z-e^{-2T})} = -1 \tag{15.105}$$

where

$$K = 0.5K_G(T+0.5e^{-2T}-0.5)$$

For $T=0.1$ s,

$$G_z(z) = \frac{K(z+0.9355)}{(z-1)(z-0.81873)} = -1 \tag{15.106}$$

and $K=0.004683\,K_G$. The root locus for Eq. (15.106) is shown in Fig. 15.20. For $T=0.1$ the maximum value of K for a stable

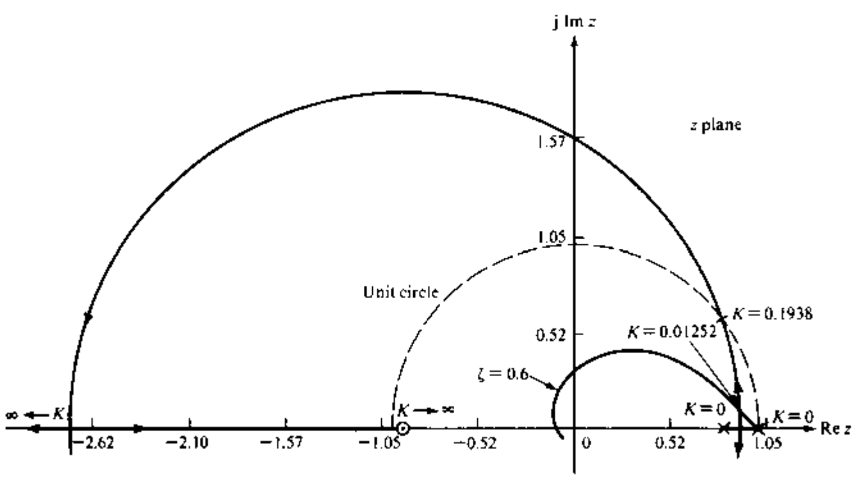


FIGURE 15.20 A root-locus sketch for Eq. (15.106) where $T=0.1$ s. (From Ref. 1, with permission of the McGraw-Hill Companies.)

response is $K \approx 0.1938$, which results in $K_{G_{max}} \approx 41.38$. An analysis of Eq. (15.105) reveal that:

1. The pole at e^{-2T} approaches the UC as $T \to 0$ and approaches the origin as $T \to \infty$.
2. The zero approaches -2 as $T \to 0$ (this can be determined by applying L'Hôpital's rule twice) and approaches the origin as $T \to \infty$.
3. Based upon the plot scale chosen, it may be difficult to interpret or secure accurate values from the root locus in the vicinity of the $1+j0$ point as $T \to 0$. Both poles will "appear" to be superimposed.

As a consequence of items 1 and 2 and considering only the root locus, one may jump to the conclusion that the range of K_G for a stable system decreases as $T \to 0$. This is not the case for this example since K is a function of T, as shown in the next section. As pointed out in item 3, for an open-loop transfer function having a number of poles and zeros in the vicinity of $z=1$, it may be difficult to obtain an accurate root-locus plot in the vicinity of $z=1$ if the plotting area is too large. This accuracy aspect can best be illustrated if the ζ contours of Fig. 15.8 are used graphically to locate a set of dominant complex roots $p_{1,2}$ corresponding to a desired ζ. Trying to graphically determine the values of the roots at the intersection of the desired ζ contours and the dominant root-locus branches is most difficult. Any slight error in the values of $p_{1,2}$ may result in a pair of dominant roots having a value of ζ that is larger or smaller than the desired value. This problem is also involved even if a computer-aided program is used to locate this intersection, especially if the program is not implemented with the necessary degree of calculation accuracy. Also, the word length (number of binary digits) of the selected digital control processor may not be sufficient to provide the desired damping performance without extended precision. It may be necessary to reduce the plotting area to a sufficiently small region about $z=1$ and then to reduce the calculation step size in order to obtain an accurate picture of the range of K for a stable system performance. This aspect of accuracy is amplified at the end of this section.

c. For a step input $[R(z)=z/(z-1)]$, $C(z)$ is solved from Eq. (15.104). Applying the final-value theorem to $C(z)$ yields

$$c(\infty) = \lim_{z\to 1}[(1-z^{-1})C(z)] = 1$$

Therefore, $e(\infty) = 0$ for a stable system.

For other than step inputs, the steady-state performance characteristics of $E(z)$ for a unity-feedback system must be analysed. Thus,

$$E(z) = \frac{1}{1 + G_z(z)} R(z) \tag{15.107}$$

Considering the ramp input $R(z) = Tz/(z-1)^2$, $E(z)$ for this example is

$$E(z) = \left\{ \frac{2(z-1)(z-e^{-2T})}{2(z-1)(z-e^{-2T}) + K_G[(T + 0.5e^{-2T} - 0.5)z + (0.5 - 0.5e^{-2T} - Te^{-2T})]} \right\} \frac{Tz}{(z-1)^2}$$

Thus, for a stable system,

$$e^*(\infty) = \lim_{z \to 1} [(1 - z^{-1})E(z)] = \left.\frac{1}{K}\right|_{K>0} > 0$$

Therefore, a sampled-data unity-feedback stable control system whose plant $G_x(s)$ is Type 1 has the same steady-state performance characteristics as does a stable unity-feedback continuous-time control system. In a similar manner, an analysis with other polynomial inputs can be made for other Type m plants.

d. The roots of the characteristic equation for $\zeta = 0.6$ are $p_{1,2} = 0.90311 \pm j0.12181$ (where $K = 0.01252$). The output time-response function is

$$c(kT) = 0.01252000r[(k-1)T] + 0.01171246r[(k-2)T] + 1.80621000c[(k-1)T] - 0.83044246c[(k-2)T]$$

and the FOM are $M_p \approx 1.113$, $t_p \approx 2.35$ s, $t_s = 3.6$ s, and $K_1 = 1.3368\ \mathrm{s}^{-1}$. Note that $K_G = 2.6735$. In addition, it should be noted that the roots can also be determined graphically by the use of the ζ contours of Fig. 15.8 as shown in Fig. 5.9 with limited accuracy.

As discussed previously in this section, Figs. 15.21 and 15.22 illustrate the care that must be exercised in performing a mathematical analysis for a sample-data control system for small values of T. For the value of $T = 0.01$ s, the three poles of $G_z(z)$ in Fig. 15.21 appear to be on top of one another (for the plotting scale used in figure). Thus, it is most difficult to locate accurately the dominant poles $p_{1,2}$ for $\zeta = 0.45$ in this figure. An error in the graphical interpretation of the values of $p_{1,2}$ can easily put these poles outside the UC or on another ζ contour. The root-locus plot in Fig. 15.22 corresponds to that portion of the root locus in Fig. 15.21 in the vicinity of the $1 + j0$ point.

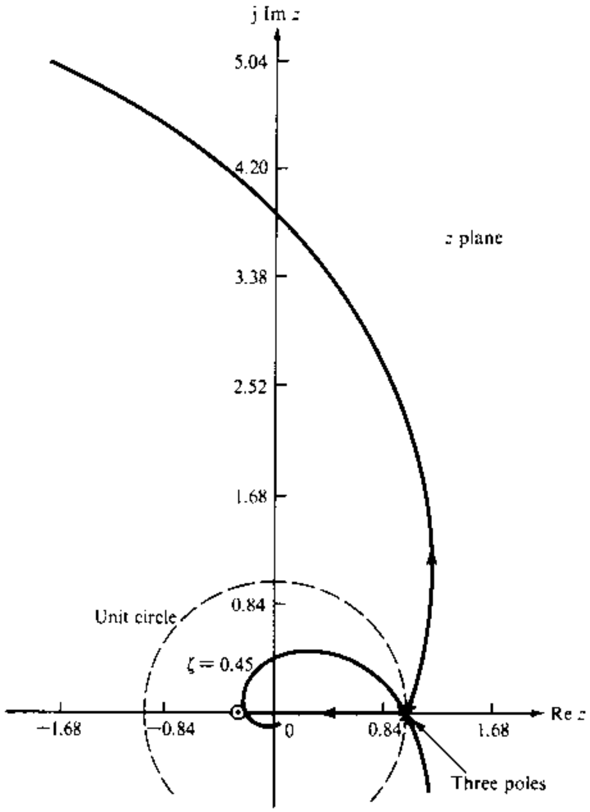


FIGURE 15.21 Root-locus for $G_z(z) = K_z(z + 0.26395)(z + 3.6767)/(z - 1)(z - 0.99005)(z - 0.95123)$, where $T = 0.001$ s. (From Ref. 1, with permission of the McGraw-Hill Companies.)

That is, Fig. 15.22 is an enlargement of the area about the $1 + j0$ point of Fig. 15.21. By "blowing up" this region, one can plot the root locus accurately in the vicinity of the $1 + j0$ point and accurately determine $p_{1,2}$.

15.12 SUMMARY

In this chapter the sampling process associated with an LTI system is analyzed by using the impulse-function representation of the sampled quantity. Linear difference equations are introduced in this chapter and used to model a continuous-time or a sampled-data (S-D) control system based upon the approximation of differentiation. The effective use of the concept of an ideal

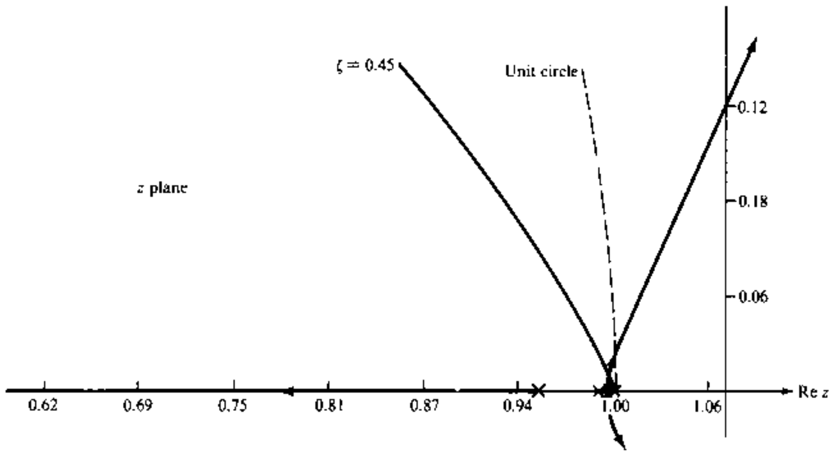


FIGURE 15.22 Enlargement of the $1 + j0$ area of the root-locus plot of Fig. 15.21 ($T = 0.01$ s). (From Ref. 1, with permission of the McGraw-Hill Companies.)

sampler in analyzing sampled-data systems is introduced. An important aspect of a S-D control system is the data conversion process (reconstruction or construction process), which is modelled by a zero-order hold (ZOH) device and the analog-to-digital (A/D) and the digital-to-analog (D/A) conversion devices.

The synthesis in the *z* domain and the associated stability analysis in the *z* domain is presented in this chapter. This is followed by the steady-state analysis of stable S-D control systems. The chapter concludes with the root-locus guidelines for the design of S-D control systems. Numerous examples are also presented.

REFERENCES

1. Houpis, C. H., and Lamont, G. B.: *Digital Control Systems, Theory, Hardware, Software*, 2nd ed., McGraw-Hill, New York, 1992.
2. Franklin, G. G., and Powell, J. D.: *Digital Control of Dynamic Systems*, 2nd ed., Addison-Wesley, Reading, Mass., 1990.
3. Philips, C. L., and Nagle, H. T.: *Digital Control System Analysis and Design*, Prentice-Hall, Englewood Cliffs, N.J., 1984.
4. Ogata, K.: *Discrete-Time Computer Control Systems*, Prentice-Hall, Englewood Cliffs, N.J., 1987.
5. Houpis, C. H., and Lamont, G. B.: *Digital Control Systems: Theory, Hardware, Software*, McGraw-Hill, New York, 1988.

16

Digital Control Systems*

16.1 INTRODUCTION [1,2]

Chapter 15 presents the direct technique for analyzing the basic sampled-data control system. Generally, gain adjustment alone is not sufficient to achieve the desired system performance specifications, i.e., the figures of merit (FOM) (see Secs. 3.9 and 3.10). A cascade and/or feedback compensator (controller) can be used to accomplish the design objectives. This chapter presents the design process for satisfying the specified values of the conventional control-theory FOM by using a cascade or a feedback controller.

As indicated in Sec. 15.1, there are two approaches that may be used in analyzing and improving the performance of an S-D control system; the *direct* (DIR) or the *digitization* (DIG) control design techniques. The main advantage of the DIR design is that the performance specifications can be met with less stringent requirements on the controller parameters. The disadvantage of the method is that there is a limited amount of knowledge, experience, or engineering tools for effecting a suitable $D_c(z)$ controller. Whereas, the DIG design utilizes the well established s domain knowledge, experience, or

*Much of the material in this chapter is taken from Ref. 1 with permission of the McGraw-Hill Companies.

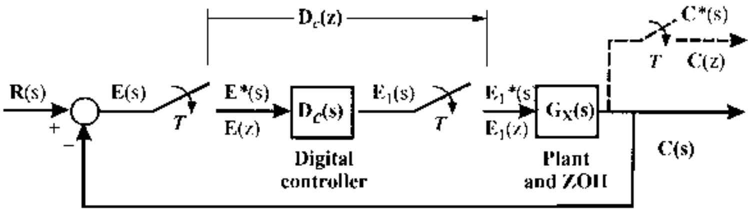


FIGURE 16.1 A compensated digital control system.

engineering design tools. Cascade compensation can be achieved by using a digital controller $D_c(z)$, as shown in Fig. 16.1.

16.2 COMPLEMENTARY SPECTRA [3]

In Sec. 15.2 it is stated that the selection of the sampling time T determines the width of the primary and complementary components of the frequency spectra shown in Fig. 15.6. To avoid the overlapping of these components, see Fig. 15.6*b*, the sampling time should be selected small enough to satisfy the Shannon sampling theorem:

$$\omega_s > 2\omega_c \qquad \text{or} \qquad \frac{\pi}{T} > \omega_c$$

To illustrate the important role that the selection of the sampling time plays, consider the sampling of two sinuoids, which are being sampled with a sampling time T,

$$e_a(t) = \sin(\omega_1 t + \theta) \qquad \text{where} \qquad \omega_1 = 2\pi/T \qquad \text{and} \qquad 0 \le \omega_1 < \omega_s/2 \tag{16.1}$$

and

$$e_b(t) = \sin\left[\left(\frac{j2\pi}{T} - \omega_1\right)t + \pi - \theta\right] \tag{16.2}$$

where $j = 1, 2, \ldots$. Their corresponding sampled expressions, with $t = kT$, are

$$e_a(kT) = \sin(k\omega_1 T + \theta) \tag{16.3}$$

$$e_b(kT) = \sin(j2k\pi - kT\omega_1 + \pi - \theta) = \sin(kT\omega_1 + \theta) \tag{16.4}$$

Equations (16.3) and (16.4) reveal that sampling either Eq. (16.1) or (16.2) results in the identical sequence. Thus, this example reveals it is impossible to differentiate between two sinuoids whose sum (or difference) of radian frequencies is an integral multiple of $2\pi/T$. The frequency π/T rad/s is referred to as the *folding* or *Nyquist frequency.*

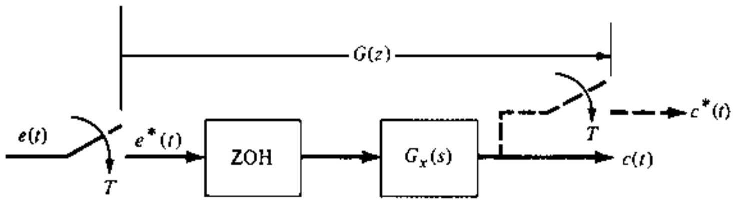


FIGURE 16.2 An open-loop sampled-data system.

To illustrate the effect of folding (or *aliasing*) [3] two sampling times, $T = 0.1$ s and $T = 1$ s, are consider for the open-loop S-D system of Fig. 16.2 where $G_x(s) = K_x/[s(s + 2 \pm j5)]$. For $T = 0.1$ s, the poles $p_{1,2} = -2 \pm j5$ lie inside the primary strip that exists between $+j\omega_s/2 = +j\pi/T = +j31.4$ and $-j\omega_s/2 = -j\pi/T = -j31.4$; thus no folding action exists. For $T = 1$ s, the poles $p_{1,2} = -2 \pm j5$ lie in a complementary strip (outside the primary strip) that exists between $+j\omega_s/2 = j\pi/T = j3.14$ and $-j\omega_s/2 = -j\pi/T = -j3.14$. Thus, the poles $p_1 = -2 + j5$ and $p_2 = -2 - j5$ are folded into $p_{1f} = -2 - j1.283$ and $p_{2f} = -2 + j1.283$, respectively.

16.3 TUSTIN TRANSFORMATION: s TO z PLANE TRANSFORMATION [1]

The transformation from the s domain to the z domain given by $z = e^{Ts}$ is a transcedental relationship. A linear approximation for this transformation is obtained by use of the Tustin algorithm. Solving for s in terms of z yields the function of z that is substituted for s in implementing the Tustin transformation:

$$s = \frac{1}{T}\ln z \tag{16.5}$$

The natural logarithm $\ln z$ is expanded into the series

$$\ln z = 2\left(x + \frac{1}{3}x^3 + \frac{1}{5}x^5 + \cdots\right) \tag{16.6}$$

where

$$x = \frac{1 - z^{-1}}{1 + z^{-1}} \tag{16.7}$$

Using only the first term in the infinite series of Eq. (16.6) yields the Tustin transformation

$$s \equiv \frac{2}{T}\frac{1 - z^{-1}}{1 + z^{-1}} = \frac{2}{T}\frac{z - 1}{z + 1} \tag{16.8}$$

The expression for s in Eq. (16.8) can be inserted into a function such as $G(s)$, which represents a continuous-time function. Rationalizing the expression yields the function $G(z)$, which represents a discrete function. The root-locus and frequency response methods that have been developed for continuous-time systems, can now be applied to the design of discrete-time systems.

An advantage of the Tustin algorithm is that it is comparatively easy to implement. Also, the accuracy of the response of the Tustin z-domain transfer function is good compared with the response of the exact z-domain transfer function; i.e., the accuracy increases as the frequency increases.

16.3.1 Tustin Transformation Properties

Ideally, the z-domain transfer function obtained by digitizing an s-domain transfer function should retain all the properties of the s-domain transfer function. However, this is not achieved completely. The transformation has four properties that are of particular interest to the designer. These four properties are the cascading property, stability, dc gain, and the impulse response. The first three properties are maintained by the Tustin transformation; i.e., these three properties are invariant. The Tustin algorithm maintains the cascading property of the s-domain transfer function. Thus, cascading two z-domain Tustin transfer functions yields the same results as cascading the s-domain counterparts and Tustin-transforming the result to the z-domain. This cascading property implies that no ZOH device is required when cascading s-domain transfer functions. The Tustin transformation of a stable s-domain transfer function is also stable. The dc gains for the s-domain and the z-domain Tustin transfer functions are identical; that is, $F(s)|_{s=0} = F(z)|_{z=1}$. However, the dc gains for the exact $\mathcal{Z}$ transfer functions are not identical.

Example 16.1. This example illustrates the properties of the Tustin transformation of a function $e(t)$ that is sampled as shown in Fig. 16.3. Given $E(s) = E_x(s) = 1/[(s+1)(s+2)]$, *with no ZOH involved*, and $T = 0.1$ s, the exact $\mathcal{Z}$ transform obtained by using Eq. (15.12) yields

$$E(z) = \mathcal{Z}[e^*(t)] = \frac{0.086107z}{(z - 0.81873)(z - 0.90484)}$$

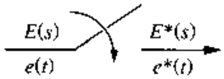


FIGURE 16.3 A sampled function.

The Tustin transformation, indicated by the subscripts TU, is obtained by inserting for s from Eq. (16.8) into $E(s)$ and simplifying the resultant equation yields

$$[E(z)]_{\mathrm{TU}} = \frac{0.0021645(z+1)^2}{(z-0.81818)(z-0.90476)}$$

For each transform, the initial- and final-value theorems yield, respectively, the results shown in the table below.

	$E(s)$	$E(z)$	$[E(z)]_{\mathrm{TU}}$	$\frac{1}{T}[E(z)]_{\mathrm{TU}}$
Initial value	0	0	0.0021645	0.021646
Final value	0	0	0	0
dc gain	0.5	4.9918	0.499985	4.99985
n (degree of denominator)	2	2	2	2
w (degree of numerator)	0	1	2	2

This example demonstrates that the Tustin transformation results in a $\mathcal{Z}$ transform for which, in general, the degrees of the numerator and denominator are the same, i.e., $w = n$. Thus, since $n = w$, the initial value of $[E(z)]_{\mathrm{TU}}$ is always different from zero; that is,

$$\lim_{z\to\infty} E(z) \neq \lim_{z\to\infty} [E(z)]_{\mathrm{TU}} \tag{16.9}$$

Based upon the dc gain, it is evident that the values of $E(z)$ and $[E(z)]_{\mathrm{TU}}$ differ by a factor of $1/T$; that is,

$$E(z) = \mathcal{Z}[e^*(t)] \approx \frac{1}{T}[E(z)]_{\mathrm{TU}} \tag{16.10}$$

The reason for this difference is that the Tustin transformation does not take into account the attenuation factor $1/T$ introduced by the sampling process (see Fig. 15.6), whereas $E(z) = \mathcal{Z}[e^*(t)]$ does take this attenuation factor into account. This factor must be included when the DIG technique is used for system analysis and design.

Example 16.2. To illustrate further the Tustin properties, consider the open-loop sampled-data system of Fig. 16.2 with $T = 0.1$ s that includes a ZOH unit and

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

Thus,

$$G(z) = \mathcal{Z}[G_{zo}(s)G_x(s)] = \frac{0.105(z-0.8171)}{(z-1)(z-0.904)}$$

Applying Eq. (16.8) to $G_x(s)$ yields

$$[G_x(z)]_{TU} = \frac{0.0524(z+1)(z-0.8182)}{(z-1)(z-0.9048)} \approx G(z)$$

The initial-value and final-value theorems yield, respectively, for each transform, the results shown in the following table.

Theorem	$G(z)$	$[G(z)]_{TU}$
I.V.	0	0.0524
F.V.	0.20047	0.200133

Therefore, for this example, since a ZOH unit is involved, the exact $\mathcal{Z}$ transform of $G(s)$, takes into account the factor $1/T$ shown in Fig. 15.6, As a result, as noted in Eq. (16.10), $G(z) \cong (1/T)[G(z)]_{TU}$.

16.3.2 Tustin Mapping Properties

The following discussion is useful in understanding the Tustin mapping properties. To represent functionally the s to z plane mapping, Eq. (16.8) is rearranged to yield

$$z = \frac{1+sT/2}{1-sT/2} \tag{16.11}$$

Utilizing the mathematical expression,

$$\frac{1+a}{1-a} = e^{j2\tan^{-1}a}$$

and letting $s = j\hat{\omega}_{sp}$ in Eq. (16.11) the following expression is obtained [1]:

$$z = \frac{1+j\hat{\omega}_{sp}T/2}{1-j\hat{\omega}_{sp}T/2} = \left(e^{j2\tan^{-1}\hat{\omega}_{sp}T/2}\right) \tag{16.12}$$

The exact $\mathcal{Z}$ transform yields $z = e^{j\omega_{sp}T}$, where ω_{sp} is as equivalent s plane frequency. Thus, the following equation is obtained from Eq. (16.12):

$$e^{j\omega_{sp}T} = \exp\left(e^{j2\tan^{-1}}\frac{\hat{\omega}_{sp}T}{2}\right) \tag{16.13}$$

Equating the exponents yields

$$\frac{\omega_{sp}T}{2} = \tan^{-1}\frac{\hat{\omega}_{sp}T}{2} \tag{16.14}$$

or

$$\tan\frac{\omega_{sp}T}{2} = \frac{\hat{\omega}_{sp}T}{2} \tag{16.15}$$

When $\omega_{sp}T/2 < 17^\circ$, or ≈ 0.30 rad, then

$$\omega_{sp} \approx \hat{\omega}_{sp} \tag{16.16}$$

This means that in the frequency domain the Tustin approximation is good for small values of $\omega_{sp}T/2$.

Returning to Eq. (16.13), it is easy to realize that the *s*-plane imaginary axis is mapped into the *unit circle* (UC) in the *z* plane, as shown in Fig. 15.10. The left-half (LH) *s* plane is mapped into the inside of the UC. The same stability regions exist for the exact $\mathcal{Z}$ transform and the Tustin approximation.

Also, in this approximation, the entire *s*-plane imaginary axis is mapped once and only once onto the UC. The Tustin approximation prevents pole and zero aliasing since the folding phenomenon does not occur with this method. However, there is a warping penalty [1], that is, the bilinear transformation does not result in a 1:1 relationship between the location of the corresponding poles and zeros in the *s* plane and the *z* plane. In other words, the location of the transformed *z*-plane pole or zero corresponds to the *warped* location of the corresponding pole or zero in the *s* plane. This warping phenomena is illustrated in Fig. 16.4. Compensation can be accomplished by using Eq. (16.15), which is depicted in Fig. 16.4. To compensate for the warping, pre-warping of ω_{sp} by using Eq. (16.15) generates $\hat{\omega}_{sp}$, which will result in a 1:1 relationship between the location of the corresponding poles and zeros in the *s* plane and the *z* plane. The continuous controller is mapped into the *z* plane by means of Eq. (16.8) using the prewarped frequency $\hat{\omega}_{sp}$. The digital compensator (controller) must be tuned (i.e., its numerical coefficients adjusted) to finalize the design since approximations have been employed. As seen from Fig. 16.4, Eq. (16.16) is a good approximation when $\omega_{sp}T/2$ and $\hat{\omega}_{sp}T/2$ are both less than 0.3 rad.

The prewarping approach for the Tustin approximation takes the *s* plane imaginary axis and folds it back to the range $\pi/2$ to $-\pi/2$, as seen in Fig. 16.4. The spectrum of the input must also be taken into consideration when selecting an approximation procedure with or without prewarping. It should be noted that in the previous discussion only the frequency has been prewarped due to the interest in the controller frequency response. The real part of the *s* plane pole influences such parameters as rise time, overshoot,

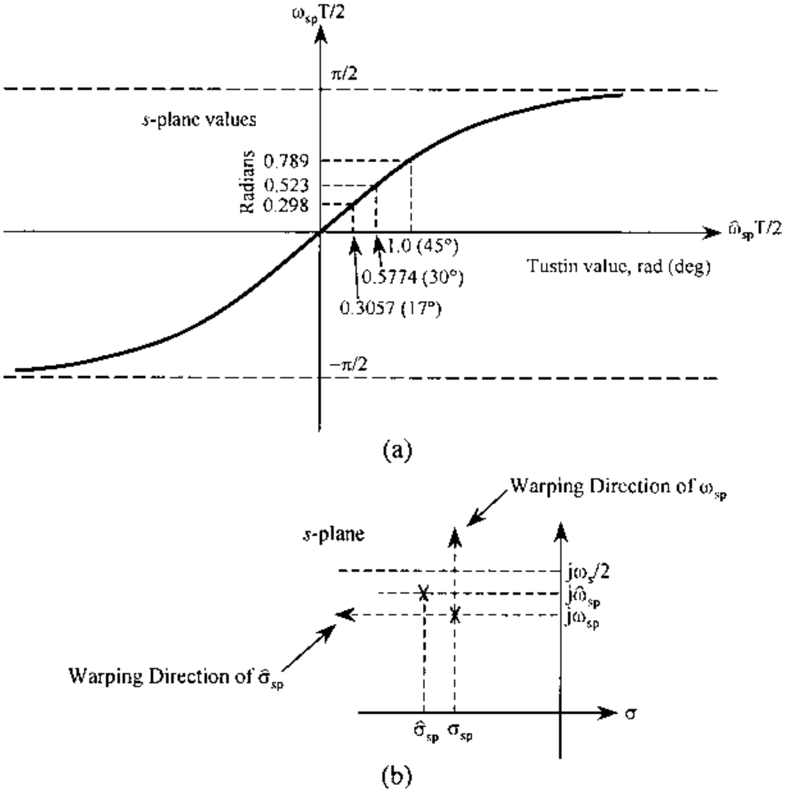


FIGURE 16.4 Map of $\hat{\omega} = 2(\tan\omega_{sp}T/2)/T$. (*a*) Plot of Eq. (16.13) and (*b*) warping effect.

and settling time. Thus, consideration of the warping of the real pole component is now analyzed as a fine-tuning approach. Proceeding in the same manner as used in deriving Eq. (16.16), substitute $z = e^{\sigma_{sp}T}$ and $s = \hat{\sigma}_{sp}$ into Eq. (16.11) to yield

$$e^{\sigma_{sp}T} = \frac{1 + \hat{\sigma}_{sp}T/2}{1 - \hat{\sigma}_{sp}T/2} \tag{16.17}$$

Replacing $e^{\sigma_{sp}T}$ by its exponential series and dividing the numerator by the denominator in Eq. (16.17) results in the expression

$$1 + \sigma_{sp}T + \frac{(\sigma_{sp}T)^2}{2} + \cdots = 1 + \frac{\hat{\sigma}_{sp}T}{1 - \hat{\sigma}_{sp}T/2} \tag{16.18}$$

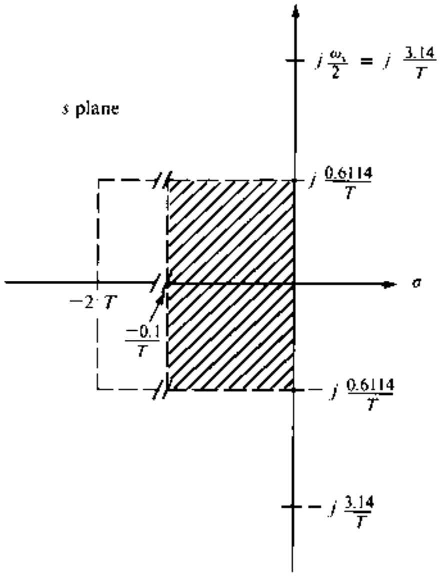


FIGURE 16.5 Allowable location (crosshatched area) of dominant poles and zeros in *s* plane for a good Tustin approximation.

If $|\sigma_{sp}T| \gg (\sigma_{sp}T)^2/2$ (or $1 \gg |\sigma_{sp}T/2|$) and $1 \gg |\hat{\sigma}_{sp}T/2|$, then

$$|\hat{\sigma}_{sp}| \approx |\sigma_{sp}| \ll \frac{2}{T} \tag{16.19}$$

Thus, with Eqs. (16.16) and (16.19) satisfied, the Tustin approximation in the *s* domain is good for small magnitudes of the real and imaginary components of the variable *s*. The shaded area in Fig. 16.5 represents the allowable location of the poles and zeros in the *s* plane for a good Tustin approximation. Because of the mapping properties and its ease of use, the Tustin transformation is employed for the *DIG* technique in this text. Fig. 16.4*b* illustrates the warping effect of a pole (or zero) when the approximations are not satisfied.

A *matched $\mathcal{Z}$ transform* can be defined as a direct mapping of each *s*-domain root to a *z*-domain root: $s + a \rightarrow 1 - e^{-aT}z^{-1}$. The poles of $G_z(z)$ using this approach are identical to those resulting from the exact $\mathcal{Z}$ transformation of a given $G(s)$. However, the zeros and the dc gain are usually different.

A characteristic of a bilinear transformation is that, in general, it transforms an unequal-order transfer function ($n_s \neq w_s$) in the *s* domain into one for which the order of the numerator is equal to the order of its denominator ($n_z = w_z$) in the *z* domain. This characteristic must be kept in mind when synthesizing $G(s)$ and $F(s)$.

16.4 z-DOMAIN TO THE w- AND w'-DOMAIN TRANSFORMATIONS [1]

The initial use of the DIG technique involved the bilinear transformation from the z domain to the w domain by use of

$$z = \frac{w+1}{-w+1} \tag{16.20}$$

A disadvantage of this mapping is that it lacks the desirable property that as the sampling time $T \to 0$, $w \to s$; i.e.,

$$w|_{T\to 0} = \lim_{T\to 0}\left[\frac{z-1}{z+1} = \frac{e^{sT}-1}{e^{sT}+1} = \frac{sT + (sT)^2/2! + \cdots}{2 + sT + (sT)^2/2! + \cdots}\right] = 0 \tag{16.21}$$

The desired property that as the sampling time $T \to 0$, $w \to s$ is achieved by defining

$$w' \equiv \frac{2}{T} w \equiv \frac{2}{T}\left[\frac{sT + (sT)^2/2! + \cdots}{2 + sT + (sT)^2/2! + \cdots}\right] \tag{16.22}$$

This desirable property yields a quantity in the w' domain that is analogous to a quantity in the s domain. To achieve the z to w' transformation, substitute $w = Tw'/2$ into Eq. (16.20). This results in

$$z = \frac{Tw' + 2}{-Tw' + 2} \tag{16.23}$$

Therefore, the w' DIG design technique utilizes this equation. *Henceforth the prime designator on* w' *is omitted*; i.e., it is designated as w.

The bilinear transformation from the z domain to the w domain may be accomplished by using Eq. (16.11) with s replaced *by* w or by use of Eq. (16.23). Both s and w bilinear transformations are referred to as a *Tustin transformation*. The w-domain mapping is illustrated in Fig. 16.6 where the mapping from the s domain into the z domain is accomplished by means of $z = e^{Ts}$. The transformation into the w domain results in a minimum-phase (m.p.) s domain transfer function becoming a nonminimum-phase (n.m.p.) w domain transfer function. In performing a w-domain control system design, the w-domain control system can be treated as being a "continuous-time" system. Thus, the same CAD packages utilized for s-domain analysis and design can be used.

Besides the n.m.p. characteristic of the w-domain transfer functions, another disadvantage in transforming the resultant digital controller into the z domain is that the resulting warping effect is in the opposite direction to that

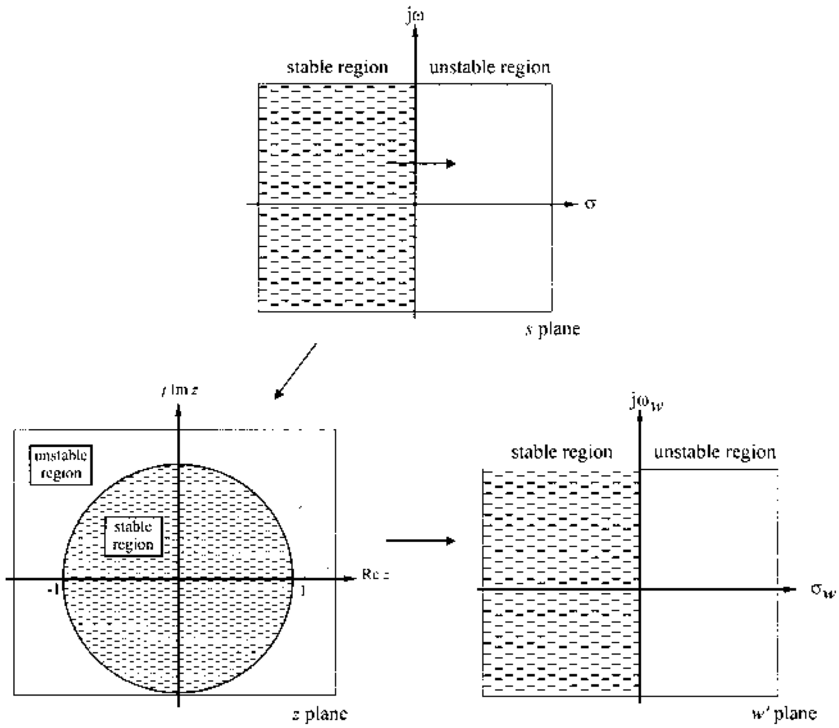


FIGURE 16.6 Mapping of the s plane into the z plane into the w' plane.

shown in Fig. 16.4*b*. That is, the warping direction is toward the imaginary axis, resulting in a less stable system.

16.5 DIGITIZATION (DIG) TECHNIQUE

There are two approaches available for utilizing the DIG method. The first approach is to transform the $\mathcal{Z}$-transform function by use of a bilinear transformation, as discussed in Sec. 16.3. The second approach is to transform into the w domain as discussed in Sec. 16.4. In the DIG method, once the controller design is achieved in the s or w domain, one of several transformation methods can be employed to transform the cascade or feedback controller, with or without the s- or w-domain controller gain included, into its equivalent z-domain discrete controller. The reason for not including the gain is discussed in the following section. The advantage of this technique, discussed in detail in this chapter, is that tried and proven continuous-time domain methods are used for designing an acceptable $D_c(s)$ or $D_c(w)$ controller.

A disadvantage is that, if T is not small enough, the poles and zeros of the controller lay outside the good bilinear transformation region, which may result in not achieving the desired performance.

After $D_c(z)$ is included in the system model, as shown in Fig. 16.1, a system analysis is again performed. If the modified system does not meet the desired specifications, the controller is modified accordingly. This process is continued until the desired system specifications are achieved. To illustrate this design process, the basic system of Sec. 15.6.3 is used in this chapter for the design of lead, lag, and lag-lead or lead-lag cascade controllers.

The disadvantage of performing the analysis and design in the w domain is that, when transforming a m.p. s-domain transfer function, it becomes a non-minimum-phase (n.m.p.) transfer function in the w domain. Another disadvantage is the warping of the poles and zeros that occurs when transforming the w domain transfer function into the z domain. Performing the PCT analysis and design in the s domain eliminates the first disadvantage. Although warping also occurs with the s-domain approach, its effect tends to make the system more stable, as shown in Fig. 16.4*b*. Also, the warping effect is minimized the smaller that the value of the sampling time T becomes. The direction of the warping effect when using the w-domain design is opposite to that which occurs with the s domain design; i.e., the w-domain poles and zeros are warped toward the imaginary axis, thus decreasing the system's degree of stability. Thus, the PCT DIG technique, discussed in Sec. 16.7, is used for the compensation designs done in the remainder portion of this chapter.

16.6 DIGITIZATION (DIG) DESIGN TECHNIQUE

An important tool for any design technique is a CAD package. As the value of T decreases, the use of a computer becomes essential in order to achieve an accurate design of the digital controller. The presentation of the cascade-compensator design procedure outlined and discussed in the following sections is intended to outline clearly the manner in which the DIG technique is applied to the design of the compensator $D_c(\cdot) = (\text{gain})D'_c(\cdot)$.

Figure 16.7 represents the trial-and-error design philosophy in applying the DIG technique in the design of a cascade compensator via the analysis of the open-loop system. If path A does not result in the specifications set forth by the sampled-data control system of Fig. 16.1, then path B is used to try o determine a satisfactory value of K_{zc}. Thus, the choice of the design path selected from Fig. 16.7 is based upon the following guidelines:

1. Follow path A if the dominant poles and zeros of $C(\bullet)/R(\bullet)$ lie in the crosshatched area of Fig. 16.5. (Tustin approximation is good!)

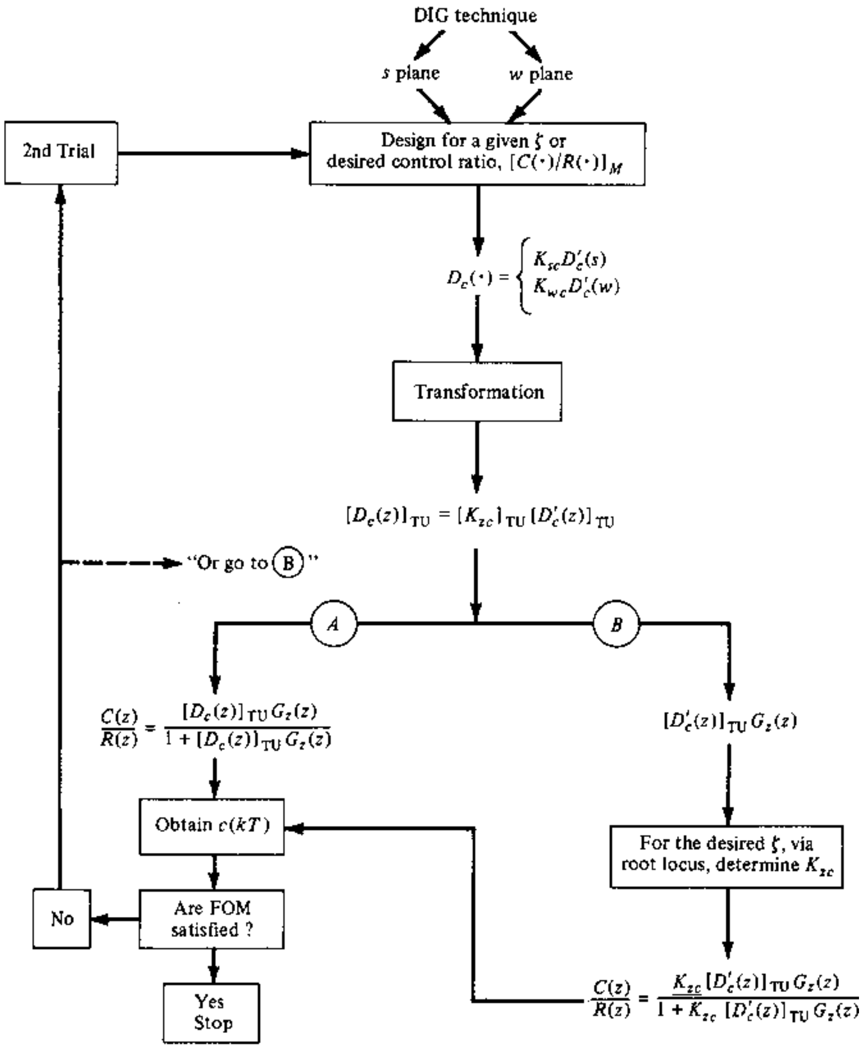


FIGURE 16.7 Digitization (DIG) design philosophy for Fig. 16.1.

2. Follow path *A* when the degree of warping is deemed not to affect negatively the achievement of the desired design results (see Sec. 16.12). If the desired results are not achieved, try path *B*.
3. Follow path *B* when severe warping exists.

The DIG design procedure is as follows:

Step 1. Convert the basic sampled-data control system to a PCT control system or transform the basic system into the *w* plane. (Try both approaches for determining the best design.)

Step 2. By means of a root-locus analysis, determine $D_c(s) = K_{sc}D_c(s)$.

Step 3. Obtain the control ratio of the compensated system and the corresponding time response for the desired forcing function. If the desired FOM are not achieved, repeat step 2 by selecting a different value of ζ, σ, ω_d, etc., or a different desired control ratio.

Step 4. When an acceptable $D_c(s)$ or $D_c(w)$ has been achieved, transform the compensator, via the Tustin transformation, into the *z* domain. *Note*: In order for real poles (or zeros) [p_i (or z_i)] that lie on the negative real axis of the *s* plane to lie between 0 and 1 in the *z* plane, *when utilizing the Tustin transformation of Eq. (16.8)*, $|p_i| \le 2/T$ (or $|z_i| \le 2/T$) must be satisfied.

Step 5. Obtain the *z*-domain control ratio of the compensated system and the corresponding time response for the desired forcing function. If the FOM for the sampled-data control system have been achieved via path *A* or *B*, then the design of the compensator is complete. If not, return to step 2 and repeat the steps with a new compensator design or proceed to the DIR technique.

The following sections present the DIG PCT design technique that requires the utilization of the Tustin tranformation.

16.7 THE PSUEDO-CONTINUOUS-TIME (PCT) CONTROL SYSTEM

As noted in Sec. 16.1, the transformation of a m.p. *s*-domain transfer function results in a n.m.p. transfer function in the *w* domain. When the requirements in Sec. 16.3.2 for a *psuedo-continuous-time* (PCT) representation of a S-D system are satisfied and if a *plant* is m.p., then the simpler analog design procedures of this text can be applied to a m.p. PCT system. With the sampling rates that are now available, the dominant *s*-plane poles and zeros that determine the system's FOM are essentially not warped when transformed into the *z* domain. These twenty-first-century sampling rates enable the design of a digital controller by first designing as *s* domain compensator $D_c(s)$, by the well-established *s*-domain cascade and feedback compensation design techniques. Once $D_c(s)$ that yields the desired system FOM is synthesized, then it is transformed into the *z* domain by the use of the Tustin transformation to obtain $[D_c(z)]_{TU}$. Thus, using the minimal practical sampling time allowable assures that the desired FOM for the designed digital control system will be achieved. This PCT design approach is also referred to as a DIG design method.

16.7.1 Introduction to Psuedo-Continuous-Time System DIG Technique [1]

The DIG method of designing an S-D system, in the complex-frequency s plane, requires a satisfactory pseudo-continuous-time (PCT) model of the S-D system. In other words, the sampler and the ZOH unit for the S-D system of Fig. 16.8 (also shown in Fig. 16.9*a*) must be approximated by the linear continuous-time unit $G_A(s)$ shown in Fig. 16.9*c*. The DIG method requires that the dominant poles and zeros of the PCT model lie in the shaded area of Fig. 16.5 in order to achieve a high level of correlation with the S-D system. To determine $G_A(s)$, first consider the frequency component E*($j\omega$) representing the continuous-time signal E($j\omega$), where all its side-bands are multiplied by the factor $1/T$ [see Eq. (6.9) of Ref. 1]. Because of the low-pass filtering characteristics of a S-D system, only the primary component needs to be considered in the analysis of the system. Therefore, the PCT approximation of the sampler of Fig. 16.8 is shown in the Fig. 16.9*b*.

Using the first-order Pade′ approximation (see Ref. 1), the transfer function of the ZOH device, when the value of T is small enough, is approximated as follows:

$$G_{zo}(s) = \frac{1 - e^{-Ts}}{s} \approx \frac{2T}{Ts + 2} = G_{pa}(s) \tag{16.24}$$

Thus, the Padé approximation $G_{pa}(s)$ is used to replace $G_{zo}(s)$, as shown in Fig. 16.9*a*. This approximation is good for $\omega_c \le \omega_s/10$, whereas the second-order approximation is good for $\omega_c \le \omega_s/3$ (Ref. 1). Therefore, the sampler

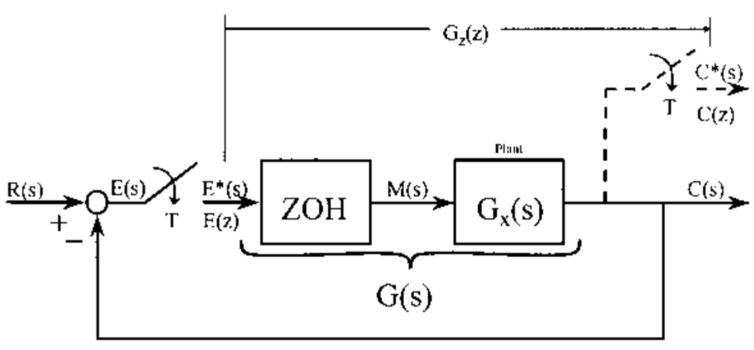


FIGURE 16.8 The uncompensated sampled-data control system.

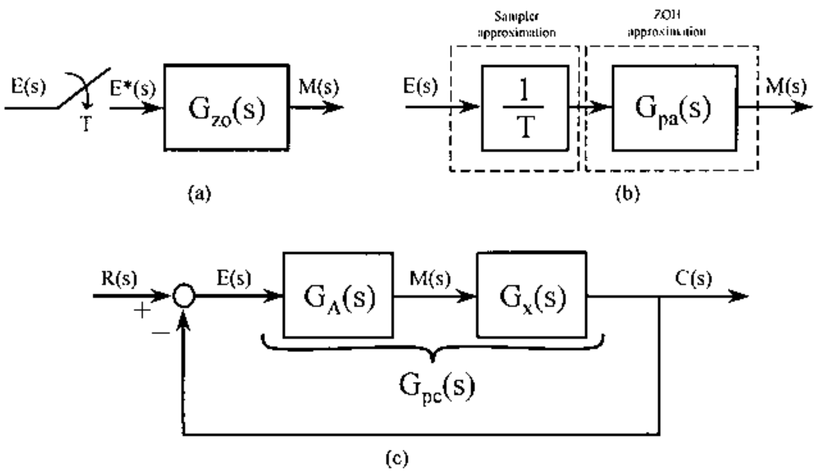


FIGURE 16.9 (*a*) Sampler and ZOH; (*b*) approximations of the sampler and ZOH; (*c*) the approximate continuous-time control system equivalent of Fig. 16.8.

and ZOH units of a S-D system are approximated in the PCT system of Fig. 16.9*c* by the transfer function.

$$G_A(s) = \frac{1}{T} G_{\text{pa}}(s) = \frac{2}{Ts + 2} \tag{16.25}$$

Since Eq. (16.25) satisfies the condition $\lim_{T\to 0} G_A(s) = 1$, it is an accurate PCT representation of the sampler and ZOH units, because it satisfies the requirement that as $T \to 0$ the output of $G_A(s)$ must equal its input. Further, note that in the frequency domain as $\omega_s \to \infty (T \to 0)$, then the primary strip becomes the entire frequency-spectrum domain, which is the representation for the continuous-time system [1].

Note that in obtaining the PCT model for a digital control system, the factor $1/T$ replaces only the sampler that is sampling the continuous-time signal. This multiplier of $1/T$ attenuates the fundamental frequency of the sampled signal, and all its harmonics are attenuated. To illustrate the effect of the value of T on the validity of the results obtained by the DIG method, consider the S-D closed-looped control system of Fig. 16.8, where

$$G_x(s) = \frac{4.2}{s(s+1)(s+5)} \tag{16.26}$$

The closed-loop system performance for three values of T is determined in both the s and z domains, i.e., the DIG technique and the *direct* (DIR) *technique* (the z analysis), respectively. Table 16.1 presents the required value of K_x and time-response characteristics for each value of T obtained by use a CAD

TABLE 16.1 Analysis of a PCT System Representing a Sampled-Data Control System for $\zeta = 0.45$ (Values obtained by use of TOTAL-PC)

Method	T, s	Domain	K_x	M_p	T_p, s	T_s, s
DIR	0.01	z	4.147	1.202	4.16	9.53
DIG		s	4.215	1.206	4.11	9.478
DIR	0.1	z	3.892	1.202	4.2–4.3	9.8+
DIG		s	3.906	1.203	4.33	9.90+
DIR	1	z	2.4393	1.199	6	13–14
DIG		s	2.496	1.200	6.18	13.76

package. Note that for $T \leq 0.1$ s there is a high level of correlation between the DIG and DIR models. For $T \leq 1$ s there is still a relatively good correlation. (The designer needs to specify, for a given application, what is considered to be "good correlation.")

16.7.2 MATLAB Design for Sec. 16.7.1

MATLAB provides a number of options for design. The root locus tool described in Appendix C allows a user to design compensators and adjust gain on a root locus plot. The following MATLAB commands are used to launch the root locus tool with the models from this section and create the closed-loop time responses shown in Fig. 16.10. The following detailed MATLAB commands, for $T = 0.1$ s, involves both the DIG (PCT) and the DIR (exact z-domain) designs.

DIG (PCT) Design

```
echo on
gxnum=4.2;                           % Create numerator
gxden=conv([1 1 0],[1 5]);           % Create denominator
Gx=tf(gxnum,gxden);                  % Create transfer function
                                     %  from num and den
zpk(Gx)                              % Show zero-pole-gain form

Zero/pole/gain:                      % Eq. (16.26)
     4.2
----------
 s(s+5)(s+1)

Tsamp=0.1;                           % Tsamp is changed for
                                     % sample time=0.01 or 1.0
Ga=tf([2],[Tsamp 2])                 % Create sampler and ZOH
```

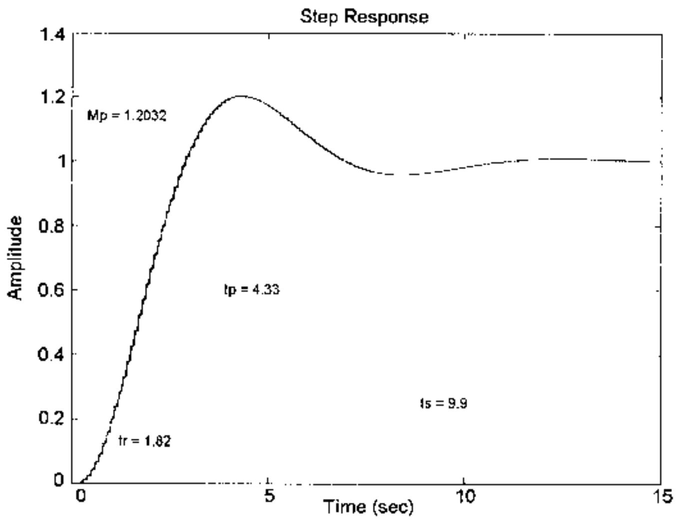


FIGURE 16.10 MATLAB step response of DIG (PCT) vs. DIR technique for $T=0.1$ s.

```
Transfer function:
   2
-----
0.1 s+2

GDIG=series(GA, Gx);      % Multiply GA and Gx
zpk(GDIG)                 % Show zero-pole-gain form

Zero/pole/gain:
          84
---------------
s (s+20) (s+5) (s+1)

rltool(GDIG)              % Open root locus tool window

% Root locus tool is used to design compensator and
%   export the closed loop transfer function T_r2y

T_r2y                     % Closed loop transfer function

Zero/pole/gain from input ''r'' to output ''y'':
                    78.12
----------------------------------------
(s+5.238)  (s+19.99)  (s^2+0.7754s+0.7462)

cltf=tf(T_r2y)            % Create transfer function form
                          % of cltf
```

```
Transfer function from input ''r'' to output ''y''
            78.12
- - - - - - - - - - - - - - - - - - - - - -
s^4 + 26 s^3 + 125 s^2 + 100 s + 78.12

fom(cltf.num{1, 1},cltf.den{1,1})   % Get figures of merit

Figures of merit for a unit step input

rise time          = 1.82
peak value         = 1.2032
peak time          = 4.33
settling time      = 9.9
final value        = 1
```

DIR Design

```
echo on
gxnum = 4.2;                              % Creat numerator
gxden=conv([1 1 0],[1 5])                 % Create denominator
Gx=tf (gxnum, gxden);                     % Create transfer function
                                          % from num and den
zpk(Gx)                                   % Show zero-pole-gain form
                                          % Ref. Eq. (16.26)

Zero/pole/gain:
     4.2
- - - - - - - - - -
 s(s+5)(s+1)

Tsamp = 0.1;                    % Tsamp is changed for
                                % sample time = 0.01, 0.1, and 1.0
dirgx = c2d(Gx,Tsamp,'zoh')     % Transform Gx to the z domain
                                % using a zero order hold

Transfer function:
0.000605 z^2 + 0.002092 z + 0.0004483
- - - - - - - - - - - - - - - - - - - - - - - - -
  z^3 - 2.511 z^2 + 2.06 z - 0.5488

Sampling time: 0.1
zpk(dirgx)                      % Show zero - pole - gain form

Zero/pole/gain:
0.000605 (z + 3.228) (z + 0.2295)
- - - - - - - - - - - - - - - - - - - - - -
 (z - 1) (z - 0.9048) (z - 0.6065)

Sampling time: 0.1
rltool(dirgx)                   % Open root locus tool
% Root locus tool is used to design compensator and
% export the closed loop transfer function T_r2y
%
```

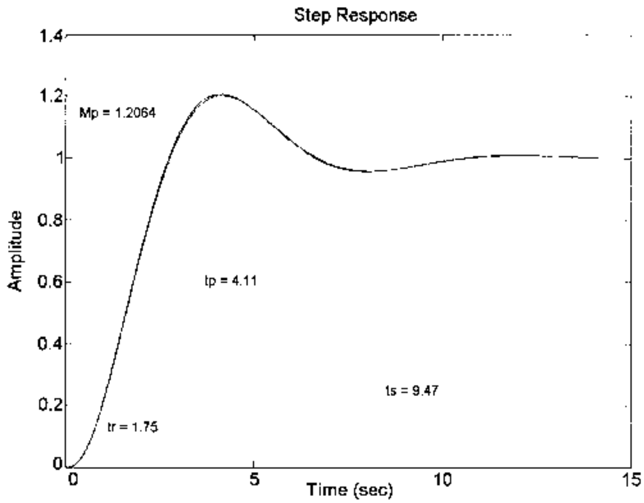


FIGURE 16.11 MATLAB step response of DIG (PCT) vs. DIR technique for $T = 0.01$ s.

```
% The following commands complete the plot shown
%in Fig. 16.10

T_r2y                          % Closed loop transfer function

Zero/pole/gain from input ''r'' to output ''y'':
0.00056063 (z+3.228) (z+0.2295)
- - - - - - - - - - - - - - - - - - - - - -
(z-0.5927) (z^2-1.918 z+0.9253)

Sampling time: 0.1
hold on
step (T_r2y)
```

The results of the MATLAB design for $T = 0.01$ and 1.0 s are shown in Figs. 16.11 and 16.12, respectively. These results agree with the values given in Table 10.1 which were obtained by use of the TOTAL PC CAD package. Figure 16.10 and 16.11 for $T \leq 0.1$ s illustrate that the PCT designs agree very closely to the DIR design results.

16.7.3 Simple PCT Example

Figure 16.8 represents a basic or uncompensated S-D control system where

$$G_x(s) = \frac{K_x}{s(s+1)} \tag{16.27}$$

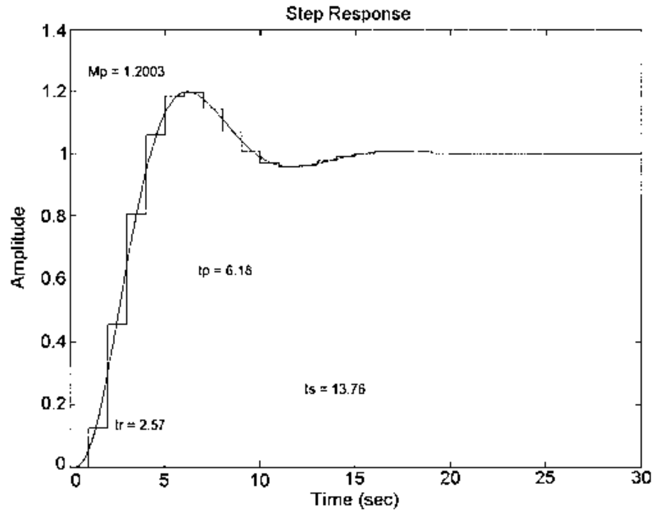


FIGURE 16.12 MATLAB step response of DIG (PCT) vs. DIR technique for $T = 1.0$ s.

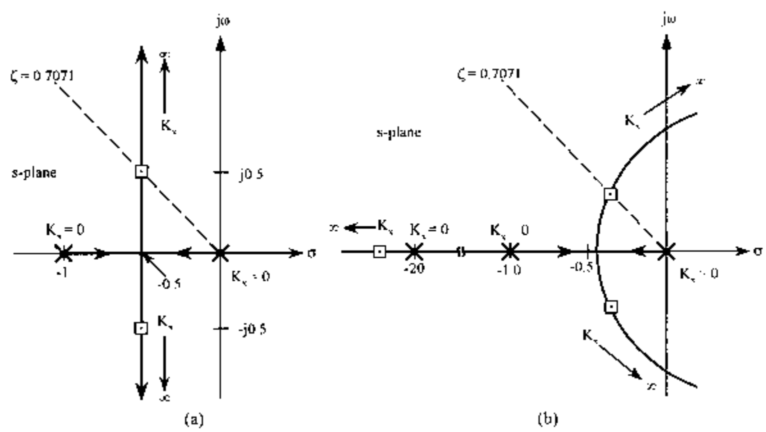


FIGURE 16.13 Root locus for (*a*) Case 1, Eq. (16.27); (*b*) Case 2, Eq. (16.28).

is used to illustrate the approaches for improving the performance of a basic system.

Case 1: PCT Open-Loop Control System

For $\zeta = 0.7071$, the root-locus plot $G_x(s) = -1$ shown in Fig. 16.13*a* yields $K_x = 0.4767$.

One approach for designing a S-D unity-feedback control system is first to obtain a suitable closed-loop model $[C(s)/R(s)]_{\mathrm{TU}}$ for the PCT unity-feedback control system of Fig. 16.8, based on the plant of the S-D control system. This model is then used as a guide for selecting an acceptable $C(z)/R(z)$. Thus, for the plant of Eq. (16.27) the open-loop transfer function is

$$G_{\mathrm{PC}}(s) = G_A(s)G_x(s) = \frac{2K_x/T}{s(s+1)(s+2/T)} \tag{16.28}$$

For $T = 0.1$ s, the closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{G_{\mathrm{PC}}(s)}{1+G_{\mathrm{PC}}(s)} = \frac{20K_x}{s^3 + 21s^2 + 20s + 20K_x} \tag{16.29}$$

The root locus for $G_{\mathrm{PC}}(s) = -1$ is shown in Fig. 16.13*b*. For comparison purpose the root-locus plot of $G_x(s) = -1$ is shown in Fig. 16.13*a*. These figures illustrate the effect of inserting a lag network in cascade in a feedback control system; i.e., the lag characteristic of $G_{zo}(s)$ reduces the degree of system stability, as illustrated by Fig. 16.13, which transforms a completely stable system into a conditionally stable system. Thus for a given value of ζ, the values of t_p and t_s (and T_s) are increased. Therefore, as stated previously, the ZOH unit degrades the degree of system stability.

Case 2: PCT System's Output Transfer Function

The desired control ratio model, given by Eq. (16.21), is based upon a desired damping ratio $\zeta = 0.707$ for the dominant roots. For a unit-step forcing function, the PCT system's output transfer function is obtained from Eq. (16.21) as follows:

$$\begin{aligned} [C(s)]_{\mathrm{M}} &= \frac{9.534}{s(s^3 + 21s^2 + 20s + 9.534)} \\ &= \frac{9.534}{s(s + 0.4875 \pm j0.4883)(s + 20.03)} \end{aligned} \tag{16.30}$$

where $K_x = 0.4767$. The real and imaginary parts of the desired roots of Eq. (16.30), for $T = 0.1$ s, lie in the acceptable region of Fig. 16.5 for a good Tustin approximation. Note that the pole at -20.03 is due to the Padé approximation for $G_{zo}(s)$ [see Eq. (16.27)].

16.7.4 Sampled-Data Control System Example

The determination of the time-domain performance of the S-D control system of Fig. 16.8 may be achieved by either obtaining the exact expression for $C(z)$ or applying the Tustin transformation to Eq. (16.29) to obtain the approximate expression for $[C(z)/R(z)]_{TU}$. Thus, the output expression $C(z)_{TU}$ for a step forcing function is given by

$$[C(z)]_{TU} = \left[\frac{C(z)}{R(z)}\right]_{TU} R(z) \qquad \text{where} \quad R(z) = \frac{z}{z-1} \tag{16.31}$$

Case 3: The DIR Approach Using the Exact $\mathcal{Z}$ Transformation

Proceeding with the exact approach first requires the $\mathcal{Z}$-transfer function of the forward-loop transfer function of Fig. 16.8. For the plant transfer function of Eq. (16.27),

$$G_z(z) = \mathcal{Z}\left[\frac{K_x(1-e^{-sT})}{s^2(s+1)}\right] = (1-z^{-1})\mathcal{Z}\left[\frac{K_x}{s^2(s+1)}\right]$$
$$= \frac{K_x[(T-1+e^{-T})z + (1-Te^{-T}-e^{-T})]}{z^2-(1+e^{-T}+K_x-TK_x-K_xe^{-T})z+e^{-T}+K_x-K_x(T+1)e^{-T}} \tag{16.32}$$

Thus, for $T = 0.1$ s and $K_x = 0.4767$,

$$G_z(z) = \frac{0.002306(z+0.9672)}{(z-1)(z-0.9048)} \tag{16.33}$$

which yields

$$\frac{C(z)}{R(z)} = \frac{0.002306(z+0.9672)}{(z-0.9513 \pm j0.04649)} \tag{16.34}$$

The DIG technique requires that the s-domain model control ratio be transformed into a z-domain model. Applying the Tustin transformation to

$$\left[\frac{C(s)}{R(s)}\right] = \frac{G_{PC}(s)}{1+G_{PC}(s)} = M(s) \tag{16.35}$$

yields

$$\frac{[C(z)]_{TU}}{[R(z)]_{TU}} = [M(z)]_{TU} \tag{16.36}$$

This equation is rearranged to

$$[C(z)]_{TU} = [M(z)]_{TU}[R(z)]_{TU} \tag{16.37}$$

As stated in Sec. 16.7.1,

$$R(z) = \mathcal{Z}[r^*(t)] = \frac{1}{T}[R(z)]_{TU} \tag{16.38}$$

$$C(z) = \mathcal{Z}[c^*(t)] = \frac{1}{T}[C(z)]_{TU} \tag{16.39}$$

Inserting Eqs. (16.38) and (16.39) into Eq. (16.36) yields

$$\frac{C(z)}{R(z)} = [M(z)]_{\mathrm{TU}} \tag{16.40}$$

Substituting from Eq. (16.39) into Eq. (16.37) and rearranging yields

$$C(z) = \frac{1}{T}[M(z)]_{\mathrm{TU}}[R(z)]_{\mathrm{TU}} = \frac{1}{T}[\text{Tustin of } M(s)R(s)] \tag{16.41}$$

Case 4: The Tustin Control-Ratio Transfer Function

Based upon Eq. (16.40), the Tustin transformation of Eq. (16.40), with $K_x = 0.4767$, results in a Tustin model of the control ratio as follows:

$$\begin{aligned}\frac{C(z)}{R(z)} &= \left[\frac{C(z)}{R(z)}\right]_{\mathrm{TU}} \\ &= \frac{5.672 \times 10^{-4}(z+1)^3}{(z - 0.9513 \pm j0.04651)(z + 6.252 \times 10^{-4})}\end{aligned} \tag{16.42}$$

Note that the dominant poles of Eq. (16.42) are essentially the same as those of Eq. (16.34). This result is due to using the value of T that produced dominant roots lying in the good Tustin region of Fig. 16.5. The nondominant pole is due to the *Padé* approximation of $G_{zo}(s)$. In using the exact $\mathcal{Z}$ transformation the order of the numerator polynomial of $C(z)/R(z)$, Eq. (16.27) is 1 less than the order of its denominator polynomial. When using the Tustin transformation, the order of the numerator polynomial of the resulting $[C(z)/R(z)]_{\mathrm{TU}}$ is in general equal to the order of its corresponding denominator polynomial [see Eq. (16.35)]. Thus, $[C(z)]_{\mathrm{TU}}$ results in a value of $c^*(t) \neq 0$ at $t = 0$, which is in error based upon zero intial conditions. Table 16.2 illustrates the effect of this characteristic of the Tustin transformation on the time response due to a unit-step forcing function. The degradation of the time response by use of the Tustin transformation is minimal; i.e., the resulting values of the FOM are in close agreement to those obtained by using the exact $\mathcal{Z}$ transformation. Therefore, the Tustin transformation is a valid design tool when the dominant zeros and poles of $[C(s)/R(s)]_{\mathrm{TU}}$ lie in the acceptable Tustin region of Fig. 16.5.

Table 16.3 summarizes the time-response FOM for a unit-step forcing function of the continuous-time system of Fig. 16.9 for the two cases of (1) $G(s) = G_x(s)$ [with $G_A(s)$ removed] and (2) $G(s) = G_A(s)G_x(s)$ and the S-D system in Fig. 16.8 based upon the (3) exact and (4) Tustin expressions for $C(z)$. Note that the value of M_p occurs between 6.4 and 6.5 s and the value of t_s occurs between 8.6 and 8.7 s. The table reveals that:

1. In converting a continuous-time system into a sample-data system the time-response characteristics are degraded.

TABLE 16.2 Comparison of Time Responses Between $C(z)$ and $[C(z)]_{TU}$ for a Unit-Step Input and $T = 0.1$ s

	c(kT)	
k	Case 3 (exact), $C(z)$	Case 4 (Tustin), $[C(z)]_{TU}$
0	0.	0.5672E-03
2	0.8924E-02	0.9823E-02
4	0.3340E-01	0.3403E-01
6	0.7024E-01	0.7064E-01
8	0.1166	0.1168
10	0.1701	0.1701
12	0.2284	0.2283
14	0.2897	0.2894
18	0.4153	0.4148
22	0.5370	0.5364
26	0.6485	0.6478
30	0.7461	0.7453
34	0.8281	0.8273
38	0.8944	0.8936
42	0.9460	0.9452
46	0.9844	0.9836
50	1.011	1.011
54	1.029	1.028
58	1.039	1.038
60	1.042	1.041
61	1.042	1.042
62	1.043	1.043
63	1.043	1.043
64	1.043 M_p	1.043 M_p
65	1.043	1.043
66	1.043	1.043
67	1.043	1.043
68	1.042	1.042
85	1.022	1.022
86	1.021 t_s	1.021 t_s
87	1.019	1.019

2. The time-response characteristics of the S-D system, using the gain values obtained from the continuous-time model, agree favorably with those of the continuous-time model. As may be expected, there is some variation in the values obtained when utilizing the exact $C(z)$ and $[C(z)]_{TU}$.

TABLE 16.3 Time-Response Characteristics of the Uncompensated System

	M_p	t_p, s	t_s, s	K_1, s^{-1}	Case
Continuous-time system					
$G(s) = G_x(s)$	1.04821	6.3	8.40 to 8.45	0.4767	1
$G_M(s) = G_A(s)G_x(s)$	1.04342	6.48	8.69		2
Sampled-data system					
$C(z)$	1.043	6.45	8.75	0.4765	3
$[C(z)]_{TU}$	1.043	6.45	8.75		4

16.7.5 The PCT System of Fig. 16.1

Comparing Fig. 16.1 to Fig. 15.1, it should be noted that the portion of the forward loop of Fig. 16.14d, excluding the plant $G_x(s)$, corresponds to the forward loop of Fig. 15.1 excluding the plant. Thus, in obtaining the PCT system representing the MISO digital control system of Fig. 16.14d, *only the sampling of the analog function $e(t)$ is replaced by a factor of* $1/T$. The sampling switch in the output of the digital controller is shown only to remind the reader that the output of the controller is the result of the digital-to-analog conversion. Therefore, in obtaining the PCT system, since this sampler *does not involve the sampling of an analog signal*, *it is not replaced by the factor* $1/T$. This diagram is simplified to the one shown in Fig. 16.14e.

16.7.6 PCT Design Summary

Once a satisfactory controller $D_c(s)$ (see Fig. 16.14d) has been achieved, where the degrees of the numerator and denominator are, respectively, w_s and n_s then (1) if $n_s \geq w_s + 2$, use the exact $\mathcal{Z}$-transform to obtain the discrete controller $D_c(z)$, or (2) if $w_s < n_s < w_s + 2$, use the Tustin transformation to obtain $[D_c(z)]_{TU}$ [1]. For the latter case, *insert $n_s - w_s$, nondominant zeros so that $n_s = w_s$*. If these nondominant zeros are not inserted, the Tustin transformation of $D_c(s)$ will result in $n_s - w_s$ *z-domain dominant zeros that can affect* the system's performance.

A final check should be made before simulating the discrete design; i.e., the Bode plots of $D_c(s)|_{s=j\omega}$ and $D_c(z)|_{z=e^{(j\omega T)}}$ should be essentially the same within the desired bandwidth (BW) in order to ascertain that the discrete-time system response characteristics are essentially the same as those obtained for the analog PCT system response. If the plots differ appreciably, it implies that warping has occurred and the desired discrete-time response characterstics may not be achieved (depending on the degree of the warping).

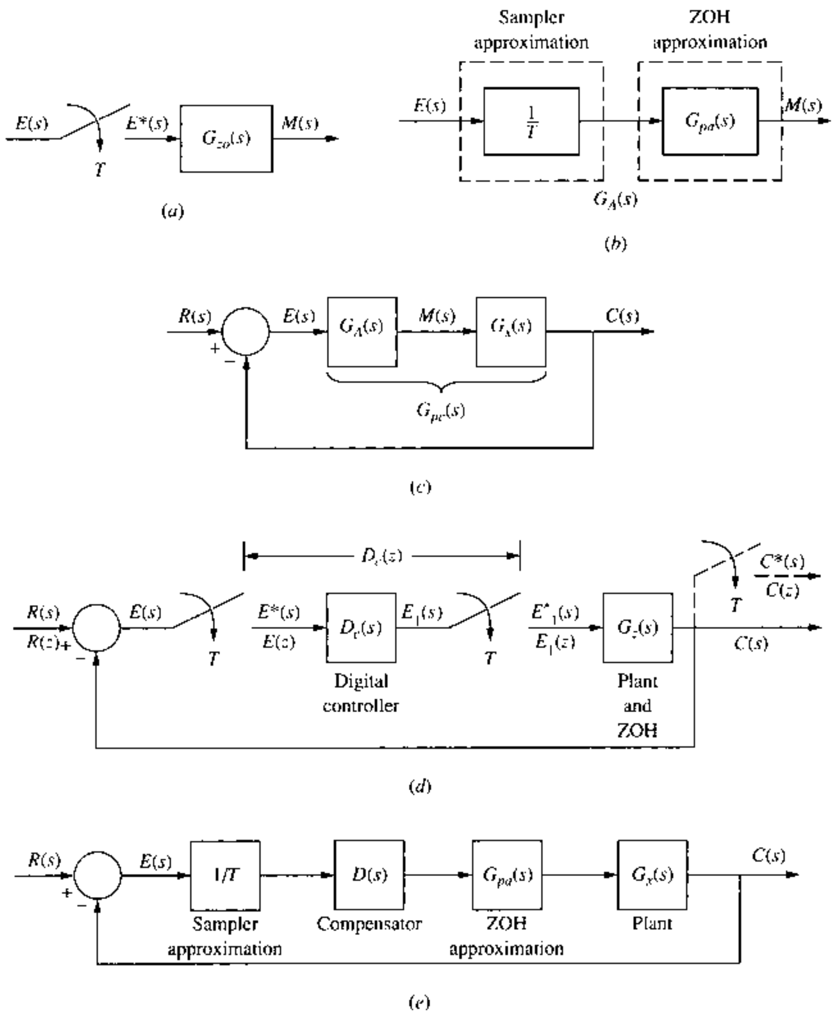


FIGURE 16.14 The PCT equivalent of Fig. 16.1.

The result of this section demonstrate that when the Tustin approximation is valid, the PCT approximation of the sampled-data system is a practical design method. When a cascade compensator $D_c(s)$ (see Fig. 16.11) is designed to yield the desired performance specifications in the pseudo-continuous-time domain, the discrete cascade compensator $D_c(z)$ (see Fig. 16.1 where the controller does not include a ZOH) is accurately obtained by using the Tustin transformation.

16.8 DESIGN OF DIGITAL CONTROL SYSTEM [1]

The previous section present the DIR and DIG techniques for analyzing and achieving the desired performance of the basic system. Generally, gain adjustment alone is not sufficient to achieve all the desired system performance specifications, i.e., the figures of merit (FOM) (see Secs. 3.9 and 3.10). A cascade and/or feedback compensator (controller) can be used to achieve the design objectives. The following sections discuss the design objective of satisfying values of the desired conventional control theory FOM by using cascade or feedback compensators.

The analysis of a system's performance may be carried out in either the time and/or frequency domains. If the performance of the basic system is unsatisfactory, then, based on the results of this analysis, a compensator can be designed. Two approaches may be used to design the digital compensator (controller) of Fig. 16.1—the DIR and DIG techniques. *Both techniques rely heavily on a trial-and-error approach that requires a firm understanding of the fundamentals of compensation.* To facilitate this trial-and-error approach, a computer-aided design (CAD) package such as MATLAB (see App. B) or TOTAL-PC (see App. D) is used. The main advantage of the DIR design is that the performance specification can be met with less stringent requirements on the controller parameters.

In the DIG method, the controller is first designed in the s domain, and then the bilinear transformation is employed to transform the controller into its equivalent z domain discrete controller. The advantage of this technique is that tried and proven continuous-time domain methods are used for designing an acceptable $D_c(s)$ controller. However, if T is not small enough, the poles and zeros of the controller lie outside the good bilinear transformation region, which may result in not achieving the desired performance.

After $D_c(z)$ is included in the system model, as shown in Fig. 16.1, a system analysis is again performed. If the modified system does not meet the desired specifications, the controller is modified accordingly. To illustrate this design process, the basic system of Sec. 16.7.3 is used for the design of a lead, lag, and lead-lag cascade compensator.

16.9 DIRECT (DIR) COMPENSATOR

This section presents the basic concept of the DIR technique for cascade compensation. The standard lead and lag s domain transfer function is transformed to the z domain where the specific design is to be applied. This simple lead ($\alpha < 1$) and lag ($\alpha > 1$) compensators in the s domain have the form

$$D_c(s) = \frac{K_{sc}(s - z_s)}{s - p_s} \tag{16.43}$$

where $p_s = z_s/\alpha$. These s plane zeros and poles are transformed into the z domain poles and zeros as follows:

$$z_z = e^{sT}\big|_{s=z_S} = e^{z_s T} \tag{16.44}$$

$$p_z = e^{sT}\big|_{s=p_s} = e^{p_s^T} = e^{z_s T/\alpha} \tag{16.45}$$

Thus, the corresponding first-order z domain compensator (digital filter) is

$$D_c(z) = \frac{K_{zc}(z - z_z)}{z - p_z} = \frac{K_{zc}(z - z_z)}{z - z_z/\beta} \tag{16.46}$$

where $p_z = z_z/\beta$.

A relationship between α and β is obtained by taking the natural log of Eqs. (16.44) and (16.45) as follows:

$$z_s T = \ln z_z \tag{16.47}$$

$$\frac{z_s T}{\alpha} = \ln p_z \tag{16.48}$$

Taking the ratio of these equations and rearranging yields $\alpha \ln p_z = \ln z_z$. Thus,

$$p_z^{\alpha} = z_z \tag{16.49}$$

and

$$\beta = \frac{z_z}{p_z} = p_z^{\alpha-1} \tag{16.50}$$

For a lead controller $\alpha < 1$ and $p_z < 1$. *Therefore, for a lead filter,* $\beta > 1$ *from Eq. (16.50). For a lag digital filter,* $\beta < 1$. (Note that the magnitude condition on β is just the opposite of that on α.)

In the z domain $p_z > z_z > 0$ for a lag controller and $z_z > p_z$ for a lead controller. Thus, the standard lag-lead controller is of the form

$$D_c(z) = K_c\left[\frac{z - z_1}{z - p_1}\right]\left[\frac{z - z_2}{z - p_2}\right] \tag{16.51}$$

where $p_1 > z_1 > 0$ (the lag portion) and $z_2 > p_2 > 0$ (the lead portion). The cascade controller design rules given in Chap. 10 can be used for the design of these z domain compensators (see Ref. 1).

16.10 PCT LEAD CASCADE COMPENSATION

In this section the PCT DIG technique is utilized to achieve the desired improvement in the performance of the basic S-D control system by the use of conventional s-plane cascade compensation methods. *It is assumed in this chapter that the digital controller of Fig. 16.1 does not use a ZOH device in the analog-to-digital (A/D) model in the input to the controller.*

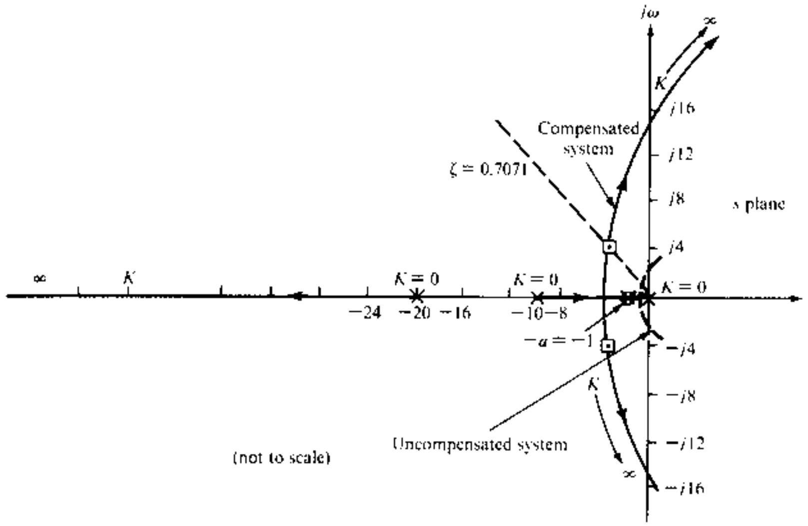


FIGURE 16.15 Root locus of a compensated system: $D_c(s)G_{pc}(s) = -1$.

Case 5: PCT Lead Compensator Design

The magnitude of the real part of the dominant poles of Eq. (16.30), for the basic model system, is $|\sigma_{1,2}| = 0.4875$. Thus, it is necessary to at least double this magnitude for the real part $|\sigma_{1,2}|$ of the dominant roots of the compensated model system (Fig. 16.14e), in order to reduce the settling time by one-half. To accomplish this, a lead network of the form

$$D_c(s) = K_{sc}\frac{s+a}{s+b} \qquad a < b \tag{16.52}$$

can be inserted in cascade with $G_{pc}(s)$ (see Fig. 16.14e). Since the settling time for the basic system is 8.6 s (see Table 16.2), then the desired $t_s \approx T_s$ is 4.3 s. Inserting this value into $T_s = 4/|\sigma'_{1,2}|$ yields

$$|\sigma'_{1,2}| \geq \frac{4}{4.3} = 0.93 \tag{16.53}$$

that is in the region for a good Tustin approximation. Using the cancellation rule of Chap. 10 for the design of Eq. (16.52), i.e., $a = 1$, and assuming $b = 10$ yields, for $\zeta = 0.7071$ in the s domain (see the root locus of Fig. 16.15), the actual value of

$$|\sigma'_{1,2}| = 3.8195 \tag{16.54}$$

This value lies just *outside the shaded region* of Fig. 16.5, for $T = 0.1$ s, but inside the boundary of $\sigma = -2/T = -20$. For this example, this does not

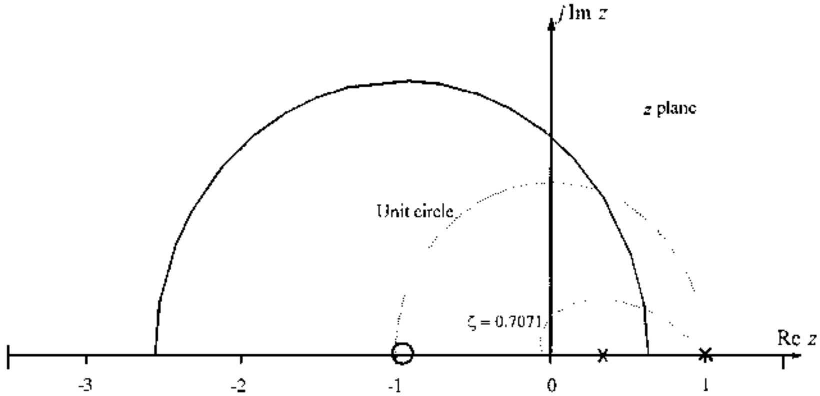


FIGURE 16.16 Lead compensation: root locus for Eq. (16.59).

present a problem with respect to T_s since 3.8195 is at least four times greater than the desired value of 0.93. This factor of 4 does not yield $T_s = 4/3.8195$ because $\sigma'_{1,2}$ is not in the shaded region of Fig. 16.5. Since $|\sigma'_{1,2}| > 2|\sigma_{1,2}|$, a definite improvement in T_s, is achieved as demonstrated in this example. Thus, the compensator

$$D_c(s) = \frac{K_{sc}(s+1)}{s+10} \tag{16.55}$$

for $\zeta = 0.7071$, yields the root locus of Fig. 16.15 and

$$\left[\frac{C(s)}{R(s)}\right]_{TU} = \frac{D_c(s)G_{PC}(s)}{1 + D_c(s)G_{PC}(s)} = \frac{9.534K_{sc}}{s^3 + 30s^2 + 200s + 652.6} \tag{16.56}$$

$$= \frac{9.534K_{sc}}{(s + 3.8195 \pm j3.8206)(s + 22.36)} \tag{16.57}$$

where $K_{sc} = 68.45$ for the continuous-time model. For this system model $M_p \approx 1.04169$, $t_p \approx 0.875$ s, $t_s \approx 1.15$ s, and $K_1 = 3.263\ \text{s}^{-1}$.

Substituting from Eq. (16.8) into Eq. (16.55), with $T = 0.1$ s, yields the digital compensator or controller form of Fig. 16.1:

$$[D_c(z)]_{TU} = \frac{E_1(z)}{E(z)} = \frac{K_c(21z - 19)}{30z - 10} = \frac{K_{zc}(z - \frac{19}{21})}{z - \frac{1}{3}} \tag{16.58}$$

where $K_{sc} = K_c$, and $K_{zc} = 0.7K_c$, and $\beta_{1e} = 2.7143$. Thus, for the system of Fig. 16.1,

$$G(z) = [D_c(z)]_{TU}G_z(z) = \frac{1.6142 \times 10^{-3}K_c(z + 0.9672)}{(z-1)(z - 0.3333)} \tag{16.59}$$

TABLE 16.4 FOM for a Cascade Lead-Compensated Designed System: s-plane Design (DIG) (Values obtained by use of the TOTAL-PC CAD package)

System	M_p	t_p, s	t_s, s	K_1, s^{-1}	Case
$[C(s)]_T$					
Uncompensated	1.0482	6.3	8.40–8.45	0.4767	1
Compensated	1.0417	0.875	1.15	3.263	5
$C(z)$					
Uncompensated	1.043	6.4+	8.6+	0.4765	3
Compensated	1.043	0.8	1+	3.3208	6

Case 6: The z-domain PCT Control Ratio for the Lead Controller Design

The root-locus plot for this design is shown in Fig. 16.16 and, for $\zeta = 0.7071$ in the z domain, the control ratio designed via path B of Fig. 16.6 is

$$\frac{C(z)}{R(z)} = \frac{0.1125(z + 0.9672)}{(z - 0.6104 \pm j0.2638)} \tag{16.60}$$

where $K_c = 69.694$ is close to the s-plane model value of 68.45 for Eq. (16.57). Thus, for this example, paths A and B yield essentially the same system performance. This good degree of correlation is due to the fact that the imaginary part of the dominant poles in the s plane lies within the shaded area of Fig. 16.5. Table 16.4 compares the FOM of the continuous-time and sampled-data uncompensated and compensated systems. Thus, the resulting design of case 6 achieves the desired design specifications. Taking the $\mathcal{Z}^{-1}$ of Eq. (16.58) yields the difference equation for the resulting controller as follows:

$$\mathcal{Z}^1[30E_1(z) - 10z^{-1}E_1(z)] = \mathcal{Z}^{-1}[K_c(21 - 19z^{-1})E(z)]$$

$$e_1(kT) = 48.7858e(kT) - 44.13953e[(k-1)T] + 0.333333e_1[(k-1)T]$$

This discrete control law (algorithm) is translated into a software program to be implemented on a digital computer (controller) [1].

16.10.1 MATLAB Design for Sec. 16.10

The following detailed MATLAB m-file, for $T = 0.1$ s, involves both the DIG (PCT) and the DIR (exact z-domain) designs.

DIG (PCT) Design

```
echo on
gxnum=1;                          % Create numerator
gxden=[1 1 0];                    % Create denominator
```

```
Gx=tf(gxnum, gxden);          % Create transfer function
                              % from num and den
zpk(Gx)                       % Show zero - pole - gain form

Zero/pole/gain:               % [Eq. (16.27)]
    1
-----
 s (s+1)

Tsamp=0.1;                    % Tsamp is sample time=0.1 sec
GA=tf([2], [Tsamp 2])         % Create sampler and ZOH
Transfer function:
    2
-----
0.1 s+2
Gpc=series(GA, Gx);           % Multiply GA and GX
zpk(Gpc)                      % Show zero - pole - gain form
Zero/pole/gain:
       20
----------
 s (s+20) (s+1)
Dc=tf([1  1], [1  10]);       % Create lead compensator
GDIG=series(Dc, Gpc);         % Multiply Dc and Gpc
zpk (GDIG)                    % Show zero - pole - gain form
Zero/pole/gain:
        20 (s+1)
----------------
 s (s+20) (s+10) (s+1)
GDIG=minreal (GDIG);          % Cancel pole-zero pairs
[k, r]=rloczeta(GDIG.num {1,1},GDIG.den{1,1},.7071,.0000 1)

                              % Find gain for zeta=.7071
k=
  32.6243
r=
  -22.3607
    -3.8196-3.8197i
    -3.8196+3.8197i

cltf=feedback(k*GDIG, 1);     % Close the loop
fom(cltf.num {1,1}, cltf.den {1,1})
                              % Calculate figures of merit
Figures of merit for a unit step input

rise time      =  0.41
peak value     =  1.0417
peak time      =  0.88
settling time  =  1.14
final value    =  1
```

DIR Design (Exact z Domain Design): $T = 0.1$ s

```
echo on
gxnum=1;                          % Create numerator
gxden=[1  1  0];                  % Create denominator
Gx=tf(gxnum, gxden);              % Create transfer function
                                  % from num and den
zpk(Gx)                           % Show zero - pole - gain  form
Zero/pole/gain:
   1
- - - - -
s (s+1)

Tsamp=0.1;                         % Tsamp is sample time=0.1
dirgx=c2d(Gx, Tsamp, 'zoh')        % Transform Gx to the z domain
Transfer function:
0.004837 z+0.004679
- - - - - - - - - - - - - -
z^2-1.905 z+0.9048
Sampling time: 0.1
zpk(dirgx)                        % Show zero - pole - gain form

Zero/pole/gain:
 0.0048374 (z+0.9672)
- - - - - - - - - - - - - - -
  (z-1) (z-0.9048)

Sampling time: 0.1
Dc=tf ([1  1], [1  10]);           % Create lead compensator
dirdc=c2d(Dc, Tsamp,'tustin');     % Transform Dc to the z domain
zpk(dirdc)                         % Show zero - pole - gain form

Zero/pole/gain:
 0.7 (z-0.9048)
- - - - - - - - - - -
  (z-0.3333)

Sampling time: 0.1
GDIR=series (dirdc, dirgx);        % Multiply Dc(z) and Gz(z)
zpk (GDIR)                         % Show zero - pole - gain form
Zero/pole/gain:
 0.0033862 (z+0.9672) (z-0.9048)
- - - - - - - - - - - - - - - - - - - - - -
  (z-1) (z-0.9048) (z-0.3333)

Sampling time: 0.1
GDIR=mineral (GDIR, .0001);        % Cancel pole - zero pair
rltool (GDIR)                      % Open root locus tool
                                     window
```

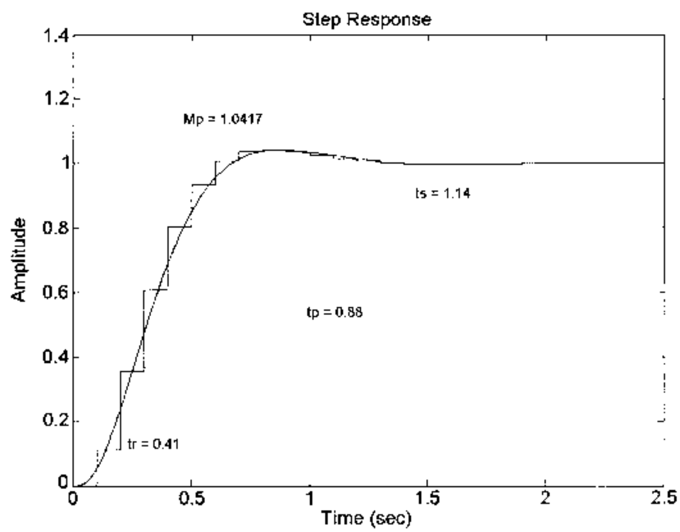


FIGURE 16.17 MATLAB step response $T = 0.1$ s and figures of merit for Eqs. (16.57) and (16.60).

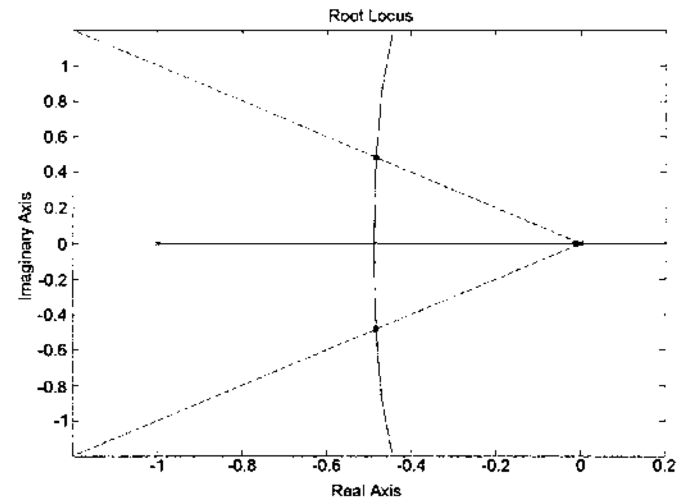
```
% Root locus tool is used to design compensator and
% export the closed loop transfer function T_r2y

hold on
Step (T_r2y)
```

As noted in Fig. 16.17, the FOM for the DIG design for $T = 0.1$ s agree very closely with those for the DIR design. It is left to the reader to obtain corresponding MATLAB plots for $T = 0.01$ and 1.0 s.

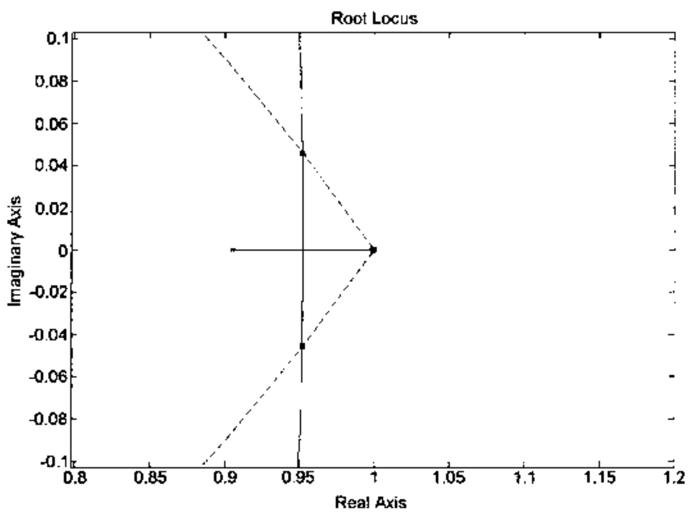
16.11 PCT LAG COMPENSATION

When a sizable improvement in the value of the static error coefficient is desired, assuming that the transient characteristics are satisfactory (category I of Sec. 10.5), a lag compensator can be inserted in cascade with the basic plant. The s plane lag compensator is of the form

$$D_c(s) = K_{sc}\left[\frac{s + 1/T_1}{s + 1/\alpha T_1}\right] \qquad \text{where} \quad K_{sc} = \frac{A}{\alpha} \tag{16.61}$$



(a)



(b)

FIGURE 16.18 Root locus of (*a*) Eq. (16.63) and (*b*) Eq. (16.66).

The same basic system and the value of $T = 0.1$ s of the previous section is also used in this section to illustrate the design of a lag compensator. Applying the standard design procedure (see Sec. 10.8) for a lag compensator in the s plane for the plant of Eq. (16.27) yields the lag compensator

$$D_c(s) = \frac{K_{sc}(s + 0.01)}{s + 0.001} \tag{16.62}$$

and

$$G(s) = D_{sc}(s)G_A(s)G_x(s) = \frac{K_s(s + 0.01)}{s(s + 0.001)(s + 1)(s + 20)} \tag{16.63}$$

where $K_s = 20K_{sc}K_x$. The root-locus method, for $\zeta = 0.7071$ (see Fig. 16.18*a*), yields

$$\frac{C(s)}{R(s)} = \frac{9.524(s + 0.01)}{(s + 0.48291 \pm j0.48304)(s + 0.010195)(s + 20.025)} \tag{16.64}$$

where $K_{sc} = 1.001$ and $K_s = 9.524$. The root-locus branches in the vicinity of the origin, for the scale used in Fig. 16.18*a*, are indistinguishable from the axes.

Applying the Tustin transformation of Eq. (16.8) to Eq. (16.62) yields

$$[D_c(z)]_{TU} = \frac{K_{zc}(z - 0.999)}{z - 0.9999} \tag{16.65}$$

where $K_{zc} = 1.00045$, $K_{sc} = 0.9994$ and

$$G(z) = [D_c(z)]_{TU}G_z(z) = \frac{K_z(z + 0.9672)(z - 0.999)}{(z - 1)(z - 0.9048)(z - 0.9999)} \tag{16.66}$$

where $K_z = 2.303 \times 10^{-3}$ (path A of Fig. 16.7). The z-domain root locus is shown in Fig. 16.18*b* and the corresponding control ratio is given in the m-file in Sec. 16.11.1.

The resultant FOM for a unit-step forcing function are given in Table 16.5. As expected, an increase in the value of K_1, by use of lag compensator is achieved at the expense of increasing the values of M_p, t_p, and t_s.

TABLE 16.5 The PCT DIG Lag Compensator Design FOM

System	Domain	M_p	M_m	t_p, s	t_s, s	K_1, s^{-1}	ω_m, rad/s	ω_b, rad/s
Uncompensated	S	1.043	1.00	6.48	8.69	0.4767	...	0.6900
	Z	1.043	1.00+	6.4+	8.6+	0.4767	0+	0.691−
Compensated	S	1.063	1.019	6.545	10.85	4.762	0.09	0.6960
	Z	1.063	1.02−	6.4+	10.7	4.8	0.060	0.644+

16.11.1 MATLAB Design for Sec. 16.11

The following detailed MATLAB m-file, for $T = 0.1$ s, involves both the DIG (PCT) and the DIR (exact z-domain) designs.

DIG (PCT) Design

```
echo on
gxnum=1;                        % Create numerator
gxden=[1  1  0];                % Create denominator
Gx=tf(gxnum, gxden);            % Create transfer function
                                % from num and den
zpk(Gx)                         % Show zero - pole - gain form

Zero/pole/gain:
   1
-----
s (s+1)

Tsamp=0.1;                      % Tsamp is sample time=0.1 sec
GA=tf([2], [Tsamp 2])           % Create sampler and ZOH

Transfer function:
   2
-----
0.1 s+2

Gpc=series(GA, Gx);             % Multiply GA and Gx
zpk (Gpc)                       % Show zero - pole - gain form

Zero/pole/gain:
     20
----------
s (s+20) (s+1)

Dc=tf([1  0.01], [1  0.001]);   % Create lag compensator
GDIG=series (Dc, Gpc);          % Multiply Dc and Gpc
zpk(GDIG)                       % Show zero - pole - gain form

Zero/pole/gain:
     20 (s+0.01)
-----------------
s (s+20) (s+1) (s+0.001)

rltool(GDIG)                    % Use root locus tool to design
                                % and export T_r2y
% Resulting gain is .4762 to give Ks=9.524
T_r2y                           % Closed loop transfer function
```

```
Zero/pole/gain from input ''r'' to output ''y'':
                9.524 (s+0.01)
- - - - - - - - - - - - - - - - - - - - - - - - - -
(s+0.01019) (s=20.02) (s^2+0.9658 s+0.4665)

cltf=tf (T_r2y)        % Create transfer function form of cltf

Transfer function from input ''r'' to output ''y'':
          9.524 s+0.09524
- - - - - - - - - - - - - - - - - - - - - -
 s^4+21 s^3+20.02 s^2+9.544 s+0.09524

fom(cltf.num {1,1}, cltf.den {1,1}, .01,15);
% Stop time set to 15 sec because auto stop senses the slow modes.
Figures of merit for a unit step input

rise time       =  3.03
peak value      =  1.0627
peak time       =  6.55
settling time   =  10.84
final value     =  1
```

DIR Design (Exact z Domain Design)

```
echo on
gxnum=1;                          % Create numerator
gxden=[1  1  0];                  % Create denominator
Gx=tf (gxnum, gxden);             % Create transfer function
                                  % from num and den
zpk(Gx)                           % Show zero - pole - gain form

Zero/pole/gain:
   1
- - - - -
s (s+1)
Tsamp=0.1;                        % Tsamp is sample time=0.1
Gz=c2d(Gx, Tsamp, 'zoh')          % Transform Gx to the z domain

Transfer function:
0.004837 z+0.004679
- - - - - - - - - - - -
 z^2-1.905 z+0.9048
Sampling time: 0.1
zpk (Gz)                          % Show zero - pole - gain form

Zero/pole/gain:
 0.0048374 (z+0.9672)
- - - - - - - - - - - - -
  (z-1) (z-0.9048)
```

```
Sampling time: 0.1
Dc = tf ([1  - 0.999], [1  - 0.9999], Tsamp);
                              % Create lead compensator
GDIR = series (Dc, Gz);       % Multiply Dc(z) and Gz(z)
zpk (GDIR)                    % Show zero - pole - gain form
Zero/pole/gain:
0.0048374 (z - 0.999) (z + 0.9672)
- - - - - - - - - - - - - - - - - - - - -
 (z - 1) (z - 0.9999) (z - 0.9048)

Sampling time: 0.1
rltool(GDIR)                  % Open root locus tool window
T_r2y                         % Show closed loop
                              % transfer function
Zero/pole/gain from input ''r'' to output ''y'':
0.0022902 (z + 0.9672) (z - 0.999)
- - - - - - - - - - - - - - - - - - - - -
(z - 0.999) (z^2 - 1.903z + 0.9079)

Sampling time: 0.1
hold on                       % Hold plot from above
Step(T_r2y)                   % Plot step response
axis([0  15  0  1.4])         % Reset axis scale
```

As noted in Fig. 16.19, the FOM for the DIG design for $T = 0.1$ s agree very closely with those for the DIR design and with the FOM in Table 16.5. It is left to the reader to obtain corresponding MATLAB plots for $T = 0.01$ and 1.0 s.

16.12 PCT LAG-LEAD COMPENSATION

When both a large improvement in the static error coefficient and the transient-response characteristics are desired, a lag-lead cascade compensator of the form

$$D_c(s) = A\frac{(s + 1/T_1)(s + 1/T_2)}{(s + 1/\alpha T_1)(s + \alpha/T_2)} \qquad \alpha > 1 \tag{16.67}$$

can be used, where a nominal value of $\alpha = 10$ is chosen. Note that the portion

$$\frac{s + 1/T_1}{s + 1/\alpha T_1}$$

of Eq. (16.67) corresponds to the lag compensator and the remaining portion corresponds to the lead compensator. The following example involves the PCT design of a lag-lead controller.

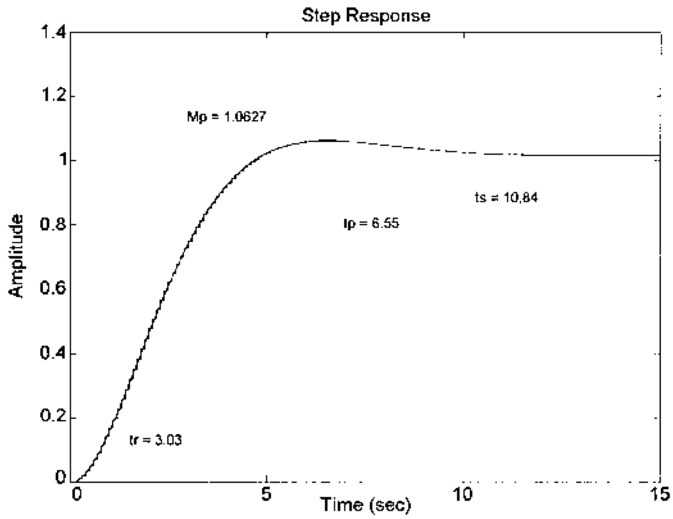


FIGURE 16.19 MATLAB step response for $T = 0.1$ s, figures of merit for Eq. (16.64) (DIG), and the DIR control ratio given in the m-file.

Example 16.3. The basic system of Sec. 10.12, which involved the design of an analog lag-lead compensator, is utilized to illustrate the effectiveness of the PCT technique. The plant transfer function is

$$G_x(s) = \frac{4.2}{s(s+1)(s+5)} \tag{16.68}$$

The desired FOM are $1 < M_p \leq 1.163$, $t_p \leq 0.3$ s, and $t_s < 0.5$ s. Using the equality conditions and Eqs. (3.60), (3.61), and (3.64) results in

$$\zeta_a = 0.5 \quad \text{and} \quad |\sigma_a| = |\zeta_a \omega_n| = 8$$

which in turn yields

$$\omega_n = 16 \qquad \omega_d = 13.869 \qquad \text{and} \qquad p_{1,2} = -8 \pm j13.86 \tag{16.69}$$

Because the lag characteristic of the ZOH unit has the same effect on increasing t_s as does a lag compensator, it is best to choose $|\sigma_a| > 8$. The desired dominant complex of poles of Eq. (16.69) are modified to

$$p_{1,2_s} = -10 \pm j14 \tag{16.70}$$

where $T_s = 4/10 = 0.25$ s and $\zeta = 0.581$. Utilizing $z = e^{sT}$ results in

$$p_{1,2_z} = 0.90483 \underline{/8.02^\circ} = 0.89598 \pm j0.12626 \tag{16.71}$$

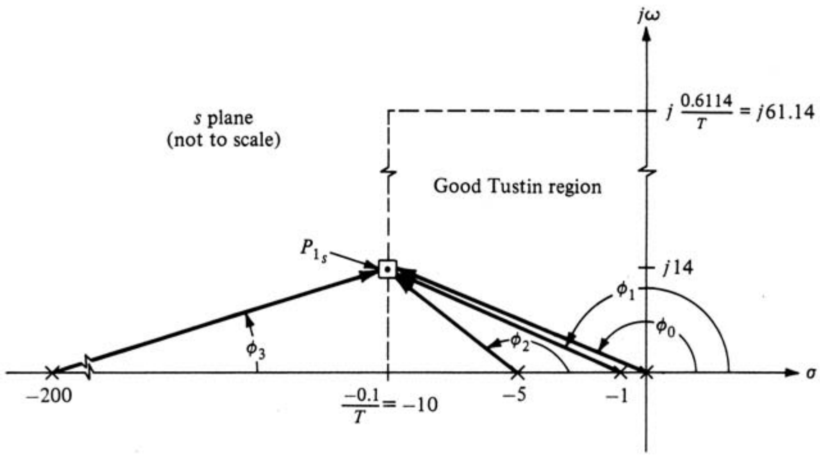


FIGURE 16.20 Plot of known poles of $G(s)$ and p_{1_s}.

The open-loop transfer function for $T = 0.01$ s of the system of Fig. 16.14e is

$$G(s) = D_c(s)G_A(s)G_x(s) = \frac{840D_c(s)}{s(s+1)(s+5)(s+200)} \tag{16.72}$$

A plot of the known poles of Eq. (16.72) and the plot of the desired dominant closed-loop pole of p_{1_s} are shown in Fig. 16.20. The dominant poles of $G(s)$ along with p_{1_s} all lie in the good Tustin bilinear transformation region. Thus, a good PCT design is achievable. First, it is necessary to ascertain whether a compensator of the simple form of Eq. (16.52) provides the required compensation to achieve the desired dominant closed-loop poles of $p_{1,2_s}$. Thus, the root-locus angle condition yields

$$\begin{aligned}(\phi_0 + \phi_1 + \phi_2 + \phi_3) - [\text{angle of } D_c(s)] &= 180^\circ \\ (125.55^\circ + 122.750^\circ + 109.65^\circ + 4.2^\circ) - [\text{angle of } D_c(s)]_{s=p_{1_s}} &= 180^\circ \\ 362.15^\circ - [\text{angle of } D_c(s)]_{s=p_{1_s}} &= 180^\circ\end{aligned} \tag{16.73}$$

Therefore,

$$[\text{angle of } D_c(s)]_{s=p_{1_s}} = 182.15^\circ \tag{16.74}$$

Since, theoretically, the largest possible positive angle that the simple compensator of Eq. (16.52) can provide is 90°, a compensator of the form

$$D_c(s) = \frac{K_{sc}(s+a)(s+b)}{(s+c)(s+d)} \tag{16.75}$$

must be utilized. Assuming perfect cancellation, let $a=1$ and $b=5$. Thus, only the values of c and d need to be determined. Utilizing $D_c(s)$ of Eq. (16.52) and the selected values of a and b results in

$$G(s) = \frac{804K_{sc}}{s(s+200)(s+c)(s+d)} \tag{16.76}$$

Thus, for Eq. (16.76),

$$\begin{gathered}(\phi_0 + \phi_3 + \phi_c + \phi_d) = 129.75 + (\phi_c + \phi_d) = 180^\circ \\ (\phi_c + \phi_d) = 50.25^\circ\end{gathered} \tag{16.77}$$

Selecting $c=d$ yields $\phi_c = \phi_d = 25.125^\circ$:

$$\tan\phi_d = \frac{14}{\Delta} = 0.46897 \qquad \rightarrow \qquad \Delta = 29.85$$

which yields $c=d=10+\Delta=39.85$. The resultant controller, for $\zeta=0.581$, generates a value of $K_{sc}=4243.8$ and the following control ratio and its associated FOM:

$$\frac{C(s)}{R(s)} = \frac{3.565 \times 10^6}{(s+9.997 \pm j14.01)(s+60.41)(s+199.3)} \tag{16.78}$$

$$M_p = 1.10 \quad t_p = 0.249\,\text{s} \quad t_s = 0.366\,\text{s} \quad \text{and} \quad K_j = 11.22\,\text{s}^{-1}$$

The FOM of Eq. (16.78) are essentially those specified by Eq. (16.69).

Since the orders of the numerator and the denominator of Eq. (16.75) are the same, then its Tustin transformation yields

$$[D_c(z)]_{TU} = \frac{3040(z-0.99005)(z-0.95123)}{(z-0.6677)^2} \tag{16.79}$$

and results in

$$\frac{C(z)}{R(z)} = \frac{[D_c(z)]_T G_z(z)}{1+[D_c(z)]_T G_z(z)]} = \frac{2.096 \times 10^{-3}(z+0.26395)(z+3.6767)}{(z-0.8953 \pm j0.1270)(z-0.5428)} \tag{16.80}$$

The FOM are

$$M_p = 1.101 \qquad t_p = 0.24\,\text{s} \qquad t_s = 0.36-\text{s} \qquad \text{and} \qquad K_l = 11.22\,\text{s}^{-1}$$

These FOM are essentially the same as those for Eq. (16.80).

Because the dominant poles of $G(s)$ and the desired s-plane closed-loop poles lie in the good Tustin region for the given value of T, the results of the PCT approach, for this example, yield essentially the same FOM in the s and z domains. If the FOM in the z domain, via path A of Fig. 16.7, were not quite satisfied, then a z-domain "fine-tuning" (gain adjustment) via path B may yield the desired FOM.

16.12.1 MATLAB Design for Sec. 16.12

The following detailed MATLAB m-file, for $T = 0.1$ s, involves both the DIG (PCT) and the DIR (exact z-domain) designs.

DIG Design

```
echo on
gxnum=4.2;                    % Create numerator
gxden=conv([1  1  0],[1  5]);  % Create denominator
Gx=tf(gxnum, gxden);          % Create transfer function
                              % from num and den
zpk(Gx)                       % Show zero - pole - gain form

Zero/pole/gain:
      4.2
---------
s (s+5) (s+1)

Tsamp=0.01;                   % Tsamp is sample time=0.01 sec
GA=tf([2], [Tsamp 2])         % Greate sampler and ZOH

Transfer function:
     2
------
0.01 s+2

Gs=series (GA, Gx);           % Multiply GA and Gx
zpk (Gs)                      % Show zero - pole - gain form

Zero/pole/gain:
          840
----------------
s (s+200) (s+5) (s+1)

Dcnum=conv([1  1], [1  5]);  % Place zeros
Dcden=conv([1  39.85], [1  39.85]);    % Place poles
Dc=tf(Dcnum, Dcden);          % Assemble lead-lag compensator
GDIG=series (Dc, Gs);         % Multiply Dc and Gpc
GDIG=minreal (GDIG);          % Cancel pole-zero pairs
zpk (GDIG)                    % Show zero - pole - gain form
```

```
Zero/pole/gain:
          840
- - - - - - - - - - - - - - -
s (s+200) (s+39.85)^2

rltool(GDIG)           % Use root locus tool to design
                       % and export T_r2y
% Resulting gain is Ksc=4243.8
% These commands are used to determine
% the figures of merit of the closed loop system
T_r2y                  % Show closed loop transfer function

Zero/pole/gain from input ''r'' to output ''y'':
                  3564792
- - - - - - - - - - - - - - - - - - - - - - - - - -
(s+199.3) (s+60.41) (s^2+19.99s+296.1)

cltf=tf(T_r2y)         % Create transfer function form of clft

Transfer function from input ''r'' to output ''y'':
                      3.565e006
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
s^4+279.7 s^3+1.753e004 s^2+3.176e005 s+3.565e006

fom(cltf.num{1, 1}, cltf.den{1, 1});
                       % Calculate figures of merit
Figures of merit for a unit step input

rise time     =  0.11
peak value    =  1.1006
peak time     =  0.25
setting time  =  0.36
final value   =  1
```

DIR Design (Exact z Domain Design)

```
echo on
gxnum=4.2;                        % Create numerator
gxden=conv([1  1  0], [1  5]);    % Create denominator
Gx=tf(gxnum, gxden);              % Create transfer function
                                  % from num and den
zpk(Gx)                           % Show zero - pole - gain form

Zero/pole/gain:
     4.2
- - - - - - - - -
s (s+5) (s+1)

Tsamp=0.01;                      % Tsamp is sample time=0.1 sec
Gz=c2d(Gx, Tsamp, 'zoh')         % Transform Gx to the z domain
```

```
Transfer function:
6.896e-007 z^2+2.717e-006 z+6.692e-007
--------------------------------------
       z^3-2.941 z^2+2.883 z-0.9418

Sampling time: 0.01
zpk(Gz)                              % Show zero - pole - gain form

Zero/pole/gain:
6.8961e-007 (z+3.677) (z+0.2639)
--------------------------------
   (z-1) (z-0.99) (z-0.9512)

Sampling time: 0.01
Dcnum=conv([1  1], [1  5]);     % Place zeros
Dcden=conv([1  39.85], [1  39.85]); % Place poles
Dc=tf(Dcnum, Dcden);          % Assemble lead-lag compensator
Dcz=c2d(Dc, Tsamp, 'tustin');  % Transform compensator
zpk(Dcz)                        % Show compensator for Eq. 16.79

Zero/pole/gain:
0.71626 (z-0.99) (z-0.9512)
---------------------------
      (z-0.6677)^2

Sampling time: 0.01
GDIR=series(Dcz, Gz);           % Multiply Dc(z) and Gz(z)
GDIR=minreal(GDIR, 0.00001)     % Cancel pole-zero pairs

Transfer function:
4.939e-007 z^2+1.946e-006 z+4.793e-007
--------------------------------------
      z^3-2.335 z^2+1.781 z-0.4458

Sampling time: 0.01
zpk(GDIR)                        % Show zero - pole - gain form
Zero/pole/gain:
4.9394e-007 (z+3.677) (z+0.2639)
--------------------------------
      (z-1) (z-0.6677)^2

Sampling time: 0.01
rltool(GDIR)                    % Open root locus tool window

% These commands are used to plot the closed loop system
T_r2y                       % Show closed loop transfer function

Zero/pole/gain from input ''r'' to output ''y'':
 0.002097 (z+3.677) (z+0.2639)
------------------------------
(z-0.5428) (z^2-1.791 z+0.8177)
```

```
Sampling time: 0.01
hold on                                    % Hold plot from above
step(T_r2y)                                % Plot step response
```

As noted in Fig. 16. 21, the FOM for the DIG design for $T = 0.1$ s agree very closely with those for the DIR design. It is left to the reader to obtain corresponding MATLAB plots for $T = 0.01$ and 1.0 s.

16.13 FEEDBACK COMPENSATION: TRACKING

System performance, based upon the system output *tracking or following* the desired system input, can be improved by using either cascade or feedback compensation. The factors listed in Sec. 11.1 must be considered when making that decision. The effectiveness of cascade compensation is illustrated in the preceding sections. The use of feedback compensation is demonstrated in the next three sections, not only for a control system whose output must track a desired system input, but also for a control system whose output should not respond to a system disturbance input. The feedback compensation design procedures for continuous-time systems can be applied to S-D systems. These procedures are not so straightforward and easy to apply as the design procedures for cascade compensation.

The feedback compensation procedures of Chap. 12, for tracking and disturbance rejection, can be readily applied to the PCT representation of the digital control system. Once the compensator in the s domain has been synthesized it then can be readily transformed into the z-domain controller by use of the Tustin transformation.

16.13.1 General Analysis

The plant transfer for the digital control system of Fig. 16. 22 is

$$G_x(s) = \frac{K_x}{s(s+1)} \tag{16.81}$$

The following equations describe the control system of Fig. 16. 22:

$$E(s) = R(s) - B(s) \tag{16.82}$$

$$C(s) = G_x(s)E(s) = G_x(s)R(s) - G_x(s)B(s) \tag{16.83}$$

$$B(s) = H_{zo}(s)I^*(s) = H_{zo}(s)H_c^*(s)C^*(s) \tag{16.84}$$

Inserting $B(s)$ into Eq. (16.83) yields

$$C(s) = G_x(s)R(s) - G_x(s)H_{zo}(s)H_c^*(s)C^*(s) \tag{16.85}$$

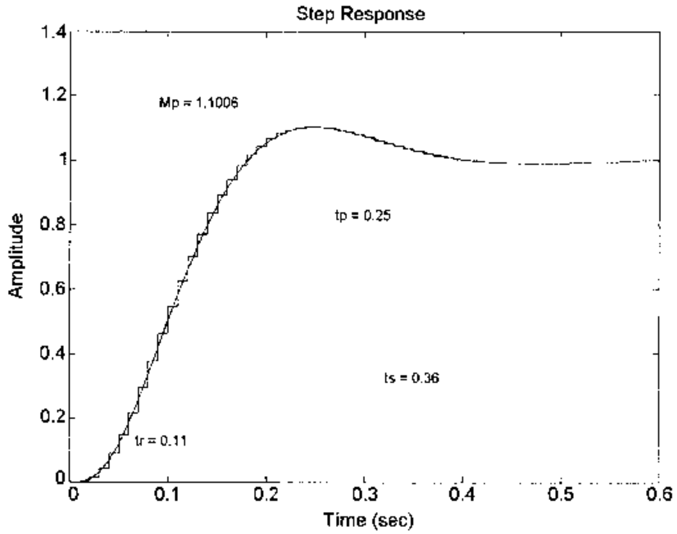


FIGURE 16.21 MATLAB step response for $T = 0.1$ s and figures of merit for Eqs. (16.78) and (16.80).

The impulse transform of this equation is

$$C^*(s) = G_x R^*(s) - H_{zo} G_x^*(s) H_c^*(s) C^*(s) \tag{16.86}$$

Solving for $C^*(s)$ yields

$$C^*(s) = \frac{G_x R^*(s)}{1 + H_{zo} G_x^*(s) H_c^*(s)} \tag{16.87}$$

and the resulting z-domain model is

$$C(z) = \frac{G_x R(z)}{1 + H_{zo} G_x(z) H_c(z)} \tag{16.88}$$

A root-locus analysis for the digital nonunity-feedback control system of Fig. 16.22 requires the determination of its characteristic equation

$$Q(z) = 1 + H_{zo} G_x(z) H_c(z) = 0 \tag{16.89}$$

which is rearranged to

$$H_{zo} G_x(z) H_c(z) = -1 \tag{16.90}$$

to yield the mathematical format of Eq. (15.92) that is required for a root-locus analysis.

For the given $G_x(s)$ of Eq. (16.81) and $R(s) = 1/s$,

$$G_x R(z) = \frac{K_x T z}{(z-1)^2} - \frac{K_x(1-e^{-T})z}{(z-1)(z-e^{-T})} \tag{16.91}$$

$$H_{zo} G_x(z) = \frac{K_x T}{z-1} - \frac{K_x(1-e^{-T})}{z-e^{-T}} \tag{16.92}$$

Let the controller take the form

$$H_c(z) \equiv K_{zc} \frac{N(z)}{D(z)} \tag{16.93}$$

Substituting Eqs. (16.92) and (16.93) into Eq. (16.90) yields

$$\begin{aligned} &\frac{K_{zc} K_x N(z)[T(z-e^{-T}) - (1-e^{-T})(z-1)]}{(z-1)(z-e^{-T})D(z)} \\ &\quad = \frac{K_{zc} K_x (T - 1 + e^{-T})(z-d)N(z)}{(z-1)(z-e^{-T})D(z)} = -1 \end{aligned} \tag{16.94}$$

where $d = (1 - e^{-T} - Te^{-T})/(1 - e^{-T} - T)$.

For a more complicated nonunity-feedback system, containing minor loops, it may be simpler to substitute Eqs. (16.91) to (16.93) into Eq. (16.90) to yield

$$\begin{aligned} C(z) &= \frac{K_x \dfrac{Tz(z-e^{-T}) - z(1-e^{-T})(z-1)}{(z-1)^2(z-e^{-T})}}{1 + K_{zc}K_x \left[\dfrac{T(z-e^{-T}) - (1-e^{-T})(z-1)}{(z-1)(z-e^{-T})}\right] \dfrac{N(z)}{D(z)}} \\ &= \frac{K_x[Tz(z-e^{-T}) - z(1-e^{-T})(z-1)]D(z)}{\underbrace{(z-1)}_{\substack{\text{Forcing} \\ \text{function pole}}} \{(z-1)(z-e^{-T})D(z) + K_{zc}K_x[T(z-e^{-T}) - (1-e^{-T})(z-1)]N(z)\}} \end{aligned} \tag{16.95}$$

The characteristic equation

$$(z-1)(z-e^{-T})D(z) + K_{zc}K_x[T(z-e^{-T}) - (1-e^{-T})(z-1)]N(z) = 0 \tag{16.96}$$

is partitioned [5] and rearranged to yield Eq. (16.94).

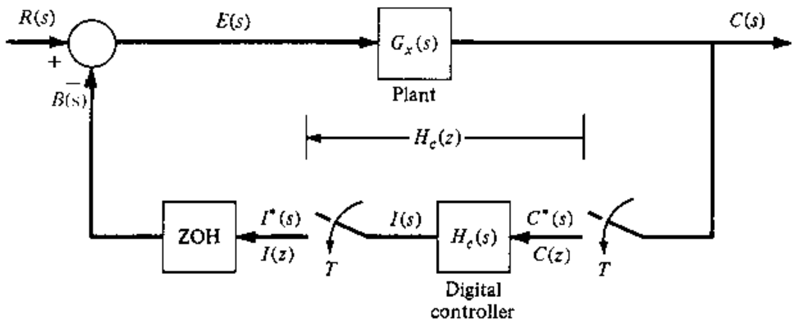


FIGURE 16.22 A feedback compensated digital control system.

Equation (16.94) permits a root-locus analysis for the sampled-data control system of Fig. 16.22. Assuming that

$$H_c(z) = K_{zc}\frac{z-c}{z-b} = K_{zc}\frac{N(z)}{D(z)} \tag{16.97}$$

the plot of the poles and zeros of Eq. (16.94) is shown in Fig. 16.23. For a desired ζ and settling time $T_s \le 4/|\sigma|$, the desired dominant complex pole p_1 is located as illustrated in this figure. If it is not possible for the basic system to produce the desired dominant poles, the angle condition may be used to locate the zero c and the pole b of H_c that can produce the desired dominant poles. If the simple feedback compensator of Eq. (16.97) is not satisfactory, then the angle condition is again applied but with a more complicated $H_c(z)$ function. Use of computer-aided-design (CAD) programs can expedite the root-locus analysis.

If a zero(s) [or pole(s)] of $H_c(\cdot)$ are selected to cancel a pole(s) [or zero(s)] of the basic plant $G(\cdot)$, then they become poles of $C(\cdot)$ (see Chap. 11). This characteristic can be observed in this example by selecting the zero of $H_c(z)$ [Eq. (16.97)] equal to the pole at $-e^{-T}$ of $H_{zo}G_x(z)$ [Eq. (16.92)]. Thus, by this design method, the term $(z-e^{-T})$ can be factored out of the bracketed term in the denominator of Eq. (16.95), resulting in this term becoming a pole of $C(z)$.

In correlating the root-locus plot with respect to the corresponding time-response characteristics, one must be aware of which poles and zeros of Eq. (16.94) are zeros of $C(z)$. As mentioned previously, a pole of $C(z)$ that lies very close to a zero of $C(z)$ can have little effect on $c^*(t)$.

The feedback compensation design example of this section is based upon achieving an $M_p \approx 1.1$, $t_s \le 1$ s, and $c(t)_{ss} = 1$, for a unit-step forcing function and $T = 0.1$ s. Solving Eqs. (3.60) and (3.64) for these values of M_p and t_s

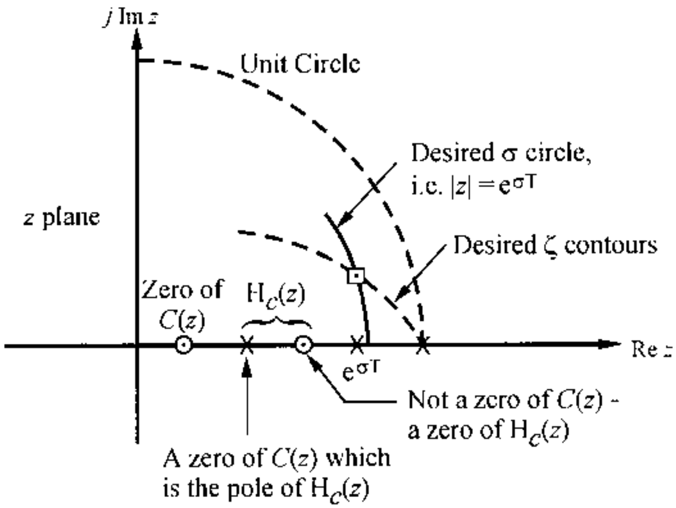


FIGURE 16.23 A plot of the poles and zeros of Eq. (16.94).

yields $\zeta \approx 0.59$, $|\sigma| \geq 4$, and $\omega_n \approx 6.7795$. Thus, the desired dominant second-order s-plane poles are

$$p_{1,2_s} = \sigma \pm j\omega_n\sqrt{1-\zeta^2} = -4 \pm j5.4738 \tag{16.98}$$

which map into the z plane as follows:

$$p_{1,2_z} = e^{p_{1,2}T} = 0.6703\angle \pm 31.363^\circ = 0.5721 \pm j0.3487 \tag{16.99}$$

These desired dominant z-plane poles, for this example, cannot be achieved by use of only $H_c(z) = K_{zc}$; i.e., the desired dominant poles cannot be obtained by only a gain adjustment (see Sec. 16.7.2). It may be possible to achieve these desired dominant poles with a simple lead, lag, or lag-lead controller.

16.13.2 DIG Technique for Feedback Control

The digitization (DIG) approach to digital design requires the pseudo-continuous-time (PCT) control-system approximation of the sampled-data system of Fig. 16.22 shown in Fig. 16.24, where, for $T = 0.1$ s, the plant transfer function is given by Eq. (16.81), and

$$H_x(s) = H_A(s)H_c(s) = \frac{20K_{sc}N(s)}{(s+20)D(s)} \tag{16.100}$$

$H_{pa}(s) = 2/(s+20)$ is the Padé approximation of $H_{zo}(s)$, and $H_A(s) = H_{pa}(s)/T$. For the approach of this subsection the $H_c(s)$ is assumed to be of the form:

$$H_c(s) = \frac{K_{sc}(s+a)^2}{(s+b)(s+c)} \tag{16.101}$$

Normally, as a first trial, one should start with the simple compensator $H_c(s) = K_{sc}(s+a)/(s+b)$ in order to try to achieve the desired FOM.

Although the desired dominant roots $p_{1,2_s} = -4 \pm j5.4738$ are not within the shaded area of Fig. 16.5 for a very good Tustin approximation, they still lie within the overall region where the condition of Eq. (16.15), $|\sigma_{1,2}| \ll 2/T = 20$, is satisfied and the Tustin approximation can yield a satisfactory system design. In general, the poles and zeros of $H_c(s)$ are selected to be real. For $H_c(s)$ to act effectively as two "simple lead networks" in cascade the poles of Eq. (16.101) are located as far to the left of the zeros as practical. They are also selected in such a manner that when they are mapped into the z domain, via the Tustin transformation, they lie between 0 and +1 on the positive real axis of the z domain. Substituting from Eq. (16.8) into Eq. (16.101) yields

$$H_c(z) = \frac{(2+aT)^2(K_{sc})}{(2+bT)(2+cT)}\left[\frac{\left(z-\frac{2-aT}{2+aT}\right)^2}{\left(2-\frac{2-bT}{2+bT}\right)\left(\frac{z-2-cT}{2+cT}\right)}\right] \tag{16.102}$$

For positive real poles and zeros to exist between 0 and +1 for Eq. (16.102), then $aT < 2$, $bT < 2$, and $cT < 2$. Assume that the poles in Eq. (16.102) are at the origin; i.e.,

$$2 - bT = 2 - cT = 0 \tag{16.103}$$

that results in $c = b = 2/T$. Note that poles at $z = 0$ in the z domain, by means of the transformation $z = e^{sT}$, correspond to poles at $-\infty$ in the s plane.

The open-loop transfer function of Fig. 16.24, with $T = 0.1$ s and $c = b = 2/T$, is

$$G_x(s)H_x(s) = \frac{20K_{sc}K_x(s+a)^2}{s(s+1)(s+20)^3} \tag{16.104}$$

The poles of Eq. (16.104) and the desired dominant closed-loop roots $p_{1,2}$ are plotted in Fig. 16.25. The angle condition is then applied to determine where the zeros at $-a$ must be placed in order for the desired dominant roots to be poles of $C(s)/R(s)$. The application of this condition results in $a \approx 7.062$. Once the value of a has been determined, the magnitude condition or a computer

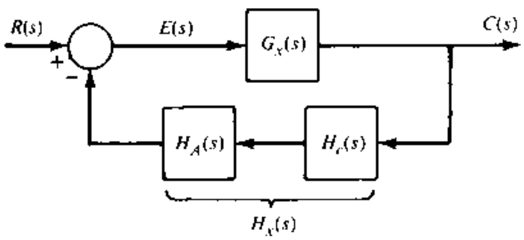


FIGURE 16.24 PCT system approximation of the sampled-data control system of Fig. 16.22.

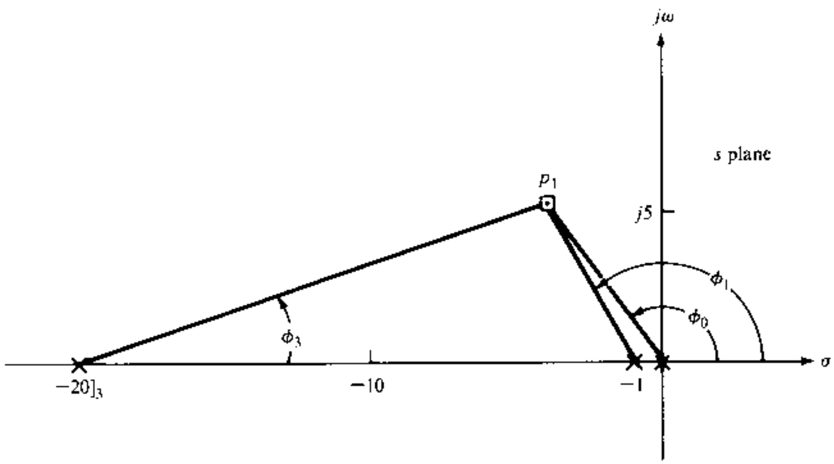


FIGURE 16.25 The plot of the poles of Eq. (16.104).

can then be used to determine the static loop sensitivity at $p_{1,2}$ and the remaining closed-loop poles. For this example, $20K_{sc}K_x \approx 5201$, $p_{3,4} \approx -8.9772 \pm j8.976$, $p_5 \approx -35.06$, and

$$\frac{C(s)}{R(s)} = \frac{K_x(s+20)^3}{(s-p_1)(s-p_2)(s-p_3)(s-p_4)(s-p_5)} \tag{16.105}$$

To satisfy the requirement of $c(t)_{ss} = 1$, for $r(t) = u_{-1}(t)$, the final value theorem is applied to the expression for $C(s)$ to generate the value of $K_x \approx 32.425$. This value of K_x results in $K_{sc} \approx 8.02$. The FOM for the control ratio of Eq. (16.105) are given in Table 16.6.

TABLE 16.6 Figures of Merit of DIG Control System of Figs. 16.22 and 16.25

Domain	M_p	t_p, s	t_s, s	K_1, s^{-1}
s	1.134	0.554	0.86+	6.096
z	1.088	0.6	0.9+	4.51

To determine the ramp error coefficient, it is necessary to solve for $G_{eq}(s)$ from

$$\frac{C(s)}{R(s)} = \frac{32.425(s+20)^3}{s^5 + 61s^4 + 1260s^3 + 14{,}400s^2 + 81{,}460s + 259{,}400} = \frac{N}{D}$$

$$= \frac{G_{eq}(s)}{1 + G_{eq}(s)}$$

Thus,

$$G_{eq}(s) = \frac{N}{D-N} = \frac{32.425(s+20)^3}{s^5 + 61s^4 + 12227.6s^3 + 12{,}455s^2 + 42{,}550s} \tag{16.106}$$

and

$$K_1 = \lim_{s \to 0} sG_{eq}(s) = 6.096 \text{ s}^{-1}$$

Although M_p is larger than desired (1.1) for the PCT system, thus Tustin transformation of $H_c(s)$ is obtained next to determine if this first trial design results in a sampled-data system whose system performance is "close" to the desired specifications. Substituting the values T, a, b, c, and K_{sc} into Eq. (16.102) yields

$$H_c(z) = \frac{K_{zc}(z - 0.48710)^2}{z^2} = K_{zc}\frac{N(z)}{D(z)} \tag{16.107}$$

where $K_{zc} = 3.6709$. Substituting Eq. (16.107) into Eq. (16.94), with $T = 0.1$, yields

$$\frac{0.0048374 K_x K_{zc}(z + 0.967255)(z - 0.48710)^2}{z^2(z-1)(z-0.90480)} = -1 \tag{16.108}$$

where $K = 0.0048374 K_x K_{zc} = 0.0048374(32.425)(3.6709) = 0.575791$. With this value of K the roots of the characteristic equation can be determined.

To determine the system's FOM it is necessary to obtain the *pseudo-control ratio* for a nonunity feedback digital control system. To accomplish this the input forcing function $r(t)$ must be specified. For this example, it is noted that the right-hand side of Eq. (16.95) contains the factor $z/(z-1)$ that is the $\mathcal{Z}$ transform of the specified input forcing function, a step forcing function. By dividing both sides of this equation by this factor yields the pseudo-control ratio as follows:

$$\left[\frac{C(z)}{R(z)}\right]_p = \frac{K_x[T_z(z-e^{-T}) - z(1-e^{-T})(z-1)]D(z)}{(z-1)(z-e^{-T})D(z) + K_{zc}K_x[T(z-e^{-T}) - (1-e^{-T})(z-1)]N(z)} \tag{16.109}$$

For the value of $T = 0.1$ and by substituting from Eq. (16.107) into Eq. (16.109) with K_x and K_{zc} unspecified and inserting the poles as determined by Eq. (16. 108) yields

$$\begin{aligned}\left[\frac{C(z)}{R(z)}\right]_p &= \frac{0.0048374K_x z^2(z+0.967255)}{z^2(z-1)(z-0.9048) + K(z+0.967255)(z-0.4871)^2} \\ &= \frac{0.0048374K_x z^2(z+0967255)}{(z-0.60086 \pm j0.31144)(z-0.063643 \pm j0.53202)} \\ &= \frac{N}{D} = \frac{G_{\text{eq}}(z)}{1+G_{\text{eq}}(z)}\end{aligned} \tag{16.110}$$

where $K = 0.0048374K_xK_{sc} = 0.575791$. The value of K_x determined in the PCT system design is not used because of the warping effect. This effect is noted by the comparison of the desired z-domain roots $p_{1,2_z} = e^{p_{1,2}T} = 0.5721 \pm j0.3487$ obtained by selecting path B of Fig. 16.7 with the roots $p_{1,2_z} = 0.6001 \pm j0.3112$ obtained by selecting path A. The difference between these roots result from the DIG technique for which the $p_{1,2_s}$ do not lie in the shaded area of Fig. 16.5.

The value of K_x is determined by satisfying the condition

$$c(\infty) = \lim_{z\to 1}\,[(1-z^{-1})C(z)] = 1$$

where $R(z) = z/(z-1)$. Thus, Eq. (16.110) requires that $K_x = 31.09728$. To satisfy the required value of $K = 0.57579$ with this value of K_x, $K_{zc} = 3.82763$.

The expression of $G_{\text{eq}}(z)$ is determined from Eq. (16.110) to be

$$G_{\text{eq}}(z) = \frac{N}{D-N} = \frac{0.15115z^2(z+0.967255)}{(z-1)(z-0.482043)(z+0.001003 \pm j0.522266)}$$

Thus,

$$K_1 = \frac{1}{T}\lim_{z\to 1}\left[\frac{z-1}{z}G_{\mathrm{eq}}(z)\right] = 4.50998\ \mathrm{s}^{-1}$$

As noted from Table 16. 6, the z-domain FOM have satisfied the desired performance specifications; thus, the design is satisfactory. The required digital controller is

$$H_c(z) = \frac{I(z)}{C(z)} = \frac{3.82763(z-0.4871)^2}{z^2} = 3.82763(1-0.4871z^{-1})^2 \tag{16.111}$$

the results in the difference equation

$$i(kT) = 3.82763c(kT) - 3.7288877c[(k-1)T] + 0.90816803c[(k-2)T] \tag{16.112}$$

This discrete control law (algorithm) can be translated into a software program to be implemented by the digital computer (controller). Note that the third coefficient requires at least a 30-bit integer computer word.

16.14 CONTROLLING UNWANTED DISTURBANCES

The design approach for minimizing unwanted disturbances for continuous-time systems is presented in Chap. 12. In this section the design approach for continuous-time systems is applied to minimize unwanted disturbances for sample-data systems by use of the DIG technique. The design procedure of this section has the objective that the system output should not follow the disturbance input This is the opposite situation from the design procedures given in Chaps. 10 and 11 for analog systems where the output tracks the command input.

16.14.1 PCT DIG Technique

Figure 16.26 is the PCT control system representation of the sampled data system of Fig. 16.22. In Example 2 of Sec. 12.7 the feedback compensator $H_c(s) = K_c/s$ resulted in an unstable system. An analysis of the root locus of Fig. 12.10 indicates that $H_c(s)$ must contain at least one zero in order to pull the root-locus branches into the left-half s plane.

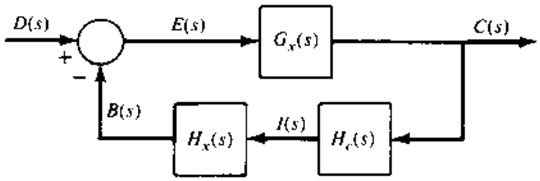


FIGURE 16.26 A control system with a disturbance input.

The result of Prob. 12.7 indicate that to minimize the magnitudes of all coefficients of

$$c(t) = A_1 e^{p_1 t} + \cdots + A_n e^{p_n t} \tag{16.113}$$

that are desired for minimizing the effect of an unwanted disturbance input on $c(t)$, it is necessary to locate as many of the poles of $C(s)/D(s)$ as far to the left of the s plane as possible. Figure 16.27 graphically describes the result of the analysis made in Prob. 12.7 for the desired placement of the control ratio poles. To achieve the desired location of the poles and yet pull the root-locus branches as far as possible to the left, initially assume the following structure for the feedback compensator:

$$H_c(s) = \frac{K_{sc}(s + a_1)(s + a_2)}{s(s + b_1)(s + b_2)} \tag{16.114}$$

As a trial, let $a_1 = 4$ and $a_2 = 8$. The poles b_1 and b_2 are chosen as far left as possible and yet should lie in the Tustin region 0 to $-2/T$ of Fig. 16.5. For this example, $T = 0.01$ s and $2/T = 200$. The values of b_1 and b_2 are both set equal to 200; therefore, the open-loop transfer function of Fig. 16.26 with $b_1 = b_2$ is

$$G_x(s)H_x(s)H_c(s) = \frac{200K_{sc}(s + 4)(s + 8)}{s^2(s + 1)(s + 200)^3} \tag{16.115}$$

where $G_x(s) = 1/[s(s+1)]$ and $H_x(s) = 200/(s + 200)$. The poles and zeros of Eq. (16.115) are plotted in Fig. 16.27. For the plot Lm $1/\mathrm{H}_x(j\omega)\mathrm{H}_c(j\omega) \le -54$ dB, where $|\mathrm{C}(j\omega)/\mathrm{D}(j\omega)1| \approx |1/\mathrm{H}_x(j\omega)\mathrm{H}_c(j\omega)|$ over the range $0 \le \omega \le \omega_b = 10$ rad/s, requires that $200K_{sc} \ge 50 \times 10^7$. The static loop sensitivity value of 50×10^7 yields the following set of roots:

The associated root locus is shown in Fig. 16.28. This choice of gain results in a satisfactory design, since $M_p = 0.001076 \le 0.002(-54\ \text{dB})$, $t_p = 0.155$ s (< 0.2 s), $c(t)_{ss} = 0$, and $0 \le \omega \le \omega_b = 10$ rad/s.

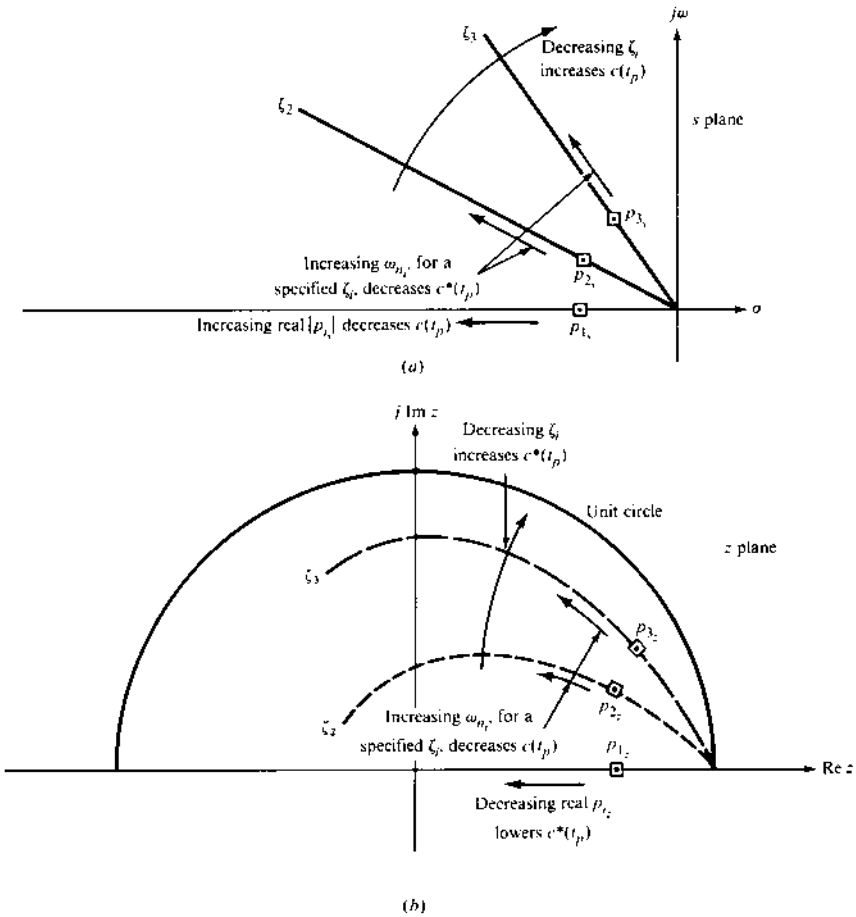


FIGURE 16.27 Desired directional changes in the parameters ζ_i and ω_{n_i} and the control ratio's real pole values for decreasing the peak overshoot to a disturbance input: (*a*) *s* plane; (*b*) *z* plane.

Applying the Tustin transformation to Eq. (16.114), with $a_1 = 4$, $a_2 = 8$, and the two poles at -200, yields

$$H_c(z) = \frac{0.3315 \times 10^{+4}(z - 0.9231)(z - 0.9608)(z + 1)}{z^2(z - 1)} \tag{16.116}$$

where $K_{sc} = 0.3315 \times 10^{+4}$. The zero at $z = -1(s = \pm j\omega_s T/2 = \pm j\pi/T)$ can cause oscillation due to the disturbances. For the system of Fig. 16.22, with

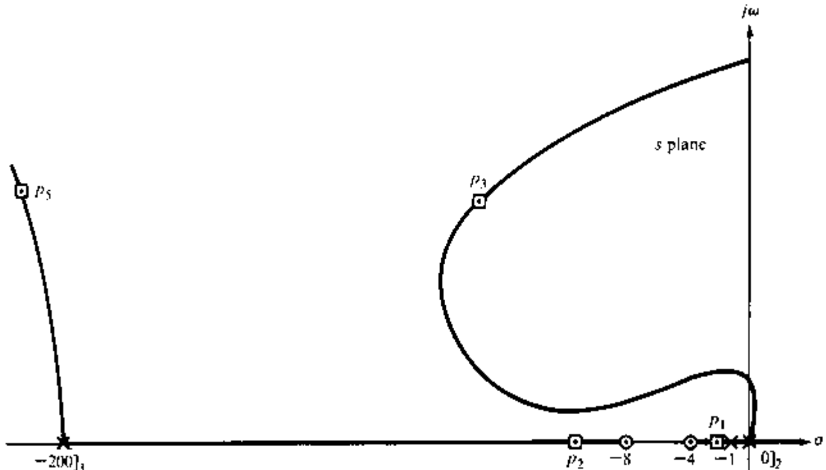


FIGURE 16.28 A possible root locus for Eq. (16.115) (not to scale).

$r(t) = u_{-1}(t)$ and $T = 0.01$ s, the resulting model is

$$G_x R(z) = \mathcal{Z}[G_x(s)R(s)] = \frac{0.4983 \times 10^{-4} z(z + 0.9967)}{(z-1)^2(z - 0.9900)} \tag{16.117}$$

$$H_{zo}G_x(z) = \mathcal{Z}[H_{zo}(s)G_x(s)] = \frac{0.4983 \times 10^{-4}(z + 0.9967)}{(z-1)(z - 0.9900)} \tag{16.118}$$

The characteristic equation, from Eq. (16.89), yields

$$H_{zo}G_x(z)H_c(z) = -1 \tag{16.119}$$

Substituting from Eqs. (16.116) and (16.118) into Eq. (16.119) yields

$$\frac{0.16519(z - 0.9231)(z - 0.9608)(z + 0.9967)(z + 1)}{z^2(z - 0.99)(z - 1)^2} = -1 \tag{16.120}$$

from which the root locus may be obtained. The roots for the specified value of static loop sensitivity yields the following expression for $C(z)$:

$$C(z) = \frac{4.983 \times 10^{-5}(z + 0.9967)z^2}{(z - 0.90403)(z - 0.196227)(z - 0.60787 \pm j0.53203)(z + 0.25723)} \tag{16.121}$$

The resulting performance characteristics for the sampled-data control system of Fig. 16.22 are

$$M_p = 1.08 \times 10^{-3} (\approx -59.3\text{dB}) \qquad \text{and} \qquad t_p \approx 0.16 \text{ s}$$
$$M_m = 1.304 \times 10^{-3} (\approx -57.7\text{dB}) \qquad \text{for} \qquad 0 \leq \omega \leq 10 \text{ rad/s}$$

The gain $K_{zc} = 3315$ can be considered as "high," and thus present a problem with respect to its implementation on the digital controller. If this is the case, then a portion of this gain can be allocated to other components in the feedback path. Also, the gain can be reduced until either $M_p = 0.002$ with $t_p \leq 0.2$ s or $t_p = 0.2$ s with $M_p \leq 0.002$ is achieved. Although not specified in this problem, the loop transmission frequency ω_ϕ must be taken into account in the design of a practical system.

16.15 EXTENSIVE DIGITAL FEEDBACK COMPENSATOR EXAMPLE

The basic second-order system and the associated desired FOM of Example 16.3 of Sec. 16.12 are utilized in this section to illustrate the design of a feedback compensator via the PCT DIG technique. The cancellation of poles (or zeros) of the plant by zeros (or poles) of $D_c(s)$ that can be done in the design of a cascade compensator, cannot be done in the design of the feedback compensator $H_c(s)$. As stated in Sec. 16.13.1, the canceled term(s) become pole(s) of $C(s)$, which in general is counterproductive to the achievement of the desired FOM.

16.15.1 PCT DIG Example

For Fig. 16.24, where $T = 0.01$ s, the open-loop PCT transfer function is

$$G_x(s)H_x(s) = \frac{200K_xK_{sc}N(s)}{s(s+1)(s+5)(s+200)D(s)} \tag{16.122}$$

This equation is identical to Eq. (16.72), except $D_c(s)$ is replaced by $H_c(s)$. Thus, the root-locus analysis of Example 16.3 holds for Eq. (16.122). Initially, consider a second-order continuous controller

$$H_x(s) = \frac{200K_{sc}(s+a)^2}{(s+200)(s+b)^2} \tag{16.123}$$

In order to minimize the effect of a dominant real root, the zeros of Eq. (16.122) at $-a$ are located to the left of the pole at -5. Thus, as a first trial, select $a = 6$. In order to achieve the desired closed-loop poles, $p_{1,2_s} = -10 \pm j14$, the application of the angle and magnitude conditions in Fig. 16.22 yield a value of $b = 62.89$ for the compensator pole and a static loop sensitivity value of $200K_xK_{sc} = 1.151 \times 10^5$. The resulting control ratio is

$$\frac{C(s)}{R(s)} = \frac{K_x(s+62.89)^2(s+200)}{(s+5.338)(s+13.70)(s+10.08 \pm j14.12)(s+95.76)(s+196.8)} \tag{16.124}$$

Applying the final value theorem to the expression for $C(s)$, with $R(s) = 1/s$, yields a value of $K_x = 523.97$ for $e(t)_{ss} = 0$. Thus, $K_{sc} = 109.83$ and the FOM for Eq. (16.124) are $M_p = 1.000$ and $t_s = 0.86$ s. As Eq. (16.124) predicts (real pole at -5.338 is more dominant than $p_{1,2}$), an overdamped response is achieved with $t_s = 0.867 > 0.5$ s (desired).

To move the dominant real pole of Eq. (16.124) farther to the left in order to decrease the value of t_s and achieve an underdamped response, it is necessary to increase the number of poles and zeros of $H_c(s)$ to 3. As a second trial, let $a = 11$ for the three zeros of $H_c(s)$. For this choice of a, the following design information is obtained:

$$b = 39.702 \qquad 200K_xK_{sc} = 1.038 \times 10^7 \qquad K_x = 11049 \qquad K_{sc} = 47$$

$$\frac{C(s)}{R(s)} = \frac{K_x(s+39.702)^3(s+200)}{(s+7.109)(s+10.05 \pm j14.08)(s+10.79 \pm j1736)(s+78.6)(s+197.7)} \tag{16.125}$$

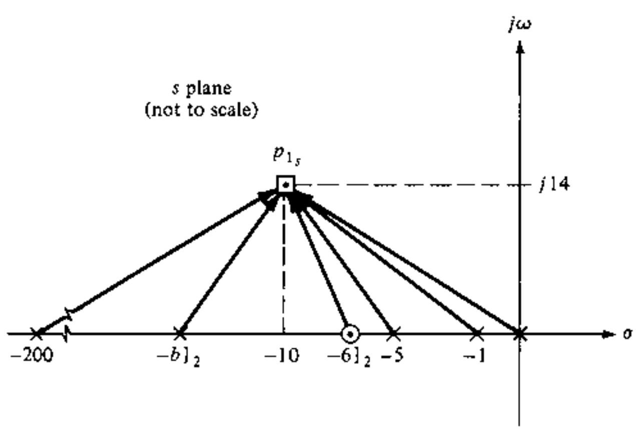


FIGURE 16.29 The plot of the poles and zeros of Eq. (16.122): first trial.

$M_p = 1.000$ and $t_s = 0.603 > 0.5$ s (desired). As noted by this design example, in order to achieve the desired underdamped response, $p_{1,2_s}$ and FOM requires a high order $H_c(s)$. Based upon this conclusion, the cascade compensator designs of Sec. 16.12 are more practical in the sense of lower-order controllers.

16.16 CONTROLLER IMPLEMENTATION [1]

In designing a digital controller it is very important that the implementation issue, as stressed in Fig. 1.16, be kept in mind. In designing the digital controller to achieve the desired control system performance requirements, the engineer must be aware of the factors that play an important role in the implementation of the controller. For example, how is maximum computational accuracy achieved?

The factors considered in the implementation of a digital controller are discussed in detail in Ref. 1. This section is intended only to stress to digital control system designers a few important factors in implementation of digital controllers.

As an example, consider a lead controller that was designed to yield the desired system FOM, whose transfer function is

$$D_c(z) = \frac{K_{zc}(z - z_a)(z - z_b)}{(z - p_c)(z - p_d)} \tag{16.126}$$

The parameter values of the controller are:

$$K_{zc} = 2002.6 \qquad z_a = 0.99005 \qquad z_b = 0.95123 \qquad p_c = p_d = 0.6696$$

Since it is desired to implement this controller by a theoretical 7-bit digital computer, it is necessary that the parameters of Eq. (16.118) be rounded off to two decimal digits ($2 \times 3.32 = 6.64 \approx 7$); i.e.,

$$D_c(z) = \frac{200(z - 0.95)(z - 0.99)}{(z - 0.67)^2} \tag{16.127}$$

Thus, the resulting control ratio now becomes

$$\frac{C(z)}{R(z)} = \frac{1.3792 \times 10^{-3}(z + 0.26395)(z + 3.6767)(z - 0.95)(z - 0.99)}{(z - 0.8883 \pm j0.07586)(z - 0.5647)(z - 0.9487)(z - 0.9901)} \tag{16.128}$$

This yields the following FOM using essentially infinite arithmetic operation accuracy:

$$M_p = 1.023 \qquad t_p = 0.38 \text{ s} \qquad t_s = 0.43 + \text{s} \qquad K_1 = 7.714 \text{ s}^{-1}$$

Comparing these FOM with those obtained without the roundoff show that the values of M_p and K_1 are decreased but the values for t_p and t_s have increased. The control engineer has to make the decision as to whether the theoretical FOM resulting from using the controller of Eq. (16.128) are or are not acceptable before taking into account factors of controller implementation including word length.

As another example, a digital control system is designed utilizing the lag controller

$$D_c(z) = \frac{z - z_a}{z - p_b} \tag{16.129}$$

where $z_a = 0.999$ and $p_b = 0.9999$ that satisfied the desired FOM. Since it is desired to implement this controller by a theoretical 7-bit digital computer, it is necessary that the parameters of Eq. (16.129) be rounded off to two decimal digits, i.e.,

$$D_c(z) = \frac{z - 0.99}{z - 1} \tag{16.130}$$

Thus, the resulting control ratio now becomes

$$\frac{C(z)}{R(z)} = \frac{2.1 \times 10^{-3}(z + 0.26396)(z + 3.6767)(z - 0.99)}{(z - 1.090 \pm j0.2140)(z - 0.9900+)(z - 0.7696)} \tag{16.131}$$

which results in an unstable system because of underflow (see Chap. 10 of Ref. 1).

Thus, in rounding up or down the parameters of a controller care must be exercised in the degree of accuracy that must be maintained in selecting a pole-zero placement [1]. The following general guidelines need to be kept in mind when rounding up or down the controller coefficients:

1. Rounding up the controller zeros (closer to the UC) and rounding down the poles (further away from the UC), regardless of the magnitude of the "excess" digits, results in a more stable system.
2. Controller implementation fine-tuning of the controller gain in the simulation phase of the design process can result in the achievement of the desired FOM.

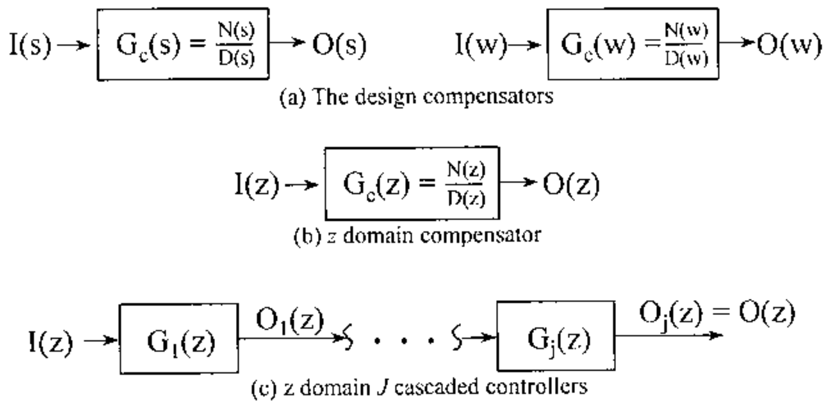


FIGURE 16.30 *s*- or *w*-domain to *z*-domain bilinear transformation: formulation for implementation of the $G(z)$ controller.

Another important factor is the software implementation of the digital controller design. Tight performance specifications and a high degree of uncertainty require small sampling intervals T. Unfortunately, the smaller the value of T, the greater the degree of accuracy that is required to be maintained. The numerical accuracy is enhanced by a factored representation of the controller and prefilter [1]. For example, by use of the bilinear transformation the controller $G_z(z)$ of Fig. 16.30*b* is obtained from the compensator $G_c(s)$ or $G_c(w)$ of Fig. 16.30*a*. Therefore, the equivalent cascaded transfer function representation (factored representation) of $G_z(z)$, shown in Fig. 16.30*c*, is utilized to obtained the algorithm for the software implementation of $G_z(z)$ in order to improve the accuracy of the system's performance due to this controller.

The reader is referred to the technical literature that discuss the other factors involved in controller implementation, e.g., Ref. 1.

16.17 SUMMARY

The availability of inexpensive microprocessors has led to their extensive use in digital control systems. Although there is a degradation of performance by the introduction of sampling, this is more than overcome by the greater range of digital controllers that can be achieved. The flexibility provided by the use of microprocessors in synthesizing digital controllers yields greatly improved control system performance.

Two approaches for analysing a sampled-data control system are presented: the DIG (digitization) and DIR (direct) techniques. The former requires the use of the Padé approximation and the Tustin transformation for an initial design in the s plane (PCT control-system configuration). If the Tustin approximation criteria of Sec. 16.3 are satisfied, the correlation between the two planes is very good. If the approximations are not valid, the analysis and design of the digital control system should be done by the DIR technique.

The conventional procedures (see Chap. 10) for the design of analog cascade compensators are used for the design of digital control system in the examples of this chapter. They involve a certain amount of trial and error and the exercise of judgement based upon past experience. The improvements produced by each controller are shown by application to a specific system. In cases where the conventional design procedures for cascade digital controllers do not produce the desired results, the method of Sec. 16.15 can be used. The feed-back compensation design procedures of this chapter can be utilized when a cascade compensator cannot produce the desired system performance, or when a "simpler" compensator can be achieved. The respective design procedures for cascade and feedback controllers depend upon the system output *tracking or following* of the system input.

When designing a digital controller, the designer is not restricted to the use of the nominal upper bound of $a = 10$ for a lag compensator and the nominal lower bound of $a = 0.1$ for a lead compensator in the s or w domain. Larger upper and lower bounds, respectively, can be used, based upon the flexibility of the computer hardware in terms of word-length accuracy (see Ref. 1).

The examples of this section are designed to provide a scenario in which a control-system designer needs to select a method based on:

1. The factors listed in Sec. 11.1 with respect to the choice of cascade vs. feedback compensation.
2. The complexity of the controller (the number of poles and zeros).
3. The achievement of the desired FOM.

The use of several compensators in cascade and/or in feedback may be required to achieve the desired pole-zero locations for the control ratio.

REFERENCES

1. Houpis, C. H., and G. B. Lamont: *Digital Control Systems: Theory, Hardware, Software*, 2nd ed., McGraw-Hill, New York, 1992.
2. Franklin, G. F., and J. D. Powell: *Digital Control of Dynamic Systems*, 2nd ed., Addison-Wesley, Reading, Mass., 1990.

3. Cadzow, J. A., and H. Martens: *Discrete-Time and Computer Control Systems*, Prentice-Hall, Englewood Cliffs, N. J., 1974.
4. Tustin, A.: "A Method of Analyzing the Behavior of Linear Systems in Terms of Time Series," *JIEE (Lond.)*, vol. 94, pt. IIA, 1947.
5. D'Azzo, J. J. and C. H. Houpis: *Feedback Control System Analysis and Synthesis*, 2nd ed., McGraw-Hill, New York, 1966.

Appendix A

Table of Laplace Transform Pairs

$F(s)$	$f(t) \quad 0 \le t$
1. 1	$u_0(t)$ unit impulse at $t = 0$
2. $\dfrac{1}{s}$	1 or $u_{-1}(t)$ unit step starting at $t = 0$
3. $\dfrac{1}{s^2}$	$tu_{-1}(t)$ ramp function
4. $\dfrac{1}{s^n}$	$\dfrac{1}{(n-1)!}t^{n-1}$ $n =$ positive integer
5. $\dfrac{1}{s}e^{-as}$	$u_{-1}(t-a)$ unit step starting at $t = a$
6. $\dfrac{1}{s}(1 - e^{-as})$	$u_{-1}(t) - u_{-1}(t-a)$ rectangular pulse
7. $\dfrac{1}{s+a}$	e^{-at} exponential decay
8. $\dfrac{1}{(s+a)^n}$	$\dfrac{1}{(n-1)!}t^{n-1}e^{-at}$ $n =$ positive integer

(*continued*)

$F(s)$		$f(t) \quad 0 \le t$
9.	$\dfrac{1}{s(s+a)}$	$\dfrac{1}{a}(1-e^{-at})$
10.	$\dfrac{1}{s(s+a)(s+b)}$	$\dfrac{1}{ab}\left(1-\dfrac{b}{b-a}e^{-at}+\dfrac{a}{b-a}e^{-bt}\right)$
11.	$\dfrac{s+a}{s(s+a)(s+b)}$	$\dfrac{1}{ab}\left[\alpha-\dfrac{b(\alpha-a)}{b-a}e^{-at}+\dfrac{a(\alpha-b)}{b-a}e^{-bt}\right]$
12.	$\dfrac{1}{(s+a)(s+b)}$	$\dfrac{1}{b-a}(e^{-at}-e^{-bt})$
13.	$\dfrac{s}{(s+a)(s+b)}$	$\dfrac{1}{a-b}(ae^{-at}-be^{-bt})$
14.	$\dfrac{s+\alpha}{(s+a)(s+b)}$	$\dfrac{1}{b-a}[(\alpha-a)e^{-at}-(\alpha-b)e^{-bt}]$
15.	$\dfrac{1}{(s+a)(s+b)(s+c)}$	$\dfrac{e^{-at}}{(b-a)(c-a)}+\dfrac{e^{-bt}}{(c-b)(a-b)}+\dfrac{e^{-ct}}{(a-c)(b-c)}$
16.	$\dfrac{s+\alpha}{(s+a)(s+b)(s+c)}$	$\dfrac{(\alpha-a)e^{-at}}{(b-a)(c-a)}+\dfrac{(\alpha-b)e^{-bt}}{(c-b)(a-b)}+\dfrac{(\alpha-c)e^{-ct}}{(a-c)(b-c)}$
17.	$\dfrac{\omega}{s^2+\omega^2}$	$\sin \omega t$
18.	$\dfrac{s}{s^2+\omega^2}$	$\cos \omega t$
19.	$\dfrac{s+\alpha}{s^2+\omega^2}$	$\dfrac{\sqrt{\alpha^2+\omega^2}}{\omega}\sin(\omega t+\phi) \qquad \phi=\tan^{-1}\dfrac{\omega}{\alpha}$
20.	$\dfrac{s\sin\theta+\omega\cos\theta}{s^2+\omega^2}$	$\sin(\omega t+\theta)$
21.	$\dfrac{1}{s(s^2+\omega^2)}$	$\dfrac{1}{\omega^2}(1-\cos\omega t)$
22.	$\dfrac{s+\alpha}{s(s^2+\omega^2)}$	$\dfrac{\alpha}{\omega^2}-\dfrac{\sqrt{\alpha^2+\omega^2}}{\omega^2}\cos(\omega t+\phi) \qquad \phi=\tan^{-1}\dfrac{\omega}{\alpha}$
23.	$\dfrac{1}{(s+a)(s^2+\omega^2)}$	$\dfrac{e^{-at}}{a^2+\omega^2}+\dfrac{1}{\omega\sqrt{a^2+\omega^2}}\sin(\omega t-\phi) \quad \phi=\tan^{-1}\dfrac{\omega}{a}$
24.	$\dfrac{1}{(s+a)^2+b^2}$	$\dfrac{1}{b}e^{-at}\sin bt$
24*a*.	$\dfrac{1}{s^2+2\zeta\omega_n s+\omega_n^2}$	$\dfrac{1}{\omega_n\sqrt{1-\zeta^2}}e^{-\zeta\omega_n t}\sin \omega_n\sqrt{1-\zeta^2}\ t$

(*continued*)

$F(s)$		$f(t) \quad 0 \le t$
25.	$\dfrac{s+a}{(s+a)^2+b^2}$	$e^{-at}\cos bt$
26.	$\dfrac{s+\alpha}{(s+a)^2+b^2}$	$\dfrac{\sqrt{(\alpha-a)^2+b^2}}{b}e^{-at}\sin(bt+\phi) \quad \phi=\tan^{-1}\dfrac{b}{\alpha-a}$
27.	$\dfrac{1}{s[(s+a)^2+b^2]}$	$\dfrac{1}{a^2+b^2}+\dfrac{1}{b\sqrt{a^2+b^2}}e^{-at}\sin(bt-\phi)$ $\phi=\tan^{-1}\dfrac{b}{-a}$
27a.	$\dfrac{1}{s(s^2+2\zeta\omega_n s+\omega_n^2)}$	$\dfrac{1}{\omega_n^2}-\dfrac{1}{\omega_n^2\sqrt{1-\zeta^2}}e^{-\zeta\omega_n t}\sin(\omega_n\sqrt{1-\zeta^2}\,t+\phi)$ $\phi=\cos^{-1}\zeta$
28.	$\dfrac{s+\alpha}{s[(s+a)^2+b^2]}$	$\dfrac{\alpha}{a^2+b^2}+\dfrac{1}{b}\sqrt{\dfrac{(\alpha-a)^2+b^2}{a^2+b^2}}e^{-at}\sin(bt+\phi)$ $\phi=\tan^{-1}\dfrac{b}{\alpha-a}-\tan^{-1}\dfrac{b}{-a}$
29.	$\dfrac{1}{(s+c)[(s+a)^2+b^2]}$	$\dfrac{e^{-ct}}{(c-a)^2+b^2}+\dfrac{e^{-at}\sin(bt-\phi)}{b\sqrt{(c-a)^2+b^2}}$ $\phi=\tan^{-1}\dfrac{b}{c-a}$
30.	$\dfrac{1}{s(s+c)[(s+a)^2+b^2]}$	$\dfrac{1}{c(a^2+b^2)}-\dfrac{e^{-ct}}{c[(c-a)^2+b^2]}$ $+\dfrac{e^{-at}\sin(bt-\phi)}{b\sqrt{a^2+b^2}\sqrt{(c-a)^2+b^2}}$ $\phi=\tan^{-1}\dfrac{b}{-a}+\tan^{-1}\dfrac{b}{c-a}$
31.	$\dfrac{s+\alpha}{s(s+c)[(s+a)^2+b^2]}$	$\dfrac{\alpha}{c(a^2+b^2)}+\dfrac{(c-\alpha)e^{-ct}}{c[(c-a)^2+b^2]}$ $+\dfrac{\sqrt{(\alpha-a)^2+b^2}}{b\sqrt{a^2+b^2}\sqrt{(c-a)^2+b^2}}e^{-at}\sin(bt+\phi)$ $\phi=\tan^{-1}\dfrac{b}{\alpha-a}-\tan^{-1}\dfrac{b}{-a}-\tan^{-1}\dfrac{b}{c-a}$
32.	$\dfrac{1}{s^2(s+a)}$	$\dfrac{1}{a^2}(at-1+e^{-at})$

(*continued*)

$F(s)$	$f(t) \quad 0 \le t$
33. $\dfrac{1}{s(s+a)^2}$	$\dfrac{1}{a^2}(1 - e^{-at} - ate^{-at})$
34. $\dfrac{s+\alpha}{s(s+a)^2}$	$\dfrac{1}{a^2}[\alpha - \alpha e^{-at} + a(a-\alpha)te^{-at}]$
35. $\dfrac{s^2+\alpha_1 s+\alpha_0}{s(s+a)(s+b)}$	$\dfrac{\alpha_0}{ab} + \dfrac{a^2 - \alpha_1 a + \alpha_0}{a(a-b)}e^{-at} - \dfrac{b^2 - \alpha_1 b + \alpha_0}{b(a-b)}e^{-bt}$
36. $\dfrac{s^2+\alpha_1 s+\alpha_0}{s[(s+a)^2+b^2]}$	$\dfrac{\alpha_0}{c^2} + \dfrac{1}{bc}[(a^2 - b^2 - \alpha_1 a + \alpha_0)^2 + b^2(\alpha_1 - 2a)^2]^{1/2} e^{-at} \sin(bt+\phi)$ $\phi = \tan^{-1}\dfrac{b(\alpha_1 - 2a)}{a^2 - b^2 - \alpha_1 a + \alpha_0} - \tan^{-1}\dfrac{b}{-a}$ $c^2 = a^2 + b^2$
37. $\dfrac{1}{(s^2+\omega_2)[(s+a)^2+b^2]}$	$\dfrac{(1/\omega)\sin(\omega t + \phi_1) + (1/b)e^{-at}\sin(bt+\phi_2)}{[4a^2\omega^2 + (a^2+b^2-\omega^2)^2]^{1/2}}$ $\phi_1 = \tan^{-1}\dfrac{-2a\omega}{a^2+b^2-\omega^2} \quad \phi_2 = \tan^{-1}\dfrac{2ab}{a^2-b^2+\omega^2}$
38. $\dfrac{s+\alpha}{(s^2+\omega^2)[(s+a)^2+b^2]}$	$\dfrac{1}{\omega}\left(\dfrac{\alpha^2+\omega^2}{c}\right)^{1/2}\sin(\omega t+\phi_1) + \dfrac{1}{b}\left[\dfrac{(\alpha-a)^2+b^2}{c}\right]^{1/2} e^{-at}\sin(bt+\phi_2)$ $c = (2a\omega)^2 + (a^2+b^2-\omega^2)^2$ $\phi_1 = \tan^{-1}\dfrac{\omega}{a} - \tan^{-1}\dfrac{2a\omega}{a^2+b^2+\omega^2}$ $\phi_2 = \tan^{-1}\dfrac{b}{\alpha-a} + \tan^{-1}\dfrac{2ab}{a^2-b^2+\omega^2}$
39. $\dfrac{s+\alpha}{s^2[(s+a)^2+b^2]}$	$\dfrac{1}{c}\left(\alpha t + 1 - \dfrac{2\alpha a}{c}\right) + \dfrac{[b^2+(\alpha-a)^2]^{1/2}}{bc}e^{-at}\sin(bt+\phi)$ $c = a^2+b^2$ $\phi = 2\tan^{-1}\left(\dfrac{b}{a}\right) + \tan^{-1}\dfrac{b}{\alpha-a}$
40. $\dfrac{s^2+\alpha_1 s+\alpha_0}{s^2(s+a)(s+b)}$	$\dfrac{\alpha_1+\alpha_0 t}{ab} - \dfrac{\alpha_0(a+b)}{(ab)^2} - \dfrac{1}{a-b}\left(1 - \dfrac{\alpha_1}{a} + \dfrac{\alpha_0}{a^2}\right)e^{-at} - \dfrac{1}{1-b}\left(1 - \dfrac{\alpha_1}{b} + \dfrac{\alpha_0}{b^2}\right)e^{-bt}$

Appendix B

Matrix Linear Algebra

B.1 INTRODUCTION

Some basic matrix concepts needed in the development of the state method of describing and analyzing physical systems are presented in this appendix [1–3].

B.2 MATRIX

A *matrix* is a rectangular array of elements. The elements may be real or complex numbers or variables of time or frequency. A matrix with α rows and β columns is called an $\alpha \times \beta$ matrix or is said to be of *order* $\alpha \times \beta$. The matrix is also said to have size or dimension $\alpha \times \beta$. If $\alpha = \beta$, the matrix is called a *square matrix*. Boldface capital letters are used to denote rectangular matrices, and boldface small letters are used to denote column matrices. A general expression for an $\alpha \times \beta$ matrix is

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1\beta} \\ m_{21} & m_{22} & \cdots & m_{2\beta} \\ \vdots & \vdots & \ddots & \vdots \\ m_{\alpha 1} & m_{\alpha 2} & \cdots & m_{\alpha\beta} \end{bmatrix} = \begin{bmatrix} \\ m_{ij} \\ \\ \end{bmatrix} \tag{B.1}$$

The elements are denoted by lowercase letters. A double-subscript notation is used to denote the location of the element in the matrix; thus, the element m_{ij} is located in the ith row and the jth column.

B.3 TRANSPOSE

The *transpose* of a matrix $\mathbf{M}$ is denoted by $\mathbf{M}^T$. The matrix $\mathbf{M}^T$ is obtained by interchanging the rows and columns of the matrix $\mathbf{M}$. In general, with $\mathbf{M} = [m_{ij}]$, the transpose matrix is $\mathbf{M}^T = [m_{ji}]$.

B.4 VECTOR

A *vector* is a matrix that has either one row or one column. An $\alpha \times 1$ matrix is called a *column vector,* written

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_\alpha \end{bmatrix} \tag{B.2}$$

and a $1 \times \beta$ matrix is called a *row vector,* written

$$\mathbf{x}^T = [x_1 \quad x_2 \quad \cdots \quad x_\beta] \tag{B.3}$$

Thus, the transpose of a column vector $\mathbf{x}$ yields the row vector $\mathbf{x}^T$.

B.5 ADDITION AND SUBTRACTION OF MATRICES

The sum or difference of two matrices $\mathbf{M}$ and $\mathbf{N}$, both of order $\alpha \times \beta$, is a matrix $\mathbf{W}$ of order $\alpha \times \beta$. The element w_{ij} of $\mathbf{W} = \mathbf{M} \pm \mathbf{N}$ is

$$w_{ij} = m_{ij} \pm n_{ij} \tag{B.4}$$

These operations are commutative and associative; i.e.,

$$\text{Commutative: } \mathbf{M} \pm \mathbf{N} = \pm\mathbf{N} + \mathbf{M} \tag{B.5}$$

$$\text{Associative: } (\mathbf{M} + \mathbf{N}) + \mathbf{Q} = \mathbf{M} + (\mathbf{N} + \mathbf{Q}) \tag{B.6}$$

Example 1. Addition of matrices

$$\mathbf{W} = \begin{bmatrix} 3 & 1 & 2 \\ 1 & 0 & -4 \\ 0 & 5 & 7 \end{bmatrix} + \begin{bmatrix} 1 & -3 & 1 \\ 2 & 1 & 5 \\ -4 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 4 & -2 & 3 \\ 3 & 1 & 1 \\ -4 & 4 & 7 \end{bmatrix} \tag{B.7}$$

B.6 MULTIPLICATION OF MATRICES

The multiplication of two matrices **MN** can be performed *if and only if* (iff) they *conform*. If the orders of **M** and **N** are $\alpha \times \beta$ and $\gamma \times \delta$, respectively, they are conformable iff $\beta = \gamma$, that is, the number of columns of **M** must be equal to the number of rows of **N**. Under this condition **MN** = **W**, where the elements of **W** are defined by

$$w_{ij} = \sum_{k=1}^{\beta} m_{ik} n_{kj} \tag{B.8}$$

That is, each element of the *i*th row of **M** is multiplied by the corresponding element of the *j*th column of **N**, and these products are summed to yield the *ij*th element of **W**. The dimension or order of the resulting matrix **W** is $\alpha \times \delta$.

Matrix multiplication operations are summarized as follows:

1. An $\alpha \times \beta$ **M** matrix times a $\beta \times \gamma$ **N** matrix yields an $\alpha \times \gamma$ **W** matrix; that is, **MN** = **W**.
2. An $\alpha \times \beta$ **M** matrix times a $\beta \times \alpha$ **N** matrix yields an $\alpha \times \alpha$ **W** square matrix; that is, **MN** = **W**. Note that although **NM** = **Y** is also a square matrix, it is of order $\beta \times \beta$.
3. When **M** is of order $\alpha \times \beta$ and **N** is of order $\beta \times \alpha$, each of the products **MN** and **NM** is conformable. However, in general,

$$\mathbf{MN} \neq \mathbf{NM} \tag{B.9}$$

   Thus the product of **M** and **N** is said to be *noncommutable*, that is, the order of multiplication cannot be interchanged.
4. The product **MN** can be referred to as "**N** *premultiplied* by **M**" or as "**M** *postmultiplied* by **N**." In other words, premultiplication or postmultiplication is used to indicate whether one matrix is multiplied by another from the left or from the right.
5. A $1 \times \alpha$ row vector times an $\alpha \times \beta$ matrix yields a $1 \times \beta$ row vector; that is $\mathbf{x}^T\mathbf{M} = \mathbf{y}^T$.
6. A $\beta \times \alpha$ matrix times an $\alpha \times 1$ column vector yields a $\beta \times 1$ column vector; that is, $\mathbf{Mx} = \mathbf{y}$.
7. A $1 \times \alpha$ row vector times an $\alpha \times 1$ column vector yields a 1×1 matrix, that is, $\mathbf{x}^T\mathbf{y} = \mathrm{w}$, which contains a single scalar element.
8. The *k*-fold multiplication of a square matrix **M** by itself is indicted by $\mathbf{M}^k$.

These operations are illustrated by the following examples.

Example 2. Matrix multiplication

$$\begin{bmatrix} 2 & 1 & 3 \\ 1 & 4 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 & 0 \\ 0 & 3 & 1 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 7 & 11 & 4 \\ 3 & 15 & 5 \end{bmatrix} \qquad \begin{bmatrix} 2 & 1 & 3 \\ 1 & 4 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 0 & 3 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 7 & 11 \\ 3 & 15 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 2 & 1 & 3 \\ 1 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 6 & 7 \\ 3 & 12 & 3 \\ 4 & 9 & 5 \end{bmatrix} \qquad \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 4 & 3 \\ 3 & 3 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ 9 \end{bmatrix} \qquad [1 \quad 3]\begin{bmatrix} 2 \\ 1 \end{bmatrix} = [5]$$

B.7 MULTIPLICATION OF A MATRIX BY A SCALAR

The multiplication of matrix **M** *by a scalar k* is effected by multiplying each element m_{ij} by k, that is,

$$k\mathbf{M} = \mathbf{M}k = [km_{ij}] \tag{B.10}$$

Example 3.

$$2\begin{bmatrix} 1 & 0 \\ 5 & -7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 5 & -7 \end{bmatrix}2 = \begin{bmatrix} 2 & 0 \\ 10 & -14 \end{bmatrix}$$

B.8 UNIT OR IDENTITY MATRIX

A *unit matrix*, denoted **I**, is a *diagonal* matrix in which each element on the principal diagonal is unity. Sometimes the notation $\mathbf{I}_n$ is used to indicate an identity matrix of order n. An example of a unit matrix is

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{B.11}$$

The premultiplication or postmultiplication of a matrix **M** by the unit matrix **I** leaves the matrix unchanged.

$$\mathbf{MI} = \mathbf{IM} = \mathbf{M} \tag{B.12}$$

B.9 DIFFERENTIATION OF A MATRIX

The *differentiation* of a matrix with respect to a scalar is effected by differentiating each element of the matrix with respect to the indicated variable:

$$\frac{d}{dt}[\mathbf{M}(t)] = \dot{\mathbf{M}}(t) = \left[\dot{m}_{ij}\right] = \begin{bmatrix} \dot{m}_{11} & \dot{m}_{12} & \cdots & \dot{m}_{1\beta} \\ \dot{m}_{21} & \dot{m}_{22} & \cdots & \dot{m}_{2\beta} \\ \vdots & \vdots & \ddots & \vdots \\ \dot{m}_{\alpha 1} & \dot{m}_{\alpha 2} & \cdots & \dot{m}_{\alpha\beta} \end{bmatrix} \tag{B.13}$$

The derivative of a product of matrices follows rules similar to those for the derivative of a scalar product, with preservation of order. Thus, typically,

$$\frac{d}{dt}[\mathbf{M}(t)\mathbf{N}(t)] = \mathbf{M}(t)\dot{\mathbf{N}}(t) + \dot{\mathbf{M}}(t)\mathbf{N}(t) \tag{B.14}$$

B.10 INTEGRATION OF A MATRIX

The *integration* of a matrix is effected by integrating each element of the matrix with respect to the indicated variable:

$$\int \mathbf{M}(t)\,dt = \left[\int m_{ij}\,dt\right] \tag{B.15}$$

B.11 ADDITIONAL MATRIX OPERATIONS AND PROPERTIES

Additional characteristics of matrices required to solve matrix state equations by using the Laplace transform are presented in this section.

PRINCIPAL DIAGONAL. The *principal diagonal* of a square matrix $\mathbf{M} = [m_{ij}]$ consists of the elements m_{ii}.

DIAGONAL MATRIX. A *diagonal matrix* is a square matrix in which all the elements off the principal diagonal are zero. This fact can be expressed as $m_{ij} = 0$ for $i \neq j$ and the principal diagonal elements $m_{ii} = m_i$ are not all zero. When the diagonal elements are all equal, the matrix is called a *scalar* matrix.

TRACE. The *trace* of a square matrix $\mathbf{M}$ of order n is the sum of all the elements along the principal diagonal; that is,

$$\text{trace } \mathbf{M} = m_{11} + m_{22} + \cdots + m_{nn} = \sum_{i=1}^{n} m_{ii} \tag{B.16}$$

For the state equation $\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu}$,

$$\text{trace } A = \sum \sigma(A) = \sum_{i=1}^{n} \lambda_i \tag{B.17}$$

where $\sigma(\mathbf{A})$ is the spectrum of **A**.

DETERMINANT. The *determinant* of a square matrix **M** of order n is the sum of all possible signed products of n elements containing one and only one element from every row and column in the matrix. The determinant of a matrix can be represented by replacing the brackets by vertical bars. The determinant of **M** may be denoted by $|\mathbf{M}|$, or det **M**, or $\Delta_{\mathbf{M}}$. It is assumed that the reader knows how to evaluate determinants.

Example 4.

$$\mathbf{M} = \begin{bmatrix} 3 & 1 & 2 \\ 1 & 0 & -4 \\ 0 & 5 & 7 \end{bmatrix} \qquad |\mathbf{M}| = \begin{vmatrix} 3 & 1 & 2 \\ 1 & 0 & -4 \\ 0 & 5 & 7 \end{vmatrix} = 63 \tag{B.18}$$

Characteristics of determinants are as follows:

1. The determinant of a unit matrix is $|\mathbf{I}| = 1$.
2. The determinant of a matrix is zero if (*a*) any row or column contains all zeros or (*b*) the elements of any two rows (or columns) have a common ratio.

SINGULAR MATRIX. A square matrix is said to be *singular* if the value of its determinant is zero. If the value of its determinant is not zero, the matrix is *nonsingular.*

MINOR. The *minor* M_{ij} of a square matrix **M** of order n is the determinant formed after the ith row and jth column are deleted from **M**.

PRINCIPAL MINOR. A *principal* minor is a minor M_{ii} whose diagonal elements are also the diagonal elements of the square matrix **M**.

COFACTOR. A *cofactor* is a *signed* minor and is given by

$$C_{ij} = \Delta_{ij} = (-1)^{i+j} M_{ij} \tag{B.19}$$

Example 5. For the matrix **M** of Example 4, the cofactor C_{21} is obtained by deleting the second row and first column, giving

$$C_{21} = (-1)^{2+1} \begin{vmatrix} 1 & 2 \\ 5 & 7 \end{vmatrix} = +3 \tag{B.20}$$

B.12 GENERALIZED DETERMINANT

The determinant of a square matrix **M** of order n can also be computed from the sum of terms obtained by an expansion along any row i or along any column j as follows:

1. The cofactors C_{ij} are formed for each element m_{ij} of any row of the matrix **M**. Then

$$|\mathbf{M}| = \sum_{j=1}^{n} m_{ij} C_{ij} \qquad \text{for any row } i \tag{B.21}$$

2. The cofactors C_{ij} are formed for each element m_{ij} of any column j of the matrix **M**. Then

$$|\mathbf{M}| = \sum_{i=1}^{n} m_{ij} C_{ij} \qquad \text{for any column } j \tag{B.22}$$

ADJOINT MATRIX. The adjoint of square matrix **M**, denoted as adj **M**, is the transpose of the cofactor matrix. The cofactor matrix is formed by replacing each element of **M** by its cofactor.

$$\text{Adj}\,\mathbf{M} = \begin{bmatrix} C_{ji} \end{bmatrix} = \begin{bmatrix} C_{ij} \end{bmatrix}^T = \begin{bmatrix} \text{array of} \\ \text{cofactor} \end{bmatrix}^T \tag{B.23}$$

Example 6. The adj **M** for the matrix **M** of Example 4 is

$$\text{Adj}\,\mathbf{M} = \begin{bmatrix} 20 & -7 & 8 \\ 3 & 21 & -15 \\ -4 & 14 & -1 \end{bmatrix}^T = \begin{bmatrix} 20 & 3 & -4 \\ -7 & 21 & 14 \\ 5 & -15 & -1 \end{bmatrix} \tag{B.24}$$

INVERSE MATRIX. The product of a matrix and its adjoint is a scalar matrix, as illustrated by the following example.

Example 7. The product **M** adj **M**, using values in Examples 4 and 6, is

$$\mathbf{M}\,\text{adj}\,\mathbf{M} = \begin{bmatrix} 3 & 1 & 2 \\ 1 & 0 & -4 \\ 0 & 5 & 7 \end{bmatrix} \begin{bmatrix} 20 & 3 & -4 \\ -7 & 21 & 14 \\ 5 & -15 & -1 \end{bmatrix} = \begin{bmatrix} 63 & 0 & 0 \\ 0 & 63 & 0 \\ 0 & 0 & 63 \end{bmatrix} = 63\mathbf{I} = |\mathbf{M}|\mathbf{I} \tag{B.25}$$

The inverse of a square matrix **M** is denoted by $\mathbf{M}^{-1}$ and has the property

$$\mathbf{M}\mathbf{M}^{-1} = \mathbf{M}^{-1}\mathbf{M} = \mathbf{I} \tag{B.26}$$

The inverse matrix $\mathbf{M}^{-1}$ is defined from the result of Example 7 as

$$\mathbf{M}^{-1} = \frac{\operatorname{adj} \mathbf{M}}{|\mathbf{M}|} \tag{B.27}$$

The inverse exists only if $|\mathbf{M}| \neq 0$; that is, the matrix $\mathbf{M}$ is nonsingular.

INVERSE OF A PRODUCT OF MATRICES. The inverse of a product of matrices is equal to the product of the inverses of the individual matrices in reverse order:

$$[\mathbf{A} \quad \mathbf{B}]^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{B.28}$$

PRODUCT OF DETERMINANTS. The determinant of the product of square matrices is equal to the product of the individual determinants:

$$|\mathbf{ABC}| = |\mathbf{A}| \cdot |\mathbf{B}| \cdot |\mathbf{C}| \tag{B.29}$$

RANK OF A MATRIX. The rank r of a matrix $\mathbf{M}$, not necessarily square, is the order of the largest square array contained in $\mathbf{M}$ that has a nonzero determinant.

Example 8.

$$\mathbf{M} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 5 & 7 \end{bmatrix} \tag{B.30}$$

The determinant $|\mathbf{M}| = 0$; that is, $\mathbf{M}$ is a singular matrix. The square array obtained by deleting the first row and second column has a nonzero determinant:

$$\begin{vmatrix} 2 & 4 \\ 3 & 7 \end{vmatrix} = 2 \tag{B.31}$$

Therefore, $\mathbf{M}$ has a rank of 2.

DEGENERACY (OR NULLITY) OF A MATRIX. When a matrix $\mathbf{M}$ of order n has rank r, there are $q = n - r$ rows or columns which are linear combinations of the r rows or columns. The matrix $\mathbf{M}$ is then said to have *degeneracy q*. In Example 8 the matrix $\mathbf{M}$ is of order $n = 3$, the rank is $r = 2$, and the degeneracy is $q = 1$ (simple degeneracy). Note that the third row of $\mathbf{M}$ is the sum of the first two rows. Also, the third column is two times the second column minus the first column.

SYMMETRIC MATRIX. A square matrix containing only real elements is symmetric if it is equal to its transpose, that is, $\mathbf{A} = \mathbf{A}^T$. The relationship between the elements is

$$a_{ij} = a_{ji}$$

A symmetric matrix is symmetrical about its principal diagonal.

TRANSPOSE OF A PRODUCT OF MATRICES. The transpose of a product of matrices is the product of the transposed matrices in reverse order:

$$[\mathbf{AB}]^T = \mathbf{B}^T\mathbf{A}^T \tag{B.32}$$

This relationship can be described by the statement that the product of transposed matrices is equal to the transpose of the product in reverse order of the original matrices.

B.13 HERMITE NORMAL FORM

A number of properties of a matrix can be determined by transforming the matrix to *Hermite normal form* (HNF). These properties include the determination of the rank of a matrix, the inverse of a matrix of full rank, the characteristic polynomial of a matrix, the eigenvectors associated with the eigenvalue, the controllability of a system, and the observability of a system. The transformation is accomplished on a matrix **M** by performing the following basic or *elementary* operations:

1. Interchanging the ith and jth rows
2. Multiplying the ith row by a nonzero scalar k
3. Adding to the elements of the ith row the corresponding elements of the jth row multiplied by a scalar k

The purpose of these operations is to produce a unit element as the first nonzero element in each row (from the left), with the remaining elements of the column all zero. These operations can be accomplished by premultiplying by *elementary matrices*, where the elementary matrices are formed by performing the desired basic operations on a unit matrix. For example, for a third-order matrix **M**, premultiplying by the following elementary matrices $\mathbf{E}_i$ produces the results indicated:

$$\underset{\text{Interchanging rows 1 and 2}}{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}} \quad \underset{\text{Multiplying row 3 by } k}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & k \end{bmatrix}} \quad \underset{\text{Adding } k \text{ times row 3 to row 2}}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & k \\ 0 & 0 & 1 \end{bmatrix}} \tag{B.33}$$

Example 9. The matrix of Example 8 is reduced to HNF by the following operations:

1. Subtract 2 times row 1 from row 2 and 3 times row 1 from row 3. Premultiplying by the corresponding elementary matrix $\mathbf{E}_1$ produces

$$\mathbf{E}_1\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 5 & 7 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & -1 & -2 \end{bmatrix} = \mathbf{M}_1 \tag{B.34}$$

2. Add 2 times row 2 to row 1, subtract row 2 from row 3, and multiply row 2 by -1. Premultiplying by the corresponding elementary matrix $\mathbf{E}_2$ produces

$$\mathbf{E}_2\mathbf{M}_1 = \begin{bmatrix} 1 & 2 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & -1 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} = \mathbf{M}_2 \tag{B.35}$$

The matrix $\mathbf{M}_2$ is in Hermite normal form and has the same rank as the matrix $\mathbf{M}$. Since $\mathbf{M}_2$ has one zero row, it has a rank of 2, which agrees with the result of Example 8.

B.14 MATRIX INVERSION BY ROW OPERATIONS

The inverse $\mathbf{M}^{-1}$ of a nonsingular matrix $\mathbf{M}$ can be obtained as follows:

1. Form the augmented matrix $[\mathbf{M} \,\vdots\, \mathbf{I}]$.
2. Perform row operations on the augmented matrix so that the left portion $\mathbf{M}$ becomes the identity matrix $\mathbf{I}$. When the left portion is so transformed, the right portion $\mathbf{I}$ is transformed into the matrix inverse $\mathbf{M}^{-1}$. Thus, the transformed augmented matrix becomes $[\mathbf{I} \,\vdots\, \mathbf{M}^{-1}]$.

Example 10. For the matrix of Example 4, the augmented matrix is

$$[\mathbf{M} \,\vdots\, \mathbf{I}] = \left[\begin{array}{ccc:ccc} 3 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & -4 & 0 & 1 & 0 \\ 0 & 5 & 7 & 0 & 0 & 1 \end{array}\right] \tag{B.36}$$

A suitable set of basic row operations includes subtracting 3 times row 2 from row 1 and then interchanging rows 1 and 2, yielding

$$\left[\begin{array}{ccc:ccc} 1 & 0 & -4 & 0 & 1 & 0 \\ 0 & 1 & 14 & 1 & -3 & 0 \\ 0 & 5 & 7 & 0 & 0 & 1 \end{array}\right]$$

Subtracting 5 times row 2 from row 3 yields

$$\left[\begin{array}{ccc:ccc} 1 & 0 & -4 & 0 & 1 & 0 \\ 0 & 1 & 14 & 1 & -3 & 0 \\ 0 & 0 & -63 & -5 & 15 & 1 \end{array}\right]$$

Dividing row 3 by -63, adding 4 times the new row 3 to row 1, and subtracting 14 times the new row 3 from row 2 yields

$$\left[\begin{array}{ccc:ccc} 1 & 0 & 0 & 20/63 & 3/63 & -4/63 \\ 0 & 1 & 0 & -7/63 & 21/63 & 14/63 \\ 0 & 0 & 1 & 5/63 & -15/63 & -1/63 \end{array}\right] = [\mathbf{I} \,\vdots\, \mathbf{M}^{-1}] \tag{B.37}$$

The inverse matrix $\mathbf{M}^{-1}$ agrees with the value obtained from Eq. (B.27) and the results of Examples 4 and 6.

B.15 EVALUATION OF THE CHARACTERISTIC POLYNOMIAL

The characteristic polynomial for the matrix A appearing in the state equation

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}\mathbf{u} \tag{B.38}$$

is

$$Q(\lambda) = |\lambda\mathbf{I} - \mathbf{A}| = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 \tag{B.39}$$

When the $n \times (n+1)$ matrix

$$\mathbf{M}_c = \begin{bmatrix} \mathbf{b} & \mathbf{A}\mathbf{b} & \mathbf{A}^2\mathbf{b} & \cdots & \mathbf{A}^n\mathbf{b} \end{bmatrix} \tag{B.40}$$

is reduced to HNF, which must be full rank, the coefficients of the characteristic polynomial appear in the last column, [4] i.e.,

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & -a_0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & -a_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & -a_{n-1} \end{bmatrix} \tag{B.41}$$

This procedure may be simpler than forming the determinant $Q(\lambda) = |\lambda\mathbf{I} - \mathbf{A}|$.

Example 11.

$$\mathbf{A} = \begin{bmatrix} 1 & 6 & -3 \\ -1 & -1 & 1 \\ -2 & 2 & 0 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathbf{M}_c = \begin{bmatrix} 1 & 4 & -2 & 10 \\ 1 & -1 & -3 & -5 \\ 1 & 0 & -10 & -2 \end{bmatrix} \tag{B.42}$$

Reducing $\mathbf{M}_c$ to HNF yields

$$\begin{array}{l} \begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{l} \leftarrow -a_0 \\ \leftarrow -a_1 \\ \leftarrow -a_2 \end{array} \tag{B.43}$$

The characteristic polynomial is therefore

$$Q(\lambda) = \lambda^3 + a_2\lambda^2 + a_1\lambda + a_0 = \lambda^3 - 3\lambda + 2 \tag{B.44}$$

B.16 LINEAR INDEPENDENCE

A set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k, \ldots, \mathbf{v}_n$ is linearly independent provided there is no set of scalars $\alpha_i (i = 1, 2, \ldots, n)$, not all zero, that satisfies

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k + \cdots + \alpha_n\mathbf{v}_n = 0 \tag{B.45}$$

If this equation is satisfied with the α_i not all zero, then the vectors are *linearly dependent*. Linear independence can be checked by forming the matrix

$$\begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} \tag{B.46}$$

If this matrix has rank n, the vectors are linearly independent; otherwise they are linearly dependent. The rank of this matrix can be determined by reducing it to Hermite normal form.

REFERENCES

1. Gantmacher, F.R.: *Applications of the Theory of Matrices*, Wiley-Interscience, New York, 1959.

2. Strang, G.: *Linear Algebra and Its Applications*, 3rd ed., Harcourt, Brace, Jovanovich, San Diego, CA, 1988.
3. Webster, J.G., ed., *Encyclopedia of Electrical and Electronics Engineering*, John Wiley & Sons, New York, 1999.
4. Reid, J.G.: Unpublished notes, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, 1979.

Appendix C

Introduction to MATLAB and Simulink

C.1 INTRODUCTION

MATLAB® is a technical computing environment (program) designed for numerical computation and visualization. Over the years it has become a standard tool in many universities and research organizations for performing mathematical calculations. Originally written to provide access to matrix manipulation software, MATLAB has developed into a powerful tool to simplify mathematical analysis.

In order to use MATLAB successfully, the student should understand how to enter and manipulate variables, how to model control systems in MATLAB, and how to produce plots and graphs depicting the performance of control systems. This appendix is intended to guide the student through the MATLAB functions which are useful in understanding and accomplishing these tasks. Students are encouraged to use the documentation provided with MATLAB to explore more advanced problems.

The commands shown here are valid for MATLAB version 6.1, Release 12.1, for Windows 98. There may be variations in other versions, or releases for other platforms. Many of the commands shown here are in the Control Systems Toolbox, which is available as an add-on to the basic MATLAB program.

C.2 BASICS

MATLAB is a window-based workspace environment in which variables are entered and manipulated. The MATLAB program is started by double clicking the MATLAB icon in the Windows desktop. This opens a window similar to that shown in Fig. C.1. Commands are entered in the command window, which is the inner window as shown on the right side of Fig. C.1. Two right arrows (>>) is the command prompt, indicating that the program is ready for the next command. Variables are created as they are entered, and are stored in the workspace. Variable names must begin with a letter. A variable to represent $\alpha = 1.7$ may be entered as follows:

```
alpha = 1.7;
```

Placing the variable name "`alpha`" to the left of the equal sign defines alpha as whatever is on the right side of the equal sign.

MATLAB displays the variable upon entry unless a semicolon is placed at the end of the line, that is,

```
beta = alpha*3.7  %Optional comment
beta =
       6.2900
```

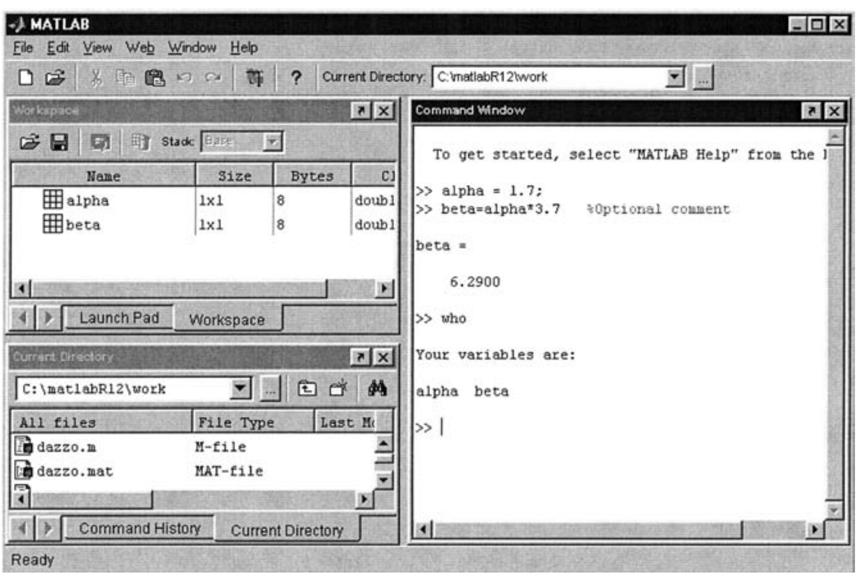


FIGURE C.1 MATLAB window.

A percent sign provides a mean to include a comment. Everything to the right of a percent sign is ignored by MATLAB.

In MATLAB version 6, a workspace window lists the variables that are defined. The "who" command also allows the user to see what variables are defined in the workspace.

```
Who
Your variables are:
alpha beta
```

C.2.1 Matrices

Matrices are entered a row at a time as follows:

```
A = [11 12 13
21 22 23
31 32 33]
```

The carriage return is implied at the end of each line and is not explicitly shown in this text. MATLAB displays the variables upon entry unless a semicolon is placed at the end of the line. Semicolons are also used to simplify and conserve space. The same 3×3 matrix may be entered as follows:

```
A = [11 12 13; 21 22 23; 31 32 33];  %  Optional Comment
```

Matrices are indexed in (row, column) format. For example, a single element of a matrix can be referenced as follows:

```
A(2,3)
```

This gives the result

```
ans =
  23
```

C.2.2 Matrix Transpose

Matrices are transposed using a prime notation.

```
B = A'
```

produces

```
B =

     11  21  31
     12  22  32
     13  23  33
```

C.2.3 Range of Values

A colon is used to denote a range of values, e.g., `A(1:2,:)` is a matrix consisting of the first two rows of `A`. A colon notation also can be used to generate a vector of equally spaced elements. The format for defining a vector x whose elements range from an initial value to a final value is

```
x=[0:2*pi];  %[Initial value:Final valueaf]
```

The step size is assumed to be 1 unless an optional step size is specified between the initial and final values. A vector, used to plot data, may be defined as follows:

```
x=[0: .25:2*pi]  %[Initial value:Step size:Final value]
```

```
MATLAB generates the vector:

x=
  Columns 1 through 7

    0   0.2500   0.5000   0.7500   1.0000   1.2500   1.5000

  Columns 8 through 14

    1.7500   2.0000   2.2500   2.5000   2.7500   3.0000   3.2500

  Columns 15 through 21

    3.5000   3.7500   4.0000   4.2500   4.5000   4.7500   5.0000

  Columns 22 through 26

    5.2500   5.5000   5.7500   6.0000   6.2500
```

Numerous common mathematical functions are available, such as

```
y=sin(x)
```

which generate a vector corresponding to the sine of each element of the vector x. Some other trigonometric functions available in MATLAB are listed in Table C.1.

The data are plotted with the command

```
plot (x, y)
```

This command opens a plot window as shown in Fig. C.2. The window has menu buttons across the top which allows the user to rescale, annotate, edit, save, and print the plot.

Table C.1 MATLAB Trigonometric Functions

sin	Sine.	sec	Secant.
sinh	Hyperbolic sine.	sech	Hyperbolic secant.
asin	Inverse sine.	asec	Inverse secant.
asinh	Inverse hyperbolic sine.	asech	Inverse hyperbolic secant.
cos	Cosine.	csc	Cosecant.
cosh	Hyperbolic cosine.	csch	Hyperbolic cosecant.
acos	Inverse cosine.	acsc	Inverse cosecant.
acosh	Inverse hyperbolic cosine.	acsch	Inverse hyperbolic cosecant.
tan	Tangent.	cot	Cotangent.
tanh	Hyperbolic tangent.	coth	Hyperbolic cotangent.
atan	Inverse tangent.	acot	Inverse cotangent.
atan2	Four quadrant inverse tangent.	acoth	Inverse hyperbolic cotangent.
atanh	Inverse hyperbolic tangent.		

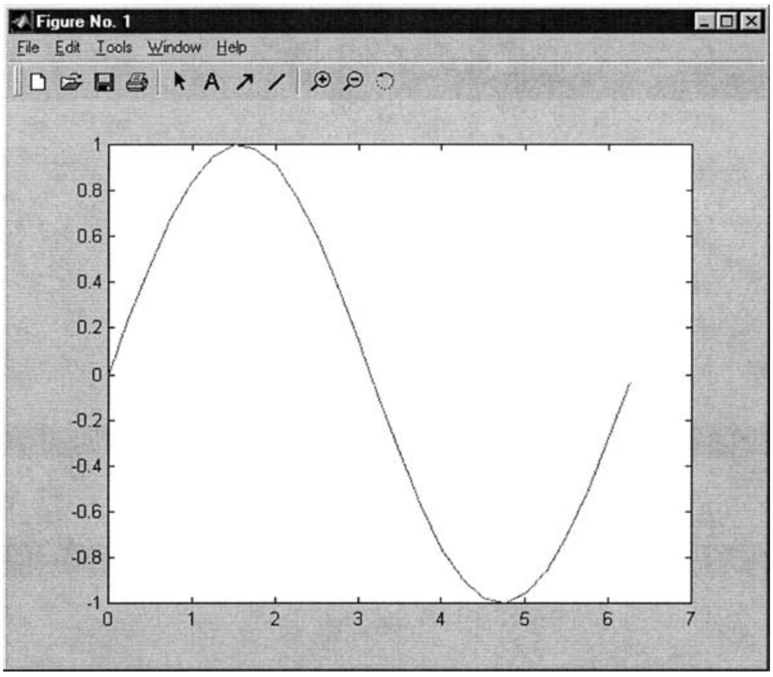


FIGURE C.2 MATLAB plot window.

Multiple traces can be shown on one plot as illustrated by the following commands:

```
z=cos(x);
subplot (2, 1, 1), plot (x, y)   %  Plot 2 rows 1 column and
                                 %  select the first subplot
title ('Multiple Plots')         %  Set title above first plot
grid                             %  Turn grid on in first plot
ylabel ('sin(x)')                %  Set Y axis label in  first plot
subplot (2, 1, 2), plot (x, z)   %  Select second subplot
grid                             %  Turn grid on in second plot
xlabel ('x')                     %  Set X axis label in second
                                 %  plot
ylabel ('cos (x)')               %  Set Y axis label in
                                 %  second plot
```

These commands generate the plot shown in Fig. C.3.

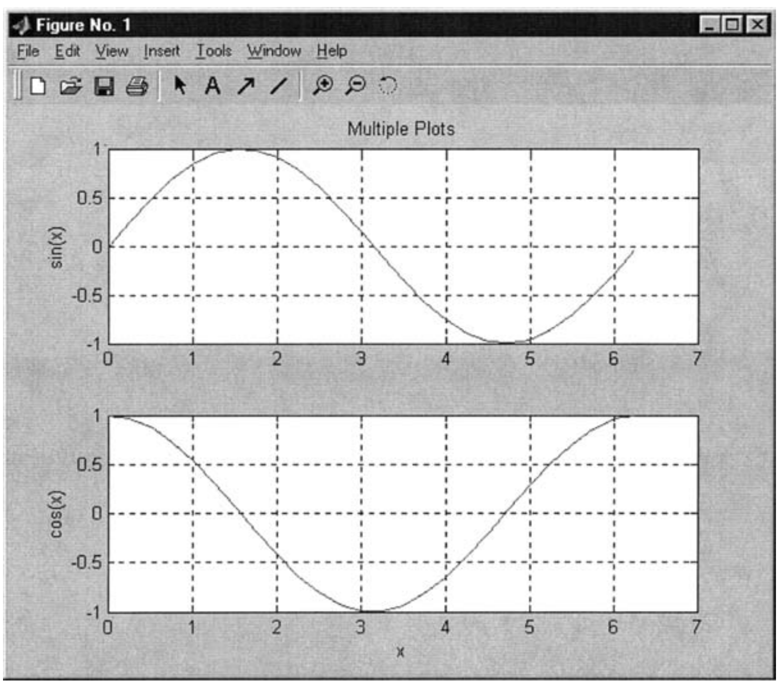


FIGURE C.3 Multiple plots in one window: sine plot (top figure); cosine plot (bottom figure).

Note that the titles and labels can be set from the command line or within the plot window using the buttons across the top of the window. A number of line styles and colors may be specified as described by the command:

```
help plot options
```

Typing "`help`" at the prompt will provide an outline of the help available in the command window. More extensive help features are accessed through the help menu button near the top of the MATLAB window.

C.2.4 m-Files and MAT-Files

MATLAB allows the user to define scripts containing several commands in m-Files. m-Files are stored in the current directory as "filename.m" and are executed from the command line by typing the file name without the .m extension. m-Files may be created and edited from the MATLAB window by using the "New m-File" button on the toolbar.

Workspace variables also may be saved for latter use by typing "`save filename`". This creates a MAT-File in the current directory named "filename.mat". The variables are loaded back into the workspace by typing "`load filename`".

C.3 DEFINING SYSTEMS

Polynomials are represented by a vector of coefficients. For example, the polynomial $s^3 + 2s^2 + 5$ is represented by the vector [1 2 0 5]. All coefficients including the zero must be included in the vector. Polynomials are multiplied by convolving the representative vectors; hence the transfer function

$$\text{poly} = s(s+3)(s+10) = s^3 + 13s^2 + 30s \tag{C.1}$$

is formed as follows:

```
poly = conv([1 0], conv([1 3], [1  10]))

poly =

     1  13  30  0
```

The roots of the polynomial are calculated using the "`roots`" command.

```
roots(poly)
ans =
       0
    - 10
      - 3
```

C.3.1 Transfer Function Model

A transfer function is a ratio of polynomials in s; hence the transfer function

$$\frac{\text{num}}{\text{den}} = \frac{s+5}{s(s+3)(s+10)} = \frac{s+5}{s^3+13s^2+30s} \tag{C.2}$$

is formed as follows:

```
num = [1  5];
den = conv([1  0], conv([1  3], [1  10]))
den =
    1   13   30   0
```

The "TF" command is used to generate a linear time-variant (LTI) transfer function system model object, "sys" from the numerator and denominator polynomials. This transfer function system model object is subsequently used for analysis as shown in Sec. C.4.

```
sys = tf(num, den)

Transfer function:

      s + 5
- - - - - - - - - - - -
s^3 + 13 s^2 + 30 s
```

C.3.2 State Space Model

State space models are entered as matrices. In Sec. 2.13, a DC servomotor is modeled by Eq. (2.135). Augmenting the system with $\dot{\theta}_\mu = \omega$ yields the third-order state space equation:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} = \begin{bmatrix} \dot{\theta}_\mu \\ \dot{\omega}_m \\ \dot{i}_m \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & \dfrac{-B_m}{J_m} & \dfrac{K_t}{J_m} \\ 0 & \dfrac{-K_b}{L_m} & \dfrac{-R_m}{L_m} \end{bmatrix} \begin{bmatrix} \theta_\mu \\ \omega_m \\ i_m \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \dfrac{1}{L_m} \end{bmatrix} e_a \tag{C.3}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u} = \theta_\mu = [1 \quad 0 \quad 0] \begin{bmatrix} \theta_\mu \\ \omega_m \\ i_m \end{bmatrix} + [0]e_a \tag{C.4}$$

A very small 12 volt DC servomotor has the characteristics provided below, which are converted to SI units to develop the **A** and **B** matrices. The units of the variables are: θ is in radians, ω is in rad/s, i is in amperes, and e is in volts.

$$B_m = 0.013 \text{ oz.-in.} \qquad K_t = 1.223 \frac{\text{oz-in.}}{\text{A}}$$

$$L_m = 0.75\,\text{mH} \qquad K_b = 0.905 \frac{\text{mV}}{\text{rpm}}$$

$$J = 0.085 \times 10^{-4} \text{ oz.-in.-s}^2 \qquad R_m = 24\,\Omega$$

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -1530 & 144000 \\ 0 & -11.5 & -32000 \end{bmatrix} \tag{C.5}$$

$$\mathbf{B} = \begin{bmatrix} 0 \\ 0 \\ 1330 \end{bmatrix} \tag{C.6}$$

The following commands are used to enter the state space model, to generate a linear time-invariant system model object, "`sys2`." This state space model object may be used for analysis like the transfer function system model object. These commands have been saved in an m-file named "dcservo.m" which is on the accompanying CDROM and can be executed by typing "`dcservo`" at the command prompt.

```
A = [0    1       0        % Define the plant matrix
    0  -1530   144000
    0  -11.5  -32000];

B = [0;  0;  1330];        % Define the input matrix
C = [1  0  0];             % Define the output matrix
D = [0];                   % Define the direct
                           % feedthrough matrix
sys2 = ss(A,B,C,D)         % Generates a state space
                           % model object
```

which produces the state and output equations:

```
a=

          x1       x2          x3
    x1     0        1           0
    x2     0    −1530    1.44e+005
    x3     0    −11.5      −32000
```

```
b=
            u1
    x1       0
    x2       0
    x3    1330
c=
          x1    x2    x3
    y1     1     0     0
d=
          u1
    y1     0
continuous-time model.
```

C.4 ANALYSIS

C.4.1 Root Locus

The LTI system model object is used for analysis. The root locus for a unity feedback system is presented in Chap. 7. The root locus is plotted in MATLAB using the "`rlocus`" command.

```
rlocus(sys)                  % Plots a root locus for sys
```

or

```
rlocus(sys, K)               % Plots a root locus with
                             % user-specified vector
                             % of gains K
```

A more advanced interface for root locus analysis and design is the "SISOTOOL" interface. The "SISOTOOL" is a graphical user interface that facilitates compensator design and is started with the command:

```
sisotool('rlocus',sys)       % Startroot locus design tool
```

This command opens a new window, plots the root locus as shown in Fig. C.4, and allows the user to specify a gain or add compensator components by placing poles and zeros on the plot. Design constraints such as desired damping ratio or peak overshoot can be set within the tool and appear as bounds on the plot.

A root-locus plot for a digital control system (see Chap. 15) is shown in Fig. C.5.

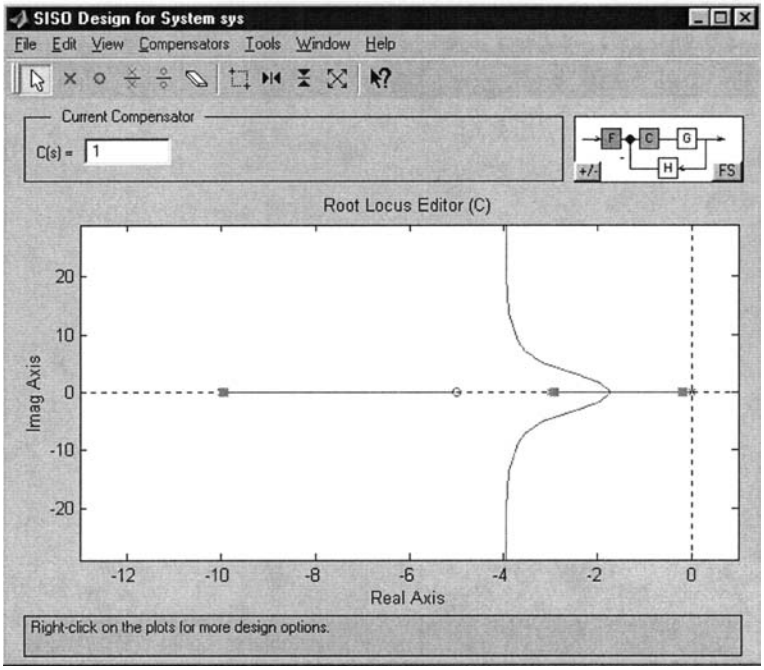


FIGURE C.4 Root locus SISOTOOL window.

C.4.2 Frequency Response

Frequency response is presented in Chaps. 8 and 9. Frequency response plots are generated for LTI system models using the command

```
bode(SYS)
```

which generates a Bode diagram as shown in Fig. C.6 for Eq. (C.2). Another example is shown in Sec. 9.9. Nichols and Nyquist plots are generated with the commands:

```
nichols(SYS)
```

or

```
nyquist(SYS)
```

The "SISOTOOL" described above also provides options for designing on a Bode or Nichols chart. The command

```
sisotool('bode',sys)          %  Start bode plot design tool
```

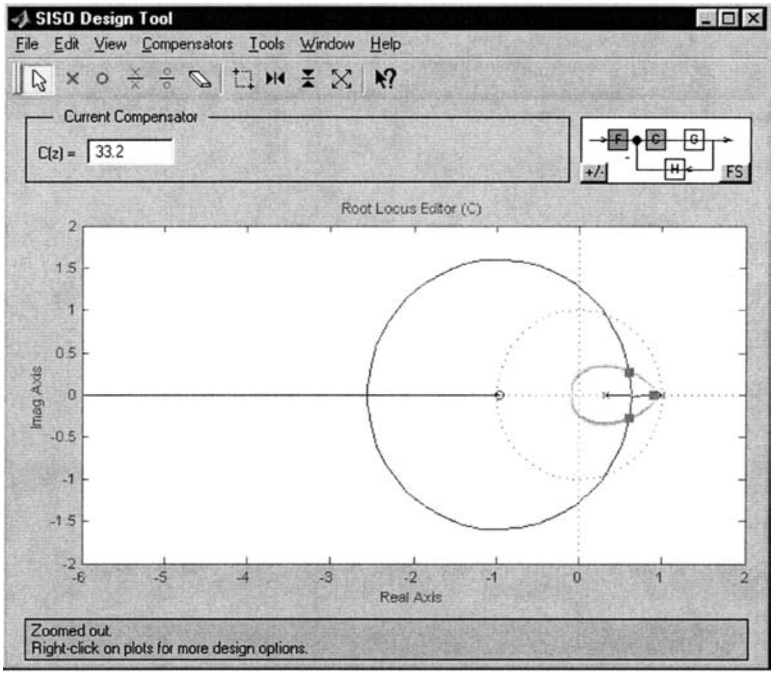


FIGURE C.5 Digital root locus SISOTOOL window.

opens the "SISOTOOL" with the Bode plots as shown in Fig. C.7. The command

```
sisotool('nichols',sys)        %  Start nichols chart
                                  design tool
```

opens the "SISOTOOL" with the Nichols chart as shown in Fig. C.8. With either view, the user can adjust gain or add compensator poles and zeros and immediately see the resulting loop characteristics.

Another "SISOTOOL" example is the Nichols chart shown in Fig. C.9 that depicts the plot of Lm G vs ϕ_G being tangent to an M-contour.

C.5 SIMULATION

Control systems are commonly evaluated by simulating the response to a step and an impulse inputs. The "`step`" and "`impulse`" commands generate plots of the step and impulse response respectively. When invoked with arguments to the left of the equal sign as shown,

```
[y, t] = step(sys)
```

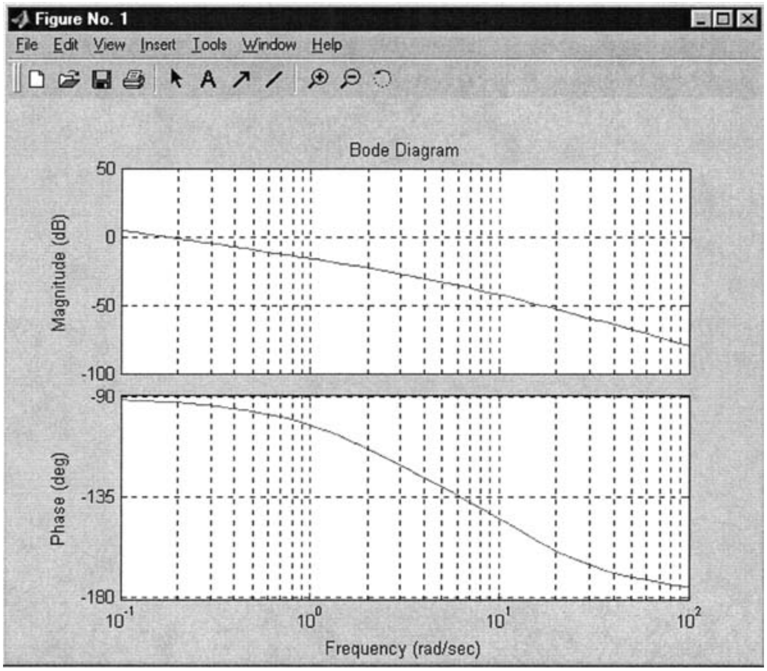


FIGURE C.6 Bode diagram of an example transfer function.

or

```
[y, t] = impulse(sys)
```

the output vector "`y`" and time vector "`t`" that are generated automatically for the plot, are defined in the workspace. These vectors are available then for use elsewhere.

It is often desirable to specify the input as a function of time to generate simulations that are more complex than the step and impulse inputs. This is done by defining an input vector "`u`" and the corresponding time vector "`t`". These vectors are used as inputs to the "`lsim`" command that simulates the response of a linear system as follows:

```
t = [0:.1:10];                  %  Define a simulation time
                                %  vector
u = [ones(1,51) zeros(1,50)];   %  Define vector representing
                                %  a pulse input
[y, t] = lsim(sys,u,t);         %  Simulate the response
plot(t, [y, u])                 %  Plot the response
```

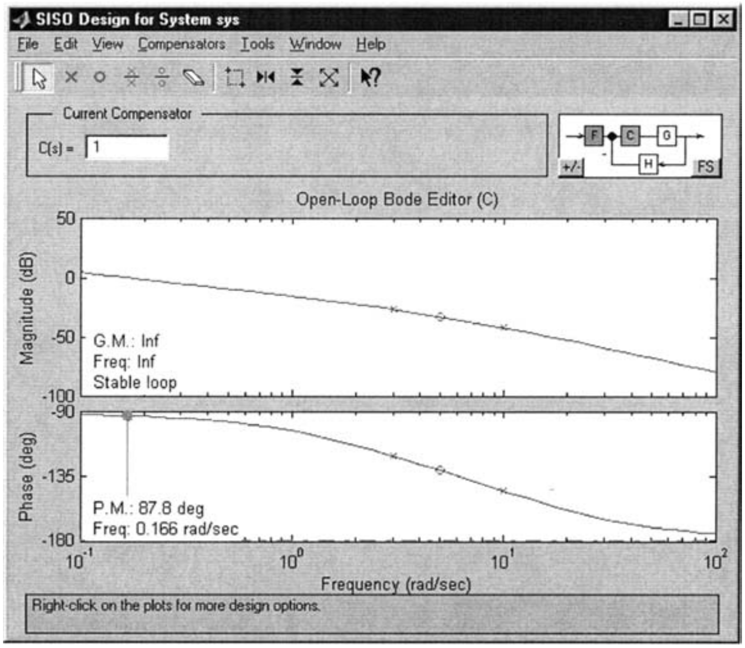


FIGURE C.7 Bode diagram SISOTOOL window.

This example also illustrates the `ones` and `zeros` commands. A vector or matrix whose elements are all `ones` or `zeros` is defined with these commands, using arguments which are the size of the vector or matrix desired. A related function is "`eye (rows, columns)`" which defines an identity matrix or a nonsquare matrix with ones on the principal diagonal and zeros elsewhere.

```
eye(3)
ans=

    1    0    0
    0    1    0
    0    0    1

eye(3,5)
ans=

    1    0    0    0    0
    0    1    0    0    0
    0    0    1    0    0
```

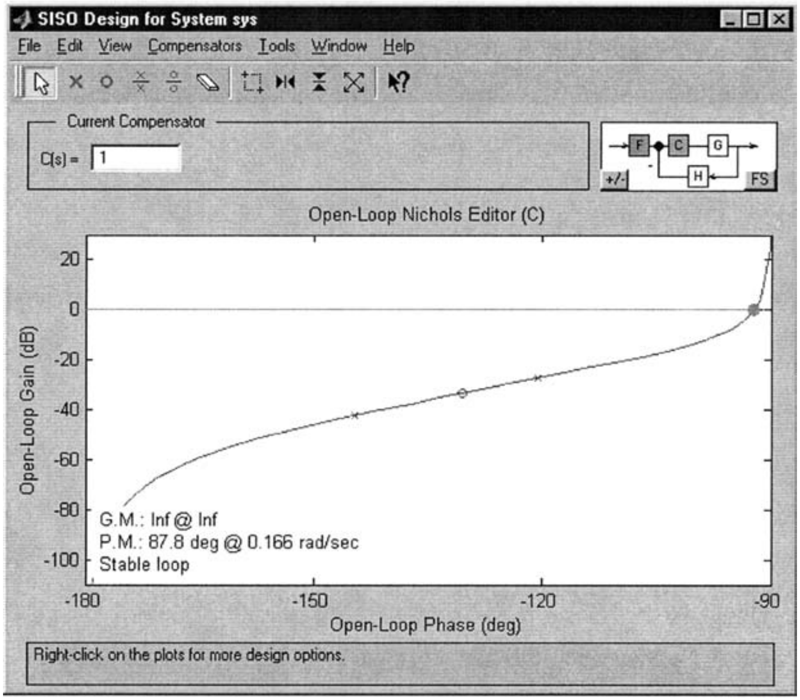


FIGURE C.8 Nichols chart SISOTOOL window.

C.5.1 Simulink

For simulations of more advanced system models, including nonlinear models, it is useful to run Simulink. Simulink is a graphical simulation tool which is started from the MATLAB command window by typing "`simulink`" at the prompt. This opens a window with the Simulink block library and provides an option for creating a new Simulink model. Blocks are selected and placed on the model workspace to create a system block diagram as shown in Fig. C.10.

C.6 IMPLEMENTATION OF RESULTS

Compensators are typically designed as continuous time transfer functions. They may be implemented with analog components, but it is more common to implement them in a digital control system. MATLAB provides commands for transforming analog designs into the discrete domain. To implement the

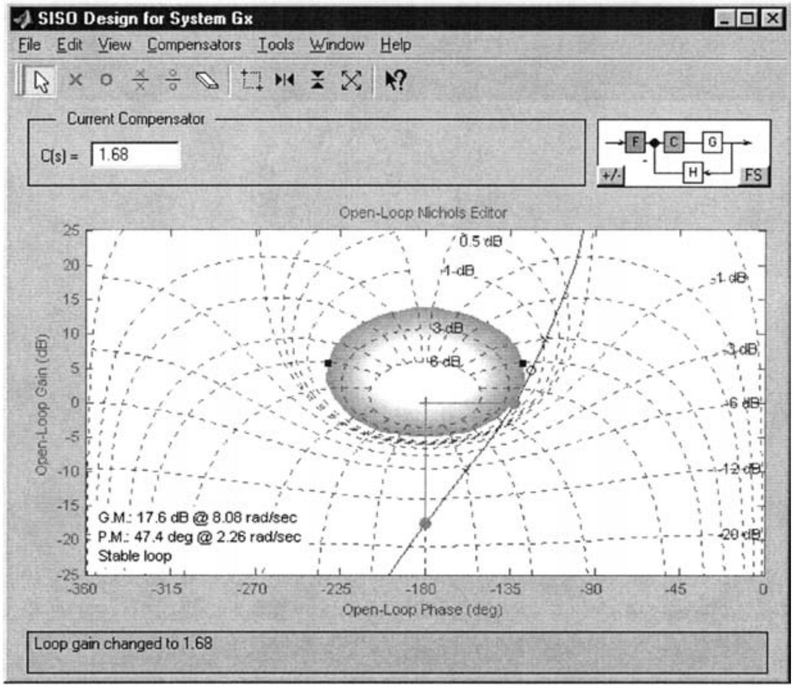


FIGURE C.9 Nichols chart diagram SISOTOOL window with grid and bounds.

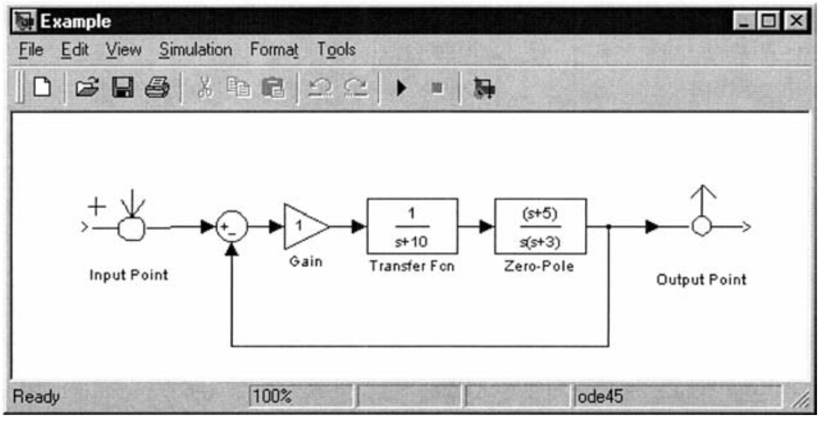


FIGURE C.10 Simulink block diagram.

lead filter $G_c = (s+3)/(s+30)$ in a digital system with a sampling rate of 40 Hz, the "`c2d`" function is used.

```
num = [1  3];
den = [1  30];
ts = 1/40;
sys = tf(num,den)
Transfer function:
s + 3
- - - -
 s + 30

sysd = c2d(sys,ts,'zoh')

Transfer function:
z - 0.9472
- - - - - -
z - 0.4724

Sampling time: 0.025
```

The discrete transfer function is used as described in Chap. 15. The "`c2d`" function provides several transformation methods including the zero order hold method shown here and the Tustin transformation. See the MATLAB documentation for more information.

Appendix D

TOTAL-PC CAD Package

D.1 INTRODUCTION

TOTAL-PC is ideally suited as an educational tool while others are more suited for the practicing engineer. In general, most packages are command-driven interactive programs through the keyboard or a file designation. Some software products include extensive system building capabilities, very high-order plant modeling, nonlinear system construction, and a multitude of data displays as the *MIMO QFT* CAD package of Ref. 5. Some of these packages require extensive knowledge to use their full capabilities, whereas others are not comprehensive in their capabilities.

The computer programs called TOTAL-PC is discussed in this appendix as a specific example of a CAD package. This CAD package along with the MATLAB® and the MIMO QFT CAD packages are the ones employed in the text examples. The disk attached to this text contains the TOTAL-PC CAD packages and its associated USERS Manual. This CAD package can also be used by students for their basic control theory courses. Questions in the use of this package should be addressed to Dr. Robert Ewing whose e-mail address is: Robert.ewing@wpafb.af.mil Users of this CAD package should periodically check with Dr. Ewing for updated versions.

D.2 OVERVIEW OF TOTAL-PC

TOTAL [1, 2] is an option-number package that reflects a hand calculator or line-terminal environment for interface speed and agility. This *FORTRAN* package was originally developed at the Air Force Institute of Technology (AFIT) in the late 1970s; provides an extensive set of control system analysis and design CAD capabilities. During the early 1990s the improved version of TOTAL-PC enhanced the analysis and design of control systems. This CAD package contains over 150 commands (see Table D.1), which are divided according to general functional categories. It divides each of its options (commands) into various groups. It contains the conventional analog and discrete control system analysis and design options and the QFT design options.

TABLE D.1 Some TOTAL-PC Options

	Transfer-function input options
0	List options
1	Recover all data from file memory
2	Polynomial form—GTF (forward transfer function)
3	Polynomial form—HTF (feedback transfer function)
4	Polynomial form—OLTF (open-loop transfer function)
5	Polynomial form—CLTF (closed-loop transfer function)
6	Factored form—GTF
7	Factored form—HTF
8	Factored form—OLTF
9	Factored form—CLTF
	Matrix input options for state equations
10	List options
11	AMAT—Continuous plant matrix
12	BMAT—Continuous input matrix
13	CMAT—Output matrix
14	DMAT—Direct transmission matrix
15	KMAT—State variable feedback matrix
16	FMAT—Discrete plant matrix
17	GMAT—Discrete input matrix
18	Set up stage-space model of system
19	Explain use of above matrices

(*continued*)

Table D.1 Continued

	Block diagram manipulation and state-space options
20	List options
21	Form OLTF=GTF*HTF (in cascade)
22	Form CLTF=(GAIN*GTF)/(1+GAIN*GTF*HTF)
23	Form CLTF=(GAIN*OLTF)/(1+GAIN*OLTF)
24	Form CLTF=GTF+HTF (in parallel)
25	GTF(s) and HTF(s) from continuous state-space model
26	GTF(z) and HTF(z) from discrete state-space model
27	Write adjoint (*sI-AMAT) to file answer
28	Find HTF from CLTF and GTF for CLTF=GTF*HTF/(1+GTF*HTF)
29	Find HTF from CLTF and GTF for CLTF=GTF/(1+GTF*HTF)
	Time-response options, continuous $F(t)$ and discrete $F(kT)$
30	List options
31	Tabular listing of $F(t)$ or $F(kT)$
32	Plot $F(t)$ or $F(kT)$ at user's terminal
33	Printer plot (written to file answer)
34	Calcomp plot (written to file plot)
35	Print time or difference equation [$F(t)$ or $F(kT)$]
36	Partial fraction expansion of CLTF (or OLTF)
37	LIST T-PEAK, T-RISE, T-SETTLING, T-DUP, M-PEAK, final value
38	Quick sketch at user's terminal
39	Select input: step, ramp, pulse, impulse, sin ωT
	Root-locus options
40	List options
41	General root locus
42	Root locus with a gain of interest
43	Root locus with a zeta (damping ratio) of interest
44	List n points on a branch of interest
45	List all points on a branch of interest
46	List locus roots at a gain of interest
47	List locus roots at a zeta of interest
48	Plot root locus at user's terminal
49	List current values of all root locus variables
	Frequency-response options
50	List options
51	Tabular listing
52	Two-cycle scan of magnitude (or dB)
53	Two-cycle scan of phase (degrees or radians)

(*continued*)

TABLE D.1 Continued

	Frequency-response options
54	Plot $F(j\omega)$ at user's terminal
55	Create GNUPlot-Frequency plot
56	Create GNUPlot-Root locus plot
57	Tabulate points of interest: peaks, breaks, etc.
58	Create GNUPLOT-Nichols Log magnitude/angle plot
59	Chalk Pitch Axis HQ Criterion Analysis
	Polynomial operations
60	List options
61	Factor polynomial (POLYA)
62	Add polynomials (POLYC=POLYA+POLYB)
63	Subtract polynomials (POLYC=POLYA-POLYB)
64	Multiply polys (POLYC=POLYA*POLYB)
65	Divide polys (POLYC+REM=POLYA/POLYB)
66	Store Polynomial (POLY_) into POLYD
67	Expand roots into a polynomial
68	$(s+a)^n$ expansion into a polynomial
69	Activate polynomial calculator
	Matrix operations
70	List options
71	ROOTA=eigenvalues of AMAT
72	CMAT=AMAT+BMAT
73	CMAT=AMAT-BMAT
74	CMAT=AMAT*BMAT
75	CMAT=AMAT inverse
76	CMAT=AMAT transposed
77	CMAT=identity matrix **I**
78	DMAT=zero matrix **0**
79	Copy one matrix to another
	Digitization options
80	List options
82	CLTF(s) to CLTF(z) by first difference approximation
83	CLTF(s) to CLTF(z) by Tustin transformation
84	CLTF(z) TO CLTF(s) by impulse invariance
85	CLTF(z) TO CLTF(s) by inverse first difference
86	CLTF(z) TO CLTF(s) by inverse Tustin
87	Find FMAT and GMAT from AMAT and BMAT
89	CLTF(X) to CLTF(Y) by X=ALPHA*(Y+A)/(Y+B)

(*continued*)

TABLE D.1 Continued

Miscellaneous options	
90	LIST OPTIONS
91	Rewind and update Memory file with current data
93	List current switch settings (ECHO, ANSWER, etc.)
96	List special commands allowed in option mode
97	List variable name directory
98	List main options of total
99	Print new features bulletin
119	Augment AMAT:[AMAT]=[AMAT]_GAIN*[BMAT]*[KMAT]
121	Form OLTF=CLK*CLNPOLY/[CLDPOLY-CLK*CLNPOLY](G-EQUIVALENT)
129	(Integral of (CLTF SQUARED))/2PI=
Double-precision discrete transform options*	
140	LIST OPTIONS
141	CLTF(s) TO CLTF(z) (IMPULSE VARIANCE)
142	CLTF(s) TO CLTF(w) $W=(Z-1)/(Z+1)$
143	CLTF(s) TO CLTF(w') $W'=(2/T)(Z-1)/(Z+1)$
144	HI-RATE CLTF(z) TO LO-RATE CLTF(z)
145	HI-RATE CLTF(w) TO LO-RATE CLTF(w)
146	HI-RATE CLTF(w') TO LO-RATE CLTF(w')
147	OPTION 144 (AVOIDING INTERNAL FACTORING)
148	OPTION 144 (ALL CALCULATIONS IN Z PLANE)
Quantitative feedback technique (QFT)	
150	Define-Plant
151	List options
152	Define-TLTF: (Lower Tracking Ratio)
153	Define-TUTF: (Upper Tracking Ratio)
154	Define-TDTF: (Disturbance)
155	Select-Nominal plant: (1–10)
156	Define-FTF: (Pre-filter function)
157	Define-LOTF: (Nominal loop function)
158	Display Plant, TLTF, TUTF, TDTF, FTF, LOTF
159	QFT design/report macro

*Option 37 is not available for the discrete domain.

D.3 QFT CAD PACKAGE

Up to about 1986 there were essentially no Quantitative Feedback Theory (QFT) CAD packages designed specifically to assist in doing a complete

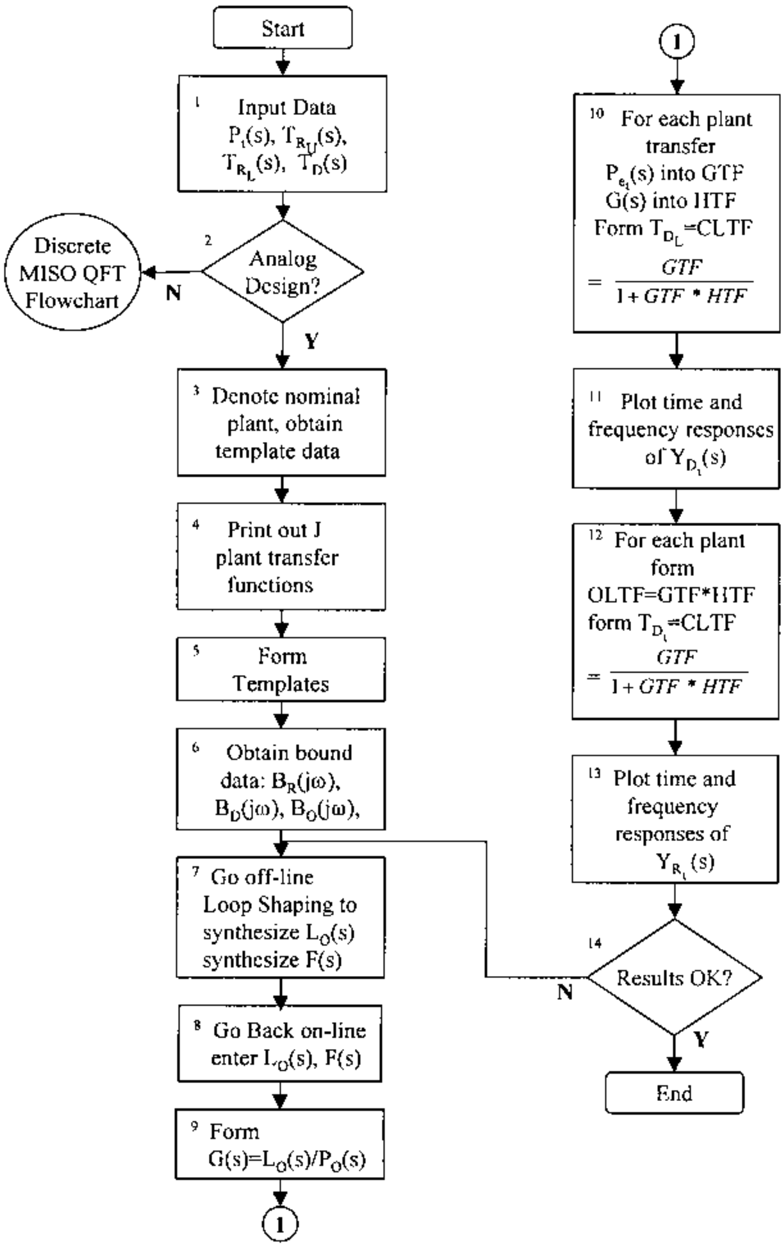


FIGURE D.1 CAD flowchart for MISO analog QFT design.

QFT design for a control system. The TOTAL QFT CAD package [3] for MISO systems was designed as an educational tool as illustrated in Fig. D.1. In 1992 the MIMO QFT CAD package [4] developed at AFIT accelerated the utilization of the QFT technique for the design of robust multivariable control systems.

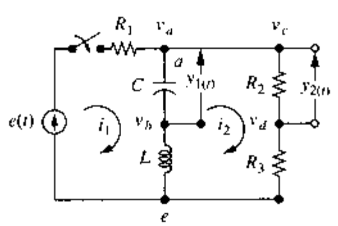
REFERENCES

1. Larimer, S.J.: "An Interactive Computer-Aided Design Program for Digital and Continuous System Analysis and Synthesis (TOTAL)," MS thesis, GE/GGC/EE/78-2, School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB, OH, 1978.
2. Ewing, R.: "TOTAL/PC," School of Engineering, Air Force Institute of Technology, P Street, Wright-Patterson AFB, OH 45433-7765, 1992.
3. Lamont, G.B.: "ICECAP/QFT," School of Engineering, Air Force Institute of Technology, P Street, Wright-Patterson AFB, OH 45433-7765, 1986.
4. Sating, R.R.: "Development of an Analog MIMO Quantitative Feedback Theory (QFT) CAD Package," MS Thesis, AFIT/GE/ENG/92J-04, Graduate School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB, OH, June 1992.
5. Houpis, C.H., and Rasmussen, S.J.: Quantitative Feedback Theory, Fundamentals and Applications, Marcel Dekker, New York, 1999.

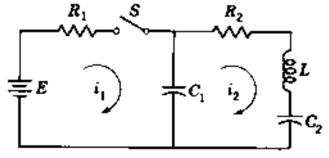
Problems

CHAPTER 2

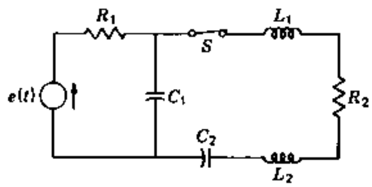
2.1. Write the (*a*) loop, (*b*) node, and (*c*) state equations for the circuit shown after the switch is closed. Let $u = e, y_1 = v_C$, and $y_2 = v_{R_2}$. (*d*) Determine the transfer function y_1/e and $y_2/u = G_2$.



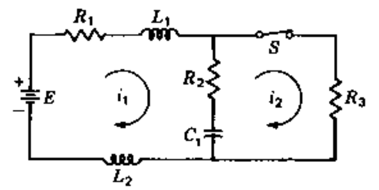
2.2. Write the (*a*) loop, (*b*) node, and (*c*) state equations for the circuit shown after the switch *S* is closed.



2.3. The circuits shown are in the steady state with the switch S closed. At time $t = 0$, S is opened. (a) Write the necessary differential equations for determining $i_1(t)$. (b) Write the state equations.

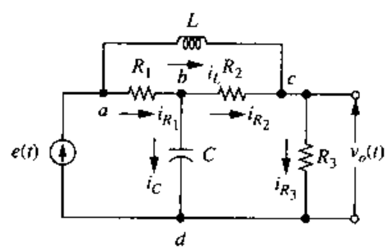


Circuit 1

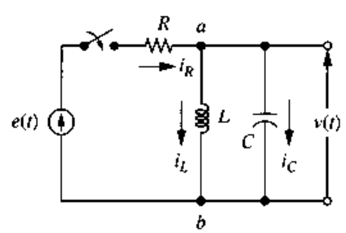


Circuit 2

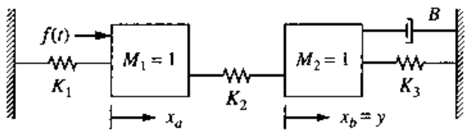
2.4. Write all the necessary equations to determine v_0. (a) Use nodal equations. (b) Use loop equations. (c) Write the state and output equations.



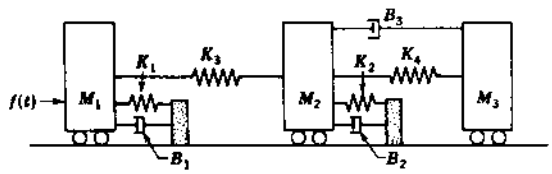
2.5. Derive the state equations. Note that there are only two independent state variables.



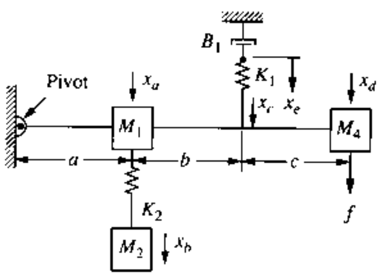
2.6. (a) Derive the differential equation relating the position $y(t)$ and the force $f(t)$. (b) Draw the mechanical network. (c) Determine the transfer function $G(D) = y/f$. (d) Identify a suitable set of independent state variables. Write the state equation in matrix form.



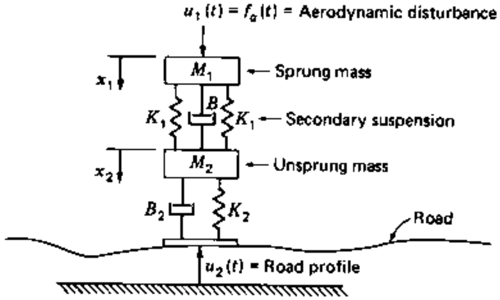
2.7. (*a*) Draw the mechanical network for the mechanical system shown. (*b*) Write the differential equations of performance. (*c*) Draw the analogous electric circuit in which force is analogous to current. (*d*) Write the state equations.



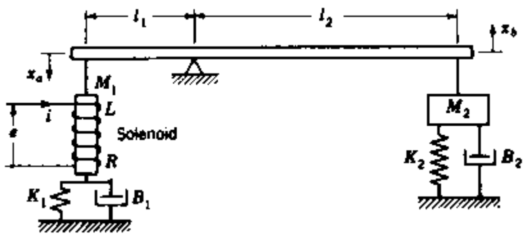
2.8. (*a*) Write the differential equations describing the motion of the following system, assuming small displacements. (*b*) Write the state equations.



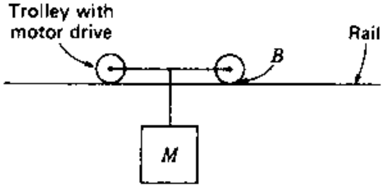
2.9. A simplified model for the vertical suspension of an automobile is shown in the figure. (*a*) Draw the mechanical network; (*b*) write the differential equations of performance; (*c*) derive the state equations; (*d*) determine the transfer function $x_1/\hat{u}_2$, where $\hat{u}_2 = (B_2D + K_2)u_2$ is the force exerted by the road.



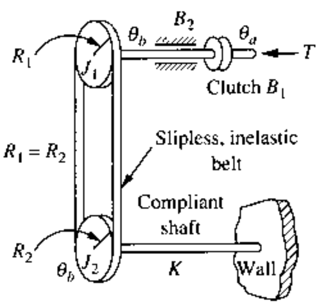
2.10. An electromagnetic actuator contains a solenoid, which produces a magnetic force proportional to the current in the coil, $f = K_i i$. The coil has resistance and inductance. (*a*) Write the differential equations of performance. (*b*) Write the state equations.



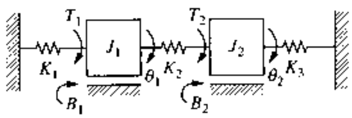
2.11. A warehouse transportation system has a motor drive moving a trolley on a rail. Write the equation of motion.



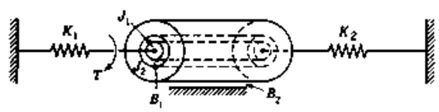
2.12. For the system shown, the torque T is transmitted through a noncompliant shaft and a hydraulic clutch to a pulley #1, which has a moment of inertial J_1. The clutch is modeled by the damping coefficient B_1 (it is assumed to be massless). A bearing between the clutch and the pulley #1 has a damping coefficient B_2. Pulley #1 is connected to pulley #2 with a slipless belt drive. Pulley #2, which has a moment of inertial J_2, is firmly connected to a wall with a compliant shaft that has a spring constant K. The desired outputs are: $y_1 = \theta_b$, the angular position of the J_1, and $y_2 = \dot{\theta}_a$, the angular velocity of the input shaft, before the clutch. (*a*) Draw the mechanical network; (*b*) write the differential equations of performance; (*c*) write the state equations; (*d*) determine the transfer function y_1/u and y_2/u.



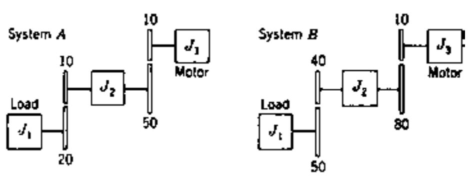
2.13. (*a*) Write the equations of motion for this system. (*b*) Using the physical energy variables, write the matrix state equation.



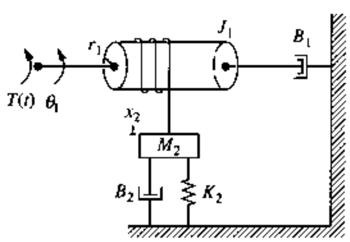
2.14. The figure represents a cylinder of intertia J_1 inside a sleeve of intertia J_2. There is viscous damping B_1 between the cylinder and the sleeve. The springs K_1 and K_2 are fastened to the inner cylinder. (*a*) Draw the mechanical network. (*b*) Write the system equations. (*c*) Draw the analogous circuit. (*d*) Write the state and output equations. The outputs are θ_1 and θ_2. (*e*) Determine the transfer function $G = \theta_2/T$.



2.15. The two gear trains have an identical net reduction, have identical inertias at each stage, and are driven by the same motor. The number of teeth on each gear is indicated on the figures. At the instant of starting, the motor develops a torque T. Which system has the higher initial load acceleration? (*a*) Let $J_1 = J_2 = J_3$; (*b*) let $J_1 = 40J_2 = 4J_3$.

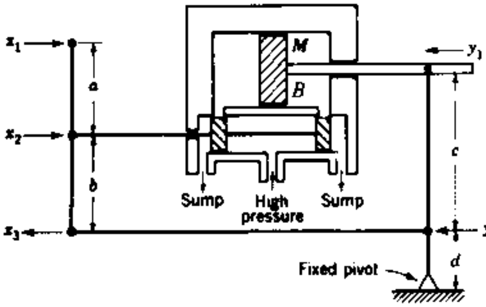


2.16. In the mechanical system shown, r_1 is the radius of the drum. (*a*) Write the necessary differential equations of performance of this system. (*b*) Obtain a differential equation expressing the relationship of the output x_2 in terms of the input $T(t)$ and the corresponding transfer function. (*c*) Write the state equations with $T(t)$ as the input.

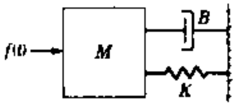


2.17. Write the state and system output equations for the rotational mechanical system of Fig. 2.16. (*a*) With T as the input, use $x_1 = \theta_3$, $x_2 = D\theta_3$, $x_3 = D\theta_2$ and (*b*) with θ_1 as the input, use $x_1 = \theta_3$, $x_2 = D\theta_3$, $x_3 = \theta_2$, $x_4 = D\theta_2$.

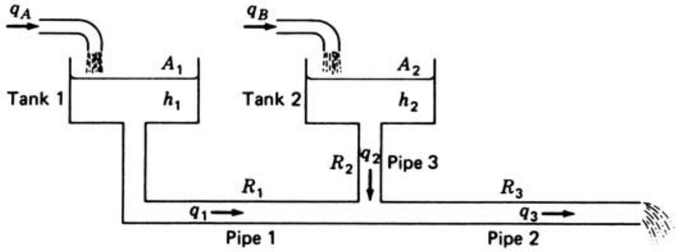
2.18. For the hydraulic preamplifier shown, write the differential equations of performance relating x_1 to y_1. (*a*) Neglect the load reaction. (*b*) Do not neglect load reaction. There is only one fixed pivot, as shown in figure.



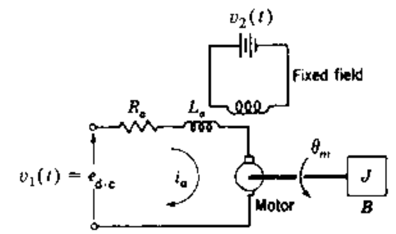
2.19. The mechanical load on the hydraulic translational actuator is Sec. 2.9, Fig. 2.22, is changed to that in the accompanying figure. Determine the new state equations and compare with Eq. (2.114).



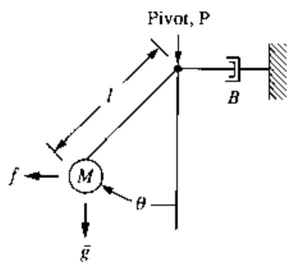
2.20. The sewage system leading to a treatment plant is shown. The variables q_A and q_B are input flow rates into tanks 1 and 2, respectively. Pipes 1, 2, and 3 have resistances as shown. Derive the state equations.



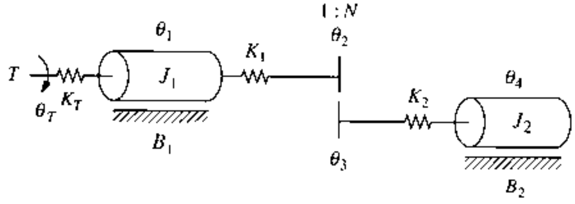
2.21. Most control systems require some type of motive power. One of the most commonly used units is the electric motor. Write the differential equations for the angular displacement of a moment of inertia, with damping, connected directly to a dc motor shaft when a voltage is suddenly applied to the armature terminals with the field separately energized.



2.22. The damped simple pendulum shown in the figure is suspended in a uniform gravitational field g. There is a horizontal force f and a rotational viscous friction B. (*a*) Determine the equation of motion (note: $J = M\ell^2$). (*b*) Linearize the equation of part (*a*) by assuming $\cos\theta \approx 1$ and $\sin\theta \approx \theta$ for small values of θ. (*c*) Write the state and output equations of the system using $x_1 = \theta$, $x_2 = \dot{\theta}$, and $y = x_1$.



2.23. Given the mechanical system shown, the load J_2 is coupled to the rotor J_1 of a motor through a reduction gear and springs K_1 and K_2. The gear-to-teeth ratio is $1:N$. The torque T is the mechanical torque generated by the motor. B_1 and B_2 are damping coefficients, and θ_1, θ_2, θ_3, and θ_4 are angular displacements. (*a*) Draw a mechanical network for this system. (*b*) Write the necessary differential equations in order to solve for θ_4.

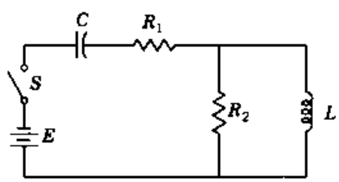


2.24. In some applications the motor shaft of Prob. 2.21 is of sufficient length so that its elastance K must be taken into account. Consider that $v_2(t)$ is constant and that the velocity $\omega_m(t) = D\theta_m(t)$ is being controlled. (a) Write the necessary differential equation for determining the velocity of the load $\omega_J = D\theta_j$. (b) Determine the transfer function $\omega_J(t)/v_1(t)$.

CHAPTER 3

3.1. (a) With $v_c(0^-) = 0$, what are the initial values of current in all elements when the switch is closed? (b) Write the state equations. (c) Solve for the voltage across C as a function of time from the state equations. (d) Find M_p, t_p, and t_s.

$$R_1 = R_2 = 2\ \text{k}\Omega \qquad C = 50\ \mu\text{F} \qquad L = 1\ \text{H} \qquad E = 100\ \text{V}$$



3.2. In Prob. 2.12, the parameters have the following values:

$$J_1 = 1.0\ \text{lb}\cdot\text{ft}\cdot\text{s}^2 \qquad B_1 = 0.5\ \text{lb}\cdot\text{ft/(rad/s)} \qquad K = 0.5\ \text{lb}\cdot\text{ft/rad}$$
$$J_1 = 1.0\ \text{slug}\cdot\text{ft}^2 \qquad B_2 = 3.35\ \text{oz}\cdot\text{ln/(deg/s)}$$

Solve for $\theta_b(t)$ if $T(t) = tu_{-1}(t)$.

3.3. In Prob. 2.10, the parameters have the following values:

$$M_1 = M_2 = 0.05\ \text{slug} \quad L = 1\ \text{H} \quad l_1 = 10\ \text{in.} \quad K_1 = K_2 = 1.2\ \text{lb/in.}$$
$$R = 10\ \Omega \quad l_2 = 20\ \text{in.} \quad B_1 = B_2 = 15\ \text{oz/(in./s)} \quad K_i = 24\ \text{oz/A}$$

With $e(t) = u_{-1}(t)$ and the system initially at rest, solve for $x_b(t)$.

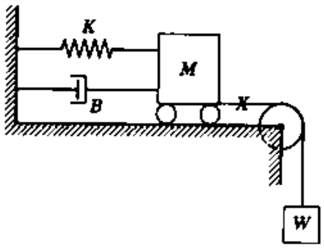
3.4. In part (a) of Prob. 2.18, the parameter have the following values:

$$a = b = 12\ \text{in.} \quad C_1 = 9.0\ \text{(in./s)/in.} \quad c = 10\ \text{in.} \quad d = 2\ \text{in.}$$

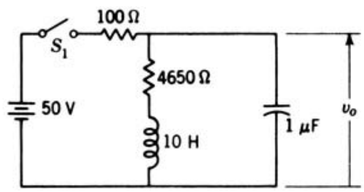
Solve for $y_1(t)$ with $x_1(t) = 0.1u_{-1}(t)$ in. and zero initial conditions.

3.5. (a) $r(t) = (D^3 + 6D^2 + 11D + 6)c(t)$. With $r(t) = \sin t$ and zero initial conditions, determine $c(t)$. (b) $r(t) = [(D+1)(D^2 + 4D + 5)]c(t)$. With $r(t) = tu_{-1}(t)$ and all initial conditions zero, determine the complete solution with all constants evaluated. (c) $D^3c + 10D^2c + 32Dc + 32c = 10r$. Find $c(t)$ for $r(t) = u_{-1}(t)$, $D^2c(0) = Dc(0) = 0$, and $c(0) = -2$. One of the eigenvalues is $\lambda = -2$. (d) For each equation determine $\mathbf{C}(j\omega)/\mathbf{R}(j\omega)$ and $c(t)_{ss}$ for $r(t) = 10 \sin 5t$.

3.6. The system shown is initially at rest. At time $t = 0$ the string connecting M to W is severed at X. Measure $x(t)$ from the initial position. Find $x(t)$.



3.7. Switch S_1 is open, and there is no energy stored in the circuit. (a) Write the state equations. (b) Find $v_o(t)$ after switch S_1 is closed. (c) After the circuit reaches steady state, the switch S_1 is open. Find $v_o(t)$.



3.8. Solve the following differential equations. Assume zero initial conditions. Sketch the solutions.

(a) $D^2x + 16x = 1$ (b) $D^2x + 4Dx + 3x = 9$

(c) $D^2x + Dx + 4.25x = t + 1$ (d) $D^3x + 3D^2x + 4Dx + 2x = 10\sin 10t$

3.9. For Prob. 2.2, solve for $i_2(t)$, where

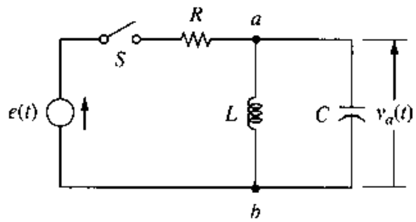
$$E = 10\ \text{V} \quad L = 1\ \text{H} \quad C_1 = C_2 = 0.001\ \text{F} \quad R_1 = 10\ \Omega \quad R_2 = 15\ \Omega$$

3.10. For the circuit of Prob. 2.2: (a) obtain the differential equation relating $i_2(t)$ to the input E; (b) write the state equations using the physical energy variables; (c) solve the state and output equations and compare with the solution of Prob. 3.9; (d) write the state and output equations using the phase variables, starting with the differential equation from part (a).

3.11. In Prob. 2.3, the parameters have the values $R_1 = R_2 = R_3 = 10\ \Omega$, $C_1 = C_2 = 1\,\mu F$, and $L_1 = L_2 = 50$ H. If $E = 100$ V, (a) find $i_1(t)$; (b) sketch $i_1(t)$ and compute M_o, t_p, and t_s; (c) determine the value of T_s (± 2 percent of $i(0)$); (d) solve for $i_1(t)$ from the state equation.

3.12. For $L = 0.4$, $C = 0.5$, and (1) $R = 1$, (2) $R = 0.5$; (a) show that the differential equation is $(RCD^2 + D + R/L)v_a = De$; ($b$) find $v_a(t)_{ss}$ for $e(t) = u_{-1}(t)$; (c) find the roots m_1 and m_2; (d) find $v_a(0^+)$; (e) find

$Dv_a(0^+)$ [use $i_c(0^+)$]; (*f*) determine the complete solution $v_a(t)$; (*g*) sketch and dimension the plot of $v_a(t)$.



3.13. Solve the following equations. Show explicitly $\Phi(t)$, $\mathbf{x}(t)$, and $\mathbf{y}(t)$.

$$\dot{\mathbf{x}} = \begin{bmatrix} -6 & 4 \\ -2 & 0 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}u \qquad \mathbf{y} = [1 \quad 0]\mathbf{x}$$

$$x_1(0) = 2 \qquad x_2(0) = 0 \qquad u(t) = u_{-1}(t)$$

3.14. A single-degree-of-freedom representation of the rolling dynamics of an aircraft, together with a first-order representation of the aileron servomotor, is given by

$$p(t) = \dot{\phi}(t) \qquad J_x\dot{p}(t) = L_{\delta_A}\delta_A(t) + L_p p(t)$$

$$T\dot{\delta}_A(A) = (\delta_A)_{\mathrm{comm}}(t) - \delta_A(t)$$

(*a*) Derive a state-variable representation of the aircraft and draw a block diagram of the control system with

$$(\delta_A)_{\mathrm{comm}}(t) = AV_{\mathrm{ref}} - \mathbf{P}\mathbf{x}(t)$$

where

$x(t)$ = state vector of the aircraft system = $[\phi \quad p \quad \delta_A]^T$
$\mathbf{P}$ = matrix of correct dimensions and with constant nonzero elements p_{ij}.
A = scalar gain
V_{ref} = the reference input
$(\delta_A)_{\mathrm{comm}}$ = the input to the servo

(*b*) Determine the aircraft state equation in response to the reference input V_{ref}.

3.15. Use the Sylvester expansion of Sec. 3.14 to obtain $\mathbf{A}^{10}$ for

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{bmatrix}$$

3.16. For the mechanical system of Fig. 2.11*a* the state equation for Example 2, Sec. 2.6, is given in *phase-variable form*. With the input as $u = x_a$, the state variables are $x_1 = x_b$ and $x_2 = \dot{x}_b$. Use $M = 5$, $K = 10$, and $B = 15$. The initial conditions are $x_b(0) = 1$ and $\dot{x}_b(0) = -2$. (*a*) Find the homogeneous solution of $\mathbf{x}(t)$. (*b*) Find the complete solution with $u(t) = u_{-1}(t)$.

3.17. For the autonomous system

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -5 & -7 & -3 \end{bmatrix} \mathbf{x} \qquad \mathbf{y} = [1 \quad 0 \quad 0]\mathbf{x}$$

(*a*) Find the system eigenvalues. (*b*) Evaluate the transmission matrix $\phi(t)$ by means of Sylvester's expansion theorem.

CHAPTER 4

4.1. Determine the inverse Laplace transform by use of partial fraction expansion:

$$(a)\ F(s) = \frac{9}{(s+1)^2(s^2+2s+10)} \qquad (b)\ F(s) = \frac{90}{s(s^2+4s+13)(s+5)}$$

$$(c)\ F(s) = \frac{20(s+5)}{(s+2)^2(s^2+4s+5)}$$

4.2. Find $x(t)$ for

$$(a)\ X(s) = \frac{4}{s^3+6s^2+16s+16} \qquad (b)\ X(s) = \frac{10(s^2+4s+13)}{s(s+2)(s^2+4s+5)}$$

$$(c)\ X(s) = \frac{0.969426(s^2+4.2s+13.44)}{s(s+1)(s^2+4s+13)}$$

(*d*) Sketch $x(t)$ for parts (*a*) and (*b*).

4.3. Given an ac servomotor with inertia load, find $\omega(t)$ by (*a*) the classical method; (*b*) the Laplace-transform method.

$$\omega(0) = 5 \times 10^3 \text{ rad/s} \quad K_\omega = -6 \times 10^{-3} \text{ oz} \cdot \text{in./(rad/s)}$$

$$K_c = 1.4 \text{ oz} \cdot \text{in./V}$$

$$e(t) = 5u_{-1}(t)\text{V} \quad J = 15.456 \text{ oz} \cdot \text{in.}^2$$

4.4. Repeat the following problems using the Laplace transform:

(a) Prob. 3.1 (b) Prob. 3.3 (c) Prob. 3.5

(d) Prob. 3.6 (e) Prob. 3.10 (f) Prob. 3.11(d)

(g) Prob. 3.13 (g) Prob. 3.14

4.5. Find the partial-fraction expansions of the following:

(a) $F(s)=\dfrac{6}{(s+2)(s+6)}$ (b) $F(s)=\dfrac{10}{s(s+2)(s+5)}$

(c) $F(s)=\dfrac{26}{s(s^2+6s+13)}$ (d) $F(s)=\dfrac{0.5(s+6)}{s^2(s+1)(s+3)}$

(e) $F(s)=\dfrac{10(s+4)}{s^2(s^2+4s+20)}$ (f) $F(s)=\dfrac{4(s+5)}{(s+1)(s^2+4s+20)}$

(g) $F(s)=\dfrac{20}{s^2(s+10)(s^2+8s+20)}$ (h) $F(s)=\dfrac{15(s+2)}{s(s+3)(s^2+6s+10)}$

(i) $F(s)=\dfrac{13(s+1.01)}{s(s+1)(s^2+4s+13)}$ (j) $F(s)=\dfrac{0.9366(s^2+6.2s+19.22)}{s(s+1)(s^2+6s+18)}$

Note: For parts (g) through (j) use a CAD program.

4.6. Solve the differential equations of Prob. 3.8 by means of the Laplace transform.

4.7. Write the Laplace transforms of the following equations and solve for $x(t)$; the initial conditions are given to the right.

(a) $Dx + 4x = 0$ $x(0) = 5$

(b) $D^2x + 2.8Dx + 4x = 10$ $x(0) = 2, Dx(0) = 3$

(c) $D^2x + 4Dx + 13x = t$ $x(0) = 0, Dx(0) = -4$

(d) $D^3x + 4D^2x + 9Dx + 10x = \sin 5t$

$x(0) = -4, Dx(0) = 1, D^2x(0) = 0$

4.8. Determine the finial value for:
(a) Prob. 4.1 (b) Prob. 4.2 (c) Prob. 4.5

4.9. Determine the initial value for:
(a) Prob. 4.1 (b) Prob. 4.2 (c) Prob. 4.5

4.10. For the functions of Prob. 4.5, plot M vs. ω and α vs. ω.

4.11. Find the complete solution of $x(t)$ with zero initial conditions for
(1) $(D^2+2D+2)(D+5)x=(D+3)f(t)$
(2) $(D^2+3D+2)(D+5)x=(D+6)f(t)$

The forcing function $f(t)$ is
(a) $u_0(t)$ (b) $10u_{-1}(t)$ (c) $tu_{-1}(t)$ (d) $2t^2u_{-1}(t)$

4.12. A linear system is described by

$$(1)\quad \dot{\mathbf{x}} = \begin{bmatrix} -5 & 1 \\ 2 & -2 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\mathbf{u} \qquad \mathbf{y} = [0 \quad 1]\,\mathbf{x}$$

$$(2)\quad \dot{\mathbf{x}} = \begin{bmatrix} 0 & 2 \\ -4 & -6 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}\mathbf{u} \qquad \mathbf{y} = [1 \quad 1]\mathbf{x}$$

where $u(t) = u_{-1}(t)$ and the initial conditions are $x_1(0) = 1$, $x_2(0) = 1$. (a) Using Laplace transforms, find $\mathbf{X}(s)$. Put the elements of this vector over a common denominator. (b) Find the transfer function $G(s)$. (c) Find $y(t)$.

4.13. A linear system is represented by

$$(1)\quad \dot{\mathbf{x}} = \begin{bmatrix} -3 & 1 \\ -3 & -7 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1 \\ 1 \end{bmatrix}\mathbf{u} \qquad \mathbf{y} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\mathbf{x} \qquad \mathbf{u} = u_{-1}(t)$$

$$(2)\quad \dot{\mathbf{x}} = \begin{bmatrix} 0 & 4 \\ -2 & -6 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}\mathbf{u} \qquad \mathbf{y} = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}\mathbf{x} \qquad \mathbf{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}u_{-1}(t)$$

$$(3)\quad \dot{\mathbf{x}} = \begin{bmatrix} 0 & -2 \\ 0 & -6 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1 \\ 1 \end{bmatrix}\mathbf{u} \qquad \mathbf{y} = \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix}\mathbf{x} \qquad u = u_{-1}(t)$$

(a) Find the complete solution for $y(t)$ when $x_1(0) = 0$ and $x_2(0) = 1$. (b) Determine the transfer functions. (c) Draw a block diagram representing the system.

4.14. A system is described by

$$(1)\quad \dot{\mathbf{x}} = \begin{bmatrix} -4 & -1 \\ 3 & 0 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\mathbf{u} \qquad y = x_1$$

$$(2)\quad \dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 \\ -6 & -5 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1 \\ 1 \end{bmatrix}\mathbf{u} \qquad y = x_1$$

(a) Find $x(t)$ with $x(0) = 0$ and $u = u_{-1}(t)$. (b) Determine the transfer function $G(s) = Y(s)/U(s)$.

4.15. Repeat Prob. 3.17 using the Laplace transform.

4.16. Shown that Eqs. (4.52) and (4.53) are equivalent expressions.

4.17. Given

$$Y(s) = \frac{K(s + a_1)\cdots(s + a_w)}{s(s + \zeta_1\omega_{n_1} \pm j\omega_{d_1})(s + \zeta_3\omega_{n_3} \pm j\omega_{d_3})(s + b_5)\cdots(s + b_n)}$$

all the a_i's and b_k's are positive and real and $\zeta_1\omega_{n_1} = \zeta_3\omega_{n_3}$ with $\omega_{n_3} > \omega_{n_1}$. Determine the effect on the Heaviside partial-fraction coefficients associated with the pole $p_3 = -\zeta_3\omega_{n_3} + j\omega_{d_3}$ when ω_{n_3} is allowed to approach infinity. *Hint*: Analyze

$$\lim_{\omega\to\infty} A_3 = \lim_{\omega\to\infty} [(s-p_k)Y(s)]_{s=p_3=-\zeta_3\omega_{n_3}+j\omega_{d_3}}$$

K has the value required for $y(t)_{ss} = 1$ for a unit step input. Consider the cases (1) $n = w$, (2) $n > w$.

4.18. Repeat Prob. 4.17 with $a_1 = 0, w = 2, n = 6$, and K is a fixed value.

4.19. For the linearized system of Prob. 2.22 determine (*a*) the transfer function and (*b*) the value of B that makes this system critically damped.

4.20. Determine the transfer function and draw a block diagram.

$$\dot{\mathbf{x}} = \begin{bmatrix} -3 & 0 \\ 4 & -2 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}\mathbf{u} \qquad \mathbf{y} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\mathbf{x}$$

CHAPTER 5

5.1. For the temperature-control system of Fig. 5.1, some of the pertinent equations are $b = K_b\theta, f_s = K_s i_s, q = K_q x, \theta = K_c q/(D + a)$. The solenoid coil has resistance R and inductance L. The solenoid armature and valve have mass M, damping B, and a restraining spring K. (*a*) Determine the transfer function of each block in Fig. 5.1*b*. (*b*) Determine the forward transfer function $G(s)$. (*c*) Write the state and output equations in matrix form. Use $x_1 = \theta, x_2 = x, x_3 = \dot{x}, x_4 = i_s$.

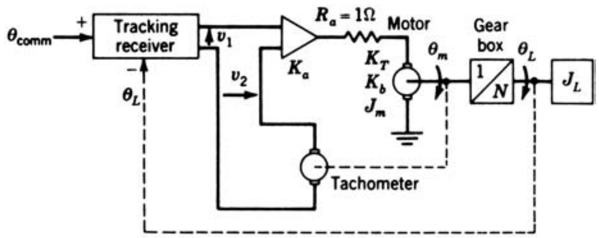
5.2. Find an example of a practical closed-loop control system not covered in this book. Briefly describe the system and show a block diagram.

5.3. A satellite-tracking system is shown in the accompanying schematic diagram. The transfer function of the tracking receiver is

$$\frac{V_1}{\Theta_{comm} - \Theta_L} = \frac{6}{1 + s/42\pi}$$

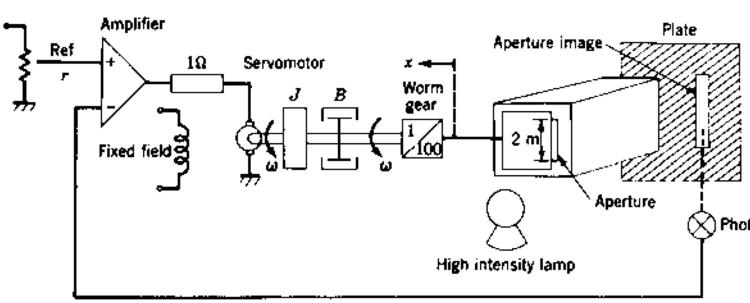
The following parameters apply:

K_a = gain of servoamplifier = 60
K_t = tachometer constant = 0.05 V · s
K_T = motor torque constant = 0.6 N · m/A
K_b = motor back-emf constant = 0.75 V · s
J_L = antenna inerita = 3,000 kg · m^2
J_m = motor inertia = 8×10^{-3} kg · m^2
$1:N$ = gearbox stepdown ratio = 1 : 12,200



(a) Draw a detailed block diagram showing all the variables. (*b*) Derive the transfer function $\Theta_L(s)/\Theta_{comm}(s)$.

5.4. A photographic control system is shown in simplified form in the diagram. The aperture slide moves to admit the light from the high-intensity lamp to the sensitive plate, the illumination of which is a linear function of the exposed area of the aperture. The maximum area of the aperture is 4 m^2. The plate is illuminated by 1 candela (cd) for every square meter of aperture area. This luminosity is detected by the photocell, which provides an output voltage of 1 V/cd. The motor torque constant is 2 N·m/A, the motor inertia is 0.25 kg·m^2, and the motor back-emf constant is 1.2 V·s. The viscous friction is 0.25 N·m·s, and the amplifier gain is 50 V/V. (*a*) Draw the block diagram of the system with the appropriate transfer function inserted in each block. (*b*) Derive the overall transfer function.



5.5. The longitudinal motion of an aircraft is represented by the vector differential equation

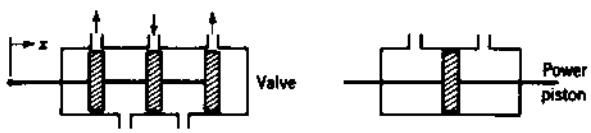
$$\dot{\mathbf{x}} = \begin{bmatrix} -0.09 & 1 & -0.02 \\ -8.0 & -0.06 & -6.0 \\ 0 & -0 & -10 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ 0 \\ 10 \end{bmatrix} \delta_E$$

where

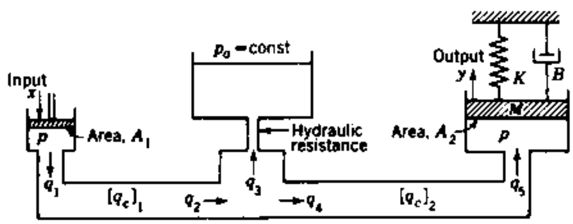
x_1 = angle of attack
x_2 = rate of change of pitch angle
x_3 = incremental elevator angle
δ_E = control input into the elevator actuator

Derive the transfer function relating the system output, rate of change of pitch angle, to the control input into the elevator actuator, x_2/δ_E.

5.6. (*a*) Use the hydraulic valve and power piston in a closed-loop position control system, and derive the transfer function of the system. (*b*) Draw a diagram of a control system for the elevators on an airplane. Use a hydraulic actuator.



5.7. In the diagram $[q_c]$ represents compressibility flow, and the pressure p is the same in both cylinders. Assume there is no leakage flow around the pistons. Draw a block diagram that relates the output $Y(s)$ to the input $X(s)$.



5.8. The dynamic equations that describe an aircraft in the landing configuration are as follows:

$$\dot{h}(t) = b_{33}h(t) + b_{32}\theta(t)$$
$$\ddot{\theta}(t) = b_{11}\dot{\theta}(t) + b_{12}\theta(t) + b_{13}h(t) + c_{11}\delta_e(t)$$

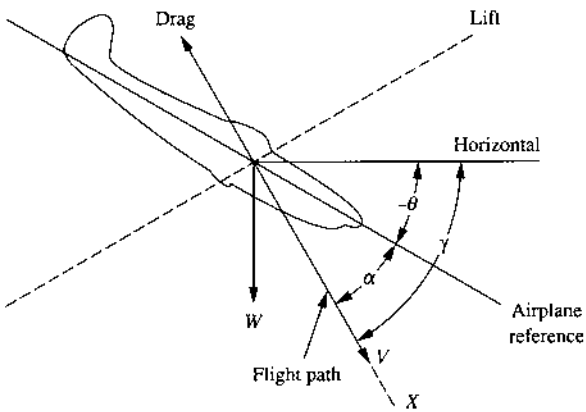
where the coefficients are defined as follows:

$$b_{11} = \frac{1}{T_s} - 2\zeta\omega_s \qquad b_{13} = \frac{1}{VT_s^2} - \frac{2\zeta\omega_s}{VT_s} + \frac{\omega_s^2}{V} \qquad b_{33} = -\frac{1}{T_s}$$
$$b_{12} = \frac{2\zeta\omega_s}{T_s} - \omega_s^2 - \frac{1}{T_s^2} \qquad b_{32} = \frac{V}{T_s} \qquad c_{11} = \omega_s^2 K_s T_s$$

These coefficients are constant for a given flight condition and aircraft configuration. The parameters are defined as follows:

K_s = short-period aircraft gain	T_s = flight path time constant
ω_s = short-period natural frequency	h = altitude
ζ = short-period damping ratio	θ = pitch angle

(a) Draw a detailed block diagram that explicitly shows each time-varying quantity and the constant coefficients (left in terms of the b's and c's). The input is $\delta_e(t)$ and the output is $h(t)$. (b) Determine the transfer function $G(s) = H(s)/\Delta_e(s)$.

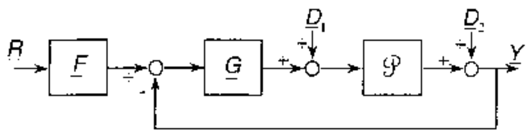


5.9. A multiple-input multiple-output (MIMO) digital robot arm control system (see Fig. 1.6) may be decomposed into sets of analog multiple-input single-output (MISO) control systems by the pseudo-continuous time (PCT) technique. The figure shown represents one of these equivalent MISO systems, where $r(t)$ is the input that must be tracked, and $d_1(t)$ and $d_2(t)$ are unwanted disturbance inputs. For this problem consider $d_2(t) = 0$ and

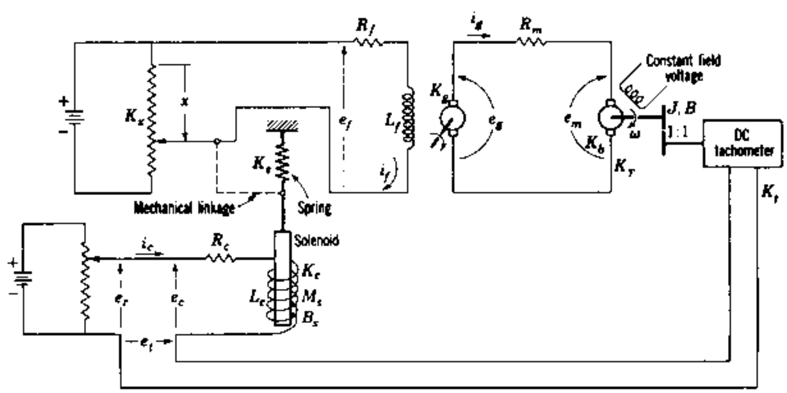
$$F(s) = \frac{0.9677615(s + 13)}{(s + 3.2 \pm j1.53)}$$

$$G(s) = \frac{824000(s + 8)(s + 450)}{(s + 110)(s + 145 \pm j285)}$$

$$\mathscr{P}(s) = P(s) = \frac{125}{s(s+450)(s-3)}$$



5.10. For the control system shown, (1) the force of attraction on the solenoid is given by $f_c = K_c i_c$, where K_c has the units pounds per ampere. (2) The voltage appearing across the generator field is given by $e_f = K_x x$, where K_x has the units volts per inch and x is in inches. (3) When the voltage across the solenoid coil is zero, the spring K_s is unstretched and $x = 0$. (*a*) Derive all necessary equations relating all the variables in the system. (*b*) Draw a block diagram for the control system. The diagram should include enough blocks to indicate specifically the variables $I_c(s)$, $X(s)$, $I_f(s)$, $E_g(s)$, and $T(s)$. Give the transfer function for each block in the diagram. (*c*) Draw an SFG for this system. (*d*) Determine the overall transfer function. (*e*) Write the system state and output equation in matrix form.



5.11. Given

(1) $D^3y + 10D^2y + 31Dy + 3y = D^3u + 4D^2u + 8Du + 2u$
(2) $D^4y + 14D^3y + 71D^2y + 154Dy + 120y = 4D^2u + 8u$
(3) $D^3y + 6D^2y + 5Dy - 12y = 3D^2u + 6Du + u$

Obtain state and output equations using (*a*) phase variables, (*b*) canonical variables.

5.12. For the system described by the state equation

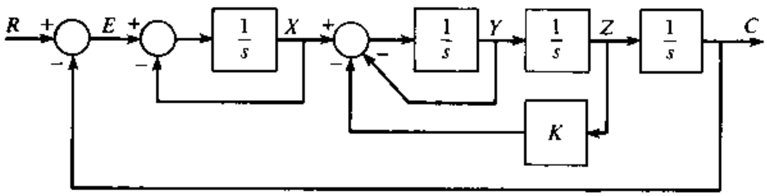
$$(1)\quad \dot{\mathbf{x}} = \begin{bmatrix} 0 & 4 \\ -1 & -5 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 2 \\ 1 \end{bmatrix}\mathbf{u}$$

$$(2)\quad \dot{\mathbf{x}} = \begin{bmatrix} -2 & -1 \\ 0 & 3 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 0 \\ 10 \end{bmatrix}\mathbf{u}$$

(*a*) Derive the system transfer function if $y = x_1$. (*b*) Draw an appropriate state-variable diagram. (*c*) For zero initial state and a unit-step input evaluate $x_1(t)$ and $x_2(t)$.

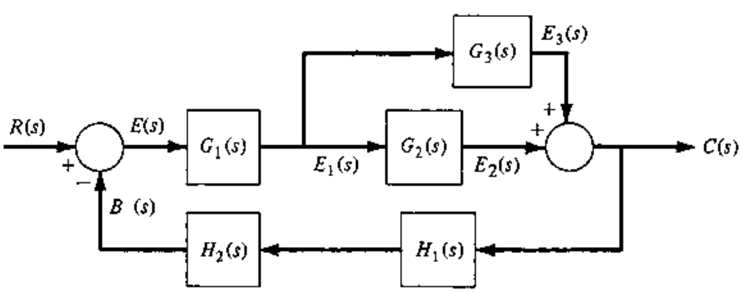
5.13. Draw a simulation diagram for the state equations of Prob. 3.7 using physical variables.

5.14. Given the following system:



(*a*) Draw an equivalent signal flow graph. (*b*) Apply Mason's gain rule to find $C(s)/R(s)$. (*c*) Write the four differential and three algebraic equations that form the math model depicted by the block diagram above. Use only the signals $r(t)$, $e(t)$, $x(t)$, $y(t)$, $z(t)$, and $c(t)$ in these equations.

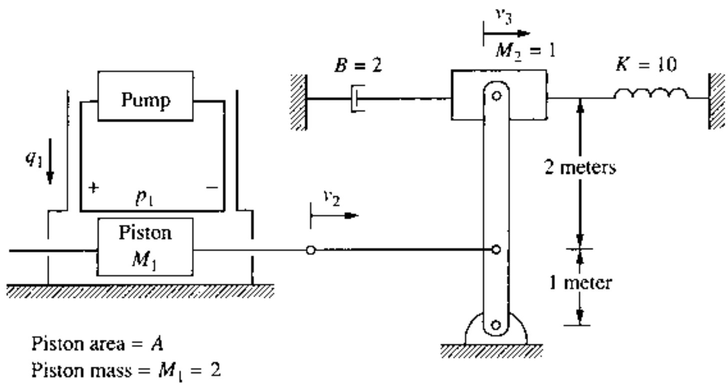
5.15. For the following system,



(*a*) Draw a signal flow graph. (*b*) Derive transfer functions for $E(s)/R(s)$, $C(s)/R(s)$, and $B(s)/R(s)$.

5.16. The pump in the system shown below supplies the pressure $p_1(t) = 4u_{-1}(t)$. Calculate the resulting velocity $v_2(t)$ of the mass M_2.

Include the effect of the mass of the power piston. (*a*) Neglect load reaction. (*b*) Include load reaction.



CHAPTER 6

6.1. For Prob. 5.3, (*a*) use the Routh criterion to check the stability of the system. (*b*) Compute the steady-state tracking error of the system in response to a command input which is a ramp of 0.0317°/s. (*c*) Repeat (*b*) with the gain of the tracking receiver increased from 16 to 40.

6.2. For each of the following cases, determine the range of values of K for which the response $c(t)$ is stable, where the driving function is a step function. Determine the roots on the imaginary axis that yield sustained oscillations.

(*a*) $$C(s) = \frac{K}{s[s(s+2)(s^2+4s+20)+K]}$$

(*b*) $$C(s) = \frac{K(s+1)}{s[s^2(s^2+3s+2)+K(s+0.1)]}$$

(*c*) $$C(s) = \frac{20K}{s[s(s+1)(s+5)+20K]}$$

(*d*) $$C(s) = \frac{K(s+5)}{s[s(s+1)(s+2)(s+3)+K(s+5)]}$$

(*e*) For Prob. 5.14.

6.3. Factor the following equations:

(*a*) $s^3 + 6.4s^2 + 18.48s + 19.36 = 0$

(*b*) $s^4 + 13s^2 + 36 = 0$

(*c*) $s^3 + 6s^2 + 15.25s + 18.75 = 0$

(*d*) $s^4 + 9s^3 + 37s^2 + 81s + 52 = 0$

(*e*) $s^4 + 5s^3 + 82s^2 + 208s + 240 = 0$

(*f*) $s^4 + 3s^3 - 15s^2 - 19s + 30 = 0$

(*g*) $s^5 + 10s^4 + 50s^3 + 140s^2 + 209s + 130 = 0$

6.4. Use Routh's criterion to determine the number of roots in the right-half *s* plane for the equations of Prob. 6.3.

6.5. A unity-feedback system has the forward transfer function

$$G(s) = \frac{K(2s/3 + 1)}{s(s+1)(s+5)(s+10)}$$

In order to obtain the best possible ramp error coefficient, the highest possible gain K is desirable. Do stability requirements limit this choice of K?

6.6. The equation relating the output $y(t)$ of a control system to its input is

(*a*) $[s^4 + 16s^3 + 65s^2 + (50 + K)s + 1.2K]\,Y(s) = 5X(s)$

(*b*) $[s^3 - 7s^2 + 10s + 2 + K(s+2)]\,Y(s) = X(s)$

Determine the range of K for stable operation of the system. Consider both positive and negative values of K.

6.7. The output of a control system is related to its input $r(t)$ by

$$[s^4 + 12s^3 + (2.1 + K)s^2 + (20 + 10K)s + 25K]C(s) = K(s+1)R(s)$$

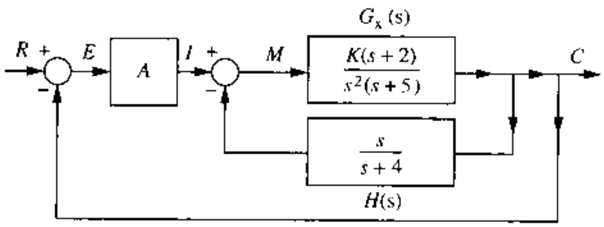
Determine the range of K for stable operation of the system. Consider both positive and negative values of K. Find K that yields imaginary roots.

6.8.

$$F(s) = \frac{1}{s[(s^4 + 0.5s^3 + 4.5s^2 + s + 2) + K(s+1)]}$$

K is real but may be positive or negative. (*a*) Find the range of values of K for which the time response is stable. (*b*) Select a value of K that will produce imaginary poles for $F(s)$. Find these poles. What is the physical significance of imaginary poles as far as the time response is concerned?

6.9. For the system shown, (a) find $C(s)/R(s)$. (b) What type of system does $C(s)/E(s)$ represent? (c) Find the step-, ramp-, and parabolic-error coefficients. (d) Find the steady-state value of $c(t)$, $e(t)$, and $m(t)$ if $r(t)=4u_{-1}(t)$, $K=1$, and $A=1$. (e) A steady-state error less than 0.5 is required. Determine the value for A and K that meet this condition. What effect does K have on the steady-state error? (f) For parts (d) and (e), is the closed-loop system stable?



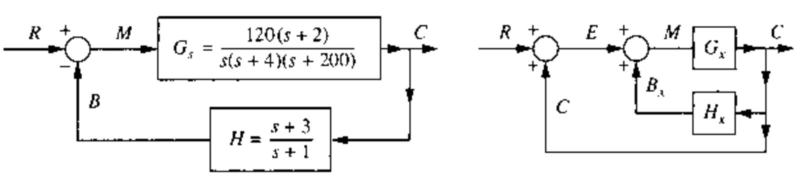
6.10. For a unity-feedback control system:

$$(1)\quad G(s)=\frac{10(s+2)}{s(s+1)(s+4)(s^2+4s+13)}$$

$$(2)\quad G(s)=\frac{400(s+2)}{s^2(s+5)(s^2+2s+10)}$$

(a) Determine the step-, ramp-, and parabolic-error coefficients (K_p, K_v, and K_a) for this system. (b) Using the appropriate error coefficients from part (a), determine the steady-state actuating signal $e(t)_{ss}$ for $r_1(t)=(18+t)u_{-1}(t)$ and for $r_2(t)=4t^2u_{-1}(t)$. (c) Is the closed-loop system table? (d) Use the results of part (b) to determine $c(t)_{ss}$.

6.11. (a) Determine the open-loop transfer function $G(s)H(s)$ of Fig. (a). (b) Determine the overall transfer function. (c) Figure (b) is an equivalent block diagram of Fig. (a). What must the transfer function $H_x(s)$ be in order for Fig. (b) to be equivalent to Fig. (a)? (d) Figure (b) represents what type of system? (e) Determine the system error coefficients of Fig. (b). (f) If $r(t)=u_{-1}(t)$, determine the final value of $c(t)$. (g) What are the values of $e(t)_{ss}$ and $m(t)_{ss}$?



(a) (b)

www.4electron.com

6.12. Find the step-, ramp-, and parabolic-error coefficients for unity-feedback systems that have the following forward transfer functions:

$$(a)\quad G(s) = \frac{20}{(0.2s+1)(0.5s+1)}$$

$$(b)\quad G(s) = \frac{200}{s^2(s^2+4s+13)(s^2+6s+25)}$$

$$(c)\quad G(s) = \frac{24(s+2)}{s(s^2+4s+6)}$$

$$(d)\quad G(s) = \frac{48(s+3)}{s(s+6)(s^2+4s+4)}$$

$$(e)\quad G(s) = \frac{25(s+3)}{(s+5)(s^2+4s+24)}$$

6.13. For Prob. 6.12, find $e(\infty)$ by use of the error coefficients, with the following inputs:

(a) $r(t)=5$ (b) $r(t)=2t$ (c) $r(t)=t^2$

6.14. A unity-feedback control system has

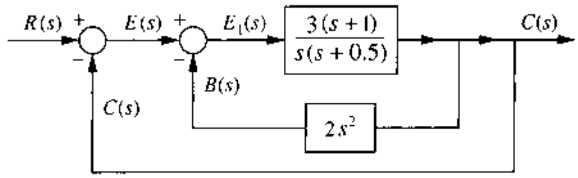
$$(1)\quad G(s) = \frac{12K}{s[(s+2)(s+4)+10]}$$

$$(2)\quad G(s) = \frac{K(s+3)}{s(s+1)(s+6)(s+9)}$$

where $r(t)=2t$. (a) If $K=2$, determine $e(t)_{ss}$. (b) It is desired that for a ramp input $e(t)_{ss} \le 1$. What minimum value must K_1 have for the condition to be satisfied? (c) For the value of K_1 determined in part (b), is the system stable?

6.15. (a) Derive the ratio $G(s)=C(s)/E(s)$. (b) Based upon $G(s)$, the figure has the characteristic of what type of system? (c) Derive the control ratio for this control system. (d) Repeat with the minor loop feedback replaced by K_h. (e) What conclusion can be reached about the effect of minor loop feedback on system type? (f) For (c) and (d)

determine $G_{eq}(s)$.



6.16. A unity-feedback system has the forward transfer function

$$G(s) = \frac{K(s+4)}{s(s+2)(s^2+4s+8)}$$

The input $r(t) = 6 + 8t$ is applied to the system. (a) It is desired that the steady-state value of the error be equal to or less than 1.6 for the given input function. Determine the minimum value that K_1 must have to satisfy this requirement. (b) Using Routh's stability criterion, determine whether the system is stable for the minimum value of K_1 determined in part (a).

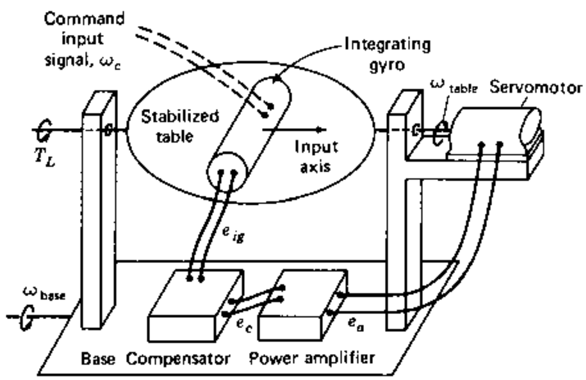
6.17. The angular position θ of a radar antenna is required to follow a command signal θ_c. The command signal e_c is proportional to θ_c with a proportionality constant K_θ V/deg. The positioning torque is applied by an ac motor that is activated by a position error signal e. The shaft of the electric motor is geared to the antenna with a gear ratio n. A potentiometer having a proportionality constant K_θ V/deg produces a feedback signal e_θ proportional to θ. The actuating signal $e = e_c - e_\theta$ is amplified with gain K_a V/V to produce the voltage e_α that drives the motor. The amplifier output e_a is an ac voltage. With the motor shaft clamped, the motor stall torque measured at the motor shaft is proportional to e_a with a proportionality constant K_c in 1b/V. The motor torque decreases linearly with increasing motor speed $\dot{\theta}_m$ and is zero at the no-load speed $\dot{\theta}_0$. The no-load speed is proportional to e_a with proportionality constant K_0. The moment of inertia of the motor, gearing, and antenna, referred to the output θ, is J and the viscous friction B is negligible. The motor torque is therefore $T = J\ddot{\theta} + T_L$, where T_L is a wind torque. (a) Draw a completely labeled detailed block diagram for this system. (b) Neglecting T_L, find the open-loop transfer function $\theta(s)/E(s)$ and the closed-loop transfer function $\theta(s)/\theta_c(s)$. (c) For a step input θ_c, describe the response characteristics as the gain K_a is increased. (d) For a constant wind torque T_L and $\theta_c(t) = 0$ find the steady-state position error as a function of K_a.

6.18. The components of a single-axis stable platform with direct drive can be described by the following equations:

Controlled platform: $T = JD\omega + B\omega + T_L$

Integrating gyro: $De_{ig} = K_{ig}(\omega_c - \omega + \omega_d + \omega_b)$

Servomotor: $T = \left(\dfrac{K_a}{K_{ig}}\right)e_{ig}$



where

J = moment of platform and servomotor
B = damping
ω = angular velocity of platform with respect to inertial frame of reference
T = servomotor torque
T_L = interfering torque
e_{ig} = gyro output voltage
ω_c = command angular velocity
ω_d = drift rate of gyro
ω_b = angular velocity of the base
K_{ig} = gyro gain
K_a/K_{ig} = amplifier gain

Consider that the base has a given orientation and that $\omega_b = 0$. (*a*) Show that the overall equation of performance for the stable platform is

$$(JD^2 + BD + K_a)\omega = K_a(\omega_c + \omega_d) - DT_L$$

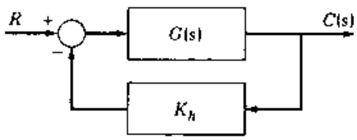
(*b*) Draw a detailed block diagram representing the system. (*c*) Using e_{ig} and ω as state variables, write the state equations of the system. (*d*) With $\omega_c = 0$ and $T_L = 0$ determine the effect of a constant draft rate ω_d.

(*e*) With $\omega_c = 0$ and $\omega_d = 0$, determine the effect of a constant interfering torque T_L on ω.

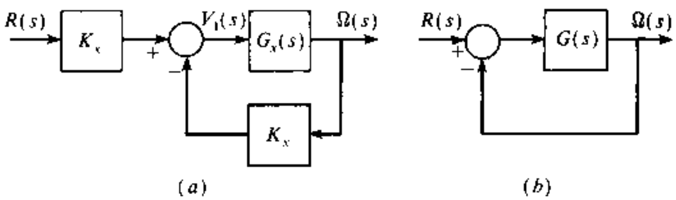
6.19. In the feedback system shown, K_h is adjustable and

$$G(s) = \frac{K_G(s^w + \cdots + c_0)}{s^n + \cdots + a_1 s + a_0}$$

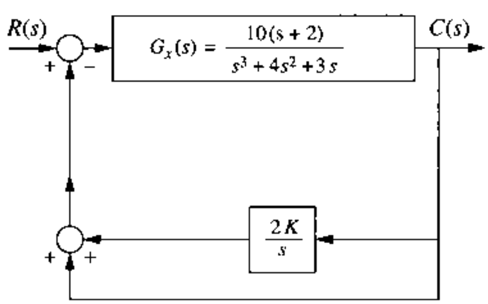
The output is required to follow a step input with no steady-state error. Determine the necessary conditions on K_h for a stable system.



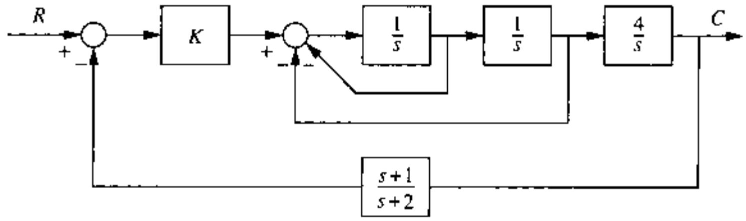
6.20. For Prob. 2.24 (*a*) derive the transfer function $G_x(s) = \Omega(s)/V_1(s)$. Figure (*a*) shows a velocity control system, where K_x is the velocity sensor coefficient. The equivalent unity-feedback system is shown in Fig. (*b*), where $\Omega(s)/R(s) = K_xG_x(s)/[1 + K_xG_x(s)]$. (*b*) What system type does $G(s) = K_xG_x(s)$ represent? (*c*) Determine $\omega(t)_{ss}$, when $r(t) = R_0u_{-1}(t)$. Assume the system is stable.



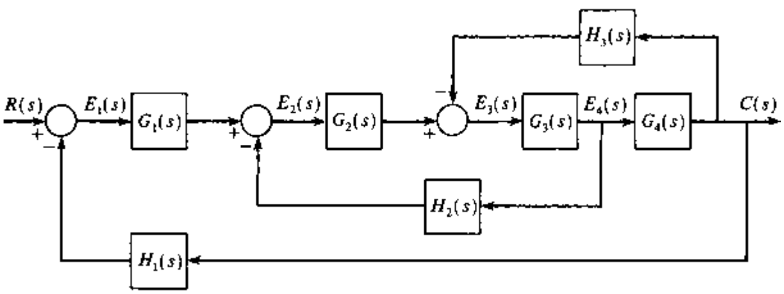
6.21. Given the following system find (*a*) $C(s)/R(s)$. (*b*) For each of the following values of K : 1 and 0.1, determine the poles of the control ratio and if the system is stable, (*c*) determine $G_{eq}(s)$, and (*d*) the equivalent system type.



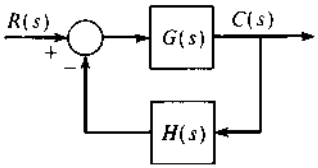
6.22. Given the following system, find $C(s)/R(s)$. (*a*) Find the range of values of K for which the system is stable. (*b*) Find the *positive* value of K that yields pure oscillations in the homogeneous system and find the frequency associated with these oscillations.



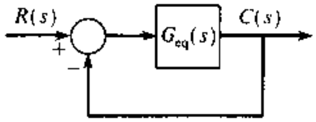
6.23. Based on this figure, do the following:



(*a*) By block diagram reduction find $G(s)$ and $H(s)$ such that the system can be represented in this form:

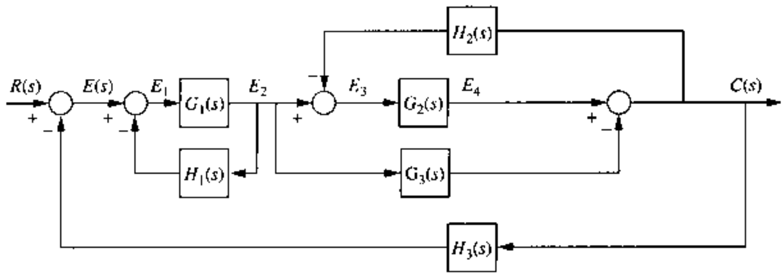


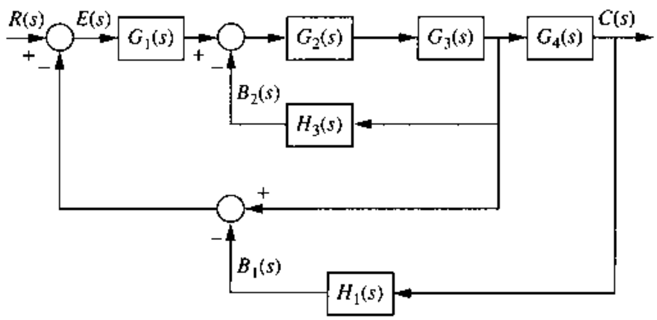
(*b*) For this system find $G_{eq}(s)$ for an equivalent unity-feedback system of this form:



www.4electron.com

6.24. By block diagram manipulations or the use of SFG, simplify the following systems to the form of Fig. 6.1.





6.25. A unity-feedback system has the forward transfer function

$$G(s) = \frac{20K}{s(s^2 + 10s + 14 + K)}$$

(a) What is the maximum value of K for stable operation of the system? (b) For K equal to one-half the maximum value, what is the $e(t)_{ss}$ for the ramp input $r(t) = 2tu_{-1}(t)$?

6.26. A unity-feedback system has the forward transfer function

$$G(s) = \frac{K(s + 30)}{(s + 1)(s^2 + 20s + 116)}$$

(a) What is the maximum value of K for stable operation of the system? (b) For a unit-step input, $e(t)_{ss}$ must be less than 0.1 for a stable system. Find K_m to satisfy this condition. (c) For the value of K_m of part (b), determine $c(t)$ and the figures of merit by use of a CAD package.

6.27. A unity-feedback system has the forward transfer function

$$G(s) = \frac{20K}{s(s^2 + 10s + 29)}$$

where the ramp input is $r(t) = 2tu_{-1}(t)$.
(*a*) For $K = 100$, assuming the system is stable, what is $e(t)_{ss}$? (*b*) For $K = 1000$ what is $e(t)_{ss}$? (*c*) Is the system stable for each value? If so, what are the figures of merit for a unit-step forcing function?

6.28. A unity-feedback system has the forward transfer function

$$G(s) = \frac{K(s + 10)}{(s + 1)(s^2 + 10s + 50)}$$

(*a*) For a unit-step input, $e(t)_{ss}$ must be less than 0.1 for a stable system. Find K_m to satisfy this condition. (*b*) For the value of K_m that satisfies this condition determine $c(t)$ and the figures of merit, for a unit-step forcing function, by use of a CAD package.

CHAPTER 7

7.1. Determine the pertinent geometrical properties and *sketch* the root-locus for the following transfer functions for both positive and negative values of K:

(*a*) $$G(s)H(s) = \frac{K}{s(s + 1)^2}$$

(*b*) $$G(s)H(s) = \frac{K(s - 2)}{(s + 1)(s^2 + 4s + 20)}$$

(*c*) $$G(s)H(s) = \frac{K(s + 4)}{s(s^2 + 4s + 29)}$$

(*d*) $$G(s)H(s) = \frac{K}{s^2(s + 1.5)(s + 5)}$$

(*e*) $$G(s)H(s) = \frac{K_0}{(1 + 0.5s)(1 + 0.2s)(1 + 0.125s)^2}$$

(*f*) $$G(s)H(s) = \frac{K(s + 3)^2}{s(s^2 + 4s + 8)}$$

(*g*) $$G(s)H(s) = \frac{K(s^2 + 6s + 34)}{(s + 1)(s^2 + 4s + 5)}$$

Determine the range of values of K for which the closed-loop system is stable.

7.2. **(CAD Problem)** Obtain the root locus of the following for a unity-feedback system with $K \geq 0$:

$$(a) \quad G(s) = \frac{K}{(1+s)(0.5+s)^2}$$

$$(b) \quad G(s) = \frac{K}{s(s^2+8s+15)(s+2)}$$

$$(c) \quad G(s) = \frac{K}{(s^2+4)(s^2+8s+20)}$$

$$(d) \quad G(s) = \frac{K(s+3)}{s^2(s^2+4s+13)}$$

$$(e) \quad G(s) = \frac{K}{s(s^2+2s+2)(s^2+6s+13)}$$

$$(f) \quad G(s) = \frac{K(s+15)}{(s+3)(s^2+12s+100)(s+15)}$$

$$(g) \quad G(s) = \frac{K(s+8)}{s(s^2+6s+25)}$$

Does the root cross any of the asymptotes?

7.3. A system has the following transfer functions:

$$G(s) = \frac{K(s+9)}{s(s^2+6s+13)} \qquad H(s) = 1 \qquad K > 0$$

(*a*) Plot the root locus. (*b*) A damping ratio of 0.6 is required for the dominant roots. Find $C(s)/R(s)$. The denominator should be in factored form. (*c*) With a unit step input, find $c(t)$, (*d*) Plot the magnitude and angle of $\mathbf{C}(j\omega)/\mathbf{R}(j\omega)$ vs. ω.

7.4. **(CAD Problem)** A feedback control system with unity feedback has a transfer function

$$G(s) = \frac{100K(s+10)}{s(s^2+10s+29)}$$

(*a*) Plot the locus of the roots of $1 + G(s) = 0$ as the loop sensitivity $K(>0)$ is varied. (*b*) The desired figures of merit for this system are $M_p = 1.0432$, $t_p \leq 18$ s, and $t_s \leq 2.27$ s. Based on these values, determine the dominant closed-loop poles that the control ratio must have in order to be able to achieve these figures of merit. Are these closed-loop poles achievable?

If so, determine the value of K. (*c*) For the value of K_1 found in part (*b*), determine $e(t)$ for $r(t) = u_{-1}(t)$ and determine the actual values of the figures of merit that are achieved. (*d*) Plot the magnitude and angle of $M(j\omega)$ vs. ω for the closed-loop system.

7.5. (*a*) Sketch the root locus for the control system having the following open-loop transfer function. (*b*) Calculate the value of K_1 that causes instability. (*c*) Determine $C(s)/R(s)$ for $\zeta = 0.3$. (*d*) Use a CAD package to determine $c(t)$ for $r(t) = u_{-1}(t)$.

$$(1) \quad G(s) = \frac{K_1(1 + 0.05s)}{s(1 + 0.0714s)(1 + 0.1s + 0.0125s^2)}$$

$$(2) \quad G(s) = \frac{K(s - 5)(s + 4)}{s(s + 1)(s + 3)}$$

7.6. (*a*) Sketch the root locus for a control system having the following forward and feedback transfer functions:

$$G(s) = \frac{K_2(1 + s/4)}{s^2(1 + s/15)} \qquad H(s) = K_H(1 + s/10)$$

(*b*) Express $C(s)/R(s)$ in terms of the unknown gains K_2 and K_H. (*c*) Determine the value of K_H what will yield $c(\infty) = 1$ for $r(t) = u_{-1}(t)$. (*d*) Choose closed-loop pole locations for a $\zeta = 0.6$. (*e*) Write the factored form of $C(s)/R(s)$. (*f*) Determine the figures of merit M_p, t_p, t_s and K_m.

7.7. A unity-feedback control system has the transfer function

$$G(s) = \frac{K}{(s + 1)(s + 7)(s^2 + 8s + 25)}$$

(*a*) Sketch the root locus for positive and negative values of K. (*b*) For what values of K does the system become unstable? (*c*) Determine the value of K for which all the roots are equal. (*d*) Determine the value of K from the root locus that makes the closed-loop system a perfect oscillator.

7.8. **(CAD Problem)** A unity-feedback control system has the transfer function

$$(1) \quad G(s) = \frac{K_1(1 - 0.5s)}{(1 + 0.5s)(1 + 0.2s)(1 + 0.1s)}$$

$$(2) \quad G(s) = \frac{K_1}{s(1 + 0.1s)(1 + 0.08s + 0.008s^2)}$$

(*a*) Obtain the root locus for positive and negative values of K_1. (*b*) What range of values of K_1 makes the system unstable?

7.9. **(CAD problem)** For positive values of gain, obtain the root locus for unity-feedback control systems having the following open-loop transfer functions. For what value or values of gain does the system become unstable in each case?

$$(a)\quad G(s) = \frac{K(s^2 + 8s + 25)}{s(s+1)(s+2)(s+5)}$$

$$(b)\quad G(s) = \frac{K(s^2 + 4s + 13)}{s^2(s+3)(s+6)(s+20)}$$

$$(c)\quad G(s) = \frac{K(s^2 + 2s + 5)}{s(s^2 - 2s + 2)(s+10)}$$

7.10. (CAD Problem) For each system of Prob. 7.9 a stable system with $M_p = 1.14$ is desired. If this value of M_p is achievable, determine the control ratio $C(s)/R(s)$. Select the *best* roots in each case.

7.11. A nonunity-feedback control system ha the transfer functions

$$G(s) = \frac{K_G(1 + s/3.2)}{(1 + s/5)[1 + 2(0.7)s/23 + (s/23)^2][1 + 2(0.49)s/7.6 + (s/76)^2]}$$

$$H(s) = \frac{K_H(1 + s/5)}{1 + 2(0.89)s/42.7 + (s/42.7)^2}$$

(*a*) Sketch the root locus for the system, using as few trial points as possible. Determine the angles and locations of the asymptotes and the angles of departure of the branches from the open-loop poles. (*b*) Determine the gain $K_G K_H$ at which the system becomes unstable, and determine the approximate locations of all the closed-loop poles for this value of gain. (*c*) Using the closed-loop configuration determined in part (*b*), write the expression for the closed-loop transfer function of the system.

7.12. (CAD Problem) A nonunity-feedback control system has the transfer functions

$$(1)\quad G(s) = \frac{10A(s^2 + 8s + 20)}{s(s+4)} \qquad H(s) = \frac{0.2}{s+2}$$

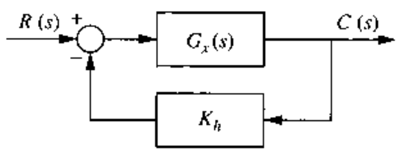
$$(2)\quad G(s) = \frac{K(s^2 + 6s + 13)}{s(s+3)} \qquad H(s) = \frac{1}{s+1}$$

For system (1): (*a*) determine the value of the amplifier gain A that will produce complex roots having the *minimum* possible value of ζ. (*b*) Express $C(s)/R(s)$ in terms of its poles, zeros, and constant term.

For system (2) an $M_p = 1.0948$ is specified: (*a*) determine the values of K that will produce complex roots for a ζ corresponding to this value of M_p. Note that there are two intersections of the ζ line. (*b*) Express $C(s)/R(s)$ in terms of its poles, zeros, and constant term for each intersection. (*c*) For each control ratio determine the figures of merit. Compare these values. Note the differences, even though the pair of complex poles in each case has the same value of ζ.

7.13. For the figure shown, only the forward and feedback gains, K_x and K_h, are adjustable. (*a*) Express $C(s)/R(s)$ in terms of the poles and zeros of $G_x(s)$ and $H(s)$. (*b*) Determine the value K_h must have in order that $c(\infty) = 1.0$ for a unit step input. (*c*) Determine the value of $K_x K_h$ that will produce complex roots having a $\zeta = 0.6$. (*d*) Determine the corresponding $C(s)/R(s)$. (*e*) Determine $G_{eq}(s)$ for a unity feedback system that is equivalent to this nonunity feedback system. (*f*) Determine the figures of merit M_p, t_p, t_s, and K_m.

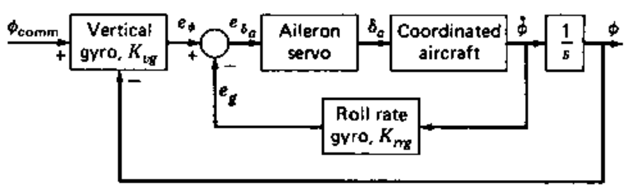
$$G_x = \frac{K_x(s+1)}{s^2(s+4)} \qquad H_s = \frac{K_h(s+2)}{s+3}$$



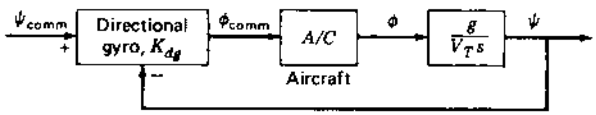
7.14. Yaw rate feedback and sideslip feedback have been added to an aircraft for turn coordination. The transfer functions of the resulting aircraft and the aileron servo are, respectively,

$$\frac{\dot{\phi} s}{\delta_a(s)} = \frac{16}{s+3} \qquad \text{and} \qquad \frac{\delta_a(s)}{e_{\delta_a}(s)} = \frac{20}{s+20}$$

For the bank-angle control system shown in the figure, design a suitable system by drawing the root-locus plots on which your design is based. (*a*) Use $\zeta = 0.8$ for the inner loop and determine K_{rrg}. (*b*) Use $\zeta = 0.7$ for the outer loop and determine K_{vg}. (*c*) Determine $\phi(s)/\phi_{\text{comm}}(s)$.



7.15. A heading control system using the bank-angle control system is shown in the following figure, where g is the acceleration of gravity and V_T is the true airspeed. Use the design of Prob. 7.14 for $\phi(s)/\phi_{comm}(s)$. (*a*) Draw a root locus for this system and use $\zeta = 0.6$ to determine K_{dg}. (*b*) Determine $\psi(s)/\psi_{comm}(s)$ and the time response with a step input for $V_T = 500$ ft/s. (The design can be repeated using $\zeta = 1.0$ for obtaining K_{rrg}.)



7.16. **(CAD Problem)** The simplified open-loop transfer function for a T-38 aircraft at 35,000 ft and 0.9 Mach is

$$G(s) = \frac{\theta(s)}{\delta_e(s)} = \frac{22.3K_x(0.995s + 1)}{(s + 45)(s^2 + 1.76s + 7.93)}$$

The open-loop transfer function contains the simplified longitudinal aircraft dynamics (a second-order system relating pitch rate to elevator deflection). The $K_x/(s + 45)$ term models the hydraulic system between the pilot's control and the elevator. Obtain a plot of the root locus for this system. Select characteristic roots that have a damping ratio of 0.42. Find the gain value K_x that yields these roots. Determine the system response to an impulse. What is the minimum value of damping ratio available at this flight condition?

CHAPTER 8

8.1. For each of the transfer functions

$$(1) \quad G(s) = \frac{1}{s(1 + 0.5s)(1 + 4s)}$$

$$(2) \quad G(s) = \frac{0.5}{s(1 + 0.2s)(1 + 0.004s + 0.0025s^2)}$$

$$(3) \quad G(s) = \frac{100(s + 2)}{s(s^2 + 6s + 25)}$$

$$(4) \quad G(s) = \frac{10(1 - 0.5s)}{s(1 + s)(1 + 0.5s)}$$

(*a*) draw the log magnitude (exact and asymptotic) and phase diagrams; (*b*) determine ω_ϕ and ω_c for each transfer function.

8.2. For each of the transfer functions

$$(1)\quad G(s)=\frac{2(1+0.25s)}{s^2(1+s\sqrt{60})(1+0.04s)} \qquad (2)\quad G(s)=\frac{5}{(1+5s/3)(1+1.25s)^2}$$

$$(3)\quad G(s)=\frac{0.2}{s^2(1+2s/9)(1+s/15)} \qquad (4)\quad G(s)=\frac{10(1+0.125s)}{s(1+0.5s)(1+0.25s)}$$

(*a*) Draw the log magnitude (exact and asymptotic) and phase diagrams.

8.3. **(CAD Problem)** Plot to scale the log magnitude and angle vs. log ω curves for $\mathbf{G}(j\omega)$. Is it an integral or a derivative compensating network?

$$(a)\quad \mathbf{G}(j\omega)=\frac{20(1+j8\omega)}{1+j0.2\omega} \qquad (b)\quad \mathbf{G}(j\omega)=\frac{20(1+j0.125\omega)}{1+j10\omega}$$

8.4. A control system with unity feedback has the forward transfer function, with $K_m=1$,

$$(1)\quad G(s)=\frac{K_1(1+0.4s)}{s(1+0.1s)(1+8s/150\sqrt{2}+s^2/900)}$$

$$(2)\quad G(s)=\frac{K_0(1+0.85)}{(1+s)(1+0.4s)(1+0.2s)^2}$$

(*a*) Draw the log magnitude and angle diagrams. (*b*) Draw the polar plot of $G'(j\omega)$. Determine all key points of the curve.

8.5. A system has

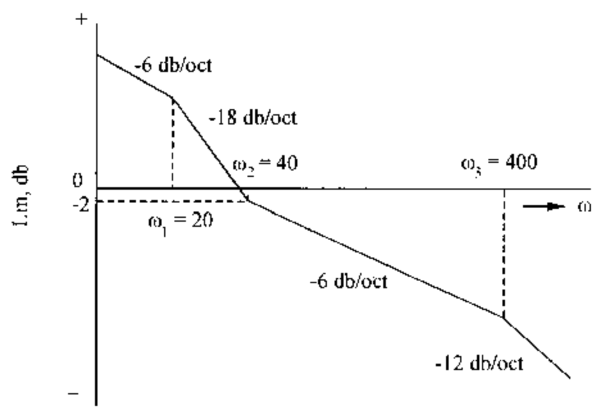
$$G(s)=\frac{8(1+T_2s)(1+T_3s)}{s^2(1+T_1s)^2(1+T_4s)}$$

where $T_1=1$, $T_2=1/2$, $T_3=1/8$, $T_4=1/16$. (*a*) Draw the asymptotes of Lm $\mathbf{G}(j\omega)$ on a decibel vs. log ω plot. Label the corner frequencies on the graph. (*b*) What is the *total* correction from the asymptotes at $\omega=2$?

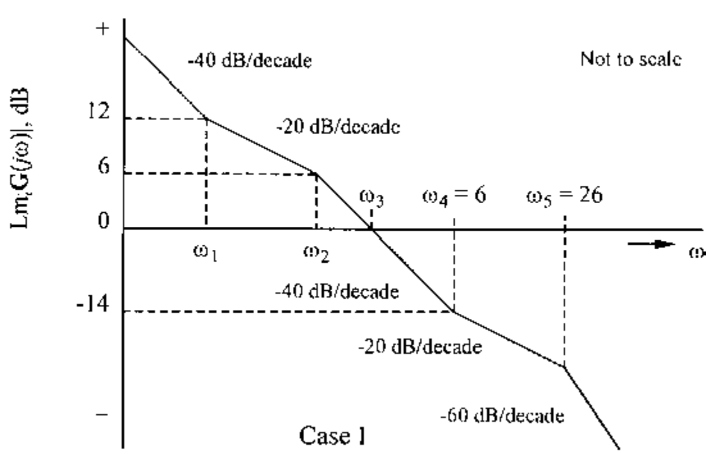
8.6. (*a*) What characteristic must the plot of magnitude in decibels vs. log ω possess if a velocity servo system (ramp input) is to have no *steady-state velocity error* for a constant velocity input $dr(t)/dt$? (*b*) What is true of the corresponding phase-angle characteristic?

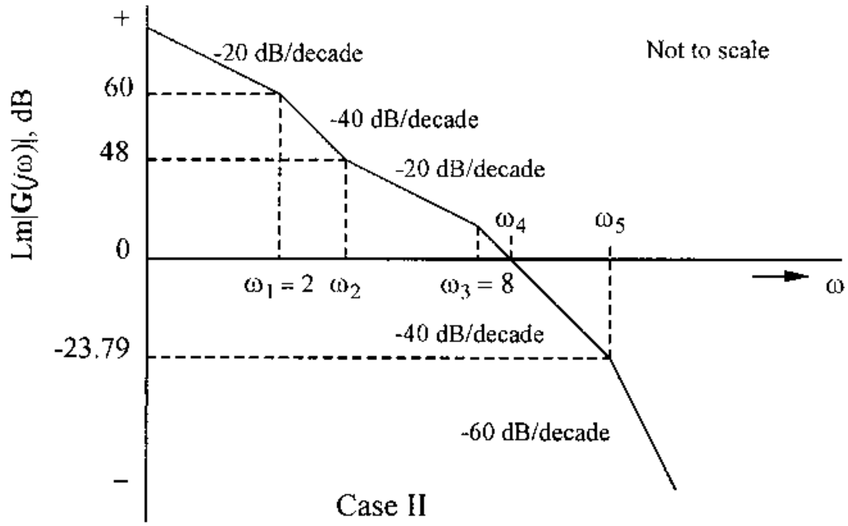
8.7. Explain why the phase-angle curve cannot be calculated from the plot of $|\mathbf{G}(j\omega)|$ in decibels vs. log ω if some of the factors are not minimum phase.

8.8. The asymptotic gain vs. frequency curve of the open-loop minimum-phase transfer function is shown for a unity-feedback control system. (*a*) Evaluate the open-loop transfer function. Assume only first-order factors. (*b*) What is the frequency at which $|G(j\omega)|$ is unity? What is the phase angle at this frequency? (*c*) Draw the polar diagram of the open-loop control system.

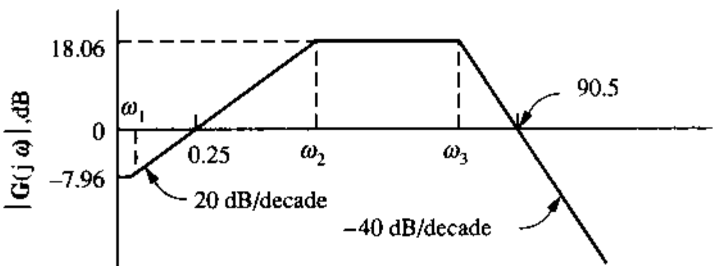


8.9. For each plot shown, (*a*) evaluate the transfer function; (*b*) find the correction that should be applied to the straight-line curve at $\omega = 4$; (*c*) determine K_m for Cases I and II. Assume only first-order factors.



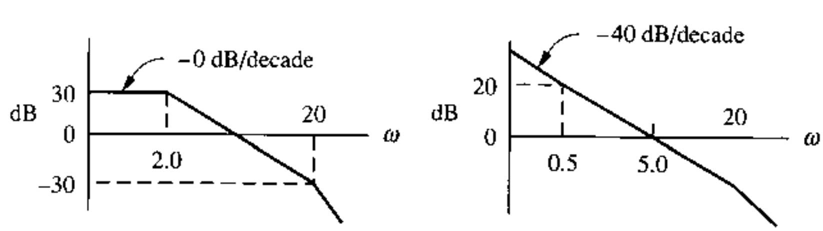


Case II



Case III

8.10. **Determine the value of the error efficient from parts (*a*) and (*b*) of the figure.**



(*a*) (*b*)

www.4electron.com

8.11. An experimental transfer function gave, respectively, the following results:

Plant 1

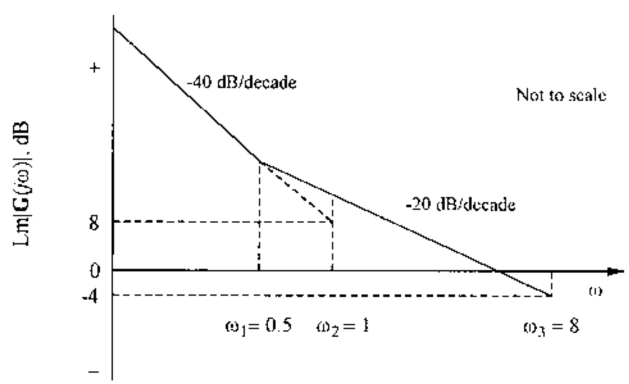
ω	Lm $G(j\omega)$, dB	Angle, deg
0.20	−18.57	−91.4
0.60	−27.97	−94.1
1.0	−32.11	−97.1
2.0	−36.65	−107.1
3.0	−37.34	−126.9
4.0	−38.13	−172.9
4.2	−38.93	−184.4
4.4	−40.02	−195.0
4.8	−42.68	−212.2
5.0	−44.09	−218.7
5.4	−46.83	−228.4
6.0	−50.55	−237.7
7.0	−55.72	−246.4
9.0	−63.50	−254.3
14.0	−76.04	−261.1
20.0	−85.69	−264.0
30.0	−96.44	−266.1
49.0	−109.74	−267.7

Plant 2

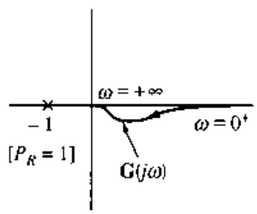
ω	Lm $G(j\omega)$, dB	Angle, deg
0.01	18.13	−0.247
0.02	18.13	−0.491
0.20	18.09	−4.893
0.50	17.89	−11.99
1.000	17.25	−22.52
2.000	15.49	−37.18
4.000	12.54	−49.65
6.000	11.05	−54.89
7.000	10.73	−57.29
8.000	10.62	−60.11
16.000	11.15	−123.9
17.000	10.32	−135.3
18.000	9.26	−145.3
19.000	8.06	−153.8
20.000	6.82	−160.8
40.000	−10.01	−205.3
50.000	−15.03	−213.6
55.000	−17.18	−217.0

(*a*) Determine the transfer function represented by the above data. (*b*) What type of system does it represent?

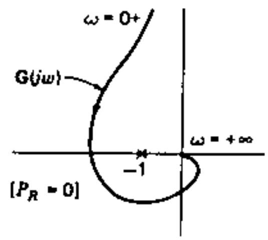
8.12. Determine the transfer function by using the straight-line asymptotic log plot shown and the fact that the correct angle is -93.576° at $\omega = 8$. Assume a minimum-phase transfer function.



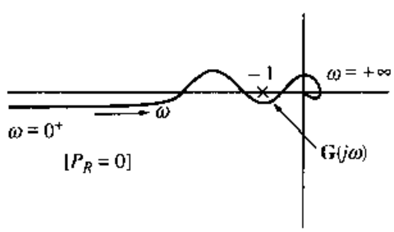
8.13. Determine whether each system shown is stable or unstable in the absolute sense by sketching the *complete* Nyquist diagrams. $H(s) = 1$.



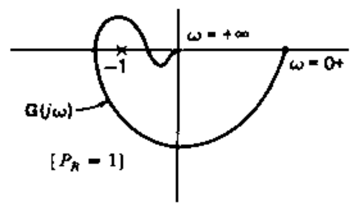
(*a*)



(*b*)



(*c*)



(*d*)

8.14. Use the Nyquist stability criterion and the polar plot to determine, with the aid of a CAD package, the range of values K (positive or negative) for which the closed-loop system is stable.

(*a*) $$G(s)H(s) = \frac{K(s+1)^2(s+2)}{s^3(s+10)}$$

(*b*) $$G(s)H(s) = \frac{K}{s^2(-1+5s)(1+s)}$$

(*c*) $$G(s)H(s) = \frac{K}{s^2(1-0.5s)}$$

(*d*) $$G(s)H(s) = \frac{K}{s^2(s+15)(s^2+6s+10)}$$

(*e*) $$G(s)H(s) = \frac{K}{s(s^2+4s+5)}$$

(*f*) $$G(s)H(s) = \frac{K}{s^2(s+9)}$$

8.15. For the following transfer functions, sketch a direct Nyquist locus to determine the closed-loop stability. Determine, with the aid of CAD package, the range of values of K that produce stable closed-loop operation and those which produce unstable closed-loop operation. $H(s) = 1$.

$$(a) \quad G(s) = \frac{K}{2 + s}$$

$$(b) \quad G(s) = \frac{K(2 + s)}{s(1 - s)}$$

$$(c) \quad G(s) = \frac{K(1 + 0.5s)}{s^2(8 + s)}$$

$$(d) \quad G(s) = \frac{K}{s(7 + s)(2 + s)}$$

$$(e) \quad G(s) = \frac{K}{(s + 3)(s - 1)(s + 6)}$$

8.16. **(CAD Problem)** For the control systems having the transfer functions

$$(1) \quad G(s) = \frac{K_1}{s(1 + 0.02s)(1 + 0.05s)(1 + 0.10s)}$$

$$(2) \quad G(s) = \frac{K(1 + 0.25s)}{(1 + 0.1s)(2 + 3s + s^2)}$$

$$(3) \quad G(s) = \frac{K(s + 3)(s + 40)}{s(s^2 + 20s + 1000)(s + 80)(s + 100)}$$

$$(4) \quad (G(s) = \frac{K(s + 4)(s + 8)}{(s + 1)(s^2 + 4s + 5)(s + 10)}$$

determine from the logarithmic curves the required value of K_m and the phase-margin frequency so that each system will have (a) a positive phase margin angle of 45°; (b) a positive phase margin angle of 60°. (c) From these curves determine the maximum permissible value of K_m for stability.

8.17. A system has the transfer functions

$$G(s) = \frac{K(s + 8)}{s(s + 4)(s^2 + 16s + 164)(s + 40)} \qquad H(s) = 1$$

(a) Draw the log magnitude and phase diagram of $\mathbf{G}'(j\omega)$. Draw both the straight-line and the corrected log magnitude curves. (b) Draw the log magnitude-angle diagram. (c) Determine the maximum value of K_1 for stability.

8.18. What gain values would just make the systems in Prob. 8.2 unstable?

8.19. For the following plant, where $K_x = 3.72$, determine the phase margin angle γ and the phase margin frequency ω_x.

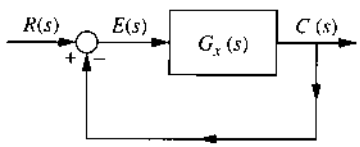
$$G_x(s) = \frac{K_x(1 + s/4)}{s(1 + s)(1 + s/8)}$$

CHAPTER 9

9.1. For Prob. 8.1, find K_1, M_m, ω_c, ω_ϕ, and ω_m for a $\gamma = 45^\circ$ by (a) The direct-polar-plot method; (b) The log magnitude–angle diagram; (c) A computer program. (d) Repeat (a) through (c) by determining the value of K_1 for an $M_m = 1.26$. (e) Compare the values of ω_m, ω_c, ω_ϕ, and K_1 obtained in (d) with those obtained in (a) and (b).

9.2. **(CAD Problem)** For the feedback control system of Prob. 7.4: (a) Determine, by use of the polar-plot method, the value of K_1 that just makes the system unstable. (b) Determine the value of K_1 that makes $M_m = 1.16$. (c) For the value of K_1 found in part (b), find $c(t)$ for $r(t) = u_{-1}(t)$. (d) Obtain the data for plotting the curve of M vs. ω for the closed-loop system and plot this curve. (e) Why do the results of this problem differ from those obtained in Prob. 7.4? [See Eq. (9.16).]

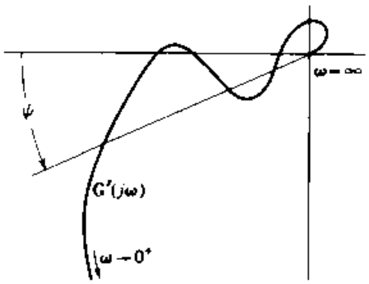
9.3. **(CAD Problem)** Determine the value K_1 must have for an $M_m = 1.3$.



$$G_x(s) = \frac{K_x}{s(1 + 0.2s)(1 + 0.125s)}$$

9.4. Using the plot of $\mathbf{G}'(j\omega)$ shown, determine the number of values of gain K_m which produce the same value of M_m. Which of these values yields the best system performance? Give the reasons for your answer.

$$G(s) = \frac{K_1(1 + 0.1s)^2}{s(1 + s)^2(1 + s/150)^3}$$



9.5. **(CAD Problem)** (*a*) Determine the values of M_m and ω_m for each of the transfer functions of Prob. 8.16 with $K_m = 2$. (*b*) Repeat with the gain-constant values obtained in part (*a*) of Prob. 8.16. (*c*) Repeat with the gain constant value in part (*b*) of Prob. 8.16. (*d*) For part (*c*), plot M vs. ω and obtain $c(t)$ for a step input.

9.6. **(CAD Problem)** (*a*) In Prob. 8.17, adjust the gain for Lm $M_m = 2$ dB and determine ω_m. For this value of gain find the phase margin angle, and plot M vs. ω and α vs. ω. (*b*) Repeat for Prob. 8.4.

9.7. For

$$G_x(s) = \frac{K}{s(s+8)(s+16)} \qquad \text{and} \qquad H(s) = 1$$

determine the required gain to achieve an $M_m = 1.26$. What are the values of ω_m, ω_ϕ, and γ for this value of M_m? (*a*) Use log plots. (*b*) Use polar plots. (*c*) Use a computer.

9.8. (*a*) For Prob. 8.2, determine, by use of the Nichols chart, the values of M_m and ω_m. (*b*) What must the gain be in order to achieve an $M_m = 1.12$? (*c*) Refine the values of parts (*a*) and (*b*) by use of a CAD package. (*d*) What is the value of ω_m and the phase margin angle for this value of M_m for each case?

9.9. **(CAD Problem)** Refer to Prob. 8.1.
(*a*) Determine the values of K_1, ω_m, and γ corresponding to the following values of M_m: 1.05, 1.1, 1.2, 1.4, 1.6, 1.8, and 2.0. For each value of M_m, determine ζ_{eff} from Eq. (9.16) and plot M_m vs. ζ_{eff}. (*b*) For each value of K_1 determined in part (*a*), calculate the value of M_p. For each value of M_p, determine ζ_{eff} from Fig. 3.7 and plot M_p vs. ζ_{eff}. (*c*) What is the correlation between M_m vs. ζ_{eff} and M_p vs. ζ_{eff}? What is the effect of a third real root that is also dominant? (*d*) What is the correlation between γ and both M_m and M_p?

9.10. For the given transfer function: (*a*) With the aid of the Nichols chart, specify the value of K_2 that will make the peak M_m in the frequency response as small as possible. (*b*) At what frequency does this peak occur? (*c*) What value does $\mathbf{C}(j\omega)/\mathbf{R}(j\omega)$ have at the peak?

$$(1) \quad G(s) = \frac{K_2(1 + 2.5s)}{s^2(1 + 0.125s)(1 + 0.04s)}$$

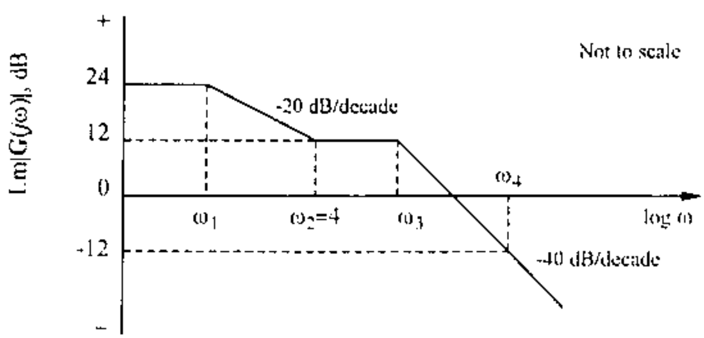
$$(2) \quad G(s) = \frac{K(s + 0.08)}{s^2(s + 2)}$$

9.11. By use of Nichols chart for

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

(*a*) What is the maximum value of the unity-feedback closed-loop frequency response magnitude M_m with $K = 1$? (*b*) Determine the value of K for a phase margin angle of $\gamma = 45^\circ$. (*c*) What is the corresponding gain margin *a* for this value of K?

9.12. Evaluate the transfer function for the frequency-response plot shown below. The plot is not drawn to scale.



CHAPTER 10

For the design of all problems of Chap. 10, obtain the figures of merit (M_p, t_p, t_s, and K_m) for the uncompensated and compensated systems. Where appropriate, use a CAD package in solving these problems.

10.1. It is desired that a control system have a damping ratio of 0.5 for the dominant complex roots. Using the root-locus method, (*a*) add a lag compensator, with $\alpha = 10$, so that this value of ζ can be obtained; (*b*) add a lead compensator with $\alpha = 0.1$; (*c*) add a lag-lead compensator with $\alpha = 10$. Indicate the time constants of the compensator in each case. Obtain $c(t)$ with a step input in each case. Compare the results obtained by the use of each type of compensator with respect to the error coefficient K_m, ω_d, T_s, and M_o.

$$(1) \quad G(s) = \frac{K}{(s+1)(s+3)(s+6)}$$

$$(2) \quad G(s) = \frac{K}{s(s+2)(s+5)(s+6)}$$

Note: For plant (1), the 5° rule for a lag compensator does not apply. By trial and error determine the best location of the compensator's zero and pole.

10.2. A control system has the forward transfer function

$$G_x(s) = \frac{K_2}{s^2(1+0.05s)}$$

The closed-loop system is to be made stable by adding a compensator $G_c(s)$ and an amplifier A in cascade with $G_x(s)$. A ζ of 0.6 is desired with a value of $\omega_n \approx 3.75$ rad/s. By use of the root-locus method, determine the following: (*a*) What kind of compensator is needed? (*b*) Select an appropriate α and T for the compensator. (*c*) Determine the value of the error coefficient K_2. (*d*) Plot the compensated root locus. (*e*) Plot M vs. ω for the compensated closed-loop system. (*f*) From the plot of part (*e*) determine the values of M_m and ω_m. With this value of M_m determine the effective ζ of the system by the use of $M_m = (2\zeta\sqrt{1-\zeta^2})^{-1}$. Compare the effective ζ and ω_m with the values obtained from the dominant pair of complex roots. *Note*: The effective $\omega_m = \omega_n\sqrt{1 = 2\zeta^2}$. (*g*) Obtain $c(t)$ for a unit step input.

10.3. For a unity feedback system whose basic plant is

$$(1) \quad G_x = \frac{K_x}{s(s^2+8s+20)}$$

$$(2) \quad G_x = \frac{K_x(s+8)}{s(s+3)(s^2+10s+50)}$$

Use $\zeta = 0.5$ for the dominant roots. Determine the figures of merit M_p, t_p, t_s, and K_m for each of the following cases: (*a*) original system; (*b*) lag compensator added, $\alpha = 10$; (*c*) lead compensator added, $\alpha = 0.1$; (*d*) lag-lead compensator added, $\alpha = 10$. *Note*: Except for the original system, it is not necessary to obtain the complete root locus for each type of compensation. When the lead and lag-lead compensator are added to the original system, there may be other dominant roots in addition to the complex pair. When a real root is also dominant, a $\zeta = 0.3$ may produce the desired improvement (see Sec. 10.3).

10.4. A unity-feedback system has the forward transfer function

$$G_x(s) = \frac{K_2}{s^2}$$

(*a*) Design a cascade compensator that will produce a stable system and that meets the following requirements without reducing the system type: (1) The dominant poles of the closed-loop control ratio are to have a damping ratio $\zeta = 0.6$. (2) The settling time is to be $T_s = 3.5$s. Is it a lag or lead compensator? (*b*) Using the cascade compensator, determine the control ratio $C(s)/R(s)$. (*c*) Find $c(t)$ for a step input. What is the effect of the real pole of $C(s)/R(s)$ on the transient response?

10.5. Using the root-locus plot of Prob. 7.7, adjust the damping ratio to $\zeta = 0.5$ for the dominant roots of the system. Find K_0, ω_n, M_o, T_p, T_s, N, $C(s)/R(s)$ for (*a*) the original system and (*b*) the original system with the cascade lag compensator using $\alpha = 10$; (*c*) design a cascade compensator that will improve the response time, i.e., will move the dominant branch to the left in the s plane.

10.6. A unity-feedback system has the transfer function $G_x(s)$. The closed-loop roots must satisfy the specifications $\zeta = 0.707$ and $T_s = 2$ s. A suggested compensator $G_c(s)$ must maintain the same degree as the characteristic equation for the basic system:

$$G_x(s) = \frac{K(s+6)}{s(s+4)} \qquad G_c(s) = \frac{A(s+a)}{s+b}$$

(*a*) Determine the values of a and b. (*b*) Determine the value of α. (*c*) Is this a lag or a lead compensator?

10.7. A unity-feedback control system contains the forward transfer function shown below. The system specifications are $\omega_n = 4$ and $T_s \leq 2$ s, which are to be achieved by the proposed cascade compensator which has the form indicated.

$$G_x = \frac{K(s+6)}{s(s+2)(s+5)} \qquad G_c = A\frac{s+a}{s+b}$$

(*a*) For the uncompensated system, based upon the given performance specifications, determine the value of ζ which yields the desired dominant complex-conjugate poles. For this value of ζ determine the value of K and the corresponding figures of merit. (*b*) Determine the values of a and b such that the desired complex-conjugate poles of $C(s)/R(s)$, for the compensated system, are truly dominant and the degree of the compensated system is 3. (*c*) Determine the values of A and α for this compensator. (*d*) Is this $G_c(s)$ a lag or lead compensator? (*e*) Determine the control ratio and the figures of merit for the compensated system.

10.8. For the transfer function

$$G_x(s) = \frac{K_x}{(s+2)(s+4)(s+6)(s+8)}$$

(*a*) sketch the root locus. (*b*) For $\zeta = 0.6$ the system's performance specifications are $T_s \leq 2.5$ s and $K_0 \geq 0.6$. Determine the roots and the value of K_x for $\zeta = 0.6$. (*c*) Add a lead compensator that cancels the pole at $s = -1$. (*d*) Add a lead compensator that cancels the pole at $s = -2$. Determine K. (*e*) Compare the results of parts (*c*) and (*d*). Establish a "rule" for adding a lead compensator to a Type 0 system.

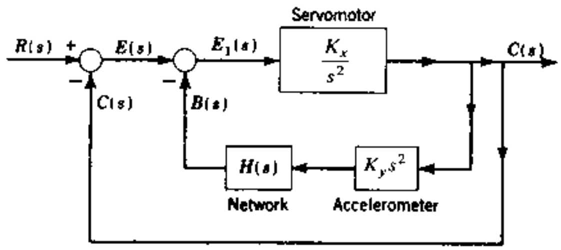
10.9. For the system of Prob. 7.4, the desired dominant roots are $-2.5 \pm j4.5$. (*a*) Design a compensator that will achieve these characteristics. (*b*) Determine $c(t)$ with a unit step input. (*c*) Are the desired complex roots dominant? (*d*) Compare the results of the basic system for the same value of ζ with those of the compensated system.

10.10. A unity-feedback system has the transfer function

$$G_x(s) = \frac{K}{s(s+3)(s+10)}$$

The dominant poles of the closed-loop system must be $s = -1.5 \pm j2$. (*a*) Design a lead compensator with the maximum possible value of α that will produce these roots. (*b*) Determine the control ratio for the compensated system. (*c*) Add a compensator to increase the gain without increasing the settling time.

10.11. In the figure the servomotor has inertia but no viscous friction. The feedback through the accelerometer is proportional to the acceleration of the output shaft. (*a*) When and $H(s) = 1$, is the servo system stable? (*b*) Is the system stable if $K_x = 15$, $K_y = 2$, and $H(s) = 1/s$? (*c*) Use $H(s) = 1/(1 + Ts)$ and show that it produces a stable system. Show that the system is stable with this value of $H(s)$ by obtaining the transfer function $C(j\omega)/E(j\omega)$ and sketching the root locus.



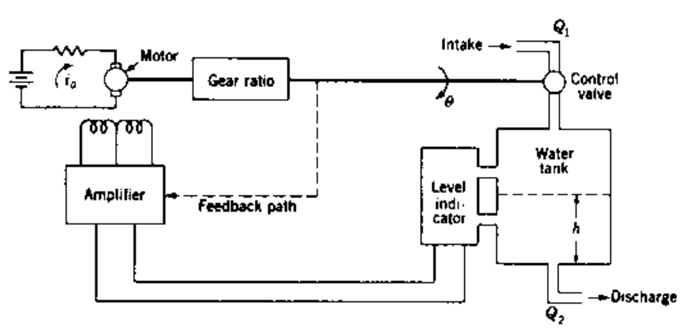
10.12. The accompanying schematic shows a method for maintaining a constant rate of discharge from a water tank by regulating the level of the water in the tank. The relationship governing the dynamics of the flow into and out of the tank are:

$$Q_1 - Q_2 = 16\frac{dh}{dt} \qquad Q_2 = 4h \qquad Q_1 = 10\theta$$

where

$Q_1 =$ volumetric flow into tank
$Q_2 =$ volumetric flow out of tank
$h =$ pressure head in tank
$\theta =$ angular rotation of control valve

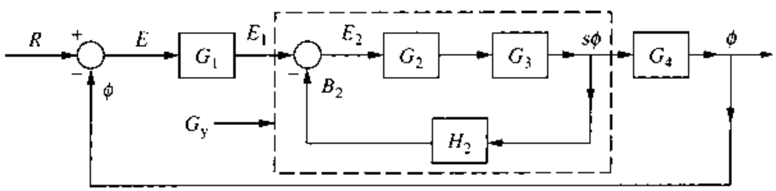
The motor damping is much smaller than the inertia. The system shown is stable only for very small values of system gain. Therefore, feedback from the control valve to the amplifier is proposed. Show conclusively which of the following feedback functions would produce the best results: (*a*) feedback signal proportional to control-valve position; (*b*) feedback signal proportional to rate of change of control-valve position; (*c*) feedback signal comprising a component proportional to valve position and a component proportional to rate of change of valve position.



www.4electron.com

For Probs. 10.13 and 10.14, the block diagram shows a simplified form of roll control for an airplane. Overall system specifications with a step input are $t_s \leq 1.2$ s and $1 < M_p \leq 1.3$. A_h is the gain of an amplifier in the H_2 feedback loop. $G_2 = A_2$ is an amplifier of adjustable gain with a maximum value of 100. Restrict b to values between 15 and 50. With $G_1(s) = 1$, derive general expressions for $G_y(s)$ and $\Phi(s)/R(s)$ before solving these problems.

$$G_3(s) = \frac{1.8}{(1 + s/6)(1 + s)} \qquad H_2(s) = 0.2A_h\frac{s+a}{s+b} \qquad G_4(s) = \frac{1}{s}$$



10.13. (*a*) By use of the root locus method 1, with $G_1(s) = 1$, determine a set of parameters A_h, a, and b in $H_2(s)$ to meet the overall system specifications. [*Note*: Problem 10.15 requires the design of the cascade compensator $G_1(s)$ to improve K_1 by a factor of 5. When a lag compensator is used to meet this requirement, it is known that it results in an increase in the values of M_p and t_s. Therefore, modify the specifications for this problem to $M_p < 1.2$ and $1\text{ s} < t_s < 1.1$ s in order to allow for the degradation in transient performance produced by this compensator.] (*b*) Determine $G_y(s) = s\Phi(s)/E_1(s)$ and $G(s) = A_1 G_y(s) G_4(s)$ in factored form. Then with $A_1 = 1$ compute K_1. (*c*) Determine the figures of merit (M_p, t_p, t_s, and K_1) for the closed-loop system. (*d*) With $e_1(t) = u_1(t)$, determine the values of M_p and t_s for the inner loop represented by $G_y(s)$.

10.14. (*a*) By use of the root locus method 2, determine the parameters, A_h, a, and b to meet the inner-loop specifications $M_p = 1.05$ and $T_s = 1.2$ s. [The same note given in Prob. 10.13 (*a*) also applies to this problem.] Select A_2 so that the steady-state value of $\dot{\phi}(t)$ follows a step input $e_1(t) = u_1(t)$. Determine the necessary values of A_h and A_2. (*b*) Determine $G_y(s)$ in factored form. (*c*) Determine the figures of merit for $G_y(s)$ and compare them with those of Prob. 10.13(*d*).

10.15. (*a*) Using the values determined in Prob. 10.13, derive the expression $G_z(s) = G_y(s) G_4(s)$. (*b*) Design $G_1(s)$ to increase the value of K_1 of Prob. 10.13 by a factor of 5 while maintaining the desired overall system specifications. (*c*) Determine the resulting $\Phi(s)/R(s)$ in factored form.

(d) Determining the figures of merit for a step input, and compare them with those of Prob. 10.13.

10.16. Repeat Prob. 10.15 with the values determined in Prob. 10.14. In tabular form, summarize the results of Probs. 10.13 through 10.16.

10.17. For the feedback-compensated system of Fig. 10.29,

$$G_x(s) = \frac{4}{s(s+2)} \qquad H(s) = K_h \frac{s(s+a)}{s+4}$$

(a) Use method 1 to determine $C(s)/R(s)$ using $T_s = 1.6$ s (2 percent criterion) and $\zeta = 0.65$ for the dominant roots. (b) Design a cascade compensator that yields the same dominant root as part (a). (c) Compare the gains of the two systems.

10.18. A unity-feedback system has

$$G_x(s) = \frac{K}{s(s+5)}$$

(a) For $\zeta = 0.5$, determine K and $C(s)/R(s)$. (b) Minor-loop and cascade compensation are to be added (see Fig. 10.29) with the amplifier replaced by $G_c(s)$:

$$G_c(s) = \frac{A(s+c)}{s+b} \qquad H(s) = \frac{K_h s}{s+a}$$

Use the value of K determined in part (a). By method **1** of the root-locus design procedure, determine values of a, b, K_h, and A to produce dominant roots having $\zeta = 0.5$, $\sigma = -7$, and $c(t)_{ss} = 1.0$. Only positive values of K_h and A are acceptable. To minimize the effect of the third real root, it should be located near a zero of $C(s)/R(s)$. (c) Determine $C(s)/E(s)$ and $C(s)/R(s)$ in factored form.

10.19. For $G(s)$ given in Prob. 10.3, add feedback rate (tachometer) compensation. Use root-locus method **1** or **2** as assigned. (a) Design the tachometer so that an improvement in system performance is achieved while maintaining a $\zeta = 0.5$ for the dominant roots. (b) Repeat part (a) for a $\zeta = 0.3$. Compare with the results of Prob. 10.3. *Note*: Use Fig. 10.29.

10.20. Refer to Fig. 10.29. It is desired that the dominant poles of $C(s)/R(s)$ have $\zeta = 0.5$ and $T_s = 1$ s. For the plant of Prob. 7.8(2), using the method **1** with $A = 5$, determine the values of b and K_t to meet these specifications. Be careful in selecting b. Check the time response to determine the effect of the real on the settling time.

$$H(s) = \frac{K_t s}{s+b}$$

10.21. Repeat Prob. 10.20 using method **2**. Compare the results with those of Probs 10.5 and 10.20.

10.22. The forward loop transfer function for a unity-feedback control system is

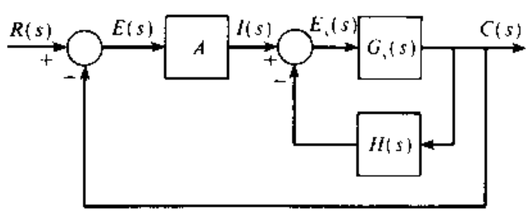
$$G_x(s) = \frac{K_x(s+10)}{s(s^2+4s+13)(s+50)}$$

The specifications for the closed-loop performance, with a step input, are $1 < M_p \le 1.095$, $t_p \le 1.6$ s, $t_s \le 2.9$ s, $K_1 \ge 2$. (*a*) Based upon these specifications determine ζ_D for the desired dominant closed-loop poles. (*b*) For this value of ζ_D are these dominant poles achievable? (*c*) Show that the desired specifications can be achieved without the use of a compensator. Hint: the complex poles of $G_x(s)$ have a $\zeta \approx 0.555$; therefore, the approach utilized in the Second Design of Sec. 10.4 may yield a $\zeta_D < 0.555$, for this problem, for the dominant closed-loop poles that may result in achieving the performance specifications. By trial and error, select a value of $\zeta_D < 0.555$ until a satisfactory solution is achieved.

10.23. Given the system of Fig. 10.29 with

$$G_x(s) = \frac{K_x(s+5)}{s(s+1.5)} \qquad H(s) = \frac{K_h(s+0.25)}{s+10}$$

(*a*) The specification for the inner loop is that $c(t)_{ss} = 1$ for $i(t) = u_{-1}(t)$. Determine the value of K_h in order to satisfy this condition. *Hint*: Analyze $C(s)/I(s)$. (*b*) It is desired that the dominant poles of $C(s)/R(s)$ be exactly at $-4 \pm j4$. By use of method 1, with K_h having the value determined in part (*a*), determine the values of A, K_m, and K_x that yield these dominant poles.



10.24. Draw the root locus for the system containing the PID cascade compensator of Sec. 10.14. Verify the closed-loop transfer function of Eq. (10.45) and the figures of merit.

CHAPTER 11

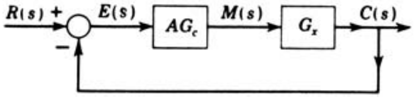
For the design problems of Chap. 11, obtain the time and frequency domain figure of merit M_p, t_p, t_s, K_m, M_m, ω_m, ω_ϕ, and γ for the uncompensated and compensated systems. Where appropriate, use a CAD package in solving these problems.

11.1. (*a*) For the basic control system of Prob. 8.17 adjust K to achieve an $M_m = 1.16$ and determine the corresponding values of ω_{ϕ_x} and γ, and all the other figures of merit. The value of γ is to be maintained for parts (*b*) through (*d*). (*b*) Add a lag compensator, with $\alpha = 10$. (*c*) Add a lead compensator with $\alpha = 0.1$. (*d*) Add a lag-lead compensator with $\alpha = 10$. (*e*) Form a table comparing both frequency and time domain figures of merit for all parts of this problem. Specify the time constants of the compensator used in each case.

11.2. A control system has the forward transfer function

$$G_x(s) = \frac{K}{s^2(1 + 0.05s)}$$

The closed-loop system is to be made stable by adding a compensator G_c and an amplifier A in cascade with $G_x(s)$. An $M_m = 1.41$ is desired with $\omega_m \approx 1.5$ rad/s. (*a*) What kind of compensator is needed? (*b*) Select an appropriate α and T for the compensator (*c*) Select the necessary value of amplifier gain A. (*d*) Plot the compensated curve. (*e*) Compare the results of this problem with those obtained in Prob. 10.2.



11.3. A basic (uncompensated) control system unity feedback has a forward open-loop transfer function

$$G_x(s) = \frac{K_1}{s(1 + s/4)(1 + s/12)^2}$$

(*a*) For the basic system, find the gain K_1 for $\gamma = 45^\circ$, and determine the corresponding phase-margin frequency ω_{ϕ_x}. (*b*) For the same phase-margin angle as in (*a*), it is desired to increase the phase-margin frequency to a value of $\omega_\phi = 3.2$ with the maximum possible improvement in gain. To accomplish this, a lead compensator is to be used. Determine the values of α and T that will satisfy these requirements. For these values of α and T, determine the new value of gain. (*c*) Repeat part (*b*) with the lag-lead compensator of Fig. 11.16. Select an appropriate value for T. With $G_c(s)$ inserted in cascade with $G_x(s)$, find the gain needed for 45° phase margin. (*d*) Show how the compensator has improved the system performance; i.e., determine all figures of merit (see note for Chap. 11 problems) for each part of this problem.

11.4. (*a*) Using the $G_x(s)$ of Prob. 11.3 wit $K_1 = 2.6271$ determine ω_{ϕ_x} and γ_x for the basic system. (*b*) Design a cascade lead compensator, using $\alpha = 0.1$, in order to achieve an $\omega_\phi = 3.2$ rad/s, for this value of $\gamma = \gamma_x$. Determine the required value of T and A. (*c*) For the basic (uncompensated) unity-feedback control system it is desired to increase K_m by a factor of approximately 10 while maintaining $\gamma = \gamma_x$. Design a cascade compensator that yields this increase in K_m. Specify the values of α and T so that $|\angle G_c(j\omega)| \leq 0.6^\circ$ at $\omega = \omega_{\phi_x}$. Determine the value of A and K_m for the compensated system. (*d*) It is desired that a phase margin angle $\gamma = 45^\circ$ and a phase-margin frequency $\omega_\phi = 3.2$ rad/s be achieved by use of the lag-lead compensator

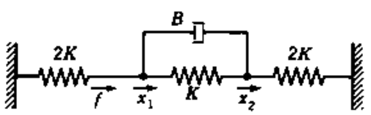
$$G_c'(s) = \frac{(1 + 100s)(1 + 0.5s)}{(1 + 1000s)(1 + 0.5\alpha s)}$$

Determine the values of α and A that yield these values of γ and ω_ϕ for the compensated system. Determine the ramp error coefficient of the compensated system. (*e*) It is desired, for the compensated system of Fig. 11.20, that $\omega_\phi = 3.2$ and $\gamma = 45^\circ$ (note that the phase-margin angle is determined *essentially* by the characteristics of a feedback unit). Determine if the following proposed feedback unity will result in a stable system. If a stable response can be achieved, determine the value of A that yields the desired value of γ and the corresponding value of ω_ϕ.

$$H(s) = \frac{K_H s^2}{1 + s/\alpha} \qquad K_H = 100$$

Note: For all parts determine the figures of merit: M_p, t_p, t_s, K_m, and M_m, and the actual values of ω_ϕ and γ that are achieved.

11.5. The mechanical system shown that been suggested for use as a compensating component in a mechanical system. (*a*) Determine whether it will function as a lead or a lag compensator by finding $X_2(s)/F(s)$. (*b*) Sketch $X_2(j\omega)/F(j\omega)$.



11.6. A unity-feedback control system has the transfer function

$$G(s) = \frac{K}{s(s+4)^3} G_c(s)$$

(*a*) What is the value of the gain of the basic system for an $M_m = 1.26$? (*b*) Design a cascade compensator $G_c(s)$ that will increase the step-error coefficient by a factor of 8 while maintaining the same M_m. (*c*) What effect does the compensation have on the closed-loop response of the system? (*d*) Repeat the design using a minor-loop feedback compensator.

11.7. A unity-feedback control system has the transfer function

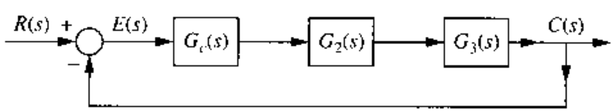
$$G(s) = G_c(s)G_x(s)$$

where $G_x(s) = 1/s^2$. It is desired to have an $M_m = 1$ with damped natural frequency of about 12.0 rad/s. Design a compensator $G_c(s)$ that will help to meet these specifications.

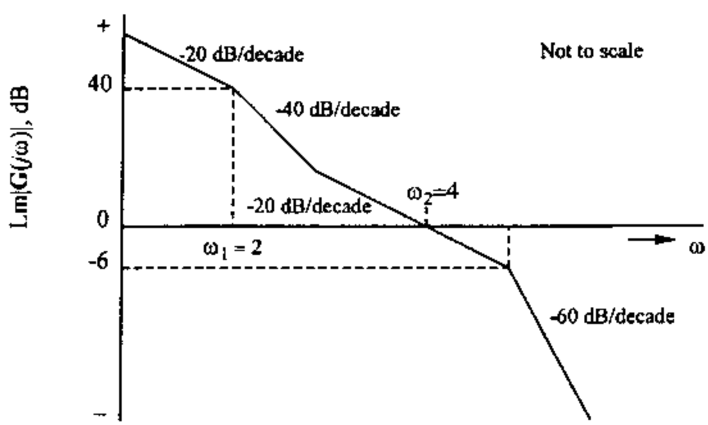
11.8. A system has

$$G_2(s) = \frac{0.8(1 + 0.0625s)}{1 + 0.01s} \qquad G_3(s) = \frac{1}{s}$$

For the control system shown in Fig. (*a*), determine the cascade compensator $G_c(s)$ required to make the system meet the desired open-loop frequency-response characteristics shown in Fig. (*b*). What is the value of M_m for the compensated system?



(*a*)



(*b*)

11.9. A system has the transfer function

$$G_x(s) = \frac{K_x(s + 0.4)}{s(s + 4)(s + 0.61 \pm j0.91)}$$

(*a*) For $M_m = 1.12$ determine ω_{ϕ_x}, γ, K_1, and ω_m for the original system. For parts (*b*) and (*c*) this value of γ is to be maintained. (*b*) Add a lag compensator with $\alpha = 10$, determine T, and adjust the gain to achieve the desired value of γ. (*c*) Add a lead compensator with $\alpha = 0.1$, determine T, and adjust the gain to achieve the desired value of γ. (*d*) Determine both frequency and time domain figures of merit, for all part of this problem, and compare these values in a tabular format.

11.10. Repeat Prob. 10.1(2), using $M_m = 1.16$ as the basis of design. Compare the results.

11.11. Repeat Prob. 10.14 but solve by the use of Bode plots. *Note*: Limit G_1 and G_2 to less than 100. (*a*) Assume that the forward transfer function for the uncompensated unity-feedback system is given by $G_2G_3G_4$, with $G_1 = A_1 = 1$. Thus the control ratio is

$$\frac{\Phi(s)}{R(s)} = \frac{A_2G_3(s)G_4(s)}{1 + A_2G_3(s)G_4(s)}$$

Solve for the value of A_2 that yields the desired value of $M_m \approx M_p = 1.2$. Determine the values of ω_{ϕ_x}, γ, M_p, t_p, t_s, and K_1, (*b*) For the compensated system, with $A_1 \neq 1$, incorporate G_4 into the minor loop. Determine $H_2(s)$ and A_1 in order to achieve the value of γ determined in (*a*) and determine all figures of merit for the compensated system, including ω_ϕ. (*c*) Compare the results of this problem with those of Prob. 10.13.

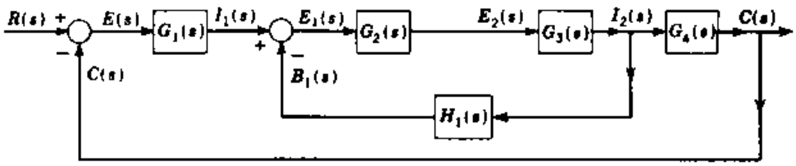
For Probs. 11.12 through 11.17, analyze the 0-dB crossing slope of the straight-line Bode plot of Lm C/E of system stability.

11.12. A system has

$$G_1(s) = 25 \quad G_2(s) = 5 \quad G_3(s) = \frac{4}{s+2} \quad G_4(s) = \frac{1}{s} \quad H_1(s) = \frac{s}{s+1}$$

System specifications are $K_1 = 200s^{-1}$, $\gamma \geq +40^\circ$, and $\omega_\phi = 8$ rad/s. Using approximate techniques, (*a*) determine whether the feedback-compensated system satisfies all the specifications; (*b*) find a cascade compensator to be added between $G_1(s)$ and $G_2(s)$, with the feedback loop omitted, to produce the same $C(j\omega)/E(j\omega)$ as in part (*a*);

(*c*) determine whether the system of part (*b*) satisfies all the specifications.



11.13. For the feedback-compensated system of Fig. 11.20,

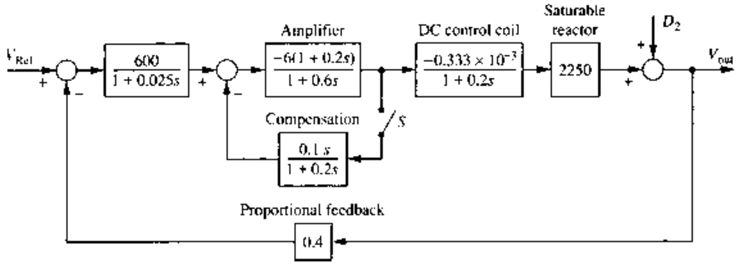
$$G_x(s) = \frac{4(1 + s/5)}{s(1 + 2)(1 + s/10)} \qquad H(s) = \frac{K_t T s^2}{1 + Ts}$$

(*a*) For the original system, without feedback compensation, determine the values of A and ω_{ϕ_x} for a phase-margin angle $\gamma = 50^\circ$. (*b*) Add the minor-loop compensator $H(s)$ and determine, for $\gamma = 50^\circ$, $\omega_\phi = 2.4$ rad/s, and $K_1 = 140$, the new value of A and the values of K_t and T. The specifications are to be dictated by $H(s)$. *Hint*: Adjust K_t so that the following conditions are satisfied:

$$\omega_1 \le \omega_\phi/20 \qquad \omega_2 \ge 20\omega_\phi$$

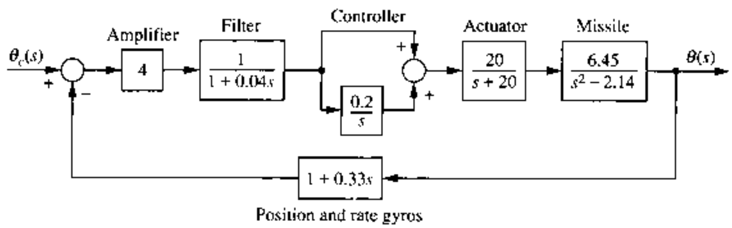
where ω_1 and ω_2 are the intersection frequencies of Lm G_x vs. ω and Lm $[1/\mathrm{H}(j\omega)]$.
(*c*) Determine the frequency and time domain figures of merit for parts (*a*) and (*b*). (*d*) Compare the results.

11.14. The system represented in the figure is an ac voltage regulator. (*a*) Calculate the response time of the uncompensated system, when the switch S is open, to an output disturbance $d_2 = u_{-1}(t)$. Calculate $v_{\text{out}}(t)$ for $t = T_s = 4/\sigma$, where σ is with the switch S open. (*b*) Repeat part (*a*) with the switch S closed. (*c*) Compare the values obtained in parts (*a*) and (*b*) and comment on any difference in performance.



www.4electron.com

11.15. The simplified pitch-attitude-control system used in a space-launch vehicle is represented in the block diagram. (*a*) Determine the upper and lower gain margin. (*b*) Is the closed system stable?



11.16. Repeat Prob. 10.19 using the log plots for $M_m = 1.16$. (*a*) For the uncompensated system adjust K_x to yield the value of $M_m = 1.16$ and determine the resulting values of ω_{ϕ_x}, γ, and all figures of merit. (*b*) In determining the value of K_t to use, first determine the range of values it can have in order to have an intersection between Lm $\mathbf{G}_x$ and Lm [1/H] and to maintain a stable open-loop $C(s)/I(s)$, transfer function. A root-locus sketch for $G_x(s)H(s) = -1$ reveals that $C(s)/I(s)$ can have right-half plane poles. Thus, for a stable $C(s)/I(s)$ the 0 dB crossing slope of Lm $\mathbf{G}_x\mathbf{H}$ must be greater than $-$ 40 dB/dec. (*c*) Select a value of K_t in the allowable range and determine the value of A that yields the desired value of γ. (*d*) Obtain the exact expressions for $C(s)/E(s)$ and $C(s)/R(s)$. (*e*) Determine all figures of merit for the compensated system. (*f*) Compare with the results in Prob. 10.3. Obtain the time response. *Hint*: since $|\angle 1/\mathbf{H}(j\omega)| = -90^\circ$, then, for $\gamma < 90^\circ$, the new ω_ϕ is close to the interaction of $1/\mathbf{H}(j\omega)$ and $\mathbf{G}(j\omega)$. *Note*: For Prob. 10.19 (1), for $K_t = 1$, the system is unstable if the value of K_x is set at the value obtained for the uncompensated system to meet the desired value of M_m.

11.17. (*a*) For the uncompensated unity-feedback control system where

$$G_x(s) = \frac{K_x(1+s)}{s^2(1+s/8)}$$

determine, for $\gamma = 50^\circ$, K_{2_x} and ω_{ϕ_x}. (*b*) Assume K_{2_x} has the value determined in part (*a*) and is not adjustable. For the feedback system of Fig. 11.20 the feedback compensator is given by $\mathbf{H}(j\omega) = K_H\mathbf{H}''(j\omega)$, where $K_H = K_t T$ and $\mathbf{H}''(j\omega) = (j\omega)^2/(1 + j\omega T)$. Draw plots of Lm $[1/\ \mathbf{H}''(j\omega)]$ and $\angle 1/\mathbf{H}''(j\omega)$. (*c*) It is desired, for the compensated system, that $\gamma = 50^\circ$ and the phase-margin frequency $\omega_\phi = 4\omega_{\phi_x}$ should be determined *essentially* by the characteristics of the feedback unit. Using straight-line approximations, determine graphically, as a first trial, the values of T, K_t, and A that yield this value of γ and the corresponding value of ω_ϕ. Analyze the slope of Lm $C(j\omega)/E(j\omega)$ at the 0 dB crossing for system stability. (*d*) Compare the values of M_p, M_m, ω_m, t_p, t_s, and K_2 of part (*c*) with part (*a*). For the value of A determined graphically in part (*c*), what are the actual values of ω_ϕ and γ of part (*c*)?

CHAPTER 12

Use a CAD package in solving the design problem of Chap. 12, where appropriate.

12.1. For the control system of Prob. 10.22 the desired specifications are $1 < M_p \le 1.123$, $t_s \le 3$ s, $t_p \le 1.6$ s, and $K_1 \ge 1.5\ \text{s}^{-1}$. (*a*) Synthesize a desired control ratio that satisfies these specifications. (*b*) Using the Guillemin-Truxal method (see Sec. 12.3), determine the required compensator $G_c(s)$.

12.2. The desired closed-loop time response for a step input is achieved by the system having the control ratio:

$$M_T(s) = \frac{23.1(s+20)}{(s^2+2s+10)(s+2.2)(s+21)}$$

Given

$$G_x(s) = \frac{10}{s(s^2+4.2s+14.4)} \qquad G_c(s) = A\frac{s+a}{s+b}$$

(*a*) Determine the required passive cascade compensator $G_c(s)$ for the unity-feedback system.

(*b*) Compare M_p, t_p, and t_s for the desired and actual $M(s)$.

12.3. The desired closed-loop time response for a step input is achieved by the system having the control ratio

$$M_T(s) = \frac{19.50(s+2)}{(s^2+4s+13)(s+3)} \quad G_x(s) = \frac{1}{s(s+1)} \quad G_c(s) = A\frac{(s+a)}{s+b}$$

(*a*) Determine, by the Guillemin-Truxal method, a passive cascade compensator $G_c(s)$ for unity-feedback system with the $G_x(s)$ shown. (*b*) It is desired to simplify the compensator to the form shown. Determine the values of A, a, and b that will yield the best results. (*c*) Compare M_p, t_p, t_s, and K_1 for the desired and the actual systems.

12.4. Determine a cascade compensator for a unity-feedback system that has the open-loop transfer function $G_x(s) = 10/[s(s+4)(s+6)]$. Determine the simplest possible approximate compensator transfer function that yields the time response, for a step input, that is achieved by the following desired control ratio:

$$M_T(s) = \frac{29.94(s+2.1)}{s^4 + 12s^3 + 43.99s^2 + 79.42s + 62.86}$$

12.5. For the nonunity-feedback control system of Fig. 10.3,

$$G_x(s) = \left[\frac{K_x \prod_{i=1}^{w} (s+a_i)}{\prod_{j=1}^{n} (s+b_j)}\right]_{n \geq w}$$

$$M_T(s) = \frac{G_x(s)}{1 + G_x(s)H(s)} = \left[\frac{K' \prod_{m=1}^{w'} (s-z_m)}{\prod_{k=1}^{n'} (s-p_k)}\right]_{n' \geq w'}$$

Determine whether a feedback compensator $H(s)$ (other than just gain) can be evaluated by use of the Guillemin-Truxal method such that the order of its denominator is equal to or greater than the order of its numerator.

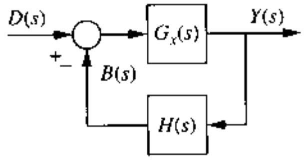
12.6. Determine a cascade compensator by the Guillemin-Truxal method where

$$G_x(s) = \frac{20}{s(s+3)(s+7)} \qquad M_T(s) = \frac{28.86}{(s+1.999)(s+7.305)(s^2+4s+8.84)}$$

12.7. Repeat Prob. 4.17, where, for disturbance rejection, it is desired that $y(t)_{ss} = 0$ for $D(s) = D_o/s$ with $a_1 = 0$. For this problem $Y(s)$ represents the output of a nonunity-feedback control system, where $G_x(s) = K_x G'(s) = N_1/D_1$ is fixed (the degree of numerator is w_1 and the degree of the denominator is n_1) and $H(s) = K_h H'(s) = N_2/D_2$ is to be determined (the degree of the numerator is w_2 and the degree of the denominator is n_2). *Hint*: First determine $Y(s)$ in terms of N_1, N_2, D_1, and D_2 and the degree of its numerator and denominator in terms of w_1, w_2, n_1, and n_2. For case (2) of Prob. 4.17, consider $n = w+1$ and $n > w+1$.

12.8. The figure below represents a disturbance-rejection control system, where

$$G_x(s) = \frac{10}{s(0.25s+1)}$$



(a) The following feedback unit

$$H(s) = \frac{K_H(s+4)(s+8)}{s(s+a)} \qquad K_H > 0$$

is proposed to meet the specifications of $y(t)_{ss} = 0$ for $d(t) = u{-}1(t)$, $|y(t_p)| \le 0.01$, and $\mathrm{Lm}|\mathbf{Y}(j\omega)/\mathbf{D}(j\omega)| \le -40\,\mathrm{dB} = -\mathrm{Lm}\,\alpha_p$ for $0 \le \omega \le 10$ rad/s. Select the value of a that satisfies these specifications. (b) *Graphically*, by use of straight-line asymptotes, determine a value of K_H that meets these specifications with the specified $H(s)$. (c) For your value of K_H, determine whether the system is stable by use of Routh's stability criterion. Correlate this with the slope of Lm $\mathbf{G}_x\mathbf{H}$ at the 0-dB crossing. (d) Obtain a plot of $y(t)$ vs. t and determine α_p and t_p. *Note*: Adjust K_H so that $|y(t_p)| = \alpha_p$.

12.9. For the disturbance-rejection control system of Prob. 12.8

$$G_x(s) = \frac{30}{s(s+3)} \qquad \text{and} \qquad H(s) = H_x(s)H_y(s)$$

where

$$H_x(s) = \frac{200}{s+200} \qquad H_y(s) = K_y\frac{N_y(s)}{D_y(s)} = \frac{K_y(s-z_1)(s-z_2)}{s(s-p_1)}$$

Determine the feedback compensator $H_y(s)$ such that the control system can satisfy the following specifications for a unit step disturbance input $d(t)$.

$$|y(t_p)| \le 0.01 = \alpha_p \qquad t_p \le 0.15\mathrm{s} \qquad y(t)_{ss} = 0$$

and

$$\mathrm{Lm}\left|\frac{\mathbf{Y}(j\omega)}{\mathbf{D}(j\omega)}\right| \le -40\,\mathrm{dB} = -\mathrm{Lm}\,\alpha_p$$

within the bandwidth of $0 < \omega \le 15$ rad/s. Analyze the 0-dB crossing slope of the straight-line Bode plot of Lm$\mathbf{G}_x\mathbf{H}$ for system stability and correlate this to a root-locus sketch of $G_x(s)H(s) = -1$.

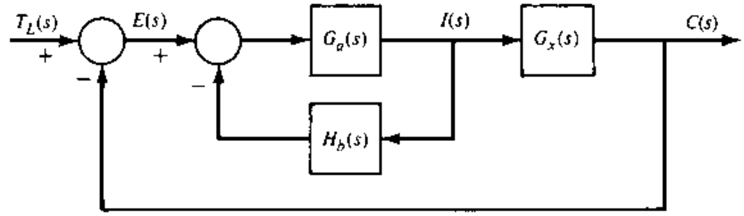
12.10. Analyze the time response characteristics of the control ratio

$$\frac{Y(s)}{R(s)} = \frac{K(s+a)}{(s+1.5)(s+6)}$$

with a unit step forcing function, for the following values of a: 0.01, 0.1, 1, 10, 50. For each value of a, adjust the value of K so that $Ka = 9$.

[See Eq. (4.68).] Discuss the characteristics of overdamped versus underdamped responses for an all pole control ratio.

12.11. The following block diagram is proposed to minimize the effect of the unwanted disturbance T_L on the output response $c(t)$:



(*a*) Determine $C(s)/T_L(s)$ only in terms of the symbols $G_a(s)$, $G_x(s)$ and $H_b(s)$. That is, do not use the actual transfer functions.

(*b*) *Determine the condition*, in the frequency domain, that must be satisfied so that $H_b(s)$ essentially determines whether the following specifications can be met:

$$\text{Lm}\left[\frac{C(j\omega)}{T_L(j\omega)}\right] \le -60\,\text{dB}$$

for frequency range $0 \le \omega \le 4\,\text{rad/s}$

Use *only* the straight-line approximation technique. *Hint*: Determine $C(s)/T_L(s)$ so that it is directly a function of $1/H_b(s)$, and again do not use the actual transfer functions.

(*c*) Determine the value of K_b that can satisfy the frequency-domain specifications where

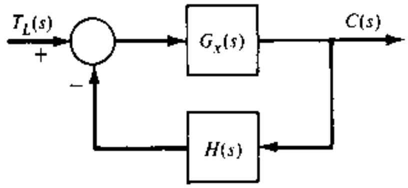
$$G_x(s) = \frac{1}{s+1} \qquad G_a(s) = \frac{1}{s} \qquad H_b(s) = \frac{K_b(s+2)}{s(s+100)}$$

(*d*) Does this value of K_b yield a stable system? Illustrate your answer by a root-locus sketch.

(*e*) If the value of K_b results in a stable system, determine $c(t_p)$, t_p, t_s, and $c(t)_{ss}$ for $T_L = 1$.

12.12. The following figure represents a disturbance-rejection control system where

$$G_x(s) = \frac{50(s+5)}{s(s+0.5)(s+10)}$$



The feedback unit

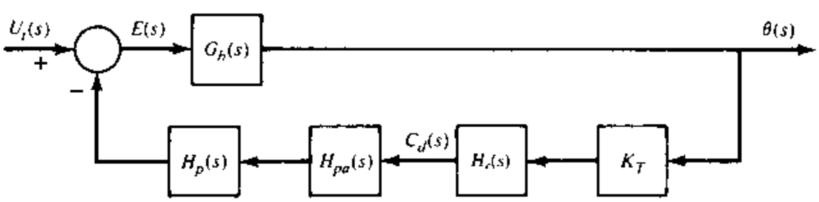
$$H(s) = \frac{K_h(s+1)}{s(s+5)} \quad K_h > 0$$

is proposed to meet the specifications of $c(t)_{ss} = 0$ for a unit-step disturbance and

$$\text{Lm}\left[\frac{\mathbf{C}(j\omega)}{\mathbf{T}_L(j\omega)}\right] \le -40\,\text{dB} \qquad \text{for frequency range } 0 \le \omega \le 5\,\text{rad/s}$$

(*a*) *Graphically*, by use of straight-line asymptotes, determine a value of K_h that results in these specifications being met with the specified $H(s)$. (*b*) With this value of K_h, plot Lm $\mathbf{C}(j\omega)/\mathbf{T}_L(j\omega)$vs.$\omega$. (*c*) Does this value of K_h yield a stable system?

12.13. Given the nonunity-feedback system:



where

$$H_p(s) = \frac{10}{s+1} \qquad H_{pa} = \frac{100}{s+100}$$

$$G_b(s) = \frac{2.2557}{s^2 + 2.148s + 51.266} \qquad K_T = 7.4$$

The system specifications are $\theta(t_p) \le 0.01$ for $u_t(t) = 1.25u_{-1}(t)$, $\theta(t)_{ss} = 0$, $\omega_b \approx 125$ rad/s (bandwidth), and $t_s \le 4$ s [t_s is based on $\theta(t_s) = 0.0002$, i.e., 2 percent of the maximum allowable $\theta(t_p)$]. $H_c(s)$ is a third-order over a third-order feedback compensator. (*a*) Determine $H_c(s)$ that results in the system achieving the desired specifications. Your solution should follow the design procedure of Sec. 12.6 and include a figure containing plots for your design that correspond to those in Fig. 12.9. (*b*) For your final design obtain the following plots: $\theta(t)$ vs. t, $\mathrm{Lm}[\theta(j\omega)/U_t(j\omega)]$ vs. ω and the root locus.

CHAPTER 13

Where appropriate use a CAD package to solve these problems.

13.1. For a single-input single-output system the open-loop matrix state and output equations are $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u$ and $y = \mathbf{c}^T\mathbf{x}$. Derive the overall closed-loop control ratio for a state-variable feedback system with $u = \mathbf{k}^T\mathbf{x}$:

$$\frac{Y(s)}{R(s)} = \mathbf{c}^T[s\mathbf{I} - (\mathbf{A} - \mathbf{b}\mathbf{k}^T)]^{-1}\mathbf{b}$$

13.2. For the system described by

$$D^3y + 9D^2y + 26Dy + 24y = 5u$$

(*a*) Draw the simulation diagram. (*b*) Write the matrix state and output equations for the simulation diagram. (*c*) Rewrite the state equation and the output equation in uncoupled form. (*d*) From part (*c*) determine whether the system is completely controllable, observable, or both.

13.3. (*a*) Determine whether the following systems are completely observable. (*b*) Determine whether they are completely controllable. (*c*) Obtain the transfer functions. (*d*) Determine how many observable states and how many controllable states are present in each system. (*e*) Are the systems stable?

$$(1)\quad \dot{\mathbf{x}} = \begin{bmatrix} 1 & 0 \\ 2 & 2 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}\mathbf{u} \qquad \mathbf{y} = [2 \quad 1]\mathbf{x}$$

$$(2)\quad \dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -5 & -6 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\mathbf{u} \qquad \mathbf{y} = [0 \quad 1 \quad 1]\mathbf{x}$$

$$(3)\quad \dot{\mathbf{x}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & -2 & -1 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\mathbf{u} \qquad \mathbf{y} = [1 \quad 1 \quad 0]\mathbf{x}$$

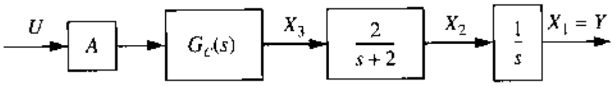
$$(4)\quad \dot{\mathbf{x}} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & -2 & -2 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\mathbf{u} \qquad \mathbf{y} = [1 \quad 1 \quad 0]\mathbf{x}$$

13.4. A single-input single-output system is described by the transfer function

$$G(s) = \frac{s^2 + 10s + 24}{s^4 + 9s^3 + 26s^2 + 24s}$$

(*a*) Derive the phase-variable state equation. Draw the block diagram. The A_c matrix is singular. How is this interpreted physically? (*b*) Represent the system in terms of canonical variables. Draw the corresponding block diagram. (*c*) Is the system observable and controllable?

13.5. A control system is to use state-variable feedback. The plant has the given transfer function. The system is to meet the following requirements: (1) it must have zero steady-state error with a step input, (2) dominant poles of the closed-loop control ratio are to be $-2 \pm j3$, (3) the system is to be stable for all values $A > 0$, and (4) the control ratio is to have low sensitivity to increase in gain A. A cascade element $G_c(s)$ must be added to meet this requirement.



(*a*) Draw the block diagram showing the state-variable feedback. (*b*) Determine $H_{eq}(s)$. (*c*) Find $Y(s)/R(s)$ in terms of the state-variable feedback coefficients. (*d*) Determine the desired control ratio $Y(s)/R(s)$. (*e*) Determine the necessary values of the feedback coefficients. (*f*) Sketch the root locus for $G(s)H_{eq}(s) = -1$ and show that the system is insensitive to variations in A. (*g*) Determine $G_{eq}(s)$ and K_1. (*h*) Determine M_p, t_p, and t_s with a step input.

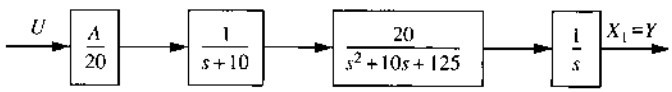
13.6. A control system is to use state-variable feedback. The plant has the given transfer function. The system is to meet the following requirements: (1) it must have zero steady-state error with a step input, (2) the dominant poles of the closed-loop control ratio are supposed to be $-1 \pm j2$, (3) the system is to be stable for all values $A > 0$, and (4) the control ratio is to have low sensitivity to increase in gain A and is to contain a zero at $s = -1.5$. A cascade element may be added to meet this requirement.

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

(*a*) Determine the desired control ratio $Y(s)/R(s)$. (*b*) Draw the block diagram showing the state-variable feedback. Use the form illustrated in Fig. 13.17 where the term $(s+2)/(s+1)$ in the figure is changed to

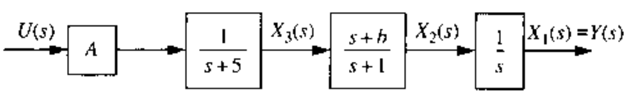
$(s+3)/(s+1)$ and the transfer function $1/(s+5)$ is replaced by $(s+1.5)/(s+6)$ for this problem. (*c*) Determine $H_{eq}(s)$. (*d*) Find $Y(s)/R(s)$ in terms of the state-variable feedback coefficients. (*e*) Determine the necessary values of the feedback coefficients. (*f*) Sketch the root locus for $G(s)H_{eq}(s) = -1$ and show that the system is insensitive to variations in A. (*g*) Determine $G_{eq}(s)$ and K_1. (*h*) Determine M_p, t_p, and t_s with a step input.

13.7. Design a state-variable feedback system for the given plant $G_x(s)$. The desired complex dominant roots are to have a $\zeta = 0.5$. For a unit-step function the approximate specifications are $M_p = 1.2$, $t_p = 0.2$ s, and $t_s = 0.5$ s. (*a*) Determine a desirable $M(s)$ that will satisfy these specifications. Use steps 1 to 6 of the design procedure given in Sec. 13.6 to determine **k**. Draw the root locus for $G(s)H_{eq}(s) = -1$. From it show the "good" properties of state feedback. (*b*) Obtain $y(t)$ for the final design. Determine the values of the figures of merit and the ramp-error coefficient.



13.8. For the plants of Prob. 10.3 design a state-variable feedback system utilizing the phase-variable representation. The specifications are that dominant roots must have $\zeta = 0.5$, a zero steady-state error for a step input, and $t_s \le \frac{1}{3}$ s. (*a*) Determine a desirable $M(s)$ that will satisfy the specifications. Use steps 1 to 6 of the design procedure given in Sec. 13.6 to determine **k**. Draw the root locus for $G(s)H_{eq}(s) = -1$. From it show the "good" properties of state feedback. (*b*) Obtain $y(t)$ for the final design. Determine the values of the figures of merit and the ramp-error coefficient.

13.9. Design a state-variable feedback system that satisfies the following specifications: (1) for a unit-step input, $M_p = 1.10$ and $T_s \le 0.6$ s; and (2) the system follows a ramp input with zero steady-state error. (*a*) Determine the expression for $y(t)$ for a step and for a ramp input. (*b*) For the step input determine M_p, t_p, and t_s. (*c*) What are the values of K_p, K_v, and K_a?



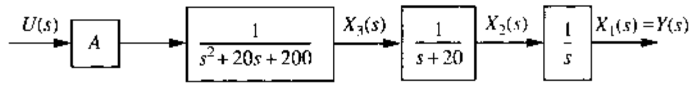
13.10. A closed-loop system with state-variable feedback is to have a control ratio of the form

$$\frac{Y(s)}{R(s)} = \frac{A(s+2)(s+0.4)}{(s^2+4s+8)(s-p_1)(s-p_2)}$$

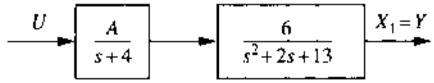
where p_1 and p_2 are not specified. (*a*) Find the two necessary relationships between A, p_1, and p_2 so that the system has zero steady-state error with both a step and a ramp input. (*b*) With $A = 100$, use the relationships developed in (*a*) to find p_1 and p_2. (*c*) Use the system shown in Fig. 13.17 with the transfer function $1/(s+5)$ replaced by $(s+0.4)/(s+5)$. Determine the values of the feedback coefficients. (*d*) Draw the root locus for $G_x(s)H_{eq}(s) = -1$. Does this root locus show that the system is insensitive to changes of gain A? (*e*) Determine the time response with a step and with a ramp input. Discuss the response characteristics.

13.11. Design a state-variable feedback system, for each of the plants shown, that satisfy the following specifications: for plant (1) the dominant roots must have a $\zeta = 0.55$, a zero steady-state error for a step input, and $T_s = 0.93$ s; and for plant (2) the dominant roots must have a $\zeta = 0.6$, a zero steady-state error for a step input, and $T_s = 0.8$ s.

(1)



(2)



13.12. For the system of Fig. 13.15 it is desired to improve t_s from that obtained with the control ratio of Eq. (13.119) while maintaining an underdamped response with a smaller overshoot and faster settling time. A desired control ratio that yields this improvement is

$$\frac{Y(s)}{R(s)} = \frac{1.467(s+1.5)}{(s+1.1)(s^2+2s+2)}$$

The necessary cascade compensator can be inserted only at the output of the amplifier A. (*a*) Design a state-variable feedback system that yields the desired control ratio. (*b*) Draw the root locus for $G(s)H_{eq}(s)$ and comment on its acceptability for gain variation. (*c*) For a step input determine M_p, t_p, t_s, and K_1. Compare with the results of Examples 1 to 3 of Sec. 13.12.

13.13. For Example 3, Sec. 13.12, determine $G_{eq}(s)$ and the system type. Redesign the state-variable feedback system to make it Type 2.

13.14. The desired control ratio $M_T(s)$ is to satisfy the following requirements: the desired dominant poles are given by $s^2 + 4s + 8$; there are no zeros; $e(t)_{ss} = 0$ for a step input; and the system is insensitive to variations in gain A. Determine an $M_T(s)$ that achieves these specifications for the plant

$$G_x(s) = \frac{A(s+3)}{s(s+1.2)(s+2.5)(s+4)}$$

13.15. Repeat Prob. 13.14 with the added stipulation that $e(t)_{ss} = 0$ for a ramp input. A cascade compensator may be added to achieve the desired specifications. A zero may be required in $M_T(s)$ to achieve $e(t)_{ss}$ for a ramp input. Compare the figures of merit with those of Prob. 13.14.

13.16. For the following plant

$$G_x(s) = \frac{A(s+2)}{s(s^2 + 4s + 3.84)}$$

(*a*) Determine a desired control ratio $M_T(s)$ such that the following specifications are met: (1) the dominant roots are at $-1 \pm j1$, (2) the zero at -2 does not affect the output response, (3) $e(t)_{ss} = 0$ for a step input, and (4) the system sensitivity to gain variation is minimized. (*b*) To achieve the desired $M_T(s)$, will a cascade unit in conjunction with $G_x(s)$ be needed? If your answer is yes, specify the transfer function(s) of unit(s) that must be added to the system to achieve the desired system performance.

13.17. Given the following closed-loop control ratio

$$M(s) = \frac{Y(s)}{R(s)} = \frac{A(s+a)^2}{(s^2 + 4s + 8)(s+2)(s+b)}$$

(*a*) With A, a, and b unspecified, determine $G_{eq}(s)$. (*b*) For part (*a*), determine the condition that must be satisfied for $G_{eq}(s)$ to be a Type 2 transfer function. (*c*) It is desired that $e(t)_{ss} = 0$ for step and ramp inputs. Determine the values of A, a, and b that will satisfy these specifications.

13.18. An open-loop plant transfer function is

$$G(s) = \frac{5}{s(s+2)(s+4)}$$

(*a*) Write state and output equations in control canonical (phase-variable) form. (*b*) Compute the feedback matrix $\mathbf{k}_p^T$, which assigns the set of closed-loop eigenvalues $\{-0.5 \pm j0.6, -60\}$. (*c*) Compute the state and output responses for the unit-step input $r(t) = u_{-1}(t)$ and $\mathbf{x}(0) = [1 \quad 0 \quad 1]^T$. Plot the responses.

13.19. For the plant represented by

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 \\ -2 & -4 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ 2 \end{bmatrix} \mathbf{u}$$

Design the feedback matrix **K** that assigns the set of closed-loop plant eigenvalues as $\{-3, -6\}$. Determine the state responses with the initial values $\mathbf{x}(0) = [1 \quad -1]^T$.

13.20. A SISO system represented by the state equation $\dot{\mathbf{x}}_p = \mathbf{A}_p \mathbf{x}_p + \mathbf{b}_p \mathbf{u}$ has the matrices

$$\mathbf{A}_p = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -2 & -2 \\ -1 & 0 & -3 \end{bmatrix} \qquad \mathbf{b}_{p_1} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad \mathbf{b}_{p_2} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

For each $\mathbf{b}_p$: (*a*) Determine controllability. (*b*) Find the eigenvalues. (*c*) Find the transformation matrix **T** that transforms the state equation to the control canonical (phase-variable) form. (*d*) Determine the state-feedback matrix $\mathbf{k}_c^T$ required to assign the eigenvalue spectrum $\sigma(\mathbf{A}_{cl}) = \{-2, \quad -4, \quad -6\}$.

13.21. For the system of Prob. 13.20, determine the feedback matrix **k** that assigns the closed-loop eigenvalue spectrum $\sigma(\mathbf{A}_{cl}) = \{-1 + j2, -1 - j2, -2\}$.

13.22. Repeat Prob. 13.20 with

$$\mathbf{A}_p = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -2 & -3 \\ 0 & 0 & -6 \end{bmatrix} \quad \mathbf{b}_p = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \quad \sigma\{\mathbf{A}_{cl}\} = \{-1 + j1 - 1 - j1, -4\}$$

13.23. $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$, where

$$\mathbf{A} = \begin{bmatrix} -3 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

(a) Determine controllability. (b) Transform the state equation to the control canonical form. Determine the matrix T. (c) Design the feedback matrix **K** that assigns the closed-loop eigenvalue spectrum

$$\sigma(\mathbf{A}_{cl}) = \{-3 + j2, \ -3 - j2, \ -5 - 10\}$$

(d) Obtain the time response of the closed-loop system with zero input and the initial conditions $\mathbf{x}(0) = [1 \quad 0 \quad 0 \quad 1]^T$.

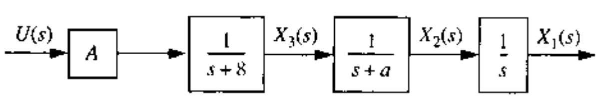
CHAPTER 14

14.1. A unity-feedback control system has the following forward transfer function:

$$G(s) = \frac{K(s+a)}{s^m(s+b)(s^2+cs+d)}$$

The nominal values are $K = 10$, $a = 4$, $b = 2$, $c = 2$, and $d = 5$. (a) With $m = 1$, determine the sensitivity with respect to (1) K, (2) a, (3) b, (4) c, and (5) d. (b) Repeat with $m = 0$.

14.2. The range of values of a is $3 \geq a \geq 1$, $M_p \leq 1.095$ for a unit-step input, and $T_s \leq 0.9$ s. Design a state-variable feedback system that satisfies the given specifications and is insensitive to variation of a. Assume that all states x_i are accessible and are fed back directly through the gains k_i. Obtain plots of $y(t)$ for $a = 1$, $a = 2$, and $a = 3$ for the final design. Compare M_p, t_p, t_s, and K_1. Determine the sensitivity functions for variation of A and a (at $a = 1$, 2, and 3).



14.3. Repeat Prob. 14.2 but with state X_3 inaccessible.

14.4. Determine the sensitivity function for Example 1 in Sec. 13.12 for variations in A. Insert the values of k_i in Eq. (13.118) and then differentiate with respect to A to obtain S_A^M.

14.5. For Example 2, Sec. 13.12, obtain $Y(s)/R(s)$ as a function of K_G [see Fig. 13.17 with $a = 1$]. Determine the sensitivity function $S_{K_G}^M$.

14.6. The desired control ratio for a tracking system is

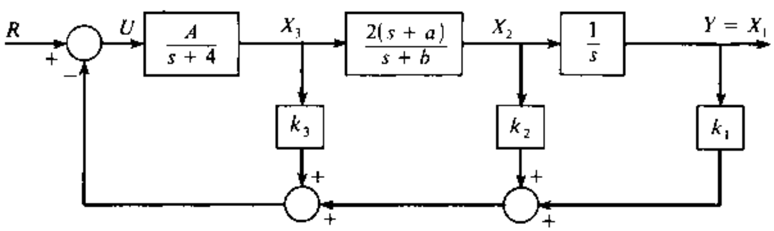
$$M_T(s) = \frac{K_G(s+1.5)}{(s+1.1)(s^2+2s+2)(s-p)}$$

This control ratio must satisfy the following specifications: $e(t)_{ss} = 0$ for $r(t) = R_0 u_{-1}(t)$, and $|S_{K_G}^{M}(j\omega)|$ must be a minimum over the range $0 \le \omega \le \omega_b$. (*a*) Determine the values of K_G and p that satisfy these specifications. The transfer function of the basic plant is

$$G_x(s) = \frac{K_x(s+3)}{s(s+1)(s+5)} = \frac{K_x(s+3)}{s^3 + 6s^2 + 5s}$$

A controllable and observable state-variable feedback-control system is to be designed for this plant so that it produces the desired control ratio above. (*b*) Is it possible to achieve $M_T(s)$ using only the plant in the final design? If not, what simple unit(s) must be added to the system to achieve the desired system performance? Specify the parameters of these unit(s). Using phase-variable representation, determine $\mathbf{k}^T$ for the state-variable feedback-control system. Assume appropriate values to perform the calculations.

14.7. For the state-feedback system shown, all the states are accessible, the parameters a and b vary, and sensors are available that can sense all state variables. The values of k_1, k_2, k_3, and A are known for a desired $M_T(s)$ based on nominal values for a and b. It is desired that the system response $y(t)$ be insensitive to parameter variations. (*a*) By means of a block diagram indicate an implementation of this design to minimize the effect of parameter variations on $y(t)$ by use of physically realizable networks. (*b*) Specify the required feedback transfer functions, ignoring sensor dynamics.



14.8. For the state-feedback control system of Prob. 14.7 with $a = 4$, $b = 3$ and a and b do not vary, the specifications are as follows: $e(t)_{ss} = 0$ for $r(t) = u_{-1}(t)$, $|S_A^M(j\omega)|$ must be a minimum over the range $0 \le \omega < \omega_b$, and the dominant poles of $M_T(s)$ are given by $s^2 + 4s + 13$. Determine (*a*) an $M_T(s)$ that achieves these specifications and (*b*) the values for K_G, k_1, k_2, and k_3 that yield this $M_T(s)$.

14.9. Use the Jacobian matrix to determine the stability in the vicinity of the equilibrium points for the system described by:

$$(a)\ \dot{x}_1 = 3x_1 = x_1^2 - 2x_1x_2 + u_1$$
$$\dot{x}_2 = -x_1 + x_2 + u_1u_2, \text{ where } \mathbf{u}_0 = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$$
$$(b)\ \dot{x}_1 = 2x_1 - 4x_1x_2$$
$$\dot{x}_2 = +x_1^2 - 2x_2$$
$$(c)\ \dot{x}_1 = -4x_1^2 + x_2$$
$$\dot{x}_2 = x_1 - 2x_2$$

14.10. For the systems described by the state equations

$$(1)\ \dot{x}_1 = x_1 - x_1^2 - x_1x_2 - 2u_1,\ \dot{x}_2 = +x_1 + 2x_2 - u_1u_2,\ \mathbf{u}_0 = [1\ \ 0]^T$$
$$(2)\ \dot{x}_1 = x_2, \quad \dot{x}_2 = 2x_1 - 3x_2 + 0.25x_3^2, \quad \dot{x}_3 = 2x_1x_3 + x_3$$
$$(3)\ \dot{x}_1 = 4x_2, \quad \dot{x}_2 = +2x_2 - \sin x_1$$

(*a*) Determine the equilibrium points for each system. (*b*) Obtain the linearized equations about the equilibrium points. (*c*) Determine whether the system is asymptotically stable or unstable in the neighborhood of its equilibrium points.

CHAPTER 15

15.1. By use of Eqs. (15.13) and (15.14), verify the following $\mathscr{Z}$ transforms in Table 15.1: (*a*) entry 4, (*b*) entry 7, (*c*) entry 12, (*d*) entry 14.

15.2. Determine the inverse $\mathscr{Z}$ transform, in closed form, of

$$(a)\ F(z) = \frac{z(z-0.2)}{(z-1)(z^2-z+0.41)} \quad (b)\ F(z) = \frac{-4(z^2+z)}{(z-1)^2(z-0.6)(z-0.5)}$$

15.3. For the sampled-data control system of Fig. 15.14 the output expression is:

$$C(z) = \frac{0.5K(z^2-0.48z)}{(z-1)(z^2-z+0.26)}$$

where $T = 1$ s, $r(t) = u_{-1}(t)$.
(*a*) Apply the partial fraction expansion to the expression $C(z)/z$ by the method of Sec. 15.7. (*b*) Using the resulting $C(z)$ expression, i.e., $z[C(z)/z]$, of part (*a*) determine $c(kT)$ by use of Table 15.1. For the complex poles use entry 12 of Table 15.1. (*c*) Determine $c(0)$ and $c(\infty)$.

15.4. Repeat Prob. 15.3 for

$$C(z) = \frac{z(z - 0.1)(z + 0.2)}{(z - 1)(z - 0.6)(z - 0.8)}$$

15.5. Repeat Prob. 15.2 but use the power-series method to obtain the open form for $f(kT)$.

15.6. Determine the value of $c(2T)$ for

$$C(z) = \frac{z^4 + z^3 + z^2}{z^4 - 2.8z^3 + 3.4z^2 - 2.24z + 0.63}$$

by use of the power series expansion method.

15.7. Determine the initial and final values of the $\mathscr{Z}$ transforms given in Prob. 15.2, i.e., solve for $c(0)$ and $c(\infty)$, where $R(z) = z/(z-1)$.

15.8. Let $F(z) = C(z)/E(z)$ in Prob. 15.2. Determine the difference equation for $c(kT)$ by use of the translation theorem of Sec. 15.4.

15.9. Based upon the time domain to the s- to the z-plane correlation determine the anticipated values of t_p and T_s for $c(kT)$ for the function

$$C(z) = \frac{Kz(z + 0.2)}{(z - 1)(z^2 - 1.4z + 0.74)}$$

where $T = 0.1$ s and $r(t) = u_{-1}(t)$.

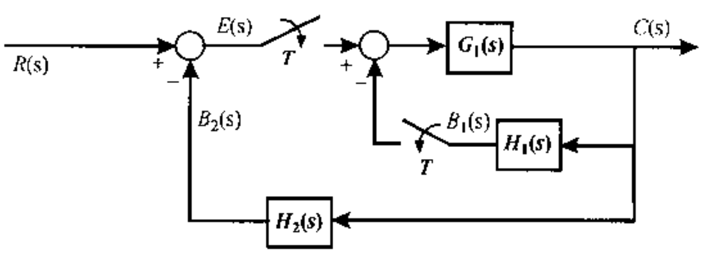
15.10. Determine the values of $c(0)$ and $c(\infty)$ for

$$C(z) = \frac{Kz(z + 0.1)}{(z - 1)(z^2 - 1.6z + 1)}$$

where $r(t) = u_{-1}(t)$.

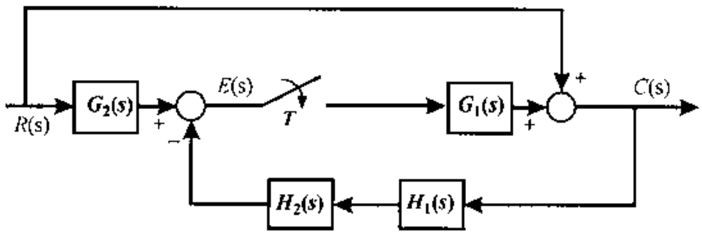
15.11. For the sampled-data systems shown, determine $C(z)$ and $C(z)/R(z)$, if possible, from the block diagram. *Note*: Write *all* basic equations that relate *all* variables shown in the figure.

(1)

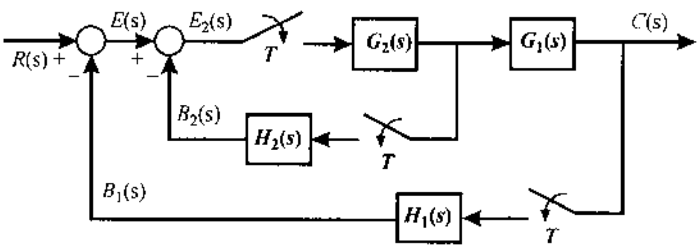


www.4electron.com

(2)



(3)



15.12. **(CAD Problem)** For the sampled-data control system of Fig. 15.14, $G_x(s)$ is given by Eq. (15.15), where $a=4$ and T is unspecified. (*a*) Determine $G(z)$ and $C(z)/R(z)$. (*b*) Assume a sufficient number of values of T between 0.01 and 10 s, and for each value of T determine the maximum value that K can have for a stable response. (*c*) Plot $K_{\max}$ vs. T and indicate the stable region. (*d*) Obtain the root locus for this system with $T=0.1$ s. (*e*) Use Eq. (15.19) to locate the roots on the root locus which have a $\zeta=0.65$ and determine the corresponding value of K. (*f*) For a unit step input determine $c(kT)$, M_p, t_p, and t_s.

15.13. For the sampled-data system of Fig. 15.14 the plant transfer function is

$$G_x(s) = \frac{K_x}{s(s+2)(s+10)}$$

where $T=0.05$ s. The desired performance specifications are: $M_p=1.095$ and $t_s=2$ s. (*a*) Determine the value ζ_D that the desired dominant roots must have. (*b*) For this value of ζ_D, by use of the root-locus method, determine the value of K_x and the roots of the characteristic equation. (*c*) For the value of K_x of part (*b*) determine the resulting $C(z)/R(z)$. (*d*) For a unit step forcing function determine the actual values of M_p and t_s for the uncompensated system.

CHAPTER 16

16.1. Given

$$(1) \quad G(s) = \frac{K(s+3)(s+20)}{s(s+1.5)(s+2\pm j3)(s+10\pm j25)}$$

$$(2) \quad G(s) = \frac{K(s+5)(s+10)}{s(s+1)(s+2\pm j2)(s+10\pm j25)}$$

and $T = 0.04$ s. Map each pole and zero of $G(s)$ by use of Eq. (16.8) and $z = e^{sT}$. Compare the results with respect to Fig. 16.5 and analyze the warping effect.

16.2. Repeat Prob. 16.1, with $T = 0.01$ s and 1 s, for:

(1) $s = -1$ (2) $s = -500$ (3) $s = -1000 \pm j500$
(4) $s = -2 + j0.5$ (5) $s = -0.1 + j0.1$ (6) $s = -0.8 + j0.6$

16.3. The poles and zeros of the following desired control ratio of the PCT system have been specified:

$$(1) \quad \left[\frac{C(s)}{R(s)}\right]_M = \frac{K(s+8)}{(s+5\pm j7)(s+10)} \qquad (2) \quad \left[\frac{C(s)}{R(s)}\right]_M = \frac{K(s+8)}{(s+4\pm j8)(s+6)}$$

(*a*) What must be the value of K in order for $y(t)_{ss} = r(t)$ for a step input? (*b*) A value of $T = 0.1$ s has been proposed for the sampled-data system whose design is to be based upon the above control ratio. Do all the poles and zeros lie in the good Tustin approximation region of Fig. 16.5? If the answer is no, what should be the value of T so that all the poles and zeros do lie in this region?

16.4. (*a*) Obtain the PCT equivalent for the sampled-data system of Prob. 15.12 for $T = 0.1$ s. (*b*) Repeat parts *d* through *f* of Prob. 15.12 for the PCT equivalent. (*c*) Compare the results of part *b* with those of parts *d* through *f* of Prob. 15.12. This problem illustrates the degradation in system stability that results in converting a continuous-time system into a sampled-data system.

16.5. **(CAD Problem)** The sampled-data control system of Prob. 16.4 is to be compensated in the manner shown in Fig. 16.1. Determine a controller by the DIR technique that reduces t_s of Prob. 16.4, by one-half. Determine M_p, t_p, t_s, and K_m.

16.6. Repeat Prob. 16.5 by the PCT DIG technique. (*a*) By use of Eq. (16.8) obtain $[D_c(z)]_{TU}$ and the control ratio $[C(z)/R(z)]_{TU}$. (*b*) Obtain the values of M_p, t_p, t_s, and K_m compare these values with those of Prob. 16.5.

16.7. Repeat Prob. 16.4 for $T = 0.04$.

16.8. Repeat Prob. 16.5 for $T = 0.04$.

16.9. Repeat Prob. 16.6 for $T = 0.04$ and compare the results with those of Prob. 16.8.

16.10. **(CAD Problem)** For the basic system of Fig. 16.7, where $G_x(s) = K_x/s(s+2)$: (*a*) Determine M_p, t_p, t_s, and K_1 for $\zeta = 0.65$ and $T = 0.01$ s by the DIR method. (*b*) Obtain the PCT control system of part *a* and the corresponding figures of merit. (*c*) Generate a controller $D_c(z)$ (with no ZOH) that reduces the value of t_s of part *a* by approximately one-half for $\zeta = 0.65$ and $T = 0.01$ s. Do first by the DIG method (*s* plane) and then by the DIR method. Obtain M_p, t_p, and t_s for $c(t)$ of the PCT and $c^*(t)$ (for both the DIG and the DIR designs) and K_m. Draw the plots of $c^*(t)$ vs. $*kT$ for both designs. (*d*) Find a controller $D_c(z)$ (with no ZOH) that increases the ramp error coefficient of part *a*, with $\zeta = 0.65$ and $T = 0.01$ s with a minimal degradation of the transient-response characteristics of part *a*. Do first by the DIG method (*s* plane) and then by the DIR method. Obtain M_p, t_p, and t_s for $c(t)$ of the PCT and $c^*(t)$ for both the DIG and the DIR designs and K_m. Draw a plot of $c^*(t)$ vs. kT for both designs. (*e*) Summarize the results of this problem in tabular form and analyze. For part (*b*) use 4-decimal-digit accuracy; for parts (*c*) and (*d*) 4-decimal-digit accuracy; then repeat for 8-decimal-digit accuracy. Compare results.

16.11. By use of the PCT DIG technique design a controller $[D_c(z)]_{\mathrm{TU}}$ [Eq. (16.46)] for Prob. 15.13 which will achieve the desired performance specifications.

16.12. For the digital control system of Fig. 16.1 the plant and the ZOH transfer function is:

$$G_z(s) = G_{zo}(s)G_p(s) = \frac{2(1 - e^{-Ts})(s+3)}{s^2(s+1)(s+2)}$$

The desired figures of merits (FOM) are $M_p = 1.043$ and $t_s = 2$ s. (*a*) By use of the PCT DIG technique, where $T = 0.02$ s, determine $D_c(s)$ that will yield the desired dominant poles, based upon satisfying the specified FOM, for the system's control ratio. (*b*) Determine the resulting FOM for the compensated system of part *a*. (*c*) By use of Eq. (16.8) obtain $[D_c(z)]_{\mathrm{TU}}$ and the control ratio $[C(z)/R(z)]_{\mathrm{TU}}$. (*d*) Obtain the values of M_p, t_p, t_s, and K_m for part (*c*) and compare these values with those of part (*b*).

16.13. **(CAD Problem)** For the basic system of Fig. 15.17, where $G_x(s) = K_x/s(s+2)$: (*a*) Determine M_p, t_p, t_s, and K_1 for $\zeta = 0.65$ and $T = 0.01$ s by the DIR method. (*b*) Obtain the PCT control system of part (*a*) and the corresponding figures of merit. (*c*) Generate a digital controller $D_c(z)$ (with no ZOH) that reduces the value of t_s of part (*a*) by approximately one-half for $\zeta = 0.65$ and $T = 0.01$ s. Do first by the DIG method (*s* plane) and then by the DIR method. Obtain M_p, t_p, and t_s for $c(t)$ of the PCT and $c^*(t)$ (for both the DIG and the DIR designs) and K_m. Draw the plot of $c^*(t)$ vs. kT for both designs. (*d*) Find a digital controller $D_c(z)$ (with no ZOH) that increases the ramp error coefficient of part (*a*), with $\zeta = 0.65$ and $T = 0.01$ s with a minimal degradation of the transient response characteristics of part (*a*). Do first by the DIG method (*s* plane) and then by the DIR method. Obtain M_p, t_p, and t_s for $c(t)$ of the PCT and $c^*(t)$ for both the DIG and the DIR designs) and K_m. Draw the plots of $c^*(t)$ vs. kT for both designs. (*e*) Summarize the results of this problem in tabular form and analyze. *Note*: For part (*b*) use 4-decimal-digit accuracy; for parts (*c*) and (*d*) use 4-decimal-digit accuracy; then repeat using 8-digit accuracy. Compare the results.

Answers to Selected Problems

CHAPTER 2

2.1. (*a*) *Loop equations*: Let $R = R_2 + R_3$ and $R' = R_1/R$

$$e(t) = \left(R_1 + LD + \frac{1}{CD}\right)i_1 - \left(LD + \frac{1}{CD}\right)i_2 \tag{1}$$

$$0 = -\left(LD + \frac{1}{CD}\right)i_1 + \left(R + LD + \frac{1}{CD}\right)i_2 \tag{2}$$

(*b*) *Node equations*:
Node v_a:

$$\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{LD}\right)v_a - (CD)v_b - \left(\frac{1}{R_2}\right)v_c = \left(\frac{1}{R_1}\right)e \tag{3}$$

Node v_b:

$$-(CD)v_a + \left(\frac{1}{LD} + CD\right)v_b = 0 \tag{4}$$

Node v_c:

$$-\left(\frac{1}{R_2}\right)v_a + \left(\frac{1}{R_2} + \frac{1}{R_3}\right)v_c = 0 \tag{5}$$

(*c*) *State equations*: Let $x_1 = y_1 = i/CD$, $x_2 = i_L$, $x_2 = Di_L$, $i_c = CDy_1 = Cx_1$, $i_1 - i_2 = i_c = i_L$

$$\mathbf{A} = \begin{bmatrix} 0 & \dfrac{1}{C} \\ -\dfrac{1}{L} & -\dfrac{R'}{(R'+1)L} \end{bmatrix} \tag{6}$$

$$\mathbf{b}^T = \begin{bmatrix} 0 \\ \dfrac{1}{(R'+1)L} \end{bmatrix} \qquad \mathbf{c}^T = [1 \quad 0] \tag{7}$$

(*d*) *Note*: Utilizing Eqs. (1), (2), (8), and (9), obtain

$$x_2 = C\dot{x}_1 \tag{8}$$

$$\dot{x}_2 = C\ddot{x}_2 \tag{9}$$

$$G(D) = \frac{y_1(t)}{u(t)} = \frac{1}{[(R'+1)LC]D^2 + [R'C]D + [R'+1]} \tag{10}$$

2.2. (*a*) *Loop equations*:

$$E = \left(R_1 + \frac{1}{C_1 D}\right) i_1 - \left(\frac{1}{C_1 D}\right) i_2 \tag{1}$$

$$0 = -\left(\frac{1}{C_1 D}\right) i_1 + \left(R_2 + LD + \frac{1}{C_1 D} + \frac{1}{C_2 D}\right) i_2 \tag{2}$$

(*b*) Node 1 between R_1 and R_2, node 2 between R_2 and L, and node 3 between L and C_2

$$\text{Node 1:} \quad \left(\frac{1}{R_1} + C_1 D + \frac{1}{R_2}\right) v_1 - \left(\frac{1}{R_2}\right) v_2 = \left(\frac{1}{R_1}\right) E$$

$$\text{Node 2:} \quad -\left(\frac{1}{R_2}\right) v_1 + \left(\frac{1}{R_2} + \frac{1}{LD}\right) v_2 - \left(\frac{1}{LD}\right) v_3 = 0$$

$$\text{Node 3:} \quad -\left(\frac{1}{LD}\right) v_2 + \left(\frac{1}{LD} + C_2 D\right) v_3 = 0$$

(c) Let $x_1 = v_1 = v_{c1}, x_2 = v_3 = v_{c2}, x_3 = i_L$

$$\mathbf{A} = \begin{bmatrix} -\dfrac{1}{R_1 C_1} & 0 & -\dfrac{1}{C_1} \\ 0 & 0 & \dfrac{1}{C_2} \\ \dfrac{1}{L} & -\dfrac{1}{L} & -\dfrac{R_2}{L} \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} \dfrac{1}{R_1 C_1} \\ 0 \\ 0 \end{bmatrix}$$

2.3. Circuit 2: (b)

$$\dot{\mathbf{x}} = \begin{bmatrix} -\dfrac{R_1 + R_2}{L} & -\dfrac{1}{L} \\ \dfrac{1}{C} & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} \dfrac{1}{L} \\ 0 \end{bmatrix} \mathbf{u}$$

2.6. (c)

$$G(D) = \frac{K_2}{D^4 + BD^3 + KD^2 + BK_a D + (K_a K_b - K^2)}$$

where $K_a = K_1 + K_2, K_b = K_2 + K_3, K = K_a + K_b$
(d) Let

$$\begin{aligned} x_1 &= x_a \qquad x_2 = \dot{x}_1 \qquad x_3 = x_b \\ x_4 &= \dot{x}_3 \qquad u = f(t) \quad y(t) = x_b \end{aligned}$$

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -K & K_2 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ K_2 & 0 & K_b & -1 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \qquad \mathbf{c} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

2.10. (a) $(LD + R)i = e, \quad k_i i = f, \quad \dfrac{x_a}{l_1} = \dfrac{x_b}{l_2}$

$$(M_2 D^2 + B_2 D + K_2)x_b = f_2$$

$$(M_1 D^2 + B_1 D + K_1)x_a + \frac{l_2}{l_1} f_2 = f = k_i i$$

Let

$$\frac{l_2}{l_1} = a, \quad \frac{M_1}{a} + aM_2 = b, \quad \frac{B_1}{a} + aB_2 = c, \quad \text{and} \quad \frac{K_1}{a} + aK_2 = d$$

$$(bD^2 + cD + d)x_b = k_i i$$

(b) There are only three independent states:

$$x_1 = x_b, \quad x_2 = \dot{x}_b, \quad x_3 = i, \quad \text{and} \quad u = e$$

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

$$\dot{x}_2 = -\frac{d}{b}x_1 - \frac{c}{b}x_2 + \frac{K_1}{b}x_3$$

$$\dot{x}_3 = \frac{R}{L}x_3 + \frac{l}{L}u$$

2.11. Traction force $= MD^2x + BDx$

2.12. (b) $T(t) = (B_1D)\theta_a - (B_1D)\theta_b$

$$0 = -(B_1D)\theta_a + [JD^2 + (B_2 - B_1)D + K_1]\theta_b$$

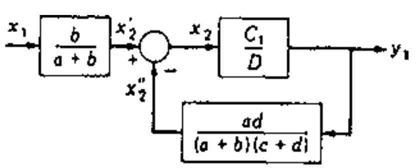
(c) Let $J = J_1 + J_2$, $y_1 = x_1 = \theta_a$, $y_2 = x_2 = \theta_b$, $x_3 = x_2$, and $u = T$

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & \dfrac{-K}{J} & \dfrac{B_1 - B_2}{J} \end{bmatrix}\mathbf{x} + \begin{bmatrix} \dfrac{1}{B_1} \\ 0 \\ \dfrac{1}{J} \end{bmatrix}\mathbf{u} \qquad \mathbf{y} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\mathbf{x}$$

(d) $$\frac{y_1}{u} = \frac{JD^2 + B_2D + K}{B_1D[JD^2 + (B_2 - B_1)D + K]}$$

$$\frac{y_2}{u} = \frac{1}{JD^2 + (B_2 - B_1)D + K}$$

2.18. (a) $$\left[\frac{C_1ad}{(a+b)(c+d)} + D\right]y_1 = \frac{C_1b}{a+b}x_1$$



2.21. $K_Tv_1(t) = JL_aD^3\theta_m + (JR_a + BL_a)D^2\theta_m + (BR_a + K_TK_b)\theta_m$

CHAPTER 3

3.2. $\theta_b(t) = -2 + 2t + A_1e^{m_1t} + A_2e^{m_2t}$
where

$$A_1 = 2.01015 \qquad A_2 = -0.01015$$

Therefore

$$\theta_b(t) = -2 + 2t + 2.010105e^{-1.071t} - 0.01015e^{-15.06t}$$

Check at $t = 0$: $\theta_b(0) = -2 + 2.01015 - 0.01015 = 0$

3.5. (*b*) $c(t) = -0.36 + 0.2t + A_1 e^{m_1 t} + A_2 e^{m_2 t} + A_3 e^{m_3 t}$
Utilizing the initial conditions yields

$$A_1 = 0.5 \qquad A_2 = (\sqrt{2}/10)\angle -171.9^\circ. = A_3^*$$

3.7. (*a*) Physical variables: $x_1 = v_c$ and $x_2 = i_L$.

$$\mathbf{A} = \begin{bmatrix} -10^4 & -10^6 \\ 0.1 & -465 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 10^4 \\ 0 \end{bmatrix} \qquad \mathbf{c}^T = [1 \quad 0]$$

(*b*) $v_c(t) = 48.95 + A_1 e^{m_1 t} + A_2 e^{m_2 t}$ where $Dv_c = m_1 A_1 e^{m_1 t} + m_2 A_2 e^{m_2 t}$
Inserting the initial conditions yields

$$0 = 48.95 + A_1 + A_2 \qquad 0.5 \times 10^6 = -475.5A_1 - 9989.5A_2,$$
$$A_1 = 1.15 \qquad A_2 = -50.1$$

3.8. (*b*) $x(t) = 3 + 1.5e^{-3t} - 4.5e^{-t}$

(*c*) $x(t) = 0.18 + 0.2353t + e^{-0.5t}[-0.18\cos(2t) - 0.16\sin(2t)]$
$= 0.18 + 0.2353t + 0.242e^{-0.5t}\sin(2t - 132^\circ)$

3.13.

$$\boldsymbol{\Phi}(t) = \begin{bmatrix} -e^{-2t} + 2e^{-4t} & 2e^{-2t} - 2e^{-4t} \\ -e^{-2t} + e^{-4t} & 2e^{-2t} - e^{-4t} \end{bmatrix}$$

$$\mathbf{x}(t) = \begin{bmatrix} \frac{1}{2} - 3e^{-2t} + \frac{9}{2}e^{-4t} \\ \frac{3}{4} - 3e^{-2t} + \frac{9}{4}e^{-4t} \end{bmatrix}$$

$$y(t) = x_1 = \frac{1}{2} - 3e^{-2t} + \frac{9}{2}e^{-4t}$$

3.16. (*a*) $\mathbf{x}(t)_h = \boldsymbol{\Phi}(t)\mathbf{x}(0) = \begin{bmatrix} e^{-2t} \\ -2e^{-2t} \end{bmatrix}$

(*b*) $\mathbf{x}(t) = \begin{bmatrix} 1 - 2e^{-t} + 2e^{-2t} \\ -2e^{-t} \end{bmatrix}$

3.17. (1) (*a*) $\lambda_1 = -1$, $\lambda_{2,3} = -1 \pm j2$

$$\boldsymbol{\Phi}(t) = \mathbf{Z}_1 e^{-t} + \mathbf{Z}_2 e^{-t} e^{jt} + \mathbf{Z}_3 e^{-t} e^{-jt}$$

CHAPTER 4

4.1. (a) $f(t) = e^{-t}[t + 0.333 \sin(3t + 180^\circ)]$
(b) $f(t) = 1.3846 - e^{-5t} + 1.9612 e^{-2t} \sin(3t + 191.3^\circ)$

4.2. (c) $x(t) = 1 - 0.9898 e^{-t} + 0.01701 e^{-2t} \sin(3t + 216.9^\circ)$

4.4. (b) With zero initial conditions, the Laplace transform yields:

$$X_b(s) = \frac{12}{s(s+10)(s+1.2874)(s+223.71)} = \frac{A}{s} + \frac{B}{s+10} + \frac{C}{s+1.2874} + \frac{D}{s+223.71}$$

where

$$A = 4.167 \times 10^{-3} \qquad B = 0.06445 \times 10^{-3}$$
$$C = -4.81 \times 10^{-3} \qquad D = -1.128 \times 10^{-6}$$

$$x_b(t) = 4.167 \times 10^{-3} + 0.6445 \times 10^{-4} e^{-10t} - 4.81 \times 10^{-3} e^{-1.2874t} - 1.128 \times 10^{-6} e^{-223.7t}$$

4.5. (d) $F(s) = \dfrac{1}{s^2} - \dfrac{7}{6s} + \dfrac{5}{4(s+1)} - \dfrac{0.5}{6(s+3)}$
(h) $F(s) = \dfrac{1}{s} + \dfrac{9}{s+3} + \dfrac{28}{s+3-j1} + \dfrac{30}{s+3+j1}$

4.6. (b) $i_L(t) = 0.4833 e^{-3.684t} - 0.5534 \sin(\omega_d t + 61^\circ)$

4.7. (b) $(s^2 + 2.8s + 4)X(s) = \dfrac{2(s^2 + 4.3s + 6.13)}{s}$

$$x(t) = 3.0669 + 1.5 e^{-1.4t} \sin(1.416t - 45.32^\circ)$$

(d) $(s^5 + 4s^4 + 34s^3 + 110s^2 + 225s + 250)X(s) + 4s^2 - s - 56 = \dfrac{5}{s^2 + 25}$

$$x(t) = 10.837 e^{-t} \sin(2t - 85.233^\circ) + 0.99974 \sin(5t + 180^\circ) + 14.799 e^{-2t}$$

4.8. (a) Prob. 4.1(a): 0, Prob. 4.1(b): $\frac{18}{13}$
(b) Prob. 4.2(a): 0, Prob. 4.2(c): 1

4.11. (1) $(b)\ x(t) = 3 + 3.8348e^{-t}\sin(t + 237.53^\circ) + 0.2353e^{-5t}$

$(c)\ x(t) = 0.30t - 0.260 + 0.2712e^{-t}\sin(t + 102.53^\circ) - 0.0047e^{-5t}$

4.12. System (1)

$$(a)\ \mathbf{\Phi}(s) = [s\mathbf{I}-\mathbf{A}]^{-1} = \begin{bmatrix} s+5 & -1 \\ -2 & s+2 \end{bmatrix}^{-1} = \frac{1}{\Delta}\begin{bmatrix} s+2 & 1 \\ 2 & s+5 \end{bmatrix}$$

$$\Delta = s^2 + 7s + 8 = (s + 5.62)(s + 1.438)$$

$$\mathbf{X}(s) = \begin{bmatrix} X_1(s) \\ X_2(s) \end{bmatrix} = \frac{1}{\Delta}\begin{bmatrix} s^2 + 3s + 1 \\ s^3 + 12s^2 + 35s + 1 \end{bmatrix}$$

$$\mathbf{X}(s) = [s\mathbf{I} - \mathbf{A}]^{-1}\mathrm{x}(0) + [s\mathbf{I} - \mathbf{A}]^{-1}\mathbf{b}\mathrm{U(s)}$$

$(b)\ G(s) = \mathbf{C}\Phi(s)\mathbf{B} + \mathbf{D}$

$$G(s) = \frac{1}{\Delta}[01]\begin{bmatrix} s+2 & 1 \\ 2 & s+5 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} + 0$$
$$= \frac{s+5}{s^2 + 7s + 8} = \frac{s+5}{(s + 5.62)(s + 1.438)}$$

$$(c)\ \mathbf{Y}(s) = \mathbf{CX}(s) = [\,0 \quad 1\,]\mathbf{X}(s) = X_2 \to y(t) = x_2(t)$$
$$= 0.74836e^{-1.4384t} - 0.37336e^{-5.5616t} + 0.6250$$

As a check: $x_i(0) = \lim_{s\to\infty}[sX_i(s)] = 1$ and $x_2(0) = \lim_{s\to\infty}[sX_1(s)]$
$= 1[sX_2(s)] = 1$ and $y(\infty) = 0.6250.$

4.15. (1) $\sigma(\mathbf{A}) = \{-1 + j2, -1 - j_2, -1\}$

CHAPTER 5

5.5. $$G(s) = \frac{-(60s + 3.8)}{(s + 10)(s^2 + 0.15s + 8.0054)}$$

5.11. (2) $\lambda = -2, \quad \lambda = -3, \quad \lambda = -4, \quad \lambda = -5$

$(a)\ y = [8 \quad 0 \quad 4 \quad 0]\mathbf{x}$

$(b)\ y = [4 \quad -22 \quad 36 \quad -18]\mathbf{x}$

5.12. (*a*) $G(s) = \dfrac{-10}{(s+2)(s+3)}$ (*c*) $\mathbf{x}(t) = \begin{bmatrix} -\frac{5}{3}(1 - 3e^{-2t} + 2e^{-3t}) \\ \frac{10}{3}(1 - e^{-3t}) \end{bmatrix}$

CHAPTER 6

6.2. (*a*) $0 < K < 1280/9$; $s_{1,2} = \pm j2.5827$; for $K = 142.22$

(*c*) $0 < K < 30$; $s_{1,2} = \pm j2.2356$; for $K = 30$

6.3. (*a*) $(s+2)(s+2.2 \pm j2.2)$ (*c*) $(s+1 \pm j1.414)(s-1 \pm j1.414)$

(*f*) $(s-1)(s+2)(s-3)(s+5)$

6.4. (*a*) None (*c*) None (*f*) 2

6.10. (*a*) For (1): $K_p = \infty, K_v = \frac{5}{13}, K_a = 0$, (*b*) For $r_1(t)$: (1) $e(t)_{ss} = \frac{13}{5}$ (2) $e(t)_{ss} = 0$

6.12. (*a*) 20, 0, 0 (*d*) ∞, 6, 0

6.13. (*a*): (*a*) $\frac{5}{2}$ (*b*) ∞ (*c*) ∞
(*d*): (*a*) 0 (*b*) $\frac{1}{3}$ (*c*) ∞

6.19. $K_G c_i = a_i + K_h K_G c_i > 0$ for a stable system and $K_G c_0 = a_0 + K_h K_G c_0$ for $e(t)_{ss} = 0$

6.25. (*b*) $K = 4$: (*b*) $K = 7$: and $K_1 = 6.666\ \text{s}^{-1}$ $e(t)_{ss} = 0.3$;

6.27. (*a*) For $K = 100$: $K_1 = 68.97\ \text{s}^{-1}$ and $e(t)_{ss} = 0.0145$; (*b*) for $K = 1000$: $K_1 = 689.7\ \text{s}^{-1}$ and $e(t)_{ss} = 0.00145$; (*c*) unstable for $K = 1000$; stable for $K = 100$ and the corresponding FOM are $M_p = 1.345$, $t_p = 1.06$ s, and $t_s = 3.286$ s.

CHAPTER 7

7.2. (*c*) For $K > 0$, real-axis branches: none; four branches; $\gamma = \pm 45^\circ$, $\pm 135^\circ$; angles of departure: $\pm 45^\circ$, $\pm 135^\circ$; imaginary-axis crossing: $\pm j2$ with $K = 0$. Branches are straight lines and coincide with the asymptotes.

(*e*) Real-axis branch: 0 to $-\infty$; five branches; $\gamma = \pm 36^\circ$, $\pm 108^\circ$, 180°; $\sigma_0 = -1.6$; no breakaway or break-in points; angles of departure: $\pm 71.7^\circ$, $\pm 33.8^\circ$; and the imaginary-axis crossing: $\pm j1$ with $K = 30$. Locus crosses the $\pm 36^\circ$ asymptotes.

7.4. (*a*) For $K > 0$, real-axis branches: 0 to -10; three branches; $\gamma = \pm 90^\circ$; $\sigma_0 = 0$; break-in point: -3.9113 ($K = 3.33095$); breakaway point: -2.78698 ($K = 3.4388$); and imaginary-axis crossing: none.

(*b*) $K = 0.05427$

(c) $e(t) = 1.5951e^{-2.2164t}\sin\,(2.1992t + 207.274^\circ)$
(d)

ω	M	α, deg	ω	M	α, deg
0	1	0	1.4	0.9586	−44.7
0.2	0.9995	−6.1	1.8	0.9153	−58.5
0.4	0.998	−12.3	2.0	0.8854	−65.5
0.6	0.9951	−18.5	2.4	0.8103	−79.3
0.8	0.9903	−24.9	2.8	0.7202	−92.3
1.0	0.9832	−31.3	3.2	0.6252	−104.1

7.7. Poles: $p_1 = -1,\ p_2 = -7,\ p_{3,4} = 4 \pm j3$
(*a*) For $K > 0$, real-axis branch: -1 to -7; four branches; $\gamma = \pm 45^\circ$, $\pm 135^\circ$; angles of departure: 0°, -180°, $\pm 90^\circ$; imaginary-axis crossing $\pm j4$ with $K = 1105$; break-in and breakaway pints coincide at $s = -4$. For $K < 0$, real-axis branches: -1 to $+\infty$ and -7 to $-\infty$; $\gamma = \pm 90^\circ, 180^\circ$; angles of departure; 0°, $\pm 90^\circ$, 180°; imaginary-axis crossing: $s = 0$ with $K = -175$.
(b) $-175 < K < 1105$ (c) $K = 81$

7.8. (1) **(*a*)** For $K = -50K_1$: three branches; $\gamma = \pm 90^\circ$ for $K_1 < 0$; $\gamma = 180^\circ$ for $K_1 > 0$; $\sigma_0 = -9.5$; breakaway points: -7.72 for $K > 0$, -3.139 for $K < 0$; break-in point: 5.362 for $K < 0$; imaginary-axis crossings: at $s = 0$ for $K = 50$ and $\pm j3.6993$ for $K = -66.32$.
(*b*) Stable for $-66.32 < K < 50$

7.11. **(*a*)** ***Real-axis branches***: -3.2 to $-\infty$ (***Note***: The canceled pole and zero at $s = -5$ is a root of the characteristic equation and is not utilized in determining the root locus for locating the remaining roots.) **6-*branches***: $\gamma = \pm 36^\circ,\ \pm 108^\circ,\ 180^\circ,\ \sigma_0 = -22.5$; ***Break-in point***: $s = -16.5$; ***Angle of departures*** (given in order of closeness to the imaginary axis in the 2nd quadrant): 144.4°, -94.4°, 4.2°; ***Imaginary-axis crossing:*** $\pm j17.20$ for $K = 1.869 \times 10^7$.
(*b*) $K_G K_H = 1.0735$; $p_{1,2} = \pm j17.21$, $p_{3,4} = 33.28 \pm j28.05$, $p_5 = -4.634$, $p_6 = -44.46$, $p_7 = -5$

(c) $$\frac{C(s)}{R(s)} = \frac{1.869 \times 10^7 (s + 3.2)}{(s + 4.634)(s + 5)(s + 44.46)(s \pm j17.21)(s + 33.28 \pm j28.05)}$$

7.12. (1) (a) $A = 1.44$

CHAPTER 8

8.1. (*b*) (1) $\mathbf{G}(j0+) = \infty\angle{-90^\circ}$, $\mathbf{G}(j\infty) = 0\angle{-270^\circ}$, $\omega_x = \omega_c = 2.236$, $\mathbf{G}(j\omega_x) = -0.2223$, $\omega_\phi = 0.867$
(2) $\mathbf{G}(j0+) = \infty\angle{-90^\circ}$, $\mathbf{G}(j\infty) = 0\angle{-360^\circ}$, $\omega_x = \omega_c = 9.75$
$\mathbf{G}(j\omega_x) = -0.0952$, $\omega_\phi = 0.4977$

8.4. (1) (*b*) $\mathbf{G}(j0+) = \infty\angle{-90^\circ}$, $\mathbf{G}(j\infty) = 0\angle{-270^\circ}$, $\omega_x = 35.4$, $\mathbf{G}(j\omega_x) = -0.05645$

8.6. (*a*) It must have an initial slope of -40 dB/decade
(*b*) The phase-angle curve approaches an angle of 180° as $\omega \to 0$

8.9. Case II:

$$G(s) = \frac{2337.5(1 + 0.25s)}{s(1 + 5s)(1 + 0.125s)(1 + 0.05s)}$$

8.10. (*a*) $K_0 = 31.7$ (*b*) $K_2 = 2.56 = (\omega_y)^2$

8.11. (*a*) Plant 1

$$G(s) = \frac{0.4}{s(s + 1 \pm j4)}$$

8.13. (*a*) $N = -2$, $Z_R = 2$, unstable (c) $N = 0$, $Z_R = 0$, stable

8.14. (*a*) For stability $K > 1.26759$

8.15. (*a*) Stable for $K > -2$ (*e*) Stable for $18 < K < 90.12$

8.16. (1) (*a*) $K_1 = 5.5654$, $\omega_\phi = 4.845$ (*b*) $K_1 = 3.345$, $\omega_\phi = 3.15$
(*c*) maximum $K_1 = 1.9429 \times 10^5$
(2) (*a*) $K_0 = 20.13$, $\omega_\phi = 8.3$ (*b*) $K_0 = 8.58$, $\omega_\phi = 4.61$
(*c*) Stable for $0 < K_0 < \infty$

CHAPTER 9

9.1. (2) $K_1 = 7.94$, $M_m = 1.326$, $\omega_m = 7.81$, $\omega_\phi = 6.94$, $\omega_c = 15.81$
(4) $K_1 = 0.7075$, $M_m = 1.318$, $\omega_m = 0.658$, $\omega_\phi = 0.6054$, $\omega_c = 1.508$

9.3. $M_m = 1.30$, $K_1 = 2.975$, $\omega_m = 2.75$

9.5. (1) (a) $M_m = 1$, $\omega_m = 0$ (*b*) $M_m = 1.326$, $\omega_m = 5.4$ (*c*) $M_m = 1.016$, $\omega_m = 2.24$

9.8. (4) (*a*) $M_m = 5.8904$, $\omega_m = 3.82$ (*c*) $K_0 = 0.18$ for $M_m = 1.12$
(*d*) $\omega_m = 0.39$, $\gamma = 56.16^\circ$

9.11. Using a CAD program: (*a*) $M_m = 1.11293$, $\omega_m = 0.435$ (*b*) $K = 1.338$
(*c*) $a = 6$

CHAPTER 10

10.3, 10.18.

System (1)	Dominant roots	Other roots	$K_{1,\,s^{-1}}$	T_p,s	t_s,s	M_o
Basic	$-1.25 \pm j2.166$	-6.501	1.72	1.67	3.36	0.14
Lag-compensated:						
$\alpha = 10,\ T = 2$	$-1.03 \pm j1.787$	-5.267 -0.726	16.16	1.8	5.37	0.41
Lead-compensated:						
$\alpha = 0.1,\ T = 0.25$	$-2.23 \pm j3.86$	-3.123 -40.42	3.13	0.916	1.78	0.10
Lag-lead-compensated:						
$\alpha = 10,\ T_1 = 2,\ T_2 = 0.25$	$-2.08 \pm j3.6$	-0.598 -2.897 -40.40	30.18	0.967	4.15	0.244
Tachometer-compensated:						
$A = 1.61,\ K_t = 1,\ \zeta = 0.5$	$-3.40 \pm j5.89$	-1.2	1.02	—	3.39	0.0
$A = 4.79,\ K_t = 1,\ \zeta = 0.3$	$-1.86 \pm j5.92$	-4.28	2.97	0.756	1.89	0.13

10.6. (*a*) $a = 4, b = 2.6653$ (*b*) $\alpha = 1.126$ (*c*) lag compensator

10.7. (*a*) For the uncompensated system: from the performance specifications, for $\zeta = 0.5$, are $M_p = 1.16$, $t_p = 2$ s, $t_s = 4.39$ s, $K = 2.933$, $K_1 = 1.76\ \mathrm{s}^{-1}$, and the control ratio poles are $-0.9235 \pm j1.601$ and -5.153.

(*b*) $|\sigma| \geq 4/T_s = 2$; assuming $a = 2$, $\zeta = 0.5$, and $|\sigma| = 2$ yields $b = 4.726$

(*c*) $A = 5.21$ and $\alpha = 0.4232$

(*d*) A lead compensator; $M_p = 1.59$, $t_p = 0.917$ s, $t_s = 2.02$ s, and $K_1 = 3.88\ \mathrm{s}^{-1}$

10.18. (*a*) For $\zeta = 0.5$ the uncompensated system yields the following data: $K = 25.01, K_1 = 5\ \mathrm{s}^{-1}, M_p = 1.163, t_p = 0.725$ s, $t_s = 1.615$ s.

(*b*) Based upon the desired values of ζ and $\sigma = -7$, the desired dominant poles are $P_{1,2D} = -7 \pm j12.124$. One possible solution is: let $a = 40$, $b = 0.01$, $c = 0.1$; by applying the angle condition and the use of a CAD package the following results are obtained: $A = 6.0464$, $K_h = 11.18$, $K_1 = 126.1\ \mathrm{s}^{-1}$, $M_p = 1.161$, $t_p = 0.27$ s, $t_s = 0.426$ s.

10.20. See Prob. 10.3 for basic system performance data. Based on the desired performance specifications, the desired dominant poles are: $P_{1,2D} = -4 \pm j6.928$. Let $K_x = 1250\,K_1$ be the gain term of $G(s)$. By applying method 1, the following results are obtained: $K_x = 760.71$, $K_t = 0.5619$, $b = 11.895$, $M_p = 1.092$, $t_p = 0.725$ s, $t_s = 0.97$ s, $K_1 = 2.369\ \text{s}^{-1}$. The other poles are: $-6.296 \pm j4.486$ and -11. The plot of $c(t)$ has the desired simple second-order time response characteristics due to the effect of the real root.

CHAPTER 11

11.1. (*a*) $K = 11935$, $\omega_\phi = 3.10$, $\gamma = 51.16^\circ$, $M_p = 1.161$, $t_p = 0.8654$ s, $t_s = 1.807$ s, $K_1 = 3.6387\ \text{s}^{-1}$.

(*b*) $1/T = 0.302$, $K = 11935$, $A = 7.7084$, $M_m = 1.18$, $M_p = 1.203$, $t_p = 1.104$ s, $t_s = 5.455$ s, $K_1 = 28.05\ \text{s}^{-1}$; $\omega_\phi = 2.53, \gamma = 51.1^\circ$.
Results for an $M_m = 1.16$ are not as good as for $M_m = 1.18$.

(*c*) $T = 0.25$, $K = 213200$, $M_m = 1.564$, $M_p = 1.186$, $t_p = 0.3125$ s, $t_s = 1.273$ s, $K_1 = 6.50\ \text{s}^{-1}$; for $M_m = 1.16$: $K_x = 175800$, $M_p = 1.076$, $t_p = 0.3372$ s, $t_s = 1.105$ s, $K_1 = 5.36\ \text{s}^{-1}$, $\omega_\phi = 7.10$, $\gamma = 66.46^\circ$.

(*d*) $T_1 = 3.311$, $T_2 = 0.25$, $K_x = 209000$, $M_m = 1.564$, $M_p = 1.212$, $t_p = 0.3180$ s, $t_s = 2.857$ s, $K_1 = 63.72\ \text{s}^{-1}$; for $M_m = 1.16$: $K = 175800$, $M_p = 1.104$, $t_p = 0.3442$ s, $t_s = 3.526$ s, $K_1 = 54.87\ \text{s}^{-1}$, $\omega_\phi = 7.35, \gamma = 62.71^\circ$.

11.5. (*a*) $G(s) = \dfrac{1}{8K}\left[\dfrac{1+Ts}{1+\alpha Ts}\right]$, where $T = \dfrac{B}{K}$ and $\alpha = 0.5$

11.12. (*a*) At $\omega = 8$: $\mathbf{G}(j\omega) = \mathbf{C}(j\omega)/\mathbf{E}(j\omega) \approx 6.3\angle{-119^\circ}$, thus $K_1 = 250$, $\omega_\phi = 8$, and $\gamma \geq 40^\circ$.

(*b*) $G_c = \dfrac{(s+1)(0.5s+1)}{(10s+1)(0.05s+1)}$

11.17. (*a*) $K_{2x} = 2.02$, $\omega_{\phi x} = 2.15$, $M_m = 1.398$, $M_p = 1.29$, $t_p = 1.39$ s, $t_s = 3.03$ s

(*c*) For $\gamma = 50^\circ$: $u = \omega_\phi T = 8.6T = 1.19$ yields $T = 0.1384$. From the graphical construction, with $\omega_1 \leq \omega_\phi/20$ and $\omega_2 = 20\,\omega_\phi = 172$, obtain $K_t = 24.28$ and $A = 180$, thus

$$\frac{C(s)}{E(s)} = \frac{2907(s+1)(s+7.225)}{s^2(s+1.108)(s+406.3)}$$

The 0-dB crossing slope is -12 dB/oct at $\omega = 6$, but the zero at -7.225 will have a moderating effect on the degree of stability.

(*d*) $M_m = 1.47$, $M_p = 1.30$, $t_p = 0.34$ s, $t_s = 1.03$ s, $K_2 = 46.65$; since the values of M_p are essentially the same, a good improvement in system performance has been achieved. $\gamma = 51^\circ$ and $\omega_\phi = 9.12$.

CHAPTER 12

12.2. $$G_c(s)\frac{2.31(s+20)}{s+21}$$

With this cascade compensator $M_p = 1.12$, $t_p = 1.493$ s, $t_s = 2.941$ s, $K_1 = 1.53$. The specified $M_T(s)$ yields $M_p = 1.123$, $t_p = 1.493$, $t_s = 2.948$, $K_1 = 1.424$.

12.4. As a first trial, use

$$G_c \approx \frac{2.994(s+2.1)}{s+2.223}$$

With this cascade compensator $M_p = 1.03$, $t_p = 2.37$ s, $t_s = 2.94$ s, $K_t = 0.75$. The specified $M_T(s)$ yields $M_p = 1.06$, $t_p = 2.2$ s, $t_s = 3.04$ s, $K_1 = 0.79$.

12.8. (*a*) The pole at the origin of $H(s)$ ensures $c(t)_{ss} = 0$; make pole $-b$ nondominant: $b = 100$

(*b*) $K_H = 2500$

(*c*) Chxaracteristic equation: $s^4 + 104s^3 + 10{,}400s^2 + 120{,}000s + 320{,}000 = 0$. Routh stability criterion reveals that all roots lie in the left-half s plane. A plot of Lm G_xH reveals that its 0-dB axis crossing slope is -6 dB/oct, indicating a stable system. The system poles are $p_1 = -4$, $p_2 = -8.06$, $p_{3,4} = -45.97 \pm j311.7$. $|y(t_p)| = 0.0025$, $t_p = 0.163$ s. For $|y(t_p)| = 0.01$, $K_H = 630$, which produces very little change in the value of t_p.

CHAPTER 13

13.3. (1) (*a*) Not observable, (*b*) controllable, (*c*) $G(s) = 2/(s-2)$, (*d*) one observable and two controllable states, (*e*) unstable

(4) (*a*) Observable, (*b*) controllable, (*c*) $G(s) = (s-1)/(s+1)\,(s+2)$, (*d*) Two observable and two controllable states, and (*e*) stable

13.5. (*a*) Add a cascade compensator $G_c(s) = 1/(s+1)$

(*b*) $H_{eq}(s) = k_3s^2 + (2k_3 + k_2)s + k_1$

(*d*) A desired control ratio is

$$M_T(s) = \frac{Y(s)}{R(s)} = \frac{A}{(s^2+4s+13)(s+25)}$$

(*e*) With $A = 162.5, \mathbf{k}^T = [1 \quad 0.16923 \quad 0.08]$
(*g*) $G_{\text{eq}}(s) = 325/[s(s^2+29s+113)],\ K_1 = 2.876$
(*h*) $M_p = 1.12, t_p = 1.09$ s, $t_s = 1.66$ s

13.8. For system (1) (*a*)

$$M(s) = \frac{11{,}700}{(s+12 \pm j21)(s+20)}$$

$$k_1 = 1 \qquad k_2 = 0.091026 \qquad k_3 = 0.003077$$

(*b*) $y(t) = 1 + 1.025e^{-12t}\sin(21t + 171.1^\circ) - 1.1584e^{-20t}$
$M_p = 1.057, \quad t_p = 0.215\,\text{s}, t_s = 0.273\,\text{s}, K_1 = 11\,\text{s}^{-1}$

13.12. (*a*)

$$M_T(s) = \frac{Y(s)}{R(s)} = \frac{1.467(s+1.5)(s+2)}{(s^2+2s+2)(s+1.1)(s+2)}$$

where $G_c(s)\dfrac{s+1.5}{s+1}$, $A = 1.467$

$Y(s)/R(s)$

$$= \frac{A(s+1.5)(s+2)}{s^4 + [7 + A(0.5k_4 + k_3 + k_2)]s^3 + [11 + A(3k_4 + 2.5k_3 + 3.5k_2 + k_1)]s^2 + [5 + A(2.5k_4 + 1.5k_3 + 3k_2 + 3.5k_1)]s + 3Ak_1}$$

where $k_1 = 1,\ k_2 = 0.8635,\ k_3 = -2.43541,\ k_4 = 0.5521$
(*c*) $M_p = 1.015,\ t_p = 3.75$ s, $t_s = 2.79$ s, $K_1 = 0.524\ \text{s}^{-1}$

13.16. (*a*)

$$M_T(s) = \frac{195(s+2)}{(s^2+2s+2)(s+1.95)(s+100)}$$

(*b*) By applying the Guillemin-Truxal method, a reduced-order compensator is obtained by dividing the denominator of $G_c(s)$ by its numerator to obtain $G_c(s) \approx 1/(s+99.9)$. The figures of merit for $M_T(s)$ are $M_p = 1.042$, $t_p = 3.18$ s, and $t_s = 4.23$ s, where for the system implemented with the reduced-order compensator $M_p = 1.056$, $t_p = 3.12$ s, and $t_s = 4.44$ s.

13.19. $\mathbf{k}^T = \begin{bmatrix} -8 & -2.5 \end{bmatrix}$

$$x_1(t) = \frac{5}{3}e^{-3t} - \frac{1}{3}e^{-6t} \qquad x_2(t) = -5e^{-3t} + 4e^{-6t} \qquad \mathbf{x}(0) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

13.20. (1) (*a*) Completely controllable, (*b*) $\sigma(\mathbf{A}_p) = \{-1, -2, -3\}$,

$$(c)\ \mathbf{T} = \begin{bmatrix} 6 & 5 & 1 \\ 3 & 2 & 1 \\ 0 & 2 & 1 \end{bmatrix}, \quad (d)\ \mathbf{k}^T = \begin{bmatrix} -42 & -33 & -6 \end{bmatrix}$$

CHAPTER 14

14.1. For $m = 1$: (*a*)

$$S_k^T(s)|_{k=15} = \frac{s^4 + 7s^3 + 25s^2 + 39s}{s^4 + 7s^3 + 25s^2 + 54s + 75}$$

14.6. (*a*) $p = -21$, $K_G = 30.8$ (*b*) No; need a compensator of the form

$$G_c = \frac{A(s+a)}{(s+b)(s+c)}$$

14.10. For system (2), (*a*) Three equilibrium points are

$$\mathbf{x}_a = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad \mathbf{x}_b = \begin{bmatrix} -\frac{1}{2} \\ 0 \\ 2 \end{bmatrix} \qquad \mathbf{x}_c = \begin{bmatrix} -\frac{1}{2} \\ 0 \\ -2 \end{bmatrix}$$

(*b*) For $\mathbf{x}_b$,

$$\mathbf{J}_{\mathbf{x}_b} = \begin{bmatrix} 0 & 1 & 0 \\ 2 & -3 & 0.5x_3 \\ \frac{2}{3} & 0 & 2x_1 + 1 \end{bmatrix}_{x_o}$$

CHAPTER 15

15.5. (*a*) $z^{-1} + 1.8z^{-2} + 2.19z^{-3} + 2.252z^{-4} + \cdots$

15.7. (*a*) $f(0) = 0$; system is stable, thus $f(\infty) = 1.95122$

15.8. (*a*) $(1 - 2z^{-1} + 1.41z^{-2} - 0.41z^{-3})C(z) = (z^{-1} - 0.2z^{-2})E(z)$

By use of Theorem 1:

$$\begin{aligned} c(kT) &= 2c[(k-1)T] - 1.41c[(k-2)T] + 0.41c[(k-3)T] \\ &\quad + e[(k-1)T] - 0.2e[(k-2)T] \end{aligned}$$

15.11. *System 2*

$$C^*(s) = G_1G_2^*(s)M^*(s)$$

$$\frac{C^*(s)}{R^*(s)} = \frac{G_1 G_2^*(s)}{1 + G_1 H^*(s) + G_1 G_2^*(s)}$$

$$\frac{C(z)}{R(z)} = \frac{G_1 G_2(z)}{1 + G_1 H(z) + G_1 G_2(z)}$$

CHAPTER 16

16.1. (1)

	z-domain value	
s Plane	Eq. (16.8)	$Z = e^{sT}$
-1.5	0.94176	0.94176
-3	0.8868	0.88692
-20	0.4286	0.44933
$-2 \pm j3$	$0.9167 \pm j0.1106$	$0.91671 \pm j0.11054$
$10 \pm j25$	$0.4201 \pm j0.5917$	$0.39208 \pm j0616063$

16.2. (2) $z_{TU} = -0.2857, z = 0.006738$—difference due to warping
(3) $z_{TU} = -0.71598 \pm j0.11834, z = 0.20586 \pm j0.69589$—difference due to warping

16.3. (1)

$$\left[\frac{C(s)}{R(s)}\right]_T = \frac{Ks + 8K}{(s^2 + 10s + 174)(s + 10)} \to 8K$$

$$= (10)(174) \to K = 217.5$$

(*b*) *For* $T = 0.1 \to \pm j(0.6114/T) = \pm j6.114$ and $(-0.1)/T = -1 \to$ thus, the zero and poles of the desired control ratio do not lie in the good Tustin region of Fig. 16.5. Thus:

$$\frac{0.6114}{T} \geq 7 \to T \leq 0.08734, \text{ and } \frac{0.1}{T} \geq 10 \to T \leq \frac{0.1}{10} \leq 0.01;$$

therefore, $T \leq 0.01$.

16.5.

$$D_c(z) = \frac{K_c(z - 0.6703)}{(z - 0.06703)} \quad \text{where}$$

$$K = 3.795 \times 10^{-2} K_1 - 0.2566 \to K_c = 6.7615$$

$$M_p = 1.062,\ t_p = 0.4\,\text{s},\ t_s = 0.6^+\text{s}, \quad \text{and} \quad K_1 = 5.1577\ \text{s}^{-1}$$

$$\frac{C(s)}{R(s)} = \frac{251.2(s+0.005)}{(s+0.4061 \pm j0.4747)(s+0.0005)(s+3.183)(s+200^+)}$$

$$M_p = 1.078,\ t_p = 6.98\ \text{s},\ t_s = 10.7\,\text{s},\ K_1 = 4.187,\ \text{and A} = 10$$

The remaining solutions are left to the instructor to obtain.

16.10. (*a*)

$$G_z(z) = \frac{4.967 \times 10^{-5} K_x(z+0.9934)}{(z-1)(z-0.9802)} \rightarrow K_x = 2.34105$$

$$\frac{C(z)}{R(z)} = \frac{1.163 \times 10^{-4}(z+0.9934)}{(z-0.9900 \pm j0.01152)}$$

(*b*) The PCT block diagram is similar to that of Fig. 16.14 *d* with $D_c(s) = 1$ and $G_{pa}(s)/T = 200/(s+200) = G_A(s)$.

$$G_{PC}(s) = G_A G_x(s) = \frac{200K_x}{s(s+2)(s+200)}, K_x = 2.34105$$

$$\frac{C(s)}{R(s)} = \frac{G_{PC}(s)}{1+G_{PC}(s)} = \frac{466.21}{(s+0.99408 \pm j1.159)(s+200.012)}$$

The continuous-time system control ratio, for $\zeta = 0.65$, is:

$$\frac{C(s)}{R(s)} = \frac{G_x(s)}{1+G_x(s)} = \frac{2.368}{s+1 \pm j1.170}$$

(*c*) *DIG*—To reduce t_s by about 1/2, select $D_c(s) = A(s+2)/(s+5)$; $G(s) = D_c(s)G_p(s) = -1$ for $\zeta = 0.65$: $K = 200AK_x = 2874.6$; $K_x = 2.368$, and $A = 6.0705$.

$$\frac{C(s)}{R(s)} = \frac{K}{(s+2.463 \pm j2.881)(s+200.1)};$$

$$[D_c(z)]_{TU} = \frac{5.982(z-0.9802)}{z-0.9512}$$

$$\frac{C(z)}{R(z)} = \frac{[D_c(z)]_{TU}G_z(z)}{1+[D_c(z)]_{TU}G_z(z)} = \frac{7.0359 \times 10^{-4}(z+0.9934)}{z-0.9752 \pm j0.02810}$$

DIR—For $K_x = 2.368$ and $\zeta = 0.65$, $D_c(z) = A(z-0.9802)/(z-0.09802)$ where $A = 1865.3$.

$$\frac{C(z)}{R(z)} = \frac{D_c(z)G_z(z)}{1+D_c(z)G_z(z)} = \frac{0.2194(z+0.9934)}{z-0.4393 \pm j0.3507}$$

(*d*) *DIG*—For $K_x = 2.368$ and $\zeta = 0.65$, $D_c(s) = 0.1A(s + 0.01)/(s + 0.001)$ and $A = 9.384893$:

$$\frac{C(s)}{R(s)} = \frac{D_c(s)G_{PC}(s)}{1 + D_c(s)G_{PC}(s)}$$

$$\frac{C(s)}{R(s)} = \frac{468.4(s + 0.01)}{(s + 0.98955 \pm j1.1595)(s + 0.01008)(s + 200.01)}$$

$$[D_c(z)]_{TU} = \frac{0.9385315(z - 0.9999)}{z - 0.99999}$$

$$\frac{C(z)}{R(z)} = \frac{[D_c(z)]_{TU} G_z(z)}{1 + [D_c(z)]_{TU} G_z(z)}$$
$$= \frac{1.1633 \times 10^{-4}(z + 0.9934)(z - 0.9999)}{(z - 0.990087 \pm j0.01148126)(z - 0.99989923)}$$

DIR—For $K_x = 2.368$, $\zeta = 0.65$, and

$$[D_c(z)]_{TU} = \frac{K_{zc}(z - 0.9999)}{z - 0.99999} \rightarrow K_{zc} = 0.986259 \text{ obtain:}$$

$$\frac{C(z)}{R(z)} = \frac{D_c(z)G_z(z)}{1 + D_c(z)G_z(z)} = \frac{1.1163 \times 10^{-4}(z + 0.9934)(z - 0.9999)}{(z - 0.990087 \pm j0.01145146)(z - 0.99989923)}$$

(*e*) The tables summarize the results of the various designs presented on the following page. The desired improvements using *s*-domain compensation design rules (for both *s*- and *z*-domain analysis) are obtained. Since the dominant poles and zeros lie in the good Tustin region for the lag and lead units, there is a good correlation between the DIG and DIR results.

System		Domain	Zeros of C/R	Poles of C/R
Basic (*a*) and (*b*)	Continuous	s	—	$-1.000 \pm j1.170$
	DIG	s	—	$-0.9941 \pm j1.159$, -200.012
	DIR	z	-0.9934	$0.9900 \pm j0.01152$
(*c*) Lead	DIG	s	—	$-2.463 \pm j2.881$, -200.1
		z	-0.9934	$0.9752 \pm j0.02810$
	DIR	z	-0.9934	$0.4393 \pm j0.3507$
(*d*) Lag	DIG	s	-0.01	$-0.9895 \pm j1.160$, -0.01008, -200.01
		z	-0.9934	$\pm 0.990087 \pm j0.011481$, 0.999899
	DIR	z	0.9999	$0.990087 \pm j0.001145146$, 0.999899

System		Domain	K_x	M_p	t_p s	t_s s	K_1 s^{-1}
Basic (*a*) and (*b*)	Continuous	s	2.368	1.0682	2.69	3.90	1.184
	DIG	s	2.341	1.0675	2.72	3.94	1.165
	DIR	z	2.341	1.068	2.70	3.93	1.171
(*c*) Lead	DIG	s	2.368	1.068	1.096	1.59	2.875
		z		1.068	1.09	1.58	2.874
	DIR	z		1.067	0.05	0.07	2.84
(*d*) Lag	DIG	s		1.0767	2.71	4.15	11.71
		z		1.077	2.71	4.16	11.71
	DIR	z		1.076	2.72	4.17	11.68

16.13. (*a*)

$$G_z(z) = G_{zo}G_x(z) = \frac{2.073 \times 10^{-7}(z+3.695)(z+0.2653)}{(z-1)(z-0.9900)(z-0.9704)}$$

$$\frac{C(z)}{R(z)} = \frac{2.073 \times 10^{-7}(z+3.695)(z+0.2653)}{(z-0.9959 \pm j0.0047756)(z-0.9687)}$$

$$M_p = 1.067,\ t_p = 6.94 \text{ s},\ t_s = 9.92 \text{ s},\ \text{and } K_1 = 0.41604 \text{ s}^{-1}$$

(*b*)

$$G(s) = G_A(s)G_x(s) = \frac{200K_x}{s(s+1)(s+3)(s+200)}$$

$$\frac{C(s)}{R(s)} = \frac{251.2}{(s+0.4082 \pm j0.4773)(s+3.184)(s+200^+)}$$

$$M_p = 1.067,\ t_p = 6.94 \text{ s},\ t_s = 9.90 \text{ s},\ \text{and } K_1 = 0.4187 \text{ s}^{-1}$$

(*c*) *PCT lead design*—$D_c(s) = K_{sc}(s)(s+1)/(s+3)$

$$\frac{C(s)}{R(s)} = \frac{1454}{(s+0.8439 \pm j0.9866)(s+4.312)(s+200^+)}$$

$M_p = 1.064,\ t_p = 3.47$ s, $t_s = 4.88$ s, $K_1 = 0.808\,\text{s}^{-1}$, and A = 5.7882

(*d*) *PCT lag design*—$D_c(s) = (A/10)[(s+0.005)/(s+0.0005)]$

www.4electron.com