МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет кораблебудування
імені адмірала Макарова

І. С. Білюк, О. В. Савченко,
С. О. Ткаченко, Г. А. Білюк

ТЕОРІЯ АВТОМАТИЧНОГО КЕРУВАННЯ
THEORY OF AUTOMATIC CONTROL

Конспект лекцій

Рекомендовано Методичною радою НУК

ВИДАВНИЦТВО
НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ
КОРАБЛЕБУДУВАННЯ
ІМ. АДМІРАЛА МАКАРОВА

2025

УДК 681.5(076)
Т33

Автори:
І. С. Білюк, канд. техн. наук, доцент;
О. В. Савченко, зав. лабораторіями;
С. О. Ткаченко, зав. лабораторіями;
Г. А. Білюк, методист

Рецензент
С. М. Новогрецький, канд. техн. наук, доцент, доцент кафедри суднових електроенергетичних систем

Рекомендовано Методичною радою НУК

Білюк І. С.
Т33 Теорія автоматичного керування (Theory of automatic control) : конспект лекцій / І. С. Білюк, О. В. Савченко, С. О. Ткаченко, Г. А. Білюк. – Миколаїв : НУК, 2025. – 154 с.

Подано основи теорії автоматичного керування. Матеріал скомпоновано у порядку проходження відповідних розділів лекційного курсу, розділи супроводжуються прикладами розв'язків, тому конспект лекцій може бути корисним для самостійної та індивідуальної роботи студентів.

Призначено для студентів спеціальності 141 «Електроенергетика, електротехніка та електромеханіка» при вивченні дисципліни «Теорія автоматичного керування» (Theory of automatic control).

УДК 681.5(076)

ЗМІСТ

PREFACE

Theory of automatic control is a field of science that studies technical systems using the ideas and methods of cybernetics. The main task of controlling a technical object is to find and implement a control algorithm under given conditions, thanks to which the requirements for the process will be met. Choosing the most rational control algorithm based on existing information is a clearly expressed extreme task.

Thus, the analysis of information processes in the system with the aim of their algorithmization and the synthesis of control systems that implement this algorithm is the task of the theory of automatic control.

The discipline “Theory of Automatic Control” forms in future bachelors system thinking and a holistic vision of the phenomena of the world of technology, nature, and the social environment, synthesizes students' knowledge of mathematics, physics, and other natural sciences, and gives illustrative examples of their constructive use.

1. INTRODUCTION TO CONTROL SYSTEMS

1.1 Introduction

Engineering is concerned with understanding and controlling the materials and forces of nature for the benefit of humankind. Control system engineers are concerned with understanding and controlling segments of their environment, often called systems, to provide useful economic products for society. The twin goals of understanding and controlling are complementary because effective systems control requires that the systems be understood and modeled. Furthermore, control engineering must often consider the control of poorly understood systems such as chemical process systems. The present challenge to control engineers is the modeling and control of modern, complex, interrelated systems such as traffic control systems, chemical processes, and robotic systems. Simultaneously, the fortunate engineer has the opportunity to control many useful and interesting industrial automation systems. Perhaps the most characteristic quality of control engineering is the opportunity to control machines and industrial and economic processes for the benefit of society.

Control engineering is based on the foundations of feedback theory and linear system analysis, and it integrates the concepts of network theory and communication theory. Therefore, control engineering is not limited to any engineering discipline but is equally applicable to aeronautical, chemical, mechanical, environmental, civil, and electrical engineering. For example, a control system often includes electrical, mechanical, and chemical components. Furthermore, as the understanding of the dynamics of business, social, and political systems increases, the ability to control these systems will also increase.

A control system is an interconnection of components forming a system configuration that will provide a desired system response. The basis for analysis of a system is the foundation provided by linear system theory, which assumes a cause-effect relationship for the components of a system. Therefore, a component or process to be controlled can be represented by a block, as shown in Figure 1.1. The input-output relationship represents the cause-and-effect relationship of the process, which in turn represents a processing of the input signal to provide an output signal variable, often with a power amplification. An open-loop control system uses a controller and an actuator to obtain the desired response, as shown in Figure 1.2. An open-loop system is a system without feedback.

An open-loop control system utilizes an actuating device to control the process directly without using feedback.

Figure 1.1 – Process to be controlled

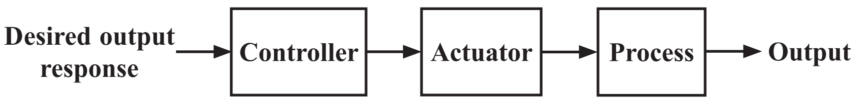


Figure 1.2 – Open-loop control system (without feedback)

In contrast to an open-loop control system, a closed-loop control system utilizes an additional measure of the actual output to compare the actual output with the desired output response. The measure of the output is called the feedback signal. A simple closed-loop feedback control system is shown in Figure 1.3. A feedback control system is a control system that tends to maintain a prescribed relationship of one system variable to another by comparing functions of these variables and using the difference as a means of control. With an accurate sensor, the measured output is a good approximation of the actual output of the system.

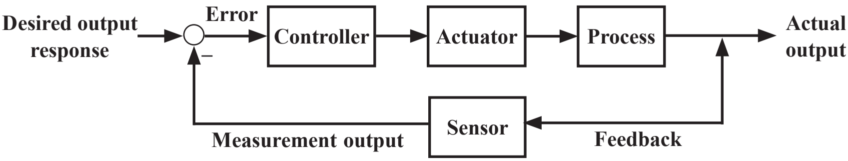


Figure 1.3 – Closed-loop feedback control system (with feedback)

A feedback control system often uses a function of a prescribed relationship between the output and reference input to control the process. Often the difference between the output of the process under control and the reference input is amplified and used to control the process so that the difference is continually reduced. In general, the difference between the desired output and the actual output is equal to the error, which is then adjusted by the controller. The output of the controller causes the actuator to modulate the process in order to reduce the error. The sequence is such, for instance, that if a ship is heading incorrectly to the right, the rudder is actuated to direct the ship to the left. The system shown in Figure 1.3 is a negative feedback control system, because the output is subtracted from the input and the difference is used as the input signal to the controller. The feedback concept has been the foundation for control system analysis and design.

A closed-loop control system uses a measurement of the output and feedback of this signal to compare it with the desired output (reference or command).

Closed-loop control has many advantages over open-loop control including the ability to reject external disturbances and improve measurement noise attenuation. We incorporate the disturbances and measurement noise in the block diagram as external inputs, as illustrated in Figure 1.4. External disturbances and measurement noise are inevitable in real-world applications and must be addressed in practical control system designs.

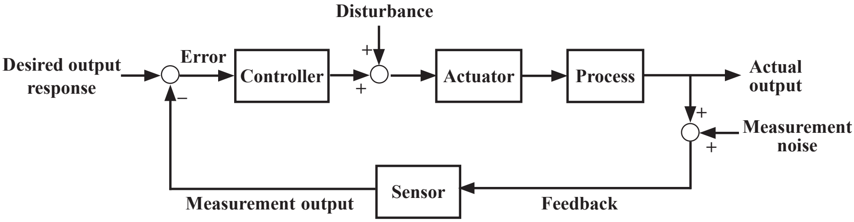


Figure 1.4 – Closed-loop feedback system with external disturbances and noise

The feedback systems in Figures 1.3 and 1.4 are single-loop feedback systems. Many feedback control systems contain more than one feedback loop. A common multiloop feedback control system is illustrated in Figure 1.5 with an inner loop and an outer loop. In this scenario, the inner loop has a controller and a sensor and the outer loop has a controller and sensor. Other varieties of multiloop feedback systems are considered throughout the book as they represent more practical situations found in real-world applications. However, we use the single-loop feedback system for learning about the benefits of feedback control systems since the outcomes readily extend to multiloop systems.

Due to the increasing complexity of the system under control and the interest in achieving optimum performance, the importance of control system engineering has grown in the past decade. Furthermore, as the systems become more complex, the interrelationship of many controlled variables must be considered in the control scheme. A block diagram depicting a multivariable control system is shown in Figure 1.6.

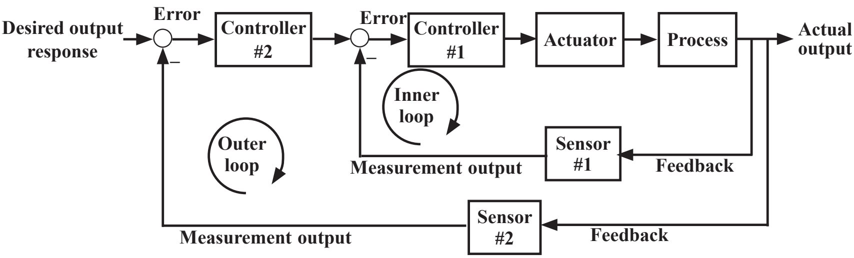


Figure 1.5 – Multiloop feedback system with an inner loop and an outer loop

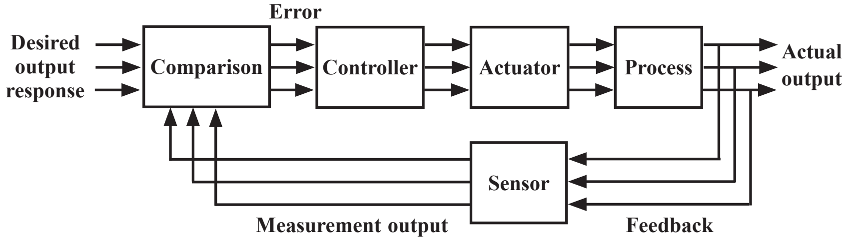


Figure 1.6 – Multivariable control system

A common example of an open-loop control system is a microwave oven set to operate for a fixed time. An example of a closed-loop control system is a person steering an automobile (assuming his or her eyes are open) by looking at the auto's location on the road and making the appropriate adjustments.

The introduction of feedback enables us to control a desired output and can improve accuracy, but it requires attention to the issue of stability of response.

1.2 Brief history of automatic control

The use of feedback to control a system has a fascinating history. The first applications of feedback control appeared in the development of float regulator mechanisms in Greece in the period

300 to 1 B.C. The water clock of Ktesibios used a float regulator. An oil lamp devised by Philon in approximately 250 B.C. used a float regulator in an oil lamp for maintaining a constant level of fuel oil. Heron of Alexandria, who lived in the first century A.D., published a book entitled Pneumatics, which outlined several forms of water-level mechanisms using float regulators.

The first feedback system to be invented in modern Europe was the temperature regulator of Cornelis Drebbel (1572–1633) of Holland. Dennis Papin (1647–1712) invented the first pressure regulator for steam boilers in 1681. Papin's pressure regulator was a form of safety regulator similar to a pressure-cooker valve.

The first automatic feedback controller used in an industrial process is generally agreed to be James Watt's flyball governor, developed in 1769 for controlling the speed of a steam engine. The all-mechanical device, shown in Figure 1.7, measured the speed of the output shaft and utilized the movement of the flyball to control the steam valve and therefore the amount of steam entering the engine.

As depicted in Figure 1.7, the governor shaft axis is connected via mechanical linkages and beveled gears to the output shaft of the steam engine. As the steam engine output shaft speed increases, the ball weights rise and move away from the shaft axis and through mechanical linkages the steam valve closes and the engine slows down.

The first historical feedback system, claimed by Russia, is the water-level float regulator said to have been invented by I. Polzunov in 1765. The level regulator system is shown in Figure l.8. The float detects the water level and controls the valve that covers the water inlet in the boiler. The next century was characterized by the development of automatic control systems through intuition and invention. Efforts to increase the accuracy of the control system led to slower attenuation of the transient oscillations and even to unstable systems. It then became imperative to develop a theory of automatic control.

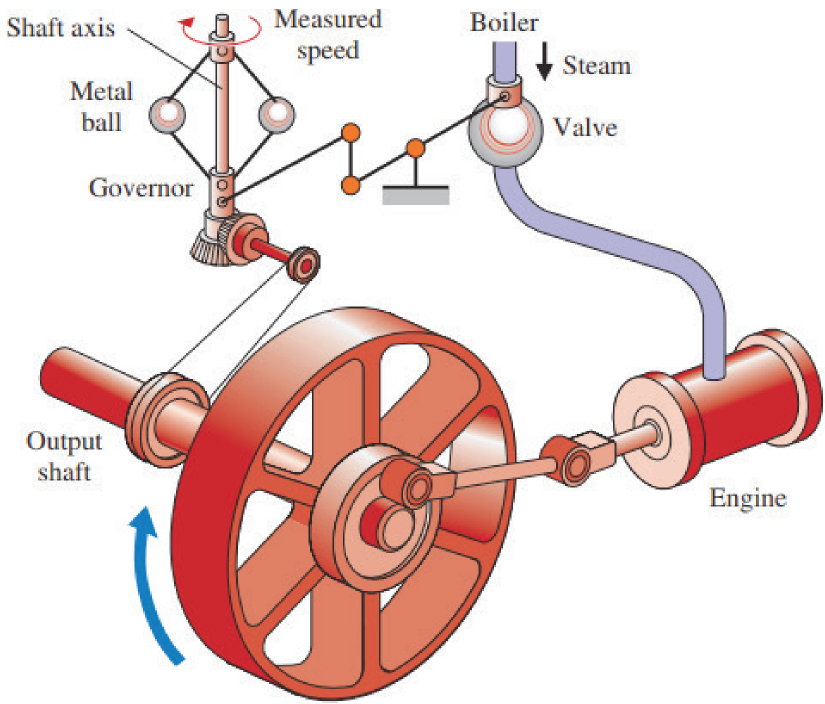


Figure 1.7 – Watt's flyball governor

In 1868, J. C. Maxwell formulated a mathematical theory related to control theory using a differential equation model of a governor. Maxwell's study was concerned with the effect various system parameters had on the system performance. During the same period, I. A. Vyshnegradskii formulated a mathematical theory of regulators.

Prior to World War II, control theory and practice developed differently in the United States and western Europe than in Russia and eastern Europe. The main impetus for the use of feedback in the United States was the development of the telephone system and electronic feedback amplifiers by Bode, Nyquist, and Black at Bell Telephone Laboratories.

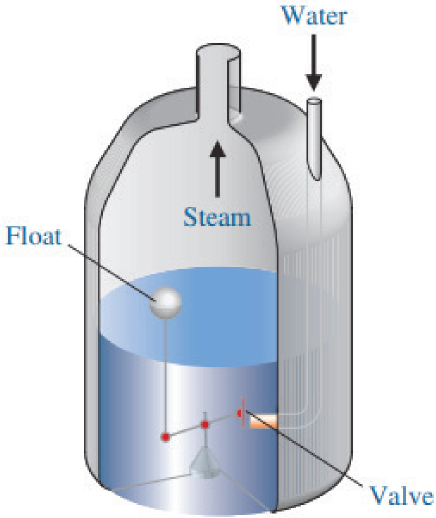


Figure 1.8 – Water-level float regulator

Harold S. Black graduated from Worcester Polytechnic Institute in 1921 and joined Bell Laboratories of American Telegraph and Telephone (AT&T). In 1921, the major task confronting Bell Laboratories was the improvement of the telephone system and the design of improved signal amplifiers. Black was assigned the task of linearizing, stabilizing, and improving the amplifiers that were used in tandem to carry conversations over distances of several thousand miles.

The frequency domain was used primarily to describe the operation of the feedback amplifiers in terms of bandwidth and other frequency variables. In contrast, the eminent mathematicians and applied mechanicians in the former Soviet Union inspired and dominated the field of control theory. Therefore, the Russian theory tended to utilize a time-domain formulation using differential equations.

The control of an industrial process (manufacturing, production, and so on) by automatic rather than manual means is often called automation. Automation is prevalent in the chemical, electric power, paper, automobile, and steel industries, among others. The concept of automation is central to our industrial society. Automatic machines are used to increase the production of a plant per worker in order to offset rising wages and inflationary costs. Thus industries are concerned with the productivity per worker of their plants. Productivity is defined as the ratio of physical output to physical input. In this case, we are referring to labor productivity, which is real output per hour of work.

A large impetus to the theory and practice of automatic control occurred during World War II when it became necessary to design and construct automatic airplane piloting, gun-positioning systems, radar antenna control systems, and other military systems based on the feedback control approach. The complexity and expected performance of these military systems necessitated an extension of the available control techniques and fostered interest in control systems and the development of new insights and methods. Prior to 1940, for most cases, the design of control systems was an art involving a trial-and-error approach. During the 1940s, mathematical and analytical methods increased in number and utility, and control engineering became an engineering discipline in its own right.

Frequency-domain techniques continued to dominate the field of control following World War II with the increased use of the Laplace transform and the complex frequency plane. During the 1950s, the emphasis in control engineering theory was on the development and use of the s-plane methods and, particularly, the root locus approach.

Furthermore, during the 1980s, the use of digital computers for control components became routine. The technology of these new control elements to perform accurate and rapid calculations was formerly unavailable to control engineers.

With the advent of Sputnik and the space age, another new impetus was imparted to control engineering. It became necessary to design complex, highly accurate control systems for missiles and space probes. Furthermore, the necessity to minimize the weight of satellites and to control them very accurately has spawned the important field of optimal control. Due to these requirements, the time-domain methods developed by Liapunov, Minorsky, and others have been met with great interest in the last two decades. Recent theories of optimal control developed by L. S. Pontryagin in the former Soviet Union and R. Bellman in the United States, as well as recent studies of robust systems, have contributed to the interest in time-domain methods. It now is clear that control engineering must consider both the time-domain and the frequency domain approaches simultaneously in the analysis and design of control systems.

1.3 Examples of control systems

Control engineering is concerned with the analysis and design of goal-oriented systems. Therefore, the mechanization of goal-oriented policies has grown into a hierarchy of goal-oriented control systems. Modern control theory is concerned with systems that have self-organizing, adaptive, robust, learning, and optimum qualities.

Feedback control is a fundamental fact of modern industry and society. Driving an automobile is a pleasant task when the auto responds rapidly to the driver's commands. Many cars have power steering and brakes, which utilize hydraulic amplifiers for amplification of the force to the brakes or the steering wheel. A simple block diagram of an automobile steering control system is shown in Figure 1.9(*a*).

The desired course is compared with a measurement of the actual course in order to generate a measure of the error, as shown in Figure 1.9(*b*). This measurement is obtained by visual and tactile (body movement) feedback, as provided by the feel of the steering wheel by the hand (sensor). This feedback system is a familiar version of the steering control system in an ocean liner or the flight controls in a large airplane. A typical dircction-of-travel response is shown in Figure 1.9(*c*).

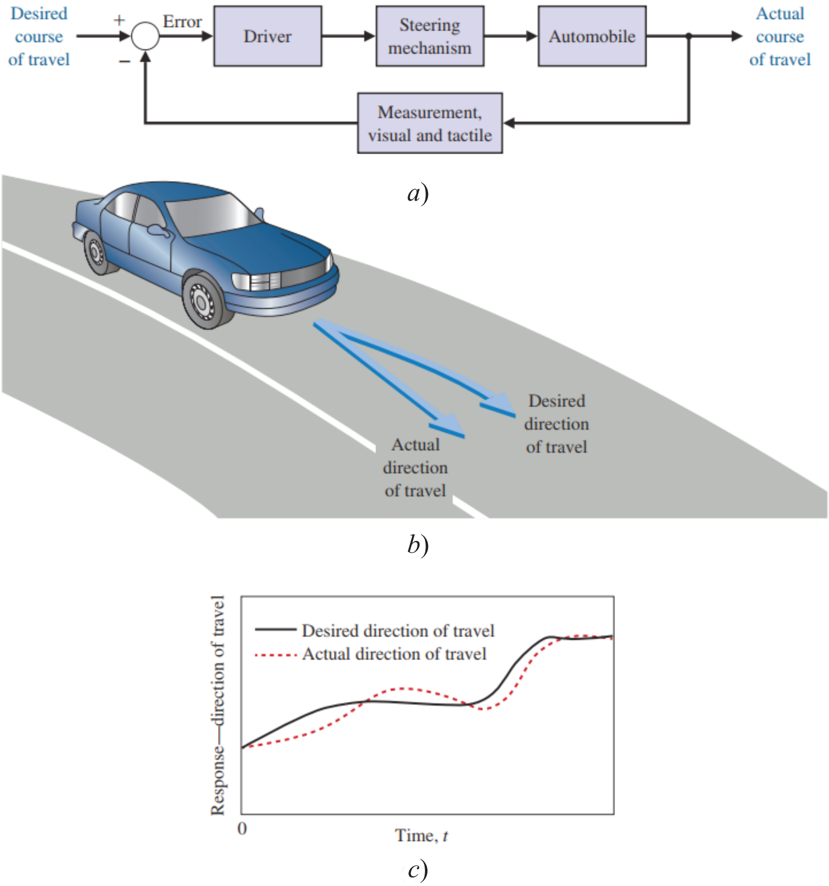


Figure 1.9 – (*a*) Automobile steering control system,
(*b*) The driver uses the difference between the actual and the desired direction of travel to generate a controlled adjustment of the steering wheel, (*c*) Typical direction of travel response

A basic, manually controlled closed-loop system for regulating the level of fluid in a tank is shown in Figure 1.10. The input is a reference level of fluid that the operator is instructed to maintain. (This reference is memorized by the operator.) The power amplifier is the operator, and the sensor is visual. The operator compares the actual level with the desired level and opens or closes the valve (actuator), adjusting the fluid flow out. to maintain the desired level.

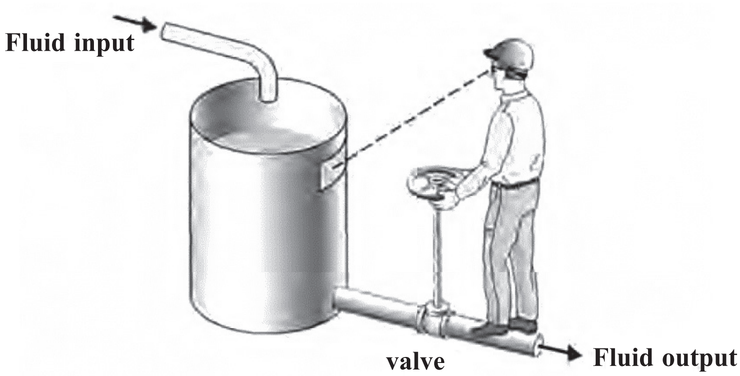


Figure 1.10 – A manual control system for regulating the level of fluid in a tank by adjusting the output valve. The operator views the level of fluid through a port in the side of the tank

Other familiar control systems have the same basic elements as the system shown in Figure 1.3. A refrigerator has a temperature setting or desired temperature, a thermostat to measure the actual temperature and the error, and a compressor motor for power amplification. Other examples in the home are the oven, furnace, and water heater. In industry, there are many examples, including speed controls; process temperature and pressure controls; and position, thickness, composition, and quality controls.

In its modern usage, automation can be defined as a technology that uses programmed commands to operate a given process, combined with feedback of information to determine that the commands have been properly executed. Automation is often used for processes that were previously operated by humans. When automated, the process can operate without human assistance or interference. In fact, most automated systems are capable of performing their functions with greater accuracy and precision, and in less time, than humans are able to do.

A semiautomated process is one that incorporates both humans and robots. For instance, many automobile assembly line operations require cooperation between a human operator and an intelligent robot.

Feedback control systems are used extensively in industrial applications. Thousands of industrial and laboratory robots are currently in use. Manipulators can pick up objects weighing hundreds of pounds and position them with an accuracy of one tenth of an inch or better. Automatic handling equipment for home, school, and industry is particularly useful for hazardous, repetitious, dull, or simple tasks.

Machines that automatically load and unload, cut, weld, or cast are used by industry to obtain accuracy, safety, economy, and productivity. The use of computers integrated with machines that perform tasks like a human worker has been foreseen by several authors. In his famous 1923 play, entitled R.U.R., Karel Capek called artificial workers robots, deriving the word from the Czech noun roboia, meaning “work.”

A robot is a computer-controlled machine and involves technology closely associated with automation. Industrial robotics can be defined as a particular field of automation in which the automated machine (that is, the robot) is designed to substitute for human labor. Thus robots possess certain humanlike characteristics.

Today, the most common humanlike characteristic is a mechanical manipulator that is patterned somewhat after the human arm and wrist. Some devices even have anthropomorphic mechanisms, including what we might recognize as mechanical arms, wrists, and hands. We recognize that the automatic machine is well suited to some tasks, as noted in Table 1.1, and that other tasks are best carried out by humans.

Another very important application of control technology is in the control of the modern automobile. Control systems for suspension, steering, and engine control have been introduced. Many new autos have a four-wheel-steering system, as well as an antiskid control system.

Table 1.1 – Task Difficulty: Human Versus Automatic Machine

Tasks Difficult for a Machine	**Tasks Difficult for a Human**
Inspect seeding in a nursery	Inspect a system in a hot, toxic environment
Drive a vehicle through rugged terrain	Repetitively assemble a clock
Identify the most expensive jewels on a tray of jewels	Land an airliner at night, in bad weather

A three-axis control system for inspecting individual semiconductor wafers is shown in Figure 1.11. This system uses a specific motor to drive each axis to the desired position in the x-y-z-axis, respectively. The goal is to achieve smooth, accurate movement in each axis. This control system is an important one for the semiconductor manufacturing industry.

There has been considerable discussion recently concerning the gap between practice and theory in control engineering. However, it is natural that theory precedes the applications in many fields of control engineering. Nonetheless, it is interesting to note that in the electric power industry, the largest industry in the United States, the gap is relatively insignificant. The electric power industry is primarily interested in energy conversion, control, and distribution. It is critical that computer control be increasingly applied to the power industry in order to improve the efficient use of energy resources. Also, the control of power plants for minimum waste emission has become increasingly important.

The modern, large-capacity plants, which exceed several hundred megawatts, require automatic control systems that account for the interrelationship of the process variables and optimum power production.

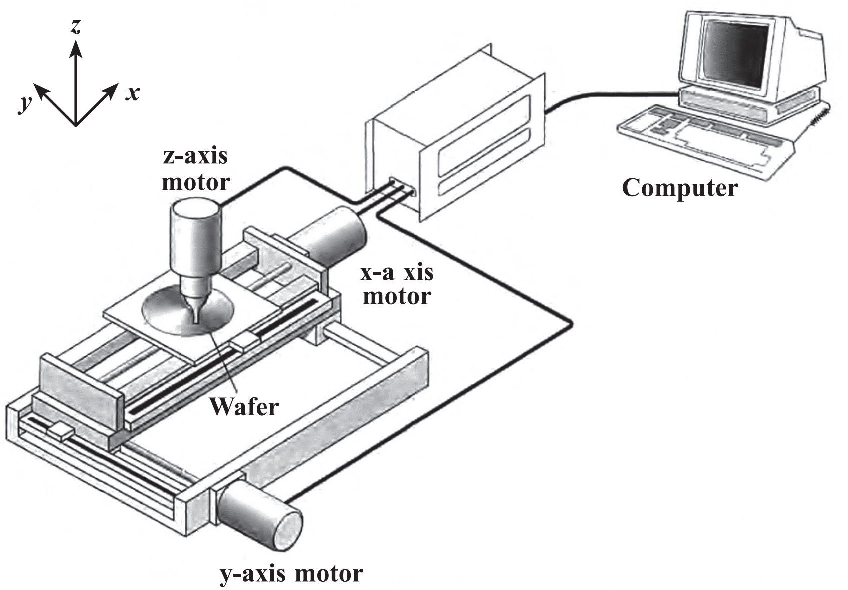


Figure 1.11 – A three-axis control system for inspecting individual semiconductor wafers with a highly sensitive camera

It is common to have 90 or more manipulated variables under coordinated control. A simplified model showing several of the important control variables of a large boiler-generator system is shown in Figure 1.12. This is an example of the importance of measuring many variables, such as pressure and oxygen, to provide information to the computer for control calculations.

The electric power industry has used the modern aspects of control engineering for significant and interesting applications. It appears that in the process industry, the factor that maintains the applications gap is the lack of instrumentation to measure all the important process variables, including the quality and composition of the product. As these instruments become available, the applications of modern control theory to industrial systems should increase measurably.

Another important industry, the metallurgical industry, has had considerable success in automatically controlling its processes. In fact, in many cases, the control theory is being fully implemented. For example, a hot-strip steel mill, which involves a $100-million investment, is controlled for temperature, strip width, thickness, and quality.

Rapidly rising energy costs coupled with threats of energy curtailment are resulting in new efforts for efficient automatic energy management. Computer controls are used to control energy use in industry and to stabilize and connect loads evenly to gain fuel economy.

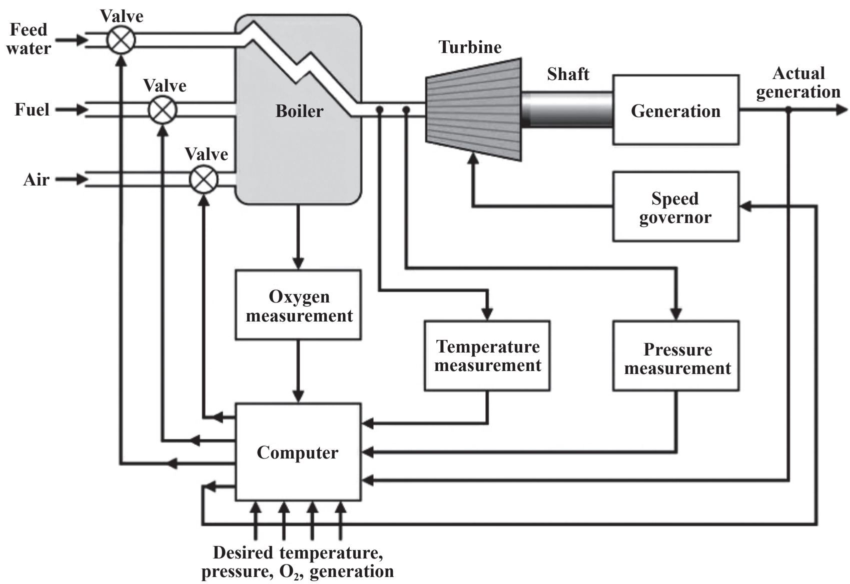


Figure 1.12 – Coordinated control system for a boiler-generator

There has been considerable interest recently in applying the feedback control concepts to automatic warehousing and inventory control. Furthermore, automatic control of agricultural systems (farms) is receiving increased interest. Automatically controlled silos and tractors have been developed and tested. Automatic control of wind turbine generators, solar heating and cooling, and automobile engine performance are important modern examples.

Also, there have been many applications of control system theory to biomedical experimentation, diagnosis, prosthetics, and biological control systems.

The control systems under consideration range from the cellular level to the central nervous system and include temperature regulation and neurological, respiratory, and cardiovascular control. Most physiological control systems are closed-loop systems.

However, we find not one controller but rather control loop within control loop, forming a hierarchy of systems. The modeling of the structure of biological processes confronts the analyst with a high-order model and a complex structure.

1.4 Engineering design

Engineering design is the central task of the engineer. It is a complex process in which both creativity and analysis play major roles.

Design is the process of conceiving or inventing the forms, parts, and details of a system to achieve a specified purpose.

Design activity can be thought of as planning for the emergence of a particular product or system. Design is an innovative act whereby the engineer creatively uses knowledge and materials to specify the shape, function, and material content of a system. The design steps are (1) to determine a need arising from the values of various groups, covering the spectrum from public policy makers to the consumer; (2) to specify in detail what the solution to that need must be and to embody these values; (3) to develop and evaluate various alternative solutions to meet these specifications; and (4) to decide which one is to be designed in detail and fabricated.

An important factor in realistic design is the limitation of time. Design takes place under imposed schedules, and we eventually settle for a design that may be less than ideal but considered "good enough." In many cases, time is the only competitive advantage.

A major challenge for the designer is writing the specifications for the technical product. Specifications are statements that explicitly state what the device or product is to be and do. The design of technical systems aims to provide appropriate design specifications and rests on four characteristics: complexity, trade-offs, design gaps, and risk.

Complexity of design results from the wide range of tools, issues, and knowledge to be used in the process. The large number of factors to be considered illustrates the complexity of the design specification activity, not only in assigning these factors their relative importance in a particular design, but also in giving them substance either in numerical or written form, or both.

The concept of trade-off involves the need to resolve conflicting design goals, all of which are desirable. The design process requires an efficient compromise between desirable but conflicting criteria. In making a technical device, we generally find that the final product does not appear as originally visualized. For example, our image of the problem we are solving does not appear in written description and ultimately in the specifications. Such design gaps are intrinsic in the progression from an abstract idea to its realization.

This inability to be absolutely sure about predictions of the performance of a technological object leads to major uncertainties about the actual effects of the designed devices and products. These uncertainties are embodied in the idea of unintended consequences or risk. The result is that designing a system is a risk-taking activity.

Complexity, trade-off, gaps, and risk are inherent in designing new systems and devices. Although they can be minimized by considering all the effects of a given design, they are always present in the design process.

Within engineering design, there is a fundamental difference between the two major types of thinking that must take place: engineering analysis and synthesis. Attention is focused on models of the physical systems that are analyzed to provide insight and that indicate directions for improvement. On the other hand, synthesis is the process by which these new physical configurations are created.

Design is a process that may proceed in many directions before the desired one is found. It is a deliberate process by which a designer creates something new in response to a recognized need while recognizing realistic constraints. The design process is inherently iterative–we must start somewhere. Successful engineers learn to simplify complex systems appropriately for design and analysis purposes. A gap between the complex physical system and the design model is inevitable.

Design gaps are intrinsic in the progression from the initial concept to the final product. We know intuitively that it is easier to improve an initial concept incrementally than to try to create a final design at the start. In other words, engineering design is not a linear process. It is an iterative, nonlinear, creative process.

The main approach to the most effective engineering design is parameter analysis and optimization. Parameter analysis is based on (1) identification of the key parameters, (2) generation of the system configuration, and (3) evaluation of how well the configuration meets the needs. These three steps form an iterative loop. Once the key parameters are identified and the configuration synthesized, the designer can optimize the parameters. Typically, the designer strives to identify a limited set of parameters to be adjusted.

1.5 Control system design

The design of control systems is a specific example of engineering design. The goal of control engineering design is to obtain the configuration, specifications, and identification of the key parameters of a proposed system to meet an actual need.

The control system design process is illustrated in Figure 1.13. The design process consists of seven main building blocks, which we arrange into three groups:

1. Establishment of goals and variables to be controlled, and definition of specifications (metrics) against which to measure performance.

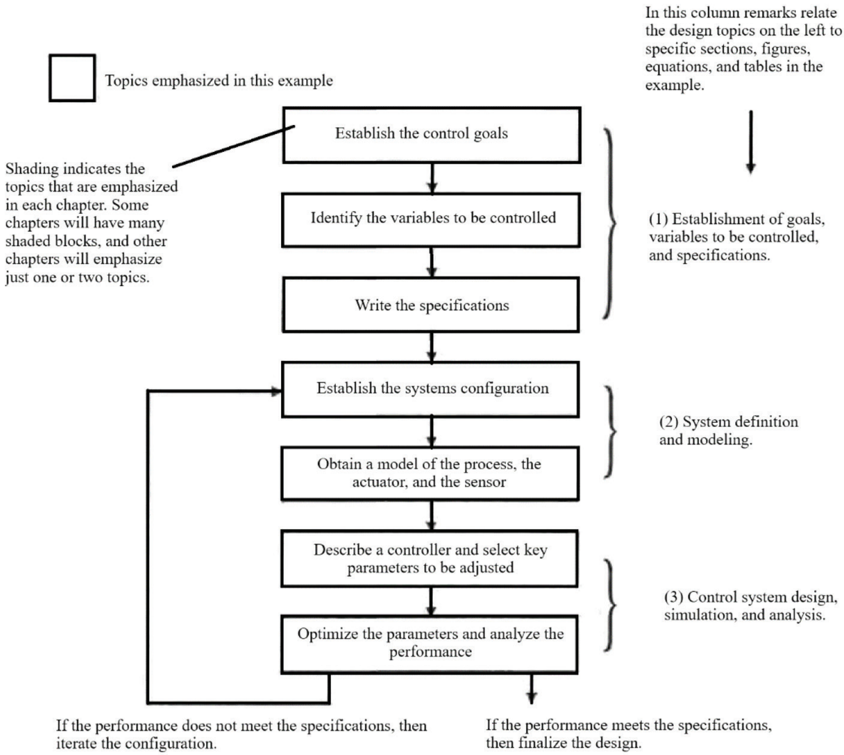


Figure 1.13 – The control system design process

2. System definition and modeling.
3. Control system design and integrated system simulation and analysis.

The first step in the design process consists of establishing the system goals. For example, we may state that our goal is to control the velocity of a motor accurately.

The second step is to identify the variables that we desire to control (for example, the velocity of the motor). The third step is to write the specifications in terms of the accuracy we must attain. This required accuracy of control will then lead to the identification of a sensor to measure the controlled variable. The performance specifications will describe how the closed-loop system should perform and will include (1) good regulation against disturbances, (2) desirable responses to commands, (3) realistic actuator signals, (4) low sensitivities, and (5) robustness.

As designers, we proceed to the first attempt to configure a system that will result in the desired control performance. This system configuration will normally consist of a sensor, the process under control, an actuator, and a controller, as shown in Figure 1.3. The next step consists of identifying a candidate for the actuator. This will, of course, depend on the process, but the actuation chosen must be capable of effectively adjusting the performance of the process. For example, if we wish to control the speed of a rotating flywheel, we will select a motor as the actuator. The sensor, in this case, must be capable of accurately measuring the speed. We then obtain a model for each of these elements.

The next step is the selection of a controller, which often consists of a summing amplifier that will compare the desired response and the actual response and then forward this error-measurement signal to an amplifier.

The final step in the design process is the adjustment of the parameters of the system to achieve the desired performance.

If we can achieve the desired performance by adjusting the parameters, we will finalize the design and proceed to document the results. If not, we will need to establish an improved system configuration and perhaps select an enhanced actuator and sensor. Then we will repeat the design steps until we are able to meet the specifications, or until we decide the specifications are too demanding and should be relaxed.

In summary, the controller design problem is as follows: Given a model of the system to be controlled (including its sensors and actuators) and a set of design goals, find a suitable controller, or determine that none exists. As with most of engineering design, the design of a feedback control system is an iterative and nonlinear process. A successful designer must consider the underlying physics of the plant under control, the control design strategy, the controller design architecture (that is, what type of controller will be employed), and effective controller tuning strategies. In addition, once the design is completed, the controller is often implemented in hardware, and hence issues of interfacing with hardware can appear. When taken together, these different phases of control system design make the task of designing and implementing a control system quite challenging.

1.6 Mechatronic systems

A natural stage in the evolutionary process of modern engineering design is encompassed in the area known as mechatronics. The term mechatronics was coined in Japan in the 1970s. Mechatronics is the synergistic integration of mechanical, electrical, and computer systems and has evolved over the past 30 years, leading to a new breed of intelligent products. Feedback control is an integral aspect of modern mechatronic systems. One can understand the extent that mechatronics reaches into various disciplines by considering the components that make up mechatronics.

The key elements of mechatronics are (1) physical systems modeling, (2) sensors and actuators, (3) signals and systems, (4) computers and logic systems, and (5) software and data acquisition. Feedback control encompasses aspects of all five key elements of mechatronics, but is associated primarily with the element of signals and systems, as illustrated in Figure 1.14.

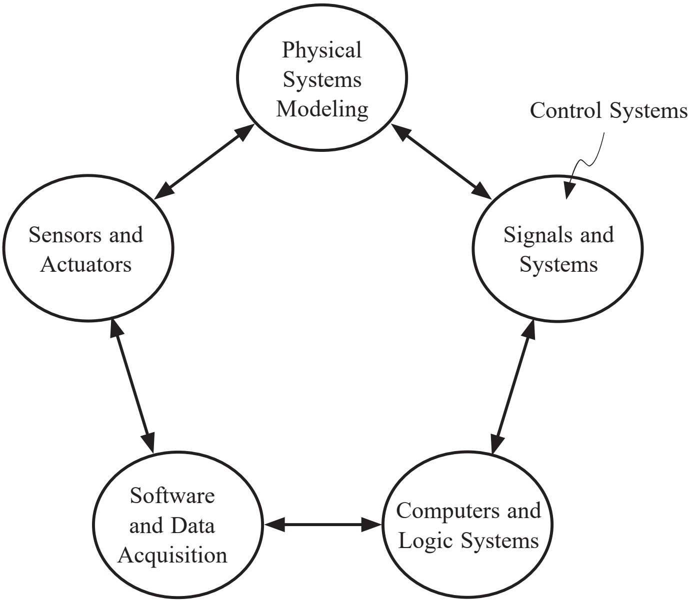


Figure 1.14 – The key elements of mechatronic

Advances in computer hardware and software technology coupled with the desire to increase the performance-to-cost ratio has revolutionized engineering design.

New products are being developed at the intersection of traditional disciplines of engineering, computer science, and the natural sciences. Advancements in traditional disciplines are fueling the growth of mechatronics systems by providing "enabling technologies". A critical enabling technology was the microprocessor which has had a profound effect on the design of consumer products. We should expect continued advancements in cost-effective microprocessors and microcontrollers, novel sensors and actuators enabled by advancements in applications of microelectromechanical systems (MEMS), advanced control methodologies and real-time programming methods, networking and wireless technologies, and mature computer-aided engineering (CAE) technologies for advanced system modeling, virtual prototyping, and testing. The continued rapid development in these areas will only accelerate the pace of smart (that is, actively controlled) products.

An exciting area of future mechatronic system development in which control systems will play a significant role is the area of alternative energy production and consumption. Hybrid fuel automobiles and efficient wind power generation are two examples of systems that can benefit from mechatronic design methods. In fact, the mechatronic design philosophy can be effectively illustrated by the example of the evolution of the modern automobile. Before the 1960s, the radio was the only significant electronic device in an automobile. Today, many automobiles have 30–60 microcontrollers, up to 100 electric motors, about 200 pounds of wiring, a multitude of sensors, and thousands of lines of software code. A modern automobile can no longer be classified as a strictly mechanical machine – it has been transformed into a comprehensive mechatronic system.

Recent research and development have led to the next-generation hybrid fuel automobile. The hybrid fuel vehicle utilizes a conventional internal combustion engine in combination with a battery (or other energy storage device such as a fuel cell or flywheel) and an electric motor to provide a propulsion system capable of doubling the fuel economy over conventional automobiles. Although these hybrid vehicles will never be zero-emission vehicles (since they have internal combustion engines), they can reduce the level of harmful emissions by one-third to one-half, and with future improvements, these emissions may reduce even further. As slated earlier, the modern automobile requires many advanced control systems to operate. The control systems must regulate the performance of the engine, including fuel-air mixtures, valve timing, transmissions, wheel traction control, antilock brakes, and electronically controlled suspensions, among many other functions. On the hybrid fuel vehicle, there are additional control functions that must be satisfied. Especially necessary is the control of power between the internal combustion engine and the electric motor, determining power storage needs and implementing the battery charging, and preparing the vehicle for low-emission start-ups. The overall effectiveness of the hybrid fuel vehicle depends on the combination of power units that are selected (e.g., battery versus fuel cell for power storage). Ultimately, however, the control strategy that integrates the various electrical and mechanical components into a viable transportation system strongly influences the acceptability of the hybrid fuel vehicle concept in the marketplace.

The second example of a mechatronic system is the advanced wind power generation system.

Many nations in the world today are faced with unstable energy supplies, often leading to rising fuel prices and energy shortages. Additionally, the negative effects of fossil fuel utilization on the quality of our air are well documented. Many nations have an imbalance in the supply and demand of energy, consuming more than they produce. To address this imbalance, many engineers are considering developing advanced systems to access other sources of energy, such as wind energy. New developments are required in materials and aerodynamics so that longer turbine rotors can operate efficiently in the lower winds, and in a related problem, the towers that support the turbine must be made taller without increasing the overall costs. In addition, advanced controls will be required to achieve the level of efficiency required in the wind generation drive train.

1.7 Future evolution of control systems

The continuing goal of control systems is to provide extensive flexibility and a high level of autonomy. Two system concepts are approaching this goal by different evolutionary pathways, as illustrated in Figure 1.15.

Today's industrial robot is perceived as quite autonomous – once it is programmed, further intervention is not normally required.

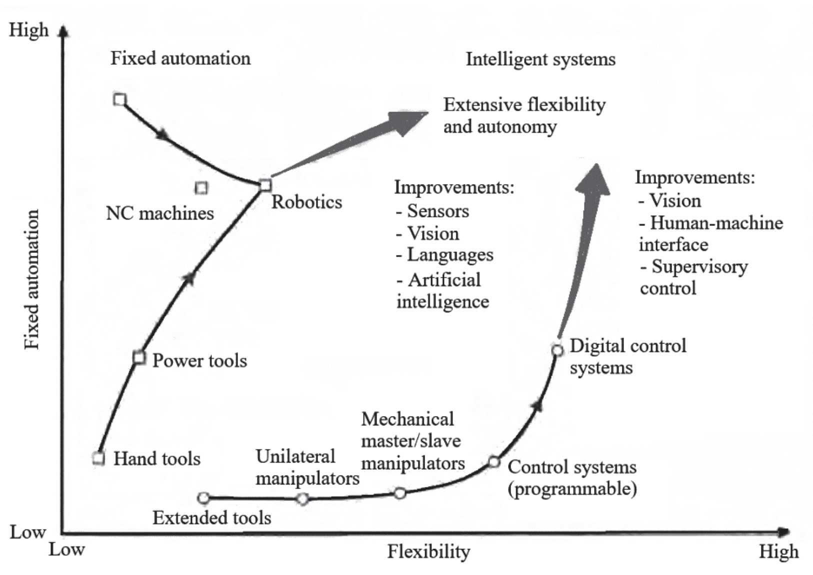


Figure 1.15 – Future evolution of control systems and robotics

Because of sensory limitations, these robotic systems have limited flexibility in adapting to work environment changes; improving perception is the motivation of computer vision research. The control system is very adaptable, but it relies on human supervision. Advanced robotic systems are striving for task adaptability through enhanced sensory feedback. Research areas concentrating on artificial intelligence, sensor integration, computer vision, and off-line CAD/CAM programming will make systems more universal and economical. Control systems are moving toward autonomous operation as an enhancement to human control. Research in supervisory control, human-machine interface methods, and computer database management are intended to reduce operator burden and improve operator efficiency. Many research activities are common to robotics and control systems and are aimed at reducing implementation cost and expanding the realm of application. These include improved communication methods and advanced programming languages.

The easing of human labor by technology, a process that began in prehistory, is entering a new stage. The acceleration in the pace of technological innovation inaugurated by the Industrial Revolution has until recently resulted mainly in the displacement of human muscle power from the tasks of production. The current revolution in computer technology is causing an equally momentous social change, the expansion of information gathering and information processing as computers extend the reach of the human brain.

Control systems are used to achieve (1) increased productivity and (2) improved performance of a device or system. Automation is used to improve productivity and obtain high-quality products. Automation is the automatic operation or control of a process, device, or system. We use automatic control of machines and processes to produce a product reliably and with high precision. With the demand for flexible, custom production, a need for flexible automation and robotics is growing.

The theory, practice, and application of automatic control is a large, exciting, and extremely useful engineering discipline. One can readily understand the motivation for a study of modern control systems.

2. MATHEMATICAL MODELS OF SYSTEMS

2.1 Introduction

Mathematical models of physical systems are key elements in the design and analysis of control systems. The dynamic behavior is generally described by ordinary differential equations. We will consider a wide range of systems, including mechanical, hydraulic, and electrical. Since most physical systems are nonlinear, we will discuss linearization approximations, which allow us to use Laplace transform methods. We will then proceed to obtain the input-output relationship for components and subsystems in the form of transfer functions. The transfer function blocks can be organized into block diagrams or signal-flow graphs to graphically depict the interconnections. Block diagrams (and signal-flow graphs) are very convenient and natural tools for designing and analyzing complicated control systems.

To understand and control complex systems, one must obtain quantitative mathematical models of these systems. It is necessary therefore to analyze the relationships between the system variables and to obtain a mathematical model. Because the systems under consideration are dynamic in nature, the descriptive equations are usually differential equations. Furthermore, if these equations can be linearized, then the Laplace transform can be used to simplify the method of solution.

In practice, the complexity of systems and our ignorance of all the relevant factors necessitate the introduction of assumptions concerning the system operation. Therefore, we will often find it useful to consider the physical system, express any necessary assumptions, and linearize the system. Then, by using the physical laws describing the linear equivalent system, we can obtain a set of linear differential equations. Finally, using mathematical tools, such as the Laplace transform, we obtain a solution describing the operation of the system. In summary, the approach to dynamic system modeling can be listed as follows:

1. Define the system and its components.
2. Formulate the mathematical model and fundamental necessary assumptions based on basic principles.
3. Obtain the differential equations representing the mathematical model.
4. Solve the equations for the desired output variables.
5. Examine the solutions and the assumptions.
6. If necessary, reanalyze or redesign the system.

2.2 Differential equations of physical systems

The differential equations describing the dynamic performance of a physical system are obtained by utilizing the physical laws of the process. This approach applies equally well to mechanical, electrical, fluid, and thermodynamic systems.

Consider the torsional spring-mass system in Figure 2.1 with applied torque $T_a(t)$. Assume the torsional spring element is massless. Suppose we want to measure the torque $T_s(t)$ transmitted to the mass m. Since the spring is massless, the sum of the torques acting on the spring itself must be zero, or

$$T_a(t) - T_s(t) = 0,$$

which implies that $T_s(t) = T_a(t)$. We see immediately that the external torque $T_a(t)$ applied at the end of the spring is transmitted through the torsional spring. Because of this, we refer to the torque as a through-variable. In a similar manner, the angular rate difference associated with the torsional spring element is

$$\omega(t) = \omega_s(t) - \omega_a(t).$$

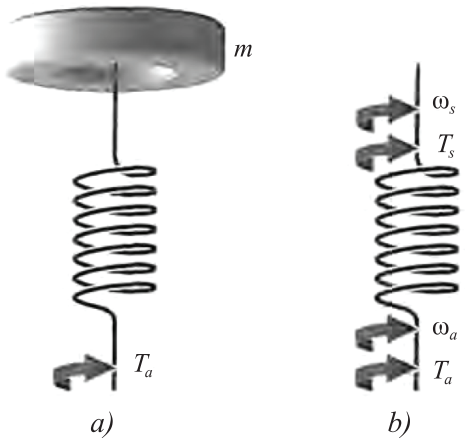


Figure 2.1 – (*a*) Torsional spring-mass system, (*b*) Spring element

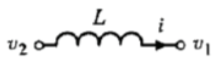
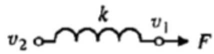
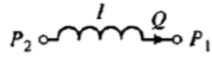
Thus, the angular rate difference is measured across the torsional spring element and is referred to as an across-variable. These same types of arguments can be made for most common physical variables (such as force, current, volume, flow rate, etc.). A summary of the through- and across-variables of dynamic systems is given in Table 2.1.

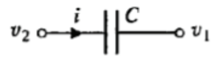
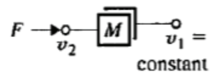
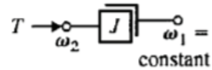
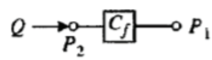
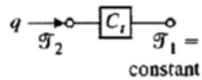
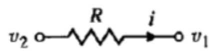
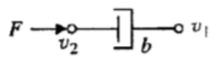
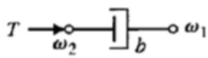
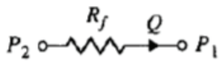
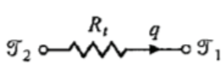
A summary of the describing equations for lumped, linear, dynamic elements is given in Table 2.2. The equations in Table 2.2 are idealized descriptions and only approximate the actual conditions (for example, when a linear, lumped approximation is used for a distributed element).

Table 2 – Summary of Through – and Across-Variables for Physical Systems

System	**Variables Through Element**	**Integrated Through-Variable**	**Variable Across Element**	**Integrated Across-Variable**
Electrical	Current, i	Charge, q	Voltage difference, v_{21}	Flux linkage, λ_{21}
Mechanical translational	Force, F	Translational momentum, P	Velocity difference, v_{21}	Displacement difference, y_{21}
Mechanical rotational	Torque, T	Angular momentum, h	Angular velocity difference, ω_{21}	Angular displacement difference, θ_{21}
Fluid	Fluid volumetric rate of flow, Q	Volume, V	Pressure difference, P_{21}	Pressure momentum, γ_{21}
Thermal	Heat flow rate, q	Heat energy, H	Temperature difference, F_{21}	

Table 2.2 – Summary of Governing Differential Equations for Ideal Elements

Type of Element	Physical Element	Goverming Equation	Energy E or Power P	Symbol
1	2	3	4	5
Inductive storage	Electrical inductance	$v_{21} = L\dfrac{di}{dt}$	$E = \dfrac{1}{2}Li^2$	
	Translational spring	$v_{21} = \dfrac{1}{k}\dfrac{dF}{dt}$	$E = \dfrac{1}{2}\dfrac{F^2}{k}$	
	Rotational spring	$\omega_{21} = \dfrac{1}{k}\dfrac{dT}{dt}$	$E = \dfrac{1}{2}\dfrac{T^2}{k}$	
	Fluid inertia	$P_{21} = I\dfrac{dQ}{dt}$	$E = \dfrac{1}{2}IQ^2$	

1	2	3	4	5
Capacitive storage	Electrical capacitance	$i = C\dfrac{dv_{21}}{dt}$	$E = \dfrac{1}{2}Cv_{21}^2$	
	Translational mass	$F = M\dfrac{dv_2}{dt}$	$E = \dfrac{1}{2}Mv_2^2$	
	Rotational mass	$T = J\dfrac{d\omega_2}{dt}$	$E = \dfrac{1}{2}J\omega_2^2$	
	Fluid capacitance	$Q = C_f\dfrac{dP_{21}}{dt}$	$E = \dfrac{1}{2}C_f P_{21}^2$	
	Thermal capacitance	$q = C_t\dfrac{dP_2}{dt}$	$E = C_1 P_2$	
Energy dissipators	Electrical resistance	$i = \dfrac{1}{R}v_{21}$	$P = \dfrac{1}{R}v_{21}^2$	
	Translational damper	$F = bv_{21}$	$P = bv_{21}^2$	
	Rotational damper	$T = b\omega_{21}$	$P = b\omega_{21}^2$	
	Fluid resistance	$Q = \dfrac{1}{R_f}P_{21}$	$P = \dfrac{1}{R_f}P_{21}^2$	
	Thermal resistance	$q = \dfrac{1}{R_t}\mathrm{P}_{21}$	$P = \dfrac{1}{R_t}\Gamma_{21}$	

– *Through-variable*: F = force, T = torque, i = current, Q = fluid volumetric flow rate, q = heat flow rate.

– *Across-variable*: v = translational velocity, ω = angular velocity, v = voltage, P = pressure, Γ = temperature.

– *Inductive storage*: L = inductance, $1/k$ = reciprocal translational or rotational stiffness, I = fluid intertance.

– *Capacitive storage*: C = capacitance, M = mass, J = moment of inertia, C_f = fluid capacitance, C_1 = thermal capacitance.

– *Energy dissipators*: R = resistance, b = viscous friction, R_f = fluid resistance, R_1 = thermal resistance.

The symbol v is used for both voltage in electrical circuits and velocity in translational mechanical systems and is distinguished within the context of each differential equation. For mechanical systems, one uses Newton's laws; for electrical systems, Kirchhoff's voltage laws. For example, the simple spring-mass-damper mechanical system shown in Figure 2.2(*a*) is described by Newton's second law of motion. (This system could represent, for example, an automobile shock absorber.)

The free-body diagram of the mass M is shown in Figure 2.2(*b*). In this spring-mass-damper example, we model the wall friction as a viscous damper, that is, the friction force is linearly proportional to the velocity of the mass. In reality the friction force may behave in a more complicated fashion. For example, the wall friction may behave as a Coulomb damper. Coulomb friction, also known as dry friction, is a nonlinear function of the mass velocity and possesses a discontinuity around zero velocity. For a well-lubricated, sliding surface, the viscous friction is appropriate and will be used here and in subsequent spring-mass-damper examples. Summing the forces acting on M and utilizing Newton's second law yields

$$M\frac{d^2y(t)}{dt^2}+b\frac{dy(t)}{dt}+ky(t)=r(t), \tag{2.1}$$

where k is the spring constant of the ideal spring and b is the friction constant. Equation (2.1) is a second-order linear constant-coefficient differential equation.

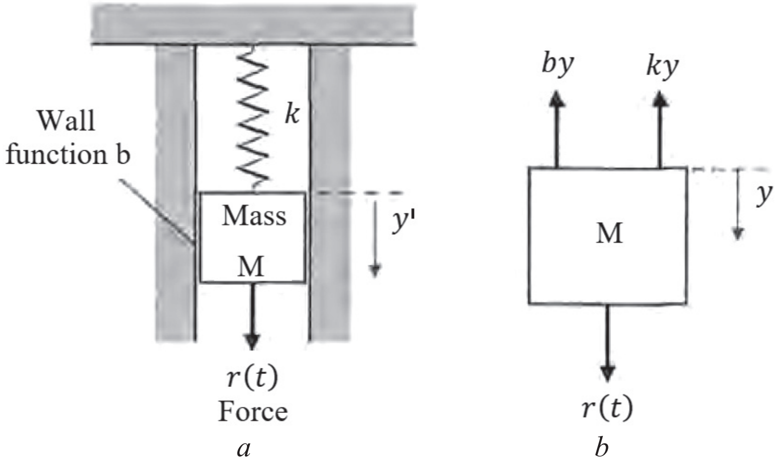


Figure 2.2 – (*a*) Spring-massdamper system, (*b*) Free-body diagram.

Alternatively, one may describe the electrical RLC circuit of Figure 2.3 by utilizing Kirchhoff s current law. Then we obtain the following integrodifferential equation:

$$\frac{v(t)}{R}+C\frac{dv(t)}{dt}+\frac{1}{L}\int_0^t v(t)\,dt=r(t). \qquad (2.2)$$

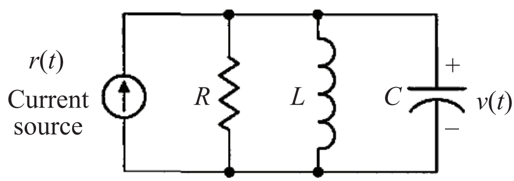


Figure 2.3 – RLC circuit

The solution of the differential equation describing the process may be obtained by classical methods such as the use of integrating factors and the method of undetermined coefficients. For example, when the mass is initially displaced a distance $y(0) = y_0$ and released, the dynamic response of the system can be represented by an equation of the form

$$y(t) = K_1 e^{-\alpha_1 t} \sin(\beta_1 t + \theta_1). \qquad (2.3)$$

A similar solution is obtained for the voltage of the RLC circuit when the circuit is subjected to a constant current $r(t) = I$. Then the voltage is

$$v(t) = K_2 e^{-\alpha_2 t} \cos(\beta_2 t + \theta_2). \qquad (2.4)$$

A voltage curve typical of an RLC circuit is shown in Figure 2.4.

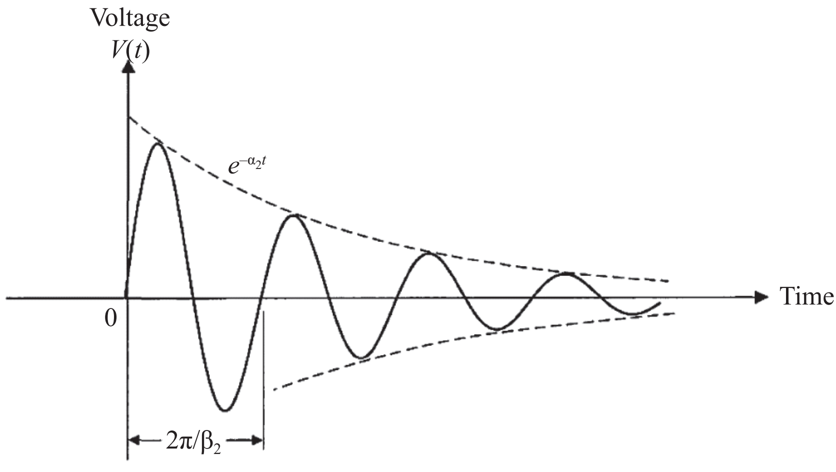


Figure 2.4 – Typical voltage response for an RLC circuit

To reveal further the close similarity between the differential equations for the mechanical and electrical systems, we shall rewrite Equation (2.1) in terms of velocity:

$$v(t) = \frac{dy(t)}{dt},$$

then we have

$$M\frac{dv(t)}{dt}+bv(t)+k\int_0^t v(t)\,dt=r(t). \qquad (2.5)$$

One immediately notes the equivalence of Equations (2.5) and (2.2) where velocity $v(t)$ and voltage $v(t)$ are equivalent variables, usually called analogous variables, and the systems are analogous systems. Therefore, the solution for velocity is similar to Equation (2.4), and the response for an underdamped system is shown in Figure 2.4.

The concept of analogous systems is a very useful and powerful technique for system modeling. The voltage-velocity analogy, often called the force-current analogy, is a natural one because it relates the analogous through- and across-variables of the electrical and mechanical systems. Another analogy that relates the velocity and current variables is often used and is called the force-voltage analogy.

Analogous systems with similar solutions exist for electrical, mechanical, thermal, and fluid systems. The existence of analogous systems and solutions provides the analyst with the ability to extend the solution of one system to all analogous systems with the same describing differential equations. Therefore, what one learns about the analysis and design of electrical systems is immediately extended to an understanding of fluid, thermal, and mechanical systems.

2.3 Linear approximations of physical systems

A great majority of physical systems are linear within some range of the variables. In general, systems ultimately become nonlinear as the variables are increased without limit. For example, the spring-mass-damper system of Figure 2.2 is linear and described by Equation (2.1) as long as the mass is subjected to small deflections $y(t)$. However, if $y(t)$ were continually increased, eventually the spring would be overextended and break. Therefore, the question of linearity and the range of applicability must be considered for each system.

A system is defined as linear in terms of the system excitation and response. In the case of the electrical network, the excitation is the input current $r(t)$ and the response is the voltage $v(t)$. In general, a necessary condition for a linear system can be determined in terms of an excitation $x(t)$ and a response $y(t)$. When the system at rest is subjected to an excitation $X_1(t)$, it provides a response $y_2(t)$. Furthermore, when the system is subjected to an excitation $x_2(t)$, it provides a corresponding response $y_2(t)$. For a linear system, it is necessary that the excitation $X_1(t) + x_2(t)$ result in a response $y_1(t) + y_2(i)$. This is usually called the principle of superposition.

Furthermore, the magnitude scale factor must be preserved in a linear system. Again, consider a system with an input $x(t)$ that results in an output $y(t)$. Then the response of a linear system to a constant multiple β of an input x must be equal to the response to the input multiplied by the same constant so that the output is equal to βy. This is called the property of homogeneity.

A linear system satisfies the properties of superposition and homogeneity.

A system characterized by the relation $y = x^2$ is not linear, because the superposition property is not satisfied. A system represented by the relation $y = mx + b$ is not linear, because it does not satisfy the homogeneity property. However, this second system may be considered linear about an operating point y_0, x_0 for small changes Δx and Δy.

When $x = x_0 + \Delta x$ and $y = y_0 + \Delta y$, we have

$$y = mx + b$$

or

$$y_0 + \Delta y = mx_0 + m\Delta x + b.$$

Therefore, $\Delta y = m\Delta x$, which satisfies the necessary conditions. The linearity of many mechanical and electrical elements can be assumed over a reasonably large range of the variables. This is not usually the case for thermal and fluid elements, which are more frequently nonlinear in character. Fortunately, however, one can often linearize nonlinear elements assuming small-signal conditions. This is the normal approach used to obtain a linear equivalent circuit for electronic circuits and transistors. Consider a general element with an excitation (through-) variable $x(t)$ and a response (across-) variable $y(t)$. Several examples of dynamic system variables are given in Table 2.1.

The relationship of the two variables is written as

$$y(t) = g\big(x(t)\big), \tag{2.6}$$

where $g(x(t))$ indicates $y(t)$ is a function of $x(t)$. *The* normal operating point is designated by x_0. Because the curve (function) is continuous over the range of interest, a **Taylor series** expansion about the operating point may be utilized. Then we have

$$y = g(x) = g(x_0) + \left.\frac{dg}{dx}\right|_{x=x_0} \frac{(x-x_0)}{1!} + \left.\frac{d^2 g}{dx^2}\right|_{x=x_0} \frac{(x-x_0)^2}{2!} + \ldots . \tag{2.7}$$

The slope at the operating point,

$$\left.\frac{dg}{dx}\right|_{x=x_0},$$

is a good approximation to the curve over a small range of $(x - x_0)$, the deviation from the operating point. Then, as a reasonable approximation, Equation (2.7) becomes

$$y = g(x_0) + \left.\frac{dg}{dx}\right|_{x=x_0} (x - x_0) = y_0 + m(x - x_0), \tag{2.9}$$

where m is the slope at the operating point.

Finally, Equation (2.8) can be rewritten as the linear equation

$$\left(y-y_0\right)=m\left(x-x_0\right)$$

or

$$\Delta y = m\Delta x. \tag{2.9}$$

Consider the case of a mass, M, sitting on a nonlinear spring, as shown in Figure 2.5(*a*). The normal operating point is the equilibrium position that occurs when the spring force balances the gravitational force *Mg*, where g is the gravitational constant. Thus, we obtain $f_0 = Mg$, as shown. For the nonlinear spring with $f = y^2$, the equilibrium position is $y_0 = \left(Mg\right)^{\frac{1}{2}}$. The linear model for small deviation is

$$\Delta f = m\Delta y,$$

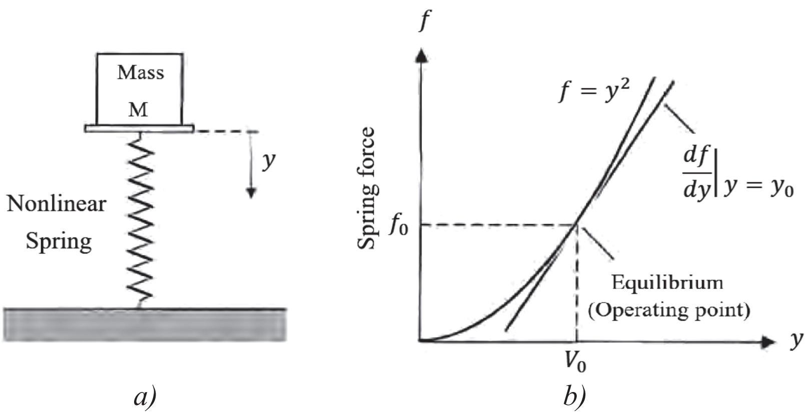


Figure 2.5 – (*a*) A mass sitting on a nonlinear spring,
(*b*) The spring force versus *y*

where

$$m = \left.\frac{df}{dy}\right|_{y_0},$$

as shown in Figure 2.5(b). Thus, $m = 2y_0$. A **linear approximation** is as accurate as the assumption of small signals is applicable to the specific problem.

If the dependent variable y depends upon several excitation variables, x_1, x_2,..., x_n, then the functional relationship is written as

$$y = g\left(x_1, x_2, \ldots, x_n\right). \tag{2.10}$$

The Taylor series expansion about the operating point x_1, x_2,..., x_n is useful for a linear approximation to the nonlinear function. When the higher-order terms are neglected, the linear approximation is written as

$$y = g\left(x_{10}, x_{20}, \ldots, x_{n0}\right) + \left.\frac{dg}{dx_1}\right|_{x=x_0}\left(x_1 - x_{10}\right) + \left.\frac{dg}{dx_2}\right|_{x=x_0}\left(x_2 - x_{2n}\right) + \ldots +$$

$$+ \left.\frac{dg}{dx_n}\right|_{x=x_0}\left(x_n - x_{n0}\right), \tag{2.11}$$

where x_0 is the operating point.

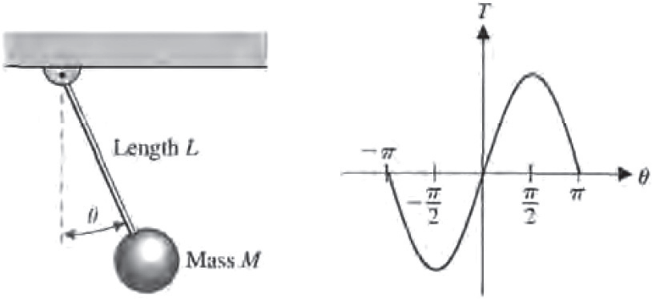
Example 2.1 will clearly illustrate the utility of this method.

EXAMPLE 2.1 **Pendulum oscillator model**

Consider the pendulum oscillator shown in Figure 2.6(a). The torque on the mass is

$$T = MgL\sin\theta, \tag{2.12}$$

where g is the gravity constant.



a) *b)*

Figure 2.6 – Pendulum oscillator

The equilibrium condition for the mass is $\Theta_0 = 0°$. The nonlinear relation between T and Θ is shown graphically in Figure 2.6(*b*).The first derivative evaluated at equilibrium provides the linear approximation, which is

$$T - T_0 \cong MgL \left.\frac{d\sin\theta}{d\theta}\right|_{\theta=\theta_0} (\theta - \theta_0),$$

where $T_0 = 0$. Then, we have

$$T = MgL(\cos 0°)(\theta - 0°) = MgL\theta. \qquad (2.13)$$

This approximation is reasonably accurate for $-\frac{\pi}{4} < \Theta < \frac{\pi}{4}$. For example, the response of the linear model for the swing through ±30° is within 5 % of the actual nonlinear pendulum response.

2.4 The laplace transform

The ability to obtain linear approximations of physical systems allows the analyst to consider the use of the Laplace transformation. The Laplace transform method substitutes relatively easily solved algebraic equations for the more difficult differential equations. The time-response solution is obtained by the following operations:

1. Obtain the linearized differential equations.
2. Obtain the Laplace transformation of the differential equations.
3. Solve the resulting algebraic equation for the transform of the variable of interest.

The Laplace transform exists for linear differential equations for which the transformation integral converges. Therefore, for $f(t)$ to be transformable, it is sufficient that

$$\int_{0^-}^{\infty} \left| f(t) \right| e^{-\sigma_1 t} dt < \infty,$$

for some real, positive σ_1. The 0^- indicates that the integral should include any discontinuity, such as a delta function at $t = 0$. If the magnitude of $f(t)$ is $|f(t)| < Me^{\alpha t}$ for all positive t, the integral will converge for $\sigma_1 > a$.

The region of convergence is therefore given by $\infty > \sigma_1 > \alpha$, and σ_1 is known as the abscissa of absolute convergence. Signals that are physically realizable always have a Laplace transform. The Laplace transformation for a function of time $f(t)$, is

$$F(s) = \int_{0^-}^{\infty} f(t) e^{-st} dt = \mathcal{L}\{f(t)\}. \tag{2.14}$$

The inverse Laplace transform is written as

$$f(t) = \frac{1}{2\pi j} \int_{\alpha - j\infty}^{\alpha + j\infty} F(s) e^{+st} ds. \tag{2.15}$$

The transformation integrals have been employed to derive tables of Laplace transforms that are used for the great majority of problems. A table of important Laplace transform pairs is given in Table 2.3.

Table 2.3 – Important Laplace Transform Pairs

***f*(*t*)**	***F*(*s*)**
1	2
Step function, $u(t)$	$\frac{1}{s}$
e^{-at}	$\frac{1}{s+a}$
$\sin\omega t$	$\frac{\omega}{s^2+\omega^2}$
$\cos\omega t$	$\frac{s}{s^2+\omega^2}$
t^n	$\frac{n!}{s^{n+1}}$
$f^{(k)}(t)=\frac{d^k f(t)}{dt^k}$	$s^k F(s)-s^{k-1}f(0^-)-$ $-s^{k-2}f'(0^-)-\ldots-f^{(k-1)}(0^-)$
$\int_{-\infty}^{t} f(t)dt$	$\frac{F(s)}{s}+\frac{1}{s}\int_{-\infty}^{0} f(t)dt$
Impulse function $\delta(t)$	1
$e^{-at}\sin\omega t$	$\frac{\omega}{(s+a)^2+\omega^2}$
$e^{-at}\cos\omega t$	$\frac{s+a}{(s+a)^2+\omega^2}$
$\frac{1}{\omega}\left[(\alpha-a)^2+\omega^2\right]^{\frac{1}{2}} e^{-at}\sin(\omega t+\phi),$	$\frac{s+\alpha}{(s+a)^2+\omega^2}$
$\phi=\tan^{-1}\frac{\omega}{\alpha-a}$	
$\frac{\omega_n}{\sqrt{1-\zeta^2}}e^{-\zeta\omega_n t}\sin\omega_n\sqrt{1-\zeta^2}t, \zeta<1$	$\frac{\omega_n^2}{s^2+2\zeta\omega_n s+\omega_n^2}$

Continuation of the table 2.3

1	2
$\frac{1}{a^2+\omega^2}+\frac{1}{\omega\sqrt{a^2+\omega^2}}e^{-at}\sin(\omega t-\phi),$	$\frac{1}{s\left[(s+a)^2+\omega^2\right]}$
$\phi=\tan^{-1}\frac{\omega}{-a}$	
$1-\frac{1}{\sqrt{1-\zeta^2}}e^{-\zeta\omega_n t}\sin\left(\omega_n\sqrt{1-\zeta^2}t+\phi\right),$	$\frac{\omega_n^2}{s\left(s^2+2\zeta\omega_n s+\omega_n^2\right)}$
$\phi=\cos^{-1}\zeta,\zeta<1$	
$\frac{\alpha}{a^2+\omega^2}+\frac{1}{\omega}\left[\frac{(\alpha-a)^2+\omega^2}{a^2+\omega^2}\right]^{\frac{1}{2}}e^{-at}\sin(\omega t+\phi).$	$\frac{s+\alpha}{s\left[(s+a)^2+\omega^2\right]}$
$\phi=\tan^{-1}\frac{\omega}{\alpha-a}-\tan^{-1}\frac{\omega}{-a}$	

Alternatively, the Laplace variable s can be considered to be the differential operator so that

$$s\equiv\frac{d}{dt}. \tag{2.16}$$

Then we also have the integral operator

$$\frac{1}{s}\equiv\int_{0^-}^{t}dt. \tag{2.17}$$

The inverse Laplace transformation is usually obtained by using the Heaviside partial fraction expansion. This approach is particularly useful for systems analysis and design because the effect of each characteristic root or eigenvalue can be clearly observed.

To illustrate the usefulness of the Laplace transformation and the steps involved in the system analysis, reconsider the spring-mass-damper system described by Equation (2.1), which is

$$M\frac{d^2y}{dt^2}+b\frac{dy}{dt}+ky=r(t). \qquad (2.18)$$

We wish to obtain the response, y, as a function of time. The Laplace transform of Equation (2.18) is

$$M\left(s^2Y(s)-sy\left(0^-\right)-\frac{dy}{dt}\left(0^-\right)\right)+$$

$$+b\left(sY(s)-y\left(0^-\right)\right)+kY(s)=R(s). \qquad (2.19)$$

When

$$r(t)=0,\ \text{and}\ \ y\left(0^-\right)=y_0,\ \ \text{and}\ \ \left.\frac{dy}{dt}\right|_{t=0^-}=0,$$

we have

$$Ms^2Y(s)-Msy_0+bsY(s)-by_0+kY(s)=0. \qquad (2.20)$$

Solving for $Y(s)$, we obtain

$$Y(s)=\frac{(Ms+b)y_0}{Ms^2+bs+k}=\frac{p(s)}{q(s)}. \qquad (2.21)$$

The denominator polynomial $q(s)$, when set equal to zero, is called the characteristic equation because the roots of this equation determine the character of the time response. The roots of this characteristic equation are also called the poles of the system. The roots of the numerator polynomial $p(s)$ are called the zeros of the system;

for example, $s=-\dfrac{b}{M}$ is a zero of Equation (2.21). Poles and zeros are critical frequencies. At the poles, the function $Y(s)$

becomes infinite, whereas at the zeros, the function becomes zero. The complex frequency s-plane plot of the poles and zeros graphically portrays the character of the natural transient response of the system.

For a specific case, consider the system when $\frac{k}{M} = 2$ and $\frac{b}{M} = 3$. Then Equation (2.21) becomes

$$Y(s) = \frac{(s+3)y_0}{(s+1)(s+2)}. \tag{2.22}$$

The poles and zeros of $Y(s)$ are shown on the s-plane in Figure 2.7.

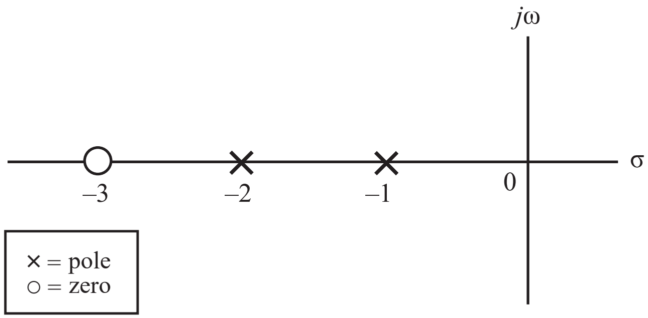


Figure 2.7 – An s-plane pole and zero plot

Expanding Equation (2.22) in a partial fraction expansion, we obtain

$$Y(s) = \frac{k_1}{s+1} + \frac{k_2}{s+2}, \tag{2.23}$$

where k_1 and k_2 are the coefficients of the expansion. The coefficients k_i are called residues and are evaluated by multiplying through by the denominator factor of Equation (2.22) corresponding to k_i and setting s equal to the root. Evaluating k_1 when $y_0 = 1$, we have

$$k_1 = \left.\frac{(s-s_1)\,p(s)}{q(s)}\right|_{s=s_1} = \left.\frac{(s+1)(s+3)}{(s+1)(s+2)}\right|_{s_1=-1} = 2 \qquad (2.24)$$

and $k_2 = -1$. Alternatively, the residues of $Y(s)$ at the respective poles may be evaluated graphically on the .s-plane plot, since Equation (2.24) may be written as

$$k_1 = \left.\frac{(s+3)}{(s+2)}\right|_{s=s_1=-1} = \left.\frac{(s_1+3)}{(s_1+2)}\right|_{s_1=-1} = 2. \qquad (2.25)$$

The graphical representation of Equation (2.25) is shown in Figure 2.8. The graphical method of evaluating the residues is particularly valuable when the order of the characteristic equation is high and several poles are complex conjugate pairs.

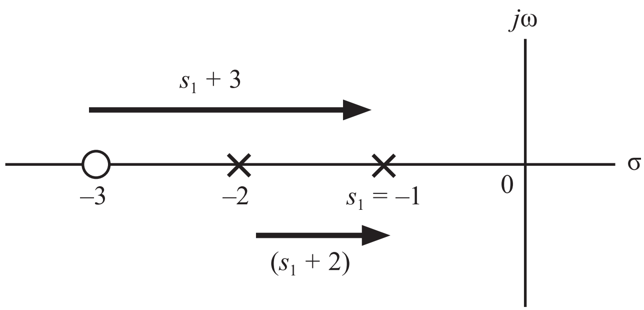


Figure 2.8 – Graphical evaluation of the residues

2.5 The transfer function of linear systems

The transfer function of a linear system is defined as the ratio of the Laplace transform of the output variable to the Laplace transform of the input variable, with all initial conditions assumed to be zero. The transfer function of a system (or element) represents the relationship describing the dynamics of the system under consideration.

A transfer function may be defined only for a linear, stationary (constant parameter) system. A nonstationary system, often called a time-varying system, has one or more time-varying parameters, and the Laplace transformation may not be utilized. Furthermore, a transfer function is an input-output description of the behavior of a system. Thus, the transfer function description does not include any information concerning the internal structure of the system and its behavior. The transfer function of the spring-mass-damper system is obtained from the original Equation (2.19), rewritten with zero initial conditions as follows:

$$Ms^2Y(s) + bsY(s) + kY(s) = R(s). \tag{2.26}$$

Then the transfer function is

$$\frac{Output}{Input} = G(s) = \frac{Y(s)}{R(s)} = \frac{1}{Ms^2 + bs + k}. \tag{2.27}$$

The transfer function of the RC network shown in Figure 2.9 is obtained by writing the Kirchhoff voltage equation, yielding

$$V_1(s) = \left(R + \frac{1}{Cs}\right) I(s), \tag{2.28}$$

expressed in terms of transform variables. We shall frequently refer to variables and their transforms interchangeably. The transform variable will be distinguishable by the use of an uppercase letter or the argument (s).

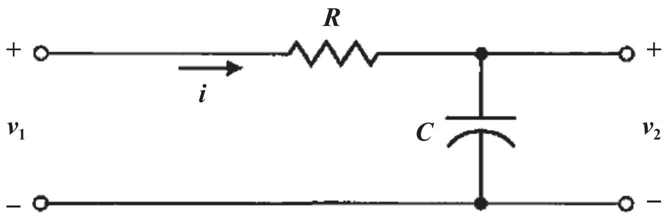


Figure 2.9 – An RC network

The output voltage is

$$V_2(s) = I(s)\left(\frac{1}{Cs}\right). \tag{2.29}$$

Therefore, solving Equation (2.28) for $I(s)$ and substituting in Equation (2.29), we have

$$V_2(s) = \frac{\left(\frac{1}{Cs}\right)V_1(s)}{R + \frac{1}{Cs}}.$$

Then the transfer function is obtained as the ratio $\frac{V_2(s)}{V_1(s)}$, which is

$$G(s) = \frac{V_2(s)}{V_1(s)} = \frac{1}{RCs+1} = \frac{1}{\tau s+1} = \frac{\frac{1}{\tau}}{s+\frac{1}{\tau}}, \tag{2.30}$$

where $\tau = RC$, the time constant of the network. The single pole of $G(s)$ is $s = -\frac{1}{\tau}$. Equation (2.30) could be immediately obtained if one observes that the circuit is a voltage divider, where

$$\frac{V_2(s)}{V_1(s)} = \frac{Z_2(s)}{Z_1(s)+Z_2(s)}, \tag{2.31}$$

and $Z_1(s) = R, Z_2 = \dfrac{1}{Cs}$.

A multiloop electrical circuit or an analogous multiple-mass mechanical system results in a set of simultaneous equations in the Laplace variable. It is usually more convenient to solve the simultaneous equations by using matrices and determinants.

2.6 Block diagram models

The dynamic systems that comprise automatic control systems are represented mathematically by a set of simultaneous differential equations. As we have noted in the previous sections, the Laplace transformation reduces the problem to the solution of a set of linear algebraic equations. Since control systems are concerned with the control of specific variables, the controlled variables must relate to the controlling variables. This relationship is typically represented by the transfer function of the subsystem relating the input and output variables. Therefore, one can correctly assume that the transfer function is an important relation for control engineering.

The importance of this cause-and-effect relationship is evidenced by the facility to represent the relationship of system variables by diagrammatic means. The block diagram representation of the system relationships is prevalent in control system engineering. Block diagrams consist of unidirectional, operational blocks that represent the transfer function of the variables of interest. A block diagram of a field-controlled DC motor and load is shown in Figure 2.10. The relationship between the displacement $\theta(s)$ and the input voltage $V_f(s)$ is clearly portrayed by the block diagram. To represent a system with several variables under control, an interconnection of blocks is utilized. For example, the system shown in Figure 2.11 has two input variables and two output variables.

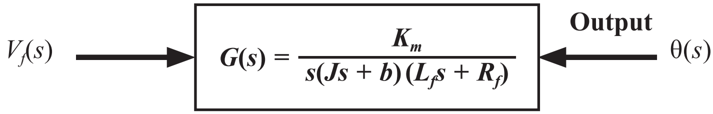


Figure 2.10 – Block diagram of a DC motor

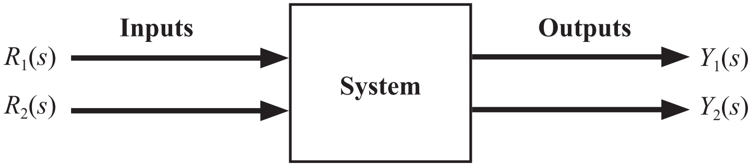


Figure 2.11 – General block representation of two-input, two-output system

Using transfer function relations, we can write the simultaneous equations for the output variables as

$$Y_1(s) = G_{11}(s)R_1(s) + G_{12}(s)R_2(s) \tag{2.32}$$

and

$$Y_2(s) = G_{21}(s)R_1(s) + G_{22}(s)R_2(s), \tag{2.33}$$

where $G_{ij}(s)$ is the transfer function relating the ith output variable to the jth input variable. The block diagram representing this set of equations is shown in Figure 2.12. In general, for J inputs and I outputs, we write the simultaneous equation in matrix form as

$$\begin{bmatrix} Y_1(s) \\ Y_2(s) \\ . \\ . \\ Y_I(s) \end{bmatrix} = \begin{bmatrix} G_{11}(s)\ldots .G_{1J}(s) \\ G_{21}(s)\ldots .G_{2J}(s) \\ . \\ . \\ . \\ G_{I1}(s)\ldots G_{IJ}(s) \end{bmatrix} \begin{bmatrix} R_1(s) \\ R_2(s) \\ . \\ . \\ . \\ R_J(s) \end{bmatrix} \tag{2.34}$$

or simply

$$Y = GR. \tag{2.35}$$

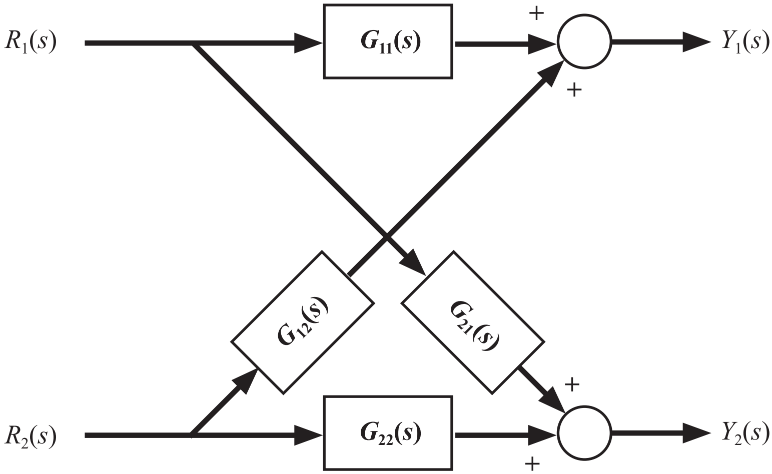
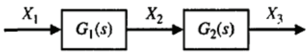
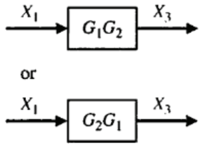
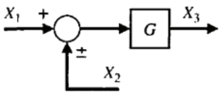
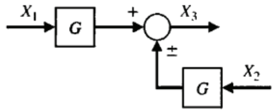
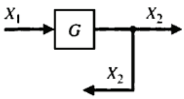
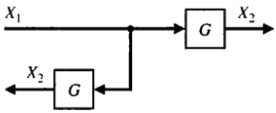
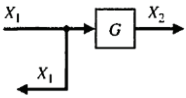
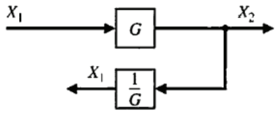
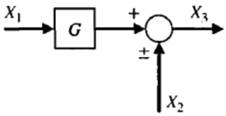
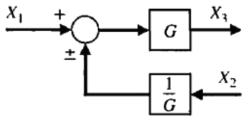
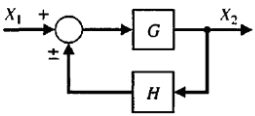
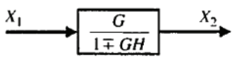


Figure 2.12 – Block diagram of interconnected system

Here the Y and R matrices are column matrices containing the I output and the J input variables, respectively, and G is an I by J transfer function matrix. The matrix representation of the interrelationship of many variables is particularly valuable for complex multivariable control systems.

The block diagram representation of a given system often can be reduced to a simplified block diagram with fewer blocks than the original diagram. Since the transfer functions represent linear systems, the multiplication is commutative.

Block Diagram Transformations

Transformation	Original Diagram	Equivalent Diagram
1. Combining blocks in cascade		
2. Moving a summing point behind a block		
3. Moving a pickoff point ahead of a block		
4. Moving a pickoff point behind a block		
5. Moving a summing point ahead of a block		
6. Eliminating a feedback loop		

3. THE STATE VARIABLES MODELS

In the preceding chapter, we developed and studied several useful approaches to the analysis and design of feedback systems. The Laplace transform was used to transform the differential equations representing the system to an algebraic equation expressed in terms of the complex variable s. Using this algebraic equation, we were able to obtain a transfer function representation of the input-output relationship.

The ready availability of digital computers makes it practical to consider the time domain formulation of the equations representing control systems. The time-domain techniques can be used for nonlinear, time-varying, and multivariable systems.

A time-varying control system is a system in which one or more of the parameters of the system may vary as a function of time.

For example, the mass of a missile varies as a function of time as the fuel is expended during flight. A multivariable system, is a system with several input and output signals.

The solution of a time-domain formulation of a control system problem is facilitated by the availability and ease of use of digital computers. Therefore we are interested in reconsidering the time-domain description of dynamic systems as they are represented by the system differential equation.

The time domain is the mathematical domain that incorporates the response and description of a system in terms of time, t. The time-domain representation of control systems is an essential basis for modern control theory and system optimization.

3.1 The state variables of a dynamic system

The time-domain analysis and design of control systems uses the concept of the state of a system.

The state of a system is a set of variables whose values, together with the input signals and the equations describing the dynamics, will provide the future state and output of the system.

For a dynamic system, the state of a system is described in terms of a set of state variables [$x_1(t)$, $x_2(t)$, …., $x_n(t)$]. The state variables are those variables that determine the future behavior of a system when the present state of the system and the excitation signals are known. Consider the system shown in Figure 3.1, where $y_1(t)$ and $y_2(t)$ are the output signals and $u_1(t)$ and $u_2(t)$ are the input signals.

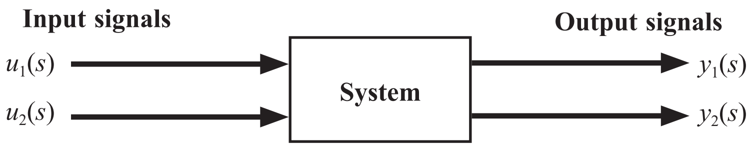


Figure 3.1 – System block diagram

A set of state variables (x_1, x_2, …, x_n) for the system shown in the figure is a set such that knowledge of the initial values of the state variables [$x_1(t_0)$, $x_2(t_0)$, …, $x_n(t_0)$] at the initial time t_0, and of the input signals $u_1(t)$ and $u_2(t)$ for $t \geq t_0$, suffices to determine the future values of the outputs and state variables.

The state variables describe the present configuration of a system and can be used to determine the future response, given the excitation inputs and the equations describing the dynamics.

The general form of a dynamic system is shown in Figure 3.2. A simple example of a state variable is the state of an on-off light switch. The switch can be in either the on or the off position, and thus the state of the switch can assume one of two possible values. Thus, if we know the present state (position) of the switch at to and if an input is applied, we are able to determine the future value of the state of the element.

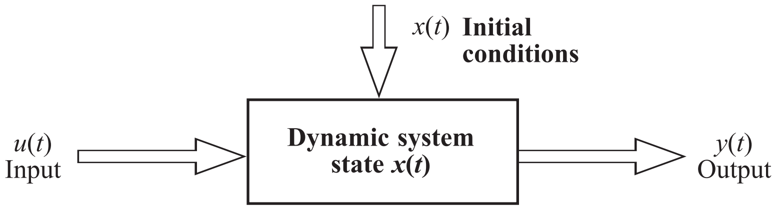


Figure 3.2 – Dynamic system

The concept of a set of state variables that represent a dynamic system can be illustrated in terms of the spring-mass-damper system shown in Figure 3.3.

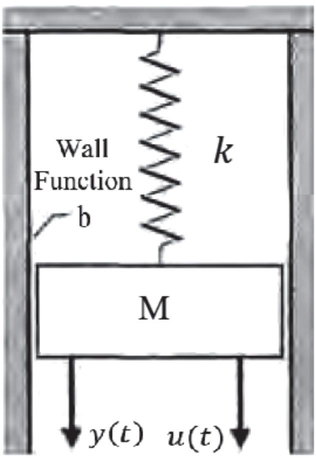


Figure 3.3 – A spring-mass-damper system

The number of state variables chosen to represent this system should be as small as possible in order to avoid redundant state variables. A set of state variables sufficient to describe this system includes the position and the velocity of the mass. Therefore, we will define a set of state variables as (x_1, x_2), where

$$x_1(t) = y(t) \text{ and } x_2(t) = \frac{dy(t)}{dt}.$$

The differential equation describes the behavior of the system and is usually written as

$$M\frac{d^2y}{dt^2} + b\frac{dy}{dt} + ky = u(t). \tag{3.1}$$

To write Equation (3.1) in terms of the state variables, we substitute the state variables as already defined and obtain

$$M\frac{dx_2}{dt} + bx_2 + kx_1 = u(t). \tag{3.2}$$

Therefore, we can write the equations that describe the behavior of the spring-massdamper system as the set of two first-order differential equations

$$\frac{dx_1}{dt} = x_2 \tag{3.3}$$

and

$$\frac{dx_2}{dt} = \frac{-b}{M}x_1 - \frac{k}{M}x_2 + \frac{1}{M}u. \tag{3.4}$$

This set of differential equations describes the behavior of the state of the system in terms of the rate of change of each state variable.

As another example of the state variable characterization of a system, consider the RLC circuit shown in Figure 3.4. The state of this system can be described by a set of state variables (x_1, x_2) where x_1 is the capacitor voltage $v_c(t)$ and x_2 is the inductor current $i_L(t)$.

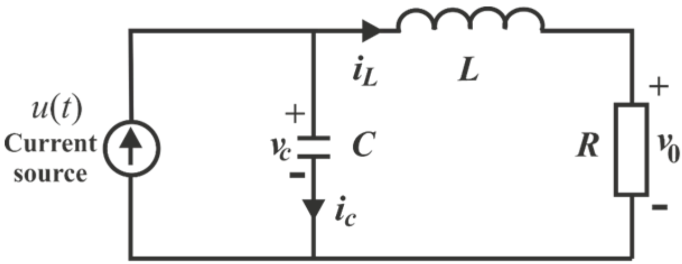


Figure 3.4 – An RLC circuit

This choice of state variables is intuitively satisfactory because the stored energy of the network can be described in terms of these variables as

$$\mathcal{E} = \frac{1}{2} L i_L^{\,2} + \frac{1}{2} C v_c^2 . \tag{3.5}$$

Therefore $x_1(t_0)$ and $x_2(t_0)$ provide the total initial energy of the network and the state of the system at $t = t_0$. For a passive RLC network, the number of state variables required is equal to the number of independent energy-storage elements. Utilizing Kirchhoff's current law at the junction, we obtain a first-order differential equation by describing the rate of change of capacitor voltage as

$$i_c = C \frac{dv_c}{dt} = +u(t) - i_L . \tag{3.6}$$

Kirchhoff's voltage law for the right-hand loop provides the equation describing the rate of change of inductor current as

$$L \frac{di_L}{dt} = -R i_L + v_c . \tag{3.7}$$

The output of this system is represented by the linear algebraic equation

$$v_0 = R i_L(t) .$$

We can rewrite Equations (3.6) and (3.7) as a set of two first-order differential equations in terms of the state variables x_1 and x_2 as follows:

$$\frac{dx_1}{dt} = -\frac{1}{C}x_2 + \frac{1}{C}y(t) \tag{3.8}$$

and

$$\frac{dx_2}{dt} = +\frac{1}{L}x_1 - \frac{R}{L}x_2\,. \tag{3.9}$$

The output signal is then

$$y_1(t) = v_0(t) = Rx_2\,. \tag{3.10}$$

Utilizing Equations (3.8) and (3.9) and the initial conditions of the network represented by $[x_1(t_0, x_2(t_0)]$ we can determine the system's future behavior and its output.

The state variables that describe a system are not a unique set, and several alternative sets of state variables can be chosen. For example, for a second-order system, such as the spring-mass-damper or RLC circuit, the state variables may be any two independent linear combinations of $x_1(t_0)$ and $x_2(t_0)$ For the RLC circuit, we might choose the set of state variables as the two voltages, $v_c(t)$ and $v_L(t)$ where v_L is the voltage drop across the inductor. Then the new state variables, x_1^* and x_2^* are are related to the old state variables, x_1 and, x_2 as

$$x_1^* = v_c = x_1 \tag{3.11}$$

and

$$x_2^* = v_L = v_c - Ri_L = x_1 - Rx_2\,. \tag{3.12}$$

Equation (3.12) represents the relation between the inductor voltage and the former state variables v_c and i_L. In a typical system, there are several choices of a set of state variables that specify the energy stored in a system and therefore adequately describe the dynamics of the system. It is usual to choose a set of state variables that can be readily measured.

An alternative approach to developing a model of a device is the use of the bond graph. Bond graphs can be used for electrical, mechanical, hydraulic, and thermal devices or systems as well as for combinations of various types of elements. Bond graphs produce a set of equations in the state variable form.

The state variables of a system characterize the dynamic behavior of a system. The engineer's interest is primarily in physical systems, where the variables are voltages, currents, velocities, positions, pressures, temperatures, and similar physical variables. However, the concept of system state is not limited to the analysis of physical systems and is particularly useful in analyzing biological, social, and economic systems. For these systems, the concept of state is extended beyond the concept of the current configuration of a physical system to the broader viewpoint of variables that will be capable of describing the future behavior of the system.

3.2 The state differential equation

The response of a system is described by the set of first-order differential equations written in terms of the state variables

$$\left(x_1, x_2, \ldots, x_n\right)$$

and the inputs

$$\left(u_1, u_2, \ldots, u_m,\right).$$

These first-order differential equations can be written in general form as

$$\dot{x}_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + b_{11}u_1 + \ldots + b_{1m}u_m,$$

$$\dot{x}_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + b_{21}u_1 + \ldots + b_{2m}u_m,$$
$$\vdots$$
$$\dot{x}_n = a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n + b_{n1}u_1 + \ldots + b_{nm}u_m, \tag{3.13}$$

where $\dot{x} = \dfrac{dx}{dt}$. Thus, this set of simultaneous differential equations can be written in matrix form as follows:

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{1n} \\ \vdots & & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nm} \end{bmatrix}\begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}. \tag{3.14}$$

The column matrix consisting of the state variables is called the state vector and is written as

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \tag{3.15}$$

where the boldface indicates a vector. The vector of input signals is defined as u.

Then the system can be represented by the compact notation of the state differential equation as

$$\dot{x} = Ax + Bu\,. \tag{3.16}$$

The differential equation (3.16) is also commonly called the state equation.

The matrix A is an n x n square matrix, and B is an n x m matrix. The state differential equation relates the rate of change of the state of the system to the state of the system and the input signals. In general, the outputs of a linear system can be related to the state variables and the input signals by the output equation.

4. FEEDBACK CONTROL SYSTEM CHARACTERISTICS

4.1 Introduction

A control system is defined as an interconnection of components forming a system that will provide a desired system response. Because the desired system response is known, a signal proportional to the error between the desired and the actual response is generated. The use of this signal to control the process results in a closed-loop sequence of operations that is called a feedback system. This closedloop sequence of operations is shown in Figure 4.1. The introduction of feedback to improve the control system is often necessary. It is interesting

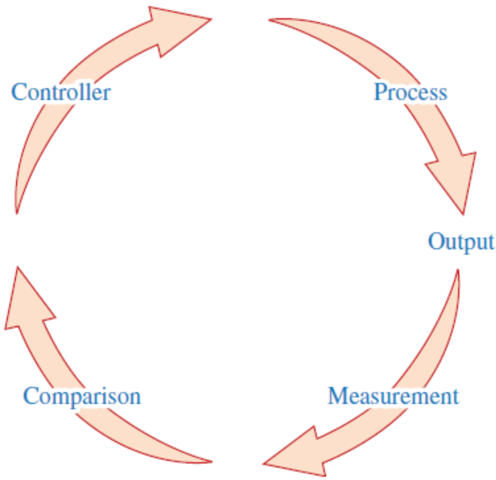


Figure 4.1 – A closed-loop system

that this is also the case for systems in nature, such as biological and physiological systems; feedback is inherent in these systems. For example, the human heartrate control system is a feedback control system. To illustrate the characteristics and advantages of introducing feedback, we will consider a single-loop feedback system. Although many control systems are multiloop, a single-loop system is illustrative. A thorough comprehension of the benefits of feedback can best be obtained from the single-loop system and then extended to multiloop systems.

A system without feedback, often called an open-loop system, is shown in Figure 4.2. The disturbance, *Td*(*s*), directly influences the output, *Y*(*s*). In the absence of feedback, the control system is highly sensitive to disturbances and to changes in parameters of *G*(*s*).

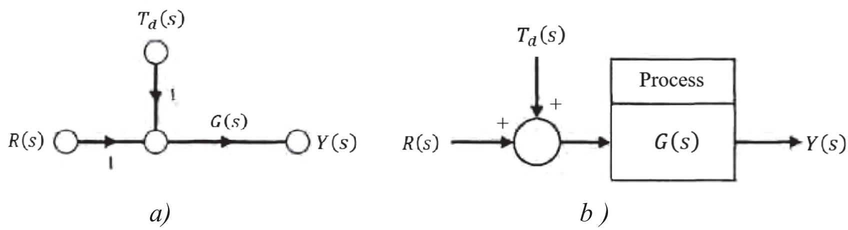


Figure 4.2 – An open-loop system with a disturbance input, $T_d(s)$, (*a*) Signal-flow graph, (*b*) Block diagram

An open-loop system operates without feedback and directly generates the output in response to an input signal.

By contrast, a closed-loop, negative feedback control system is shown in Figure 4.3.

A closed-loop system uses a measurement of the output signal and a comparison with the desired output to generate an error signal that is used by the controller to adjust the actuator.

The two forms of control systems are shown in both block diagram and signal-flow graph form. Despite the cost and increased system complexity, closed-loop feedback control has the following advantages:

- Decreased sensitivity of the system to variations in the parameters of the process.
- Improved rejection of the disturbances.
- Improved measurement noise attenuation,
- Improved reduction of the steady-state error of the system.
- Easy control and adjustment of the transient response of the system.

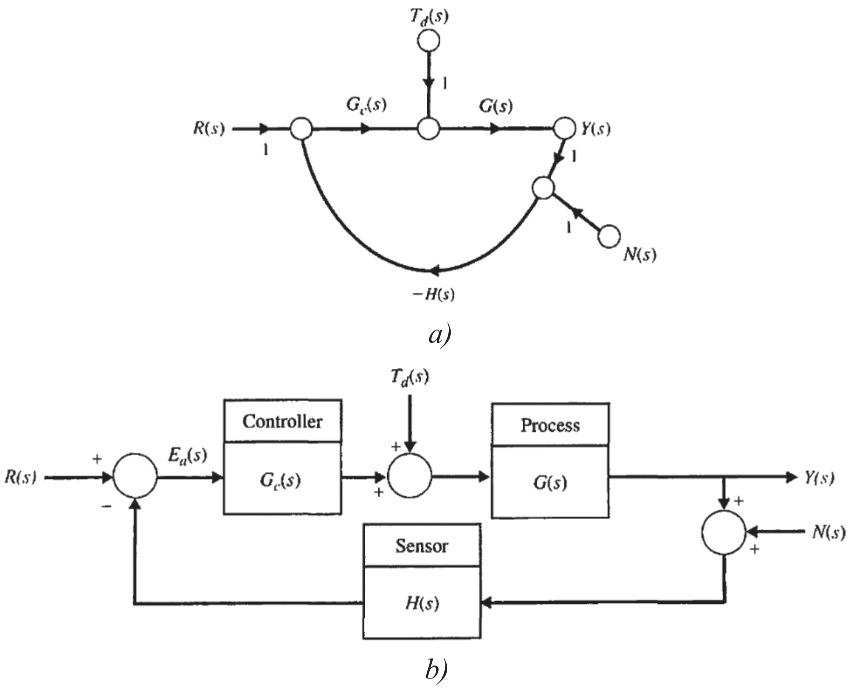


Figure 4.3 – A closed-loop control system:
(*a*) Signal-flow graph, (*b*) Block diagram

4.2 Error signal analysis

The closed-loop feedback control system shown in Figure 4.3 has three inputs – *R*(*s*), *Td*(*s*), and *N*(*s*) – and one output, *Y*(*s*). The signals *Td*(*s*) and *N*(*s*) are the disturbance and measurement noise signals, respectively. Define the tracking error as

$$E(s) = R(s) - Y(s). \quad (4.1)$$

For ease of discussion, we will consider a unity feedback system, that is, *H*(*s*)= 1, in Figure 4.3.

After some block diagram manipulation, we find that the output is given by

$$Y(s) = \frac{G_c(s)G(s)}{1+G_c(s)G(s)}R(s) + \frac{G(s)}{1+G_c(s)G(s)}T_d(s) - \frac{G_c(s)G(s)}{1+G_c(s)G(s)}N(s). \quad (4.2)$$

Therefore, with *E*(*s*) = *R*(*s*) – *Y*(*s*), we have

$$E(s) = \frac{1}{1+G_c(s)G(s)}R(s) - \frac{G(s)}{1+G_c(s)G(s)}T_d(s) - \frac{G_c(s)G(s)}{1+G_c(s)G(s)}N(s). \quad (4.3)$$

Define the function

$$L(s) = G_c(s)G(s).$$

The function, *L*(*s*), is known as the loop gain and plays a fundamental role in control system analysis [12]. In terms of *L*(*s*) the tracking error is given by

$$E(s) = \frac{1}{1+L(s)}R(s) - \frac{G(s)}{1+L(s)}T_d(s) + \frac{L(s)}{1+L(s)}N(s). \quad (4.4)$$

We can define the function

$$F(s) = 1 + L(s).$$

Then, in terms of $F(s)$, we define the sensitivity function as

$$S(s)=\frac{1}{F(s)}=\frac{1}{1+L(s)}. \tag{4.5}$$

Similarly, in terms of the loop gain, we define the complementary sensitivity function as

$$C(s)=\frac{L(s)}{1+L(s)}. \tag{4.6}$$

In terms of the functions $S(s)$ and $C(s)$, we can write the tracking error as

$$E(s)=S(s)R(s)-S(s)G(s)T_d(s)+C(s)N(s). \tag{4.7}$$

Examining Equation (4.7), we see that (for a given $G(s)$), if we want to minimize the tracking error, we want both $S(s)$ and $C(s)$ to be small. Remember that $S(s)$ and $C(s)$ are both functions of the controller, $Gc(s)$, which the control design engineer must select. However, the following special relationship between $S(s)$ and $C(s)$ holds

$$S(s)+C(s)=1. \tag{4.8}$$

We cannot simultaneously make $S(s)$ and $C(s)$ small. Obviously, design compromises must be made.

To analyze the tracking error equation, we need to understand what it means for a transfer function to be "large" or to be "small". However, for our purposes here, we describe the magnitude of the loop gain $L(s)$ by considering the magnitude $|L(j\omega)|$ over the range of frequencies, $\boldsymbol{\omega}$, of interest.

Considering the tracking error in Equation (4.4), it is evident that, for a given $G(s)$, to reduce the influence of the disturbance, $Td(s)$, on the tracking error, $E(s)$, we desire $L(s)$ to be large over the range of frequencies that characterize the disturbances.

That way, the transfer function $\dfrac{G(s)}{1+L(s)}$ will be small, thereby reducing the influence of *Td*(*s*). Since L(s)= *Gc*(*s*)*G*(*s*), this implies that we need to design the controller Gc(s) to have a large magnitude. Conversely, to attenuate the measurement noise, *N*(*s*), and reduce the influence on the tracking error, we desire *L*(*s*) to be small over the range of frequencies that characterize the measurement noise. The transfer function $\dfrac{L(s)}{1+L(s)}$ will be small, thereby reducing the influence of *Td*(*s*). Again, since *L*(*s*)= *Gc*(*s*) *G*(*s*), that implies that we need to design the controller *Gc*(*s*) to have a small magnitude. Fortunately, the apparent conflict between wanting to make *Gc*(*s*) large to reject disturbances and the wanting to make *Gc*(*s*) small to attenuate measurement noise can be addressed in the design phase by making the loop gain, *L*(*s*), large at low frequencies (generally associated with the frequency range of disturbances), and making *L*(*s*) small at high frequencies (generally associated with measurement noise).

4.3 Disturbance signals in a feedback control system

An important effect of feedback in a control system is the control and partial elimination of the effect of disturbance signals. A disturbance signal is an unwanted input signal that affects the output signal. Many control systems are subject to extraneous disturbance signals that cause the system to provide an inaccurate output. Electronic amplifiers have inherent noise generated within the integrated circuits or transistors; radar antennas are subjected to wind gusts; and many systems generate unwanted distortion signals due to nonlinear elements. The benefit of feedback systems is that the effect of distortion, noise, and unwanted disturbances can be effectively reduced.

Disturbance Rejection

When $R(s) = N(s) = 0$, it follows from Equation (4.4) that

$$E(s) = -S(s)G(s)T_d(s) = \frac{G(s)}{1+L(s)}T_d(s).$$

For a fixed $G(s)$ and a given $T_d(s)$, as the loop gain $L(s)$ increases, the effect of $T_d(s)$ on the tracking error decreases. In other words, the sensitivity function $S(s)$ is small when the loop gain is large. We say that large loop gain leads to good disturbance rejection. More precisely, for good disturbance rejection, we require a large loop gain over the frequencies of interest associated with the expected disturbance signals.

In practice, the disturbance signals are often low frequency. When that is the case, we say that we want the loop gain to be large at low frequencies. This is equivalent to stating that we want to design the controller $Gc(s)$ so that the sensitivity function $S(s)$ is small at low frequencies. As a specific example of a system with an unwanted disturbance, let us consider again the speed control system for a steel rolling mill. The rolls, which process steel, are subjected to large load changes or disturbances. As a steel bar approaches the rolls (see Figure 4.4), the rolls are empty. However, when the bar engages in the rolls, the load on the rolls increases immediately to a large value. This loading effect can be approximated by a step change of disturbance torque. Alternatively, the response can be seen from the speed-torque curves of a typical motor, as shown in Figure 4.6.

The transfer function model of an armature-controlled DC motor with a load torque disturbance was determined in Example 2.5 and is shown in Figure 4.5, where it is assumed that La is negligible. Let $R(s) = 0$ and examine $E(s) = -\omega(s)$, for a disturbance $T_d(s)$.

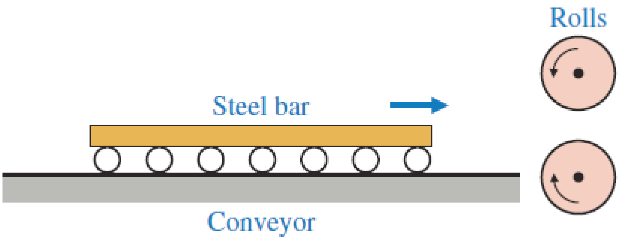


Figure 4.4 – Steel rolling mill

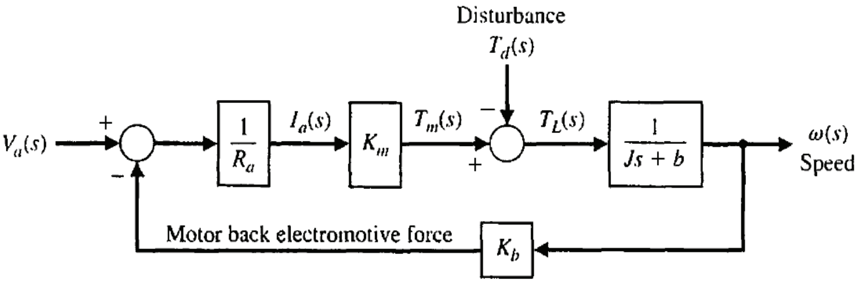


Figure 4.5 – Open-loop speed control system
(without tachometer feedback)

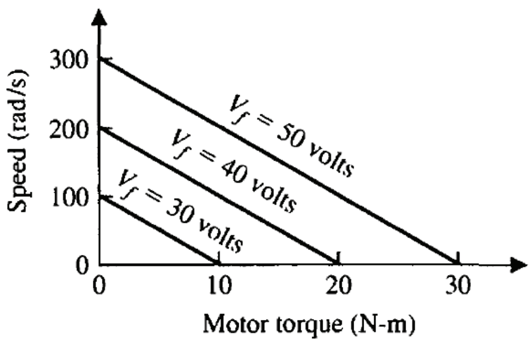


Figure 4.6 – Motor speed-torque curves

The change in speed due to the load disturbance is then

$$E(s) = -\omega(s) = \frac{1}{Js + b + \dfrac{K_m K_b}{R_a}} T_d(s). \tag{4.9}$$

The steady-state error in speed due to the load torque, $T_d(s) = \dfrac{D}{s}$, is found by using the final-value theorem. Therefore, for the open-loop system, we have

$$\lim_{t\to\infty} E(t) = \lim_{s\to 0} sE(s) = \lim_{s\to 0} s \frac{1}{Js + b + \dfrac{K_m K_b}{R_a}} \left(\frac{D}{s}\right) =$$

$$= \frac{D}{b + \dfrac{K_m K_b}{R_a}} = -\omega(\infty). \tag{4.10}$$

The closed-loop speed control system is shown in block diagram form in Figure 4.7.

The closed-loop system is shown in signal-flow graph and block diagram form in Figure 4.8, where $Gx(s) = \dfrac{KaKm}{Ra}$, $G2(s) = \dfrac{1}{Js + b}$, and $H(s) = Kt + \dfrac{Kb}{Ka}$. The error, $E(s) = -\omega(s)$, of the closed-loop system of Figure 4.10 is:

$$E(s) = -\omega(s) = \frac{G_2(s)}{1 + G_1(s) G_2(s) H(s)} T_d(s). \tag{4.11}$$

Then, if $GiG2H(s)$ is much greater than 1 over the range of s, we obtain the approximate result

$$E(s) \approx \frac{1}{G_1(s) H(s)} T_d(s). \tag{4.12}$$

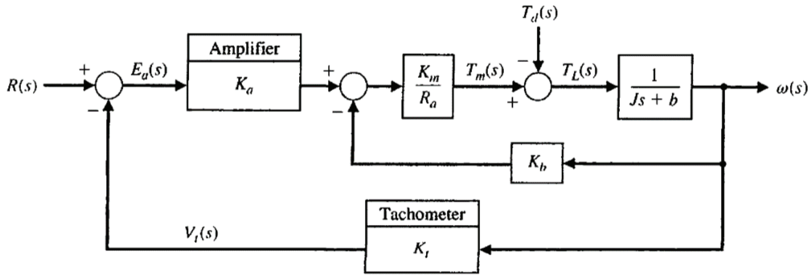
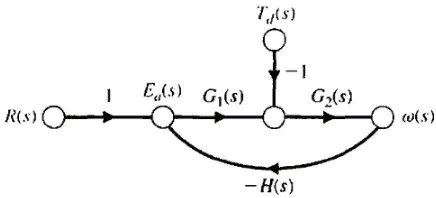
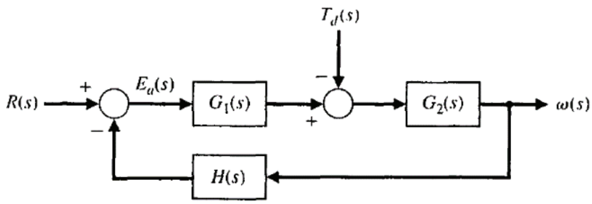


Figure 4.7 – Closed-loop speed tachometer control system



a)



b)

Figure 4.8 – Closed-loop system:
(*a*) Signal-flow graph mode, (*b*) Block diagram model

Therefore, if $Gx(s)H(s)$is made sufficiently large, the effect of the disturbance can be decreased by closed-loop feedback. Note that

$$C_1(s)H(s) = \frac{K_a K_m}{R_a}\left(K_1 + \frac{K_b}{K_a}\right) \approx \frac{K_a K_m K_t}{R_a}$$

since $Ka \gg Kb$. Thus, we strive to obtain a large amplifier gain, Ka, and keep $Ra < 2\ \Omega$. The error for the system shown in Figure 4.10 is

$$E(s) = R(s) - \omega(s)$$

and $R(s) = \omega_d(s)$, the desired speed. For calculation ease, we let $R(s) = 0$ and examine $\omega(s)$.

To determine the output for the speed control system of Figure 4.7, we must consider the load disturbance when the input $R(s) = 0$. This is written as

$$\omega(s) = \frac{-\dfrac{1}{Js+b}}{1+\left(\dfrac{K_t K_a K_m}{R_a}\right)\left[\dfrac{1}{(Js+b)}\right]+(K_m K_b / R_a\left[\dfrac{1}{(Js+b)}\right]} T_d(s) =$$

$$= \frac{-1}{Js+b+\left(\dfrac{K_m}{R_a}\right)\left(K_t K_a + K_b\right)} T_d(s). \qquad (4.13)$$

The steady-state output is obtained by utilizing the final-value theorem, and we have

$$\lim_{t\to\infty}\omega(t) = \lim_{s\to 0} s\omega(s)) = \frac{-1}{b+\left(\dfrac{K_m}{R_a}\right)\left(K_1 K_a + K_b\right)} D \qquad (4.14)$$

when the amplifier gain K_a is sufficiently high, we have

$$\omega(\infty) \approx \frac{-R_a}{K_a K_m K_t} D = \omega_c(\infty). \qquad (4.32)$$

The ratio of closed-loop to open-loop steady-state speed output due to an undesired disturbance is

$$\frac{\omega_c(\infty)}{\omega_0(\infty)} = \frac{R_a b + K_m K_b}{K_a K_m K_t} \tag{4.33}$$

and is usually less than 0.02.

This advantage of a feedback speed control system can also be illustrated by considering the speed-torque curves for the closed-loop system, which are shown in

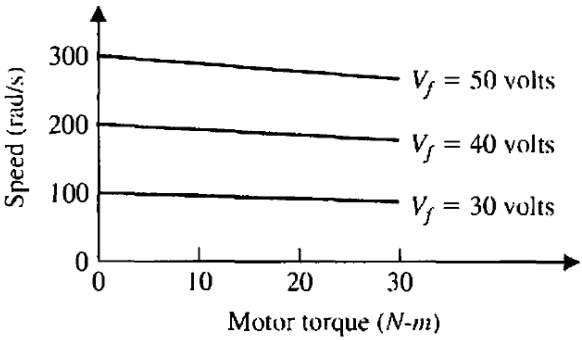


Figure 4.9 – The speed-torque curves for the closed-loop system

Figure 4.9. The improvement of the feedback system is evidenced by the almost horizontal curves, which indicate that the speed is almost independent of the load torque.

Measurement Noise Attenuation. When $R(s) = Td(s) = 0$, it follows from Equation (4.4) that

$$E(s) = C(s)N(s) = \frac{L(s)}{1+L(s)}N(s).$$

As the loop gain $L(s)$ decreases, the effect of $N(s)$ on the tracking error decreases. In other words, the complementary sensitivity function $C(s)$ is small when the loop gain $L(s)$ is small. If we design $Gc(s)$ such that $L(s) \ll 1$, then the noise is attenuated because

$$C(s) \approx L(s).$$

We say that small loop gain leads to good noise attenuation. More precisely, for effective measurement noise attenuation, we need a small loop gain over the frequencies associated with the expected noise signals. In practice, measurement noise signals are often high frequency. Thus we want the loop gain to be low at high frequencies. This is equivalent to a small complementary sensitivity function at high frequencies. The separation of disturbances (at low frequencies) and measurement noise (at high frequencies) is very fortunate because it gives the control system designer a way to approach the design process: the controller should be high gain at low frequencies and low gain at high frequencies. Remember that by low and high we mean that the loop gain magnitude is low/high at the various high/low frequencies. It is not always the case that the disturbances are low frequency or that the measurement noise is high frequency. For example, an astronaut running on a treadmill on a space station may impart disturbances to the spacecraft at high frequencies. If the frequency separation does not exist, the design process usually becomes more involved (for example, we may have to use notch filters to reject disturbances at known high frequencies). A noise signal that is prevalent in many systems is the noise generated by the measurement sensor. This noise, $N(s)$, can be represented as shown in Figure 4.3. The effect of the noise on the output is

$$Y(s) = \frac{-G_c(s)G(s)}{1+G_c(s)G(s)} N(s) \tag{4.17}$$

which is approximately

$$Y(s) \cong -N(s) \tag{4.18}$$

for large loop gain $L(s)=\dfrac{Gc(s)}{G(s)}$. This is consistent with the earlier discussion that smaller loop gain leads to measurement noise attenuation. Clearly, the designer must shape the loop gain appropriately.

The equivalency of sensitivity, S_G^T, and the response of the closed-loop system tracking error to a reference input can be illustrated by considering Figure 4.3. The sensitivity of the system to *G*(*s*) is

$$S_G^T=\frac{1}{1+G_c(s)G(s)}=\frac{1}{1+L(s)}. \tag{4.19}$$

The effect of the reference on the tracking error (with *Td*(*s*) = 0 and *N*(*s*) = 0) is

$$\frac{E(s)}{R(s)}=\frac{1}{1+G_c(s)G(s)}=\frac{1}{1+L(s)}. \tag{4.20}$$

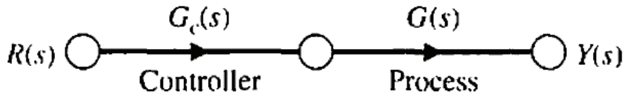
In both cases, we find that the undesired effects can be alleviated by increasing the loop gain. Feedback in control systems primarily reduces the sensitivity of the system to parameter variations and the effect of disturbance inputs. Note that the measures taken to reduce the effects of parameter variations or disturbances are equivalent, and fortunately, they reduce simultaneously. As a final illustration, consider the effect of the noise on the tracking error:

$$\frac{E(s)}{T_d(s)}=\frac{G_c(s)G(s)}{1+G_c(s)G(s)}=\frac{L(s)}{1+L(s)}. \tag{4.21}$$

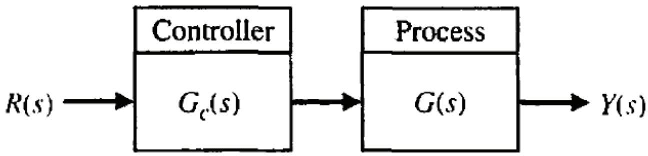
We find that the undesired effects of measurement noise can be alleviated by decreasing the loop gain. Keeping in mind the relationship *S*(*s*) + *C*(*s*) = 1, the trade-off in the design process is evident.

4.4 Control of the transient response

One of the most important characteristics of control systems is their transient response. The transient response is the response of a system as a function of time. Because the purpose of control systems is to provide a desired response, the transient response of control systems often must be adjusted until it is satisfactory. If an open loop control system does not provide a satisfactory response, then the process, $G(s)$, must be replaced with a more suitable process. By contrast, a closed-loop system can often be adjusted to yield the desired response by adjusting the feedback loop parameters. It is often possible to alter the response of an open-loop system by inserting a suitable cascade controller, $G_c(s)$, preceding the process, $G(s)$, as shown in Figure 4.10. Then it is necessary to design the cascade transfer function, $G_c(s)$ $G(s)$, so that the resulting transfer function provides the desired transient response.



a)



b)

Figure 4.10 – Cascade controller system (without feedback): (*a*) Signal-flow graph, (*b*) Block diagram

To make this concept more comprehensible, consider a specific control system, which may be operated in an open- or closed-loop manner. A speed control system, as shown in Figure 4.11, is often used in industrial processes to move materials and products. Several important speed control systems are used in steel mills for rolling the steel sheets and moving the steel through the mill [19]. The transfer function of the open-loop system (without feedback) was obtained earlier. For $\frac{\omega(s)}{V_a(s)}$, we have

$$\frac{\omega(s)}{V_a(s)} = G(s) = \frac{K_1}{\tau_1 s + 1} \tag{4.22}$$

where

$$K_1 = \frac{K_m}{R_a b + K_b K_m} \; and \; \tau_1 = \frac{R_a J}{R_a b + K_b K_m}.$$

In the case of a steel mill, the inertia of the rolls is quite large, and a large armature-controlled motor is required. If the steel rolls are subjected to a step command for a speed change of

$$V_a(s) = \frac{k_2 E}{s} \tag{4.23}$$

the output response is

$$\omega(s) = G(s) V_a(s). \tag{4.24}$$

The transient speed change is then

$$\omega(t) = K_1 (k_2 E)(1 - e^{-\frac{t}{\tau_1}}. \tag{4.25}$$

If this transient response is too slow, we must choose another motor with a different time constant τ_1 if possible. However, because τ_1 is dominated by the load inertia, J, it may not be possible to achieve much alteration of the transient response.

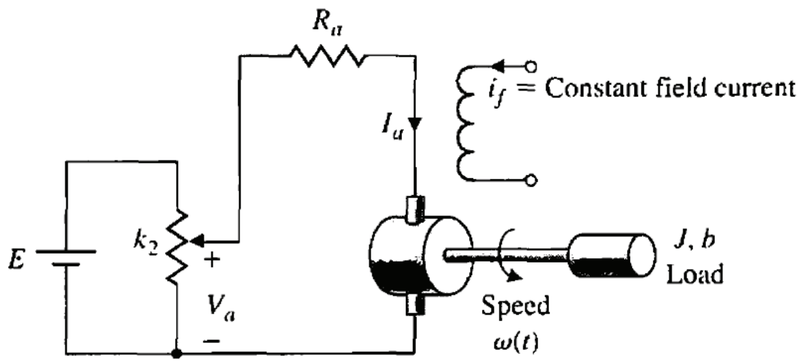


Figure 4.11 – Open-loop speed control system (without feedback)

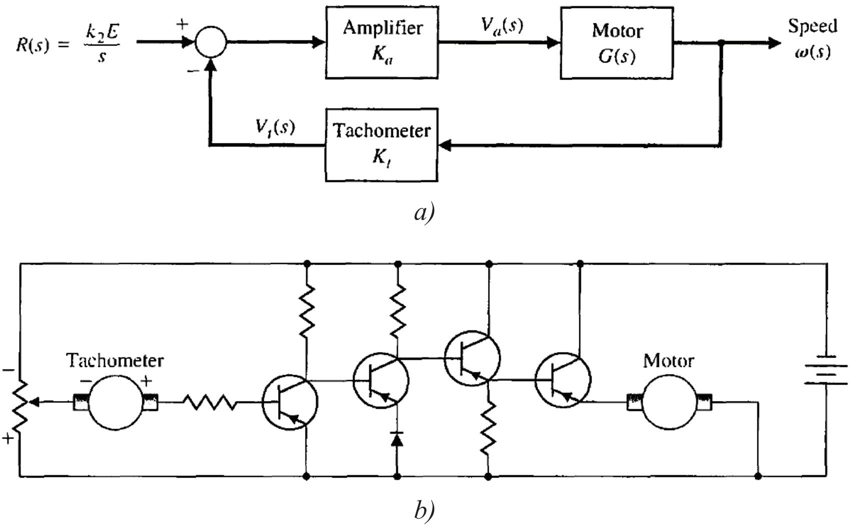


Figure 4.12 – (*a*) Closed-loop speed control system,
(*b*) Transistorized closed-loop speed control system

A closed-loop speed control system is easily obtained by using a tachometer to generate a voltage proportional to the speed, as shown in Figure 4.12(*a*). This voltage is subtracted from the potentiometer voltage and amplified as shown in Figure 4.12(*a*).

A practical transistor amplifier circuit for accomplishing this feedback in low-power applications is shown in Figure 4.12(*b*) [1, 5, 7]. The closed-loop transfer function is

$$\frac{\omega(s)}{R(s)}=\frac{K_a G(s)}{1+K_a K_t G(s)}=\frac{K_a K_1}{\tau_1 s+1+K_a K_t K_1}=\frac{K_a K_1 \frac{1}{\tau 1}}{s+\frac{1+K_a K_t K_1}{\tau_1}}. \quad (4.26)$$

The amplifier gain, K_a, may be adjusted to meet the required transient response specifications. Also, the tachometer gain constant, K_t, may be varied, if necessary.

The transient response to a step change in the input command is then

$$\omega(t)=\frac{K_a K_1}{1+K_a K_1 K_1}(k_2 E)(1-e^{-pt}), \quad (4.27)$$

where $p=\frac{1+K_a K_t K_1}{T_i}$. Because the load inertia is assumed to be very large, we alter the response by increasing *Ka*. Thus, we have the approximate response

$$\omega(t)\approx\frac{1}{K_t}(k_2 E)\left[1-\exp\left(\frac{-(K_a K_1 K_1}{\tau_1}\right)\right]. \quad (4.28)$$

For a typical application, the open-loop pole might be $\frac{1}{\tau_1}=0.10$, whereas the closed-loop pole could be at least $\frac{K_a K_t K_1}{\tau_1}=10$, a factor of one hundred in the improvement of the speed of response. To attain the gain $K_a K_t K_1$ the amplifier gain *Ka* must be reasonably large, and the armature voltage signal to the motor and its associated torque signal must be larger for the closed-loop than for the open-loop operation. Therefore, a higher-power motor will be required to avoid saturation of the motor. The responses of the closed-loop system and the open-loop system are shown in Figure 4.13. Note the rapid response of the closed-loop system.

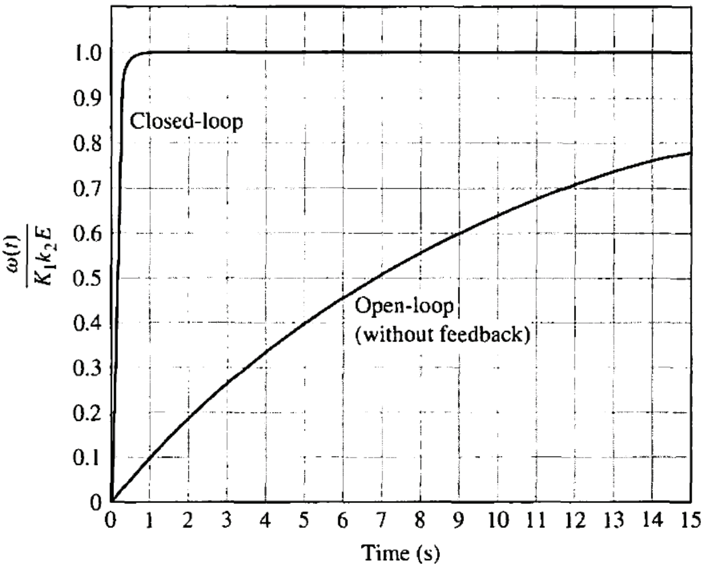


Figure 4.13 – The response of the open-loop and closed-loop speed control system when τ = 10 and K_1KaK_t = 100. The time to reach 98 % of the final value for the open-loop and closed-loop system is 40 seconds and 0.4 seconds, respectively

While we are considering this speed control system, it will be worthwhile to determine the sensitivity of the open- and closed-loop systems. As before, the sensitivity of the open-loop system to a variation in the motor constant or the potentiometer constant k_2 is unity. The sensitivity of the closed-loop system to a variation in K_m is

$$S_{K_m}^T = S_G^T S_{K_m}^G \approx \frac{\left[s + \left(\dfrac{1}{\tau_1}\right)\right]}{s + \dfrac{K_a K_t K_1 + 1}{\tau_1}}.$$

Using the typical values given in the previous paragraph, we have

$$S_{K_m}^T \approx \frac{(s + 0.10)}{s + 10}.$$

We find that the sensitivity is a function of s and must be evaluated for various values of frequency. This type of frequency analysis is straightforward but will be deferred until a later chapter. However, it is clearly seen that at a specific low frequency – for example, $s = j\omega = j_1$ the magnitude of the sensitivity is approximately $\left|S_{K_m}^T\right| \cong 0.1$.

4.5 Steady-state error

A feedback control system is valuable because it provides the engineer with the ability to adjust the transient response. In addition, as we have seen, the sensitivity of the system and the effect of disturbances can be reduced significantly. However, as a further requirement, we must examine and compare the final steady-state error for an open-loop and a closed-loop system. The steady-state error is the error after the transient response has decayed, leaving only the continuous response.

The error of the open-loop system shown in Figure 4.2 is

$$E_0(s) = R(s) - Y(s) = \big(1 - G(s)\big)R(s), \qquad (4.29)$$

when $T_d(s) = 0$. Figure 4.3 shows the closed-loop system. When $T(i(s) = 0$ and $N(s) = 0$, and we let $H(s) = 1$, the tracking error is given by (Equation 4.3)

$$E_c(s) = \frac{1}{1+G_c(s)G(s)}R(s). \tag{4.30}$$

To calculate the steady-state error, we use the final-value theorem

$$\lim_{t\to\infty} e(t) = \lim_{s\to 0} sE(s). \tag{4.32}$$

Therefore, using a unit step input as a comparable input, we obtain for the open loop system

$$e_0(\infty) = \lim_{s\to 0} s\left(1-G(s)\right)\left(\frac{1}{s}\right) = \lim_{s\to 0}\left(1-G(s)\right) = 1G(0). \tag{4.33}$$

For the closed-loop system we have

$$e_c(\infty) = \lim_{s\to 0} s\left(\frac{1}{1+G_c(s)G(s)}\right)\left(\frac{1}{s}\right) = \frac{1}{1+G_c(0)G(0)}.$$

The value of $G(s)$ when $s = 0$ is often called the DC gain and is normally greater than one. Therefore, the open-loop system will usually have a steady-state error of significant magnitude. By contrast, the closed-loop system with a reasonably large DC loop gain $L(0) = G_c(0)G(0)$ will have a small steady-state error.

4.6 The cost of feedback

Adding feedback to a control system results in the advantages outlined in the previous sections. Naturally, however, these advantages have an attendant cost. The first cost of feedback is an increased number of components and complexity in the system. To add the feedback, it is necessary to consider several feedback components; the measurement component (sensor) is the key one. The sensor is often the most expensive component in a control system. Furthermore, the sensor introduces noise and inaccuracies into the system.

The second cost of feedback is the loss of gain. For example, in a single-loop system, the open-loop gain is $G_c(s)G(s)$ and is reduced to $\frac{G_c(s)G(s)}{1+G_c(s)G(s)}$ in a unity negative feedback system. The closed-loop gain is smaller by a factor of $\frac{1}{1+G_c(s)G(s)}$, which is exactly the factor that reduces the sensitivity of the system to parameter variations and disturbances. Usually, we have extra open-loop gain available, and we are more than willing to trade it for increased control of the system response.

We should note that it is the gain of the input-output transmittance that is reduced. The control system does retain the substantial power gain of a power amplifier and actuator, which is fully utilized in the closed-loop system.

The final cost of feedback is the introduction of the possibility of instability. Whereas the open-loop system is stable, the closed-loop system may not be always stable.

The addition of feedback to dynamic systems causes more challenges for the designer. However, for most cases, the advantages far outweigh the disadvantages, and a feedback system is desirable. Therefore, it is necessary to consider the additional complexity and the problem of stability when designing a control system.

5. THE PERFORMANCE OF FEEDBACK CONTROL SYSTEMS

5.1 Introduction

The ability to adjust the transient and steady-state performance is a distinct advantage of feedback control systems. To analyze and design a control system, we must define and measure its performance. Based on the desired performance of the control system, the system parameters may be adjusted to provide the desired response. Because control systems are inherently dynamic, their performance is usually specified in terms of both the transient response and the steady-state response. The transient response is the response that disappears with time. The steady-state response is the response that exists for a long time following an input signal initiation. The design specifications for control systems normally include several time response indices for a specified input command, as well as a desired steady-state accuracy. In the course of any design, the specifications are often revised to effect a compromise. Therefore, specifications are seldom a rigid set of requirements, but rather a first attempt at listing a desired performance. The effective compromise and adjustment of specifications are graphically illustrated in Figure 5.1.

The parameter p may minimize the performance measure M_2 if we select p as a very small value. However, this results in large measure M_1, an undesirable situation. If the performance measures are equally important, the crossover point at $p_{\min}$ provides the best compromise. This type of compromise is normally encountered in control system design. It is clear that if the

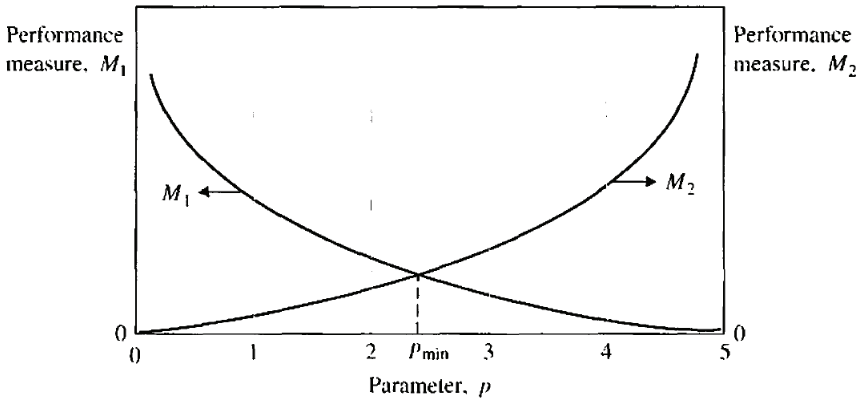


Figure 5.1 – Two performance measures versus parameter ρ

original specifications called for both M_1 and M_2 to be zero, the specifications could not be simultaneously met; they would then have to be altered to allow for the compromise resulting with p_{min}. The specifications, which are stated in terms of the measures of performance, indicate the quality of the system to the designer. In other words, the performance measures help to answer the question: How well does the system perform the task for which it was designed?

5.2 Test input signals

The time-domain performance specifications are important indices because control systems are inherently time-domain systems. That is, the system transient or time performance is the response of prime interest for control systems. It is necessary to determine initially whether the system is stable; we can achieve this goal by using the techniques of ensuing chapters. If the system is stable, the response to a specific input signal will provide several measures of the performance. However, because the actual input signal of the system is usually unknown, a standard **test input**

signal is normally chosen. This approach is quite useful because there is a reasonable correlation between the response of a system to a standard test input and the system's ability to perform under normal operating conditions. Furthermore, using a standard input allows the designer to compare several competing designs. Many control systems experience input signals that are very similar to the standard test signals.

The standard test input signals commonly used are the step input, the ramp input, and the parabolic input. These inputs are shown in Figure 5.2. The equations representing these test signals are given in Table 5.1, where the Laplace transform can be obtained by using Table 2.3. The ramp signal is the integral of the step input, and the parabola is simply the integral of the ramp input.

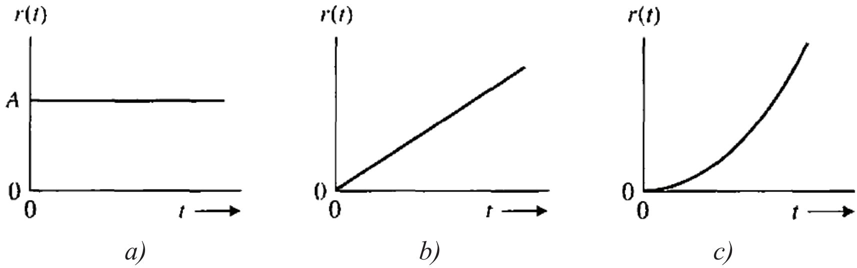


Figure 5.2 – Test input signals: (*a*) step, (*b*) ramp, and (*c*) parabolic

Table 5.1 – Test Signal Inputs

Test Signal	***r*(*t*)**	***R*(*s*)**
Step	$r(t) = A, t > 0 = 0, t < 0$	$R(s) = A/s$
Ramp	$r(t) = A, t > 0 = 0, t < 0$	$R(s) = A/s^2$
Parabolic	$r(t) = At^2, t > 0 = 0, t < 0$	$R(s) = 2A/s^3$

A **unit impulse** function is also useful for test signal purposes. The unit impulse is based on a rectangular function

$$f_{\in}(t)=\begin{cases}\frac{1}{\in},-\frac{\in}{2}\le t\le\frac{\in}{2},\\ 0, otherwise\end{cases}$$

where ∈ > 0. As e approaches zero, the function $f_{\in}(t)$ approaches the unit impulse function 5(*f*), which has the following properties:

$$\int_{-\infty}^{\infty}\delta(t)dt=1 and \int_{-\infty}^{\infty}\delta(t-a)g(t)dt=g(a). \quad (5.1)$$

The impulse input is useful when we consider the convolution integral for the output $y(t)$ in terms of an input $r(t)$, which is written as

$$y(t)=\int_{-\infty}^{t}g(t-\tau)r(\tau)d\tau=\mathcal{L}^{-1}\{G(s)R(s)\}. \quad (5.2)$$

This relationship is shown in block diagram form in Figure 5.3. If the input is a unit impulse function, we have

$$y(t)=\int_{-\infty}^{t}g(t-\tau)\delta(\tau)d\tau. \quad (5.3)$$

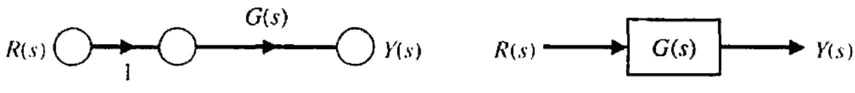


Figure 5.3 – Open-loop control system

The integral has a value only at τ = 0; therefore

$$y(t)=g(t)$$

the impulse response of the system $G(s)$. The impulse response test signal can often be used for a dynamic system by subjecting the system to a large-amplitude, narrowwidth pulse of area A.

The standard test signals are of the general form

$$r(t) = t^n$$

and the Laplace transform is

$$R(s) = \frac{n!}{s^{n+1}}. \tag{5.5}$$

Hence, the response to one test signal may be related to the response of another test signal of the form of Equation (5.4). The step input signal is the easiest to generate and evaluate and is usually chosen for performance tests.

Consider the response of the system shown in Figure 5.3 for a unit step input when

$$G(s) = \frac{9}{s+10}.$$

Then the output is

$$Y(s) = \frac{9}{s(s+10)}$$

the response during the transient period is

$$y(t) = 0.9\left(1 - e^{-10t}\right)$$

and the steady-state response is

$$y(\infty) = 0.9.$$

If the error is $E(s) = R(s) - Y(s)$, then the steady-state error is

$$e_{ss} = \lim_{s \to 0} sE(s) = 0.1.$$

6. THE STABILITY OF LINEAR FEEDBACK SYSTEMS

6.1 The concept of stability

When considering the design and analysis of feedback control systems, stability is of the utmost importance. From a practical point of view, a closed-loop feedback system that is unstable is of little value. As with all such general statements, there are exceptions; but for our purposes, we will declare that all our control designs must result in a closed-loop stable system. Many physical systems are inherently open-loop unstable, and some systems are even designed to be open-loop unstable. Most modern fighter aircraft are open-loop *unstable by design*, and without active feedback control assisting the pilot, they cannot fly. Active control is introduced by engineers to stabilize the unstable system– that is, the aircraft – so that other considerations, such as transient performance, can be addressed. Using feedback, we can stabilize unstable systems and then with a judicious selection of controller parameters, we can adjust the transient performance. For open-loop stable systems, we still use feedback to adjust the closed-loop performance to meet the design specifications. These specifications take the form of steady-state tracking errors, percent overshoot, settling time, time to peak, and the other indices discussed in Chapters 4 and 5.

We can say that a closed-loop feedback system is either stable or it is not stable. This type of stable/not stable characterization is referred to as absolute stability. A system possessing absolute stability is called a stable system – the label of absolute is dropped.

Given that a closed-loop system is stable, we can further characterize the degree of stability. This is referred to as relative stability. The pioneers of aircraft design were familiar with the notion of relative stability – the more stable an aircraft was, the more difficult it was to maneuver (that is, to turn). One outcome of the relative instability of modern fighter aircraft is high maneuverability. A fighter aircraft is less stable than a commercial transport, hence it can maneuver more quickly. In fact, the motions of a fighter aircraft can be quite violent to the "passengers." As we will discuss later in this section, we can determine that a system is stable (in the absolute sense) by determining that all transfer function poles lie in the left-half s-plane, or equivalently, that all the eigenvalues of the system matrix A lie in the left-half s-plane. Given that all the poles (or eigenvalues) are in the left-half s-plane, we investigate relative-stability by examining the relative locations of the poles (or eigenvalues).

A stable system is defined as a system with a bounded (limited) system response. That is, if the system is subjected to a bounded input or disturbance and the response is bounded in magnitude, the system is said to be stable.

A stable system is a dynamic system with a bounded response to a bounded input.

The concept of stability can be illustrated by considering a right circular cone placed on a plane horizontal surface. If the cone is resting on its base and is tipped slightly, it returns to its original equilibrium position. This position and response are said to be stable. If the cone rests on its side and is displaced slightly, it rolls with no tendency to leave the position on its side. This position is designated as the neutral stability. On the other hand, if the cone is placed on its tip and released, it falls onto its side. This position is said to be unstable. These three positions are illustrated in Figure 6.1.

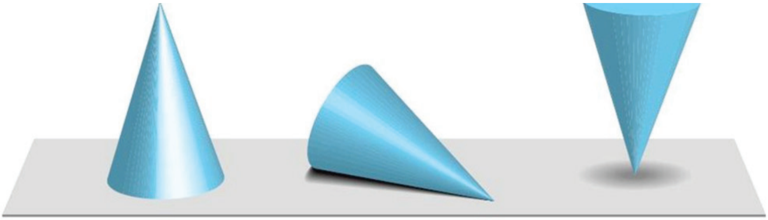


Figure 6.1 – The stability of a cone

The stability of a dynamic system is defined in a similar manner. The response to a displacement, or initial condition, will result in either a decreasing, neutral, or increasing response. Specifically, it follows from the definition of stability that a linear system is stable if and only if the absolute value of its impulse response g(t), integrated over an infinite range, is finite. That is, in terms of the convolution integral Equation (5.2) for a bounded input, $\int_0^\infty |g(t)| dt$ must be finite.

The location in the s-plane of the poles of a system indicates the resulting transient response. The poles in the left-hand portion of the s-plane result in a decreasing response for disturbance inputs. Similarly, poles on the *j*ω-*axis* and in the right-hand plane result in a neutral and an increasing response, respectively, for a disturbance input. This division of the s-plane is shown in Figure 6.2. Clearly, the poles of desirable dynamic systems must lie in the left-hand portion of the s-plane.

A common example of the potential destabilizing effect of feedback is that of feedback hi audio amplifier and speaker systems used for public address in auditoriums. In this case, a loudspeaker produces an audio signal that is an amplified version of the sounds picked up by a microphone. In addition to other audio inputs, the sound coming from the speaker itself may be sensed by the microphone. The strength of this particular

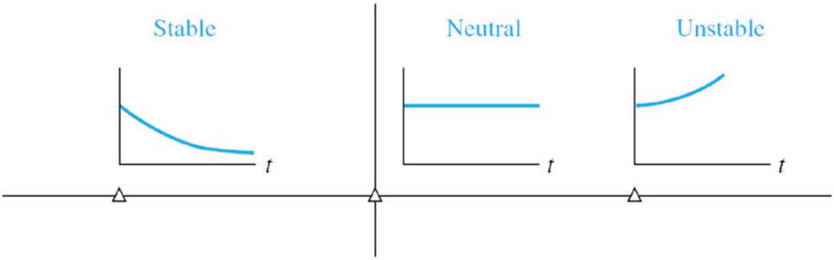


Figure 6.2 – Stability in the s-plane

signal depends upon the distance between the loudspeaker and the microphone. Because of the attenuating properties of air, a larger distance will cause a weaker signal to reach the microphone. Due to the finite propagation speed of sound waves, there will also be a time delay between the signal produced by the loudspeaker and the signal sensed by the microphone. In this case, the output from the feedback path is added to the external input. This is an example of positive feedback.

As the distance between the loudspeaker and the microphone decreases, we find that if the microphone is placed too close to the speaker, then the system will be unstable. The result of this instability is an excessive amplification and distortion of audio signals and an oscillatory squeal.

In terms of linear systems, we recognize that the stability requirement may be defined in terms of the location of the poles of the closed-loop transfer function. The closed-loop system transfer function is written as

$$T(s)=\frac{p(s)}{q(s)}=\frac{K\prod_{i=1}^{M}(s+z_i}{s^N\prod_{k=1}^{Q}(s+\sigma_k)\prod_{m=1}^{R}\left[s^2+2\alpha_m s+(a_m^2+\omega_m^2\right]}, \quad (6.1)$$

where $q(s) = \Delta(s) = 0$ is the characteristic equation whose roots are the poles of the closed-loop system. The output response for an impulse function input (when $N = 0$) is then

$$y(t)=\sum_{k=1}^{Q}A_k e^{-\sigma_k 1}+\sum_{m=1}^{R}B_m\left(\frac{1}{\omega_m}\right)e^{-a_m t}\sin(\omega_m t+\theta_m), \qquad (6.2)$$

where A_k and B_m are constants that depend on σ_k, z_i, α_m, K, and ω_m. To obtain a bounded response, the poles of the closed-loop system must be in the left-hand portion of the s-plane. Thus, a necessary and sufficient condition for a feedback system to be stable is that all the poles of the system transfer function have negative real parts. A system is stable if all the poles of the transfer function are in the left-hand s-plane. A system is not stable if not all the roots are in the left-hand plane. If the characteristic equation has simple roots on the imaginary axis ($j\omega$-*axis*) with all other roots in the left half-plane, the steady-state output will be sustained oscillations for a bounded input, unless the input is a sinusoid (which is bounded) whose frequency is equal to the magnitude of the ($j\omega$-*axis*) roots. For this case, the output becomes unbounded. Such a system is called marginally stable, since only certain bounded inputs (sinusoids of the frequency of the poles) will cause the output to become unbounded. For an unstable system, the characteristic equation has at least one root in the right half of the s-plane or repeated ($j\omega$-*axis*); for this case, the output will become unbounded for any input.

For example, if the characteristic equation of a closed-loop system is

$$(s+10)(s^2+16)=0$$

then the system is said to be marginally stable. If this system is excited by a sinusoid of frequency $\omega = 4$, the output becomes unbounded.

To ascertain the stability of a feedback control system, we could determine the roots of the characteristic polynomial $q(s)$. However, we are first interested in determining the answer to the question, Is the system stable? If we calculate the roots of the characteristic equation in order to answer this question, we have determined much more information than is necessary. Therefore, several methods have been developed that provide the required yes or no answer to the stability question. The three approaches to the question of stability are (1) the s-plane approach, (2) the frequency plane (*j*ω) approach, and (3) the time-domain approach.

6.2 The Routh Hurwitz stability criterion

The discussion and determination of stability has occupied the interest of many engineers. Maxwell and Vyshnegradskii first considered the question of stability of dynamic systems. In the late 1800s, A. Hurwitz and E. J. Routh independently published a method of investigating the stability of a linear system. The Routh-Hurwitz stability method provides an answer to the question of stability by considering the characteristic equation of the system. The characteristic equation in the Laplace variable is written as

$$\Delta(s) = q(s) = a_n s^n + a_{n-z} s^{n-1} + \ldots + a_1 s + a_0 . \qquad (6.3)$$

To ascertain the stability of the system, it is necessary to determine whether any one of the roots of $q(s)$ lies in the right half of the s-plane. If Equation (6.3) is written in factored form, we have

$$a_n (s - r_1)(s - r_2)\ldots(s - r_n) = 0 , \qquad (6.4)$$

where r_i = *ith* root of the characteristic equation. Multiplying the factors together, we find that

$$\begin{gathered} q(s) = a_n s^n - a_n (r_1 + r_2 + \ldots + r_n) s^{n-1} + a_n (r_1 r_2 + r_2 r_3 + \ldots) s^{n-2} - \\ - a_n (r_1 r_2 r_3 + r_1 r_2 r_4 \ldots\ldots) s^{n-3} + \ldots\ldots + a_n (-1)^n r_1 r_2 r_3 \ldots r_n = 0 \end{gathered} . \qquad (6.5)$$

In other words, for an nth-degree equation, we obtain

$$\begin{aligned} q(s) &= \\ &= a_n s^n - a_n (\mathit{sum\ of\ all\ the\ roots}) s^{n-1} + \\ &+ a_n (\mathit{sum\ of\ the\ products\ of\ the\ roots\ taken\ 2\ at\ a\ time}) s^{n-2} - \\ &- a_n (\mathit{sum\ of\ the\ products\ of\ the\ roots\ taken\ 3\ at\ a\ time}) s^{n-3} + \ldots + \\ &+ a_n (-1)^n (\mathit{product\ of\ all\ n\ roots}) = 0. \end{aligned} \tag{6.6}$$

Examining Equation (6.5), we note that all the coefficients of the polynomial must have the same sign if all the roots are in the left-hand plane. Also, it is necessary that all the coefficients for a stable system be nonzero. These requirements are necessary but not sufficient. That is, we immediately know the system is unstable if they are not satisfied; yet if they are satisfied, we must proceed further to ascertain the stability of the system. For example, when the characteristic equation is

$$q(s) = (s+2)(s^2 - s + 4) = (s^3 + s^2 + 2s + 8) \tag{6.7}$$

the system is unstable, and yet the polynomial possesses all positive coefficients.

The Routh-Hurwitz criterion is a necessary and sufficient criterion for the stability of linear systems. The method was originally developed in terms of determinants, but we shall use the more convenient array formulation.

The Routh-Hurwitz criterion is based on ordering the coefficients of the characteristic equation

$$a_n s^n + a_{n-1} s^{n-1} + a_{n-2} s^{n-2} + \ldots + a_1 s + a_0 = 0 \tag{6.8}$$

into an array or schedule as follows [4]:

$$\begin{matrix} s^n & a_n & a_{n-2} & a_{n-4} \\ s^{n-1} & a_{n-1} & a_{n-3} & a_{n-5} \end{matrix}.$$

Further rows of the schedule are then completed as

$$\begin{array}{c|cccc} s^n & a_n & a_{n-2} & a_{n-4} \\ s^{n-1} & a_{n-1} & a_{n-3} & a_{n-5} \\ s^{n-2} & b_{n-1} & b_{n-3} & b_{n-5} \\ s^{n-3} & c_{n-1} & c_{n-3} & n_{n-5} \\ \vdots & \vdots & \vdots & \vdots \\ s^0 & h_{n-1} & & \end{array},$$

where

$$b_{n-1} = \frac{a_{n-1}a_{n-2} - a_n a_{n-3}}{a_{n-1}} = \frac{-1}{a_{n-1}} \begin{vmatrix} a_n & \cdots & a_{n-2} \\ \vdots & \ddots & \vdots \\ a_{n-1} & \cdots & a_{n-3} \end{vmatrix}$$

$$b_{n-3} = -\frac{1}{a_{n-1}} \begin{vmatrix} a_n & a_{n-4} \\ a_{n-1} & a_{n-5} \end{vmatrix}$$

$$c_{n-1} = \frac{-1}{b_{m-1}} \begin{vmatrix} a_{n-1} & a_{n-3} \\ b_{n-1} & a_{n-5} \end{vmatrix}$$

and so on. The algorithm for calculating the entries in the array can be followed on a determinant basis or by using the form of the equation for b_{n-1}.

The Routh-Hurwitz criterion states that the number of roots of $q(s)$ with positive real parts is equal to the number of changes in sign of the first column of the Routh array. This criterion requires that there be no changes in sign in the first column for a stable system. This requirement is both necessary and sufficient.

Four distinct cases or configurations of the first column array must be considered, and each must be treated separately and requires suitable modifications of the array calculation procedure: (1) No element in the first column is zero; (2) there is a zero in the first column, but some other elements of the row containing the zero in the first column are nonzero; (3) there is a zero in the first column, and the other elements of the row containing the zero are also zero; and (4) as in the third case, but with repeated roots on the *jw-axis*.

6.3 The relative stability of feedback control systems

The verification of stability using the Routh-Hurwitz criterion provides only a partial answer to the question of stability. The Routh-Hurwitz criterion ascertains the absolute stability of a system by determining whether any of the roots of the characteristic equation lie in the right half of the j-plane. However, if the system satisfies the Routh-Hurwitz criterion and is absolutely stable, it is desirable to determine the relative stability; that is, it is necessary to investigate the relative damping of each root of the characteristic equation. The relative stability of a system can be defined as the property that is measured by the relative real part of each root or pair of roots. Thus, root r_2 is relatively more stable than the roots $r_1 r_2$ as shown in Figure 6.3. The relative stability of a system can also be defined in terms of the relative damping coefficients ς of each complex root pair and, therefore, in terms of the speed of response and overshoot instead of settling time.

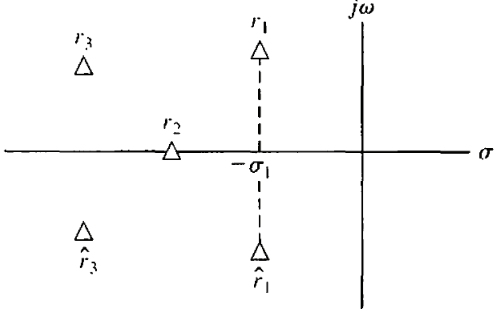


Figure 6.3 – Root locations in the s-plane

Hence, the investigation of the relative stability of each root is clearly necessary because, as we found in Chapter 5, the location of the closed-loop poles in the 5-plane determines the performance of the system.

Thus, it is imperative that we reexamine the characteristic polynomial $q(s)$ and consider several methods for the determination of relative stability.

Because the relative stability of a system is dictated by the location of the roots of the characteristic equation, a first approach using an s-plane formulation is to extend the Routh-Hurwitz criterion to ascertain relative stability. This can be simply accomplished by utilizing a change of variable, which shifts the s-plane axis in order to utilize the Routh-Hurwitz criterion. Examining Figure 6.5, we notice that a shift of the vertical axis in the s-plane to $-\sigma_1$ will result in the roots r_1, r_2 appearing on the shifted axis. The correct magnitude to shift the vertical axis must be obtained on a trial-and-error basis. Then, without solving the fifth-order polynomial $q(s)$, we may determine the real part of the dominant roots r_1, r_2.

7. FREQUENCY RESPONSE METHODS

7.1 Introduction

In preceding chapters, the response and performance of a system have been described in terms of the complex frequency variable s and the location of the poles and zeros on the s-plane. A very practical and important alternative approach to the analysis and design of a system is the frequency response method.

The frequency response of a system is defined as the steady-state response of the system to a sinusoidal input signal. The sinusoid is a unique input signal, and the resulting output signal for a linear system, as well as signals throughout the system, is sinusoidal in the steady state; it differs from the input waveform only in amplitude and phase angle.

For example, consider the system

$$Y(s) = T(s)R(s) \, with \, r(t) = A \sin \omega t .$$

We have

$$R(s) = \frac{A\omega}{s^2 + \omega^2}$$

and

$$T(s) = \frac{m(s)}{q(s)} = \frac{m(s)}{\prod_{i=1}^{n}(s + p_i)}$$

where $-p_i$ are assumed to be distinct poles. Then, in partial fraction form, we have

$$Y(s) = \frac{k_1}{s + p_1} + \ldots + \frac{k_n}{s + p_n} + \frac{\alpha s + \beta}{s^2 + \omega^2}.$$

Taking the inverse Laplace transform yields

$$y(t) = k_1 e^{-p_1 t} + \ldots + k_n e^{-p_n^t} + \mathcal{L}^{-1}\left\{\frac{as + \beta}{s^2 + \omega^2}\right\},$$

where α and β are constants which are problem dependent. If the system is stable, then all p_i. have positive real parts and

$$\lim_{t\to\infty} y(t) = \lim_{t\to\infty} \mathcal{L}^{-1}\left\{\frac{as + \beta}{s^2 + \omega^2}\right\},$$

since each exponential term $k_i e^{-p_i^t}$ decays to zero as $t \to \infty$. In the limit for $y(t)$, it can be shown, for $t \to \infty$ (the steady state)

$$\begin{aligned} y(t) = \mathcal{L}^{-1}\left[\frac{as + \beta}{s^2 + \omega^2}\right] &= \frac{1}{\omega}\left|A\omega T(j\omega)\right|\sin(\omega t + \phi) = \\ &= A\left|T(j\omega)\right|\sin(\omega t + \phi). \end{aligned} \tag{7.1}$$

Where $\phi = /T(j\omega)$.

Thus, the steady-state output signal depends only on the magnitude and phase of $T(j\omega)$ at a specific frequency w. Notice that the steady-state response, as described in Equation (7.1), is true only for stable systems, $T(s)$.

One advantage of the frequency response method is the ready availability of sinusoid test signals for various ranges of frequencies and amplitudes. Thus, the experimental determination of the system's frequency response is easily accomplished; it is the most reliable and uncomplicated method for the experimental analysis of a system. The unknown transfer function of a system can be deduced from the experimentally determined frequency response of a system. Furthermore, the design of a system in the frequency domain provides the designer with control of the bandwidth of a system, as well as some measure of the response of the system to undesired noise and disturbances.

A second advantage of the frequency response method is that the transfer function describing the sinusoidal steady-state behavior of a system can be obtained by replacing s with $j\omega$ in the system transfer function $T(s)$. The transfer function representing the sinusoidal steady-state behavior of a system is then a function of the complex variable $j\omega$ and is itself a complex function $T(j\omega)$ that possesses a magnitude and phase angle. The magnitude and phase angle of $T(j\omega)$ are readily represented by graphical plots that provide significant insight into the analysis and design of control systems.

The basic disadvantage of the frequency response method for analysis and design is the indirect link between the frequency and the time domain. Direct correlations between the frequency response and the corresponding transient response characteristics are somewhat tenuous, and in practice the frequency response characteristic is adjusted by using various design criteria that will normally result in a satisfactory transient response.

The Laplace transform pair is written as

$$F(s) = \mathcal{L}\{f(t)\} = \int_0^{\infty} f(t) e^{-st} dt \tag{7.2}$$

and

$$f(t) = \mathcal{L}^{-1}\{F(s)\} = \frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} F(s) e^{st} ds\,, \tag{7.3}$$

where the complex variable $s = \sigma + j\omega$. Similarly, the **Fourier transform pair** is written as

$$F(\omega) = \mathfrak{F}\{f(t)\} = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \tag{7.4}$$

and

$$f(t) = \mathfrak{F}^{-1}\{F(\omega)\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{-j\omega t}d\omega . \qquad (7.5)$$

The Fourier transform exists for $f(t)$ when

$$\int_{-\infty}^{\infty} |f(t)| dt < \infty .$$

The Fourier and Laplace transforms are closely related, as we can see by examining Equations (7.2) and (7.4). When the function $f(t)$ is defined only for $t \geq 0$, as is often the case, the lower limits on the integrals are the same. Then we note that the two equations differ only in the complex variable. Thus, if the Laplace transform of a function $fi(t)$ is known to be $F_1(s)$, we can obtain the Fourier transform of this same time function by setting $s = j\omega$ in $F_1(s)$.

Again, we might ask, Since the Fourier and Laplace transforms are so closely related, why can't we always use the Laplace transform? Why use the Fourier transform at all? The Laplace transform permits us to investigate the s-plane location of the poles and zeros of a transfer function $T(s)$. However, the frequency response method allows us to consider the transfer function $T(j\omega)$ and to concern ourselves with the amplitude and phase characteristics of the system. This ability to investigate and represent the character of a system by amplitude, phase equations, and curves is an advantage for the analysis and design of control systems. If we consider the frequency response of the closed-loop system, we might have an input $r(t)$ that has a Fourier transform in the frequency domain as follows:

$$R(j\omega) = \int_{-\infty}^{\infty} r(t)e^{-j\omega t}dt .$$

Then the output frequency response of a single-loop control system can be obtained by substituting $s = j\omega$ in the closed-loop system relationship, $Y(s) = T(s)R(s)$, so that we have

$$Y(j\omega) = T(j\omega)R(j\omega) = \frac{G(j\omega)}{1 + G(j\omega)H(j\omega)}R(j\omega). \quad (7.6)$$

Using the inverse Fourier transform, the output transient response would be

$$y(t) = \mathfrak{F}^{-1}\{j\omega)\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} Y(j\omega)e^{j\omega t}d\omega. \quad (7.7)$$

However, it is usually quite difficult to evaluate this inverse transform integral for all but the simplest systems, and a graphical integration may be used. Alternatively, as we will note in succeeding sections, several measures of the transient response can be related to the frequency characteristics and utilized for design purposes.

7.2 Frequency response plots

The transfer function of a system $G(s)$ can be described in the frequency domain by the relation

$$G(j\omega) = G(s) \mid s = j\omega = R(\omega) + jX(\omega), \quad (8.8)$$

where

$$R(\omega) = Re\big[Gj\omega)\big]\, and\; X(\omega) = Im\big[Gj\omega)\big].$$

Alternatively, the transfer function can be represented by a magnitude $|G(j\omega)|$ and a phase $\phi(j\omega)$ as

$$G(j\omega) = \left|G(j\omega)\right|e^{j\phi(\omega)} = \left|G(j\omega)\right| / \phi(\omega) \quad (7.9)$$

Where

$$\phi(\omega) = tan^{-1}\frac{X(\omega)}{R(\omega)} and \left|G(j\omega)\right|^2 = \left[R(\omega)\right]^2 + \left[X(\omega)\right]^2.$$

The graphical representation of the frequency response of the system $G(j\omega)$ can utilize either Equation (7.8) or Equation (7.9). The **polar plot** representation of the frequency response is obtained by using Equation (7.8). The coordinates of the polar plot are the real and imaginary parts of $G(j\omega)$, as shown in Figure 7.1. An example of a polar plot will illustrate this approach.

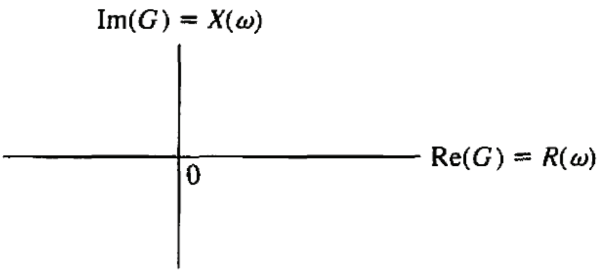


Figure 7.1 – The polar plane

7.3 Bode diagram of an RC filter

The transfer function of Example 8.1 is

$$G(j\omega) = \frac{1}{j\omega(RC)+1} = \frac{1}{j\omega\tau+1}. \qquad (7.10)$$

Where

$$\tau = RC$$

the time constant of the network. The logarithmic gain is

$$20log\left|G\left(j\omega\right)\right| = 20log\left(\frac{1}{1+\left(\omega\tau\right)^2}\right)^{1/2} = -10\log\left(1+\left(\omega\tau\right)^2\right). \quad (7.11)$$

For small frequencies – that is, $\omega \ll 1/\tau$ – the logarithmic gain is

$$20\log\left|G\left(j\omega\right)\right| = -10\log\left(1\right) = 0\,dB, \quad \omega \ll 1/\tau. \quad (7.12)$$

For large frequencies – that is, – the logarithmic gain is

$$20logG\left(j\omega\right) = -20\log\left(\omega\tau\right), \;\; \omega \gg 1/\tau. \quad (7.13)$$

And at $\omega = \dfrac{1}{\tau}$, we have

$$20\log\left|G\left(j\omega\right)\right| = -10\log 2 = -3.01\,dB.$$

The magnitude plot for this network is shown in Figure 7.2(*a*). The phase angle of the network is

$$\phi\left(\omega\right) = -tan^{-1}\left(\omega\tau\right). \quad (7.14)$$

The phase plot is shown in Figure 7.2(*b*). The frequency $\omega = 1/T$ is often called the **break frequency** or **corner frequency**.

A linear scale of frequency is not the most convenient or judicious choice, and we consider the use of a logarithmic scale of frequency. The convenience of a logarithmic scale of frequency can be seen by considering Equation (8.21) for large frequencies $\omega \gg 1/\tau$, as follows:

$$20\log\left|G\left(j\omega\right)\right| = -20\log\left(\omega\tau\right) = -20\log\tau - 20\log\omega\,. \quad (7.15)$$

Then, on a set of axes where the horizontal axis is log ω, the asymptotic curve for $\omega \gg 1/\tau$ is a straight line, as shown in Figure 7.3.

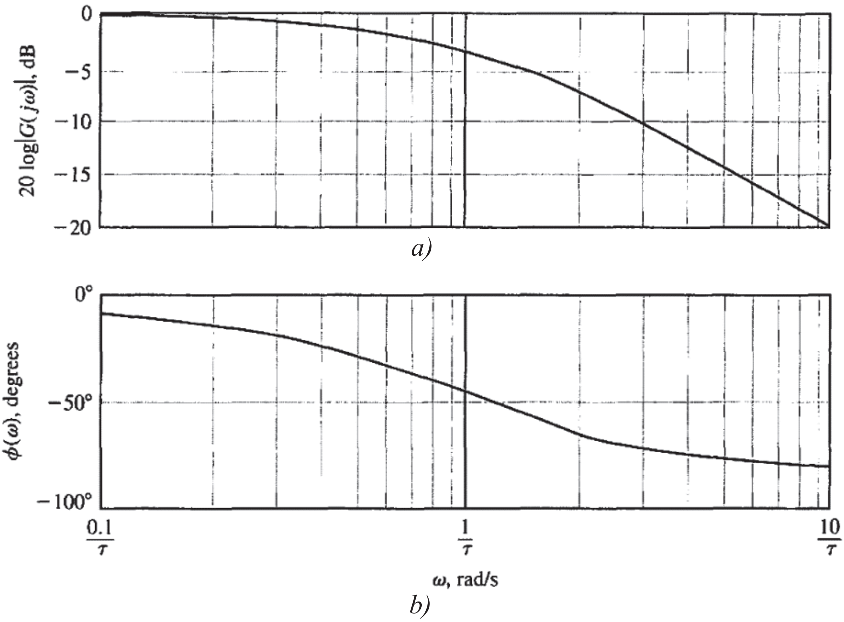


Figure 7.2 – Bode diagram for $G(j\omega) = 1/(j\omega\tau + 1)$:
(*a*) magnitude plot and (*b*) phase plot

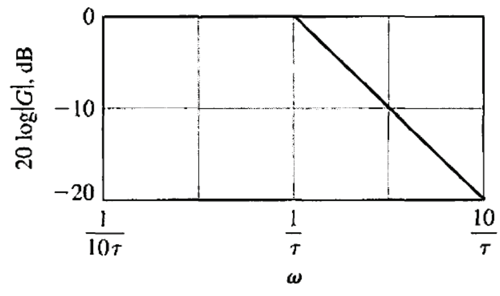


Figure 7.3 – Asymptotic curve for $(j\omega\tau + 1)^{-1}$

The slope of the straight line can be ascertained from Equation (8.21). An interval of two frequencies with a ratio equal to 10 is called a decade, so that the range of frequencies from

ω_1 to ω_2, where $\omega_2 = 10\omega_1$ is called a decade. The difference between the logarithmic gains, for $\omega \gg 1/\tau$, over a decade of frequency is

$$20\,log\left|G\left(j\omega_1\right)\right| - 20\,log\left|G\left(j\omega_2\right)\right| =$$
$$= -20\log\left(\omega_1\tau\right) - \left(-20\log\left(\omega_2\tau\right)\right) =$$
$$= -20\,log\frac{\omega_1\tau}{\omega_2\tau} = -20\,log\frac{1}{10} = +20\,dB$$

that is, the slope of the asymptotic line for this first-order transfer function is -20 dB/decade, and the slope is shown for this transfer function in Figure 8.7. Instead of using a horizontal axis of log ω and linear rectangular coordinates, it is easier to use semilog paper with a linear rectangular coordinate for dB and a logarithmic coordinate for ω. Alternatively, we could use a logarithmic coordinate for the magnitude as well as for frequency and avoid the necessity of calculating the logarithm of the magnitude.

The frequency interval $\omega_2 = 2\omega_1 s$ often used and is called an octave of frequencies. The difference between the logarithmic gains for $\gg 1/\tau$, for an octave, is

$$20\,log\left|G\left(j\omega_1\right)\right| - 20\,log\left|G\left(j\omega_2\right)\right| = -20\log\frac{\omega_1\tau}{\omega_2\tau} = -20\log\frac{1}{2} = 6.02\,dB\,.$$

Therefore, the slope of the asymptotic line is – 6 dB/octave. The primary advantage of the logarithmic plot is the conversion of multiplicative factors, such as ($j\omega\tau$ + 1), into additive factors, 20 log($j\omega\tau$ + 1), by virtue of the definition of logarithmic gain. This can be readily ascertained by considering the generalized transfer function

$$G\left(j\omega\right) =$$
$$= \frac{K_b\prod_{i=1}^{Q}\left(1 + j\omega\tau_i\right)}{\left(j\omega\right)^N\prod_{m=1}^{M}\left(1 + j\omega\tau_m\right)\prod_{k=1}^{R}\left[\left(1 + \left(2\zeta_k / \omega_{nk}\right)j\omega + \left(j\omega / \omega_{nk}\right)^2\right)\right]}.$$

This transfer function includes Q zeros, N poles at the origin, M poles on the real axis, and R pairs of complex conjugate poles. Obtaining the polar plot of such a function would be a formidable task indeed. However, the logarithmic magnitude of $G(j\omega)$ is

$$20\log\left|G(j\omega)\right| = 20\log K_b + 20\sum_{i=1}^{Q}\log\left|1 + j\omega\tau_i\right|$$

$$-20\log\left|(j\omega)^N\right| - 20\sum_{m=1}^{M}\log\left|1 + j\omega\tau_m\right|$$

$$-20\sum_{k=1}^{R}\log\left|1 + \frac{2\zeta_k}{\omega_{nk}} j\omega + \left(\frac{j\omega}{\omega_{nk}}\right)^2\right|, \tag{7.16}$$

and the Bode diagram can be obtained by adding the plot due to each individual factor. Furthermore, the separate phase angle plot is obtained as

$$\phi(\omega) = +\sum_{i=1}^{Q}\tan^{-1}(\omega\tau_i) - N(90°) - \sum_{m=1}^{M}\tan^{-1}(\omega\tau_m)$$

$$-\sum_{k=1}^{R}\tan^{-1}\frac{2\zeta_k\omega_{nk}\omega}{\omega_{nk}^2 - \omega^2}, \tag{7.17}$$

which is simply the summation of the phase angles due to each individual factor of the transfer function.

Therefore, the four different kinds of factors that may occur in a transfer function are as follows:

1. Constant gain K_b.
2. Poles (or zeros) at the origin ($j\omega$).
3. Poles (or zeros) on the real axis ($j\omega\tau + 1$).
4. Complex conjugate poles (or zeros) $[1+\left(\frac{2\zeta}{\omega_n}\right)j\omega+\left(\frac{j\omega}{\omega_n}\right)^2$.

We can determine the logarithmic magnitude plot and phase angle for these four factors and then use them to obtain a Bode diagram for any general form of a transfer function. Typically, the curves for each factor are obtained and then added together graphically to obtain the curves for the complete transfer function. Furthermore, this procedure can be simplified by using the asymptotic approximations to these curves and obtaining the actual curves only at specific important frequencies.

Constant Gain K_b. The logarithmic gain for the constant K_b is 20 logK_b = *constant in dB* and the phase angle is $\phi(\omega) = 0$.

The gain curve is a horizontal line on the Bode diagram. If the gain is a negative value, – K_b, the logarithmic gain remains 20 *log* – K_b.The negative sign is accounted for by the phase angle, –80°.

Poles (or Zeros) at the Origin, (ja)). A pole at the origin has a logarithmic magnitude

$$20\log\left|\frac{1}{j\omega}\right| = -20\log\omega\, dB \tag{7.18}$$

and a phase angle

$$\phi(\omega) = -90°.$$

The slope of the magnitude curve is -20 dB/decade for a pole. Similarly, for a multiple pole at the origin, we have

$$20\log\left|\frac{1}{(j\omega)^N}\right| = -20N\log\omega, \tag{7.19}$$

and the phase is

$$\phi(\omega) = -90° N.$$

In this case, the slope due to the multiple pole is -20iV dB/decade. For a zero at the origin, we have a logarithmic magnitude

$$20\log|j\omega| = +20\log\omega, \quad \phi(\omega) = +90°. \tag{7.20}$$

The Bode diagram of the magnitude and phase angle of $(j\omega)) \pm N$ is shown in Figure 7.4 for $N = 1$ and $N = 2$.

Poles or Zeros on the Real Axis. The pole factor $(1 + (j\omega\tau)^{-1}$ has been considered previously, and we found that, for a pole on the real axis

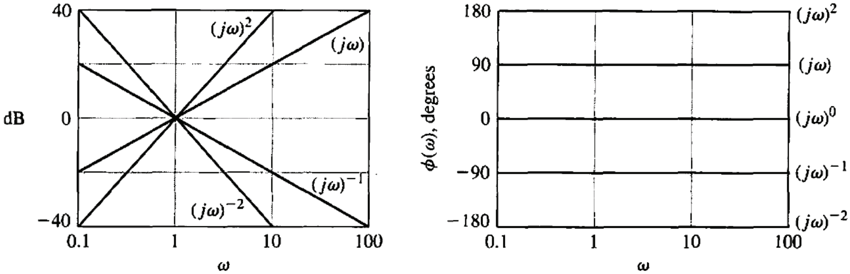


Figure 7.4 – Bode diagram for $(j\omega)^{\pm N}$

$$20\log\left|\frac{1}{1+j\omega\tau}\right| = -10\log\left(1+\omega^2\tau^2\right). \tag{7.20}$$

The asymptotic curve for $\omega \ll 1/\tau$ is $20\ log\ 1 = 0\ dB$, and the asymptotic curve for $\omega \gg 1/\tau$ is $-20\ log(\omega\tau)$, which has a slope of –20 dB/decade. The intersection of the two asymptotes occurs when or when $\omega = 1/\tau$, the break frequency. The actual logarithmic gain when $\omega = 1/\tau$ is –3 dB for this factor. The phase angle is $\phi(\omega) = -tan^{-1}(\omega\tau)$ for the denominator factor. The Bode diagram of a pole factor $(1 + j\omega\tau)^{-1}$ is shown in Figure 7.5

The Bode diagram of a zero factor $1 + j\omega\tau$ is obtained in the same manner as that of the pole. However, the slope is positive at +20 dB/decade, and the phase angle is $\phi(\omega) = +\ tan^{-1}(\omega\tau)$.

A piecewise linear approximation to the phase angle curve can be obtained as shown in Figure 7.6. This linear approximation, which passes through the correct phase at the break frequency, is within 6° of the actual phase curves for all frequencies. This

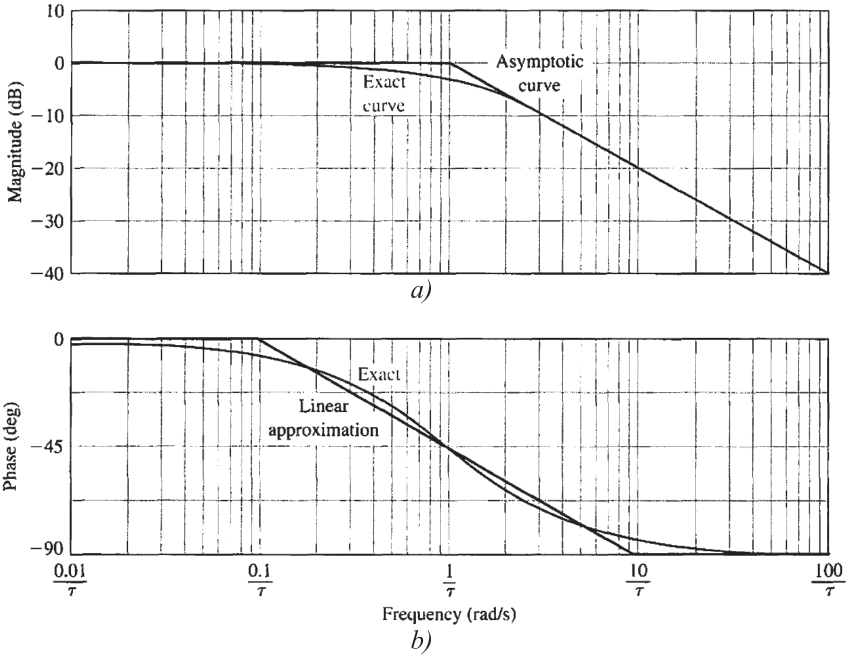


Figure 7.5 – Bode diagram for $(1 + j\omega\tau)^{-1}$

approximation will provide a useful means for readily determining the form of the phase angle curves of a transfer function $G(s)$. However, often the accurate phase angle curves are required, and the actual phase curve for the first-order factor must be obtained via a computer program. The exact values of the frequency response for the pole $(1 + j\omega\tau)^{-1}$, as well as the values obtained by using the approximation for comparison, are given in Table 7.1.

Table 7.1

$\omega\tau$	**0.10**	**0.50**	**0.76**	**1**	**1.31**	**2**	**5**	**10**
$20\log\left\|(1+j\omega\tau)^{-1}\right\|$, dB	-0.04	-1.0	-2.0	-3.0	-4.3	-7.0	-14.2	-20.04
Asymptotic approximation, dB	0	0	0	0	-2.3	-6.0	-14.0	-20.0
$\phi(\omega)$, degrees	-5.7	-26.6	-37.4	-45.0	-52.7	-63.4	-78.7	-84.3
Linear approximation, degrees	0	-31.50	-39.5	-45.0	-50.3	-58.5	-76.5	-90.0

Complex Conjugate Poles or Zeros $1+\left(\frac{2\zeta}{\omega_n}\right)j\omega+\left(\frac{j\omega}{\omega_n}\right)^2$. The quadratic factor for a pair of complex conjugate poles can be written in normalized form as

$$\left[1+j2\zeta u-u^2\right]^{-1}, \tag{7.22}$$

where $u=\omega/\omega_n$. Then the logarithmic magnitude for a pair of complex conjugate poles is

$$20\log\left|G(j\omega)\right|=-10\log\left(\left(1-u^2\right)^2+4\zeta^2u^2\right) \tag{7.23}$$

and the phase angle is

$$\phi(\omega)=-\tan^{-1}\frac{2\zeta u}{1-u^2}. \tag{7.24}$$

When $\ll 1$, the magnitude is

$$20\log\left|G(j\omega)\right|=-10\log 1=0\,dB,$$

and the phase angle approaches 0°. When $u \gg 1$, the logarithmic magnitude approaches

$$20\log\left|G(j\omega)\right|=-10\log u^4=-40\log u,$$

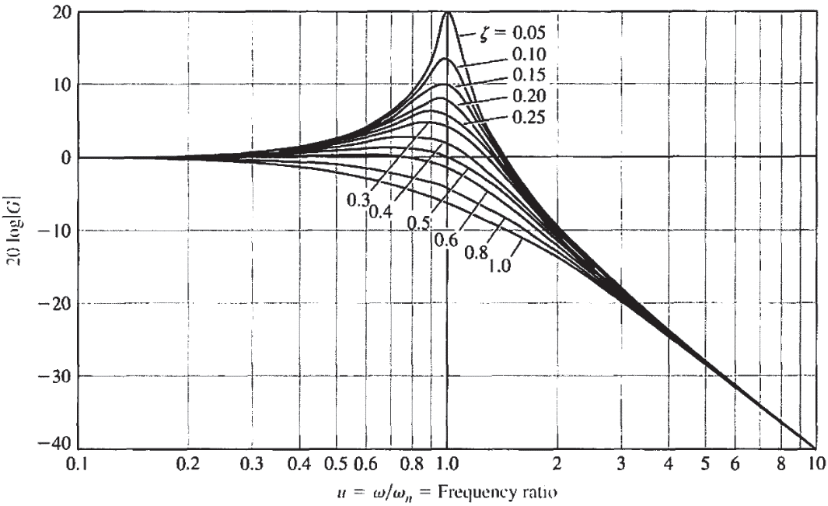
which results in a curve with a slope of -40 dB/decade. The phase angle, when $u \gg 1$, approaches $-180°$. The magnitude asymptotes meet at the 0 dB line when $u = \frac{\omega}{\omega_n} = 1$. However, the difference between the actual magnitude curve and the asymptotic approximation is a function of the damping ratio and must be accounted for when £ < 0.707. The Bode diagram of a quadratic factor due to a pair of complex conjugate poles is shown in Figure 7.6. The maximum value $M_{p\omega}$ of the frequency response occurs at the **resonant frequency** ω_r. When the damping ratio approaches zero, then ω_r approaches ω_n, the **natural frequency**. The resonant frequency is determined by taking the derivative of the magnitude of Equation (7.23) with respect to the normalized frequency, u, and setting it equal to zero. The resonant frequency is given by the relation

$$\omega_r = \omega_n\sqrt{1-2\zeta^2}, \quad \zeta < 0.707, \tag{7.25}$$

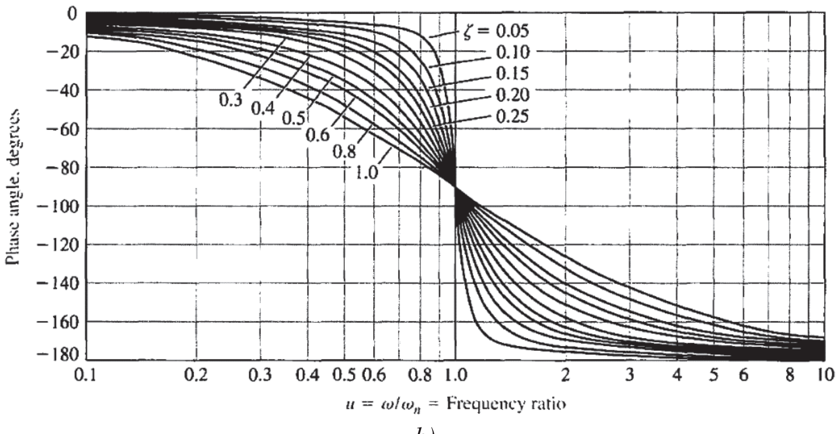
and the maximum value of the magnitude is $M_{p\omega} = \left|G\left(j\omega_r\right)\right| = (2\zeta\sqrt{1-\zeta^2})^{-1}$ $\zeta < 0.707$ for a pair of complex poles.

The maximum value of the frequency response, $M_{p\omega}$, and the resonant frequency ω_r are shown as a function of the damping ratio ζ for a pair of complex poles in Figure 7.7. Assuming the dominance of a pair of complex conjugate closed-loop poles, we find that these curves are useful for estimating the damping ratio of a system from an experimentally determined frequency response. The frequency response curves can be evaluated on the s-plane by determining the vector lengths and angles at various frequencies ω along the (s = +jω)axis. For example, considering the second-order factor with complex conjugate poles, we have

$$G\left(s\right) = \frac{1}{\left(s/\omega_n\right)^2 + 2\zeta s/\omega_n + 1}\frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}. \tag{7.26}$$



a)



b)

Figure 7.6 – Bode diagram for

$$G\left(j\omega\right)=\left[1+\left(2\zeta/\omega_n\right)j\omega+\left(j\omega/\omega_n\right)^2\right]^{-1}$$

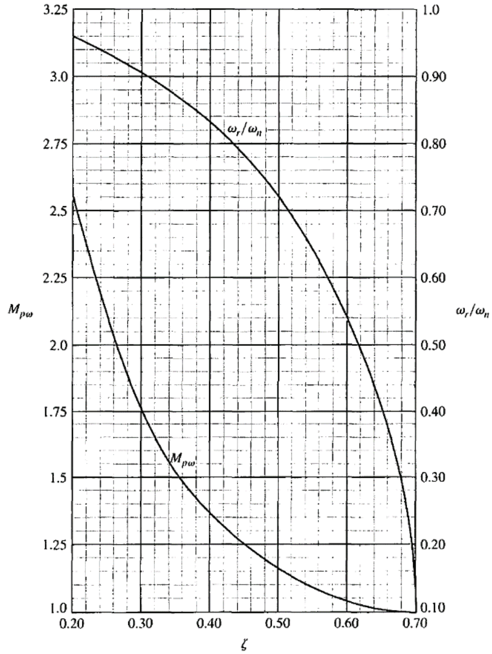


Figure 7.7 – The maximum $M_{p\omega}$ of the frequency response and the resonant frequency ω_r versus ζ for a pair of complex conjugate poles

The poles for varying £ lie on a circle of radius ω_n. The transfer function evaluated for real frequency $s = j\omega$ is written as

$$G(j\omega) = \left.\frac{\omega_n^2}{(s - s_1)(s - \hat{s}_1)}\right|_{s=j\omega} = \frac{\omega_n^2}{(j\omega - s_1)(j\omega - \hat{s}_1)},$$

where s_1 and $\hat{s}_1$ are the complex conjugate poles. The vectors $j\omega - s_1$ and $j\omega - \hat{s}_1$ are the vectors from the poles to the frequency $j\omega$. Then the magnitude and phase may be evaluated for various specific frequencies. The magnitude is

$$G(j\omega)\Big| = \frac{\omega_n^2}{|j\omega - s_1||j\omega - s_1^*|}. \tag{7.27}$$

The magnitude and phase may be evaluated for three specific frequencies, namely,

$$\omega = 0,\ \omega = \omega_r \ \text{ and } \ \omega = \omega_d.$$

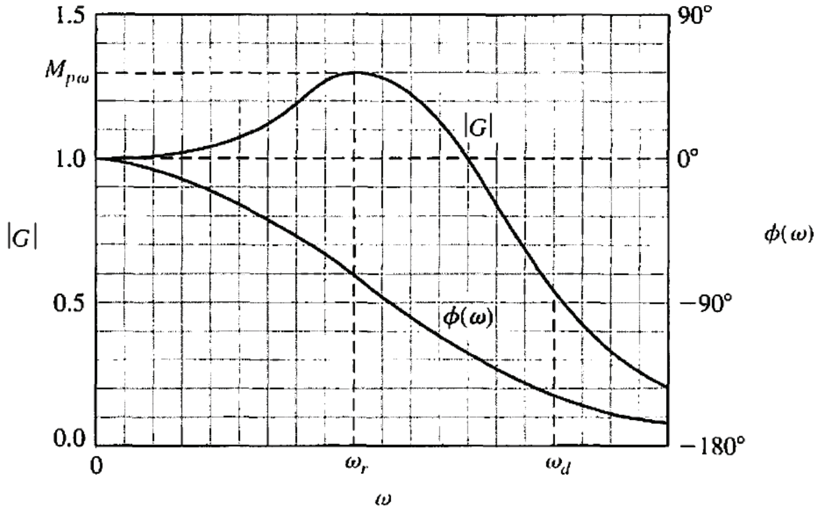


Figure 7.8 – Bode diagram for complex conjugate poles

8. STABILITY IN THE FREQUENCY DOMAIN

8.1 The Nyquist criterion

To investigate the stability of a control system, we consider the characteristic equation, which is F(s)=0, so that

$$F(s) = 1 + L(s) = \frac{K\prod_{i=1}^{n}(s+z_i)}{\prod_{k=1}^{M}(s+p_k)} = 0.$$

For a system to be stable, all the zeros of $F(s)$ must lie in the left-hand s-plane. Thus, we find that the roots of a stable system (the zeros of $F(s)$) must lie to the left of the $j\omega$-axis in the s-plane. Therefore, we choose a contour Γ_s in the s-plane that encloses the entire right-hand s-plane, and we determine whether any zeros of $F(s)$ lie within T_s by utilizing Cauchy's theorem. That is, we plot Γ_F in the $F(s)$-plane and determine the number of encirclements of the origin N. Then the number of zeros of $F(s)$ with in the Γ_s contour (and therefore, the unstable zeros of $F(s)$) is

$$Z = N + P. \tag{8.2}$$

Thus, if $P = 0$, as is usually the case, we find that the number of unstable roots of the system is equal to N, the number of encirclements of the origin of the $F(s)$-plane.

The Nyquist contour that encloses the entire right-hand s-plane is shown in Figure 8.1. The contour Γ_s passes along the $j\omega$ > -axis from $-j\infty$ to $+j\infty$, and this part of the contour provides the familiar $F(j\omega)$. The contour is completed by a semicircular path of radius r, where r approaches infinity so this part of the contour typically maps to a point. This contour *TF* is known as the Nyquist diagram or polar plot.

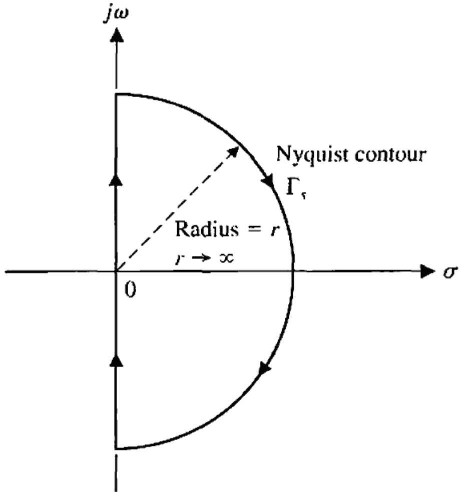


Figure 8.1 – Nyquist contour is shown as the heavy line

Now, the Nyquist criterion is concerned with the mapping of the characteristic equation

$$F(s) = 1 + L(s) \tag{8.3}$$

and the number of encirclements of the origin of the $F(s)$ – plane. Alternatively, we may define the function

$$F'(s) = F(s) - 1 = L(s). \tag{8.4}$$

The change of functions represented by Equation (8.4) is very convenient because $L(s)$ is typically available in factored form, while $1 + L(s)$ is not. Then, the mapping of *Ts* in the s-plane will be through the function $F'(s) = L(s)$ into the $L(s)$-plane. In this case, the number of clockwise encirclements of the origin of the $F(s)$-plane becomes the number of clockwise encirclements of the –1 point in the $F'(s) = L(s)$-plane because

$F'(s) = F(s) - 1$. Therefore, the Nyquist stability criterion can be stated as follows:

A feedback system is stable if and only if the contour *TL* in the *JL*(s)-plane does not encircle the (–1,0) point when the number of poles of $L(s)$ in the righthand s-plane is zero ($P = 0$).

When the number of poles of $L(s)$ in the right-hand s-plane is other than zero, the Nyquist criterion is stated as follows:

A feedback control system is stable if and only if, for the contour *FL*, the number of counterclockwise encirclements of the (–1,0) point is equal to the number of poles of $L(s)$ with positive real parts.

The basis for the two statements is the fact that, for the $F'(s) = L(s)$ mapping, the number of roots (or zeros) of $1 + L(s)$ in the right-hand s-plane is represented by the expression

$$Z = N + P.$$

Clearly, if the number of poles of $L(s)$ in the right-hand s-plane is zero ($P = 0$), we require for a stable system that $N = 0$, and the contour T_p must not encircle the –1 point. Also, if P is other than zero and we require for a stable system that $Z = 0$, then we must have $N = -P$, or P counterclockwise encirclements. It is best to illustrate the use of the Nyquist criterion by completing several examples.

8.2 Relative stability and the nyquist criterion

For the s-plane, we defined the relative stability of a system as the property measured by the relative settling time of each root or pair of roots. Therefore, a system with a shorter settling time is considered more relatively stable. We would like to determine a similar measure of relative stability useful for the frequency response method. The Nyquist criterion provides us with suitable information concerning the absolute stability and, furthermore, can be utilized to define and ascertain the relative stability of a system.

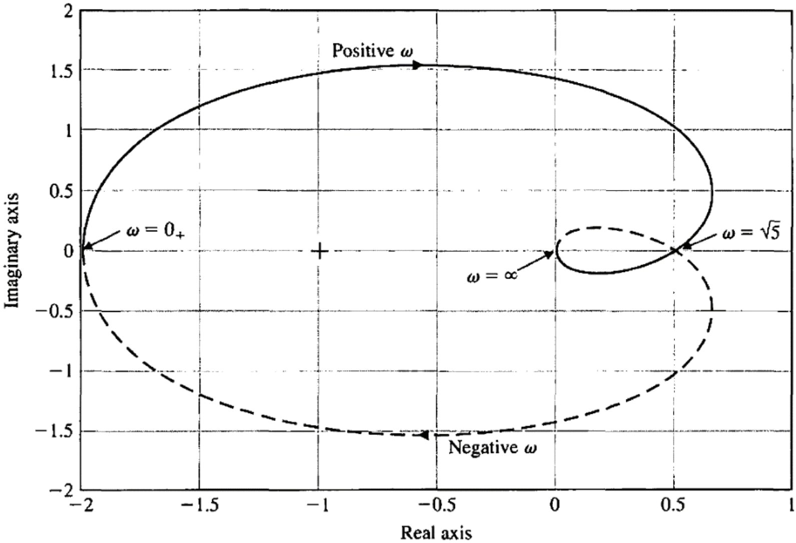


Figure 8.2 – Nyquist diagram for Example 9.6 for $L(j\omega)/K$

The Nyquist stability criterion is defined in terms of the (–1,0) point on thepolar plot or the 0-dB, –180° point on the Bode diagram or log-magnitude-phase diagram. Clearly, the proximity of the $L(j\omega)$-locus to this stability point is a measure of the relative stability of a system. The polar plot for $L(j\omega)$ for several values of K and

$$L(j\omega) = G_c(j\omega)G(j\omega)H\omega = \frac{K}{j\omega(j\omega\tau_1 + 1)(j\omega\tau_2 + 1)}$$

is shown in Figure 8.3.

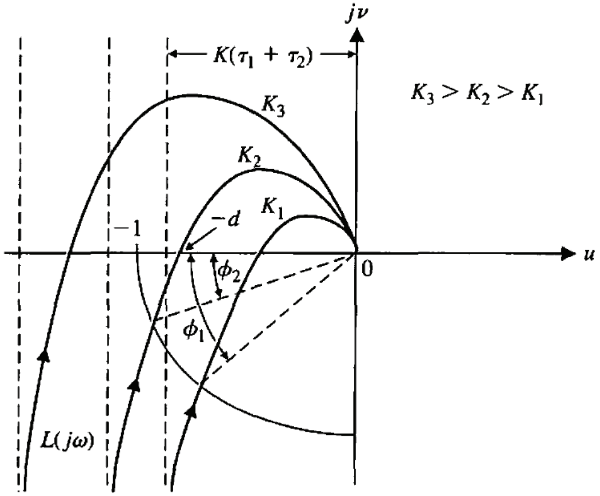


Figure 8.3 – Polar plot for $L(j\omega)$ for three values of gain

As K increases, the polar plot approaches the –1 point and eventually encircles the –1 point for a gain $K = K_3$. The locus intersects the $j\omega$-axis at a point

$$u = \frac{-K\tau_1\tau_2}{\tau_1 + \tau_2}.$$

Therefore, the system has roots on the $j\omega$-axis when

$$u = -1 \text{ or } K = \frac{\tau_1 + \tau_2}{\tau_1\tau_2}.$$

As K is decreased below this marginal value, the stability is increased, and the margin between the critical gain $K = (\tau_1 + \tau_2)/\tau_1\tau_2$ and a gain $K = K_2$ is a measure of the relative stability. This measure of relative stability is called the gain margin and is defined as the reciprocal of the gain $|L(j\omega)|$ at the frequency at which the phase angle reaches –180° (that is, $v = 0$). The gain margin is a measure of the factor by which the system gain would have to be increased for the $L(j\omega)$) locus to pass through the $u = -1$ point. Thus, for a gain $K = K_2$ in Figure 8.3, the gain margin is equal to the reciprocal of $L(j\omega)$ when $v = 0$. Because $\omega = 1/\sqrt{\tau_1\tau_2}$ when the phase shift is $-180°$, we have a gain margin equal to

$$\frac{1}{|L(j\omega)|} = \left[\frac{K_2\tau_1\tau_2}{\tau_1+\tau_2}\right]^{-1} = \frac{1}{d}.$$

The gain margin can be defined in terms of a logarithmic (decibel) measure as

$$20\log\frac{1}{d} = -20\log d \text{ dB}.$$

For example, when $\tau_1 = \tau_2 = 1$, the system is stable when $K \le 2$. Thus, when $K = K2 = 0.5$, the gain margin is equal to

$$\frac{1}{d} = \left[\frac{K_2\tau_1\tau_2}{\tau_1+\tau_2}\right]^{-1} = 4,$$

or, in logarithmic measure

$$20\log 4 = 12dB.$$

Therefore, the gain margin indicates that the system gain can be increased by a factor of four (12 *dB*) before the stability boundary is reached.

The gain margin is the increase in the system gain when phase = $-180°$ that will result in a marginally stable system with intersection of the $-1 + j\omega$ point on the Nyquist diagram.

The gain and phase margins are easily evaluated from the Bode diagram, and because it is preferable to draw the Bode diagram in contrast to the polar plot, it is worthwhile to illustrate the relative stability measures for the Bode diagram. The critical point for stability is $u = -1$, $v = 0$ in the $L(j\omega)$-plane, which is equivalent to a logarithmic magnitude of 0 *dB* and a phase angle of 180° (or $-180°$) on the Bode diagram.

It is relatively straightforward to examine the Nyquist diagram of a minimum phase system. Special care is required with a nonminimum-phase system, however, and the complete Nyquist diagram should be studied to determine stability.

The gain margin and phase margin can be readily calculated by utilizing a computer program, assuming the system is minimum phase. In contrast, for nonminimum-phase systems, the complete Nyquist diagram must be constructed.

The Bode diagram of

$$L(j\omega) = G_c(j\omega)G(j\omega)H(j\omega) = \frac{1}{j\omega(j\omega+1)(0.2j\omega+1)}$$

is shown in Figure 8.4.

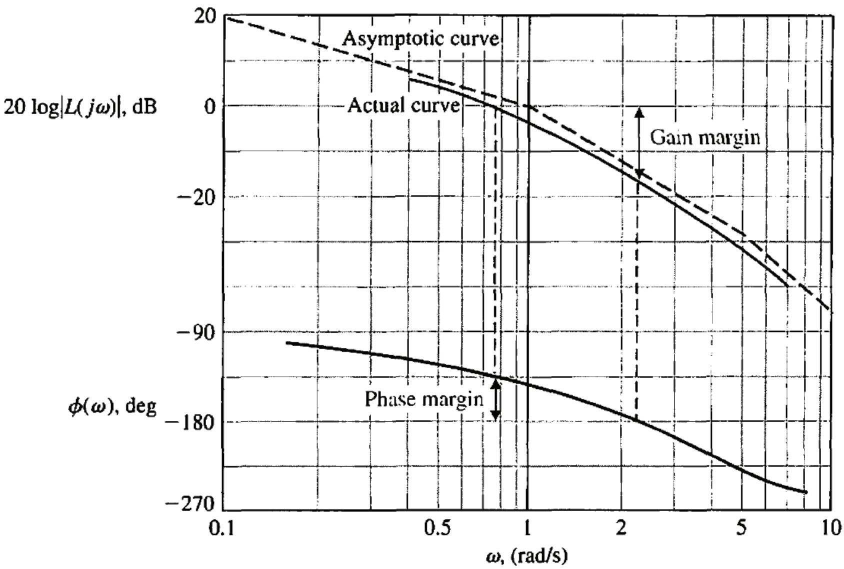


Figure 8.4 – Bode diagram for $L(j\omega) = 1/(j\omega(j\omega + 1)(0.2j\omega + 1))$

The phase angle when the logarithmic magnitude is 0 dB is equal to –137 °. Thus, the phase margin is 180 ° – 137 ° = 43 °, as shown in Figure 8.4. The logarithmic magnitude when the phase angle is $-180°$ is -15 *dB*, and therefore the gain margin is equal to 15 *dB*, as shown in Figure 8.4.

The frequency response of a system can be graphically portrayed on the logarithmic-magnitude-phase-angle diagram. For the log-magnitude-phase diagram, the critical stability point is the 0-*dB*, $-180°$ point, and the gain margin and phase margin can be easily determined and indicated on the diagram. The log-magnitude phase locus of

$$L_1(j\omega) = G_c(j\omega)G(j\omega)H_1(j\omega) = \frac{1}{j\omega(j\omega+1)(0.2j\omega+1)}$$

is shown in Figure 8.5.

The indicated phase margin is $43°$, and the gain margin is 15 *dB*. For comparison, the locus for

$$L_2(j\omega) = G_c(j\omega)G(j\omega)H_2(j\omega) = \frac{1}{j\omega(j\omega+1)^2}$$

is also shown in Figure 8.5.

The gain margin for L_2 is equal to 5.7 *dB*, and the phase margin for L_2 is equal to 20°. Clearly, the feedback system $L_2(j\omega)$ is relatively less stable than the system $L_1(j\omega)$. However, the question still remains: How much less stable is the system $L_2(j\omega)$ in comparison to the system $L_1(j\omega)$? In the following, we answer this question for a second-order system, and the general usefulness of the relation that we develop will depend on the presence of dominant roots.

Let us now determine the phase margin of a second-order system and relate the phase margin to the damping ratio ζ of an underdamped system. Consider the looptransfer function of the system shown in Figure 8.6, where

$$L(s) = G_c(s)G(s)H(s) = \frac{\omega_n^2}{s(s+2\zeta\omega_n)}.$$

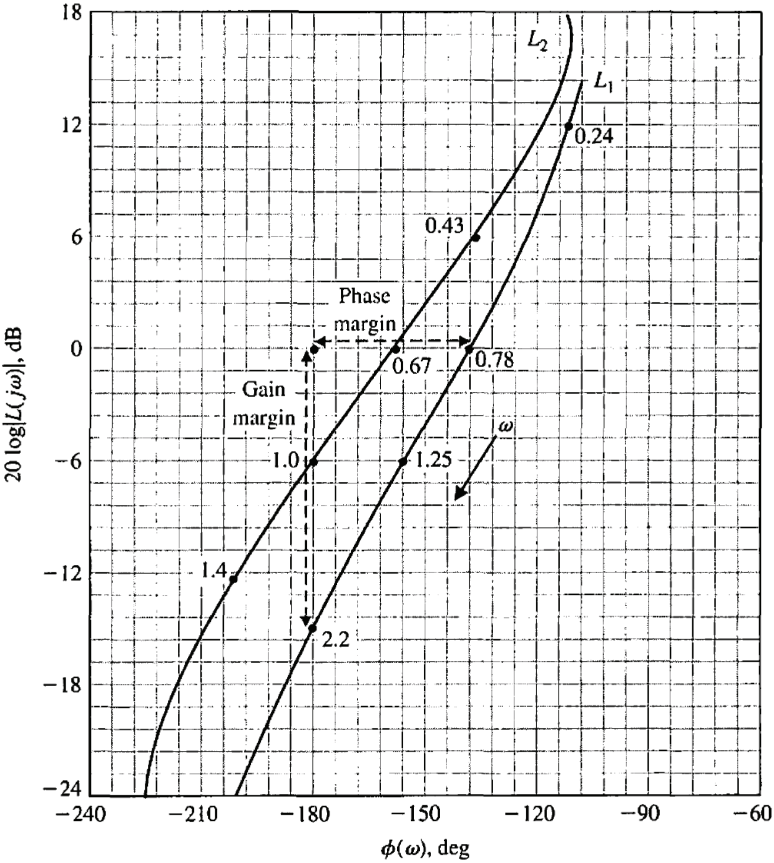


Figure 8.5 – Log-magnitude-phase curve for L_1 and L_2

The characteristic equation for this second-order system is

$$s^2 + 2\zeta\omega_n s + \omega_n^2 = 0.$$

Therefore, the closed-loop roots are

$$s = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2}.$$

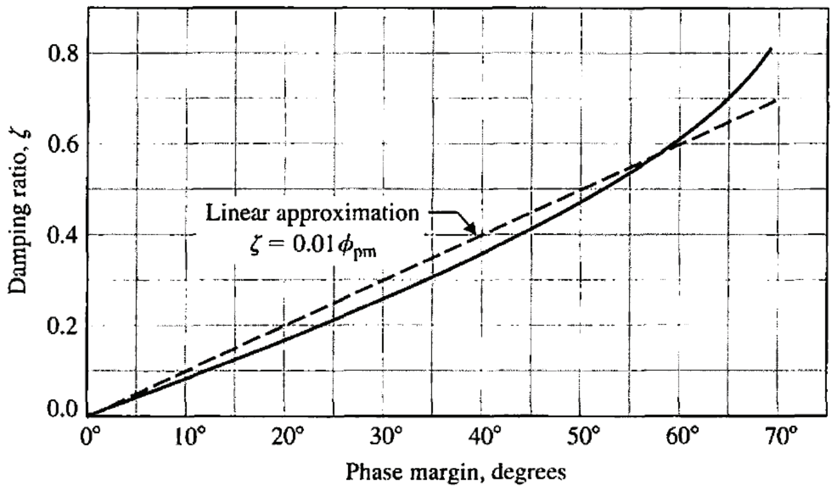


Figure 8.6 – Damping ratio versus phase margin
for a second-order system

The frequency domain form is

$$L(j\omega) = \frac{\omega_n^2}{j\omega(j\omega + 2\zeta\omega_n)}.$$

The magnitude of the frequency response is equal to 1 at a frequency ω_c thus,

$$\frac{\omega_n^2}{\omega_c\left(\omega_c^2 + 4\zeta^2\omega_n^2\right)^{1/2}} = 1.$$

Rearranging Equation (8.15), we obtain

$$\left(\omega_c^2\right)^2 + 4\zeta^2\omega_n^2\left(\omega_c^2\right) - \omega_n^4 = 0.$$

Solving for ω_c, we find that

$$\frac{\omega_c^2}{\omega_n^2} = \left(4\zeta^4 + 1\right)^{1/2} - 2\zeta^2.$$

The phase margin for this system is

$$\phi_{pm} = 180° - 90° - \tan^{-1}\frac{\omega_c}{2\zeta\omega_n} =$$

$$= 90° - \tan^{-1}\left(\frac{1}{2\zeta}\left[\left(4\zeta^4+1\right)^{1/2} - 2\zeta^2\right]^{1/2}\right) = \tan^{-1}\frac{2}{\left[\left(4+\frac{1}{\zeta^4}\right)^{\frac{1}{2}} - 2\right]^{\frac{1}{2}}}.$$

Therefore, for the system with a loop transfer function

$$L(j\omega) = \frac{1}{j\omega(j\omega+1)(0.2j\omega+1)},$$

we found that the phase margin was 43°, as shown in Figure 8.4. Thus, the damping ratio is approximately

$$\zeta \simeq 0.01\phi_{pm} = 0.43.$$

Then the percent overshoot to a step input for this system is approximately

$$P.O. = 22\,\%.$$

9. ROBUST CONTROL SYSTEMS

9.1 Introduction

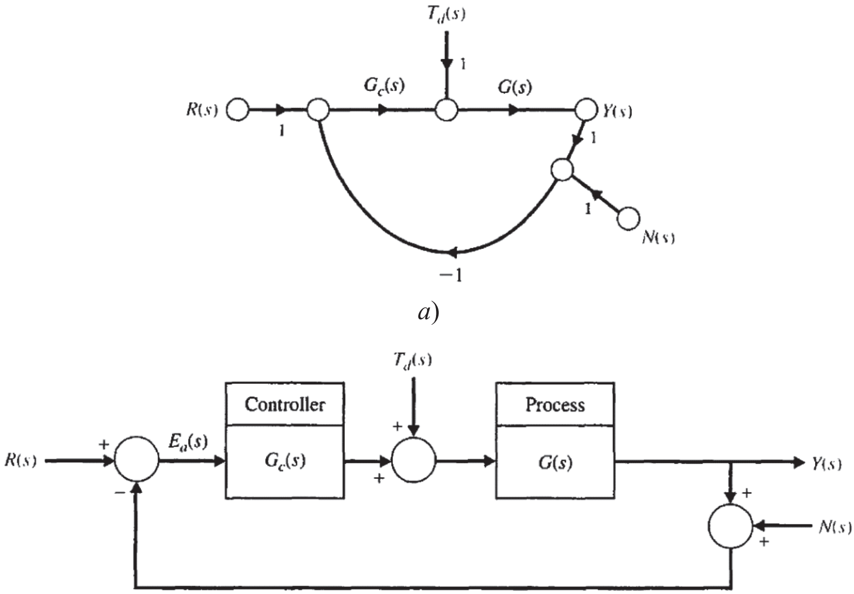
A control system designed using the methods and concepts of the preceding chapters assumes knowledge of the model of the process and controller and constant parameters. The process model will always be an inaccurate representation of the actual physical system because of

- parameter changes;
- unmodeled dynamics;
- unmodeled time delays;
- changes in equilibrium point (operating point);
- sensor noise;
- unpredicted disturbance inputs.

The goal of robust systems design is to retain assurance of system performance in spite of model inaccuracies and changes. A system is robust when the system has acceptable changes in performance due to model changes or inaccuracies.

A robust control system exhibits the desired performance despite the presence of significant process uncertainty.

A system structure that incorporates potential system uncertainties is shown in Figure 9.1. This model includes the sensor noise $N(s)$, the unpredicted disturbance input $Td(s)$, and a process $G(s)$ with potentially unmodeled dynamics or parameter changes. The unmodeled dynamics and parameter changes may be significant or very large, and for these systems the challenge is to create a design that retains the desired performance.



a)

b)

Figure 9.1 – Closed-loop control system:
(*a*) Signal flow graph, (*b*) Block diagram

9.2 Robust control systems and system sensitivity

Designing highly accurate systems in the presence of significant plant uncertainty is a classical feedback design problem. The theoretical bases for the solution of this problem date back to the works of H. S. Black and H. W. Bode in the early 1930s, when this problem was referred to as the sensitivities design problem. A significant amount of literature has been published since then regarding the design of systems subject to large process uncertainty. The designer seeks to obtain a system that performs adequately over a large range of uncertain parameters. A system is said to be robust when it is durable, hardy, and resilient.

A control system is robust when (1) it has low sensitivities, (2) it is stable over the range of parameter variations, and (3) the performance continues to meet the specifications in the presence of a set of changes in the system parameters [3, 4]. Robustness is the low sensitivity to effects that are not considered in the analysis and design phase – for example, disturbances, measurement noise, and unmodeled dynamics. The system should be able to withstand these neglected effects when performing the tasks for which it was designed.

The system sensitivity is defined as

$$S_{\alpha}^{T} = \frac{\partial T / T}{\partial \alpha / \alpha}, \tag{9.1}$$

where α is the parameter and T the transfer function of the system. The root sensitivity is defined as

$$S_{\alpha}^{r_1} = \frac{\partial r_i}{\partial \alpha / \alpha}. \tag{9.2}$$

When the zeros of $T(s)$ are independent of the parameter α, we showed that

$$S_{\alpha}^{T} = -\sum_{i=1}^{n} S_{\alpha}^{r_1} \cdot \frac{1}{s + r_1}, \tag{9.3}$$

for an nth-order system. For example, if we have a closed-loop system, as shown in Figure 9.2, where the variable parameter is α, then $T\left(s\right)=l/\left[s+\left(a+1\right)\right]$, and

$$S_{\alpha}^{T} = \frac{-\alpha}{s + \alpha + 1}. \tag{9.4}$$

This follows because $r_1 = +\left(a+1\right)$, and

$$-S_{\alpha}^{r_1} = -\alpha. \tag{9.5}$$

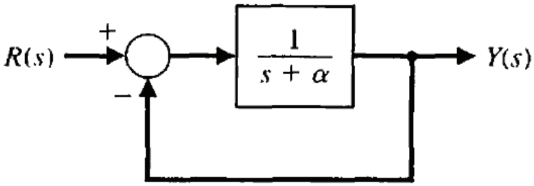


Figure 9.2 – A first-order system

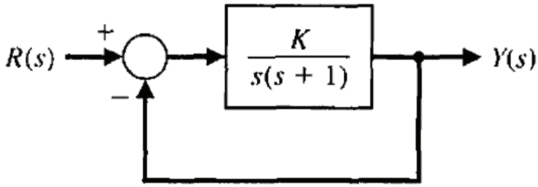


Figure 9.3 – A second-order system

Therefore,

$$S_{\alpha}^{T} = -S_{\alpha}^{r_1}\frac{1}{s+\alpha+1} = \frac{-\alpha}{s+\alpha+1}. \tag{9.6}$$

Let us examine the sensitivity of the second-order system shown in Figure 9.3. The transfer function of the closed-loop system is

$$T(s) = \frac{K}{s^2+s+K}. \tag{9.7}$$

The system sensitivity for K is

$$S(s) = S_K^T = \frac{s(s+1)}{s^2+s+K}. \tag{9.8}$$

A Bode plot of the asymptotes of $20\log\left|T(j\omega)\right|$ and $20\log\left|S(j\omega)\right|$ is shown in Figure 9.4 for $K = 1/4$ (critical damping). Note that the sensitivity is small for lower frequencies, while the transfer function primarily passes low frequencies.

Of course, the sensitivity $S(s)$ only represents robustness for small changes in gain. If K changes from 1/4 within the range

$K = 1/16$ to $K = 1$, the resulting range of step response is shown in Figure 9.4. This system, with an expected wide range of K, may not be considered adequately robust. A robust system would be expected to yield essentially the same (within an agreed-upon variation) response to a selected input.

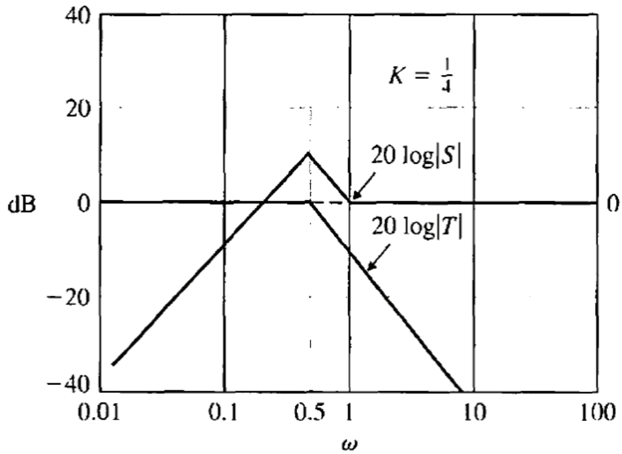


Figure 9.4 – Sensitivity and 20 log|$T(j\omega)$| for the second-order system

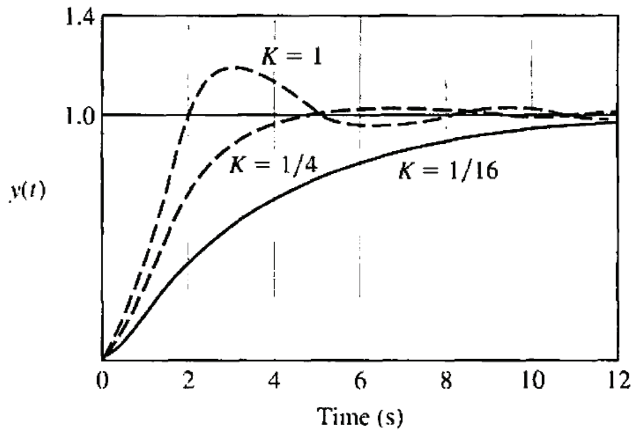


Figure 9.5 – The step response for selected gain K

9.3 Analysis of robustness

Consider the closed-loop system shown in Figure 9.1. System goals include maintaining a small tracking error [$E(s) = R(s) - Y(s)$] for an input $R(s)$ and keeping the output $Y(s)$ small for a disturbance $Td(s)$.

The **sensitivity function** is

$$S(s) = \left[1 + G_c(s)G(s)\right]^{-1},$$

and the **complementary sensitivity function** is

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

We also have the relationship

$$S(s) + C(s) = 1. \tag{9.9}$$

For physically realizable systems, the loop gain $L(s) = Gc(s)$ $G(s)$ must be small for high frequencies. This means that $S(j\omega)$ approaches 1 at high frequencies.

An **additive perturbation** characterizes the set of possible processes as follows (here we assume that $G_c(s) = 1$):

$$G_a(s) = G(s) + A(s),$$

where $G(s)$ is the nominal process, and $A(s)$ is the perturbation that is bounded in magnitude. We assume that $G_a(s)$ and $G(s)$ have the same number of poles in the right-hand s-plane (if any). Then the system stability will not change if

$$\left|A(j\omega)\right| < \left|1 + G(j\omega)\right| \quad \text{for all}\,\omega. \tag{9.10}$$

This assures stability but not dynamic performance.

A **multiplicative perturbation** is modeled as

$$G_m(s) = G(s)\left[1 + M(s)\right].$$

The perturbation is bounded in magnitude, and it is again assumed that $G_m(s)$ and $G(s)$ have the same number of poles in the right-hand s-plane. Then the system stability will not change if

$$\left|M(j\omega)\right| < \left|1 + \frac{1}{G(j\omega)}\right| \quad \text{for all } \omega.$$

Equation (12.15) is called the **robust stability criterion**. This is a test for robustness with respect to a multiplicative perturbation. This form of perturbation is often used because it satisfies the intuitive properties of (1) being small at low frequencies, where the nominal process model is usually well known, and (2) being large at high frequencies, where the nominal model is always inexact.

9.4 The design of robust pid-controlled systems

The PID controller has the transfer function

$$G_c(s) = K_p + \frac{K_1}{s} + K_D s.$$

The popularity of PID controllers can be attributed partly to their robust performance in a wide range of operating conditions and partly to their functional simplicity, which allows engineers to operate them in a simple straightforward manner. To implement such a controller, three parameters must be determined for the given process: proportional gain, integral gain, and derivative gain.

Consider the PID controller

$$G_c(s) = K_p + \frac{K_1}{s} + K_D s = \frac{K_D s^2 + K_p s + K_1}{s} =$$

$$= \frac{K_D\left(s^2 + as + b\right)}{s} = \frac{K_D\left(s + z_1\right)\left(s + z_2\right)}{s},$$

where $a = K_p/K_D$ and $b = K_I/K_D$. Therefore, a PID controller introduces a transfer function with one pole at the origin and two zeros that can be located anywhere in the left-hand s-plane.

Recall that a root locus begins at the poles and ends at the zeros. If we have a system as shown in Figure 9.6 with

$$G(s) = \frac{1}{(s+2)(s+5)},$$

and we use a K_D PID controller with complex zeros, we can plot the root locus as shown in Figure 9.7.

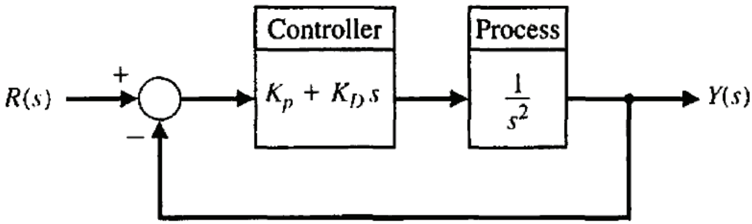


Figure 9.6 – A system with a PD controller

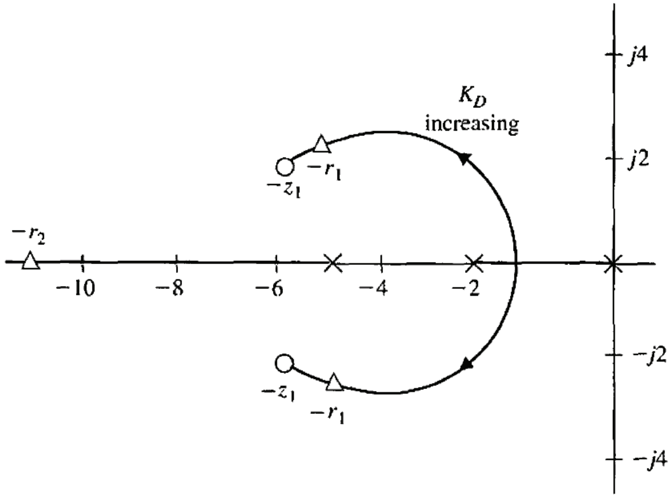


Figure 9.7 – Root locus with $-z_1 = -6 + j2$

As the gain of the controller is increased, the complex roots approach the zeros. The closed-loop transfer function is

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

$$= \frac{K_D(s+z_1)\left(s+\hat{z}_1\right)}{(s+r_2)(s+r_1)\left(s+\hat{r}_1\right)} G_p(s) =$$

$$\simeq \frac{K_D G_p(s)}{s+r_2},$$

because the zeros and the complex roots are approximately equal ($r_1 \approx r_2$) Setting $Gp(s) = 1$, we have

$$T(s) = \frac{K_D}{s+r_2} \approx \frac{K_D}{s+K_D}$$

when $K_D \gg 1$. The only limiting factor is the allowable magnitude of $U(s)$ (Figure 9.6) when K_D is large. If K_D is 100, the system has a fast response and zero steady state error. Furthermore, the effect of the disturbance is reduced significantly.

In general, we note that PID controllers are particularly useful for reducing steady-state error and improving the transient response when $G(s)$ has one or two poles (or may be approximated by a second-order process).

The selection of the three coefficients of PID controllers is basically a search problem in a three-dimensional space. Points in the search space correspond to different selections of a PID controller parameters. By choosing different points of the parameter space, we can produce, for example, different step responses for a step input. A PID controller can be determined by moving in this search space on a trial-and-error basis.

The main problem in the selection of the three coefficients is that these coefficients do not readily translate into the desired performance and robustness characteristics that the control system designer has in mind. Several rules and methods have been proposed to solve this problem.

10. DIGITAL CONTROL SYSTEMS

10.1 Introduction

The use of digital computer compensator (controller) devices has grown during the past three decades as the price and reliability of digital computers have improved dramatically. A block diagram of a single-loop digital control system is shown in Figure 10.1. The digital computer in this system configuration receives the error in digital form and performs calculations in order to provide an output in digital form. The computer may be programmed to provide an output so that the performance of the process is near or equal to the desired performance. Many computers are able to receive and manipulate several inputs, so a digital computer control system can often be a multivariable system.

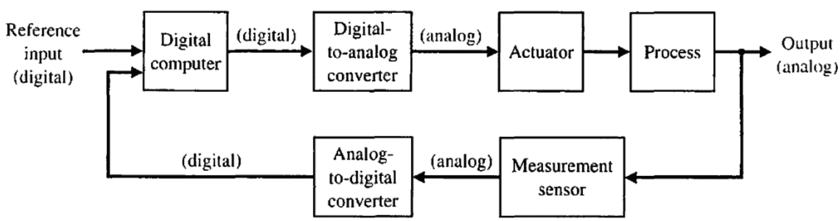


Figure 10.1 – A block diagram of a computer control system, including the signal converters. The signal is indicated as digital or analog

A digital computer receives and operates on signals in digital (numerical) form, as contrasted to continuous signals. A digital control system uses digital signals and a digital computer to control a process. The measurement data are converted from analog form to digital form by means of the analog-to-digital converter shown in Figure 10.1. After processing the inputs, the digital computer provides an output in digital form. This output is then converted to analog form by the digital-to-analog converter shown in Figure 10.1.

10.2 Digital computer control system applications

The total number of computer control systems installed in industry has grown over the past three decades. Currently, there are approximately 100 million control systems using computers, although the computer size and power may vary significantly. If we consider only computer control systems of a relatively complex nature, such as chemical process control or aircraft control, the number of computer control systems is approximately 20 million.

A digital computer consists of a central processing unit (CPU), input-output units, and a memory unit. The size and power of a computer will vary according to the size, speed, and power of the CPU, as well as the size, speed, and organization of the memory unit. Small computers, called **minicomputers**, have become increasingly common since 1980. Powerful but inexpensive computers, called **microcomputers**, which use a 16-bit word or 32-bit word, have become readily available. These systems use a microprocessor as a CPU. Therefore, the nature of the control task, the extent of the data required in memory, and the speed of calculation required will dictate the selection of the computer within the range of available computers.

The size of computers and the cost for the active logic devices used to construct them have both declined exponentially. The active components per cubic centimeter have increased so that the actual computer can be reduced in size to the point where relatively inexpensive, powerful laptop computers are providing mobile high performance computational capability to students and professionals alike, and are, in many instances, replacing traditional desktop microcomputers. The speed of computers has also increased exponentially. The transistor density (a measure of computational performance) on INTEL microprocessor integrated circuits has increased exponentially over the last 30 years.

Digital control systems are used in many applications: for machine tools, metalworking processes, chemical processes, aircraft control, and automobile traffic control, and others. Automatic computer-controlled systems are used for purposes as diverse as measuring the objective refraction of the human eye and controlling the engine spark timing or air-fuel ratio of automobile engines. The latter innovations are necessary to reduce automobile emissions and increase fuel economy.

The advantages of using digital control include: improved measurement sensitivity; the use of digitally coded signals, digital sensors and transducers, and microprocessors; reduced sensitivity to signal noise; and the capability to easily reconfigure the control algorithm in software. Improved sensitivity results from the low-energy signals required by digital sensors and devices. The use of digitally coded signals permits the wide application of digital devices and communications. Digital sensors and transducers can effectively measure, transmit, and couple signals and devices. In addition, many systems are inherently digital because they send out pulse signals. Examples of such a digital system are a radar tracking system and a space satellite.

10.3 Sampled-data systems

Computers used in control systems are interconnected to the actuator and the process by means of signal converters. The output of the computer is processed by a digital-to-analog converter.

We will assume that all the numbers that enter or leave the computer do so at the same fixed period T, called the sampling period. Thus, for example, the reference input shown in Figure 10.2 is a sequence of sample values $r(kT)$.The variables $r(kT)$, $m(kT)$, and $u(kT)$ are discrete signals in contrast to $m(t)$ and $y(t)$, which are continuous functions of time.

Sampled data (or a discrete signal) are data obtained for the system variables only at discrete intervals and are denoted as $x(kT)$.

A system where part of the system acts on sampled data is called a sampled-data system. A sampler is basically a switch that closes every T seconds for one instant of time. Consider an ideal sampler, as shown in Figure 10.3. The input is r(t), and the output is $r^*(t)$, where nT is the current sample time, and the current value of $r^*(t)$ is $r(nT)$.We then have $r^*(t) = r(nT)\,\delta(t - nT)$, where 8 is the impulse function.

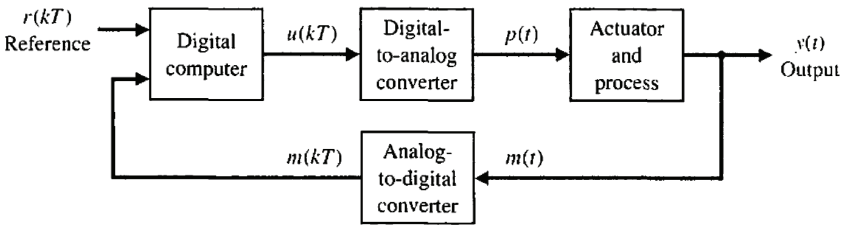


Figure 10.2 – A digital control system

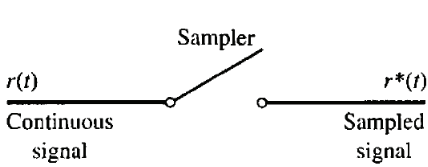
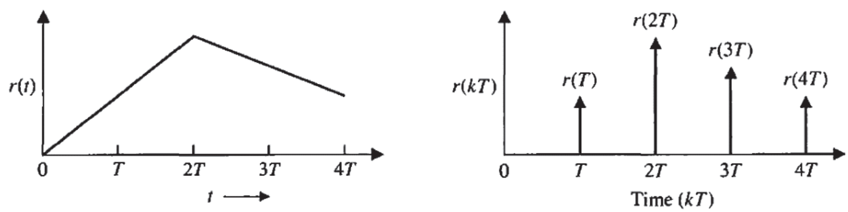


Figure 10.3 – An ideal sampler with an input $r(t)$

Let us assume that we sample a signal $r(r)$, as shown in Figure 10.3, and obtain $r^*(t)$. Then, we portray the series for $r^*(t)$ as a string of impulses starting at $t = 0$, spaced at T seconds, and of amplitude $r(kT)$. For example, consider the input signal $r(t)$ shown in Figure 10.4(a). The sampled signal is shown in Figure 10.4(b) with an impulse represented by a vertical arrow of magnitude $r(kT)$.

A digital-to-analog converter serves as a device that converts the sampled signal $r^*(t)$ to a continuous signal $p(t)$. The digital-to-analog converter can usually be represented by a zero-order hold circuit, as shown in Figure 10.5. The zero-order hold takes the value $r(kT)$ and holds it constant for $kT < t < (k + 1)$T, as shown in Figure 13.8 for $k = 0$. Thus, we use $r(kT)$ during the sampling period.

A sampler and zero-order hold can accurately follow the input signal if T is small compared to the transient changes in the signal. The response of a sampler and zero-order hold for a ramp input is shown in Figure 10.4. Finally, the response of a sampler and zero-order hold for an exponentially decaying signal is shown in Figure 10.5 for two values of the sampling period. Clearly, the output $p(t)$ will approach the input $r(t)$ as T approaches zero, meaning that we sample frequently.



a) b)

Figure 10.4 – (a) An input signal $r(t)$, (b) The sampled signal $r^*(t) = \sum_{k=0}^{x} r(kT)\delta(t - kT)$. The vertical arrow represents an impulse

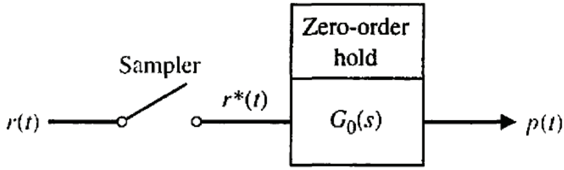


Figure 10.5 – A sampler and zero-order hold circuit

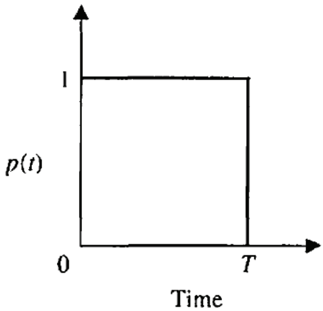


Figure 10.6 – The response of a zero-order hold to an impulse input $r(kT)$, which equals unity when $k = 0$ and equals zero when $k \neq 0$, so that $r^*(t) = (0)\delta(t)$

The impulse response of a zero-order hold is shown in Figure 10.6. The transfer function of the zero-order hold is

$$G_0(s) = \frac{1}{s} - \frac{1}{s}e^{-sT} = \frac{1-e^{-sT}}{s}.$$

The precision of the digital computer and the associated signal converters is limited (see Figure 10.2). Precision is the degree of exactness or discrimination with which a quantity is stated. The precision of the computer is limited by a finite word length. The precision of the analog-to-digital converter is limited by an ability to store its output only in digital logic composed of a finite number of binary digits. The converted signal $m(kT)$ is then said to include an amplitude quantization error. When the quantization error and the error due to a computer's finite word size are small relative to the amplitude of the signal, the system is sufficiently precise, and the precision limitations can be neglected.

10.4 The z-transform

Because the output of the ideal sampler, $r^*(t)$, is a series of impulses with values $r(kT)$, we have

$$r^*(t) = \sum_{k=0}^{\infty} r(kT)\delta(t - kT),$$

for a signal for $t > 0$. Using the Laplace transform, we have

$$\mathcal{L}\{r^*(t)\} = \sum_{k=0}^{\infty} r(kT)e^{-ksT}.$$

We now have an infinite series that involves multiples of e^{sT} and its powers. We define

$$z = e^{sT},$$

where this relationship involves a conformal mapping from the s-plane to the z-plane. We then define a new transform, called the z-transform, so that

$$Z\{r(t)\} = Z\{r^*(t)\} = \sum_{k=0}^{\infty} r(kT)z^{-k}.$$

As an example, let us determine the z-transform of the unit step function $u(t)$ (not to be confused with the control signal $u(t)$). We obtain

$$Z\{u(t)\} = \sum_{k=0}^{\infty} u(kT)z^{-k} = \sum_{k=0}^{\infty} z^{-k},$$

Since $u(kT) = 1$ for $k \geq 0$. This series can be written in closed form *as*[1]

$$U(z) = \frac{1}{1 - z^{-1}} = \frac{z}{z - 1}.$$

In general, we will define the z-transform of a function $f(t)$ as

$$Z\{f(t)\} = F(z) = \sum_{k=0}^{\infty} f(kT) z^{-k}.$$

10.5 Closed-loop feedback sampled-data systems

In this section, we consider closed-loop, sampled-data control systems. Consider the system shown in Figure 10.7(*a*). The sampled-data z-transform model of this figure with a sampled-output signal $Y(z)$ is shown in Figure 10.7(*b*). The closed-loop transfer function (using block diagram reduction) is

$$\frac{Y(z)}{R(z)} = T(z) = \frac{G(z)}{1+G(z)}.$$

Here, we assume that the $G(z)$ is the z-transform of $G(s) = GQ(s)Gp(s)$, where $GQ(s)$ is the zero-order hold, and $Gp(s)$ is the process transfer function

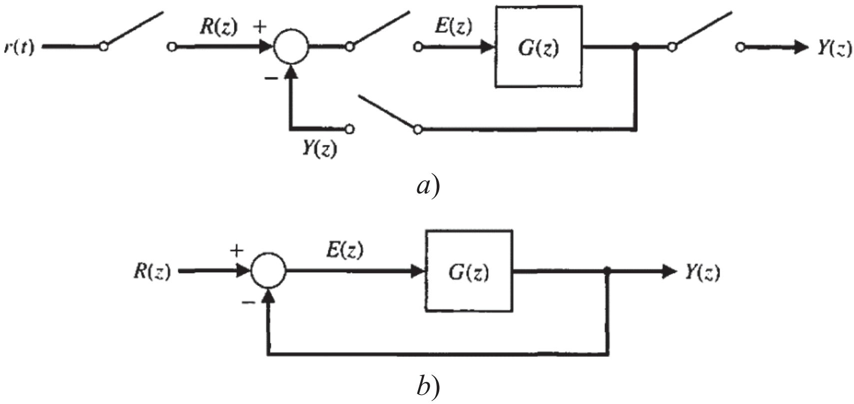


Figure 10.7 – Feedback control system with unity feedback.
$G(z)$ is the z-transform corresponding to $G(s)$, which represents the process and the zero-order hold

10.6 Implementation of digital controllers

We will consider the PID controller with an s-domain transfer function

$$\frac{U(s)}{X(s)} = G_c(s) = K_p + \frac{K_1}{s} + K_D s.$$

We can determine a digital implementation of this controller by using a discrete approximation for the derivative and integration. For the time derivative, we use the backward difference rule

$$u(kT) = \left.\frac{dx}{dt}\right|_{t=kT} = \frac{1}{T}\Big(x(kT) - x\big[(k-1)T\big]\Big).$$

The z-transform of Equation (13.55) is then

$$U(z) = \frac{1-z^{-1}}{T}X(z) = \frac{z-1}{Tz}X(z).$$

The integration of $x(i)$ can be represented by the forward-rectangular integration at $t = kT$ as

$$u(kT) = u\big[(k-1)T\big] + Tx(kT),$$

where $u(kT)$ is the output of the integrator at $t = kT$. The z-transform of Equation (10.10) is

$$U(z) = z^{-1}U(z) + TX(z)$$

and the transfer function is then

$$\frac{U(z)}{X(z)} = \frac{Tz}{z-1}.$$

Hence, the z-domain transfer function of the PID controller is

$$G_c(z) = K_p + \frac{K_1 Tz}{z-1} + K_D\frac{z-1}{Tz}.$$

The complete difference equation algorithm that provides the PID controller is obtained by adding the three terms to obtain [we use $x(kT) = x(k)$)

$$u(k) = K_p x(k) + K_I\left[u(k-1) + Tx(k)\right] + (K_D/T)\left[x(k) - x(k-1)\right] =$$

$$= \left[K_p + K_1 T + (K_D/T)\right]x(k) - K_D T x(k-1) + K_I u(k-1).$$

This equation can be implemented using a digital computer or microprocessor. Of course, we can obtain a PI or PD controller by setting an appropriate gain equal to zero.

REFERENCES

1. Dorf, Richard C. Modern control systems / Richard C. Dorf, Robert H. Bishop. – 12th ed. – Pearson Education, 2014. – 1080 p.

2. Brogan, W. L. Modern Control Theory // Brogan, W. L. – 3rd ed., Prentice Hall, Englewood Cliffs, NJ, 1990. – 400 p.

3. Kuo, B.C. Automatic Control Systems / Kuo, B.C. Prentice Hall, NJ, 2001. – 458 p.

4. Kuo, B.C. Analysis and Synthesis of Sampled-Data Control Systems / Kuo, B.C. Prentice Hall, NJ, 2003. – 564 p.

5. Tou, J.T. Modern Control Theory / Tou, J.T. New York : McGraw-Hill Book Company, 1998. – 541 p.

6. Langill, A.W. Automatic Control Systems Engineering / Langill, A.W. Prentice Hall, NJ, 1995. – 674 p.

Навчальне видання

БІЛЮК Іван Сергійович
САВЧЕНКО Олег Валерійович
ТКАЧЕНКО Сергій Олександрович
БІЛЮК Ганна Анатоліївна

ТЕОРІЯ АВТОМАТИЧНОГО КЕРУВАННЯ
THEORY OF AUTOMATIC CONTROL

Конспект лекцій

Комп'ютерне верстання *Н. М. Ковальчук*
Коректор *О. Є. Вакула*

Формат 60×84/16. Ум. друк. арк. 8,95. Вид. № 7. Зам. № 0311-57.
Видавець і виготівник Національний університет кораблебудування
імені адмірала Макарова
просп. Героїв України, 9, м. Миколаїв, 54007
E-mail : publishing@nuos.edu.ua
Свідоцтво суб'єкта видавничої справи ДК № 6402 від 19.09.2018 р.