1

Introduction to Control Systems

1-1 INTRODUCTION

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

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

Historical Review. The first significant work in automatic control was James Watt's centrifugal governor for the speed control of a steam engine in the eighteenth century. Other significant works in the early stages of development of control theory were due to Minorsky, Hazen, and Nyquist, among many others. In 1922, Minorsky worked on automatic controllers for steering ships and showed how stability could be determined from the differential equations describing the system. In 1932, Nyquist developed a relatively simple procedure for determining the stability of closed-loop systems on the basis of open-loop response to steady-state sinusoidal inputs. In 1934, Hazen, who introduced the term servomechanisms for position control systems, discussed the design of relay servomechanisms capable of closely following a changing input.

During the decade of the 1940s, frequency-response methods (especially the Bode diagram methods due to Bode) made it possible for engineers to design linear closed-loop control systems that satisfied performance requirements. From the end of the 1940s to the early 1950s, the root-locus method due to Evans was fully developed.

The frequency-response and root-locus methods, which are the core of classical control theory, lead to systems that are stable and satisfy a set of more or less arbitrary performance requirements. Such systems are, in general, acceptable but not optimal in any meaningful sense. Since the late 1950s, the emphasis in control design problems has been shifted from the design of one of many systems that work to the design of one optimal system in some meaningful sense.

As modern plants with many inputs and outputs become more and more complex, the description of a modern control system requires a large number of equations. Classical control theory, which deals only with single-input–single-output systems, becomes powerless for multiple-input–multiple-output systems. Since about 1960, because the availability of digital computers made possible time-domain analysis of complex systems, modern control theory, based on time-domain analysis and synthesis using state variables, has been developed to cope with the increased complexity of modern plants and the stringent requirements on accuracy, weight, and cost in military, space, and industrial applications.

During the years from 1960 to 1980, optimal control of both deterministic and stochastic systems, as well as adaptive and learning control of complex systems, were fully investigated. From 1980 to the present, developments in modern control theory centered around robust control, H_∞ control, and associated topics.

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

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

Controlled Variable and Manipulated Variable. The *controlled* variable is the quantity or condition that is measured and controlled. The *manipulated* variable is the quantity or condition that is varied by the controller so as to affect the value of the controlled variable. Normally, the controlled variable is the output of the system. *Control* means measuring the value of the controlled variable of the system and applying the manipulated variable to the system to correct or limit deviation of the measured value from a desired value.

In studying control engineering, we need to define additional terms that are necessary to describe control systems.

Plants. A plant may be a piece of equipment, perhaps just a set of machine parts functioning together, the purpose of which is to perform a particular operation. In this book, we shall call any physical object to be controlled (such as a mechanical device, a heating furnace, a chemical reactor, or a spacecraft) a plant.

Processes. The *Merriam–Webster Dictionary* defines a process to be a natural, progressively continuing operation or development marked by a series of gradual changes that succeed one another in a relatively fixed way and lead toward a particular result or end; or an artificial or voluntary, progressively continuing operation that consists of a series of controlled actions or movements systematically directed toward a particular result or end. In this book we shall call any operation to be controlled a *process*. Examples are chemical, economic, and biological processes.

Systems. A system is a combination of components that act together and perform a certain objective. A system is not limited to physical ones. The concept of the system can be applied to abstract, dynamic phenomena such as those encountered in economics. The word system should, therefore, be interpreted to imply physical, biological, economic, and the like, systems.

Disturbances. A disturbance is a signal that tends to adversely affect the value of the output of a system. If a disturbance is generated within the system, it is called *internal*, while an *external* disturbance is generated outside the system and is an input.

Feedback Control. Feedback control refers to an operation that, in the presence of disturbances, tends to reduce the difference between the output of a system and some reference input and does so on the basis of this difference. Here only unpredictable disturbances are so specified, since predictable or known disturbances can always be compensated for within the system.

1–2 EXAMPLES OF CONTROL SYSTEMS

In this section we shall present several examples of control systems.

Speed Control System. The basic principle of a Watt's speed governor for an engine is illustrated in the schematic diagram of Figure 1–1. The amount of fuel admitted

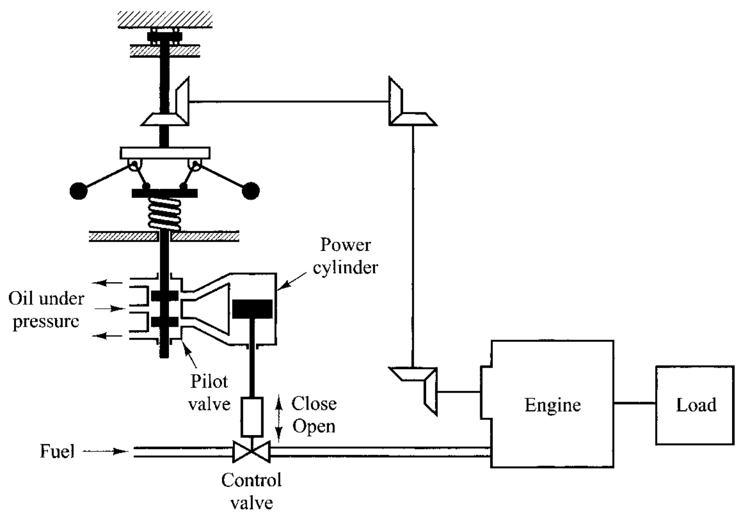

Figure 1–1
Speed control system.

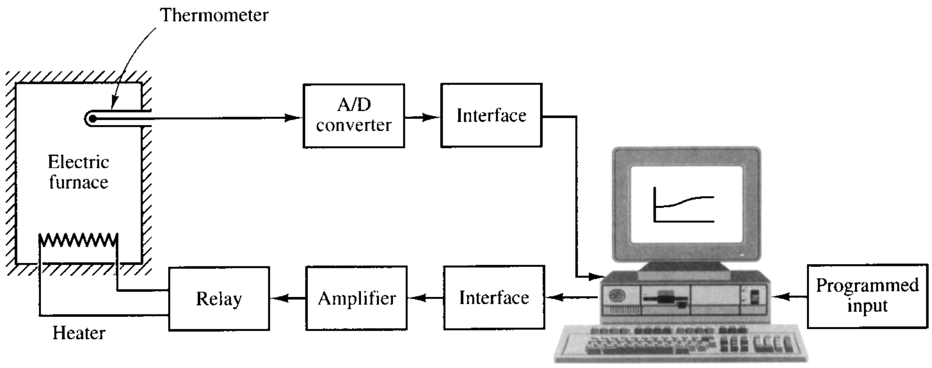

Figure 1–2
Temperature control system.

to the engine is adjusted according to the difference between the desired and the actual engine speeds.

The sequence of actions may be stated as follows: The speed governor is adjusted such that, at the desired speed, no pressured oil will flow into either side of the power cylinder. If the actual speed drops below the desired value due to disturbance, then the decrease in the centrifugal force of the speed governor causes the control valve to move downward, supplying more fuel, and the speed of the engine increases until the desired value is reached. On the other hand, if the speed of the engine increases above the desired value, then the increase in the centrifugal force of the governor causes the control valve to move upward. This decreases the supply of fuel, and the speed of the engine decreases until the desired value is reached.

In this speed control system, the plant (controlled system) is the engine and the controlled variable is the speed of the engine. The difference between the desired speed and the actual speed is the error signal. The control signal (the amount of fuel) to be applied to the plant (engine) is the actuating signal. The external input to disturb the controlled variable is the disturbance. An unexpected change in the load is a disturbance.

Temperature Control System. Figure 1–2 shows a schematic diagram of temperature control of an electric furnace. The temperature in the electric furnace is measured by a thermometer, which is an analog device. The analog temperature is converted to a digital temperature by an A/D converter. The digital temperature is fed to a controller through an interface. This digital temperature is compared with the programmed input temperature, and if there is any discrepancy (error), the controller sends out a signal to the heater, through an interface, amplifier, and relay, to bring the furnace temperature to a desired value.

EXAMPLE 1–1 Consider the temperature control of the passenger compartment of a car. The desired temperature (converted to a voltage) is the input to the controller. The actual temperature of the passenger compartment must be converted to a voltage through a sensor and fed back to the controller for comparison with the input.

Figure 1–3 is a functional block diagram of temperature control of the passenger compartment of a car. Note that the ambient temperature and radiation heat transfer from the sun, which are not constant while the car is driven, act as disturbances.

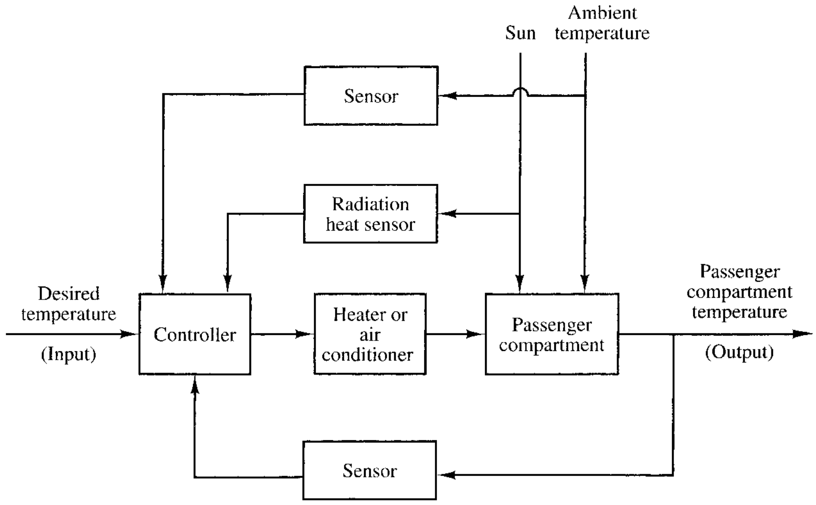

Figure 1–3 Temperature control of passenger compartment of a car.

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

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

Business Systems. A business system may consist of many groups. Each task assigned to a group will represent a dynamic element of the system. Feedback methods of reporting the accomplishments of each group must be established in such a system for proper operation. The cross-coupling between functional groups must be made a minimum in order to reduce undesirable delay times in the system. The smaller this cross-coupling, the smoother the flow of work signals and materials will be.

A business system is a closed-loop system. A good design will reduce the managerial control required. Note that disturbances in this system are the lack of personnel or materials, interruption of communication, human errors, and the like.

The establishment of a well-founded estimating system based on statistics is mandatory to proper management. Note that it is a well-known fact that the performance of such a system can be improved by the use of lead time, or *anticipation*.

To apply control theory to improve the performance of such a system, we must represent the dynamic characteristic of the component groups of the system by a relatively simple set of equations.

Although it is certainly a difficult problem to derive mathematical representations of the component groups, the application of optimization techniques to business systems significantly improves the performance of the business system.

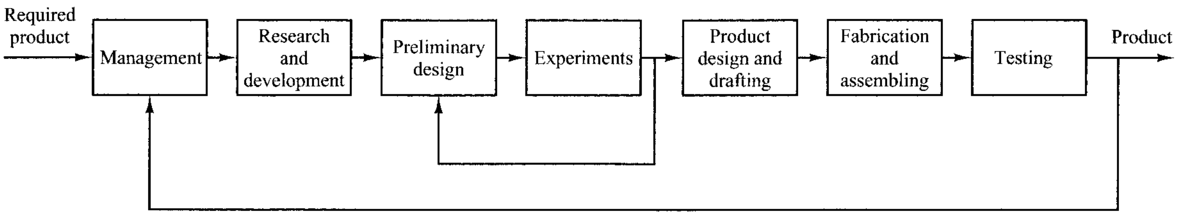

Figure 1–4
Block diagram of an engineering organizational system.

EXAMPLE 1–2 An engineering organizational system is composed of major groups such as management, research and development, preliminary design, experiments, product design and drafting, fabrication and assembling, and testing. These groups are interconnected to make up the whole operation.

Such a system may be analyzed by reducing it to the most elementary set of components necessary that can provide the analytical detail required and by representing the dynamic characteristics of each component by a set of simple equations. (The dynamic performance of such a system may be determined from the relation between progressive accomplishment and time.) Draw a functional block diagram showing an engineering organizational system.

A functional block diagram can be drawn by using blocks to represent the functional activities and interconnecting signal lines to represent the information or product output of the system operation. A possible block diagram is shown in Figure 1–4.

1–3 CLOSED-LOOP CONTROL VERSUS OPEN-LOOP CONTROL

Feedback Control Systems. A system that maintains a prescribed relationship between the output and the reference input by comparing them and using the difference as a means of control is called a *feedback control system*. An example would be a room-temperature control system. By measuring the actual room temperature and comparing it with the reference temperature (desired temperature), the thermostat turns the heating or cooling equipment on or off in such a way as to ensure that the room temperature remains at a comfortable level regardless of outside conditions.

Feedback control systems are not limited to engineering but can be found in various nonengineering fields as well. The human body, for instance, is a highly advanced feedback control system. Both body temperature and blood pressure are kept constant by means of physiological feedback. In fact, feedback performs a vital function: It makes the human body relatively insensitive to external disturbances, thus enabling it to function properly in a changing environment.

Closed-Loop Control Systems. Feedback control systems are often referred to as *closed-loop control* systems. In practice, the terms feedback control and closed-loop control are used interchangeably. In a closed-loop control system the actuating error signal, which is the difference between the input signal and the feedback signal (which may be the output signal itself or a function of the output signal and its derivatives and/or integrals), is fed to the controller so as to reduce the error and bring the output of the system to a desired value. The term closed-loop control always implies the use of feedback control action in order to reduce system error.

Open-Loop Control Systems. Those systems in which the output has no effect on the control action are called *open-loop control systems*. In other words, in an open-loop control system the output is neither measured nor fed back for comparison with the input. One practical example is a washing machine. Soaking, washing, and rinsing in the washer operate on a time basis. The machine does not measure the output signal, that is, the cleanliness of the clothes.

In any open-loop control system the output is not compared with the reference input. Thus, to each reference input there corresponds a fixed operating condition; as a result, the accuracy of the system depends on calibration. In the presence of disturbances, an open-loop control system will not perform the desired task. Open-loop control can be used, in practice, only if the relationship between the input and output is known and if there are neither internal nor external disturbances. Clearly, such systems are not feedback control systems. Note that any control system that operates on a time basis is open loop. For instance, traffic control by means of signals operated on a time basis is another example of open-loop control.

Closed-Loop versus Open-Loop Control Systems. An advantage of the closed-loop control system is the fact that the use of feedback makes the system response relatively insensitive to external disturbances and internal variations in system parameters. It is thus possible to use relatively inaccurate and inexpensive components to obtain the accurate control of a given plant, whereas doing so is impossible in the open-loop case.

From the point of view of stability, the open-loop control system is easier to build because system stability is not a major problem. On the other hand, stability is a major problem in the closed-loop control system, which may tend to overcorrect errors and thereby can cause oscillations of constant or changing amplitude.

It should be emphasized that for systems in which the inputs are known ahead of time and in which there are no disturbances it is advisable to use open-loop control. Closed-loop control systems have advantages only when unpredictable disturbances and/or unpredictable variations in system components are present. Note that the output power rating partially determines the cost, weight, and size of a control system. The number of components used in a closed-loop control system is more than that for a corresponding open-loop control system. Thus, the closed-loop control system is generally higher in cost and power. To decrease the required power of a system, open-loop control may be used where applicable. A proper combination of open-loop and closed-loop controls is usually less expensive and will give satisfactory overall system performance.

EXAMPLE 1–3 Most analyses and designs of control systems presented in this book are concerned with closed-loop control systems. Under certain circumstances (such as where no disturbances exist or the output is hard to measure) open-loop control systems may be desired. Therefore, it is worthwhile to summarize the advantages and disadvantages of using open-loop control systems.

The major advantages of open-loop control systems are as follows:

1. Simple construction and ease of maintenance.
2. Less expensive than a corresponding closed-loop system.
3. There is no stability problem.
4. Convenient when output is hard to measure or measuring the output precisely is economically not feasible. (For example, in the washer system, it would be quite expensive to provide a device to measure the quality of the washer's output, cleanliness of the clothes.)

The major disadvantages of open-loop control systems are as follows:

1. Disturbances and changes in calibration cause errors, and the output may be different from what is desired.
2. To maintain the required quality in the output, recalibration is necessary from time to time.

1–4 OUTLINE OF THE BOOK

We briefly describe here the organization and contents of the book.

Chapter 1 has given introductory materials on control systems. Chapter 2 presents basic Laplace transform theory necessary for understanding the control theory presented in this book. Chapter 3 deals with mathematical modeling of dynamic systems in terms of transfer functions and state-space equations. It discusses mathematical modeling of mechanical systems and electrical and electronic systems. This chapter also includes the signal flow graphs and linearization of nonlinear mathematical models. Chapter 4 treats mathematical modeling of liquid-level systems, pneumatic systems, hydraulic systems, and thermal systems. Chapter 5 treats transient-response analyses of first- and second-order systems as well as higher-order systems. Detailed discussions of transient-response analysis with MATLAB are presented. Routh's stability criterion and steady-state errors in unity-feedback control systems are also presented in this chapter.

Chapter 6 gives a root-locus analysis of control systems. General rules for constructing root loci are presented. Detailed discussions for plotting root loci with MATLAB are included. Chapter 7 deals with the design of control systems via the root-locus method. Specifically, root-locus approaches to the design of lead compensators, lag compensators, and lag–lead compensators are discussed in detail. Chapter 8 gives the frequency-response analysis of control systems. Bode diagrams, polar plots, Nyquist stability criterion, and closed-loop frequency response are discussed. Chapter 9 treats control systems design via the frequency-response approach. Here Bode diagrams are used to design lead compensators, lag compensators, and lag–lead compensators. Chapter 10 discusses the basic and modified PID controls. In this chapter two-degrees-of-freedom control systems are introduced. We design high-performance control systems using two-degrees-of-freedom configuration. MATLAB is extensively used in the design of such systems.

Chapter 11 presents basic materials for the state-space analysis of control systems. The solution of the time-invariant state equation is derived and concepts of controllability and observability are discussed. Chapter 12 treats the design of control systems in state space. This chapter begins with the pole-placement problems, followed by the design of state observers, and the design of regulator systems with observers and control systems with observers. Finally, quadratic optimal control is discussed.

2

The Laplace Transform*

2–1 INTRODUCTION

The Laplace transform method is an operational method that can be used advantageously for solving linear differential equations. By use of Laplace transforms, we can convert many common functions, such as sinusoidal functions, damped sinusoidal functions, and exponential functions, into algebraic functions of a complex variable s. Operations such as differentiation and integration can be replaced by algebraic operations in the complex plane. Thus, a linear differential equation can be transformed into an algebraic equation in a complex variable s. If the algebraic equation in s is solved for the dependent variable, then the solution of the differential equation (the inverse Laplace transform of the dependent variable) may be found by use of a Laplace transform table or by use of the partial-fraction expansion technique, which is presented in Section 2–5 and 2–6.

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

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

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

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

2–2 REVIEW OF COMPLEX VARIABLES AND COMPLEX FUNCTIONS

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

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

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

where σ is the real part and ω is the imaginary part.

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

$$G(s) = G_x + jG_y$$

where G_x and G_y are real quantities. The magnitude of $G(s)$ is $\sqrt{G_x^2 + G_y^2}$, and the angle θ of $G(s)$ is $\tan^{-1}(G_y/G_x)$. The angle is measured counterclockwise from the positive real axis. The complex conjugate of $G(s)$ is $\bar{G}(s) = G_x - jG_y$.

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

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

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

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

For a particular path $\Delta s = \Delta\sigma$ (which means that the path is parallel to the real axis).

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