1.4-1 Unit Step Function $u(t)$

In much of our discussion, the signals begin at $t = 0$ (causal signals). Such signals can be conveniently described in terms of unit step function $u(t)$ shown in Fig. 1.14a. This function is defined by

$$u(t) = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{1.20}$$

If we want a signal to start at $t = 0$ (so that it has a value of zero for $t < 0$), we need only multiply the signal by $u(t)$. For instance, the signal e^{-at} represents an everlasting exponential that starts at $t = -\infty$. The causal form of this exponential (Fig. 1.14b) can be described as $e^{-at}u(t)$.

The unit step function also proves very useful in specifying a function with different mathematical descriptions over different intervals. Examples of such functions appear in Fig. 1.7. These functions have different mathematical descriptions over different segments of time as seen from Eqs. (1.14), (1.15a), and (1.15b). Such a description often proves clumsy and inconvenient in mathematical treatment. We can use the unit step function to describe such functions by a single expression that is valid for all t.

Consider, for example, the rectangular pulse depicted in Fig. 1.15a. We can express such a pulse in terms of familiar step functions by observing that the pulse $x(t)$ can be expressed as the sum of the two delayed unit step functions as shown in Fig. 1.15b. The unit step function $u(t)$

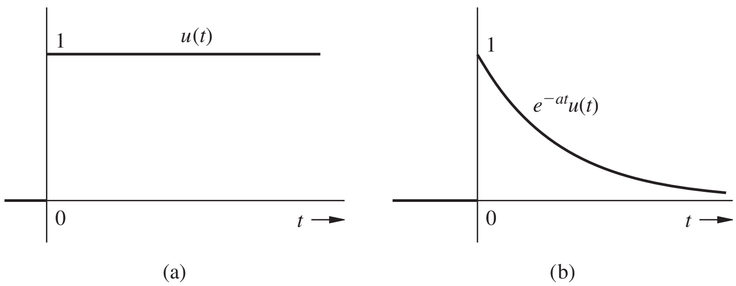


Figure 1.14 **(a)** Unit step function $u(t)$. **(b)** Exponential $e^{-at}u(t)$.

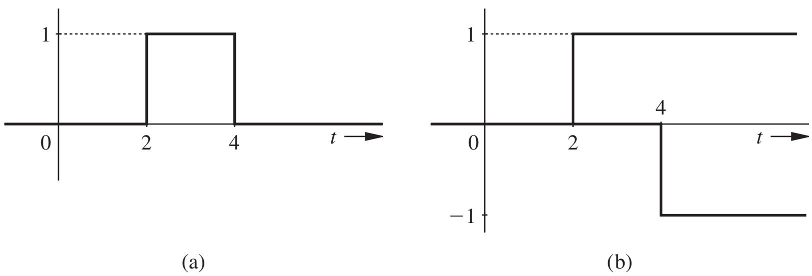


Figure 1.15 Representation of a rectangular pulse by step functions.

delayed by T seconds is $u(t - T)$. From Fig. 1.15b, it is clear that

$$x(t) = u(t - 2) - u(t - 4)$$

EXAMPLE 1.6

Describe the signal in Fig. 1.16a.

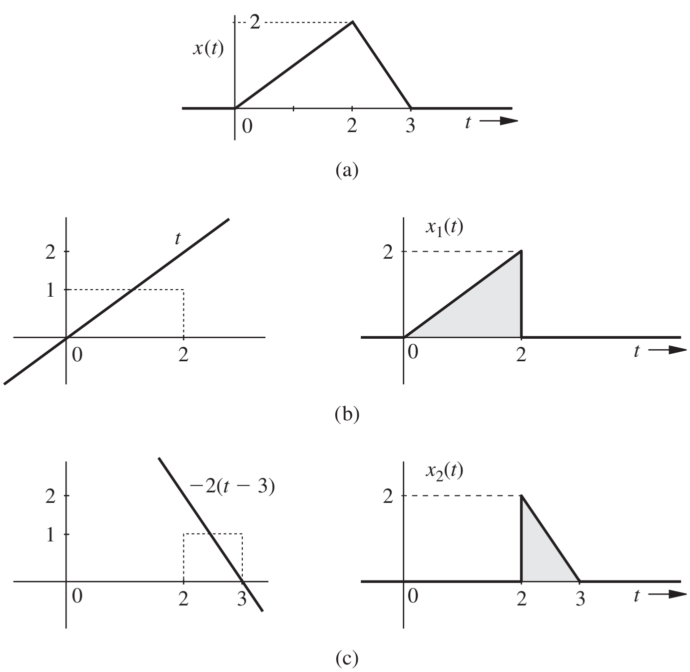


Figure 1.16 Representation of a signal defined interval by interval.

The signal illustrated in Fig. 1.16a can be conveniently handled by breaking it up into the two components $x_1(t)$ and $x_2(t)$, depicted in Fig. 1.16b and 1.16c, respectively. Here, $x_1(t)$ can be obtained by multiplying the ramp t by the gate pulse $u(t) - u(t - 2)$, as shown in Fig. 1.16b. Therefore

$$x_1(t) = t[u(t) - u(t - 2)]$$

The signal $x_2(t)$ can be obtained by multiplying another ramp by the gate pulse illustrated in Fig. 1.16c. This ramp has a slope -2; hence it can be described by $-2t + c$. Now, because the ramp has a zero value at $t = 3$, the constant $c = 6$, and the ramp can be described by $-2(t-3)$. Also, the gate pulse in Fig. 1.16c is $u(t-2)-u(t-3)$. Therefore

$$x_2(t) = -2(t-3)[u(t-2)-u(t-3)]$$

and

$$\begin{aligned} x(t) &= x_1(t) + x_2(t) \\ &= t\,[u(t)-u(t-2)] - 2(t-3)\,[u(t-2)-u(t-3)] \\ &= tu(t) - 3(t-2)u(t-2) + 2(t-3)u(t-3) \end{aligned}$$

EXAMPLE 1.7

Describe the signal in Fig. 1.7a by a single expression valid for all t.

Over the interval from -1.5 to 0, the signal can be described by a constant 2, and over the interval from 0 to 3, it can be described by $2\,e^{-t/2}$. Therefore

$$\begin{aligned} x(t) &= \underbrace{2[u(t+1.5)-u(t)]}_{x_1(t)} + \underbrace{2e^{-t/2}[u(t)-u(t-3)]}_{x_2(t)} \\ &= 2u(t+1.5) - 2(1-e^{-t/2})u(t) - 2e^{-t/2}u(t-3) \end{aligned}$$

Compare this expression with the expression for the same function found in Eq. (1.14).

EXERCISE E1.7

Show that the signals depicted in Fig. 1.17a and 1.17b can be described as $u(-t)$ and $e^{-at}u(-t)$, respectively.

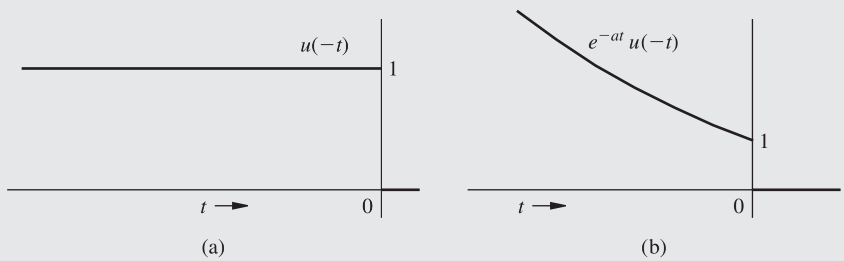


Figure 1.17

EXERCISE E1.8

Show that the signal shown in Fig. 1.18 can be described as

$$x(t) = (t-1)u(t-1) - (t-2)u(t-2) - u(t-4)$$

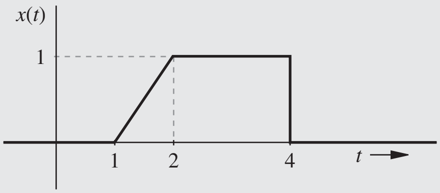


Figure 1.18

1.4-2 The Unit Impulse Function $\delta(t)$

The unit impulse function $\delta(t)$ is one of the most important functions in the study of signals and systems. This function was first defined by P. A. M. Dirac as

$$\delta(t) = 0 \qquad t \neq 0$$
$$\int_{-\infty}^{\infty} \delta(t)\,dt = 1 \tag{1.21}$$

We can visualize an impulse as a tall, narrow, rectangular pulse of unit area, as illustrated in Fig. 1.19b. The width of this rectangular pulse is a very small value $\epsilon \to 0$. Consequently, its height is a very large value $1/\epsilon \to \infty$. The unit impulse therefore can be regarded as a rectangular pulse with a width that has become infinitesimally small, a height that has become infinitely large, and an overall area that has been maintained at unity. Thus $\delta(t) = 0$ everywhere except at $t = 0$, where it is undefined. For this reason a unit impulse is represented by the spearlike symbol in Fig. 1.19a.

Other pulses, such as the exponential, triangular, or Gaussian types, may also be used in impulse approximation. The important feature of the unit impulse function is not its shape but the fact that its effective duration (pulse width) approaches zero while its area remains at unity. For example, the exponential pulse $\alpha e^{-\alpha t}u(t)$ in Fig. 1.20a becomes taller and narrower as α increases. In the limit as $\alpha \to \infty$, the pulse height $\to \infty$, and its width or duration $\to 0$. Yet,

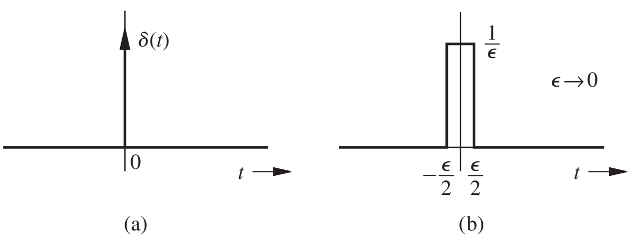


Figure 1.19 A unit impulse and its approximation.

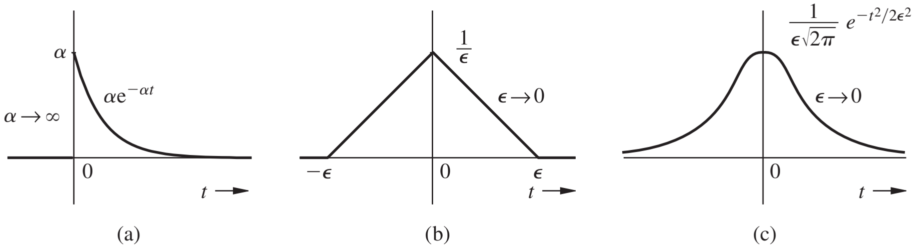


Figure 1.20 Other possible approximations to a unit impulse.

the area under the pulse is unity regardless of the value of α because

$$\int_0^{\infty} \alpha e^{-\alpha t}\, dt = 1 \tag{1.22}$$

The pulses in Fig. 1.20b and 1.20c behave in a similar fashion. Clearly, the exact impulse function cannot be generated in practice; it can only be approached.

From Eq. (1.21), it follows that the function $k\delta(t) = 0$ for all $t \neq 0$, and its area is k. Thus, $k\delta(t)$ is an impulse function whose area is k (in contrast to the unit impulse function, whose area is 1).

Multiplication of a Function by an Impulse

Let us now consider what happens when we multiply the unit impulse $\delta(t)$ by a function $\phi(t)$ that is known to be continuous at $t = 0$. Since the impulse has nonzero value only at $t = 0$, and the value of $\phi(t)$ at $t = 0$ is $\phi(0)$, we obtain

$$\phi(t)\delta(t) = \phi(0)\delta(t) \tag{1.23a}$$

Thus, multiplication of a continuous-time function $\phi(t)$ with an unit impulse located at $t = 0$ results in an impulse, which is located at $t = 0$ and has strength $\phi(0)$ (the value of $\phi(t)$ at the location of the impulse). Use of exactly the same argument leads to the generalization of this result, stating that provided $\phi(t)$ is continuous at $t = T$, $\phi(t)$ multiplied by an impulse $\delta(t - T)$ (impulse located at $t = T$) results in an impulse located at $t = T$ and having strength $\phi(T)$ [the value of $\phi(t)$ at the location of the impulse].

$$\phi(t)\delta(t - T) = \phi(T)\delta(t - T) \tag{1.23b}$$

Sampling Property of the Unit Impulse Function

From Eq. (1.23a) it follows that

$$\begin{aligned} \int_{-\infty}^{\infty} \phi(t)\delta(t)\, dt &= \phi(0) \int_{-\infty}^{\infty} \delta(t)\, dt \\ &= \phi(0) \end{aligned} \tag{1.24a}$$

provided $\phi(t)$ is continuous at $t = 0$. This result means that *the area under the product of a function with an impulse $\delta(t)$ is equal to the value of that function at the instant at which the unit impulse is located.* This property is very important and useful and is known as the *sampling* or *sifting property* of the unit impulse.

From Eq. (1.23b) it follows that

$$\int_{-\infty}^{\infty} \phi(t)\delta(t-T)\,dt = \phi(T) \tag{1.24b}$$

Equation (1.24b) is just another form of sampling or sifting property. In the case of Eq. (1.24b), the impulse $\delta(t-T)$ is located at $t = T$. Therefore, the area under $\phi(t)\delta(t-T)$ is $\phi(T)$, the value of $\phi(t)$ at the instant at which the impulse is located (at $t = T$). In these derivations we have assumed that the function is continuous at the instant where the impulse is located.

Unit Impulse as a Generalized Function

The definition of the unit impulse function given in Eq. (1.21) is not mathematically rigorous, which leads to serious difficulties. First, the impulse function does not define a unique function: for example, it can be shown that $\delta(t) + \dot{\delta}(t)$ also satisfies Eq. (1.21).[1] Moreover, $\delta(t)$ is not even a true function in the ordinary sense. An ordinary function is specified by its values for all time t. The impulse function is zero everywhere except at $t = 0$, and at this, the only interesting part of its range, it is undefined. These difficulties are resolved by defining the impulse as a generalized function rather than an ordinary function. A *generalized function* is defined by its effect on other functions instead of by its value at every instant of time.

In this approach the impulse function is defined by the sampling property [Eqs. (1.24)]. We say nothing about what the impulse function is or what it looks like. Instead, the impulse function is defined in terms of its effect on a test function $\phi(t)$. We define a unit impulse as a function for which the area under its product with a function $\phi(t)$ is equal to the value of the function $\phi(t)$ at the instant at which the impulse is located. It is assumed that $\phi(t)$ is continuous at the location of the impulse. Therefore, either Eq. (1.24a) or (1.24b) can serve as a definition of the impulse function in this approach. Recall that the sampling property [Eqs. (1.24)] is the consequence of the classical (Dirac) definition of impulse in Eq. (1.21). In contrast, *the sampling property [Eqs. (1.24)] defines the impulse function in the generalized function approach.*

We now present an interesting application of the generalized function definition of an impulse. Because the unit step function $u(t)$ is discontinuous at $t = 0$, its derivative du/dt does not exist at $t = 0$ in the ordinary sense. We now show that this derivative *does* exist in the generalized sense, and it is, in fact, $\delta(t)$. As a proof, let us evaluate the integral of $(du/dt)\phi(t)$, using integration by parts:

$$\int_{-\infty}^{\infty} \frac{du}{dt}\phi(t)\,dt = u(t)\phi(t)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} u(t)\dot{\phi}(t)\,dt \tag{1.25}$$

$$= \phi(\infty) - 0 - \int_{0}^{\infty} \dot{\phi}(t)\,dt$$

$$= \phi(\infty) - \phi(t)|_0^{\infty}$$

$$= \phi(0) \tag{1.26}$$

This result shows that du/dt satisfies the sampling property of $\delta(t)$. Therefore it is an impulse $\delta(t)$ in the generalized sense—that is,

$$\frac{du}{dt} = \delta(t) \tag{1.27}$$

Consequently

$$\int_{-\infty}^{t} \delta(\tau)\, d\tau = u(t) \tag{1.28}$$

These results can also be obtained graphically from Fig. 1.19b. We observe that the area from $-\infty$ to t under the limiting form of $\delta(t)$ in Fig. 1.19b is zero if $t < -\epsilon/2$ and unity if $t \geq \epsilon/2$ with $\epsilon \to 0$. Consequently

$$\begin{aligned} \int_{-\infty}^{t} \delta(\tau)\, d\tau &= \begin{cases} 0 & t < 0 \\ 1 & t \geq 0 \end{cases} \\ &= u(t) \end{aligned} \tag{1.29}$$

This result shows that the unit step function can be obtained by integrating the unit impulse function. Similarly the unit ramp function $x(t) = tu(t)$ can be obtained by integrating the unit step function. We may continue with unit parabolic function $t^2/2$ obtained by integrating the unit ramp, and so on. On the other side, we have derivatives of impulse function, which can be defined as generalized functions (see Prob. 1-28). All these functions, derived from the unit impulse function (successive derivatives and integrals) are called *singularity functions*.†

EXERCISE E1.9

Show that

(a) $(t^3 + 3)\delta(t) = 3\delta(t)$

(b) $\left[\sin\left(t^2 - \frac{\pi}{2}\right)\right]\delta(t) = -\delta(t)$

(c) $e^{-2t}\delta(t) = \delta(t)$

(d) $\frac{\omega^2 + 1}{\omega^2 + 9}\delta(\omega - 1) = \frac{1}{5}\delta(\omega - 1)$

†Singularity functions were defined by late Prof. S. J. Mason as follows. A singularity is a point at which a function does not possess a derivative. Each of the singularity functions (or if not the function itself, then the function differentiated a finite number of times) has a singular point at the origin and is zero elsewhere.[2]

EXERCISE E1.10

Show that

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

(b) $\int_{-\infty}^{\infty} \delta(t-2)\cos\left(\frac{\pi t}{4}\right)dt = 0$

(c) $\int_{-\infty}^{\infty} e^{-2(x-t)}\delta(2-t)\,dt = e^{-2(x-2)}$

1.4-3 The Exponential Function e^{st}

Another important function in the area of signals and systems is the exponential signal e^{st}, where s is complex in general, given by

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

Therefore

$$e^{st} = e^{(\sigma+j\omega)t} = e^{\sigma t}e^{j\omega t} = e^{\sigma t}(\cos\,\omega t + j\,\sin\,\omega t) \tag{1.30a}$$

Since $s^* = \sigma - j\omega$ (the conjugate of s), then

$$e^{s^*t} = e^{\sigma - j\omega} = e^{\sigma t}e^{-j\omega t} = e^{\sigma t}(\cos\,\omega t - j\,\sin\,\omega t) \tag{1.30b}$$

and

$$e^{\sigma t}\cos\,\omega t = \tfrac{1}{2}(e^{st} + e^{s^*t}) \tag{1.30c}$$

Comparison of this equation with Euler's formula shows that e^{st} is a generalization of the function $e^{j\omega t}$, where the frequency variable $j\omega$ is generalized to a complex variable $s = \sigma + j\omega$. For this reason we designate the variable s as the *complex frequency*. From Eqs. (1.30) it follows that the function e^{st} encompasses a large class of functions. The following functions are either special cases of or can be expressed in terms of e^{st}:

1. A constant $k = ke^{0t}$ $\quad (s = 0)$
2. A monotonic exponential $e^{\sigma t}$ $\quad (\omega = 0,\ s = \sigma)$
3. A sinusoid $\cos\,\omega t$ $\quad (\sigma = 0,\ s = \pm j\omega)$
4. An exponentially varying sinusoid $e^{\sigma t}\cos\,\omega t$ $\quad (s = \sigma \pm j\omega)$

These functions are illustrated in Fig. 1.21.

The complex frequency s can be conveniently represented on a *complex frequency plane* (s plane) as depicted in Fig. 1.22. The horizontal axis is the real axis (σ axis), and the vertical axis is the imaginary axis ($j\omega$ axis). The absolute value of the imaginary part of s is $|\omega|$ (the *radian* frequency), which indicates the frequency of oscillation of e^{st}; the real part σ (the *neper* frequency) gives information about the rate of increase or decrease of the amplitude of e^{st}. For signals whose complex frequencies lie on the real axis (σ axis, where $\omega = 0$), the frequency

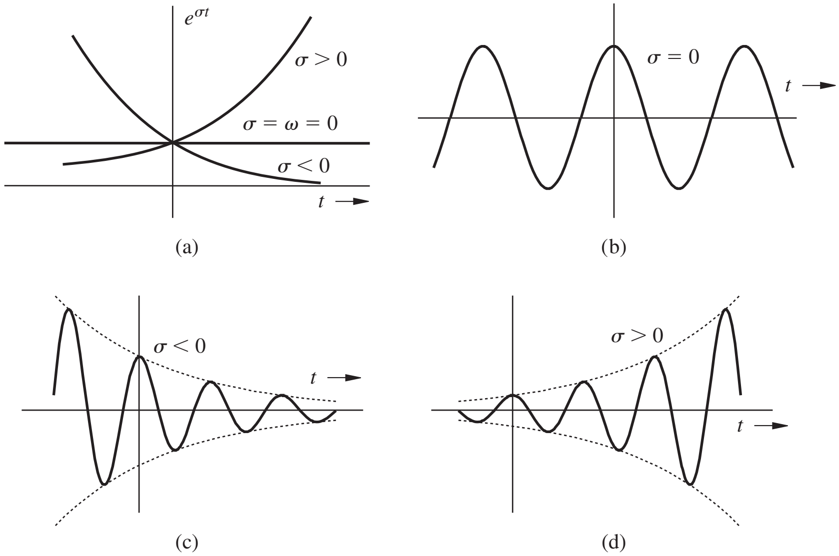


Figure 1.21 Sinusoids of complex frequency $\sigma + j\omega$.

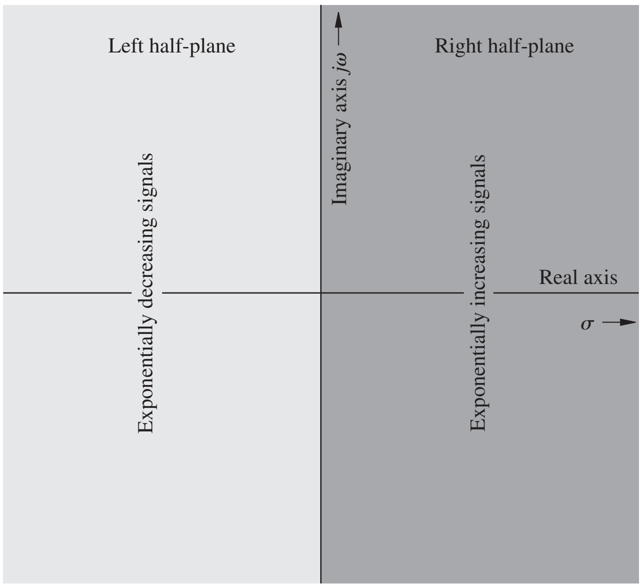


Figure 1.22 Complex frequency plane.

of oscillation is zero. Consequently these signals are monotonically increasing or decreasing exponentials (Fig. 1.21a). For signals whose frequencies lie on the imaginary axis ($j\omega$ axis where $\sigma = 0$), $e^{\sigma t} = 1$. Therefore, these signals are conventional sinusoids with constant amplitude (Fig. 1.21b). The case $s = 0$ ($\sigma = \omega = 0$) corresponds to a constant (dc) signal because $e^{0t} = 1$. For the signals illustrated in Fig. 1.21c and 1.21d, both σ and ω are nonzero; the frequency s is complex and does not lie on either axis. The signal in Fig. 1.21c decays exponentially. Therefore, σ is negative, and s lies to the left of the imaginary axis. In contrast, the signal in Fig. 1.21d *grows* exponentially. Therefore, σ is positive, and s lies to the right of the imaginary axis. Thus the s plane (Fig. 1.21) can be separated into two parts: the *left half-plane* (LHP) corresponding to exponentially decaying signals and the *right half-plane* (RHP) corresponding to exponentially growing signals. The imaginary axis separates the two regions and corresponds to signals of constant amplitude.

An exponentially growing sinusoid $e^{2t}\cos 5t$, for example, can be expressed as a linear combination of exponentials $e^{(2+j5)t}$ and $e^{(2-j5)t}$ with complex frequencies $2+j5$ and $2-j5$, respectively, which lie in the RHP. An exponentially decaying sinusoid $e^{-2t}\cos 5t$ can be expressed as a linear combination of exponentials $e^{(-2+j5)t}$ and $e^{(-2-j5)t}$ with complex frequencies $-2+j5$ and $-2-j5$, respectively, which lie in the LHP. A constant amplitude sinusoid $\cos 5t$ can be expressed as a linear combination of exponentials e^{j5t} and e^{-j5t} with complex frequencies $\pm j5$, which lie on the imaginary axis. Observe that the monotonic exponentials $e^{\pm 2t}$ are also generalized sinusoids with complex frequencies ± 2.

1.5 Even and Odd Functions

A real function $x_e(t)$ is said to be an *even function* of t if[†]

$$x_e(t) = x_e(-t) \tag{1.31}$$

and a real function $x_o(t)$ is said to be an *odd function* of t if

$$x_o(t) = -x_o(-t) \tag{1.32}$$

An even function has the same value at the instants t and $-t$ for all values of t. Clearly, $x_e(t)$ is symmetrical about the vertical axis, as shown in Fig. 1.23a. On the other hand, the value of an odd function at the instant t is the negative of its value at the instant $-t$. Therefore, $x_o(t)$ is antisymmetrical about the vertical axis, as depicted in Fig. 1.23b.

1.5-1 Some Properties of Even and Odd Functions

Even and odd functions have the following properties:

$$\begin{aligned}
\text{even function} \times \text{odd function} &= \text{odd function} \\
\text{odd function} \times \text{odd function} &= \text{even function} \\
\text{even function} \times \text{even function} &= \text{even function}
\end{aligned}$$

The proofs are trivial and follow directly from the definition of odd and even functions [Eqs. (1.31) and (1.32)].

[†]A complex signal $x(t)$ is said to be *conjugate symmetrical* if $x(t) = x^*(-t)$. A real conjugate symmetrical signal is an even signal. A signal is *conjugate antisymmetrical* if $x(t) = -x^*(-t)$. A real conjugate antisymmetrical signal is an odd signal.

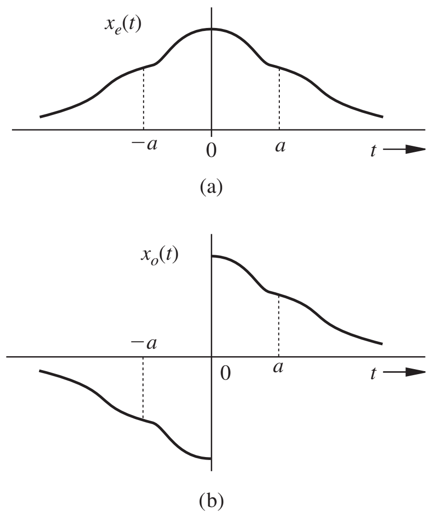


Figure 1.23 Functions of t: **(a)** even and **(b)** odd.

AREA

Because $x_e(t)$ is symmetrical about the vertical axis, it follows from Fig. 1.23a that

$$\int_{-a}^{a} x_e(t)\,dt = 2\int_{0}^{a} x_e(t)\,dt \tag{1.33a}$$

It is also clear from Fig. 1.23b that

$$\int_{-a}^{a} x_o(t)\,dt = 0 \tag{1.33b}$$

These results are valid under the assumption that there is no impulse (or its derivatives) at the origin. The proof of these statements is obvious from the plots of the even and the odd function. Formal proofs, left as an exercise for the reader, can be accomplished by using the definitions in Eqs. (1.31) and (1.32).

Because of their properties, study of odd and even functions proves useful in many applications, as will become evident in later chapters.

1.5-2 Even and Odd Components of a Signal

Every signal $x(t)$ can be expressed as a sum of even and odd components because

$$x(t) = \underbrace{\tfrac{1}{2}[x(t) + x(-t)]}_{\text{even}} + \underbrace{\tfrac{1}{2}[x(t) - x(-t)]}_{\text{odd}} \tag{1.34}$$

From the definitions in Eqs. (1.31) and (1.32), we can clearly see that the first component on the right-hand side is an even function, while the second component is odd. This is apparent from the fact that replacing t by $-t$ in the first component yields the same function. The same maneuver in the second component yields the negative of that component.

Consider the function

$$x(t) = e^{-at}u(t)$$

Expressing this function as a sum of the even and odd components $x_e(t)$ and $x_o(t)$, we obtain

$$x(t) = x_e(t) + x_o(t)$$

where [from Eq. (1.34)]

$$x_e(t) = \tfrac{1}{2}[e^{-at}u(t) + e^{at}u(-t)] \tag{1.35a}$$

and

$$x_o(t) = \tfrac{1}{2}[e^{-at}u(t) - e^{at}u(-t)] \tag{1.35b}$$

The function $e^{-at}u(t)$ and its even and odd components are illustrated in Fig. 1.24.

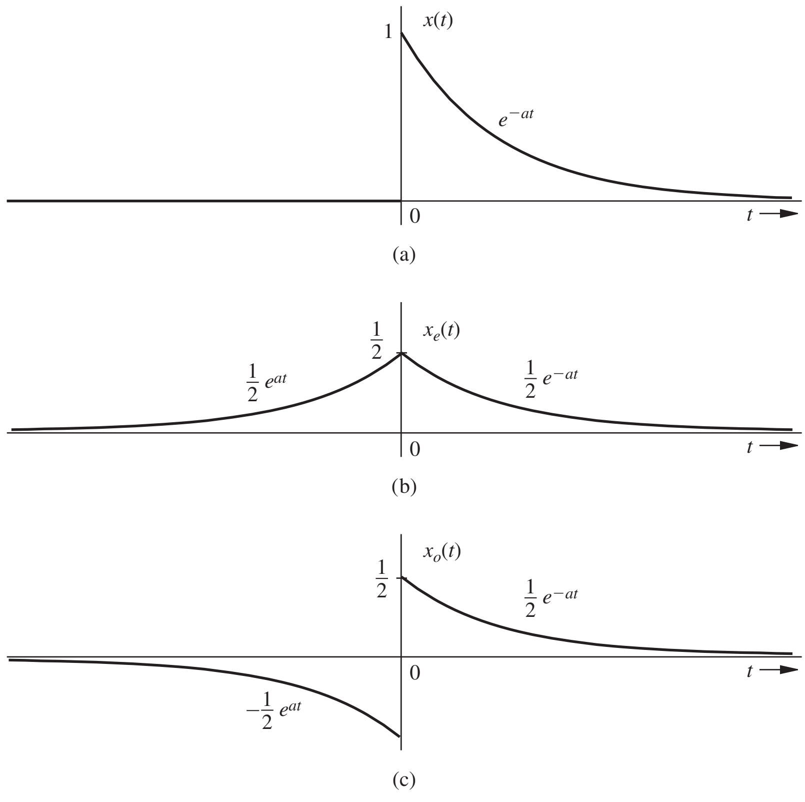


Figure 1.24 Finding even and odd components of a signal.

EXAMPLE 1.8

Find the even and odd components of e^{jt}.

From Eq. (1.34)

$$e^{jt} = x_e(t) + x_o(t)$$

where

$$x_e(t) = \tfrac{1}{2}[e^{jt} + e^{-jt}] = \cos t$$

and

$$x_o(t) = \tfrac{1}{2}[e^{jt} - e^{-jt}] = j \sin t$$

1.6 Systems

As mentioned in Section 1.1, systems are used to process signals to allow modification or extraction of additional information from the signals. A system may consist of physical components (hardware realization) or of an algorithm that computes the output signal from the input signal (software realization).

Roughly speaking, a physical system consists of interconnected components, which are characterized by their terminal (input–output) relationships. In addition, a system is governed by laws of interconnection. For example, in electrical systems, the terminal relationships are the familiar voltage–current relationships for the resistors, capacitors, inductors, transformers, transistors, and so on, as well as the laws of interconnection (i.e., Kirchhoff's laws). Using these laws, we derive mathematical equations relating the outputs to the inputs. These equations then represent a *mathematical model* of the system.

A system can be conveniently illustrated by a "black box" with one set of accessible terminals where the input variables $x_1(t), x_2(t), \ldots, x_j(t)$ are applied and another set of accessible terminals where the output variables $y_1(t), y_2(t), \ldots, y_k(t)$ are observed (Fig. 1.25).

The study of systems consists of three major areas: mathematical modeling, analysis, and design. Although we shall be dealing with mathematical modeling, our main concern is with analysis and design. The major portion of this book is devoted to the analysis problem—how to determine the system outputs for the given inputs and a given mathematical model of the system

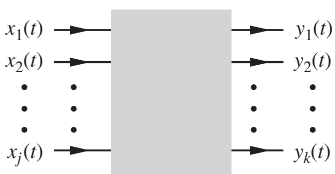


Figure 1.25 Representation of a system.

(or rules governing the system). To a lesser extent, we will also consider the problem of design or synthesis—how to construct a system that will produce a desired set of outputs for the given inputs.

Data Needed to Compute System Response

To understand what data we need to compute a system response, consider a simple *RC* circuit with a current source $x(t)$ as its input (Fig. 1.26). The output voltage $y(t)$ is given by

$$y(t) = Rx(t) + \frac{1}{C}\int_{-\infty}^{t} x(\tau)\,d\tau \tag{1.36a}$$

The limits of the integral on the right-hand side are from $-\infty$ to t because this integral represents the capacitor charge due to the current $x(t)$ flowing in the capacitor, and this charge is the result of the current flowing in the capacitor from $-\infty$. Now, Eq. (1.36a) can be expressed as

$$y(t) = Rx(t) + \frac{1}{C}\int_{-\infty}^{0} x(\tau)\,d\tau + \frac{1}{C}\int_{0}^{t} x(\tau)\,d\tau \tag{1.36b}$$

The middle term on the right-hand side is $v_C(0)$, the capacitor voltage at $t = 0$. Therefore

$$y(t) = v_C(0) + Rx(t) + \frac{1}{C}\int_{0}^{t} x(\tau)\,d\tau \qquad t \geq 0 \tag{1.36c}$$

This equation can be readily generalized as

$$y(t) = v_C(t_0) + Rx(t) + \frac{1}{C}\int_{t_0}^{t} x(\tau)\,d\tau \qquad t \geq t_0 \tag{1.36d}$$

From Eq. (1.36a), the output voltage $y(t)$ at an instant t can be computed if we know the input current flowing in the capacitor throughout its entire past ($-\infty$ to t). Alternatively, if we know the input current $x(t)$ from some moment t_0 onward, then, using Eq. (1.36d), we can still calculate $y(t)$ for $t \geq t_0$ from a knowledge of the input current, provided we know $v_C(t_0)$, the initial capacitor voltage (voltage at t_0). Thus $v_C(t_0)$ contains all the relevant information about the circuit's entire past ($-\infty$ to t_0) that we need to compute $y(t)$ for $t \geq t_0$. Therefore, the response of a system at $t \geq t_0$ can be determined from its input(s) during the interval t_0 to t and from certain *initial conditions* at $t = t_0$.

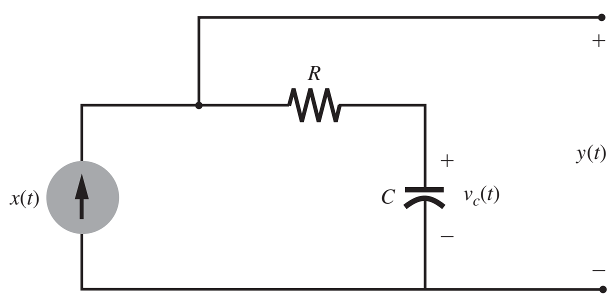


Figure 1.26 Example of a simple electrical system.

In the preceding example, we needed only one initial condition. However, in more complex systems, several initial conditions may be necessary. We know, for example, that in passive *RLC* networks, the initial values of all inductor currents and all capacitor voltages[†] are needed to determine the outputs at any instant $t \geq 0$ if the inputs are given over the interval $[0, t]$.

1.7 Classification of Systems

Systems may be classified broadly in the following categories:

1. Linear and nonlinear systems
2. Constant-parameter and time-varying-parameter systems
3. Instantaneous (memoryless) and dynamic (with memory) systems
4. Causal and noncausal systems
5. Continuous-time and discrete-time systems
6. Analog and digital systems
7. Invertible and noninvertible systems
8. Stable and unstable systems

Other classifications, such as deterministic and probabilistic systems, are beyond the scope of this text and are not considered.

1.7-1 Linear and Nonlinear Systems

The Concept of Linearity

A system whose output is proportional to its input is an *example* of a linear system. But linearity implies more than this; it also implies the *additivity property:* that is, if several inputs are acting on a system, then the total effect on the system due to all these inputs can be determined by considering one input at a time while assuming all the other inputs to be zero. The total effect is then the sum of all the component effects. This property may be expressed as follows: for a linear system, if an input x_1 acting alone has an effect y_1, and if another input x_2, also acting alone, has an effect y_2, then, with both inputs acting on the system, the total effect will be $y_1 + y_2$. Thus, if

$$x_1 \longrightarrow y_1 \qquad \text{and} \qquad x_2 \longrightarrow y_2 \tag{1.37}$$

then for all x_1 and x_2

$$x_1 + x_2 \longrightarrow y_1 + y_2 \tag{1.38}$$

In addition, a linear system must satisfy the *homogeneity* or scaling property, which states that for arbitrary real or imaginary number k, if an input is increased k-fold, the effect also increases k-fold. Thus, if

$$x \longrightarrow y$$

[†]Strictly speaking, this means independent inductor currents and capacitor voltages.

then for all real or imaginary k

$$kx \longrightarrow ky \tag{1.39}$$

Thus, linearity implies two properties: homogeneity (scaling) and additivity.† Both these properties can be combined into one property (*superposition*), which is expressed as follows: If

$$x_1 \longrightarrow y_1 \quad \text{and} \quad x_2 \longrightarrow y_2$$

then for all values of constants k_1 and k_2,

$$k_1x_1 + k_2x_2 \longrightarrow k_1y_1 + k_2y_2 \tag{1.40}$$

This is true for all x_1 and x_2.

It may appear that additivity implies homogeneity. Unfortunately, homogeneity does not always follow from additivity. Exercise E1.11 demonstrates such a case.

EXERCISE E1.11

Show that a system with the input $x(t)$ and the output $y(t)$ related by $y(t) = \text{Re}\{x(t)\}$ satisfies the additivity property but violates the homogeneity property. Hence, such a system is not linear. [Hint: Show that Eq. (1.39) is not satisfied when k is complex.]

Response of a Linear System

For the sake of simplicity, we discuss only *single-input, single-output* (*SISO*) systems. But the discussion can be readily extended to *multiple-input, multiple-output* (*MIMO*) systems.

A system's output for $t \geq 0$ is the result of two independent causes: the initial conditions of the system (or the system state) at $t = 0$ and the input $x(t)$ for $t \geq 0$. If a system is to be linear, the output must be the sum of the two components resulting from these two causes: first, the *zero-input response* component that results only from the initial conditions at $t = 0$ with the input $x(t) = 0$ for $t \geq 0$, and then the *zero-state response* component that results only from the input $x(t)$ for $t \geq 0$ when the initial conditions (at $t = 0$) are assumed to be zero. When all the appropriate initial conditions are zero, the system is said to be in *zero state*. The system output is zero when the input is zero only if the system is in zero state.

In summary, a linear system response can be expressed as the sum of the zero-input and the zero-state component:

$$\textbf{total response} = \textbf{zero-input response} + \textbf{zero-state response} \tag{1.41}$$

This property of linear systems, which permits the separation of an output into components resulting from the initial conditions and from the input, is called the *decomposition property*.

† A linear system must also satisfy the additional condition of *smoothness,* where small changes in the system's inputs must result in small changes in its outputs.[3]

For the *RC* circuit of Fig. 1.26, the response $y(t)$ was found to be [see Eq. (1.36c)]

$$y(t) = \underbrace{v_C(0)}_{z-i \text{ component}} + \underbrace{Rx(t) + \frac{1}{C}\int_0^t x(\tau)\,d\tau}_{z-s \text{ component}} \tag{1.42}$$

From Eq. (1.42), it is clear that if the input $x(t) = 0$ for $t \geq 0$, the output $y(t) = v_C(0)$. Hence $v_C(0)$ is the zero-input component of the response $y(t)$. Similarly, if the system state (the voltage v_C in this case) is zero at $t = 0$, the output is given by the second component on the right-hand side of Eq. (1.42). Clearly this is the zero-state component of the response $y(t)$.

In addition to the decomposition property, linearity implies that both the zero-input and zero-state components must obey the principle of superposition with respect to each of their respective causes. For example, if we increase the initial condition k-fold, the zero-input component must also increase k-fold. Similarly, if we increase the input k-fold, the zero-state component must also increase k-fold. These facts can be readily verified from Eq. (1.42) for the *RC* circuit in Fig. 1.26. For instance, if we double the initial condition $v_C(0)$, the zero-input component doubles; if we double the input $x(t)$, the zero-state component doubles.

EXAMPLE 1.9

Show that the system described by the equation

$$\frac{dy}{dt} + 3y(t) = x(t) \tag{1.43}$$

is linear.[†]

Let the system response to the inputs $x_1(t)$ and $x_2(t)$ be $y_1(t)$ and $y_2(t)$, respectively. Then

$$\frac{dy_1}{dt} + 3y_1(t) = x_1(t)$$

and

$$\frac{dy_2}{dt} + 3y_2(t) = x_2(t)$$

[†]Equations such as (1.43) and (1.44) are considered to represent linear systems in the classical definition of linearity. Some authors consider such equations to represent *incrementally linear* systems. According to this definition, *linear system* has only a zero-state component. The zero-input component is absent. Hence, incrementally linear system response can be represented as a response of a linear system (linear in this new definition) plus a zero-input component. We prefer the classical definition to this new definition. It is just a matter of definition and makes no difference in the final results.

Multiplying the first equation by k_1, the second with k_2, and adding them yields

$$\frac{d}{dt}[k_1y_1(t) + k_2y_2(t)] + 3[k_1y_1(t) + k_2y_2(t)] = k_1x_1(t) + k_2x_2(t)$$

But this equation is the system equation [Eq. (1.43)] with

$$x(t) = k_1x_1(t) + k_2x_2(t)$$

and

$$y(t) = k_1y_1(t) + k_2y_2(t)$$

Therefore, when the input is $k_1x_1(t) + k_2x_2(t)$, the system response is $k_1y_1(t) + k_2y_2(t)$. Consequently, the system is linear. Using this argument, we can readily generalize the result to show that a system described by a differential equation of the form

$$a_0\frac{d^Ny}{dt^N} + a_1\frac{d^{N-1}y}{dt^{N-1}} + \cdots + a_Ny(t) = b_{N-M}\frac{d^Mx}{dt^M} + \cdots + b_{N-1}\frac{dx}{dt} + b_Nx(t) \qquad (1.44)$$

is a linear system. The coefficients a_i and b_i in this equation can be constants or functions of time. Although here we proved only zero-state linearity, it can be shown that such systems are also zero-input linear and have the decomposition property.

EXERCISE E1.12

Show that the system described by the following equation is linear:

$$\frac{dy}{dt} + t^2y(t) = (2t+3)x(t)$$

EXERCISE E1.13

Show that the system described by the following equation is nonlinear:

$$y(t)\frac{dy}{dt} + 3y(t) = x(t)$$

More Comments on Linear Systems

Almost all systems observed in practice become nonlinear when large enough signals are applied to them. However, it is possible to approximate most of the nonlinear systems by linear systems for small-signal analysis. The analysis of nonlinear systems is generally difficult. Nonlinearities

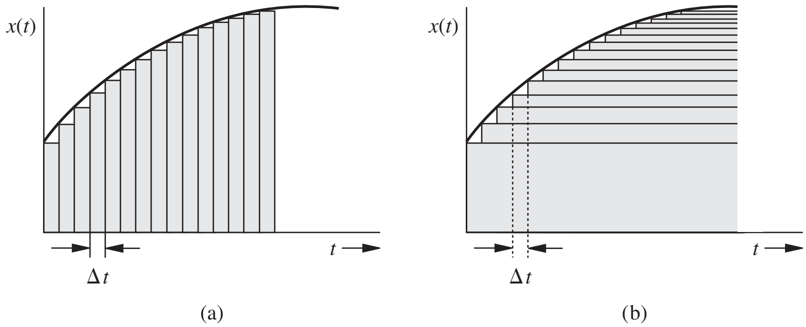


Figure 1.27 Signal representation in terms of impulse and step components.

can arise in so many ways that describing them with a common mathematical form is impossible. Not only is each system a category in itself, but even for a given system, changes in initial conditions or input amplitudes may change the nature of the problem. On the other hand, the superposition property of linear systems is a powerful unifying principle that allows for a general solution. The superposition property (linearity) greatly simplifies the analysis of linear systems. Because of the decomposition property, we can evaluate separately the two components of the output. The zero-input component can be computed by assuming the input to be zero, and the zero-state component can be computed by assuming zero initial conditions. Moreover, if we express an input $x(t)$ as a sum of simpler functions,

$$x(t) = a_1x_1(t) + a_2x_2(t) + \cdots + a_mx_m(t)$$

then, by virtue of linearity, the response $y(t)$ is given by

$$y(t) = a_1y_1(t) + a_2y_2(t) + \cdots + a_my_m(t) \tag{1.45}$$

where $y_k(t)$ is the zero-state response to an input $x_k(t)$. This apparently trivial observation has profound implications. As we shall see repeatedly in later chapters, it proves extremely useful and opens new avenues for analyzing linear systems.

For example, consider an arbitrary input $x(t)$ such as the one shown in Fig. 1.27a. We can approximate $x(t)$ with a sum of rectangular pulses of width Δt and of varying heights. The approximation improves as $\Delta t \to 0$, when the rectangular pulses become impulses spaced Δt seconds apart (with $\Delta t \to 0$).[†] Thus, an arbitrary input can be replaced by a weighted sum of impulses spaced Δt ($\Delta t \to 0$) seconds apart. Therefore, if we know the system response to a unit impulse, we can immediately determine the system response to an arbitrary input $x(t)$ by adding the system response to each impulse component of $x(t)$. A similar situation is depicted in Fig. 1.27b, where $x(t)$ is approximated by a sum of step functions of varying magnitude and spaced Δt seconds apart. The approximation improves as Δt becomes smaller. Therefore, if we know the system response to a unit step input, we can compute the system response to any arbitrary input $x(t)$ with relative ease. Time-domain analysis of linear systems (discussed in Chapter 2) uses this approach.

[†]Here, the discussion of a rectangular pulse approaching an impulse at $\Delta t \to 0$ is somewhat imprecise. It is explained in Section 2.4 with more rigor.

Chapters 4 through 7 employ the same approach but instead use sinusoids or exponentials as the basic signal components. We show that any arbitrary input signal can be expressed as a weighted sum of sinusoids (or exponentials) having various frequencies. Thus a knowledge of the system response to a sinusoid enables us to determine the system response to an arbitrary input $x(t)$.

1.7-2 Time-Invariant and Time-Varying Systems

Systems whose parameters do not change with time are *time-invariant* (also *constant-parameter*) systems. For such a system, if the input is delayed by T seconds, the output is the same as before but delayed by T (assuming initial conditions are also delayed by T). This property is expressed graphically in Fig. 1.28. We can also illustrate this property, as shown in Fig. 1.29. We can delay

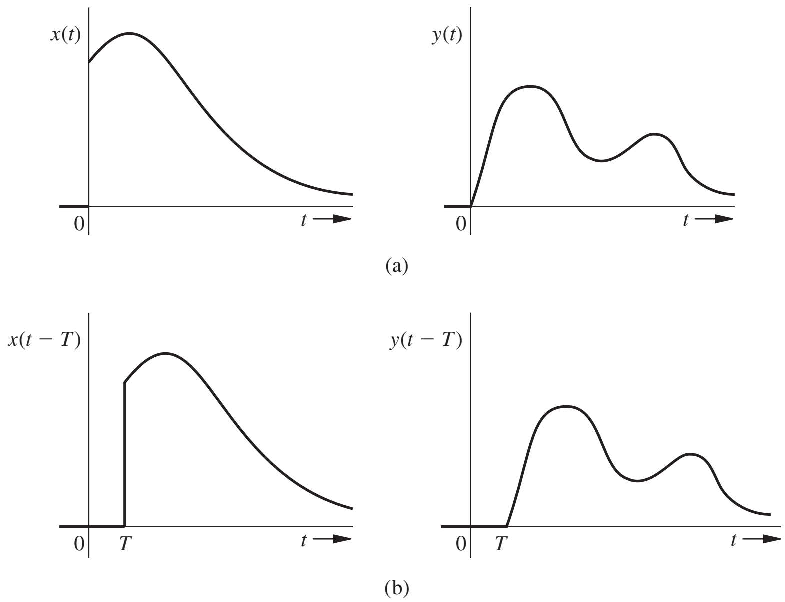


Figure 1.28 Time-invariance property.

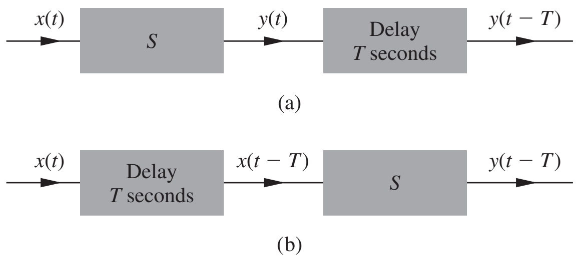


Figure 1.29 Illustration of time-invariance property.

the output $y(t)$ of a system $\mathcal{S}$ by applying the output $y(t)$ to a T second delay (Fig. 1.29a). If the system is time invariant, then the delayed output $y(t-T)$ can also be obtained by first delaying the input $x(t)$ before applying it to the system, as shown in Fig. 1.29b. In other words, the system $\mathcal{S}$ and the time delay commute if the system $\mathcal{S}$ is time invariant. This would not be true for time-varying systems. Consider, for instance, a time-varying system specified by $y(t) = e^{-t}x(t)$. The output for such a system in Fig. 1.29a is $e^{-(t-T)}x(t-T)$. In contrast, the output for the system in Fig. 1.29b is $e^{-t}x(t-T)$.

It is possible to verify that the system in Fig. 1.26 is a time-invariant system. Networks composed of *RLC* elements and other commonly used active elements such as transistors are time-invariant systems. A system with an input–output relationship described by a linear differential equation of the form given in Example 1.9 [Eq. (1.44)] is a linear time-invariant (LTI) system when the coefficients a_i and b_i of such equation are constants. If these coefficients are functions of time, then the system is a linear *time-varying* system.

The system described in Exercise E1.12 is linear time varying. Another familiar example of a time-varying system is the carbon microphone, in which the resistance R is a function of the mechanical pressure generated by sound waves on the carbon granules of the microphone. The output current from the microphone is thus modulated by the sound waves, as desired.

EXERCISE E1.14

Show that a system described by the following equation is time-varying-parameter system:

$$y(t) = (\sin\, t)x(t-2)$$

[Hint: Show that the system fails to satisfy the time-invariance property.]

1.7-3 Instantaneous and Dynamic Systems

As observed earlier, a system's output at any instant t generally depends on the entire past input. However, in a special class of systems, the output at any instant t depends only on its input at that instant. In resistive networks, for example, any output of the network at some instant t depends only on the input at the instant t. In these systems, past history is irrelevant in determining the response. Such systems are said to be *instantaneous* or *memoryless* systems. More precisely, a system is said to be instantaneous (or memoryless) if its output at any instant t depends, at most, on the strength of its input(s) at the same instant t, and not on any past or future values of the input(s). Otherwise, the system is said to be *dynamic* (or a system with memory). A system whose response at t is completely determined by the input signals over the past T seconds [interval from $(t-T)$ to t] is a *finite-memory system* with a memory of T seconds. Networks containing inductive and capacitive elements generally have infinite memory because the response of such networks at any instant t is determined by their inputs over the entire past $(-\infty, t)$. This is true for the *RC* circuit of Fig. 1.26.

In this book we will generally examine dynamic systems. Instantaneous systems are a special case of dynamic systems.

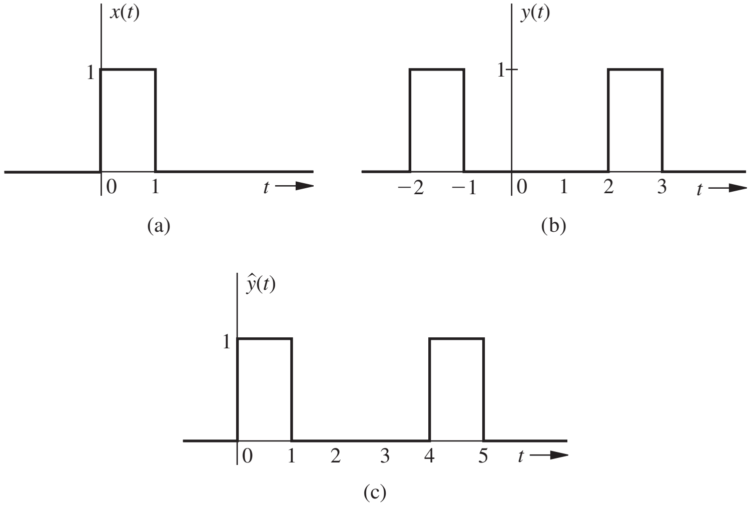


Figure 1.30 A noncausal system and its realization by a delayed causal system.

1.7-4 Causal and Noncausal Systems

A *causal* (also known as a *physical* or *nonanticipative*) system is one for which the output at any instant t_0 depends only on the value of the input $x(t)$ for $t \leq t_0$. In other words, the value of the output at the present instant depends only on the past and present values of the input $x(t)$, not on its future values. To put it simply, in a causal system the output cannot start before the input is applied. If the response starts before the input, it means that the system knows the input in the future and acts on this knowledge before the input is applied. A system that violates the condition of causality is called a *noncausal* (or *anticipative*) system.

Any practical system that operates in real time† must necessarily be causal. We do not yet know how to build a system that can respond to future inputs (inputs not yet applied). A noncausal system is a prophetic system that knows the future input and acts on it in the present. Thus, if we apply an input starting at $t = 0$ to a noncausal system, the output would begin even before $t = 0$. For example, consider the system specified by

$$y(t) = x(t-2) + x(t+2) \tag{1.46}$$

For the input $x(t)$ illustrated in Fig. 1.30a, the output $y(t)$, as computed from Eq. (1.46) (shown in Fig. 1.30b), starts even before the input is applied. Equation (1.46) shows that $y(t)$, the output at t, is given by the sum of the input values 2 seconds before and 2 seconds after t (at $t - 2$ and $t + 2$, respectively). But if we are operating the system in real time at t, we do not know what the value of the input will be 2 seconds later. Thus it is impossible to implement this system in real time. For this reason, noncausal systems are unrealizable in *real time*.

†In real-time operations, the response to an input is essentially simultaneous (contemporaneous) with the input itself.

Why Study Noncausal Systems?

The foregoing discussion may suggest that noncausal systems have no practical purpose. This is not the case; they are valuable in the study of systems for several reasons. First, noncausal systems *are* realizable when the independent variable is other than "time" (e.g., *space*). Consider, for example, an electric charge of density $q(x)$ placed along the x axis for $x \geq 0$. This charge density produces an electric field $E(x)$ that is present at every point on the x axis from $x = -\infty$ to ∞. In this case the input [i.e., the charge density $q(x)$] starts at $x = 0$, but its output [the electric field $E(x)$] begins before $x = 0$. Clearly, this space-charge system is noncausal. This discussion shows that only temporal systems (systems with time as independent variable) must be causal to be realizable. The terms "before" and "after" have a special connection to causality only when the independent variable is time. This connection is lost for variables other than time. Nontemporal systems, such as those occurring in optics, can be noncausal and still realizable.

Moreover, even for temporal systems, such as those used for signal processing, the study of noncausal systems is important. In such systems we may have all input data prerecorded. (This often happens with speech, geophysical, and meteorological signals, and with space probes.) In such cases, the input's future values are available to us. For example, suppose we had a set of input signal records available for the system described by Eq. (1.46). We can then compute $y(t)$ since, for any t, we need only refer to the records to find the input's value 2 seconds before and 2 seconds after t. Thus, noncausal systems can be realized, although not in real time. We may therefore be able to realize a noncausal system, provided we are willing to accept a time delay in the output. Consider a system whose output $\hat{y}(t)$ is the same as $y(t)$ in Eq. (1.46) delayed by 2 seconds (Fig. 1.30c), so that

$$\begin{aligned}\hat{y}(t) &= y(t-2) \\ &= x(t-4) + x(t)\end{aligned}$$

Here the value of the output $\hat{y}$ at any instant t is the sum of the values of the input x at t and at the instant 4 seconds earlier [at $(t-4)$]. In this case, the output at any instant t does not depend on future values of the input, and the system is causal. The output of this system, which is $\hat{y}(t)$, is identical to that in Eq. (1.46) or Fig. 1.30b except for a delay of 2 seconds. Thus, a noncausal system may be realized or satisfactorily approximated in real time by using a causal system with a delay.

Noncausal systems are realizable with time delay!

A third reason for studying noncausal systems is that they provide an upper bound on the performance of causal systems. For example, if we wish to design a filter for separating a signal from noise, then the optimum filter is invariably a noncausal system. Although unrealizable, this noncausal system's performance acts as the upper limit on what can be achieved and gives us a standard for evaluating the performance of causal filters.

At first glance, noncausal systems may seem to be inscrutable. Actually, there is nothing mysterious about these systems and their approximate realization through physical systems with delay. If we want to know what will happen one year from now, we have two choices: go to a prophet (an unrealizable person) who can give the answers instantly, or go to a wise man and allow him a delay of one year to give us the answer! If the wise man is truly wise, he may even be able, by studying trends, to shrewdly guess the future very closely with a delay of less than a year. Such is the case with noncausal systems—nothing more and nothing less.

EXERCISE E1.15

Show that a system described by the following equation is noncausal:

$$y(t) = \int_{t-5}^{t+5} x(\tau)\, d\tau$$

Show that this system can be realized physically if we accept a delay of 5 seconds in the output.

1.7-5 Continuous-Time and Discrete-Time Systems

Signals defined or specified over a continuous range of time are *continuous-time signals,* denoted by symbols $x(t)$, $y(t)$, and so on. Systems whose inputs and outputs are continuous-time signals are *continuous-time systems*. On the other hand, signals defined only at discrete instants of time $t_0, t_1, t_2, \ldots, t_n, \ldots$ are *discrete-time signals,* denoted by the symbols $x(t_n)$, $y(t_n)$, and so on, where n is some integer. Systems whose inputs and outputs are discrete-time signals are *discrete-time systems*. A digital computer is a familiar example of this type of system. In practice, discrete-time signals can arise from sampling continuous-time signals. For example, when the sampling is uniform, the discrete instants $t_0, t_1, t_2, \ldots$ are uniformly spaced so that

$$t_{k+1} - t_k = T \qquad \text{for all } k$$

In such case, the discrete-time signals represented by the samples of continuous-time signals $x(t)$, $y(t)$, and so on can be expressed as $x(nT)$, $y(nT)$, and so on; for convenience, we further simplify this notation to $x[n]$, $y[n]$, $\ldots$, where it is understood that $x[n] = x(nT)$ and that n is some integer. A typical discrete-time signal is shown in Fig. 1.31. A discrete-time signal may also be viewed as a sequence of numbers $\ldots, x[-1], x[0], x[1], x[2], \ldots$. Thus a discrete-time system may be seen as processing a sequence of numbers $x[n]$ and yielding as an output another sequence of numbers $y[n]$.

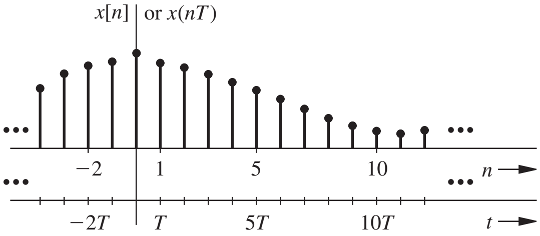


Figure 1.31 A discrete-time signal.

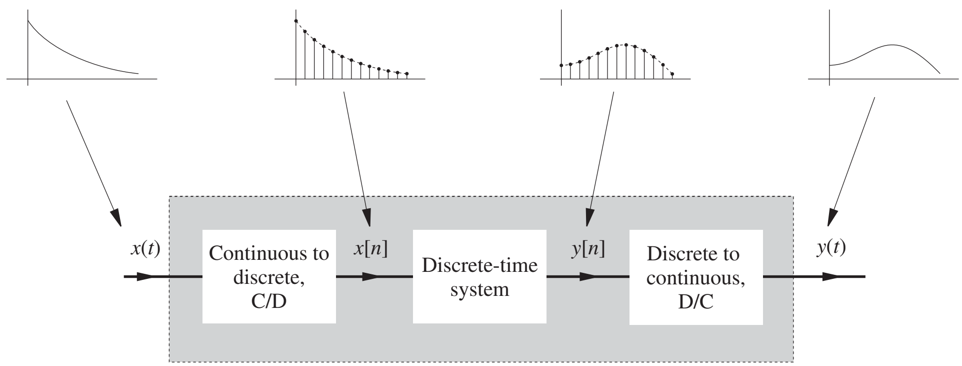


Figure 1.32 Processing continuous-time signals by discrete-time systems.

Discrete-time signals arise naturally in situations that are inherently discrete time, such as population studies, amortization problems, national income models, and radar tracking. They may also arise as a result of sampling continuous-time signals in sampled data systems, digital filtering, and the like. Digital filtering is a particularly interesting application in which continuous-time signals are processed by using discrete-time systems as shown in Fig. 1.32. A continuous-time signal $x(t)$ is first sampled to convert it into a discrete-time signal $x[n]$, which then is processed by the discrete-time system to yield a discrete-time output $y[n]$. A continuous-time signal $y(t)$ is finally constructed from $y[n]$. In this manner we can process a continuous-time signal with an appropriate discrete-time system such as a digital computer. Because discrete-time systems have several significant advantages over continuous-time systems, there is an accelerating trend toward processing continuous-time signals with discrete-time systems.

1.7-6 Analog and Digital Systems

Analog and digital signals are discussed in Section 1.3-2. A system whose input and output signals are analog is an *analog system;* a system whose input and output signals are digital is a *digital system*. A digital computer is an example of a digital (binary) system. Observe that a digital computer is digital as well as a discrete-time system.

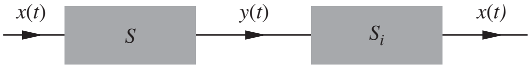


Figure 1.33 A cascade of a system with its inverse results in an identity system.

1.7-7 Invertible and Noninvertible Systems

A system $\mathcal{S}$ performs certain operation(s) on input signal(s). If we can obtain the input $x(t)$ back from the corresponding output $y(t)$ by some operation, the system $\mathcal{S}$ is said to be *invertible*. When several different inputs result in the same output (as in a rectifier), it is impossible to obtain the input from the output, and the system is *noninvertible*. Therefore, for an invertible system, it is essential that every input have a unique output so that there is a one-to-one mapping between an input and the corresponding output. The system that achieves the inverse operation [of obtaining $x(t)$ from $y(t)$] is the *inverse system* for $\mathcal{S}$. For instance, if $\mathcal{S}$ is an ideal integrator, then its inverse system is an ideal differentiator. Consider a system $\mathcal{S}$ connected in tandem with its inverse $\mathcal{S}_i$, as shown in Fig. 1.33. The input $x(t)$ to this tandem system results in signal $y(t)$ at the output of $\mathcal{S}$, and the signal $y(t)$, which now acts as an input to $\mathcal{S}_i$, yields back the signal $x(t)$ at the output of $\mathcal{S}_i$. Thus, $\mathcal{S}_i$ undoes the operation of $\mathcal{S}$ on $x(t)$, yielding back $x(t)$. A system whose output is equal to the input (for all possible inputs) is an *identity* system. Cascading a system with its inverse system, as shown in Fig. 1.33, results in an identity system.

In contrast, a rectifier, specified by an equation $y(t) = |x(t)|$, is noninvertible because the rectification operation cannot be undone.

Inverse systems are very important in signal processing. In many applications, the signals are distorted during the processing, and it is necessary to undo the distortion. For instance, in transmission of data over a communication channel, the signals are distorted owing to nonideal frequency response and finite bandwidth of a channel. It is necessary to restore the signal as closely as possible to its original shape. Such equalization is also used in audio systems and photographic systems.

1.7-8 Stable and Unstable Systems

Systems can also be classified as *stable* or *unstable* systems. Stability can be *internal* or *external*. If every *bounded input* applied at the input terminal results in a *bounded output,* the system is said to be stable *externally*. The external stability can be ascertained by measurements at the external terminals (input and output) of the system. This type of stability is also known as the stability in the BIBO (bounded-input/bounded-output) sense. The concept of internal stability is postponed to Chapter 2 because it requires some understanding of internal system behavior, introduced in that chapter.

EXERCISE E1.16

Show that a system described by the equation $y(t) = x^2(t)$ is noninvertible but BIBO stable.

1.8 System Model: Input–Output Description

A system description in terms of the measurements at the input and output terminals is called the *input–output description*. As mentioned earlier, systems theory encompasses a variety of systems, such as electrical, mechanical, hydraulic, acoustic, electromechanical, and chemical, as well as social, political, economic, and biological. The first step in analyzing any system is the construction of a system model, which is a mathematical expression or a rule that satisfactorily approximates the dynamical behavior of the system. In this chapter we shall consider only continuous-time systems. (Modeling of discrete-time systems is discussed in Chapter 3.)

1.8-1 Electrical Systems

To construct a system model, we must study the relationships between different variables in the system. In electrical systems, for example, we must determine a satisfactory model for the voltage–current relationship of each element, such as Ohm's law for a resistor. In addition, we must determine the various constraints on voltages and currents when several electrical elements are interconnected. These are the laws of interconnection—the well-known Kirchhoff laws for voltage and current (KVL and KCL). From all these equations, we eliminate unwanted variables to obtain equation(s) relating the desired output variable(s) to the input(s). The following examples demonstrate the procedure of deriving input-output relationships for some LTI electrical systems.

EXAMPLE 1.10

For the series RLC circuit of Fig. 1.34, find the input–output equation relating the input voltage $x(t)$ to the output current (loop current) $y(t)$.

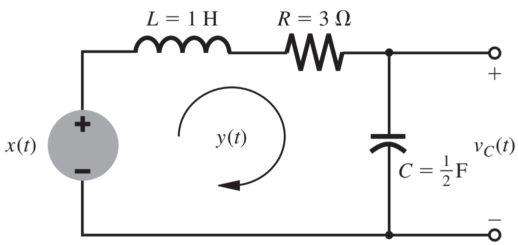


Figure 1.34

Application of Kirchhoff's voltage law around the loop yields

$$v_L(t) + v_R(t) + v_C(t) = x(t) \tag{1.47}$$

By using the voltage–current laws of each element (inductor, resistor, and capacitor), we can express this equation as

$$\frac{dy}{dt} + 3y(t) + 2\int_{-\infty}^{t} y(\tau)\,d\tau = x(t) \tag{1.48}$$

Differentiating both sides of this equation, we obtain

$$\frac{d^2y}{dt^2} + 3\frac{dy}{dt} + 2y(t) = \frac{dx}{dt} \tag{1.49}$$

This differential equation is the input–output relationship between the output $y(t)$ and the input $x(t)$.

It proves convenient to use a compact notation D for the differential operator d/dt. Thus

$$\frac{dy}{dt} \equiv Dy(t) \tag{1.50}$$

$$\frac{d^2y}{dt^2} \equiv D^2y(t) \tag{1.51}$$

and so on. With this notation, Eq. (1.49) can be expressed as

$$(D^2 + 3D + 2)y(t) = Dx(t) \tag{1.52}$$

The differential operator is the inverse of the integral operator, so we can use the operator $1/D$ to represent integration.†

$$\int_{-\infty}^{t} y(\tau)\,d\tau \equiv \frac{1}{D}y(t) \tag{1.53}$$

†Use of operator $1/D$ for integration generates some subtle mathematical difficulties because the operators D and $1/D$ do not commute. For instance, we know that $D(1/D) = 1$ because

$$\frac{d}{dt}\left[\int_{-\infty}^{t} y(\tau)\,d\tau\right] = y(t)$$

However, $(1/D)D$ is not necessarily unity. Use of Cramer's rule in solving simultaneous integro-differential equations will always result in cancellation of operators $1/D$ and D. This procedure may yield erroneous results when the factor D occurs in the numerator as well as in the denominator. This happens, for instance, in circuits with all-inductor loops or all-capacitor cut sets. To eliminate this problem, avoid the integral operation in system equations so that the resulting equations are differential rather than integro-differential. In electrical circuits, this can be done by using charge (instead of current) variables in loops containing capacitors and choosing current variables for loops without capacitors. In the literature this problem of commutativity of D and $1/D$ is largely ignored. As mentioned earlier, such procedure gives erroneous results only in special systems, such as the circuits with all-inductor loops or all-capacitor cut sets. Fortunately such systems constitute a very small fraction of the systems we deal with. For further discussion of this topic and a correct method of handling problems involving integrals, see Ref. 4.

Consequently, the loop equation (1.48) can be expressed as

$$\left(D + 3 + \frac{2}{D}\right) y(t) = x(t) \tag{1.54}$$

Multiplying both sides by D, that is, differentiating Eq. (1.54), we obtain

$$(D^2 + 3D + 2)y(t) = Dx(t) \tag{1.55}$$

which is identical to Eq. (1.52).

Recall that Eq. (1.55) is not an algebraic equation, and $D^2 + 3D + 2$ is not an algebraic term that multiplies $y(t)$; it is an operator that operates on $y(t)$. It means that we must perform the following operations on $y(t)$: take the second derivative of $y(t)$ and add to it 3 times the first derivative of $y(t)$ and 2 times $y(t)$. Clearly, a polynomial in D multiplied by $y(t)$ represents a certain differential operation on $y(t)$.

EXAMPLE 1.11

Find the equation relating the input to output for the series RC circuit of Fig. 1.35 if the input is the voltage $x(t)$ and output is

(a) the loop current $i(t)$

(b) the capacitor voltage $y(t)$

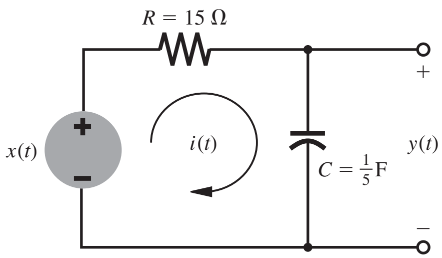


Figure 1.35

(a) The loop equation for the circuit is

$$R\,i(t) + \frac{1}{C}\int_{-\infty}^{t} i(\tau)\,d\tau = x(t) \tag{1.56}$$

or

$$15\,i(t) + 5\int_{-\infty}^{t} i(\tau)\,d\tau = x(t) \tag{1.57}$$

With operational notation, this equation can be expressed as

$$15\,i(t) + \frac{5}{D} i(t) = x(t) \tag{1.58}$$

(b) Multiplying both sides of Eqs. (1.58) by D (i.e., differentiating the equation), we obtain

$$(15D + 5)\, i(t) = Dx(t) \tag{1.59a}$$

or

$$15\frac{di}{dt} + 5\, i(t) = \frac{dx}{dt} \tag{1.59b}$$

Moreover,

$$\begin{aligned} i(t) &= C\,\frac{dy}{dt} \\ &= \tfrac{1}{5}Dy(t) \end{aligned}$$

Substitution of this result in Eq. (1.59a) yields

$$(3D + 1)y(t) = x(t) \tag{1.60}$$

or

$$3\frac{dy}{dt} + y(t) = x(t) \tag{1.61}$$

EXERCISE E1.17

For the RLC circuit in Fig. 1.34, find the input–output relationship if the output is the inductor voltage $v_L(t)$.

ANSWER
$(D^2 + 3D + 2)v_L(t) = D^2x(t)$

EXERCISE E1.18

For the RLC circuit in Fig. 1.34, find the input–output relationship if the output is the capacitor voltage $v_C(t)$.

ANSWER
$(D^2 + 3D + 2)v_C(t) = 2x(t)$

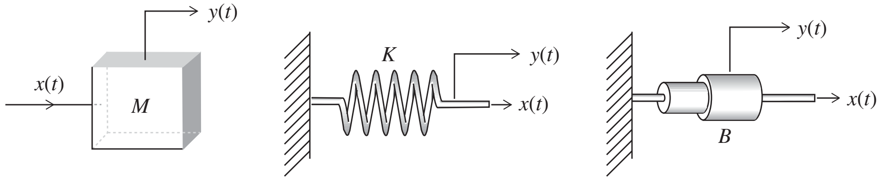


Figure 1.36 Some elements in translational mechanical systems.

1.8-2 Mechanical Systems

Planar motion can be resolved into translational (rectilinear) motion and rotational (torsional) motion. Translational motion will be considered first. We shall restrict ourselves to motions in one dimension.

Translational Systems

The basic elements used in modeling translational systems are ideal masses, linear springs, and dashpots providing viscous damping. The laws of various mechanical elements are now discussed.

For a *mass* M (Fig. 1.36a), a force $x(t)$ causes a motion $y(t)$ and acceleration $\ddot{y}(t)$. From Newton's law of motion,

$$x(t) = M\ddot{y}(t) = M\frac{d^2y}{dt^2} = MD^2y(t) \tag{1.62}$$

The force $x(t)$ required to stretch (or compress) a *linear spring* (Fig. 1.36b) by an amount $y(t)$ is given by

$$x(t) = Ky(t) \tag{1.63}$$

where K is the *stiffness* of the spring.

For *a linear dashpot* (Fig. 1.36c), which operates by virtue of viscous friction, the force moving the dashpot is proportional to the relative velocity $\dot{y}(t)$ of one surface with respect to the other. Thus

$$x(t) = B\dot{y}(t) = B\frac{dy}{dt} = BDy(t) \tag{1.64}$$

where B is the *damping coefficient* of the dashpot or the viscous friction.

EXAMPLE 1.12

Find the input–output relationship for the translational mechanical system shown in Fig. 1.37a or its equivalent in Fig. 1.37b. The input is the force $x(t)$, and the output is the mass position $y(t)$.

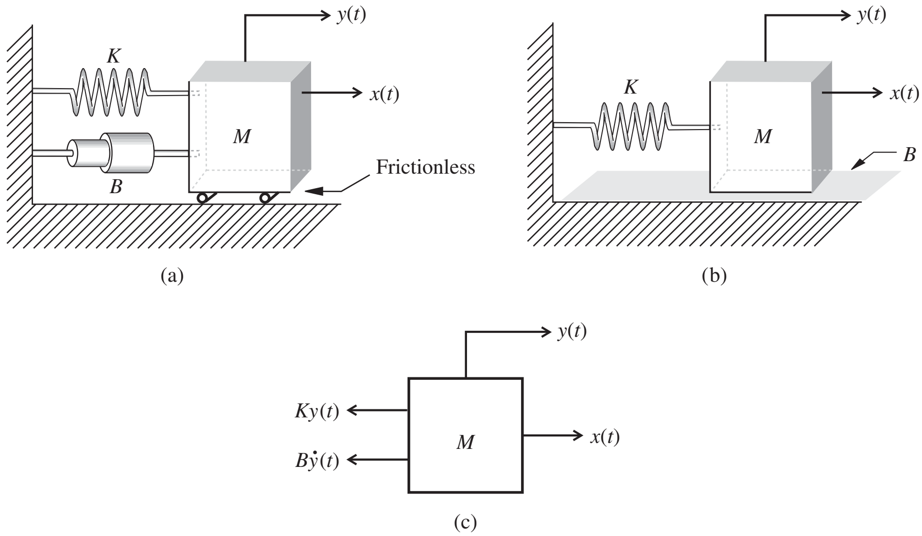


Figure 1.37

In mechanical systems it is helpful to draw a free-body diagram of each junction, which is a point at which two or more elements are connected. In Fig. 1.37, the point representing the mass is a junction. The displacement of the mass is denoted by $y(t)$. The spring is also stretched by the amount $y(t)$, and therefore it exerts a force $-Ky(t)$ on the mass. The dashpot exerts a force $-B\dot{y}(t)$ on the mass as shown in the free-body diagram (Fig. 1.37c). By Newton's second law, the net force must be $M\ddot{y}(t)$. Therefore

$$M\ddot{y}(t) = -B\dot{y}(t) - Ky(t) + x(t)$$

or

$$(MD^2 + BD + K)y(t) = x(t) \tag{1.65}$$

Rotational Systems

In rotational systems, the motion of a body may be defined as its motion about a certain axis. The variables used to describe rotational motion are torque (in place of force), angular position (in place of linear position), angular velocity (in place of linear velocity), and angular acceleration (in place of linear acceleration). The system elements are *rotational mass* or *moment of inertia* (in place of mass) and *torsional springs* and *torsional dashpots* (in place of linear springs and dashpots). The terminal equations for these elements are analogous to the corresponding equations for translational elements. If J is the moment of inertia (or rotational mass) of a rotating body about a certain axis, then the external torque required for this motion is equal to J (rotational mass) times the angular acceleration. If θ is the angular position of the body, $\ddot{\theta}$ is its angular acceleration, and

$$\text{torque} = J\ddot{\theta} = J\frac{d^2\theta}{dt^2} = JD^2\theta(t) \tag{1.66}$$

Similarly, if K is the stiffness of a torsional spring (per unit angular twist), and θ is the angular displacement of one terminal of the spring with respect to the other, then

$$\text{torque} = K\theta \tag{1.67}$$

Finally, the torque due to viscous damping of a torsional dashpot with damping coefficient B is

$$\text{torque} = B\dot{\theta}(t) = BD\theta(t) \tag{1.68}$$

EXAMPLE 1.13

The attitude of an aircraft can be controlled by three sets of surfaces (shown shaded in Fig. 1.38): elevators, rudder, and ailerons. By manipulating these surfaces, one can set the aircraft on a desired flight path. The roll angle φ can be controlled by deflecting in the opposite direction the two aileron surfaces as shown in Fig. 1.38. Assuming only rolling motion, find the equation relating the roll angle φ to the input (deflection) θ.

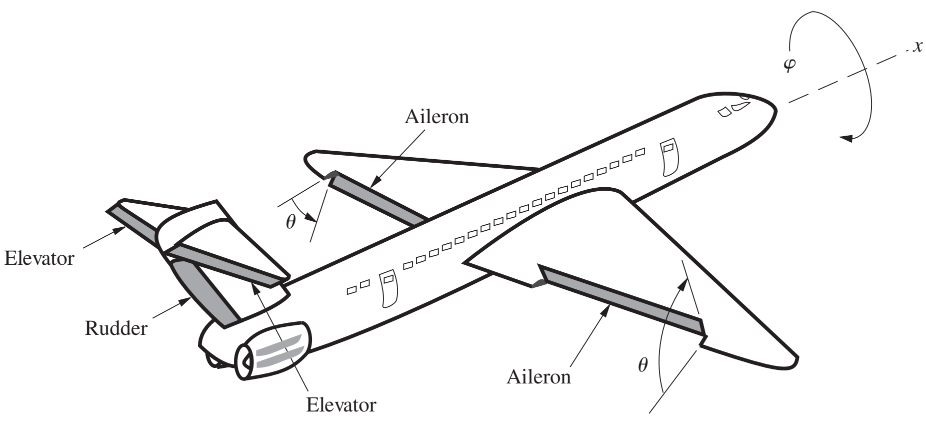


Figure 1.38 Attitude control of an airplane.

The aileron surfaces generate a torque about the roll axis proportional to the aileron deflection angle θ. Let this torque be $c\theta$, where c is the constant of proportionality. Air friction dissipates the torque $B\dot{\varphi}(t)$. The torque available for rolling motion is then $c\theta(t) - B\dot{\varphi}(t)$. If J is the moment of inertia of the plane about the x axis (roll axis), then

$$\begin{aligned} J\ddot{\varphi}(t) &= \text{net torque} \\ &= c\theta(t) - B\dot{\varphi}(t) \end{aligned} \tag{1.69}$$

and

$$J\frac{d^2\varphi}{dt^2} + B\frac{d\varphi}{dt} = c\,\theta(t) \tag{1.70}$$

or

$$(JD^2 + BD)\varphi(t) = c\theta(t) \tag{1.71}$$

This is the desired equation relating the output (roll angle φ) to the input (aileron angle θ).

The roll velocity ω is $\dot{\varphi}(t)$. If the desired output is the roll velocity ω rather than the roll angle φ, then the input–output equation would be

$$J\frac{d\omega}{dt} + B\omega = c\theta \tag{1.72}$$

or

$$(JD + B)\omega(t) = c\theta(t) \tag{1.73}$$

EXERCISE E1.19

Torque $\mathcal{T}(t)$ is applied to the rotational mechanical system shown in Fig. 1.39a. The torsional spring stiffness is K; the rotational mass (the cylinder's moment of inertia about the shaft) is J; the viscous damping coefficient between the cylinder and the ground is B. Find the equation relating the output angle θ to the input torque $\mathcal{T}$. [Hint: A free-body diagram is shown in Fig. 1.39b.]

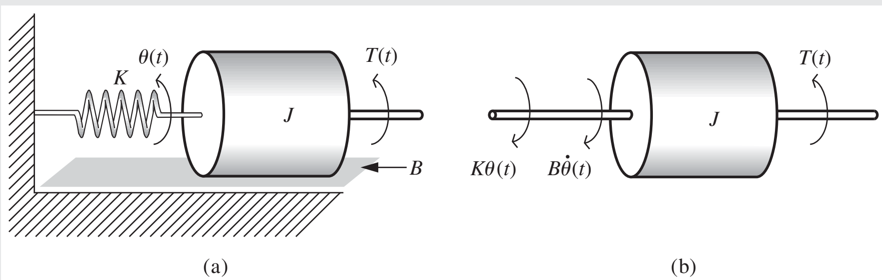


Figure 1.39 Rotational system.

ANSWER

$$J\frac{d^2\theta}{dt^2} + B\frac{d\theta}{dt} + K\theta(t) = \mathcal{T}(t)$$

or

$$(JD^2 + BD + K)\theta(t) = \mathcal{T}(t)$$

1.8-3 Electromechanical Systems

A wide variety of electromechanical systems convert electrical signals into mechanical motion (mechanical energy) and vice versa. Here we consider a rather simple example of an armature-controlled dc motor driven by a current source $x(t)$, as shown in Fig. 1.40a. The torque $\mathcal{T}(t)$ generated in the motor is proportional to the armature current $x(t)$. Therefore

$$\mathcal{T}(t) = K_T x(t) \tag{1.74}$$

where K_T is a constant of the motor. This torque drives a mechanical load whose free-body diagram is shown in Fig. 1.40b. The viscous damping (with coefficient B) dissipates a torque $B\dot{\theta}(t)$. If J is the moment of inertia of the load (including the rotor of the motor), then the net torque $\mathcal{T}(t) - B\dot{\theta}(t)$ must be equal to $J\ddot{\theta}(t)$:

$$J\ddot{\theta}(t) = \mathcal{T}(t) - B\dot{\theta}(t) \tag{1.75}$$

Thus

$$\begin{aligned}(JD^2 + BD)\theta(t) &= \mathcal{T}(t) \\ &= K_T x(t)\end{aligned} \tag{1.76}$$

which in conventional form can be expressed as

$$J\frac{d^2\theta}{dt^2} + B\frac{d\theta}{dt} = K_T\, x(t) \tag{1.77}$$

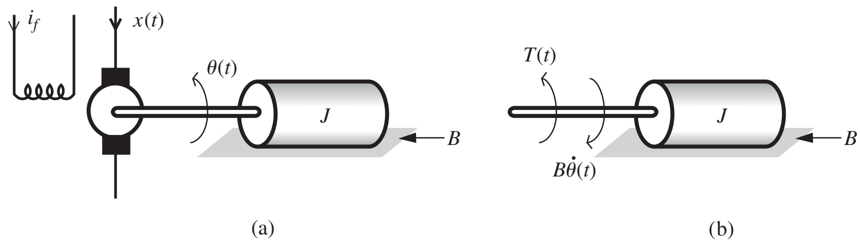


Figure 1.40 Armature-controlled dc motor.

1.9 Internal and External Description of a System

The input–output relationship of a system is an *external description* of that system. We have found an external description (not the *internal description*) of systems in all the examples discussed so far. This may puzzle the reader because in each of these cases, we derived the input–output relationship by analyzing the internal structure of that system. Why is this not an internal description? What makes a description internal? Although it is true that we did find the input–output description by internal analysis of the system, we did so strictly for convenience. We could have obtained the input–output description by making observations at the external (input and output) terminals, for example, by measuring the output for certain inputs such as an impulse or a sinusoid. A description that can be obtained from measurements at the external terminals (even when the rest of the system is sealed inside an inaccessible black box) is an external description. Clearly, the input–output description is an external description. What, then, is an internal description? Internal description is capable of providing the complete information about all possible signals in the system. An external description may not give such complete information. An external description can always be found from an internal description, but the converse is not necessarily true. We shall now give an example to clarify the distinction between an external and an internal description.

Let the circuit in Fig. 1.41a with the input $x(t)$ and the output $y(t)$ be enclosed inside a "black box" with only the input and the output terminals accessible. To determine its external description, let us apply a known voltage $x(t)$ at the input terminals and measure the resulting output voltage $y(t)$.

Let us also assume that there is some initial charge Q_0 present on the capacitor. The output voltage will generally depend on both, the input $x(t)$ and the initial charge Q_0. To compute the output resulting because of the charge Q_0, assume the input $x(t) = 0$ (short across the input).

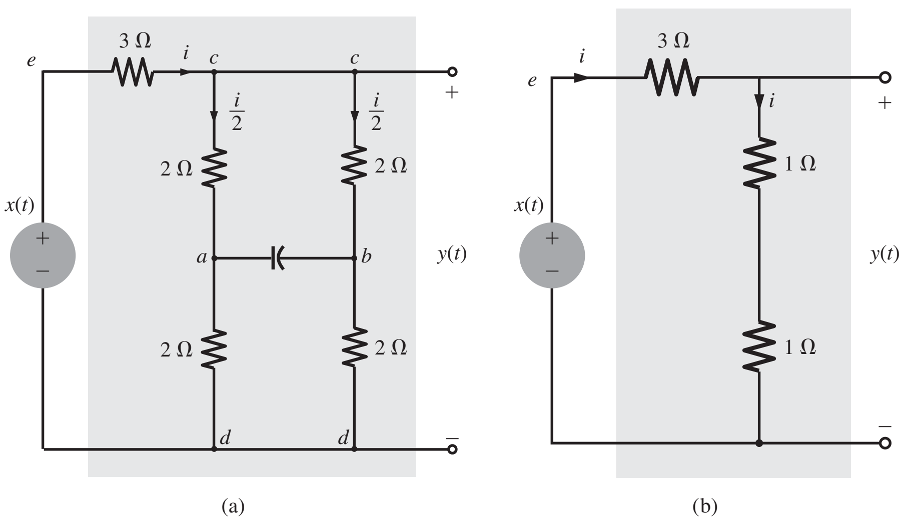


Figure 1.41 A system that cannot be described by external measurements.

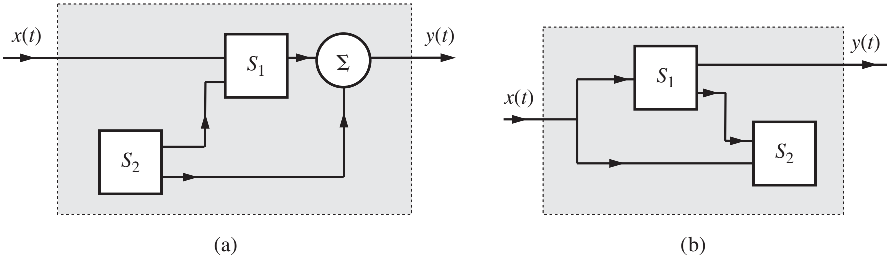


Figure 1.42 Structures of uncontrollable and unobservable systems.

In this case, the currents in the two 2 Ω resistors in the upper and the lower branches at the output terminals are equal and opposite because of the balanced nature of the circuit. Clearly, the capacitor charge results in zero voltage at the output.†

Now, to compute the output $y(t)$ resulting from the input voltage $x(t)$, we assume zero initial capacitor charge (short across the capacitor terminals). The current $i(t)$ (Fig. 1.41a), in this case, divides equally between the two parallel branches because the circuit is balanced. Thus, the voltage across the capacitor continues to remain zero. Therefore, for the purpose of computing the current $i(t)$, the capacitor may be removed or replaced by a short. The resulting circuit is equivalent to that shown in Fig. 1.41b, which shows that the input $x(t)$ sees a load of 5 Ω, and

$$i(t) = \tfrac{1}{5}x(t)$$

Also, because $y(t) = 2\,i(t)$,

$$y(t) = \tfrac{2}{5}x(t) \tag{1.78}$$

This is the total response. Clearly, for the external description, the capacitor does not exist. No external measurement or external observation can detect the presence of the capacitor. Furthermore, if the circuit is enclosed inside a "black box" so that only the external terminals are accessible, it is impossible to determine the currents (or voltages) inside the circuit from external measurements or observations. An internal description, however, can provide every possible signal inside the system. In Example 1.15, we shall find the internal description of this system and show that it is capable of determining every possible signal in the system.

For most systems, the external and internal descriptions are equivalent, but there are a few exceptions, as in the present case, where the external description gives an inadequate picture of the systems. This happens when the system is *uncontrollable* and/or *unobservable*.

Figure 1.42 shows structural representations of simple uncontrollable and unobservable systems. In Fig. 1.42a, we note that part of the system (subsystem $\mathcal{S}_2$) inside the box cannot be controlled by the input $x(t)$. In Fig. 1.42b, some of the system outputs (those in subsystem $\mathcal{S}_2$) cannot be observed from the output terminals. If we try to describe either of these systems by applying an external input $x(t)$ and then measuring the output $y(t)$, the measurement will not

†The output voltage $y(t)$ resulting because of the capacitor charge [assuming $x(t) = 0$] is the zero-input response, which, as argued above, is zero. The output component due to the input $x(t)$ (assuming zero initial capacitor charge) is the zero-state response. Complete analysis of this problem is given later in Example 1.15.

characterize the complete system but only the part of the system (here $\mathcal{S}_1$) that is both controllable and observable (linked to both the input and output). Such systems are undesirable in practice and should be avoided in any system design. The system in Fig. 1.41a can be shown to be neither controllable nor observable. It can be represented structurally as a combination of the systems in Fig. 1.42a and 1.42b.

1.10 Internal Description: The State-Space Description

We shall now introduce the *state-space* description of a linear system, which is an internal description of a system. In this approach, we identify certain key variables, called the *state variables,* of the system. These variables have the property that every possible signal in the system can be expressed as a linear combination of these state variables. For example, we can show that every possible signal in a passive *RLC* circuit can be expressed as a linear combination of independent capacitor voltages and inductor currents, which, therefore, are state variables for the circuit.

To illustrate this point, consider the network in Fig. 1.43. We identify two state variables; the capacitor voltage q_1 and the inductor current q_2. If the values of q_1, q_2, and the input $x(t)$ are known at some instant t, we can demonstrate that every possible signal (current or voltage) in the circuit can be determined at t. For example, if $q_1 = 10$, $q_2 = 1$, and the input $x = 20$ at some instant, the remaining voltages and currents at that instant will be

$$
\begin{aligned}
i_1 &= (x - q_1)/1 = 20 - 10 = 10\text{ A} \\
v_1 &= x - q_1 = 20 - 10 = 10\text{ V} \\
v_2 &= q_1 = 10\text{ V} \\
i_2 &= q_1/2 = 5\text{ A} \\
i_C &= i_1 - i_2 - q_2 = 10 - 5 - 1 = 4\text{ A} \\
i_3 &= q_2 = 1\text{ A} \\
v_3 &= 5q_2 = 5\text{ V} \\
v_L &= q_1 - v_3 = 10 - 5 = 5\text{ V}
\end{aligned}
\tag{1.79}
$$

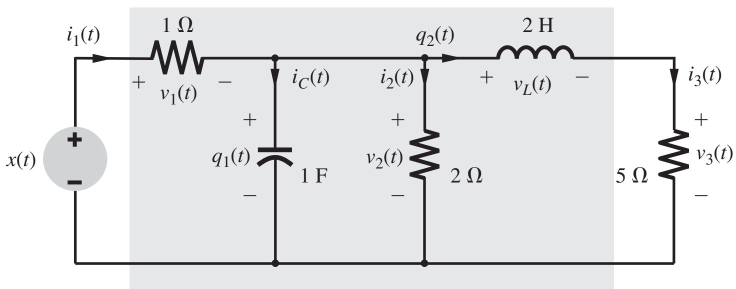


Figure 1.43 Choosing suitable initial conditions in a network.

Thus all signals in this circuit are determined. Clearly, state variables consist of the *key variables* in a system; a knowledge of the state variables allows one to determine every possible output of the system. Note that the *state-variable description is an internal description* of a system because it is capable of describing all possible signals in the system.

EXAMPLE 1.14

This example illustrates how state equations may be natural and easier to determine than other descriptions, such as loop or node equations. Consider again the network in Fig. 1.43 with q_1 and q_2 as the state variables and write the state equations.

This can be done by simple inspection of Fig. 1.43. Since $\dot{q}_1$ is the current through the capacitor,

$$\begin{aligned}\dot{q}_1 = i_C &= i_1 - i_2 - q_2 \\ &= (x - q_1) - 0.5\,q_1 - q_2 \\ &= -1.5\,q_1 - q_2 + x\end{aligned}$$

Also $2\dot{q}_2$, the voltage across the inductor, is given by

$$\begin{aligned}2\dot{q}_2 &= q_1 - v_3 \\ &= q_1 - 5q_2\end{aligned}$$

or

$$\dot{q}_2 = 0.5\,q_1 - 2.5\,q_2$$

Thus the state equations are

$$\begin{aligned}\dot{q}_1 &= -1.5\,q_1 - q_2 + x \\ \dot{q}_2 &= 0.5\,q_1 - 2.5\,q_2\end{aligned} \tag{1.80}$$

This is a set of two simultaneous first-order differential equations. This set of equations is known as the *state equations*. Once these equations have been solved for q_1 and q_2, everything else in the circuit can be determined by using Eqs. (1.79). The set of output equations (1.79) is called the *output equations*. Thus, in this approach, we have two sets of equations, the state equations and the output equations. Once we have solved the state equations, all possible outputs can be obtained from the output equations. In the input–output description, an Nth-order system is described by an Nth-order equation. In the state-variable approach, the same system is described by N simultaneous first-order state equations.†

† This assumes the system to be controllable and observable. If it is not, the input–output description equation will be of an order lower than the corresponding number of state equations.

EXAMPLE 1.15

In this example, we investigate the nature of state equations and the issue of controllability and observability for the circuit in Fig. 1.41a. This circuit has only one capacitor and no inductors. Hence, there is only one state variable, the capacitor voltage $q(t)$. Since $C = 1$ F, the capacitor current is $\dot{q}$. There are two sources in this circuit: the input $x(t)$ and the capacitor voltage $q(t)$. The response due to $x(t)$, assuming $q(t) = 0$, is the zero-state response, which can be found from Fig. 1.44a, where we have shorted the capacitor $[q(t) = 0]$. The response due to $q(t)$ assuming $x(t) = 0$, is the zero-input response, which can be found from Fig. 1.44b, where we have shorted $x(t)$ to ensure $x(t) = 0$. It is now trivial to find both the components.

Figure 1.44a shows zero-state currents in every branch. It is clear that the input $x(t)$ sees an effective resistance of 5 Ω, and, hence, the current through $x(t)$ is $x/5$ A, which divides in the two parallel branches resulting in the current $x/10$ through each branch.

Examining the circuit in Fig. 1.44b for the zero-input response, we note that the capacitor voltage is q and the current is $\dot{q}$. We also observe that the capacitor sees two loops in parallel, each with resistance 4 Ω and current $\dot{q}/2$. Interestingly, the 3 Ω branch is effectively shorted because the circuit is balanced, and thus the voltage across the terminals *cd* is zero. The total current in any branch is the sum of the currents in that branch in Fig. 1.44a and 1.44b (principle of superposition).

Branch	Current	Voltage
ca	$\frac{x}{10}+\frac{\dot{q}}{2}$	$2\left(\frac{x}{10}+\frac{\dot{q}}{2}\right)$
cb	$\frac{x}{10}-\frac{\dot{q}}{2}$	$2\left(\frac{x}{10}-\frac{\dot{q}}{2}\right)$
ad	$\frac{x}{10}-\frac{\dot{q}}{2}$	$2\left(\frac{x}{10}-\frac{\dot{q}}{2}\right)$
bd	$\frac{x}{10}+\frac{\dot{q}}{2}$	$2\left(\frac{x}{10}+\frac{\dot{q}}{2}\right)$
ec	$\frac{x}{5}$	$3\left(\frac{x}{5}\right)$
ed	$\frac{x}{5}$	x

(1.81)

To find the state equation, we note that the current in branch *ca* is $(x/10) + \dot{q}/2$ and the current in branch *cb* is $(x/10) - \dot{q}/2$. Hence, the equation around the loop *acba* is

$$q = 2\left[-\frac{x}{10} - \frac{\dot{q}}{2}\right] + 2\left[\frac{x}{10} - \frac{\dot{q}}{2}\right] = -2\dot{q}$$

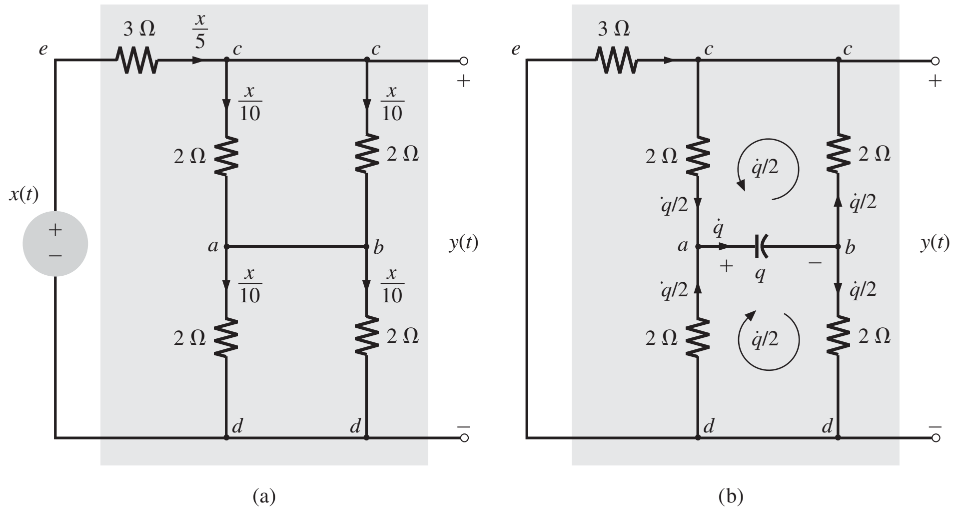


Figure 1.44 Analysis of a system that is neither controllable nor observable.

or

$$\dot{q} = -0.5q \tag{1.82}$$

This is the desired state equation.

Substitution of $\dot{q} = -0.5q$ in Eqs. (1.81) shows that every possible current and voltage in the circuit can be expressed in terms of the state variable q and the input x, as desired. Hence, the set of Eqs. (1.81) is the output equation for this circuit. Once we have solved the state equation (1.82) for q, we can determine every possible output in the circuit.

The output $y(t)$ is given by

$$\begin{aligned} y(t) &= 2\left[\frac{x}{10} - \frac{\dot{q}}{2}\right] + 2\left[\frac{x}{10} + \frac{\dot{q}}{2}\right] \\ &= \frac{2}{5}x(t) \end{aligned} \tag{1.83}$$

A little examination of the state and the output equations indicates the nature of this system. The state equation (1.82) shows that the state $q(t)$ is independent of the input $x(t)$, and hence, the system state q cannot be controlled by the input. Moreover, Eq. (1.83) shows that the output $y(t)$ does not depend on the state $q(t)$. Thus, the system state cannot be observed from the output terminals. Hence, the system is neither controllable nor observable. Such is not the case of other systems examined earlier. Consider, for example, the circuit in Fig. 1.43. The state equation (1.80) shows that the states are influenced by the input directly or indirectly. Hence, the system is controllable. Moreover, as the output Eqs. (1.79) show, every possible output is expressed in terms of the state variables and the input. Hence, the states are also observable.

State-space techniques are useful not just because of their ability to provide internal system description, but for several other reasons, including the following.

1. State equations of a system provide a mathematical model of great generality that can describe not just linear systems, but also nonlinear systems; not just time-invariant systems, but also time-varying parameter systems; not just SISO (single-input/single-output) systems, but also multiple-input/multiple-output (MIMO) systems. Indeed, state equations are ideally suited for the analysis, synthesis, and optimization of MIMO systems.
2. Compact matrix notation and the powerful techniques of linear algebra greatly facilitates complex manipulations. Without such features, many important results of the modern system theory would have been difficult to obtain. State equations can yield a great deal of information about a system even when they are not solved explicitly.
3. State equations lend themselves readily to digital computer simulation of complex systems of high order, with or without nonlinearities, and with multiple inputs and outputs.
4. For second-order systems ($N = 2$), a graphical method called *phase-plane analysis* can be used on state equations, whether they are linear or nonlinear.

The real benefits of the state-space approach, however, are realized for highly complex systems of large order. Much of the book is devoted to introduction of the basic concepts of linear systems analysis, which must necessarily begin with simpler systems without using the state-space approach. Chapter 10 deals with the state-space analysis of linear, time invariant, continuous-time, and discrete-time systems.

1.11 Summary

A *signal* is a set of data or information. A *system* processes input signals to modify them or extract additional information from them to produce output signals (response). A system may be made up of physical components (hardware realization), or it may be an algorithm that computes an output signal from an input signal (software realization).

A convenient measure of the size of a signal is its energy, if it is finite. If the signal energy is infinite, the appropriate measure is its power, if it exists. The signal power is the time average of its energy (averaged over the entire time interval from $-\infty$ to ∞). For periodic signals the time averaging need be performed only over one period in view of the periodic repetition of the signal. Signal power is also equal to the mean squared value of the signal (averaged over the entire time interval from $t = -\infty$ to ∞).

Signals can be classified in several ways.

1. A *continuous-time signal* is specified for a continuum of values of the independent variable (such as time t). A *discrete-time signal* is specified only at a finite or a countable set of time instants.
2. An *analog signal* is a signal whose amplitude can take on any value over a continuum. On the other hand, a signal whose amplitudes can take on only a finite number of values is a *digital signal*. The terms *discrete-time* and *continuous-time* qualify the nature of a signal along the time axis (horizontal axis). The terms *analog* and *digital,* on the other hand, qualify the nature of the signal amplitude (vertical axis).

3. A *periodic signal* $x(t)$ is defined by the fact that $x(t) = x(t + T_0)$ for some T_0. The smallest value of T_0 for which this relationship is satisfied is called the *fundamental period*. A periodic signal remains unchanged when shifted by an integer multiple of its period. A periodic signal $x(t)$ can be generated by a periodic extension of any contiguous segment of $x(t)$ of duration T_0. Finally, a periodic signal, by definition, must exist over the entire time interval $-\infty < t < \infty$. A signal is *aperiodic* if it is not periodic.

   An *everlasting signal* starts at $t = -\infty$ and continues forever to $t = \infty$. Hence, periodic signals are everlasting signals. A *causal signal* is a signal that is zero for $t < 0$.

4. A signal with finite energy is an *energy signal*. Similarly a signal with a finite and nonzero power (mean square value) is a *power signal*. A signal can be either an energy signal or a power signal, but not both. However, there are signals that are neither energy nor power signals.

5. A signal whose physical description is known completely in a mathematical or graphical form is a *deterministic signal*. A *random signal* is known only in terms of its probabilistic description such as mean value or mean-square value, rather than by its mathematical or graphical form.

A signal $x(t)$ delayed by T seconds (right-shifted) can be expressed as $x(t - T)$; on the other hand, $x(t)$ advanced by T (left-shifted) is $x(t + T)$. A signal $x(t)$ time-compressed by a factor a ($a > 1$) is expressed as $x(at)$; on the other hand, the same signal time-expanded by factor a ($a > 1$) is $x(t/a)$. The signal $x(t)$ when time reversed can be expressed as $x(-t)$.

The unit step function $u(t)$ is very useful in representing causal signals and signals with different mathematical descriptions over different intervals.

In the classical (Dirac) definition, the unit impulse function $\delta(t)$ is characterized by unit area and is concentrated at a single instant $t = 0$. The impulse function has a sampling (or sifting) property, which states that the area under the product of a function with a unit impulse is equal to the value of that function at the instant at which the impulse is located (assuming the function to be continuous at the impulse location). In the modern approach, the impulse function is viewed as a generalized function and is defined by the sampling property.

The exponential function e^{st}, where s is complex, encompasses a large class of signals that includes a constant, a monotonic exponential, a sinusoid, and an exponentially varying sinusoid.

A real signal that is symmetrical about the vertical axis ($t = 0$) is an *even* function of time, and a real signal that is antisymmetrical about the vertical axis is an *odd* function of time. The product of an even function and an odd function is an odd function. However, the product of an even function and an even function or an odd function and an odd function is an even function. The area under an odd function from $t = -a$ to a is always zero regardless of the value of a. On the other hand, the area under an even function from $t = -a$ to a is two times the area under the same function from $t = 0$ to a (or from $t = -a$ to 0). Every signal can be expressed as a sum of odd and even function of time.

A system processes input signals to produce output signals (response). The input is the cause, and the output is its effect. In general, the output is affected by two causes: the internal conditions of the system (such as the initial conditions) and the external input.

Systems can be classified in several ways.

1. Linear systems are characterized by the linearity property, which implies superposition; if several causes (such as various inputs and initial conditions) are acting on a linear system, the total output (response) is the sum of the responses from each cause, assuming that all the remaining causes are absent. A system is nonlinear if superposition does not hold.
2. In time-invariant systems, system parameters do not change with time. The parameters of time-varying-parameter systems change with time.
3. For memoryless (or instantaneous) systems, the system response at any instant t depends only on the value of the input at t. For systems with memory (also known as dynamic systems), the system response at any instant t depends not only on the present value of the input, but also on the past values of the input (values before t).
4. In contrast, if a system response at t also depends on the future values of the input (values of input beyond t), the system is noncausal. In causal systems, the response does not depend on the future values of the input. Because of the dependence of the response on the future values of input, the effect (response) of noncausal systems occurs before the cause. When the independent variable is time (temporal systems), the noncausal systems are prophetic systems, and therefore, unrealizable, although close approximation is possible with some time delay in the response. Noncausal systems with independent variables other than time (e.g., space) are realizable.
5. Systems whose inputs and outputs are continuous-time signals are continuous-time systems; systems whose inputs and outputs are discrete-time signals are discrete-time systems. If a continuous-time signal is sampled, the resulting signal is a discrete-time signal. We can process a continuous-time signal by processing the samples of the signal with a discrete-time system.
6. Systems whose inputs and outputs are analog signals are analog systems; those whose inputs and outputs are digital signals are digital systems.
7. If we can obtain the input $x(t)$ back from the output $y(t)$ of a system $\mathcal{S}$ by some operation, the system $\mathcal{S}$ is said to be invertible. Otherwise the system is noninvertible.
8. A system is stable if bounded input produces bounded output. This defines the external stability because it can be ascertained from measurements at the external terminals of the system. The external stability is also known as the stability in the BIBO (bounded-input/bounded-output) sense. The internal stability, discussed later in Chapter 2, is measured in terms of the internal behavior of the system.

The system model derived from a knowledge of the internal structure of the system is its internal description. In contrast, an external description is a representation of a system as seen from its input and output terminals; it can be obtained by applying a known input and measuring the resulting output. In the majority of practical systems, an external description of a system so obtained is equivalent to its internal description. At times, however, the external description fails to describe the system adequately. Such is the case with the so-called uncontrollable or unobservable systems.

A system may also be described in terms of certain set of key variables called state variables. In this description, an Nth-order system can be characterized by a set of N simultaneous first-order differential equations in N state variables. State equations of a system represent an internal description of that system.

REFERENCES

1. Papoulis, A. *The Fourier Integral and Its Applications*. McGraw-Hill, New York, 1962.
2. Mason, S. J. *Electronic Circuits, Signals, and Systems*. Wiley, New York, 1960.
3. Kailath, T. *Linear Systems*. Prentice-Hall, Englewood Cliffs, NJ, 1980.
4. Lathi, B. P. *Signals and Systems*. Berkeley-Cambridge Press, Carmichael, CA, 1987.

MATLAB SESSION 1: WORKING WITH FUNCTIONS

Working with functions is fundamental to signals and systems applications. MATLAB provides several methods of defining and evaluating functions. An understanding and proficient use of these methods is therefore necessary and beneficial.

M1.1 Inline Functions

Many simple functions are most conveniently represented by using MATLAB inline objects. An inline object provides a symbolic representation of a function defined in terms of MATLAB operators and functions. For example, consider defining the exponentially damped sinusoid $f(t) = e^{-t}\cos(2\pi t)$.

```
>> f = inline('exp(-t).*cos(2*pi*t)','t')
f =  Inline function:
      f(t) = exp(-t).*cos(2*pi*t)
```

The second argument to the `inline` command identifies the function's input argument as `t`. Input arguments, such as `t`, are local to the inline object and are not related to any workspace variables with the same names.

Once defined, $f(t)$ can be evaluated simply by passing the input values of interest. For example,

```
>> t = 0;
>> f(t)
ans = 1
```

evaluates $f(t)$ at $t = 0$, confirming the expected result of unity. The same result is obtained by passing $t = 0$ directly.

```
>> f(0)
ans = 1
```

Vector inputs allow the evaluation of multiple values simultaneously. Consider the task of plotting $f(t)$ over the interval $(-2 \leq t \leq 2)$. Gross function behavior is clear: $f(t)$ should oscillate four times with a decaying envelope. Since accurate hand sketches are cumbersome, MATLAB-generated plots are an attractive alternative. As the following example illustrates, care must be taken to ensure reliable results.

Suppose vector t is chosen to include only the integers contained in $(-2 \leq t \leq 2)$, namely $[-2, -1, 0, 1, 2]$.

```
>> t = (-2:2);
```

This vector input is evaluated to form a vector output.

```
>> f(t)
ans = 7.3891    2.7183    1.0000    0.3679    0.1353
```

The plot command graphs the result, which is shown in Fig. M1.1.

```
>> plot(t,f(t));
>> xlabel('t'); ylabel('f(t)'); grid;
```

Grid lines, added by using the grid command, aid feature identification. Unfortunately, the plot does not illustrate the expected oscillatory behavior. More points are required to adequately represent $f(t)$.

The question, then, is how many points is enough?† If too few points are chosen, information is lost. If too many points are chosen, memory and time are wasted. A balance is needed. For oscillatory functions, plotting 20 to 200 points per oscillation is normally adequate. For the present case, t is chosen to give 100 points per oscillation.

```
>> t = (-2:0.01:2);
```

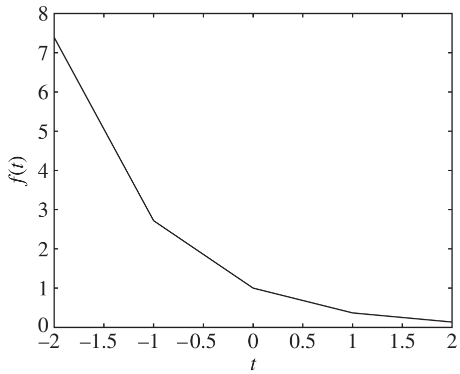


Figure M1.1 $f(t) = e^{-t}\cos(2\pi t)$ for t = (-2:2).

†Sampling theory, presented later, formally addresses important aspects of this question.

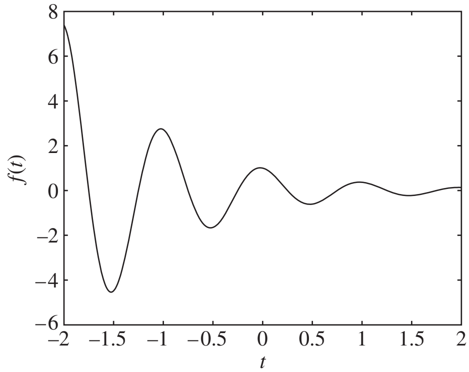


Figure M1.2 $f(t) = e^{-t}\cos(2\pi t)$ for `t = (-2:0.01:2)`.

Again, the function is evaluated and plotted.

```
>> plot(t,f(t));
>> xlabel('t'); ylabel('f(t)'); grid;
```

The result, shown in Fig. M1.2, is an accurate depiction of $f(t)$.

M1.2 Relational Operators and the Unit Step Function

The unit step function $u(t)$ arises naturally in many practical situations. For example, a unit step can model the act of turning on a system. With the help of relational operators, inline objects can represent the unit step function.

In MATLAB, a relational operator compares two items. If the comparison is true, a logical true (`1`) is returned. If the comparison is false, a logical false (`0`) is returned. Sometimes called indicator functions, relational operators indicates whether a condition is true. Six relational operators are available: <, >, <=, >=, ==, and ~=.

The unit step function is readily defined using the >= relational operator.

```
>> u = inline('(t>=0)','t')
u =  Inline function:
     u(t) = (t>=0)
```

Any function with a jump discontinuity, such as the unit step, is difficult to plot. Consider plotting $u(t)$ by using `t = (-2:2)`.

```
>> t = (-2:2);
>> plot(t,u(t));
>> xlabel('t'); ylabel('u(t)');
```

Two significant problems are apparent in the resulting plot, shown in Fig. M1.3. First, MATLAB automatically scales plot axes to tightly bound the data. In this case, this normally desirable feature obscures most of the plot. Second, MATLAB connects plot data with lines,

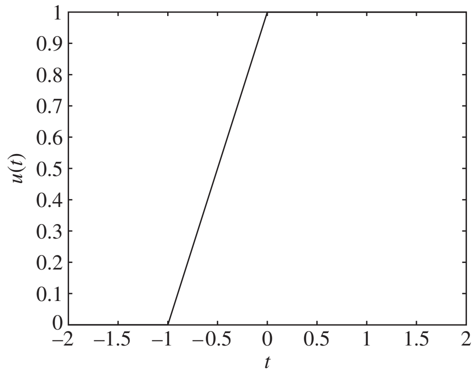


Figure M1.3 $u(t)$ for `t = (-2:2).`

making a true jump discontinuity difficult to achieve. The coarse resolution of vector `t` emphasizes the effect by showing an erroneous sloping line between $t = -1$ and $t = 0$.

The first problem is corrected by vertically enlarging the bounding box with the `axis` command. The second problem is reduced, but not eliminated, by adding points to vector `t`.

```
>> t = (-2:0.01:2);
>> plot(t,u(t));
>> xlabel('t'); ylabel('u(t)');
>> axis([-2 2 -0.1 1.1]);
```

The four-element vector argument of `axis` specifies x-axis minimum, x-axis maximum, y-axis minimum, and y-axis maximum, respectively. The improved results are shown in Fig. M1.4.

Relational operators can be combined using logical AND, logical OR, and logical negation: `&`, `|`, and `~`, respectively. For example, `(t>0)&(t<1)` and `~((t<=0)|(t>=1))` both test if

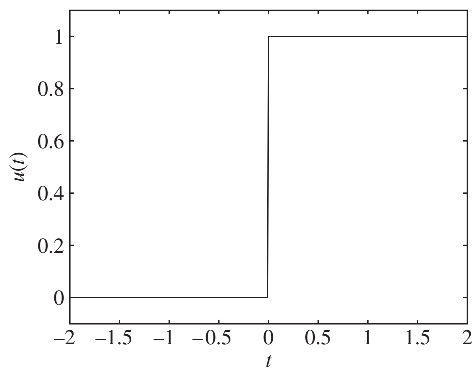


Figure M1.4 $u(t)$ for `t = (-2:0.01:2)` with axis modification.

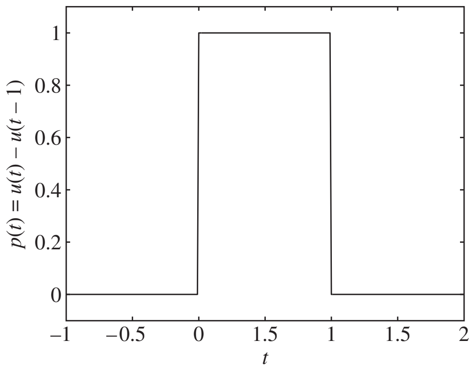


Figure M1.5 $p(t) = u(t) - u(t-1)$ over $(-1 \leq t \leq 2)$.

$0 < t < 1$. To demonstrate, consider defining and plotting the unit pulse $p(t) = u(t) - u(t-1)$, as shown in Fig. M1.5:

```
>> p = inline('(t>=0)&(t<1)','t');
>> t = (-1:0.01:2); plot(t,p(t));
>> xlabel('t'); ylabel('p(t) = u(t)-u(t-1)');
>> axis([-1 2 -.1 1.1]);
```

For scalar operands, MATLAB also supports two short-circuit logical constructs. A short-circuit logical AND is performed by using `&&`, and a short-circuit logical OR is performed by using `||`. Short-circuit logical operators are often more efficient than traditional logical operators because they test the second portion of the expression only when necessary. That is, when scalar expression `A` is found false in `(A&&B)`, scalar expression `B` is not evaluated, since a false result is already guaranteed. Similarly, scalar expression `B` is not evaluated when scalar expression `A` is found true in `(A||B)`, since a true result is already guaranteed.

M1.3 Visualizing Operations on the Independent Variable

Two operations on a function's independent variable are commonly encountered: shifting and scaling. Inline objects are well suited to investigate both operations.

Consider $g(t) = f(t)u(t) = e^{-t}\cos(2\pi t)u(t)$, a realizable version of $f(t)$.† Unfortunately, MATLAB cannot multiply inline objects. That is, MATLAB reports an error for `g = f*u` when `f` and `u` are inline objects. Rather, $g(t)$ needs to be explicitly defined.

```
>> g = inline('exp(-t).*cos(2*pi*t).*(t>=0)','t')
g =  Inline function:
     g(t) = exp(-t).*cos(2*pi*t).*(t>=0)
```

A combined shifting and scaling operation is represented by $g(at + b)$, where a and b are arbitrary real constants. As an example, consider plotting $g(2t + 1)$ over $(-2 \leq t \leq 2)$. With

†The function $f(t) = e^{-t}\cos(2\pi t)$ can never be realized in practice; it has infinite duration and, as $t \to -\infty$, infinite magnitude.

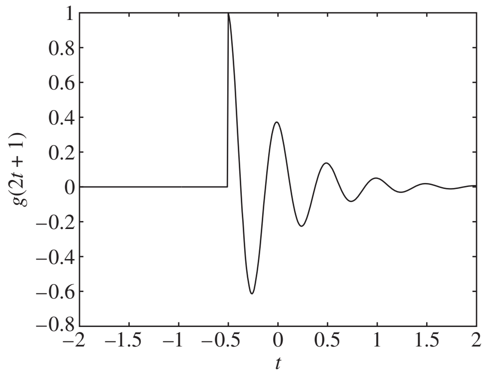


Figure M1.6 $g(2t+1)$ over $(-2 \leq t \leq 2)$.

$a = 2$, the function is compressed by a factor of 2, resulting in twice the oscillations per unit t. Adding the condition $b > 0$, the waveform shifts to the left. Given inline function `g`, an accurate plot is nearly trivial to obtain.

```
>>t = (-2:0.01:2);
>> plot(t,g(2*t+1)); xlabel('t'); ylabel('g(2t+1)'); grid;
```

Figure M1.6 confirms the expected waveform compression and left shift. As a final check, realize that function $g(\cdot)$ turns on when the input argument is zero. Therefore, $g(2t+1)$ should turn on when $2t + 1 = 0$ or at $t = -0.5$, a fact again confirmed by Fig. M1.6.

Next, consider plotting $g(-t+1)$ over $(-2 \leq t \leq 2)$. Since $a < 0$, the waveform will be reflected. Adding the condition $b > 0$, the final waveform shifts to the right.

```
>> plot(t,g(-t+1)); xlabel('t'); ylabel('g(-t+1)'); grid;
```

Figure M1.7 confirms both the reflection and the right shift.

Up to this point, Figs. M1.6 and M1.7 could be reasonably sketched by hand. Consider plotting the more complicated function $h(t) = g(2t+1) + g(-t+1)$ over $(-2 \leq t \leq 2)$ (Fig. M1.8). In this case, an accurate hand sketch is quite difficult. With MATLAB, the work is much less burdensome.

```
>> plot(t,g(2*t+1)+g(-t+1)); xlabel('t'); ylabel('h(t)'); grid;
```

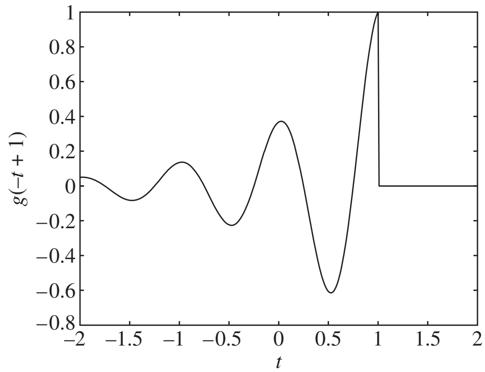


Figure M1.7 $g(-t+1)$ over $(-2 \leq t \leq 2)$.

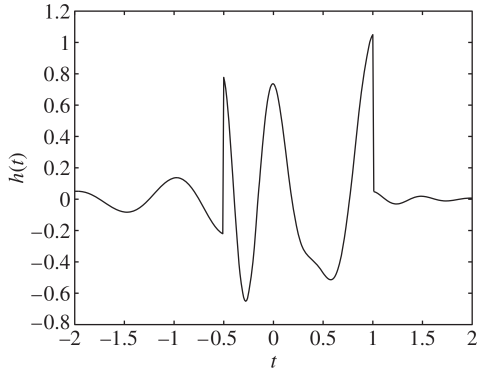


Figure M1.8 $h(t) = g(2t+1) + g(-t+1)$ over $(-2 \leq t \leq 2)$.

M1.4 Numerical Integration and Estimating Signal Energy

Interesting signals often have nontrivial mathematical representations. Computing signal energy, which involves integrating the square of these expressions, can be a daunting task. Fortunately, many difficult integrals can be accurately estimated by means of numerical integration techniques. Even if the integration appears simple, numerical integration provides a good way to verify analytical results.

To start, consider the simple signal $x(t) = e^{-t}(u(t) - u(t-1))$. The energy of $x(t)$ is expressed as $E_x = \int_{-\infty}^{\infty} |x(t)|^2\, dt = \int_0^1 e^{-2t}\, dt$. Integrating yields $E_x = 0.5(1 - e^{-2}) \approx 0.4323$. The energy integral can also be evaluated numerically. Figure 1.27 helps illustrate the simple method of rectangular approximation: evaluate the integrand at points uniformly separated by Δt, multiply each by Δt to compute rectangle areas, and then sum over all rectangles. First, we create function $x(t)$.

```
>> x = inline('exp(-t).*((t>=0)&(t<1))','t');
```

Letting $\Delta t = 0.01$, a suitable time vector is created.

```
>> t = (0:0.01:1);
```

The final result is computed by using the `sum` command.

```
>> E_x = sum(x(t).*x(t)*0.01)
E_x = 0.4367
```

The result is not perfect, but at 1% relative error it is close. By reducing Δt, the approximation is improved. For example, $\Delta t = 0.001$ yields `E_x = 0.4328`, or 0.1% relative error.

Although simple to visualize, rectangular approximation is not the best numerical integration technique. The MATLAB function `quad` implements a better numerical integration technique called recursive adaptive Simpson quadrature.† To operate, `quad` requires a function describing

† A comprehensive treatment of numerical integration is outside the scope of this text. Details of this particular method are not important for the current discussion; it is sufficient to say that it is better than the rectangular approximation.

the integrand, the lower limit of integration, and the upper limit of integration. Notice, that no Δt needs to be specified.

To use quad to estimate E_x, the integrand must first be described.

```
>> x_squared = inline('exp(-2*t).*((t>=0)&(t<1))','t');
```

Estimating E_x immediately follows.

```
>> E_x = quad(x_squared,0,1)
E_x = 0.4323
```

In this case, the relative error is -0.0026%.

The same techniques can be used to estimate the energy of more complex signals. Consider $g(t)$, defined previously. Energy is expressed as $E_g = \int_0^\infty e^{-2t}\cos^2(2\pi t)\,dt$. A closed-form solution exists, but it takes some effort. MATLAB provides an answer more quickly.

```
>> g_squared = inline('exp(-2*t).*(cos(2*pi*t).^2).*(t>=0)','t');
```

Although the upper limit of integration is infinity, the exponentially decaying envelope ensures $g(t)$ is effectively zero well before $t = 100$. Thus, an upper limit of $t = 100$ is used along with $\Delta t = 0.001$.

```
>> t = (0:0.001:100);
>> E_g = sum(g_squared(t)*0.001)
E_g = 0.2567
```

A slightly better approximation is obtained with the quad function.

```
>> E_g = quad(g_squared,0,100)
E_g = 0.2562
```

As an exercise, confirm that the energy of signal $h(t)$, defined previously, is $E_h = 0.3768$.

PROBLEMS

1-1 Find the energies of the signals illustrated in Fig. P1-1. Comment on the effect on energy of sign change, time shifting, or doubling of the signal. What is the effect on the energy if the signal is multiplied by k?

1-2 Repeat Prob. 1-1 for the signals in Fig. P1-2.

1-3 (a) Find the energies of the pair of signals $x(t)$ and $y(t)$ depicted in Fig. P1-3a and P1-3b. Sketch and find the energies of signals $x(t)+y(t)$ and $x(t)-y(t)$. Can you make any observation from these results?

(b) Repeat part (a) for the signal pair illustrated in Fig. P1-3c. Is your observation in part (a) still valid?

1-4 Find the power of the periodic signal $x(t)$ shown in Fig. P1-4. Find also the powers and the rms values of:

(a) $-x(t)$
(b) $2x(t)$
(c) $cx(t)$.

Comment.

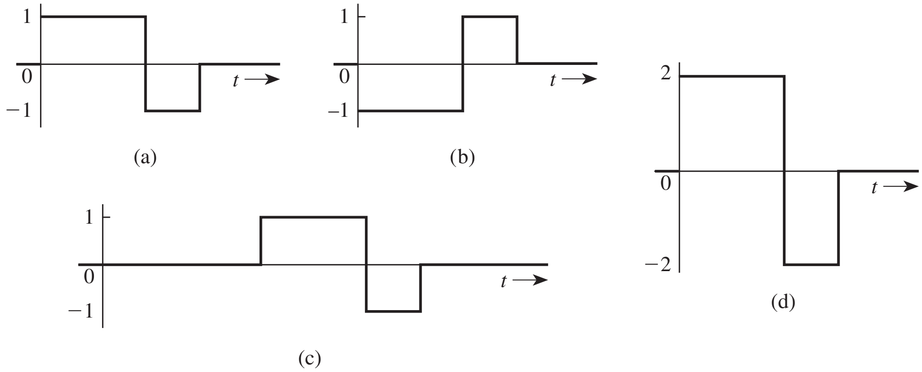


Figure P1-1

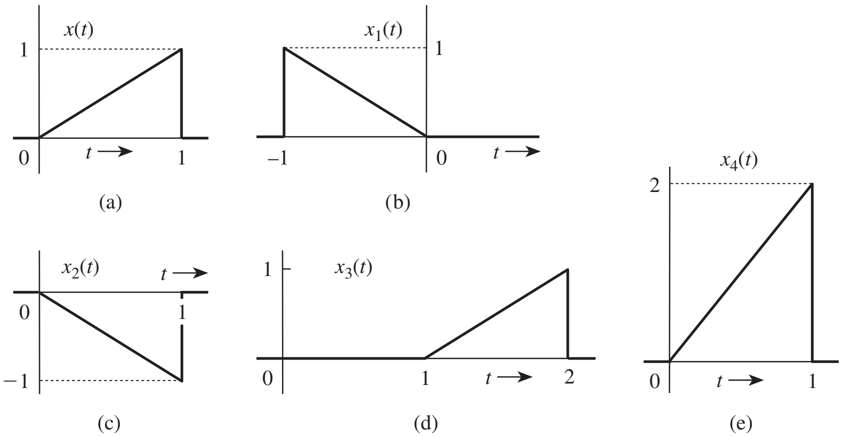


Figure P1-2

1-5 Figure P1-5 shows a periodic 50% duty cycle dc-offset sawtooth wave $x(t)$ with peak amplitude A. Determine the energy and power of $x(t)$.

1-6 (a) There are many useful properties related to signal energy. Prove each of the following statements. In each case, let energy signal $x_1(t)$ have energy $E[x_1(t)]$, let energy signal $x_2(t)$ have energy $E[x_2(t)]$, and let T be a nonzero, finite, real-valued constant.

(i) Prove $E[Tx_1(t)] = T^2 E[x_1(t)]$. That is, amplitude scaling a signal by constant T scales the signal energy by T^2.

(ii) Prove $E[x_1(t)] = E[x_1(t-T)]$. That is, shifting a signal does not affect its energy.

(iii) If $(x_1(t) \neq 0) \Rightarrow (x_2(t) = 0)$ and $(x_2(t) \neq 0) \Rightarrow (x_1(t) = 0)$, then prove $E[x_1(t) + x_2(t)] = E[x_1(t)] + E[x_2(t)]$. That is, the energy of the sum of two nonoverlapping signals is the sum of the two individual energies.

(iv) Prove $E[x_1(Tt)] = (1/|T|)E[x_1(t)]$. That is, time-scaling a signal by T reciprocally scales the signal energy by $1/|T|$.

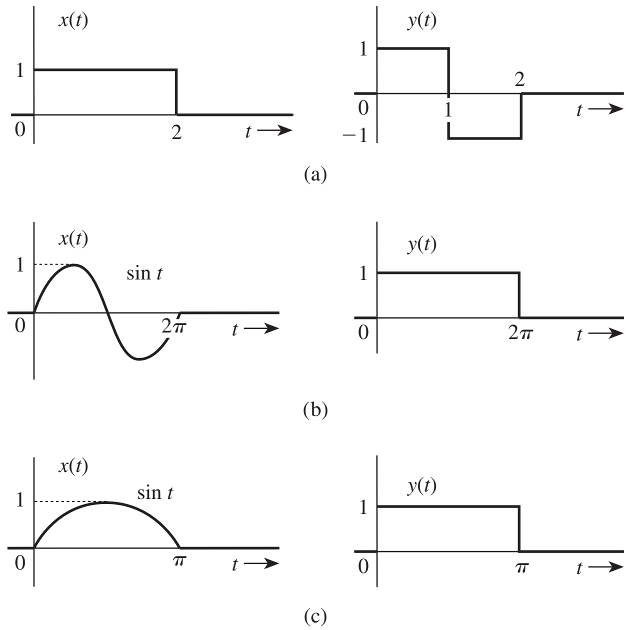


Figure P1-3

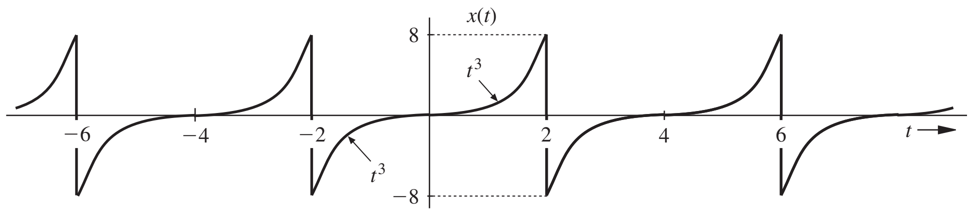


Figure P1-4

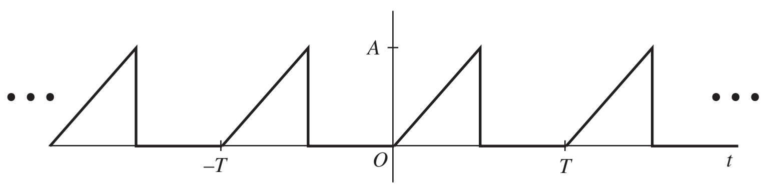


Figure P1-5 50% duty cycle dc-offset sawtooth wave $x(t)$.

(b) Consider the signal $x(t)$ shown in Fig. P1-6. Outside the interval shown, $x(t)$ is zero. Determine the signal energy $E[x(t)]$.

1-7 A binary signal $x(t) = 0$ for $t < 0$. For positive time, $x(t)$ toggles between one and zero as follows: one for 1 second, zero for 1 second, one for 1 second, zero for 2 seconds, one for 1

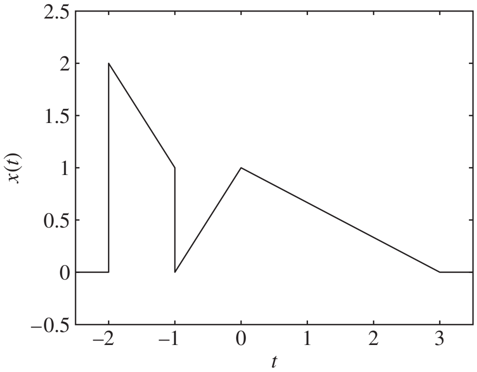


Figure P1-6 Energy signal $x(t)$.

second, zero for 3 seconds, and so forth. That is, the "on" time is always one second but the "off" time successively increases by one second between each toggle. A portion of $x(t)$ is shown in Fig. P1-7. Determine the energy and power of $x(t)$.

1-8 For the signal $x(t)$ depicted in Fig. P1-8, sketch the signals

(a) $x(-t)$

(b) $x(t+6)$

(c) $x(3t)$

(d) $x(t/2)$

1-9 For the signal $x(t)$ illustrated in Fig. P1-9, sketch

(a) $x(t-4)$

(b) $x(t/1.5)$

(c) $x(-t)$

(d) $x(2t-4)$

(e) $x(2-t)$

1-10 In Fig. P1-10, express signals $x_1(t), x_2(t), x_3(t), x_4(t)$, and $x_5(t)$ in terms of signal $x(t)$ and its time-shifted, time-scaled, or time-reversed versions.

1-11 For an energy signal $x(t)$ with energy E_x, show that the energy of any one of the signals $-x(t)$, $x(-t)$, and $x(t-T)$ is E_x. Show also that the energy of $x(at)$ as well as $x(at-b)$ is E_x/a, but the energy of $ax(t)$ is a^2E_x. This shows that time inversion and time shifting do not affect signal energy. On the other hand, time compression of a signal ($a > 1$) reduces the energy, and time expansion of a signal ($a < 1$) increases the energy. What is the

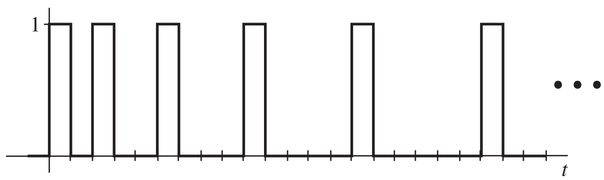


Figure P1-7 Binary signal $x(t)$.

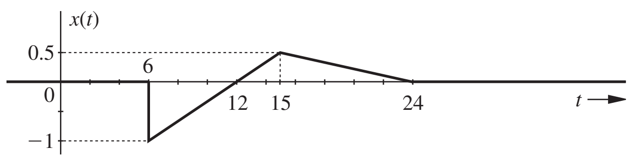


Figure P1-8

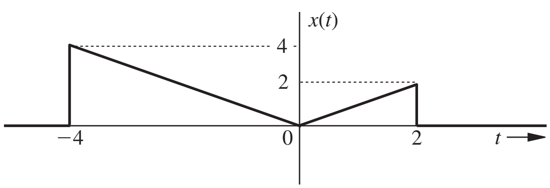


Figure P1-9

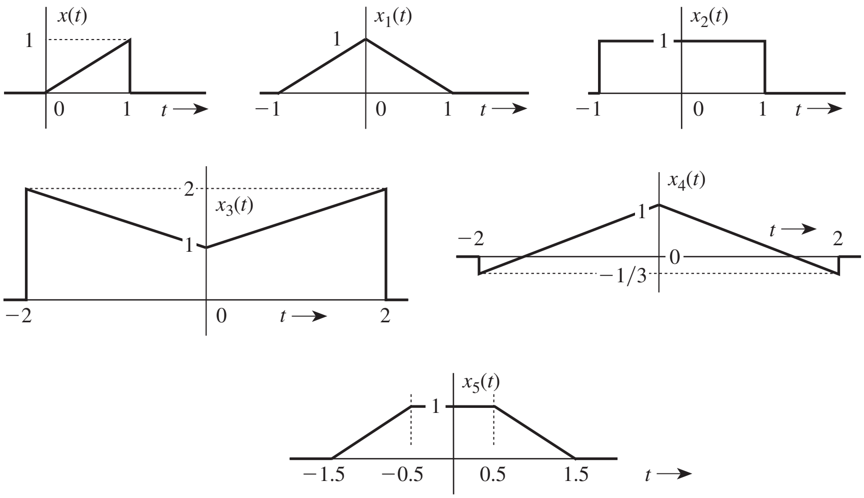


Figure P1-10

effect on signal energy if the signal is multiplied by a constant a?

1-12 Define $2x(-3t+1) = t(u(-t-1) - u(-t+1))$, where $u(t)$ is the unit step function.

(a) Plot $2x(-3t+1)$ over a suitable range of t.
(b) Plot $x(t)$ over a suitable range of t.

1-13 Consider the signal $x(t) = 2^{-tu(t)}$, where $u(t)$ is the unit step function.

(a) Accurately sketch $x(t)$ over $(-1 \leq t \leq 1)$.
(b) Accurately sketch $y(t) = 0.5x(1-2t)$ over $(-1 \leq t \leq 1)$.

1-14 Determine whether each of the following statements is true or false. If the statement is false, demonstrate this by proof or example.

(a) Every continuous-time signal is analog signal.
(b) Every discrete-time signal is digital signal.
(c) If a signal is not an energy signal, then it must be a power signal and vice versa.
(d) An energy signal must be of finite duration.
(e) A power signal cannot be causal.
(f) A periodic signal cannot be anticausal.

1-15 Determine whether each of the following statements is true or false. If the statement is false, demonstrate by proof or example why the statement is false.

(a) Every bounded periodic signal is a power signal.
(b) Every bounded power signal is a periodic signal.
(c) If an energy signal $x(t)$ has energy E, then the energy of $x(at)$ is E/a. Assume a is real and positive.
(d) If a power signal $x(t)$ has power P, then the power of $x(at)$ is P/a. Assume a is real and positive.

1-16 Given $x_1(t) = \cos(t)$, $x_2(t) = \sin(\pi t)$, and $x_3(t) = x_1(t) + x_2(t)$.

(a) Determine the fundamental periods T_1 and T_2 of signals $x_1(t)$ and $x_2(t)$.
(b) Show that $x_3(t)$ is not periodic, which requires $T_3 = k_1 T_1 = k_2 T_2$ for some integers k_1 and k_2.
(c) Determine the powers P_{x_1}, P_{x_2}, and P_{x_3} of signals $x_1(t)$, $x_2(t)$, and $x_3(t)$.

1-17 For any constant ω, is the function $f(t) = \sin(\omega t)$ a periodic function of the independent variable t? Justify your answer.

1-18 The signal shown in Fig. P1-18 is defined as

$$x(t) = \begin{cases} t & 0 \leq t < 1 \\ 0.5 + 0.5\cos(2\pi t) & 1 \leq t < 2 \\ 3 - t & 2 \leq t < 3 \\ 0 & \text{otherwise} \end{cases}$$

The energy of $x(t)$ is $E \approx 1.0417$.

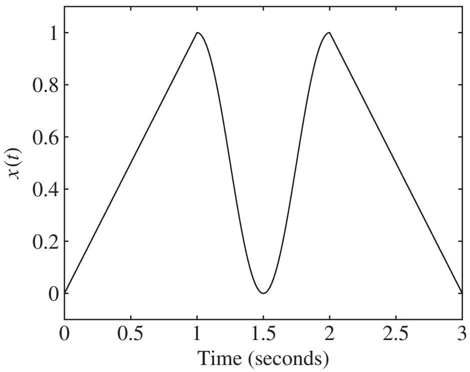


Figure P1-18 Energy signal $x(t)$.

(a) What is the energy of $y_1(t) = (1/3)x(2t)$?

(b) A periodic signal $y_2(t)$ is defined as

$$y_2(t) = \begin{cases} x(t) & 0 \le t < 4 \\ y_2(t+4) & \forall t \end{cases}$$

What is the power of $y_2(t)$?

(c) What is the power of $y_3(t) = (1/3)y_2(2t)$?

1-19 Let $y_1(t) = y_2(t) = t^2$ over $0 \le t \le 1$. Notice, this statement does not require $y_1(t) = y_2(t)$ for all t.

(a) Define $y_1(t)$ as an even, periodic signal with period $T_1 = 2$. Sketch $y_1(t)$ and determine its power.

(b) Design an odd, periodic signal $y_2(t)$ with period $T_2 = 3$ and power equal to unity. Fully describe $y_2(t)$ and sketch the signal over at least one full period. [Hint: There are an infinite number of possible solutions to this problem—you need to find only one of them!]

(c) We can create a complex-valued function $y_3(t) = y_1(t) + jy_2(t)$. Determine whether this signal is periodic. If yes, determine the period T_3. If no, justify why the signal is not periodic.

(d) Determine the power of $y_3(t)$ defined in part (c). The power of a complex-valued function $z(t)$ is

$$P = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} z(\tau)z^*(\tau)\, d\tau$$

1-20 Sketch the following signal:

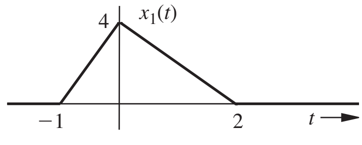
(a) $u(t-5) - u(t-7)$

(b) $u(t-5) + u(t-7)$

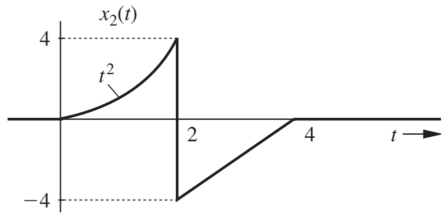
(c) $t^2[u(t-1) - u(t-2)]$

(d) $(t-4)[u(t-2) - u(t-4)]$

1-21 Express each of the signals in Fig. P1-21 by a single expression valid for all t.



(a)



(b)

Figure P1-21

1-22 Simplify the following expressions:

(a) $\left(\dfrac{\sin t}{t^2+2}\right)\delta(t)$

(b) $\left(\dfrac{j\omega+2}{\omega^2+9}\right)\delta(\omega)$

(c) $[e^{-t}\cos(3t-60^\circ)]\delta(t)$

(d) $\left(\dfrac{\sin\left[\frac{\pi}{2}(t-2)\right]}{t^2+4}\right)\delta(1-t)$

(e) $\left(\dfrac{1}{j\omega+2}\right)\delta(\omega+3)$

(f) $\left(\dfrac{\sin k\omega}{\omega}\right)\delta(\omega)$

[Hint: Use Eq. (1.23). For part (f) use L'Hôpital's rule.]

1-23 Evaluate the following integrals:

(a) $\displaystyle\int_{-\infty}^{\infty} \delta(\tau)x(t-\tau)\, d\tau$

(b) $\displaystyle\int_{-\infty}^{\infty} x(\tau)\delta(t-\tau)\, d\tau$

(c) $\int_{-\infty}^{\infty} \delta(t)e^{-j\omega t}\,dt$

(d) $\int_{-\infty}^{\infty} \delta(2t-3)\sin \pi t\,dt$

(e) $\int_{-\infty}^{\infty} \delta(t+3)e^{-t}\,dt$

(f) $\int_{-\infty}^{\infty} (t^3+4)\delta(1-t)\,dt$

(g) $\int_{-\infty}^{\infty} x(2-t)\delta(3-t)\,dt$

(h) $\int_{-\infty}^{\infty} e^{(x-1)}\cos\left[\frac{\pi}{2}(x-5)\right]\delta(x-3)\,dx$

1-24 (a) Find and sketch dx/dt for the signal $x(t)$ shown in Fig. P1-9.
(b) Find and sketch d^2x/dt^2 for the signal $x(t)$ depicted in Fig. P1-21a.

1-25 Find and sketch $\int_{-\infty}^{t} x(t)\,dt$ for the signal $x(t)$ illustrated in Fig. P1-25.

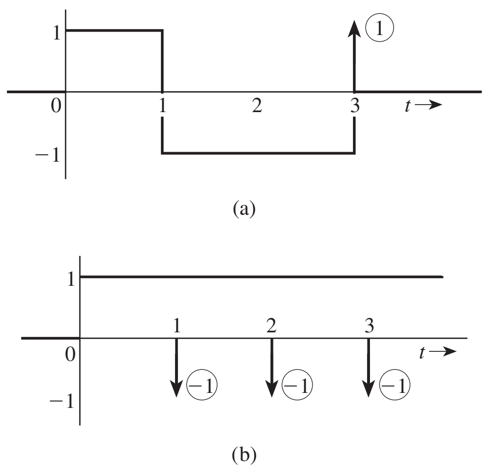


Figure P1-25

1-26 Using the generalized function definition of impulse [Eq. (1.24a)], show that $\delta(t)$ is an even function of t.

1-27 Using the generalized function definition of impulse [Eq. (1.24a)], show that

$$\delta(at) = \frac{1}{|a|}\delta(t)$$

1-28 Show that

$$\int_{-\infty}^{\infty} \dot{\delta}(t)\phi(t)\,dt = -\dot{\phi}(0)$$

where $\phi(t)$ and $\dot{\phi}(t)$ are continuous at $t=0$, and $\phi(t) \to 0$ as $t \to \pm\infty$. This integral defines $\dot{\delta}(t)$ as a generalized function. [Hint: Use integration by parts.]

1-29 A sinusoid $e^{\sigma t}\cos \omega t$ can be expressed as a sum of exponentials e^{st} and e^{-st} [Eq. (1.30c)] with complex frequencies $s = \sigma + j\omega$ and $s = \sigma - j\omega$. Locate in the complex plane the frequencies of the following sinusoids:
(a) $\cos 3t$
(b) $e^{-3t}\cos 3t$
(c) $e^{2t}\cos 3t$
(d) e^{-2t}
(e) e^{2t}
(f) 5

1-30 Find and sketch the odd and the even components of the following:
(a) $u(t)$
(b) $tu(t)$
(c) $\sin \omega_0 t$
(d) $\cos \omega_0 t$
(e) $\cos(\omega_0 t + \theta)$
(f) $\sin \omega_0 t u(t)$
(g) $\cos \omega_0 t u(t)$

1-31 (a) Determine even and odd components of the signal $x(t) = e^{-2t}u(t)$.
(b) Show that the energy of $x(t)$ is the sum of energies of its odd and even components found in part (a).
(c) Generalize the result in part (b) for any finite energy signal.

1-32 (a) If $x_e(t)$ and $x_o(t)$ are even and the odd components of a real signal $x(t)$, then show that

$$\int_{-\infty}^{\infty} x_e(t)x_o(t)\,dt = 0$$

(b) Show that

$$\int_{-\infty}^{\infty} x(t)\,dt = \int_{-\infty}^{\infty} x_e(t)\,dt$$

1-33 An aperiodic signal is defined as $x(t) = \sin(\pi t)u(t)$, where $u(t)$ is the continuous-time step function. Is the odd portion of this signal, $x_o(t)$, periodic? Justify your answer.

1-34 An aperiodic signal is defined as $x(t) = \cos(\pi t)u(t)$, where $u(t)$ is the continuous-time step function. Is the even portion of this signal, $x_e(t)$, periodic? Justify your answer.

1-35 Consider the signal $x(t)$ shown in Fig. P1-35.
(a) Determine and carefully sketch $v(t) = 3x(-(1/2)(t+1))$.
(b) Determine the energy and power of $v(t)$.
(c) Determine and carefully sketch the even portion of $v(t)$, $v_e(t)$.
(d) Let $a = 2$ and $b = 3$, sketch $v(at+b)$, $v(at)+b$, $av(t+b)$, and $av(t)+b$.
(e) Let $a = -3$ and $b = -2$, sketch $v(at+b)$, $v(at)+b$, $av(t+b)$, and $av(t)+b$.

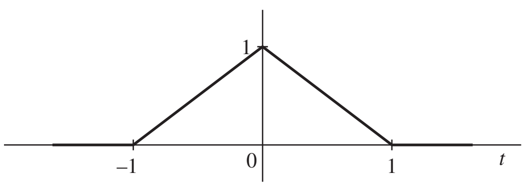


Figure P1-35 Input $x(t)$.

1-36 Consider the signal $y(t) = (1/5)x(-2t-3)$ shown in Figure P1-36.
(a) Does $y(t)$ have an odd portion, $y_o(t)$? If so, determine and carefully sketch $y_o(t)$. Otherwise, explain why no odd portion exists.
(b) Determine and carefully sketch the original signal $x(t)$.

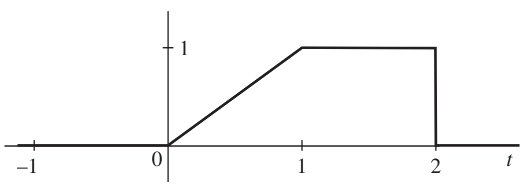


Figure P1-36 $y(t) = \frac{1}{5}x(-2t-3)$.

1-37 Consider the signal $-(1/2)x(-3t+2)$ shown in Fig. P1-37.
(a) Determine and carefully sketch the original signal $x(t)$.
(b) Determine and carefully sketch the even portion of the original signal $x(t)$.
(c) Determine and carefully sketch the odd portion of the original signal $x(t)$.

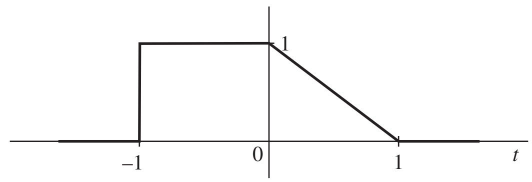


Figure P1-37 $-\frac{1}{2}x(-3t+2)$.

1-38 The conjugate symmetric (or Hermitian) portion of a signal is defined as $w_{cs}(t) = (w(t) + w^*(-t))/2$. Show that the real portion of $w_{cs}(t)$ is even and that the imaginary portion of $w_{cs}(t)$ is odd.

1-39 The conjugate antisymmetric (or skew-Hermitian) portion of a signal is defined as $w_{ca}(t) = (w(t) - w^*(-t))/2$. Show that the real portion of $w_{ca}(t)$ is odd and that the imaginary portion of $w_{ca}(t)$ is even.

1-40 Figure P1-40 plots a complex signal $w(t)$ in the complex plane over the time range $(0 \le t \le 1)$. The time $t = 0$ corresponds with the origin, while the time $t = 1$ corresponds with the point (2,1).

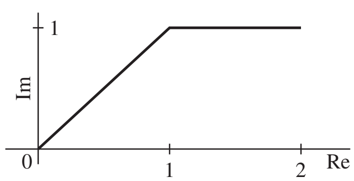


Figure P1-40 $w(t)$ for $(0 \le t \le 1)$.

(a) In the complex plane, plot $w(t)$ over $(-1 \le t \le 1)$ if:
(i) $w(t)$ is an even signal.
(ii) $w(t)$ is an odd signal.
(iii) $w(t)$ is a conjugate symmetric signal. [Hint: See Prob. 1-38.]
(iv) $w(t)$ is a conjugate antisymmetric signal. [Hint: See Prob. 1-39.]
(b) In the complex plane, plot as much of $w(3t)$ as possible.

1-41 Define complex signal $x(t) = t^2(1+j)$ over interval $(1 \le t \le 2)$. The remaining portion is defined such that $x(t)$ is a minimum-energy, skew-Hermitian signal.
(a) Fully describe $x(t)$ for all t.
(b) Sketch $y(t) = \mathrm{Re}\{x(t)\}$ versus the independent variable t.

(c) Sketch $z(t) = \text{Re}\{jx(-2t+1)\}$ versus the independent variable t.

(d) Determine the energy and power of $x(t)$.

1-42 Write the input–output relationship for an ideal integrator. Determine the zero-input and zero-state components of the response.

1-43 A force $x(t)$ acts on a ball of mass M (Fig. P1-43). Show that the velocity $v(t)$ of the ball at any instant $t > 0$ can be determined if we know the force $x(t)$ over the interval from 0 to t and the ball's initial velocity $v(0)$.

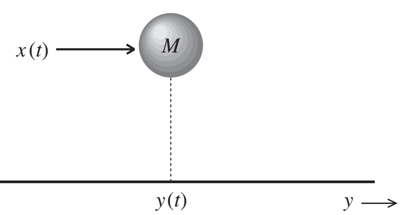


Figure P1-43

1-44 For the systems described by the following equations, with the input $x(t)$ and output $y(t)$, determine which of the systems are linear and which are nonlinear.

(a) $\dfrac{dy}{dt} + 2y(t) = x^2(t)$

(b) $\dfrac{dy}{dt} + 3ty(t) = t^2x(t)$

(c) $3y(t) + 2 = x(t)$

(d) $\dfrac{dy}{dt} + y^2(t) = x(t)$

(e) $\left(\dfrac{dy}{dt}\right)^2 + 2y(t) = x(t)$

(f) $\dfrac{dy}{dt} + (\sin t)y(t) = \dfrac{dx}{dt} + 2x(t)$

(g) $\dfrac{dy}{dt} + 2y(t) = x(t)\dfrac{dx}{dt}$

(h) $y(t) = \displaystyle\int_{-\infty}^{t} x(\tau)\,d\tau$

1-45 For the systems described by the following equations, with the input $x(t)$ and output $y(t)$, explain with reasons which of the systems are time-invariant parameter systems and which are time-varying-parameter systems.

(a) $y(t) = x(t-2)$

(b) $y(t) = x(-t)$

(c) $y(t) = x(at)$

(d) $y(t) = t\,x(t-2)$

(e) $y(t) = \displaystyle\int_{-5}^{5} x(\tau)\,d\tau$

(f) $y(t) = \left(\dfrac{dx}{dt}\right)^2$

1-46 For a certain LTI system with the input $x(t)$, the output $y(t)$ and the two initial conditions $q_1(0)$ and $q_2(0)$, following observations were made:

$x(t)$	$q_1(0)$	$q_2(0)$	$y(t)$
0	1	−1	$e^{-t}u(t)$
0	2	1	$e^{-t}(3t+2)u(t)$
$u(t)$	−1	−1	$2u(t)$

Determine $y(t)$ when both the initial conditions are zero and the input $x(t)$ is as shown in Fig. P1-46. [Hint: There are three causes: the input and each of the two initial conditions. Because of the linearity property, if a cause is increased by a factor k, the response to that cause also increases by the same factor k. Moreover, if causes are added, the corresponding responses add.]

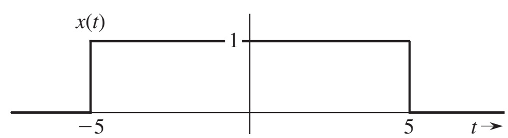


Figure P1-46

1-47 A system is specified by its input–output relationship as

$$y(t) = \frac{x^2(t)}{dx/dt}$$

Show that the system satisfies the homogeneity property but not the additivity property.

1-48 Show that the circuit in Fig. P1-48 is zero-state linear but not zero-input linear. Assume all diodes to have identical (matched) characteristics. The output is the current $y(t)$.

1-49 The inductor L and the capacitor C in Fig. P1-49 are nonlinear, which makes the circuit nonlinear. The remaining three elements are linear. Show that the output $y(t)$ of this nonlinear circuit satisfies the linearity

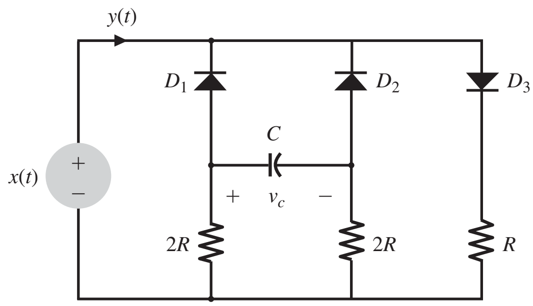


Figure P1-48

conditions with respect to the input $x(t)$ and the initial conditions (all the initial inductor currents and capacitor voltages).

1-50 For the systems described by the following equations, with the input $x(t)$ and output $y(t)$, determine which are causal and which are noncausal.

(a) $y(t) = x(t-2)$
(b) $y(t) = x(-t)$
(c) $y(t) = x(at) \quad a > 1$
(d) $y(t) = x(at) \quad a < 1$

1-51 For the systems described by the following equations, with the input $x(t)$ and output $y(t)$, determine which are invertible and which are noninvertible. For the invertible systems, find the input–output relationship of the inverse system.

(a) $y(t) = \int_{-\infty}^{t} x(\tau)\,d\tau$
(b) $y(t) = x^n(t) \quad x(t)$ real and n integer
(c) $y(t) = \dfrac{dx(t)}{dt}$
(d) $y(t) = x(3t-6)$
(e) $y(t) = \cos[x(t)]$
(f) $y(t) = e^{x(t)} \quad x(t)$ real

1-52 Consider a system that multiplies a given input by a ramp function, $r(t) = tu(t)$. That is, $y(t) = x(t)r(t)$.

(a) Is the system linear? Justify your answer.
(b) Is the system memoryless? Justify your answer.
(c) Is the system causal? Justify your answer.
(d) Is the system time invariant? Justify your answer.

1-53 A continuous-time system is given by

$$y(t) = 0.5\int_{-\infty}^{\infty} x(\tau)[\delta(t-\tau) - \delta(t+\tau)]\,d\tau$$

Recall that $\delta(t)$ designates the Dirac delta function.

(a) Explain what this system does.
(b) Is the system BIBO stable? Justify your answer.
(c) Is the system linear? Justify your answer.
(d) Is the system memoryless? Justify your answer.
(e) Is the system causal? Justify your answer.
(f) Is the system time invariant? Justify your answer.

1-54 A system is given by

$$y(t) = \frac{d}{dt}x(t-1)$$

(a) Is the system BIBO stable? [Hint: Let system input $x(t)$ be a square wave.]
(b) Is the system linear? Justify your answer.
(c) Is the system memoryless? Justify your answer.
(d) Is the system causal? Justify your answer.
(e) Is the system time invariant? Justify your answer.

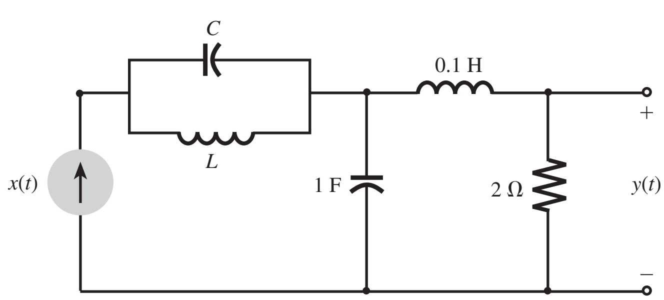


Figure P1-49

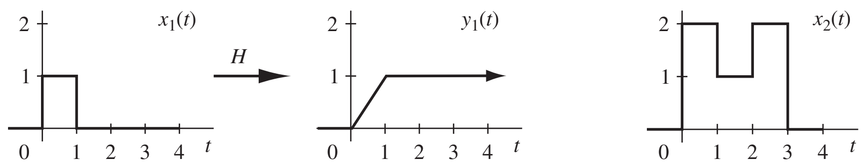


Figure P1-56 $x_1(t) \xrightarrow{H} y_1(t)$ and $x_2(t)$.

1-55 A system is given by

$$y(t) = \begin{cases} x(t) & \text{if} \quad x(t) > 0 \\ 0 & \text{if} \quad x(t) \leq 0 \end{cases}$$

(a) Is the system BIBO stable? Justify your answer.
(b) Is the system linear? Justify your answer.
(c) Is the system memoryless? Justify your answer.
(d) Is the system causal? Justify your answer.
(e) Is the system time invariant? Justify your answer.

1-56 Figure P1-56 displays an input $x_1(t)$ to a linear time-invariant (LTI) system H, the corresponding output $y_1(t)$, and a second input $x_2(t)$.
(a) Bill suggests that $x_2(t) = 2x_1(3t) - x_1(t-1)$. Is Bill correct? If yes, prove it. If not, correct his error.
(b) Hoping to impress Bill, Sue wants to know the output $y_2(t)$ in response to the input $x_2(t)$. Provide her with an expression for $y_2(t)$ in terms of $y_1(t)$. Use MATLAB to plot $y_2(t)$.

1-57 For the circuit depicted in Fig. P1-57, find the differential equations relating outputs $y_1(t)$ and $y_2(t)$ to the input $x(t)$.

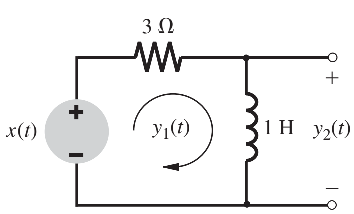


Figure P1-57

1-58 Repeat Prob. 1-57 for the circuit in Fig. P1-58.

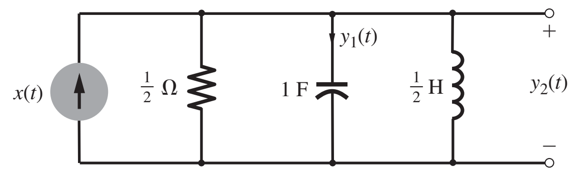


Figure P1-58

1-59 A simplified (one-dimensional) model of an automobile suspension system is shown in Fig. P1-59. In this case, the input is not a force but a displacement $x(t)$ (the road contour). Find the differential equation relating the output $y(t)$ (auto body displacement) to the input $x(t)$ (the road contour).

1-60 A field-controlled dc motor is shown in Fig. P1-60. Its armature current i_a is maintained constant. The torque generated by this motor is proportional to the field current i_f (torque $= K_f i_f$). Find the differential equation relating the output position θ to the input voltage $x(t)$. The motor and load together have a moment of inertia J.

1-61 Water flows into a tank at a rate of q_i units/s and flows out through the outflow valve at a rate of q_0 units/s (Fig. P1-61). Determine the equation relating the outflow q_0 to the input q_i. The outflow rate is proportional to the head h. Thus $q_0 = Rh$, where R is the valve resistance. Determine also the differential equation relating the head h to the input q_i. [Hint: The net inflow of water in time Δt is $(q_i - q_0)\Delta t$. This inflow is also $A\Delta h$ where A is the cross section of the tank.]

1-62 Consider the circuit shown in Fig. P1-62, with input voltage $x(t)$ and output currents $y_1(t)$, $y_2(t)$, and $y_3(t)$.
(a) What is the order of this system? Explain your answer.
(b) Determine the matrix representation for this system.

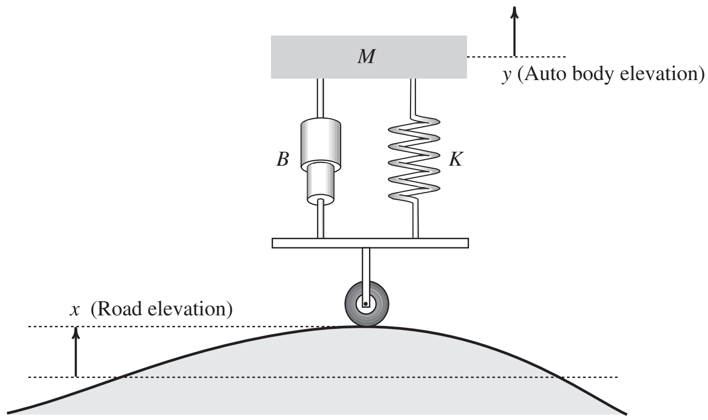


Figure P1-59

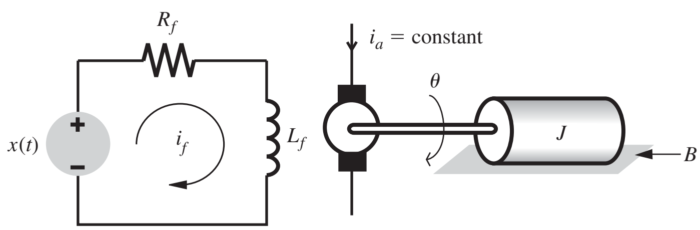


Figure P1-60

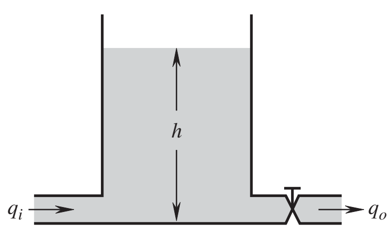


Figure P1-61

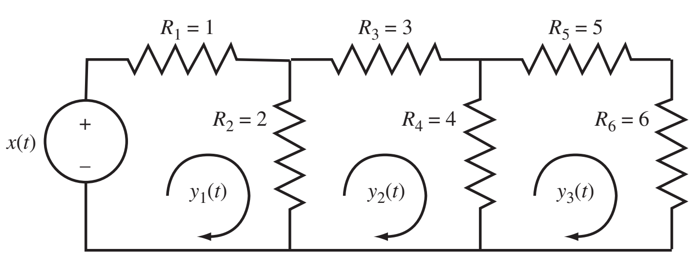


Figure P1-62 Resistor circuit.

(c) Use Cramer's rule to determine the output current $y_3(t)$ for the input voltage $x(t) = (2 - |\cos(t)|)u(t-1)$.

1-63 Write state equations for the parallel RLC circuit in Fig. P1-58. Use the capacitor voltage q_1 and the inductor current q_2 as your state variables. Show that every possible current or voltage in the circuit can be expressed in terms of q_1, q_2 and the input $x(t)$.

1-64 Write state equations for the third-order circuit shown in Fig. P1-64, using the inductor currents q_1, q_2 and the capacitor voltage q_3 as state variables. Show that every possible voltage or current in this circuit can be expressed as a linear combination of q_1, q_2, q_3, and the input $x(t)$. Also, at some instant t it was found that $q_1 = 5$, $q_2 = 1$, $q_3 = 2$, and $x = 10$. Determine the voltage across and the current through every element in this circuit.

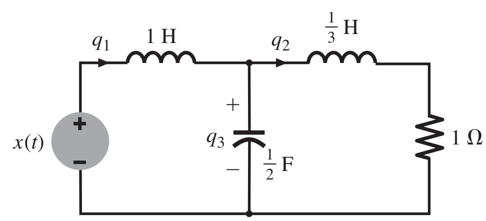


Figure P1-64