Chapter 1 Introduction to Signals

Signals are detectable quantities used to convey information about time-varying physical phenomena. Common examples of signals are human voice, temperature, pressure, and stock prices. Electrical signals, normally expressed in the form of voltage or current wave, are some of the easiest one to generate and process.

Mathematically, signals are modeled as functions of one or more independent variables. Examples of independent variables used to represent signals are time, frequency, or spatial coordinates. Before introducing the mathematical representation of signals, let us consider a few physical systems associated with the generation of signals. Fig. 1.1 illustrates some common signals and systems encountered in different fields of engineering, with the physical systems represented in the left-hand column and the associated signals included in the right-hand column. Fig. 1.1(a) is a simple electrical circuit consisting of three passive components: a capacitor C, an inductor L, and a resistor R. A voltage $v(t)$ is applied at the input of the RLC circuit, which produces an output voltage $y(t)$ across the capacitor. A possible waveform for $y(t)$ is the sinusoidal signal shown in Fig. 1.1(b). The notations $v(t)$ and $y(t)$ include the dependent variables, v and y, respectively, in the two expressions, and the independent variable t. The $v(t)$ indicates that the voltage v is a function of time t. Fig. 1.1(c) shows an audio recording system where the input signal is an audio or a speech waveform. The function of the audio recording system is to convert the audio signal into an electrical waveform, recorded on a magnetic tape or a compact disc. A possible resulting waveform for the recorded electrical signal is shown in Fig. 1.1(d). Fig. 1.1(e) shows a charge-coupled device (CCD) based on the digital camera where the input signal is the light emitted from a scene. The incident light charges a CCD panel located inside the camera, thereby storing the external scene in terms of the spatial variations of the charges on the CCD panel. Fig. 1.1(g) illustrates a thermometer that measures the ambient temperature of its environment. Electronic thermometers typically use a thermal resistor, known as a thermistor, whose resistance varies with temperature. The fluctuations in the resistance are used to measure the temperature. Fig. 1.1(h) plots the readings of the thermometer as a function of discrete time. In the aforementioned examples of Fig. 1.1, the RLC circuit, audio recorder, CCD camera, and thermometer represent different systems, while the information-bearing waveforms, such as the voltage, audio, charges, and fluctuations in resistance, do signals. The output waveforms, for example the voltage in the case of the electrical circuit, current for the microphone, and the fluctuations in the resistance for the thermometer, vary with respect to only one variable (time) and are classified as one-dimensional (1D) signals. On the other hand, the charge distribution in the CCD panel of the camera varies spatially in two dimensions. The independent variables are the two spatial coordinates (m, n). The charge distribution signal is therefore classified as a two-dimensional (2D) one.

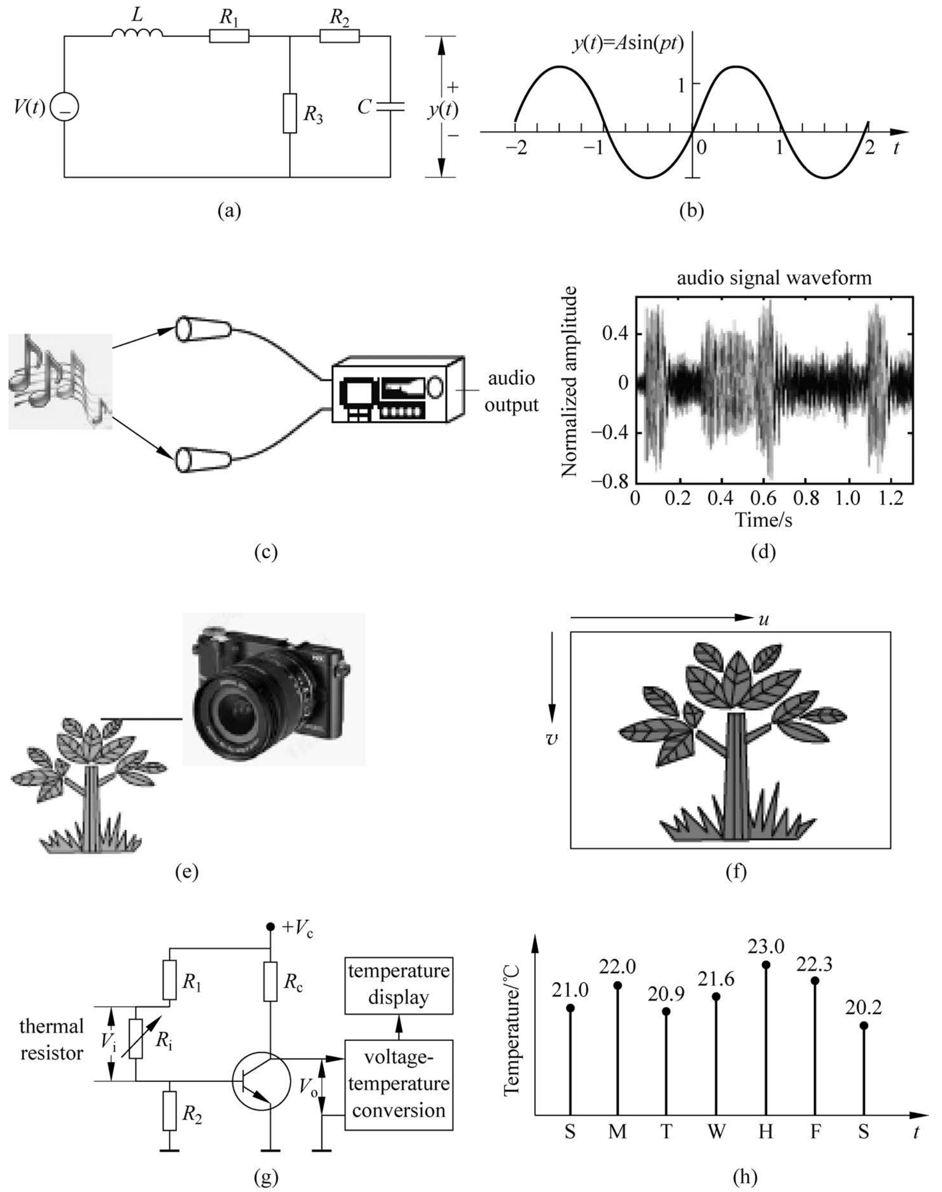

Fig. 1. 1　Examples of signals and systems

(a) An electrical circuit; (b) Output signal generated by the electrical circuit; (c) An audio recording system; (d) Output signal generated by the audio recording system; (e) A digital camera; (f) Output signal generated by the digital camera; (g) A digital thermometer; (h) Output signal generated by the digital thermometer

The examples shown in Fig. 1. 1 illustrate that typically every system has one or more signals associated with it. A system is therefore defined as an entity that processes a set of signals (called the input signals) and produces another set of signals (called the output signals). The voltage source in Fig. 1. 1(a), the sound in Fig. 1. 1(c), the light entering the camera in Fig. 1. 1(e), and the ambient heat in Fig. 1. 1(g) provide examples of the input signals. The voltage across capacitor C in Fig. 1. 1(b), the voltage generated by the microphone in Fig. 1. 1(d), the charge stored on the CCD panel of the digital camera, displayed as an image in Fig. 1. 1(f), and the voltage generated by the thermistor, used to measure the room temperature, in Fig. 1. 1(h) are examples of output signals.

Fig. 1. 2 shows a simplified schematic representation of a signal-processing system. The system processes an input signal $x(t)$ producing an output $y(t)$. This model may be used to represent a range of physical processes including electrical circuits, mechanical devices, hydraulic systems, and computer algorithms with a single input and a single output. More complex systems have multiple inputs and multiple outputs.

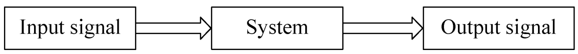

Fig. 1. 2 Processing of a signal by a system

1. 1 Classification of Signals

Signals are classified into several categories depending upon the criteria used for their classification. In this section, we cover the following categories for signals:

(i) continuous-time and discrete-time signals;

(ii) analog and digital signals;

(iii) deterministic and nondeterministic signals;

(iv) energy and power signals;

(v) even and odd signals.

1. 1. 1 Continuous-and discrete-time signals

If a signal is defined for continuous values of the independent variable t, it is called a continuous-time (CT) signal. Consider the signals shown in Figs. 1. 1(b) and (d). Since these signals vary continuously with time t and have known magnitudes for all time instants, they are classified as CT signals. On the other hand, if a signal is defined only at discrete values of time, it is called a discrete-time (DT) signal. Fig. 1. 1(h) shows the output temperature of a room measured at the same hour every day for one week. No information is available for the temperature between the daily readings. Fig. 1. 1(h) is therefore an example of a DT signal. A CT signal is denoted by $x(t)$ with regular parenthesis, and a DT signal with square parenthesis as follows:

$$x(kT), \quad k = 0, \pm 1, \pm 2, \pm 3, \ldots, \tag{1.1}$$

where T denotes the time interval between two consecutive samples. In the example of Fig. 1.1(h), the value of T is one day. To keep the notation simple, we denote a 1D DT signal x by $x(k)$. Though the sampling interval is not explicitly included in $x(k)$, it will be incorporated if and when required.

Note that every DT signal is not the function of time. Fig. 1.1(f), for example, shows the output of a CCD camera, where the discrete output varies spatially in two dimensions. Here, the independent variables are denoted by (u, v), where u and v are the discretized horizontal and vertical coordinates of the picture element. In this case, the 2D DT signal representing the spatial charge is denoted by $x(u, v)$.

1.1.2 Analog and digital signals

A second classification of signals is based on their amplitudes. The amplitudes of many real-world signals, such as voltage, current, temperature, and pressure, change continuously, called analog signals. For example, the ambient temperature of a house is an analog number that requires an infinite number of digits (e.g., 24.763578...) to record the readings precisely. Digital signals, on the other hand, can only have a finite number of amplitude values. For example, if a digital thermometer, with a resolution of 1℃ and a range of [10℃, 30℃], is used to measure the room temperature at discrete time instants, $t = kT$, then the recordings constitute a digital signal. An example of a digital signal was shown in Fig. 1.1(h), which plots the temperature readings taken once a day for one week. This digital signal has an amplitude resolution of 0.1℃, and a sampling interval of one day.

Fig. 1.3 shows an analog signal with its digital approximation. Its waveform is shown with a line plot. The analog signal has a limited dynamic range between [−1, 1] but can assume any real value (rational or irrational). If the analog signal is sampled at time instants $t = kT$ and the magnitude of the resulting samples are quantized to a set of finite number of known values within the range [−1, 1], the resulting signal becomes a digital signal. Using the following set of nine uniformly distributed values, [0.000, 0.625, 0.875, 0.875, 0.625, 0.000, −0.625, −0.875, −0.875].

Another example of a digital signal is the music recorded on an audio compact disc (CD). On a CD, the music signal is first sampled at a rate of 44,100 samples per second. The sampling interval T is given by 1/44,100 s, or 22.68 microseconds (μs). Each sample is then quantized with a 16-bit uniform quantizer. In other words, a sample of the recorded music signal is approximated from a set of uniformly distributed values that can be represented by a 16-bit binary number. The total number of values in the discretized set is therefore limited to 2^{16} entries.

Digital signals may also occur naturally. For example, the price of a product is a multiple of the lowest denomination of a currency. The grades of students on a course are

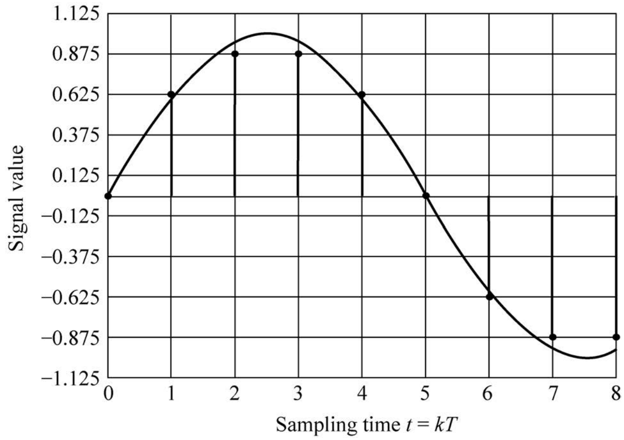

Fig. 1. 3 Analog signal with its digital approximation

also discrete, e. g. 8 out of 10, or 3. 6 out of 4 on a 4-point grade point average. The number of employees in an organization is a nonnegative integer and is also digital by nature.

1. 1. 3 Deterministic and nondeterministic signals

Signals can be classified as either the deterministic or the nondeterministic (random). A deterministic signal can be described by an explicit mathematical relation. Its future behavior, therefore, is predictable. Each time history record of a random signal is unique. Its future behavior cannot be determined exactly but to within some limits with a certain confidence.

Deterministic signals can be classified further into the static and the dynamic, as shown in Fig. 1. 4. Static signals are steady in time. Their amplitude remains constant. Dynamic signals are either periodic or aperiodic. A periodic signal, $y(t)$, repeats itself at regular intervals, nT, where $n=1,2,3,\ldots$ Analytically, this is expressed as

$$y(t+nT)=y(t) \tag{1.2}$$

for all t. The smallest value of T for which Eq. (1. 2) holds true is called the fundamental period. If $y(t)$ and $z(t)$ are periodic signals and the period ratio of $y(t)$ to $z(t)$ is a rational number, then their product $y(t)z(t)$ and the sum of any linear combination of them, $c_1 y(t)+c_2 z(t)$, are also periodic.

A simple periodic signal has one period. A complex periodic signal has more than one period. An almost-periodic signal is comprised of two or more sinusoids of arbitrary frequencies. However, if the ratios of all possible pairs of frequencies are rational numbers, then an almost-periodic signal is also periodic.

Nondeterministic signals are classified as shown in Fig. 1. 5. Properties of the ensemble of the nondeterministic signals can be computed by taking the average of the instantaneous values acquired from each of the time histories at an arbitrary time, t_1. The ensemble mean value, $\mu_x(t_1)$, and the ensemble autocorrelation function (see Chapter 2

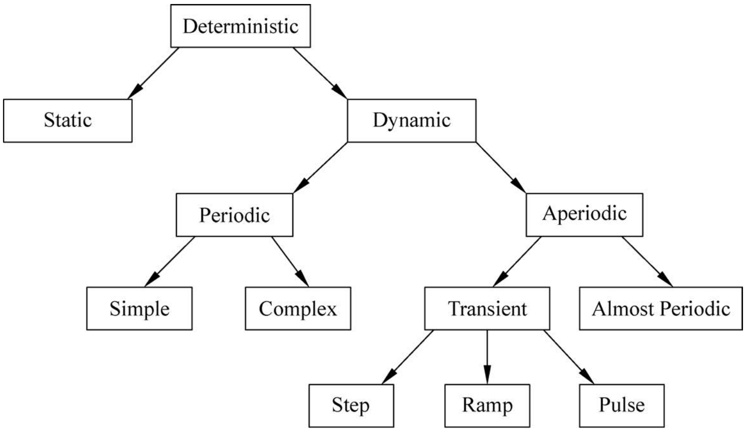

Fig. 1.4　Deterministic signal subdivisions

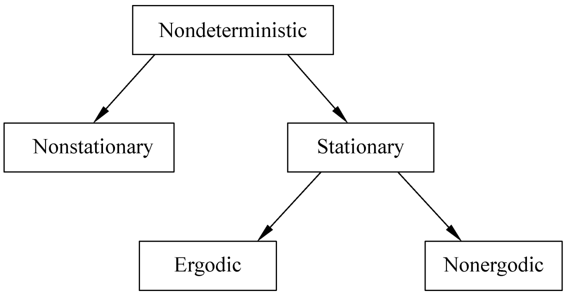

Fig. 1.5　Nondeterministic signal subdivisions

for more on the autocorrelation), $R_x(t_1, t_1+\tau)$, are

$$\mu_x(t_1) = \lim_{N \to +\infty} \frac{1}{N} \sum_{i=1}^{N} x_i(t_1) \tag{1.3}$$

and

$$R_x(t_1, t_1+\tau) = \lim_{N \to +\infty} \frac{1}{N} \sum_{i=1}^{N} x_i(t_1) x_i(t_1+\tau), \tag{1.4}$$

in which τ denotes an arbitrary time measured from time t_1. Both equations represent ensemble averages. This is because $\mu_x(t_1)$ and $R_x(t_1, t_1+\tau)$ are determined by performing averages over the ensemble at time t_1.

If the values of $\mu_x(t_1)$ and $R_x(t_1, t_1+\tau)$ change with t_1, then the signal is nonstationary. Otherwise, it is stationary (in the wide sense). A nondeterministic signal is considered to be weakly stationary only when $\mu_x(t_1)=\mu_x$ and $R_x(t_1, t_1+\tau)=R_x(\tau)$, that is, only when the signal's ensemble mean and autocorrelation function are time invariant. In a more restrictive sense, if all other ensemble higher-order moments and joint moments are also time invariant, the signal is strongly stationary (stationary in the *strict* sense). So, the term stationary means that each of a signal's ensemble-averaged statistical properties are constant with respect to t_1. It does not mean that the amplitude of the signal is constant over time. In fact, a random signal is never completely stationary in timing.

For a single time history, the temporal mean value, μ_x, and the temporal autocorrelation coefficient, $R_x(\tau)$, are

$$\mu_x = \lim_{T\to+\infty} \frac{1}{T}\int_0^T x(t)\mathrm{d}t \tag{1.5}$$

and

$$R_x(\tau) = \lim_{T\to+\infty} \frac{1}{T}\int_0^T x(t)x(t+\tau)\mathrm{d}t. \tag{1.6}$$

For most stationary data, the ensemble averages at an arbitrary time, t_1, will equal their corresponding temporal averages computed for an arbitrary single time history in the ensemble. When this is true, the signal is ergodic. If the signal is periodic, then the limits in Eqs. (1.5) and (1.6) do not exist because averaging over one time period is sufficient. Ergodic signals are important because all of their properties can be determined by performing time averages over a single time history record. This greatly simplifies data acquisition and reduction. Most random signals representing stationary physical phenomena are ergodic.

A finite record of data of an ergodic random process can be used in conjunction with probabilistic methods to quantify the statistical properties of an underlying process. For example, it can be used to determine a random variable's true mean value within a certain confidence limit. These methods can also be applied to deterministic signals, which are considered next below.

1.1.4 Energy and power signals

Before presenting the conditions for classifying a signal as an energy or a power one, we present the formulas for calculating the energy and power in a signal.

The instantaneous power at time $t=t_0$ of a real-valued CT signal $x(t)$ is given by $x^2(t_0)$. Similarly, the instantaneous power of a real-valued DT signal $x(k)$ at time instant $k=k_0$ is given by $x^2(k)$. If the signal is complex-valued, the expressions for the instantaneous power are modified to $|x(t_0)|^2$ or $|x(k_0)|^2$, where the symbol $|\cdot|$ represents the absolute value of a complex number.

The energy present in a CT or DT signal within a given time interval is given as follows:

CT signal $E_{(T_1,T_2)} = \int_{T_1}^{T_2} |x(t)|^2\mathrm{d}t$ in interval $t=(T_1, T_2)$ with $T_2 > T_1$ (1.7a)

DT sequence $E_{[N_1,N_2]} = \sum_{k=N_1}^{N_2} |x(k)|^2$ in interval $k=[N_1, N_2]$ with $N_2 > N_1$. (1.7b)

The total energy of a CT signal is calculated over the interval $t=[-\infty, +\infty]$. Likewise, a DT signal's is done over the range $k=[-\infty, +\infty]$. The expressions for the total energy are therefore given as

$$\text{CT signal}\quad E_x = \int_{-\infty}^{+\infty} |x(t)|^2 \mathrm{d}t, \tag{1.8a}$$

$$\text{DT sequence}\quad E_x = \sum_{k=-\infty}^{+\infty} |x(k)|^2. \tag{1.8b}$$

Since power is defined as energy per unit time, the average power of a CT signal $x(t)$ over the interval $t=(-\infty,+\infty)$ and of a DT signal $x(k)$ over the range $k=[-\infty,+\infty]$ are expressed as follows:

$$\text{CT signal}\quad P_x = \lim_{T\to+\infty} \frac{1}{T}\int_{-T/2}^{T/2} |x(t)|^2 \mathrm{d}t, \tag{1.9}$$

$$\text{DT sequence}\quad P_x = \frac{1}{2K+1}\sum_{k=-K}^{K} |x(k)|^2. \tag{1.10}$$

Eqs. (1.9) and (1.10) are simplified considerably for periodic signals. Since a periodic signal repeats itself, the average power is calculated from one period of the signal as follows:

$$\text{CT signal}\quad P_x = \frac{1}{T_0}\int_{\langle T_0\rangle} |x(t)|^2 \mathrm{d}t = \frac{1}{T_0}\int_{t_1}^{t_1+T_0} |x(t)|^2 \mathrm{d}t. \tag{1.11}$$

$$\text{DT sequence}\quad P_x = \frac{1}{K_0}\sum_{k=K_0} |x(k)|^2 = \frac{1}{K_0}\sum_{k=k_1}^{k_1+K_0-1} |x(k)|^2. \tag{1.12}$$

where t_1 is an arbitrary real number and k_1 is an arbitrary integer. The symbols T_0 and K_0 are, respectively, the fundamental periods of the CT signal $x(t)$ and the DT signal $x(k)$. In Eq. (1.11), the duration of integration is one complete period over the range $[t_1, t_1+T_0]$, where t_1 can take any arbitrary value. In other words, the lower limit of integration can have any value provided that the upper limit is one fundamental period apart from the lower limit. To illustrate this mathematically, we introduce the notation $\int_{\langle T_0\rangle}$ to show that the integration is performed over a complete period T_0 and is independent of the lower limit. Likewise, while computing the average power of a DT signal $x(k)$, the upper and lower limits of the summation in Eq. (1.12) can take any values as long as the duration of summation equals one fundamental period K_0.

A signal $x(t)$, or $x(k)$, is called an energy signal if the total energy E_x has a nonzero finite value, i. e. $0<E_x<+\infty$. On the other hand, a signal is called a power signal if it has nonzero finite power, i. e. $0<P_x<+\infty$. Note that a signal cannot be both an energy and a power signals simultaneously. The energy signals have zero average power whereas the power signals have infinite total energy. Some signals, however, can be classified as neither power's nor energy's. For example, the signal $\mathrm{e}^{2t}u(t)$ is a growing exponential whose average power cannot be calculated. Such signals are generally of little interest to us.

Most periodic signals are typically power ones. For example, the average power of the CT sinusoidal signal, or $A\sin(\omega_0 t+\theta)$, is given by $A^2/2$. Similarly, the average power of the complex exponential signal $A\exp(\mathrm{j}\omega_0 t)$ is given by A^2.

Example 1.1

Statement: Consider the following DT sequence:

$$f(k)=\begin{cases}e^{-0.5k}, & k\geqslant 0,\\ 0, & k<0.\end{cases}$$

Determine if the signal is a power or an energy signal.

Solution: The total energy of the DT sequence is calculated as follows:

$$E_f=\sum_{k=-\infty}^{+\infty}|f(k)|^2=\sum_{k=0}^{+\infty}|e^{-0.5k}|^2=\sum_{k=0}^{+\infty}|e^{-1}|^k=\frac{1}{1-e^{-1}}\approx 1.582.$$

Because E_f is finite, the DT sequence $f(k)$ is an energy signal.

In computing E_f, we make use of the geometric progression (GP) to calculate the summation.

Example 1.2

Statement: Determine if the DT sequence $g(k)=3\cos(\pi k/10)$ is a power or an energy signal.

Solution: The DT sequence $g(k)=3\cos(\pi k/10)$ is a periodic signal with a fundamental period of 20. All periodic signals are power's, hence the DT sequence $g(k)$.

Using Eq. (1.15), the average power of $g(k)$ is given by

$$\begin{aligned}P_g&=\frac{1}{20}\sum_{k=0}^{19}9\cos^2\frac{\pi k}{10}=\frac{9}{20}\sum_{k=0}^{19}\frac{1}{2}\left(1+\cos\frac{2\pi k}{10}\right)\\&=\underbrace{\frac{9}{40}\sum_{k=0}^{19}1}_{\text{term I}}+\underbrace{\frac{9}{40}\sum_{k=0}^{19}\cos\frac{2\pi k}{10}}_{\text{term II}}\end{aligned}$$

Clearly, the summation represented by term Ⅰ equals $9\times 20/40=4.5$. To compute the summation in term Ⅱ, we express the cosine as follows:

$$\text{term II}=\frac{9}{40}\sum_{k=0}^{19}\frac{1}{2}(e^{j\pi k/5}+e^{-j\pi k/5})=\frac{9}{80}\sum_{k=0}^{19}(e^{j\pi/5})^k+\frac{9}{80}\sum_{k=0}^{19}(e^{-j\pi/5})^k.$$

Using the formulas for the GP series yields

$$\sum_{k=0}^{19}(e^{j\pi/5})^k=\frac{1-(e^{j\pi/5})^{20}}{1-(e^{j\pi/5})}=\frac{1-e^{-j4\pi}}{1-(e^{j\pi/5})}=\frac{1-1}{1-(e^{j\pi/5})}=0$$

and

$$\sum_{k=0}^{19}(e^{-j\pi/5})^k=\frac{1-(e^{-j\pi/5})^{20}}{1-(e^{j\pi/5})}=\frac{1-e^{-j4\pi}}{1-(e^{j\pi/5})}=\frac{1-1}{1-(e^{j\pi/5})}=0.$$

Term Ⅱ, therefore, equals zero. The average power of $g(k)$ is then given by

$$P_g=4.5+0=4.5.$$

In general, a periodic DT sinusoidal signal of the form $x(k)=A\cos(\omega_0 k+\theta)$ has an average power $P_x=A^2/2$.

1.1.5 Even and odd signals

A CT signal $x_e(t)$ is said to be an even signal if

$$x_e(t)=x_e(-t). \tag{1.13}$$

Conversely, a CT signal $x_o(t)$ is to be an odd signal if

$$x_o(t) = -x_o(-t). \tag{1.14}$$

A DT signal $x_e(k)$ is said to be an even signal if

$$x_e(k) = x_e(-k). \tag{1.15}$$

Conversely, a DT signal $x_o(k)$ is said to be an odd signal if

$$x_o(k) = -x_o(-k). \tag{1.16}$$

The even signal property, Eq. (1.13) for CT signals or Eq. (1.15) for DT signals, implies that an even signal is symmetric about the vertical axis ($t=0$). Likewise, the odd signal property, Eq. (1.14) for CT signals or Eq. (1.16) for DT signals, implies that an odd signal is antisymmetric about the vertical axis ($t=0$). The symmetry characteristics of both signals are illustrated in Fig. 1.6. The waveform in Fig 1.6(a) is even as it is symmetric about the y-axis and the waveform in Fig. 1.6(b) is odd as it is antisymmetric about the y-axis.

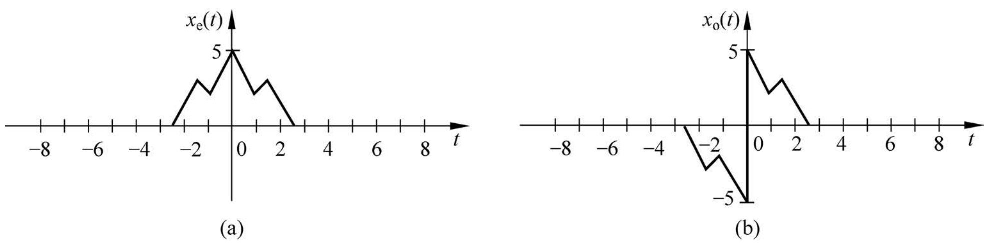

Fig. 1.6 Example of signals

(a) An even signal; (b) An odd signal

Most practical signals are neither odd nor even. The signals do not exhibit any symmetry about the y-axis. Such signals are classified in the "neither odd nor even" category.

Neither odd nor even signals can be expressed as a sum of even and odd signals as follows:

$$x(t) = x_e(t) + x_o(t), \tag{1.17}$$

where the even component $x_e(t)$ is given by

$$x_e(t) = \frac{1}{2}[x(t) + x(-t)], \tag{1.18a}$$

while the odd component $x_o(t)$ is by

$$x_o(t) = \frac{1}{2}[x(t) - x(-t)]. \tag{1.18b}$$

1.2 Elementary Signals

In this section, we define some elementary functions that will be used frequently to represent more complicated signals. Representing signals in terms of the elementary functions simplifies the analysis and design of linear systems.