11

INTRODUCTION TO DIGITAL FILTERS

A digital filter is a digital signal processor that converts a sequence of numbers called the *input* to another sequence of numbers called the *output*. Many theoretical concepts of digital filtering have been known since the days of Laplace. However, the technology of that time could not utilize this body of knowledge. As digital computers came onto the scene, digital filters began to proliferate. Seismic scientists made notable use of digital filtering concepts to solve many interesting problems. Picture processing uses digital filtering techniques to improve the clarity of pictures obtained from remote sensings, interplanetary communications, and x-ray films. Other areas of applications include speech processing, mapping, radar, sonar, and various fields of medical technology.

A digital filter can be implemented as software, such as a subroutine on a digital computer, or as hardware, such as a circuit containing registers, multipliers, and summers. For a number of years, software implementation was the only possible mode of performing digital filtering. Today, software implementation is still the dominant mode. Large-scale digital filters are invariably implemented on a general-purpose or a special-purpose digital computer. However, the rapid development of very large-scale integrated circuit technologies have opened up the area of hardware implementation of digital filters. Currently, the industry can produce adders, shift registers, and multiplier chips needed for the hardware implementation of digital filters at reasonable cost. In addition, general-purpose digital signal processing chips and number-crunching microprocessors are on the horizon. In view of the past history of the IC industry, it is foreseeable that these components may cost much less and perform much better in the future. Consequently, hardware and software implementations may be combined together to yield low-cost and yet efficient digital filters.

11-1 DIGITAL SIGNALS AND SYSTEMS

As mentioned in Chapter 1, a filter is a signal processor that enhances some signals and attenuates others. A signal may be a continuous function of an independent variable, which we usually call *time*, such as voltage and current waveforms in analog filters. These signals are called *continuous-time* signals. On the other hand, a signal may be defined for a finite or at most a countably infinite number of time instants only. This type of signal is called a *discrete-time* signal. Some examples of discrete-time signals are: the annual GNP of a nation shown in Fig. 11-1(a), the monthly unemployment rate in Fig. 11-1(b), the population chart of a small village shown in Fig. 11-1(c), and the monthly automobile production of a company in Fig. 11-1(d). Among the main sources of discrete-time signals are those obtained by sampling a continuous-time signal. A case in point is shown in Fig. 11-2.

Digital signals are discrete-time signals whose values are *quantized.* The output of an A/D converter, which samples a continuous-time input signal and generates a sequence of finite-length binary numbers, is a typical digital signal. The essence of an A/D converter is shown in Fig. 11-3(a). If the sampler samples at the rate of one sample per μsec. and the quantizer has an input–output relationship as given by Fig. 11-3(b), then given a continuous-time signal $\hat{x}(t)$ as in Fig. 11-3(c), the corresponding discrete-time signal $x_1(nT)$ and the output digital signal $x(nT)$ are shown, respectively, in Figs. 11-3(d) and (e). Some other typical digital signals are those shown in Fig. 11-1(c) and (d), where the quantized levels are respectively per person and per car. Strictly speaking, digital computers can handle digital signals only.

Because there are only a finite number of quantized signal levels, errors arise in any system that handles digital signals. Consequently, one of the design considerations of a digital filter is the number of bits or the number of quantized levels needed to represent a digital signal. The larger the number of bits used, the more accurate the representation of the signal and the costlier the filter. Clearly, there is a trade-off between accuracy and cost.

In this book, we do not consider the quantization effect of a digital filter. This essentially means that we have an *infinite-bit* representation of numbers. Thus, we treat digital signals as if they are discrete-time signals. In other words, we make no distinction between the words "discrete-time" and "digital," and we use the word "digital" hereafter.

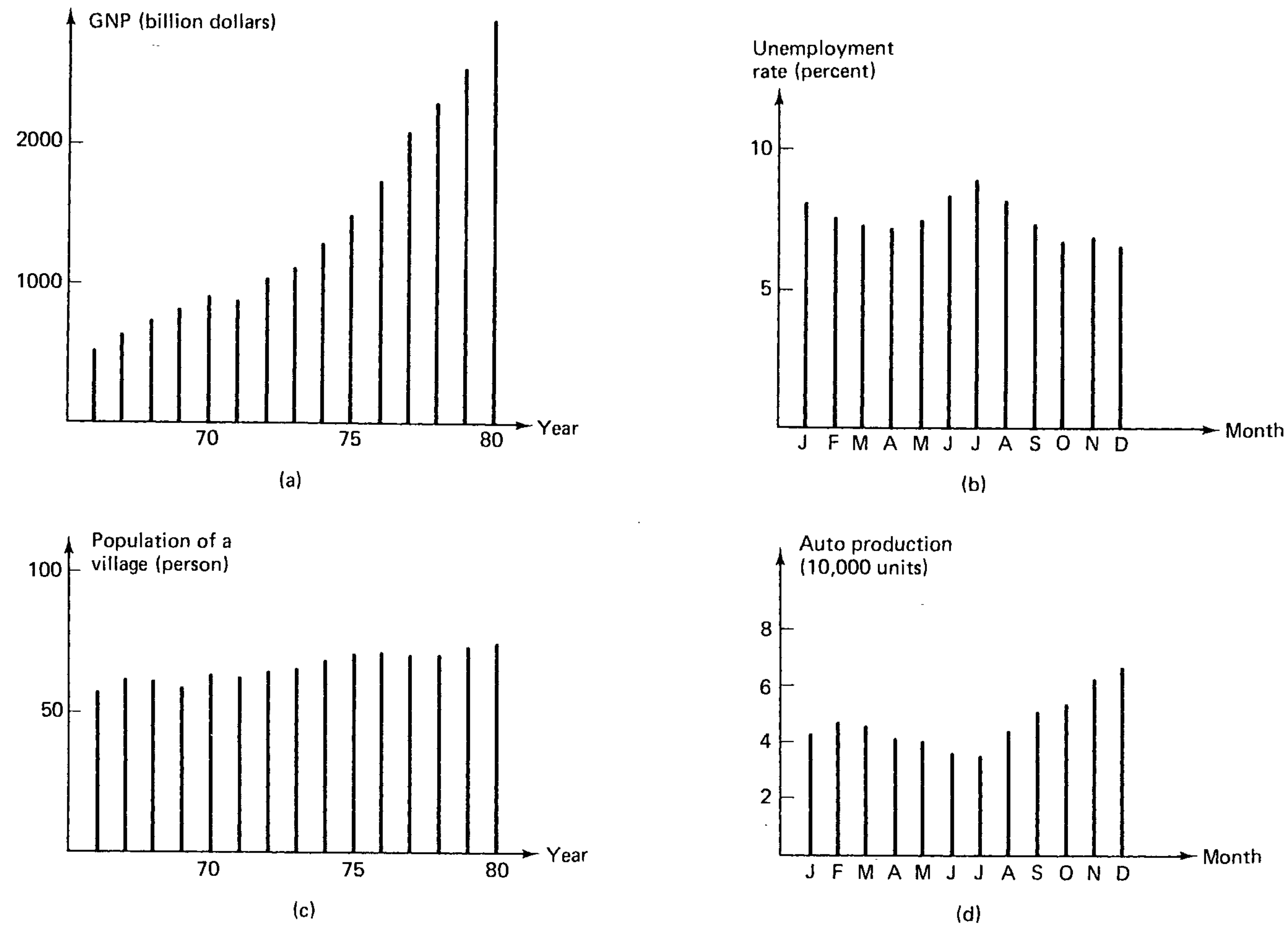


Fig. 11-1 Some examples of discrete-time signals.

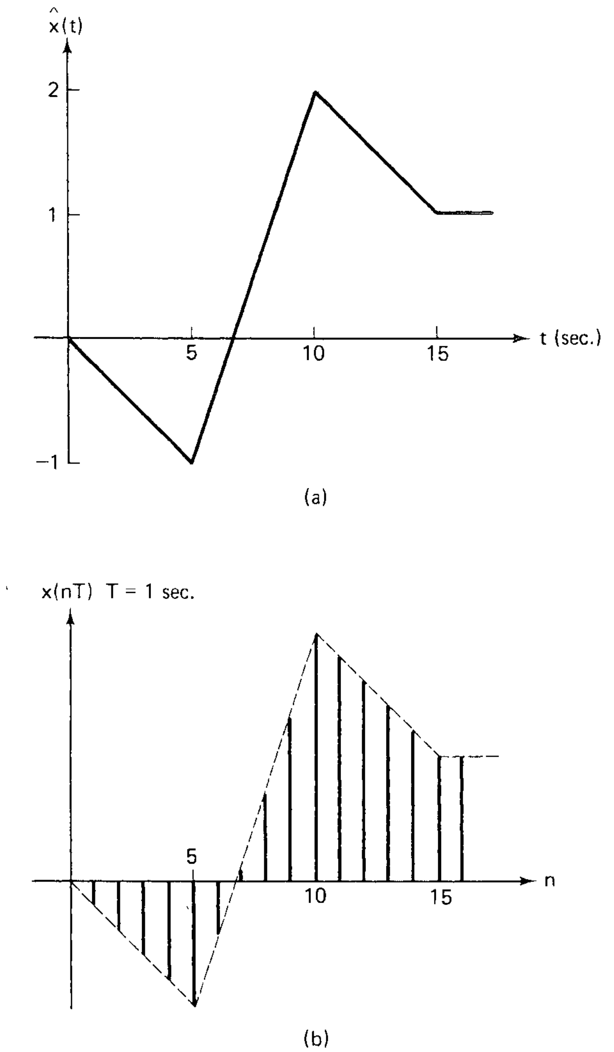


Fig. 11-2 Sampling of a continuous signal. (a) A continuous signal. (b) Its corresponding sampled sequence.

No matter how they arise, digital signals can be considered as sequences of numbers. The notations used to describe digital signals are[1]

$$x(n) \text{ or } \{x(n)\} \tag{11-1a}$$

[1]Strictly speaking, $\{x(n)\}$ denotes the complete sequence, and $x(n)$ denotes the sequence value at the nth point. However, for convenience, we use both $x(n)$ and $\{x(n)\}$ to denote the sequence of x.

In this book, we consider one-dimensional digital sequences only. That is, the values of the sequences depend on one independent variable only.

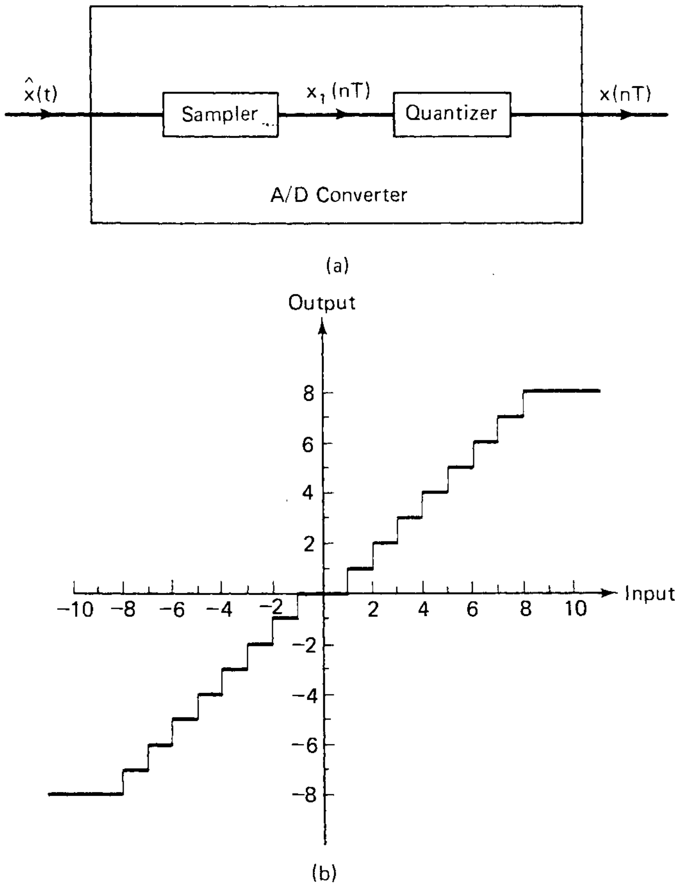


Fig. 11-3 The function of an A/D converter. (a) Schematic. (b) Input–output relationship of the quantizer.

and

$$x(nT) \text{ or } \{x(nT)\} \tag{11-1b}$$

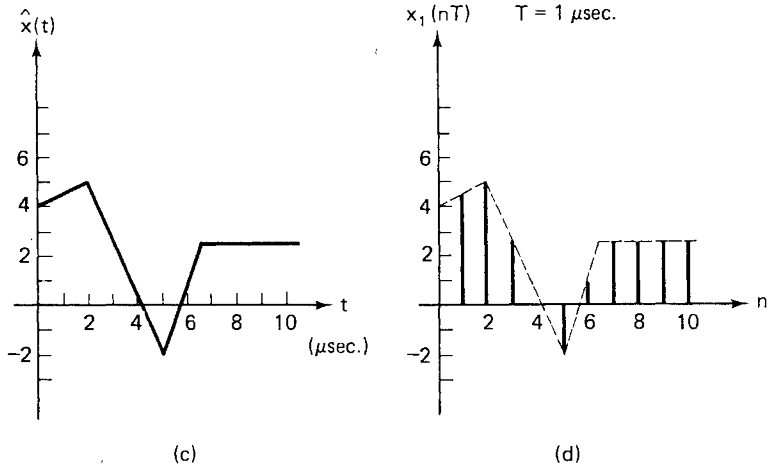
Note that (11-1b) applies to signals with uniform time intervals, whereas (11-1a) applies to signals with uniform as well as nonuniform time spacings.[2]

Some important sequences are:

1. the *unit impulse* sequence $\delta(n)$ defined by

$$\begin{aligned} \delta(n) &= 0 \quad n \neq 0 \\ &= 1 \quad \text{when } n = 0 \end{aligned} \tag{11-2}$$

[2]We consider exclusively digital signals with uniform time spacings only. For those who are interested in systems where the time intervals between signal samples are not identical, please consult Reference [23].



(e)

Legend: $\hat{x}(t)$ = continuous-time signal, input to the A/D converter.
x(nT) = discrete-time signal, output of the sampler.
x(nT) = digital signal, output of the A/D converter.

Fig. 11-3 (c), (d), and (e) An example.

Observe that a sequence $x(n)$ given by

$$\{x(n)\} = \{\ldots, x(-1), x(0), x(1), \ldots\}$$

can be written in terms of the unit impulse sequence as

$$x(n) = \sum_{k=-\infty}^{\infty} x(k)\,\delta(n-k) \tag{11-3}$$

2. the *unit step* sequence $u(n)$ defined by

$$\begin{aligned} u(n) &= 1 \quad \text{when } n \geq 0 \\ &= 0 \quad \text{when } n < 0 \end{aligned} \tag{11-4}$$

Based on the definitions in (11-2) and (11-4), the relationships between the unit impulse and the unit step sequences are

$$u(n) = \sum_{k=-\infty}^{n} \delta(k) \tag{11-5a}$$

$$\delta(n) = u(n) - u(n-1) \tag{11-5b}$$

3. an *exponential* sequence

$$\begin{aligned} x(n) &= a^n \quad \text{when } n \geq 0 \\ &= 0 \quad \text{when } n < 0 \end{aligned} \tag{11-6a}$$

where a may be real or complex. Note that an exponential sequence can be expressed as

$$x(n) = a^n u(n) \tag{11-6b}$$

4. *sinusoidal* sequences with period P

$$x_1(n) = A_1 \cos (2\pi n/P) \tag{11-7a}$$

$$x_2(n) = A_2 \sin (2\pi n/P) \tag{11-7b}$$

If P is a positive rational number, say $P = \alpha/\beta$, where both α and β are two relatively primed positive integers, then the sequences in (11-7) repeat every α sample. That is,

$$x_k(n) = x_k(n + m\alpha) \tag{11-8}$$

where $k = 1, 2$ and m is an integer. On the other hand, if P is an irrational positive number, then the sequences in (11-7) do not repeat themselves. Therefore, sinusoidal digital sequences are *not* necessarily periodic sequences.

Just like continuous-time functions, digital signals or sequences are subjected to arithmetic operations. Let $x \triangleq \{x(n)\}$ and $y \triangleq \{y(n)\}$ be two sequences, and let α be a scalar. Then we define:

1. sum and difference of two sequences

$$x \pm y \triangleq \{x(n) \pm y(n)\} \tag{11-9a}$$

2. multiplication of a sequence by a scalar

$$\alpha x \triangleq \{\alpha x(n)\} \tag{11-9b}$$

3. multiplication and division of two sequences

$$xy \triangleq \{x(n)\, y(n)\} \tag{11-9c}$$

$$\frac{x}{y} \triangleq \{x(n)/y(n)\} \tag{11-9d}$$

In the "time" domain, a digital system is characterized by a set of *difference* equations.[3] This means that given an input sequence and the initial conditions of the system, the set of difference equations will yield a unique output sequence.[4] For example, consider the system characterized by

$$y(n) - ay(n-1) = x(n) \tag{11-10a}$$

$$y(0) = 1 \tag{11-10b}$$

where $x(n)$ and $y(n)$ are, respectively, the input and the output sequences as shown in Fig. 11-4. If the input sequence is a unit step

$$\begin{aligned} x(n) &= 1 \quad \text{for } n \geq 0 \\ &= 0 \quad \text{for } n < 0 \end{aligned} \tag{11-11}$$

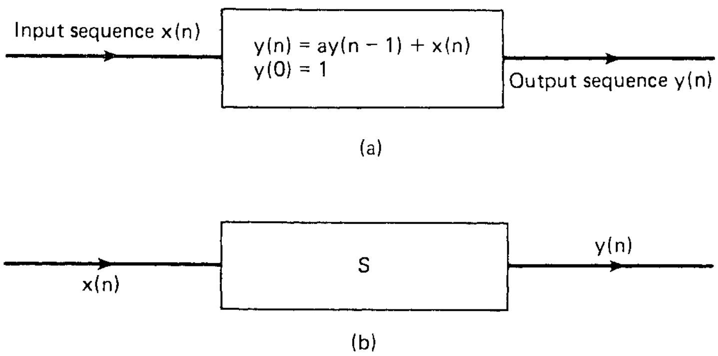


Fig. 11-4 A digital system. (a) A specific case. (b) A general case.

then the output sequence can be computed from (11-10) for $n = 1, 2, \ldots$ as

$$\begin{aligned} y(1) &= ay(0) + x(1) = a + 1 \\ y(2) &= ay(1) + x(2) = a(a+1) + 1 \\ &= a^2 + a + 1 \\ &\;\vdots \\ y(k) &= ay(k-1) + x(k) \\ &= a^k + a^{k-1} + a^{k-2} + \ldots + 1 = \left(\sum_{i=0}^{k} a^i\right) \end{aligned} \tag{11-12}$$

---

[3] Recall that a continuous-time system such as that of an active or a passive RLC circuit is characterized by a set of *differential* equations in the time domain.

[4] In this book, all difference equations are assumed to be linear and time-invariant.

If $|a| < 1$, then (11-12) can be written as

$$\begin{aligned} y(k) &= \left(\sum_{i=0}^{k} a^i\right) = \left(\sum_{i=0}^{\infty} a^i\right) - \left(\sum_{i=k+1}^{\infty} a^i\right) \\ &= \left(\sum_{i=0}^{\infty} a^i\right) - a^{k+1}\left(\sum_{j=0}^{\infty} a^j\right) \\ &= \left(1 - a^{k+1}\right)\left(\sum_{i=0}^{\infty} a^i\right) = \frac{1 - a^{k+1}}{1 - a} \end{aligned} \tag{11-13}$$

Basically, a single-input single-output digital system S is an algorithm for converting one sequence of numbers to another sequence of numbers, as shown in Fig. 11-4(b), where the input sequence is called $x(n)$ and the output sequence is called $y(n)$. Let $y_1(n)$ and $y_2(n)$ be, respectively, the zero-state responses[5] due to the input sequences $x_1(n)$ and $x_2(n)$. Then S is said to be *linear* if the zero-state output sequence $y(n)$ due to the input sequence

$$x(n) \triangleq a_1 x_1(n) + a_2 x_2(n) \tag{11-14a}$$

is given by

$$y(n) \triangleq a_1 y_1(n) + a_2 y_2(n) \tag{11-14b}$$

S is said to be *time-invariant* if the zero-state output $y(n)$ due to the input sequence

$$x(n) \triangleq x_1(n - n_0) \tag{11-15a}$$

is given by

$$y(n) \triangleq y_1(n - n_0) \tag{11-15b}$$

Let $h(n)$ be the zero-state response to $\delta(n)$. The time-invariant property of the system leads us to conclude that $h(n - k)$ is the zero-state response to $\delta(n - k)$. By (11-3) and by the linearity property of the system, the zero-state output sequence due to an input sequence of $x(n)$ written as

$$x(n) = \sum_{k=-\infty}^{\infty} x(k)\,\delta(n - k) \tag{11-16}$$

is given by

$$y(n) = \sum_{k=-\infty}^{\infty} x(k)\,h(n - k) \tag{11-17a}$$

This means that a linear and time-invariant digital system S can be characterized by an *impulse response* $h(n)$, which is the output sequence of S when the input is a unit impulse sequence and all initial conditions of S are zero. By a change of variable, (11-17a) can also be written as

$$y(n) = \sum_{k=-\infty}^{\infty} x(n - k)\,h(k) \tag{11-17b}$$

[5] A zero-state response is an output of the system when all initial conditions of the system are zero.

Both equations of (11-17) are called the *convolution sum* of the two sequences $x(n)$ and $h(n)$ and are denoted by

$$y(n) = x(n) * h(n) \tag{11-17c}$$

Finally, a linear and time-invariant digital system S is *stable* if its impulse response $h(n)$ satisfies the condition

$$\sum_{n=-\infty}^{\infty} |h(n)| < \infty \tag{11-18}$$

and is *causal* if

$$h(n) = 0 \quad \text{for } n < 0 \tag{11-19}$$

Note that if (11-18) is violated, then we can find a bounded input sequence $x(n)$, where

$$\sum_{n=-\infty}^{\infty} |x(n)|^2 = K < \infty, \tag{11-20a}$$

to yield an unbounded output sequence $y(n)$ such that

$$\sum_{n=-\infty}^{\infty} |y(n)|^2 \longrightarrow \infty \tag{11-20b}$$

Example 11-1 Let the system S be characterized by[6]

$$y(n) - ay(n-1) = x(n) \tag{11-21a}$$

$$y(-1) = 0 \tag{11-21b}$$

Find the impulse response $h(n)$ of S, and discuss the stability and causality conditions of S.

Solution: Because the initial condition of S is zero, as given by (11-21b), when

$$x(n) = \delta(n) \tag{11-22}$$

the output sequence $y(n)$ will be the impulse response $h(n)$. From (11-21), we obtain

$$y(0) = ay(-1) + \delta(0) = 0 + 1 = 1$$
$$y(1) = ay(0) + \delta(1) = a + 0 = a$$
$$y(2) = ay(1) + \delta(2) = a^2 + 0 = a^2$$

Progressing inductively, we obtain

$$y(n) = a^n \quad \text{for } n \geq 0 \tag{11-23a}$$

To consider the case when $n < -1$, we write (11-21) and (11-22) as

$$y(n-1) = a^{-1}[y(n) - \delta(n)]$$

[6]The zero initial condition of the system S is given by (11-21b), where we have assumed that the initial time is when $n = 0$. Recall that for the continuous-time case, the initial conditions are given at the point $t = 0-$.

with

$$y(-1) = 0$$

This gives

$$y(-2) = a^{-1}[y(-1) - \delta(-1)] = a^{-1}(0 - 0) = 0$$
$$y(-3) = a^{-1}[y(-2) - \delta(-2)] = 0$$

Clearly, we have

$$y(n) = 0 \quad \text{for } n < 0 \tag{11-23b}$$

Hence, the impulse response $h(n)$ of the system S characterized by (11-21) is given by

$$h(n) = a^n u(n) \tag{11-24}$$

In view of (11-24), the system S is causal for all a and is stable when $|a| < 1$. ▮

11-2 Z-TRANSFORM

The z-transform method is a very useful tool in solving linear difference equations. It reduces the solutions of such equations into those of algebraic equations. The application of z-transforms to a set of difference equations is analogous to the application of Laplace transforms to a set of differential equations.

The z-transform $X(z)$ of a sequence $x(n)$ is defined to be[7]

$$X(z) \triangleq \sum_{n=-\infty}^{\infty} x(n)z^{-n} \tag{11-25}$$

where z is a complex variable. Hence, $X(z)$ is complex.

Example 11-2 Find the z-transform of the sequence $x(n)$ given by

$$x(n) = (\cos n\phi + \sin n\phi)u(n) \tag{11-26}$$

Solution: From (11-25), we have

$$\begin{aligned} X(z) &= \sum_{n=-\infty}^{\infty} x(n)z^{-n} = \sum_{n=0}^{\infty} (\cos n\phi + \sin n\phi)z^{-n} \\ &= \sum_{n=0}^{\infty} \left[\frac{e^{jn\phi} + e^{-jn\phi}}{2} + \frac{e^{jn\phi} - e^{-jn\phi}}{2j}\right] z^{-n} \\ &= \sum_{n=0}^{\infty} \left[\frac{1-j}{2} e^{jn\phi} z^{-n}\right] + \sum_{n=0}^{\infty} \left[\frac{1+j}{2} e^{-jn\phi} z^{-n}\right] \end{aligned} \tag{11-27}$$

If

$$|z^{-1}| < 1 \quad \text{or} \quad |z| > 1 \tag{11-28a}$$

then

$$|e^{\pm j\phi}z^{-1}| < 1 \tag{11-28b}$$

---

[7]We use the convention that the z-transform of a time sequence $x(n)$ is denoted by $X(z)$; a time sequence is denoted by a lower-case letter, and its z-transform is denoted by the corresponding upper-case letter.

and (11-27) can be simplified to

$$\begin{aligned} X(z) &= \frac{1-j}{2}\frac{1}{1-e^{j\phi}z^{-1}} + \frac{1+j}{2}\frac{1}{1-e^{-j\phi}z^{-1}} \\ &= \frac{1-(\cos\phi)z^{-1}+(\sin\phi)z^{-1}}{1-(e^{j\phi}+e^{-j\phi})z^{-1}+z^{-2}} \\ &= \frac{1+(\sin\phi-\cos\phi)z^{-1}}{1-2(\cos\phi)z^{-1}+z^{-2}} \end{aligned} \tag{11-29}$$

∎

Clearly, $X(z)$ is defined for those values of z or z^{-1} for which the power series in (11-25) converges. For example, $X(z)$ of (11-27) is defined only if (11-28a) is satisfied. By writing z in its polar form

$$z = re^{j\theta} \tag{11-30}$$

(11-25) becomes

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)r^{-n}e^{-j\theta n} \tag{11-31}$$

Hence, $X(z)$ is defined for those values of z with radius r in the z-plane such that

$$\sum_{n=-\infty}^{\infty} |x(n)r^{-n}| < \infty \tag{11-32}$$

The totality of all z such that (11-32) holds is called the *region of convergence* for the sequence $x(n)$. In the case of Example 11-2, the region of convergence is $r > 1$ in the z-plane.

Example 11-3 Find the region of convergence for the pulse sequence

$$\begin{aligned} h(n) &= a \quad 0 \le n < N-1 \\ &= 0 \quad \text{elsewhere} \end{aligned} \tag{11-33}$$

where a is real.

Solution:

$$H(z) = \sum_{n=-\infty}^{\infty} h(n)z^{-n} = \sum_{n=0}^{N-1} az^{-n}$$

Hence,

$$H(re^{j\theta}) = \sum_{n=0}^{N-1} ar^{-n}e^{-j\theta n} \tag{11-34}$$

Because (11-34) involves only a finite sum (the number of terms in the summation is finite), $H(z)$ is defined for all $r < \infty$. Hence, the region of convergence is the entire z-plane. ∎

Example 11-4 Find the region of convergence for the exponential sequence

$$\begin{aligned} h(n) &= a^n \quad \text{for } 0 \le n < \infty \\ &= 0 \quad \text{for } n < 0 \end{aligned} \tag{11-35}$$

Solution: Because

$$H(re^{j\theta}) = \sum_{n=-\infty}^{\infty} h(n)r^{-n}e^{-j\theta n} = \sum_{n=0}^{\infty} a^n r^{-n} e^{-j\theta n} = \sum_{n=0}^{\infty} (ar^{-1})^n e^{-j\theta n} \quad (11\text{-}36)$$

the region of convergence is the values of z with radius r such that

$$\sum_{n=0}^{\infty} \left|\frac{a}{r}\right|^n < \infty \quad (11\text{-}37)$$

Clearly, (11-37) is satisfied if and only if

$$\left|\frac{a}{r}\right| < 1 \quad (11\text{-}38)$$

Hence, the region of convergence of $h(n)$ of (11-35) is the exterior of a circle with radius $|a|$ in the z-plane, as shown in Fig. 11-5(a). ∎

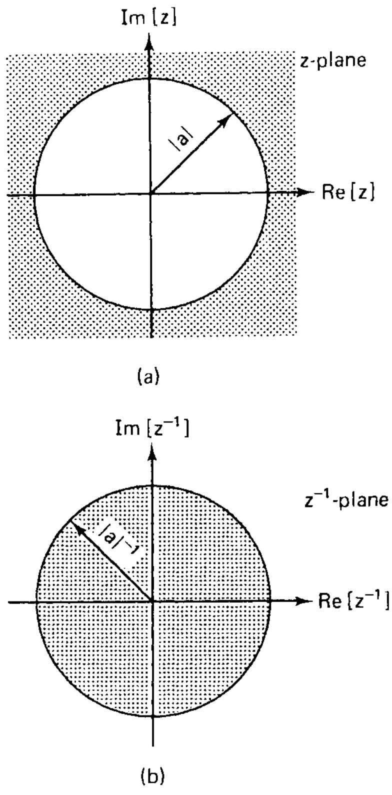


Fig. 11-5 Region of convergence for an exponential sequence. (a) In the z-plane. (b) In the z^{-1}-plane.

The region of convergence for a causal sequence $x(n)$—with $x(n) = 0$ for $n < 0$—is everywhere outside a certain circle with radius R in the z-plane.[8] A case in point is given by Example 11-4. The value of R depends on the pole locations of $X(z)$.[9] For the sequence considered in Example 11-4, the z-transform $H(z)$ of the sequence $h(n)$ is given by

$$\begin{aligned} H(z) &= \sum_{n=0}^{\infty} a^n z^{-n} \\ &= \frac{1}{1 - az^{-1}} \end{aligned} \tag{11-39}$$

Hence, the pole of $H(z)$ is located at the point $z = a$, which is the boundary of the region of convergence for the sequence.

In most physical digital systems including digital filters, causal sequences form the basis of all signals involved in the processing. For convenience, the z-transforms of some of the frequently used causal sequences are listed in Table 11-1, together with their regions of convergence. In general, we will assume that we are working within the area in the z-plane where the z-transforms of all sequences involved are defined, and hence we can ignore the problems associated with the regions of convergence.

From Table 11-1, we observe that the z-transform of a sequence is a rational function of either z or z^{-1}. Thus, if we know the poles and zeros of the z-transform $X(z)$ of a sequence $x(n)$, we can construct $X(z)$ up to a constant multiple rather easily. For example, if $X(z)$ has poles $p_1, p_2, \ldots, p_N$ and zeros $z_1, z_2, \ldots, z_M$, then $X(z)$ can be written in the factored form as:

$$X(z) = \frac{\alpha \prod_{i=1}^{M} (1 - z_i z^{-1})}{\prod_{k=1}^{N} (1 - p_k z^{-1})} \tag{11-40a}$$

or

$$X(z) = \frac{\alpha z^{(N-M)} \prod_{i=1}^{M} (z - z_i)}{\prod_{k=1}^{N} (z - p_k)} \tag{11-40b}$$

where α is a constant. In digital filter applications, (11-40a) is preferred, because a shift register or a unit of a tapped delay line is an implementation

---

[8]The region of convergence can be located in the z^{-1}-plane also. For a causal sequence, the region of convergence is everywhere inside a certain circle with radius $\tilde{R}$ in the z^{-1}-plane. For example, the region of convergence of the exponential sequence in Example 11-4 is everywhere inside the circle with radius $|a|^{-1}$ in the z^{-1}-plane, as shown in Fig. 11-5(b).

[9]A pole {zero} of a z-transformed function $X(z)$ is the location z_1 in the z-plane, where $X(z_1) = \infty$ $\{X(z_1) = 0\}$.