CHAPTER 4

Discrete Equivalents to Continuous Transfer Functions: The Digital Filter

4.1 INTRODUCTION

One of the exciting fields of application of digital systems[1] is in signal processing and digital filtering. A filter is a device designed to pass desirable elements and hold back or reject undesirable ones; in signal processing it is common to represent signals as a sum of sinusoids and to define the "desirable elements" as those signals whose frequency components are in a specified band. Thus a radio receiver filter passes the band of frequencies transmitted by the station we want to hear and rejects all others. We would call such a filter a *bandpass filter.* In electrocardiography it often happens that power-line frequency signals are strong and unwanted, so we design a filter to pass signals between 1 and 500 Hz but to eliminate those at 60 Hz. The magnitude of the transfer function for this purpose may look like Fig. 4.1 on a log-frequency scale, where the amplitude response between 59.5 and 60.5 Hz might reach 10^{-3}. Here we have a band-reject filter with a 60-dB rejection ratio in a 1-Hz band centered at 60 Hz.

In long-distance telephony some filters play a conceptually different role. There the issue is that transmission media—wires or microwaves—introduce distortion in the amplitude and phase of a sinusoid which must be removed. Filters to accomplish this correction are called *equalizers.* And in control we must control systems whose dynamics require modification in order that the complete system have satisfactory dynamic response. We call the devices that make these changes *compensators.*

---

[1]Including microprocessors and special-purpose, very large-scale integration (VLSI) digital chips for signal processing, called DSP chips.

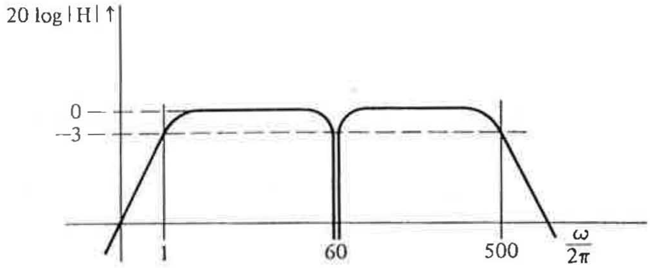

Figure 4.1 Magnitude of a low-frequency bandpass filter with a narrow rejection band.

Whatever the name—filter, equalizer, or compensator—many fields have found use for devices having specified characteristics of amplitude and phase transmission, and the trend is to perform these functions by digital means. The design of electronic filters is a well-established subject that includes not only very sophisticated techniques but also well-tested computer programs [Van Valkenburg(1982)]. Much of the effort in digital filter design has been directed toward the design of digital filters that have the same characteristics (as nearly as possible) as those of a satisfactory continuous design. For digital control systems we have much the same motivation: Because continuous-control designs are well established, we should like to know how to take advantage of a good continuous design and cause a digital computer to produce a discrete equivalent to the continuous compensator. This method of design is called *emulation.* Although much of our presentation in this book is oriented toward direct digital design and away from emulation of continuous designs with digital equivalents, it is important to understand the techniques of discrete equivalents for purposes of comparison and because it is widely used by practicing engineers.

Thus we are led to the specific problem of this chapter: Given a transfer function, $H(s)$, what discrete transfer function will have approximately the same characteristics? We will present here three approaches to this task:

Method 1: *numerical integration.*

Method 2: *pole-zero mapping.*

Method 3: *hold equivalence.*

4.2 DESIGN OF DISCRETE EQUIVALENTS BY NUMERICAL INTEGRATION

The topic of numerical integration of differential equations is quite complex, and only the most elementary techniques are presented here. For example, we only consider formulas of low complexity and fixed step-size. The fundamental concept is to represent the given filter transfer function $H(s)$ as a differential equation and to derive a difference equation whose solution is an approximation to that of the differential equation. For example, the system

$$\frac{U(s)}{E(s)} = H(s) = \frac{a}{s+a} \tag{4.1}$$

is equivalent to the differential equation

$$\dot{u} + au = ae. \tag{4.2}$$

Now, if we write (4.2) in integral form, we have a development much like that of (2.5) in Chapter 2, except that the integral is more complex here:

$$\begin{aligned} u(t) &= \int_0^t [-au(\tau) + ae(\tau)]d\tau, \\ u(kT) &= \int_0^{kT-T} [-au + ae]d\tau + \int_{kT-T}^{kT} [-au + ae]d\tau \\ &= u(kT-T) + \left\{ \begin{array}{l} \text{area of } -au + ae \\ \text{over } kT - T \le \tau < kT \end{array} \right\}. \end{aligned} \tag{4.3}$$

We can now develop many rules based on our selection of the approximation of the incremental area term. The first approximation leads to the forward rectangular rule[2] wherein we approximate the area by the rectangle looking forward from $kT-T$ and take the amplitude of the rectangle to be the value of the integrand at $kT-T$. The width of the rectangle is T. The result is an equation in the first approximation, u_1:

$$\begin{aligned} u_1(kT) &= u_1(kT-T) + T[-au_1(kT-T) + ae(kT-T)] \\ &= (1-aT)u_1(kT-T) + aTe(kT-T). \end{aligned} \tag{4.4a}$$

[2]Also known as *Euler's rule*.

The transfer function corresponding to the forward rectangular rule in this case is

$$\begin{aligned} H_F(z) &= \frac{aTz^{-1}}{1-(1-aT)z^{-1}} \\ &= \frac{a}{(z-1)/T+a} \quad \text{(forward rectangular rule).} \end{aligned} \tag{4.4b}$$

A second rule follows from taking the amplitude of the approximating rectangle to be the value looking backward from kT toward $kT-T$, namely, $-au(kT) + ae(kT)$. The equation for u_2, the second approximation,[3] is

$$\begin{aligned} u_2(kT) &= u_2(kT-T) + T[-au_2(kT) + ae(kT)] \\ &= \frac{u_2(kT-T)}{1+aT} + \frac{aT}{1+aT}e(kT). \end{aligned} \tag{4.5a}$$

Again we take the z-transform and compute the transfer function of the backward rule:

$$\begin{aligned} H_B(z) &= \frac{aT}{1+aT}\frac{1}{1-z^{-1}/(1+aT)} = \frac{aTz}{z(1+aT)-1} \\ &= \frac{a}{(z-1)/Tz+a} \quad \text{(backward rectangular rule).} \end{aligned} \tag{4.5b}$$

Our final version of integration rules is the *trapezoid rule* found by taking the area approximated in (4.3) to be that of the trapezoid formed by the average of the previously selected rectangles. The approximating difference equation is

$$\begin{aligned} u_3(kT) &= u_3(kT-T) \\ &\quad + \frac{T}{2}[-au_3(kT-T) + ae(kT-T) - au_3(kT) + ae(kT)] \\ &= \frac{1-(aT/2)}{1+(aT/2)}u_3(kT-T) + \frac{aT/2}{1+(aT/2)}[e_3(kT-T) + e_3(kT)]. \end{aligned} \tag{4.6a}$$

---

[3]It is worth noting that in order to solve for (4.5a) we had to eliminate $u(kT)$ from the right-hand side where it entered from the integrand. Had (4.2) been nonlinear, the result would have been an implicit equation requiring an iterative solution. This topic is the subject of predictor-corrector rules, which are beyond our scope of interest. A discussion is found in most books on numerical analysis. See Golub and Van Loan(1983).

The corresponding transfer function from the trapezoid rule is

$$\begin{aligned} H_T(z) &= \frac{aT(z+1)}{(2+aT)z+aT-2} \\ &= \frac{a}{(2/T)[(z-1)/(z+1)]+a} \quad \text{(trapezoid rule).} \end{aligned} \tag{4.6b}$$

Suppose we tabulate our results obtained thus far.

$H(s)$	Method	Transfer function
$\frac{a}{s+a}$	Forward rule	$H_F = \frac{a}{(z-1)/T+a}$
$\frac{a}{s+a}$	Backward rule	$H_B = \frac{a}{(z-1)/Tz+a}$
$\frac{a}{s+a}$	Trapezoid rule	$H_T = \frac{a}{(2/T)\{(z-1)/(z+1)\}+a}$

(4.7)

From direct comparison of $H(s)$ with the three approximations in this tabulation we can see that the effect of each of our methods is to present a discrete transfer function that can be obtained from the given Laplace transfer function $H(s)$ by substitution of an approximation for the frequency variable as shown below:

Method	Approximation
Forward rule	$s \leftarrow \frac{z-1}{T}$
Backward rule	$s \leftarrow \frac{z-1}{Tz}$
Trapezoid rule	$s \leftarrow \frac{2}{T}\frac{z-1}{z+1}$

(4.8)

The trapezoid-rule substitution is also known, especially in digital and sampled-data control circles, as *Tustin's method* [Tustin (1947)] after the British engineer whose work on nonlinear circuits stimulated a great deal of interest in this approach. The transformation is also called the *bilinear transformation* from consideration of its mathematical form. The design method can be summarized by stating the rule: Given a continuous transfer function (filter), $H(s)$, a discrete equivalent can be found by the substitution

$$H_T(z) = H(s)|_{s=(2/T)[(z-1)/(z+1)]}. \tag{4.9}$$

Each of the approximations given in (4.8) can be viewed as a map from the s-plane to the z-plane. A further understanding of the maps can be obtained by considering them graphically. For example, because the $(s = j\omega)$-axis is the boundary between poles of stable systems and poles of unstable systems, it would be interesting to know how the $j\omega$-axis is mapped by the three rules and where the left (stable) half of the s-plane appears in the z-plane. For this purpose we must solve the relations in (4.8) for z in terms of s. We find

$$\begin{aligned} &\text{i)} \quad z = 1 + Ts, && \text{forward rectangular rule,} \\ &\text{ii)} \quad z = \frac{1}{1 - Ts}, && \text{backward rectangular rule,} \\ &\text{iii)} \quad z = \frac{1 + Ts/2}{1 - Ts/2}, && \text{bilinear rule.} \end{aligned} \tag{4.10}$$

If we let $s = j\omega$ in these equations, we obtain the boundaries of the regions in the z-plane which originate from the stable portion of the s-plane. The shaded areas sketched in the z-plane in Fig. 4.2 are these stable regions for each case. To show that rule (ii) results in a circle, $\frac{1}{2}$ is added to and subtracted from the right-hand side to yield

$$\begin{aligned} z &= \frac{1}{2} + \left\{ \frac{1}{1 - Ts} - \frac{1}{2} \right\} \\ &= \frac{1}{2} - \frac{1}{2} \frac{1 + Ts}{1 - Ts}. \end{aligned} \tag{4.11}$$

Now it is easy to see that with $s = j\omega$, the magnitude of $z - \frac{1}{2}$ is constant,

$$|z - \frac{1}{2}| = \frac{1}{2},$$

and the curve is thus a circle as drawn in Fig. 4.2(b). Because the unit circle is the stability boundary in the z-plane, it is apparent from Fig. 4.2 that the forward rectangular rule could cause a stable continuous filter to be mapped into an unstable digital filter.

It is especially interesting to notice that the bilinear rule maps the stable region of the s-plane exactly into the stable region of the z-plane although the entire $j\omega$-axis of the s-plane is stuffed into the 2π-length of the unit circle! Obviously a great deal of distortion takes place in the mapping in spite of the congruence of the stability regions. As our final rule deriving from numerical

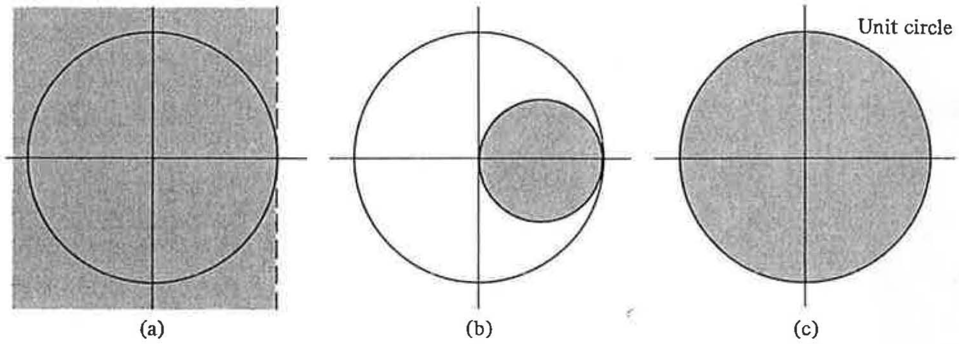

Figure 4.2 Maps of the left-half s-plane to the z-plane by the integration rules of (4.8). Stable s-plane poles map into the shaded regions in the z-plane. The unit circle is shown for reference. (a) Forward rectangular rule. (b) Backward rectangular rule. (c) Trapezoid or bilinear rule.

integration ideas, we discuss a formula that extends Tustin's rule one step in an attempt to correct for the inevitable distortion of real frequencies mapped by the rule. We begin with our elementary transfer function (4.1) and consider the bilinear rule approximation

$$H_T(z) = \frac{a}{(2/T)[(z-1)/(z+1)] + a}.$$

The original $H(s)$ had a pole at $s = -a$, and for real frequencies, $s = j\omega$, the magnitude of $H(j\omega)$ is given by

$$\begin{aligned} |H(j\omega)|^2 &= \frac{a^2}{\omega^2 + a^2} \\ &= \frac{1}{\omega^2/a^2 + 1}. \end{aligned} \tag{4.12}$$

Thus our reference filter has a half-power point, $|H|^2 = \frac{1}{2}$, at $\omega = a$. It will be interesting to know where $H_T(z)$ has a half-power point.

As we saw in Chapter 2, signals with poles on the imaginary axis in the s-plane (sinusoids) map into signals on the unit circle of the z-plane. A sinusoid of frequency ω_1 corresponds to $z_1 = e^{j\omega_1 T}$, and the response of $H_T(z)$ to a sinusoid of frequency ω_1 is $H_T(z_1)$. We consider now (4.6b) for $H_T(z_1)$ and manipulate it into a more convenient form for our present purposes:

$$H_T(z_1) = a/\left(\frac{2}{T}\frac{e^{j\omega_1 T}-1}{e^{j\omega_1 T}+1}+a\right)$$

$$= a/\left(\frac{2}{T}\frac{e^{j\omega_1 T/2}-e^{-j\omega_1 T/2}}{e^{j\omega_1 T/2}+e^{-j\omega_1 T/2}}+a\right)$$

$$= a/\left(\frac{2}{T}j\tan\frac{\omega_1 T}{2}+a\right). \tag{4.13}$$

The magnitude squared of H_T will be $\frac{1}{2}$ when

$$\frac{2}{T}\tan\frac{\omega_1 T}{2} = a$$

or

$$\tan\frac{\omega_1 T}{2} = \frac{aT}{2}. \tag{4.14}$$

Equation (4.14) is a measure of the frequency distortion or warping caused by Tustin's rule. Whereas we wanted to have a half-power point at $\omega = a$, we realized a half-power point at $\omega_1 = (2/T)\tan^{-1}(aT/2)$. ω_1 will be approximately correct only if $aT/2 \ll 1$ so that $\tan^{-1}(aT/2) \cong aT/2$, that is, if $\omega_s(= 2\pi/T) \gg a$ and the sample rate is much faster than the half-power frequency. We can turn our intentions around and suppose that we really want the half-power point to be at ω_1. Equation (4.14) can be made into an equation of prewarping: If we select a according to (4.14), then, using Tustin's bilinear rule for the design, the half-power point will be at ω_1. A statement of a complete set of rules for filter design via bilinear transformation with prewarping is:

a) Write the desired filter characteristic with transform variable s and critical frequency ω_1 in the form $H(s/\omega_1)$.[4]

[4]The critical frequency need not be the band edge. We can use the band center of a bandpass filter or the crossover frequency of a Bode plot compensator. However, we must have $\omega_1 < \pi/T$ if a stable filter is to remain stable after warping.

b) Replace ω_1 by a such that

$$a = \frac{2}{T}\tan\frac{\omega_1 T}{2},$$

and in place of $H(s/\omega_1)$, consider the prewarped function $H(s/a)$. For more complicated shapes, such as bandpass filters, the specification frequencies, such as band edges and center frequency, should be prewarped before the continuous design is done; and then the bilinear transformation will bring all these points to their correct frequencies in the digital filter.

c) Substitute

$$s = \frac{2}{T}\frac{z-1}{z+1}$$

in $H(s/a)$ to obtain the prewarped equivalent $H_p(z)$.

As a frequency substitution the result can be expressed as

$$H_p(z) = H(s/\omega_1)\big|_{s=\frac{\omega_1}{\tan(\omega_1 T/2)}\frac{z-1}{z+1}} \tag{4.15}$$

It is clear from (4.15) that when $\omega = \omega_1$, $H_p(z_1) = H(j1)$ and the discrete filter has exactly the same transmission at ω_1 as the continuous filter has at this frequency. This is the consequence of prewarping. We also note that as the sampling period gets small, $H_p(z)$ approaches $H(j\omega/\omega_1)$.

Example 4.1: The frequency responses of discrete filters designed by the four rules studied in this section are plotted in Fig. 4.3. The filters are discrete equivalents of a third-order lowpass Butterworth[5] filter designed to have unity pass bandwidth ($\omega_p = 1$). A simple frequency scaling would of course translate these curves to any desired passband frequency. The continuous filter transfer function is

$$H(s) = \frac{1}{s^3 + 2s^2 + 2s + 1}.$$

[5]A description of the properties of Butterworth filters is given in most books on filter design and briefly in Franklin, Powell and Emami-Naeini(1986).

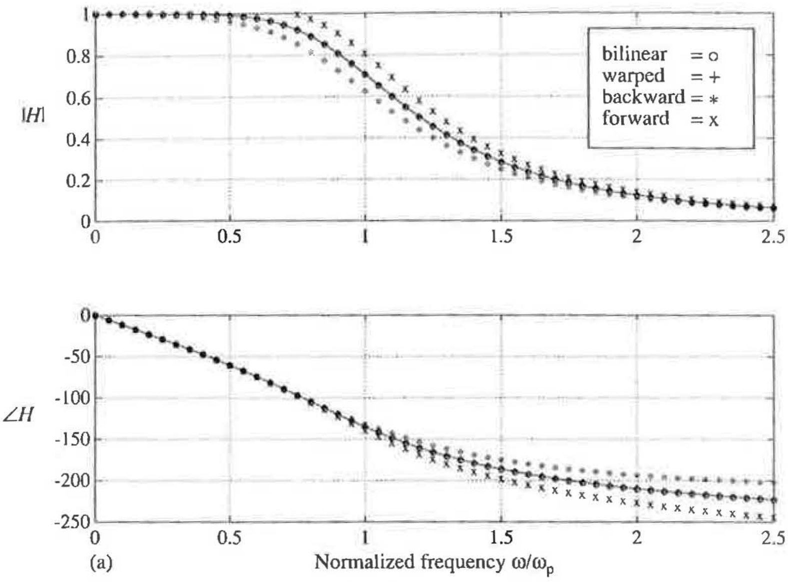

Figure 4.3(a) Response of third-order lowpass filter and digital equivalents for $\omega_s/\omega_p = 20\pi$.

Computation of the discrete equivalents is numerically tedious, and the state-space algorithms described below were used to generate the functions and the response curves. In Fig 4.3(a) we see that at a high sample rate ($T = 0.1$), where the ratio of sampling frequency to passband frequency is $\omega_s/\omega_p \approx 63$, all the rules do reasonably well but the rectangular rules are already showing some deviation.

From Fig. 4.3(b) we see that at $\omega_s/\omega_p = 2\pi$ the rectangular rules are useless (the forward rule is unstable).

Finally, in Fig. 4.3(c) at $\omega_s/\omega_p = \pi$ with a sampling period of $T = 2$ sec we see that only with prewarping do we have a design that comes even close to the continuous response. Notice in this plot that at the Nyquist radian frequency, $\omega = \pi/T$, the magnitude response of the discrete filter starts to repeat according to the periodic nature of a discrete-transfer-function frequency response. Notice also that the magnitude and phase of the prewarped designs match those of the continuous filter exactly at the band edge, $\omega = 1$ for all these cases. This is no surprise, because such matching was the whole idea of prewarping.

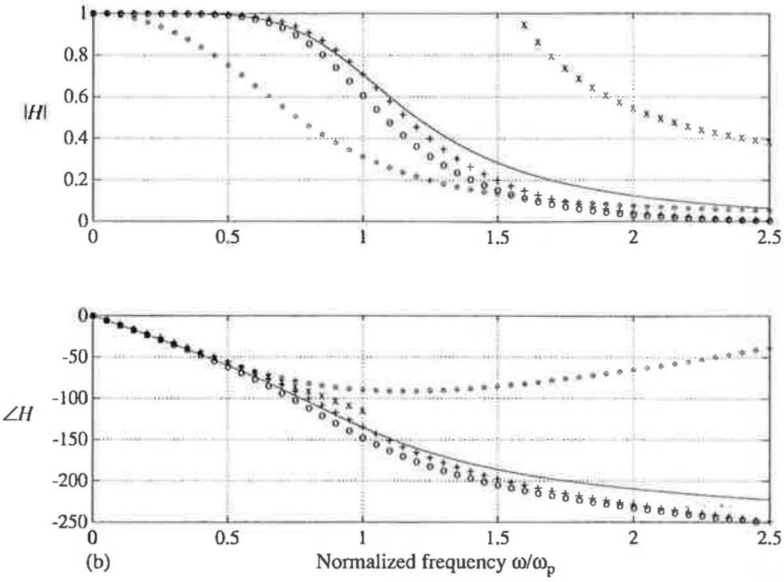

Figure 4.3(b) Response of third-order lowpass filter and digital equivalents for $\omega_s/\omega_p = 2\pi$.

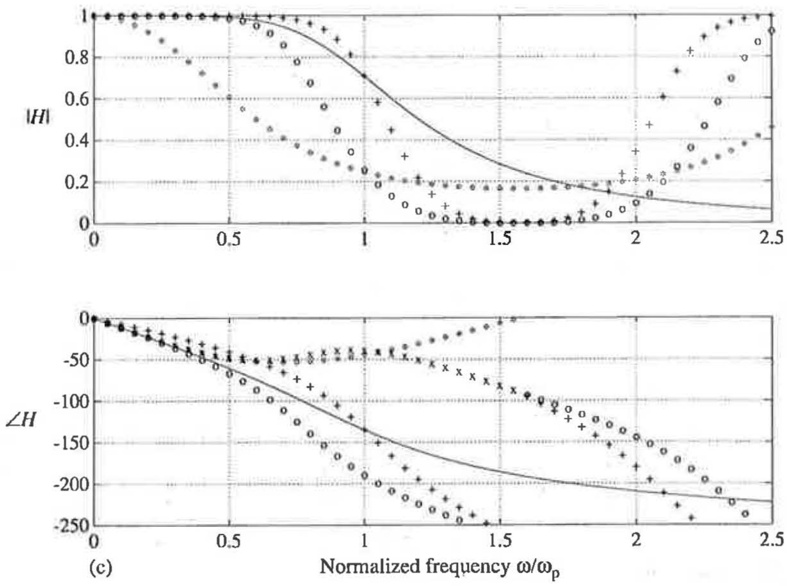

Figure 4.3(c) Response of third-order lowpass filter and digital equivalents for $\omega_s/\omega_p = \pi$. Bilinear = 0, warp = +, backward = $*$, and forward = $\times$.

These formulas for discrete equivalents are particularly simple and convenient when expressed in state-variable form and used with a computer-aided design package. For example, suppose we have a vector-matrix description of a continuous design in the form of the equations

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}e,\\ u &= \mathbf{C}\mathbf{x} + \mathbf{D}e.\end{aligned} \tag{4.16}$$

The Laplace transform of this equation is

$$\begin{aligned}s\mathbf{X} &= \mathbf{A}\mathbf{X} + \mathbf{B}E,\\ U &= \mathbf{C}\mathbf{X} + \mathbf{D}E.\end{aligned} \tag{4.17}$$

We can now make the substitution for s in (4.17) of any of the forms in z corresponding to an integration rule. Suppose we start with the forward rectangular rule, for which we replace s with $(z-1)/T$ from (4.8):

$$\begin{aligned}\frac{z-1}{T}\mathbf{X} &= \mathbf{A}\mathbf{X} + \mathbf{B}E,\\ U &= \mathbf{C}\mathbf{X} + \mathbf{D}E.\end{aligned} \tag{4.18}$$

In the time domain, the operator z corresponds to forward shift; that is, $zx(k) = x(k+1)$. Thus the corresponding discrete equations in the time domain are

$$\begin{aligned}\mathbf{x}(k+1) - \mathbf{x}(k) &= T\mathbf{A}\mathbf{x}(k) + T\mathbf{B}e(k),\\ \mathbf{x}(k+1) &= (\mathbf{I} + T\mathbf{A})\mathbf{x}(k) + T\mathbf{B}e(k),\\ u &= \mathbf{C}\mathbf{x} + \mathbf{D}e.\end{aligned} \tag{4.19}$$

Equation (4.19) is a state-space formula for the forward rule equivalent.

For the backward rule, we substitute $s \leftarrow (z-1)/zT$ with the result

$$\frac{z-1}{Tz}\mathbf{X} = \mathbf{A}\mathbf{X} + \mathbf{B}E,$$

and in the time domain,

$$\mathbf{x}(k+1) - \mathbf{x}(k) = T\mathbf{A}\mathbf{x}(k+1) + T\mathbf{B}e(k+1). \tag{4.20}$$

In this equation, we have terms in $k+1$ on both the right- and left-hand sides. In order to get an equation with such terms only on the left, we transpose all $k+1$ terms to the left and define them as a new state vector:

$$\begin{aligned}\mathbf{x}(k+1) - T\mathbf{A}\mathbf{x}(k+1) - T\mathbf{B}e(k+1) &= \mathbf{x}(k) \\ &\triangleq \mathbf{w}(k+1). \end{aligned} \tag{4.21}$$

From this equation, solving for $\mathbf{x}$ in terms of $\mathbf{w}$ and e, we obtain

$$\begin{aligned}(\mathbf{I} - \mathbf{A}T)\mathbf{x} &= \mathbf{w} + T\mathbf{B}e \\ \mathbf{x} &= (\mathbf{I} - \mathbf{A}T)^{-1}\mathbf{w} + (\mathbf{I} - \mathbf{A}T)^{-1}\mathbf{B}Te. \end{aligned} \tag{4.22}$$

With this expression for $\mathbf{x}$, we can rewrite (4.21) in standard form as

$$\mathbf{w}(k+1) = (\mathbf{I} - \mathbf{A}T)^{-1}\mathbf{w}(k) + (\mathbf{I} - \mathbf{A}T)^{-1}\mathbf{B}Te(k), \tag{4.23a}$$

and the output equation is now

$$u(k) = \mathbf{C}(\mathbf{I} - \mathbf{A}T)^{-1}\mathbf{w} + \{\mathbf{D} + \mathbf{C}(\mathbf{I} - \mathbf{A}T)^{-1}\mathbf{B}T\}e. \tag{4.23b}$$

Equation (4.23) is a state-space description of the backward rule equivalent to (4.16).

Finally, for the trapezoid or bilinear rule, the z-transform equivalent is obtained from

$$\begin{aligned}\frac{2(z-1)}{T(z+1)}\mathbf{X} &= \mathbf{A}\mathbf{X} + \mathbf{B}E \\ (z-1)\mathbf{X} &= \frac{\mathbf{A}T}{2}(z+1)\mathbf{X} + \frac{\mathbf{B}T}{2}(z+1)E \\ U &= \mathbf{C}\mathbf{X} + \mathbf{D}E, \end{aligned} \tag{4.24}$$

and the time domain equation for the state is

$$\mathbf{x}(k+1) - \mathbf{x}(k) = \frac{\mathbf{A}T}{2}(\mathbf{x}(k+1) + \mathbf{x}(k)) + \frac{\mathbf{B}T}{2}(e(k+1) + e(k)). \tag{4.25}$$

Once more, we collect all the $k+1$ terms onto the left and define these as $\mathbf{w}(k+1)$ as follows:[6]

$$\mathbf{x}(k+1) - \frac{\mathbf{A}T}{2}\mathbf{x}(k+1) - \frac{\mathbf{B}T}{2}e(k+1) = \mathbf{x}(k) + \frac{\mathbf{A}T}{2}\mathbf{x}(k) + \frac{\mathbf{B}T}{2}e(k)$$
$$\triangleq \sqrt{T}\mathbf{w}(k+1). \tag{4.26}$$

Writing the definition of $\mathbf{w}$ at time k, we can solve for $\mathbf{x}$ as before:

$$\left(\mathbf{I} - \frac{\mathbf{A}T}{2}\right)\mathbf{x} = \sqrt{T}\mathbf{w} + \frac{\mathbf{B}T}{2}e$$
$$\mathbf{x} = \left(\mathbf{I} - \frac{\mathbf{A}T}{2}\right)^{-1}\sqrt{T}\mathbf{w} + \left(\mathbf{I} - \frac{\mathbf{A}T}{2}\right)^{-1}\frac{\mathbf{B}T}{2}e. \tag{4.27}$$

Substituting (4.27) into (4.26), we get

$$\sqrt{T}\mathbf{w}(k+1) = \left(\mathbf{I} + \frac{\mathbf{A}T}{2}\right)\left(\mathbf{I} - \frac{\mathbf{A}T}{2}\right)^{-1}\left\{\sqrt{T}\mathbf{w}(k) + \frac{\mathbf{B}T}{2}e\right\} + \frac{\mathbf{B}T}{2}e(k)$$
$$\mathbf{w}(k+1) = \left(\mathbf{I} + \frac{\mathbf{A}T}{2}\right)\left(\mathbf{I} - \frac{\mathbf{A}T}{2}\right)^{-1}\mathbf{w}(k)$$
$$+ \left\{\left(\mathbf{I} + \frac{\mathbf{A}T}{2}\right)\left(\mathbf{I} - \frac{\mathbf{A}T}{2}\right)^{-1} + \mathbf{I}\right\}\frac{\mathbf{B}\sqrt{T}}{2}e(k)$$
$$= \left(\mathbf{I} + \frac{\mathbf{A}T}{2}\right)\left(\mathbf{I} - \frac{\mathbf{A}T}{2}\right)^{-1}\mathbf{w}(k) + \left(\mathbf{I} - \frac{\mathbf{A}T}{2}\right)^{-1}\mathbf{B}\sqrt{T}e(k). \tag{4.28}$$

In following this algebra, it is useful to know that in deriving the last part of (4.28), we expressed the identity $\mathbf{I}$ as $(\mathbf{I} - \frac{\mathbf{A}T}{2})(\mathbf{I} - \frac{\mathbf{A}T}{2})^{-1}$ and factored out $(\mathbf{I} - \frac{\mathbf{A}T}{2})^{-1}$ on the right.

To obtain the output equation for the bilinear equivalent, we substitute (4.27) into the second part of (4.24):

$$u(k) = \sqrt{T}\mathbf{C}\left(\mathbf{I} - \frac{\mathbf{A}T}{2}\right)^{-1}\mathbf{w}(k) + \left\{\mathbf{D} + \mathbf{C}\left(\mathbf{I} - \frac{\mathbf{A}T}{2}\right)^{-1}\frac{\mathbf{B}T}{2}\right\}e(k).$$

[6] The scale factor of $\sqrt{T}$ is introduced so that the gain of the discrete equivalent will be balanced between input and output, a rather technical condition. See Al Saggaf and Franklin (1986) for many more details.

These results can be tabulated for convenient reference. Suppose we have a continuous system described by

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}e(t),$$
$$u(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}e(t).$$

Then a discrete equivalent at sampling period T will be described by

$$\mathbf{w}(k+1) = \mathbf{\Phi}\mathbf{w}(k) + \mathbf{\Gamma}e(k),$$
$$u(k) = \mathbf{H}\mathbf{w}(k) + \mathbf{J}e(k),$$

where $\mathbf{\Phi}$, $\mathbf{\Gamma}$, $\mathbf{H}$, and $\mathbf{J}$ are given as follows:[7]

	Forward	Backward	Bilinear
$\mathbf{\Phi}$	$(\mathbf{I}+\mathbf{A}T)$	$(\mathbf{I}-\mathbf{A}T)^{-1}$	$(\mathbf{I}+\frac{\mathbf{A}T}{2})(\mathbf{I}-\frac{\mathbf{A}T}{2})^{-1}$
$\mathbf{\Gamma}$	$\mathbf{B}T$	$(\mathbf{I}-\mathbf{A}T)^{-1}\mathbf{B}T$	$(\mathbf{I}-\frac{\mathbf{A}T}{2})^{-1}\mathbf{B}\sqrt{T}$
$\mathbf{H}$	$\mathbf{C}$	$\mathbf{C}(\mathbf{I}-\mathbf{A}T)^{-1}$	$\sqrt{T}\mathbf{C}(\mathbf{I}-\frac{\mathbf{A}T}{2})^{-1}$
$\mathbf{J}$	$\mathbf{D}$	$\mathbf{D}+\mathbf{C}(\mathbf{I}-\mathbf{A}T)^{-1}\mathbf{B}T$	$\mathbf{D}+\mathbf{C}(\mathbf{I}-\frac{\mathbf{A}T}{2})^{-1}\mathbf{B}T/2$

4.3 ZERO-POLE MAPPING EQUIVALENTS

A very simple but effective method of obtaining a discrete equivalent to a continuous transfer function is to be found by extrapolation of the relation derived in Chapter 2 between the s- and z-planes. If we take the z-transform of samples of a continuous signal $e(t)$, then the poles of the discrete transform $E(z)$ are related to the poles of $E(s)$ according to $z = e^{sT}$. We must go through the z-transform process to locate the zeros of $E(z)$, however. The idea of the zero-pole mapping technique is that the map $z = e^{sT}$ could reasonably be applied to the zeros also. The technique consists of a set of heuristic rules for locating the zeros and poles and setting the gain of a z-transform that will describe a discrete, equivalent transfer function that approximates the given $H(s)$. The rules are as follows:

1. All poles of $H(s)$ are mapped according to $z = e^{sT}$. If $H(s)$ has a pole at $s = -a$, then $H_{zp}(z)$ has a pole at $z = e^{-aT}$. If $H(s)$ has a pole at $-a + jb$ then $H_{zp}(z)$ has a pole at $re^{j\theta}$, where $r = e^{-aT}$ and $\theta = bT$.

[7]These formulas are easily implemented in CAD packages; for example, see EQUIVNT in Table E.1 in Appendix E.

2. All *finite* zeros are also mapped by $z = e^{sT}$. If $H(s)$ has a zero at $s = -a$, then $H_{zp}(z)$ has a zero at $z = e^{-aT}$, and so on.
3. The zeros of $H(s)$ at $s = \infty$ are mapped in $H_{zp}(z)$ to the point $z = -1$.
   a) One zero of $H(s)$ at $s = \infty$ is mapped into $z = \infty$. That is, $H_{zp}(z)$ is left with the number of zeros one less than the number of poles in the finite plane. The series expansion of $H(z)$ in powers of z^{-1} will have no constant term, and thus the corresponding pulse response, $h(k)$, has a one-unit delay. This choice means that the computer has one sample period to do the computation that corresponds to the transfer function because there is no direct transmission term.
4. The gain of the digital filter is selected to match the gain of $H(s)$ at the band center or a similar critical point. In most control applications, the critical frequency is $s = 0$, and hence we typically select the gain so that

$$H(s)|_{s=0} = H_{zp}(z)|_{z=1}.$$

The rationale behind rule 3 is that the map of real frequencies from $j\omega = 0$ to increasing ω is onto the unit circle at $z = e^{j0} = 1$ until $z = e^{j\pi} = -1$. Thus the point $z = -1$ represents, in a real way, the highest frequency possible in the discrete transfer function, so it is appropriate that if $H(s)$ is zero at the highest (continuous) frequency, $|H_{zp}(z)|$ should be zero at $z = -1$, the highest frequency that can be processed by the digital filter.

Example 4.2: Application of these rules to $H(s) = a/(s+a)$ gives

$$H_{zp}(z) = \frac{(z+1)(1-e^{-aT})}{2(z-e^{-aT})}, \tag{4.29}$$

or, using rule 3(a), we get

$$H_{zp}(z) = \frac{1-e^{-aT}}{z-e^{-aT}}. \tag{4.30}$$

A state-space algorithm to generate the zero-pole equivalent is also readily constructed with the utilities of a control design package.[8] The frequency

[8]See EQUIVNT in Table E.1.

response of the zero-pole equivalent was found for the third-order case of Example 4.1 and is shown in Fig. 4.8 along with other equivalents for purposes of comparison.

4.4 HOLD EQUIVALENTS

For this technique, we construct the situation sketched in Fig. 4.4. The purpose of the samplers in Fig. 4.4(b) is to require that $H_{h0}(z)$ has a discrete signal at its input and produces a discrete signal at its output, and thus $H_{h0}(z)$ can be realized as a discrete transfer function. The philosophy of the design is the following. We are asked to design a discrete system that, with an input consisting of *samples* of $e(t)$, has an output that approximates the output of the continuous filter $H(s)$ whose input is the *continuous* $e(t)$. We generate the discrete equivalent by first approximating $e(t)$ from the samples $e(k)$ and then putting this $\hat{e}(t)$ through the given $H(s)$. First consider the possibilities for approximation. These are techniques for taking a sequence of samples and extrapolating or holding them to produce a continuous signal.[9] Suppose we have the $e(t)$ as sketched in Fig. 4.5. This figure also shows a sketch of a piecewise constant approximation to $e(t)$ obtained by the operation of holding $\hat{e}(t)$ constant at $e(k)$ over the interval from kT to $(k+1)T$. This operation is the *zero-order hold* (or ZOH) we've discussed before. If we use a first-order polynomial for extrapolation, we have a *first-order hold* (or FOH), and so on for second-, and nth-order holds.

4.4.1 Zero-Order Hold Equivalent

If the approximating hold is the zero-order hold, then we have for our approximation exactly the same situation that in Chapter 2 was analyzed as a sampled-data system.[10] Therefore, the zero-order-hold equivalent to $H(s)$ is given by (2.39), or

$$H_{h0}(z) = (1 - z^{-1})\mathcal{Z}\left\{\frac{H(s)}{s}\right\}. \tag{4.31}$$

[9]Some books on digital-signal processing suggest using no hold at all, using the equivalent $H(z) = \mathcal{Z}\{H(s)\}$. This choice is called the z-transform equivalent.

[10]Recall that we noticed in Chapter 3 that the signal $\hat{e}$ is, on the average, delayed from e by $T/2$ sec. The size of this delay is one measure of the quality of the approximation and can be used as a guide to the selection of T.

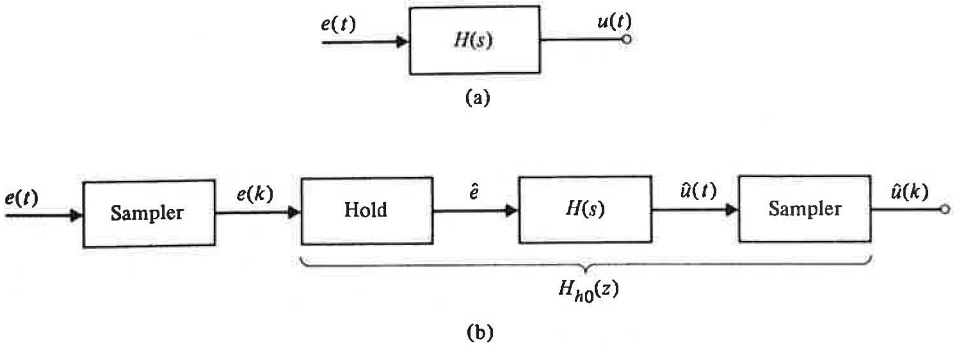

Figure 4.4 System construction for hold equivalents. a) A continuous transfer function. b) Block diagram of an equivalent system.

An example will fix ideas.

Example 4.3: Suppose we again take the first-order filter

$$H(s) = \frac{a}{s+a}.$$

Then

$$\frac{H(s)}{s} = \frac{a}{s(s+a)} = \frac{1}{s} - \frac{1}{s+a}$$

and

$$\mathcal{Z}\left\{\frac{H(s)}{s}\right\} = \mathcal{Z}\left\{\frac{1}{s}\right\} - \mathcal{Z}\left\{\frac{1}{s+a}\right\}, \tag{4.32}$$

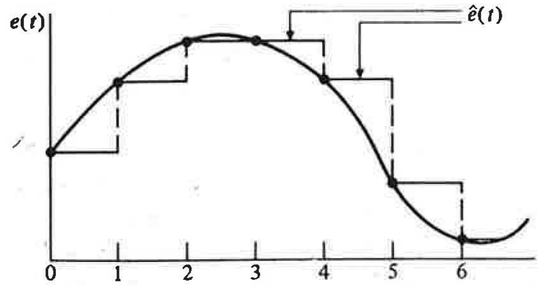

Figure 4.5 A signal, its samples, and its approximation by a zero-order hold.

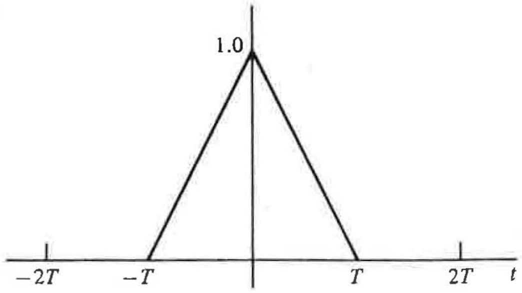

Figure 4.6 Impulse response of the extrapolation filter for the triangle hold.

and, by definition of the operation given in (4.32),

$$\begin{aligned}\mathcal{Z}\left\{\frac{H(s)}{s}\right\} &= \sum_0^\infty z^{-k} - \sum_0^\infty z^{-k}e^{-akT} \\ &= \frac{1}{1-z^{-1}} - \frac{1}{1-e^{-aT}z^{-1}} \\ &= \frac{(1-e^{-aT}z^{-1})-(1-z^{-1})}{(1-z^{-1})(1-e^{-aT}z^{-1})}. \qquad (4.33)\end{aligned}$$

Finally, substituting (4.33) in (4.31), we get the zero-order-hold equivalent of $H(s)$, namely,

$$H_{h0}(z) = \frac{(1-e^{-aT})}{z-e^{-aT}}. \qquad (4.34)$$

We note that for the trivial example given, the zero-order-hold equivalent of (4.34) is identical to the matched zero-pole equivalent given by (4.30). However, this is not generally true as is evident in the comparison with other equivalents for the third-order example (4.1) in Fig. 4.8.

4.4.2 Triangle Hold Equivalent

An interesting hold equivalent can be constructed by imagining that we have a noncausal hold impulse response, as sketched in Fig. 4.6. The result is called the triangle-hold equivalent. The effect of this hold filter is to extrapolate the samples so as to connect sample to sample in a straight line. Although the continuous system is noncausal, the discrete equivalent is not.

The Laplace transform of the extrapolation filter that follows the impulse sampling is

$$\frac{e^{Ts} - 2 + e^{-Ts}}{Ts^2}.$$

Therefore the discrete equivalent that corresponds to (4.31) is

$$H_{tri}(z) = \frac{(z-1)^2}{Tz} \mathcal{Z}\left\{\frac{H(s)}{s^2}\right\}. \tag{4.35}$$

Example 4.4: As a first example, consider $H(s) = 1/s^2$. In this case, from the tables, we have

$$\begin{aligned}\mathcal{Z}\left\{\frac{H(s)}{s^2}\right\} &= \mathcal{Z}\left\{\frac{1}{s^4}\right\} \\ &= \frac{T^2}{6}\frac{(z^2+4z+1)z}{(z-1)^4}\end{aligned} \tag{4.36}$$

and

$$\begin{aligned}H_{\text{tri}}(z) &= \frac{(z-1)^2}{Tz}\frac{T^2}{6}\frac{(z^2+4z+1)z}{(z-1)^4} \\ &= \frac{T^2}{6}\frac{z^2+4z+1}{(z-1)^2}.\end{aligned} \tag{4.37}$$

An alternative, convenient way to compute the triangle-hold equivalent is again to consider the state-space formulation. The block diagram is shown in Fig. 4.7.

The continuous equations are

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{F}\mathbf{x} + \mathbf{G}v, \\ \dot{v} &= w/T, \\ \dot{w} &= u(t+T)\delta(t+T) - 2u(t)\delta(t) + u(t-T)\delta(t-T),\end{aligned} \tag{4.38}$$

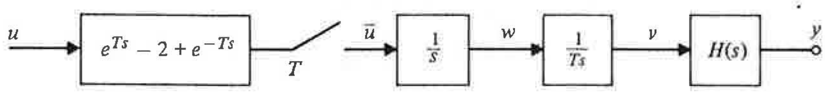

Figure 4.7 Block diagram of the triangle-hold equivalent.

and, in matrix form,

$$\begin{bmatrix} \dot{\mathbf{x}} \\ \dot{v} \\ \dot{w} \end{bmatrix} = \begin{bmatrix} \mathbf{F} & \mathbf{G} & 0 \\ 0 & 0 & 1/T \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ v \\ w \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \bar{u}, \tag{4.39}$$

where $\bar{u}$ represents the input impulse functions. We define the large matrix in (4.39) as $\mathbf{F_T}$, and the one-step solution to this equation is

$$\zeta(kT + T) = e^{\mathbf{F_T}T}\zeta(kT),$$

because $\bar{u}$ consists only of impulses at the sampling instants. If we define

$$\exp(\mathbf{F_T}T) = \begin{bmatrix} \mathbf{\Phi} & \mathbf{\Gamma}_1 & \mathbf{\Gamma}_2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \tag{4.40}$$

then the equation in $\mathbf{x}$ becomes

$$\mathbf{x}(k+1) = \mathbf{\Phi}\mathbf{x}(k) - \mathbf{\Gamma}_1 v(k) + \mathbf{\Gamma}_2 w(k).$$

With care, the last two equations of (4.38) can be integrated to show that $v(k) = u(k)$ and that $w(k) = u(k+1) - u(k)$. Now if we define a new state $\boldsymbol{\xi}(k) = \mathbf{x}(k) - \mathbf{\Gamma}_2 u(k)$, then the state equation for the triangle equivalent is

$$\begin{aligned} \boldsymbol{\xi}(k+1) &= \mathbf{\Phi}(\boldsymbol{\xi}(k) + \mathbf{\Gamma}_2 u(k)) + (\mathbf{\Gamma}_1 - \mathbf{\Gamma}_2)u(k) \\ &= \mathbf{\Phi}\boldsymbol{\xi}(k) + (\mathbf{\Gamma}_1 + \mathbf{\Phi}\mathbf{\Gamma}_2 - \mathbf{\Gamma}_2)u(k). \end{aligned} \tag{4.41}$$

The output equation is

$$\begin{aligned} y(k) &= \mathbf{H}\mathbf{x}(k) + \mathbf{J}u(k) \\ &= H(\boldsymbol{\xi}(k) + \mathbf{\Gamma}_2 u(k)) + \mathbf{J}u(k) \\ &= \mathbf{H}\boldsymbol{\xi}(k) + (\mathbf{J} + \mathbf{H}\mathbf{\Gamma}_2)u(k). \end{aligned} \tag{4.42}$$

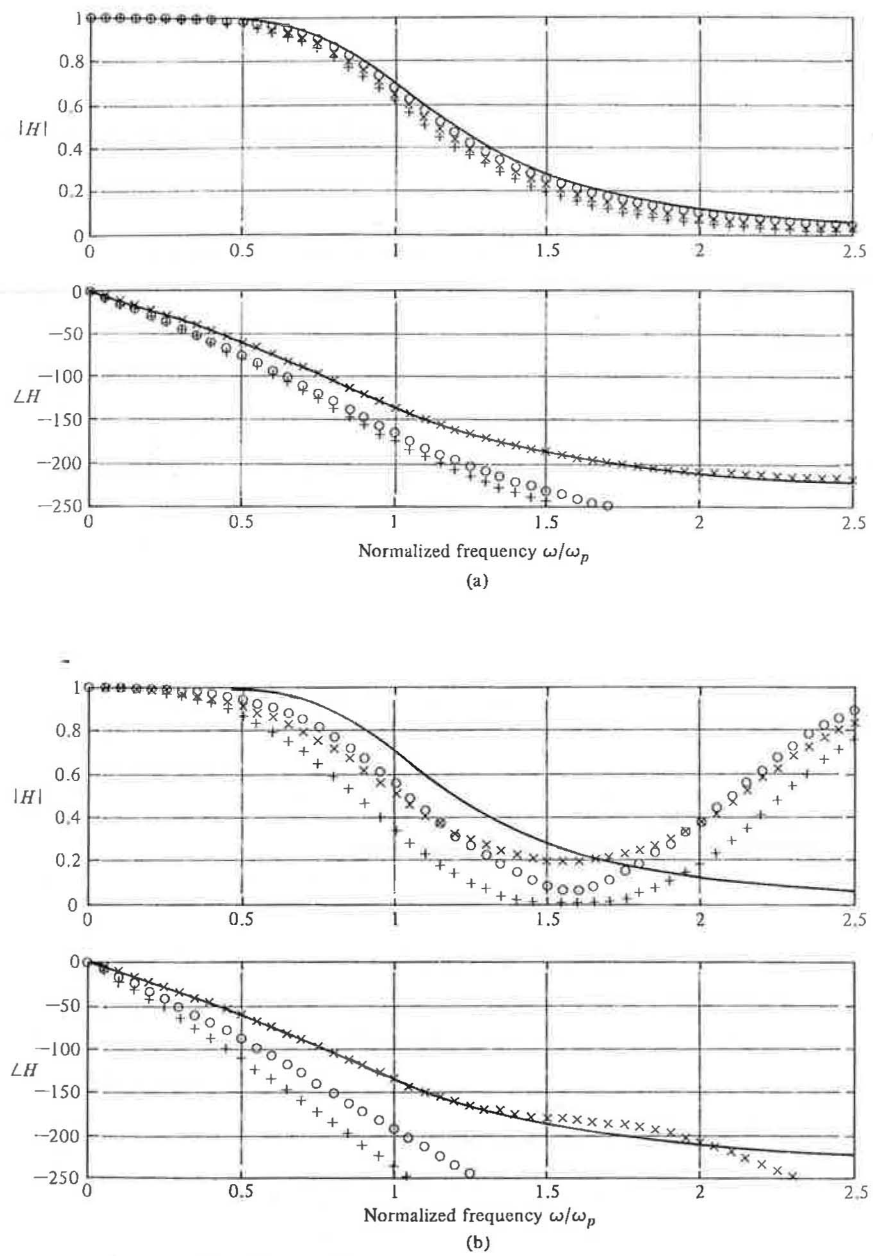

Figure 4.8 Comparison of digital equivalents for sampling period (a) $T = 1$ and $\omega_s/\omega_p = 2\pi$ and (b) $T = 2$ and $\omega_s/\omega_n = \pi$; where ZOH = o, zero-pole = +, and triangle = ×.

Thus the triangle equivalent of $[\mathbf{F}, \mathbf{G}, \mathbf{H}, \mathbf{J}]$ with sample period T is given by

$$\begin{aligned}
\mathbf{A} &= \mathbf{\Phi}, \\
\mathbf{B} &= \mathbf{\Gamma}_1 + \mathbf{\Phi}\mathbf{\Gamma}_2 - \mathbf{\Gamma}_2, \\
\mathbf{C} &= \mathbf{H}, \\
\mathbf{D} &= \mathbf{J} + \mathbf{H}\mathbf{\Gamma}_2,
\end{aligned} \tag{4.43}$$

where $\mathbf{\Phi}$, $\mathbf{\Gamma}_1$, and $\mathbf{\Gamma}_2$ are defined by (4.40).[11]

In Fig. 4.8 the responses of the zero-pole, the zero-order hold, and the triangle-hold equivalents are compared, again for the third-order lowpass filter. We notice in particular that the triangle hold has excellent phase responses, even with sampling period of $T = 2$, which corresponds to a sampling frequency to passband frequency ratio of only $\omega_s/\omega_p = \pi$.

4.5 SUMMARY

In this chapter we have presented several techniques for the construction of discrete equivalents to continuous transfer functions so that known design methods for continuous systems—controls and filters—can be used as a basis for the design of discrete systems. The methods presented were:

1. *Numerical integration*
   a) Forward rectangular rule
   b) Backward rectangular rule
   c) Trapezoid or Tustin's rule
   d) Bilinear transformation with prewarping
2. *Zero-pole mapping*
3. *Hold equivalence*

All methods, except the forward rectangular rule, guarantee a stable, discrete system from a stable, continuous prototype. The bilinear transformation with prewarping affords exact control over the transmission at a selected critical frequency, which must be less than $1/2T$. Zero-pole mapping is the simplest method to apply computationally if the desired filter is given

[11] See TRIANG in Table E.1.

in terms of its zeros and poles; but with a reasonable computer-aided-design tool, the designer can select that method that best meets the requirements of the design.

PROBLEMS AND EXERCISES

4.1 Sketch the zone in the z-plane where poles corresponding to the left half of the s-plane will be mapped by the zero-pole mapping technique and the zero-order-hold technique.

4.2 Show that (4.15) is true.

4.3 **a)** The following transfer function is a lead network designed to add about 60° phase lead at $\omega_1 = 3$ rad:

$$H(s) = \frac{s+1}{0.1s+1}.$$

For each of the following design methods compute and plot in the z-plane the pole and zero locations and compute the amount of phase lead given by the equivalent network at $z_1 = e^{j\omega_1 T}$ if $T = 0.25$ sec and the design is via:

i) Forward rectangular rule
ii) Backward rectangular rule
iii) Bilinear rule
iv) Bilinear with prewarping (Use ω_1 as the warping frequency.)
v) Zero-pole mapping
vi) Zero-order-hold equivalent
vii) Triangular-hold equivalent

b) Plot over the frequency range $\omega_l = 0.1 \rightarrow \omega_h = 100$ the amplitude and phase Bode plots for each of the above equivalents.

4.4 **a)** The following transfer function is a lag network designed to introduce gain attenuation of a factor of 10 (20 dB) at $\omega_1 = 3$:

$$H(s) = \frac{10s+1}{100s+1}.$$

For each of the following design methods, compute and plot on the z-plane the zero-pole patterns of the resulting discrete equivalents and give the gain attenuation at $z_1 = e^{j\omega_1 T}$. Let $T = 0.25$ sec.

i) Forward rectangular rule
ii) Backward rectangular rule
iii) Bilinear rule

- **iv)** Bilinear with prewarping (Use $\omega_1 = 3$ radians as the warping frequency.)
- **v)** Zero-pole mapping
- **vi)** Zero-order-hold equivalent
- **vii)** Triangle-hold equivalent

b) For each case computed, plot the Bode amplitude and phase curves over the range $\omega_l = 0.01 \rightarrow \omega_h = 10$ rad.