7
FILTER DESIGN TECHNIQUES

7.0 INTRODUCTION

Filters are a particularly important class of linear time-invariant systems. Strictly speaking, the term *frequency-selective filter* suggests a system that passes certain frequency components and totally rejects all others, but in a broader context any system that modifies certain frequencies relative to others is also called a filter. While the primary emphasis in this chapter is on the design of frequency-selective filters, some of the techniques are more broadly applicable. Also, we concentrate on the design of causal filters, although in many contexts filters need not be restricted to causal designs. Very often, noncausal filters are designed and implemented by modifying causal designs.

The design of filters involves the following stages: (1) the specification of the desired properties of the system, (2) the approximation of the specifications using a causal discrete-time system, and (3) the realization of the system. Although these three steps are certainly not independent, we focus our attention primarily on the second step, the first being highly dependent on the application and the third dependent on the technology to be used for the implementation. In a practical setting, the desired filter is generally implemented with digital computation and used to filter a signal that is derived from a continuous-time signal by means of periodic sampling followed by analog-to-digital conversion. For this reason, it has become common to refer to discrete-time filters as *digital filters,* even though the underlying design techniques most often relate only to the discrete-time nature of the signals and systems.

When a discrete-time filter is to be used for discrete-time processing of continuous-time signals in the configuration of Figure 7.1, the specifications for both the discrete-time filter and the effective continuous-time filter are typically (but not always) given in the frequency domain. This is especially common for frequency-selective filters such

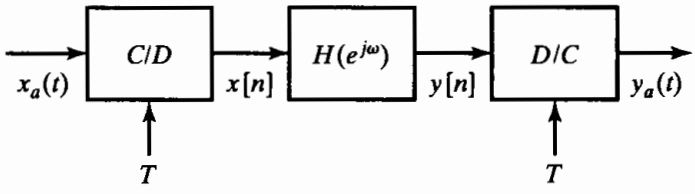


Figure 7.1 Basic system for discrete-time filtering of continuous-time signals.

as lowpass, bandpass, and highpass filters. As shown in Section 4.4, if a linear time-invariant discrete-time system is used as in Figure 7.1, and if the input is bandlimited and the sampling frequency is high enough to avoid aliasing, then the overall system behaves as a linear time-invariant continuous-time system with frequency response

$$H_{\text{eff}}(j\Omega) = \begin{cases} H(e^{j\Omega T}), & |\Omega| < \pi/T, \\ 0, & |\Omega| > \pi/T. \end{cases} \tag{7.1a}$$

In such cases, it is straightforward to convert from specifications on the effective continuous-time filter to specifications on the discrete-time filter through the relation $\omega = \Omega T$. That is, $H(e^{j\omega})$ is specified over one period by the equation

$$H(e^{j\omega}) = H_{\text{eff}}\left(j\frac{\omega}{T}\right), \qquad |\omega| < \pi. \tag{7.1b}$$

This type of conversion is illustrated in Example 7.1.

Example 7.1 Determining Specifications for a Discrete-Time Filter

Consider a discrete-time filter that is to be used to lowpass filter a continuous-time signal using the basic configuration of Figure 7.1. Specifically, we want the overall system of that figure to have the following properties when the sampling rate is 10^4 samples/s ($T = 10^{-4}$ s):

1. The gain $|H_{\text{eff}}(j\Omega)|$ should be within ± 0.01 of unity in the frequency band $0 \le \Omega \le 2\pi(2000)$.
2. The gain should be no greater than 0.001 in the frequency band $2\pi(3000) \le \Omega$.

Such a set of lowpass specifications on $|H_{\text{eff}}(j\Omega)|$ can be depicted as in Figure 7.2(a), where the limits of tolerable approximation error are indicated by the shaded horizontal lines. For this specific example, the parameters would be

$$\delta_1 = 0.01,$$

$$\delta_2 = 0.001,$$

$$\Omega_p = 2\pi(2000),$$

$$\Omega_s = 2\pi(3000).$$

Therefore, in this case, the ideal passband gain is unity. The passband gain varies between $(1 + \delta_1)$, and $(1 - \delta_1)$, and the stopband gain varies betwen 0 and δ_2. It is common to express the maximum passband and stopband gains in units of decibels. For this example:

$$\begin{aligned} &\text{ideal passband gain in decibels} &&= 20\log_{10}(1) &&= 0 \text{ dB} \\ &\text{maximum passband gain in decibels} &&= 20\log_{10}(1.01) &&= 0.086 \text{ dB} \\ &\text{maximum stopband gain in decibels} &&= 20\log_{10}(0.001) &&= -60 \text{ dB} \end{aligned}$$

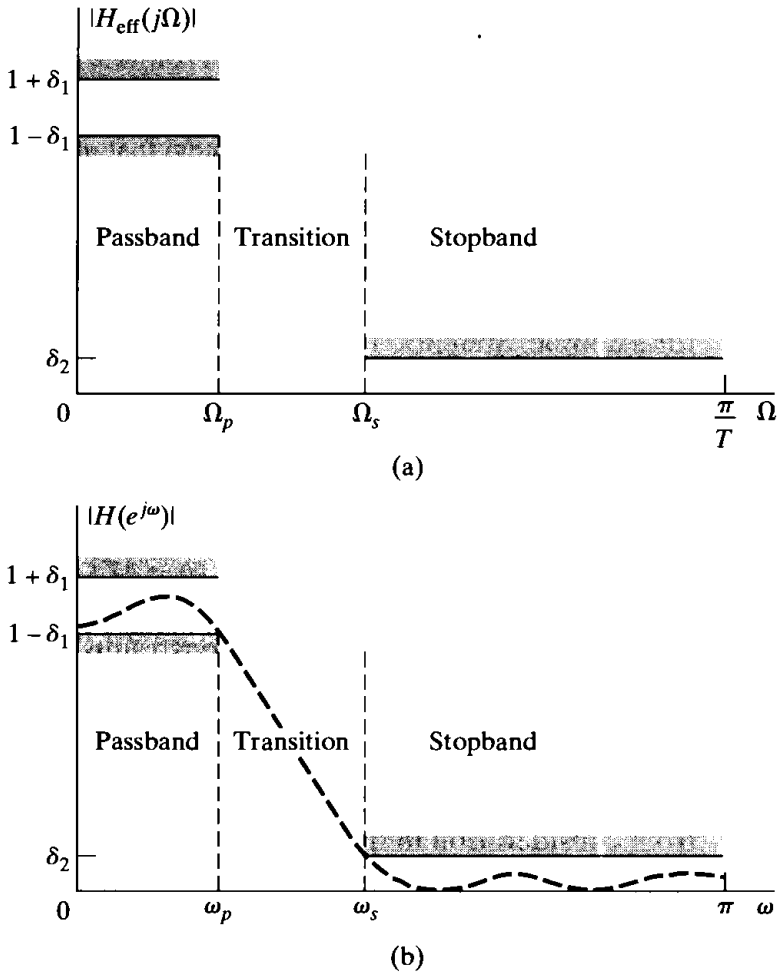


Figure 7.2 (a) Specifications for effective frequency response of overall system in Figure 7.1 for the case of a lowpass filter. (b) Corresponding specifications for the discrete-time system in Figure 7.1.

Since the sampling rate is 10^4 samples/s, the gain of the overall system is identically zero above $\Omega = 2\pi(5000)$, due to the ideal discrete-to-continuous (D/C) converter in Figure 7.1.

The tolerance scheme for the discrete-time filter is shown in Figure 7.2(b). It is the same as that in Figure 7.2(a), except that it is plotted as a function of normalized frequency ($\omega = \Omega T$), and it need only be plotted in the range $0 \leq \omega \leq \pi$, since the remainder can be inferred from symmetry properties (assuming that $h[n]$ is real) and the periodicity of $H(e^{j\omega})$. From Eq. (7.1b), it follows that in the *passband* the magnitude of the frequency response must approximate unity within an error of $\pm\delta_1$, i.e.,

$$(1-\delta_1) \leq |H(e^{j\omega})| \leq (1+\delta_1), \qquad |\omega| \leq \omega_p, \tag{7.2}$$

where $\delta_1 = 0.01$ and $\omega_p = 2\pi(2000) \cdot 10^{-4} = 0.4\pi$ radians. The other approximation band is the *stopband,* in which the magnitude response must approximate zero with an error less than δ_2; i.e.,

$$|H(e^{j\omega})| \leq \delta_2, \qquad \omega_s \leq |\omega| \leq \pi. \tag{7.3}$$

In this example, $\delta_2 = 0.001$ and $\omega_s = 2\pi(3000) \cdot 10^{-4} = 0.6\pi$ radians. The passband cutoff frequency ω_p and the stopband cutoff frequency ω_s are given in terms

of normalized radian frequency or equivalently in terms of angle in the z-plane. To approximate the ideal lowpass filter in this way with a realizable system, we must provide a transition band of nonzero width $(\omega_s - \omega_p)$ in which the magnitude response changes smoothly from passband to stopband. The dashed curve in Figure 7.2(b) is the magnitude response of a system that meets the prescribed specification.

There are many applications in which a discrete-time signal to be filtered is not derived from a continuous-time signal, and there are a variety of means besides periodic sampling for representing continuous-time signals in terms of sequences. (See, for example, Steiglitz, 1965, and Oppenheim and Johnson, 1972.) Also, in most of the design techniques that we discuss, the sampling period plays no role whatsoever in the approximation procedure. For these reasons, we take the point of view that the filter design problem begins from a set of desired specifications in terms of the discrete-time frequency variable ω. Depending on the specific application or context, these specifications may or may not have been obtained from a consideration of filtering in the framework of Figure 7.1.

Many of the filters used in practice are specified by a tolerance scheme similar to that in Example 7.1, with no constraints on the phase response other than those imposed implicitly by stability and causality requirements. For example, the poles of the system function for a causal and stable infinite impulse response (IIR) filter must lie inside the unit circle. Similarly, in designing finite impulse response (FIR) filters, we often impose the constraint of a linear phase. This again removes the phase of the signal from consideration in the design process.

Given a set of specifications in the form of Figure 7.2(b), we must determine the system function of a discrete-time linear system whose frequency response falls within the prescribed tolerances. This is a problem in functional approximation. Designing IIR filters implies approximation by a rational function of z, while designing FIR filters implies polynomial approximation. Our discussion distinguishes between design techniques that are appropriate for IIR filters and those that are appropriate for FIR filters. We discuss a variety of design techniques for both types of filter, ranging from closed-form procedures, which involve only substitution of design specifications into design formulas, to algorithmic techniques, in which a solution is obtained by an iterative procedure.

7.1 DESIGN OF DISCRETE-TIME IIR FILTERS FROM CONTINUOUS-TIME FILTERS

The traditional approach to the design of discrete-time IIR filters involves the transformation of a continuous-time filter into a discrete-time filter meeting prescribed specifications. This is a reasonable approach for several reasons:

- The art of continuous-time IIR filter design is highly advanced, and since useful results can be achieved, it is advantageous to use the design procedures already developed for continuous-time filters.
- Many useful continuous-time IIR design methods have relatively simple closed-form design formulas. Therefore, discrete-time IIR filter design methods based on such standard continuous-time design formulas are rather simple to carry out.

- The standard approximation methods that work well for continuous-time IIR filters do not lead to simple closed-form design formulas when these methods are applied directly to the discrete-time IIR case.

The fact that continuous-time filter designs can be mapped to discrete-time filter designs is totally unrelated to, and independent of, whether the discrete-time filter is to be used in the configuration of Figure 7.1 for processing continuous-time signals. We emphasize again that the design procedure for the discrete-time system begins from a set of *discrete-time* specifications. Henceforth, we assume that these specifications have been determined by an analysis like that of Example 7.1 or by some other method. We will use continuous-time filter approximation methods only as a convenience in determining the discrete-time filter that meets the desired specifications. Indeed, the continuous-time filter on which the approximation is based may have a frequency response that is vastly different from the effective frequency response when the discrete-time filter is used in the configuration of Figure 7.1.

In designing a discrete-time filter by transforming a prototype continuous-time filter, the specifications for the continuous-time filter are obtained by a transformation of the specifications for the desired discrete-time filter. The system function $H_c(s)$ or impulse response $h_c(t)$ of the continuous-time filter is then obtained through one of the established approximation methods used for continuous-time filter design, examples of which are discussed in Appendix B. Next, the system function $H(z)$ or impulse response $h[n]$ for the discrete-time filter is obtained by applying to $H_c(s)$ or $h_c(t)$ a transformation of the type discussed in this section.

In such transformations, we generally require that the essential properties of the continuous-time frequency response be preserved in the frequency response of the resulting discrete-time filter. Specifically, this implies that we want the imaginary axis of the s-plane to map onto the unit circle of the z-plane. A second condition is that a stable continuous-time filter should be transformed to a stable discrete-time filter. This means that if the continuous-time system has poles only in the left half of the s-plane, then the discrete-time filter must have poles only inside the unit circle in the z-plane. These constraints are basic to all the techniques discussed in this section.

7.1.1 Filter Design by Impulse Invariance

In Section 4.4.2 we discussed the concept of *impulse invariance,* wherein a discrete-time system is defined by sampling the impulse response of a continuous-time system. We showed that impulse invariance provides a direct means of computing samples of the output of a bandlimited continuous-time system for bandlimited input signals. Alternatively, in the context of filter design, we can think of impulse invariance as a method for obtaining a discrete-time system whose frequency response is determined by the frequency response of a continuous-time system.

In the impulse invariance design procedure for transforming continuous-time filters into discrete-time filters, the impulse response of the discrete-time filter is chosen proportional to equally spaced samples of the impulse response of the continuous-time filter; i.e.,

$$h[n] = T_d h_c(nT_d), \tag{7.4}$$

where T_d represents a sampling interval. As we will see, because we begin the design problem with the discrete-time filter specifications, the parameter T_d in Eq. (7.4) in fact has no role whatsoever in the design process or the resulting discrete-time filter. However, since it is customary to specify this parameter in defining the procedure, we include it in the following discussion. As we will see, even if the filter is used in the basic configuration of Figure 7.1, the design sampling period T_d need not be the same as the sampling period T associated with the C/D and D/C conversion.

When impulse invariance is used as a means for designing a discrete-time filter with a specified frequency response, we are especially interested in the relationship between the frequency responses of the discrete-time and continuous-time filters. From the discussion of sampling in Chapter 4, it follows that the frequency response of the discrete-time filter obtained through Eq. (7.4) is related to the frequency response of the continuous-time filter by

$$H(e^{j\omega}) = \sum_{k=-\infty}^{\infty} H_c\left(j\frac{\omega}{T_d} + j\frac{2\pi}{T_d}k\right). \tag{7.5}$$

If the continuous-time filter is bandlimited, so that

$$H_c(j\Omega) = 0, \qquad |\Omega| \geq \pi/T_d, \tag{7.6}$$

then

$$H(e^{j\omega}) = H_c\left(j\frac{\omega}{T_d}\right), \qquad |\omega| \leq \pi; \tag{7.7}$$

i.e., the discrete-time and continuous-time frequency responses are related by a linear scaling of the frequency axis, namely, $\omega = \Omega T_d$ for $|\omega| < \pi$. Unfortunately, any practical continuous-time filter cannot be exactly bandlimited, and consequently, interference between successive terms in Eq. (7.5) occurs, causing aliasing, as illustrated in Figure 7.3. However, if the continuous-time filter approaches zero at high frequencies, the aliasing may be negligibly small, and a useful discrete-time filter can result from the sampling of the impulse response of a continuous-time filter.

In the impulse invariance design procedure, the discrete-time filter specifications are first transformed to continuous-time filter specifications through the use of Eq. (7.7). Assuming that the aliasing involved in the transformation from $H_c(j\Omega)$ to $H(e^{j\omega})$ will be negligible, we obtain the specifications on $H_c(j\Omega)$ by applying the relation

$$\Omega = \omega/T_d \tag{7.8}$$

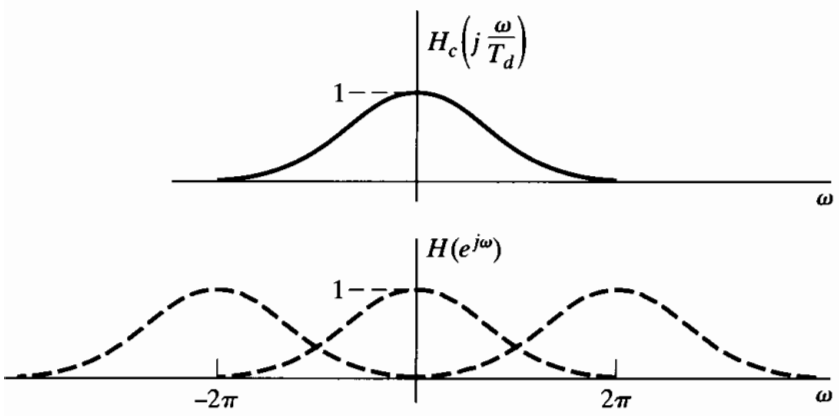


Figure 7.3 Illustration of aliasing in the impulse invariance design technique.

to obtain the continuous-time filter specifications from the specifications on $H(e^{j\omega})$. After obtaining a suitable continuous-time filter based on these specifications, the continuous-time filter with system function $H_c(s)$ is transformed to the desired discrete-time filter with system function $H(z)$. We develop the algebraic details of the transformation from $H_c(s)$ to $H(z)$ shortly. Note, however, that in the transformation back to discrete-time frequency, $H(e^{j\omega})$ will be related to $H_c(j\Omega)$ through Eq. (7.5), which again applies the transformation of Eq. (7.8) to the frequency axis. As a consequence, the "sampling" parameter T_d cannot be used to control aliasing. Since the basic specifications are in terms of discrete-time frequency, if the sampling rate is increased (i.e., if T_d is made smaller), then the cutoff frequency of the continuous-time filter must increase in proportion. In practice, to compensate for aliasing that might occur in the transformation from $H_c(s)$ to $H(z)$, the continuous-time filter may be somewhat overdesigned, i.e., designed to exceed the specifications, particularly in the stopband.

While the impulse invariance transformation from continuous time to discrete time is defined in terms of time-domain sampling, it is easy to carry out as a transformation on the system function. To develop this transformation, let us consider the system function of the continuous-time filter expressed in terms of a partial fraction expansion, so that[1]

$$H_c(s) = \sum_{k=1}^{N} \frac{A_k}{s - s_k}. \tag{7.9}$$

The corresponding impulse response is

$$h_c(t) = \begin{cases} \displaystyle\sum_{k=1}^{N} A_k e^{s_k t}, & t \geq 0, \\ 0, & t < 0. \end{cases} \tag{7.10}$$

The impulse response of the discrete-time filter obtained by sampling $T_d h_c(t)$ is

$$\begin{aligned} h[n] = T_d h_c(nT_d) &= \sum_{k=1}^{N} T_d A_k e^{s_k n T_d} u[n] \\ &= \sum_{k=1}^{N} T_d A_k (e^{s_k T_d})^n u[n]. \end{aligned} \tag{7.11}$$

The system function of the discrete-time filter is therefore given by

$$H(z) = \sum_{k=1}^{N} \frac{T_d A_k}{1 - e^{s_k T_d} z^{-1}}. \tag{7.12}$$

In comparing Eqs. (7.9) and (7.12), we observe that a pole at $s = s_k$ in the s-plane transforms to a pole at $z = e^{s_k T_d}$ in the z-plane and the coefficients in the partial fraction expansions of $H_c(s)$ and $H(z)$ are equal, except for the scaling multiplier T_d. If the continuous-time filter is stable, corresponding to the real part of s_k being less than zero, then the magnitude of $e^{s_k T_d}$ will be less than unity, so that the corresponding pole in the discrete-time filter is inside the unit circle. Therefore, the causal discrete-time filter is

[1]For simplicity, we assume in the discussion that all poles of $H(s)$ are single order. In Problem 7.24, we consider the modifications required for multiple-order poles.

also stable. While the poles in the s-plane map to poles in the z-plane according to the relationship $z_k = e^{s_k T_d}$, it is important to recognize that the impulse invariance design procedure does not correspond to a simple mapping of the s-plane to the z-plane by that relationship. In particular, the zeros in the discrete-time system function are a function of the poles and the coefficients $T_d A_k$ in the partial fraction expansion, and they will not in general be mapped in the same way the poles are mapped. We illustrate the impulse invariance design procedure with the following example.

Example 7.2 Impulse Invariance with a Butterworth Filter

Let us consider the design of a lowpass discrete-time filter by applying impulse invariance to an appropriate Butterworth continuous-time filter.[2] The specifications for the discrete-time filter are

$$0.89125 \leq |H(e^{j\omega})| \leq 1, \qquad 0 \leq |\omega| \leq 0.2\pi, \tag{7.13a}$$

$$|H(e^{j\omega})| \leq 0.17783, \qquad 0.3\pi \leq |\omega| \leq \pi. \tag{7.13b}$$

Since the parameter T_d cancels in the impulse invariance procedure, we can choose $T_d = 1$, so that $\omega = \Omega$. In Problem 7.2, this same example is considered, but with the parameter T_d explicitly included to illustrate how and where it cancels.

In designing the filter using impulse invariance on a continuous-time Butterworth filter, we must first transform the discrete-time specifications to specifications on the continuous-time filter. Recall that impulse invariance corresponds to a linear mapping between Ω and ω in the absence of aliasing. For this example, we will assume that the effect of aliasing is negligible. After the design is complete, we can evaluate the resulting frequency response against the specifications in Eqs. (7.13a) and (7.13b).

Because of the preceding considerations, we want to design a continuous-time Butterworth filter with magnitude function $|H_c(j\Omega)|$ for which

$$0.89125 \leq |H_c(j\Omega)| \leq 1, \qquad 0 \leq |\Omega| \leq 0.2\pi, \tag{7.14a}$$

$$|H_c(j\Omega)| \leq 0.17783, \qquad 0.3\pi \leq |\Omega| \leq \pi. \tag{7.14b}$$

Since the magnitude response of an analog Butterworth filter is a monotonic function of frequency, Eqs. (7.14a) and (7.14b) will be satisfied if

$$|H_c(j0.2\pi)| \geq 0.89125 \tag{7.15a}$$

and

$$|H_c(j0.3\pi)| \leq 0.17783. \tag{7.15b}$$

Specifically, the magnitude-squared function of a Butterworth filter is of the form

$$|H_c(j\Omega)|^2 = \frac{1}{1 + (\Omega/\Omega_c)^{2N}}, \tag{7.16}$$

so that the filter design process consists of determining the parameters N and Ω_c to meet the desired specifications. Using Eq. (7.16) in Eqs. (7.15) with equality leads

[2]Continuous-time Butterworth and Chebyshev filters are discussed in Appendix B.

to the equations

$$1+\left(\frac{0.2\pi}{\Omega_c}\right)^{2N}=\left(\frac{1}{0.89125}\right)^2 \tag{7.17a}$$

and

$$1+\left(\frac{0.3\pi}{\Omega_c}\right)^{2N}=\left(\frac{1}{0.17783}\right)^2. \tag{7.17b}$$

The solution of these two equations is $N = 5.8858$ and $\Omega_c = 0.70474$. The parameter N, however, must be an integer. Therefore, so that the specifications are met or exceeded, we must round N up to the nearest integer, $N = 6$. Because we have rounded N up to the next highest integer, the filter will not exactly satisfy both Eqs. (7.17a) and (7.17b) simultaneously. With $N = 6$, the filter parameter Ω_c can be chosen to exceed the specified requirements in either the passband, the stopband, or both. Specifically, as the value of Ω_c varies, there is a trade-off in the amount by which the stopband and passband specifications are exceeded. If we substitute $N = 6$ into Eq. (7.17a), we obtain $\Omega_c = 0.7032$. With this value, the passband specifications (of the continuous-time filter) will be met exactly, and the stopband specifications (of the continuous-time filter) will be exceeded. This allows some margin for aliasing in the discrete-time filter. With $\Omega_c = 0.7032$ and with $N = 6$, the 12 poles of the magnitude-squared function $H_c(s)H_c(-s) = 1/[1 + (s/j\Omega_c)^{2N}]$ are uniformly distributed in angle on a circle of radius $\Omega_c = 0.7032$, as indicated in Figure 7.4. Consequently, the poles of $H_c(s)$ are the three pole pairs in the left half of the s-plane with the following coordinates:

Pole pair 1: $-0.182 \pm j(0.679)$,

Pole pair 2: $-0.497 \pm j(0.497)$,

Pole pair 3: $-0.679 \pm j(0.182)$.

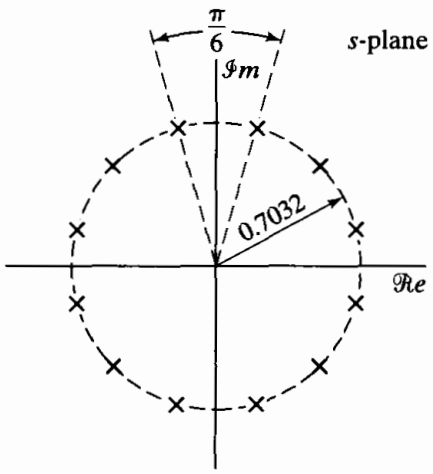


Figure 7.4 s-plane locations for poles of $H_c(s)H_c(-s)$ for sixth-order Butterworth filter in Example 7.2.

Therefore,

$$H_c(s)=\frac{0.12093}{(s^2+0.3640s+0.4945)(s^2+0.9945s+0.4945)(s^2+1.3585s+0.4945)} \tag{7.18}$$

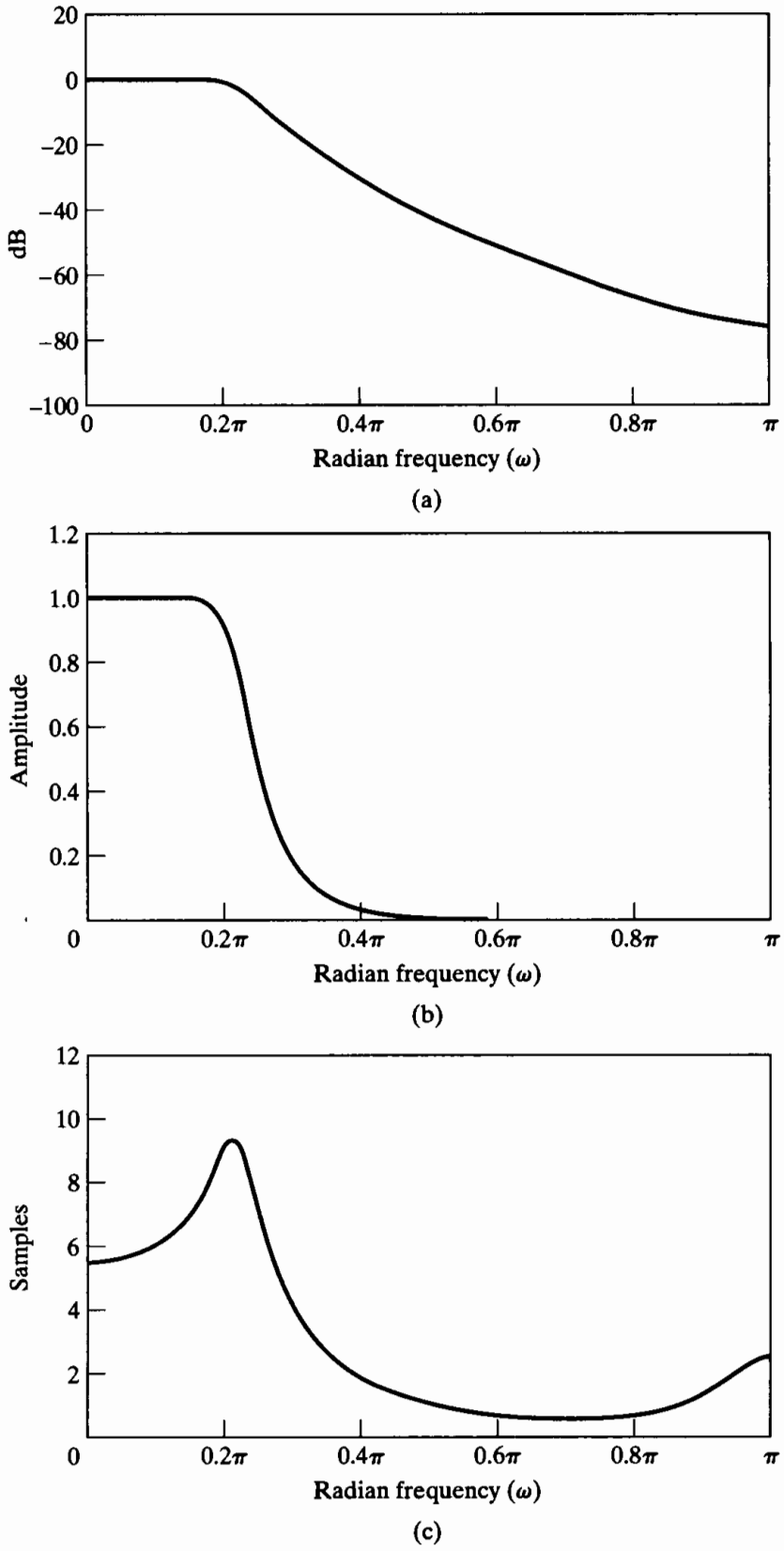


Figure 7.5 Frequency response of sixth-order Butterworth filter transformed by impulse invariance. (a) Log magnitude in dB. (b) Magnitude. (c) Group delay.

If we express $H_c(s)$ as a partial fraction expansion, perform the transformation of Eq. (7.12), and then combine complex-conjugate terms, the resulting system function of the discrete-time filter is

$$H(z) = \frac{0.2871 - 0.4466z^{-1}}{1 - 1.2971z^{-1} + 0.6949z^{-2}} + \frac{-2.1428 + 1.1455z^{-1}}{1 - 1.0691z^{-1} + 0.3699z^{-2}} + \frac{1.8557 - 0.6303z^{-1}}{1 - 0.9972z^{-1} + 0.2570z^{-2}}. \tag{7.19}$$

As is evident from Eq. (7.19), the system function resulting from the impulse invariance design procedure may be realized directly in parallel form. If cascade or direct form is desired, the separate second-order terms must be combined in an appropriate way.

The frequency-response functions of the discrete-time system are shown in Figure 7.5. Recall that the prototype continuous-time filter was designed to meet the specifications exactly at the passband edge and to exceed the specifications at the stopband edge, and this turns out to be true for the resulting discrete-time filter. This is an indication that the continuous-time filter was sufficiently bandlimited so that aliasing presented no problem. Indeed, the difference between $20\log_{10}|H(e^{j\omega})|$ and $20\log_{10}|H_c(j\Omega)|$ would not be visible on this plotting scale, except for a slight deviation around $\omega = \pi$. (Recall that $T_d = 1$, so $\Omega = \omega$.) Sometimes, aliasing is much more of a problem. If the resulting discrete-time filter fails to meet the specifications because of aliasing, there is no alternative with impulse invariance but to try again with a higher order filter or with different filter parameters, holding the order fixed.

The basis for impulse invariance is to choose an impulse response for the discrete-time filter that is similar in some sense to the impulse response of the continuous-time filter. The use of this procedure is often motivated not so much by a desire to maintain the shape of the impulse response as by the knowledge that if the continuous-time filter is bandlimited, then the discrete-time filter frequency response will closely approximate the continuous-time frequency response. However, in some filter design problems, a primary objective may be to control some aspect of the time response, such as the impulse response or the step response. In these cases, a natural approach might be to design the discrete-time filter by impulse invariance or by step invariance. In the latter case, the response of the filter to a sampled unit step function is defined to be the sequence obtained by sampling the continuous-time step response. If the continuous-time filter has good step response characteristics, such as a small rise time and low peak overshoot, these characteristics will be preserved in the discrete-time filter. Clearly, this concept of waveform invariance can be extended to the preservation of the output waveshape for a variety of inputs, as illustrated in Problem 7.1. The problem points up the fact that transforming the same continuous-time filter by impulse invariance and also by step invariance (or some other waveform invariance criterion) does not lead to the same discrete-time filter in the two cases.

In the impulse invariance design procedure, the relationship between continuous-time and discrete-time frequency is linear; consequently, except for aliasing, the shape of the frequency response is preserved. This is in contrast to the procedure discussed next, which is based on an algebraic transformation. We note, in concluding this subsection, that the impulse invariance technique is appropriate only for bandlimited filters; highpass or bandstop continuous-time filters, for example, would require additional bandlimiting to avoid severe aliasing distortion if impulse invariance design is used.

7.1.2 Bilinear Transformation

The technique discussed in this subsection avoids the problem of aliasing by using the bilinear transformation, an algebraic transformation between the variables s and z that maps the entire $j\Omega$-axis in the s-plane to one revolution of the unit circle in the z-plane. Since $-\infty \leq \Omega \leq \infty$ maps onto $-\pi \leq \omega \leq \pi$, the transformation between the continuous-time and discrete-time frequency variables must be nonlinear. Therefore, the use of this technique is restricted to situations in which the corresponding warping of the frequency axis is acceptable.

With $H_c(s)$ denoting the continuous-time system function and $H(z)$ the discrete-time system function, the bilinear transformation corresponds to replacing s by

$$s = \frac{2}{T_d}\left(\frac{1 - z^{-1}}{1 + z^{-1}}\right); \tag{7.20}$$

that is,

$$H(z) = H_c\left[\frac{2}{T_d}\left(\frac{1 - z^{-1}}{1 + z^{-1}}\right)\right]. \tag{7.21}$$

As in impulse invariance, a "sampling" parameter T_d is included in the definition of the bilinear transformation. Historically, this parameter has been included because the difference equation corresponding to $H(z)$ can be obtained by applying the trapezoidal integration rule to the differential equation corresponding to $H_c(s)$, with T_d representing the step size of the numerical integration. (See Kaiser, 1966, and Problem 7.43.) However, in filter design, our use of the bilinear transformation is based on the properties of the algebraic transformation given in Eq. (7.20). As with impulse invariance, the parameter T_d is of no consequence in the design procedure, since we assume that the design problem always begins with specifications on the discrete-time filter $H(e^{j\omega})$. When these specifications are mapped to continuous-time specifications and the continuous-time filter is then mapped back to a discrete-time filter, the effect of T_d will cancel. Although we will retain the parameter T_d in our discussion, in specific problems and examples any convenient value of T_d can be chosen.

To develop the properties of the algebraic transformation specified in Eq. (7.20), we solve for z to obtain

$$z = \frac{1 + (T_d/2)s}{1 - (T_d/2)s}, \tag{7.22}$$

and, substituting $s = \sigma + j\Omega$ into Eq. (7.22), we obtain

$$z = \frac{1 + \sigma T_d/2 + j\Omega T_d/2}{1 - \sigma T_d/2 - j\Omega T_d/2}. \tag{7.23}$$

If $\sigma < 0$, then, from Eq. (7.23), it follows that $|z| < 1$ for any value of Ω. Similarly, if $\sigma > 0$, then $|z| > 1$ for all Ω. That is, if a pole of $H_c(s)$ is in the left-half s-plane, its image in the z-plane will be inside the unit circle. Therefore, causal stable continuous-time filters map into causal stable discrete-time filters.

Next, to show that the $j\Omega$-axis of the s-plane maps onto the unit circle, we substitute $s = j\Omega$ into Eq. (7.22), obtaining

$$z = \frac{1 + j\Omega T_d/2}{1 - j\Omega T_d/2}. \tag{7.24}$$

From Eq. (7.24), it is clear that $|z| = 1$ for all values of s on the $j\Omega$-axis. That is, the $j\Omega$-axis maps onto the unit circle, so Eq. (7.24) takes the form

$$e^{j\omega} = \frac{1 + j\Omega T_d/2}{1 - j\Omega T_d/2}. \tag{7.25}$$

To derive a relationship between the variables ω and Ω, it is useful to return to Eq. (7.20) and substitute $z = e^{j\omega}$. We obtain

$$s = \frac{2}{T_d}\left(\frac{1 - e^{-j\omega}}{1 + e^{-j\omega}}\right), \tag{7.26}$$

or, equivalently,

$$s = \sigma + j\Omega = \frac{2}{T_d}\left[\frac{2e^{-j\omega/2}(j \sin \omega/2)}{2e^{-j\omega/2}(\cos \omega/2)}\right] = \frac{2j}{T_d}\tan(\omega/2). \tag{7.27}$$

Equating real and imaginary parts on both sides of Eq. (7.27) leads to the relations $\sigma = 0$ and

$$\Omega = \frac{2}{T_d}\tan(\omega/2), \tag{7.28}$$

or

$$\omega = 2\arctan(\Omega T_d/2). \tag{7.29}$$

These properties of the bilinear transformation as a mapping from the s-plane to the z-plane are summarized in Figures 7.6 and 7.7. From Eq. (7.29) and Figure 7.7, we see that the range of frequencies $0 \leq \Omega \leq \infty$ maps to $0 \leq \omega \leq \pi$, while the range $-\infty \leq \Omega \leq 0$ maps to $-\pi \leq \omega \leq 0$. The bilinear transformation avoids the problem of aliasing encountered with the use of impulse invariance, because it maps the entire imaginary axis of the s-plane onto the unit circle in the z-plane. The price paid for this, however, is the nonlinear compression of the frequency axis depicted in Figure 7.7. Consequently, the design of discrete-time filters using the bilinear transformation is useful only when this compression can be tolerated or compensated for, as in the case of filters that approximate ideal piecewise-constant magnitude-response characteristics. This is illustrated in Figure 7.8, where we show how a continuous-time frequency response and tolerance scheme maps to a corresponding discrete-time frequency response and tolerance scheme through the frequency warping of Eqs. (7.28) and (7.29). If the critical frequencies (such as the passband and stopband edge frequencies) of the continuous-time filter are prewarped according to Eq. (7.28) then, when the continuous-time filter is transformed to the discrete-time filter using Eq. (7.21), the discrete-time filter will meet the desired specifications.

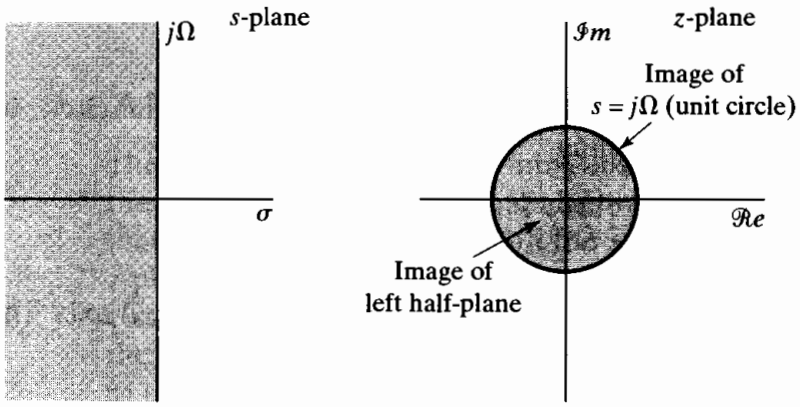


Figure 7.6 Mapping of the *s*-plane onto the *z*-plane using the bilinear transformation.

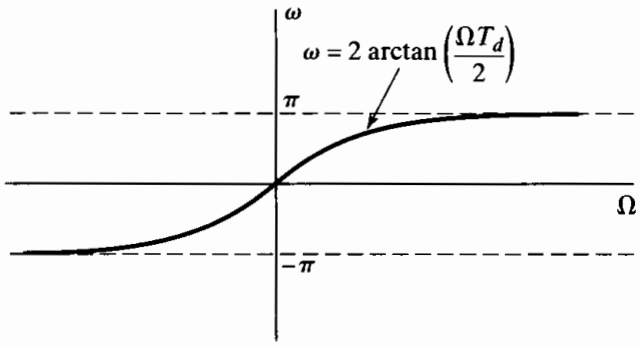


Figure 7.7 Mapping of the continuous-time frequency axis onto the discrete-time frequency axis by bilinear transformation.

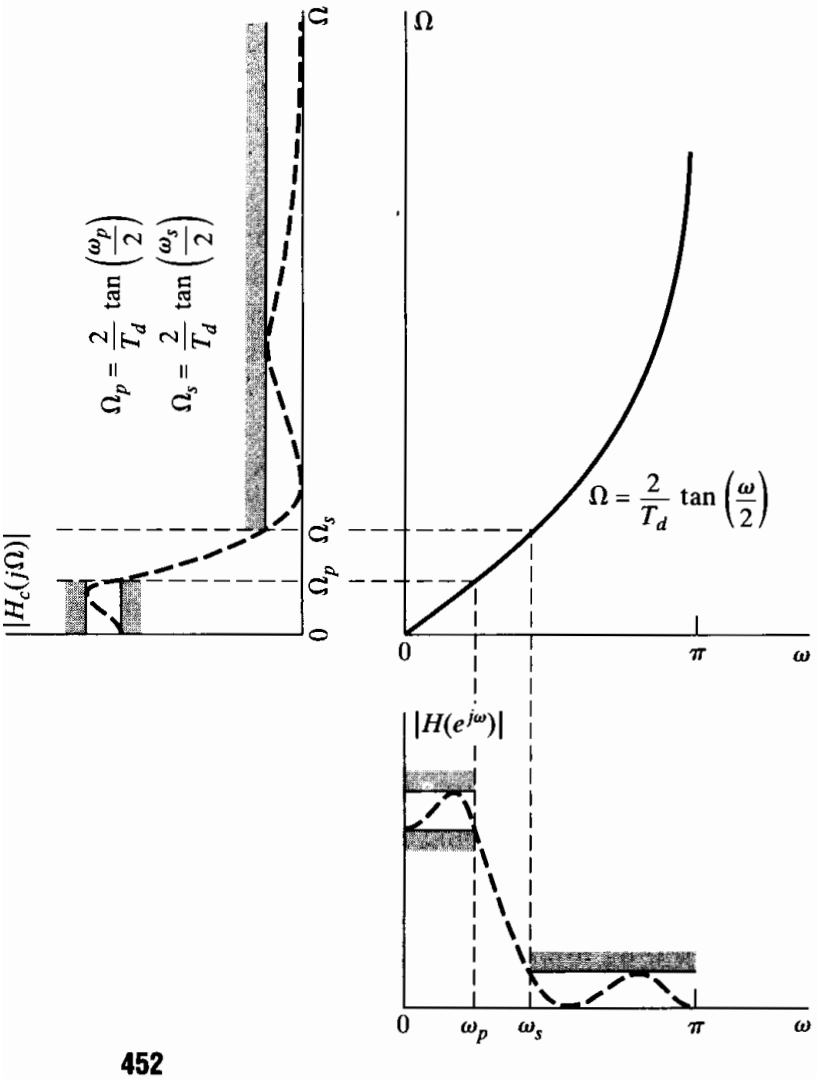


Figure 7.8 Frequency warping inherent in the bilinear transformation of a continuous-time lowpass filter into a discrete-time lowpass filter. To achieve the desired discrete-time cutoff frequencies, the continuous-time cutoff frequencies must be prewarped as indicated.

Typical frequency-selective continuous-time approximations are Butterworth, Chebyshev, and elliptic filters. The closed-form design formulas of these continuous-time approximation methods make the design procedure rather straightforward. As discussed in Appendix B a Butterworth continuous-time filter is monotonic in the passband and in the stopband. A type I Chebyshev filter has an equiripple characteristic in the passband and varies monotonically in the stopband. A type II Chebyshev filter is monotonic in the passband and equiripple in the stopband. An elliptic filter is equiripple in both the passband and the stopband. Clearly, these properties will be preserved when the filter is mapped to a digital filter with the bilinear transformation. This is illustrated by the dashed approximation shown in Figure 7.8.

Although the bilinear transformation can be used effectively in mapping a piecewise-constant magnitude-response characteristic from the s-plane to the z-plane, the distortion in the frequency axis also manifests itself as a warping of the phase response of the filter. For example, Figure 7.9 shows the result of applying the bilinear transformation to an ideal linear phase factor $e^{-s\alpha}$. If we substitute Eq. (7.20) for s and evaluate the result on the unit circle, the phase angle is $-(2\alpha/T_d)\tan(\omega/2)$. In Figure 7.9, the solid curve shows the function $-(2\alpha/T_d)\tan(\omega/2)$, and the dotted curve is the periodic linear phase function $-(\omega\alpha/T_d)$, which is obtained by using the small angle approximation $\omega/2 \approx \tan(\omega/2)$. From this, it should be evident that if we were interested in a discrete-time lowpass filter with a linear phase characteristic, we could not obtain such a filter by applying the bilinear transformation to a continuous-time lowpass filter with a linear phase characteristic.

As mentioned previously, because of the frequency warping, the use of the bilinear transformation is restricted to the design of approximations to filters with

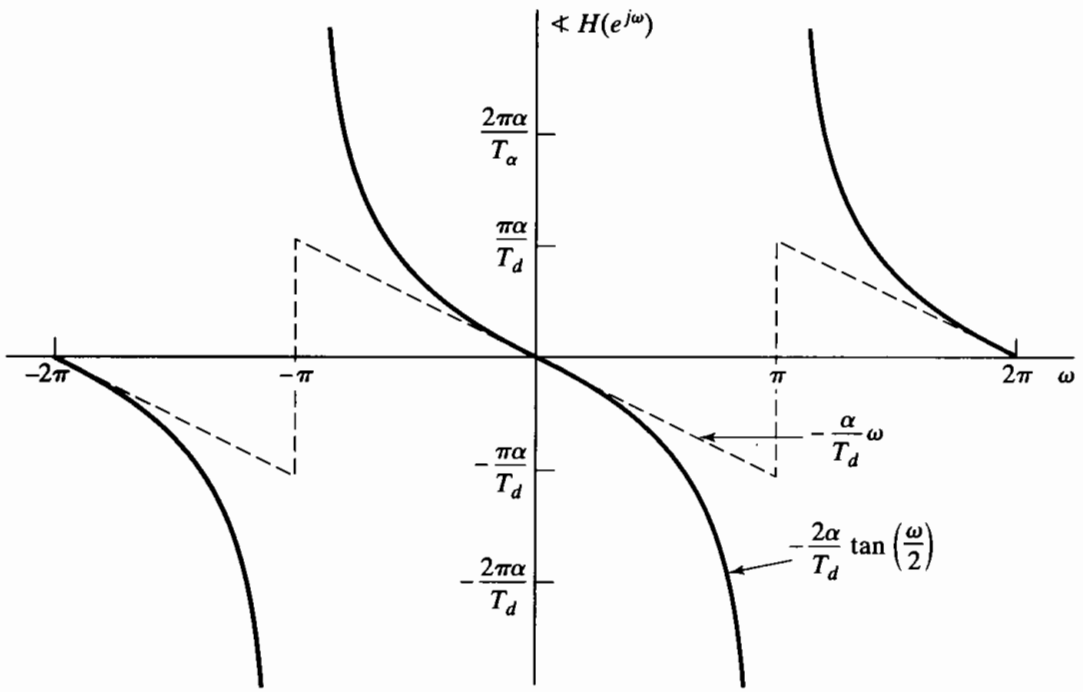


Figure 7.9 Illustration of the effect of the bilinear transformation on a linear phase characteristic. (Dashed line is linear phase and solid line is phase resulting from bilinear transformation.)

piecewise-constant frequency magnitude characteristics, such as highpass, lowpass and bandpass filters. As demonstrated in Example 7.2, impulse invariance can also be used to design lowpass filters. However, impulse invariance cannot be used to map highpass continuous-time designs to highpass discrete-time designs, since highpass continuous-time filters are not bandlimited.

In Example 4.5, we discussed a class of filters often referred to as discrete-time differentiators. A significant feature of the frequency response of this class of filters is that it is linear with frequency. The nonlinear warping of the frequency axis introduced by the bilinear transformation will not preserve that property. Consequently, the bilinear transformation applied to a continuous-time differentiator will not result in a discrete-time differentiator. However, impulse invariance applied to an appropriately bandlimited continuous-time differentiator will result in a discrete-time differentiator.

7.1.3 Examples of Bilinear Transformation Design

In the following discussion, we present a number of examples to illustrate IIR filter design using the bilinear transformation. Example 7.3 serves to illustrate the design procedure based on the bilinear transformation, in comparison with the use of impulse invariance. Examples 7.4, 7.5, and 7.6 illustrate a Butterworth, Chebyshev, and elliptic filter, respectively, each designed to the same specifications using the bilinear transformation.

Example 7.3 Bilinear Transformation of a Butterworth Filter

Consider the discrete-time filter specifications of Example 7.2, in which we illustrated the impulse invariance technique for the design of a discrete-time filter. The specifications on the discrete-time filter are

$$0.89125 \leq |H(e^{j\omega})| \leq 1, \qquad 0 \leq \omega \leq 0.2\pi, \tag{7.30a}$$

$$|H(e^{j\omega})| \leq 0.17783, \qquad 0.3\pi \leq \omega \leq \pi. \tag{7.30b}$$

In carrying out the design using the bilinear transformation, the critical frequencies of the discrete-time filter must be prewarped to the corresponding continuous-time frequencies using Eq. (7.28) so that the frequency distortion inherent in the bilinear transformation will map the continuous-time frequencies back to the correct discrete-time critical frequencies. For this specific filter, with $|H_c(j\Omega)|$ representing the magnitude-response function of the continuous-time filter, we require that

$$0.89125 \leq |H_c(j\Omega)| \leq 1, \qquad 0 \leq \Omega \leq \frac{2}{T_d}\tan\left(\frac{0.2\pi}{2}\right), \tag{7.31a}$$

$$|H_c(j\Omega)| \leq 0.17783, \qquad \frac{2}{T_d}\tan\left(\frac{0.3\pi}{2}\right) \leq \Omega \leq \infty. \tag{7.31b}$$

For convenience, we choose $T_d = 1$. Also, as with Example 7.2, since a continuous-time Butterworth filter has a monotonic magnitude response, we can equivalently require that

$$|H_c(j2\tan(0.1\pi))| \geq 0.89125 \tag{7.32a}$$

and

$$|H_c(j2\tan(0.15\pi))| \leq 0.17783. \tag{7.32b}$$

The form of the magnitude-squared function for the Butterworth filter is

$$|H_c(j\Omega)|^2 = \frac{1}{1+(\Omega/\Omega_c)^{2N}}. \tag{7.33}$$

Solving for N and Ω_c with the equality sign in Eqs. (7.32a) and (7.32b), we obtain

$$1+\left(\frac{2\tan(0.1\pi)}{\Omega_c}\right)^{2N} = \left(\frac{1}{0.89}\right)^2 \tag{7.34a}$$

and

$$1+\left(\frac{2\tan(0.15\pi)}{\Omega_c}\right)^{2N} = \left(\frac{1}{0.178}\right)^2, \tag{7.34b}$$

and solving for N in Eqs. (7.34a) and (7.34b) gives

$$N = \frac{\log\left[\left(\left(\frac{1}{0.178}\right)^2-1\right)\Big/\left(\left(\frac{1}{0.89}\right)^2-1\right)\right]}{2\log[\tan(0.15\pi)/\tan(0.1\pi)]} \tag{7.35}$$
$$= 5.305.$$

Since N must be an integer, we choose $N = 6$. Substituting $N = 6$ into Eq. (7.34b), we obtain $\Omega_c = 0.766$. For this value of Ω_c, the passband specifications are exceeded and the stopband specifications are met exactly. This is reasonable for the bilinear transformation, since we do not have to be concerned with aliasing. That is, with proper prewarping, we can be certain that the resulting discrete-time filter will meet the specifications exactly at the desired stopband edge.

In the s-plane, the 12 poles of the magnitude-squared function are uniformly distributed in angle on a circle of radius 0.766, as shown in Figure 7.10. The system function of the continuous-time filter obtained by selecting the left half-plane

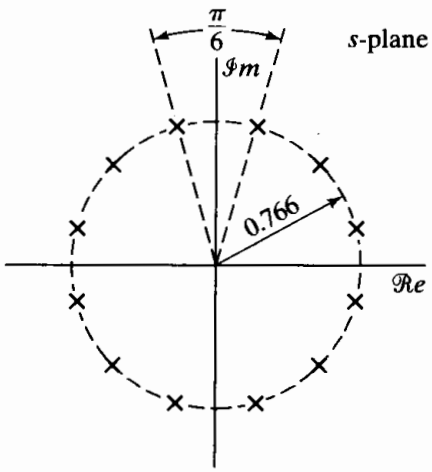


Figure 7.10 s-plane locations for poles of $H_c(s)H_c(-s)$ for sixth-order Butterworth filter in Example 7.3.

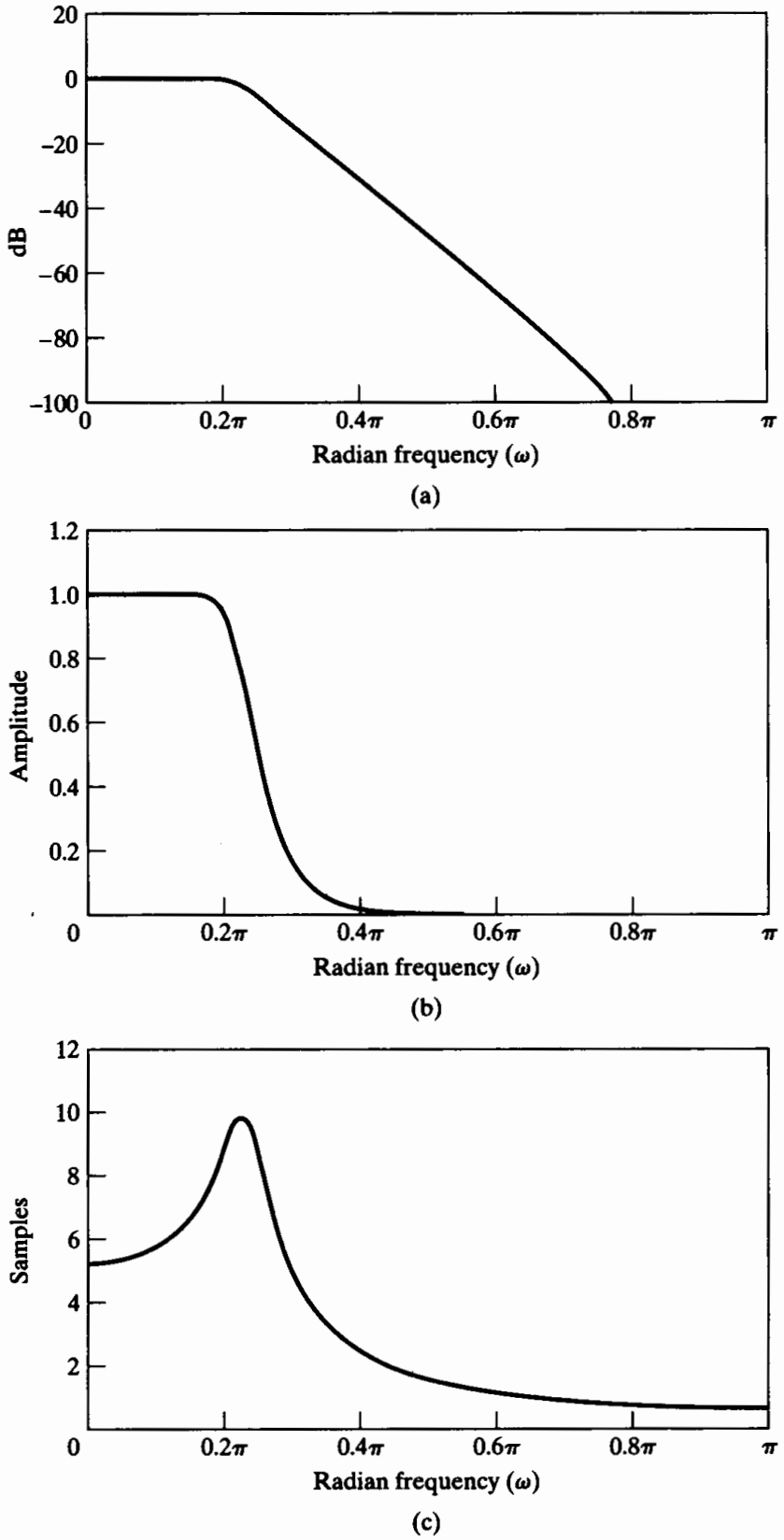


Figure 7.11 Frequency response of sixth-order Butterworth filter transformed by bilinear transform. (a) Log magnitude in dB. (b) Magnitude. (c) Group delay.

poles is

$$H_c(s) = \frac{0.20238}{(s^2 + 0.3996s + 0.5871)(s^2 + 1.0836s + 0.5871)(s^2 + 1.4802s + 0.5871)}. \tag{7.36}$$

The system function for the discrete-time filter is then obtained by applying the bilinear transformation to $H_c(s)$ with $T_d = 1$. The result is

$$H(z) = \frac{0.0007378(1 + z^{-1})^6}{(1 - 1.2686z^{-1} + 0.7051z^{-2})(1 - 1.0106z^{-1} + 0.3583z^{-2})} \times \frac{1}{(1 - 0.9044z^{-1} + 0.2155z^{-2})}. \tag{7.37}$$

The magnitude, log magnitude, and group delay of the frequency response of the discrete-time filter are shown in Figure 7.11. At $\omega = 0.2\pi$ the log magnitude is -0.56 dB, and at $\omega = 0.3\pi$ the log magnitude is exactly -15 dB.

Since the bilinear transformation maps the entire $j\Omega$-axis of the s-plane onto the unit circle in the z-plane, the magnitude response of the discrete-time filter falls off much more rapidly than that of the original continuous-time filter. In particular, the behavior of $H(e^{j\omega})$ at $\omega = \pi$ corresponds to the behavior of $H_c(j\Omega)$ at $\Omega = \infty$. Therefore, since the continuous-time Butterworth filter has a sixth-order zero at $s = \infty$, the resulting discrete-time filter has a sixth-order zero at $z = -1$.

It is interesting to note that, since the general form of the Nth-order Butterworth continuous-time filter is as given by Eq. (7.33), and since ω and Ω are related by Eq. (7.28), it follows that the general Nth-order Butterworth discrete-time filter has magnitude-squared function

$$|H(e^{j\omega})|^2 = \frac{1}{1 + \left(\dfrac{\tan(\omega/2)}{\tan(\omega_c/2)}\right)^{2N}}, \tag{7.38}$$

where $\tan(\omega_c/2) = \Omega_c T_d/2$.

The frequency-response function of Eq. (7.38) has the same properties as the continuous-time Butterworth response; i.e., it is maximally flat[3] and $|H(e^{j\omega_c})|^2 = 0.5$. However, the function in Eq. (7.38) is periodic with period 2π and falls off more sharply than the continuous-time Butterworth response.

We do not design discrete-time Butterworth filters directly by starting with Eq. (7.38), because it is not straightforward to determine the z-plane locations of the poles (all the zeros are at $z = -1$) associated with the magnitude-squared function of Eq. (7.38). It is necessary to determine the poles so as to factor the magnitude-squared

[3]The first $(2N - 1)$ derivatives of $|H(e^{j\omega})|^2$ are zero at $\omega = 0$.

function into $H(z)H(z^{-1})$ and thereby determine $H(z)$. It is much easier to find the s-plane pole locations (all the zeros are at infinity), factor the continuous-time system function, and then transform the left half-plane poles by the bilinear transformation as we did in Example 7.3.

Equations of the form of Eq. (7.38) may also be obtained for discrete-time Chebyshev filters, but the same difficulties arise in their use. Thus, the two-step approach just described has become the established method of designing IIR frequency-selective filters.

The major approximation methods for frequency-selective IIR analog filters are the Butterworth, Chebyshev, and elliptic function approximation methods. The details of these methods can be found in Guillemin (1957), Daniels (1974), Weinberg (1975), and Lam (1979). The methods are generally explained and developed in terms of lowpass filter approximations. This is the approach followed in Appendix B, where we summarize the essential features of some of the methods. In the next three examples, we illustrate the realization of a set of filter specifications for each of these classes of filters. The details of the design computations are not presented, since they are tedious and lengthy and are best carried out by computer programs that incorporate the appropriate closed-form design equations.

The lowpass discrete-time filter specifications for these examples are those used in Example 7.1, i.e.,

$$0.99 \leq |H(e^{j\omega})| \leq 1.01, \qquad |\omega| \leq 0.4\pi, \tag{7.39a}$$

and

$$|H(e^{j\omega})| \leq 0.001, \qquad 0.6\pi \leq |\omega| \leq \pi. \tag{7.39b}$$

In terms of the tolerance scheme of Figure 7.2(b), $\delta_1 = 0.01$, $\delta_2 = 0.001$, $\omega_p = 0.4\pi$, and $\omega_s = 0.6\pi$. These specifications are sufficient to determine the input parameters to the Butterworth, Chebyshev, and elliptic design formulas. Note that the specifications are only on the magnitudes of the frequency response. The phase is implicitly determined by the nature of the approximating functions.

Example 7.4 Butterworth Approximation

For the specification of Eqs. (7.39a) and (7.39b), the Butterworth approximation method requires a system of 14th order. The frequency response of the discrete-time filter that results from the bilinear transformation of the appropriate prewarped Butterworth filter is shown in Figure 7.12. Figure 7.12(a) shows the log magnitude in dB, Figure 7.12(b) shows the magnitude of $H(e^{j\omega})$ in the passband only, and Figure 7.12(c) shows the group delay of the filter. From these plots, we see that the Butterworth frequency response decreases monotonically with frequency and the gain of the filter becomes very small above about $\omega = 0.7\pi$. Note from Figure 7.12(b) that in this example the Butterworth frequency response has been normalized so that it has gain greater than unity in the passband as is allowed in the specifications in Eqs. (7.39).

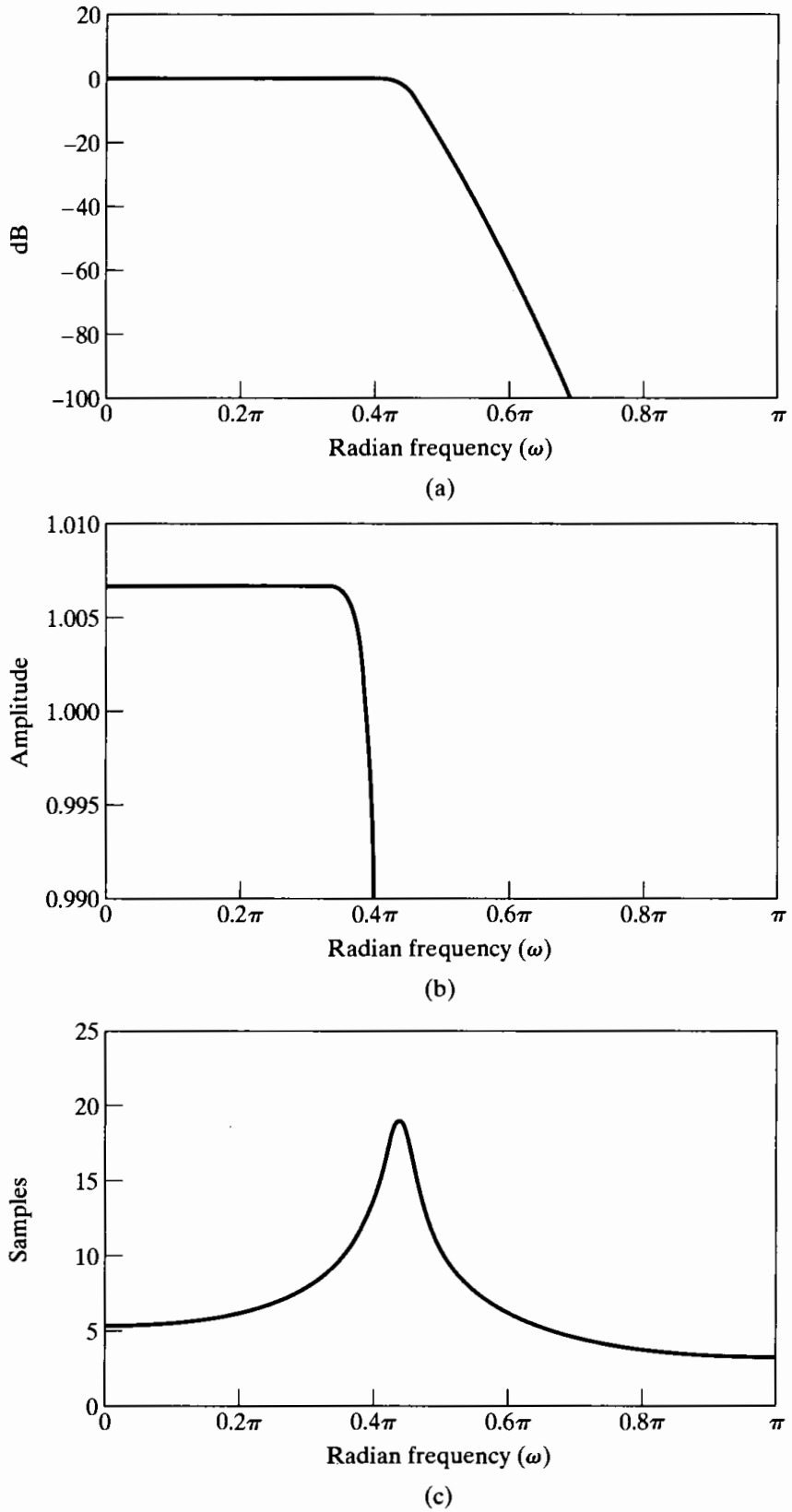


Figure 7.12 Frequency response of 14th-order Butterworth filter in Example 7.4. (a) Log magnitude in dB. (b) Detailed plot of magnitude in passband. (c) Group delay.

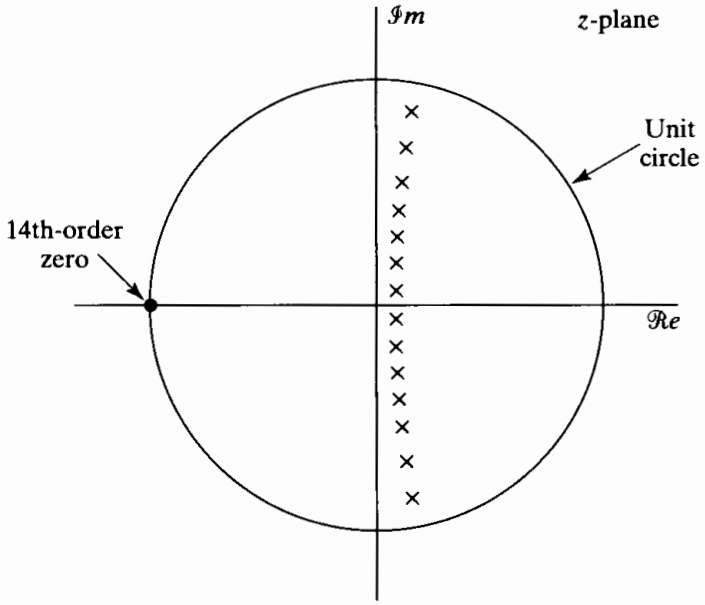


Figure 7.13 Pole–zero plot of 14th-order Butterworth filter in Example 7.4.

In the Butterworth example, the specifications are exceeded at the passband and stopband edges because of rounding the order up to the next integer. However, the specifications are *far* exceeded in the stopband. The reason for this is evident from Figure 7.13, which shows the pole–zero plot for the 14th-order Butterworth filter. Because the continuous-time Butterworth filter has 14 zeros at $s = \infty$, the bilinear transformation creates 14 zeros at $z = -1$ for the discrete-time filter. It is reasonable to expect that a lower order filter might still satisfy the specifications, even if it did not exceed them so greatly in the stopband. This expectation motivates the use of Chebyshev or equiripple approximation.

Example 7.5 Chebyshev Approximation

This method has two forms. Chebyshev type I approximations have equiripple behavior in the passband, and Chebyshev type II approximations have equiripple behavior in the stopband. Both methods lead to the same order for a given set of specifications. For the specifications of Eqs. (7.39a) and (7.39b), the required order is 8 rather than 14, as for the Butterworth approximation. Figure 7.14 shows the log magnitude, passband magnitude, and group delay for the type I approximation to the specifications of Eqs. (7.39a) and (7.39b). Note that the frequency response oscillates with equal maximum error on either side of the desired gain of unity in the passband.

Figure 7.15 shows the frequency-response functions for the Chebyshev type II approximation to the specifications of Eqs. (7.39a) and (7.39b). In this case, the equiripple approximation behavior is in the stopband. The pole–zero plots for the Chebyshev filters are shown in Figure 7.16. Note that the Chebyshev type I system is similar to the Butterworth system in that it has all eight of its zeros at $z = -1$. On the other hand, the type II system has its zeros arrayed on the unit circle. These zeros

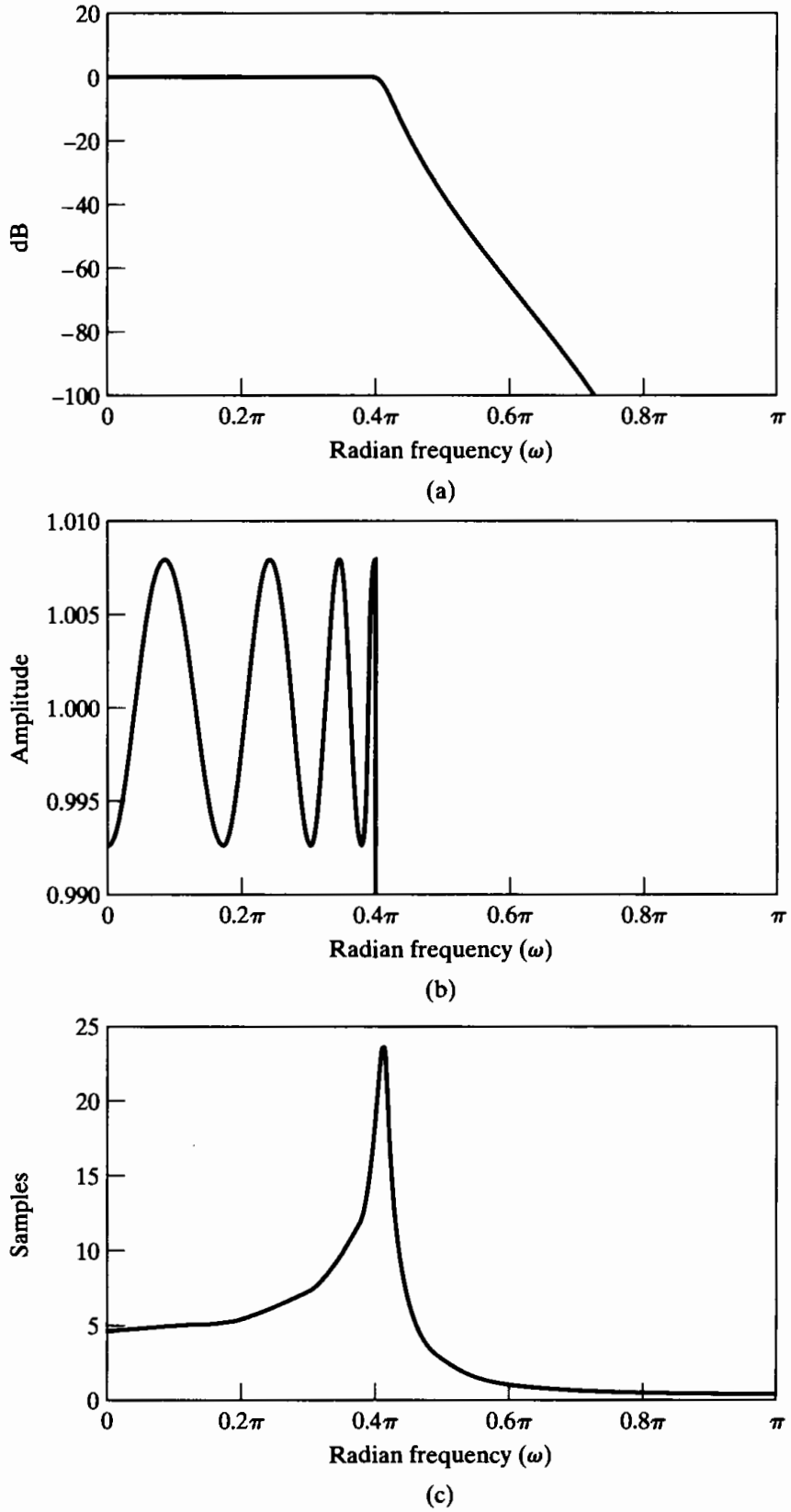


Figure 7.14 Frequency response of eighth-order Chebyshev type I filter in Example 7.5. (a) Log magnitude in dB. (b) Detailed plot of magnitude in passband. (c) Group delay.

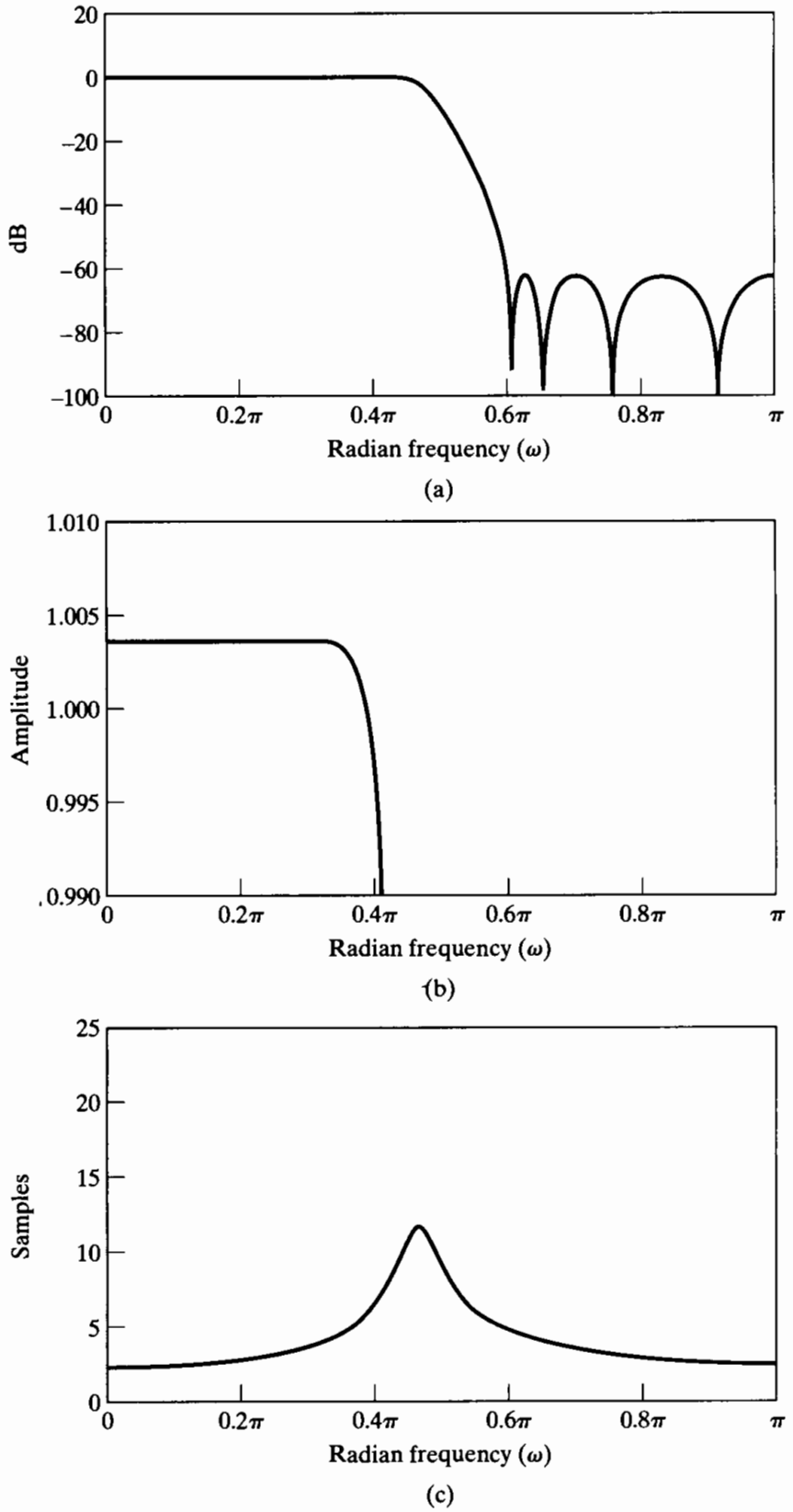


Figure 7.15 Frequency response of eighth-order Chebyshev type II filter in Example 7.5. (a) Log magnitude in dB. (b) Detailed plot of magnitude in passband. (c) Group delay.

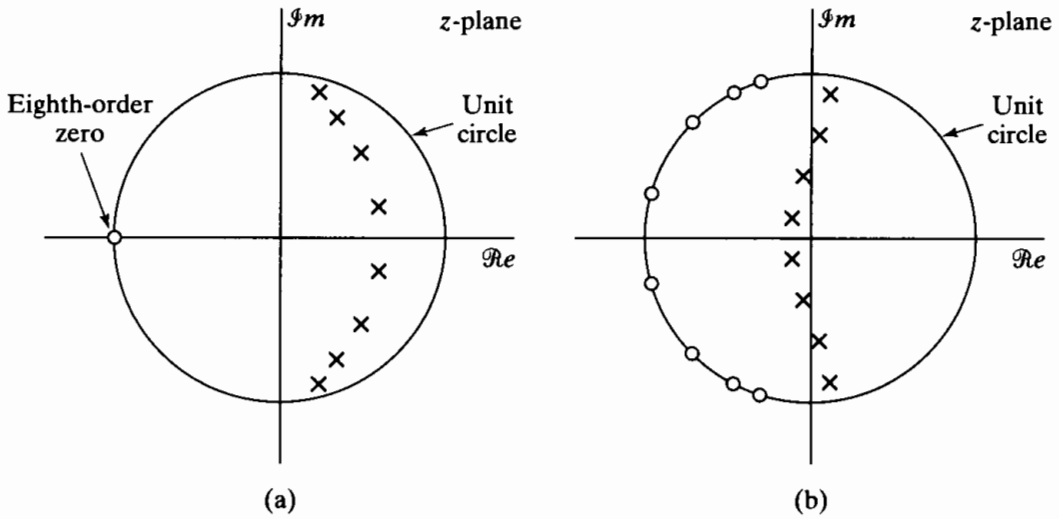


Figure 7.16 Pole–zero plot of eighth-order Chebyshev filters in Example 7.5. (a) Type I. (b) Type II.

are positioned by the design equations so as to achieve the equiripple behavior in the stopband.

In both cases of Chebyshev approximation, the monotonic behavior in either the stopband or the passband suggests that perhaps a lower order system might be obtained if equiripple approximation were used in both the passband and the stopband. Indeed, it can be shown (see Papoulis, 1957) that for fixed values of δ_1, δ_2, ω_p, and ω_s in the tolerance scheme of Figure 7.2(b), the lowest order filter is obtained when the approximation error ripples equally between the extremes of the two approximation bands. Since this equiripple behavior is achieved with a rational function that involves elliptic functions, such systems are generally called elliptic filters.

Example 7.6 Elliptic Approximation[4]

The specifications of Eqs. (7.39a) and (7.39b) are met by an elliptic filter of order six. This is the *lowest* order rational function approximation to the specifications. Figure 7.17 clearly shows the equiripple behavior in both approximation bands. Figure 7.18 shows that the elliptic filter, like the Chebyshev type II, has its zeros arrayed in the stopband region of the unit circle.

[4]The design equations for elliptic filters are too involved to be appropriately summarized in Appendix B. They can be found in Storer (1957), Weinberg (1975), and Parks and Burrus (1987). A program for elliptic filter design was given by Gray and Markel (1976), and extensive tables for elliptic filter designs are available in Zverev (1967). All three types of filters can be designed using functions in the Signal Processing Toolbox of MATLAB®.

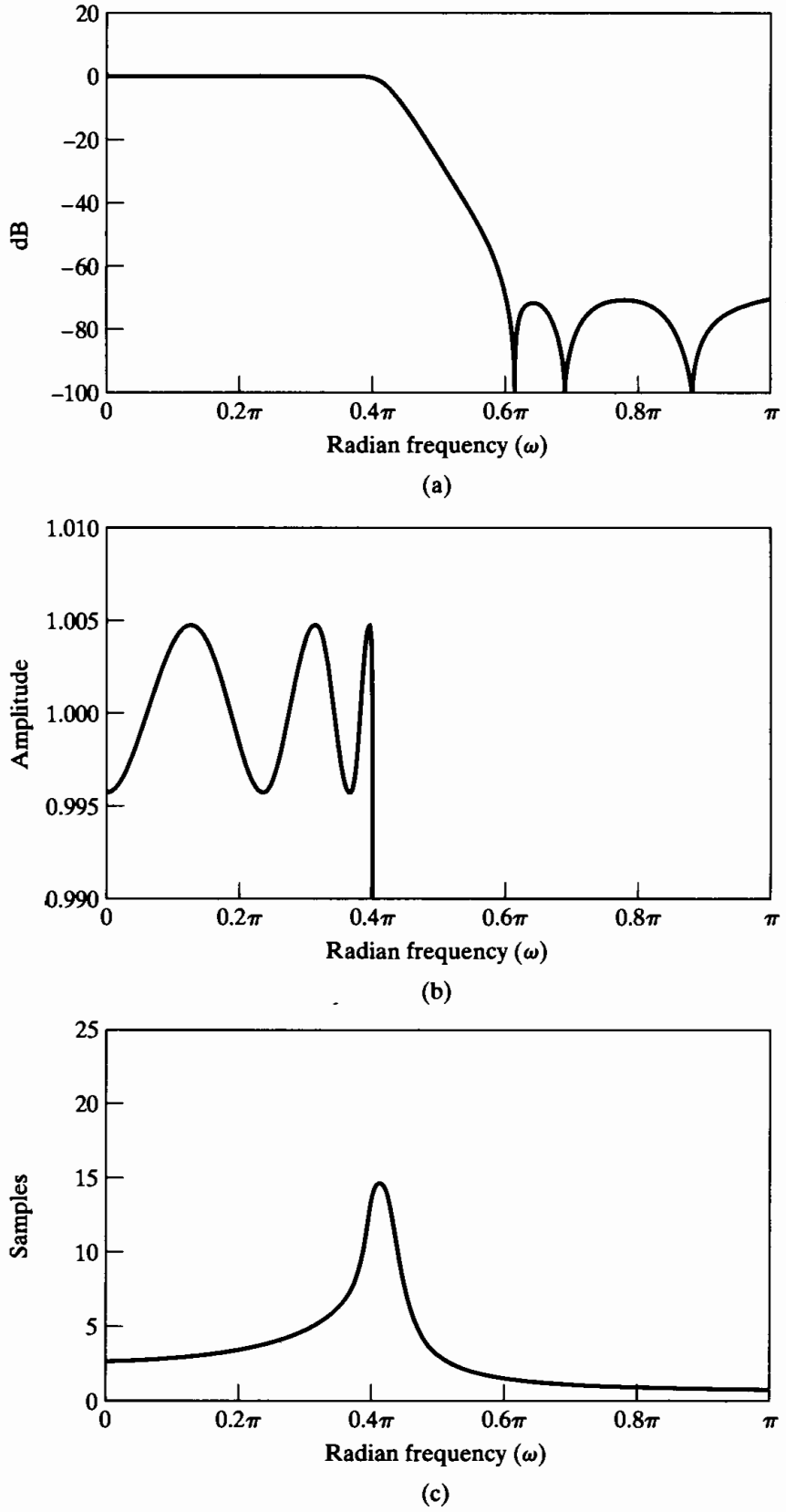


Figure 7.17 Frequency response of sixth-order elliptic filter in Example 7.6. (a) Log magnitude in dB. (b) Detailed plot of magnitude in passband. (c) Group delay.

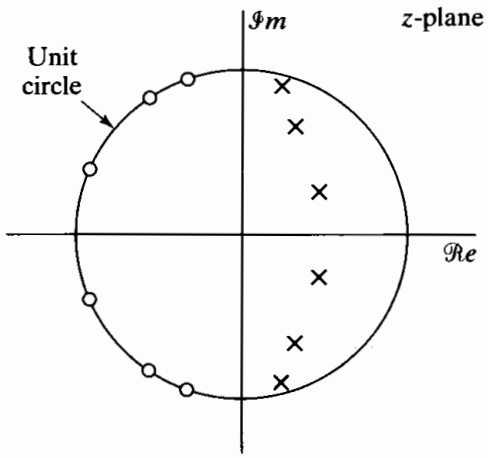


Figure 7.18 Pole–zero plot of sixth-order elliptic filter in Example 7.6.

Bilinear transformation of analog filters designed by Butterworth, Chebyshev, or elliptic approximation methods is a standard method of design of IIR discrete-time filters. The previous examples illustrate several important general features of such filters. In all cases, the resulting system function $H(z)$ has all its zeros on the unit circle and (for stability) all its poles inside the unit circle. As a result, all the approximation methods yield digital filters with nonconstant group delay or, equivalently, nonlinear phase. The greatest deviation from constant group delay occurs in all cases at the edge of the passband or in the transition band. In general, the Chebyshev type II approximation method yields the smallest delay in the passband and the widest region of the passband over which the group delay is approximately constant. However, if phase linearity is not an issue, then elliptic approximation yields the lowest order system function, and therefore, elliptic filters will generally require the least computation to implement a given filter specification.

7.2 DESIGN OF FIR FILTERS BY WINDOWING

As discussed in Section 7.1, commonly used techniques for the design of IIR filters are based on transformations of continuous-time IIR systems into discrete-time IIR systems. This is partly because continuous-time filter design was a highly advanced art before discrete-time filters were of interest and partly because of the difficulty of implementing a noniterative direct design method for IIR filters.

In contrast, FIR filters are almost entirely restricted to discrete-time implementations. Consequently, the design techniques for FIR filters are based on directly approximating the desired frequency response of the discrete-time system. Furthermore, most techniques for approximating the magnitude response of an FIR system assume a linear phase constraint, thereby avoiding the problem of spectrum factorization that complicates the direct design of IIR filters.

The simplest method of FIR filter design is called the *window method.* This method generally begins with an ideal desired frequency response that can be represented as

$$H_d(e^{j\omega}) = \sum_{n=-\infty}^{\infty} h_d[n]e^{-j\omega n}, \tag{7.40}$$

where $h_d[n]$ is the corresponding impulse response sequence, which can be expressed in terms of $H_d(e^{j\omega})$ as

$$h_d[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_d(e^{j\omega})e^{j\omega n}d\omega. \tag{7.41}$$

Many idealized systems are defined by piecewise-constant or piecewise-functional frequency responses with discontinuities at the boundaries between bands. As a result, these systems have impulse responses that are noncausal and infinitely long. The most straightforward approach to obtaining a causal FIR approximation to such systems is to truncate the ideal response. Equation (7.40) can be thought of as a Fourier series representation of the periodic frequency response $H_d(e^{j\omega})$, with the sequence $h_d[n]$ playing the role of the Fourier coefficients. Thus, the approximation of an ideal filter by truncation of the ideal impulse response is identical to the issue of the convergence of Fourier series, a subject that has received a great deal of study. A particularly important concept from this theory is the Gibbs phenomenon, which was discussed in Example 2.22. In the following discussion, we will see how this nonuniform convergence phenomenon manifests itself in the design of FIR filters.

The simplest way to obtain a causal FIR filter from $h_d[n]$ is to define a new system with impulse response $h[n]$ given by[5]

$$h[n] = \begin{cases} h_d[n], & 0 \le n \le M, \\ 0, & \text{otherwise.} \end{cases} \tag{7.42}$$

More generally, we can represent $h[n]$ as the product of the desired impulse response and a finite-duration "window" $w[n]$; i.e.,

$$h[n] = h_d[n]w[n], \tag{7.43}$$

where, for simple truncation as in Eq. (7.42), the window is the *rectangular window*

$$w[n] = \begin{cases} 1, & 0 \le n \le M, \\ 0, & \text{otherwise.} \end{cases} \tag{7.44}$$

It follows from the modulation, or windowing, theorem (Section 2.9.7) that

$$H(e^{j\omega}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_d(e^{j\theta})W(e^{j(\omega-\theta)})d\theta. \tag{7.45}$$

That is, $H(e^{j\omega})$ is the periodic convolution of the desired ideal frequency response with the Fourier transform of the window. Thus, the frequency response $H(e^{j\omega})$ will be a "smeared" version of the desired response $H_d(e^{j\omega})$. Figure 7.19(a) depicts typical functions $H_d(e^{j\theta})$ and $W(e^{j(\omega-\theta)})$, as required in Eq. (7.45).

If $w[n] = 1$ for all n (i.e., if we do not truncate at all), $W(e^{j\omega})$ is a periodic impulse train with period 2π, and therefore, $H(e^{j\omega}) = H_d(e^{j\omega})$. This interpretation suggests

[5]The notation for FIR systems was established in Chapter 5. That is, M is the order of the system function polynomial. Thus, $(M+1)$ is the length, or duration, of the impulse response. Often in the literature, N is used for the length of the impulse response of an FIR filter; however, we have used N to denote the order of the denominator polynomial in the system function of an IIR filter. Thus, to avoid confusion and maintain consistency throughout this book, we will always consider the length of the impulse response of an FIR filter to be $(M+1)$.

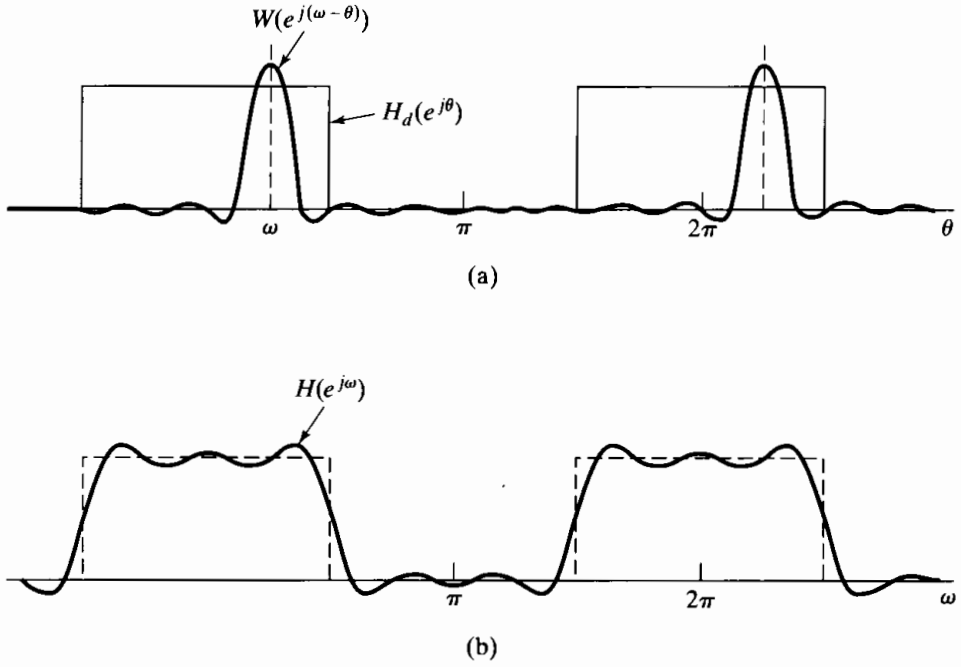


Figure 7.19 (a) Convolution process implied by truncation of the ideal impulse response. (b) Typical approximation resulting from windowing the ideal impulse response.

that if $w[n]$ is chosen so that $W(e^{j\omega})$ is concentrated in a narrow band of frequencies around $\omega = 0$, then $H(e^{j\omega})$ will "look like" $H_d(e^{j\omega})$, except where $H_d(e^{j\omega})$ changes very abruptly. Consequently, the choice of window is governed by the desire to have $w[n]$ as short as possible in duration, so as to minimize computation in the implementation of the filter, while having $W(e^{j\omega})$ approximate an impulse; that is, we want $W(e^{j\omega})$ to be highly concentrated in frequency so that the convolution of Eq. (7.45) faithfully reproduces the desired frequency response. These are conflicting requirements, as can be seen in the case of the rectangular window of Eq. (7.44), where

$$W(e^{j\omega}) = \sum_{n=0}^{M} e^{-j\omega n} = \frac{1 - e^{-j\omega(M+1)}}{1 - e^{-j\omega}} = e^{-j\omega M/2} \frac{\sin[\omega(M+1)/2]}{\sin(\omega/2)}. \tag{7.46}$$

The magnitude of the function $\sin[\omega(M+1)/2]\sin(\omega/2)$ is plotted in Figure 7.20 for the case $M = 7$. Note that $W(e^{j\omega})$ for the rectangular window has a generalized linear phase. As M increases, the width of the "main lobe" decreases. The main lobe is usually defined as the region between the first zero-crossings on either side of the origin. For the rectangular window, the width of the main lobe is $\Delta_{\omega_m} = 4\pi/(M+1)$. However, for the rectangular window, the side lobes are large, and in fact, as M increases, the peak amplitudes of the main lobe and the side lobes grow in a manner such that the area under each lobe is a constant while the width of each lobe decreases with M. Consequently, as $W(e^{j(\omega-\theta)})$ "slides by" a discontinuity of $H_d(e^{j\theta})$ with increasing ω, the integral of $W(e^{j(\omega-\theta)})H_d(e^{j\theta})$ will oscillate as each side lobe of $W(e^{j(\omega-\theta)})$ moves past the discontinuity. This result is depicted in Figure 7.19 (b). Since the area under

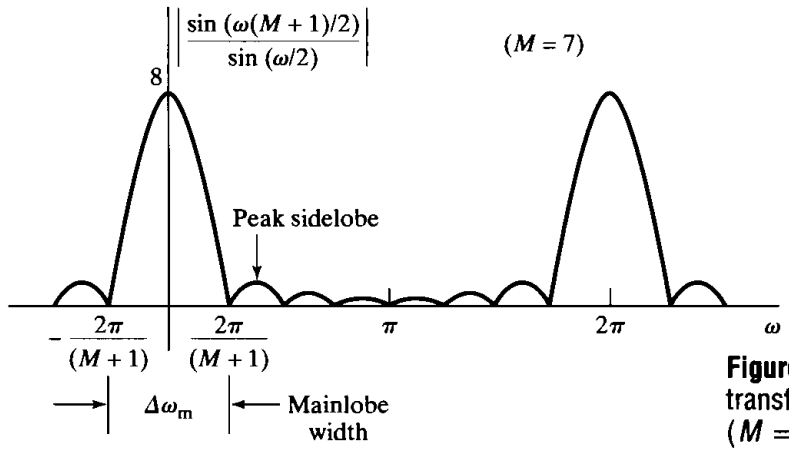


Figure 7.20 Magnitude of the Fourier transform of a rectangular window ($M = 7$).

each lobe remains constant with increasing M, the oscillations occur more rapidly, but do not decrease in amplitude as M increases.

In the theory of Fourier series, it is well known that this nonuniform convergence, the *Gibbs phenomenon,* can be moderated through the use of a less abrupt truncation of the Fourier series. By tapering the window smoothly to zero at each end, the height of the side lobes can be diminished; however, this is achieved at the expense of a wider main lobe and thus a wider transition at the discontinuity.

7.2.1 Properties of Commonly Used Windows

Some commonly used windows are shown in Figure 7.21.[6] These windows are defined by the following equations:

Rectangular

$$w[n] = \begin{cases} 1, & 0 \le n \le M, \\ 0, & \text{otherwise} \end{cases} \tag{7.47a}$$

Bartlett (triangular)

$$w[n] = \begin{cases} 2n/M, & 0 \le n \le M/2, \\ 2 - 2n/M, & M/2 < n \le M, \\ 0, & \text{otherwise} \end{cases} \tag{7.47b}$$

Hanning

$$w[n] = \begin{cases} 0.5 - 0.5\cos(2\pi n/M), & 0 \le n \le M, \\ 0, & \text{otherwise} \end{cases} \tag{7.47c}$$

Hamming

$$w[n] = \begin{cases} 0.54 - 0.46\cos(2\pi n/M), & 0 \le n \le M, \\ 0, & \text{otherwise} \end{cases} \tag{7.47d}$$

[6]The Bartlett, Hanning, Hamming, and Blackman windows are all named after their originators. The Hanning window is associated with Julius von Hann, an Austrian meteorologist, and is sometimes referred to as the Hann window. The term "hanning" was used by Blackman and Tukey (1958) to describe the operation of applying this window to a signal and has since become the most widely used name for the window, with varying preferences for the choice of "Hanning" or "hanning."

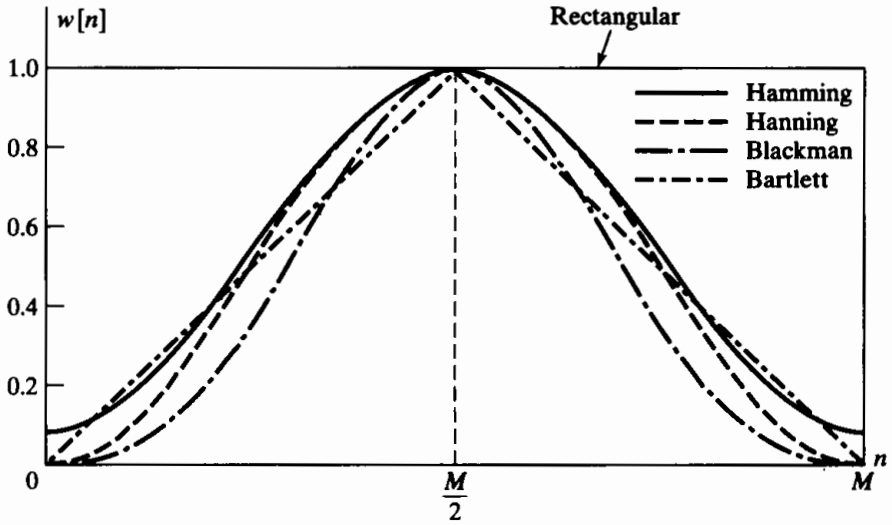


Figure 7.21 Commonly used windows.

Blackman

$$w[n] = \begin{cases} 0.42 - 0.5\cos(2\pi n/M) + 0.08\cos(4\pi n/M), & 0 \le n \le M, \\ 0, & \text{otherwise} \end{cases} \tag{7.47e}$$

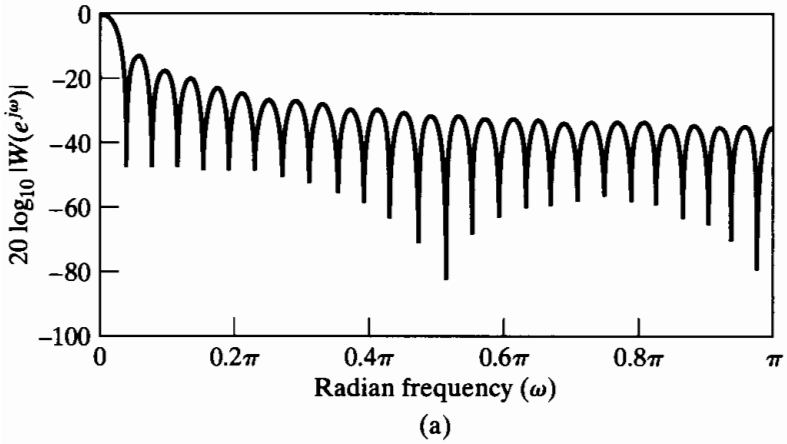
(For convenience, Figure 7.21 shows these windows plotted as functions of a continuous variable; however, as specified in Eqs. (7.47), the window sequence is defined only at integer values of n.)

As will be discussed in Chapter 10, the windows defined in Eqs. (7.47) are commonly used for spectrum analysis as well as for FIR filter design. They have the desirable property that their Fourier transforms are concentrated around $\omega = 0$, and they have a simple functional form that allows them to be computed easily. The Fourier transform of the Bartlett window can be expressed as a product of Fourier transforms of rectangular windows, and the Fourier transforms of the other windows can be expressed as sums of frequency-shifted Fourier transforms of the rectangular window, as given by Eq. (7.46). (See Problem 7.34.)

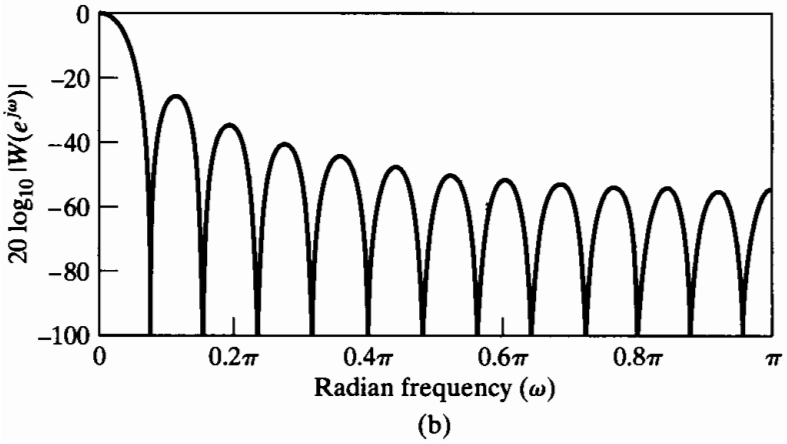
The function $20\log_{10}|W(e^{j\omega})|$ is plotted in Figure 7.22 for each of these windows with $M = 50$. The rectangular window clearly has the narrowest main lobe, and thus, for a given length, it should yield the sharpest transitions of $H(e^{j\omega})$ at a discontinuity of $H_d(e^{j\omega})$. However, the first side lobe is only about 13 dB below the main peak, resulting in oscillations of $H(e^{j\omega})$ of considerable size around discontinuities of $H_d(e^{j\omega})$. Table 7.1, which compares the windows of Eqs. (7.47), shows that, by tapering the window smoothly to zero, as with the Hamming, Hanning, and Blackman windows, the side lobes (second column) are greatly reduced in amplitude; however, the price paid is a much wider main lobe (third column) and thus wider transitions at discontinuities of $H_d(e^{j\omega})$. The other columns of Table 7.1 will be discussed later.

7.2.2 Incorporation of Generalized Linear Phase

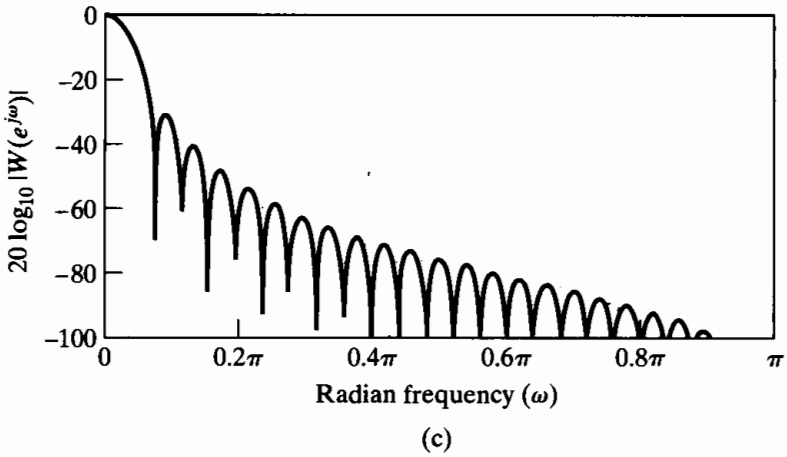
In designing many types of FIR filters, it is desirable to obtain causal systems with a generalized linear phase response. All the windows of Eqs. (7.47) have been defined in



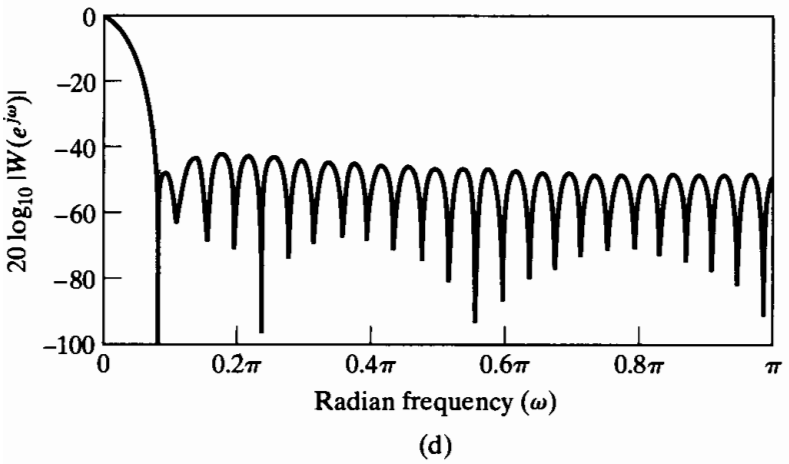
(a)



(b)

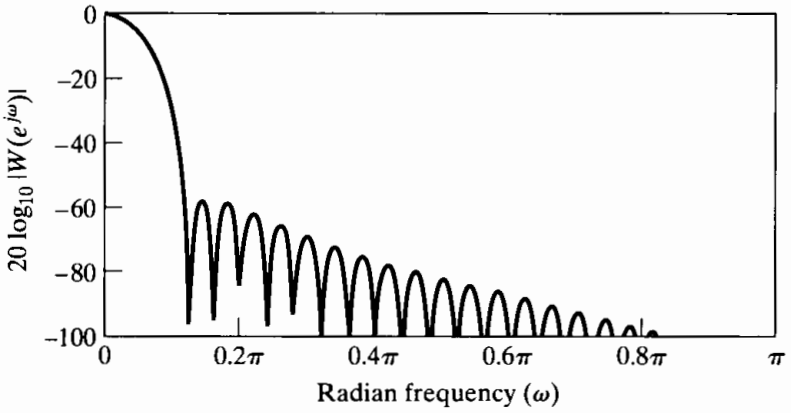


(c)



(d)

Figure 7.22 Fourier transforms (log magnitude) of windows of Figure 7.21. with $M = 50$. (a) Rectangular. (b) Bartlett. (c) Hanning. (d) Hamming. (e) Blackman.



(e)

Figure 7.22 (*Continued*) (e) Blackman.

TABLE 7.1 COMPARISON OF COMMONLY USED WINDOWS

Type of Window	Peak Side-Lobe Amplitude (Relative)	Approximate Width of Main Lobe	Peak Approximation Error, $20 \log_{10} \delta$ (dB)	Equivalent Kaiser Window, β	Transition Width of Equivalent Kaiser Window
Rectangular	−13	$4\pi/(M+1)$	−21	0	$1.81\pi/M$
Bartlett	−25	$8\pi/M$	−25	1.33	$2.37\pi/M$
Hanning	−31	$8\pi/M$	−44	3.86	$5.01\pi/M$
Hamming	−41	$8\pi/M$	−53	4.86	$6.27\pi/M$
Blackman	−57	$12\pi/M$	−74	7.04	$9.19\pi/M$

anticipation of this need. Specifically, note that all the windows have the property that

$$w[n] = \begin{cases} w[M-n], & 0 \le n \le M, \\ 0, & \text{otherwise;} \end{cases} \tag{7.48}$$

i.e., they are symmetric about the point $M/2$. As a result, their Fourier transforms are of the form

$$W(e^{j\omega}) = W_e(e^{j\omega})e^{-j\omega M/2}, \tag{7.49}$$

where $W_e(e^{j\omega})$ is a real, even function of ω. This is illustrated by Eq. (7.46). The convention of Eq. (7.48) leads to causal filters in general, and if the desired impulse response is also symmetric about $M/2$, i.e., if $h_d[M-n] = h_d[n]$, then the windowed impulse response will also have that symmetry, and the resulting frequency response will have a generalized linear phase; that is,

$$H(e^{j\omega}) = A_e(e^{j\omega})e^{-j\omega M/2}, \tag{7.50}$$

where $A_e(e^{j\omega})$ is real and is an even function of ω. Similarly, if the desired impulse response is antisymmetric about $M/2$, i.e., if $h_d[M-n] = -h_d[n]$, then the windowed impulse response will also be antisymmetric about $M/2$, and the resulting frequency response will have a generalized linear phase with a constant phase shift of ninety degrees; i.e.,

$$H(e^{j\omega}) = jA_o(e^{j\omega})e^{-j\omega M/2}, \tag{7.51}$$

where $A_o(e^{j\omega})$ is real and is an odd function of ω.

Although the preceding statements are straightforward if we consider the product of the symmetric window with the symmetric (or antisymmetric) desired impulse response, it is useful to consider the frequency-domain representation. Suppose $h_d[M-n] = h_d[n]$. Then

$$H_d(e^{j\omega}) = H_e(e^{j\omega})e^{-j\omega M/2}, \tag{7.52}$$

where $H_e(e^{j\omega})$ is real and even.

If the window is symmetric, we can substitute Eqs. (7.49) and (7.52) into Eq. (7.45) to obtain

$$H(e^{j\omega}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_e(e^{j\theta})e^{-j\theta M/2}W_e(e^{j(\omega-\theta)})e^{-j(\omega-\theta)M/2}d\theta. \tag{7.53}$$

A simple manipulation of the phase factors leads to

$$H(e^{j\omega}) = A_e(e^{j\omega})e^{-j\omega M/2}, \tag{7.54}$$

where

$$A_e(e^{j\omega}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_e(e^{j\theta})W_e(e^{j(\omega-\theta)})d\theta. \tag{7.55}$$

Thus, we see that the resulting system has a generalized linear phase and, moreover, the real function $A_e(e^{j\omega})$ is the result of the periodic convolution of the real functions $H_e(e^{j\omega})$ and $W_e(e^{j\omega})$.

The detailed behavior of the convolution of Eq. (7.55) determines the magnitude response of the filter that results from windowing. The following example illustrates this for a linear-phase lowpass filter.

Example 7.7 Linear-Phase Lowpass Filter

The desired frequency response is defined as

$$H_{lp}(e^{j\omega}) = \begin{cases} e^{-j\omega M/2}, & |\omega| < \omega_c, \\ 0, & \omega_c < |\omega| \le \pi, \end{cases} \tag{7.56}$$

where the generalized linear phase factor has been incorporated into the definition of the ideal lowpass filter. The corresponding ideal impulse response is

$$h_{lp}[n] = \frac{1}{2\pi}\int_{-\omega_c}^{\omega_c} e^{-j\omega M/2}e^{j\omega n}d\omega = \frac{\sin[\omega_c(n-M/2)]}{\pi(n-M/2)} \tag{7.57}$$

for $-\infty < n < \infty$. It is easily shown that $h_{lp}[M-n] = h_{lp}[n]$, so if we use a symmetric window in the equation

$$h[n] = \frac{\sin[\omega_c(n-M/2)]}{\pi(n-M/2)}w[n], \tag{7.58}$$

then a linear-phase system will result.

The upper part of Figure 7.23 depicts the character of the amplitude response that would result for all the windows of Eqs. (7.47), except the Bartlett window, which is rarely used for filter design. (The Bartlett window would produce a monotonic function $A_e(e^{j\omega})$, because $W_e(e^{j\omega})$ is a positive function.) The figure displays the important properties of window method approximations to desired frequency responses that have step discontinuities. It applies accurately when ω_c is not close to zero or to π and when the width of the main lobe is smaller than $2\omega_c$. At the bottom of the figure

is a typical Fourier transform for a symmetric window (except for the linear phase). This function should be visualized in different positions as an aid in understanding the shape of the approximation $A_e(e^{j\omega})$ in the vicinity of ω_c.

When $\omega = \omega_c$, the symmetric function $W_e(e^{j(\omega-\theta)})$ is centered on the discontinuity, and about one-half its area contributes to $A_e(e^{j\omega})$. Similarly, we can see that the peak overshoot occurs when $W_e(e^{j(\omega-\theta)})$ is shifted such that the first negative side lobe on the right is just to the right of ω_c. Similarly, the peak negative undershoot occurs when the first negative side lobe on the left is just to the left of ω_c. This means that the distance between the peak ripples on either side of the discontinuity is approximately the main-lobe width $\Delta\omega_m$, as shown in Figure 7.23. The transition width $\Delta\omega$ as defined in the figure is therefore somewhat less than the main-lobe width. Finally, due to the symmetry of $W_e(e^{j(\omega-\theta)})$, the approximation tends to be symmetric around ω_c; i.e., the approximation overshoots by an amount δ in the passband and undershoots by the same amount in the stopband.

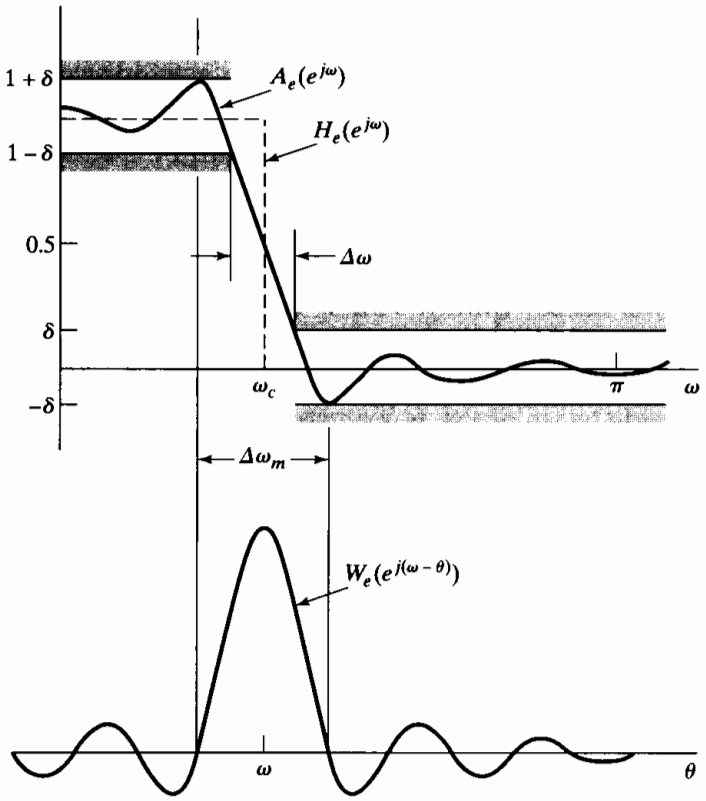


Figure 7.23 Illustration of type of approximation obtained at a discontinuity of the ideal frequency response.

The fourth column of Table 7.1 shows the peak approximation error (in dB) for the windows of Eqs. (7.47). Clearly, the windows with the smaller side lobes yield better approximations of the ideal response at a discontinuity . Also, the third column, which shows the width of the main lobe, suggests that narrower transition regions can be achieved by increasing M. Thus, through the choice of the shape and duration of the window, we can control the properties of the resulting FIR filter. However, trying different windows and adjusting lengths by trial and error is not a very satisfactory way to design filters. Fortunately, a simple formalization of the window method has been developed by Kaiser (1974).

7.2.3 The Kaiser Window Filter Design Method

The trade-off between the main-lobe width and side-lobe area can be quantified by seeking the window function that is maximally concentrated around $\omega = 0$ in the frequency domain. The issue was considered in depth in a series of classic papers by Slepian et al. (1961). The solution found in this work involves prolate spheroidal wave functions, which are difficult to compute and therefore unattractive for filter design. However, Kaiser (1966, 1974) found that a near-optimal window could be formed using the zeroth-order modified Bessel function of the first kind, a function that is much easier to compute. The Kaiser window is defined as

$$w[n] = \begin{cases} \dfrac{I_0[\beta(1-[(n-\alpha)/\alpha]^2)^{1/2}]}{I_0(\beta)}, & 0 \le n \le M, \\ 0, & \text{otherwise}, \end{cases} \tag{7.59}$$

where $\alpha = M/2$, and $I_0(\cdot)$ represents the zeroth-order modified Bessel function of the first kind. In contrast to the other windows in Eqs. (7.47), the Kaiser window has two parameters: the length $(M+1)$ and a shape parameter β. By varying $(M+1)$ and β, the window length and shape can be adjusted to trade side-lobe amplitude for main-lobe width. Figure 7.24(a) shows continuous envelopes of Kaiser windows of length $M+1 = 21$ for $\beta = 0, 3$, and 6. Notice from Eq. (7.59) that the case $\beta = 0$ reduces to the rectangular window. Figure 7.24(b) shows the corresponding Fourier transforms of the Kaiser windows in Figure 7.24(a). Figure 7.24(c) shows Fourier transforms of Kaiser windows with $\beta = 6$ and $M = 10$, 20, and 40. The plots in Figs. 7.24(b) and (c) clearly show that the desired trade-off can be achieved. If the window is tapered more, the side lobes of the Fourier transform become smaller, but the main lobe becomes wider. Figure 7.24(c) shows that increasing M while holding β constant causes the main lobe to decrease in width, but does not affect the amplitude of the side lobes. In fact, through extensive numerical experimentation, Kaiser obtained a pair of formulas that permit the filter designer to predict in advance the values of M and β needed to meet a given frequency-selective filter specification. Figure 7.23 is also typical of approximations obtained using the Kaiser window, and Kaiser (1974) found that, over a usefully wide range of conditions, the peak approximation error (δ in Figure 7.23) is determined by the choice of β. Given that δ is fixed, the passband cutoff frequency ω_p of the lowpass filter is defined to be the highest frequency such that $|H(e^{j\omega})| \ge 1 - \delta$. The stopband cutoff frequency ω_s is defined to be the lowest frequency such that $|H(e^{j\omega})| \le \delta$. Therefore, the transition region has width

$$\Delta\omega = \omega_s - \omega_p \tag{7.60}$$

for the lowpass filter approximation. Defining

$$A = -20\log_{10}\delta, \tag{7.61}$$

Kaiser determined empirically that the value of β needed to achieve a specified value of A is given by

$$\beta = \begin{cases} 0.1102(A-8.7), & A > 50, \\ 0.5842(A-21)^{0.4} + 0.07886(A-21), & 21 \le A \le 50, \\ 0.0, & A < 21. \end{cases} \tag{7.62}$$

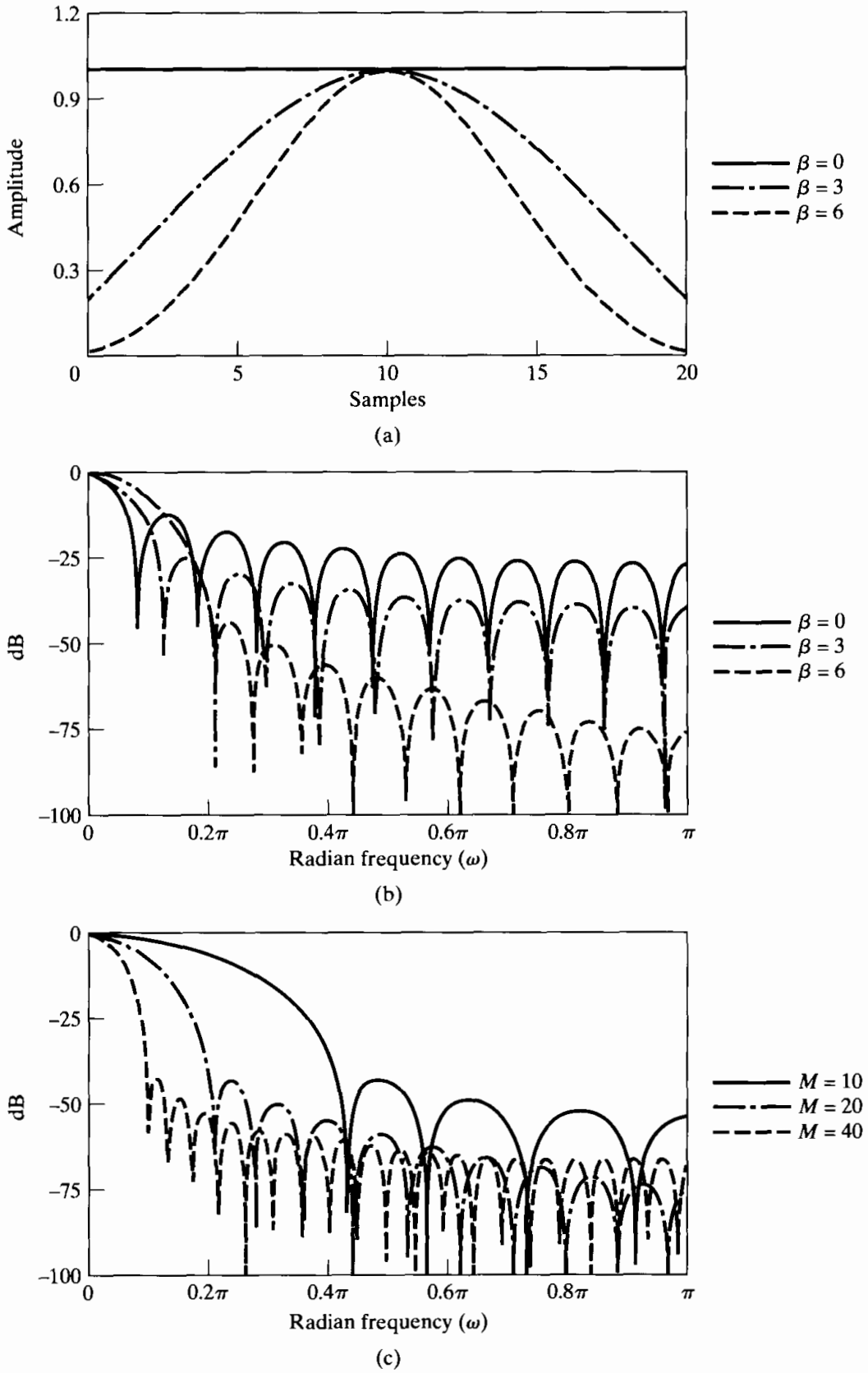


Figure 7.24 (a) Kaiser windows for $\beta = 0$, 3, and 6 and $M = 20$. (b) Fourier transforms corresponding to windows in (a). (c) Fourier transforms of Kaiser windows with $\beta = 6$ and $M = 10$, 20, and 40.

(Recall that the case $\beta = 0$ is the rectangular window for which $A = 21$.) Furthermore, Kaiser found that to achieve prescribed values of A and $\Delta\omega$, M must satisfy

$$M = \frac{A - 8}{2.285\Delta\omega}. \tag{7.63}$$

Equation (7.63) predicts M to within ± 2 over a wide range of values of $\Delta\omega$ and A. Thus, with these formulas, the Kaiser window design method requires almost no iteration or trial and error. Example 7.8 outlines and illustrates the procedure.

Example 7.8 Kaiser Window Design of a Lowpass Filter

With the use of the design formulas for the Kaiser window, it is straightforward to design an FIR lowpass filter to meet prescribed specifications. The procedure is as follows:

1. First the specifications must be established. This means selecting the desired ω_p and ω_s and the maximum tolerable approximation error. For window design, the resulting filter will have the same peak error δ in both the passband and the stopband. For this example, we use the same specifications as in Examples 7.4, 7.5, and 7.6, i.e., $\omega_p = 0.4\pi$, $\omega_s = 0.6\pi$, $\delta_1 = 0.01$, and $\delta_2 = 0.001$. Since filters designed by the window method inherently have $\delta_1 = \delta_2$, we must set $\delta = 0.001$.
2. The cutoff frequency of the underlying ideal lowpass filter must be found. Due to the symmetry of the approximation at the discontinuity of $H_d(e^{j\omega})$, we would set

$$\omega_c = \frac{\omega_p + \omega_s}{2} = 0.5\pi.$$

3. To determine the parameters of the Kaiser window, we first compute

$$\Delta\omega = \omega_s - \omega_p = 0.2\pi, \qquad A = -20\log_{10}\delta = 60.$$

We substitute these two quantities into Eqs. (7.62) and (7.63) to obtain the required values of β and M. For this example the formulas predict

$$\beta = 5.653, \qquad M = 37.$$

4. The impulse response of the filter is computed using Eqs. (7.58) and (7.59). We obtain

$$h[n] = \begin{cases} \dfrac{\sin\omega_c(n-\alpha)}{\pi(n-\alpha)} \cdot \dfrac{I_0[\beta(1-[(n-\alpha)/\alpha]^2)^{1/2}]}{I_0(\beta)}, & 0 \le n \le M, \\ 0, & \text{otherwise}, \end{cases}$$

where $\alpha = M/2 = 37/2 = 18.5$. Since $M = 37$ is an odd integer, the resulting linear-phase system would be of type II. (See Section 5.7.3 for the definitions of the four types of FIR systems with generalized linear phase.) The response characteristics of the filter are shown in Figure 7.25. Figure 7.25 (a), which shows the impulse response, displays the characteristic symmetry of a type II system. Figure 7.25(b), which shows the log magnitude response in dB, indicates that

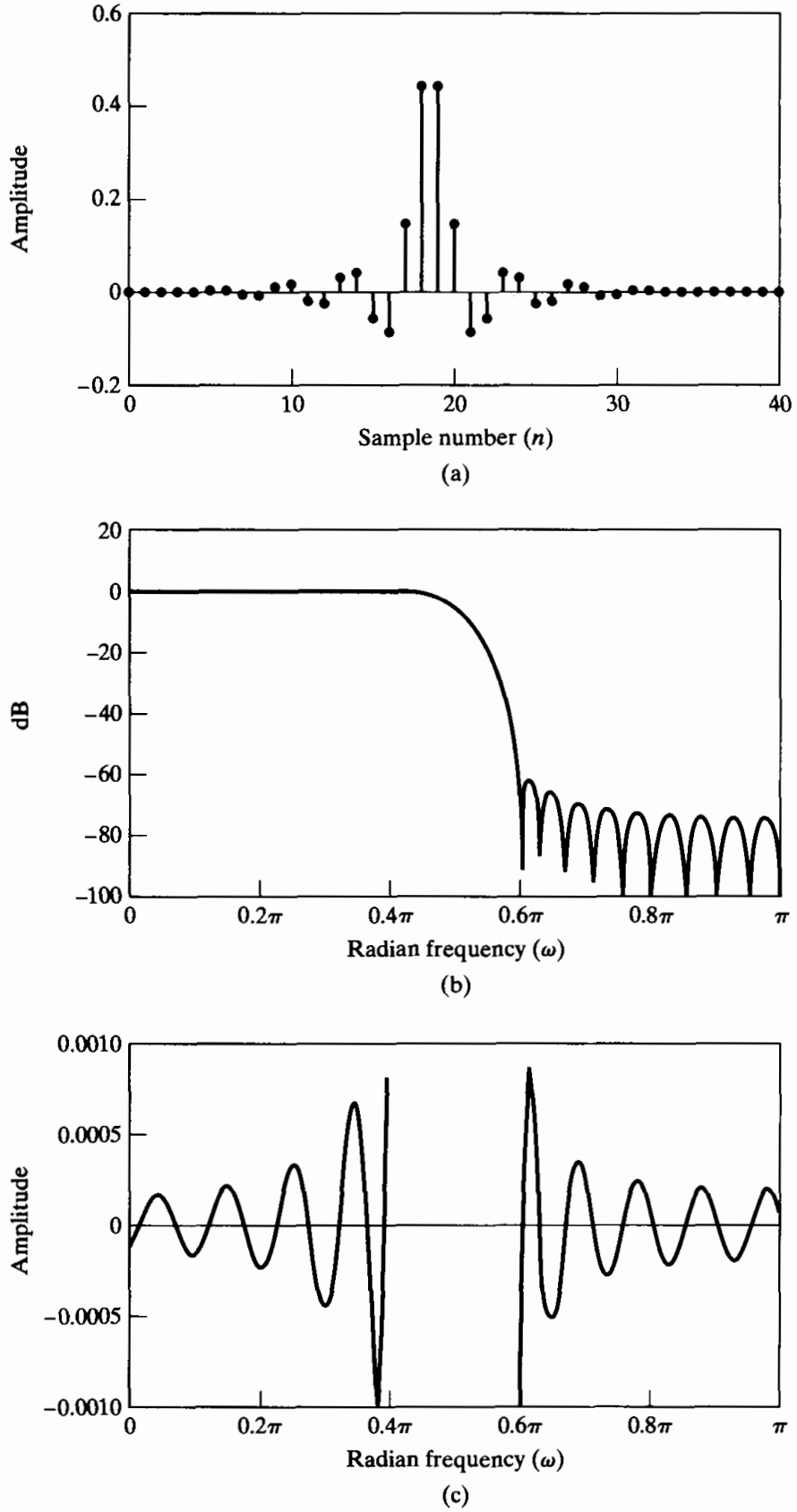


Figure 7.25 Response functions for Example 7.8. (a) Impulse response ($M = 37$). (b) Log magnitude. (c) Approximation error.

$H(e^{j\omega})$ is zero at $\omega = \pi$ or, equivalently, that $H(z)$ has a zero at $z = -1$, as required for a type II FIR system. Figure 7.25(c) shows the approximation error in the passband and stopbands. This error function is defined as

$$E_A(\omega) = \begin{cases} 1 - A_e(e^{j\omega}), & 0 \le \omega \le \omega_p, \\ 0 - A_e(e^{j\omega}), & \omega_s \le \omega \le \pi. \end{cases} \tag{7.64}$$

(The error is not defined in the transition region, $0.4\pi < \omega < 0.6\pi$.) Note the symmetry of the approximation error, and note also that the peak approximation error is slightly greater than $\delta = 0.001$. Increasing M to 38 results in a type I filter for which $\delta = 0.0008$.

Finally, observe that it is not necessary to plot either the phase or the group delay, since we know that the phase is precisely linear and the delay is $M/2 = 18.5$ samples.

7.2.4 Relationship of the Kaiser Window to Other Windows

The basic principle of the window design method is to truncate the ideal impulse response with a finite-length window such as one of those discussed in this section. The corresponding effect in the frequency domain is that the ideal frequency response is convolved with the Fourier transform of the window. If the ideal filter is a lowpass filter, the discontinuity in its frequency response is smeared as the main lobe of the Fourier transform of the window moves across the discontinuity in the convolution process. To a first approximation, the width of the resulting transition band is determined by the width of the main lobe of the Fourier transform of the window, and the passband and stopband ripples are determined by the side lobes of the Fourier transform of the window. Because the passband and stopband ripples are produced by integration of the symmetric window side lobes, the ripples in the passband and the stopband are approximately the same. Furthermore, to a very good approximation, the maximum passband and stopband deviations are not dependent on M and can be changed only by changing the shape of the window used. This is illustrated by Kaiser's formula, Eq. (7.62), for the window shape parameter, which is independent of M. The last two columns of Table 7.1 compare the Kaiser window with the windows of Eqs. (7.47). The fifth column gives the Kaiser window shape parameter that yields the same peak approximation error (δ) as the window indicated in the first column. The sixth column shows the corresponding transition width (from Eq. 7.63) for filters designed with the Kaiser window. This formula would be a much better predictor of the transition width for the other windows than would the main-lobe width given in the third column of the table.

7.3 EXAMPLES OF FIR FILTER DESIGN BY THE KAISER WINDOW METHOD

The use of the window method is, of course, not restricted to lowpass filters: Windows can be used to truncate *any* ideal impulse response to obtain a causal FIR approximation. In this section, we give several examples that illustrate the technique using the Kaiser

window. These examples also serve to point out some important properties of FIR systems.

7.3.1 Highpass Filter

The ideal highpass filter with generalized linear phase has the frequency response

$$H_{\text{hp}}(e^{j\omega}) = \begin{cases} 0, & |\omega| < \omega_c, \\ e^{-j\omega M/2}, & \omega_c < |\omega| \leq \pi. \end{cases} \tag{7.65}$$

The corresponding impulse response can be found by evaluating the inverse transform of $H_{\text{hp}}(e^{j\omega})$, or we can observe that

$$H_{\text{hp}}(e^{j\omega}) = e^{-j\omega M/2} - H_{\text{lp}}(e^{j\omega}), \tag{7.66}$$

where $H_{\text{lp}}(e^{j\omega})$ is given by Eq. (7.56). Thus, $h_{\text{hp}}[n]$ is

$$h_{\text{hp}}[n] = \frac{\sin \pi(n - M/2)}{\pi(n - M/2)} - \frac{\sin \omega_c(n - M/2)}{\pi(n - M/2)}, \qquad -\infty < n < \infty. \tag{7.67}$$

To design an FIR approximation to the highpass filter, we can proceed in a manner similar to that of Example 7.8.

Example 7.9 Kaiser Window Design of a Highpass Filter

Suppose that we wish to design a filter to meet the highpass specifications

$$|H(e^{j\omega})| \leq \delta_2, \qquad |\omega| \leq \omega_s$$

$$1 - \delta_1 \leq |H(e^{j\omega})| \leq 1 + \delta_1, \qquad \omega_p \leq |\omega| \leq \pi$$

where $\omega_s = 0.35\pi$, $\omega_p = 0.5\pi$, and $\delta_1 = \delta_2 = \delta = 0.021$. Since the ideal response also has a discontinuity, we can apply Kaiser's formulas in Eqs. (7.62) and (7.63) to estimate the required values of $\beta = 2.6$ and $M = 24$. Figure 7.26 shows the response characteristics that result when a Kaiser window with these parameters is applied to $h_{\text{hp}}[n]$ with $\omega_c = (0.35\pi + 0.5\pi)/2$. Note that, since M is an even integer, the filter is a type I FIR system with linear phase, and the delay is precisely $M/2 = 12$ samples. In this case, the actual peak approximation error is $\delta = 0.0213$ rather than 0.021, as specified. Since the error is less than 0.021 everywhere except at the stopband edge, it is tempting to simply increase M to 25, keeping β the same, thereby narrowing the transition region. This type II filter, which is shown in Figure 7.27, is highly unsatisfactory, due to the zero of $H(z)$ that is forced by the linear phase constraint to be at $z = -1$, i.e., $\omega = \pi$. Although increasing the order by 1 leads to a worse result, increasing M to 26 would, of course, lead to a type I system that would exceed the specifications. Clearly, type II FIR linear-phase systems are generally not appropriate approximations for either highpass or bandstop filters.

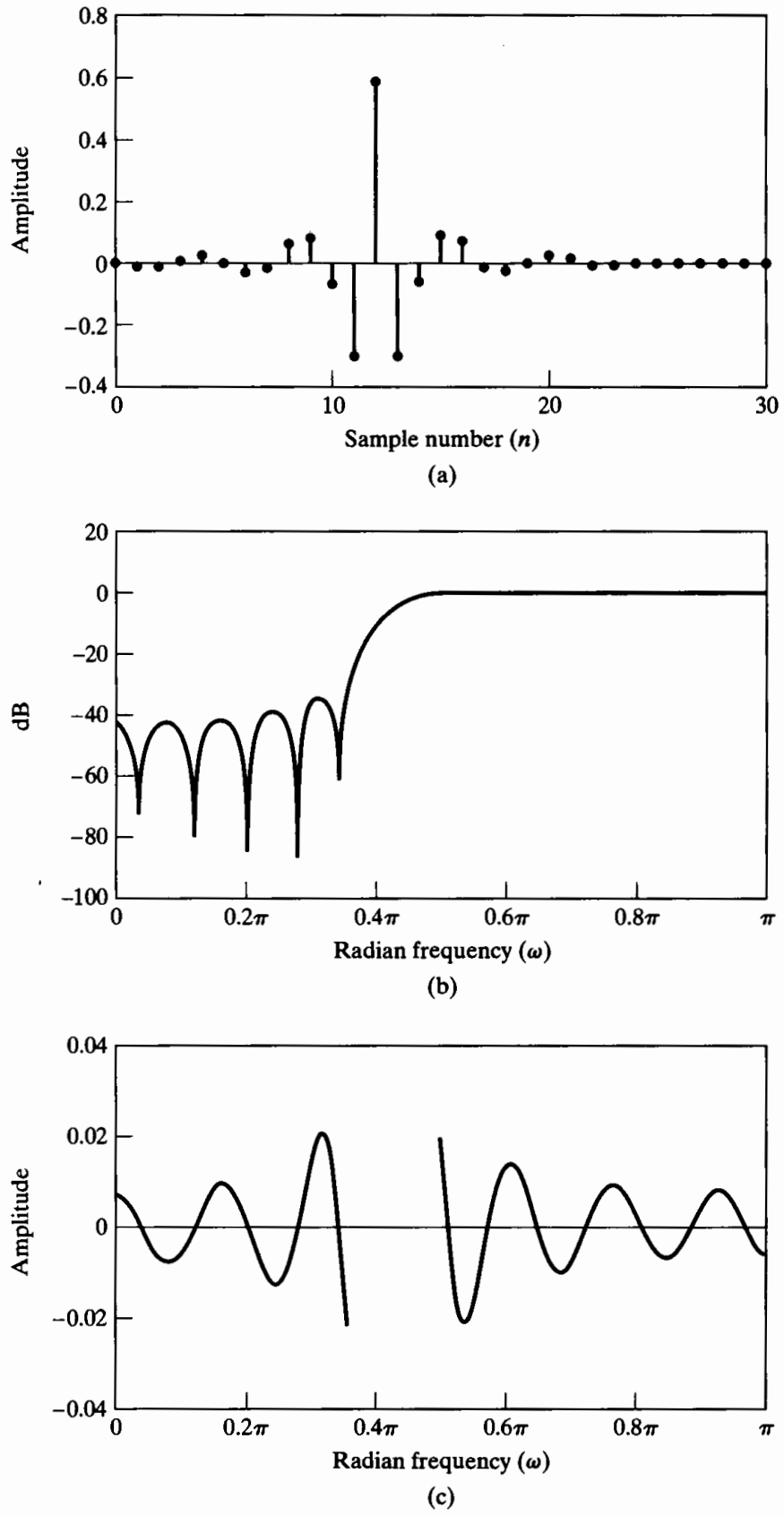


Figure 7.26 Response functions for type I FIR highpass filter. (a) Impulse response ($M = 24$). (b) Log magnitude. (c) Approximation error.

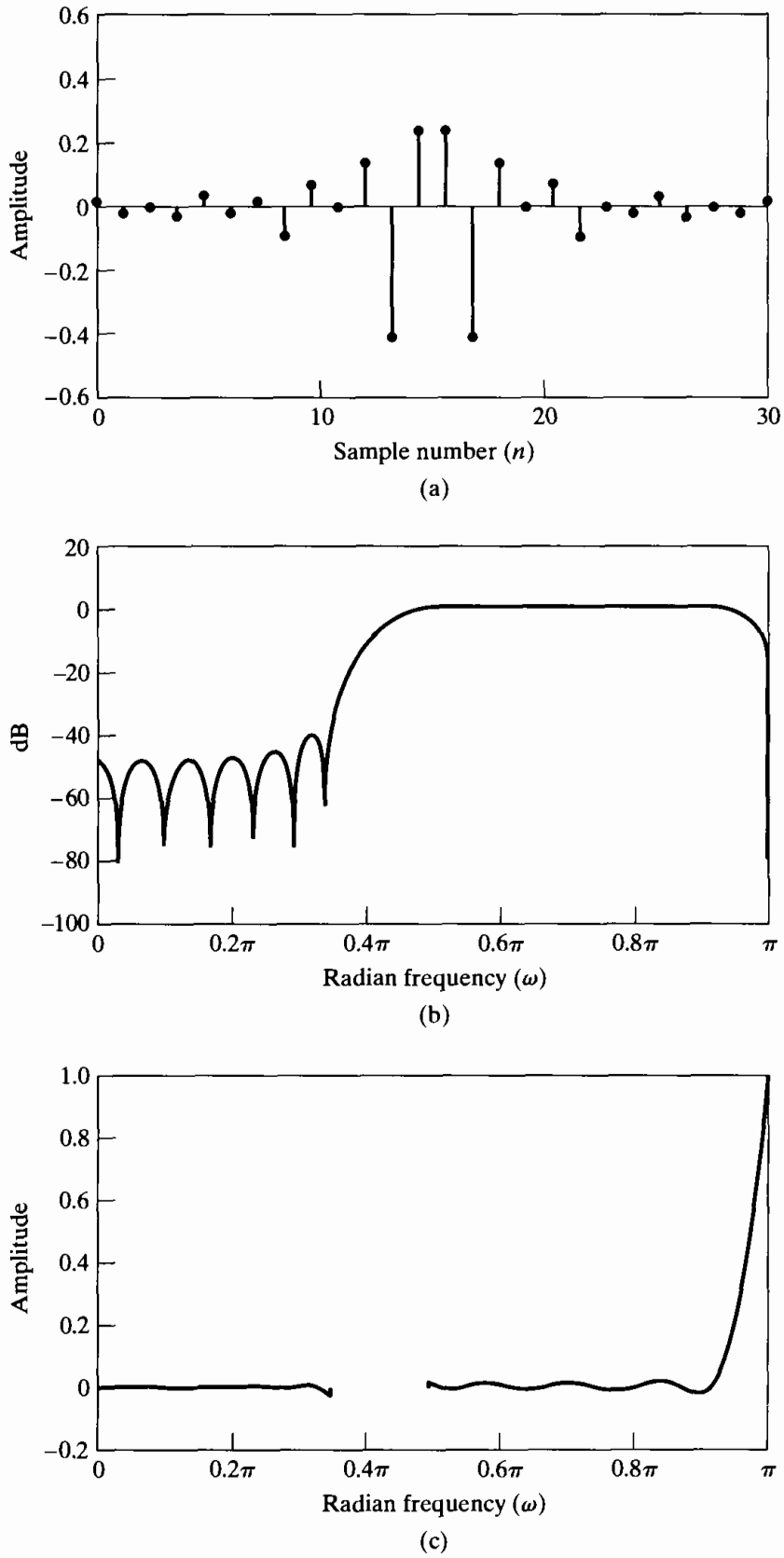


Figure 7.27 Response functions for type II FIR highpass filter. (a) Impulse response ($M = 25$). (b) Log magnitude of Fourier transform. (c) Approximation error.

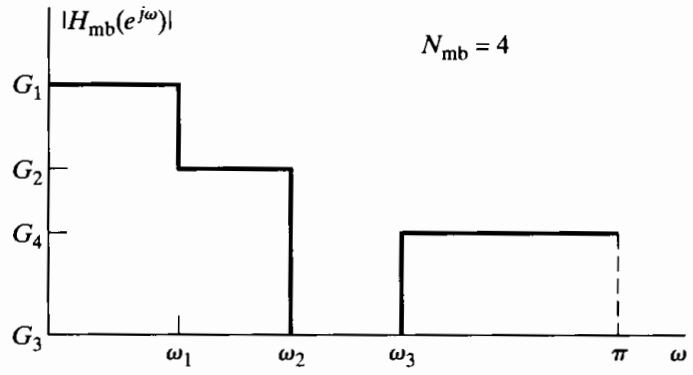


Figure 7.28 Ideal frequency response for multiband filter.

The previous discussion of highpass filter design can be generalized to the case of multiple passbands and stopbands. Figure 7.28 shows an ideal multiband frequency-selective frequency response. This generalized multiband filter includes lowpass, highpass, bandpass, and bandstop filters as special cases. If such a magnitude function is multiplied by a linear phase factor $e^{-j\omega M/2}$, the corresponding ideal impulse response is

$$h_{mb}[n] = \sum_{k=1}^{N_{mb}} (G_k - G_{k+1}) \frac{\sin \omega_k (n - M/2)}{\pi (n - M/2)}, \tag{7.68}$$

where N_{mb} is the number of bands and $G_{N_{mb}+1} = 0$. If $h_{mb}[n]$ is multiplied by a Kaiser window, the type of approximations that we have observed at the single discontinuity of the lowpass and highpass systems will occur at *each* of the discontinuities. The behavior will be the same at each discontinuity, provided that the discontinuities are far enough apart. Thus, Kaiser's formulas for the window parameters can be applied to this case to predict approximation errors and transition widths. Note that the approximation errors will be scaled by the size of the jump that produces them. That is, if a discontinuity of unity produces a peak error of δ, then a discontinuity of one-half will have a peak error of $\delta/2$.

7.3.2 Discrete-Time Differentiators

As illustrated in Example 4.5, sometimes it is of interest to obtain samples of the derivative of a bandlimited signal from samples of the signal itself. Since the Fourier transform of the derivative of a continuous-time signal is $j\Omega$ times the Fourier transform of the signal, it follows that, for bandlimited signals, a discrete-time system with frequency response $(j\omega/T)$ for $-\pi < \omega < \pi$ (and that is periodic, with period 2π) will yield output samples that are equal to samples of the derivative of the continuous-time signal. A system with this property is referred to as a discrete-time differentiator.

For an ideal discrete-time differentiator with a linear phase, the appropriate frequency response is

$$H_{diff}(e^{j\omega}) = (j\omega)e^{-j\omega M/2}, \qquad -\pi < \omega < \pi. \tag{7.69}$$