Since (5.3.24) has the form of a convolution, the frequency-domain equivalent expression is

$$\Gamma_{yx}(\omega) = H(\omega)\Gamma_{xx}(\omega) \tag{5.3.25}$$

In the special case where $x(n)$ is white noise, (5.3.25) reduces to

$$\Gamma_{yx}(\omega) = \sigma_x^2 H(\omega) \tag{5.3.26}$$

where σ_x^2 is the input noise power. This result means that an unknown system with frequency response $H(\omega)$ can be identified by exciting the input with white noise, crosscorrelating the input sequence with the output sequence to obtain $\gamma_{yx}(m)$, and finally, computing the Fourier transform of $\gamma_{yx}(m)$. The result of these computations is proportional to $H(\omega)$.

5.4 Linear Time-Invariant Systems as Frequency-Selective Filters

The term *filter* is commonly used to describe a device that discriminates, according to some attribute of the objects applied at its input, what passes through it. For example, an air filter allows air to pass through it but prevents dust particles that are present in the air from passing through. An oil filter performs a similar function, with the exception that oil is the substance allowed to pass through the filter, while particles of dirt are collected at the input to the filter and prevented from passing through. In photography, an ultraviolet filter is often used to prevent ultraviolet light, which is present in sunlight and which is not a part of visible light, from passing through and affecting the chemicals on the film.

As we have observed in the preceding section, a linear time-invariant system also performs a type of discrimination or filtering among the various frequency components at its input. The nature of this filtering action is determined by the frequency response characteristics $H(\omega)$, which in turn depends on the choice of the system parameters (e.g., the coefficients $\{a_k\}$ and $\{b_k\}$ in the difference equation characterization of the system). Thus, by proper selection of the coefficients, we can design frequency-selective filters that pass signals with frequency components in some bands while they attenuate signals containing frequency components in other frequency bands.

In general, a linear time-invariant system modifies the input signal spectrum $X(\omega)$ according to its frequency response $H(\omega)$ to yield an output signal with spectrum $Y(\omega) = H(\omega)X(\omega)$. In a sense, $H(\omega)$ acts as a *weighting function* or a *spectral shaping function* to the different frequency components in the input signal. When viewed in this context, any linear time-invariant system can be considered to be a frequency-shaping filter, even though it may not necessarily completely block any or all frequency components. Consequently, the terms "linear time-invariant system" and "filter" are synonymous and are often used interchangeably.

We use the term *filter* to describe a linear time-invariant system used to perform spectral shaping or frequency-selective filtering. Filtering is used in digital signal processing in a variety of ways, such as removal of undesirable noise from desired signals, spectral shaping such as equalization of communication channels, signal detection in radar, sonar, and communications, and for performing spectral analysis of signals, and so on.

5.4.1 Ideal Filter Characteristics

Filters are usually classified according to their frequency-domain characteristics as lowpass, highpass, bandpass, and bandstop or band-elimination filters. The ideal magnitude response characteristics of these types of filters are illustrated in Fig. 5.4.1. As shown, these ideal filters have a constant-gain (usually taken as unity-gain) passband characteristic and zero gain in their stopband.

Another characteristic of an ideal filter is a linear phase response. To demonstrate this point, let us assume that a signal sequence $\{x(n)\}$ with frequency components confined to the frequency range $\omega_1 < \omega < \omega_2$ is passed through a filter with

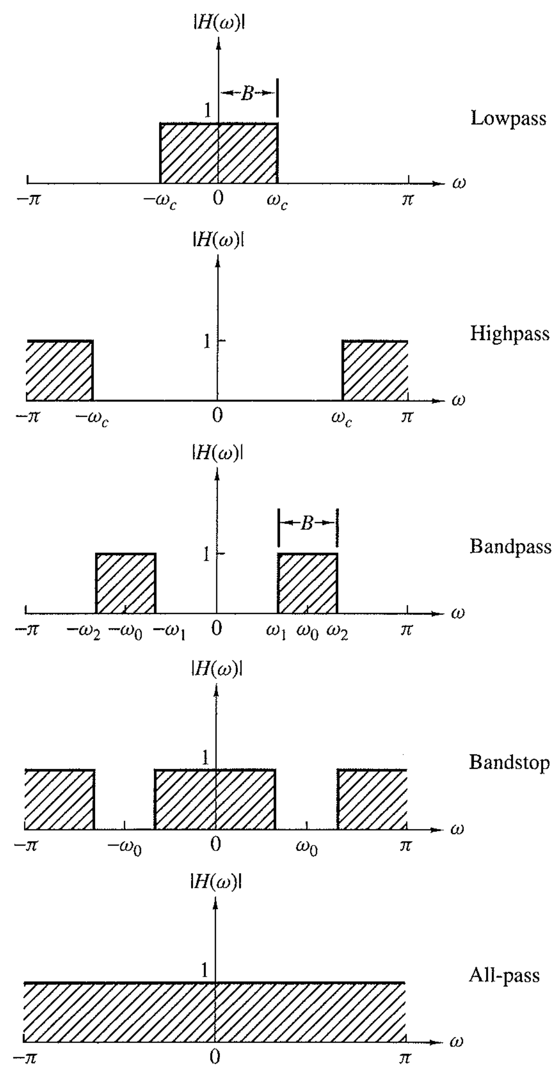


Figure 5.4.1 Magnitude responses for some ideal frequency-selective discrete-time filters.

frequency response

$$H(\omega) = \begin{cases} Ce^{-j\omega n_0}, & \omega_1 < \omega < \omega_2 \\ 0, & \text{otherwise} \end{cases} \tag{5.4.1}$$

where C and n_0 are constants. The signal at the output of the filter has a spectrum

$$\begin{aligned} Y(\omega) &= X(\omega)H(\omega) \\ &= CX(\omega)e^{-j\omega n_0}, \qquad \omega_1 < \omega < \omega_2 \end{aligned} \tag{5.4.2}$$

By applying the scaling and time-shifting properties of the Fourier transform, we obtain the time-domain output

$$y(n) = Cx(n - n_0) \tag{5.4.3}$$

Consequently, the filter output is simply a delayed and amplitude-scaled version of the input signal. A pure delay is usually tolerable and is not considered a distortionof the signal. Neither is amplitude scaling. Therefore, ideal filters have a linear phase characteristic within their passband, that is,

$$\Theta(\omega) = -\omega n_0 \tag{5.4.4}$$

The derivative of the phase with respect to frequency has the units of delay. Hence we can define the signal delay as a function of frequency as

$$\tau_g(\omega) = -\frac{d\Theta(\omega)}{d\omega} \tag{5.4.5}$$

$\tau_g(\omega)$ is usually called the *envelope delay*or the *group delay* of the filter. We interpret $\tau_g(\omega)$ as the time delay that a signal component of frequency ω undergoes as it passes from the input to the output of the system. Note that when $\Theta(\omega)$ is linear as in (5.4.4), $\tau_g(\omega) = n_0 =$ constant. In this case all frequency components of the input signal undergo the same time delay.

In conclusion, ideal filters have a constant magnitude characteristic and a linear phase characteristic within their passband. In all cases, such filters are not physically realizable but serve as a mathematical idealization of practical filters. For example, the ideal lowpass filter has an impulse response

$$h_{lp}(n) = \frac{\sin \omega_c \pi n}{\pi n}, \qquad -\infty < n < \infty \tag{5.4.6}$$

We note that this filter is not causal and it is not absolutely summable and therefore it is also unstable. Consequently, this ideal filter is physically unrealizable. Nevertheless, its frequency response characteristics can be approximated very closely by practical, physically realizable filters, as will be demonstrated in Chapter 10.

In the following discussion, we treat the design of some simple digital filters by the placement of poles and zeros in the z-plane. We have already described how the location of poles and zeros affects the frequency response characteristics of the system. In particular, in Section 5.2.2 we presented a graphical method for computing the frequency response characteristics from the pole–zero plot. This same approach can be used to design a number of simple but important digital filters with desirable frequency response characteristics.

The basic principle underlying the pole–zero placement method is to locate poles near points of the unit circle corresponding to frequencies to be emphasized, and to place zeros near the frequencies to be deemphasized. Furthermore, the following constraints must be imposed:

1. All poles should be placed inside the unit circle in order for the filter to be stable. However, zeros can be placed anywhere in the z-plane.
2. All complex zeros and poles must occur in complex-conjugate pairs in order for the filter coefficients to be real.

From our previous discussion we recall that for a given pole–zero pattern, the system function $H(z)$ can be expressed as

$$H(z) = \frac{\sum_{k=0}^{M} b_k z^{-k}}{1 + \sum_{k=1}^{N} a_k z^{-k}} = b_0 \frac{\prod_{k=1}^{M}(1 - z_k z^{-1})}{\prod_{k=1}^{N}(1 - p_k z^{-1})} \tag{5.4.7}$$

where b_0 is a gain constant selected to normalize the frequency response at some specified frequency. That is, b_0 is selected such that

$$|H(\omega_0)| = 1 \tag{5.4.8}$$

where ω_0 is a frequency in the passband of the filter. Usually, N is selected to equal or exceed M, so that the filter has more nontrivial poles than zeros.

In the next section, we illustrate the method of pole–zero placement in the design of some simple lowpass, highpass, and bandpass filters, digital resonators, and comb filters. The design procedure is facilitated when carried out interactively on a digital computer with a graphics terminal.

5.4.2 Lowpass, Highpass, and Bandpass Filters

In the design of lowpass digital filters, the poles should be placed near the unit circle at points corresponding to low frequencies (near $\omega = 0$) and zeros should be placed near or on the unit circle at points corresponding to high frequencies (near $\omega = \pi$). The opposite holds true for highpass filters.

Figure 5.4.2 illustrates the pole–zero placement of three lowpass and three highpass filters. The magnitude and phase responses for the single-pole filter with system function

$$H_1(z) = \frac{1 - a}{1 - az^{-1}} \tag{5.4.9}$$

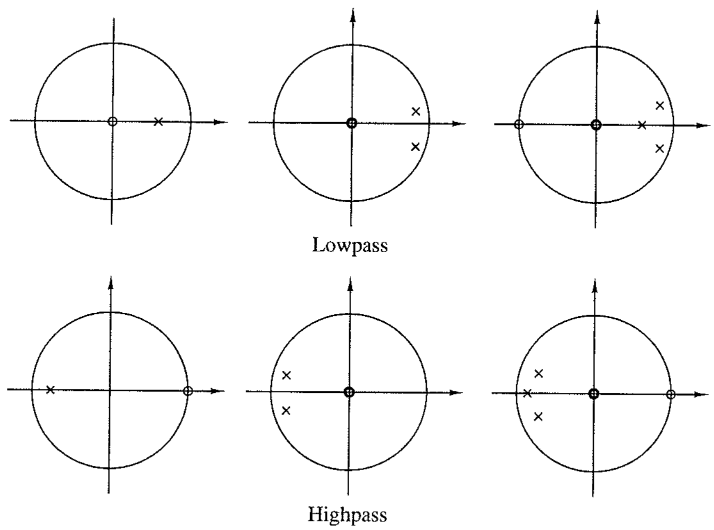


Figure 5.4.2 Pole–zero patterns for several lowpass and highpass filters.

are illustrated in Fig. 5.4.3 for $a = 0.9$. The gain G was selected as $1 - a$, so that the filter has unity gain at $\omega = 0$. The gain of this filter at high frequencies is relatively small.

The addition of a zero at $z = -1$ further attenuates the response of the filter at high frequencies. This leads to a filter with a system function

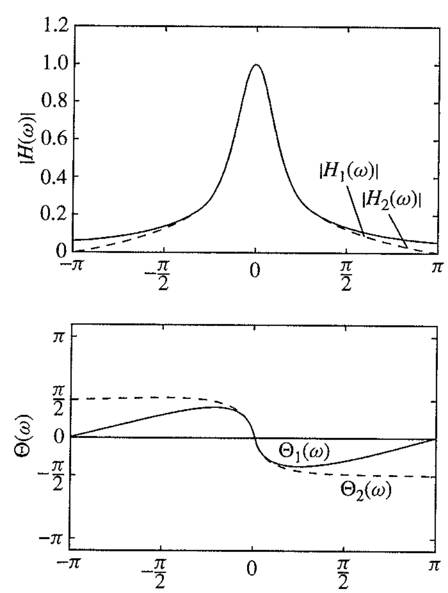


Figure 5.4.3 Magnitude and phase response of (1) a single-pole filter and (2) a one-pole, one-zero filter; $H_1(z) = (1-a)/(1-az^{-1})$, $H_2(z) = [(1-a)/2][(1+z^{-1})/(1-az^{-1})]$ and $a = 0.9$.

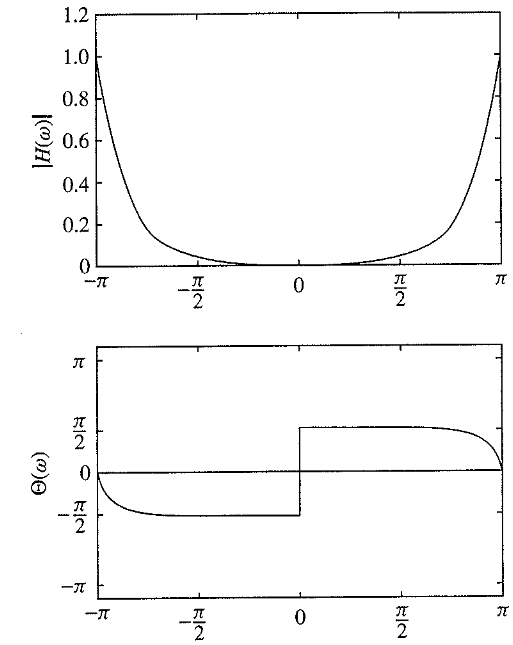


Figure 5.4.4
Magnitude and phase response of a simple highpass filter; $H(z) = [(1 - a)/2][(1 - z^{-1})/(1 + az^{-1})]$ with $a = 0.9$.

$$H_2(z) = \frac{1-a}{2}\frac{1+z^{-1}}{1-az^{-1}} \tag{5.4.10}$$

and a frequency response characteristic that is also illustrated in Fig. 5.4.3. In this case the magnitude of $H_2(\omega)$ goes to zero at $\omega = \pi$.

Similarly, we can obtain simple highpass filters by reflecting (folding) the pole–zero locations of the lowpass filters about the imaginary axis in the z-plane. Thus we obtain the system function

$$H_3(z) = \frac{1-a}{2}\frac{1-z^{-1}}{1+az^{-1}} \tag{5.4.11}$$

which has the frequency response characteristics illustrated in Fig. 5.4.4 for $a = 0.9$.

EXAMPLE 5.4.1

A two-pole lowpass filter has the system function

$$H(z) = \frac{b_0}{(1 - pz^{-1})^2}$$

Determine the values of b_0 and p such that the frequency response $H(\omega)$ satisfies the conditions

$$H(0) = 1$$

and

$$\left|H\left(\frac{\pi}{4}\right)\right|^2 = \frac{1}{2}$$

Solution. At $\omega = 0$ we have

$$H(0) = \frac{b_0}{(1-p)^2} = 1$$

Hence

$$b_0 = (1-p)^2$$

At $\omega = \pi/4$,

$$\begin{aligned} H\left(\frac{\pi}{4}\right) &= \frac{(1-p)^2}{(1-pe^{-j\pi/4})^2} \\ &= \frac{(1-p)^2}{(1-p\cos(\pi/4)+jp\sin(\pi/4))^2} \\ &= \frac{(1-p)^2}{(1-p/\sqrt{2}+jp/\sqrt{2})^2} \end{aligned}$$

Hence

$$\frac{(1-p)^4}{[(1-p/\sqrt{2})^2+p^2/2]^2} = \frac{1}{2}$$

or, equivalently,

$$\sqrt{2}(1-p)^2 = 1+p^2-\sqrt{2p}$$

The value of $p = 0.32$ satisfies this equation. Consequently, the system function for the desired filter is

$$H(z) = \frac{0.46}{(1-0.32z^{-1})^2}$$

The same principles can be applied for the design of bandpass filters. Basically, the bandpass filter should contain one or more pairs of complex-conjugate poles near the unit circle, in the vicinity of the frequency band that constitutes the passband of the filter. The following example serves to illustrate the basic ideas.

EXAMPLE 5.4.2

Design a two-pole bandpass filter that has the center of its passband at $\omega = \pi/2$, zero in its frequency response characteristic at $\omega = 0$ and $\omega = \pi$, and a magnitude response of $1/\sqrt{2}$ at $\omega = 4\pi/9$.

Solution. Clearly, the filter must have poles at

$$p_{1,2} = re^{\pm j\pi/2}$$

and zeros at $z = 1$ and $z = -1$. Consequently, the system function is

$$\begin{aligned} H(z) &= G\frac{(z-1)(z+1)}{(z-jr)(z+jr)} \\ &= G\frac{z^2-1}{z^2+r^2} \end{aligned}$$

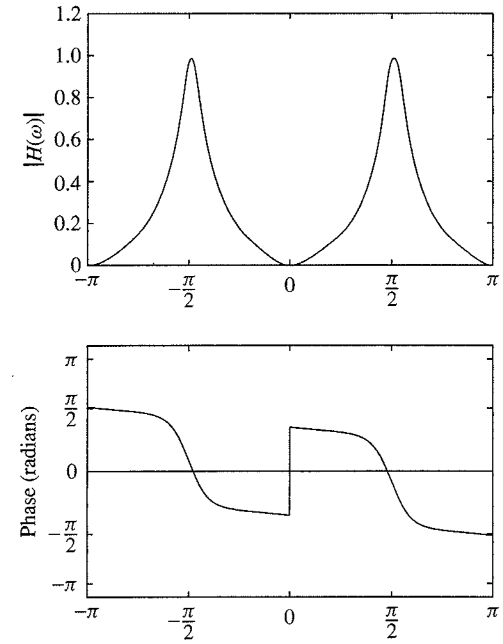


Figure 5.4.5
Magnitude and phase response of a simple bandpass filter in Example 5.4.2; $H(z) = 0.15[(1 - z^{-2})/(1 + 0.7z^{-2})]$.

The gain factor is determined by evaluating the frequency response $H(\omega)$ of the filter at $\omega = \pi/2$. Thus we have

$$H\left(\frac{\pi}{2}\right) = G\frac{2}{1 - r^2} = 1$$

$$G = \frac{1 - r^2}{2}$$

The value of r is determined by evaluating $H(\omega)$ at $\omega = 4\pi/9$. Thus we have

$$\left|H\left(\frac{4\pi}{9}\right)\right|^2 = \frac{(1 - r^2)^2}{4}\frac{2 - 2\cos(8\pi/9)}{1 + r^4 + 2r^2\cos(8\pi/9)} = \frac{1}{2}$$

or, equivalently,

$$1.94(1 - r^2)^2 = 1 - 1.88r^2 + r^4$$

The value of $r^2 = 0.7$ satisfies this equation. Therefore, the system function for the desired filter is

$$H(z) = 0.15\frac{1 - z^{-2}}{1 + 0.7z^{-2}}$$

Its frequency response is illustrated in Fig. 5.4.5.

It should be emphasized that the main purpose of the foregoing methodology for designing simple digital filters by pole–zero placement is to provide insight into the effect that poles and zeros have on the frequency response characteristic of systems. The methodology is not intended as a good method for designing digital filters with well-specified passband and stopband characteristics. Systematic methods for the design of sophisticated digital filters for practical applications are discussed in Chapter 10.

A simple lowpass-to-highpass filter transformation. Suppose that we have designed a prototype lowpass filter with impulse response $h_{\text{lp}}(n)$. By using the frequency translation property of the Fourier transform, it is possible to convert the prototype filter to either a bandpass or a highpass filter. Frequency transformations for converting a prototype lowpass filter into a filter of another type are described in detail in Section 10.3. In this section we present a simple frequency transformation for converting a lowpass filter into a highpass filter, and vice versa.

If $h_{\text{lp}}(n)$ denotes the impulse response of a lowpass filter with frequency response $H_{\text{lp}}(\omega)$, a highpass filter can be obtained by translating $H_{\text{lp}}(\omega)$ by π radians (i.e., replacing ω by $\omega - \pi$). Thus

$$H_{\text{hp}}(\omega) = H_{\text{lp}}(\omega - \pi) \tag{5.4.12}$$

where $H_{\text{hp}}(\omega)$ is the frequency response of the highpass filter. Since a frequency translation of π radians is equivalent to multiplication of the impulse response $h_{\text{lp}}(n)$ by $e^{j\pi n}$, the impulse response of the highpass filter is

$$h_{\text{hp}}(n) = (e^{j\pi})^n h_{\text{lp}}(n) = (-1)^n h_{\text{lp}}(n) \tag{5.4.13}$$

Therefore, the impulse response of the highpass filter is simply obtained from the impulse response of the lowpass filter by changing the signs of the odd-numbered samples in $h_{\text{lp}}(n)$. Conversely,

$$h_{\text{lp}}(n) = (-1)^n h_{\text{hp}}(n) \tag{5.4.14}$$

If the lowpass filter is described by the difference equation

$$y(n) = -\sum_{k=1}^{N} a_k y(n-k) + \sum_{k=0}^{M} b_k x(n-k) \tag{5.4.15}$$

its frequency response is

$$H_{\text{lp}}(\omega) = \frac{\sum_{k=0}^{M} b_k e^{-j\omega k}}{1 + \sum_{k=1}^{N} a_k e^{-j\omega k}} \tag{5.4.16}$$

Now, if we replace ω by $\omega - \pi$, in (5.4.16), then

$$H_{\text{hp}}(\omega) = \frac{\sum_{k=0}^{M} (-1)^k b_k e^{-j\omega k}}{1 + \sum_{k=1}^{N} (-1)^k a_k e^{-j\omega k}} \tag{5.4.17}$$

which corresponds to the difference equation

$$y(n) = -\sum_{k=1}^{N}(-1)^k a_k y(n-k) + \sum_{k=0}^{M}(-1)^k b_k x(n-k) \tag{5.4.18}$$

EXAMPLE 5.4.3

Convert the lowpass filter described by the difference equation

$$y(n) = 0.9y(n-1) + 0.1x(n)$$

into a highpass filter.

Solution. The difference equation for the highpass filter, according to (5.4.18), is

$$y(n) = -0.9y(n-1) + 0.1x(n)$$

and its frequency response is

$$H_{\text{hp}}(\omega) = \frac{0.1}{1 + 0.9e^{-j\omega}}$$

The reader may verify that $H_{\text{hp}}(\omega)$ is indeed highpass.

5.4.3 Digital Resonators

A *digital resonator* is a special two-pole bandpass filter with the pair of complex-conjugate poles located near the unit circle as shown in Fig. 5.4.6(a). The magnitude of the frequency response of the filter is shown in Fig. 5.4.6(b). The name resonator refers to the fact that the filter has a large magnitude response (i.e., it resonates) in the vicinity of the pole location. The angular position of the pole determines the resonant frequency of the filter. Digital resonators are useful in many applications, including simple bandpass filtering and speech generation.

In the design of a digital resonator with a resonant peak at or near $\omega = \omega_0$, we select the complex-conjugate poles at

$$p_{1,2} = re^{\pm j\omega_0}, \qquad 0 < r < 1$$

In addition, we can select up to two zeros. Although there are many possible choices, two cases are of special interest. One choice is to locate the zeros at the origin. The other choice is to locate a zero at $z = 1$ and a zero at $z = -1$. This choice completely eliminates the response of the filter at frequencies $\omega = 0$ and $\omega = \pi$, and it is useful in many practical applications.

The system function of the digital resonator with zeros at the origin is

$$H(z) = \frac{b_0}{(1 - re^{j\omega_0}z^{-1})(1 - re^{-j\omega_0}z^{-1})} \tag{5.4.19}$$

$$H(z) = \frac{b_0}{1 - (2r\cos\omega_0)z^{-1} + r^2z^{-2}} \tag{5.4.20}$$

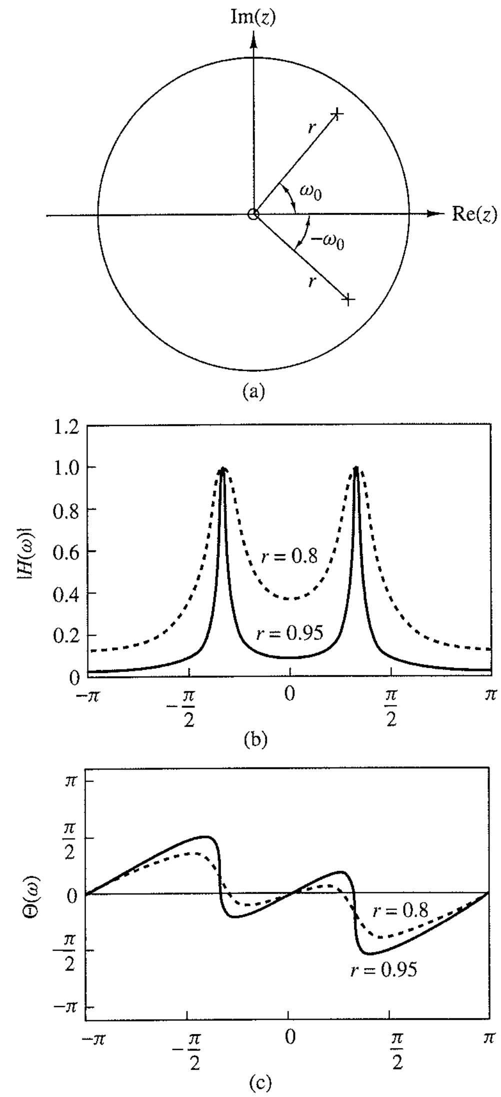


Figure 5.4.6
(a) Pole–zero pattern and (b) the corresponding magnitude and phase response of a digital resonator with (1) $r = 0.8$ and (2) $r = 0.95$.

Since $|H(\omega)|$ has its peak at or near $\omega = \omega_0$, we select the gain b_0 so that $|H(\omega_0)| = 1$. From (5.4.19) we obtain

$$\begin{aligned} H(\omega_0) &= \frac{b_0}{(1 - re^{j\omega_0}e^{-j\omega_0})(1 - re^{-j\omega_0}e^{-j\omega_0})} \\ &= \frac{b_0}{(1-r)(1 - re^{-j2\omega_0})} \end{aligned} \tag{5.4.21}$$

and hence

$$|H(\omega_0)| = \frac{b_0}{(1-r)\sqrt{1 + r^2 - 2r\cos 2\omega_0}} = 1$$

Thus the desired normalization factor is

$$b_0 = (1 - r)\sqrt{1 + r^2 - 2r\cos 2\omega_0} \tag{5.4.22}$$

The frequency response of the resonator in (5.4.19) can be expressed as

$$\begin{aligned} |H(\omega)| &= \frac{b_0}{U_1(\omega)U_2(\omega)} \\ \Theta(\omega) &= 2\omega - \Phi_1(\omega) - \Phi_2(\omega) \end{aligned} \tag{5.4.23}$$

where $U_1(\omega)$ and $U_2(\omega)$ are the magnitudes of the vectors from p_1 and p_2 to the point ω in the unit circle and $\Phi_1(\omega)$ and $\Phi_2(\omega)$ are the corresponding angles of these two vectors. The magnitudes $U_1(\omega)$ and $U_2(\omega)$ may be expressed as

$$\begin{aligned} U_1(\omega) &= \sqrt{1 + r^2 - 2r\cos(\omega_0 - \omega)} \\ U_2(\omega) &= \sqrt{1 + r^2 - 2r\cos(\omega_0 + \omega)} \end{aligned} \tag{5.4.24}$$

For any value of r, $U_1(\omega)$ takes its minimum value $(1 - r)$ at $\omega = \omega_0$. The product $U_1(\omega)U_2(\omega)$ reaches a minimum value at the frequency

$$\omega_r = \cos^{-1}\left(\frac{1 + r^2}{2r}\cos\omega_0\right) \tag{5.4.25}$$

which defines precisely the resonant frequency of the filter. We observe that when r is very close to unity, $\omega_r \approx \omega_0$, which is the angular position of the pole. We also observe that as r approaches unity, the resonance peak becomes sharper because $U_1(\omega)$ changes more rapidly in relative size in the vicinity of ω_0. A quantitative measure of the sharpness of the resonance is provided by the 3-dB bandwidth $\Delta\omega$ of the filter. For values of r close to unity,

$$\Delta\omega \approx 2(1 - r) \tag{5.4.26}$$

Figure 5.4.6 illustrates the magnitude and phase of digital resonators with $\omega_0 = \pi/3$, $r = 0.8$ and $\omega_0 = \pi/3$, $r = 0.95$. We note that the phase response undergoes its greatest rate of change near the resonant frequency.

If the zeros of the digital resonator are placed at $z = 1$ and $z = -1$, the resonator has the system function

$$\begin{aligned} H(z) &= G\frac{(1 - z^{-1})(1 + z^{-1})}{(1 - re^{j\omega_0}z^{-1})(1 - re^{-j\omega_0}z^{-1})} \\ &= G\frac{1 - z^{-2}}{1 - (2r\cos\omega_0)z^{-1} + r^2z^{-2}} \end{aligned} \tag{5.4.27}$$

and a frequency response characteristic

$$H(\omega) = b_0 \frac{1 - e^{-j2\omega}}{[1 - re^{j(\omega_0 - \omega)}][1 - re^{-j(\omega_0 + \omega)}]} \tag{5.4.28}$$

We observe that the zeros at $z = \pm 1$ affect both the magnitude and phase response of the resonator. For example, the magnitude response is

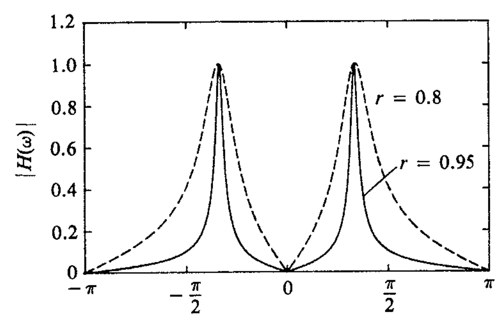
$$|H(\omega)| = b_0 \frac{N(\omega)}{U_1(\omega) U_2(\omega)} \tag{5.4.29}$$

where $N(\omega)$ is defined as

$$N(\omega) = \sqrt{2(1 - \cos 2\omega)}$$

Due to the presence of the zero factor, the resonant frequency is altered from that given by the expression in (5.4.25). The bandwidth of the filter is also altered. Although exact values for these two parameters are rather tedious to derive, we can easily compute the frequency response in (5.4.28) and compare the result with the previous case in which the zeros are located at the origin.

Figure 5.4.7 illustrates the magnitude and phase characteristics for $\omega_0 = \pi/3$, $r = 0.8$ and $\omega_0 = \pi/3$, $r = 0.95$. We observe that this filter has a slightly smaller bandwidth than the resonator, which has zeros at the origin. In addition, there appears to be a very small shift in the resonant frequency due to the presence of the zeros.



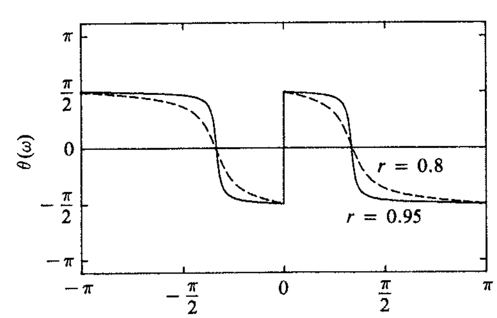


Figure 5.4.7 Magnitude and phase response of digital resonator with zeros at $\omega = 0$ and $\omega = \pi$ and (1) $r = 0.8$ and (2) $r = 0.95$.

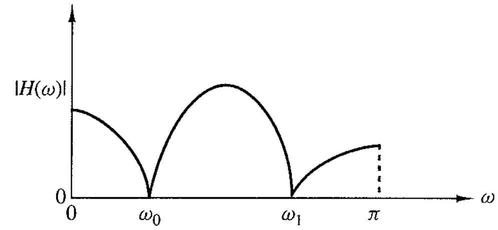


Figure 5.4.8
Frequency response characteristic of a notch filter.

5.4.4 Notch Filters

A notch filter is a filter that contains one or more deep notches or, ideally, perfect nulls in its frequency response characteristic. Figure 5.4.8 illustrates the frequency response characteristic of a notch filter with nulls at frequencies ω_0 and ω_1. Notch filters are useful in many applications where specific frequency components must be eliminated. For example, instrumentation and recording systems require that the power-line frequency of 60 Hz and its harmonics be eliminated.

To create a null in the frequency response of a filter at a frequency ω_0, we simply introduce a pair of complex-conjugate zeros on the unit circle at an angle ω_0. That is,

$$z_{1,2} = e^{\pm j\omega_0}$$

Thus the system function for an FIR notch filter is simply

$$\begin{aligned} H(z) &= b_0(1 - e^{j\omega_0}z^{-1})(1 - e^{-j\omega_0}z^{-1}) \\ &= b_0(1 - 2\cos\omega_0 z^{-1} + z^{-2}) \end{aligned} \tag{5.4.30}$$

As an illustration, Fig. 5.4.9 shows the magnitude response for a notch filter having a null at $\omega = \pi/4$.

The problem with the FIR notch filter is that the notch has a relatively large bandwidth, which means that other frequency components around the desired null are severely attenuated. To reduce the bandwidth of the null, we can resort to a more sophisticated, longer FIR filter designed according to criteria described in Chapter 10. Alternatively, we could, in an ad hoc manner, attempt to improve on the frequency response characteristics by introducing poles in the system function.

Suppose that we place a pair of complex-conjugate poles at

$$p_{1,2} = re^{\pm j\omega_0}$$

The effect of the poles is to introduce a resonance in the vicinity of the null and thus to reduce the bandwidth of the notch. The system function for the resulting filter is

$$H(z) = b_0 \frac{1 - 2\cos\omega_0 z^{-1} + z^{-2}}{1 - 2r\cos\omega_0 z^{-1} + r^2 z^{-2}} \tag{5.4.31}$$

The magnitude response $|H(\omega)|$ of the filter in (5.4.31) is plotted in Fig. 5.4.10 for $\omega_0 = \pi/4$, $r = 0.85$, and for $\omega_0 = \pi/4$, $r = 0.95$. When compared with the frequency response of the FIR filter in Fig. 5.4.9, we note that the effect of the poles is to reduce the bandwidth of the notch.

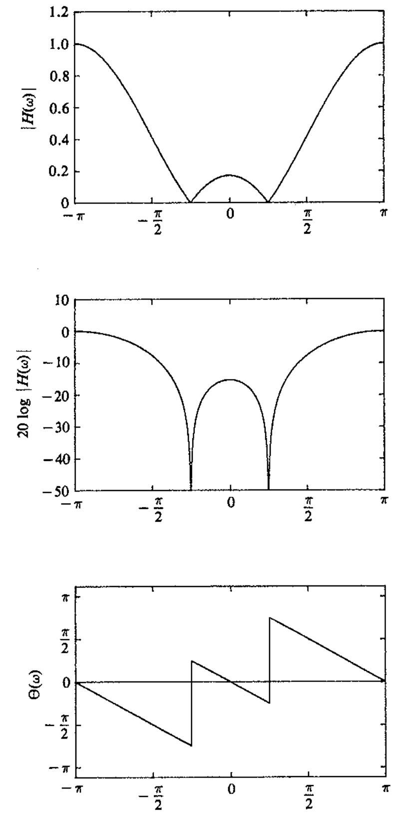


Figure 5.4.9
Frequency response characteristics of a notch filter with a notch at $\omega = \pi/4$ or $f = 1/8$; $H(z) = G[1 - 2\cos\omega_0 z^{-1} + z^{-2}]$.

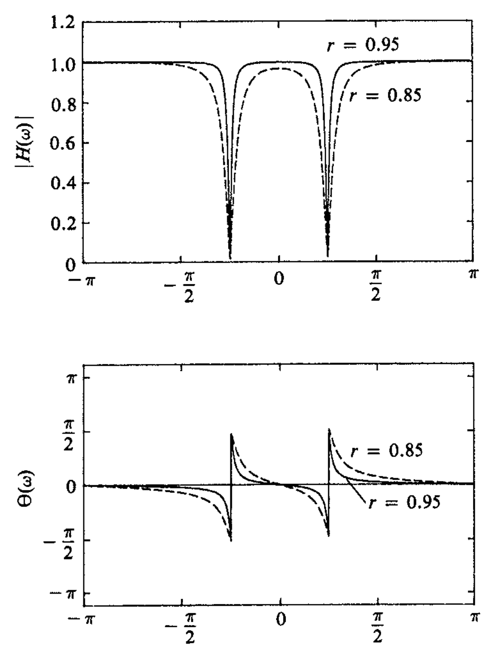


Figure 5.4.10
Frequency response characteristics of two notch filters with poles at (1) $r = 0.85$ and (2) $r = 0.95$; $H(z) = b_0[(1-2\cos\omega_0 z^{-1}+z^{-2})/(1-2r\cos\omega_0 z^{-1} + r^2 z^{-2})]$.

In addition to reducing the bandwidth of the notch, the introduction of a pole in the vicinity of the null may result in a small ripple in the passband of the filter due to the resonance created by the pole. The effect of the ripple can be reduced by introducing additional poles and/or zeros in the system function of the notch filter. The major problem with this approach is that it is basically an ad hoc, trial-and-error method.

5.4.5 Comb Filters

In its simplest form, a comb filter can be viewed as a notch filter in which the nulls occur periodically across the frequency band, hence the analogy to an ordinary comb that has periodically spaced teeth. Comb filters find applications in a wide range of practical systems such as in the rejection of power-line harmonics, in the separation of solar and lunar components from ionospheric measurements of electron concentration, and in the suppression of clutter from fixed objects in moving-target-indicator (MTI) radars.

To illustrate a simple form of a comb filter, consider a moving average (FIR) filter described by the difference equation

$$y(n) = \frac{1}{M+1}\sum_{k=0}^{M} x(n-k) \tag{5.4.32}$$

The system function of this FIR filter is

$$\begin{aligned} H(z) &= \frac{1}{M+1}\sum_{k=0}^{M} z^{-k} \\ &= \frac{1}{M+1}\frac{[1-z^{-(M+1)}]}{(1-z^{-1})} \end{aligned} \tag{5.4.33}$$

and its frequency response is

$$H(\omega) = \frac{e^{-j\omega M/2}}{M+1}\frac{\sin \omega(\frac{M+1}{2})}{\sin(\omega/2)} \tag{5.4.34}$$

From (5.4.33) we observe that the filter has zeros on the unit circle at

$$z = e^{j2\pi k/(M+1)}, \qquad k = 1, 2, 3, \ldots, M \tag{5.4.35}$$

Note that the pole at $z = 1$ is actually canceled by the zero at $z = 1$, so that in effect the FIR filter does not contain poles outside $z = 0$.

A plot of the magnitude characteristic of (5.4.34) clearly illustrates the existence of the periodically spaced zeros in frequency at $\omega_k = 2\pi k/(M+1)$ for $k = 1, 2, \ldots, M$. Figure 5.4.11 shows $|H(\omega)|$ for $M = 10$.

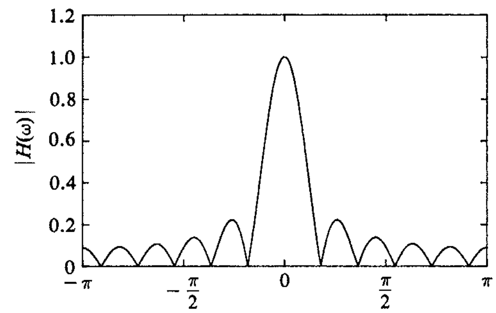


Figure 5.4.11 Magnitude response characteristic for the comb filter given by (5.4.34) with $M = 10$.

In more general terms, we can create a comb filter by taking an FIR filter with system function

$$H(z) = \sum_{k=0}^{M} h(k) z^{-k} \tag{5.4.36}$$

and replacing z by z^L, where L is a positive integer. Thus the new FIR filter has a system function

$$H_L(z) = \sum_{k=0}^{M} h(k) z^{-kL} \tag{5.4.37}$$

If the frequency response of the original FIR filter is $H(\omega)$, the frequency response of the FIR in (5.4.37) is

$$\begin{aligned} H_L(\omega) &= \sum_{k=0}^{M} h(k) e^{-jkL\omega} \\ &= H(L\omega) \end{aligned} \tag{5.4.38}$$

Consequently, the frequency response characteristic $H_L(\omega)$ is simply an L-order repetition of $H(\omega)$ in the range $0 \le \omega \le 2\pi$. Figure 5.4.12 illustrates the relationship between $H_L(\omega)$ and $H(\omega)$ for $L = 5$.

Now, suppose that the original FIR filter with system function $H(z)$ has a spectral null (i.e., a zero), at some frequency ω_0. Then the filter with system function $H_L(z)$ has periodically spaced nulls at $\omega_k = \omega_0 + 2\pi k/L$, $k = 0, 1, 2, \ldots, L-1$. As an illustration, Fig. 5.4.13 shows an FIR comb filter with $M = 3$ and $L = 3$. This FIR filter can be viewed as an FIR filter of length 10, but only four of the 10 filter coefficients are nonzero.

Let us now return to the moving average filter with system function given by (5.4.33). Suppose that we replace z by z^L. Then the resulting comb filter has the system function

$$H_L(z) = \frac{1}{M+1} \frac{1 - z^{-L(M+1)}}{1 - z^{-L}} \tag{5.4.39}$$

and a frequency response

$$H_L(\omega) = \frac{1}{M+1} \frac{\sin[\omega L(M+1)/2]}{\sin(\omega L/2)} e^{-j\omega LM/2} \tag{5.4.40}$$

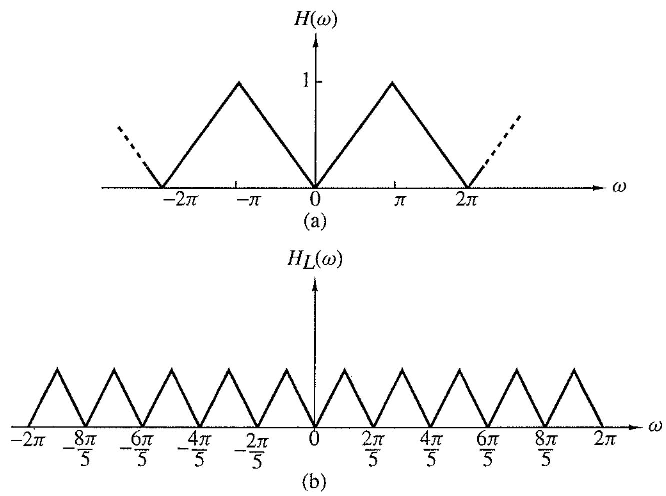


Figure 5.4.12 Comb filter with frequency response $H_L(\omega)$ obtained from $H(\omega)$.

This filter has zeros on the unit circle at

$$z_k = e^{j2\pi k/L(M+1)} \tag{5.4.41}$$

for all integer values of k except $k = 0, L, 2L, \ldots, ML$. Figure 5.4.14 illustrates $|H_L(\omega)|$ for $L = 3$ and $M = 10$.

The comb filter described by (5.4.39) finds application in the separation of solar and lunar spectral components in ionospheric measurements of electron concentration as described in the paper by Bernhardt et al. (1976). The solar period is $T_s = 24$ hours and results in a solar component of one cycle per day and its harmonics. The lunar period is $T_L = 24.84$ hours and provides spectral lines at 0.96618 cycle per day and its harmonics. Figure 5.4.15(a) shows a plot of the power density spectrum of the unfiltered ionospheric measurements of the electron concentration. Note that the weak lunar spectral components are almost hidden by the strong solar spectral components.

The two sets of spectral components can be separated by the use of comb filters. If we wish to obtain the solar components, we can use a comb filter with a narrow passband at multiples of one cycle per day. This can be achieved by selecting L such that $F_s/L = 1$ cycle per day, where F_s is the corresponding sampling frequency. The

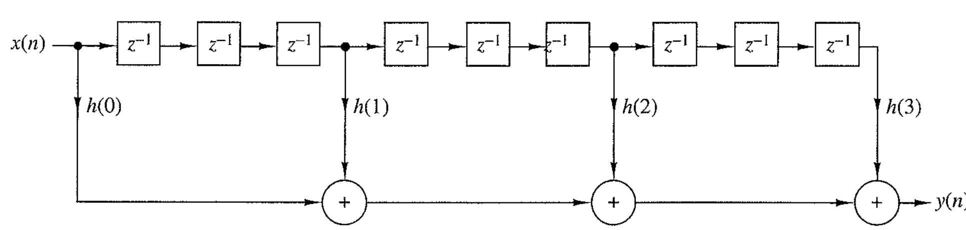


Figure 5.4.13 Realization of an FIR comb filter having $M = 3$ and $L = 3$.

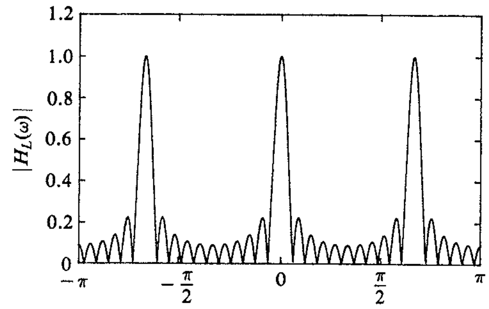


Figure 5.4.14 Magnitude response characteristic for a comb filter given by (5.4.40), with $L = 3$ and $M = 10$.

result is a filter that has peaks in its frequency response at multiples of one cycle per day. By selecting $M = 58$, the filter will have nulls at multiples of $(F_s/L)/(M+1) = 1/59$ cycle per day. These nulls are very close to the lunar components and result in good rejection. Figure 5.4.15(b) illustrates the power spectral density of the output of the comb filter that isolates the solar components. A comb filter that rejects the solar components and passes the lunar components can be designed in a similar manner. Figure 5.4.15(c) illustrates the power spectral density at the output of such a lunar filter.

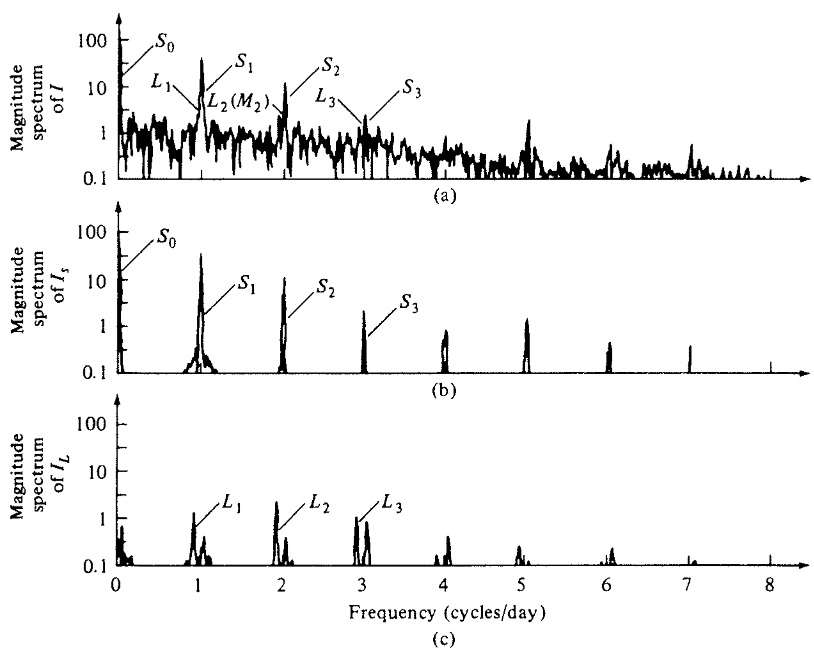


Figure 5.4.15 (a) Spectrum of unfiltered electron content data; (b) spectrum of output of solar filter; (c) spectrum of output of lunar filter. [From paper by Bernhardt et al. (1976). Reprinted with permission of the American Geophysical Union.]

5.4.6 All-Pass Filters

An all-pass filter is defined as a system that has a constant magnitude response for all frequencies, that is,

$$|H(\omega)| = 1, \qquad 0 \le \omega \le \pi \tag{5.4.42}$$

The simplest example of an all-pass filter is a pure delay system with system function

$$H(z) = z^{-k}$$

This system passes all signals without modification except for a delay of k samples. This is a trivial all-pass system that has a linear phase response characteristic.

A more interesting all-pass filter is described by the system function

$$\begin{aligned} H(z) &= \frac{a_N + a_{N-1}z^{-1} + \cdots + a_1 z^{-N+1} + z^{-N}}{1 + a_1 z^{-1} + \cdots + a_N z^{-N}} \\ &= \frac{\sum_{k=0}^{N} a_k z^{-N+k}}{\sum_{k=0}^{N} a_k z^{-k}}, \qquad a_0 = 1 \end{aligned} \tag{5.4.43}$$

where all the filter coefficients $\{a_k\}$ are real. If we define the polynomial $A(z)$ as

$$A(z) = \sum_{k=0}^{N} a_k z^{-k}, \qquad a_0 = 1$$

then (5.4.43) can be expressed as

$$H(z) = z^{-N}\frac{A(z^{-1})}{A(z)} \tag{5.4.44}$$

Since

$$|H(\omega)|^2 = H(z)H(z^{-1})|_{z=e^{j\omega}} = 1$$

the system given by (5.4.44) is an all-pass system. Furthermore, if z_0 is a pole of $H(z)$, then $1/z_0$ is a zero of $H(z)$ (i.e., the poles and zeros are reciprocals of one another). Figure 5.4.16 illustrates typical pole–zero patterns for a single-pole, single-zero filter and a two-pole, two-zero filter. A plot of the phase characteristics of these filters is shown in Fig. 5.4.17 for $a = 0.6$ and $r = 0.9$, $\omega_0 = \pi/4$.

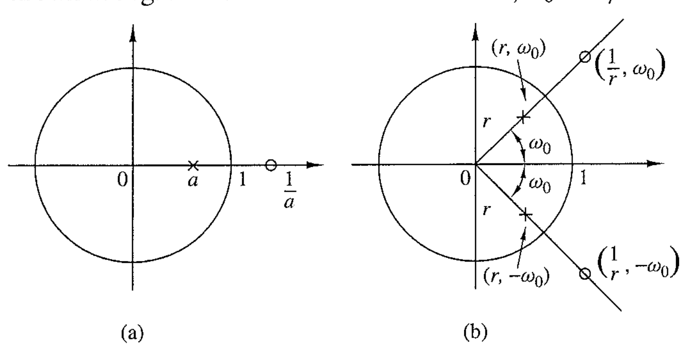


Figure 5.4.16 Pole–zero patterns of (a) a first-order and (b) a second-order all-pass filter.

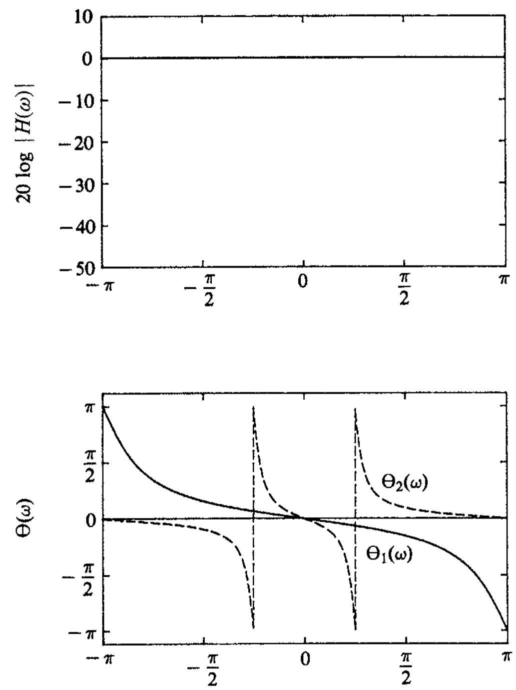


Figure 5.4.17 Frequency response characteristics of an all-pass filter with system functions (1) $H(z) = (0.6 + z^{-1})/(1 + 0.6z^{-1})$, (2) $H(z) = (r^2 - 2r\cos\omega_0 z^{-1} + z^{-2})/(1 - 2r\cos\omega_0 z^{-1} + r^2 z^{-2})$, $r = 0.9$, $\omega_0 = \pi/4$.

The most general form for the system function of an all-pass system with real coefficients, expressed in factored form in terms of poles and zeros, is

$$H_{\text{ap}}(z) = \prod_{k=1}^{N_R} \frac{z^{-1} - \alpha_k}{1 - \alpha_k z^{-1}} \prod_{k=1}^{N_c} \frac{(z^{-1} - \beta_k)(z^{-1} - \beta_k^*)}{(1 - \beta_k z^{-1})(1 - \beta_k^* z^{-1})} \tag{5.4.45}$$

where there are N_R real poles and zeros and N_c complex-conjugate pairs of poles and zeros. For causal and stable systems we require that $-1 < \alpha_k < 1$ and $|\beta_k| < 1$.

Expressions for the phase response and group delay of all-pass systems can easily be obtained using the method described in Section 5.2.1. For a single pole–single zero all-pass system we have

$$H_{\text{ap}}(\omega) = \frac{e^{j\omega} - re^{-j\theta}}{1 - re^{j\theta}e^{-j\omega}}$$

Hence

$$\Theta_{\text{ap}}(\omega) = -\omega - 2\tan^{-1}\frac{r\sin(\omega - \theta)}{1 - r\cos(\omega - \theta)}$$

and

$$\tau_g(\omega) = -\frac{d\Theta_{\text{ap}}(\omega)}{d\omega} = \frac{1 - r^2}{1 + r^2 - 2r\cos(\omega - \theta)} \tag{5.4.46}$$

We note that for a causal and stable system, $r < 1$ and hence $\tau_g(\omega) \geq 0$. Since the group delay of a higher-order pole–zero system consists of a sum of positive terms as in (5.4.46), the group delay will always be positive.

All-pass filters find application as phase equalizers. When placed in cascade with a system that has an undesired phase response, a phase equalizer is designed to compensate for the poor phase characteristics of the system and therefore to produce an overall linear-phase response.

5.4.7 Digital Sinusoidal Oscillators

A *digital sinusoidal oscillator*can be viewed as a limiting form of a two-pole resonator for which the complex-conjugate poles lie on the unit circle. From our previous discussion of second-order systems, we recall that a system with system function

$$H(z) = \frac{b_0}{1 + a_1 z^{-1} + a_2 z^{-2}} \tag{5.4.47}$$

and parameters

$$a_1 = -2r\cos\omega_0 \quad \text{and} \quad a_2 = r^2 \tag{5.4.48}$$

has complex-conjugate poles at $p = re^{\pm j\omega_0}$, and a unit sample response

$$h(n) = \frac{b_0 r^n}{\sin\omega_0}\sin(n+1)\omega_0 u(n) \tag{5.4.49}$$

If the poles are placed on the unit circle ($r = 1$) and b_0 is set to $A\sin\omega_0$, then

$$h(n) = A\sin(n+1)\omega_0 u(n) \tag{5.4.50}$$

Thus the impulse response of the second-order system with complex-conjugate poles on the unit circle is a sinusoid and the system is called a digital sinusoidal oscillator or a *digital sinusoidal generator.*

A digital sinusoidal generator is a basic component of a digital frequency synthesizer.

The block diagram representation of the system function given by (5.4.47) is illustrated in Fig. 5.4.18. The corresponding difference equation for this system is

$$y(n) = -a_1 y(n-1) - y(n-2) + b_0\delta(n) \tag{5.4.51}$$

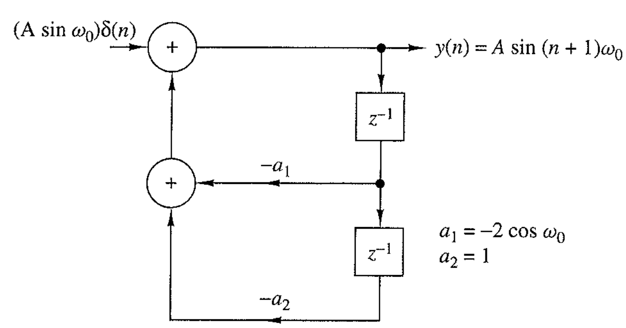


Figure 5.4.18
Digital sinusoidal generator.

where the parameters are $a_1 = -2\cos\omega_0$ and $b_0 = A\sin\omega_0$, and the initial conditions are $y(-1) = y(-2) = 0$. By iterating the difference equation in (5.4.51), we obtain

$$
\begin{aligned}
y(0) &= A\sin\omega_0 \\
y(1) &= 2\cos\omega_0 y(0) = 2A\sin\omega_0\cos\omega_0 = A\sin 2\omega_0 \\
y(2) &= 2\cos\omega_0 y(1) - y(0) \\
&= 2A\cos\omega_0\sin 2\omega_0 - A\sin\omega_0 \\
&= A(4\cos^2\omega_0 - 1)\sin\omega_0 \\
&= 3A\sin\omega_0 - 4\sin^3\omega_0 = A\sin 3\omega_0
\end{aligned}
$$

and so forth. We note that the application of the impulse at $n = 0$ serves the purpose of beginning the sinusoidal oscillation. Thereafter, the oscillation is self-sustaining because the system has no damping (i.e., $r = 1$).

It is interesting to note that the sinusoidal oscillation obtained from the system in (5.4.51) can also be obtained by setting the input to zero and setting the initial conditions to $y(-1) = 0$, $y(-2) = -A\sin\omega_0$. Thus the zero-input response to the second-order system described by the homogeneous difference equation

$$
y(n) = -a_1 y(n-1) - y(n-2) \tag{5.4.52}
$$

with initial conditions $y(-1) = 0$ and $y(-2) = -A\sin\omega_0$, is exactly the same as the response of (5.4.51) to an impulse excitation. In fact, the difference equation in (5.4.52) can be obtained directly from the trigonometric identity

$$
\sin\alpha + \sin\beta = 2\sin\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2} \tag{5.4.53}
$$

where, by definition, $\alpha = (n+1)\omega_0$, $\beta = (n-1)\omega_0$, and $y(n) = \sin(n+1)\omega_0$.

In some practical applications involving modulation of two sinusoidal carrier signals in phase quadrature, there is a need to generate the sinusoids $A\sin\omega_0 n$ and $A\cos\omega_0 n$. These signals can be generated from the so-called *coupled-form oscillator,* which can be obtained from the trigonometric formulas

$$
\begin{aligned}
\cos(\alpha+\beta) &= \cos\alpha\cos\beta - \sin\alpha\sin\beta \\
\sin(\alpha+\beta) &= \sin\alpha\cos\beta + \cos\alpha\sin\beta
\end{aligned}
$$

where, by definition, $\alpha = n\omega_0$, $\beta = \omega_0$, and

$$
y_c(n) = \cos n\omega_0 u(n) \tag{5.4.54}
$$

$$
y_s(n) = \sin n\omega_0 u(n) \tag{5.4.55}
$$

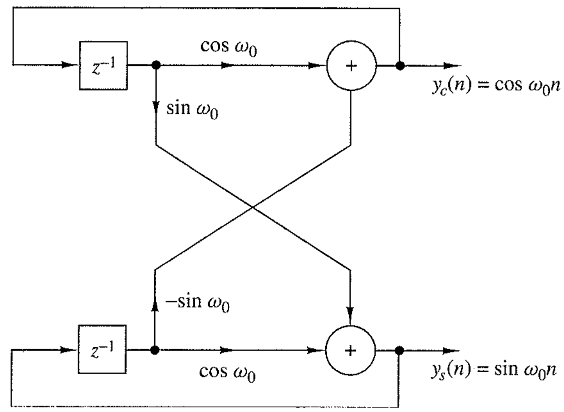


Figure 5.4.19 Realization of the coupled-form oscillator.

Thus we obtain the two coupled difference equations

$$y_c(n) = (\cos\omega_0)y_c(n-1) - (\sin\omega_0)y_s(n-1) \tag{5.4.56}$$

$$y_s(n) = (\sin\omega_0)y_c(n-1) + (\cos\omega_0)y_s(n-1) \tag{5.4.57}$$

which can also be expressed in matrix form as

$$\begin{bmatrix} y_c(n) \\ y_s(n) \end{bmatrix} = \begin{bmatrix} \cos\omega_0 & -\sin\omega_0 \\ \sin\omega_0 & \cos\omega_0 \end{bmatrix} \begin{bmatrix} y_c(n-1) \\ y_s(n-1) \end{bmatrix} \tag{5.4.58}$$

The structure for the realization of the coupled-form oscillator is illustrated in Fig. 5.4.19. We note that this is a two-output system which is not driven by any input, but which requires the initial conditions $y_c(-1) = A\cos\omega_0$ and $y_s(-1) = -A\sin\omega_0$ in order to begin its self-sustaining oscillations.

Finally, it is interesting to note that (5.4.58) corresponds to vector rotation in the two-dimensional coordinate system with coordinates $y_c(n)$ and $y_s(n)$. As a consequence, the coupled-form oscillator can also be implemented by use of the so-called CORDIC algorithm [see the book by Kung et al. (1985)].

5.5 Inverse Systems and Deconvolution

As we have seen, a linear time-invariant system takes an input signal $x(n)$ and produces an output signal $y(n)$, which is the convolution of $x(n)$ with the unit sample response $h(n)$ of the system. In many practical applications we are given an output signal from a system whose characteristics are unknown and we are asked to determine the input signal. For example, in the transmission of digital information at high data rates over telephone channels, it is well known that the channel distorts the signal and causes intersymbol interference among the data symbols. The intersymbol interference may cause errors when we attempt to recover the data. In such a case the problem is to design a corrective system which, when cascaded with the channel, produces an output that, in some sense, corrects for the distortion caused by the channel, and thus yields a replica of the desired transmitted signal. In digital communications such a corrective system is called an *equalizer*. In the general context of linear systems theory, however, we call the corrective system an *inverse system*, because the corrective system has a frequency response which is basically the reciprocal of the frequency response of the system that caused the distortion. Furthermore, since the distortive system yields an output $y(n)$ that is the convolution of the input $x(n)$ with the impulse response $h(n)$, the inverse system operation that takes $y(n)$ and produces $x(n)$ is called *deconvolution*.

If the characteristics of the distortive system are unknown, it is often necessary, when possible, to excite the system with a known signal, observe the output, compare it with the input, and in some manner, determine the characteristics of the system. For example, in the digital communication problem just described, where the frequency response of the channel is unknown, the measurement of the channel frequency response can be accomplished by transmitting a set of equal-amplitude sinusoids, at different frequencies with a specified set of phases, within the frequency band of the channel. The channel will attenuate and phase shift each of the sinusoids. By comparing the received signal with the transmitted signal, the receiver obtains a measurement of the channel frequency response which can be used to design the inverse system. The process of determining the characteristics of the unknown system, either $h(n)$ or $H(\omega)$, by a set of measurements performed on the system is called *system identification*.

The term "deconvolution" is often used in seismic signal processing, and more generally, in geophysics to describe the operation of separating the input signal from the characteristics of the system which we wish to measure. The deconvolution operation is actually intended to identify the characteristics of the system, which in this case, is the earth, and can also be viewed as a system identification problem. The "inverse system," in this case, has a frequency response that is the reciprocal of the input signal spectrum that has been used to excite the system.

5.5.1 Invertibility of Linear Time-Invariant Systems

A system is said to be *invertible* if there is a one-to-one correspondence between its input and output signals. This definition implies that if we know the output sequence $y(n)$, $-\infty < n < \infty$, of an invertible system $\mathcal{T}$, we can uniquely determine its input $x(n)$, $-\infty < n < \infty$. The *inverse system* with input $y(n)$ and output $x(n)$ is denoted by $\mathcal{T}^{-1}$. Clearly, the cascade connection of a system and its inverse is equivalent to the identity system, since

$$w(n) = \mathcal{T}^{-1}[y(n)] = \mathcal{T}^{-1}\{\mathcal{T}[x(n)]\} = x(n) \tag{5.5.1}$$

as illustrated in Fig. 5.5.1. For example, the systems defined by the input–output relations $y(n) = ax(n)$ and $y(n) = x(n-5)$ are invertible, whereas the input–output relations $y(n) = x^2(n)$ and $y(n) = 0$ represent noninvertible systems.

As indicated above, inverse systems are important in many practical applications, including geophysics and digital communications. Let us begin by considering the problem of determining the inverse of a given system. We limit our discussion to the class of linear time-invariant discrete-time systems.

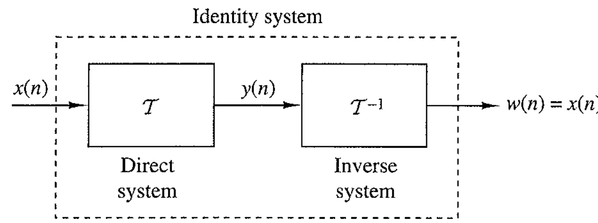


Figure 5.5.1
System $\mathcal{T}$ in cascade with its inverse $\mathcal{T}^{-1}$.

Now, suppose that the linear time-invariant system $\mathcal{T}$ has an impulse response $h(n)$ and let $h_I(n)$ denote the impulse response of the inverse system $\mathcal{T}^{-1}$. Then (5.5.1) is equivalent to the convolution equation

$$w(n) = h_I(n) * h(n) * x(n) = x(n) \tag{5.5.2}$$

But (5.5.2) implies that

$$h(n) * h_I(n) = \delta(n) \tag{5.5.3}$$

The convolution equation in (5.5.3) can be used to solve for $h_I(n)$ for a given $h(n)$. However, the solution of (5.5.3) in the time domain is usually difficult. A simpler approach is to transform (5.5.3) into the z-domain and solve for $\mathcal{T}^{-1}$. Thus in the z-transform domain, (5.5.3) becomes

$$H(z)H_I(z) = 1$$

and therefore the system function for the inverse system is

$$H_I(z) = \frac{1}{H(z)} \tag{5.5.4}$$

If $H(z)$ has a rational system function

$$H(z) = \frac{B(z)}{A(z)} \tag{5.5.5}$$

then

$$H_I(z) = \frac{A(z)}{B(z)} \tag{5.5.6}$$

Thus the zeros of $H(z)$ become the poles of the inverse system, and vice versa. Furthermore, if $H(z)$ is an FIR system, then $H_I(z)$ is an all-pole system, or if $H(z)$ is an all-pole system, then $H_I(z)$ is an FIR system.

EXAMPLE 5.5.1

Determine the inverse of the system with impulse response

$$h(n) = (\tfrac{1}{2})^n u(n)$$

Solution. The system function corresponding to $h(n)$ is

$$H(z) = \frac{1}{1 - \frac{1}{2}z^{-1}}, \qquad \text{ROC: } |z| > \frac{1}{2}$$

This system is both causal and stable. Since $H(z)$ is an all-pole system, its inverse is FIR and is given by the system function

$$H_I(z) = 1 - \frac{1}{2}z^{-1}$$

Hence its impulse response is

$$h_I(n) = \delta(n) - \frac{1}{2}\delta(n-1)$$

EXAMPLE 5.5.2

Determine the inverse of the system with impulse response

$$h(n) = \delta(n) - \frac{1}{2}\delta(n-1)$$

This is an FIR system and its system function is

$$H(z) = 1 - \frac{1}{2}z^{-1}, \qquad \text{ROC: } |z| > 0$$

The inverse system has the system function

$$H_I(z) = \frac{1}{H(z)} = \frac{1}{1 - \frac{1}{2}z^{-1}} = \frac{z}{z - \frac{1}{2}}$$

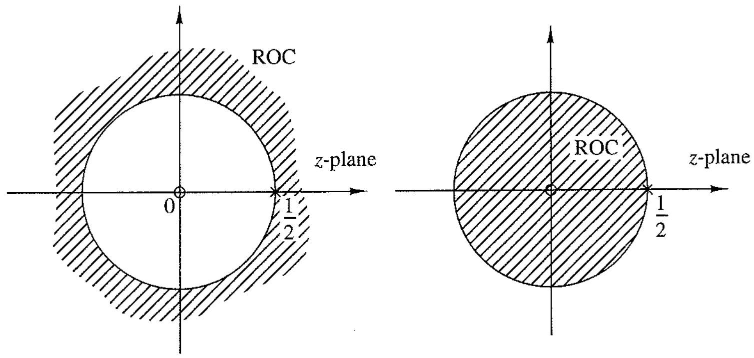
Thus $H_I(z)$ has a zero at the origin and a pole at $z = \frac{1}{2}$. In this case there are two possible regions of convergence and hence two possible inverse systems, as illustrated in Fig. 5.5.2. If we take the ROC of $H_I(z)$ as $|z| > \frac{1}{2}$, the inverse transform yields

$$h_I(n) = (\frac{1}{2})^n u(n)$$

which is the impulse response of a causal and stable system. On the other hand, if the ROC is assumed to be $|z| < \frac{1}{2}$, the inverse system has an impulse response

$$h_I(n) = -\left(\frac{1}{2}\right)^n u(-n-1)$$

In this case the inverse system is anticausal and unstable.



(a) (b)

Figure 5.5.2 Two possible regions of convergence for $H(z) = z/(z - \frac{1}{2})$.

We observe that (5.5.3) cannot be solved uniquely by using (5.5.6) unless we specify the region of convergence for the system function of the inverse system.

In some practical applications the impulse response $h(n)$ does not possess a z-transform that can be expressed in closed form. As an alternative we may solve (5.5.3) directly using a digital computer. Since (5.5.3) does not, in general, possess a unique solution, we assume that the system and its inverse are causal. Then (5.5.3) simplifies to the equation

$$\sum_{k=0}^{n} h(k)h_I(n-k) = \delta(n) \tag{5.5.7}$$

By assumption, $h_I(n) = 0$ for $n < 0$. For $n = 0$ we obtain

$$h_I(0) = 1/h(0) \tag{5.5.8}$$

The values of $h_I(n)$ for $n \geq 1$ can be obtained recursively from the equation

$$h_I(n) = \sum_{k=1}^{n} \frac{h(n)h_I(n-k)}{h(0)}, \qquad n \geq 1 \tag{5.5.9}$$

This recursive relation can easily be programmed on a digital computer.

There are two problems associated with (5.5.9). First, the method does not work if $h(0) = 0$. However, this problem can easily be remedied by introducing an appropriate delay in the right-hand side of (5.5.7), that is, by replacing $\delta(n)$ by $\delta(n-m)$, where $m = 1$ if $h(0) = 0$ and $h(1) \neq 0$, and so on. Second, the recursion in (5.5.9) gives rise to round-off errors which grow with n and, as a result, the numerical accuracy of $h(n)$ deteriorates for large n.

EXAMPLE 5.5.3

Determine the causal inverse of the FIR system with impulse response

$$h(n) = \delta(n) - \alpha\delta(n-1)$$

Since $h(0) = 1$, $h(1) = -\alpha$, and $h(n) = 0$ for $n \geq \alpha$, we have

$$h_I(0) = 1/h(0) = 1$$

and

$$h_I(n) = \alpha h_I(n-1), \qquad n \geq 1$$

Consequently,

$$h_I(1) = \alpha, \quad h_I(2) = \alpha^2, \quad \ldots, \quad h_I(n) = \alpha^n$$

which corresponds to a causal IIR system as expected.

5.5.2 Minimum-Phase, Maximum-Phase, and Mixed-Phase Systems

The invertibility of a linear time-invariant system is intimately related to the characteristics of the phase spectral function of the system. To illustrate this point, let us consider two FIR systems, characterized by the system functions

$$H_1(z) = 1 + \frac{1}{2}z^{-1} = z^{-1}(z + \frac{1}{2}) \tag{5.5.10}$$

$$H_2(z) = \frac{1}{2} + z^{-1} = z^{-1}(\frac{1}{2}z + 1) \tag{5.5.11}$$

The system in (5.5.10) has a zero at $z = -\frac{1}{2}$ and an impulse response $h(0) = 1$, $h(1) = 1/2$. The system in (5.5.11) has a zero at $z = -2$ and an impulse response $h(0) = 1/2$, $h(1) = 1$, which is the reverse of the system in (5.5.10). This is due to the reciprocal relationship between the zeros of $H_1(z)$ and $H_2(z)$.

In the frequency domain, the two systems are characterized by their frequency response functions, which can be expressed as

$$|H_1(\omega)| = |H_2(\omega)| = \sqrt{\frac{5}{4} + \cos\omega} \tag{5.5.12}$$

and

$$\Theta_1(\omega) = -\omega + \tan^{-1}\frac{\sin\omega}{\frac{1}{2} + \cos\omega} \tag{5.5.13}$$

$$\Theta_2(\omega) = -\omega + \tan^{-1}\frac{\sin\omega}{2 + \cos\omega} \tag{5.5.14}$$