Chapter 9

Finite Impulse Response Filters

In this chapter we study the structure, properties, and design methods of digital FIR filters. Digital FIR filters have several favorable properties, thanks to which they are popular in digital signal processing. First and foremost of those is the linear-phase property, which, as we saw in Chapter 8, provides distortionless response (or nearly so) for signals in the pass band. We therefore begin this chapter with an expanded discussion of the linear-phase property, and study its manifestations in the time and frequency domains.

The simplest design method for FIR filters is impulse response truncation, so this is the first to be presented. This method is not quite useful by itself, since it has undesirable frequency-domain characteristics. However, it serves as a necessary introduction to the second design method to be presented—windowing. We have already encountered windows in Section 6.2, in the context of spectral analysis. Here we shall learn how windows can be used for mitigating the adverse effects of impulse response truncation in the same way as they mitigate the effects of signal truncation in spectral analysis.

The windowing design method, although simple and convenient, is not optimal. By this we mean that, for given pass-band and stop-band specifications, their order is not the minimum possible. We present two design methods based on optimality criteria. The first is *least-squares design*, which minimizes an integral-of-square-error criterion in the frequency domain. The second is *equiripple design*, which minimizes the maximum ripple in each band. Equiripple design is intricate in its basic theory and details of implementation. Fortunately, there exist well-developed computer programs for this purpose, which take much of the burden of the design from the individual user. We shall therefore devote relatively little space to this topic, presenting only its principles.

9.1 Generalized Linear Phase Revisited

Practical FIR filters are usually designed to have a linear phase, either exact or generalized. We shall omit the modifiers "exact" and "generalized," calling filters having either constant phase delay or constant group delay *linear-phase filters.* The transfer function of an FIR filter is usually expressed in terms of its impulse response coefficients, that is,

$$H^z(z) = h[0] + h[1]z^{-1} + \ldots + h[N]z^{-N}. \tag{9.1}$$

Although FIR filters are a special case of rational filters, we do not use the polynomial notation $b(z)$ for these filters. We use N to denote the *order* of the filter; in some books N is used for the *length* of the filter, so watch out for differences resulting from this notation. An Nth-order filter has $N+1$ nonzero coefficients, so its length is $N+1$. Thus, if the order is even, the length is odd, and vice versa.

In this section we expand the discussion of linear-phase filters and derive expressions for the frequency responses of the four filter types. Then we discuss the restrictions on the zero locations resulting from the linear-phase property.

9.1.1 Type-I Filters

Type-I filters have even order, $N = 2M$, and initial phase $\phi_0 = 0$. Consequently, they have constant phase delay. Their impulse response satisfies the symmetry condition

$$h[n] = h[N-n]. \tag{9.2}$$

Their frequency response is derived as follows:

$$\begin{aligned}
H^{\mathrm{f}}(\theta) &= \sum_{n=0}^{2M} h[n]e^{-j\theta n} = e^{-j\theta M}\sum_{n=0}^{2M} h[n]e^{j\theta(M-n)} \\
&= e^{-j\theta M}\left\{h[M] + \sum_{n=0}^{M-1} h[n]e^{j\theta(M-n)} + \sum_{n=M+1}^{2M} h[n]e^{j\theta(M-n)}\right\} \\
&= e^{-j\theta M}\left\{h[M] + \sum_{n=0}^{M-1} h[n]e^{j\theta(M-n)} + \sum_{n=0}^{M-1} h[N-n]e^{-j\theta(M-n)}\right\} \\
&= e^{-j\theta M}\left\{h[M] + 2\sum_{n=0}^{M-1} h[n]\cos[\theta(M-n)]\right\}. \qquad (9.3)
\end{aligned}$$

In passing from the second to the third line, we substituted $n' = N-n$ in the second term and then renamed n' as n. In the fourth line we used Euler's formula for $\cos[\theta(M-n)]$.

The amplitude function of a type-I filter can be written in the form

$$A_{\mathrm{I}}(\theta) = \sum_{n=0}^{M} g[n]\cos(\theta n), \tag{9.4}$$

where

$$g[n] = \begin{cases} h[M], & n = 0, \\ 2h[M-n], & 1 \le n \le M. \end{cases} \tag{9.5}$$

The amplitude function is symmetric in θ and has period 2π. Figure 9.1 shows typical impulse and amplitude responses of a type-I filter.

Example 9.1 The filter

$$H^{z}(z) = 1 + 0.5z^{-1} - 0.3z^{-2} + 1.2z^{-3} - 0.3z^{-4} + 0.5z^{-5} + z^{-6}$$

is type I. Its amplitude function is

$$A(\theta) = 1.2 - 0.6\cos\theta + \cos(2\theta) + 2\cos(3\theta).$$

□

9.1.2 Type-II Filters

Type-II filters have odd order, $N = 2M+1$, and initial phase $\phi_0 = 0$. Like type-I filters, they have constant phase delay. Their impulse response satisfies the symmetry

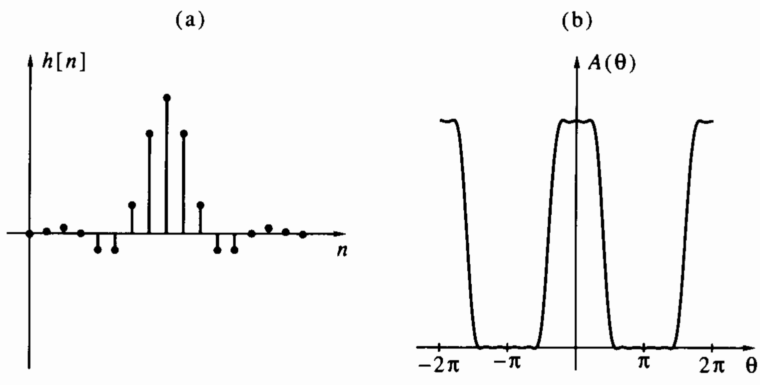

Figure 9.1 Type-I FIR filter: (a) impulse response; (b) amplitude response.

condition (9.2). Their frequency response is derived as follows:

$$\begin{aligned}
H^{\mathrm{f}}(\theta) &= \sum_{n=0}^{2M+1} h[n]e^{-j\theta n} = e^{-j\theta(M+0.5)} \sum_{n=0}^{2M+1} h[n]e^{j\theta(M+0.5-n)} \\
&= e^{-j\theta(M+0.5)} \left\{ \sum_{n=0}^{M} h[n]e^{j\theta(M+0.5-n)} + \sum_{n=M+1}^{2M+1} h[n]e^{j\theta(M+0.5-n)} \right\} \\
&= e^{-j\theta(M+0.5)} \left\{ \sum_{n=0}^{M} h[n]e^{j\theta(M+0.5-n)} + \sum_{n=0}^{M} h[N-n]e^{-j\theta(M+0.5-n)} \right\} \\
&= e^{-j\theta(M+0.5)} \left\{ 2\sum_{n=0}^{M} h[n]\cos[\theta(M+0.5-n)] \right\}.
\end{aligned} \tag{9.6}$$

The amplitude function of a type-II filter can be written in the form

$$A_{\mathrm{II}}(\theta) = \cos(0.5\theta) \sum_{n=0}^{M} g[n]\cos(\theta n). \tag{9.7}$$

This is proved as follows:

$$\begin{aligned}
&\cos(0.5\theta) \sum_{n=0}^{M} g[n]\cos(\theta n) = \cos(0.5\theta) \sum_{n=0}^{M} g[M-n]\cos[\theta(M-n)] \\
&= 0.5 \sum_{n=0}^{M} g[M-n]\{\cos[\theta(M+0.5-n)] + \cos[\theta(M-0.5-n)]\} \\
&= 0.5 \sum_{n=1}^{M} \{g[M-n] + g[M+1-n]\}\cos[\theta(M+0.5-n)] \\
&\quad + 0.5g[M]\cos[\theta(M+0.5)] + 0.5g[0]\cos(0.5\theta).
\end{aligned} \tag{9.8}$$

Thus we can make (9.8) equal to $A_{\mathrm{II}}(\theta)$ by choosing

$$2h[n] = \begin{cases} 0.5g[M], & n = 0, \\ 0.5\{g[M+1-n] + g[M-n]\}, & 1 \le n \le M-1, \\ 0.5g[1] + g[0], & n = M. \end{cases} \tag{9.9}$$

The coefficients $g[n]$ can be computed from the impulse response $h[n]$ by inverting the relationships (9.9). This yields the iterative formulas

$$g[M] = 4h[0], \tag{9.10a}$$
$$g[M-n] = 4h[n] - g[M+1-n], \quad 1 \le n \le M-1, \tag{9.10b}$$
$$g[0] = 2h[M] - 0.5g[1]. \tag{9.10c}$$

The amplitude function of a type-II filter is symmetric in θ and has period 4π. As we see from (9.7), $A_{\mathrm{II}}(\pi) = 0$, because $\cos(0.5\pi) = 0$. Therefore, type-II filters are not suitable for high-pass or band-stop filters. This is in contrast with type-I filters, which are suitable for filters of all kinds. Figure 9.2 shows typical impulse and amplitude responses of a type-II filter.

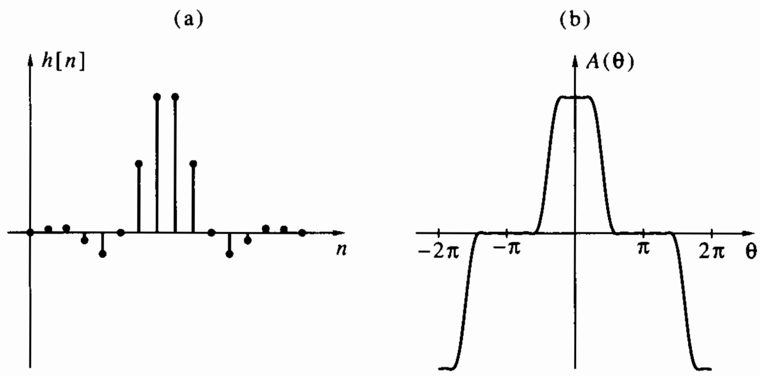

Figure 9.2 Type-II FIR filter: (a) impulse response; (b) amplitude response.

Example 9.2 The filter

$$H^z(z) = 0.4 + 0.6z^{-1} + 1.5z^{-2} + 1.5z^{-3} + 0.6z^{-4} + 0.4z^{-5}$$

is type II. Its amplitude function is

$$A(\theta) = \cos(0.5\theta)[2.6 + 0.8\cos\theta + 1.6\cos(2\theta)].$$ □

9.1.3 Type-III Filters

Type-III filters have even order, $N = 2M$, and initial phase $\phi_0 = \pi/2$. Consequently, they have constant group delay, but not constant phase delay. Their impulse response satisfies the antisymmetry condition

$$h[n] = -h[N-n]. \tag{9.11}$$

Their frequency response is derived similarly to (9.3) (note that $h[M] = 0$, because of the antisymmetry condition):

$$\begin{aligned} H^{\mathrm{f}}(\theta) &= e^{-j\theta M}\left\{\sum_{n=0}^{M-1} h[n]e^{j\theta(M-n)} + \sum_{n=M+1}^{2M} h[n]e^{j\theta(M-n)}\right\} \\ &= e^{-j\theta M}\left\{\sum_{n=0}^{M-1} h[n]e^{j\theta(M-n)} + \sum_{n=0}^{M-1} h[N-n]e^{-j\theta(M-n)}\right\} \\ &= e^{j(0.5\pi-\theta M)}\left\{2\sum_{n=0}^{M-1} h[n]\sin[\theta(M-n)]\right\}. \end{aligned} \tag{9.12}$$

The amplitude function of a type-III filter can be written in the form

$$A_{\text{III}}(\theta) = \sin\theta \sum_{n=0}^{M-1} g[n]\cos(\theta n). \tag{9.13}$$

This is proved as follows:

$$\begin{aligned}
&\sin\theta \sum_{n=0}^{M-1} g[n]\cos(\theta n) = \sin\theta \sum_{n=1}^{M} g[M-n]\cos[\theta(M-n)] \\
&= 0.5 \sum_{n=1}^{M} g[M-n]\{\sin[\theta(M+1-n)] - \sin[\theta(M-1-n)]\} \\
&= 0.5 \sum_{n=0}^{M-1} g[M-1-n]\sin[\theta(M-n)] - 0.5 \sum_{n=2}^{M} g[M+1-n]\sin[\theta(M-n)] \\
&\quad + 0.5g[0]\sin\theta.
\end{aligned} \tag{9.14}$$

Thus we can make (9.14) equal to $A_{\text{III}}(\theta)$ by choosing

$$2h[n] = \begin{cases} 0.5g[M-1], & n = 0, \\ 0.5g[M-2], & n = 1, \\ 0.5\{g[M-1-n] - g[M+1-n]\}, & 2 \le n \le M-2, \\ g[0] - 0.5g[2], & n = M-1. \end{cases} \tag{9.15}$$

The coefficients $g[n]$ can be computed from the impulse response $h[n]$ by inverting the relationships (9.15). This gives the iterative formulas

$$g[M-1] = 4h[0], \tag{9.16a}$$

$$g[M-2] = 4h[1], \tag{9.16b}$$

$$g[M-n] = 4h[n-1] + g[M+2-n], \quad 3 \le n \le M-1, \tag{9.16c}$$

$$g[0] = 2h[M-1] + 0.5g[2]. \tag{9.16d}$$

The amplitude function of a type-III filter is antisymmetric in θ and has period 2π. As we see from (9.13), $A_{\text{III}}(0) = A_{\text{III}}(\pi) = 0$, because $\sin 0 = \sin\pi = 0$. Therefore, type-III filters are not suitable for low-pass, high-pass, or band-stop filters. They can be used for band-pass filters, but since they do not have exact linear phase, only constant group delay, they are not common for this application. Type III is sometimes used for differentiators and Hilbert transformers, as we shall illustrate later. Figure 9.3 shows typical impulse and amplitude responses of a type-III filter.

Example 9.3 The filter

$$H^z(z) = 1 + 0.5z^{-1} - 0.3z^{-2} + 0.3z^{-4} - 0.5z^{-5} - z^{-6}$$

is type III. Its amplitude function is

$$A(\theta) = \sin\theta[1.4 + 2\cos\theta + 4\cos(2\theta)].$$

□

9.1.4 Type-IV Filters

Type-IV filters have odd order, $N = 2M + 1$, and initial phase $\phi_0 = \pi/2$. Like type-III filters, they have constant group delay, but not constant phase delay. Their impulse response satisfies the antisymmetry condition (9.11). Their frequency response is derived similarly to (9.6):

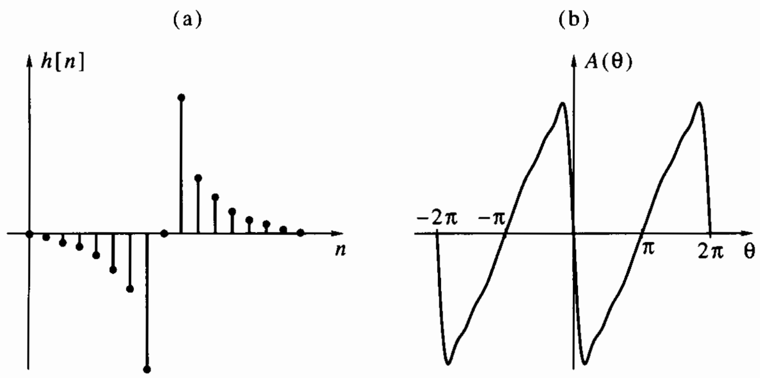

Figure 9.3 Type-III FIR filter: (a) impulse response; (b) amplitude response.

$$\begin{aligned} H^{\mathrm{f}}(\theta) &= e^{-j\theta(M+0.5)}\left\{\sum_{n=0}^{M} h[n]e^{j\theta(M+0.5-n)} + \sum_{n=0}^{M} h[N-n]e^{-j\theta(M+0.5-n)}\right\} \\ &= e^{j[0.5\pi-\theta(M+0.5)]}\left\{2\sum_{n=0}^{M} h[n]\sin[\theta(M+0.5-n)]\right\}. \end{aligned} \tag{9.17}$$

The amplitude function of a type-IV filter can be written in the form

$$A_{\mathrm{IV}}(\theta) = \sin(0.5\theta)\sum_{n=0}^{M} g[n]\cos(\theta n). \tag{9.18}$$

This is proved as follows:

$$\begin{aligned} \sin(0.5\theta)\sum_{n=0}^{M} g[n]\cos(\theta n) &= \sin(0.5\theta)\sum_{n=0}^{M} g[M-n]\cos[\theta(M-n)] \\ &= 0.5\sum_{n=0}^{M} g[M-n]\{\sin[\theta(M+0.5-n)] - \sin[\theta(M-0.5-n)]\} \\ &= 0.5\sum_{n=1}^{M} \{g[M-n] - g[M+1-n]\}\sin[\theta(M+0.5-n)] \\ &\quad + 0.5g[M]\sin[\theta(M+0.5)] + 0.5g[0]\sin(0.5\theta). \end{aligned} \tag{9.19}$$

Thus we can make (9.19) equal to $A_{\mathrm{IV}}(\theta)$ by choosing

$$2h[n] = \begin{cases} 0.5g[M], & n = 0, \\ 0.5\{g[M-n] - g[M+1-n]\}, & 1 \le n \le M-1, \\ g[0] - 0.5g[1], & n = M. \end{cases} \tag{9.20}$$

The coefficients $g[n]$ can be computed from the impulse response $h[n]$ by inverting the relationships (9.20). This gives the iterative formulas

$$g[M] = 4h[0], \tag{9.21a}$$

$$g[M-n] = 4h[n] + g[M+1-n], \quad 1 \le n \le M-1, \tag{9.21b}$$

$$g[0] = 2h[M] + 0.5g[1]. \tag{9.21c}$$

The amplitude function of a type-IV filter is antisymmetric in θ and has period 4π. As we see from (9.18), $A_{\mathrm{IV}}(0) = 0$, because $\sin 0 = 0$. Therefore, type-IV filters are not suitable for low-pass or band-stop filters. They can be used for high-pass and band-pass filters, but this is not common. Similarly to type-III filters, they are useful for differentiators and Hilbert transformers. Figure 9.4 shows typical impulse and amplitude responses of a type-IV filter.

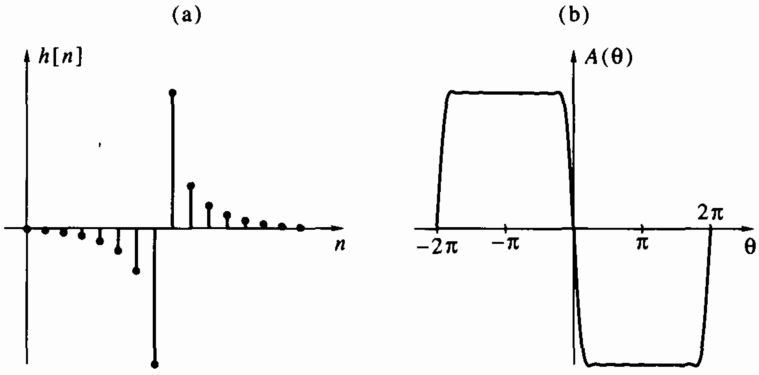

Figure 9.4 Type-IV FIR filter: (a) impulse response; (b) amplitude response.

Example 9.4 The filter

$$H^z(z) = 0.4 + 0.6z^{-1} + 1.5z^{-2} - 1.5z^{-3} - 0.6z^{-4} - 0.4z^{-5}$$

is type IV. Its amplitude function is

$$A(\theta) = \sin(0.5\theta)[5 + 4\cos\theta + 1.6\cos(2\theta)].$$ □

9.1.5 Summary of Linear-Phase Filter Types

As we have seen, the frequency responses of linear-phase FIR filters of all four types are of the form

$$H^f(\theta) = A(\theta)e^{j(\phi_0 - 0.5\theta N)}, \tag{9.22}$$

where the amplitude response has the form

$$A(\theta) = F(\theta)G(\theta), \quad G(\theta) = \sum_{k=0}^{K} g[k]\cos(\theta k), \tag{9.23}$$

and $F(\theta)$ is one of the functions 1, $\cos(0.5\theta)$, $\sin\theta$, $\sin(0.5\theta)$, depending on the type. Table 9.1 summarizes the properties and parameters of the four types as a function of the order N and the impulse response coefficients $h[n]$.

9.1.6 Zero Locations of Linear-Phase Filters

The zeros of a linear-phase FIR filter cannot be completely arbitrary, but must satisfy certain constraints. Let $H^z(z)$ be the transfer function of an Nth-order, linear-phase filter with real coefficients. Then

$$H^z(z^{-1}) = \sum_{n=0}^{N} h[n]z^n = z^N \sum_{n=0}^{N} h[n]z^{-(N-n)} = z^N \sum_{n=0}^{N} h[N-n]z^{-n} = \pm z^N H^z(z), \tag{9.24}$$

where the sign of the right side is positive for a symmetric filter and negative for an antisymmetric filter. Let β be a zero of $H^z(z)$, so $H^z(\beta) = 0$. Then we get from (9.24)

Type	I	II	III	IV
Order	even	odd	even	odd
Symmetry of $h[n]$	symmetric	symmetric	anti-symmetric	anti-symmetric
Symmetry of $A(\theta)$	symmetric	symmetric	anti-symmetric	anti-symmetric
Period of $A(\theta)$	2π	4π	2π	4π
ϕ_0	0	0	0.5π	0.5π
$F(\theta)$ in (9.23)	1	$\cos(0.5\theta)$	$\sin\theta$	$\sin(0.5\theta)$
K in (9.23)	$N/2$	$(N-1)/2$	$(N-2)/2$	$(N-1)/2$
$g[n]$ in (9.23)	see (9.5)	see (9.10)	see (9.16)	see (9.21)
$H^{\mathrm{f}}(0)$	arbitrary	arbitrary	0	0
$H^{\mathrm{f}}(\pi)$	arbitrary	0	0	arbitrary
Uses	LP, HP, BP, BS, Multiband filters	LP, BP	Differentiators, Hilbert transformers	

Table 9.1 Properties and parameters of the four FIR filter types.

that $H^z(\beta^{-1}) = 0$ (note that necessarily $\beta \neq 0$, since a minimal FIR filter cannot have a zero at the origin). This shows that, whenever β is a zero of $H(z)$, so is β^{-1}. Therefore, the zeros of a linear-phase filter must obey the following restrictions:

1. A zero $\beta = 1$ can appear any number of times, because it is self reciprocal.
2. A zero $\beta = -1$ can appear any number of times, because it is self reciprocal.
3. If there is a zero $\beta = e^{j\zeta}$, where $\zeta \neq 0$, $\zeta \neq \pi$, then $\bar{\beta} = e^{-j\zeta}$ is also a zero, and the two are the reciprocals of each other. Such pairs can appear any number of times.
4. If there is a zero $\beta = r$, where r is real and $|r| \neq 1$, then $\beta^{-1} = r^{-1}$ must also be a zero. Such pairs can appear any number of times.
5. If there is a zero $\beta = re^{j\zeta}$, where $r \neq 1$, $\zeta \neq 0$, $\zeta \neq \pi$, then

$$\bar{\beta} = re^{-j\zeta}, \quad \beta^{-1} = r^{-1}e^{-j\zeta}, \quad \bar{\beta}^{-1} = r^{-1}e^{j\zeta}$$

must also be zeros. Such quadruples can appear any number of times.

Figure 9.5 illustrates the possible zero locations of a linear-phase FIR filter in the z-plane. The numbers in the figure match the numbers in the preceding list.

Recalling the discussion in Section 8.4.7, we conclude that a linear-phase filter is not minimum phase in general, since it has zeros outside the unit circle. The only exception occurs when *all* the zeros of the filter are on the unit circle, but such a case is rare in practice. We also recall that a minimum-phase filter can be obtained from any linear-phase filter by replacing the zeros outside the unit circle by their conjugate inverses. Such a procedure will leave the magnitude response unchanged. The minimum-phase filter obtained by this procedure has the following property: All its zeros inside the unit circle have multiplicity 2 at least.

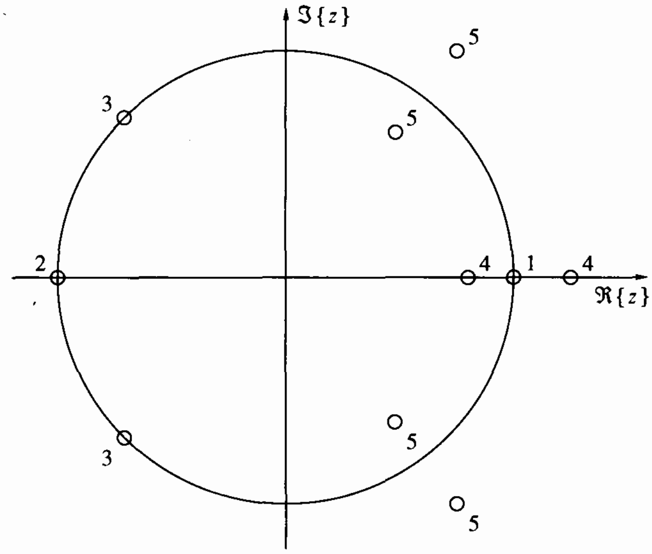

Figure 9.5 Possible zero locations of a linear-phase FIR filter.

Example 9.5 Let

$$H^z(z) = 1 - 0.5z^{-1} + 0.25z^{-2} + 0.875z^{-3} - 0.5z^{-4} + 0.25z^{-5} - 0.125z^{-6}.$$

We wish to find a linear-phase filter $H_1^z(z)$ of minimal order such that the zeros of $H^z(z)$ will be a subset of the zeros of $H_1^z(z)$. We find, using the function `roots` of MATLAB, that the zeros of $H^z(z)$ are

$$\beta_1 = -1, \quad \beta_2 = 0.5, \quad \beta_{3,4} = 0.5(1 \pm j\sqrt{3}), \quad \beta_{5,6} = \pm j0.5.$$

The filter $H_1^z(z)$ must include, in addition to these zeros, zeros at

$$\beta_7 = 2, \quad \beta_{8,9} = \pm j2.$$

Note that β_1 need not be duplicated because it is self reciprocal, and $\beta_{3,4}$ need not be duplicated because they are the reciprocals of each other. Thus, $H_1^z(z)$ is given by

$$\begin{aligned} H_1(z) = H^z(z)(1 - 2z^{-1})(1 + 4z^{-2}) &= 1 - 2.5z^{-1} + 5.25z^{-2} - 9.625z^{-3} \\ &+ 2.75z^{-4} + 2.75z^{-5} - 9.625z^{-6} + 5.25z^{-7} - 2.5z^{-8} + z^{-9}. \end{aligned}$$

Next we wish to find a minimum-phase filter $H_2^z(z)$ having the same magnitude response as $H_1^z(z)$. The filter $H_2^z(z)$ includes $H^z(z)$ as a factor, because $H^z(z)$ is minimum phase, as well as the conjugate inverses of $\beta_{7,8,9}$, because these zeros are outside the unit circle. Therefore,

$$\begin{aligned} H_2(z) &= K \cdot H^z(z)(1 - 0.5z^{-1})(1 + 0.25z^{-2}) \\ &= K(1 - z^{-1} + 0.75z^{-2} + 0.5z^{-3} - 0.8125z^{-4} + 0.6875z^{-5} \\ &\quad - 0.4844z^{-6} + 0.1875z^{-7} - 0.0625z^{-8} + 0.0156z^{-9}). \end{aligned}$$

The factor K should be chosen so as to make the DC gains of the two filters equal. This is equivalent to saying that the sum of the coefficients must be equal. We get from this condition that $K = -8$. □

9.2 FIR Filter Design by Impulse Response Truncation

9.2.1 Definition of the IRT Method

Ideal frequency responses, such as the ones shown in Figure 8.1, have infinite impulse responses. However, by Parseval's theorem, the impulse response $h[n]$ has finite energy. Truncating the impulse response of the ideal filter on both the right and the left thus yields a finite impulse response whose associated frequency response approximates that of the ideal filter. Furthermore, by shifting the truncated impulse response to the right (i.e., by delaying it), we can make it causal. This is the basic idea of the impulse response truncation (IRT) design method.

The phase response of an ideal filter is usually either identically zero, in which case the impulse response is symmetric, or identically $\pi/2$, in which case the impulse response is antisymmetric. In either case, the larger (in magnitude) impulse response coefficients are the ones whose indices n are close to zero. It is therefore reasonable to truncate the impulse response symmetrically around $n = 0$, before shifting it for causality. The filter thus obtained will be symmetric (in the case of zero phase) or antisymmetric (in the case of phase $\pi/2$), so it will have linear phase. This, however, limits the filter's length to an odd number, hence its order to an even number, hence its type to I or III. An alternative approach, which frees us from the constraint of even order, is to incorporate a linear-phase factor into the ideal response, that is, a factor $e^{-j0.5\theta N}$, in which N reflects the desired order after truncation. Then, after the impulse response has been computed, it is truncated to the range $0 \le n \le N$. The filter thus obtained is causal, has linear phase, and approximates the ideal frequency response. Its order can be even or odd, depending on the choice of N in the linear-phase factor. Thus, the FIR filter can be of any of the four types.[1] In summary, the impulse response truncation method consists of the following steps:

Impulse-response truncation FIR filter design procedure

1. Write the desired amplitude response $A_{\mathrm{d}}(\theta)$, according to the filter class: low pass, high pass, etc.
2. Choose the filter's phase characteristic: integer or fractional group delay, initial phase 0 or $\pi/2$.
3. Choose the filter's order N. The ideal desired frequency response, including the phase, can now be written as
$$H_{\mathrm{d}}^{\mathrm{f}}(\theta) = A_{\mathrm{d}}(\theta)e^{j(\mu 0.5\pi - 0.5\theta N)}, \tag{9.25}$$
where $\mu = 0$ for a symmetric filter, and $\mu = 1$ for an antisymmetric one.
4. Compute the impulse response of the ideal filter, using the inverse Fourier transform formula
$$h_{\mathrm{d}}[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} A_{\mathrm{d}}(\theta)e^{j(\mu 0.5\pi - 0.5\theta N)}e^{j\theta n}d\theta. \tag{9.26}$$
5. Truncate the impulse response by taking
$$h[n] = \begin{cases} h_{\mathrm{d}}[n], & 0 \le n \le N, \\ 0, & \text{otherwise.} \end{cases} \tag{9.27}$$

In most practical cases, the form of $A_{\mathrm{d}}(\theta)$ is simple, so the integral can be computed in a closed form. This is illustrated in the following sections, which discuss the design of the most common kinds of FIR filters.

9.2.2 Low-Pass, High-Pass, and Band-Pass Filters

Let the desired amplitude response be

$$A_{\mathrm{d}}(\theta) = \begin{cases} 1, & \theta_1 \le |\theta| \le \theta_2, \\ 0, & \text{otherwise.} \end{cases} \tag{9.28}$$

This is the ideal amplitude response of a band-pass filter. A low-pass filter is obtained as a special case, by taking $\theta_1 = 0$. Similarly, a high-pass filter is obtained as a special case, by taking $\theta_2 = \pi$. For these three kinds of filter it is reasonable to require constant phase delay, rather than constant group delay. This amounts to choosing $\mu = 0$, that is, filters of type I or II. Therefore, the ideal frequency response, including the linear-phase term, is

$$H_{\mathrm{d}}^{\mathrm{f}}(\theta) = \begin{cases} e^{-j0.5\theta N}, & \theta_1 \le |\theta| \le \theta_2, \\ 0, & \text{otherwise.} \end{cases} \tag{9.29}$$

The desired impulse response is computed as

$$\begin{aligned} h_{\mathrm{d}}[n] &= \frac{1}{2\pi}\int_{-\theta_2}^{-\theta_1} e^{j\theta(n-0.5N)}d\theta + \frac{1}{2\pi}\int_{\theta_1}^{\theta_2} e^{j\theta(n-0.5N)}d\theta \\ &= \frac{\theta_2}{\pi}\operatorname{sinc}\left[\frac{\theta_2(n-0.5N)}{\pi}\right] - \frac{\theta_1}{\pi}\operatorname{sinc}\left[\frac{\theta_1(n-0.5N)}{\pi}\right]. \end{aligned} \tag{9.30}$$

In the special case of a low-pass filter, we get

$$h_{\mathrm{d}}[n] = \frac{\theta_2}{\pi}\operatorname{sinc}\left[\frac{\theta_2(n-0.5N)}{\pi}\right]. \tag{9.31}$$

In the special case of a high-pass filter, we get

$$h_{\mathrm{d}}[n] = \delta[n-0.5N] - \frac{\theta_1}{\pi}\operatorname{sinc}\left[\frac{\theta_1(n-0.5N)}{\pi}\right], \tag{9.32}$$

provided N is even. Recall that odd N (i.e., type II) is not suitable for high-pass filters.

Example 9.6 Design a band-pass filter with band-edge frequencies 0.2π and 0.6π, and order $N = 40$. Figure 9.6 shows the impulse and amplitude responses of the filter. The oscillatory nature of the amplitude response, which results from the impulse response truncation, is discussed in detail later. □

9.2.3 Multiband Filters

The ideal amplitude response of a multiband filter is a superposition of band-pass amplitude responses. Suppose that the desired frequency response has K bands, where the band-edge frequencies of the kth band are $\theta_{1,k}$, $\theta_{2,k}$ and its gain is C_k. Then,

$$A_{\mathrm{d}}(\theta) = \sum_{k=1}^{K} A_{\mathrm{d},k}(\theta), \tag{9.33}$$

where

$$A_{\mathrm{d},k}(\theta) = \begin{cases} C_k, & \theta_{1,k} \le |\theta| \le \theta_{2,k}, \\ 0, & \text{otherwise.} \end{cases} \tag{9.34}$$

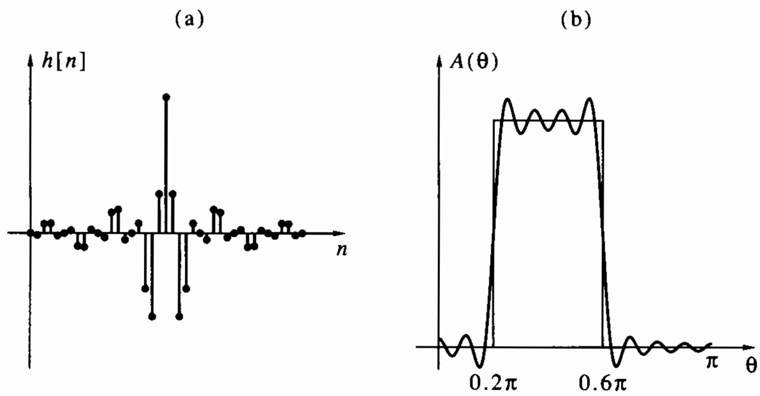

Figure 9.6 Band-pass filter designed by the IRT method, of order $N = 40$: (a) impulse response; (b) amplitude response.

Using superposition, we get from (9.30)

$$h_{\mathrm{d}}[n] = \sum_{k=1}^{K} \frac{C_k}{\pi} \left\{ \theta_{2,k} \operatorname{sinc}\left[\frac{\theta_{2,k}(n - 0.5N)}{\pi}\right] - \theta_{1,k} \operatorname{sinc}\left[\frac{\theta_{1,k}(n - 0.5N)}{\pi}\right] \right\}. \tag{9.35}$$

A band-stop filter can be designed as a special case of a multiband filter having two bands and parameters $C_1 = C_2 = 1$, $\theta_{1,1} = 0$, $\theta_{2,2} = \pi$. Recall that type-II filters (odd N) are not suitable for band-stop filters.

The procedure `firdes` in Program 9.1 implements the design of a general multiband FIR filter by impulse response truncation. The program accepts the desired filter order N and the filter specifications. The specifications are entered as a $K \times 3$ matrix. The kth row of the matrix describes the kth pass band. Its first two entries are the lower and upper band-edge frequencies, and its third entry is the gain of the band. The program accepts a third optional window, which should be ignored for now (we shall discuss this parameter in Section 9.3). The program can handle the three single-band filter types as special cases.

Example 9.7 Design a two-band filter with $N = 80$, $\theta_{1,1} = 0.2\pi$, $\theta_{2,1} = 0.4\pi$, $\theta_{1,2} = 0.7\pi$, $\theta_{2,2} = 0.8\pi$, $C_1 = 1$, $C_2 = 0.5$. Figure 9.7 shows the impulse response and the amplitude response of the resulting filter. As we see, the shape of the amplitude response approximates the desired low-pass characteristic, but the undesirable oscillations are again apparent. □

9.2.4 Differentiators

The frequency response of an ideal analog differentiator is

$$H^{\mathrm{F}}(\omega) = j\omega.$$

A digital differentiator can approximate an analog differentiator only over a limited range of frequencies, determined by the sampling rate. The ideal frequency response of a digital differentiator is therefore

$$H^{\mathrm{f}}(\theta) = j\frac{\theta}{T}, \quad -\pi \le \theta \le \pi.$$

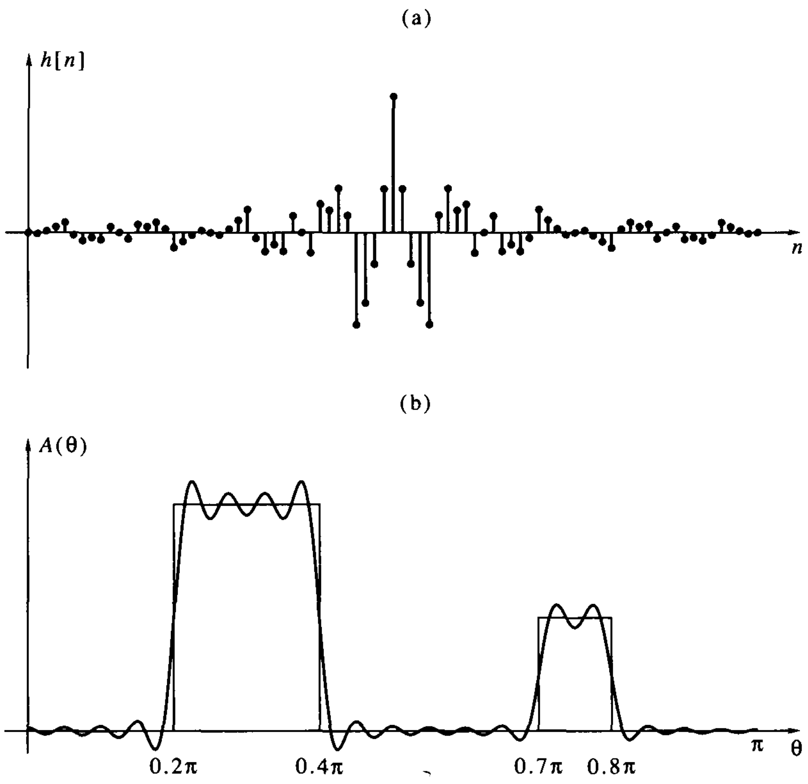

Figure 9.7 Multiband filter designed by the IRT method, of order $N = 80$: (a) impulse response; (b) frequency response.

The phase response of an ideal differentiator is 0.5π at all frequencies. Therefore, either a type-III or a type-IV filter are the natural choices in this case. We include a linear-phase term, so the desired response is

$$H_{\mathrm{d}}^{\mathrm{f}}(\theta) = \frac{\theta}{T}e^{j(0.5\pi - 0.5\theta N)}. \tag{9.36}$$

The impulse response is therefore

$$h_{\mathrm{d}}[n] = \frac{j}{2\pi T}\int_{-\pi}^{\pi} \theta e^{j\theta(n-0.5N)}d\theta = \begin{cases} \dfrac{(-1)^{(n-0.5N)}}{(n-0.5N)T}, & N \text{ even, } n \neq 0.5N, \\ 0, & N \text{ even, } n = 0.5N, \\ \dfrac{(-1)^{(n-0.5N+0.5)}}{\pi(n-0.5N)^2 T} & N \text{ odd.} \end{cases} \tag{9.37}$$

We observe that the magnitude of $h_{\mathrm{d}}[n]$ decays in proportion to $|n - 0.5N|^{-1}$ for even N, but in proportion to $|n - 0.5N|^{-2}$ for odd N. This has a significant influence on the oscillations of the frequency response, as is illustrated next.

Example 9.8 Design a differentiator by the IRT method, first of order $N = 16$, and then of order $N = 15$, with $T = 1$ in both cases. The coefficients are obtained by straightforward substitution in (9.37) and are shown in Figure 9.8. The amplitude responses are shown in Figure 9.9. As we see, the coefficients decay much more rapidly at the ends in the case of odd N than in the case of even N. Correspondingly, the amplitude response in the former case is much smoother, although the difference in the orders is only 1. The conclusion is that an odd-order (i.e., type-IV) filter is preferred to an even-order filter for FIR differentiators. □

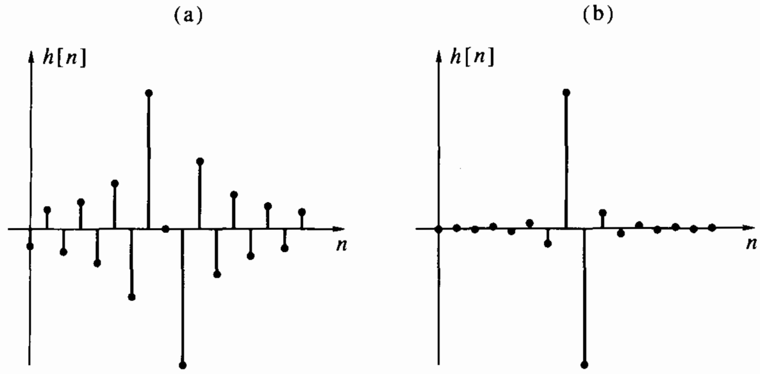

Figure 9.8 Impulse response of a differentiator designed by the IRT method: (a) $N = 16$; (b) $N = 15$.

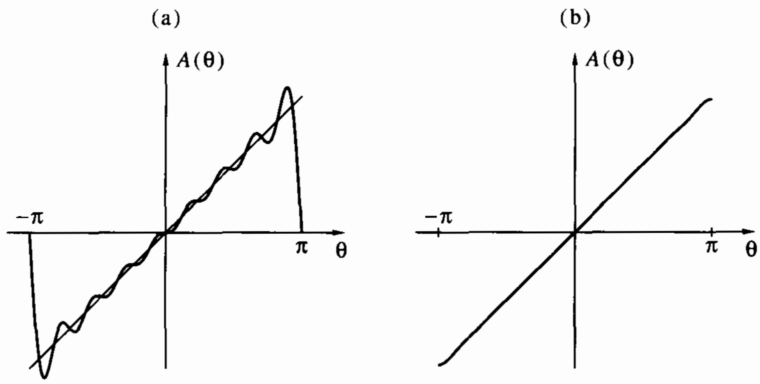

Figure 9.9 Amplitude response of a differentiator designed by the IRT method: (a) $N = 16$; (b) $N = 15$.

9.2.5 Hilbert Transformers

The frequency response of an ideal analog Hilbert transformer is

$$H^{\mathrm{F}}(\omega) = \begin{cases} -j, & \omega > 0, \\ 0, & \omega = 0, \\ j, & \omega < 0. \end{cases}$$

We provide a brief description of Hilbert transformers here, to encourage interest in them; however, we shall not explore their applications.[2]

Suppose that $x(t)$ is a real-valued signal; then its Fourier transform satisfies the conjugate symmetry property $X^{\mathrm{F}}(\omega) = \bar{X}^{\mathrm{F}}(-\omega)$. Let $y(t)$ be the signal obtained by passing $x(t)$ through an ideal Hilbert transformer. Then the Fourier transform of $y(t)$

is given by

$$Y^{\mathrm{F}}(\omega) = H^{\mathrm{F}}(\omega)X^{\mathrm{F}}(\omega) = \begin{cases} -jX^{\mathrm{F}}(\omega), & \omega > 0, \\ 0, & \omega = 0, \\ jX^{\mathrm{F}}(\omega), & \omega < 0. \end{cases}$$

Let $z(t)$ be the complex signal

$$z(t) = x(t) + jy(t). \tag{9.38}$$

Then the Fourier transform of $z(t)$ is

$$Z^{\mathrm{F}}(\omega) = [1 + jH^{\mathrm{F}}(\omega)]X^{\mathrm{F}}(\omega) = \begin{cases} 2X^{\mathrm{F}}(\omega), & \omega > 0, \\ X^{\mathrm{F}}(\omega), & \omega = 0, \\ 0, & \omega < 0. \end{cases}$$

The signal $z(t)$ is called the *analytic signal* of $x(t)$. Its frequency response is equal to that of $x(t)$ in the positive frequency range (except for a factor 2), and is zero in the negative frequency range. Thus, the spectrum of the analytic signal occupies only half the bandwidth and is free of the redundancy that exists in the spectrum of the original signal. It follows that the analytic signal can be sampled at half the sampling rate of the original signal. The analytic signal is complex: Its real part is equal to the original signal, and its imaginary part is obtained by passing the original signal through a Hilbert transformer.

As in the case of a differentiator, a digital Hilbert transformer can approximate an analog Hilbert transformer only over a limited range of frequencies, determined by the sampling rate. The ideal frequency response of a digital Hilbert transformer, including a linear-phase term, is

$$H_{\mathrm{d}}^{\mathrm{f}}(\theta) = \begin{cases} -e^{j(0.5\pi - 0.5\theta N)}, & \theta > 0, \\ 0, & \theta = 0, \\ e^{j(0.5\pi - 0.5\theta N)}, & \theta < 0. \end{cases} \tag{9.39}$$

As we see, $\phi_0 = 0.5\pi$ for a Hilbert transformer, so either type-III or type-IV filter is appropriate. The impulse response corresponding to (9.39) is given by

$$h_{\mathrm{d}}[n] = \begin{cases} \dfrac{1 - \cos[(n - 0.5N)\pi]}{\pi(n - 0.5N)}, & n \neq 0.5N, \\ 0, & n = 0.5N. \end{cases} \tag{9.40}$$

The procedure `diffhilb` in Program 9.2 implements the design of an FIR differentiator or an FIR Hilbert transformer by impulse response truncation. The program accepts the filter type (differentiator or Hilbert) and the desired filter order N. The program accepts a third optional window, which should be ignored for now (we shall discuss this parameter in Section 9.3).

Example 9.9 Design a Hilbert transformer by the IRT method, first of order $N = 16$, and then of order $N = 15$. The coefficients are obtained by substitution in (9.40) and shown in Figure 9.10. The amplitude responses are shown in Figure 9.11. As we see, the amplitude responses are similar near $\theta = 0$, but the response of the odd-order (type-IV) filter is better behaved near $\theta = \pm\pi$. □

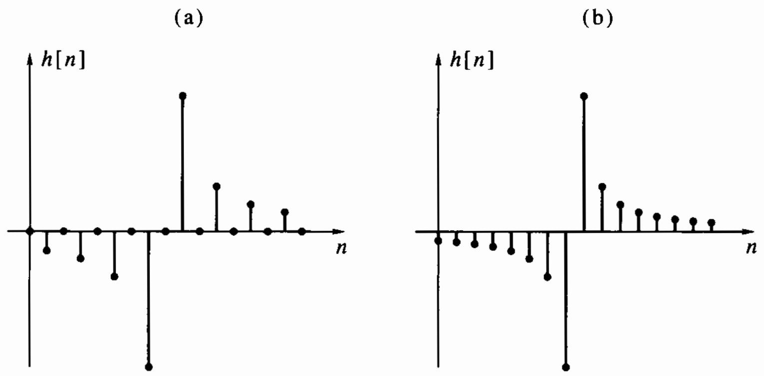

Figure 9.10 Impulse response of a Hilbert transformer designed by the IRT method: (a) $N = 16$; (b) $N = 15$.

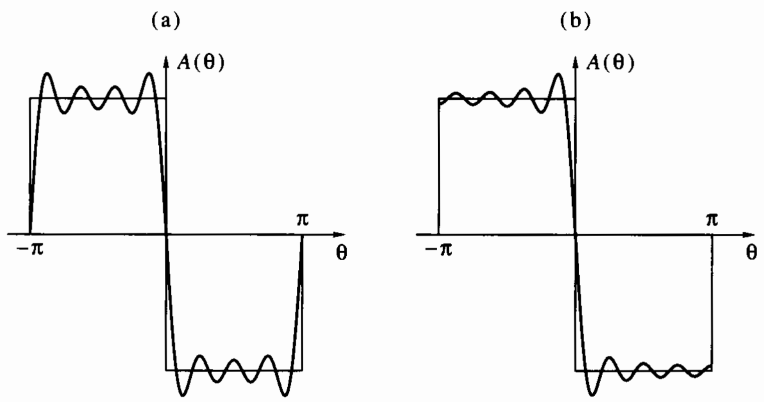

Figure 9.11 Amplitude response of a Hilbert transformer designed by the IRT method: (a) $N = 16$; (b) $N = 15$.

9.2.6 Optimality of the IRT Method

A measure of the error between the frequency response of the ideal filter $H_d^f(\theta)$ and that of the designed filter $H^f(\theta)$ is the integral of the square error, defined by

$$\varepsilon^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |H_d^f(\theta) - H^f(\theta)|^2 d\theta. \tag{9.41}$$

Filters designed by the impulse response truncation method are optimal in the sense of minimizing the integral of the square error, as we now prove.

Theorem 9.1 For a given ideal frequency response $H_d^f(\theta)$ and a given order N, the filter obtained by impulse response truncation has a minimum integral of square error among all causal FIR filters of the given order.

Proof Let $\{h[n],\ 0 \le n \le N\}$ be the coefficients of any causal FIR filter of order N. By Parseval's theorem, ε^2 can be expressed as

$$\varepsilon^2 = \sum_{n=-\infty}^{\infty} (h_d[n] - h[n])^2 = \sum_{n=-\infty}^{-1} h_d^2[n] + \sum_{n=N+1}^{\infty} h_d^2[n] + \sum_{n=0}^{N} (h_d[n] - h[n])^2. \tag{9.42}$$

The first two terms on the right side of (9.42) are determined only by $H_d(\theta)$ and are not affected by the choice of filter's parameters. The third term is nonnegative, so ε^2 is minimized if and only if this term is zero. But this happens if and only if

$$h[n] = h_d[n], \quad 0 \le n \le N,$$

which is precisely the IRT design criterion. □

Despite its optimality in the sense of integral-of-square-error, IRT is not considered a good method, because of the oscillatory nature of the frequency response near discontinuity points of the desired response $H_d^f(\theta)$. As N increases, the oscillations become more localized, therefore ε^2 decreases. However, the magnitude of the oscillations relative to the size of the discontinuity does not decrease to zero, but tends to a finite limit. This property, known as the *Gibbs phenomenon*, is explained next.

9.2.7 The Gibbs Phenomenon

Consider the amplitude response of an IRT low-pass filter having band-edge frequency 0.5π, for different orders N. Figure 9.12 illustrates the responses for $N = 10$ and $N = 40$. The filter of the larger order has a narrower transition band, as expected. However, the pass-band and stop-band tolerance parameters, δ_p and δ_s, are approximately the same for both, and their value is about 0.09. It turns out that the value 0.09 is characteristic of the IRT design method, and is almost independent of the filter's order and shape. This phenomenon is named in honor of J. W. Gibbs,[3] and is mathematically explained as follows.

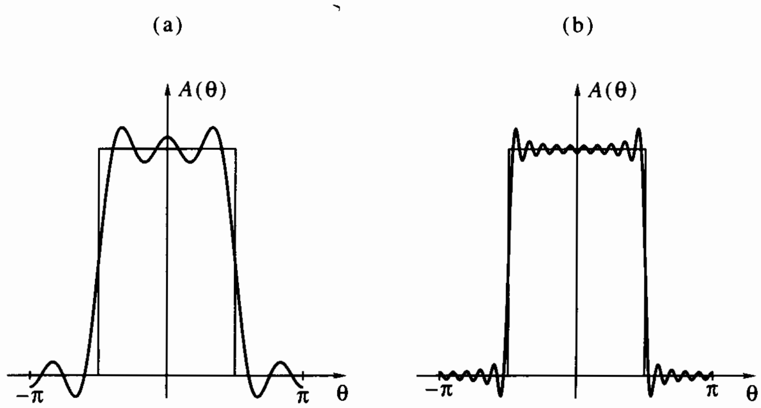

Figure 9.12 Amplitude response of a low-pass filter designed by the IRT method: (a) $N = 10$; (b) $N = 40$.

Let $A_d(\theta)$ be the amplitude response of an ideal low-pass filter with band-edge frequency θ_0. Denote for convenience $L = N + 1$ (the parameter L is the length of the sequence $h[n]$), and assume that

$$L \gg \frac{2\pi}{\theta_0} \quad \text{and} \quad L \gg \frac{2\pi}{\pi - \theta_0}. \tag{9.43}$$

The truncated impulse response is related to the ideal impulse response by

$$h[n] = h_d[n] w_r[n], \tag{9.44}$$

where $w_r[n]$ is a rectangular window of length L. Therefore, the amplitude response of the truncated filter is the convolution of $A_d(\theta)$ with the Dirichlet kernel $D(\theta, L)$, that is

$$A(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} A_d(\lambda)D(\theta-\lambda, L)d\lambda = \frac{1}{2\pi}\int_{-\theta_0}^{\theta_0} D(\theta-\lambda, L)d\lambda. \tag{9.45}$$

We ignore the linear-phase terms in both $H_d^f(\theta)$ and $H^f(\theta)$, since they do not affect the behavior of the amplitude response.

Let us now concentrate on values of θ in the vicinity of θ_0. Specifically, we require that $(L|\theta-\theta_0|)/(2\pi)$ be a small number, say, not greater than 5. Assume first that $\theta < \theta_0$ and compute the integral in (9.45) as follows:

$$\begin{aligned}\frac{1}{2\pi}\int_{-\theta_0}^{\theta_0} D(\theta-\lambda, L)d\lambda &= \frac{1}{2\pi}\int_{-\theta_0}^{\theta} D(\theta-\lambda, L)d\lambda + \frac{1}{2\pi}\int_{\theta}^{\theta_0} D(\theta-\lambda, L)d\lambda \\ &= \frac{1}{2\pi}\int_{0}^{\theta+\theta_0} D(\mu, L)d\mu + \frac{1}{L\pi}\int_{0}^{0.5L(\theta_0-\theta)} D\left(\frac{2\mu}{L}, L\right)d\mu,\end{aligned} \tag{9.46}$$

where we substituted $\mu = \theta - \lambda$ in the first integral, and $\mu = 0.5(\lambda - \theta)L$ in the second (we also used the symmetry of the Dirichlet kernel). Since θ is close to θ_0, and by assumption (9.43), the first integral is approximately equal to the integral of the Dirichlet kernel from 0 to π, the value of which is π. In the second integral, the argument of the Dirichlet kernel is a small number, so

$$D\left(\frac{2\mu}{L}, L\right) = \frac{\sin\mu}{\sin(\mu/L)} \approx \frac{L\sin\mu}{\mu}. \tag{9.47}$$

Substitution of these approximations in (9.46) gives

$$A(\theta) \approx 0.5 + \frac{1}{\pi}\int_{0}^{0.5L(\theta_0-\theta)} \frac{\sin\mu}{\mu}d\mu. \tag{9.48}$$

The function

$$\mathrm{Si}(x) = \int_0^x \frac{\sin\mu}{\mu}d\mu \tag{9.49}$$

is called the *sine integral*. Using this function in (9.48), we get

$$A(\theta) \approx 0.5 + \frac{1}{\pi}\mathrm{Si}[0.5(\theta_0-\theta)L], \quad \theta < \theta_0. \tag{9.50}$$

Using a similar derivation, we obtain

$$A(\theta) \approx 0.5 - \frac{1}{\pi}\mathrm{Si}[0.5(\theta-\theta_0)L], \quad \theta > \theta_0. \tag{9.51}$$

The sine integral is shown in Figure 9.13. The salient properties of this function are as follows:

1. For small values of x, $\mathrm{Si}(x) \approx x$.
2. $\lim_{x\to\infty} \mathrm{Si}(x) = 0.5\pi$.
3. The global maximum is at $x = \pi$, and its value is approximately $1.179 \times 0.5\pi$.
4. After the function has reached its global maximum, it oscillates around its limiting value in a decaying manner.

The properties of the sine integral explain the Gibbs phenomenon. As we see from (9.50) and (9.51), $A(\theta)$ takes the value 0.5 at $\theta = \theta_0$. At $\theta = \theta_0 - 2\pi/L$ it takes the value $0.5 + \pi^{-1}\mathrm{Si}(\pi)$, which is approximately 1.0895. Similarly, at $\theta = \theta_0 + 2\pi/L$ it takes the value $0.5 - \pi^{-1}\mathrm{Si}(\pi)$, which is approximately -0.0895. Figure 9.14 shows the details of the amplitude function $A(\theta)$ in the vicinity of θ_0. As we see, both the

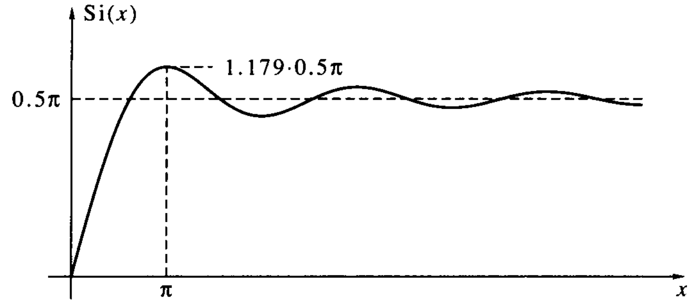

Figure 9.13 The sine integral function.

stop-band attenuation and the pass-band ripple are about 0.09, not depending on the filter's order. The width of the transition band is smaller than $4\pi/L$, so it decreases as the filter's order increases. The upper limit on the width of the transition band is the width of the main lobe of the Dirichlet kernel.

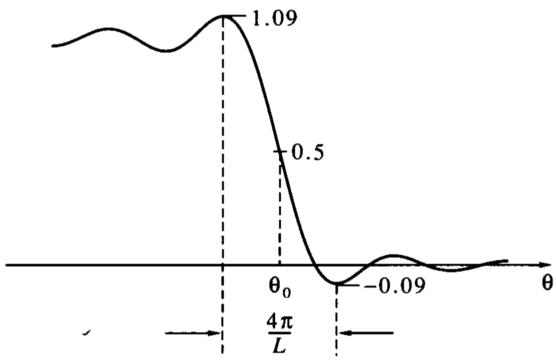

Figure 9.14 The Gibbs phenomenon.

The practical implication of the preceding discussion is that the IRT design method is suitable only for filters whose tolerances are not smaller than 0.09, or about 21 dB in the stop band and 0.75 dB in the pass band. Practical filters are almost always required to have smaller tolerances, so the IRT method is not suitable for their design. In the next section we shall see how the Gibbs phenomenon can be mitigated with the aid of windows.

9.3 FIR Filter Design Using Windows

In Chapter 6 we used windows for reducing the side-lobe interference resulting from truncating an infinite-duration signal to a finite length. We recall that windowing performs convolution in the frequency domain; hence it can be used to attenuate the Gibbs oscillations seen in the amplitude response of FIR filters. Windowing is applied to an impulse response of an FIR filter in the same way as to a finite-duration signal. Specifically, let $w[n]$ denote a window of length $L = N + 1$. The window design method of FIR filters is then as follows:

Windowed FIR filter design procedure

1. Define the ideal frequency response $H_d^f(\theta)$ as in the IRT method; see (9.25). For each pair $\{\theta_p, \theta_s\}$ (depending on the individual bands), take the corresponding band-edge frequency of the ideal response as the midpoint, that is, $0.5(\theta_p + \theta_s)$.
2. Obtain the ideal impulse response $h_d[n]$ as in the IRT method.
3. Compute the coefficients of the filter by

$$h[n] = \begin{cases} w[n]h_d[n], & 0 \le n \le N, \\ 0, & \text{otherwise.} \end{cases} \tag{9.52}$$

Windows are always symmetric, that is, they satisfy

$$w[n] = w[N - n],$$

for either even or odd N. Therefore, the windowed impulse response $h[n]$ is either symmetric or antisymmetric, in agreement with $h_d[n]$. It follows that the FIR filter has linear phase, and the window does not affect its type.

The frequency response of the FIR filter is the convolution of that of the ideal filter with the kernel function of the window, that is,

$$H^f(\theta) = \frac{1}{2\pi}\{W^f * H_d^f\}(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} W^f(\lambda)H_d^f(\theta - \lambda)d\lambda. \tag{9.53}$$

For a given ideal frequency response $H_d^f(\theta)$, the properties of the actual response—stop-band attenuation, pass-band ripple, and transition bandwidth—depend on the chosen window. As we learned in Section 6.2, a window is characterized by two primary parameters:

1. The width of the main lobe, equal to $4\pi\eta/L$, where η is a proportionality constant characteristic of the window. For example, recall from Section 6.2 that $\eta = 1$ for the rectangular window, $\eta = 2$ for the Bartlett, Hann, and Hamming windows, etc.
2. The magnitude of the highest side lobe relative to the main lobe.

The effect of the window on the amplitude response $A(\theta)$ can be analyzed similarly to the analysis of the Gibbs phenomenon. Let $A_w(\theta)$ be the amplitude function of the kernel $W^f(\theta)$ and define

$$\Gamma_w(x) = \frac{1}{L}\int_0^x A_w\left(\frac{2\mu}{L}\right)d\mu. \tag{9.54}$$

Suppose that $A_d(\theta)$ is discontinuous at θ_0 such that $A_d(\theta_0^-) = 1$, $A_d(\theta_0^+) = 0$. Then we have, using the same derivation that has led to (9.50), (9.51),

$$A(\theta) \approx \begin{cases} 0.5 + \frac{1}{\pi}\Gamma_w[0.5(\theta_0 - \theta)L], & \theta < \theta_0, \\ 0.5 - \frac{1}{\pi}\Gamma_w[0.5(\theta - \theta_0)L], & \theta > \theta_0. \end{cases} \tag{9.55}$$

Let us examine the shape of (9.55) for the windows studied in Section 6.2. Consider first the Bartlett window and recall that the amplitude response of its kernel is nonnegative [cf. (6.11)]. Therefore, the function $\Gamma_w(x)$ is monotone nondecreasing in this case. This implies that $A(\theta)$ is monotone in the vicinity of the discontinuity point, as illustrated in Figure 9.15. Since $A(\theta)$ is monotone, the transition bandwidth and the tolerance parameters are not well defined as separate entities, only jointly. Nevertheless, it is common to regard $8\pi/L$ as the transition bandwidth of the filter and, correspondingly, the values $\delta_s = \delta_p = 0.05$ as the tolerance parameters; see Figure 9.15. The respective dB values are

$$A_s \approx 26\,\text{dB}, \quad A_p \approx 0.45\,\text{dB}.$$

As we see, the Bartlett window offers only a small improvement over the rectangular window in terms of the tolerance parameters, whereas its transition bandwidth is twice as large. Therefore, the use of the Bartlett window for FIR filter design is rare. It can be recommended only if a monotone amplitude response is desired and if the tolerance specifications are not tight.

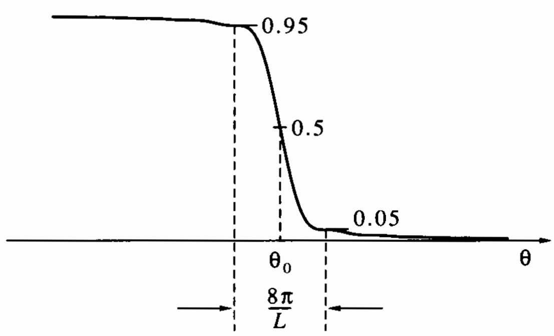

Figure 9.15 The amplitude response of an FIR filter based on the Bartlett window near a discontinuity point.

Let us now proceed to the Hann and Hamming windows, whose kernel functions are given by (6.13) and (6.17), respectively. The details of the amplitude functions of the corresponding FIR filters are shown in Figures 9.16 and 9.17. In these figures, as in the ones to follow, we omitted the transition band for better display of the pass-band and stop-band ripples. As we see, the transition bandwidth is slightly less than $8\pi/L$ for both windows. The tolerance parameters of the Hann window are $\delta_p = \delta_s = 0.0063$, and the respective dB values are

$$A_s \approx 44\,\text{dB}, \quad A_p \approx 0.055\,\text{dB}.$$

The tolerance parameters of the Hamming window are $\delta_p = \delta_s = 0.0022$, and the respective dB values are

$$A_s \approx 53\,\text{dB}, \quad A_p \approx 0.019\,\text{dB}.$$

The Hamming window has the advantage of smaller tolerances, whereas the Hann window has the advantage of faster decay of the ripple away from the discontinuity.

The details of the amplitude function of an FIR filter based on the Blackman window are shown in Figure 9.18. The kernel function of the Blackman window, given in (6.19), has a peculiar property: Its first side lobe is nonnegative, but subsequent side lobes have alternating signs. As a result, the transition bandwidth is $12\pi/L$, but the extremal points are at $\theta_0 \pm 8\pi/L$, as shown in Figure 9.18. The tolerance parameters of the Blackman window are $\delta_p = \delta_s = 0.0002$, and the respective dB values are

$$A_s \approx 74\,\text{dB}, \quad A_p \approx 0.0017\,\text{dB}.$$

Finally, let us examine filters based on the Kaiser window. The properties of a Kaiser window—main-lobe width and side-lobe level—can be controlled by the window

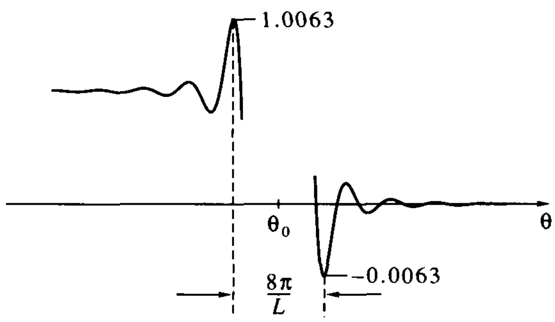

Figure 9.16 The amplitude response of an FIR filter based on the Hann window near a discontinuity point.

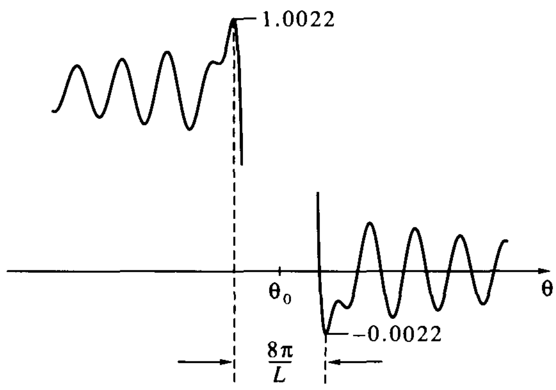

Figure 9.17 The amplitude response of an FIR filter based on the Hamming window near a discontinuity point.

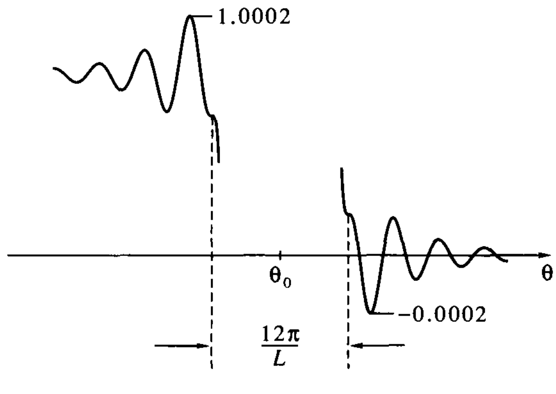

Figure 9.18 The amplitude response of an FIR filter based on the Blackman window near a discontinuity point.

parameter α. In the context of filter design, α is used for controlling the transition bandwidth and the tolerance parameters. Figure 9.19 shows the details of the amplitude response of a filter based on a Kaiser window with $\alpha = 10$. The transition bandwidth is about $13\pi/L$ in this case. The tolerance parameters are $\delta_p = \delta_s = 0.00001$, and the corresponding dB values are

$$A_s \approx 100\,\text{dB}, \quad A_p \approx 0.000087\,\text{dB}.$$

Table 9.2 summarizes the main properties of the windows we have described.

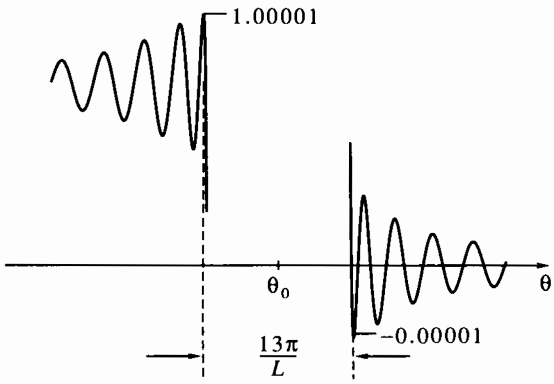

Figure 9.19 The amplitude response of an FIR filter based on the Kaiser window, with $\alpha = 10$, near a discontinuity point.

Window	eqn.	main-lobe width	side-lobe level, dB	δ_p, δ_s	pass-band ripple A_p [dB]	stop-band attn. A_s [dB]
rectangular	(6.2)	$4\pi/L$	-13.5	0.09	0.75	21
Bartlett	(6.10)	$8\pi/L$	-27	0.05	0.45	26
Hann	(6.14)	$8\pi/L$	-32	0.0063	0.055	44
Hamming	(6.16)	$8\pi/L$	-43	0.0022	0.019	53
Blackman	(6.18)	$12\pi/L$	-57	0.0002	0.0017	74
Kaiser	(6.21)	Depend on α				

Table 9.2 Windows and their properties.

For FIR filter design using the Kaiser window, there exist empirical formulas that give the parameter α and the order N as a function of the transition bandwidth $|\theta_p - \theta_s|$ and the tolerance parameters δ_p, δ_s, as follows:

$$A = -20\log_{10}(\min\{\delta_p, \delta_s\}), \tag{9.56a}$$

$$\alpha = \begin{cases} 0.1102(A - 8.7), & A > 50, \\ 0.5842(A - 21)^{0.4} + 0.07886(A - 21), & 21 < A \le 50, \\ 0, & A \le 21, \end{cases} \tag{9.56b}$$

$$N = \frac{A - 7.95}{2.285|\theta_p - \theta_s|}. \tag{9.56c}$$

Note that if $A \le 21$ dB, a rectangular window is sufficient; therefore $\alpha = 0$ in this case. The order N should be rounded up to an even or odd integer, according to the desired filter type. Since the formulas are empirical, the filter is not guaranteed to meet the specifications. It is therefore necessary to check the response of the resulting filter and, if found unsatisfactory, to increase N or α (or both) and repeat the design.

The Kaiser window is better than the other windows we have mentioned, since for given tolerance parameters its transition band is always narrower. For this reason, and because of the convenience of controlling the filter tolerances via the parameter α, the Kaiser window has become the preferred choice in window-based filter design.

Program 9.1 can be used for window design of multiband FIR filters, by multiplying the IRT filter by a window. The window is entered as a third optional parameter, in which case it must have length equal to that of the filter. The same applies to Program 9.2 for differentiators and Hilbert transformers design.

The procedures `firkais`, `kaispar`, `verspec` in Programs 9.3, 9.4, and 9.5 implement Kaiser window FIR filter design according to given specifications. The program is limited to the six basic filter types: low pass, high pass, band pass, band stop, differentiator, and Hilbert transformer. The program accepts the filter type, the parity of the order (even or odd), the band-edge frequencies, and the tolerance parameters. It operates as follows:

1. Initial guesses for N and α are obtained by calling `kaispar`. This routine implements Kaiser's formulas (9.56). The requested parity is honored, except when the filter type is high pass or band stop, whereupon an even-order filter is forced regardless of the input parameter.
2. The filter is designed by calling `firdes` or `diffhilb`, according to the desired type.
3. The filter is tested against the specifications, by calling `verspec`. The test can result in three possible outcomes: 0 means that the filter meets the specifications; 1 means that N needs to be increased; 2 means that α needs to be increased. In the first case the program exits; in the second it increases N by 2 (to preserve the parity) and repeats the procedure; in the third it increases α by 0.05 and repeats the procedure.
4. The routine `verspec` computes the magnitude response on intervals near the edges of the bands, by calling `frqresp`. Each interval starts at a band-edge frequency, stretches over half the transition band in a direction away from the transition region, and contains 100 test points. If any band-edge frequency does not meet the specifications, the output is set to 1 to indicate a need for order increase. If there is a deviation from the specifications any interval, the output is set to 2 to indicate a need for increasing α. Otherwise the output is set to 0.

9.4 FIR Filter Design Examples

We now illustrate the window-based FIR design procedure by a few examples.

Example 9.10 Design a type-I low-pass filter according to the specifications:

$$\theta_p = 0.2\pi, \quad \theta_s = 0.3\pi, \quad \delta_p = \delta_s = 0.01.$$

The required stop-band attenuation is 40 dB, so the Hann window should be adequate. The band-edge frequency of the ideal response is the midpoint between θ_p and θ_s, that is, 0.25π. The transition bandwidth is $8\pi/(N+1) = 0.1\pi$, so the filter's order is chosen as $N = 80$. However, as we recall from Section 6.2, the Hann window has the property $w_{hn}[0] = w_{hn}[N] = 0$, so the actual order is only $N = 78$.

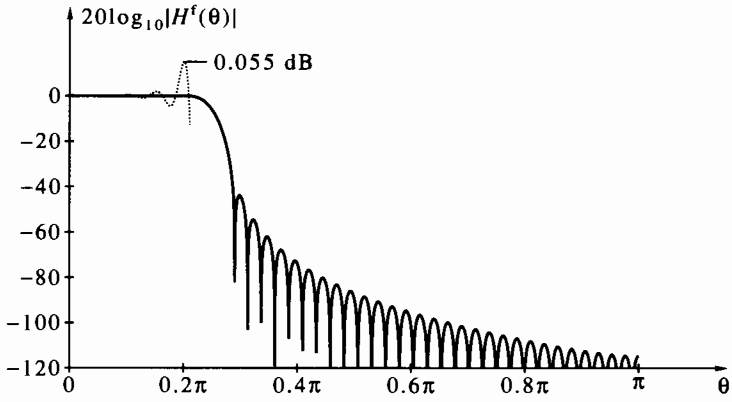

Figure 9.20 The magnitude response of the filter based on the Hann window in Example 9.10.

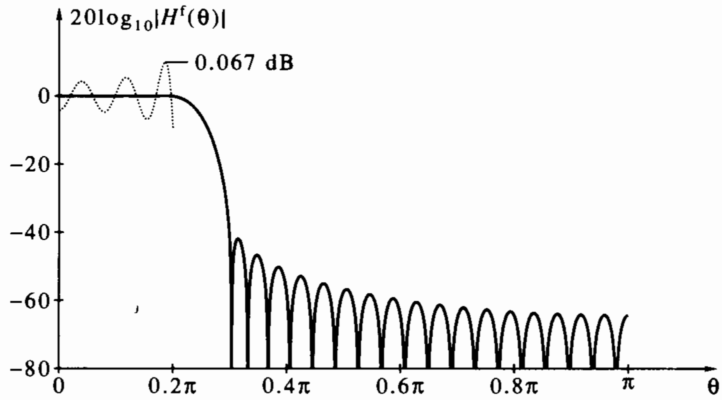

Figure 9.21 The magnitude response of the filter based on the Kaiser window in Example 9.10.

The magnitude response of the filter is shown in Figure 9.20. The figure also shows, as a dotted line, the details of the pass-band response at an expanded scale. For comparison, we design a filter based on a Kaiser window that meets the same specification. According to (9.56), the window parameter is $\alpha = 3.395$, and the order is $N = 46$. The magnitude response of the Kaiser window filter is shown in Figure 9.21. As we see, both filters meet the specifications, and the filter based on the Kaiser window does so at a considerably smaller order. However, the ripple decays more rapidly in the filter based on the Hann window. □

Example 9.11 Design a type-II band-pass filter according to the specifications:

$$\theta_{s,1} = 0.1\pi, \quad \theta_{p,1} = 0.25\pi, \quad \theta_{p,2} = 0.6\pi, \quad \theta_{s,2} = 0.8\pi,$$

$$\delta_{s,1} = 0.005, \quad \delta_{s,2} = 0.0025, \quad \delta_p = 0.005.$$

The stop-band attenuation is determined by the smaller of the two tolerances, giving 52 dB. Therefore, a Hamming window is adequate here. The design transition bandwidth is the narrower of the two specified widths, so $8\pi/(N+1) = 0.15\pi$, which gives $N = 53$. Since the actual filter will have its two transition bands equal, we take the band-edge frequencies of the ideal response as 0.175π and 0.675π.

The magnitude response of the filter is shown in Figure 9.22. For comparison, we design a filter based on a Kaiser window that meets the same specification. According to (9.56), the window parameter is $\alpha = 4.772$ and the order is $N = 41$. The magnitude response of the Kaiser window filter is shown in Figure 9.23. As we see, both filters meet the specifications, and the filter based on the Kaiser window does so at a smaller order. In general, the magnitude responses of the two filters are quite similar. □

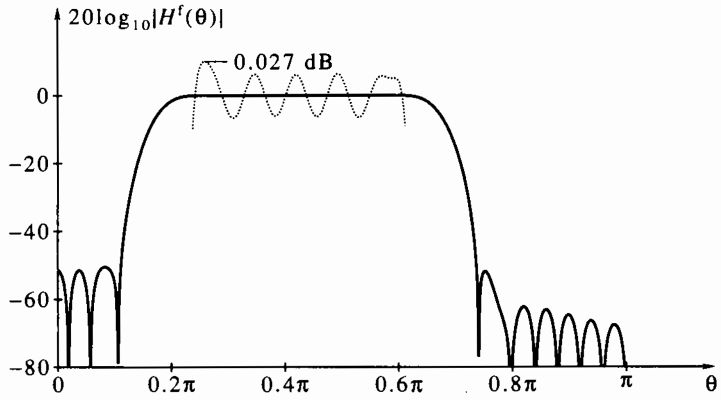

Figure 9.22 The magnitude response of the filter based on the Hamming window in Example 9.11.

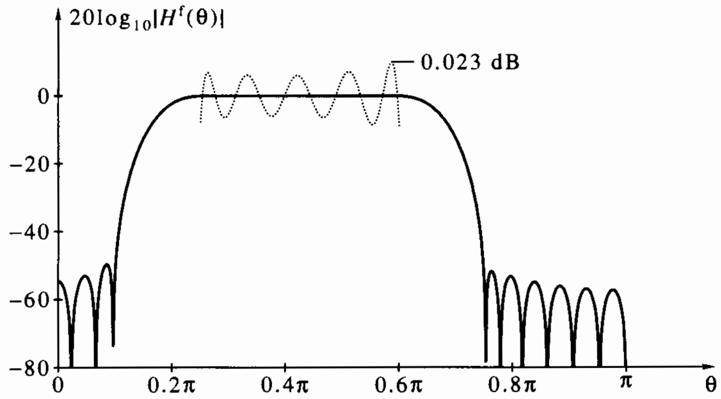

Figure 9.23 The magnitude response of the filter based on the Kaiser window in Example 9.11.

Example 9.12 Design a type-I high-pass filter according to the specifications:

$$\theta_s = 0.7\pi, \quad \theta_p = 0.8\pi, \quad \delta_s = 0.0002, \quad \delta_p = 0.001.$$

The required stop-band attenuation is 74 dB, so the Blackman window should be adequate, although marginal. The transition bandwidth is $12\pi/(N+1) = 0.1\pi$, so we choose the filter's order as $N = 120$. Like the Hann window, the Blackman window has the property $w_b[0] = w_b[N] = 0$, so the actual filter order is only $N = 118$. The band-edge frequency of the ideal response is 0.75π.

The magnitude response of the filter is shown in Figure 9.24. For comparison, we design a Kaiser window filter that meets the same specifications. According to (9.56), the window parameter is $\alpha = 7.196$ and the order is $N = 92$. The magnitude response of the Kaiser window filter is shown in Figure 9.25. As we see, both filters meet the specifications, and the filter based on the Kaiser window does so at a smaller order. However, the ripple decays more rapidly in the filter based on the Blackman window.□

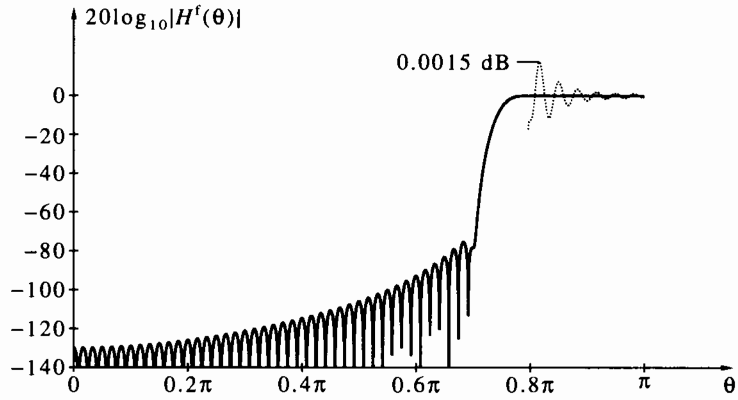

Figure 9.24 The magnitude response of the filter based on the Blackman window in Example 9.12.

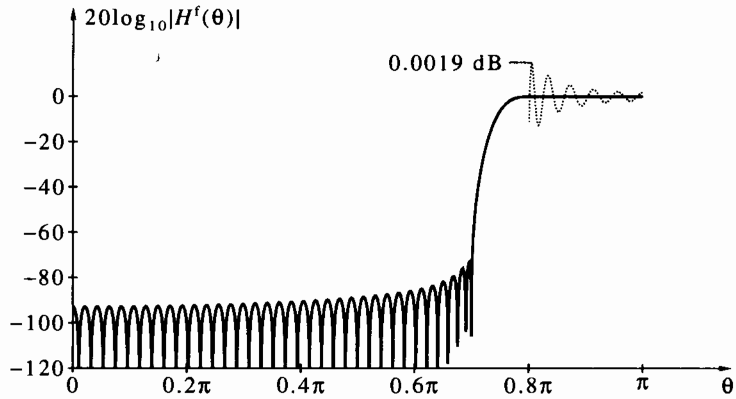

Figure 9.25 The magnitude response of the filter based on the Kaiser window in Example 9.12.

Example 9.13 Design a type-I band-stop filter according to the specifications:

$$\theta_{p,1} = 0.3\pi, \quad \theta_{s,1} = 0.4\pi \quad, \theta_{s,2} = 0.6\pi, \quad \theta_{p,2} = 0.7\pi,$$

$$\delta_s = 0.00001, \quad \delta_{p,1} = \delta_{p,2} = 0.0002.$$

The stop-band attenuation is 100 dB, so only a Kaiser window is adequate in this case. According to (9.56), the window parameter is $\alpha = 10.061$, and the order is $N = 130$. The magnitude response of the filter is shown in Figure 9.26. □

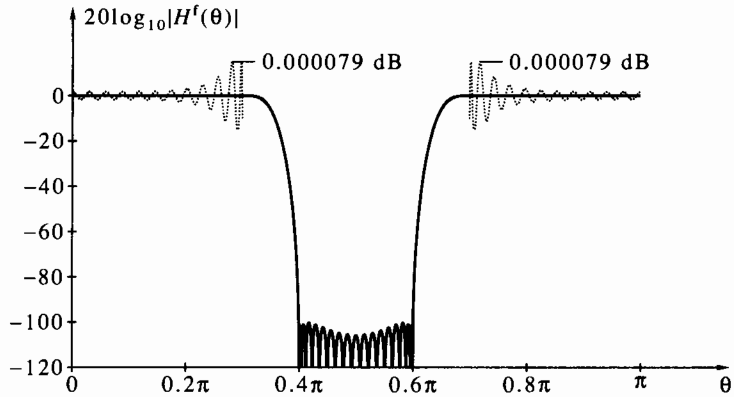

Figure 9.26 The magnitude response of the filter based on the Kaiser window in Example 9.13.

Example 9.14 A *digital raised-cosine* filter has an ideal frequency response like the one shown in Example 3.9. In particular, if $f_0T = 0.5$ (where T is the sampling interval), then

$$H^f(\theta) = \begin{cases} 2, & |\theta| \le \theta_1, \\ 1 + \cos\left(\dfrac{|\theta| - \theta_1}{\alpha}\right), & \theta_1 < |\theta| \le \theta_2, \\ 0, & |\theta| > \theta_2, \end{cases} \tag{9.57}$$

where $0 < \alpha < 1$, and

$$\theta_1 = 0.5(1-\alpha)\pi, \quad \theta_2 = 0.5(1+\alpha)\pi. \tag{9.58}$$

In this special case, the filter is zero-phase, half-band, as defined in Problem 8.11, and its frequency response is as shown in Figure 8.11. Its noncausal impulse response is obtained from (3.22) as

$$h_d[n] = \begin{cases} \dfrac{\sin(0.5\pi n)\cos(0.5\pi\alpha n)}{0.5\pi n[1-(\alpha n)^2]}, & n \ne 0, \\ 1, & n = 0. \end{cases} \tag{9.59}$$

As we see, the even-index coefficients are zero except $h_d[0]$, as necessary for half-band filters [cf. (8.74)].

A causal FIR filter can be obtained from $h_d[n]$ by shifting it $2M$ points to the right, truncating to $0 \le n \le 4M$, and applying a window. The resulting filter will have a frequency response different from (9.57), but it will still be a half-band filter, because (8.74) continues to hold when multiplied by a window. The FIR filter will have zero coefficients in positions 0 and $4M$. Therefore, we can reduce the order to $4M-2$ if we define

$$h[n] = \begin{cases} h_d[n-2M+1]w[n], & 0 \le n \le 4M-2, \\ 0, & \text{otherwise}, \end{cases} \tag{9.60}$$

where $w[n]$ is any window of length $4M-1$.

Consider a specific example. Let $M = 10$, so the filter's order is $N = 38$. Let the bandwidth excess be $\alpha = 0.4$. Figure 9.27 shows the magnitude response of a rectangularly windowed FIR filter. This result is perhaps surprising: Instead of the familiar 21 dB stop-band attenuation, which we would expect from a rectangular window, we get attenuation of nearly 60 dB. However, the value 21 dB results from points

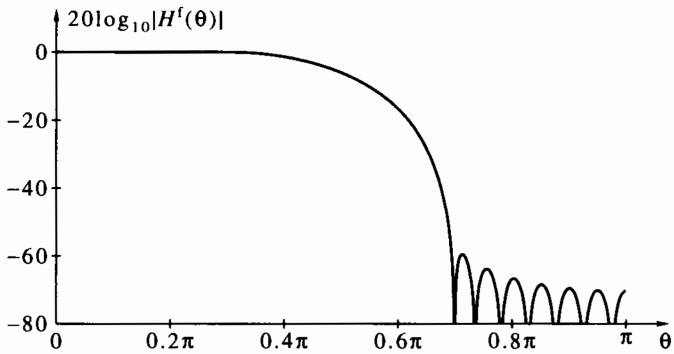

Figure 9.27 The magnitude response of a digital, raised-cosine, half-band filter; $N = 38$, rectangular window.

of discontinuity of the ideal frequency response. The raised-cosine response (9.57) is continuous and has a continuous first derivative at all points; only its second derivative is discontinuous at $\theta = 0.5(1 \pm \alpha)\pi$. The Gibbs phenomenon occurs whenever the frequency response or one of its derivatives is discontinuous, but such that the higher the order of the discontinuous derivative, the less pronounced the ripple around the discontinuity point. Another way to see this is from the rate of decay of the impulse response coefficients as a function of n. For a discontinuous function, the rate is $|n|^{-1}$, for discontinuous first derivative the rate is $|n|^{-2}$, and for discontinuous second derivative it is $|n|^{-3}$. High decay rate acts, in a sense, like a window to mitigate the Gibbs effect. As we see from (9.59), the rate of decay is $|n|^{-3}$ for the raised-cosine function.

In the range $0 \le \theta \le 0.7\pi$, the magnitude response of the FIR filter is very close to that of the ideal filter; it is plotted in Figure 9.27, but is not visible, because the two graphs coincide.

The conclusion from the preceding discussion is that a raised-cosine, half-band filter can be designed with a rectangular window if the required stop-band attenuation is 60 dB or less. If the attenuation is higher, another window should be employed. A window would not affect the half-band property, but would increase the transition band. It is recommended, in this case, to choose a Kaiser window and try different values of α, until the specifications are met. □

9.5 Least-Squares Design of FIR Filters*

FIR filter design based on windows is simple and robust, yielding filters with good performance. However, in two respects it is not optimal:

1. The resulting pass-band and stop-band parameters δ_p, δ_s are equal (or almost so), even if we do not require them to be equal a priori. Often, the specification is more stringent in the stop band than in the pass band (i.e., $\delta_s < \delta_p$), so we obtain an unnecessarily high accuracy in the pass band.
2. The ripple of windows is not uniform in either the pass band or the stop band, but decays as we move away from the discontinuity points, according to the side-lobe pattern of the window (except for the Dolph window, which is rarely used for filter design). By allowing more freedom in the ripple behavior, we may be able to reduce the filter's order, thereby reducing its complexity.

In this section we present a design method that approximates the desired amplitude response $A_d(\theta)$ by a linear-phase FIR amplitude response function $A(\theta)$ according to an optimality criterion. The optimality criterion is a modified version of the frequency-domain integral of square error (9.41): To each frequency θ we assign a *weight* $V(\theta)$. The weight is a nonnegative number, reflecting the relative importance of the error between $A_d(\theta)$ and $A(\theta)$ at the particular frequency θ. For example, if we wish to design a low-pass filter, we can assign a constant weight δ_p^{-1} to all frequencies in the pass band, a different constant weight δ_s^{-1} to all frequencies in the stop band, and zero weight to all frequencies in the transition band. Then, the smaller the tolerance, the larger the weight given to the error in the corresponding band.

After choosing a weight function $V(\theta)$, we define the weighted frequency-domain error as

$$E(\theta) = V(\theta)[A_d(\theta) - A(\theta)], \tag{9.61}$$

and the integral of weighted square frequency-domain error as

$$\varepsilon^2 = \int_0^{\pi} E^2(\theta)d\theta \tag{9.62}$$

(because of symmetry, it is sufficient to integrate only over the positive frequencies). We assume that order N of the filter to be designed and its type are known. Under these assumptions, we can express $A(\theta)$ as in (9.23):

$$A(\theta) = F(\theta)G(\theta), \quad G(\theta) = \sum_{k=0}^{K} g[k]\cos(\theta k), \tag{9.63}$$

where K and $F(\theta)$ depend on the filter type, as given in Table 9.1. Substitution of (9.63) in (9.61) gives

$$E(\theta) = V(\theta)\left[A_d(\theta) - F(\theta)\sum_{k=0}^{K} g[k]\cos(\theta k)\right]. \tag{9.64}$$

Designing the FIR filter now reduces to determining the coefficient values $g[k]$ that would minimize ε^2 in (9.62).

To find the optimal coefficients, we differentiate ε^2 with respect to all $g[k]$ and equate the derivatives to zero. In general, this gives a local extremum point. However, only one extremum is possible in this case: the global minimum of the function.[4] Differentiation of (9.62) thus yields

$$\frac{\partial \varepsilon^2}{\partial g[m]} = \int_0^{\pi} 2E(\theta)\frac{\partial E(\theta)}{\partial g[m]}d\theta, \tag{9.65}$$

where

$$\frac{\partial E(\theta)}{\partial g[m]} = -V(\theta)F(\theta)\cos(\theta m). \tag{9.66}$$

Substitution of (9.64) and (9.66) in (9.65) gives

$$-2\int_0^{\pi} V^2(\theta)F(\theta)\cos(\theta m)\left[A_d(\theta) - F(\theta)\sum_{k=0}^{K} g[k]\cos(\theta k)\right]d\theta = 0, \quad 0 \le m \le K. \tag{9.67}$$

We therefore get a set of $K+1$ equations in $K+1$ unknowns

$$\sum_{k=0}^{K} c_{m,k}\, g[k] = d_m, \quad 0 \le m \le K, \tag{9.68}$$

where

$$c_{m,k} = \int_0^{\pi} V^2(\theta)F^2(\theta)\cos(\theta m)\cos(\theta k)d\theta, \tag{9.69a}$$

$$d_m = \int_0^{\pi} V^2(\theta)F(\theta)A_{\mathrm{d}}(\theta)\cos(\theta m)d\theta. \tag{9.69b}$$

Solution of the linear equations gives the desired coefficients $g[k]$.

The least-squares design procedure for FIR filters is then as follows:

Least-squares FIR filter design procedure

1. Choose the filter order N and its type; these determine K and $F(\theta)$.
2. Choose the weight function $V(\theta)$. As we have said, a common choice for $V(\theta)$ is

$$V(\theta) = \begin{cases} (\delta_{\mathrm{p},l})^{-1}, & \theta \text{ in the } l\text{th pass band,} \\ (\delta_{\mathrm{s},l})^{-1}, & \theta \text{ in the } l\text{th stop band,} \\ 0, & \theta \text{ in any of the transition bands.} \end{cases} \tag{9.70}$$

3. Compute the coefficients $c_{m,k}$ and d_m as given by (9.69). Since $A_{\mathrm{d}}(\theta)$ and $V(\theta)$ are typically piecewise constant, often the integrals can be computed in a closed form. However, it is more common to approximate these integrals by sums. The frequency range $[0, \pi]$ is sampled uniformly, at a number of points I proportional to the filter's order (say $I = 16N$ points). Then $c_{m,k}$ and d_m are approximated by

$$c_{m,k} \approx \frac{\pi}{I}\sum_{i=0}^{I-1} V^2(\theta_i)F^2(\theta_i)\cos(\theta_i m)\cos(\theta_i k), \tag{9.71a}$$

$$d_m \approx \frac{\pi}{I}\sum_{i=0}^{I-1} V^2(\theta_i)F(\theta_i)A_{\mathrm{d}}(\theta_i)\cos(\theta_i m). \tag{9.71b}$$

4. Solve the set of equations (9.68).
5. Compute the coefficients $h[n]$ of the filter, using one of (9.5), (9.9), (9.15), (9.20), according to the filter type.

The procedures `firls` and `firlsaux` in Programs 9.6 and 9.7 implement least-squares design of the four basic frequency responses in MATLAB. The program `firls` accepts the filter's order N, the desired frequency response characteristic (LP, HP, BP, BS), the band-edge frequencies, and the ripple tolerances. It uses piecewise-constant weighting and samples the frequency range $[0, \pi]$ at $16N$ points. It then implements the procedure in a straightforward manner. The program `firlsaux` is an auxiliary program used by `firls`.

Least-squares design often leads to filters of orders lower than those of filters based on windows, especially when there is a large difference between the tolerances δ_{p} and δ_{s}. This design method is highly flexible. The amplitude function $A(\theta)$ can have an almost arbitrary shape, and is not required to be expressed by a mathematical formula: We need only its numerical values on a sufficiently dense grid. There is much flexibility in choosing the weighting function, allowing much freedom in shaping the frequency response. The computational requirements are modest, and programming the method is straightforward. The main drawbacks of the least-squares method are as follows:

1. Meeting the specifications is not guaranteed a priori, and trial and error is often required. To help the design procedure meet the specifications at the desired band-edge frequencies, it is often advantageous to set the transition bands

slightly narrower than needed (e.g., increasing θ_p and decreasing θ_s for a low-pass filter). Also, it is often necessary to experiment with the weights until satisfactory results are achieved.

2. Occasionally, the resulting frequency response may be peculiar. For example, the transition-band response may be nonmonotonic, or the ripple may be irregular. In such cases, changing the weighting function usually solves the problem.

Example 9.15 Recall Example 9.10, but assume that the pass-band tolerance is $\delta_p = 0.1$ instead of 0.01. The Hann and Kaiser designs cannot benefit from this relaxation, so they remain as designed in Example 9.10. A least-squares design gives a filter of order $N = 33$. Figure 9.28 shows the magnitude response of the filter; here it was obtained by artificially decreasing the transition band to $[0.21\pi, 0.29\pi]$. □

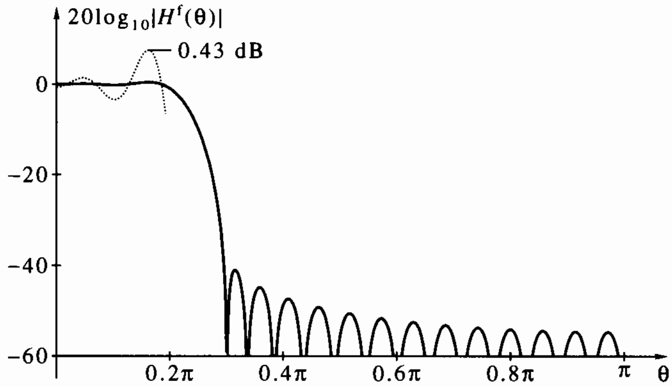

Figure 9.28 The magnitude response of the filter in Example 9.15.

9.6 Equiripple Design of FIR Filters*

The least-squares criterion, presented in the preceding section, often is not entirely satisfactory. A better criterion is the minimization of the maximum error at each band. This criterion leads to an *equiripple filter*, that is, a filter whose amplitude response oscillates uniformly between the tolerance bounds of each band. The design method we study in this section is optimal in the sense of minimizing the maximum magnitude of the ripple in all bands of interest, under the constraint that the filter order N be fixed. A computational procedure for solving this mathematical optimization problem was developed by Remez [1957] and is known as the *Remez exchange algorithm*. The algorithm in common use is by Parks and McClellan [1972a,b] and is known as the *Parks-McClellan algorithm*.

9.6.1 Mathematical Background

We begin as in Section 9.5, by defining the weighted frequency domain error (9.64). The desired amplitude response $A_d(\theta)$ and the weighting function $V(\theta)$ are assumed to be specified on a *compact* subset of $[0, \pi]$; that is, a set $\mathbb{S}$ that is a finite union of closed intervals. These intervals correspond to the pass bands and stop bands, and the complement of $\mathbb{S}$ in $[0, \pi]$ is the union of transition bands, at which the response is not specified. The weighting function fulfills the same role as in least-squares design and is chosen in a similar manner (i.e., proportional to the inverse of the tolerance at each band). We wish to find the coefficients $g[k]$ that minimize the maximum absolute weighted error $|E(\theta)|$ over the entire set $\mathbb{S}$.

Since $F(\theta)$ is a fixed nonnegative function, depending only on the filter type, we can rewrite (9.64) as

$$E(\theta) = \tilde{V}(\theta)[\tilde{A}_{\mathrm{d}}(\theta) - G(\theta)], \quad \theta \in \mathbb{S}, \tag{9.72}$$

where

$$\tilde{V}(\theta) = V(\theta)F(\theta), \quad \tilde{A}_{\mathrm{d}}(\theta) = \frac{A_{\mathrm{d}}(\theta)}{F(\theta)}. \tag{9.73}$$

Then, the optimization problem to be solved is to minimize the nonnegative quantity ε defined by

$$\varepsilon = \max_{\theta \in \mathbb{S}} |E(\theta)| \tag{9.74}$$

over all choices of real coefficients $\{g[k],\ 0 \le k \le K\}$.

The solution of the optimization problem is based on the following theorem by Remez [1957], known as the *alternation theorem.*

Theorem 9.2 The function $G(\theta)$ is optimal in the minimax sense if and only if there exist at least $K+2$ frequencies $\{\theta_i,\ 0 \le i \le K+1\}$ in $\mathbb{S}$ (arranged in increasing order) such that

$$|E(\theta_i)| = \varepsilon, \quad 0 \le i \le K+1, \tag{9.75a}$$

$$E(\theta_{i+1}) = -E(\theta_i), \quad 0 \le i \le K. \tag{9.75b}$$

□

The theorem essentially says that the absolute error achieves its maximum $K+2$ times or more, and does so at alternating signs (hence the name *alternation theorem*). In other words, the optimal approximation is equiripple. The reason for having at least $K+2$ extrema can be intuitively explained as follows. The function $\cos(\theta k)$ is known to be kth-order polynomial in $\cos\theta$, called the kth-order Chebyshev polynomial,[5] and denoted by $T_k(\cos\theta)$. Therefore,

$$G(\theta) = \sum_{k=0}^{K} g[k]T_k(\cos\theta), \tag{9.76}$$

so $G(\theta)$ is a Kth-order polynomial in $\cos\theta$. Therefore, $G(\theta)$ has $K-1$ extrema. If $\mathbb{S}$ consists of B disjoint bands (where B must be at least 2), there may be up to $2B$ additional extrema at the end points of the bands. The alternation theorem says that for the optimal solution, at least three end points must be extrema of the error function. The total number of extrema may be larger, up to a maximum of $K-1+2B$. For example, in the case of a low-pass filter $B=2$, so there may be either $K+2$ or $K+3$ extrema.

9.6.2 The Remez Exchange Algorithm

The Remez exchange algorithm exploits the alternation theorem in the following manner. Assume, temporarily, that we *know* the frequencies $\{\theta_i,\ 0 \le i \le K+1\}$ of the extremal points. Then the $K+2$ parameters $\{g[k],\ 0 \le k \le K\}$ and ε can be obtained by solving the set of linear equations

$$E(\theta_i) = \tilde{V}(\theta_i)[\tilde{A}_{\mathrm{d}}(\theta_i) - G(\theta_i)] = (-1)^i\varepsilon, \quad 0 \le i \le K+1, \tag{9.77}$$

or, equivalently,

$$\sum_{k=0}^{K} g[k]\cos(\theta_i k) + \frac{(-1)^i \varepsilon}{\tilde{V}(\theta_i)} = \tilde{A}_{\mathrm{d}}(\theta_i), \quad 0 \le i \le K+1. \tag{9.78}$$

The Remez exchange algorithm proceeds according to the following steps.

1. Choose the order N of the filter. The following empirical formula, proposed by Kaiser, can be used for estimating N:
$$N = \frac{-20\log_{10}\sqrt{\delta_{\mathrm{p}}\delta_{\mathrm{s}}} - 13}{2.32|\theta_{\mathrm{p}} - \theta_{\mathrm{s}}|}. \tag{9.79}$$
The parameter K is derived from N according to filter type; see Table 9.1.
2. Choose initial values for $\{\theta_i,\ 0 \le i \le K+1\}$. This can be done, for example, by dividing the set $\mathbb{S}$ uniformly.
3. Solve the linear equations (9.78) and obtain $\{g[k],\ 0 \le k \le K\}$ and ε.
4. Find all the extremal points of $E(\theta),\ \theta \in \mathbb{S}$. Both local extrema and end-point extrema must be examined. If there are more than $K+2$ such points, retain those corresponding to the $K+2$ largest values of $|E(\theta)|$. Denote the points thus found by $\{\tilde{\theta}_i\}$.
5. If $|\theta_i - \tilde{\theta}_i| < \mu$, where μ is a small user-chosen number (e.g., 10^{-4}), go to step 6. If not, replace $\{\theta_i\}$ by $\{\tilde{\theta}_i\}$, and go back to step 3 (this is the *exchange step*).
6. Get the coefficients $g[k]$ for the last values of the frequencies, and compute the frequency response. If it meets the specifications, go to step 7. If not, increase the order N (thereby increasing K) and go back to step 2.
7. Convert from the coefficients $g[k]$ to the coefficients $h[n]$, using one of the formulas (9.5), (9.9), (9.15), (9.20), according to the filter type.

Remarks

1. It turns out that ε in (9.78) can be computed explicitly, yielding
$$\varepsilon = \frac{\sum_{i=0}^{K+1}\eta_i\tilde{A}_{\mathrm{d}}(\theta_i)}{\sum_{i=0}^{K+1}\frac{(-1)^i\eta_i}{\tilde{V}(\theta_i)}}, \tag{9.80}$$
where
$$\eta_i = \prod_{\substack{k=0\\k\ne i}}^{K+1}[\cos\theta_i - \cos\theta_k]^{-1}. \tag{9.81}$$
2. Step 3 is performed by computing $E(\theta)$ on a dense grid in $\mathbb{S}$. The number of grid points is chosen by the user; a common choice is $16N$.
3. To compute $E(\theta)$ on the grid, it is necessary to evaluate $G(\theta)$ for each grid point. This can be done without solving the system of equations (9.78). Instead, we can use the Lagrange interpolation formula
$$G(\theta) = \frac{\sum_{i=0}^{K+1}G(\theta_i)\beta_i[\cos\theta - \cos\theta_i]^{-1}}{\sum_{i=0}^{K+1}\beta_i[\cos\theta - \cos\theta_i]^{-1}}, \tag{9.82}$$
where
$$\beta_i = \prod_{\substack{k=0\\k\ne i}}^{K+1}[\cos\theta_i - \cos\theta_k]^{-1}. \tag{9.83}$$

The values $G(\theta_i)$ are known, since by (9.77)

$$G(\theta_i) = \tilde{A}_{\text{d}}(\theta_i) - \frac{(-1)^i \varepsilon}{\tilde{V}(\theta_i)}. \tag{9.84}$$

Only after the algorithm stops, we need to solve (9.78) once and obtain the final values of $g[k]$.

We do not include a program for equiripple FIR filter design in this book, due to the considerable complexity of such a program. If you have access to the MATLAB Signal Processing Toolbox, you can use the program `remez` for this purpose. The design examples given in Section 9.6.3 were prepared using this program.

9.6.3 Equiripple FIR Design Examples

Example 9.16 Consider the low-pass filter whose specifications were given in Example 8.1. An equiripple FIR filter that meets these specifications has order $N = 146$. Figure 9.29 shows the magnitude response of the filter. Figure 9.30(a) shows an input signal to the filter, consisting of a sinusoid at frequency $f_0 = 500$ Hz (i.e., in the middle of the pass band) and white noise at SNR 10 dB. Figure 9.30(b) shows the corresponding output signal. We choose a segment of the output signal delayed by 73 samples with respect to the input signal, to compensate for the phase delay of the filter. As we see, the output signal is considerably cleaner than the input signal, as a result of the noise attenuation performed by the filter. However, the amplitude of the output signal is not constant. This effect is due to the residual noise: The noise at the filter's output is not white any more, but approximately band limited, and its bandwidth is the same as the pass band of the filter. □

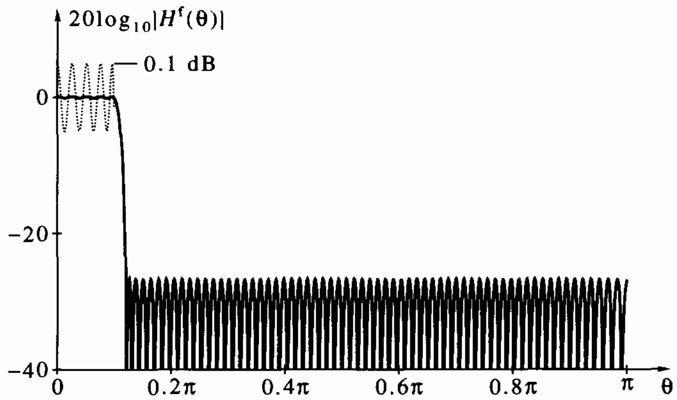

Figure 9.29 Magnitude response of the filter in Example 9.16.

Example 9.17 Consider the high-pass filter whose specifications were given in Example 8.2. An equiripple FIR filter that meets these specifications has order $N = 248$. Figure 9.31 shows the magnitude response of the filter. □

Example 9.18 Consider the five filters whose specifications were given in Example 8.3. Here we design the band-pass filter for the second channel. An equiripple FIR filter

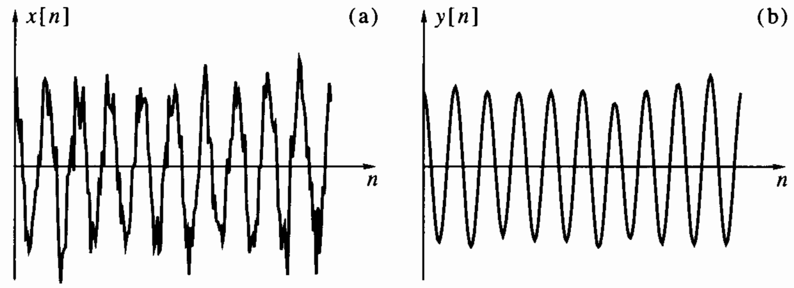

Figure 9.30 Input (a) and output (b) signals in Example 9.16.

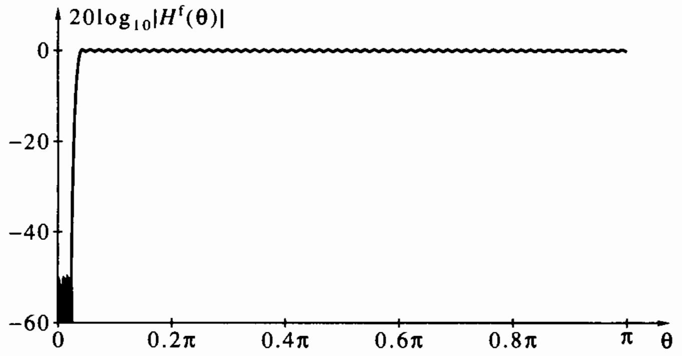

Figure 9.31 Magnitude response of the filter in Example 9.17.

that meets these specifications has order $N = 116$. Figure 9.32 shows the magnitude response of the filter. □

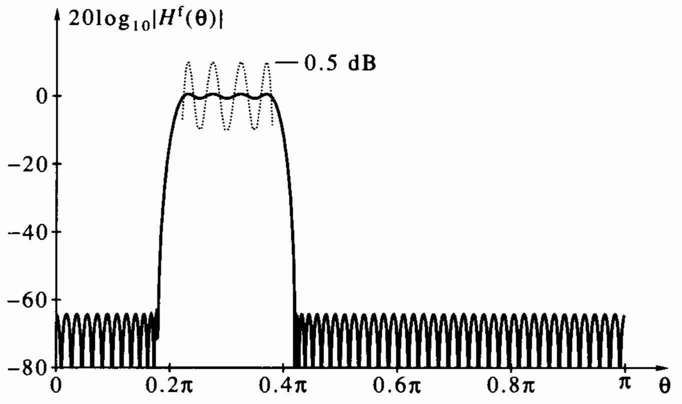

Figure 9.32 Magnitude response of the filter in Example 9.18.

Example 9.19 Consider the band-stop filter whose specifications were given in Example 8.4. An equiripple FIR filter that meets these specifications has order $N = 272$. Figure 9.33 shows the magnitude response of the filter. To illustrate the operation of this filter, we feed it with a simulated physiological signal $x[n]$, as shown in

Figure 9.34, part a. The response to this signal (delayed by 136 samples to account for the phase delay) is shown in part b. As we see, the response is almost identical to the input, since the filter eliminates only a negligible percentage of the energy. Part c shows the same signal as in part a, with an added 60 Hz sinusoidal interference, whose energy is equal to that of the signal. Part d shows the response to the signal in part c. As we see, the filter eliminates the sinusoidal interference almost completely, and the signal in part d is almost identical to those shown in parts a and b. □

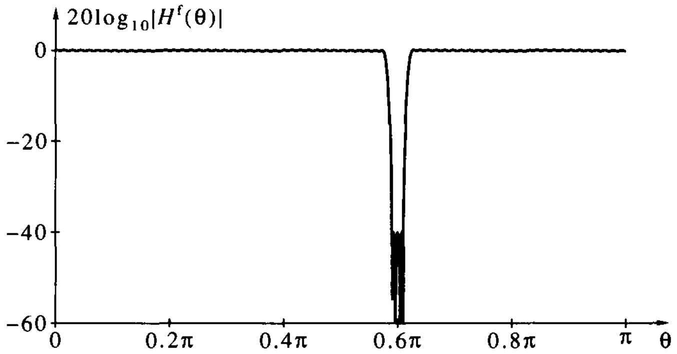

Figure 9.33 Magnitude response of the filter in Example 9.19.

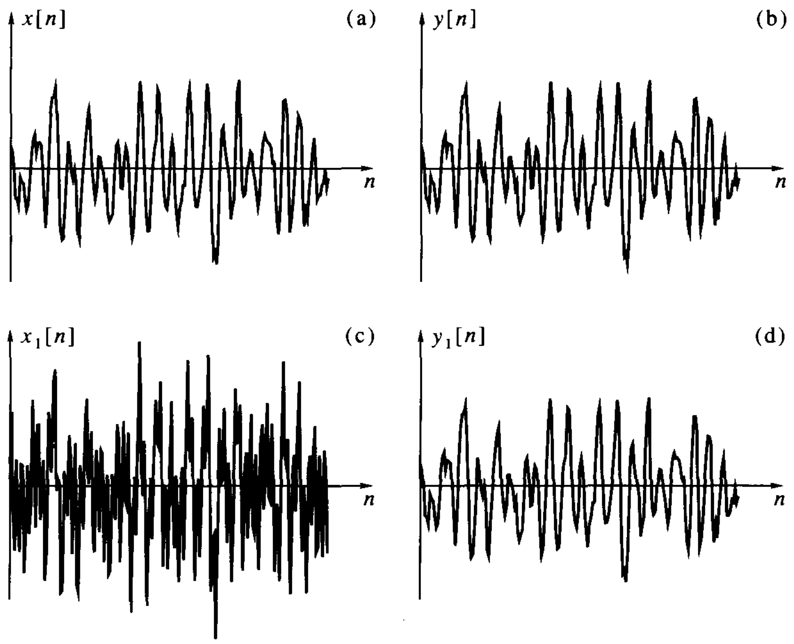

Figure 9.34 Signals in Example 9.19: (a) input signal; (b) output signal; (c) input signal with sinusoidal interference; (d) response to signal with sinusoidal interference.

9.7 Summary and Complements

9.7.1 Summary

This chapter was devoted to the properties and design of FIR filters. FIR filters are almost always designed so as to have linear phase (exact or generalized). There are four types of linear-phase filter: Types I and II have even symmetry of the impulse response coefficients, and even and odd orders, respectively; types III and IV have odd symmetry of the coefficients, and even and odd orders, respectively. A type-I filter is suitable for LP, HP, BP, and BS filters, as well as for multiband filters. Type II, on the other hand, is limited to LP and BP filters. Types III and IV are used mainly for differentiators and Hilbert transformers.

The zero locations of linear-phase FIR filters are not arbitrary, since for every zero at $z = \beta$, there must be zero at $z = \beta^{-1}$. This implies that linear-phase filters usually have zeros both inside and outside (as well as on) the unit circle.

The simplest design method for FIR filters is impulse response truncation. This method is based on computing the (infinite) impulse response of an ideal filter that has the desired frequency response by the inverse Fourier transform integral, and then truncating the impulse response to a finite length. It is applicable to LP, HP, BP, and BS filters, as well as to multiband filters, differentiators, and Hilbert transformers.

The IRT method is mathematically identical to truncating the Fourier series of a periodic function, hence it shares similar properties: On one hand, it is optimal in the sense of minimizing the integral of frequency-domain square error; on the other hand, it suffers from the Gibbs phenomenon. Windowing of the IRT filter attenuates the Gibbs oscillations, thereby reducing the pass-band ripple and increasing the stop-band attenuation. The tolerances of an FIR filter obtained by windowing depend on the window, but we always get $\delta_p \approx \delta_s$ in this design method. The width of the transition band(s) depends on the window and on the order of the filter. Of the various windows, the Kaiser window is most commonly used for filter design. The parameter α is determined by the specified ripple, whereas the order is determined by both the ripple and the specified transition bandwidth.

The least-squares design method is a convenient alternative to the windowing method when the tolerance parameters δ_p, δ_s differ widely, or when the desired amplitude response has a nonstandard shape (such as when it is only tabulated, not defined by a formula). Least-squares design requires trial and error in setting up the specification parameters and the weighting function.

Equiripple design is an optimal method, in the sense of providing the minimum-order filter that meets a given set of specifications. However, this design method requires sophisticated software tools, which are not easy to develop. Such software tools are available for standard filter types (e.g., in the Signal Processing Toolbox of MATLAB). Although the equiripple principle (in the form of the alternation theorem) applies to general amplitude responses and general weighting functions, this method is rarely used in its full generality, due to the absence of widely accessible software tools.

We summarize the subject by reiterating the main advantages and disadvantages of FIR filters:

1. Advantages:
   (a) Linear phase.
   (b) Inherent stability.
   (c) Flexibility in achieving almost any desired amplitude response.

   (d) Existence of convenient design techniques and sophisticated design tools.
   (e) Low sensitivity to finite word length effects (to be studied in Chapter 11).

2. Disadvantages:
   (a) High complexity of implementation, since large orders are needed to achieve tight tolerances and narrow transition bands.
   (b) Large delays, which may be undesirable in certain applications.

9.7.2 Complements

1. [p. 284] Many textbooks on digital signal processing teach the design of even-order filters first, assuming zero phase in the design process and then shifting the impulse response to make it causal. Then, special tricks are employed in designing odd-order filters. The method presented here avoids the need to learn such tricks.
2. [p. 288] See almost any book on communication systems for Hilbert transforms, analytic signals, and their applications. Also see the example in Section 14.4 of this book.
3. [p. 291] Josiah Willard Gibbs (1839–1903), a distinguished physicist, did not discover the phenomenon (it was the physicist Albert Abraham Michelson who did), but offered a mathematical explanation for it.
4. [p. 304] The solution of (9.67) is the unique global minimizer of ε^2 because the matrix of second partial derivatives of ε^2 with respect to the $g[k]$ is positive definite.
5. [p. 307] We shall discuss the Chebyshev polynomials in detail in Section 10.3, when we present Chebyshev filters.

9.8 MATLAB Programs

Program 9.1 Design of a multiband FIR filter.

```
function h = firdes(N,spec,win);
% Synopsis: h = firdes(N,spec,win).
% Design of a general multiband FIR filter by truncated
% impulse response, with optional windowing.
% Input parameters:
% N: the filter order (the number of coefficients is N+1)
% spec: a table of K rows and 3 columns, a row to a band:
%       spec(k,1) is the low cutoff frequency,
%       spec(k,2) is the high cutoff frequency,
%       spec(k,3) is the gain.
% win: an optional window of length N+1.
% Output:
% h: the impulse response coefficients.

flag = rem(N,2); [K,m] = size(spec);
n = (0:N) - N/2; if (~flag), n(N/2+1) = 1; end, h = zeros(1,N+1);
for k = 1:K,
   temp = (spec(k,3)/pi)*(sin(spec(k,2)*n) - sin(spec(k,1)*n))./n;
   if (~flag), temp(N/2+1) = spec(k,3)*(spec(k,2) - spec(k,1))/pi; end
   h = h + temp; end
if  nargin == 3, h = h.*reshape(win,1,N+1); end
```

Program 9.2 Design of FIR differentiators and Hilbert transformers.

```
function h = diffhilb(typ,N,win);
% Synopsis: h = diffhilb(typ,N,win).
% Design of an FIR differentiator or an FIR Hilbert transformer
% by truncated impulse response, with optional windowing.
% Input parameters:
% typ: 'd' for differentiator, 'b' for Hilbert
% N: the filter order (the number of coefficients is N+1)
% win: an optional window.
% Output:
% h: the filter coefficients.

flag = rem(N,2); n = (0:N)-(N/2); if (~flag), n(N/2+1) = 1; end
if (typ == 'd'),
   if (~flag), h = ((-1).^n)./n; h(N/2+1) = 0;
   else, h = ((-1).^(round(n+0.5)))./(pi*n.^2); end
elseif (typ == 'b'),
   h = (1-cos(pi*n))./(pi*n); if (~flag), h(N/2+1) = 0; end, end
if  nargin == 3, h = h.*win; end
```

Program 9.3 Kaiser window FIR filter design according to prescribed specifications.

```
function h = firkais(typ,par,theta,deltap,deltas);
% Synopsis: h = firkais(typ,par,theta,deltap,deltas).
% Designs an FIR filter of one of the six basic types by
% Kaiser window, to meet prescribed specifications.
% Input parameters:
% typ: the filter type:
%      'l','h','p','s' for LP, HP, BP, BS, respectively,
%      'b' for Hilbert transformer, 'd' for differentiator
% par: 'e' for even order (type I or III),
%      'o' for odd order (type II or IV)
% theta: vector of band-edge frequencies in increasing order.
% deltap: one or two pass-band tolerances
% deltas: one or two stop-band tolerances; not needed for
%         typ = 'b' or typ = 'd'
% Output:
% h: the filter coefficients.

if (nargin == 4), deltas = deltap; end
if (typ == 'p' | typ == 's'),
   if (length(deltap) == 1), deltap = deltap*[1,1]; end
   if (length(deltas) == 1), deltas = deltas*[1,1]; end
end
[N,alpha] = kaispar(typ,par,theta,deltap,deltas);
while (1),
   if (alpha == 0), win = window(N+1,'rect');
   else, win = window(N+1,'kais',alpha); end
   if (typ=='l'), h = firdes(N,[0,mean(theta),1],win);
   elseif (typ=='h'), h = firdes(N,[mean(theta),pi,1],win);
   elseif (typ=='p'),
   h = firdes(N,[mean(theta(1:2)),mean(theta(3:4)),1],win);
   elseif (typ=='s'),
   h = firdes(N, ...
       [0,mean(theta(1:2)),1; mean(theta(3:4)),pi,1],win);
   elseif (typ=='b' | typ=='d'),
   h = diffhilb(typ,N,win); end
   res = verspec(h,typ,theta,deltap,deltas);
   if (res==0), break;
   elseif (res==1), N = N+2;
   else, alpha = alpha+0.05; end
end
```

Program 9.4 Computation of N and α to meet the specifications of a Kaiser window FIR filter.

```
function [N,alpha] = kaispar(typ,par,theta,deltap,deltas);
% Synopsis: kaispar(typ,par,theta,deltap,deltas).
% Estimates parameters for FIR Kaiser window filter design.
% Input parameters: see description in firkais.m.
% Output parameters:
% N: the filter order
% alpha: the Kaiser window parameter.

A = -20*log10(min([deltap,deltas]));
if (A > 50), alpha = 0.1102*(A-8.7);
elseif (A > 21), alpha = 0.5842*(A-21)^(0.4)+0.07886*(A-21);
else, alpha = 0; end
if (typ == 'b'), dt = theta;
elseif (typ == 'd'), dt = (pi-theta);
else,
   if (length(theta) == 2), dt = theta(2)-theta(1);
   else, dt = min(theta(2)-theta(1),theta(4)-theta(3)); end
end
N = ceil((A-7.95)/(2.285*dt)); Npar = rem(N,2);
oddpermit = (par=='o') & (typ~='h') & (typ~='s');
if (Npar ~= oddpermit), N = N+1; end
```

Program 9.5 Verification that a given FIR filter meets the specifications.

```
function res = verspec(h,typ,t,dp,ds);
% Synopsis: res = verspec(h,typ,t,dp,ds).
% Verifies that an FIR filter meets the design specifications.
% Input parameters:
% h: the FIR filter coefficients
% other parameters: see description in firkais.m.
% Output:
% res: 0: OK, 1: increase order, 2: increase alpha.

if (typ=='l'), ntest = 1;
Hp = abs(frqresp(h,1,100,[max(0,1.5*t(1)-0.5*t(2)),t(1)]));
Hs = abs(frqresp(h,1,100,[t(2),min(pi,1.5*t(2)-0.5*t(1))]));
elseif (typ=='h'), ntest = 1;
Hp = abs(frqresp(h,1,100,[t(2),min(pi,1.5*t(2)-0.5*t(1))]));
Hs = abs(frqresp(h,1,100,[max(0,1.5*t(1)-0.5*t(2)),t(1)]));
elseif (typ=='p'), ntest = 2;
Hp1 = abs(frqresp(h,1,100,[t(2),min(t(3),1.5*t(2)-0.5*t(1))]));
Hs1 = abs(frqresp(h,1,100,[max(0,1.5*t(1)-0.5*t(2)),t(1)]));
Hp2 = abs(frqresp(h,1,100,[max(t(2),1.5*t(3)-0.5*t(4)),t(3)]));
Hs2 = abs(frqresp(h,1,100,[t(4),min(pi,1.5*t(4)-0.5*t(3))]));
Hp = [Hp1; Hp2]; Hs = [Hs1; Hs2];
elseif (typ=='s'), ntest = 2;
Hp1 = abs(frqresp(h,1,100,[max(0,1.5*t(1)-0.5*t(2)),t(1)]));
Hs1 = abs(frqresp(h,1,100,[t(2),min(t(3),1.5*t(2)-0.5*t(1))]));
Hp2 = abs(frqresp(h,1,100,[t(4),min(pi,1.5*t(4)-0.5*t(3))]));
Hs2 = abs(frqresp(h,1,100,[max(t(2),1.5*t(3)-0.5*t(4)),t(3)]));
Hp = [Hp1; Hp2]; Hs = [Hs1; Hs2];
elseif (typ=='b'), ntest = 1;
Hp = abs(frqresp(h,1,100,[t,2*t])); Hs = zeros(1,100);
end

res = 0;
for i = 1:ntest,
   if(max(abs(Hp(i,1)-1),abs(Hp(i,100)-1)) > dp(i)), res = 1;
   elseif(max(Hs(i,1),Hs(i,100)) > ds(i)), res = 1; end
end
if (res), return, end
for i = 1:ntest,
   if(max(abs(Hp(i,:)-1)) > dp(i)), res = 2;
   elseif(max(Hs(i,:)) > ds(i)), res = 2; end
end
```

Program 9.6 Least-squares design of linear-phase FIR filters.

```
function h = firls(N,typ,theta,deltap,deltas);
% Synopsis: h = firls(N,typ,theta,deltap,deltas).
% Designs an FIR filter of one of the four basic types by
% least-squares.
% Input parameters:
% N: the filter order
% typ: the filter type:
%      'l','h','p','s' for LP, HP, BP, BS, respectively,
% theta: vector of band-edge frequencies in increasing order.
% deltap: one or two pass-band tolerances
% deltas: one or two stop-band tolerances.
% Output:
% h: the filter coefficients.

thetai = (pi/(32*N)) + (pi/(16*N))*(0:(16*N-1));
if (rem(N,2)),
   F = cos(0.5*thetai); K = (N-1)/2;
else,
   F = ones(1,16*N); K = N/2;
end
if (typ == 'p' | typ == 's'),
   if (length(deltap) == 1), deltap = deltap*[1,1]; end
   if (length(deltas) == 1), deltas = deltas*[1,1]; end
end
[V,Ad] = firlsaux(typ,theta,deltap,deltas,thetai);
carray = cos(thetai'*(0:K)).*((F.*V)'*ones(1,K+1));
darray = (V.*Ad)';
g = (carray\darray)';
if (rem(N,2)),
   h = 0.25*[g(K+1),fliplr(g(3:K+1))+fliplr(g(2:K))];
   h = [h,0.25*g(2)+0.5*g(1)];
   h = [h,fliplr(h)];
else,
   h = [0.5*fliplr(g(2:K+1)),g(1),0.5*g(2:K+1)];
end
```

Program 9.7 An auxiliary subroutine for `firls`.

```
function [V,Ad] = firlsaux(typ,theta,deltap,deltas,thetai);
% Synopsis: [V,Ad] = firlsaux(typ,theta,deltap,deltas,thetai).
% An auxiliary function for FIRLS.
% Input parameters: see firls.m
% Output parameters:
% V, Ad: variables needed in firls.m

ind1 = find(thetai < theta(1));
ind3 = find(thetai > theta(length(theta)));
if (typ == 'p' | typ == 's'),
ind2 = find(thetai > theta(2) & thetai < theta(3)); end
V = zeros(1,length(thetai)); Ad = zeros(1,length(thetai));
if (typ == 'l'),
   Ad(ind1) = ones(1,length(ind1));
   Ad(ind3) = zeros(1,length(ind3));
   V(ind1) = (1/deltap)*ones(1,length(ind1));
   V(ind3) = (1/deltas)*ones(1,length(ind3));
elseif (typ == 'h'),
   Ad(ind1) = zeros(1,length(ind1));
   Ad(ind3) = ones(1,length(ind3));
   V(ind1) = (1/deltas)*ones(1,length(ind1));
   V(ind3) = (1/deltap)*ones(1,length(ind3));
elseif (typ == 'p'),
   Ad(ind1) = zeros(1,length(ind1));
   Ad(ind2) = ones(1,length(ind2));
   Ad(ind3) = zeros(1,length(ind3));
   V(ind1) = (1/deltas(1))*ones(1,length(ind1));
   V(ind2) = (1/deltap(1))*ones(1,length(ind2));
   V(ind3) = (1/deltas(2))*ones(1,length(ind3));
elseif (typ == 's'),
   Ad(ind1) = ones(1,length(ind1));
   Ad(ind2) = zeros(1,length(ind2));
   Ad(ind3) = ones(1,length(ind3));
   V(ind1) = (1/deltap(1))*ones(1,length(ind1));
   V(ind2) = (1/deltas(1))*ones(1,length(ind2));
   V(ind3) = (1/deltap(2))*ones(1,length(ind3));
end
```

9.9 Problems

9.1 Show that a type-II filter has a factor $(1 + z^{-1})$ in its transfer function, a type-III filter has a factor $(1 - z^{-2})$, and a type-IV filter has a factor $(1 - z^{-1})$.

9.2 Let $H^z(z)$ be a general FIR filter (not necessarily linear phase).

(a) Show that $H^z(z)$ can be expressed as a sum of two FIR filters

$$H^z(z) = G_1^z(z) + G_2^z(z),$$

where $G_1^z(z)$ and $G_2^z(z)$ have linear phase (exact or generalized) and are of the same order as $H^z(z)$ or less.

(b) Of what types are $G_1^z(z)$ and $G_2^z(z)$ when N is even? When N is odd?

(c) Express $|H^f(\theta)|$ in terms of the amplitude functions of $G_1^z(z)$ and $G_2^z(z)$.

9.3 The filters in Figure 9.35 are FIR, with linear phase, real and causal. Their orders are N_1, N_2, and $N_3 = N_1 + N_2$, respectively. Each of the three filters can be either symmetric or antisymmetric, so there are eight cases in total. For which of these eight cases will the equivalent filter have linear phase (either exact or generalized)? In each case state whether the equivalent filter is symmetric or antisymmetric.

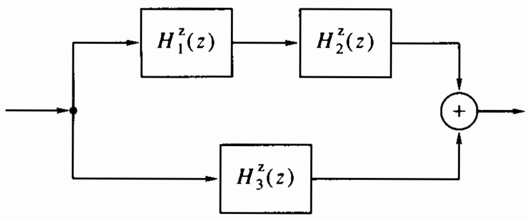

Figure 9.35 Pertaining to Problem 9.3.

9.4 The impulse response of a linear-phase FIR filter starts at the values

$$h[0] = 1, \quad h[1] = 3, \quad h[2] = -2.$$

For each of the four filter types, find the coefficients of the smallest order FIR filter that satisfies this condition.

9.5 Let $H^z(z)$ be the transfer function of a linear-phase FIR filter with real coefficients. The filter is known to have zeros in the following locations:

$$\beta_1 = 1, \quad \beta_2 = 0.5e^{j\pi/3}, \quad \beta_3 = -5, \quad \beta_4 = j.$$

What is the lowest possible order of the filter? What is the transfer function of the filter of the lowest order, and what is its type?

9.6 Suppose that $H^z(z)$ is a linear-phase FIR filter. Can a filter whose transfer function is $1/H^z(z)$ be causal and stable? Can $1/H^z(z)$ be stable if it is not required to be causal?

9.7 This problem explores the preservation of linear phase in cascade and parallel connections of FIR filters.

(a) Show that a cascade connection of two (generalized) linear-phase filters has (generalized) linear phase. Find the order of the equivalent filter, its amplitude function, and its initial phase as a function of the types of the two filters.

(b) Is the same true for a parallel connection? What if the two filters in parallel have the same order?

9.8 In Section 9.1.6 we saw that the zeros of a linear-phase filter must satisfy certain symmetry conditions, divided into five possible cases. Show that the converse is also true, that is: If every zero of an FIR filter satisfies one of the five symmetry conditions, then the filter has linear phase.

9.9 A real, causal, linear-phase FIR filter of order N is known to satisfy

$$\sum_{n=0}^{N} h[n] \cdot (0.8)^{-n} e^{-j\pi n/3} = 0.$$

Find the order and the coefficients of the filter of minimal order satisfying these conditions. Assume that $h[0] = 1$.

9.10 A real FIR filter is required to block all real sinusoidal signals of frequencies $\theta_k = 0.125k\pi$, for all integer k. Subject to this requirement, the filter should have a minimal order. There are no other requirements from the filter. Specify the order and the transfer function of the filter. Does the filter have exact linear phase, generalized linear phase, or no linear phase at all?

9.11 You are given a fourth-order FIR filter $H_1^z(z)$ that has a double zero at $\beta_{1,3} = 0.5 + j0.75$, and a double zero at $\beta_{2,4} = 0.5 - j0.75$. Suggest a linear-phase FIR filter $H_2^z(z)$ whose magnitude response $|H_2^f(\theta)|$ is as close as possible to $|H_1^f(\theta)|$.

9.12 Let

$$x[n] = 0.3\cos(0.9273n + 0.25\pi) - 0.4\sin(\pi n - 0.3).$$

Find an FIR filter $h[n]$ of minimal order that satisfies

$$\{h * x\}[n] = 0, \quad n \in \mathbb{Z}.$$

9.13 A type-II FIR filter of order 5 is known to satisfy

$$H^z(1) = 6, \quad H^z(0.5 - 0.5j) = 0.$$

Find $H^z(-j)$.

9.14 The DFT of the coefficients of a causal FIR filter is $\{6, -1-j, 0, -1+j\}$. State as many properties of the filter as you can think of.

9.15 You are given two causal filters whose frequency responses are

$$H_1^f(\theta) = \frac{1}{1 - 0.5e^{-j\theta}}, \quad H_2^f(\theta) = \left(\frac{\sin 2\theta}{\sin 0.5\theta} e^{-j1.5\theta}\right)^2.$$

Let

$$H^f(\theta) = \frac{1}{2\pi}\{H_1^f * H_2^f\}(\theta).$$

Compute the coefficients of $H^f(\theta)$.

9.16 Let

$$x_1[n] = 0.2\text{sinc}(0.2n), \quad y_1[n] = \delta[n] - 0.6\text{sinc}(0.6n),$$
$$x_2[n] = 0.6\text{sinc}(0.6n), \quad y_2[n] = \delta[n] - 0.2\text{sinc}(0.2n).$$

Compute $\{x_1 * y_1\}[n]$ and $\{x_2 * y_2\}[n]$.

9.17 Design a symmetric FIR filter with group delay $N/2$ according to the ideal magnitude response

$$|H_{\mathrm{d}}^{\mathrm{f}}(\theta)| = \begin{cases} 1, & 0 \le |\theta| \le \frac{\pi}{3}, \\ 0, & \frac{\pi}{3} < |\theta| < \frac{2\pi}{3}, \\ 0.5, & \frac{2\pi}{3} \le |\theta| \le \pi. \end{cases}$$

(a) Compute $h_{\mathrm{d}}[n]$.

(b) If the filter is designed with the Hamming window and its order is $N = 40$, what are the values of

$$\theta_{\mathrm{p},1},\ \theta_{\mathrm{s},1},\ \theta_{\mathrm{s},2},\ \theta_{\mathrm{p},2},\ \delta_{\mathrm{p},1},\ \delta_{\mathrm{s}},\ \delta_{\mathrm{p},2}?$$

(c) Suppose we want to have $\delta_{\mathrm{p},1} = \delta_{\mathrm{p},2} = 0.01$, and $\delta_{\mathrm{s}} = 0.005$. Is it possible to achieve this with a Hamming window filter of order $N = 39$?

9.18 Consider the following filter specifications:

$$\theta_{\mathrm{p}} = 0.3\pi,\ \theta_{\mathrm{s}} = 0.5\pi,\ \delta_{\mathrm{p}} = 0.0025,\ \delta_{\mathrm{s}} = 0.0021.$$

(a) Design a low-pass FIR type-I filter $H^{\mathrm{z}}(z)$, using the Kaiser window, according to these specifications. Give the order N and the parameter α (there is no need to write down the coefficient values).

(b) Let $\tilde{H}^{\mathrm{z}}(z)$ be the noncausal filter

$$\tilde{H}^{\mathrm{z}}(z) = z^{N/2}H^{\mathrm{z}}(z).$$

Define

$$\tilde{G}_1^{\mathrm{z}}(z) = [\tilde{H}^{\mathrm{z}}(z)]^2,$$
$$\tilde{G}_2^{\mathrm{z}}(z) = 2\tilde{H}^{\mathrm{z}}(z) - \tilde{G}_1^{\mathrm{z}}(z).$$

Is $\tilde{G}_1^{\mathrm{z}}(z)$ low pass, high pass, band pass, or band stop? What about $\tilde{G}_2^{\mathrm{z}}(z)$?

(c) Compute the pass-band ripple and the stop-band attenuation of $\tilde{G}_1^{\mathrm{z}}(z)$ and $\tilde{G}_2^{\mathrm{z}}(z)$.

(d) Now assume that $\tilde{H}^{\mathrm{z}}(z)$ is not available, only $H^{\mathrm{z}}(z)$. We wish to realize two causal filters $G_1^{\mathrm{z}}(z)$ and $G_2^{\mathrm{z}}(z)$ having the same amplitude response as $\tilde{G}_1^{\mathrm{z}}(z)$ and $\tilde{G}_2^{\mathrm{z}}(z)$, respectively, using copies of $H^{\mathrm{z}}(z)$ as building blocks (and other elements as needed). Show how to do this.

9.19 We are given the system shown in Figure 9.36. The signal $x(t)$ is limited to $|f| \le 0.5$ Hz, and the sampling interval is $T = 1$ second. The filter is FIR of even-order N. We wish to design the filter such that the output will satisfy

$$y[n] \approx x(n - L - 0.25),$$

where L is a given positive integer.

(a) What is the ideal frequency response $H_{\mathrm{d}}^{\mathrm{f}}(\theta)$ of the filter?

(b) Choose an appropriate N (which depends on L), and design the filter by the IRT method.

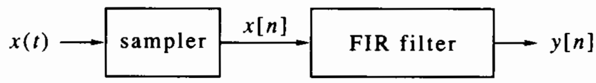

Figure 9.36 Pertaining to Problem 9.19.

(c) Does $H^z(z)$ have exact linear phase, generalized linear phase, or no linear phase at all?

9.20 Figure 9.37 shows the ideal magnitude and phase responses of a filter that is a differentiator at low frequencies and high pass at high frequencies.

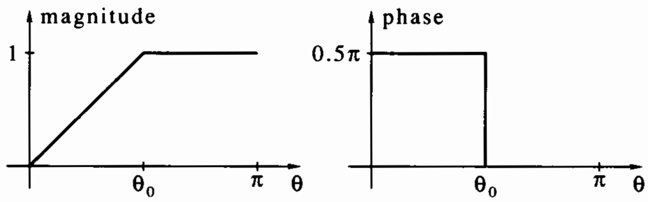

Figure 9.37 Pertaining to Problem 9.20.

(a) Compute the desired impulse response $h_d[n]$. Does the truncated impulse response have linear phase? Explain why or why not.

(b) Design an FIR filter of order $N = 128$ and having $\theta_0 = 0.5\pi$, using: (1) a rectangular window; (2) a Hamming window; (3) a Kaiser window with $\alpha = 6$; (4) a Kaiser window with $\alpha = 12$. Compute and plot the magnitude and phase responses of the four filters.

9.21 We are given a low-pass type-I FIR filter $h[n]$, with given parameters θ_p, θ_s, δ_p, δ_s. Define

$$g[n] = (-1)^{N/2}\delta[n - 0.5N] - (-1)^n h[n].$$

(a) What type is the filter $g[n]$ and what is the nature of its frequency response?

(b) Express the tolerance and band-edge parameters of the filter $g[n]$ in terms of the parameters of $h[n]$.

9.22 Consider a digital filter whose frequency response is like the right side of (3.67), with $f_0 T = 0.5$ and $\theta = \omega T$.

(a) Compute the ideal impulse response $h_d[n]$. Is it a half-band filter?

(b) Is

$$g_d[n] \triangleq \{h_d * h_d\}[n]$$

a half-band filter?

(c) Suggest a causal FIR filter $h[n]$ that will approximate this frequency response. Compute and plot the magnitude response of the FIR filter for $N = 38$ and $\alpha = 0.4$.

(d) Is

$$g[n] \triangleq \{h * h\}[n]$$

a half-band filter?

9.23 Suggest a definition of a *quarter-band filter.* Write a necessary and sufficient condition for a noncausal filter to be quarter band, and then for a causal filter. Suggest a procedure for designing a quarter-band FIR filter. Show the magnitude response of a quarter-band, raised-cosine filter with α, N, and window of your choice.

9.24 It is required to design a linear-phase FIR filter according to the desired amplitude response

$$A_d(\theta) = \cos(0.5\theta)[1 + \cos\theta].$$

The order of the filter N is required to be no larger than 8. Subject to this requirement, the filter must have the smallest possible integral of square error ε^2; see its definition (9.41). Find the filter's coefficients $h[n]$ and the associated value of ε^2. Hint: Solve without computing any integrals.

9.25 Find an FIR filter $H^z(z)$ whose amplitude response is

$$A(\theta) = \sin(2\theta).$$

You are free to choose the phase response $\phi(\theta)$.

9.26 Let $w[n]$ be an arbitrary window, and define a new window $v[n] = \{w * w\}[n]$. Show that the function $\Gamma_v(x)$ corresponding to $v[n]$ is monotone, similar to that of the Bartlett window.

9.27 Suggest a procedure for designing a low-pass, half-band filter using windows. Hint: What is the ideal frequency response?

9.28 Suppose we wish to design a low-pass filter with band-edge frequencies θ_p, θ_s. Instead of using the boxcarlike amplitude response (9.28) for $A_d(\theta)$, we can use the function shown in Figure 9.38.

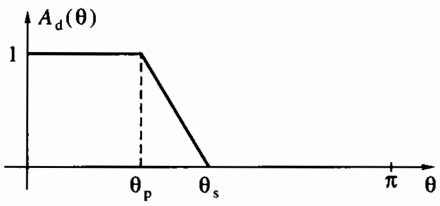

Figure 9.38 Pertaining to Problem 9.28.

(a) Compute the ideal impulse response $h_d[n]$ corresponding to $A_d(\theta)$ shown in the figure.

(b) Use MATLAB for designing a low-pass, linear-phase filter using the IRT method with $h_d[n]$ found in part a. Take $N = 20$, $\theta_p = 0.2\pi$, $\theta_s = 0.4\pi$. Give the filter coefficients $h[0]$ through $h[10]$ to four decimal digits as an answer.

(c) Plot the frequency response of the filter and use it for finding the pass-band ripple and the stop-band attenuation.

(d) Suggest an extension of this idea to band-pass filters (there is no need to work out the details for this case).

9.29 Suppose we wish to form the analytic signal corresponding to a given real signal $x[n]$, as in (9.38), where $y[n]$ is the Hilbert transform of $x[n]$. Since $y[n]$ is not given, we generate it by passing $x[n]$ through a linear-phase FIR Hilbert transformer, as explained in Section 9.2.5. Explain why we must delay $x[n]$ by $N/2$ samples before combining it with $y[n]$ (where N is the order of the Hilbert transformer). Hence state whether it is more convenient to use a type-III or type-IV transformer in this case.

9.30 A linear-phase FIR filter of an even-order N, whose initial conditions are set to zero, is fed with an input signal of length L. As we know, this results in a signal of length $L + N$. It is often desired to retain only L output values, the same as the number of input values. Consider the following three options:

- Deleting the first N output points.
- Deleting the last N output points.
- Deleting $N/2$ points from the beginning and $N/2$ points from the end of the output signal.

(a) Which of the three is used by the MATLAB function `filter`? Try before you give an answer.

(b) Which of the three makes more sense to you? Give reasons. Hint: Design a low-pass filter using `firdes`, and feed it with the signal $x[n] = 1,\ 0 \le n \le L - 1$. Try the three options with MATLAB, then form an opinion.

(c) Will your answer to part b be different if $h[n]$ is a minimum-phase (rather than linear-phase) filter?

9.31 Design an even-order band-pass Hilbert transformer whose ideal frequency response is

$$H_{\mathrm{d}}^{\mathrm{f}}(\theta) = \begin{cases} -je^{-j0.5\theta N}, & 0.25\pi \le \theta \le 0.75\pi, \\ je^{-j0.5\theta N}, & -0.75\pi \le \theta \le -0.25\pi, \\ 0, & \text{otherwise.} \end{cases}$$

The transition bands of the filter should not be wider than 0.1π and its ripple should not be higher than 0.01 in any of the stop bands. As an answer, give a table of the coefficients, accurate to five decimal digits.

9.32* The signal $x_1(t)$ is low pass in the frequency range $|\omega| < 2\pi \cdot 20$, and the magnitude of its Fourier transform is known to be not larger than 1 in this range. The signal $x_2(t)$ is band pass in the frequency range $2\pi \cdot 40 < |\omega| < 2\pi \cdot 60$, and the magnitude of its Fourier transform is known to be not larger than 1 in this range. We measure the signal

$$y(t) = x_1(t) + x_2(t)$$

and sample it at frequency f_{sam}. The discrete-time signal $y[n]$ is filtered by a digital low-pass filter, whose aim is:

- to attenuate the component resulting from $x_2(t)$, such that the magnitude of its Fourier transform in the θ domain will not exceed 0.001;
- to pass the component resulting from $x_1(t)$, with tolerance ± 0.001.

The filter is to be designed by means of a Kaiser window. Two options are considered for the sampling frequency: (1) $f_{\mathrm{sam}} = 120\,\mathrm{Hz}$; (2) $f_{\mathrm{sam}} = 100\,\mathrm{Hz}$.

Which sampling rate will lead to a filter of lower order? Hints: Do not neglect the aliasing in case 1; use (9.56c) for the order of a Kaiser window filter.

9.33* Let $H^z(z)$ be a type-I, Nth-order, low-pass FIR filter with tolerances δ_p, δ_s. Define

$$G^z(z) = [3z^{-N/2} - 2H^z(z)][H^z(z)]^2.$$

What are the tolerances of $G^z(z)$? What is its order? Suggest a possible use of this idea.

9.34* Suppose we wish to design a differentiator using least-squares design.

(a) Explain why it makes sense to use the weighting function

$$W(\theta) = \delta \cdot \theta,$$

where δ is a small number.

(b) Write a MATLAB program for least-squares type-IV differentiator design. Use parts of `firls` as needed.

(c) Use the program to design a differentiator with $\delta = 0.01$. Find and report the order N that meets this specification.

9.35* Suppose we design an equiripple low-pass filter $G^z(z)$ according to the following specifications:

- Odd-order N, that is, type II.
- $\theta_p < \pi$ and a corresponding δ_p.
- $\theta_s = \pi$, that is, no stop band (and therefore no δ_s).

Such a filter is called *one-band filter.* Define

$$H^z(z) = 0.5[G^z(z^2) + z^{-N}].$$

(a) Show that $H^z(z)$ is a half-band filter.

(b) What is the order of $H^z(z)$?

(c) What are the pass-band and stop-band tolerances of $H^z(z)$, and what are the band-edge frequencies?

(d) Carry out the design for $\theta_p = 0.9\pi$, $\delta_p = 0.1$, and $N = 11$. Draw the amplitude response plots of $G^z(z)$ and $H^z(z)$.

This method of equiripple half-band filter design was developed by Vaidyanathan and Nguyen [1987].

9.36* The following method, called *frequency sampling,* is proposed for FIR filter design:

- Sample the desired frequency response $H_d^f(\theta)$ at the $N + 1$ points

$$\theta_k = \frac{2\pi k}{N+1}, \quad 0 \le k \le N.$$

- Let the filter coefficients $h[n]$ be the inverse DFT of $\{H_d^f(\theta_k),\ 0 \le k \le N\}$.

It is clear from the construction of $h[n]$ that its frequency response coincides with the desired one. This, however, does not imply proximity of $H^f(\theta)$ to $H_d^f(\theta)$ at other frequencies.

(a) Prove that for frequencies different from θ_k,

$$H^f(\theta) = \frac{1}{N+1}\sum_{k=0}^{N} H_d^f(\theta_k)\frac{1 - e^{-j\theta(N+1)}}{1 - e^{-j[\theta - 2\pi k/(N+1)]}}.$$

(b) Compute $h[n]$ in the special case of a low-pass filter with desired cutoff frequency θ_c. Be careful to maintain the conjugate symmetry property of $H_d^f(\theta_k)$; otherwise the impulse response $h[n]$ will not be real valued.

(c) Choose $N = 63$, $\theta_c = 0.3\pi$. Compute the impulse response and the frequency response of the filter. Plot the magnitude response.

(d) Conclude from part c on the properties of the frequency sampling design method.