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

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

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

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

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

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

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

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

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

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

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

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

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

Consequently, the choice of window is governed by the desire to have $w[n]$ as short as possible in duration, so as to minimize computation in the implementation of the filter, while having $W(e^{j\omega})$ approximate an impulse.

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

As M increases, the width of the "main lobe" decreases. The main lobe is usually defined as the region between the first zero-crossings on either side of the origin. For the rectangular window, the width of the main lobe is $\Delta_{\omega_m} = 4\pi/(M+1)$. However, for the rectangular window, the side lobes are large, and in fact, as M increases, the peak amplitudes of the main lobe and the side lobes grow in a manner such that the area under each lobe is a constant while the width of each lobe decreases with M. Consequently, as $W(e^{j(\omega-\theta)})$ "slides by" a discontinuity of $H_d(e^{j\theta})$ with increasing ω, the integral of $W(e^{j(\omega-\theta)})H_d(e^{j\theta})$ will oscillate as each side lobe of $W(e^{j(\omega-\theta)})$ moves past the discontinuity.

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

Rectangular

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

Bartlett (triangular)

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

Hanning

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

Hamming

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

Blackman

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

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

Incorporation of Generalized Linear Phase

In designing many types of FIR filters, it is desirable to obtain causal systems with a generalized linear phase response.

Specifically, note that all the windows have the property that

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

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

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

where $W_e(e^{j\omega})$ is a real, even function of ω.

if $h_d[M-n] = h_d[n]$ $\qquad H(e^{j\omega}) = A_e(e^{j\omega})e^{-j\omega M/2}, \qquad (7.50)$

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

if $h_d[M-n] = -h_d[n]$ $\qquad H(e^{j\omega}) = jA_o(e^{j\omega})e^{-j\omega M/2}, \qquad (7.51)$

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

Linear-Phase Lowpass Filter

The desired frequency response is defined as

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

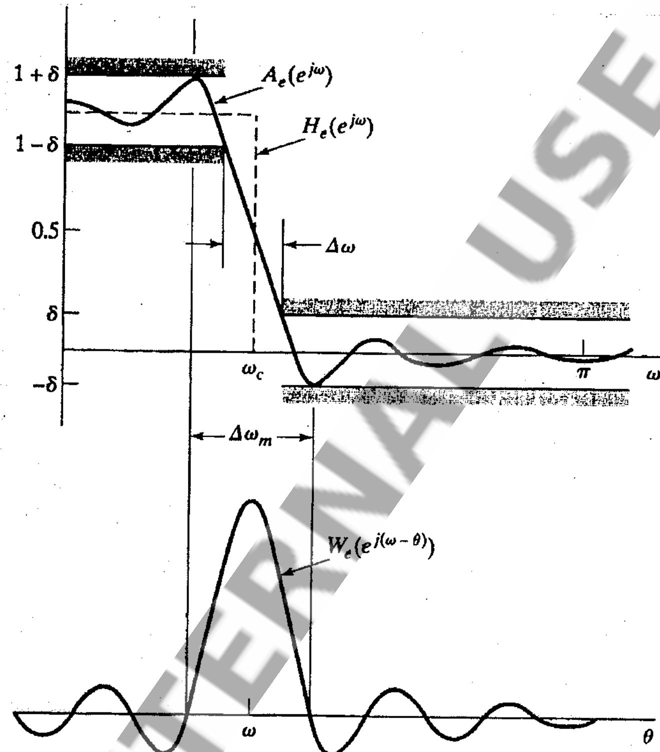
where the generalized linear phase factor has been incorporated into the definition of the ideal lowpass filter. The corresponding ideal impulse response is

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

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

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

then a linear-phase system will result.



Clearly, the windows with the smaller side lobes yield better approximations of the ideal response at a discontinuity . Also, the third column, which shows the width of the main lobe, suggests that narrower transition regions can be achieved by increasing M. Thus, through the choice of the shape and duration of the window, we can control the properties of the resulting FIR filter.

The Kaiser Window Filter Design Method

The trade-off between the main-lobe width and side-lobe area can be quantified by seeking the window function that is maximally concentrated around $\omega = 0$ in the frequency domain.

The Kaiser window is defined as

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

where $\alpha = M/2$, and $I_0(\cdot)$ represents the zeroth-order modified Bessel function of the first kind.

114

Given that δ is fixed, the passband cutoff frequency ω_p of the lowpass filter is defined to be the highest frequency such that $|H(e^{j\omega})| \geq 1-\delta$. The stopband cutoff frequency ω_s is defined to be the lowest frequency such that $|H(e^{j\omega})| \leq \delta$. Therefore, the transition region has width

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

Defining

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

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

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

Furthermore, Kaiser found that to achieve prescribed values of A and $\Delta\omega$, M must satisfy

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

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

Kaiser Window Design of a Lowpass Filter

For this example, we use the same specifications as in Examples 7.4, 7.5, and 7.6, i.e., $\omega_p = 0.4\pi$, $\omega_s = 0.6\pi$, $\delta_1 = 0.01$, and $\delta_2 = 0.001$. Since filters designed by the window method inherently have $\delta_1 = \delta_2$, we must set $\delta = 0.001$. The cutoff frequency of the underlying ideal lowpass filter must be found. Due to the symmetry of the approximation at the discontinuity of $H_d(e^{j\omega})$, we would set

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

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

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

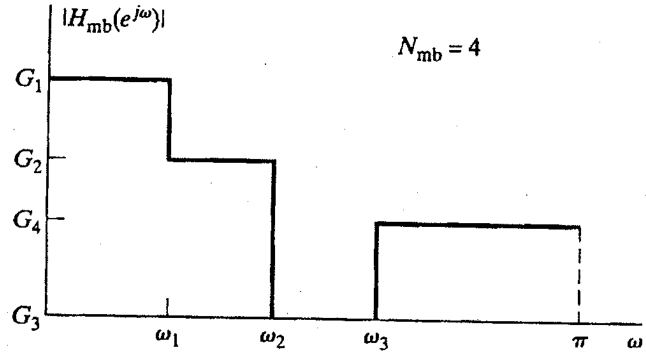
We substitute these two quantities into Eqs. (7.62) and (7.63) to obtain the required values of β and M. For this example the formulas predict

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

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

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

where $\alpha = M/2 = 37/2 = 18.5$. Since $M = 37$ is an odd integer, the resulting linear-phase system would be of type II.



This generalized multiband filter includes lowpass, highpass, bandpass, and bandstop filters as special cases. If such a magnitude function is multiplied by a linear phase factor $e^{-j\omega M/2}$, the corresponding ideal impulse response is

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

where N_{mb} is the number of bands and $G_{N_{mb}+1} = 0$. If $h_{mb}[n]$ is multiplied by a Kaiser window, the type of approximations that we have observed at the single discontinuity of the lowpass and highpass systems will occur at *each* of the discontinuities.

OPTIMUM APPROXIMATIONS OF FIR FILTERS

That is,

$$h[n] = \begin{cases} h_d[n], & 0 \le n \le M, \\ 0, & \text{otherwise}, \end{cases} \tag{7.72}$$

minimizes the expression

$$\varepsilon^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H_d(e^{j\omega}) - H(e^{j\omega})|^2 d\omega. \tag{7.73}$$

However, as we have seen, this approximation criterion leads to adverse behavior at discontinuities of $H_d(e^{j\omega})$.

In designing a causal type I linear-phase FIR filter, it is convenient first to consider the design of a zero-phase filter, i.e., one for which

$$h_e[n] = h_e[-n], \tag{7.74}$$

and then to insert a delay sufficient to make it causal.

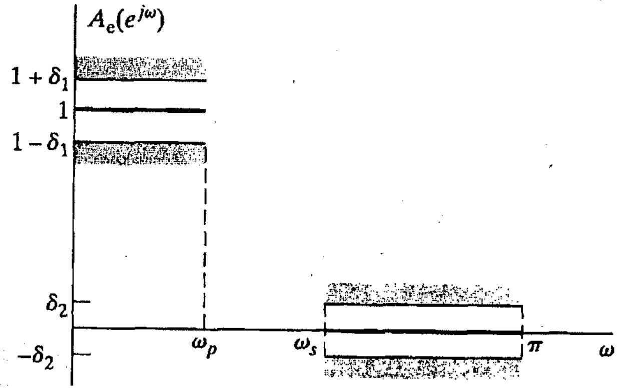
The corresponding frequency response is given by

$$A_e(e^{j\omega}) = \sum_{n=-L}^{L} h_e[n] e^{-j\omega n}, \tag{7.75}$$

with $L = M/2$ an integer, or

$$A_e(e^{j\omega}) = h_e[0] + \sum_{n=1}^{L} 2h_e[n] \cos(\omega n). \tag{7.76}$$

Note that $A_e(e^{j\omega})$ is a real, even, and periodic function of ω. A causal system can be obtained from $h_e[n]$ by delaying it by $L = M/2$ samples.



The Parks–McClellan algorithm is based on reformulating the filter design problem as a problem in polynomial approximation. Specifically, the terms $\cos(\omega n)$ in Eq. (7.76) can be expressed as a sum of powers of $\cos\omega$ in the form

$$\cos(\omega n) = T_n(\cos\omega), \tag{7.79}$$

$T_n(x)$ is the nth-order Chebyshev polynomial, defined as $T_n(x) = \cos(n\cos^{-1} x)$.

Consequently, Eq. (7.76) can be rewritten as an Lth-order polynomial in $\cos\omega$, namely,

$$A_e(e^{j\omega}) = \sum_{k=0}^{L} a_k(\cos\omega)^k, \tag{7.80}$$

where the a_k's are constants that are related to $h_e[n]$, the values of the impulse response. With the substitution $x = \cos\omega$, we can express Eq. (7.80) as

$$A_e(e^{j\omega}) = P(x)|_{x=\cos\omega}, \tag{7.81}$$

where $P(x)$ is the Lth-order polynomial

$$P(x) = \sum_{k=0}^{L} a_k x^k. \tag{7.82}$$

To formalize the approximation problem in this case, let us define an approximation error function

$$E(\omega) = W(\omega)[H_d(e^{j\omega}) - A_e(e^{j\omega})], \tag{7.83}$$

where the weighting function $W(\omega)$ incorporates the approximation error parameters into the design process. In this design method, the error function $E(\omega)$, the weighting function $W(\omega)$, and the desired frequency response $H_d(e^{j\omega})$ are defined only over closed subintervals of $0 \le \omega \le \pi$.

For example, suppose that we wish to obtain an approximation as in Figure 7.31, where L, ω_p, and ω_s are fixed design parameters. For this case,

$$H_d(e^{j\omega}) = \begin{cases} 1, & 0 \le \omega \le \omega_p, \\ 0, & \omega_s \le \omega \le \pi. \end{cases} \tag{7.84}$$

The weighting function $W(\omega)$ allows us to weight the approximation errors differently in the different approximation intervals.

$$W(\omega) = \begin{cases} \dfrac{1}{K} & 0 \le \omega \le \omega_p, \\ 1, & \omega_s \le \omega \le \pi, \end{cases} \tag{7.85}$$

where $K = \delta_1/\delta_2$.

The particular criterion used in this design procedure is the so-called minimax or Chebyshev criterion, where

$$\min_{\{h_e[n]:0\le n\le L\}} \left(\max_{\omega \in F} |E(\omega)| \right),$$

where F is the closed subset of $0 \le \omega \le \pi$ such that $0 \le \omega \le \omega_p$ or $\omega_s \le \omega \le \pi$.

Alternation Theorem: Let F_P denote the closed subset consisting of the disjoint union of closed subsets of the real axis x. Then

$$P(x) = \sum_{k=0}^{r} a_k x^k$$

is an rth-order polynomial. Also, $D_P(x)$ denotes a given desired function of x that is continuous on F_P; $W_P(x)$ is a positive function, continuous on F_P, and

$$E_P(x) = W_P(x)[D_P(x) - P(x)].$$

is the weighted error. The maximum error is defined as

$$\|E\| = \max_{x \in F_P} |E_P(x)|.$$

A necessary and sufficient condition that $P(x)$ be the unique rth-order polynomial that minimizes $\|E\|$ is that $E_p(x)$ exhibit *at least* $(r+2)$ alternations; i.e., there must exist at least $(r+2)$ values x_i in F_p such that $x_1 < x_2 < \cdots < x_{r+2}$ and such that $E_p(x_i) = -E_P(x_{i+1}) = \pm\|E\|$ for $i = 1, 2, \ldots, (r+1)$.

The Parks–McClellan Algorithm

The alternation theorem gives necessary and sufficient conditions on the error for optimality in the Chebyshev or minimax sense. Although the theorem does not state explicitly how to find the optimum filter, the conditions that are presented serve as the basis for an efficient algorithm for finding it.

From the alternation theorem, we know that the optimum filter $A_e(e^{j\omega})$ will satisfy the set of equations

$$W(\omega_i)[H_d(e^{j\omega_i}) - A_e(e^{j\omega_i})] = (-1)^{i+1}\delta, \qquad i = 1, 2, \ldots, (L+2), \tag{7.99}$$

where δ is the optimum error .

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^L & \dfrac{1}{W(\omega_1)} \\ 1 & x_2 & x_2^2 & \cdots & x_2^L & \dfrac{-1}{W(\omega_2)} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & x_{L+2} & x_{L+2}^2 & \cdots & x_{L+2}^L & \dfrac{(-1)^{L+2}}{W(\omega_{L+2})} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ \delta \end{bmatrix} = \begin{bmatrix} H_d(e^{j\omega_1}) \\ H_d(e^{j\omega_2}) \\ \vdots \\ H_d(e^{j\omega_{L+2}}) \end{bmatrix}, \tag{7.100}$$

where $x_i = \cos\omega_i$. This set of equations serves as the basis for an iterative algorithm for finding the optimum $A_e(e^{j\omega})$. The procedure begins by guessing a set of alternation frequencies ω_i for $i = 1, 2, \ldots, (L+2)$. Note that ω_p and ω_s are fixed and are necessarily members of the set of alternation frequencies. Specifically, if $\omega_\ell = \omega_p$, then $\omega_{\ell+1} = \omega_s$.

Parks and McClellan (1972a, 1972b) found that, for the given set of the extremal frequencies,

$$\delta = \frac{\sum_{k=1}^{L+2} b_k H_d(e^{j\omega_k})}{\sum_{k=1}^{L+2} \frac{b_k(-1)^{k+1}}{W(\omega_k)}}, \tag{7.101}$$

where

$$b_k = \prod_{\substack{i=1 \\ i\neq k}}^{L+2} \frac{1}{(x_k - x_i)} \tag{7.102}$$

and, as before, $x_i = \cos\omega_i$.

Now, since $A_e(e^{j\omega})$ is known to be an Lth-order trigonometric polynomial, we can interpolate a trigonometric polynomial through $(L+1)$ of the $(L+2)$ known values $E(\omega_i)$ (or equivalently, $A_e(e^{j\omega_i})$). Parks and McClellan used the Lagrange interpolation formula to obtain

$$A_e(e^{j\omega}) = P(\cos\omega) = \frac{\sum_{k=1}^{L+1} [d_k/(x - x_k)]C_k}{\sum_{k=1}^{L+1} [d_k/(x - x_k)]}, \tag{7.103a}$$

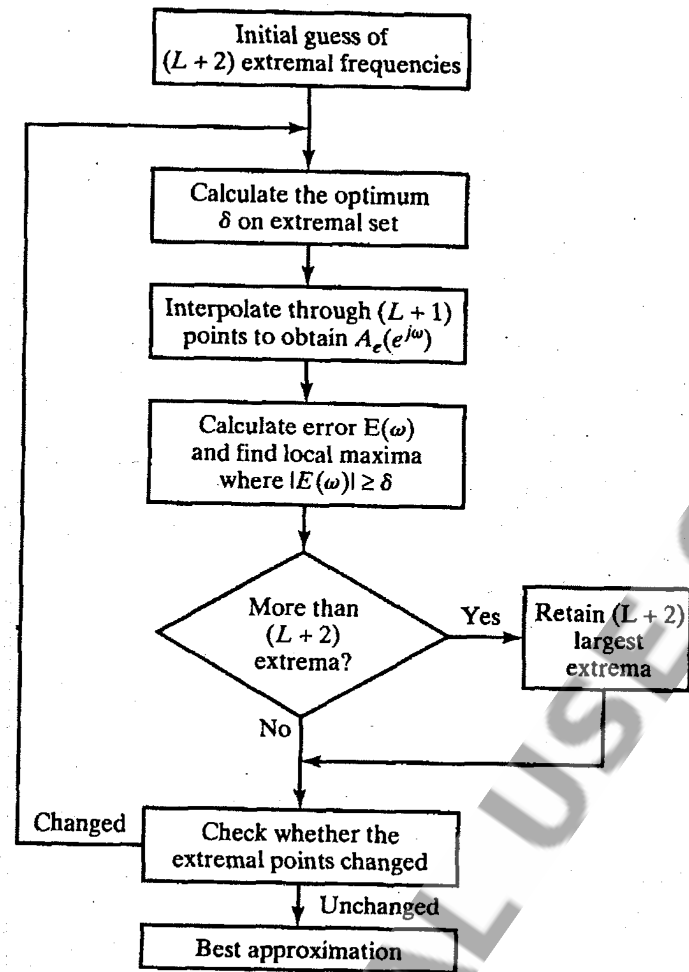
where $x = \cos\omega$, $x_i = \cos\omega_i$,

$$C_k = H_d(e^{j\omega_k}) - \frac{(-1)^{k+1}\delta}{W(\omega_k)}, \tag{7.103b}$$

and

$$d_k = \prod_{\substack{i=1 \\ i\neq k}}^{L+1} \frac{1}{(x_k - x_i)} = b_k(x_k - x_{L+2}). \tag{7.103c}$$

The polynomial of Eq. (7.103a) can be used to evaluate $A_e(e^{j\omega})$ and also $E(\omega)$ on a dense set of frequencies in the passband and stopband. If $|E(\omega)| \leq \delta$ for all ω in the passband and stopband, then the optimum approximation has been found. Otherwise we must find a new set of extremal frequencies.



After the algorithm has converged, the impulse response can be computed from samples of the polynomial representation using the discrete Fourier transform.