VI. Design of FIR Filters

This chapter deals with the design of digital filters with finite impulse response. In a similar way to IIR filter design, the design of FIR filters begins with the filter specifications. These include constraints on the magnitude and/or phase of the frequency response and the filter order. This is followed by the determination of FIR filter coefficients that produce an acceptable filter. Then, the filter can be implemented in hardware or software with an appropriate structure.

Consider the design of a filter with a cutoff frequency w_c. The frequency response of an ideal low-pass filter with linear phase and a cutoff frequency w_c is

$$H_d(e^{jw}) = \begin{cases} e^{-j\alpha w} & , \ |w| \leq w_c \\ 0 & , \ w_c < |w| \leq \pi \end{cases}$$

which has an impulse response given by

$$h_d[n] = \frac{\sin(n-\alpha)w_c}{\pi(n-\alpha)}$$

Because this filter is noncausal and unstable, it cannot be realized. Therefore, we need to relax the constraints on the frequency response and allow some deviation from the ideal response.

One way of developing a realizable approximation of an FIR digital filter is to truncate the impulse response, resulting in a magnitude response with acceptable tolerances in the passband and in the stopband.

The specifications for a low-pass filter are :

$$1 - \delta_p < |H(e^{j\omega})| \leq 1 + \delta_p , \quad 0 \leq |\omega| < \omega_p$$

$$|H(e^{j\omega})| \leq \delta_s , \quad \omega_s \leq |\omega| < \pi ,$$

where ω_p is the passband cutoff frequency, ω_s is the stopband cutoff frequency, δ_p is the passband deviation and δ_s is the stopband deviation, as illustrated in Fig. 1 below.

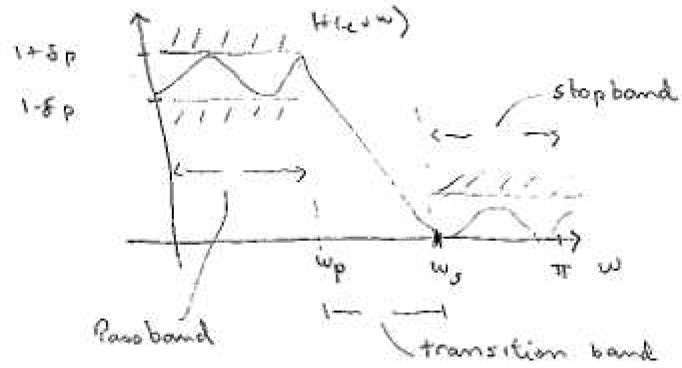


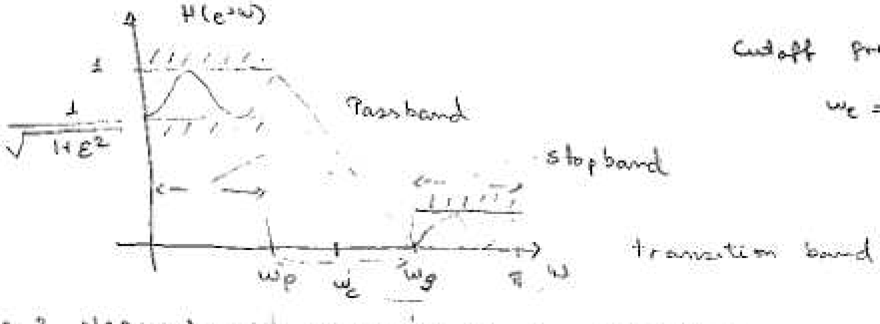
Fig. 1 Typical magnitude specifications for a digital lowpass filter

The passband and stopband deviations are given in decibels (dB).

$$\alpha_p = -20 \log (1 - \delta_p)$$

$$\alpha_s = -20 \log (\delta_s)$$

The interval $[\omega_p, \omega_s]$ is called the transition band. The frequency response $H(e^{j\omega})$ of a digital filter is a periodic function of ω with a period 2π. In most cases, the filter coefficients are real and the magnitude is an even function of ω. As a result, the filter specifications are given for the range $0 \leq |\omega| \leq \pi$. The magnitude response specifications can also be given in a normalized form, as in Fig. 2 below



Cutoff frequency:

$$\omega_c = \frac{(\omega_p + \omega_s)}{2}$$

Fig. 2 Normalized magnitude specifications

Unlike IIR filter designs, the FIR filter designs do not have any connection with the design of analog filters. Specifically, the design of FIR filters is based on approximations of the specified magnitude response. The frequency response of an FIR filter is given by

$$H(e^{jw}) = \sum_{n=0}^{N} h[n] \cdot e^{-jwn}$$

An important step in the design of FIR filters is the estimation of the filter order N. In particular, the order of a lowpass FIR filter can be estimated from the following specifications: normalized edge angular frequency w_p, normalized edge angular frequency w_s, peak passband ripple δ_p and peak stopband ripple δ_s. A formula developed by Kaiser is given by

$$N \cong \frac{-20 \log_{10}\left(\sqrt{\delta_p \delta_s}\right) - 13}{14.6\,(w_s - w_p)/2\pi}$$

which works well for filter with moderate passband. A formula for a narrowband filter is described by

$$N \cong \frac{-20\log_{10}(\delta_s) + 0.221}{27\,(w_s - w_p)/2\pi}$$

Ex: Estimate the order of a linear-phase lowpass FIR filter with the following specifications:

$F_p = 1.8$ kHz $\rightarrow$ $\delta_p = 1 - 10^{-\alpha_p/20} = 0.11$

$F_{sb} = 2$ KHz $\rightarrow$ $\delta_{sb} = 10^{-\alpha_s/20} = 0.017$

$\alpha_p = 0.1$ dB

$\alpha_s = 35$ dB

$F_s = 12$ KHz

$$N \cong \frac{-20 \log_{10}\left(\sqrt{\delta_p \delta_s}\right) - 13}{14.6\,(w_s - w_p)/2\pi} = 98.27$$

$\rightarrow$ $\boxed{N = 99}$ $\rightarrow$ Type 2 FIR filter (odd order)

$\rightarrow$ Type 1 FIR filter $\rightarrow$ N=100 (even order)

In the normalized case, the maximum value of the magnitude in the passband is set to unity, and the maximum passband deviation, denoted as $1/\sqrt{1+\varepsilon^2}$, is given by the minimum value of the magnitude in the passband, whereas the maximum stopband magnitude is $1/A$.

For the normalized specification, the maximum value of the gain function or the minimum value of the loss function is 0 dB. The quantity α_{max} is called maximum passband attenuation and is given by

$$\alpha_{max} = 20 \log_{10} \left(\sqrt{1+\varepsilon^2} \right) dB$$

The passband and stopband frequencies are specified in Hz in most applications along with the sampling rate used. Since all filter designs employ normalized frequencies ω_p and ω_s, the specified critical frequencies need to be normalized before a specific design algorithm is applied.

Let F_s denote the sampling frequency in Hz and F_p and F_{sb}, respectively, the passband and stopband edge frequencies in Hz. The normalized angular frequencies are then given by

$$\omega_p = \frac{\Omega_p}{F_s} = \frac{2\pi F_p}{F_s} = 2\pi F_p \cdot T_s$$

$$\omega_{sb} = \frac{\Omega_{sb}}{F_s} = 2\pi \frac{F_{sb}}{F_s} = 2\pi F_{sb} T_s$$

An FIR filter has the system function:

$$H[z] = \sum_{n=0}^{N} h[n] \cdot z^{-n},$$

where the order N should be kept to the smallest possible. Moreover, if a linear phase is desired the FIR filter coefficients must satisfy the constraint $h[n] = \pm h[n-N]$.

Ex: Estimate the order of a linear-phase bandpass FIR filter with the following specifications:

$F_{p_1} = 0.35$ KHz , $\delta_p = 0.002$

$F_{p_2} = 1$ KHz , $\delta_s = 0.001$

$F_{s_1} = 0.3$ KHz , $f_s = 10$ KHz

$F_{s_2} = 1.1$ KHz

Since the widths of the transition bands are not equal, we use the width of the smallest transition band to compute N.

$$N \cong \frac{-20\log_{10}\left(\sqrt{\delta_s \delta_p}\right) - 13}{14.6\left(\omega_s - \omega_p\right)/2\pi} = 424.37 \rightarrow$$

$$\rightarrow \boxed{N = 425} \longrightarrow \text{Type 2 filter.}$$

After a digital filter is designed, the corresponding system function H[z] has to be scaled in magnitude before implementation. In magnitude scaling, the transfer function is multiplied by a constant K, so that the maximum magnitude of K · H[z] is unity in the passband.

A. FIR Filter Design using Windows

Let $h_d[n]$ be the impulse response of an ideal frequency selective filter with linear phase whose frequency response is given by

$$H_d(e^{jw}) = A(e^{jw})\, e^{-j(\alpha w - \beta)}$$

Because $h_d[n]$ will be generally infinite in length and unrealizable, it is necessary to find an FIR approximation to $H_d(e^{jw})$.

For example, the ideal lowpass filter has a frequency response given by

$$H_{LP}(e^{jw}) = \begin{cases} 1, & |w| \leq w_c \\ 0, & w_c < w \leq \pi \end{cases} \qquad \text{with} \quad w_c = \frac{(w_s + w_p)}{2}$$

and the corresponding impulse response is described by

$$h_{LP}[n] = \frac{\sin w_c n}{\pi n}, \quad -\infty \leq n \leq \infty,$$

which is doubly infinite, not absolutely summable and unrealizable.

A possible approach is to set all the coefficients of the filter outside the range $-M \leq n \leq M$ to zero, which results in a noncausal approximation of length $N = 2M+1$. If we shift these coefficients to the right, we obtain a causal FIR lowpass filter:

$$\hat{h}_{LP}[n] = \begin{cases} \dfrac{\sin(w_c(n-M))}{\pi(n-M)}, & 0 \leq n \leq N-1 \\ 0, & \text{otherwise} \end{cases}$$

Other filters can be described in a similar way.

The highpass filter $h_{HP}[n]$ has an ideal impulse response given by

$$h_{HP}[n] = \begin{cases} 1 - \dfrac{w_c}{\pi}, & \text{for } n = 0 \\ -\dfrac{\sin(w_c n)}{\pi n}, & \text{for } |n| > 0 \end{cases}$$

The ideal bandpass filter $h_{BP}[n]$ with cutoff frequencies w_{c1} and w_{c2} has an impulse response described by

$$h_{BP}[n] = \frac{\sin(w_{c2} n)}{\pi n} - \frac{\sin(w_{c1} n)}{\pi n}, \quad |n| > 0,$$

and the bandstop filter with ideal characteristics and cutoff frequencies w_{c1} and w_{c2} has an impulse response given by

$$h_{BS}[n] = \begin{cases} 1 - \dfrac{(w_{c2} - w_{c1})}{\pi}, & \text{for } n=0 \\ \dfrac{\sin(w_{c1} n)}{\pi n} - \dfrac{\sin(w_{c2} n)}{\pi n}, & \text{for } |n|>0 \end{cases}$$

Two other types of FIR system that are useful in various applications are the Hilbert transformer and the differentiator. The ideal Hilbert transformer, also called a 90-degree phase shifter, is characterized by the following frequency response:

$$H_{HT}(e^{jw}) = \begin{cases} j, & -\pi < w < 0 \\ -j, & 0 < w < \pi \end{cases}$$

The impulse response $h_{HT}[n]$ of the Hilbert transformer is

$$h_{HT}[n] = \mathcal{F}^{-1}\{H_{HT}(e^{jwn})\}$$

$$= \begin{cases} 0, & \text{for } n \text{ even} \\ \dfrac{2}{\pi n}, & \text{for } n \text{ odd} \end{cases}$$

The ideal discrete-time differentiator is characterized by the frequency response:

$$H_{DIF}(e^{jw}) = jw, \quad 0 \leq |w| \leq \pi,$$

and its impulse response $h_{DIF}[n]$ is described by

$$h_{DIF}(n) = \begin{cases} 0, & n=0 \\ \dfrac{\cos \pi n}{n}, & |n|>0 \end{cases}$$

Similarly to the case of the ideal lowpass filter, the above FIR filters and systems are not absolutely summable and are therefore unrealizable. They can be realized by truncating their impulse responses and shifting the truncated coefficients to the right appropriately.

The truncation of the coefficients of the impulse response of filters result in an oscillatory behaviour of their magnitude responses known as the Gibbs phenomenon. An illustration of the Gibbs phenomenon is shown in fig. 3 below

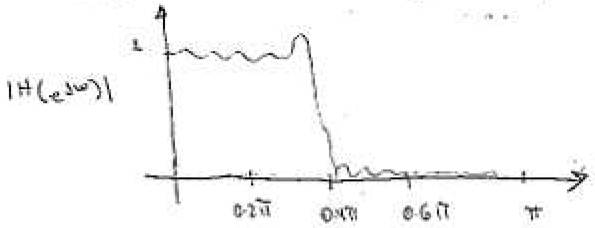


Fig. 3 Magnitude responses of lowpass filters using the truncated response of $\hat{h}_{LP}[n] = \begin{cases} \frac{\sin(w_c(n-M))}{\pi(n-M)} & , 0 \leq n \leq N-1 \\ 0 & , \text{otherwise} \end{cases}$ with $N = 61$ and $M = \frac{N}{2}$.

The height of the ripples in both passband and stopband remain the same regardless of the filter length. A similar behaviour occurs with other types of FIR filters and systems.

In order to mitigate the Gibbs phenomenon and reduce the height of the ripples, FIR filter design techniques based on windows have been developed. In the window design method, the filter is designed by windowing the impulse response of the ideal filter $h_d[n]$ as described by

$$h[n] = h_d[n] \cdot w[n] ,$$

where $w[n]$ is a finite-length window that is equal to zero outside the interval $0 \leq n \leq N$ and is symmetric about its midpoint:

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

The effect of the window on the frequency response may be seen by the complex convolution theorem:

$$H(e^{jw}) = \frac{1}{2\pi} H_d(e^{jw}) * W(e^{jw})$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} H_d(e^{j\theta}) \cdot W(e^{j(w-\theta)})\, d\theta ,$$

where the ideal frequency response is smoothed by the DTFT of $W(e^{jw})$.

There are different types of windows that can be used in the filter design, which are listed in Table I below

Table I – Windows for use in filter design

Rectangular	$w[m] = \begin{cases} 1, & 0 \le m \le N \\ 0, & \text{else} \end{cases}$
Hanning or Hann	$w[m] = \begin{cases} 0.5 - 0.5\cos\left(\frac{2\pi m}{N}\right), & 0 \le m \le N \\ 0, & \text{else} \end{cases}$
Hamming	$w[m] = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi m}{N}\right), & 0 \le m \le N \\ 0, & \text{else} \end{cases}$
Blackman	$w[m] = \begin{cases} 0.42 - 0.5\cos\left(\frac{2\pi m}{N}\right) + 0.08\cos\left(\frac{4\pi m}{N}\right), & 0 \le m \le N \\ 0, & \text{else} \end{cases}$

The quality or accuracy of the approximation to a desired response, $H_d(e^{jw})$, is determined by two factors:

1. The width of the main lobe of $W(e^{jw})$
2. The peak side-lobe amplitude of $W(e^{jw})$.

This is illustrated in Fig. 4 below.

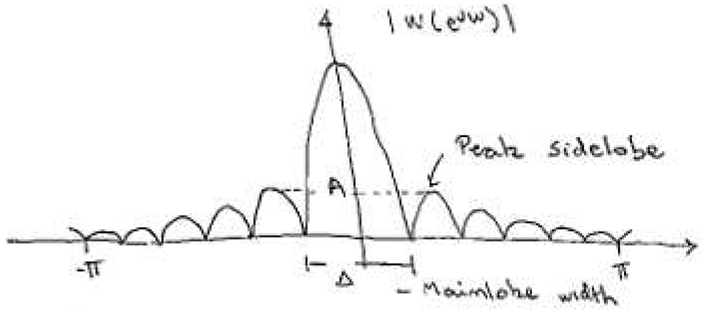


Fig. 4. The DTFT of a typical window characterised by the width of its main lobe, Δ, and the peak amplitude of its side lobes, A, relative to the amplitude of $W(e^{jw})$ at $w=0$.

Ideally, the main lobe should be narrow, and the side-lobe amplitude should be small. However, for a fixed-length window, these cannot be minimized independently. Some general properties of windows are as follows:

1. As the length N of the window increases, the width of the main lobe decreases, which results in a decrease in the transition width between passbands and stopbands. This relationship is given approximately by

$$N \Delta f = c,$$

where Δf is the transition width and c is a parameter that depends on the window.

2. The peak side-lobe amplitude of the window is determined by the shape of the window, and it is essentially independent of the window length.

3. If the window shape is changed to decrease the side-lobe amplitude, the width of the main lobe will generally increase.

The sidelobe amplitudes of several windows along with the approximate transition width and stopband attenuation are shown in Table II below

Table II. Peak side-lobe amplitude and approximate transition width and stopband attenuation of windows.

Window	Side-lobe Amplitude (dB)	Transition Width (Δf)	Stopband Attenuation (dB)
Rectangular	-13	0.9 /N	-21
Hanning	-31	3.1 /N	-44
Hamming	-41	3.3 /N	-53
Blackman	-57	5.5 /N	-74

Ex: Suppose that we would like to design an FIR linear phase low-pass filter according to the specifications:

$$0.99 \leq |H(e^{jw})| \leq 1.01 \quad , \quad 0 \leq |w| \leq 0.19\pi$$

$$|H(e^{jw})| \leq 0.01 \quad , \quad 0.21\pi \leq |w| \leq \pi$$

For a stopband attenuation of $20 \log (0.01) = -40$ dB, we may use a Hanning window. The Hamming and Blackman windows are also possible but they would over design the filter and produce a larger stopband attenuation at the expense of an increase in the transition width.

Because the specification calls for a transition width of

$\Delta w = w_s - w_p = 0.02\pi$ or $\Delta f = \frac{\Delta w}{2\pi} = 0.01$ with

$$N \cdot \Delta f = 3.1 \quad (\text{see Table II})$$

For a Hanning window, an estimate of the required order of the filter is

$$N = \frac{3.1}{\Delta f} = 310$$

The last step is to find the impulse response of the ideal low-pass filter that is to be windowed. With a cutoff frequency of $\boxed{w_c = \left(\frac{w_s + w_p}{2}\right)} = 0.2\pi$ and a delay of $\alpha = \frac{N}{2} = 155$, the impulse response is

$$h_d[n] = \frac{\sin\left(0.2\pi\,[n - 155]\right)}{(n-155)\pi}$$

An alternative to using fixed windows is to employ an adjustable window that can meet different sets of specifications. The Kaiser family of adjustable windows is defined by

$$w[n] = \frac{I_0\left(\beta\left[1 - \{[n-\alpha]/\alpha^2\}^2\right]^{1/2}\right)}{I_0(\beta)}, \quad 0 \le n \le N,$$

where $\alpha = \frac{N}{2}$ and $I_0(\cdot)$ is a zeroth-order modified Bessel function of the first kind, which may be easily generated using the power series expansion

$$I_0(x) = 1 + \sum_{k=1}^{\infty}\left[\frac{(x/2)^k}{k!}\right]^2$$

The parameter β determines the shape of the window and thus controls the trade-off between main-lobe width and side-lobe amplitude. A Kaiser window is nearly optimum in the sense of having the most energy in its main lobe for a given side-lobe amplitude. Table III illustrates the effect of changing β.

Table III - Characteristics of the Kaiser Window as a function of β

Parameter β	Sidelobe (db)	Transition Width ($N\Delta f$)	Stopband Attenuation (dB)
2.0	-19	1.5	-29
3.0	-24	2.0	-37
4.0	-30	2.6	-45
5.0	-37	3.2	-54

There are two empirically derived relationships for the Kaiser window that facilitate its to design FIR filters. The first relates the stopband ripple of a lowpass filter, $\alpha_s = -20\log(\delta_s)$, to β as given by

$$\beta = \begin{cases} 0.11\,(\alpha_s - 8.7), & \alpha_s > 50 \\ 0.54\,(\alpha_s - 21)^{0.4} + 0.07(\alpha_s - 21), & 21 \le \alpha_s \le 50 \\ 0.0, & \alpha_s < 21 \end{cases}$$

The second relates N to the transition width Δf and the stopband attenuation α_s :

$$N = \frac{\alpha_s - 7.95}{14.36\,\Delta f}\ , \quad \alpha_s \geq 21$$

Note that if $\alpha_s < 21$ dB, a rectangular window may be used ($\beta = 0$) and $N = 0.9/\Delta f$.

Ex: Design a low-pass filter with a cutoff frequency $w_c = \frac{\pi}{4}$, a transition width $\Delta\omega = 0.02\pi$, and a stopband ripple $\delta_s = 0.01$

Because $\alpha_s = -20 \log(0.01) = -40$, the Kaiser window parameter is

$$\beta = 0.58\,(40-21)^{0.4} + 0.07\,(40-21) = 3.4$$

With $\Delta f = \frac{\Delta\omega}{2\pi} = 0.01$, we have

$$N = \frac{40 - 7.95}{14.36 \cdot (0.01)} = 224$$

Therefore, we have $h[m] = h_d[m] \cdot w[m]$,

where
$$h_d[m] = \frac{\sin\left([m-112]\,\pi/4\right)}{[m-112]\pi}$$

and
$$w[m] = \frac{I_0\left(3.4\left(1 - \left([m-\frac{N}{2}]/\frac{N}{2}\right)^2\right)^{1/2}\right)}{I_0(\beta)}\ , \quad 0 \leq m \leq N$$

· Although it is simple to design FIR filters with the window design method, there are some limitations with it. First, it is necessary to find a closed-form expression for $h_d[m]$. Second, the transition widths between frequency bands and their ripples will be approximately the same. As a result, the windows method requires the filter to be designed to the tightest tolerances in all the bands. Moreover, they are not optimum in the sense that the resulting ripples are not the smallest given a filter order and a set of cutoff frequencies.

B. Computer-Aided Design of Digital Filters

In this section, we consider the computed-aided design of digital filters which rely on optimization techniques to minimize the error between the desired frequency response $H_d(e^{jw})$ and that of the computer generated filter.

Let $H(e^{jw})$ denote the frequency response of the digital transfer function $H[z]$ to be designed so that it approximates the desired frequency response $H_d(e^{jw})$, given as a piecewise linear function of w in some sense. The goal is to determine iteratively the coefficients of the transfer function so that the difference between $H(e^{jw})$ and $H_d(e^{jw})$ for all values of w over closed subintervals of $0 \le w \le \pi$ is minimized. This difference is often specified as a weighted error function $E(w)$ given by

$$E(w) = W(e^{jw})\left[H(e^{jw}) - H_d(e^{jw}) \right],$$

where $W(e^{jw})$ is a positive weighting function specified by the designer. There are several criteria that use $E(w)$ to design filters.

The minimax criterion is an approximation measure that minimizes the peak absolute value of the weighted error $E(w)$:

$$\varepsilon = \max_{w \in R} |E(w)|,$$

where R is the set of disjoint frequency bands in the range $0 \le w \le \pi$, on which the desired frequency response is defined. In filtering applications, R is composed of the passbands and stopbands of the filter to be designed. For a lowpass filter design, R is the disjoint union of the freq. ranges $(0, w_p)$ and (w_s, π), where w_p and w_s are the passband and stopband edges, respectively.

The least-p criterion is the approximation measure that minimizes the integral of the p^{th} power of the weighted error function $E(\omega)$:

$$\varepsilon = \int_{\omega \in R} \left| W(e^{j\omega}) \left(H(e^{j\omega}) - H_d(e^{j\omega}) \right) \right|^p d\omega,$$

where R is the specified frequency range and p is a positive integer. For simplicity, the least-squares criterion obtained with $p=2$ is often used. It can be shown that as $p \to \infty$, the least pth solution approaches the minimax solution. Note also that $W(e^{j\omega}) = 1$ is usually not used to avoid large peak errors.

In practice, the integral error measure above is replaced by a finite sum given by

$$\varepsilon = \sum_{i=1}^{K} \left\{ W(e^{j\omega_i}) \left(H(e^{j\omega_i}) - H_d(e^{j\omega_i}) \right) \right\}^p,$$

where ω_i, $1 \le i \le K$, is a suitably chosen dense grid of digital angular frequencies. The least-squares criterion obtained from the above equation with $p=2$ is often used for simplicity.

In the case of linear-phase FIR filter designs, $H(e^{j\omega})$ and $H_d(e^{j\omega})$ are zero-phase frequency responses. On the other hand, for IIR filter designs, these functions are replaced with their magnitude functions. The design objective is thus to iteratively adjust the filter parameters so that ε is a minimum. For equiripple linear phase filters, we can employ the Parks-McClellan algorithm which results in an error function $E(\omega)$ with an equiripple behaviour in the frequency range of interest. Other designs include criteria such as the least-mean-square error and the constrained least-square design.

i) Design based on the least-mean-square error

For the design of a linear-phase FIR filter with a minimum mean-square error criterion, we employ the error measure given by

$$\mathcal{E} = \sum_{i=1}^{K} \left[W(\omega_i) \left(H(\omega_i) - H_d(\omega_i) \right) \right]^2,$$

where $H(\omega)$ is the amplitude response of the designed filter, $H_d(\omega_i)$ is the desired amplitude response and $W(\omega_i)$ is the weighting function.

The amplitude response for linear-phase FIR filters can be expressed in the form

$$H(\omega) = Q(\omega) \sum_{k=0}^{L} a[k] \cos(\omega k),$$

where
$$Q(\omega) = \begin{cases} 1, & \text{for type 1} \\ \cos(\omega/2), & \text{for type 2} \\ \sin(\omega), & \text{for type 3} \\ \sin(\omega/2), & \text{for type 4.} \end{cases}$$

$$\text{Type I} \begin{cases} a[0] = h[M] \\ a[k] = 2\,h[M-k], & 1 \le k \le M \end{cases} \qquad \text{Type II} \begin{cases} a[k] = 2\,h\left[\frac{2M+1}{2} - k\right], & 1 \le k \le \frac{2M+1}{2}. \end{cases}$$

$$\text{Type III} \begin{cases} a[k] = 2\,h[M-k], & 1 \le k \le M \end{cases} \qquad \text{Type IV} \begin{cases} a[k] = 2\,h\left[\frac{2M+1}{2} - k\right], & 1 \le k \le \frac{2M+1}{2} \end{cases}$$

and
$$L = \begin{cases} M & \text{for type 1} \\ \frac{2M-1}{2} & \text{for type 2} \\ M-1 & \text{for type 3} \\ \frac{2M-1}{2} & \text{for type 4} \end{cases}$$

The mean-square error of $\mathcal{E}$ is a function of the filter parameters $a[k]$. To arrive at the minimum value of $\mathcal{E}$, we set

$$\frac{\partial \mathcal{E}}{\partial a[k]} = 0, \quad 0 \le k \le L,$$

which results in a set of $(L+1)$ linear equations that can be solved for $a[k]$.

Let us consider the design of a Type 1 linear-phase FIR filter. In this case, $Q(\omega) = 1$, and $L = M$. The expression for the mean-square error then takes the form

$$\mathcal{E} = \sum_{i=1}^{K} \left\{ W(\omega_i) \left[\sum_{k=0}^{M} a[k] \cos(\omega_i k) - H_d(\omega_i) \right] \right\}^2$$

$$= \sum_{i=1}^{K} \left\{ \sum_{k=0}^{M} W(\omega_i)\, a[k] \cos(\omega_i k) - W(\omega_i) H_d(\omega_i) \right\}^2$$

Using the notation described by

$$\underline{H} = \begin{bmatrix} W(\omega_1) & W(\omega_1)\cos(\omega_1) & \cdots & W(\omega_1)\cos(M\omega_1) \\ W(\omega_2) & W(\omega_2)\cos(\omega_2) & \cdots & W(\omega_2)\cos(M\omega_2) \\ \vdots & \vdots & \ddots & \vdots \\ W(\omega_K) & W(\omega_K)\cos(\omega_K) & \cdots & W(\omega_K)\cos(M\omega_K) \end{bmatrix}$$

$$\underline{a} = [\, a[0] \;\; a[1] \;\; \ldots \;\; a[M] \,]^T$$

and

$$\underline{d} = [\, W(\omega_1) H_d(\omega_1) \;\; W(\omega_2) H_d(\omega_2) \;\; \ldots \;\; W(\omega_K) H_d(\omega_K) \,]^T.$$

We can then express $\mathcal{E}$ in the form

$$\mathcal{E} = \underline{e}^T \underline{e} = \| \underline{e} \|^2,$$

where $\underline{e} = \underline{H}\,\underline{a} - \underline{d}$

The minimum mean-square solution is then obtained by solving the normal equations:

$$\underline{H}^T \underline{H}\, \underline{a} = \underline{H}^T \underline{d}$$

$$\Rightarrow \boxed{\underline{a} = (\underline{H}^T \underline{H})^{-1} \underline{H}^T . \underline{d}}$$

If $K \geq M$ the above should be solved using an iterative method as the direct solution is often ill-conditioned. This design approach can be used to design a linear-phase FIR filter meeting any arbitrary shaped desired response.

ii) Design based on the constrained least-squares criterion

In many design problems, the amplitude response of the filter must satisfy some constraints which may include spectral nulls and specific requirements. In order to illustrate this approach, let us assume that the filter to be designed is a type I linear-phase filter of order $N = 2M$ subject to a set of r constraint described by

$$\underset{r\times(M+1)}{\underline{G}}\ \underset{(M+1)\times 1}{\underline{a}} = \underset{r\times 1}{\underline{d}}$$

The constrained least-squares design method minimizes the squared error given by

$$\varepsilon = \left(\frac{1}{\pi} \int_0^{\pi} W(\omega) \left[H(\omega) - H_d(\omega) \right]^2 d\omega \right)^{1/2}$$

subject to $\underline{G}\,\underline{a} = \underline{d}$,

where $H_d(\omega)$ is the desired amplitude response and $W(\omega)$ is a weighting function

To minimize ε^2 subject to the constraints, we first form the Lagrangian

$$\mathcal{L} = \varepsilon^2 + \underline{\mu}^T . (\underline{G}.\underline{a} - \underline{d}),$$

where $\underline{\mu} = [\mu_1\ \mu_2 \ldots \mu_r]^T$ is the vector with Lagrange multipliers.

We can derive the necessary conditions for the minimization of ε^2 by setting the derivatives of $\mathcal{L}$ with respect to the filter parameters $a[n]$ and the Lagrange multipliers μ_i to zero resulting in the following equations:

$$\underline{R}.\underline{a} + \underline{G}^T.\underline{\mu} = \underline{c}$$

$$\underline{G}.\underline{a} = \underline{d},$$

where $c[0] = \frac{1}{\pi}\int_0^{\pi} W(\omega) H_d(\omega)\, d\omega$ and $c[k] = \frac{1}{\pi}\int_0^{\pi} W(\omega) H_d(\omega) \cos(k\omega) d\omega$, $1 \le k \le M$

and the (i,k)th element $R_{i,k}$ of $\underline{R}$ is given by

$$R_{i,k} = \int_0^{\pi} W(\omega) \cdot \cos(i\omega) \cdot \cos(k\omega)\, d\omega$$

The two matrix equations can be written as

$$\begin{bmatrix} \underline{R} & \underline{G}^T \\ \underline{G} & \underline{0} \end{bmatrix} \begin{bmatrix} \underline{a} \\ \underline{\mu} \end{bmatrix} = \begin{bmatrix} \underline{c} \\ \underline{d} \end{bmatrix}$$

Solving the above equations we obtain

$$\underline{\mu} = (\underline{G}\,\underline{R}^{-1}\underline{G}^T)\,(\underline{G}\,\underline{R}^{-1}\underline{c} - \underline{d})$$

$$\underline{a} = \underline{R}^{-1}(\underline{c} - \underline{G}^T \cdot \underline{\mu}).$$

When the integrals needed to form $\underline{R}$ and $\underline{c}$ cannot be calculated in a simple way, we approximate them as

$$\underline{R} \cong \underline{H}^T \cdot \underline{H} \quad \text{and} \quad \underline{c} \cong \underline{H}^T \cdot \underline{d}.$$

One useful application of the constrained least-squares method to design filters is an approach that takes into account both the squared error and the peak-ripple error. This constrained least-squares approach allows a compromise between the squared error and the peak-ripple error. Other sophisticated design techniques include inequality constraints which require the use of iterative algorithms to obtain a suitable filter design.