# Digital Filters, Chapter 8

Basically, there are two categories for the design of filters: IR and IIR.  FIR can have linear phase within their passband.  In general, an IIR filter has lower side lobes in the stop-band. Therefore, if some phase distortion can be tolerated, the IIR filter is preferable, mainly because its implementation is easier, fewer parameters and less memory is required.

## Causality

An ideal low pass filter is: $H(\omega) = \begin{cases} 1 & |\omega| \le \omega_c \\ 0 & \omega_c \le |\omega| \le \pi \end{cases}$
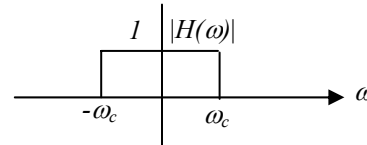


$$h(n) = \begin{cases} \dfrac{\omega_c}{\pi} & n = 0 \\ \dfrac{\omega_c}{\pi} \dfrac{\sin \omega_c n}{\omega_c n} & n \ne 0 \end{cases}$$

Clearly this *h(n)* is non-casual and hence it is impractical to build.

Paley and Wiener proved that if $\int_{-\pi}^{\pi} |\ell n|H(\omega)||\, d\omega < \infty$, then *h(n)* would be casual.  This means that |*H(ω)*| can be zero at some frequencies but cannot be zero over a finite band of frequencies. Therefore, any ideal filter is non-causal.

Causality also makes the |*H(ω)*| and *θ(ω)* to be interdependent.  In general, it imposes the following:

1)  |*H(ω)*| cannot be zero over a band of frequencies.

2)  |*H(ω)*| cannot have an infinitely sharp cutoff from pass-band to stop-band.

3)  *H_r(ω)* and *H_I(ω)* are interdependent and therefore |*H(ω)*| and *θ(ω)* are also interdependent.
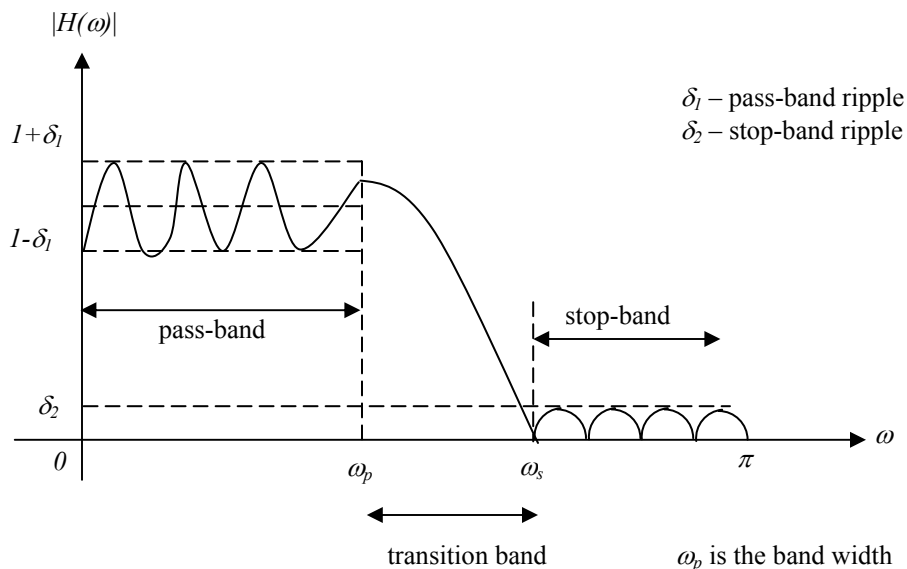
To ensure that the FIR has a linear phase, it should satisfy:

$h(n) = \pm h(M - 1 - n)$ for $n = 0,....., M - 1$ (to be symmetric/anti-symmetric about mid-point).

$z^{-(M-1)}H(z^{-1}) = \pm H(z)$ → means that the roots of $H(z)$ are also roots of $H(z^{-1})$ – the root (or zeros) have to be in reciprocal or complex conjugates.

## Design of the Linear Phase FIR Filters Using Windows

In this method, we begin with the desired response features of $H_d(\omega)$ and determine its $h_d(n)$ and

then truncate it to become FIR. $H_d(\omega) = \sum_{n=0}^{\infty} h_d(n)e^{-j\omega n} \rightarrow h_d(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_d(\omega)e^{j\omega n}d\omega$. Now

truncation of $h_d(n)$ is as if we multiply it by a rectangle window: $w(n) = \begin{cases} 1 & n = o,...M - 1 \\ o & else \end{cases}$
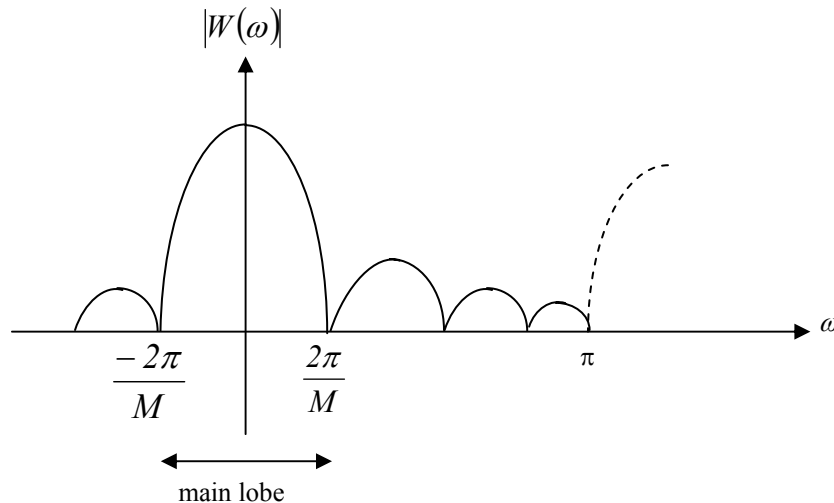
Then $h(n) = h_d(n) \cdot w(n) \rightarrow H(\omega) = H_d(\omega) * W(\omega)$.

For Rectangular window $W(\omega) = \sum_{n=0}^{M-1} e^{j\omega n} = \frac{1 - e^{-j\omega M}}{1 - e^{-j\omega}} \Rightarrow W(\omega) = e^{-j\omega(M-1)/2} \frac{\sin(\omega M/2)}{\sin(\omega/2)}$.

Therefore the magnitude of the window is $|W(\omega)| = \frac{|\sin(\omega M/2)|}{|\sin(\omega/2)|}$ for $-\pi \le \omega \le \pi$ and it has a

piecewise linear phase:

$\theta(\omega) = \begin{cases} -\omega\left(\dfrac{M-1}{2}\right) & \sin\dfrac{\omega M}{2} > 0 \qquad \because \sin\dfrac{\omega}{2} \text{ is positive for } \omega < \pi \text{ anyway} \\ -\omega\left(\dfrac{M-1}{2}\right) + \pi & else \end{cases}$

As the *M* increases, the main lobe becomes narrower. However the peak of side lobes remain almost unaffected because their width decreases as *M* increases but their peak also increase. Therefore, their area and their normalized peak remain unaffected. As discussed before, there are many other windows to remedy the side-lobe effects of the Rectangular window. Summary of some windows is in page 626 – 627.

Starting with ideal filters in time domain and then truncate it by a window. Let's consider an ideal LPF:

$$H_d(\omega) = \begin{cases} 1 \cdot e^{-j\omega\alpha} & |\omega| \le \omega_c \\ 0 & \omega_c \le |\omega| \le \pi \end{cases}$$     $\alpha$ is delay or shift to make the $h_d(n)$ having linear phase.

$$h_d(n) = F^{-1}[H_d(\omega)] = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_d(\omega)e^{j\omega n}d\omega = \frac{1}{2\pi}\int_{-\omega_c}^{\omega_c} e^{j\omega(n-\alpha)}d\omega$$

$$h_d(n) = \frac{\sin\omega_c(n-\alpha)}{\pi(n-\alpha)}$$

Now lets truncate it by a rectangle window:

$$h(n) = \begin{cases} h_d(n) & o \le n \le M-1 \\ o & else \end{cases}$$     and     $\alpha = \dfrac{M-1}{2}$

$h(n) = h_d(n) \cdot w(n)$, where $w(n)$ in general is:

$$w(n) = \begin{cases} some\ symmetric\ function\ with\ respect\ to\ \alpha & o \le n \le M-1 \\ 0 & else \end{cases}$$

For rectangle window   $w^R(n) = \begin{cases} 1 & o \le n \le M-1 \\ o & else \end{cases}$.

In frequency domain:  $H(\omega) = H_d(\omega) * W(\omega) = \dfrac{1}{2\pi}\int_{-\pi}^{\pi} W(\lambda)H_d(W-\lambda)d\lambda$.
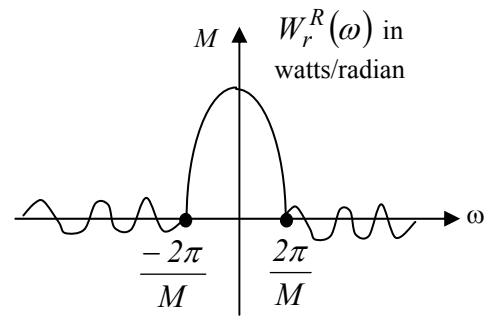
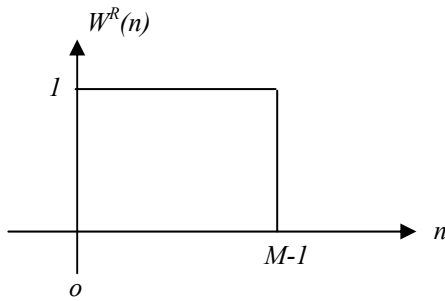Note that since *w(n)* has a finite length equal to M, its frequency response has a main lobe with the width proportional to $1/M$. The main lobe produces a transition band in *H(ω)* whose width is responsible for the transition width.

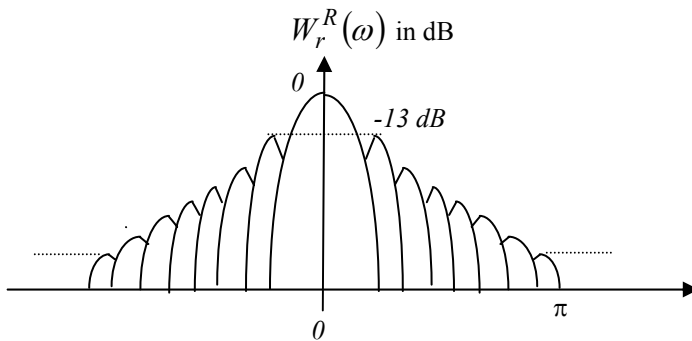The side lobes produce ripples that have similar shapes both in the pass-band and stop-band. Now for a rectangular window, we have:

$$W^R(\omega) = \left[\frac{\sin\dfrac{\omega M}{2}}{\sin\dfrac{\omega}{2}}\right] e^{-j\omega\frac{M-1}{2}} \Rightarrow \underset{Real\ part}{W_r^R(\omega)} = \frac{\sin\dfrac{\omega M}{2}}{\sin\dfrac{\omega}{2}}$$

$$H_r(\omega) \cong \frac{1}{2\pi}\int_{-\pi}^{\omega+\omega_c} W_r^R(\lambda)d\lambda = \frac{1}{2\pi}\frac{\sin\dfrac{\lambda M}{2}}{\sin\dfrac{\lambda}{2}}d\lambda \quad M >> 1$$





Transition band $\simeq \dfrac{4\pi}{M}$ for Rectangular window



The magnitude of side lobe $\approx \dfrac{2M}{3\pi}$ for $M >> 1$ and attenuation $\approx 21$ dB from plot of $H_r(\omega)$

| Window | Transition Width | | Min. Stop-band Attenuation |
|---|---|---|---|
| | Approximate | Exact | |
| Rectangle | $\dfrac{4\pi}{M}$ | $\dfrac{1.8\pi}{M}$ | 21 dB |
| Barlet | $8\pi/M$ | $\dfrac{6.1\pi}{M}$ | 25 dB |
| Hanning | $8\pi/M$ | $\dfrac{6.2\pi}{M}$ | 44 dB |
| Hamning | $8\pi/M$ | $\dfrac{6.6\pi}{M}$ | 53 dB |
| Blackman | $12\pi/M$ | $\dfrac{11\pi}{M}$ | 74 dB |

```matlab
%% Example of designing a digital FIR LPF filter with
%% wp=0.2pi, ws=0.3pi, delta1=0.25 dB and As=50 dB

%% creating an ideal hd(n) with M=6.6pi/(ws-wp)
clear
wp=0.2*pi; ws=0.3*pi;
tr_width=ws-wp;
M=ceil(6.6*pi/tr_width)+1;   %M=67;

wc=(wp+ws)/2;
alpha=(M-1)/2;

n=0:M-1;
hd=sin((n-alpha+eps)*wc)./((n-alpha+eps)*pi); %% hd is the ideal LPF filter
with a shift of alpha to be causal.
W_ham=hamming(M); %% Hamming Window
h=hd.*W_ham';     %% the actual filter

[H,w]=freqz(h,1,1000,'whole');
H=H(1:501);
H_dB=20*log10(abs(H)+eps)/max(abs(H));
w=w(1:501);
delta_w=2*pi/1000;    %% scaling the w axis according to sampling rate
Rp=-(min(H_dB(1:wp/delta_w+1)));   % Actual passband ripple =0.019
As=-round(max(H_dB(ws/delta_w+1:501)));   % Actual stopband ripple =51

%%plotting
subplot(2,2,1); stem(n,hd); title('Ideal Impulse Response of the LPF')
axis([0 M-1 -0.1 0.3]); xlabel('n'); ylabel('hd(n)')

subplot(2,2,2); stem(n, W_ham); title('Hamming window')
axis([0 M-1 0 1.1]); xlabel('n'); ylabel('W(n)')

subplot(2,2,3); stem(n,h); title('Actual Impulse Response of LPF');
axis([0 M-1 -0.1 0.3]); xlabel('n'); ylabel('h(n)')

subplot(2,2,4); plot(w/pi,H_dB); title('Magnitude of H(w) in dB');
axis([0 1 -100 10]); xlabel('frequency in pi unit'); ylabel('dB');
```

## Design of Optimum Equiripple Linear Phase FIR Filter

In previous method, the most important disadvantage is the lack of control on $\omega_p$ and $\omega_c$ and $\omega_s$. This method is formulated as Cheby-Chev approximation problem. It is viewed as an optimum design criterion in the sense that the weighted approximation error between the desired $H_d(\omega)$ and the actual $H(\omega)$ is spread evenly across the pass-band the stop-band minimizing the maximum error. So, our filter in pass-band must satisfy:

$$1 - \delta_1 \le H_r(\omega) \le 1 + \delta_1 \qquad\qquad |\omega| \le \omega_p$$

and in stop-band $\qquad -\delta_2 \le H_r(\omega) \le \delta_2 \qquad\qquad |\omega| > \omega_s$

Now consider the case of a symmetric $h(n) = h(M-1-n)$ with $M$ to be an odd number.

$$\underbrace{h(n) = h(M-1-n)}_{n=0,1,\ldots,M-1} \Rightarrow H(\omega) = H_r(\omega)e^{-j\omega\left(\frac{M-1}{2}\right)}$$

$$\Rightarrow H_r(\omega) = H(\omega)e^{j\omega\left(\frac{M-1}{2}\right)} = \sum_{n=0}^{M-1} h(n)e^{-j\omega n} e^{j\omega\left(\frac{M-1}{2}\right)} = \sum_{n=o}^{M-1} h(n)e^{j\omega\left(\frac{M-1}{2}-n\right)}$$

$$\Rightarrow H_r(\omega) = h\left(\frac{M-1}{2}\right) + 2\sum_{n=o}^{\frac{M-3}{2}} h(n)\cos\left[\omega\left(\frac{M-1}{2}-n\right)\right] \qquad \text{for an odd M.}$$

Let $k = \dfrac{M-1}{2} - n \to n$ from $\left(0, \dfrac{M-3}{2}\right)$ makes $k$ to change in $\left(\dfrac{M-1}{2}, 1\right)$ and lets define

$$a_k = \begin{cases} h\left(\dfrac{M-1}{2}\right) & k = 0 \\ 2h\left(\dfrac{M-1}{2}-k\right) & k = 1,2\ldots,\dfrac{M-1}{2} \end{cases}$$

$$\Rightarrow H_r(\omega) = \sum_{k=0}^{\frac{M-1}{2}} a(k)\cos\omega k$$

Now lets choose a weighting function for the ripples.

Let $W(\omega) = \begin{cases} \dfrac{\delta_2}{\delta_1} & \omega \text{ in } pass-band \\ 1 & \omega \text{ in } stop-band \end{cases}$

Now we can define the weighted approximation error as $E(\omega) = W(\omega)[H_{dr}(\omega) - H_r(\omega)]$. The Cheby-Chev approximation is basically to determine $a(k)$ that minimizes the maximum absolute

value of $E(\omega)$ over the frequency band which approximation is to be performed. In a sense, we seek solution to the problem

$$\text{Min error over } a(k) = \underbrace{\left[Max|E(\omega)|\right]}_{\omega \in s} = \min\left[Max|W(\omega)[H_{dr}(\omega) - H_r(\omega)]|\right]$$

(s consists of pass-band and stop-band)

Since $E(\omega)$ alternates in sign between two successive external frequencies, it is called "alternation theorem".

$$\frac{dE(\omega)}{d\omega} = \frac{d}{d\omega}[W(\omega)[H_{dr}(\omega) - H_r(\omega)]] = -\frac{d}{d\omega}H_r(\omega) = 0 \text{ Let } L = \frac{M-1}{2}$$

$$\Rightarrow H_r(\omega) = \sum_{k=0}^{L} a(k)\cos\omega k = \sum_{k=0}^{L} a(k)\sum_{n=0}^{k} \beta_{nk}(\cos\omega)^n$$

$$= \sum_{k=0}^{L} \alpha'(k)(\cos\omega)^k = 0$$

This means that $H_r(\omega)$ can have at most $L - 1$ local Maxima and Minima on the interval $0 < \omega < \pi$. In addition, $\omega = 0, \pi$ are usually extrema of $H_r(\omega)$ and also of $E(\omega)$. Therefore $H_r(\omega)$ has at most $L - 1 + 2 = L + 1$ extremal frequencies. Furthermore, the band-edge frequencies $\omega_p$ and $\omega_s$ are also extrema of $E(\omega)$, since $|E(\omega)|$ is maximum at $\omega = \omega_p$ and $\omega = \omega_s$.

Therefore, there are at most $L + 3$ extremal frequencies in $E(\omega)$ for the unique and best approximation of the ideal low-pass filter. So by selecting the desired alternation frequencies, we have a set of linear equations:

$$\hat{W}(\omega_n)\left[\hat{H}_{dr}(\omega_n) - H_r(\omega_n)\right] = (-1^n)\delta \qquad n = 0,...., L+1$$

$$\hat{H}_{dr}(\omega_n) = H_r(\omega_n) + \frac{(-1)^n \delta}{\hat{W}(\omega_n)}$$

$$\sum_{k=0}^{L} a_k(\cos\omega)^k = \sum_{k=0}^{L} a_k \cos\omega k$$

$$\begin{bmatrix} 1 & \cos\omega_0 & \cos 2\omega_0 & \cos L\omega_0 & 1/\hat{W}(\omega_0) \\ 1 & \cos\omega_1 & \cos 2\omega_1 & \cos L\omega_1 & -1/\hat{W}(\omega_1) \\ 1 & & & & \\ 1 & \cos\omega_{L+1} & \cos 2\omega_{L+1} & & \end{bmatrix} \begin{bmatrix} \alpha(0) \\ \alpha(1) \\ \alpha(L) \\ \delta \end{bmatrix} = \begin{bmatrix} \hat{H}_{dr}(\omega_0) \\ \hat{H}_{dr}(\omega_1) \\ \\ \hat{H}_{dr}(\omega_{L+1}) \end{bmatrix}$$

$\Gamma$

Remez algorithm solves the above equation recursively. (Matlab function: "remez")

Using Direct Method to solve the equations with MatLab

*Example*

Assume: $L = 4$ and goal is to design a low pass filter with $\omega_p = 1$ and $\omega_s = 1.5$ rad/s.

Try this code:
$\qquad\qquad\qquad \omega_p \qquad \omega_s \qquad\qquad \pi$

$\qquad\qquad\qquad \downarrow \qquad \downarrow \qquad\qquad \downarrow$

*freq = [0,    0.5    1     1.5    2.3    3.14]';*

* D = [1     1     1     0     0     0]';*

*FREQ = freq * [0: 4];*

*Cos_Matrix = cos(FREQ);*

*wt = [1 1 1 1 1 1]';   % using 1 as weights for     $\dfrac{\delta_2}{\delta_1}$*

*wtt = [1     -1    1    -1    1     -1]'./wt;*

*Gamma = [cos-Matrix wtt];*

*a= Gamma\D; % this is equivalent to a = Gamma$^{-1}$*D*

*h1 = [a(5)/2, a(4)/2, a(3)/2, a(2)/2];*

*h = [h1   a(1)   flip lr(h1)];*

*delta = a(6);*

*[H, f] = freqz(h, 1);*

*Plot (f, abs(H));*

Solving same problem with Remez algorithm

*freq $= \left[0, \quad \dfrac{1}{pi}, \quad \dfrac{1.5}{pi}, \quad 1\right];$*

*d = [1    1    0     0];*

*h = remez (8, freq, d);*

*[H, F] = freqz (h, 1);*

*Plot (F, abs (H))*

If you use *wt = [2 1]* and then remez (8, *freq, d, wt*), then you will get twice the tolerance in stop-band than in pass-band as *wt* is the weighting function.

Note that is these $\omega_1$ to $\omega_6$ are all extremal frequencies, the delta will be the maximum error defined in the formula. But if $\omega_1$ to $\omega_6$ are not all extremal frequencies, still there is a solution but $\delta$ will not be the maximum. Remez exchange algorithm first selects $L + 2$ arbitrary

frequencies in $[0 \quad \omega_p]$ and $[\omega_s \quad \pi]$. The band edge frequencies $\omega_p$ and $\omega_s$ must be included in the set. Including $0$ and $\pi$ are unimportant. Then the algorithm runs as direct method in Program in the example and checks if $|e(\omega)| \leq e_m = \delta$ is true for all $\omega$ in the set or not. If yes, the selected frequencies are all extremal and the filter is optimum. If not, Remez algorithm selects four frequencies besides $\omega_p$ and $\omega_s$ that makes the slope of $e(\omega)$ to zero or equivalently has a peak ripple locally and runs Program again. This is also called Parks-McClellan algorithm.

Note that selecting the frequency set as the one in the program of the example, $\delta = e_m = 0.09605$ which is not the largest error. Next you may select a different set of frequencies and retain *0, 1, 1.5*.