

Lecture 9

Filter Design: FIR Filters

Reading: Sections 7.2 - 7.6 in Oppenheim, Schafer & Buck (OSB).

Unlike discrete-time IIR filters which are generally obtained by transforming continuous-time IIR systems, FIR filters are almost always implemented in discrete time: a desired magnitude response is approximated directly in the DT domain. In addition, a linear phase constraint is often imposed during FIR filter design. In this lecture, we will look at two FIR filter design techniques: by windowing and through the Parks-McClellan Algorithm.

FIR Filter Design by Windowing

In designing FIR filter, given the frequency response $H_d(e^{j\omega})$ and impulse response $h_d[n]$ of an ideal system, we would like to approximate the infinitely long $h_d[n]$ with a finite sequence $h[n]$, where $h[n] = 0$ except for $0 \leq n \leq M$. Consider an ideal lowpass filter whose frequency response is finite and rectangular. A possible approximation error criterion can be defined as:

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

To minimize ϵ , use Parseval's theorem:

$$\begin{aligned} \epsilon &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |H_d(e^{j\omega}) - H(e^{j\omega})|^2 d\omega \\ &= \sum_{n=-\infty}^{\infty} |h_d[n] - h[n]|^2 = \sum_{n=0}^M |h_d[n] - h[n]|^2 + \sum_{n=\mathbb{Z} \setminus [0, M]} |h_d[n] - h[n]|^2 \\ \Rightarrow h[n] &= \begin{cases} h_d[n] & 0 \leq n \leq M \\ 0 & \text{otherwise} \end{cases} . \end{aligned}$$

The optimal FIR approximation using the mean square error criterion gives the truncation of the ideal impulse response. We can represent $h[n]$ as the product of the ideal impulse response with a finite-duration rectangular window $w[m]$:

$$\begin{aligned} h[n] &= h_d[n]w[n], \quad w[n] = \begin{cases} 1 & 0 \leq n \leq M \\ 0 & \text{otherwise} \end{cases} \\ \Rightarrow H(e^{j\omega}) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} H_d(e^{j\theta})W(e^{j(\omega-\theta)})d\theta . \end{aligned}$$

OSB Figure 7.19(a) depicts this convolution process implied by truncation of the ideal impulse response. The resulting magnitude approximation as the sinc pulse $W(e^{j(\omega-\theta)})$ slides past the ideal frequency response $H_d(e^{j\omega})$ is shown in Figure 7.19(b). When $W(e^{j(\omega-\theta)})$ moves across the discontinuity of $H_d(e^{j\omega})$, a transition band results and ripples occur on both sides. As shown in OSB Figure 7.23, the main lobe of the window frequency response controls transition bandwidth $\Delta\omega \approx 2\pi(2/(M+1))$. The mainlobe is defined as the region between the first zero crossings on either side of the origin. It is desirable to have $W(e^{j\omega})$ as concentrated in frequency as possible. On the other hand, sidelobes control passband and stopband ripples. The larger the area under the sidelobes, the larger the ripples. Passband and stopband ripples are approximately equal over a wide range of frequencies.

For any given window length, we would like $W(e^{j\omega})$ to be “most like an impulse,” with a narrow main lobe and low sidelobes. For example, OSB Figure 7.20 shows a magnitude response of a rectangular window where $M = 7$. Although increasing the rectangular window length M narrows the lobe widths, areas under each lobe remains constant, thus the ripples occur more rapidly, but with the same amplitude. To reduce the height of the sidelobes, we will need to taper the rectangular window smoothly to zero at the ends. Some commonly used windows are displayed in OSB Figure 7.21. Their properties are quantitatively compared in OSB Table 7.1. Corresponding analytical definitions can be found in Section 7.2.1.

Comparison among these commonly used windows shows that at a given length, the rectangular window has the narrowest mainlobe; it gives the sharpest transition at the discontinuity of $H_d(e^{j\omega})$. The Bartlett window is triangular, with a slope discontinuity at its center, while the Hanning, Hamming and Blackman windows are smoother. By tapering the window smoothly to zero, the sidelobes can be reduced in amplitudes, but the trade-off is larger main lobes. All of these windows are symmetric about $M/2$, hence their frequency responses have generalized linear phase. The Kaiser windows, as defined below, are a family of near optimal windows that allow controlled trade-offs between the sidelobe amplitudes and mainlobe widths. By varying the β parameter, it is possible to approximately duplicate the other types of windows using a Kaiser window.

The Kaiser window family:

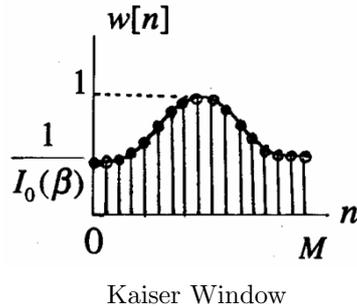
$$w[n] = \begin{cases} \frac{I_0 \left[\beta \sqrt{1 - \left(\frac{n-\alpha}{\alpha} \right)^2} \right]}{I_0(\beta)} & 0 \leq n \leq M \\ 0 & \text{otherwise} \end{cases}$$

where

$$\alpha = \frac{M}{2}, \quad \beta : \text{shape parameter}$$

$I_0(\cdot)$: zeroth-order modified Bessel Function of the first kind

The definition implies that the Kaiser class of windows is parameterized by length $(M+1)$ and shape (β) . The following figure illustrates its general form. As β increases, $I_0(\beta)$ also increases to give a smoother window.



Given a set of filter specifications, the values of M and β needed can be determined numerically. As in OSB Figure 7.23, define:

Transition bandwidth: $\Delta\omega = \omega_s - \omega_p$

Peak approximation error: δ (passband: $1 \pm \delta$ stopband: $\pm \delta$)

Stopband attenuation: $A = -20 \log_{10} \delta$

Then the values of β, M needed to achieve A are

$$\beta = \begin{cases} 0.1102(A - 8.7) & A > 50 \\ 0.5842(A - 21)^{0.4} + 0.07886(A - 21) & 21 \leq A \leq 50 \\ 0 & A < 21 \end{cases}, \quad M \geq \frac{A - 8}{2.285\Delta\omega}.$$

See OSB Section 7.3 for examples of the Kaiser window method.

Optimum FIR Filter Approximation: The Parks-McClellan Algorithm

Although the rectangular windowing method provides the best mean-square approximation to a desired frequency response, just because it is optimal does not mean it is good. FIR filters designed with windows exhibit oscillatory behavior around the discontinuity of the ideal frequency response and does not allow separate control of the passband and stopband ripples. An alternative FIR design technique is the Parks-McClellan algorithm which is based on polynomial approximations.

Filter Design as Polynomial Approximation

Consider the DTFT of a causal FIR system of length $M + 1$:

$$H(e^{j\omega}) = \sum_{n=0}^M h[n]e^{-j\omega n} = \sum_{n=0}^M h[n]x^n \Big|_{x=e^{-j\omega}}.$$

Let's further constrain our system to a type I generalized linear-phase filter defined by a symmetric impulse response with M even:

$$h[n] = h[M - n] \quad 0 \leq n \leq M, M \text{ even.}$$

Its corresponding frequency response is

$$H(e^{j\omega}) = \sum_{n=0}^M h[n]e^{-j\omega n} = A_e(e^{j\omega})e^{j\omega M/2},$$

where

$$\begin{aligned} A_e(e^{j\omega}) &= \sum_{n=-M/2}^{M/2} h_e[n]e^{j\omega n} & h_e[n] &= h[n + M/2] = h_e[-n] \\ &= h_e[0] + \sum_{n=1}^{M/2} h_e[n](e^{-j\omega n} + e^{j\omega n}) \\ &= h_e[0] + 2 \sum_{n=1}^{M/2} h_e[n] \cos \omega n \end{aligned}$$

Here $h[n]$ is just $h_e[n]$ with an integer delay of $M/2$. Rewrite the $\cos \omega n$ terms using Chebyshev polynomials:

$$\begin{aligned} T_n(x) &= \cos(n \cos^{-1} x) \quad \Rightarrow \quad \cos \omega n = T_n(\cos \omega) \\ A_e(e^{j\omega}) &= \sum_{k=0}^{L=M/2} a_k (\cos \omega)^k = \sum_{k=0}^{M/2} a_k x^k \Big|_{x=\cos \omega} = P(x) \Big|_{x=\cos \omega}. \end{aligned}$$

Similar procedures can also be carried out for type II, III, IV systems (See Section 7.4 of OSB). In particular, when M is odd:

$$H(e^{j\omega}) = e^{-j\omega M/2} \cos\left(\frac{\omega}{2}\right) P(\cos \omega)$$

The Alternation Theorem

With $A_e(e^{j\omega})$ written as a trigonometric polynomial, Parks and McClellan applied the Alternation Theorem from approximation theory to the filter design problem. The theorem is stated in Section 7.4 of OSB (p.489). Although the alternation theorem does not lead to any optimal filter design directly, what it allows us to do is to determine whether a filter response written as $P(x)$ is uniquely optimal under the *minimax* error criterion, i.e., if it minimizes the maximum weighted approximation error. The Parks-McClellan Algorithm uses the alternation theorem to iteratively determine an optimal equiripple approximation.

OSB Figures 7.35 and 7.36 show a typical example of an optimal lowpass filter frequency response for $L = 7$ and its equivalent polynomial approximation function $P(x)$. The alternating points are $\omega_1, \omega_2, \dots, \omega_8, \pi$, where $E_p(x) = E_p(\cos \omega)$ alternates between its maximum and minimum for $L + 2$ times. Such an approximation is said to be equiripple. The weighting function in this case satisfies:

$$W_p(\cos \omega) = \begin{cases} \frac{1}{K} & \cos \omega_p \leq \cos \omega \leq 1 \\ 1 & -1 \leq \cos \omega \leq \omega_s \end{cases},$$

where $K = \delta_1/\delta_2$. $P(x)$ turns into a hyperbolic cosine function outside of $F_p = [-1, 1]$. Since we are only interested in $\omega \in [0, \pi]$, we do not care about the behavior of $P(x)$ outside of F_p . Comparison between the Figures 7.35 and 7.36 shows that for lowpass filters (or any piecewise constant filters), the alternations in $P(x)$ can be counted directly from the frequency response $A_e(e^{j\omega})$.

In OSB Figure 7.37 are other examples of optimum lowpass filter approximations for $L = 7$. All four satisfy the alternation theorem. Nonetheless, the uniqueness of optimal filters implied by the alternation theorem is not violated because ω_p and/or ω_s are different for each. By comparison, OSB Figure 7.38 and 7.39 show approximations that are not optimal according to the alternation theorem. There are 8 alternations in each case.

More generally, it can be shown that for type I lowpass filters, the following properties are implied by the alternation theorem:

- The maximum possible number of alternations of the error is $L + 3$.
- Alternations will always occur at ω_p and ω_s .
- The passband and stopband have equal ripples, and the transition band is monotonic. All points with zero slope inside the passband and all points with zero slope inside the stopband (for $0 < \omega < \omega_p$ and $\omega_s < \omega < \pi$) will correspond to alternations, i.e., the filter is equiripple, except possibly at $\omega = 0, \pi$. See Section 7.4 of OSB for proofs.

So far we have focused on only optimal lowpass FIR filters, which have two approximation bands. For multi-band filters such as the bandpass filter, the alternation theorem still holds, but it does imply a different set of properties on the approximating polynomial. For example, the optimal approximation can have more than $L + 3$ alternations and have ripples (local maxima/minima) in the transition band. OSB Figure 7.47 is such an example. This system is optimal since the alternation theorem is satisfied, but the ripple in its transition band makes it undesirable in practical applications. Sections 7.5 and 7.6 of OSB give more examples of FIR equiripple approximations and further discuss practical implementation issues that should be considered during the filter design process.

Parks-McClellan Algorithm

The alternation theorem provides us with a tool for efficiently finding the optimum filter in the minimax sense. For a type I lowpass filter, the following set of equations is satisfied by the

optimum filter according to the alternation theorem:

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

where δ is the optimum error. Parks and McClellan found that given fixed ω_p, ω_s , an iterative algorithm finds the optimal approximating $A_e(e^{j\omega_i})$:

Algorithm

1. Initial guess of $(L + 2)$ extremal frequencies $\omega_0, \omega_1, \dots, \omega_{L+1}$.
2. Solve for polynomial coefficients and $(L + 2)$ equations in $(L + 2)$ unknowns.
3. Evaluate $A_e(e^{j\omega_i})$ or $E(\omega)$ on a dense set of frequencies and choose a new set of extremal frequencies.
4. Check whether extremal frequencies changed. If yes go to step 2. If no, algorithm has converged.

This algorithm is described in more detail and illustrated by a flowchart in OSB Section 7.4.3. Keep in mind that the optimality of filters designed by the PM algorithm is only in the minimax error sense, and the equiripple characteristic may be less desirable when compared with filters designed by other techniques.