Digital Signal Processing
Lecture 9 - Design of Digital Filters - FIR

Electrical Engineering and Computer Science
University of Tennessee, Knoxville

November 3, 2015

Overview

Roadmap

- Introduction
- Discrete-time signals and systems - LTI systems
  - Unit sample response $h[n]$: uniquely characterizes an LTI system
  - Linear constant-coefficient difference equation
  - Frequency response: $H(e^{j\omega})$
  - Complex exponentials being eigenvalues of an LTI system: $y[n] = H(e^{j\omega})x[n]$
  - Fourier transform
- z transform
  - The z-transform, $X(z) = \sum_{n=-\infty}^{\infty} x[n]z^{-n}$
  - Region of convergence - the z-plane
  - System function, $H(z)$
  - Properties of the z-transform
  - The significance of zeros
  - The inverse z-transform, $x[n] = \frac{1}{2\pi j} \oint_C X(z)z^{n-1}dz$: inspection, power series, partial fraction expansion
- Relationships between the n, ω, and z domains

Review - Design structures

- Different representations of causal LTI systems
  - LCDE with initial rest condition
  - $H(z)$ with $|z| > R_+$ and starts at $n = 0$
- Block diagram vs. Signal flow graph and how to determine system function (or unit sample response) from the graphs
- Design structures
  - Direct form I (zeros first)
  - Direct form II (poles first) - Canonic structure
  - Transposed form (zeros first)
- IIR: cascade form, parallel form, feedback in IIR (computable vs. noncomputable)
- FIR: direct form, cascade form, parallel form, linear phase
- Metric: computational resource and precision
- Sources of errors: coefficient quantization error, input quantization error, product quantization error, limit cycles
  - Pole sensitivity of 2nd-order structures: coupled form
  - Coefficient quantization examples: direct form vs. cascade form

Review - Filter Design - IIR

- Design guidelines
  - mapping of the good behavior of $j\Omega$ axis to the unit circle $|z| = 1$
  - mapping of the stable poles
- CT –> DT
  - Impulse invariance
  - Bilinear transformation

FIR filters with generalized linear phase

- General FIR filters

$$y[n] = \sum_{k=0}^{M-1} h[k]x[n-k], H(z) = \sum_{k=0}^{M-1} h[k]z^{-k}$$

  - All poles are at $z = 0$
  - ROC is the entire z-plane except for $z = 0$
- FIR filters with generalized linear phase

$$H(e^{j\omega}) = A(e^{j\omega})e^{j(\beta-\alpha\omega)}$$
$$arg[H(e^{j\omega})] = \beta - \omega\alpha, 0 < \omega < \pi$$
$$grd[H(e^{j\omega})] = -\frac{d}{d\omega}\{arg[H(e^{j\omega})]\} = \alpha$$

- Necessary condition on $h[n]$ to have constant group delay

$$\sum_{n=-\infty}^{\infty} h[n] \sin[\omega(n-\alpha)+\beta] = 0 \text{ for all } \omega$$

Four types of causal FIR systems with generalized linear phase

- Sufficient condition: $h[n]$ satisfies the symmetry and antisymmetry conditions

$$h[n] = \pm h[M-n], n = 0, 1, \cdots, M$$

- Type 1: $h[n] = h[M-n]$, M even. $\beta = 0$ or π. Delay is even as $\alpha = M/2$.

$$H(e^{j\omega}) = e^{-j\omega M/2}(h[M/2] + 2\sum_{k=1}^{M/2} h[M/2-k]\cos\omega k)$$

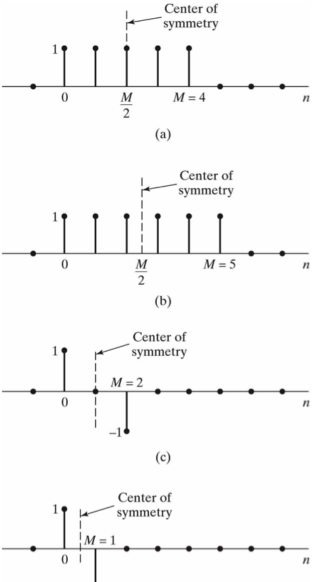
- Type 2: $h[n] = h[M-n]$, M odd. $\beta = 0$ or π. Delay is integer plus a half.

$$H(e^{j\omega}) = e^{-j\omega M/2}(2\sum_{k=1}^{(M+1)/2} h[(M+1)/2-k]\cos[\omega(k-1/2)])$$

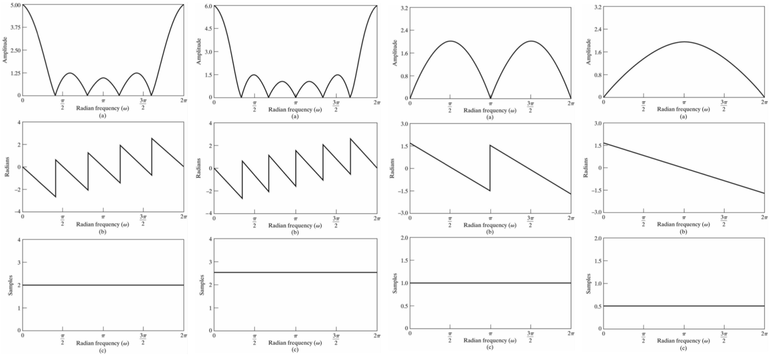
Four types of causal FIR systems (cont')

- Type 3: $h[n] = -h[M-n]$, M even. $\beta = \pi/2$ or $3\pi/2$. Delay is integer.
- Type 4: $h[n] = -h[M-n]$, M odd. $\beta = \pi/2$ or $3\pi/2$. Delay is integer plus a half.

Examples $h[n]$


Center of symmetry
1
0
M/2
M = 4
n
(a)
Center of symmetry
1
0
M/2
M = 5
n
(b)
Center of symmetry
1
M = 2
0
–1
n
(c)
Center of symmetry
1
M = 1
0
n

Examples $H(e^{j\omega})$, α, and group delay



Zeros of FIR filters with generalized linear phase

- The roots of $H(z)$ must occur in reciprocal pairs, and
- If $h[n]$ is real, the complex-valued roots must occur in complex-conjugate pairs

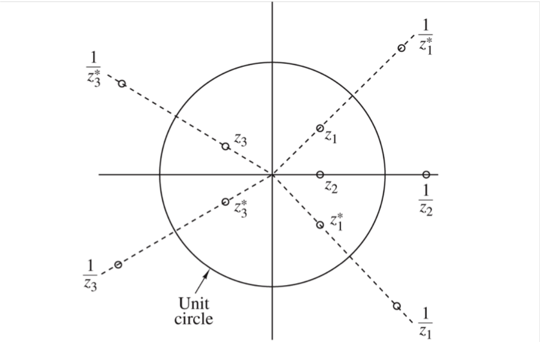
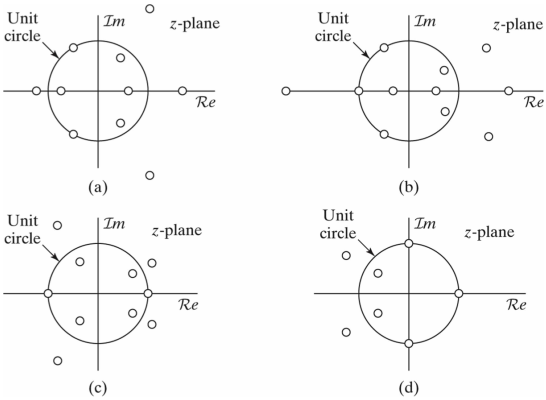


Figure 10.2.1 Symmetry of zero locations for a linear-phase FIR filter.

Zeros of the four types of FIR systems

- Type I and II: $z = -1$
- Type III and IV: $z = \pm 1$



Design techniques

- Windows
- Equiripple design

FIR filter design by windowing

- Desired frequency response, $H_d(e^{j\omega})$ vs. the impulse response, $h_d[n]$

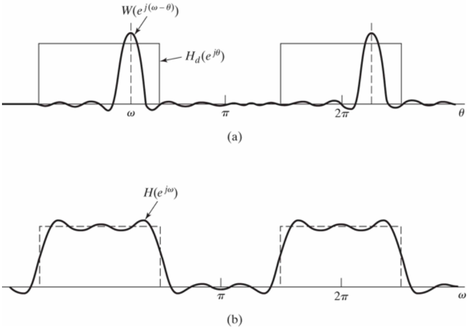
$$H_d(e^{j\omega}) = \sum_{n=0}^{\infty} h_d[n]e^{-j\omega n},\ h_d[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_d(e^{j\omega})e^{j\omega n}d\omega$$

- The need for applying a window function

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

- $H(e^{j\omega})$ is the periodic convolution of the desired ideal frequency response with the Fourier transform of the window, a “smeared” version of $H_d(e^{j\omega})$
- What kind of window function, $w[n]$, will make $H(e^{j\omega}) = H_d(e^{j\omega})$?

FIR filter design by windowing (cont')



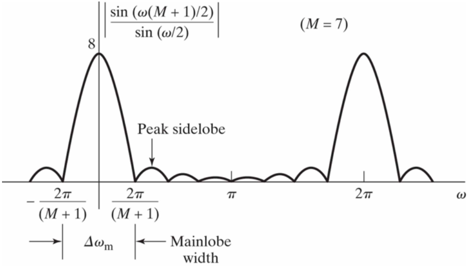
(a)

(b)

$$W(e^{j\omega}) \quad = \quad \sum_{k=0}^{M-1} e^{-j\omega n} = \frac{1-e^{-j\omega M}}{1-e^{-j\omega}} = e^{-j\omega(M-1)/2}\frac{\sin(\omega M/2)}{\sin(\omega/2)}$$

FIR filter design by windowing (cont’)

- Width of the main lobe, $\Delta\omega_m = 4\pi/(M+1)$



Different windows

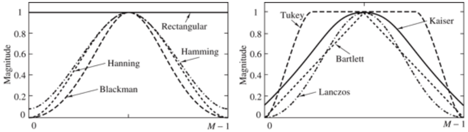
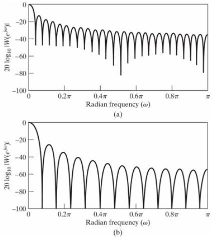
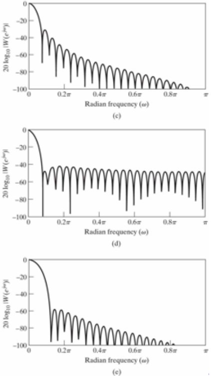


Figure 10.2.3 Shapes of several window functions.





Different windows (cont')

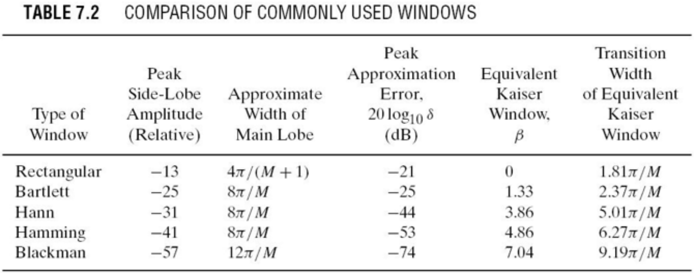
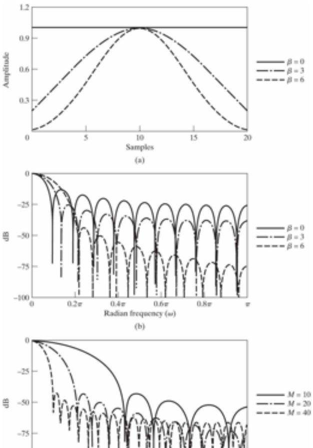


TABLE 7.2 COMPARISON OF COMMONLY USED WINDOWS

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	$2.37\pi/M$
Hann	-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	$9.19\pi/M$

The Kaiser window filter design method

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



The Kaiser window filter design method (cont')

- Getting a trade-off between the main-lobe width and side-lobe area
- If the window is tapered more, the side lobes of the FT become smaller, but the main lobe becomes wider If holding β constant while increasing M, the width of the main lobe is smaller but the peak amplitude of side load does not change.
- Through extensive numerical experimentation, Kaiser obtained a pair of formula to meet a given frequency-selective filter specification

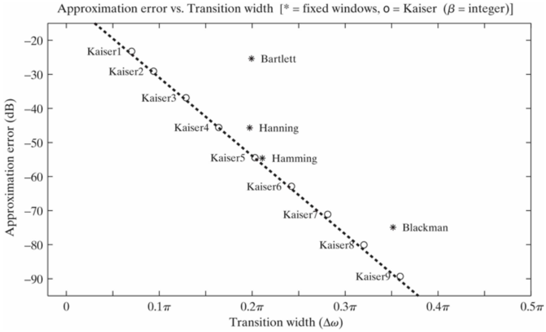
The Kaiser window filter design method (cont')

$$A = -20\log_{10}\delta, \delta: \text{ the peak approximation error}$$

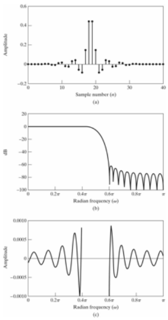
$$\beta = \begin{cases} 0.1102(A-8.7), & A > 50 \\ 0.5842(A-21)^{0.4} + 0.07886(A-21), & 21 \le A \le 50 \\ 0.0, & A < 21 \end{cases}$$

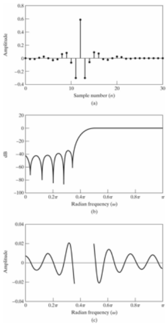
$$M = \frac{A-8}{2.285\Delta\omega}, \Delta\omega = \omega_s - \omega_p$$

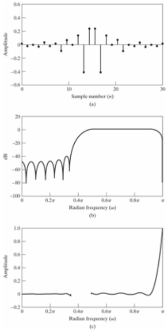
Relationship of the Kaiser window to other windowes



Examples


Amplitude
Sample number (n)
(a)
dB
Radian frequency (ω)
(b)
Amplitude
Radian frequency (ω)
(c)


Amplitude
Sample number (n)
(a)
dB
Radian frequency (ω)
(b)
Amplitude
Radian frequency (ω)
(c)


Amplitude
Sample number (n)
(a)
dB
Radian frequency (ω)
(b)
Amplitude
Radian frequency (ω)
(c)

Limitations

- Potential problems in linear-phase FIR filter design
  - The error is greatest on either side of a discontinuity of the ideal frequency response and becomes smaller for frequencies away from the discontinuity
  - Approximately equal errors in the passband and stopband
  - No control over the exact location of ω_p and ω_s
- Intuitive solutions
  - Design the best for a given value of M
  - Make the approximation error spread out uniformly in frequency
  - Separately adjust passband and stopband ripples
  - The minmax strategy (or frequency-weighted error criterion): minimize the maximum error

Parks-McClellan algorithm

- Reformulating the filter design problem as a problem in polynomial approximation

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

$$
\begin{aligned}
A_e(e^{j\omega}) &= \textstyle\sum_{k=0}^{L} a_k(\cos\omega)^k \\
&= P(x)|_{x=\cos\omega}
\end{aligned}
$$

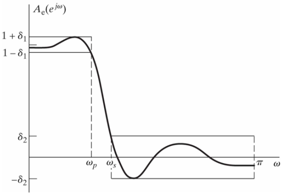
- Parks and McClellan showed that with M, ω_p, ω_s fixed, the filter design problem becomes a problem in Chebyshev approximation over disjoint sets

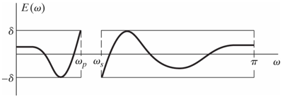
The Chebyshev approximation problem and the minmax (or Chebyshev) criterion

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

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

$$W(\omega) = \begin{cases} 1/K, & 0 \le \omega \le \omega_p \\ 1, & \omega_s \le \omega \le \pi \end{cases}, K = \delta_1/\delta_2$$





$$\min_{\{h_e[n]:0\le n\le L\}} (\max |E(\omega)|)$$

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

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

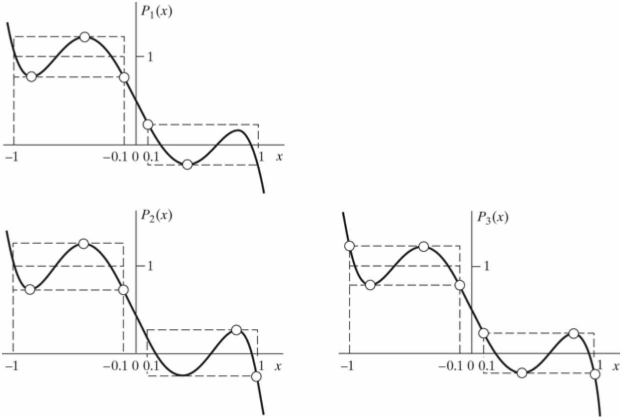
is the weighted error. The maximum error is defined as

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

Alternation theorem (cont')

- A necessary and sufficient condition that $P(x)$ be the unique rth-order polynomial that minimizes $\|E\|$ is that $E_P(x)$ exhibits 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 $i = 1, 2, \cdots, (r+1)$

Example

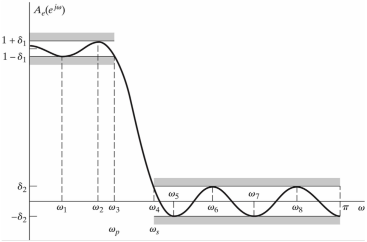


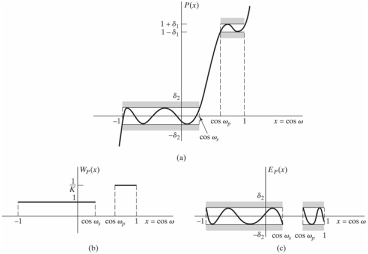
Equiripple approximations - Optimal type I lowpass filters

$$x = \cos\omega, L = M/2, P(\cos\omega) = \sum_{k=0}^{L} a_k(\cos\omega)^k$$
$$D_P(\cos\omega) = \begin{cases} 1, & \cos\omega_p \le \cos\omega \le 1, \\ 0, & -1 \le \cos\omega \le \cos\omega_s \end{cases}$$
$$W_P(\cos\omega) = \begin{cases} 1/K, & \cos\omega_p \le \cos\omega \le 1 \\ 1, & -1 \le \cos\omega \le \cos\omega_s \end{cases}$$
$$E_P(\cos\omega) = W_P(\cos\omega)[D_P(\cos\omega) - P(\cos\omega)]$$
$$F_P : 0 \le \omega \le \omega_p \text{ and } \omega_s \le \omega \le \pi$$

- The alternation theorem states that a set of coefficients a_k will correspond to the filter representing the unique best approximation to the ideal lowpass filter with the ratio δ_1/δ_2 fixed at K and with passband and stopband edges ω_p and ω_s if and only if $E_P(\cos\omega)$ exhibits at least $(L+2)$ alternations on F_P. Such approximations are called equiripple approximations.

Equiripple approximations - example

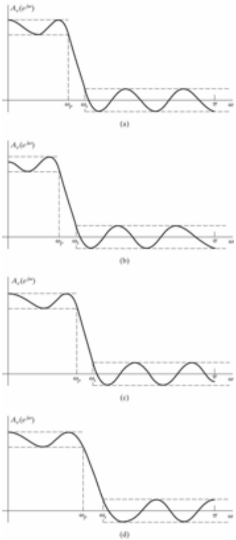




Properties of type I lowpass filters

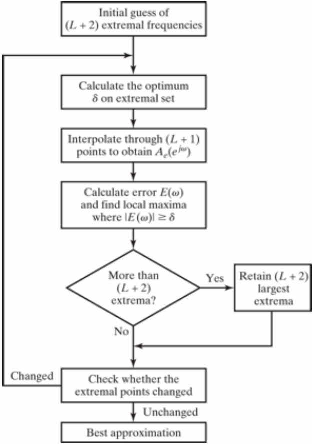
- The maximum possible number of alternations of the error is $(L+3)$
- Alternations will always occur at ω_p and ω_s
- All points with zero slope inside the passband and all points with zero slope inside the stopband will correspond to alternations, except possibly at $\omega = 0$ and $\omega = \pi$ — equiripple

Examples - Possible optimum lowpass filter approximations for $L = 7$



The Parks-McClellan algorithm - polynomial interpolation

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



Example, $\omega_p = 0.4\pi$, $\omega_s = 0.6\pi$, $\delta_1 = 0.01$, $\delta_2 = 0.001$

$$M = \frac{-10\log_{10}(\delta_1\delta_2) - 13}{2.324\Delta\omega}$$