FIR Filter Design

ECE 320 Lecture Notes
Kathleen E. Wage

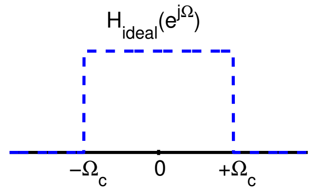
November 18, 2008

Frequency-selective filter design

Focus on designing linear time-invariant (LTI) frequency-selective filters, *i.e.*, filters that

- pass one band of frequencies
- reject other frequencies

Consider the ideal lowpass filter:



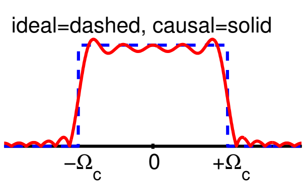
- $h_{\mathrm{ideal}}[n]$ is infinite length spans $n = -\infty$ to $n = +\infty$
- Ideal LPF is not causal since $h_{\mathrm{ideal}}[n] \neq 0$ for $n < 0$

We cannot implement a filter whose impulse response spans all time

Focus on causal filters $\Rightarrow$

Causal filters

Causality imposes some constraints on the filter response:

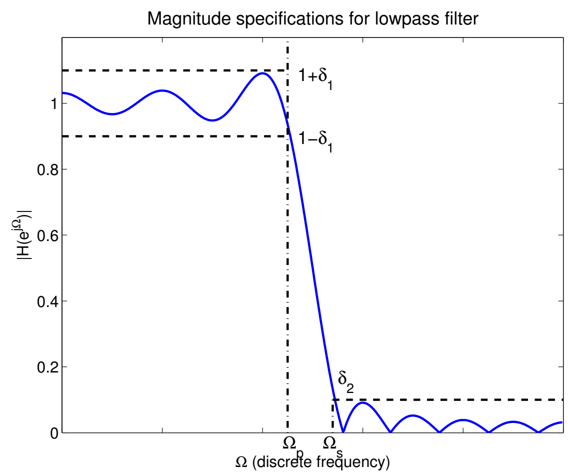


- Stopband ripple: $|H(e^{j\Omega})|$ can't be zero everywhere in stopband
- Passband ripple $|H(e^{j\Omega})|$ can't be one everywhere in passband
- Transition band: filter can't have infinitely sharp transition

Fortunately, in real applications we can tolerate some ripple in the frequency response

Start with a set of design specifications $\Rightarrow$

Design specifications



- Ω_p = Passband edge
- Ω_s = Stopband edge
- δ_1 = Passband ripple
- δ_2 = Stopband ripple

Note: only half of frequency response is shown. It is assumed that $h[n]$ is real, so $H(e^{j\Omega})$ is conjugate symmetric (magnitude is even function of Ω).

Typically specifications are given in terms of CT frequency $\Rightarrow$

CT to DT conversion of design specifications

Specifications are often given in terms of the continuous time (analog) frequency.

The Oppenheim/Willsky book uses ω for continuous time (CT) frequency and Ω for discrete time (DT) frequency.

Can convert between CT and DT frequencies as follows (see discussion of sampling in Ch. 7 of Oppenheim/Willsky)

$$\Omega = \omega T = \frac{\omega}{f_s},$$

where T is the sample period (in seconds) and $f_s = \frac{1}{T}$ is the sampling frequency (in Hz).

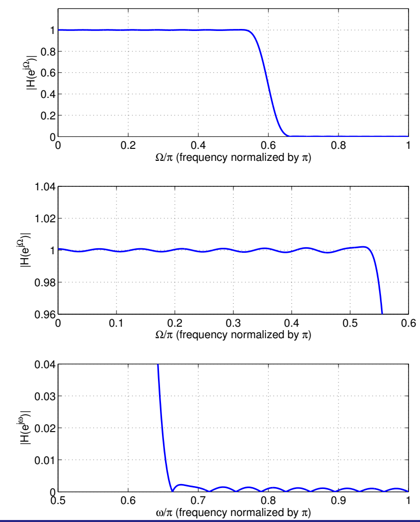
In-class problem $\Rightarrow$

In-class problem

Sample rate is 11,025 Hz
LPF design must meet the following specifications:

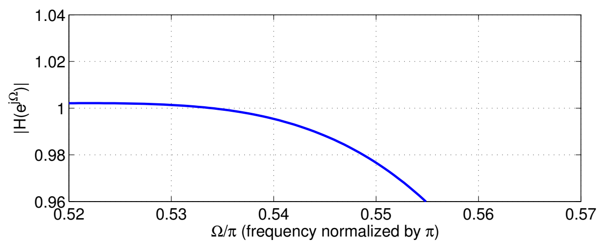
- LPF must pass all frequencies up to 3,000 Hz. The allowable amplitude distortion (ripple) in the passband is $\pm 2\%$., *i.e.*, $\delta_1 = 0.02$.
- Above 3,600 Hz, filter must have attenuation of at least 50 dB, *i.e.*, $20 \log_{10}(\delta_2) = -50$.

Does this filter meet specs ? $\Rightarrow$

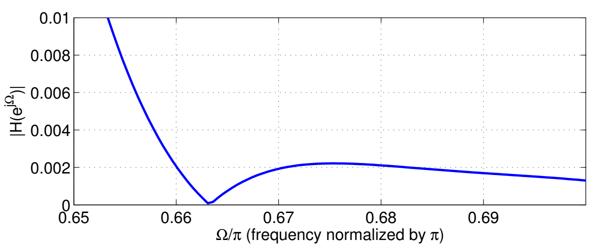


Zoomed plots for in-class problem

Passband $\longrightarrow$



Stopband $\longrightarrow$



In-class problem solution

We compute the design parameters from the given information.
The pass and stopband frequencies can be computed as follows:

$$\Omega_p = \omega_p T = \frac{2\pi(3000)}{11025} = 0.544 \qquad \Omega_s = \omega_s T = \frac{2\pi(3600)}{11025} = 0.653$$

The passband ripple is given: $\delta_1 = 0.02$
The stopband ripple can be computed from the given info as follows:

$$20 \log_{10}(\delta_2) = -50$$

$$\delta_2 = 10^{-50/20} = 0.0032$$

Looking at the plots, we see that the passband constraints are met, but the stopband constraints are not. Specifically the response is too high at the stopband edge.

Causal filters specified by difference equations

We want a causal filter that satisfies a linear constant-coefficient difference equation:

$$\sum_{k=0}^{N} a_k y[n-k] = \sum_{m=0}^{M} b_m x[n-m]$$

A filter of this type can be implemented recursively, *i.e.*, by rearranging the difference equation:

$$y[n] = \frac{1}{a_0}\left(-\sum_{k=1}^{N} a_k y[n-k] + \sum_{m=0}^{M} b_m x[n-m]\right)$$

For system to be LTI the initial conditions are specified as *initial rest*

What is the frequency response of this system?

Frequency response of LCCDE systems

The frequency response of this system can be computed by transforming the difference equation:

$$Y(e^{j\Omega})\sum_{k=0}^{N} a_k e^{-jk\Omega} = X(e^{j\Omega})\sum_{m=0}^{M} b_m e^{-jm\Omega}$$

$$H(e^{j\Omega}) = \frac{Y(e^{j\Omega})}{X(e^{j\Omega})} = \frac{\sum_{m=0}^{M} b_m e^{-jm\Omega}}{\sum_{n=0}^{N} a_k e^{-jk\Omega}}$$

Design problem: pick a_k and b_m coefficients to get as close to desired $|H(e^{j\Omega})|$ as possible

Important design choice: IIR vs. FIR $\Rightarrow$

Filter design choices: IIR vs. FIR

Infinite Impulse Response (IIR)

$$y[n] = \frac{1}{a_0}\left(-\sum_{k=1}^{N} a_k y[n-k] + \sum_{m=0}^{M} b_m x[n-m]\right)$$

- Must implement recursively! (a_k and b_m coefficients are non-zero)
- Phase is difficult to control
- Fewer multiplies and adds for same $|H(e^{j\omega})|$
- Stability can be a problem
- Can build on analog filter design techniques

Finite Impulse Response (FIR)

$$y[n] = \sum_{m=0}^{M} b_m x[n-m]$$

- Recursion not required! (only b_m coefficients are non-zero)
- Linear phase easy to obtain
- More multiplies and adds for same $|H(e^{j\omega})|$
- FIR filters always stable
- No analog history to build upon

Focus on FIR design via window method in ECE 320 $\Rightarrow$

Window design of FIR filters

Basic idea: obtain an FIR lowpass filter by truncating the ideal lowpass filter impulse response

Truncation is viewed as multiplication of the ideal $h[n]$ with a finite-length window $w[n]$, *i.e.*,

$$h[n] = w[n]h_{\mathrm{ideal}}[n]$$

In the simplest case a rectangular window is used

$$w_{\mathrm{rect}}[n] = \begin{cases} 1 & -M \le n \le +M \\ 0 & \text{otherwise.} \end{cases}$$

The window is $L = 2M + 1$ points long

Use the window method to compute a 5-point filter $\Rightarrow$

In-class problem

(a) Consider an ideal LPF with cutoff frequency $\Omega_c = \frac{\pi}{2}$. Sketch $H_{\text{ideal}}(e^{j\Omega})$ and determine an expression for $h_{\text{ideal}}[n]$.

(b) Design a 5-point FIR filter by windowing $h_{\text{ideal}}[n]$ with $w_{\text{rect}}[n]$ where $M = 2$. Determine and sketch $h[n] = w_{\text{rect}}[n] h_{\text{ideal}}[n]$.

Solution to in-class problem

(a) Consider an ideal LPF with cutoff frequency $\Omega_c = \frac{\pi}{2}$. Sketch $H_{\text{ideal}}(e^{j\Omega})$ and determine an expression for $h_{\text{ideal}}[n]$.

See sketch of $H(e^{j\Omega})$ on slide 2:

$$h_{\text{ideal}}[n] = \frac{1}{2\pi}\int_{-\pi}^{+\pi} H_{\text{ideal}}(e^{j\Omega}) e^{j\Omega n} d\Omega = \frac{1}{2\pi}\int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} 1 e^{+j\Omega n} d\Omega = \frac{\sin\left(\frac{\pi}{2}n\right)}{\pi n}$$

(b) Design a 5-point FIR filter by windowing $h_{\text{ideal}}[n]$ with $w_{\text{rect}}[n]$ where $M = 2$. Determine and sketch $h[n] = w_{\text{rect}}[n] h_{\text{ideal}}[n]$.

$$h[0] = \frac{\frac{\pi}{2}\cos(0)}{\pi} = \frac{1}{2} \qquad h[\pm 1] = \frac{\sin\left(\pm\frac{\pi}{2}\right)}{\pm\pi} = \frac{1}{\pi} \qquad h[\pm 2] = \frac{\sin(\pm\pi)}{\pm 2\pi} = 0$$

Note: use L'Hopital's to get $h[0]$. See class notes for sketch.

How does $H(e^{j\Omega})$ compare to $H_{\text{ideal}}(e^{j\Omega})$?

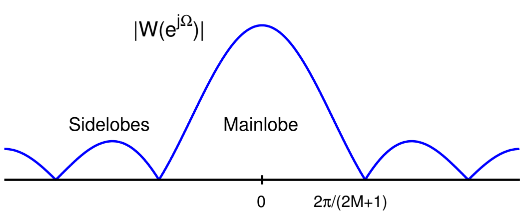
Frequency response of window-designed filter

Multiplication in time is convolution in frequency:

$$h[n] = w[n] h_{\text{ideal}}[n] \qquad H(e^{j\Omega}) = \frac{1}{2\pi} W(e^{j\Omega}) * H_{\text{ideal}}(e^{j\Omega})$$

Consider frequency response of rectangular window $w_{\text{rect}}[n]$:

$$W_{\text{rect}}(e^{j\Omega}) = \sum_{n=-M}^{+M} 1 e^{-j\Omega n} = \frac{\sin\left(\Omega\left(\frac{2M+1}{2}\right)\right)}{\sin\left(\frac{\Omega}{2}\right)}$$



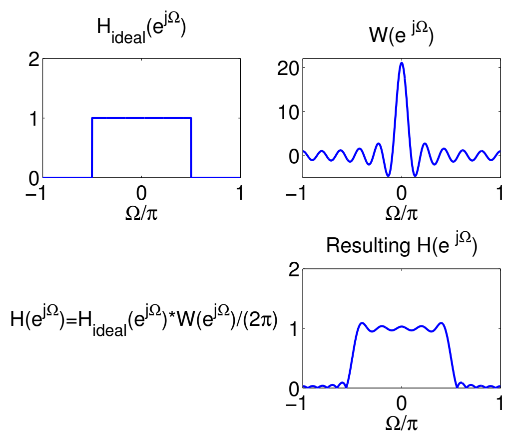
As $M \to \infty$:

- mainlobe shrinks
- sidelobes stay same height

Consider what happens when we convolve ⇒

Frequency response of window-designed filter

Convolution in frequency results in a smearing of the ideal filter frequency response



- Response oscillates as window moves past discontinuity
- Width of transition band relates to the width of the mainlobe of the window
- Recall that mainlobe width is determined by length of window

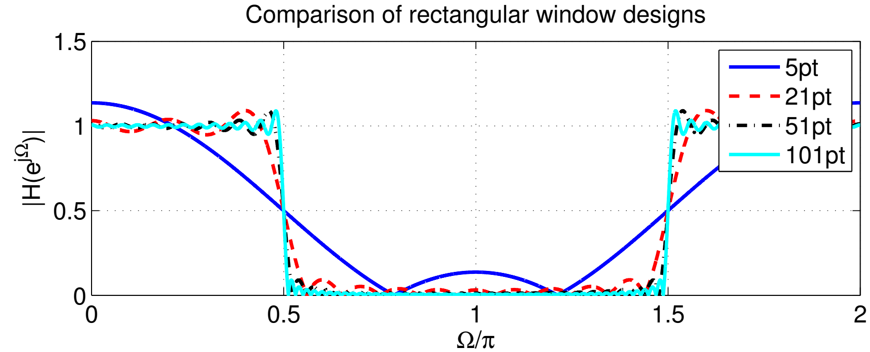
Frequency response of window-designed filter

Recall that mainlobe width is determined by length of window. There are 2 extremes:

- $w[n] = 1 \forall n \longrightarrow W(e^{j\Omega}) = 2\pi \sum_k \delta(w + 2\pi k)$
  No truncation: $h[n] = h_{\text{ideal}}[n]$
  In this case there is no smearing. We obtain the ideal filter.
- $w[n] = \delta[n] \longrightarrow W(e^{j\Omega}) = 1$
  $h[n] = 1$ sample of $h_{\text{ideal}}[n]$
  In this case the frequency response totally smeared, and the filter is ineffective.

Compare rectangular window designs of different lengths ⇒

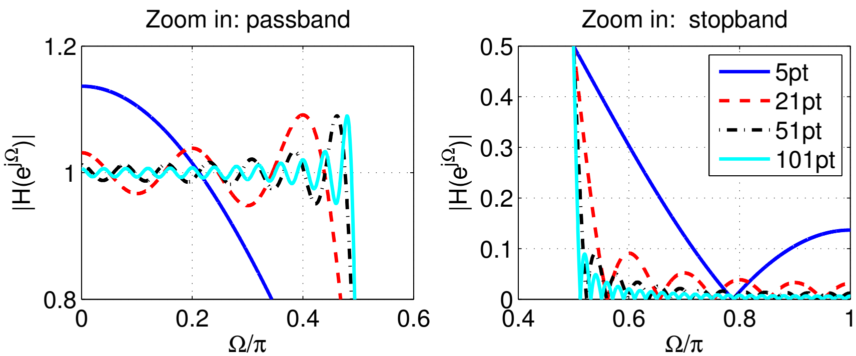
Comparison of rectangular window designs



- As expected, frequency response improves as length of window increases

Zoom in on pass and stop bands ⇒

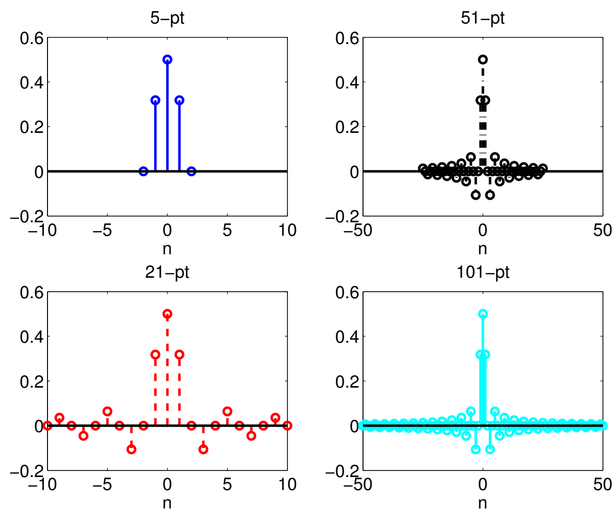
Rectangular window designs: pass/stop bands



- Transition band gets shorter as window gets longer
- Ripples stay approximately the same as window gets longer

Summary of rectangular window design ⇒

Rectangular window designs: impulse responses



Recall that:
$h[n] = w_{\text{rect}}[n] h_{\text{ideal}}[n]$

Summary of rect. window design ⇒

Rectangular window design

- Width of transition band related to the width of the mainlobe of the window
- Ripples related to the *area* under the sidelobes of the window

  Why area? because we're doing convolution!

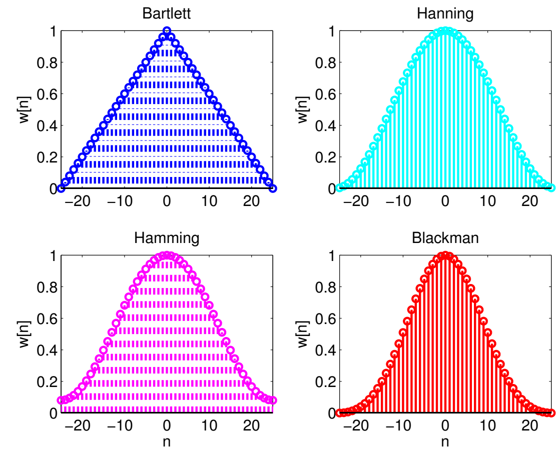
  As length increases the area under the sidelobes doesn't change (oscillations occur more frequently, but area basically unchanged)

Thus changing the window length can't change the ripple in the filter (related to Gibbs phenomenon)

This motivates us to consider other (smoother) window functions ⇒

Common windows

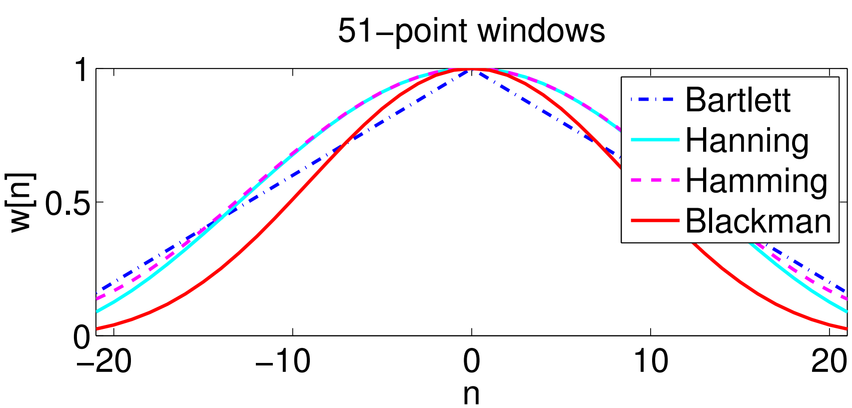
How to get famous in digital signal processing? Design a window!



Show on one plot for better comparison ⇒

Common windows

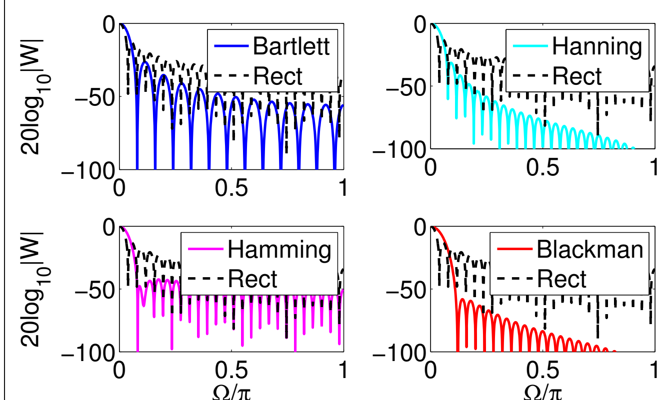
$w[n]$ shown using "plot" instead of "stem", but windows are still discrete!



Window frequency responses ⇒

Window frequency responses (51-point windows)

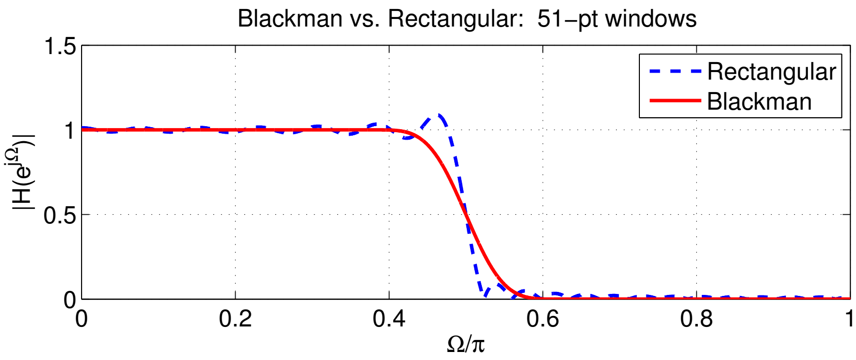
Note: log-scale plots, *i.e.* $20\log_{10}|W(e^{j\Omega})|$



- Other windows have wider mainlobe than rectangular
- Other windows have lower sidelobes than rectangular

Compare 51-point rectangular and Blackman designs ⇒

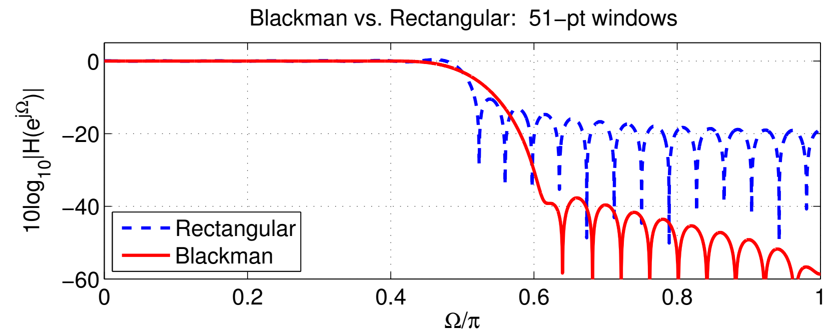
Rectangular vs. Blackman comparison



- Blackman has less ripple than rectangular

Log scale shows more details ⇒

Rectangular vs. Blackman comparison: log scale

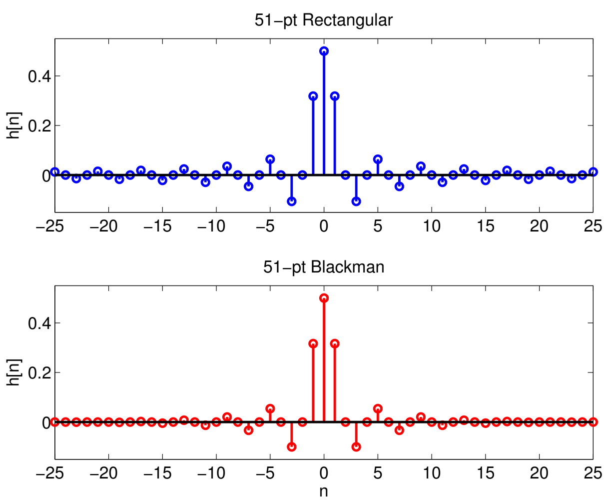


Compared to rectangular:

- Blackman has less ripple ⟶ much better stopband rejection
- Blackman has wider transition bandwidth

Comparison of impulse responses ⇒

Rectangular vs. Blackman: impulse responses



- Blackman response is smoother (as expected from shape of window $w[n]$)

Summary of window characteristics ⇒

Characteristics of some commonly-used windows

The following table is taken from Oppenheim/Schafer/Buck (Table 7.1)

Windows are assumed to be L points long (recall $L = 2M + 1$)

Window Type	Peak Sidelobe	Approximate ML width	Peak approx. error $20\log_{10}\delta$	Equiv. Trans. width (Kaiser)
Rectangular	-13 dB	$4\pi/L$	-21 dB	$1.81\pi/(L-1)$
Bartlett	-25 dB	$8\pi/(L-1)$	-25 dB	$2.37\pi/(L-1)$
Hanning	-31 dB	$8\pi/(L-1)$	-44 dB	$5.01\pi/(L-1)$
Hamming	-41 dB	$8\pi/(L-1)$	-53 dB	$6.27\pi/(L-1)$
Blackman	-57 dB	$12\pi/(L-1)$	-74 dB	$9.19\pi/(L-1)$

Table can be used to define initial set of design parameters ⇒

In-class problem

We wish to design an FIR lowpass filter satisfying the specifications

$$\begin{aligned} 0.95 \le |H(e^{j\Omega})| \le 1.05, \qquad & 0 \le |\Omega| \le 0.25\pi \\ |H(e^{j\Omega})| \le 0.1, \qquad & 0.35\pi \le |\Omega| \le \pi \end{aligned}$$

by applying a window $w[n]$ to the impulse response $h_{\text{ideal}}[n]$ for the ideal discrete-time lowpass filter with cutoff $\Omega_c = 0.3\pi$. Which of the five windows listed in the table can be used to meet this specification? For each window that you claim will satisfy this specification, give the minimum length L required for the filter.

Solution to in-class problem

First decide window type based on ripple constraint. For this specification $\delta_1 = 0.05 < \delta_2 = 0.1$. Thus δ_1 is the more stringent constraint:

$$20\log_{10}(\delta_1) = 20\log_{10}(0.05) = -26\text{dB}$$

Based on the table, the Hanning, Hamming, and Blackman windows meet this criteria (using peak approx. error column).
The desired transition width is $\Delta\Omega = 0.1\pi$. Use approximate transition width of Kaiser window to get approximate lengths for the 3 types of filters:

$$\text{Hanning: } \Delta\Omega = 0.1\pi = \frac{5.01\pi}{L-1} \quad \longrightarrow L \approx 51$$

$$\text{Hamming: } \Delta\Omega = 0.1\pi = \frac{6.27\pi}{L-1} \quad \longrightarrow L \approx 64$$

$$\text{Blackman: } \Delta\Omega = 0.1\pi = \frac{9.19\pi}{L-1} \quad \longrightarrow L \approx 93$$

Summary: window design method

Basic idea: obtain an FIR lowpass filter by truncating the ideal lowpass filter impulse response

$$h[n] = w[n]h_{\text{ideal}}[n] \overset{\mathcal{F}}{\longleftrightarrow} H(e^{j\Omega}) = \frac{1}{2\pi} W(e^{j\Omega}) * H_{\text{ideal}}(e^{j\Omega})$$

- Window type determines the ripple
  Ripple is essentially the same in pass/stop bands so have to design to meet the more stringent specification.
- Window length determines the transition width.

Window design method guarantees that filter has linear phase $\Rightarrow$

Linear phase characteristic

Note that the filters obtained through the window method (as defined in these slides) are finite length, but not causal

- $h[n] = w[n]h_{\text{ideal}}[n]$ is a real and even signal, so $h[n] \neq 0$ for $n < 0$
- Since $h[n]$ is real and even, the phase of $H(e^{j\Omega})$ is restricted to be 0 or π (it is π when $H(e^{j\Omega})$ is negative

Can shift $h[n]$ to make the FIR system causal, *i.e*,

$$h_{\text{causal}}[n] = h[n - M/2]$$

Thus we multiply $H(e^{j\Omega})$ by a complex exponential (time-shifting property):

$$H_{\text{causal}}(e^{j\Omega}) = H(e^{j\Omega})e^{-j\Omega\frac{M}{2}}$$

This guarantees that the causal filter has a linear phase characteristic

Other FIR design methods

- Kaiser Window Method
  Tradeoff in ML width and SL area can be quantified by looking for a window function concentrated around $\Omega = 0$ in freq. domain:
  - Slepian: prolate spheroidal functions. Hard to work with!
  - Kaiser: near-optimal window can be created from a zeroth order Bessel function. Easier to work with! Kaiser window has 2 parameters: length$= M + 1$, shape$= \beta$

  Kaiser developed empirical formulas to choose M and β given the transition bandwidth and the ripple

  Note: transition bandwidth of equivalent Kaiser window often a better predictor of transition width than mainlobe width of window.
- Parks McClellan
  - Design criteria: minimize the maximum approximation error $\longrightarrow$ results in equiripple designs

References

1. A. V. Oppenheim and R. W. Schafer with J. R. Buck, *Discrete-Time Signal Processing*, Prentice Hall, Second Edition, 1999.
2. J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, Pearson/Prentice Hall, Fourth Edition, 2007.
3. F. J. Harris, “On the use of windows for harmonic analysis with the discrete Fourier transform,” *Proceedings of the IEEE*, vol. 66, pp. 51-83, January 1978.

Project-related questions

The following questions have been raised during office hours:

- How to define the cutoff frequency of the ideal filter ($h_{\text{ideal}}[n]$)?
- What is the length of the filter?
- What is the meaning of the dB scale?
- How to use Matlab to get the phase plots? How to interpret Matlab’s phase plots?
- How to construct bandpass filters with correct magnitude?

How to define the cutoff frequency Ω_c?

The filter specifications generally give values for the continuous-time passband frequency ω_p and the continuous-time stopband frequency ω_s. Using the frequency scaling discussed on slide 5 ($\Omega = \omega T$), we can determine values for the discrete-time passband frequency Ω_p and discrete-time stopband frequency Ω_s.

As In-class Problem 1 on 11/25/08 showed, the symmetry of the Fourier transform of the window guarantees that the transition band is centered at Ω_c. Thus half the transition band is to the left of Ω_c and half to the right.

This implies that we should choose the following value for Ω_c:

$$\Omega_c = \frac{\Omega_p + \Omega_s}{2}$$

What is the length of the filter?

As defined in these notes, the length of the filter is $2M + 1$. In the table containing the window properties, the window length is defined to be L. The window length must be the same as the filter length, thus $L = 2M + 1$.

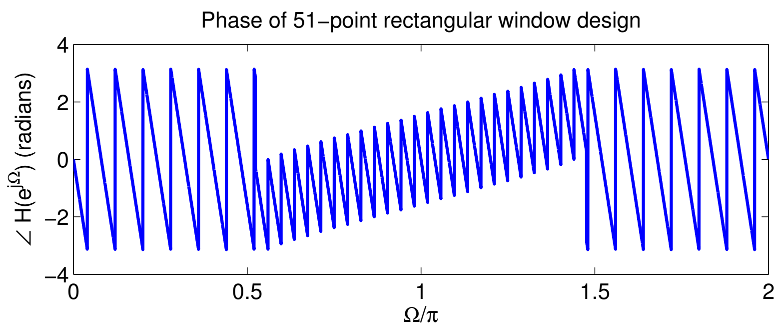
What is the meaning of the dB scale?

See footnote on page 437 of textbook by Oppenheim/Willsky/Nawab.

- Square of the magnitude of the Fourier transform can be interpreted as power.
- The square of the magnitude of the filter frequency response $|H(e^{j\Omega})|^2$ can be thought of as the power ratio between input and output of system.
- The term *bel* indicates a factor of 10 in a power ratio (in honor of Alexander Graham Bell).
- *decibel* denotes one-tenth of a *bel* on a log scale.
- $10\log_{10}|H(e^{j\Omega})|^2 = 20\log_{10}|H(e^{j\Omega})|$ is the number of decibels of power amplification for the system with frequency response $H(e^{j\Omega})$.

How to get phase plots? What do they mean?

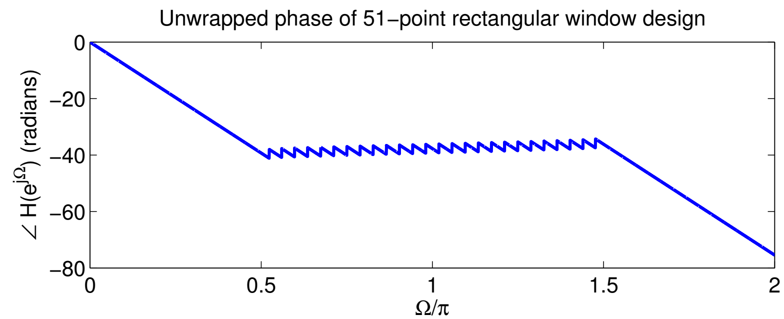
In Matlab, use the `angle` command to compute the phase of a complex number, *e.g.*, `plot(omega/pi,angle(H));` will produce a plot of the phase of the transform returned by the `freqz` command.



Is this linear? Unwrap the phase ⇒

Unwrapped phase

In Matlab, use the `unwrap` command to get the phase without wrap-around at the 2π point, *e.g.*, *e.g.*, `plot(omega/pi,unwrap(angle(H)));` will produce a plot of the unwrapped phase of the transform returned by the `freqz` command.



How to construct bandpass filters?

Slight change in project definition for Section 3, *i.e.*, incorporate a scaling factor into the definition:

$$h_{\mathrm{new}}[n] = h_{\mathrm{proto}}[n] A \cos(\omega_0 n)$$

A can be chosen to get the correct amplitude of the resulting bandpass filter.