

4. Linear-Phase FIR Digital Filter Design

4.1. Window (Windowing) Method

4.1.1. Basic Principles and Algorithms

Since $H(e^{j\omega})$, the frequency response of any digital filter, is a periodic in frequency, it can be expanded in a Fourier series. The resultant series is in the form

$$H(e^{j\omega}) = \sum_{k=-\infty}^{\infty} h(k)e^{-j\omega k}$$

where

$$h(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(e^{j\omega}) e^{j\omega n} d\omega$$

The coefficients of the Fourier series $h(k)$ are easily recognized as being identical to the impulse response of a digital filter. There are two difficulties with the application of the above given expressions for designing of FIR digital filters:

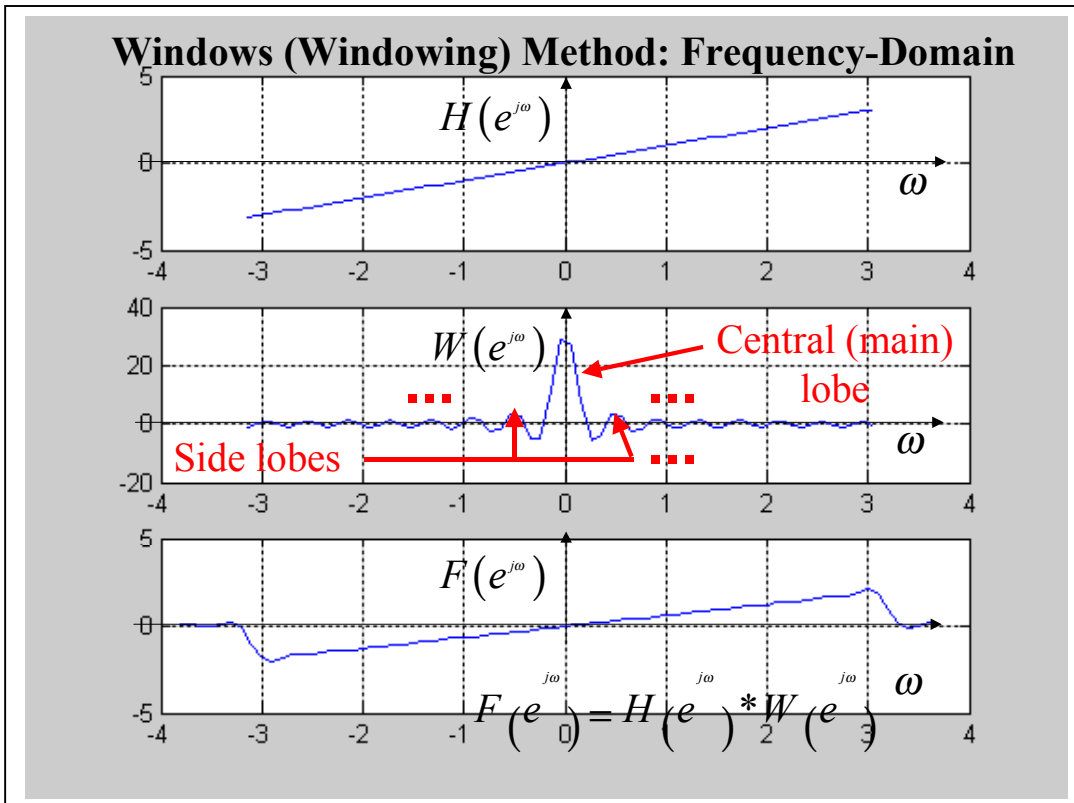
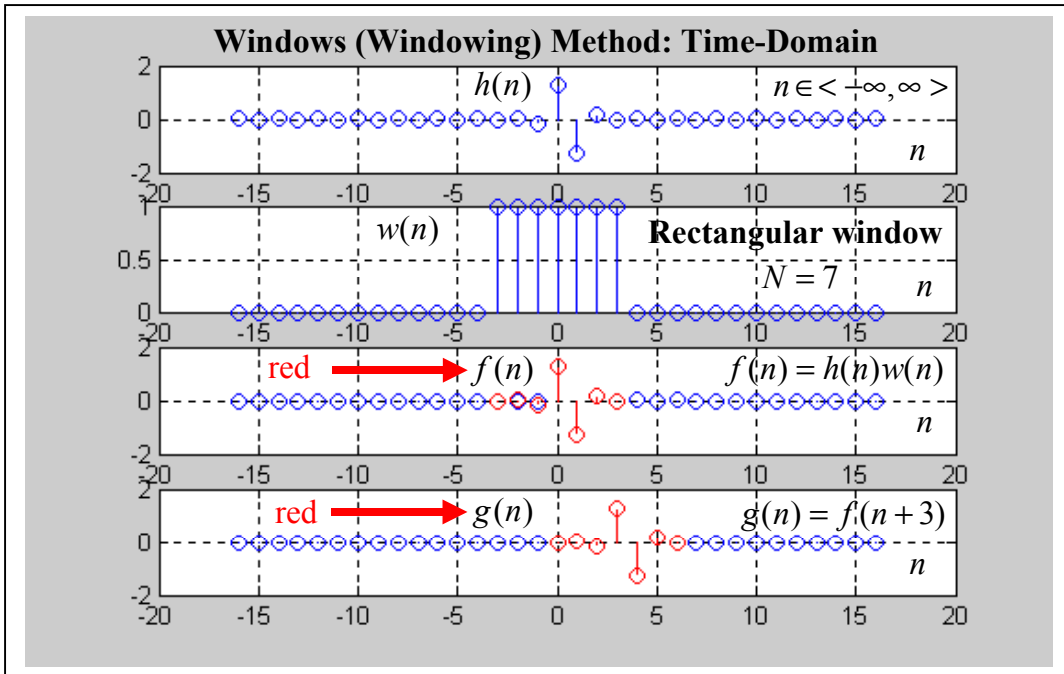
- The filter impulse response is infinite in duration, since the above given summation extends to $\pm\infty$.
- The filter is unrealizable (non-causal) because the impulse response begins at $-\infty$; i.e. no finite amount of delay can make the impulse response realizable.

Hence the filter resulting from a Fourier series representation of $H(e^{j\omega})$ is an unrealizable (non-causal) IIR filter. In spite of that fact, the causal FIR filter can be design by the approach illustrated at the next figures.

One possible way of obtaining an FIR filter that approximates $H(e^{j\omega})$ would be to truncate the infinite impulse response at $n = \pm N$. Direct truncation (see the previous figures) of the series leads to well known Gibbs phenomenon, however, which manifests itself as a fixed percentage overshoot and ripple before and after discontinuity in the frequency response. Thus for example, in the approximation of such standard filters as the ideal low-pass or band-pass filter, the largest ripple in the frequency response is about 9% of the size of discontinuity and its amplitude does not decrease with increasing impulse response duration – i.e. including more and more terms in the Fourier series does not decrease the amplitude of the largest ripple. Instead, the overshoot is confined to a smaller and smaller frequency range as N is increased. Since any reasonable design technique must be capable of designing good approximations to ideal low-pass filters, direct truncation is not a reasonable way of obtaining an FIR filter.

A possible way of obtaining an FIR filter is to use a finite weighting sequence $w(n)$, called a window, to modify the Fourier coefficients $h(n)$ to control the convergence of the Fourier series. The technique of windowing is illustrated in the previous figure.

At the top of these figures is shown the desired periodic frequency response $H(e^{j\omega})$ and its Fourier series coefficients $h(n)$. The next figure shows a finite duration weighing sequence $w(n)$ (in this case so-called rectangular window) with Fourier transform $W(e^{j\omega})$. $W(e^{j\omega})$, for most reasonable windows consists of a central lobe which contains most of the energy of the window and side lobes which generally decay rapidly. To produce an FIR approximation to $H(e^{j\omega})$, the sequence $f(n) = w(n)h(n)$ is formed. Outside the interval $-N \leq n \leq N$, $f(n)$ is zero exactly. The next figure shows $f(n)$ and its Fourier transform $F(e^{j\omega})$, which is readily seen to be the circular



convolution of $H(e^{j\omega})$ and $W(e^{j\omega})$, since $f(n)$ is the product of $h(n)$ and $w(n)$. Finally, the last figure shows the realizable sequence $g(n)$, which is shifted version of $f(n)$ and may be used as the desired filter impulse response.

Example:

By the impulse response truncation method (by the windowing method at rectangular window application) design a low-pass filter of order $N = 15$ with pass-band cut off frequency (pass-band edge frequency) $f_0 = 1 \text{ kHz}$. Frequency sampling is $f_S = 4 \text{ kHz}$.

Solution:

$$f_S = 4 \text{ kHz}; \quad f_0 = 1 \text{ kHz}, \quad \text{normalized cut off frequency: } \omega_0 = \frac{2\pi}{4 \cdot 10^3} \cdot 1 \cdot 10^3 = \frac{\pi}{2}$$

$$\text{Low-pass filter: } H(e^{j\omega}) = \begin{cases} 1 & \text{for } |\omega| \leq \omega_0 \\ 0 & \text{for } |\omega| > \omega_0 \end{cases}$$

$$\begin{aligned} h(n) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} H(e^{j\omega}) e^{j\omega n} d\omega = \frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} 1 e^{j\omega n} d\omega = \frac{1}{2\pi} \left[\frac{e^{j\omega n}}{jn} \right]_{-\pi/2}^{\pi/2} = \\ &= \frac{1}{\pi n} \left[\frac{e^{jn\frac{\pi}{2}}}{2j} - \frac{e^{-jn\frac{\pi}{2}}}{2j} \right] = \frac{1}{\pi n} \frac{e^{jn\frac{\pi}{2}} - e^{-jn\frac{\pi}{2}}}{2j} = \frac{\sin\left(n\frac{\pi}{2}\right)}{\pi n} \end{aligned}$$

$$h(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(e^{j\omega}) e^{j\omega 0} d\omega = \frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} 1 d\omega = \frac{1}{2\pi} [\omega]_{-\pi/2}^{\pi/2} = \frac{1}{2\pi} \left[\frac{\pi}{2} - \left(-\frac{\pi}{2} \right) \right] = 0.5$$

Another solution:

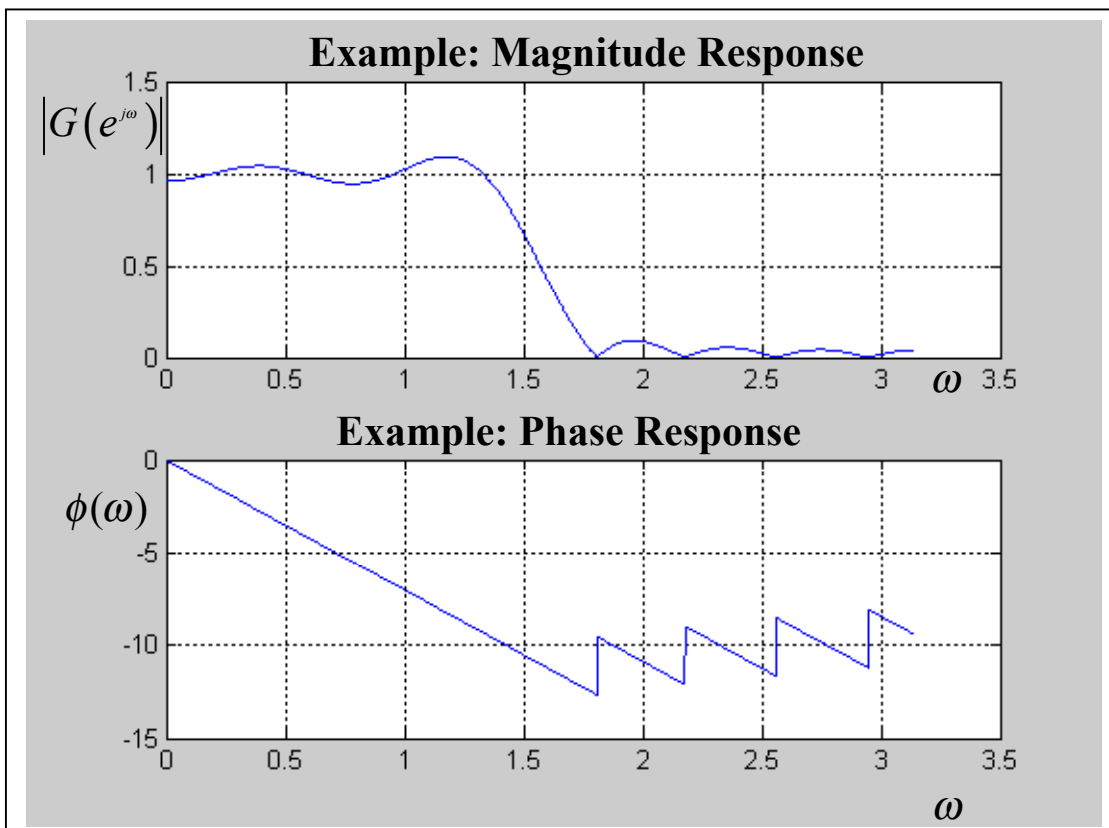
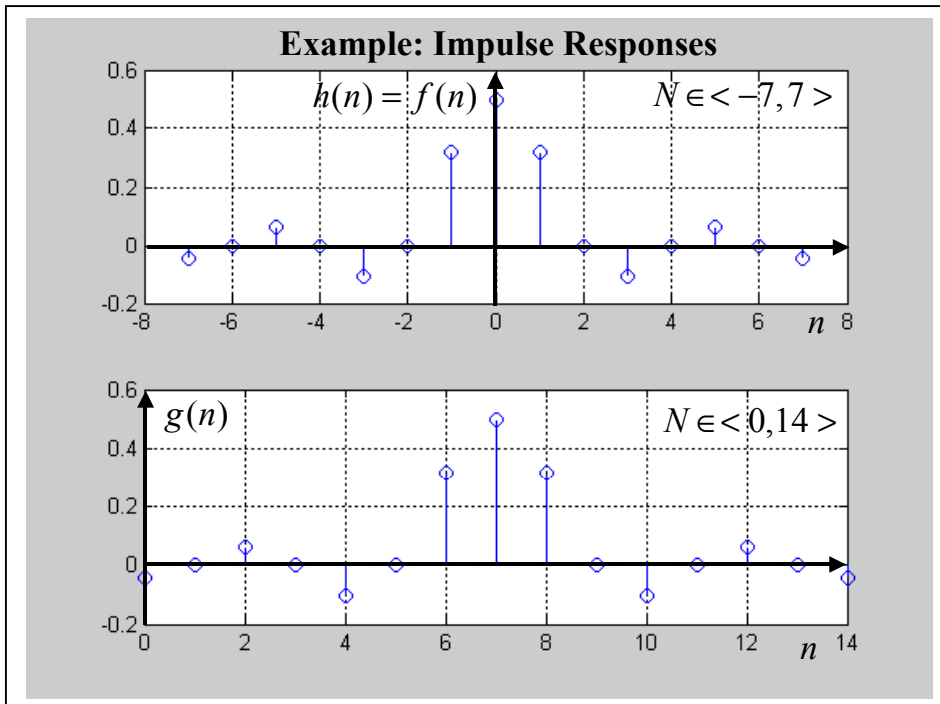
$$h(0) = \lim_{n \rightarrow 0} h(n) = \lim_{n \rightarrow 0} \frac{\sin\left(n\frac{\pi}{2}\right)}{\pi n} = \lim_{n \rightarrow 0} \frac{\sin\left(n\frac{\pi}{2}\right)}{n\frac{\pi}{2}} \cdot \frac{1}{2} = \frac{1}{2} = 0.5$$

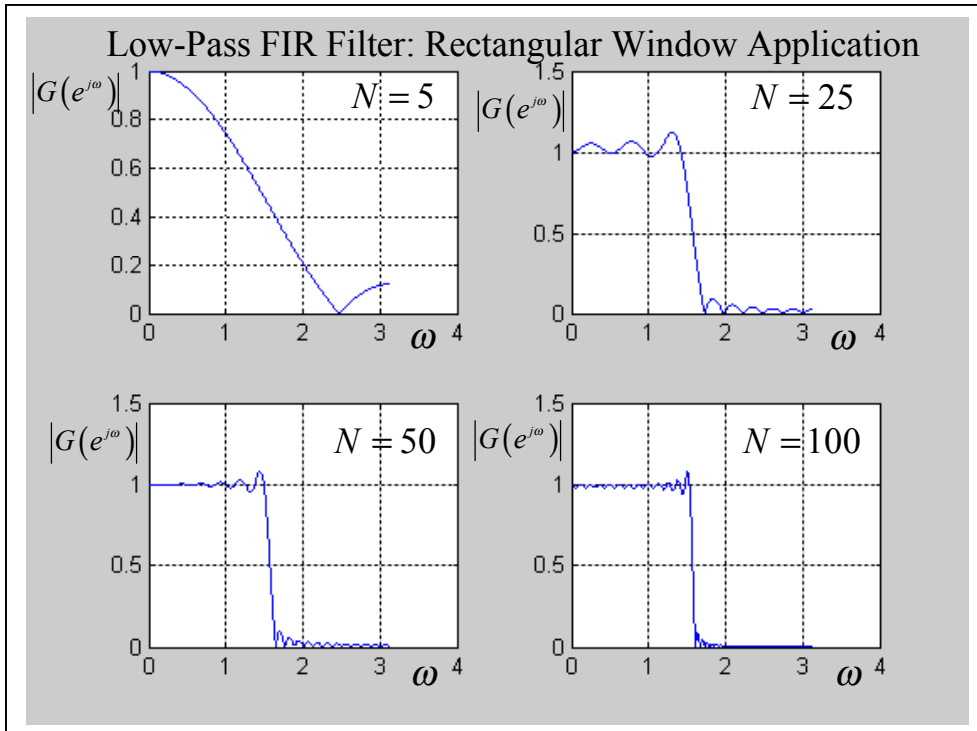
Because of rectangular window of application: $f(n) = h(n)$ for $n \in \langle -7, 7 \rangle$

n	$h(n) = f(n)$	n	$g(n)$
-7	-0.0455	0	-0.0455
-6	0.0000	1	0.0000
-5	0.0637	2	0.0637
-4	0.0000	3	0.0000
-3	-0.1061	4	-0.1061
-2	0.0000	5	0.0000
-1	0.3183	6	0.3183
0	0.5000	7	0.5000
1	0.3183	8	0.3183
2	0.0000	9	0.0000
3	-0.1061	10	-0.1061
4	-0.0000	11	-0.0000
5	0.0637	12	0.0637
6	0.0000	13	0.0000
7	-0.0455	14	-0.0455

The impulse responses, magnitude response and phase response are given in the next figures.

End.





4.1.2. Gibbs Phenomenon and Different Windowing

The above mentioned well known Gibbs phenomenon manifesting itself as a fixed percentage overshoot and ripple before and after discontinuity in the frequency response is illustrating in the next figures (magnitude response of the N -th order FIR low-pass filters with normalized cut off frequency 0.5π , for $N=5, 25, 50, 100$). The figures confirms the above given statements concerning the Gibbs phenomenon.

A major effect is that discontinuities of $H(e^{j\omega})$ became transition bands between values on the either side of the discontinuity. Since the final frequency response of the filter is the circular convolution of the ideal frequency response with the window's frequency response ($F(e^{j\omega}) = H(e^{j\omega}) * W(e^{j\omega})$), it is clear that the width of these transition bands depends on the width of the main (central) lobe of $W(e^{j\omega})$.

A second effect of the windowing is that the ripple from the side lobes of $W(e^{j\omega})$ produces approximation errors (ripple in the resulting frequency response) for all ω . Finally since the filter frequency response is obtained via convolution relation, it is clear that the resulting filters are never optimal in any sense, even though the windows from which they are obtained may satisfy some reasonable optimality criterion.

The discussion above leads to the questions of what are desirable window characteristics. This question can be answered relatively simply. Desirable window characteristics are:

- a) Small width of the main lobe of the frequency response of the window containing as much the total energy as possible.
- b) Side lobes of the frequency response that decrease in energy rapidly as ω tends to π .

There have been many windows proposed that approximate the desired characteristics. In the following sections several of them will be described.

4.1.2.1. Some Commonly Used Windows for FIR Filter Design

Window Type	Window Functions, $w(n)$, $-M \leq n \leq M$, $M = \frac{N-1}{2}$, $ w(n) = 0$ for $n > M$
Rectangular	$w(n) = 1$
Bartlett	$w(n) = 1 - \frac{ n }{M+1}$
Hann	$w(n) = \frac{1}{2} \left[1 + \cos \frac{2\pi n}{2M+1} \right]$
Hamming	$w(n) = 0.54 + 0.46 \cos \frac{2\pi n}{2M+1}$
Blackmann	$w(n) = 0.42 + 0.5 \cos \frac{2\pi n}{2M+1} + 0.08 \cos \frac{4\pi n}{2M+1}$
Kaiser (adjustable window) parameter: α	$w(n) = \frac{I_0 \left(\alpha \sqrt{1 - \left(\frac{n}{M} \right)^2} \right)}{I_0(\alpha)}$

Comments on Kaiser Window: $I_0(x)$ is the modified zero-th-order Bessel function of the first kind, which has the simple power series expansion

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

For most practical applications, about 20 terms in the above summation are sufficient to arrive at reasonably accurate values of $w(n)$.

The window functions are illustrating for $N=55$ in the next figures.

Example 1:

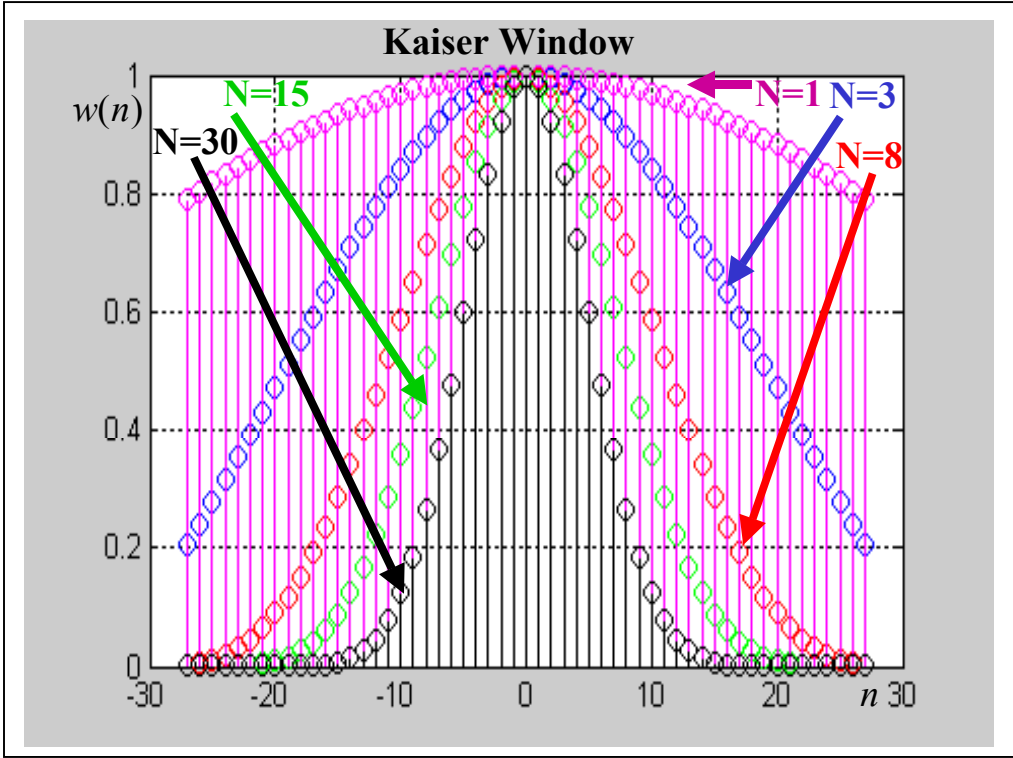
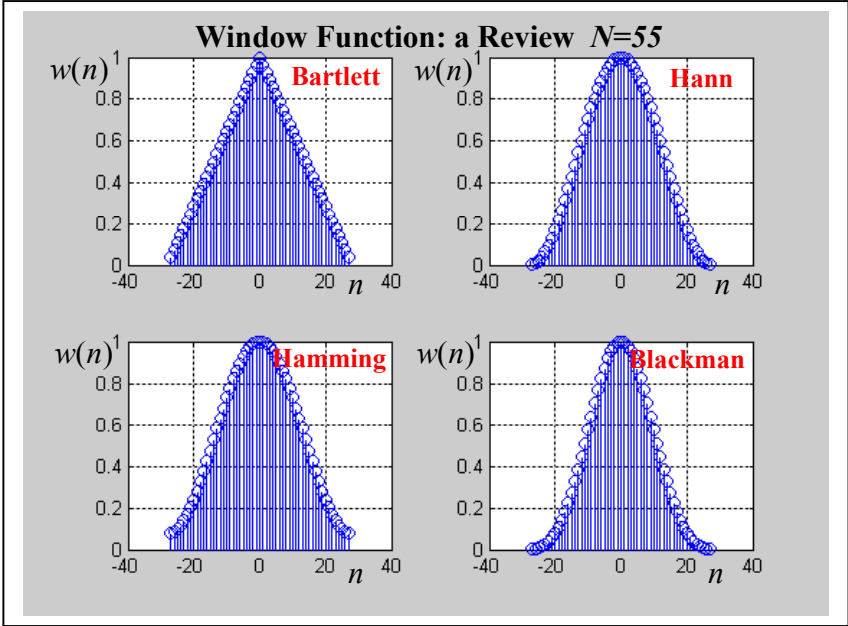
At the next figures, the illustration of the different windows is given. For the details concerning the window methods see the references.

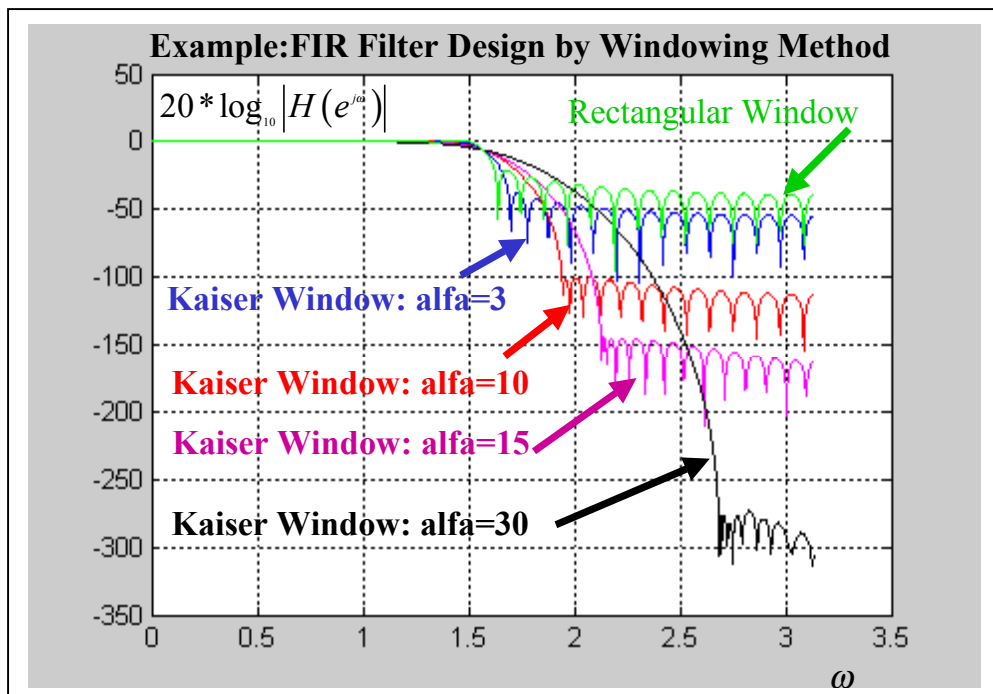
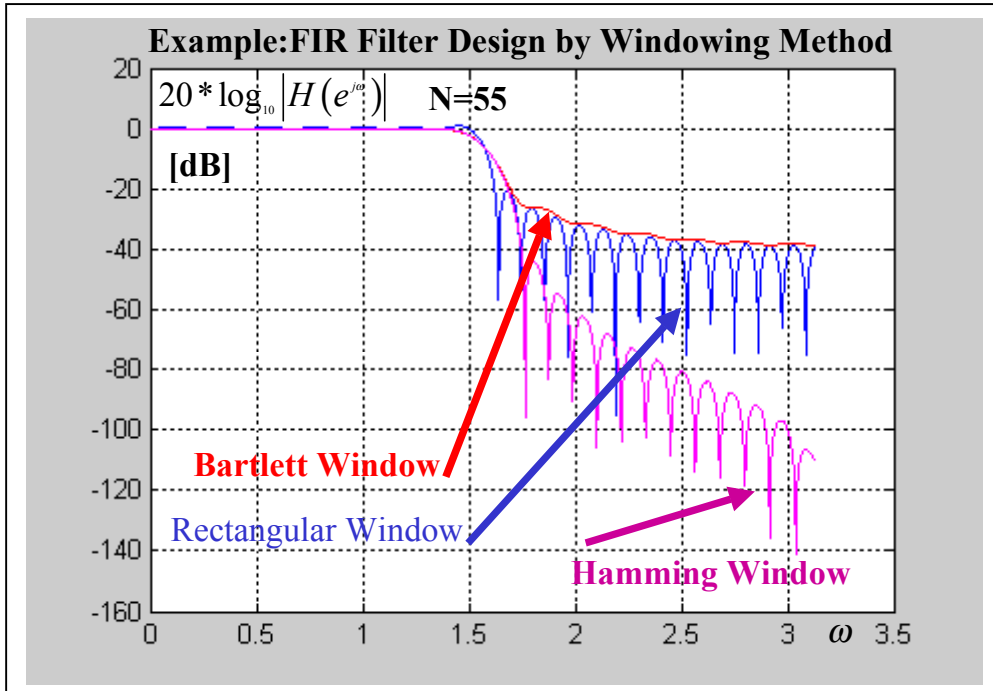
End.

Example2:

By the windowing design a low-pass filter of order $N = 55$ with pass-band cut off frequency (pass-band edge frequency) $f_0 = 1 \text{ kHz}$. Frequency sampling is $f_s = 4 \text{ kHz}$.

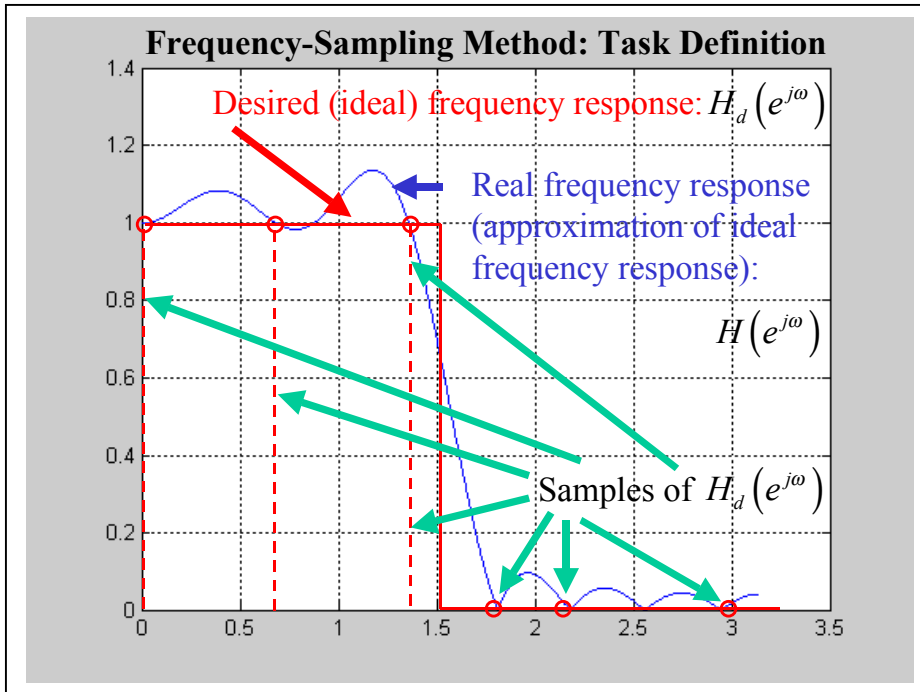
Solution: MATLAB function fir1.m. For the result see the next figures.





4.2. Frequency-Sampling Methods

To approximate any continuous frequency response, one could sample frequency response at N points along the interval frequencies $\langle 0, \pi \rangle$ and evaluate the continuous frequency response as an interpolation of the sampled frequency response. The approximation error would then be exactly zero at the sampling frequencies and be finite between them. The smoother the frequency response being approximated, the smaller the error of interpolation between the sample points. An example of this process is shown in the next figures. For the solution of this approximation – interpolation task (i.e. the evaluation of a FIR impulse response in order to fulfill the conditions of approximation) a number of the methods will be given in the next parts of our presentation.



4.2.1. Non-Uniform Frequency-Sampling Method

Each of the equations in the previous table constitutes a set of linear equations for determining the coefficients of a linear phase FIR filter. Consequently, if we specify response at either $(M+1)/2$ or $(M-1)/2$ or $M/2$ points in ω , we can solve the corresponding set of linear equations for the coefficients $h(n)$. Generally, the values of ω_k and $H(e^{j\omega_k})$ can be chosen arbitrarily. Therefore, The method is called as non-uniform frequency sampling method.

Example:

Build up a set of linear equations for the design of the 6-th order linear phase FIR filter, which has a symmetric unit sample response and a frequency response that satisfy the conditions:

$$\begin{aligned} \omega_1 = 0 \quad H_r(\omega_1) &= H_r(0) = 1 \\ \omega_2 = \frac{\pi}{3} \quad H_r(\omega_2) &= H_r\left(\frac{\pi}{3}\right) = .8 \\ \omega_3 = \frac{\pi}{2} \quad H_r(\omega_3) &= H_r\left(\frac{\pi}{2}\right) = .1 \end{aligned}$$

Solution:

If $M=6$, then the following expression for $H_r(\omega)$ is validate:

$$H_r(\omega) = 2 \sum_{k=0}^{\frac{M-1}{2}} h(k) \cos \omega \left(\frac{M-1}{2} - k \right) = 2 \sum_{k=0}^{\frac{6-1}{2}} h(k) \cos \omega \left(\frac{M-1}{2} - k \right) = 2 \sum_{k=0}^2 h(k) \cos \omega \left(\frac{6-1}{2} - k \right)$$

$$H_r(\omega) = 2h(0) \cos \omega \left(\frac{5}{2} \right) + 2h(1) \cos \omega \left(\frac{5}{2} - 1 \right) + 2h(2) \cos \omega \left(\frac{5}{2} - 2 \right)$$

$$H_r(\omega_k) = 2h(0) \cos \frac{5\omega_k}{2} + 2h(1) \cos \frac{3\omega_k}{2} + 2h(2) \cos \frac{\omega_k}{2}$$

$$\omega_1 = 0 \quad H_r(\omega_1) = H_r(0) = 1:$$

$$1 = 2h(0) \cos \frac{5 \cdot 0}{2} + 2h(1) \cos \frac{3 \cdot 0}{2} + 2h(2) \cos \frac{1 \cdot 0}{2}$$

$$1 = 2h(0) + 2h(1) + 2h(2)$$

$$\omega_2 = \frac{\pi}{3} \quad H_r(\omega_2) = H_r\left(\frac{\pi}{3}\right) = .8$$

$$0.8 = 2h(0) \cos \frac{5\pi}{6} + 2h(1) \cos \frac{3\pi}{6} + 2h(2) \cos \frac{\pi}{6}$$

$$\omega_3 = \frac{\pi}{2} \quad H_r(\omega_3) = H_r\left(\frac{\pi}{2}\right) = .1$$

$$0.1 = 2h(0) \cos \frac{5\pi}{8} + 2h(1) \cos \frac{3\pi}{8} + 2h(2) \cos \frac{\pi}{4}$$

The set of desired equations:

$$1 = 2h(0) + 2h(1) + 2h(2)$$

$$0.8 = 2h(0) \cos \frac{5\pi}{6} + 2h(1) \cos \frac{3\pi}{6} + 2h(2) \cos \frac{\pi}{6}$$

$$0.1 = 2h(0) \cos \frac{5\pi}{8} + 2h(1) \cos \frac{3\pi}{8} + 2h(2) \cos \frac{\pi}{4}$$

By solving of the above given set of linear equations we can obtain: $h(0), h(1), h(2)$

Because of symmetrical impulse response: $h(0) = h(5), \quad h(1) = h(4), \quad h(2) = h(3)$

Therefore, by using these results, the complete impulse response can be obtained.

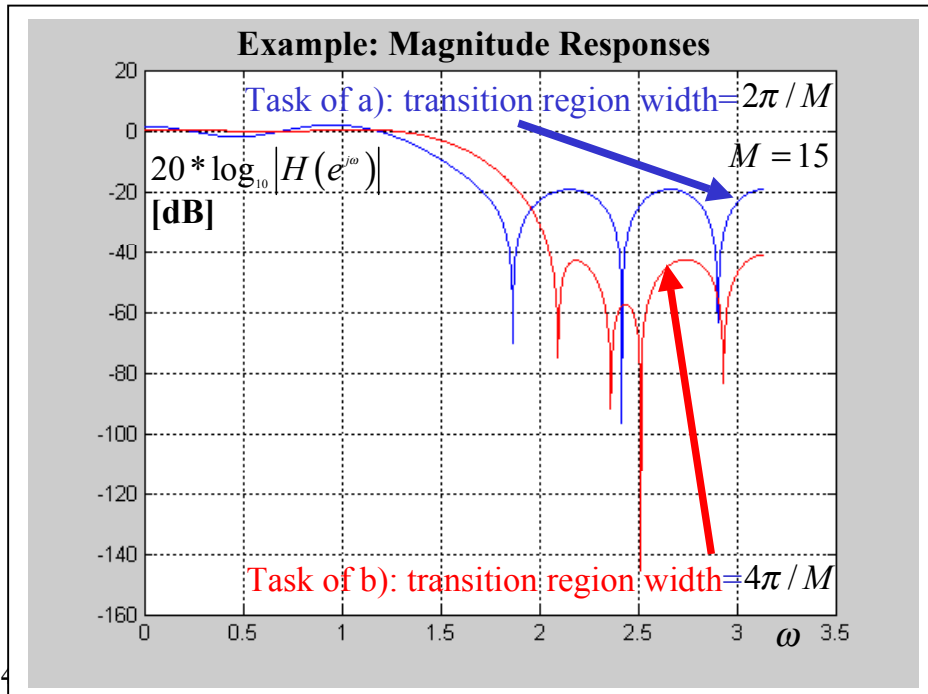
End.

4.2.2. Uniform Frequency-Sampling Method

Although the values of ω_k and $H(e^{j\omega_k})$ in the equations in the Table can be chosen arbitrarily, it is usually desirable to select equally spaced points in the frequency, in the range $0 \leq \omega \leq \pi$. Following this idea the frequencies can be selected in accordance with the next Table:

Comments on frequency specification for the linear phase FIR filter with antisymmetric unit sample response:

In this case, it is required specifications of the frequency response at $(M-1)/2$ points when M is odd and at $M/2$ points when M is even. Because in this case $H_r(0) = 0$, independent of our choice of $h(n)$, it is obvious that the frequency $\omega = 0$ cannot be used in the specification. Thus we may define the frequencies in accordance with Table 4.2.2.1 for $k = 1, 2, \dots$ (method A). An alternative choice of equally spaced frequencies that completely avoids the zeros at $\omega = 0$ and $\omega = \pi$ is a set of frequencies selected in accordance with methods B.



In the case of a linear phase FIR filter design by frequency sampling methods, the values of $H_r(\omega_k)$ have to be specified. The simplest approach is based on this method:

- $H_r(\omega_k) = 1$ for ω_k from the filter pass-band.
- $H_r(\omega_k) = 0$ for ω_k from the filter stop-band.
- $H_r(\omega_k)$ is not specified in the transient-band, i.e. the transient-band width is $\frac{2\pi}{M}$.

In this form of $H_r(\omega)$ specification, $H_r(\omega)$ is changed abruptly from the pass-band, where $H_r(\omega) = 1$, to the stop-band, where $H_r(\omega) = 0$. Instead of having an abrupt change, if we specify an intermediate value of $H_r(\omega)$ in the transition band, the resulting frequency response has significantly smaller side lobes in the stop-band. The following example illustrates this point.

Example:

Determine the coefficients $h(n)$ of a linear phase FR filter of length $M=15$ which has a symmetric unit sample response and a frequency response that satisfies the condition:

$$a) H_r\left(\frac{2\pi k}{15}\right) = \begin{cases} 1 & k = 0, 1, 2, 3 \\ 0 & k = 4, 5, 6, 7 \end{cases} \quad b) H_r\left(\frac{2\pi k}{15}\right) = \begin{cases} 1 & k = 0, 1, 2, 3 \\ 0.4 & k = 4 \\ 0 & k = 5, 6, 7 \end{cases}$$

The magnitude response of the resulting filter are illustrated in the next figures. In the case of the task of a), we observe that the filter has an overshoot at the edge of the pass-band, just prior to the transition region. It also has relatively large side lobes in the stop-band. In the case of the task b) it can be noted that the side lobes are significantly lower than in the previous design.

End.

4.2.2.2. Solution for Optimization

This example clearly illustrates the advantages in allowing a small transition in the design of filters. The only disadvantage is that the width of the transition region is increased. Specifically, the transition region is now about twice as wide as in the above example. However, the benefits usually outweigh the one disadvantage. Further reduction in the stop-band side-lobes can be obtained by allowing for one or more (usually one or two) additional frequency specifications in the transition bands. In effect, the transition band is widened further to achieve additional reduction in the stop-band side lobes.

The suitable values of the specifications in the transition band can be found experimentally (e.g. by using MATLAB software). The problem has been considered in the technical literature (e.g. ref. [4]). The optimum values of the specification in the transition band are tabulated available in some references.

In order to solve for optimum values of the unconstrained frequency samples, a series of equation must be written and solved that mathematically expresses the desired optimization. Let us assume that the frequency samples in the transition band are labeled T1 and T2 for convenience. These samples are ones to be optimized. In order to find optimum values T1 and T2 , a set of constraint equations must be written for frequencies in the pass-band and stop-band. Typical sets of such constraints might be:

a)

$$\left| H(e^{j\omega}) - H_d(e^{j\omega}) \right| \leq \varepsilon \quad \text{for } \omega \text{ in the pass-band}$$

$$\min \left\{ \max \left| H(e^{j\omega}) - H_d(e^{j\omega}) \right| \right\} \quad \text{for } \omega \text{ in the stop-band by selection of T1 and T2,}$$

where $\varepsilon > 0$ is a fixed tolerance, $H_d(e^{j\omega})$ and $H(e^{j\omega})$ are the desired (ideal) frequency response and the real frequency response function, respectively.

b)

$$\min \left\{ \max \left| W(e^{j\omega}) \left[H(e^{j\omega}) - H_d(e^{j\omega}) \right] \right| \right\} \quad \text{by selection of T1 and T2,}$$

where $W(e^{j\omega})$ is a known weighting function on the frequency response approximation error.

The sets of these constraints may be solved by the methods linear programming.