FIR Filter Design: Part I

1. Introduction

In this set of notes, we continue our exploration of the frequency response of FIR filters. First, we consider some "intuitive" examples of *low-pass* and *high-pass* FIR filters; we had previously considered these examples in our course introduction (see 1/16 lecture notes). Second, we consider the design of FIR filters with more desirable frequency response characteristics. We consider two approaches. In the first approach, we derive low-pass FIR filters from the impulse response of and *ideal* low-pass filter; that is, a filter with sharp cut-off frequencies between the passband (frequency range for which frequencies are allowed through the system), and the rejection-band (frequency range for which frequencies are zeroed). As we will see, this approach leads to FIR filters with overshoot characteristics near the cutoff frequency. To mitigate this problem, we then consider a second approach, where we allow a finite-length transition between the passband and rejection-band, and observe that the resulting FIR filters more closely approximate an idealized frequency response.

2. Simple low-pass FIR filters[1]

A. Introduction

Previously (1/16 lecture notes), we had considered *low-pass* FIR filters of the following form:

$$y[n] = \sum_{k=0}^{L-1} \frac{1}{L} x[n-k] \tag{1}$$

for which the output $y[n]$ is an average of the current input $x[n]$ and the $(L-1)$ previous inputs $x[n-k]$, $k \in \{1, 2, \ldots, L-1\}$. This type of FIR filter is known as a *running-average filter*, and tends to attenuate high frequencies in an input signal, while leaving lower frequencies relatively untouched (hence, the label "low-pass filter").

For these filters, the frequency response function $H(e^{\mathbf{j}\theta})$ is given by,

$$H(e^{\mathbf{j}\theta}) = \sum_{n=0}^{L-1} \frac{1}{L} e^{-\mathbf{j}n\theta}. \tag{2}$$

B. Examples

Here we consider the frequency response of filters of type (1), for $L = 10$, $L = 30$ and $L = 100$. Figure 1 below plots the magnitude and phase part of the frequency response for each of these filters as a function of the frequency variable $\theta \in [-\pi, \pi]$. Note that the larger the value of L, the more narrow the central peak about $\theta = 0$; that is, longer running averages will tend to suppress frequency components in the input signal of lower and lower frequency. The zero-frequency component ($\theta = 0$), however, remains unaffected no matter how large the value of L, since averaging a constant input sequence will not change that constant value.

To verify the frequency response characteristics of these filters, we now conduct the following experiment. We sample a continuous-time, noisy 4Hz cosine wave at $f_s = 500$ Hz and then pass the resulting discrete-time sequence $x[n]$ through the three filters above. That is,

$$x(t) = \cos(8\pi t) \tag{3}$$

$$x[n] = x(n/f_s) + noise, \; -\infty < n < \infty, \tag{4}$$

where, for each sample, the *noise* component of the signal consists of a uniformly distributed random number in the interval $[-1/2, 1/2]$. In Figures 2, 3 and 4, for each of the three filters, we plot a 500-length (1second) part of the input sequence $x[n]$ and a 500-length part of the output sequence $y[n]$. Moreover, we plot the magnitude FFT of each 500-length discrete-time sequence as a function of frequency f. Note that for

1. You may want to review the 1/16 lecture notes to complement the materials in this section.

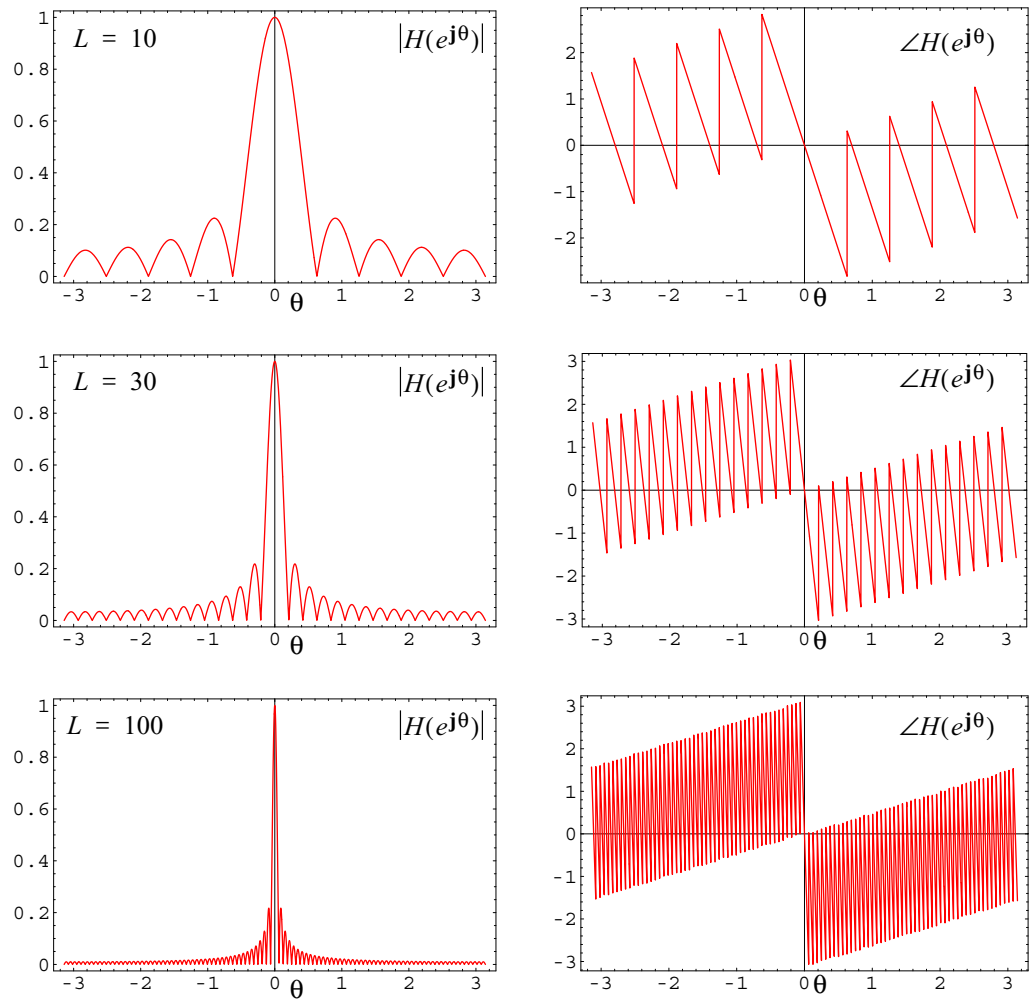

Figure 1

this sampling process, the relationship between the frequency variable θ in Figure 1 and the real frequency f (in Hertz) is given by,

$$\theta = \frac{2\pi f}{f_s} \tag{5}$$

so that the frequency variable range,

$$\theta \in [-\pi, \pi] \tag{6}$$

corresponds to the frequency range,

$$f \in [-f_s/2, f_s/2] = [-250\text{Hz}, 250\text{Hz}]\,. \tag{7}$$

A few observations: first, note that the uniform noise added to the 4Hz cosine waveform at the input shows up as non-zero frequencies throughout the frequency range $[-f_s/2, f_s/2]$; second, note that each filter tends to smooth out the noisy cosine waveform by suppressing most of the noise-induced frequency components — the longer the filter, the more the noise-induced frequencies are suppressed, and, consequently, the smoother the resulting output waveform $y[n]$.

In the following section, we reconsider some high-pass filters that we derived "intuitively" in an earlier lecture (1/16).

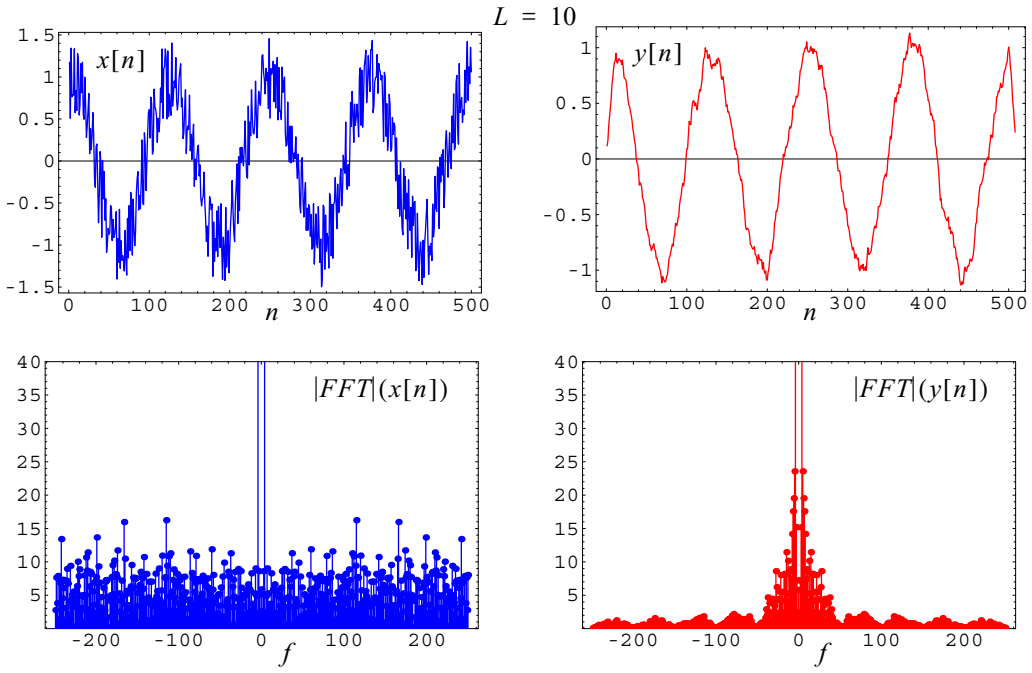

Figure 2

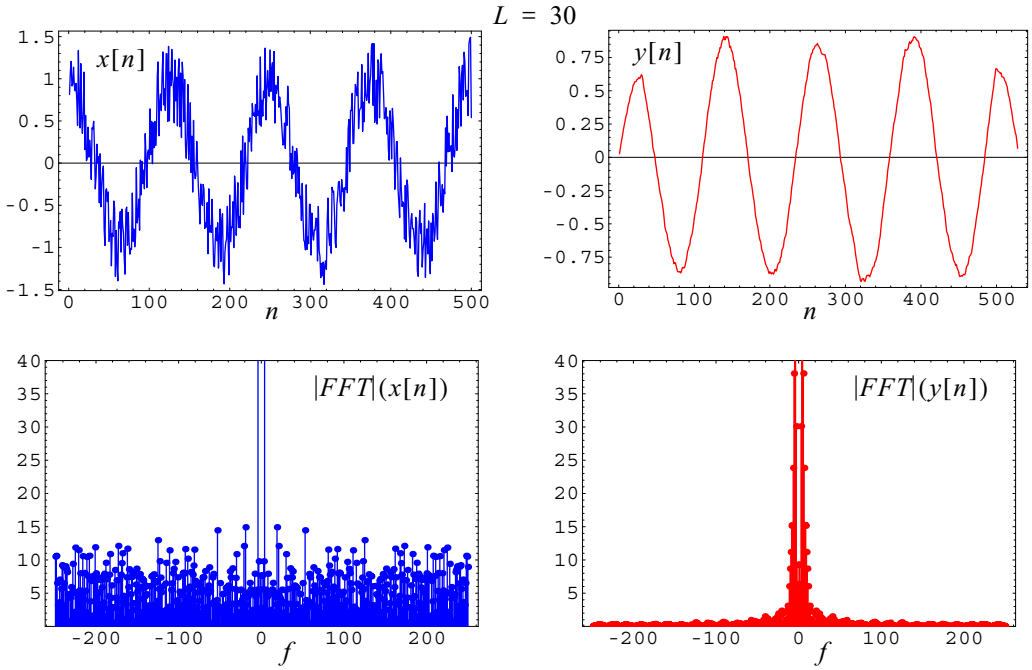

Figure 3

3. Simple high-pass FIR filters[1]

A. Introduction

Previously (1/16 lecture notes), we had considered the following *high-pass* FIR filters:

1. You may want to review the 1/16 lecture notes to complement the materials in this section.

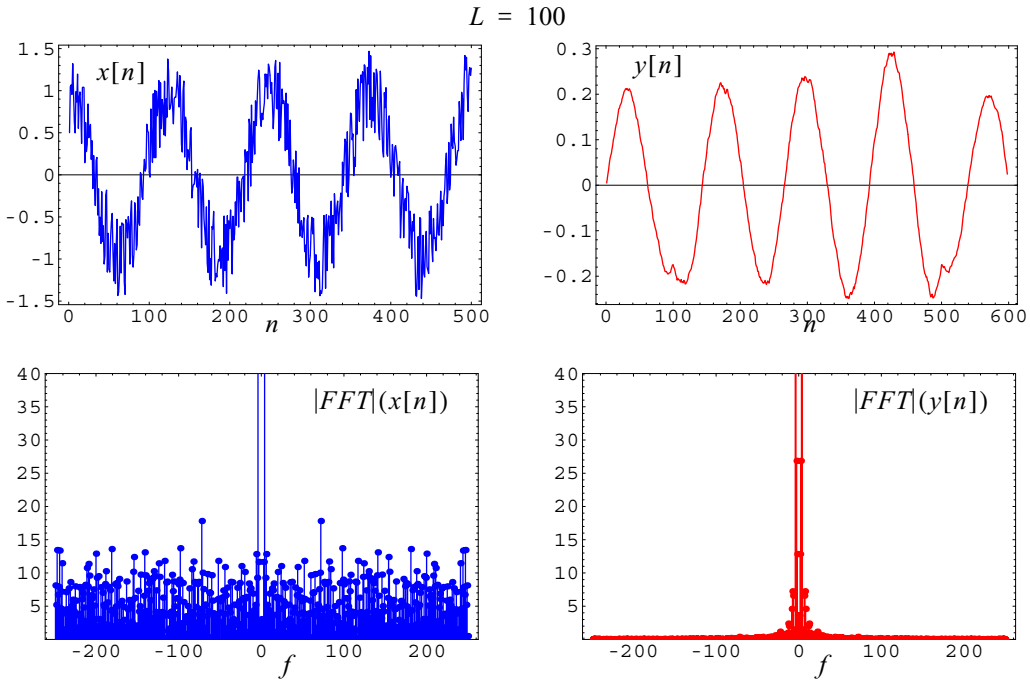

Figure 4

$$y_1[n] = 1/4x[n] - 1/2x[n-1] + 1/4x[n-2] \text{ (second-order)} \tag{8}$$

$$y_2[n] = 1/8x[n] - 3/8x[n-1] + 3/8x[n-2] - 1/8x[n-3] \text{ (third-order)} \tag{9}$$

$$y_3[n] = 1/16x[n] - 1/4x[n-1] + 3/8x[n-2] - 1/4x[n-3] + 1/16x[n-4] \text{ (fourth-order)} \tag{10}$$

which we derived by taking consecutive discrete-time derivatives of the first-order high-pass filter:

$$y[n] = 1/2x[n] - 1/2x[n-1]\,. \tag{11}$$

These filters will tend to attenuate low frequencies (zeroing any DC or zero-frequency component), while passing through higher frequencies relatively untouched (hence, the label “high-pass filter”).

For these filters, the frequency response functions are straightforward to compute:

$$H_1(e^{\mathbf{j}\theta}) = \frac{1}{4} - \frac{1}{2}e^{-\mathbf{j}\theta} + \frac{1}{4}e^{-\mathbf{j}2\theta} \tag{12}$$

$$H_2(e^{\mathbf{j}\theta}) = \frac{1}{8} - \frac{3}{8}e^{-\mathbf{j}\theta} + \frac{3}{8}e^{-\mathbf{j}2\theta} - \frac{1}{8}e^{-\mathbf{j}3\theta} \tag{13}$$

$$H_3(e^{\mathbf{j}\theta}) = \frac{1}{16} - \frac{1}{4}e^{-\mathbf{j}\theta} + \frac{3}{8}e^{-\mathbf{j}2\theta} - \frac{1}{4}e^{-\mathbf{j}3\theta} + \frac{1}{16}e^{-\mathbf{j}4\theta}\,. \tag{14}$$

Figure 5 below plots the magnitude and phase part of the frequency response for each of these filters as a function of the frequency variable $\theta \in [-\pi, \pi]$. Note that as the length of the filter is increased, more and more low frequencies are virtually zeroed out entirely, while the highest frequencies remain untouched.

B. Experiments

To verify the frequency response characteristics of the high-pass filters in equations (12) through (14), we now conduct the same experiment as for the low-pass filters in the previous section. We sample a continuous-time, noisy 4Hz cosine wave at $f_s = 500$ Hz and then pass the resulting discrete-time sequence $x[n]$ through the three high-pass filters above. That is,

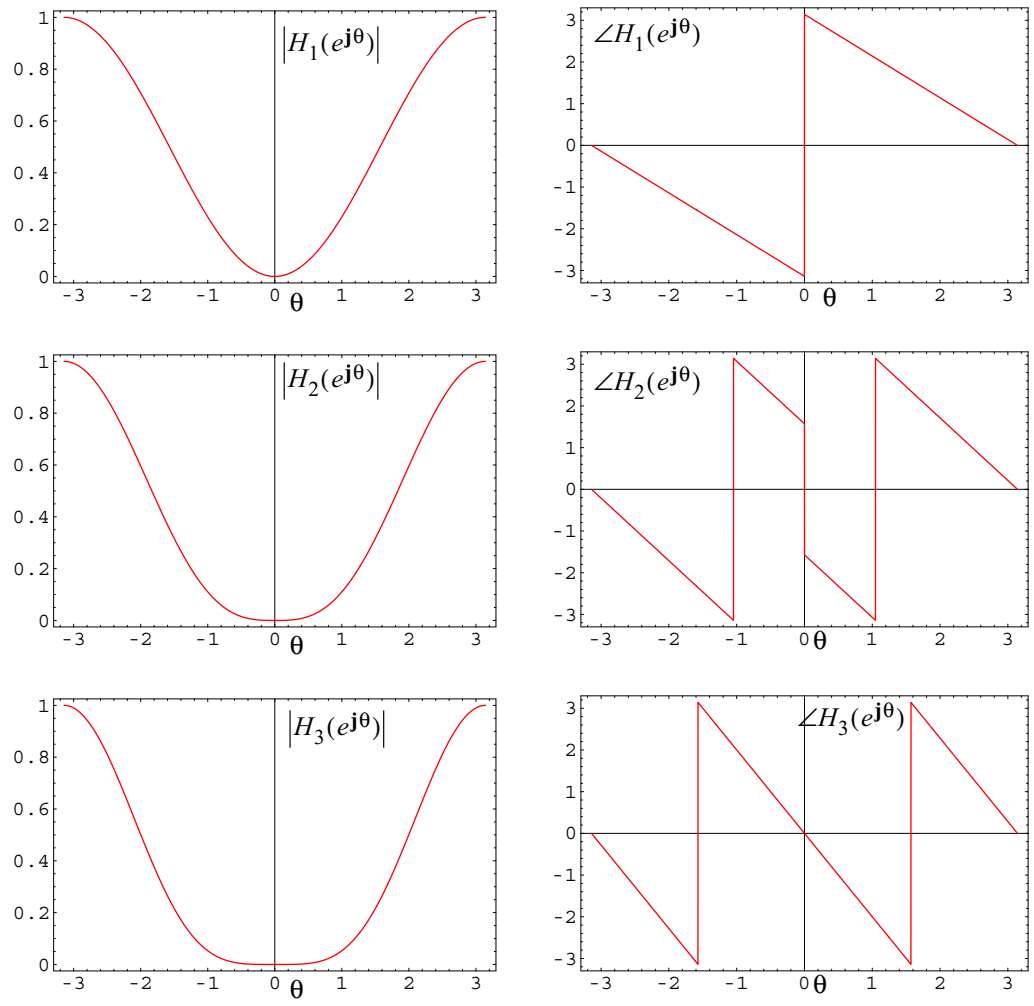

Figure 5

$$x(t) = \cos(8\pi t) \tag{15}$$

$$x[n] = x(n/f_s) + noise\,,\ -\infty < n < \infty\,, \tag{16}$$

where, for each sample, the *noise* component of the signal consists of a uniformly distributed random number in the interval $[-1/2,\ 1/2]$. In Figures 6, 7 and 8, for each of the three high-pass filters, we plot a 500-length (1second) part of the input sequence $x[n]$ and a 500-length part of the output sequence $y[n]$. Moreover, we plot the magnitude FFT of each 500-length discrete-time sequence as a function of frequency f. Note that for this sampling process, the relationship between the frequency variable θ in Figure 5 and the real frequency f (in Hertz) is again given by,

$$\theta = \frac{2\pi f}{f_s} \tag{17}$$

so that the frequency variable range,

$$\theta \in [-\pi,\ \pi] \tag{18}$$

corresponds to the frequency range,

$$f \in [-f_s/2, f_s/2] = [-250\text{Hz},\ 250\text{Hz}]\,. \tag{19}$$

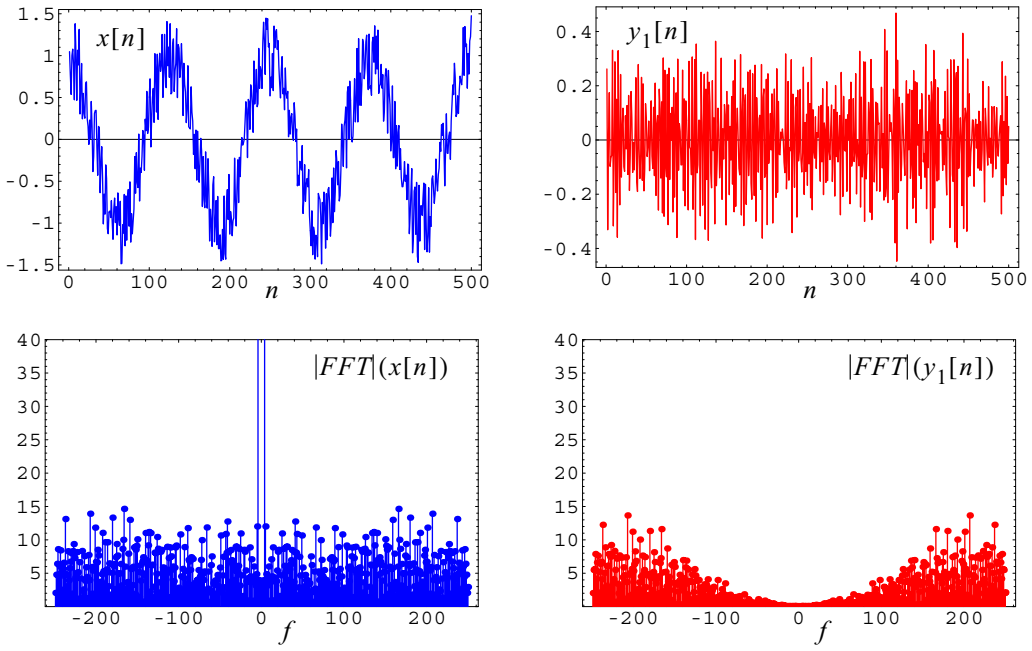

Figure 6

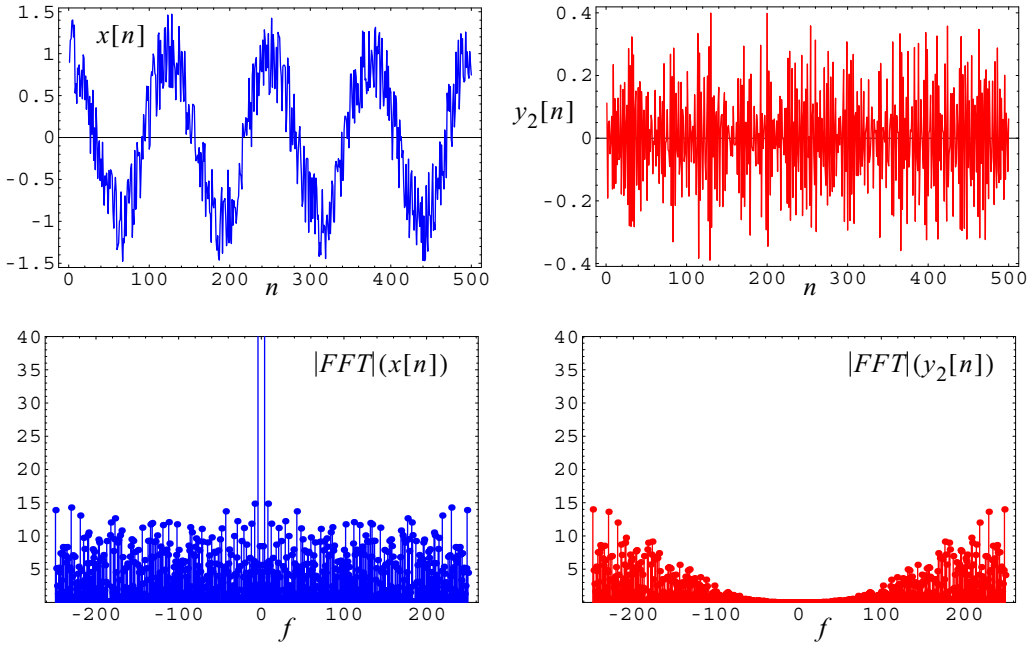

Figure 7

Note that each filter removes the low-frequency cosine wave from the signal and preserves only high-frequency components (to varying degrees). Also note that these experimental results closely match the magnitude frequency response functions in Figure 5.

The *Mathematica* notebook "fir_filter_design_part1a.nb" was used to generate the examples in the previous two sections. Additional speech and image processing examples for the low-pass and high-pass filters discussed so far can be found in the notes corresponding to the 1/16 lecture. In the next two sections, we illustrate how one might go about designing low-pass FIR filters, where you, as the designer, have much more control over the filter's resulting frequency response.

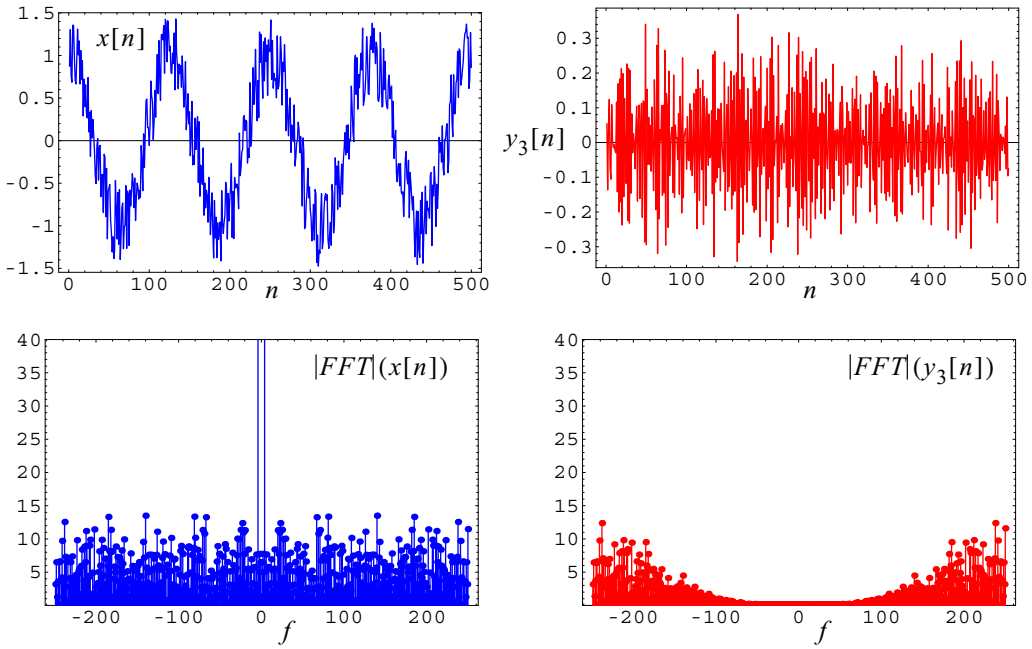

Figure 8

4. Design of a low-pass FIR filter: approach #1

A. Introduction

In the previous two sections, we saw examples of two types of simple FIR filters: (1) low-pass and (2) high-pass. Although these filters were relatively simple to derive, we had little control over the precise shape of the frequency response function corresponding to each filter. Here, we consider the design of a low-pass filter, beginning with an ideal low-pass filter with zero-width transitions between the passband and rejection-band.

B. Ideal low-pass filter

Figure 9 illustrates the frequency response of an ideal low-pass filter with normalized cutoff frequency at $\theta = \theta_c$. Note that a filter with these characteristics will not change the magnitude of any frequencies in the input signal for the range $\theta \in [-\theta_c, \theta_c]$, but will zero out all frequencies outside that range. Moreover, the linear-phase property of such a filter will shift input frequencies in the range $\theta \in [-\theta_c, \theta_c]$ by an amount proportional to each frequency and, therefore, will not introduce any phase distortion into the output of the system. The slope of the line $(-a)$ in the phase plot will determine the extent of the shift; a slope of zero will correspond to no phase shift (i.e. no delay).

Below, we will use the inverse DTFT to determine the time-domain impulse response $h[n]$ of an ideal filter with $a = 0$ (i.e. no delay). Recall that the inverse DTFT is given by,

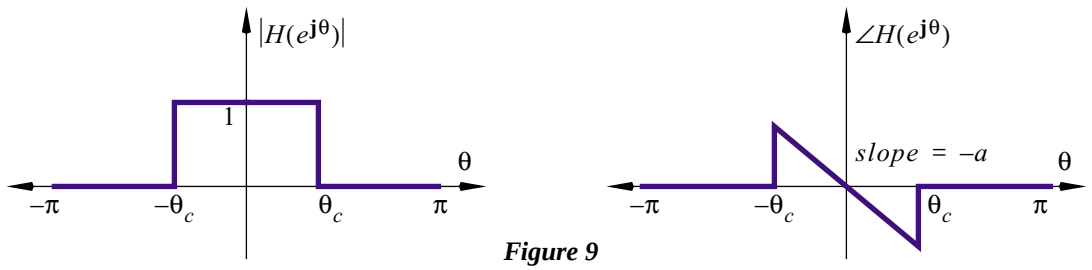

Figure 9

$$h[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} H(e^{\mathbf{j}\theta})e^{\mathbf{j}n\theta}d\theta\,. \tag{20}$$

For the ideal filter in Figure 9 (with no delay), equation (20) reduces to,

$$h[n] = \frac{1}{2\pi}\int_{-\theta_c}^{\theta_c} e^{\mathbf{j}n\theta}d\theta \tag{21}$$

$$h[n] = \frac{1}{2\pi}\frac{e^{\mathbf{j}n\theta}}{\mathbf{j}n}\bigg|_{\theta=-\theta_c}^{\theta=\theta_c} = \frac{1}{2\pi}\left(\frac{e^{\mathbf{j}n\theta_c}}{\mathbf{j}n} - \frac{e^{-\mathbf{j}n\theta_c}}{\mathbf{j}n}\right) = \frac{1}{n\pi}\left(\frac{e^{\mathbf{j}n\theta_c} - e^{-\mathbf{j}n\theta_c}}{\mathbf{j}2}\right) = \frac{\sin(n\theta_c)}{n\pi} \tag{22}$$

$$h_{ideal}[n] = \frac{\sin(n\theta_c)}{n\pi},\ -\infty < n < \infty\,. \tag{23}$$

Note that the impulse response for an ideal filter with no delay is infinite in length, both forwards and backwards in time. For example, in Figure 10, we plot $h[n]$, $-50 \le n \le 50$, for $\theta_c = 1$.

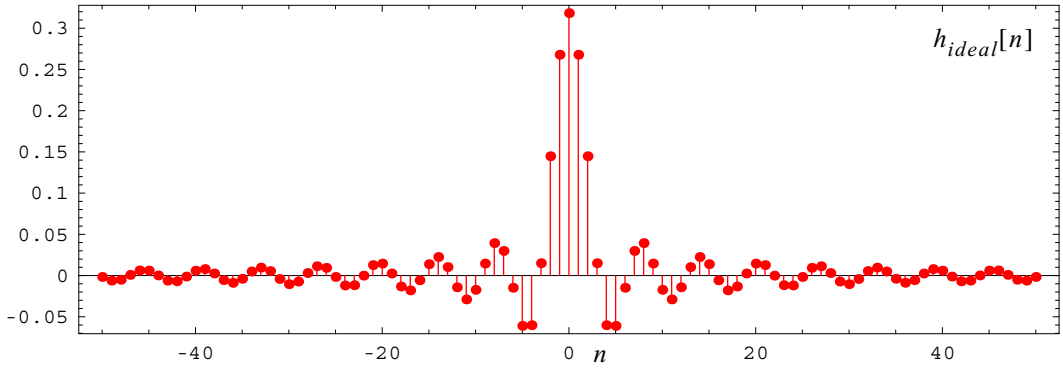

Figure 10

C. Approximating an ideal filter with an FIR filter

Now, we will try to approximate the ideal filter with a causal FIR filter using the following procedure: First we will retain values for $h_{ideal}[n]$ only for a limited range of n,

$$-n_{max} \le n \le n_{max} \tag{24}$$

since values of $h[n]$ approach zero as $|n| \to \infty$. Let us denote this finite impulse response as $h_1[n]$, such that,

$$h_1[n] = \begin{cases} h_{ideal}[n] & -n_{max} \le n \le n_{max} \\ 0 & elsewhere \end{cases} \tag{25}$$

Second, we will shift $h_1[n]$ to be causal; let us denote the resulting impulse response as $h_2[n]$ such that,

$$h_2[n] = h_1[n - n_{max}]\,. \tag{26}$$

Note that this shift will not change the steady-state frequency response of the resulting system, except to introduce a delay at the output. Thus, the difference equation corresponding to $h_2[n]$ can be written as:

$$y[n] = \sum_{k=0}^{2n_{max}} b_k x[n-k] \tag{27}$$

where,

$$b_k = h_{ideal}[-n_{max}+k] = \frac{\sin((-n_{max}+k)\theta_c)}{(-n_{max}+k)\pi} \tag{28}$$

In Figure 11, we plot the impulse response $h_2[n]$ and the corresponding magnitude frequency response $\left|H_2(e^{j\theta})\right|$ for various values of n_{max} (10, 20 and 50) and a numeric cutoff frequency of $\theta_c = 1$. For each frequency response plot, the blue line indicates the ideal magnitude frequency response.

Note that as we increase the length of the FIR filter (i.e. n_{max}) the transition region at the cutoff frequency becomes more well-defined; however, also note the overshoot phenomenon (Gibbs phenomenon) at the cutoff frequency, which is similar to what we observed earlier in the course for the finite-sum Fourier series approximation of periodic waveforms with discontinuities (such as a square wave). While increasing the FIR filter length will make the transition region in the frequency response function increasingly narrow, the overshoot at the cutoff frequency will remain. Therefore, this approach for designing low-pass FIR filters will yield filters that approximate the ideal filter characteristic arbitrarily well, *except* near $\theta = \pm\theta_c$, where the overshoot problem will remain. This problem motivates our second approach to filter design in the following section, where we explicitly allow some narrow, nonzero transition region between the passband and rejection band of the filter.

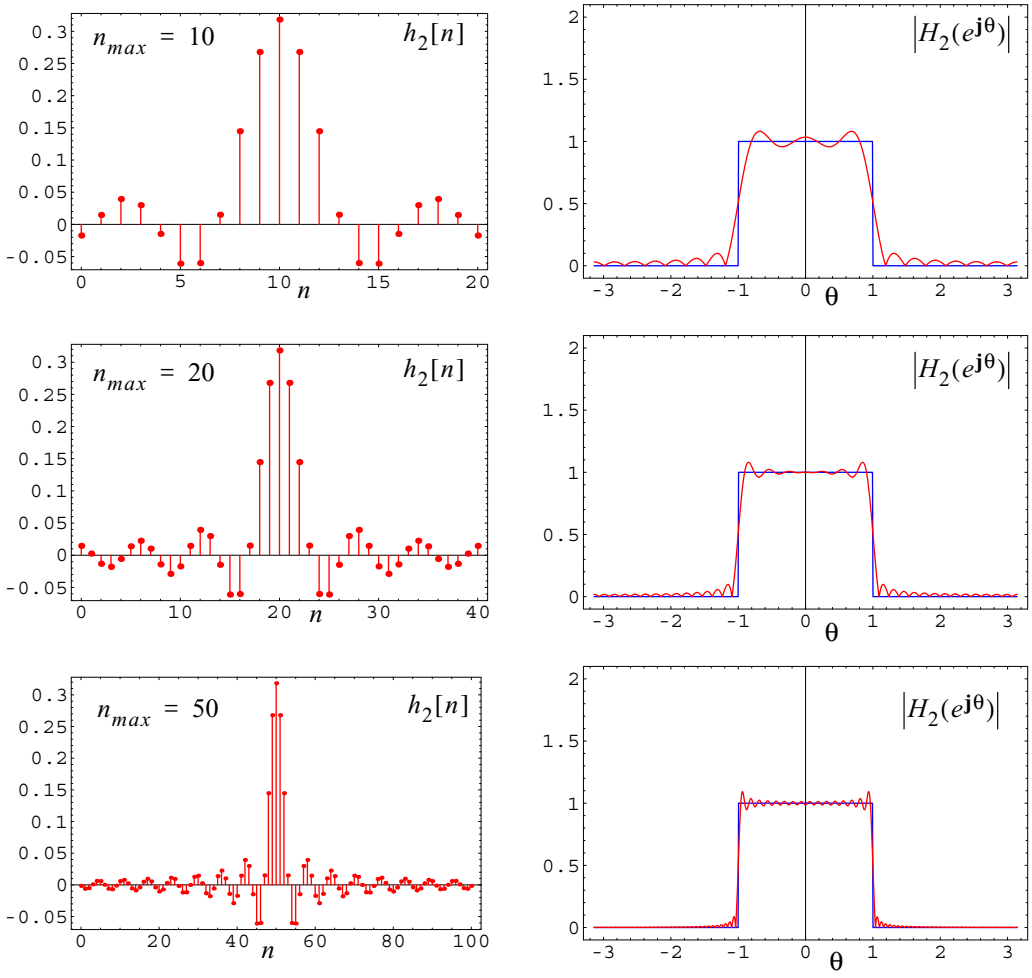

Figure 11

5. Design of a low-pass FIR filter: approach #2

A. Introduction

In this section, we will follow exactly the same approach to designing a low-pass FIR filter as in the previous section, except that we will start off with a slightly different magnitude frequency response, as illustrated in Figure 12 below. Note that unlike before, we now allow a finite-width transition region of width Δ from one to zero in the magnitude frequency response.

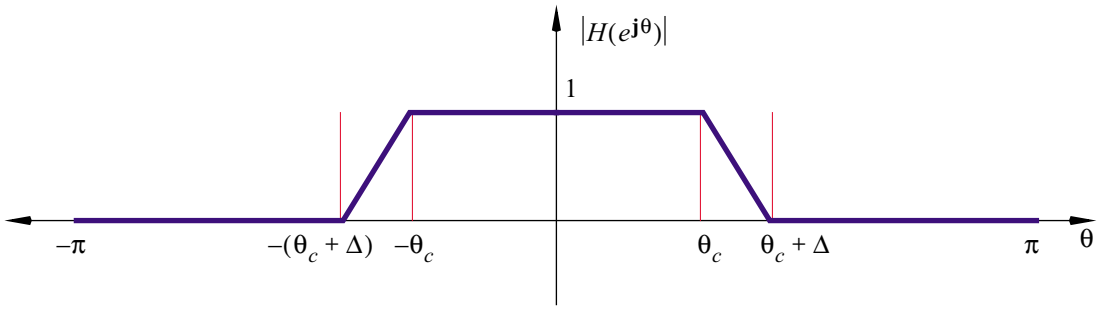

Figure 12

Below, we will use the inverse DTFT to determine the time-domain impulse response $h[n]$ of this filter, assuming no delay. For this filter, the inverse DTFT is given by,

$$h[n] = \frac{1}{2\pi}\int_{-(\theta_c+\Delta)}^{-\theta_c}\left(\frac{1}{\Delta}\cdot(\theta+\theta_c)+1\right)e^{jn\theta}d\theta + \frac{1}{2\pi}\int_{-\theta_c}^{\theta_c}e^{jn\theta}d\theta \\ \frac{1}{2\pi}\int_{\theta_c}^{(\theta_c+\Delta)}\left(\frac{1}{\Delta}\cdot(\theta_c-\theta)+1\right)e^{jn\theta}d\theta \tag{29}$$

Note that,

$$H(e^{j\theta}) = \frac{1}{\Delta}\cdot(\theta+\theta_c)+1\ ,\ \theta\in[-(\theta_c+\Delta),-\theta_c] \tag{30}$$

$$H(e^{j\theta}) = \frac{1}{\Delta}\cdot(\theta_c-\theta)+1\ ,\ \theta\in[\theta_c,(\theta_c+\Delta)] \tag{31}$$

are the equations of the two lines in the transition regions. We solve the integration in equation (29) using *Mathematica* (see "fir_filter_design_part1b.nb") and arrive at the following solution for $h[n]$:

$$h[n] = \frac{\cos(n\theta_c)-\cos(n(\theta_c+\Delta))}{n^2\pi\Delta}\ ,-\infty<n<\infty\ . \tag{32}$$

Using L'Hopital's rule, we can show that this result is, in fact, identical to the one previously derived for the ideal filter [equation (23)], as $\Delta\to 0$:

$$\lim_{\Delta\to 0}h[n] = \lim_{\Delta\to 0}\left[\frac{\cos(n\theta_c)-\cos(n(\theta_c+\Delta))}{n^2\pi\Delta}\right] \tag{33}$$

$$\lim_{\Delta\to 0}h[n] = \lim_{\Delta\to 0}\left[\frac{\frac{d}{d\Delta}\{\cos(n\theta_c)-\cos(n(\theta_c+\Delta))\}}{\frac{d}{d\Delta}\{n^2\pi\Delta\}}\right] \tag{34}$$

$$\lim_{\Delta\to 0}h[n] = \lim_{\Delta\to 0}\left[\frac{n\sin(n(\theta_c+\Delta))}{n^2\pi}\right] = \frac{\sin(n\theta_c)}{n\pi}\ . \tag{35}$$

As for the case of the ideal filter, the impulse response $h[n]$ in equation (32) is infinite in length, both forwards and backwards in time. For example, in Figure 13, we plot $h[n]$, $-50 \le n \le 50$, for $\theta_c = 1$ and $\Delta = 1/10$ .However, when comparing Figures 10 and 13, we notice that the impulse response for non-zero transition regions (i.e. $\Delta \neq 0$, Figure 13), decays to zero values much quicker than for $\Delta = 0$ (Figure 10) as $n \to \infty$. This is due to the fact that the $h[n]$ values vary as $1/n$ for the ideal filter in the previous section, while they vary as $1/n^2$ for the filter with nonzero-width transition regions in this section. Therefore, the effects of approximating the infinite impulse response in Figure 13 by retaining only values of $h[n]$ for $-n_{max} \le n \le n_{max}$ should be less severe.

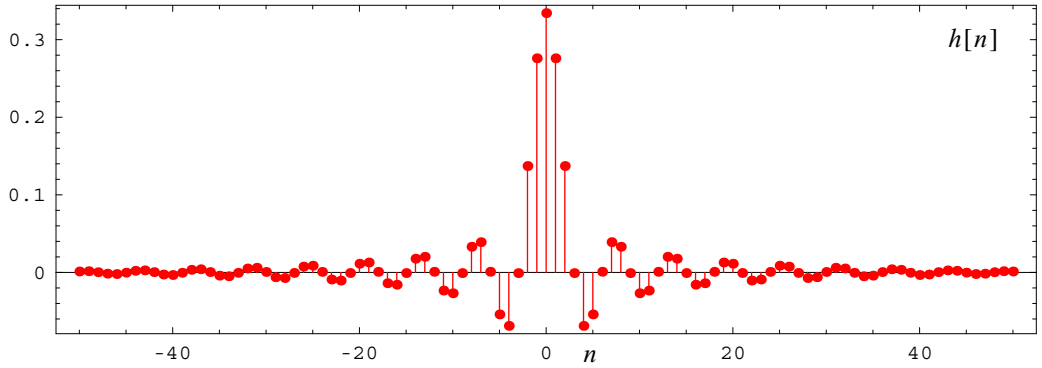

Figure 13

B. Approximation with an FIR filter

Here we will approximate the infinite impulse response in Figure 13, corresponding to the frequency response in Figure 12 (with no delay), with a causal FIR filter, following the same procedure as for the ideal filter in the previous section. First we will retain values for $h[n]$ [see equation (32)] only for a limited range of n,

$$-n_{max} \le n \le n_{max} \tag{36}$$

since values of $h[n]$ approach zero as $|n| \to \infty$. Let us denote this finite impulse response as $h_1[n]$, such that,

$$h_1[n] = \begin{cases} h[n] & -n_{max} \le n \le n_{max} \\ 0 & elsewhere \end{cases} \tag{37}$$

Second, we will shift $h_1[n]$ to be causal; let us denote the resulting impulse response as $h_2[n]$ such that,

$$h_2[n] = h_1[n - n_{max}]. \tag{38}$$

Note that this shift will not change the steady-state frequency response of the resulting system, except to introduce a delay at the output. Thus, the difference equation corresponding to $h_2[n]$ can be written as:

$$y[n] = \sum_{k=0}^{2n_{max}} b_k x[n-k] \tag{39}$$

where,

$$b_k = h[-n_{max} + k] = \frac{\cos((-n_{max}+k)\theta_c) - \cos((-n_{max}+k)(\theta_c + \Delta))}{(-n_{max}+k)^2 \pi \Delta} \tag{40}$$

In Figure 14, we plot the impulse response $h_2[n]$ and the corresponding magnitude frequency response $\left|H_2(e^{j\theta})\right|$ for various values of n_{max} (10, 20 and 50) and numeric values, $\theta_c = 1$ and $\Delta = 1/10$. For each

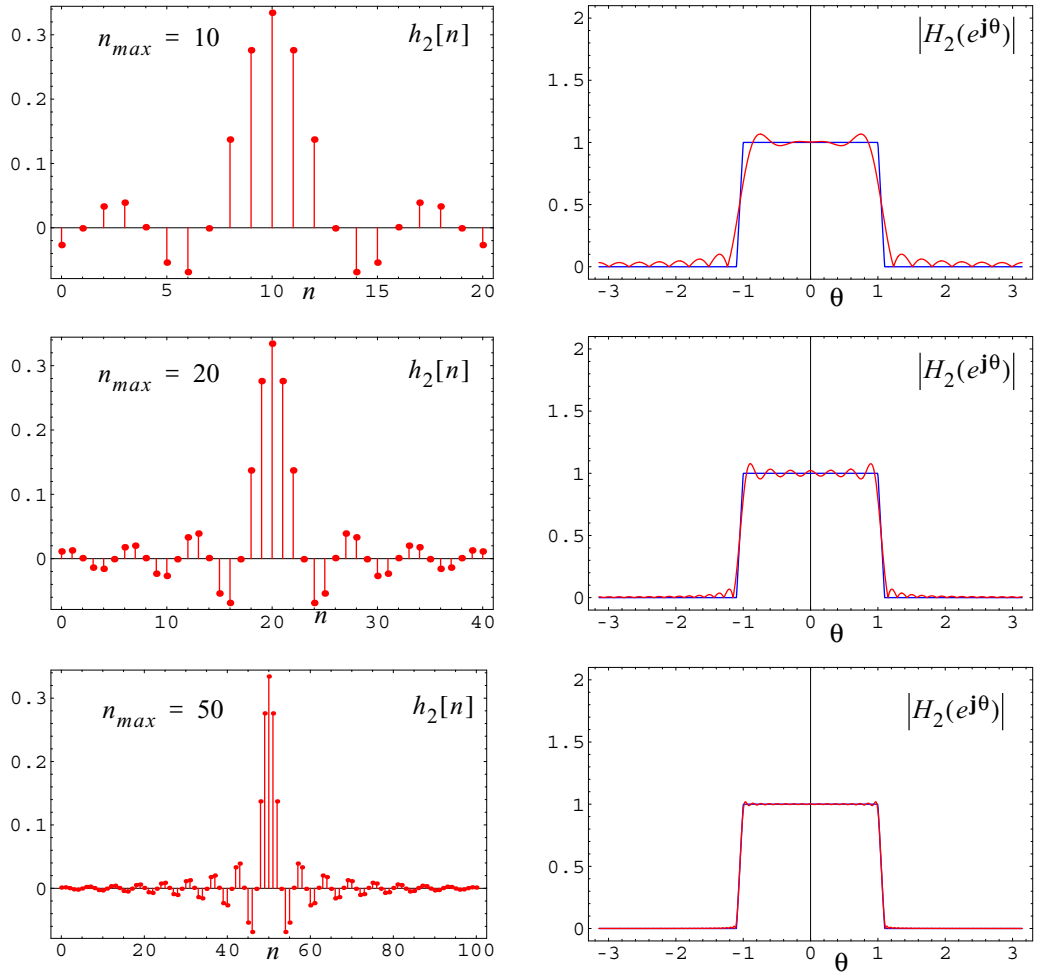

Figure 14

frequency response plot, the blue line indicates the infinite impulse magnitude frequency response.Note that unlike the previous examples in Figure 11, we no longer have the same overshoot problem. In fact, for $n_{max} = 50$, we see that the FIR filter is a very good approximation of the IIR filter. Again, this is due to the fact that we no longer require the transition from one to zero in the frequency response to occur instantaneously, but rather that we allow for that transition region to be of some small, non-zero width (i.e. Δ).

C. Filtering examples

In this section, we illustrate the output $y[n]$ for an FIR filter of type (40) on some sample input signals $x[n]$. For these examples, we assume a sampling rate of $f_s = 500$ Hz, a desired cutoff frequency $f_c = 100$ Hz, and a transition width of $\Delta_f = 10$ Hz. We also, limit our FIR filter to a length $(2n_{max} + 1)$, where $n_{max} = 50$. Using the conversion rule between θ and f,

$$\theta = \frac{2\pi f}{f_s} \tag{41}$$

we can compute our normalized cutoff frequency θ_c and transition-region width Δ:

$$\theta_c = 2\pi(f_c/f_s) = 2\pi(100/500) = 2\pi/5 \tag{42}$$

$$\Delta = 2\pi(\Delta_f/f_s) = 2\pi(10/500) = \pi/25\,. \tag{43}$$

Using equations (32), (37) and (38), we can now compute the impulse response $h_2[n]$ for the values of θ_c and Δ in (42) and (43) above. In Figure 15 below, we plot that impulse response and the corresponding magnitude frequency response as a function of frequency f. Note how well the FIR frequency response approximates our desired cutoff frequency and transition width.

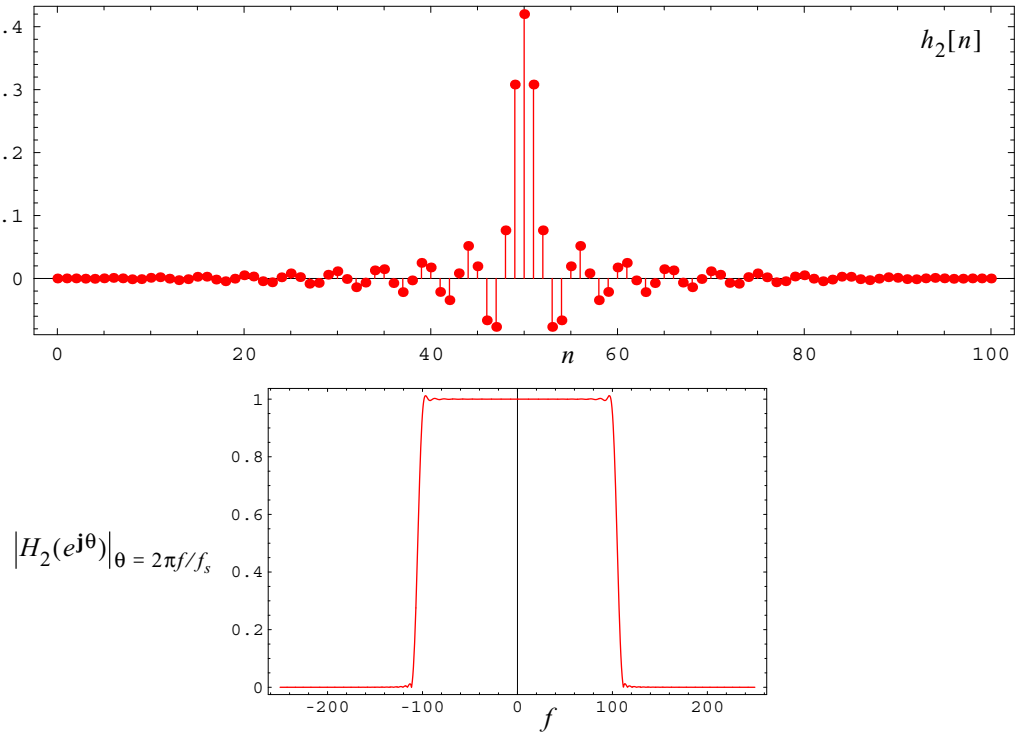

Figure 15

We now apply the following four test inputs to this system:

$$x_1[n] = \cos\left(\frac{2\pi \cdot 50n}{f_s}\right) = \cos(0.2\pi n) \text{ (i.e. sampled 50Hz cosine wave)} \tag{44}$$

$$x_2[n] = \cos\left(\frac{2\pi \cdot 50n}{f_s}\right) + 3\cos\left(\frac{2\pi \cdot 75n}{f_s}\right) = \cos(0.2\pi n) + 3\cos(0.3\pi n) \tag{45}$$

$$x_3[n] = \cos\left(\frac{2\pi \cdot 50n}{f_s}\right) + 3\cos\left(\frac{2\pi \cdot 150n}{f_s}\right) = \cos(0.2\pi n) + 3\cos(0.6\pi n) \tag{46}$$

$$x_4[n] = \cos\left(\frac{2\pi \cdot 50n}{f_s}\right) + 3\cos\left(\frac{2\pi \cdot 105n}{f_s}\right) = \cos(0.2\pi n) + 3\cos(0.42\pi n) \tag{47}$$

For each test input, we compute the output as the convolution between $x_i[n]$ and $h_2[n]$:

$$y_i[n] = x_i[n] * h_2[n],\ i \in \{1, 2, 3, 4\}, \tag{48}$$

and compute the magnitude FFT of 500-length segments of the steady-state input and corresponding output (see *Mathematica* notebook "fir_filter_design_part1b.nb" for time-domain plots with transient responses as well).These FFTs are plotted in Figures 16 through 19 below.

Note that for the first three test signals, the filter leaves the magnitude of the input frequencies less than 100Hz unchanged, while zeroing out frequencies above 110Hz. For the fourth signal, we see that the 105Hz frequency component is approximately halved, since it lies exactly halfway inside the transition region from 100Hz to 110Hz. Thus, this low-pass FIR filter appears to perform exactly as designed.

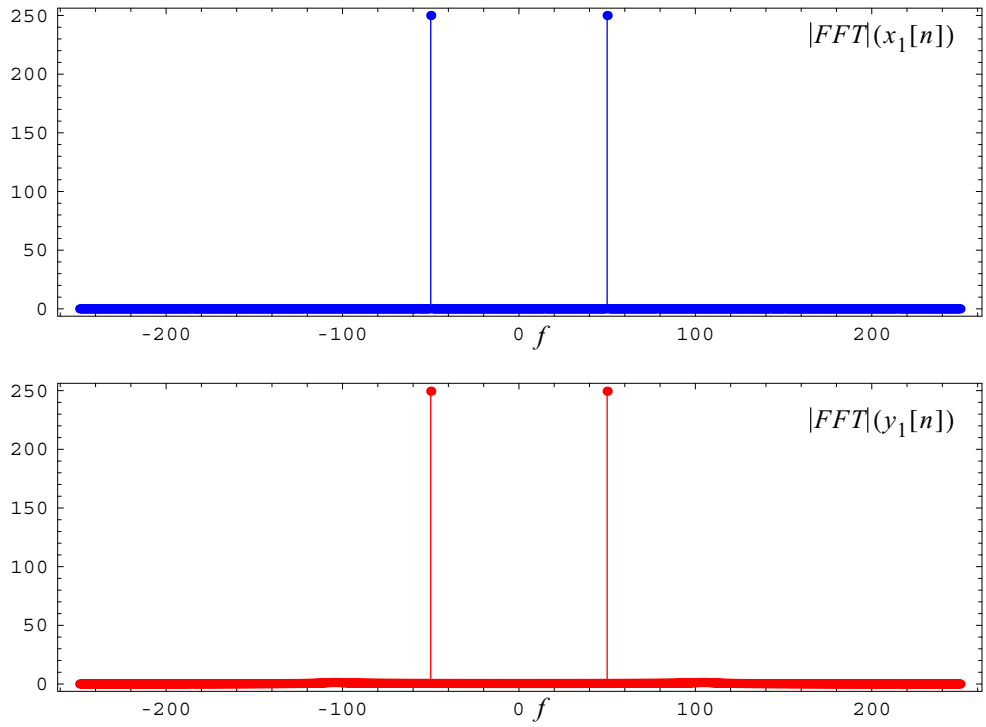

Figure 16

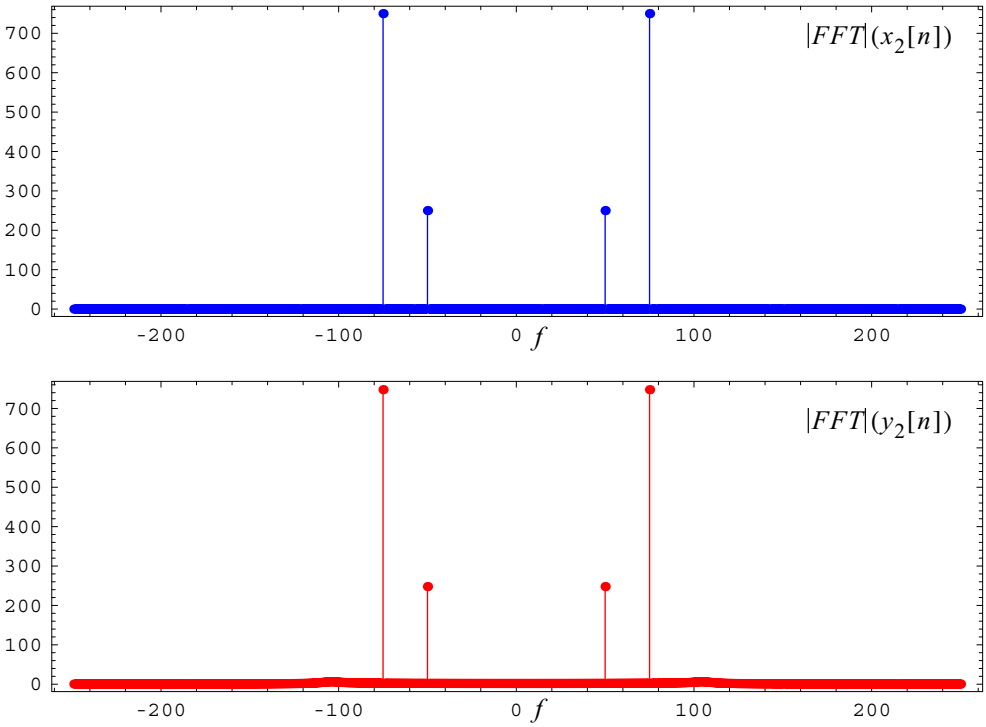

Figure 17

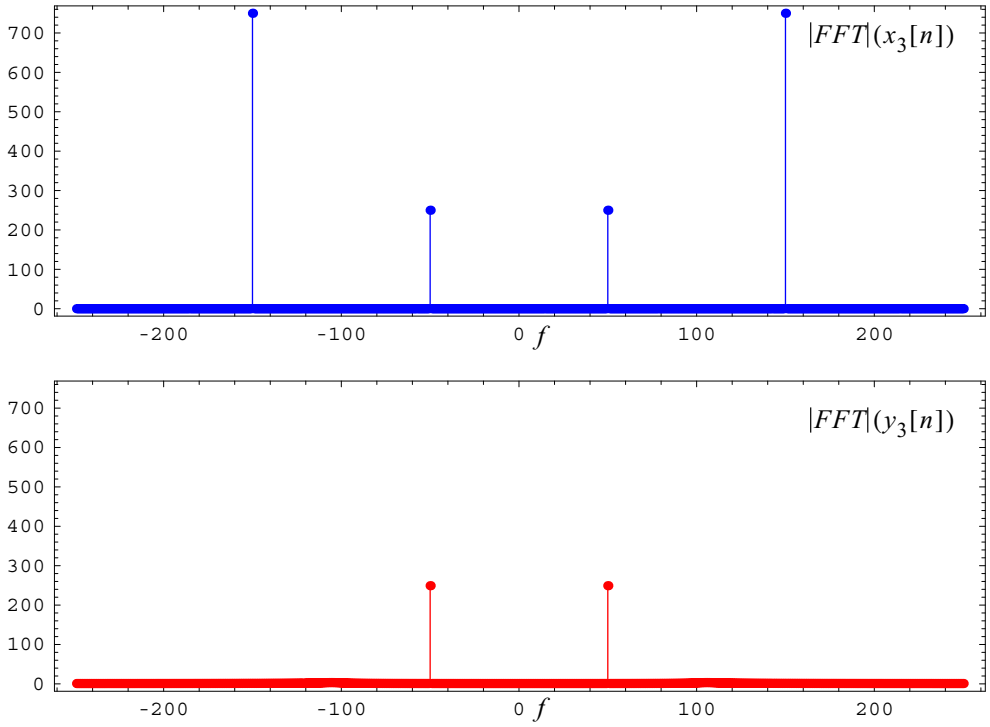

Figure 18

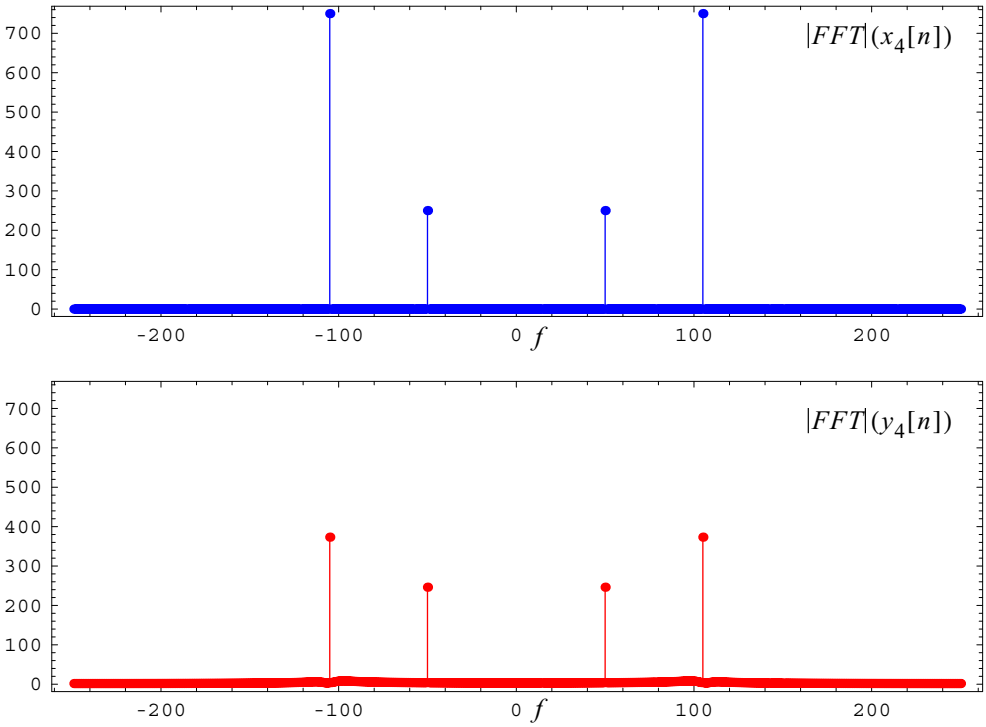

Figure 19

6. Conclusion

The *Mathematica* notebooks “fir_filter_design_part1a.nb” and “fir_filter_design_part1b.nb” were used to generate all the examples in this set of notes. In a companion set of notes, we conclude our discussion on FIR-filter design.