

Lecture 9

FIR filters 1

Introduction

In our study of the z -transform we saw that a rational transfer function

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_M z^{-M}}{1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_N z^{-N}} \quad (1)$$

can be realized in the time-domain by the difference equation

$$y(n) = \sum_{k=0}^M b_k x(n-k) - \sum_{k=1}^N a_k y(n-k) \quad (2)$$

For $N=0$, the $H(z)$ denominator is simply 1, and

$$H(z) = b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_M z^{-M} \quad (3)$$

It follows that

$$y(n) = \sum_{k=0}^M b_k x(n-k) \quad (4)$$

This has the form of a convolution with impulse response $h(k) = b_k$, $0 \leq k \leq M$.

Definition

A (causal) finite impulse response (FIR) filter is described by

$$y(n) = \sum_{k=0}^M b_k x(n-k) \quad (5)$$

This has $M+1$ coefficients. The transfer function

$$H(z) = \sum_{k=0}^M b_k z^{-k} \quad (6)$$

is a polynomial of order M in z^{-1} . It has zeros but no poles. For this reason it is also called an *all-zero* filter.

Determining the coefficients

In this lecture we limit consideration to filters with a *desired frequency response* of the form

$$H_d(\omega) = \begin{cases} 1 & \text{passband} \\ 0 & \text{stopband} \end{cases} \quad (7)$$

This includes ideal *lowpass*, *highpass*, *bandpass* and *bandstop* filters (Fig. 1). The corresponding impulse response is the inverse Fourier transform

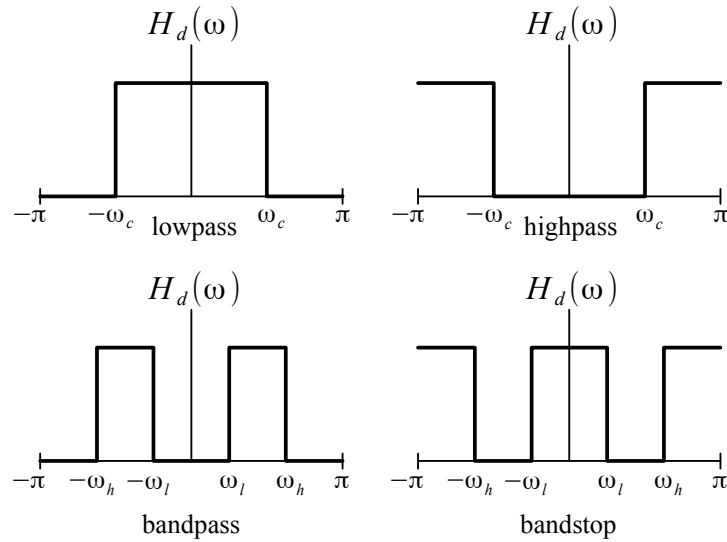


Fig. 1: Four ideal filter frequency responses.

$$h_d(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_d(\omega) e^{j\omega n} d\omega \quad (8)$$

We also limit consideration to filters with real impulse response.

$$H_d(-\omega) = H_d(\omega) \quad (9)$$

The frequency response is an even function. Another way to express this is $H_d(\omega) = H_d(|\omega|)$.

With this condition we have

$$H_d(-\omega)e^{-j\omega n} + H_d(\omega)e^{j\omega n} = H_d(\omega)(e^{j\omega n} + e^{-j\omega n}) = 2H_d(\omega)\cos(\omega n) \quad (10)$$

so we can write

$$h_d(n) = \frac{1}{\pi} \int_0^{\pi} H_d(\omega)\cos(\omega n) d\omega \quad (11)$$

In this form $h_d(n)$ is manifestly real if $H_d(\omega)$ is. Also, note that the impulse response is even in n , $h_d(-n) = h_d(n)$, because $\cos(\omega n)$ is.

A useful integral

If $H_d(\omega)$ is 1 in the interval $\omega_1 \leq \omega \leq \omega_2$ then (11) will include the integral

$$\begin{aligned} I &= \frac{1}{\pi} \int_{\omega_1}^{\omega_2} \cos(\omega n) d\omega \\ &= \frac{1}{\pi n} \sin(\omega n) \Big|_{\omega_1}^{\omega_2} \\ &= \frac{\sin(\omega_2 n) - \sin(\omega_1 n)}{\pi n} \end{aligned} \quad (12)$$

where we have assumed $n \neq 0$. For the special case $n = 0$ we have

$$I = \frac{1}{\pi} \int_{\omega_1}^{\omega_2} d\omega = \frac{\omega_2 - \omega_1}{\pi} \quad (13)$$

Lowpass filter

The frequency response of an ideal lowpass filter is

$$H_d(\omega) = \begin{cases} 1 & 0 \leq |\omega| \leq \omega_c \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

where $\omega_c = 2\pi f_c$ is the cutoff frequency. To make use of (12) and (13) we set

$$\omega_1 = 0, \omega_2 = \omega_c \quad (15)$$

with the result

$$h_d(n) = \frac{\sin(\omega_c n)}{\pi n}, \quad h_d(0) = \frac{\omega_c}{\pi} \quad (16)$$

As expected, this impulse response is even, $h_d(-n) = h_d(n)$. Note that

$$\frac{\sin(\omega_c n)}{\pi n} = 2f_c \frac{\sin(2\pi f_c n)}{2\pi f_c n} = 2f_c \operatorname{sinc}(2f_c n) \quad (17)$$

and $\omega_c/\pi = 2f_c$. Therefore, we can write

$$h_d(n) = 2f_c \operatorname{sinc}(2f_c n) \quad (18)$$

Highpass filter

The frequency response of an ideal highpass filter is

$$H_d(\omega) = \begin{cases} 1 & \omega_c \leq |\omega| \leq \pi \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

where, again, $\omega_c = 2\pi f_c$ is the cutoff frequency. In (12) and (13) we set

$$\omega_1 = \omega_c, \omega_2 = \pi \quad (20)$$

with result

$$h_d(n) = \frac{-\sin(\omega_c n)}{\pi n}, \quad h_d(0) = \frac{\pi - \omega_c}{\pi} = 1 - \frac{\omega_c}{\pi} \quad (21)$$

This impulse response is also even, $h_d(-n) = h_d(n)$.

For a given ω_c call the lowpass and highpass impulse responses respectively $h_d^{LP}(n)$ and $h_d^{HP}(n)$. From (16) and (21) we see that

$$h_d^{LP}(n) + h_d^{HP}(n) = \delta(n) \quad (22)$$

This *perfect reconstruction* property of FIR filters can be very useful. We will come back to this in a future lecture.

Causal FIR filters

The ideal lowpass and highpass filter impulse responses we have derived are both infinite and non-causal. Therefore they cannot be realized in practice. We will convert an ideal impulse response into a practical one by a two-step process (Fig. 2).

We first make the impulse response finite by truncating the desired impulse response as follows

$$h_{trunc}(n) = \begin{cases} h_d(n) & |n| \leq M/2 \\ 0 & |n| > M/2 \end{cases} \quad (23)$$

where M is a positive even integer. We now have an FIR filter, but it is still non-causal. Truncating the impulse response obviously changes the impulse response which in turn changes the frequency response. So, the price we pay for truncation is that the actual frequency response $H(\omega)$ will differ from the desired response $H_d(\omega)$.

Next we make the filter causal by shifting the impulse response so that the first non-zero sample occurs at $n=0$ instead of at $n=-M/2$

$$h(n) = h_{trunc}(n - M/2) \quad (24)$$

The final result is

$$h(n) = \begin{cases} h_d(n - M/2) & 0 \leq n \leq M \\ 0 & \text{otherwise} \end{cases} \quad (25)$$

Shifting the impulse response simply causes a corresponding shift in the output of the filter

$$h(n) * x(n) = y(n) \rightarrow h(n - n_0) * x(n) = y(n - n_0) \quad (26)$$

The output is delayed relative to the non-causal filter.

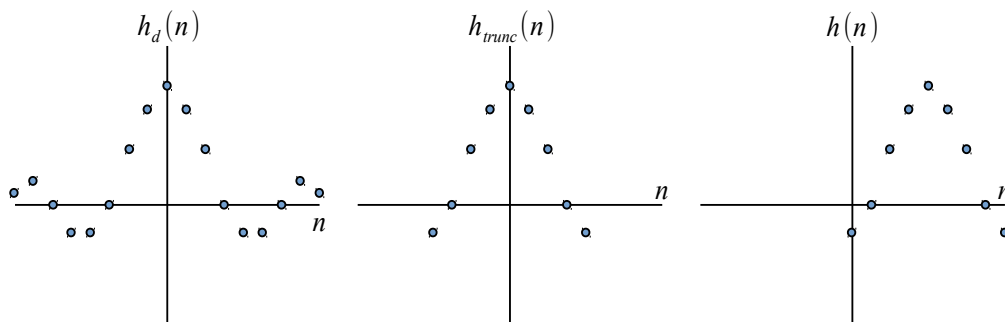


Fig. 2: (Left) the desired impulse response is generally infinite and non-causal. (Center) truncating makes it finite. (Right) shifting makes it causal.

Scilab code to generate lowpass and highpass FIR impulse responses and calculate the frequency response is given in the Appendix.

As we noted above, truncation of the impulse response means the frequency response will differ from the desired $H_d(\omega)$. Let's look at some examples of this effect.

Example 1: Determine the coefficients of a lowpass FIR filter with cutoff frequency $f_c = 0.1$ and $M = 30$.

We have $\omega_c = 2\pi(0.1)$. The $n=0$ desired impulse response value is

$$h_d(0) = \frac{2\pi(0.1)}{\pi} = 0.2$$

The $n \neq 0$ values are

$$h_d(n) = \frac{\sin(0.2\pi n)}{\pi n}$$

The finite, causal approximation is

$$h(n) = \begin{cases} h_d(n-20) & 0 \leq n \leq 40 \\ 0 & \text{otherwise} \end{cases}$$

This, and the corresponding frequency response magnitude

$$|H(\omega)| = \left| \sum_{k=0}^{40} h(n) e^{-j\omega n} \right|$$

with $\omega = 2\pi f$, are shown in Fig. 3.

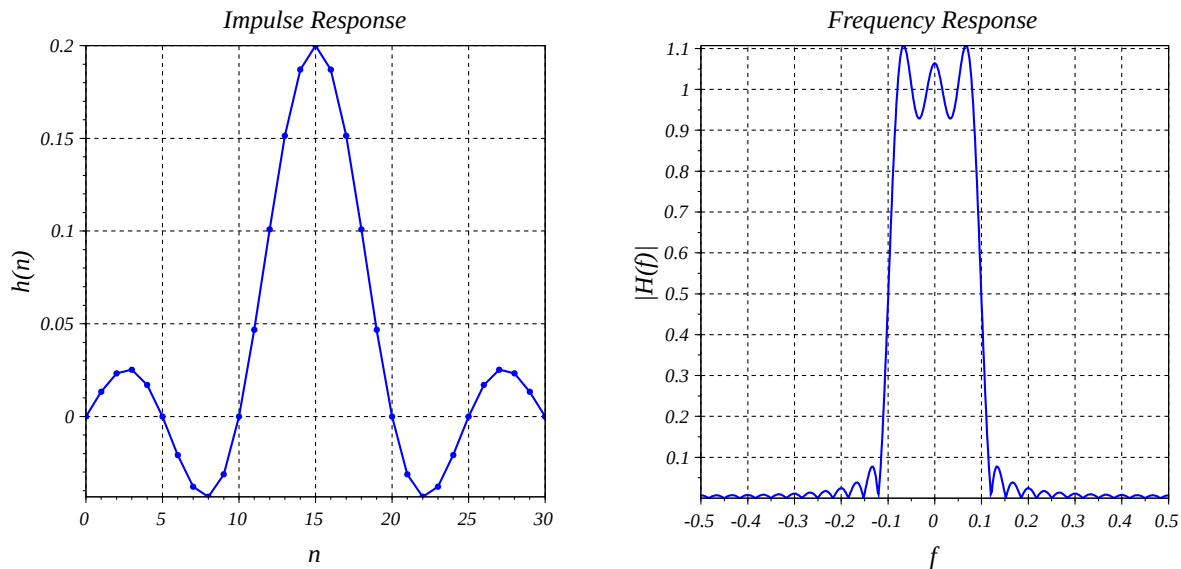


Fig. 3: Lowpass filter with $f_c=0.1$ and $M=30$.

Let's try a highpass filter.

Example 2: Determine the coefficients of a highpass FIR filter with cutoff frequency $f_c=0.1$ and $M=30$.

As in the previous example $\omega_c = 2\pi(0.1)$. We have

$$h_d(0) = 1 - \frac{2\pi(0.1)}{\pi} = 0.8$$

$$h_d(n) = -\frac{\sin(0.2\pi n)}{\pi n}$$

$$h(n) = \begin{cases} h_d(n-20) & 0 \leq n \leq 40 \\ 0 & \text{otherwise} \end{cases}$$

This, and the corresponding frequency response magnitude are shown in Fig. 4.

The frequency responses shown in Fig. 3 and Fig. 4 are clearly not ideal. Two things stand out. First, the response has *ripples* in both the passband and stopband. Second, the transition between passband and stopband is not instantaneous. Instead the response magnitude is 0.5 at the cutoff frequency and has a finite slope.

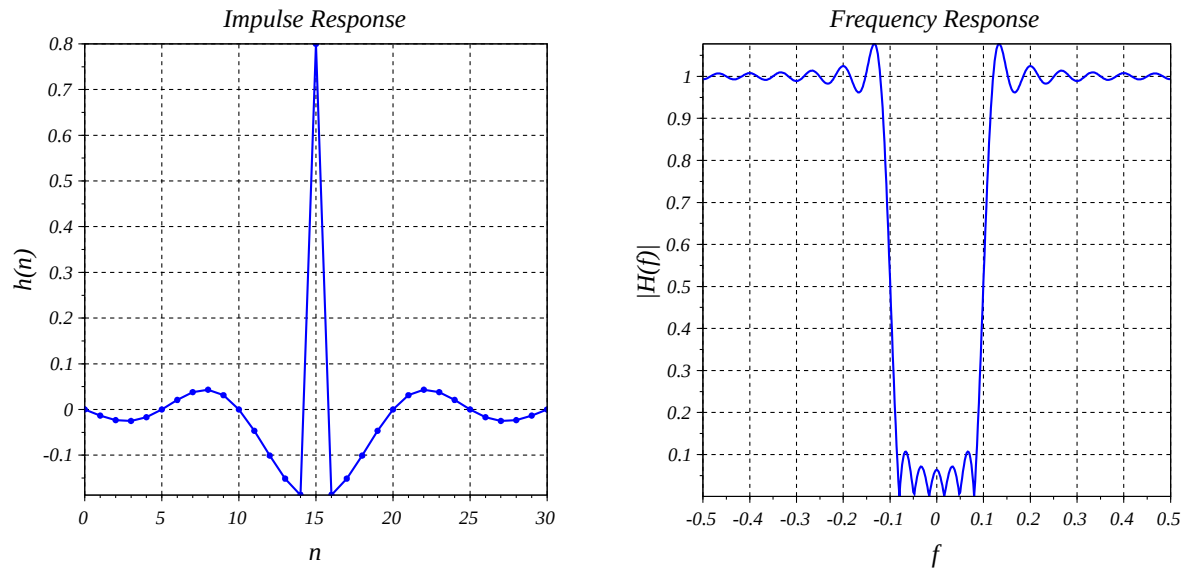


Fig. 4: Highpass filter with $f_c=0.1$ and $M=30$.

These two effects arise from truncating the ideal impulse response. Let's see what happens if we keep more impulse response values. Fig. 5 shows the lowpass filter response with M increased to 60. The transition between passband and stopband is more steeply sloped. Apparently we can make the transition between passband and stopband arbitrarily rapid by proper selection of the FIR filter length. However, the ripples have not decreased in amplitude. Instead, there are more of them and they are more closely spaced. This is the so-called *Gibbs phenomenon* [1]. We will look at a solution of this problem in our lecture on *windowing*.

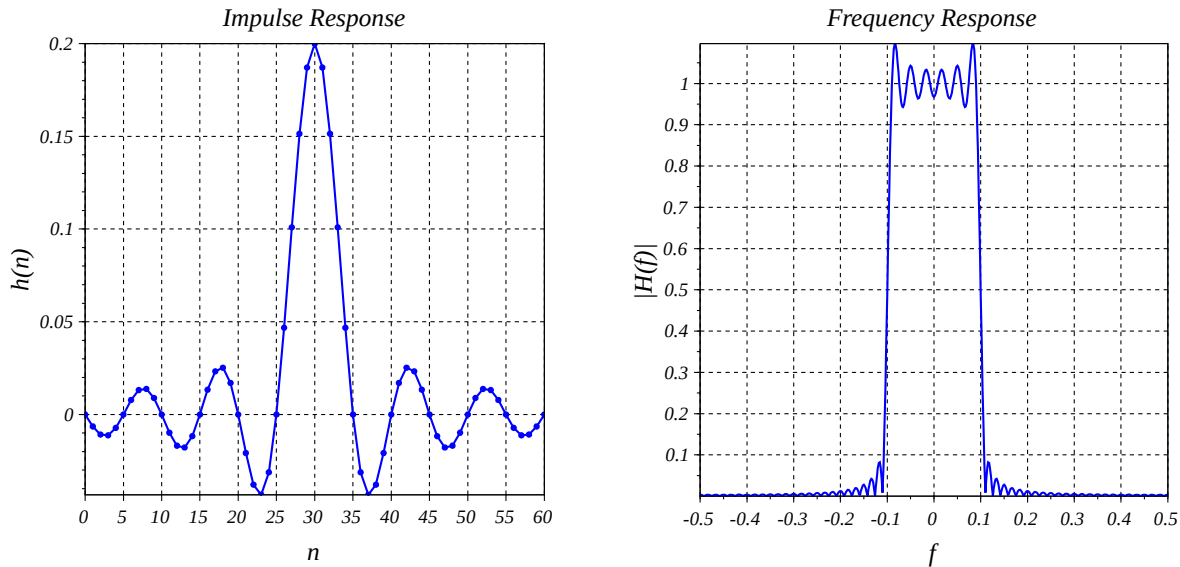


Fig. 5: Lowpass filter with $f_c=0.1$ and $M=60$.

Choosing the filter length

If we truncate the impulse response of a lowpass filter as in (23) the frequency response is

$$H(\omega) = \sum_{n=-M/2}^{M/2} h_d(n) e^{-j\omega n} = h_d(0) + 2 \sum_{n=1}^{M/2} h_d(n) \cos(\omega n) \quad (27)$$

The slope of this at the cutoff frequency, $\omega = \omega_c$, is

$$\left. \frac{d}{d\omega} H(\omega) \right|_{\omega=\omega_c} = -2 \sum_{n=1}^{M/2} h_d(n) n \sin(\omega_c n) \quad (28)$$

Since

$$h_d(n) n \sin(\omega_c n) = \frac{\sin^2(\omega_c n)}{\pi} \quad (29)$$

we have

$$\left. \frac{d}{d\omega} H(\omega) \right|_{\omega=\omega_c} = -\frac{2}{\pi} \sum_{n=1}^{M/2} \sin^2(\omega_c n) \quad (30)$$

If we replace $\sin^2(\omega_c n)$ with its average value

$$\langle \sin^2(\omega_c n) \rangle = \frac{1}{2} \quad (31)$$

then

$$\left. \frac{d}{d\omega} H(\omega) \right|_{\omega=\omega_c} \approx -\frac{2}{\pi} \frac{M}{2} \frac{1}{2} = -\frac{M}{2\pi} \quad (32)$$

For $H(\omega)$ to change by -1 (from 1 to 0) we require a frequency interval of about

$$\Delta \omega \approx \frac{2\pi}{M} \quad (33)$$

Therefore

$$M \approx \frac{1}{\Delta f} \quad (34)$$

where Δf is (roughly) the width of the *transition band* between the passband and stopband. We can use this value of M to come up with an initial design and then “tweak” it by trial and error to meet our design specifications (keeping in mind that M needs to be an even integer).

Example 3: Design a lowpass FIR filter with cutoff frequency $f_c=0.2$ and $\Delta f=0.02$.

We need $M \approx \frac{1}{0.02} = 50$. The frequency response magnitude of the filter generated using

$$h = \text{hLP}(0.2, 25);$$

is shown in Fig. 6.

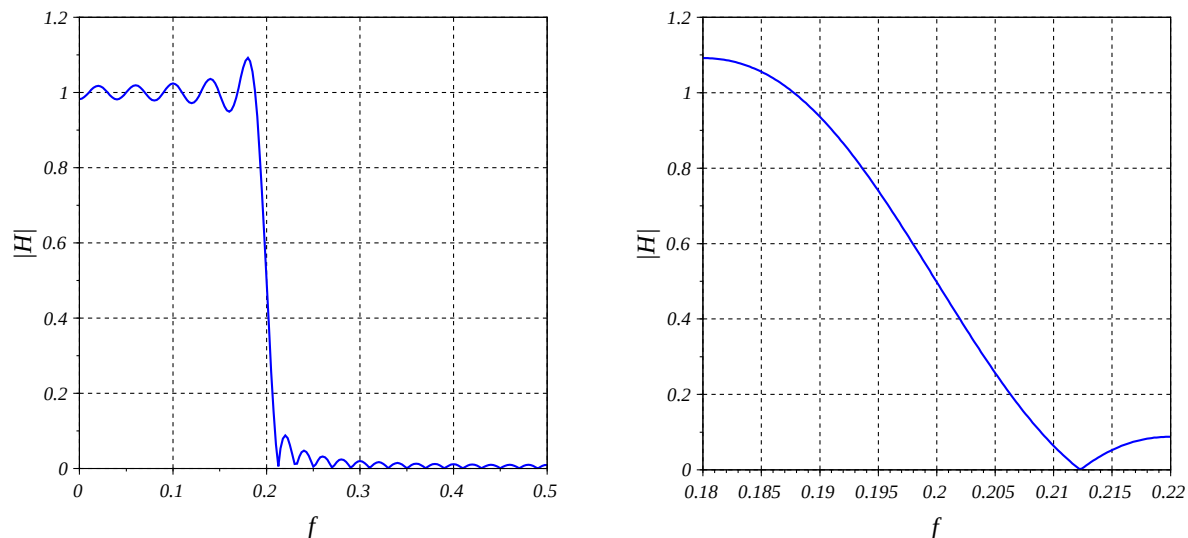


Fig. 6: Frequency response magnitude for Example 3. $|H|$ drops from a bit more than 0.9 to a bit less than 0.1 over the interval $0.19 \leq f \leq 0.21$.

References

1. https://en.wikipedia.org/wiki/Gibbs_phenomenon

Appendix – Scilab code

```
//FIR lowpass impulse response with 2*K+1 samples
//and cutoff frequency fc. (K=M/2)
function h = hLP(fc,K)
    wc = 2*pi*fc;
    h = zeros(1,2*K+1);
    h(K+1) = wc/pi;
    for n=1:K
        h(K+1+n) = sin(wc*n)/(pi*n);
        h(K+1-n) = h(K+1+n);
    end
endfunction

//FIR highpass impulse response with 2*K+1 samples
//and cutoff frequency fc. (K=M/2)
function h = hHP(fc,K)
    h = -hLP(fc,K);
    h(K+1) = h(K+1)+1;
endfunction

//calculate freq. resp. mag of transfer function H(z)
//H(z) = ( b(1)+b(2)*z^(-1)+b(3)*z^(-2)+... )
// / (a(1)+a(2)*z^(-1)+a(3)*z^(-2)+... )
//at Np points over interval f1<=f<=f2
function [f,Hm] = freqRespMag(b,a,f1,f2,Np)
    M = length(b)-1;
    N = length(a)-1;
    f = linspace(f1,f2,Np);
    Hm = zeros(f);
    for k=1:Np
        zi = exp(-i*2*pi*f(k));
        num = b(1);
        for n=1:M
            num = num+b(n+1)*zi^n;
        end
        den = a(1);
        for n=1:N
            den = den+a(n+1)*zi^n;
        end
        H = num/den;
        Hm(k) = abs(H);
    end
endfunction
```