Tutorial 5: Discrete Filter Design

1. Filter Specifications

There are two complementary ways to specify a filter: in the frequency and in the time domains. The frequency domain specification of the desired filter's magnitude and phase is probably more common for IIR filters, while the time domain specification in terms of the impulse response of the filter is more common in FIR filter design.

The magnitude specifications consist in determining frequency ranges where the attenuation or gain has specific values. Consider the case of a low-pass filter. First of all, since the filter is discrete and it is to have real coefficients, the specifications need to be made only for frequencies $[0, \pi]$ due to the even and odd characteristics of the magnitude and the phase functions, and to their periodicity. In the low-pass filter we would like the magnitude $|H_d(e^{jw})|$ to be close to unity in a passband frequency region, and close to zero in the stopband frequency region. Since, it is not possible to design a filter that changes from unity to zero, a transition frequency band is provided. Thus, the magnitudespecifications are displayed in Fig. 1. The passband is the set of frequencies for which the attenuation specification is the smallest; the stopband is the set of frequencies where the attenuation specification is the greatest.

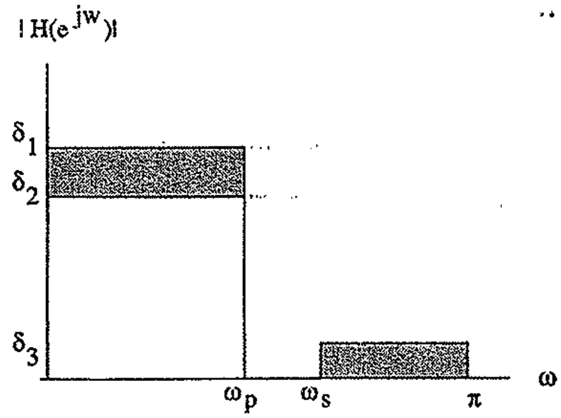

Figure 1: Low-pass Filter Magnitude Specifications

For reasons that will be made clear shortly, and especially when the filters are for acoustic applications, the phase is specified to be linear. This condition can be satisfied in the design of FIR filters, but it is not possible in IIR design. In IIR filter design, the condition of linear phase is downgraded to require that the phase be close to linear in the passband (in the stopband the magnitude is small so the phase in that band is not important).

1.1. Gain and Loss Functions

Since the magnitude of a filter can be expressed in either the normal or the logarithmic scale, the magnitude filter specifications can be given in two forms. As an example,

65

consider again the low-pass filter. The graphic specifications given in Fig. 1 can be expressed as

$$\delta_2 \leq |H(e^{jw})| \leq \delta_1 \quad 0 \leq w \leq w_p \tag{1}$$

$$0 \leq |H(e^{jw})| \leq \delta_3 \quad w_s \leq w \leq \pi \tag{2}$$

The gain function is defined as

$$A(w) = 20 \log |H(e^{jw})| \text{ (dB)}$$

and the loss or attenuation function is given by

$$\alpha(w) = -A(w) = -20 \log |H(e^{jw})| \text{ (dB)}$$

Although $|H(e^{jw})|, A(w)$, and $\alpha(w)$ provide equivalent specifications, the gain and the loss functions are more commonly used.

Example. Consider the following specifications for a low-pass filter:

$$0.9 \leq |H(e^{jw})| \leq 1.0 \quad 0 \leq w \leq \pi/2$$
$$0 \leq |H(e^{jw})| \leq 0.1 \quad 3\pi/2 \leq w \leq \pi$$

If we express these specifications in dB, we get

$$-0.92 \leq A(w) \leq 0 \quad (dB) \quad 0 \leq w \leq \pi/2..$$
$$-\infty \leq A(w) \leq -20 \quad (dB) \quad 3\pi/2 \leq w \leq \pi$$

The loss specifications are then

$$0.92 \geq \alpha(w) \geq 0 \quad (dB) \quad 0 \leq w \leq \pi/2$$
$$\infty \geq \alpha(w) \geq 20 \quad (dB) \quad 3\pi/2 \leq w \leq \pi$$

1.2. Frequency Scales

There are different equivalent ways in which the frequency of a discrete filter can be expressed. In many of the analog/discrete applications the sampling frequency (f_s in Hz or Ω_s in rad/sec) is known, and given that $H(e^{jw})$ is periodic and the magnitude is an even function of frequency and the phase is an odd function, then the specifications can be equivalently made in different scales.

One scale is the f (Hz) scale from 0 to $f_s/2$, the fold-over or Nyquist frequency that comes from the sampling theory; the scale Ω (rad/sec) is the previous scale multiplied by 2π so that the specifications are made from 0 to $\Omega_s/2$. The discrete frequency scale w (rad) can be obtained by multiplying the Ω scale by T_s, the sampling period, and the specifications are made from 0 to π.

Other scales are possible, but less used than the ones given. One of these consist in dividing by the sampling frequency either in Hz or in radians/sec: the f/f_s (no units) scale goes from 0 to 1/2, and so does the Ω/Ω_s (no units) scale. It is clear that when the specifications are given in any scale, it can be easily transformed into any other desired scale.

66

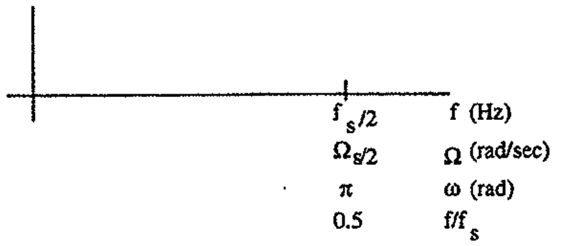

Figure 2: Discrete Filter Frequency Scales

1.3. Linear Phase

In acoustic filtering the signal needs to be amplified in a band of frequencies without distortions in the phase. Phase distortions cause changes in the acoustic signal that are captured by the hearing system. In general, when we are processing a signal and we wish to filter out certain components we would like the phase to be linear in the passband.

Let's consider a simple example to illustrate the effect of phase. The simplest communication system would be the one that takes the signal at the input and puts the same signal at the output. To allow for possible delays in the transmission, the output can be the signal at the input delayed. This would be accomplished by an all-pass filter with linear phase (See Fig 3), e.g., an all-pass filter with a transfer function $H(z) = z^{-1}$ so that the input $x(n)$ is simply delayed one sample to give the output $x(n-1)$. The frequency response of this filter is

$$H(e^{jw}) = 1e^{-jw}$$

Let's assume the input is

$$x(n) = 1 + \cos(w_o n) + \cos(w_1 n) \quad w_1 > w_o > 0, \quad w_1 = 2w_o$$

Since the filter is all-pass, to calculate the steady state response of the filter we only need to consider the effect of the phase in the input signal, so that

$$\begin{aligned} y_{ss}(n) &= 1\cos(0n + 0) + 1\cos(w_o n - w_o) + 1\cos(w_1 n - w_1) \\ &= 1 + \cos(w_o(n-1)) + 1\cos(w_1(n-1)) \\ &= x(n-1) \end{aligned}$$

Thus, in this case where the phase is linear for the passband of the all-pass filter (all frequencies), the output is simply the input delayed by one sample. Notice that the delay coincides with the negative of the derivative of the phase with respect to frequency, $-d\theta(w)/dw$, which is called the phase delay and which in this case is unity. One could think of the frequency components as waves that are being transmitted over a medium and the phase change causing that all components travel at the same speed in the case of linear phase. Thus whenever the phase delay is a constant there is no distortion caused by the filter in the frequency components in the passband.

Consider then the case of non-linear phase. Instead of the all-pass filter to have linear phase assume that the phase is given by the second plot in Fig. 3.

67

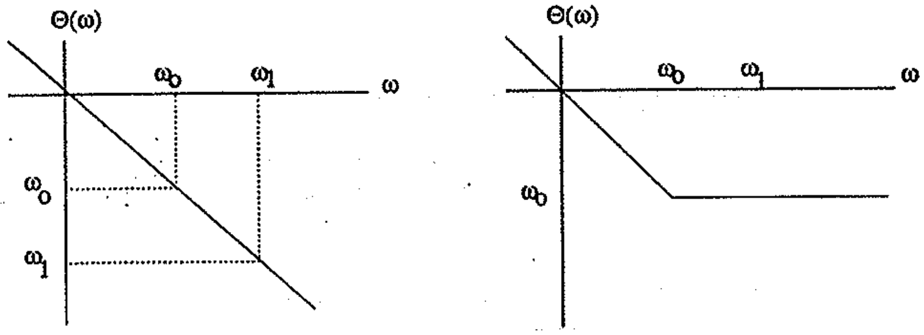

Figure 3: Linear and Non-linear Phases

If we input $x(n)$ given above into this filter, the steady state output is

$$\begin{aligned} y_{ss}(n) &= 1\cos(0n+0) + 1\cos(w_o n - w_o) + 1\cos(w_1 n - 0.5w_1) \\ &= 1 + \cos(w_o(n-1)) + 1\cos(w_1(n-0.5)) \\ &\neq x(n-1) \end{aligned}$$

In this case the phase delay is not uniform (it is 1 for frequencies between 0 and $w_1/2$, and 0 for frequencies in $(w_1/2, \pi]$) causing distortion in the signal: the higher frequency component appears before the low and middle frequency components which would be perceived as distorted by the ear.

When designing FIR filters the linear phase specification can be easily satisfied; however, when designing IIR filters this specification is only – at best – approximated in the passband. The phase of IIR filters is in general nonlinear, although in optimal designs such as Butterworth, Chebyshev and Elliptic is approximately linear in the passband.

2. IIR Filter Design

There are basically two ways to design discrete filters. One uses analog filter designs and by sampling or transformations obtains discrete filters. The other way is to design the filter directly in the discrete domain using approximation and optimizatiom techniques. The former method is commonly used and will be introduced in this section.

Among the different analog filters, the most commonly used are the Butterworth, Chebyshev and elliptic. The main idea in filter design is to design prototype filters (typically low-pass, unit gain and normalized frequency) and then use frequency transformations to obtain the desired filter. The low-pass Butterworth filter, also called maximally-flat, has a flat response at the low frequencies and decays slowly in the transition and stop bands. The low-pass Chebyshev filter displays ripple in the pass-band, but it sharply decays in the transition and stop bands. The ripple in the low-pass elliptic filter occurs both in the pass and the stop bands. In general, low-pass magnitude specifications are satisfied by an elliptic filter of the lowest order, followed by the Chebyshev and then the Butterworth.

The analog low-pass filter has a general form

$$|H(j\Omega)|^2 = \frac{1}{1+f(-\Omega^2)} \tag{3}$$

such that for low frequencies $f(-\Omega^2) \approx 0$ so that $|H(j\Omega)|^2 \approx 1$, and for high frequencies $f(-\Omega^2) \to \infty$ so that $|H(j\Omega)|^2 \to 0$. For instance, the Butterworth magnitude response function is of the form

$$|H_N(j\Omega)|^2 = \frac{1}{1+\left[\frac{\Omega}{\Omega_{HP}}\right]^{2N}}$$

and the Chebyshev magnitude response is given by

$$|H_N(j\Omega)|^2 = \frac{1}{1+\epsilon\, C_N\left(\frac{\Omega}{\Omega_P}\right)^2}$$

where the $C(.)$ are the Chebyshev polynomials. The design problem is then: given the magnitude specifications in the passband and stopband, solve for the parameters of the filters (i.e., (N, Ω_{HP}) for the Butterworth and (ϵ, N) for the Chebyshev). Once these parameters are found, one needs to do a factorization of the denominator and choose the poles on the left-hand s-plane to guarantee the filter stability. The final form of the filter is

$$H(s) = \frac{K}{D(s)}$$

where K is a positive gain, and $N(s)$ has zeros on the left-hand s-plane.

To obtain the discrete filter it is then necessary to use a transformation that maps the left-hand and right-hand s-planes into the inside and outside of the unit circle in the z-plane, respectively. This will preserve the stability of the filters. Also the $j\Omega$ axis needs to be mapped into the unit circle so that the frequency response in the analog plane is mapped into the frequency response in the discrete plane. The bilinear transformation

$$s = K\frac{z-1}{z+1} \tag{4}$$

$$z = \frac{1+s/K}{1-s/K} \tag{5}$$

does the desired mapping. The constant K is a positive gain. To illustrate the mapping consider: the point (A) $s = 0$ on the $j\Omega$-axis is mapped into $z = 1$ on the unit circle; points (B) and (B') $s = \pm j\infty$ are mapped into $z = -1$ which is on the unit circle; a point (C) $s = -1$ is mapped into $z = (1 - 1/K)/(1 + K) < 1$ which is inside the unit circle; finally a point (D) $s = 1$ is mapped into $z = (1 + 1/K)/(1 - K) > 1$ which is located outside the unit circle.

Using the bilinear transformation one can then transform a continuous filter $H(s)$ into a discrete filter

$$H(z) = H(s)\Big|_{s=K\frac{z-1}{z+1}}$$

63

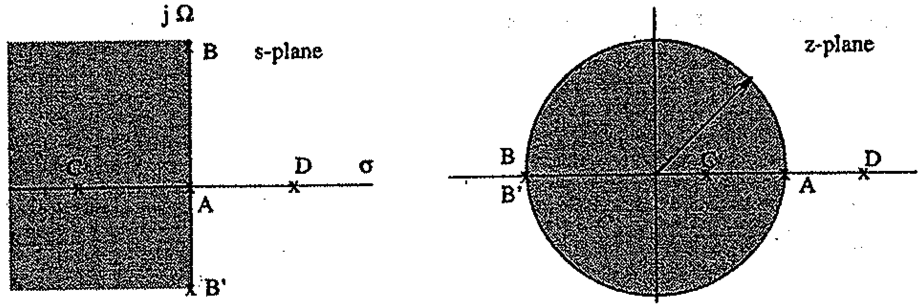

Figure 4: Mapping of s to z Plane by Bilinear Transformation

To understand the discrete filter design based on analog prototypes, we need to look at the relation between the analog frequency Ω (rad/sec) and the discrete frequency w (rad) given by the bilinear transformation. If we let $s = j\Omega$ and $z = e^{jw}$ in the bilinear transformation we have

$$\begin{aligned} j\Omega &= K\frac{e^{jw}-1}{e^{jw}+1} \\ &= K\frac{e^{jw/2}-e^{-jw/2}}{e^{jw/2}+e^{-jw/2}} = \frac{jK\sin(w/2)}{\cos(w/2)} \\ &= jK\tan(w/2) \end{aligned}$$

so that

$$\Omega = K\tan\left(\frac{w}{2}\right) \tag{6}$$

This gives a warped relation between the discrete and the continous frequencies, and so the procedure to design a discrete filter using analog prototypes consists of the following steps:

- Give the desired discrete filter magnitude specifications
- Convert using the above relation the discrete frequencies w into analog frequencies Ω and keep the magnitude specifications
- Design an analog filter prototype that satisfies the specifications obtained above (analog frequency transformations might be needed)
- Convert the analog filter $H(s)$ into a discrete filter $H(z)$ by means of the bilinear transformation

Despite omiting all the details left out in the above steps, one can still see that the procedure is basically to obtain the specifications of an analog filter from the discrete specifications, and after the analog filter is designed the bilinear transformation is used to obtain the discrete filter from the analog filter. The following examples illustrate the procedure.

Example We wish to design a discrete Butterworth filter that satisfies (or exceeds) the following specifications:

- maximum passband attenuation: 3 dB
- passband frequency $f_p = 1$ kHz
- minimum stopband attenuation: 10 dB
- stopband frequency $f_{st} = 2$ kHz
- sample frequency $f_s = 10$ kHz

Notice that in this case since the maximum passband attenuation is 3 dB then the passband frequency is the half-power frequency. The design gives a transfer function

$$H(z) = 0.0676\frac{(z+1)^2}{z^2 - 1.142z + 0.412}$$

which gives the frequency response shown in Fig. 5

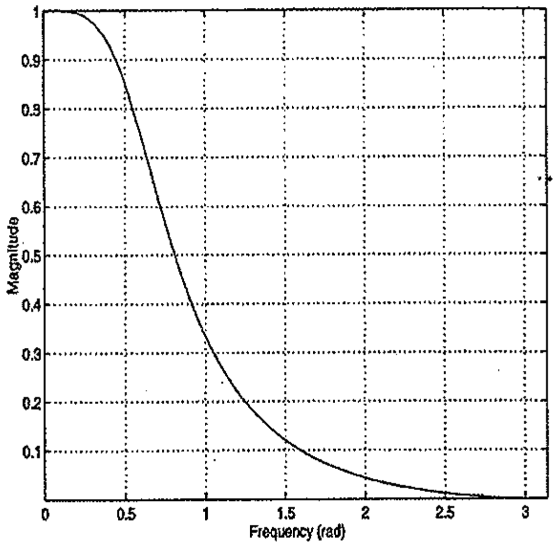

Figure 5: Magnitude of Low-pass Filter

Example A discrete filter based on Butterworth analog filter is to satisfy the following specifications:

- maximum passband attenuation: 1 dB
- passband $f \in [2, 3]$ kHz
- minimum stopband attenuation: 20 dB
- stopband-1: $f \in [0, 1]$ kHz
- stopband-2: $f \in [4, 5]$ kHz
- sampling frequency $f_s = 10$ kHz

71

These specifications are for a band-pass filter. The transfer function of the designed filter is

$$H(z) = 0.0304\frac{z^6 - 3z^4 + 3z^2 - 1}{z^6 + 1.48z^4 + 0.92z^2 + 0.2}$$

which gives the magnitude response shown in Fig. 6

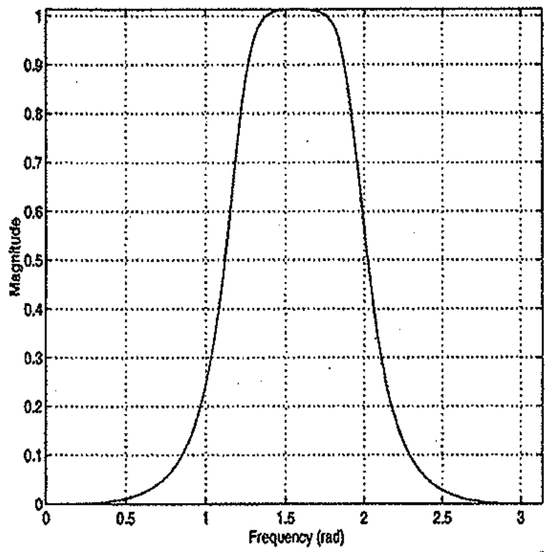

Figure 6: Magnitude of Band-pass Filter

Example In this example we compare three designs of a discrete filter with the following specifications

$$20\log_{10}|H(e^{j0.2\pi})| \geq -1 \quad (dBs)$$
$$20\log_{10}|H(e^{j0.3\pi})| \leq -15 \quad (dBs)$$

The minimum order of a Butterworth filter that satisfies these specifications was found to be 6. A 4th-order Chebyshev filter was found to be the minimum order that satisfies the specifications. Finally, a third-order elliptic filter was found to satisfy the specifications. Magnitude response of the the three filters are shown in Fig. 7.

In general the above situation is typical, the elliptic usually satisfies the specifications with the lowest order, followed by the Chebyshev and the Butterworth's order is the largest of the three. The Butterworth is the flatest, however.

3. FIR Filter Design

The design of FIR filters is typically discrete. There is no equivalent of FIR filters in the analog domain, since it would require the implementation of higher order derivatives which would be very undesirable when dealing with non-smooth signals. The specifications of FIR filters are usually given in the time domain rather than in magnitude and phase. As mentioned before, two advantages of FIR filters are their

(72)