7 Filter Design Techniques

- Filters are a particularly important class of LTI systems. A *Frequency-selective filter* means the LTI system that passes certain frequency components of its input signal and totally rejects others. In general, any system modifies certain frequencies relative to others is also called a *filter*.

- This chapter concentrates on the design of causal discrete-time LTI systems, or commonly referred to as digital filters, to realize analog filters. In particular, we shall focus on frequency-selective lowpass filters, which suppress high-frequency components.

- The design of causal digital filters involves the following stages:

  1. Specifying the desired properties of the target analog system.
  2. Approximating the specifications using a causal discrete-time LTI system.
  3. Realizing the causal discrete-time LTI system.

  This chapter focuses on the second stage.

7.1 Filter Specifications

- *Fig. 7.1* depicts the typical representation of the tolerance limits associated with approximating a discrete-time lowpass filter having unity gain in the passband and zero gain in the stopband ideally

  Fig. 7.1

  which is called a *tolerance scheme*. Here, ω_p and ω_s are the passband edge frequency and the beginning frequency of the stopband, respectively. With both frequencies, the frequency band from $\omega = 0$ to π is partitioned into passband, transition band, and stopband, where the filter magnitude response $|H(e^{j\omega})|$ is constrained be within $[1-\delta_{p2}, 1+\delta_{p1}]$ for the passband, $[0, \delta_s]$ for the stopband, and $[\delta_s, 1-\delta_{p2}]$ for the transition band.

  In most digital filters, the frequency response is usually constrained to support stability and causality. Thus, the poles of the system function have to lie inside the unit circle for the causal and stable IIR filters. The constraint of linear phase is generally imposed in designing the FIR filters.

- Ex: Determining Specifications for a Discrete-Time Filter

  Consider the analog filter realized in the form of discrete-time filtering

  Fig. 7.2

  The overall continuous-time system is an LTI continuous-time system with effective frequency response

  $$H_{eff}(j\varpi) = \begin{cases} H(e^{j\varpi T}), & |\varpi| < \pi/T \\ 0, & |\varpi| \geq \pi/T \end{cases}.$$

  Note that the sampling frequency $1/T$ should be high enough when compared with the bandwidth of the input bandlimited signal.

  In such cases, one can convert from specifications on the effective continuous-time filter to specifications on the discrete-time filter through $\omega = \varpi T$. Thus, given $H_{eff}(j\varpi)$, $H(e^{j\omega})$ is specified by

  $$H(e^{j\omega}) = H_{eff}(j\frac{\omega}{T}) \qquad |\omega| < \pi.$$

  Using *Fig. 7.1*, the specifications on $H(e^{j\omega})$ can be directly obtained from those on $H_{eff}(j\varpi)$.

  For example, if the sampling time is $T = 10^{-4}$ seconds, and $H_{eff}(j\varpi)$ is constrained by (1) $|H_{eff}(j\varpi)| \in [0.9, 1.1]$ for $0 \leq \varpi \leq 4000\pi$ (passband specification) and (2) $|H_{eff}(j\varpi)| \in [0, 0.001]$ for $6000\pi \leq \varpi$ (stopband specification), then the specification parameters for $H(e^{j\omega})$ are given by (1) $\delta_{p1} = \delta_{p2} = 0.1$ and $\omega_p = 0.4\pi$ (passband specification) and (2) $\delta_s = 0.001$ and $\omega_s = 0.6\pi$ (stopband specification).

  Thus, in approximating an analog filter by a digital filter, it suffices to deal with the discrete-time LTI system with $H(e^{j\omega})$ even when the objective is to design a continuous-time LTI system with $H_{eff}(j\varpi)$ provided with a small enough sampling time T.

7.2 Design of Discrete-Time IIR Filters From Continuous-Time Filters

- A part of the following material comes from "J. G. Proakis and D. G. Manolakis, Introduction to Digital Signal Processing, MacMillan, 1988, Section 8.2."

- Our purpose here is to convert an analog filter to a digital filter with infinite impulse response (which yields good amplitude response design in general).

  Note that the techniques for the design of IIR filters have evolved from applying transformations of continuous-time IIR systems into discrete-time IIR systems.

- Consider an analog filter with the rational system function

$$H_c(s) = \frac{\sum_{k=0}^{M} \beta_k s^k}{\sum_{k=0}^{N} \alpha_k s^k} = \int_{-\infty}^{\infty} h_c(t) e^{-st} dt$$

  where $H_c(s)$ is the Laplace transform of $h_c(t)$ with complex variable s and the impulse response $h_c(t)$. This filter can be described by the linear constant-coefficient differential equation

$$\sum_{k=0}^{N} \alpha_k \frac{d^k y(t)}{dt^k} = \sum_{k=0}^{M} \beta_k \frac{d^k x(t)}{dt^k}$$

  where

  Fig. 106-F1

- Note that $H_c(s)$ is stable if all its poles lie in the left-half of the s-plane.

  Thus, if the conversion is to be effective, it should possess the following desirable properties:

  1. The $j\varpi$ axis in the s-plane should map into the unit-circle in the z-plane.
  2. The left-half plane of the s-plane should map into the inside of the unit-circle in the z-plane.

  This converts a stable analog filter to a stable digital filter in the bounded-input bounded-output sense.

7.2.1 IIR Filter Design by Approximation of Derivatives

- Let us sample $y(t)$ by T and have

$$y[n] = y(nT).$$

  Then, approximate $\frac{dy(t)}{dt}$ by a backward difference equation

$$\left.\frac{dy(t)}{dt}\right|_{t=nT} \cong \frac{y(nT) - y((n-1)T)}{T} = \frac{y[n] - y[n-1]}{T}.$$

- Note that

Fig. 107-F1

where $H(z) = \frac{1-z^{-1}}{T}$. Second, approximate

$$\begin{aligned}
\left.\frac{d^2y(t)}{dt^2}\right|_{t=nT} &= \frac{d}{dt}\left[\frac{dy(t)}{dt}\right]_{t=nT} \\
&\cong \frac{(y[n]-y[n-1])/T-(y[n-1]-y[n-2])/T}{T} \\
&= \frac{1}{T^2}(y[n]-2y[n-1]+y[n-2])
\end{aligned}$$

Fig. 107-F2

where $H(z) = \frac{\left(1-z^{-1}\right)^2}{T^2}$. This implies that the transformation $s = \frac{1-z^{-1}}{T}$ can serve as a conversion rule for finding a digital filter to approximate $H_c(s)$. That is, we can find the digital filter by

$$\boxed{H(z) = H_c(s)|_{s=\frac{1-z^{-1}}{T}}}.$$

Now, $s = \frac{1-z^{-1}}{T} \Rightarrow z = \frac{1}{1-sT}$.
If we let $s = j\varpi$, $z = \frac{1}{1-j\varpi T} = \frac{1}{1+\varpi^2T^2} + j\frac{\varpi T}{1+\varpi^2T^2}$ and $|z-\frac{1}{2}|^2 = \frac{1}{4}$ is a circle with center $\frac{1}{2}$ and radius $\frac{1}{2}$.

Fig.107-B1

Thus, a stable analog filter can be converted into a stable digital filter.

- If we approximate $\frac{dy(t)}{dt}$ by a forward difference equation as

$$\frac{dy(t)}{dt} \cong \frac{y((n+1)T)-y(nT)}{T} = \frac{y[n+1]-y[n]}{T}$$

then $s = \frac{z-1}{T}$ will be the conversion rule. In this case,

$$H(z) = H_c(s)|_{s=\frac{z-1}{T}}.$$

If we let $s = j\varpi$, $z = Ts+1 = j\varpi T+1$.

Fig. 108-F1

We can not convert a stable analog filter into a stable digital filter! Thus, this is NOT a good method.

- Ex: Find the digital filter for $H_c(s) = \frac{1}{s+1}$ by the backward difference equation approach.

$$\begin{aligned} H(z) &= H_c(s)\big|_{s=\frac{1-z^{-1}}{T}} = \frac{1}{\frac{1-z^{-1}}{T}+1} \\ &\qquad\qquad\qquad\quad = \frac{zT/(1+T)}{z-1/(1+T)} \end{aligned}$$

The digital filter has a pole at $\frac{1}{1+T}$.

Now, pick $T = 0.1$,

$$H(z) = \frac{0.0909}{1-0.909z^{-1}}.$$

Fig. 108-B1

You can check the quality of the digital filter by $\omega = \varpi T$ and comparing the two charts!

7.2.2 IIR Filter Design by Impulse Invariance

- Our objective is to design a causal digital filter with an impulse response

$$h[n] = T_d h_c(nT_d), \; n = 0, 1, 2, ...$$

with T_d representing a sampling interval. Recall that the frequency response can be related to $H_c(s)$ by

Fig. 109-F1

$$H\left(e^{j\omega}\right) = T_d \cdot \frac{1}{T_d} \sum_{k=-\infty}^{\infty} H_c\left(j\varpi - j\frac{2\pi k}{T_d}\right) \qquad \text{with } \omega = \varpi T_d$$

$$\Rightarrow \quad H(z)\big|_{z=e^{j\omega}} = \sum_{k=-\infty}^{\infty} H_c\left(j\frac{\omega}{T_d} - j\frac{2\pi k}{T_d}\right)$$

$$\Rightarrow \quad H(z)\big|_{z=e^{j\varpi T_d}} = \sum_{k=-\infty}^{\infty} H_c\left(j\varpi - j\frac{2\pi k}{T_d}\right)$$

$$\Rightarrow \quad H(z)\big|_{z=e^{sT_d}} = \sum_{k=-\infty}^{\infty} H_c\left(s - j\frac{2\pi k}{T_d}\right) \quad (s = j\varpi)$$

$$= \sum_{n=0}^{\infty} h[n]\, z^{-n}.$$

This implies that $z = e^{sT_d}$ or $\omega = \varpi T_d$ is the mapping rule.

Fig. 109-F2

The mapping is not one-to-one, but desirable in view of the aforementioned equivalence properties. This many-to-one mapping reflects the effect of aliasing due to sampling.

Fig. 109-B1

If the continuous-time filter is bandlimited with $H_c(j\varpi) = 0$ for $|\varpi| \geq \pi/T_d$, then

$$H\left(e^{j\omega}\right) = H_c\left(j\frac{\omega}{T_d}\right) \qquad |\omega| < \pi$$

and the mapping becomes one-to-one without aliasing. Unfortunately, any practical continuous-time filter can not be exactly bandlimited.

Fig. 7.3

In the band-unlimited case, the sampling time T_d can not be used to eliminate the aliasing! In some applications, the continuous-time filter has to be over-designed to exceed the specifications, particularly in stopband, in a way that the aliasing effect is suppressed.

- Let us consider a causal real-valued continuous-time filter with the system function $H_c(s)$ which has distinct poles s_k's and

$$H_c(s) = \sum_{k=1}^{N} \frac{A_k}{s - s_k} \text{ for } |s| > \max_k |s_k| \qquad (+)$$

$$\Rightarrow \quad h_c(t) = \begin{cases} \sum_{k=1}^{N} A_k e^{s_k t}, & t \geq 0 \\ 0, & t < 0 \end{cases}.$$

Now, sampling $T_d h_c(t)$ periodically at $t = nT_d$ yields

$$\begin{aligned}
h[n] &= T_d h_c(nT_d) \\
&= \sum_{k=1}^{N} T_d A_k e^{n s_k T_d} u[n] \\
\Rightarrow H(z) &= \sum_{n=0}^{\infty} h[n] z^{-n} \\
&= \sum_{n=0}^{\infty} \left(\sum_{k=1}^{N} T_d A_k e^{n s_k T_d} u[n] \right) z^{-n} \\
&= \sum_{k=1}^{N} T_d A_k \sum_{n=0}^{\infty} \left(e^{s_k T_d} z^{-1} \right)^n \\
&= \sum_{k=1}^{N} \frac{T_d A_k}{1 - e^{s_k T_d} z^{-1}}. \qquad \text{(-)}
\end{aligned}$$

Notes:

1. In comparing (+) and (-), we observe that a pole at $s = s_k$ in the s-plane transforms to a pole at $z = e^{s_k T_d}$ in the z-plane, and the coefficients in the partial fraction expansions of $H_c(s)$ and $H(z)$ are equal except for a scaling factor T_d.
2. If the poles s_k's are all real, $H(z)$ can be implemented as a bank of one-pole filters. If some poles are complex, $H(z)$ can be implemented as a bank of two-pole filters. Note that $e^{s_k T_d}$'s may not be all distinct!

- Ex: Consider

$$H_c(s) = \frac{s + 0.1}{(s + 0.1)^2 + 9}.$$

Now,

$$H_c(s) = \frac{\frac{1}{2}}{s + 0.1 - j3} + \frac{\frac{1}{2}}{s + 0.1 + j3}$$
$$(s_k = -0.1 \pm j3)$$
$$\Rightarrow \quad H(z) = \frac{\frac{1}{2} T_d}{1 - e^{-0.1 T_d} e^{j3T_d} z^{-1}} + \frac{\frac{1}{2} T_d}{1 - e^{-0.1 T_d} e^{-j3T_d} z^{-1}}.$$

Notes:

1. T_d has to be small to mitigate aliasing effect!
2. The impulse invariance approach is appropriate only for bandlimited filters. Due to aliasing effect, the impulse invariance approach is approximate for the design of band-unlimited LPF and HPF.

- Ex: Impulse Invariance With a Butterworth Filter

We want to design a continuous-time Butterworth Filter with frequency response $H_c(j\varpi)$ for which

$$\begin{aligned} 0.89125 \leq |H_c(j\varpi)| \leq 1, &\quad 0 \leq |\varpi| \leq 0.2\pi \\ |H_c(j\varpi)| \leq 0.17783, &\quad 0.3\pi \leq |\varpi| \leq \pi. \end{aligned}$$

Now, the magnitude response of a continuous-time Butterworth Filter has the form

$$|H_c(j\varpi)|^2 = \frac{1}{1 + (\varpi/\varpi_c)^{2N}}.$$

Now, we note that $|H_c(j\varpi)|$ is even with $|H_c(j0)| = 1$ and choose that $|H_c(j0.2\pi)| = 0.89125$ and $|H_c(j0.3\pi)| = 0.17783$. Solving both equations yield $N = 5.8858$ and $\varpi_c = 0.70474$. Rounding N by the nearest integer gives $N = 6$ and we can choose $\varpi_c = 0.7032$ to meet the constraints $0.89125 \leq |H_c(j0.2\pi)|$ and $|H_c(j0.3\pi)| \leq 0.17783$.

With $N = 6$ and $\varpi_c = 0.7032$,

$$H_c(s)\, H_c(-s) = \frac{1}{1 + (s/j\varpi_c)^{2N}}$$

$$\Rightarrow \quad H_c(s) = \frac{0.12093}{(s^2 + 0.364s + 0.4945)\,(s^2 + 0.9945s + 0.4945)(s^2 + 1.3585s + 0.4945)}.$$

$H_c(s)$ has poles at

$$-0.182 \pm j0.679, \; -0.497 \pm j0.497, \; -0.679 \pm j0.182$$

$H_c(s)\, H_c(-s)$ has the pole-zero diagram

Fig. 7.4

Expressing $H_c(s)$ in a partial fraction expansion and employing (-) with $T_d = 1$ give

$$\begin{aligned} H(z) &= \frac{0.2871 - 0.4466z^{-1}}{1 - 1.2971z^{-1} + 0.6949z^{-2}} + \frac{-2.1428 + 1.1455z^{-1}}{1 - 1.0691z^{-1} + 0.3699z^{-2}} \\ &\quad + \frac{1.8557 - 0.6303z^{-1}}{1 - 0.9972z^{-1} + 0.257z^{-2}}. \end{aligned}$$

Fig. 7.5

7.2.3 IIR Filter Design by The Bilinear Transformation

- Previous two methods are only good for lowpass, highpass, and a limited class of bandpass filters. This limitation is due to the mapping rules used there.

- The bilinear transform uses a mapping from s-plane to z-plane such that

  1. $j\varpi$-axis is mapped to the unit circle only once;
  2. All points in the right-half-plane (RHP) of the s-plane are mapped into the corresponding points outside the unit circle in the z-plane;
  3. All points in the left-half-plane (LHP) of the s-plane are mapped into the corresponding points inside the unit circle in the z-plane.

- Consider, for example, $H_c(s) = \frac{b}{s+a}$ and

  Fig. 111-F1

  Now,

$$\frac{d}{dt}y(t) + ay(t) = bx(t) \tag{1}$$

  and

$$y(t) = \int_{t_0}^{t} y^{'}(\tau)\,d\tau + y(t_0).$$

  Let us approximate the integral by the trapezoidal formula as follows

$$\begin{aligned} y(nT_d) &= \frac{T_d}{2}\left[y^{'}(nT_d) + y^{'}((n-1)T_d)\right] + y((n-1)T_d) \\ &\overset{(1)}{=} \frac{T_d}{2}\left[-ay(nT_d) + bx(nT_d) - ay((n-1)T_d) + bx((n-1)T_d)\right] \\ &\quad + y((n-1)T_d) \end{aligned}$$

$$\Rightarrow \quad \left(1 + \frac{aT_d}{2}\right)y[n] - \left(1 - \frac{aT_d}{2}\right)y[n-1] = \frac{bT_d}{2}\{x[n] + x[n-1]\}$$

$$\Rightarrow \quad H(z) = \frac{\frac{bT_d}{2}(1+z^{-1})}{1 + \frac{aT_d}{2} - \left(1 - \frac{aT_d}{2}\right)z^{-1}} = \frac{b}{\frac{2}{T_d}\left(\frac{1-z^{-1}}{1+z^{-1}}\right) + a}.$$

  This implies the mapping $s = \frac{2}{T_d}(\frac{1-z^{-1}}{1+z^{-1}})$, which is called the bilinear transformation.

- Let us investigate the characteristic of $s = \frac{2}{T_d}(\frac{1-z^{-1}}{1+z^{-1}})$.

  Now, let $z = re^{j\omega}$ and $s = \sigma + j\varpi$

$$
\begin{aligned}
\Rightarrow \quad s &= \frac{2}{T_d}\frac{z-1}{z+1} = \frac{2}{T_d}\frac{re^{j\omega}-1}{re^{j\omega}+1} \\
&= \frac{2}{T_d}\left(\frac{r^2-1}{1+r^2+2r\cos\omega} + j\frac{2r\sin\omega}{1+r^2+2r\cos\omega}\right) \\
\Rightarrow \quad & \boxed{\sigma = \frac{2}{T_d}\frac{r^2-1}{1+r^2+2r\cos\omega}} \\
& \boxed{\varpi = \frac{2}{T_d}\frac{2r\sin\omega}{1+r^2+2r\cos\omega}}
\end{aligned}
$$

  Notes:

  1. if $r < 1$ (inside the unit circle of the z-plane), then $\sigma < 0$ (in the LHP of s-plane).
  2. if $r > 1$ (outside the unit circle of the z-plane), then $\sigma > 0$ (in the RHP of s-plane).
  3. if $r = 1$ (on the unit circle of the z-plane), then $\sigma = 0$ (on the $j\varpi$-axis of s-plane) and

$$
\begin{aligned}
\varpi &= \frac{2}{T_d}\frac{\sin\omega}{1+\cos\omega} \\
&= \frac{2}{T_d}\tan\frac{\omega}{2} \\
\Rightarrow \quad \omega &= 2\tan^{-1}\frac{\varpi T_d}{2}
\end{aligned}
$$

  which maps

  Fig. 7.6

  Fig. 7.7

  In this case, the entire range in ϖ is mapped only "once" into the range $-\pi \leq \omega < \pi$. However, the mapping is highly nonlinear. In this $\varpi \to \omega$ mapping, one observes a frequency compression, or sometimes called a "frequency warping". For example,

  Fig. 7.8

  Fig. 7.9

- Ex: Consider $H_c(s) = \frac{\varpi_c}{s+\varpi_c}$ where ϖ_c is the 3-dB bandwidth. Our goal is to obtain a digital filter with a 3-dB bandwidth 0.2π.

  Now,

$$\begin{aligned} \omega_c = 0.2\pi &= 2\arctan\frac{\varpi_c T}{2} \\ \Rightarrow \quad \varpi_c &= \frac{2}{T_d}\tan 0.1\pi = \frac{0.65}{T_d}. \end{aligned}$$

  Thus, $H_c(s) = \frac{0.65}{sT_d+0.65}$, which is the corresponding analog filter. Now, let $s = \frac{2}{T_d}\left(\frac{1-z^{-1}}{1+z^{-1}}\right)$, we have

$$H(z) = \frac{0.65\left(1+z^{-1}\right)}{2\left(1-z^{-1}\right)+0.65\left(1+z^{-1}\right)} = \frac{0.65\left(1+z^{-1}\right)}{2.65-1.35z^{-1}}$$

  which is the digital filter with a 3-dB bandwidth 0.2π.

7.3 Discrete-Time Butterworth, Chebyshev and Elliptic Filters

- This subsection presents a number of examples to illustrate IIR filter design. Reading the following examples yourselves.

- Example 7.3: Bilinear Transformation of a Butterworth Filter

  Eqs. (7.28)-(7.36)
  Figs. 7.10-7.11

- Example 7.4: Design Comparisons

  Figs. 7.12-7.16

- Example 7.5 Design Example for Comparison With FIR Designs

  Eqs. (7.37)
  Figs. 7.17-7.22

7.4 Frequency Transformations of Lowpass Filters

- Here, we are interested in composing frequency-selective filters of lowpass, highpass, bandpass, and bandstop types from a lowpass discrete-time filter by use of bilinear-like transformations.

  Figs. 7.23

- Now, given a lowpass system function $H_{lp}(Z)$, we wish to transform to a new system function $H(z)$ with desired characteristics when evaluated on the unit circle $z = e^{j\omega}$.

  First, define a mapping from the Z-plane to the z-plane of the form

  $$Z^{-1} = G(z^{-1})$$

  such that

  $$H(z) = H_{lp}(Z)|_{Z^{-1}=G(z^{-1})}.$$

  In addition, we place the following constraints on the mapping G:

  1. $G(z^{-1})$ must be a rational function of z^{-1}.
  2. The inside of the unit circle of the Z-plane must map to the inside of the unit circle of the z-plane.
  3. The unit circle of the Z-plane must map onto the unit circle of the z-plane.

- Let $Z = e^{j\theta}$ and $z = e^{j\omega}$. Then, constraint 3 implies that

  $$\begin{aligned} |e^{-j\theta}| &= |G(e^{-j\omega})| = 1 \\ e^{-j\theta} &= e^{j\angle G(e^{-j\omega})} \end{aligned}$$

  Thus, $\angle G(e^{-j\omega}) = -\theta$ and $|G(e^{-j\omega})| = 1$. It has been shown by Constantinides that the most general form satisfying both equalities is

  $$G(z^{-1}) = \pm \prod_{k=1}^{N} \frac{z^{-1} - \alpha_k}{1 - \alpha_k z^{-1}}$$

  which maps the inside of the unit circle of the Z-plane to the inside of the unit circle of the z-plane iff $|\alpha_k| < 1$ for all k.

  Special Case: Consider $N = 1$ and

  $$Z^{-1} = G(z^{-1}) = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}}.$$

  With $Z = e^{j\theta}$ and $z = e^{j\omega}$, we have

  $$e^{-j\theta} = \frac{e^{-j\omega} - \alpha}{1 - \alpha e^{-j\omega}}$$

  $$\Longrightarrow \quad \omega = \arctan\left[\frac{(1-\alpha^2)\sin\theta}{2\alpha + (1+\alpha^2)\cos\theta}\right].$$

  Fig. 7.24

  Notes:

1. As indicated, there is a frequency warping except $\alpha = 0$.
2. If $H_{lp}(Z)$ has a piecewise-constant lowpass frequency response with cutoff frequency θ_p, then $H(z)$ has a similar lowpass response with cutoff frequency ω_p determined by α. In fact, α can be expressed in terms of θ_p and ω_p as

$$\alpha = \frac{\sin[(\theta_p - \omega_p)/2]}{\sin[(\theta_p + \omega_p)/2]}.$$

   This α in turn defines the transformation

$$H(z) = H_{lp}(Z)|_{Z^{-1}=\frac{z^{-1}-\alpha}{1-\alpha z^{-1}}}.$$

3. Table 7.1 lists various transformations from $H_{lp}(Z)$ with cutoff frequency θ_p to other types of filters.

Table 7.1

- Example 7.6: Transformation of a Lowpass Filter to a Highpass Filter

Eqs. (7.49)-(7.52)
Figs. 7.25-7.26

7.5 Design of FIR Filters by Windowing

- The design techniques for FIR filters are based on directly approximating the ideal desired frequency response $H_d(e^{j\omega})$ or impulse response $h_d[n]$ of the the discrete-time system, where $H_d(e^{j\omega})$ and $h_d[n]$ are related by

$$\begin{aligned} H_d(e^{j\omega}) &= \sum_{n=-\infty}^{\infty} h_d[n]e^{-j\omega n} \\ h_d[n] &= \frac{1}{2\pi}\int_{-\pi}^{\pi} H_d(e^{j\omega})e^{j\omega n}d\omega. \end{aligned}$$

  Notes:

  1. $h_d[n]$ can be thought of as the coefficient of the Fourier series expansion of periodic frequency response $H_d(e^{j\omega})$ corresponding to $e^{-j\omega n}$.
  2. Truncating $h_d[n]$ can provide an approximation to the ideal impulse response. Such a method is called the *windowing* method.

- By windowing, we define a new system with impulse response $h[n]$

$$h[n] = h_d[n]w[n]$$

where $w[n]$ is a finite-duration window defined by

$$w[n] = \begin{cases} 1, & n = 0, 1, ..., M \\ 0, & \text{otherwise} \end{cases}.$$

From the modulation theorem in Section 2.9.7, the frequency response $H(e^{j\omega})$ corresponding to $h[n]$ is obtained as

$$H(e^{j\omega}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_d(e^{j\theta})W(e^{j(\omega-\theta)})d\theta \qquad (*)$$

where $W(e^{j\omega}) = \sum_{n=-\infty}^{\infty} w[n]e^{-j\omega n}$.

Fig. 7.27

Notes:

1. If $w[n] = 1$ for all n, then $W(e^{j\omega}) = \sum_{n=-\infty}^{\infty} e^{-j\omega n} = 2\pi \sum_{n=-\infty}^{\infty} \delta(\omega + 2\pi k)$ (from Poisson's sum formula or Table 2.3) which is a periodic impulse train in ω and $H(e^{j\omega}) = H_d(e^{j\omega})$. In the case, $W(e^{j\omega})$ is highly concentrated in frequency, but $w[n]$ has an extremely long duration.
2. When $w[n]$ has a finite duration with length $M + 1$,

$$W(e^{j\omega}) = \sum_{n=0}^{M} e^{-j\omega n} = \frac{1 - e^{-j\omega(M+1)}}{1 - e^{-j\omega}} = e^{-j\omega M/2}\frac{\sin(\omega(M+1)/2)}{\sin(\omega/2)}.$$

Fig. 7.28

The area under each lobe is constant. Because of the use of rectangular window, sidelobes have large peak amplitudes decaying asymptotically as $|\omega|^{-1}$ when $|\omega|$ is increased. As M increases, the mainlobe width $\Delta\omega_m$ and all sidelobe widths decreases, while peak amplitudes of all lobes grow. This increases the oscillation of the integration in (*) when there is discontinuities in $H_d(e^{j\omega})$ and cause nonuniform convergence for a large M. Such a phenomenon is called the *Gibbs phenomenon.* The phenomenon is the most significant when a rectangular window is used. This phenomenon can be mitigated by using *smoother* windows with *continuous* window edges.

7.5.1 Properties of Commonly Used Windows

- Some commonly used windows are given here:

1. Rectangular Window

$$w[n] = \begin{cases} 1, & n = 0, 1, ..., M \\ 0, & \text{otherwise} \end{cases}.$$

2. Bartlett (Triangular) Window (With an even M)

$$w[n] = \begin{cases} 2n/M, & n = 0, 1, ..., M/2 \\ 2 - 2n/M, & n = M/2+1, M/2+2, ..., M \\ 0, & \text{otherwise} \end{cases}.$$

3. Hann Window

$$w[n] = \begin{cases} \frac{1}{2}[1 - \cos(\frac{2\pi n}{M})], & n = 0, 1, ..., M \\ 0, & \text{otherwise} \end{cases}.$$

4. Hamming Window

$$w[n] = \begin{cases} 0.54 - 0.46\cos(\frac{2\pi n}{M}), & n = 0, 1, ..., M \\ 0, & \text{otherwise} \end{cases}.$$

5. Blackman Window

$$w[n] = \begin{cases} 0.42 - 0.5\cos(\frac{2\pi n}{M}) + 0.08\cos(\frac{4\pi n}{M}), & n = 0, 1, ..., M \\ 0, & \text{otherwise} \end{cases}.$$

Notes:

1. Bartlett, Hann, and Blackman windows have zero edges at $n = 0$ and M, and the corresponding window length is really only $M-1$ samples. Thus, these windows are seemingly continuous at window edges, and thus smoother than the rectangular window.

   Fig. 7.29

2. The magnitude responses $|W(e^{j\omega})|$ are plotted in

   Fig. 7.30
   Table 7.2

   As indicated, by tapering the window smoothly to zero, Bartlett, Hamming, Hann, and Blackman windows exhibit wider mainlobe and smaller sidelobes than the rectangular window.

7.5.2 Incorporation of Generalized Linear Phase

- All previously introduced windows have the symmetric property that

$$w[n] = \begin{cases} w[M-n], & n = 0, 1, ..., M \\ 0, & \text{otherwise} \end{cases}$$

and their $W(e^{j\omega})$ have the common form of

$$W(e^{j\omega}) = W_e(e^{j\omega})e^{-j\omega M/2}$$

where $W_e(e^{j\omega})$ is even and real-valued. This property results in the following.

- If the desired impulse response $h_d[n]$ is symmetric about $M/2$, i.e., $h_d[M-n] = h_d[n]$, then the windowed impulse response $h[n] = h_d[n]w[n]$ is also symmetric about $M/2$ and its frequency response has a generalized linear phase, as

$$H(e^{j\omega}) = A_e(e^{j\omega})e^{-j\omega M/2}$$

where $A_e(e^{j\omega})$ is even and real-valued. This is because $h_d[n]$ has the frequency response

$$H_d(e^{j\omega}) = H_e(e^{j\omega})e^{-j\omega M/2}$$

where $H_e(e^{j\omega})$ is even and real-valued, and thus

$$\begin{aligned} H(e^{j\omega}) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} H_d(e^{j\theta})W(e^{j(\omega-\theta)})d\theta \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} H_e(e^{j\theta})e^{-j\theta M/2}W_e(e^{j(\omega-\theta)})e^{-j(\omega-\theta)M/2}d\theta \\ &= \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} H_e(e^{j\theta})W_e(e^{j(\omega-\theta)})d\theta\right] e^{-j\omega M/2} \\ &= A_e(e^{j\omega})e^{-j\omega M/2}. \end{aligned}$$

- Similarly, if the desired impulse response $h_d[n]$ is antisymmetric about $M/2$, i.e., $h_d[M-n] = -h_d[n]$, then the windowed impulse response $h[n] = h_d[n]w[n]$ is also antisymmetric about $M/2$ and its frequency response has a generalized linear phase, as

$$H(e^{j\omega}) = jA_o(e^{j\omega})e^{-j\omega M/2}$$

where $A_o(e^{j\omega})$ is odd and real-valued.

- Ex: Linear-Phase Lowpass Filter (Self-Reading)

  Consider the desired linear-phase frequency response

$$H_{d,lp}(e^{j\omega}) = \begin{cases} e^{-j\omega M/2}, & |\omega| < \omega_c \\ 0, & \omega_c < |\omega| \le \pi \end{cases}.$$

  The corresponding desired impulse response is

$$h_{d,lp}[n] = \frac{\sin(\omega_c(n - M/2))}{\pi(n - M/2)}$$

  which is symmetric about $M/2$. With a symmetric window $w[n]$,

$$h[n] = \frac{\sin(\omega_c(n - M/2))}{\pi(n - M/2)} w[n]$$

  is also symmetric about $M/2$ and results in a linear-phase lowpass system. The upper part of

  Fig. 7.31

  depicts the character of the amplitude response of $H(e^{j\omega})$ that would result for all considered window functions except rarely used Bartlett window.

  Table 7.2

7.5.3 The Kaiser Window Filter Design Method

- Sidelobe magnitude and mainlobe width are two important factors in designing the window function that is maximally concentrated around $\omega = 0$ in the frequency domain. Complicated prolate spheroidal wave functions and simple Kaiser function are some well known solutions for trading off the two factors. Here, we consider the Kaiser window function.

- Kaiser window is defined as

$$w[n] = \begin{cases} I_0[\beta(1 - [(n-\alpha)/\alpha]^2)^{1/2}], & n = 0, 1, ..., M \\ 0, & \text{otherwise} \end{cases}$$

  where $\alpha = M/2$ and $I_0[\cdot]$ represents the zeroth-order modified Bessel function of the first kind defined by

$$I_0[x] \triangleq \frac{1}{2\pi} \int_{\theta}^{\theta+2\pi} e^{x \cos\phi} d\phi \qquad \text{for any } \theta.$$

Note that $I_0[x]$ is even and monotonically increasing with respect to its nonnegative argument and with $I_0[0] = 1$. Since $I_0[x]$ is even, we can restrict the parameter β to be nonnegative. By varying $M + 1$ and β, the window length and shape can be adjusted to trade sidelobe magnitude for mainlobe width.

Fig. 7.32

Note that the Kaiser window specializes to the rectangular window when $\beta = 0$.

- In *Fig. 7.31* with $\delta_1 = \delta_2 = \delta$, the passband cutoff frequency ω_p and the stopband cutoff frequency ω_s are related to the peak approximation error δ by the highest positive frequency satisfying $|H(e^{j\omega})| \geq 1 - \delta$ and the lowest positive frequency satisfying $|H(e^{j\omega})| \leq \delta$, respectively. The width $\Delta\omega$ of the transition region from the passband to the stopband is

$$\Delta\omega = \omega_s - \omega_p.$$

For the Kaiser window, in order to achieve specified $A = -20 \log_{10} \delta$ and $\Delta\omega$, β and M are determined empirically by

$$\begin{aligned} \beta &= \begin{cases} 0.1102(A - 8.7), & A > 50 \\ 0.5842(A - 21)^{0.4} + 0.07886(A - 21), & 21 \leq A \leq 50 \\ 0 & A < 21 \end{cases} \\ M &\cong \frac{A - 8}{2.285\Delta\omega}. \end{aligned}$$

Both empirical formulae facilitate the design of the Kaiser window.

- Various windows can be compared in terms of the A versus $\Delta\omega$ characteristic as

Fig. 7.33

Note that the Kaiser window can be designed to provide the smallest $\Delta\omega$ for a given A among all considered windows.

7.6 Examples of FIR Filter Design By the Kaiser Window Method

- This section gives some design examples of FIR filters approximated from ideal ones by use of the Kaiser window.

- Lowpass Filter Design

  We want to approximate the desired linear-phase frequency response

$$H_{d,lp}(e^{j\omega}) = \begin{cases} e^{-j\omega M/2}, & |\omega| < \omega_c \\ 0, & \omega_c < |\omega| \le \pi \end{cases}$$

  with the corresponding desired impulse response

$$h_{d,lp}[n] = \frac{\sin(\omega_c(n - M/2))}{\pi(n - M/2)}.$$

  1. Specify $\omega_p = 0.4\pi$, $\omega_s = 0.6\pi$, and $\delta = 0.001$.
  2. Compute the cutoff frequency $\omega_c = \frac{\omega_p+\omega_s}{2} = 0.5\pi$, the transition width $\Delta\omega = \omega_s - \omega_p = 0.2\pi$, and the decibel approximation error $A = -20\log_{10}\delta = 60$.
  3. Using the aforementioned empirical formulae, we predict $\beta = 5.5653$ and $M = 37$.
  4. Obtain the impulse response

$$h[n] = \begin{cases} h_{d,lp}[n]\frac{I_0[\beta(1-[(n-\alpha)/\alpha]^2)^{1/2}]}{I_0[\beta]}, & n = 0, 1, ..., M \\ 0, & \text{otherwise} \end{cases}$$

     which is a type-II linear-phase FIR filter with an odd M.

     Fig. 7.34

  5. Compute the approximation error based on the function

$$E_A(\omega) = \begin{cases} 1 - A_e(e^{j\omega}), & 0 \le \omega \le \omega_p \\ 0 - A_e(e^{j\omega}), & \omega_s \le \omega \le \pi \end{cases}.$$

     The actual peak approximation error is 0.00113.
  6. The group delay is given by $M/2 = 18.5$ samples.

- Highpass Filter Design

  We want to approximate the desired linear-phase frequency response

$$H_{d,hp}(e^{j\omega}) = \begin{cases} 0, & |\omega| < \omega_c \\ e^{-j\omega M/2}, & \omega_c < |\omega| \le \pi \end{cases} = e^{-j\omega M/2} - H_{d,lp}(e^{j\omega})$$

  with the corresponding desired impulse response

$$h_{d,hp}[n] = \frac{\sin(\pi(n - M/2))}{\pi(n - M/2)} - \frac{\sin(\omega_c(n - M/2))}{\pi(n - M/2)}.$$

Suppose that the filter to be designed should meet the highpass specifications

$$\begin{array}{ll} |H(e^{j\omega})| \le \delta & \text{for } |\omega| < \omega_s \\ 1-\delta \le |H(e^{j\omega})| \le 1+\delta & \text{for } \omega_p \le |\omega| \le \pi \end{array} .$$

1. Specify $\omega_s = 0.35\pi$, $\omega_p = 0.5\pi$, and $\delta = 0.02$.
2. Compute the cutoff frequency $\omega_c = \frac{\omega_p+\omega_s}{2} = 0.425\pi$, the transition width $\Delta\omega = \omega_p - \omega_s = 0.15\pi$, and the decibel approximation error $A = -20\log_{10}\delta = 33.98$.
3. Using the aforementioned empirical formulae, we predict $\beta = 2.65$ and $M = 24$.
4. Obtain the impulse response

$$h[n] = \begin{cases} h_{d,hp}[n]\frac{I_0[\beta(1-[(n-\alpha)/\alpha]^2)^{1/2}]}{I_0[\beta]}, & n = 0, 1, ..., M \\ 0, & \text{otherwise} \end{cases}$$

which is a type-I linear-phase FIR filter with an even M.

Fig. 7.35

5. The actual peak approximation error is 0.0209.
6. The group delay is given by $M/2 = 12$ samples.
7. We can also choose $M = 25$ and come out with a type-II linear-phase FIR filter.

Fig. 7.36

This filter is unsatisfactory due to the zero of $H(z)$ at $z = -1$, i.e., $\omega = \pi$. This illustrates that type-II FIR linear-phase systems are generally not appropriate approximations for either highpass or bandstop filters.

Note: The generalization to a filter design with multiple passbands and stopbands is also possible.

Fig. 7.37

See pp. 579 for self reading.

- Self-read Section 7.6.3 about the design of discrete-time differentiators based on the Kaiser window.