Chapter 5: FIR and IIR Filters

Sem. II, 2010 E.C Electrical and Computer Engineering By: Waltengus

Outline

- Filter Design Methods
- Interpolation and Decimation

4.2 Filter Design

- Filter design process begins with the filter specifications, which may include constraints on the magnitude and/or phase of the frequency response, constraints on the unit sample response or step response of the filter, specification of the type of filter (e.g., finite-length impulse response (FIR) or IIR), and the filter order.
- Once the specifications have been defined,
- The next step is to find a set of filter coefficients that produce an acceptable filter.
- After the filter has been designed, the last step is to implement the system in hardware or software, quantizing the filter coefficients if necessary, and choosing an appropriate filter structure

2.2.1 Filter specifications

- Before a filter can be designed, a set of filter specifications must be defined.
- For example, suppose that we would like to design a low-pass filter with a **cutoff frequency** ω_c.
- The frequency response of an ideal low-pass filter with linear phase and a cutoff frequency ω_c is

$$H_d(e^{j\omega}) = \begin{cases} e^{-j\alpha\omega} & |\omega| \le \omega_c \\ 0 & \omega_c < |\omega| \le \pi \end{cases}$$

which has a unit sample response

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

- Because this filter is unrealizable (non causal and unstable), it is necessary to relax the ideal constraints on the frequency response and allow some deviation from the ideal response.
- **The specifications for a low-pass filter will typically have the form**

$$1 - \delta_p < |H(e^{j\omega})| \le 1 + \delta_p \qquad 0 \le |\omega| < \omega_p$$

$$|H(e^{j\omega})| \le \delta_s \qquad \omega_s \le |\omega| < \pi$$

- As illustrated in Fig. 4-1. Thus, the specifications include the **pass band cutoff frequency, ω_p the stop band cutoff frequency, ω_s the pass band deviation, δ_p. and the stop band deviation, δ_s.**
- The **pass band** and **stop band deviations** are often given in decibels (dB) as follows: $\alpha_p = -20\log(1 - \delta_p)$

$$\alpha_s = -20\log(\delta_s)$$

- Where α_p is pass band attenuation and α_s the stop band attenuation

❖ The interval [ω_p, ω_s] is called the **transition band**.

❖ Once the filter specifications have been defined, the next step is to design a filter that meets these specifications.

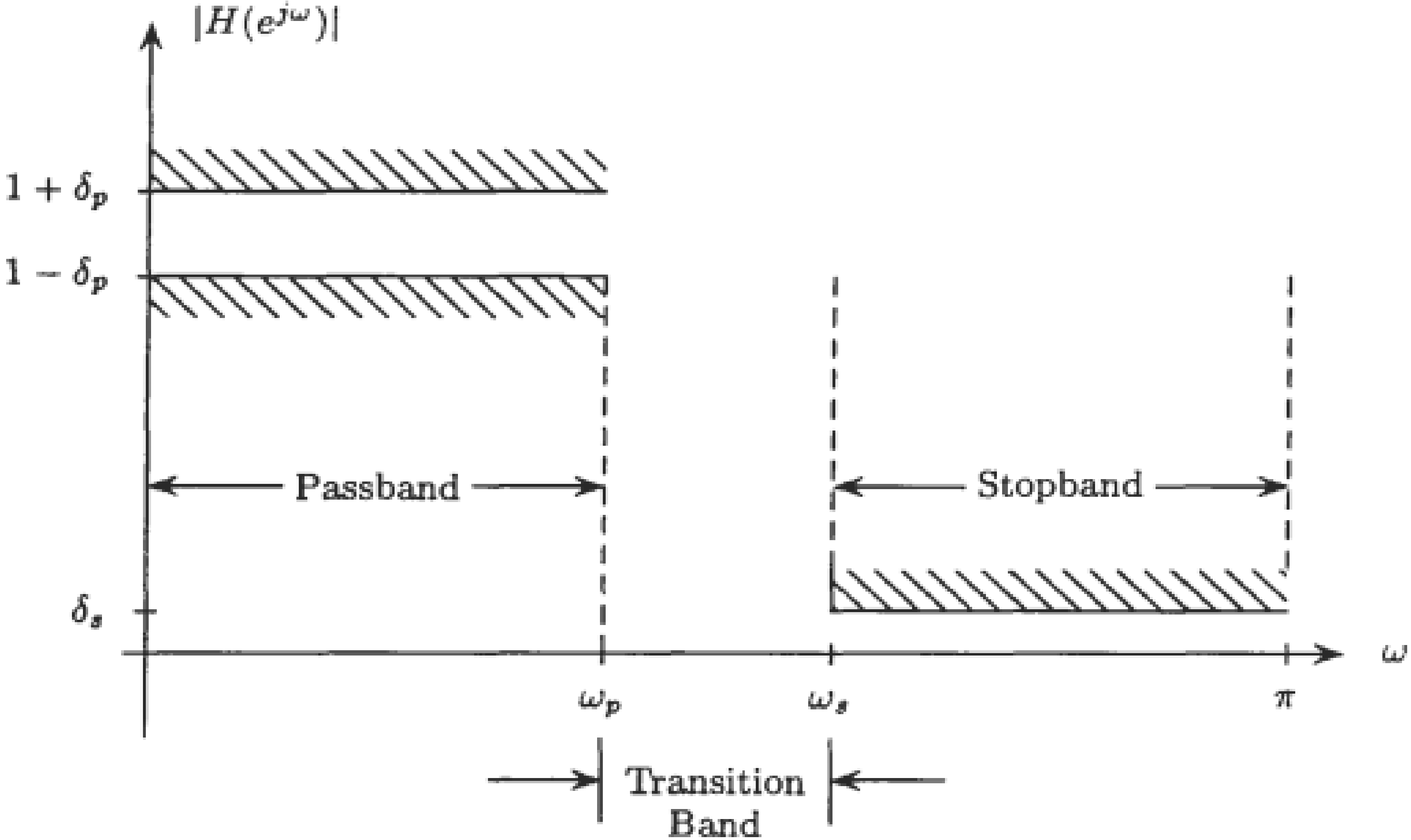


Fig. 4-1. Filter specifications for a low-pass filter,

4.3 FIR Filter Design

The frequency response of an Nth-order causal FIR filter is

$$H(e^{j\omega}) = \sum_{n=0}^{N} h(n)e^{-jn\omega}$$

- The design of an FIR filter involves finding the coefficients h(n) that result in a frequency response that satisfies a given set of filter specifications.
- **FIR filters** have **two important advantages** over **IIR filters**.

✓ **First**, they are guaranteed to be **stable**, even after the filter coefficients have been **quantized**.

✓ **Second,** they may be easily constrained to have (generalized) linear phase. Because FIR filters are generally designed to have **linear phase**, in the following we consider the design of linear phase FIR filters.

4.3.1 Linear Phase FIR Design Using Windows

Let $h_d(n)$ be the unit sample response of an ideal frequency selective filter with linear phase,

$$H_d(e^{j\omega}) = A(e^{j\omega})e^{-j(\alpha\omega-\beta)}$$

Because $h_d(n)$ will generally be infinite in length, it is necessary to find an FIR approximation to $H_d(e^{j\omega})$. With the window design method, the filter is designed by windowing the unit sample response,

$$h(n) = h_d(n)w(n)$$

where $w(n)$ is a finite-length window that is equal to zero outside the interval $0 \le n \le N$ and is symmetric about its midpoint:

$$w(n) = w(N-n)$$

The effect of the window on the frequency response may be seen from the complex convolution theorem,

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

- The ideal frequency response is smoothed by the **discrete-time Fourier transform** of the window, $W(e^{j\omega})$.
- **There are many different types of windows that may be used in the window design method, a few of which are listed in Table 4-1.**

Table 4-1 Some Common Windows

Rectangular	$w(n) = \begin{cases} 1 & 0 \le n \le N \\ 0 & \text{else} \end{cases}$
Hanning	$w(n) = \begin{cases} 0.5 - 0.5\cos\left(\frac{2\pi n}{N}\right) & 0 \le n \le N \\ 0 & \text{else} \end{cases}$
Hamming	$w(n) = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{N}\right) & 0 \le n \le N \\ 0 & \text{else} \end{cases}$
Blackman	$w(n) = \begin{cases} 0.42 - 0.5\cos\left(\frac{2\pi n}{N}\right) + 0.08\cos\left(\frac{4\pi n}{N}\right) & 0 \le n \le N \\ 0 & \text{else} \end{cases}$

- **This window is also called a Hann window or a von Hann window**

❖ **How well the frequency response of a filter designed with the window design method approximates a desired response, $H_d(e^{j\omega})$, is determined by two factors (see Fig. 4-2):**

1. The width of the main lobe of $W(e^{j\omega})$.
2. The peak side-lobe amplitude of $W(e^{j\omega})$.

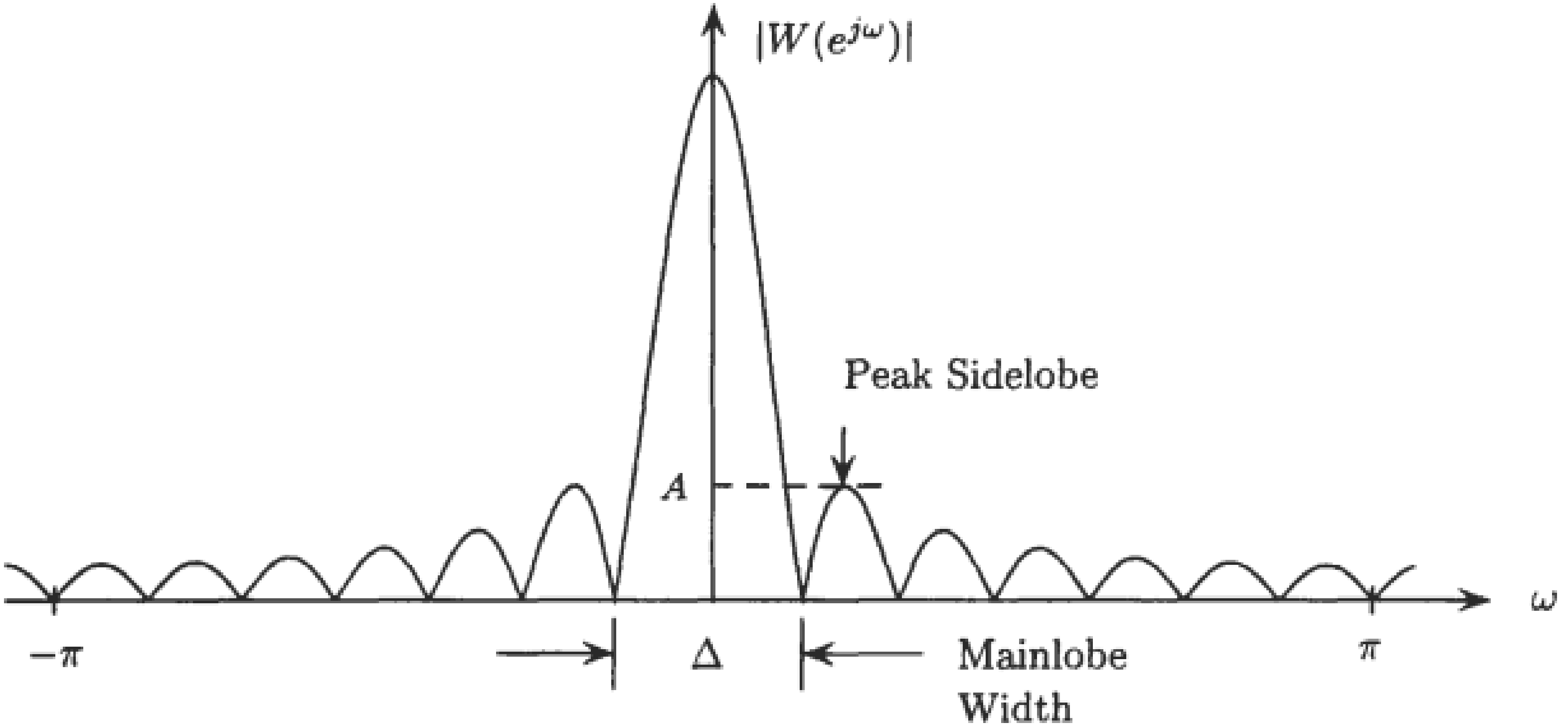


Fig. 4-2. The DTFT of a typical window, which is characterized by the width of its **main lobe Δ** and the **peak amplitude of its side lobes, A**, relative to the amplitude of $W(e^{j\omega})$ at $\omega = 0$.

- Ideally, the **main-lobe width** should be **narrow**, and the **side-lobe amplitude** should be **small**.
- However, for a fixed-length window, these cannot be minimized independently. Some general properties of windows are as follows:

1. As the length **N** of the window **increases**, the **width** of the **main lobe decreases**, which results in a **decrease** in the transition width between pass bands and stop bands. This relationship is given approximately by

$$N\,\Delta f = c$$

- ✓ Where **Δf** is the transition width, and **c** is a parameter that depends on the window.

2. The peak side-lobe amplitude of the window is determined by the **shape of the window**, and it is essentially independent of the window length.
3. If the window shape is changed to **decrease the side-lobe amplitude**, the width of the main lobe will generally **increase**.

❖ Table 4.2 below are the side-lobe amplitudes of several windows along with the approximate transition width and stop band attenuation that results when the given window is used to design an **Nth order low pass filter.**

❖ The Peak Side-Lobe Amplitude of Some Common Windows and the Approximate. Transition Width and Stop band Attenuation of an Nth-Order Low-Pass Filter Designed Using the Given Window.

Window	Side-Lobe Amplitude (dB)	Transition Width (Δf)	Stopband Attenuation (dB)
Rectangular	−13	$0.9/N$	−21
Hanning	−31	$3.1/N$	−44
Hamming	−41	$3.3/N$	−53
Blackman	−57	$5.5/N$	−74

Table 4.2

Example 4.3.1 Suppose that we would like to design an FIR linear phase low-pass filter according to the following specifications:

$$0.99 \le |H(e^{j\omega})| \le 1.01 \qquad 0 \le |\omega| \le 0.19\pi$$

$$|H(e^{j\omega})| \le 0.01 \qquad 0.21\pi \le |\omega| \le \pi$$

For a stopband attenuation of $20\log(0.01) = -40$ dB, we may use a Hanning window. Although we could also use a Hamming or a Blackman window, these windows would overdesign the filter and produce a larger stopband attenuation at the expense of an increase in the transition width. Because the specification calls for a transition width of $\Delta\omega = \omega_s - \omega_p = 0.02\pi$, or $\Delta f = 0.01$, with

$$N\,\Delta f = 3.1$$

for a Hanning window an estimate of the required filter order is

$$N = \frac{3.1}{\Delta f} = 310$$

The last step is to find the unit sample response of the ideal low-pass filter that is to be windowed. With a cutoff frequency of $\omega_c = (\omega_s + \omega_p)/2 = 0.2\pi$, and a delay of $\alpha = N/2 = 155$, the unit sample response is

$$h_d(n) = \frac{\sin[0.2\pi(n-155)]}{(n-155)\pi}$$

Kaiser developed a *family* of windows that are defined by

$$w(n) = \frac{I_0[\beta(1-[(n-\alpha)/\alpha]^2)^{1/2}]}{I_0(\beta)} \qquad 0 \le n \le N$$

where $\alpha = N/2$, and $I_0(\cdot)$ is a zeroth-order modified Bessel function of the first kind, which may be easily generated using the power series expansion

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

- The parameter β determines the **shape of the window** and thus controls the trade-off between main-lobe width and side-lobe amplitude.
- A Kaiser window is nearly optimum in the sense of having the most energy in its main lobe for a given side-lobe amplitude. Table 4-3 illustrates the effect of changing the parameter β.

- There are two empirically derived relationships for the Kaiser window that facilitate the use of these windows to design FIR filters.
- The first relates the stop band ripple of a low-pass filter,

$\alpha_s = -20\log(\delta_s)$, to the parameter β,

$$\beta = \begin{cases} 0.1102(\alpha_s - 8.7) & \alpha_s > 50 \\ 0.5842(\alpha_s - 21)^{0.4} + 0.07886(\alpha_s - 21) & 21 \le \alpha_s \le 50 \\ 0.0 & \alpha_s < 21 \end{cases}$$

Table 4-3 Characteristics of the Kaiser Window as a Function of β

Parameter β	Side Lobe (dB)	Transition Width ($N\Delta f$)	Stopband Attenuation (dB)
2.0	−19	1.5	−29
3.0	−24	2.0	−37
4.0	−30	2.6	−45
5.0	−37	3.2	−54
6.0	−44	3.8	−63
7.0	−51	4.5	−72
8.0	−59	5.1	−81
9.0	−67	5.7	−90
10.0	−74	6.4	−99

The second relates N to the transition width Δf and the stopband attenuation α_s,

$$N = \frac{\alpha_s - 7.95}{14.36\Delta f} \qquad \alpha_s \geq 21$$

Note that if $\alpha_s < 21$ dB, a rectangular window may be used ($\beta = 0$), and $N = 0.9/\Delta f$.

Example 4.3.2

Suppose that we would like to design a low-pass filter with a cutoff frequency $\omega_c = \pi/4$, a transition width $\Delta\omega = 0.02\pi$, and a stopband ripple $\delta_s = 0.01$. Because $\alpha_s = -20\log(0.01) = -40$, the Kaiser window parameter is

$$\beta = 0.5842(40-21)^{0.4} + 0.07886(40-21) = 3.4$$

With $\Delta f = \Delta\omega/2\pi = 0.01$, we have

$$N = \frac{40 - 7.95}{14.36 \cdot (0.01)} = 224$$

Therefore,

$$h(n) = h_d(n)w(n)$$

where

$$h_d(n) = \frac{\sin[(n-112)\pi/4]}{(n-112)\pi}$$

is the unit sample response of the ideal low-pass filter.

- Although it is simple to design a filter using the window design method, there are some **limitations** with this method.
- **First**, it is necessary to find a closed-form expression for $h_d(n)$ (or it must be approximated using a very long DFT).
- **Second**, for a frequency selective filter, the transition widths between frequency bands, and the ripples with in these bands, will be approximately the same.
- As a result, the window design method requires that the filter be designed to the tightest tolerances in all of the bands by selecting the smallest transition width and the smallest ripple.
- **Finally**, window design filters are **not, in general, optimum** in the sense that they do not have the smallest possible ripple for a given filter order and a given set of cutoff frequencies.

4.3.2 Frequency Sampling Filter Design

Another method for FIR filter design is the frequency sampling approach. In this approach, the desired frequency response, $H_d(e^{j\omega})$, is first uniformly sampled at N equally spaced points between 0 and 2π:

$$H(k) = H_d\left(e^{j2\pi k/N}\right) \qquad k = 0, 1, \ldots, N-1$$

These frequency samples constitute an N-point DFT, whose inverse is an FIR filter of order $N-1$:

$$h(n) = \frac{1}{N}\sum_{k=0}^{N-1} H(k)e^{j2\pi nk/N} \qquad 0 \le n \le N-1$$

The relationship between $h(n)$ and $h_d(n)$ (see Chap. 3) is

$$h(n) = \sum_{k=-\infty}^{\infty} h_d(n+kN) \qquad 0 \le n \le N-1$$

- Although the frequency samples match the ideal frequency response exactly, there is no control on how the samples are interpolated between the samples.
- Because filters designed with the frequency sampling method are not generally very good, this method is often modified by introducing one or more transition samples as illustrated in Fig. 4-3.
- These transition samples are optimized in an iterative manner to maximize the stop band attenuation or minimize the pass band ripple.

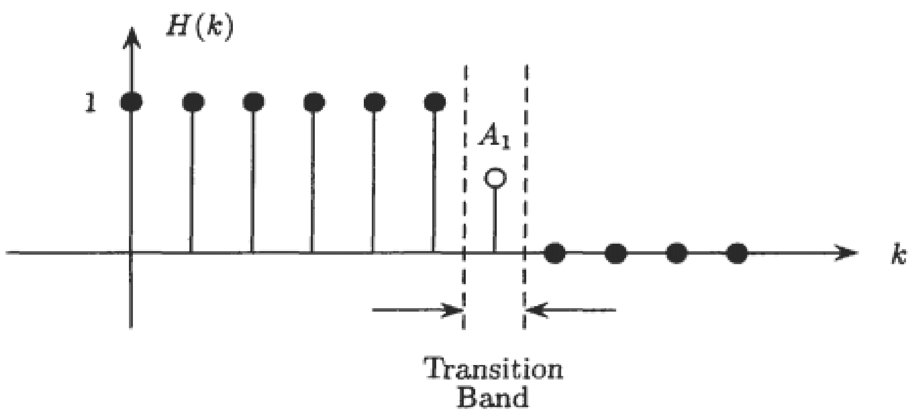


Fig. 4-3. Introducing a transition sample with an amplitude of A_1 in the frequency sampling method.

4.3.3 Equiripple Linear Phase Filters

The design of an FIR low-pass filter using the window design technique is simple and generally results in a filter with relatively good performance. However, in two respects, these filters are not optimal:

1. First, the passband and stopband deviations, δ_p and δ_s, are approximately equal. Although it is common to require δ_s to be much smaller than δ_p, these parameters cannot be independently controlled in the window design method. Therefore, with the window design method, it is necessary to *overdesign* the filter in the passband in order to satisfy the stricter requirements in the stopband.
2. Second, for most windows, the ripple is not uniform in either the passband or the stopband and generally decreases when moving away from the transition band. Allowing the ripple to be uniformly distributed over the entire band would produce a smaller *peak ripple*.

An equiripple linear phase filter, on the other hand, is optimal in the sense that the magnitude of the ripple is minimized in all bands of interest for a given filter order, N. In the following discussion, we consider the design of a type I linear phase filter. The results may be easily modified to design other types of linear phase filters.

The frequency response of an FIR linear phase filter may be written as

$$H(e^{j\omega}) = A(e^{j\omega})e^{-j\alpha\omega}$$

where the amplitude, $A(e^{j\omega})$, is a real-valued function of ω. For a type I linear phase filter,

$$h(n) = h(N - n)$$

where N is an even integer. The symmetry of $h(n)$ allows the frequency response to be expressed as

$$A(e^{j\omega}) = \sum_{k=0}^{L} a(k)\cos(k\omega)$$

where $L = N/2$ and

$$a(0) = h\left(\frac{N}{2}\right)$$

$$a(k) = h\left(k + \frac{N}{2}\right) \qquad k = 1, 2, \ldots, \frac{N}{2}$$

The terms $\cos(k\omega)$ may be expressed as a sum of powers of $\cos\omega$ in the form

$$\cos(k\omega) = T_k(\cos\omega)$$

where $T_k(x)$ is a kth-order Chebyshev polynomial [see Eq. (9.9)]. Therefore, Eq. (9.4) may be written as

$$A(e^{j\omega}) = \sum_{k=0}^{L} \alpha(k)(\cos\omega)^k$$

Thus, $A(e^{j\omega})$ is an Lth-order polynomial in $\cos\omega$.

With $A_d(e^{j\omega})$ a desired amplitude, and $W(e^{j\omega})$ a positive weighting function, let

$$E(e^{j\omega}) = W(e^{j\omega})[A_d(e^{j\omega}) - A(e^{j\omega})]$$

be a weighted approximation error. The equiripple filter design problem thus involves finding the coefficients $a(k)$ that minimize the maximum absolute value of $E(e^{j\omega})$ over a set of frequencies, $\mathcal{F}$,

$$\min_{a(k)} \left\{ \max_{\omega \in \mathcal{F}} |E(e^{j\omega})| \right\}$$

For example, to design a low-pass filter, the set $\mathcal{F}$ will be the frequencies in the passband, $[0, \omega_p]$, and the stopband, $[\omega_s, \pi]$, as illustrated in Fig. 9-4. The transition band, (ω_p, ω_s), is a *don't care* region, and it is not

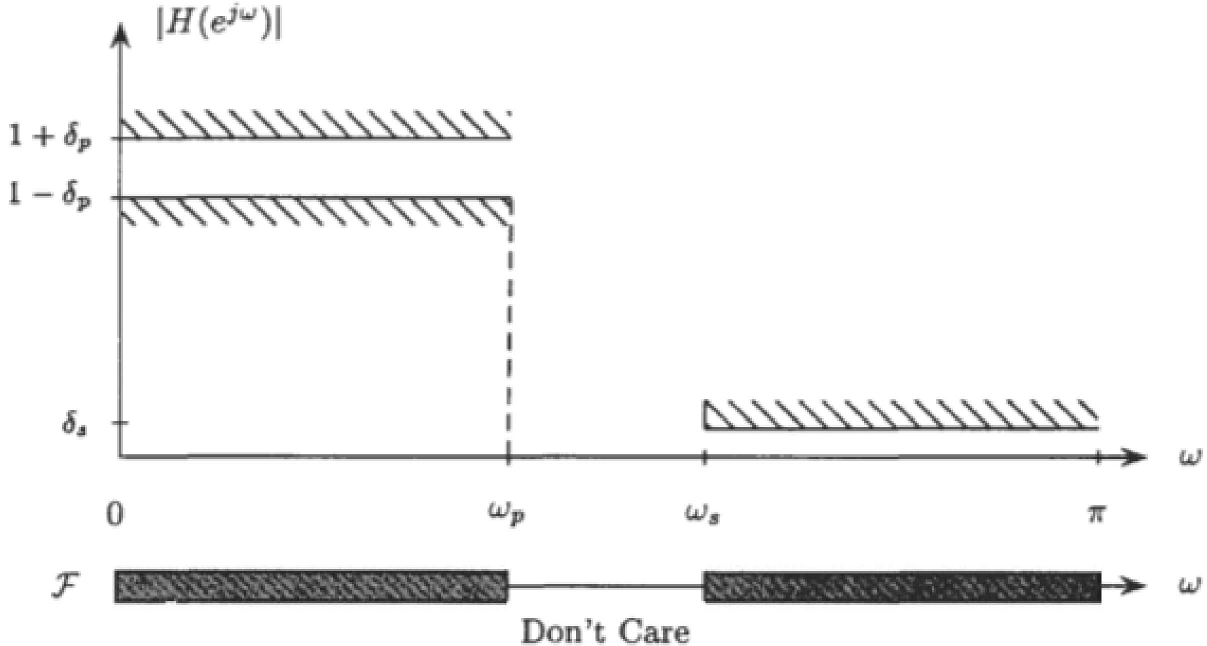


Fig. 4-4. The set R in the equiripple filter design problem, consisting of the pass band

$[0, \omega_p]$ and the stopband $[\omega_p, \omega_s]$. The transition band (ω_p, ω_s) is a don't care region.

considered in the minimization of the weighted error. The solution to this optimization problem is given in the *alternation theorem*, which is as follows:

Alternation Theorem: Let $\mathcal{F}$ be a union of closed subsets over the interval $[0, \pi]$. For a positive weighting function $W(e^{j\omega})$, a necessary and sufficient condition for

$$A(e^{j\omega}) = \sum_{k=0}^{L} a(k)\cos(k\omega)$$

to be the unique function that minimizes the maximum value of the weighted error $|E(e^{j\omega})|$ over the set $\mathcal{F}$ is that the $E(e^{j\omega})$ have at least $L + 2$ *alternations*. That is to say, there must be at least $L + 2$ *extremal frequencies*,

$$\omega_0 < \omega_1 < \cdots < \omega_{L+1}$$

over the set $\mathcal{F}$ such that

$$E(e^{j\omega_k}) = -E(e^{j\omega_{k+1}}) \qquad k = 0, 1, \ldots, L$$

and

$$|E(e^{j\omega_k})| = \max_{\omega \in \mathcal{F}} |E(e^{j\omega})| \qquad k = 0, 1, \ldots, L + 1$$

Thus, the alternation theorem states that the optimum filter is equiripple. Although the alternation theorem specifies the minimum number of extremal frequencies (or ripples) that the optimum filter must have, it may have more. For example, a low-pass filter may have either $L + 2$ or $L + 3$ extremal frequencies. A low-pass filter with $L + 3$ extrema is called an *extraripple filter*.

From the alternation theorem, it follows that

$$W(e^{j\omega_k})[A_d(e^{j\omega_k}) - A(e^{j\omega_k})] = (-1)^k \epsilon \qquad k = 0, 1, \ldots, L+1$$

where
$$\epsilon = \pm \max_{\omega \in \mathcal{F}} |E(e^{j\omega})|$$

is the maximum absolute weighted error. These equations may be written in matrix form in terms of the unknowns $a(0), \ldots, a(L)$ and ϵ as follows:

$$\begin{bmatrix} 1 & \cos(\omega_0) & \cdots & \cos(L\omega_0) & 1/W(e^{j\omega_0}) \\ 1 & \cos(\omega_1) & \cdots & \cos(L\omega_1) & -1/W(e^{j\omega_1}) \\ \vdots & \vdots & \vdots & \vdots & \\ 1 & \cos(\omega_L) & \cdots & \cos(L\omega_L) & (-1)^L/W(e^{j\omega_L}) \\ 1 & \cos(\omega_{L+1}) & \cdots & \cos(L\omega_{L+1}) & (-1)^{L+1}/W(e^{j\omega_{L+1}}) \end{bmatrix} \begin{bmatrix} a(0) \\ a(1) \\ \vdots \\ a(L) \\ \epsilon \end{bmatrix} = \begin{bmatrix} A_d(e^{j\omega_0}) \\ A_d(e^{j\omega_1}) \\ \vdots \\ A_d(e^{j\omega_L}) \\ A_d(e^{j\omega_{L+1}}) \end{bmatrix}$$

Given the extremal frequencies, these equations may be solved for $a(0), \ldots, a(L)$ and ϵ. To find the extremal frequencies, there is an efficient iterative procedure known as the Parks-McClellan algorithm, which involves the following steps:

1. Guess an initial set of extremal frequencies.
2. Find ϵ by solving Eq. (9.5). The value of ϵ has been shown to be

$$\epsilon = \frac{\sum_{k=0}^{L+1} b(k) D(e^{j\omega_k})}{\sum_{k=0}^{L+1} (-1)^k b(k)/W(e^{j\omega_k})}$$

where $$b(k) = \prod_{i=1, i \neq k}^{L+1} \frac{1}{\cos(\omega_k) - \cos(\omega_i)}$$

3. Evaluate the weighted error function over the set $\mathcal{F}$ by interpolating between the extremal frequencies using the Lagrange interpolation formula.
4. Select a new set of extremal frequencies by choosing the $L + 2$ frequencies for which the interpolated error function is maximum.
5. If the extremal frequencies have changed, repeat the iteration from step 2.

A design formula that may be used to estimate the equiripple filter order for a low-pass filter with a transition width Δf, passband ripple δ_p, and stopband ripple δ_s is

$$N = \frac{-10\log(\delta_p \delta_s) - 13}{14.6\Delta f}$$

Example 4.3.3

Suppose that we would like to design an equiripple low-pass filter with a passband cutoff frequency

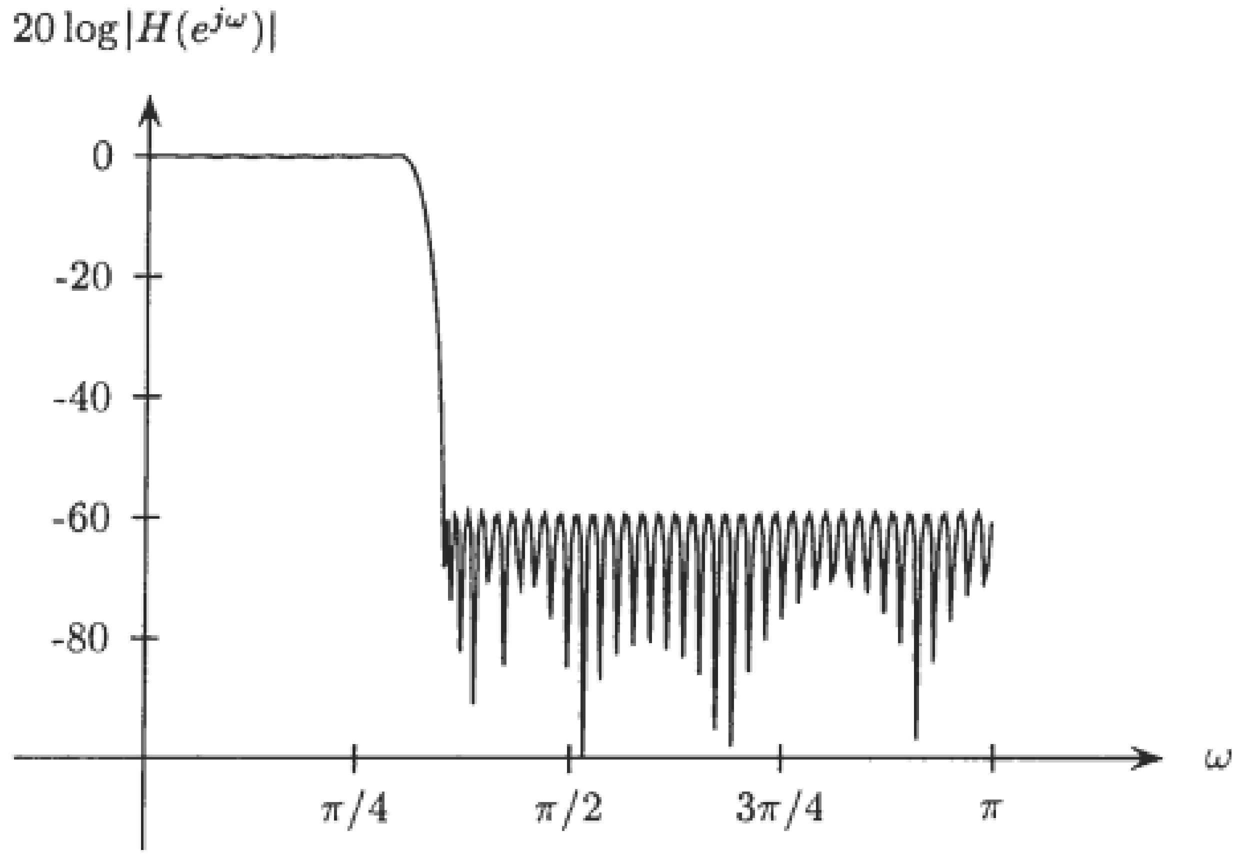
$\omega_p = 0.3\pi$, a stopband cutoff frequency $\omega_s = 0.35\pi$, a passband ripple of $\delta_p = 0.01$, and a stopband ripple of $\delta_s = 0.001$. we find

$$N = \frac{-10\log(\delta_p\delta_s) - 13}{14.6\Delta f} = 102$$

Because we want the ripple in the stopband to be 10 times smaller than the ripple in the passband, the error must be weighted using the weighting function

$$W(e^{j\omega}) = \begin{cases} 1 & 0 \le |\omega| \le 0.3\pi \\ 10 & 0.35\pi \le |\omega| \le \pi \end{cases}$$

Using the Parks-McClellan algorithm to design the filter, we obtain a filter with the frequency response magnitude shown below.



$20 \log|H(e^{j\omega})|$
0
-20
-40
-60
-80
$\pi/4$
$\pi/2$
$3\pi/4$
π
ω

4.4 IIR Filter Design

- There are two general approaches used to design IIR digital filters. The most common is to design an analog IIR filter and then map it into an equivalent digital filter because the art of analog filter design is highly advanced.
- Therefore, it is prudent to consider optimal ways for mapping these filters into the discrete-time domain.
- Furthermore, because there are **powerful design procedures** that facilitate the design of **analog filters**, this approach to IIR filter design is relatively simple.
- The **second** approach to design IIR digital filters is to use an algorithmic design procedure, which generally requires the use of a computer to solve a set of linear or nonlinear equations.

- These methods may be used to design digital filters with arbitrary frequency response characteristics for which no analog filter prototype exists or to design filters when other types of constraints are imposed on the design.
- In this section, we consider the approach of mapping analog filters into digital filters. Initially, the focus will be on the **design of digital low-pass filters from analog low-pass filters**.
- Techniques for transforming these designs into more general frequency selective filters will then be discussed.

4.4.1 Analog Low-Pass Filter Prototypes

- To design an IIR digital low-pass filter from an analog low-pass filter, we must first know how to design an analog low-pass filter.
- Most analog filter approximation methods were developed for the design of **passive systems** having a gain less than or equal to 1.

Therefore, a typical set of specifications for these filters is as shown in Fig. 4-5(a), with the pass band specifications having the form

$$1 - \delta_p \le |H_a(j\Omega)| \le 1$$

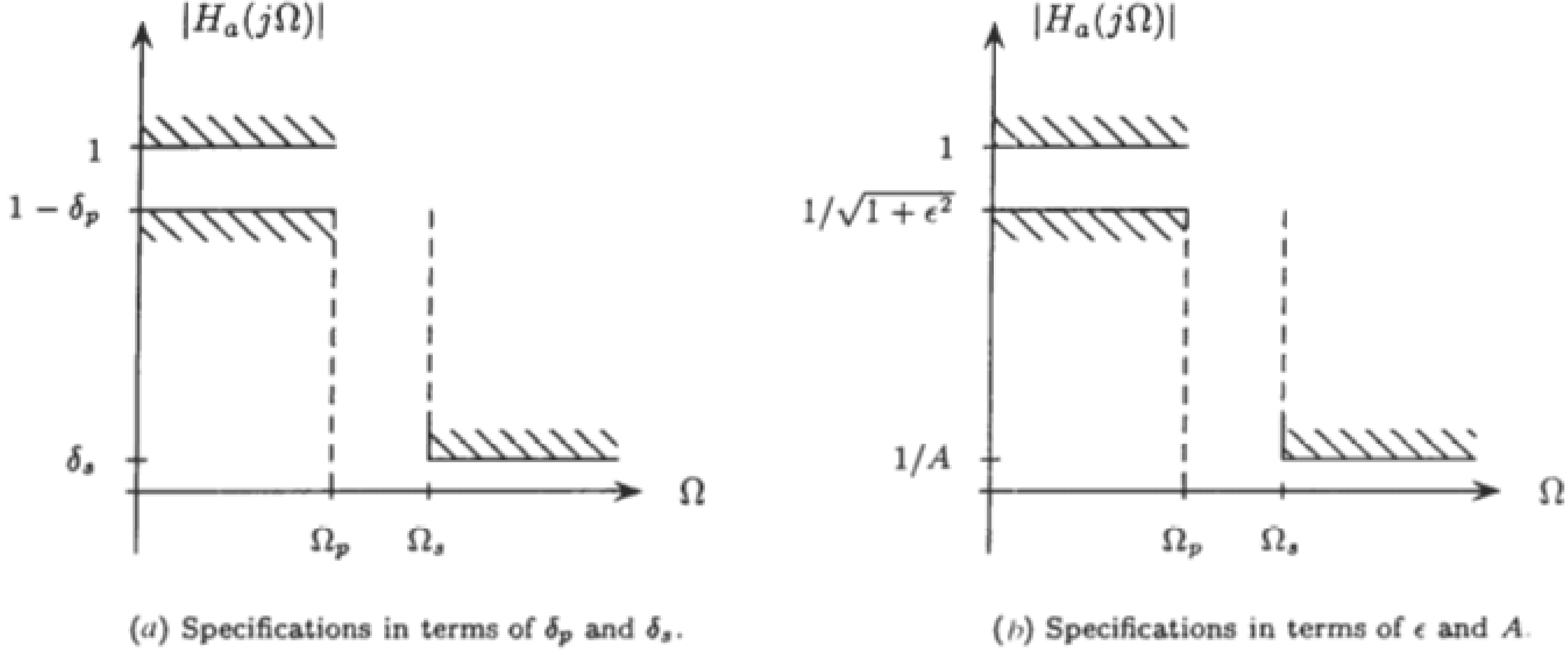


Fig. 4-5. Two different conventions for specifying the pass band and stop band deviations for an analog low-pass filter.

Another convention that is commonly used is to describe the passband and stopband constraints in terms of the parameters **E** and **A** as illustrated in Fig. 4-5(h). Two auxiliary parameters of interest are the discriminational factor

$$d = \left[\frac{(1-\delta_p)^{-2}-1}{\delta_s^{-2}-1}\right]^{1/2} = \frac{\epsilon}{\sqrt{A^2-1}}$$

and the *selectivity factor*

$$k = \frac{\Omega_p}{\Omega_s}$$

The three most commonly used analog low-pass filters are the Butterworth, Chebyshev, and elliptic filters. These filters are described below.

Butterworth Filter

A low-pass Butterworth filter is an all-pole filter with a squared magnitude response given by

$$|H_a(j\Omega)|^2 = \frac{1}{1+(j\Omega/j\Omega_c)^{2N}}$$

The parameter N is the order of the filter (number of poles in the system function), and Ω_c is the 3-dB cutoff frequency. The magnitude of the frequency response may also be written as

$$|H_a(j\Omega)|^2 = \frac{1}{1+\epsilon^2(j\Omega/j\Omega_p)^{2N}}$$

where

$$\epsilon = \left(\frac{\Omega_p}{\Omega_c}\right)^N$$

The frequency response of the Butterworth filter decreases monotonically with increasing Ω, and as the filter order increases, the transition band becomes narrower. These properties are illustrated in Fig. 9-6, which shows $|H_a(j\Omega)|$ for Butterworth filters of orders $N = 2, 4, 8$, and 12. Because

$$|H_a(j\Omega)|^2 = H_a(s)H_a(-s)\big|_{s=j\Omega}$$

from the magnitude-squared function, we may write

$$G_a(s) = H_a(s)H_a(-s) = \frac{1}{1+(s/j\Omega_c)^{2N}}$$

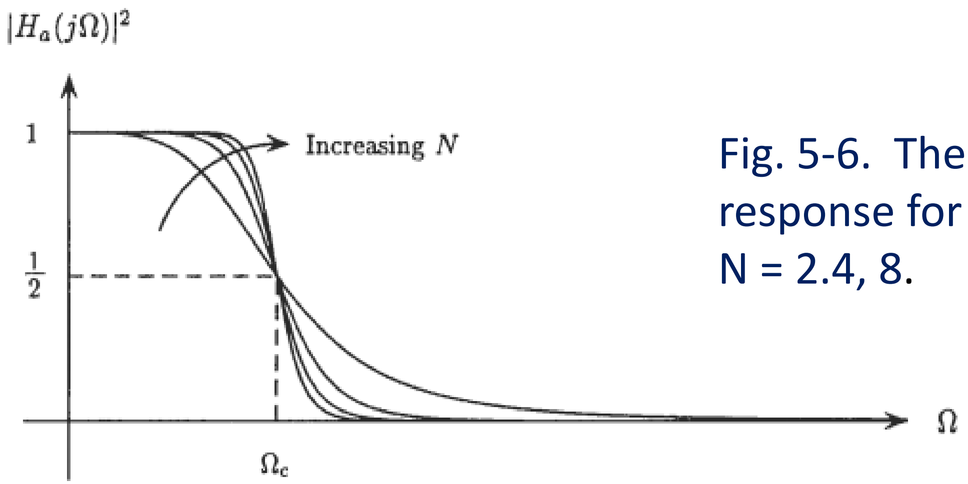


Fig. 5-6. The magnitude of the frequency response for Butterworth filters of orders N = 2.4, 8.

Therefore, the poles of $G_a(s)$ are located at $2N$ equally spaced points around a circle of radius Ω_c,

$$s_k = (-1)^{1/2N}(j\Omega_c) = \Omega_c \exp\left\{ j\frac{(N+1+2k)\pi}{2N} \right\} \qquad k = 0, 1, \dots, 2N-1$$

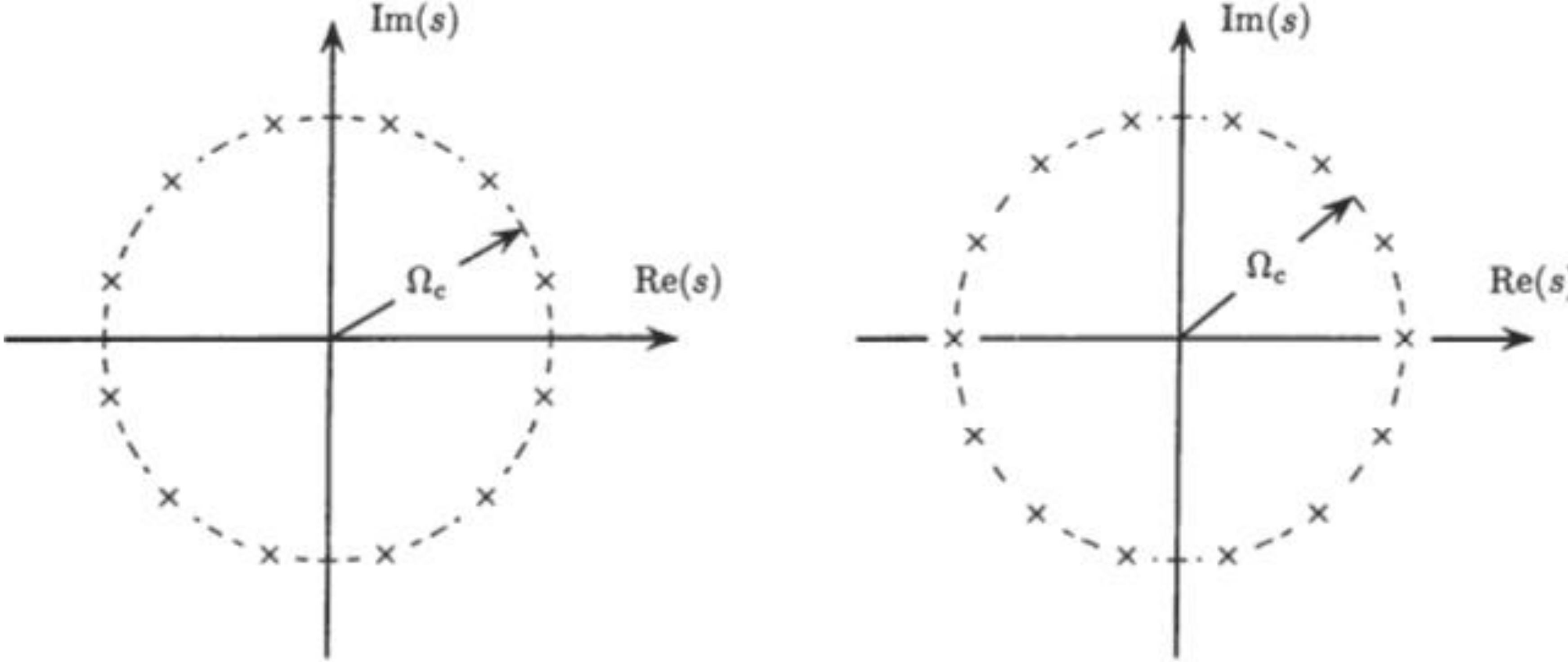
and are symmetrically located about the $j\Omega$-axis. Figure 9-7 shows these pole positions for $N = 6$ and $N = 7$. The system function, $H_a(s)$, is then formed from the N roots of $H_a(s)H_a(-s)$ that lie in the left-half s-plane. For a *normalized* Butterworth filter with $\Omega_c = 1$, the system function has the form

$$H_a(s) = \frac{1}{A_N(s)} = \frac{1}{s^N + a_1 s^{N-1} + \cdots + a_{N-1}s + a_N}$$

Table 5-4 lists the coefficients of $A_N(s)$ for $1 \le N \le 8$. Given Ω_p, Ω_s, δ_p, and δ_s, the steps involved in designing a Butterworth filter are as follows:

1. Find the values for the selectivity factor, k, and the discrimination factor, d, from the filter specifications.
2. Determine the order of the filter required to meet the specifications using the design formula

$$N \ge \frac{\log d}{\log k}$$



(a) Order $N = 6$.　　　　(b) Order $N = 7$.

Fig. 5.7. The poles of $H_a(a)H_a(-s)$ for a Butterworth filter of order $N = 6$ and $N = 7$.

Table 5.4 The Coefficients in the System Function of a Normalized Butterworth Filter ($\Omega_c = 1$) for Orders $1 \le N \le 8$

N	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
1	1.0000							
2	1.4142	1.0000						
3	2.0000	2.0000	1.0000					
4	2.6131	3.4142	2.6131	1.0000				
5	3.2361	5.2361	5.2361	3.2361	1.0000			
6	3.8637	7.4641	9.1416	7.4641	3.8637	1.0000		
7	4.4940	10.0978	14.5918	14.5918	10.0978	4.4940	1.0000	
8	5.1258	13.1371	21.8462	25.6884	21.8462	13.1372	5.1258	1.0000

3. Set the 3-dB cutoff frequency, Ω_c, to any value in the range

$$\Omega_p[(1-\delta_p)^{-2}-1]^{-1/2N} \le \Omega_c \le \Omega_s\left[\delta_s^{-2}-1\right]^{-1/2N}$$

4. Synthesize the system function of the Butterworth filter from the poles of

$$G_a(s) = H_a(s)H_a(-s) = \frac{1}{1+(s/j\Omega_c)^{2N}}$$

that lie in the left-half s-plane. Thus,

$$H_a(s) = \prod_{k=0}^{N-1} \frac{-s_k}{s-s_k}$$

where $$s_k = \Omega_c \exp\left\{ j\frac{(N+1+2k)\pi}{2N} \right\} \qquad k = 0, 1, \ldots, N-1$$

Example Let us design a low-pass Butterworth filter to meet the following specifications:

$$f_p = 6\text{ kHz} \qquad f_s = 10\text{ kHz} \qquad \delta_p = \delta_s = 0.1$$

First, we compute the discrimination and selectivity factors:

$$d = \left[\frac{(1-\delta_p)^{-2} - 1}{\delta_s^{-2} - 1}\right]^{1/2} = 0.0487 \qquad k = \frac{\Omega_p}{\Omega_s} = \frac{f_p}{f_s} = 0.6$$

Because

$$N \geq \frac{\log d}{\log k} = 5.92$$

it follows that the minimum filter order is $N = 6$. With

$$f_p[(1-\delta_p)^{-2} - 1]^{-1/2N} = 6770$$

and

$$f_s\left[\delta_s^{-2} - 1\right]^{-1/2N} = 6819$$

the center frequency, f_c, may be any value in the range

$$6770 \leq f_c \leq 6819$$

$$H_a(s) = \frac{1}{s^6 + 3.8637s^5 + 7.4641s^4 + 9.1416s^3 + 7.4641s^2 + 3.8637s + 1}$$

and then replacing s with s/Ω_c so that the cutoff frequency is Ω_c instead of unity

4.4.2 Design of IIR Filters from Analog Filters

The design of a digital filter from an analog prototype requires that we transform $h_a(t)$ to $h(n)$ or $H_a(s)$ to $H(z)$. A mapping from the s-plane to the z-plane may be written as

$$H(z) = H_a(s)\Big|_{s=m(z)}$$

where $s = m(z)$ is the mapping function. In order for this transformation to produce an acceptable digital filter, the mapping $m(z)$ should have the following properties:

1. The mapping from the $j\Omega$-axis to the unit circle, $|z| = 1$, should be one to one and *onto* the unit circle in order to preserve the frequency response characteristics of the analog filter.
2. Points in the left-half s-plane should map to points *inside* the unit circle to preserve the stability of the analog filter.
3. The mapping $m(z)$ should be a rational function of z so that a rational $H_a(s)$ is mapped to a rational $H(z)$.

Described below are two approaches that are commonly used to map analog filters into digital filters.

Impulse Invariance

With the *impulse invariance* method, a digital filter is designed by sampling the impulse response of an analog filter:

$$h(n) = h_a(nT_s)$$

From the sampling theorem, it follows that the frequency response of the digital filter, $H(e^{j\omega})$, is related to the frequency response $H_a(j\Omega)$ of the analog filter as follows:

$$H(e^{j\omega}) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} H_a\left(j\frac{\omega}{T_s} + j\frac{2\pi k}{T_s} \right)$$

More generally, this may be extended into the complex plane as follows:

$$H(z)\Big|_{z=e^{sT_s}} = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} H_a\left(s + j\frac{2\pi k}{T_s} \right)$$

The mapping between the s-plane and the z-plane is illustrated in Fig. 5-11. Note that although the $j\Omega$-axis maps *onto* the unit circle, the mapping is not one to one. In particular, each interval of length $2\pi/T_s$ along the $j\Omega$-axis is mapped onto the unit circle (i.e., the frequency response is aliased). In addition, each point in the left-half s-plane is mapped to a point *inside* the unit circle. Specifically, strips of width $2\pi/T_s$ map *onto* the z-plane. If the frequency response of the analog filter, $H_a(j\Omega)$, is sufficiently bandlimited, then

$$H(e^{j\omega}) \approx \frac{1}{T_s} H_a\left(\frac{j\omega}{T_s}\right)$$

Although the impulse invariance may produce a reasonable design in some cases, this technique is essentially limited to bandlimited analog filters.

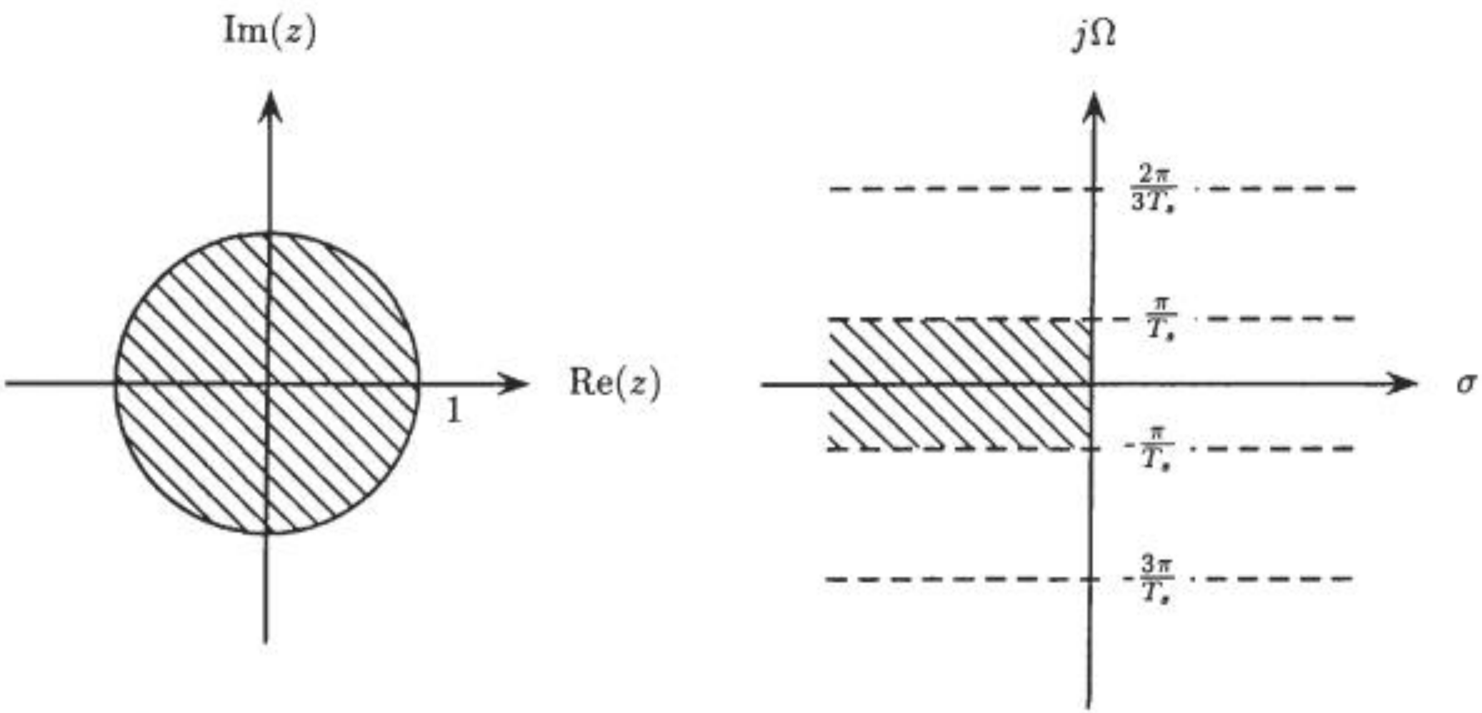


Fig. 5 11. Properties of the s-plane to z-plane mapping in the impulse invariance method.

To see how poles and zeros of an analog filter are mapped using the impulse invariance method, consider an analog filter that has a system function

$$H_a(s) = \sum_{k=1}^{p} \frac{A_k}{s - s_k}$$

The impulse response, $h_a(t)$, is

$$h_a(t) = \sum_{k=1}^{p} A_k e^{s_k t} u(t)$$

Therefore, the digital filter that is formed using the impulse invariance technique is

$$h(n) = h_a(nT_s) = \sum_{k=1}^{p} A_k e^{s_k n T_s} u(nT_s) = \sum_{k=1}^{p} A_k (e^{s_k T_s})^n u(n)$$

and the system function is

$$H(z) = \sum_{k=1}^{p} \frac{A_k}{1 - e^{s_k T_s} z^{-1}}$$

Thus, a pole at $s = s_k$ in the analog filter is mapped to a pole at $z = e^{s_k T_s}$ in the digital filter,

$$\frac{1}{s - s_k} \Longrightarrow \frac{1}{1 - e^{s_k T_s} z^{-1}}$$

The zeros, however, do not get mapped in any obvious way.

The Bilinear Transformation

The bilinear transformation is a mapping from the s-plane to the z-plane defined by

$$s = \frac{2}{T_s} \frac{1 - z^{-1}}{1 + z^{-1}}$$

Given an analog filter with a system function $H_a(s)$, the digital filter is designed as follows:

$$H(z) = H_a\left(\frac{2}{T_s} \frac{1 - z^{-1}}{1 + z^{-1}}\right)$$

The bilinear transformation is a rational function that maps the left-half s-plane *inside* the unit circle and maps the $j\Omega$-axis in a one-to-one manner *onto* the unit circle. However, the relationship between the $j\Omega$-axis and the unit circle is highly nonlinear and is given by the *frequency warping function*.

$$\omega = 2\arctan\left(\frac{\Omega T_s}{2}\right)$$

As a result of this warping, the bilinear transformation will only preserve the magnitude response of analog filters that have an ideal response that is piecewise constant. Therefore, the bilinear transformation is generally only used in the design of frequency selective filters.

The parameter T_s in the bilinear transformation is normally included for historical reasons. However, it does not enter into the design process, because it only scales the $j\Omega$-axis in the frequency warping function, and this scaling may be done in the specification of the analog filter. Therefore, T_s may be set to any value to simplify the design procedure. The steps involved in the design of a digital low-pass filter with a passband cutoff frequency ω_p, stopband cutoff frequency ω_s, passband ripple δ_p, and stopband ripple δ_s are as follows:

1. *Prewarp* the passband and stopband cutoff frequencies of the digital filter, ω_p and ω_s, using the inverse of Eq. (*9.12*) to determine the passband and cutoff frequencies of the analog low-pass filter. With $T_s = 2$, the prewarping function is

$$\Omega = \tan\left(\frac{\omega}{2}\right)$$

2. Design an analog low-pass filter with the cutoff frequencies found in step 1 and a passband and stopband ripple δ_p and δ_s, respectively.
3. Apply the bilinear transformation to the filter designed in step 2.

Example Let us design a first-order digital low-pass filter with a 3-dB cutoff frequency of $\omega_c = 0.25\pi$ by applying the bilinear transformation to the analog Butterworth filter

$$H_a(s) = \frac{1}{1 + s/\Omega_c}$$

Because the 3-dB cutoff frequency of the Butterworth filter is Ω_c, for a cutoff frequency $\omega_c = 0.25\pi$ in the digital filter, we must have

$$\Omega_c = \frac{2}{T_s}\tan\left(\frac{0.25\pi}{2}\right) = \frac{0.828}{T_s}$$

Therefore, the system function of the analog filter is

$$H_a(s) = \frac{1}{1 + sT_s/0.828}$$

Applying the bilinear transformation to the analog filter gives

$$H(z) = H_a(s)\Big|_{s=\frac{2}{T_s}\frac{1-z^{-1}}{1+z^{-1}}} = \frac{1}{1 + (2/0.828)[(1-z^{-1})/(1+z^{-1})]} = 0.2920\frac{1+z^{-1}}{1-0.4159z^{-1}}$$

Note that the parameter T_s does not enter into the design.

Table 5.5 The Transformation of an Analog Low-pass Filter with a 3-dB Cutoff Frequency Ω_p to Other Frequency Selective Filters

Transformation	Mapping	New Cutoff Frequencies
Low-pass	$s \rightarrow \frac{\Omega_p}{\Omega'_p} s$	Ω'_p
High-pass	$s \rightarrow \frac{\Omega_p \Omega'_p}{s}$	Ω'_p
Bandpass	$s \rightarrow \Omega_p \frac{s^2 + \Omega_l \Omega_u}{s(\Omega_u - \Omega_l)}$	Ω_l, Ω_u
Bandstop	$s \rightarrow \Omega_p \frac{s(\Omega_u - \Omega_l)}{s^2 + \Omega_l \Omega_u}$	Ω_l, Ω_u

4.5 Filter design based on a least squares approach

The design techniques described in the previous section are based on converting an analog filter into a digital filter. It is also possible to perform the design directly in the time domain without any reference to an analog filter. This section describes several methods for designing a digital filter directly.

Table 5.6 The Transformation of a Digital Low-Pass Filter with a Cutoff Frequency ω_c to Other Frequency Selective Filters

Filter Type	Mapping	Design Parameters
Low-pass	$z^{-1} \rightarrow \frac{z^{-1}-\alpha}{1-\alpha z^{-1}}$	$\alpha = \frac{\sin[(\omega_c - \omega_c')/2]}{\sin[(\omega_c + \omega_c')/2]}$ ω_c' = desired cutoff frequency
High-pass	$z^{-1} \rightarrow -\frac{z^{-1}+\alpha}{1+\alpha z^{-1}}$	$\alpha = -\frac{\cos[(\omega_c + \omega_c')/2]}{\cos[(\omega_c - \omega_c')/2]}$ ω_c' = desired cutoff frequency
Bandpass	$z^{-1} \rightarrow -\frac{z^{-2} - [2\alpha\beta/(\beta+1)]z^{-1} + [(\beta-1)/(\beta+1)]}{[(\beta-1)/(\beta+1)]z^{-2} - [2\alpha\beta/(\beta+1)]z^{-1} + 1}$	$\alpha = \frac{\cos[(\omega_{c2} + \omega_{c1})/2]}{\cos[(\omega_{c2} - \omega_{c1})/2]}$ $\beta = \cot[(\omega_{c2} - \omega_{c1})/2]\tan(\omega_c/2)$ ω_{c1} = desired lower cutoff frequency ω_{c2} = desired upper cutoff frequency
Bandstop	$z^{-1} \rightarrow \frac{z^{-2} - [2\alpha/(\beta+1)]z^{-1} + [(1-\beta)/(1+\beta)]}{[(1-\beta)/(1+\beta)]z^{-2} - [2\alpha/(\beta+1)]z^{-1} + 1}$	$\alpha = \frac{\cos[(\omega_{c1} + \omega_{c2})/2]}{\cos[(\omega_{c1} - \omega_{c2})/2]}$ $\beta = \tan[(\omega_{c2} - \omega_{c1})/2]\tan(\omega_c/2)$ ω_{c1} = desired lower cutoff frequency ω_{c2} = desired upper cutoff frequency

4.5.1 FIR Least-Squares Inverse

The inverse of a linear shift-invariant system with unit sample response $g(n)$ and system function $G(z)$ is the system that has a unit sample response, $h(n)$, such that

$$h(n) * g(n) = \delta(n)$$

or

$$H(z)G(z) = 1$$

In most applications, the system function $H(z) = 1/G(z)$ is not a viable solution. One of the reasons is that, unless $G(z)$ is minimum phase, $1/G(z)$ cannot be both causal and stable. Another consideration comes from the fact that, in some applications, it may be necessary to constrain $H(z)$ to be an FIR filter. Because $1/G(z)$ will be infinite in length unless $G(z)$ is an all-pole filter, constraining $h(n)$ to be FIR would only be an approximation to the inverse filter.

In the FIR least-squares inverse filter design problem, the goal is to find the FIR filter $h(n)$ of length N such that

$$h(n) * g(n) \approx \delta(n)$$

The filter that minimizes the squared error

$$\mathcal{E} = \sum_{n=0}^{\infty} |e(n)|^2$$

where $$e(n) = \delta(n) - h(n) * g(n) = \delta(n) - \sum_{l=0}^{N-1} h(l)g(n-l)$$

may be found by solving the linear equations

$$\sum_{l=0}^{N-1} h(l) r_g(k-l) = \begin{cases} g(0) & k = 0 \\ 0 & k = 1, 2, \ldots, N-1 \end{cases}$$

where $$r_g(k) = \sum_{n=0}^{\infty} g(n)g(n-k)$$

In many cases, constraining the least-squares inverse filter to minimize the difference between $h(n) * g(n)$ and $\delta(n)$ is overly restrictive. For example, if a delay may be tolerated, we may consider finding the filter $h(n)$ so that

$$h(n) * g(n) \approx \delta(n - n_0)$$

for some delay n_0. In most cases, a nonzero delay will produce a better approximate inverse filter and, in many cases, the improvement will be substantial. The least-squares inverse filter with delay is found by solving the linear equations

$$\sum_{l=0}^{N-1} h(l) r_g(k-l) = \begin{cases} g(n_0 - k) & k = 0, 1, \ldots, n_0 \\ 0 & k = n_0 + 1, \ldots, N \end{cases}$$

Examples

1 Use the window design method to design a linear phase FIR filter of order $N = 24$ to approximate the following ideal frequency response magnitude:

$$|H_d(e^{j\omega})| = \begin{cases} 1 & |\omega| \leq 0.2\pi \\ 0 & 0.2\pi < |\omega| \leq \pi \end{cases}$$

The ideal filter that we would like to approximate is a low-pass filter with a cutoff frequency $\omega_p = 0.2\pi$. With $N = 24$, the frequency response of the filter that is to be designed has the form

$$H(e^{j\omega}) = \sum_{n=0}^{24} h(n)e^{-jn\omega}$$

Therefore, the delay of $h(n)$ is $\alpha = N/2 = 12$, and the ideal unit sample response that is to be windowed is

$$h_d(n) = \frac{\sin[0.2\pi(n-12)]}{(n-12)\pi}$$

All that is left to do in the design is to select a window. With the length of the window fixed, there is a trade-off between the width of the transition band and the amplitude of the passband and stopband ripple. With a rectangular window, which provides the smallest transition band,

$$\Delta\omega = 2\pi \cdot \frac{0.9}{24} = 0.075\pi$$

and the filter is

$$h(n) = \begin{cases} \dfrac{\sin[0.2\pi(n-12)]}{(n-12)\pi} & 0 \leq n \leq 24 \\ 0 & \text{otherwise} \end{cases}$$

However, the stopband attenuation is only 21 dB, which is equivalent to a ripple of $\delta_s = 0.089$. With a Hamming window, on the other hand,

$$h(n) = \left[0.54 - 0.46\cos\left(\frac{2\pi n}{24}\right)\right] \cdot \frac{\sin[0.2\pi(n-12)]}{(n-12)\pi} \qquad 0 \le n \le 24$$

and the stopband attenuation is 53 dB, or $\delta_s = 0.0022$. However, the width of the transition band increases to

$$\Delta\omega = 2\pi \cdot \frac{3.3}{24} = 0.275\pi$$

which, for most designs, would be too wide.

2 Use the window design method to design a minimum-order high-pass filter with a stopband cutoff frequency $\omega_s = 0.22\pi$, a passband cutoff frequency $\omega_p = 0.28\pi$, and a stopband ripple $\delta_s = 0.003$.

A stopband ripple of $\delta_s = 0.003$ corresponds to a stopband attenuation of $\alpha_s = -20\log\delta_s = 50.46$. For the minimum-order filter, we use a Kaiser window with

$$\beta = 0.1102(\alpha_s - 8.7) = 4.6$$

Because the transition width is $\Delta\omega = 0.06\pi$, or $\Delta f = 0.03$, the required window length is

$$N = \frac{\alpha_s - 7.95}{14.36\Delta f} = 98.67$$

Rounding this up to $N = 99$ results in a type II linear phase filter, which will have a zero in its system function at $z = -1$. Because this produces a null in the frequency response at $\omega = \pi$, this is not acceptable. Therefore, we increase the order by 1 to obtain a type I linear phase filter with $N = 100$.

In order to have a transition band that extends from $\omega_s = 0.22\pi$ to $\omega_p = 0.28\pi$, we set the cutoff frequency of the ideal high-pass filter equal to the midpoint:

$$\omega_c = \frac{\omega_p + \omega_s}{2} = 0.25\pi$$

The unit sample response of an ideal zero-phase high-pass filter with a cutoff frequency $\omega_c = 0.25\pi$ is

$$h_{\rm hp}(n) = \delta(n) - \frac{\sin(0.25\pi n)}{n\pi}$$

where the second term is a low-pass filter with a cutoff frequency $\omega_c = 0.25\pi$. Delaying $h_{\rm hp}(n)$ by $N/2 = 50$, we have

$$h_d(n) = \delta(n-50) - \frac{\sin[0.25\pi(n-50)]}{(n-50)\pi}$$

and the resulting FIR high-pass filter is

$$h(n) = h_d(n) \cdot w(n)$$

where $w(n)$ is a Kaiser window with $N = 100$ and $\beta = 4.6$.

3 As the order of an analog Butterworth filter is increased, the slope of $|H_a(j\Omega)|^2$ at the 3-dB cutoff frequency, Ω_c, increases. Derive an expression for the slope of $|H_a(j\Omega)|^2$ at Ω_c as a function of the filter order, N.

The magnitude squared of the Butterworth filter's frequency response is

$$|H_a(j\Omega)|^2 = \frac{1}{1 + (j\Omega/j\Omega_c)^{2N}}$$

To evaluate the slope of $|H_a(j\Omega)|^2$ at $\Omega = \Omega_c$, we may set $\Omega_c = 1$ and evaluate the derivative at $\Omega = 1$. Therefore, with

$$|H_a(j\Omega)|^2 = \frac{1}{1 + \Omega^{2N}}$$

we have

$$\frac{d}{d\Omega}|H_a(j\Omega)|^2 = \frac{-2N\Omega^{2N-1}}{(1 + \Omega^{2N})^2}$$

Evaluating this at $\Omega = 1$, we have

$$\frac{d}{d\Omega}|H_a(j\Omega)|^2\bigg|_{\Omega=1} = -\frac{N}{2}$$

4 Design a low-pass Butterworth filter that has a 3-dB cutoff frequency of 1.5 kHz and an attenuation of 40 dB at 3.0 kHz.

Given the 3-dB cutoff frequency of the Butterworth filter, all that is needed is to find the filter order, N, that will give 40 dB of attenuation at 3 kHz, or $\Omega_s = 2\pi \cdot 3000$. At the stopband cutoff frequency Ω_s, the magnitude of the frequency response squared is

$$|H_a(j\Omega)|^2_{\Omega=2\pi\cdot 3000} = \left.\frac{1}{1+(j\Omega/j\Omega_c)^{2N}}\right|_{\Omega=2\pi\cdot 3000} = \frac{1}{1+2^{2N}}$$

Therefore, if we want the magnitude of the frequency response to be down 40 dB at $\Omega_s = 2\pi \cdot 3000$, the magnitude *squared* must be no larger than 10^{-4}, or

$$\frac{1}{1+2^{2N}} \leq 10^{-4}$$

Thus, we want

$$2N = \frac{\log(10^4 - 1)}{\log 2} = 13.29$$

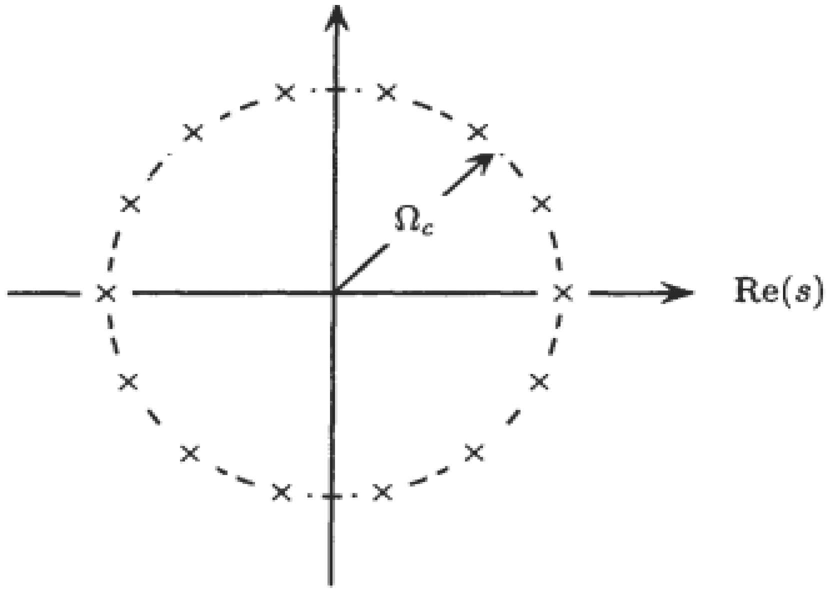
or $N = 7$. For a seventh-order Butterworth filter, the 14 poles of

$$H_a(s)H_a(-s) = \frac{1}{1+(s/j\Omega_c)^{2N}}$$

lie on a circle of radius $\Omega_c = 2\pi \cdot 3000$, at angles of

$$\theta_k = \frac{(N+1+2k)\pi}{N} = \frac{(4+k)\pi}{7} \qquad k = 0, 1, \ldots, 13$$

as illustrated in the following figure:



The poles of $H_a(s)$ are the seven poles of $H_a(s)H_a(-s)$ that lie in the left-half s-plane, that is,

$$s_k = -\Omega_c e^{\pm jk\pi/7} \qquad k = 0, 1, 2, 3$$

Except for the isolated pole at $s = -\Omega_c$, the remaining six poles occur in complex conjugate pairs. The conjugate pairs may be combined to form second-order factors with real coefficients to yield factors of the form

$$H_k(s) = \frac{1}{s^2 - 2\Omega_c \cos(k\pi/7)s + \Omega_c^2} \qquad k = 1, 2, 3$$

Thus, the system function of the seventh-order Butterworth filter is

$$H_a(s) = \prod_{k=0}^{N-1} \frac{-s_k}{s - s_k} = \frac{\Omega_c}{s + \Omega_c} \cdot \prod_{k=1}^{3} \frac{\Omega_c^2}{s^2 - 2\Omega_c \cos(k\pi/7)s + \Omega_c^2}$$