

4.2.6. Design of Equiripple Linear-Phase FIR Digital Filter

The windowing method and the frequency-sampling method are relatively simple techniques for designing linear-phase FIR filter. However, they also possess some minor disadvantages. A major problem, is a lack of precise control of the critical frequencies such cut off frequencies of pass-band and stop-band.

The filter design method described in this section is formulated as a Chebyshev approximation problem. It is viewed as an optimum design criterion in the sense that the weighted approximation error between the desired frequency response and the actual frequency response is spread evenly across the pass-band and evenly across the stop-band of the filter and the maximum error is minimized. The resulting filter designs have ripples in both the pass-band and the stop-band.

To describe the design procedure, let us consider the design of a low-pass filter with pass-band cut off frequency (edge frequency) ω_p and stop-band band cut off frequency (edge frequency) ω_s . From the general specification given in the next figure, in the pass-band, the filter frequency response satisfies the condition

$$1 - \delta_1 \leq H_r(\omega) \leq 1 + \delta_1 \quad \text{for } |\omega| \leq \omega_p$$

Similarly, in the stop-band, the filter response is specified to fall between the limits $\pm\delta_2$, that is

$$-\delta_2 \leq H_r(e^{j\omega}) \leq \delta_2 \quad \text{for } |\omega| > \omega_s$$

Thus δ_1 represents the ripple in the pass-band and δ_2 represents the attenuation or ripple in the stop-band. The remaining filter parameter is M , the filter length or the number of coefficients. Then the filter is described by

$$y(n) = \sum_{k=0}^{M-1} h(k)x(n-k)$$

As we recall, there are four different cases that result in a linear phase FIR filter. They are summarized below.

Case 1: Symmetric unit sample response and $h(n)=h(M-1-n)$ and M odd.

Case 2: Symmetric unit sample response and $h(n)=h(M-1-n)$ and M even.

Case 3: Antisymmetric unit sample response and $h(n)=-h(M-1-n)$ and M odd.

Case 4: Antisymmetric unit sample response and $h(n)=-h(M-1-n)$ and M even.

It can be shown (see e.g. references), that $H_r(\omega)$ can be in these four cases expressed in the form,

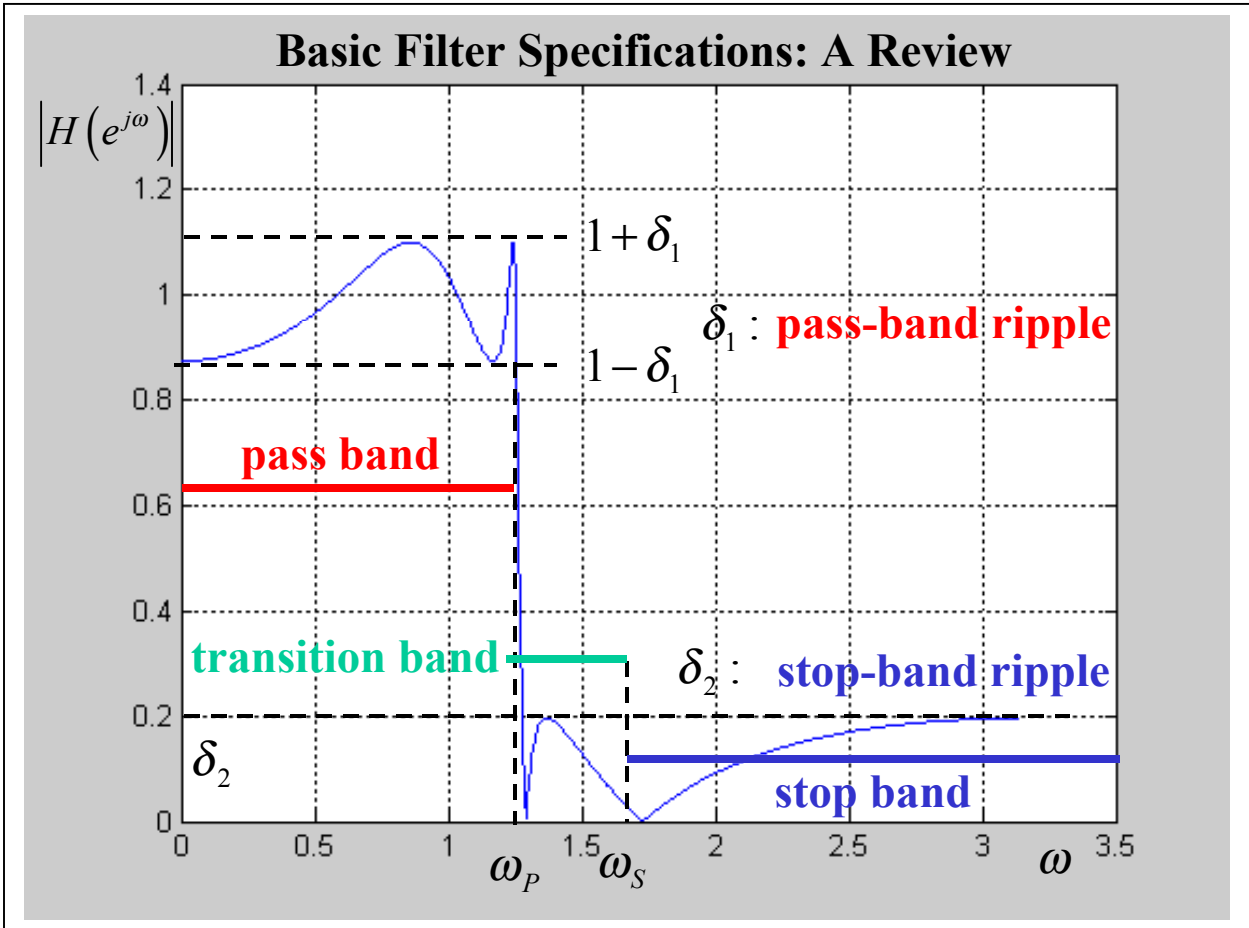
$$H_r(e^{j\omega}) = Q(\omega)P(\omega)$$

where

$$P(\omega) = \sum_{k=0}^L \alpha(k) \cos \omega k$$

and L is given in the next Table. The $\alpha(k)$ representing the parameters of the filter, which are linearly related to the unit sample response $h(n)$.

Filter Type	$Q(\omega)$	L
Case 1	1	(M-1)/2
Case 2	$\cos \frac{\omega}{2}$	(M/2)-1
Case 3	$\sin \frac{\omega}{2}$	(M-3)/2
Case 4	$\sin \frac{\omega}{2}$	(M/2)-1



In addition to the common framework given above for the representation of $H_r(\omega)$, we also define the real-valued desired (ideal) frequency response $H_{dr}(\omega)$ and the weighting function $W(\omega)$ on the estimation error. The real-valued desired (ideal) frequency response $H_{dr}(\omega)$ is simply defined to be unity in the pass-band and zero in the stop-band i.e.

$$H_{dr}(\omega) = \begin{cases} 1 & \text{for } \omega \text{ in pass-band} \\ 0 & \text{for } \omega \text{ in stop-band} \end{cases}$$

The weighting function on the approximation error allows us to choose the relatively size of the errors in the different frequency bands (i.e. in the pass-band and in the stop-band). In particular, it is convenient to normalize $W(\omega)$ to the unity in the stop-band and set $W(\omega) = \delta_2 / \delta_1$ in the pass-band, i.e.

$$W(\omega) = \begin{cases} \delta_2 / \delta_1 & \text{for } \omega \text{ in pass-band} \\ 1 & \text{for } \omega \text{ in stop-band} \end{cases}$$

Then we simply select $W(\omega)$ in the pass-band to reflect our emphasis on the relative size of the ripple in the stop-band to the ripple in the pass-band.

With the specification of $H_{dr}(\omega)$ and $W(\omega)$, we can now define the weighted approximation error as

$$E(\omega) = W(\omega)[H_{dr}(\omega) - H_r(\omega)] = W(\omega)[H_{dr}(\omega) - Q(\omega)P(\omega)] = W(\omega)Q(\omega)\left[\frac{H_{dr}(\omega)}{Q(\omega)} - P(\omega)\right]$$

For mathematical convenience, we define a modified weighting function $\widehat{W}(\omega)$ and the modified desired frequency response $\widehat{H}_{dr}(\omega)$ as

$$\widehat{W}(\omega) = W(\omega)Q(\omega)$$

$$\widehat{H}_{dr}(\omega) = \frac{H_{dr}(\omega)}{Q(\omega)}$$

Then the weighted approximation error may be expressed as

$$E(\omega) = \widehat{W}(\omega)\left[\widehat{H}_{dr}(\omega) - P(\omega)\right]$$

for all different types of linear-phase FIR filter.

Given the error function $E(\omega)$, the Chebyshev approximation problem is basically to determine the filter parameter $\alpha(k)$, that minimize the maximum absolute value of $E(\omega)$ over the frequency bands in which the approximation is to be performed. In mathematical terms we seek the solution to the problem

$$\begin{aligned} \min_{\text{over } \{\alpha(k)\}} \left[\max_{\omega \in S} |E(\omega)| \right] &= \min_{\text{over } \{\alpha(k)\}} \left[\max_{\omega \in S} \left| \widehat{W}(\omega) \left[\widehat{H}_{dr}(\omega) - P(\omega) \right] \right| \right] = \\ &= \min_{\text{over } \{\alpha(k)\}} \left[\max_{\omega \in S} \left| \widehat{W}(\omega) \left[\widehat{H}_{dr}(\omega) - P(\omega) = \sum_{k=0}^L \alpha(k) \cos \omega k \right] \right| \right] \end{aligned}$$

where S represents the set (disjoint union) of frequency bands over which the optimization is to be performed. Basically, the set S consists of the pass-bands and stop-bands of the desired filter. The criterion leads to an equiripple filter, that is, a filter whose magnitude response (or $H_r(\omega)$) oscillates uniformly between the tolerance bounds of each band.

The solution to this problem is due to Parks and McClellan, who applied the alternation theorem in the theory of Chebyshev approximation in combination with the Remez exchange algorithm. As a result of this work, a program for designing the FIR filter has been reported by Parks and McClellan. This program represented by sophisticated software tools (e.g., in the Signal Processing Toolbox of MATLAB, m-file `remez.m`: Parks-McClellan optimal equiripple FIR filter design) can be applied for designing linear phase FIR filter based on the Chebyshev approximation criterion and implemented with the Remez exchange algorithm. This program may be used to design low-pass, band-pass, high-pass or stop-band filters, differentiators and Hilbert transformers.

4.2.6.1. MATLAB function for optimal equiripple FIR filter design: `b=remez(N,F,A)`

REMEZ Parks-McClellan optimal equiripple FIR filter design. `b=remez(N,F,A)` returns a length $N+1$ linear phase (real, symmetric coefficients) FIR filter which has the best approximation to the desired frequency response described by F and A in the minimax sense. F is a vector of frequency band edges in pairs, in ascending order between 0 and 1. 1 corresponds to the Nyquist frequency or half the sampling frequency. A is a real vector the same size as F which specifies the desired amplitude of the frequency response of the resultant filter B . The desired

response is the line connecting the points $(F(k),A(k))$ and $(F(k+1),A(k+1))$ for odd k ; REMEZ treats the bands between $F(k+1)$ and $F(k+2)$ for odd k as "transition bands" or "don't care" regions. Thus the desired amplitude is piecewise linear with transition bands. The maximum error is minimized.

For filters with a gain other than zero at $F_s/2$, e.g., highpass and bandstop filters, N must be even. Otherwise, N will be incremented by one.

$B=\text{REMEZ}(N,F,A,W)$ uses the weights in W to weight the error. W has one entry per band (so it is half the length of F and A) which tells REMEZ how much emphasis to put on minimizing the error in each band relative to the other bands.

$B=\text{REMEZ}(N,F,A,\text{'Hilbert'})$ and $B=\text{REMEZ}(N,F,A,W,\text{'Hilbert'})$ design filters that have odd symmetry, that is, $B(k) = -B(N+2-k)$ for $k = 1, \dots, N+1$. A special case is a Hilbert transformer which has an approx. Amplitude of 1 across the entire band, e.g. $B=\text{REMEZ}(30,[.1 \ .9],[1 \ 1],\text{'Hilbert'})$.

$B=\text{REMEZ}(N,F,A,\text{'differentiator'})$ and $B=\text{REMEZ}(N,F,A,W,\text{'differentiator'})$ also design filters with odd symmetry, but with a special weightingscheme for non-zero amplitude bands. The weight is assumed to be equal to the inverse of frequency times the weight W . Thus the filter has a much better fit at low frequency than at high frequency. This designs FIR differentiators.

Example 1:

Design a low-pass filters of length $M=21$, $M=41$, $M=61$, $M=101$ with a pass-band edge frequency $\omega_p = \pi/4$ and stop-band edge frequency $\omega_s = \pi/2$.

Solution and results: application of MATLAB function $b=\text{remez}(N,F,A)$.

$F=[0 \ .25 \ .5 \ 1]$;

$A=[1 \ 1 \ 0 \ 0]$;

$b_{21}=\text{remez}(20,f,a)$; $b_{40}=\text{remez}(40,f,a)$; $b_{61}=\text{remez}(60,f,a)$; $b_{101}=\text{remez}(100,f,a)$;

4.3. Comparison of Design Methods for Linear-Phase FIR Digital Filter

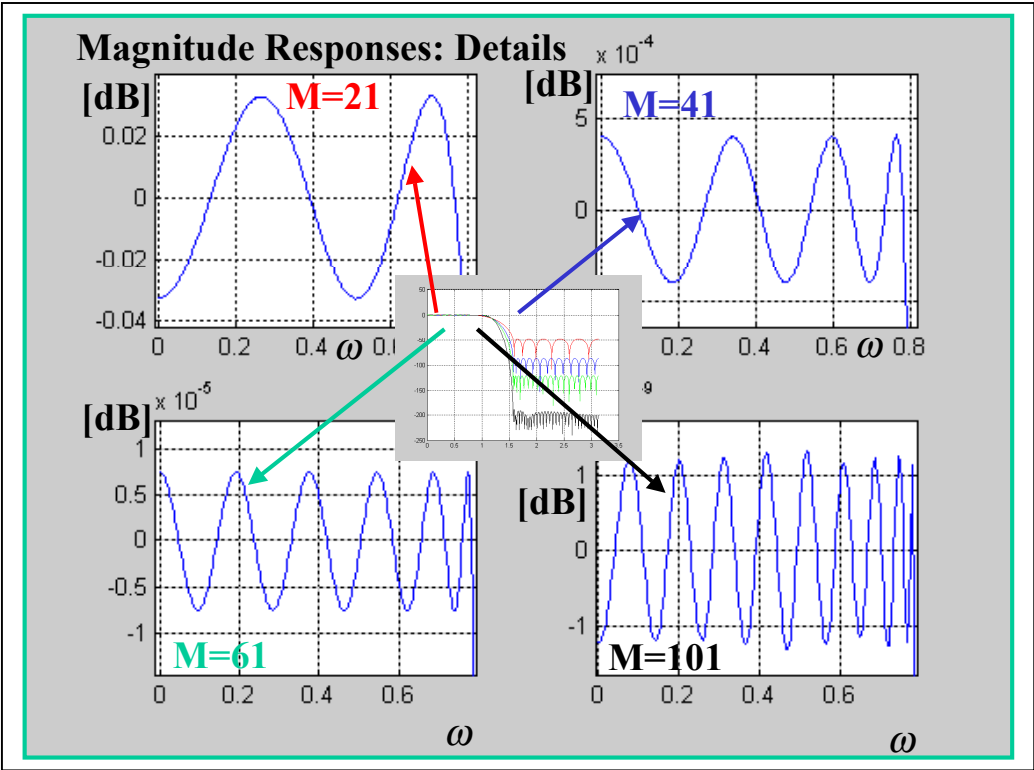
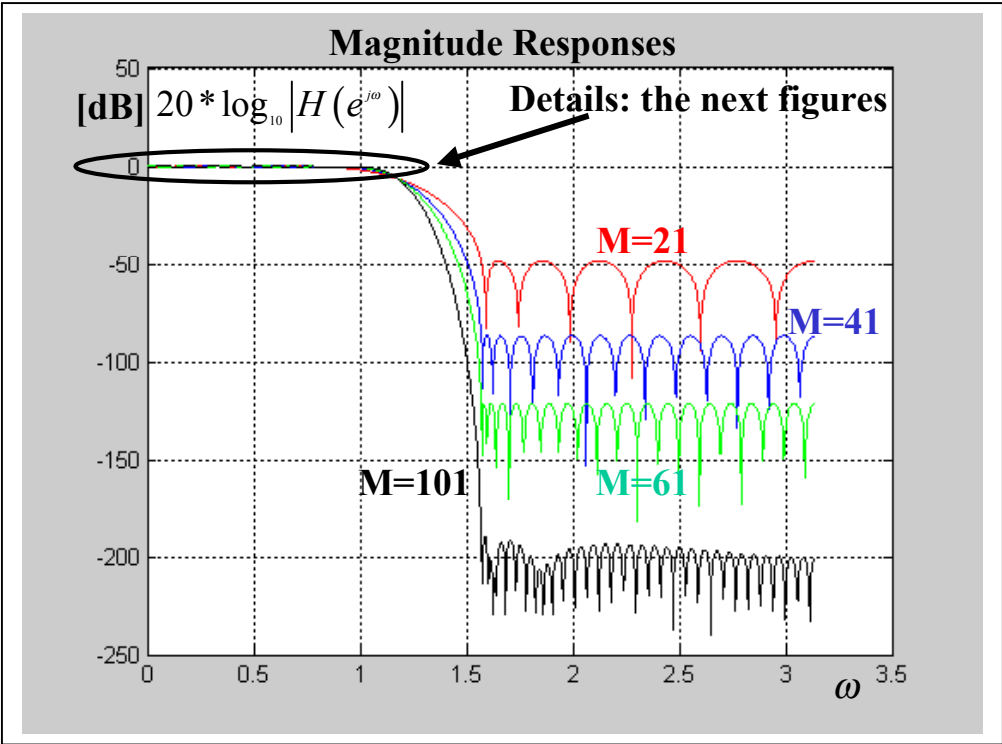
Historically, the design method based on the use of windows to truncate the impulse response and to obtain the desired spectral shaping was the first method proposed for designing linear-phase FIR filters. The frequency-sampling method and the Chebyshev approximation method were developed in the 1970s and have since become very popular in the design of practical linear-phase filters.

The major disadvantage of the window design method is the lack of precise control of the critical frequencies such as pass-band cut off frequency (edge frequency) ω_p and stop-band band cut off frequency (edge frequency) ω_s . The values of ω_p and ω_s , in general, depend on the type of window and the filter length M .

The frequency sampling method provides an improvement over the window design method, since $H_r(\omega)$ is specified at the frequencies $\omega_k = 2\pi k/M$, and the transition band is a multiple of $2\pi k/M$. The filter design method is particularly attractive when the FIR filter is realized either in the frequency domain by means of the DFT or in any of the frequency sampling realizations. The attractive feature of these realizations is that $H_r(\omega_k)$ is either zero or unity at all frequencies, except in the transition band.

The Chebyshev approximation method provides total control of the filter specifications, and, as a consequence, it is usually preferable over the other two methods. For a low-pass filter, the specifications are given in terms of the parameters $\omega_p, \omega_s, \delta_1, \delta_2$ and M .

The Chebyshev design procedure based on the Remez exchange requires that we specify the length of the filter M , the critical frequencies ω_p and ω_s , and the ratio δ_2/δ_1 . However, it is more natural in filter design to specify $\omega_p, \omega_s, \delta_1$ and δ_2 and to determine the filter length M that satisfies the specifications. Although there is no simple formula to determine the filter length from these specifications, a number of approximations have been proposed for estimating M from $\omega_p, \omega_s, \delta_1$ and δ_2 . A particularly, simple formula attributed to Kaiser for approximating M is



$$\widehat{M} = \frac{-20 \log_{10} \sqrt{\delta_1 \delta_2} - 13}{14.6 \Delta f} + 1$$

where Δf is the transition band, defined as $\Delta f = (\omega_s - \omega_p) / 2\pi$ (Rabiner's formula). A more accurate formula proposed by Herrmann et al. is

$$\widehat{M} = \frac{D_\infty(\delta_1, \delta_2) - f(\delta_1, \delta_2)(\Delta f)^2}{\Delta f} + 1$$

where, by definition,

$$D_\infty(\delta_1, \delta_2) = \left[0.005309(\log_{10} \delta_1)^2 + 0.07114 \log_{10} \delta_1 - 0.4761 \right] (\log_{10} \delta_2) - \left[0.00266(\log_{10} \delta_1)^2 + 0.5941 \log_{10} \delta_1 + 0.4278 \right]$$

$$f(\delta_1, \delta_2) = 11.012 + 0.51244(\log_{10} \delta_1 - \log_{10} \delta_2)$$

These formulas are extremely useful in obtaining a good estimate of the filter length required to achieve the given specifications Δf , δ_1 and δ_2 .

5. Design of IIR Digital Filters from Analog Filters

5.1. Introduction

Just as in the design of FIR filters, there are several methods that can be used to design digital filters having an infinite-duration unit sample response. The techniques to be described in this section are all based on taking an analog filter and converting it to a digital filter. Analog filter design is a mature and well-developed field, so it is not surprising that we begin the design of a digital filter in the analog domain and then convert the design into the digital domain.

An analog filter may be described by its system (transfer) function

$$H_A(p) = \frac{L[y(t)]}{L[x(t)]} = \frac{B(p)}{A(p)} = \frac{\sum_{k=0}^M \beta_k p^k}{\sum_{k=0}^N \alpha_k p^k}$$

where α_k and β_k are the filter coefficients, or by its impulse response $h(t)$, which is related to $H_A(p)$ by its Laplace transform

$$H_A(p) = \int_{-\infty}^{\infty} h(t) e^{-pt} dt$$

Alternatively, the analog filter having the rational system function $H_A(p)$ 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 $x(t)$ denotes the input signal and $y(t)$ denotes the output of the filter.

Each of these three equivalent characterizations of an analog filter leads to alternative method for converting the filter into the digital domain, as will be described below. We recall that an analog linear time-invariant system with system function $H_A(p)$ is stable if all its poles lies in the left-half p -plane. Consequently, if the conversion technique is to be effective, it should be possess the following desirable properties:

1. The $j\Omega$ axis in the p -plane should map into the unit circle in the z -plane. Thus there will be a direct relationship between the two frequency variables in the two domains.
2. The left-half p -plane should map into the inside of the unite circle in the z -plane. Thus the stable analog filter will be converted to a stable digital filter

It can be shown that physically realizable (causal) and stable IIR filters cannot have linear phase. If the restriction on causality is removed, it is possible to obtain a linear phase IIR, at least in principle.

In the design of IIR filters, we shall specify the desired filter characteristics for the magnitude response only. This does not mean that we consider the phase response unimportant. Since the magnitude and phase characteristics are related, we specify the desired magnitude characteristics and accept the phase response that is obtained from the design methodology.

In the remainder of this section, the following four most widely used procedures for digitizing the transfer function $H_A(p)$ include

1. Method of approximation of derivatives (differentials).
2. Impulse-invariant method (impulse invariant transformation).
3. The matched z -transform.
4. Bilinear transformation method.

5.2. Method of Approximation of Derivatives (Differentials)

One of the simplest methods for converting an analog filter into a digital filter is to approximate the differential equation

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

by an equivalent difference equation. This approach is often used to solve a linear constant-coefficient differential equation numerically on a digital computer.

For the derivative $\frac{dy(t)}{dt}$ at time $t=nT$, we substitute the backward difference $[y(nT) - y(nT - T)]/T$.

Thus

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

where T represents the sampling interval and $y(n)=y(nT)$. The analog differentiator with output $\frac{dy(t)}{dt}$ has the system (transfer) function $H(p)=p$, while the digital system produces the output $[y(nT) - y(nT - T)]/T$ has the system function

$$H(z) = \frac{1 - z^{-1}}{T}$$

Consequently, the transform-domain equivalent is

$$p = \frac{1 - z^{-1}}{T}$$

The second derivative $\frac{d^2 y(t)}{dt^2}$ is replaced by the second difference, which is derived as follows

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

Then the transform-domain, it is equivalent to

$$p^2 = \frac{1 - 2z^{-1} + z^{-2}}{T} = \left[\frac{1 - z^{-1}}{T} \right]^2$$

It is easily follows from the above that the substitution for the k-th derivative of $y(t)$ results in the equivalent transfer-domain relationship

$$p^k = \left[\frac{1 - z^{-1}}{T} \right]^k$$

Consequently, the system (transfer) function for the digital IIR filter obtained as a result of the approximation of the derivatives by the finite backward difference is

$$H(z) = H_A(p) \Big|_{p=(1-z^{-1})/T}$$

where $H_A(p)$ is the system function of the analog filter.

Now, let us investigate the implications of the mapping from the p -plane to z -plane given by

$$p = \frac{1 - z^{-1}}{T} \longrightarrow z = \frac{1}{1 - pT}$$

When $p = j\Omega$, then

$$z = \frac{1}{1-pT} = \frac{1}{1-j\Omega T} = \frac{1}{2} \frac{(1-j\Omega T) + (1+j\Omega T)}{1-j\Omega T} = \frac{1}{2} \left(1 + \frac{1+j\Omega T}{1-j\Omega T} \right) =$$

$$= \frac{1}{2} \left(1 + \frac{\sqrt{1+(-\Omega T)^2} e^{j \arctg \Omega T}}{\sqrt{1+(-\Omega T)^2} e^{-j \arctg \Omega T}} \right) = \frac{1}{2} (1 + e^{j 2 \arctg \Omega T})$$

Taking real and imaginary parts of z gives

$$\operatorname{Re}[z] = \frac{1}{2} + \frac{\cos[2 \arctg(\Omega T)]}{2} \quad \operatorname{Im}[z] = \frac{\sin[2 \arctg(\Omega T)]}{2}$$

Thus the mapping of the line $p = j\Omega$ for $\Omega = (-\infty, \infty)$ in the z -plane is the circle described

$$\left[\operatorname{Re}[z] - \frac{1}{2} \right]^2 + \operatorname{Im}[z]^2 = \left(\frac{1}{2} \right)^2$$

i.e. a circle with the center at $\operatorname{Re}[z]=1/2$ and radius $1/2$ as shown in the next figure. **Therefore, the $j\Omega$ axis maps onto perimeter of the circle of radius $1/2$ in the z -plane.** Except for extremely small values of ΩT , the image of the $j\Omega$ axis in the p -plane is off the unit circle in the z -plane. Therefore, the property 1 is not satisfied.

To see whether property 2 of mapping is satisfied we set

$$pT = a + jb$$

where $a, b \in R$ and $a < 0$. Because

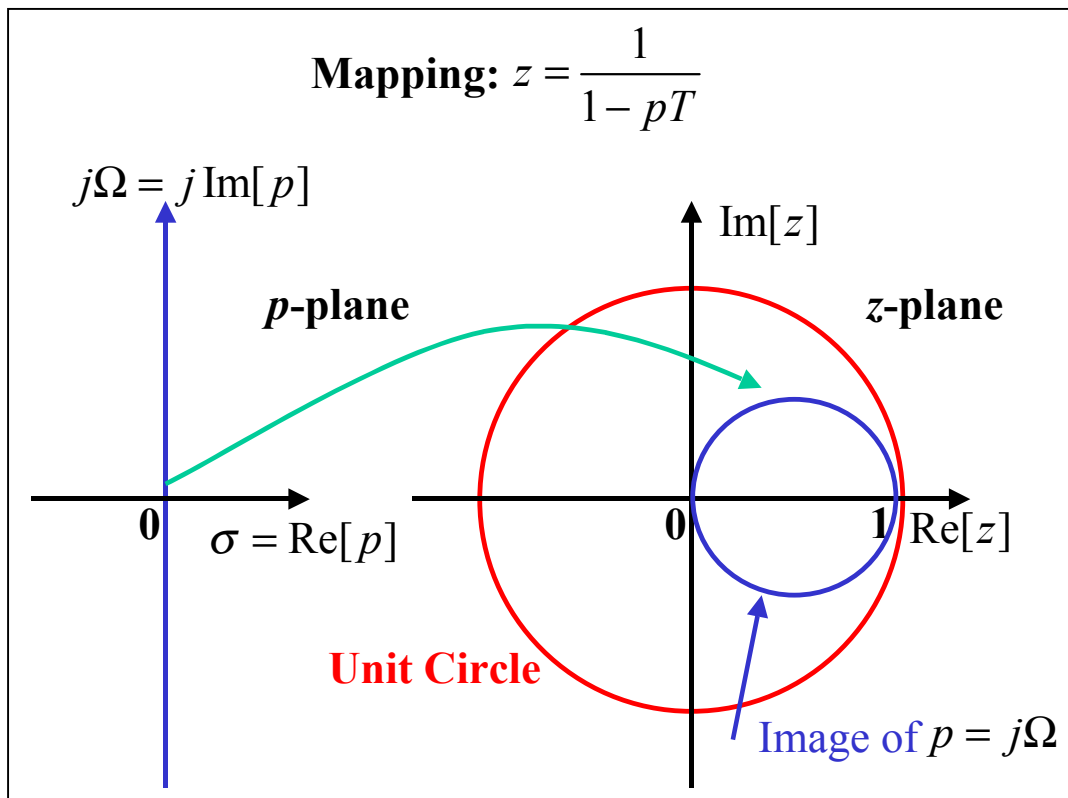
$$z = \frac{1}{1-pT}$$

we can get

$$z = \frac{1}{1-a-bj} = \frac{1}{\sqrt{(1-a)^2 + b^2} e^{-j \arctg \frac{b}{1-a}}}$$

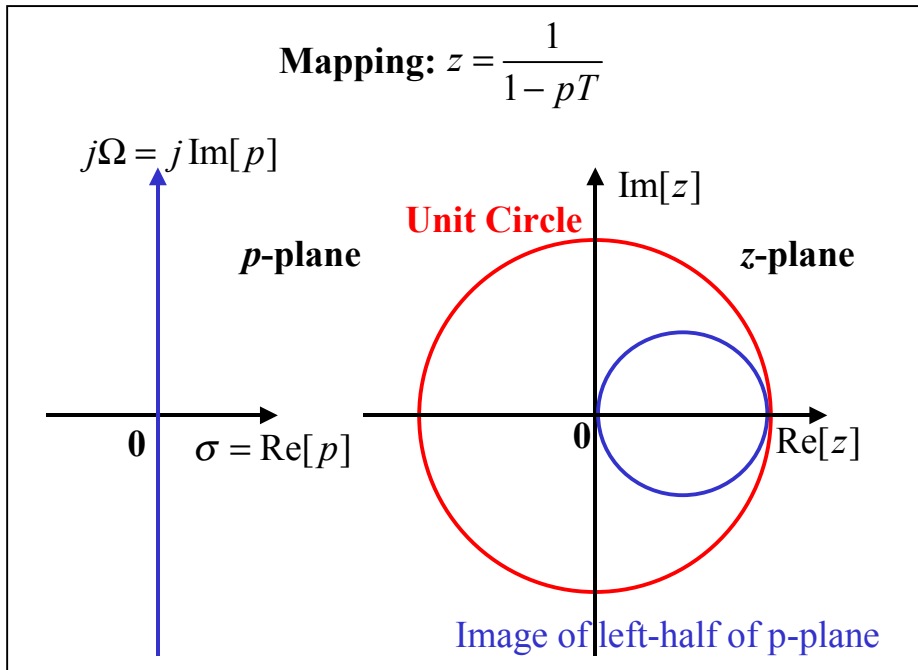
Thus,

$$|z| = \frac{1}{\sqrt{(1-a)^2 + b^2}} < 1$$



Therefore, stable analog filters maps into stable digital filters using backward differences but their frequency properties are not maintained, i.e. this mapping does not preserve the shape of the analog frequency characteristics (the $j\Omega$ axis does not even map onto the unit circle). **It can be shown also , that the left-half *p*-plane is mapped inside a circle with the center at $\text{Re}[z]=1/2$ and radius $1/2$ The right-half *p*-plane maps into the region outside the circle with the center at $\text{Re}[z]=1/2$ and radius $1/2$ (see the next figure).**

If T is decreased, more of the frequency response will be concentrated near $z=1$. For a low-pass filter, this improved the matching between the analog and digital filter frequency responses, but significant distortion will still exists and T may have to be inordinately small. Besides, the possible location of the poles of the digital filter are confined to relatively small frequencies and, as a consequence, the mapping is restricted to design of low-pass filters and band-pass filters having relatively small resonant frequencies. Then, if the desired filter is not low-pass (e.g. if it is high-pass or band-stop filter), the above mentioned procedure (T decreasing) typically cannot be applied. As a result, the backward difference transformation is seldom used.



Example:

Use the backward difference for the derivative to convert the analog low-pass filter with the system (transfer) function

$$H_A(p) = \frac{1}{p+1}$$

into a digital IIR filter.

Solution:

The mapping from p to z is

$$p = \frac{1 - z^{-1}}{T}$$

Hence

$$H(z) = \frac{1}{\frac{1 - z^{-1}}{T} + 1} = \frac{T}{1 - z^{-1} + T} = \frac{T}{(1 + T) - z^{-1}} = \frac{\frac{T}{1 + T}}{1 - \frac{T}{1 + T} z^{-1}}$$