Digital Signal Processing
Lecture 7 - Structures for Discrete-Time Systems

Electrical Engineering and Computer Science
University of Tennessee, Knoxville

Overview

Discrete-time systems

- Special properties: linearity, TI, stability, causality
- LTI systems: the unit sample response $h[n]$ uniquely characterizes an LTI system

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k] = x[n] * h[n]$$

- Frequency response: $H(e^{j\omega})$ is eigenvalues of LTI systems, complex exponentials are eigenfunctions of LTI systems, i.e., if $x[n] = e^{j\omega n}$,

$$y[n] = H(e^{j\omega})x[n] = (\sum_{k=-\infty}^{\infty} h[k]e^{-j\omega k})e^{j\omega n}$$

- Fourier transform: Generalization of frequency response (a periodic continous function of ω)

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n}, x[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(e^{j\omega})e^{j\omega n}d\omega$$

- The z-transform as a generalization to the Fourier transform, $X(z) = \sum_{n=-\infty}^{\infty} x[n]z^{-n}$, and the system function $H(z)$
- Sampling, Aliasing, Reconstruction

Transform-domain Analysis

- FIR vs. IIR
- FIR filters with generalized linear phase (special pattern for zeros)
- Minimum phase systems (special pole-zero properties)
- All-pass systems (special pole/zero properties)
- $H = H_{min} H_{ap}$
- Geometric interpretation of the pole-zero plot

Different system representations

- Using LCDE with initial rest condition

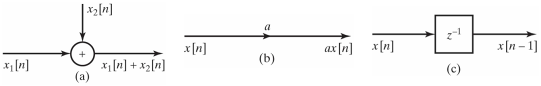
$$y[n] = \sum_{k=0}^{M} b_k x[n-k] + \sum_{k=1}^{N} a_k y[n-k]$$

- Using system function with ROC $|z| > R_+$

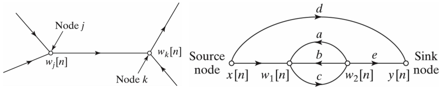
$$H(z) = \frac{\sum_{k=0}^{M} b_k z^{-k}}{1 - \sum_{k=1}^{N} a_k z^{-k}}$$

Block diagram vs. Signal flow graph

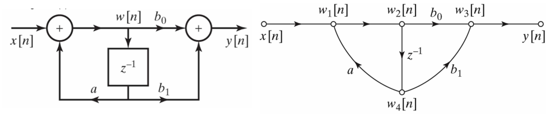
- Block diagram symbols: adder, multiplier, unit delay (memory)



- Signal flow graph: directed branches (branch gain, delay branch), nodes (source node, sink node)

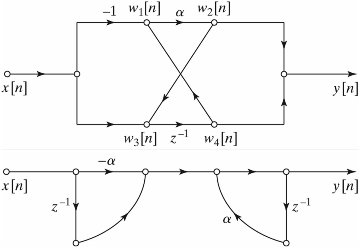


- A comparison: nodes in the flow graph represent both branching points and adders, whereas in the block diagram a special symbol is used for adders



Determination of the system function from a flow graph

- Different flow graph representations require different amounts of computational resources



Direct form I and II

$$H(z) = \frac{\sum_{k=0}^{M} b_k z^{-k}}{1 - \sum_{k=1}^{N} a_k z^{-k}}$$

- Direct form I: implementing zeros first

$$H(z) = H_2(z)H_1(z) = \left(\frac{1}{1 - \sum_{k=1}^{N} a_k z^{-k}}\right)\left(\sum_{k=0}^{M} b_k z^{-k}\right) \quad (1)$$

$$V(z) = H_1(z)X(z) = \left(\sum_{k=0}^{M} b_k z^{-k}\right)X(z) \quad (2)$$

$$Y(z) = H_2(z)V(z) = \left(\frac{1}{1 - \sum_{k=1}^{N} a_k z^{-k}}\right)V(z) \quad (3)$$

$$y[n] = \sum_{k=0}^{M} b_k x[n-k] + \sum_{k=1}^{N} a_k y[n-k] \quad (4)$$

- Direct form II: implementing poles first

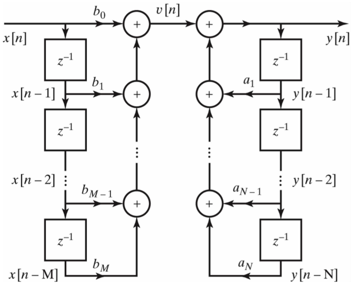
$$H(z) = H_1(z)H_2(z) = \left(\sum_{k=0}^{M} b_k z^{-k}\right)\left(\frac{1}{1 - \sum_{k=1}^{N} a_k z^{-k}}\right) \quad (5)$$

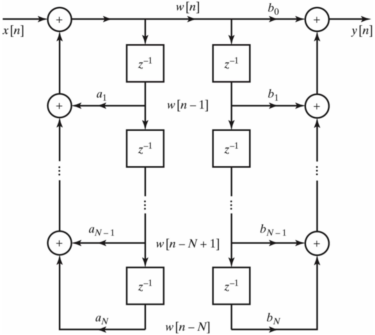
$$W(z) = H_2(z)X(z) = \left(\frac{1}{1 - \sum_{k=1}^{N} a_k z^{-k}}\right)X(z) \quad (6)$$

$$Y(z) = H_1(z)W(z) = \left(\sum_{k=0}^{M} b_k z^{-k}\right)W(z) \quad (7)$$

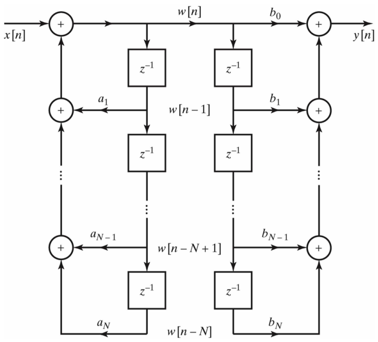
$$y[n] = \sum_{k=1}^{N} a_k y[n-k] + \sum_{k=0}^{M} b_k x[n-k] \quad (8)$$

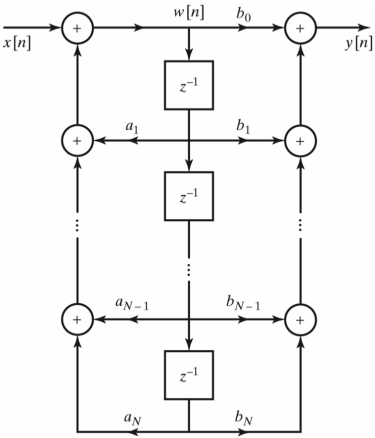
Comparison


b_0
$v[n]$
$x[n]$
$y[n]$
z^{-1}
$x[n-1]$
b_1
a_1
$y[n-1]$
$x[n-2]$
b_{M-1}
a_{N-1}
$y[n-2]$
$x[n-M]$
b_M
a_N
$y[n-N]$

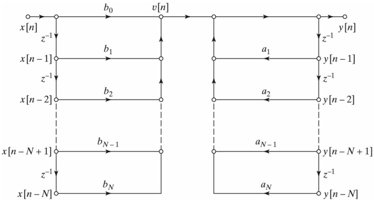

$x[n]$
$w[n]$
b_0
$y[n]$
z^{-1}
a_1
$w[n-1]$
b_1
a_{N-1}
$w[n-N+1]$
b_{N-1}
a_N
$w[n-N]$
b_N

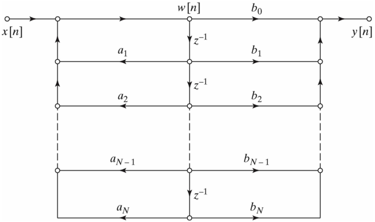
Comparison (cont')





Comparison (cont')

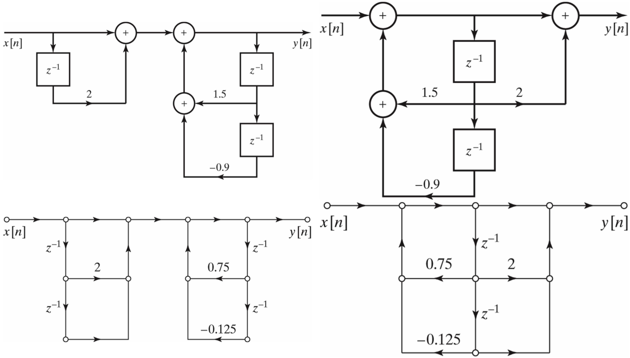




Exercises

- Ex1: $H(z) = \frac{1+2z^{-1}}{1-1.5z^{-1}+0.9z^{-2}}$
- Ex2: $H(z) = \frac{1+2z^{-1}+z^{-2}}{1-0.75z^{-1}+0.125z^{-2}}$

Solution

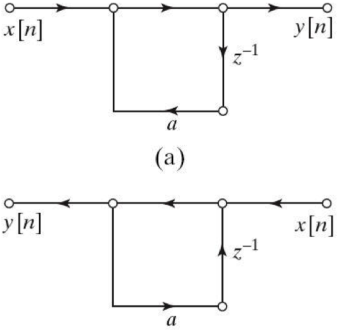


Canonic vs. Noncanonic structures

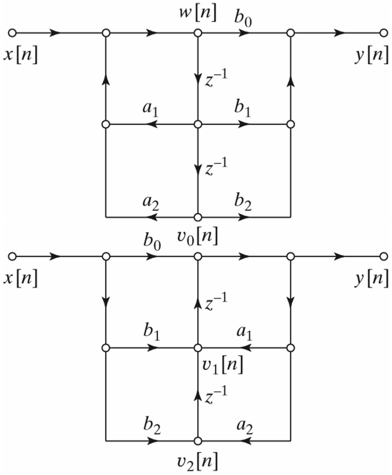
- A digital filter structure is said to be *canonic* if the number of delays is equal to the order of the difference equation. Otherwise, it is a *noncanonic* structure. That is, minimum number of delays required is $max(N, M)$.

Transposed form

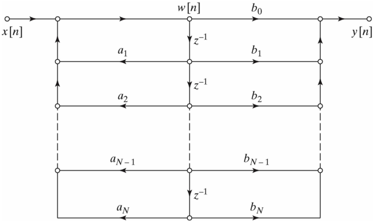
- The transposition theorem: For single-input, single-output systems, the resulting flow graph has the same system function as the original graph if the input and output nodes are interchanged.
  - reverse direction of all branches
  - interchange input and output
- Implement zeros first, then poles as compared to the direct II form

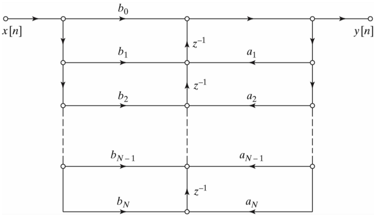


Example



Comparison - Direct form II vs. Transposed direct form II



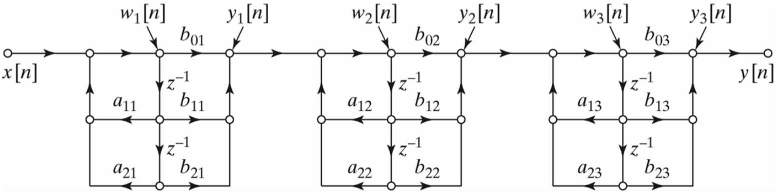


Cascade form

$$H(z) = \frac{\sum_{k=0}^{M} b_k z^{-k}}{1 - \sum_{k=1}^{N} a_k z^{-k}}$$

$$H(z) = \prod_{k=1}^{N_s} \frac{b_{0k} + b_{1k} z^{-1} + b_{2k} z^{-2}}{1 - a_{1k} z^{-1} - a_{2k} z^{-2}}$$

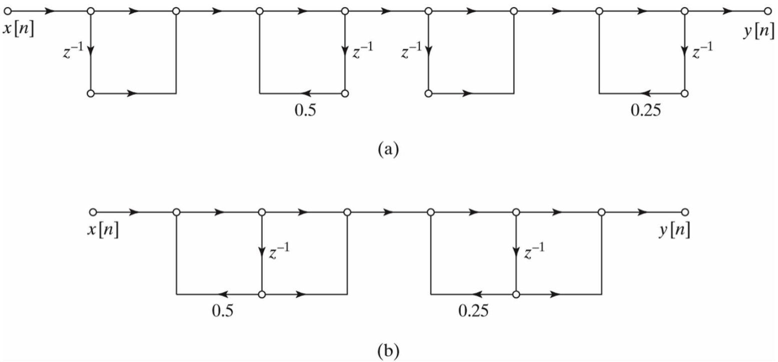
$$N_s = \lfloor (N+1)/2 \rfloor$$



Exercises

- Ex: $H(z) = \frac{1+2z^{-1}+z^{-2}}{1-0.75z^{-1}+0.125z^{-2}} = \frac{(1+z^{-1})(1+z^{-1})}{(1-0.5z^{-1})(1-0.25z^{-1})}$

Solution



Why cascading?

- Use of computation resource
  - Direct form II structure: $2N+1$ constant multipliers

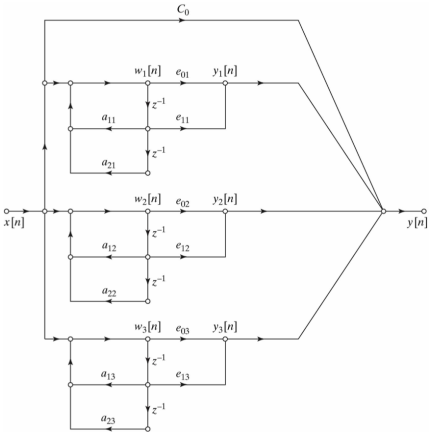
$$H(z) = \frac{\sum_{k=0}^{M} b_k z^{-k}}{1 - \sum_{k=1}^{N} a_k z^{-k}}$$

  - Cascade form structure: $5N/2$ constant multipliers (assume $M = N$ and N is even)

$$H(z) = \prod_{k=1}^{N_s} \frac{b_{0k} + b_{1k} z^{-1} + b_{2k} z^{-2}}{1 - a_{1k} z^{-1} - a_{2k} z^{-2}}$$

- Precision

Parallel form

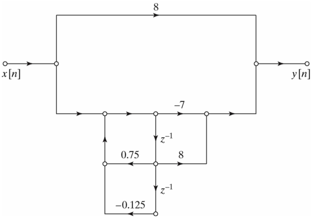


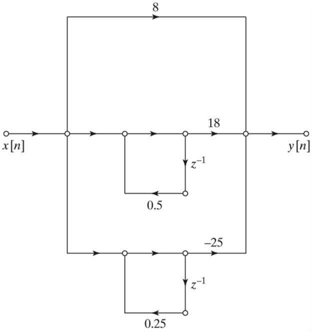
$$H(z) = \sum_{k=0}^{N_p} C_k z^{-k} + \sum_{k=1}^{N_s} \frac{e_{0k} + e_{1k} z^{-1}}{1 - a_{1k} z^{-1} - a_{2k} z^{-2}}$$

Exercises

- Ex: $H(z) = \frac{1+2z^{-1}+z^{-2}}{1-0.75z^{-1}+0.125z^{-2}} = 8 + \frac{-7+8z^{-1}}{1-0.75z^{-1}+0.125z^{-2}}$

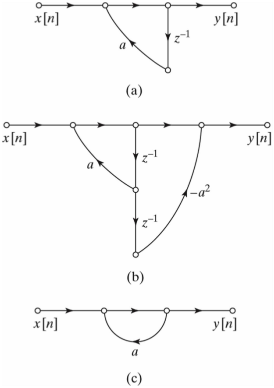
Solution



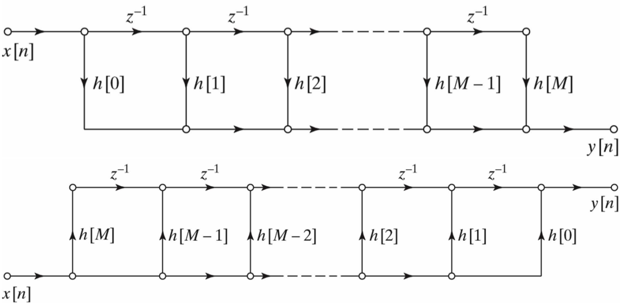


Feedback in IIR systems

- Closed path: necessary to generate infinite long impulse responses (not sufficient)
- The computability of a flow graph is that all loops must contain at least one unit delay element



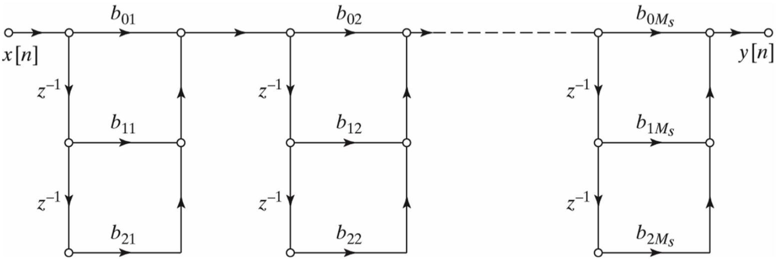
FIR - Direct and transposed direct form



- tapped delay line structure (transversal filter structure)
- discrete convolution

$$y[n] = \sum_{k=0}^{M} h[k]x[n-k]$$

Cascade form

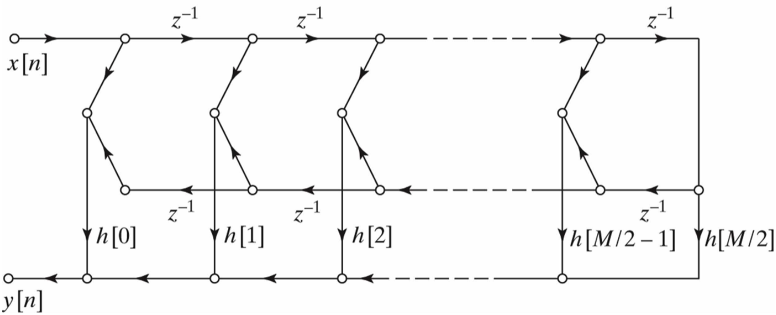


$$H(z) = \sum_{n=0}^{M} h[n]z^{-n} = \prod_{k=1}^{M_s} (b_{0k} + b_{1k}z^{-1} + b_{2k}z^{-2})$$

$$M_s = \lfloor (M+1)/2 \rfloor$$

Linear phase FIR systems

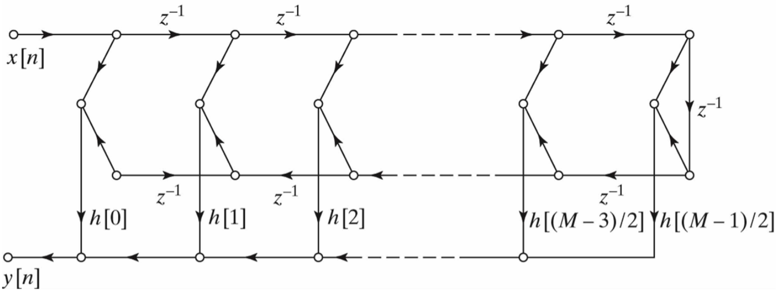
$$h[M-n] = h[n], n = 0, 1, \cdots, M$$



When M is even

Linear phase FIR systems (cont')

$$h[M-n] = h[n], n = 0, 1, \cdots, M$$



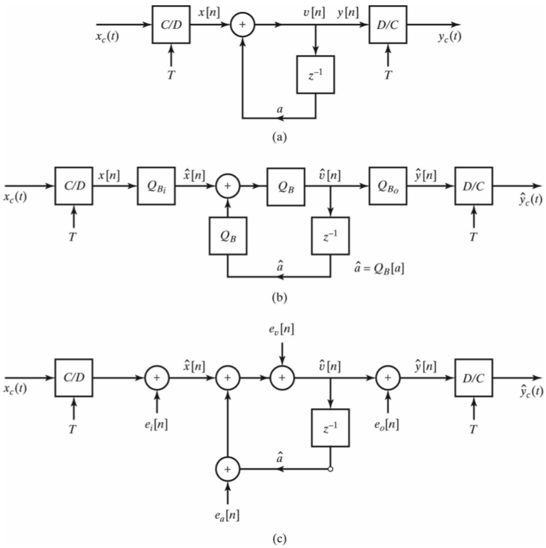
When M is odd

Sources of errors

$$y[n] = ay[n-1] + x[n]$$

- Coefficient quantization problem: $a \to \hat{a}$
- Input quantization error: $x[n] \to \hat{x}[n] = x[n] + e[n]$
- Product quantization error:
  $v[n] = ay[n-1] \to \hat{v}[n] = v[n] + e_a[n]$
- Limit cycles: caused by the nonlinearity by the quantization of arithmetic operations. When the input is absent or constant input or sinusoidal input signals are present, the output is in the form of oscillation

Quantization problem in Implementation



Number representations

- The two's complement format

$$x = X_m(-b_0 + \sum_{i=1}^{\infty} b_i 2^{-i})$$

  - X_m: an arbitrary scale factor
  - b_0: the sign bit. $0 \leq x \leq X_m$ if $b_0 = 0$; $-X_m \leq x < 0$ if $b_0 = 1$

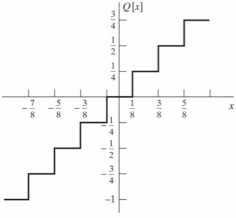
- Fix-point binary numbers

$$\hat{x} = Q_B[x] = X_m(-b_0 + \sum_{i=1}^{B} b_i 2^{-i}) = X_m \hat{x}_B$$

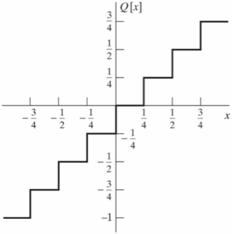
  - Quantizing a number to $B+1$ bits. Quantization error: $e = Q_b[x] - x$

Quantization error

- Rounding: $-\Delta/2 < e \leq \Delta/2$
- Truncating: $-\Delta < e \leq 0$

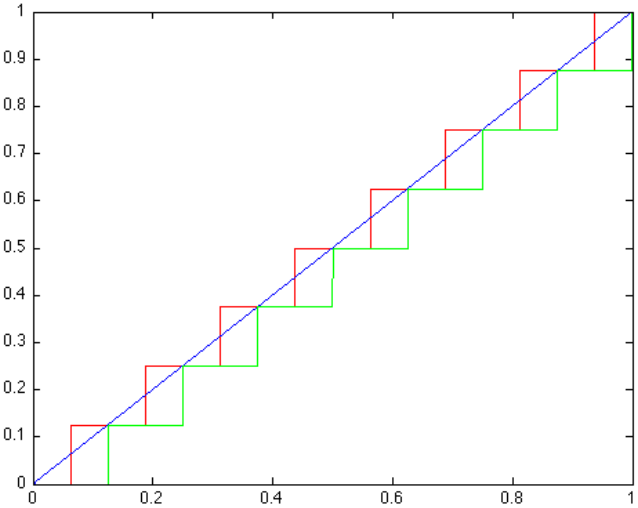


(a)



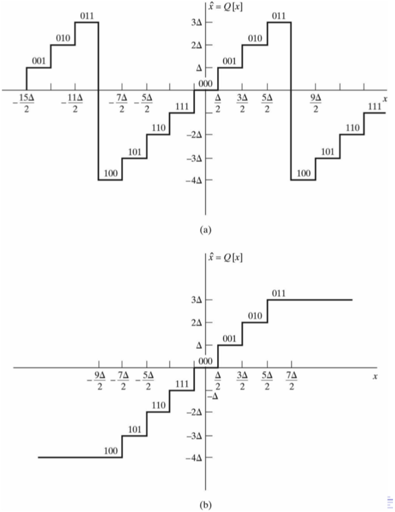
(b)

Quantization error (cont')



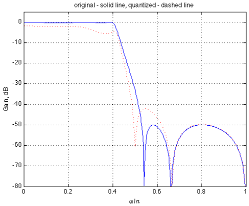
Overflow

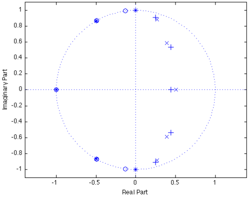
- When $x > X_m$
- Saturation overflow (Clipping)



Coefficient quantization - IIR

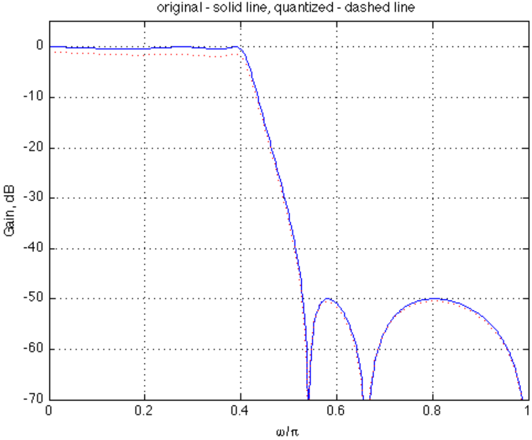
- Effect of coefficient quantization of an IIR digital filter implemented in direct form (5th-order IIR elliptic lowpass filter)





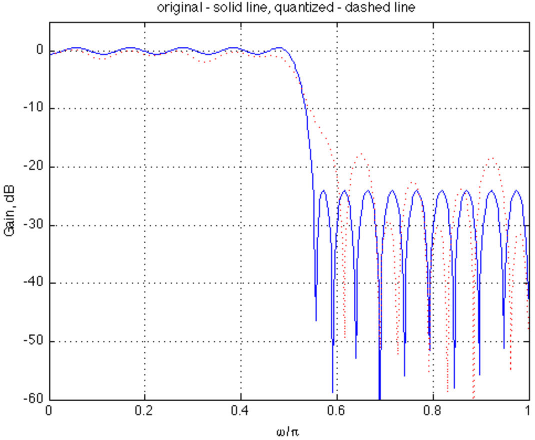
Coefficient quantization - IIR (cont')

- Effect of coefficient quantization of an IIR digital filter implemented in cascade form (5th-order IIR elliptic lowpass filter)



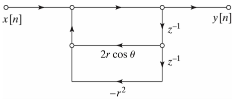
Coefficient quantization - FIR

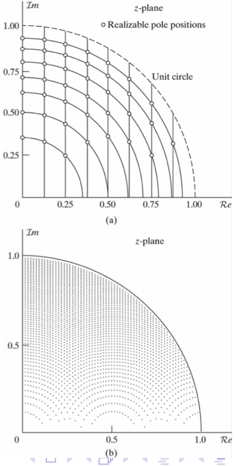
- Effect of coefficient quantization of an FIR digital filter implemented in direct form (39th-order FIR equiripple lowpass filter)



Pole sensitivity of second-order structures (Product quantization)

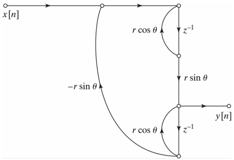
- The direct form structure exhibits high pole sensitivity with poles closer to the real axis and low pole sensitivity with poles closer to $z = \pm j$

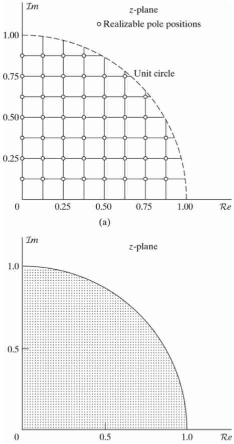




Pole sensitivity of second-order structures (cont')

- The coupled form structure is more suitable for implementing any type of second-order transfer function.





(a)

(b)