7.9 Coefficient Quantization error

In the design of a digital filter the coefficients are evaluated with infinite precision. But when they are quantized, the frequency response of the actual filter deviates from that which would have been obtained with an infinite word length representation and the filter may actually fail to meet the desired specifications. If the poles of the desired filter are close to the unit circle, then those of the filter with quantized coefficients may lie just outside the unit circle leading to instability.

Example 7.8 Consider a second order IIR filter with

$$H(z) = \frac{1.0}{(1-0.5z^{-1})(1-0.45z^{-1})}$$

find the effect on quantization on pole locations of the given system function in direct form and in cascade form . Take $b = 3$ bits.

Solution

Direct Form I

We can write $H(z) = \dfrac{1}{(1-0.95z^{-1}+0.225z^{-2})}$

$$(0.95)_{10} = (0.1111001...)_2$$

$$(-0.95)_{10} = (1.1111001...)_2$$

After truncation we have $(1.111)_2 = -0.875$. Similarly

$$(0.225)_{10} = (0.001110...)_2$$

After truncation we have $(0.001)_2 = 0.125$

$$\text{So} \quad H(z) = \frac{1}{(1-0.875z^{-1}+0.125z^{-2})}$$

Cascade form

$$H(z) = \frac{1}{(1-0.5z^{-1})(1-0.45z^{-1})}$$

$$(-0.5)_{10} = (1.100)_2$$

$$(-0.45)_{10} = (1.01110\ldots)_2$$

After truncation we have $(1.011)_2 = (-0.375)_{10}$

$$\text{So} \quad H(z) = \frac{1}{(1 - 0.5z^{-1})(1 - 0.375z^{-1})}$$

Practice Problem 7.3 A causal LTI system has a system function

$$H(z) = \frac{1}{1 - 1.02z^{-1} + 0.97z^{-2}}$$

(a) Is the system stable?

(b) If the coefficients are rounded to the nearest tenth, would the resulting system be stable?
