

PROPERTIES OF IIR FILTERS*

C. Sidney Burrus

This work is produced by OpenStax-CNX and licensed under the Creative Commons Attribution License 3.0[†]

Digital filters with an Infinite-duration Impulse Response (IIR) have characteristics that make them useful in many applications. This section develops and discusses the properties and characteristics of these filters[3].

Because of the feedback necessary in an implementation, the Infinite Impulse Response (IIR) filter is also called a recursive filter or, sometimes, an autoregressive moving-average filter (ARMA). In contrast to the FIR filter with a polynomial transfer function, the IIR filter has a rational transfer function. The transfer function being a ratio of polynomials means it has finite poles as well as zeros, and the frequency-domain design problem becomes a rational-function approximation problem in contrast to the polynomial approximation for the FIR filter[4]. This gives considerably more flexibility and power, but brings with it certain problems in both design and implementation[3], [2], [1].

The defining relationship between the input and output variables for the IIR filter is given by

$$y(n) = \sum_{k=1}^N a(k) y(n-k) + \sum_{m=0}^M b(m) x(n-m). \quad (1)$$

The second summation in (1) is exactly the same moving average of the present plus past M values of the input that occurs in the definition of the FIR filter. The difference arises from the first summation, which is a weighted sum of the previous N output values. This is the feedback or recursive part which causes the response to an impulse input theoretically to endure forever. The calculation of each output term $y(n)$ from (1) requires $N + M + 1$ multiplications and $N + M$ additions. There are other algorithms or structures for calculating $y(n)$ that may require more or less arithmetic.

In addition to the number of calculations required to calculate each output term being a measure of efficiency, the amount of storage for coefficients and intermediate calculations is important. DSP chips are designed to efficiently implement calculations such as (1) by having a single cycle operation that multiplies a variable by a constant and accumulates it. In parallel with that operation, it is simultaneously calculating the address of the next variable.

Just as in the case of the FIR filter, the output of an IIR filter can also be calculated by convolution.

$$y(n) = \sum_{k=0}^{\infty} h(k) x(n-k) \quad (2)$$

In this case, the duration of the impulse response $h(n)$ is infinite and, therefore, the number of terms in (2) is infinite. The $N + M + 1$ operations required in (1) are clearly preferable to the infinite number required by (2). This gives a hint as to why the IIR filter is very efficient. The details will become clear as the characteristics of the IIR filter are developed in this section.

*Version 1.2: Nov 17, 2012 6:17 pm +0000

[†]<http://creativecommons.org/licenses/by/3.0/>

1 Frequency-Domain Formulation of IIR Filters

The transfer function of a filter is defined as the ratio $Y(z)/X(z)$, where $Y(z)$ and $X(z)$ are the z-transforms of the output $y(n)$ and input $x(n)$, respectively. It is also the z-transform of the impulse response. Using the definition of the z-transform in Equation 32 from Discrete-Time Signals, the transfer function of the IIR filter defined in (1) is

$$H(z) = \sum_{n=0}^{\infty} h(n) z^{-n} \quad (3)$$

This transfer function is also the ratio of the z-transforms of the $a(n)$ and $b(n)$ terms.

$$H(z) = \frac{\sum_{n=0}^M b(n) z^{-n}}{\sum_{n=0}^N a(n) z^{-n}} = \frac{B(z)}{A(z)} \quad (4)$$

The frequency response of the filter is found by setting $z = e^{j\omega}$, which gives (1) the form

$$H(\omega) = \sum_{n=0}^{\infty} h(n) e^{-j\omega n} \quad (5)$$

It should be recalled that this form assumes a sampling rate of $T = 1$. To simplify notation, $H(\omega)$ is used to denote the frequency response rather than $H(e^{j\omega})$.

This frequency-response function is complex-valued and consists of a magnitude and phase. Even though the impulse response is a function of the discrete variable n , the frequency response is a function of the continuous-frequency variable ω and is periodic with period 2π .

Unlike the FIR filter case, exactly linear phase is impossible for the IIR filter. It has been shown that linear phase is equivalent to symmetry of the impulse response. This is clearly impossible for the IIR filter with an impulse response that is zero for $n < 0$ and nonzero for n going to infinity.

The FIR linear-phase filter allowed removing the phase from the design process. The resulting problem was a real-valued approximation problem requiring the solution of linear equations. The IIR filter design problem is more complicated. Linear phase is not possible, and the equations to be solved are generally nonlinear. The most common technique is to approximate the magnitude of the transfer function and let the phase take care of itself. If the phase is important, it becomes part of the approximation problem, which then is often difficult to solve.

2 Calculation of the IIR Filter Frequency Response

As shown in another module, L equally spaced samples of $H(\omega)$ can be approximately calculated by taking an L -length DFT of $h(n)$ given in (5). However, unlike for the FIR filter, this requires that the infinitely long impulse response be truncated to at least length- L . A more satisfactory alternative is to use the DFT to evaluate the numerator and denominator of (4) separately rather than to approximately evaluate (5). This is accomplished by appending $L - N$ zeros to the $a(n)$ and $L - M$ zeros to the $b(n)$ from (1), and taking length- L DFTs of both to give

$$H(2\pi k/L) = \frac{\mathcal{DFT}\{b(n)\}}{\mathcal{DFT}\{a(n)\}} \quad (6)$$

where the division is a term-wise division of each of the L values of the DFTs as a function of k . This direct method of calculation is a straightforward and flexible technique that does not involve truncation of $h(n)$ and the resulting error. Even nonuniform spacing of the frequency samples can be achieved by altering the DFT as was suggested for the FIR filter. Because IIR filters are generally lower in order than FIR filters, direct use of the DFT is usually efficient enough and use of the FFT is not necessary. Since the $a(n)$ and

$b(n)$ do not generally have the symmetries of the FIR $h(n)$, the DFTs cannot be made real and, therefore, the shifting and stretching techniques of other modules are not applicable.

As an example, the frequency-response plot of a third-order elliptic-function lowpass filter with a transfer function of

$$H(z) = \frac{0.1335z^3 + 0.056z^2 + 0.056z + 0.1335}{z^3 - 1.507z^2 + 1.2646z - 0.3786} \quad (7)$$

is given in Figure 1a. The details for designing this filter are discussed in elsewhere. A similar performance for the magnitude response would require a length of 18 for a linear-phase FIR filter.

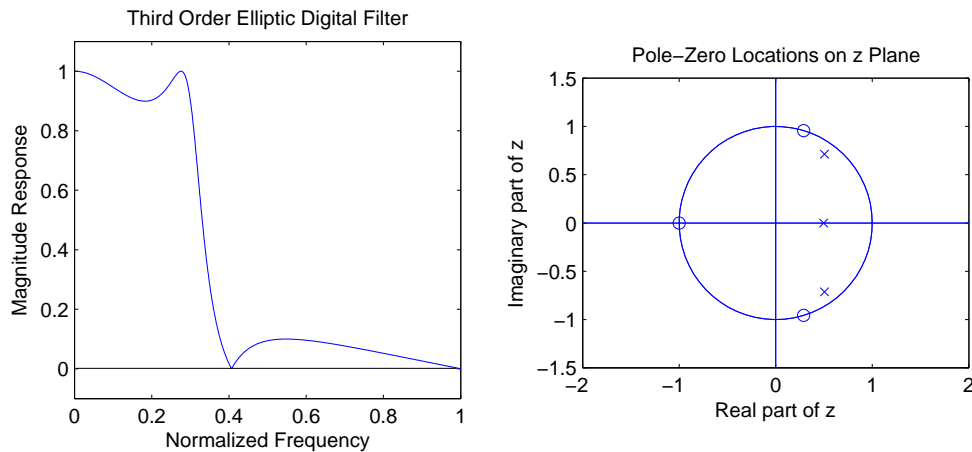


Figure 1: Frequency Response and Pole-Zero Locations of a Third-Order IIR Filter

3 Pole-Zero Locations for IIR Filters

The possible locations of the zeros of the transfer function of an FIR linear-phase filter were analyzed elsewhere. For the IIR filter, there are poles as well as zeros. For most applications, the coefficients $a(n)$ and $b(n)$ are real and, therefore, the poles and zeros occur in complex conjugate pairs or are real. A filter is stable if for any bounded input, the output is bounded. This implies the poles of the transfer function must be strictly inside the unit circle of the complex z plane. Indeed, the possibility of an unstable filter is a serious problem in IIR filter design, which does not exist for FIR filters. An important characteristic of any design procedure is the guarantee of stable designs, and an important ability in the analysis of a given filter is the determination of stability. For a linear filter analysis, this involves the zeros of the denominator polynomial of (4). The location of the zeros of the numerator, which are the zeros of $H(z)$, are important to the performance of the filter, but have no effect on stability.

If both the poles and zeros of a transfer function are all inside or on the unit circle of the z -plane, the filter is called minimum phase. The effects of a pole or zero at a radius of r from the origin of the z -plane on the magnitude of the transfer function are exactly the same as one at the same angle but at a radius of $1/r$. However, the effect on the phase characteristics is different. Since only stable filters are generally used in practice, all the poles must be inside the unit circle. For a given magnitude response, there are two possible locations for each zero that is not on the unit circle. The location that is inside gives the least phase shift, hence the name "minimum-phase" filter.

The locations of the poles and zeros of the example in (7) are given in Figure 1b.

Since evaluating the frequency response of a transfer function is the same as evaluating $H(z)$ around the unit circle in the z -plane, a comparison of the frequency-response plot in Figure 1a and the pole-zero locations in Figure 1b gives insight into the effects of pole and zero location on the frequency response. In the case where it is desirable to reject certain bands of frequencies, zeros of the transfer function will be located on the unit circle at locations corresponding to those frequencies.

By having both poles and zeros to describe an IIR filter, much more can be done than in the FIR filter case where only zeros exist. Indeed, an FIR filter is a special case of an IIR filter with a zero-order denominator. This generality and flexibility does not come without a price. The poles are more difficult to realize than the zeros, and the design is more complicated.

4 Summary

This section has given the basic definition of the IIR or recursive digital filter and shown it to be a generalization of the FIR filter described in the previous chapters. The feedback terms in the IIR filter cause the transfer function to be a rational function with poles as well as zeros. This feedback and the resulting poles of the transfer function give a more versatile filter requiring fewer coefficients to be stored and less arithmetic. Unfortunately, it also destroys the possibility of linear phase and introduces the possibility of instability and greater sensitivity to the effects of quantization. The design methods, which are more complicated than for the FIR filter, are discussed in another section, and the implementation, which also is more complicated, is discussed in still another section.

References

- [1] Sanjit K. Mitra. *Digital Signal Processing, A Computer-Based Approach*. McGraw-Hill, New York, third edition, 2006. First edition in 1998, second in 2001.
- [2] A. V. Oppenheim and R. W. Schaffer. *Discrete-Time Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ, second edition, 1999. Earlier editions in 1975 and 1989.
- [3] T. W. Parks and C. S. Burrus. *Digital Filter Design*. John Wiley & Sons, New York, 1987.
- [4] Charles M. Rader. The rise and fall of recursive digital filters. *IEEE Signal Processing Magazine*, 23(6):468211;49, November 2006.