PHY408

Lecture 8: Z-Transform, Digital Filters

March 8, 2023

LTI systems

1. Recall LTI systems which can be described by a weight function:

$$f(t) * w(t) = g(t) \tag{1}$$

   We sometimes need to design $w(t)$ based on given characteristics to 'filter' the input signal $f(t)$. These characteristics a lot of times are specified in the frequency domain (frequency-selective filter, such as low-pass, high-pass, band-pass, band-stop, etc)

2. This can be realized in both time and frequency domain. Which is more efficient?

3. Time-domain convolution $g = f * w$ (f of length N, w of length M, $N \geq M$),

$$g_n = \Delta \sum_{k=-\infty}^{\infty} w_k f_{n-k} \tag{2}$$

   which requires $\sim NM$ number of operations.

4. frequency-domain filter: order $N \log_2 N$. Is it significantly less than NM?

Filters in practice

If $N \sim M$, $Nlog_2 N \ll NM$, frequency implementation is more efficient. However, in practice, time-domain convolution may be desirable sometimes:

1. f and w are of disparate lengths: $N \gg M$,
   - freq. domain: $\sim N \log_2 N$ operations
   - time domain: $\sim NM$ operations, which may be significantly less!
2. real-time filter: adding one extra input sample point requires:
   - freq. domain: additional $\sim N \log_2 N$ operations
   - time domain: additional N operations

Z transform

For a time series $f_0, f_1, \cdots$, define its z-transform as

$$F(z) = f_0 + f_1 z + f_2 z^2 + \cdots = \sum_{k=0}^{\infty} f_k z^k \tag{3}$$

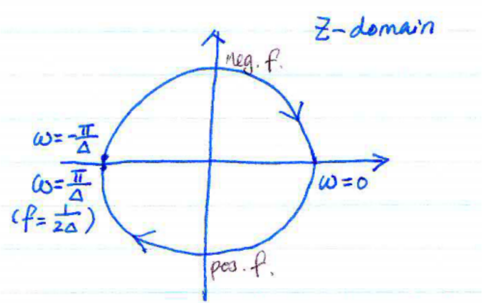
1. express f_k as coefficients of a polynomial (or power series) of z, an analytic function of complex variable z, well-behaved within a region of convergence (ROC); a convenient theoretical tool)
2. why z-transform? If $f_k \to F(z)$ and $g_m \to G(z)$, then

$$\begin{aligned} H(z) &= F(z)G(z) = \sum_k \sum_m f_k g_m z^{k+m} = \sum_n \sum_k f_k g_{n-k} z^n \\ &= \sum_n h_n z^n \leftarrow h_n = \sum_k f_k g_{n-k} \end{aligned} \tag{4}$$

   This is the convolution theorem for the z-transform.
3. $F(z) \Rightarrow f_n$: polynomial series expansion and reading off the coefficients in front of z^n.
4. Note, some books follow the definition of $F(z) = \sum_{k=0}^{\infty} f_k z^{-k}$.

Relationship between z-transform and FT

Let $z = e^{-i\omega\Delta}$, i.e. z clock-wisely goes along the unit circle as $\omega \in [-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}]$ or freq $f \in [-f_N, f_N]$, where $f_N = \frac{1}{2\Delta}$ is the Nyquist frequency.



1. $F(z = e^{-i\omega\Delta}) \Rightarrow F(\omega) = \sum_k f_k e^{-i\omega k\Delta}$ is the FT of time series f_k
2. implicit periodicity $e^{i\phi} = e^{i(\phi+2\pi)}$; one cycle $f \in [-f_N, f_N]$.

Properties of Z transform

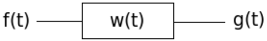
Properties of z transform

- linearity: $ax_n + by_n \to aX(z) + bY(z)$
- time shifting: $w_n \to W(z)$, then $w_{n-k} \to W(z)z^k$.
- complex conjugation: $x_n \to X(z)$, then $x_n^* \to X^*(z)$, i.e., only conjugate the coefficients of $X(z)$, not z itself ($a + bz$ becomes $a^* + b^* z$).
- time reversal: $x_n \to X(z)$, then $x_{-n} \to X(1/z)$

Often, system function $W(z)$ can be written as $\frac{Q(z)}{P(z)}$, where $P(z)$ and $Q(z)$ are polynomials of z. How should $W(z)$ be applied to an input signal $F(z)$?

- two-step operation: $G(z) = F(z)W(z) = [F(z)Q(z)]\frac{1}{P(z)}$
- $D(z) = [F(z)Q(z)]$ is a simple convolution of f_n and q_n.
- $G(z) = \frac{D(z)}{P(z)}$ involves deconvolution and inverse filtering, or $D(z) = P(z)G(z)$, i.e. what is the 'input' $G(z)$ given 'output' $D(z)$ and known system function $P(z)$.

Deconvolution by Polynomial Division



If $f_k * w_k = g_k$, then $G(z) = W(z)F(z)$. Given output $G(z)$, input

$$F(z) = \frac{G(z)}{W(z)} \tag{5}$$

which can be realized by polynomial long-division

1. Example: $w = [1, 1]$, $g = [1, 4, 6, 4, 1]$, therefore

$$F(z) = \frac{1 + 4z + 6z^2 + 4z^3 + z^4}{1 + z} = 1 + 3z + 3z^2 + z^3 \tag{6}$$

2. Limited to very simple time series of finite-length

Deconvolution by Inverse Filtering

If $y(t) * a(t) = x(t)$, $Y(z)A(z) = X(z)$, then given observation of output signal $x(t)$, how do we retrieve the input signal $y(t)$ if $a(t)$ is known as $[a_0, a_1, \cdots]$? Define $B(z) = 1/A(z)$,

$$Y(z) = \frac{X(z)}{A(z)} = \frac{X(z)}{a_0 + a_1 z + \cdots} = X(z)B(z) \tag{7}$$

Zeros of $A(z)$ are the poles of $B(z)$.

1. Example: $A(z) = 1 - \frac{1}{2}z$ and $B(z) = \frac{1}{1-\frac{z}{2}}$ (with pole at 2), then

$$B(z) = \frac{1}{A(z)} = \frac{1}{1 - \frac{z}{2}} = (1 - \frac{z}{2})^{-1} = 1 + \frac{z}{2} + \frac{z^2}{4} + \frac{z^3}{8} + \frac{z^4}{16} + \cdots \tag{8}$$

   and $b = [1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \cdots]$ converges (ratio test) and is causal!
2. pole for $B(z)$ at $z = 2$; ROC for z: $|z| < 2$, a circular region inside the pole that includes the unit circle.
3. convergence means terms at large n can be ignored, i.e. filter b_n can be truncated.
4. realizable time series b_n: stable (absolutely summable) and causal

Inverse Filtering: series that are NOT realizable

Another example: If $A(z) = 1 - 2z$ and $B(z) = \frac{1}{1-2z}$ (with pole at $\frac{1}{2}$):

1. $$B(z) \sim (1-2z)^{-1} = 1 + 2z + 4z^2 + 8z^3 + \cdots \leftrightarrow b = [1, 2, 4, 8, \cdots]$$

   causal (right-sided), but does not converge, not 'realizable' time series.

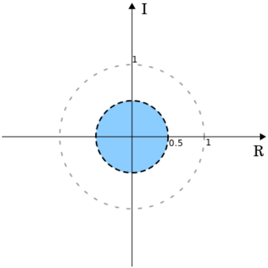
2. but if we write

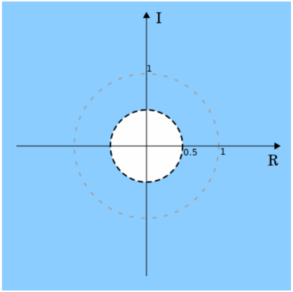
   $$B(z) = -\frac{1}{2z}(1-\frac{1}{2z})^{-1} = -\frac{1}{2}z^{-1} - \frac{1}{4}z^{-2} - \frac{1}{8}z^{-3} - \cdots \leftrightarrow b = [\cdots, -\frac{1}{8}, -\frac{1}{4}, -$$

   non-causal (left-sided) time series, converges as $n \to -\infty$, not 'realizable'.

3. ROC also needs to be specified: $|z| < \frac{1}{2}$ (causality, right-sided) and $|z| > \frac{1}{2}$ (left-sided).
4. stability: ROC includes unit circle
5. in this case, for pole at $\frac{1}{2}$, causality and stability can not be simultaneously satisfied, not realizable

Series that are not realizable: continued





Based on our definition of z-transform, we have:
Left: causal (right-sided) time series; ROC does not include unit circle (diverging)
Right: anti-causal (left-sided) time series, ROC includes unit circle (converging)

Inverse Filtering

For a general two point time series $A(z) = 1 - z/p$, where p is the zero of $A(z)$

$$B(z) = \frac{1}{1 - \frac{z}{p}} = 1 + \frac{z}{p} + \frac{z^2}{p^2} + \frac{z^3}{p^3} + \cdots \tag{9}$$

which corresponds to a right-sided time series $b = [1, \frac{1}{p}, \frac{1}{p^2}, \frac{1}{p^3}, \cdots]$

1. b_i converges $\Rightarrow |\frac{1}{p}| < 1$, in other words, zeros of $A(z)$ or poles of $B(z)$ p should be outside the unit circle. And the ROC for $B(z)$ is $|\frac{z}{p}| < 1$, i.e. $|z| < |p|$, the circular region inside the pole.
2. realizable time series: stable (poles outside unit circle) and causal (ROC is the circular region inside poles)
3. causality and stability may not be compatible
4. A simple $B(z)$ may correspond to a time series of infinite length: infinite impulse response (IIR); if b_i is of finite length: finite impulse response (FIR).

Inverse Filtering

For a general

$$A(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_N z^N = a_N (z - r_1)(z - r_2) \cdots (z - r_N) \quad (10)$$

i.e. the input signal goes through N filters in series, the inverse filter

$$B(z) = \frac{1}{a_N} \frac{1}{z - r_1} \frac{1}{z - r_2} \cdots \frac{1}{z - r_N} \quad (11)$$

1. $B(z)$ realizable: $|r_i| > 1$, for $i = 1, \cdots, N$, i.e. all poles are outside the unit circle.
2. Implementation, $x(t)$ goes through each individual $B_i(z)$, however $B_i(z)$ may correspond to a long time series, making the entire filter time-consuming.
3. If $A(z)$ is relatively short: feedback filtering

Inverse Filtering

Suppose $y(t) * a(t) \Rightarrow x(t)$, and now from the output $x(t)$ how can we retrieve the input time series $y(t)$, given $A(z) = a_0 + a_1 z + \cdots + a_{M-1}$?

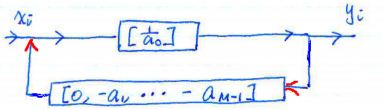
1. From convolution

$$x_i = \sum_{k=0}^{M-1} a_k y_{i-k} = a_0 y_i + \sum_{k=1}^{M-1} a_k y_{i-k} \tag{12}$$

   Now we can retrieve y_i given x_i and $y_j, j = 0, \cdots i-1$, provided $a_0 \neq 0$:

$$y_i = \frac{1}{a_0}\left[x_i - \sum_{k=1}^{M-1} a_k y_{i-k}\right] \tag{13}$$

2. iterative computation: obtain the current value of y_i based on current value of x_i and previous values of y, involving only extra M computation per point, i.e., the shorter the a series, the more efficient the filter.



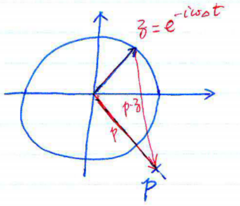
Filter Design

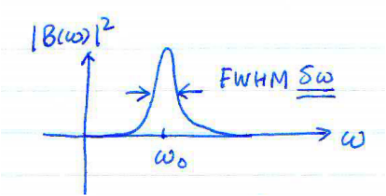
For given characteristics of a filter in the freq domain, choose a proper function form and solve for related parameters

1. real filters are often required $\Rightarrow$ real-even and imaginary odd spectrum.
2. For example, a narrow band filter at ω_0, we use the function form

$$B(z) = \frac{p}{p - z} \tag{14}$$

where $p = (1 + \epsilon)e^{-i\omega_0}$ (outside unit circle!), and assuming $\Delta t = 1$





3. The full-width-half-maximum (i.e. the sharpness of the narrow-band filter) is determined by parameter ϵ

Example: Narrow-band filter

Its frequency spectrum:

$$|B(z = e^{-i\omega})|^2 = \frac{(1+\epsilon)^2}{\epsilon^2 + 4(1+\epsilon)\sin^2\frac{\omega-\omega_0}{2}} \tag{15}$$

1. usually $\epsilon << 1$,
2. at $\omega = \omega_0$, $|B(\omega)|^2$ takes the maximum value $\frac{(1+\epsilon)^2}{\epsilon^2}$
3. if at $\omega = \omega_0 \pm \delta\omega/2$, $|B(\omega)|^2 = \frac{1}{2}|B(\omega_0)|^2$ ($\delta\omega$ is so-called Full Width Half Maximum(FWHM)), then

$$\epsilon^2 = 4(1+\epsilon)\sin^2(\delta\omega/4) \sim (1+\epsilon)\delta^2\omega/4 \tag{16}$$

and $\epsilon \sim \frac{1}{2}\delta\omega$, relating function parameters to filter characteristics.

Narrow-band filter

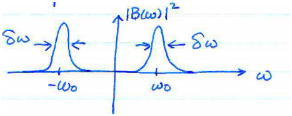
In order to generate a real time filter, we need $B(-\omega) = B^*(\omega)$. Given $B(\omega) = \frac{p}{p-e^{-i\omega}}$ for $\omega > 0$, we have $B(-\omega) = B^*(\omega) = \frac{p^*}{p^*-e^{i\omega}}$, i.e.

$$B(\omega) = \frac{p^*}{p^* - e^{-i\omega}}, \quad \omega < 0 \tag{17}$$

or $B(z) = \frac{p^*}{p^*-z}$ for a series corresponding to neg freq. spectrum, i.e., we need to add poles and zeros that are complex conjugate of the poles and zeros for original $B(z)$.
Multiply together the positive and negative freq. spectrum:

$$B(z) = \frac{p}{p - z}\frac{p^*}{p^* - z} = \frac{1}{1 - \frac{(p+p^*)z}{|p|^2} + \frac{z^2}{|p|^2}} \tag{18}$$

which can be implemented as an inverse filter of $a = \left[1, -\frac{p+p^*}{|p|^2}, \frac{1}{|p|^2}\right]$, a 3-pt real time series.

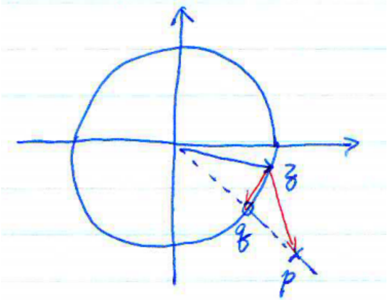


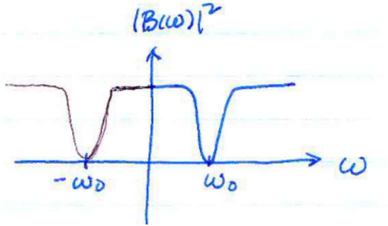
More Examples: Notch filter

1. bandpass filter: composed of a series of narrow band filters to cover the desired frequency range
2. notch filter:

$$B(z) = \frac{z-q}{z-p}, \quad \text{where } p = (1+\epsilon)e^{-i\omega_0} \text{ and } q = e^{-i\omega_0} \tag{19}$$

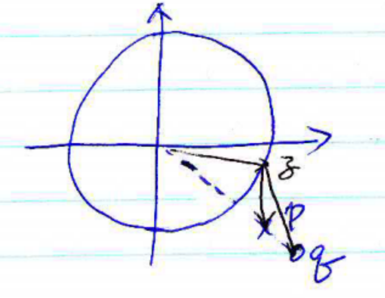
i.e. pole p outside the unit circle and zero q on the unit circle

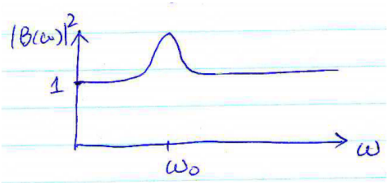




1. Pedestal filter

$$B(z) = \frac{z-q}{z-p}, \quad \text{where } p = (1+\epsilon)e^{-i\omega_0} \text{ and } q = (1+2\epsilon)e^{-i\omega_0} \tag{20}$$





Filter design

1. If we require our filter to be zero-phase, i.e. $B(\omega)$ to be real, then b_n has to be real-even and imaginary-odd, which means b_n is non-causal
2. On the other hand, if b_n is causal, then $B(\omega)$ can not be a pure real function, therefore, phase shift is introduced to the input signal after filtering.
3. Sometimes filter characteristics are given in terms of $|B(\omega)|^2$ (square amplitude response). Through bilinear transformation: $i\omega\Delta t \to 2\frac{1-z}{1+z}$, it may be converted to

$$|B(i\omega)|^2 = B(i\omega)B^*(-i\omega) \to B(z)B^*(1/z) = S(z) = \frac{z-q}{z-p}\frac{\frac{1}{z}-q^*}{\frac{1}{z}-p^*} \tag{21}$$

whose poles and zeros are conjugate reciprocal pairs (different from the real filter requirement!). Then there exists several $B(z)$ that may have the same amplitude spectrum.