Notes 22 largely plagiarized by %khc

1 LT v. ZT

Using the conformal map $z = e^{sT}$, we can transfer our knowledge of the ROC of the bilateral laplace transform to the ROC of the bilateral z transform.

Laplace transform of:	ROC is:
right sided function	right of rightmost pole (not necessarily causal)
left sided function	left of leftmost pole
two sided function	vertical strip in s plane
finite duration	entire s plane

Note also that the ROC of a transfer function corresponding to a stable system contains the $j\omega$ axis.

Z transform of:	ROC is:
right sided sequence	outside of outermost pole
left sided sequence	inside of innermost pole
two sided sequence	ring in z plane centered on origin
finite duration	entire z plane perhaps except for $z = 0$ or $z = \infty$

Note also that the ROC of a transfer function corresponding to stable system contains the unit circle.

In other words, through the use of $z = e^{sT}$, the left half s plane maps to the inside of the unit circle, the $j\omega$ axis maps to the unit circle, and the right half s plane maps to the outside of the unit circle.

2 ω and Ω, Revisited

From the pecularities of discrete time [cf page 33 of OWY], we note that:

- for normalized frequency $\Omega = \omega T$:
  - low frequencies are about even multiples of π
  - high frequencies are about odd multiples of π
- for unnormalized frequency ω:
  - low frequencies are about even multiples of $\frac{\pi}{T}$
  - high frequencies are about odd multiples of $\frac{\pi}{T}$

Note that some books use Ω to denote unnormalized frequency and ω to denote normalized frequency, such as the ee123 textbook.

Why use normalized frequency? The units of ω are radians/second, while the units of $\Omega = \omega T$ are (radians/second)(seconds/cycle), or radians/cycle. Note that time units have been completely normalized away.

3 Solving Difference Equations

To solve this type of equation, we need a few properties. We'll assume that our systems are causal, and use the unilateral Z transform instead of the bilateral one.

We then note that:

$$\begin{aligned}
\mathcal{Z}[x[n-1]] &= \sum_{n=0}^{\infty} x[n-1]z^{-n} \\
&= \sum_{n'+1=0}^{\infty} x[n']z^{-(n'+1)} \\
&= z^{-1}\sum_{n'=-1}^{\infty} x[n']z^{-n'} \\
&= z^{-1}[\sum_{n'=0}^{\infty} x[n']z^{-n'} + zx[-1]] \\
&= z^{-1}[\sum_{n'=0}^{\infty} x[n']z^{-n'}] + x[-1] \\
&= z^{-1}X(z) + x[-1]
\end{aligned}$$

We can similarly apply the above to a sequence delayed by 2 instead of 1:

$$\mathcal{Z}[x[n-2]] = z^{-2}X(z) + z^{-1}x[-1] + x[-2]$$

Let's try these out on the following equation:

$$y[n] - 3y[n-1] + 2y[n-2] = x[n] - x[n-1]$$

Transforming gives us:

$$\begin{aligned}
Y(z) - 3(z^{-1}Y(z) + y[-1]) + 2(z^{-2}Y(z) + z^{-1}y[-1] + y[-2]) &= X(z) - (z^{-1}X(z) + x[-1]) \\
(1 - 3z^{-1} + 2z^{-2})Y(z) &= (1 - z^{-1})X(z) - x[-1] + 3y[-1] - 2z^{-1}y[-1] - 2y[-2]
\end{aligned}$$

We recognize the stuff with the X in it as the ZSR and the stuff with the $y[-1]$ and $y[-2]$ as the ZIR.

Exercise Grind through the math and figure out what $y[n]$. Build some character.

Exercise Use the properties you derived in ps13, problem 2 to find an expression for $y[n]$ in the following equation:

$$y[n+2] - 3y[n+1] + 2y[n] = x[n+2] - x[n+1]$$

What does this tell you about the solution technique that you use?

4 Transfer Function, Unit Sample Response, Unit Step Response

Continuing the above example, to find the unit sample response, we first kill the initial conditions and note that $x[-1] = 0$. That leaves:

$$\frac{Y(z)}{X(z)} = \frac{1 - z^{-1}}{1 - 3z^{-1} + 2z^{-2}}$$

which we recognize as the transfer function. Why is this true?

Well, we're assuming the the LDE [linear difference equation] describes some LTI system. The output $y[n]$ of this LTI system is related to the input $x[n]$ by the equation $y[n] = x[n] * h[n]$, where $h[n]$ is the impulse response of the system. If we take the Z transform of the equation, we end up with $Y(z) = H(z)X(z)$, from which we find $H(z) = \frac{Y(z)}{X(z)}$. This is the fact we used above.

Note that we have the transfer function $H(z)$ is found by assuming zero initial conditions, just as in CT.

Now, $X(z) = 1$ since we're looking for the unit sample response or the impulse response [in the time domain, that means we have as input $x[n] = \delta[n]$ and we're trying to find $y[n]$]. So

$$Y(z) = \frac{1 - z^{-1}}{1 - 3z^{-1} + 2z^{-2}}$$

We then factor the numerator and denominator and apply the various partial fraction tricks that we learned previously. Sometimes it is more useful to use the heaviside expansion theorem [a fancy name for "partial fraction tricks" that you can use to impress your friends at cocktail parties] with z^{-1} as the variable for which substitution should be performed, eg:

$$\frac{1}{1-3z^{-1}+2z^{-2}} = \frac{A}{1-2z^{-1}} + \frac{B}{1-z^{-1}}$$

Multiply both sides by $1-2z^{-1}$ and substitute $z^{-1} = \frac{1}{2}$ to obtain $A = 2$. Multiply both sides by $1-z^{-1}$ and substitute $z^{-1} = 1$ to obtain $B = -1$.

Some people prefer to multiply by enough $\frac{z}{z}$ terms to get rid of all negative powers of z and then perform the partial fraction expansion.

However, the partial fraction expansion only works if the degree of the numerator is less than the degree of the denominator. If this is not so, then you have to long divide the numerator by the denominator and then transform whatever you have left.

For causal systems, in the worst case, the degree of the numerator can be equal to the degree of the denominator, but we can just factor out a z term from the numerator, do the partial fraction expansion, and then put the z back in.

Sometimes, the numerator breaks more elegantly; sometimes, the denominator is well-known so that you don't have to do this [cf ps13, 6cd].

To find the unit step response, let $X(z) = \frac{1}{1-z^{-1}}$. This might be kind of messy though. For some simpler systems, it is perhaps more useful to consider the DT convolution $y[n] = u[n] * h[n]$ instead.

5 Finding Poles and Zeros

Take the transfer function and multiply it by sufficient numbers of $\frac{z}{z}$ factors until all the annoying $z^{-\text{whatevers}}$ factors disappear. That way you're guaranteed that you find all the poles and zeros.

Proof by example: it's not readily apparent that $H_1(z) = \frac{1}{1+\frac{1}{2}z^{-1}}$ has a zero at $z = 0$.

Note that for a stable causal DT system, the poles should all be inside the unit circle. The ROC will then be all z with a magnitude outside the outermost pole.

6 Determining Magnitude and Phase Response

Recall that when we were looking for the frequency response of a CT system with transfer function $H(s)$, we said that if the axis was in the ROC, the frequency response was $H(s)$ evaluated on the $j\omega$ axis. This was another way to arrive at the Fourier transform of the impulse response.

Similarly, when we look for the frequency response of a DT system with transfer function $H(z)$, if the unit circle is in the ROC [or the outermost pole is contained in the unit circle], then $H(z)$ evaluated on the unit circle gives you the frequency response. But this is just the corresponding DTFT.

Consider the system with Z transform:

$$\begin{aligned} H_1(z) &= \frac{1}{1+\frac{1}{2}z^{-1}} \\ &= \frac{z}{z+\frac{1}{2}} \end{aligned}$$

This has a pole at $-\frac{1}{2}$ and a zero at 0.

If we restrict ourselves to the unit circle:

$$\begin{aligned} H_1(e^{j\omega T}) &= \frac{e^{j\omega T}}{e^{j\omega T}+\frac{1}{2}} \\ |H_1(e^{j\omega T})|e^{j\angle H_1(e^{j\omega T})} &= \frac{e^{j\omega T}}{|e^{j\omega T}+\frac{1}{2}|e^{j\angle(e^{j\omega T}+\frac{1}{2})}} \end{aligned}$$

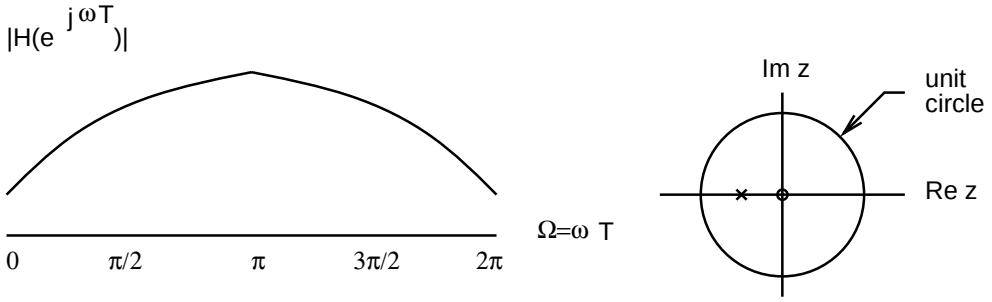

Figure 1: High pass filter: pole zero diagram and magnitude response.

where we have rewritten everything in polar form. Equating the magnitudes and the phases gives us:

$$\begin{aligned} |H_1(e^{j\omega T})| &= \frac{1}{|e^{j\omega T} + \frac{1}{2}|} \\ \angle H_1(e^{j\omega T}) &= \omega T - \angle(e^{j\omega T} + \frac{1}{2}) \end{aligned}$$

This looks like a mess.

Consider a testpoint on the unit circle. What is this testpoint? It's the point $e^{j\omega T}$. If you write $e^{j\omega T}$ in polar, you're going to get $e^{j\omega T}$, implying unit magnitude; this is why the testpoint is on the unit circle. Its angle is ωT.

We can then interpret the magnitude as the length of the vector from the zero to a testpoint on the unit circle at angle ωT divided by the length of the vector from the pole to the testpoint. Similarly, the angle is the angle that the vector from the zero to the testpoint forms with respect to the horizontal line, minus the angle that the vector from the pole to the testpoint forms with respect to the horizontal line.

So if we restrict ourselves to the unit circle, draw vectors from the pole and the zero, and whip around on the circle by changing the value of ωT from 0 to 2π, the magnitude response is as sketched in Figure 1.

The main idea here is that the magnitude (and the phase response) are going to be periodic in frequency. Consider restricting yourself to the unit circle; you will come back to the same place every $\frac{2\pi}{T}$ in unnormalized frequency. For this particular system, the low frequencies are attenuated compared to the high frequencies, so this thing is a high pass filter.

Exercise Repeat the above analysis for normalized frequency. Make sure that your magnitude and phase response is periodic with period 2π.

Now consider the system with Z transform:

$$\begin{aligned} H_2(z) &= \frac{1}{1 - \frac{1}{2}z^{-1}} \\ &= \frac{z}{z - \frac{1}{2}} \end{aligned}$$

This has pole zero diagram and magnitude response as sketched in Figure 2. It's a low pass filter.

Let's check out the sequences that correspond to these two transfer functions.

$$\begin{aligned} h_1[n] &= (-\frac{1}{2})^n u[n] \\ h_2[n] &= (\frac{1}{2})^n u[n] \end{aligned}$$

This is cool. If i alternate the sign on the sequence, i can turn a low pass filter into a high pass filter. Why does this work? Using the DTFT frequency shift property, we find that multiplying by $(-1)^n$ corresponds to a frequency shift of $\frac{\pi}{T}$.

Check out OWY pages 646-648 for some much better pics on how to find phase and magnitude response.

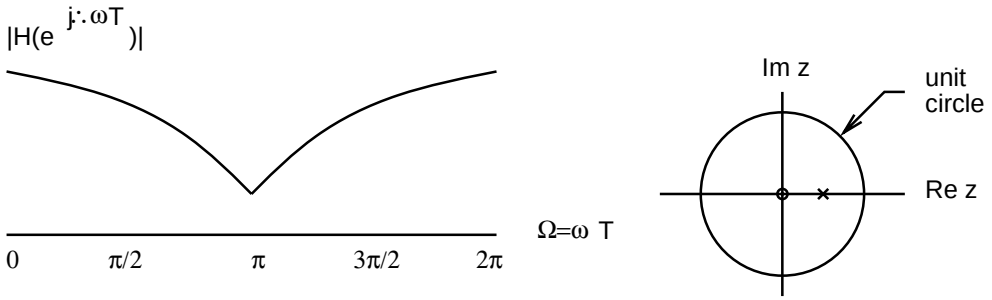

Figure 2: Low pass filter: pole zero diagram and magnitude response.

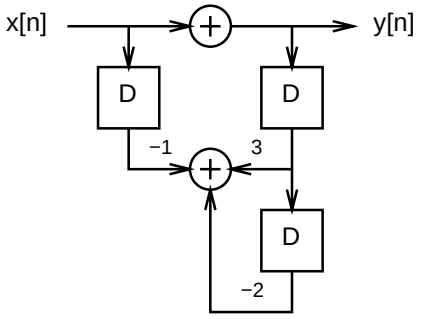

Figure 3: A direct form I implementation.

7 Block Diagrams for DT Systems

The trick here for DT systems is to notice that passing $x[n]$ into a delay block D produces $x[n-1]$.

Back to the same equation from above:

$$y[n] - 3y[n-1] + 2y[n-2] = x[n] - x[n-1]$$

Rewrite this as:

$$y[n] = 3y[n-1] - 2y[n-2] + x[n] - x[n-1]$$

Using that equation, draw the system around the uppermost summer, as in Figure 3. This is the direct form I implementation of this system.

Exercise Convert this to direct form II. Prove that the direct form I and direct form II are equivalent. See OWY page 116 for block diagrams of these two forms.

These ideas can be utilized also for CT systems by taking the relevant linear differential equation, integrating to get rid of the differentiations, and then translating the resulting equation into a direct form I or direct form II realization.

Exercise Prove that the direct form I and direct form II for CT systems on OWY page 119 are equivalent. [Hint: replace integrators by $\frac{1}{s}$.]

8 FIR Filters v. IIR Filters

Why bother having two types of filters? Why not choose one over the other? Well, each type has its own strengths and weaknesses. The type which best fits one situation may not be the one that you would end up choosing next time.

The general linear difference equation for FIR filters is:

$$a_0 y[n] = \sum_{j=0}^{M} b_j x[n-j]$$

FIR stands for "finite-duration impulse response". One way to see that the impulse response is finite is to just put in $x[n] = \delta[n]$. Because there are at most $M + 1$ terms on the right hand side, after $M + 1$ steps the impulse falls out of the chain of delays and the output goes to zero.

The general linear difference equation for IIR filters is:

$$\sum_{i=0}^{N} a_i y[n-i] = \sum_{j=0}^{M} b_j x[n-j]$$

IIR stands for "infinite-duration impulse response". This is because of the recursion implied by all the $y[n-i]$ terms on the left hand side— if you put an impulse in, you're going to have to wait a long time before the output goes to zero [assuming the resulting system is stable, of course].

Anyway, FIR filters have some nice properties:

- they are always stable, because all their poles are at the origin (take Z transform of the FIR general form above, multiply through by $\frac{z^M}{z^M}$) [as opposed to IIR filters, which may or may not be stable].
- they can be used to construct filters that have no CT equivalents [four of the popular IIR filter design techniques that we have seen rely on the existence of CT equivalents; if you're interested in these, look up "impulse invariance", "box integration", "trapezoidal integration", or "backward difference" in a DSP textbook].
- they can be made quite easily to have linear phase [see next section].

IIR filters have some nice properties:

- the equivalent FIR filter would have many more filter coefficients (a large number of b_j) than the equivalent IIR filter (a smaller number of a_i and b_j); an equivalent statement is:

$$\text{FIR order} M > \text{IIR order} N + \text{IIR order} M$$

- fewer filter coefficients means fewer multiplications and additions, which means higher throughput
- fewer filter coefficients means smaller rom sizes, which means cheaper filters

9 Linear Phase for FIR Filters

Why bother with linear phase? Well, in CT, when you say that a filter with frequency response $H(\omega)$ has linear phase, it means that $\angle H(\omega)$ is a linear function of ω, or in symbols:

$$\angle H(\omega) = -\omega T$$

Note the negative sign. If this is true, then $H(\omega) = |H(\omega)|e^{-j\omega T}$. This last factor corresponds to a delay in time by T. If there was no negative sign, we would have an advance in time, which has the potential of causing noncausality.

What does this linear phase imply? The best argument that i have run across is the graphical one on OWY page 227. The gist of the argument is that nonlinear phase leads to distortion of the output signal in the time domain. Linear phase distorts everything linearly, so that the output signal ends up being a delayed version of the input signal.

Well, all that motivation was nice, but what are the conditions under which FIR filters have linear phase?

Let's consider an FIR filter with a real impulse response $h[n]$ defined as:

$$h[n] = \begin{cases} h[M-n] & \text{where } M \text{ is an even integer and } 0 \le n \le M \\ 0 & \text{otherwise} \end{cases}$$

We can rewrite h[n] as:

$$h[n] = h_{even}[n] * \delta[n - M/2]$$

where $h_{even}[n]$ is:

$$h_{even}[n] = \begin{cases} h[n - \frac{M}{2}] & \text{where } M \text{ is an even integer and } -\frac{M}{2} \le n \le \frac{M}{2} \\ 0 & \text{otherwise} \end{cases}$$

Note that h_{even} is both real and even.

Taking the DTFT, we get:

$$H(e^{j\omega T}) = H_{even}(e^{j\omega T})e^{-j\omega TM/2}$$

Since $h_{even}[n]$ is real and even, then $H_{even}(e^{j\omega T})$ is real and even [symmetry properties that you know from the FT also apply to the DTFT]. The only contribution we have to phase [modulo changes of sign in $H_{even}(e^{j\omega T})$] is then the $e^{-j\omega TM/2}$ term, so consequently the phase is linear.

The above condition on $h[n]$ is one of the conditions under which an FIR filter can have linear phase.

Exercise Prove that:

$$h[n] = \begin{cases} -h[M-n] & \text{where } M \text{ is an even integer and } 0 \leq n \leq M \\ 0 & \text{otherwise} \end{cases}$$

is another such condition.

10 FIR Filter Design Techniques

So far we have seen two FIR filter design methods: windowing and [for lack of a better name] the inverse DTFT.

We have seen one version of windowing, and that windowing was performed with a rectangular window. This corresponds to taking an impulse response that you like, and chopping off the tail after a certain point. In lecture, we considered the DTFT of a delayed pulse in DT, and examined its corresponding DTFT [recall that the DTFT is what you get when you take the Z transform and evaluate it by running around on the unit circle]. Note that the frequency response was kind of ripply. Where does that rippling come from?

In general, the FIR windowing technique can be thought of trying to construct $h[n] = h_d[n]w[n]$, where $h_d[n]$ is the impulse response that you desire, and $w[n]$ is the shape of the window. In our case above, the truncating corresponds to multiplying our desired impulse response by a rectangular window. Well, in the frequency domain, multiplication in time is equivalent to convolution in frequency. But here the catch is that we only need to convolve over one period [the DTFT is periodic in frequency— it doesn't really make much sense to convolve over more than one period, because then what we're going to get is going to blow up]. So the ripply looking $H(e^{j\omega T})$ is the periodic convolution of $H_d(e^{j\omega T})$ and $W(e^{j\omega T})$. i'm not going to sketch this out, because my eyeballs are dropping out of my head right now. Drop by during office hours if you want to talk about it.

The other method that we discussed was this: if somebody gives you $H(e^{j\omega T})$ and says, "Hey! Your money or $h[n]$!" All you have to do is take the inverse DTFT:

$$h[n] = \frac{T}{2\pi}\int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} H(e^{j\omega T})e^{jn\omega T}d\omega$$

This formula is also repeated on OWY page 308.

The remez exchange algorithm is a way of designing fir filters without too much pain and effort, mainly because it has already been implemented in various packages, such as matlab. And that's all i want to say about it, mainly because explaining it is going to take an entire lecture in ee123.

In otherwords, the interested reader is strongly encouraged to take ee123.[1]

[1] ee123 discusses DT in great depth, spending one entire semester on what we've covered in two weeks. Call telebears now.