Lecture 4

The z transform

Introduction

The *z* transform is a discrete version of the Laplace transform. Recall that the Laplace transform derives its usefulness from the fact that a linear differential equation with constant coefficients, such as

$$\ddot{x}+b\,\dot{x}+c\,x=0 \tag{1}$$

always has solutions of the form $x(t)=A\,e^{st}$. Since $\frac{d}{dt}e^{st}=s\,e^{st}$ (1) becomes

$$s^2+b\,s+c=0 \tag{2}$$

We have converted a differential equation in *t* into a polynomial equation in *s*.

The "two-sided" Laplace transform of a signal $x(t)$ is

$$X(s)=\int_{-\infty}^{\infty} x(t)\,e^{-st}\,dt \tag{3}$$

If $x(t)$ is a causal signal, $x(t<0)\equiv 0$, then

$$X(s)=\int_{0}^{\infty} x(t)\,e^{-st}\,dt \tag{4}$$

which is the "normal" Laplace transform.

Motivation

The behavior of any LTI system is fully characterized by its impulse response $h(n)$. The output $y(n)$ is obtained from the input $x(n)$ by the convolution

$$y(n)=\sum_{k=-\infty}^{\infty} h(k)\,x(n-k) \tag{5}$$

or, equivalently

$$y(n)=\sum_{k=-\infty}^{\infty} x(k)\,h(n-k) \tag{6}$$

For an arbitrary input *x*, the relation between input and output is literally "convoluted.

However, there is one type of input for which the convolution results in the input merely being multiplied by a constant. Let $z=a+j\,b$ be any complex number, and let $x(n)=z^n$ be the system input. The output is

$$y(n)=\sum_{k=-\infty}^{\infty} h(k)\,z^{n-k}=z^n\left[\sum_{k=-\infty}^{\infty} h(k)\,z^{-k}\right] \tag{7}$$

This is just the input $x(n)=z^n$ times the number

$$H(z) = \sum_{k=-\infty}^{\infty} h(k) z^{-k} \tag{8}$$

We say that z^n is an *eigenfunction* ("characteristic function") of the system. It passes through the system unchanged except for multiplication by a constant. This constant, $H(z)$, we call the *z-transform* of the impulse response $h(n)$ (Fig. 1).

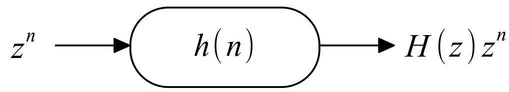

Fig. 1: Exponentials are eigenfunctions of an LTI system.

Of course, $H(z)$ is a constant only for a constant value of *z*. If we let *z* vary, then $H(z)$ is a function of this variable.

Using the superposition principle, if $x(n)$ is any linear combination of exponential functions

$$x(n) = A_1 z_1^n + A_2 z_2^n + \cdots \tag{9}$$

the output is simply

$$y(n) = A_1 H(z_1) z_1^n + A_2 H(z_2) z_2^n + \cdots \tag{10}$$

Nearly all the discrete signals we are interested in *can* be written as a linear combination of exponential functions.

Geometric series

The geometric series plays a central role in application of the *z* transform. Consider the following identity

$$1 + a + a^2 + a^3 = \frac{1-a^4}{1-a} \tag{11}$$

This can be verified by clearing fractions.

$$(1-a)(1+a+a^2+a^3) = (1+a+a^2+a^3) - (a+a^2+a^3+a^4) = 1-a^4 \tag{12}$$

It generalizes to

$$\sum_{k=0}^{N-1} a^k = \frac{1-a^N}{1-a} \tag{13}$$

If $|a| < 1$ then $a^N \to 0$ as $N \to \infty$ and we define

$$\sum_{k=0}^{\infty} a^k = \lim_{N\to\infty} \sum_{k=0}^{N-1} a^k = \lim_{N\to\infty} \frac{1-a^N}{1-a} = \frac{1}{1-a} \quad \text{if } |a| < 1 \tag{14}$$

This allows us to calculate the *z*-transform of a signal

$$x(n) = a^n u_s(n) \tag{15}$$

as

$$X(z)=\sum_{k=0}^{\infty} a^k z^{-k}=\sum_{k=0}^{\infty}(a z^{-1})^k=\frac{1}{1-a z^{-1}} \quad \text{if } |az^{-1}|<1 \rightarrow |z|>|a| \tag{16}$$

We can express this *z-transform pair* as

$$x(n)=a^n u_s(n) \Leftrightarrow X(z)=\frac{1}{1-a z^{-1}} \tag{17}$$

This gives us two representations of a signal: one, $x(n)$, in the *time domain*, and one, $X(z)$, in the *z domain*. This is analogous to how the Laplace transform allows us to represent an analog signal as $x(t)$ in the time domain and as $X(s)$ in the *s* domain.

Sinusoids

Any complex number can be expressed in polar format

$$z=a+jb=r e^{j\omega}=r\cos(\omega)+j r\sin(\omega) \tag{18}$$

Raising this to the power *n* we have

$$z^n=r^n e^{j\omega n}=r^n\cos(\omega n)+j r^n\sin(\omega n) \tag{19}$$

It follows that we can write

$$A r^n\cos(\omega n+\phi)=\text{Re}\{A e^{j\phi} z^n\} \ , \ z=r e^{j\omega} \tag{20}$$

The transfer function

The input-output relation for any LTI system is completely characterized by the convolution

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

Taking the z transform of both sides we have

$$\sum_{n=-\infty}^{\infty} y(n)z^{-n}=\sum_{n=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} h(k)x(n-k)z^{-n} \tag{21}$$

The left side is simply $Y(z)$. On the right side we write $z^{-n}=z^{-k}z^{-(n-k)}$ to get

$$Y(z)=\sum_{n=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} h(k)x(n-k)z^{-k}z^{-(n-k)} \tag{22}$$

The change of variable $m=n-k$ results in

$$Y(z)=\sum_{k=-\infty}^{\infty} h(k)z^{-k}\sum_{m=-\infty}^{\infty} x(m)z^{-m}=H(z)X(z) \tag{23}$$

We have two ways to describe the behavior of a LTI discrete system. One is a convolution in the time domain; the other is a product in the *z* domain.

$$y(n)=h(n)*x(n) \Leftrightarrow Y(z)=H(z)X(z) \tag{24}$$

The function $H(z)$, the *z* transform of the impulse response, is called the *transfer function*.

System realization

In principle we can implement ("realize") any LTI system using the impulse response $h(n)$ and the convolution (5). If the transfer function $H(z)$ is a rational function of z^{-1} or z then it also suggests a direct realization. This is due to the following interpretation of the factors z^{-1} and z.

Delay and advance operators

A system with impulse response

$$h(n)=\delta(n-1)$$

produces output

$$y(n)=h(n)*x(n)=x(n-1)$$

which is simply the input delayed by one sample. The transfer function is

$$H(z)=\sum_{k=-\infty}^{\infty}\delta(k-1)z^{-k}=z^{-1}$$

Therefore, we can think of z^{-1} as the *delay operator*. If we see $z^{-1}X(z)$ in a *z*-domain expression we know it corresponds $x(n-1)$ in the time domain. This can easily be verified

$$\sum_{k=-\infty}^{\infty}x(k-1)z^{-k}=\sum_{m=-\infty}^{\infty}x(m)z^{-(m+1)}=z^{-1}\sum_{m=-\infty}^{\infty}x(m)z^{-m}=z^{-1}X(z)$$

Since $z^{-2}=z^{-1}z^{-1}$, $z^{-2}X(z)$ corresponds to $x(n-2)$, and, in general $z^{-k}X(z)$ corresponds to $x(n-k)$. The delay operator is causal and can be implemented in real time because it does not refer to the future.

Now consider a system with impulse response

$$h(n)=\delta(n+1)$$

This produces output

$$y(n)=h(n)*x(n)=x(n+1)$$

which is simply the input advanced by one sample. The transfer function is

$$H(z)=\sum_{k=-\infty}^{\infty}\delta(k+1)z^{-k}=z$$

We can think of z as the *advance operator*. The *z*-domain expression $z^{k}X(z)$ corresponds to the time-domain expression $x(n+k)$. The advance operator is *not* causal and *cannot* be implemented in real time because it refers to the future.

Implementing a rational transfer function

Suppose we have a transfer function

$$H(z)=\frac{b_0+b_1z^{-1}+b_2z^{-2}}{1+a_1z^{-1}+a_2z^{-2}} \tag{25}$$

Since $Y(z)=H(z)X(z)$ we can write

$$\left[1+a_1 z^{-1}+a_2 z^{-2}\right]Y(z)=\left[b_0+b_1 z^{-1}+b_2 z^{-2}\right]X(z) \tag{26}$$

Because z^{-1} is the delay operator, the corresponding time-domain expression is

$$y(n)+a_1 y(n-1)+a_2 y(n-2)=b_0 x(n)+b_1 x(n-1)+b_2 x(n-2) \tag{27}$$

Solving for $y(n)$ we find

$$y(n)=b_0 x(n)+b_1 x(n-1)+b_2 x(n-2)-a_1 y(n-1)-a_2 y(n-2) \tag{28}$$

This is a formula for the current output as a linear combination of a finite number of input and previous output values.

Generalizing this, the transfer function

$$H(z)=\frac{b_0+b_1 z^{-1}+b_2 z^{-2}+\cdots+b_M z^{-M}}{1+a_1 z^{-1}+a_2 z^{-2}+\cdots+a_N z^{-N}} \tag{29}$$

has time-domain realization

$$y(n)=\sum_{k=0}^{M} b_k x(n-k)-\sum_{k=1}^{N} a_k y(n-k) \tag{30}$$

This is easily programmable. For instance

```
y(n) = b0*x(n);
for k=1:M
  y(n) = y(n)+b(k)*x(n-k);
end
for k=1:N
  y(n) = y(n)-a(k)*y(n-k);
end
```

Now let's consider a subtle point. Multiplying the numerator and denominator of (25) by z^2 we get

$$H(z)=\frac{b_0 z^2+b_1 z+b_2}{z^2+a_1 z+a_2} \tag{31}$$

We have converted a rational function of z^{-1} into a rational function of *z*. For any finite, non-zero value of *z* (25) and (31) will give the same value for $H(z)$, and in principle we can use either one to represent the system. However, let's look at the time-domain form of (31). We have

$$\left[z^2+a_1 z+a_2\right]Y(z)=\left[b_0 z^2+b_1 z+b_2\right]X(z) \tag{32}$$

Since *z* is the advance operation we have

$$y(n+2)+a_1 y(n+1)+a_2 y(n)=b_0 x(n+2)+b_1 x(n+1)+b_2 x(n) \tag{33}$$

This is a non-causal expression in which the current output $y(n)$ depends on future inputs and outputs. Of course we can simply subtract 2 from every index to convert this to the causal expression (27). This is simply a "relabeling of the time axis," and it corresponds to multiplying both sides of (32) by z^{-2}.

The point is that a transfer function expressed as a rational function of z^{-1} (29) leads directly to a causal realization (30). If we express the transfer function as a rational function of z it does not directly lead to a causal realization, although it can be manipulated into one. For this reason we will almost always work with transfer functions expressed in terms of negative powers of z. This "looks weird" since in our continuous-time theory courses we've become used to rational transfer functions with non-negative powers of a variable (usually s).

This confronts us with a wrinkle. Many computer tools (including Scilab) will automatically convert an expression into a form with non-negative powers of the variable. Suppose

$$H(z)=\frac{1+0.5z^{-1}+0.25z^{-2}}{1-0.3z^{-1}-0.1z^{-2}} \tag{34}$$

Entering this symbolic expression in Scilab we get

```
--> z = poly(0,'z'); //make z a symbolic variable

--> H = (1+0.5*z^(-1)+0.25*z^(-2))/(1-0.3*z^(-1)-0.1*z^(-2))

 H  =
                     2
   0.25 + 0.5z + z
   ----------------
                     2
   -0.1 - 0.3z + z
```

Scilab has expressed this as a rational function in non-negative powers of z (it's multiplied numerator and denominator by z^2). We can get around this with a change of variable. We write

$$H(z)=\frac{1+0.5w+0.25w^2}{1-0.3w-0.1w^2} \quad , \quad w=z^{-1} \tag{35}$$

In Scilab

```
--> w = poly(0,'w');

--> H = (1+0.5*w+0.25*w^2)/(1-0.3*w-0.1*w^2)

 H  =
                      2
   1 + 0.5w + 0.25w
   -----------------
                     2
   1 - 0.3w - 0.1w
```

and our expression has the desired form.

Important transform pairs

Let's work out the z transforms of some important signals. The impulse (delta function) is very simple. If

$$x(n)=\delta(n) \tag{36}$$

then

$$X(z)=\sum_{k=-\infty}^{\infty} x(k)z^{-k}=1 \tag{37}$$

We can express this *z* transform pair as

$$x(n)=\delta(n)\Leftrightarrow X(z)=1 \tag{38}$$

For a delayed impulse

$$x(n)=\delta(n-k) \tag{39}$$

we have

$$X(z)=\sum_{i=-\infty}^{\infty} \delta(i-k)z^{-i}=z^{-k} \tag{40}$$

So

$$x(n)=\delta(n-k)\Leftrightarrow X(z)=z^{-k} \tag{41}$$

For a causal exponential sequence

$$x(n)=a^n u_s(n) \tag{42}$$

we can evaluate

$$X(z)=\sum_{n=0}^{\infty} a^n z^{-n}=\sum_{n=0}^{\infty}(a/z)^n \tag{43}$$

using (14) to obtain

$$X(z)=\frac{1}{1-a/z} \tag{44}$$

(valid if $|a/z|<1$). We express this as the transform pair

$$x(n)=a^n u_s(n)\Leftrightarrow X(z)=\frac{1}{1-a z^{-1}} \tag{45}$$

This is arguably the most important *z* transform pair in this course.
Writing this as

$$(1-a z^{-1})^{-1}=\sum_{n=0}^{\infty} a^n z^{-n} \tag{46}$$

and taking the *z* derivative of both sides we obtain

$$-(1-a z^{-1})^{-2} a z^{-2}=\sum_{n=0}^{\infty} -n a^n z^{-n-1} \tag{47}$$

Multiplying both sides by $-z/a$ results in

$$(1-a z^{-1})^{-2} z^{-1}=\sum_{n=0}^{\infty} n a^{(n-1)} z^{-n} \tag{48}$$

This gives us the transform pair

$$x(n)=n a^{n-1} u_s(n)\Leftrightarrow X(z)=\frac{z^{-1}}{(1-a z^{-1})^2} \tag{49}$$

Consider the most general exponentially growing/decaying sinusoid that "turns on" at $n=0$

$$x(n)=a^n\cos(\omega n+\phi)u_s(n) \tag{50}$$

Writing this as

$$x(n)=a^n\cos(\omega n+\phi)u_s(n)=\left[\frac{1}{2}e^{j\phi}\left(ae^{j\omega}\right)^n+\frac{1}{2}e^{-j\phi}\left(ae^{-j\omega}\right)^n\right]u_s(n) \tag{51}$$

we can use (45) to write

$$X(z)=\frac{1}{2}e^{j\phi}\frac{1}{1-ae^{j\omega}z^{-1}}+\frac{1}{2}e^{-j\phi}\frac{1}{1-ae^{-j\omega}z^{-1}} \tag{52}$$

Putting both terms over a common denominator

$$X(z)=\frac{1}{2}e^{j\phi}\frac{1-ae^{-j\omega}z^{-1}}{(1-ae^{j\omega}z^{-1})(1-ae^{-j\omega}z^{-1})}+\frac{1}{2}e^{-j\phi}\frac{1-ae^{j\omega}z^{-1}}{(1-ae^{-j\omega}z^{-1})(1-ae^{j\omega}z^{-1})} \tag{53}$$

and doing some algebra results in

$$X(z)=\frac{1}{2}\frac{e^{j\phi}+e^{-j\phi}-a\left(e^{-j(\omega-\phi)}+e^{j(\omega-\phi)}\right)z^{-1}}{1-2a\cos\omega z^{-1}+a^2z^{-2}}=\frac{\cos\phi-a\cos(\omega-\phi)z^{-1}}{1-2a\cos\omega z^{-1}+a^2z^{-2}} \tag{54}$$

This is an important, but rather formidable transform pair

$$x(n)=a^n\cos(\omega n+\phi)u_s(n)\Leftrightarrow X(z)=\frac{\cos\phi-a\cos(\omega-\phi)z^{-1}}{1-2a\cos\omega z^{-1}+a^2z^{-2}} \tag{55}$$

The special case $\phi=0$ reduces to

$$x(n)=a^n\cos(\omega n)u_s(n)\Leftrightarrow X(z)=\frac{1-a\cos\omega z^{-1}}{1-2a\cos\omega z^{-1}+a^2z^{-2}} \tag{56}$$

Since $\cos(x-\pi/2)=\sin x$, setting $\phi=-\pi$ results in

$$x(n)=a^n\sin(\omega n)u_s(n)\Leftrightarrow X(z)=\frac{a\sin\omega z^{-1}}{1-2a\cos\omega z^{-1}+a^2z^{-2}} \tag{57}$$

Taylor series method

For a causal signal ($n<0\rightarrow h(n)=0$), $H(z)=h(0)+h(1)z^{-1}+h(2)z^{-2}+h(3)z^{-2}+\cdots$ has the form of a Taylor series expansion of $H(z)$ in positive powers of (z^{-1}). It follows for a causal system that if we compute the Taylor series of $H(z)$ in terms of the variable (z^{-1}), the series coefficients will be the impulse response values. Suppose we have

$$H(z)=\frac{b_0+b_1z^{-1}}{1+a_1z^{-1}+a_2z^{-2}} \tag{58}$$

and we want to calculate $h(0),h(1),h(2),h(3),h(4)$. As previously mentioned, working with the variable z^{-1} is awkward, so let's define $w=z^{-1}$. The transfer function is then

$$\frac{b_0+b_1w}{1+a_1w+a_2w^2}=\frac{b_0+b_1w}{1-u} \tag{59}$$

where $u=-(a_1 w+a_2 w^2)$. Since

$$\frac{1}{1-u}=1+u+u^2+u^3+u^4+\cdots \tag{60}$$

(for $|u|<1$) we have

$$\frac{b_0+b_1 w}{1+a_1 w+a_2 w^2}=[b_0+b_1 w][1+u+u^2+u^3+u^4+\cdots] \tag{61}$$

The lowest power of w in u^k is w^k (why?). So, to get a series accurate up to and including $h(k)$ we calculate

$$\frac{b_0+b_1 w}{1+a_1 w+a_2 w^2}\approx[b_0+b_1 w][1+u+u^2+u^3+u^4+\cdots+u^k] \tag{62}$$

Let's see how this works with an example.

Example 1: For $H(z)=\dfrac{1-z^{-1}}{1-0.1z^{-1}-0.72z^{-2}}$ use the Taylor series method to find $h(0),h(1),h(2)$.

Let $w=z^{-1}$ and $u=0.1w+0.72w^2$. Then to 2nd order in w

$$\begin{aligned}H(w^{-1})&=(1-w)(1+u+u^2)\\&=(1-w)[1+(0.1w+0.72w^2)+(0.1w+0.72w^2)^2]\\&=(1-w)[1+0.1w+0.72w^2+0.01w^2+0.144w^3+0.5184w^4]\end{aligned}$$

Expanding and reducing

$$\begin{aligned}H(w^{-1})&=(1-w)[1+0.1w+0.73w^2+0.144w^3+0.5184w^4]\\&=[1+0.1w+0.73w^2+0.01w^2+0.144w^3+0.5184w^4]\\&\quad-[w+0.1w^2+0.73w^3+0.01w^4+0.144w^5+0.5184w^6]\\&=1-0.9w+0.63w^2+\cdots\end{aligned}$$

and $h(0)=1,h(1)=-0.9,h(2)=0.63$. We can do the calculations with Scilab

```
-->w = poly(0,'w'); //defines w to be a symbolic variable

-->u = 0.1*w+0.72*w^2;

-->H = (1-w)*(1+u+u^2)
 H  =
                       2          3           4           5
    1 - 0.9w + 0.63w - 0.586w + 0.3744w - 0.5184w
```

Try this for yourself.

Exercise 1: For $H(z)=\dfrac{1+z^{-1}}{1-0.25z^{-1}-0.125z^{-2}}$ use the Taylor series method to find $h(0),h(1),h(2)$. First perform the calculations by hand. Then confirm your result with Scilab.

The Appendix lists Scilab code that uses this idea to check the correctness of a z transform or inverse z transform result.

Table of z transform pairs

$x(n)$	$X(z)$
$\delta(n)$	1
$\delta(n-k)$	z^{-k}
$u_s(n)$	$\frac{1}{1-z^{-1}}$
$a^n u_s(n)$	$\frac{1}{1-az^{-1}}$
$n a^{n-1} u_s(n)$	$\frac{z^{-1}}{(1-az^{-1})^2}$
$a^n \cos(\omega n) u_s(n)$	$\frac{1-a\cos\omega\, z^{-1}}{1-2a\cos\omega\, z^{-1}+a^2 z^{-2}}$
$a^n \sin(\omega n) u_s(n)$	$\frac{a\sin\omega\, z^{-1}}{1-2a\cos\omega\, z^{-1}+a^2 z^{-2}}$
$a^n \cos(\omega n+\phi) u_s(n)$	$\frac{\cos\phi-a\cos(\omega-\phi) z^{-1}}{1-2a\cos\omega\, z^{-1}+a^2 z^{-2}}$

Appendix – z-transform/inverse transform check with Scilab

```
//ztransCheck checks the correctness of a z-transform or inverse
//z-transform calculation. Modify the USER DEFINED sections as
//appropriate.
clear;
j = %i;
w = poly(0,'w');

////////////////////////////////////////////////////////////////////////
//USER DEFINED
nMax = 20; //check agreement of x(0),x(1),...,x(nMax)
//X is the z transform expression with w=z^(-1)
X = (1-w)*(1-exp(j*%pi/2)*w)*(1-exp(-j*%pi/2)*w);
X = X/((1-0.9*exp(j*2*%pi/3)*w)*(1-0.9*exp(-j*2*%pi/3)*w));
////////////////////////////////////////////////////////////////////////

function y = delta(n) //impulse or delta function
  if (n==0)
    y = 1;
  else
    y = 0;
  end
endfunction

function y = us(n) //unit step function
  if (n>=0)
    y = 1;
  else
    y = 0;
  end
endfunction

////////////////////////////////////////////////////////////////////////
//USER DEFINED
//x(n) is the inverse z transform of X(z)
function y = x(n)
  y = 2.61*delta(n)-1.24*delta(n-1)..
    +2.4*0.9^n*cos(2*%pi/3*n+2.3)*us(n);
endfunction
////////////////////////////////////////////////////////////////////////

u = 1-X(3);
uf = 1;
for n=1:nMax
  uf = uf+u^n;
end
Xs = X(2)*uf;
mprintf('%2s %10s %10s %10s\n','n','x(n)','x(n)TS','%err');
for n=0:nMax
  mprintf("%2d %10f %10f %10f\n",n,x(n),coeff(Xs,n),..
          100*(1-x(n)/coeff(Xs,n)));
end
```