Chapter 5

z-transform

5.1 The z-transform of sequences

Laplace transforms are used extensively to analyze continuous-time (analog) signals as well as systems that process continuous-time signals. As you may recall, the role of the Laplace transform was to represent a large class of continuous-time signals as a superposition of many simpler signals, sometimes called "basis functions" or "kernels". For the Laplace transform, the kernels were complex exponential signals of the form, e^{st} , and we represented signals for which the Laplace transform existed according to the formula

$$x(t) = \frac{1}{2\pi j} \oint_C X(s) e^{st} ds,$$

where the integral is taken as a line integral along a suitable closed contour C in the complex plane. While the integral form of the inverse Laplace transform can be a powerful tool in the analysis of continuous-time signals and systems, we can often avoid its direct evaluation by algebraically manipulating the expression for $X(s)$ such that it can be represented as a sum of terms, each of which can be immediately recognized as the Laplace transform of a known signal $x(t)$. Then, using linearity of the Laplace transform, we can construct the inverse transform, term by term. We can view the inverse Laplace transform as a way of constructing $x(t)$, piece by piece, from many (an uncountably infinite number, actually) simpler signals of the form e^{st}, where the amount of each such signal contained in the signal $x(t)$ is given by $X(s)ds$. To determine how much of each complex exponential signal e^{st} is contained in $x(t)$, we have the Laplace transform formula given by

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

For signals that are zero, for negative time, this integral can be taken over positive time, giving the one-sided, or unilateral Laplace transform,

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

For many linear time-invariant (LTI) continuous-time systems, the relationship between the input and output signals can be expressed in terms of linear constant coefficient differential equations. The one-sided Laplace transform can be a useful tool for solving these differential equations. For such systems, the Laplace transform of the input signal and that of the output signal can be expressed in terms of a "transfer function" or "system function." In fact, many of the properties, such as causality or stability, of LTI systems can be conveniently explored by considering the system function of the continous-time system. Another helpful property of the Laplace transform is that it maps the convolution relationship between the input and output signals in the time domain to a conceptually simpler multiplicative relationship. In this form, LTI systems can be thought of in terms of how they change the magnitude and phase of each of the kernel signals e^{st} individually, and then the output of the system is given by a superposition of each of these scaled kernel signals.

For discrete-time signals, we will see that an analogous relationship can be developed between signals and systems using the z-transform. The discrete-time complex exponential signal, z^n , where z is a complex number, plays a similar role to the continuous-time complex exponential signal e^{st}. We have already seen that discrete-time signals of this form play an important role in the analysis of linear, constant coefficient difference equations (LCCDEs), through the their aid in developing the characteristic equation and finding solutions to homogenous LCCDEs. There is great elegance in the mathematics linking discrete-time signals and systems through the z-transform and we could delve deeply into this theory, devoting much more time than we will be able to here. While our treatment of the z-transform will be limited in scope, we will see that it is an equally valuable tool for the analysis of discrete-time signals and systems. We will use the z-transform to solve linear constant-coefficient difference equations, as well as develop the notion of discrete-time transfer functions. We can then use it to readily compute convolution and to analyze properties of discrete-time linear shift-invariant systems.

We note that as with the Laplace transform, the z-transform is a function of a complex variable. The transform itself can also take on complex values. As a result, it is a complex function of a complex variable.

5.2 Unilateral (one-sided) z-transform

Now, we will begin our study of the z-transform by first considering the one-sided, or unilateral, version of the transform. The unilateral z-transform of a sequence $\{x[n]\}_{n=-\infty}^{\infty}$ is given by the sum

$$X(z) = \sum_{n=0}^{\infty} x[n] z^{-n} \tag{5.1}$$

for all z such that (5.1) converges. Here, z is a complex variable and the set of values of z for which the sum (5.1) converges is called the region of convergence (ROC) of the z-transform. The z-transform maps sequences to functions and their associated region of convergence, such that X (z) is the z-transform of the sequence $\{x[n]\}_{n=0}^{\infty}$. When it is clear that we are discussing sequences defined for non-negative values of the independent "time" axis, or n-axis, we will write $x[n]$ simply, and omit the brace notation $\{\}_{n=0}^{\infty}$ indicating the positive n axis. The sequences for which the z-transform is defined can be real-valued, or complex valued. Note that the summation (5.1) multiplies $x[n]$ by a complex geometric sequence of the form z^{-n}, such that the series will converge whenever $|x[n]|$ grows no faster than exponentially. The region of convergence will be all z such that the geometrically-weighted series (5.1) converges. This region will be all values of z outside of some circle in the complex z-plane of radius R, the "radius of convergence" for the series (5.1) as depicted in Figure (5.1).

When we call $X(z)$ the transform of the sequence $\{x[n]\}_{n=0}^{\infty}$, we imply a form of uniqueness for the z-transform. Namely, we imply that for a given sequence $\{x[n]\}_{n=0}^{\infty}$, there exists one and only one z-transform $X(z)$ and its associated region of convergence. Similarly, for a given z-transform $X(z)$, there exists one and only one sequence $\{x[n]\}_{n=0}^{\infty}$ for which the series in (5.1)converges for $|z| > R$. The uniqueness for the z-transform derives from properties of power series expansions of complex functions of complex variables.

Example Consider the sequence $x[n] = 2^n$, defined for non-negative n as shown in Figure .

This discrete-time sequence has a z-transform given by

$$X(z) = \sum_{n=0}^{\infty} 2^n z^{-n},$$

which can be re-written as

$$X(z) = \sum_{n=0}^{\infty} \left(\frac{2}{z}\right)^n.$$

To determine the region of convergence of this z-transform, we simply need to consider the values of z for which the power series converges. This can be accomplished by recalling the method for summing an infinite geometric series. Recall that for a series of the form

$$S = \sum_{n=0}^{\infty} a^n$$

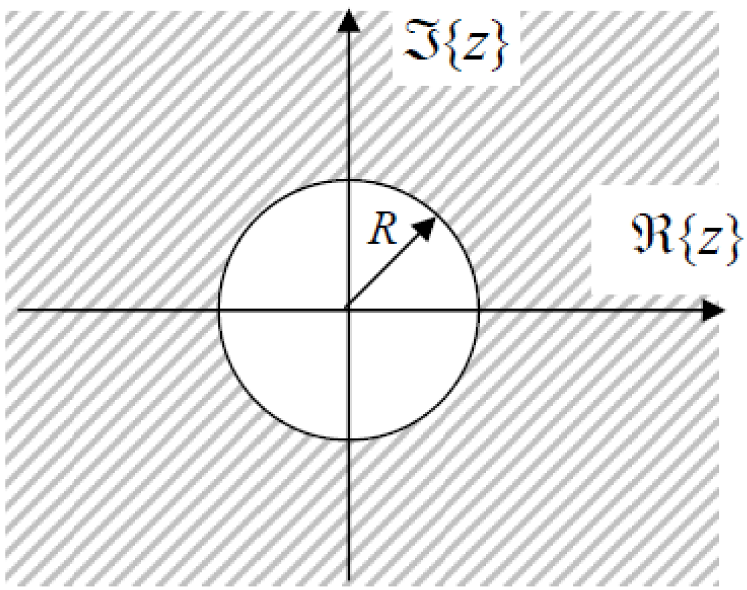

Figure 5.1: A typical region of convergence (ROC) for a unilateral z-transform. The radius of convergence, R, is shown and the ROC is all values of z such that $|z| > R$.

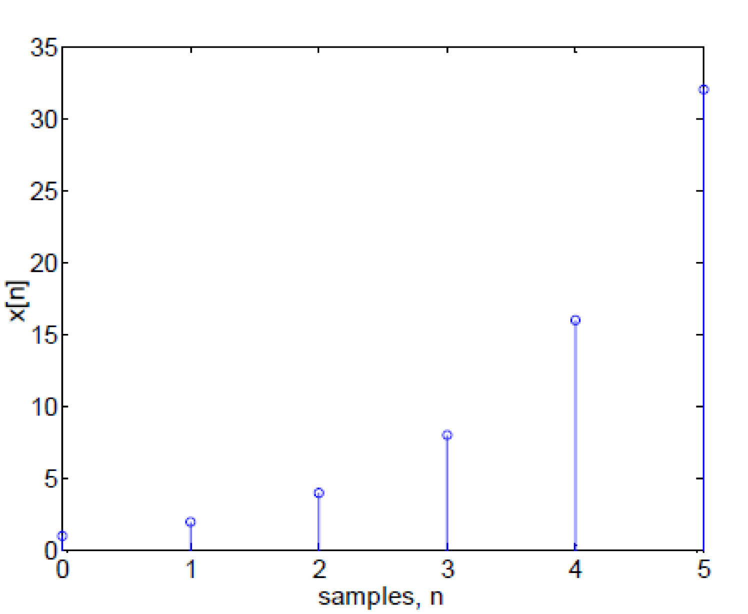

Figure 5.2: Discrete time sequence $x[n] = 2^n$ for $n \geq 0$.

where a is a complex number, we note that this is really shorthand notation for the limit

$$S = \lim_{N\to\infty} S_N = \lim_{N\to\infty} \sum_{n=0}^{N} a^n.$$

For the finite geometric series defining S_N, we write

$$S_N = (1 + a + a^2 + \ldots + a^N).$$

Since this is a finite series, we can multiply both sides by a to obtain

$$aS_N = (a + a^2 + \ldots + a^N + a^{N+1}).$$

Subtracting, we obtain

$$\begin{aligned} S_N - aS_N &= (1 - a^{N+1}) \\ S_N(1-a) &= (1 - a^{N-1}). \end{aligned}$$

Now, if $a = 1$, we know that $S_N = N + 1$. When $a \neq 1$, can divide both sides by $(1 - a)$ to obtain

$$S_N = \frac{1 - a^{N+1}}{1 - a}$$

which is valid for all $a \neq 1$. Returning to the definition of S, we have that

$$S = \lim_{N\to\infty} S_N = \lim_{N\to\infty} \frac{1 - a^{N+1}}{1 - a},$$

which will only be finite when $|a| < 1$, for which we have

$$S = \frac{1}{1-a}.$$

This is a special case of the series

$$\begin{aligned} S &= \sum_{n=N_1}^{N_2} a^n = (a^{N_1} + a^{N_1+1} + \ldots + a^{N_2}) \\ aS &= (a^{N_1+1} + a^{N_1+2} + \ldots + a^{N_2} + a^{N_2+1}), \end{aligned}$$

leading to

$$S(1-a) = (a^{N_1} - a^{N_2+1})$$

or

$$S = \frac{a^{N_1} - a^{N_2+1}}{1 - a}$$

so long as $a \neq 1$.Note that this holds even for values of a that have magnitude greater than one. When $N_2 = \infty$, we may consider

$$\lim_{N_2\to\infty} S = \lim_{N_2\to\infty} \frac{a^{N_1} - a^{N_2+1}}{1 - a} = \frac{a^{N_1}}{1-a},$$

so long as $|a| < 1$. When $N_1 = 0$, this takes the form $S = \frac{1}{1-a}$ seen above. To summarize, we have seen that

$$\boxed{\sum_{n=N_1}^{N_2} a^n = \frac{a^{N_1} - a^{N_2+1}}{1-a}, \quad \text{for } a \neq 1} \tag{5.2}$$

and

$$\boxed{\sum_{n=N_1}^{\infty} a^n = \frac{a^{N_1}}{1-a}, \quad \text{for } |a| < 1}. \tag{5.3}$$

Now returning to our example, for $x[n] = 2^n, \;\; n \geq 0$, let us find the ROC for $X(z)$, the z-transform of $x[n]$. Is $z = 1$ in the ROC of $X(z)$? Is $z = 3$ in the ROC? First consider $z = 1$.

$$\begin{aligned} X(1) &= \sum_{n=0}^{\infty} 2^n 1^{-n} \\ &= \sum_{n=0}^{\infty} 2^n, \end{aligned}$$

which clearly diverges. Therefore, $z = 1$ is not in the ROC. Now consider $z = 3$.

$$\begin{aligned} Z(3) &= \sum_{n=0}^{\infty} 2^n 3^{-n} \\ &= \sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^n \\ &= \frac{1}{1-\frac{2}{3}} \\ &= 3. \end{aligned}$$

Thus, $X(z)$ is well-defined at $z = 3$ and therefore $z = 3$ is a point in the ROC of $X(z)$.

In this example, we saw that a larger value of z was in the ROC, whereas a smaller value was not. It should not be a surprise that larger values of z are more likely to be in the ROC. Why so? Because, in the definition of the z-transform, z is raised to a negative power and multiplied by the sequence $x[n]$. Therefore, the z-transform is essentially a sum of the signal $x[n]$ multiplied by either a damped or a growing complex exponential signal z^{-n}. Thus, larger values of z offer greater likelihood for convergence of the z-transform sum, since these correspond to more rapidly decaying exponential signals. In general, $X(z)$ converges for all z that are large enough, that is, when z is sufficiently large, that the signal $x[n]z^{-n}$ becomes summable. Specifically, $X(z)$ converges for all z such that $|z| > R$ (for some R). Thus, the ROC of $X(z)$ includes all points z lying outside a circle of radius R, as illustrated in Figure 5.1. To discover the value of R for a given sequence, we need only consider the convergence test that we need to apply when we try to compute the z-transform sum.

For our example, we have

$$\sum_{n=0}^{\infty} \left(\frac{2}{z}\right)^n,$$

which, when applying the formula (5.3)for a geometric series, yields

$$\begin{aligned} X(z) &= \sum_{n=0}^{\infty} \left(\frac{2}{z}\right)^n. \\ &= \frac{1}{1-\left(\frac{2}{z}\right)}, \quad \left|\frac{2}{z}\right| < 1 \\ &= \frac{z}{z-2}, \quad |z| > 2, \end{aligned}$$

that is the ROC of $X(z)$ is $|z| > 2$. We can look at a more general example, such as that considered next.

Example

Consider the sequence $x[n] = a^n$, for $n \geq 0$, where a is a possibly complex constant. To determine $X(z)$, we consider the sum

$$\begin{aligned} X(z) &= \sum_{n=0}^{\infty} a^n z^{-n} \\ &= \sum_{n=0}^{\infty} \left(\frac{a}{z}\right)^n, \end{aligned}$$

which for $|z| > |a|$ converges to
$$X(z) = \frac{z}{z-a}.$$
Note that
$$\left|\frac{a}{z}\right| < 1 \Leftrightarrow \frac{|a|}{|z|} < 1 \Leftrightarrow |a| < |z| \Leftrightarrow |z| > |a|.$$
This, we have
$$X(z) = \frac{z}{z-a}, |z| > |a|.$$
What happens for $|z| < |a|$? Although the algebraic expression $z/(z-a)$ can be evaluated for any value of $z \neq a$, this is most certainly not the z-transform, since we know that the infinite sum defining $X(z)$ does not converge for such values of z. Therefore, $X(z)$ is defined only on its ROC and is not defined elsewhere. Hence, when we mention the z-transform of a sequence, we need to not only provide an expression for $X(z)$, but to also define the values of z for which this expression holds, i.e. the ROC.

Linearity

We can also use some elementary calculus to extend some of the relationships developed thus far. First, let us show that the z-transform is linear, that is if $X_1(z)$ is the z-transform for the sequence $x_1[n]$ and $X_2(z)$ is the z-transform for the signal $x_2[n]$, then the signal $x_3[n] = ax_1[n] + bx_2[n]$ is given by $X_3(z) = aX_2(z) + bX_1(z)$. This superposition property can be shown directly from the definition of the z-transform,

$$\begin{aligned} X_3(z) &= \sum_{n=0}^{\infty}(ax_1[n] + bx_2[n])z^{-n} \\ &= \sum_{n=0}^{\infty} ax_1[n]z^{-n} + \sum_{n=0}^{\infty} bx_2[n]z^{-n} \\ &= a\sum_{n=0}^{\infty} x_1[n]z^{-n} + b\sum_{n=0}^{\infty} x_2[n]z^{-n} \\ &= aX_1(z) + bX_2(z). \end{aligned}$$

Example

Now, to determine the z-transform of a sequence of the form $x[n] = na^n$, we can use linearity of the transform to obtain the desired result. We know that for the sequence $x[n] = a^n$ we have

$$X(z) = \sum_{n=0}^{\infty} a^n z^{-n} = \frac{z}{z-a}, |z| > |a|$$

and that if we differentiate this expression with respect to z we have

$$\begin{aligned} \frac{d}{dz}X(z) &= \frac{d}{dz}\left(\sum_{n=0}^{\infty} a^n z^{-n}\right) = \frac{d}{dz}\left(\frac{z}{z-a}\right), |z| > |a| \\ &= -\sum_{n=0}^{\infty} na^n z^{-n-1} = \frac{-a}{(z-a)^2}, |z| > |a|. \end{aligned}$$

From this expression, we can multiply by $-z$ and obtain

$$-z\frac{d}{dz}X(z) = \sum_{n=0}^{\infty} na^n z^{-n} = \frac{az}{(z-a)^2}, |z| > |a|.$$

That is, we have the relation

$$\sum_{n=0}^{\infty} na^n z^{-n} = \frac{az}{(z-a)^2}, |z| > |a|.$$

In a similar manner, we can obtain the more general result

$$nx[n] \Leftrightarrow -z\left(\frac{d}{dz}X(z)\right),$$

for $X(z)$ the z-transform of $x[n]$. We can continue to differentiate to obtain the relation

$$\frac{1}{2}n(n-1)x[n] \Leftrightarrow \frac{a^2 z}{(z-a)^3}, |z| > |a|,$$

and m-fold differentiation leads to the relation

$$\frac{1}{m!}n(n-1)\cdots(n-m+1)a^n \Leftrightarrow \frac{a^m z}{(z-a)^{m+1}}, |z| > |a|.$$

Example

We can use linearity of the z-transform to compute the z-transform of trigonometric functions, such as $x[n] = \cos(\omega n)$, for $n \geq 0$. Note that rather than using $x[n] = \cos(\omega n)u[n]$, we instead use the notation $n \geq 0$, since the unilateral z-transform for both sequences would be the same. From Euler's relation, we have

$$\begin{aligned}
X(z) &= \sum_{n=0}^{\infty} \cos(\omega n) z^{-n} \\
&= \sum_{n=0}^{\infty} \frac{1}{2}(e^{j\omega n} + e^{-j\omega n}) z^{-n} \\
&= \frac{1}{2}\sum_{n=0}^{\infty}(e^{j\omega}z^{-1})^n + \frac{1}{2}\sum_{n=0}^{\infty}(e^{-j\omega}z^{-1})^n \\
&= \frac{1}{2}\frac{1}{1-e^{j\omega}z^{-1}} + \frac{1}{2}\frac{1}{1-e^{-j\omega}z^{-1}}, \quad |z| > |e^{j\omega}| = 1 \\
&= \frac{1}{2}\frac{z}{z-e^{j\omega}} + \frac{1}{2}\frac{z}{z-e^{-j\omega}}, \quad |z| > 1 \\
&= \frac{1}{2}\left(\frac{z(z-e^{-j\omega})}{z^2 - z(e^{j\omega}+e^{-j\omega})+1} + \frac{z(z-e^{j\omega})}{z^2 - z(e^{j\omega}+e^{-j\omega})+1}\right), \quad |z| > 1 \\
&= \frac{z^2 - z\cos(\omega)}{z^2 - 2z\cos(\omega)+1}, \quad |z| > 1.
\end{aligned}$$

We could have shortened the derivation by using our knowledge that $\cos(\omega n)$ is a sum of two complex exponentials of the form a^n where $a = e^{\pm j\omega}$ and then use linearity together with our knowledge of the z-transform for a^n. Let us now use this approach to find the z-transform for $x[n] = \sin(\omega n)$. We have that

$$\begin{aligned}
x[n] &= \frac{1}{2j}\left(e^{j\omega n} - e^{-j\omega n}\right) \\
&= \frac{1}{2j}\left(e^{j\omega}\right)^n - \frac{1}{2j}\left(e^{-j\omega}\right)^n
\end{aligned}$$

to which we can apply transform pairs we already know. From the z-transform of a single complex exponential, we have

$$\begin{aligned}
X(z) &= \frac{1}{2j}\frac{z}{z-e^{j\omega}} - \frac{1}{2j}\frac{z}{z-e^{-j\omega}}, \quad |z| > 1 \\
&= \frac{1}{2j}\frac{z(z-e^{-j\omega}) - z(z-e^{j\omega})}{z^2 - 2z\cos(\omega)+1}, \quad |z| > 1 \\
&= \frac{z\sin(\omega)}{z^2 - 2z\cos(\omega)+1}, \quad |z| > 1.
\end{aligned}$$

Example

From the definition of the z-transform, it should be clear that the unit sample function, i.e. the discrete-time impulse, has a z-transform

$$\delta[n] \Leftrightarrow 1.$$

Similarly, directly from the definition of the z-transform, a discrete-time impulse at $n = k$, i.e. $\delta[n-k]$ has the z-transform

$$\delta[n-k] \Leftrightarrow z^{-k},$$

so long as $k \geq 0$. Note that if $k < 0$, then the summation for the unilateral z-transform will never "see" the only non-zero term, and hence the z-transform will be zero for $\delta[n+k]$ for $k > 0$.

Another sequence for which we can apply knowledge of an existing transform is the unit step, $u[n]$. Note that for $n \geq 0$, the unit step is a complex exponential sequence of the form a^n for the specific case $a = 1$. As a result, we know that the z-transform for $u[n]$ is given by

$$u[n] \Leftrightarrow \frac{z}{z-1}, \quad |z| > 1.$$

5.3 Properties of the unilateral z-transform

We will discuss a few properties of the unilateral z-transform. To facilitate this discussion, we will use the following operator notation for the z-transform, $Z(y[n]) \triangleq Y(z)$. The first property has already been shown, and is that of linearity.

5.3.1 Linearity

The unilateral z-transform is a linear operation, i.e. it satisfies superposition. This has been shown previously, and we have that

$$Z(ay_1[n] + by_2[n]) = aY_1(z) + bY_2(z).$$

This is readily shown from the definition of the z-transform, i.e.

$$\begin{aligned} \sum_{n=0}^{\infty}(ay_1[n] + by_2[n])z^{-n} &= a\sum_{n=0}^{\infty} y_1[n] + b\sum_{n=0}^{\infty} y_2[n] \\ &= aY_1(z) + bY_2(z). \end{aligned}$$

5.3.2 Delay Property #1

The next property is the first delay property, that is, when a sequence is delayed by a positive amount. If a sequence is delayed by k samples, then we have that

$$Z(y[n-k]u[n-k]) = z^{-k}Y(z).$$

In words, this property states that truncating a sequence at the origin, and then shifting to the right by a positive integer k, is equivalent to multiplying the z-transform of the un-shifted sequence by z^{-k}. This can be proven from the defition of the z-transform. We have that

$$\begin{aligned} Z(y[n-k]u[n-k]) &= \sum_{n=0}^{\infty} y[n-k]u[n-k]z^{-n} \\ &= \sum_{n=k}^{\infty} y[n-k]z^{-n} \\ &= \sum_{m=0}^{\infty} y[m]z^{-(m+k)} \\ &= z^{-k}Y(z), \end{aligned}$$

where the second line follows from $u[n-k]$ being zero for $n < k$ and the third line follows from making the substitution $m = n - k$.

5.3.3 Delay Property #2

For cases where $y[-1], y[-2], \ldots, y[-k]$ are known or defined ($k > 0$), we have the following property. Here, the sequence $y[n]$ is not truncated at the origin, prior to shifting.

$$Z(y[n-k]) = z^{-k}\left[Y(z) + \sum_{m=1}^{k} y[-m]z^{m}\right].$$

This can be shown from linearity and delay property #1. Specifically, we note that for $n \geq 0$, we have that

$$y[n-k] = y[n-k]u[n-k] + \sum_{m=1}^{k} y[-m]\delta[n-k+m]$$

by simply adding back into the sequence the "new" values that shift into the region $n \geq 0$ from the left. We now can use linearity together with delay property #1 and the z-transform for a shifted discrete-time impulse to obtain

$$\begin{aligned} Z(y[n-k]) &= z^{-k}Y(z) + \sum_{m=1}^{k} y[-m]z^{-(k-m)} \\ &= z^{-k}\left[Y(z) + \sum_{m=1}^{k} y[-m]z^{m}\right]. \end{aligned}$$

5.3.4 Advance Property

The following advance property can also be used in the solution of difference equations with initial conditions. We have that

$$Z(y[n+k]u[n]) = z^{k}\left[Y(z) - \sum_{m=0}^{k-1} y[m]z^{-m}\right].$$

This property is also readily shown by noting that

$$\begin{aligned} Z(y[n+k]u[n]) &= \sum_{n=0}^{\infty} y[n+k]z^{-n} \\ &= z^{k}\sum_{n=0}^{\infty} y[n+k]z^{-(n+k)} \\ &= z^{k}\sum_{m=k}^{\infty} y[m]z^{-m} \\ &= z^{k}\left(\sum_{m=0}^{\infty} y[m]z^{-m} - \sum_{m=0}^{k-1} y[m]z^{-m}\right) \\ &= z^{k}\left(Y(z) - \sum_{m=0}^{k-1} y[m]z^{-m}\right). \end{aligned}$$

5.3.5 Convolution

One of the useful properties of the z-transform is that it maps convolution in the time domain into multiplication in the z-transform domain. We will show this here for the unilateral z-tranform and sequences that are only nonzero for $n \geq 0$ and revisit the more general case when we explore the two-sided z-transform. Spefically, we assume that $x[n] = h[n] = 0, n < 0$ and consider the convolution

$$y[n] = \sum_{m=-\infty}^{\infty} h[m]x[n-m].$$

Taking the z-transform of both sides, we have

$$\begin{aligned}
Y(z) &= \sum_{n=0}^{\infty} y[n]z^{-n} \\
&= \sum_{n=0}^{\infty} \sum_{m=-\infty}^{\infty} h[m]x[n-m]z^{-n} \\
&= \sum_{m=-\infty}^{\infty} h[m] \sum_{n=0}^{\infty} x[n-m]z^{-n} \\
&= \sum_{m=-\infty}^{\infty} h[m]X(z)z^{-m} \\
&= X(z) \sum_{m=-\infty}^{\infty} h[m]z^{-m} \\
&= X(z)H(z),
\end{aligned}$$

where in the third line we used the delay property and that both sequences were zero for $n < 0$. When the sequences $x[n]$ and $h[n]$ are not both zero for $n < 0$, then multiplication of one-sided z-transforms cam be shown to be equivalent to convolution of the sequences $x[n]u[n]$ and $h[n]u[n]$, i.e.

$$\sum_{k=-\infty}^{\infty} x[n-k]y[n-k]h[k]u[k] = \sum_{k=0}^{n} x[n-k]h[k] \longleftrightarrow X(z)H(z),$$

where $X(z)$ and $H(z)$ are the one-sided z-transforms of the sequences $x[n]$ and $h[n]$.

5.3.6 Inverse unilateral z-transform

One method that can be used to solve difference equations, is to take the z-transform of both sides of the difference equation, and solve the resulting algebraic equation for $Y(z)$, and then find the inverse transform to obtain $y[n]$. A formula for the inverse unilateral z-transform can be written

$$y[n] = \frac{1}{2\pi j} \oint Y(z)z^{n-1}dz$$

which is an integral taken over a closed contour in a counter clockwise direction in the region of converge of $Y(z)$, as shown in Figure . Other inversion methods exist if $Y(z)$ is a rational function (i.e., a ratio of polynomials), e.g.,

$$Y(z) = \frac{b_0 + b_1 z + \ldots + b_M z^M}{a_0 + a_1 z + \ldots + a_N z^N}.$$

Direct long division

A straightforward, but not entirely practical method, since it does not produce a closed-form expression for $y[n]$, is to employ long-division of the polynomials directly. This is a simple method for obtaining a power-series expansion for $Y(z)$ from the rational expression, and then from the definition of the z-transform, the terms of the sequence can be identified one at a time.

Example

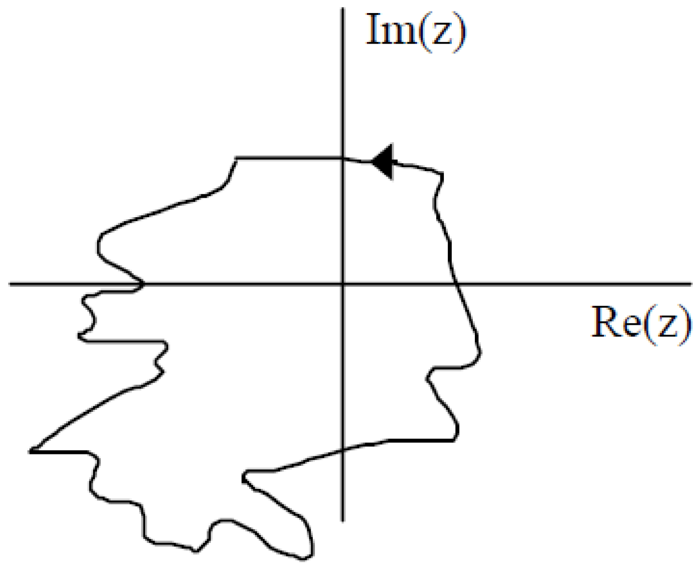

Figure 5.3: Contour integral for taking an inverse z-transform.

For the expression $Y(z) = z/(z-a)$, we have that

$$Y(z) = \frac{z}{z-a}$$

$$\frac{z}{z-a} = \begin{array}{r} 1+\frac{a}{z}+\frac{a^2}{z^2} \\ z-a\overline{)z \qquad\quad} \\ \underline{z-a} \\ 0+a+0 \\ \underline{0+a-\frac{a^2}{z}} \\ 0+0+\frac{a^2}{z} \\ 0+0+\frac{a^2}{z}-\frac{a^3}{z^2} \end{array}$$

Note that from the above series expansion, together with the definition of the unilateral z-transform, i.e. $Y(z) = y[0] + y[1]z^{-1} + y[2]z^{-2} + \cdots$, we can immediately identify all of the terms of the sequence $y[n]$. That is we have that

$$\begin{aligned} Y(z) &= 1 + az^{-1} + a^2z^{-2} + a^3z^{-3} + \cdots \\ &= y[0] + y[1]z^{-1} + y[2]z^{-2} + y[3]z^{-3} + \cdots, \end{aligned}$$

from which we may infer that $y[n] = a^n, n \geq 0$.

5.3.7 z-transform properties

A short table of z-transform properties is given in Table (5.1) . These can be proven either directly from the definition of the z-transform, or through application of other known properties.

5.3.8 Table of unilateral z-transform pairs

A short table of unilateral z-transforms is given in Table (5.2) below. These can also be derived directly from the definition of the unilateral z-transform, or through application of the theorems listed in Table (5.1).

Superposition	$ax_1[n] + bx_2[n]$	$\Leftrightarrow$	$aX_1(z) + bX_2(z)$
Advance	$x[n+1]u[n]$	$\Leftrightarrow$	$z(X(z) - x[0])$
Modulation	$a^n x[n]$	$\Leftrightarrow$	$X(a^{-1}z)$
Multiplication by n	$nx[n]$	$\Leftrightarrow$	$-z\left(\frac{dX(z)}{dz}\right)$
Convolution	$\sum_{k=0}^{n} x[k]y[n-k]$	$\Leftrightarrow$	$X(z)Y(z)$
Convolution when $x[n] = y[n] = 0, \ \ n < 0$	$\sum_{k=-\infty}^{\infty} x[k]y[n-k]$	$\Leftrightarrow$	$X(z)Y(z)$
Advance by k	$y[n+k]u[n]$	$\Leftrightarrow$	$z^k \left[Y(z) - \sum_{m=0}^{k-1} y[m]z^{-m}\right]$
Delay property #1	$y[n-k]u[k]$	$\Leftrightarrow$	$z^k Y(z)$
Delay property #2	$y[n-k]$	$\Leftrightarrow$	$z^{-k}\left[Y(z) + \sum_{m=1}^{k} y[-m]z^m\right]$

Table 5.1: Table of unilateral z-transform properties.

$x[n]$	$\Leftrightarrow$	$X(z) = \sum_{n=0}^{\infty} x[n]z^{-n}$	ROC_X
$\delta[n] = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases}$	$\Leftrightarrow$	1	all z
$\delta[n-k]$	$\Leftrightarrow$	$\begin{cases} z^{-k}, & k \geq 0 \\ 0, & k < 0 \end{cases}$	$z \neq 1$
a^n	$\Leftrightarrow$	$\frac{z}{z-a}$	$\|z\| > \|a\|$
na^n	$\Leftrightarrow$	$\frac{az}{(z-a)^2}$	$\|z\| > \|a\|$
$a^n \sin(\omega n)$	$\Leftrightarrow$	$\frac{az\sin(\omega)}{z^2 - 2az\cos(\omega) + a^2}$	$\|z\| > \|a\|$
$a^n \cos(\omega n)$	$\Leftrightarrow$	$\frac{1 - az\cos(\omega)}{z^2 - 2az\cos(\omega) + a^2}$	$\|z\| > \|a\|$
$u[n] = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases}$	$\Leftrightarrow$	$\frac{z}{z-1}$	$\|z\| > 1$
1	$\Leftrightarrow$	$\frac{z}{z-1}$	$\|z\| > 1$

Table 5.2: Table of unilateral z-transform pairs.

5.4 Inverse z-transform by partial fraction expansion

One method for finding the inverse of a unilateral (one-sided) z-transform is to recognize the transform of interest as the transform of a signal whose z-transform you already know, or have access to via a lookup table, such as that found at the end of the last chapter. For example, if you know that the transform of the unit-sample (or discrete-time impulse) is $X(z) = 1$, then given a transform of the form 1 $X_2(z) = 2 + z^{-1}$, you might use the linearity property of the z-transform together with the delay property to identify $x[n] = 2\delta[n] + \delta[n-1]$. This method is sometimes referred to as the "table lookup method". We can generalize this idea to find the inverse transform of more elaborate functions by learning how to decompose complex expressions into a linear combination of terms, each of which we might be able to identify their inverse by inspection. This simply amounts to using linearity to break a complex transform into a sum of simpler terms, and then using a lookup table to find the inverse of each of the terms independently. The overall inverse transform would then be the sum of the inverses of each of the simpler terms, exploiting the linearity of the transform.

The inverse transform method we will describe will work well in the case when $X(z)$ is a rational function, that is, when it can be expressed as a ratio of finite order polynomials in z . The method is based on the notion that every rational function can be expanded in terms of partial fractions. If the rational function $X(z)$ is proper, that is, the degree of the numerator polynomial is less than the degree of the denominator polynomial, and if the roots of the denominator polynomial are distinct, then we can factor $X(z)$ in the form

$$X(z) = \frac{b_0 + b_1 z + \cdots b_M z^M}{a_0 + a_1 z + \cdots + a_N z^N} = \frac{b_0 + b_1 z + \cdots b_M z^M}{(z - r_1)(z - r_2)\cdots(z - r_N)},$$

where, here, $X(z)$ is proper if $M < N$, and the roots of denominator polynomial are ${r_k}_{k=1}^N$. When the r_k are distinct (or "simple"), then, we can write

$$X(z) = \sum_{k=1}^{N} \frac{A_k}{z - z_k},$$

where the constants A_k are called the residues of $X(z)$. In this form, we can use a simple method to find the residues when all of the roots are distinct. We see that they can be obtained by the formula

$$A_k = (z - r_k)X(z)\,|_{z=r_k},$$

since the term $(z - r_k)$ makes each term in the sum become zero when evaluated at $z = r_k$, except for the one term in the sum that had$(z - r_k)$ in the denominator. This term is has A_k in the numerator, and hence yields the formula above.

Once we have expanded $X(z)$ in this form, we can then read off the inverse transform as

$$X(z) = \sum_{k=1}^{N} \frac{A_k}{z - z_k} = \sum_{k=1}^{N} z^{-1} \frac{A_k z}{z - r_k} \Leftrightarrow x[n] = \sum_{k=1}^{N} A_k r_k^{n-1} u[n-1],$$

once again using a combination of the linearity property of the unilateral z-transform and the delay property. We can see how this works in practice by looking at an example.

Example

We can use this approach to find the inverse transform for the following unilateral z-transform:

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

Now, we wish to find the sequence $y[n]$, for $n \geq 0$. We have that

$$Y(z) = \frac{z - 1}{(z - 2)(z - 3)} = \frac{A_1}{z - 2} + \frac{A_2}{z - 3},$$

so that when we multiply $Y(z)$ by $z - 2$ we obtain

$$(z - 2)Y(z) = A_1 + \frac{(z - 2)A_2}{z - 3}.$$

Now, setting $z = 2$,we have

$$z - 2Y(z)|_{z=2} = A_1,$$

since the second term on the right hand side becomes zero. We then find that

$$A_1 = \left.\frac{z-1}{z-3}\right|_{z=2} = \frac{1}{-1} = -1.$$

Similarly we find that

$$A_2 = \left.\frac{z-1}{z-2}\right|_{z=3} = \frac{2}{1} = 2.$$

Putting these together yeilds that

$$\begin{aligned} Y(z) &= \frac{-1}{z-2} + \frac{2}{z-3} \\ &= -z^{-1}\left[\frac{z}{z-2}\right] + 2z^{-1}\left[\frac{z}{z-3}\right]. \end{aligned}$$

From the table of unilateral z-transform pairs, we have that

$$a^n \Leftrightarrow \frac{z}{z-a},$$

and applying Delay Property #1, we ahve that

$$a^{n-1}u[n-1] \Leftrightarrow z^{-1}\frac{z}{z-a}.$$

From the linearity of the z-transform, we can now invert each of the terms individually, and then put them together to obtain

$$y[n] = -(2)^{n-1}u[n-1] + 2(3)^{n-1}u[n-1].$$

If we prefer, we can re-write this as

$$y[n] = \begin{cases} -\frac{1}{2}(2)^n + \frac{2}{3}3^n, & n \geq 1 \\ 0, & \text{n=0.} \end{cases}$$

We do not evaluate $y[n]$ for values of $n < 0$, since the unilateral z-transform does not tell us anything about this region. In this example, we needed to apply both linearity and Delay Property #1. We can avoid the need to apply the delay property to each term, by expanding $z^{-1}Y(z)$ in a PFE as

$$\frac{Y(z)}{z} = \frac{A_1}{z} + \frac{A_2}{z-2} + \frac{A_3}{z-3}.$$

Then we can obtain

$$Y(z) = A_1 + \frac{A_2 z}{z-2} + \frac{A_3 z}{z-3},$$

and each of the terms in this expansion can be inverted directly, without the need for the delay property. Working out the details for this example, we have

$$\begin{aligned} \frac{Y(z)}{z} &= \frac{z-1}{z(z-2)(z-3)} \\ &= \frac{A_1}{z} + \frac{A_2}{z-2} + \frac{A_3}{z-3}, \end{aligned}$$

and that

$$\begin{aligned} A_1 &= \left.\frac{z-1}{(z-2)(z-3)}\right|_{z=0} = \frac{-1}{(-2)(-3)} = -\frac{1}{6}, \\ A_1 &= \left.\frac{z-1}{z(z-3)}\right|_{z=2} = \frac{1}{(2)(-1)} = -\frac{1}{2}, \\ A_3 &= \left.\frac{z-1}{z(z-2)}\right|_{z=3} = \frac{2}{(3)(1)} = \frac{2}{3}. \end{aligned}$$

Putting these together, yields

$$\begin{aligned}\frac{Y(z)}{z} &= \frac{-\frac{1}{6}}{z} + \frac{-\frac{1}{2}}{z-2} + \frac{\frac{2}{3}}{z-3} \\ Y(z) &= \frac{-\frac{1}{6}z}{z} + \frac{-\frac{1}{2}z}{z-2} + \frac{\frac{2}{3}z}{z-3}.\end{aligned}$$

We can again invert each term, term by term, to obtain

$$y[n] = -\frac{1}{6}\delta[n] - \frac{1}{2}(2)^n u[n] + \frac{2}{3}(3)^n u[n].$$

Here we have identified that the inverse transform of a constant is a discrete-time impulse. This can be obtained either from the table of transforms, or by noting that if a z-transform is constant, say $X(z) = C$, then we have that

$$X(z) = x[0] + x[1]z^{-1} + x[2]z^{-2} + x[3]z^{-2} + \cdots$$

and we see that the only way that $X(z)$ can be a constant (i.e. the only power of z in the expression is z^0) is for $x[0] = C$, i.e. we have that

$$X(z) = -\frac{1}{6} \Leftrightarrow x[n] = -\frac{1}{6}\delta[n].$$

Putting all of the terms together yields,

$$\begin{aligned} y[n] &= \begin{cases} -\frac{1}{2}(2)^n + \frac{2}{3}(3)^n, & n \geq 1 \\ -\frac{1}{6} - \frac{1}{2} + \frac{2}{3}, & n = 0 \end{cases} \\ &= \begin{cases} -\frac{1}{2}(2)^n + \frac{2}{3}(3)^n, & n \geq 1 \\ 0, & n = 0, \end{cases}\end{aligned}$$

as we had before. In this example, the PFE for $z^{-1}Y(z)$ was more complicated (involved one more term) than the PFE for $Y(z)$. In many cases this extra complication does not arise. If the numerator of $Y(z)$ contains a power of z (say z or z^2), then the z in the denominator of $z^{-1}Y(z)$ is cancelled, in which case the PFE for $z^{-1}Y(z)$ has exactly the same form as the PFE of $Y(z)$.

If the r_i are not distinct, we will need to modify the partial fraction expansion slightly. Suppose r_j is a root that is repeated q times. We then must replace the single term corresponding to $r_j$with a set of q terms, one for each occurrence of the root, where the denominator is raised to each power, starting from the first power up to the qth power, i.e. we replace

$$\frac{A_k}{(z - r_k)} \Rightarrow \sum_{\ell=1}^{q} \frac{B_\ell}{(z - r_k)^\ell}$$

in the partial fraction expansion, where the new constants satisfy

$$B_\ell = \frac{1}{(q-\ell)!}\left[\frac{d^{q-\ell}}{dz^{q-\ell}}(z - r_k)^q Y(z)\right]\Bigg|_{z=r_k}.$$

While it is important to know that this formula exists, in practice, the form of the expansion is more important than the explicit formula for determination of the constants. For example, you can determine the constants by simply matching terms in the expansion as shown in the next example.

Example

Determine the partial fraction expansion of the z-transform

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

To accomplish this, we need only know the form of the expansion, and not dwell on the formula for the constants of the repeated roots. First, we obtain

$$\frac{Y(z)}{z} = \frac{1}{(z-1)(z-3)^2} = \frac{A_1}{(z-1)} + \frac{A_2}{(z-3)} + \frac{A_3}{(z-3)^2}$$

as the form of the partial fraction expansion. We can now obtain the first term directly, using the non-repeated roots formula

$$A_1 = \left.\frac{1}{(z-3)^2}\right|_{z=1} = \frac{1}{4},$$

to get started. Now, we find A_3 before we find A_2. In general, if we find the coefficient over the highest power denominator first, the resulting algebra will be simplified. By multiplying both sides of the PFE by $(z–3)^2$ we obtain

$$\frac{(z-3)^2 Y(z)}{z} = \frac{1}{(z-1)} = \frac{A_1(z-3)^2}{(z-1)} + A_2(z-3) + A_3.$$

Setting $z = 3$, we have

$$A_3 = \left.\frac{1}{z-1}\right|_{z=3} = \frac{1}{2}.$$

There are a few ways to determine A_2. One is to first differentiate the expression $(z-3)^2Y(z)/z$ with respect to z, which yields

$$\frac{-1}{(z-1)^2} = \frac{2A_1(z-3)(z-1) - A_1(z-3)^2}{(z-1)^2} + A_2,$$

which upong setting $z = 3$, yields

$$-\frac{1}{4} = A_2.$$

Another way to find A_2would be to simply fill in the known constants, yielding

$$\begin{aligned}\frac{1}{(z-1)(z-3)^2} &= \frac{\frac{1}{4}}{(z-1)} + \frac{A_2}{(z-3)} + \frac{\frac{1}{2}}{(z-3)^2} \\ &= \frac{\frac{1}{4}(z-3)^2 + \frac{1}{2}(z-1) + A_2(z-1)(z-3)}{(z-1)(z-3)^2}.\end{aligned}$$

Now, the numerators must match, so we must have

$$1 = \frac{1}{4}(z-3)^2 + \frac{1}{2}(z-1) + A_2(z-1)(z-3),$$

which can be easily solved for A_2. For example, both sides must have the same coefficient to the term z^2,which, on the left hand side is zero, and on the right hand side is

$$0 = \frac{1}{4} + A_2,$$

which yields that

$$A_2 = -\frac{1}{4}$$

as before. Substituting these values into the original PFE yields

$$Y(z) = \frac{\frac{1}{4}z}{(z-1)} - \frac{\frac{1}{4}z}{(z-3)} + \frac{\frac{1}{2}z}{(z-3)^2}.$$

The first two terms are easy to invert from our table of known transforms. For the third term, we recall that

$$na^n \Leftrightarrow \frac{az}{(z-a)^2}.$$

Therefore we have that

$$\begin{aligned} y[n] &= \frac{1}{4}(1)^n - \frac{1}{4}(3)^n + \frac{1}{2}\left(\frac{1}{3}\right)n(3)^n, \quad n \geq 0 \\ &= \frac{1}{4} - \frac{1}{4}(3)^n + \frac{1}{6}n(3)^n, \quad n \geq 0.\end{aligned}$$

Let us consider another example.

Example

Given the unilateral z-transform of the seqeunce $y[n]$ is given by

$$Y(z) = \frac{2z^3 + z^2 - z + 4}{(z-2)^3},$$

find $y[n]$. Recall that for a "strictly proper" rational function, we require that the degree of the numerator polynomial be strictly less than the degree of the denominator polynomial. This condition is necessary for us to use the form of the partial fraction expansion we have considered thus far. We can use the PFE form if we choose to expand $Y(z)/z$ in PFE, since this will be a strictly proper rational function. We begin with

$$\begin{aligned}\frac{Y(z)}{z} &= \frac{2z^3 + z^2 - z + 4}{z(z-2)^3} \\ &= \frac{A_1}{z} + \frac{A_2}{(z-2)} + \frac{A_3}{(z-2)^2} + \frac{A_4}{(z-2)^3}\end{aligned}$$

and immediately note that

$$A_1 = \left.\frac{2z^3 + z^2 - z + 4}{(z-2)^3}\right|_{z=0} = \frac{4}{-8} = -\frac{1}{2}.$$

Now, we again find the coefficient of repeated-root term with highest power denominator first. Mutliplying $Y(z)/z$ by $(z–2)^3$, we obtain

$$\frac{2z^3 + z^2 - z + 4}{z} = -\frac{(z-2)^3}{2z} + A_2(z-2)^2 + A_3(z-2) + A_4,$$

which when evaluated for $z = 2$, yields

$$\begin{aligned}\frac{16 + 4 - 2 + 4}{2} &= A_4 \\ 11 &= A_4.\end{aligned}$$

Now putting the PFE into a common demonimator and setting the numerators equal yields,

$$2z^3 + z^2 - z + 4 = -\frac{1}{2}(z-2)^3 + A_2 z(z-2)^2 + A_3 z(z-2) + 11z.$$

We can now match terms with corresponding powers of z to obtain

$$\begin{aligned}2z^3 &= -\frac{1}{2}z^3 + A_2 \\ \frac{5}{2} &= A_2,\end{aligned}$$

and

$$\begin{aligned}z^2 &= \frac{1}{2}6z^2 + \frac{5}{2}(-4)z^2 + A_3 z^2 \\ 8 &= A_3.\end{aligned}$$

Putting all of the terms together, yields,

$$Y(z) = -\frac{1}{2} + \frac{\frac{5}{2}z}{z-2} + \frac{8z}{(z-2)^2} + \frac{11z}{(z-2)^3}.$$

Now we can invert each of the terms, one at a time, to yield,

$$y[n] = -\frac{1}{2}\delta[n] + \frac{5}{2}(2)^n + 4n(2)^n + \frac{11}{2}(n-1)n(2)^{n-2}, \quad n \geq 0,$$

where for the last term, we used the transform pair

$$\frac{1}{2}n(n-1)a^n \Leftrightarrow \frac{a^2 z}{(z-a)^3}.$$

We could combine all of the results to obtain

$$y[n] = \begin{cases} 2, & n = 0 \\ \frac{1}{8}\left(11n^2 + 21n + 20\right)(2)^n, & n \geq 1. \end{cases}$$

5.5 Difference equations and the z-transform

Just as the Laplace transform was used to aid in the solution of linear differential equations, the z-transform can be used to aid in the solution of linear difference equations. Recall that linear, constant coefficient differential equations could be converted into algebraic equations by transforming the signals in the equation using the Laplace transform. Derivatives could be mapped into functions of the Laplace transform variable s, through the derivative law for Laplace transforms. Similarly, delayed versions of a sequence can be mapped into algebraic functions of z, using one of the delay rules for z-transforms.

In the case of continuous-time linear systems described by differential equations, in order to find the response of such a linear system to an particular input, the differential equations needed to be solved, using either time-domain or Laplace transform methods. For an Nth-order differential equation, in general N conditions on the output were needed in order to specify the output in response to a given input. Similarly, for linear difference equations of Nth-order, N pieces of information are needed to find the output for a given input. Unlike the continuous-time case, difference equations can often be simply iterated forward in time if these N conditions are consecutive. That is, given $y[-N+1], ...y[-1]$, then re-writing

$$\sum_{k=0}^{N} a_k y[n-k] = \sum_{k=0}^{M} b_k x[n-k]$$

in the form

$$a_0 y[0] = -\sum_{k=1}^{N} a_k y[n-k] = \sum_{k=0}^{M} b_k x[n-k],$$

from which $y[0]$ could be found. Iterating this process forward could find each value of the output without ever explicitly obtaining a general expression for $y[n]$.

In this chapter we will explore the z-transform for the explicit solution of linear constant coefficient difference equations. The properties of the z-transform that we have developed can be used to map the difference equations describing the relationship between the input and the output, into a simple set of linear algebraic equations involving the z-transforms of the input and output sequences. By solving the resulting algebraic equations for the z-transform of the output, we can then use the methods we've developed for inverting the transform to obtain an explicit expression for the output. We begin with an example.

Example

We revisit this simple linear, homogeneous difference equation, now using the unilateral ztransform. Again consider the difference equation

$$y[n] - 3y[n-1] = 0, \quad n \geq 0, \quad y[-1] = 2$$

Taking unilateral z-transform of both sides, and using the delay property, we obtain

$$\begin{aligned} Y(z) - 3z^{-1}[Y(z) + zy[-1]] &= 0 \\ Y(z)[1 - 3z^{-1}] &= 6, \end{aligned}$$

which can be solved for $Y(z)$ directly, yielding

$$\begin{aligned} Y(z) &= \frac{6z}{z-3}, \\ y[n] &= 6(3)^n u[n]. \end{aligned}$$

Another, slightly more involved example, repeats another example as well.

Example

Consider the following homogenous, linear constant coefficient difference equation, defined for nonnegative n and with initial conditions shown

$$y[n] + 4y[n\text{-}1] + 4y[n-2] = 0, \quad n{\ge}0, \quad y[\text{-}1] = y[-2] = 1.$$

Taking the z-transform of both sides, again using the delay property and including the initial conditions, we obtain

$$\begin{aligned}
Y(z) + 4z^{-1}\left[Y(z) + zy[-1]\right] + 4z^{-2}\left[Y(z) + zy[-1] + z^2 y[-2]\right] &= 0 \\
Y(z)\left[1 + 4z^{-1} + 4z^{-2}\right] &= -4y[-1] - 4z^{-1}y[-1] - 4y[-2] \\
Y(z) &= \frac{-8 - 4z^{-1}}{1 + 4z^{-1} + 4z^{-2}} \\
&= \frac{-8z^2 - 4z}{z^2 + 4z + 4}.
\end{aligned}$$

This is not in strictly proper rational form, i.e. the degree of the numerator is not strictly less than that of the denominator, however when we expand $z^{-1}Y(z)$, we have

$$\begin{aligned}
\frac{Y(z)}{z} &= \frac{-8z - 4}{z^2 + 4z + 4} = \frac{-8z - 4}{(z+2)^2} \\
&= \frac{A_1}{(z+2)} + \frac{A_2}{(z+2)^2}.
\end{aligned}$$

Since we have repeated roots, we first seek the coefficient of the highest order root, A_2. By cross multiplying, we obtain

$$-8z - 4 = A_1(z+2) + A_2.$$

Setting $z = -2$ on both sides, we have that

$$16 - 4 = 12 = A_2.$$

We can also immediately see from the cross multiplication that

$$A_1 = -8,$$

by matching the terms on both sides that each multiply z. Putting these terms together, we have the full partial fraction expansion for $Y(z)$,

$$Y(z) = \frac{-8z}{(z+2)} + \frac{12z}{(z+2)^2}.$$

Using linearity to invert each term of the z-transform independently, we obtain

$$y[n] = -8(-2)^n u[n] - 6n(-2)^n u[n].$$

We now consider a case where the difference equation contains an input, or drive term, such that we no longer have a homogenous difference equation.

Example

Consider the following linear constant coefficient difference equation.

$$y[n+2] - \frac{3}{2}y[n+1] + \frac{1}{2}y[n] = \left(\frac{1}{3}\right)^n u[n], \;\; y[0] = 4, \; y[1] = 0.$$

Taking the unilateral z-transform of both sides and using the advance property, we obtain

$$\begin{aligned}
z^2\left[Y(z) - y[0] - z^{-1}y[1]\right] - \frac{3}{2}z\left[Y(z) - y[0]\right] + \frac{1}{2}Y(z) &= \frac{z}{z - \frac{1}{3}} \\
z^2\left[Y(z) - 4\right] - \frac{3}{2}z\left[Y(z) - 4\right] + \frac{1}{2}Y(z) &= \frac{z}{z - \frac{1}{3}} \\
Y(z)\left[z^2 - \frac{3}{2}z + \frac{1}{2}\right] &= \frac{z}{z - \frac{1}{3}} + 4z^2 - 6z.
\end{aligned}$$

We can now solve for $Y(z)$ and keep the terms on the right hand side separated into two distinct groups, namely,

$$Y(z) = \frac{1}{\left(z^2 - \frac{3}{2}z + \frac{1}{2}\right)} \left(\underbrace{\frac{z}{z - \frac{1}{3}}}_{\text{term due to input}} + \underbrace{4z^2 - 6z}_{\text{term due to initial conditions}} \right).$$

We can now write the z-transform as a sum of two terms, one due to the input, and one due to the initial conditions. Recall from our analysis of linear constant coefficient difference equations that these correspond to the zero-state response and the zero-input response of the system. Taking these two terms separately, again through linearity of the transform, we have that

$$Y(z) = T_z(z) + T_2(z)$$

where

$$\begin{aligned} T_1(z) &= \frac{z}{\left(z^2 - \frac{3}{2}z + \frac{1}{2}\right)\left(z - \frac{1}{3}\right)}, \\ T_2(z) &= \frac{4z^2 - 6z}{\left(z^2 - \frac{3}{2}z + \frac{1}{2}\right)}. \end{aligned}$$

Here, $T_1(z)$ is the z-transform of the zero-state repsonse, and $T_2(z)$ is the z-transform of the zero-input response. We can then take a pratial fraction expanson of each of the terms independently. For the first term, we find it convenient to express the partial fraction expansion as

$$\frac{T_1(z)}{z} = \frac{1}{(z - \frac{1}{2})(z - 1)(z - \frac{1}{3})} = \frac{A_1}{(z - \frac{1}{2})} + \frac{A_2}{(z - 1)} + \frac{A_3}{(z - \frac{1}{3})}.$$

This leads to

$$\begin{aligned} A_1 &= -12,\ A_2 = 3,\ A_3 = 9, \\ T_1(z) &= \frac{-12z}{(z - \frac{1}{2})} + \frac{3z}{(z - 1)} + \frac{9z}{(z - \frac{1}{3})}, \end{aligned}$$

and the resulting zero state response is given by

$$y_x[n] = -12\left(\frac{1}{2}\right)^n + 3 + 9\left(\frac{1}{3}\right)^n, n \geq 0.$$

For the zero-input response term, we have that

$$\frac{T_2(z)}{z} = \frac{4z - 6}{(z - \frac{1}{2})(z - 1)} = \frac{B_1}{(z - \frac{1}{2})} + \frac{B_2}{(z - 1)},$$

from which we can quickly solve for the constants, yielding

$$B_1 = 8,\ \ B_2 = -4,$$

which gives the partial fraction expansion for the zero-input response as

$$T_2 = \frac{8z}{(z - \frac{1}{2})} - \frac{4z}{(z - 1)}.$$

The resulting zero-input response is then given by

$$y_s[n] = 8\left(\frac{1}{2}\right)^n - 4,\ n \geq 0.$$

Putting the zero-state response and the zero-input response together, we obtain the total response

$$y[n] = y_x[n] + y_s[n] = -4\left(\frac{1}{2}\right)^n - 1 + 9\left(\frac{1}{3}\right)^n,\ n \geq 0.$$

In general, this method of solution can be applied to linear constant coefficient difference equations of arbitrary order. Note that while in this particular case, we applied the time advance property of the unilateral z-transform, when solving difference equations of the form

$$y[n] + a_1 y[n-1] + \cdots + a_N y[n-N] = x[n],\ n \geq 0,$$

which initial conditions $y[-k]$,$k = 1, \ldots, N$, we can use the Delay Property #2.

5.5.1 General form of solution of linear constant coefficient difference equations (LCCDE)s

In this section, we will derive the general form of a solution to a linear constant coefficient difference equation. We will prove that the zero-state response (response to the input, when state is initially zero) is given by a convolution. Consider the following difference equation

$$y[n+K] + a_1 y[n+K-1] + \cdots + a_K y[n] = x[n],\ n \geq 0$$

together with initial conditions $y[k],\ k = 0, 1, \ldots K-1$. Taking the one-sided z-transform of both sides, and using the Advance Property, we obtain

$$z^K \left[Y(z) - \sum_{m=0}^{K-1} y[m] z^{-m} \right] + a_1 z^{K-1} \left[Y(z) - \sum_{m=0}^{K-2} y[m] z^{-m} \right] + \cdots + a_{K-1} z \left[Y(z) - y[0] \right] + a_K Y(z) = X(z).$$

By defining

$$S(z) = z^K \left[\sum_{m=0}^{K-1} y[m] z^{-m} \right] + a_1 z^{K-1} \left[\sum_{m=0}^{K-2} y[m] z^{-m} \right] + \cdots + a_{K-1} z y[0],$$

we have that

$$Y(z)[z^K + a_1 z^{K-1} + \cdots + a_K] = X(z) + S(z),$$

where the *characteristic polynomial* is given by

$$z^K + a_1 z^{K-1} + \cdots + a_K.$$

We now define the *transfer function* $H(z)$ of the system described by the LCCDE as

$$H(z) = \frac{1}{z^K + a_1 z^{K-1} + \cdots + a_K}.$$

We then obtain that

$$Y(z) = H(z) \left[\underbrace{X(z)}_{\text{term do to the input}} + \underbrace{S(z)}_{\text{term due to initial conditions}} \right].$$

Notice that the decomposition property holds with

$$\begin{aligned} y_s[n] &= Z^{-1}\{H(z)S(z)\} \\ y_x[n] &= Z^{-1}\{H(z)X(z)\}. \end{aligned}$$

Both homogeneity and superposition hold with respect to $y_s[n]$ and $y_x[n]$ because the z-tranform is linear. Linear constant coefficient difference equations (LCCDE)s describe linear systems, which have already explored the time-domain (sequence-domain). It is worthwhile to consider the form of the solution that $y_s[n]$ will take.

Consider first the case when the roots of the characteristic polynomial are distinct. In this case, we have

$$\frac{S(z)H(z)}{z}=\frac{B_1}{(z-r_1)}+\frac{B_2}{(z-r_2)}+\cdots+\frac{B_K}{(z-r_K)}.$$

From the definition of $S(z)$, z is a factor in $S(z)$, so there is no need for a $z^{-1}B_0$term in the partial fraction expansion. Multiplying by z, we have

$$S(z)H(z)=\frac{B_1z}{(z-r_1)}+\frac{B_2z}{(z-r_2)}+\cdots+\frac{B_Kz}{(z-r_K)},$$

from which we can easily recover the sequence

$$y_s[n]=\sum_{i=1}^{K}B_i(r_i)^n,\ n\geq 0,$$

which is in the same form as the homogeneous solution that would be obtained from a time-domain solution of the LCCDE.

We can now observe the form of $y_x[n]$. Since we have that

$$y_x[n]=Z^{-1}\left\{H(z)X(z)\right\},$$

the partial fraction expansion shows that $y_x[n]$ will involve terms in both $y[n]$ and $x[n]$. We can also rewrite $y_x[n]$ using the convolution property:

$$y_x[n]=\sum_{m=0}^{n}h[m]x[n-m],$$

where

$$\begin{aligned} h[m] &= Z^{-1}\left\{H(z)\right\}=Z^{-1}\left\{z\frac{H(z)}{z}\right\} \\ &= Z^{-1}\left\{z\left(\frac{D_0}{z}+\frac{D_1}{z-r_1}++\frac{D_K}{z-r_K}\right)\right\} \\ &= \begin{cases} D_0+\sum_{i=1}^{K}D_i(r_i)^n, & n=0 \\ \sum_{i=1}^{K}D_i(r_i)^n, & n\geq 1. \end{cases} \end{aligned}$$

So, we see that $y_x[n]$ is given by a convolution of the input with $h[n]=Z^{-1}\{H(z)\}$. Note that the sequence $h[n], n\geq 0$, can be interpreted as the system unit pulse response (u.p.r), or impulse response, assuming zero initial conditions.

Definition

The unit-pulse seqeunce, or the discrete-time impulse, is given by

$$\delta[n]=\begin{cases}1, & n=0\\ 0, & n\neq 0.\end{cases}$$

The system response to a unit pulse, or discrete-time impulse, is given by

$$y[n]=y_x[n]|_{\text{assuming zero initial conditions}}=\sum_{m=0}^{n}h[m]\delta[n-m]=h[n].$$

We can explore the use of the impulse response to derive the response to more general signals through another example.

Example

Consider the following linear system with input $x[n]$ and output $y[n]$ as shown in Figure 5.4 .

$$x[n] \longrightarrow \boxed{\text{LSI System}} \longrightarrow y[n]$$

Figure 5.4: A linear shift-invariant system.

Suppose that when the input $x[n] = \delta[n]$ with zero initial conditions, then the output satisfies $y[n] = a^n$ for $n \geq 0$. Again, assuming zero initial conditions (i.e. that the system is *initially at rest*), determine $y[n]$ due to the input $x[n] = b^n, n \geq 0$.

Solution

Given $h[n] = a^n, n \geq 0$, we know that the output satisfies $y[n] = y_x[n]$, since the initial conditions are all zero, i.e. the system is initially at rest. We know from the convolution property that

$$\begin{aligned} y[n] &= \sum_{m=0}^{n} a^m b^{n-m} = b^n \sum_{m=0}^{n} \left(\frac{a}{b}\right)^m \\ &= \frac{b^{n+1}}{b} \frac{1-\left(\frac{a}{b}\right)^{n+1}}{1-\frac{a}{b}}, \quad a \neq b \\ &= \frac{b^{n+1}-a^{n+1}}{b-a}, \quad a \neq b. \end{aligned}$$

Comments

This discussion and these examples lead us to a number of conclusions about the solutions to linear constant coefficient difference equations. First, we can show (and we will see in the next sections) that the solution to a linear constant coefficient difference equation will have a essentially the same form when the input is merely shifted in time. Also, we will see that a similar form is maintained for inputs that are linear combinations of shifted versions of the input. For example, the response to an input of the form $x[n]$ will be similar in form to the response to the input $x[n] - 2x[n-1]$. We will also see that the solution methods developed here, as well as the unilateral z-transform, can be modified to accommodate situations when the input is applied earlier or later than for $n = 0$. While we discussed situations here that included both the zero-input response and the zero-state response, in practice we are generally interested in the zero-state response, or equivalently, we are interested in the response to an input when the system of interest is initially at rest. The reason for this is that we either have a system where the initial conditions are all zero, or for a stable system, such that the roots of the characteristic polynomial are all of modulus less than unity, $|r_i| < 1$, and that after some time, $y_s[n]$ has sufficiently decayed, such that for time scales of interest for a given application, $y[n] \approx y_x[n]$. As a result, from this point forward, we will assume that systems under discussion are initially at rest, and that all initial conditions are set to zero. As a result, the output of a linear system will be taken as the zero-state response, and we will be interested in the convolution relationship between the input and the output.

5.6 Two-sided z-transform

When the input to a discrete-time LSI system is of the form z^n for all n, i.e. the two-sided sequence that has non-zero terms for arbitrarily large positive and negative n, the output of the system is simply a scaled version of the input. This is the *eigenfunction* property of LSI systems in discrete-time. The eigenfunction property of continuous-time systems tells us that when the input to a continuous-time LTI system is of the form e^{st} for all t, then the output will be a scaled version of the input. This is easily shown as a consequence of the convolution integral for LTI systems

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau,$$

where $h(\tau)$ is the impulse repsonse of the continuous-time LTI system. Letting the input take the form of a complex exponential, we have

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(\tau)e^{s(t-\tau)}d\tau \\ &= e^{st}\int_{-\infty}^{\infty} h(\tau)e^{-s\tau}d\tau \\ &= e^{st}H(s), \end{aligned}$$

where $H(s)$ is the Laplace transform of the impulse response, when the integral exists. We call the signals of the form e^{st} eigenfunctions of continuous-time LTI systems, since they satisfy the property that, when taken as input to an LTI system, they produce an output that is identical except for a (possibly complex) scale factor. The scale factor $H(s)$ is called the eigenvalue associated with the eigenfunction. Note that eigenvalue for a given s is the same as the Laplace transform of the impulse response, evaluated at that value of s. The only signals that have this property, i.e. the only eigenfunctions for LTI systems, are signals of the form e^{st}, for different possible values of the complex parameter s. Note that sinusoids are not eigenfunctions for LTI systems! That means that if a sinusoid is input to an LTI system, the output will not be a simple scaled version of the input. However, since a sinusoid can be simply constructed as a sum of two such eigenfunctions, we can easily see what the output will be:

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(\tau)\cos(\omega(t-\tau))d\tau \\ &= \int_{-\infty}^{\infty} h(\tau)\frac{1}{2}\left(e^{j\omega(t-\tau)} + e^{-j\omega(t-\tau)}\right)d\tau \\ &= \int_{-\infty}^{\infty} h(\tau)\frac{1}{2}\left(e^{j\omega(t-\tau)}\right)d\tau + \int_{-\infty}^{\infty} h(\tau)\frac{1}{2}\left(e^{-j\omega(t-\tau)}\right)d\tau \\ &= \frac{1}{2}\left(e^{j\omega t}H(j\omega) + e^{-j\omega t}H(-j\omega)\right). \end{aligned}$$

Now, if the impulse response is a purely real-valued function, then its Fourier transform will have complex conjugate symmetry, such that

$$\begin{aligned} y(t) &= \frac{1}{2}\left(e^{j\omega t}H(j\omega) + e^{-j\omega t}H^*(j\omega)\right) \\ &= \frac{1}{2}\left(e^{j\omega t}|H(j\omega)|e^{j\angle H(j\omega)} + e^{-j\omega t}|H(j\omega)|e^{-j\angle H(j\omega)}\right) \\ &= \frac{1}{2}|H(j\omega)|\left(e^{j\omega t}e^{j\angle H(j\omega)} + e^{-j\omega t}e^{-j\angle H(j\omega)}\right) \\ &= |H(j\omega)|\cos\left(\omega t + \angle H(j\omega)\right). \end{aligned}$$

While the output is not simply a scaled version of the input, when we decompose the sinusoid into a sum of two eigenfunctions, we can use linearity of the LTI system to construct the output as a sum of the two eigenfunction outputs.

Returning to discrete-time LSI systems, when the input to an LSI system is of the form z^n for all n, the convolution sum yields that

$$\begin{aligned} y[n] &= \sum_{m=-\infty}^{\infty} h[m] z^{(n-m)} \\ &= z^n \sum_{m=-\infty}^{\infty} h[m] z^{-m} \\ &= z^n H(z), \end{aligned}$$

when the sum converges. Once again, we call signals of the form z^n eigenfunctions of discrete time LSI systems, and the associated eigenvalues, $H(z)$, correspond to the two-sided z-transform of the impulse response, evaluated at the particular value of z.

We define the two-sided z-transform of a sequence $y[n]$ as follows

$$Y(z) = \sum_{n=-\infty}^{\infty} y[n] z^{-n},$$

for values of z for which the sum converges. We call the values of z for which the sum converges the region of convergence of $Y(z)$, or simply the ROC_Y. Note that as with the unilateral z-transform, the two-sided (or bilateral) z-transform is again a complex function of a complex variable, meaning that it can take on complex values and that its argument is itself a complex variable.

For the two-sided transform, we can consider again a few example sequences for which the sequence values are non-zero for both positive and negative index values.

Example

Consider the following sequence,

$$y[n] = a^n u[n] + b^n u[-n-1] = \begin{cases} a^n, & n \geq 0 \\ b^n, & n < 0. \end{cases}$$

Now, using the definition of the z-transform, we have for this sequence,

$$\begin{aligned} Y(z) &= \sum_{n=-\infty}^{\infty} \left(a^n u[n] + b^n u[-n-1]\right) z^{-n} \\ &= \sum_{n=-\infty}^{\infty} \left(a^n u[n]\right) z^{-n} \sum_{n=-\infty}^{\infty} \left(b^n u[-n-1]\right) z^{-n} \\ &= \sum_{n=0}^{\infty} a^n z^{-n} \sum_{n=-\infty}^{-1} b^n z^{-n} \\ &= \sum_{n=0}^{\infty} \left(\frac{a}{z}\right)^n + \sum_{n=-\infty}^{-1} \left(\frac{b}{z}\right)^n \\ &= \frac{z}{z-a}, |z| > |a| + \sum_{n=-\infty}^{-1} \left(\frac{b}{z}\right)^n \\ &= \frac{z}{z-a}, |z| > |a| + \sum_{m=1}^{\infty} \left(\frac{z}{b}\right)^m \\ &= \frac{z}{z-a}, |z| > |a| + \frac{z}{b-z}, |z| < |b|, \end{aligned}$$

where we must combine the two conditions on $|z|$, to ensure convergence of both of the summations in the expression. Otherwise, one of the terms in the expression will be invalid, and the resulting algebraic expression will not be meaningful. Hence, we have

$$Y(z) = \frac{z}{z-a} + \frac{z}{b-z}, |a| < |z| < |b|.$$

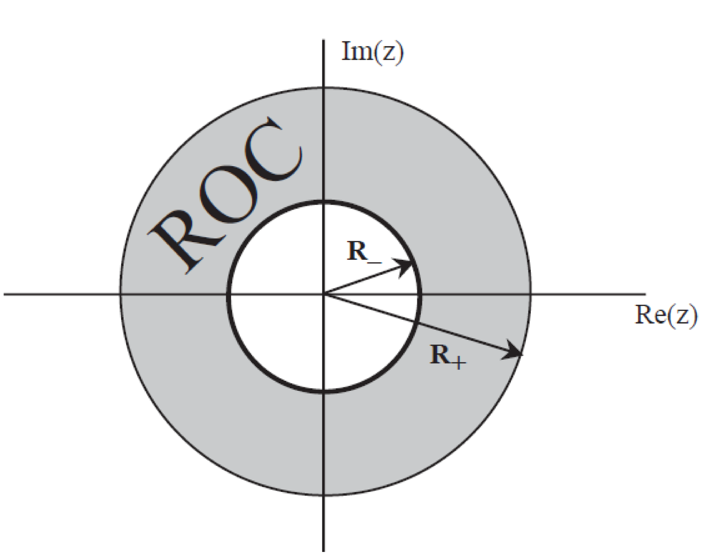

Figure 5.5: Region of convergence of the two-sided z-transform for a two-sided sequence.

Note that the region of convergence, ROC_Y, in this case is a ring, or annulus, in the complex plane as shown in Figure 5.5.

In this example,

$$R_- = |a|, \quad R_+ = |b|.$$

If $|a| \geq |b|$ then ROC_Y would be the empty set and z-transform would be undefined (i.e. is infinite) for all z. The reason that the region of convergence turns out to be a ring in the complex plane comes from properties of the summations that were assumed to converge in deriving the algebraic expression for the resulting z-transform. Specifically, looking at the definition of the z-transform, we obtain

$$Y(z) = \underbrace{\sum_{n=-\infty}^{-1} y[n]z^{-1}}_{\text{converges for } z\underline{small} \text{ enough, i.e. } |z| < R_+} + \underbrace{\sum_{n=0}^{\infty} y[n]z^{-1}}_{\text{converges for } z\underline{large} \text{ enough, i.e. } |z| > R_-}.$$

Note that $R_-$is determined by $y[n], n \geq 0$ and $R_+$is determined by $y[n], n < 0$. If $y[n] = 0$ for $n < 0$, then we have

$$Y(z) = \sum_{n=0}^{\infty} y[n]z^{-n}$$

and $R_+ = \infty$, which is essentially a one-sided (unilateral) z-transform. As a result, the region of convergence corresponds to $|z| > R_-$, as in Figure 5.6. If $y[n] = 0$ for $n > 0$, then we have that

$$Y(z) = \sum_{n=-\infty}^{0} y[n]z^{-n}$$

and $R_- = 0$, which implies that the region of convergence corresponds to a solid disk in the complex plain, i.e. we have $|z| < R_+$as in Figure 5.7. Note that in contrast to the one-sided z-transform, the two-sided z-transform can accommodate a wider range of signal behaviors, since they can be left-sided, right-sided, or two-sided and still have a bilateral z-transform. As such, we must state the ROC for $Y(z)$ to uniquely identify $y[n]$.

A *right sided sequence* is one that is zero for all n before some time index, i.e. $y[n] = 0, n < n_0$, for some n_0. A *left-sided sequence* is one that is zero for all n after some index, i.e. $y[n] = 0, n > n_0$, for some n_0, and a *two-sided sequence* is one that is neither left-sided nor right sided, i.e. it has non-zero terms for arbitrarily

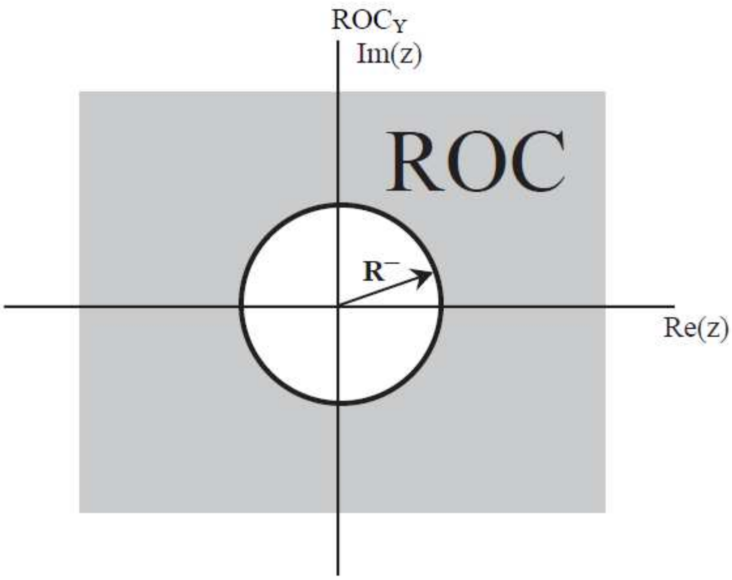

Figure 5.6: Region of convergence for a right sided sequence.

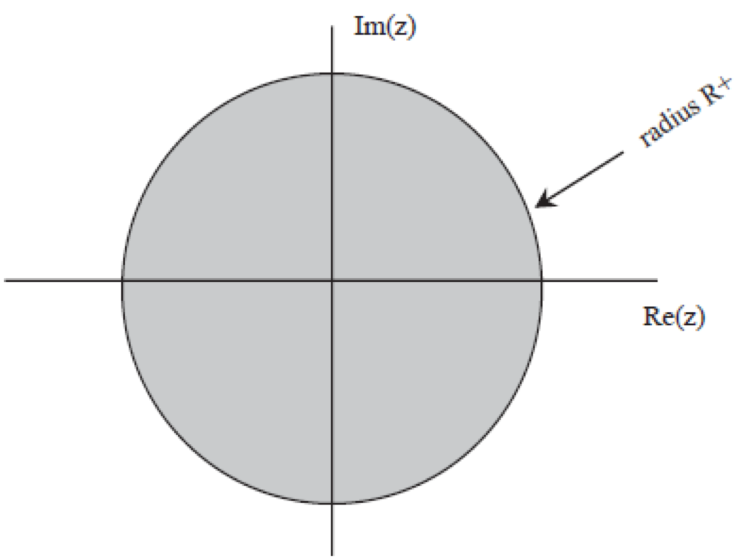

Figure 5.7: Region of convergence for a left-sided sequence.

large positive and negative indices. Examples of a right-sided sequence, include the unit step sequence, $u[n]$, and the complex exponential sequence $a^n u[n]$. An example left-sided sequence could be $u[-n]$ or $a^n u[-n-1]$. A two-sided sequence is one such as $a^{-|n|}$, where, for $|a| < 1$ is a decaying geometric seqeunce for positive and negative n. Since the two-sided z-transform multiplies the sequence $y[n]$ by z^{-n} and then sums the resulting modulated sequence for each value of z, in $Y(z)$, then whether a sequence is left-sided, rightsided or two-sided play an important role in the convergence (and the ROC) of the z-transform. Specifically, a right-sided sequence will have an infinite number of terms for large positive n, and, hence, the z-transform can converge when the magnitude of z is sufficiently large that z^{-n} dominates, making the sequence convergent. Therefore, right-sided sequences will have a ROC that is the entire z-plane outside of a circle of some radius (with the possible exception of infinity). Similarly, a left-sided sequence can converge when the magnitude of z is sufficiently small, such that z^{-n}, for large negative n decays sufficiently rapidly to dominate, making the series convergent. Therefore, a left-sided sequence will have a ROC for a disc-shaped region in the complex plane (with the possible exception of zero). A two-sided sequence, having both left-sided and right-sided elements must balance the effects such that the ROC will result in an annulus (ring) in the complex plane.

Example

Consider the following two sequences,

$$\begin{aligned} x[n] &= -(a^n)u[-n-1] = \begin{cases} -(a^n), & n < 0 \\ 0, & n \geq 0 \end{cases} \\ y[n] &= a^n u[n] = \begin{cases} a^n, & n \geq 0 \\ 0, & n < 0. \end{cases} \end{aligned}$$

For $x[n]$, we have

$$\begin{aligned} X(z) &= \sum_{n=-\infty}^{\infty} x[n] z^{-n} \\ &= -\sum_{n=-\infty}^{-1} a^n z^{-n} \\ &= -\sum_{n=-\infty}^{-1} \left(\frac{a}{z}\right)^n \\ &= -\sum_{k=1}^{\infty} \left(\frac{z}{a}\right)^k \\ &= \frac{\frac{z}{a}}{\frac{z}{a} - 1}, \left|\frac{z}{a}\right| < 1 \\ &= \frac{z}{z-a}, |z| < |a|. \end{aligned}$$

Similarly, we have already seen that

$$Y(z) = \frac{z}{z-a}, |z| > |a|.$$

So, we see that the *algebraic form* of $X(z)$ and $Y(z)$ are identical, but they are not the same functions, since they are defined on completely different regions of the complex plane. The z-transform of a sequence is not simply defined by the algebraic expression alone, but rather, the combination of the algebraic expression together with the region of convergence. In order to uniquely specify a sequence from its z-transform, we must include both the algebraic form as well as the region of the complex plane over which the form is valid. This leads to the following set of relations.

$$\begin{aligned} \text{uniquely defined sequence} &\iff \text{z-transform and region of converge} \\ a^n u[n] &\iff \frac{z}{z-a}, |z| > |a| \\ -(a^n)u[-n-1] &\iff \frac{z}{z-a}, |z| < |a| \end{aligned}$$

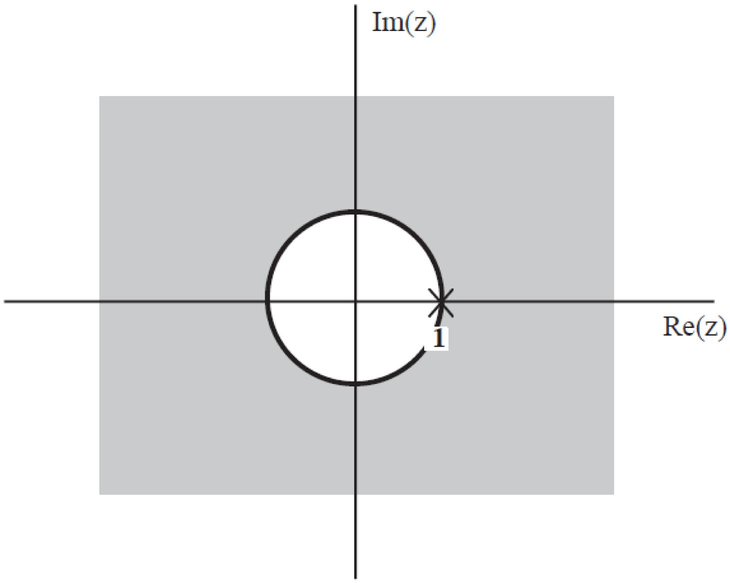

Figure 5.8: Region of convergence of the two-sided z-transform of $u[n]$.

Poles and Zeros

When sequences correspond to z-transforms that are rational functions (ratios of finite-order polynomials in z), we can explore some of the properties of the sequences and their z-transforms by examining the roots of the numerator and denominator polynomials. These are referred to as the zeros and the poles, respectively, of a rational z-transform. Specifically, for a z-transform given by

$$X(z) = \frac{B(z)}{A(z)}, z \in ROC_X,$$

we refer to the values of z such that $B(z) = 0$, as the zeros of $X(z)$, and the values of z for which $A(z) = 0$, as the poles of $X(z)$. That is,

$$\begin{aligned} \text{zeros:} &= \{z : B(z) = 0\} \\ \text{poles:} &= \{z : A(z) = 0\}, \end{aligned}$$

for rational $X(z)$. Rational z-transforms always have ROCs that are bounded by poles. This means that the ROC is either a disc, an annulus, or the entire plane minus a disc, with the possible exclusion of zero and infinity.

Example

Consider the rational transform

$$Y(z) = \frac{z}{z-1}, |z| > 1,$$

which has a pole at $z = 1$. This corresponds to the seqeunce $x[n] = u[n]$. The region of convergence for the z-transform is given by $|z| > 1$ as shown in Figure 5.8.

Example

Consider the sequence with rational transform

$$Y(z) = \frac{z}{z-2}, \quad |z| < 2,$$

which has a pole at $z = 2$. This corresponds to the seqeunce $y[n] = -(2^n)u[-n-1]$. The region of convergence is now the disk shown in Figure 5.9 .

Example

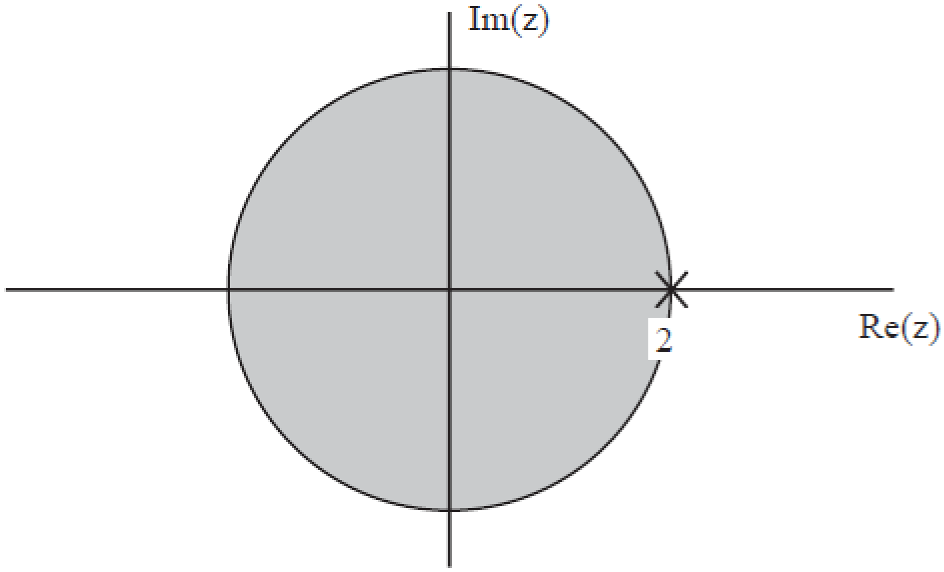

Figure 5.9: Region of convergence of the sequence $-(2^n)u[-n-1]$.

Now consider the sequence with rational transform

$$Y(z) = \frac{z}{(z-2)^2}, \quad |z| < 2,$$

which has a second-order pole at $z = 2$. For muiltiple poles and a left-sided seqeunce, we use the same methods we did for the right-sided case. We can easily show that

$$na^n u[-n-1] \Longleftrightarrow \frac{-az}{(z-a)^2}, \quad |z| < |a|.$$

Thus, we have that

$$y[n] = -\frac{1}{2}n(2^n)u[-n-1].$$

Example

Now consider the sequence with rational transform given by

$$Y(z) = \frac{z}{z-1} + \frac{z}{z-2}, \quad 1 < |z|| < 2,$$

which has poles at $z = 1$ and $z = 2$. The region of convergence is therefore an annulus in the complex plane, and the sequence will turn out to be two sided,

$$\begin{aligned} y[n] &= \begin{cases} 1, & n \geq 0 \\ -(2^n), & n < 0 \end{cases} \\ &= u[n] - (2^n)u[-n-1]. \end{aligned}$$

The region of convergence is depicted in Figure 5.10.

Example

Consider the seqeunce given by

$$x[n] = \left(\frac{1}{3}\right)^n, \quad -\infty < n < \infty.$$

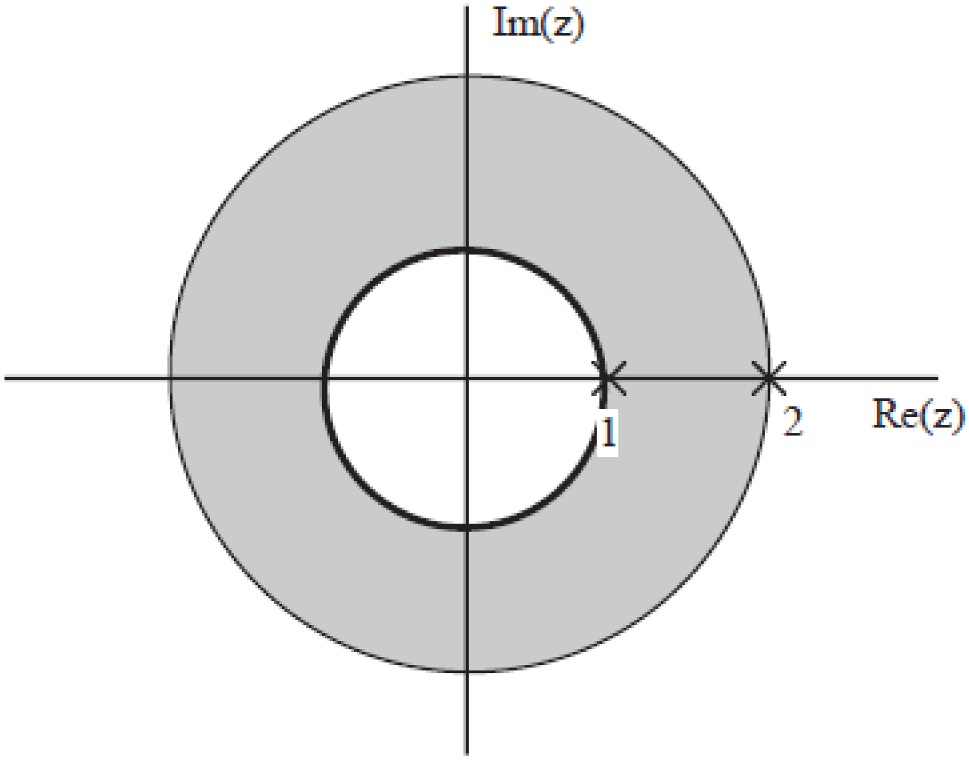

Figure 5.10: Region of convergence for the sequence $y[n] = u[n] - (2^n)u[-n-1]$.

For such a two-sided sequence, does the two-sided z-transform, $X(z)$ exist? Let us examine the z-transform of the seqeunce from the definition, from which we have

$$\begin{aligned} X(z) &= \sum_{n=-\infty}^{\infty} \left(\frac{1}{3}\right)^n z^{-n} \\ &= \sum_{n=-\infty}^{-1} \left(\frac{1}{3}\right)^n z^{-n} + \sum_{n=0}^{\infty} \left(\frac{1}{3}\right)^n z^{-n}. \end{aligned}$$

From here, we can see that the first sum will converge for $|z| < \frac{1}{3}$, but the second sum will only converge for $|z| > \frac{1}{3}$. As such, there is no value of z for which both sums will converge. Thus, $X(z)$ does not exist for any z. The z-tranform of this sequence cannot be defined, since the sums do not converge.

Example

Now let us consider a slightly different variation on the two-sided above, let

$$x[n] = \left(\frac{1}{3}\right)^{|n|}, \quad -\infty < n < \infty.$$

For this sequence, we might have some hope of finding a range of values of z for which the z-transform will converge, since the sequence remains bounded for all n. In this case we write

$$x[n] = \left(\frac{1}{3}\right)^n u[n] + 3^n u[-n-1].$$

We can transform the right-sided and left-sided pieces individually, and add the results, by linearity of the transform, taking into account the regions in the complex plane for which the series will converge. Since each series has a different region of convergence, we need to consider, for the total sequence, only that portion of the complex plane that is common to both the ROC for the right-sided part and the left-sided part. That is, we need to know for which values of the complex plain will the total z-transform converge. This leads us to the following transform for the sequence:

$$X(z) = \frac{z}{z-\frac{1}{3}} - \frac{z}{z-3}, \quad \frac{1}{3} < |z| < 3.$$

This transform brings to bear an important property of the region of convergence for a two-sided z-transform, i.e. the two-sided transform of a two-sided sequence. If the algebraic form for a z-transform is $A(z)$, e.g. $X(z) = A(z), z \in ROC_X$, where

$$A(z) = \frac{N(z)}{(z-p_1)(z-p_2)\cdots(z-p_N)},$$

then ROC_X is generally smaller than the set of z where $A(z)$ alone is well defined. Indeed, $A(z)$ is well defined at all z except the pole locations $z = p_i, 1 \le i \le N$, whereas ROC_X must be a ring in the complex plane. It is important to remember that the z-transform of a sequence is not defined solely by an algebraic expression, but rather by the combination of an algebraic expression and the region of the complex plane over which the expression is correct. Outside of this region, the algebraic expression is not the z-transform of the sequence of interest. Some points to remember are that

1. Poles cannot lie in ROC_X (because even $A(z)$ is undefined at the pole locations).
2. ROC_X is generally smaller than the set of z where $A(z)$ is defined.
3. The z-transform, $X(z)$, is given by the pair of $A(z)$ and ROC_X.

Another example that will illustrate this point follows.

Example

Let the sequence $x[n]$ be defined as $x[n] = \left(\frac{1}{2}\right)^n u[n]$. The z-transform of the seqeunce can readily be found to be

$$X(z) = \frac{z}{z-\frac{1}{2}}, \quad |z| > \frac{1}{2}.$$

The algebraic form for $X(z)$ is defined everywhere except at $z = \frac{1}{2}$, and yet, the z-transform is not defined for $|z| < \frac{1}{2}$. For example, consider when $z = \frac{1}{4}$, for which we can evaluate the algebraic expression to be

$$\left.\frac{z}{z-\frac{1}{2}}\right|_{z=\frac{1}{4}} = -1.$$

However, this does not imply that $X(\frac{1}{4}) = -1$. Indeed, at $z = \frac{1}{4}$, $X(z)$ is not defined, since this is not in the region of convergence of the z-transform, i.e.,

$$X\left(\frac{1}{4}\right) = \left.\sum_{n=-\infty}^{\infty} \left(\frac{1}{2}\right)^n u[n] z^{-n}\right|_{z=\frac{1}{4}} = \sum_{n=0}^{\infty} 2^n,$$

which clearly fails to converge.

5.7 Properties of the two-sided z-transform

5.7.1 Linearity

When two sequences $x[n]$ and $y[n]$ have a two-sided z-transforms, $X(z)$ and $Y(z)$, respectively, then the superposition of these sequences will also have a two-sided z-transform, so long as $X(z)$ and $Y(z)$ are jointly defined on a non-null subset of the z-plane. Specifically, we have

$$w[n] = ax[n] + by[n] \Longleftrightarrow W(z) = aX(z) + bY(z), \quad ROC_W \supseteq ROC_X \cap ROC_Y,$$

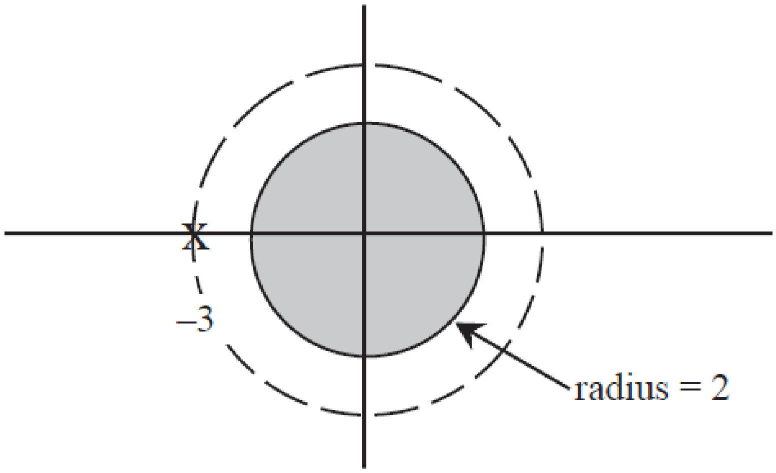

Figure 5.11: Region of convergence for the sum of two seqeunces.

that is, the region of convergence is at least as large as the intersection $ROC_X \cap ROC_Y$.

Example

Let us consider $w[n] = x[n] + y[n]$ with

$$\begin{aligned} X(z) &= \frac{z}{(z+2)(z+3)}, \quad |z| < 2, \\ Y(z) &= \frac{2}{z+2}, \quad |z| < 2, \end{aligned}$$

from which we have that

$$\begin{aligned} W(z) &= X(z) + Y(z) \\ &= \frac{z + 2(z+3)}{(z+2)(z+3)} \\ &= \frac{3(z+2)}{(z+2)(z+3)} \\ &= \frac{3}{z+3}. \end{aligned}$$

Now, the region of convergence of this expression must be determined. We know two things,

1. The ROC is bounded by poles
2. The ROC contains $ROC_X \cap ROC_y$.

For this example, there is a pole at $z = -3$. We also have that $ROC_X \cap ROC_Y = \{z : |z| < 2\}$as shown in Figure 5.11 .

We now can see that the proper region of convergence must be $ROC_W = \{z : |z| < 3\}$. So, the ROC can be larger than the intersection if we have pole-zero cancellation on the boundary of intersection, in which case, the ROC expands outward or inward to be bounded by another pole.

5.7.2 Shifting property

For the two-sided z-transform, the shifting properties are much simpler than their counterparts in the unilateral z-transform, since we do not need to worry about terms shifting in-to or out-of the summation defining the z-transform. We simply have

$$x[n] \longleftrightarrow X(z) \Longleftrightarrow x[n-k] \longleftrightarrow z^{-k}X(z)$$

and the region of convergence of the shifted sequence remains unchanged, except for the possible addition or deletion of $z = 0$ or $|z| = \infty$.

Example

Consder the seqeunce $x[n] = \delta[n-2]$ for which we have $Y(z) = z^{-2},\ |z| > 0$. now, if we let $y[n] = x[n+3] = \delta[n+1]$, then we have $Y(z) = z,\ |z| < \infty$. In this case, we see that $z = 0$ was added to the region of convergence and $|z| = \infty$ was removed from the region of convergence. The proof of the shifting property follows that for the unilateral z-transform, only simpler. We have

$$\begin{aligned} X(z) &= \sum_{n=-\infty}^{\infty} x[n]z^{-n} \\ Y(z) &= \sum_{n=-\infty}^{\infty} y[n]z^{-n} \\ &= \sum_{n=-\infty}^{\infty} x[n-k]z^{-n} \\ &= \sum_{m=-\infty}^{\infty} x[m]z^{-(m+k)} \\ &= z^{-k} \sum_{m=-\infty}^{\infty} x[m]z^{-m} \\ &= z^{-k}X(z), \end{aligned}$$

where in the fourth line, the change of variable $m = n - k$ was made.

5.7.3 Convolution

The convolution property for the two-sided z-transform follows similary from the unilateral case, for which we have

$$y[n] = \sum_{m=-\infty}^{\infty} h[m]x[n-m] \Longleftrightarrow Y(z) = H(z)X(z),\ ROC_Y \supseteq ROC_X \cap ROC_H,$$

so long as there exists a non-null intersection $ROC_X \cap ROC_H$. Just as with linearity, if there is pole-zero cancellation on a boundary of the intersection, then ROC_Y expands to the next pole.

Example

Consider the seqeunces $x[n]$ and $h[n]$ for which we have z-transforms $X(z)$ and $H(z)$ and define $Y(z)$ as follows

$$Y(z) = H(z)X(z),$$

where

$$\begin{aligned} H(z) &= \frac{1}{(z+1)(z+2)},\ \ 1 < |z| < 2, \\ X(z) &= \frac{z+1}{z+2},\ \ |z| < 2. \end{aligned}$$

Note that $ROC_H \cap ROC_X = \{z : 1 < |z| < 2\}$, however we have that $ROC_Y = \{z : |z| < 2\}$.

The convolution formula can be readily shown by taking the z-transform of both sides of the convolution sum. Since each of the steps in this derivation is reversible, this shows the if and only if nature of the convolution property. Specifically, we have

$$\begin{aligned}
y[n] &= \sum_{m=-\infty}^{\infty} h[m]x[n-m] \\
Y(z) &= \sum_{n=-\infty}^{\infty} \left(\sum_{m=-\infty}^{\infty} h[m]x[n-m] \right) z^{-n} \\
&= \sum_{m=-\infty}^{\infty} h[m] \left(\sum_{n=-\infty}^{\infty} x[n-m]z^{-n} \right) \\
&= \left(\sum_{m=-\infty}^{\infty} h[m]z^{-m} \right) X(z) \\
&= H(z)X(z).
\end{aligned}$$

5.8 The system function and poles and zeros of an LSI system

The transfer function of an LSI system with input $x[n]$ and output $y[n]$ is defined for two-sided z-transforms using

$$H(z) = \left. \frac{Y(z)}{X(z)} \right|_{\text{zero initial conditions.}}$$

Indeed, we have seen that $H(z)$ is independent of $X(z)$, and therefore independent of $x[n]$. For an LSI system, we can find $H(z)$ by a number of means. For example, we can

1. Directly compute the z-transform of $h[n]$ using the two-sided z-transform.
2. Compute the quantity $H(z) = Y(z)/X(z)$, for a given pair of input and output sequences $x[n]$ and $y[n]$.
3. Determine $H(z)$ directly from a block diagram description of the LSI system.

To further examine the last option, we will consider in more detail the methods used for analysis of LSI systems using a block diagram comprising delay, adder, and gain elements in Section 5.10.

5.9 Inverse two-sided z-transform

When taking an inverse two-sided z-transform, we can, once again, consider the complex contour-integral that defines its direct inversion, or, more simply, use methods such as partial fraction expansion to reduce a rational z-transform into a superposition of simpler terms, each of which can be inverted one at a time. Unlinke the unilateral z-transform, for each term in the partial fraction expansion, we now must consider the region of convergence of the overall transform and select the appropriate inverse transform sequence whose ROC would intersect with that of the overall transform to be inverted. To capture this notion graphically, consider Figure 5.12.

The poles that lie outside the ROC, i.e. those poles located such that $|p_i| > R_+$ correspond to terms in the partial fraction expation for which a left-sided inverse must be selected. The poles that lie inside the ROC, that is those poles located such that $|p_k| < R_-$correspond to terms in the partial fraction expation for which a right-sided inverse must be selected. These facts can be readily deduced as follows. The poles that lie inside the inner ring, i.e. those for which $|p_i| < R_-$must have a term in the partial fraction expansion for which the ROC for each pole intersects that of the overall z-transform. Since the poles are insize the ROC, the only possibility (out of the two choices, $|z| < |p_i|$ and $|z| > |p_i|$) that could possibly overlap with that of the overall ROC, $R_- < |z| < R_+$is $|z| > |p_i|$, which imples that each of these poles, labeled p_i^{RHS} correspond to right-sided inverse transforms, of the form $p_i^n u[n]$, assuming that the poles are not repeated roots. Similarly, the poles that lie outside the outer ring, i.e. those for which $|p_k| > R_+$must have a term in the partial fraction expansion for which the ROC for each pole intersects that of the overall z-transform.

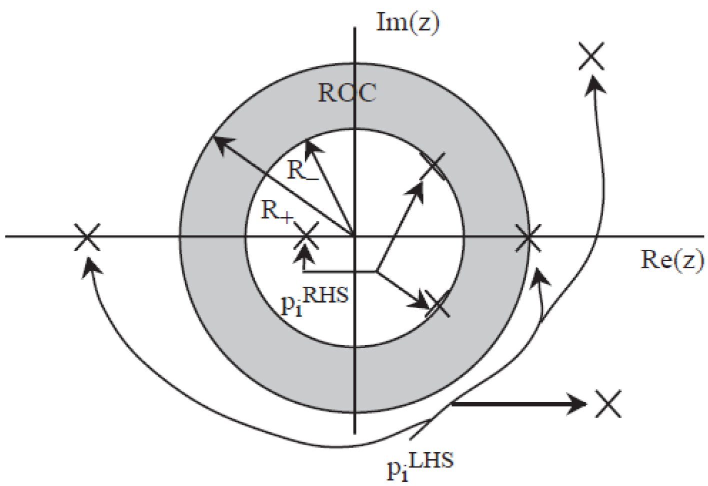

Figure 5.12: A graphical representation of the ROC for a two-sided rational z-transform that includes the locations of the poles.

Since the poles are outside the ROC, the only possibility (out of the two choices, $|z| < |p_k|$ and $|z| > |p_k|$) that could possibly overlap with that of the overall ROC, $R_- < |z| < R_+$is $|z| < |p_k|$, which imples that each of these poles, labeled p_k^{LHS} correspond to right-sided inverse transforms, of the form $-\left(p_k^n\right) u[-n-1]$, assuming again that the poles are not repeated roots.

Example

Let us consider a two-sided z-transform to invert as an example. Let $Y(z)$ be given such that the algebraic form is as follows

$$Y(z) = \frac{z}{(z-1)(z-2)}.$$

From this information alone, we are unable to compute $y[n]$, since there are three different regions of convergence that could be possible for this algebraic expression, and each would lead to a distinct, and different, $y[n]$. The three possibilities are $ROC_1 = \{z : |z| < 1\}$, $ROC_2 = \{z : 1 < |z| < 2\}$, and $ROC_3 = \{z : |z| > 2\}$, depicted in Figure 5.13.

These three possible ROCs lead to three different sequences, since we know that ROC_1 yields a left-sided sequence, $y_1[n]$, ROC_2 yields a two-sided sequence, $y_2[n]$, and ROC_3 yields a right sided sequence, $y_3[n]$. From the partial fraction expansion, we have

$$\begin{aligned} \frac{Y(z)}{z} &= \frac{A}{z-1} + \frac{B}{z-2} \\ Y(z) &= \frac{-z}{z-1} + \frac{z}{z-2}. \end{aligned}$$

The corresponding three inverse transforms would yield,

$$\begin{aligned} y_1[n] &= u[-n-1] - \left(2^n\right) u[-n-1], \\ y_2[n] &= -u[n] - \left(2^n\right) u[-n-1], \\ y_3[n] &= -u[n] + \left(2^n\right) u[n]. \end{aligned}$$

Example

Let us consider another example, this time with the ROC given. Let

$$Y(z) = \frac{z}{(z-2)(z-3)(z-4)}, \quad 2 < |z| < 3.$$

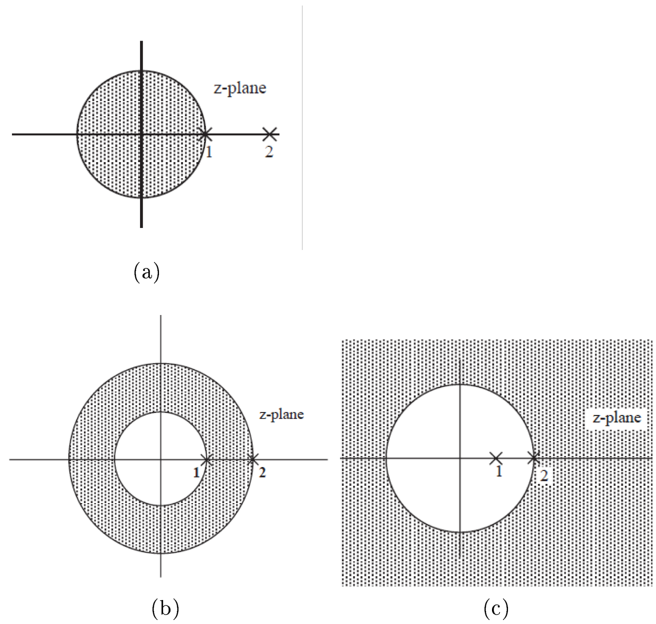

Figure 5.13: Three possible regions of convergence for the algebraic expression for $Y(z)$. Shown in (a) is ROC_1 corresponding to $y_1[n]$, in (b) is ROC_2 for $y_2[n]$ and in (c) is ROC_3 for $y_3[n]$.

From the partial fraction expansion, we have

$$Y(z) = \underbrace{\frac{\frac{1}{2}z}{z-2}}_{\text{right sided}} - \underbrace{\frac{z}{z-3}}_{\text{left sided}} + \underbrace{\frac{\frac{1}{2}z}{z-4}}_{\text{left sided}},$$

which yields,

$$y[n] = \frac{1}{2}(2^n)u[n] + 3^n u[-n-1] - \frac{1}{2}\left(4^n\right)u[-n-1].$$

Example

For another example, we consider a sequence with complex poles, i.e.

$$\begin{aligned} X(z) &= \frac{1}{(z+1)^2}, \quad |z| < 1. \\ &= \frac{1}{(z+j)(z-j)}, \end{aligned}$$

for which we have

$$\begin{aligned} \frac{1}{z(z+j)(z-j)} &= \frac{A}{z} + \frac{B}{(z+j)} + \frac{C}{(z-j)} \\ &= \frac{1}{z} + \frac{-\frac{1}{2}}{(z+j)} + \frac{-\frac{1}{2}}{(z-j)}/ \end{aligned}$$

This yields,

$$X(z) = 1 - \underbrace{\frac{\frac{1}{2}z}{z+j} - \frac{\frac{1}{2}z}{z-j}}_{\text{left sided}},$$

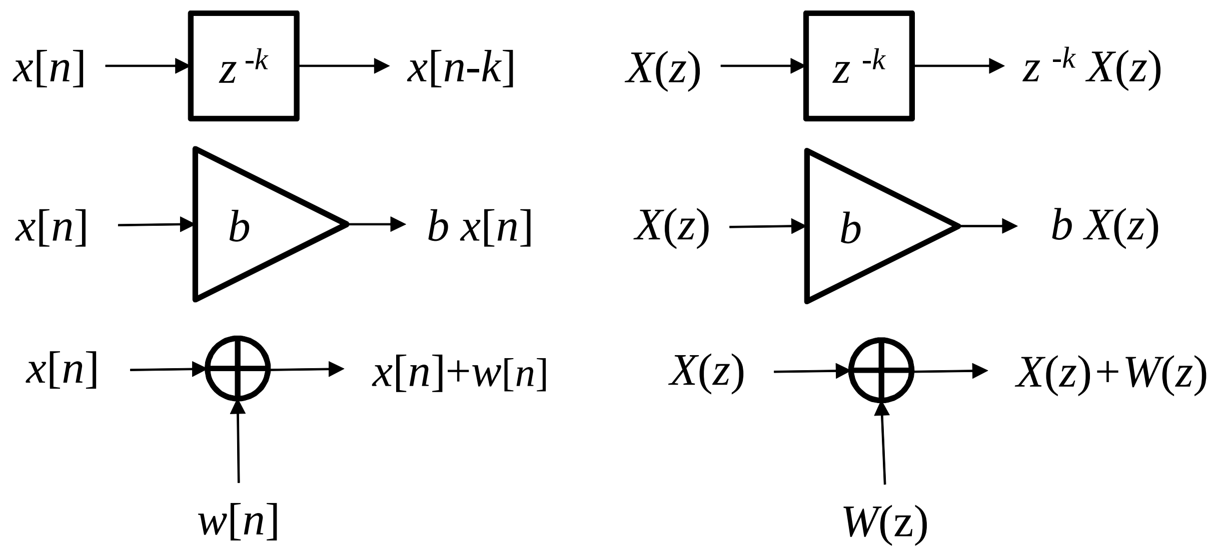

Figure 5.14: Basic elements of a delay-adder-gain flowgraph. To the left, the delay, gain, and adder elements are shown with their corresponding time-domain representation. To the right, the delay, gain and adder blocks are indicated with their corresponding z-transform representation.

from which we can onbain

$$\begin{aligned} x[n] &= \delta[n] + \frac{1}{2}(-j)^n u[-n-1] + \frac{1}{2}(j)^n u[-n-1] \\ &= \delta[n] + \frac{1}{2}\left(e^{-j(\pi/2)n} + e^{j(\pi/2)n}\right) u[-n-1] \\ &= \delta[n] + \cos\left(\frac{\pi}{2}n\right) u[-n-1] \\ &= \cos\left(\frac{\pi}{2}n\right) u[-n]. \end{aligned}$$

5.10 System Block Diagrams

To explore some of the methods for analyzing LSI system properties together with their implementation in hardware, we often use a delay-adder-gain model or flowgraph model for discrete-time LSI structures. In Figure

Shown in Figure 5.15 is a common delay-adder-gain block diagram for a second-order LSI system. In the figure, the notation for a delay element is that of a box labeled with z^{-1}inside. This is to denote that the operation of a delay element in the z-transform domain (through the delay property of z-transforms) is to multiply the input by z^{-1}. For example, the first delay element in the flowgraph, to the left, takes as input $x[n]$, which we depict in the z-transform domain as $X(z)$. The output of the delay element is the signal $x[n-1]$, i.e. the signal $x[n]$ delayed by one time unit. In the z-transform domain we write $x[n-1]$ as $z^{-1}X(z)$.

The transfer function of the LSI system shown in Figure 5.15 can be shown to be

$$H(z) = \frac{Y(z)}{X(z)} = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 - a_1 z^{-1} - a_2 z^{-2}}.$$

This can be shown as follows. First, we note that the flowgraph structure has only one adder node. If we write an equation for the output of the adder node as a function of its inputs, and do so using z-transform

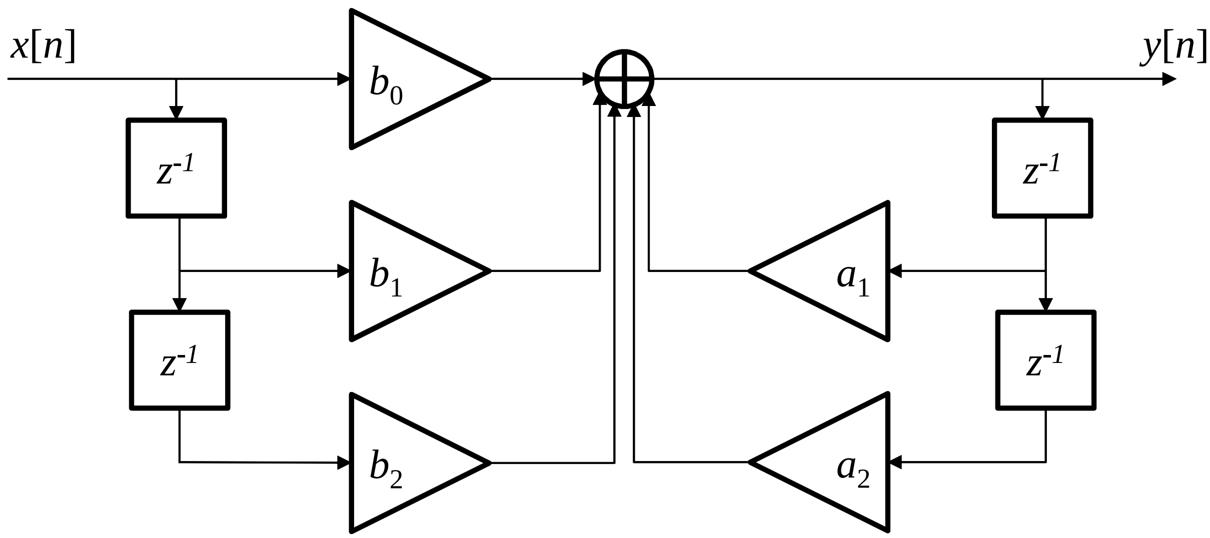

Figure 5.15: A direct-form I structure is a common delay-adder-gain model. Shown is a second-order DFI structure.

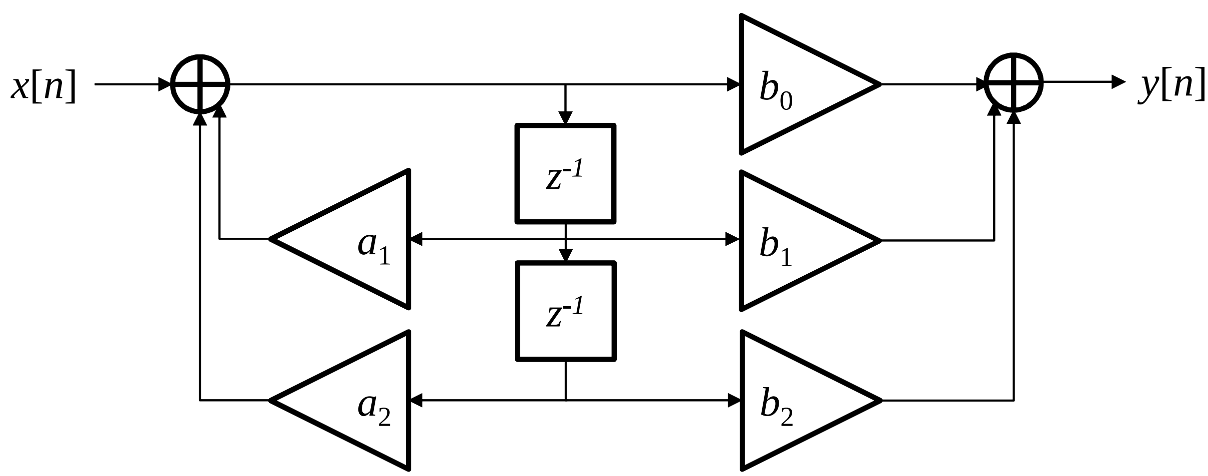

Figure 5.16: A delay-adder-gain model for a second order direct form II structure.

domain representation, using linearity and the delay property, we obtain

$$\begin{aligned}
Y(z) &= b_0X(z) + b_1z^{-1}X(z) + b_2z^{-2}X(z) + a_1z^{-1}Y(z) + a_2z^{-2}Y(z). \\
Y(z)\left[1 - a_1z^{-1} - a_2z^{-2}\right] &= X(z)\left[b_0 + b_1z^{-1} + b_2z^{-2}\right] \\
H(z) = \frac{Y(z)}{X(z)} &= \frac{b_0 + b_1z^{-1} + b_2z^{-2}}{1 - a_1z^{-1} - a_2z^{-2}}.
\end{aligned}$$

A second structure, called a direct form II structure is shown in Figure 5.16.

This structure can also be shown to have the same transfer function given by

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

through a method similar to that employed for the direct form I structure. Here we introduce a three-step method that is systematic and guaranteed to determine $H(z)$ for any cycle-free delay adder gain flowgraph. A cycle-free delay adder gain flowgraph is one in which all closed cycles contain at least one delay element. The three steps are as follows.

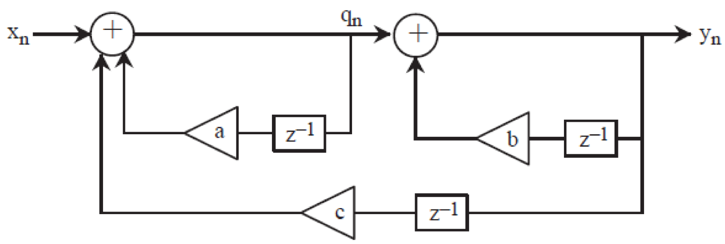

Figure 5.17: Flowgraph of a causal LSI system.

1. Label the output of each adder node in the flowgraph with a unique z-transform domain label.
2. Write an equation setting the output of each adder node in the flowgraph to the sum of the inputs to the adder node.
3. Use the resulting equations to remove all labels except for $X(z)$ and $Y(z)$, to obtain a single input-output relation from which $H(z)$ can be obtained by setting $H(z) = Y(z)/X(z)$.

The three steps are illustrated here for the direct for II structure. First, we note that there are two adder nodes in the flowgraph. The adder node to the left does not have a label, so we introduce a new sequence $q[n]$ as its output and label this $Q(z)$ in the z-transform domain. For this node, we obtain

$$Q(z) = X(z) + a_1 z^{-1} Q(z) + a_2 z^{-2} Q(z).$$

The output of the adder node to the right has already been labeled $y[n]$, so that in the z-transform domain we obtain

$$Y(z) = b_0 Q(z) + b_1 z^{-1} Q(z) + b_2 z^{-2} Q(z).$$

Finally, from these two equations, we can eliminate $Q(z)$ as follows

$$\begin{aligned} Q(z)\left[1 - a_1 z^{-1} - a_2 z^{-2}\right] &= X(z) \\ Q(z) &= \frac{X(z)}{1 - a_1 z^{-1} - a_2 z^{-2}} \end{aligned}$$

which can then be substituted into the expression for $Y(z)$ to yield

$$\begin{aligned} Y(z) &= \left[b_0 + b_1 z^{-1} + b_2 z^{-2}\right] Q(z) \\ Y(z) &= \left[b_0 + b_1 z^{-1} + b_2 z^{-2}\right] \frac{X(z)}{1 - a_1 z^{-1} - a_2 z^{-2}} \\ \frac{Y(z)}{X(z)} &= \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 - a_1 z^{-1} - a_2 z^{-2}}, \end{aligned}$$

as before. To futher illustrate this method, we consider another example.

Example

Consider the LSI system shown in Figure 5.17 .

The first step in our three step method is to label the outputs of each of the adder nodes. The first adder node to the left has $q[n]$ as its output and the second adder node has $y[n]$ as its output. For the first adder node, we have

$$Q(z) = X(z) + az^{-1} Q(z) + cz^{-1} Y(z)$$

and for the second adder node, we have

$$Y(z) = Q(z) + bz^{-1} Y(z).$$

Solving for $Q(z)$, we have

$$Q(z) = Y(z)\left(1 - bz^{-1}\right).$$

Plugging this into the other expression, we have

$$\begin{aligned} Y(z)\left(1-bz^{-1}\right)\left(1-az^{-1}\right) &= X(z)+cz^{-1}Y(z) \\ Y(z)\left(1-(a+b-c)z^{-1}+abz^{-2}\right) &= X(z) \\ H(z)=\frac{Y(z)}{X(z)} &= \frac{1}{1-(a+b-c)z^{-1}+abz^{-2}}. \end{aligned}$$

Note that the impulse response $h[n]$ and the system transfer function $H(z)$ are input-output descriptions of discrete-time LSI systems. These are also called "digital filters." Given an input $x[n]$ we can use either the impulse response to determine the output $y[n]$ through the convolution sum or we can use the system transfer function to compute the output through the z-transform. In this sense, both $h[n]$ and $H(z)$ summarize the behavior of the LSI system. However neither tells use what the internal structure of the digital filter is. Indeed, for any given system transfer function $H(z)$, there are an unlimited number of possible filter structures that have this same transfer function. For a second-order transfer function of the form

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

just two of the possible realizations are the direct form I and direct form II structures we have just visited. At this point, you may wonder how the filter structure or delay-adder-gain flowgraph relates to the actual filter implementaiton. The answer to this is multifacted. For example, let us consider the direct form I structure of Figure 5.15.

If the direct form I structure is implemeted in a digital signal processing microprocessor, then we note that there is a system clock that guides the operation of the filter. While the clock is not shown in the flowgraph, we know that the operation of the system depends on shifting values of the input into the system and computing values of the output that are then shifted out. It may take several clock cycles (microprocessor instructions) to compute each single value of the output sequence $y[n]$. For example, if the DSP has a single multiplier/accumulator (MAC), then the clock might trigger the following sequence of instructions

1. multiply $x[n]$ by b_0
2. multiply $x[n-1]$ by b_1 and add the result to 1)
3. multiply $x[n-2]$ by $b_2$and add the result to 2)
4. multiply $y[n-1]$ by $b_1$and add the result to 3)
5. multiply $y[n-2]$ by $b_2$and add the result to 4) to give $y[n]$.

The values of $x[n], x[n-1], x[n-2], y[n-1], y[n-2]$ are each stored in memory locations. You might expect that after $y[n]$ is computed, then in preparation for computing $y[n+1]$ we would use a sequence of instructions to move $x[n+1]$ into the old location for $x[n]$, move $x[n]$ into the old location for $x[n-1]$, move $x[n-1]$ into the old location for $x[n-2]$, move $y[n]$ into the old location for $y[n-1]$, and move $y[n-1]$ into the old location for $y[n-2]$. However, especially in higher order filters, this would be a huge waste of clock cycles. Instead, a pointer is used to address the proper memory location at each clock cycle. Therefore, it is not necessary to move data from memory location to memory location after computer each $y[n]$.

Just as there are a large number of filter structures that implement the same transfer, there are many algorithms (for a specific DSP) that can implement a given filter structure. Two important factors that you might consider in selecting a particular algorithm are the speed (number of clock cycles required to compute each output value) and the errors introduced through finite-precision effects, due to finite length registers used to represent the real-valued coefficients of the filter as well as the sequence values. We have not yet discussed finite register length effects, i.e. that the DSP has finite length registers for both memory locations as well as for the computations in the arithmetic units. This means that the digital filtering algorithm is not implemented in an exact manner. There will be error at the filter output due to coefficient quantization, and arithmetic roundoff. Of course, longer register lengths will reduce the error at the filter output. Generally, there is a tradeoff between algorithm speed and numerical precision. For a fixed register length, error usually can be reduced by using a more complicated (than Direct Form I or II) filter structure, requiring

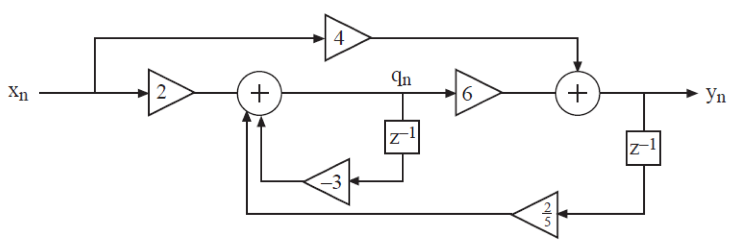

Figure 5.18: System flowgraph example.

more multiplications, additions, and memory locations. This in turn reduces the speed of the filter. The filter structure used in practice depends on $H(z)$ (some transfer functions are more difficult to implement with low error), on the available register length, and on the number of clock cycles available per output.

Example

Find the transfer function of the system in Figure 5.18 and construct a Direct Form II filter structure that implements the same transfer function.

We immediately label the output of the two adder nodes with the labels $y[n]$ and $q[n]$. From these we can then write

$$\begin{aligned} Y(z) &= 6Q(z) + 4X(z) \\ Q(z) &= 2X(z) - 3z^{-1}Q(z) + \frac{2}{5}z^{-1}Y(z). \end{aligned}$$

We can reduce these equations using

$$\begin{aligned} Q(z)\left(1+3z^{-1}\right) &= 2X(z) + \frac{2}{5}z^{-1}Y(z) \\ Q(z) &= \frac{2X(z)+\frac{2}{5}z^{-1}Y(z)}{(1+3z^{-1})} \end{aligned}$$

which yields

$$\begin{aligned} Y(z) &= 6\frac{2X(z)+\frac{2}{5}z^{-1}Y(z)}{(1+3z^{-1})} + 4X(z) \\ Y(z)\left(1-\frac{\frac{12}{5}z^{-1}}{1+3z^{-1}}\right) &= \left(\frac{12}{(1+3z^{-1})}+4\right)X(z) \\ H(z) = \frac{Y(z)}{X(z)} &= \frac{\left(\frac{12}{(1+3z^{-1})}+4\right)}{\left(1-\frac{\frac{12}{5}z^{-1}}{1+3z^{-1}}\right)} \\ H(z) &= \frac{\left(16+12z^{-1}\right)}{\left(1+\frac{3}{5}z^{-1}\right)}. \end{aligned}$$

The Direct Form II structure having this transfer function is now given in Figure 5.19 .

This structure is far simpler than the previous one and it computes exactly the same output $y[n]$. It is important to note that digital filter structures cannot have delay-free loops.

Example

Consider the filter structure shown in Figure 5.20.

This flowgraph depicts a system that is unrealizable. If we attempt to determine the input-output relation, we find

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

however the adder node has a delay-free loop which implies that the output at time n requires the addition of terms that include the output at time n. It is impossible therefore to compute $y[n]$ at any n.

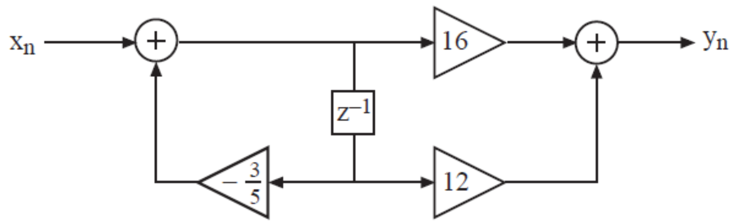

Figure 5.19: Direct Form II structure for this example.

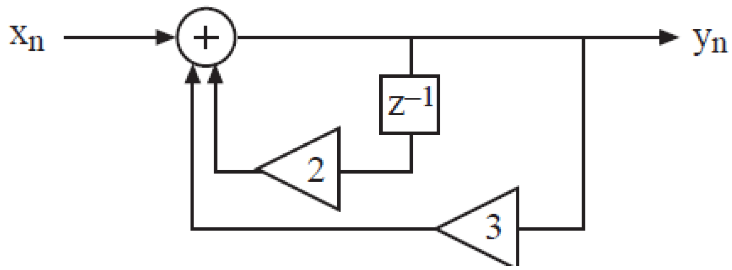

Figure 5.20: An unrealizable digital filter structure.

Consider the system shown in Figure 5.21below.

We can immediately write that

$$W(z) = H_1(z)X(z)$$

and that

$$Y(z) = H_2(z)W(z)$$

which leads to

$$\begin{aligned} Y(z) &= H_2(z)H_1(z)X(z) \\ \frac{Y(z)}{X(z)} = H(z) &= H_2(z)H_1(z) = H_1(z)H_2(z), \end{aligned}$$

where the last line follows from commutativity of multiplication of z-tranforms. This is known as a cascade combination of two LSI systems.

Consider the system shown in Figure 5.22below.

We can immediately write that

$$\begin{aligned} Y(z) &= H_1(z)X(z) + H_2(z)X(z) \\ &= (H_1(z) + H_2(z))\,X(z) \end{aligned}$$

which yields that

$$H(z) = \frac{Y(z)}{X(z)} = (H_1(z) + H_2(z))\,X(z).$$

This is known as a parallel combination of two LSI systems.

A feedback connection of two LSI systems is depicted in Figure5.23 .

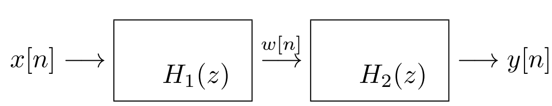

Figure 5.21: A cascade of two LSI systems.

$$x[n] \longrightarrow | \begin{array}{c} \longrightarrow \boxed{H_1(z)} \longrightarrow \\ \longrightarrow \boxed{H_2(z)} \longrightarrow \end{array} \oplus \longrightarrow y[n]$$

Figure 5.22: A cascade of two LSI systems.

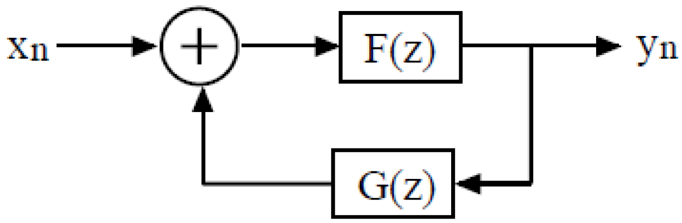

Figure 5.23: A feedback connection of two LSI systems.

The transfer function for a feedback connection of LSI systems can readily be obtained by again labeling the output of the adder node and writing an equation for its output. In this case, we have

$$W(z) = X(z) + G(z)Y(z)$$

and we have that

$$Y(z) = F(z)W(z)$$

which leads to

$$\begin{aligned} Y(z) &= F(z)\left(X(z) + G(z)Y(z)\right) \\ Y(z)\left(1 - F(z)G(z)\right) &= F(z)X(z) \\ Y(z) &= \frac{F(z)}{1 - F(z)G(z)}X(z) \end{aligned}$$

and finally,

$$H(z) = \frac{F(z)}{1 - F(z)G(z)}.$$

We see that for a feedback connection, the overall transfer function is given by the so-called "open loop gain" $F(z)$ divided by one minus the "closed loop gain", i.e. $1 - F(z)G(z)$.

5.11 Flowgraph representations of complex-valued systems

5.12 System analysis

As we have seen, the input-output relationship of a linear-shift invariant (LSI) system is captured through its response to a single input, that due to a discrete-time impulse, or the impulse response of the system. There are a number of important properties of LSI systems that we can study by observing properties of its impulse response directly. Perhaps one of the more important properties of such systems is whether or not they are stable, that is, whether or not the output of the system will remain bounded for all time when the input to the system is bounded for all time. While for continuous-time systems and circuits stability may be required for ensuring that components do not become damaged as voltages or currents grow unbounded in a system, for discrete-time systems, stability can be equally important. For example, practical implementations of many discrete-time systems involve digital representations of the signals. To ensure proper implementation of the operations involved, the numerical values of the signal levels must remain within the limits of the

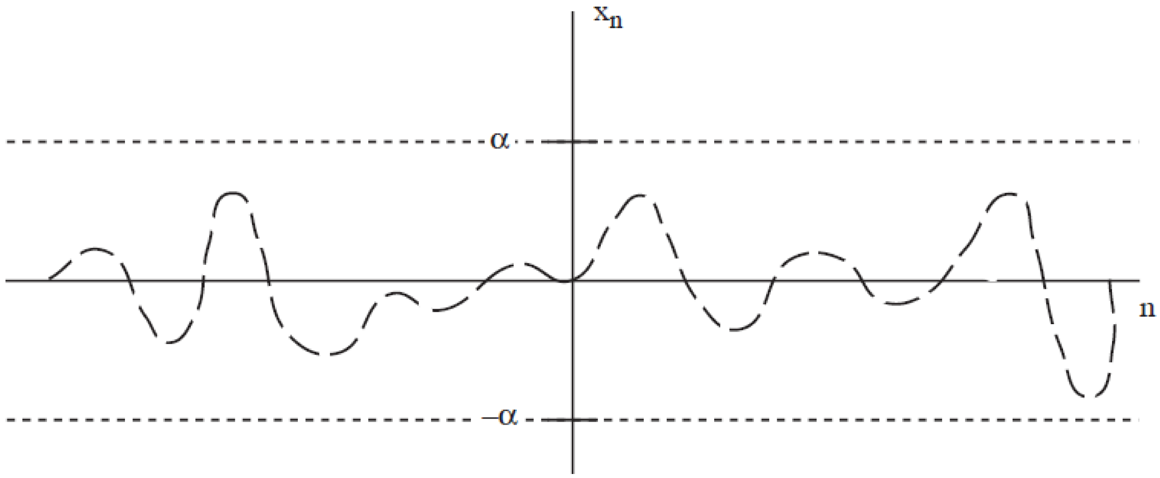

Figure 5.24: Bounded input $x[n]$, such that $|x[n]| < \alpha$.

number system used to represent the signals. If the signals are represented using fixed-point arithmetic, there may be strict bounds on the dynamic range of the signals involved. For example, any real number $-1 \leq x[n] \leq 1$ can be represented as an infinite binary string in two's complement notation as

$$x = -b_0 + \sum_{k=1}^{N} b_k 2^{-k}.$$

In a practical implementation, only finite-precision representations are available, such that all values might be represented and computed using fixed-point two's compliment arithmetic where any signal at a given point in time would be represented as a $B+1$-bit binary string $-1 \leq x[n] < 1$,

$$x = -b_0 + \sum_{k=1}^{B} b_k 2^{-k}.$$

Now, if the input signal such a system was carefully conditioned such that it was less than 1 in magnitude, it is important that not only does the output remain less than 1 in magnitude, but also all intermediate calculations must also. If not, then the numbers would overflow, and produce incorrect results, i.e. they would not represent the true output of the LSI system to the given input. If the discrete-time system were used to control a mechanical system such as an aircraft, such miscalculations due to instability of the discrete-time system could produce erratic or even catastrophic results.

5.13 BIBO stability

A system is bounded-input, bounded-output (BIBO) stable if for every bounded input, $x[n]$, the resulting output, $y[n]$, is bounded. That is, if there exists a fixed positive constant α, such that

$$|x[n]| < \alpha < \infty, \text{ for all } n,$$

then there exists a fixed positive constant β, such that

$$|y[n]| < \beta < \infty, \text{ for all } n,$$

where the constants α and β are fixed, meaning that they do not depend on n. Graphically, if every bounded input $x[n]$ as shown in Figure 5.24

causes a bounded output $y[n]$ as shown in Figure 5.25

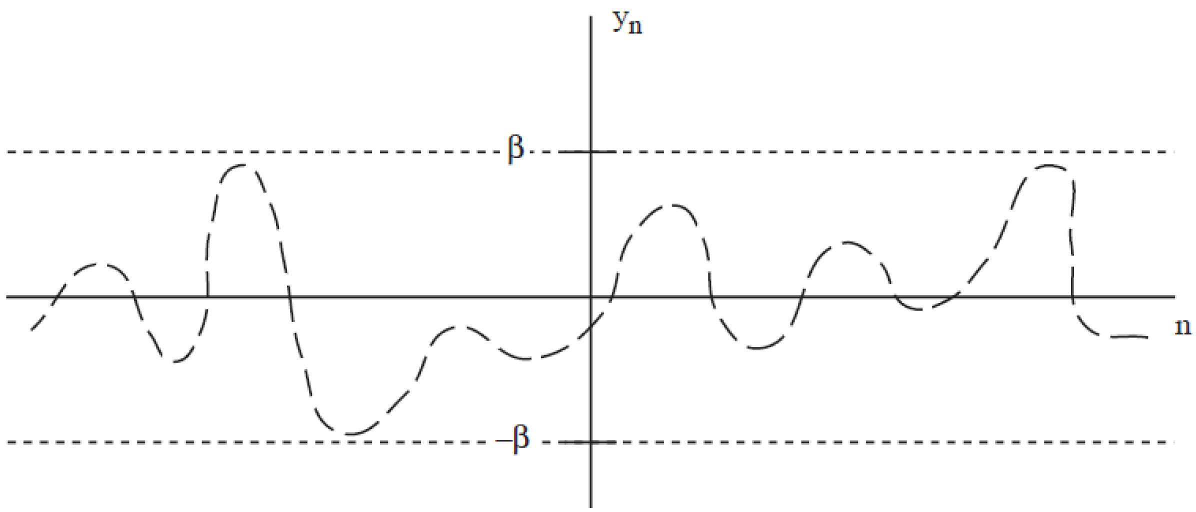

Figure 5.25: Bounded output $y[n]$, such that $|y[n]| < \beta$.

then the system is BIBO stable.then system is BIBO stable. Note that BIBO stability is a property of the system and not the inputs or outputs. While it may be possible to find specific bounded inputs such that the outputs remain bounded, a system is only BIBO stable if the output remains stable for all possible inputs. If there exists even one input for which the output grows unbounded, then the system is not stable in the BIBO sense.

How do we check if a system is BIBO stable? We cannot possibly try every bounded input and check that the resulting outputs are bounded. Rather, the input-output relationship must be used to prove that BIBO stability holds. Similarly, the following theorems can be used to provide simple tests for BIBO stability. It turns out that we can show that BIBO stability can be determined directly from the impulse respnse of an LSI system. Specifically, an LSI system with impulse response $h[n]$ is BIBO stable if and only if the impulse response is absolutely summable. That is,

$$\text{LSI system is BIBO stable} \iff \sum_{n=-\infty}^{\infty} |h[n]| < \infty.$$

To show both sides of the if and only if relationship, we start with assuming that $h[n]$ is absolutely summable, and seek to show that the output is bounded (sufficiency). This can be shown directly from the definition of an LSI system, i.e. from the convolution sum. We can write

$$y[n] = \sum_{m=-\infty}^{\infty} x[n-m]h[m].$$

Now, we take the absolute value of both sides and obtain

$$|y[n]| = \left| \sum_{m=-\infty}^{\infty} x[n-m]h[m] \right|,$$

which can be upper bounded by

$$|y[n]| \leq \sum_{m=-\infty}^{\infty} |x[n-m]||h[m]|.$$

Now we want to see that if $|x[n]| < \alpha$that we can find a suitable β such that $|y[n]| < \beta$. We have that

$$|y[n]| \leq \alpha \sum_{m=-\infty}^{\infty} |h[m]|,$$

and since we assumed that

$$\sum_{n=-\infty}^{\infty} |h[n]| = \gamma < \infty,$$

we have

$$|y[n]| \leq \alpha\gamma = \beta < \infty.$$

To show the other direction of the if and only if relation (necessity), we need to show that when the impulse response is not absolutely summable, then there exists a sequence $x[n]$ that is bounded, but for which the output of the system is not bounded. That is, given that the sum $\sum_{m=-\infty}^{\infty} |h[m]|$ diverges, we need to show that there exists a bounded sequence $x[n]$ that produces an output $y[n]$ such that for some fixed n_0 the convolution sum diverges, i.e., $y[n_0]$ is not bounded. From the convolution sum, we have

$$y[n_0] = \sum_{m=-\infty}^{\infty} x[m]h[n_0 - m].$$

By selecting the sequence $x[n]$ to be such that $x[m] = h^*[n_0 - m]/|h[n_0 - m]|$, (for real-valued $h[n]$, this amounts to $x[m] = \text{sgn}\ (h[n_0 - m]) = \pm 1$), then we have that

$$\begin{aligned} y[n_0] &= \sum_{m=-\infty}^{\infty} \frac{h^*[n_0 - m]h[n_0 - m]}{|h[n_0 - m]|} \\ &= \sum_{m=-\infty}^{\infty} \frac{|h[n_0 - m]|^2}{|h[n_0 - m]|} \\ &= \sum_{m=-\infty}^{\infty} |h[n_0 - m]| \end{aligned}$$

and letting $k = n_0 - m$, we obtain that

$$y[n_0] = \sum_{k=-\infty}^{\infty} |h[k]|,$$

which diverges, completing the proof.

BIBO stablility of a system can also be directly determined from the transfer function $H(z)$, relating the z-transform of the input to the z-transform of the output. Specifically, we have that for an LSI system with a rational transfer function, the system is BIBO stable if and only if the region of convergence includes the unit circle. For causal systems, this means that all of the poles of the system are inside the unit circle. Specifically, we have that

$$\text{An LSI system with transfer function } H(z) \text{is BIBO stable} \Longleftrightarrow ROC_H \subseteq |z| = 1.$$

We will show this result specifically for causal systems, noting that generalizing the result to left-sided and two-sided seqeunces is straightforward. First, to prove sufficiency, assume the region of convergence ROC_H includes the unit circle. Next, to illustrate that this implies absolute summability, i.e. $\sum_{n=-\infty}^{\infty} |h[n]| < \infty$, we consider the poles of the system function. First, the poles (roots of the denominator polynomial) must lie inside the unit circle since we have assumed that the region of convergence includes the unit circle, and for causal systems, i.e. systems for which $h[n] = 0$ for $n < 0$, we know ROC_H is given by $|z| > R$ for some $R > 0$. Since this must include the unit circle, then we have that $R < 1$ and all of the poles lie inside the unit circle.

The inverse z-transform, as determined by the partial fraction expansion of the system function $H(z)$ takes the form

$$h[n] = \sum_{k=0}^{N} b_k (p_k)^n, \ n \geq 0,$$

assuming there are no repeated roots in the denominator polynomial. Since we have that $|p_k| < 1$ for all of the poles, we know that

$$\begin{aligned}
\sum_{n=-\infty}^{\infty} |h[n]| &= \sum_{n=-\infty}^{\infty} \sum_{k=0}^{N} |b_k||p_k|^n u[n] \\
&= \sum_{k=0}^{N} |b_k| \sum_{n=-\infty}^{\infty} |p_k|^n u[n] \\
&= \sum_{k=0}^{N} \frac{|b_k|}{1-|p_k|} < \infty.
\end{aligned}$$

For the case of repeated roots, we would simply have to show that series of the form

$$\sum_{k=0}^{\infty} n^L (p_k)^n$$

are convergent. This is readily shown by the ratio test, where we compare the $(n+1)$th term to the nth term in the series. Here we have

$$\lim_{n\to\infty} \frac{(n+1)^L |p_k|^{n+1}}{n^L |p_k|^n} = \lim_{n\to\infty} \frac{(n+1)^L}{n^L} |p_k| = |p_k| < 1,$$

which implies that these series also all converge, indicating that even for repeated roots, we have that a causal LSI system whose ROC includes the unit circle will have an absolutely summable impulse response, and therefore will be BIBO stable.

To show necessity, we assume BIBO stability, and hence absolute summability of the impulse response, and then, for any point z on the unit circle, we have that

$$\begin{aligned}
|H(z)|_{|z|=1} &= \left| \sum_{n=0}^{\infty} h[n] z^{-n} \right|_{|z|=1} \\
&\leq \sum_{n=0}^{\infty} |h[n]||z|^{-n} \Big|_{|z|=1} \\
&\leq \sum_{n=0}^{\infty} |h[n]||1|^{-n} \\
&\leq \sum_{n=0}^{\infty} |h[n]| < \infty,
\end{aligned}$$

which implies that the region of convergence includes the unit circle and completes the proof. This indeed implies that for a causal LSI system with a rational transfer function (in minimal form), the system is BIBO stable if and only if all of its poles are inside the unit circle.

5.14 System properties from the system function

Some of the properties we have developed are explored in several examples.

Example

Consider the following LSI system with impulse response $h[n]$, we have that

$$h[n] = \cos(\theta n) u[n]$$

which leads to

$$\sum_{n=-\infty}^{\infty} |h[n]| = \sum_{n=0}^{\infty} |\cos(\theta n)|,$$

which diverges. Therefore, the system is not BIBO stable.

Example

Consider the following transfer function for a causal LSI system,

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

which after factoring the demoninator, yields,

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

We see that $H(z)$ has poles at $z = \pm \frac{j}{\sqrt{2}}$. The system is therefore causal and has all of its poles inside the unit circle. Therefore the system is BIBO stable. Note that as done in this example, factors that are common to the numerator and denominator must be cancelled before applying the stability test.

Example

Consider the following system function of an LSI system,

$$H(z) = \frac{z}{z + 100}, |z| < 100.$$

Note that this is a non-causal system, with a left-sided impulse response. The ROC in this case includes the unit circle, and therefore the system is BIBO stable.

Example

Consider the following impulse response of an LSI system,

$$h[n] = \begin{cases} 4^n, & 0 \le n \le 10^6, \\ n\left(\frac{1}{2}\right)^n, & n > 10^6 \\ 0, & n < 0. \end{cases}$$

Testing for absolute summability of the impulse response, we see that

$$\sum_{n=0}^{\infty} |h[n]| = \sum_{n=0}^{10^6} 4^n + \sum_{n=10^6+1}^{\infty} n\left(\frac{1}{2}\right)^n < \infty,$$

and therefore the system is BIBO stable.

We continue exploring the properties of LSI systems through observation of their system functions (that is, the z-transform of the impulse response), with a focus on the relationship between the region of convergence of the z-transform and the stability and causality of the system.

Example

Consider the following system function of a stable LSI system,

$$H(z) = \frac{z}{(z - \frac{1}{4})(z-2)},$$

can it be causal?

Answer: No, it cannot be causal. First, note that although the region of convergence is not explicitly stated, it is implicitly determined. Noting that the system is stable, we know that the region of convergence must include the unit circle. Given the pole locations, we know that the region of convergence must be $z : \frac{1}{4} < |z| < 2$ implying that the impulse response will have leftsided and right-sided components and that $h[n]$ must be two-sided, i.e. that $H(z)$ is a two-sided z-transform. Since the impulse response is two-sided, this implies that the system cannot be causal, i.e. $h[n]$ is non-zero for $n < 0$ and from the convolution sum,

$$y[n] = \sum_{k=-\infty}^{\infty} h[m]x[n-m],$$

we see that this implies that $y[n]$ depends on values of $x[m]$ for $m > n$.

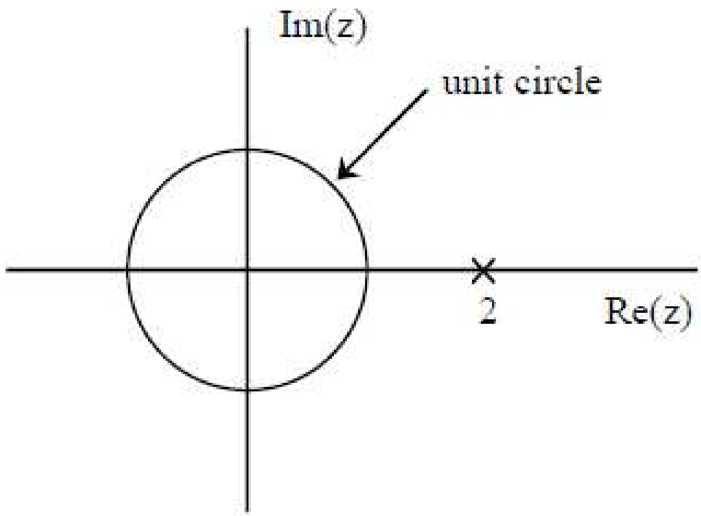

Figure 5.26: Pole-zero plot for an LSI system.

Example

Consider the following discrete time system,

$$y[n] = (x[n])^2.$$

Is this system stable?

Answer: This system is not linear. Therefore, we cannot apply a stability test involving either the impulse response or transfer function, since the tests discussed so far apply only to LSI systems. Since this system is not LSI, the convolution sum does not hold, so that the input output relationship does not satisfy $y[n] = x[n] * h[n]$or $Y(z) = H(z)X(z)$. Instead, we appeal to the definition of BIBO stability.Since $|x[n]| < \alpha$for all n, then we have that $|y[n]| < \alpha^2 < \infty$ for all n. Therefore, the system is indeed stable, albeit nonlinear.

Unbounded Outputs

Given an unstable LSI system, how do we find a bounded input that will cause an unbounded output? This will be illustrated by example for some causal systems in the following examples.

Example

Consider the following causal LSI system with pole-zero plot shown in Figure 5.26 and with system function $H(z)$ given by

$$H(z) = \frac{z}{z-2}, |z| > 2.$$

The impulse response is therefore given by $h[n] = 2^n u[n]$ and is itself unbounded. Since $h[n]$ grows without bound, almost any bounded input will cause the output to be unbounded. For example, taking $x[n] = \delta[n]$would yield $y[n] = h[n]$.

Example

Now consider the following LSI system with system function

$$H(z) = \frac{z}{z-1}, |z| > 1.$$

Although the system is not stable, the impulse response remains bounded, as $h[n] = u[n]$, in this case. Here we could choose $x[n] = u[n]$ (which is bounded) so that $y[n]$ will be a linear ramp in time. Looking at the z-transform of the output, this corresponds to forcing $Y(z)$ to have a double pole at $z = 1$, i.e.

$$Y(z) = H(z)X(z) = \frac{z^2}{(z-1)},$$

which for the region of convergence of this output corresponds to a sequence that grows linearly in time.

Example

Here we consider an LSI system with a complex-conjugate pole pair on the unit circle. Let

$$H(z) = \frac{z^2 - z\cos(\alpha)}{(z - e^{j\alpha})(z - e^{-j\alpha})}, |z| > 1.$$

The complex conjugate pair of poles on the unit circle corresponds to a sinusoidal oscillating impulse response,

$$h[n] = \cos(\alpha n)u[n].$$

Thinking of the z-transform of the output, note that choosing $x[n] = h[n]$ will cause $Y(z)$ to have double poles at $z = e^{\pm j\alpha}$, which will in turn cause $y[n]$ to have the form of n times $\cos(\alpha n)$, which grows unbounded. From these examples with causal systems, we see that for systems with poles outside the unit circle, since the impulse response itself grows unbounded, substantial effort would be required to find a bounded input that will not cause an unbounded output. For poles on the unit circle, it is more difficult to find bounded inputs that ultimately cause the output to be unbounded. In some fields, such as dynamic systems or control, LSI systems with poles on the unit circle are called “marginally stable” systems. In our terminology, they are simply unstable systems.