2

Review of Discrete-Time Systems

2.0 INTRODUCTION

This chapter provides a brief review of the fundamentals of discrete-time systems. This serves as quick reference material throughout the text, and also familiarizes the reader with the notations we will use. In this chapter, we will present basic facts and results without detailed justification. Detailed treatements can be found in Oppenheim and Schafer [1989]. Other related references are Rabiner and Gold [1975], and Jackson [1991].

2.1 DISCRETE-TIME SIGNALS

Discrete-time signals are typically denoted as $u(n)$, $x(n)$, and so on, where n is an integer called the *time index.* We will use notations such as $x(n)$ to indicate the entire sequence (i.e., with $-\infty \leq n \leq \infty$) or, on occasions, just to denote the nth sample $x(n)$. The context will clarify the exact meaning. All sequences are taken to be complex valued unless mentioned otherwise. Figure 2.1-1 shows some typical sequences.

1. The *unit-pulse,* denoted $\delta(n)$, is defined according to

$$\delta(n) = \begin{cases} 1, & n = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{2.1.1}$$

   This is sometimes called the *impulse function,* and should not be confused with the impulse function $\delta_a(\alpha)$ of the real continuous variable α. The function $\delta_a(\alpha)$ is usually called the *Dirac delta function,* and is defined to be zero everywhere except $\alpha = 0$, and such that $\int_p^q \delta_a(\alpha)d\alpha = 1$ if, and only if, $p < 0 < q$.
2. The *unit-step* sequence is defined as

$$\mathcal{U}(n) = \begin{cases} 1, & n \geq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{2.1.2}$$

3. *Exponentials.* A sequence of the form ca^n is said to be an exponential. Here c and a are arbitrary (possibly complex) constants. Sequences such as $ca^n\mathcal{U}(n)$ and $cb^n\mathcal{U}(-n)$ are called *one-sided exponentials* (or truncated exponentials). Thus, $ca^n\mathcal{U}(n)$ is right-sided, whereas $cb^n\mathcal{U}(-n)$ is left-sided.

4. *Single-frequency sequence.* The sequence $ce^{j\omega_0 n}$ is said to be a single-frequency sequence. This is an exponential sequence with $a = e^{j\omega_0}$. Here, ω_0 is real, but can have either sign. In less formal terms this is sometimes called a *sinusoid* with frequency ω_0. This is periodic if, and only if, the frequency ω_0 is a rational multiple of 2π, that is, $\omega_0 = 2\pi k/L$ for integer k and L.

5. A sequence of the form $A\cos(\omega_0 n + \theta)$ is a true *sinusoid.* Since we can write $\cos(\omega_0 n + \theta) = 0.5(e^{j(\omega_0 n+\theta)} + e^{-j(\omega_0 n+\theta)})$, it contains two frequencies, that is, ω_0 and $-\omega_0$. So it is not a single-frequency signal.

6. *Bounded sequence.* A sequence $u(n)$ is said to be bounded if there exists a finite B such that $|u(n)| \leq B$ for all n. Examples: (a) $a^n\mathcal{U}(n), |a| < 1$, (b) $\cos\omega_0 n$ (real ω_0). Note that an exponential a^n is not bounded unless $a = 0$ or $|a| = 1$.

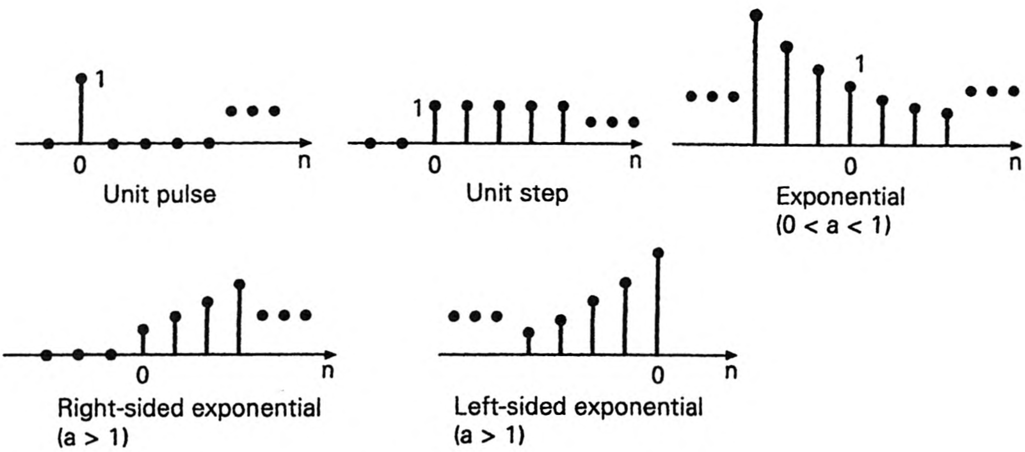

Figure 2.1-1 Demonstration of well known sequences.

2.1.1 Transform Domain Analysis

It is often convenient to work with transformed versions of signals such as the z-transform and Fourier transform. These are defined next.

The z-Transform

The z-transform of a sequence $x(n)$ is defined as

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n}. \tag{2.1.3}$$

If this summation does not converge for any z, the z-transform does not exist; an example is the exponential sequence $a^n, a \neq 0$.

In general, the summation converges in an annulus defined as $R_1 < |z| < R_2$ in the z-plane. This is called the *region of convergence* (ROC). For example if $x(n) = a^n\mathcal{U}(n)$, then $X(z) = 1/(1 - az^{-1})$ with ROC given by $|z| > |a|$. This same $X(z)$ with ROC specified as $|z| < |a|$ would result in the inverse transform $x(n) = -a^n\mathcal{U}(-n-1)$. Given $X(z)$ and its associated ROC, $x(n)$ can be uniquely recovered from $X(z)$.

For a finite length sequence (with finite sample values) the z-transform converges *everywhere* except possibly at $z = 0$ and/or $z = \infty$.

The Fourier Transform (FT)

If the ROC of $X(z)$ includes the unit circle (i.e., points of the form $z = e^{j\omega}$ where ω is real), we say that $X(e^{j\omega})$ is the Fourier transform (FT) of $x(n)$. Thus,

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n}. \tag{2.1.4a}$$

The inverse transform relation is given by

$$x(n) = \frac{1}{2\pi}\int_0^{2\pi} X(e^{j\omega})e^{j\omega n}d\omega. \tag{2.1.4b}$$

Note that the frequency variable ω is in radians. The FT $X(e^{j\omega})$ is periodic in ω with period 2π, so that the region $\pi < \omega < 2\pi$ is considered to be the negative frequency region (equivalent to $-\pi < \omega < 0$).

Fourier transform of a single-frequency signal. Since the z-transform of a^n does not converge anywhere (unless $a = 0$), the Fourier transform of $e^{j\omega_0 n}$, in particular, does not exist in the usual sense. However, by using a Dirac delta function $\delta_a(\omega)$, one can write the FT of this sequence as $2\pi\delta_a(\omega - \omega_0)$ in the range $0 \leq \omega < 2\pi$ (and periodically repeating with period 2π).

Parseval's Relation

Let $U(e^{j\omega})$ and $V(e^{j\omega})$ be the Fourier transforms of $u(n)$ and $v(n)$. Parseval's relation says

$$\sum_{n=-\infty}^{\infty} u(n)v^*(n) = \frac{1}{2\pi}\int_0^{2\pi} U(e^{j\omega})V^*(e^{j\omega})d\omega. \tag{2.1.5a}$$

In particular, if we set $u(n) = v(n)$, then

$$\sum_{n=-\infty}^{\infty} |u(n)|^2 = \frac{1}{2\pi}\int_0^{2\pi} |U(e^{j\omega})|^2 d\omega. \tag{2.1.5b}$$

Energy of a sequence. The energy of a sequence $u(n)$ is defined as $E_u = \sum_{n=-\infty}^{\infty} |u(n)|^2$. If this summation does not converge, the energy is taken to be infinite. Eq. (2.1.5b) gives us two ways to express the energy.

Tables of z-transform and Fourier transform pairs, and Tables of their properties can be found in Chap. 2 and 4 of Oppenheim and Schafer [1989]. We will make use of these throughout the text.

2.1.2 Discrete-Time Systems

A discrete-time system operates on an input sequence $u(n)$ to produce an output sequence $y(n)$. It is assumed that the value of $u(n)$ in the range $-\infty \leq n \leq \infty$ uniquely determines the output $y(n)$ in the range $-\infty \leq n \leq \infty$. Of great interest to us in this text are linear systems and shift invariant (or time invariant) systems. The properties of linearity and shift invariance enable us to characterize the system using the notion of transfer functions.

Linearity. Suppose the input sequences $u_0(n)$ and $u_1(n)$ produce the output sequences $y_0(n)$ and $y_1(n)$ respectively. If the system output in response to the input $a_0u_0(n) + a_1u_1(n)$ is equal to $a_0y_0(n) + a_1y_1(n)$, and if this is true for every pair of constants a_0, a_1 and every possible $u_0(n)$ and $u_1(n)$, then we say that the system is linear.

Shift-invariance. Let $y(n)$ denote the output of a system in response to the input $u(n)$. If the output in response to the shifted version $u(n-N)$ is equal to $y(n-N)$, and if this holds for all integers N and all input sequences $u(n)$, we say that the system is shift-invariant or time-invariant.

LTI Systems

A system is said to be linear and shift-invariant (abbreviated LSI or LTI) if it is both linear and shift-invariant. Such a system can be completely characterized by the *impulse response* sequence $h(n)$ (also called the *unit-pulse response*) which is the output $y(n)$ in response to an unit-pulse input $\delta(n)$. For LTI systems, the input-output relation is given by

$$y(n) = \sum_{m=-\infty}^{\infty} h(m)u(n-m), \tag{2.1.6}$$

which is called the *convolution summation.* This can be expressed in the transform domain as

$$Y(z) = H(z)U(z), \tag{2.1.7}$$

where $H(z)$ is the z-transform of $h(n)$, that is,

$$H(z) = \sum_{n=-\infty}^{\infty} h(n)z^{-n}. \tag{2.1.8}$$

$H(z)$ is called the *transfer function* of the LTI system. To physically visualize the meaning of $H(z)$, note that if we apply an exponential input a^n, the output is also an exponential, given by $y(n) = H(a)a^n$ [provided "a" belongs to the region of convergence of $H(z)$].

Eigenfunctions of LTI systems. If a nonzero input $f(n)$ to an LTI system $H(z)$ produces an output of the form $cf(n)$, where c is a constant, we say that $f(n)$ is an eigenfunction (and c an eigenvalue) of $H(z)$. Thus, *exponentials* are eigenfunctions of LTI systems.

Causality

A discrete-time system is said to be causal if the output $y(n)$ at time n does not depend on the future values of the input sequence, that is, does not depend on $u(m), m > n$. An LTI system is causal if and only if the impulse response satisfies the condition

$$h(n) = 0, \quad n < 0 \quad \text{(causality)}. \tag{2.1.9}$$

This has given rise to the term 'causal sequence' for any sequence $x(n)$ which is zero for $n < 0$. We say that a sequence $x(n)$ is *anticausal* if $x(n) = 0$ for $n \geq 0$. An example is $\mathcal{U}(-n-1)$.

From (2.1.8), we see that a causal LTI system has $H(\infty) = h(0)$, and that an LTI system is causal if and only if $H(\infty)$ is finite. For convenience of language, we often use phrases such as '$H(z)$ is causal'. This means that the associated ROC has been so chosen that the inverse transform $h(n)$ is zero for $n < 0$.

Rational Transfer Functions

All transfer functions in this text are rational, that is, of the form

$$H(z) = \frac{A(z)}{B(z)}, \tag{2.1.10}$$

with

$$A(z) = \sum_{n=0}^{N} a_n z^{-n}, \quad B(z) = \sum_{n=0}^{N} b_n z^{-n}. \tag{2.1.11}$$

Here a_n and b_n are (possibly complex) finite numbers. If there are no common factors of the form $(\beta - \alpha z^{-1}), \alpha \neq 0$ between $A(z)$ and $B(z)$ (i.e., if $A(z)$ and $B(z)$ are relatively prime), we say that (2.1.10) is an *irreducible* rational form. Under this condition, N is called the order of the system (assuming that at least one of a_N or b_N is nonzero).

Zeros and poles. If $A(z)/B(z)$ is irreducible, the zeros of $A(z)$ and $B(z)$ are said to be the zeros and poles, respectively, of $H(z)$. The time-domain significance of poles and zeros is well known, and is discussed in Problems 2.4 to 2.6.

Realness. A system is said to be 'real' if the output $y(n)$ is real for real inputs $u(n)$. For LTI systems this is equivalent to the condition that $h(n)$ be real for all n. For rational LTI system in the irreducible form (2.1.10), this in turn is equivalent to the condition that a_n and b_n be real for all n. Real systems are also referred to as real coefficient systems.

FIR and IIR Systems

A finite impulse response (FIR) system is one for which b_n in (2.1.11) is nonzero for only one value of n. As an example, let $N = 3$, and let $b_2 = 1$. Then, $H(z) = a_0 z^2 + a_1 z + a_2 + a_3 z^{-1}$, and is FIR.

A *causal* Nth order FIR filter can be represented as

$$H(z) = \sum_{n=0}^{N} h(n) z^{-n}, \quad h(N) \neq 0. \tag{2.1.12}$$

[This corresponds to $B(z) = 1$ and $A(z) = H(z)$.] The quantity $N+1$, which is the number of impulse response coefficients, is said to be the *length* of the filter (i.e., $H(z)$ is an $(N+1)$-point filter). In Sec. 2.4.2 we will see that FIR systems can be designed to have exactly linear phase response, which is required in some applications.

An LTI system which is not FIR is said to be an IIR (Infinite impulse response) system. An example is the system with impulse response $a^n \mathcal{U}(n)$, which has transfer function $H(z) = 1/(1 - az^{-1})$.

All-zero and all-pole systems. An FIR system is also said to be an all-zero system (because poles are located only at $z = 0$ and/or ∞). An IIR system of the form $H(z) = cz^{-K}/B(z)$ is said to be an all-pole system. The zeros for such a system are at $z = 0$ and/or ∞.

FIR sequences. A finite-duration (or finite-length) sequence $u(n)$ is often referred to as an FIR sequence. We often use the terms "FIR input", "FIR output" and so on, where the term FIR actually stands for "finite-length".

Stability

If a discrete-time system is such that every bounded input produces a bounded output, we say that the system is stable (more precisely *bounded input bounded output* stable or BIBO stable). For the case of LTI systems, it can be shown that BIBO stability is equivalent to the condition

$$\sum_{n=-\infty}^{\infty} |h(n)| < \infty. \tag{2.1.13}$$

In other words, the impulse response must be *absolutely summable.*

Stability condition in terms of poles. If $H(z)$ is rational and $h(n)$ causal, then (2.1.13) is equivalent to the condition that all poles p_k of $H(z)$ be inside the unit circle, that is, $|p_k| < 1$. Unless mentioned otherwise, the statement "stable" in this text will imply this condition (i.e., $|p_k| < 1$).

Causality of $h(n)$ and rationality of $H(z)$ will be implicit. The only exception to this convention will be noncausal filters of the form $z^K H(z)$, where $K > 0$ and $H(z)$ is causal and stable. This system has K poles at $z = \infty$, yet it is stable.

2.1.3 Implementations of Rational Transfer Functions

It is well known that the system with transfer function (2.1.10) can be described in the time domain by a difference equation of the form

$$b_0 y(n) = -\sum_{m=1}^{N} b_m y(n-m) + \sum_{m=0}^{N} a_m u(n-m). \tag{2.1.14}$$

Since the output $y(n)$ depends in general on past outputs $y(n-m)$, this is called a *recursive difference equation.* Without loss of generality, we can assume that at least one of a_0, b_0 in (2.1.11) is nonzero. If $b_0 = 0$ and $a_0 \neq 0$ this implies a noncausal system [since $H(\infty)$ is then not finite]. For causal systems $b_0 \neq 0$, and we can assume $b_0 = 1$ without loss of generality.

Direct form structure. With $b_0 = 1$ we obtain the structure of Fig. 2.1-2(a) (demonstrated for $N = 2$) for the implementation of this difference equation. This is called the *direct form structure,* and requires $2N+1$ multiplications and $2N$ additions for the computation of each output sample $y(n)$. The number of delays is N, which is the filter order.

Figure 2.1-2(b) shows the common notations and building blocks (multipliers, adders and delays) used to draw digital filter structures. Multipliers with values $\pm 2^{\pm K}$ are often not counted as multipliers, as these can be implemented with binary shifts on a digital machine.

FIR direct form. For the special case of FIR filters the structure reduces to the form shown in Fig. 2.1-3(a) (assuming causality), requiring $N+1$ multipliers, N adders, and N delays. An equivalent structure called the *transposed* direct form is shown in Fig. 2.1-3(b). FIR structures do not have any feedback paths (unlike Fig. 2.1-2(a)). Equivalently, the difference equation (2.1.14) has only the input terms $u(n-m)$ and no $y(n-m)$ terms on the right hand side, that is, the difference equation is nonrecursive.

Cascade form structures. Another popular structure used in digital filtering is the cascade form. This is obtained by expressing $A(z)$ and $B(z)$ in factored form as

$$A(z) = a_0 \prod_{n=1}^{N} (1 - z_n z^{-1}), \quad B(z) = \prod_{n=1}^{N} (1 - p_n z^{-1}), \tag{2.1.15}$$

so that

$$H(z) = \frac{a_0 \prod_{n=1}^{N} (1 - z_n z^{-1})}{\prod_{n=1}^{N} (1 - p_n z^{-1})}. \tag{2.1.16}$$

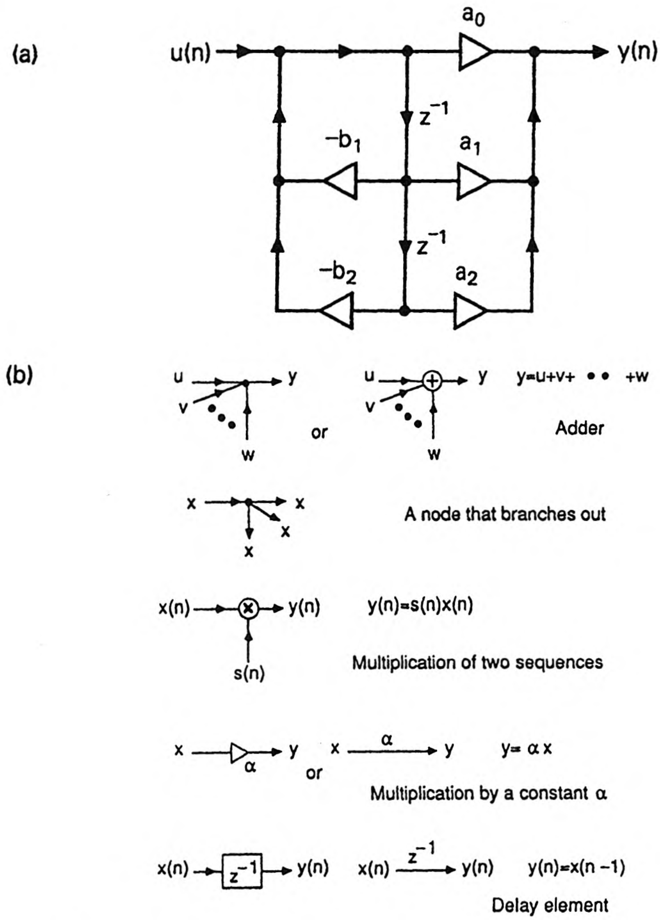

Figure 2.1-2 (a) The direct form structure for $N = 2$. (b) Meanings of signal flow graph notations used.

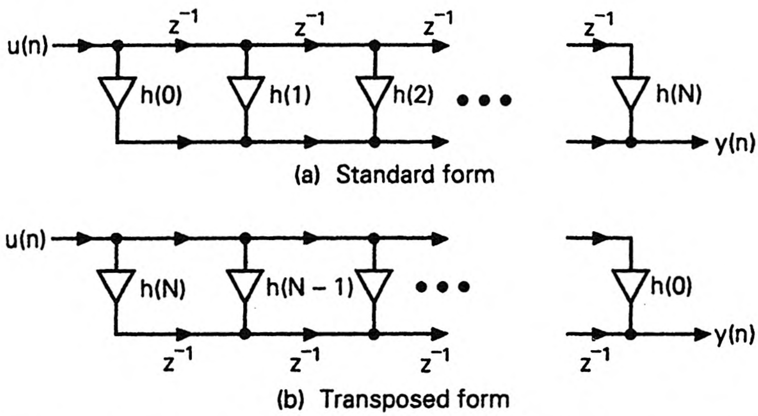

Figure 2.1-3 Direct form structures for FIR filters. (a) Standard form, and (b) transposed form.

Figure 2.1-4(a) shows the cascade form structure, where the building blocks are as in Fig. 2.1-4(b). Notice that p_n and z_n are, respectively, the poles and zeros of $H(z)$.

There exist more complicated structures for filters. A useful tool to compute the transfer functions of arbitrary structures is provided by *Mason's formula*, reviewed in Appendix E.

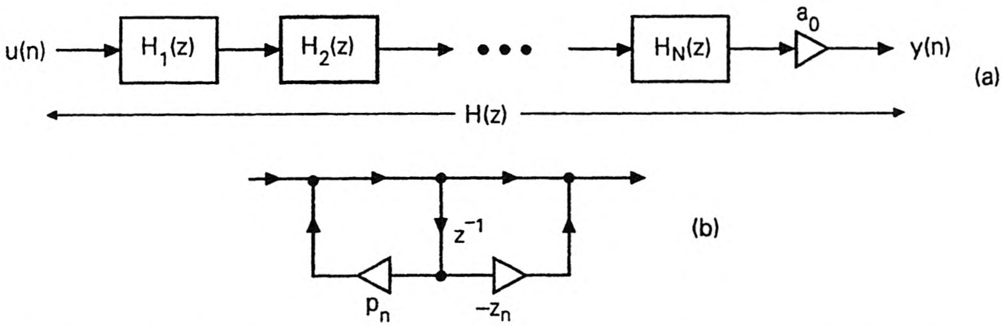

Figure 2.1-4 (a) The cascade form structure, and (b) the first order building blocks. Here z_n and p_n can be complex.

Real Coefficient Case

In the above cascade form structure, the multipliers can in general be complex even if a_n and b_n are real. For the special case where b_n are real, the poles are either real or occur in complex conjugate pairs. (The same is true of zeros if a_n are real.)

If p_k is complex, we can combine the factors $1 - p_k z^{-1}$ and $1 - p_k^* z^{-1}$ to produce the real factor $1 - c_k z^{-1} - d_k z^{-2}$. In this way, we can obtain a real coefficient cascade form. The transfer function can now be expressed as

$$H(z) = \frac{a_0(1 + az^{-1})^i}{(1 + bz^{-1})^\ell} \frac{\prod_{k=1}^{m}(1 + t_k z^{-1} + q_k z^{-2})}{\prod_{k=1}^{m}(1 - c_k z^{-1} - d_k z^{-2})}, \tag{2.1.17}$$

where i and ℓ can take the values 0 or 1. All coefficients in (2.1.17) are real. The second order sections can be implemented as in Fig. 2.1-5, using the direct form structure.

A complex conjugate pair of poles can also be represented as $R_k e^{\pm j\phi_k}$, as demonstrated in Fig. 2.1-6(a), where R_k is the pole radius and ϕ_k the pole angle. This gives rise to a factor in the denominator of $H(z)$, of the form

$$(1 - 2R_k \cos\phi_k z^{-1} + R_k^2 z^{-2}). \tag{2.1.18}$$

The same is true of zeros. In particular, a complex conjugate pair of zeros on the unit circle [Fig. 2.1-6(b)] can be represented by the factor

$$(1 - 2\cos\omega_k z^{-1} + z^{-2}). \tag{2.1.19}$$

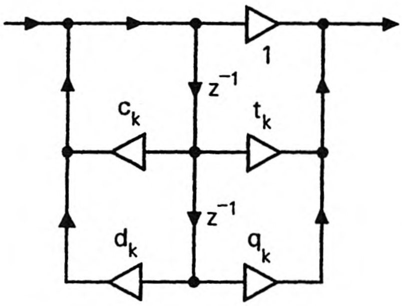

Figure 2.1-5 Implementation of a second order section in (2.1.17), using the direct form structure.

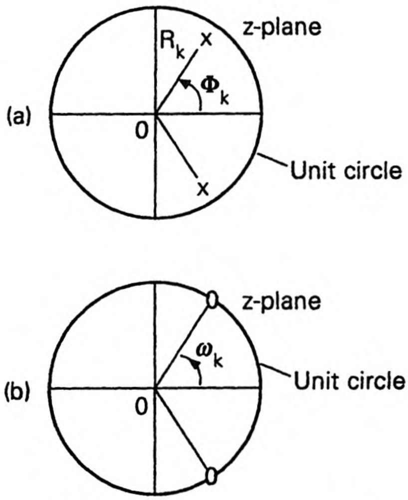

Figure 2.1-6 (a) A complex conjugate pair of poles inside the unit circle, and (b) a complex conjugate pair of zeros on the unit circle.

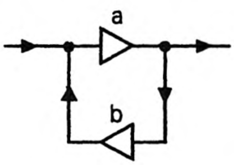

Figure 2.1-7 Example of a delay-free loop.

The number of delays in all the above structures is equal to N, which is the smallest possible. So these structures are minimal. †

Delay-free loops. A loop in which there is no delay is said to be a delay-free loop. Fig. 2.1-7 demonstrates the idea. Discrete-time structures with delay free loops cannot be implemented in practice. Such structures are, therefore, of no practical interest.

† Digital filter structures which use the smallest possible number of delays are said to be minimal in delays, or just *minimal* or *canonical.*

2.1.4 Continuous-Time Systems

Continuous-time functions are denoted as $x_a(t)$, $y_a(t)$ and so on. The subscript a , which stands for "analog", is deleted if the context makes it clear. The Fourier transform of $x_a(t)$, if it exists, is defined as

$$X_a(j\Omega) = \int_{-\infty}^{\infty} x_a(t)e^{-j\Omega t}dt, \tag{2.1.20}$$

and the inverse transform relation is

$$x_a(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X_a(j\Omega)e^{j\Omega t}d\Omega. \tag{2.1.21}$$

Here, the *frequency variable* Ω has the dimension of *radians per second.*

Many of the concepts described earlier carry over to continuous-time systems in an obvious manner. A continuous-time LTI system is characterized by an impulse response $h_a(t)$ and transfer function $H_a(s)$. The transfer function is the Laplace's transform of $h_a(t)$, that is,

$$H_a(s) = \int_{-\infty}^{\infty} h_a(t)e^{-st}\, dt$$

One has to specify a region of convergence for this integral in the s-plane [Oppenheim, Willsky, and Young, 1983]. If the region of convergence includes the imaginary axis, then $H_a(j\Omega)$ [the Fourier transform of $h_a(t)$] is defined, and is called the frequency response of the system. The system is causal if, and only if, $h_a(t) = 0$ for $t < 0$. A causal system is (BIBO) stable if, and only if, all the poles of $H_a(s)$ are in the open left half plane (abbreviated LHP) which is the region characterized by $\text{Re}[s] < 0$. In this case the region of convergence includes the closed right half plane, that is, the region $\text{Re}[s] \geq 0$.

Sampling

We say that $x(n)$ is the sampled version of $x_a(t)$ if $x(n) = x_a(nT)$ for some $T > 0$. The quantity T is the sampling period (or sample spacing), and $2\pi/T$ the sampling frequency or sampling rate. Denote the Fourier transforms of $x(n)$ and $x_a(t)$ as $X(e^{j\omega})$ and $X_a(j\Omega)$. It can be shown that

$$X(e^{j\omega}) = \frac{1}{T}\sum_{k=-\infty}^{\infty} X_a\big(j(\Omega - \frac{2\pi k}{T})\big)\Big|_{\Omega=\omega/T} \tag{2.1.22}$$

Thus, $X(e^{j\omega})$ is obtained as follows: (a) duplicate $X_a(j\Omega)$ at uniform intervals separated by $2\pi/T$, (b) add these copies and divide by T, and (c) replace Ω with ω/T. Figure 2.1-8(a) demonstrates this idea. In part (b) we demonstrate the physical dimensions of the frequency axis, by assuming that the sampling period T is one millisecond [i.e., $2\pi/T$ is 2π Kilo radians per second (Kr/s)]. Figure 2.1-8(c) shows the correspondence with the frequency variable ω associate with the sequence $x(n)$.

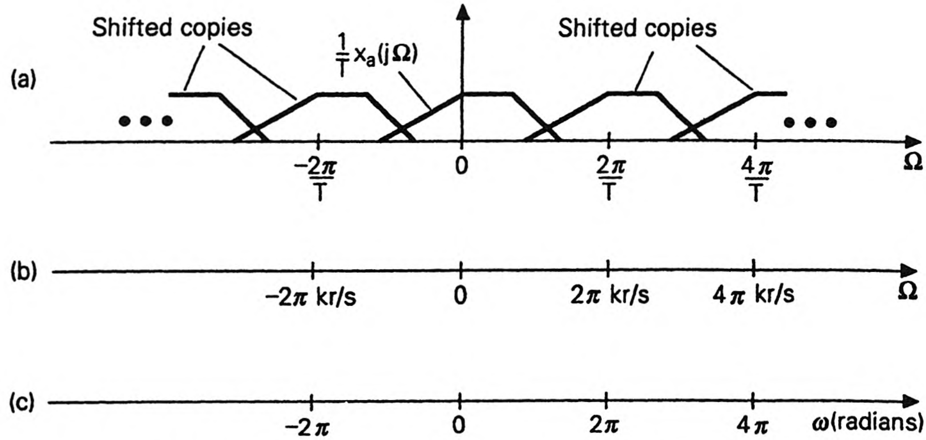

Figure 2.1-8 (a) Fourier transform of a sampled version of $x_a(t)$. (b) Example of frequency dimensions in kiloradians/second, assuming 1 kHz sampling rate, and (c) correspondence with discrete-time frequency variable ω (radians).

Aliasing. If there is no overlap between $X_a(j\Omega)$ and the shifted versions, we can recover $x_a(t)$ from the sampled version $x(n)$ by retaining only one copy. This is accomplished by filtering. If we have the apriori knowledge that the signal is lowpass, then an ideal lowpass filter is used. Otherwise a bandpass filter with appropriate center frequency is required.

The overlap-free condition can be ensured by requiring that $X_a(j\Omega)$ be zero for $|\Omega| \geq \pi/T$. (This is the lowpass case; see Problem 2.15 for other possibilities.) If there is overlap between $X_a(j\Omega)$ and any of its shifted versions, we say that there is *aliasing*.

Bandlimited signals and Nyquist rate. If $X_a(j\Omega)$ is zero for $|\Omega| \geq \sigma$, we say that $x_a(t)$ is σ-bandlimited or σ-BL. We see that if $x_a(t)$ is σ-BL, then we can avoid aliasing by sampling at the rate $\Theta \triangleq 2\sigma$ (*Shannon* or *Nyquist sampling theorem*). This is called the *Nyquist rate* for $x_a(t)$.

Samplers and A/D Converters

We often refer to a continuous-time signal as an "analog" signal, even though there is a fine distinction between these two [Oppenheim and Schafer, 1989]. The amplitude of an analog signal can take continuous values (possibly complex). A digital signal can take values only from a preassigned discrete set (e.g., the values represented by the binary number system). Either of these signals could be continuous-time or discrete-time. In this text, we often use the term "analog signal" to imply both analog and continuous-time signals. Similarly, terms such as "digital signals" and "digital filters" are used to imply "digital" as well as "discrete-time" versions. Whenever a finer distinction is appropriate, it is either mentioned or will be clear from the context.

We often refer to black boxes in figures, with the labels "A/D converters" and "D/A converters" (or just A/D and D/A). These stand for "analog-to-digital" and "digital-to-analog," respectively. In most cases, these black boxes really imply conversion between continuous-time and discrete time, and a notation such as C/D and D/C (as in Oppenheim and Schafer, [1989]) would have been appropriate. We will, however, use A/D and D/A everywhere, and the precise meaning will be clear from the context.

2.2 MULTI-INPUT MULTI-OUTPUT SYSTEMS

Consider a system with r inputs and p outputs, with a transfer function connected from every input to every output. Thus, let $H_{km}(z)$ denote the transfer function from the mth input to the kth output. This is demonstrated in Fig. 2.2-1 for $p = r = 2$. The kth output in response to all the inputs is given by

$$Y_k(z) = \sum_{m=0}^{r-1} H_{km}(z)U_m(z), \quad 0 \leq k \leq p-1. \tag{2.2.1}$$

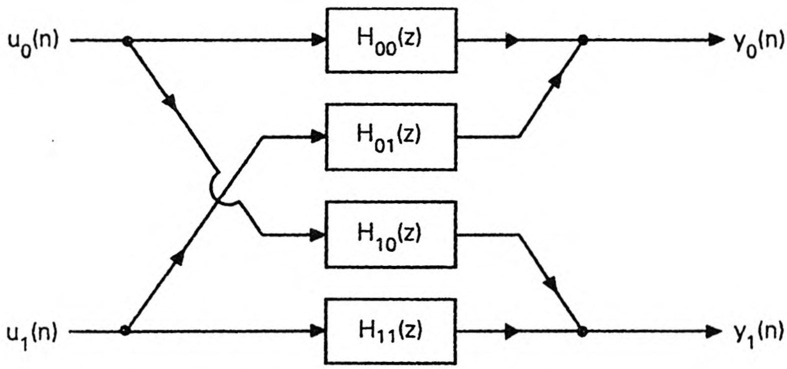

Figure 2.2-1 A two-input two-output system.

The entire system is said to be a multi-input multi-output (MIMO) LTI sytstem, and can be characterized by the set of pr transfer functions $H_{km}(z)$. In order to compactly represent the system, we define the input and output vectors

$$\begin{aligned} \mathbf{u}(n) &= [\,u_0(n) \quad u_1(n) \quad \ldots \quad u_{r-1}(n)\,]^T, \\ \mathbf{y}(n) &= [\,y_0(n) \quad y_1(n) \quad \ldots \quad y_{p-1}(n)\,]^T, \end{aligned} \tag{2.2.2}$$

and their z-transforms

$$\begin{aligned}\mathbf{U}(z) &= [\,U_0(z) \quad U_1(z) \quad \ldots \quad U_{r-1}(z)\,]^T = \sum_{n=-\infty}^{\infty} \mathbf{u}(n)z^{-n},\\ \mathbf{Y}(z) &= [\,Y_0(z) \quad Y_1(z) \quad \ldots \quad Y_{p-1}(z)\,]^T = \sum_{n=-\infty}^{\infty} \mathbf{y}(n)z^{-n}.\end{aligned} \tag{2.2.3}$$

Then the system can be described as

$$\mathbf{Y}(z) = \mathbf{H}(z)\mathbf{U}(z), \tag{2.2.4}$$

where

$$\mathbf{H}(z) = [H_{km}(z)]. \tag{2.2.5}$$

Note the use of bold letters to indicate matrices and vectors. The $p \times r$ matrix $\mathbf{H}(z)$ is called the *transfer matrix* of the system. We will use the terms "r-input p-output system" and "$p \times r$ system" interchangeably. A system with $p = r = 1$ is said to be a single-input single-output (SISO) system, or a *scalar* system.

Fig. 2.2-2 indicates two ways of representing the system. The input and output lines are indicated either by heavy arrows, or by double-line arrows according to convenience. The double lines do *not* imply that there are only two inputs or two outputs.

Figure 2.2-2 Two ways to represent a multi-input multi-output LTI system.

The impulse Response Matrix

Let $h_{km}(n)$ denote the impulse response of the transfer function $H_{km}(z)$. Define the $p \times r$ matrix of impulse response sequences as

$$\mathbf{h}(n) = [h_{km}(n)]. \tag{2.2.6}$$

Then, the relation (2.2.4) can be expressed in the time-domain as

$$\mathbf{y}(n) = \sum_{i=-\infty}^{\infty} \mathbf{h}(i)\mathbf{u}(n-i) = \sum_{i=-\infty}^{\infty} \mathbf{h}(n-i)\mathbf{u}(i), \tag{2.2.7}$$

which is the matrix version of the familiar convolution summation (2.1.6). From the definitions of $\mathbf{H}(z)$ and $\mathbf{h}(n)$ it is evident that they are related as

$$\mathbf{H}(z) = \sum_{n=-\infty}^{\infty} \mathbf{h}(n)z^{-n}. \tag{2.2.8}$$

In general, the above infinite summation converges only in certain regions of the z-plane.

The matrix sequence $\mathbf{h}(n)$ is said to be the "impulse response" or "unit-pulse response" of the system $\mathbf{H}(z)$. For example, let

$$\mathbf{H}(z) = \begin{bmatrix} 1+z^{-1} & 2+z^{-2} \\ 1+z^{-1}+z^{-2} & 2 \end{bmatrix}. \tag{2.2.9}$$

This can be rewritten as

$$\mathbf{H}(z) = \underbrace{\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}}_{\mathbf{h}(0)} + z^{-1} \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}}_{\mathbf{h}(1)} + z^{-2} \underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{\mathbf{h}(2)}, \tag{2.2.10}$$

and the sequence $\mathbf{h}(n)$ can be readily identified, as indicated.

Transfer matrices which are row or column vectors. A system with one input and p outputs has a $p \times 1$ transfer matrix, that is, a column vector. A system with r inputs and one output has a $1 \times r$ transfer matrix, that is, a row vector. Figure 2.2-3 shows both types of examples, where

$$\mathbf{H}_1(z) = \begin{bmatrix} 1+z^{-1} \\ 2+3z^{-1} \end{bmatrix}, \quad \mathbf{H}_2(z) = [\, 1+z^{-1} \quad 2+3z^{-1} \,].$$

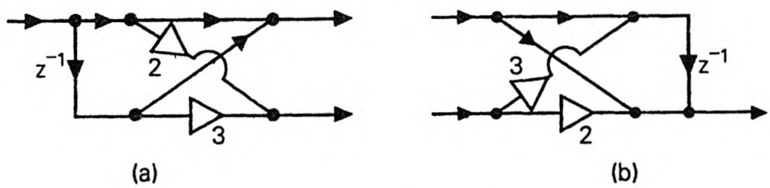

Figure 2.2-3 Examples of transfer matrices (a) a column vector, and (b) a row vector.

Such vector-transfer functions arise in the study of digital filter banks (e.g., Chap. 5).

The frequency response matrix. The discrete-time Fourier transform of a sequence is obtainable from the z-transform by setting $z = e^{j\omega}$. This is given by

$$\mathbf{H}(e^{j\omega}) = \sum_{n=-\infty}^{\infty} \mathbf{h}(n) e^{-j\omega n}, \tag{2.2.11}$$

and is the frequency response matrix of the system.

Stability and causality. The system $\mathbf{H}(z)$ is said to be 'causal' if $\mathbf{h}(n)$ is causal [that is, $\mathbf{h}(n) = \mathbf{0}$ for $n < 0$]. This is equivalent to saying that each $h_{km}(n)$ is causal. We say that α is a pole of $\mathbf{H}(z)$ if it is a pole of some element $H_{km}(z)$. The system $\mathbf{H}(z)$ is stable (in the BIBO sense) if each of the functions $H_{km}(z)$ is stable. So, $\mathbf{H}(z)$ is causal and stable if, and only if, each of the systems $H_{km}(z)$ is causal with all poles strictly inside the unit circle. This is equivalent to the condition that the region of convergence of the summation (2.2.8) includes all points on and outside the unit circle of the z-plane. In particular, this means that the summation in (2.2.11) converges.

No poles at infinity. For a causal rational system $\mathbf{H}(z)$, the region of convergence is everywhere outside a certain circle in the z-plane. In particular, therefore, there are no poles at $z = \infty$, whether the system is stable or not. Since the ROC of $\mathbf{H}(z)$ includes $z = \infty$, the value $\mathbf{H}(\infty)$ can be obtained from the infinite power series $\sum_{n=0}^{\infty} \mathbf{h}(n) z^{-n}$. Thus $\mathbf{H}(\infty) = \mathbf{h}(0)$. In contrast, the value of $\mathbf{H}(0)$ cannot be found by using the infinite power series, since the ROC of the causal power series does not include the origin.

Parseval's relation for vector signals. Let $\mathbf{x}(n)$ and $\mathbf{y}(n)$ be vector sequences with Fourier transforms $\mathbf{X}(e^{j\omega})$ and $\mathbf{Y}(e^{j\omega})$. We then have (Problem 13.1)

$$\sum_{n=-\infty}^{\infty} \mathbf{y}^{\dagger}(n)\mathbf{x}(n) = \frac{1}{2\pi} \int_0^{2\pi} \mathbf{Y}^{\dagger}(e^{j\omega})\mathbf{X}(e^{j\omega})d\omega. \qquad (2.2.12)$$

Degree of a system. The degree or "McMillan degree" of an LTI system $\mathbf{H}(z)$ is defined to be the smallest number of delay elements (z^{-1} elements) required to implement the system. Unlike in the scalar case, the degree of an MIMO system cannot be determined just by inspection of $\mathbf{H}(z)$. In Chap. 13 we will study this topic carefully.

Exponential Inputs Produce Exponential Outputs

For a scalar LTI system $H(z)$, we know that an exponential input a^n produces the output $H(a)a^n$. Now consider a $p \times r$ system $\mathbf{H}(z)$, and apply the input $\mathbf{v}a^n$ where a is an arbitrary scalar and $\mathbf{v}$ an arbitrary vector. Using (2.2.7) we find

$$\mathbf{y}(n) = \sum_{i=-\infty}^{\infty} \mathbf{h}(i)\mathbf{v}a^{n-i} = a^n \sum_{i=-\infty}^{\infty} \mathbf{h}(i)\mathbf{v}a^{-i} = \mathbf{H}(a)\mathbf{v}a^n. \qquad (2.2.13)$$

In other words, an exponential input $\mathbf{v}a^n$ aligned in the direction of the vector $\mathbf{v}$ produces the exponential output $\mathbf{H}(a)\mathbf{v}a^n$, which is aligned in the direction of the vector $\mathbf{H}(a)\mathbf{v}$. This gives us a beautiful 'physical' significance for the transfer matrix $\mathbf{H}(z)$.

Chapter 13 is dedicated to a thorough review of MIMO LTI systems, and is a preparation for some of the deeper results shown in the later sections of Chap. 14 on paraunitary system.

2.3 NOTATIONS

In what follows we summarize the notations used in the text. The reader may wish to glance through this section during first reading (a section on notations can hardly be entertaining!), and then use this primarily as a reference.

2.3.1 Preliminaries

1. The variables Ω and ω are the frequency variables for the continuous and discrete-time cases respectively.
2. $\mathcal{U}(n)$ denotes the unit step sequence and should not be confused with $u(n)$ which sometimes represents the input signal.
3. $\delta(n)$ is the unit-pulse (n is discrete) and $\delta_a(t)$ is the Dirac delta function (t is continuous). Both $\delta(n)$ and $\delta_a(t)$ are often termed as the "impulse functions," and the distinction is usually clear from the context.
4. The terms "inside the unit circle" and "outside the unit circle" are often abbreviated as "the region $|z| < 1$" and "region $|z| > 1$" respectively.
5. Superscript asterik, as in $H^*(z)$, denotes complex conjugation of $H(z)$, whereas subscript asterik, as in $H_*(z)$, means that only the coefficients are conjugated. For example, if $H(z) = a + bz^{-1}$ then, $H_*(z) = a^* + b^*z^{-1}$.

2.3.2 Polynomials

A polynomial in x has the form $\sum_{n=0}^{N} a_n x^n$, that is, has only nonnegative powers of x. Here N is a finite integer. If $a_N \neq 0$, N is said to be its order. Usually we do not use the word "degree," which is reserved for the number of delays required to implement a causal LTI system.

Let $H(z) = \sum_{n=0}^{N} h(n)z^{-n}$. This is a polynomial in z^{-1}, and represents a causal FIR filter. It is common to attach various adjectives to $H(z)$ depending upon its zero locations. Here are some commonly used ones.

1. $H(z)$ is strictly minimum-phase (or strictly Hurwitz, abbreviated SH) if all zeros are inside the unit circle.
2. $H(z)$ is strictly maximum-phase if all zeros are outside the unit circle.
3. $H(z)$ is minimum-phase if all the zeros satisfy $|z_k| \leq 1$ (i.e., are inside or on the unit circle) .
4. $H(z)$ is maximum-phase if all the zeros satisfy $|z_k| \geq 1$.
5. $H(z)$ is a mixed phase polynomial if none of the above holds.

2.3.3 The "Tilde" Notation (Paraconjugation)

The notation $\widetilde{H}(z)$ plays a crucial role in our discussion. This is defined such that, on the unit circle, $\widetilde{H}(z) = [H(z)]^*$ (that is, complex conjugation). Examples:

Let $H(z) = 1 + 2z^{-1}$, then $\widetilde{H}(z) = 1 + 2z$.

Let $H(z) = (a + bz^{-1})/(c + dz^{-1})$, then $\widetilde{H}(z) = (a^* + b^* z)/(c^* + d^* z)$.

More generally, we define $\widetilde{H}(z)$ for a rational function $H(z)$ as follows: first conjugate the coefficients, and then replace z with z^{-1}. Using the subscript asterik notation defined earlier, we see that

$$\widetilde{H}(z) = H_*(z^{-1}).$$

As an application, if $H(z) = \sum_{n=0}^{N} h(n) z^{-n}$, then

$$z^{-N}\widetilde{H}(z) = h^*(N) + h^*(N-1)z^{-1} + \ldots + h^*(0)z^{-N},$$

that is, the coefficients are time-reversed and conjugated. A number of points about the "tilde notation" are worth noting.

1. The function $H_*(z^*)$ *also* reduces to $H^*(z)$ on the unit circle (in fact for any z), but it is not a rational function of z [unlike $H(z)$]. The function $\widetilde{H}(z)$, on the other hand, continues to be rational in z (hence analytic), and it is mathematically more convenient to use. $\widetilde{H}(z)$ is also called the *paraconjugate* of $H(z)$ and can be regarded as an analytic extension of unit-circle conjugation.
2. Given a function $H(z)$, the quantity $\widetilde{H}(z)H(z)$, evaluated on the unit circle, is the magnitude squared response $|H(e^{j\omega})|^2$.
3. We can write $\widetilde{H}(z) = H^*(1/z^*)$ for any z. This has been used, for example, in Oppenheim and Schafer [1989] when discussing magnitude squared responses of digital filters.
4. If $H(z) = 0$ for $z = \alpha$, then $\widetilde{H}(z) = 0$ for $z = 1/\alpha^*$ (the reciprocal conjugate point).
5. If $H(z)$ has (strictly) minimum phase, then $\widetilde{H}(z)$ has (strictly) maximum phase, and conversely.
6. If $H(z) = H_1(z)H_2(z)$, then $\widetilde{H}(z) = \widetilde{H}_1(z)\widetilde{H}_2(z)$. If $H(z) = H_1(z) + H_2(z)$, then $\widetilde{H}(z) = \widetilde{H}_1(z) + \widetilde{H}_2(z)$.

2.3.4 Matrices and Matrix Functions

Bold faced letters such as $\mathbf{A}, \mathbf{v}$ denote matrices and vectors. See Appendix A for a brief review of matrices, and matrix operations such as transpose, transpose conjugate, and so on. We often encounter matrix functions $\mathbf{H}(z)$. These are matrices in which each element is a rational function (e.g., polynomial) in z or z^{-1}. Once again the "tilde" notation plays a major role. Here is a summary of key matrix notations.

1. $\mathbf{A}^T$ denotes the transpose of $\mathbf{A}$, and $\mathbf{H}^T(z)$ stands for $[\mathbf{H}(z)]^T$.
2. $\mathbf{A}^\dagger$ denotes transpose-conjugate of $\mathbf{A}$, and $\mathbf{H}^\dagger(e^{j\omega})$ denotes $[\mathbf{H}(e^{j\omega})]^\dagger$.

3. $\mathbf{H}_*(z)$ denotes conjugation of *coefficients* without changing z (see example below).
4. $\widetilde{\mathbf{H}}(z)$ denotes $\mathbf{H}_*^T(z^{-1})$.

As an example, let $\mathbf{H}(z) = \mathbf{h}(0) + \mathbf{h}(1)z^{-1}$. Then

$$
\begin{aligned}
\mathbf{H}^T(z) &= \mathbf{h}^T(0) + \mathbf{h}^T(1)z^{-1}, \\
\mathbf{H}_*(z) &= \mathbf{h}^*(0) + \mathbf{h}^*(1)z^{-1}, \\
\widetilde{\mathbf{H}}(z) &= \mathbf{h}^\dagger(0) + \mathbf{h}^\dagger(1)z, \\
\mathbf{H}(e^{j\omega}) &= \mathbf{h}(0) + \mathbf{h}(1)e^{-j\omega}, \\
\mathbf{H}^\dagger(e^{j\omega}) &= \mathbf{h}^\dagger(0) + \mathbf{h}^\dagger(1)e^{j\omega}.
\end{aligned}
\tag{2.3.1}
$$

The matrix $\widetilde{\mathbf{H}}(z)$ is said to be the paraconjugate of $\mathbf{H}(z)$. For rational $\mathbf{H}(z)$ it continues to be rational. Notice that $\widetilde{\mathbf{H}}(z) = \mathbf{H}^\dagger(z)$ for $z = e^{j\omega}$, that is, paraconjugation and transpose conjugation are identical on the unit circle. In general, one can write $\widetilde{\mathbf{H}}(z) = \mathbf{H}^\dagger(1/z^*)$ for any z.

The word 'scalar' corresponds to a matrix with $p = r = 1$. Thus a 'scalar system' is a single-input single-output (SISO) system. All the notations introduced above apply to the scalar case; just remember that transposition leaves the scalar quantity unchanged.

2.3.5 Notations for FIR Functions

An FIR function is any function of the form

$$G(z) = \sum_{n=n_1}^{n_2} g(n)z^{-n}, \tag{2.3.2}$$

where $-\infty < n_1 \leq n_2 < \infty$. A causal FIR function (or a polynomial in z^{-1}) is of the form $H(z) = \sum_{n=0}^{N} h(n)z^{-n}$. Several types of causal FIR filters can be distinguished:

1. *Hermitian and skew-Hermitian polynomials.* We say that $H(z)$ is Hermitian [or $h(n)$ is Hermitian] if $h(n) = h^*(N-n)$ for all n, and skew-Hermitian if $h(n) = -h^*(N-n)$. In terms of $H(z)$ this means $H(z) = z^{-N}\widetilde{H}(z)$ (Hermitian) and $H(z) = -z^{-N}\widetilde{H}(z)$ (skew-Hermitian). For example, $1 + 2j + (1-2j)z^{-1}$ is Hermitian whereas $2 + j - (2-j)z^{-1}$ is skew-Hermitian.
2. *Generalized-Hermitian polynomial.* We say that $H(z)$ is generalized Hermitian if $h(n) = ch^*(N-n)$, [i.e., $H(z) = cz^{-N}\widetilde{H}(z)$] for some c with $|c| = 1$. Examples: (a) $a + 2z^{-1} + a^*z^{-2}$, (b) $a + jz^{-1} + z^{-2} + ja^*z^{-3}$. Evidently, Hermitian and skew-Hermitian polynomials are special cases of this.

3. *Symmetric and antisymmetric polynomials.* $H(z)$ [or $h(n)$] is said to be symmetric if $h(n) = h(N-n)$ and antisymmetric if $h(n) = -h(N-n)$; [this definition is particularly useful if $h(n)$ is real]. These are equivalent, respectively, to $H(z) = z^{-N}H(z^{-1})$ and $H(z) = -z^{-N}H(z^{-1})$. Examples: $1+2z^{-1}+z^{-2}$ is symmetric; $1+2z^{-1}-2z^{-2}-z^{-3}$ is antisymmetric.

4. *Hermitain image and mirror image.* Let $A(z)$ and $B(z)$ be two polynomials in z^{-1} with order N. We say that $B(z)$ is the generalized Hermitian image of $A(z)$ if $B(z) = cz^{-N}\widetilde{A}(z)$ for some c with $|c| = 1$ (Hermitian image if $c = 1$, skew-Hermitian image if $c = -1$). Also $B(z)$ is the *mirror image* of $A(z)$ if $B(z) = z^{-N}A(z^{-1})$. Here are some examples, with $N = 1$.

   a) $1+jz^{-1}$ and $j+z^{-1}$ (mirror images)
   b) $1+jz^{-1}$ and $-j+z^{-1}$ (Hermitian images)
   c) $1+jz^{-1}$ and $1+jz^{-1}$ (Hermitian images; why?)
   d) $1+jz^{-1}$ and $j-z^{-1}$ (Hermitian images)
   e) $1+2z^{-1}$ and $2+z^{-1}$ (mirror *and* Hermitian images)

The above terminology can also be used for the more general form (2.3.2). For example, we say $G(z)$ is symmetric if $H(z)$ defined as $z^{-n_1}G(z) = \sum_{n=0}^{N} h(n)z^{-n}$ is symmetric. Thus, $z+2+z^{-1}$ is symmetric, and so is the transfer function $z^{-1}+5z^{-2}+z^{-3}$.

2.3.6 Miscellaneous Mathematical Symbols

The following symbols are sometimes employed for economy: $\exists$ (there exists); $\iff$ (if and only if); $\forall$ (for all); $\approx$ (approximately equal to); $\Rightarrow$ (implies); $\in$ (belongs to); $\stackrel{\Delta}{=}$ (defined as).

2.4 DISCRETE-TIME FILTERS (DIGITAL FILTERS)

We use the term 'digital filter' for discrete-time filters even though digitization (quantization) effects will be considered only in Chap. 9. A digital filter, then, is an LTI system with rational transfer function as in (2.1.10). The quantity $H(e^{j\omega})$ is called the *frequency response.* Its meaning is clear from (2.1.7) which yields

$$Y(e^{j\omega}) = H(e^{j\omega})U(e^{j\omega}). \tag{2.4.1}$$

From the time domain viewpoint, if we apply an input with frequency ω_0, that is, $u(n) = e^{j\omega_0 n}$, then the output is $y(n) = H(e^{j\omega_0})e^{j\omega_0 n}$. This follows because $e^{j\omega_0 n}$ is an eigenfunction of the system. $H(e^{j\omega_0})$ is the weighting function (or eigenvalue) or simply "the gain of the system" at frequency ω_0.

Transmission Zeros

If $H(e^{j\omega_0}) = 0$ then the frequency ω_0 is rejected by the filter. So the filter offers infinite attenuation at this frequency. We say that ω_0 is a transmission zero of $H(z)$.

Transmission zeros come from zeros of $H(z)$ on the unit circle. Let $H(z) = A(z)/B(z)$, with $A(z)$ and $B(z)$ as in (2.1.11). It is clear that ω_0 is a transmission zero if, and only if, $(1 - e^{j\omega_0}z^{-1})$ is a factor of $A(z)$. If all zeros of $H(z)$ are on the unit circle, then all the factors of the numerator $A(z)$ have the form $(1 - e^{j\omega_k}z^{-1})$.

For the real coefficient case, each factor $(1 - e^{j\omega_k}z^{-1})$ is paired with $(1 - e^{-j\omega_k}z^{-1})$ unless $\omega_k = 0$ or π. The factor of $A(z)$ which represents the complex conjugate pair of zeros is, therefore, $1 - 2\cos\omega_k z^{-1} + z^{-2}$. This is a symmetric polynomial so that the product of such factors is symmetric. Transmission zeros at $\omega = 0$ and π give rise to the factors $(1 - z^{-1})$ (antisymmetric) and $(1 + z^{-1})$ (symmetric). Thus, for a real filter with all zeros on the unit circle, the numerator $A(z)$ is either symmetric or antisymmetric. That is,

$$a_n = \begin{cases} a_{N-n} & \text{(even number of zeros at } \omega = 0) \\ -a_{N-n} & \text{(odd number of zeros at } \omega = 0). \end{cases} \tag{2.4.2}$$

For the more general case of complex systems, if all zeros are on the unit circle then $A(z)$ is generalized-Hermitian (Problem 2.3). Notice, however, that these conditions on $A(z)$ are not sufficient to ensure that all zeros are on the unit circle.

2.4.1 Magnitude Response and Phase Response

The frequency response which in general is a complex quantity, can be expressed as

$$H(e^{j\omega}) = |H(e^{j\omega})|e^{j\phi(\omega)}. \tag{2.4.3}$$

The real-valued quantities $|H(e^{j\omega})|$ and $\phi(\omega)$ are, respectively, called the magnitude response and the phase response of the filter. The quantity $\tau(\omega) = -d\phi(\omega)/d\omega$ is said to be the group delay of the system $H(z)$.

Depending on the nature of $|H(e^{j\omega})|$, filters are typically classified as lowpass, bandpass and so on. Figure 2.4-1 demonstrates this for real coefficient filters (see below).

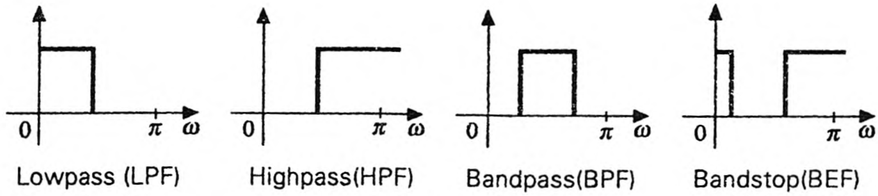

Figure 2.4-1 Various types of magnitude responses for real-coefficient filters.

Real Coefficient Filters

For real $h(n)$, the magnitude $|H(e^{j\omega})|$ is an even function of ω whereas the phase response $\phi(\omega)$ is an odd function. For example, let $H(z) = 1-az^{-1}$ with real a. Then, $H(e^{j\omega}) = 1 - ae^{-j\omega}$, and

$$|H(e^{j\omega})| = (1 - 2a\cos\omega + a^2)^{0.5}, \quad \phi(\omega) = \tan^{-1}\Big(\frac{a\sin\omega}{1 - a\cos\omega}\Big),$$

so that $|H(e^{j\omega})|$ is even, and $\phi(\omega)$ odd indeed. So the response needs to be shown only for the region $0 \le \omega \le \pi$. If $H(z)$ has real coefficients, an input $\cos(\omega_0 n)$ produces the output $y(n) = |H(e^{j\omega_0})|\cos(\omega_0 n + \phi(\omega_0))$. †

Unwrapped and Wrapped Phase

Consider the filter $H(z) = (\frac{1+z^{-1}}{2})^6$. The frequency response is given by $H(e^{j\omega}) = e^{-j3\omega}\cos^6(\omega/2)$. So, the phase response is $\phi_u(\omega) = -3\omega$ and varies from 0 to -6π as ω changes from 0 to 2π. If we replace $\phi_u(\omega)$ with its principal value $\phi_w(\omega) = \phi_u(\omega)$ mod 2π, this does not change the value of $H(e^{j\omega})$, as seen from (2.4.3). The quantities $\phi_u(\omega)$ and $\phi_w(\omega)$ are called the *unwrapped* and *wrapped* phase responses, respectively (this also explains the introduction of subscripts for this discussion).

The unwrapped phase $\phi_u(\omega)$ can have any value whereas the wrapped phase $\phi_w(\omega)$ is always within a range of length 2π, for example, $-2\pi < \phi_w(\omega) \le 0$ or $-\pi < \phi_w(\omega) \le \pi$. The unwrapped phase $\phi_u(\omega)$ is related to the group delay $\tau(\omega)$ according to the integral

$$\phi_u(\omega) = -\int_0^{\omega} \tau(\theta)\,d\theta + \phi_u(0).$$

Most computer programs that evaluate the phase response actually return the wrapped phase $\phi_w(\omega)$. There exist good algorithms to obtain the unwrapped phase from the wrapped phase [Tribolet, 1977], [Oppenheim and Schafer, 1989].

If the distinction between the wrapped and unwrapped phases is not necessary (as in the majority of our discussions), the subscript will be omitted. The distinction is sometimes essential, for example, when dealing with the so-called *complex cepstrum.* We will have occasion to describe this in Appendix D, where we study the spectral factorization problem.

Decibel (dB) Plots

The plot of $20\log_{10}|H(e^{j\omega})|$ as a function of ω is particularly useful in revealing the stopband details of the response. This is often referred to as the dB plot of the (magnitude) response. Figure 2.4-2 demonstrates this for

† Notice that, in general, signals of the form $\cos(\omega_0 n)$ are *not* eigenfunctions of LTI systems.

a lowpass filter. It is helpful to remember that $|H(e^{j\omega})| = 10^{-k}$ implies a level of $-20k$ dB on this plot. For example, $|H(e^{j\omega_1})| = 0.01$ implies that the filter provides 40 dB attenuation at ω_1 (see Sec. 3.1 later).

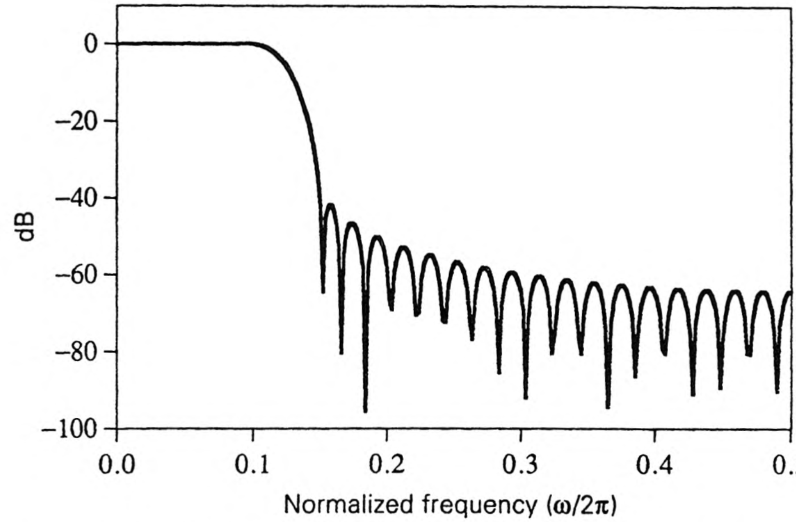

Figure 2.4-2 Demonstration of decibel (dB) plot of the magnitude response.

2.4.2 Linear Phase Filters

Strictly speaking, a digital filter is said to have linear phase if the phase response $\phi(\omega)$ is linear in ω (i.e., of the form $\alpha\omega$ where α is constant). In engineering practice, a less stringent definition is used, which we employ in this text. According to this definition, $H(z)$ has linear phase if

$$H(e^{j\omega}) = ce^{-jK\omega}H_R(\omega), \tag{2.4.4}$$

where c is a possibly complex constant, K is real, and $H_R(\omega)$ is a real valued function of ω. Note that $H_R(\omega)$ does not necessarily have period 2π [for example, try $H(z) = 1 + z^{-1}$], and we have to avoid the notation $H_R(e^{j\omega})$.

The quantity $H_R(\omega)$ is called the *amplitude response* or *zero-phase response.* Linear phase filters for which c is real and $K = 0$ are called *zero-phase filters.* For a zero-phase filter $H(e^{j\omega})$ itself is real, but it can become negative (this typically happens in the stopband).

In a region where $H_R(\omega)$ has fixed sign, the group delay of a linear-phase filter is constant, that is, $\tau(\omega) = K$. For filters with nonlinear phase, it is common practice to plot the group delay $\tau(\omega)$ to show the nonlinearity. The degree to which $\tau(\omega)$ is a nonconstant reveals the degree of phase nonlinearity.

Four Types of Real Coefficient Linear Phase Filters

Let $H(z) = \sum_{n=0}^{N} h(n)z^{-n}$, with real $h(n)$. It is well known that the response has the form (2.4.4) if $h(n)$ is symmetric or antisymmetric. Depending on whether N is even or odd, and whether $h(n)$ is symmetric or antisymmetric, we obtain four types of real coefficient linear phase filters. These are summarized in Table 2.4.1.

Notice that some of these filters have transmission zeros at $\omega = 0$ and/or π, so that they cannot be used for certain applications. For example, Types 3 and 4 (antisymmetric cases) cannot be used for lowpass filter design, and Type 2 cannot be used for highpass design. If $H(z)$ is Type 1 or 3, then the filter $G(z) = z^M H(z)$ (where $M = N/2$) is a zero-phase filter. In this text whenever we refer to linear phase filters of Type 1–4, it is implicit that the coefficients are real so that the properties in Table 2.4.1 hold.

TABLE 2.4.1 Four types of real coefficient linear phase FIR filters. Here $H(z) = \sum_{n=0}^{N} h(n)z^{-n}$, with $h(n)$ real

Type	1	2	3	4
Symmetry	$h(n) = h(N-n)$	$h(n) = h(N-n)$	$h(n) = -h(N-n)$	$h(n) = -h(N-n)$
Parity of N	N even	N odd	N even	N odd
Expression for frequency response $H(e^{j\omega})$	$e^{-j\omega N/2} H_R(\omega)$	$e^{-j\omega N/2} H_R(\omega)$	$je^{-j\omega N/2} H_R(\omega)$	$je^{-j\omega N/2} H_R(\omega)$
Amplitude response or zero-phase response $H_R(\omega)$	$\sum_{n=0}^{M} b_n \cos(\omega n)$ $M = N/2$	$\cos\frac{\omega}{2} \sum_{n=0}^{M} b_n \cos(\omega n)$ $M = (N-1)/2$	$\sin\omega \sum_{n=0}^{M} b_n \cos(\omega n)$ $M = (N-2)/2$	$\sin\frac{\omega}{2} \sum_{n=0}^{M} b_n \cos(\omega n)$ $M = (N-1)/2$
Special features		Zero at $\omega = \pi$	Zero at $\omega = 0$ and π	Zero at $\omega = 0$
Can be used for	Any type of bandpass response (LPF, HPF, etc.)	Any bandpass response except Highpass	Differentiators and and Hilbert transformers†	Differentiators, Hilbert transformers, and high pass filters

†See Rabiner and Gold, 1975

For Type 1 filters, $H_R(\omega)$ does have period 2π, whereas for Type 2 filters the period is 4π. (This can be deduced from Table 2.4.1.) Notice that $H_R(\omega)$ changes sign at $\omega = 0$ in some cases. Note also that $[H_R(\omega)]^2$ has period 2π in all cases.

Efficient structures for linear phase filters. Because of the property $h(n) = \pm h(N-n)$, Type 1–4 linear phase FIR filters can be implemented with only about $(N+1)/2$ multipliers. For example, Fig. 2.4-3 shows how we can implement a fifth order (Type 2) filter with only 3 multipliers.

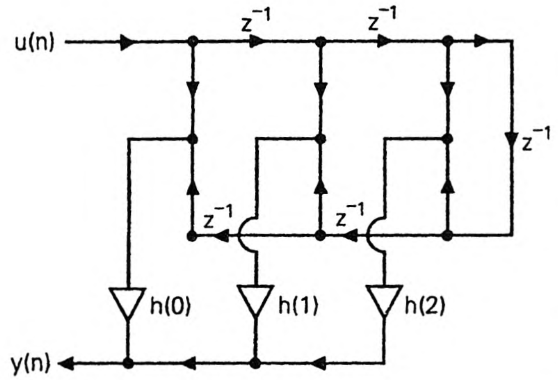

Figure 2.4-3 Efficient direct form implementation of a linear phase filter.

Advantage of Linear Phase Property

Suppose we have a signal $u(n)$ whose energy is dominantly in the region $|\omega| < \sigma$. Suppose this is passed through a real lowpass filter with passband edge at σ. (This is often done to attenuate the out-of-band noise.) The output signal is

$$Y(e^{j\omega}) = H(e^{j\omega})U(e^{j\omega}). \tag{2.4.5}$$

If we assume that the filter $H(z)$ is a 'good' lowpass filter, then $|H(e^{j\omega})| \approx 1$ in the passband. Moreover, both $|Y(e^{j\omega})|$ and $|U(e^{j\omega})|$ are very small in the stopband. So we have

$$|Y(e^{j\omega})| \approx |U(e^{j\omega})|, \tag{2.4.6}$$

for all ω. The approximate nature of this relation is due to the facts that (a) the filter is not ideal, and (b) $u(n)$ is not perfectly bandlimited. Nevertheless, (2.4.6) implies that the output signal $y(n)$ tends to resemble $u(n)$ provided that there is no phase distortion.

The phase distortion, in turn, is eliminated if $U(e^{j\omega})$ and $Y(e^{j\omega})$ have same phase (except for a linear offset term). This can be satisfied if $H(z)$ has linear phase. In this case (2.4.4) holds so that (2.4.5) can be replaced with

$$Y(e^{j\omega}) \approx ce^{-j\omega K}U(e^{j\omega}). \tag{2.4.7}$$

For the case where c is real and K an integer, this implies $y(n) \approx cu(n-K)$, which is a (scaled) and delayed version of $u(n)$.

Summarizing, if the input has energy confined to the passband of the filter then the output signal is approximately equal to (a scaled and shifted version of) this input provided the filter has linear phase and 'good' passband and stopband. (The above discussion assumes K is an integer. But this is not always the case, e.g., when the filter order is odd.)

If the filter has nonlinear phase, then we still have (2.4.6), but due to phase distortion, the time domain relation $y(n) \approx cu(n-K)$ does not hold. Whether this distortion is acceptable or not depends largely on applications. For example, in speech processing a certain degree of phase distortion can be tolerated, but in image processing, phase distortion is often disastrous [Lim, 1990].

Most General Conditions for the Linear Phase Property

Let $H(z) = \sum_{n=0}^{N} h(n)z^{-n}$ with $h(0) \neq 0$ and $h(N) \neq 0$. Then it is a linear phase filter if and only if the impulse response is generalized-Hermitian, that is, satisfies the condition

$$h(n) = dh^*(N-n), \tag{2.4.8}$$

for some d with $|d| = 1$ (Problem 2.12). The following points are worth noting:

1. For the real coefficient case, the above condition reduces to $h(n) = \pm h(N-n)$. So, the impulse response has to be symmetric or antisymmetric.
2. In the transform domain, (2.4.8) is equivalent to the condition

$$H(z) = dz^{-N}\widetilde{H}(z). \tag{2.4.9}$$

3. *Zero-locations of linear-phase filters.* One consequence of (2.4.9) is that, if z_k is a zero then so is $1/z_k^*$. So, the zeros of a linear phase FIR filter $H(z)$ occur in reciprocal conjugate pairs. This also explains why it is not possible to design causal stable IIR linear phase filters (Problem 2.11 requests a more rigorous discussion.)

2.4.3 Analytic Continuation

Let $H_0(z)$ and $H_1(z)$ be two transfer functions such that $H_0(e^{j\omega}) = H_1(e^{j\omega})$ for all ω. This implies that their impulse responses are identical, that is, $h_0(n) = h_1(n)$, and, therefore, that the transfer functions are identical. Thus, two filters with identical frequency responses must have identical transfer functions. In other words, if

$$H_0(z) = H_1(z), \qquad z = e^{j\omega}, \tag{2.4.10}$$

then

$$H_0(z) = H_1(z), \qquad \text{for all } z. \tag{2.4.11}$$

This is called the analytic continuation property. If $H_0(z)$ and $H_1(z)$ in the above discussion are replaced with $p \times r$ transfer matrices $\mathbf{H}_0(z)$ and $\mathbf{H}_1(z)$, then the analytic continuation property still holds.

The above mentioned property might appear to be trivial but it holds because all practical transfer functions are rational (hence analytic), and the unit circle is in the region of analyticity. If $H_0(z)$ and $H_1(z)$ were arbitrary (nonanalytic) functions, then (2.4.10) would not imply (2.4.11). Consider, for example, a function $P(z)$, defined to be zero in an annulus around the unit-circle, but unity otherwise (Fig. 2.4-4). This satisfies $P(e^{j\omega}) = 0$, but $P(z)$ is not identically zero for all z.

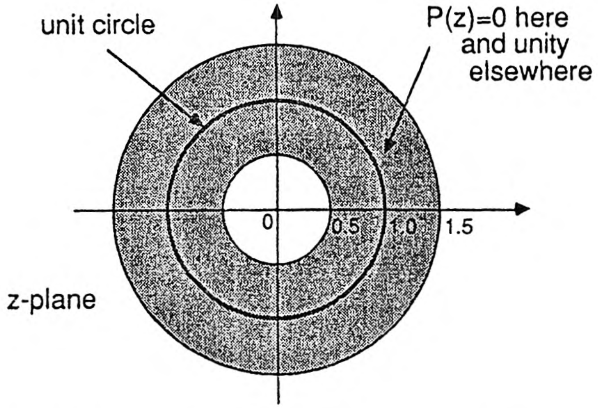

Figure 2.4-4 A function $P(z)$, defined to be zero everywhere on the unit circle, but not identically zero for all z.

PROBLEMS

2.1. In Sec. 2.1.2 we stated that exponential sequences are eigenfunctions of LTI systems. Conversely, suppose $s(n)$ is an eigenfunction of a rational IIR transfer function $H(z)$ (with order > 0 to avoid trivial answers). Does this necessarily mean that $s(n)$ is an exponential? Justify (that is, prove if *yes;* give counter example if *no*).

2.2. Suppose we apply the truncated exponential $a^n\mathcal{U}(-n)$ to a causal stable LTI system $H(z)$. Assume $|a| > 1$ so the input does not blow up for $n \to -\infty$. (a) What is the output $y(n)$ for $n \leq 0$? (b) Suppose $y(n)$ is zero for $n > 0$, i.e., the output becomes zero as soon as the input becomes zero. This leads us to suspect that the system might be *memoryless,* [i.e., $H(z) =$ constant]. This is indeed true. Prove this.

2.3. Let $H(z) = \sum_{n=0}^{N} h(n)z^{-n}$ be a transfer function with all zeros on unit circle. Show that $h(n)$ is generalized-Hermitian. [In particular if $h(n)$ is real, this reduces to the fact that $h(n)$ is symmetric or antisymmetric.]

2.4. Let $H(z) = \sum_{n=0}^{\infty} h(n)z^{-n}, h(0) \neq 0$, be a rational transfer function representing a causal, stable LTI system. So, $H(z) = A(z)/B(z)$ where $A(z)$ and $B(z)$ are relatively prime polynomials in z^{-1}. We shall now develop a time domain interpretation of "poles", which is more appealing than the definition which says that $H(z)$ "blows up" at a pole.

a) Suppose $p \neq 0$ is a pole of $H(z)$. Show that there exists a causal finite length input $x(n)$ and a finite integer L such that the output $y(n)$ has the form p^n for $n > L$.

b) Conversely suppose there exists a causal finite length input $x(n)$ and a finite integer L such that the output $y(n)$ has the form p^n for $n > L$. Show that p is a pole of $H(z)$.

2.5. We know that if z_k is a zero of $H(z)$ then the output in response to the input z_k^n is zero for all n. This input is noncausal (and doubly infinitely long). In this problem we develop another engineering insight for the meaning of a zero of $H(z)$. This is based on causal inputs, and might be more appealing. Assume $H(z)$ is in irreducible rational form as in (2.1.10). Also, assume that it is causal so that it can be implemented with the difference equation (2.1.14) (with $b_0 = 1$). Assume $a_0 \neq 0$, and $b_N \neq 0$.

a) Consider the first order case ($N = 1$). We know the system has a zero at $z_1 = -a_1/a_0$. Suppose, we apply an input of the form $z_1^n\mathcal{U}(n)$ where $\mathcal{U}(n)$ is the unit step. Find an initial value $y(-1)$ such that $y(n) = 0$ for all $n \geq 0$.

b) More generally, for the Nth order system suppose z_1 is a zero and we apply the causal input $z_1^n\mathcal{U}(n)$. Show how you can find an initial state $y(-1), y(-2), \ldots, y(-N)$ such that $y(n) = 0$ for all $n \geq 0$.

c) Conversely, suppose there exists an initial state $y(-1), y(-2), \ldots, y(-N)$ such that the input $z_1^n\mathcal{U}(n)$ produces zero output for all $n \geq 0$. Show that z_1 is a zero of $H(z)$.

2.6. Let $H(z) = \sum_{n=0}^{\infty} h(n)z^{-n}, h(0) \neq 0$, be a rational transfer function representing a causal, stable LTI system. So, we can write $H(z) = A(z)/B(z)$ where $A(z)$ and $B(z)$ are relatively prime polynomials in z^{-1}.

a) Let z_0 be a *zero* of the system. Then show that there exists a causal finite length sequence $s(n)$ such that the input

$$x(n) = z_0^n \mathcal{U}(n) + s(n) \tag{P2.6}$$

produces a causal, finite-length output.

b) Conversely, let there exist an input of the form in (P2.6) where $s(n)$ is causal and finite-length, such that the output is of finite length. Then show that z_0 is indeed a zero, assuming $z_0 \neq 0$.

Note. This gives the following engineering interpretation of a *zero*: there exists a causal input such that, if you wait for finite time after applying the input, the input will look like an exponential z_0^n whereas the output will become zero and stay zero!

2.7. You are given a black box, which you can imagine to be a computer program. This black box takes an input sequence $x(n)$ and computes each sample of the output $y(n)$ in finite amount of time. You are given the additional information that any exponential input a^n produces an exponential output $H(a)a^n$. Can you conclude that the black box is an LTI system? Justify (i.e., prove if *yes;* give counter example if *no*).

2.8. Let $H(z) = \sum_{n=0}^{N} h(n)z^{-n}$ be a Type 1 linear phase FIR filter. Define a new filter $G(z)$ with $g(n) = h(n)\cos[\omega_0(n-K)]$ where K is an integer. How would you choose K so that $G(z)$ also has linear phase?

2.9. Let $x(n) = \cos(\omega_0 n)$ be the input to a Type 1 linear phase FIR filter $H(z)$. Find an expression for the output $y(n)$ and simplify as best as you can. Can you say that $y(n) = cx(n-K)$ for some c and K? What if the filter were Type 3?

2.10. Let $H(z) = 1/[1 + \sum_{n=1}^{N} b_n z^{-n}]$ represent a causal filter with linear phase. Prove that it is unstable (unless $b_n = 0$ for all n).

2.11. Let $H(z)$ be causal stable with irreducible form $A(z)/B(z)$. Suppose this has linear phase, that is, satisfies (2.4.4). Show then that $H(z)$ is FIR! [This is a generalization of Problem 2.10. It is somewhat subtle because, you have to wonder whether $H(z)$ might have linear phase even if $A(z)$ and $B(z)$ are not, individually, linear phase functions.]

2.12. This is a continuation of (2.4.8) and the paragraph preceding it. We stated that (2.4.8) is necessary and sufficient for linear phase property. Prove this.

2.13. This pertains to stability.

a) Consider a causal IIR filter with transfer function $H(z) = 1/D(z)$ where

$$D(z) = 1 + d_1 z^{-1} + d_2 z^{-2} + \ldots + d_N z^{-N},$$

with d_n real for all n. Show that this is BIBO stable *only if* $D(1) > 0$ and $D(-1) > 0$.

b) As a special case, consider a second order causal IIR filter with transfer function $H(z) = 1/(1 + az^{-1} + bz^{-2})$, with a, b real. Show that this is BIBO stable only if the values of a and b are restricted to be *strictly* inside the triangular area in Fig. P2-13 (called the *stability triangle*).

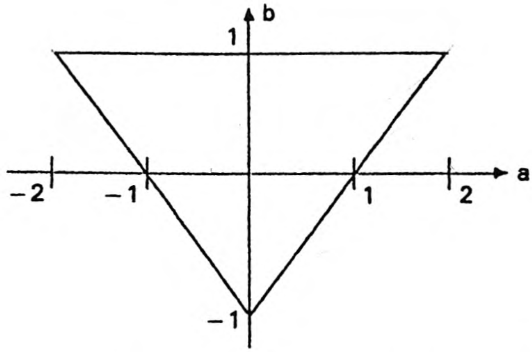

Figure P2-13

c) Finally, show that, in the above second-order case, the converse is also true, that is, if a, b are stritctly inside the triangular region then $H(z)$ is BIBO stable.

2.14 The deterministic cross correlation between two sequences $x(n)$ and $y(n)$ is defined as

$$R_{xy}(k) = \sum_{n=-\infty}^{\infty} x(n)y^*(n-k). \tag{P2.14}$$

The integer k is called the lag variable. Let $S_{xy}(z)$ denote the z-transform of $R_{xy}(k)$. Show that $S_{xy}(z) = X(z)\widetilde{Y}(z)$. [*Note:* A special case of this result arises when $y(n) = x(n)$. The quantity $R_{xx}(k)$ is called the *deterministic autocorrelation* of $x(n)$. Its z-transform is $S_{xx}(z) = X(z)\widetilde{X}(z)$, so that $S_{xx}(e^{j\omega}) = |X(e^{j\omega})|^2$.]

2.15. Give example of a function $x_a(t)$ such that (a) it is not σ-BL, and (b) if it is sampled at the rate $2\pi/T = 2\sigma$, no two terms in (2.1.22) overlap.