Fatemeh Hadaeghi

Principles of Digital Signal Processing

Lecture Notes for BSP, Chapter 3

Master Program Data Engineering

3 Principles of Digital Signal Processing

As were introduced in chapter 1, the discrete-time signals are defined at discrete times and represented as a sequences of numbers.

In this chapter, we consider the fundamental concepts of discrete-time signals and systems for one-dimensional signals. We highlight the class of linear time-invariant discrete-time systems.

The main reference for this chapter is:

[1] Alan, V. Oppenheim, W. Schafer Ronald, and R. B. John. "Discrete-time signal processing." New Jersey, Printice Hall Inc (1989).

3.1 Discrete-Time Signals

Discrete-time signals are represented as sequences of numbers. A sequence of numbers, x, in which the nth number in the sequence is denoted $x[n]$ is formally written as:

$$x = \{x[n]\} \qquad -\infty < \mathrm{n} < \infty \tag{3-1}$$

where n is an integer. In practice, the discrete time signals are obtained by sampling from continuous signals such that the nth number of the sequence is equal to the value of the continuous signal, $x_a(t)$, at time nT_s (i.e, $x[n] = x_a(nT_s)$). The quantity T_s is then called *sampling period* and its reciprocal $F_s = \frac{1}{T_s}$ is the *sampling frequency*.

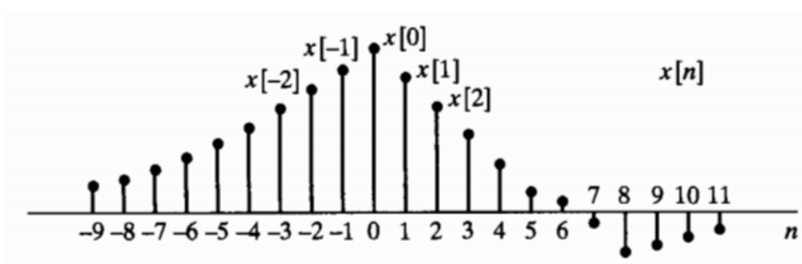


Figure 3-1: Graphical representation of a discrete- time signal [1].

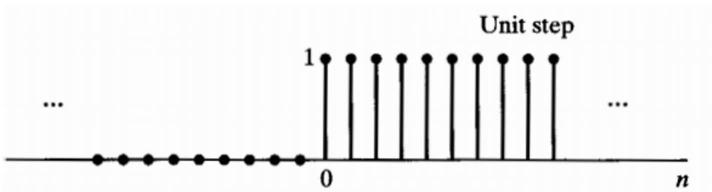
In discussing the theory of discrete-time signals and systems, several basic sequences are of particular importance. These sequences are:

- unit sample sequence (also called discrete- time impulse):

$$\delta[n] = \begin{cases} 0, & n \neq 0 \\ 1, & n = 0 \end{cases} \tag{3-2}$$

- unit step sequence:

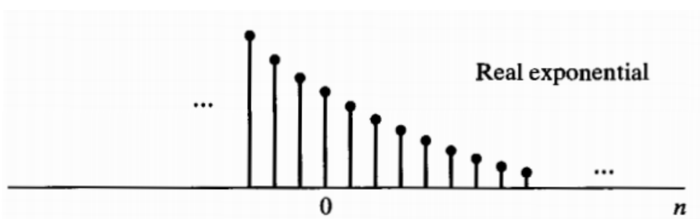
$$u[n] = \begin{cases} 0, & n < 0 \\ 1, & n \geq 0 \end{cases} \tag{3-3}$$



- exponential sequences (with α, a real number):

$$x[n] = A\alpha^n \tag{3-4}$$

$0 < \alpha < 1$



- sinusoidal sequences:

$$x[n] = A\cos(\omega_0 n + \varphi), \text{ for all } n \tag{3-5}$$

- complex exponential sequences:

$$x[n] = A\alpha^n, \alpha = |\alpha|e^{j\omega_0}, A = |A|e^{j\varphi} \tag{3-6}$$

$$\begin{aligned} x[n] &= |A|e^{j\varphi}\,(|\alpha|e^{j\omega_0})^n \\ &= |A||\alpha|^n e^{(j\omega_0 n+\varphi)} \\ &= |A||\alpha|^n \cos(j\omega_0 n + \varphi) + j\,|A||\alpha|^n \sin(j\omega_0 n + \varphi) \end{aligned}$$

A sequence $y[n] = x[n - n_0]$ is said to be a delayed or shifted version of a sequence $x[n]$. Any arbitrary sequence can be represented as a sum of scaled, delayed impulses. For example, the sequence $p[n]$ in Figure 3-2 can be expressed as:

$$p[n] = a_{-3}\delta[n+3] + a_1\delta[n-1] + a_2\delta[n-2] + a_7\delta[n-7]$$

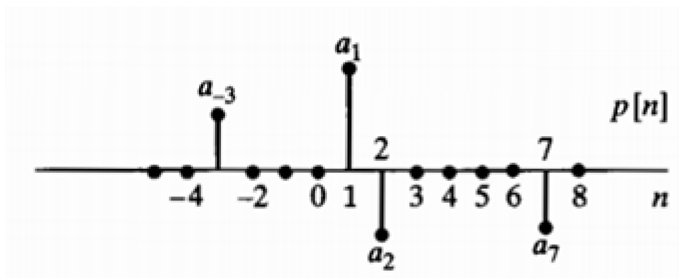


Figure 3-2: Example of a sequence [1].

More generally, any sequence $x[n]$ can be expressed as:

$$x[n] = \sum_{k=-\infty}^{\infty} x[k]\delta[n-k] \tag{3-7}$$

The unit sample is related to impulse by:

$$x[n] = \sum_{k=0}^{\infty} \delta[n-k]. \tag{3-8}$$

Conversely, the impulse sequence is expressed as the first backward difference of the unit step sequence:

$$\delta[n] = u[n] - u[n-1]. \tag{3-9}$$

3.2 Discrete-Time Systems

A discrete-time system is mathematically defined as a transformation that maps an input sequence with values $x[n]$ into an output sequence with values $y[n]$:

$$y[n] = T\{x[n]\}. \tag{3-10}$$

The following important classes of systems are defined by placing constraints on the properties of the transformation T {•}:

- **Memoryless systems:**

  A system is referred to as memoryless if the output $y[n]$ at every value of n depends only on the input $x[n]$ at the same value of n. An example is $y[n] = (x[n])^2$, for each value of n. In contrast, a system which shifts the input samples by n_d samples (i.e. $y[n] = x[n - n_d]$) is not memoryless unless $n_d = 0$.

- **Linear systems**

  The class of linear systems is defined by the principle of superposition. If the system satisfies both the additivity and the homogeneity or scaling properties, it is called a linear system. Mathematically speaking, if $y_1[n]$ and $y_2[n]$ are responses of system to $x_1[n]$ and $x_2[n]$, respectively, then the system is linear if and only if:

$$T\{x_1[n] + x_2[n]\} = T\{x_1[n]\} + T\{x_2[n]\} = y_1[n] + y_2[n] \qquad \text{(3-11a)}$$

  and

$$T\{ax[n]\} = aT\{x[n]\} = ay[n] \qquad \text{(3-11b)}$$

  where a is an arbitrary constant. The first property stated in Eq. (3-11a) is called additivity and the second (i.e. Eq. (3-11b)) is called homogeneity property. A linear system is then a system for which the principle of superposition is satisfied:

$$T\{ax_1[n] + bx_2[n]\} = aT\{x_1[n]\} + b\,T\{x_2[n]\} \qquad \text{(3-12)}$$

  for arbitrary constants a and b. Eq. (3-12) can be generalized to the superposition of many inputs.

  An example of linear system is the "Accumulator system" defined by the following input-output equation:

$$y[n] = \sum_{k=-\infty}^{n} x[k] \qquad \text{(3-13)}$$

  The accumulator system, which at each time step, n calculates the sum of the present and all previous samples, is a linear system. In order to prove this, we must show that it satisfies the superposition principle for all inputs. Therefore, we define, $y_1[n] = \sum_{k=-\infty}^{n} x_1[k]$ and $y_2[n] = \sum_{k=-\infty}^{n} x_2[k]$ as the responses of the system to two inputs $x_1[n]$ and $x_2[n]$, respectively. Now, if we define the input $x_3[n] = ax_1[n] + bx_2[n]$, the superposition requires the output, $y_3[n] = ay_1[n] + by_2[n]$ for all possible choices of a and b. We can show this by starting from the transformation function in Eq. (3-13):

$$\begin{aligned} y_3[n] &= \sum_{k=-\infty}^{n} x_3[k] \\ &= \sum_{k=-\infty}^{n} ax_1[k] + bx_2[k] \\ &= a\sum_{k=-\infty}^{n} x_1[k] + b\sum_{k=-\infty}^{n} x_2[k] \\ &= ay_1[n] + by_2[n]. \end{aligned}$$

Thus the accumulator system satisfies the superposition principle for all possible choices of a and b.

The system defined by $w[n] = log_{10}(|x[n]|)$ can be shown to be nonlinear by giving a counter example, $x_1[n] = 1, x_2[n] = 10$.

- **Time invariant systems**

A time-invariant system is a system for which a time shift or delay of the input sequence causes a corresponding shift in the output sequence. Specifically, suppose that a system transforms the input sequence with values $x[n]$ into the output sequence with values $y[n]$. Then the system is said to be time invariant if, for all n_τ, the input sequence with values $x_1[n] = x[n - n_\tau]$ produces the output sequence with values $y_1[n] = y[n - n_\tau]$. The accumulator system (Eq. (3-13)) can be shown to be time invariant. To prove that, we need to show $y_1[n] = y[n - n_\tau]$.

$$y[n - n_\tau] = \sum_{k=-\infty}^{n-n_\tau} x[k]$$

$$and$$

$$y_1[n] = \sum_{k=-\infty}^{n} x_1[k] = \sum_{k=-\infty}^{n} x[k - n_\tau]$$

Substituting the changes of variables $k_1 = k - n_\tau$ into the second summation gives:

$$y_1[n] = \sum_{k=-\infty}^{n-n_\tau} x[k_1] = y[n - n_\tau]$$

Thus the accumulator is a time-invariant system.

Following the same procedure, it is easy to show the compressor system defined by $y[n] = x[Mn],\ -\infty < \mathrm{n} < \infty$ is not time-invariant.

- **Causality**

  A system is causal if, for every choice of n_0, the output sequence value at the index $n = n_0$ depends only on the input sequence values for $n < n_0$. This implies that if for $n < n_0$, $x_1[n] = x_2[n]$, then $y_1[n] = y_2[n]$ for $n < n_0$. That is the system is nonanticipative. a system which shifts the input samples by n_d samples (i.e. $y[n] = x[n - n_d]$) is causal if $n_d > 0$. The backward difference system defined as:

  $$y[n] = x[n] - x[n-1] \tag{3-14}$$

  is another example of causal systems. In contrast, the forward difference system defined by the following relationship is not causal:

  $$y[n] = x[n+1] - x[n] \tag{3-15}$$

- **Stability**

  A system is stable in the bounded-input, bounded-output (BIBO) sense if and only if every bounded input sequence produces a bounded output sequence. The input $x[n]$ is bounded if there exists a fixed positive finite value B_x such that $|x[n]| < B_x < \infty$ for all n. Stability requires that, for every bounded input, there exist a fixed positive finite value B_y such that $|y[n]| < B_y < \infty$ for all n.

  The system $y[n] = (x[n])^2$ is stable. To see this, assume that the input $x[n]$ is bounded such that $|x[n]| < B_x$ for all n. Then $|y[n]| = |x[n]|^2 < B_x^2$. Thus by choosing $B_y = B_x^2$, it is shown that $y[n]$ is bounded. We can show the system defined by $w[n] = log_{10}(|x[n]|)$ is not bounded. How?

  The accumulator system defined by Eq. (3-13) is not stable either. Why?

3.3 Linear Time Invariant Systems

A particularly important class of systems consists of those that are linear and time invariant. This class of systems has significant signal-processing applications. The class of linear systems is defined by the principle of superposition in Eq. (3-12). If the linearity property is combined with the representation of a general sequence as a linear combination of delayed impulses as in Eq. (3-7), it follows that a linear system can be completely characterized by its impulse response.

Specifically, let $h_k[n]$ be the response of the system to $\delta[n-k]$, an impulse occurring at $n = k$. Then, from Eq. (3-7),

$$y[n] = T\{\sum_{k=-\infty}^{\infty} x[k]\delta[n-k]\}. \quad (3\text{-}16)$$

From the principle superposition in Eq. (3-12), we can write

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]\, T\{\delta[n-k]\} = \sum_{k=-\infty}^{\infty} x[k]h_k[n]\,. \quad (3\text{-}17)$$

According to Eq. (3-17), the system response to any input can be expressed in terms of the responses of the system to the sequences $\delta[n-k]$. If only linearity is imposed, $h_k[n]$ will depend on both n and k, in which case the computational usefulness of Eq. (3-17) is limited. We obtain a more useful result if we impose the additional constraint of time invariance.

The property of time invariance implies that if $h[n]$ is the response to $\delta[n]$, then the response to $\delta[n-k]$ is $h[n-\ k]$. With this additional constraint, Eq. (3-17) becomes

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k]\,. \quad (3\text{-}18)$$

Equation (3-18) is commonly called the convolution sum. If $y[n]$ is a sequence whose values are related to the values of two sequences $h[n]$ and $x[n]$ as in Eq. (3-18), it is said that $y[n]$ is the convolution of $x[n]$ with $h[n]$ and represent this by the notation:

$$y[n] = x[n] * h[n]\,. \quad (3\text{-}19)$$

As a consequence of Eq. (3-17), a linear time-invariant (also called LTI) system is completely characterized by its impulse response $h[n]$ in the sense that, given $h[n]$, it is possible to use Eq. (3-18) to compute the output $y[n]$ in response to any input $x[n]$.

To implement discrete-time convolution, the two sequences $x[k]$ and $h[n — k]$ are multiplied together for $-\infty < k < \infty$, and the products are summed to compute the output sample $y[n]$. To obtain another output sample, the origin of the sequence $h[n-k]$ is shifted to the new sample position, and the process is repeated. This computational procedure applies whether the computations are carried out numerically on sampled data or analytically with sequences for which the sample values have simple formulas.

Example 3-1: suppose $h[k]$ is the sequence shown in Figure 3-3(a) and we wish to find $h[n-\ k] = h[-\,(k-n)]$. Define $h_1[k]$ to be $h[-k]$, which is shown in Figure 3-3(b). Next, define $h_2[k]$ to be $h_1[k]$ delayed, by n samples on the k axis, i.e., $h_2[k] = h_1[k-n]$. Figure 3-

3(c) shows the sequence that results from delaying the sequence in Figure 3-3(b) by n samples. Using the relationship between $h_1[k]$ and $h[k]$, we can show that $h_2[k] = h_1[k-n] = h[-(k-n)] = h[n-k]$, and thus, the bottom figure is the desired signal.

To summarize, to compute $h[n-k]$ from $h[k]$, we first reverse $h[k]$ in time about $k = 0$ and then delay the time-reversed signal by n samples.

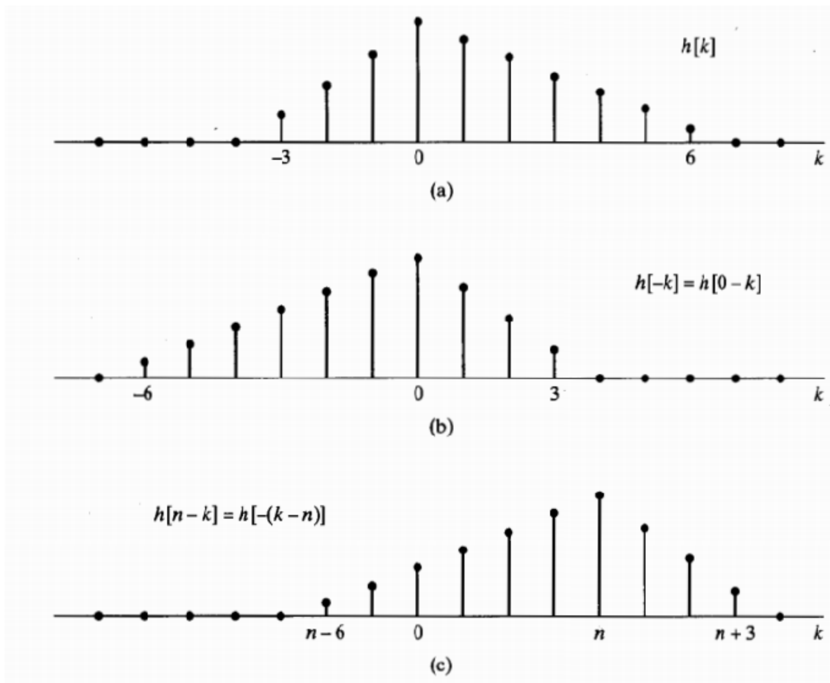


Figure 3-3: Forming the sequence $h[n-k]$ [1].

Example 3-2: Consider a system with impulse response:

$$h[n] = u[n] - u[n-N] = \begin{cases} 1 & 0 \le n \le N-1 \\ 0 & otherwise \end{cases}. \qquad (3\text{-}20)$$

and the input $x[n] = \alpha^n u[n]$. In order to calculate $y[n] = x[n] * h[n]$, three distinct regions must be considered (see Figure 3-4):

1- For $n < 0$, the non-zero portion of the sequences $x[k]$ and $h[n-k]$ do not overlap so that $y[n] = 0$ (Figure 3-4(a)).

2- For $0 \le n \le N-1$, as depicted in Figure 3-4(b), since $x[k]h[n-k] = a^k$, it follows that $y[n] = \sum_{k=0}^{n} a^k = \frac{1-a^{n+1}}{1-a}$.

In this range, $y[n]$ exponentially approaches the asymptote, $\frac{1}{1-a}$.

3- For $n > N$, as it is clear from (Figure 3-4(c)), $y[n] = \sum_{k=n-N+1}^{n} a^k = a^{n-N+1}(\frac{1-a^N}{1-a})$.

Therefore, the output, $y[n]$ as a function of index, n, is described by the following closed-form expression:

$$y[n] = \begin{cases} 0 & n < 0 \\ \frac{1-a^{n+1}}{1-a} & 0 \le n \le N-1 \\ a^{n-N+1}(\frac{1-a^N}{1-a}) & n > N-1 \end{cases} . \tag{3-21}$$

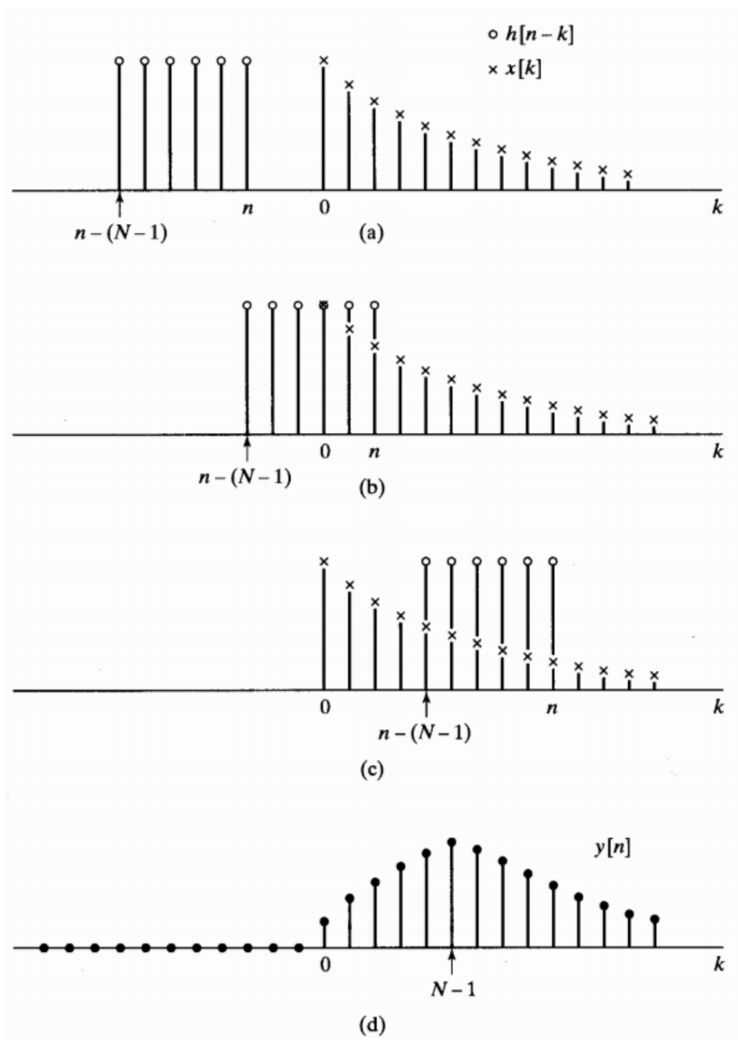


Figure 3-4: Sequences involved in computing the discrete convolution in example 3-2 [1].

3.4 Properties of Linear Time Invariant Systems

Since all linear time-invariant systems are described by the convolution sum of Eq. (3-19), the properties of this class of systems are defined by the properties of discrete-time convolution. Therefore, the impulse response is a complete characterization of the properties of a specific linear time-invariant system. Some general properties of the class of LTI systems are:

1- Commutative Property:

$$x[n] * h[n] = h[n] * x[n]\,. \tag{3-22}$$

Following this property, in a cascade connection of systems (let's say three systems with impulse responses, $h_1[n], h_2[n]$ and $h_3[n]$, respectively), the impulse response of the whole system is the convolution of the impulse responses of the sub-systems, that is, $h_1[n] * h_2[n] * h_3[n]$. As a consequence of the commutative property of convolution, the impulse response of a cascade combination of linear time-invariant systems is independent of the order in which they are cascaded. This is illustrated in Figure 3-5.

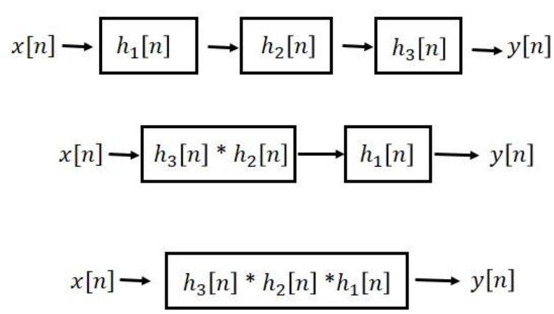


Figure 3-5: Three LTI systems with identical impulse response.

2- Distributive property over addition:

The convolution operation also distributes over addition:

$$x[n] * (h_1[n] + h_2[n]) = x[n] * h_1[n] + x[n] * h_2[n]\ . \tag{3-23}$$

Following this property, in a parallel connection of systems (let's say three systems with impulse responses, $h_1[n], h_2[n]$ and $h_3[n]$, respectively), the sub-systems have the same input, and their outputs are summed to produce an overall output. In this case, the impulse response of the whole system is equivalent to the sum of the individual impulse responses; that is $h[n] = h_1[n] + h_2[n] + h_3[n]$. This is depicted in Figure 3-6.

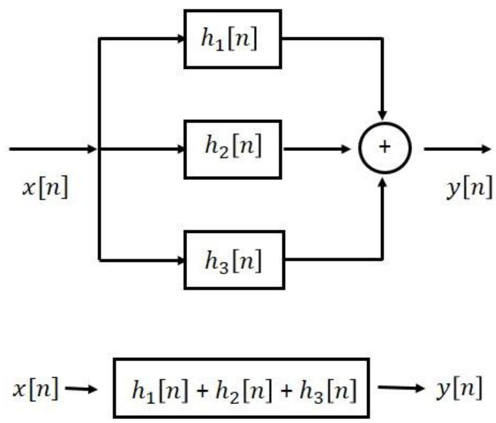


Figure 3-6: Parallel combination of LTI systems and the equivalent system.

3.5 Linear Constant- Coefficient Differential Equations

An important subclass of linear time-invariant systems consists of those systems for which the input $x[n]$ and the output $y[n]$ satisfy an Nth-order linear constant-coefficient difference equation of the form:

$$\sum_{k=0}^{N} a_k y[n-k] = \sum_{m=0}^{M} b_m x[n-m] \,. \qquad (3\text{-}24)$$

Writing the equation stated in Eq. (3-24) in the form of Eq. (3-25), the output, $y[n]$, at time n can be calculated from a finite number of previous values of the input (i.e., $x[n]$) and the output:

$$y[n] = \sum_{k=1}^{N} a'_k y[n-k] + \sum_{m=0}^{M} b'_m x[n-m] \,. \qquad (3\text{-}25)$$

The order of the system is defined by the maximum of the numbers M and K. If the input signal is defined after a certain initial time (e.g., for $n \geq n_0$) then values of both the input and output for a short time prior to n_0 must be known in order to initialize the difference equation. Specifically, $y[n]$ must be known for $n_0 - K \leq n \leq n_0 - 1$, and $x[n]$ for $n_0 - M \leq n \leq n_0 - 1$. In many applications, it is justified to assume that these values are zero if the system has not received any input for a long time before n_0, so that its response to any previous inputs has decayed to zero.

Here are examples of discrete-time systems governed by a linear constant-coefficient difference equation which are commonly served as digital filters:

- A plain amplifier $y[n] = Gx[n]$ (3-26)
- A delay system $y[n] = x[n - n_0]$ (3-27)
- Two-point moving average $y[n] = \frac{1}{2}(x[n] + x[n - 1])$ (3-28)
- Moving Average System $y[n] = \frac{1}{M+1}\sum_{k=0}^{M} x[n-k]$ (3-29)
- Digital "Leaky Integrator" $y[n] = \alpha y[n-1] + x[n], 0 < \alpha < 1$ (3-30)

There exist two classes of systems described by Eq. (3-25), the systems with *finite impulse response (FIR)* and those with *infinite impulse response (IIR)*.

If all the a'_k coefficients in Eq. (3-25) are zero, then the output depends only on a finite samples of the input. Such filters are called finite impulse response (FIR), or moving average (MA) filters. In a MA system, the output is a weighted sum (or average) of samples of the input. In contrast, if at least one of the a'_k coefficients in Eq. (3-25) is non-zero, the systems is known to have an IIR. Autoregressive (AR) and Autoregressive, moving-average (ARMA) filters are examples of such systems. If all of the b'_m coefficients except b'_0 are zero, the output depends only on the current value of the input and a finite number of past values of the output. Such systems are called purely recursive, or autoregressive (AR) filters. The term "autoregressive" means that the output is approximately a sum of its own past values. Digital "leaky integrator" is an example of AR filters. In the general case, both a'_k and b'_m in Eq. (3-25) are nonzero, with $N > 1$ and $M > 0$. Such systems are called autoregressive, moving-average (ARMA) filters. An example for ARMA filters is:

$$y[n] = \frac{y[n-1]+(x[n]+x[n-1])T_s}{2}. \quad (3\text{-}31)$$

3.6 Frequency- Domain Representation of Discrete Time Signals and Systems

LTI systems have two practically useful properties:

1- A representation of the input sequence as a weighted sum of delayed impulses leads to a representation of the output as a weighted sum of delayed impulse responses.

2- Complex exponential sequences are eigenfunctions of linear time-invariant systems and the response to a sinusoidal input is sinusoidal with the same frequency as the input and with amplitude and phase determined by the system. Regarding this fundamental property, representation of the signals in terms of sinusoids or complex exponentials is practically useful.

To demonstrate the eigenfunction property of complex exponentials for discrete-time systems, consider a complex exponential of radian frequency ω as an input sequence, i.e., $x[n] = e^{j\omega n}$, for $\infty < n < \infty$. If the impulse response of the system is $h[n]$, regarding the commutative property of LTI systems (Eq. (3-22)), the corresponding output of the system is

$$y[n] = \sum_{k=-\infty}^{\infty} h[k]e^{j\omega(n-k)} = e^{j\omega n}\left(\sum_{k=-\infty}^{\infty} h[k]e^{-j\omega k}\right). \tag{3-32}$$

Now, if we define

$$H\left(e^{j\omega}\right) = \sum_{k=-\infty}^{\infty} h[k]e^{-j\omega k}, \tag{3-33}$$

Eq. (3-32) becomes

$$y[n] = H\left(e^{j\omega}\right)e^{j\omega n} = H\left(e^{j\omega}\right)x[n] \ , \tag{3-34}$$

As a result, $e^{j\omega n}$is an eigenfunction of the system and the associated eigenvalue is $H\left(e^{j\omega}\right)$. The eigenvalue $H\left(e^{j\omega}\right)$ is called the *frequency response of the system* and describes the changes in complex amplitude of a complex exponential input signal as a function of the frequency, ω.

In general, $H\left(e^{j\omega}\right)$ is complex and commonly expressed in terms of its magnitude and phase as:

$$H\left(e^{j\omega}\right) = \left|H\left(e^{j\omega}\right)\right| e^{j\measuredangle H\left(e^{j\omega}\right)} \tag{3-35}$$

or in terms of its real and imaginary parts as:

$$H\left(e^{j\omega}\right) = H_R\left(e^{j\omega}\right) + jH_I\left(e^{j\omega}\right) \tag{3-36}$$

Example 3-3: Frequency response of the delay system:

The delay system is defined by

$$y[n] = x[n - n_d] \tag{3-37}$$

where n_d is a fixed integer. As we consider the $e^{j\omega n}$ as the input to this system, then the output is

$$y[n] = e^{j\omega(n- n_d)} = e^{-j\omega n_d}\, e^{j\omega n} \tag{3-38}$$

Therefore, $H(e^{j\omega}) = e^{-j\omega n_d}$. That is for a given frequency ω, we obtain the output that is the input multiplied by a complex constant. The value of this constant depends on the frequency ω and the delay n_d. For this frequency response, the magnitude and phase are:

$$|H(e^{j\omega})| = 1, \measuredangle H(e^{j\omega}) = -\omega n_d \tag{3-39}$$

From the Euler relation, its real and imaginary parts are as follows:

$$H_R(e^{j\omega}) = \cos(\omega n_d), H_I(e^{j\omega}) = -\sin(\omega n_d) \tag{3-40}$$

Example 3-4: Frequency response of the general moving average system:

The impulse response of the moving average system is defined by Eq. (3-41).

$$h[n] = \begin{cases} \frac{1}{M_1+M_2+1} & -M_1 < n < M_2 \\ 0 & otherwise \end{cases} \tag{3-41}$$

Based on its definition (i.e., Eq. (3-33)), the frequency response is:

$$H(e^{j\omega}) = \frac{1}{M_1+M_2+1} \sum_{n=-M_1}^{M_2} e^{-j\omega n}, \tag{3-42}$$

which can be expressed by the following closed-form description:

$$H(e^{j\omega}) = \frac{1}{M_1 + M_2 + 1} \frac{e^{j\omega M_1} - e^{-j\omega(M_2+1)}}{1 - e^{-j\omega}} \tag{3-43}$$

$$\ldots \quad = \frac{1}{M_1+M_2+1} \frac{\sin(\omega(M_1+M_2+1)/2)}{\sin(\omega/2)} e^{-j\omega(M_2-M)/2}.$$

To obtain this closed-form representation, note that $\sum_{n=-M_1}^{M_2} \alpha^n = \frac{\alpha^{M_1} - \alpha^{M_2+1}}{1-\alpha}$ and $\sin\omega = \frac{e^{j\omega}-e^{-j\omega}}{2}$.

The magnitude and the phase of this frequency response are illustrated in Figure 3-7.

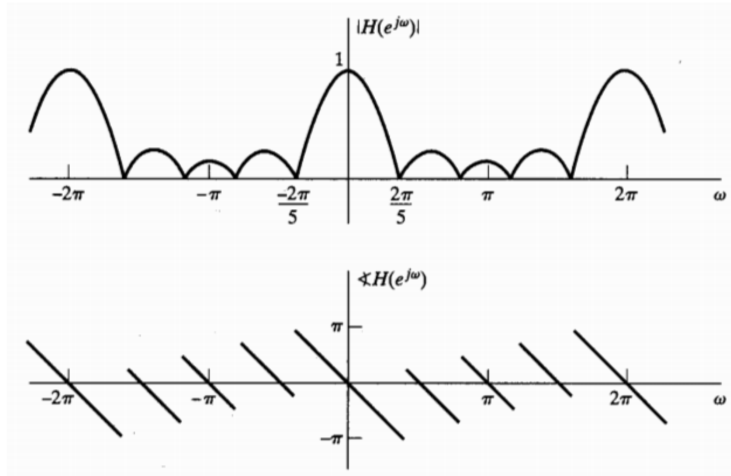


Figure 3-7: magnitude and phase of the frequency response of the moving average system with $M_1 = 0, M_2 = 4$ [1].

An important class of linear time-invariant systems includes those systems for which the frequency response is unity over a certain range of frequencies and is zero at the remaining frequencies. These correspond to *ideal frequency-selective* filters. The frequency responses of these commonly used filters are shown in Figure 3-8.

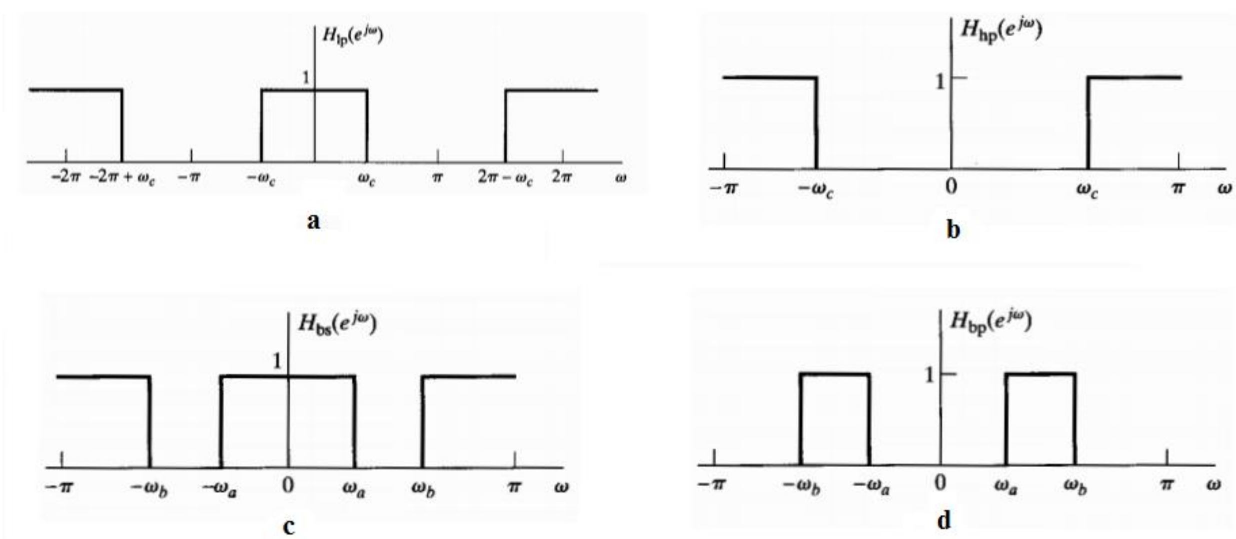


Figure 3-8: frequency response of ideal selective filters. a) low-pass, b) high-pass, c) band-stop and d) band-pass filters [1].