ECE 431
Digital Signal Processing
Lecture Notes

Prof. Dan Cobb

Contents

1 Introduction

Digital Signal Processing (DSP) is the application of a digital computer to modify an analog or digital signal. Typically, the signal being processed is either temporal, spatial, or both. For example, an audio signal is temporal, while an image is spatial. A movie is both temporal and spatial. The analysis of temporal signals makes heavy use of the Fourier transform in one time variable and one frequency variable. Spatial signals require two independent variables. Analysis of such signals relies on the Fourier transform in two frequency variables (e.g. ECE 533). In ECE 431, we will restrict ourselves to temporal signal processing.

Our main goal is to be able to design digital LTI filters. Such filters are using widely in applications such as audio entertainment systems, telecommunication and other kinds of communication systems, radar, video enhancement, and biomedical engineering. The first half of the course will be spent reviewing and developing the fundamentals necessary to understand the design of digital filters. Then we will examine the basic types of filters and the myriad of design issues surrounding them.

From the outset, the student should recognize that there are two distinct classes of applications for digital filters. *Real-time* applications are those where data streams into the filter and **must be processed immediately**. A significant delay in generating the filter output data cannot be tolerated. Such applications include communication networks of all sorts, musical performance, public address systems, and patient monitoring. Real-time filtering is sometimes called *on-line* processing and is based on the theory of causal systems.

Non-real-time applications are those where a filter is used to process a pre-existing (i.e. stored) file of data. In this case, the engineer is typically allotted a large amount of time over which the processing of data may be performed. Such applications include audio recording and mastering, image processing, and the analysis of seismic data. Non-real-time filtering is sometimes called *off-line* processing and is based on the theory of noncausal systems. In these applications, the fact that noncausal filters may be employed opens the door to a much wider range of filters and commensurately better results. For example, one problem typical of real-time filtering is phase distortion, which we will study in detail in this course. Phase distortion can be eliminated completely if noncausal filters are permitted.

The first part of the course will consist of review material from signals and systems. Throughout the course, we will rely heavily on the theory of Fourier transforms, since much of signal processing and filter theory is most easily addressed in the frequency domain. It will be convenient to refer to commonly used transform concepts by the following acronyms:

CTFT: Continuous-Time Fourier Transform
DTFT: Discrete-Time Fourier Transform
CFS: Continuous-Time Fourier Series
DFS: Discrete-Time Fourier Series
LT: Laplace Transform
DFT: Discrete Fourier Transform
ZT: z-Transform

An "I" preceding an acronym indicates "Inverse" as in IDTFT and IDFT. All of these concepts should be familiar to the student, except the DFT and ZT, which we will define and study in detail.

2 Review of the DT Fourier Transform

2.1 Definition and Properties

The *CT Fourier transform (CTFT)* of a CT signal $x(t)$ is

$$\mathcal{F}\{x(t)\} = X(j\omega) = \int_{-\infty}^{\infty} x(t)\, e^{-j\omega t} dt.$$

The *Inverse CT Fourier Transform (ICTFT)* is

$$\mathcal{F}^{-1}\{X(j\omega)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(j\omega)\, e^{j\omega t} d\omega.$$

Recall the CT unit impulse $\delta(t)$, the DT unit impulse $\delta[n]$, and their basic properties:

$$\int_{-\infty}^{\infty} \delta(t)\,dt = 1, \qquad \sum_{n=-\infty}^{\infty} \delta[n] = 1$$

$$x(t)\,\delta(t-\tau) = x(\tau)\,\delta(t-\tau), \quad x[n]\,\delta[n-m] = x[m]\,\delta[n-m]$$

$$x(t) * \delta(t-\tau) = x(t-\tau), \quad x[n] * \delta[n-m] = x[n-m] \quad \text{(sifting property)}.$$

For any DT signal $x[n]$, we may define its *DT Fourier transform (DTFT)* by associating with $x[n]$ the CT impulse train

$$x(t) = \sum_{n=-\infty}^{\infty} x[n]\,\delta(t-n)$$

and taking the transform

$$\begin{aligned} X(j\omega) &= \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x[n]\,\delta(t-n)\,e^{-j\omega t}dt \\ &= \sum_{n=-\infty}^{\infty} x[n]\,e^{-j\omega n} \int_{-\infty}^{\infty} \delta(t-n)\,dt \\ &= \sum_{n=-\infty}^{\infty} x[n]\,e^{-j\omega n}. \end{aligned}$$

Thus we may write

$$X(j\omega) = \sum_{n=-\infty}^{\infty} x[n]\left(e^{j\omega}\right)^{-n},$$

expressing X as a function of $e^{j\omega}$. For this reason, the DTFT is normally written

$$X\left(e^{j\omega}\right) = \sum_{n=-\infty}^{\infty} x[n]\,e^{-j\omega n}.$$

Technically, this is an abuse of notation, since the two X's are actually different functions, but the meaning will usually be clear from context. In order to help distinguish between CT and DT transforms, we will henceforth denote the frequency variable in DT transforms as Ω:

$$X\left(e^{j\Omega}\right) = \sum_{n=-\infty}^{\infty} x[n]\,e^{-j\Omega n}. \tag{2.1}$$

Although your text writes frequency as ω for both CT and DT transforms, the Ω notation has numerous advantages. For example, it keeps the units of frequency straight: ω is in rad/sec, while Ω is in radians.

By Euler's formula,

$$e^{j\Omega} = \cos\Omega + j\sin\Omega,$$

so $e^{j\Omega}$ is periodic with fundamental period 2π. Hence, $X\left(e^{j\Omega}\right)$ has period 2π. We also write

$$\mathcal{F}\{x[n]\} = X\left(e^{j\Omega}\right)$$

and

$$x[n] \longleftrightarrow X\left(e^{j\Omega}\right).$$

The *Inverse DTFT* is

$$\mathcal{F}^{-1}\left\{X\left(e^{j\Omega}\right)\right\} = x[n] = \frac{1}{2\pi}\int_0^{2\pi} X\left(e^{j\Omega}\right)e^{j\Omega n}d\Omega.$$

The integral may be evaluated over any interval of length 2π.

Properties: (See O&S Table 2.1 on p. 55 and Table 2.2 on p. 58.)

Periodicity:

$$X\left(e^{j(\Omega+2\pi)}\right) = X\left(e^{j\Omega}\right)$$

Linearity:

$$\begin{cases} \alpha x[n] \longleftrightarrow \alpha X\left(e^{j\Omega}\right) \\ x_1[n] + x_2[n] \longleftrightarrow X_1\left(e^{j\Omega}\right) + X_2\left(e^{j\Omega}\right) \end{cases}$$

Time Shift:

$$x[n-n_0] \longleftrightarrow e^{-j\Omega n_0}X\left(e^{j\Omega}\right)$$

Frequency Shift:

$$e^{j\Omega_0 n}x[n] \longleftrightarrow X\left(e^{j(\Omega-\Omega_0)}\right)$$

Time/Frequency Scaling:

$$x_{(N)}[n] = \begin{cases} x\left[\frac{n}{N}\right], & \frac{n}{N} \text{ an integer} \\ 0, & \text{else} \end{cases}$$

$$x_{(N)}[n] \longleftrightarrow X\left(e^{j\Omega N}\right)$$

Convolution:

$$x_1[n] * x_2[n] = \sum_{m=-\infty}^{\infty} x_1[n-m]\,x_2[m]$$

$$x_1[n] * x_2[n] \longleftrightarrow X_1\left(e^{j\Omega}\right)X_2\left(e^{j\Omega}\right)$$

Multiplication:

$$x_1[n]\,x_2[n] \longleftrightarrow \frac{1}{2\pi}\int_0^{2\pi} X_1\left(e^{j(\Omega-\theta)}\right)X_2\left(e^{j\theta}\right)d\theta$$

Time Differencing:

$$x[n] - x[n-1] \longleftrightarrow \left(1-e^{-j\Omega}\right)X\left(e^{j\Omega}\right)$$

Accumulation:

$$\sum_{m=-\infty}^{n} x[m] \longleftrightarrow \frac{1}{1-e^{-j\Omega}}X\left(e^{j\Omega}\right)$$

Frequency Differentiation:

$$nx[n] \longleftrightarrow j\frac{dX\left(e^{j\Omega}\right)}{d\Omega}$$

Conjugation:

$$x^*[n] \longleftrightarrow X^*\left(e^{-j\Omega}\right)$$

Reflection:

$$x[-n] \longleftrightarrow X\left(e^{-j\Omega}\right)$$

Real Time Signal:

$$x[n] \text{ real } \Longleftrightarrow \left\{ \begin{array}{ll} \left|X\left(e^{j\Omega}\right)\right| & \text{even} \\ \angle X\left(e^{j\Omega}\right) & \text{odd} \end{array} \right.$$

Even-Odd:

$$\left\{ \begin{array}{l} x[n] \text{ even } \Longleftrightarrow X\left(e^{j\Omega}\right) \text{ real} \\ x[n] \text{ odd } \Longleftrightarrow X\left(e^{j\Omega}\right) \text{ imaginary} \end{array} \right.$$

Parseval's Theorem:

$$\sum_{n=-\infty}^{\infty} x_1[n]\, x_2^*[n] = \frac{1}{2\pi} \int_0^{2\pi} X\left(e^{j\Omega}\right) Y^*\left(e^{j\Omega}\right) d\Omega$$

Example 2.1 *The DT unit impulse*

$$\delta[n] = \left\{ \begin{array}{ll} 1, & n = 0 \\ 0, & n \neq 0 \end{array} \right.$$

has DTFT

$$\mathcal{F}\{\delta[n]\} = \sum_{n=-\infty}^{\infty} \delta[n]\, e^{-j\Omega n} = 1.$$

Example 2.2 *The unit impulse train in frequency*

$$X\left(e^{j\Omega}\right) = \sum_{k=-\infty}^{\infty} \delta(\Omega - 2\pi k)$$

has Inverse DTFT

$$\begin{aligned} x[n] &= \frac{1}{2\pi} \int_{0^-}^{2\pi^-} \left(\sum_{k=-\infty}^{\infty} \delta(\Omega - 2\pi k) \right) e^{j\Omega n} d\Omega \\ &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \int_{0^-}^{2\pi^-} \delta(\Omega - 2\pi k)\, e^{j2\pi k n} d\Omega \\ &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \int_{0^-}^{2\pi^-} \delta(\Omega - 2\pi k)\, d\Omega. \end{aligned}$$

But

$$\int_{0^-}^{2\pi^-} \delta(\Omega - 2\pi k)\, d\Omega = \left\{ \begin{array}{ll} 1, & k = 0 \\ 0, & k \neq 0 \end{array} \right. ,$$

so

$$x[n] = \frac{1}{2\pi}$$

and

$$1 \longleftrightarrow 2\pi \sum_{k=-\infty}^{\infty} \delta(\Omega - 2\pi k).$$

Example 2.3 *Define the DT rectangular window*

$$w_N[n] = \begin{cases} 1, & \leq n \leq N-1 \\ 0, & else \end{cases}.$$

The DTFT is

$$\begin{aligned}
W_N\left(e^{j\Omega}\right) &= \sum_{n=-\infty}^{\infty} w_N[n]\, e^{-j\Omega n} \\
&= \sum_{n=0}^{N-1} e^{-j\Omega n} \\
&= \sum_{n=0}^{N-1} \left(e^{-j\Omega}\right)^n \\
&= \frac{1-e^{-jN\Omega}}{1-e^{-j\Omega}} \\
&= \frac{e^{j\frac{\Omega}{2}}}{e^{j\frac{\Omega}{2}}} \frac{\left(e^{j\frac{(N-1)\Omega}{2}} - e^{-j\frac{(N+1)\Omega}{2}}\right) e^{-j\frac{(N-1)\Omega}{2}}}{1-e^{-j\Omega}} \\
&= \frac{e^{j\frac{N\Omega}{2}} - e^{-j\frac{N\Omega}{2}}}{e^{j\frac{\Omega}{2}} - e^{-j\frac{\Omega}{2}}} e^{-j\frac{(N-1)\Omega}{2}} \\
&= \frac{\sin\frac{N\Omega}{2}}{\sin\frac{\Omega}{2}} e^{-j\frac{(N-1)\Omega}{2}}.
\end{aligned}$$

The real factor in $W_N\left(e^{j\Omega}\right)$ *is the "periodic* sinc*" function:*

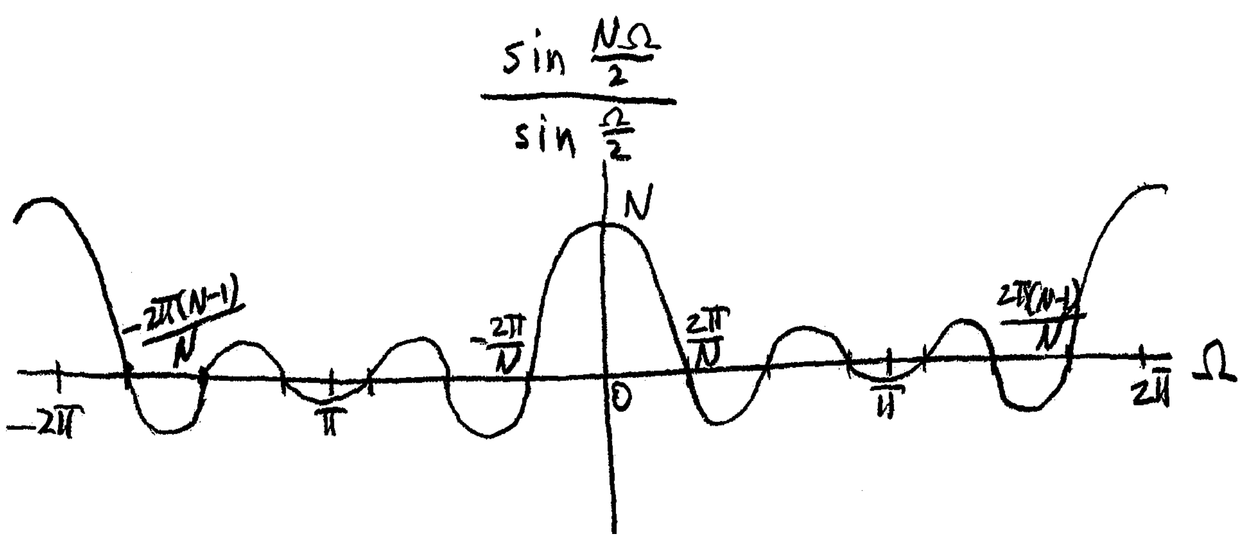


Figure 2.1

(See O&S Table 2.3 on p. 62 for further examples.)

2.2 Periodic Convolution

The multiplication property involves the *periodic convolution*

$$X_1\left(e^{j\Omega}\right) * X_2\left(e^{j\Omega}\right) = \int_0^{2\pi} X_1\left(e^{j(\Omega-\theta)}\right) X_2\left(e^{j\theta}\right) d\theta.$$

Since $X\left(e^{j\Omega}\right)$ and $Y\left(e^{j\Omega}\right)$ both have period 2π, the linear (i.e. ordinary) convolution blows up (except in trivial cases):

$$\begin{aligned}\int_{-\infty}^{\infty} X_1\left(e^{j(\Omega-\theta)}\right) X_2\left(e^{j\theta}\right) d\theta &= \sum_{i=-\infty}^{\infty} \int_{2\pi i}^{2\pi(i+1)} X_1\left(e^{j(\Omega-\theta)}\right) X_2\left(e^{j\theta}\right) d\theta \\ &= \sum_{i=-\infty}^{\infty} \int_{0}^{2\pi} X_1\left(e^{j(\Omega-\theta)}\right) X_2\left(e^{j\theta}\right) d\theta \\ &= \infty.\end{aligned}$$

On the other hand, the periodic convolution is well-defined with period 2π.

Example 2.4 *Consider the square wave*

$$X\left(e^{j\Omega}\right) = \begin{cases} 1, & 0 \le \Omega < \pi \\ 0, & \pi \le \Omega < 2\pi \end{cases}$$

with period 2π. *We wish to convolve* $X\left(e^{j\Omega}\right)$ *with itself. We need to look at two cases:*
1) $0 \le \Omega < \pi$

$$\int_0^{2\pi} X\left(e^{j(\Omega-\theta)}\right) X\left(e^{j\theta}\right) d\theta = \int_0^{\Omega} 1 d\theta = \Omega$$

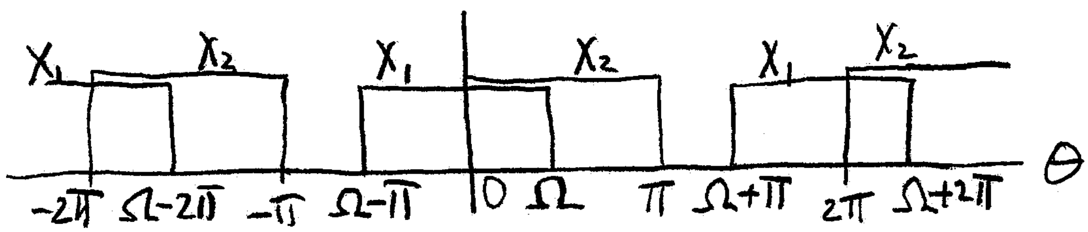


Figure 2.2

2) $\pi \le \Omega < 2\pi$

$$\int_0^{2\pi} X\left(e^{j(\Omega-\theta)}\right) X\left(e^{j\theta}\right) d\theta = \int_{\Omega-\pi}^{\pi} 1 d\theta = 2\pi - \Omega$$

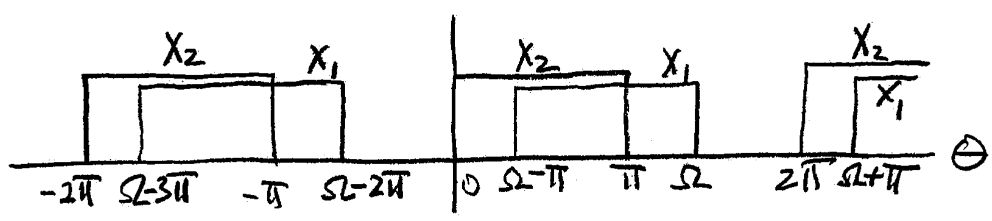


Figure 2.3

The periodic convolution is the triangle wave

$$X\left(e^{j\Omega}\right) * X\left(e^{j\Omega}\right) = \begin{cases} \Omega, & 0 \le \Omega < \pi \\ 2\pi - \Omega, & \pi \le \Omega < 2\pi \end{cases}$$

with period 2π.

Periodic convolution may also be defined for sequences. If $x_1[n]$ and $x_2[n]$ have period N, then

$$x_1[n] * x_2[n] = \sum_{m=0}^{N-1} x_1[n-m]\, x_2[m]$$

has period N.

2.3 Fourier Series

Let a_k be a sequence of complex numbers with period N and

$$\Omega_0 = \frac{2\pi}{N}.$$

Suppose we restrict attention to DT signals whose DTFT's are impulse trains of the form

$$X\left(e^{j\Omega}\right) = 2\pi \sum_{k=-\infty}^{\infty} a_k \delta\left(\Omega - \Omega_0 k\right). \tag{2.2}$$

Then

$$\begin{aligned} x[n] &= \frac{1}{2\pi}\int_0^{2\pi} X\left(e^{j\Omega}\right) e^{j\Omega n} d\Omega \\ &= \int_{0^-}^{2\pi^-} \sum_{k=-\infty}^{\infty} a_k \delta\left(\Omega - \Omega_0 k\right) e^{j\Omega n} d\Omega \\ &= \sum_{k=-\infty}^{\infty} a_k e^{j\Omega_0 kn} \int_{0^-}^{2\pi^-} \delta\left(\Omega - \Omega_0 k\right) d\Omega. \end{aligned}$$

But

$$\int_{0^-}^{2\pi^-} \delta\left(\Omega - \Omega_0 k\right) = \begin{cases} 1, & 0 \leq k \leq N-1 \\ 0, & \text{else} \end{cases},$$

so

$$x[n] = \sum_{k=0}^{N-1} a_k e^{j\Omega_0 kn}. \tag{2.3}$$

Note that

$$\begin{aligned} e^{j\Omega_0 k(n+N)} &= e^{j\Omega_0 kn} + e^{j\Omega_0 kN} \\ &= e^{j\Omega_0 kn} + e^{j2\pi k} \\ &= e^{j\Omega_0 kn}, \end{aligned}$$

so $e^{j\Omega_0 kn}$ and, therefore, $x[n]$ have period N.

Formula (2.3) is the *DT Fourier series (DFS)* representation of the periodic signal $x[n]$. The (complex) numbers a_k are the Fourier coefficients of $x[n]$. In this case, we write

$$x[n] \longleftrightarrow a_k.$$

Every DT signal $x[n]$ with period N has DTFT (2.2) and DFS (2.3). The Fourier coefficients also have period N and may be derived from $x[n]$ via the summation

$$a_k = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-j\Omega_0 kn}. \tag{2.4}$$

In both the DFS (2.3) and its inverse (2.4), the sum may be taken over any interval of length N.

The properties of the DFS are similar to those of the DTFT. (See O&S Table 8.1 on p. 634.)

Linearity:

$$\begin{cases} \alpha x[n] \longleftrightarrow \alpha a_k \\ x_1[n] + x_2[n] \longleftrightarrow a_k + b_k \end{cases}$$

Time-Shift:

$$x[n-n_0] \longleftrightarrow e^{-j\Omega_0 k n_0} a_k$$

Frequency Shift

$$e^{j\Omega_0 k_0 n} x[n] \longleftrightarrow a_{k-k_0}$$

Time/Frequency Scaling:

$$x_{(M)}[n] \longleftrightarrow \frac{1}{M} a_k \text{ (period } MN)$$

Convolution:

$$\sum_{m=0}^{N-1} x_1[n-m] x_2[m] \longleftrightarrow N a_k b_k$$

Multiplication:

$$x_1[n] x_2[n] \longleftrightarrow \sum_{i=0}^{N-1} a_{k-i} b_i$$

Time Differencing:

$$x[n] - x[n-1] \longleftrightarrow \left(1 - e^{-j\Omega_0 k}\right) a_k$$

Accumulation:

$$\sum_{m=-\infty}^{n} x[m] \longleftrightarrow \frac{1}{1 - e^{-j\Omega_0 k}} a_k \quad \text{(only for } a_0 = 0)$$

Frequency Differencing:

$$\left(1 - e^{j\Omega_0 n}\right) x[n] \longleftrightarrow a_k - a_{k-1}$$

Conjugation:

$$x^*[n] \longleftrightarrow a_{-k}^*$$

Reflection:

$$x[-n] \longleftrightarrow a_{-k}$$

Real Time Signal:

$$x[n] \text{ real} \iff \begin{cases} |a_k| \text{ even} \\ \angle a_k \text{ odd} \end{cases}$$

Even-Odd:

$$\begin{cases} x[n] \text{ even} \iff a_k \text{ real} \\ x[n] \text{ odd} \iff a_k \text{ imaginary} \end{cases}$$

Parseval's Theorem:

$$\frac{1}{N}\sum_{n=0}^{N-1} x_1[n]\, x_2^*[n] = \sum_{k=0}^{N-1} a_k b_k^*$$

Many of the properties of the DFS appear to be "mirror images" of one another. This principle is called *duality* and is the result of the similarity of equations (2.3) and (2.4). The same phenomenon can be seen with regard to transforms of specific signals.

Example 2.5 *Find the DTFT and DFS of*

$$x[n] = \sum_{m=-\infty}^{\infty} \delta[n - mN].$$

The coefficients are

$$a_k = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=-\infty}^{\infty} \delta[n-mN]\, e^{-j\Omega_0 kn} = \frac{1}{N}\sum_{m=-\infty}^{\infty} e^{-j\Omega_0 kmN}\left(\sum_{n=0}^{N-1}\delta[n-mN]\right).$$

But

$$\sum_{n=0}^{N-1}\delta[n-mN] = \begin{cases} 1, & m = 0 \\ 0, & m \neq 0 \end{cases},$$

so $a_k = \frac{1}{N}$ *for every* k*. The DFS is*

$$x[n] = \frac{1}{N}\sum_{k=0}^{N-1} e^{j\Omega_0 kn}$$

and the DTFT is

$$X\left(e^{j\Omega}\right) = 2\pi\sum_{k=-\infty}^{\infty} a_k\delta(\Omega - \Omega_0 k) = \Omega_0\sum_{k=-\infty}^{\infty}\delta(\Omega-\Omega_0 k).$$

Example 2.6 *From Example 2.5, the Fourier coefficients corresponding to an impulse train are constant. Now find the Fourier coefficients of* $x[n] = 1$*. By duality, we should get an impulse train.*

$$\begin{aligned} a_k &= \frac{1}{N}\sum_{n=0}^{N-1} x[n]\, e^{-j\Omega_0 kn} \\ &= \frac{1}{N}\sum_{n=0}^{N-1}\left(e^{-j\Omega_0 k}\right)^n \\ &= \begin{cases} 1, & k = mN \\ \frac{1}{N}\frac{1-\left(e^{-j\Omega_0 k}\right)^N}{1-e^{-j\Omega_0 k}}, & else \end{cases} \end{aligned}$$

But

$$\left(e^{-j\Omega_0 k}\right)^N = e^{-j2\pi k} = 1,$$

so

$$a_k = \begin{cases} 1, & k = mN \\ 0, & else \end{cases} = \sum_{i=-\infty}^{\infty}\delta[k - iN].$$

3 Sampling

3.1 Time and Frequency Domain Analysis

For any $T > 0$, we may *sample* a CT signal $x(t)$ to generate the DT signal

$$x[n] = x(nT).$$

This amounts to evaluating $x(t)$ at uniformly spaced points on the t-axis. The number T is the *sampling period,*

$$f_s = \frac{1}{T}$$

is the *sampling frequency,* and

$$\omega_s = 2\pi f_s = \frac{2\pi}{T}$$

is the *radian sampling frequency.* Normally, the units of f_s are Hertz or samples/sec. The units of ω_s are rad/sec. The time interval $[nT, (n+1)T]$ is called the nth sampling interval. The process of sampling is sometimes depicted as a switch which closes momentarily every T units of time:

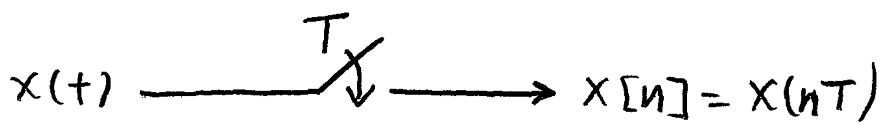


Figure 3.1

A useful expression for the DTFT of $x[n]$ can be obtained by writing $x(t)$ in terms of its inverse transform:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(j\omega) e^{j\omega t} d\omega$$

$$\begin{aligned}
x[n] &= x(nT) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} X(j\omega) e^{j\omega nT} d\omega \\
&= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \int_{k\omega_s}^{(k+1)\omega_s} X(j\omega) e^{j\omega nT} d\omega \\
&= \frac{1}{2\pi T} \sum_{k=-\infty}^{\infty} \int_{0}^{2\pi} X\left(j\frac{\Omega + 2\pi k}{T}\right) e^{j(\Omega + 2\pi k)n} d\Omega \quad (\Omega = \omega T - 2\pi k) \\
&= \frac{1}{2\pi} \int_{0}^{2\pi} \left(\frac{1}{T} \sum_{k=-\infty}^{\infty} X\left(j\frac{\Omega + 2\pi k}{T}\right)\right) e^{j\Omega n} d\Omega
\end{aligned}$$

The analysis shows that

$$X_{DT}\left(e^{j\Omega}\right) = \frac{1}{T} \sum_{k=-\infty}^{\infty} X_{CT}\left(j\frac{\Omega + 2\pi k}{T}\right). \tag{3.1}$$

Expression (3.1) is referred to as the *Poisson summation formula.*

3.2 Aliasing

We say a CT signal $x(t)$ is *bandlimited* if there exists $\omega_B < \infty$ such that $X_{CT}(j\omega) = 0$ for $|\omega| > \omega_B$. Suppose X_{CT} has transform depicted (very roughly) in Figure 3.2. (We use a signal with $X_{CT}(0) = 1$ for reference.)

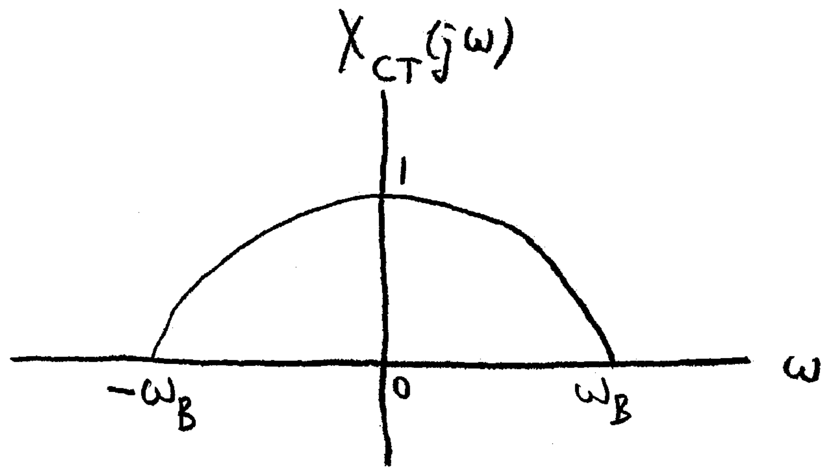


Figure 3.2

The number ω_B is the *bandwidth* of the signal. If $x(t)$ is bandlimited, (3.1) indicates that $X_{DT}\left(e^{j\Omega}\right)$ looks like Figure 3.3

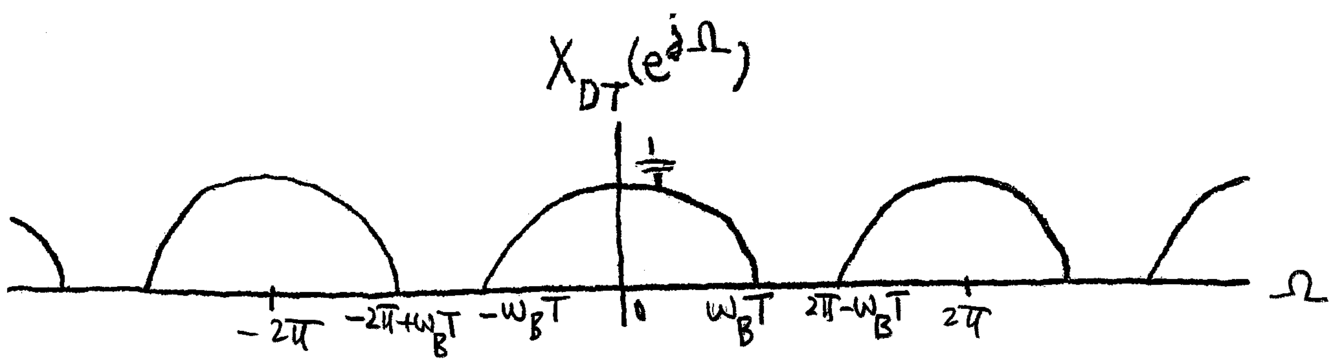


Figure 3.3

or Figure 3.4.

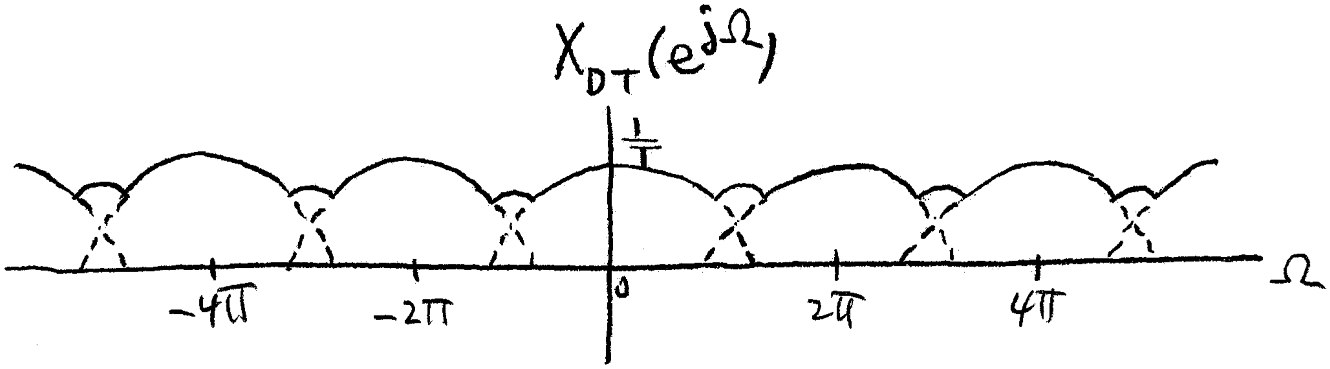


Figure 3.4

Figure 3.3 is drawn assuming

$$2\pi - \omega_B T > \omega_B T$$

or, equivalently,

$$\omega_s > 2\omega_B.$$

In this case, (3.1) indicates that

$$X_{DT}\left(e^{j\Omega}\right) = X_{CT}\left(j\frac{\Omega}{T}\right)$$

for $-\pi \leq \Omega < \pi$. This establishes the fundamental relationship between the CT and DT frequency variables ω and Ω under sampling:

$$\Omega = \omega T. \tag{3.2}$$

We will encounter equation (3.2) under a variety of circumstances when sampling is involved.

For $\omega_s < 2\omega_B$, the picture reverts to Figure 3.4. In this case, the shifts of $X_{CT}\left(j\frac{\Omega}{T}\right)$ overlap – a phenomenon called *aliasing*. As we will see, aliasing is undesirable in most signal processing applications. The minimum radian sampling frequency $\omega_s = 2\omega_B$ required to avoid aliasing is called the *Nyquist rate.*

3.3 The Nyquist Theorem

Consider the set Σ_{CT} of all CT signals $x(t)$ and the set Σ_{DT} of all DT signals $x[n]$. For a given sampling period T, the process of sampling may be viewed as a mapping from Σ_{CT} into Σ_{DT} :

$$x(t) \longmapsto x[n] = x(nT).$$

That is, each CT signal generates exactly one DT signal. The following example shows that the mapping changes if the sampling period changes.

Example 3.1 *Let $x(t) = \sin t$ and $T = \frac{\pi}{2}$. Then*

$$x[n] = \sin\left(n\frac{\pi}{2}\right) = \begin{cases} (-1)^{\frac{n-1}{2}}, & n \text{ odd} \\ 0, & n \text{ even} \end{cases}.$$

On the other hand, setting $T = \pi$ yields

$$x[n] = \sin(n\pi) = 0.$$

Thus sampling $\sin t$ *results in different signals for different T.*

The next example shows that the sampling map may not be $1-1$.

Example 3.2 *Let* $x_1(t) = \sin t$ *and* $x_2(t) = 0$. *For* $T = \pi$,

$$x_1[n] = x_2[n] = 0,$$

so the distinct CT signals $x_1(t)$ *and* $x_2(t)$ *map into the same DT signal.*

Now let $\Sigma_{\omega_B} \subset \Sigma_{CT}$ be the set of all CT signals with bandwidth at most ω_B. In Example 3.2, both $x_1(t)$ and $x_2(t)$ belong to Σ_1. Yet, they map into the same DT signal for $T = \pi$. In other words, the sampling map may not be $1-1$ even on Σ_{ω_B}. Also, note that in Example 3.2,

$$\omega_s = \frac{2\pi}{T} = 2 = 2\omega_B,$$

so we are sampling at exactly the Nyquist rate. The situation is clarified by the Nyquist Sampling Theorem:

Theorem 3.1 *The sampling map is* $1-1$ *on* Σ_{ω_B} *iff* $\omega_s > 2\omega_B$.

The Nyquist theorem states that, if we are given a signal $x(t)$ and we sample at greater than the Nyquist rate, then there is no loss of information in replacing $x(t)$ by its samples $x[n]$. In other words, $x(t)$ can be recovered from $x[n]$. However, if we sample at or below the Nyquist rate, then knowledge of $x[n]$ is insufficient to determine $x(t)$ uniquely.

3.4 Anti-Aliasing Filters

In order to avoid aliasing, we may set the sampling rate $\omega_s > 2\omega_B$. However, in certain applications it is desirable to achieve the same end by reducing the bandwidth of $x(t)$ prior to sampling. This can be done by passing $x(t)$ through a CT filter. Define the *ideal CT low-pass filter (LPF)* to be the CT LTI system with transfer function

$$H_{LP}(j\omega) = \begin{cases} 1, & |\omega| \leq 1 \\ 0, & |\omega| > 1 \end{cases}. \tag{3.3}$$

If we pass $x(t)$ through the frequency-scaled filter $H_{LP}\left(j\frac{\omega}{\omega_B}\right)$, then $X(j\omega)$ is "chopped" down to bandwidth ω_B. An LPF used in this way is called an *anti-aliasing filter* and **must be built from analog components**.

The impulse response of $H_{LP}(j\omega)$ is

$$\begin{aligned} h_{LP}(t) &= \mathcal{F}^{-1}\{H_{LP}(j\omega)\} \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} H_{LP}(j\omega)\, e^{j\omega t} d\omega \\ &= \frac{1}{2\pi}\int_{-1}^{1} e^{j\omega t} d\omega \\ &= \frac{1}{j2\pi t}\left(e^{jt} - e^{-jt}\right) \\ &= \frac{\sin t}{\pi t}. \end{aligned}$$

Let

$$\operatorname{sinc} t = \begin{cases} 1, & t = 0 \\ \frac{\sin(\pi t)}{\pi t}, & t \neq 0 \end{cases}.$$

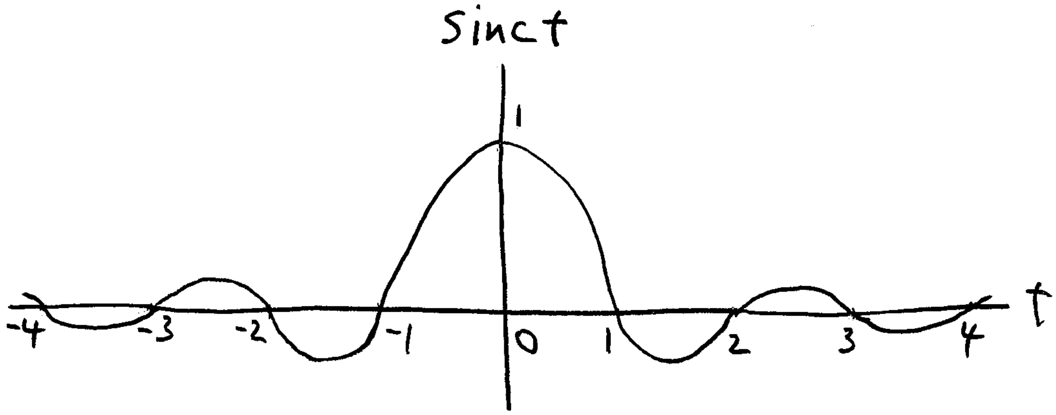


Figure 3.5

We may write

$$h_{LP}(t) = \frac{1}{\pi} \operatorname{sinc} \frac{t}{\pi} \tag{3.4}$$

Note that $h_{LP}(t)$ has a couple of unfortunate features:

1) $h_{LP}(t) \neq 0$ for $t < 0$, so the ideal LPF is **noncausal**.
2)

$$\begin{aligned}
\int_{-\infty}^{\infty} |h_{LP}(t)|\, dt &= \frac{1}{\pi} \int_{-\infty}^{\infty} \left| \frac{\sin(\pi t)}{t} \right| dt \\
&= \frac{2}{\pi} \int_{0}^{\infty} \left| \frac{\sin(\pi t)}{t} \right| dt \\
&= \frac{2}{\pi} \sum_{n=1}^{\infty} \int_{n-1}^{n} \left| \frac{\sin(\pi t)}{t} \right| dt \\
&\geq \frac{2}{\pi} \sum_{n=1}^{\infty} \left(\frac{1}{n} \int_{n-1}^{n} |\sin(\pi t)|\, dt \right) \\
&= \frac{2}{\pi} \sum_{n=1}^{\infty} \left(\frac{1}{n} \int_{0}^{1} \sin(\pi t)\, dt \right) \\
&= \frac{2}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n} \\
&= \infty.
\end{aligned}$$

Hence, the ideal LPF is **not BIBO stable**.

Although an ideal LPF cannot be realized in practice, we will eventually study approximations to the ideal LPF that can actually be built.

3.5 Downsampling

Let $x[n]$ be a DT signal and $N > 0$ an integer, and define

$$x_d[n] = x[nN].$$

$x_d[n]$ is obtained from $x[n]$ be selecting every N values and discarding the rest. Hence, we may view downsampling as "sampling a DT signal". If $x[n]$ was obtained by sampling a CT signal $x(t)$ with period T, then

$$x_d[n] = x[nN] = x(nNT)$$

corresponds to sampling $x(t)$ at a lower rate with period NT.

Downsampling leads to a version of the Poisson summation formula. This may be derived by mimicking the analysis leading to (3.1), but in DT:

$$x[n] = \frac{1}{2\pi}\int_0^{2\pi} X\left(e^{j\Omega}\right)e^{j\Omega n}d\Omega$$

$$\begin{aligned}
x_d[n] &= x(nN) \\
&= \frac{1}{2\pi}\int_0^{2\pi} X\left(e^{j\Omega}\right)e^{j\Omega nN}d\Omega \\
&= \frac{1}{2\pi}\sum_{k=0}^{N-1}\int_{\frac{2\pi}{N}k}^{\frac{2\pi}{N}(k+1)} X\left(e^{j\Omega}\right)e^{j\Omega nN}d\Omega \\
&= \frac{1}{2\pi N}\sum_{k=0}^{N-1}\int_0^{2\pi} X\left(e^{j\frac{\theta+2\pi k}{N}}\right)e^{j(\theta+2\pi k)n}d\theta \quad (\theta = \Omega N - 2\pi k) \\
&= \frac{1}{2\pi}\int_0^{2\pi}\left(\frac{1}{N}\sum_{k=0}^{N-1} X\left(e^{j\frac{\theta+2\pi k}{N}}\right)\right)e^{j\theta n}d\theta
\end{aligned}$$

$$X_d\left(e^{j\Omega}\right) = \frac{1}{N}\sum_{k=0}^{N-1} X\left(e^{j\frac{\Omega+2\pi k}{N}}\right). \tag{3.5}$$

Expression (3.5) is analogous to (3.1). They both state that sampling in time corresponds to adding shifts of the original transform. However, unlike sampling a CT signal, downsampling $x[n]$ results in a **finite** sum of shifts of $X\left(e^{j\Omega}\right)$.

Suppose $x[n]$ has bandwidth $\Omega_B < \pi$:

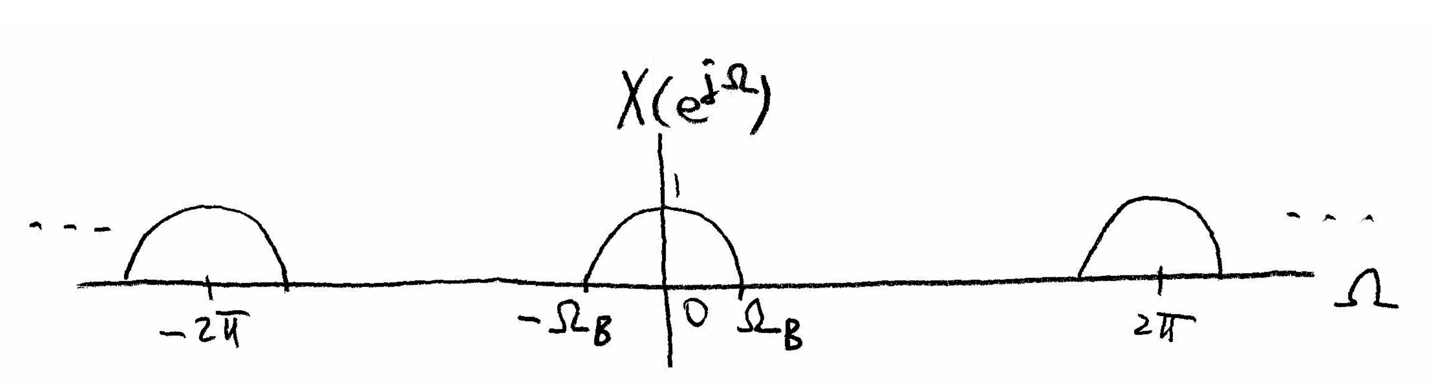


Figure 3.6

Then $x_d[n]$ has spectrum

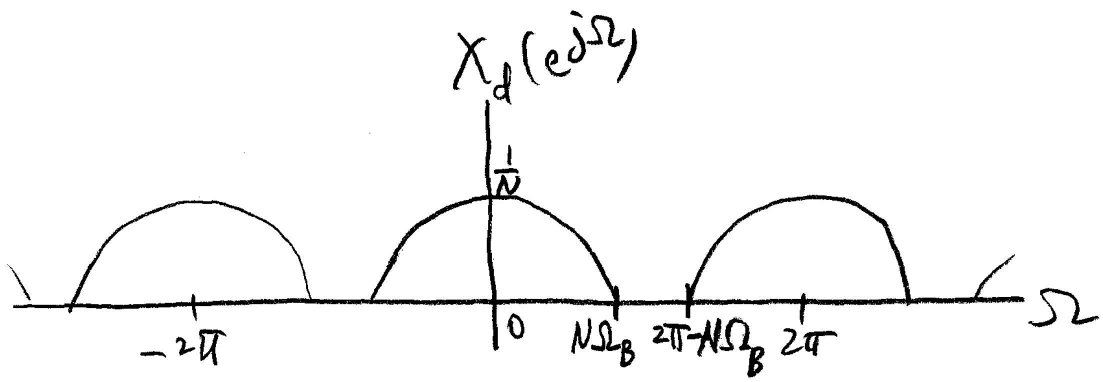


Figure 3.7

To avoid aliasing, we need

$$N\Omega_B < 2\pi - N\Omega_B$$

or

$$N < \frac{\pi}{\Omega_B}.$$

If $x[n]$ comes from sampling a CT signal $x(t)$ with bandwidth ω_B, then $\Omega_B = \omega_B T$ and we need

$$N < \frac{\pi}{\omega_B T} = \frac{\omega_s}{2\omega_B}.$$

3.6 Upsampling

For any DT signal $x[n]$, the Time/Frequency Scaling property of the DTFT states that the *expanded* signal

$$x_{(N)}[n] = \begin{cases} x\left[\frac{n}{N}\right], & \frac{n}{N} \text{ an integer} \\ 0, & \text{else} \end{cases}$$

has DTFT

$$\mathcal{F}\left\{x_{(N)}[n]\right\} = X\left(e^{j\Omega N}\right).$$

This is only one kind of *interpolation* scheme obtained by setting missing data values to 0. An alternative is to assume that $x[n]$ was obtained by sampling a CT signal $x(t)$ with bandwidth ω_B at some sampling frequency $\omega_s > 2\omega_B$. Resampling $x(t)$ at N times the original rate yields the *upsampled* signal

$$x_u[n] = x\left(n\frac{T}{N}\right). \tag{3.6}$$

We would like to develop a method for computing $x_u[n]$ directly from $x[n]$, without resorting to explicit construction of $x(t)$.

Suppose $x(t)$ has CTFT $X_{CT}(j\omega)$ as in Figure 3.2. Then from the Poisson formula (3.1),

$$\begin{aligned} \mathcal{F}\left\{x_{(N)}[n]\right\} &= X_{DT}\left(e^{j\Omega N}\right) \qquad (3.7) \\ &= \frac{1}{T}\sum_{k=-\infty}^{\infty} X_{CT}\left(j\frac{\Omega N + 2\pi k}{T}\right) \\ &= \frac{1}{T}\sum_{k=-\infty}^{\infty} X_{CT}\left(j\frac{\Omega + 2\pi\frac{k}{N}}{\frac{T}{N}}\right). \end{aligned}$$

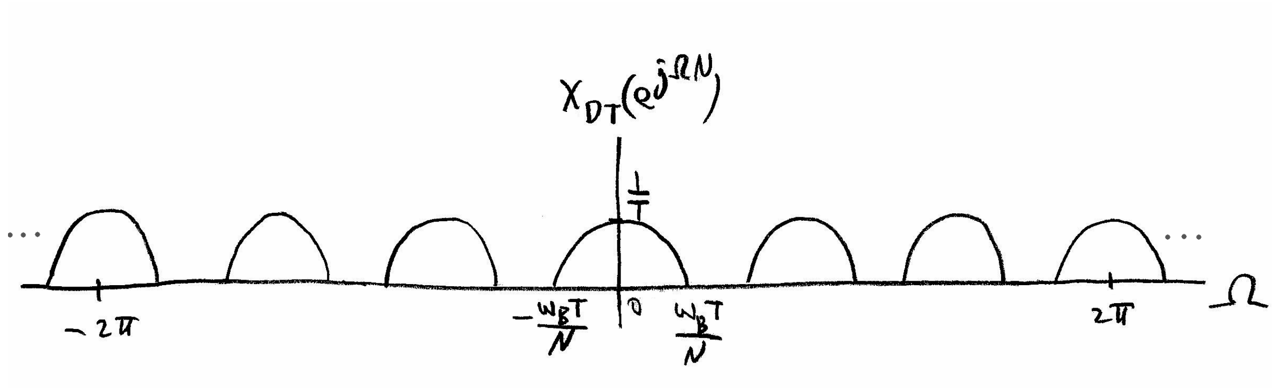


Figure 3.8

and

$$\mathcal{F}\left\{x_u\left[n\right]\right\} = X_u\left(e^{j\Omega}\right) = \frac{N}{T}\sum_{k=-\infty}^{\infty} X_{CT}\left(j\frac{\Omega + 2\pi k}{\frac{T}{N}}\right) \tag{3.8}$$

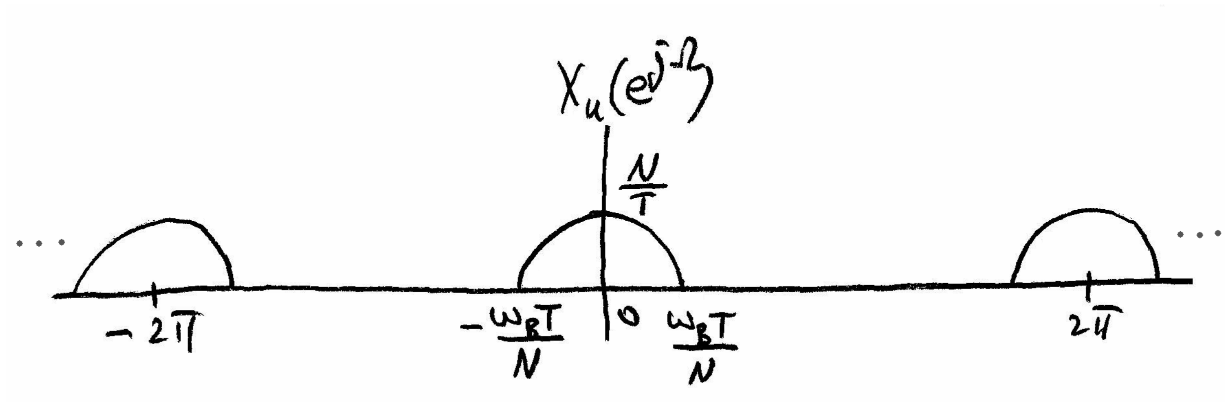


Figure 3.9

Note that (3.8) is a scaled version of (3.7), but containing only 1 out of every N terms.

Define the *ideal DT low-pass filter*:

$$H_{LP}\left(e^{j\Omega}\right) = \begin{cases} 1, & |\Omega| \leq \frac{\pi}{N} \\ 0, & \frac{\pi}{N} < |\Omega| \leq \pi \end{cases} \qquad \text{(period } 2\pi\text{)}$$

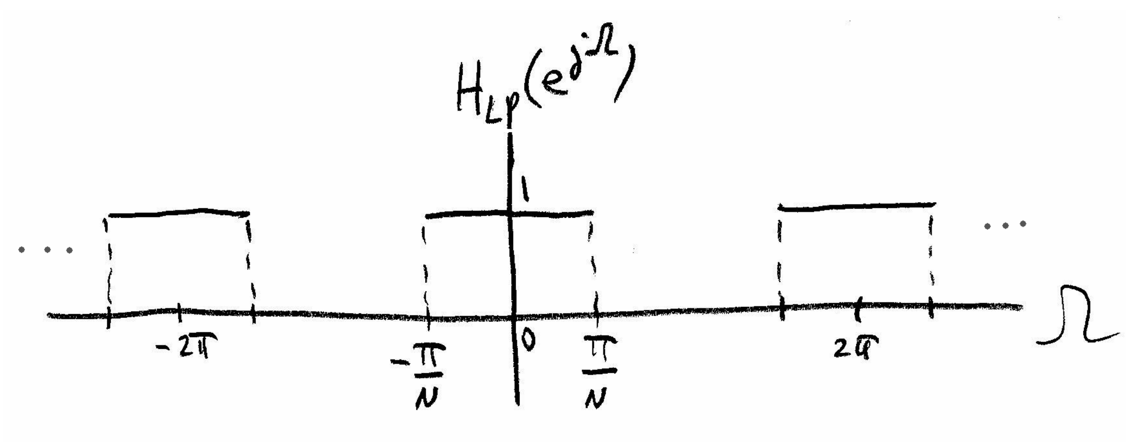


Figure 3.10

Passing from (3.7) to (3.8) is the same as applying $NH_{LP}\left(e^{j\Omega}\right)$ to $X_{DT}\left(e^{j\Omega N}\right)$:

$$X_u\left(e^{j\Omega}\right) = NH_{LP}\left(e^{j\Omega}\right) X_{DT}\left(e^{j\Omega N}\right)$$

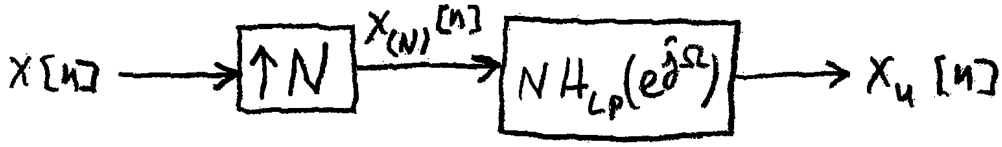


Figure 3.11

Upsampling from $x[n]$ to $x_u[n]$ amounts to Figure 3.11, where the first block indicates expansion by N. As with the ideal CT LPF, the ideal DT LPF is not realizable in practice. However, close approximations are achievable.

3.7 Change of Sampling Frequency

More generally, suppose we are given a DT signal $x[n]$ and we wish to replace every N consecutive values by M without changing the "character" of the signal. A common application of this idea is that of resampling a CT signal: Suppose $x[n]$ consists of samples of a CT signal $x(t)$ at frequency $\omega_s > 2\omega_B$. We may wish to resample $x(t)$ at a new rate $r\omega_s$, where r is a rational number. The number r must be chosen to avoid aliasing – i.e. $r\omega_s > 2\omega_B$.

Resampling can be achieved through upsampling and downsampling. Write

$$r = \frac{M}{N},$$

where M and N are coprime integers. Upsampling by M, we obtain

$$x_u[n] = x\left(n\frac{T}{M}\right).$$

Downsampling by N yields

$$x_r[n] = x_u[nN] = x\left(nN\frac{T}{M}\right) = x\left(n\frac{T}{r}\right).$$

The block diagram in Figure 3.12 depicts the process, where the last block indicates downsampling by N.

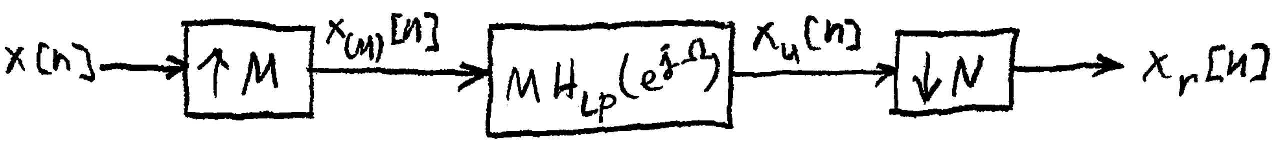


Figure 3.12

Examination of Figures 3.6-3.10 shows that no aliasing occurs at any step. The final spectrum is shown in Figure 3.13:

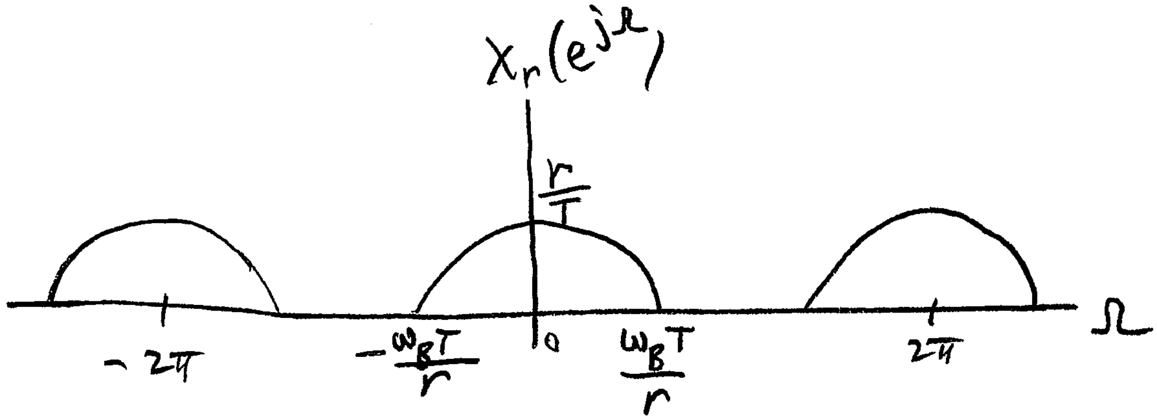


Figure 3.13

Obviously, if $r = \frac{M}{1}$ (i.e. an integer), then downsampling by $N = 1$ is unnecessary. If $r = \frac{1}{N}$, then upsampling by M is unnecessary.

4 CT Signal Reconstruction

4.1 Hybrid Systems

A system that takes a DT input $x[n]$ into a CT output $y(t)$ is called a *hybrid system.* (Actually, CT $\rightarrow$ DT is also hybrid, but we will not pursue this idea.) The definitions of system properties such as linearity, causality, and BIBO stability carry over word-for-word:

Linearity:

$$\begin{cases} \alpha x[n] \rightarrow \alpha y(t) \\ x_1[n] + x_1[n] \rightarrow y_1(t) + y_2(t) \end{cases}$$

BIBO Stability:

$$x[n] \quad \text{bounded} \Longrightarrow y(t) \quad \text{bounded}$$

We also need to define the concepts of time-invariance and causality. This must be done relative to a sampling period $T > 0$:

Time-Invariance:

$$x[n-m] \rightarrow y(t-mT) \quad \text{for every integer } m$$

Causality:

$$x_1[n] = x_2[n] \quad \text{for } n < m \Longrightarrow y_1(t) = y_2(t) \quad \text{for } t < mT$$

This will allow us to use hybrid system theory to analyze sampling and signal reconstruction for any given sampling rate f_s. Note that a system that is time-invariant relative to one value of T generally will not be time-invariant relative to other values.

In the case of linear, time-invariant (LTI) hybrid systems, the concept of impulse response is useful: The *impulse response* $h(t)$ is the output generated by the input $\delta[n]$. Exploiting linearity, time-invariance, and the DT sifting property, we obtain the following statements about how the system processes signals:

$$\delta[n] \rightarrow h(t)$$

$$\delta[n-m] \rightarrow h(t-mT)$$

$$x[m]\,\delta[n-m] \rightarrow x[m]\,h(t-mT)$$

$$\sum_{m=-\infty}^{\infty} x[m]\,\delta[n-m] \to \sum_{m=-\infty}^{\infty} x[m]\,h(t-mT)$$

$$x[n] \to \sum_{m=-\infty}^{\infty} h(t-mT)\,x[m] \tag{4.1}$$

The sum on the right side of (4.1) is *hybrid convolution*:

$$h(t) * x[n] = \sum_{m=-\infty}^{\infty} h(t-mT)\,x[m]\,. \tag{4.2}$$

As with purely CT or DT systems, a hybrid LTI system convolves the input with its impulse response.

We define the *transfer function* of a hybrid system with impulse response $h(t)$ to be

$$H(j\omega) = \mathcal{F}\{h(t)\}\,.$$

Then the output $y(t)$ has Fourier transform

$$\begin{aligned} Y(j\omega) &= \int_{-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h(t-mT)\,x[m]\,e^{-j\omega t}dt \\ &= \sum_{m=-\infty}^{\infty} x[m] \int_{-\infty}^{\infty} h(t-mT)\,e^{-j\omega t}dt \\ &= \sum_{m=-\infty}^{\infty} x[m] \int_{-\infty}^{\infty} h(\tau)\,e^{-j\omega(\tau+mT)}dt \quad (\tau = t-mT) \\ &= \left(\int_{-\infty}^{\infty} h(\tau)\,e^{-j\omega\tau}dt\right)\left(\sum_{m=-\infty}^{\infty} x[m]\,e^{-j(\omega T)m}\right) \\ &= H(j\omega)\,X\left(e^{j\omega T}\right), \end{aligned} \tag{4.3}$$

so $H(j\omega)$ may be viewed as the ratio of the Fourier transforms of the input and output. Equation (4.3) is just the hybrid version of the convolution property of the Fourier transform:

$$h(t) * x[n] \longleftrightarrow H(j\omega)\,X\left(e^{j\omega T}\right)$$

Note that here again we encounter the frequency equivalence $\Omega = \omega T$ as in (3.2).

Example 4.1 *For any DT signal $x[n]$, we define*

$$x_T(t) = \sum_{n=-\infty}^{\infty} x[n]\,\delta(t-nT)\,. \tag{4.4}$$

If $x[n]$ is obtained by sampling a CT signal $x(t)$ with period T, then $x_T(t)$ is called the "impulse sampled" signal corresponding to $x(t)$. The map from $x[n]$ to $x_T(t)$ is a hybrid system, which is easily shown to be LTI. The corresponding impulse response is

$$h_T(t) = \sum_{n=-\infty}^{\infty} \delta[n]\,\delta(t-nT) = \delta(t)\,.$$

The transfer function is

$$H_T(j\omega) = \mathcal{F}\{\delta(t)\} = 1.$$

Thus

$$X_T(j\omega) = H_T(j\omega)\, X\left(e^{j\omega T}\right) = X\left(e^{j\omega T}\right). \tag{4.5}$$

$X_T(j\omega)$ is just a copy of $X\left(e^{j\Omega}\right)$, but with the frequency axis scaled by T.

Consider a causal LTI hybrid system. Setting $\alpha = 0$ in the definition of linearity, we obtain $0 \to 0$. Since $\delta[n] = 0$ for $n < 0$, the definition of causality says that $h(t) = 0$ for $t < 0$. Conversely, suppose $h(t) = 0$ for $t < 0$. If $x_1[n] = x_2[n]$ for $n < m$, then

$$\begin{aligned} y_1(t) - y_2(t) &= h(t) * (x_1[n] - x_2[n]) \\ &= \sum_{l=-\infty}^{\infty} h(t - lT)(x_1[l] - x_2[l]) \\ &= \sum_{l=m}^{\infty} h(t - lT)(x_1[l] - x_2[l]). \end{aligned}$$

But

$$h(t - lT) = 0 \quad \text{for } t < lT,$$

so

$$y_1(t) - y_2(t) = 0 \quad \text{for } t < mT,$$

proving causality. Thus an LTI hybrid system is causal iff $h(t) = 0$ for $t < 0$.

Now consider BIBO stability. It turns out that an LTI hybrid system is BIBO stable iff its impulse response $h(t)$ satisfies

$$\max_t \sum_{n=-\infty}^{\infty} |h(t - nT)| < \infty. \tag{4.6}$$

To show that (4.6) implies stability, let $|x[n]| \leq M_1$ for every n and

$$M_2 = \max_t \sum_{n=-\infty}^{\infty} |h(t - nT)|.$$

Define the *floor* function

$$\lfloor t \rfloor = \text{largest integer} \leq t.$$

Invoking the triangle inequality,

$$\begin{aligned} |y(t)| &= \left| \sum_{n=-\infty}^{\infty} h(t - nT)\, x[n] \right| \\ &\leq \sum_{n=-\infty}^{\infty} |h(t - nT)|\, |x[n]| \\ &\leq M_1 \sum_{n=-\infty}^{\infty} |h(t - nT)| \\ &= M_1 \sum_{n=-\infty}^{\infty} |h(t - nT)| \quad \left(\tau = t - \left\lfloor \frac{t}{T} \right\rfloor T, \quad m = n - \left\lfloor \frac{t}{T} \right\rfloor\right) \\ &= M_1 \sum_{m=-\infty}^{\infty} |h(\tau - mT)|. \end{aligned}$$

But $0 \leq \tau < T$, so

$$|y(t)| \leq M_1 M_2.$$

In other words, bounded inputs produce bounded outputs, so the system is BIBO stable. One may also prove the converse – i.e. that stability implies (4.6). However, this is **far** more difficult.

Hybrid convolution shares most of the basic algebraic properties of CT and DT convolution. Unfortunately, hybrid convolution is not commutative, since one signal in the convolution (4.2) must be CT and the other must be DT. It is easy to see that hybrid convolution is distributive:

$$(h_1(t) + h_2(t)) * x[n] = (h_1(t) * x[n]) + (h_2(t) * x[n])$$

This tells us that the impulse response of a parallel combination of hybrid systems is just the sum of the individual impulse responses.

Hybrid convolution is also associative:

$$\begin{aligned}
h_2(t) * (h_1(t) * x[n]) &= \int_{-\infty}^{\infty} \left(h_2(t-\tau) \sum_{n=-\infty}^{\infty} h_1(\tau - nT)\, x[n] \right) d\tau \\
&= \sum_{n=-\infty}^{\infty} \left(\int_{-\infty}^{\infty} h_2(t-\tau)\, h_1(\tau - nT)\, d\tau \right) x[n] \\
&= \sum_{n=-\infty}^{\infty} \left(\int_{-\infty}^{\infty} h_2((t-nT) - \mu)\, h_1(\mu)\, d\mu \right) x[n] \qquad (\mu = \tau - nT) \\
&= \left(\int_{-\infty}^{\infty} h_2(t-\mu)\, h_1(\mu)\, d\mu \right) * x[n] \\
&= (h_2(t) * h_1(t)) * x[n]
\end{aligned}$$

Associativity corresponds to systems in series. Actually, connecting hybrid systems in series makes no sense, since the output of the first is CT and input of the second is DT. However, we can connect a CT system to the output of a hybrid system:

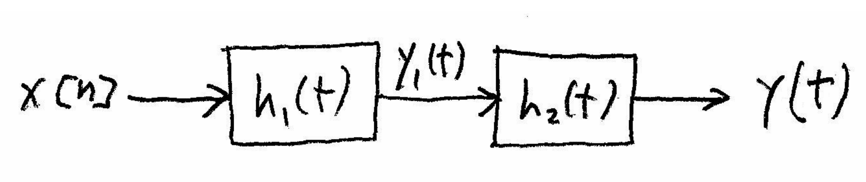


Figure 4.1

Associativity tells us that the series impulse response is the convolution of the individual impulse responses:

$$h(t) = h_2(t) * h_2(t).$$

Thus the series transfer function is

$$H(j\omega) = H_2(j\omega)\, H_1(j\omega).$$

Suppose we place two causal systems in series as in Figure 4.1. Then $h_1(t) = h_2(t)$ for $t < 0$, so the convolution $h(t)$ inherits the same property and the combines system is causal. Now suppose we place two BIBO stable systems in series. If $x[n]$ is bounded, then stability of the first system implies boundedness of $y_1(t)$. Stability of the second system in turn implies boundedness of $y(t)$.

Hence, the composite system is BIBO stable. Recall that a CT system with impulse response $h(t)$ is BIBO stable iff

$$\int_{-\infty}^{\infty} |h(t)|\, dt < \infty. \tag{4.7}$$

So, if $h_1(t)$ satisfies (4.6) and $h_2(t)$ satisfies (4.7), then $h_2(t) * h_1(t)$ satisfies (4.6). One may derive similar results when a DT system is followed by a hybrid system.

4.2 Ideal Signal Reconstruction

CT signal reconstruction of $x(t)$ from its samples $x[n]$ is a fundamental problem in DSP. This process must take place whenever information from a digital system (e.g. a computer, CD player, voice synthesizer) is converted to a usable form in the analog world. We will first study the problem of exactly reproducing $x(t)$.

We begin the reconstruction by converting $x[n]$ to its impulse sampled version $x_T(t)$. (See Example 4.1.). This process amounts to passing $x[n]$ through the hybrid system with transfer function $H_T(j\omega) = 1$. Now pass $x_T(t)$ through the ideal LPF with transfer function

$$H_r(j\omega) = TH_{LP}\left(j\frac{2\omega}{\omega_s}\right).$$

Assuming $\omega_s > 2\omega_B$, the LPF "selects" the lowest frequency lobe of $X_T(j\omega)$, yielding $X_{CT}(j\omega)$. Formally,

$$X_{CT}(j\omega) = TH_{LP}\left(j\frac{2\omega}{\omega_s}\right) H_T(j\omega)\, X\left(e^{j\omega T}\right) = H_r(j\omega)\, X_T(j\omega).$$

Comparing Figures 3.1, 4.1, and 4.3 illustrates the idea.

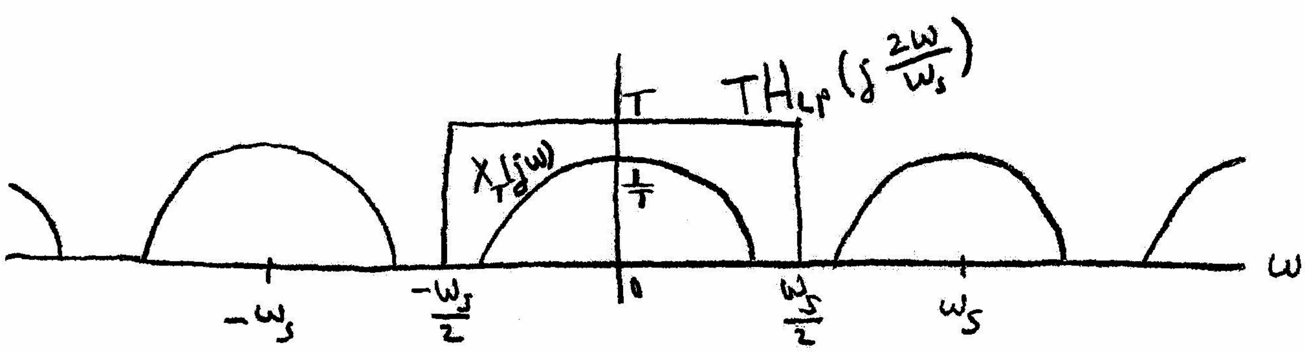


Figure 4.2

Used in this way, an LPF is a *reconstruction filter.*

We may also analyze this process in time-domain. Recall from Example 4.1 that the impulse sampler amounts to the convolution

$$x_T(t) = \delta(t) * x[n].$$

The LPF has impulse response

$$\begin{aligned} h_r(t) &= \mathcal{F}^{-1}\{H_r(j\omega)\} \\ &= \pi h_{LP}\left(\frac{\omega_s}{2}t\right) \\ &= \operatorname{sinc}\frac{t}{T}. \end{aligned}$$

Connecting these two systems in series yields the impulse response

$$h_r(t) * \delta(t) = h_r(t),$$

so the output is

$$x(t) = h_r(t) * x[n] = \sum_{n=-\infty}^{\infty} x[n] \operatorname{sinc}\frac{t-nT}{T}. \tag{4.8}$$

The reconstruction of $x(t)$ is illustrated in Figure 4.4:

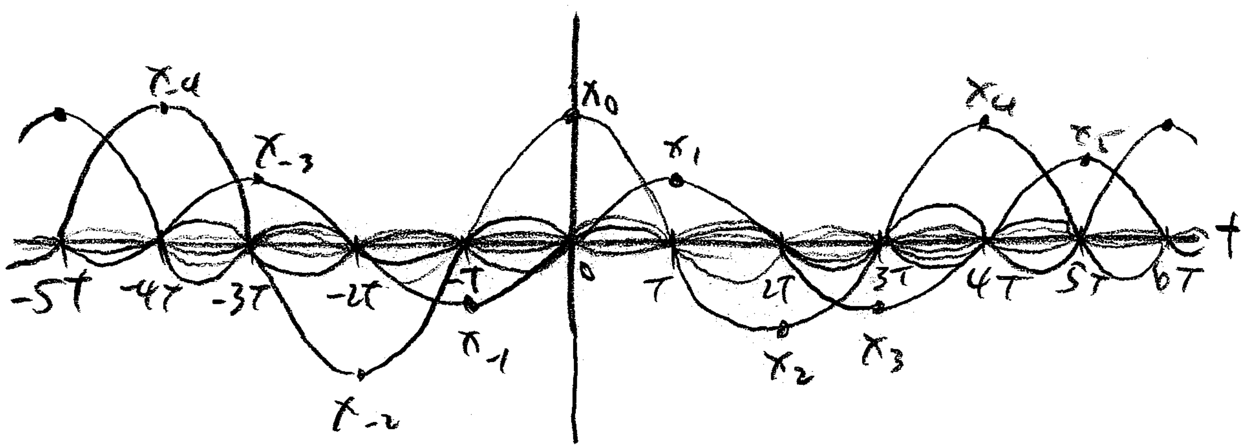


Figure 4.3

The interplay between sampling and the Fourier transform is summarized in Figure 4.5:

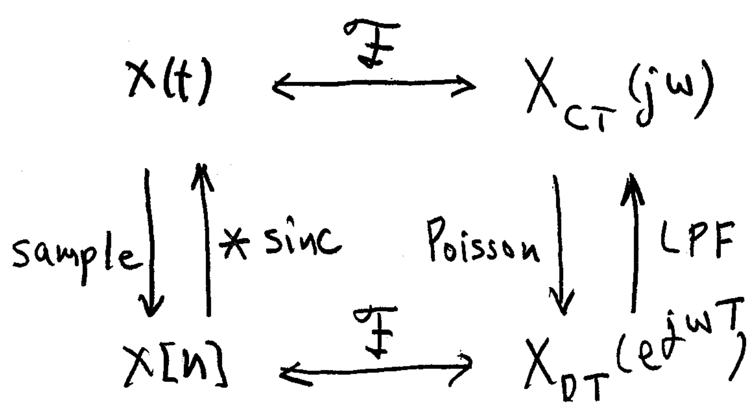


Figure 4.4

Unfortunately, there are a couple of problems associated with ideal reconstruction. First note that $h_r(t) \neq 0$ for most $t < 0$, so the system is **noncausal**. Also,

$$\begin{aligned}
\max_t \sum_{n=-\infty}^{\infty} |h_r(t - nT)| &= \max_t \sum_{n=-\infty}^{\infty} \left| \operatorname{sinc} \frac{t - nT}{T} \right| \\
&\geq \sum_{n=-\infty}^{\infty} \left| \operatorname{sinc} \left(\frac{1}{2} - n \right) \right| \quad (t = \frac{T}{2}) \\
&= \sum_{n=-\infty}^{\infty} \left| \operatorname{sinc} \left(n - \frac{1}{2} \right) \right| \\
&= 2 \sum_{n=1}^{\infty} \left| \operatorname{sinc} \left(n - \frac{1}{2} \right) \right| \\
&= \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{\left| \sin \pi \left(n - \frac{1}{2} \right) \right|}{n - \frac{1}{2}} \\
&= \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{1}{n - \frac{1}{2}} \\
&\geq \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{1}{n},
\end{aligned} \tag{4.9}$$

so the system is **not BIBO stable**. Thus there exists a bounded input $x[n]$ that produces an unbounded output $x(t)$. Note that, if we apply the scaled input $\varepsilon x[n]$ even for small $\varepsilon \neq 0$, linearity guarantees that the output $\varepsilon x(t)$ is still unbounded. This shows that even the tiniest amount of random noise in the system has the tendency to produce extremely large error in the reconstruction, making ideal reconstruction unachievable in practice. Nevertheless, we will see that it is possible to closely approximate the ideal.

4.3 The Zero-Order Hold

Define the *Zero-Order Hold (ZOH)* to be the hybrid system that takes $x[n]$ into the CT signal $x(t)$ defined by

$$x(t) = x[n], \quad nT \leq t < (n+1)T. \tag{4.10}$$

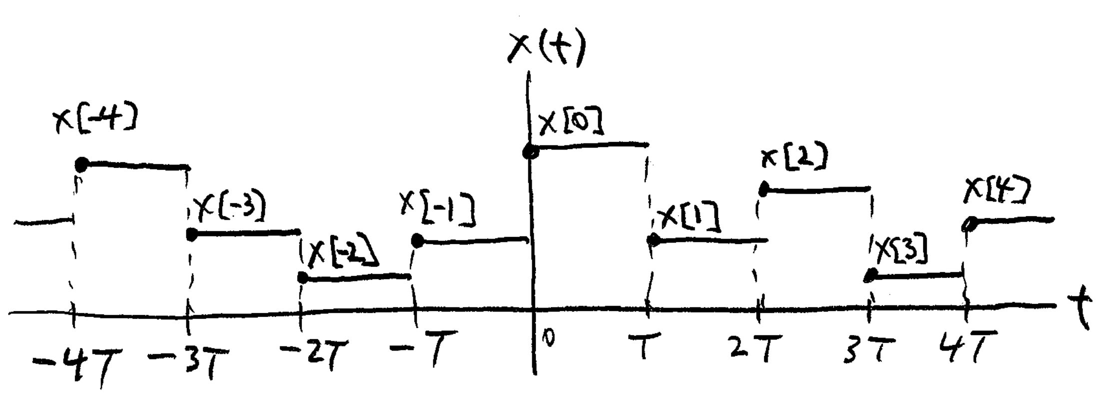


Figure 4.5

In other words, each input value is held throughout the subsequent sampling interval.

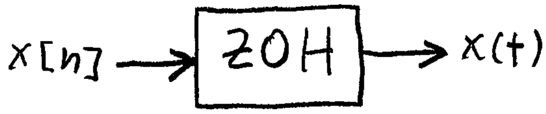


Figure 4.6

Consider the *unit rectangular pulse*

$$w(t) = \begin{cases} 1, & 0 \le t \le 1 \\ 0, & \text{else} \end{cases}.$$

The ZOH processes signals according to

$$x_0(t) = \sum_{n=-\infty}^{\infty} x[n]\, w\left(\frac{t-nT}{T}\right) = w\left(\frac{t}{T}\right) * x[n],$$

so the ZOH is an LTI hybrid system with impulse response

$$h_0(t) = w\left(\frac{t}{T}\right) = \begin{cases} 1, & 0 \le t < T \\ 0, & \text{else} \end{cases}. \tag{4.11}$$

The Fourier transform of $w(t)$ is

$$\begin{aligned} W(j\omega) &= \int_{-\infty}^{\infty} w(t)\, e^{-j\omega t} dt \\ &= \int_{0}^{1} e^{-j\omega t} dt \\ &= \frac{1-e^{-j\omega}}{j\omega} \\ &= \frac{e^{j\frac{\omega}{2}} - e^{-j\frac{\omega}{2}}}{j\omega} e^{-j\frac{\omega}{2}} \\ &= \frac{2\sin\frac{\omega}{2}}{\omega} e^{-j\frac{\omega}{2}} \\ &= \left(\operatorname{sinc}\frac{\omega}{2\pi}\right) e^{-j\frac{\omega}{2}}, \end{aligned}$$

so the ZOH has transfer function

$$H_0(j\omega) = \mathcal{F}\left\{w\left(\frac{t}{T}\right)\right\} = TW(j\omega T) = T\left(\operatorname{sinc}\frac{\omega}{\omega_s}\right) e^{-j\frac{\omega T}{2}}.$$

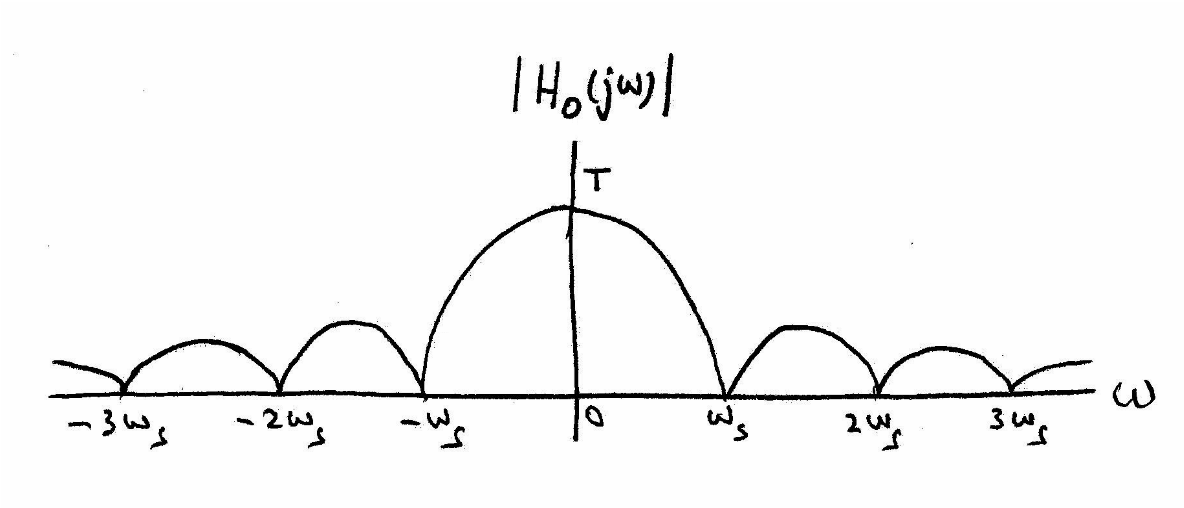


Figure 4.7

The output $x_0(t)$ of a ZOH has Fourier transform satisfying

$$X_0(j\omega) = H_0(j\omega)\, X_{DT}\left(e^{j\omega T}\right) = \frac{1}{T} H_0(j\omega) \sum_{k=-\infty}^{\infty} X_{CT}\left(j\left(\omega + k\omega_s\right)\right).$$

$|X_0(j\omega)|$ is depicted in Figure 4.8, obtained by combining Figures 3.2 and 4.7:

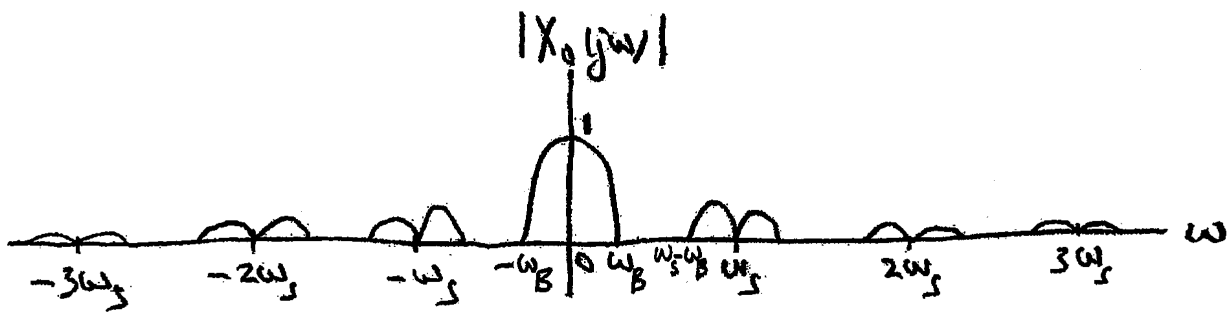


Figure 4.8

Keeping in mind that the bandwidth ω_B is fixed, we may consider the effect of letting the sampling rate f_s become large (i.e. T becomes small). As $f_s \to \infty$, all but the central lobe of $X_T(j\omega) = X_{DT}\left(e^{j\omega T}\right)$ slides off the picture. Furthermore, $H_0(j\omega) \to T$ for $-\omega_B \leq \omega \leq \omega_B$ so the central lobe is multiplied by a transfer function close to unity. In this sense, $X_0(j\omega) \approx X_{CT}(j\omega)$ for large f_s. This fact can also be seen in the time domain:

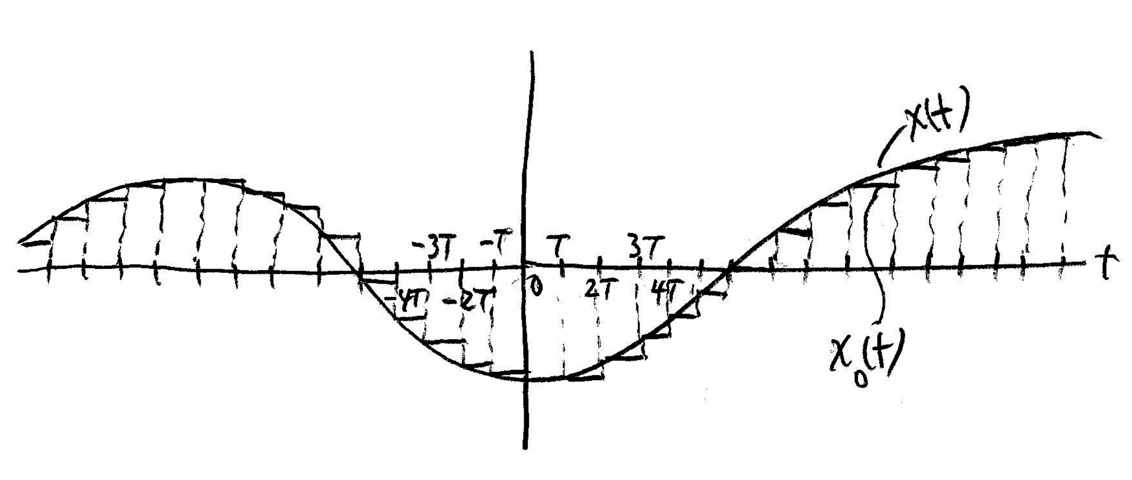


Figure 4.9

As $T \to 0$ the "staircase" function $x_0(t)$ produced by the ZOH converges to $x(t)$.

The point of using a ZOH is that **it can actually be built**. This is true because a unit rectangular pulse (or at least a very close approximation) can be produced using conventional electronics. Try building an electric circuit that generates anything close to an impulse! The ZOH alone does not perform ideal reconstruction, but only a close approximation. For many applications (e.g. digital control systems), this approximation is entirely adequate. For others, the distortion caused by the higher-frequency lobes shown in Figure 4.8 can create serious problems. For example, in audio systems these lobes are not audible, but can damage other equipment such as the power amplifier and speakers. Another consideration is that, in the interest of economy, we usually prefer to use the smallest sampling rate possible, so an extremely close approximation in Figure 4.9 may be wasteful. For these reasons, the higher-frequency lobes of $x_0(t)$ are often filtered out using an LPF. In this context, the LPF is sometimes referred to as a "smoothing" filter.

If $\omega_s > 2\omega_B$, passing the output of the ZOH through the filter with transfer function

$$\begin{aligned} H_{0r}(j\omega) &= \frac{TH_{LP}\left(j\frac{2\omega}{\omega_s}\right)}{H_0(j\omega)} \qquad (4.12) \\ &= \frac{H_{LP}\left(j\frac{2\omega}{\omega_s}\right)}{W(j\omega T)} \\ &= \begin{cases} \frac{e^{j\frac{\omega T}{2}}}{\operatorname{sinc}\frac{\omega}{\omega_s}}, & |\omega| \leq \frac{\omega_s}{2} \\ 0, & \text{else} \end{cases} \end{aligned}$$

yields the output

$$\begin{aligned} Y(j\omega) &= H_{0r}(j\omega) H_0(j\omega) X_{DT}\left(e^{j\omega T}\right) \\ &= H_{LP}\left(j\frac{2\omega}{\omega_s}\right) \sum_{k=-\infty}^{\infty} X_{CT}(j(\omega + k\omega_s)) \\ &= X_{CT}(j\omega). \end{aligned}$$

Hence, we may view ideal reconstruction as the series combination of the ZOH with the filter having transfer function $H_{0r}(j\omega)$.

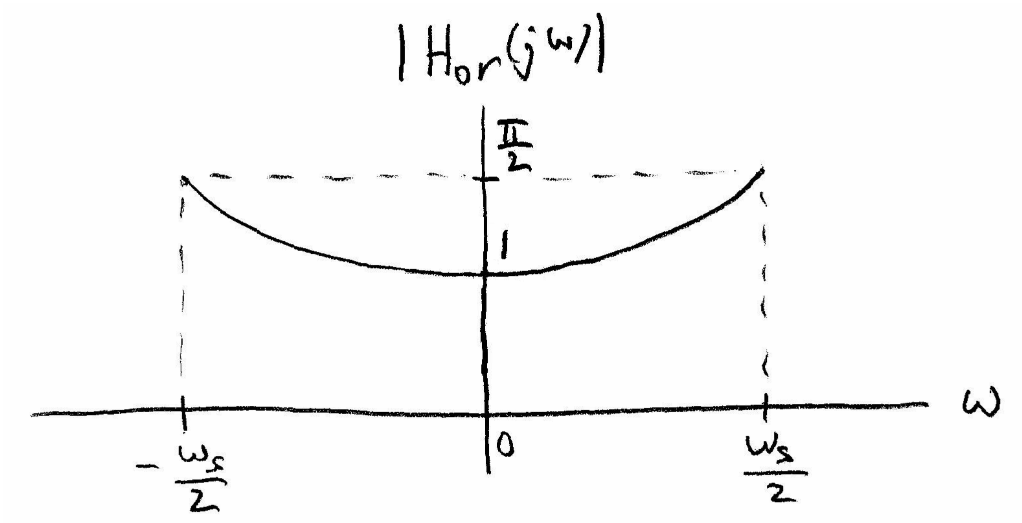


Figure 4.10

From (4.11), the ZOH is causal. Furthermore,

$$\begin{aligned}\max_t \sum_{n=-\infty}^{\infty} \left|h_0\left(t-nT\right)\right| &= \max_t \sum_{n=-\infty}^{\infty} \left|w\left(\frac{t-nT}{T}\right)\right| \\ &= \max_t 1 \\ &= 1 \\ &< \infty,\end{aligned}$$

so the ZOH is BIBO stable (which is also clear from (4.10)). As with $H_r(j\omega)$, $H_{0r}(j\omega)$ determines a noncausal, unstable filter. However, we will see that $H_{0r}(j\omega)$ can be approximated by a stable filter. Hence, the reconstruction method of choice in practice is to process samples through a ZOH and then, if necessary, through a smoothing filter approximating $H_{0r}(j\omega)$.

4.4 A/D and D/A Converters

The actual electronic device that performs sampling is called an analog-to-digital (A/D) converter. An A/D converter actually consists of two parts: First, the CT input is sampled. Then the resulting value (a voltage) is *quantized* – i.e. approximated by the nearest value taken from a given finite set. Quantization is depicted in Figure 4.11:

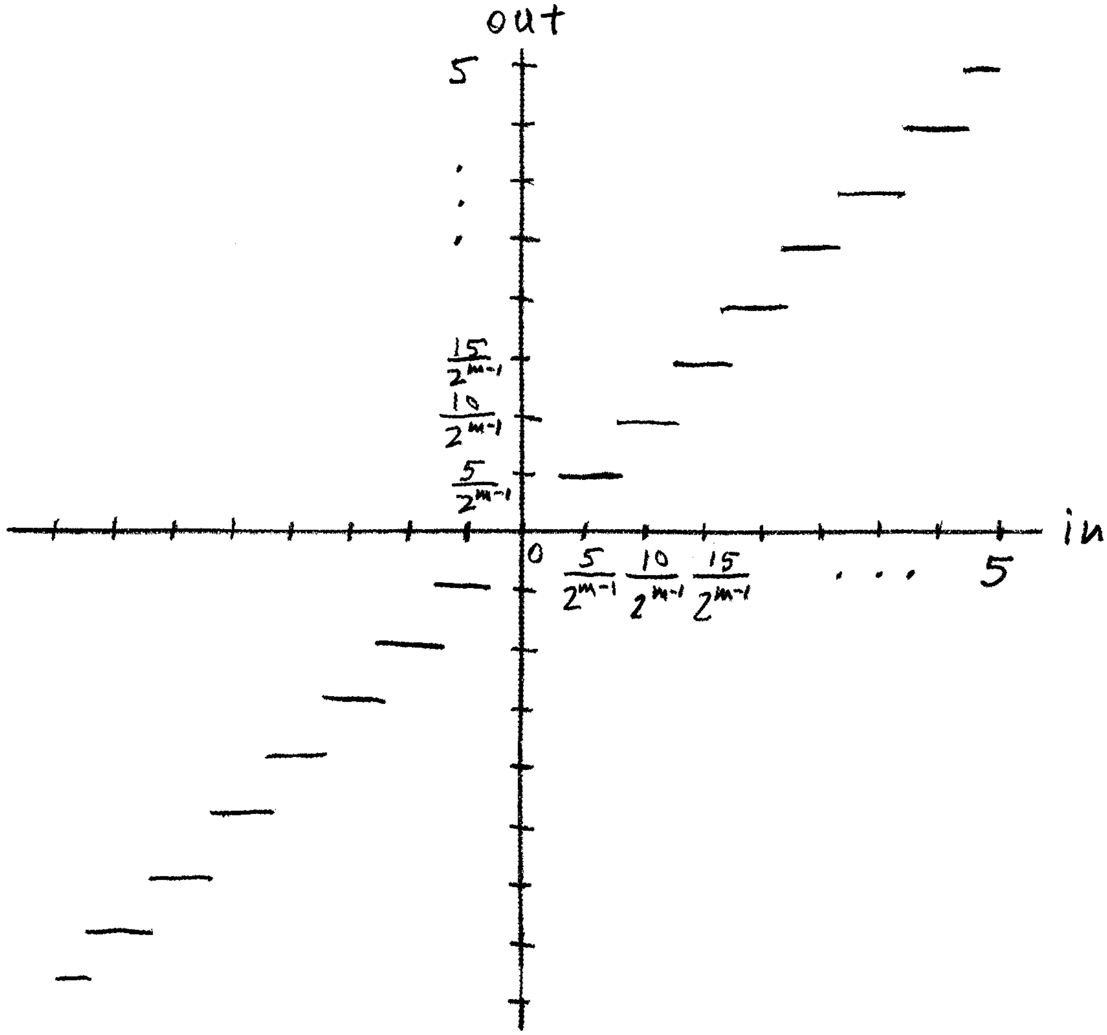


Figure 4.11

Here we assume the input voltage stays within ± 5V and that the output can achieve 2^m possible values. The quantized signal is then represented in binary form in an m-bit computer register. Quantization contributes a certain degree of distortion to a signal, which can be made smaller by increasing m. In practice, the value of m depends on the application. In control systems, $m = 12$ is common, while in high-fidelity audio systems, $m = 16$ is minimum. The distortion caused by the nonlinear nature of Figure 4.11 is called *quantization noise*. An analytic treatment of quantization noise is possible, but it is mathematically difficult and requires the study of random processes. This is beyond our scope, so we will henceforth assume that the number of bits m is sufficiently large to ensure that the effects of quantization are negligible. In other words, we will approximate the A/D converter as an *ideal sampler*, depicted in Figure 3.1.

A D/A converter performs signal reconstruction. The m-bit binary value is converted back to a voltage with 2^m possible values and then passed through a ZOH. A smoothing filter may then be applied to the output. The only idealization required here is that m be very large, so the ZOH can accept any DT signal $x[n]$.

4.5 Digital Filters

At this point, we can pull together several concepts we have already studied and describe the general framework in which digital filtering may be carried out. Consider a series combination of a sampler, DT system with transfer function $H_{DT}\left(e^{j\Omega}\right)$, and ideal reconstruction device:

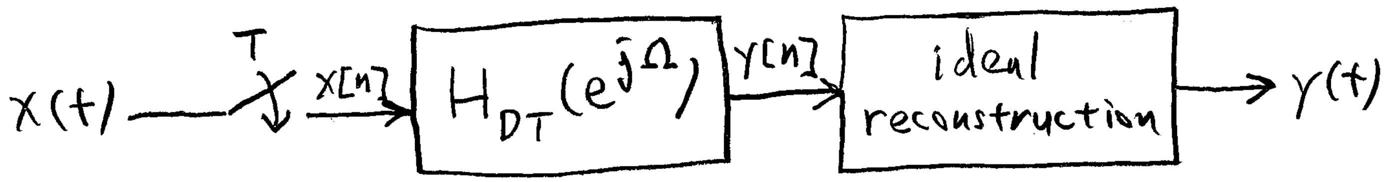


Figure 4.12

We assume that all input signals have bandwidth less than $\frac{\omega_s}{2}$ to avoid aliasing. The sampled input $x[n]$ has DTFT $X_{DT}\left(e^{j\Omega}\right)$ given by the Poisson formula (3.1). The DT output $y[n]$ thus has DTFT

$$Y_{DT}\left(e^{j\Omega}\right) = H_{DT}\left(e^{j\Omega}\right) X_{DT}\left(e^{j\Omega}\right) = \frac{1}{T}\sum_{k=-\infty}^{\infty} H_{DT}\left(e^{j\Omega}\right) X_{CT}\left(j\frac{\Omega+2\pi k}{T}\right). \tag{4.13}$$

As discussed above, ideal reconstruction may be envisioned as either 1) conversion of $y[n]$ to the impulse train $y_T(t)$ followed by an ideal LPF or 2) passing $y[n]$ through a ZOH and then through the filter $H_{0r}(j\omega)$. Both have the same effect, isolating the $k=0$ term on the right side of (4.13) as in Figure 4.2. This yields the CT output $y(t)$ with FT

$$Y_{CT}(j\omega) = TH_{LP}\left(j\frac{2\omega}{\omega_s}\right) Y_{DT}\left(e^{j\omega T}\right) = H_{DT}\left(e^{j\omega T}\right) X_{CT}(j\omega).$$

The system in Figure 4.12 therefore has frequency response

$$H_{CT}(j\omega) = H_{DT}\left(e^{j\omega T}\right)$$

for $-\frac{\omega_s}{2} \le \omega < \frac{\omega_s}{2}$:

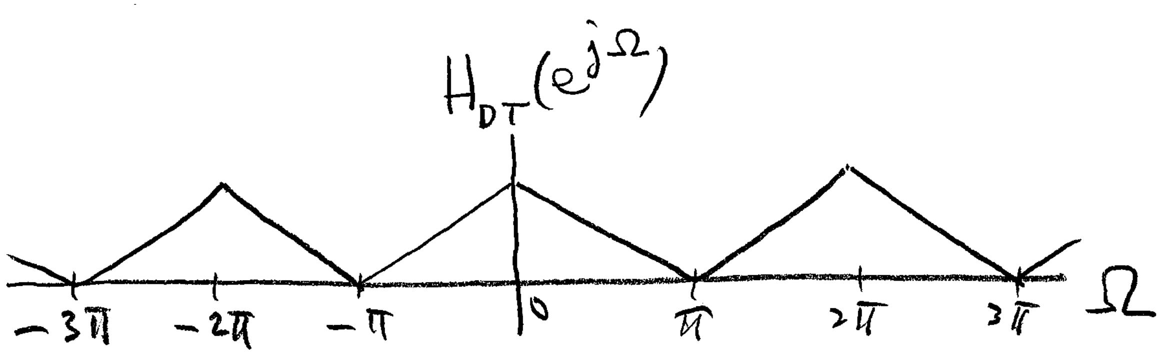


Figure 4.13

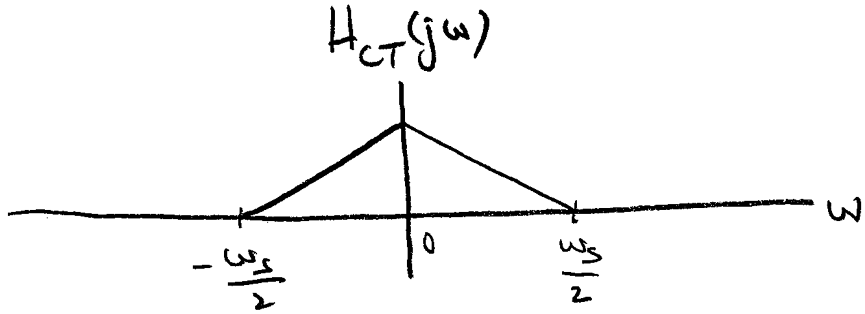


Figure 4.14

The CT system $H_{CT}(j\omega)$ is just a copy of the DT system $H_{DT}\left(e^{j\Omega}\right)$, restricted to the frequency interval $-\frac{\omega_s}{2} \leq \omega < \frac{\omega_s}{2}$, and with the familiar frequency scaling $\Omega = \omega T$.

5 The Discrete Fourier Transform

5.1 Definition and Properties

The *Discrete Fourier Transform (DFT)* is a variant of the DT Fourier series. The DFT is the only kind of Fourier transform that can actually be evaluated on a computer. Since a computer can only process a finite amount of data, we must be able to represent every time signal and transform as a finite array. This requirement does not sit well with respect to ordinary transforms: If a DT signal $x[n]$ has a DTFT concentrated at discrete frequencies, then $x[n]$ is periodic. (The transform is essentially the sequence of Fourier coefficients a_k.) In this case, the DTFT is also periodic, so neither array is finite.

Another way to look at this problem is to consider the signals $x[n]$ which can be expressed as a finite array of numbers. We say $x[n]$ is a *finite-duration signal* if there exists $N < \infty$ such that $x[n] = 0$ for $|n| > N$. Otherwise, $x[n]$ is *infinite-duration.*

Theorem 5.1 *If $x[n]$ is finite-duration and bandlimited, then $x[n] = 0$ for all n.*

(Theorem 5.1 also holds for *CT* signals.) Clearly, our notion of the Fourier transform must be modified in working with computers.

Fortunately, there is a simple trick that resolves the issue. Suppose $x[n]$ is finite-duration with

$$x[n] = 0 \quad \text{for } n < 0 \text{ and } n \geq N, \tag{5.1}$$

and consider the signal

$$x_p[n] = \sum_{m=-\infty}^{\infty} x[n - mN]. \tag{5.2}$$

Then

$$\begin{aligned} x_p[n+N] &= \sum_{m=-\infty}^{\infty} x[n - (m-1)N] \\ &= \sum_{q=-\infty}^{\infty} x[n - qN] \quad (q = m-1) \\ &= x_p[n], \end{aligned}$$

so $x_p[n]$ has period N. In other words, $x_p[n]$ extends $x[n]$ periodically. Let a_k be the sequence of DT Fourier coefficients for $x_p[n]$, also with period N. Although $x_p[n]$ is not bandlimited, we only need N values a_k to represent $x_p[n]$. **We define the DFT of $x[n]$ to be one period of** Na_k. This way, both $x[n]$ and a_k are finite arrays. Instead of Na_k, we denote the DFT by $X[k]$. Formally,

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\Omega_0 kn}, \quad 0 \leq k \leq N-1, \tag{5.3}$$

where $\Omega_0 = \frac{2\pi}{N}$ is the *fundamental frequency.* The DFS recovers $x[n]$:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{j\Omega_0 kn}, \quad 0 \leq n \leq N-1 \tag{5.4}$$

Expression (5.4) is the *Inverse Discrete Fourier Transform (IDFT)*. Note that the IDFT is just the DFT reflected and divided by N.

For $x[n]$ satisfying (5.1), the DFT and DTFT are closely related. Evaluating $X\left(e^{j\Omega}\right)$ at $\Omega = \Omega_o k$, we obtain

$$X\left(e^{j\Omega_0 k}\right) = \sum_{n=0}^{N-1} x[n] e^{-j\Omega_0 kn} = X[k].$$

Hence, the DFT is simply the array of samples of the DTFT taken with sampling period Ω_0 and indices $k = 0, \ldots, N-1$. Furthermore, expression (5.4) tells us that the N samples $X[0], \ldots, X[N-1]$ of $X\left(e^{j\Omega}\right)$ are sufficient to exactly reconstruct $x[n]$ and, therefore, $X\left(e^{j\Omega}\right)$ for all Ω. This is the dual of the Nyquist theorem, made possible by the fact that $x[n]$ is finite-duration. Specifically, $X\left(e^{j\Omega}\right)$ is recovered by the formula

$$\begin{aligned} X\left(e^{j\Omega}\right) &= \sum_{n=0}^{N-1} x[n] e^{-j\Omega n} \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \left(\sum_{k=0}^{N-1} X[k] e^{j\Omega_0 kn} \right) e^{-j\Omega n} \\ &= \frac{1}{N} \sum_{k=0}^{N-1} \left(X[k] \sum_{n=0}^{N-1} e^{-j(\Omega - \Omega_0 k)n} \right) \\ &= \frac{1}{N} \sum_{k=0}^{N-1} X[k] W_N\left(e^{j(\Omega - \Omega_0 k)}\right), \end{aligned} \tag{5.5}$$

where $W_N\left(e^{j\Omega}\right)$ is the periodic sinc function. (See Example 2.3.) Note that (5.5) is actually hybrid convolution (in frequency) and is dual to the ideal signal reconstruction formula (4.8) for $x(t)$. Ideal reconstruction of $X\left(e^{j\Omega}\right)$ is possible, since the "Nyquist rate" in this setting is N and the "sampling rate" is

$$\frac{2\pi}{\Omega_0} = N.$$

Since $x[n]$ and $X[k]$ are defined on a fixed, finite interval, an ordinary time or frequency shift is not allowed. However, there is an appropriate operation that serves the same purpose for DFT's. We need to invoke *mod N* arithmetic: For any integer n, there exists a unique integer

$$((n))_N \in \{0, \ldots, N-1\} \tag{5.6}$$

such that $n - ((n))_N$ is an integer multiple of N. In other words, an integer m can be found such that

$$((n))_N = n - mN$$

and (5.6) holds.

Properties: (See O&S Table 8.2 on p. 660.)

Linearity:

$$\begin{cases} \alpha x[n] \longleftrightarrow \alpha X[k] \\ x_1[n] + x_2[n] \longleftrightarrow X_1[k] + X_2[k] \end{cases}$$

Time Shift:

$$x[((n-n_0))_N] \longleftrightarrow e^{-j\Omega_0 n_0 k} X[k]$$

Frequency Shift:

$$e^{jin} x[n] \longleftrightarrow X[((k-i))_N]$$

Convolution:

$$x_1[n] * x_2[n] = \sum_{m=0}^{N-1} x_1[((n-m))_N] x_2[m]$$

$$x_1[n] * x_2[n] \longleftrightarrow X_1[k] X_2[k]$$

Multiplication:

$$X_1[k] * X_2[k] = \sum_{i=0}^{N-1} X_1[((k-i))_N] X_2[i]$$

$$x_1[n] x_2[n] \longleftrightarrow \frac{1}{N} X_1[k] * X_2[k]$$

Time Differencing:

$$x[n] - x[((n-1))_N] \longleftrightarrow \left(1 - e^{-j\Omega_0 k}\right) X[k]$$

Accumulation:

$$\sum_{m=0}^{n} x[m] \longleftrightarrow \frac{1}{1 - e^{-j\Omega_0 k}} X[k] \quad \text{(only for } X[0] = 0\text{)}$$

Frequency Differencing:

$$\left(1 - e^{j\Omega_0 n}\right) x[n] \longleftrightarrow X[k] - X[((k-1))_N]$$

Conjugation:

$$x^*[n] \longleftrightarrow X^*[((-k))_N]$$

Reflection:

$$x[((-n))_N] \longleftrightarrow X[((-k))_N]$$

Real Time Signal

$$x[n] \text{ real } \Longleftrightarrow \begin{cases} |X[((k))_N]| \text{ even} \\ \angle X[((k))_N] \text{ odd} \end{cases}$$

Even-Odd:

$$\begin{cases} x[((n))_N] \text{ even } \Longleftrightarrow X[k] \text{ real} \\ x[((n))_N] \text{ odd } \Longleftrightarrow X[k] \text{ imaginary} \end{cases}$$

Parseval's Theorem:

$$\frac{1}{N} \sum_{n=0}^{N-1} x_1[n] x_2^*[n] = \sum_{k=0}^{N-1} X_1[k] X_2^*[k]$$

5.2 Circular Operations

Mod N arithmetic may be used to define operations on signals that are appropriate for the DFT. The *circular reflection* of $x[n]$ is $x[((-n))_N]$. The *circular shift* of $x[n]$ by any integer n_0 is $x[((n-n_0))_N]$. From circular shift, we define the *circular convolution* of $x_1[n]$ and $x_2[n]$ is

$$x_1[n] * x_2[n] = \sum_{m=0}^{N-1} x_1[((n-m))_N]\, x_2[m].$$

Periodic and circular convolution are similar, but apply to different circumstances: Periodic convolution requires periodic signals and produces a periodic signal; circular convolution requires finite-duration signals and produces a finite-duration signal. Nevertheless,

$$x_1[n] *_{circ} x_2[n] = x_{1p}[n] *_{per} x_{2p}[n]$$

for $0 \le n \le N-1$. Circular shift and periodic convolution may also be applied to $X[k]$. Note that circular shift and convolution appear in several of the DFT properties above.

Example 5.1 *Let N be even and*

$$x[n] = \begin{cases} 1, & n = 0, \ldots, \frac{N}{2}-1 \\ 0, & \frac{N}{2} \le n \le N-1 \end{cases}.$$

Find the circular convolution of $x[n]$ with itself.

$$x[n] * x[n] = \sum_{m=0}^{N-1} x_p[n-m]\, x_p[m]$$

For $0 \le n \le \frac{N}{2}-1$,

$$x[n] * x[n] = \sum_{m=0}^{n} 1 = n+1.$$

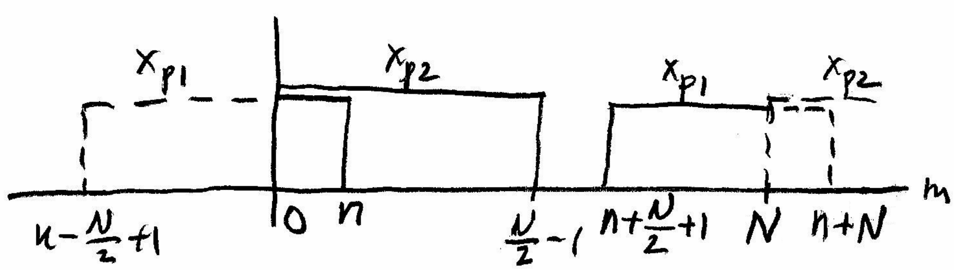


Figure 5.1

For $\frac{N}{2} \le n \le N-1$,

$$x[n] * x[n] = \sum_{m=n-\frac{N}{2}+1}^{\frac{N}{2}-1} 1 = N-1-n.$$

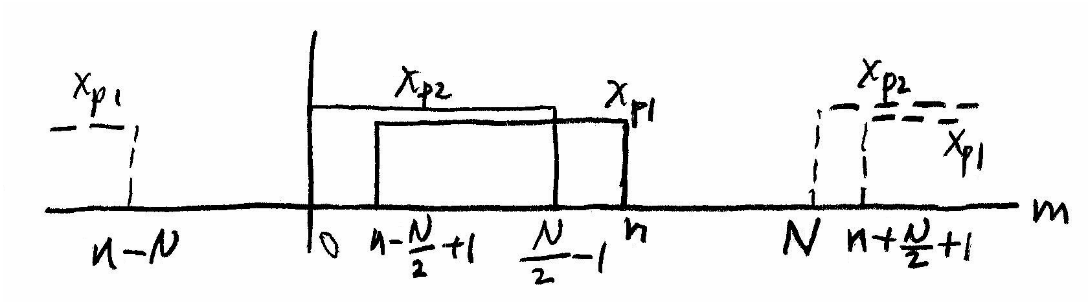


Figure 5.2

(For additional examples, see O&S pp. 655-659.)

5.3 Fast Fourier Transform Algorithms

Computational efficiency is paramount in applying the DFT (5.3) and IDFT (5.4). If we compute the DFT (5.3) in the most direct way, we are faced with multiplying the arrays

$$x = \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix}, \quad e = \begin{bmatrix} 1 \\ e^{-j\Omega_0 k} \\ \vdots \\ e^{-j\Omega_0 k(N-1)} \end{bmatrix}$$

together entry-by-entry and then adding the results to produce $X[k]$. For each k, this requires N multiplications and $N-1$ additions. To generate the DFT, this must be done N times, so N^2 multiplications and $N(N-1)$ additions are required. We say that such an algorithm requires $o(N^2)$ operations, meaning that the number of operations "increases like N^2" for large N.

There are a variety of ways to reduce the number of operations in computing the DFT. The most efficient algorithms require $o(N \ln N)$ operations. These are collectively referred to as *Fast Fourier Transform (FFT)* algorithms. Note that $N \ln N$ grows more slowly than N^2, resulting in substantial computational savings. To give a sense of how such algorithms work, we will study one FFT algorithm, called *Decimation in Time:*

Suppose N is even. We may split the DFT into 2 parts:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\Omega_0 kn} = \sum_{\substack{n=0 \\ n \text{ even}}}^{N-1} x[n] e^{-j\Omega_0 kn} + \sum_{\substack{n=0 \\ n \text{ odd}}}^{N-1} x[n] e^{-j\Omega_0 kn}.$$

Substituting $m = \frac{n}{2}$ in the first sum and $m = \frac{n-1}{2}$ in the second,

$$X[k] = \sum_{m=0}^{\frac{N}{2}-1} x[2m] e^{-j(2\Omega_0)km} + e^{-j\Omega_0 k} \sum_{m=0}^{\frac{N}{2}-1} x[2m+1] e^{-j(2\Omega_0)km} = H_0[k] + e^{-j\Omega_0 k} H_1[k]. \quad (5.7)$$

We note that $H_0[k]$ and $H_1[k]$ are the DFT's of $x[2m]$ and $x[2m+1]$, each having length $\frac{N}{2}$ and fundamental frequency

$$\frac{2\pi}{\frac{N}{2}} = 2\Omega_0.$$

Normally, $H_0[k]$ and $H_1[k]$ are defined only for $0 \le k \le \frac{N}{2} - 1$, rendering (5.7) meaningless for $\frac{N}{2} \le k \le N-1$. This problem can be avoided by extending $H_0[k]$ and $H_1[k]$ periodically. In other words, we evaluate the 2 sums in (5.7) over the entire frequency range $0 \le k \le N-1$.

If $\frac{N}{2}$ is even, we may apply the same technique to $H_0[k]$ and $H_1[k]$ to split the DFT of $X[k]$ into 4 parts:

$$\begin{aligned}
H_0[k] &= \sum_{\substack{m=0 \\ m \text{ even}}}^{\frac{N}{2}-1} x[2m]\, e^{-j(2\Omega_0)km} + \sum_{\substack{m=0 \\ m \text{ odd}}}^{\frac{N}{2}-1} x[2m]\, e^{-j(2\Omega_0)km} \\
&= \sum_{p=0}^{\frac{N}{4}-1} x[4p]\, e^{-j(4\Omega_0)kp} + e^{-j(2\Omega_0)k} \sum_{p=0}^{\frac{N}{4}-1} x[4p+2]\, e^{-j(4\Omega_0)kp} \\
&= G_0[k] + e^{-j(2\Omega_0)k} G_1[k]
\end{aligned} \tag{5.8}$$

$$\begin{aligned}
H_1[k] &= \sum_{\substack{m=0 \\ m \text{ even}}}^{\frac{N}{2}-1} x[2m+1]\, e^{-j(2\Omega_0)km} + \sum_{\substack{m=0 \\ m \text{ odd}}}^{\frac{N}{2}-1} x[2m+1]\, e^{-j(2\Omega_0)km} \\
&= \sum_{p=0}^{\frac{N}{4}-1} x[4p+1]\, e^{-j(4\Omega_0)kp} + e^{-j(2\Omega_0)k} \sum_{p=0}^{\frac{N}{4}-1} x[4p+3]\, e^{-j(4\Omega_0)kp} \\
&= G_2[k] + e^{-j(2\Omega_0)k} G_3[k]
\end{aligned} \tag{5.9}$$

If $N = 2^Q$, (5.7) may be applied recursively a total of

$$\sum_{q=0}^{Q-2} 2^q = \frac{2^{Q-1}-1}{2^{Q-2}-1}$$

times. It then remains to find the DFT of 2^{Q-1} signals of length 2. A careful accounting of products and sums shows that this approach requires $o\left(Q \cdot 2^Q\right)$ operations. Since

$$Q = \log_2 N = \frac{\ln N}{\ln 2},$$

we may also express the computational burden as $o(N \ln N)$, proving that decimation in time qualifies as an FFT algorithm.

Example 5.2 *The following signal flow graph illustrates decimation in time for $N = 8$:*

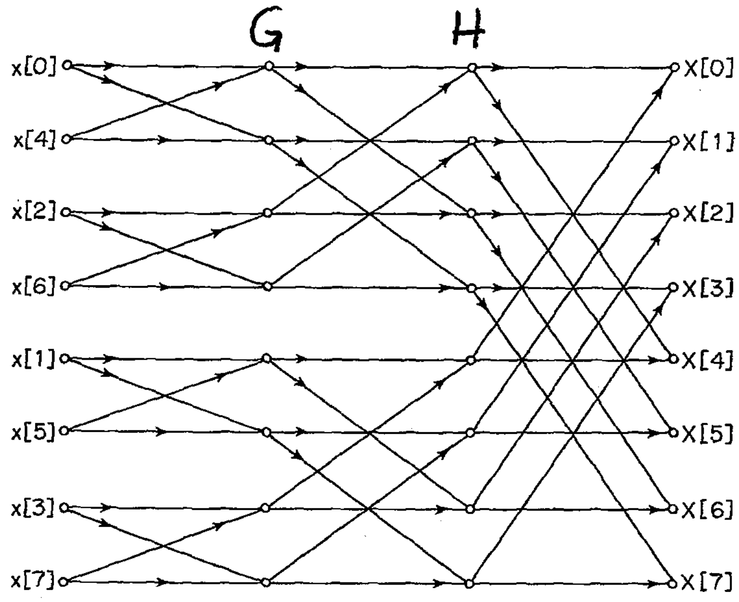


Figure 5.3

Applying (5.8) and (5.9),

$$\begin{bmatrix} G_0[0] \\ G_0[1] \end{bmatrix} = \begin{bmatrix} x[0]+x[4] \\ x[0]-x[4] \end{bmatrix}$$

$$\begin{bmatrix} G_1[0] \\ G_1[1] \end{bmatrix} = \begin{bmatrix} x[2]+x[6] \\ x[2]-x[6] \end{bmatrix}$$

$$\begin{bmatrix} G_2[0] \\ G_2[1] \end{bmatrix} = \begin{bmatrix} x[1]+x[5] \\ x[1]-x[5] \end{bmatrix}$$

$$\begin{bmatrix} G_3[0] \\ G_3[1] \end{bmatrix} = \begin{bmatrix} x[3]+x[7] \\ x[3]-x[7] \end{bmatrix}$$

$$\begin{bmatrix} H_0[0] \\ H_0[1] \\ H_0[2] \\ H_0[3] \end{bmatrix} = \begin{bmatrix} G_0[0]+G_1[0] \\ G_0[1]-jG_1[1] \\ G_0[2]-G_1[2] \\ G_0[3]+jG_1[3] \end{bmatrix} = \begin{bmatrix} G_0[0]+G_1[0] \\ G_0[1]-jG_1[1] \\ G_0[0]-G_1[0] \\ G_0[1]+jG_1[1] \end{bmatrix}$$

$$\begin{bmatrix} H_1[0] \\ H_1[1] \\ H_1[2] \\ H_1[3] \end{bmatrix} = \begin{bmatrix} G_2[0]+G_3[0] \\ G_2[1]-jG_3[1] \\ G_2[2]-G_3[2] \\ G_2[3]+jG_3[3] \end{bmatrix} = \begin{bmatrix} G_2[0]+G_3[0] \\ G_2[1]-jG_3[1] \\ G_2[0]-G_3[0] \\ G_2[1]+jG_3[1] \end{bmatrix}.$$

From (5.7),

$$\begin{bmatrix} X[0] \\ X[1] \\ X[2] \\ X[3] \\ X[4] \\ X[5] \\ X[6] \\ X[7] \end{bmatrix} = \begin{bmatrix} H_0[0]+H_1[0] \\ H_0[1]+\frac{1-j}{\sqrt{2}}H_1[1] \\ H_0[2]-jH_1[2] \\ H_0[3]-\frac{1+j}{\sqrt{2}}H_1[3] \\ H_0[4]-H_1[4] \\ H_0[5]-\frac{1-j}{\sqrt{2}}H_1[5] \\ H_0[6]+jH_1[6] \\ H_0[7]+\frac{1+j}{\sqrt{2}}H_1[7] \end{bmatrix} = \begin{bmatrix} H_0[0]+H_1[0] \\ H_0[1]+\frac{1-j}{\sqrt{2}}H_1[1] \\ H_0[2]-jH_1[2] \\ H_0[3]-\frac{1+j}{\sqrt{2}}H_1[3] \\ H_0[0]-H_1[0] \\ H_0[1]-\frac{1-j}{\sqrt{2}}H_1[1] \\ H_0[2]+jH_1[2] \\ H_0[3]+\frac{1+j}{\sqrt{2}}H_1[3] \end{bmatrix}.$$

Many other FFT algorithms have been devised. (See O&S Sections 9.2-9.5.)

Exploiting circular reflection, the IDFT can be written as

$$x[n] = \frac{1}{N}\left(\sum_{k=0}^{N-1} X[k]\, e^{-j\Omega_0 k((-n))_N}\right).$$

This says that the IDFT can be obtained by taking the DFT of $X[k]$, performing a circular reflection, and dividing by N. Starting with $X[0], \ldots, X[N-1]$, FFT provides the values $\widetilde{x}[0], \ldots, \widetilde{x}[N-1]$. Then

$$x[n] = \frac{\widetilde{x}[((-n))_N]}{N} = \begin{cases} \frac{\widetilde{x}[0]}{N}, & n = 0 \\ \frac{\widetilde{x}[N-n]}{N}, & n = 1, \ldots, N-1 \end{cases}.$$

5.4 Zero-Padding

If N is not a power of 2, we need to reconfigure $x[n]$ to apply FFT algorithms. Choose Q such that

$$2^{Q-1} < N < 2^Q, \tag{5.10}$$

define $M = 2^Q$, and consider $x[n]$ on the interval $0 \leq n \leq M-1$, recalling that $x[n] = 0$ for $N \leq n \leq M-1$. This method is called *zero-padding*, since it extends the $x[n]$ with 0's. Zero-padding produces a DFT of length M and changes the values of $X[k]$ for $0 \leq k \leq N-1$. Nevertheless, we will see in the next section that the zero-padded DFT is still a viable representation of the DTFT of $x[n]$. Taking ln of (5.10),

$$Q - 1 < \ln N < Q$$

$$2^{Q-1}(Q-1)\ln 2 < N \ln N < 2^Q Q \ln 2$$

$$o\left(2^{Q-1}(Q-1)\ln 2\right) = o\left(\frac{\ln 2}{2} Q 2^Q - \frac{\ln 2}{2} 2^Q\right) = o\left(Q 2^Q\right)$$

$$o\left(2^Q Q \ln 2\right) = o\left(Q 2^Q\right),$$

so

$$o(M \ln M) = o\left(Q 2^Q\right) = o(N \ln N).$$

Hence, the computational efficiency of the algorithm is unchanged.

6 Applications of the DFT

6.1 Spectral Analysis

Recalling that $\Omega_0 = \frac{2\pi}{N}$, (5.3) states

$$X[k] = X\left(e^{j\frac{2\pi}{N}k}\right). \tag{6.1}$$

Now suppose we employ zero-padding by choosing an arbitrary $M > N$ and computing the DFT on that basis. This amounts to extending $x[n]$ with 0's for $N \leq n \leq M-1$. The new DFT is

$$\begin{aligned} X_M[k] &= \sum_{n=0}^{M-1} x[n] e^{j\frac{2\pi}{M}kn} \qquad (6.2) \\ &= \sum_{n=0}^{N-1} x[n] e^{j\frac{2\pi}{M}kn} \\ &= X\left(e^{j\frac{2\pi}{M}k}\right). \end{aligned}$$

Comparing (6.1) and (6.2), we see that the effect of zero-padding is to resample $X\left(e^{j\Omega}\right)$ at more closely spaced frequencies (M points instead of N). Thus zero-padding gives a more refined picture of the signal spectrum.

6.2 Linear Convolution

Suppose $x[n]$ and $h[n]$ are finite-duration signals with

$$\begin{aligned} x[n] &= 0 \quad \text{for } n < 0 \text{ and } n \geq N, \\ h[n] &= 0 \quad \text{for } n < 0 \text{ and } n \geq M. \end{aligned}$$

Then the (linear) convolution of $x[n]$ and $h[n]$ is

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

But

$$x[n-m] = 0, \quad m > n$$

and

$$h[m] = 0, \quad m < 0,$$

so

$$x[n] * h[n] = \sum_{m=0}^{n} x[n-m] h[m], \quad 0 \leq n \leq M+N-1.$$

A careful accounting of operations reveals a computational burden of $o(MN)$ for linear convolution.

For non-real-time applications, an alternative approach is to use the convolution property of the DFT. Unfortunately, the convolution property applies to circular, not linear, convolution.

Example 6.1 *Let $M = N$ and*

$$x[n] = \begin{cases} 1, & 0 \leq n \leq N-1 \\ 0, & \text{else} \end{cases}.$$

For $0 \leq n \leq N-1$, the linear convolution is

$$\begin{aligned} x[n] *_{lin} x[n] &= \sum_{m=-\infty}^{\infty} x[n-m] x[m] \\ &= \sum_{m=0}^{n} 1 \\ &= n+1. \end{aligned}$$

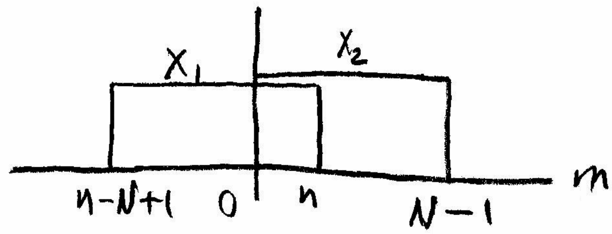


Figure 6.1

For $N \leq n \leq 2N-2$,

$$x[n] *_{lin} x[n] = \sum_{m=n-N+1}^{N-1} 1 = 2N - n - 1.$$

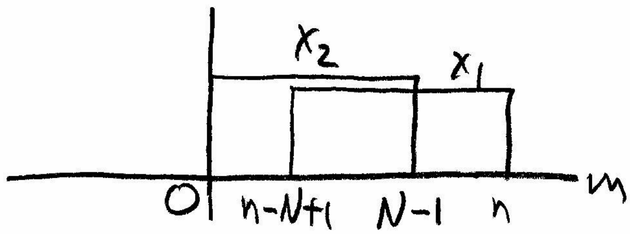


Figure 6.2

But the circular convolution is

$$\begin{aligned} x[n] *_{circ} x[n] &= \sum_{m=0}^{N-1} x\left[((n-m))_N\right] x[m] \\ &= \sum_{m=0}^{N-1} 1 \\ &= N \end{aligned}$$

for $0 \leq n \leq N-1$.

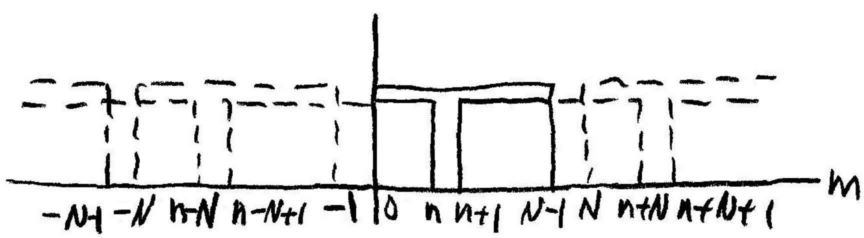


Figure 6.3

Fortunately, we can change linear into circular convolution through zero-padding: Extending both signals with 0's to $M+N-1$ points, the circular convolution is

$$\begin{aligned} h[n] *_{circ} x[n] &= \sum_{m=0}^{M+N-2} h\left[((n-m))_{M+N-1}\right] x[m] \\ &= \sum_{m=0}^{n} h\left[((n-m))_{M+N-1}\right] x[m] + \sum_{m=n+1}^{M+N-2} h\left[((n-m))_{M+N-1}\right] x[m]. \end{aligned}$$

The range of m and n may be decomposed into several regions as shown in Figure 6.4:

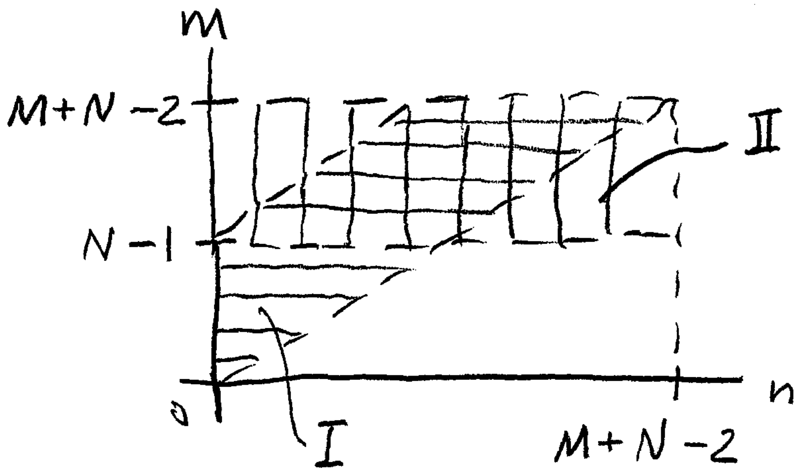


Figure 6.4

In region I,

$$n < m \leq n + N - 1$$

or

$$-N + 1 \leq n - m < 0.$$

Hence,

$$((n - m))_{M+N-1} = n - m + M + N - 1$$

and

$$((n - m))_{M+N-1} \geq M,$$

so

$$h\left[((n - m))_{M+N-1}\right] = 0.$$

In region II,

$$N - 1 < m \leq M + N - 2,$$

so

$$x[m] = 0.$$

Note that the triangular region

$$\left\{(m, n) \,\middle|\, 0 \leq n < m \leq M + N - 2\right\} \subset I \cup II,$$

so

$$\sum_{m=n+1}^{M+N-2} h\left[((n - m))_{M+N-1}\right] x[m] = 0$$

and

$$h[n] *_{circ} x[n] = \sum_{m=0}^{n} h[n - m]\, x[m] = h[n] *_{lin} x[n]\,.$$

By the convolution property of DFT's, we may compute the linear convolution $h[n] * x[n]$ via the following algorithm:

Algorithm 6.1 *1) Zero-pad $h[n]$ and $x[n]$ to $M+N-1$ points.*
2) Apply FFT to $x[n]$ and $h[n]$.
3) Compute $H[k]X[k]$.
4) Apply FFT to $H[k]X[k]$.
5) Reflect the result (mod $M+N-1$) and divide by $M+N-1$.

This approach requires $o((M+N)\ln(M+N))$ operations in steps 2) and 4) and $o(M+N)$ in steps 3) and 5). Hence, convolution using the DFT is performed in $o((M+N)\ln(M+N))$ operations. Compare this to the $o(MN)$ required for direct convolution. Which approach is faster depends on the values of M and N.

6.3 Windowing

In many applications, one needs to compute the DFT of a signal $x[n]$ that is infinite-duration. For example, this can occur when $x[n]$ is specified by the mathematical constructions encountered in a digital filter design. Another possibility is that $x[n]$ actually is finite-duration, but too long to process as a whole. A common approach to such problems is to truncate $x[n]$ to a manageable size. This operation can be done in a variety of ways and is generally referred to as *windowing*.

Windowing may be thought of as an approximation technique. As long as $x[n]\to 0$ as $|n|\to\infty$, the approximation can be made arbitrarily close by extending the width of the window. However, computer memory and processing speed is limited, so the windowed signal must be kept to a reasonable length.

In principle, a *window* is any finite-duration signal $w[n]$ satisfying (5.1). However, the most useful windows are designed so that the windowed signal is a close approximation to the original. The most commonly used windows are

1) $w[n]=w_N[n]=\begin{cases}1, & 0\le n\le N-1\\ 0, & \text{else}\end{cases}$ (rectangular)

2) $w[n]=\begin{cases}\frac{2}{N-1}n, & 0\le n\le\frac{N-1}{2}\\ 2-\frac{2}{N-1}n, & \frac{N-1}{2}<n\le N-1\\ 0, & \text{else}\end{cases}$ (Bartlett or triangular)

3) $w[n]=\begin{cases}\frac{1}{2}-\frac{1}{2}\cos\left(\frac{2\pi}{N-1}n\right), & 0\le n\le N-1\\ 0, & \text{else}\end{cases}$ (Hann)

4) $w[n]=\begin{cases}.54-.46\cos\left(\frac{2\pi}{N-1}n\right), & 0\le n\le N-1\\ 0, & \text{else}\end{cases}$ (Hamming)

5) $w[n]=\begin{cases}.42-.5\cos\left(\frac{2\pi}{N-1}n\right)+.08\cos\left(\frac{4\pi}{N-1}n\right), & 0\le n\le N-1\\ 0, & \text{else}\end{cases}$ (Blackman)

(See O&S pp. 536-538.) These windows are depicted in Figure 6.5:

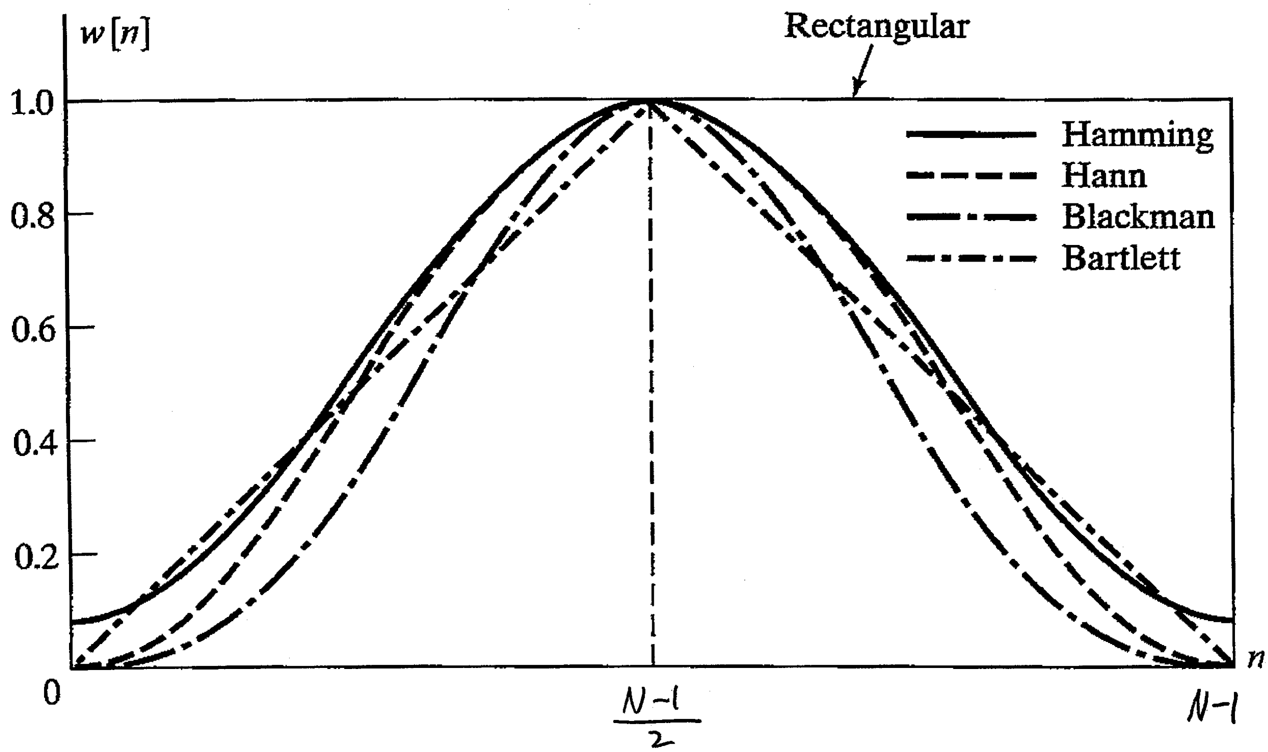


Figure 6.5

The windows may be scaled in amplitude and shifted in time as the application dictates.

The only perfect "window" is the constant $w[n] = 1$. But a nonzero constant does not qualify as a window, since it is infinite-duration. Nevertheless, it is useful to note that, from the multiplication property of the DTFT, multiplication of $x[n]$ by 1 corresponds to convolution of $X\left(e^{j\Omega}\right)$ by

$$W\left(e^{j\Omega}\right) = 2\pi \sum_{k=-\infty}^{\infty} \delta\left(\Omega - 2\pi k\right).$$

Thus we may approach the problem of window selection as that of finding one whose DTFT best approximates a periodic impulse train.

Consider first the rectangular window. Its DTFT is

$$W\left(e^{j\Omega}\right) = \frac{\sin\left(\frac{N\Omega}{2}\right)}{\sin\frac{\Omega}{2}} e^{-j\frac{(N-1)\Omega}{2}}$$

(See Figure 2.1.) It is more instructive to graph the "normalized gain"

$$20\log\left|\frac{W\left(e^{j\Omega}\right)}{W(1)}\right|,$$

obtained by dividing $\left|W\left(e^{j\Omega}\right)\right|$ by its value at $\Omega = 0$ and converting to dB:

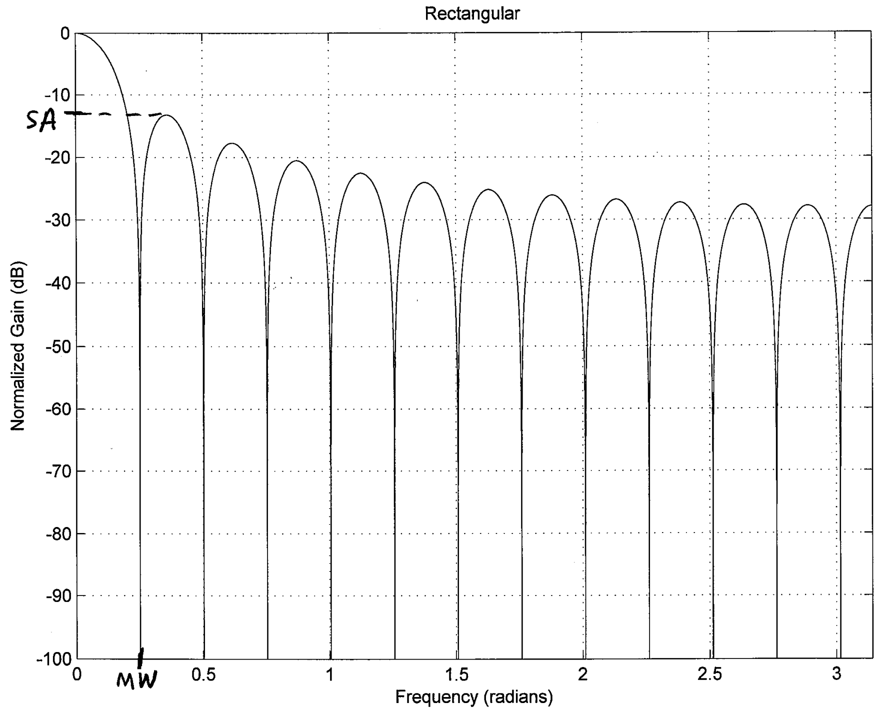
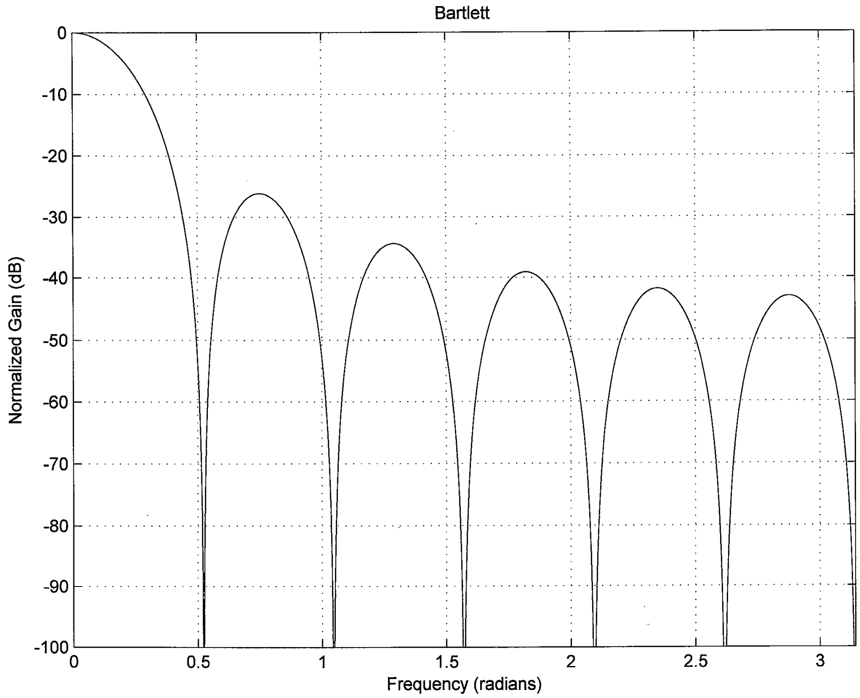
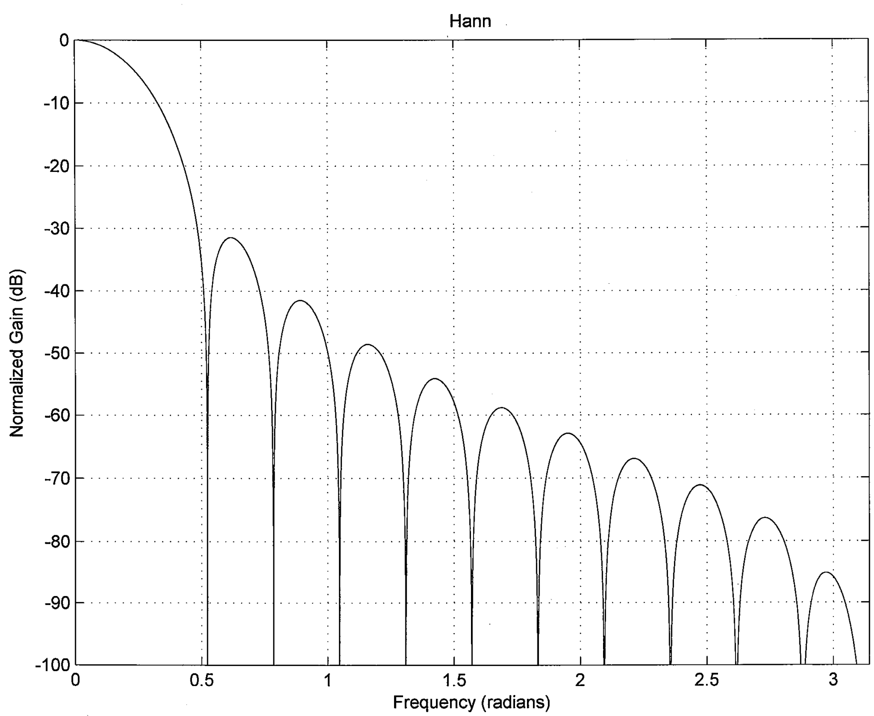


Figure 6.6 corresponds to $N = 25$. The most important quantities to observe in such a graph are the width MW of the "main lobe" and the amplitude SA of the first "side-lobe". For the rectangular window, the main lobe has width about $\frac{2\pi}{N}$, while the side lobe has maximum amplitude about -13 dB (relative to $W(1)$). These measures are compared to those of the ideal impulse train: $MW = 0$, $SA = -\infty$ dB. Thus we wish to find a window whose spectrum is concentrated near $\Omega = 0$. The phase shift in $W\left(e^{j\Omega}\right)$ is not important, since it can be eliminated by time-shifting $w[n]$. (See the time-shift property of the DTFT.)

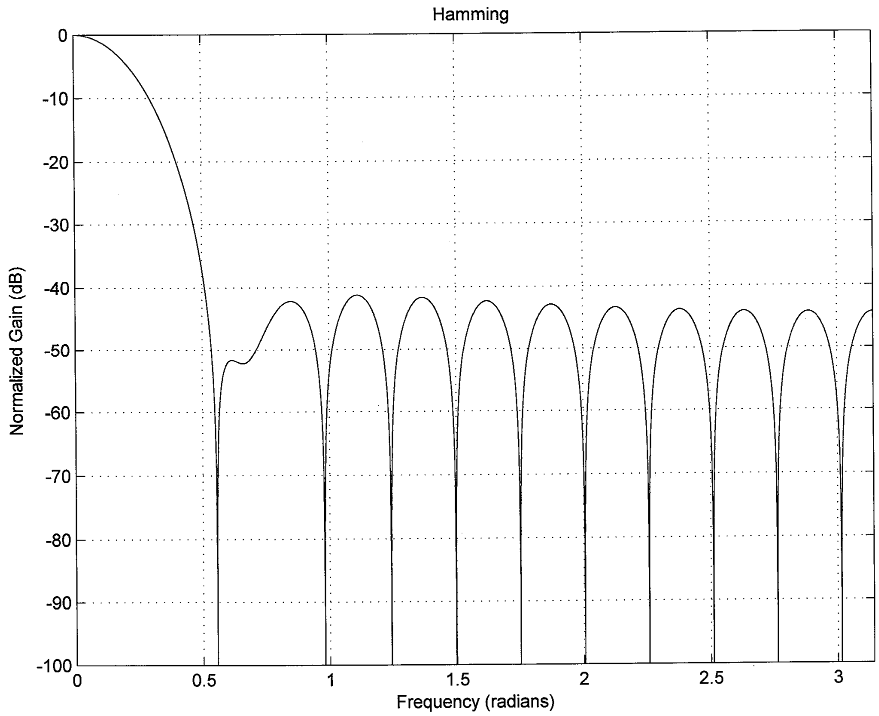
Similar graphs for the Bartlett, Hann, Hamming, and Blackman windows are shown in Figures 6.7-6.10, respectively.

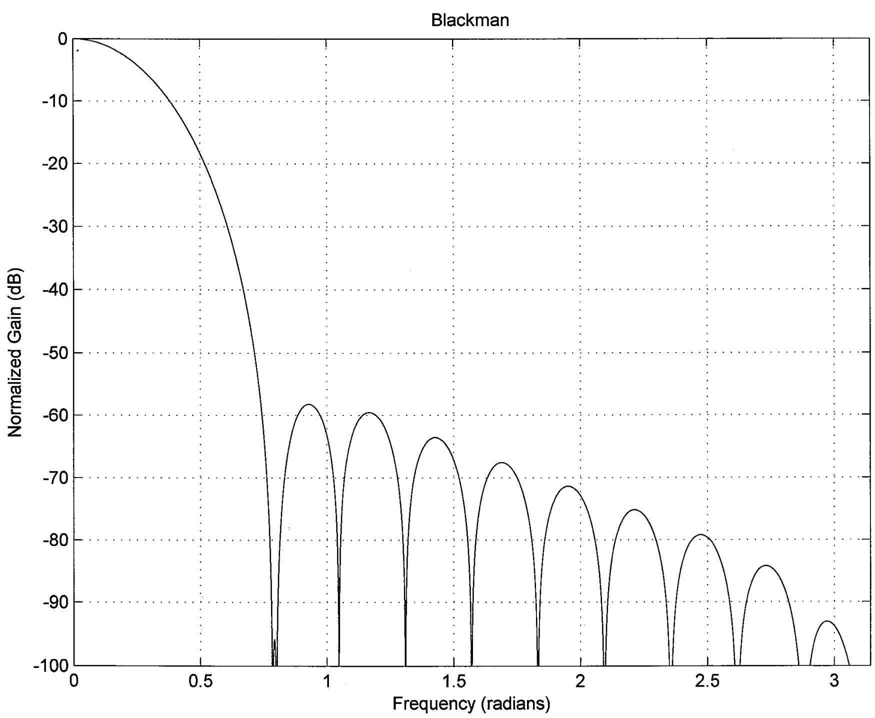


Bartlett
Normalized Gain (dB)
Frequency (radians)
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-100
0
0.5
1
1.5
2
2.5
3



Hann
Normalized Gain (dB)
Frequency (radians)
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-100
0
0.5
1
1.5
2
2.5
3





The values of MW and SA for the 5 windows are tabulated below:

Window	MW (rad)	SA (dB)
Rectangular	$\frac{2\pi}{N}$	-13
Bartlett	$\frac{4\pi}{N-1}$	-25
Hann	$\frac{4\pi}{N-1}$	-31
Hamming	$\frac{4\pi}{N-1}$	-41
Blackman	$\frac{6\pi}{N-1}$	-57

Note the trade-off between MW and SA.

Another way to look at the effect of windowing is to examine a prototype signal. Let $0 < \Omega_B < \pi$ and $X\left(e^{j\Omega}\right)$ have period 2π with

$$X\left(e^{j\Omega}\right) = \begin{cases} 1, & |\Omega| \leq \Omega_B \\ 0, & \Omega_B < |\Omega| \leq \pi \end{cases}.$$

Then

$$\begin{aligned} x[n] &= \frac{1}{2\pi}\int_{-\pi}^{\pi} X\left(e^{j\Omega}\right) e^{j\Omega n} d\Omega \\ &= \frac{1}{2\pi}\int_{-\Omega_B}^{\Omega_B} e^{j\Omega n} d\Omega \\ &= \frac{\sin \Omega_B n}{\pi n} \\ &= \frac{\Omega_B}{\pi} \operatorname{sinc}\left(\frac{\Omega_B}{\pi} n\right). \end{aligned}$$

Let $w[n]$ be the rectangular window time-shifted to the left by $\frac{N-1}{2}$. Then

$$W\left(e^{j\Omega}\right) = \frac{\sin\left(\frac{N\Omega}{2}\right)}{\sin\left(\frac{\Omega}{2}\right)}.$$

Multiplying $x[n]\, w[n]$ corresponds to the periodic convolution

$$\frac{1}{2\pi} X\left(e^{j\Omega}\right) * W\left(e^{j\Omega}\right) = \frac{1}{2\pi}\int_{\Omega-\pi}^{\Omega+\pi} X\left(e^{j(\Omega-\theta)n}\right) W\left(e^{j\theta}\right) d\theta = \frac{1}{2\pi}\int_{\Omega-\Omega_B}^{\Omega+\Omega_B} W\left(e^{j\theta}\right) d\theta$$

shown in Figure 6.11:

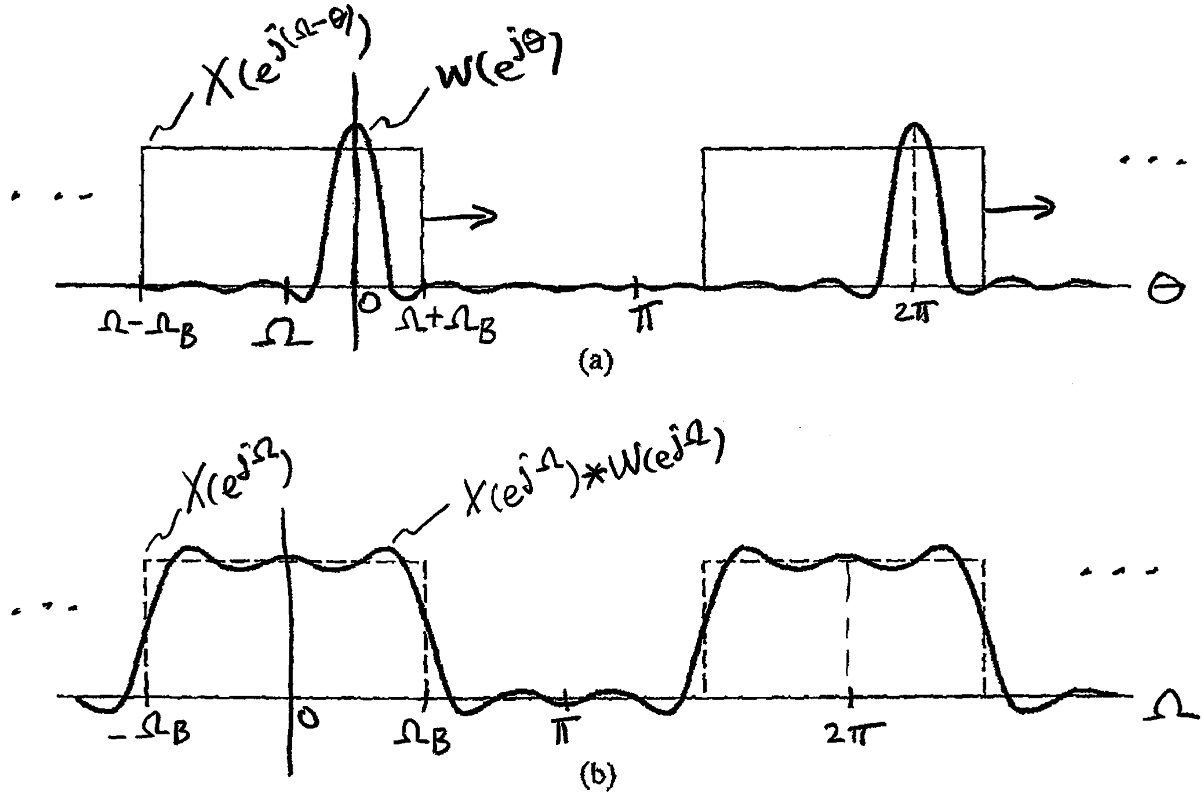


Figure 6.11

Part (a) depicts the process of convolution and (b) the result. The oscillations in (b) are called *Gibbs phenomena*. They are due to the interaction of the sinc function with the discontinuity in $X\left(e^{j\Omega}\right)$ in the convolution process. Note that the "overshoot" increases with SA, while the "sharpness" of the approximation improves with smaller MW. The choice of window $w[n]$ may be viewed as the quest to improve these effects.

Example 6.2 *Real-Time Spectrum Analyzer*
We wish to perform a running computation of the "Fourier transform versus time" for an incoming signal $x[n]$. *Since* $x[n]$ *has unknown duration, it must be windowed before we can apply the DFT. Choosing a particular window* $w[n]$ *of length* N, *we position the window at time* n *to enclose input values* $x[n-N+1], \ldots, x[n]$. *As* n *increases, the window slides from left to right:*

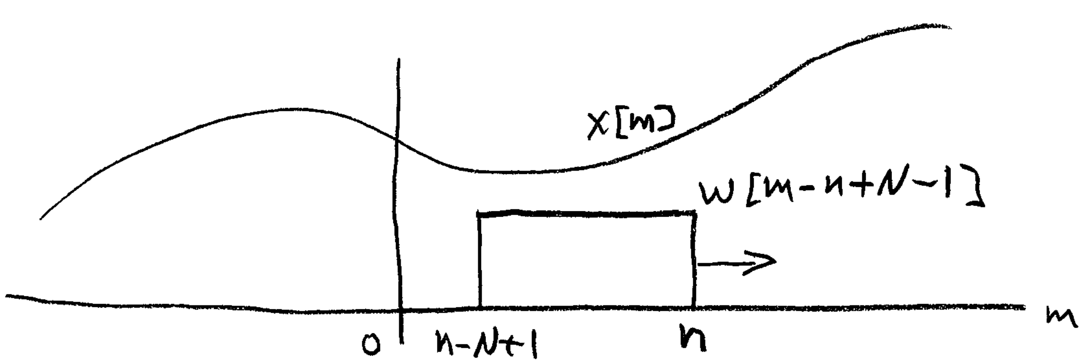


Figure 6.12

The nth windowed signal is

$$x_n[m] = x[m]\,w[m-n+N-1].$$

For each n, we shift to the interval $0 \le m \le N-1$

$$y_n[m] = x_n[m+n-N+1] = x[m+n-N+1]\,w[m]$$

and take the DFT:

$$Y_n[k] = \sum_{m=0}^{N-1} y_n[m]\,e^{j\Omega_0 km} = \sum_{m=0}^{N-1} x[m+n-N+1]\,w[m]\,e^{j\Omega_0 km}$$

For each n, the arrays $|Y_n[k]|$ and $\angle Y_n[k]$ can be displayed. In practice, in order to save computation, $Y_n[k]$ is computed for only about 1 in $\frac{N}{5}$ values of n. This corresponds to 80% overlap of adjacent windows.

7 The z-Transform

7.1 The CT Laplace Transform

One of our main goals is to be able to design digital filters. An indispensable tool in filter design is the z-Transform. First we review the Laplace Transform. For a CT signal $x(t)$, its *Laplace Transform (LT)* is

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

where $s = \sigma + j\omega$ is *complex frequency.* In general, the Laplace integral can be evaluated on only a subset of the complex plane.

Example 7.1 *Let $u(t)$ be the unit step function and*

$$x(t) = e^t u(t).$$

Then

$$\begin{aligned} X(s) &= \int_0^{\infty} e^t e^{-st}dt \\ &= \int_0^{\infty} e^{(1-s)t}dt \\ &= \frac{1}{1-s} e^{(1-s)t}\Big|_0^{\infty}. \end{aligned}$$

But

$$e^{(1-s)t} = e^{(1-\sigma)t}e^{-j\omega t} = e^{(1-\sigma)t}(\cos\omega t + j\sin\omega t),$$

so the upper limit can be evaluated iff $\sigma > 1$. In other words, $X(s)$ is defined only for the right half-plane $\operatorname{Re} s > 1$. *In this case,*

$$X(s) = \frac{1}{1-s}(0-1) = \frac{1}{s-1}.$$

The set of s on which $X(s)$ is defined is called the *Region of Convergence (ROC)*. One can show that, in general, the ROC is an open vertical strip in the complex plane

$$ROC = \left\{ s \,\middle|\, a < \operatorname{Re} s < b \right\},$$

where $-\infty \leq a < b \leq \infty$. In other words, the ROC is either a vertical strip, bounded on both sides, a left half-plane ($a = -\infty$), a right half-plane ($b = \infty$), or the entire plane ($a = -\infty$, $b = \infty$). If the ROC contains the imaginary axis, then we may evaluate $s = j\omega$, yielding the CTFT $X(j\omega)$. (This explains the notation $X(j\omega)$.)

The *Inverse Laplace Transform (ILT)* is

$$\mathcal{L}^{-1}\{X(s)\} = x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\sigma + j\omega)\, e^{(\sigma+j\omega)t} d\omega.$$

The integral evaluates to $x(t)$ as long as σ lies in the ROC.

7.2 The DT Laplace Transform and the z-Transform

To each DT signal $x[n]$ we may associate an impulse train

$$x(t) = \sum_{n=-\infty}^{\infty} x[n]\,\delta(t-n)$$

and apply the LT:

$$\begin{aligned} X(s) &= \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x[n]\,\delta(t-n)\, e^{-st} dt \qquad (7.1)\\ &= \sum_{n=-\infty}^{\infty} x[n]\, e^{-sn} \int_{-\infty}^{\infty} \delta(t-n)\, dt \\ &= \sum_{n=-\infty}^{\infty} x[n]\, e^{-sn}. \end{aligned}$$

We denote the complex frequency variable in DT problems as

$$S = \Sigma + j\Omega.$$

The *DT Laplace transform* is

$$X\left(e^{S}\right) = \sum_{n=-\infty}^{\infty} x[n]\, e^{-Sn}.$$

Since we may write

$$X\left(e^{S}\right) = \sum_{n=-\infty}^{\infty} x[n]\left(e^{-S}\right)^{n},$$

$X\left(e^{S}\right)$ may be viewed as a function of e^{S}.Note that

$$e^{S+j2\pi} = e^{S} e^{j2\pi} = e^{S}.$$

Hence, e^{S} and, therefore, $X\left(e^{S}\right)$ have period $j2\pi$.

Since $X\left(e^{S}\right)$ is still an LT, the rules surrounding the ROC continue to apply. A DT Laplace transform need only be specified on the rectangle bounded horizontally by a and b and vertically by 0 and 2π :

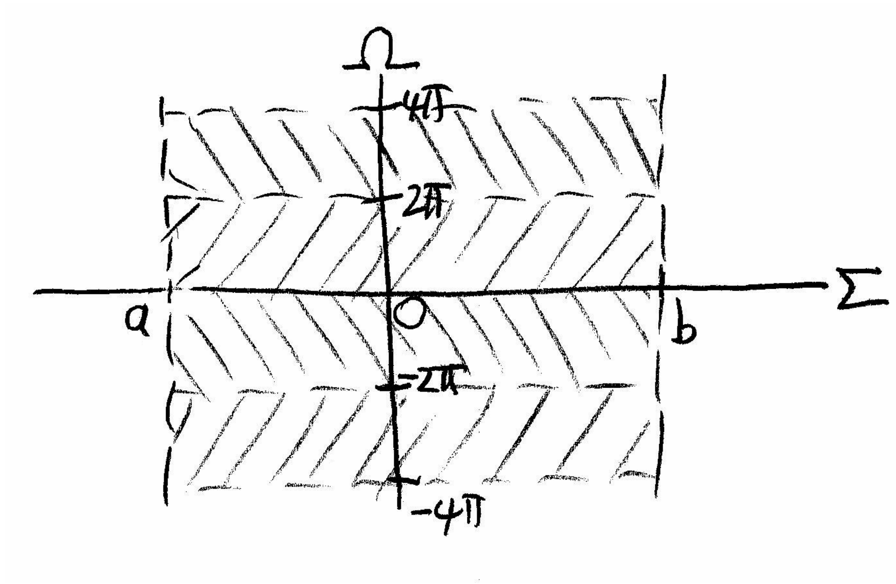


Figure 7.1

The value of $X\left(e^S\right)$ on the remaining rectangles is determined completely by periodicity.

Example 7.2 *Let ρ be any complex number, $u[n]$ be the DT unit step, and*

$$x[n] = \rho^n u[n].$$

Then

$$\begin{aligned} X\left(e^S\right) &= \sum_{n=0}^{\infty} \rho^n e^{-Sn} \\ &= \sum_{n=0}^{\infty} \left(\rho e^{-S}\right)^n \\ &= \frac{1}{1-\rho e^{-S}}. \end{aligned}$$

The geometric series converges for

$$|\rho|\, e^{-\Sigma} = \left|\rho e^{-S}\right| < 1$$

or

$$\Sigma > \ln|\rho|.$$

Hence,

$$ROC = \left\{ S \,\middle|\, \operatorname{Re} S > \ln|\rho| \right\}.$$

In order to avoid writing e^S repeatedly, it is conventional to introduce the frequency map

$$z = e^S. \tag{7.2}$$

The complex variable z may also be considered complex frequency. (Both S and z are dimensionless.) From (7.2),

$$|z| = e^{\Sigma}, \quad \angle z = \Omega. \tag{7.3}$$

Making the substitution (7.2), the DT Laplace transform becomes the *z-Transform (ZT)*

$$\mathcal{Z}\left\{x\left[n\right]\right\} = X\left(z\right) = \sum_{n=-\infty}^{\infty} x\left[n\right] z^{-n}.$$

In many cases, $X\left(z\right)$ is defined at $z = 0$ in spite of the fact that $e^{S} \neq 0$ for all finite S. Thus, in effect, the z-transform extends the LT to $S = \infty$.

It is important to understand the geometry of the exponential map (7.2). Consider the vertical line

$$VL\left(\Sigma\right) = \left\{S \,\middle|\, \operatorname{Re} S = \Sigma\right\}$$

in the S-plane. From (7.3), $VL\left(\Sigma\right)$ maps onto the circle

$$C\left(e^{\Sigma}\right) = \left\{z \,\middle|\, |z| = e^{\Sigma}\right\}$$

in the z-plane. Owing to periodicity, the exponential wraps $VL\left(\Sigma\right)$ infinitely many times around $C\left(e^{\Sigma}\right)$. In particular, setting $\Sigma = 0$ shows that the imaginary axis $VL\left(0\right)$ maps onto the *unit circle* $C\left(1\right)$.

We may extend this idea to all ROC's in the S-plane: Consider the vertical strip

$$VS\left(a, b\right) = \left\{S \,\middle|\, a < \operatorname{Re} S < b\right\},$$

where $0 \leq \alpha < \beta \leq \infty$. It is conventional to denote

$$LHP = VS\left(-\infty, 0\right),$$

$$RHP = VS\left(0, \infty\right).$$

The exponential function maps $VS\left(a, b\right)$ onto the annulus

$$\operatorname{ann}\left(e^{a}, e^{b}\right) = \left\{z \,\middle|\, e^{a} < |z| < e^{b}\right\}.$$

In particular, LHP maps onto $\operatorname{ann}\left(0, 1\right)$.

If the ROC of $X\left(z\right)$ contains the unit circle, then we may evaluate $z = e^{j\Omega}$, yielding the DT Fourier transform $X\left(e^{j\Omega}\right)$. (This explains the notation $X\left(e^{j\Omega}\right)$.)

Example 7.3 *Find the ZT of $x\left[n\right] = \rho^{n} u\left[n\right]$. From Example 7.2, the DT Laplace transform is*

$$X\left(S\right) = \frac{1}{1 - \rho e^{-S}} = \frac{1}{1 - \rho\left(e^{S}\right)^{-1}}$$

with

$$ROC = VS\left(\ln|\rho|, \infty\right),$$

so

$$X\left(z\right) = \frac{1}{1 - \rho z^{-1}}$$

with

$$ROC = \operatorname{ann}\left(|\rho|, \infty\right).$$

Now, let us calculate $X(z)$ directly:

$$X(z) = \sum_{n=0}^{\infty} \rho^n z^{-n}$$
$$= \sum_{n=0}^{\infty} \left(\rho z^{-1}\right)^n$$
$$= \frac{1}{1-\rho z^{-1}}.$$

The geometric series converges iff $\left|\rho z^{-1}\right| < 1$ or, equivalently, $|z| > |\rho|$, so ROC $= \operatorname{ann}(|\rho|, \infty)$.

Example 7.4 *Setting $\rho = 1$ in Example 7.3 yields the ZT of the unit step:*

$$u[n] \longleftrightarrow \frac{1}{1-z^{-1}} = \frac{z}{z-1}$$

Example 7.5 *Let $x[n] = \delta[n]$. Then*

$$X(z) = \sum_{n=-\infty}^{\infty} \delta[n] z^{-n}.$$

The only nonzero term in the series is the $n = 0$ term.

$$X(z) = z^{-0} = 1.$$

The ROC is the entire z-plane.

The *Inverse z-transform (IZT)* is

$$\mathcal{Z}^{-1}\{X(z)\} = x[n] = \frac{r^n}{2\pi} \int_0^{2\pi} X\left(re^{j\Omega}\right) e^{j\Omega n} d\Omega,$$

where r lies in the ROC. The integral may be evaluated on any interval of length 2π. (Additional examples are listed in O&S Table 3.1 on p. 110.)

7.3 Properties

The ZT has the following list of properties:

Properties: (See O&S Table 3.2 on p. 132.)

$$R_x = \text{ the ROC of } X(z) = \operatorname{ann}(\alpha, \beta)$$

Linearity:

$$\begin{cases} \alpha x[n] \longleftrightarrow \alpha X(z) & (ROC = R_x) \\ x_1[n] + x_2[n] \longleftrightarrow X_1(z) + X_2(z) & (ROC \supset R_{x_1} \cap R_{x_2}) \end{cases}$$

Time Shift:

$$x[n-n_0] \longleftrightarrow z^{-n_0} X(z) \quad (ROC \supset R_x - \{0\})$$

Frequency Shift:

$$z_0^n x[n] \longleftrightarrow X\left(z_0^{-1}z\right) \quad (ROC = |z_0|\, R_x)$$

Time/Frequency Scaling:

$$R_x^{\frac{1}{m}} = \left\{ z \,\middle|\, z^m \in R_x \right\} = \operatorname{ann}\left(\alpha^{\frac{1}{m}}, \beta^{\frac{1}{m}}\right)$$

$$x_m[n] \longleftrightarrow X(z^m) \quad (ROC = R_x^{\frac{1}{m}})$$

Convolution:

$$x_1[n] * x_2[n] \longleftrightarrow X_1(z)\, X_2(z) \quad (ROC \supset R_{x_1} \cap R_{x_2})$$

Time Differencing:

$$x[n] - x[n-1] \longleftrightarrow \left(1 - z^{-1}\right) X(z) \quad (ROC \supset R_x - \{0\}\,)$$

Accumulation:

$$\sum_{m=-\infty}^{n} x[m] \longleftrightarrow \frac{1}{1 - z^{-1}} X(z) \quad (ROC \supset R_x \cap \operatorname{ann}(1, \infty)\,)$$

Frequency Differentiation:

$$nx[n] \longleftrightarrow -z\frac{dX(z)}{dz} \quad (ROC = R_x)$$

Conjugation:

$$x^*[n] \longleftrightarrow X^*(z^*) \quad (ROC = R_x)$$

Reflection:

$$\frac{1}{R_x} = \left\{ z \,\middle|\, \frac{1}{z} \in R_x \right\} = \operatorname{ann}\left(\frac{1}{\beta}, \frac{1}{\alpha}\right)$$

$$x[-n] \longleftrightarrow X\left(z^{-1}\right) \quad (ROC = \frac{1}{R_x})$$

Real Time Signal

$$x[n] \text{ real } \iff X(z^*) = X^*(z)$$

8 DT Systems and the ZT

8.1 LTI Systems

Recall that a DT system is LTI iff it maps the input $x[n]$ to the output $y[n]$ via (linear) convolution:

$$y[n] = h[n] * x[n].$$

The sifting property of $\delta[n]$ tells us that the input $x[n] = \delta[n]$ produces the output

$$y[n] = h[n] * \delta[n] = h[n].$$

For this reason, $h[n]$ is called the *impulse response* of the system. By the convolution property of the ZT,

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

$H(z)$ is the *transfer function* of the system.

Since DT convolution is a linear operation, impulse responses and transfer functions of systems in parallel add:

$$h[n] = h_1[n] + h_2[n],$$
$$H(z) = H_1(z) + H_2(z).$$

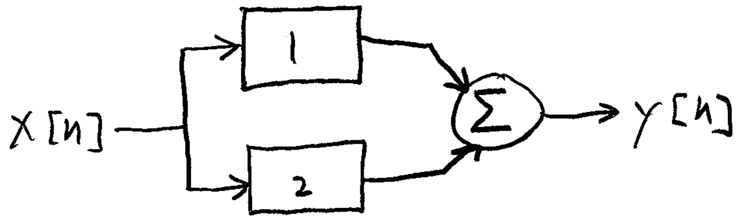


Figure 8.1

Since DT convolution is associative, impulse responses of systems in series must be (linearly) convolved

$$h[n] = h_1[n] * h_2[n],$$

while transfer functions must be multiplied

$$H(z) = H_1(z)\, H_2(z).$$

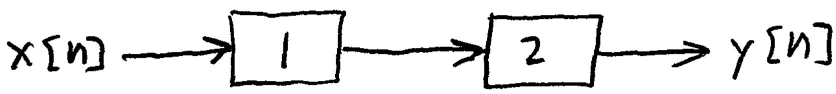


Figure 8.2

Recall that a DT system is *causal* if

$$x_1[n] = x_2[n] \quad \text{for } n < m \Longrightarrow y_1[n] = y_2[n] \quad \text{for } n < m.$$

We also define a system to be *anti-causal* if

$$x_1[n] = x_2[n] \quad \text{for } n > m \Longrightarrow y_1[n] = y_2[n] \quad \text{for } n > m.$$

In terms of the impulse response $h[n]$, an LTI system is causal iff $h[n] = 0$ for $n < 0$. It is easy to show that an LTI system is anti-causal iff $h[n] = 0$ for $n > 0$. Hence, the only LTI systems that are both causal and anti-causal are those with impulse responses of the form

$$h[n] = A\delta[n].$$

These are the *static* systems.

A DT system is *BIBO stable* if

$$x[n] \text{ bounded} \implies y[n] \text{ bounded}.$$

For LTI systems, this is equivalent to

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

8.2 Difference Equations

An important class of digital filters is implemented through the use of *difference equations*:

$$a_N y[n+N] + \ldots + a_0 y[n] = b_M x[n+M] + \ldots + b_0 x[n].$$

There is no harm in assuming $a_N \neq 0$ and $b_M \neq 0$, since otherwise M and N can be redefined. In fact, we may divide through the equation by a_N and redefine coefficients accordingly. This makes $a_N = 1$. Now suppose

$$a_0 = \ldots = a_{K-1} = 0.$$

In other words, K is the smallest index such that $a_K \neq 0$. The difference equation becomes

$$y[n+N] + a_{N-1} y[n+N-1] + \ldots + a_K y[n+K] = b_M x[n+M] + \ldots + b_0 x[n]. \tag{8.1}$$

The number $N-K$ is the *order* of the equation. For now, we will restrict ourselves to equations with order $N-K>0$.

A difference equation is very similar to a differential equation in that it expresses a relationship between shifts of the input $x[n]$ and the output $y[n]$, rather than derivatives of $x(t)$ and $y(t)$. Like a differential equation, a difference equation has infinitely many solutions corresponding to a given input signal $x[n]$. A single solution is determined uniquely by specifying $N-K$ initial conditions, typically adjacent values such as $y[-1], \ldots, y[K-N]$. For a given $x[n]$ and set of initial conditions, a difference equation can be solved using the same analytic methods as for differential equations: Find a particular solution, add the general homogeneous solution with $N-K$ free parameters, and apply the $N-K$ initial conditions.

Example 8.1 *Solve*

$$y[n+2] - \frac{5}{2} y[n+1] + y[n] = x[n]$$

for the input

$$x[n] = n$$

and initial conditions

$$y[-1] = 1, \quad y[-2] = 0.$$

As with differential equations, a polynomial input admits at least one polynomial solution. Applying "variation of parameters", we try

$$y[n] = An + B.$$

Then

$$A(n+2) + B - \frac{5}{2}(A(n+1) + B) + (An + B) = n$$

or

$$\left(-\frac{A}{2} - 1\right) n - \frac{A}{2} - \frac{B}{2} = 0.$$

which yields

$$A = -2, \quad B = 2$$

and the particular solution

$$y_p[n] = -2n + 2.$$

The "homogeneous equation" is

$$y[n+2] - \frac{5}{2}y[n+1] + y[n] = 0.$$

The polynomial

$$z^2 - \frac{5}{2}z + 1 = 0$$

has roots

$$\rho_1 = \frac{1}{2} \quad \rho_2 = 2,$$

so the general homogeneous solution is

$$y_h[n] = C\left(\frac{1}{2}\right)^n + D \cdot 2^n,$$

where C and D are free parameters. All solutions of the difference equation have the form

$$\begin{aligned} y[n] &= y_p[n] + y_h[n] \\ &= -2n + 2 + C\left(\frac{1}{2}\right)^n + D \cdot 2^n. \end{aligned}$$

Applying the initial conditions, we obtain

$$y[-1] = 4 + 2C + \frac{D}{2} = 1,$$

$$y[-2] = 6 + 4C + \frac{D}{4} = 0$$

or

$$C = -\frac{3}{2}, \quad D = 0.$$

The final answer is

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

Equation (8.1) can also be solved recursively – a method that has no counterpart in the study of differential equations. There are 2 ways of rewriting the equation that will be useful. First, *forward recursive form* is obtained by shifting (8.1) N steps to the right and solving for $y[n]$:

$$y[n] = -a_{N-1}y[n-1] - \ldots - a_K y[n+K-N] + b_M x[n+M-N] + \ldots + b_0 x[n-N]. \quad (8.2)$$

The solution is obtained via the following algorithm:

Forward Recursion (given $y[-1], \ldots, y[K-N]$)
1) Apply $x[n]$ and the initial conditions to (8.2) to find $y[0]$.
2) Apply $x[n]$, the initial conditions, and $y[0]$ to (8.2) to find $y[1]$.
3) Apply $x[n]$, the initial conditions, $y[0]$, and $y[1]$ to (8.2) to find $y[2]$.
4) Continue indefinitely.

Backward recursive form is achieved by shifting (8.1) K steps to the right and solve for $y[n]$:

$$y[n] = -\frac{1}{a_K}y[n+N-K] - \frac{a_{N-1}}{a_K}y[n+N-K-1] - \ldots - \frac{a_{K+1}}{a_K}y[n+1] + \frac{b_M}{a_K}x[n+M-K] + \ldots + \frac{b_0}{a_K}x[n-K]. \tag{8.3}$$

Backward Recursion (given $y[-1], \ldots, y[K-N]$)

1) Apply $x[n]$ and the initial conditions to (8.3) to find $y[K-N-1]$.

2) Apply $x[n]$, the initial conditions, and $y[K-N-1]$ to the right side of (8.3) to find $y[K-N-2]$.

3) Apply $x[n]$, the initial conditions, $y[K-N-1]$, and $y[K-N-2]$ to the right side of (8.3) to find $y[K-N-3]$.

4) Continue indefinitely.

Example 8.2 *Solve the problem in Example 8.1 recursively. Iterating forward,*

$$y[n] = \frac{5}{2}y[n-1] - y[n-2] + x[n-2]$$

$$\begin{aligned} y[0] &= \frac{5}{2}y[-1] - y[-2] + x[-2] = \frac{1}{2} \\ y[1] &= \frac{5}{2}y[0] - y[-1] + x[-1] = -\frac{3}{4} \\ y[2] &= \frac{5}{2}y[1] - y[0] + x[0] = -\frac{19}{8} \\ y[3] &= \frac{5}{2}y[2] - y[1] + x[1] = -\frac{67}{16} \\ &\vdots \end{aligned}$$

Iterating backward,

$$y[n] = -y[n+2] + \frac{5}{2}y[n+1] + x[n]$$

$$\begin{aligned} y[-3] &= -y[-1] + \frac{5}{2}y[-2] + x[-3] = -1 - 3 = -4 \\ y[-4] &= -y[-2] + \frac{5}{2}y[-3] + x[-4] = -10 - 4 = -14 \\ y[-5] &= -y[-3] + \frac{5}{2}y[-4] + x[-5] = 4 - 35 - 5 = -36 \\ &\vdots \end{aligned}$$

Note that these values are consistent with the analytic solution of Example 8.1.

In some problems, backward recursion is based on initial conditions $y[1], \ldots, y[N-K]$:

Backward Recursion (given $y[1], \ldots, y[N-K]$)
1) Apply $x[n]$ and the initial conditions to (8.3) to find $y[0]$.
2) Apply $x[n]$, the initial conditions, and $y[0]$ to the right side of (8.3) to find $y[-1]$.
3) Apply $x[n]$, the initial conditions, $y[0]$, and $y[-1]$ to the right side of (8.3) to find $y[-2]$.
4) Continue indefinitely.

The recursive approach can easily be written into a computer program.

Difference equations can be used to describe a certain class of LTI systems. Let $x[n] = \delta[n]$ in (8.1) and find any solution $h[n]$. This determines an LTI system with impulse response $h[n]$. If $x[n]$ is any other input, then we may check that

$$y[n] = h[n] * x[n]$$

is a corresponding solution. To do so, note that

$$\begin{aligned} y[n+m] &= \delta[n+m] * y[n] \\ &= (\delta[n+m] * h[n]) * x[n] \\ &= h[n+m] * x[n]. \end{aligned}$$

Substituting $y[n]$ into (8.1) yields

$$\begin{aligned} &y[n+N] + a_{N-1} y[n+N-1] + \ldots + a_K y[n+K] \\ &= h[n+N] * x[n] + a_{N-1} h[n+N-1] * x[n] + \ldots + a_K h[n+K] * x[n] \\ &= (h[n+N] + a_{N-1} h[n+N-1] + \ldots + a_K h[n+K]) * x[n] \\ &= (b_M \delta[n+M] + \ldots + b_0 \delta[n]) * x[n] \\ &= b_M \delta[n+M] * x[n] + \ldots + b_0 \delta[n] * x[n] \\ &= b_M x[n+M] + \ldots + b_0 x[n], \end{aligned}$$

proving that the LTI system with impulse response $h[n]$ is consistent with (8.1).

As we have seen, each input $x[n]$ leads to infinitely many solutions, so **there are infinitely many impulse responses corresponding to each difference equation**. Each impulse response determines a distinct LTI system.

Example 8.3

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

For $x[n] = \delta[n]$, a particular solution is $h_p[n] = u[n]$, since

$$u[n+1] - u[n] = \delta[n+1].$$

The homogeneous solutions are just the constants $h_h[n] = A$. The general form of the impulse response is

$$h[n] = h_p[n] + h_h[n] = u[n] + A,$$

where A is arbitrary.

Having infinitely many impulses responses may seem daunting, but it will turn out that only one of these can be used in applications.

8.3 Rational Transfer Functions

The time-shift property transforms the difference equation (8.1) into the form

$$z^{N}Y(z)+a_{N-1}z^{N-1}Y(z)+\ldots+a_{K}z^{K}Y(z)=b_{M}z^{M}X(z)+\ldots+b_{0}X(z).$$

Let

$$\Delta(z)=z^{N}+a_{N-1}z^{N-1}+\ldots+a_{K}z^{K}$$

and

$$\Gamma(z)=b_{M}z^{M}+\ldots+b_{0}.$$

The transformed equation may be written

$$\Delta(z)\,Y(z)=\Gamma(z)\,X(z). \tag{8.4}$$

$\Delta(z)$ is called the *characteristic polynomial* of the equation. Its roots are the *eigenvalues*. Basic to the study of difference equations is that the eigenvalues are the "natural frequencies" of the equation. Note that the order $N-K$ of the difference equation is positive iff there exists at least one nonzero eigenvalue.

It is tempting to divide (8.4) through by $\Delta(z)\,X(z)$, leading to a transfer function

$$\frac{Y(z)}{X(z)}=\frac{\Gamma(z)}{\Delta(z)}.$$

But this would entail dividing by zero whenever z is an eigenvalue. It is an unfortunate consequence of this fact that **not every impulse response of the difference equation has a z-transform**.

Example 8.4 *In Example 8.3, setting $A=0$ yields $h[n]=u[n]$. From Table 3.1, line 2,*

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

For $A=-1$,

$$h[n]=u[n]-1=-u[-n-1].$$

From line 3,

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

No other value of A yields a z-transform $H(z)$.

Nevertheless, the function

$$H(z)=\frac{\Gamma(z)}{\Delta(z)}=\frac{b_{M}z^{M}+\ldots+b_{0}}{z^{N}+a_{N-1}z^{N-1}+\ldots+a_{K}z^{K}} \tag{8.5}$$

is fundamental to the study of (8.1). Since $H(z)$ is the ratio of two polynomials, $H(z)$ is called a *rational function*. Starting from $H(z)$, we can recover the difference equation, since both forms are completely defined by the coefficients $a_K,\ldots,a_{N-1}$ and $b_0,\ldots,b_M$. We say $H(z)$ is *proper* if $M\leq N$, *strictly proper* if $M<N$, *improper* if $M>N$, and *biproper* if $M=N$. Note that strict properness is equivalent to $|H(z)|\to 0$ as $|z|\to\infty$, while improperness is the same as $|H(z)|\to\infty$ as $|z|\to\infty$.

Sometimes it will be advantageous to write $H(z)$ as a rational function in z^{-1}. If $H(z)$ is proper, we multiply the numerator and denominator in (8.5) by z^{-N} to obtain

$$H(z) = \frac{b_M z^{M-N} + \ldots + b_0 z^{-N}}{1 + a_{N-1} z^{-1} + \ldots + a_K z^{K-N}}. \tag{8.6}$$

If $H(z)$ is improper, multiply numerator and denominator by z^{-M} :

$$H(z) = \frac{b_M + b_{M-1} z^{-1} + \ldots + b_0 z^{-M}}{z^{N-M} + a_{N-1} z^{N-M-1} + \ldots + a_K z^{K-M}} \tag{8.7}$$

(Note that O&S Table 3.1 uses the form (8.6)-(8.7).)

8.4 Poles and Zeros

The *Fundamental Theorem of Algebra* tells us that every polynomial has a unique factorization into linear factors. For example,

$$\Gamma(z) = \prod_{i=1}^{q} (z - \eta_i)^{M_i},$$

$$\Delta(z) = \prod_{i=1}^{r} (z - \rho_i)^{N_i}.$$

The *multiplicities* M_i and N_i must add up to $\sum M_i = M$ and $\sum N_i = N$. If $\Gamma(z)$ and $\Delta(z)$ have a common root, then there is at least one factor common to both. In forming $H(z)$, all common may be factors cancelled, leaving a reduced numerator and denominator

$$H(z) = \frac{\widetilde{\Gamma}(z)}{\widetilde{\Delta}(z)}, \tag{8.8}$$

where $\widetilde{\Gamma}(z)$ and $\widetilde{\Delta}(z)$ do not have a common root. The roots of $\widetilde{\Gamma}(z)$ are the *zeros* of $H(z)$, and the roots of $\widetilde{\Delta}(z)$ are the *poles* of $H(z)$. If $H(z)$ is strictly proper, we say it has a *zero at* ∞ with multiplicity $N - M$. If $H(z)$ is improper, it has a *pole at* ∞ with multiplicity $M - N$. ($H(z)$ never has both poles and zeros at ∞.) Counting poles or zeros at ∞, the total number zeros always equals the total number of poles.

Example 8.5 *Let*

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

From the quadratic formula, the roots of $\Gamma(z)$ *are* ± 1 *and the roots of* $\Delta(z)$ *are* $-1, -2$. *Hence,*

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

$H(z)$ *has a single zero* $\eta = 1$ *and a single pole* $\rho = -2$. *The number* -1 *is neither a pole nor a zero of* $H(z)$.

In practice, filter design usually proceeds by choosing a rational function $H(z)$, transforming it to a difference equation, and writing the equation into a computer program. Representing $H(z)$ with numerator and denominator $\Gamma(z)$ and $\Delta(z)$ having a common root would unnecessarily increase the order of the difference equation and the computational burden of solving it. For this reason, **we will henceforth assume that the difference equation (8.1) has coefficients such that $\Gamma(z)$ and $\Delta(z)$ have no root in common**. In particular, this assumption implies that a difference equation has positive order $N-K$ iff $H(z)$ has at least one nonzero pole.

Each pole ρ of $H(z)$ determines a circle centered at 0 and passing through ρ. Suppose these circles have radii $0 \leq r_1 < \ldots < r_p$. Then each annulus $\operatorname{ann}(r_i, r_{i+1})$ contains no pole of $H(z)$, but does have at least one pole on each of its boundaries. If $r_1 > 0$, we also need to consider the disk

$$D(0, r_1) = \left\{ z \,\middle|\, |z| < r_1 \right\} = \operatorname{ann}(0, r_1) \cup \{0\}.$$

These regions play a special role in describing the impulse responses corresponding to the difference equation.

Theorem 8.1 *If an impulse response $h[n]$ of the difference equation (8.1) has a z-transform $H(z)$, then $H(z)$ is given by (8.5) with ROC either $\operatorname{ann}(r_i, r_{i+1})$ for some i, $\operatorname{ann}(r_p, \infty)$, or $D(0, r_1)$. Each such ROC corresponds to an impulse response $h[n]$.*

Example 8.6 *Let*

$$H(z) = \frac{1}{z^2 + \frac{5}{2}z + 1}.$$

The poles are $\rho_{1,2} = -\frac{1}{2}, -2$, so $r_1 = \frac{1}{2}$ and $r_2 = 2$. The possible regions of convergence are $D\left(0, \frac{1}{2}\right)$, $\operatorname{ann}\left(\frac{1}{2}, 2\right)$, or $\operatorname{ann}(2, \infty)$.

In applications, we need only consider systems with real impulse response. In other words, we want the IZT of $H(z)$ to be real. From the "real time-signal" property of the ZT, this is true iff $H(z^*) = H^*(z)$. For rational functions, one can show that this condition holds iff all coefficients $a_0, \ldots a_{N-1}, b_0, \ldots, b_M$ of $H(z)$ are real. Hence, we will restrict attention to rational functions with real coefficients. A consequence of this assumption is that poles and zeros are always distributed symmetrically about the real axis in the z-plane.

8.5 Partial Fraction Expansion

$H(z)$ can be decomposed into a sum of smaller terms using *partial fraction expansion (PFE)*: If $M \geq N$ in $H(z)$, then $\Delta(z)$ may be divided into $\Gamma(z)$ yielding quotient $Q(z)$ and remainder $R(z)$. Polynomial division produces a remainder with degree strictly less than that of the the divisor $\Delta(z)$. Thus

$$H(z) = Q(z) + \frac{R(z)}{\Delta(z)},$$

where the second term is the strictly proper part

$$H_s(z) = \frac{R(z)}{\Delta(z)}.$$

If $Q(z)$ is not a constant, then The strictly proper term may be further decomposed into the form

$$H_s(z) = \sum_{i=1}^{r} \sum_{k=1}^{N_i} \frac{A_{ik}}{(z - \rho_i)^k}. \tag{8.9}$$

The coefficients A_{ik} are given by the formula

$$A_{ik} = \frac{1}{(N_i - k)!} \frac{d^{N_i - k}}{dz^{N_i - k}} \left((z - \rho_i)^{N_i} H_s(z) \right) \Big|_{z=\rho_i}; \quad k = 1, \ldots, N_i; \quad i = 1, \ldots, r. \tag{8.10}$$

For $N_i = 1$,

$$A_{i1} = (z - \rho_i) H_s(z) \Big|_{z=\rho_i}.$$

The A_{ik} can also be found by multiplying through (8.9) by $\Delta(z)$ and equating the coefficients for each power of z.

Example 8.7 *Find the PFE of*

$$H(z) = \frac{1}{z^2 + \frac{5}{2}z + 1}.$$

No polynomial division is required. The poles are $\rho_{1,2} = -\frac{1}{2}, -2$, *both with multiplicity* 1. *Since none of the poles is repeated, no differentiation in (8.10) is required.*

$$A_{11} = \left(z + \frac{1}{2}\right) \frac{1}{\left(z + \frac{1}{2}\right)(z+2)} \Big|_{z=-\frac{1}{2}} = \frac{1}{z+2} \Big|_{z=-\frac{1}{2}} = \frac{2}{3}$$

$$A_{21} = (z+2) \frac{1}{\left(z + \frac{1}{2}\right)(z+2)} \Big|_{z=-2} = \frac{1}{z+\frac{1}{2}} \Big|_{z=-2} = -\frac{2}{3}$$

Alternatively, multiplication of (8.9) by $\left(z + \frac{1}{2}\right)(z+2)$ *yields*

$$A_{11}(z+2) + A_{21}\left(z + \frac{1}{2}\right) = 1,$$

$$A_{11} + A_{21} = 0, \quad 2A_{11} + \frac{1}{2}A_{21} = 1,$$

$$A_{11} = \frac{2}{3}, \quad A_{21} = -\frac{2}{3}.$$

The PFE is

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

It is also useful to consider the PFE of $H(z)$ in the form (8.6)-(8.7). An easy way to do so is to define $v = z^{-1}$ and $\widetilde{H}(v) = H(v^{-1})$ and use the techniques above. Here it is common practice to write the PFE as

$$\widetilde{H}(v) = \widetilde{Q}(v) + \widetilde{H}_s(v),$$

$$\widetilde{H}_s(v) = \sum_{i=1}^{r} \sum_{k=1}^{N_i} \frac{\widetilde{A}_{ik}}{(1 - \rho_i v)^k}.$$

The coefficients are

$$\widetilde{A}_{ik} = \frac{1}{(N_i - k)!} \left(-\frac{1}{\rho_i}\right)^{N_i - k} \frac{d^{N_i - k}}{dv^{N_i - k}} \left((1 - \rho_i v)^{N_i} \widetilde{H}(v) \right) \Big|_{v=\frac{1}{\rho_i}}; \quad k = 1, \ldots, N_i; \quad i = 1, \ldots, r,$$

where the ρ_i are still the poles of $H(z)$. For $N_i = 1$,

$$\widetilde{A}_{i1} = (1 - \rho_i v)\, \widetilde{H}(v)\Big|_{v=\frac{1}{\rho_i}}.$$

This approach facilitates the use of tables such as O&S Table 3.1 to find the inverse transform of $H(z)$.

Example 8.8 *For $H(z)$ as in Example 8.7, (8.6) has the form*

$$H(z) = \frac{z^{-2}}{1 + \frac{5}{2}z^{-1} + z^{-2}},$$

so

$$\widetilde{H}(v) = \frac{v^2}{1 + \frac{5}{2}v + v^2}.$$

By polynomial division,

$$\widetilde{H}(v) = 1 - \frac{1 + \frac{5}{2}v}{1 + \frac{5}{2}v + v^2}.$$

The coefficients are

$$A_{11} = -\left(1 + \frac{1}{2}v\right) \frac{1 + \frac{5}{2}v}{\left(1 + \frac{1}{2}v\right)(1 + 2v)}\Big|_{v=-2} = -\frac{1 + \frac{5}{2}v}{1 + 2v}\Big|_{v=-2} = \frac{4}{3},$$

$$A_{21} = -(1 + 2v) \frac{1 + \frac{5}{2}v}{\left(1 + \frac{1}{2}v\right)(1 + 2v)}\Big|_{v=-\frac{1}{2}} = -\frac{1 + \frac{5}{2}v}{1 + \frac{1}{2}v}\Big|_{v=-\frac{1}{2}} = -\frac{1}{3}.$$

The PFE is

$$\widetilde{H}(v) = 1 + \frac{\frac{4}{3}}{1 + \frac{1}{2}v} - \frac{\frac{1}{3}}{1 + 2v},$$

or

$$H(z) = \widetilde{H}\left(z^{-1}\right) = 1 + \frac{\frac{4}{3}}{1 + \frac{1}{2}z^{-1}} - \frac{\frac{1}{3}}{1 + 2z^{-1}}.$$

Table 3.1, p. 110, in O&S yields the inverse transforms:

Case I: *ROC* $= \operatorname{ann}(2, \infty)$
From line 5,

$$\frac{1}{1 + \frac{1}{2}z^{-1}} \longleftrightarrow \left(\frac{1}{2}\right)^n u[n], \quad \frac{1}{1 + 2z^{-1}} \longleftrightarrow 2^n u[n],$$

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

Case II: *ROC* $= \operatorname{ann}\left(\frac{1}{2}, 2\right)$
From lines 5 and 6,

$$\frac{1}{1 + \frac{1}{2}z^{-1}} \longleftrightarrow \left(\frac{1}{2}\right)^n u[n], \quad \frac{1}{1 + 2z^{-1}} \longleftrightarrow -2^n u[-n-1],$$

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

Case III: *ROC* $= D\left(0, \frac{1}{2}\right)$
From line 6,

$$\frac{1}{1+\frac{1}{2}z^{-1}} \longleftrightarrow -\left(\frac{1}{2}\right)^n u[-n-1], \quad \frac{1}{1+2z^{-1}} \longleftrightarrow -2^n u[-n-1],$$

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

Closely related to the PFE is the *Inner-Outer Decomposition*: If no pole of $H(z)$ lies on the unit circle then we may separate the poles into those inside the unit circle and those outside. The terms in the PFE may be grouped accordingly and recombined to yield the functions $H_i(z)$ with all poles inside the unit circle and $H_o(z)$ with all poles outside the unit circle. Then

$$H(z) = H_i(z) + H_o(z).$$

If the PFE is performed in z, we include the $Q(z)$ term in $H_o(z)$, since poles at ∞ lie outside the unit circle. If the PFE is done in z^{-1}, then $Q(z^{-1})$ has poles at 0 and is therefore grouped with $H_i(z)$.

Example 8.9 *In Example 8.7, $H(z)$ is already in inner-outer form:*

$$H_i(z) = \frac{\frac{2}{3}}{z+\frac{1}{2}}, \quad H_o(z) = -\frac{\frac{2}{3}}{z+2}.$$

In Example 8.8,

$$H_i(z) = 1 + \frac{\frac{4}{3}}{1+\frac{1}{2}z^{-1}} = \frac{\frac{7}{3}+\frac{1}{2}z^{-1}}{1+\frac{1}{2}z^{-1}},$$

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

8.6 Causality and Stability of Difference Equations

As we have seen, each difference equation leads to infinitely many LTI systems. The transfer function $H(z)$ corresponds to those which have a z-transform. If we are to use difference equations to implement a digital filter, the system must be BIBO stable. We may or may not be restricted to causal systems, depending on the application. The following results clarify the picture.

Theorem 8.2 *The family of all the LTI systems corresponding to a given difference equation contains at most one BIBO stable system. Such a system exists iff $H(z)$ has no pole on the unit circle.*

Suppose the condition of Theorem 8.2 holds. Then we may construct an (open) annulus containing the unit circle and none of the poles of $H(z)$. Let Λ be the largest such annulus. That is, Λ contains the unit circle, contains no pole of $H(z)$, and has at least one pole of $H(z)$ on each of its boundaries:

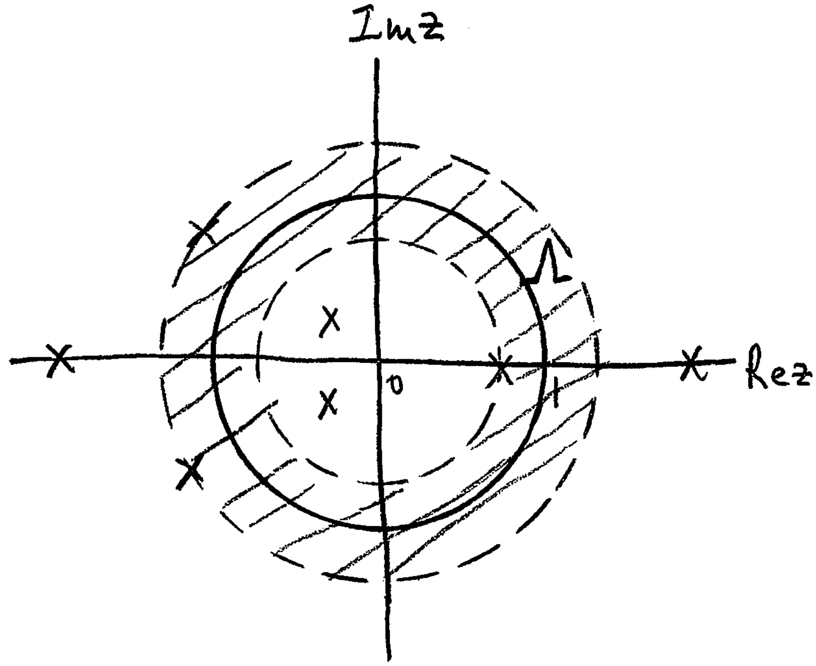


Figure 8.3

Theorem 8.3 *Suppose $H(z)$ has no pole on the unit circle, and consider the unique BIBO stable system determined by the difference equation.*
1) The impulse response of the system has ZT equal to $H(z)$ with ROC $= \Lambda$.
2) The system is causal iff $H(z)$ is proper and every pole of $H(z)$ satisfies $|\rho| < 1$.
3) The system is anti-causal iff every pole of $H(z)$ satisfies $|\rho| > 1$.

Note that the condition for stability is exactly the same as the condition for applying the inner-outer decomposition – i.e. no pole on the unit circle. Under this assumption, we may write

$$H(z) = H_i(z) + H_o(z).$$

The inner term $H_i(z)$ satisfies the conditions of Theorems 8.2 and 8.3 for stability and causality. The corresponding ROC is

$$\Lambda_i = \operatorname{ann}(R_i, \infty),$$

where R_i is the largest radius over the poles lying inside the unit circle. Similarly, $H_o(z)$ satisfies the conditions for stability and anti-causality. Its ROC is

$$\Lambda_o = D(0, R_o),$$

where R_o is the smallest radius over the poles lying outside the unit circle. From the linearity property of the ZT, $H(z)$ has ROC

$$\Lambda = \Lambda_i \cap \Lambda_o.$$

The next example shows how to find $h[n]$ using this decomposition.

Example 8.10 *For $H(z)$ as in Examples 8.7 and 8.8, neither pole lies on the unit circle, so there must be a stable system associated with the equation*

$$y[n+2] + \frac{5}{2}y[n+1] + y[n] = x[n].$$

Since

$$|\rho_1| = \frac{1}{2} < 1, \quad |\rho_2| = 2 > 1,$$

the stable system is neither causal nor anti-causal. The stable impulse response $h[n]$ *has transform* $H(z)$ *with ROC* $= \operatorname{ann}\left(\frac{1}{2}, 2\right)$. *Example 8.9 yields* $H_i(z)$ *and* $H_o(z)$. *The BIBO stable impulse response is calculated in Example 8.8, part II.*

We note that the impulse response calculated in Example 8.10 is infinite-duration. In fact, a casual glance at O&S Table 3.1 shows that none of the terms in a partial fraction expansion corresponding to a nonzero pole has an inverse transform which is finite-duration. From this we conclude that **every transfer function having at least one nonzero pole corresponds to an infinite-duration impulse response**. Equivalently, every difference equation with positive order has an infinite-duration impulse response. This fact will have major implications to the design of digital filters.

8.7 Choice of Initial Conditions

A given difference equation determines at most one BIBO LTI system. In applications, this is the system of interest. It should not be surprising that passing an input signal $x[n]$ through the system is equivalent to solving the difference equation for an appropriate choice of initial conditions. Thus the system may be implemented by encoding the difference equation in a computer program. In order to start the recursion, we need to choose initial conditions so that the equation generates solutions consistent with a BIBO stable system. The next result shows how to proceed.

Theorem 8.4 *Consider a difference equation with input* $x[n]$.
1) If $H(z)$ *is proper,* $x[n] = 0$ *for* $n < N_1$, *and* $y[N_1 - 1] = \ldots = y[N_1 - N + K] = 0$, *then the solution of the difference equation corresponds to a causal system.*
2) If $\rho = 0$ *is not a pole of* $H(z)$, $x[n] = 0$ *for* $n > N_2$, *and* $y[N_2 + 1] = \ldots = y[N_2 + N] = 0$, *then the solution of the difference equation corresponds to an anti-causal system.*

The assumptions on $x[n]$ in Theorem 8.4 are not a major concession. Part 1) applies to real-time applications. In this case, input data begins streaming at some initial time, which we call N_1. In non-real-time applications, both parts 1) and 2) are relevant. Here the input consists of a data file of finite length. We may identify the first entry of the file as $x[N_1]$ and the last entry as $x[N_2]$.

If the stable system determined by a difference equation is causal, then we apply part 1) of Theorem 8.4. For example, since $x[n] = \delta[n]$ has $N_1 = 0$, the causal impulse response may be calculated by setting $y[-1] = \ldots = y[-N + K] = 0$. If the system is anti-causal, then part 2) applies. For $x[n] = \delta[n]$, $N_2 = 0$ and we apply $y[1] = \ldots = y[N] = 0$. In either case, for an arbitrary input $x[n]$, the recursion produces the same output as the linear convolution

$$y[n] = h[n] * x[n].$$

If $H(z)$ has poles both inside and outside the unit circle, the stable system will be neither causal nor anti-causal. Here we must invoke the inner-outer decomposition

$$H(z) = H_i(z) + H_o(z)$$

to generate a pair of difference equations. These must be solved using the initial conditions described in Theorem 8.4, part 1) for $H_i(z)$ and part 2) for $H_o(z)$. Adding the two solutions together generates $y[n]$ corresponding to the stable system. The causal (inner) part requires forward recursion, while the anti-causal (outer) part requires backward recursion.

Example 8.11 *As in Examples 8.7-8.10, let*

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

This is already the inner-outer decomposition with

$$H_i(z) = \frac{\frac{2}{3}}{z + \frac{1}{2}}, \quad H_o(z) = -\frac{\frac{2}{3}}{z+2}.$$

$H_i(z)$ *determines the difference equation*

$$y_i[n+1] + \frac{1}{2}y_i[n] = \frac{2}{3}x[n],$$

while $H_o(z)$ *determines*

$$y_o[n+1] + 2y_o[n] = -\frac{2}{3}x[n].$$

Let $x[n] = w_4[n]$. *We want causality for the first equation, so we apply the initial condition* $y_i[-1] = 0$. *For the second, we want anti-causality, so we set* $y_o[4] = 0$. *Forward recursion of the first equation yields*

$$y_i[n] = -\frac{1}{2}y_i[n-1] + \frac{2}{3}w_4[n-1]$$

$$y_i[0] = -\frac{1}{2}y_i[-1] = 0$$

$$y_i[1] = -\frac{1}{2}y_i[0] + \frac{2}{3} = \frac{2}{3}$$

$$y_i[2] = -\frac{1}{2}y_i[1] + \frac{2}{3} = \frac{1}{3}$$

$$y_i[3] = -\frac{1}{2}y_i[2] + \frac{2}{3} = \frac{1}{2}$$

$$y_i[4] = -\frac{1}{2}y_i[3] + \frac{2}{3} = \frac{5}{12}$$

$$\vdots$$

with $y_i[n] = 0$ *otherwise. Backward recursion of the second equation yields*

$$y_o[n] = -\frac{1}{2}y_o[n+1] - \frac{1}{3}w_4[n]$$

$$y_o[3] = -\frac{1}{2}y_o[4] - \frac{1}{3} = -\frac{1}{3}$$

$$y_o[2] = -\frac{1}{2}y_o[3] - \frac{1}{3} = -\frac{1}{6}$$

$$y_o[1] = -\frac{1}{2}y_o[2] - \frac{1}{3} = -\frac{1}{4}$$

$$y_o[0] = -\frac{1}{2}y_o[1] - \frac{1}{3} = -\frac{5}{24}$$

$$y_o[-1] = -\frac{1}{2} y_o[0] = \frac{5}{48}$$
$$\vdots$$

with $y_o[n] = 0$ *otherwise. Adding the results,*

$$y[n] = y_i[n] + y_o[n] = \begin{cases} \vdots & \vdots \\ \frac{5}{48}, & n = -1 \\ -\frac{5}{24}, & n = 0 \\ \frac{5}{12}, & n = 1 \\ \frac{1}{6}, & n = 2 \\ \frac{1}{6}, & n = 3 \\ \frac{5}{12}, & n = 4 \\ \vdots & \vdots \end{cases}.$$

8.8 Zeroth-Order Difference Equations

Until now, we have restricted attention to difference equations with positive order – i.e. $N - K > 0$. If $N - K = 0$, the difference equation (8.1) becomes

$$y[n+N] = b_M x[n+M] + \ldots + b_0 x[n]. \tag{8.11}$$

The corresponding rational function is

$$H(z) = \frac{b_M z^M + \ldots + b_0}{z^N} = b_M z^{M-N} + \ldots + b_0 z^{-N}.$$

As always, we assume that no cancellation is possible between numerator and denominator. Thus $b_0 \neq 0$. Right-shifting (8.11) by N yields

$$y[n] = b_M x[n+M-N] + \ldots + b_0 x[n-N]. \tag{8.12}$$

Note that, for any input signal $x[n]$, (8.12) has exactly one solution $y[n]$. Initial conditions play no role in determining the solution of the equation.

Setting $x[n] = \delta[n]$ yields the impulse response

$$\begin{aligned} h[n] &= b_M \delta[n+M-N] + \ldots + b_0 \delta[n-N] \\ &= \begin{cases} b_{N-n}, & n = N-M, \ldots, N \\ 0, & \text{else} \end{cases} \end{aligned} \tag{8.13}$$

and transfer function

$$\mathcal{Z}\{h[n]\} = H(z).$$

If $N = 0$, $H(z)$ is a polynomial and the ROC is the entire complex plane. If $N > 0$, $ROC = \text{ann}(0, \infty)$. In contrast to systems with nonzero poles, **zeroth-order systems always determine finite-duration impulse responses**.

Stability and causality are easy to characterize for zeroth-order systems. Since the pole $z = 0$ lies inside the unit circle, **every zeroth-order difference equation determines a BIBO stable system**. If $N = 0$, expression (8.13) reduces to

$$h[n] = \begin{cases} b_{-n}, & n = -M, \ldots, 0 \\ 0, & \text{else} \end{cases}.$$

Since $h[n] = 0$ for $n > 0$, the system is anti-causal. If $M \le N$, (8.13) shows that $h[n] = 0$ for $n < 0$, so the system is causal. In all other cases ($M > N > 0$), the system is neither causal nor anti-causal. Here we could invoke the inner-outer decomposition

$$H_i(z) = \frac{b_{N-1}z^{N-1} + \ldots + b_0}{z^N},$$

$$H_o(z) = b_M z^{M-N} + \ldots + b_N,$$

but this has little use in applications.

9 Analog Filter Design

9.1 Introduction

The design of a digital filter usually begins with the choice of a CT transfer function $H(s)$. This approach is an historical artifact, owing to the fact that CT filter design is a mature subject, thoroughly developed during the first half of the 20th century. As we will see, converting from CT to DT is a simple matter, so it is not necessary to reinvent DT filter theory from scratch.

Most filter design is based on rational transfer functions. For analog filters, this is necessary because filters are built with electronic components: operational amplifiers, resistors, and capacitors. The physics of such devices dictate that circuits are governed by differential equations, which in turn lead to rational functions. For digital filters, rational functions correspond to difference equations, which may be solved recursively.

We begin by examining two important classes of CT filters.

9.2 The Butterworth Filter

The $2N$th roots of -1 are

$$\eta_k = \left(e^{j\pi}e^{j2\pi k}\right)^{\frac{1}{2N}} = e^{j\frac{1+2k}{2N}\pi}; \quad k = -N, \ldots, N-1.$$

Let

$$\lambda_k = j\eta_k = e^{j\frac{\pi}{2}}\eta_k = e^{j\frac{N+1+2k}{2N}\pi}.$$

We are interested in those λ_k for which $\operatorname{Re}\lambda_k < 0$. This requires

$$\frac{1}{2} < \frac{N+1+2k}{2N} < \frac{3}{2}$$

or

$$0 \le k \le N-1.$$

Define the degree N *Butterworth polynomial*

$$B_N(s) = s^N + a_{N-1}s^{N-1} + \ldots + a_0$$

to be the polynomial with roots $\lambda_0, \ldots, \lambda_{N-1}$. The cases $N = 4, 5$ are shown in Figure 9.1:

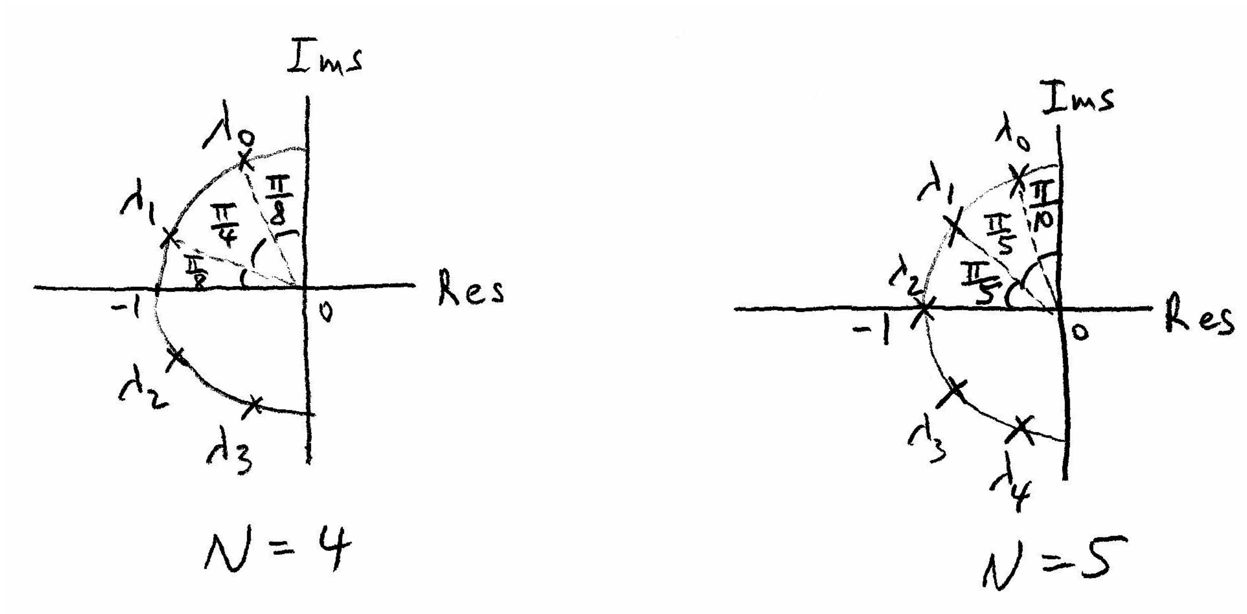


Figure 9.1

Coefficients of the first 5 Butterworth polynomials are provided below:

N	$B_N(s)$
1	$s+1$
2	$s^2+1.41s+1$
3	$s^3+2.00s^2+2.00s+1$
4	$s^4+2.61s^3+3.41s^2+2.61s+1$
5	$s^5+3.24s^4+5.24s^3+5.24s^2+3.24s+1$

Note that

$$\lambda_{k-N} = -\lambda_k; \quad k = 0, \ldots, N-1,$$

so the roots of $B_N(-s)$ are just those λ_k that lie in *RHP* (i.e. $k = -N, \ldots, -1$). It follows that the roots of $B_N(s)\,B_N(-s)$ are the complete set of λ_k. In other words,

$$B_N(s)\,B_N(-s) = (-js)^{2N} + 1 = (-1)^N s^{2N} + 1.$$

The Nth order *Butterworth LPF* is the (causal and stable) CT system with transfer function

$$H_N(s) = \frac{1}{B_N(s)}.$$

The magnitude frequency response of the filter (squared) is

$$\begin{aligned} |H_N(j\omega)|^2 &= H_N(j\omega)\,H_N^*(j\omega) \\ &= H_N(j\omega)\,H_N(-j\omega) \\ &= \frac{1}{B_N(j\omega)\,B_N(-j\omega)} \\ &= \frac{1}{(-1)^N (j\omega)^{2N} + 1} \\ &= \frac{1}{\omega^{2N}+1}, \end{aligned}$$

so

$$|H_N(j\omega)| = \frac{1}{\sqrt{\omega^{2N}+1}}.$$

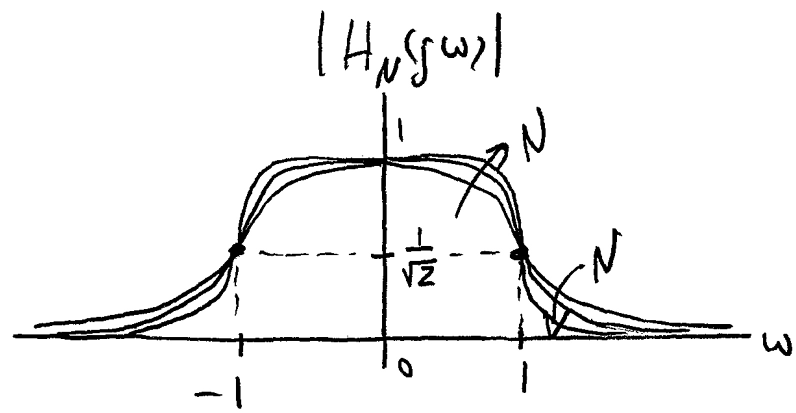


Figure 9.2

Figure 9.2 shows that $|H_N(j\omega)|$ converges to the magnitude of the ideal LPF as $N \to \infty$. $|H_N(j\omega)|$ has additional properties of interest. For example, it is straightforward to calculate

$$\frac{d}{d\omega}|H_N(j\omega)| = -N\omega^{2N-1}|H_N(j\omega)|^3,$$

which is negative for $\omega > 0$ and positive for $\omega < 0$. Hence, $|H_N(j\omega)|$ "rolls off" monotonically with frequency. Further differentiation shows that

$$\frac{d^k}{d\omega^k}|H_N(j\omega)|\Big|_{\omega=0} = 0; \quad k = 1, \ldots, 2N-1.$$

For this reason, $|H_N(j\omega)|$ is said to be "maximally flat". It is important to note that **only** the magnitude $|H_N(j\omega)|$ converges to the ideal LPF. The phase of $H_N(j\omega)$ rolls off from 0 to $-90N°$ asymptotically as $\omega \to \infty$, so taking $N \to \infty$ sends $\angle H_N(j\omega) \to -\infty$. In some applications, the phase is not an important issue, so approximating the ideal LPF in magnitude only is adequate. In other applications, this may not be acceptable.

9.3 The Chebyshev Filter

Consider the (Type I) *Chebyshev polynomials*

$$T_0(\omega) = 1,$$

$$T_1(\omega) = \omega,$$

$$T_{N+1}(\omega) = 2\omega T_N(\omega) - T_{N-1}(\omega).$$

Associated with these are the functions

$$\Gamma_N(s) = 1 + \varepsilon^2 T_N^2\left(\frac{s}{j}\right),$$

where $\varepsilon > 0$ is a design parameter. The $2N$ roots λ_k of Γ_N are symmetrically distributed around an ellipse with major axis equal to the imaginary axis:

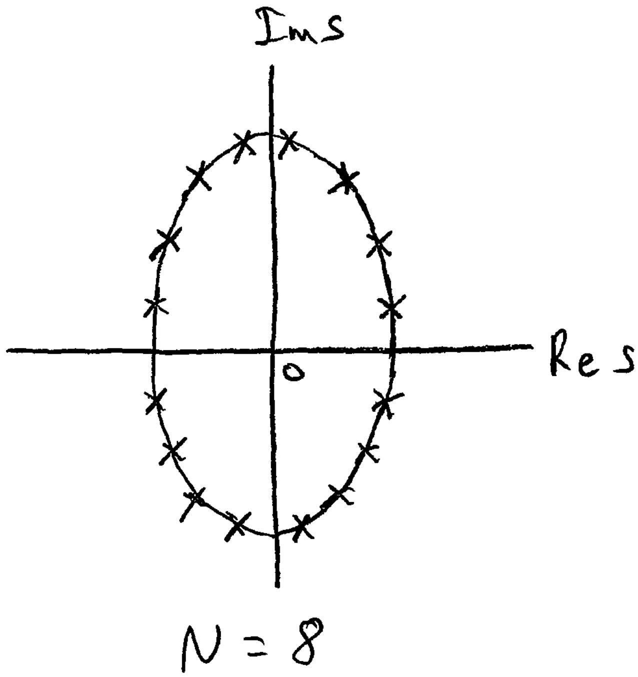


Figure 9.3

Let

$$C_N(s) = 2^{N-1}\varepsilon s^N + a_{N-1}s^{N-1} + \ldots + a_0$$

be the polynomial whose roots are those of $\Gamma_N(s)$ that lie in LHP. Then

$$\Gamma_N(s) = C_N(s)\,C_N(-s)\,.$$

The transfer function of the Nth-order *Chebyshev LPF* is

$$H_N(s) = \frac{1}{C_N(s)}.$$

The frequency response is

$$\begin{aligned} |H_N(j\omega)|^2 &= \frac{1}{C_N(j\omega)\,C_N(-j\omega)} \\ &= \frac{1}{\Gamma_N(j\omega)} \\ &= \frac{1}{1+\varepsilon^2 T_N^2(\omega)}, \end{aligned}$$

$$|H_N(j\omega)| = \frac{1}{\sqrt{1+\varepsilon^2 T_N^2(\omega)}}.$$

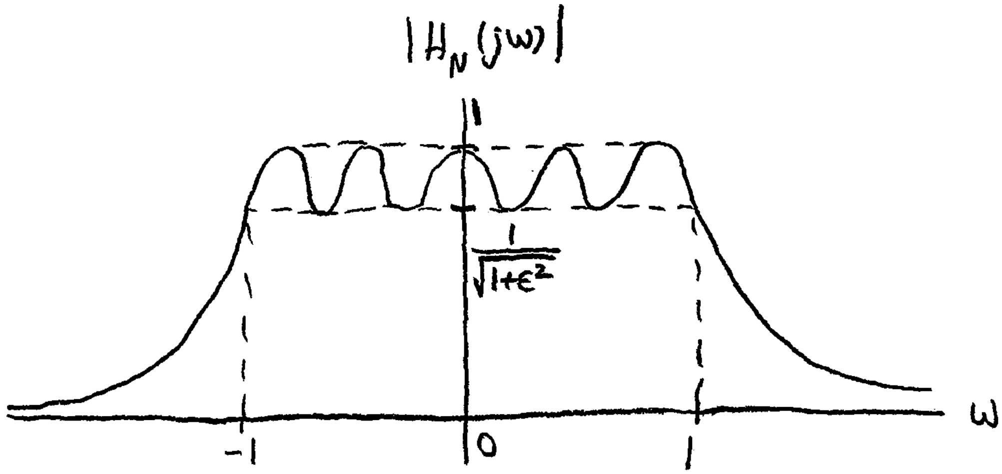


Figure 9.4

The Chebyshev filter has a steeper cutoff than the Butterworth filter for any value of N and is, therefore, a more efficient design. However, it suffers from "ripple" in the passband. This effect can be made arbitrarily small by choosing ε small, since the maximum deviation satisfies

$$\lim_{\varepsilon \to 0} \left(1 - \frac{1}{\sqrt{1+\varepsilon^2}} \right) = 0.$$

The trade-off here is that N must be increased as ε is decreased in order to maintain the filter bandwidth near $\omega_B = 1$. For the Chebyshev filter, taking $N \to \infty$ and $\varepsilon \to 0$ makes $|H_N(j\omega)|$ tend to the ideal LPF. However, $\angle H_N(j\omega) \to -\infty$ for each $\omega > 0$.

9.4 Causality

The issue of causality is just as important to analog filters as it is to digital filters. The theory is completely analogous: Differential equations replace difference equations, vertical strips in the plane replace annuli, and the imaginary axis replaces the unit circle. Suppose a rational transfer function $H(s)$ has no pole on the imaginary axis and Λ is the largest open vertical strip containing the axis but no pole of $H(s)$:

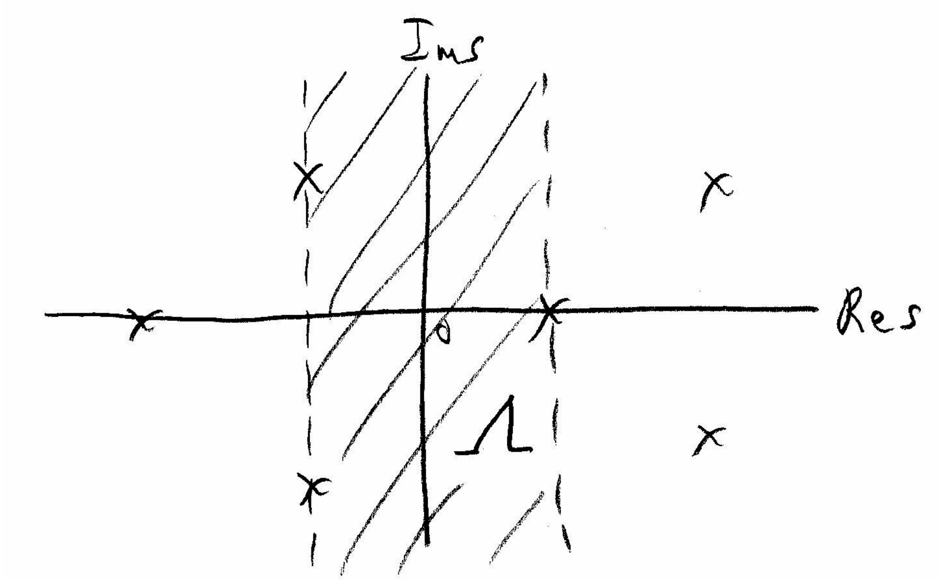


Figure 9.5

Then $H(s)$ along with $ROC = \Lambda$ determines a CT LTI system. Compare the following result with Theorem 8.3.

Theorem 9.1 *1) The system is BIBO stable iff $H(s)$ is proper.*
2) The system is causal iff every pole of $H(s)$ satisfies $\operatorname{Re}\lambda < 0$.
3) The system is anti-causal iff every pole of $H(s)$ satisfies $\operatorname{Re}\lambda > 0$.

For any stable CT system with rational $H(s)$, we may separate $H(s)$ into causal and anti-causal terms in a manner analogous to the inner-outer decomposition from DT analysis: Perform PFE and group terms with poles satisfying $\operatorname{Re}\lambda < 0$ and terms with $\operatorname{Re}\lambda > 0$. This yields the *left-right decomposition*

$$H(s) = H_l(s) + H_r(s).$$

$H_l(s)$ represents a causal stable system, while $H_r(s)$ represents an anti-causal stable system.

From Theorem 9.1, the Butterworth and Chebyshev filters are BIBO stable and causal. The anti-causal version of each can be constructed by reflecting the poles across the imaginary axis. Since the poles are already symmetric relative to the real axis, it suffices to map $s \longmapsto -s$. In other words, the Nth-order anti-causal Butterworth filter is

$$H_N(s) = \frac{1}{B_N(-s)}$$

and the Nth-order anti-causal Chebyshev filter is

$$H_N(s) = \frac{1}{C_N(-s)}.$$

We may also consider the $2N$th-order filters obtained by combining the poles of the causal and anti-causal cases. For Butterworth filters,

$$H(s) = \frac{1}{B_N(s)\,B_N(-s)} = \frac{1}{(-1)^N s^{2N} + 1}, \tag{9.1}$$

$$H(j\omega) = \frac{1}{\omega^{2N} + 1}.$$

For Chebyshev,

$$H(s) = \frac{1}{C_N(s)\,C_N(-s)} = \frac{1}{1 + \varepsilon^2 T_N^2\left(\frac{s}{j}\right)}, \tag{9.2}$$

$$H(j\omega) = \frac{1}{1 + \varepsilon^2 T_N^2(\omega)}.$$

9.5 Frequency Scaling, Highpass, and Bandpass Transformations

Let $H(s)$ be any filter with (roughly speaking) a bandwidth of 1 rad/sec. In practice, we will need to modify $H(s)$ in order to achieve certain design specifications. The simplest such modification is to set the bandwidth of the filter to a value other than unity. This is easily done by the *frequency scaling*

$$s \longmapsto \frac{s}{\omega_B},$$

which "stretches" the frequency response by ω_B. In other words, $H\left(\frac{s}{\omega_B}\right)$ has bandwidth ω_B, rather than 1. The pole locations are scaled outward by a factor of ω_B.

Example 9.1 *Design a causal 4th-order Butterworth LPF with bandwidth 5 Krad/sec. The 4th-order Butterworth polynomial is*

$$\begin{aligned} B_4(s) &= \left(s - e^{j\frac{5}{8}\pi}\right)\left(s - e^{j\frac{7}{8}\pi}\right)\left(s - e^{j\frac{9}{8}\pi}\right)\left(s - e^{j\frac{11}{8}\pi}\right) \\ &= \left(s^2 + \left(2\sin\frac{\pi}{8}\right)s + 1\right)\left(s^2 + \left(2\cos\frac{\pi}{8}\right)s + 1\right). \end{aligned}$$

Interpreting s as rad/sec, we need $\omega_B = 5000$. Then

$$\begin{aligned} B_4\left(\frac{s}{5000}\right) &= \left(\left(\frac{s}{5000}\right)^2 + \left(2\sin\frac{\pi}{8}\right)\left(\frac{s}{5000}\right) + 1\right)\left(\left(\frac{s}{5000}\right)^2 + \left(2\cos\frac{\pi}{8}\right)\left(\frac{s}{5000}\right) + 1\right) \\ &= \frac{s^2 + 3827s + 2.5\times 10^7}{2.5\times 10^7}\cdot\frac{s^2 + 9239s + 2.5\times 10^7}{2.5\times 10^7}, \end{aligned}$$

so

$$H\left(\frac{s}{5000}\right) = \frac{1}{B_4\left(\frac{s}{5000}\right)} = \frac{6.25\times 10^{14}}{\left(s^2 + 3827s + 2.5\times 10^7\right)\left(s^2 + 9239s + 2.5\times 10^7\right)}.$$

The poles lie on the circle centered at 0 with radius 5 Krad/sec.
It is sometimes convenient to reinterpret s as Krad/sec, Mrad/sec, etc. in order to make the coefficients easier to work with. In our example, we can write the transfer function with s in Krad/sec. This amounts to setting $\omega_B = 5$:

$$H\left(\frac{s}{5}\right) = \frac{625}{\left(s^2 + 3.827s + 25\right)\left(s^2 + 9.239s + 25\right)}.$$

In some applications it is desired to block components of a signal $x(t)$ **below** a certain frequency ω_0. Define the *ideal highpass filter (HPF)* by

$$H_{HP}(j\omega) = \begin{cases} 1, & |\omega| \geq \omega_0 \\ 0, & |\omega| < \omega_0 \end{cases}.$$

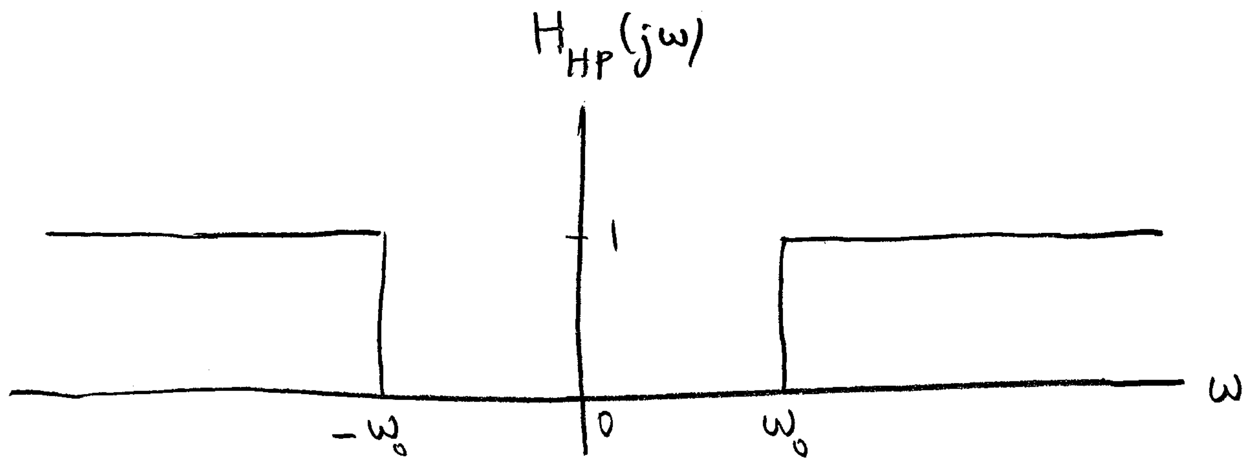


Figure 9.6

The interval $[0, \omega_0]$ is called the *stopband*.

Consider the frequency map

$$s \longmapsto \frac{\omega_0}{s}$$

and the transformed filter $H\left(\frac{\omega_0}{s}\right)$. Specializing to $s = j\omega$ yields

$$\omega \longmapsto -\frac{\omega_0}{\omega},$$

$$\pm\omega_0 \longmapsto \mp 1,$$

$$|\omega| > \omega_0 \iff \left|-\frac{\omega_0}{\omega}\right| < 1.$$

Hence, if $|H(j\omega)|$ approximates $H_{LP}(j\omega)$, then $\left|H\left(-j\frac{\omega_0}{\omega}\right)\right|$ approximates $H_{HP}(j\omega)$.

Example 9.2 *Design a causal 4th-order Butterworth HPF with stopband* $[0,5]$ *Krad/sec. Interpreting* s *as Krad/sec, we need* $\omega_0 = 5$. *Then*

$$\begin{aligned} B_4\left(\frac{5}{s}\right) &= \left(\left(\frac{5}{s}\right)^2 + \left(2\sin\frac{\pi}{8}\right)\left(\frac{5}{s}\right) + 1\right)\left(\left(\frac{5}{s}\right)^2 + \left(2\cos\frac{\pi}{8}\right)\left(\frac{5}{s}\right) + 1\right) \\ &= \frac{s^2 + \left(10\sin\frac{\pi}{8}\right)s + 25}{s^2} \cdot \frac{s^2 + \left(10\cos\frac{\pi}{8}\right)s + 25}{s^2}, \end{aligned}$$

$$H\left(\frac{5}{s}\right) = \frac{1}{B_4\left(\frac{5}{s}\right)} = \frac{s^4}{\left(s^2 + 3.827 + 25\right)\left(s^2 + 9.239s + 25\right)}.$$

Sometimes it is required to design a filter which passes components of a signal over some interval of frequencies and attenuates the signal elsewhere. For example the *ideal bandpass filter (BPF)* is given by

$$H_{BP}(j\omega) = \begin{cases} 1, & \omega_1 \leq |\omega| \leq \omega_2 \\ 0, & \text{else} \end{cases},$$

where $0 < \omega_1 < \omega_2 < \infty$.

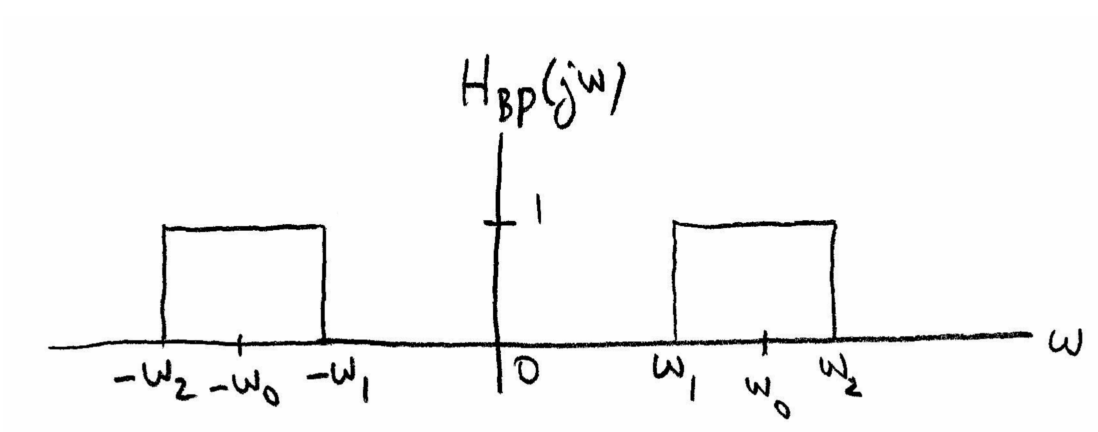


Figure 9.7

The filter has *bandwidth*

$$\omega_B = \omega_2 - \omega_1$$

and *center frequency*

$$\omega_0 = \sqrt{\omega_1\omega_2}.$$

The interval $[\omega_1, \omega_2]$ is called the *passband*.

Consider the frequency map

$$s \longmapsto \beta(s) = \frac{\omega_0}{\omega_B}\left(\frac{s}{\omega_0} + \frac{\omega_0}{s}\right)$$

and the transformed filter $H(\beta(s))$. Writing

$$\beta(s) = \frac{s^2 + \omega_0^2}{\omega_B s}$$

exposes the fact that $H(\beta(s))$ has order $2N$. Specializing to $s = j\omega$, we obtain

$$\omega \longmapsto \frac{\omega_0}{\omega_B}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right),$$

$$\omega_0 \longmapsto 0,$$

$$\pm\omega_1 \longmapsto \frac{\omega_0}{\omega_B}\left(\pm\frac{\omega_1}{\omega_0} \mp \frac{\omega_0}{\omega_1}\right) = \mp 1,$$

$$\pm\omega_2 \longmapsto \frac{\omega_0}{\omega_B}\left(\frac{\omega_2}{\omega_0} \mp \frac{\omega_0}{\omega_2}\right) = \pm 1,$$

$$\omega_1 < |\omega| < \omega_2 \iff |\beta(j\omega)| < 1.$$

If $|H(j\omega)|$ is approximates $H_{LP}(j\omega)$, then $|H(\beta(j\omega))|$ approximates $H_{BP}(j\omega)$.

Example 9.3 *Design a causal 8th-order Butterworth BPF with passband* $[5, 10]$ *Krad/sec. Interpreting* s *as Krad/sec, we merely need to calculate*

$$H(s) = \frac{1}{B_4(\beta(s))}$$

with

$$\omega_0 = \sqrt{\omega_1\omega_2} = \sqrt{50} = 5\sqrt{2},$$

$$\omega_B = \omega_2 - \omega_1 = 5.$$

$$\begin{aligned}
B_4(\beta(s)) &= B_4\left(\frac{s^2+50}{5s}\right) \\
&= \left(\left(\frac{s^2+50}{5s}\right)^2 + \left(2\sin\frac{\pi}{8}\right)\left(\frac{s^2+50}{5s}\right) + 1\right)\left(\left(\frac{s^2+50}{5s}\right)^2 + \left(2\cos\frac{\pi}{8}\right)\left(\frac{s^2+50}{5s}\right) + 1\right) \\
&= \frac{(s^4 + 5.412s^3 + 150s^2 + 270.6s + 2500)(s^4 + 13.07s^3 + 150s^2 + 653.3s + 2500)}{625s^4}
\end{aligned}$$

$$H(s) = \frac{625s^4}{(s^4 + 3.827s^3 + 125s^2 + 191.3s + 2500)(s^4 + 9.239s^3 + 125s^2 + 461.9s + 2500)}.$$

9.6 Zero Phase Filters

A filter $H(j\omega) \neq 0$ has *zero phase* if $\angle H(j\omega) = 0$ for all ω. This is the same as saying that $H(j\omega)$ is real and nonnegative for all ω or, equivalently,

$$H(j\omega) = |H(j\omega)|. \tag{9.3}$$

From the even-odd property of the CTFT, a zero phase filter has an even impulse response $h(t)$. If, in addition, the filter is causal, then $h(t) = 0$ for $|t| > 0$. This implies $h(t) = \alpha\delta(t)$ for some constant α. A similar analysis holds for anti-causal zero phase filters. These facts may be summaraized as follows:

Theorem 9.2 *If a CT filter has zero phase and is either causal or anti-causal, then $H(j\omega)$ is constant.*

Unfortunately, $h(t)$ being even does not imply zero phase:

Example 9.4 *Let*

$$h(t) = w\left(\frac{t+1}{2}\right) - \begin{cases} 1, & |t| \leq 1 \\ 0, & else \end{cases}.$$

Note that $h(t)$ is even, but

$$H(j\omega) = 2\operatorname{sinc}\frac{\omega}{\pi}$$

is negative for certain ω, violating (9.3). In this example, $\angle H(j\omega)$ oscillates between 0 and $180°$.

Example **??** demonstrates that, in general, one must check $H(j\omega)$ directly to determine whether the a filter actually has zero phase. However, for rational functions $H(s)$, the zero phase property has a simple characterization in terms of poles and zeros: For any real number λ, let

$$Q_\lambda(s) = -\left(s^2 - \lambda^2\right).$$

The roots of $Q_\lambda(s)$ are $\pm\lambda$. Furthermore,

$$Q_\lambda(j\omega) = \omega^2 + \lambda^2,$$

so

$$\angle Q_\lambda(j\omega) = 0$$

for all ω. Now let λ be complex and set

$$Q_\lambda(s) = s^4 - 2\left(\operatorname{Re}\lambda^2\right)s^2 + |\lambda|^4.$$

The roots of $Q_\lambda(s)$ are the 4 symmetrically positioned complex numbers $\pm\lambda$ and $\pm\lambda^*$:

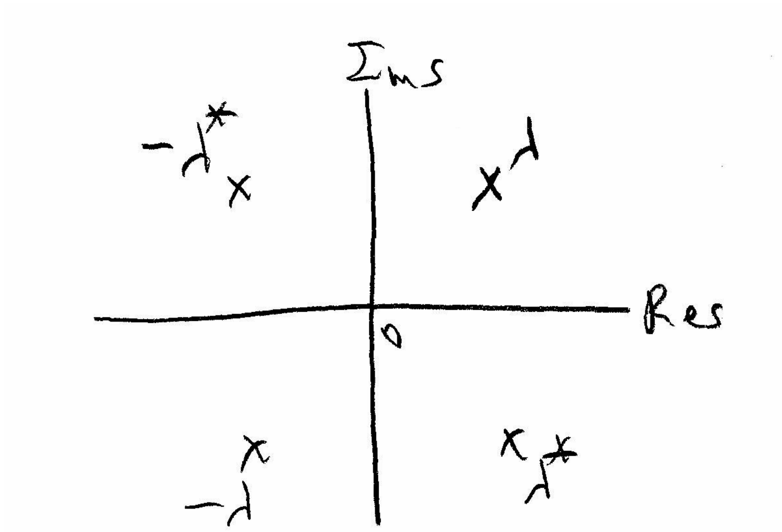


Figure 9.8

Setting $s = j\omega$,

$$Q_\lambda(j\omega) = \omega^4 + 2\left(\operatorname{Re}\lambda^2\right)\omega^2 + |\lambda|^4$$

is real. Since

$$|\lambda|^2 = \left|\lambda^2\right| \geq -\operatorname{Re}\lambda^2,$$

we obtain

$$Q_\lambda(j\omega) \geq \omega^4 - 2|\lambda|^2\omega^2 + |\lambda|^4 = \left(\omega^2 - |\lambda|^2\right)^2,$$

so

$$\angle Q_\lambda(j\omega) = 0.$$

More generally, suppose $H(s)$ has poles and zeros which may grouped into pairs and 4-tuples as described above. That is,

$$H(s) = A\frac{\prod Q_{\zeta_i}(s)}{\prod Q_{\lambda_i}(s)}. \tag{9.4}$$

If A is real and positive, then

$$\angle H(j\omega) = \angle A + \sum \angle Q_{\zeta_i}(j\omega) - \sum \angle Q_{\lambda_i}(j\omega) = 0,$$

so $H(s)$ has zero phase. A moment's reflection reveals that the form (9.4) is equivalent to having the poles and zeros of $H(s)$ symmetric with respect to both axes. In the case of imaginary poles or zeros, symmetry reduces to even multiplicity.

Theorem 9.3 *A CT rational transfer function $H(s) \neq 0$ has zero phase iff $H(j\omega)$ is real and positive for small ω and the poles and zeros of $H(s)$ are symmetric relative to both the real and imaginary axes.*

The $2N$th order noncausal Butterworth and Chebyshev filters (9.1) and (9.2) have $H(0) > 0$, no zeros, and the required symmetry for poles. Hence, these filters have zero phase.

Example 9.5 *Design an 8th-order zero phase Butterworth LPF with bandwidth* 5 *Krad/sec. Here we use the form (9.1) with* $N = 4$ *and perform a frequency scaling:*

$$H(s) = \frac{1}{(-1)^4 s^8 + 1} = \frac{1}{s^8 + 1},$$

$$H\left(\frac{s}{5}\right) = \frac{1}{\left(\frac{s}{5}\right)^8 + 1} = \frac{390625}{s^8 + 390625}.$$

In applications where approximating an ideal LPF, BPF, or HPF in magnitude only is acceptable, a nonzero phase phase filter will work. In non-real-time applications, approximation in both magnitude and phase is feasible using a zero phase filter. Since frequency scaling, highpass, and bandpass transformations merely reassign the values of $H(j\omega)$ to different frequencies, the zero phase property is not affected by such modifications of $H(j\omega)$.

9.7 Phase Delay, Linear Phase, and Phase Distortion

Consider a single-frequency signal

$$x(t) = e^{j(\omega_0 t + \phi)}.$$

If we write $\phi = -\omega_0 \tau$, we obtain

$$x(t) = e^{j\omega_0 (t - \tau)}.$$

In other words, the phase shift ϕ is equivalent to a time delay

$$\tau = -\frac{\phi}{\omega_0}.$$

For an arbitrary signal $x(t)$, the phase $\angle X(j\omega)$ varies with frequency. We define the *phase delay* to be

$$\Delta(\omega) = -\frac{\angle X(j\omega)}{\omega}.$$

Suppose $\angle X(j\omega)$ is a linear function of frequency

$$\angle X(j\omega) = -\tau\omega,$$

where τ is a constant. Then the phase delay is

$$\Delta(\omega) = -\frac{\angle X(j\omega)}{\omega} = \tau.$$

In other words, each sinusoid that makes up the signal is shifted right by τ, so **linear phase shift corresponds to a time-delay (or advance) of the signal**.

Now consider systems. We say an LTI system with transfer function $H(s)$ has *linear phase* if

$$\angle H(j\omega) = -\omega\tau$$

for some constant τ. In this case, let

$$H_1(s) = H(s) e^{s\tau}.$$

Then

$$\angle H_1(j\omega) = \angle \left(H(j\omega) e^{j\omega\tau}\right) = \angle H(j\omega) + \omega\tau = 0$$

and

$$H(s) = H_1(s)\, e^{-s\tau}.$$

From the time-shift property of the LT, the impulse responses satisfy

$$h(t) = h_1(t-\tau).$$

We conclude that **a linear phase filter is just a zero phase filter with a time-shift**. Since the impulse response of a zero phase filter is even, the impulse response of a linear phase filter is the shift of an even function.

Rational transfer functions $H(s)$ generally have a phase $\angle H(j\omega)$ that is nonlinear. This can be seen by writing $H(s)$ in terms of poles and zeros:

$$H(s) = A\frac{\prod (s-\zeta_i)}{\prod (s-\rho_i)}.$$

Then

$$\begin{aligned}\angle H(j\omega) &= \angle A + \sum \angle (j\omega - \zeta_i) - \sum \angle (j\omega - \rho_i) \\ &= \angle A - \sum \arctan\left(\frac{\omega - \operatorname{Im}\zeta_i}{\operatorname{Re}\zeta_i}\right) + \sum \arctan\left(\frac{\omega - \operatorname{Im}\rho_i}{\operatorname{Re}\rho_i}\right),\end{aligned}$$

which is a nonlinear function, except when poles and zeros satisfy the kind of symmetry discussed above. One can show that for CT systems, **every rational function with linear phase has zero phase**.

Suppose a system has nonlinear phase. Then the phase delay $\Delta(\omega)$ is not constant. This means that at some frequencies the system imposes more delay than at others. This effect is called *phase distortion*. Phase distortion is a serious problem in applications such as high-speed communication networks, where the "shape" of signals must be maintained. It has also been argued that phase distortion causes problems in audio systems. This claim is still controversial, except in unrealistically high-order filters.

The next example shows how signal shape can be corrupted by nonlinear phase.

Example 9.6 *Let*

$$H(s) = -\frac{s-1}{s+1}.$$

Then

$$|H(j\omega)| = \left|-\frac{j\omega - 1}{j\omega + 1}\right| = \frac{\omega^2+1}{\omega^2+1} = 1,$$

so any change in signals passing through the system is due entirely to phase distortion. Systems with $|H(j\omega)| = 1$ are called "all-pass". The phase

$$\angle H(j\omega) = \angle\left(-\frac{j\omega - 1}{j\omega + 1}\right) = -2\arctan\omega$$

is nonlinear and

$$\Delta(\omega) = 2\frac{\arctan\omega}{\omega}.$$

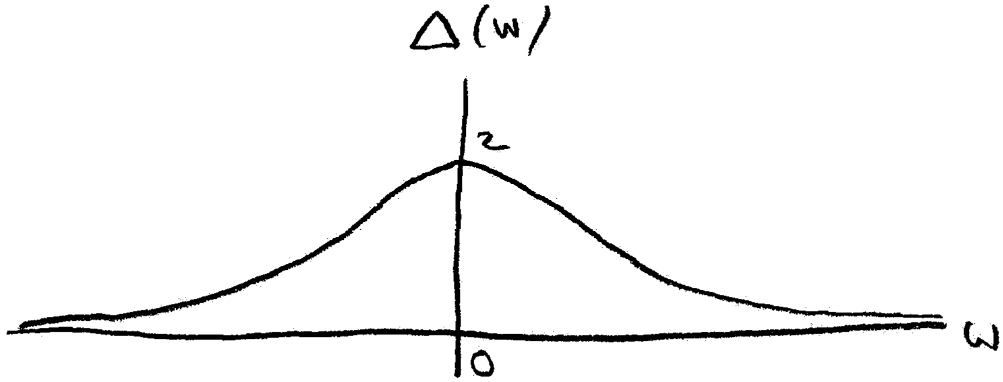


Figure 9.9

Low frequencies are delayed more than high frequencies. To see what phase distortion looks like, consider

$$x(t) = u(t).$$

Then

$$\begin{aligned} Y(s) &= H(s)X(s) \\ &= -\frac{s-1}{s+1} \cdot \frac{1}{s} \\ &= -\frac{2}{s+1} + \frac{1}{s}, \end{aligned}$$

$$y(t) = \left(1 - 2e^{-t}\right) u(t).$$

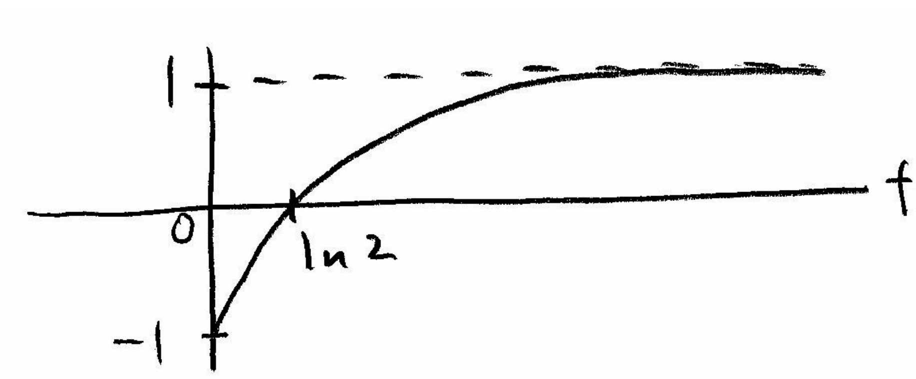


Figure 9.10

For DT filters $H\left(e^{j\Omega}\right)$, phase delay is the same as in CT:

$$\Delta(\Omega) = -\frac{\angle H\left(e^{j\Omega}\right)}{\Omega}.$$

Thus linear phase means

$$\angle H\left(e^{j\Omega}\right) = -\Omega\tau.$$

Although the definitions of linear phase in CT and DT are the same, there are a couple of important differences:

1) Since

$$e^{j\angle H\left(e^{j\Omega}\right)} = e^{-j\Omega\tau}$$

must have period 2π,

$$-\left(\Omega + 2\pi\right)\tau = -\Omega\tau - 2\pi N$$

for some integer N. Hence, $\tau = N$. This is merely a reflection of the fact that delay in DT systems must be integer.

2) A DT filter with nonzero linear phase can have a rational transfer function. For example, the system that delays the input by N time steps has transfer function

$$H\left(z\right) = z^{-N}.$$

10 IIR Filters

10.1 Conversion of CT to DT Filters

We now have methods for designing CT low-pass, high-pass, and band-pass filters. These can be either causal or zero phase. The next step is to convert the CT filter to DT. The simplest way to do this utilizes a function $\phi\left(z\right)$ which maps the z-plane into the s-plane:

$$s = \phi\left(z\right).$$

Ideally, $\phi\left(z\right)$ should satisfy 3 properties:

1) ϕ is a $1-1$ map between the s and z planes.
2) $\phi\left(z\right)$ maps the unit circle onto the imaginary axis and the unit disk onto LHP.
3) ϕ is a rational function.

In view of 1), $\phi\left(z\right)$ has an inverse

$$z = \phi^{-1}\left(s\right).$$

Applying $\phi\left(z\right)$ to the CT filter $H_{CT}\left(s\right)$ produces the DT filter

$$H_{DT}\left(z\right) = H_{CT}\left(\phi\left(z\right)\right).$$

The poles of $H_{DT}\left(z\right)$ are just $\rho = \phi^{-1}\left(\lambda\right)$, where λ ranges over the poles of $H_{CT}\left(s\right)$. (The same holds for zeros.) Property 2) ensures that $\phi\left(z\right)$ preserves BIBO stability, causality, and anti-causality. Property 3) ensures that $H_{DT}\left(z\right)$ is a rational function whenever $H_{CT}\left(s\right)$ is rational. This guarantees that we can implement the digital filter by encoding difference equations.

A common choice of $\phi\left(z\right)$ is the *bilinear transformation*

$$\phi\left(z\right) = \frac{2}{T}\frac{z-1}{z+1}, \tag{10.1}$$

where $T > 0$ will be the sampling period. The inverse is

$$\phi^{-1}\left(s\right) = \frac{1+\frac{T}{2}s}{1-\frac{T}{2}s}, \tag{10.2}$$

proving 1). Writing z in polar form, we obtain the rectangular form of $\phi(z)$:

$$\begin{aligned}
\phi\left(re^{j\Omega}\right) &= \frac{2}{T}\frac{re^{j\Omega}-1}{re^{j\Omega}+1} \\
&= \frac{2}{T}\frac{\left(re^{j\Omega}-1\right)\left(re^{-j\Omega}+1\right)}{\left|re^{j\Omega}+1\right|^2} \\
&= \frac{2}{T}\frac{r^2-1+j2r\sin\Omega}{\left|re^{j\Omega}+1\right|^2},
\end{aligned}$$

so

$$\sigma = \operatorname{Re}\phi\left(re^{j\Omega}\right) = \frac{2}{T}\frac{r^2-1}{\left|re^{j\Omega}+1\right|^2}.$$

Note that $\sigma = 0$ iff $r = 1$, so $\phi(z)$ maps the unit circle onto the imaginary axis. Also, $\sigma < 0$ iff $r < 1$, so $\phi(z)$ maps the unit disk onto LHP, establishing 2). Property 3) is obvious. Converting from CT to DT using the bilinear transformation is called *Tustin's method.*

In filter design, we are interested in frequency response $H_{CT}(j\omega)$ and $H_{DT}\left(e^{j\Omega}\right)$. Restricting to $s = j\omega$ and $z = e^{j\Omega}$, the bilinear transformation becomes

$$\begin{aligned}
\omega &= -j\phi\left(e^{j\Omega}\right) \\
&= \frac{4}{T}\frac{\sin\Omega}{\left|e^{j\Omega}+1\right|^2} \\
&= \frac{2}{T}\frac{\sin\Omega}{1+\cos\Omega} \\
&= \frac{2}{T}\tan\frac{\Omega}{2}.
\end{aligned} \tag{10.3}$$

Note that (10.3) and, therefore,

$$H_{DT}\left(e^{j\Omega}\right) = H_{CT}\left(j\frac{2}{T}\tan\frac{\Omega}{2}\right)$$

have period 2π. The inverse is

$$\Omega = 2\arctan\frac{\omega T}{2}. \tag{10.4}$$

In cases where the CT filter has transfer function defined only for $s = j\omega$, applying the bilinear transformation requires (10.3) rather than (10.1).

Example 10.1 *Applying (10.3) to the CT ideal LPF* $H_{LP}\left(j\frac{\omega}{\omega_B}\right)$ *with bandwidth* ω_B *yields the DT filter in Figure 10.1:*

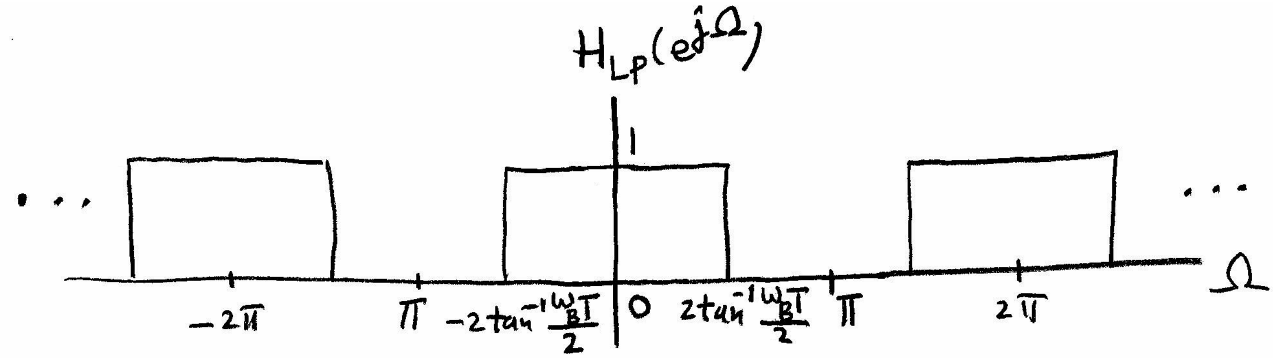


Figure 10.1

The CT ideal BPF converts to Figure 10.2:

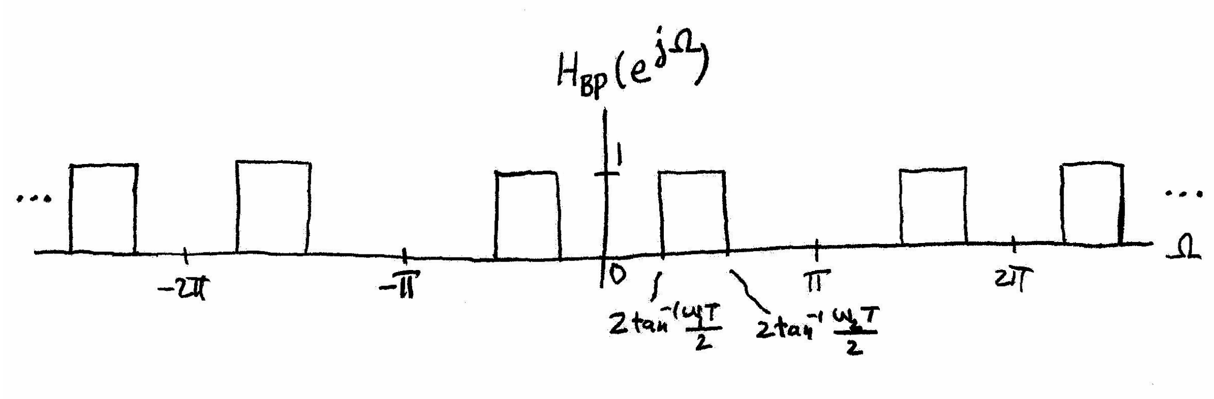


Figure 10.2

One might be troubled by the fact that we now have two very different transformations from ω to Ω. These are $\Omega = \omega T$, which results from sampling, and the bilinear transformation (10.4). It is important not to confuse the meaning and usage of these transformations. Sampling is the physical (and mathematical) process of converting **signals** from CT to DT. Tustin's method is one of many mathematical operations that convert **systems** from CT to DT.

This difference is made clearer if we apply each transformation to a single CT filter. First, sample the filter impulse response $h(t)$ to produce $h(nT)$. Then convert the transfer function $H_{CT}(s)$ to $H_{DT}(z)$ via Tustin's method, and find the impulse response $h[n]$ of $H_{DT}(z)$. In virtually all cases, it will turn out that $h(nT) \neq h[n]$.

Example 10.2 *Let*

$$H_{CT}(s) = \frac{1}{s+1}.$$

The CT impulse response is

$$h(t) = e^{-t}u(t).$$

For $T = 2$, sampling yields

$$h(nT) = e^{-2n}u[n].$$

Applying Tustin's method,

$$H_{DT}(z) = H_{CT}\left(\frac{z-1}{z+1}\right) = \frac{1}{2}\left(1 + z^{-1}\right)$$

and

$$h[n] = \frac{1}{2}(\delta[n] + \delta[n-1]).$$

Note that

$$h(nT) \neq h[n]$$

for every $n \geq 0$.

The following *Filter Design Algorithm* should help keep things straight:

1) Identify the desired bandwidth or passband in CT frequency ω.
2) Apply $\Omega = \omega T$ to determine the corresponding bandwidth Ω_B or passband $[\Omega_1, \Omega_2]$.

3) Convert to the CT bandwidth ω_B or passband $[\omega_1, \omega_2]$ by applying the bilinear formula (10.3).
4) Choose a CT filter design method (Butterworth, Chebyshev, etc.) and associated parameters (N, ε) to meet frequency and other specifications in CT.
5) Transform the CT filter to DT using the bilinear formula (10.1) and graph $H_{DT}\left(e^{j\omega T}\right)$.
6) If the specifications on $H_{DT}\left(e^{j\omega T}\right)$ are not met or if the filter is over-designed, modify N accordingly and repeat 5).

Under Tustin's method, the two frequency responses $H_{CT}(j\omega)$ and $H_{DT}\left(e^{j\omega T}\right)$ will be similar over $0 \leq \omega \leq \frac{\omega_s}{2}$ but not quite identical because of the distortion of the frequency axis described by

$$\omega \longmapsto \frac{2}{T}\tan\left(\frac{T}{2}\omega\right). \tag{10.5}$$

This is a consequence of the nonlinear nature of the bilinear transformation. Steps 2) and 3) of the filter design algorithm are constructed to cancel out this effect, at least at the critical cutoff frequencies. Applying this scheme with modern computers and software such as MATLAB, it is easy to iteratively adjust the design parameters N and ε, view the resulting graph, and ultimately converge to acceptable values.

Example 10.3 *Design an 8th-order, causal, digital Butterworth BPF for sampling rate* 44.1 *KHz with passband* $[10, 15]$ *KHz. The DT passband is*

$$[\Omega_1, \Omega_2] = \left[\frac{20\pi}{44.1}, \frac{30\pi}{44.1}\right] = [1.425, 2.137].$$

From (10.3), the modified CT passband is

$$[\omega_1, \omega_2] = \left[\frac{2}{T}\tan\frac{\Omega_1}{2}, \frac{2}{T}\tan\frac{\Omega_2}{2}\right] = [76.18, 160.6].$$

Note that the CT frequencies have increased substantially, due to (10.5):

$$[f_1, f_2] = [12.12, 25.56] \ \mathit{KHz}$$

KHz. Working as in Example 9.3, we obtain the CT filter

$$H_{CT}(s) = \frac{5.079 \times 10^7 s^4}{\left(\begin{array}{c} s^8 + 220.6s^7 + 7.327 \times 10^4 s^6 + 9.669 \times 10^6 s^5 + 1.544 \times 10^9 s^4 \\ +1.183 \times 10^{11} s^3 + 1.097 \times 10^{13} s^2 + 4.039 \times 10^{14} s + 2.240 \times 10^{16} \end{array}\right)}$$

to DT with

$$T = \frac{1}{44.1} = 0.02268 \ ms.$$

Invoking Tustin's method,

$$\phi(z) = 88.2\frac{z-1}{z+1},$$

$$\begin{aligned} H_{DT}(z) &= \frac{5.079 \times 10^7 \phi^4}{\left(\begin{array}{c} s^8 + 220.6\phi^7 + 7.327 \times 10^4 \phi^6 + 9.669 \times 10^6 \phi^5 + 1.544 \times 10^9 \phi^4 \\ +1.183 \times 10^{11} \phi^3 + 1.097 \times 10^{13} \phi^2 + 4.039 \times 10^{14} \phi + 2.240 \times 10^{16} \end{array}\right)} \\ &= \frac{7.374 \times 10^{-3}\left(z^2-1\right)^4}{z^8 - 1.370z^7 + 2.871z^6 - 2.493z^5 + 2.848z^4 - 1.566z^3 + 1.136z^2 - 0.3244z + 0.1481} \end{aligned}$$

Since (10.3) merely reassigns values of $H_{CT}(j\omega)$ to different frequencies to produce $H_{DT}\left(e^{j\Omega}\right)$, the zero phase property is preserved under the bilinear transformation. In both CT and DT, zero phase is equivalent to evenness of the impulse response, so the bilinear transformation takes even $h(t)$ into even $h[n]$. As in CT, **the only causal zero phase filters are the constants**.

It is worth noting that the bilinear transformation does not preserve **linear** phase:

Example 10.4 *Consider the unit delay system*

$$H_{CT}(s) = e^{-sT}.$$

The phase is

$$\angle H_{CT}(j\omega) = \angle e^{-j\omega T} = -\omega T,$$

so the system has linear, but nonzero phase. Applying the bilinear transformation,

$$H_{DT}\left(e^{j\Omega}\right) = H_{CT}\left(j\frac{2}{T}\tan\frac{\Omega}{2}\right) = e^{-j2\tan\frac{\Omega}{2}},$$

which has nonlinear phase

$$\angle H_{DT}\left(e^{j\Omega}\right) = -2\tan\frac{\Omega}{2}.$$

For rational functions, we note that the bilinear transformation maps CT poles and zeros into DT poles and zeros according to

$$\rho = \frac{1+\frac{T}{2}\lambda}{1-\frac{T}{2}\lambda}. \tag{10.6}$$

In the zero phase case, poles and zeros are symmetric about the real and imaginary axes. Note that symmetry about the real axis is retained under (10.6):

$$\frac{1+\frac{T}{2}\lambda^*}{1-\frac{T}{2}\lambda^*} = \left(\frac{1+\frac{T}{2}\lambda}{1-\frac{T}{2}\lambda}\right)^* = \rho^*$$

Symmetry about the imaginary axis is more complicated:

$$\frac{1+\frac{T}{2}(-\lambda^*)}{1-\frac{T}{2}(-\lambda^*)} = \left(\frac{1-\frac{T}{2}\lambda}{1+\frac{T}{2}\lambda}\right)^* = \frac{1}{\rho^*}. \tag{10.7}$$

We say complex numbers ρ_1 and ρ_2 are *symmetric about the unit circle* if

$$\rho_2 = \frac{1}{\rho_1^*}. \tag{10.8}$$

For real numbers, (10.8) states that ρ_1 and ρ_2 are reciprocal. If $\rho_1 = 0$, then (10.8) reduces to $\rho_2 = \infty$. Hence, the number of poles or zeros at $z = 0$ must equal the number at ∞.

Under the bilinear transformation, symmetry of CT zeros about the imaginary axis is equivalent to symmetry of DT zeros about the unit circle. The same holds for poles. This leads to the DT counterpart to Theorem 9.3:

Theorem 10.1 *A DT rational transfer function $H(z) \neq 0$ has zero phase iff $H\left(e^{j\Omega}\right)$ is real and positive for small Ω and the poles and zeros of $H(z)$ are symmetric relative to both the real axis and the unit circle.*

10.2 Recursive Structures for Causal IIR Filters

Having designed a DT filter $H(z)$ that we wish to implement on a computer, we need to decide between several methods for doing so. If $H(z)$ has at least one nonzero pole, then the impulse response $h[n]$ does not have finite-duration. Such systems are called *infinite impulse response (IIR)* filters. In order to implement an IIR filter, we must encode the difference equations corresponding to

$$H(z) = H_i(z) + H_o(z)$$

with the appropriate initial conditions. This can be done using a variety of computational schemes.

Several issues must be considered when deciding which scheme to use. For example, the method chosen can affect computational speed, memory requirements, and numerical stability. A detailed study of these issues is beyond the scope of this course. A "quick and dirty" approach is to design several different computational structures and decide through simulation which works best. The best choice turns out to be quite application-dependent.

Suppose we are given a difference equation

$$y[n+N] + a_{N-1}y[n+N-1] + \ldots + a_K y[n+K] = b_M x[n+M] + \ldots + b_0 x[n] \tag{10.9}$$

or rational function

$$H(z) = \frac{b_M z^{M-N} + \ldots + b_0 z^{-N}}{1 + a_{N-1}z^{-1} + \ldots + a_K z^{K-N}}$$

corresponding to a BIBO stable system. Recall that, if the system is neither causal nor anti-causal, then we may decompose $H(z)$ into inner and outer parts, resulting in the sum of a causal system and an anti-causal system. Hence, it suffices to consider the implementation of systems that are either causal or anti-causal. We begin with causal systems.

We may rewrite (10.9) for forward recursion as

$$y[n] = -a_{N-1}y[n-1] - \ldots - a_K y[n+K-N] + b_M x[n+M-N] + \ldots + b_0 x[n-N]. \tag{10.10}$$

For any input $x[n]$ with $x[n] = 0$ for $n < N_1$, $y[n]$ may be computed directly by recursively evaluating (10.10) using the initial conditions

$$y[N_1 - 1] = \ldots = y[N_1 - N] = 0.$$

This approach corresponds to the signal flow graph in Figure 10.3:

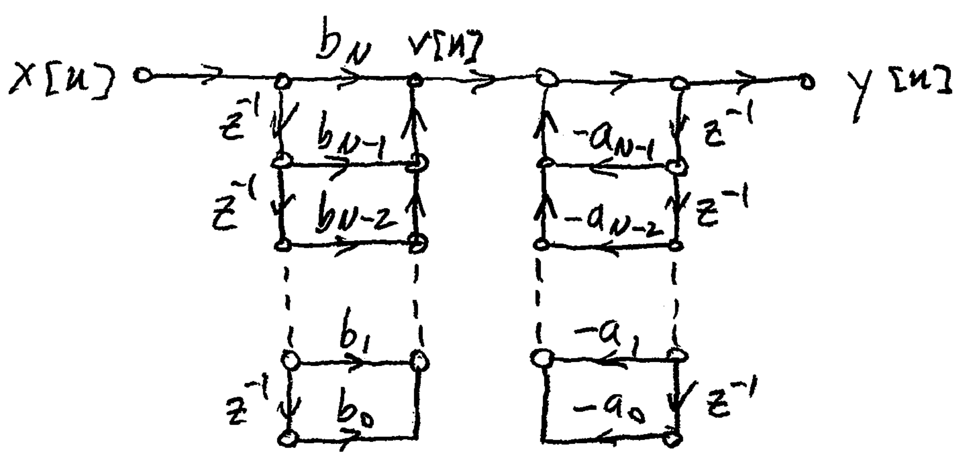


Figure 10.3

(Set $a_i = 0$ or $b_i = 0$ if it is not defined.) At each time step,

$$v[n] = b_M x[n + M - N] + \ldots + b_0 x[n - N]$$

is computed. Then $y[n]$ is generated by the recursion

$$y[n] = -a_{N-1} y[n-1] - \ldots - a_K y[n + K - N] + v[n].$$

This method is called *Direct Form I* and may be viewed as factoring

$$H(z) = \left(\frac{1}{1 + a_{N-1} z^{-1} + \ldots + a_K z^{K-N}}\right) \left(b_M z^{M-N} + \ldots + b_0 z^{-N}\right).$$

A related method is *Direct Form II* and is defined by the factorization

$$H(z) = \left(b_M z^{M-N} + \ldots + b_0 z^{-N}\right) \left(\frac{1}{1 + a_{N-1} z^{-1} + \ldots + a_K z^{K-N}}\right).$$

In other words, we first compute an intermediate signal $v[n]$ through the recursion

$$v[n] = -a_{N-1} v[n-1] - \ldots - a_K v[n + K - N] + x[n].$$

Then find

$$y[n] = b_M v[n + M - N] + \ldots + b_0 v[n - N].$$

The signal flow graph is

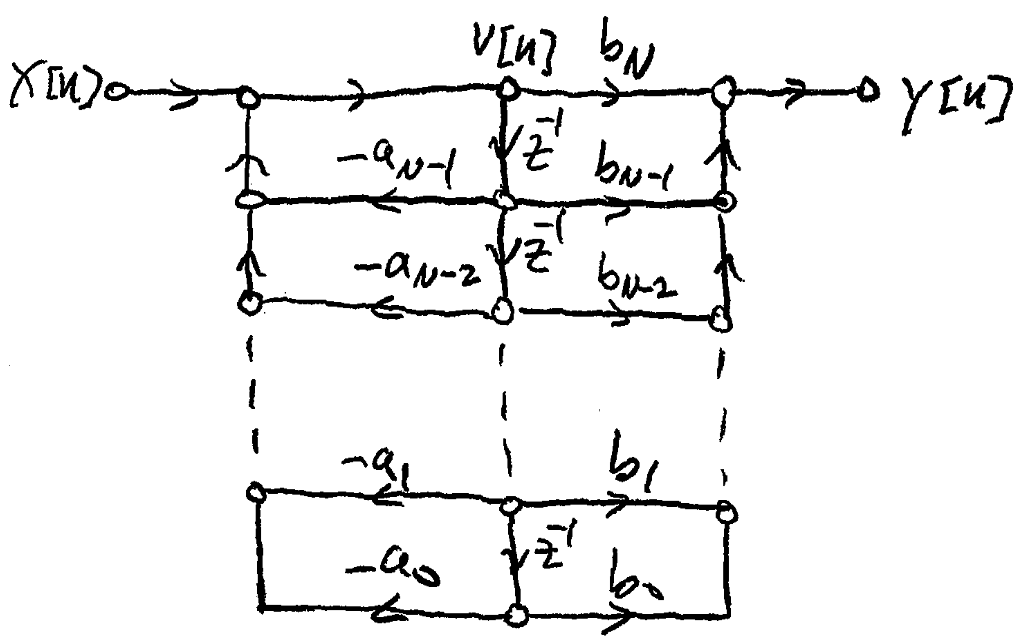


Figure 10.4

For any input and initial conditions, Direct Forms I and II yield the same output (ignoring roundoff error). One difference between the two structures is that Figure 10.3 uses $2N - K$ delays, while Figure 10.4 requires only N. Each delay corresponds to one unit of computer memory, so Direct Form II involves a savings in hardware.

A third structure is obtained through PFE (in z^{-1}):

$$H(z) = \widetilde{Q}\left(z^{-1}\right) + \sum_{i=1}^{r}\sum_{k=1}^{N_i} \frac{\widetilde{A}_{ik}}{\left(1-\rho_i z^{-1}\right)^k}.$$

The polynomial $\widetilde{Q}\left(z^{-1}\right)$ may be implemented with a series of delays. Recall that the coefficients in the strictly proper term may be complex. In order to avoid the difficulties of complex computation, we may combine conjugate terms into terms with real coefficients. For $k=1$,

$$\frac{\widetilde{A}_{i1}}{1-\rho_i z^{-1}} + \frac{\widetilde{A}_{i1}^*}{1-\rho_i^* z^{-1}} = 2\frac{\left(\operatorname{Re}\widetilde{A}_{i1}\right) z^{-1} - \operatorname{Re}\left(\widetilde{A}_{i1}^*\rho_i\right) z^{-2}}{1-\left(2\operatorname{Re}\rho_i\right) z^{-1} + \left|\rho_i\right|^2 z^{-2}}.$$

For arbitrary k,

$$\frac{\widetilde{A}_{ik}}{\left(1-\rho_i z^{-1}\right)^k} + \frac{\widetilde{A}_{ik}^*}{\left(1-\rho_i^* z^{-1}\right)^k} = \frac{2\sum_{l=0}^{k}(-1)^l\binom{k}{l}\operatorname{Re}\left(\widetilde{A}_{ik}^*\rho^l\right) z^{-l}}{\left(1-\left(2\operatorname{Re}\rho_i\right) z^{-1} + \left|\rho_i\right|^2 z^{-2}\right)^k}.$$

Each term in the resulting PFE corresponds to a smaller difference equation, which may be solved recursively according to either Direct Form I or II. Summing the outputs gives a realization of the original system. This approach is called *Parallel Form.*

Example 10.5 *Let*

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

The PFE is

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

Combining conjugate terms,

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

Using Direct Form II, the first term yields the equation

$$y_1[n] = \frac{1}{2}y_1[n-1] + x[n].$$

The second term corresponds to

$$v_2[n] = -v_2[n-1] - \frac{1}{2}v_2[n-2] + x[n],$$

$$y_2[n] = 2v_2[n] + v_2[n-1].$$

The overall system output is

$$y[n] = y_1[n] + y_2[n].$$

The signal flow graph is

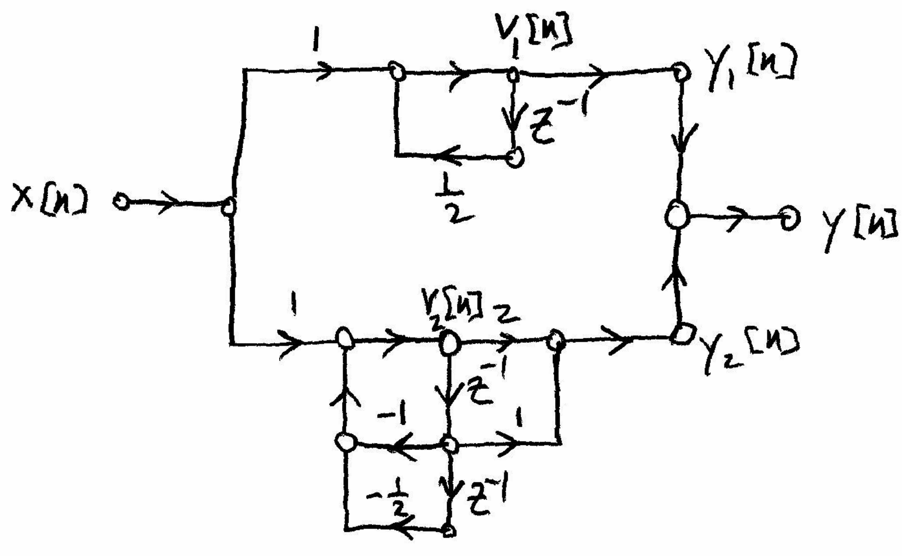


Figure 10.5

The final technique we will study requires factoring $H(z)$ into linear and quadratic factors with real coefficients

$$H(z) = H_p(z) \cdots H_1(z).$$

This can be done in a variety of ways. Each factor $H_i(z)$ determines a difference equation which can be implemented using Direct Forms I and II. The resulting systems are connected in series to generate the overall system. This is called *Cascade Form*.

Example 10.6 *The transfer function in Example 10.5 may be factored*

$$H(z) = \frac{3 + z^{-1}}{\left(1 - \frac{1}{2}z^{-1}\right)\left(1 + z^{-1} + \frac{1}{2}z^{-2}\right)} = H_2(z) H_1(z),$$

where

$$H_1(z) = \frac{3 + z^{-1}}{1 + z^{-1} + \frac{1}{2}z^{-2}},$$

$$H_2(z) = \frac{1}{1 - \frac{1}{2}z^{-1}}.$$

Of course, this is only one of several possibilities. $H_1(z)$ *and* $H_2(z)$ *determine the Direct Form II equations*

$$v_1[n] = -v_1[n-1] - \frac{1}{2}v_1[n-2] + x[n],$$

$$y_1[n] = 3v_1[n] + v_1[n-1],$$

$$y[n] = \frac{1}{2}y[n-1] + y_1[n].$$

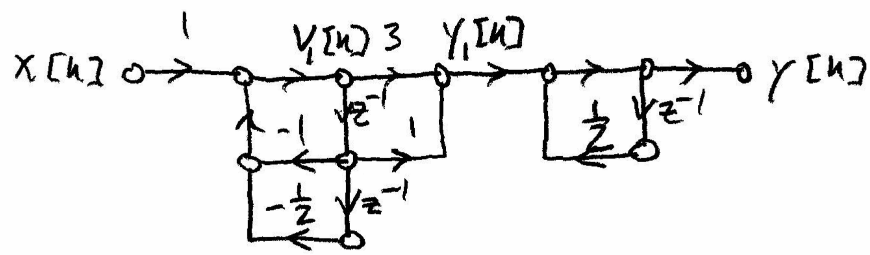


Figure 10.6

10.3 The Anti-Causal Case

The same computational structures may be used for anti-causal systems. The only change is that the recursion is backward, rather than forward. Suppose (10.9) corresponds to an anti-causal system. Backward recursion is based on the form

$$\begin{aligned}y[n] = &-\frac{1}{a_K}y[n+N-K]-\frac{a_{N-1}}{a_K}y[n+N-K-1]-\ldots-\frac{a_{K+1}}{a_K}y[n+1]\\&+\frac{b_M}{a_K}x[n+M-K]+\ldots+\frac{b_0}{a_K}x[n-K]\end{aligned}$$

and

$$H(z)=\frac{\frac{b_M}{a_K}z^{M-K}+\ldots+\frac{b_0}{a_K}z^{-K}}{\frac{1}{a_K}z^{N-K}+\frac{a_{N-1}}{a_K}z^{N-K-1}+\ldots+\frac{a_{K-1}}{a_K}z+1}.$$

For Direct Form I, we factor

$$H(z)=\left(\frac{1}{\frac{1}{a_K}z^{N-K}+\frac{a_{N-1}}{a_K}z^{N-K-1}+\ldots+\frac{a_{K-1}}{a_K}z+1}\right)\left(\frac{b_M}{a_K}z^{M-K}+\ldots+\frac{b_0}{a_K}z^{-K}\right),$$

which yields the equations

$$v[n]=\frac{b_M}{a_K}x[n+M-K]+\ldots+\frac{b_0}{a_K}x[n-K],$$

$$y[n]=-\frac{1}{a_K}y[n+N-K]-\frac{a_{N-1}}{a_K}y[n+N-K-1]-\ldots-\frac{a_{K+1}}{a_K}y[n+1]+v[n].$$

For Direct Form II,

$$H(z)=\left(\frac{b_M}{a_K}z^{M-K}+\ldots+\frac{b_0}{a_K}z^{-K}\right)\left(\frac{1}{\frac{1}{a_K}z^{N-K}+\frac{a_{N-1}}{a_K}z^{N-K-1}+\ldots+\frac{a_{K-1}}{a_K}z+1}\right),$$

$$v[n]=-\frac{1}{a_K}v[n+N-K]-\frac{a_{N-1}}{a_K}v[n+N-K-1]-\ldots-\frac{a_{K+1}}{a_K}v[n+1]+x[n],$$

$$y[n]=\frac{b_M}{a_K}v[n+M-K]+\ldots+\frac{b_0}{a_K}v[n-K].$$

For Parallel form, perform PFE (in z, not z^{-1}), combine conjugate terms, and use Direct Form I or II on each term. For Cascade Form, factor $H(z)$ and use Direct Form I or II on each factor, combining conjugate terms. In each case, the resulting signal flow graph involves forward shift elements z. Although mathematically a forward shift is a noncausal operation, on a computer each shift is just a memory element. The recursion proceeds forward in actual time, but represents a backward progression through conceptual time. Of course, this is possible only in non-real-time applications.

Example 10.7 *Let*

$$H(z)=\frac{3z^2-2}{z^3-2z-4}.$$

Find the Parallel Form. The poles are

$$\rho_1 = 2, \quad \rho_{2,3} = -1 \pm j.$$

Since the poles lie outside the unit circle, $H(z)$ corresponds to a BIBO stable, but anticausal, filter. The PFE is

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

For the first term,

$$y_1[n] = \frac{1}{2}y_1[n+1] - \frac{1}{2}x[n].$$

For the second, we use Direct Form II:

$$v_2[n] = -\frac{1}{2}v_2[n+2] - v_2[n+1] + x[n],$$

$$y_2[n] = v_2[n+1] + v_2[n].$$

Finally,

$$y[n] = y_1[n] + y_2[n].$$

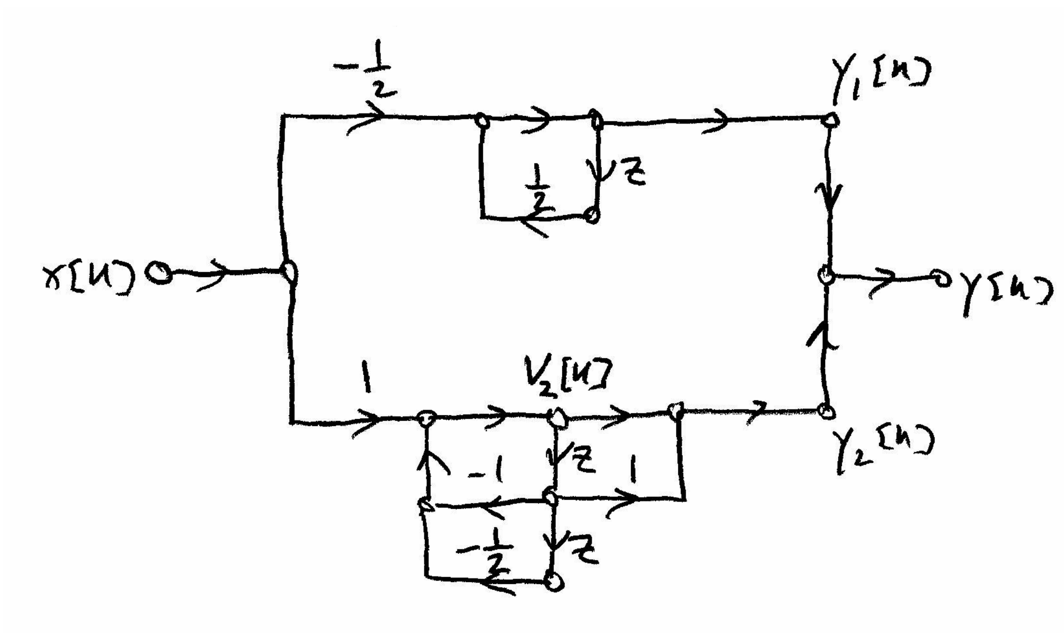


Figure 10.7

11 FIR Filters

11.1 Causal FIR Filters

If a DT filter $H_{DT}\left(e^{j\Omega}\right)$ has a finite-duration impulse response $h[n]$, it is called a *finite impulse response (FIR)* filter. FIR filters are always BIBO stable, since

$$\sum_{n=-\infty}^{\infty}|h[n]|=\sum_{n=-N_1}^{N_2}|h[n]|<\infty.$$

An FIR filter is often obtained by windowing the impulse response of an IIR filter.

First consider the causal case. We start with an IIR filter with impulse response $h[n]$. Then $h[n]$ is infinite-duration and $h[n]=0$ for $n<0$. For a window $w[n]$ with length N, we set

$$\widetilde{h}[n]=\alpha w[n]\,h[n],$$

where α is a scaling factor. Then $\widetilde{h}[n]$ is nonzero only for $0\leq n\leq N-1$ and thus determines a causal FIR filter. If the IIR design is based on a CT LPF $H_{CT}(j\omega)$, the bilinear transformation (10.3) indicates

$$H_{DT}\left(e^{j\Omega}\right)=H_{CT}\left(j\frac{2}{T}\tan\frac{\Omega}{2}\right).$$

In this context, windowing has some unfortunate consequences. One of these is that it changes the gain of the filter at critical frequencies. For example, in designing a low-pass FIR filter, we wish to achieve a "DC gain" which matches the original IIR filter. Setting $\Omega=0$, we want

$$H_{DT}(1)=H_{CT}(0).$$

(Typically, $H_{CT}(0)=1$.) Windowing generally results in the filter with impulse response $w[n]\,h[n]$ having DC gain other than $H_{CT}(0)$. Hence, we need to choose α to achieve

$$\widetilde{H}(1)=H_{CT}(0).$$

The DTFT of the windowed filter is

$$\widetilde{H}\left(e^{j\Omega}\right)=\sum_{n=0}^{N-1}\widetilde{h}[n]\,e^{-j\Omega n},$$

so

$$\widetilde{H}(1)=\sum_{n=0}^{N-1}\widetilde{h}[n]=\alpha\sum_{n=0}^{N-1}w[n]\,h[n],$$

$$\alpha=\frac{H_{CT}(0)}{\sum\limits_{n=0}^{N-1}w[n]\,h[n]}.$$

Example 11.1 *Starting with the* 2nd *order Butterworth, causal, CT LPF with bandwidth* 1, *convert to DT using sampling rate* $f_s=\frac{1}{2}$ *and window to FIR using a Hann window with* $N=7$.

$$H_{CT}(s)=\frac{1}{s^2+\sqrt{2}s+1}$$

$$\begin{aligned}
H_{DT}(z) &= \frac{1}{\left(\frac{z-1}{z+1}\right)^2 + \sqrt{2}\left(\frac{z-1}{z+1}\right) + 1} \\
&= \frac{1}{2+\sqrt{2}} \frac{z^2+2z+1}{z^2+3-2\sqrt{2}} \\
&= \frac{1}{2+\sqrt{2}} \frac{1+2z^{-1}+z^{-2}}{1+\left(3-2\sqrt{2}\right)z^{-2}} \\
&= 1 + \frac{1}{\sqrt{2}} - \sqrt{2}\frac{1-\sqrt{3-2\sqrt{2}}z^{-1}}{1+\left(3-2\sqrt{2}\right)z^{-2}}
\end{aligned}$$

From O&S Table 3.1, lines 11 and 12,

$$h[n] = \left(1+\frac{1}{\sqrt{2}}\right)\delta[n] - \sqrt{2}\left(\cos\left(\frac{\pi}{2}n\right) - \sin\left(\frac{\pi}{2}n\right)\right)\left(3-2\sqrt{2}\right)^{\frac{n}{2}} u[n].$$

$$w[n] = \begin{cases} \frac{1}{2} - \frac{1}{2}\cos\left(\frac{\pi}{3}n\right), & 0 \le n \le 6 \\ 0, & \textit{else} \end{cases}$$

$$\widetilde{h}[n] = \alpha w[n] h[n]$$

$$\begin{aligned}
\alpha &= \frac{1}{\sum\limits_{n=0}^{6} w[n] h[n]} \\
&= -\frac{\sqrt{2}}{\sum\limits_{n=1}^{6} \left(\cos\left(\frac{\pi}{2}n\right) - \sin\left(\frac{\pi}{2}n\right)\right)\left(3-2\sqrt{2}\right)^{\frac{n}{2}}\left(1-\cos\left(\frac{\pi}{3}n\right)\right)} \\
&= 4.975
\end{aligned}$$

The constant α may be designed to correct the gain at other critical frequencies. For example, in a band-pass design, we may wish to set α so that the FIR and IIR gains match at $\omega_0 = \sqrt{\omega_1\omega_2}$:

$$\widetilde{H}\left(e^{j\omega_0 T}\right) = H_{CT}(j\omega_0).$$

Then

$$\alpha = \frac{H_{CT}(j\omega_0)}{\sum\limits_{n=0}^{N-1} w[n] h[n] e^{-j\omega_0 T n}}.$$

11.2 Zero-Phase FIR Filters

Beginning with a zero phase (noncausal) CT filter $H_{CT}(j\omega)$, we may convert it to a DT $H_{DT}\left(e^{j\Omega}\right)$ and take the IDTFT, obtaining an even impulse response $h[n]$. To maintain zero phase after windowing, the window must also be even. The standard windows $w[n]$ are nonzero only for $0 \le n \le N-1$, but may be shifted to the even function $w\left[n+\frac{N-1}{2}\right]$. ($N$ must be odd.). In this case, set

$$\widetilde{h}[n] = \alpha w\left[n+\frac{N-1}{2}\right] h[n],$$

with

$$\alpha = \frac{H_{CT}(0)}{\sum\limits_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w\left[n+\frac{N-1}{2}\right] h[n]}.$$

As we noted in Example 9.4, an even impulse response does not always give rise to a zero phase frequency response. This leads to another unfortunate consequence of windowing: **A windowed zero phase filter may not have zero phase**.

Example 11.2 *Let*

$$h[n] = \left(\frac{2}{3}\right)^{|n|}.$$

The poles and zeros of

$$H(z) = -\frac{5}{6}\frac{z}{\left(z-\frac{2}{3}\right)\left(z-\frac{3}{2}\right)}$$

guarantee that the IIR filter has zero phase. Let $w[n]$ be the rectangular window with length $N=3$. Then

$$\widetilde{h}[n] = \alpha w[n+1]\, h[n] = \begin{cases} \alpha\left(\frac{2}{3}\right)^{|n|}, & n=-1,0,1 \\ 0, & else \end{cases}$$

is even, but

$$\widetilde{H}\left(e^{j\Omega}\right) = \alpha \sum_{n=-1}^{1}\left(\frac{2}{3}\right)^{|n|} e^{-j\Omega} = \alpha\left(\frac{2}{3}e^{j\Omega}+1+\frac{2}{3}e^{-j\Omega}\right) = \alpha\left(1+\frac{4}{3}\cos\Omega\right).$$

Since $\widetilde{H}\left(e^{j\Omega}\right)$ oscillates between $+$ and $-$, the FIR filter does not have zero phase.

In most cases windowing zero phase IIR to FIR does maintain zero phase. But, as demonstrated in Example 11.2, the zero phase property must be confirmed explicitly.

Example 11.3 *Starting with a 2nd order Butterworth, zero phase, CT LPF with bandwidth 1, convert to DT using sampling rate $f_s = 1$ and window to FIR using a Hann window with $N = 11$. Confirm that the FIR filter has zero phase.*

$$H_{CT}(s) = \frac{1}{-s^2+1}$$

$$H_{DT}(z) = \frac{1}{-\left(2\frac{z-1}{z+1}\right)^2+1} = \frac{-\frac{1}{3}z^2-\frac{2}{3}z-\frac{1}{3}}{z^2-\frac{10}{3}z+1}$$

$$H_i(z) = \frac{\frac{2}{9}}{z-\frac{1}{3}} = \frac{\frac{2}{9}z^{-1}}{1-\frac{1}{3}z^{-1}}$$

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

$$h[n] = \frac{2}{9}3^{-n+1}u[n-1]-\frac{1}{3}\delta[n]+2\cdot 3^{n-1}u[-n] = \frac{1}{3}\left(2\cdot 3^{-|n|}-\delta[n]\right)$$

$$\widetilde{h}[n] = \alpha w[n+5]\, h[n]$$

$$w[n+5] = \frac{1}{2}\left(1 - \cos\left(\frac{\pi}{5}(n+5)\right)\right) = \frac{1}{2}\left(1 + \cos\left(\frac{\pi}{5}n\right)\right)$$

$$\begin{aligned}
\alpha &= \frac{1}{\sum\limits_{n=-5}^{5} w[n+5]\, h[n]} \\
&= \frac{6}{\sum\limits_{n=-5}^{5} \left(2 \cdot 3^{-|n|} - \delta[n]\right)\left(1 + \cos\left(\frac{\pi}{5}n\right)\right)} \\
&= \frac{3}{1 + 2\sum\limits_{n=1}^{5} 3^{-n}\left(1 + \cos\left(\frac{\pi}{5}n\right)\right)} \\
&= 1.175
\end{aligned}$$

$$\widetilde{h}[n] = \begin{cases} 0.1959\left(2 \cdot 3^{-|n|} - \delta[n]\right)\left(1 + \cos\left(\frac{\pi}{5}n\right)\right), & -5 \le n \le 5 \\ 0, & else \end{cases} = \begin{cases} 0.0009238, & n = \pm 4 \\ 0.01003, & n = \pm 3 \\ 0.05699, & n = \pm 2 \\ 0.2362, & n = \pm 1 \\ 0.3918, & n = 0 \\ 0, & else \end{cases}$$

$$\begin{aligned}
\widetilde{H}(z) = {} & 0.0009238z^{-4} + 0.01003z^{-3} + 0.05699z^{-2} + 0.2362z^{-1} + 0.3918 \\
& + 0.2362z + 0.05699z^{2} + 0.01003z^{3} + 0.0009238z^{4}
\end{aligned}$$

$$\begin{aligned}
\widetilde{H}\left(e^{j\Omega}\right) &= 0.0009238e^{-j4\Omega} + 0.01003e^{-j3\Omega} + 0.05699e^{-j2\Omega} + 0.2362e^{-j\Omega} + 0.3918 \\
&\quad + 0.2362e^{j\Omega} + 0.05699e^{j2\Omega} + 0.01003e^{j3\Omega} + 0.0009238e^{j4\Omega} \\
&= 0.001848\cos(4\Omega) + 0.02006\cos(3\Omega) + 0.1140\cos(2\Omega) + 0.4724\cos\Omega + 0.3918 \\
&\ge 0.01519
\end{aligned}$$

Since $\widetilde{H}\left(e^{j\Omega}\right) > 0$ *for all* Ω, $\angle\widetilde{H}\left(e^{j\Omega}\right) = 0$.

11.3 Choice of Window Length

For any BIBO stable LTI system, the impulse response satisfies $|h[n]| \to 0$ as $|n| \to \infty$. Hence, using a sufficiently long window ensures that

$$\widetilde{h}[n] = \alpha w[n]\, h[n]$$

closely approximates $h[n]$. But how long is long enough? If the window is too long, we are essentially windowing 0's for large n, and the filter implementation will require unnecessary computations and memory. If the window is too short, the approximation to the IIR response will be poor.

A simple answer to this question comes from a study of the exponential function, which should be familiar from elementary circuit analysis. The following table is easy to compute:

m	e^{-m}
0	1
1	.37
2	.14
3	.050
4	.018
5	.0067
6	.0025
7	.00091
8	.00034
9	.00012

In CT problems, one often encounters the decaying exponential

$$x(t) = e^{-\lambda t}.$$

The magnitude is

$$|x(t)| = e^{-(\operatorname{Re}\lambda)t}.$$

Setting

$$\tau = \frac{1}{\operatorname{Re}\lambda},$$

we may write

$$|x(t)| = e^{-\frac{t}{\tau}}.$$

The number τ is the *time constant*. The left column of the table represents an integer number of time constants, while the right column indicates the extent to which $x(t)$ has decayed after that length of time. For example, 1 time constant translates into $x(t)$ falling to 37% of its initial value. After 5 time constants, $x(t)$ has fallen below 1% of its initial value $x(0)$. For this reason, 5 time constants is a common measure of how long an exponentially decaying signal takes to become negligible.

For DT, exponentials are usually written in the form

$$x[n] = \rho^n.$$

Assuming $|\rho| < 1$, we may draw a comparison to the CT case by writing

$$|x[n]| = |\rho|^n = e^{-\frac{n}{\tau}},$$

where

$$\tau = \frac{1}{\ln \frac{1}{|\rho|}}$$

is a (dimensionless) "time constant". Again, for $n \geq 5\tau$, $|x[n]|$ is less than 1% of $|x[0]|$. For $n < 5\tau$, $x[n]$ is large enough that it should not be ignored.

Now consider an IIR filter with poles $\rho_1, \ldots, \rho_k$ inside the unit circle and poles $\rho_{k+1}, \ldots, \rho_N$ outside the unit circle. For $n > 0$, the impulse response $h_i[n]$ of the inner poles is a sum of right-sided exponentials ρ_l^n, and the worst-case "inner time constant" is

$$\tau_i = \max_{l \leq k} \frac{1}{\ln \frac{1}{|\rho_l|}}.$$

The impulse response $h_o[n]$ of the outer poles is a sum of left-sided exponentials ρ_l^n. The decay occurs as $n \to -\infty$. This is equivalent to letting $n \to +\infty$ in

$$\rho_l^{-n} = \left(\frac{1}{\rho_l}\right)^n.$$

Hence, we define the "outer time constant"

$$\tau_o = \max_{l>k} \frac{1}{\ln|\rho_l|}.$$

The inner time constant relates to decay as $n \to \infty$ and the outer to $n \to -\infty$. To avoid wasting computer resources, windows should be nonzero only on the interval $-5\tau_o \le n \le 5\tau_i$. We refer to this as the 5 *time constant rule.*

For causal filters, only inner poles appear, so we can calculate N according to

$$N - 1 = \lceil 5\tau_i \rceil$$

or

$$N = \lceil 5\tau_i \rceil + 1,$$

where

$$\lceil t \rceil = \text{smallest integer} \ge t$$

is the *ceiling* function. In the case of zero phase filters, $h[n]$ is even so, again, we need only consider the inner poles. Since $h[n] = 0$ except for

$$-\frac{N-1}{2} \le n \le \frac{N-1}{2},$$

we may set

$$\frac{N-1}{2} = \lceil 5\tau_i \rceil$$

or

$$N = 2\lceil 5\tau_i \rceil + 1.$$

Example 11.4 *In Example 11.1, the poles are*

$$\rho_{1,2} = \pm j\sqrt{3 - 2\sqrt{2}},$$

so

$$\tau_i = \frac{1}{\ln \frac{1}{\sqrt{3-2\sqrt{2}}}} = 1.135.$$

The appropriate window length is

$$N = \lceil 5\tau_i \rceil + 1 = \lceil 5.675 \rceil + 1 = 7.$$

In Example 11.3,

$$\rho_1 = \frac{1}{3}, \quad \rho_2 = 3,$$

$$\tau_i = \frac{1}{\ln \frac{1}{1/3}} = .9102,$$

$$N = 2\lceil 5\tau_i \rceil + 1 = 2\lceil 4.551 \rceil + 1 = 11.$$

11.4 Linear Phase FIR Filters

A DT filter with transfer function $H\left(e^{j\Omega}\right)$ is a *linear phase* filter if there exists an integer n_0 such that

$$\angle H\left(e^{j\Omega}\right) = -n_0\Omega, \quad -\pi < \Omega < \pi.$$

Thus

$$H\left(e^{j\Omega}\right) = \left|H\left(e^{j\Omega}\right)\right| e^{-jn_0\Omega}.$$

In particular, a zero phase filter has linear phase with $n_0 = 0$. Linear phase filters exhibit no phase distortion, since the phase delay

$$\Delta\left(\Omega\right) = -\frac{\angle H\left(e^{j\Omega}\right)}{\Omega} = -n_0$$

is constant.

Let $h[n]$ be the impulse response of a linear-phase filter. Then

$$h[n+n_0] \longleftrightarrow e^{jn_0\Omega}H\left(e^{j\Omega}\right) = \left|H\left(e^{j\Omega}\right)\right|,$$

so a filter has linear phase iff its impulse response is the time-shift of that of a zero phase filter. In particular, $h[n+n_0]$ must be an even function. If, in addition, the linear phase filter is causal, then $h[n] = 0$ for $n < 0$, so $h[n] = 0$ for $n > 2n_0 + 1$. This proves the following result:

Theorem 11.1 *If a DT linear phase filter is causal, then it is FIR.*

The main application of linear-phase is in real-time problems where phase distortion is a significant issue. As we have seen, real-time IIR filters have nonlinear phase. Unfortunately, zero phase IIR filters are noncausal and thus cannot be implemented in real-time. The alternative is to window the IIR filter to convert it to zero phase FIR. But such a filter has an even impulse response, making the filter noncausal. Set

$$n_0 = \max\left\{n \,\middle|\, h[n] \neq 0\right\}.$$

Then

$$\min\left\{n \,\middle|\, h[n] \neq 0\right\} = -n_0,$$

so we may simply delay $h[n]$ by n_0 time steps to force causality. This achieves a real-time linear phase filter without changing $\left|H\left(e^{j\Omega}\right)\right|$. Although the resulting filter does not suffer from phase distortion, it does exhibit a delay of $n_0 T$. Clearly, n_0 must be kept small enough to make the delay negligible. This constitutes a second upper bound on window length, which may be significantly smaller than that determined by the system time constants. Such a design involves a trade-off between small delay and good approximation to the IIR response.

Example 11.5 *Starting with a 2nd order Butterworth, zero phase, CT LPF with bandwidth* 1, *convert to DT using sampling rate* $f_s = 1$ *and window to FIR using a Hann window with* $N = 11$. *Shift the impulse response to generate a causal, linear phase LPF.*
From Example 11.3,

$$\widetilde{h}[n] = \begin{cases} 0.1959\left(2\cdot 3^{-|n|} - \delta[n]\right)\left(1+\cos\left(\frac{\pi}{5}n\right)\right), & -5 \leq n \leq 5 \\ 0, & else \end{cases}$$

meets all the conditions, except causality. Delaying by $n_0 = 5$ *yields a linear-phase, causal filter with impulse response* $\widetilde{h}[n-5]$. *The resulting CT delay is* $n_0 T = 5$.

11.5 Difference Equation Implementation

Suppose an FIR filter has impulse response $h[n]$, where $h[n] = 0$ for $n < N_1$ and $n > N_2$. For any input $x[n]$,

$$\begin{aligned} y[n] &= h[n] * x[n] \qquad (11.1) \\ &= \sum_{m=N_1}^{N_2} h[m]\, x[n-m] \\ &= h[N_1]\, x[n-N_1] + \ldots + h[N_2]\, x[n-N_2], \end{aligned}$$

which is just the zeroth-order difference equation (8.12) with

$$M = N_2 - N_1,$$

$$N = N_2,$$

$$b_i = h[N_2 - i].$$

As with IIR filters, one may program the filter in a variety of ways for the sake of computational efficiency and stability. We will examine two methods. The first is *Direct Form*:

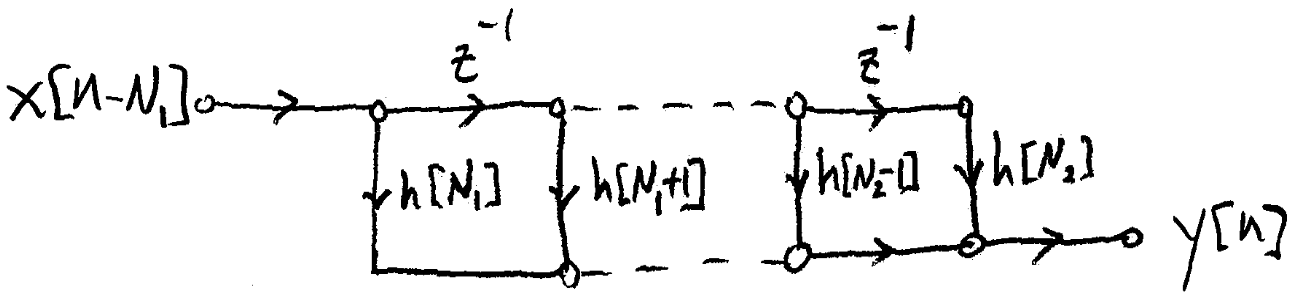


Figure 11.1

Here the difference equation (11.1) is computed directly, requiring $N_2 - N_1 + 1$ multiplications and N_2 delays.

In the case of causal, linear phase filters, a more efficient variant of Direct Form is possible. Suppose M is even and

$$h[n] = \begin{cases} h[M-n], & 0 \le n \le M \\ 0, & n < 0 \text{ or } n > M \end{cases}.$$

Then $N_1 = 0$ and $N_2 = M$. The filter is FIR with phase

$$\angle H\left(e^{j\Omega}\right) = -\frac{M}{2}\Omega.$$

The following structure requires only $\frac{M}{2} + 1$ multiplications:

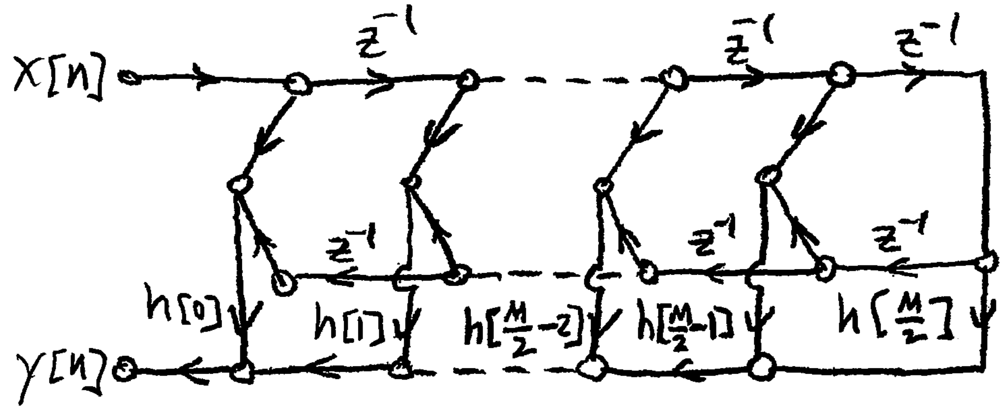


Figure 11.2

The second method requires factorization of the polynomial

$$Q\left(z^{-1}\right) = h\left[N_1\right] + h\left[N_1+1\right]z^{-1} + h\left[N_1+2\right]z^{-2} + \ldots + h\left[N_2\right]\left(z^{-1}\right)^{N_2-N_1}$$

according to the Fundamental Theorem of Algebra. The complex conjugate factors must be combined to yield a product of linear and quadratic factors, all with real coefficients. Each of these factors may be programmed in Direct Form and connected in series. Such a structure is called *Cascade Form.* Note the similarity between Cascade Form for FIR filters and IIR filters.

Example 11.6 *Consider the FIR filter with impulse response*

$$h[n] = \begin{cases} 1, & n=-2 \\ 3, & n=-1,0,1 \\ 2, & n=2 \\ 0, & else \end{cases}.$$

Then $N_1 = -2$, $N_2 = 2$, *and*

$$\begin{aligned} Q\left(z^{-1}\right) &= h[-2] + h[-1]z^{-1} + h[0]z^{-2} + h[1]z^{-3} + h[2]z^{-4} \\ &= 1 + 3z^{-1} + 3z^{-2} + 3z^{-3} + 2z^{-4} \\ &= \left(1+z^{-1}\right)\left(1+2z^{-1}\right)\left(1+jz^{-1}\right)\left(1-jz^{-1}\right) \\ &= \left(1+z^{-1}\right)\left(1+2z^{-1}\right)\left(1+z^{-2}\right). \end{aligned}$$

The Cascade Form is

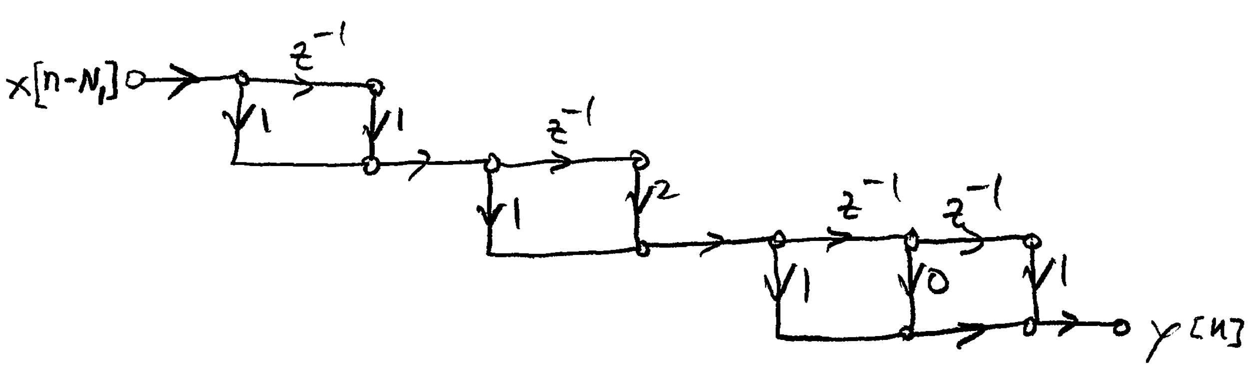


Figure 11.3

11.6 DFT Implementation

Given the impulse response $h[n]$ of an FIR filter and an input signal $x[n]$, the convolution (11.1) may also be computed using the DFT as outlined in Section 6. This constitutes another method of FIR filter implementation. Recall Algorithm 6.1:

Algorithm 11.1 *1) Zero-pad $h[n]$ and $x[n]$ to $M+N-1$ points.*
2) Apply FFT to $x[n]$ and $h[n]$.
3) Compute $H[k]X[k]$.
4) Apply FFT to $H[k]X[k]$.
5) Reflect the result (mod $M+N-1$) and divide by $M+N-1$.

Since the entire input signal $x[n]$ must be known in order to compute $X[k]$, this method applies only to non-real-time applications.