Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632

ALAN V. OPPENHEIM
ALAN S. WILLSKY

with
IAN T. YOUNG

SIGNALS and SYSTEMS

Library of Congress Cataloging in Publication Data

OPPENHEIM, ALAN V. (date)
Signals and systems.

(Prentice-Hall signal processing series)
Includes index.
1. System analysis. 2. Signal theory
(Telecommunication) I. Willsky, Alan S.
II. Young, Ian T. III. Title. IV. Series.
QA402.063 003 81-22652
ISBN 0-13-809731-3 AACR2

Editorial production/supervision
by Gretchen K. Chenenko
Chapter opening design
by Dawn Stanley
Cover executed by Judy Matz
Manufacturing buyers: Joyce Levatino
and Anthony Caruso

Printed in the United States of America

10 9

ISBN 0-13-809731-3

Prentice-Hall International, Inc., *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Prentice-Hall of Southeast Asia Pte. Ltd., *Singapore*
Whitehall Books Limited, *Wellington, New Zealand*

7
Modulation *447*

8
Sampling *513*

9
The Laplace Transform *573*

8.0 INTRODUCTION

Under certain conditions a continuous-time signal can be completely represented by and recoverable from knowledge of its instantaneous values or *samples* equally spaced in time. This somewhat surprising property follows from a basic result which is referred to as the *sampling theorem.* This theorem is extremely important and useful. It is exploited, for example, in moving pictures, which consist of a time sequence of individual frames, each of which represents an instantaneous view (i.e., a time sample) of a continuously changing scene. When these samples are viewed in time sequence at a sufficiently fast rate, we perceive an accurate representation of the original continuously moving scene. As another example, printed pictures typically consist of a very fine grid of points, each corresponding to a sample of the spatially continuous picture to be represented. If the samples are sufficiently close together the picture appears to be spatially continuous, although under a magnifying glass its representation in terms of samples becomes evident.

Much of the importance of the sampling theorem also lies in its role as a bridge between continuous-time signals and discrete-time signals. As we develop in some detail, the ability under certain conditions to completely represent a continuous-time signal by a sequence of instantaneous samples provides a mechanism for representing a continuous-time signal by a discrete-time signal. In many contexts, processing of discrete-time signals is more flexible and is often preferable to processing of continuous-time signals, in part because of the increasing availability of inexpensive, lightweight, programmable and easily reproducible digital and discrete-time systems. This technology also offers the possibility of exploiting the concept of sampling to convert

Sampling

8

a continuous-time signal to a discrete-time signal. After processing the discrete-time signal using a discrete-time system, we can then convert back to continuous time.

In the following discussion, we first introduce and develop the concept of sampling and the process of reconstructing a continuous-time signal from its samples. We then explore the processing of continuous-time signals that have been converted to discrete-time signals through sampling. Next we consider the dual concept to time-domain sampling, specifically sampling in the frequency domain. Finally, we develop the concept and some applications of sampling applied to discrete-time signals.

8.1 REPRESENTATION OF A CONTINUOUS-TIME SIGNAL BY ITS SAMPLES: THE SAMPLING THEOREM

In general, we could not expect that in the absence of any additional conditions or information, a signal could be uniquely specified by a sequence of equally spaced samples. For example, in Figure 8.1 we illustrate three different continuous-time signals, all of which have identical values at integer multiples of T, that is,

$$x_1(kT) = x_2(kT) = x_3(kT)$$

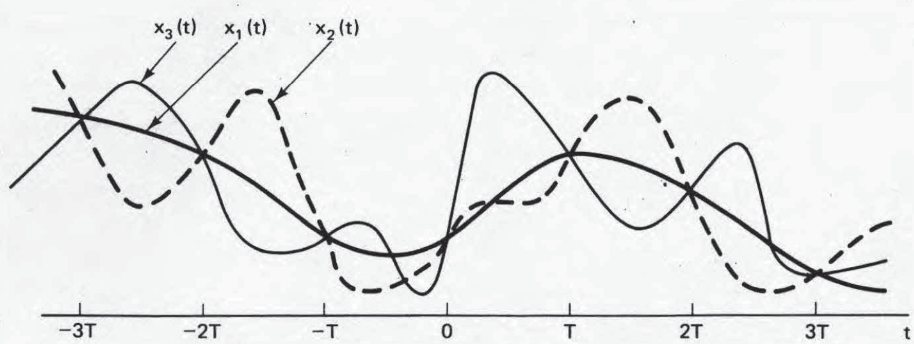


Figure 8.1 Three continuous-time signals with identical values at integer multiples of T.

In general, there are an infinite number of signals that can generate a given set of samples. As we will see, however, if a signal is bandlimited and if the samples are taken sufficiently close together, in relation to the highest frequency present in the signal, then the samples *uniquely* specify the signal and we can reconstruct it perfectly. The basic result was suggested in Section 7.4 in the context of pulse amplitude modulation. Specifically, if a bandlimited signal $x(t)$ is amplitude-modulated with a periodic pulse train, corresponding to extracting equally spaced time segments, it can be recovered exactly by lowpass filtering if the fundamental frequency of the modulating pulse train is greater than twice the highest frequency present in $x(t)$. Furthermore, the ability to recover $x(t)$ is independent of the time duration of the individual pulses. Thus, as suggested by Figures 8.2 and 8.3 as this duration becomes arbitrarily small, pulse amplitude modulation is, in effect, representing $x(t)$ by instantaneous samples equally spaced in time. In the pulse-amplitude-modulation system in Figure 8.2, we

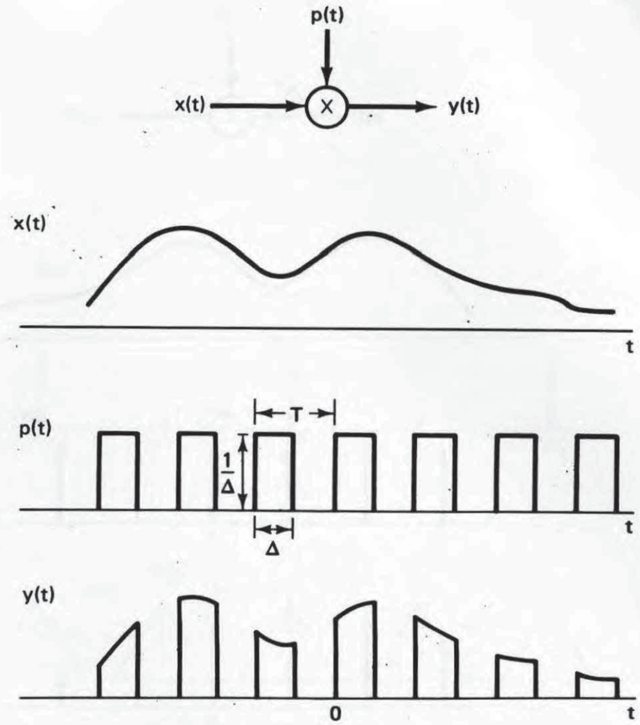


Figure 8.2 Pulse amplitude modulation. As $\Delta \longrightarrow 0$, $p(t)$ approaches an impulse train.

have scaled the amplitude of the pulse train to be inversely proportional to the pulse width Δ. In any practical pulse-amplitude-modulation system, it is particularly important as Δ becomes small to maintain a constant time-average power in the modulated signal. As illustrated in Figure 8.3, as Δ approaches zero the modulated signal then becomes an impulse train for which the individual impulses have values corresponding to instantaneous samples of $x(t)$ at time instants spaced T seconds apart.

8.1.1 Impulse-Train Sampling

In a manner identical to that used to analyze the more general case of pulse amplitude modulation, let us consider the specific case of impulse-train sampling depicted in Figure 8.3. The impulse train $p(t)$ is referred to as the *sampling function*, the period T as the *sampling period*, and the fundamental frequency of $p(t)$, $\omega_s = 2\pi/T$, as the *sampling frequency*. In the time domain we have

$$x_p(t) = x(t)p(t) \tag{8.1a}$$

where

$$p(t) = \sum_{n=-\infty}^{+\infty} \delta(t - nT) \tag{8.1b}$$

$x_p(t)$ is an impulse train with the amplitudes of the impulses equal to the samples of $x(t)$ at intervals spaced by T, that is,

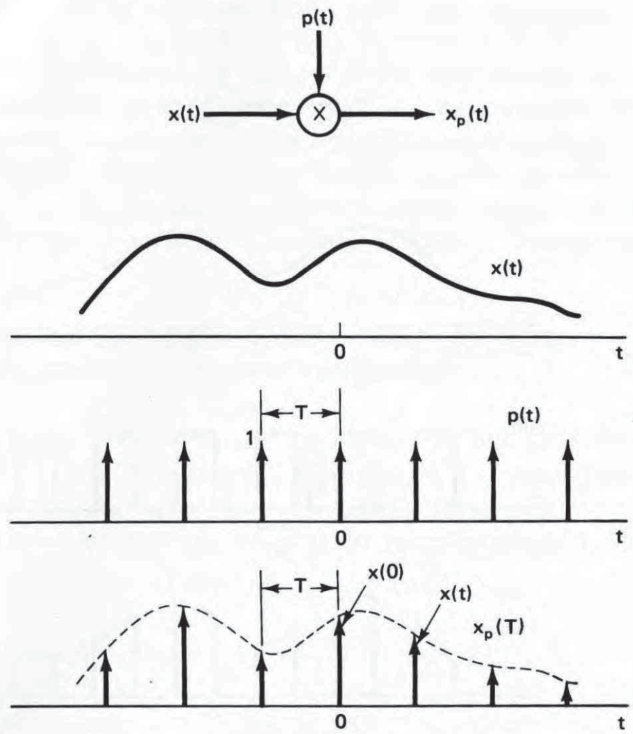


Figure 8.3 Pulse amplitude modulation with an impulse train.

$$x_p(t) = \sum_{n=-\infty}^{+\infty} x(nT)\,\delta(t - nT) \tag{8.2}$$

From the modulation property [Sec. 4.8],

$$X_p(\omega) = \frac{1}{2\pi}[X(\omega) * P(\omega)] \tag{8.3}$$

and from Example 4.15,

$$P(\omega) = \frac{2\pi}{T} \sum_{k=-\infty}^{+\infty} \delta(\omega - k\omega_s) \tag{8.4}$$

so that

$$X_p(\omega) = \frac{1}{T} \sum_{k=-\infty}^{+\infty} X(\omega - k\omega_s) \tag{8.5}$$

That is, $X_p(\omega)$ is a periodic function of frequency consisting of a sum of shifted replicas of $X(\omega)$, scaled by $1/T$ as illustrated in Figure 8.4. In Figure 8.4(c), $\omega_M < (\omega_s - \omega_M)$ or equivalently $\omega_s > 2\omega_M$, and thus there is no overlap between the shifted replicas of $X(\omega)$, whereas in Figure 8.4(d) with $\omega_s < 2\omega_M$, there is overlap. For the case illustrated in Figure 8.4(c), $X(\omega)$ is faithfully reproduced at integer multiples of the sampling frequency. Consequently, if $\omega_s > 2\omega_M$, $x(t)$ can be recovered exactly from $x_p(t)$ by means of a lowpass filter with gain T and a cutoff

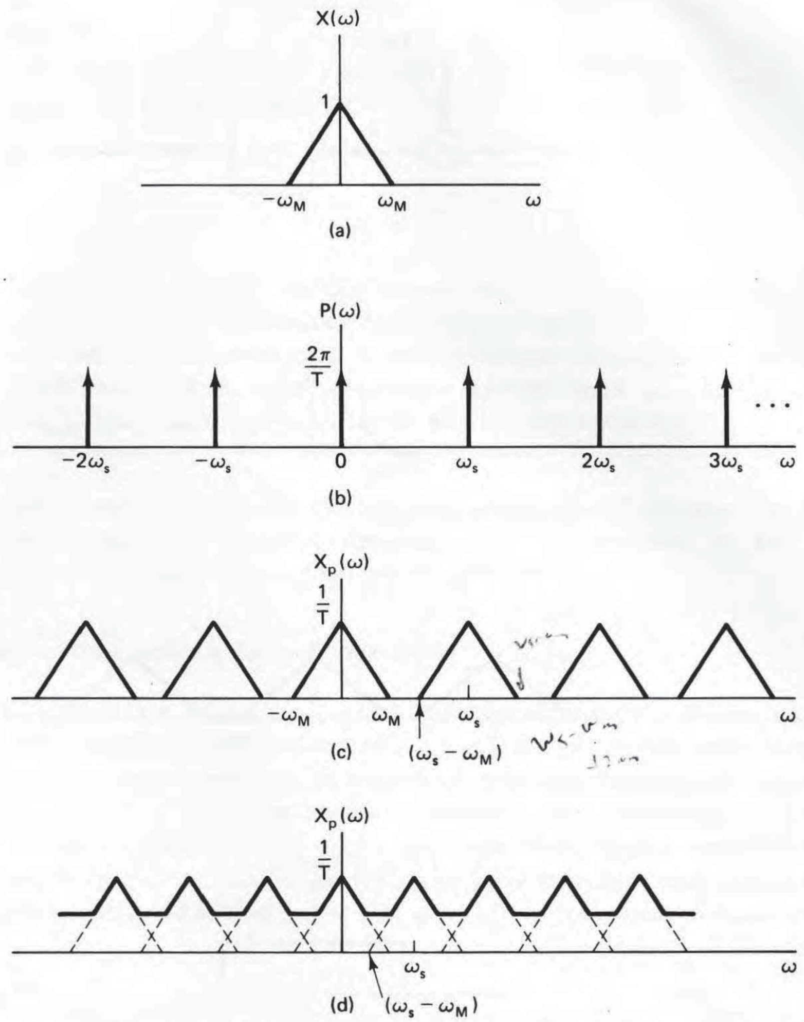


Figure 8.4 Effect in the frequency domain of sampling in the time domain: (a) spectrum of original signal; (b) spectrum of sampling function; (c) spectrum of sampled signal with $\omega_s > 2\omega_M$; (d) spectrum of sampled signal with $\omega_s < 2\omega_M$.

frequency greater than ω_M and less than $\omega_s - \omega_M$, as indicated in Figure 8.5. This basic result, referred to as the *sampling theorem*, can be stated as follows:†

†This important and elegant theorem was available for many years in a variety of forms in the mathematics literature. See, for example, J. M. Whittaker, "Interpolatory Function Theory," *Cambridge Tracts in Mathematics and Mathematical Physics*, no. 33 (Cambridge, 1935), chap. 4. It did not appear explicitly in the literature of communication theory until the publication in 1949 of the classic paper by Shannon entitled "Communication in the Presence of Noise" (*Proceedings of the IRE*, January, 1949, pp. 10–21). However, H. Nyquist in 1928 and D. Gabor in 1946 had pointed out, based on the use of Fourier Series, that 2TW numbers are sufficient to represent a function of time duration T and highest frequency W. [H. Nyquist, "Certain Topics in Telegraph Transmission Theory," *AIEE Transactions*, 1946, p. 617; D. Gabor, "Theory of Communication," *Journal of IEE* 93, no. 26 (1946): 429.]

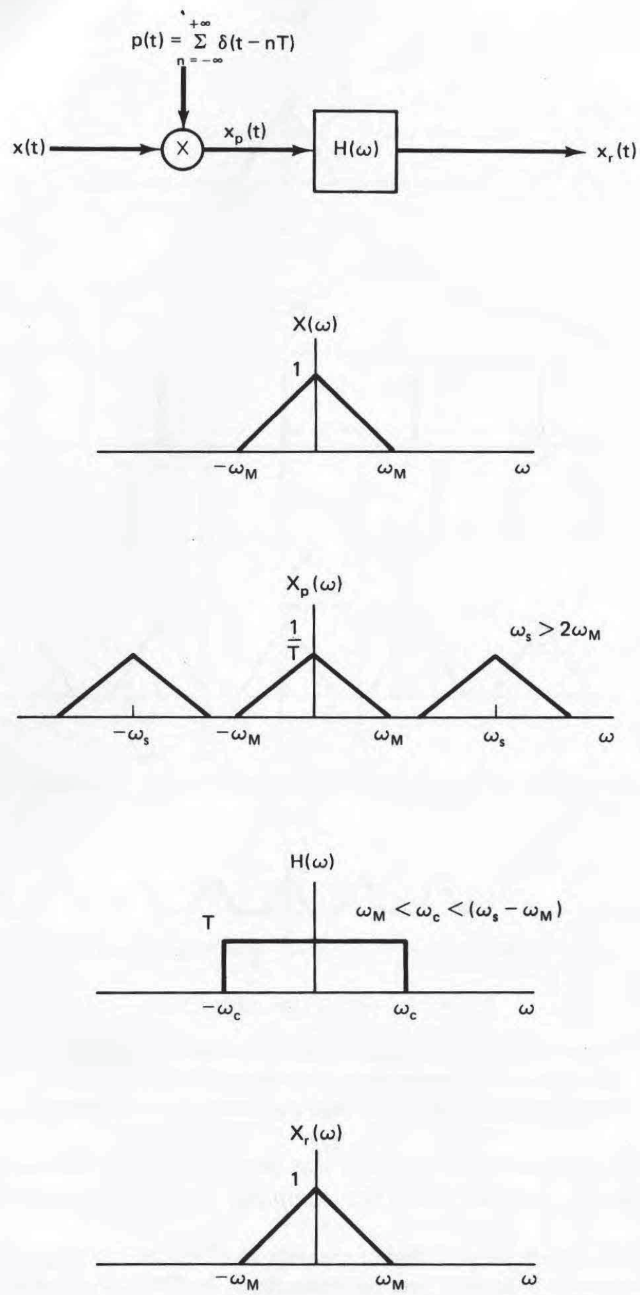


Figure 8.5 Exact recovery of a continuous-time signal from its samples using an ideal lowpass filter.

Sampling Theorem:

Let $x(t)$ be a bandlimited signal with $X(\omega) = 0$ for $|\omega| > \omega_M$. Then $x(t)$ is uniquely determined by its samples $x(nT)$, $n = 0, \pm 1, \pm 2, \ldots$ if

$$\omega_s > 2\omega_M$$

where

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

Given these samples, we can reconstruct $x(t)$ by generating a periodic impulse train in which successive impulses have amplitudes that are successive sample values. This impulse train is then processed through an ideal lowpass filter with gain T and cutoff frequency greater than ω_M and less than $(\omega_s - \omega_M)$. The resulting output signal will exactly equal $x(t)$.

The sampling frequency ω_s is also referred to as the *Nyquist frequency*. The frequency $2\omega_M$, which, under the sampling theorem, must be exceeded by the sampling frequency, is commonly referred to as the *Nyquist rate*.

8.1.2 Sampling with a Zero-Order Hold

The sampling theorem establishes the fact that a bandlimited signal is uniquely represented by its samples, and is motivated on the basis of impulse-train sampling. In practice, narrow large-amplitude pulses, which approximate impulses, are relatively difficult to generate and transmit, and it is often more convenient to generate the sampled signal in a form referred to as a *zero-order hold*. Such a system samples $x(t)$ at a given sampling instant and holds that value until the succeeding sampling instant, as illustrated in Figure 8.6. Reconstruction of $x(t)$ from the output of a zero-order hold

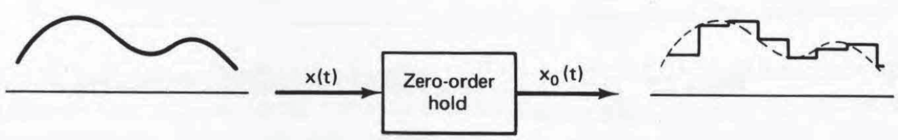


Figure 8.6 Sampling utilizing a zero-order hold.

can again be carried out by lowpass filtering. However, in this case, the required filter no longer has constant gain in the passband. To develop the required filter characteristic, we first note that the output $x_0(t)$ of the zero-order hold can in principle be generated by impulse-train sampling followed by an LTI system with a rectangular impulse response as depicted in Figure 8.7. To reconstruct $x(t)$ from $x_0(t)$, we consider processing $x_0(t)$ with an LTI system with impulse response $h_r(t)$ and frequency response $H_r(\omega)$. The cascade of this system with the system of Figure 8.7 is shown in Figure 8.8, where we wish to specify $H_r(\omega)$ so that $r(t) = x(t)$. Comparing the system in Figure 8.8 with that in Figure 8.5, we see that $r(t) = x(t)$ if the cascade combination of $h_0(t)$ and $h_r(t)$ is the ideal lowpass filter $H(\omega)$ used in Figure 8.5. Since, from

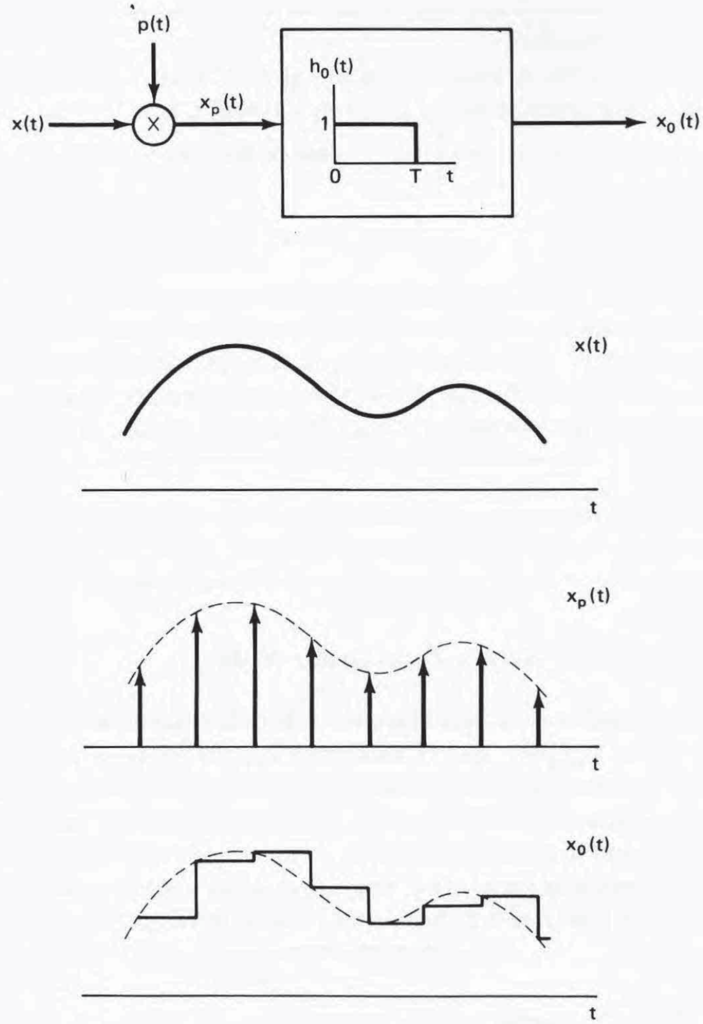


Figure 8.7 Zero-order hold as impulse train sampling followed by convolution with a rectangular pulse.

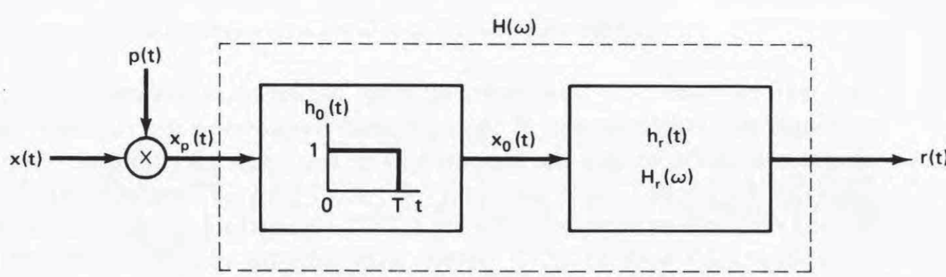


Figure 8.8 Cascade of the representation of a zero-order hold (Figure 8.7) with a reconstruction filter.

Example 4.10 and the time-shifting property 4.6.3

$$H_0(\omega) = e^{-j\omega T/2}\left[\frac{2 \sin (\omega T/2)}{\omega}\right] \tag{8.6}$$

This requires that

$$H_r(\omega) = \frac{e^{j\omega T/2} H(\omega)}{\left[\frac{2 \sin (\omega T/2)}{\omega}\right]} \tag{8.7}$$

For example with the cutoff frequency of $H(\omega)$ as $\omega_s/2$, the ideal magnitude and phase for the reconstruction filter following a zero-order hold is that shown in Figure 8.9.

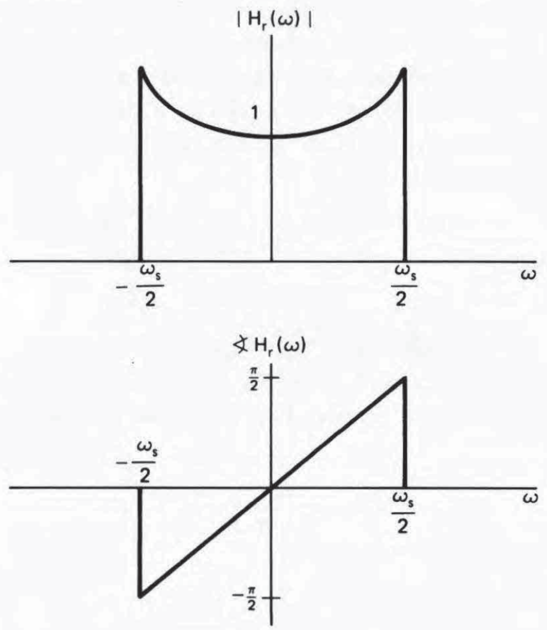


Figure 8.9 Magnitude and phase for reconstruction filter for zero-order hold.

In many situations the zero-order hold is considered to be an adequate approximation to the original signal without any additional lowpass filtering and in essence represents a possible, although admittedly very coarse, interpolation between the sample values. In the next section we explore in more detail the general concept of interpreting the reconstruction of a signal from its samples as a process of interpolation.

8.2 RECONSTRUCTION OF A SIGNAL FROM ITS SAMPLES USING INTERPOLATION

Interpolation is a commonly used procedure for reconstructing a function either approximately or exactly from samples. One simple interpolation procedure is the zero-order hold discussed in Section 8.1. Another simple and useful form of interpolation is *linear interpolation*, whereby adjacent sample points are connected by a straight line as illustrated in Figure 8.10. In more complicated interpolation for-

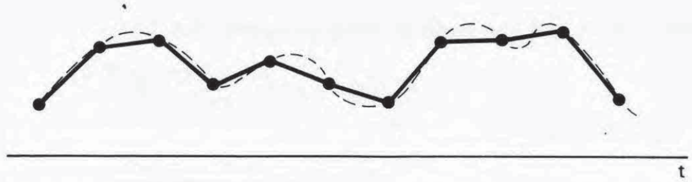


Figure 8.10 Linear interpolation between sample points. The dashed curve represents the original signal and the solid curve the linear interpolation.

mulas, sample points may be connected by higher-order polynomials or other mathematical functions.

As we have seen in Section 8.1, for a bandlimited signal, if the sampling instants are sufficiently close, then the signal can be reconstructed exactly, i.e., through the use of a lowpass filter exact interpolation can be carried out between the sample points. The interpretation of the reconstruction of $x(t)$ as a process of interpolation becomes evident when we consider the effect in the time domain of the lowpass filter in Figure 8.5. In particular, the output $x_r(t)$ is

$$x_r(t) = x_p(t) * h(t)$$

or with $x_p(t)$ given by eq. (8.2),

$$x_r(t) = \sum_{n=-\infty}^{+\infty} x(nT)h(t - nT) \tag{8.8}$$

Equation (8.8) represents an interpolation formula since it describes how to fit a continuous curve between the sample points. For the ideal lowpass filter $H(\omega)$ in Figure 8.5, $h(t)$ is given by

$$h(t) = T\frac{\omega_c}{\pi} \operatorname{sinc}\left(\frac{\omega_c t}{\pi}\right) \tag{8.9}$$

so that

$$x_r(t) = \sum_{n=-\infty}^{+\infty} x(nT)T\frac{\omega_c}{\pi} \operatorname{sinc}\left[\frac{\omega_c(t - nT)}{\pi}\right] \tag{8.10}$$

The reconstruction according to eq. (8.10) with $\omega_c = \omega_s/2$ is illustrated in Figure 8.11.

Interpolation using the sinc function as in eq. (8.10) is commonly referred to as *bandlimited interpolation*, since it implements exact reconstruction if $x(t)$ is bandlimited and the sampling frequency satisfies the conditions of the sampling theorem. Since a very good approximation to an ideal lowpass filter is relatively difficult to implement, in many cases it is preferable to use a less accurate but simpler filter (or equivalently interpolating function) $h(t)$. For example, as we previously indicated, the zero-order hold can be viewed as a form of interpolation between sample values in which the interpolating function $h(t)$ is the impulse response $h_0(t)$ depicted in Figure 8.7. In that sense, with $x_0(t)$ in Figure 8.7 corresponding to the approximation to $x(t)$, the system $h_0(t)$ represents an approximation to the ideal lowpass filter required for the exact interpolation. Figure 8.12 shows the magnitude of the transfer function of the zero-order-hold interpolating filter, superimposed on the desired transfer function of the exact interpolating filter. Both from Figure 8.12 and from Figure 8.7 we see

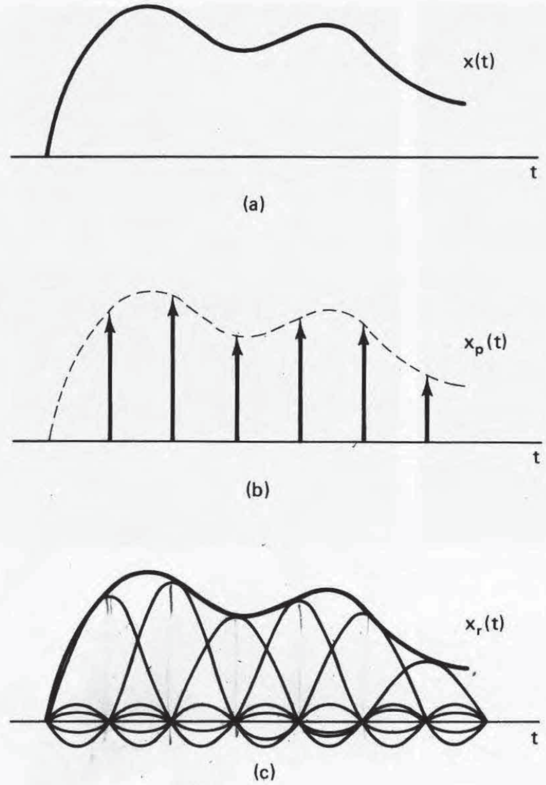


Figure 8.11 Ideal bandlimited interpolation using the sinc function.

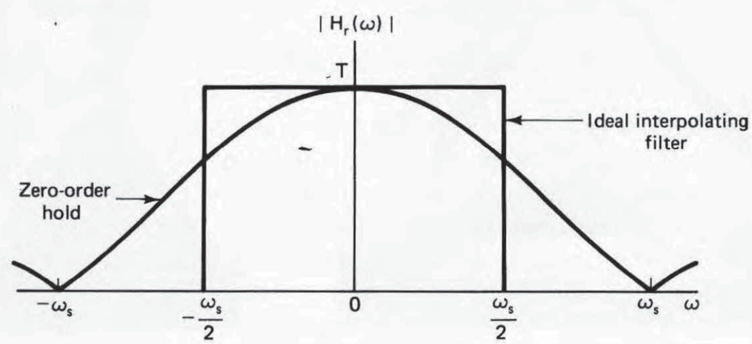
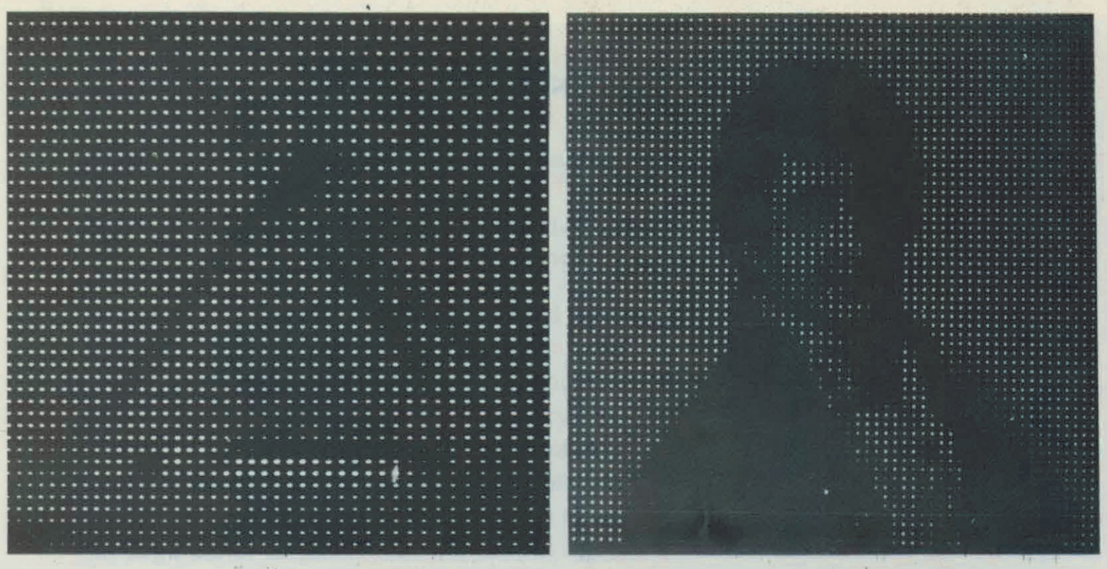


Figure 8.12 Transfer function for the zero-order hold and for the ideal interpolating filter.

that the zero-order hold is a very rough approximation, although in some cases it is sufficient. For example, if, in a given application, there is additional lowpass filtering that is naturally applied, this will tend to improve the overall interpolation. This is illustrated in the case of pictures in Figure 8.13. Figure 8.13(a) shows a picture with "impulse" sampling (i.e., sampling with spatially narrow pulses). Figure 8.13(b) is the result of applying a two-dimensional zero-order hold to Figure 8.13(a) with a resulting mosaic effect when viewed at close range. However, the human visual system inherently



(a)

(b)

Figure 8.13 (a) The original pictures of Figs. 2.2 and 4.2 with impulse sampling; (b) zero-order hold applied to the pictures in (a). The visual system naturally introduces lowpass filtering with a cutoff frequency that increases with distance. Thus, when viewed at a distance, the discontinuities in the mosaic in Figure 8.13(b) are not resolved; (c) result of applying a zero-order hold after impulse sampling with one-half the horizontal and vertical spacing used in (a) and (b).

(c)

Figure 8.13 (cont.)

imposes lowpass filtering, and consequently when viewed at a distance, the discontinuities in the mosaic are not resolved. In Figure 8.13(c) a zero-order hold is again used, but here the sample spacing in each direction is half that in Figure 8.13(a). With normal viewing, considerable lowpass filtering is naturally applied although, particularly with a magnifying glass, the mosaic effect is still somewhat evident.

Another approximate form of interpolation often used is linear interpolation, for which the reconstructed signal is continuous, although its derivative is not. Linear interpolation, sometimes referred to as a first-order hold, was illustrated in Figure 8.10 and can also be viewed as an interpolation in the form of Figure 8.5 and eq. (8.8) with $h(t)$ triangular, as illustrated in Figure 8.14. The associated transfer function $H(\omega)$ is also shown in Figure 8.14 and is given by

$$H(\omega) = \frac{1}{T}\left[\frac{\sin(\omega T/2)}{\omega/2}\right]^2 \qquad (8.11)$$

The transfer function of the first-order hold in Figure 8.14 is shown superimposed on the transfer function for the ideal interpolating filter. Figure 8.15 corresponds to the same pictures as in Figure 8.13 but with a first-order hold applied to the sampled picture.

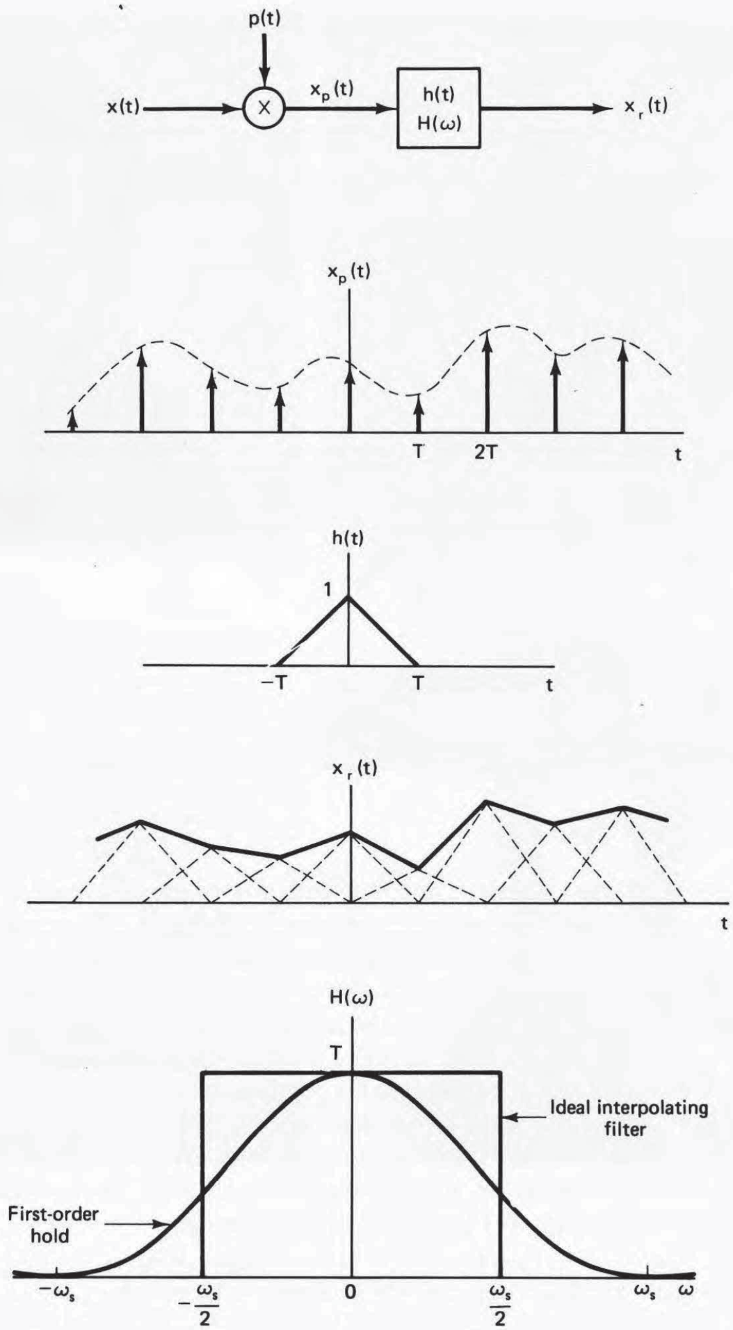


Figure 8.14 Linear interpolation (first-order hold) as impulse-train sampling followed by convolution with a triangular impulse response.

Figure 8.15 Figure 8.13 with a first-order hold applied to the sampled pictures.

8.3 THE EFFECT OF UNDERSAMPLING: ALIASING

In the discussion in previous sections, it was assumed that the sampling frequency was sufficiently high so that the conditions of the sampling frequency were met. As was illustrated in Figure 8.4, with $\omega_s > 2\omega_M$, the spectrum of the sampled signal consists of exact replications of the spectrum of $x(t)$, and this forms the basis for the sampling theorem. When $\omega_s < 2\omega_M$, $X(\omega)$, the spectrum of $x(t)$, is no longer replicated in $X_p(\omega)$ and thus is no longer recoverable by lowpass filtering. This effect, in which the individual terms in eq. (8.5) overlap, is referred to as *aliasing*, and in this section we explore its effect and consequences.

Clearly, if the system of Figure 8.5 is applied to a signal with $\omega_s < 2\omega_M$, the reconstructed signal $x_r(t)$ will no longer be equal to $x(t)$. However, as explored in Problem 8.4 the original signal and the signal $x_r(t)$ which is reconstructed using bandlimited interpolation will always be equal at the sampling instants; that is, for any choice of ω_s,

$$x_r(nT) = x(nT), \qquad n = 0, \pm 1, \pm 2, \ldots \tag{8.12}$$

Some insight into the relationship between $x(t)$ and $x_r(t)$ when $\omega_s < 2\omega_M$ is provided by considering in more detail the comparatively simple case of a sinusoidal signal. Thus, let $x(t)$ be given by

$$x(t) = \cos \omega_0 t \tag{8.13}$$

with Fourier transform $X(\omega)$ as indicated in Figure 8.16(a). In this figure, we have graphically distinguished the impulse at ω_0 from that at $-\omega_0$ for convenience as the discussion proceeds. Let us consider $X_p(\omega)$, the spectrum of the sampled signal and focus in particular on the effect of a change in the frequency ω_0 with the sampling frequency ω_s fixed. In Figure 8.16(b) — (e) we illustrate $X_p(\omega)$ for several values of