Cambridge University Press
978-1-107-05932-0 - Essentials of Digital Signal Processing
B. P. Lathi and Roger A. Green
Excerpt
More information

Chapter 1

Review of Continuous-Time Signals and Systems

This chapter reviews the basics of continuous-time (CT) signals and systems. Although the reader is expected to have studied this background as a prerequisite for this course, a thorough yet abbreviated review is both justified and wise since a solid understanding of continuous-time concepts is crucial to the study of digital signal processing.

Why Review Continuous-Time Concepts?

It is natural to question how continuous-time signals and systems concepts are relevant to digital signal processing. To answer this question, it is helpful to first consider elementary signals and systems structures.

In the most simplistic sense, the study of signals and systems is described by the block diagram shown in Fig. 1.1a. An input signal is fed into a system to produce an output signal. Understanding this block diagram in a completely general sense is quite difficult, if not impossible. A few well-chosen and reasonable restrictions, however, allow us to fully understand and mathematically quantify the character and behavior of the input, the system, and the output.

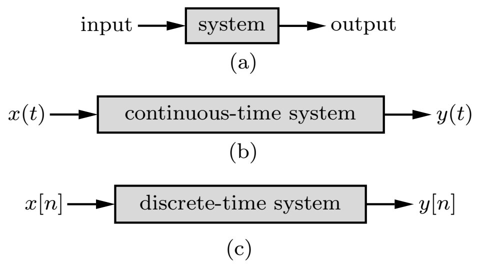

Figure 1.1: Elementary block diagrams of (a) general, (b) continuous-time, and (c) discrete-time signals and systems.

Introductory textbooks on signals and systems often begin by restricting the input, the system, and the output to be continuous-time quantities, as shown in Fig. 1.1b. This diagram captures the basic structure of continuous-time signals and systems, the details of which are reviewed later in this chapter and the next. Restricting the input, the system, and the output to be discrete-time (DT) quantities, as shown in Fig. 1.1c, leads to the topic of discrete-time signals and systems.

Typical digital signal processing (DSP) systems are hybrids of continuous-time and discrete-time systems. Ordinarily, DSP systems begin and end with continuous-time signals, but they process signals using a digital signal processor of some sort. Specialized hardware is required to bridge the continuous-time and discrete-time worlds. As the block diagram of Fig. 1.2 shows, general DSP systems are more complex than either Figs. 1.1b or 1.1c allow; both CT and DT concepts are needed to understand complete DSP systems.

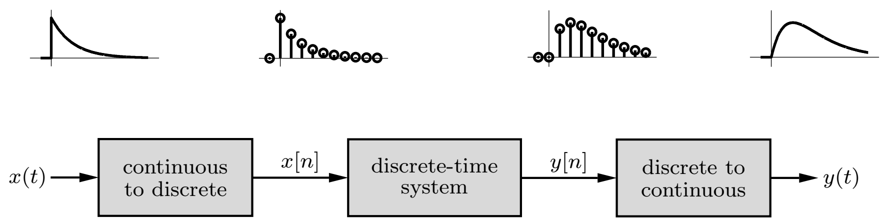

Figure 1.2: Block diagram of a typical digital signal processing system.

A more detailed explanation of Fig. 1.2 helps further justify why it is important for us to review continuous-time concepts. The continuous-to-discrete block converts a continuous-time input signal into a discrete-time signal, which is then processed by a digital processor. The discrete-time output of the processor is then converted back to a continuous-time signal.† Only with knowledge of continuous-time signals and systems is it possible to understand these components of a DSP system. Sampling theory, which guides our understanding of the CT-to-DT and DT-to-CT converters, can be readily mastered with a thorough grasp of continuous-time signals and systems. Additionally, the discrete-time algorithms implemented on the digital signal processor are often synthesized from continuous-time system models. All in all, continuous-time signals and systems concepts are useful and necessary to understand the elements of a DSP system.

Nearly all basic concepts in the study of continuous-time signals and systems apply to the discrete-time world, with some modifications. Hence, it is economical and very effective to build on the previous foundations of continuous-time concepts. Although discrete-time math is inherently simpler than continuous-time math (summation rather than integration, subtraction instead of differentiation), students find it difficult, at first, to grasp basic discrete-time concepts. The reasons are not hard to find. We are all brought up on a steady diet of continuous-time physics and math since high school, and we find it easier to identify with the continuous-time world. It is much easier to grasp many concepts in continuous-time than in discrete-time. Rather than fight this reality, we might use it to our advantage.

1.1 Signals and Signal Categorizations

A *signal* is a set of data or information. Examples include telephone and television signals, monthly sales of a corporation, and the daily closing prices of a stock market (e.g., the Dow Jones averages). In all of these examples, the signals are functions of the independent variable *time*. This is not always the case, however. When an electrical charge is distributed over a body, for instance, the signal is the charge density, a function of *space* rather than time. In this book we deal almost exclusively with signals that are functions of time. The discussion, however, applies equally well to other independent variables.

Signals are categorized as either continuous-time or discrete-time and as either analog or digital. These fundamental signal categories, to be described next, facilitate the systematic and efficient analysis and design of signals and systems.

†As we shall later see, the continuous-to-discrete block is typically comprised of a signal conditioning circuit followed by a CT-to-DT converter and an analog-to-digital converter (ADC). Similarly, the discrete-to-continuous block is typically comprised of a digital-to-analog converter (DAC) followed by a DT-to-CT converter and finally another conditioning circuit.

1.1.1 Continuous-Time and Discrete-Time Signals

A signal that is specified for every value of time t is a *continuous-time signal*. Since the signal is known for every value of time, precise event localization is possible. The tidal height data displayed in Fig. 1.3a is an example of a continuous-time signal, and signal features such as daily tides as well as the effects of a massive tsunami are easy to locate.

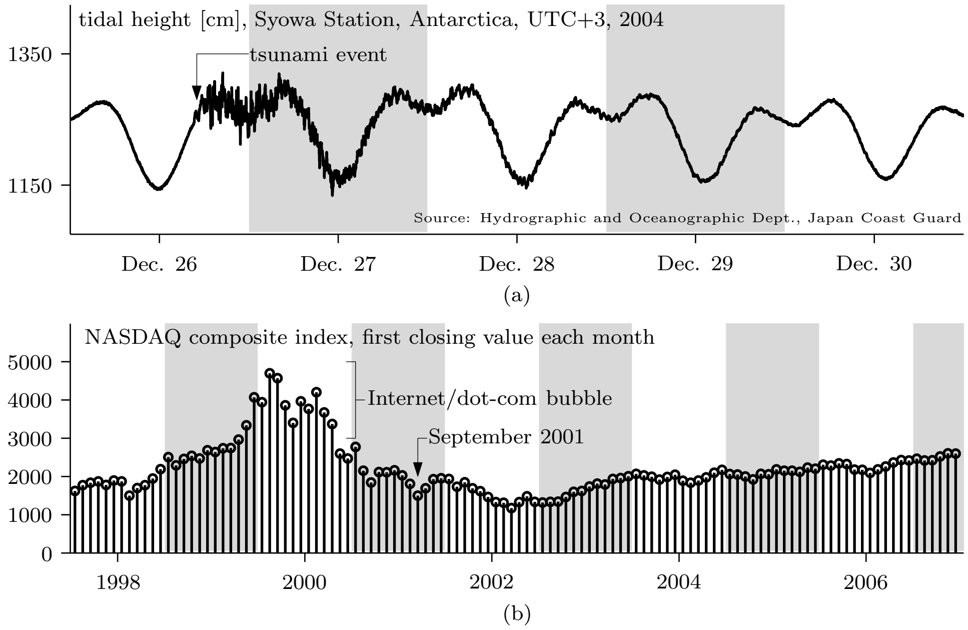

Figure 1.3: Examples of (a) continuous-time and (b) discrete-time signals.

A signal that is specified only at discrete values of time is a *discrete-time signal*. Ordinarily, the independent variable for discrete-time signals is denoted by the integer n. For discrete-time signals, events are localized within the sampling period. The technology-heavy NASDAQ composite index displayed in Fig. 1.3b is an example of a discrete-time signal, and features such as the Internet/dot-com bubble as well as the impact of the September 11 terrorist attacks are visible with a precision that is limited by the one month sampling interval.

1.1.2 Analog and Digital Signals

The concept of continuous-time is often confused with that of analog. The two are not the same. The same is true of the concepts of discrete-time and digital. A signal whose amplitude can take on any value in a continuous range is an *analog signal*. This means that an analog signal amplitude can take on an infinite number of values. A *digital signal*, on the other hand, is one whose amplitude can take on only a finite number of values. Signals associated with typical digital devices take on only two values (binary signals). A digital signal whose amplitudes can take on L values is an *L-ary* signal of which binary ($L = 2$) is a special case.

The terms "continuous-time" and "discrete-time" qualify the nature of a signal along the time (horizontal) axis. The terms "analog" and "digital," on the other hand, qualify the nature of the signal amplitude (vertical axis). Using a sinusoidal signal, Fig. 1.4 demonstrates the various differences. It is clear that analog is not necessarily continuous-time and that digital need not be discrete-time. Figure 1.4c shows, for example, an analog, discrete-time signal. We shall discuss later a systematic procedure for A/D conversion, which involves quantization (rounding off), as explained in Sec. 3.6.

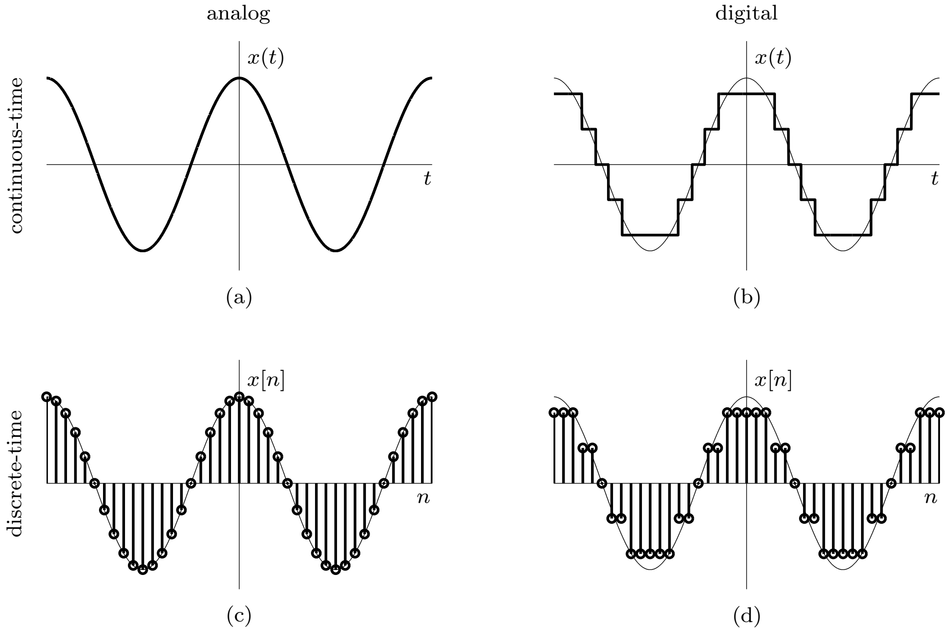

Figure 1.4: Examples of (a) analog, continuous-time, (b) digital, continuous-time, (c) analog, discrete-time, and (d) digital, discrete-time sinusoids.

Signals in the physical world tend to be analog and continuous-time in nature (Fig. 1.4a). Digital, continuous-time signals (Fig. 1.4b) are not common in typical engineering systems. As a result, when we refer to a continuous-time signal, an analog continuous-time signal is implied.

Computers operate almost exclusively with digital, discrete-time data (Fig. 1.4d). Digital representations can be difficult to mathematically analyze, so we often treat computer signals as if they were *analog* rather than digital (Fig. 1.4c). Such approximations are mathematically tractable and provide needed insights into the behavior of DSP systems and signals.

1.2 Operations on the Independent CT Variable

We shall review three useful operations that act on the independent variable of a CT signal: shifting, scaling, and reversal. Since they act on the independent variable, these operations do not change the shape of the underlying signal. Detailed derivations of these operations can be found in [1]. Although the independent variable in our signal description is time, the discussion is valid for functions having continuous independent variables other than time (e.g., frequency or distance).

1.2.1 CT Time Shifting

A signal $x(t)$ (Fig. 1.5a) delayed by $b > 0$ seconds (Fig. 1.5b) is represented by $x(t-b)$. Similarly, the signal $x(t)$ advanced by $b > 0$ seconds (Fig. 1.5c) is represented by $x(t+b)$. Thus, to time shift a signal $x(t)$ by b seconds, we replace t with $t-b$ everywhere in the expression for $x(t)$. If b is positive, the shift represents a time delay; if b is negative, the shift represents a time advance by $|b|$. This is consistent with the fact that a time delay of b seconds can be viewed as a time advance of $-b$ seconds.

Notice that the time-shifting operation is on the independent variable t; the function itself remains unchanged. In Fig. 1.5, the function $x(\cdot)$ starts when its argument equals T_1. Thus, $x(t-b)$ starts

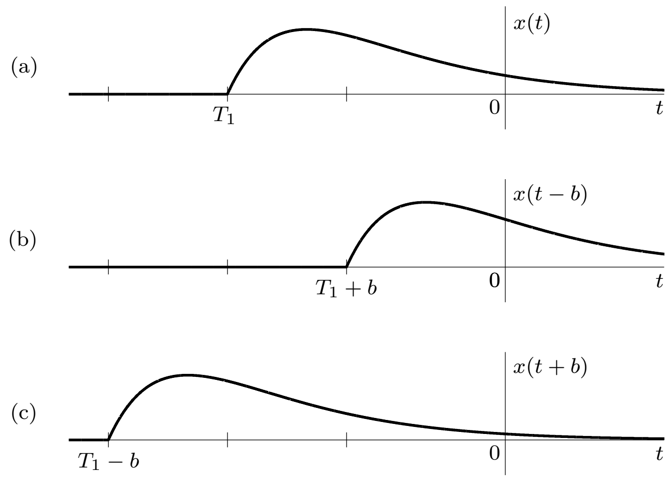

Figure 1.5: Time shifting a CT signal: (a) original signal, (b) delay by b, and (c) advance by b.

when its argument $t-b$ equals T_1, or $t = T_1 + b$. Similarly, $x(t+b)$ starts when $t + b = T_1$, or $t = T_1 - b$.

1.2.2 CT Time Scaling

A signal $x(t)$, when time compressed by factor $a > 1$, is represented by $x(at)$. Similarly, a signal time expanded by factor $a > 1$ is represented by $x(t/a)$. Figure 1.6a shows a signal $x(t)$. Its factor-2 time-compressed version is $x(2t)$ (Fig. 1.6b), and its factor-2 time-expanded version is $x(t/2)$ (Fig. 1.6c). In general, to time scale a signal $x(t)$ by factor a, we replace t with at everywhere in the expression for $x(t)$. If $a > 1$, the scaling represents time compression (by factor a), and if $0 < a < 1$, the scaling represents time expansion (by factor $1/a$). This is consistent with the fact that time compression by factor a can be viewed as time expansion by factor $1/a$.

As in the case of time shifting, time scaling operates on the independent variable and does not change the underlying function. In Fig. 1.6, the function $x(\cdot)$ has a maximum value when its argument equals T_1. Thus, $x(2t)$ has a maximum value when its argument $2t$ equals T_1, or $t = T_1/2$. Similarly, $x(t/2)$ has a maximum when $t/2 = T_1$, or $t = 2T_1$.

▷ **Drill 1.1 (CT Time Scaling)**

Show that the time compression of a sinusoid by a factor a ($a > 1$) results in a sinusoid of the same amplitude and phase, but with the frequency increased a-fold. Similarly, show that the time expansion of a sinusoid by a factor a ($a > 1$) results in a sinusoid of the same amplitude and phase, but with the frequency reduced by a factor a. Verify your conclusions by sketching the sinusoid $\sin(2t)$ and the same sinusoid compressed by a factor 3 and expanded by a factor 2.

◁

1.2.3 CT Time Reversal

Consider the signal $x(t)$ in Fig. 1.7a. We can view $x(t)$ as a rigid wire frame hinged at the vertical axis. To time reverse $x(t)$, we rotate this frame 180° about the vertical axis. This time reversal, or reflection of $x(t)$ about the vertical axis, gives us the signal $x(-t)$ (Fig. 1.7b); whatever happens in Fig. 1.7a at some instant t also happens in Fig. 1.7b at the instant $-t$. Thus, the mirror image of

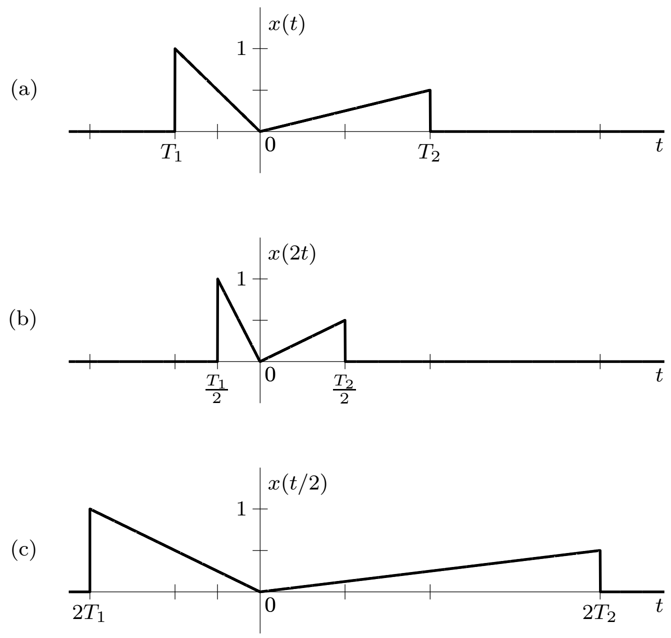

Figure 1.6: Time scaling a CT signal: (a) original signal, (b) compress by 2, and (c) expand by 2.

$x(t)$ about the vertical axis is $x(-t)$. Notice that time reversal is a special case of the time-scaling operation $x(at)$ where $a = -1$.

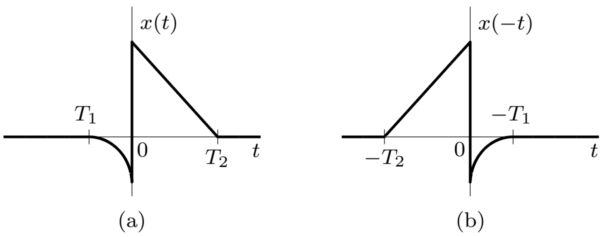

Figure 1.7: Time reversing a CT signal: (a) original signal and (b) its time reverse.

1.2.4 Combined CT Time Shifting and Scaling

Many circumstances require simultaneous use of more than one of the previous operations. The most general case is $x(at - b)$, which is realized in two possible sequences of operations:

1. Time shift $x(t)$ by b to obtain $x(t - b)$. Now time scale the shifted signal $x(t - b)$ by a (i.e., replace t with at) to obtain $x(at - b)$.
2. Time scale $x(t)$ by a to obtain $x(at)$. Now time shift $x(at)$ by $\frac{b}{a}$ (i.e., replace t with $[t - \frac{b}{a}]$) to obtain $x\left(a[t - \frac{b}{a}]\right) = x(at - b)$.

For instance, the signal $x(2t - 6)$ can be obtained in two ways. First, delay $x(t)$ by 6 to obtain $x(t - 6)$ and then time compress this signal by factor 2 (replace t with $2t$) to obtain $x(2t - 6)$. Alternately, we first time compress $x(t)$ by factor 2 to obtain $x(2t)$; next, replace t with $t - 3$ to delay this signal and produce $x(2t - 6)$.

When a is negative, $x(at)$ involves time scaling as well as time reversal. The procedure, however, remains the same. Consider the case of a signal $x(-2t+3)$ where $a=-2$. This signal can be generated by advancing the signal $x(t)$ by 3 seconds to obtain $x(t+3)$. Next, compress and reverse this signal by replacing t with $-2t$ to obtain $x(-2t+3)$. Alternately, we may compress and reverse $x(t)$ to obtain $x(-2t)$; next, replace t with $t-3/2$ to delay this signal by 3/2 and produce $x(-2[t-3/2]) = x(-2t+3)$.

▷ **Drill 1.2 (Combined CT Operations)**

Using the signal $x(t)$ shown in Fig. 1.6a, sketch the signal $y(t) = x(-3t-4)$. Verify that $y(t)$ has a maximum value at $t = \frac{T_1+4}{-3}$.

◁

1.3 CT Signal Models

In the area of signals and systems, the unit step, the unit gate, the unit triangle, the unit impulse, the exponential, and the interpolation functions are very useful. They not only serve as a basis for representing other signals, but their use benefits many aspects of our study of signals and systems. We shall briefly review descriptions of these models.

1.3.1 CT Unit Step Function $u(t)$

In much of our discussion, signals and processes begin at $t=0$. Such signals can be conveniently described in terms of unit step function $u(t)$ shown in Fig. 1.8a. This function is defined by

$$u(t) = \begin{cases} 1 & t > 0 \\ \frac{1}{2} & t = 0 \\ 0 & t < 0 \end{cases}. \tag{1.1}$$

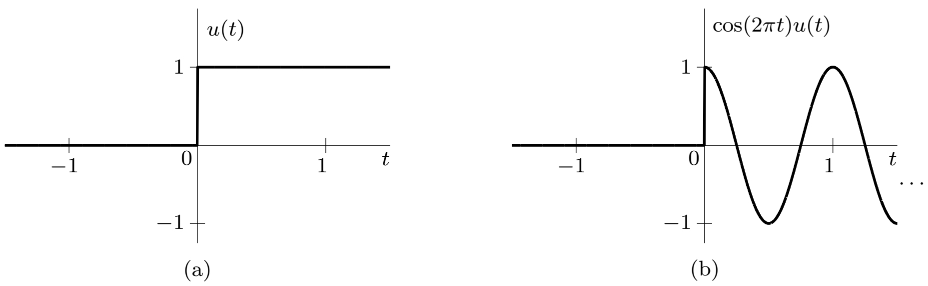

Figure 1.8: (a) CT unit step $u(t)$ and (b) $\cos(2\pi t)u(t)$.

If we want a signal to start at $t=0$ and have a value of zero for $t<0$, we only need to multiply the signal by $u(t)$. For instance, the signal $\cos(2\pi t)$ represents an everlasting sinusoid that starts at $t=-\infty$. The causal form of this sinusoid, illustrated in Fig. 1.8b, can be described as $\cos(2\pi t)u(t)$. The unit step function and its shifts also prove very useful in specifying functions with different mathematical descriptions over different intervals (piecewise functions).

A Meaningless Existence?

It is worth commenting that not everyone defines the point $u(0)$ as 1/2. Some texts define $u(0)$ as 1, others define it as 0, and still others refuse to define it at all. While each definition has its own advantages, $u(0) = 1/2$ is particularly appropriate from a theoretical signals and systems perspective. For real-world signals applications, however, it makes no practical difference how the point $u(0)$ is defined as long as the value is finite. A single point, $u(0)$ or otherwise, is just one among an uncountably infinite set of peers. Lost in the masses, any single, finite-valued point simply does not matter; its individual existence is meaningless.

Further, notice that since it is everlasting, a true unit step cannot be generated in practice. One might conclude, given that $u(t)$ is physically unrealizable and that individual points are inconsequential, that the whole of $u(t)$ is meaningless. This conclusion is false. *Collectively* the points of $u(t)$ are well behaved and dutifully carry out the desired function, which is greatly needed in the mathematics of signals and systems.

1.3.2 CT Unit Gate Function $\Pi(t)$

We define a unit gate function $\Pi(x)$ as a gate pulse of unit height and unit width, centered at the origin, as illustrated in Fig. 1.9a. Mathematically,[†]

$$\Pi(t) = \begin{cases} 1 & |t| < \frac{1}{2} \\ \frac{1}{2} & |t| = \frac{1}{2} \\ 0 & |t| > \frac{1}{2} \end{cases}. \tag{1.2}$$

The gate pulse in Fig. 1.9b is the unit gate pulse $\Pi(t)$ expanded by a factor τ and therefore can be expressed as $\Pi(t/\tau)$. Observe that τ, the denominator of the argument of $\Pi(t/\tau)$, indicates the width of the pulse.

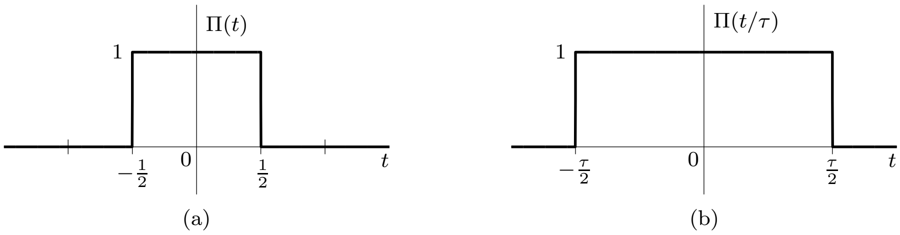

Figure 1.9: (a) CT unit gate $\Pi(t)$ and (b) $\Pi(t/\tau)$.

▷ **Drill 1.3 (CT Unit Gate Representations)**

The unit gate function $\Pi(t)$ can be represented in terms of time-shifted unit step functions. Determine the value b that ensures $u(t + b) - u(t - b)$ is equal to $\Pi(t)$. Next, represent $\Pi(t)$ using only time-shifted and *reflected* unit step functions.

◁

1.3.3 CT Unit Triangle Function $\Lambda(t)$

We define a unit triangle function $\Lambda(t)$ as a triangular pulse of unit height and unit width, centered at the origin, as shown in Fig. 1.10a. Mathematically,

$$\Lambda(t) = \begin{cases} 1 - 2|t| & |t| \leq \frac{1}{2} \\ 0 & |t| > \frac{1}{2} \end{cases}. \tag{1.3}$$

[†]At $|t| = \frac{1}{2}$, we desire $\Pi(t) = 0.5$ because the inverse Fourier transform of a discontinuous signal converges to the mean of the two values at either side of the discontinuity. As in the case of the unit step, the particular value assigned to a point of discontinuity, while perhaps theoretically convenient, has little practical significance.

The pulse in Fig. 1.10b is $\Lambda(t/\tau)$. Observe that here, as for the gate pulse, the denominator τ of the argument of $\Lambda(t/\tau)$ indicates the pulse width.

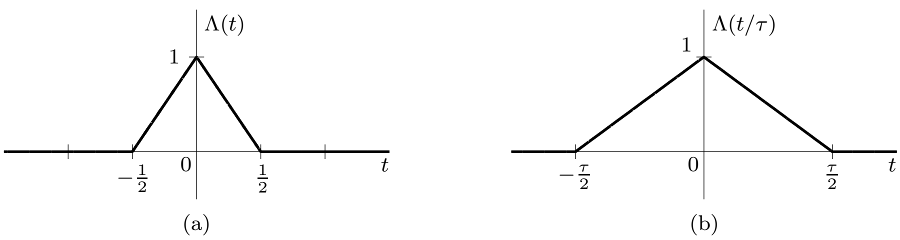

Figure 1.10: (a) CT unit triangle $\Lambda(t)$ and (b) $\Lambda(t/\tau)$.

1.3.4 CT Unit Impulse Function $\delta(t)$

The CT unit impulse function $\delta(t)$ is one of the most important functions in the study of signals and systems. Often called the Dirac delta function, $\delta(t)$ was first defined by P. A. M. Dirac as

$$\begin{gathered} \delta(t) = 0 \text{ for } t \neq 0 \\ \text{and} \\ \int_{-\infty}^{\infty} \delta(t)\, dt = 1. \end{gathered} \tag{1.4}$$

We can visualize this impulse as a tall, narrow rectangular pulse of unit area, as illustrated in Fig. 1.11b. The width of this rectangular pulse is a very small value ϵ, and its height is a very large value $1/\epsilon$. In the limit $\epsilon \to 0$, this rectangular pulse has infinitesimally small width, infinitely large height, and unit area, thereby conforming exactly to the definition of $\delta(t)$ given in Eq. (1.4). Notice that $\delta(t) = 0$ everywhere except at $t = 0$, where it is undefined. For this reason a unit impulse is represented by the spear-like symbol in Fig. 1.11a.

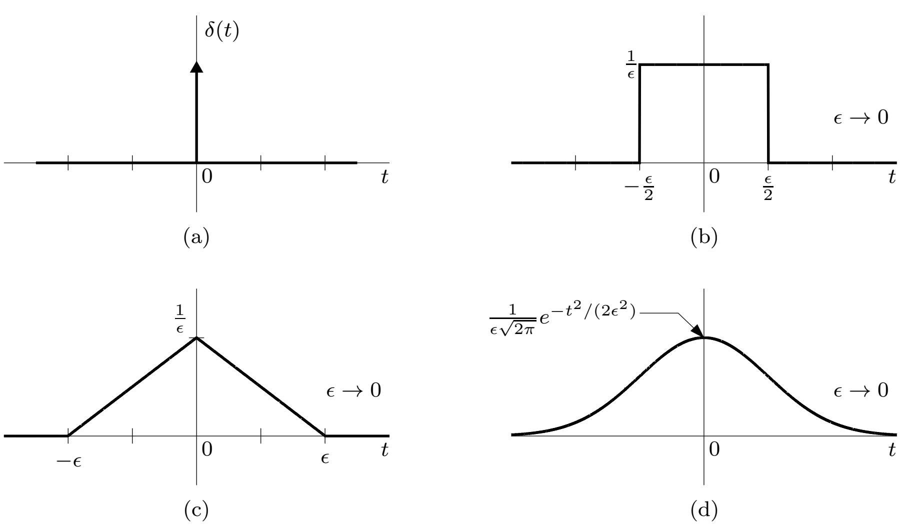

Figure 1.11: (a) CT unit impulse $\delta(t)$ and (b)–(d) visualizing $\delta(t)$ using various functions in the limit $\epsilon \to 0$.

Other pulses, such as the triangle pulse shown in Fig. 1.11c or the Gaussian pulse shown in Fig. 1.11d, may also be used to develop the unit impulse function. The important feature of $\delta(t)$

is not its shape but the fact that its effective duration (pulse width) approaches zero while its area remains at unity. Both the triangle pulse (Fig. 1.11c) and the Gaussian pulse (Fig. 1.11d) become taller and narrower as ϵ becomes smaller. In the limit as $\epsilon \to 0$, the pulse height $\to \infty$, and its width or duration $\to 0$. Yet, the area under each pulse is unity regardless of the value of ϵ.

From Eq. (1.4), it follows that the function $k\delta(t) = 0$ for all $t \neq 0$, and its area is k. Thus, $k\delta(t)$ is an impulse function whose area is k (in contrast to the unit impulse function, whose area is 1). Graphically, we represent $k\delta(t)$ by either scaling our representation of $\delta(t)$ by k or by placing a k next to the impulse.

Properties of the CT Impulse Function

Without going into the proofs, we shall enumerate properties of the unit impulse function. The proofs may be found in the literature (see, for example, [1]).

1. **Multiplication by a CT Impulse:** If a function $\phi(t)$ is continuous at $t = 0$, then

$$\phi(t)\delta(t) = \phi(0)\delta(t).$$

Generalizing, if $\phi(t)$ is continuous at $t = b$, then

$$\phi(t)\delta(t-b) = \phi(b)\delta(t-b). \tag{1.5}$$

2. **The Sampling Property:** If $\phi(t)$ is continuous at $t = 0$, then Eqs. (1.4) and (1.5) yield

$$\int_{-\infty}^{\infty} \phi(t)\delta(t)\,dt = \phi(0)\int_{-\infty}^{\infty} \delta(t)\,dt = \phi(0).$$

Similarly, if $\phi(t)$ is continuous at $t = b$, then

$$\int_{-\infty}^{\infty} \phi(t)\delta(t-b)\,dt = \phi(b). \tag{1.6}$$

Equation (1.6) states that *the area under the product of a function with a unit impulse is equal to the value of that function at the instant where the impulse is located.* This property is very important and useful and is known as the *sampling* or *sifting property* of the unit impulse.

3. **Relationships between $\boldsymbol{\delta(t)}$ and $\boldsymbol{u(t)}$:** Since the area of the impulse is concentrated at one point $t = 0$, it follows that the area under $\delta(t)$ from $-\infty$ to 0^- is zero, and the area is unity once we pass $t = 0$. The symmetry of $\delta(t)$, evident in Fig. 1.11, suggests the area is $1/2$ at $t = 0$. Hence,

$$\int_{-\infty}^{t} \delta(\tau)\,d\tau = u(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{2} & t = 0 \\ 1 & t > 0 \end{cases}. \tag{1.7}$$

From Eq. (1.7) it follows that

$$\delta(t) = \frac{d}{dt}u(t). \tag{1.8}$$

The Unit Impulse as a Generalized Function

The definition of the unit impulse function given in Eq. (1.4) is not rigorous mathematically, which leads to serious difficulties. First, the impulse function does not define a unique function. For example, it can be shown that $\delta(t) + \dot{\delta}(t)$ also satisfies Eq. (1.4). Moreover, $\delta(t)$ is not even a true function in the ordinary sense. An ordinary function is specified by its values for all time t. The