CHAPTER 3

Analog Systems and Signal Spaces

This chapter extends linear systems and Hilbert space ideas to continuous domain signals, filling the gap Chapter 2 left conspicuous. Indeed, noting that an integral over the real line displaces an integral summation, the definitions, theorems, and examples are quite similar in form to their discrete-world cousins. We mainly verify that after replacing summations with integrations we can still construct analog signal spaces that support signal theory.

The initial presentation is informal, not rigorous, and takes a quicker pace. We require normed vector space operations for analog signals: the capability to add, scalar multiply, and measure the size of a signal. The signal spaces should also support limit operations, which imply that arbitrarily precise signal approximations are possible; we find that we can construct analog Banach spaces, too. More important is the measure of similarity between two analog signals—the inner product relation—and we take some care in showing that our abstract structures survive the transition to the analog world. There is an analog Hilbert space theory, for which many of the purely algebraic results of Chapter 2 remain valid. This is convenient, because we can simply quote the same results for analog signals. Hilbert spaces, principally represented by the discrete and analog finite energy signals, will prove to be the most important abstract structure in the sequel. Introductory texts that cover analog signal theory include Refs. 1–5.

The last two sections are optional reading; they supply rigor to the analog theory. One might hope that it is only necessary to replace the infinite summations with infinite integrations, but subtle problems thwart this optimistic scheme. The Riemann integral, familiar from college calculus, cannot handle signals with an infinite number of discontinuities, for example. Its behavior under limit operations is also problematic. Mathematicians faced this same problem at the end of the nineteenth century when they originally developed function space theories. The solution they found—the Lebesgue integral—works on exotic signals, has good limit operation properties, and is identical to the Riemann integral on piecewise continuous functions. Lebesgue measure and integration theory is covered in Section 3.4. In another area, the discrete-world results do not straightforwardly generalize to the analog world: There is no bona fide continuous-time delta function. The intuitive

Signal Analysis: Time, Frequency, Scale, and Structure, by Ronald L. Allen and Duncan W. Mills
ISBN: 0-471-23441-9

treatment, which we offer to begin with, does leave the theory of linear, translation-invariant systems with a glaring hole. It took mathematicians some time to put forward a theoretically sound alternative to the informal delta function concept as well. The concept of a distribution was worked out in the 1930s, and we introduce the theory in Section 3.5.

The chapter contains two important applications. The first, called the matched filter, uses the ideas of inner product and orthogonalization to construct an optimal detector for an analog waveform. The second application introduces the idea of a frame. Frames generalize the concept of a basis, and we show that they are the basic tool for numerically stable pattern detection using a family of signal models. The next chapter applies matched filters to the problem of recognizing signal shapes. Chapter 10 develops frame theory further in the context of time-frequency signal transforms.

3.1 ANALOG SYSTEMS

This section introduces operations on analog signals. Analog signals have a continuous rather than discrete independent time-domain variable. Analog systems operate on analog systems, and among them we find the familiar amplifiers, attenuators, summers, and so on. This section is a quick read for readers who are already familiar with analog systems.

3.1.1 Operations on Analog Signals

For almost every discrete system there corresponds an analog system. We can skim quickly over these ideas, so similar they are to the discrete-world development in the previous chapter.

Definition (Analog System). An analog system H is a partial function from the set of all analog signals to itself. If $x(t)$ is a signal and $y(t)$ is the signal output by H from the input $x(t)$, then $y = Hx$, $y(t) = (Hx)(t)$, or $y(t) = H(x(t))$. As with discrete signals, we call $y(t)$ the response of the system H to input $x(t)$. The set of signals $x(t)$ for which some $y = Hx$ is determined is the domain of the system H. The set of signals y for which $y = Hx$ for some signal x is the range of H.

3.1.2 Extensions to the Analog World

Let us begin by enumerating some common, but not unimportant, analog operations that pose no theoretical problems. As with discrete signals, the operations of scaling (in the sense of amplification and attenuation) and translation (or shifting) an analog signal are central.

We may amplify or attenuate an analog signal by multiplying its values by a constant:

$$y = Ax(t). \tag{3.1}$$

This is sometimes called a scaling operation, which introduces a possible confusion with the notion of dilation $y(t) = x(At)$, which is also called "scaling." The scaling operation inverts the input signal when $A < 0$, amplifies the signal when $|A| > 1$, and attenuates the signal when $|A| < 1$. The domain of a scaling system is all analog signals, as is the range, as long as $A = 0$.

An analog signal may be translated or shifted by any real time value:

$$y = x(t - t_0). \tag{3.2}$$

When $t_0 > 0$ the translation is a delay, and when $t_0 < 0$ the system can more precisely be called an advance. As in the discrete world, translations cause no domain and range problems. If T is an analog time shift, then $\mathrm{Dom}(T) = \mathrm{Ran}(T) = \{s(t): s$ is an analog signal$\}$.

Analog signal reflection reverses the order of signal values: $y(t) = x(-\mathrm{t})$. For analog time signals, this time reversal system reflects the signal values $x(t)$ around the time $t = 0$. As with discrete signals, the reflection and translation operations do not commute.

The basic arithmetic operations on signals exist for the analog world as well. Signal addition or summation adds a given signal to the input, $y(t) = x(t) + x_0(t)$, where $x_0(t)$ is a fixed signal associated with the system H. We can also consider the system that takes the termwise product of a given signal with the input, $y(t) = x(t)x_0(t)$.

One benefit of a continuous-domain variable is that analog signals allow some operations that were impossible or at least problematic in the discrete world.

Dilation always works in the analog world. We can form $y(t) = x(at)$ whatever the value of $a \in \mathbb{R}$. The corresponding discrete operation, $y(n) = x(bn)$, works nicely only if $b \in \mathbb{Z}$ and $|b| \geq 1$; when $0 < |b| < 1$ and $b \in \mathbb{Q}$ we have to create special values (typically zero) for those $y(n)$ for which $b \in \mathbb{Z}$. As noted earlier, dilation is often called scaling, because it changes the scale of a signal. Dilation enlarges or shrinks signal features according to whether $|a| < 1$ or $|a| > 1$, respectively.

Another analog operation is differentiation. If it is smooth enough, we can take the derivative of an analog signal:

$$y(t) = \frac{dx}{dt} = x'(t). \tag{3.3}$$

If the signal $x(t)$ is only piecewise differentiable in (3.3), then we can assign some other value to $y(t)$ at the points of nondifferentiability.

3.1.3 Cross-Correlation, Autocorrelation, and Convolution

The correlation and convolution operations depend on signal integrals. In the discrete world, systems that implement these operations have input–output relations that involve infinite summations over the integers. In continuous-domain signal processing, the corresponding operations rely on integrations over the entire real line.

These do pose some theoretical problems—just as did the infinite summations reminiscent of Chapter 2; we shall address them later when we consider signal spaces of analog signals.

The analog *convolution* operation is once again denoted by the * operator: $y = x*h$. We define:

$$y(t) = (x*h)(t) = \int_{-\infty}^{\infty} x(s)h(t-s)\,ds. \tag{3.4}$$

The *cross-correlation* system is defined by the rule $y = x\circ h$, where

$$y(t) = (x\circ h)(t) = \int_{-\infty}^{\infty} x(s)h(t+s)\,ds. \tag{3.5}$$

The analog *autocorrelation* operation on a signal is $y = x\circ x$, and when the signals are complex-valued, we use the complex conjugate of the kernel function $h(t)$:

$$y(t) = (x\circ h)(t) = \int_{-\infty}^{\infty} x(s)\overline{h(t+s)}\,ds. \tag{3.6}$$

The autocorrelation is defined by $y(t) = (x\circ x)(t)$. One of the applications of functional analysis ideas to signal processing, which we shall provide below, is to show the existence of the correlation and autocorrelation functions for square-integrable signals $x(t)$ and $h(t)$.

We can show that linear translation invariant analog systems are again characterized by the convolution operation. This is not as easy as it was back in the discrete realm. We have no analog signal that corresponds to the discrete impulse, and discovering the right generalization demands that we invent an entirely new theory: distributions.

3.1.4 Miscellaneous Operations

Let us briefly survey other useful analog operations.

A subsampling or downsampling system continuously expands or contracts an analog signal: $y(t) = x(at)$, where $a > 0$ is the scale or dilation factor. Tedious as it is to say, we once more have a terminology conflict; the term “scale” also commonly refers in the signal theory literature to the operation of amplifying or attenuating a signal: $y(t) = ax(t)$.

Analog thresholding is a just as simple as in the discrete world:

$$y(t) = \begin{cases} 1 & \text{if } x(t) \geq T, \\ 0 & \text{if } x(t) < T. \end{cases} \tag{3.7}$$

The accumulator system, $y = Hx$, is given by

$$y(t) = \int_{-\infty}^{t} x(s)\, ds. \tag{3.8}$$

The accumulator outputs a value that is the sum of all input values to the present signal instant. As already noted, not all signals are in the domain of an accumulator system. The exercises explore some of these ideas further.

The moving average system is given by

$$y(t) = \frac{1}{2a} \int_{t-a}^{t+a} x(s)\, ds, \tag{3.9}$$

where $a > 0$. This system averages $x(s)$ around in an inteval of width $2a$ to output $y(t)$.

3.2 CONVOLUTION AND ANALOG LTI SYSTEMS

The characterization of a linear, translation-invariant analog system as one given by the convolution operation holds for the case of continuous domain signals too. We take aim at this idea right away. But, the notion of an impulse response—so elementary it is an embarassment within discrete system theory—does not come to us so readily in the analog world. It is not an understatement to say that the proper explication of the analog delta requires supplementary theory; what it demands is a complete alternative conceptualization of the mathematical representation of analog signals. We will offer an informal definition for the moment, and this might be the prudent stopping point for first-time readers. We shall postpone the more abstruse development, known as distribution theory, until Section 3.5.

3.2.1 Linearity and Translation-Invariance

Analog systems can be classified much like discrete systems. The important discrete signal definitions of linearity and translation- (or shift- or time-) invariance extend readily to the analog world.

Definition (Linear System). An analog system H is linear if $H(ax) = aH(x)$ and $H(x + y) = H(x) + H(y)$. Often the system function notation drops the parentheses; thus, we write Hx instead of $H(x)$. $H(x)$ is a signal, a function of a time variable, and so we use the notation $y(t) = (Hx)(t)$ to include the independent variable of the output signal.

Definition (Translation-Invariant). An analog system H is translation-invariant if whenever $y = Hx$ and $s(t) = x(t - a)$, then $H(s) = y(t - a)$.

A linear system obeys the principles of scaling and superposition. When a system is translation-invariant, then the output of the shifted input is precisely the shifted output.

Definition (LTI System). An LTI system is both linear and translation-invariant.

Let us consider some examples.

Example. Let the system $y = Hx$ be given by $y(t) = x(t)\cos(t)$. The cosine term is a nonlinear distraction, but this system is linear. Indeed, $H(x_1 + x_2)(t) = [x_1(t) + x_2(t)]\cos(t) = x_1(t)\cos(t) + x_2(t)\cos(t) = H(x_1)(t) + H(x_2)(t)$. Also, $H(ax)(t) = [ax(t)]\cos(t) = a[x(t)\cos(t)] = a(Hx)(t)$.

Example. Let $y = Hx$ by given by $y(t) = tx(t)$. Then H is not translation-invariant. The decision about whether a system is or is not translation-invariant can sometimes bedevil signal processing students. The key idea is to hide the shift amount inside a new signal's definition: Let $w(t) = x(t - a)$. Then $w(t)$ is the shifted input signal. $(Hw)(t) = tw(t)$ by definition of the system H. But $tw(t) = tx(t - a)$. Is this the shifted output? Well, the shifted output is $y(t - a) = (t - a)x(t - a)$. In general, this will not equal $tx(t - a)$, so the system H is not shift-invariant.

Example. Let $y = Hx$, where $y(t) = x^2(t) + 8$. This system is translation-invariant. Again, let $w(t) = x(t - a)$, so that $(Hw)(t) = w^2(t) + 8 = x(t - a)x(t - a) + 8$ is the output of the translated input signal. Is this the translated output signal? Yes, because $y(t - a) = x^2(t - a) + 8 = (Hw)(t)$.

Example (Moving Average). Let $T > 0$ and consider the system $y = Hx$:

$$y(t) = \int_{t-T}^{t+T} x(s)\, ds. \tag{3.10}$$

Then H is LTI. The integration is a linear operation, which is easy to show. So let us consider a translated input signal $w(t) = x(t - a)$. Then $(Hw)(t)$ is

$$\int_{t-T}^{t+T} w(s)\, ds = \int_{t-T}^{t+T} x(s-a)\, ds = \int_{t-T-a}^{t+T-a} x(u)\, du, \tag{3.11}$$

where we have changed the integration variable with $u = s - a$. But note that the shifted output is

$$y(t-a) = \int_{t-a-T}^{t-a+T} x(s)\, ds, \tag{3.12}$$

and (3.12) is identical to (3.11).

3.2.2 LTI Systems, Impulse Response, and Convolution

Putting aside mathematical formalities, it is possible to characterize analog linear, translation-invariant systems by convolution of the input signal with the system impulse response.

3.2.2.1 Analog Delta and Impulse Response. Let us begin by developing the idea of the analog delta function, or Dirac[1] delta, $\delta(t)$. This signal should—like the discrete delta, $\delta(n)$—be zero everywhere except at time $t = 0$. Discrete convolution is a discrete sum, so $\delta(0) = 1$ suffices for sifting out values of discrete siganls $x(n)$. Analog convolution is an integral, and if $\delta(t)$ is a signal which is nonzero only at $t = 0$, then the integral of any integrand $x(s)\delta(t - s)$ in the convolution integral (3.4) is zero. Consequently, it is conventional to imagine the analog impulse signal as being infinite at $t = 0$ and zero otherwise; informally, then,

$$\delta(t) = \begin{cases} \infty & \text{if } t \neq 0, \\ 0 & \text{if otherwise.} \end{cases} \tag{3.13}$$

Another way to define the analog delta function is through the following convolutional identity:

Sifting Property. The analog impulse is the signal for which, given analog signal $x(t)$,

$$x(t) = (x * \delta)(t) = \int_{-\infty}^{\infty} x(s)\delta(t-s)\, ds. \tag{3.14}$$

No signal satisfying (3.13) or having the property (3.14) exists, however. The choice seems to be between the Scylla of an impossible function or the Charybdis of an incorrect integration.

To escape this quandary, let us try to approximate the ideal, unattainable analog sifting property by a local average. Let $\delta_n(t)$ be defined for $n > 0$ by

$$\delta_n(t) = \begin{cases} n & \text{if } t \in \left[-\frac{1}{2n}, \frac{1}{2n}\right], \\ 0 & \text{if otherwise.} \end{cases} \tag{3.15}$$

[1]P. A. M. Dirac (1902–1984) applied the delta function to the discontinuous energy states found in quantum mechanics (*The Principles of Quantum Mechanics*, Oxford: Clarendon, 1930). Born in England, Dirac studied electrical engineering and mathematics at Bristol and then Cambridge, respectively. He developed the relativistic theory of the electron and predicted the existence of the positron.

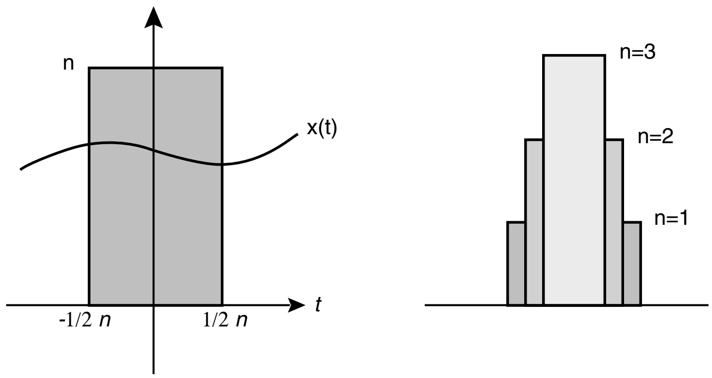

Fig. 3.1. Approximating the analog impulse, $\delta_n(t) = n$ on the interval of width $1/n$ around the origin.

Then, referring to Figure 3.1, $\delta_n(t)$ integrates to unity on the real line,

$$\int_{-\infty}^{\infty} \delta_n(t)\, dt = 1. \tag{3.16}$$

Furthermore, $\delta_n(t)$ provides a rudimentary sifting relationship,

$$\int_{-\infty}^{\infty} x(t)\delta_n(t)\, dt = \text{Average value of } x(t) \text{ on} \left[-\frac{1}{2n}, \frac{1}{2n}\right]. \tag{3.17}$$

To verify this, note that

$$\int_{-\infty}^{\infty} x(t)\delta_n(t)\, dt = n \int_{-1/2n}^{1/2n} x(t)\, dt, \tag{3.18}$$

while

$$\int_{a}^{b} x(t)\, dt = (b-a) \times \text{Average value of } x(t) \text{ on } \left[-\frac{1}{2n}, \frac{1}{2n}\right]. \tag{3.19}$$

Combining (3.18) and (3.19) proves that $\delta_n(t)$ has a sifting-like property. Note that as $n \to \infty$ the square impluse in Figure 3.1 grows higher and narrower, approximating an infinitely high spike. Under this limit the integral (3.17) is the average of $x(t)$ within an increasingly minute window around $t = 0$:

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} x(t)\delta_n(t)\, dt = x(0). \tag{3.20}$$

This argument works whatever the location of the tall averaging rectangles defined by $\delta_n(t)$. We can conclude that

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} x(s)\delta_n(t-s)\ ds = x(t). \tag{3.21}$$

An interchange of limit and integration in (3.21) gives

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} x(s)\delta_n(t-s)\ ds = \int_{-\infty}^{\infty} x(s) \lim_{n \to \infty} \delta_n(t-s)\ ds. \tag{3.22}$$

Finally, and if we assume $\delta_n(t) \to \delta(t)$, that the square spikes converge to the Dirac delta, then the sifting property follows from (3.22).

Rigorously speaking, of course, there is no signal $\delta(t)$ to which the $\delta_n(t)$ converge, no such limit exists, and our interchange of limits is quite invalid [6]. It is possible to formally substantiate delta function theory with the theory of generalized functions and distributions [7–10]. Some applied mathematics texts take time to validate their use of Diracs [11–13]. In fact, the amendments follow fairly closely the informal motivation that we have provided above. Despite our momentary neglect of mathematical justification, at this point these ideas turn out to be very useful in analog signal theory; we shamelessly proceed to feed delta "functions" into linear, translation-invariant systems.

Definition (Impulse Response). Let H be an analog LTI system and $\delta(t)$ be the Dirac delta function. Then the impulse response of H is $h(t) = (H\delta)(t)$. Standard notation uses a lowercase "h" for the impulse response of the LTI system "H."

The next section applies these ideas toward a characterization of analog LTI systems.

3.2.2.2 LTI System Characterization. The most important idea in analog system theory is that a convolution operation—the input signal with the system's impulse response—characterizes LTI systems. We first rewrite an analog signal $x(t)$ as a scaled sum of shifted impulses. Why take a perfectly good—perhaps infinitely differentiable—signal and write it as a linear combination of these spikey, problematic components? This is how, in Chapter 2, we saw that convolution governs the input–output relation of an LTI discrete system.

To decompose $x(t)$ into Diracs, note that any integral is a limit of Riemann sums of decreasing width:

$$\int_{-\infty}^{\infty} f(s)ds = \lim_{Len(I) \to 0} \sum_{n=-\infty}^{\infty} f(a_n)Len(I_n), \tag{3.23}$$

where $I = \{I_n: n \in \mathbb{Z}\}$ is a partition of the real line, $I_n = [a_n, b_n]$, the length of interval I_n is $\mathrm{Len}(I_n) = b_n - a_n$, and $\mathrm{Len}(I) = \max\{\mathrm{Len}(I_n): n \in \mathbb{Z}\}$. If $f(s) = x(s)\delta(t - s)$, then

$$x(t) = \int_{-\infty}^{\infty} f(s)\, ds = \lim_{Len(I) \to 0} \sum_{n=-\infty}^{\infty} x(a_n)\delta(t - a_n)\, Len(I_n), \tag{3.24}$$

so that

$$y(t) = H(x)(t) = H\left[\lim_{Len(I) \to 0} \sum_{n=-\infty}^{\infty} x(a_n)\delta(t - a_n) Len(I_n)\right]$$
$$= \lim_{Len(I) \to 0} \sum_{n=-\infty}^{\infty} x(a_n) H[\delta(t - a_n)] Len(I_n) \tag{3.25}$$

Besides blithely swapping the limit and system operators, (3.25) applies linearity twice: scaling with the factor $x(a_n)$ and superposition with the summation. By translation-invariance, $H[\delta(t - a_n)] = (H\delta)(t - a_n)$. The final summation above is an integral itself, precisely the LTI system characterization we have sought:

Theorem (Convolution for LTI Systems). Let H be an LTI system, $y = Hx$, and $h = Hd$. Then

$$y(t) = (x*h)(t) = \int_{-\infty}^{\infty} x(s)h(t - s)\, ds. \tag{3.26}$$

Now let us consider a basic application of the theorem to the study of stable and causal systems.

3.2.2.3 Stable and Causal Systems.

The same ideas of stability and causality, known from discrete system theory, apply to analog systems.

Definition (Bounded). A signal $x(t)$ is *bounded* if there is a constant M such that $|x(t)| < M$ for all $t \in \mathbb{R}$.

Definition (Absolutely Integrable). A signal $x(t)$ is *absolutely integrable* if

$$\int_{-\infty}^{\infty} |x(t)|\, dt < \infty. \tag{3.27}$$

The equivalent concept in Chapter 2 is *absolutely summable*.

Definition (Stable System). If $y = Hx$ is a bounded output signal whenever $x(t)$ is a bounded input signal, then the system H is *stable*.

Proposition (Stability Characterization). The LTI system $y = Hx$ is stable if and only if its impulse response $h = H\delta$ is absolutely integrable.

Proof: Suppose h is absolutely integrable and $|x(t)| < M$ for all $t \in \mathbb{R}$. Then

$$|y(t)| = \left| \int_{-\infty}^{\infty} x(s)h(t-s)\,ds \right| \le \int_{-\infty}^{\infty} |x(s)h(t-s)|\,ds \le M \int_{-\infty}^{\infty} |h(t-s)|\,ds. \tag{3.28}$$

Since the final integral in (3.28) is finite, $y(t)$ is bounded and H must be stable. Conversely, suppose that H is stable but $h = H\delta$ is not absolutely integrable:

$$\int_{-\infty}^{\infty} |h(t)|\,dt = \infty; \tag{3.29}$$

we seek a contradiction. If

$$x(t) = \begin{cases} \dfrac{\overline{h(-t)}}{|h(-t)|} & \text{if } h(-t) \ne 0, \\ 0 & \text{if otherwise,} \end{cases} \tag{3.30}$$

where we are allowing for the situation that $h(t)$ is complex-valued, then $x(t)$ is clearly bounded. But convolution with $h(t)$ governs the input-output relation of H, and that is the key. We can write $y = Hx$ as

$$y(t) = \int_{-\infty}^{\infty} x(s)h(t-s)\,ds = \int_{-\infty}^{\infty} \frac{\overline{h(-s)}}{|h(-s)|} h(t-s)\,ds. \tag{3.31}$$

So $y(t)$ is the output of stable system H given bounded input $x(t)$. $y(t)$ should be bounded. What is $y(0)$? Well,

$$\begin{aligned} y(0) &= \int_{-\infty}^{\infty} x(s)h(0-s)\,ds = \int_{-\infty}^{\infty} \frac{\overline{h(-s)}}{|h(-s)|} h(0-s)\,ds = \int_{-\infty}^{\infty} \frac{|h(-s)|^2}{|h(-s)|}\,ds \\ &= \int_{-\infty}^{\infty} \frac{|h(-s)|^2}{|h(-s)|}\,ds = \int_{-\infty}^{\infty} |h(-s)|\,ds. \end{aligned} \tag{3.32}$$

However, the final integral in (3.32) is infinite by (3.29). So $y(t)$ cannot be bounded, and this contradicts the converse's assumption. ■

To the discrete theory, there also corresponds an analog notion of causality. A system is causal if the future of the input signals have no bearing on their response.

Definition (Causal System). The system H is *causal* if $y = Hx$ can be found using only present and past values of $x(t)$.

Proposition (Causality Criterion). The LTI system $y = Hx$ is causal if its impulse response $h = H\delta$ satisfies $h(r) = 0$ for $r < 0$.

Proof: If $h(t - s) = 0$ for $s > t$, then jotting down the convolution integral shows

$$y(t) = \int_{-\infty}^{\infty} x(s)h(t-s)\, ds = \int_{-\infty}^{t} x(s)h(t-s)\, ds. \tag{3.33}$$

We have written $y(t)$ knowing only earlier $x(s)$ values, and so H must be causal. ■

Remark. For analog signals, the converse of the causality criterion is not true. The impulse response $h(t)$ could be nonzero at an isolated point, say $t = -2$, and so the convolution integral (3.33) does not depend on negative times.

Example. Let H be the system with impulse response

$$h(t) = \begin{cases} e^{-t} & \text{if } t \geq 0, \\ 1 & \text{if } t = -1, \\ 0 & \text{if otherwise.} \end{cases} \tag{3.34}$$

The system $y(t) = (x*h)(t)$ is LTI, and H is causal, but it does not obey the converse of the causality criterion.

The causality criterion's lack of a converse distinguishes analog from discrete theory. In discrete system theory, an LTI system H is causal if and only if its impulse response $h(n)$ is zero for $n < 0$. We can improve on the proposition, by developing a stronger integral. Indeed, the fact that we might allow signals that have isolated finite impulses, or even a countably infinite number of finite impulse discontinuities points to the Riemann integral's inadequacy. This same defect troubled mathematicians in the early 1900s, when they first drew set theoretic concepts into analysis and began to demand limit operations from the integral. Modern measure and integration theory was the outcome, and we cover it in Section 3.4.

3.2.3 Convolution Properties

Convolution is the most important signal operation, since $y = h*x$ gives the input–output relation for the LTI system H, where $h = H\delta$. This section develops basic theorems, these coming straight from the properties of the convolution integral.

Proposition (Linearity). The convolution operation is linear: $h*(ax) = ah*x$, and $h*(x + y) = h*x + h*y$.

Proof: Easy by the linearity of integration (exercise). ■

Proposition (Translation-Invariance). The convolution operation is translation invariant: $h*[x(t-a)] = (h*x)(t-a)$.

Proof: It is helpful to hide the time shift in a signal with a new name. Let $w(t) = x(t-a)$, so that the translated input is $w(t)$. We are asking, what is the convolution of the shifted input? It is $(h*w)(t)$. Well,

$$(h*w)(t)= \int_{-\infty}^{\infty} h(s)w(t-s)\, ds = \int_{-\infty}^{\infty} h(s)x((t-a)-s)\, ds = (h*x)(t-a), \quad (3.35)$$

which is the translation of the output by the same amount. ■

These last two propositions comprise a converse to the convolution theorem (3.26) for LTI systems: a system $y = h*x$ is an LTI system. The next property shows that the order of LTI processing steps does not matter.

Proposition (Commutativity). The convolution operation is commutative: $x*y = y*x$.

Proof: Let $u = t - s$ for a change of integration variable. Then $du = -ds$, and

$$(x*y)(t) = \int_{-\infty}^{\infty} x(s)y(t-s)\, ds = -\int_{\infty}^{-\infty} x(t-u)y(u)\, du = \int_{-\infty}^{\infty} x(t-u)y(u)\, du. \quad (3.36)$$

The last integral in (3.36) we recognize to be the convolution $y*x$. ■

Proposition (Associativity). The convolution operation is associative: $h*(x*y) = h*(x*y)$.

Proof: Exercise. ■

Table 3.1 summarizes the above results.

TABLE 3.1. Convolution Properties

Signal Expression	Property Name
$(x*y)(t) = \int_{-\infty}^{\infty} x(s)y(t-s)\, ds$	Definition
$h*[ax + y] = ah*x + h*y$	Linearity
$h*[x(t-a)] = (h*x)(t-a)$	Translation- or shift-invariance
$x*y = y*x$	Commutativity
$h*(x*y) = (h*x)*y$	Associativity
$h*(\delta(t-a)) = h(a)$	Sifting

3.2.4 Dirac Delta Properties

The Dirac delta function has some unusual properties. Although they can be rigorously formulated, they also follow from its informal description as a limit of ever higher and narrower rectangles. We maintain a heuristic approach.

Proposition. Let $u(t)$ be the unit step signal. Then $\delta(t) = \frac{d}{dt}[u(t)]$.

Proof: Consider the functions $u_n(t)$:

$$u_n(t) = \begin{cases} 1 & \text{if } t \geq \frac{1}{2n}, \\ 0 & \text{if } t \leq -\frac{1}{2n}, \\ nt + \frac{1}{2} & \text{if otherwise.} \end{cases} \tag{3.37}$$

Notice that the derivatives of $u_n(t)$ are precisely the $\delta_n(t)$ of (3.15) and that as $n \to \infty$ we have $u_n(t) \to u(t)$. Taking the liberty of assuming that in this case differentiation and the limit operations are interchangeable, we have

$$\delta(t) = \lim_{n \to \infty} \delta_n(t) = \lim_{n \to \infty} \frac{d}{dt} u_n(t) = \frac{d}{dt} \lim_{n \to \infty} u_n(t) = \frac{d}{dt} u(t). \tag{3.38}$$

■

Remarks. The above argument does stand mathematical rigor on its head. In differential calculus [6] we need to verify several conditions on a sequence of signals and their derivatives in order to conclude that the limit of the derivatives is the derivative of their limit. In particular, one must verify the following in some interval I around the origin:

(i) Each function in $\{x_n(t) \mid n \in \mathbb{N}\}$ is continuously differentiable on I.
(ii) There is an $a \in I$ such that $\{x_n(a)\}$ converges.
(iii) The sequence $\{x_n'(t)\}$ converges uniformly on I.

Uniform convergence of $\{y_n(t) \mid n \in \mathbb{N}\}$ means that for every $\varepsilon > 0$ there is an $N_\varepsilon > 0$ such that $m, n > N_\varepsilon$ implies $|y_n(t) - y_m(t)| < \varepsilon$ for all $t \in I$. The key distinction between uniform convergence and ordinary, or *pointwise*, convergence is that uniform convergence requires that an N_ε be found that pinches $y_n(t)$ and $y_m(t)$ together throughout the interval. If the N_ε has to depend on $t \in I$ (such as could happen if the derivatives of the $y_n(t)$ are not bounded on I), then we might find pointwise convergence exists, but that uniform convergence is lost. The exercises delve further into these ideas. Note also that a very important instance of this distinction occurs in Fourier series convergence, which we detail in Chapters 5 and 7. There we show that around a discontinuity in a signal $x(t)$ the Fourier series is pointwise but not uniformly convergent. Connected with this idea is the famous ringing artifact of Fourier series approximations, known as the *Gibbs phenomenon*.

The next property describes the scaling or time dilation behavior of $\delta(t)$. Needless to say, it looks quite weird at first glance.

Proposition (Scaling Property). Let $u(t)$ be the unit step signal and $a \in \mathbb{R}$, $a \neq 0$. Then

$$\delta(at) = \frac{1}{|a|}\delta(t). \tag{3.39}$$

Proof: Consider the rectangles approaching $\delta(t)$ as in Figure 3.1. If the $\delta_n(t)$ are dilated to form $\delta_n(at)$, then in order to maintain unit area, we must alter their height; this makes the integrals an average of $x(t)$. This intuition explains away the property's initial oddness. However, we can also argue for the proposition by changing integration variables: $s = at$. Assume first that $a > 0$; thus,

$$\int_{-\infty}^{\infty} x(t)\delta(at)\,dt = \int_{-\infty}^{\infty} x\left(\frac{s}{a}\right)\delta(s)\frac{1}{a}\,ds = \frac{1}{a}\int_{-\infty}^{\infty} x\left(\frac{s}{a}\right)\delta(s)\,ds = \frac{1}{a}x(0), \tag{3.40}$$

and $\delta(at)$ behaves just like $\frac{1}{|a|}\delta(t)$. If $a < 0$, then $s = at = -|a|t$, and

$$\int_{-\infty}^{\infty} x(t)\delta(at)\,dt = \int_{\infty}^{-\infty} x\left(\frac{s}{a}\right)\delta(s)\frac{-1}{|a|}\,ds = \frac{1}{|a|}\int_{-\infty}^{\infty} x\left(\frac{s}{a}\right)\delta(s)\,ds = \frac{1}{|a|}x(0). \tag{3.41}$$

Whatever the sign of the scaling parameter, (3.39) follows. ■

The Dirac is even: $\delta(-t) = \delta(t)$, as follows from the Dirac scaling proposition. The next property uses this corollary to show how to sift out the signal derivative (Table 3.2).

TABLE 3.2. Dirac Delta Properties

Signal Expression	Property Name
$x(t) = \int_{-\infty}^{\infty} x(s)\delta(t-s)\,ds$	Sifting
$\delta(t) = \frac{d}{dt}[u(t)]$	Derivative of unit step
$\delta(at) = \frac{1}{\lvert a\rvert}\delta(t)$	Scaling
$\delta(t) = \delta(-t)$	Even
$\int_{-\infty}^{\infty} x(t)\frac{d}{dt}\delta(t)\,dt = -\frac{d}{dt}x(t)\Big\rvert_{t=0}$	Derivative of Dirac

Proposition. Let $\delta(t)$ be the Dirac delta and $x(t)$ a signal. Then,

$$\int_{-\infty}^{\infty} x(t)\frac{d}{dt}\delta(t)\,dt = -\frac{d}{dt}(x(t))\bigg|_{t=0}. \tag{3.42}$$

Proof: We differentiate both sides of the sifting property equality to obtain

$$\frac{d}{dt}x(t) = \frac{d}{dt}\int_{-\infty}^{\infty} x(s)\delta(t-s)\,ds = \int_{-\infty}^{\infty} x(s)\frac{d}{dt}\delta(t-s)\,ds = \int_{-\infty}^{\infty} x(s)\frac{d}{dt}\delta(s-t)\,ds, \tag{3.43}$$

and the proposition follows by taking $t = 0$ in (3.43). ■

3.2.5 Splines

Splines are signals formed by interpolating discrete points with polynomials, and they are particularly useful for signal modeling and analysis. Sampling converts an analog signal into a discrete signal, and we can effect the reverse by stretching polynomials between isolated time instants. The term "spline" comes from a jointed wood or plastic tool of the same name used by architects to hand draw elongated, smooth curves for the design of a ship or building. Here we provide a brief overview of the basic splines, or *B-splines*. There are many other types of splines, but what now interests us in B-splines is their definition by repetitive convolution.

Definition (Spline). The analog signal $s(t)$ is a *spline* if there is a set of points $K = \{k_m: m \in \mathbb{Z}\}$ such that $s(t)$ is continuous and equal to a polynomial on each interval $[k_m, k_{m+1}]$. Elements of K are called *knots*. If $s(t)$ is a polynomial of degree n on each $[k_m, k_{m+1}]$, then $s(t)$ has degree n. A spline $s(t)$ of degree n is called *smooth* if it is $n - 1$ times continuously differentiable.

Splines are interesting examples of analog signals, because they are based on discrete samples—their set of knots—and yet they are very well-behaved between the knots. Linear interpolation is probably everyone's first thought to get from a discrete to an analog signal representation. But the sharp corners are problematic. Splines of small order, say $n = 2$ or $n = 3$, are a useful alternative; their smoothness is often precisely what we want to model a natural, continuously occurring process.

Splines accomplish function approximation in applied mathematics [14, 15] and shape generation in computer graphics [16, 17]. They have become very popular signal modeling tools in signal processing and analysis [18, 19] especially in connection with recent developments in wavelet theory [20]. A recent tutorial is Ref. 21.

For signal processing, splines with uniformly spaced knots $K = \{k_m: m \in \mathbb{Z}\}$ are most useful. The knots then represent samples of the spline $s(t)$ on intervals of length $T = k_1 - k_0$. We confine the discussion to smooth splines.

Definition (B-Spline). The zero-order B-spline $\beta_0(t)$ is given by

$$\beta_0(t) = \begin{cases} 1 & \text{if } -\frac{1}{2} < t < \frac{1}{2}, \\ \frac{1}{2} & \text{if } |t| = \frac{1}{2}, \\ 0 & \text{if otherwise.} \end{cases} \tag{3.44}$$

The *B-spline of order n*, $\beta_n(t)$, is

$$\beta_n(t) = \underbrace{\beta_0(t)*\beta_0(t)*\ldots\beta_0(t)}_{n+1 \text{ times}}. \tag{3.45}$$

The B-splines are isolated pulses. The zeroth-order B-spline is a square, $\beta_1(t)$ is a triangle, and higher-order functions are Gaussian-like creatures (Figure 3.2).

Before proving some theorems about splines, let us recall a few ideas from the topology of the real number system $\mathbb{R}$ [6]. Open sets on the real line have soft edges; $S \subset \mathbb{R}$ is *open* if every $s \in S$ is contained in an open interval $I = (a, b)$ that is completely within S: $s \in I \subset S$. Closed sets have hard edges: S is *closed* if its complement is open. Unions of open sets are open, and intersections of closed sets are closed. S is *bounded* if it is contained in some finite interval (a, b). A set of open sets $O = \{O_n\}$ is an *open covering* of S if $\bigcup_n O_n \supset S$. Set S is *compact* if for every open

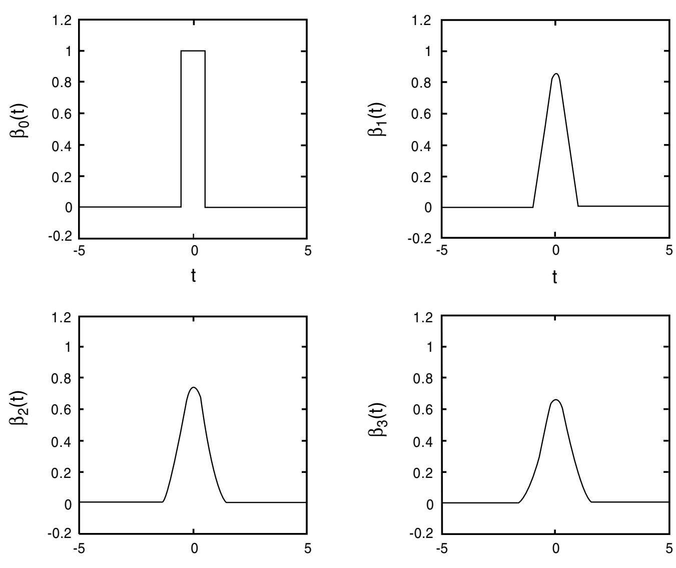

Fig. 3.2. The first four B-splines.

covering O of S there is a subset $P \subset O$ such that P is finite and P covers S. The famous Heine–Borel theorem[2] states that a set is compact if and only if it is closed and bounded [22, 23].

Definition (Support). Let $x(t)$ be an analog signal. Then $x(t)$ is *compactly supported* if $x(t) = 0$ outside some interval $[a, b]$. The *support* of $x(t)$, written Support(x), is the smallest closed set S such that $x(t) = 0$ if $t \notin S$.

Proposition. The n th-order B-spline $\beta_n(t)$ has compact support; that is, the smallest closed set which contains the domain over which set $\beta_n(t)$ is nonzero is closed and bounded.

Proof: By induction on the order (exercise). ■

The next result shows that general splines are linear combinations of shifted B-splines [14].

Theorem (Schoenberg). Let $x(t)$ be a spline having degree n and integral knots $K = \{m = k_m : m \in \mathbb{Z}\}$. Then there are constants c_m such that

$$s(t) = \sum_{m=-\infty}^{\infty} c_m \beta_n(t-m). \tag{3.46}$$

Proof: The proof is outside our present scope; we need Fourier transform tools (Chapters 5 and 6) for the arguments; interested readers will find some of the steps sketched in the latter chapter's problems. ■

The Schoenberg theorem shows that sums of simple shifts and scalings (amplification or attentuation) of the B-spline $\beta_n(t)$ will generate any spline of order n. Thus, the B-splines constitute a kind of signal atom. If signal $x(t)$ admits a spline model, then we can decompose the model into a sum of atoms. The relative abundance of each atom in the original signal at a time value $t = m$ is given by a simple coefficient c_m. This data reduction can be a great help to signal processing and analysis applications, such as filtering and pattern recognition. Notice too that as models of naturally occurring signals, B-splines have the realistic property that they decay to zero as time increases toward $\pm\infty$.

Once we develop more signal theory we shall explore B-spline ideas further. We require a more powerful tool: the full theory of signal frequency, which the Fourier transform provides. Looking forward, we will apply B-splines to construct windowed Fourier transforms (Chapter 10) and scaling functions for wavelet multiresolution analysis structures (Chapter 11).

[2]German analyst Heinrich Eduard Heine (1821–1881) first defined uniform continuity. Emile Borel (1871–1956) is mostly known for his contributions to topology and modern analysis. But he was also Minister of the French Navy (1925–1940) and received the Resistance Medal in 1945 for his efforts opposing the Vichy government during the war.

3.3 ANALOG SIGNAL SPACES

We must once again find sufficiently powerful mathematical structures to support the analog signal processing operations that we have been working with. For instance, convolution characterizes linear, translation-invariant analog systems. But how does one know that the convolution of two signals produces a genuine signal? Since convolution involves integration between infinite limits, there is ample room for doubt about the validity of the result. Are there classes of signals that are broad enough to capture the behavior of all the real-world signals we encounter and yet are closed under the convolution operation? We are thus once again confronted with the task of finding formal mathematical structures that support the operations we need to study systems that process analog signals.

3.3.1 L^p Spaces

Let us begin by defining the p-integrable signal spaces, the analog equivalents of the p-summable spaces studied in the previous chapter. Again, some aforementioned signal classes—such as the absolutely integrable, bounded, and finite-energy signals—fall into one of these families. The theory of signal frequency (Fourier analysis) in Chapters 5 and 7 uses these concepts extensively. These ideas also lie at the heart of recent developments in mixed-domain signal analysis: Gabor transforms, wavelets, and their applications, which comprise the final three chapters of the book.

3.3.1.1 Definition and Basic Properties. We begin with the definition of the p norm for analog signals. The norm is a measure of analog signal size. It looks a lot like the definition of the l^p norm for discrete signals, with an infinite integral replacing the infinite sum. Again, our signals may be real- or complex-valued.

Definition (L^p Norm or p Norm). If $x(t)$ is an analog signal and $p \geq 1$ is finite, then its L^p *norm* is

$$\|x\|_p = \left[\int_{-\infty}^{\infty} |x(t)|^p dt\right]^{\frac{1}{p}}, \tag{3.47}$$

if the Riemann integral exists. If $p = \infty$, then $\|x\|_\infty$ is the least upper bound of $\{|x(t)| \mid x \in \mathbb{R}\}$, if it exists. The L^p norm of $x(t)$ restricted to the interval $[a, b]$ is

$$\|x\|_{L^p[a,\,b]} = \left[\int_a^b |x(t)|^p dt\right]^{\frac{1}{p}}. \tag{3.48}$$

Other notations are $\|x\|_{p,\mathbb{R}}$ and $\|x\|_{p,[a,\,b]}$. The basic properties of the p-norm are the Hölder, Schwarz, and Minkowski inequalities; let us extend them to integrals.

Theorem (Hölder Inequality). Suppose $1 \leq p \leq \infty$ and p and q are conjugate exponents: $p^{-1} + q^{-1} = 1$. If $\|x\|_p < \infty$ and $\|y\|_q < \infty$, then $\|xy\|_1 \leq \|x\|_p \|y\|_q$.

Proof: Recall the technical lemma from Section 2.5.4: $ab \le p^{-1}a^p + q^{-1}b^q$, where p and q are conjugate exponents and $a, b > 0$. So by the same lemma,

$$\frac{|x(t)|}{\|x\|_p}\frac{|y(t)|}{\|y\|_q} \le \frac{|x(t)|^p}{p\|x\|_p^p} + \frac{|y(t)|^q}{q\|y\|_q^q}. \tag{3.49}$$

Integrating (3.49) on both sides of the inequality gives

$$\frac{1}{\|x\|_p\|y\|_q}\int_{-\infty}^{\infty}|x(t)||y(t)|\ dt \le \frac{1}{p\|x\|_p^p}\int_{-\infty}^{\infty}|x(t)|^p dt + \frac{1}{q\|y\|_q^q}\int_{-\infty}^{\infty}|y(t)|^q\ dt = \frac{1}{p}+\frac{1}{q} = 1. \tag{3.50}$$

The case $p = 1$ and $q = \infty$ is straightforward and left as an exercise. ■

Corollary (Schwarz Inequality). If $\|x\|_2 < \infty$ and $\|y\|_2 < \infty$, then $\|xy\|_1 \le \|x\|_2\|y\|_2$.

Proof: Because $p = q = 2$ are conjugate exponents. ■

Theorem (Minkowski Inequality). Let $1 \le p \le \infty$, $\|x\|_p < \infty$, and $\|y\|_p < \infty$. Then $\|x + y\|_p \le \|x\|_p + \|y\|_p$.

Proof: We prove the theorem for $1 < p < \infty$ and leave the remaining cases as exercises. Because $|x(t) + y(t)| \le |x(t)| + |y(t)|$ and $1 \le p$, we have $|x(t) + y(t)|^p \le (|x(t)| + |y(t)|)^p = (|x| + |y|)^{p-1}|x| + (|x| + |y|)^{p-1}|y|$. Integration gives

$$\int_{-\infty}^{\infty}|x(t)+y(t)|^p dt \le \int_{-\infty}^{\infty}(|x|+|y|)^p dt = \int_{-\infty}^{\infty}(|x|+|y|)^{p-1}|x|\ dt + \int_{-\infty}^{\infty}(|x|+|y|)^{p-1}|y|\ dt. \tag{3.51}$$

Let $q = \frac{p}{p-1}$ so that p and q are conjugate exponents. Hölder's inequality then applies to both integrals on the right-hand side of (3.51):

$$\int_{-\infty}^{\infty}|x|(|x|+|y|)^{p-1}dt = \int_{-\infty}^{\infty}|x|(|x|+|y|)^{\frac{p}{q}}dt \le \|x\|_p\left[\int_{-\infty}^{\infty}(|x|+|y|)^p dt\right]^{\frac{1}{q}} \tag{3.52a}$$

and

$$\int_{-\infty}^{\infty}|y|(|x|+|y|)^{p-1}dt = \int_{-\infty}^{\infty}|y|(|x|+|y|)^{\frac{p}{q}}dt \le \|y\|_p\left[\int_{-\infty}^{\infty}(|x|+|y|)^p dt\right]^{\frac{1}{q}}. \tag{3.52b}$$

If the term in square brackets is zero, the theorem is trivially true; we assme otherwise. Putting (3.52a) and (3.52b) together into (3.51) and dividing through by the square-bracketed term gives

$$\frac{\int_{-\infty}^{\infty}(|x|+|y|)^p\,dt}{\left[\int_{-\infty}^{\infty}(|x|+|y|)^p\,dt\right]^{\frac{1}{q}}} = \left[\int_{-\infty}^{\infty}(|x|+|y|)^p\,dt\right]^{p-\frac{1}{q}} \le \|x\|_p + \|y\|_p\,. \tag{3.53}$$

Since $p - \frac{1}{q} = \frac{1}{p}$, the middle term in (3.53) is $\|\,|x|+|y|\,\|_p$. But $\|x+y\|_p \le \|\,|x|+|y|\,\|_p$ and we are done. ■

Now we can define the principal abstract spaces for signal theory. The definition is provisional only, since it relies upon the Riemann integral for the idea that $\|x\|_p < \infty$. We offer two refinements in what follows.

Definition ($L^p(\mathbb{R})$, $L^p[a, b]$). Let $1 \le p \le \infty$. For $p < \infty$, the p-integrable space of analog signals or functions defined on the real numbers is $L^p(\mathbb{R}) = \{x(t) \mid x: \mathbb{R} \to \mathbb{K}$ and $\|x\|_p < \infty\}$, where $\|\cdot\|$ is the L^p norm and $\mathbb{K}$ is either the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$. Also if $a < b$, then $L^p[a, b] = \{x(t) \mid x: [a, b] \to \mathbb{K}$ and $\|x\|_{p,[a,b]} < \infty\}$. If $p = \infty$, then $L^\infty(\mathbb{R})$ and $L^\infty[a, b]$ are the bounded signals on $\mathbb{R}$ and $[a, b]$, respectively.

It is conventional to use uppercase letters for the analog p-integrable spaces and lowercase letters for the discrete p-summable signal spaces. It is also possible to consider half-infinite L^p spaces: $L^p(-\infty, a]$ and $L^p[a, +\infty)$.

These ideas have signal processing significance. The absolutely integrable signals can be used with other L^p signals under the convolution operation [12]. The following proposition tells us that as long as its impulse response $h = H\delta$ is absolutely integrable, an LTI system will produce an L^1 output from an L^1 input.

Proposition. If $x, h \in L^1(\mathbb{R})$, then $y = x * h \in L^1(\mathbb{R})$, and $\|y\|_1 \le \|x\|_1\|h\|_1$.

Proof

$$\int_{-\infty}^{\infty}|y(t)|\,dt = \int_{-\infty}^{\infty}\left|\int_{-\infty}^{\infty}x(s)h(t-s)\,ds\right|dt \le \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|x(s)h(t-s)|\,ds dt. \tag{3.54}$$

From the Fubini–Tonelli theorem [24], if the two-dimensional integrand is either absolutely integrable or non-negative, then the double integral equals either iterated integral. Thus,

$$\|y\|_1 \le \int_{-\infty}^{\infty} |x(s)| \left[\int_{-\infty}^{\infty} |h(t-s)|\ dt \right] ds = \|x\|_1 \|h\|_1 . \tag{3.55}$$

■

The next proposition concerns the concept of uniform continuity, which readers may recall from calculus [6]. A function $y(t)$ is *uniformly continuous* means that for any $e > 0$ there is a $\delta > 0$ such that $|t - s| < \delta$ implies $|y(t) - y(s)| < \varepsilon$. The key idea is that for any $\varepsilon > 0$, it is possible to find a $\delta > 0$ that works for all time values. When the interval width δ must depend on $t \in \mathbb{R}$, then we may have ordinary continuity, but not necessarily uniform continuity. An example of a signal that is uniformly continuous on $\mathbb{R}$ is $\sin(t)$. A signal that is continuous, but not uniformly so, is t^2.

Proposition. If $x \in L^2(\mathbb{R})$ and $y = x \circ x$, then $|y(t)| \le \|x\|_2{}^2$. Furthermore, $y(t)$ is uniformly continuous.

Proof: We apply the Schwarz inequality, $\|fg\|_1 \le \|f\|_2 \|g\|_2$:

$$|y(t)| \le \int_{-\infty}^{\infty} |x(s)| \left|\overline{x(t+s)}\right|\ ds \le \|x(s)\|_2 \|x(t+s)\|_2 = \|x\|_2^2 . \tag{3.56}$$

To show uniform continuity, let us consider the magnitude $|y(t + \Delta t) - y(t)|$:

$$|y(t+\Delta t) - y(t)| \le \int_{-\infty}^{\infty} |x(s)| \left|\overline{x(t+\Delta t+s) - x(t+s)}\right|\ ds. \tag{3.57}$$

Invoking the Schwarz inequality on the right-hand side of (3.57) and changing the integration variable with $\tau = t + s$, we obtain

$$|y(t+\Delta t) - y(t)| \le \|x\|_2 \left[\int_{-\infty}^{\infty} |x(\tau+\Delta t) - x(\tau)|^2 d\tau \right]^{\frac{1}{2}} . \tag{3.58}$$

The limit, $\lim_{\Delta t \to 0}\ y(t + \Delta t) - y(t)$, concerns us. From integration theory—for instance, Ref. 24, p. 91—we know that

$$\lim_{\tau \to 0} \int_{-\infty}^{\infty} |x(\tau + \Delta t) - x(\tau)|\ d\tau = 0, \tag{3.59}$$

and since this limit does not depend upon t, $y(t)$ is uniform continuous. ■

Proposition. Let $1 \le p \le \infty$, $x \in L^p(\mathbb{R})$, and $h \in L^1(\mathbb{R})$. Then $y = x * h \in L^p(\mathbb{R})$, and $\|y\|_p \le \|x\|_p \|h\|_1$.

Proof: Exercise. ■

The criterion upon which the definition rides depends completely on whether $|x(t)|^p$ is integrable or not; that is, it depends how powerful an integral we can roust up. The Riemann integral, of college calculus renown, is good for functions that are piecewise continuous. Basic texts usually assume continuity of the integrand, but their theory generalizes easily to those functions having a finite number of discontinuities; it is only necessary to count the pieces and perform separate Riemann integrals on each segment. A refinement of the definition is possible, still based on Riemann integration. We make this refinement after we discover how to construct Banach spaces out of the L^p-normed linear spaces given by the first defintion of $\|x\|_p$. We issue the following warning: The Riemann integral will fail to serve our signal theoretic needs. We will see this as soon as we delve into the basic abstract signal structures: normed linear, Banach, inner product, and Hilbert spaces. The modern Lebesgue integral replaces it. Our final definition, which accords with modern practice, will provide the same finite integral criterion, but make use of the modern Lebesgue integral instead. This means that although the above definition will not change in form, when we interpret it in the light of Lebesgue's rather than Riemann's integral, the L^p spaces will admit a far wider class of signals.

3.3.1.2 Normed Linear Spaces.

A normed linear space allows basic signal operations such as summation and scalar multiplication (amplification or attenuation) and in addition provides a measure of signal size, the norm operator, written $\|\cdot\|$. Normed spaces can be made up of abstract elements, but generally we consider those that are sets of analog signals.

Definition (Norm, Normed Linear Space). Let X be a vector space of analog signals over $\mathbb{K}$ ($\mathbb{R}$ or $\mathbb{C}$). Then a *norm*, written $\|\cdot\|$, is a map $\|\cdot\|: X \to \mathbb{R}$ such that

(i) (Non-negative) $0 \le \|x\|$ for all $x \in X$.
(ii) (Zero) $\|x\| = 0$ if and only if $x(t) = 0$ for all t.
(iii) (Scalar multiplication) $\|ax\| = |c|\,\|x\|$ for every scalar $c \in \mathbb{K}$.
(iv) (Triangle inequality) $\|x + y\| \le \|x\| + \|y\|$.

If X is a vector space of analog signals and $\|\cdot\|$ is a norm on X, then $(X, \|\cdot\|)$ is a *normed linear space*. Other common terms are normed vector space or simply normed space.

One can show that the norm is a continuous map. That is, for any $x \in X$ and $\varepsilon > 0$ there is a $\delta > 0$ such that for all $y \in X$, $\|y - x\| < \delta$ implies $|\,\|y\| - \|x\|\,| < \varepsilon$. The algebraic operations, addition and scalar multiplication, are also continuous.

Example. Let $a < b$ and consider the set of continuous functions $x(t)$ on $[a, b]$. This space, denoted $C^0[a, b]$ is a normed linear space with the following norm: $\|x(t)\| = \sup\{|x(t)|: t \in [a, b]\}$. Since the closed interval $[a, b]$ is compact, it is closed and bounded (by the Heine–Borel theorem), and a continuous function therefore achieves a maximum [6]. Thus, the norm is well-defined.

Example. Let $C^0(\mathbb{R})$ the set of bounded continuous analog signals. This too is a normed linear space, given the supremum norm: $\|x(t)\| = \sup\{|x(t)|: t \in \mathbb{R}\}$.

Example. Different norms can be given for the same underlying set of signals, and this results in different normed vector spaces. For example, we can choose the energy of a continuous signal $x(t) \in C^0[a, b]$ as a norm:

$$\|x\| = E_x = \left[\int_a^b |x(t)|^2 dt\right]^{\frac{1}{2}}. \tag{3.60}$$

The next proposition ensures that the L^p-norm is indeed a norm for continuous signals.

Proposition. Let X be the set of continuous, p-integrable signals $x : \mathbb{R} \to \mathbb{K}$, where $\mathbb{K}$ is either $\mathbb{R}$ or $\mathbb{C}$. Then $\|x\|_p$ is a norm, and $(X, \|x\|_p)$ is a normed linear space.

Proof: The continuous signals are clearly an Abelian (commutative) group under addition, and obey the scalar multiplication rules for a vector space. Norm properties (i) and (iii) follow from basic integration theory. For (ii), note that if $x(t)$ is not identically zero, then there must be some t_0 such that $x(t_0) = \varepsilon \neq 0$. By continuity, in an interval $I = (a, b)$ about t_0, we must have $|x(t)| > \varepsilon/2$. But then the norm integral is at least $[\varepsilon(\mathrm{b} - a)/2]^p > 0$. The last property follows from Minkowski's inequality for analog signals. ■

Proposition. Let X be the set of continuous, p-integrable signals x: $[a, b] \to \mathbb{K}$, where $a < b$. Then $(X, \|x\|_{p,[a, b]})$ is a normed linear space.

Proof: Exercise. ■

Must analog signal theory confine itself to continuous signals? Some important analog signals contain discontinuities, such as the unit step and square pulse signals, and we should have enough confidence in our theory to apply it to signals with an infinite number of discontinuities. Describing signal noise, for example, might demand just as much. The continuity assumption enforces the zero property of the norm, (ii) above; without presupposing continuity, signals that are zero except on a finite number of points, for example, violate (ii). The full spaces, $L^p(\mathbb{R})$ and $L^p[a,b]$, are not—from discussion so far—normed spaces.

A *metric space* derives naturally from a normed linear space. Recall from the Chapter 2 exercises that a *metric* $d(x, y)$ is a map from pairs of signals to real numbers. Its four properties are:

(i) $d(x, y) \geq 0$ for all x, y.
(ii) $d(x, y) = 0$ if and only if $x = y$.

(iii) $d(u, v) = d(v, u)$ for all u, v.
(iv) For any s, $d(x, y) \le d(x, s) + d(s, y)$.

Thus, the L^p-norm generates a metric. In signal analysis matching applications, a common application requirement is to develop a measure of match between candidate signals and prototype signals. The candidate signals are fed into the analysis system, and the prototypes are models or library elements which are expected among the inputs. The goal is to match candidates against prototypes, and it is typical to require that the match measure be a metric.

Mappings between normed spaces are also important in signal theory. Such maps abstractly model the idea of filtering a signal: signal-in and signal-out. When the normed spaces contain analog signals, or rather functions on the real number line, then such maps are precisely the analog systems covered earlier in the chapter. For applications we are often interested in linear maps.

Definition (Linear Operator). Let X and W be normed spaces over $\mathbb{K}$ ($\mathbb{R}$ or $\mathbb{C}$) and T: $X \to W$ such that for all $x, y \in X$ and any $a \in \mathbb{K}$ we have

(i) $T(x + y) = T(x) + T(y)$.
(ii) $T(ax) = aT(x)$.

Then T is called a *linear operator* or *linear map*. Dropping parentheses is widespread: $Tx \equiv T(x)$. If the operator's range is included in $\mathbb{R}$ or $\mathbb{C}$, then we more specifically call T a *linear functional*.

Proposition (Properties). Let X and W be normed spaces over $\mathbb{K}$ and let T: $X \to W$ be a linear operator. Then

(i) Range(T) is a normed linear subspace of W.
(ii) The null space of T, $\{x \in X \mid Tx = 0\}$ is a normed linear subspace of X.
(iii) The inverse map T^{-1}: Range(T) $\to X$ exists if and only if the null space of T is precisely $\{0\}$ in X.

Proof: Exercise. ■

Definition (Continuous Operator). Let X and W be normed spaces over $\mathbb{K}$ and T: $X \to W$. Then T is *continuous at* x if for any $\varepsilon > 0$ there is a $\delta > 0$ such that if $\|y - x\| < \delta$, then $\|Ty - Tx\| < \varepsilon$. T is *continuous* if it is continuous at every $x \in X$.

Definition (Norm, Bounded Linear Operator). Let X and W be normed spaces over $\mathbb{K}$ and T: $X \to W$ be a linear operator. Then we define the *norm* of T, written $\|T\|$, by $\|T\| = \sup\{\|Tx\|/\|x\|: x \in X, x \neq 0\}$. If $\|T\| < \infty$, then T is a *bounded linear operator*.

Theorem (Boundedness). Let X and W be normed spaces over $\mathbb{K}$ and T: $X \to W$ be a linear operator. Then the following are equivalent:

(i) T is bounded.
(ii) T is continuous.
(iii) T is continuous at $0 \in X$.

Proof: The easy part of this proof is not hard to spot: (ii) obviously implies (iii). Let us therefore assume continuity at zero (iii) and show that T is bounded. Let $\delta > 0$ such that $\|Tx - 0\| = \|Tx\| < 1$ when $\|x - 0\| = \|x\| < \delta$. Let $y \in X$ be nonzero. Then $\left\|\frac{\delta y}{2\|y\|}\right\| = \frac{\delta}{2} < \delta$, so that $\left\|\frac{T(\delta y)}{2\|y\|}\right\| < 1$ and $\|Ty\| < \frac{2\|y\|}{\delta}$; T is bounded. Now we assume T is bounded (i) and show continuity. Note that $\|Tx - Ty\| = \|T(x - y)\| \le \|T\| \, \|x - y\|$, from which continuity follows. ■

The boundedness theorem seems strange at first glance. But what $\|T\| < \infty$ really says is that T amplifies signals by a limited amount. So there cannot be any sudden jumps in the range when there are only slight changes in the domain. Still, it might seem that a system could be continuous without being bounded, since it could allow no jumps but still amplify signals by arbitrarily large factors. The linearity assumption on T prevents this, however.

3.3.1.3 Banach Spaces. Analog Banach spaces are normed linear spaces for which every Cauchy sequence of signals converges to a limit signal also in the space. Again, after formal developments in the previous chapter, the main task here is to investigate how analog spaces using the L^p norm can be complete. Using familiar Riemann integration, we can solve this problem with an abstract mathematical technique: forming the completion of a given normed linear space. But this solution is unsatisfactory because it leaves us with an abstract Banach space whose elements are quite different in nature from the simple analog signals with which we began. Interested readers will find that the ultimate solution is to replace the Riemann with the modern Lebesgue integral.

Recall that a sequence of signals $\{x_n(t): n \in \mathbb{Z}\}$ is Cauchy when for all $\varepsilon > 0$ there is an N such that if $m, n > N$, then $\|x_m - x_n\| < \varepsilon$. Note that the definition depends on the choice of norm on the space X. That is, the signals get arbitrarily close to one another; as the sequence continues, signal perturbations become less and less significant—at least as far as the norm can measure. A signal $x(t)$ is the limit of a sequence $\{x_n(t)\}$ means that for any $\varepsilon > 0$ there is an $N > 0$ such that $n > N$ implies $\|x_n - x\| < \varepsilon$. A normed space X is complete if for any Cauchy sequence $\{x_n(t) : n \in \mathbb{N}\}$ there is an $x(t) \in X$ such that $x(t)$ is the limit of $\{x_n(t)\}$. A complete normed space is also called a Banach space.

In the previous section, we considered the continuous analog signals on the real line, or on an interval, and showed that with the L^p norm, they constituted normed linear spaces. Are they also Banach spaces? The answer is no, unfortunately;

Cauchy sequences of continuous signals may converge to signals that are not continuous, as the counterexample below illustrates.

Example. If $p < \infty$, then the signal space $(C^0[-1, 1], \|\cdot\|_{p,[-1,1]})$, consisting of all continuous signals on $[-1, 1]$ with the L^p norm, is not complete. The claim is secure if we can exhibit a Cauchy sequence of continuous functions that converges to a discontinuous function. A sequence that approximates the unit step on $[-1, 1]$ is

$$x_n(t) = \begin{cases} 0 & \text{if } t < -\dfrac{1}{n}, \\ \dfrac{tn}{2} + \dfrac{1}{2} & \text{if } -\dfrac{1}{n} \le t \le \dfrac{1}{n}, \\ 1 & \text{if } t > \dfrac{1}{n}. \end{cases} \tag{3.61}$$

The $\{x_n(t)\}$, shown in Figure 3.3, clearly converge pointwise. Indeed, for any $t_0 < 0$, $x_n(t_0) \to 0$; for any $t_0 > 0$, $x_n(t_0) \to 1$; and for $t_0 = 0$, $x_n(t_0) = 1/2$ for all n. Now, if we assume that $n < m$, then

$$\int_{-1}^{1} |x_m(t) - x_n(t)|^p dt = \int_{-1/n}^{1/n} |x_m(t) - x_n(t)|^p dt \le \frac{2}{n}\left(\frac{1}{2^p}\right). \tag{3.62}$$

The sequence is Cauchy, but converges to a discontinuous signal. The same reasoning applies to $L^p[a, b]$, where $a < b$, and to $L^p(\mathbb{R})$.

Example. Now consider $C^0[a, b]$ the set of bounded continuous analog signals on $[a, b]$ with the supremum or L^∞ norm: $\|x(t)\|_\infty = \sup\{|x(t)| : t \in [a, b]\}$. This space's norm avoids integration, so $(C^0[a, b], \|\cdot\|_\infty)$ earns Banach space status. To see this, note that if $\{x_n(t)\}$ is Cauchy and $\varepsilon > 0$, then there is an $N > 0$ such that for all $m, n > N$ we have $\|x_n - x_m\|_\infty < \varepsilon$. Fixing $t_0 \in [a, b]$, the sequence of real numbers $\{x_n(t_0)\}$ is Cauchy in $\mathbb{R}$. Calculus teaches that Cauchy sequences of real numbers converge to a limit in $\mathbb{R}$; for each $t \in [a, b]$, we may therefore set $x(t) = \lim_{n \to \infty} x_n(t)$. We claim $x(t)$ is continuous. Indeed, since the sequence $\{x_n(t)\}$ is Cauchy in the L^∞ norm, the sequence must converge not just pointwise, but uniformly to $x(t)$. That is, for any $\varepsilon > 0$, there is an $N > 0$ such that $m, n > N$ implies $|x_m(t) - x_n(t)| < \varepsilon$ for all $t \in [a, b]$. Uniformly convergent sequences of continuous functions converge to a continuous limit [6] $x(t)$ must therefore be continuous, and $C^0[a, b]$ is a Banach space.

Analog signal spaces seem to leave us in a quandary. We need continuity in order to achieve the basic properties of normed linear spaces, which provide a basic signal size function, namely the norm. Prodded by our intuition that worldly processes—at least at our own perceptual level of objects, forces between them, and their motions—are continuously defined, we might proceed to develop analog signal theory from

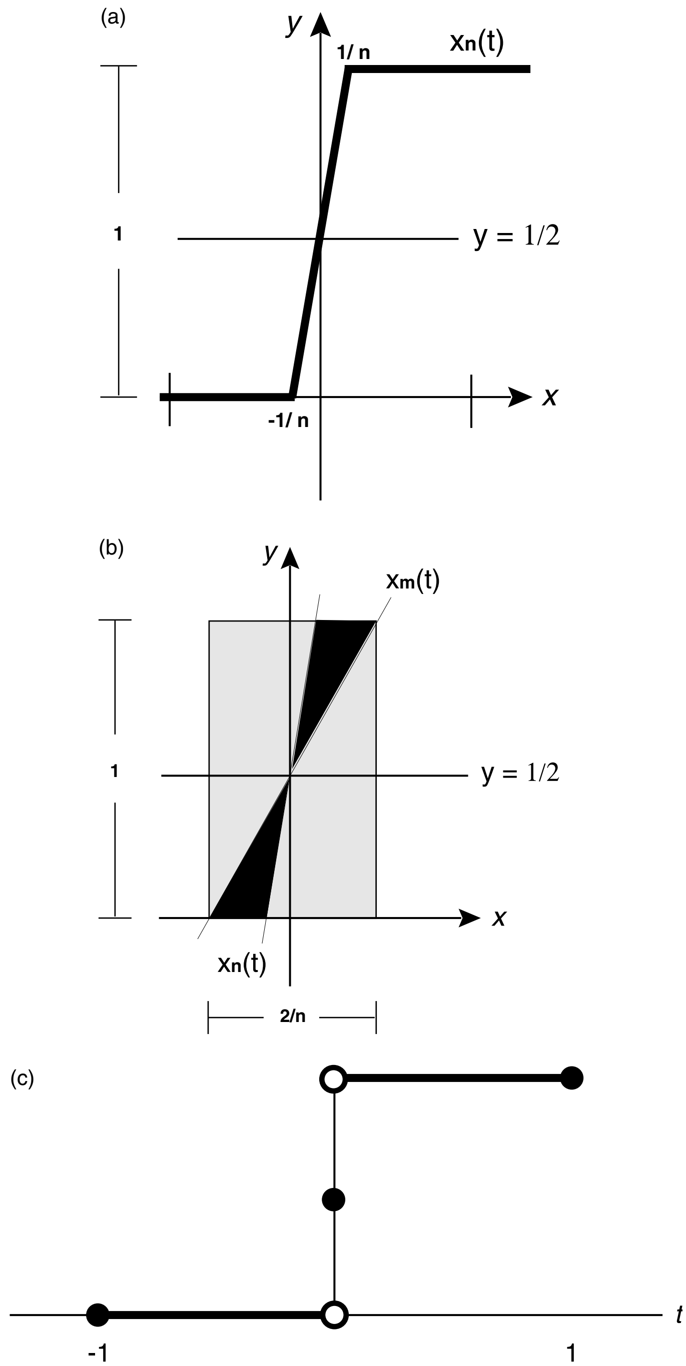

Fig. 3.3. (a) A Cauchy sequence of continuous signals in $L^p[-1, 1]$. (b) Detail of the difference between x_m and x_n. Assuming $n < m$, the signals differ by at most 1/2 within the rectangular region, which has width $2/n$. (c) Diagram showing discontinuous limit.

continuous signals alone. However, arbitrarily precise signal approximations depend on the mathematical theory of limits, which in turn begs the question of the completeness for our designated theoretical base, $(C^0[-1, 1], \|\cdot\|_p)$ for example. But the first above example shows that we cannot get completeness from families of continuous signals with the L^p-norm, where $p < \infty$.

What about working with the supremum norm, $\|\cdot\|_\infty$? This is a perfectly good norm for certain signal analysis applications. Some matching applications rely on it, for example. Nonetheless, the next section shows that L^∞ does not support an inner product. Inner products are crucial for much of our later development: Fourier analysis, windowed Fourier (Gabor) transforms, and wavelet transforms. Also many applications presuppose square-integrable physical quantities. It would appear that the supremum norm would confine signal theory to a narrow range of processing and analysis problems.

Let us persist: Can we only allow sequences that do converge? As the exercises explain, uniformly convergent sequences of continuous signals converge to a continuous limit. The problem is when the signal values, say $x_n(t_0)$, converges for every $t_0 \in [-1, 1]$, then we should expect that x_n itself converges to an allowable member of our signal space. Alternatively, since Cauchy sequences of continuous signals lead us to discontinuous entities, can we just incorporate into our foundational signal space the piecewise continuous signals that are p-integrable? We would allow signals with a finite number of discontinuities. The Riemann integral extends to them, and it serves the L^p-norm definition (3.47). We would have to give up one of the criteria for a normed linear space; signals would differ, but the norm of their difference would be zero. Frustratingly, this scheme fails, too, as the next example shows [25].

Example. Consider the signals $\{x_n(t)\}$ defined on $[-1, 1]$ defined by

$$x_n(t) = \lim_{m \to \infty} [\cos(n!\pi t)]^{2m} = \begin{cases} 1 & \text{if } n!t \in \mathbb{Z}, \\ 0 & \text{if otherwise.} \end{cases} \tag{3.63}$$

Note that $x_n(t)$ is zero at all but a finite number of points in $[-1, 1]$; $x_n(t)$ is Riemann-integrable, and its Riemann integral is zero on $[-1, 1]$. Also $\|x_n(t) - x_m(t)\|_{p,[-1,1]} = 0$, which means that the sequence $\{x_n(t)\}$ is Cauchy. It converges, however, to

$$x(t) = \begin{cases} 1 & \text{if } t \in \mathbb{Q}, \\ 0 & \text{if otherwise}. \end{cases} \tag{3.64}$$

which is not Riemann-integrable.

3.3.1.4 Constructing Banach Spaces. An abstract mathematical technique, called completing the normed space, eliminates most of the above aggravations. The completion of a normed linear space X, is a Banach space B having a subspace C isometric to X. Two normed spaces, M and N, are isometric means there is a one-to-one, onto map f: $M \rightarrow N$ such that f is a vector space isomorphism and for all $x \in M$, $\|x\|_M = \|f(x)\|_N$. Isometries preserve norms. The next theorem shows how to construct the completion of a general normed linear space.

Let us cover an algebraic idea which features in the completion result: equivalence relations [26].

Definition (Equivalence Relation). We say $a \sim b$ is an equivalence relation on a set S if it satisfies the following three properties:

(i) (Reflexive) For all $a \in S$, $a \sim a$.
(ii) (Symmetric) For all $a, b \in S$, if $a \sim b$, then $b \sim a$.
(iii) (Transitive) For all $a, b, c \in S$, if $a \sim b$ and $b \sim c$, then $a \sim c$.

Equivalence relations associate things in a collection that are similar in form.

Example. For example, we might consider ordered pairs of integers (m, n) with $n \neq 0$. If $p = (m, n)$ and $q = (i, k)$, then it is easy to check that the relation, $p \sim q$ if and only if $m \cdot k = n \cdot i$, is an equivalence relation. In fact, this is the equivalence relationfor different forms of the same rational number: $\frac{7}{8} = \frac{42}{48}$, for instance. Rational numbers themselves are not pairs of integers, but are represented by an infinite collection of such pairs. This is an important idea.

Definition (Equivalence Class). Let $a \sim b$ be an equivalence relation on a set S. The equivalence class of $a \in S$ is $[a] = \{b \in S \mid a \sim b\}$.

If $a \sim b$ is an equivalence relation on a set S, then the equivalence classes form a partition of S. Every $a \in S$ belongs to one and only one equivalence class.

Example (Rational Numbers). From the previous example we can let $S = \{(m, n) \mid m, n \in \mathbb{Z} \text{ and } n \neq 0\}$. Let $\mathbb{Q} = \{[q] \mid q \in S\}$. We can define addition and multiplication on elements of $\mathbb{Q}$. This is easy. If $p = (m, n)$ and $q = (i, k)$, then we define $[p] + [q] = [(M, N)]$, where $\frac{M}{N} = \frac{m}{n} + \frac{i}{k}$. Constructing a rational multiplication operator is simple, too. These steps construct the rational numbers from the integers [26].

But for signal theory, so what? Well, equivalence relations are precisely what we need to deal with the problem that arises when making a normed linear space based on the L^p norm. Just a moment ago, we noted that the zero property of the norm—namely, that $\|x\|_p = 0$ if and only if $x(t) = 0$ identically—compelled us to use continuous signals for normed linear spaces. But we use and need piecewise continuous entities, such as the unit step, in signal processing. Also, signals with point discontinuities are also useful for modeling noise spikes and the like. The strategy is to begin with piecewise discontinuous signals, but to assemble them into equivalence classes for making the normed linear space.

Example (Signal Equivalence Classes). Let $1 \le p < \infty$, and suppose that we have extended the Riemann integral to piecewise continuous analog signals (that is, having at most a finite number of discontinuities) on the interval $[a, b]$. Strictly speaking, this class of signals is not a normed linear space using $\|\cdot\|_p$; the zero property of the norm fails. However, we let $[x] = \{y(t): \|y\|_p = \|x\|_p\}$ and define $\underline{L}^p[a, b] = \{[x] \mid x \in L^p[a, b]\}$. Evidently, we have identified all signals that differ from one another by a finite number of discontinuities as being essentially the same. That is what lumping them into equivalence classes accomplishes. Now we can define signal addition, scalar multiplication, and the norm on equivalence classes. This reinterpretation of $L^p[a, b]$ in terms of equivalence classes is often implicit in many signal processing treatments. The same idea applies to $L^p(\mathbb{R})$. The exercises explore these ideas in greater depth.

Now let us turn to the problem of constructing a Banach space from a given normed linear space [22, 27]. This leads to a refined definition for the L^p spaces.

Theorem (Completion of Normed Linear Space). Let X be a normed linear space. Then there is a Banach space B and an isometry $f: X \to B$ such that $C = \text{Range}(f)$ is dense in B.

Proof: Let $S = \{x_n\}$ and $T = \{y_n\}$ be Cauchy sequences in X. We define the relation $S \sim T$ if and only if $\lim_{n\to\infty}\|x_n - y_n\| = 0$. It is not hard to show this is an equivalence relation (exercise). Let $[S] = \{T: T \sim S\}$ and set $B = \{[S]: \ S = \{x_n\}$ is Cauchy in $X\}$. Let $x \in X$ and define $f(x) = [\{x_n\}]$, where $x_n = x$ for all n; the image $f(x)$ is a constant sequence. There are several things we must show.

With appropriate definitions of addition, scalar multiplication, and norm, we can make B into a normed linear space. If $S = \{x_n\}$, $T = \{y_n\} \in B$, then we define an additon operation on B by $[S] + [T] = [\{x_n + y_n\}]$. This works, but there is a slight technical problem. Many different Cauchy sequences $\{a_n\}$ can be the source for a single equivalence class, say $[S]$. We must show that the definition of additon does not depend on which sequences in the respective equivalence classes, S and T, are taken for defining the sums in $\{x_n + y_n\}$. So suppose $S = [\{a_n\}]$ and $T = [\{b_n\}]$ so that $\{a_n\} \sim \{x_n\}$ and $\{b_n\} \sim \{y_n\}$. We claim that $[\{x_n + y_n\}] = [\{a_n + b_n\}]$; that is, our addition operation is well-defined. Because $\|(x_n + y_n) - (a_n + b_n)\| = \|(x_n - a_n) + (y_n - b_n)\| \le \|x_n - a_n\| + \|y_n - b_n\|$, and both of these last terms approach zero, we must have $\{x_n + y_n\} \sim \{a_n + b_n\}$, proving the claim. We define scalar multiplication by $c[S] = [\{cx_n\}]$. It is straightforward to show that these definitions make B into a vector space. For the norm, we define $\|[S]\| = \lim_{n\to\infty}\|x_n\|$. Justifying the definition requires that the limit exists and that the definition is independent of the sequence chosen from the equivalence class $[S]$. We have

$$\left|\|x_n\| - \|x_m\|\right| \le \|x_n - x_m\| \tag{3.65}$$

by the triangle inequality. Since $\{x_n\}$ is Cauchy in X, so must $\{\|x_n\|\}$ be in $\mathbb{R}$. Next, suppose that some other sequence $\{a_n\}$ generates the same equivalence class: $[\{a_n\}] = [S]$. We need to show that $\lim_{n\to\infty}\|x_n\| = \lim_{n\to\infty}\|a_n\|$. In fact, we know that $[\{a_n\}] \sim [\{x_n\}]$, since they are in the same equivalence class. Thus, $\lim_{n\to\infty}\|x_n - a_n\| = 0$. Since $\|x_n - a_n\| \geq |\|x_n\| - \|a_a\||$, $\lim_{n\to\infty}[\|x_n\| - \|a_n\|] = 0$, and we have shown our second point necessary for a well-defined norm. Verifying the normed space properties remains, but it is straightforward and perhaps tedious.

Notice that the map f is a normed space isomorphism that preserves norms—an *isometry*. An *isomorphism* is one-to-one and onto, $f(x + y) = f(x) + f(y)$, and $f(cx) = cf(x)$ for scalars $c \in \mathbb{R}$ (or $\mathbb{C}$). In an isometry we also have $\|x\| = \|f(x)\|$.

Our major claim is that B is complete, but a convenient shortcut is to first show that Range(f) is dense in B. Given $[T] = [\{y_n\}] \in B$, we seek an $x \in X$ such that $f(x)$ is arbitrarily close to $[T]$. Since $\{y_n\}$ is Cauchy, for any $\varepsilon > 0$, there is an N_ε such that if $m, n > N_\varepsilon$, then $\|y_m - y_n\| < \varepsilon$. Let $k > N_\varepsilon$, and set $x = x_n = y_k$ for all $n \in \mathbb{N}$. Then $f(x) = [\{x_n\}]$, and $\|[\{y_n\}] - [\{x_n\}]\| = \lim_{n\to\infty}\|y_n - x_n\| = \lim_{n\to\infty}\|y_n - y_k\| \leq \varepsilon$. Since ε is arbitrary, $f(X)$ must be dense in B.

Finally, to show that B is complete, let $\{S_n\}$ be a Cauchy sequence in B; we have to find an $S \in B$ such that $\lim_{n\to\infty}S_n = S$. Since Range(f) is dense in B, there must exist $x_n \in X$ such that $\|f(x_n) - S_n\| < 1/(n + 1)$ for all $n \in \mathbb{N}$. We claim that $\{x_n\}$ is Cauchy in X, and if we set $S = [\{x_n\}]$, then $S = \lim_{n\to\infty}S_n$. Since f is an isometry,

$$\begin{aligned} \|x_n - x_m\| &= \|f(x_n) - f(x_m)\| = \|(f(x_n) - S_n) + (S_m - f(x_m)) + (S_n - S_m)\| \\ &\leq \|f(x_n) - S_n\| + \|S_m - f(x_m)\| + \|S_n - S_m\|. \end{aligned} \tag{3.66}$$

By the choice of $\{x_n\}$, the first two terms on the bottom of (3.66) are small for sufficiently large m, n. The final term in (3.66) is small too, since $\{S_n\}$ is Cauchy. Consequently, $\{x_n\}$ is Cauchy, and $S = [\{x_n\}]$ must be a bona fide element of B. Furthermore, note that $\|S_k - S\| \leq \|f(x_k) - S_k\| + \|f(x_k) - S\|$. Again, $\|f(x_k) - S_k\| < 1/(k + 1)$; we must attend to the second term. Let $y_n = x_k$ for all $n \in N$, so that $f(x_k) = [\{y_n\}]$. Then $\|f(x_k) - S\| = \|[\{y_n\}] - [\{x_n\}]\| = \lim_{n\to\infty}\|y_n - x_n\| = \lim_{n\to\infty}\|x_k - x_n\|$. But $\{x_n\}$ is Cauchy, so this last expression is also small for large n. ∎

Corollary (Uniqueness). Let X be a normed linear space and suppose B is a Banach space with a dense subset C isometric to X. Then B is isometric to the completion of X.

Proof: Let $f: X \to C$ be the isometry and suppose $\overline{X}$ is the completion of X. Any element of B is a limit of a Cauchy sequence of elements in C: $b = \lim_{n\to\infty}c_n$. We can extend f to a map from $\overline{X}$ to B by $f(x) = b$, where $x = \lim_{n\to\infty} f^{-1}(c_n)$. We trust the further demonstration that this is an isometry to the reader (exercise). ∎

The corollary justifies our referring to *the* completion of a normed linear space. Now we can refine the definition of the p-integrable signal spaces.

Definition ($L^p(\mathbb{R})$, $L^p[a, b]$). For $1 \le p < \infty$, $L^p(\mathbb{R})$ and $L^p[a, b]$ are the completions of the normed linear spaces consisting of the continuous, Riemann p-integrable analog signals on $\mathbb{R}$ and $[a, b]$, respectively.

So the completion theorem builds up Banach spaces from normed linear spaces having only a limited complement of signals—consisting, for instance, of just continuous signals. The problem with completion is that it provides no clear picture of what the elements in completed normed space look like. We do need to expand our realm of allowable signals because limit operations lead us beyond functions that are piecewise continuous. We also seek constructive and elemental descriptions of such functions and hope to avoid invoking abstract, indirect operations such as with the completion theorem. Can we accomplish so much and still preserve closure under limit operations?

The Lebesgue integral is the key concept. Modern integration theory removes almost every burden the Riemann integral imposes, but some readers may prefer to skip the purely mathematical development; the rest of the text is quite accessible without Lebesgue integration. So we postpone the formalities and turn instead to inner products, Hilbert spaces, and ideas on orthonormality and basis expansions that we need for our later analysis of signal frequency.

3.3.2 Inner Product and Hilbert Spaces

Inner product and Hilbert spaces provide many of the theoretical underpinnings for time domain signal pattern recognition applicatins and for the whole theory of signal frequency, or Fourier analysis.

3.3.2.1 Inner Product Spaces. An inner product space X is a vector space equipped with an inner product relation $\langle x, y\rangle$. The operation $\langle \cdot, \cdot\rangle$ takes pairs of elements in X and maps them to the real numbers or, more generally, the complex numbers. The algebraic content of Chapter 2's development is still valid; again, all we need to do is define the inner product for analog signals and verify that the properties of an abstract inner product space remain true.

Definition (Inner Product). Let $x(t)$ and $y(t)$ be real- or complex-valued analog signals. Then their inner product is $\langle x, y\rangle$:

$$\langle x, y\rangle = \int_{-\infty}^{\infty} x(t)\overline{y(t)}\, dt. \tag{3.67}$$

The inner product induces a norm, $\|x\| = (\langle x, x\rangle)^{1/2}$. So any inner product space thereby becomes a normed linear space. Readers mindful of Chapter 2's theorems will rightly suspect that the converse does not hold for analog signals. Recall that the inner product norm obeys the *parallelogram law*,

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2), \tag{3.68}$$

and the *polarization identity*,

$$4\langle x, y\rangle = \|x+y\|^2 - \|x-y\|^2 + j\|x+jy\|^2 - j\|x-jy\|^2. \tag{3.69}$$

It is not hard to show that the definition (3.67) satisfies the properties of an inner product. The main difficulty—as in the discrete world—is to find out for which abstract signal spaces the integral (3.67) actually exists. As with our earlier discrete results, we find that the space L^2 is special.

Example (Square-Integrable Signals). The spaces $L^2[a, b]$ and $L^2(\mathbb{R})$ are inner product spaces. Let $x(t)$ and $y(t)$ be real- or complex-valued analog signals. By the Schwarz inequality we know that if signals x and y are square-integrable, that is $\|x\|_2 < \infty$ and $\|y\|_2 < \infty$, then $\|xy\|_1 \le \|x\|_2\|y\|_2$. We must show that their inner product integral (3.67) exists. But,

$$|\langle x, y\rangle| = \left|\int_{-\infty}^{\infty} x(t)\overline{y(t)}\,dt\right| \le \int_{-\infty}^{\infty} \left|x(t)\overline{y(t)}\right|\,dt = \int_{-\infty}^{\infty} |x(t)||y(t)|\,dt = \|xy\|_1. \tag{3.70}$$

Schwarz's inequality shows the integration works. It states that if $\|x\|_2 < \infty$ and $\|y\|_2 < \infty$, then $\|xy\|_1 \le \|x\|_2\|y\|_2$. But (3.70) shows that $|\langle x, y\rangle| \le \|xy\|_1$. Requisite properties of an inner product are:

(i) $0 \le \langle x, x\rangle$ and $\langle x, x\rangle = 0$ if and only if $x(t) = 0$ for all t.
(ii) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$.
(iii) $\langle cx, y\rangle = c\langle x, y\rangle$, for any scalar c.
(iv) $\langle x, y\rangle = \overline{\langle y, x\rangle}$.

Their verification from (3.67) follows from the basic properties of Riemann integration (exercise).

Example (L^p spaces, $p \ne 2$). The spaces $L^p[a, b]$ and $L^p(\mathbb{R})$ are *not* inner product spaces with $\langle\cdot,\cdot\rangle$ defined in (3.67). Let $x(t)$ and $y(t)$ be the signals shown in Figure 3.4. Observe that $\|x\|_p = \|y\|_p = 2^{1/p}$, but $\|x + y\|_p = \|x - y\|_p = 2$. The parallelogram law holds in an inner product space, which for these signals implies $2 = 2^{2/p}$. This is only possible if $p = 2$.

3.3.2.2 Hilbert Spaces. An inner product space that is complete with respect to its induced norm is called a *Hilbert space*. All of the L^p signal spaces are Banach spaces, but only L^2 is an inner product space. So our generalization of linear algebra to encompass vectors that are infinitely long in both directions—that is, *signals*—succeeds but at the cost of eliminating all but an apparently narrow class of signals.

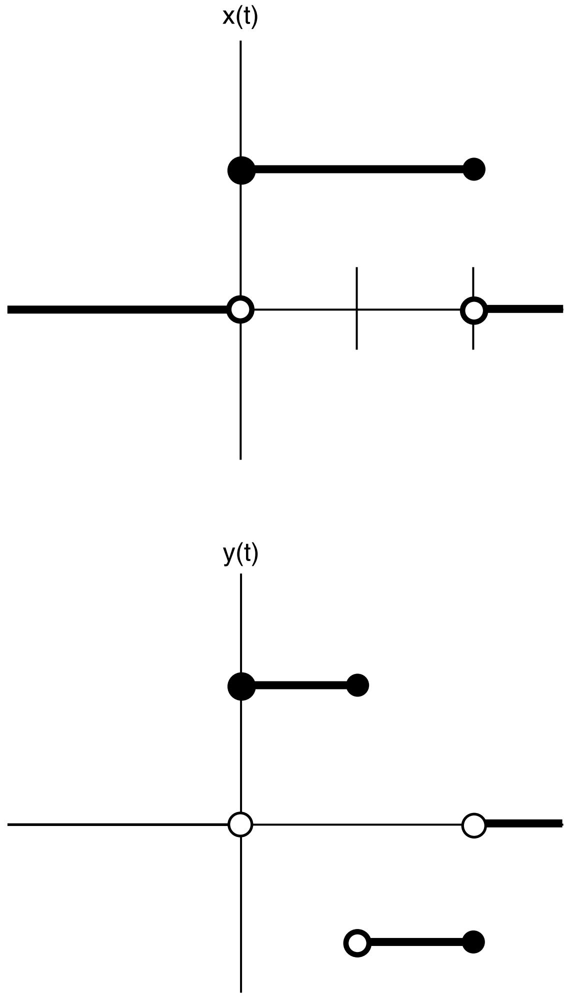

Fig. 3.4. Signals in the L^p Banach spaces: a square pulse $x(t)$ and a step signal $y(t)$.

Though it is true that square-integrable signals shall be our principal realm for signal theory, other classes do feature significantly in the sequel:

- Absolutely integrable signals, L^1;
- Bounded signals, L^∞;
- Certain subclasses of L^2, such as $L^1 \cap L^2$;
- Infinitely differentiable, rapidly decreasing signals.

It turns out that the L^1 signals constitute the best stepping-off point for constructing the Fourier transform (Chapter 5). This theory is the foundation of signal frequency analysis. With some extra effort, we can also handle L^2 signals, but we have to resort to limit operations to so extend the Fourier transform. Boundedness connects with absolute integrability in the theory of stable linear, translation-invariant systems: $y = Hx$ is stable when its impulse response $h = H\delta \in L^1(\mathbb{R})$. L^∞ is a basic, but useful, signal class. For example, if $f \in L^\infty$ and $g \in L^p$, then $fg \in L^p$.

Example (Square-Integrable Signals). The spaces $L^2[a, b]$ and $L^2(\mathbb{R})$ are Hilbert spaces. All of the L^p spaces are complete, and L^2 has an inner product that corresponds to its standard p-norm. These and their discrete cousin, $l^2(\mathbb{Z})$, the square-summable discrete signals, are the most important Hilbert spaces for signal theory.

Example (Absolutely and Square-Integrable Signals). The inner product on $L^2(\mathbb{R})$, restricted to those signals that are also absolutely integrable, furnishes a $\langle \cdot, \cdot \rangle$ operation for $L^1 \cap L^2$. We have to define $\|x\| = \langle x, x \rangle^{1/2}$ on this space. Note too that any Cauchy sequence of signals in $L^1 \cap L^2$ is still Cauchy in L^1. Thus, the sequence converges to an absolutely integrable limit. This sequence and its limit is also square-integrable, and so the limit is also in L^2. Thus, $L^1 \cap L^2$ is complete. It is easy to show that the restrictions of the signal addition and scalar multiplication operations to $L^1 \cap L^2$ are closed on that space. So $L^1 \cap L^2$ is a Hilbert space. We can say more: $L^1 \cap L^2$ is a *dense* subspace of L^2; for every square integrable signal $x(t)$ and any $\varepsilon > 0$, there is $y(t) \in L^1 \cap L^2$ such that $\|y - x\| < \varepsilon$. In fact, if $x(t) \in L^2$, then we may take

$$x_n(t) = \begin{cases} x(t) & \text{if } -n \le t \le n, \\ 0 & \text{if otherwise.} \end{cases} \tag{3.71}$$

Then $\lim_{n \to \infty} x_n = x$, and $x_n \in L^1(\mathbb{R})$ are absolutely integrable because they are compactly supported.

Example (Schwarz Space). The Schwarz space S is the class of infinitely differentiable, rapidly decreasing functions of a real variable [28]. *Infinitely differentiable* means that each $x(t) \in S$ has derivatives of all orders. Thus, $x(t)$ and its derivatives are all continuous. *Rapidly decreasing* means that $\lim_{t \to \infty} t^m x^{(n)}(t) = 0$ for all $m, n \in \mathbb{N}$, where $x^{(n)}(t) = d^n x/dt^n$. Examples of signals in S are the Gaussians of mean μ and standard deviation $\sigma > 0$:

$$g_{\mu, \sigma}(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}. \tag{3.72}$$

Examples of signals not in S include rational signals such as $(\sigma^2 + t^2)^{-1}$; these are not rapidly decreasing. The even decaying exponentials $\exp(-\sigma|t|)$ rapidly decrease, but fail differentiability, so they are not in the Schwarz class. The Schwarz space is a plausible candidate for the mathematical models of continuously defined naturally occurring signals: temperatures, voltages, pressures, elevations, and like quantitative phenomena. The Schwarz class strikes us as incredibly special and probably populated by very few signals. In fact, S is dense in both $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$. To see this, it suffices to show that elements of S are arbitrarily close to square pulse functions. Since linear combinations of square pulses are precisely the step functions, and since step functions are dense in $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$, this implies that S is dense as well. The trick is to blend the upper and lower ledge of a step together in an infinitely differentiable way [28]. The following function interpolates a unit step edge:

$$s(t) = \begin{cases} 0 & \text{if } t \le 0, \\ e^{-\frac{1}{t\exp\left[\frac{1}{t-1}\right]}} & \text{if } 0 < t < 1. \\ 1 & \text{if } 1 \le t. \end{cases} \tag{3.73}$$

Scalings and dilations of $s(t)$ interpolate a general step edge, and it is evident that arbitrary step functions are approximated to any precision by linear combinations of functions of the form (3.73).

There is a rich theory of linear operators and especially linear functionals on Hilbert spaces. Detailed explication of the ideas would take this presentation too far astray into abstract functional analysis; we refer the reader to the broad, excellent literature [13, 25, 26, 29] We shall be obliged to quote some of these results here, as elaborating some of our later signal analysis tools depends upon them.

Example (Inner Product). If we fix $h \in L^2(\mathbb{R})$, then the inner product $Tx = \langle x, h \rangle$ is a bounded linear functional with $\|T\| = \|h\|_2$. Indeed by Schwarz's inequality (3.70), $\|Tx\| = |\langle x, h \rangle| = \|xh\|_1 \le \|x\|_2 \|h\|_2$.

We have held up the inner product operation as the standard by which two signals may be compared. Is this right? It is conjugate linear and defined for square-integrable analog signals, which are desirable properties. We might well wonder whether any functional other than the inner product could better serve us as a tool for signal comparisons. The following theorem provides the answer.

Theorem (Riesz Representation). Let T be a bounded linear function on a Hilbert Space H. Then there is a unique $h \in H$ such that $Tx = \langle x, h \rangle$ for all $x \in H$. Futhermore, $\|T\| = \|h\|$.

Proof: See Ref. 29. ■

Inner products and bounded linear operators are very closely related. Using a generalization of the Riesz representation theorem, it is possible to show that every bounded linear Hilbert space operator $T : H \to K$ has a related map $S : K \to H$ which cross-couples the inner product.

Theorem. Let $T: H \to K$ be a bounded linear function on Hilbert space H and K. Then there is a bounded linear operator $S : K \to H$ such that:

(i) $\|T\| = \|S\|$.
(ii) For all $h \in H$ and $k \in K$, $\langle Th, k\rangle = \langle h, Sk\rangle$.

Proof: The idea is as follows. Let $k \in K$ and define the linear functional $L : H \to \mathbb{K}$ ($\mathbb{R}$ or $\mathbb{C}$) by $L(h) = \langle Th, k\rangle$. L is linear by the properties of the inner product. L is also bounded. The Riesz representation theorem applies, guaranteeing that there is a unique $g \in H$ such that $L(h) = \langle h, g\rangle$. Thus, we set $S(k) = g$. After verifying that S is linear and bounded, we see that it satisfies the two required properties [26]. ■

Definition (Adjoint). Let $T: H \to K$ be a bounded linear operator on Hilbert spaces H and K and S be the map identified by the previous theorem. Then S is called the *Hilbert adjoint operator* of T and is usually written $S = T^*$. If $H = K$ and $T^* = T$, then T is called *self-adjoint*.

Note that if T is self-adjoint, then $\langle Th, h\rangle = \langle h, Th\rangle$. So $\langle Th, h\rangle \in \mathbb{R}$ for all $h \in H$. This observation enables us to order self-adjoint operators.

Definition (Positive Operator, Ordering). A self-adjoint linear operator is *positive*, written $T \geq 0$, if for all h $\in H$, $0 \leq \langle Th, h\rangle$. If S and T are self-adjoint operators on a Hilbert space H with $T - S \geq 0$, then we say $S \leq \mathrm{T}$.

We shall use ordered self-adjoint linear operators when we study frame theory in Section 3.3.4.

Finally, important special cases of Hilbert operators are those that are isometries.

Definition (Isometry). If $T: H \to K$ is linear operator on Hilbert spaces H and K and for all $g, h \in H$ we have $\langle Tg, Th\rangle = \langle g, h\rangle$, then T is an *isometry.*

3.3.2.3 Application: Constructing Optimal Detectors. As an application of functional analysis ideas to signal analysis, consider the problem of finding a known or prototype signal $p(t)$ within a given, candidate signal $x(t)$. The idea is to convolve a kernel $k(t)$ with the input: $y(t) = (x * k)(t)$. Where the response $y(t)$ has a maximum, then hopefully $x(t)$ closely resembles the model signal $p(t)$. How should we choose $k(t)$ to make this work?

A commonly used approach is to let $k(t) = p(-\mathrm{t})$, the prototype pattern's reflection [30, 31]. Then $y(t) = (x * k)(t) = (x \circ p)(t)$, the correlation of $x(t)$ and $p(t)$. Since

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x(t) - p(t))^2 \, dt = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2(t)dt + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p^2(t)dt - 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(t)p(t)\, dt, \quad (3.74)$$

we see that minimizing the energy of the difference of $x(t)$ and $p(t)$ is equivalent to maximizing their inner product as long as the two signals have constant 2-norms. It is easy to do this for the prototype; we use the normalized signal $\tilde{p}(t) = p(t)/\|p\|_2$ as the model signal, for example. This step is called normalization, and so the method is often called *normalized cross-correlation*. If the entire candidate $x(t)$ is available at the moment of comparison, such as for signals acquired offline or as functions of a nontemporal independent variable, then we can similarly normalize $x(t)$ and compare it to $\tilde{p}(t)$. If, on the other hand, $x(t)$ is acquired in real time, then the feasible analysis works on past fragments of $x(t)$.

The Schwarz inequality tells us that equality exists if and only if the candidate and prototype are constant multiples of one another. If we subtract the mean of each signal before computing the normalized cross-correlation, then the normalized cross-correlation has unit magnitude if and only if the signals are related by $x(t) = Ap(t) + B$, for some constants A, B. Since $y(t) = (x * k)(t) = (x \circ p)(t)$ attains a maximum response when the prototype $p(t)$ matches the candidate $x(t)$, this technique is also known as matched filtering. It can be shown that in the presence of a random additive white noise signal, the optimal detector for a known pattern is still given by the matched filter [32].

Many signal and image processing applications depend upon matched filtering. In speech processing, one problem is to minimize the effect of reverberation from the walls and furnishings within a room on the recorded sound. This is an echo cancellation problem where there may be multiple microphones. The idea is to filter each microphone's input by the reflected impulse response of the room system [33].

Normalized cross correlation can be computationally demanding. When it is applied to images, this is especially the case. Consequently, many applications in image-based pattern matching use coarse resolution matched filtering to develop a set of likely match locations of the template against the input image. Then, finer resolution versions of the template and original image are compared [34]. Correlation techniques are also one of the cornerstones of image motion analysis [35].

3.3.3 Orthonormal Bases

In Chapter 2 a close relationship was established between general Hilbert spaces and the space of square-summable discrete signals. Here we list discuss four different orthogonal basis sets:

- Exponential signals of the form $\exp(jnt)$, where $n \in \mathbb{N}$
- Closely related sinusoids, which are the real and imaginary parts of the exponentials
- Haar basis, which consists of translations and dilations of a single, simple step function
- Sinc functions of the form $s_n(t) = \frac{1}{\sqrt{\pi}} \frac{\sin(At - n\pi)}{At - n\pi}$, where $A > 0$ and $n \in \mathbb{N}$

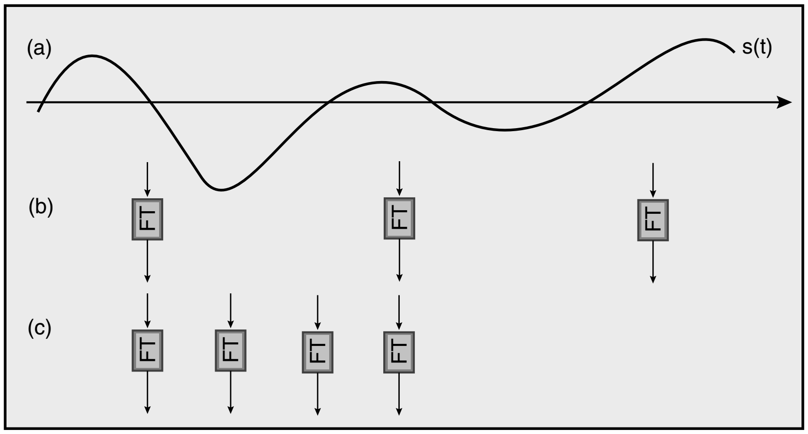

Fig. 3.5. Signal decomposition on an orthonormal basis (a). Sparse representations are better (b) than decompositions that spread signal energy across many different basis elements (c).

Orthonormal bases are fundamental signal identification tools. Given a signal $x(t)$, we project it onto the orthonormal basis set $\{e_n \mid n \in \mathbb{N}\}$ by taking the inner products $c_n = \langle x(t), e_n(t)\rangle$. Each c_n indicates the relative presence of the basis element e_n inside $x(t)$. The strategy is that—hopefully, at least—the set of coefficients $\{c_n \mid n \in \mathbb{N}\}$ is a simpler description of the original signal. If they are not, then we attempt to find an alternative basis set that better captures the character of anticipated input signals. We can say more, though. It works best for the signal recognition application if decomposition produces only a few significant c_n values for every typical $x(t)$ that we try to analyze. In other words, for the original signals $x(t)$ that we expect to feed into our analysis application, the energy of the decomposition coefficients is sparse and concentrated in a relative few values. On the other hand, dense decomposition coefficient sets make signal classification harder, because we cannot clearly distinguish which $e_n(t)$ factor most critically within $x(t)$. Figure 3.5 illustrates the idea. For signal identification, therefore, the upshot is that the statistics of the decomposition coefficients for typical system input signals are our guide for selecting an orthonormal basis set.

3.3.3.1 Exponentials. The most important basis for the L^2 Hilbert spaces is the exponential signals. We begin by considering the space $L^2[-\pi, \pi]$.

Let $\{e_n(t) \mid n \in \mathbb{N}\}$ be defined by $e_n(t) = (2\pi)^{-1/2}\exp(jnt)$. It can be easily shown that the $e_n(t)$ are indeed orthonormal (exercise). Similarly, if we set

$$e_n(t) = \frac{1}{\sqrt{b-a}} e^{2\pi jn\frac{t-a}{b-a}}, \tag{3.75}$$

then $B = \{e_n(t) \mid n \in \mathbb{N}\}$ is orthonormal in $L^2[a, b]$. We shall show in Chapter 5 that B is complete so that it is, in fact, a basis. Thus, for any square-integrable signal $x(t)$ on $[a, b]$ a linear combination of $\{e_n\}$ is arbitrarily close to $x(t)$; in other words,

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{jnt} \tag{3.76}$$

for some constants $\{c_n \mid n \in \mathbb{N}\}$, called the Fourier series coefficients for $x(t)$. Note that c_n measures the similarity of $x(t)$ to the basis element $\exp(jnt)$: $\langle x(t), \exp(jnt)\rangle = c_n$ by orthonormality.

Now consider the case of $L^2(\mathbb{R})$. We can break up the real line into 2π-wide intervals $I_m = [(2m-1)\pi, (2m+1)\pi]$. Let X_m be the characteristic function on I_m, and set $e_{m,n}(t) = X_m(2\pi)^{-1/2}\exp(jnt)$. Then clearly $\{e_{m,n}(t) \mid m, n \in \mathbb{N}\}$ is an orthonormal basis for $L^2(\mathbb{R})$.

3.3.3.2 Sinusoids. There is an orthonormal basis for $L^2[-\pi, \pi]$ consisting entirely of sinusoids. Let us break up $e_n(t) = (2\pi)^{-1/2}\exp(jnt)$ into its real and imaginary parts using $\exp(jt) = \cos(t) = j\sin(t)$. We set $a_n = c_n + c_{-n}$ and $jb_n = c_{-n} - c_n$. Thus, (3.36) becomes

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(nt) + \sum_{n=1} b_n \sin(nt), \tag{3.77}$$

and any $x(t) \in L^2[-\pi, \pi]$ can be expressed as a sum of sinusoids. Equation (3.77) shows that in addition to bona fide sinusoids on $[-\pi, \pi]$, we need one constant function to comprise a spanning set. That the sinusoids are also orthogonal follows from writing them in terms of exponentials:

$$\cos(t) = \frac{e^{jt} + e^{-jt}}{2} \tag{3.78a}$$

$$\sin(t) = \frac{e^{jt} - e^{-jt}}{2j} \tag{3.78b}$$

and using the orthonormality of the exponentials once more. As with exponentials, the sinusoids can be assembled interval-by-interval into an orthonormal basis for $L^2(\mathbb{R})$. The exercises further explore exponential and sinusoidal basis decomposition.

3.3.3.3 Haar Basis. The Haar[3] basis uses differences of shifted square pulses to form an orthonormal basis for square-integrable signals. It is a classic construction

[3]Hungarian mathematician Alfréd Haar (1885–1933) was Hilbert's graduate student at Göttingen. The results of his 1909 dissertation on orthogonal systems, including his famous basis set, were published a year later.

[36], dating from the early 1900s. It is also quite different in nature from the exponential and sinusoidal bases discussed above.

The sinusoidal basis elements consist of sinusoids whose frequencies are all integral multiples of one another—*harmonics*. As such, they all have different shapes. Thus, cos(t) follows one undulation on $[-\pi, \pi]$, and it looks like a shifted version of sin(t). Futhermore, $\cos(2t)$ and $\sin(2t)$ resemble one another as shapes, but they are certainly different from $\cos(t)$, $\sin(t)$, and any other basis elements of the form $\cos(nt)$ or $\sin(nt)$ where $n \neq 2$.

Haar's orthonormal family begins with a single step function, defined as follows:

$$h(t) = \begin{cases} 1 & \text{if } 0 \le t < \dfrac{1}{2}, \\ -1 & \text{if } \dfrac{1}{2} \le t \le 1, \\ 0 & \text{if otherwise.} \end{cases} \tag{3.79}$$

Nowadays $h(t)$ is called a *Haar wavelet*. Haar's basis contains dilations and translations of this single atomic step function (Figure 3.6). Indeed, if we set $H = \{h_{m,n}(t) = 2^{n/2}h(2^n t - m) \mid m, n \in \mathbb{N}\}$, then we claim that H is an orthonormal basis for $L^2(\mathbb{R})$.

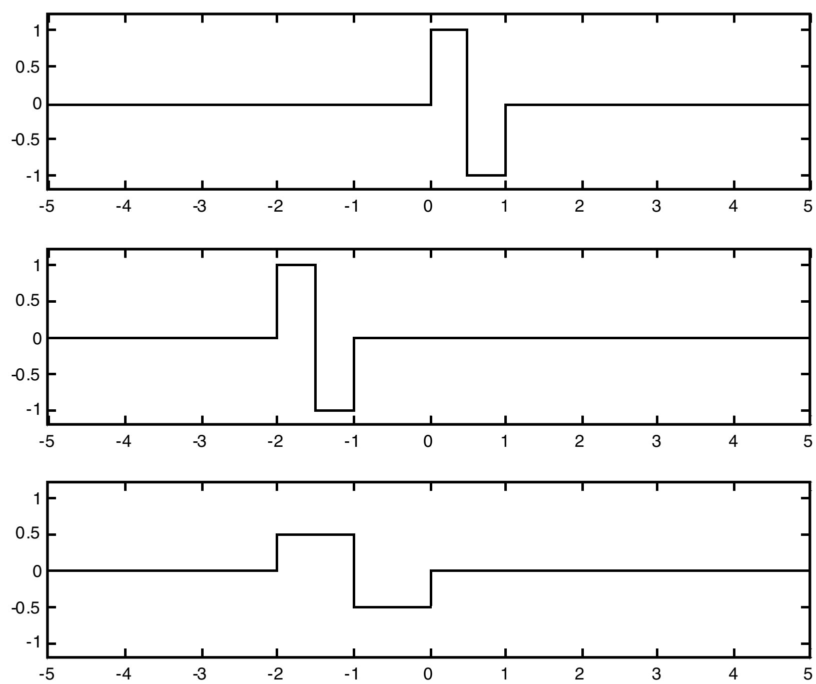

Fig. 3.6. Some examples of Haar basis elements.

Orthonormality is easy. If the supports of two unequal basis elements, $h_{m,n}(t)$ and $h_{p,q}(t)$, overlap, then we must have $n \neq q$; let us suppose $n > q$. Then the support of $h_{m,n}(t)$ will lie completely within one of the dyadic subintervals of Support $(h_{p,q}(t))$. The inner product must be zero: $\langle h_{m,n}(t), h_{p,q}(t)\rangle = 0$.

Completeness—that the closure of the span of $H = \{h_{m,n}(t) \mid m, n \in \mathbb{N}\}$ happens to be all of $L^2(\mathbb{R})$—is somewhat harder to show. Informally, any square-integrable signal can be approximated to arbitrary precision with a step function. Likewise, if we can approximate step functions with linear combinations of $\{h_{m,n}(t)\}$, then completeness follows. We first show that the unit square pulse, $s(t) = u(t-1) - u(t)$, is in Span(H). Indeed, one can check that for any $N \in \mathbb{N}$,

$$\sum_{n=-\infty}^{-1} 2^{\frac{n}{2}} h_{0,n}(t) = \begin{cases} 0 & \text{if } t \leq 0, \\ 1 & \text{if } 0 < t < 1, \\ 0 & \text{if } 2^N < t < 2^{N+1}. \end{cases} \tag{3.80}$$

So on all but a countably infinite set of points, the sum on the left-hand side of (3.80) is $s(t)$. When we study the Lebesgue integral, we shall see that these discontinuities do not affect the integral. Similarly, we can show that dyadic dilations and translations of $s(t)$ are in Span(H). Linear combinations of these dilations and translations can be arbitrarily close to a general step function, which justifies our completeness claim.

Practically, the importance of this is that it is good for finding one particular shape, which might manifest itself in different sizes or scales inside a signal. By contrast, the sinusoidal basis elements find various harmonics—distinct shapes—on the subintervals by which they divide the time domain. Of course, the elemental shape of the Haar basis is quite simple, but the basis elements are tuned to this shape at an infinite range of scales and an infinite set of positions. The two different types of basis in fact represent two different ways of analyzing signals. The sinusoidal and exponential bases find frequencies inside signals; they are *frequency-domain* tools for signal interpretation. The Haar basis finds a single atomic shape at different scales inside a signal, so it exemplifies *scale-domain* methods.

Recent books on wavelets [37, 38] cover Haar's basis. Until the mid-1980s, mathematicians believed that the only possible orthonormal bases for $L^2(\mathbb{R})$ which used dilations and translations of a single signal atom were those like Haar's construction where the atom has step discontinuities. The new wavelet theories, however, have completely toppled this long-standing conviction (Chapter 11).

3.3.3.4 Sinc Functions. Another important basis set consists of $\mathrm{sinc}(t) = \sin(t)/t$ functions. In later chapters we shall have the tools in hand to show that the following family of functions is orthonormal: $\{s_n(t) \mid n \in \mathbb{N}\}$, where

$$s_n(t) = \frac{1}{\sqrt{\pi}} \frac{\sin(At - n\pi)}{At - n\pi} \tag{3.81}$$

and $A > 0$. We shall find that $\{s_n(t)\}$ defined in (3.81) spans an important Hilbert subspace of $L^2(\mathbb{R})$: a band-limited subspace that consists of signals whose spectral content lies entirely within a finite frequency range.

3.3.4 Frames

The concept of a frame generalizes the idea of an orthonormal basis in Hilbert space. Frame representations may be *overcomplete*. This means they may represent a signal in more than one way, and therefore they are not orthogonal. To understand why this might be useful and, indeed, why a signal analysis application based on frames can still be made workable, we have to reflect for a moment on the use of orthonormal bases for signal representation.

Many signal and image analysis applications need to recognize an unknown, candidate profile as an instance or combination of certain known prototypes. Probably the first approach that comes to mind is to decompose the candidate signal as a linear combination of the prototypical building blocks: $x(t) = \Sigma c_n e_n(t)$. The c_n represent the relative weight of each $e_n(t)$ in this linear signal combination. The application's desirable features are as follows:

(i) Incoming signals are uniquely represented by the coefficients $\langle x, e_n \rangle$; $x(t)$ does not have multiple identifying strings of weighting coefficients.

(ii) Two representations in terms of decomposition coefficients should permit a straightforward comparison of source signals for the differences between them.

(iii) Any original signal should be reconstructible from the stored coefficients; this is a completeness criterion.

(iv) Finally, the reconstruction of $x(t)$ from $\{c_n\}$ should be numerically stable; a small change in the coefficients results in a small change on the rebuilt signal.

These criteria suggest orthonormal bases. If we use an orthogonal set, such as the exponential functions $e_n(t) = e^{-jnt}$, then the four conditions hold in theory. The orthogonal character of the underlying special functions eases the computation of coefficients. However, a local change in the incoming signal leads to changes in the whole range of decomposition coefficients. Noise in signals, however sparse, drastically perturbs the stored digital form of the signal. This leads to practical difficulties in comparing two outwardly similar signals.

Sometimes, in an attempt to localize changes in signals to appropriate portions of the coefficient set, the decomposition functions are chosen to be windowed or "short-time" exponentials. These may take the form $e_{n,m}(t) = C_m e^{-jnt}$, where C_m is the characteristic function of the integral interval $[m, m + 1]$. The downfall of this tactic is that it adds high-frequency components to the decomposition. Relatively smooth signals decompose into sequences with unusually large coefficients for large values of n. This problem can be ameliorated by choosing a smoother window function—a Gaussian instead of a square window, for example. This becomes a representation in terms of the Gabor elementary functions, which first chapter introduced. But a deeper problem with windowed exponentials arises. It turns out that one cannot construct an orthonormal basis out of Gaussian-windowed exponentials. Chapter 10 will cover this famous result, known as the Balian–Low theorem. Nonetheless, it is possible to make *frames* out Gabor elementary functions, and this is a prime reason for the technique's recent popularity.

Earlier, however, we noted that the statistics of the decomposition coefficients are an important consideration. Sparse sets are better than dense sets. We form vectors out of the decomposition coefficient sets of system input signals. The vectors comprise a library of signal models. Sparse decomposition coefficients imply short vectors, and short vectors mean that the library more succinctly represents the essential aspects of our signal models.

Unfortunately, it quite often happens that orthonormal bases produce nonsparse coefficient sets on fairly simple signals. For example, a decaying pulse signal, $x(t) = \exp(-At^2)\cos(Bt)$, contains a local frequency component set by the width of the Gaussian envelope. Now, $x(t)$ can be represented by the exponential or sinusoidal basis sets, but far from the origin there will always be large weighting coefficients. These distant, high-frequency sinusoidal wiggles have to cancel one another in order to correctly represent the negligible amplitude of $x(t)$ in their neighborhood. Libraries based on such orthonormal bases can also be problematic; when the signal changes a little bit, many decomposition coefficients must change globally to effect just the right weighting to cancel and reinforce the right signal components.

One surprising result from contemporary signal analysis research is that overcomplete representations—in particular, the frame decomposition that we cover here—can help in constructing sparse signal representations. Special techniques, such as the matching pursuit algorithm have been devised to cope with the overcompleteness [39, 40]. Originating in the 1950s [41, 42], frames are now widely used in connection with the recent development of the theory of time-frequency and time-scale transforms [38, 43].

First-time readers may skip this section. Visitors to the latest research literature on time-frequency and time-scale transform methods will, however, commonly encounter frame theoretic discussions. We shall cover frames more thoroughly in Chapters 10–12.

3.3.4.1 Definition and Basic Properties.

This section defines the notion of a Hilbert space frame and provides some simple connections to the more specific and familiar concept of orthonormal basis. In what follows, we shall allow Hilbert spaces defined over the complex numbers.

Definition (Frame). Let $\{f_n: n \in \mathbb{Z}\}$ be signals in a Hilbert space H. If there are positive $A, B \in \mathbb{R}$ such that for all $x \in H$ we have

$$A\|x\|^2 \le \sum_{n=-\infty}^{\infty} |\langle x, f_n\rangle|^2 \le B\|x\|^2, \tag{3.82}$$

then the $\{f_n\}$ constitute a *frame* in H. The constants A and B are the *frame bounds*, *lower* and *upper*, respectively. The frame is *tight* if $A = B$. A frame is *exact* if it is no longer a frame following the deletion of a single element.

It is immediately clear that the frame condition (3.82) generalizes the idea of an orthonormal basis, makes the convergence unconditional, and ensures that a frame is a complete set (the closure of its linear span is all of H).

Proposition. Let $\{e_n: n \in \mathbb{Z}\}$ be an orthonormal basis for a Hilbert space H. Then $\{e_n\}$ is a tight exact frame having bounds $A = B = 1$.

Proof: Let $x \in H$. Parseval's relation for Hilbert spaces implies

$$\sum_{n=-\infty}^{\infty} |\langle x, e_n\rangle|^2 = \|x\|^2. \tag{3.83}$$

Therefore,

$$1 \cdot \|x\|^2 = \sum_{n=-\infty}^{\infty} |\langle x, e_n\rangle|^2 = 1 \cdot \|x\|^2, \tag{3.84}$$

which is precisely the frame condition (3.82), with $A = B = 1$. ■

Proposition (Unconditionality). Let $\{f_n : n \in \mathbb{Z}\}$ be a frame in a Hilbert space H. Then any rearrangement of the sequence $\{f_n\}$ is also a frame.

Proof: If $x \in H$, then $\{|\langle x, f_n\rangle|^2\}$ is a sequence of positive real numbers, and the convergence of the series

$$\sum_{n=-\infty}^{\infty} |\langle x, f_n\rangle|^2 \tag{3.85}$$

is absolute [6]. This means that the above sum converges to the same value under any rearrangement of the $\{f_n\}$. ■

Remark. Thus, we are free to renumber a frame with the natural numbers. Later, in Chapter 10, we shall find it convenient to use pairs of integers to index frame elements. This same idea is at work, as long as the index set has the same cardinality as the natural numbers.

Proposition (Completeness). Let $F = \{f_n : n \in \mathbb{Z}\}$ be a frame in a Hilbert space H. Then $\{f_n\}$ is a total set: $\overline{\text{Span}}\,\{f_n\} = H$.

Proof: Let $x \in H$, $\langle x, f_n \rangle = 0$ for all n, and $A > 0$ be the lower frame bound for F. By the definition of a frame,

$$0 \le A\|x\|^2 \le \sum_{n=-\infty}^{\infty} |\langle x, f_n \rangle|^2 = 0. \tag{3.86}$$

Equation (3.86) shows that $x = 0$. Since a subset X of H is complete if and only if no nonzero element is orthogonal to all elements of X, it must be the case that F is total in H. ■

3.3.4.2 Examples. In the following examples, H is a Hilbert space, and $\{e_n: n \in \mathbb{N}\}$ is an orthonormal basis in H.

Example (Overcompleteness). Let $F = \{f_n\} = \{e_0, e_0, e_1, e_1, e_2, e_2, ...\}$. Then F is a tight frame with bounds $A = B = 2$. Of course, it is not an orthonormal basis, although the subsequence $\{f_{2n}: n \in \mathbb{N}\}$ is orthonormal. It is also not exact. Elements of H have multiple decompositions over the frame elements.

Example (Orthogonal, Yet Not a Frame). $F = \{f_n\} = \{e_0, e_1/2, e_2/3, ...\}$ is complete and orthogonal. However, it is not a frame, because it can have no positive lower frame bound. To see this, assume instead that F is a frame and let $A > 0$ be its lower bound. Let N be large enough that $N^{-2} < A$ and set $x = e_N$. Then, (3.82) gives

$$A = A\|x\|^2 \le \sum_n |\langle x, f_n \rangle|^2 = |\langle e_N, f_N \rangle|^2 = \frac{1}{N^2} < A, \tag{3.87}$$

a contradiction.

Example (Tight Frame). $F = \{f_n\} = \{e_0, 2^{-1/2}e_1, 2^{-1/2}e_1, 3^{-1/2}e_2, 3^{-1/2}e_2, 3^{-1/2}e_2, ...\}$ is a tight frame with bounds $A = B = 1$. F is not exact.

Example (Exact But Not Tight). $F = \{f_n\} = \{2e_0, e_1, e_2, ...\}$ is a frame, with lower bound $A = 1$ and upper bound $B = 2$. F is exact but not tight.

3.3.4.3 Frame Operator. There is a natural bounded linear operator associated with a frame. In fact, if the decomposition set is a frame, then the basic signal analysis system that associates signals in a Hilbert space with their decomposition coefficients is just such an operator. This section gives their definition and properties. While they are mathematically elegant and abstract, frame operators also factor critically in using overcomplete frames in signal representation.

Definition (Frame Operator). If $F = \{f_n(t): m, n \in \mathbb{Z}\}$ is a frame in a Hilbert space H, then the associated frame operator $T_F: H \to l^2(\mathbb{Z})$ is defined by

$$T_F(x)(n) = \langle x, f_n \rangle . \tag{3.88}$$

In other words, $y = T_F(x) = T_F x$ is the complex-valued function defined on the integers such that $y(n) = \langle x, f_n \rangle$. When the frame is clear by the context we may drop the subscript on the frame operator: $y = Tx$.

Proposition. If $F = \{f_n(t): n \in \mathbb{Z}\}$ is a frame in a Hilbert space H, then the associated frame operator T given by (3.88) is linear and bounded. Furthermore, if B is the upper frame bound, then $\|T\| \le B^{1/2}$.

Proof: Linearity is clear from inner product properties. If $x \in H$, then

$$\|Tx\| = \left(\sum_n \langle x, x_n \rangle \overline{\langle x, x_n \rangle}\right)^{\frac{1}{2}} = \left(\sum_n \left|\langle x, x_n \rangle\right|^2\right)^{\frac{1}{2}} \le (B\|x\|^2)^{\frac{1}{2}} = B^{\frac{1}{2}}\|x\| . \tag{3.89}$$

■

Definition (Adjoint Frame Operator). Let $F = \{f_n(t) : n \in \mathbb{Z}\}$ be a frame in a Hilbert space H and let T be its associated frame operator T (3.88). The *adjoint frame operator* $S : l^2(\mathbb{Z}) \to H$ is defined for $y(n) \in l^2(\mathbb{Z})$ by

$$S(y) = \sum_{n=-\infty}^{\infty} y(n) f_n. \tag{3.90}$$

Proposition. Let S be given by (3.90). Then S is the Hilbert adjoint operator with respect to frame operator T of F: $S = T^*$.

Proof: That S is linear is left as an exercise. Let $x \in H$, T the frame operator for $F = \{f_n(t) : n \in \mathbb{Z}\}$, and let $y = \{y_n : n \in \mathbb{Z}\}$ be some sequence in l^2. Then

$$\langle x, Sy \rangle = \left\langle x, \sum_n y_n x_n \right\rangle = \sum_n \overline{y_n} \langle x, x_n \rangle, \tag{3.91}$$

and also

$$\langle Tx, y \rangle = \sum_n \langle x, x_n \rangle \overline{y_n} = \sum_n \overline{y_n} \langle x, x_n \rangle . \tag{3.92}$$

Together, (3.91) and (3.92) show that T and S cross-couple the inner products of the two Hilbert spaces. Therefore, $S = T^*$. ■

The next theorem, one of the classic results on frames and frame operators, offers a characterization of frames [41]. It uses the idea of a positive operator. Recall that an operator T on a Hilbert space H is *positive*, written $T \ge 0$, when $\langle Tx, x \rangle \ge 0$ for all $x \in H$. Also, if S and T are operators on H, then $T \ge S$ means that $T - S \ge 0$. Positive operators are self-adjoint [26, 29].

Theorem (Frame Characterization). Let $F = \{f_n : n \in \mathbb{Z}\}$ be a sequence in a Hilbert space H; $A, B > 0$; and let I be the identity operator on H. Then F is a frame with lower and upper bounds A and B, respectively, if and only if the operator S defined by

$$Sx = \sum_n \langle x, f_n \rangle f_n \tag{3.93}$$

is a bounded linear operator with

$$AI \le S \le BI. \tag{3.94}$$

Proof: Assume that (3.93) defines S and (3.94) holds. By the definition of $\le$ for operators, for all $x \in H$ (3.94) implies

$$\langle AIx, x\rangle \le \langle Sx, x\rangle \le \langle BIx, x\rangle. \tag{3.95}$$

However,

$$\langle AIx, x\rangle = A\|x\|^2 \tag{3.96a}$$

and

$$\langle BIx, x\rangle = B\|x\|^2. \tag{3.96b}$$

The middle term of (3.95) is

$$\langle Sx, x\rangle = \left\langle \sum_n \langle x, f_n\rangle f_n, x\right\rangle = \sum_n \langle x, f_n\rangle \overline{\langle f_n, x\rangle} = \sum_n \left|\langle x, f_n\rangle\right|^2. \tag{3.97}$$

Together (3.96a), (3.96b), and (3.97) can be inserted into (3.95) to show that the frame condition is satisfied for F.

Conversely, suppose that F is a frame. We must first show that S is well-defined—that is, the series (3.93) converges. Now, the Schwarz inequality implies that the norm of any $z \in H$ is $\sup\{|\langle z, y\rangle| : y \in H \text{ and } \|y\| = 1\}$. Let s_N represent partial sums of the series (3.93):

$$s_N(x) = \sum_{n=-N}^{N} \langle x, f_n\rangle f_n. \tag{3.98}$$

When $M \le N$ the Schwarz inequality applies again:

$$\|s_N - s_M\|^2 = \sup_{\|y\| = 1} \left\{ \left|\langle s_N - s_M, y\rangle\right|^2 \right\} = \sup_{\|y\| = 1} \left\{ \left| \sum_{n=M+1}^{N} \langle x, f_n\rangle \langle f_n, y\rangle \right|^2 \right\}. \tag{3.99}$$

Algebra on the last term above gives

$$\|s_N - s_M\|^2 \le \sup_{\|y\| = 1} \left\{ \sum_{n=M+1}^{N} \left|\langle x, f_n\rangle\right|^2 \sum_{n=M+1}^{N} \left|\langle f_n, y\rangle\right|^2 \right\}. \tag{3.100}$$

By the frame condition,

$$\|s_N - s_M\|^2 \le \sup_{\|y\| = 1} \sum_{n=M+1}^{N} \left|\langle x, f_n\rangle\right|^2 B\|y\|^2 = B \sum_{n=M+1}^{N} \left|\langle x, f_n\rangle\right|^2, \tag{3.101}$$

and the final term in (3.101) must approach zero as $M, N \to \infty$. This shows that the sequence $\{s_N \mid N \in \mathbb{N}\}$ is Cauchy in the Hilbert space H. H is complete, so the $\{s_N\}$ converge, the series must converge, and the operator S is well-defined.

Similarly,

$$\|Sx\|^2 \le \sup_{\|y\| = 1} |\langle Sx, y\rangle|^2, \tag{3.102}$$

which entails $\|S\| \le B$. From the frame condition and (3.93) the operator ordering, $AI \le S \le BI$ follows immediately. ■

Remark. Notice that for a frame $F = \{f_n : n \in \mathbb{Z}\}$ the theorem's operator S is the composite T^*T, where T is the frame operator and T^* is its adjoint. The following corollaries provides further properties of T^*T.

Corollary. Let $F = \{f_n : n \in \mathbb{Z}\}$ be a frame in a Hilbert space H and let T be the frame operator. Then the map $S = T^*T : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ given by

$$(T^*T)(x) = \sum_{n=-\infty}^{\infty} \langle x, f_n\rangle f_n. \tag{3.103}$$

is positive and invertible.

Proof: Let A be the lower frame bound for F. Since $AI \le S$, $S - AI \ge 0$, by the definition of the operator ordering relation. Also, since $A > 0$, $S/A - I \ge 0$. A property of positive operators is that if an operator, say U, is positive, $U \ge 0$, then $U + I$ is invertible [27]. Therefore, $S/A - I + I = S/A$ is invertible. Clearly then, S is invertible. Moreover, adding a positive operator I to $S/A - I$ still gives a positive operator. This shows that S is indeed positive. ■

Corollary. Let $F = \{f_n : n \in \mathbb{Z}\}$ be a frame in a Hilbert space H; let A and B be the lower and upper frame bounds, respectively, for F; let T be the frame operator; and let $S = T^*T$ be given by the previous corollary. Then $I/B \le S^{-1} \le I/A$.

Proof: The previous corollary shows that S^{-1} exists. Since S^{-1} commutes with I and with S, and since $AI \le S \le BI$, it follows that $S^{-1}AI \le S^{-1}S \le S^{-1}BI$. Upon rearrangment, this yields $B^{-1}I \le S^{-1} \le A^{-1}I$. ■

These results allow us to define the concept of the dual frame. The dual frame is the key concept for applying frames in signal analysis applications.

Definition (Dual Frame). Let $F = \{f_n : n \in \mathbb{Z}\}$ be a frame in a Hilbert space H and T be the frame operator. We define the *dual frame* to F by applying the inverse of T^*T to frame elements:

$$\tilde{F} = \left\{(T^*T)^{-1}(f_n)\right\}_{n \in Z}. \tag{3.104}$$

Corollary. Let $F = \{f_n: n \in Z\}$ be a frame in a Hilbert space H; let A and B the lower and upper frame bounds, respectively, for F; let T be the frame operator; and let $S = T^*T$. Then the sequence $\{S^{-1}f_n \mid n \in \mathbb{Z}\}$ in H is a frame with lower bound B^{-1} and upper bound A^{-1}.

Proof: S^{-1} exists and is positive. Let $x \in H$ and note that

$$S^{-1}x = S^{-1}(SS^{-1}x) = S^{-1}\left(\sum_n \langle S^{-1}x, f_n\rangle f_n\right) = \sum_n \langle S^{-1}x, f_n\rangle S^{-1}f_n \tag{3.105}$$

by the linearity and continuity of S^{-1}. Since every positive operator is self-adjoint (Hermitian) [29], S^{-1} is self-adjoint. Hence,

$$S^{-1}x = \sum_n \langle S^{-1}x, f_n\rangle S^{-1}f_n = \sum_n \langle x, S^{-1}f_n\rangle S^{-1}f_n\,. \tag{3.106}$$

Notice that (3.106) is precisely the form that the operator S^{-1} takes in (3.93) of the frame characterization theorem. That $B^{-1}I \le S^{-1} \le A^{-1}I$ follows from an earlier corollary. Thus, the theorem's condition applies, and $\{S^{-1}f_n\}$ is a frame in H. ■

Corollary. Under the assumptions of the previous corollary, any $x \in H$ can be written

$$x = \sum_n \langle x, S^{-1}f_n\rangle f_n = \sum_n \langle x, f_n\rangle S^{-1}f_n. \tag{3.107}$$

Proof: Using (3.105), (3.106), and $x = SS^{-1}x = S^{-1}Sx$, the result follows easily. ■

Corollary. Further assuming that the frame T is tight, we have $S = AI$, $S^{-1} = A^{-1}I$, and, if $x \in H$, then

$$x = A^{-1}\sum_n \langle x, f_n\rangle f_n. \tag{3.108}$$

Proof: Clear from the definition of tightness and the preceding corollaries. ■

3.3.4.4 Application: Stable Modeling and Characterization. These results shed light on our proposed requirements for a typical signal analysis system. Let us list what we know so far:

- The first requirement—that the representation be unique—was demonstrated not to hold for general frames by an easy counterexample.
- The second specification—that signal representations should permit a straightforward comparison of two incoming signals—is satisfied by the frame operator that maps signals to sequences of complex numbers allowing us to use the l^2 norm for comparing signals.

- The corollary (3.107) fulfills the requirement that the original signal should be reconstructible from the decomposition coefficients.
- The fourth requirement has been left rather vague: What does numerical instability mean?

We can understand numerical instability in terms of bounded operators. Let $F = \{f_n\}$ be a frame and $T = T_F$ its frame operator. If the inverse mapping T^{-1} is unbounded, then elements of l^2 of unit norm will be mapped back to elements of H having arbitrarily large norms. This is not at all desirable; signals of enormous power as well as signals of miniscule power will map to decomposition coefficients of small l^2-norm. *This is numerical instability.* The next result shows that frame-based signal decomposition realizes the fourth requirement of a signal analysis system.

Corollary. Let $F = \{f_n : n \in \mathbb{Z}\}$ be a frame in a Hilbert space H and let $T = T_F$ be the associated frame operator. Then the inverse T^{-1} exists and is bounded.

Proof: Let $S = T^*T$, as in the previous section. Then S^{-1} exists and $S^{-1}T^* = T^{-1}$ is the bounded inverse of T. Alternatively, (3.107) explicitly maps a square-summable sequence in $\mathbb{C}$ to H, and it inverts T. A straightforward calculation shows that the map is bounded with $\|T^{-1}\| \le A^{-1/2}$. ■

Remarks. So the use of a frame decomposition for the signal analysis system allows a *stable reconstruction* of incoming signals from the coefficients obtained previously. In the case that the frame used in the signal processing system is tight, then the reconstruction is much simpler (3.108). We can reconstruct a signal from its decomposition coefficients using (3.108) alone; there is no need to invert $S = T^*T$ to get $S^{-1}f_n$ values.

We have substantiated all of the basic requirements of a signal analysis system, except for the first stipulation—that the coefficients of the decomposition be unique. The exercises elaborate some properties of exact frames that allow us to recover this uniqueness property. Briefly, if $F = \{f_n(t) : n \in \mathbb{Z}\}$ is a frame in a Hilbert space H, $T = T_F$ is its associated frame operator (3.88), and $S = T^*T$, then we know from the corollaries to the frame representation theorem that for any $x \in H$, if $a_n = \langle x, S^{-1}f_n \rangle$, then $x = \Sigma a_n f_n$. We can also show (exercises) that if there is some other representation of x, then it is no better than the the one we give in terms of the dual frame. That is, if there are $c_n \in C$ such that $x = \Sigma c_n f_n$, then

$$\sum_n |c_n|^2 = \sum_n |a_n|^2 + \sum_n |a_n - c_n|^2 ; \tag{3.109}$$

the representation by dual frame elements is the best in a least-squares sense. Later chapters (10–12) further cover frame representations and elaborate upon this idea.

To sum up. We began by listing the desired features of a signal analysis system. The notion of a frame can serve as a mathematical foundation for the decomposition, analysis, and reconstruction of signals. Orthonormal bases have very nice computational properties, but their application is often confounded by an undesirable practicality: The representations may well not be sparse. Frame theoretic approaches are a noteworthy alternative trend in recent signal analysis research, offering improved representation density over orthonormal bases [39, 40, 43].

3.4 MODERN INTEGRATION THEORY

The Lebesgue integral offers several theoretical advantages over the Riemann integral. The modern integral allows us to define Banach spaces directly, rather than in terms of the completion of a simpler space based on continuous analog signals. Also, the Lebesgue integral widens considerably the class of functions for which we can develop signal theory. It supports a powerful set of limit operations. Practically speaking, this means that we can develop powerful signal approximation techniques where signal and error magnitudes are based upon Lebesgue integrals (here we have in mind, of course, the L^p-norm).

This material has not been traditionally included in the university engineering and science curricula. Until recently, these disciplines could get along quite well without the mathematician's full toolkit. Mixed domain transform methods, Gabor and wavelet transforms in particular, have entered into widespread use in signal processing and analysis in the last several years. And carried with them has been an increased need for ideas from abstract functional analysis, Hilbert space techniques, and their mathematical underpinnings, among them the Lebesgue integral.

First-time readers and those who are content to build Banach spaces indirectly, by completing a given normed linear space, may elect to skip the material on the modern integral. Frankly, much of the sequel will still be quite understandable. Occasionally, we may worry that a signal is nonzero, but has Lebesgue integral zero; the mathematical term is that the signal is zero *almost everywhere* on an interval $[a, b]$. This is the standard, albeit homely, term for a function with so many zero points that its L^1-norm with respect to the Lebesgue integral is zero. We also say two signals are *equal almost everywhere* when their difference is zero almost everywhere.

Standard mathematical analysis texts cover measure and integration theory in far more detail and generality than we need here [24, 44, 45]. A tutorial is contained in Ref. 25. We cover the basic theoretical development only in these settings: measures on subsets of the real line and complex plane, real- and complex-valued functions, and their integrals. So restricted, this treatment follows the classic approach of Ref. 44.

Calculus defines the Riemann[4] integral as a limit of sums of areas of rectangles [6, 25]. Another approach uses trapezoids instead of rectangles; it offers somewhat

[4]Georg Friedrich Bernhard Riemann (1826–1866) studied under a number of great German mathematicians of the nineteenth century, including Gauss and Dirichlet. In 1857 he assumed a professorship at Göttingen. He contributed important results to complex variable theory, within which the Cauchy–Riemann equations are fundamental, and to non-Euclidean geometries, whose Riemannian manifolds Einstein much later appropriated for modern cosmology. Riemann is most widely known, ironically perhaps, for a relatively minor accomplishment—formalizing the definition of the conventional integral from calculus [R. Dedekind, Biography of Riemann, in H. Weber, ed., *Collected Works of Bernhard Riemann*, New York: Dover, 1953].

better numerical convergence. Either approach produces the same result. For rectangular sums, let us recall the definition.

Definition (Riemann Integral). Let $x(t)$ be continuous on the interval $I = [a, b]$; $a = t_0 < t_1 < \cdots < t_N = b$ partition I; and, for each subinterval, $I_k = [t_k, t_{k+1}]$, let r_k and s_k be the minimum and maximum values, respectively, of $x(t)$ on I_k. Then the lower and upper Riemann sums for $x(t)$ and I are

$$R_{x,I} = \sum_{k=1}^{N} r_k(t_k - t_{k-1}). \tag{3.110a}$$

$$S_{x,I} = \sum_{k=1}^{N} s_k(t_k - t_{k-1}). \tag{3.110b}$$

The Riemann integral is defined by

$$\int_a^b x(t)\, dt = \lim_{\Delta_I \to 0} R_{x,I}, \tag{3.110c}$$

where $\Delta_I = \max\{t_k - t_{k-1} \mid k = 1, 2, \cdots, N\}$.

Calculus proves that limit (3.110c) remains the same whether we use upper or lower Riemann sums. The height of the rectangle may indeed be the function value at any point within the domain interval. We have chosen to define the Riemann integral using the extreme cases, because, in fact, the modern Lebesgue integral uses sums from below similar to (3.110c).

The base of the rectangle or trapezoid is, of course, an interval. So the Riemann integral partitions the function *domain* and lets that partition determine the range values used for computing little areas. The insight of modern integration is this: Partition the *range*, not the domain, and then look at the sets in the function's domain that it maps to the range regions. There is a way to measure the area of these domain sets (next section), and then their area is weighted by the range values in much the same way as the Riemann integral. The difference seems simple. The difference seems inconsequential. But the implications are enormous.

3.4.1 Measure Theory

The standard approach to the Lebesgue integral is to develop a preliminary theory of the *measure* of a set. This generalizes the notion of simple interval length to a much wider class of sets. Although the Lebesgue integral can be defined without first building a foundation in measure theory (cf. Refs. 13 and 25), the idea of a measure is not difficult. Just as the Riemann integral is a limit of sums of areas, which are interval widths wide and function values high, the Lebesgue integral is a limit of sums of weighted measures—set measures scaled by function values.

Measurable sets, however, can be much more intricate than simple intervals. For example, the rational numbers $\mathbb{Q}$ is a measurable set, and its measure, or its *area* if you will, is zero. Furthermore, any countable set (i.e., a set that can be put into a one–one correspondence with the natural numbers $\mathbb{N}$) is measurable and has zero measure. The interval [0, 1] has unit measure, which is no doubt reassuring, and if we remove all the rational points from it, obtaining $[0, 1] \setminus \mathbb{Q}$, then the result is still measurable and still has unit measure. The rest of this section sketches the developments upon which these appealing ideas can be justified.

3.4.1.1 Rudiments of the Theory. Measure theory axiomatics are closely related to the ideas of a probability measure, which we covered in Chapter 1. We recall therefrom the concept of a σ-algebra Σ. These can be defined for abstract spaces, but we shall stick to sets of real or complex numbers, since these are the spaces for which we define analog signals. Let $\mathbb{K}$ be either $\mathbb{R}$ or $\mathbb{C}$. The four properties of Σ are:

(i) Elements of Σ are subsets of $\mathbb{R}$ (or $\mathbb{C}$): $\wp(\mathbb{K}) \supset \Sigma$.
(ii) The entire space is in Σ: $\mathbb{K} \in \Sigma$.
(iii) Closure under complement: If $A \in \Sigma$, then $\overline{A} = \{t \in \mathbb{K} \mid t \notin A\} \in \Sigma$.
(iv) Closure under countable union: If $\Sigma \supset \{A_n \mid n \in \mathbb{N}\}$, then $\bigcup_{n \in \mathbb{N}} A_n \in \Sigma$.

Examples. A couple of extreme examples of σ-algebras are as follows:

- $\Sigma_1 = \wp(\mathbb{R})$, the set of all subsets of $\mathbb{R}$.
- $\Sigma_0 = \{\varnothing, \mathbb{R}\}$.

There are equivalent examples for $\mathbb{C}$. We have a more interesting σ-algebra in mind, the Borel sets, and we will cover this class momentarily. Combining the closure under countable unions and complement rules, we can show (exercise) that σ-algebras are closed under countable intersections. Two basic concepts are that of a *measurable function*, which we interpret as a measurable signal, and of a *measure* itself, which is a map from a σ-algebra to the non-negative reals.

Definition (Measurable Function). A real- or complex-valued function $x(t)$ is measurable with respect to a σ-algebra Σ if $x^{-1}(A) = \{t \in \mathrm{R} \mid x(t) \in A\} \in \Sigma$ for all open sets A in $\mathbb{K}$.

Proposition. Let Σ be a σ-algebra and let $x : \mathbb{R} \to \mathbb{K}$ be a real- or complex-valued function. Let $\Theta = \{T \in \wp(\mathbb{K}) \mid x^{-1}(T) \in \Sigma\}$. Then Θ is a σ-algebra in $\mathbb{K}$.

Proof: Clearly, $\varnothing$ and $\mathbb{K} \in \Theta$. If $T \in \Theta$, then $S = x^{-1}(T) \in \Sigma$ and $\mathbb{R} \setminus S \in \Sigma$. But $\overline{T} = x(\mathbb{R} \setminus S)$, so $\mathbb{T} \in \Theta$. Finally, $x^{-1}(\cup_{n\in \mathbb{N}} T_n) = \cup_{n\in N} x^{-1}(T_n)$, so closure under countable unions holds as well. ■

The properties of a measure and limit theorems for the modern integral depend on an extension of the real numbers to include two infinite values: ∞ and $-\infty$. Formally, these are just symbols. Intuitively, however, ∞ is an abstract positive value that is larger than any than any real number and $-\infty$ is an abstract negative value that has larger magnitude than any real number. Let us first consider ∞. This value's arithmetic operations are limited. For example, $r + \infty = \infty$ for any $r \in$ R; $r \times \infty = \infty$ for any $r > 0$; if $r = 0$, then $r \times \infty = 0$; and addition and multiplication with ∞ are commutative. There is also a negative infinity element, $-\infty$, so that $r + (-\infty) = -\infty$ for any $r \in \mathbb{R}$. Furthermore, $r \times (-\infty) = -\infty$ for any $r > 0$; if $r = 0$, then $r \times (-\infty) = 0$; if $r < 0$, then $r \times \infty = -\infty$; and so on.

Note that subtraction, division, and cancellation operations only work with finite values. That is, $r - s$ is not defined if both r and s are infinite, and a similar restriction applies to r/s. If $rs = rt$, then we can conclude $s = t$ only if r is finite. A similar restriction applies to $r + s = r + t$.

We can also consider the extended real line, $\mathbb{R}^+ = \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$. We consider $\mathbb{R}^+$ as having the additional open sets $(r, \infty]$ and $[-\infty, r)$ for any finite $r \in \mathbb{R}$. Of course, countable unions of open sets are open in the extended reals. Analog signals can be extended so that they take on infinite values at their singularities. Thus, a signal like $x(t) = t^{-2}$ is undefined at $t = 0$. In the extended reals, however, we may set $x(0) = \infty$. This makes the later limit theorems on modern integration (Section 3.4.3) into equalities. Note that we do not use the extended reals as the domain for analog signals. We only use the extension $\mathbb{R} \cup \{\infty\} \cup \{-\infty\}$ for defining the idea of a measure on σ-algebras and for extending the range of analog signals.

Definition (Measure). A *measure* on a σ-algebra Σ is a function $\mu : \Sigma \to [0, \infty]$ such that

$$\mu\left(\bigcup_{n=-\infty}^{\infty} A_n \right) = \sum_{n=-\infty}^{\infty} \mu(A_n) \tag{3.111}$$

whenever $\{A_n\}$ are pairwise disjoint.

Thus a measure is just like a probability measure, except that its values range in $[0, \infty]$ rather than in $[0, 1]$. A measure function gives a size value for a set. Thus, the measure might indicate the relative size of part of a signal's domain. We are also limiting the discussion to the real line, even though the ideas generalize to measures on σ-algebras in abstract metric spaces. Real analysis texts formalize these notions [24, 44, 45], but we prefer to limit the scope to just what we need for analog signal theory. Here are some easy examples.

Example (All or Nothing Measure). Let $\Sigma = \wp(\mathbb{R})$, and for $A \in \Sigma$ define $\mu(\varnothing) = 0$ and $\mu(A) = \infty$ if $A \neq \varnothing$. Then μ is a measure on Σ.

Example (Counting Measure). Again let $\Sigma = \wp(\mathbb{R})$, and for $A \in \Sigma$ define $\mu(A) = N$ if A contains exactly N elements and $\mu(A) = \infty$ otherwise. Then μ is a measure on Σ.

Proposition. Let μ be a measure on the σ-algebra Σ. Then

(i) (*Null set*) $\mu(\varnothing) = 0$.
(ii) (*Additivity*) If $A_p \cap A_q = \varnothing$ when $p \neq q$, then $\mu(A_1 \cup A_2 \cup \cdots \cup A_n) = \mu(A_1) + \mu(A_2) + \cdots + \mu(A_n)$.
(iii) (*Monotonicity*) If $B \supset A$, then $\mu(B) \geq \mu(A)$.

Proof: Similar to probability measure arguments. ■

3.4.1.2 Lebesgue Measurable Sets. There are lots of σ-algebras on the real line. Analog signal theory needs only the smallest σ-algebra that contains all the open sets in $\mathbb{R}$. We must show that such a smallest σ-algebra exists.

Theorem. There is a σ-algebra $\mathcal{B}$ on $\mathbb{R}$ such that:

(i) $\mathcal{B}$ contains all open subsets of $\mathbb{R}$.
(ii) If Σ is a σ-algebra containing all the open sets, then $\Sigma \supset \mathcal{B}$.

Proof: To begin with, the power set on the real line, $\wp(\mathbb{R})$, is itself is a σ-algebra and contains all open sets. Nonconstructively, we set $\mathcal{B}$ to be the intersection of all such σ-algebras. It is straightforward to show that this intersection is still a σ-algera. It is still a subset of $\wp(\mathbb{R})$. Since $\mathbb{R}$ must be in every σ-algebra, it is must be in the intersection of those containing the open sets. Closure under complement is also easy: if $A \in \Sigma$, where Σ is a any σ-algebra containing the open sets, then $\overline{A} \in \Sigma$; thus, $\overline{A}$ is in the intersection of all such σ-algebras. Finally, let $\mathcal{B} \supset \{A_n \mid n \in \mathbb{N}\}$. Then for all $n \in \mathbb{N}$, A_n is in every σ-algebra Σ that contains all the open sets in $\mathbb{R}$. Thus the countable family $\{A_n\}$ is a subset of every such σ-algebra. Hence $\bigcup_{n \in \mathrm{N}} A_n$ is in each of these σ-algebras; and, consequently, $\bigcup_{n \in \mathrm{N}} A_n$ is in the intersection $\mathcal{B}$. ■

Definition (Borel sets). The class of *Borel* or *Lebesgue measurable* sets is the smallest σ-algebra that contains every open set.

All the sets we normally use in signal theory are Borel sets. In fact, it takes a certain amount of craftiness to exhibit a set that is not Lebesgue measurable.

Example (Intervals). All of the open and closed sets, and the intervals (a, b) and $[a, b]$ in particular, are Lebesgue measurable. That closed sets are measurable follows from the σ-algebra's complement property. Also, since we take the half-infinite intervals $(a, \infty]$ and $[-\infty, a)$ to be open, these too are measurable. Finally, we can form countable unions involving these basic measurable sets. In the complex plane, open disks $\{z: |z| < r\}$, and closed disks $\{z: |z| \le r\}$ are measurable as are half-infinite sets $\{z: |z| > r\}$, and so on.

Example (Countable Sets). Any countable set is measurable. For example, real singletons $\{a\}$ are measurable because they are closed. So any countable union of singletons is measurable.

Proposition. If $x(t)$ is measurable and T is a Lebesgue measurable set in $\mathbb{K}$ ($\mathbb{R}$ or $\mathbb{C}$), then $x^{-1}(T) \in \mathcal{B}$; in other words, $x^{-1}(T)$ is Lebesgue measurable in $\mathbb{R}$.

Proof: Let $\Theta = \{T \in \wp(\mathbb{K}) \mid x^{-1}(T) \in \mathcal{B}\}$. Since $x(t)$ is measurable, Θ contains all the open sets in $\mathbb{K}$. In the previous section, we showed that Θ is a σ-algebra, so it must contain all the Lebesgue measurable sets. ■

3.4.1.3 Lebesgue Measure. There are lots of possible measures on σ-algebras in $\mathbb{R}$ (or $\mathbb{C}$). Again, we need only one of them: the *Lebesgue measure*. It applies to the Lebesgue measurable, or Borel sets.

Definition (Open Covering). Let S be a set and $O = \{A_n \mid n \in \mathbb{N}\}$ be a family of open sets. If $S \subset \bigcup_{n \in \mathrm{N}} A_n$, then O is an *open covering* for S.

Definition (Real Lebesgue Measure). Let $\mathcal{B}$ be the Lebesgue measurable sets on $\mathbb{R}$ and let the measure function μ be defined as follows:

(i) $\mu(a, b) = b - a$.
(ii) If S is an open set in R, then

$$\mu(S) = \inf_{S \subset \cup A_n} \left[\sum_{n=0}^{\infty} \mu(A_n) \right], \tag{3.112}$$

where the greatest lower bound is taken over all open coverings of S by intervals. The function μ is called the (real) *Lebesgue measure*.

Definition (Complex Lebesgue Measure). Let $\mathcal{B}$ be the Lebesgue measurable sets on $\mathbb{C}$. Let the measure function μ be defined as follows:

(i) If $B = \{z \in \mathbb{C} \mid |z - c| < r\}$, then $\mu(B) = \pi \mathrm{r}^2$.

(ii) If S is an open set in $\mathbb{C}$, then

$$\mu(S) = \inf_{S \subset \bigcup B_n} \left[\sum_{n=0}^{\infty} \mu(B_n) \right], \tag{3.113}$$

where the greatest lower bound is taken over all open coverings of S by open balls. The function μ is called the (complex) *Lebesgue measure*.

Accepting that these definitions do produce functions that are indeed measures on the Borel sets, let us provide some examples of the measures of sets.

Example (Singletons). The measure of a singleton $\{a\}$ is zero. Let $\varepsilon > 0$. Then the single interval $I_\varepsilon = (a - \varepsilon, a + \varepsilon)$ covers $\{a\}$. The Lebesgue measure of I_ε is 2ε. Since ε was arbitrary and positive, the greatest lower bound of the lengths of all such intervals cannot be positive, so $\mu\{a\} = 0$.

Example (Intervals). The measure of a half-open interval $[a, b)$ is $b - a$. The measure of (a, b) is $b - a$, and the singleton $\{a\}$ has measure zero. Because (a, b) and $\{a\}$ are disjoint, $\mu(a, b) + \mu\{a\} = \mu[a, b) = b - a$. We also have $\mu(a, \infty] = \infty$; this is a consequence of the monotonicity property of a measure, since there are infinitely many disjoint intervals of unit length that are contained in a half-infinite interval.

Example (Countable Sets). Suppose A is a countable set $A = \{a_n \mid n \in \mathbb{N}\}$. Then for each $\varepsilon > 0$ we can find a set of intervals that covers A such that the sum of the lengths of the intervals is ε. For example, let $I_{n,\varepsilon} = (a_n - \varepsilon 2^{-n-2}, a_n + \varepsilon 2^{-n-2})$. Since ε is arbitrary, $\mu A = 0$.

Definition (Almost Everywhere). A property is said to hold *almost everywhere* on a measurable set A if the set of elements of A upon which is does not hold has Lebesgue measure zero.

Example (Nonmeasurable Set). To show that there are non-Lebesgue measurable sets, we recapitulate the example from Ref. 24. Consider the half-open unit interval $I = [0, 1)$, for which $\mu I = 1$. For $a, b \in I$, define $a \oplus b$ by

$$a \oplus b = \begin{cases} a + b & \text{if } a + b < 1, \\ a + b - 1 & \text{if } a + b \geq 1. \end{cases} \tag{3.114}$$

We can easily see that Lebesgue measure is translation invariant: If $\mu S = r$, then $\mu(S + a) = \mu\{s + a \mid s \in S\} = r$ for any $a \in \mathbb{R}$. Similarly, if we define $S \oplus a = \{s \oplus a \mid s \in S\}$, then $\mu(S \oplus a) = \mu(S)$. Now define an equivalence relation $a \sim b$ on I to mean $a - b \in \mathbb{Q}$, the rational numbers. Let $[a] = \{b \in I \mid a \sim b\}$ be the equivalence class of any $a \in I$ and $K = \{[a] \mid a \in I\}$. Define the set C to contain exactly one element from each equivalence class in K. Set theory's Axiom of Choice [46] ensures that set C exists. Since $\mathbb{Q}$ is countable, we can index $\mathbb{Q} \cap I$ by $\{q_n \mid n \in \mathbb{N}\}$, with $q_0 = 0$. We set $C_n = C \oplus q_n$. The properties of the C_n are as follows:

(i) $C_0 = C$.

(ii) The C_n are disjoint; for if $r \in C_m \cap C_n$, then $r = c + q_m = d + q_n$ for some c, $d \in C$; so $c \sim d$, and since C was a choice set containing exactly one element from disjoint equivalence classes in K, we must have $m = n$.

(iii) $I = \bigcup_{n \in \mathbb{N}} C_n$; if $r \in I$, then there is an $[a] \in K$ with $r \in [a]$ and $a \in C$; this implies $r \sim a$ or $r - a \in \mathbb{Q} \cap I$; but $\{q_n\}$ indexes such rational numbers, so $r - a = q_n$ for some $n \in \mathbb{N}$; thus, $r = a + q_n \in C_n$.

If C is Lebesgue measurable, then by the properties of the Lebesgue measure under translations, for all $n \in \mathbb{N}$, C_n is measurable and $\mu C = \mu C_n$.

Thus,

$$\mu(I) = \mu(\bigcup_{n \in \mathbb{N}} C_n) = \sum_{n=0}^{\infty} \mu(C_n) = \sum_{n=0}^{\infty} \mu C = \begin{cases} \infty & \text{if } \mu C > 0 \\ 0 & \text{if } \mu C = 0 \end{cases} \tag{3.115}$$

However, we know $\mu I = 1$, so that (3.115) is a contradiction; it must be the case that the choice set C is not Lebesgue measurable.

Our intuition might well suggest an easy generalization of the idea of an interval's length to a length measure for any subset of the real line. This last example has shown that the task demands some care. We cannot have the proposed properties of a measure and still be able to measure the size (length or area) of all sets. Some functions must be outside our theory of analog signals. The characteristic function for the choice set C in the above example is a case in point. Nevertheless, the class of measurable sets is quite large, and so too is the class of measurable functions. Let us turn to integration of measurable functions and apply the modern integral to signal theory. Here we shall see how these mathematical tools sharpen the definitions of the basic analog signal spaces. Moreover, these concepts will support our later development of signal approximation and transform techniques.

3.4.2 Lebesgue Integration

The key distinction between the Lebesgue and Riemann integrals is that the modern integral partitions the range of function values, whereas the classic integral partitions the domain of the function. A seemingly inconsequential difference at first glance, this insight is critical.

To illustrate the idea, consider the task of counting the supply of canned goods on a cabinet shelf. (Perhaps the reader lives in a seismically active region, such as this book's Californian authors, and is assessing household earthquake preparedness.) One way to do this is to iterate over the shelf, left to right, front to back, adding up the volume of each canned food item. No doubt, most people would count cans like this. And it is just how we Riemann integrate $x(t)$ by adding areas: for each interval $I = [a, b)$ in the domain partition, accumulate the area, say $x(a) \times (b - a)$; refine the partition; find a better Riemann sum; and continue to the limit. The other way to guage the stock of canned food is to first count all the cans of one size, say the small tins of tuna. Then proceed to the medium size soup cans. Next the large canned vegetables. Finally—and perhaps exhausting the reader's patience—we count the giant fruit juices. It might seem silly, but it works.

And this is also Lebesgue's insight for defining the integral. We slice up the range of $x(t)$, forming sums with the weighted measures $x(a) \times \mu[x^{-1}\{a\}]$, where $x^{-1}\{a\} = \{t \in \mathbb{R} \mid x(t) = a\}$ and μ is some measure of the size of $x^{-1}\{a\}$. Our remaining exposition omits many abstractions and details; mathematical treatises that excel in this area are readily available. Instead we seek a clear and simple statement of how the ideas on measure and measurable functions fit together to support the modern integral.

3.4.2.1 Simple Functions. Like the Riemann integral, the Lebesgue integral is also a limit. It is the limit of integrals of so-called simple functions.

Definition (Simple Function). If $x : \mathbb{R} \to \mathbb{R}$ is Lebesgue measurable and Range(x) is finite, then $x(t)$ is called a simple function.

Every Lebesgue measurable function can be approximated from below by such simple functions [44]. This theorem's proof uses the stereotypical Lebesgue integration technique of partitioning the range of a measurable function.

Theorem (Approximation by Simple Functions). If $x(t)$ is measurable, then there is a sequence of simple functions $s_n(t)$ such that:

(i) If $n \le m$, then $|s_n(t)| \le |s_m(t)|$.

(ii) $|s_n(\mathrm{t})| \le |x(t)|$.

(iii) $\lim_{n\to\infty} s_n(t) = x(t)$.

Proof: Since we can split $x(t)$ into its negative and non-negative parts, and these are still measurable, we may assume that $x(t) \in [0, \infty]$ for all $t \in \mathbb{R}$. The idea is to break Range(x) into two parts: $S_n = [0, n)$ and $T_n = [n, \infty]$ for each $n > 0$. We further subdivide S_n into $n2^n$ subintervals of length 2^{-n}: $S_{m,n} = [m2^{-n}, (m + 1)2^{-n})$, for $0 \le m < n2^n$. Set $A_n = x^{-1}(S_n)$ and $B_n = x^{-1}(T_n)$. Define

$$s_n(t) = n\chi_{B_n}(t) + \sum_{m=0}^{n2^n-1} \frac{m}{2^n}\chi_{B_n}(t), \tag{3.116}$$

where χ_S is the characteristic function on the set S. The simple functions (3.116) satisfy the conditions (i)–(iii). ■

3.4.2.2 Definition and Basic Properties. We define the modern integral in increments: first for non-negative functions, then for functions that go negative, and finally for complex-valued functions.

Definition (Lebesgue Integral, Non-negative Functions). Let μ the Lebesgue measure, let A be a Lebesgue measurable set in $\mathbb{R}$, and let $x : \mathbb{R} \to [0, \infty]$ be a measurable function. Then the Lebesgue integral with respect to μ of $x(t)$ over A is

$$\int_A x(t)\, d\mu = \sup_{0 \le s(t) \le x(t)} \left\{ \int_A s(t)\, d\mu \right\}, \tag{3.117}$$

where the functions $s(t)$ used to take the least upper bound in (3.117) are all simple.

Definition (Lebesgue Integral, Real-Valued Functions). Let μ the Lebesgue measure, let A be a Lebesgue measurable set in $\mathbb{R}$, and let $x : \mathbb{R} \to \mathbb{R}$ be a measurable function. Furthermore, let $x(t) = p(t) - n(t)$, where $p(t) > 0$ and $n(t) > 0$ for all $t \in \mathbb{R}$. Then the Lebesgue integral with respect to μ of $x(t)$ over A is

$$\int_A x(t)\, d\mu = \int_A p(t)\, d\mu - \int_A n(t)\, d\mu \tag{3.118}$$

as long as one of the integrals on the right-hand side of (3.118) is finite.

Remarks. By the elementary properties of Lebesgue measurable functions, the positive, negative, and zero parts of $x(t)$ are measurable functions. Note that the zero part of $x(t)$ does not contribute to the integral. In general, definitions and properties of the Lebesgue integral must assume that the subtractions of extended reals make sense, such as in (3.118). This assumption is implicit in what follows.

Definition (Lebesgue Integral, Complex-Valued Functions). Let μ the Lebesgue measure; let A be a Lebesgue measurable set in $\mathbb{R}$, let x: $\mathbb{R} \to \mathbb{C}$ be a measurable function, and let $x(t) = x_r(t) + jx_i(t)$, where $x_r(t)$ and $x_i(t)$ are real-valued for all $t \in \mathbb{R}$. Then the Lebesgue integral with respect to μ of $x(t)$ over A is

$$\int_A x(t)\, d\mu = \int_A x_r(t)\, d\mu + j\int_A x_i(t)\, d\mu . \tag{3.119}$$

The modern Lebesgue integral obeys all the rules one expects of an integral. It also agrees with the classic Riemann integral on piecewise continuous functions. Finally, it has superior limit operation properties.

Proposition (Linearity). Let μ the Lebesgue measure, let A be a Lebesgue measurable set in $\mathbb{R}$, and let $x, y : \mathbb{R} \to \mathbb{C}$ be measurable functions. Then,

(i) (Scaling) For any c $\in \mathbb{C}$, $\int_A cx(t)\, d\mu = c\int_A x(t)\, d\mu$.

(ii) (Superposition) $\int_A [x(t) + y(t)]\, d\mu = \int_A x(t)\, d\mu + \int_A y(t)\, d\mu$.

Proof: The integral of simple functions and the supremum are linear. ■

Remark. If $x(t) = 0$ for all $t \in \mathbb{R}$, then (even if $\mu A = \infty$) $\int_A x(t)\, d\mu = 0$ by (i).

Proposition. Let μ be the Lebesgue measure, let $[a, b]$ be an interval on $\mathbb{R}$, and let x: $\mathbb{R} \to \mathbb{C}$ be a piecewise continuous function. Then the Lebesgue and Riemann integrals of $x(t)$ are identical:

$$\int_{[a,b]} x(t)\, d\mu = \int_a^b x(t)\, dt. \tag{3.120}$$

Proof: $x(t)$ is both Riemann integrable and Lebesgue integrable. The Riemann integral, computed as a limit of lower rectangular Riemann sums (3.110a), is precisely a limit of simple function integrals. ■

Proposition (Domain Properties). Let μ be the Lebesgue measure; let A, B be Lebesgue measurable sets in $\mathbb{R}$; and let x: $\mathbb{R} \to \mathbb{C}$ be a measurable function. Then,

(i) (Subset) If $B \supset A$ and $x(t)$ is non-negative, then $\int_B x(t)\, d\mu \geq \int_A x(t)\, d\mu$.

(ii) (Union) $\int_{A \cup B} x(t)\, d\mu = \int_A x(t)\, d\mu + \int_B x(t)\, d\mu - \int_{A \cap B} x(t)\, d\mu$.

(iii) (Measure Zero Set) If $\mu(A) = 0$, then $\int_A x(t)\, d\mu = 0$.

Remark. Note that $\int_A x(t)\, d\mu = 0$ in (iii) even if $x(t) = \infty$ for all $t \in \mathbb{R}$.

Proposition (Integrand Properties). Let μ the Lebesgue measure; let A be Lebesgue measurable in $\mathbb{R}$; let $\chi_A(t)$ be the characteristic function on A; and let x, y: $\mathbb{R} \to \mathbb{C}$ be measurable functions. Then,

(i) (Monotonicity) If $y(t) \geq x(t) \geq 0$ for all $t \in \mathbb{R}$, then $\int_A y(t)\, d\mu \geq \int_A x(t)\, d\mu$.

(ii) (Characteristic Function) $\int_A x(t)\, d\mu = \int_{\mathbb{R}} x(t)\chi_A(t)\, d\mu$.

Proof: The proofs of these propositions follows from the definition of simple functions and integrals as limits thereof. ■

3.4.2.3 Limit Operations with Lebesgue's Integral. The modern integral supports much more powerful limit operations than does the Riemann integral. We recall that sequence of functions can converge to a limit that is not Riemann integrable. In order to simplify the discussion, we offer the following theorems without detailed proofs; the interested reader can find them in treatises on modern analysis, [13, 24, 25, 44, 45].

Theorem (Monotone Convergence). Let μ be the Lebesgue measure, let A be a measurable set in $\mathbb{R}$, and x_n: $\mathbb{R} \to \mathbb{R}$ be Lebesgue measurable for $n \in \mathbb{N}$. If $0 \leq x_n(t) \leq x_m(t)$ for $n < m$ and $\lim_{n\to\infty} x_n(t) = x(t)$ for all $t \in A$, then $x(t)$ is measurable and

$$\lim_{n \to \infty} \int_A x_n(t)\, d\mu = \int_A x(t)\, d\mu . \tag{3.121}$$

Proof: $x(t)$ is measurable, because for any $r \in \mathbb{R}$,

$$x^{-1}(r, \infty] = \bigcup_{n=0}^{\infty} x_n^{-1}(r, \infty] . \tag{3.122}$$

Hence, $\lim_{n \to \infty} \int_A x_n(t)\, d\mu \leq \int_A x(t)\, d\mu$. Arguing the inequality the other way [44] requires that we consider simple functions $s(t)$ such that $0 \leq s(t) \leq x(t)$ for all $t \in \mathbb{R}$. Let $0 < c < 1$ be constant and set $A_n = \{t \in \mathbb{R} | \, cs(t) \leq x_n(t)\}$. Then, $A_{n+1} \supset A_n$ for all $n \in \mathbb{N}$, and $A = \bigcup_{n=0}^{\infty} A_N$. Thus,

$$c \int_{A_n} s(t)\, d\mu \leq \int_A x_n(t)\, d\mu . \tag{3.123}$$

As $n \to \infty$ on the right-hand side of (3.123), we see

$$c \int_A s(t)\, d\mu \leq \lim_{n \to \infty} \int_A x_n(t)\, d\mu , \tag{3.124a}$$

which is true for every $0 < c < 1$. Let $c \to 1$, so that

$$\int_A s(t)\, d\mu \leq \lim_{n \to \infty} \int_A x_n(t)\, d\mu . \tag{3.124b}$$

But $s(t)$ can be any simple function bounding $x(t)$ below, so by the definition of Lebesgue integration we know $\lim_{n \to \infty} \int_A x_n(t)\, d\mu \geq \int_A x(t)\, d\mu$. ■

Corollary. Let μ the Lebesgue measure; let A be a measurable set in $\mathbb{R}$; let x_n: $\mathbb{R} \to \mathbb{R}$ be Lebesgue measurable for $n \in \mathbb{N}$; and, for all $t \in A$, suppose $\lim_{n\to\infty} x_n(t) = x(t)$. Then, $x(t)$ is measurable and

$$\lim_{n \to \infty} \int_A x_n(t)\, d\mu = \int_A x(t)\, d\mu . \tag{3.125}$$

Proof: Split $x(t)$ into negative and positive parts. ■

A similar result holds for complex-valued functions. The next corollary shows that we may interchange Lebesgue integration and series summation.

Corollary (Integral of Series). Let μ the Lebesgue measure; let A be a measurable set in $\mathbb{R}$; let x_n: $\mathbb{R} \to \mathbb{C}$ be Lebesgue measurable for $n \in \mathbb{N}$; and, for all $t \in A$, suppose $\Sigma_{n\to\infty} x_n(t) = x(t)$. Then,

$$\sum_{n=0}^{\infty} \int_A x_n(t)\, d\mu = \int_A x(t)\, d\mu . \tag{3.126}$$

Proof: Beginning with non-negative functions, apply the theorem to the partial sums in (3.126). Then extend the result to general real- and complex-valued functions. ■

The next theorem, Fatou's lemma,[5] relies on the idea of the lower limit of a sequence [23].

Definition (lim inf, lim sup). Let $A = \{a_n \mid n \in \mathbb{N}\}$ be a set of real numbers and let $A_N = \{a_n \mid n \geq \mathrm{N}\}$. Let $r_N = \inf A_N$ be the greatest lower bound of A_N in the extended real numbers $\mathbb{R} \cup \{\infty\} \cup \{-\infty\}$. Let $s_N = \sup A_N$ be the least upper bound of A_N in the extended real numbers. We define lim inf A and lim sup A by

$$\liminf \{a_n\} = \lim_{N\to\infty} r_N \tag{3.127a}$$

and

$$\limsup \{a_n\} = \lim_{N\to\infty} s_N . \tag{3.127b}$$

The main things anyone has to know are:

- $\liminf\{a_n\}$ is the smallest limit point in the sequence $\{a_n\}$;
- $\limsup\{a_n\}$ is the largest limit point in the sequence $\{a_n\}$;
- $\liminf\{a_n\} = \limsup\{a_n\}$ if and only if the sequence $\{a_n\}$ converges to some limit value $a = \lim_{n\to\infty}\{a_n\}$, in which case $a = \liminf\{a_n\} = \limsup\{a_n\}$;
- the upper and lower limits could be infinite, but they always exist;
- if $a_n \leq b_n$ then $\liminf\{a_n\} \leq \liminf\{b_n\}$;
- $\limsup\{-a_n\} = -\liminf\{a_n\}$.

Example. Consider the sequence $\{a_n\} = \{1^{-1}, 1, -1, 2^{-1}, 2, -2, 3^{-1}, 3, -3, ...\}$. This sequence has three limit points: $-\infty$, ∞, and 0. We have $\liminf\{a_n\} = -\infty$, $\limsup\{a_n\} = +\infty$, and $\lim_{n\to\infty}\{a_{3n}\} = 0$. Only this last sequence is a genuine Cauchy sequence, however.

Theorem (Fatou's Lemma). Let μ the Lebesgue measure; let A be a measurable set in $\mathbb{R}$; let x_n: $\mathbb{R} \to \mathbb{R}$ be Lebesgue measurable for $n \in \mathbb{N}$; and, for all $n \in \mathbb{N}$ and $t \in A$, $0 \leq x_n(t) \leq \infty$. Then,

$$\int_A \liminf[x_n(t)]\, d\mu \leq \liminf \int_A x_n(t)\, d\mu . \tag{3.128}$$

[5]Pierre Fatou (1878–1929), mathematician and astronomer at the Paris Observatory.

Proof: Define

$$y_n(t) = \inf_{k \ge n} \{x_k(t)\}, \tag{3.129}$$

so that the $y_n(t)$ approach $\liminf x_n(t)$ as $n \to \infty$ and $y_n(t) \le y_m(t)$ when $n < m$. Note that for all $n \in \mathbb{N}$, $y_n(t) \le x_n(t)$. Consequently,

$$\int_A y_n(t)\, d\mu \le \int_A x_n(t)\, d\mu, \tag{3.130}$$

and thus,

$$\liminf \int_A y_n(t)\, d\mu \le \liminf \int_A x_n(t)\, d\mu. \tag{3.131}$$

The left-hand side of (3.131) draws our attention. Since $\lim\{y_n(t)\} = \liminf \{x_n(t)\}$ and $\{y_n(t)\}$ are monotone increasing, Lebesgue's monotone convergence theorem implies

$$\liminf \int_A y_n(t)\, d\mu = \lim \int_A y_n(t)\, d\mu = \int_A \lim[y_n(t)]\, d\mu = \int_A \liminf[x_n(t)]\, d\mu. \tag{3.132}$$

Combining (3.131) and (3.132) completes the proof. ■

Theorem (Lebesgue's Dominated Convergence). Let μ be the Lebesgue measure, let A be a measurable set in $\mathbb{R}$, and let $x_n: \mathbb{R} \to \mathbb{C}$ be Lebesgue measurable for $n \in \mathbb{N}$, $\lim_{n\to\infty} x_n(t) = x(t)$ for all $t \in A$, and $|x_n(t)| \le g(t) \in L^1(\mathbb{R})$. Then $x(t) \in L^1(\mathbb{R})$ and

$$\lim_{n \to \infty} \int_A x_n(t)\, d\mu = \int_A x(t)\, d\mu. \tag{3.133}$$

Proof: We borrow the proof from Ref. 44. Note first that limit $x(t)$ is a measurable function and it is dominated by $g(t)$, so $x(t) \in L^1(\mathbb{R})$. Next, we have $|x(t) - x_n(t)| \le 2g(t)$ and we apply Fatou's lemma to the difference $2g(t) - |x_n(t) - x(t)|$:

$$2\int_A g(t) d\mu \le \liminf_{n \to \infty} \int_A \{2g(t) - |x(t) - x_n(t)|\} d\mu. \tag{3.134a}$$

Manipulating the lower limit on the right-hand integral in (3.134a) gives

$$2\int_A g(t)\, d\mu \le 2\int_A g(t)\, d\mu - \limsup_{n \to \infty} \int_A |x(t) - x_n(t)|\, d\mu. \tag{3.134b}$$

Subtracting $2\int_A g\, d\mu$ out of (3.134b), we have $\lim_{n\to\infty} \sup \int_A |x(t) - x_n(t)|\, d\mu \le 0$ from which $\lim_{n\to\infty} \int_A |x(t) - x_n(t)|\, d\mu = 0$ and (3.133) follows. ■

Corollary (Interchange of Limits). Let μ the Lebesgue measure, let A be a measurable set in $\mathbb{R}$, and let x_n: $\mathbb{R} \to \mathbb{C}$ be Lebesgue measurable for $n \in \mathbb{N}$. Suppose that

$$\sum_{n=-\infty}^{\infty} \int_A |x_n(t)| \, d\mu < \infty . \tag{3.135}$$

Then the series $\Sigma_n x_n(t) = x(t)$ for almost all $t \in A$, $x(t) \in L^1(\mathbb{R})$, and

$$\sum_{n=-\infty}^{\infty} \int_A x_n(t) \, d\mu = \int_A x(t) \, d\mu . \tag{3.136}$$

Proof: Apply the dominated convergence theorem to partial series sums. ■

3.4.2.4 Lebesgue Integrals in Signal Theory. To this point, our definitions of L^p signal spaces were defined abstractly as completions of more rudimentary spaces. The new integral helps avoid such stilted formulations. We define both the L^p norm and L^p signals spaces using the Lebesgue integral.

Definition (L^p, L^p norm). Let μ the Lebesgue measure and let A be a measurable set in $\mathbb{R}$. Then $L^p(A)$ is the set of all Lebesgue measurable signals $x(t)$ such that

$$\int_A |x(t)|^p d\mu < \infty . \tag{3.137}$$

We define $\|x\|_{p,A}$ to be

$$\|x\|_{p,A} = \left(\int_A |x(t)|^p d\mu \right)^{\frac{1}{p}} \tag{3.138}$$

when the integral (3.137) exists. In the case of $p = \infty$, we take $L^\infty(A)$ to be the set of all $x(t)$ for which there exists M_x with $|x(t)| < M_x$ almost everywhere on A.

Now we can recast the entire theory of L^p spaces using the modern integral. We must still identify L^p space elements with the equivalence class of all functions that differ only by a set of measure zero. The basic inequalities of Holder, Minkowski, and Schwarz still hold. For instance, Minkowski's inequality states that $\|x + y\|_p \le \|x\|_p + \|y\|_p$, where the p-norm is defined by Lebesgue integral. The more powerful limit theorems of the modern integral, however, allow us to prove the following completeness result [25, 44, 45].

Theorem. For Lebesgue measurable A, the $L^p(A)$ spaces are complete, $1 \le p \le \infty$.

Proof: We leave the case $p = \infty$ as an exercise. Let $\{x_n(t)\}$ be Cauchy in $L^p(A)$ We can extract a subsequence $\{y_n(t)\}$ of $\{x_n(t)\}$ with $\|y_{n+1} - y_n\|_p < 2^{-n}$ for all n. We then define $f_n(t) = |y_1(t) - y_0(t)| + |y_2(t) - y_1(t)| + \cdots + |y_{n+1}(t) - y_n(t)|$ and $f(t) = \lim_{n\to\infty} f_n(t)$.

The choice of the subsequence and Minkowski's inequality together imply $\|f_n(t)\|_p < 1$. Invoking Fatou's lemma (3.128) on $\{[f_n(t)]^p\}$, we obtain

$$\int_A [f(t)]^p d\mu = \int_A \liminf [f_n(t)]^p d\mu \le \liminf \int_A [f_n(t)]^p d\mu \le \liminf \{1^p\} = 1. \tag{3.139}$$

This also shows $f(t) < \infty$ almost everywhere on A. If $x(t)$ is given by

$$x(t) = y_0(t) + \sum_{n=0}^{\infty} (y_{n+1}(t) - y_n(t)), \tag{3.140}$$

then the convergence of $f(t)$ guarantees that $x(t)$ converges absolutely almost everywhere on A. Thus, $x(t) = \lim_{n\to\infty} y_n(t)$, the Cauchy sequence $\{x_n(t)\}$ has a convergent subsequence, and so $\{x_n(t)\}$ must have the same limit. ■

The next result concerns two-dimensional integrals. These occur often, even in one-dimensional signal theory. The Fubini theorem[6] provides conditions under which iterated one-dimensional integrals—which arise when we apply successive integral operators—are equal to the associated two-dimensional integral [24]. Functions defined on $\mathbb{R} \times \mathbb{R}$ are really two-dimensional signals—*analog images*—and generally outside the scope of this book. However, in later chapters our signal transform operations will mutate a one-dimensional signal into a two-dimensional transform representation. Opportunties to apply the following result will abound.

Theorem (Fubini). Let $x(s, t)$ be a measurable function on a measurable subset $A \times B$ of the plane $\mathbb{R}^2$. If either of these conditions obtains,

(i) $x(s, t) \in L^1(A \times B)$,
(ii) $0 \le x(s, t)$ on $A \times B$,

then the order of integration may be interchanged:

$$\int_A \left[\int_B x(s, t)\, dt \right] ds = \int_B \left[\int_A x(s, t)\, ds \right] dt. \tag{3.141}$$

Proof: Refer to Ref. 24.

3.4.2.5 Differentiation. It remains to explain the concept of a derivative in the context of Lebesgue integration. Modern integration seems to do everything backwards. First, it defines the sums for integration in terms of range values rather than domain intervals. Then, unlike conventional calculus courses, it leaves out the intuitively easier differentiation theory until the end. Lastly, as we shall see below, it defines the derivative in terms of an integral.

[6]Guido Fubini (1879–1943) was a mathematics professor at Genoa and Turin until anti-Semitic decrees issued by Mussolini's fascist regime compelled him to retire. Fubini moved to the United States from Italy, taking a position at the Institute for Advanced Study in Princeton, New Jersey in 1939.

Definition (Derivative). Let $y(t)$ be a Lebesgue measurable function. Suppose $x(t)$ is a Lebesgue measurable function such that

$$y(t) = \int_a^t x \, d\mu , \tag{3.142}$$

almost everywhere on any interval $[a, b]$ that contains t. Then we say $x(t)$ is the *derivative* of $y(t)$. With the usual notations, we write $\frac{dy}{dt} = y'(t) = y^{(1)}(t) = x(t)$. Second and higher derivatives may be further defined, and the notations carry through as well.

All of the differentiation properties of conventional calculus check out under this definition.

3.5 DISTRIBUTIONS

Distributions extend the utility of Hilbert space to embrace certain useful quantities which are not classically defined in Riemannian calculus. The most celebrated example of a distribution is the *Dirac delta*, which played a seminal role in the development of quantum mechanics and has been extended to other areas of quantitative science [7–11, 47–49]. The Dirac delta is easy to apply but its development within the context of distributions is often obscured, particularly at the introductory level. This section develops the foundations of distribution theory with emphasis on the Dirac delta. The theory and definitions developed here are also the basis for the generalized Fourier transform of Chapter 6.

3.5.1 From Function to Functional

Quantitative science is concerned with generating numbers and the notion of a function as a mapping from the one set of complex numbers to another is well-established. The inner product in Hilbert space is another tool for generating physically relevant data, and this Hilbert space mapping is conveniently generalized by the concept of a *functional*.

Definition (Functional). Let $\phi(t)$ be a function belonging to the class of so-called *test functions* (to be defined shortly). If the inner product

$$\langle f(t), \phi(t) \rangle \equiv \int_{-\infty}^{\infty} f(t) \phi^*(t) \, dy \tag{3.143}$$

converges, the quantity $f(t)$ is a *functional* on the space of test functions.

The test functions are defined as follows.

Definition (Rapid Descent). A function is said to be *rapidly descending* if for each positive integer N, the product

$$t^N \cdot f(t) \tag{3.144}$$

remains bounded as $|t| \to \infty$. A test function $\phi(t)$ will be classified as *rapidly decreasing* if $\phi(t)$ and all its derivatives are rapidly decreasing.

The derivatives are included in this definition to ensure that the concept of rapid descent is closed under this operation. Exponentials such as e^{-t}, and Gaussians e^{-t^2} are rapidly decreasing. On the other hand, polynomials, the exponential e^t, and $\sin t$, $\cos t$ are not rapidly decreasing. (These will be categorized shortly.) Furthermore, rapidly decreasing functions are integrable.

The condition of rapid descent can be guaranteed for a large class of functions by forcing them vanish identically for all t outside some interval $[t_1, t_2]$, that is,

$$f(t) \to f(t) \cdot [u(t-t_1) - u(t-t_2)]\,. \tag{3.145}$$

These *test functions of compact support* are a subset of all test functions of rapid descent.

Remark. The test functions are our slaves; by stipulating that they decay sufficiently rapidly, we can ensure that (3.143) converges for a sufficiently broad class of functionals $f(t)$. In many discussions, particularly general theoretical treatment of functionals, the exact form of the test function is immaterial; it is merely a vehicle for ensuring that the inner products on the space of $\phi(t)$ converge. Often, all that is required is the knowledge that a test function behaves in a certain way under selected operations, such as differentiation, translation, scaling, and more advanced operations such as the Fourier transform (Chapters 5 and 6). Whenever possible, it is advantageous to work in the space of compact support test functions since that eliminates any questions as to whether its descent is sufficiently rapid; if $\phi(t)$ has compact support, the product $\phi(t)f(t)$ follows suit, admitting a large set of $f(t)$ for which (3.143) generates good data. However, in some advanced applications, such as the generalized Fourier transform (Chapter 6), we do not have the luxury of assuming that all test functions are compactly supported.

3.5.2 From Functional to Distribution

A distribution is a subset of the class of functionals with some additional (and physically reasonable) properties imposed.

3.5.2.1 Defintion and Classification

Definition (Distribution). Let $\phi(t)$ be a test function of rapid descent. A functional $f(t)$ is a distribution if it satisfies conditions of continuity and linearity:

(i) *Continuity.* Functional $f(t)$ is continuous if when the sequence $\phi_k(t)$ converges to zero in the space of test functions, then $\lim_{k \to \infty} \langle f(t), \phi_k(t) \rangle \to 0$.

(ii) *Linearity.* Functional $f(t)$ is linear if for all complex constants c_1 and c_2 and for $\psi(t)$ of rapid descent,

$$\langle f(t), c_1\phi(t) + c_2\psi(t) \rangle = c_1 \langle f(t), \phi(t) \rangle + c_2 \langle f(t), \psi(t) \rangle. \tag{3.146}$$

Linearity plays such a central role to signal analysis that any functional that does not belong to the class of distributions is not useful for our applications. Distributions of all types are sometimes referred to as *generalized functions.*

Definition (Equivalent Distributions). Two functionals $f(t)$ and $g(t)$ are equivalent if

$$\langle f(t), \phi(t) \rangle = \langle g(t), \phi(t) \rangle \tag{3.147}$$

for all test functions.

A distribution $f(t)$ that is equivalent to a classically defined function $f_0(t)$ so that

$$\langle f(t), \phi(t) \rangle = \langle f_0(t), \phi(t) \rangle \tag{3.148}$$

is termed a *regular distribution*. Distributions that have no direct expression as a standard function belong to the class of *singular distributions*. The most celebrated example of a singular distribution is the Dirac delta, which we will study in detail.

Distributions $f(t)$ defined on the space of the rapidly descending test functions are *tempered distributions* or *distributions of slow growth.* The concept of slow growth applies equally well to regular and singular distributions. For the former, it can be illustrated with familiar concepts:

Definition (Slow Increase). A function is said to be *slowly increasing*, *tempered*, or of *slow growth* if for some positive integer M, the product

$$t^{-M} \cdot f(t) \tag{3.149}$$

remains bounded as $|t| \to \infty$.

In essence, a slowly increasing function is one that can be tamed (tempered) by a sufficiently high power of t in (3.149). Examples include the polynomials, the sine and cosine, as well as $\sin t/t$. The exponential e^t grows too rapidly to be tempered.

Remarks. Functions of slow growth are not generally integrable, a fact that later hinders the description of their spectral content via the integral Fourier transform.

We will demonstrate this difficulty in Chapter 5 and remedy the situation by defining a generalized Fourier transform in Chapter 6; of utility will be many of the concepts developed here.

Note that the product of a function of slow growth $g(t)$ and a function of rapid descent $f(t)$ is a function of rapid descent. This is easily demonstrated. According to (3.144) and (3.149), there is some M for which

$$\upsilon \cdot g(t)f(t) \tag{3.150}$$

remains bounded as $t \to \infty$ for any positive integer $\upsilon \equiv N - M$. Therefore $h(t) \equiv f(t)g(t)$ decreases rapidly. Such products are therefore integrable. For signal analysis applications, distributions of slow growth and test functions of rapid descent are the most useful set of *dual spaces* in the distribution literature. For alternative spaces, see Ref. 10.

3.5.2.2 Properties of Distributions. Many standard operations such as scaling, addition, and multiplication by constants have predictable effects on the inner product defining distributions. In selected cases these operations map distributions to distributions, giving us flexibility to add them, scale the independent variable, multiply by classically defined functions or constants, and take derivatives.

Proposition. Let $f(t)$ be a distribution. Then $\frac{df}{dt}$ is a functional which is continuous and linear.

Proof: By definition, a distribution $\frac{df}{dt}$ is a functional satisfying

$$\left\langle \frac{df}{dt}, \phi(t) \right\rangle = f(t)\phi(t)\Big|_{-\infty}^{\infty} - \left\langle f(t), \frac{d\phi}{dt} \right\rangle = -\left\langle f(t), \frac{d\phi}{dt} \right\rangle \tag{3.151}$$

Continuity is assured by noting

$$\lim_{k \to \infty} \left\langle \frac{df}{dt}, \phi_k(t) \right\rangle = -\lim_{k \to \infty} \left\langle f(t), \frac{d\phi_k}{dt} \right\rangle = 0, \tag{3.152}$$

which follows directly from the stipulations placed on the test function $\phi(t)$. Linearity is easy to establish and is left to the reader. ■

Derivatives of higher order follow in a similar manner. Consider the second derivative $\frac{d^2 f}{dt^2}$. Let $g(t) = f'(t)$, where the prime denotes differentiation. Then for $g'(t) \equiv f''(t)$ the derivative rule leads to

$$\int_{-\infty}^{\infty} g'(t)\phi(t)\, dt = -\int_{-\infty}^{\infty} g(t)\phi'(t)\, dt. \tag{3.153}$$

According to the conditions defining a test function, $\psi(t) \equiv \phi'(t)$ is also a bona fide test function, so we can reexpress the derivative rule for $g(t)$:

$$\int_{-\infty}^{\infty} g'(t)\phi(t)\ dt = -\int_{-\infty}^{\infty} g(t)\phi'(t)\ dt = -\int_{-\infty}^{\infty} g(t)\psi(t)\ dt\,. \tag{3.154}$$

But the last element in this equality can be further developed to

$$-\int_{-\infty}^{\infty} g(t)\psi(t)\ dt = -\int_{-\infty}^{\infty} f'(t)\psi(t)\ dt = -\left[\int_{-\infty}^{\infty} f'(t)\psi(t)\ dt\right] = \int_{-\infty}^{\infty} f(t)\psi'(t)\ dt \tag{3.155}$$

where we used the first derivative property for $f(t)$. Linking (3.154) and (3.155) leads to the desired result expressed in terms of the original test function $\phi(t)$:

$$\int_{-\infty}^{\infty} f''(t)\phi(t)\ dt = \int_{-\infty}^{\infty} f(t)\phi''(t)\ dt \tag{3.156}$$

This result generalizes to derivatives of all orders (exercise).

Proposition (Scaling). Let a be a constant. Then

$$\langle f(at),\phi(t)\rangle = \frac{1}{|a|}\left\langle f(t),\phi\left(\frac{t}{a}\right)\right\rangle \tag{3.157}$$

Proof: Left as an exercise. ■

Note that (3.157) does not *necessarily* imply that $f(at) = \frac{1}{|a|}f(t)$, although such equivalence may be obtained in special cases. If this is not clear, reconsider the definition of equivalence.

Proposition (Multiplication by Constant). If a is a constant, if follows that

$$\langle af(t),\phi(t)\rangle = a\langle f(t),\phi(t)\rangle. \tag{3.158}$$

Proof: Trivial. ■

In many signal analysis applications it is common to mix distributions and classically defined functions.

Proposition (Associativity). Let $f_0(t)$ be an infinitely differentiable regular distribution. Then

$$\langle f_0(t)f(t),\ \phi(t)\rangle = \langle f(t), f_0(t)\phi(t)\rangle. \tag{3.159}$$

Proof: Exercise. ■

Why must we stipulate that $f_0(t)$ be infinitely differentiable? In general, the product of two distributions is *not* defined unless at least one of them is an infinitely differentiable regular distribution. This associative law (3.159) could be established because $f_0(t)$ was a regular distribution and the product $f_0(t)\phi(t)$ has meaning as a test function.

Proposition (Derivative of Product). Let $f_0(t)$ be a regular distribution and let $f(t)$ be a distribution of arbitrary type. The derivative of their product is a distribution satisfying

$$\frac{d}{dt}[f_0(t)f(t)] = f(t)\frac{df_0}{dt} + f_0(t)\frac{df}{dt}. \tag{3.160}$$

Proof: Consider the distribution represented by the second term above, regrouping factors and then applying the derivative rule:

$$\int_{-\infty}^{\infty} f_0(t)\frac{df}{dt}\phi(t)\ dt = \int_{-\infty}^{\infty} \frac{df}{dt}\cdot f_0(t)\phi(t)\ dt = -\int_{-\infty}^{\infty} f(t)\cdot\frac{d}{dt}[f_0(t)\phi(t)]\ dt. \tag{3.161}$$

The derivative in the last equality can be unpacked using the classical product rule since both factors are regular functions,

$$-\int_{-\infty}^{\infty} f(t)\cdot\left[\phi(t)\frac{df_0}{dt} + f_0(t)\frac{d\phi}{dt}\right] dt = -\int_{-\infty}^{\infty} f(t)\cdot\phi(t)\frac{df_0}{dt}\ dt - \int_{-\infty}^{\infty} f(t)\cdot\frac{d\phi}{dt}f_0(t)\ dt. \tag{3.162}$$

The right-hand side can be rearranged by applying the derivative rule to the second term. This gives two terms with $\phi(t)$ acting as a test function:

$$-\int_{-\infty}^{\infty} f(t)\frac{df_0}{dt}\cdot\phi(t)\ dt + \int_{-\infty}^{\infty} \frac{d}{dt}[f(t)f_0(t)]\cdot\phi(t)\ dt. \tag{3.163}$$

Comparing (3.162) and (3.163) and applying the definition of equivalence gives

$$f_0(t)\frac{df}{dt} = -f(t)\frac{df_0}{dt} + \frac{d}{dt}[f(t)f_0(t)], \tag{3.164}$$

from which the desired result (3.161) follows. ■

In many applications it is convenient to scale the independent variable. Distributions admit a chain rule under differentiation.

Proposition (Chain Rule). If $f(t)$ is an arbitrary distribution, a is a constant, and $y \equiv at$, then

$$\frac{d}{dt}f(at) = a\frac{d}{dt}\ f(t). \tag{3.165}$$

Proof: Exercise. ■

3.5.3 The Dirac Delta

Certain signals $f(t)$ exhibit jump discontinuities. From a classical Riemannian perspective, the derivative df/dt is singular at the jump. The situation can be reassessed within the context of distributions. Consider the unit step $u(t)$. For a test function of rapid descent $\phi(t)$, integration by parts produces

$$\int_{-\infty}^{\infty} \frac{du}{dt}\phi(t)\, dt = u(t)\phi(t)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} u(t)\frac{d\phi}{dt}\, dt = -\int_{-\infty}^{\infty} u(t)\frac{d\phi}{dt}\, dt. \tag{3.166}$$

This reduces to

$$-\int_{0}^{\infty} u(t)\frac{du}{dt}\, dt = -[\phi(\infty) - \phi(0)] = \phi(0), \tag{3.167}$$

so that

$$\int_{-\infty}^{\infty} \frac{du}{dt}\phi(t)\, dt = \phi(0). \tag{3.168}$$

The existence of $\frac{d\phi}{dt}$ is central to the preceding argument and is guaranteed by the definition of the test function. Consider the following definition.

Definition (Dirac Delta). The Dirac delta $\delta(t)$ is a functional that is equivalent to the derivative of the unit step,

$$\delta(t) \equiv \frac{du}{dt}. \tag{3.169}$$

From (3.168), we have

$$\langle \delta(t), \phi(t) \rangle = \left\langle \frac{du}{dt}, \phi(t) \right\rangle = \phi(0) \tag{3.170}$$

so the value returned by the distribution $\delta(t)$ is the value of the test function at the origin. It is straightforward to show that the Dirac delta satisfies the conditions of continuity and linearity; it therefore belongs to the class of functionals defined as distributions. Appearances are deceiving: despite its singular nature, the Dirac delta is a distribution of slow growth since it is defined on the (dual) space consisting of the test functions of rapid descent.

By simple substitution of variables we can generalize further:

$$\langle \delta(t-\tau), \phi(t) \rangle = \int_{-\infty}^{\infty} \frac{d}{dt}u(t-\tau)\phi(t)\, dt = \phi(\tau). \tag{3.171}$$

This establishes the Dirac delta as a *sifting* operator which returns the value of the test function at an arbitrary point $t = \tau$. In signal analysis, the process of sampling, whereby a the value of a function is determined and stored, is ideally represented by an inner product of the form (3.171).

Remark. Unfortunately, it has become common to refer to the Dirac delta as the "delta function." This typically requires apologies such as "the delta function is not a function", which can be confusing to the novice, and imply that the Dirac delta is the result of mathematical sleight of hand. The term delta distribution is more appropriate. The truth is simple: The delta function is not a *function*; it is a *functional* and if the foregoing discussion is understood thoroughly, the Dirac delta is (rightly) stripped of unnecessary mathematical mystique.

The sifting property described by (3.171) can be further refined through a test compact support on the interval $t \in [a, b]$. The relevant integration by parts,

$$\int_a^b \frac{du}{dt}\phi(t)\ du = u(t)\phi(t)\Big|_a^b - \int_a^b u(t)\frac{d\phi}{dt}\ dt, \tag{3.172}$$

takes specific values depending on the relative location of the discontinuity (in this case, located at $t = 0$ for convenience) and the interval on which the test function is supported. There are three cases:

(i) $a < 0 < b$:

$$\int_a^b \frac{du}{dt}\phi(t)\ du = u(t)\phi(t)\Big|_a^b - \int_0^b 1 \cdot \frac{d\phi}{dt}\ dt = \phi(b) - [\phi(b) - \phi(0)] = \phi(0). \tag{3.173}$$

(ii) $a < b < 0$: Since the unit step is identically zero on this interval,

$$\int_a^b \frac{du}{dt}\phi(t)\ du = 0 - \int_a^b 0 \cdot \frac{d\phi}{dt}\ dt = 0\,. \tag{3.174}$$

(iii) $b > a > 0$: on this interval, $u(t) = 1$, so that

$$\int_a^b \frac{du}{dt}\phi(t)\ du = \phi(t)\Big|_a^b - \int_a^b 1 \cdot \frac{d\phi}{dt}\ dt = 0\,. \tag{3.175}$$

Remark. In general, it is not meaningful to assign a pointwise value to a distribution since by their nature they are defined by an integral over an interval specified by the test function. Cases (ii) and (iii) assert that $\delta(t)$ is identically zero on the interval $[a, b]$. The above arguments can be applied to the intervals $[-\infty, -\varepsilon]$ and $[\varepsilon, \infty]$ where ε is arbitrarily small, demonstrating that the Dirac delta is identically zero at all points along the real line except the origin, consistent with expected behavior of $\frac{du}{dt}$.

The general scaling law (3.157) relates two inner products. But when the distribution is a Dirac delta, there are further consequences.

Proposition (Scaling). If $a \neq 0$, then

$$\delta(at) = \frac{1}{|a|}\delta(t). \tag{3.176}$$

Proof: Note that

$$\int_{-\infty}^{\infty} \delta(at)\phi(t)\, dt = \frac{1}{|a|}\int_{-\infty}^{\infty} \delta(t)\phi\left(\frac{t}{a}\right) dt = \frac{1}{|a|}\phi(0). \tag{3.177}$$

The right-hand side can be expressed as

$$\frac{1}{|a|}\phi(0) = \frac{1}{|a|}\int_{-\infty}^{\infty} \delta(t)\phi(t)\, dt. \tag{3.178}$$

Applying the definition of equivalence leads to the scaling law (3.176). ■

The special case $a = -1$ leads to the relation $\delta(-t) = \delta(t)$, so the Dirac delta has even symmetry. Some useful relations follow from the application of (3.159) to the Dirac delta.

Proposition (Associative Property). We have

$$\delta(t)f_0(t) = \delta(t)f_0(0). \tag{3.179}$$

Proof: Since

$$\int_{-\infty}^{\infty} \delta(t)f_0(t)\phi(t)\, dt = f_0(0)\phi(0) \equiv \int_{-\infty}^{\infty} \delta(t)f_0(0)\phi(t)\, dt, \tag{3.180}$$

this establishes the equivalence (3.179). ■

Note this does *not* imply $f_0(t) = f_0(0)$, since division is not an operation that is naturally defined for arbitrary singular distributions such as $\delta(t)$. However, (3.179) leads to some interesting algebra, as shown in the following example.

Example. Suppose $f_0(t) = t$. According to (3.179),

$$t\delta(t) = 0 \tag{3.181}$$

for all *t*. This implies that if $f(t)$ and $g(t)$ are distributions, and $tg(t) = tf(t)$, then

$$g(t) = f(t) + a_0\delta(t), \tag{3.182}$$

where a_0 is constant. Note that (3.182) applies equally well to regular and singular distributions; it is central to establishing important relations involving the generalized Fourier transform in Chapter 6.

The inner product defining the derivative,

$$\int_{-\infty}^{\infty} \frac{df}{dt}\phi(t)\, dt = -\int_{-\infty}^{\infty} f(t)\frac{d\phi}{dt}\, dt, \tag{3.183}$$

also leads to a sifting property.

Proposition (Differentiation)

$$\int_{-\infty}^{\infty} \delta'(t-\tau)\phi(t)\, dt = -\phi'(\tau). \tag{3.184}$$

Proof: Using a prime to denote differentiation with respect to *t*, we obtain

$$\int_{-\infty}^{\infty} \delta'(t)\phi(t)\, dt = -\int_{-\infty}^{\infty} \delta(t)\phi'(t)\, dt = -\phi'(0). \tag{3.185}$$

For an arbitrary Dirac delta centered at $t = \tau$, this generalizes to (3.184). ■

3.5.4 Distributions and Convolution

The convolution operation is central to analyzing the output of linear systems. Since selected signals and system impulse responses may be expressed in terms of the Dirac delta, some of our applications may involve the convolution of two singular distributions, or singular and regular distributions. Given that the product of two singular distributions is not defined, it may come as an unexpected result to define a convolution operation. As before, we will base the development on an analogous result derived from Riemannian calculus.

First, consider the convolution of a distribution with a test function. This problem is straightforward. Let $f(t)$ be an arbitrary distribution. Then the convolution

$$(f*\phi)(u) = \int_{-\infty}^{\infty} f(t)\phi(u-t)\, dt = \int_{-\infty}^{\infty} f(u-t)\phi(t)\, dt \tag{3.186}$$

is a function in the variable u.

Next, consider the convolution of two test functions. If $\phi(t)$ and $\psi(t)$ are test functions, their convolution presents no difficulty:

$$(\phi*\psi)(u) \equiv \int_{-\infty}^{\infty} \phi(u-t)\psi(t)\,dt = \int_{-\infty}^{\infty} \phi(t)\psi(u-t)\,dt = (\psi*\phi)(u). \qquad (3.187)$$

Remark. If $(\phi*\psi)(u)$ is to be a test function of compact support, both $\phi(t)$ and $\psi(t)$ must also be compactly supported. When this consideration impacts an important conclusion, it will be noted.

Now reconsider this convolution in terms of a Hilbert space inner product. If we define

$$\psi_{\mathrm{ref}} \equiv \psi(-t), \qquad (3.188)$$

then

$$(\phi*\psi_{\mathrm{ref}})(u) = \int_{-\infty}^{\infty} \phi(t)\psi(u+t)\,dt. \qquad (3.189)$$

This leads to a standard inner product, since

$$(\phi*\psi_{\mathrm{ref}})(0) = \int_{-\infty}^{\infty} \phi(t)\psi(t)\,dt = \langle \phi(t),\psi(t)\rangle. \qquad (3.190)$$

Because $(\phi*\psi)(t)$ is a test function, it follows that

$$\langle \phi*\psi,\eta\rangle = ((\phi*\psi)*\eta_{\mathrm{ref}})(0) = (\phi*(\psi_{\mathrm{ref}}*\eta)_{\mathrm{ref}})(0). \qquad (3.191)$$

However,

$$(\phi*(\psi_{\mathrm{ref}}*\eta)_{\mathrm{ref}})(0) = \langle \phi,\psi_{\mathrm{ref}}*\eta\rangle. \qquad (3.192)$$

Comparing the last two equations gives the desired result:

$$\langle \phi*\psi,\eta\rangle = \langle \phi,\psi_{\mathrm{ref}}*\eta\rangle. \qquad (3.193)$$

The purpose of this exercise was to allow the convolution to migrate to the right-hand side of the inner product. This leads naturally to a definition that embraces the convolution of two singular distributions.

Definition (Convolution of Distributions). Using (3.193) as a guide, if $f(t)$ and $g(t)$ are distributions of any type, including singular, their convolution is defined by

$$\langle f*g,\phi\rangle \equiv \langle f, g_{\mathrm{ref}}*\phi\rangle. \qquad (3.194)$$

Example (Dirac Delta). Given two delta distributions $f(t) = \delta(t-a)$ and $g(t) = \delta(t-b)$, we have

$$g_{\mathrm{ref}} = \delta(-t-b) = \delta(-(t+b)) = \delta(t+b). \tag{3.195}$$

So

$$g_{\mathrm{ref}}*\phi = \int_{-\infty}^{\infty} \phi(u-t)\delta(t+b)\,dt = \phi(u+b), \tag{3.196}$$

and the relevant inner product (3.194) takes the form

$$\langle f*g, \phi\rangle = \int_{-\infty}^{\infty} \delta(u-a)\phi(u+b)\,dt = \phi(a+b). \tag{3.197}$$

So the desired convolution is expressed,

$$f*g = \delta(t-(a+b)). \tag{3.198}$$

Later developments will lead us derivatives of distributions. An importaint result involving convolution is the following. A similar result holds for standard analog signals (functions on the real line).

Proposition (Differentiation of Convolution). Let f and g be arbitrary distributions. Then,

$$\frac{d}{dt}(f*g) = f*\frac{dg}{dt} = \frac{df}{dt}*g\ . \tag{3.199}$$

Proof: The proof of (3.199) requires the usual inner product setting applicable to distributions and is left as an exercise. (You may assume that convolution of distributions is commutative, which has been demonstrated for standard functions and is not difficult to prove in the present context.) ∎

3.5.5 Distributions as a Limit of a Sequence

Another concept that carries over from function theory defines a functional as a limit of a sequence. For our purposes, this has two important consequences. First, limits of this type generate approximations to the Dirac delta, a convenient property that impacts the wavelet transform (Chapter 11). Second, such limits make predictions of the high-frequency behavior of pure oscillations (sinusoids) which are the foundations of Fourier analysis (Chapters 5 and 6). This section covers both topics.

Definition. Let $f(t)$ be a distribution of arbitrary type, and suppose $f_n(t)$ is a sequence in some parameter n. Then if

$$\lim_{n\to\infty} \langle f_n(t),\phi(t)\rangle \to \langle f(t),\phi(t)\rangle, \tag{3.200}$$

then the sequence $f_n(t)$ approaches $f(t)$ in this limit:

$$\lim_{n \to \infty} f_n(t) = f(t). \tag{3.201}$$

3.5.5.1 Approximate Identities for the Dirac Delta. The Dirac delta has been introduced as the derivative of the unit step, a functional which is identically zero along the entire real line except in an infinitesimal neighborhood of the origin. A surprising number of regular distributions—such as the square pulse, suitably scaled functions involving trigonometric functions, and another family generated by Gaussians—approach the Dirac delta when sufficiently scaled and squeezed in the limit in which the scaling factor becomes large. Intuitively, we need normalized, symmetric functions with maxima at the origin and that approach zero on other points along the real line as the appropriate limit (3.201) is taken. This leads to the following theorem.

Theorem (Approximating Identities). Let $f(t)$ be a regular distribution satisfying the criteria:

$$f(t) \in L^1(R), \tag{3.202}$$

$$\int_{-\infty}^{\infty} f(t)\, dt = 1. \tag{3.203}$$

Define

$$f_a(t) \equiv af(at) \tag{3.204}$$

and stipulate

$$\lim_{a \to \infty} \int_{\tilde{R}} |f_a(t)| = 0, \tag{3.205}$$

where $\tilde{R}$ denotes the real line minus a segment of radius ρ centered around the origin: $\tilde{R} \equiv R/[-\rho, \rho]$.

Then $\lim_{a \to \infty} f_a(t) = \delta(t)$, and $f_a(t)$ is said to be an *approximating identity.*

Proof: First,

$$|\langle f_a(t), \phi(t)\rangle - \langle \delta(t), \phi(t)\rangle| = \left| \int_{-\infty}^{\infty} af(at)\phi(t)\, dt - \phi(0) \right|. \tag{3.206}$$

From (3.203) a simple substitution of variables shows $\int_{-\infty}^{\infty} af(at)\, dt = 1$. We can thus recast the right-hand side of (3.206) into a more convenient form:

$$\left| \int_{-\infty}^{\infty} af(at)\phi(t)\, dt - \phi(0) \right| = \left| \int_{-\infty}^{\infty} af(at)\phi(t)\, dt - \int_{-\infty}^{\infty} af(at)\phi(0)\, dt \right|. \tag{3.207}$$

Dividing the real line into regions near and far from the origin, a change of variables $u \equiv at$ gives

$$\left| \int_{-\infty}^{\infty} f(u)\left[\phi\left(\frac{u}{a}\right) - \phi(0)\right] du \right| = I_1 + I_2, \tag{3.208}$$

where

$$I_1 = \left| \int_{-\rho/a}^{\rho/a} f(u)\left[\phi\left(\frac{u}{a}\right) - \phi(0)\right] du \right| \tag{3.209}$$

and

$$I_2 = \left| \int_{R/\left(\frac{-\rho}{a}, \frac{\rho}{a}\right)} f(u)\left[\phi\left(\frac{u}{a}\right) - \phi(0)\right] du \right|. \tag{3.210}$$

Since $\phi(t)$ is continuous, for some t suitably close to the origin, we have

$$\left|\phi\left(\frac{u}{a}\right) - \phi(0)\right| < \varepsilon. \tag{3.211}$$

Thus, I_1 is bounded above:

$$I_1 \le \varepsilon \int_{-\rho/a}^{\rho/a} |f(u)|\ du. \tag{3.212}$$

In the second integral, note that there exists some ρ such that

$$\int_{R/\left(\frac{-\rho}{a}, \frac{\rho}{a}\right)} |f(u)|\ du \le \varepsilon. \tag{3.213}$$

Equation (3.210) can be expressed as

$$I_2 \le \int_{R/\left(\frac{-\rho}{a}, \frac{\rho}{a}\right)} \phi\left(\frac{u}{a}\right)|f(u)|\ du - |\phi(0)| \int_{R/\left(\frac{-\rho}{a}, \frac{\rho}{a}\right)} |f(u)|\ du. \tag{3.214}$$

Furthermore, by definition, $\phi(t)$ is bounded above. Consequently,

$$I_2 \le \left|\phi\left(\frac{u}{a}\right)\right|_{\max} \int_{R/\left(\frac{-\rho}{a}, \frac{\rho}{a}\right)} |f(u)|\ du - |\phi(0)|\varepsilon, \tag{3.215}$$

which reduces to

$$I_2 \le \varepsilon\left[\left|\phi\left(\frac{u}{a}\right)\right|_{\max} - |\phi(0)|\right]. \tag{3.216}$$

Returning to (3.206), we have

$$\left|\langle f_a(t),\phi(t)\rangle - \langle\delta(t),\phi(t)\rangle\right| \le \varepsilon\left[\int\limits_{R/\left(\frac{-\rho}{a},\frac{\rho}{a}\right)} |f(u)|\ du + \left|\phi\left(\frac{u}{a}\right)\right|_{\max} - |\phi(0)|\right]. \quad (3.217)$$

In the limit of large a, we have

$$\lim_{a\to\infty} \varepsilon = 0, \quad (3.218)$$

so that

$$\lim_{a\to\infty}\left|\langle f_a(t),\phi(t)\rangle - \langle\delta(t),\phi(t)\rangle\right| = 0 \quad (3.219)$$

and we obtain the desired result. ■

The decay condition (3.205) is equivalent to

$$\lim_{t\to\infty} f(t) = 0, \quad (3.220)$$

which can be stated: for each $\lambda > 0$ there exists $T_\lambda > 0$ such that $|f(t)| < \lambda$ for all $|t| > T_\lambda$. This is equivalent to stating that if $\lambda/a > 0$ there exists $T_{\lambda/a} > 0$ such that $|f(at)| < \lambda/a$ for $|at| > T_{\lambda/a}$. The scaling required to convert $f(t)$ to an approximate identity implies that

$$|af(at)| = a|f(t)| < a\frac{\lambda}{a} < \lambda \quad (3.221)$$

so that

$$\lim_{|t|\to\infty} af(at) = 0. \quad (3.222)$$

Remark. Functions $f(t)$ that satisfy (3.202)–(3.205) are commonly called *weight functions*. They are of more than academic interest, since their localized atomistic character and ability to wrap themselves around a selected location make them useful for generating wavelets (Chapter 11). Such wavelets can zoom in on small-scale signal features. The class of weight functions is by no means small; it includes both uniformly and piecewise continuous functions.

There are several variations on these approximating identities. For example, (3.203) can be relaxed so that if the weight function has arbitrary finite area

$$\Gamma \equiv \int_{-\infty}^{\infty} f(t)\ dt, \quad (3.223)$$

then (3.125) reads

$$\lim_{a \to \infty} f_a(t) = \Gamma\delta(t). \tag{3.224}$$

The scaling of the amplitude can be modified so that

$$f_a(t) \equiv \sqrt{a} f(at) \tag{3.225}$$

is an approximating identity. The proofs are left as exercises. These approximating identities give the wavelet transform the ability to extract information about the local features in signals, as we will demonstrate in Chapter 11.

We consider several common weight functions below.

Example (Gaussian). Let

$$f(t) = A_0 e^{-t^2/2}, \tag{3.226}$$

where A_0 is a constant. The general Gaussian integral

$$\int_{-\infty}^{\infty} e^{-\alpha y^2} dy \equiv \sqrt{\frac{\pi}{\alpha}} \tag{3.227}$$

implies that

$$A_0 = \frac{1}{\sqrt{2\pi}} \tag{3.228}$$

if $f(t)$ is to have unit area. Note that the ability to normalize the area is proof that $f(t)$ is integrable, so two of the weight function criteria are satisfied. Figure 3.7 illustrates a Gaussian approximate identity

$$f_a(t) = \frac{a}{\sqrt{2\pi}} e^{-(at)^2/2} \tag{3.229}$$

for increasing values of a. The Gaussian is a powerful tool in the development of the continuous wavelet transform. We shall review its role as an approximating identity when we cover the small-scale resolution of the wavelet transform.

Example (Abel's Function). In this case the weight function permits an arbitrary positive-definite parameter β,

$$f(t) = A_0 \frac{\beta}{1 + \beta^2 t^2}. \tag{3.230}$$

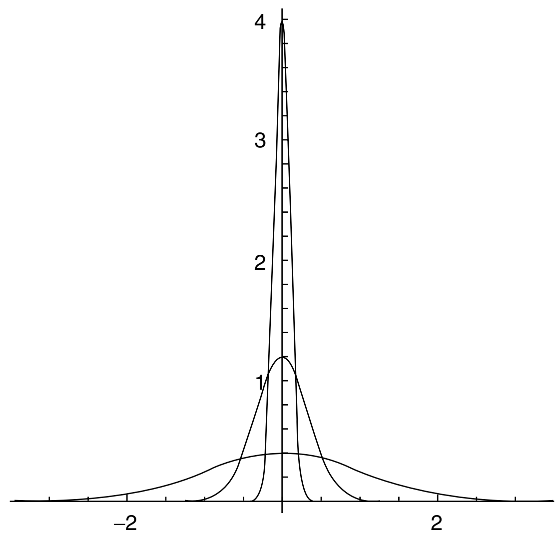

Fig. 3.7. The evolution of the scaled Gaussian (3.229) for $a = 1$ (widest), $a = 3$, and $a = 10$. In the limit $a \to \infty$, the scaled Gaussian approaches a Dirac delta.

It is left as an exercise to show that the condition of unit area requires

$$A_0 = \frac{1}{\pi}. \qquad (3.231)$$

For this purpose it is convenient to use the tabulated integral,

$$\int_0^{\infty} \frac{x^{\mu - 1}}{1 + x^{\nu}} dx = \frac{\pi}{\nu} \csc \frac{\mu\pi}{\nu}, \qquad (3.232)$$

where $Re\nu > Re\mu > 0$. The resulting approximate identity has the form

$$f_a(t) = \frac{a}{\pi} \frac{\beta}{1 + \left(\frac{\beta}{a}\right)^2 t^2}. \qquad (3.233)$$

Example (Decaying Exponential). This weight function is piecewise continuous, but fulfills the required criteria:

$$f(t) = A_0 e^{-|t|}. \qquad (3.234)$$

It is left as an exercise to show that $A_0 = 1$, so that $f_a(t) = ae^{-|t/a|}$.

Example (Rectangular Pulse). The unit-area rectangular pulse described by

$$f(t) = \frac{1}{\sigma}[u(t) - u(t-\sigma)] \tag{3.235}$$

approximates a Dirac delta (exercise).

Remark. The weight functions described in this section embody properties that are sufficient to generate approximate identities. These conditions are by no means necessary. The family of sinc pulses

$$f(t) = \frac{1}{\pi}\frac{\sin t}{t} \tag{3.236}$$

does not qualify as a weight function (why not?), yet it can be shown that

$$\lim_{a \to \infty} \frac{1}{\pi}\frac{\sin at}{t} = \delta(t)\,. \tag{3.237}$$

Note that the scale is applied only in the numerator. The properties of this approximate identity are explored in the exercises.

3.5.5.2 The Riemann–Lebesgue Lemma. Much of signal analysis involves the study of spectral content—the relative contributions from individual pure tones or oscillations at given frequency ω, represented by $\sin(\omega t)$ and $\cos(\omega t)$. The behavior of these oscillations as $\omega \to \infty$ is not predicted classically, but by treating the complex exponential $e^{j\omega t}$ as a regular distribution, the foregoing theory predicts that $\sin(\omega t)$ and $\cos(\omega t)$ vanish identically in the limit of high frequency. This result is useful in several subsequent developments.

Theorem (Riemann–Lebesgue Lemma). Let $f(t) \equiv e^{-j\omega t}$. If $\phi(t)$ is a test function of compact support on the interval $t \in [t_1, t_2]$, then

$$\lim_{\omega \to \infty} \int_{-\infty}^{\infty} f(t)\phi(t)\, dt = 0. \tag{3.238}$$

Proof: It is convenient to define a $g(t)$ such that

$$\frac{dg}{dt} \equiv f(t) = e^{-j\omega t} \tag{3.239}$$

and apply integration by parts to obtain

$$\int_{t_1}^{t_2} \frac{dg}{dt}\phi(t)\, dt = g(t)\phi(t)\Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} g(t)\frac{d\phi}{dt}\, dt\,. \tag{3.240}$$

Then

$$\int_{t_1}^{t_2} e^{-j\omega t}\phi(t)\, dt = \frac{1}{j\omega}[\phi(t_2)e^{-j\omega t_2} - \phi(t_1)e^{-j\omega t_1}] + \frac{1}{j\omega}\int_{t_1}^{t_2} e^{-j\omega t}\frac{d\phi}{dt}\, dt. \tag{3.241}$$

By definition, the test function is bounded. Hence,

$$\left|\phi(t)e^{-j\omega t}\right| \le |\phi(t)|\left|e^{-j\omega t}\right| \le |\phi(t)|, \tag{3.242}$$

and the numerator of the first term in (3.241) is bounded. Consequently, in the limit $\omega \to 0$, this first term vanishes. The integral in the second term is also finite (it is the Fourier transform of $\frac{d\phi}{dt}$), and we defer proof of its boundedness until Chapter 5. Finally,

$$\lim_{\omega \to \infty} e^{-j\omega t} = 0. \tag{3.243}$$

By extension, the real and imaginary parts of this exponential ($\cos(\omega t)$ and $\sin(\omega t)$) also vanish identically in this limit. ■

Remark. Test functions of compact support, as used herein, greatly simplify matters. When $(t_1, t_2) \to (-\infty, \infty)$, thus generalizing the result to all t, the same result obtains; we omit the complete proof, which is quite technical.

3.6 SUMMARY

Analog signal processing builds directly upon the foundation of continuous domain function theory afforded by calculus [6] and basic topology of sets of real numbers [22]. An analog system accepts an input signal and alters it to produce an output analog signal—a simple concept as long as the operations involve sums, scalar multiplications, translations, and so on. Linear, translation-invariant systems appear in nature and engineering. The notion is a mix of simple operations, and although there was a very straightforward theory of discrete LTI systems and discrete impulse response $\delta(n)$, replacing the n by a t stirs up serious complications.

Informal arguments deliver an analog equivalent of the convolution theorem, and they ought to be good enough justification for most readers. Others might worry about the analog impulse's scaling property and the supposition that an ordinary signal $x(t)$ decomposes into a linear combination of Diracs. But rigorous justification is possible by way of distribution theory [8–10].

Analog signals of certain classes, of which the L^p signals are especially convenient for signal processing, form elegant mathematical structures: normed spaces, Banach spaces, inner product, and Hilbert spaces. The inner product is a natural measure of similarity between signals. Completeness, or closure under limit operations, exists in Banach and Hilbert spaces and allows for incremental approximation of signals. Many of the results from discrete theory are purely algebraic in nature; we have been able to appropriate them once we show that analog signals—having rightly chosen a norm or inner product definition, of course—reside in one of the familiar abstract function spaces.

3.6.1 Historical Notes

Distribution theory dates from the mid-1930s with the work of S.L. Sobolev. Independent of the Soviet mathematician, L. Schwartz formalized the notion of a distribution and developed a rigorous delta function theory [7]. Schwartz's lectures are in fact the inspiration for Ref. 8.

Splines were first studied by Schoenberg in 1946 [14]. It is interesting to note that while his paper preceded Shannon's more famous results on sampling and signal reconstruction [50], the signal processing research community overlooked splines for many years.

Mathematicians developed frame theory in the early 1950s and used it to represent functions with Fourier series not involving the usual sinusoidal harmonics [41, 42]. Only some 30 years later did applied mathematicians, physicists, and engineers discover the applicability of the concept to signal analysis [43].

Details on the origins of signal theory, signal spaces, and Hilbert spaces can be found in mathematically oriented histories [51, 52]. Advanced treatments of functional analysis include Refs. 53 and 54.

Lebesgue's own presentation of his integral is given in Ref. 55. This book also contains a short biography of Lebesgue. Lebesgue's approach was the best of many competing approaches to replace the Riemannian integral [56, 57].

3.6.2 Looking Forward

The next chapter covers time-domain signal analysis. Taking representative signal interpretation problems, Chapter 4 attempts to solve them using the tools we have developed so far. The time-domain tools we can formulate, however, turn out to be deficient in analyzing signals that contain periodicities. This motivates the frequency-domain analysis of signals, beginning with the analog Fourier transforms in Chapter 5.

In Chapter 6, the formal development of the Dirac delta as a functional will help build a Fourier transform theory for periodic, constant, and otherwise untransformable analog signals.

3.6.3 Guide to Problems

This chapter includes many basic problems (1–46) that reinforce the ideas of the text. Some mathematical subtleties have been covered but casually in the main text, and the problems help the interested reader pursue and understand them. Of course, in working the problems, the student should feel free to consult the mathematical literature for help. The advanced problems (47–56) require broader, deeper investigation. For example, the last few problems advance the presentation on frames. Searching the literature should provide a number of approaches that have been considered for these problems. Indeed, investigators have not satisfactorily answered all of the questions posed.

REFERENCES

1. A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and Systems*, 2nd ed., Upper Saddle River, NJ: Prentice-Hall, 1997.
2. J. A. Cadzow and H. F. Van Landingham, *Signals, Systems, and Transforms*, Englewood Cliffs, New Jersey: Prentice-Hall, 1983.
3. H. Baher, *Analog and Dignal Signal Processing*, New York: Wiley, 1990.
4. R. E. Ziemer, W. H. Tranter, and D. R. Fannin, *Signals and Systems: Continouous and Discrete*, 3rd ed., New York: Macmillan, 1993.
5. L. B. Jackson, Signals, *Systems, and Transforms*, Reading, MA: Addison-Wesley, 1991.
6. M. Rosenlicht, *An Introduction to Analysis*, New York: Dover, 1978.
7. L. Schwartz, *Théorie des Distributions*, vols. I and II, Paris: Hermann, 1950–1951.
8. I. Halperin, *Introduction to the Theory of Distributions*, Toronto: University of Toronto Press, 1952.
9. M. J. Lighthill, *Introduction to Fourier Analysis and Generalized Functions*, New York: Cambridge University Press, 1958.
10. A. H. Zemanian, *Distribution Theory and Transform Analysis*, New York: Dover, 1965.
11. A. Friedman, *Generalized Functions and Partial Differential Equations*, Englewood Cliffs, NJ: Prentice-Hall, 1963.
12. E. M. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton, NJ: Princeton University Press, 1971.
13. L. Debnath and P. Mikusinski, *Introduction to Hilbert Spaces with Applications*, 2nd ed., San Diego, CA: Academic Press, 1999.
14. I. J. Schoenberg, Contribution to the problem of approximation of equidistant data by analytic functions, *Quarterly of Applied Mathematics*, vol. 4, pp. 45–99, 112–141, 1946.
15. J. H. Ahlberg, E. N. Nilson, and J. L. Walsh, *The Theory of Splines and Their Applications*, New York: Academic Press, 1967.
16. D. Hearn and M. P. Baker, *Computer Graphics*, Englewood Cliffs, NJ: Prentice-Hall, 1986.
17. R. H. Bartels, J. C. Beatty, and B. A. Barsky, *Splines for Use in Computer Graphics*, Los Altos, CA: Morgan Kaufmann, 1987.
18. M. Unser, A. Aldroubi, and M. Eden, B-spline signal processing: Part I—theory, *IEEE Transactions on Signal Processing*, vol. 41, no. 2, pp. 821–833, February 1993.
19. M. Unser, A. Aldroubi, and M. Eden, B-spline signal processing: Part II—efficient design and applications, *IEEE Transactions on Signal Processing*, vol. 41, no. 2, pp. 834–848, February 1993.
20. S. Sakakibara, A practice of data smoothing by B-spline wavelets, in C. K. Chui, et al., eds., *Wavelets: Theory, Algorithms, and Applications*, San Diego, CA: Academic Press, pp. 179–196, 1994.
21. M. Unser, Splines: A perfect fit for signal and image processing, *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22–38, November 1999.
22. A. N. Kolmogorov and S. V. Fomin, *Introductory Real Analysis*, translator R. A. Silverman, New York: Dover, 1975.
23. R. Beals, *Advanced Mathematical Analysis*, New York: Springer-Verlag, 1987.

24. H. L. Royden, *Real Analysis*, 2nd ed., Toronto: Macmillan, 1968.
25. A. W. Naylor and G. R. Sell, *Linear Operator Theory in Engineering and Science*, New York: Springer-Verlag, 1982.
26. G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, New York: Macmillan, 1965.
27. E. Kreysig, *Introductory Functional Analysis with Applications*, New York: Wiley, 1989.
28. H. Dym and H. P. McKean, *Fourier Series and Integrals*, New York: Academic, 1972.
29. N. Young, *An Introduction to Hilbert Space*, Cambridge: Cambridge University Press, 1988.
30. A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, 2nd ed., vol. II, Orlando, FL: Academic, 1982.
31. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, New York: Wiley, 1973.
32. S. Haykin, *Communication Systems*, 3rd ed., New York: Wiley, 1994.
33. J. L. Flanagan, A. C. Surendran, and E.-E. Jan, Spatially selective sound capture for speech and audio processing, *Speech Communication*, vol. 13, pp. 207–222, 1993.
34. F. Glazer, G. Reynolds, and P. Anandan, Scene matching by hierarchical correlation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 19–23, Washington, D.C., pp. 432–441, 1983.
35. P. J. Burt, C. Yen, and X. Xu, Local correlation measures for motion analysis: a comparative study, *Proceedings of the IEEE Conference on Pattern Recogniton and Image Processing*, pp. 269–274, 1982.
36. A. Haar, Zur theorie der orthogonalen Functionensysteme, *Mathematics Annalen*, vol. 69, pp. 331–371, 1910.
37. P. Wojtaszczyk, *A Mathematical Introduction to Wavelets*, Cambridge: Cambridge University Press, 1997.
38. I. Daubechies, *Ten Lectures on Wavelets*, Philadelphia: SIAM Press, 1992.
39. S. Mallat and Z. Zhang Matching pursuits with time-frequency dictionaries, *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, December 1993.
40. Goodwin and M. Vetterli, Matching pursuit and atomic signal models based on recursive filter banks, *IEEE Transactions on Signal Processing*, vol. 47, no. 7, pp. 1890–1902, July 1999.
41. R. J. Duffin and A. C. Schaeffler, A class of nonharmonic Fourier series, *Transactions of the American Mathematical Society*, vol. 72, pp. 341–366, 1952.
42. R. M. Young, *An Introduction to Nonharmonic Fourier Series*, New York: Academic Press, 1980.
43. I. Daubechies, A. Grossmann, and Y. Meyer, Painless nonorthogonal expansions, *Journal of Mathematical Physics*, vol. 27, pp. 1271–1283, 1986.
44. W. Rudin, *Real and Complex Analysis*, 2nd ed., New York: McGraw-Hill, 1974.
45. E. Hewitt and K. Stromberg, *Real and Abstract Analysis*, New York: Springer, 1991.
46. P. R. Halmos, *Naive Set Theory*, New York: Van Nostrand, 1960.
47. J. H. Richards and H. K. Youn, *Theory of Distributions: A Nontechnical Introduction*, New York: Cambridge University Press, 1990.
48. A. Papoulis, *The Fourier Integral and Its Application*, New York: McGraw-Hill, 1962.

49. N. Wiener, *The Fourier Integral and Certain of Its Applications*, New York: Dover Publications, 1958.
50. C. E. Shannon, Communication in the presence of noise, *Proceedings of the Institute of Radio Engineers*, vol. 37, pp. 10–21, 1949.
51. J. Dieudonne, *History of Functional Analysis*, Amsterdam: North-Holland, 1981.
52. J. Stillwell, *Mathematics and Its History*, New York: Springer-Verlag, 1989.
53. K. Yosida, *Functional Analysis*, 6th ed., Berlin: Springer-Verlag, 1980.
54. W. F. Rudin, *Functional Analysis*, 2nd ed., New York: McGraw-Hill, 1991.
55. H. L. Lebesgue, *Measure and the Integral*, edited, with a biography by K. O. May, San Francisco: Holden-Day, 1966.
56. J.-P. Pier, *Development of Mathematics: 1900–1950*, Basel: Birkhaeuser Verlag, 1994.
57. T. Hawkins, *Lebesgue's Theory of Integration: Its Origin and Development*, New York: Chelsea, 1970.

PROBLEMS

1. Find the domain and range of the following analog systems; if necessary, narrow the problem domain to a particular analog signal space.

(a) The amplifier (or attenuator, $|A| < 1$) system: $y(t) = Ax(t)$

(b) A translation system: $y(t) = x(t - a)$

(c) The system on real-valued signals, $y(t) = x(t)^{1/2}$

(d) The system on complex-valued signals, $y(t) = x(t)^{1/2}$

(e) The adder: $y(t) = x(t) + x_0(t)$

(f) Termwise multiplication: $y(t) = x(t) \times x_0(t)$

(g) Convolution: $y(t) = x(t)*h(t)$

(h) Accumulator:

$$y(t) = \int_{-\infty}^{t} x(s)\, ds. \tag{3.244}$$

2. Let $x(t)$ be an analog signal and let H be an analog accumulator system. Show that:

(a) If $x(t)$ has finite support, then $x(t)$ is in the domain of H.

(b) If $x(t)$ is absolutely integrable, then it is in the domain of H.

(c) There are finite energy signals are not in the domain of H. (*Hint*: Provide an example based on the signal $x(t) = t^{-1}$.)

3. Consider the systems in Problem 1.

(a) Which of them are linear?

(b) Which of them are translation-invariant?

(c) Which of them are stable?

4. Suppose that the analog systems H and G are LTI. Let $T = aH$ and $S = H + G$. Show that both T and S are LTI systems.

5. Which of the following systems are linear? translation-invariant? stable? causal?

   (a) $y(t) = x(t/2)$

   (b) $y(t) = x(2t)$

   (c) $y(t) = x(0) + x(t)$

   (d) $y(t) = x(-t)$

   (e) The system given by

$$y(t) = \int_{-\infty}^{\infty} x(s)\cos(s)\, ds. \tag{3.245}$$

6. Consider the cross-correlation system $y = x^{\circ}h = Hx$, where

$$y(t) = (x^{\circ}h)(t) = \int_{-\infty}^{\infty} x(s)h(t+s)\, ds. \tag{3.246}$$

   (a) Prove or disprove: H is linear.

   (b) Prove or disprove: H is translation-invariant.

   (c) Prove or disprove: H is stable.

   (d) Is $h(t)$ the impulse response of H? Explain.

   (e) Answer these same questions for the autocorrelation operation $y = x^{\circ}x$.

7. An analog LTI system H has impulse response $h(t) = 2e^{t}u(2 - t)$, where $u(t)$ is the unit step signal. What is the response of H to $x(t) = u(t + 3) - u(t - 4)$?

8. Analog LTI system G has impulse response $g(t) = u(t + 13) - u(t)$; $u(t)$ is the unit step. What is the response of G to $x(t) = u(t)e^{-t}$?

9. Analog LTI system K has impulse response $k(t) = \delta(t + 1) + 2\delta(t - 2)$, where $\delta(t)$ is the Dirac delta. What is K's response to $x(t) = u(t - 1) - u(t)$?

10. Show that convolution is linear:

    (a) $h*(ax) = ah*x$.

    (b) $h*(x + y) = h*x + h*y$.

11. Show that if H and G are LTI systems, then $H(G(x)) = G(H(x))$.

12. Show that convolution is associative: $h*(x*y) = (h*x)*y$.

13. Give an alternative argument that $\delta(t) = u'(t)$, the derivative of the unit step [3].

    (a) Let $x(t)$ be a signal; show that integration by parts implies

$$\int_{-\infty}^{\infty} x(t)\frac{d}{dt}u(t)\, dt = x(t)u(t)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} x'(t)u(t)\, dt. \tag{3.247}$$

(b) Remove $u(t)$ from the final integrand in (3.247) and correct the limits of integration.

(c) Show that

$$\int_{-\infty}^{\infty} x(t)\frac{d}{dt}u(t)\ dt = x(0)\,, \tag{3.248}$$

which is the sifting property once again; hence, $\delta(t) = \frac{d}{dt}u(t)$.

14. Suppose the function (an analog signal) $x(t)$ is uniformly continuous over some real interval $I = (a, b)$. This means that for every $\varepsilon > 0$, there is a $\Delta > 0$ such that if s, $t \in I$ with $|s - t| < \Delta$, then $|x(s) - x(t)| < \varepsilon$. We allow the cases that $a = -\infty$, $b = \infty$, or both.

(a) Show that the signal $\sin(t)$ is uniformly continuous on $\mathbb{R}$.

(b) Show that $\exp(t)$ is not uniformly continuous on $\mathbb{R}$.

(c) Show that $x(t) = t^2$ is uniformly continuous on any finite interval (a, b), but not on any unbounded interval ($a = -\infty$, $b = \infty$, or both).

(d) Show that $x(t) = \sqrt{|t|}$ is not uniformly continuous on any interval that includes the origin.

(e) Prove or disprove: If $x(t)$ is continuous and differentiable and has a bounded derivative in the interval (a, b), then $x(t)$ is uniformly continuous on (a, b).

15. Suppose the sequence $\{x_n(t) \in \mathbb{N}\}$ converges uniformly on some real time interval (a, b). That is, every $\varepsilon > 0$ there is an $N_\varepsilon > 0$ such that $m, n > N_\varepsilon$ implies that for all $t \in I$ we have $|y_n(t) - y_m(t)| < \varepsilon$.

(a) Show that if each $x_n(t)$ is continuous, then the limit $x(t) = \lim_{n \to \infty} x_n(t)$ is also continuous.

(b) Show that $x(t)$ may not be continuous if the convergence is not uniform.

(c) Prove or disprove: If each $x_n(t)$ is bounded, then $x(t)$ is also bounded.

16. This problem explores interchanging limit and integration operations [6]. Suppose the signal $x_n(t)$ is defined for $n > 0$ by

$$x_n(t) = \begin{cases} 4n^2 t & \text{if } 0 \le t \le \frac{1}{2n}, \\ 4n - 4n^2 t & \text{if } \frac{1}{2n} \le t \le \frac{1}{n}, \\ 0 & \text{if otherwise.} \end{cases} \tag{3.249}$$

(a) Show that $x_n(t)$ is continuous and integrates to unity on (0, 1).

$$\int_0^1 x_n(t)\ dt = 1 \tag{3.250}$$

(b) Let $x(t) = \lim_{n \to \infty} x_n(t)$. Show that $x(t) = 0$ for all $t \in \mathbb{R}$.

(c) Conclude that

$$\int_0^1 x(t)\, dt \neq \lim_{n \to \infty} \int_0^1 x_n(t)\, dt\,. \tag{3.251}$$

17. Let $\{x_n(t)\}$ be continuous and converge uniformly on the real interval $[a, b]$ to $x(t)$.

(a) Show that $x(t)$ is continuous and therefore Riemann integrable on $[a, b]$.

(b) Using the uniform continuity, find a bound for the integral of $x_n(t) - x(t)$.

(c) Finally, show that

$$\lim_{n \to \infty} \int_a^b x_n(t)\, dt = \int_a^b x(t)\, dt. \tag{3.252}$$

18. Prove (3.252) assuming that $\{x_n(t)\}$ converge uniformly and are Riemann-integrable (but not necessarily continuous) on $[a, b]$.

(a) Approximate $x_n(t)$ by step functions.

(b) Show that the limit $x(t)$ is Riemann-integrable on $[a, b]$.

(c) Conclude that the previous problem applies and (3.252) holds once again [6].

19. Suppose that the sequence $\{x_n(t) \in \mathbb{N}\}$ converges uniformly on some real interval $I = (a, b)$ to $x(t)$; that each $x_n(t)$ is continuously differentiable on I; that for some $c \in I$, $\{x_n(c)\}$ converges; and that the sequence $\{x_n{}'(t)\}$ converges uniformly on I.

(a) Show that for all $n \in \mathbb{N}$ and all $t \in I$,

$$x_n(t) - x_n(a) = \int_a^t x_n{}'(s)\, ds. \tag{3.253}$$

(b) By the previous problem, $\lim\, [x_n(t) - x_n(a)]$ exists and

$$\lim_{n \to \infty} [x_n(t) - x_n(a)] = \int_a^t \lim_{n \to \infty} x_n{}'(s)\, ds = x(t) - x(a)\,. \tag{3.254}$$

(c) So by the fundamental theorem of calculus [6], we have

$$\lim_{n \to \infty} x_n{}'(t) = x(t). \tag{3.255}$$

20. Using informal arguments, show the following properties of the Dirac delta:

(a) $\delta(-t) = \delta(t)$.

(b)

$$\int_{-\infty}^{\infty} x(t)\frac{d}{dt}u(t)\,dt = x(0). \tag{3.256}$$

(c) Assuming the Dirac is differentiable and that interchange of differentiation and integration is permissable, show that

$$\int_{-\infty}^{\infty} x(t)\frac{d^n}{dt^n}\delta(t)\,dt = (-1)^n\frac{d^n}{dt^n}x(t)\bigg|_{t=0}. \tag{3.257}$$

21. Show that the nth-order B–spline $\beta_n(t)$ has compact support.

22. Let $\beta_n(t)$ be the nth-order B-spline. Show the following [21]:

(a)

$$\frac{d}{dt}\beta_n(t) = \beta_{n-1}\left(t+\frac{1}{2}\right) - \beta_{n-1}\left(t-\frac{1}{2}\right). \tag{3.258}$$

(b)

$$\int_{-\infty}^{t} \beta_n(s)\,ds = \sum_{k=0}^{\infty} \beta_{n+1}\left(t-\frac{1}{2}-k\right). \tag{3.259}$$

23. Assuming that $\|x\|_1 < \infty$ and $\|y\|_\infty < \infty$, show that $\|xy\|_1 \le \|x\|_1\|y\|_\infty$.

24. If $\|x\|_\infty < \infty$ and $\|y\|_\infty < \infty$, show $\|x + y\|_\infty \le \|x\|_\infty + \|y\|_\infty$.

25. Let X be the set of continuous, p-integrable signals x: $[a, b] \to \mathbb{K}$, where $a < b$ and $\mathbb{K}$ is either $\mathbb{R}$ or $\mathbb{C}$. Show that $(X, \|x\|_{p,\,[a,b]})$ is a normed linear space.

26. Let X be a normed linear space. Show that the norm is a continuous map: For any $x \in X$ and $\varepsilon > 0$ there is a $\delta > 0$ such that for any $y \in X$, if $\|y - x\| < \delta$, then $\big|\|y\| - \|x\|\big| < \varepsilon$. Show continuity for the algebraic operations on X: addition and scalar multiplication.

27. Show that the map $d(x, y) = \|x - y\|_p$ is a metric. How must the set of signals $x(t)$ be restricted in order to rigorously show this result? Explain how to remove the restrictions.

28. Suppose $a \sim b$ is an equivalence relation on a set S. Show that $\{[a]: a \in S\}$ partitions S.

29. Consider the analog signals having at most a finite number of discontinuities on $[a, b]$, where $a < b$, and let $1 \le p < \infty$. We restrict ourselves to the Riemann integral, suitably extended to handle piecewise continuous functions.

(a) Show that the set of all such signals does not constitute a normed linear space. In particular, exhibit a signal $x(t)$ which is nonzero and yet $\|x\|_p = 0$.

(b) Show that the relation $x \sim y$ if and only if $\|y\|_p = \|x\|_p$ is an equivalence relation.

(c) Let $[x] = \{y(t): \|y\|_p = \|x\|_p\}$. Show that $[x] = [y]$ if and only if $x(t)$ and $y(t)$ are identical except at a finite number of points.

(d) Define $\underline{L}^p[a, b] = \{[x] \mid x \in L^p[a, b]\}$. Further define an addition operation on these equivalence classes by $[x] + [y] = [s(t)]$, where $s(t) = x(t) + y(t)$. Show that this addition operation makes $\underline{L}^p[a, b]$ into an additive Abelian group: it is commutative, associative, has an identity element, and each element has an additive inverse. Explain the nature of the identity element for $\underline{\mathbf{L}}^p[a, b]$. For a given $[x] \in \underline{L}^p[a, b]$, what is its additive inverse, $-[x]$? Explain.

(e) Define scalar multiplication for $\underline{L}^p[a, b]$ by $c[x] = [cx(t)]$. Show that $\underline{L}^p[a, b]$ thereby becomes a vector space.

(f) Define a norm on $\underline{L}^p[a, b]$ by $\|[x]\|_p = \|x\|_p$. Show that this makes $\underline{L}^p[a, b]$ into a normed linear space.

(g) Apply the same reasoning to $L^p(\mathbb{R})$.

30. Suppose X is a normed linear space and $x, y \in X$. Applying the triangle inequality to the expression $x_n = x_m + x_n - x_m$, show that $\big|\|x_n\| - \|x_m\|\big| \le \|x_n - x_m\|$.

31. Let $S = \{x_n\}$ and $T = \{y_n\}$ be Cauchy sequences in a normed linear space X. Define the relation $S \sim T$ to mean that S and T get arbitrarily close to one another, that is, $\lim_{n\to\infty}\|x_n - y_n\| = 0$.

(a) Show that ~ is an equivalence relation.

(b) Let $[S] = \{T: T \sim S\}$; set $B = \{[S]:\ S = \{x_n\}$ is Cauchy in $X\}$. Define addition and scalar multiplication on B. Show that these operations are well-defined; and show that B is vector space.

(c) Define a norm for B. Show that it is well-defined, and verify each property.

(d) For $x \in X$, define $f(x) = [\{x_n\}]$, where $x_n = x$ for all n, and let $Y = \mathrm{Range}(f)$. Show that $f: X \to Y$ is a normed linear space isometry.

(e) Show that if C is any other Banach space that contains X, then C contains a Banach subspace that is isometric to $Y = f(X)$, where f is given in (d).

32. If X and Y are normed spaces over $\mathbb{K}$ and $T: X \to Y$ is a linear operator, then show the following:

(a) Range(T) is a normed linear space.

(b) The null space of T is a normed linear space.

(c) The inverse map $T^{-1}: \mathrm{Range}(T) \to X$ exists if and only $Tx = 0$ implies $x = 0$.

33. Prove the following alternative version of Schwarz's inequality: If $x, y \in L^2(\mathbb{R})$, then $|\langle x, y\rangle| \le \|x\|_2\|y\|_2$.

34. Suppose $F = \{f_n(t): n \in \mathbb{Z}\}$ is a frame in a Hilbert space H, and T is the frame operator T given by (3.88). Just to review the definition of the frame operator and inner product properties, please show us that T is linear. From its definition, show that the frame adjoint operator is also linear.

35. Suppose that a linear operator U is positive: $U \ge 0$. Show the following [26]:

(a) If I is the identity map, then $U + I$ is invertible.

(b) If V is a positive operator, then $U + V$ is positive.

36. Show that the following are σ-algebras:

(a) $\wp(\mathbb{R})$, the set of all subsets of $\mathbb{R}$.

(b) $\{\emptyset, \mathbb{R}\}$.

37. Let $S \subset \mathcal{P}(\mathbb{R})$. Show that there is a smallest σ-algebra that contains S. (The Borel sets is the class where S is the family of open sets in $\mathbb{R}$.)

38. Let Σ be a σ-algebra. Show Σ is closed under countable intersections.

39. Let $\Sigma = \wp(\mathbb{R})$. Show the following are measures:

(a) $\mu(\emptyset) = 0$ and $\mu(A) = \infty$ if $A \neq \emptyset$.

(b) For $A \in \Sigma$, define $\mu(A) = N$ if A contains exactly N elements and $\mu(A) = \infty$ otherwise.

(c) Show that if μ is a measure on Σ and $c > 0$, then $c\mu$ is also a measure on Σ.

40. Show the following properties of lim inf and lim sup:

(a) $\liminf\{a_n\}$ is the smallest limit point in the sequence $\{a_n\}$.

(b) $\limsup\{a_n\}$ is the largest limit point in the sequence $\{a_n\}$.

(c) $\liminf\{a_n\} = \limsup\{a_n\}$ if and only if the sequence $\{a_n\}$ converges to some limit value $a = \lim_{n\to\infty}\{a_n\}$; show that when this limit exists a = $\liminf\{a_n\} = \limsup\{a_n\}$.

(d) If $a_n \le b_n$, then $\liminf\{a_n\} \le \liminf\{b_n\}$.

(e) Provide an example of strict inequality in the above.

(f) $\limsup\{-a_n\} = -\liminf\{a_n\}$.

41. Show that the general differential equation governing the nth derivative of the Dirac delta is

$$f(t)\delta^{(n)}(t) = \sum_{m=0}^{n} (-1)^m \cdot \frac{n!}{(n-m)!} f^{(m)}(0)\delta^{(n-m)}(t). \tag{3.260}$$

42. Derive the following:

(a) The scaling law for an arbitrary distribution (3.157).

(b) Associativity (3.159).

43. **(a)** Calculate the amplitude for the unit-area of the Abel function (3.230).

(b) Calculate the amplitude for the unit-area of the decaying exponential (3.234).

(c) Graph each of these approximate identities for scales of $a = 1, 10, 100, 1000$.

(d) Verify that the behavior is consistent with the conditions governing an approximate identity.

44. Show that

(a) The rectangular pulse (3.235) is a weight function for any value of the parameter σ.

(b) Verify that this unit-area pulse acts as an approximate identity by explicitly showing

$$\lim_{a \to \infty} \int_{-\infty}^{\infty} f_a(t)\phi(t)\, dt = \phi(0)\,. \tag{3.261}$$

45. Prove the differentiation theorem for convolution of distributions (3.199).

46. Demonstrate the validity (as approximating identities) of the alternative weight functions expressed by (3.224) and (3.225).

The following problems expand on the text's presentation, require some exploratory thinking, and are suitable for extended projects.

47. Let $1 \le p \le \infty$, $x \in L^p(\mathbb{R})$, and $h \in L^1(\mathbb{R})$.

(a) Show that $|y(t)| \le \int |x(t-s)||h(s)|\, ds$;

(b) $y = x * h \in L^p(\mathbb{R})$;

(c) $\|y\|_p \le \|x\|_p \|h\|_1$.

48. Consider a signal analysis matching application, where signal prototypes $P = \{p_1, p_2, ..., p_M\}$ are compared to candidates $C = \{c_1, c_2, ..., c_N\}$, using a distance measure $d(p, c)$. Explain why each of the following properties of $d(p, c)$ are useful to the application design.

(a) $d(p, c) \ge 0$ for all $p \in P$, $c \in C$.

(b) $d(p, c) = 0$ if and only if $p = c$.

(c) $d(p, c) = d(c, p)$ for all p, c.

(d) For any s, $d(p, c) \le d(p, s) + d(s, c)$.

Collectively, these properties make the measure $d(p, c)$ a metric. Consider matching applications where the $d(p,c)$ violates one metric property but obeys the other three. What combinations of of the three properties still suffice for a workable matching application? Explain how the deficiency might be overcome. Provide examples of such deficient match measures. Does deleting a particular metric property provide any benefit to the application—for instance, an ambiguity of analysis that could aid the application?

49. Develop the theory of half-infinite analog signals spaces. Provide formal definitions of the half-infinite L^p spaces: $L^p(-\infty, a]$ and $L^p[a, +\infty)$ and show that these are normed linear spaces. Are they complete? Are they inner product spaces? (You may use the Riemann integral to define L^p for this problem.)

50. Using Lebesgue measure to define it, show that L^∞ is complete [44].

51. Study the matched filtering technique of Section 3.3.2.3 for signal pattern detection. Assume that for computer experiments, we approximate analog convolution with discrete sums.

(a) Show that the the Schwarz inequality implies that the method gives a match measure of unit magnitude if and only if the candidate $x(t)$ and prototype $p(t)$ are constant multiples of one another.

(b) Show that we can generalize the match somewhat by subtracting the mean of each signal before computing the normalized cross-correlation, then the normalized cross-correlation has unit magnitude if and only if the signals are related by $x(t) = Ap(t) + B$, for some constants A, B.

(c) Consider what happens when we neglect to normalize the prototype signal. Show that the matching is still satisfactory but that the maximum match value must be the norm of the prototype pattern, $\|p\|$.

(d) Suppose further that we attempt to build a signal detector without normalizing the prototype. Show that this algorithm may fail because it finds false positives. Explain using examples how the match measure can be larger where $x(t)$ in fact is not a constant multiple of $p(t)$.

(e) What are the computational costs of matched filtering?

(f) How can the computational cost be reduced? What are the effects of various fast correlation methods on the matching performance? Justify your results with both theory and experimentation;

(g) Develop some algorithms and demonstrate with experiments how coarse-to-fine matching can be done using normalized cross-correlation [34].

52. Study the exponential and sinusoidal basis decompositions. Assume the exponential signals constitute an orthonormal basis for $L^2[-\pi, \pi]$.

(a) Show that any $x(t) \in L^2[-\pi, \pi]$ can be expressed as a sum of sinusoidal harmonics. From (3.76) set $a_n = c_n + c_{-n}$ and $jb_n = c_{-n} - c_n$. Show that

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(nt) + \sum_{n=1} b_n \sin(nt). \tag{3.262}$$

(b) Give the spanning set for $L^2[-\pi, \pi]$ implied by (3.262).

(c) Show that the sinusoids are also orthogonal.

(d) By dividing up the real line into 2π-wide segments, $[-\pi + 2n\pi, \pi + 2n\pi]$, give an orthonormal basis for $L^2(\mathbb{R})$.

(e) Consider a square-integrable signal on $L^2(\mathbb{R})$, such as a Gaussian $\exp(-At^2)$ for some $A > 0$. Find a formula, based on the inner product on $L^2[-\pi, \pi]$ for the Fourier series coefficients that arise from the basis elements corresponding to this central interval.

(f) Consider the Fourier series expansion on adjacent intervals, say $[-\pi, \pi]$ and $[\pi, 3\pi]$. Show that the convergence of partial sums of the Fourier series to the signal $x(t)$ exhibits artifacts near the endpoints of the intervals. Explain these anomalies in terms of the periodicity of the exponentials (or sinusoids) on each interval.

(g) Does the selection of sinusoidal or exponential basis functions affect the partial convergence anomaly discovered in (e)? Explain.

(h) What happens if we widen the intervals used for fragmenting $\mathbb{R}$? Can this improve the behavior of the convergence at endpoints of the intervals?

(i) Summarize by explaining why identifying frequency components of general square-integrable signals based on sinusoidal and exponential Fourier series expansions can be problematic.

53. Section 3.3.4.4 argued that frame decompositions support basic signal analysis systems. One desirable property left open by that discussion is to precisely characterize the relation between two sets of coefficients that represent the same incoming signal. This exercise provides a partial solution to the uniqueness problem [42]. Let $F = \{f_n(t) : n \in \mathbb{Z}\}$ be a frame in a Hilbert space H, let $T = T_F$ be its associated frame operator (3.88), and let $S = T^*T$.

(a) If $x \in H$, define $a_n = \langle x, S^{-1}f_n\rangle$, and then show that $x = \Sigma a_n f_n$.

(b) If there are $c_n \in \mathbb{C}$ such that $x = \Sigma c_n f_n$, then

$$\sum_n |c_n|^2 = \sum_n |a_n|^2 + \sum_n |a_n - c_n|^2 . \tag{3.263}$$

(c) Explain how the representation of x in terms of the dual frame for F is optimal in some sense.

(d) Develop an algorithm for deriving this optimal representation.

54. Let $F = \{f_n(t): n \in \mathbb{Z}\}$ be a frame in Hilbert space H. Prove the following:

(a) If an element of the frame is removed, then the reduced sequence is either a frame or not complete (closure of its linear span is everything) in H.

(b) Continuing, let $S = T^*T$, where T^* is the adjoint of the frame operator $T = T_F$. Show that if $\langle f_k, S^{-1}f_k\rangle \neq 1$, then $F \setminus \{f_k\}$ is still a frame.

(c) Finally, prove that if $\langle f_k, S^{-1}f_k\rangle = 1$, then $F \setminus \{f_k\}$ is not complete in H.

55. Let us define some variations on the notion of a basis. If $E = \{e_n\}$ is a sequence in a Hilbert space H, then E is a *basis* if for each $x \in H$ there are unique complex scalars a_n such that x is a series summation, $x = \Sigma_n a_n e_n$. The basis E is *bounded* if $0 \leq \inf\{e_n\} \leq \sup\{e_n\} < \infty$. The basis is *unconditional* if the series converges unconditionally for every element x in H. This last result shows that the uniqueness of the decomposition coefficients in a signal processing system can in fact be guaranteed when the frame is chosen to be *exact*. Let $F = \{f_n(t): n \in \mathbb{Z}\}$ be a sequence in a Hilbert space H. Show that F is an exact frame if and only if it is a bounded unconditional basis for H.

56. Repeat Problem 52, except allow for a frame-based signal decomposition. In what ways does the presence of redundancy positively and negatively affect the resulting decompositions?