Chapter 2
Hilbert Spaces and Orthonormal Systems

The organic unity of mathematics is inherent in the nature of this science, for mathematics is the foundation of all exact knowledge of natural phenomena.

David Hilbert

Hilbert spaces constitute at present the most important examples of Banach spaces, not only because they are the most natural and closest generalization in the realm of "infinite dimensions", of our classical Euclidean geometry, but chiefly for the fact they have been, up to now, the most useful spaces in the applications to functional analysis.

Jean Dieudonné

2.1 Introduction

Historically, the theory of Hilbert spaces originated from David Hilbert's (1862–1943) work on quadratic forms in infinitely many variables with their applications to integral equations. During the period of 1904–1910, Hilbert published a series of six papers, subsequently collected in his classical book *Grundzüge einer allemeinen Theorie der linearen integralgleichungen* published in 1912. It contains many general ideas including Hilbert spaces (ℓ^2 and L^2), the compact operators, and orthogonality, and had a tremendous influence on mathematical analysis and its applications. After many years, John von Neumann (1903–1957) first formulated an axiomatic approach to Hilbert space and developed the modern theory of operators on Hilbert spaces. His remarkable contribution to this area has provided the mathematical foundation of quantum mechanics. Von Neumann's work has also provided an almost definite physical interpretation of quantum mechanics in terms of abstract relations in an infinite dimensional Hilbert spaces. It was shown that observables of a physical system can be represented by linear symmetric operators in a Hilbert space, and the eigenvalues and eigenfunctions of the particular operator that represents energy are energy levels of an electron in an atom and corresponding stationary states of the system. The

L. Debnath, F.A. Shah, *Wavelet Transforms and Their Applications*,
DOI 10.1007/978-0-8176-8418-1_2

differences in two eigenvalues represent the frequencies of the emitted quantum of light and thus define the radiation spectrum of the substance.

The theory of Hilbert spaces plays an important role in the development of wavelet transform analysis. Although a full understanding of the theory of Hilbert spaces is not necessary in later chapters, some familiarity with the basic ideas and results is essential.

One of the nice features of normed spaces is that their geometry is very much similar to the familiar two- and three-dimensional Euclidean geometry. Inner product spaces and Hilbert spaces are even nicer because their geometry is even closer to Euclidean geometry. In fact, the geometry of Hilbert spaces is more or less a generalization of Euclidean geometry to infinite dimensional spaces. The main reason for this simplicity is that the concept of orthogonality can be introduced in any inner product space so that the familiar Pythagorean formula holds. Thus, the structure of Hilbert spaces is more simple and beautiful, and hence, a large number of problems in mathematics, science, and engineering can be successfully treated with geometric methods in Hilbert spaces.

This chapter deals with normed spaces, the L^p spaces, generalized functions (distributions), inner product spaces (also called pre-Hilbert spaces), and Hilbert spaces. The fundamental ideas and results are discussed with special attention given to orthonormal systems, linear functionals, and the Riesz representation theorem. The generalized functions and the above spaces are illustrated by various examples. Separable Hilbert spaces are discussed in Sect. 2.14. Linear operators on a Hilbert space are widely used to represent physical quantities in applied mathematics and physics. In signal processing and wavelet analysis, almost all algorithms are essentially based on linear operators. The most important operators include differential, integral, and matrix operators. In Sect. 2.15, special attention is given to different kinds of operators and their basic properties. The eigenvalues and eigenvectors are discussed in Sect. 2.16. Included are several spectral theorems for self-adjoint compact operators and other related results.

2.2 Normed Spaces

The reader is presumed to have a working knowledge of the real number system and its basic properties. The set of natural numbers (positive integers) is denoted by $\mathbb{N}$, and the set of integers (positive, negative, and zero) is denoted by $\mathbb{Z}$, and the set of rational numbers by $\mathbb{Q}$. We use $\mathbb{R}$ and $\mathbb{C}$ to denote the set of real numbers and the set of complex numbers respectively. Elements of $\mathbb{R}$ and $\mathbb{C}$ are called scalars. Both $\mathbb{R}$ and $\mathbb{C}$ form a scalar field.

We also assume that the reader is familiar with the concept of a linear space or vector space which is an example of mathematical systems that have algebraic structure only. The important examples of linear spaces in mathematics have the real or complex numbers as the scalar field. The simplest example of a real vector space is the set $\mathbb{R}$ of real numbers. Similarly, the set $\mathbb{C}$ of complex numbers is a vector space over the complex numbers.

The concept of *norm* in a vector space is an abstract generalization of the length of a vector in $\mathbb{R}^3$. It is defined axiomatically, that is, any real-valued function satisfying certain conditions is called a norm.

Definition 2.2.1 (Norm). A real-valued function $\|x\|$ defined on a vector space X, where $x \in X$, is called a *norm* on X if the following conditions hold:

(a) $\|x\| = 0$ if and only if $x = 0$,
(b) $\|ax\| = |a|\|x\|$ for every $a \in \mathbb{R}$ and $x \in X$,
(c) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in X$.

Condition (c) is usually called the *triangle inequality*. Since

$$0 = \|0\| = \|x - x\| \leq \|x\| + \|-x\| = 2\|x\|,$$

it follows that $\|x\| \geq 0$ for every $x \in X$.

Definition 2.2.2 (Normed Space). A *normed space* is a vector space X with a given norm.

So, a normed space is a pair $(X, \|.\|)$, where X is a vector space and $\|.\|$ is a norm defined on X. Of course, it is possible to define different norms on the same vector space.

Example 2.2.1. (a) $\mathbb{R}$ is a real normed space with the norm defined by the absolute values, $\|x\| = |x|$.
(b) $\mathbb{C}$ becomes a complex normed space with the norm defined by the modulus, $\|z\| = |z|$.

Example 2.2.2. (a) $\mathbb{R}^N = \{(x_1, x_2, \ldots, x_N) : x_1, x_2, \ldots, x_N \in \mathbb{R}\}$ is a vector space with a norm defined by

$$\|x\| = \sqrt{\left(x_1^2 + x_2^2 + \cdots + x_N^2\right)}, \tag{2.2.1}$$

where $x = (x_1, x_2, \ldots, x_N) \in \mathbb{R}^N$. This norm is often called the *Euclidean norm*.
(b) $\mathbb{C}^N = \{(z_1, z_2, \ldots, z_N) : z_1, z_2, \ldots, z_N \in \mathbb{C}\}$ is a vector space with a norm defined by

$$\|z\| = \sqrt{|z_1|^2 + |z_2|^2 + \cdots + |z_N|^2}, \tag{2.2.2}$$

where $z = (z_1, z_2, \ldots, z_N) \in \mathbb{C}^N$.

Example 2.2.3. The sequence space $\ell^p (1 \leq p < \infty)$ is the set of all sequences $x = \{x_n\}_{n=1}^{\infty}$ of real (complex) numbers such that $\sum_{n=1}^{\infty} |x_n|^p < \infty$ and equipped with the norm

$$\|x\|_p = \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p}. \tag{2.2.3}$$

This space is a normed space.

Example 2.2.4. The vector space $C\big([a,b]\big)$ of continuous functions on the interval $[a,b]$ is a normed space with a norm defined by

$$\|f\| = \left(\int_a^b |f(x)|^2 dx\right)^{1/2}, \tag{2.2.4}$$

or, with a norm defined by

$$\|f\| = \sup_{a \le x \le b} |f(x)|. \tag{2.2.5}$$

Remark. Every normed space $\big(X, \|.\|\big)$ is a metric space (X,d), where the norm induces a metric d defined by

$$d(x,y) = \|x-y\|.$$

But the converse is not necessarily true. In other words, a metric space (X,d) is not necessarily a normed space. This is because of the fact that the metric is not induced by a norm, as seen from the following example.

Example 2.2.5. We denote by s the set of all sequences of real numbers with the metric

$$d(x,y) = \sum_{n=1}^{\infty} \frac{|x_n - y_n|}{2^n\big(1 + |x_n - y_n|\big)}. \tag{2.2.6}$$

This is a metric space, but the metric is not generated by a norm, so the space is not a normed space.

Definition 2.2.3 (Banach Space). A normed space X is called *complete* if every Cauchy sequence in X converges to an element of X. A complete normed space is called a *Banach space*.

Example 2.2.6. The normed spaces $\mathbb{R}^N$ and $\mathbb{C}^N$ with the usual norm as given in Examples 2.2.2(a) and (b) are Banach spaces.

Example 2.2.7. The space of continuous functions $C\big([a,b]\big)$ with the norm defined (2.2.4) is not a complete normed space. Thus, it is not a Banach space.

Example 2.2.8. The sequence space ℓ^p as given in Example 2.2.3 is a Banach space for $p \ge 1$.

Example 2.2.9. The set of all bounded real-valued functions $M\big([a,b]\big)$ on the closed interval $[a,b]$ with the norm (2.2.5) is a complete normed (Banach) space.

This is left for the reader as an exercise.

The following are some important subspaces of $M\big([a,b]\big)$:

(a) $C\big([a,b]\big)$ is the space of continuous functions on the closed interval $[a,b]$,
(b) $D\big([a,b]\big)$ is the space of differentiable functions on $[a,b]$,
(c) $P\big([a,b]\big)$ is the space of polynomials on $[a,b]$,
(d) $R\big([a,b]\big)$ is the space of Riemann integrable functions on $[a,b]$.

Each of these spaces are normed spaces with the norm (2.2.5).

Example 2.2.10. The space of continuously differentiable functions $C' = C'([a,b])$ with the norm

$$\|f\| = \max_{a\le x\le b}\big|f(x)\big| + \max_{a\le x\le b}\big|f'(x)\big| \tag{2.2.7}$$

is a complete normed space.

It is easy to check that this space is complete.

2.3 The L^p Spaces

If $p \ge 1$ is any real number, the vector space of all complex-valued Lebesgue integrable functions f defined on $\mathbb{R}$ is denoted by $L^p(\mathbb{R})$ with a norm

$$\|f\|_p = \left[\int_{-\infty}^{\infty}\big|f(x)\big|^p dx\right]^{1/p} < \infty. \tag{2.3.1}$$

The number $\|f\|_p$ is called the L^p*-norm*. This function space $L^p(\mathbb{R})$ is a Banach space. Since we do not require any knowledge of the Banach space for an understanding of wavelets in this introductory book, the reader needs to know some elementary properties of the L^p-norms.

The L^p spaces for the cases $p = 1$, $p = 2$, $0 < p < 1$, and $1 < p < \infty$ are different in structure, importance, and technique, and these spaces play a very special role in many mathematical investigations.

In particular, $L^1(\mathbb{R})$ is the space of all Lebesgue integrable functions defined on $\mathbb{R}$ with the L^1*-norm* given by

$$\|f\| = \int_{-\infty}^{\infty}\big|f(x)\big|\, dx < \infty. \tag{2.3.2}$$

Definition 2.3.1 (Convergence in Norm). A sequence of functions $f_1, f_2, \ldots, f_n \cdots \in L^1(\mathbb{R})$ is said to converge to a function $f \in L^1(\mathbb{R})$ in norm if $\|f_n - f\|_1 \to 0$ as $n \to \infty$.

So, the convergence in norm is denoted by $f_n \to f$ i.n. This is the usual convergence in a normed space.

Usually, the symbol $\int_{-\infty}^{\infty} f(x)\,dx$ or $\int_{\mathbb{R}} f(x)\,dx$ is used to represent the integral over the entire real line. In applications, we often need to integrate functions over bounded intervals on $\mathbb{R}$. This concept can easily be defined using the integral $\int_{\mathbb{R}} f(x)\,dx$.

Definition 2.3.2 (Integral Over an Interval). The integral of a function f over an interval $[a, b]$ is denoted by

$$\int_a^b f(x)\,dx$$

and defined by

$$\int_a^b f(x)\,\chi_{[a,b]}(x)\,dx, \tag{2.3.3}$$

where $\chi_{[a,b]}$ denotes the *characteristic function* of $[a, b]$ defined by

$$\chi_{[a,b]}(x) = \left\{ \begin{array}{l} 1,\ a \leq x \leq b, \\ 0,\ \text{otherwise} \end{array} \right\} \tag{2.3.4}$$

and $f\,\chi_{[a,b]}$ is the product of two functions.

In other words, $\int_a^b f(x)\,dx$ is the integral of the function equal to f on $[a, b]$ and zero otherwise.

Theorem 2.3.1. *If $f \in L^1(\mathbb{R})$, then the integral $\int_a^b f(x)\,dx$ exists for every interval $[a, b]$.*

The proof is left to the reader as an exercise.

The converse of this theorem is not necessarily true. For example, for the constant function $f = 1$, the integral $\int_a^b f(x)\,dx$ exists for every $-\infty < a < x < b < \infty$, although $f \notin L^1(\mathbb{R})$. This suggests the following definition.

Definition 2.3.3 (Locally Integrable Functions). A function f defined on $\mathbb{R}$ is called *locally integrable* if, for every $-\infty < a < x < b < \infty$, the integral $\int_a^b f(x)\,dx$ exists.

Although this definition requires integrability of f over every bounded interval, it is sufficient to check that the integral $\int_{-n}^{n} f(x)\,dx$ exists for every positive integer n. The proof of this simple fact is left as an exercise.

Note that Theorem 2.3.1 implies that $L^1(\mathbb{R})$ is a subspace of the space of locally integrable functions.

Theorem 2.3.2. *The locally integrable functions form a vector space. The absolute value of a locally integrable function is locally integrable. The product of a locally integrable function and a bounded locally integrable function is a locally integrable function.*

For a proof of this theorem, the reader is referred to Debnath and Mikusinski (1999).

Theorem 2.3.3. *If f is a locally integrable function such that if $|f| \le g$ for some $g \in L^1(\mathbb{R})$, then $f \in L^1(\mathbb{R})$.*

Proof. Let $f_n = f\,\chi_{[a,b]}$ for $n = 1, 2, 3, \ldots$. Then, the sequence of functions $\{f_n\}$ converges to f everywhere and $|f| \le g$ for every $n = 1, 2, \ldots$. Thus, by the Lebesgue dominated convergence theorem, $f \in L^1(\mathbb{R})$.

The function space $L^2(\mathbb{R})$ is the space of all complex-valued Lebesgue integrable functions defined on $\mathbb{R}$ with the $L^2 - norm$ defined by

$$\|f\|_2 = \left[\int_{-\infty}^{\infty} |f(x)|^2 dx\right]^{1/2} < \infty. \tag{2.3.5}$$

Elements of $L^2(\mathbb{R})$ will be called *square integrable functions*. Many functions in physics and engineering, such as wave amplitude in classical or quantum mechanics, are square integrable, and the class of L^2 functions is of fundamental importance.

The space $L^2[a,b]$ is the space of square integrable functions over $[a,b]$ such that $\int_a^b |f(x)|^2 dx$ exists, Thus, the function $x^{-\frac{1}{3}} \in L^2[a,b]$ but $x^{-\frac{2}{3}} \notin L^2[a,b]$.

Remark. The fact that a function belongs to L^p for one particular value of p does not imply that it will belong to L^p for some other value of p.

Example 2.3.1. The function $|x|^{-\frac{1}{2}} e^{-|x|} \in L^1(\mathbb{R})$, but it does not belong to $L^2(\mathbb{R})$. On the other hand, $(1 + |x|)^{-1} \in L^2(\mathbb{R})$, but it does not belong to $L^1(\mathbb{R})$.

Example 2.3.2. Functions $x^n e^{-|x|}$ and $\left(1 + x^2\right)^{-1} \in L^1(\mathbb{R})$ for any integer n.

We add a comment here on the integrability and the local integrability. The condition of integrability is more stringent than local integrability. For example, the functions equal almost everywhere to $|x|^{-\frac{1}{2}}$ and $\left(1+x^2\right)^{-1}$, respectively, are both locally integrable, but only the latter one belongs to $L^1(\mathbb{R})$ because $|x|^{-\frac{1}{2}}$ decays very slowly as $|x| \to \infty$. The additional constraint imposed by integrability over that imposed by local integrability is associated with the nature of the function as $|x| \to \infty$. However, a function $f \in L^1(\mathbb{R})$ does not necessarily decay to zero at infinity. For example, for the function f whose graph consists of an infinite set of rectangular pulses with centers at $x = \pm 1, \pm 2, \ldots, \pm n, \ldots$, the pulse at $x = \pm n$ with height n and width n^{-3}, we obtain that $f \in L^1(\mathbb{R})$, but it does not tend to zero as $|x| \to \infty$.

We make another comment on functions in L^p spaces. If a function f belongs to $L^p(a,b)$ for some value of $p \geq 1$, then it also belongs to $L^q(a,b)$ for all q such that $1 \leq q \leq p$. In other words, raising a function to some power $p > 1$ makes the infinite singularities get "worse" as p is increased. On the other hand, if a function is bounded in $\mathbb{R}$ and belongs to L^p for some $p \geq 1$, then it does belong to L^q for all $q \geq p$. In other words, raising a bounded function to some power p makes its nature at infinity get "better" as far as integrability is concerned. For example, the function $f(x) = \left(1+|x|\right)^{-1} \in L^1(\mathbb{R})$ and is also bounded on $\mathbb{R}$, and also square integrable. However, if the condition of boundedness is relaxed, this result does not hold, even if the function is still locally bounded, that is, it is bounded on every finite interval on $\mathbb{R}$.

Definition 2.3.4 (Convolution). The convolution of two functions $f, g \in L^1(\mathbb{R})$ is defined by

$$\left(f * g\right)(x) = \int_{-\infty}^{\infty} f(x-y)\, g(y)\, dy \tag{2.3.6}$$

which exists for all $x \in \mathbb{R}$ or at least almost everywhere. Then, it defines a function which is called the *convolution* of f and g and is denoted by $f * g$.

We next discuss some basic properties of the convolution.

Theorem 2.3.4. *If $f, g \in L^1(\mathbb{R})$, then the function $f(x-y)\, g(y)$ is integrable for almost all $x \in \mathbb{R}$. Furthermore, the convolution*

$$\left(f * g\right)(x) = \int_{-\infty}^{\infty} f(x-y)\, g(y)\, dy \tag{2.3.7}$$

is an integrable function and $(f * g) \in L^1(\mathbb{R})$ and the following inequality holds:

$$\left\| f * g \right\|_1 \leq \left\| f \right\|_1 \left\| g \right\|_1. \tag{2.3.8}$$

Proof. We refer to Debnath and Mikusinski (1999) for the proof of the first part of the theorem, that is, $(f * g) \in L^1(\mathbb{R})$.

To prove inequality (2.3.8), we proceed as follows:

$$\begin{aligned}
\|f * g\|_1 &= \int_{-\infty}^{\infty} |f * g|\,dx = \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} f(x-y)\,g(y)\,dy \right| dx \\
&\le \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f(x-y)||g(y)\,dy|\,dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f(x-y)||g(y)|dy\,dx, \quad \text{by Fubini's Theorem} \\
&= \int_{-\infty}^{\infty} |f(x-y)|dx \int_{-\infty}^{\infty} |g(y)|\,dy \\
&= \int_{-\infty}^{\infty} |f(x)|dx \int_{-\infty}^{\infty} |g(y)|\,dy = \|f\|_1 \|g\|_1.
\end{aligned}$$

Thus, the proof is complete.

Theorem 2.3.5. *If $f, g \in L^1(\mathbb{R})$, then the convolution is commutative, that is,*

$$\big(f * g\big)(x) = \big(g * f\big)(x). \tag{2.3.9}$$

The proof follows easily by the change of variables.

Theorem 2.3.6. *If $f, g, h \in L^1(\mathbb{R})$, then the following properties hold:*

(a)

$$(f * g) * h = f * (g * h) \qquad \textit{(associative)}, \tag{2.3.10}$$

(b)

$$(f + g) * h = f * h + g * h \qquad \textit{(distributive)}. \tag{2.3.11}$$

We use Fubini's theorem to prove that the convolution is associative. We have

$$\begin{aligned}
\big(f * g\big) * h = (g * f) * h(x) &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g(x-z-y)\,f(y)\,dy \right] h(z)\,dz \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y)\,g(x-z-y)h(z)\,dz\,dy \\
&= f * \big(g * h\big)(x).
\end{aligned}$$

The proof of part (b) is left to the reader as an exercise.

Remarks. The properties of convolution just described above shows that the $L^1(\mathbb{R})$ is a commutative Banach algebra under ordinary addition, multiplication defined by convolution, and $\|.\|_1$ as norm. This Banach algebra is also referred to as the L^1-algebra on $\mathbb{R}$.

Theorem 2.3.7. *If f is an integrable function and g is a bounded locally integrable function, then the convolution $f * g$ is a continuous function.*

Proof. First, note that since $\left|f(x-y)\,g(y)\right| \le M\left|f(x-y)\right|$ for some constant M and every x, the integral $\int_{-\infty}^{\infty} f(x-y)\,g(y)\,dy$ is defined at every $x \in \mathbb{R}$ by Theorem 2.3.3. Next, we show that $f * g$ is a continuous function.

For any $x, h \in \mathbb{R}$, we have

$$\begin{aligned}
\left|(f*g)(x+h)-(f*g)(x)\right| &= \left|\int_{-\infty}^{\infty} f(x+h-y)\,g(y)\,dy - \int_{-\infty}^{\infty} f(x-y)\,g(y)\,dy\right| \\
&= \left|\int_{-\infty}^{\infty} \left[f(x+h-y)-f(x-y)\right]g(y)\,dy\right| \\
&\le \int_{-\infty}^{\infty} \left|f(x+h-y)-f(x-y)\right|\left|g(y)\right|dy \\
&\le M\int_{-\infty}^{\infty} \left|f(0+h-y)-f(0-y)\right|dy,
\end{aligned}$$

which tends to zero as $h \to 0$ since

$$\lim_{h\to 0}\int_{-\infty}^{\infty} \left|f(h-y)-f(-y)\right|dy = 0.$$

Thus, the proof is complete.

2.4 Generalized Functions with Examples

The Dirac delta function $\delta(x)$ is the best known of a class of entities called *generalized functions*. The generalized functions are the natural mathematical quantities which are used to describe many abstract notions which occur in the physical sciences. The impulsive force, the point mass, the point charge, the point dipole, and the frequency response of a harmonic oscillator in a nondissipating medium are all aptly represented by generalized functions. The generalized functions play an important role in the Fourier transform analysis, and they can resolve the inherent difficulties that occur in classical mathematical analysis. For example, every locally integrable function (and indeed every generalized function) can be considered as the integral of some generalized function and thus becomes infinitely differentiable in the new sense. Many sequences of functions which do not converge in the ordinary sense to a limit function can be found to converge to a generalized function. Thus, in many ways the idea of generalized functions not only simplifies the rules of mathematical analysis but also becomes very useful in the physical sciences.

In order to give a sound mathematical formulation of quantum mechanics, Dirac in 1920 introduced the delta function $\delta(x)$ having the following properties:

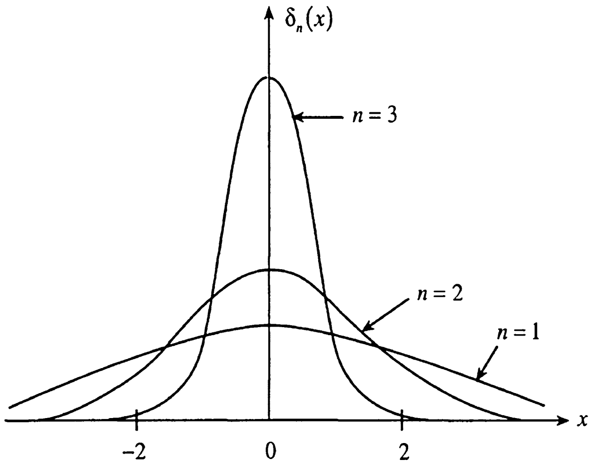

Fig. 2.1 The sequence of functions $\{\boldsymbol{\delta_n(x)}\}$ for $\mathbf{n = 1, 2, 3}\ldots$

$$\left.\begin{aligned} \delta(x) &= 0, \quad x \neq 0 \\ \int_{-\infty}^{\infty} \delta(x)\, dx &= 1 \end{aligned}\right\}. \tag{2.4.1}$$

These properties cannot be satisfied by any ordinary function in classical mathematics. Hence, the delta function is not really a function in the classical sense. However, it can be regarded as the limit of a sequence of ordinary functions. A good example of such a sequence $\delta_n(x)$ is a sequence of Gaussian functions given by

$$\delta_n(x) = \sqrt{\frac{n}{\pi}} \exp\left(-nx^2\right). \tag{2.4.2}$$

Clearly, $\delta_n(x) \to 0$ as $n \to \infty$ for any $x \neq 0$ and $\delta_n(x) \to \infty$ as $n \to -\infty$, as shown in Fig. 2.1. Also, for all $n = 1, 2, 3, \ldots,$

$$\int_{-\infty}^{\infty} \delta_n(x)\, dx = 1$$

and

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} \delta_n(x)\, dx = \int_{-\infty}^{\infty} \lim_{n\to\infty} \delta_n(x)\, dx = \int_{-\infty}^{\infty} \delta(x)\, dx = 1. \tag{2.4.3}$$

Thus, the Dirac delta function can be regarded as the limit of sequence $\delta_n(x)$ of ordinary functions, and we write

$$\delta(x) = \lim_{n\to\infty} \delta_n(x) = \lim_{n\to\infty} \sqrt{\frac{n}{\pi}} \exp\left(-nx^2\right). \tag{2.4.4}$$

This approach of defining new entities, such as $\delta(x)$, which do not exist as ordinary functions becomes meaningful mathematically and useful from a physical point of view.

Another alternative definition is based on the idea that if a function f is continuous at $x = a$, then $\delta(x)$ is defined by its fundamental property

$$\int_{-\infty}^{\infty} f(x)\,\delta(x-a)\,dx = f(a). \tag{2.4.5}$$

Or, equivalently,

$$\int_{-\infty}^{\infty} f(x)\,\delta(x)\,dx = f(0). \tag{2.4.6}$$

This is a rather more formal approach pioneered by Laurent Schwartz in the late 1940s. Thus, the concept of the delta function is clear and simple in modern mathematics. It has become very useful in science and engineering. Physically, the delta function represents a point mass, that is, a particle of unit mass is located at the origin. This means that a point particle can be regarded as the limit of a sequence of continuous mass distribution. The Dirac delta function is also interpreted as a probability measure in terms of the formula (2.4.5).

Definition 2.4.1 (Support of a Function). The support of a function $f : \mathbb{R} \to \mathbb{C}$ is $\{x : f(x) \neq 0\}$ and denoted by $\mathrm{supp}(f)$. A function has *bounded support* if there are two real numbers a, b such that $\mathrm{supp}(f) \subset (a, b)$. By a *compact support*, we mean a closed and bounded support.

Definition 2.4.2 (Smooth or Infinitely Differentiable Function). A function $f : \mathbb{R} \to \mathbb{C}$ is called *smooth* or *infinitely differentiable* if its derivatives of all orders exist and are continuous.

A function $f : \mathbb{R} \to \mathbb{C}$ is said to be *n-times continuously differentiable* if its first n derivatives exist and are continuous.

Definition 2.4.3 (Test Functions). A *test function* is an infinitely differentiable function on $\mathbb{R}$ whose support is compact. The space of all test functions is denoted by $\mathscr{D}(\mathbb{R})$ or simply by $\mathscr{D}$. The graph of a "typical" test function is shown in Fig. 2.2.

Since smooth (infinitely differentiable) functions are continuous and the support of a continuous function is always closed, test functions can be equivalently defined as follows: ϕ is a test function if it is a smooth function vanishing outside a bounded set.

Example 2.4.1. A function ϕ defined by

$$\phi(x) = \begin{cases} \exp\left[\left(x^2 - a^2\right)^{-1}\right], & \text{for } |x| < a, \\ 0, & \text{otherwise} \end{cases} \tag{2.4.7}$$

is a test function with support $(-a, a)$.

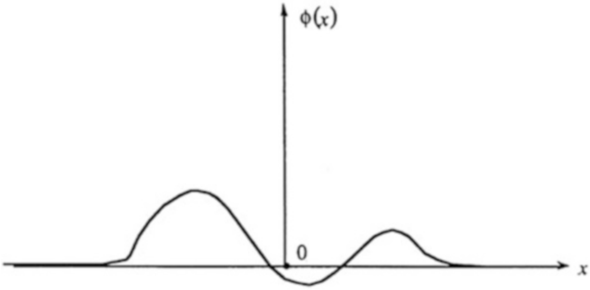

Fig. 2.2 A typical test function

Using this test function, we can easily generate a number of examples. The following are test functions:

$\phi(ax + b)$, a, b are constants and $a \neq 0$,
$f(x)\,\phi(x)$, f is an arbitrary smooth function,
$\phi^{(n)}(x)$, n is a positive integer.

Definition 2.4.4 (Convergence of Test Functions). Suppose $\{\phi_n\}$ is a sequence of test functions and ϕ is another test function. We say that the sequence $\{\phi_n\}$ converges to ϕ in $\mathscr{D}$, denoted by $\phi_n \stackrel{\mathscr{D}}{\to} \phi$, if the following two conditions are satisfied:

(a) $\phi_1, \phi_2, \ldots, \phi_n, \ldots$ and ϕ vanish outside some bounded interval $[a, b] \subset \mathbb{R}$,
(b) for each k, $\phi_n(x) \to \phi(x)$ as $n \to \infty$ uniformly for some $x \in [a, b]$, where $\phi^{(k)}(x)$ denotes the kth derivative of ϕ.

Definition 2.4.5 (Generalized Function or Distribution). A continuous linear functional F on $\mathscr{D}$ is called a *generalized function* or *distribution*. In other words, a mapping $F : \mathscr{D} \to \mathbb{C}$ is called a *generalized function* or *distribution* if

(a) $F(a\phi + b\psi) = aF(\phi) + bF(\psi)$ for every $a, b, \in \mathbb{C}$ and $\phi, \psi \in \mathscr{D}(\mathbb{R})$,
(b) $F(\phi_n) \to F(\phi)$ (in $\mathbb{C}$) whenever $\phi_n \to \phi$ in $\mathscr{D}$.

The space of all generalized functions is denoted by $\mathscr{D}'(\mathbb{R})$ or simply by $\mathscr{D}'$. It is convenient to write (F, ϕ) instead of $F(\phi)$.

Distributions generalize the concept of a function. Formally, a function on $\mathbb{R}$ is not a distribution because its domain is not $\mathscr{D}$. However, every locally integrable function f on $\mathbb{R}$ can be identified with a distribution F defined by

$$(F, \phi) = \int_{\mathbb{R}} f(x)\,\phi(x)\,dx. \tag{2.4.8}$$

The distribution F is said to be *generated* by the function f.

Definition 2.4.6 (Regular and Singular Distributions). A distribution $F \in \mathscr{D}'$ is called a *regular distribution* if there exists a locally integrable function f such that

$$(F, \phi) = \int_{\mathbb{R}} f(x)\, \phi(x)\, dx \tag{2.4.9}$$

for every $\phi \in \mathscr{D}$. A distribution that is not regular is called a *singular distribution*.

The fact that (2.4.9) defines a distribution is because of the following results.

First, the product $f\phi$ is integrable because it vanishes outside a compact support $[a, b]$. In other words,

$$(F, \phi) = \int_{-\infty}^{\infty} f(x)\, \phi(x)\, dx = \int_a^b f(x)\, \phi(x)\, dx$$

exists. Hence, F is a linear functional on $\mathscr{D}$. Also,

$$\begin{aligned} \left|(F, \phi_n) - (F, \phi)\right| &= \left|\int_a^b \left[\phi_n(x) - \phi(x)\right] f(x) dx\right| \\ &\le \int_a^b \left|\phi_n(x) - \phi(x)\right| \left|f(x)\right| dx \\ &\le \max \left|\phi_n(x) - \phi(x)\right| \int_a^b \left|f(x)\right| dx \to 0 \quad \text{as } n \to \infty, \end{aligned}$$

because $\phi_n \to \phi$ uniformly. Hence,

$$(F, \phi_n) \to (F, \phi) \quad \text{as } n \to \infty.$$

This means that F is a continuous linear functional, that is, F is a distribution.

Thus, the class of generalized functions contains elements which corresponds to ordinary functions as well as singular distributions. We now give an interpretation of (F, ϕ).

The integral $\int_{\mathbb{R}} f(x)\phi(x)\, dx$ in (2.4.9) can be interpreted, at least for some test function ϕ, as the average value of f with respect to probability whose density function is ϕ. Thus, (F, ϕ) can be regarded as an average value of F and of distributions as entities that have average values in neighborhoods of every point. However, in general, distributions may not have values at points. This interpretation is very natural from a physical point of view. In fact, when a quantity is measured, the result is not the exact value at a single point.

Example 2.4.2. If Ω is an open set in $\mathbb{R}$, then the functional F defined by

$$(F, \phi) = \int_{\Omega} \phi(x)\, dx \tag{2.4.10}$$

is a distribution. Note that it is a regular distribution since

$$(F, \phi) = \int_{-\infty}^{\infty} \phi(x)\, \chi_\Omega(x)\, dx, \tag{2.4.11}$$

where χ_Ω is the characteristic function of the set Ω.

In particular, if $\Omega = (0, \infty)$, we obtain a distribution

$$(H, \phi) = \int_0^{\infty} \phi(x)\, dx \tag{2.4.12}$$

which is called the *Heaviside function*. The symbol H is used to denote this distribution as well as the characteristic function of $\Omega = (0, \infty)$.

Example 2.4.3 (Dirac Distribution). One of the most important examples of generalized functions is the so-called Dirac delta function or, more precisely, the *Dirac distribution*. It is denoted by δ and defined by

$$(\delta, \phi) = \int_{-\infty}^{\infty} \phi(x)\delta(x)\, dx = \phi(0). \tag{2.4.13}$$

The linearity of δ is obvious. To prove the continuity, note that $\phi_n \to \phi$ in $\mathscr{D}$ implies that $\phi_n \to \phi$ uniformly on $\mathbb{R}$ and hence $\phi_n(x) \to \phi(x)$ for every $x \in \mathbb{R}$. This implies that the Dirac delta function is a singular distribution.

Example 2.4.4.

(a)

$$\big(\delta(x-a), \phi\big) = \big(\delta(x), \phi(x+a)\big) = \phi(a). \tag{2.4.14}$$

(b)

$$\big(\delta(ax), \phi\big) = \frac{1}{|a|}\, \phi(0). \tag{2.4.15}$$

We have

$$\begin{aligned}\big(\delta(x-a), \phi\big) &= \int_{-\infty}^{\infty} \delta(x-a)\, \phi(x)\, dx \\ &= \int_{-\infty}^{\infty} \delta(y)\, \phi(y+a)\, dy = \phi(a).\end{aligned}$$

This is called the *shifting property* of the delta function.

Similarly,

$$\big(\delta(ax), \phi\big) = \int_{-\infty}^{\infty} \delta(ax)\, \phi(x)\, dx = \int_{-\infty}^{\infty} \delta(y)\, \phi\left(\frac{y}{a}\right) \frac{dy}{a} = \frac{1}{a}\, \phi(0).$$

Hence, for $a \neq 0$,

$$\delta(ax) = \frac{1}{|a|}\,\phi(0). \tag{2.4.16}$$

The success of the theory of distributions is essentially due to the fact that most concepts of ordinary calculus can be defined for distributions. While adopting definitions and rules for distributions, we expect that new definitions and rules will agree with classical ones when applied to regular distributions. When looking for an extension of some operation A, which is defined for ordinary functions, we consider regular distributions defined by (2.4.9). Since we expect AF to be the same as Af, it is natural to define

$$(AF, \phi) = \int_{\mathbb{R}} Af(x)\,\phi(x)\,dx.$$

If there exists a continuous operation A^* which maps $\mathscr{D}$ into $\mathscr{D}$ such that

$$\int Af(x)\,\phi(x)\,dx = \int f(x)\,A^*\phi(x)\,dx,$$

then it makes sense to introduce, for an arbitrary distribution F,

$$(AF, \phi) = (F, A^*\phi).$$

If this idea is used to give a natural definition of a derivative of a distribution, it suffices to observe

$$\int_{\mathbb{R}} \left\{\frac{\partial}{\partial x} f(x)\right\} \phi(x)\,dx = -\int_{\mathbb{R}} f(x) \frac{\partial}{\partial x}\,\phi(x)\,dx.$$

Definition 2.4.7 (Derivatives of a Distribution). The derivative of a distribution F is a distribution F' defined by

$$\left(\frac{dF}{dx}, \phi\right) = -\left(F, \frac{d\phi}{dx}\right). \tag{2.4.17}$$

This result follows by integrating by parts. In fact, we find

$$\left(\frac{dF}{dx}, \phi\right) = \int_{-\infty}^{\infty} \frac{dF}{dx}\,\phi(x)\,dx = \big[F(x)\phi(x)\big]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} F(x)\phi'(x)\,dx = -\big(F, \phi'(x)\big),$$

where the first term vanishes because ϕ vanishes at infinity.

More generally,

$$\big(F^{(k)}, \phi\big) = (-1)^k \big(F, \phi^{(k)}\big), \tag{2.4.18}$$

where $F^{(k)}(x)$ is the kth derivative of distribution F.

Thus, the extension of the idea of a function to that of a distribution has a major success in the sense that every distribution has derivatives of all orders which are again distributions.

Example 2.4.5 (Derivative of the Heaviside Function).
(a)

$$H'(x) = \delta(x). \tag{2.4.19}$$

We have

$$\begin{aligned}\left(H', \phi\right) &= \int_0^\infty H'(x)\,\phi(x)\,dx = \left[H(x)\phi(x)\right]_0^\infty - \int_0^\infty H(x)\,\phi'(x)\,dx \\ &= -\int_0^\infty \phi'(x)\,dx = \phi(0) = (\delta, \phi), \quad \text{since } \phi \text{ vanishes at infinity.}\end{aligned}$$

This proves the result.

(b) *(Derivatives of the Dirac Delta Function).*

$$(\delta', \phi) = -(\delta, \phi') = -\phi'(0), \tag{2.4.20}$$

$$\left(\delta^{(n)}, \phi\right) = (-1)^n \phi^{(n)}(0). \tag{2.4.21}$$

We have

$$(\delta', \phi) = \int_{-\infty}^{\infty} \delta'(x)\,\phi(x)\,dx = \left[\delta(x)\phi(x)\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \delta(x)\,\phi'(x)\,dx = -\phi'(0),$$

since ϕ vanishes at infinity.

Result (2.4.21) follows from a similar argument.

Example 2.4.6. If h is a smooth function and F is a distribution, then the derivative of the product (hF) is given by

$$(hF)' = hF' + h'F. \tag{2.4.22}$$

We have, for any $\phi \in \mathscr{D}$,

$$\begin{aligned}\left((hF)', \phi\right) &= -\left(hF, \phi'\right) \\ &= -\left(F, h\phi'\right) \\ &= -\left(F, (h\phi)' - h'\phi\right) \\ &= \left(F', h\phi\right) + \left(F, h'\phi\right) \\ &= \left(hF', \phi\right) + \left(h'F, \phi\right) \\ &= \left(hF' + h'F, \phi\right).\end{aligned}$$

This proves the result.

Example 2.4.7. The function $|x|$ is locally integrable and differentiable for all $x \neq 0$ but certainly not differentiable at $x = 0$. The generalized derivative can be calculated as follows.

For any test function ϕ, we have

$$\left(|x|', \phi\right) = -(|x|, \phi')$$
$$= -\int_{-\infty}^{\infty} |x|\,\phi'(x)\,dx = \int_{-\infty}^{0} x\,\phi'(x)\,dx - \int_{0}^{\infty} x\,\phi'(x)\,dx$$

which is, integrating by parts and using the fact that ϕ vanishes at infinity,

$$= -\int_{-\infty}^{0} \phi(x)\,dx + \int_{0}^{\infty} \phi(x)\,dx. \tag{2.4.23}$$

Thus, we can write (2.4.23) in the form

$$\left(|x|', \phi\right) = \int_{-\infty}^{\infty} \operatorname{sgn}(x)\phi(x)\,dx = (\operatorname{sgn}, \phi) \quad \text{for al } \phi \in \mathscr{D}.$$

Therefore,

$$|x|' = \operatorname{sgn}(x), \tag{2.4.24}$$

where sgn (x) is called the *sign function*, defined by

$$\operatorname{sgn}(x) = \begin{cases} 1, & x > 0, \\ -1, & x < 0, \end{cases} \tag{2.4.25}$$

Obviously,

$$H(x) = \frac{1}{2}(1 + \operatorname{sgn} x). \tag{2.4.26}$$

Or, equivalently,

$$\operatorname{sgn} x = 2H(x) - 1. \tag{2.4.27}$$

Thus,

$$\frac{d}{dx}(\operatorname{sgn} x) = 2H'(x) = 2\,\delta(x). \tag{2.4.28}$$

Definition 2.4.8 (Antiderivative of a Distribution). If F is a distribution on $\mathbb{R}$ and $F \in \mathscr{D}'(\mathbb{R})$, a distribution G on $\mathbb{R}$ is called an *antiderivative* of F if $G' = F$.

Theorem 2.4.1. *Every distribution has an antiderivative.*

Proof. Suppose $\phi_0 \in \mathscr{D}(\mathbb{R})$ is a fixed test function such that

$$\int_{-\infty}^{\infty} \phi_0(x)\, dx = 1. \tag{2.4.29}$$

Then, for every test function $\phi_0 \in \mathscr{D}(\mathbb{R})$, there exists a test function $\phi_1 \in \mathscr{D}(\mathbb{R})$ such that $\phi = K\phi_0 + \phi_1$, where

$$K = \int_{-\infty}^{\infty} \phi(x)\, dx \quad \text{and} \quad \int_{-\infty}^{\infty} \phi_1(x)\, dx = 0.$$

Suppose $F \in \mathscr{D}'(\mathbb{R})$. We define a functional G on $\mathscr{D}(\mathbb{R})$ by

$$\big(G, \phi\big) = \big(G, K\phi_0 + \phi_1\big) = CK - \big(F, \psi\big),$$

where C is a constant and ψ is a test defined by

$$\psi(x) = \int_{-\infty}^{x} \phi_1(t)\, dt.$$

Then, G is a distribution and $G' = F$.

We close this section by adding an example of application to partial differential equations.

Consider a partial differential operator L of order m in N variables

$$L = \sum_{|\alpha| \le m} A_\alpha D^\alpha, \tag{2.4.30}$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)$ is a multi-index, the α_n's are nonnegative integers, $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_N$, $A_\alpha = A_{\alpha_1, \alpha_2, \ldots, \alpha_N}(x_1, x_2, \ldots, x_N)$ are functions on $\mathbb{R}^N$ (possibly constant), and

$$D^\alpha = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_N}\right)^{\alpha_N} = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \ldots \partial x_N^{\alpha_N}}. \tag{2.4.31}$$

Equations of the form

$$LG = \delta \tag{2.4.32}$$

are of particular interest. Suppose G is a distribution which satisfies (2.4.32). Then, for any distribution f with compact support, the convolution $(f * G)$ is a solution of the partial differential equation

$$Lu = f. \tag{2.4.33}$$

We have

$$\begin{aligned} L(f * G) &= \sum_{|\alpha| \le m} A_\alpha D^\alpha \big(f * G\big) \\ &= \sum_{|\alpha| \le m} A_\alpha \big(f * D^\alpha G\big) \\ &= f * \left(\sum_{|\alpha| \le m} A_\alpha D^\alpha G \right) = f * LG \\ &= f * \delta = f. \end{aligned}$$

This explains the importance of the equation $Lu = \delta$, at least in the context of the existence of solutions of partial differential equations.

2.5 Definition and Examples of an Inner Product Space

Definition 2.5.1 (Inner Product Space). A (real or complex) inner product space is a (real or complex) vector space X with an *inner product* defined in X as a mapping

$$\langle .,. \rangle : X \times X \to \mathbb{C}$$

such that, for any $x, y, z \in X$ and $\alpha, \beta \in \mathbb{C}$ (a set of complex numbers), the following conditions are satisfied:

(a) $\langle x, y \rangle = \overline{\langle y, x \rangle}$ (the bar denotes the complex conjugate),
(b) $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$,
(c) $\langle x, x \rangle \ge 0$, and $\langle x, x \rangle = 0$ implies $x = 0$.

Clearly, an inner product space is a vector space with an inner product specified. Often, an inner product space is called a *pre-Hilbert space* or a *unitary space*.

According to the above definition, the inner product of two vectors is a complex number. The reader should be aware that other symbols are sometimes used to denote the inner product: (x, y) or $\langle x/y \rangle$. Instead of $\bar{z}$, the symbol z^* is also used for the complex conjugate. In this book, we will use $\langle x, y \rangle$ and $\bar{z}$.

By (a), $\langle x, x \rangle = \overline{\langle x, x \rangle}$ which means that $\langle x, x \rangle$ is a real number for every $x \in X$. It follows from (b) that

$$\langle x, \alpha y + \beta z \rangle = \overline{\langle \alpha y + \beta z, x \rangle} = \overline{\alpha \langle y, x \rangle + \beta \langle z, x \rangle} = \bar{\alpha} \langle x, y \rangle + \bar{\beta} \langle x, z \rangle.$$

In particular,

$$\langle \alpha x, y \rangle = \alpha \langle x, y \rangle \quad \text{and} \quad \langle x, \alpha y \rangle = \bar{\alpha} \langle x, y \rangle.$$

Hence, if $\alpha = 0$,

$$\langle 0, y\rangle = \langle x, 0\rangle = 0.$$

The algebraic properties (a) and (b) are generally the same as those governing the scalar product in ordinary vector algebra with which the reader should be familiar. The only property that is not obvious is that in a complex space the inner product is not linear but *conjugate linear* with respect to the second factor; that is, $\langle x, \alpha y\rangle = \bar{\alpha}\,\langle x, y\rangle$.

Example 2.5.1. The simplest but important example of an inner product space is the space of complex numbers $\mathbb{C}$. The inner product in $\mathbb{C}$ is defined by $\langle x, y\rangle = x\bar{y}$.

Example 2.5.2. The space $\mathbb{C}^N$ of ordered N-tuples $x = (x_1, \ldots, x_N)$ of complex numbers, with the inner product defined by

$$\langle x, y\rangle = \sum_{k=1}^{N} x_k\,\bar{y}_k, \quad x = (x_1, \ldots, x_N), \quad y = (y_1, \ldots, y_N),$$

is an inner product space.

Example 2.5.3. The space ℓ^2 of all infinite sequences of complex numbers $\{x_k\}$ such that $\sum_{k=1}^{\infty} |x_k|^2 < \infty$ with the inner product defined by

$$\langle x, y\rangle = \sum_{k=1}^{\infty} x_k\,\bar{y}_k, \quad \text{where} \quad x = (x_1, x_2, x_3, \ldots), \quad y = (y_1, y_2, y_3, \ldots),$$

is an infinite dimensional inner product space. As we will see later, this space is one of the most important examples of an inner product space.

Example 2.5.4. Consider the space of infinite sequences $\{x_n\}$ of complex numbers such that only a finite number of terms are nonzero. This is an inner product space with the inner product defined as in Example 2.5.3.

Example 2.5.5. The space $\mathscr{C}\big([a, b]\big)$ of all continuous complex-valued functions on the interval $[a, b]$ with the inner product

$$\langle f, g\rangle = \int_a^b f(x)\,\overline{g(x)}\,dx \tag{2.5.1}$$

is an inner product space.

Example 2.5.6 (The Space of Square Integrable Functions). The function space $L^2\big([a, b]\big)$ of all complex-valued Lebesgue square integrable functions on the interval $[a, b]$ with the inner product defined by (2.5.1) is an inner product space.

Similarly, the function space $L^2(\mathbb{R})$ is also an inner product space with the inner product defined by

$$\langle f, g\rangle = \int_a^b f(x)\,\overline{g(x)}\,dx, \tag{2.5.2}$$

where $f, g \in L^2(\mathbb{R})$.

Since

$$fg = \frac{1}{4}\left[(f+g)^2 - (f-g)^2\right],$$

and

$$|fg| \le \frac{1}{2}\left(|f|^2 + |g|^2\right),$$

it follows that $f, g \in L^1(\mathbb{R})$.

Furthermore,

$$|f+g|^2 \le |f|^2 + 2|fg| + |g|^2.$$

Integrating this inequality over $\mathbb{R}$ shows that $(f+g) \in L^2(\mathbb{R})$.

It can be shown that $L^2(\mathbb{R})$ is a complete normed space with the norm induced by (2.5.2), that is,

$$\|f\|_2 = \left\{\int_{\mathbb{R}} |f(x)|^2 dx\right\}^{\frac{1}{2}}. \tag{2.5.3}$$

This is exactly the L^2-norm defined by (2.3.1). Both spaces $L^2([a,b])$ and $L^2(\mathbb{R})$ are of special importance in theory and applications.

Example 2.5.7. Suppose D is a compact set in $\mathbb{R}^3$ and $X = C^2(D)$ is the space of complex-valued functions that have continuous second partial derivatives in D. If $u \in D$, we assume

$$\nabla u = \left(\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \frac{\partial u}{\partial x_3}\right). \tag{2.5.4}$$

We define the inner product by the integral

$$\langle u, v\rangle = \int_D \left[u\bar{v} + \frac{\partial u}{\partial x_1}\cdot\frac{\partial \bar{v}}{\partial x_1} + \frac{\partial u}{\partial x_2}\cdot\frac{\partial \bar{v}}{\partial x_2} + \frac{\partial u}{\partial x_3}\cdot\frac{\partial \bar{v}}{\partial x_3}\right] dx, \tag{2.5.5}$$

where $x = (x_1, x_2, x_3)$.

Clearly, this is linear in u and also $\langle u, v\rangle = \overline{\langle v, u\rangle}$ and $\langle u, v\rangle \geq 0$. Furthermore, if $\langle u, u\rangle = 0$, then $\int_D |u|^2 dx = 0$. Since u is continuous, this means that $u = 0$. Hence, (2.5.5) defines an inner product in the space X. Obviously, the norm is given by

$$\|u\| = \left(\int_D \left(|u|^2 + |\nabla u|^2\right) dx\right)^{\frac{1}{2}}, \tag{2.5.6}$$

where

$$|\nabla u|^2 = \left|\frac{\partial u}{\partial x_1}\right|^2 + \left|\frac{\partial u}{\partial x_2}\right|^2 + \left|\frac{\partial u}{\partial x_3}\right|^2. \tag{2.5.7}$$

2.6 Norm in an Inner Product Space

An inner product space is a vector space with an inner product. It turns out that every inner product space is also a normed space with the norm defined by

$$\|x\| = \sqrt{\langle x, x\rangle}.$$

First notice that the norm is well defined because $\langle x, x\rangle$ is always a nonnegative (real) number. Condition (c) of Definition 2.5.1 implies that $\|x\| = 0$ if and only if $x = 0$. Moreover,

$$\|\lambda x\| = \sqrt{\langle \lambda x, \lambda x\rangle} = \sqrt{\lambda\bar{\lambda}\langle x, x\rangle} = |\lambda|\|x\|.$$

It thus remains to prove the triangle inequality. This is not as simple as the first two conditions. We first prove the so-called Schwarz's inequality, which will be used in the proof of the triangle inequality.

Theorem 2.6.1 (Schwarz's Inequality). *For any two elements x and y of an inner product space, we have*

$$|\langle x, y\rangle| \leq \|x\|\|y\|. \tag{2.6.1}$$

The equality $|\langle x, y\rangle| = \|x\|\|y\|$ holds if and only if x and y are linearly dependent.

Proof. If $y = 0$, then (2.6.1) is satisfied because both sides are equal to zero. Assume then $y \neq 0$. By (c) in Definition 2.5.1, we have

$$0 \leq \langle x + \alpha y, x + \alpha y\rangle = \langle x, x\rangle + \bar{\alpha}\langle x, y\rangle + \alpha\langle y, x\rangle + |\alpha|^2\langle y, y\rangle. \tag{2.6.2}$$

Now, put $\alpha = -\langle x, y\rangle / \langle y, y\rangle$ in (2.6.2) and then multiply by $\langle y, y\rangle$ to obtain

$$0 \le \langle x, x\rangle\langle y, y\rangle - \big|\langle x, y\rangle\big|^2.$$

This gives Schwarz's inequality.

If x and y are linearly dependent, then $y = \alpha x$ for some $\alpha \in \mathbb{C}$. Hence,

$$\big|\langle x, y\rangle\big| = \big|\langle x, \alpha x\rangle\big| = |\bar{\alpha}|\big|\langle x, x\rangle\big| = |\alpha|\,\|x\|\,\|x\| = \|x\|\,\|\alpha x\| = \|x\|\,\|y\|.$$

Now, let x and y be vectors such that $\big|\langle x, y\rangle\big| = \|x\|\,\|y\|$. Or, equivalently,

$$\langle x, y\rangle\langle y, x\rangle = \langle x, x\rangle\langle y, y\rangle. \tag{2.6.3}$$

We next show that $\langle y, y\rangle x - \langle x, y\rangle y = 0$, which shows that x and y are linearly dependent. Indeed, by (2.6.3), we have

$$\begin{aligned}
&\Big\langle \langle y, y\rangle x - \langle x, y\rangle y, \langle y, y\rangle x - \langle x, y\rangle y \Big\rangle \\
&= \langle y, y\rangle^2\langle x, x\rangle - \langle y, y\rangle\langle y, x\rangle\langle x, y\rangle - \langle x, y\rangle\langle y, y\rangle\langle y, x\rangle \\
&+ \langle x, y\rangle\langle y, x\rangle\langle y, y\rangle = 0.
\end{aligned}$$

Thus, the proof is complete.

Corollary 2.6.1 (Triangle Inequality). *For any two elements x and y of an inner product space X, we have*

$$\|x + y\| \le \|x\| + \|y\|. \tag{2.6.4}$$

Proof. When $\alpha = 1$, equality (2.6.2) can be written as

$$\begin{aligned}
\|x + y\|^2 = \langle x + y, x + y\rangle &= \langle x, x\rangle + 2\,\mathrm{Re}\,\langle x, y\rangle + \langle y, y\rangle \\
&\le \langle x, x\rangle + 2\big|\langle x, y\rangle\big| + \langle y, y\rangle \\
&\le \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2 \quad \text{(by Schwarz's inequality)} \\
&\le \big(\|x\| + \|y\|\big)^2,
\end{aligned} \tag{2.6.5}$$

where $\mathrm{Re}\, z$ denotes the real part of $z \in \mathbb{C}$. This proves the triangle inequality.

Definition 2.6.1 (Norm in an Inner Product Space). By the *norm* in an inner product space X, we mean the functional defined by

$$\|x\| = \sqrt{\langle x, x\rangle}. \tag{2.6.6}$$

We have proved that every inner product space is a normed space. It is only natural to ask whether every normed space is an inner product space. More precisely, is it possible to define in a normed space $(X, \|.\|)$ with an inner product $\langle ., . \rangle$ such that $\|x\| = \sqrt{\langle x, x \rangle}$ for every $x \in X$? In general, the answer is negative. In the following theorem, we prove a property of the norm in an inner product space that is a necessary and sufficient condition for a normed space to be an inner product space.

The next theorem is usually called the parallelogram law because of its remarkable geometric interpretation, which reveals that the sum of the squares of the diagonals of a parallelogram is the sum of the squares of the sides. This characterizes the norm in a Hilbert space.

Theorem 2.6.2 (Parallelogram Law). *For any two elements x and y of an inner product space X, we have*

$$\|x + y\|^2 + \|x - y\|^2 = 2\left(\|x\|^2 + \|y\|^2\right). \tag{2.6.7}$$

Proof. We have

$$\|x + y\|^2 = \langle x + y, x + y \rangle = \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle$$

and hence,

$$\|x + y\|^2 = \|x\|^2 + \langle x, y \rangle + \langle y, x \rangle + \|y\|^2. \tag{2.6.8}$$

Now, replace y by $-y$ to obtain

$$\|x - y\|^2 = \|x\|^2 - \langle x, y \rangle - \langle y, x \rangle + \|y\|^2. \tag{2.6.9}$$

By adding (2.6.8) and (2.6.9), we obtain the parallelogram law (2.6.7).

One of the most important consequences of having the inner product is the possibility of defining orthogonality of vectors. This makes the theory of Hilbert spaces so much different from the general theory of Banach spaces.

Definition 2.6.2 (Orthogonal Vectors). Two vectors x and y in an inner product space are called *orthogonal* (denoted by $x \perp y$) if $\langle x, y \rangle = 0$.

Theorem 2.6.3 (Pythagorean Formula). *For any pair of orthogonal vectors x and y, we have*

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2. \tag{2.6.10}$$

Proof. If $x \perp y$, then $\langle x, y \rangle = 0$ and thus the equality (2.6.10) follows immediately from (2.6.8).

In the definition of the inner product space, we assume that X is a complex vector space. However, it is possible to define a real inner product space. Then condition (b) in Definition 2.5.1 becomes $\langle x, y\rangle = \langle y, x\rangle$. All of the above theorems hold in the real inner product space. If in Examples 2.5.1–2.5.6, the word *complex* is replaced by *real* and $\mathbb{C}$ by $\mathbb{R}$, we obtain a number of examples of real inner product spaces. A finite-dimensional real inner product space is called a *Euclidean space*.

If $x = (x_1, x_2, \ldots, x_N)$ and $y = (y_1, y_2, \ldots, y_N)$ are vectors in $\mathbb{R}^N$, then the inner product $\langle x, y\rangle = \sum_{k=1}^{N} x_k\, y_k$ can be defined equivalently by

$$\langle x, y\rangle = \|x\|\,\|y\|\cos\theta,$$

where θ is the angle between vectors x and y. In this case, Schwarz's inequality becomes

$$\left|\cos\theta\right| = \frac{\left|\langle x, y\rangle\right|}{\|x\|\,\|y\|} \le 1.$$

2.7 Definition and Examples of Hilbert Spaces

Definition 2.7.1 (Hilbert Space). A complete inner product space is called a *Hilbert space*.

By the completeness of an inner product space X, we mean the completeness of X as a normed space. Now, we discuss completeness of the inner product spaces and also give some new examples of inner product spaces and Hilbert spaces.

Example 2.7.1. Since the space $\mathbb{C}$ is complete, it is a Hilbert space.

Example 2.7.2. Clearly, both $\mathbb{R}^N$ and $\mathbb{C}^N$ are Hilbert spaces.

In $\mathbb{R}^N$, the inner product is defined by $\langle x, y\rangle = \sum_{k=1}^{N} x_k\, y_k$.

In $\mathbb{C}^N$, the inner product is defined by $\langle x, y\rangle = \sum_{k=1}^{N} x_k\, \bar{y}_k$.

In both cases, the norm is defined by

$$\|x\| = \sqrt{\langle x, x\rangle} = \left(\sum_{k=1}^{N} \left|x_k\right|^2\right)^{\frac{1}{2}}$$

Since these spaces are complete, they are Hilbert spaces.

Example 2.7.3. The sequence space l^2 defined in Example 2.5.3 is a Hilbert space.

Example 2.7.4. The space X described in Example 2.5.4 is an inner product space which is not a Hilbert space because it is not complete. The sequence

$$x_n = \left(1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, 0, 0, \dots\right)$$

is a Cauchy sequence because

$$\lim_{n,m\to\infty} \|x_n - x_m\| = \lim_{n,m\to\infty} \left[\sum_{k=m+1}^{n} \frac{1}{k^2}\right]^{\frac{1}{2}} = 0 \quad \text{for } m < n.$$

However, the sequence does not converge in X because its limit $\left(1, \frac{1}{2}, \frac{1}{3}, \dots\right)$ is not in X. However, this sequence $\{x_n\}$ converges in l^2.

Example 2.7.5. The space defined in Example 2.5.5 is another example of an incomplete inner product space. In fact, we consider the following sequence of functions in $\mathscr{C}([0,1])$ (see Fig. 2.3):

$$f_n(x) = \begin{cases} 1 & \text{if } \; 0 \le x \le \dfrac{1}{2}, \\ 1 - 2n\left(x - \dfrac{1}{2}\right) & \text{if } \; \dfrac{1}{2} \le x \le \left(\dfrac{1}{2n} + \dfrac{1}{2}\right), \\ 0 & \text{if } \; \left(\dfrac{1}{2n} + \dfrac{1}{2}\right) \le x \le 1. \end{cases}$$

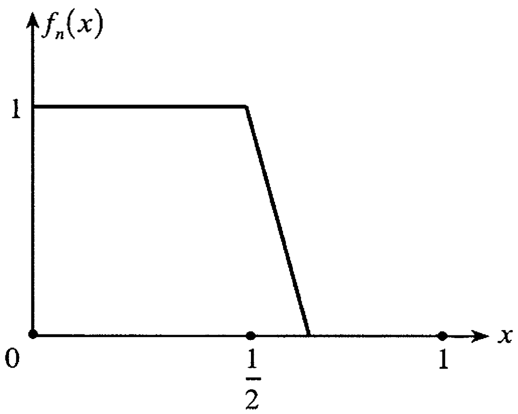

Fig. 2.3 Sequence of functions $\mathbf{f_n(x)}$

Evidently, the $\{f_n\}$ are continuous. Moreover,

$$\|f_n - f_m\| \le \left(\frac{1}{n} + \frac{1}{m}\right)^{1/2} \to 0 \qquad \text{as } m, n \to \infty.$$

Thus, $\{f_n\}$ is a Cauchy sequence. It is easy to check that this sequence converges to the limit function

$$f(x) = \begin{cases} 1 & \text{if} \quad 0 \le x \le \frac{1}{2}, \\ 0 & \text{if} \quad \frac{1}{2} < x \le 1. \end{cases}$$

The limit function is not continuous and hence is not an element of $\mathscr{C}([0,1])$. Consequently, $\mathscr{C}([0,1])$ is not a Hilbert space.

Example 2.7.6. The function space $L^2([a,b])$ is a Hilbert space. Since $L^2([a,b])$ is a normed space, it suffices to prove it is complete. Let $\{f_n\}$ be a Cauchy sequence in $L^2([a,b])$, that is,

$$\int_a^b |f_m - f_n|^2 dx \to 0 \qquad \text{as } m, n \to \infty.$$

Schwarz's inequality implies that as $m, n \to \infty$

$$\int_a^b |f_m - f_n| dx \le \sqrt{\int_a^b dx} \sqrt{\int_a^b |f_m - f_n|^2 dx} = \sqrt{b-a} \sqrt{\int_a^b |f_m - f_n|^2 dx} \to 0.$$

Thus, $\{f_n\}$ is a Cauchy sequence in $L^1([a,b])$ and hence converges to a function f in $L^1([a,b])$, that is,

$$\int_a^b |f - f_n| dx \to 0 \qquad \text{as } n \to \infty.$$

By Riesz's theorem, there exists a subsequence $\{f_{p_n}\}$ convergent to f almost everywhere. Clearly, given an $\varepsilon > 0$, we have

$$\int_a^b |f_{p_m} - f_{p_n}|^2 dx < \varepsilon$$

for sufficiently large m and n. Hence, by letting $n \to \infty$, we obtain

$$\int_a^b |f_{p_m} - f|^2 dx \le \varepsilon$$

by Fatou's lemma (see Theorem 2.8.5 in Debnath and Mikusinski 1999, p. 60). This proves that $f \in L^2([a,b])$. Moreover

$$\int_a^b |f - f_n|^2 dx \le \int_a^b |f - f_{p_n}|^2 dx + \int_a^b |f_{p_n} - f_n|^2 dx < 2\varepsilon$$

for sufficiently large n. This shows that the sequence $\{f_n\}$ converges to f in $L^2([a,b])$. Thus, the completeness is proved.

Example 2.7.7. Consider the space $\mathscr{C}_0(\mathbb{R})$ of all complex-valued continuous functions that vanish outside some finite interval. This is an inner product space with the inner product

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f(x)\,\overline{g(x)}\,dx.$$

Note that there is no problem with the existence of the integral because the product $f(x)\,\overline{g(x)}$ vanishes outside a bounded interval.

We now show that $\mathscr{C}_0(\mathbb{R})$ is not complete. We define

$$f_n(x) = \begin{cases} (\sin \pi x)/(1 + |x|) & \text{if } |x| \le n, \\ 0 & \text{if } |x| > n. \end{cases}$$

Clearly, $f_n \in \mathscr{C}_0(\mathbb{R})$ for every $n \in \mathbb{N}$. For $n > m$, we have

$$\|f_n - f_m\|^2 = \int_{-\infty}^{\infty} |f_n(x) - f_m(x)|^2 dx \le 2\int_m^n \frac{dx}{\left(1 + |x|^2\right)} \to 0 \qquad \text{as } m \to \infty.$$

This shows that $\{f_n\}$ is a Cauchy sequence. On the other hand, it follows directly from the definition of f_n that

$$\lim_{n\to\infty} f_n(x) = \frac{\sin \pi x}{(1 + |x|)},$$

which does not belong to $\mathscr{C}_0(\mathbb{R})$.

Example 2.7.8. We denote by $L^{2,\rho}([a,b])$ the space of all complex-valued square integrable functions on $[a,b]$ with a weight function ρ which is positive almost everywhere, that is, $f \in L^{2,\rho}([a,b])$ if

$$\int_a^b |f(x)|^2 \rho(x)\,dx < \infty.$$

This is a Hilbert space with the inner product

$$\langle f, g\rangle = \int_a^b f(x)\,\overline{g(x)}\,\rho(x)\,dx. \tag{2.7.1}$$

Example 2.7.9 (Sobolev Space). Let Ω be an open set in $\mathbb{R}^N$. Denote by $\tilde{H}^m(\Omega), m = 1, 2, \ldots,$ the space of all complex-valued functions $f \in \mathscr{C}^m(\Omega)$ such that $D^\alpha f \in L^2(\Omega)$ for all $|\alpha| \le m$, where

$$D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \ldots \partial x_N^{\alpha_N}}, \quad |\alpha| = \alpha_1 + \cdots + \alpha_N, \quad \text{and} \quad \alpha_1, \ldots, \alpha_N \ge 0.$$

For example, if $N = 2, \alpha = (2, 1)$, we have

$$D^\alpha f = \frac{\partial^3 f}{\partial x_1^2 \partial x_2}.$$

For $f \in \mathscr{C}^m(\Omega)$, we thus have

$$\int_\Omega \left| \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \ldots \partial x_N^{\alpha_N}} \right| < \infty$$

for every multi-index $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)$ such that $|\alpha| \le m$. The inner product in $\tilde{H}^m(\Omega)$ is defined by

$$\langle f, g \rangle = \int_\Omega \sum_{|\alpha| \le m} D^\alpha f \, \overline{D^\alpha g}. \tag{2.7.2}$$

In particular, if $\Omega \subset \mathbb{R}^2$, then the inner product in $\tilde{H}^2(\Omega)$ is given by

$$\langle f, g \rangle = \int_\Omega \big(f \, \bar{g} + f_x \, \bar{g}_x + f_y \, \bar{g}_y + f_{xx} \, \bar{g}_{xx} + f_{yy} \, \bar{g}_{yy} + f_{xy} \, \bar{g}_{xy} \big). \tag{2.7.3}$$

Or, if $\Omega = (a, b) \subset \mathbb{R}$, the inner product in $\tilde{H}^m(a, b)$ is

$$\langle f, g \rangle = \int_a^b \sum_{n=1}^m \frac{d^n f}{dx^n} \cdot \overline{\frac{d^n g}{dx^n}}. \tag{2.7.4}$$

The function space $\tilde{H}^m(\Omega)$ is an inner product space, but it is not a Hilbert space because it is not complete. The completion of $\tilde{H}^m(\Omega)$, denoted by $H^m(\Omega)$, is a Hilbert space. The function space $H^m(\Omega)$ can be defined directly if D^α in the above is understood as the distributional derivative. This approach is often used in more advanced textbooks and treatises.

The space $H^m(\Omega)$ is a particular case of a general class of spaces denoted by $W_p^m(\Omega)$ and introduced by S.L. Sobolev. We have $H^m(\Omega) = W_2^m(\Omega)$. Because of the applications to partial differential equations, space $H^m(\Omega)$ is one of the most important examples of Hilbert spaces.

2.8 Strong and Weak Convergences

Since every inner product space is a normed space, it is equipped with a convergence, and the convergence is defined by the norm, This convergence is called the *strong convergence*. Moreover, the norm induces a topology in the space. Thus, a normed space is, in a natural way, a metric space and hence a topological space.

Definition 2.8.1 (Strong Convergence). A sequence $\{x_n\}$ of vectors in an inner product space X is called *strongly convergent* to a vector x in X if

$$\|x_n - x\| \to 0 \qquad \text{as } n \to \infty.$$

The word "strong" is added in order to distinguish "strong convergence" from "weak convergence."

Definition 2.8.2 (Weak Convergence). A sequence $\{x_n\}$ of vectors in an inner product space X is called *weakly convergent* to a vector x in X if

$$\langle x_n, y \rangle \to \langle x, y \rangle \qquad \text{as } n \to \infty, \quad \text{for every } y \in X.$$

The condition in the above definition can also be stated as $\langle x_n - x, y \rangle \to 0$ as $n \to \infty$, for every $y \in X$.

It is convenient to reserve the notation "$x_n \to x$" for the strong convergence and use "$x_n \xrightarrow{w} x$" to denote weak convergence.

Theorem 2.8.1. *A strongly convergent sequence is weakly convergent (to the same limit), that is, $x_n \to x$ implies $x_n \xrightarrow{w} x$.*

Proof. Suppose that the sequence $\{x_n\}$ converges strongly to x. This means

$$\|x_n - x\| \to 0 \qquad \text{as } n \to \infty.$$

By Schwarz's inequality, we have

$$|\langle x_n - x, y \rangle| \le \|x_n - x\| \|y\| \to 0 \qquad \text{as } n \to \infty,$$

and thus,

$$\langle x_n - x, y \rangle \to 0 \qquad \text{as } n \to \infty, \quad \text{for every } y \in X.$$

This proves the theorem.

For any fixed y in an inner product space X, the mapping $\langle . , y \rangle : X \to \mathbb{C}$ is a linear functional on X. Theorem 2.8.1 states that such a functional is continuous for every $y \in X$. Obviously, the mapping $\langle x, . \rangle : X \to \mathbb{C}$ is also continuous.

In general, the converse of Theorem 2.8.1 is not true. A suitable example will be given in Sect. 2.9. On the other hand, we have the following theorem.

Theorem 2.8.2. *If* $x_n \xrightarrow{w} x$ *and* $\|x_n\| \to \|x\|$, *then* $x_n \to x$.

Proof. By the definition of weak convergence, we have

$$\langle x_n, y\rangle \to \langle x, y\rangle \qquad \text{as } n \to \infty, \quad \text{for all } y.$$

Hence,

$$\langle x_n, x\rangle \to \langle x, x\rangle = \|x\|^2.$$

Now,

$$\begin{aligned}\|x_n - x\|^2 &= \langle x_n - x, x_n - x\rangle \\ &= \langle x_n, x_n\rangle - \langle x_n, x\rangle - \langle x, x_n\rangle + \langle x, x\rangle \\ &= \|x_n\|^2 - 2\operatorname{Re}\langle x_n, x\rangle + \|x\|^2 \to \|x\|^2 - 2\|x\|^2 + \|x\|^2 = 0 \qquad \text{as } n \to \infty.\end{aligned}$$

The sequence $\{x_n\}$ is thus strongly convergent to x.

Theorem 2.8.3. *Suppose that the sequence* $\{x_n\}$ *converges weakly to* x *in a Hilbert space* H. *If, in addition,*

$$\|x\| = \lim_{n\to\infty} \|x_n\|, \tag{2.8.1}$$

then $\{x_n\}$ *converges strongly to* x *in* H.

Proof. We assume that $\langle x_n, y\rangle \to \langle x, y\rangle$ and hence, $\langle y, x_n\rangle \to \langle y, x\rangle$. We have the result

$$\|x - x_n\|^2 = \langle x - x_n, x - x_n\rangle = \|x\|^2 + \|x_n\|^2 - \langle x, x_n\rangle - \langle x_n, x\rangle. \tag{2.8.2}$$

In view of the assumption (2.8.1), result (2.8.2) gives

$$\lim_{n\to\infty} \|x - x_n\|^2 = \|x\|^2 + \|x\|^2 - \|x\|^2 - \|x\|^2 = 0.$$

This proves the theorem.

We next state an important theorem (without proof) that describes an important property of weakly convergent sequences.

Theorem 2.8.4. *Weakly convergent sequences are bounded, that is, if* $\{x_n\}$ *is a weakly convergent sequence, then there exists a number* M *such that* $\|x_n\| \le M$ *for all* $n \in \mathbb{N}$.

2.9 Orthogonal and Orthonormal Systems

By a basis of a vector space X, we mean a linearly independent family $\mathscr{B}$ of vectors from X such that any vector $x \in X$ can be written as $x = \sum_{n=1}^{m} \lambda_n x_n$, where $x_n \in \mathscr{B}$ and λ_n's are scalars. In inner product spaces, orthonormal bases are of much greater importance. Instead of finite combinations $\sum_{n=1}^{m} \lambda_n x_n$, infinite sums are allowed, and the condition of linear independence is replaced by orthogonality. One of the immediate advantages of these changes is that in all important examples it is possible to describe orthonormal bases. For example, $L^2([a,b])$ has countable orthonormal bases consisting of simple functions (see Example 2.9.2), whereas every basis of $L^2([a,b])$ is uncountable and we can only prove that such a basis exists without being able to describe its elements. In this section and the next, we give all necessary definitions and discuss basic properties of orthonormal bases.

Definition 2.9.1 (Orthogonal and Orthonormal Systems). Let X be an inner product space. A family S of nonzero vectors in X is called an *orthogonal system* if $x \perp y$ for any two distinct elements of S. If, in addition, $\|x\| = 1$ for all $x \in S$, S is called an *orthonormal system.*

Every orthogonal set of nonzero vectors can be normalized. If S is an orthogonal system, then the family

$$S_1 = \left\{ \frac{x}{\|x\|} : x \in S \right\}$$

is an orthonormal system. Both systems are equivalent in the sense that they span the same subspace of X.

Note that if x is orthogonal to each of $y_1, \ldots, y_n$, then x is orthogonal to every linear combination of vectors $y_1, \ldots, y_n$. In fact, we have

$$\langle x, y \rangle = \left\langle x, \sum_{k=1}^{n} \lambda_k y_k \right\rangle = \sum_{k=1}^{n} \bar{\lambda}_k \langle x, y_k \rangle = 0.$$

Theorem 2.9.1. *Orthogonal systems are linearly independent.*

Proof. Let S be an orthogonal system. Suppose that $\sum_{k=1}^{n} \alpha_k x_k = 0$, for some $x_1, \ldots, x_n \in S$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$. Then,

$$0 = \left\langle \sum_{k=1}^{n} \alpha_k x_k, \sum_{k=1}^{n} \alpha_k x_k \right\rangle = \sum_{k=1}^{n} |\alpha_k|^2 \|x_k\|^2.$$

This means that $\alpha_k = 0$ for each $k \in \mathbb{N}$. Thus, $x_1, \ldots, x_n$ are linearly independent.

Definition 2.9.2 (Orthonormal Sequence). A finite or infinite sequence of vectors which forms an orthonormal system is called an *orthonormal sequence*.

The condition of orthogonality of a sequence $\{x_n\}$ can be expressed in terms of the Kronecker delta symbol:

$$\langle x_m, x_n \rangle = \delta_{mn} = \begin{cases} 0 & \text{if } m \neq n, \\ 1 & \text{if } m = n. \end{cases} \tag{2.9.1}$$

Example 2.9.1. For $e_n = (0, \ldots, 0, 1, 0, \ldots)$ with 1 in the nth position, the set $S = \{e_1, e_2, \ldots\}$ is an orthonormal system in the sequence space l^2.

Example 2.9.2 (Trigonometric Functions). The sequence $\phi_n(x) = e^{inx}/\sqrt{2\pi}$, $n = 0, \pm 1, \pm 2, \ldots$ is an orthonormal system in $L^2([-\pi, \pi])$. Indeed, for $m \neq n$, we have

$$\langle \phi_m, \phi_n \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(m-n)x} dx = \frac{e^{\pi i(m-n)} - e^{-\pi i(m-n)}}{2\pi i(m-n)} = 0.$$

On the other hand,

$$\langle \phi_n, \phi_n \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(n-n)x} dx = 1.$$

Thus, $\langle \phi_m, \phi_n \rangle = \delta_{mn}$ for every pair of integers m and n.

For the real Hilbert space $L^2([-\pi, \pi])$, we can use the real and imaginary parts of the sequence $\{\phi_n\}$ and find that functions

$$\frac{1}{\sqrt{2\pi}} \cos nx, \qquad \frac{1}{\sqrt{2\pi}} \sin nx, \qquad (n = 0, 1, 2, \ldots)$$

form an orthonormal sequence.

Example 2.9.3. The *Legendre polynomials* defined by

$$P_0(x) = 1, \tag{2.9.2a}$$

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} \left(x^2 - 1\right)^n, \qquad n = 1, 2, 3, \ldots, \tag{2.9.2b}$$

form an orthogonal system in the space $L^2([-1, 1])$. It is convenient to write $\left(x^2 - 1\right)^n = p_n(x)$ so that

$$\int_{-1}^{1} P_n(x)\, x^m\, dx = \frac{1}{2^n n!} \int_{-1}^{1} p_n^{(n)}(x)\, x^m\, dx. \tag{2.9.3}$$

We evaluate this integral for $m < n$ by recursion. First, we note that

$$p_n^{(k)}(x) = 0$$

for $x = \pm 1$ and $k = 0, 1, 2, \ldots, (n-1)$. Hence, by integrating (2.9.3) by parts, we obtain

$$\int_{-1}^{1} p_n^{(n)}(x)\, x^m\, dx = -m \int_{-1}^{1} p_n^{(n-1)}(x)\, x^{m-1}\, dx.$$

Repeated application of this operation ultimately leads to

$$m!(-1)^m \int_{-1}^{1} p_n^{(n-m)}(x)\, dx = m!(-1)^m \Big[p_n^{(n-m-1)}(x)\Big]_{-1}^{1} = 0 \quad (m < n).$$

Consequently,

$$\int_{-1}^{1} P_n(x)\, x^m\, dx = 0 \qquad \text{for } m < n. \tag{2.9.4}$$

Since P_m is a polynomial of degree m, it follows that

$$\langle P_n, P_m \rangle = \int_{-1}^{1} P_n(x) P_m(x)\, dx = 0 \qquad \text{for } n \neq m. \tag{2.9.5}$$

This proves the orthogonality of the Legendre polynomials. To obtain an orthonormal system from the Legendre polynomials, we have to evaluate the norm of P_n in $L^2\big([-1, 1]\big)$:

$$\|P_n\| = \sqrt{\int_{-1}^{1} \big(P_n(x)\big)^2 dx}.$$

By repeated integration by parts, we first obtain

$$\begin{aligned}
\int_{-1}^{1} \big(1 - x^2\big)^n\, dx &= \int_{-1}^{1} (1-x)^n\, (1+x)^n\, dx \\
&= \frac{n}{n+1} \int_{-1}^{1} (1-x)^{n-1}\, (1+x)^{n+1}\, dx = \ldots \\
&= \frac{n(n-1)\ldots 2.1}{(n+1)(n+2)\ldots 2n} \int_{-1}^{1} (1+x)^{2n}\, dx \\
&= \frac{(n!)^2\, 2^{2n+1}}{(2n)!(2n+1)}.
\end{aligned} \tag{2.9.6}$$

A similar procedure gives

$$\int_{-1}^{1} \left\{p_n^{(n)}(x)\right\}^2 dx = 0 - \int_{-1}^{1} p_n^{(n-1)}(x)\, p_n^{(n+1)}(x)\, dx = \dots$$
$$= (-1)^n \int_{-1}^{1} p_n(x)\, p_n^{(2n)}(x)\, dx$$
$$= (2n)! \int_{-1}^{1} (1-x)^n\,(1+x)^n\, dx, \tag{2.9.7}$$

where we have used the fact that the $2n$th derivative of $p_n(x) = \left(x^2-1\right)^n$ is the same as the derivative of the term of exponent $2n$. The $2n$th derivatives of all the other terms of the sum are zero. From (2.9.2), (2.9.6), and (2.9.7), we obtain

$$\int_{-1}^{1} \left\{P_n(x)\right\}^2 dx = \frac{1}{(2^n n!)^2}\,(2n)!\,\frac{(n!)^2\, 2^{2n+1}}{(2n)!(2n+1)} = \frac{2}{2n+1}. \tag{2.9.8}$$

Thus, the polynomials $\sqrt{n+\frac{1}{2}}\; P_n(x)$ form an orthonormal system in the space $L^2\big([-1,1]\big)$.

Example 2.9.4. We denote by H_n the Hermite polynomials of degree n, that is,

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2}. \tag{2.9.9}$$

The functions $\phi_n(x) = e^{-x^2/2} H_n(x)$ form an orthogonal system in $L^2(\mathbb{R})$. The inner product

$$\langle \phi_n, \phi_m \rangle = (-1)^{n+m} \int_{-\infty}^{\infty} e^{x^2} \frac{d^n}{dx^n} e^{-x^2} \frac{d^m}{dx^m} e^{-x^2} dx$$

can be evaluated by integrating by parts, which gives

$$(-1)^{n+m} \langle \phi_n, \phi_m \rangle = \left[e^{x^2} \frac{d^n}{dx^n} e^{-x^2} \frac{d^{m-1}}{dx^{m-1}} e^{-x^2} \right]_{-\infty}^{\infty}$$
$$- \int_{-\infty}^{\infty} \frac{d}{dx} \left[e^{x^2} \frac{d^n}{dx^n} e^{-x^2} \right] \frac{d^{m-1}}{dx^{m-1}} e^{-x^2} dx, \tag{2.9.10}$$

and hence, all terms under the differential sign contain the factor e^{-x^2}. Since, for any $k \in \mathbb{N}$, we have

$$x^k e^{-x^2} \to 0 \qquad \text{as } x \to \infty,$$

the first term in (2.9.10) vanishes. Therefore, repeated integration by parts gives the result

$$\langle \phi_n, \phi_m \rangle = 0 \qquad \text{as } n \neq m. \tag{2.9.11}$$

To obtain an orthonormal system, we evaluate the norm:

$$\|\phi_n\|^2 = \int_{-\infty}^{\infty} e^{-x^2} \big(H_n(x)\big)^2 dx = \int_{-\infty}^{\infty} e^{-x^2} \left[e^{x^2} \frac{d^n}{dx^n} e^{-x^2} \right]^2 dx.$$

Integrating by parts n times yields

$$\|\phi_n\|^2 = (-1)^n \int_{-\infty}^{\infty} e^{-x^2} \left[e^{x^2} \frac{d^n}{dx^n} e^{-x^2} \right]^2 dx.$$

Since $H_n(x)$ is a polynomial of degree n, direct differentiation gives

$$e^{x^2} \frac{d^n}{dx^n} e^{-x^2} = (-2x)^n + \cdots$$

and

$$\frac{d^n}{dx^n} \left[e^{x^2} \frac{d^n}{dx^n} e^{-x^2} \right] = \frac{d^n}{dx^n} \left\{ (-2x)^n + \cdots \right\} = (-1)^n 2^n \, n!.$$

Consequently,

$$\|\phi_n\|^2 = 2^n \, n! \int_{-\infty}^{\infty} e^{-x^2} dx = 2^n \, n! \sqrt{\pi}. \tag{2.9.12}$$

Thus, the functions

$$\psi_n(x) = \frac{1}{\sqrt{2^n \, n! \sqrt{\pi}}} \, e^{-\frac{x^2}{2}} H_n(x)$$

form an orthonormal system in the Hilbert space $L^2(\mathbb{R})$.

In the preceding examples, the original sequence of functions is orthogonal but not orthonormal. Although the calculations involved might be complicated, it is always possible to normalize the functions and obtain an orthonormal sequence. It turns out that the same is possible if the original sequence of functions (or, in general, a sequence of vectors in an inner product space) is linearly independent, not necessarily orthogonal. The method of transforming such a sequence into an orthonormal sequence is called the *Gram–Schmidt orthonormalization process*. The process can be described as follows.

Given a sequence $\{y_n\}$ of linearly independent vectors in an inner product space, define sequences $\{w_n\}$ and $\{x_n\}$ inductively by

$$w_1 = y_1, \qquad x_1 = \frac{w_1}{\|w_1\|},$$

$$w_k = y_k - \sum_{n=1}^{k-1} \langle y_k, x_n \rangle x_n, \qquad x_k = \frac{w_k}{\|w_k\|}, \quad \text{for } k = 1, 2, \ldots$$

The sequence $\{w_n\}$ is orthogonal. Indeed,

$$\begin{aligned}\langle w_2, w_1 \rangle &= \Big\langle y_2 - \langle y_2, x_1 \rangle x_1, y_1 \Big\rangle = \langle y_2, y_1 \rangle - \langle y_2, x_1 \rangle \langle x_1, y_1 \rangle \\ &= \langle y_2, y_1 \rangle - \frac{\langle y_2, y_1 \rangle \langle y_1, y_1 \rangle}{\|y_1\|^2} = 0.\end{aligned}$$

Assume now that $w_1, \ldots, w_{k-1}$ are orthogonal. Then, for any $m < k$,

$$\begin{aligned}\langle w_k, w_m \rangle &= \langle y_k, w_m \rangle - \frac{\sum_{n=1}^{k-1} \langle y_k, w_n \rangle \langle w_n, w_m \rangle}{\|w_m\|^2} \\ &= \langle y_k, w_m \rangle - \frac{\langle y_k, w_m \rangle \langle w_m, w_m \rangle}{\|w_m\|^2} = 0.\end{aligned}$$

Therefore, vectors $w_1, \ldots, w_k$ are orthogonal. It follows, by induction, that the sequence $\{w_n\}$ is orthogonal and thus, $\{x_n\}$ is orthonormal. It is easy to check that any linear combination of vectors $x_1, \ldots, x_n$ is also a linear combination of $y_1, \ldots, y_n$ and vice versa. In other words, $\operatorname{span}\{x_1, \ldots, x_n\} = \operatorname{span}\{y_1, \ldots, y_n\}$ for every $n \in \mathbb{N}$.

2.10 Properties of Orthonormal Systems

In Sect. 2.6, we proved that the Pythagorean formula holds for any pair of orthogonal vectors in an inner product space X. It turns out that it can be generalized to any finite number of orthogonal vectors.

Theorem 2.10.1 (Pythagorean Formula). *If $x_1, \ldots, x_n$ are orthogonal vectors in an inner product space X, then*

$$\left\| \sum_{k=1}^{n} x_k \right\|^2 = \sum_{k=1}^{n} \|x_k\|^2. \tag{2.10.1}$$

Proof. If $x_1 \perp x_2$, then $\|x_1+x_2\|^2 = \|x_1\|^2+\|x_2\|^2$ by (2.6.10). Thus, the theorem is true for $n = 2$. Assume now that the (2.10.1) holds for $n-1$, that is,

$$\left\|\sum_{k=1}^{n-1} x_k\right\|^2 = \sum_{k=1}^{n-1} \|x_k\|^2.$$

Set $x = \sum_{k=1}^{n-1} x_k$ and $y = x_n$. Since $x \perp y$, we have

$$\left\|\sum_{k=1}^{n} x_k\right\|^2 = \|x+y\|^2 = \|x\|^2 + \|y\|^2 = \sum_{k=1}^{n-1} \|x_k\|^2 + \|x_n\|^2 = \sum_{k=1}^{n} \|x_k\|^2.$$

This proves the theorem.

Theorem 2.10.2 (Bessel's Equality and Inequality). *Let $x_1, \ldots, x_n$ be an orthonormal set of vectors in an inner product space X. Then, for every $x \in X$, we have*

$$\left\|x - \sum_{k=1}^{n} \langle x, x_k\rangle x_k\right\|^2 = \|x\|^2 - \sum_{k=1}^{n} |\langle x, x_k\rangle|^2 \tag{2.10.2}$$

and

$$\sum_{k=1}^{n} |\langle x, x_k\rangle|^2 \le \|x\|^2. \tag{2.10.3}$$

Proof. In view of the Pythagorean formula (2.10.1), we have

$$\left\|\sum_{k=1}^{n} \alpha_k x_k\right\|^2 = \sum_{k=1}^{n} \|\alpha_k x_k\|^2 = \sum_{k=1}^{n} |\alpha_k|^2$$

for any arbitrary complex numbers $\alpha_1, \ldots, \alpha_n$. Hence,

$$\begin{aligned}
\left\|x - \sum_{k=1}^{n} \alpha_k x_k\right\|^2 &= \left\langle x - \sum_{k=1}^{n} \alpha_k x_k, x - \sum_{k=1}^{n} \alpha_k x_k \right\rangle \\
&= \|x\|^2 - \left\langle x, \sum_{k=1}^{n} \alpha_k x_k \right\rangle - \left\langle \sum_{k=1}^{n} \alpha_k x_k, x \right\rangle + \sum_{k=1}^{n} |\alpha_k|^2 \|x_k\|^2 \\
&= \|x\|^2 - \sum_{k=1}^{n} \overline{\alpha_k} \langle x, x_k\rangle - \sum_{k=1}^{n} \alpha_k \overline{\langle x, x_k\rangle} + \sum_{k=1}^{n} \alpha_k \overline{\alpha_k} \\
&= \|x\|^2 - \sum_{k=1}^{n} |\langle x, x_k\rangle|^2 + \sum_{k=1}^{n} |\langle x, x_k\rangle - \alpha_k|^2.
\end{aligned} \tag{2.10.4}$$

In particular, if $\alpha_k = \langle x, x_k \rangle$, this result yields (2.10.2). From (2.10.2), it follows that

$$0 \le \|x\|^2 - \sum_{k=1}^{n} |\langle x, x_k \rangle|^2,$$

which gives (2.10.3). Thus, the proof is complete.

Remarks. 1. Note that expression (2.10.4) is minimized by taking $\alpha_k = \langle x, x_k \rangle$. This choice of α_k's minimizes $\left\| x - \sum_{k=1}^{n} \alpha_k x_k \right\|$ and thus provides the best approximation of x by a linear combination of vectors $x_1, \ldots, x_n$.

2. If $\{x_n\}$ is an orthonormal sequence of vectors in an inner product space X, then, from (2.10.2), by letting $n \to \infty$, we obtain

$$\sum_{k=1}^{\infty} |\langle x, x_k \rangle|^2 \le \|x\|^2. \tag{2.10.5}$$

This shows that the series $\sum_{k=1}^{\infty} |\langle x, x_k \rangle|^2$ converges for every $x \in X$. In other words, the sequence $\{\langle x, x_k \rangle\}$ is an element of l^2. We can say that an orthonormal sequence in X induces a mapping from X into l^2. The expansion

$$x \sim \sum_{n=1}^{\infty} \langle x, x_n \rangle x_n \tag{2.10.6}$$

is called a *generalized Fourier series* of x. The scalars $\alpha_n = \langle x, x_n \rangle$ are called the *generalized Fourier coefficients* of x with respect to the orthonormal sequence $\{x_n\}$. It may be observed that this set of coefficients gives the best approximation. In general, we do not know whether the series in (2.10.6) is convergent. However, as the next theorem shows, the completeness of the space ensures the convergence.

Theorem 2.10.3. *Let $\{x_n\}$ be an orthonormal sequence in a Hilbert space H and let $\{\alpha_n\}$ be a sequence of complex numbers. Then, the series $\sum_{n=1}^{\infty} \alpha_n x_n$ converges if and only if $\sum_{n=1}^{\infty} |\alpha_n|^2 < \infty$ and in that case*

$$\left\| \sum_{n=1}^{\infty} \alpha_n x_n \right\|^2 = \sum_{n=1}^{\infty} |\alpha_n|^2. \tag{2.10.7}$$

Proof. For every $m > k > 0$, we have

$$\left\| \sum_{n=k}^{m} \alpha_n x_n \right\|^2 = \sum_{n=k}^{m} |\alpha_n|^2 \qquad \text{by (2.10.1).} \tag{2.10.8}$$

If $\sum_{n=1}^{\infty} |\alpha_n|^2 < \infty$, then the sequence $s_m = \sum_{n=1}^{\infty} \alpha_n x_n$ is a Cauchy sequence by (2.10.8). This implies convergence of the series $\sum_{n=1}^{\infty} \alpha_n x_n$ because of the completeness of H.

Conversely, if the series $\sum_{n=1}^{\infty} \alpha_n x_n$ converges, then the same formula (2.10.8) implies the convergence of $\sum_{n=1}^{\infty} |\alpha_n|^2$ because the sequence of numbers $\sigma_m = \sum_{n=1}^{m} |\alpha_n|^2$ is a Cauchy sequence in $\mathbb{R}$.

To obtain (2.10.7), it is enough to take $k = 1$ and let $m \to \infty$ in (2.10.8).

The above theorem and (2.10.5) imply that in a Hilbert space H the series $\sum_{n=1}^{\infty} \langle x, x_n \rangle x_n$ converges for every $x \in H$. However, it may happen that it converges to an element different from x.

Example 2.10.1. Let $H = L^2([-\pi, \pi])$, and let $x_n(t) = \frac{1}{\sqrt{\pi}} \sin nt$ for $n = 1, 2, \ldots$. The sequence $\{x_n\}$ is an orthonormal set in H. On the other hand, for $x(t) = \cos t$, we have

$$\begin{aligned} \sum_{n=1}^{\infty} \langle x, x_n \rangle \, x_n(t) &= \sum_{n=1}^{\infty} \left[\frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} \cos t \, \sin nt \, dt \right] \frac{\sin nt}{\sqrt{\pi}} \\ &= \sum_{n=1}^{\infty} 0 \cdot \sin nt = 0 \neq \cos t. \end{aligned}$$

If $\{x_n\}$ is an orthonormal sequence in an inner product space X, then, for every $x \in X$, we have

$$\sum_{n=1}^{\infty} | \langle x, x_n \rangle |^2 < \infty,$$

and consequently,

$$\lim_{n \to \infty} \langle x, x_n \rangle = 0.$$

Therefore, orthonormal sequences are weakly convergent to zero. On the other hand, since $\|x_n\| = 1$ for all $n \in \mathbb{N}$, orthonormal sequences are not strongly convergent.

Definition 2.10.1 (Complete Orthonormal Sequence). An orthonormal sequence $\{x_n\}$ in an inner product space X is said to be *complete* if, for every $x \in X$, we have

$$x = \sum_{n=1}^{\infty} \langle x, x_n \rangle \, x_n \tag{2.10.9}$$

It is important to remember that since the right-hand side of (2.10.9) is an infinite series, the equality means

$$\lim_{n\to\infty} \left\| x - \sum_{k=1}^{n} \langle x, x_k \rangle \, x_k \right\| = 0,$$

where $\|.\|$ is the norm in X. For example, if $X = L^2([-\pi, \pi])$ and $\{f_n\}$ is an orthonormal sequence in X, then by

$$f = \sum_{n=1}^{\infty} \langle f, f_n \rangle \, f_n$$

we mean

$$\lim_{n\to\infty} \int_{-\pi}^{\pi} \left| f(t) - \sum_{k=1}^{n} \alpha_k \, f_k(t) \right|^2 dt = 0, \qquad \text{where } \alpha_k = \int_{-\pi}^{\pi} f(t) \, \overline{f_k(t)} \, dt.$$

This, in general, does not imply pointwise convergence: $f(x) = \sum_{n=1}^{\infty} \alpha_n \, f_n(x)$.

Definition 2.10.2 (Orthonormal Basis). An orthonormal system S in an inner product space X is called an *orthonormal basis* if every $x \in X$ has a unique representation

$$x = \sum_{n=1}^{\infty} \alpha_n \, x_n,$$

where $\alpha_n \in \mathbb{C}$ and x_n's are distinct elements of S.

Remarks. 1. Note that a complete orthonormal sequence $\{x_n\}$ in an inner product space X is an orthonormal basis in X. It suffices to show the uniqueness. Indeed, if

$$x = \sum_{n=1}^{\infty} \alpha_n \, x_n \quad \text{and} \quad x = \sum_{n=1}^{\infty} \beta_n \, x_n,$$

then

$$0 = \|x - x\|^2 = \left\|\sum_{n=1}^{\infty} \alpha_n x_n - \sum_{n=1}^{\infty} \beta_n x_n\right\|^2 = \left\|\sum_{n=1}^{\infty} (\alpha_n - \beta_n) x_n\right\|^2 = \sum_{n=1}^{\infty} |\alpha_n - \beta_n|^2$$

by Theorem 2.10.3. This means that $\alpha_n = \beta_n$ for all $n \in \mathbb{N}$. This proves the uniqueness.

2. If $\{x_n\}$ is a complete orthonormal sequence in an inner product space X, then the set

$$\operatorname{span}\{x_1, x_2, \ldots\} = \left\{\sum_{k=1}^{n} \alpha_k x_k : n \in \mathbb{N}, \alpha_1, \ldots, \alpha_k \in \mathbb{C}\right\}$$

is dense in X.

The following two theorems give important characterizations of complete orthonormal sequences in Hilbert spaces.

Theorem 2.10.4. *An orthonormal sequence $\{x_n\}$ in a Hilbert space H is complete if and only if $\langle x, x_n\rangle = 0$ for all $n \in \mathbb{N}$ implies $x = 0$.*

Proof. Suppose $\{x_n\}$ is a complete orthonormal sequence in H. Then, every $x \in H$ has the representation

$$x = \sum_{n=1}^{\infty} \langle x, x_n\rangle x_n.$$

Thus, if $\langle x, x_n\rangle = 0$ for every $n \in \mathbb{N}$, then $x = 0$.

Conversely, suppose $\langle x, x_n\rangle = 0$ for every $n \in \mathbb{N}$ implies $x = 0$. Let x be an element of H. We define

$$y = \sum_{n=1}^{\infty} \langle x, x_n\rangle x_n.$$

The sum y exists in H by (2.10.5) and Theorem 2.10.3. Since, for every $n \in \mathbb{N}$,

$$\begin{aligned}
\langle x - y, x_n\rangle &= \langle x, x_n\rangle - \left\langle \sum_{k=1}^{\infty} \langle x, x_k\rangle x_k, x_n \right\rangle \\
&= \langle x, x_n\rangle - \sum_{k=1}^{\infty} \langle x, x_k\rangle \langle x_k, x_n\rangle \\
&= \langle x, x_n\rangle - \langle x, x_n\rangle = 0,
\end{aligned}$$

we have $x - y = 0$ and hence,

$$x = \sum_{n=1}^{\infty} \langle x, x_n\rangle x_n.$$

Theorem 2.10.5 (Parseval's Formula). *An orthonormal sequence* $\{x_n\}$ *in a Hilbert space* H *is complete if and only if*

$$\|x\|^2 = \sum_{n=1}^{\infty} |\langle x, x_n\rangle|^2 \tag{2.10.10}$$

for every $x \in H$.

Proof. Let $x \in H$. By (2.10.2), for every $n \in \mathbb{N}$, we have

$$\left\| x - \sum_{k=1}^{n} \langle x, x_k\rangle x_k \right\|^2 = \|x\|^2 - \sum_{k=1}^{n} |\langle x, x_k\rangle|^2. \tag{2.10.11}$$

If $\{x_n\}$ is a complete sequence, then the expression on the left-hand side in (2.10.11) converges to zero as $n \to \infty$. Hence,

$$\lim_{n\to\infty} \left[\|x\|^2 - \sum_{k=1}^{n} |\langle x, x_k\rangle|^2 \right] = 0.$$

Therefore, (2.10.10) holds.

Conversely, if (2.10.10) holds, then the expression on the right-hand side of (2.10.11) converges to zero as $n \to \infty$ and thus,

$$\lim_{n\to\infty} \left\| x - \sum_{k=1}^{n} \langle x, x_k\rangle x_k \right\|^2 = 0.$$

This proves that $\{x_n\}$ is a complete sequence.

Example 2.10.2. The orthonormal system

$$\phi_n(x) = \frac{e^{inx}}{\sqrt{2\pi}}, \quad n = 0, \pm 1, \pm 2, \ldots,$$

given in Example 2.9.2, is complete in the space $L^2([-\pi, \pi])$. The proof of completeness is not simple. It will be discussed in Sect. 2.11.

A simple change of scale allows us to represent a function $f \in L^2([0, a])$ in the form

$$f(x) = \sum_{n=-\infty}^{\infty} \beta_n \, e^{2n\pi i x/a},$$

where

$$\beta_n = \frac{1}{a} \int_0^a f(t)\, e^{-2n\pi i t/a} dt.$$

Example 2.10.3. The sequence of functions

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \ldots$$

is a complete orthonormal system in $L^2([-\pi, \pi])$. The orthogonality follows from the following identities by simple integration:

$$\begin{aligned} 2\cos nx \cos mx &= \cos(n+m)x + \cos(n-m)x, \\ 2\sin nx \sin mx &= \cos(n-m)x - \cos(n+m)x, \\ 2\cos nx \sin mx &= \sin(n+m)x - \sin(n-m)x. \end{aligned}$$

Since

$$\int_{-\pi}^{\pi} \cos^2 nx\, dx = \int_{-\pi}^{\pi} \sin^2 mx\, dx = \pi,$$

the sequence is also orthonormal. The completeness follows from the completeness of the sequence in Example 2.10.2 in view of the following identities:

$$e^0 = 1 \qquad \text{and} \qquad e^{inx} = (\cos nx + i \sin nx).$$

Example 2.10.4. Each of the following two sequences of functions is a complete orthonormal system in the space $L^2([-\pi, \pi])$:

$$\frac{1}{\sqrt{\pi}}, \sqrt{\frac{2}{\pi}} \cos x, \sqrt{\frac{2}{\pi}} \cos 2x, \sqrt{\frac{2}{\pi}} \cos 3x, \ldots,$$

$$\sqrt{\frac{2}{\pi}} \sin x, \sqrt{\frac{2}{\pi}} \sin 2x, \sqrt{\frac{2}{\pi}} \sin 3x, \ldots.$$

Example 2.10.5 (Rademacher Functions and Walsh Functions). Rademacher functions $R(m, x)$ can be introduced in many different ways. We will use the definition based on the sine function,

$$R(m, x) = \operatorname{sgn}\big(\sin(2^m \pi x)\big), \quad m = 0, 1, 2, \ldots, x \in [0, 1],$$

where sgn denotes the signum function defined by

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{if} \quad x > 0, \\ 0 & \text{if} \quad x = 0, \\ -1 & \text{if} \quad x < 0. \end{cases}$$

Rademacher functions form an orthonormal system in $L^2([0,1])$. Obviously,

$$\int_0^1 \left|R(m,x)\right|^2 dx = 1 \qquad \text{for all } m.$$

To show that for $m \neq n$, we have

$$\int_0^1 R(m,x)\,\overline{R(n,x)}\,dx = 0.$$

First, notice that $\int_a^b R(m,x)\,dx = 0$ whenever $2^m(b-a)$ is an even number. Thus, for $m > n \geq 0$, we have

$$\begin{aligned}
\int_0^1 R(m,x)\,\overline{R(n,x)}\,dx &= \int_0^1 R(m,x)R(n,x)\,dx \\
&= \sum_{k=1}^{2^n} \int_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} R(m,x)R(n,x)\,dx \\
&= \sum_{k=1}^{2^n} \operatorname{sgn}\left(R\left(n, \frac{2k-1}{2}\right)\right) \int_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} R(m,x)\,dx = 0
\end{aligned}$$

because all of the integrals vanish.

The sequence of Rademacher functions is not complete. Indeed, consider the function

$$f(x) = \begin{cases} 0 & \text{if } \; 0 \leq x < \dfrac{1}{4}, \\ 1 & \text{if } \; \dfrac{1}{4} \leq x \leq \dfrac{3}{4}, \\ 0 & \text{if } \; \dfrac{3}{4} < x \leq 1. \end{cases}$$

Then

$$\int_0^1 R(0,x) f(x)\,dx = \frac{1}{2} \quad \text{and} \quad \int_0^1 R(m,x) f(x)\,dx = 0 \quad \text{for } m \geq 1,$$

but $f(x) \neq \frac{1}{2} R(0,x)$.

Rademacher functions can be used to construct Walsh functions, which form a complete orthonormal system. Walsh (1923) functions are denoted by $W(m,x), m = 0,1,2,\ldots$. For $m = 0$, we set $W(0,x) = 1$. For other values of m, we first represent m as a binary number, that is,

$$m = \sum_{k=1}^{n} 2^{k-1} a_k = a_1 + 2^1 a_2 + 2^2 a_3 + \cdots + 2^{n-1} a_n,$$

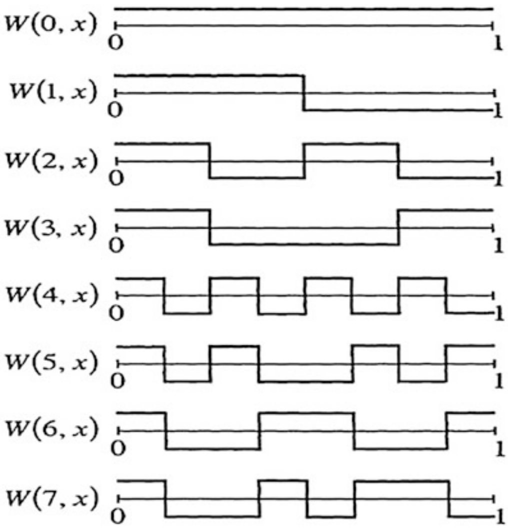

Fig. 2.4 Walsh functions $W(n,\ x)$

where $a_1, a_2, \ldots, a_n = 0$ or 1. Then, we define

$$W(m, x) = \prod_{k=1}^{n} \big(R(k, x)\big)^{a_k} = \big(R(1, x)\big)^{a_1} \big(R(2, x)\big)^{a_2} \ldots \big(R(n, x)\big)^{a_n},$$

where $\big(R(m, x)\big)^0 \equiv 1$. For instance, since 53 is written as 110101 in binary form, we have

$$W(53, x) = R(1, x)\, R(3, x)\, R(5, x)\, R(6, x).$$

Clearly, we have

$$R(n, x) = W\big(2^{n-1}, x\big), \quad n \in \mathbb{N}.$$

Several Walsh functions are shown in Fig. 2.4.

2.11 Trigonometric Fourier Series

In this section, we prove that the sequence

$$\phi_n(x) = \frac{e^{inx}}{\sqrt{2\pi}}, \quad n = 0, \pm 1, \pm 2, \ldots,$$

is a complete orthonormal sequence in $L^2([-\pi,\pi])$. The orthogonality has been established in Example 2.9.2. The proof of completeness is much more complicated. For the purpose of this proof, it is convenient to identify elements of the space $L^1([-\pi,\pi])$ with 2π-periodic locally integrable functions on $\mathbb{R}$ due to the fact that

$$\int_{-\pi}^{\pi} f(t)\,dt = \int_{-\pi-x}^{\pi-x} f(t)\,dt = \int_{-\pi}^{\pi} f(t-x)\,dt$$

for any $f \in L^1([-\pi,\pi])$ and any $x \in \mathbb{R}$.

Let $f \in L^1([-\pi,\pi])$ and

$$f_n = \sum_{k=-n}^{n} \langle f, \phi_k \rangle \phi_k, \quad n = 0, 1, 2, \ldots.$$

Then

$$f_n(x) = \sum_{k=-n}^{n} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)\, e^{-ikt} dt\, e^{ikx} = \sum_{k=-n}^{n} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)\, e^{ik(x-t)} dt.$$

We next show that, for every $f \in L^1([-\pi,\pi])$, we have

$$\lim_{n\to\infty} \frac{f_0 + f_1 + \cdots + f_n}{n+1} = f$$

in the $L^1([-\pi,\pi])$ norm. We first observe that

$$\begin{aligned} \frac{f_0(x) + f_1(x) + \cdots + f_n(x)}{n+1} &= \sum_{k=-n}^{n} \left(1 - \frac{|k|}{n+1}\right) \langle f, \phi_k \rangle \phi_k(x) \\ &= \sum_{k=-n}^{n} \frac{1}{2\pi} \left(1 - \frac{|k|}{n+1}\right) \int_{-\pi}^{\pi} f(t)\, e^{-ikt} dt\, e^{ikx} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \left(\sum_{k=-n}^{n} \left(1 - \frac{|k|}{n+1}\right) e^{ik(x-t)} \right) dt. \end{aligned} \tag{2.11.1}$$

Lemma 2.11.1. *For every $n \in \mathbb{N}$ and $x \in \mathbb{R}$, we have*

$$\sum_{k=-n}^{n} \left(1 - \frac{|k|}{n+1}\right) e^{ikx} = \frac{1}{n+1} \frac{\sin^2 \dfrac{(n+1)x}{2}}{\sin^2 \dfrac{x}{2}}.$$

Proof. We have

$$\sin^2 \frac{x}{2} = \frac{1}{2}(1 - \cos x) = -\frac{1}{4} e^{-ix} + \frac{1}{2} - \frac{1}{4} e^{ix}.$$

Then, direct calculation gives

$$\left(-\frac{1}{4} e^{-ix} + \frac{1}{2} - \frac{1}{4} e^{ix}\right) \sum_{k=-n}^{n} \left(1 - \frac{|k|}{n+1}\right) e^{ikx} = \frac{1}{n+1}\left(-\frac{1}{4} e^{-i(n+1)x} + \frac{1}{2} - \frac{1}{4} e^{i(n+1)x}\right).$$

This proves the lemma.

Lemma 2.11.2. *The sequence of functions*

$$K_n(t) = \sum_{k=-n}^{n} \left(1 - \frac{|k|}{n+1}\right) e^{ikt}$$

is a Fejér summability kernel.

Proof. Since $\int_{-\pi}^{\pi} e^{ikt} = 2\pi$ if $k = 0$ and $\int_{-\pi}^{\pi} e^{ikt} = 0$ for any other integer k, we obtain

$$\int_{-\pi}^{\pi} K_n(t)dt = \sum_{k=-n}^{n} \left(1 - \frac{|k|}{n+1}\right) \int_{-\pi}^{\pi} e^{ikt} dt = 2\pi.$$

From Lemma 2.11.1, it follows that $K_n \geq 0$ and hence

$$\int_{-\pi}^{\pi} |K_n(t)| dt = \int_{-\pi}^{\pi} K_n(t)dt = 2\pi.$$

Finally, let $\delta \in (0, \pi)$. For $t \in (\delta, 2\pi - \delta)$, we have $\sin \frac{t}{2} \geq \sin \frac{\delta}{2}$ and therefore

$$K_n(t) = \frac{1}{n+1} \frac{\sin^2 \frac{(n+1)x}{2}}{\sin^2 \frac{x}{2}} \leq \frac{1}{(n+1)\sin^2 \frac{\delta}{2}}.$$

Thus,

$$\int_{\delta}^{2\pi-\delta} K_n(t)dt \leq \frac{2\pi}{(n+1)\sin^2 \frac{\delta}{2}}.$$

For a fixed δ, the right-hand side tends to 0 as $n \to \infty$. This proves the lemma.

Theorem 2.11.1. *If* $f \in L^2([-\pi, \pi])$ *and* $|\langle f, \phi_n \rangle| = 0$ *for all integers n, then* $f = 0$ *a.e.*

Proof. If

$$\int_{-\pi}^{\pi} f(t)\, e^{-int} dt = 0$$

for all integers n, then

$$f_n(x) = \sum_{k=-n}^{n} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)\, e^{ik(x-t)}\, dt = 0.$$

Consequently,

$$\frac{f_0(x) + f_1(x) + \cdots + f_n(x)}{n+1} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \left(\sum_{k=-n}^{n} \left(1 - \frac{|k|}{n+1}\right) e^{ik(x-t)} \right) dt = 0.$$

On the other hand, since f and all the functions e^{ikx} are 2π-periodic, we have

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \left(\sum_{k=-n}^{n} \left(1 - \frac{|k|}{n+1}\right) e^{ik(x-t)} \right) dt$$
$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-t) \left(\sum_{k=-n}^{n} \left(1 - \frac{|k|}{n+1}\right) e^{ikt} \right) dt$$

and hence, by Theorem 3.8.1 (see Debnath and Mikusinski 1999) and Lemma 2.11.2,

$$\lim_{n\to\infty} \frac{f_0 + f_1 + \cdots + f_n}{n+1} = f$$

in the $L^1([-\pi, \pi])$ norm. Therefore, $f = 0$ a.e.

Theorem 2.11.2. *The sequence of functions*

$$\phi_n(x) = \frac{e^{inx}}{\sqrt{2\pi}}, \quad n = 0, \pm 1, \pm 2, \ldots,$$

is complete.

Proof. If $f \in L^2([-\pi, \pi])$, then $f \in L^1([-\pi, \pi])$. Thus, by Theorem 2.11.1 if $\langle f, \phi_n \rangle = 0$ for all integers n, then $f = 0$ a.e., that is, $f = 0$ in $f \in L^2([-\pi, \pi])$. This proves completeness of the sequence by Theorem 2.10.4.

Theorem 2.11.2 implies that, for every $f \in L^2\big([-\pi, \pi]\big)$, we have

$$f = \sum_{n=-\infty}^{\infty} \alpha_n \, \phi_n, \tag{2.11.2}$$

where

$$\phi_n(x) = \frac{e^{inx}}{\sqrt{2\pi}} \quad \text{and} \quad \alpha_n = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(t)\, e^{-int} dt.$$

In this case, Parseval's formula yields

$$\|f\|^2 = \int_{-\pi}^{\pi} |f(x)|^2 dx = \sum_{n=-\infty}^{\infty} |\alpha_n|^2.$$

The series (2.11.2) is called the *Fourier series* of f, and the numbers an are called the *Fourier coefficients* of f. It is important to point out that, in general, (2.11.2) does not imply pointwise convergence. The problem of pointwise convergence of Fourier series is much more difficult. In 1966, Carleson proved that Fourier series of functions in $L^2([-\pi, \pi])$ converge almost everywhere.

2.12 Orthogonal Complements and the Projection Theorem

By a subspace of a Hilbert space H, we mean a vector subspace of H. A subspace of a Hilbert space is an inner product space. If we additionally assume that S is a closed subspace of H, then S is a Hilbert space itself because a closed subspace of a complete normed space is complete.

Definition 2.12.1 (Orthogonal Complement). Let S be a nonempty subset of a Hilbert space H. An element $x \in H$ is said to be orthogonal to S, denoted by $x \perp S$, if $\langle x, y \rangle = 0$ for every $y \in S$. The set of all elements of H orthogonal to S, denoted by $S^\perp$, is called the *orthogonal complement* of S. In symbols,

$$S^\perp = \big\{x \in H : x \perp S\big\}.$$

The orthogonal complement of $S^\perp$ is denoted by $S^{\perp\perp} = \big(S^\perp\big)^\perp$.

Remarks. If $x \perp y$ for every $y \in H$, then $x = 0$. Thus $H^\perp = \{0\}$. Similarly, $\{0\}^\perp = H$. Two subsets A and B of a Hilbert space are said to be *orthogonal* if $x \perp y$ for every $x \in A$ and $y \in B$. This is denoted by $A \perp B$. Note that if $A \perp B$, then $A \cap B = \{0\}$ or $\emptyset$.

Theorem 2.12.1 (Orthogonal Complement). *For any subset of S of a Hilbert space H, the set $S^\perp$ is a closed subspace of H.*

Proof. If $\alpha, \beta \in \mathbb{C}$ and $x, y \in S^\perp$, then

$$\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle = 0$$

for every $z \in S$. Thus, $S^\perp$ is a vector subspace of H. We next prove that $S^\perp$ is closed.

Let $\{x_n\} \in S^\perp$ and $x_n \to x$ for some $x \in H$. From the continuity of the inner product, we find

$$\langle x, y \rangle = \left\langle \lim_{n\to\infty} x_n, y \right\rangle = \lim_{n\to\infty} \langle x_n, y \rangle = 0,$$

for every $y \in S$. This shows that $x \in S^\perp$, and thus, $S^\perp$ is closed.

The above theorem implies that $S^\perp$ is a Hilbert space for any subset S of H. Note that S does not have to be a vector space. Since $S \perp S^\perp$, we have $S \cap S^\perp = \{0\}$ or $S \cap S^\perp = \emptyset$.

Definition 2.12.2 (Convex Sets). A set S in a vector space is called *convex* if, for any $x, y \in S$ and $\alpha \in (0, 1)$, we have $\alpha x + (1 - \alpha) y \in S$.

Note that a vector subspace is a convex set.

The following theorem concerning the minimization of the norm is of fundamental importance in approximation theory.

Theorem 2.12.2 (The Closest Point Property). *Let S be a closed convex subset of a Hilbert space H. For every point $x \in H$, there exists a unique point $y \in S$ such that*

$$\|x - y\| = \inf_{z \in S} \|x - z\|. \tag{2.12.1}$$

Proof. Let $\{y_n\}$ be a sequence in S such that

$$\|x - y_n\| = \inf_{z \in S} \|x - z\|.$$

Denote $d = \inf_{z \in S} \|x - z\|$. Since $\frac{1}{2}(y_m + y_n) \in S$, we have

$$\left\| x - \frac{1}{2}(y_m + y_n) \right\| \geq d, \quad \text{for all } m, n \in \mathbb{N}.$$

Moreover, by the parallelogram law (2.6.7),

$$\begin{aligned}\|y_m - y_n\|^2 &= 4\left\|x - \frac{1}{2}(y_m + y_n)\right\|^2 + \|y_m - y_n\|^2 - 4\left\|x - \frac{1}{2}(y_m + y_n)\right\|^2 \\ &= \|(x - y_m) + (x - y_n)\|^2 + \|(x - y_m) - (x - y_n)\|^2 - 4\left\|x - \frac{1}{2}(y_m + y_n)\right\|^2 \\ &= 2\left(\|x - y_m\|^2 + \|x - y_n\|^2\right) - 4\left\|x - \frac{1}{2}(y_m + y_n)\right\|^2.\end{aligned}$$

Since

$$2\left(\|x - y_m\|^2 + \|x - y_n\|^2\right) \to 4d^2, \qquad \text{as } m, n \to \infty,$$

and

$$\left\|x - \frac{1}{2}(y_m + y_n)\right\|^2 \geq d^2,$$

it follows that $\|y_m - y_n\|^2 \to 0$ as $m, n \to \infty$. Thus, $\{y_n\}$ is a Cauchy sequence. Since H is complete and S is closed, $\lim_{n\to\infty} y_n = y$ exists and $y \in S$. It follows from the continuity of the norm that

$$\|x - y\| = \left\|x - \lim_{n\to\infty} y_n\right\| = \lim_{n\to\infty} \|x - y_n\| = d.$$

We have proved that there exists a point in S satisfying (2.12.1). It remains to prove the uniqueness. Suppose there is another point y_1 in S satisfying (2.12.1). Then, since $\frac{1}{2}(y + y_1) \in S$, we have

$$\|y - y_1\|^2 = 4d^2 - 4\left\|x - \frac{y + y_1}{2}\right\|^2 \leq 0.$$

This can only be true if $y = y_1$.

Remark. Theorem 2.12.2 gives an existence and uniqueness result which is crucial for optimization problems. However, it does not tell us how to find that optimal point. The characterization of the optimal point in the case of a real Hilbert space stated in the following theorem is often useful in such problems.

Theorem 2.12.3. *Let S be a closed convex subset of a real Hilbert space H, $y \in S$, and let $x \in H$. Then, the following conditions are equivalent:*

(a) $\qquad \|x - y\| = \inf_{z\in S} \|x - z\|,$

(b) $\qquad \langle x - y, z - y\rangle \leq 0 \qquad \text{for all } z \in S.$

Proof. Let $z \in S$. Since S is convex, $\lambda z + (1-\lambda)y \in S$ for every $\lambda \in (0,1)$. Then, by (a), we have

$$\|x-y\| \le \|x-\lambda z-(1-\lambda)y\| = \|(x-y)-\lambda(z-y)\|.$$

Since H is a real Hilbert space, we get

$$\|x-y\|^2 \le \|x-y\|^2 - 2\lambda\langle x-y, z-y\rangle + \lambda^2\|z-y\|^2.$$

Consequently,

$$|\langle x-y, z-y\rangle| \le \frac{\lambda}{2}\|z-y\|^2.$$

Thus, (b) follows by letting $\lambda \to 0$. .

Conversely, if $x \in H$ and $y \in S$ satisfy (b), then, for every $z \in S$, we have

$$\|x-y\|^2 - \|x-z\|^2 = 2\langle x-y, z-y\rangle - \|z-y\|^2 \le 0.$$

Thus, x and y satisfy (a).

If $H = \mathbb{R}^2$ and S is a closed convex subset of $\mathbb{R}^2$, then condition (b) has an important geometric meaning: the angle between the line through x and y and the line through z and y is always *obtuse*, as shown in Fig. 2.5.

Theorem 2.12.4 (Orthogonal Projection). *If S is a closed subspace of a Hilbert space H, then every element $x \in H$ has a unique decomposition in the form $x = y + z$, where $y \in S$ and $z \in S^\perp$.*

Proof. If $x \in S$, then the obvious decomposition is $x = x + 0$. Suppose now that $x \neq S$. Let y be the unique point of S satisfying $\|x-y\| = \inf_{w\in S}\|x-w\|$, as in Theorem 2.12.2. We show that $x = y + (x-y)$ is the desired decomposition.

If $w \in S$ and $\lambda \in \mathbb{C}$, then $y + \lambda w \in S$ and

$$\|x-y\|^2 \le \|x-y-\lambda w\|^2 = \|x-y\|^2 - 2\mathscr{R}\lambda|\langle w, x-y\rangle|^2 + |\lambda|^2\|w\|^2.$$

Hence,

$$-2\mathscr{R}\lambda\langle w, x-y\rangle + |\lambda|^2\|w\|^2 \ge 0.$$

If $\lambda > 0$, then dividing by λ and letting $\lambda \to 0$ gives

$$\mathscr{R}\langle w, x-y\rangle \le 0. \tag{2.12.2}$$

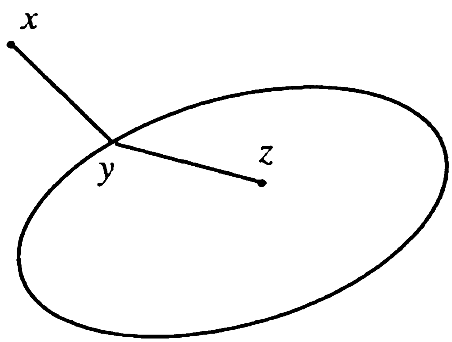

Fig. 2.5 Angle between two lines

Similarly, replacing λ by $-i\lambda(\lambda > 0)$, dividing by λ, and letting $\lambda \to 0$ yields

$$\mathscr{I}\langle w, x - y\rangle \leq 0. \tag{2.12.3}$$

Since $y \in S$ implies $-y \in S$, inequalities (2.12.2) and (2.12.3) hold also with $-w$ instead of w. Therefore, $\langle w, x - y\rangle = 0$ for every $w \in S$, which means $x - y \in S^{\perp}$

To prove the uniqueness, note that if $x = y_1 + z_1$, $y_1 \in S$, and $z_1 \in S^{\perp}$, then $y - y_1 \in S$ and $z - z_1 \in S^{\perp}$. Since $y - y_1 = z_1 - z$, we must have $y - y_1 = z_1 - z = 0$.

Remarks. 1. According to Theorem 2.12.4, every element of H can be uniquely represented as the sum of an element of S and an element of $S^{\perp}$. This can be stated symbolically as

$$H = S \oplus S^{\perp}. \tag{2.12.4}$$

We say that H is the direct sum of S and $S^{\perp}$. Equality (2.12.4) is called an *orthogonal decomposition* of H. Note that the union of a basis of S and a basis of $S^{\perp}$ is a basis of H.

2. Theorem 2.12.2 allows us to define a mapping $P_s(x) = y$, where y is as in (2.12.1). The mapping P_s is called the *orthogonal projection* onto S.

Example 2.12.1. Let $H = \mathbb{R}^2$. Figure 2.6 exhibits the geometric meaning of the orthogonal decomposition in $\mathbb{R}^2$. Here, $x \in \mathbb{R}^2$, $x = y + z$, $y \in S$, and $z \in S^{\perp}$. Note that if s_0 is a unit vector in S, then $y = \langle x, s_0\rangle s_0$.

Example 2.12.2. If $H = \mathbb{R}^3$, given a plane P, any vector x can be projected onto the plane P. Figure 2.7 illustrates this example.

Theorem 2.12.5. *If S is a closed subspace of a Hilbert space H, then $S^{\perp\perp} = S$.*

Proof. If $x \in S$, then for every $z \in S^{\perp}$ we have $\langle x, z\rangle = 0$, which means $x \in S^{\perp\perp}$. Thus, $S \subset S^{\perp\perp}$. To prove that $S^{\perp\perp} \subset S$ consider an $x \in S^{\perp\perp}$. Since S is closed, $x = y + z$ for some $y \in S$ and $z \in S^{\perp}$. In view of the inclusion $S \subset S^{\perp\perp}$, we

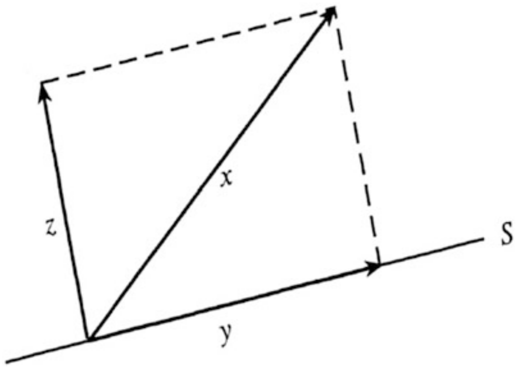

Fig. 2.6 Orthogonal decomposition in $\mathbb{R}^2$

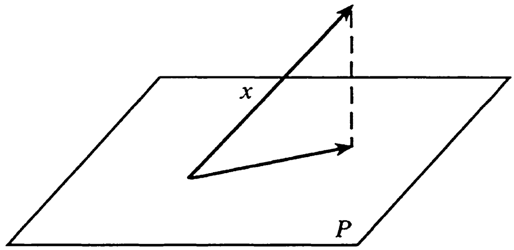

Fig. 2.7 Orthogonal projection into a plane

have $y \in S^{\perp\perp}$ and thus, $z = x - y \in S^{\perp\perp}$ because $S^{\perp\perp}$ is a vector subspace. But $z \in S^{\perp}$, so we must have $z = 0$, which means $x = y \in S$. This shows that $S^{\perp\perp} \subset S$. This completes the proof.

2.13 Linear Functionals and the Riesz Representation Theorem

In Sect. 2.7, we have remarked that for any fixed vector x_0 in an inner product space X, the formula $f(x) = \langle x, x_0 \rangle$ defines a bounded linear functional on X. It turns out that if X is a Hilbert space, then every bounded linear functional is of this form. Before proving this result, known as the Riesz representation theorem, we discuss some examples and prove a lemma.

Example 2.13.1. Let $H = L^2\big((a,b)\big), -\infty < a < b < \infty$. Define a linear functional f on H by the formula

$$f(x) = \int_a^b x(t)\,dt.$$

If x_0 denotes the constant function 1 on (a, b), then clearly $f(x) = \langle x, x_0 \rangle$ and thus, f is a bounded functional.

Example 2.13.2. Let $H = L^2(a, b)$ and let t_0 be a fixed point in (a, b). Let f be a functional on H defined by $f(x) = x(t_0)$. This functional is linear, but it is not bounded.

Example 2.13.3. Let $H = \mathbb{C}^n$ and let $n_0 \in \{1, 2, \ldots, n\}$. Define f by the

$$f\big((x_1, \ldots, x_n)\big) = x_{n_0}.$$

We have

$$f\big((x_1, \ldots, x_n)\big) = \big\langle (x_1, \ldots, x_n), e_{n_0} \big\rangle,$$

where e_{n_0} is the vector which has 1 on the n_0-th place and zeros in the remaining places. Thus, f is a bounded linear functional.

Lemma 2.13.1. *Let f be a bounded linear functional on an inner product space X. Then,* $\dim \mathcal{N}(f)^\perp \le 1$.

Proof. If $f = 0$, then $\mathcal{N}(f) = X$ and $\dim \mathcal{N}(f)^\perp = 0 \le 1$. It remains to show that $\dim \mathcal{N}(f)^\perp = 1$ when f is not zero. Continuity of f implies that $\mathcal{N}(f)$ is a closed subspace of X and thus $\mathcal{N}(f)^\perp$ is not empty. Let $x_1, x_2 \in \mathcal{N}(f)^\perp$ be nonzero vectors. Since $f(x_1) \neq 0$ and $f(x_2) \neq 0$, there exists a scalar $a \neq 0$ such that $f(x_1) + af(x_2) = 0$ or $f(x_1 + ax_2) = 0$. Thus, $x_1 + ax_2 \in \mathcal{N}(f)$. On the other hand, since $\mathcal{N}(f)^\perp$ is a vector space and $x_1, x_2 \in \mathcal{N}(f)^\perp$, we must have $x_1 + ax_2 \in \mathcal{N}(f)^\perp$. This is only possible if $x_1 + ax_2 = 0$ which shows that x_1 and x_2 are linearly dependent because $a \neq 0$.

Theorem 2.13.1 (The Riesz Representation Theorem). *Let f be a bounded linear functional on a Hilbert space H. There exists exactly one $x_0 \in H$ such that $f(x) = \langle x, x_0 \rangle$ for all $x \in H$. Moreover, we have $\|f\| = \|x_0\|$.*

Proof. If $f(x) = 0$ for all $x \in H$, then $x_0 = 0$ has the desired properties. Assume now that f is a nonzero functional. Then, $\dim \mathcal{N}(f)^\perp = 1$, by Lemma 2.13.1. Let z_0 be a unit vector in $\mathcal{N}(f)^\perp$. Then, for every $x \in H$, we have

$$x = x - \langle x, z_0 \rangle z_0 + \langle x, z_0 \rangle z_0.$$

Since $\langle x, z_0 \rangle z_0 \in \mathcal{N}(f)^\perp$, we must have $x - \langle x, z_0 \rangle z_0 \in \mathcal{N}(f)$, which means that

$$f\big(x - \langle x, z_0 \rangle z_0\big) = 0.$$

Consequently,

$$f(x) = f\big(\langle x, z_0 \rangle z_0\big) = \langle x, z_0 \rangle f(z_0) = \big\langle x, \overline{f(z_0)} z_0 \big\rangle.$$

Therefore, if we put

$$x_0 = \overline{f(z_0)} z_0,$$

then $f(x) = \langle x, x_0 \rangle$ for all $x \in H$.

Suppose now that there is another point x_1 such that $f(x) = \langle x, x_1\rangle$ for all $x \in H$. Then $\langle x, x_0 - x_1\rangle = 0$ for all $x \in H$ and thus $\langle x_0 - x_1, x_0 - x_1\rangle = 0$. This is only possible if $x_0 = x_1$.

Finally, we have

$$\|f\| = \sup_{\|x\|=1} |f(x)| = \sup_{\|x\|=1} |\langle x, x_0\rangle| \leq \sup_{\|x\|=1} \big(\|x\|\,\|x_0\|\big) = \|x_0\|$$

and

$$\|x_0\|^2 = \langle x_0, x_0\rangle = |f(x_0)| \leq \|f\|\,\|x_0\|.$$

Therefore,

$$\|f\| = \|x_0\|.$$

The collection H' of all bounded linear functionals on a Hilbert space H is a Banach space. The Riesz representation theorem states that $H' = H$ or, more precisely, that H' and H are isomorphic. The element x_0 corresponding to a functional f is sometimes called the *representer* of f.

Note that the functional f defined by $f(x) = \langle x, x_0\rangle$, where $x_0 \neq 0$ is a fixed element of a complex Hilbert space H, is not linear. Indeed, we have $f(\alpha x + \beta y) = \bar{\alpha} f(x) + \bar{\beta} f(y)$. Such functionals are often called *anti-linear* or *conjugate-linear.*

2.14 Separable Hilbert Spaces

Definition 2.14.1 (Separable Hilbert Space). A Hilbert space is called *separable* if it contains a complete orthonormal sequence. Finite-dimensional Hilbert spaces are considered separable.

Example 2.14.1. The Hilbert space $L^2\big([-\pi, \pi]\big)$ is separable. Example 2.10.2 shows a complete orthonormal sequence in $L^2\big([-\pi, \pi]\big)$.

Example 2.14.2. The sequence space l^2 is separable.

Example 2.14.3 (Nonseparable Hilbert Space). Let H be the space of all complex-valued functions defined on $\mathbb{R}$ which vanish everywhere except a countable number of points in $\mathbb{R}$ and such that

$$\sum_{f(x)\neq 0} |f(x)|^2 < \infty.$$

The inner product in H can be defined as

$$\langle f, g\rangle = \sum_{f(x)g(x)\neq 0} f(x)\,\overline{g(x)}.$$

This space is not separable because, for any sequence of functions $f_n \in H$, there are nonzero functions f such that $\langle f, f_n\rangle = 0$ for all $n \in \mathbb{N}$.

Recall that a set S in a Banach space X is called dense in X if every element of X can be approximated by a sequence of elements of S. More precisely, for every $x \in X$, there exist $x_n \in S$ such that $\|x - x_n\| \to 0$ as $n \to \infty$.

Theorem 2.14.1. *Every separable Hilbert space contains a countable dense subset.*

Proof. Let $\{x_n\}$ be a complete orthonormal sequence in a Hilbert space H. The set

$$S = \left\{(\alpha_1 + i\beta_1)x_1 + \cdots + (\alpha_n + i\beta_n)x_n : \alpha_1, \ldots, \alpha_n; \beta_1, \ldots, \beta_n \in \mathbb{Q}, n \in \mathbb{N}\right\}$$

is obviously countable. Since, for every $x \in H$,

$$\left\| \sum_{k=1}^{n} \langle x, x_k \rangle x_k - x \right\| \to 0 \qquad \text{as } n \to \infty,$$

the set S is dense in H.

The statement in the preceding theorem is often used as a definition of separability.

Theorem 2.14.2. *Every orthogonal set in a separable Hilbert space is countable.*

Proof. Let S be an orthogonal set in a separable Hilbert space H, and let S_1 be the set of normalized vectors from S, that is, $S_1 = \{x/\|x\| : x \in S\}$. For any distinct $x, y \in S_1$, we have

$$\begin{aligned}
\|x - y\|^2 &= \langle x - y, x - y \rangle \\
&= \langle x, x \rangle - \langle x, y \rangle - \langle y, x \rangle + \langle y, y \rangle \\
&= 1 - 0 - 0 + 1 \qquad \text{(by the orthogonality)} \\
&= 2.
\end{aligned}$$

This means that the distance between any two distinct elements of S_1 is $\sqrt{2}$.

Now, consider the collection of $(1/\sqrt{2})$-neighborhoods about every element of S_1. Clearly, no two of these neighborhoods can have a common point. Since every dense subset of H must have at least one point in every neighborhood and H has a countable dense subset, S_1 must be countable. Thus, S is countable.

Definition 2.14.2 (Hilbert Space Isomorphism). A Hilbert space H_1 is said to be *isomorphic* to a Hilbert space H_2 if there exists a one-to-one linear mapping T from H_1 onto H_2 such that

$$\langle T(x), T(y) \rangle = \langle x, y \rangle \tag{2.14.1}$$

for every $x, y \in H_1$. Such a mapping T is called a *Hilbert space isomorphism* of H_1 onto H_2.

Note that (2.14.1) implies $\|T\| = 1$ because $\|T(x)\| = \|x\|$ for every $x \in H_1$.

Theorem 2.14.3. *Let H be a separable Hilbert space.*

(a) If H is infinite-dimensional, then it is isomorphic to the space l^2;
(b) If H has dimension N, then it is isomorphic to the space $\mathbb{C}^N$.

Proof. Let $\{x_n\}$ be a complete orthonormal sequence in H. If H is infinite dimensional, then $\{x_n\}$ is an infinite sequence. Let x be an element of H. Define $T(x) = \{\alpha_n\}$, where $\alpha_n = \langle x, x_n\rangle, n = 1, 2, \ldots$. By Theorem 2.10.3, T is a one-to-one mapping from H onto l^2. It is clearly a linear mapping. Moreover for $\alpha_n = \langle x, x_n\rangle$ and $\beta_n = \langle y, x_n\rangle, x, y \in H, n \in \mathbb{N}$, we have

$$\begin{aligned}\big\langle T(x), T(y)\big\rangle &= \big\langle (\alpha_1, \alpha_1, \ldots), (\beta_1, \beta_1, \ldots)\big\rangle \\ &= \sum_{n=1}^{\infty} \alpha_n \overline{\beta_n} = \sum_{n=1}^{\infty} \langle x, x_n\rangle \overline{\langle y, x_n\rangle} \\ &= \sum_{n=1}^{\infty} \big\langle x, \langle y, x_n\rangle x_n\big\rangle = \left|\left\langle x, \sum_{n=1}^{\infty} \langle y, x_n\rangle x_n \right\rangle\right| = \langle x, y\rangle.\end{aligned}$$

Thus, T is an isomorphism from H onto l^2.

The proof of (b) is left as an exercise.

It is easy to check that isomorphism of Hilbert spaces is an equivalence relation.

Since any infinite dimensional separable Hilbert space is isomorphic to the space l^2, it follows that any two such spaces are isomorphic. The same is true for real Hilbert spaces; any real infinite dimensional separable Hilbert space is isomorphic to the real space l^2. In some sense, there is only one real and one complex infinite dimensional separable Hilbert space.

2.15 Linear Operators on Hilbert Spaces

The concept of an operator (or transformation) on a Hilbert space is a natural generalization of the idea of a function of a real variable. Indeed, it is fundamental in mathematics, science, and engineering. Linear operators on a Hilbert space are widely used to represent physical quantities, and hence, they are more important and useful. The most important operators include differential, integral, and matrix operators. In signal processing and wavelet analysis, almost all algorithms are mainly based on linear operators.

Definition 2.15.1 (Linear Operator). An operator T of a vector space X into another vector space Y, where X and Y have the same scalar field, is called a *linear operator* if

$$T\big(ax_1 + bx_2\big) = a\,Tx_1 + b\,Tx_2 \tag{2.15.1}$$

for all scalars a, b and for all $x_1, x_2 \in X$.

Otherwise, it is called a *nonlinear operator*.

Example 2.15.1 (Integral Operator). One of the most important operators is the *integral operator* T defined by

$$Tx(s) = \int_a^b K(s,t)\,x(t)\,dt \tag{2.15.2}$$

where a and b are finite or infinite. The function K is called the *kernel* of the operator.

Example 2.15.2 (Differential Operator). Another important operator is called the *differential* operator

$$\big(Df\big)(x) = \frac{df(x)}{dx} = f'(x) \tag{2.15.3}$$

defined on the space of all differentiable functions on some interval $[a,b] \subset \mathbb{R}$, which is a linear subspace of $L^2\big([a,b]\big)$.

Example 2.15.3 (Matrix Operator). Consider an operator T on $\mathbb{C}^n$, and let $\{e_1, e_2, \ldots, e_n\}$ be the standard base in $\mathbb{C}^n$, that is, $e_1 = (1,0,0,\ldots,0), e_2 = (0,1,0,\ldots,0), \ldots, e_n = (0,0,\ldots,1)$.

We define

$$a_{ij} = \langle Te_j, e_i\rangle \qquad \text{for all } i,j \in \{1,2,\ldots,n\}\,.$$

Then, for $x = \sum_{j=1}^{n} a_j\, e_j \in \mathbb{C}^n$, we have

$$Tx = \sum_{j=1}^{n} a_j\, Te_j \tag{2.15.4}$$

and hence

$$\langle Tx, e_i\rangle = \sum_{j=1}^{n} a_j\,\langle Te_j, e_i\rangle = \sum_{j=0}^{n} a_{ij}\, a_j\,. \tag{2.15.5}$$

Thus, every operator T on the space $\mathbb{C}^n$ is defined by an $n \times n$ matrix.

Conversely, for every $n \times n$ matrix $\big(a_{ij}\big)$, formula (2.15.5) defines an operator on $\mathbb{C}^n$. We thus have a one-to-one correspondence between operators on an n-dimensional vector space and $n \times n$ matrices.

Definition 2.15.2 (Bounded Operator). An operator $T : X \to X$ is called *bounded* if there exists a number K such that

$$\|Tx\| \le K\|x\| \qquad \text{for every } x \in X.$$

The norm of an operator T is defined as the least of all such number K or, equivalently, by

$$\|T\| = \sup_{\|x\|=1} \|T\|.$$

It follows from this definition that

$$\|Tx\| \le \|T\|\|x\|.$$

If the operator T is defined by the matrix (a_{ij}) in Example 2.15.3, then

$$\|T\| \le \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|^2}. \tag{2.15.6}$$

This means that every operator on $\mathbb{C}^n$, and thus also every operator on any finite dimensional Hilbert space is bounded.

The differential operator defined in Example 2.15.2 is *unbounded*. Consider the sequence of functions $f_n(x) = \sin nx, n = 1, 2, 3, \ldots$ defined on $[-\pi, \pi]$. Then,

$$\|f_n\| = \left\{\int_{-\pi}^{\pi} \sin^2 nx\, dx\right\}^{\frac{1}{2}} = \sqrt{\pi}$$

and

$$\|Df_n\| = \left\{\int_{-\pi}^{\pi} (n\cos nx)^2 dx\right\}^{\frac{1}{2}} = n\sqrt{\pi}.$$

Thus,

$$\|Df_n\| = n\|f_n\| \to \infty \qquad \text{as } n \to \infty.$$

Definition 2.15.3 (Continuous Operator). A linear operator $T : X \to Y$, where X and Y are normed spaces, is *continuous* at a point $x_0 \in X$, if, for any sequence $\{x_n\}$ of elements in X convergent to x_0, the sequence $\{T(x_n)\}$ converges to $T(x_0)$. In other words, T is continuous at x_0 if $\|x_n - x_0\| \to 0$ implies $\|T(x_n) - T(x_0)\| \to 0$. If T is continuous at every point $x \in X$, we simply say that T is continuous in X.

Theorem 2.15.1. *A linear operator is continuous if and only if it is bounded.*

The proof is fairly simple (see Debnath and Mikusinski 1999, p. 22) and omitted here.

Two operators T and S on a vector space X are said to be *equal*, $T = S$, if $Tx = Sx$ for every $x \in X$. The set of all operators forms a vector space with the addition and multiplication by a scalar defined by

$$(T + S)x = Tx + Sx,$$

$$(\alpha T)x = \alpha Tx.$$

The product TS of operators T and S is defined by

$$(TS)(x) = T(Sx).$$

In general, $TS \neq ST$. Operators T and S for which $TS = ST$ are called *commuting operators*.

Example 2.15.4. Consider the space of differentiable functions on $\mathbb{R}$ and the operators

$$Tf(x) = xf(x) \quad \text{and} \quad D = \frac{d}{dx}.$$

It is easy to check that $TD \neq DT$.

The square of an operator T is defined as $T^2x = T(Tx)$. Using the principle of induction, we can define any power of T by

$$T^n x = T\left(T^{n-1}x\right).$$

Theorem 2.15.2. *The product TS of bounded operators T and S is bounded and*

$$\|TS\| \leq \|T\|\|S\|.$$

Proof. Suppose T and S are two bounded operators on a normed space X; $\|T\| = k_1$ and $\|S\| = k_2$. Then,

$$\|TSx\| \leq k_1\|Sx\| \leq k_1k_2\|x\| \qquad \text{for every } x \in X.$$

This proves the theorem.

Theorem 2.15.3. *A bounded operator on a separable infinite dimensional Hilbert space can be represented by an infinite matrix.*

Proof. Suppose T is a bounded operator on a Hilbert space H and $\{e_n\}$ is a complete orthonormal sequence in H. For $i, j \in \mathbb{N}$, define

$$a_{ij} = \langle Te_j, e_i \rangle.$$

For any $x \in H$, we have

$$\begin{aligned} Tx &= T\left(\lim_{n\to\infty}\sum_{j=1}^{n}\langle x, e_j\rangle e_j\right) \\ &= \lim_{n\to\infty} T\left(\sum_{j=1}^{n}\langle x, e_j\rangle e_j\right), \quad \text{by continuity of } T \\ &= \lim_{n\to\infty}\left(\sum_{j=1}^{n}\langle x, e_j\rangle Te_j\right), \quad \text{by linearity of } T \\ &= \sum_{j=1}^{\infty}\langle x, e_j\rangle Te_j. \end{aligned}$$

Now,

$$\langle Tx, e_i\rangle = \left\langle \sum_{j=1}^{\infty}\langle x, e_j\rangle Te_j, e_i\right\rangle = \sum_{j=1}^{\infty}\langle Te_j, e_i\rangle\langle x, e_j\rangle = \sum_{j=1}^{\infty} a_{ij}\,\langle x, e_j\rangle.$$

This shows that T is represented by the matrix $\left(a_{ij}\right)$.

Suppose T is a bounded operator on a Hilbert space H. For every fixed $x_0 \in H$, the functional f defined on H by

$$f(x) = \langle Tx, x_0\rangle$$

is a bounded linear functional on H. Thus, by the Riesz representation theorem, there exists a unique $y_0 \in H$ such that $f(x) = \langle x, y_0\rangle$ for all $x \in H$. Or, equivalently, $\langle Tx, x_0\rangle = \langle x, y_0\rangle$ for all $x \in H$. If we denote by T^* the operator which to every $x_0 \in H$ assigns that unique y_0, then we have

$$\langle Tx, y\rangle = \langle x, T^* y\rangle \qquad \text{for all } x, y \in H.$$

Definition 2.15.4 (Adjoint Operator). If T is a bounded linear operator on a Hilbert space H, the operator $T^* : H \to H$ defined by

$$\left\langle Tx, y\right\rangle = \left\langle x, T^* y\right\rangle \qquad \text{for all } x, y \in H$$

is called the *adjoint operator* of T.

The following are immediate consequences of the preceding definition.

$$\left(T + S\right)^* = T^* + S^*, \qquad \left(\alpha T\right)^* = \overline{\alpha}\, T^*,$$

$$\left(T^*\right)^* = T, \quad I^* = I, \qquad \left(TS\right)^* = S^* T^*,$$

for arbitrary operators T and S, I is the identity operator and for any scalar α.

Theorem 2.15.4. *The adjoint operator T^* of a bounded operator T is bounded. Moreover,*

$$\|T\| = \|T^*\| \quad \textit{and} \quad \|TT^*\| = \|T\|^2.$$

Proof. The reader is referred to Debnath and Mikusinski (1999, p. 151).

In general, $T \neq T^*$. For example, suppose $H = \mathbb{C}^2$, and suppose T is defined by

$$T(z_1, z_2) = \langle 0, z_1 \rangle.$$

Then,

$$\big\langle T(x_1, x_2), (y_1, y_2) \big\rangle = x_1\, \bar{y}_2 \quad \text{and} \quad \big\langle (x_1, x_2), T(y_1, y_2) \big\rangle = x_2\, \bar{y}_1.$$

However, operators for which $T = T^*$ are of special interest.

Definition 2.15.5 (Self-adjoint Operator). If $T = T^*$, that is, $\langle Tx, y \rangle = \langle x, Ty \rangle$ for all $x, y \in H$, then T is called *self-adjoint* (or *Hermitian*).

Example 2.15.5. Suppose $H = \mathbb{C}^n$ and that $\{e_1, e_2, \ldots, e_n\}$ is a standard orthonormal base in H. Suppose T is an operator represented by the matrix (a_{ij}), where $a_{ij} = \langle Te_j, e_i \rangle$ (see Example 2.15.3). Then, the adjoint operator T^* is represented by the matrix $b_{kj} = \langle T^* e_j, e_k \rangle$. Consequently,

$$b_{kj} = \langle e_j, Te_k \rangle = \overline{\langle Te_k, e_j \rangle} = \overline{a_{jk}}.$$

Therefore, the operator T is self-adjoint if and only if $a_{ij} = \overline{a_{ji}}$. A matrix that satisfies this condition is often called *Hermitian*.

Example 2.15.6. Suppose H is a separable, infinite-dimensional Hilbert space, and suppose $\{e_n\}$ is a complete orthonormal sequence in H. If T is a bounded operator on H represented by an infinite matrix (a_{ij}), the operator T is self-adjoint if and only if $a_{ij} = \overline{a_{ji}}$ for all $i, j \in \mathbb{N}$.

Example 2.15.7. Suppose T is a Fredholm operator on $L^2([a,b])$ defined by (2.15.2), where the kernel K is defined on $[a,b] \times [a,b]$ such that

$$\int_a^b \int_a^b \big|K(s,t)\big|^2 ds\, dt < \infty.$$

This condition is satisfied if K is continuous. We have

$$\begin{aligned}\langle Tx, y \rangle &= \int_a^b \int_a^b K(s,t)\, x(t)\, \overline{y(s)}\, ds\, dt \\ &= \overline{\int_a^b \int_a^b \overline{K(s,t)\, x(t)}\, y(s)\, ds\, dt}\end{aligned}$$

$$= \int_a^b x(t) \overline{\int_a^b \overline{K(s,t)}\, y(s)\, ds\, dt}$$

$$= \left\langle x, \int_a^b \overline{K(s,t)}\, y(t)\, dt \right\rangle.$$

This shows that

$$\big(T^* x\big)(s) = \int_a^b \overline{K(s,t)}\, x(t)\, dt.$$

Thus, the Fredholm operator is self-adjoint if its kernel satisfies the equality $K(s,t) = \overline{K(t,s)}$.

Example 2.15.8. The operator T on $L^2\big([a,b]\big)$ defined by $\big(Tx\big)(t) = tx(t)$ is self-adjoint.

We have

$$\langle Tx, y\rangle = \int_a^b t\, x(t)\, \overline{y(t)}\, dt = \int_a^b x(t)\, \overline{t\, y(t)}\, dt = \langle x, Ty\rangle.$$

Example 2.15.9. The operator T defined on $L^2(\mathbb{R})$ defined by $\big(Tx\big)(t) = e^{-|t|}x(t)$ is bounded and self-adjoint.

The fact that T is self-adjoint follows from

$$\langle Tx, y\rangle = \int_{-\infty}^{\infty} e^{-|t|}x(t)\, \overline{y(t)}\, dt = \int_{-\infty}^{\infty} x(t)\, \overline{e^{-|t|}\, y(t)}\, dt = \langle x, Ty\rangle.$$

The proof of boundedness is left as an exercise.

Theorem 2.15.5. *If T is a bounded operator on a Hilbert space H, the operators $A = T + T^*$ and $B = T^*T$ are self-adjoint.*

Proof. For all $x, y \in H$, we have

$$\langle Ax, y\rangle = \left\langle \big(T + T^*\big)x, y\right\rangle = \left\langle x, \big(T + T^*\big)^* y\right\rangle = \left\langle x, \big(T + T^*\big)y\right\rangle = \langle x, Ay\rangle$$

and

$$\langle Bx, y\rangle = \left\langle T^* Tx, y\right\rangle = \langle Tx, Ty\rangle = \left\langle x, T^* Ty\right\rangle = \langle x, By\rangle.$$

Theorem 2.15.6. *The product of two self-adjoint operators is self-adjoint if and only if they commute.*

Proof. Suppose T and S are two self-adjoint operators. Then,

$$\langle TSx, y\rangle = \langle Sx, Ty\rangle = \langle x, STy\rangle$$

Thus, if $TS = ST$, then TS is self-adjoint.

Conversely, if TS is self-adjoint, then the above implies $TS = (TS)^* = ST$.

Example 2.15.10. Consider the differential operator D in the space of all differentiable functions on $\mathbb{R}$ vanishing at infinity. Then,

$$\begin{aligned}\langle Dx, y\rangle &= \int_{-\infty}^{\infty} \frac{d}{dt} x(t) \cdot \overline{y(t)}\, dt = -\int_{-\infty}^{\infty} x(t) \cdot \frac{d}{dt} \overline{y(t)}\, dt \\ &= \int_{-\infty}^{\infty} x(t) \cdot \overline{\left(-\frac{d}{dt} y(t)\right)}\, dt = \langle x, -Dy\rangle.\end{aligned}$$

Thus, $-D$ is the adjoint of the operator D.

Example 2.15.11. Consider the operator $T = i\dfrac{d}{dt}$ in the space of all differentiable functions on $\mathbb{R}$ vanishing at infinity.

We have

$$\begin{aligned}\langle Tx, y\rangle &= \int_{-\infty}^{\infty} i\frac{d}{dt} x(t) \cdot \overline{y(t)}\, dt = -i\int_{-\infty}^{\infty} x(t) \cdot \frac{d}{dt} \overline{y(t)}\, dt \\ &= \int_{-\infty}^{\infty} x(t) \cdot \overline{\left(i\frac{d}{dt} y(t)\right)}\, dt = \langle x, Ty\rangle.\end{aligned}$$

Therefore, T is a self-adjoint operator.

Theorem 2.15.7. *For every bounded operator T on a Hilbert space H, there exist unique self-adjoint operators A and B such that $T = A + iB$ and $T^* = A - iB$.*

Proof. Suppose T is a bounded operator on H. Define

$$A = \frac{1}{2}\left(T + T^*\right) \quad \text{and} \quad B = \frac{1}{2}\left(T - T^*\right).$$

Evidently, A and B are self-adjoint and $T = A + iB$. Moreover, for any $x, y \in H$, we have

$$\begin{aligned}\langle Tx, y\rangle &= \big\langle (A + iB)x, y\big\rangle = \big\langle Ax, y\big\rangle + i\big\langle Bx, y\big\rangle \\ &= \big\langle x, Ay\big\rangle + i\big\langle x, By\big\rangle = \big\langle x, (A - iB)y\big\rangle.\end{aligned}$$

Hence, $T^* = A - iB$.

The proof of uniqueness is left as an exercise.

In particular, if T is self-adjoint, then $T = A$ and $B = 0$. This implies that self-adjoint operators are like real numbers in $\mathbb{C}$.

We next discuss projection operators and their properties.

According to the projection Theorem 2.12.4, if S is a closed subspace of a Hilbert space H, then for every $x \in H$, there exists a unique element $y \in S$ such that $x = y + z$ and $z \in S^{\perp}$. Thus, every closed subspace induces an operator on H which assigns to x that unique y.

Definition 2.15.6 (Anti-Hermitian Operator). An operator A is called *anti-Hermitian* if $A = -A^*$.

The operator in Example 2.15.10 is anti-Hermitian.

Definition 2.15.7 (Inverse Operator). Let T be an operator defined on a vector subspace of X. An operator S defined on $R(T)$ is called the *inverse* of T if $TSx = x$ for all $x \in R(T)$ and $STx = x$ for all $x \in D(T)$. An operator which has an inverse is called *invertible*. The inverse of T will be denoted by T^{-1}.

If an operator has an inverse, then it is unique. Indeed, suppose S_1 and S_2 are inverses of T. Then

$$S_1 = S_1 I = S_1 T S_2 = I S_2 = S_2.$$

Note also that

$$D(T^{-1}) = R(T) \qquad \text{and} \qquad R(T^{-1}) = D(T).$$

First, we recall some simple algebraic properties of invertible operators.

Theorem 2.15.8. *(a) The inverse of a linear operator is a linear operator.*
(b) An operator T is invertible if and only if $Tx = 0$ implies $x = 0$.
(c) If an operator T is invertible and vectors $x_1, \dots, x_n$ are linearly independent, then $Tx_1, \dots, Tx_n$ are linearly independent.
(d) If operators T and S are invertible, then the operator TS is invertible and we have $(TS)^{-1} = S^{-1}T^{-1}$.

Proof. (a) For any $x, y \in R(T)$ and $\alpha, \beta \in \mathbb{C}$, we have

$$\begin{aligned} T^{-1}(\alpha x + \beta y) &= T^{-1}\big(\alpha\, T T^{-1} x + \beta\, T T^{-1} y\big) \\ &= T^{-1} T\big(\alpha\, T^{-1} x + \beta\, T^{-1} y\big) = \alpha\, T^{-1} x + \beta\, T^{-1} y. \end{aligned}$$

(b) If T is invertible and $Tx = 0$, then $x = T^{-1}Tx = T^{-1}0 = 0$. Assume now that $Tx = 0$ implies $x = 0$. If $Tx_1 = Tx_2$, then $T(x_1 - x_2) = 0$ and thus $x_1 - x_2 = 0$. Consequently, $x_1 - x_2 = O$. which proves that T is invertible.

(c) Suppose $\alpha_1 Tx_1 + \cdots + \alpha_n Tx_n = 0$. Then, $T\big(\alpha_1 x_1 + \cdots + \alpha_n x_n\big) = 0$, and since T is invertible, $\alpha_1 x_1 + \cdots + \alpha_n x_n = 0$. Linear independence of $x_1, \dots, x_n$ implies $\alpha_1 = \cdots = \alpha_n = 0$. Thus, vectors $Tx_1, \dots, Tx_n$ are linearly independent.

(d) In view of (b). if $T(Sx) = 0$, then $Sx = 0$ since T is invertible. If $Sx = 0$, then $x = 0$, since S is invertible. Thus, TS is invertible by (b). Moreover,

$$\big(S^{-1}T^{-1}\big)\big(TS\big) = S^{-1}\big(T^{-1}T\big)S = S^{-1}S = I.$$

Similarly, $\big(TS\big)\big(S^{-1}T^{-1}\big) = I$. This proves that $\big(TS\big)^{-1} = S^{-1}T^{-1}$.

It follows from part (c) in the preceding theorem that if X is a finite dimensional vector space and T is a linear invertible operator on X, then $R(T) = X$. As the following example shows, in infinite dimensional vector spaces this is not necessarily true.

Example 2.15.12. Let $X = l^2$. Define an operator T on X by

$$T(x_1, x_2, \dots) = (0, x_1, x_2, \dots).$$

Clearly. this is a linear invertible operator on l^2 whose range is a proper subspace of l^2.

The next example shows that the inverse of a bounded operator is not necessarily bounded.

Example 2.15.13. Let $X = l^2$. Define an operator T on X by

$$T(x_1, x_2, \dots) = \left(x_1, \frac{x_2}{2}, \frac{x_3}{3}, \dots, \frac{x_n}{n}, \dots\right).$$

Since

$$\|T(x_1, x_2, \dots)\| = \sqrt{\sum_{n=1}^{\infty} \frac{|x_n|^2}{n^2}} \le \sqrt{\sum_{n=1}^{\infty} |x_n|^2} = \|(x_1, x_2, \dots)\|,$$

T is a bounded operator. T is also invertible:

$$T^{-1}(x_1, x_2, \dots) = (x_1, 2x_2, 3x_3, \dots, nx_n, \dots).$$

However, T^{-1} is not bounded. In fact, consider the sequence $\{e_n\}$ of elements of l^2, where $\{e_n\}$ is the sequence whose nth term is 1 and all the remaining terms are 0. Then, $\|e_n\| = 1$ and $\|T^{-1}e_n\| = n$. Therefore, T^{-1} is unbounded.

If X is finite dimensional, then the inverse of any invertible operator on X is bounded because every operator on a finite dimensional space is bounded.

Theorem 2.15.9. *Let T be a bounded operator on a Hilbert space H such that $R(T) = H$. If T has a bounded inverse, then the adjoint T^* is invertible and $(T^*)^{-1} = (T^{-1})^*$.*

Proof. It suffices to show that

$$(T^{-1})^* T^* x = T^* (T^{-1})^* x = x \tag{2.15.7}$$

for every $x \in H$. Indeed, for any $y \in H$, we have

$$\left\langle y, (T^{-1})^* T^* x \right\rangle = \langle T^{-1} y, T^* x \rangle = \langle T T^{-1} y, x \rangle = \langle y, x \rangle$$

and

$$\left\langle y, T^*\left(T^{-1}\right)^* x\right\rangle = \left\langle Ty, \left(T^{-1}\right)^* x\right\rangle = \langle T^{-1}Ty, x\rangle = \langle y, x\rangle.$$

Thus,

$$\left\langle y, \left(T^{-1}\right)^* T^* x\right\rangle = \left\langle y, T^*\left(T^{-1}\right)^* x\right\rangle = \langle y, x\rangle \quad \text{for all } y \in H. \tag{2.15.8}$$

This implies (2.15.7).

Corollary 2.15.1. *If a bounded self-adjoint operator T has bounded inverse T^{-1}, then T^{-1} is self-adjoint.*

Proof. $\left(T^{-1}\right)^* = \left(T^*\right)^{-1} = T^{-1}$.

Definition 2.15.8 (Isometric Operator). A bounded operator T on a Hilbert space H is called an *isometric operator* if $\|Tx\| = \|x\|$ for all $x \in H$.

Example 2.15.14. Let $\{e_n\}, n \in \mathbb{N}$, be a complete orthonormal sequence in a Hilbert space H. There exists a unique operator T such that $Te_n = e_{n+1}$ for all $n \in \mathbb{N}$. In fact, if $x = \sum_{n=1}^{\infty} \alpha_n e_n$, then $Tx = \sum_{n=1}^{\infty} \alpha_n e_{n+1}$. Clearly, T is linear and $\|Tx\|^2 = \sum_{n=1}^{\infty} |\alpha_n|^2 = \|x\|$. Therefore, T is an isometric operator. The operator T is called a *one-sided shift operator.*

Theorem 2.15.10. *A bounded operator T defined on a Hilbert space H is isometric if and only if $T^*T = I$ on H.*

Proof. If T is isometric, then for every $x \in H$ we have $\|Tx\|^2 = \|x\|^2$ and hence,

$$\langle T^*Tx, x\rangle = \langle Tx, Tx\rangle = \langle x, x\rangle \qquad \text{for all } x \in H.$$

This implies that $T^*T = I$. Similarly, if $T^*T = I$, then

$$\|Tx\| = \sqrt{\langle Tx, Tx\rangle} = \sqrt{\langle T^*Tx, x\rangle} = \sqrt{\langle x, x\rangle} = \|x\|.$$

Note that isometric operators "preserve inner product": $\langle Tx, Ty\rangle = \langle x, y\rangle$. In particular, $x \perp y$ if and only if $Tx \perp Ty$. The operator in Example 2.15.12 is an isometric operator.

Definition 2.15.9 (Unitary Operator). A bounded operator T on a Hilbert space H is called a *unitary operator* if $T^*T = TT^* = I$ on H.

In the above definition it is essential that the domain and the range of T be the entire space H.

Theorem 2.15.11. *An operator T is unitary if and only if it is invertible and $T^{-1} = T^*$.*

Proof. Assume that T is an invertible operator on a Hilbert space H such that $T^{-1} = T^*$. Then, $T^*T = T^{-1}T = I$ and $TT^* = TT^{-1} = I$. Therefore, T is a unitary operator. The proof of the converse is similar.

Theorem 2.15.12. *Suppose T is a unitary operator. Then*

(a) T is isometric,
(b) T^{-1} and T^* are unitary.

Proof. (a) follows from Theorem 2.15.10. To prove (b), note that

$$\left(T^{-1}\right)^* T^{-1} = T^{**}T^{-1} = TT^{-1} = I.$$

Similarly, $T^{-1}\left(T^{-1}\right)^* = I$, and thus, T^{-1} is unitary. Since $T^* = T^{-1}$, by Theorem 2.15.11, T^* is also unitary.

Example 2.15.15. Let H be the Hilbert space of all sequences of complex numbers $x = \{\ldots, x_{-1}, x_0, x_1, \ldots\}$ such that $\|x\| = \sum_{n=-\infty}^{\infty} |x_n|^2 < \infty$. The inner product is defined by

$$\langle x, y\rangle = \sum_{n=-\infty}^{\infty} x_n \overline{y_n}.$$

Define an operator T by $T(x_n) = (x_{n-1})$. T is a unitary operator and hence, T is invertible and

$$\langle Tx, y\rangle = \sum_{n=-\infty}^{\infty} x_{n-1}\overline{y_n} = \sum_{n=-\infty}^{\infty} x_n \overline{y_{n+1}} = \langle x, T^{-1}y\rangle.$$

This implies that $T^* = T^{-1}$.

Example 2.15.16. Let $H = L^2([0,1])$. Define an operator T on H by $(Tx)(t) = x(1-t)$. This operator is a one-to-one mapping of H onto H. Moreover, we have $T = T^* = T^{-1}$. Thus, T is a unitary operator.

Definition 2.15.10 (Positive Operator). An operator T is called *positive* if it is self-adjoint and $\langle Tx, x\rangle \geq 0$ for all $x \in H$.

Example 2.15.17. Let ϕ be a nonnegative continuous function on $[a,b]$. The *multiplication operator* T on $L^2([a,b])$ defined by $Tx = \phi x$ is positive. In fact for any $x \in L^2([a,b])$, we have

$$\langle Tx, x\rangle = \int_a^b \phi(t)\, x(t)\, \overline{x(t)}\, dt = \int_a^b \phi(t)|x(t)|^2 dt \geq 0.$$

Example 2.15.18. Let K be a positive continuous function defined on $[a,b]\times[a,b]$. The integral operator T on $L^2([a,b])$ defined by

$$(Tx)(s) = \int_a^b K(s,t)\,x(t)\,dt$$

is positive. Indeed, we have

$$\langle Tx, x\rangle = \int_a^b \int_a^b K(s,t)\,x(t)\,\overline{x(t)}\,dt\,ds = \int_a^b \int_a^b K(s,t)\big|x(t)\big|^2 dt\,ds \geq 0$$

for all $x \in L^2([a,b])$.

Theorem 2.15.13. *For any bounded operator A on a Hilbert space H, the operators A^*A and AA^* are positive.*

Proof. For any $x \in H$, we have

$$\langle A^*Ax, x\rangle = \langle Ax, Ax\rangle = \|Ax\|^2 \geq 0$$

and

$$\langle AA^*x, x\rangle = \langle A^*x, A^*x\rangle = \|Ax\|^2 \geq 0.$$

Theorem 2.15.14. *If A is an invertible positive operator on a Hilbert space H, then its inverse A^{-1} is positive.*

Proof. If $y \in D(A^{-1})$, then $y = Ax$ for some $x \in H$, and then

$$\langle A^{-1}y, y\rangle = \langle A^{-1}Ax, Ax\rangle = \langle x, Ax\rangle \geq 0.$$

To indicate that A is a positive operator, we write $A \geq 0$. If the difference $A - B$ of two self-adjoint operators is a positive operator, that is, $A - B \geq 0$, then we write $A \geq B$. Consequently,

$$A \geq B \quad \text{if and only if} \quad \langle Ax, x\rangle \geq \langle Bx, x\rangle \quad \text{for all } x \in H.$$

This relation has the following natural properties:

If $A \geq B$ and $C \geq D$, then $A + C \geq B + D$;
If $A \geq 0$ and $\alpha \geq 0 (\alpha \in \mathbb{R})$, then $\alpha A \geq 0$;
If $A \geq B$ and $B \geq C$, then $A \geq C$.

Proofs are left as exercises.

Theorem 2.15.15. *If T is a self-adjoint operator on H and $\|T\| \leq 1$, then $T \leq I$.*

Proof. If $\|T\| \le 1$, then

$$\langle Tx, x\rangle \le \|T\| \|x\|^2 \le \langle x, x\rangle = \langle Ix, x\rangle$$

for all $x \in H$.

Definition 2.15.11 (Orthogonal Projection Operator). If S is a closed subspace of a Hilbert space H, the operator P on H defined by

$$Px = y \quad \text{if } x = y + z, \quad y \in S \quad \text{and} \quad z \in S^{\perp}, \tag{2.15.9}$$

is called the *orthogonal projection operator* onto S, or simply, the *projection operator* onto S. The vector y is called the *projection* of x onto S.

Since the decomposition $x = y + z$ is unique, it follows that projection operators are linear. The Pythagorean formula implies that

$$\|Px\|^2 = \|y\|^2 = \|x\|^2 - \|z\|^2 \le \|x\|^2.$$

This shows that projection operators are bounded and $\|Px\| \le 1$. The zero operator is a projection operator onto the zero subspace. If P is a nonzero projection operator, then $\|Px\| = 1$ because, for every $x \in S$, we have $Px = x$. The identity operation I is the projection operator onto the whole space H.

Moreover, it follows from (2.15.9) that

$$\langle Px, x - Px\rangle = 0 \qquad \text{for every } x \in H.$$

Example 2.15.19. If S is a closed subspace of a Hilbert space H and $\{e_n\}$ is a complete orthonormal system in S, then the projection operator P onto S can be defined by

$$Px = \sum_{n=1}^{\infty} \langle x, e_n\rangle e_n.$$

In particular, if the dimension of S is unity and $u \in S$, $\|u\| = 1$, then $Px = \langle x, u\rangle u$.

Example 2.15.20. Suppose that $H = L^2([-\pi, \pi])$. Every $x \in H$ can be represented as $x = y + z$, where y is an even function and z is an odd function. The operator defined by $Px = y$ is the projection operator onto the subspace of all even functions. This operator can also be defined as in Example 2.15.19:

$$Px = \sum_{n=0}^{\infty} \langle x, \phi_n\rangle \phi_n,$$

where $\phi_0 = \dfrac{1}{\sqrt{2\pi}}$ and $\phi_n(t) = \dfrac{1}{\sqrt{\pi}} \cos nt, \quad n = 1, 2, 3, \ldots.$

Example 2.15.21. Let $H = L^2([-\pi, \pi])$ and P be an operator defined by

$$(Px)(t) = \begin{cases} 0, & t \le 0 \\ x(t), & t > 0 \end{cases}$$

Then, P is the projection operator onto the space of all functions that vanish for $t \le 0$.

Definition 2.15.12 (Idempotent Operator). An operator T is called *idempotent* if $T^2 = T$.

Every projection operator is idempotent. In fact, if P is the projection operator onto a subspace S, then P is the identity operator on S. Since $Px \in S$ for every $x \in H$, it follows that $P^2 = P(Px)$ for all $x \in H$.

Example 2.15.22. Consider the operator T on $\mathbb{C}^2$ defined by $T(x, y) = \langle x - y, 0\rangle$. Obviously, T is idempotent. On the other hand, since

$$\langle T(x, y), (x, y) - T(x, y)\rangle = x\overline{y} - |y|^2,$$

$T(x, y)$ need not be orthogonal to $(x, y) - T(x, y)$ and thus T is not a projection.

Definition 2.15.13 (Compact Operator). An operator T on a Hilbert space H is called a *compact operator* (or *completely continuous operator*) if, for every bounded sequence $\{x_n\}$ in H, the sequence $\{Tx_n\}$ contains a convergent subsequence.

Compact operators constitute an important class of bounded operators. The concept originated from the theory of integral equations of the second kind. Compact operators also provide a natural generalization of operators with finite-dimensional range.

Example 2.15.23. Every operator on a finite dimensional Hilbert space is compact. Indeed, if T is an operator on $\mathbb{C}^N$, then it is bounded. Therefore, if $\{x_n\}$ is a bounded sequence, then $\{Tx_n\}$ is a bounded sequence in $\mathbb{C}^N$. By the Bolzano–Weierstrass theorem, $\{Tx_n\}$ contains a convergent subsequence.

Theorem 2.15.16. *Compact operators are bounded.*

Proof. If an operator T is not bounded, then there exists a sequence $\{x_n\}$ such that $\|x_n\| = 1$, for all $n \in \mathbb{N}$, and $\|Tx_n\| \to \infty$. Then, $\{Tx_n\}$ does not contain a convergent subsequence, which means that T is not compact.

Not every bounded operator is compact.

Example 2.15.24. The identity operator I on an infinite dimensional Hilbert space H is not compact, although it is bounded. In fact. consider an orthonormal sequence $\{e_n\}$ in H. Then, the sequence $Ie_n = e_n$ does not contain a convergent subsequence.

Example 2.15.25. Let y and z be fixed elements of a Hilbert space H. Define

$$Tx = \langle x, y\rangle z.$$

Let $\{x_n\}$ be a bounded sequence, that is, $\|x_n\| \leq M$ for some $M > 0$ and all $n \in \mathbb{N}$. Since

$$|\langle x_n, y\rangle| \leq \|x_n\| \|y\| \leq M \|y\|,$$

the sequence $\{\langle x_n, y\rangle\}$ contains a convergent subsequence $\{\langle x_{p_n}, y\rangle\}$. Denote the limit of that subsequence by α. Then,

$$T x_{p_n} = \langle x_{p_n}, y\rangle z \to \alpha z \quad \text{as } n \to \infty.$$

Therefore, T is a compact operator.

Example 2.15.26. Important examples of compact operators are integral operators T on $L^2([a, b])$ defined by

$$(Tx)(s) = \int_a^b K(s,t)\, x(t)\, dt,$$

where a and b are finite and K is continuous.

Example 2.15.27. Let S be a finite-dimensional subspace of a Hilbert space H. The projection operator P_s is a compact operator.

Theorem 2.15.17. *Let A be a compact operator on a Hilbert space H, and let B be a bounded operator on H. Then, AB and BA are compact.*

Proof. Let $\{x_n\}$ be a bounded sequence in H. Since B is bounded, the sequence $\{Bx_n\}$ is bounded. Next, since A is compact, the sequence $\{ABx_n\}$ contains a convergent subsequence, which means that the operator AB is compact. Similarly, since A is compact, the sequence $\{Ax_n\}$ contains a convergent subsequence $\{Ax_{p_n}\}$. Now, since B is bounded (and thus continuous), the sequence $\{BAx_{p_n}\}$ converges. Therefore, the operator BA is compact.

The operator defined in Example 2.15.27 is a special case of a finite-dimensional operator.

Definition 2.15.14 (Finite-Dimensional Operator). An operator is called *finite-dimensional* if its range is of finite dimension.

Theorem 2.15.18. *Finite-dimensional bounded operators are compact.*

Proof. Let A be a finite-dimensional bounded operator and let $\{z_1, \ldots, z_k\}$ be an orthonormal basis of the range of A. Define

$$T_n x = \langle Ax, z_n\rangle z_n$$

for $n = 1, \ldots, k$. Since

$$T_n x = \langle Ax, z_n\rangle z_n = \langle x, A^* z_n\rangle z_n,$$

the operators T_n are compact, as proved in Example 2.15.25. Since

$$A = \sum_{n=1}^{k} T_n,$$

A is compact because the collection of all compact operators on a Hilbert space H is a vector space.

Theorem 2.15.19. *If $T_1, T_2, \ldots$ are compact operators on a Hilbert space H and $\|T_n - T\| \to 0$ as $n \to \infty$ for some operator T on H, then T is compact.*

Proof. Let $\{x_n\}$ be a bounded sequence in H. Since T_1 is compact, there exists a subsequence $\{x_{1,n}\}$ of $\{x_n\}$ such that $\{T_1 x_{1,n}\}$ is convergent. Similarly, the sequence $\{T_2 x_{1,n}\}$ contains a convergent subsequence $\{T_2 x_{2,n}\}$. In general, for $k \geq 2$, let $\{x_{k,n}\}$ be a subsequence of $\{x_{k-1,n}\}$ such that $\{T_k x_{k,n}\}$ is convergent. Consider the sequence $\{x_{n,n}\}$. Since it is a subsequence of $\{x_n\}$, we can put $x_{p_n} = x_{n,n}$ where $\{p_n\}$ is an increasing sequence of positive integers. Obviously, the sequence $\{T_k x_{p_n}\}$ converges for every $k \in \mathbb{N}$. We will show that the sequence $\{T x_{p_n}\}$ also converges.

Let $\varepsilon > 0$. Since $\|T_n - T\| \to 0$, there exists $k \in \mathbb{N}$ such that $\|T_k - T\| \leq \frac{\varepsilon}{3M}$, where M is a constant such that $\|x_n\| \leq M$ for all $n \in \mathbb{N}$. Next, let $k_1 \in \mathbb{N}$ be such that

$$\|T_k x_{p_n} - T_k x_{p_m}\| \leq \frac{\varepsilon}{3}$$

for all $n, m > k_1$. Then,

$$\begin{aligned}\|T x_{p_n} - T x_{p_m}\| &\leq \|T x_{p_n} - T_k x_{p_n}\| + \|T_k x_{p_n} - T_k x_{p_m}\| + \|T_k x_{p_m} - T x_{p_m}\| \\ &\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon\end{aligned}$$

for sufficiently large n and m. Thus, $\{T x_{p_n}\}$ is a Cauchy sequence in H. Completeness of H implies that $\{T x_{p_n}\}$ is convergent.

Corollary 2.15.2. *The limit of a convergent sequence of finite-dimensional operators is a compact operator.*

Proof. Finite-dimensional operators are compact.

Theorem 2.15.20. *The adjoint of a compact operator is compact.*

Proof. Let T be a compact operator on a Hilbert space H, and let $\{x_n\}$ be a bounded sequence in H, that is, $\|x_n\| \leq M$ for some M for all $n \in \mathbb{N}$. Define $y_n = T^* x_n, n = 1, 2, \ldots$. Since T^* is bounded, the sequence $\{y_n\}$ is bounded. It thus contains a subsequence $\{y_{k_n}\}$ such that the sequence $\{T y_{k_n}\}$ converges in H. Now, for any $m, n \in \mathbb{N}$, we have

$$\begin{aligned}
\left\| y_{k_m} - y_{k_n} \right\|^2 &= \left\| T^* x_{k_m} - T^* x_{k_n} \right\|^2 \\
&= \left\langle T^* \left(x_{k_m} - x_{k_n} \right), T^* \left(x_{k_m} - x_{k_n} \right) \right\rangle \\
&= \left\langle T T^* \left(x_{k_m} - x_{k_n} \right), \left(x_{k_m} - x_{k_n} \right) \right\rangle \\
&\le \left\| T T^* \left(x_{k_m} - x_{k_n} \right) \right\| \left\| x_{k_m} - x_{k_n} \right\| \\
&\le 2M \left\| T y_{k_m} - T y_{k_n} \right\| \to 0, \quad \text{as } m, n \to \infty.
\end{aligned}$$

Therefore, $\{y_{k_n}\}$ is a Cauchy sequence in H, which implies that $\{y_{k_n}\}$ converges. This proves that T^* is a compact operator.

In the next theorem, we characterize compactness of operators in terms of weakly convergent sequences. Recall that we write "$x_n \to x$" to denote strong convergence and "$x_n \overset{w}{\to} x$" to denote weak convergence.

Theorem 2.15.21. *An operator T on a Hilbert space H is compact if and only if $x_n \overset{w}{\to} x$ implies $T x_n \to T x$.*

Proof. For a proof of this theorem, the reader is referred to Debnath and Mikusinski (1999).

Corollary 2.15.3. *If T is a compact operator on a Hilbert space H and $\{x_n\}$ is an orthonormal sequence in H, then* $\lim_{n\to\infty} T x_n = 0$.

Proof. Orthonormal sequences are weakly convergent to 0.

It follows from the above theorem that the inverse of a compact operator on an infinite-dimensional Hilbert space, if it exists, is unbounded.

It has already been noted that compactness of operators is a stronger condition than boundedness. For operators, boundedness is equivalent to continuity. Bounded operators are exactly those operators that map strongly convergent sequences into strongly convergent sequences. Theorem 2.15.21 states that compact operators on a Hilbert space can be characterized as those operators which map weakly convergent sequences into strongly convergent sequences. From this point of view, compactness of operators is a stronger type of continuity. For this reason, compact operators are sometimes called *completely continuous operators*. The above condition has been used by F. Riesz as the definition of compact operators. Hilbert used still another (equivalent) definition of compact operators: an operator T defined on a Hilbert space H is compact $x_n \to x$ weakly and $y_n \to y$ weakly implies $\langle T x_n, y_n \rangle \to \langle T x, y \rangle$ strongly.

2.16 Eigenvalues and Eigenvectors of an Operator

This section deals with concepts of eigenvalues and eigenvectors which play a central role in the theory of operators.

Definition 2.16.1 (Eigenvalue). Let T be an operator on a complex vector space X. A complex number A is called an *eigenvalue* of T if there is a nonzero vector $u \in X$ such that

$$Tu = \lambda u. \tag{2.16.1}$$

Every vector u satisfying (2.16.1) is called an *eigenvector* of T corresponding to the eigenvalue λ. If X is a function space, eigenvectors are often called *eigenfunctions*.

Example 2.16.1. Let S be a linear subspace of an inner product space X, and T be the projection on S. The only eigenvalues of T are 0 and 1. Indeed, if, for some $\lambda \in \mathbb{C}$ and $0 \neq u \in X$, we have $Tu = \lambda u$, then

$$\lambda u = \lambda^2 u,$$

because $T^2 = T$. Therefore, $\lambda = 0$ or $\lambda = 1$. The eigenvectors corresponding to 0 are the vectors of X which are orthogonal to S. The eigenvectors corresponding to 1 are all elements of S.

It is important to note that every eigenvector corresponds to exactly one eigenvalue, but there are always infinitely many eigenvectors corresponding to an eigenvalue. Indeed, every multiple of an eigenvector is an eigenvector. Moreover, several linearly independent vectors may correspond to the same eigenvalue. We have the following simple theorem.

Theorem 2.16.1. *The collection of all eigenvectors corresponding to one particular eigenvalue of an operator is a vector space.*

The easy proof is left as an exercise.

Definition 2.16.2 (Eigenvalue Space). The set of all eigenvectors corresponding to one particular eigenvalue λ is called the *eigenvalue space* of λ. The dimension of that space is called the *multiplicity* of λ. An eigenvalue of multiplicity one is called *simple* or *nondegenerate*. In such a case, the number of linearly independent eigenvectors is also called the *degree of degeneracy*.

Example 2.16.2. Consider the integral operator $T : L^2([0, 2\pi]) \to L^2([0, 2\pi])$ defined by

$$(Tu)(t) = \int_0^{2\pi} \cos(t - y)\, u(y)\, dy. \tag{2.16.2}$$

We will show that T has exactly one nonzero eigenvalue $\lambda = \pi$, and its eigenfunctions are

$$u(t) = a\cos t + b\sin t$$

with arbitrary a and b.

The eigenvalue equation is

$$\big(Tu\big)(t) = \int_0^{2\pi} \cos(t-y)\,u(y)\,dy = \lambda\,u(t).$$

Or,

$$\cos t \int_0^{2\pi} u(y)\cos y\,dy + \sin t \int_0^{2\pi} u(y)\sin y\,dy = \lambda\,u(t). \tag{2.16.3}$$

This means that, for $\lambda \neq 0$, u is a linear combination of cosine and sine functions, that is,

$$u(t) = a\cos t + b\sin t, \tag{2.16.4}$$

where $a, b \in \mathbb{C}$. Substituting this into (2.16.3), we obtain

$$\pi a = \lambda\,a \quad \text{and} \quad \pi b = \lambda\,b. \tag{2.16.5}$$

Hence, $\lambda = \pi$, which means that T has exactly one nonzero eigenvalue and its eigenfunctions are given by (2.16.4). This is a two-dimensional eigenspace, so the multiplicity of the eigenvalue is 2.

Equation (2.16.3) reveals that $\lambda = 0$ is also an eigenvalue of T. The corresponding eigenfunctions are all the functions orthogonal to $\cos t$ and $\sin t$. Therefore, $\lambda = 0$ is an eigenvalue of infinite multiplicity.

Note that if λ is not an eigenvalue of T, then the operator $T - \lambda I$ is invertible, and conversely. If space X is finite dimensional and λ is not an eigenvalue of T, then the operator $\big(T - \lambda I\big)^{-1}$ is bounded because all operators on a finite-dimensional space are bounded. The situation for infinite dimensional spaces is more complicated.

Definitions 2.16.3 (Resolvent, Spectrum). Let T be an operator on a normed space X. The operator

$$T_\lambda = \big(T - \lambda I\big)^{-1}$$

is called the *resolvent* of T. The values λ for which T_λ is defined on the whole space X and is bounded are called *regular points* of T. The set of all λ's which are not regular is called the *spectrum* of T.

Every eigenvalue belongs to the spectrum. The following example shows that the spectrum may contain points that are not eigenvalues. In fact, a non empty spectrum may contain no eigenvalues at all.

Example 2.16.3. Let X be the space $C\big([a,b]\big)$ of continuous functions on the interval $[a,b]$. For a fixed $u \in C\big([a,b]\big)$, consider the operator T defined by

$$\big(Tx\big)(t) = u(t)\,x(t).$$

Since

$$\big(T - \lambda I\big)^{-1} x(t) = \frac{x(t)}{u(t) - \lambda},$$

the spectrum of T consists of all λ's such that $\lambda - u(t) = 0$ for some $t \in [a,b]$. This means that the spectrum of T is exactly the range of u. If $u(t) = c$ is a constant function, then $\lambda = c$ is an eigenvalue of T. On the other hand, if u is a strictly increasing function, then T has no eigenvalues. The spectrum of T in such a case is the interval $\big[u(a), u(b)\big]$.

The problem of finding eigenvalues and eigenvectors is called the eigenvalue problem. One of the main sources of eigenvalue problems in mechanics is the theory of oscillating systems. The state of a given system at a given time t may be represented by an element $u(t) \in H$, where H is an appropriate Hilbert space of functions. The equation of motion in classical mechanics is

$$\frac{d^2u}{dt^2} = Tu, \tag{2.16.6}$$

where T is an operator in H. If the system oscillates, the time dependence of u is sinusoidal, so that $u(t) = v \sin \omega t$, where v is a fixed element of H. If T is linear, then (2.16.6) becomes

$$Tv = \big(-\omega^2\big)\,v. \tag{2.16.7}$$

This means that $-\omega^2$ is an eigenvalue of T. Physically, the eigenvalues of T correspond to possible frequencies of oscillations. In atomic systems, the frequencies of oscillations are visible as bright lines in the spectrum of light they emit. Thus, the name spectrum arises from physical considerations.

The following theorems describe properties of eigenvalues and eigenvectors for some special classes of operators. Our main interest is in self-adjoint, unitary, and compact operators.

Theorem 2.16.2. *Let T be an invertible operator on a vector space X, and let A be an operator on X. The operators A and TAT^{-1} have the same eigenvalues.*

Proof. Let λ be an eigenvalue of A. This means that there exists a nonzero vector u such that $Au = \lambda u$. Since T is invertible, $Tu \neq 0$ and

$$TAT^{-1}(Tu) = TAu = T(\lambda u) = \lambda Tu.$$

Thus, λ is an eigenvalue of TAT^{-1}.

Assume now that λ is an eigenvalue of TAT^{-1}, that is, $TAT^{-1}u = \lambda u$ for some nonzero vector $u = Tv$. Since $AT^{-1}u = \lambda T^{-1}u$ and $T^{-1}u \neq 0$, hence, λ is an eigenvalue of A.

Theorem 2.16.3. *All eigenvalues of a self-adjoint operator on a Hilbert space are real.*

Proof. Let λ be an eigenvalue of a self-adjoint operator T, and let u be a nonzero eigenvector of λ. Then,

$$\lambda\langle u, u\rangle = \langle \lambda u, u\rangle = \langle Tu, u\rangle = \langle u, Tu\rangle = \langle u, \lambda u\rangle = \bar{\lambda}\langle u, u\rangle.$$

Since $\langle u, u\rangle > 0$, we conclude $\lambda = \bar{\lambda}$.

Theorem 2.16.4. *All eigenvalues of a positive operator are nonnegative. All eigenvalues of a strictly positive operator are positive.*

Proof. Let T be a positive operator, and let $Tx = \lambda x$ for some $x \neq 0$. Since T is self-adjoint, we have

$$0 \leq \langle Tx, x\rangle = \lambda \langle x, x\rangle = \lambda \|x\|^2. \tag{2.16.8}$$

Thus, $\lambda \geq 0$. The proof of the second part of the theorem is obtained by replacing $\leq$ by $<$ in (2.16.8).

Theorem 2.16.5. *All eigenvalues of a unitary operator on a Hilbert space are complex numbers of modulus* 1.

Proof. Let λ be an eigenvalue of a unitary operator T, and let u be an eigenvector of λ, $u \neq 0$. Then,

$$\langle Tu, Tu\rangle = \langle \lambda u, \lambda u\rangle = |\lambda|^2 \|u\|^2.$$

On the other hand,

$$\langle Tu, Tu\rangle = \langle u, T^* Tu\rangle = \langle u, u\rangle = \|u\|^2.$$

Thus, $|\lambda| = 1$.

Theorem 2.16.6. *Eigenvectors corresponding to distinct eigenvalues of a self-adjoint or unitary operator on a Hilbert space are orthogonal.*

Proof. Let T be a self-adjoint operator, and let u_1 and u_2 be eigenvectors corresponding to distinct eigenvalues λ_1 and λ_2, that is, $Tu_1 = \lambda_1 u_1$ and $Tu_2 = \lambda_2 u_2$, $\lambda_1 \neq \lambda_2$. By Theorem 2.16.3. λ_1 and λ_2 are real. Then

$$\lambda_1\langle u_1, u_2\rangle = \langle Tu_1, u_2\rangle = \langle u_1, Tu_2\rangle = \langle u_1, \lambda_2 u_2\rangle = \bar{\lambda}_2\langle u_1, u_2\rangle = \lambda_2\langle u_1, u_2\rangle,$$

and hence,

$$(\lambda_1 - \lambda_2)\langle u_1, u_2\rangle = 0.$$

Since $\lambda_1 \neq \lambda_2$, we have $\langle u_1, u_2\rangle = 0$, that is, u_1 and u_2 are orthogonal.

Suppose now that T is a unitary operator on a Hilbert space H. Then, $TT^* = T^*T = I$ and $\|Tu\| = \|u\|$ for all $u \in H$. First, note that $\lambda_1 \neq \lambda_2$ implies $\lambda_1\bar{\lambda}_2 \neq 1$. Indeed, if $\lambda_1\bar{\lambda}_2 = 1$, then

$$\lambda_2 = \lambda_1\bar{\lambda}_2\lambda_2 = \lambda_1\left|\lambda_2\right|^2 = \lambda_1,$$

because $\left|\lambda_2\right| = 1$ by Theorem 2.16.5. Now,

$$\lambda_1\bar{\lambda}_2\langle u_1, u_2\rangle = \langle \lambda_1 u_1, \lambda_2 u_2\rangle = \langle Tu_1, Tu_2\rangle = \langle u_1, T^*Tu_2\rangle = \langle u_1, u_2\rangle.$$

Since $\lambda_1\bar{\lambda}_2 \neq 1$, we get $\langle u_1, u_2\rangle = 0$. This proves that the eigenvectors u_1 and u_2 are orthogonal.

Theorem 2.16.7. *For every eigenvalue λ of a bounded operator T, we have $|\lambda| \leq \|T\|$.*

Proof. Let u be a nonzero eigenvector corresponding to λ. Since $Tu = \lambda u$, we have

$$\|\lambda u\| = \|Tu\|,$$

and thus,

$$|\lambda|\,\|u\| = \|Tu\| \leq \|T\|\,\|u\|.$$

This implies that $\left|\lambda\right| \leq \|T\|$.

If the eigenvalues are considered as points in the complex plane, the preceding result implies that all the eigenvalues of a bounded operator T lie inside the circle of radius $\|T\|$.

Corollary 2.16.1. *All eigenvalues of a bounded, self-adjoint operator T satisfy the inequality*

$$\left|\lambda\right| \leq \sup_{\|x\|\leq 1}\left|\langle Tx, x\rangle\right|. \tag{2.16.9}$$

The proof follows immediately from Theorem 2.16.5, proved by Debnath and Mikusinski (1999).

Theorem 2.16.8. *If T is a nonzero, compact, self-adjoint operator on a Hilbert space H, then it has an eigenvalue λ equal to either $\|T\|$ or $-\|T\|$.*

Proof. Let $\{u_n\}$ be a sequence of elements of H such that $\|u_n\| = 1$, for all $n \in \mathbb{N}$, and

$$\|Tu_n\| \to \|T\| \quad \text{as } n \to \infty. \tag{2.16.10}$$

Then

$$\begin{aligned}
\left\|T^2u_n - \|Tu_n\|^2 u_n\right\|^2 &= \left\langle T^2u_n - \|Tu_n\|^2 u_n, T^2u_n - \|Tu_n\|^2 u_n\right\rangle \\
&= \|T^2u_n\|^2 - 2\|Tu_n\|^2 \langle T^2u_n, u_n\rangle + \|Tu_n\|^4 \|u_n\|^2 \\
&= \|T^2u_n\|^2 - \|Tu_n\|^4 \\
&\le \|T\|^2 \|Tu_n\|^2 - \|Tu_n\|^4 \\
&= \|Tu_n\|^2 \left(\|T\|^2 - \|Tu_n\|^2\right).
\end{aligned}$$

Since $\|Tu_n\|$ converges to $\|T\|$, we obtain

$$\left\|T^2u_n - \|Tu_n\|^2 u_n\right\| \to 0 \qquad \text{as } n \to \infty. \tag{2.16.11}$$

The operator T^2, being the product of two compact operators, is also compact. Hence, there exists a subsequence $\{u_{p_n}\}$ of $\{u_n\}$ such that $\{T^2u_{p_n}\}$ converges. Since $\|T\| \neq 0$, the limit can be written in the form $\|T\|v$, $v \neq 0$. Then, for every $n \in \mathbb{N}$, we have

$$\begin{aligned}
\left\|\|T\|^2 v - \|T\|^2 u_{p_n}\right\| \le \left\|\|T\|^2 v - T^2u_{p_n}\right\| &+ \left\|T^2u_{p_n} - \|Tu_{p_n}\|^2 u_{p_n}\right\| \\
&+ \left\|\|Tu_{p_n}\|^2 u_{p_n} - \|T\|^2 u_{p_n}\right\|.
\end{aligned}$$

Thus, by (2.16.10) and (2.16.11), we have

$$\left\|\|T\|^2 v - \|T\|^2 u_{p_n}\right\| \to 0 \qquad \text{as } n \to \infty.$$

Or,

$$\left\|\|T\|^2 (v - u_{p_n})\right\| \to 0 \qquad \text{as } n \to \infty.$$

This means that the sequence $\{u_{p_n}\}$ converges to v and therefore

$$T^2 v = \|T\|^2 v.$$

The above equation can be written as

$$\big(T - \|T\| I\big)\,\big(T + \|T\| I\big)\, v = 0.$$

If $w = \big(T + \|T\| I\big)\, v \neq 0$, then $\big(T - \|T\| I\big)\, w = 0$, and thus $\|T\|$ is an eigenvalue of T. On the other hand, if $w = 0$, then $-\|T\|$ is an eigenvalue of T.

Corollary 2.16.2. *If T is a nonzero compact, self-adjoint operator on a Hilbert space H, then there is a vector w such that $\|w\| = 1$ and*

$$\big|\langle Tw, w\rangle\big| = \sup_{\|x\|\le 1} \big|\langle Tx, x\rangle\big|.$$

Proof. Let w, $\|w\| = 1$, be an eigenvector corresponding to an eigenvalue λ such that $|\lambda| = \|T\|$. Then

$$\big|\langle Tw, w\rangle\big| = \big|\langle \lambda w, w\rangle\big| = \big|\lambda\big|\,\big\|w\big\|^2 = \big|\lambda\big| = \big\|T\big\| = \sup_{\|x\|\le 1} \big|\langle Tx, x\rangle\big|$$

by Theorem 4.4.5, proved by Debnath and Mikusinski (1999).

Theorem 2.16.8 guarantees the existence of at least one nonzero eigenvalue but no more in general. The corollary gives a useful method for finding that eigenvalue by maximizing certain quadratic expressions.

Theorem 2.16.9. *The set of distinct nonzero eigenvalues $\{\lambda_n\}$ of a self-adjoint compact operator is either finite or* $\lim_{n\to\infty} \lambda_n = 0$.

Proof. Suppose T is a self-adjoint, compact operator that has infinitely many distinct eigenvalues $\lambda_n, n \in \mathbb{N}$. Let u_n be an eigenvector corresponding to λ_n such that $\|u_n\| = 1$. By Theorem 2.16.6, $\{u_n\}$ is an orthonormal sequence. Since orthonormal sequences are weakly convergent to 0, Theorem 2.15.14 implies

$$\begin{aligned} 0 &= \lim_{n\to\infty} \big\|Tu_n\big\|^2 = \lim_{n\to\infty} \big\langle Tu_n, Tu_n\big\rangle \\ &= \lim_{n\to\infty} \big\langle \lambda_n u_n, \lambda_n u_n\big\rangle = \lim_{n\to\infty} \lambda_n^2 \big\|u_n\big\|^2 = \lim_{n\to\infty} \lambda_n^2. \end{aligned}$$

Example 2.16.4. We determine the eigenvalues and eigenfunctions of the operator T on $L^2\big([0, 2\pi]\big)$ defined by

$$\big(Tu\big)(x) = \int_0^{2\pi} k(x - t)\, u(t)\, dt,$$

where k is a periodic function with period 2π and square integrable on $[0, 2\pi]$.

As a trial solution, we take

$$u_n(x) = e^{inx}$$

and note that

$$\big(Tu_n\big)(x) = \int_0^{2\pi} k(x-t)\, e^{int} u(t)\, dt = e^{inx} \int_{x-2\pi}^{x} k(s)\, e^{ins}\, ds.$$

Thus,

$$Tu_n = \lambda_n u_n, \quad n \in \mathbb{Z},$$

where

$$\lambda_n = \int_0^{2\pi} k(s)\, e^{ins}\, ds.$$

The set of functions $\{u_n, n \in \mathbb{Z}\}$ is a complete orthogonal system in $L^2\big([0, 2\pi]\big)$. Note that T is self-adjoint if $k(x) = k(-x)$ for all x, but the sequence of eigenfunctions is complete even if T is not self-adjoint.

Theorem 2.16.10. *Let $\{P_n\}$ be a sequence of pairwise orthogonal projection operators on a Hilbert space H, and let $\{\lambda_n\}$ be a sequence of numbers such that $\lambda_n \to 0$ as $n \to \infty$. Then,*

(a) $\sum_{n=1}^{\infty} \lambda_n P_n$ converges in $B(H, H)$ and thus, defines a bounded operator;
(b) For each $n \in \mathbb{N}$, λ_n is an eigenvalue of the operator $T = \sum_{n=1}^{\infty} \lambda_n P_n$, and the only other possible eigenvalue of T is 0.
(c) If all λ_n's are real, then T is self-adjoint.
(d) If all projections P_n are finite-dimensional, then T is compact.

For a proof of this theorem, the reader is referred to Debnath and Mikusinski (1999).

Definition 2.16.3 (Approximate Eigenvalue). Let T be an operator on a Hilbert space H. A scalar λ is called an *approximate eigenvalue* of T if there exists a sequence of vectors $\{x_n\}$ such that $\|x_n\| = 1$ for all $n \in \mathbb{N}$ and $\|Tx_n - \lambda x_n\| \to 0$ as $n \to \infty$.

Obviously, every eigenvalue is an approximate eigenvalue.

Example 2.16.5. Let $\{e_n\}$ be a complete orthonormal sequence in a Hilbert space H. Let λ_n be a strictly decreasing sequence of scalars convergent to some λ. Define an operator T on H by

$$Tx = \sum_{n=1}^{\infty} \lambda_n \langle x, e_n \rangle e_n.$$

It is easy to see that every λ_n is an eigenvalue of T, but λ is not. On the other hand,

$$\|Te_n - \lambda e_n\| = \|\lambda_n e_n - \lambda e_n\| = \|(\lambda_n - \lambda)e_n\| = |\lambda_n - \lambda| \to 0 \quad \text{as } n \to \infty.$$

Thus, λ is an approximate eigenvalue of T. Note that the same is true if we just assume that $\lambda_n \to \lambda$ and $\lambda_n \neq \lambda$ for all $n \in \mathbb{N}$.

For further properties of approximate eigenvalues, see the exercises at the end of this chapter.

The rest of this section is concerned with several theorems involving spectral decomposition.

Let H be a finite-dimensional Hilbert space, say $H = \mathbb{C}^N$. It is known from linear algebra that eigenvectors of a self-adjoint operator on H form an orthogonal basis of H. The following theorems generalize this result to infinite-dimensional spaces.

Theorem 2.16.11 (Hilbert–Schmidt Theorem). *For every self-adjoint, compact operator T on an infinite-dimensional Hilbert space H, there exists an orthonormal system of eigenvectors $\{u_n\}$ corresponding to nonzero eigenvalues $\{\lambda_n\}$ such that every element $x \in H$ has a unique representation in the form*

$$x = \sum_{n=1}^{\infty} \alpha_n u_n + v, \tag{2.16.12}$$

where an $\alpha_n \in \mathbb{C}$ and v satisfies the equation $Tv = 0$. If T has infinitely many distinct eigenvalues $\lambda_1, \lambda_2, \ldots$, then $\lambda_n \to 0$ as $n \to \infty$.

For a proof of this theorem, the reader is referred to Debnath and Mikusinski (1999).

Theorem 2.16.12 (Spectral Theorem for Self-adjoint, Compact Operators). *Let T be a self-adjoint, compact operator on an infinite-dimensional Hilbert space H. Then, there exists in H a complete orthonormal system (an orthonormal basis) $\{v_1, v_2, \ldots\}$ consisting of eigenvectors of T. Moreover, for every $x \in H$,*

$$Tx = \sum_{n=1}^{\infty} \lambda_n \langle x, v_n \rangle v_n, \tag{2.16.13}$$

where λ_n is the eigenvalue corresponding to v_n.

Proof. Most of this theorem is already contained in Theorem 2.16.11. To obtain a complete orthonormal system $\{v_1, v_2, \ldots\}$, we must add an arbitrary orthonormal basis of $S^\perp$ to the system $\{u_1, u_2, \ldots\}$ (defined in the proof of Theorem 2.16.11). All of the eigenvalues corresponding to those vectors from $S^\perp$ are all equal to zero. Equality (2.16.13) follows from the continuity of T.

Theorem 2.16.13. *For any two commuting, self-adjoint, compact operators A and B on a Hilbert space H, there exists a complete orthonormal system of common eigenvectors.*

Proof. Let λ be an eigenvalue of A, and let X be the corresponding eigenspace. For any $x \in X$, we have

$$ABx = BAx = B(\lambda x) = \lambda Bx.$$

This means that Bx is an eigenvector of A corresponding to λ, provided $Bx \neq 0$. In any case, $Bx \in X$ and hence B maps X into itself. Since B is a self-adjoint, compact operator, by Theorem 2.16.12, X has an orthonormal basis consisting of eigenvalues of B, but these vectors are also eigenvectors of A because they belong to X. If we repeat the same procedure with every eigenspace of A, then the union of all of these eigenvectors will be an orthonormal basis of H.

Theorem 2.16.14. *Let T be a self-adjoint, compact operator on a Hilbert space H with a complete orthonormal system of eigenvectors $\{v_1, v_2, \ldots\}$ corresponding to eigenvalues $\{\lambda_1, \lambda_2, \ldots\}$. Let P_n be the projection operator onto the one-dimensional space spanned by v_n. Then, for all $x \in H$,*

$$x = \sum_{n=1}^{\infty} P_n x, \tag{2.16.14}$$

and

$$T = \sum_{n=1}^{\infty} \lambda_n P_n. \tag{2.16.15}$$

Proof. From the spectral theorem 2.16.12, we have

$$x = \sum_{n=1}^{\infty} \langle x, v_n \rangle v_n. \tag{2.16.16}$$

For every $k \in \mathbb{N}$, the projection operator P_k onto the one-dimensional subspace S_k spanned by v_k is given by

$$P_k x = \langle x, v_k \rangle v_k.$$

Now, (2.16.16) can be written as

$$x = \sum_{n=1}^{\infty} P_n x,$$

and thus, by Theorem 2.16.2,

$$Tx = \sum_{n=1}^{\infty} \lambda_n \langle x, v_n \rangle v_n = \sum_{n=1}^{\infty} \lambda_n P_n x.$$

Hence, for all $x \in H$,

$$Tx = \left(\sum_{n=1}^{\infty} \lambda_n P_n \right) x.$$

This proves (2.16.15) since convergence of $\sum_{n=1}^{\infty} \lambda_n P_n$ is guaranteed by Theorem 2.16.10.

Theorem 2.16.15 is another version of the spectral theorem. This version is important in the sense that it can be extended to noncompact operators. It is also useful because it leads to an elegant expression for powers and more general functions of an operator.

Theorem 2.16.15. *If eigenvectors $u_1, u_2, \ldots$ of a self-adjoint operator T on a Hilbert space H form a complete orthonormal system in H and all eigenvalues are positive (or nonnegative), then T is strictly positive (or positive).*

Proof. Suppose $u_1, u_2, \ldots$ is a complete orthonormal system of eigenvalues of T corresponding to real eigenvalues $\lambda_1, \lambda_2, \ldots$. Then, any nonzero vector $u \in H$ can be represented as $u = \sum_{n=1}^{\infty} \alpha_n u_n$, and we have

$$\begin{aligned}
\langle Tu, u \rangle &= \left\langle Tu, \sum_{n=1}^{\infty} \alpha_n u_n \right\rangle = \sum_{n=1}^{\infty} \overline{\alpha_n} \langle Tu, u_n \rangle = \sum_{n=1}^{\infty} \overline{\alpha_n} \langle u, Tu_n \rangle \\
&= \sum_{n=1}^{\infty} \overline{\alpha_n} \langle u, \lambda_n u_n \rangle = \sum_{n=1}^{\infty} \lambda_n \overline{\alpha_n} \langle u, u_n \rangle = \sum_{n=1}^{\infty} \lambda_n \overline{\alpha_n} \alpha_n \\
&= \sum_{n=1}^{\infty} \lambda_n |\alpha_n|^2 \geq 0,
\end{aligned}$$

if all eigenvalues are nonnegative. If all λ_n's are positive, then the last inequality becomes strict.

2.17 Exercises

1. Show that on any inner product space X

(a) $\langle x, \alpha y + \beta z \rangle = \bar{\alpha} \langle x, y \rangle + \bar{\beta} \langle x, z \rangle$ for all $\alpha, \beta \in \mathbb{C}$,
(b) $2[\langle x, y \rangle + \langle y, x \rangle] = \|x + y\|^2 - \|x - y\|^2$.

2. Prove that the space $C_0(\mathbb{R})$ of all complex-valued continuous functions that vanish outside some finite interval is an inner product space with the inner product

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f(x)\,\overline{g(x)}\,dx.$$

3. (a) Show that the space $C^1([a,b])$ of all continuously differentiable complex-valued functions on $[a,b]$ is not an inner product space with the inner product

$$\langle f, g\rangle = \int_a^b f'(x)\,\overline{g'(x)}\,dx.$$

(b) If $f \in C^1([a,b])$ with $f(a) = 0$, show that $C^1([a,b])$ is an inner product space with the inner product defined in (a).

4. (a) Show that the space $C([a,b])$ of real or complex-valued functions is a normed space with the norm $\|f\| = \max_{a\le x\le b} |f(x)|$.

(b) Show that the space $C([a,b])$ is a complete metric space with the metric induced by the norm in (a), that is,

$$d(f,g) = \|f - g\| = \max_{a\le x\le b} |f(x) - g(x)|.$$

5. Prove that the space $C_0^1(\mathbb{R})$ of all continuously differentiable complex-valued continuous functions that vanish outside some finite interval is an inner product space with the inner product

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f'(x)\,\overline{g'(x)}\,dx.$$

6. Prove that the norm in an inner product space is strictly convex, that is, if $x \neq y$ and $\|x\| = \|y\| = 1$, then $\|x + y\| \le 2$.

7. (a) Show that the space $C([-\pi,\pi])$ of continuous functions with the norm defined by (2.2.4) is an incomplete normed space.

(b) In the Banach space $L^2([-\pi,\pi])$,

$$f(x) = \sum_{n=1}^{\infty} \frac{1}{n} \sin nx,$$

where $f(x) = -\frac{\pi}{4}$ in $(-\pi, 0)$ and $f(x) = \frac{\pi}{4}$ in $(0,\pi)$. Show that f is not continuous in $C([-\pi,\pi])$, but the series converges in $L^2([-\pi,\pi])$.

8. Show that, in any inner product space X,

$$\|x-y\|+\|y-z\|=\|x-z\|$$

if and only if $y=\alpha x+(1-\alpha)z$ for some α in $0\le\alpha\le 1$.

9. (a) Prove that the *polarization identity*

$$\langle x,y\rangle=\frac{1}{4}\left(\|x+y\|^2-\|x-y\|^2+i\,\|x+iy\|^2-i\,\|x-iy\|^2\right)$$

holds in any complex inner product space.

(b) In any real inner product space, show that

$$\langle x,y\rangle=\frac{1}{4}\left(\|x+y\|^2-\|x-y\|^2\right).$$

10. Prove that, for any x in a Hilbert space, $\|x\|=\sup\limits_{\|y\|=1}|\langle x,y\rangle|$

11. Show that $L^2([a,b])$ is the only inner product space among the spaces $L^p([a,b])$.

12. Show that the *Apollonius identity* in an inner product space is

$$\|z-x\|^2+\|z-y\|^2=\frac{1}{2}\|x-y\|^2+2\left\|z-\frac{x+y}{2}\right\|^2.$$

13. Prove that any finite-dimensional inner product space is a Hilbert space.

14. Let $X=\{f\in C^1([a,b]) : f(a)=0\}$ and

$$\langle f,g\rangle=\int_a^b f'(x)\,\overline{g'(x)}\,dx.$$

Is X a Hilbert space?

15. Is the space $C_0^1(\mathbb{R})$ with the inner product

$$\langle f,g\rangle=\int_{-\infty}^{\infty} f'(x)\,\overline{g'(x)}\,dx$$

a Hilbert space?

16. Let X be an incomplete inner product space. Let H be the completion of X. Is it possible to extend the inner product from X onto H such that H would become a Hilbert space?

17. Suppose $x_n \to x$ and $y_n \to y$ as $n \to \infty$ in a Hilbert space, and $\alpha_n \to \alpha$ in $\mathbb{C}$. Prove that

(a) $x_n + y_n \to x + y$,
(b) $\alpha_n x_n \to \alpha x$,
(c) $\langle x_n, y_n \rangle \to \langle x, y \rangle$,
(d) $\|x_n\| \to \|x\|$.

18. Suppose $x_n \xrightarrow{w} x$ and $y_n \xrightarrow{w} y$ as $n \to \infty$ in a Hilbert space, and $\alpha_n \to \alpha$ in $\mathbb{C}$. Prove or give a counter example:

(a) $x_n + y_n \xrightarrow{w} x + y$,
(b) $\alpha_n x_n \xrightarrow{w} \alpha x$,
(c) $\langle x_n, y_n \rangle \to \langle x, y \rangle$,
(d) $\|x_n\| \to \|x\|$,
(e) If $x_n = y_n$ for all $n \in \mathbb{N}$, then $x = y$.

19. Show that, in a finite-dimensional Hilbert space, weak convergence implies strong convergence.

20. Is it always possible to find a norm on an inner product space X which would define the weak convergence in X?

21. If $\sum_{n=1}^{\infty} u_n = u$, show that

$$\sum_{n=1}^{\infty} \langle u_n, x \rangle = \langle u, x \rangle$$

for any x in an inner product space X.

22. Let $\{x_1, \dots, x_n\}$ be a finite orthonormal set in a Hilbert space H. Prove that for any $x \in H$ the vector

$$x - \sum_{k=1}^{n} \langle x, x_k \rangle x_k$$

is orthogonal to x_k for every $k = 1, \dots, n$.

23. In the pre-Hilbert space $\mathscr{C}([-\pi, \pi])$, show that the following sequences of functions are orthogonal

(a) $x_k(t) = \sin kt, \quad k = 1, 2, 3, \dots,$
(b) $y_n(t) = \cos nt, \quad n = 0, 1, 2, \dots.$

24. Show that the application of the Gram–Schmidt process to the sequence of functions

$$f_0(t) = 1,\ f_1(t) = t,\ f_2(t) = t^2, \ldots, f_n(t) = t^n, \ldots$$

(as elements of $L^2\big([-1,1]\big)$) yields the Legendre polynomials.

25. Show that the application of the Gram–Schmidt process to the sequence of functions

$$f_0(t) = e^{-t^2/2},\ f_1(t) = te^{-t^2/2},\ f_2(t) = t^2e^{-t^2/2}, \ldots, f_n(t) = t^n e^{-t^2/2}, \ldots$$

(as elements of $L^2(\mathbb{R})$) yields the orthonormal system discussed in Example 2.9.4.

26. Apply the Gram–Schmidt process to the sequence of functions

$$f_0(t) = 1,\ f_1(t) = t,\ f_2(t) = t^2, \ldots, f_n(t) = t^n, \ldots$$

defined on $\mathbb{R}$ with the inner product

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f(t)\,\overline{g(t)}\, \exp(-t^2)\, dt.$$

Compare the result with Example 2.9.4.

27. Apply the Gram–Schmidt process to the sequence of functions

$$f_0(t) = 1,\ f_1(t) = t,\ f_2(t) = t^2, \ldots, f_n(t) = t^n, \ldots$$

defined on $[0, \infty)$ with the inner product

$$\langle f, g\rangle = \int_0^{\infty} f(t)\,\overline{g(t)}\, e^{-t}\, dt.$$

The resulting polynomials are called the *Laguerre polynomials*.

28. Let T_n be the *Chebyshev polynomial* of degree n, that is,

$$T_0(x) = 1, \quad T_n(x) = 2^{1-n} \cos(n \arccos x).$$

Show that the functions

$$\phi_n(x) = \frac{2^n}{\sqrt{2\pi}} T_n(x), \qquad n = 0, 1, 2, \ldots,$$

form an orthonormal system in $L^2[(-1,1)]$ with respect to the inner product

$$\langle f, g\rangle = \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}\, f(x)\,\overline{g(x)}\, dx.$$

29. Prove that for any polynomial

$$p_n(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0,$$

we have

$$\max_{[-1,1]} \left|p_n(x)\right| \geq \max_{[-1,1]} \left|T_n(x)\right|,$$

where T_n denotes the Chebyshev polynomial of degree n.

30. Show that the complex functions

$$\phi_n(z) = \sqrt{\frac{n}{\pi}}\, z^{n-1}, \quad n = 1, 2, 3, \ldots,$$

form an orthonormal system in the space of continuous complex functions defined in the unit disk $D = \{z \in \mathbb{C} : \|z\| \leq 1\}$ with respect to the inner product

$$\langle f, g \rangle = \int_D f(z)\, \overline{g(z)}\, dz.$$

31. Prove that the complex functions

$$\psi_n(z) = \frac{1}{\sqrt{2\pi}}\, z^{n-1}, \quad n = 1, 2, 3, \ldots$$

form an orthonormal system in the space of continuous complex functions defined on the unit circle $C = \{z \in \mathbb{C} : \|z\| = 1\}$ with respect to the inner product

$$\langle f, g \rangle = \int_C f(z)\, \overline{g(z)}\, dz.$$

32. With respect to the inner product

$$\langle f, g \rangle = \int_{-1}^{1} f(x)\, \overline{g(x)}\, \omega(x) dx,$$

where $\omega(x) = (1-x)^\alpha (1+x)^\beta$ and $\alpha, \beta > -1$, show that the *Jacobi polynomials*

$$P_n^{(\alpha\beta)}(x) = \frac{(-1)^n}{n! 2^n} (1-x)^{-\alpha} (1+x)^{-\beta} \frac{d^n}{dx^n} \left[(1-x)^\alpha (1+x)^\beta (1-x^2)^n\right]$$

form an orthogonal system.

33. Show that the *Gegenbauer polynomials*

$$C_n^\gamma(x) = \frac{(-1)^n}{n! 2^n} (1-x^2)^{\frac{1}{2}-\gamma} \frac{d^n}{dx^n} (1-x^2)^{n+\gamma-\frac{1}{2}}$$

where $\gamma > \frac{1}{2}$ form an orthonormal system with respect to the inner product

$$\langle f, g \rangle = \int_{-1}^{1} f(x)\,\overline{g(x)}\,\left(1 - x^2\right)^{\frac{1}{2}-\gamma} dx.$$

Note that Gegenbauer polynomials are a special case of Jacobi polynomials if $\alpha = \beta = \gamma - \frac{1}{2}$.

34. If x and x_k $(k = 1, \ldots, n)$ belong to a real Hilbert space, show that

$$\left\| x - \sum_{k=1}^{n} a_k x_k \right\|^2 = \|x\|^2 - \sum_{k=1}^{n} a_k \langle x, x_k \rangle + \sum_{k=1}^{n} \sum_{l=1}^{n} a_k a_l \langle x_k, x_l \rangle.$$

Also show that this expression is minimum when $Aa = b$ where $a = (a_1, \ldots, a_n), b = (\langle x, x_1 \rangle, \ldots, \langle x, x_n \rangle)$, and the matrix $A = (a_{kl})$ is defined by $a_{kl} = \langle x_k, x_l \rangle$.

35. If $\{a_n\}$ is an orthonormal sequence in a Hilbert space H and $\{\alpha_n\}$ is a sequence in the space l^2, show that there exists $x \in H$ such that

$$\langle x, a_n \rangle = \alpha_n \quad \text{and} \quad \| \{\alpha_n\} \| = \|x\|,$$

where $\| \{\alpha_n\} \|$ denotes the norm in the sequence space l^2.

36. If α_n and β_n $(n = 1, 2, 3, \ldots)$ are generalized Fourier coefficients of vectors x and y with respect to a complete orthonormal sequence in a Hilbert space, show that

$$\langle x, y \rangle = \sum_{k=1}^{\infty} \alpha_k \overline{\beta_k}.$$

37. If $\{x_n\}$ is an orthonormal sequence in a Hilbert space H such that the only element orthogonal to all the x_n's is the null element, show that the sequence $\{x_n\}$ is complete.

38. Let $\{x_n\}$ be an orthonormal sequence in a Hilbert space H. Show that $\{x_n\}$ is complete if and only if $\mathrm{cl}(\mathrm{span}\{x_1, x_2, \ldots\}) = H$. In other words, $\{x_n\}$ is complete if and only if every element of H can be approximated by a sequence of finite combinations of x_n's.

39. Show that the sequence of functions

$$\phi_n(x) = \frac{e^{-x/2}}{n!} L_n(x), \qquad n = 0, 1, 2, \ldots,$$

where L_n is the Laguerre polynomial of degree n, that is,

$$L_n(x) = e^x \frac{d^n}{dx^n}\left(x^n e^{-x}\right),$$

form a complete orthonormal system in $L^2(0,\infty)$.

40. Let

$$\phi_n(x) = \frac{e^{inx}}{\sqrt{2\pi}}, \qquad n = 0, \pm 1, \pm 2, \ldots,$$

and let $f \in L^1\big([-\pi,\pi]\big)$. Define

$$f_n(x) = \sum_{k=-n}^{n} \langle f, \phi_k \rangle \phi_k, \qquad \text{for } n = 0, 1, 2, \ldots.$$

Show that

$$\frac{f_0(x) + f_1(x) + \cdots + f_n(x)}{n+1} = \sum_{k=-n}^{n} \left(1 - \frac{|k|}{n+1}\right) \langle f, \phi_k \rangle \phi_k(x).$$

41. Show that the sequence of functions

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \ldots$$

is a complete orthonormal sequence in $L^2\big([-\pi,\pi]\big)$.

42. Show that the following sequence of functions is a complete orthonormal system in $L^2([0,\pi])$:

$$\frac{1}{\sqrt{\pi}}, \sqrt{\frac{2}{\pi}} \cos x, \sqrt{\frac{2}{\pi}} \cos 2x, \sqrt{\frac{2}{\pi}} \cos 3x, \ldots.$$

43. Show that the following sequence of functions is a complete orthonormal system in $L^2\big([0,\pi]\big)$:

$$\sqrt{\frac{2}{\pi}} \sin x, \sqrt{\frac{2}{\pi}} \sin 2x, \sqrt{\frac{2}{\pi}} \sin 3x, \ldots.$$

44. Show that the sequence of functions defined by

$$f_n(x) = \frac{1}{\sqrt{2a}} \exp\left(\frac{in\pi x}{a}\right), \qquad n = 0, \pm 1, \pm 2, \ldots$$

is a complete orthonormal system in $L^2\big([-a,a]\big)$.

45. Show that the sequence of functions

$$\frac{1}{\sqrt{2a}}, \frac{1}{\sqrt{a}}\cos\left(\frac{n\pi x}{a}\right), \frac{1}{\sqrt{a}}\sin\left(\frac{n\pi x}{a}\right), \ldots$$

is a complete orthonormal system in $L^2([-a,a])$.

46. Show that each of the following sequences of functions is a complete orthonormal system in $L^2([0,a])$:

$$\frac{1}{\sqrt{a}}, \sqrt{\frac{2}{a}}\cos\left(\frac{\pi x}{a}\right), \sqrt{\frac{2}{a}}\cos\left(\frac{2\pi x}{a}\right), \ldots, \sqrt{\frac{2}{a}}\cos\left(\frac{n\pi x}{a}\right), \ldots$$

$$\sqrt{\frac{2}{a}}\sin\left(\frac{\pi x}{a}\right), \sqrt{\frac{2}{a}}\sin\left(\frac{2\pi x}{a}\right), \ldots, \sqrt{\frac{2}{a}}\sin\left(\frac{n\pi x}{a}\right), \ldots.$$

47. Let X be the Banach space $\mathbb{R}^2$ with the norm $\|\langle x, y\rangle\| = \max\{|x|, |y|\}$. Show that X does not have the closest-point property.

48. Let S be a closed subspace of a Hilbert space H and let $\{e_n\}$ be a complete orthonormal sequence in S. For an arbitrary $x \in H$, there exists $y \in S$ such that $\|x - y\| = \inf_{z\in S}\|x - z\|$. Define y in terms of $\{e_n\}$.

49. If S is a closed subspace of a Hilbert space H, then $H = S \oplus S^{\perp}$. Is this true in every inner product space?

50. Show that the functional in Example 2.13.2 is unbounded.

51. The Riesz representation theorem states that for every bounded linear functional $f \in H'$ on a Hilbert space H, there exists a representer $x_f \in H$ such that $f(x) = \langle x, x_f\rangle$ for all $x \in H$. Let $T : H' \to H$ be the mapping that assigns x_f to f. Prove the following properties of T:

(a) T is onto,
(b) $T(f+g) = T(f) + T(g)$,
(c) $T(\alpha f) = \bar{\alpha}\, T(f)$,
(d) $\|T(f)\| = \|f\|$,

where $f, g \in H'$ and $\alpha \in \mathbb{C}$.

52. Let f be a bounded linear functional on a closed subspace X of a Hilbert space H. Show that there exists a bounded linear functional g on H such that $\|f\| = \|g\|$ and $f(x) = g(x)$ whenever $x \in X$.

53. Show that the space l^2 is separable.

54. (a) Show that the sequence of Gaussian functions on $\mathbb{R}$ defined by

$$f_n(x) = \frac{n}{\sqrt{\pi}} \exp(-n^2x^2), \quad n = 1, 2, 3, \ldots$$

converges to the Dirac delta distribution $\delta(x)$.

(b) Show that the sequence of functions on $\mathbb{R}$ defined by

$$f_n(x) = \frac{\sin nx}{\pi x}, \quad n = 1, 2, \ldots$$

converges to the Dirac delta distribution.

55. Show that the sequence of functions on $\mathbb{R}$ defined by

$$f_n(x) = \begin{cases} 0, \text{ for } x < -\dfrac{1}{2n}, \\ n, \text{ for } -\dfrac{1}{2n} \le x \le \dfrac{1}{2n}, \\ 0, \text{ for } x > \dfrac{1}{2n} \end{cases}$$

converges to the Dirac delta distribution.

56. If f is a locally integrable function on $\mathbb{R}^N$, show that the functional F on $\mathscr{D}$ defined by

$$\langle F, \phi \rangle = \int_{\mathbb{R}^N} f\,\phi$$

is a distribution.

57. If $f_n(x) = \sin nx$, show that $f_n \to 0$ in the distributional sense.

58. Find the nth distributional derivative of $f(x) = |x|$.

59. Verify which functions belong to $L^1(\mathbb{R})$ and which do not belong to $L^1(\mathbb{R})$. Find their $L^1(\mathbb{R})$ norms when they exist.

(a) $f(x) = \left(a^2 + x^2\right)^{-1}$, (b) $f(x) = x\left(a^2 + x^2\right)^{-1}$,

(c) $f(x) = \left\{ \begin{matrix} 1, & |x| \le 1, \\ |x|^{-r}, & |x| > 1 \end{matrix} \right\}$, (d) $f(x) = x^{-1}$.

60. Let $\{e_n\}$ be a complete orthonormal sequence in a Hilbert space H, and let $\{\lambda_n\}$ be a sequence of scalars.

(a) Show that there exists a unique operator T on H such that $Te_n = \lambda_n e_n$.
(b) Show that T is bounded if and only if the sequence $\{\lambda_n\}$ is bounded.
(c) For a bounded sequence $\{\lambda_n\}$, find the norm of T.

61. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $T(x, y) = (x + 2y, 3x + 2y)$. Find the eigenvalues and eigenvectors of T.

62. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $T(x, y) = (x + 3y, 2x + y)$. Show that $T^* \neq T$.

63. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $T(x, y, z) = (3x - z, 2y, -x + 3z)$. Show that T is self-adjoint.

64. Compute the adjoint of each of the following operators:

(a) $A : \mathbb{R}^3 \to \mathbb{R}^3, \quad A(x, y, z) = (-y + z, -x + 2z, x + 2y)$,
(b) $B : \mathbb{R}^3 \to \mathbb{R}^3, \quad B(x, y, z) = (x + y - z, -x + 2y + 2z, x + 2y + 3z)$,
(c) $C : P_2(\mathbb{R}) \to P_2(\mathbb{R}), \quad C\{p(x)\} = x \dfrac{d}{dx} p(x) - \dfrac{d}{dx}\big(xp(x)\big)$,

where $P_2(\mathbb{R})$ is the space of all polynomials on $\mathbb{R}$ of degree less than or equal to 2.

65. If A is a self-adjoint operator and B is a bounded operator, show that $B^* AB$ is self-adjoint.

66. Prove that the representation $T = A + iB$ in Theorem 2.15.7 is unique.

67. If $A^* A + B^* B = 0$, show that $A = B = 0$.

68. If T is self-adjoint and $T \neq 0$, show that $T^* \neq 0$ for all $n \in \mathbb{N}$.

69. Let T be a self-adjoint operator. Show that

(a) $\|Tx + ix\|^2 = \|Tx\|^2 + \|x\|^2$,
(b) the operator $U = (T - iI)(T + iI)^{-1}$ is unitary. (U is called the *Cayley transform* of T.)

70. Show that the limit of a convergent sequence of self-adjoint operators is a self-adjoint operator.

71. If T is a bounded operator on H with one-dimensional range, show that there exists vectors $y, z \in H$ such that $Tx = \langle x, z\rangle y$ for all $x \in H$. Hence, show that

(a) $T^* x = \langle x, y\rangle z$ for all $x \in H$,
(b) $T^2 = \lambda T$, where λ is a scalar,
(c) $\|T\| = \|y\|\,\|z\|$,
(d) $T^* = T$ if and only if $y = \alpha z$ for some real scalar α.

72. Let T be a bounded self-adjoint operator on a Hilbert space H such that $\|T\| \leq 1$. Prove that $\langle x, Tx\rangle \geq (1 - \|T\|)\|x\|^2$ for all $x \in H$.

73. If A is a positive operator and B is a bounded operator, show that $B^* AB$ is positive.

74. If A and B are positive operators and $A + B = 0$, show that $A = B = 0$.

75. Show that, for any self-adjoint operator A, there exists positive operators S and T such that $A = S - T$ and $ST = 0$.

76. If P is self-adjoint and P^2 is a projection operator, is P a projection operator?

77. Let T be a multiplication operator on $L^2([a, b])$. Find necessary and sufficient conditions for T to be a projection.

78. Show that P is a projection if and only if $P = P^* P$.

79. If P, Q, and $P + Q$ are projections, show that $PQ = 0$.

80. Show that every projection P is positive and $0 \leq P \leq I$.

81. Show that, for projections P and Q, the operator $P + Q - PQ$ is a projection if and only if $PQ = QP$.

82. Show that the projection onto a closed subspace X of a Hilbert space H is a compact operator if and only if X is finite dimensional.

83. Show that the operator $T : l^2 \to l^2$ defined by $T(x_n) = (2^{-n} x_n)$ is compact.

84. Prove that the collection of all eigenvectors corresponding to one particular eigenvalue of an operator is a vector space.

85. Show that the space of all eigenvectors corresponding to one particular eigenvalue of a compact operator is finite dimensional.

86. Show that a self-adjoint operator T is compact if and only if there exists a sequence of finite-dimensional operators strongly convergent to T.

87. Show that eigenvalues of a symmetric operator are real and eigenvectors corresponding to different eigenvalues are orthogonal.

88. Give an example of a self-adjoint operator that has no eigenvalues.

89. Show that a nonzero vector x is an eigenvector of an operator T if and only if $|\langle Tx, x\rangle| = \|Tx\| \|x\|$.

90. Show that if the eigenvectors of a self-adjoint operator T form a complete orthogonal system and all eigenvalues are nonnegative (or positive), then T is positive (or strictly positive).

91. If λ is an approximate eigenvalue of an operator T, show that $|\lambda| \leq \|T\|$.

92. Show that if T has an approximate eigenvalue λ such that $|\lambda| = \|T\|$, then $\sup_{\|x\| \leq 1} |\langle Tx, x\rangle| = \|T\|$.

93. If λ is an approximate eigenvalue of T, show that $\lambda + \mu$ is an approximate eigenvalue of $T + \mu I$ and $\lambda\mu$ is an approximate eigenvalue of μT.

94. For every approximate eigenvalue λ of an isometric operator, show that we have $|\lambda| = 1$.

95. Show that every approximate eigenvalue of a self-adjoint operator is real.