Chapter 7
Function Spaces

In the preceding chapters we have seen that separation of variables can generate families of product solutions for certain PDE. For example, we found families of trigonometric solutions of the wave equation in Sect. 5.2 and the heat equation in Sect. 6.1. By the superposition principle, finite linear combinations of these functions give more general solutions.

It is natural to hope that we could push this construction farther and obtain solutions by infinite series. Solutions of PDE by trigonometric series were studied extensively in the 18th century by d'Alembert, Euler, Bernoulli, and others. However, notions of convergence were not well developed at that time, and many fundamental questions were left open.

In this chapter we will introduce some basic concepts of functional analysis, which will give us the tools to address some of these fundamental issues.

7.1 Inner Products and Norms

We assume that the reader has had a basic course in linear algebra and is familiar with the notion of a *vector space*, i.e., a set equipped with the operations of addition and scalar multiplication. The basic finite-dimensional example is the vector space $\mathbb{R}^n$. This space comes equipped with a natural inner product given by the dot product $\boldsymbol{v} \cdot \boldsymbol{w}$ for $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$. The Euclidean length of a vector $\boldsymbol{v} \in \mathbb{R}^n$ is $\|\boldsymbol{v}\| := \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}$.

In this section we will review the corresponding definitions for general real or complex vector spaces, which include function spaces. One important set of examples are the spaces $C^m(\Omega)$ introduced in Sect. 2.4, consisting of m-times continuously differentiable complex-valued functions on a domain $\Omega \subset \mathbb{R}^n$. Because differentiability and continuity of functions are preserved under linear combination and scalar multiplication, $C^m(\Omega)$ is naturally a complex vector space.

An *inner product* on a complex vector space V is a function of two variables,

$$u, v \in V \mapsto \langle u, v \rangle \in \mathbb{C},$$

D. Borthwick, *Introduction to Partial Differential Equations*,
Universitext, DOI 10.1007/978-3-319-48936-0_7

satisfying the following properties:

(I1) Positive definiteness: $\langle v, v\rangle \geq 0$ for $v \in V$, with equality only if $v = 0$.
(I2) Symmetry: for $v, w \in V$,

$$\langle v, w\rangle = \overline{\langle w, v\rangle}.$$

(I3) Linearity in the first variable: for $c_1, c_2 \in \mathbb{C}$ and $v_1, v_2, w \in V$,

$$\langle c_1 v_1 + c_2 v_2, w\rangle = c_1 \langle v_1, w\rangle + c_2 \langle v_2, w\rangle .$$

Together, (I2) and (I3) imply conjugate linearity in the second variable,

$$\langle w, c_1 v_1 + c_2 v_2\rangle = \overline{c_1} \langle w, v_1\rangle + \overline{c_2} \langle w, v_2\rangle .$$

The combination of linearity and conjugate linearity in the respective variables is called *sesquilinearity*. In the real case, the complex conjugation can be omitted, reducing sesquilinearity to *bilinearity*.

An *inner product space* is a real or complex vector space V equipped with an inner product $\langle\cdot,\cdot\rangle$. The *Euclidean* inner product on $\mathbb{C}^n$ is defined by including a conjugation in the dot product,

$$\langle \boldsymbol{v}, \boldsymbol{w}\rangle := \boldsymbol{v} \cdot \overline{\boldsymbol{w}}. \tag{7.1}$$

One way to define an inner product on function spaces is by integration. For example, on $C^0[0, 1]$ we could take

$$\langle f, g\rangle := \int_0^1 f\overline{g}\, dx.$$

Certain geometric notions are carried over from Euclidean geometry to inner product spaces. For example, vectors u, v in an inner product space V are called *orthogonal* if

$$\langle u, v\rangle = 0.$$

The analog of length for vectors in V is called a *norm*. A norm is a function $\|\cdot\| : V \to \mathbb{R}$ satisfying the following properties: for all $u, v \in V$ and scalar λ,

(N1) Positive definiteness: $\|u\| \geq 0$ with equality only if $u = 0$.
(N2) Homogeneity: $\|\lambda u\| = |\lambda|\, \|u\|$.
(N3) Triangle inequality: $\|u + v\| \leq \|u\| + \|v\|$.

For an inner product space, the definition of the Euclidean length in terms of the dot product suggests that the function

$$\|u\| := \sqrt{\langle u, u\rangle} \tag{7.2}$$

should yield a norm.

Positive definiteness of (7.2) clearly follows from positive definiteness of the inner product, and homogeneity follows from sesquilinearity. To see that (7.2) also satisfies the triangle inequality, we present a relation first derived in the Euclidean case by the great 19th century analyst Augustin-Louis Cauchy; Hermann Schwarz later generalized the result to inner product spaces.

Theorem 7.1 (Cauchy-Schwarz inequality) *For an inner product space V with $\|\cdot\|$ defined by* (7.2)*,*

$$|\langle v, w\rangle| \leq \|v\| \, \|w\|$$

for all $v, w \in V$.

Proof For $v, w \in V$ and $t \in \mathbb{R}$, consider the function

$$q(t) := \left\| v + t \langle v, w\rangle \, w \right\|^2,$$

The claimed inequality is trivial if $w = 0$, so assume $w \neq 0$. By (I2), (I3), and (7.2),

$$\begin{aligned} q(t) &= \Big\langle v + t \langle v, w\rangle \, w, v + t \langle v, w\rangle \, w \Big\rangle \\ &= \|v\|^2 + 2t \, |\langle v, w\rangle|^2 + t^2 \, |\langle v, w\rangle|^2 \, \|w\|^2 . \end{aligned}$$

The minimum of this quadratic polynomial occurs at $t_0 = -\|w\|^{-2}$. Since $q \geq 0$,

$$0 \leq q(t_0) = \|v\|^2 - \frac{|\langle v, w\rangle|^2}{\|w\|^2},$$

which gives the claimed inequality. □

The triangle inequality for (7.2) follows from the Cauchy-Schwarz inequality by

$$\begin{aligned} \|u + v\|^2 &= \langle u + v, u + v\rangle \\ &= \|u\|^2 + 2 \operatorname{Re} \langle u, v\rangle + \|v\|^2 \\ &\leq \|u\|^2 + 2 \, |\langle u, v\rangle| + \|v\|^2 \\ &\leq \|u\|^2 + 2 \, \|u\| \, \|v\| + \|v\|^2 \\ &= (\|u\| + \|v\|)^2 . \end{aligned}$$

Thus (7.2) defines a norm associated to the inner product. This definition of the norm is used by default on an inner product space.

It is possible to have a norm that is not associated to an inner product. For example, this is the case for the *sup norm*, defined for $f \in C^0(\overline{\Omega})$, with $\Omega \subset \mathbb{R}^n$ bounded, by

$$\sup_{\boldsymbol{x}\in\Omega} |f(\boldsymbol{x})| := \sup\left\{|f(\boldsymbol{x})|\,;\ \boldsymbol{x} \in \Omega\right\}. \tag{7.3}$$

We will explain how to tell that a norm does not come from an inner product in the exercises.

7.2 Lebesgue Integration

In the early 20th century, Henri Lebesgue developed an extension of the classic definition of the integral introduced by Bernhard Riemann in 1854. (Riemann's is the version commonly taught in calculus courses.) Lebesgue's definition agrees with the Riemann integral when the latter exists, but extends to a broader class of integrable functions.

A full course would be needed to develop this integration theory properly. In this section, we present only a brief sketch of the Lebesgue theory, with the focus on the features most relevant for applications to PDE.

The Lebesgue integral is based on a generalized notion of volume for subsets of $\mathbb{R}^n$, which can be defined in terms of approximation by rectangles. For a rectangular subset in $\mathcal{R} \subset \mathbb{R}^n$, let $\mathrm{vol}(\mathcal{R})$ denote the usual notion of volume, the product of the lengths of the sides. (It is conventional to use "volume" as a general term when the dimension is arbitrary.) The volume of a subset $A \subset \mathbb{R}^n$ can be overestimated by covering the set with rectangles, as illustrated in Fig. 7.1. The (n-dimensional) *measure* of A is defined by taking the infimum of these overestimates,

$$m_n(A) := \inf\left\{\sum_{j=1}^{\infty} \mathrm{vol}(\mathcal{R}_j);\ A \subset \bigcup_{j=1}^{\infty} \mathcal{R}_j\right\}. \tag{7.4}$$

For a bounded region with C^1 boundary, the definition (7.4) reproduces the notion of volume used in multivariable calculus. Note that the concept of measure is dependent on the dimension. The measure of a line segment in $\mathbb{R}^1$ is the length, but a line segment has measure zero in $\mathbb{R}^n$ for $n \geq 2$.

There is a major technicality in the application of (7.4). In order to make the definition of measure consistent with respect to basic set operations, we cannot apply it to all possible subsets of $\mathbb{R}^n$. Instead, the definition is restricted to a special class of *measurable* sets. Lebesgue gave a criterion for measurability that rules out certain exotic sets for which volume is ill-defined. Fortunately, these sets are so exotic that we are unlikely to encounter them in normal usage. All open and closed sets in $\mathbb{R}^n$ are included in the measurable category, as are any sets constructed from them by basic set operations of union and intersection.

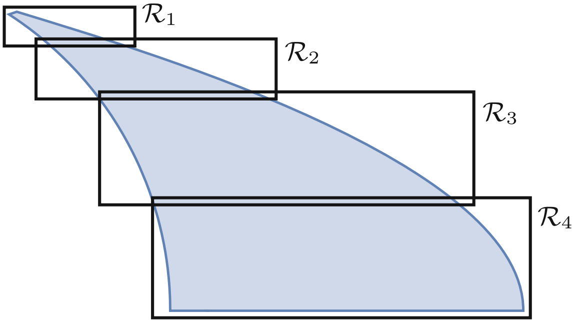


Fig. 7.1 Covering a set with *rectangles*

The *characteristic function* of a set $A \subset \mathbb{R}^n$ is defined by

$$\chi_A(\boldsymbol{x}) := \begin{cases} 1, & \boldsymbol{x} \in A, \\ 0, & \text{otherwise.} \end{cases} \tag{7.5}$$

The measure can be used to define the integral of a characteristic function,

$$\int_{\mathbb{R}^n} \chi_A \, d^n\boldsymbol{x} := m_n(A),$$

provided A is a measurable set. The integral of a general function is then built from approximations by linear combinations of characteristic functions. In order to construct these approximations, we need to use a restricted class of functions. A function $f : \Omega \to \mathbb{C}$ is called *measurable* if the preimage $f^{-1}(\mathcal{R})$ is a measurable subset of Ω for every rectangle $\mathcal{R} \subset \mathbb{C}$. Every Riemann-integrable function is measurable in the Lebesgue sense, so the measurable class includes all functions encountered in a traditional calculus class. Henceforth, whenever we write $f : \Omega \to \mathbb{C}$ or $\mathbb{R}$, we will assume implicitly that f is measurable.

With this basic picture in mind, we will ask the reader to accept certain important consequences of the Lebesgue definition without further justification. In examples and exercises we will confine our attention to functions for which ordinary Riemannian integrals exist.

It is standard practice when working with function spaces related to integration to make an equivalence:

$$f \equiv g \iff f = g \text{ except on a set of measure zero.} \tag{7.6}$$

For example, in $\mathbb{R}$ the characteristic functions of the intervals (a, b) and $[a, b]$ are equivalent. In measure theory, a property is said to hold *almost everywhere* if it fails only on a set of measure zero. The equivalence (7.6) amounts to identifying functions that agree almost everywhere.

If functions f and g satisfy

$$\int |f - g| \; d^n\boldsymbol{x} = 0,$$

then there is no way to distinguish them in terms of integration. The definition (7.6) is motivated by the following:

Lemma 7.2 *For measurable functions $f, g : \Omega \to \mathbb{C}$ with $\Omega \subset \mathbb{R}^n$,*

$$\int_\Omega |f - g| \; d^n\boldsymbol{x} = 0$$

if and only if $f \equiv g$.

7.3 L^p Spaces

A function $f : \Omega \to \mathbb{C}$ is defined to be *integrable* if its integral converges absolutely, i.e.,

$$\int_\Omega |f| \; d^n\boldsymbol{x} < \infty.$$

For $p \geq 1$, we define the space of "p-integrable" functions by

$$L^p(\Omega) := \left\{ f : \Omega \to \mathbb{C};\; \int_\Omega |f|^p \; d^n\boldsymbol{x} < \infty \right\}, \tag{7.7}$$

with the understanding that functions in L^p are identified according to the equivalence (7.6). The space $L^p(\Omega)$ is clearly closed under scalar multiplication. Closure under addition is a consequence of the convexity of the function $x \mapsto |x|^p$ for $p \geq 1$, which implies the inequality

$$\left| \frac{f+g}{2} \right|^p \leq \frac{|f|^p + |g|^p}{2}.$$

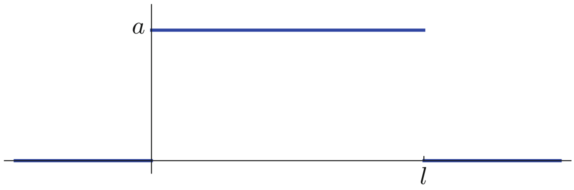
Hence $L^p(\Omega)$ is a complex vector space for $p \geq 1$.

The L^p norm is defined by

$$\|f\|_p := \left(\int_\Omega |f|^p \; d^n\boldsymbol{x} \right)^{\frac{1}{p}}.$$

To check that this is really a norm, we first note that Lemma 7.2 implies positive definiteness (N1) because of the equivalence relation (7.6). Homogeneity (N2) is satisfied because of cancellation between the powers p and $1/p$.

Fig. 7.2 Step function



The triangle inequality (N3) is clear for $p = 1$ because $|f + g| \leq |f| + |g|$. And for $p = 2$ it follows from Cauchy-Schwarz inequality, because $\|\cdot\|_2$ is associated to the inner product

$$\langle f, g\rangle := \int_\Omega f\bar{g}\, d^n\boldsymbol{x}. \tag{7.8}$$

In general the L^p triangle inequality,

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p\,,$$

is called the *Minkowski inequality* and holds for $p \geq 1$. We omit the proof because we are mainly concerned with the cases L^1 and L^2.

Example 7.3 To illustrate the distinction between the L^p norms, consider the function

$$h := a\chi_{[0,l]},$$

for $a, l > 0$, as shown in Fig. 7.2.

For general $p \geq 1$,

$$\begin{aligned}\|h\|_p &= \left[\int_0^l a^p dx\right]^{1/p} \\ &= al^{1/p}.\end{aligned} \tag{7.9}$$

If we think of h as a density function, then the L^1 norm gives the total mass $\|h\|_1 = al$. The sensitivity of $\|\cdot\|_p$ to the spread of the function decreases as p increases, as illustrated by the fact that

$$\lim_{p\to\infty} \|h\|_p = a,$$

For large p, the L^p norms increasingly become measures of local concentration rather than mass. ◊

Example 7.3 suggests the possibility of defining a space L^∞ that is a limiting case of the L^p spaces, with a norm that generalizes the sup norm (7.3). The sup norm itself does not respect the equivalence (7.6), so we must modify the definition to define a norm consistent with the other L^p spaces.

For a function $h : \Omega \to \mathbb{R}$, the *essential supremum* is

$$\text{ess-sup}(h) := \inf \Big\{ a \in \mathbb{R};\ \{h > a\} \text{ has measure zero} \Big\}. \tag{7.10}$$

Note that $\{h > a\}$ has measure zero precisely when h is equivalent to a function bounded by a. The value ess-sup(h) is thus the least upper bound among all functions equivalent to h. For continuous functions the essential supremum reduces to the supremum.

For $f : \Omega \to \mathbb{C}$, we define

$$\|f\|_\infty := \text{ess-sup } |f|. \tag{7.11}$$

The normed vector space $L^\infty(\Omega)$ consists of functions which are "essentially bounded",

$$L^\infty(\Omega) := \Big\{ f : \Omega \to \mathbb{C};\ \|f\|_\infty < \infty \Big\}, \tag{7.12}$$

subject to the equivalence (7.6).

Collectively, the L^p spaces play a vital role in the analysis of PDE. The different norms can be thought of as a collection of measuring tools. Although the full toolkit is needed for many applications, for this book we will rely on the cases $p = 1, 2$, or ∞.

Example 7.4 The Schrödinger equation in $\mathbb{R}^n$ describes the evolution of a quantum-mechanical wave function $\psi(t, \boldsymbol{x})$:

$$-i \frac{\partial \psi}{\partial t} = \Delta \psi.$$

In Exercise 4.7 we saw that solutions have constant spatial L^2 norms,

$$\|\psi(t, \cdot)\|_2 = \|\psi(0, \cdot)\|_2,$$

which corresponds to the conservation of total probability. On the other hand, solutions also satisfy a *dispersive estimate*

$$\|\psi(t, \cdot)\|_\infty \le C t^{-n/2} \|\psi(0, \cdot)\|_1,$$

for all $t > 0$, with C a dimensional constant. The norm on the left measures the peak amplitude of the wave. By the estimate on the right, this amplitude is bounded in terms of the mass and decays as a function of time. In general, dispersive estimates describe the spreading of solutions as a function of time. ♢

It is conventional to represent elements of L^p as ordinary functions, even though each element is actually an equivalence class of functions identified under (7.6). This usually causes no trouble because equivalent functions give the same results in integrals.

One point that requires clarification, however, is the issue of continuity or differentiability of functions in L^p. Under (7.6), a C^m function is equivalent to a class of functions which are not even continuous. To account for this technicality, we adopt the convention that if a function in L^p is equivalent to a continuous function, then the continuous representative is used by default. This is unambiguous because the continuous representative is unique when it exists. Under this convention, the statement that $f \in L^p$ is a C^m function really means that f admits a continuous representative which is C^m.

7.4 Convergence and Completeness

In a normed vector space V, convergence of a sequence $v_n \to v$ means

$$\lim_{n\to\infty} \|v_n - v\| = 0. \tag{7.13}$$

We might also write this as

$$v = \lim v_n,$$

provided the choice of norm is clear.

It frequently proves useful to approximate L^p functions by smooth functions. For $p \geq 1$ there is a natural inclusion

$$C^\infty_{\text{cpt}}(\Omega) \subset L^p(\Omega),$$

because continuous functions on a compact set are bounded. The Lebesgue theory gives the following:

Theorem 7.5 *Assume* $1 \leq p < \infty$. *For a function* $f \in L^p(\Omega)$ *there exists an approximating sequence* $\{\psi_k\} \subset C^\infty_{\text{cpt}}(\Omega)$, *such that*

$$\lim_{k\to\infty} \|\psi_k - f\|_p = 0.$$

A subset W of a normed vector space V is called *dense* if every $v \in V$ can be obtained as a limit of a sequence in W. Theorem 7.5 thus states that $C^\infty_{\text{cpt}}(\Omega)$ is dense in $L^p(\Omega)$ for $p \in [1, \infty)$.

In PDE applications, a common method of proving the existence of a solution is to construct a sequence of approximate solutions, and then establish convergence of this sequence with respect to an appropriate norm. We cannot simply use the definition (7.13) to check convergence in this situation, because the limiting function may not exist. It is therefore crucial to be able to deduce convergence using only the sequence itself.

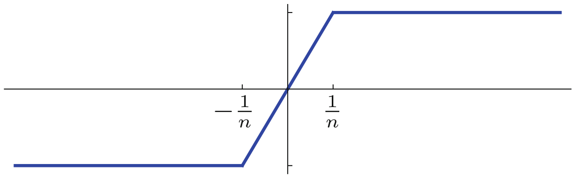


Fig. 7.3 A Cauchy sequence with respect to $\|\cdot\|_1$

The most useful tool for this purpose is a slightly weaker form of convergence. A sequence $\{v_k\} \subset V$ is said to be *Cauchy* if the difference between elements converges to zero: given $\varepsilon > 0$ there exists an N such that $k, m \geq N$ implies

$$\|v_k - v_m\| < \varepsilon.$$

This Cauchy condition is sometimes written as a double limit,

$$\lim_{k,m\to\infty} \|v_k - v_m\| = 0.$$

Every convergent sequence is Cauchy. This is because the triangle inequality implies

$$\begin{aligned} \|v_k - v_m\| &= \|v_k - v + v - v_m\| \\ &\leq \|v_k - v\| + \|v - v_m\| . \end{aligned}$$

If the sequence converges then the terms on the right are arbitrarily small for k and m sufficiently large.

In $\mathbb{R}^n$, it follows from the completeness axiom for real numbers that all Cauchy sequences are convergent. (See Theorem A.3.) This property does not necessarily hold in a general normed vector space, as the following demonstrates.

Example 7.6 Consider the space $C^0[-1, 1]$ equipped with the L^1 norm $\|\cdot\|_1$. For $n \in \mathbb{N}$ define the functions

$$f_n(x) = \begin{cases} -1, & x < -\frac{1}{n}, \\ nx & -\frac{1}{n} \leq x \leq \frac{1}{n}, \\ 1, & x > \frac{1}{n}, \end{cases}$$

as illustrated in Fig. 7.3.

We can see that the sequence $\{f_n\}$ is Cauchy by computing

$$\begin{aligned} \|f_k - f_m\|_1 &= \int_{-1}^{1} |f_k - f_m| \, dx \\ &= \left| \frac{1}{k} - \frac{1}{m} \right| . \end{aligned}$$

However, for $f \in C^0[-1, 1]$,

$$\lim_{k\to\infty} \|f_k - f\|_1 = \int_{-1}^{0} |f + 1|\, dx + \int_{0}^{1} |f - 1|\, dx.$$

This limit equals 0 only if $f(x) = -1$ for $x < 0$ and $f(x) = 1$ for $x > 0$. That is not possible for f continuous. Therefore the sequence $\{f_n\}$ does not converge in $C^0[-1, 1]$. ♢

A normed vector space V is *complete* if all Cauchy sequences in V converge within V. Theorem A.3 implies that Euclidean $\mathbb{R}^n$ is complete in this sense. For L^p spaces the Lebesgue integration theory gives the following result.

Theorem 7.7 *For a domain $\Omega \subset \mathbb{R}^n$, the normed vector space $L^p(\Omega)$ is complete for each $p \in [1, \infty]$.*

In functional analysis, a complete normed vector space is called a *Banach space* and a complete inner product space is called a *Hilbert space*. Thus Theorem 7.7 could be paraphrased as the statement that $L^p(\Omega)$ is a Banach space. The inner product space $L^2(\Omega)$ is a Hilbert space.

A subspace $W \subset V$ is *closed* if it contains the limit of every sequence in W that converges in V.

Lemma 7.8 *If V is a complete normed vector space and $W \subset V$ is a closed subspace, then W is complete with respect to the norm of V.*

Proof Suppose $\{w_k\} \subset W$ is a Cauchy sequence. The sequence is also Cauchy in V, and so converges to some $v \in V$ by the completeness of V. Since W is closed, $v \in W$. □

The L^p function spaces have discrete counterparts, denoted by ℓ^p, whose elements are sequences. To a sequence $(a_1, a_2, \dots)$ of complex numbers we associate the function $a : \mathbb{N} \to \mathbb{C}$ defined by $j \mapsto a_j$. The ℓ^p norm of this function is

$$\|a\|_{\ell^p} := \left[\sum_{j=1}^{\infty} |a_j|^p\right]^{\frac{1}{p}},$$

The corresponding vector spaces are

$$\ell^p(\mathbb{N}) := \left\{a : \mathbb{N} \to \mathbb{C};\ \|a\|_{\ell^p} < \infty\right\}, \tag{7.14}$$

for $p \geq 1$. It is possible to prove directly that $\ell^p(\mathbb{N})$ is complete, but this can also be deduced easily from Lemma 7.8. We interpret $\ell^p(\mathbb{N})$ as a closed subspace of $L^p(\mathbb{R})$ consisting of functions which are constant on each interval $[j, j + 1)$ for $j \in \mathbb{N}$ and zero on $(-\infty, 0)$. On this subspace the L^p norm reduces to the ℓ^p norm, so that Lemma 7.8 implies that $\ell^p(\mathbb{N})$ is complete. In particular, $\ell^2(\mathbb{N})$ is a Hilbert space with the inner product

$$\langle a, b\rangle_{\ell^2} := \sum_{j=1}^{\infty} a_j \overline{b_j}.$$

7.5 Orthonormal Bases

Let H be an infinite-dimensional complex Hilbert space. A sequence of vectors $\{e_1, e_2, \dots\} \subset H$ is *orthonormal* if

$$\langle e_j, e_k\rangle = \begin{cases} 1, & j = k, \\ 0, & j \neq k, \end{cases} \tag{7.15}$$

for all $j, k \in \mathbb{N}$. An *orthonormal basis* for H is an orthonormal sequence such that each $v \in H$ admits a unique representation as a convergent series,

$$v = \sum_{j=1}^{\infty} c_j e_j, \tag{7.16}$$

with $c_j \in \mathbb{C}$.

As we will see in Sect. 7.6, the sets of eigenfunctions of certain differential operators naturally form orthonormal sequences with respect to the L^2 inner product. For example the sine eigenfunctions appearing in Theorem 5.2 have this property. If a sequence of eigenfunctions forms a basis, then we can expand general functions in terms of eigenfunctions.

Suppose we are given an orthonormal sequence $\{e_j\} \subset H$, and we would like to show that this forms a basis. To represent an element $v \in H$ in the form (7.16), we must decide how to choose the coefficients c_j. This works in much the same way as it does in finite dimensions. By the orthonormality property (7.15), we can compute that

$$\left\langle \sum_{j=1}^{n} c_j e_j, e_k \right\rangle = c_k \tag{7.17}$$

for all $n \geq k$. Assuming that $\sum c_j e_j$ converges to v in H, we can take the limit $n \to \infty$ in (7.17) to compute

$$\langle v, e_k\rangle = c_k. \tag{7.18}$$

Based on this calculation, we assign coefficients to v by setting

$$c_j[v] := \langle v, e_j\rangle. \tag{7.19}$$

The corresponding partial sums for $n \in \mathbb{N}$ are denoted by

$$S_n[v] := \sum_{j=1}^{n} c_j[v] e_j. \tag{7.20}$$

The condition that $\{e_j\}$ is a basis is equivalent to the convergence of $S_n[v] \to v$ in H for every $v \in H$.

Theorem 7.9 (Bessel's inequality) *Assume that $\{e_j\}$ is an orthonormal sequence in an infinite-dimensional Hilbert space H. For $v \in H$, the series $\sum \left|c_j[v]\right|^2$ converges and the limit satisfies*

$$\sum_{j=1}^{\infty} \left|c_j[v]\right|^2 \le \|v\|^2 .$$

Equality holds if and only if $S_n[v] \to v$ in H.

Proof Using the sesquilinearity (I3) of the inner product, we can expand

$$\begin{aligned} \left\|v - S_n[v]\right\|^2 &:= \langle v - S_n[v], v - S_n[v]\rangle \\ &= \langle v, v\rangle - \langle S_n[v], v\rangle - \langle v, S_n[v]\rangle + \langle S_n[v], S_n[v]\rangle , \end{aligned}$$

for $n \in \mathbb{N}$. By the definition (7.20) of $S_n[v]$ and the orthonormality condition (7.15),

$$\langle S_n[v], v\rangle = \langle v, S_n[v]\rangle = \langle S_n[v], S_n[v]\rangle = \sum_{j=1}^{n} \left|c_j[v]\right|^2.$$

We thus conclude that

$$\left\|v - S_n[v]\right\|^2 = \|v\|^2 - \sum_{j=1}^{n} \left|c_j[v]\right|^2. \tag{7.21}$$

Since the left-hand side is positive, the identity (7.21) shows that

$$\sum_{j=1}^{n} \left|c_j[v]\right|^2 \le \|v\|^2 ,$$

for all $n \in \mathbb{N}$. The partial sums of the series $\sum \left|c_j[v]\right|^2$ are thus bounded and the terms are all positive. Hence the series converges by the monotone sequence theorem, to a limit satisfying the claimed bound,

$$\sum_{j=1}^{\infty} \left|c_j[v]\right|^2 \leq \|v\|^2 .$$

To complete the proof, note that $S_n[v] \to v$ in H means that the limit as $n \to \infty$ of the left-hand side of (7.21) is zero. Hence $S_n[v] \to v$ if and only if

$$\lim_{n\to\infty} \sum_{j=1}^{n} \left|c_j[v]\right|^2 = \|v\|^2 . \qquad \square$$

The combination of completeness and Bessel's inequality leads to an alternative characterization of a basis that is easier to apply.

Theorem 7.10 *Suppose H is an infinite-dimensional Hilbert space. An orthonormal sequence in H forms a basis if and only if* 0 *is the only element of H that is orthogonal to all vectors in the sequence.*

Proof Assume first that $\{e_j\}$ forms a basis, so that every $v \in H$ can be written as a convergent sum $\sum c_j[v]e_j$. If v is orthogonal to all of the vectors e_j, then $c_j[v] = 0$ for all j by (7.19). Hence $v = 0$.

To establish the converse statement, let $\{e_j\}$ be an orthonormal sequence. For $v \in H$, Bessel's inequality implies

$$\sum_{j=1}^{\infty} \left|c_j[v]\right|^2 \leq \|v\|^2 < \infty. \tag{7.22}$$

For $n \leq m$,

$$\begin{aligned}\left\| S_m[v] - S_n[v] \right\|^2 &= \left\| \sum_{j=n+1}^{m} c_j[v]e_j \right\|^2 \\ &= \sum_{j=n+1}^{m} \left|c_j[v]\right|^2.\end{aligned}$$

Hence (7.22) implies that

$$\lim_{m,n\to\infty} \left\| S_m[v] - S_n[v] \right\|^2 = 0,$$

meaning that the sequence $\{S_n[v]\}$ is Cauchy in H. By completeness of H this implies that $S_n[v] \to \tilde{v}$ for some $\tilde{v} \in H$.

Now assume that 0 is the only vector orthogonal to e_j for all j. For $n \geq j$ we have

$$\begin{aligned}\left\langle v - S_n[v], e_j \right\rangle &= \left\langle v, e_j \right\rangle - \left\langle S_n[v], e_j \right\rangle \\ &= c_j[v] - c_j[v] \\ &= 0.\end{aligned}$$

Taking the limit as $n \to \infty$ with j fixed gives

$$\left\langle v - \tilde{v}, e_j \right\rangle = 0.$$

Thus $v - \tilde{v}$ is orthogonal to every e_j, implying that $v = \tilde{v}$. This proves that $S_n[v] \to v$ in H for each $v \in H$, and thus $\{e_j\}$ is a basis. □

7.6 Self-adjointness

The process of forming a basis from eigenvectors of an operator should be familiar from linear algebra; for a finite-dimensional matrix this is called *diagonalization*. Let us briefly recall the basic facts for the finite-dimensional case. A complex $n \times n$ matrix A is *self-adjoint* (also called Hermitian) if the matrix is equal to its conjugate transpose. In terms of the Euclidean inner product (7.1) this means precisely that

$$\langle Au, v \rangle = \langle u, Av \rangle \tag{7.23}$$

for all $u, v \in \mathbb{C}^n$. (In the real case self-adjoint is the same as *symmetric*.)

The spectral theorem in linear algebra says that for a self-adjoint matrix A there exists an orthonormal basis for $\mathbb{C}^n$ consisting of eigenvectors for A, with real eigenvalues. Functional analysis allows a powerful extension of this result, that applies in particular to certain differential operators acting on L^2 spaces. The full spectral theorem for Hilbert spaces is too technical for us to state here, but we will prove a version of this for the Laplacian on bounded domains later in Sect. 11.5.

Self-adjointness remains important as a hypothesis for the more general spectral theorem, but even this condition becomes rather technical in the Hilbert space setting. The issues arise from the fact that differentiable operators cannot act on the whole space $L^2(\Omega)$ because L^2 functions need not be differentiable. We will avoid these complexities, by focusing on the Laplacian and restricting our attention to C^2 functions.

Lemma 7.11 *Suppose that $\Omega \in \mathbb{R}^n$ is a bounded domain with C^1 boundary. If $u, v \in C^2(\overline{\Omega})$ both satisfy either Dirichlet or Neumann boundary conditions on $\partial\Omega$, then*

$$\langle \Delta u, v \rangle = \langle u, \Delta v \rangle . \tag{7.24}$$

Proof By Green's first identity (Theorem 2.10),

$$\int_{\Omega} \left[u\overline{\Delta v} - \bar{v}\Delta u \right] d^n\boldsymbol{x} = \int_{\partial\Omega} \left[u\frac{\partial \bar{v}}{\partial \nu} - \bar{v}\frac{\partial u}{\partial \nu} \right] dS. \tag{7.25}$$

The Dirichlet conditions require that

$$u|_{\partial\Omega} = v|_{\partial\Omega} = 0,$$

implying the vanishing of the right-hand side of (7.25). Similarly, the Neumann conditions

$$\left.\frac{\partial u}{\partial \nu}\right|_{\partial\Omega} = \left.\frac{\partial v}{\partial \nu}\right|_{\partial\Omega} = 0 \tag{7.26}$$

also imply that the integrand on the right vanishes. □

Boundary conditions for which (7.24) holds are called *self-adjoint* boundary conditions (for the Laplacian). Formally, (7.24) resembles the matrix condition (7.23), but of course there is no analog of boundary conditions in the matrix case. The proper definition of self-adjointness in functional analysis involves a more precise specification of the domain on which Δ acts and (7.24) holds. Even without going into these details, we can still draw some meaningful conclusions from Lemma 7.11.

Lemma 7.12 *Suppose* $\{\lambda_j\}$ *is a sequence of eigenvalues of* $-\Delta$ *on a bounded domain* $\Omega \subset \mathbb{R}^n$*, with eigenvectors in* $C^2(\overline{\Omega})$ *subject to a self-adjoint boundary condition. Then* $\lambda_j \in \mathbb{R}$ *and, after possible rearrangement, the eigenvectors form an orthonormal sequence in* $L^2(\Omega)$.

Furthermore, $\lambda_j > 0$ *for Dirichlet conditions, and* $\lambda_j \geq 0$ *for Neumann.*

Proof Suppose we have a sequence $\{\phi_j\} \subset C^2(\overline{\Omega})$ satisfying

$$-\Delta\phi_j = \lambda_j\phi_j.$$

The condition (7.24) implies that for $j, k \in \mathbb{Z}$,

$$\langle \Delta\phi_j, \phi_k \rangle = \langle \phi_j, \Delta\phi_k \rangle.$$

By the eigenvalue property this reduces to

$$\left(\lambda_j - \overline{\lambda_k}\right) \langle \phi_j, \phi_k \rangle = 0. \tag{7.27}$$

For $j = k$ the inner product equals $\|\phi_j\|_2^2 > 0$, implying that $\lambda_j \in \mathbb{R}$ for all j. We can thus drop the conjugation in (7.27). If $\lambda_j \neq \lambda_k$, then this now implies that $\langle \phi_j, \phi_k \rangle = 0$.

If some of the λ_j's are equal, then every linear combination of the corresponding of eigenfunctions will still be an eigenfunction for the same value of λ_j. Hence we can rearrange the eigenfunctions sharing a common eigenvalue into an orthogonal set using the Gram-Schmidt procedure from linear algebra.

By multiplying the eigenfunctions by constants we can normalize so that $\|\phi_j\|_2 = 1$. The divergence theorem (Theorem 2.6) then implies

$$\begin{aligned}\lambda_j &= \langle -\Delta\phi_j, \phi_j\rangle \\ &= \int_\Omega \left|\nabla\phi_j\right|^2 d^n\boldsymbol{x} - \int_{\partial\Omega} \overline{\phi_j}\frac{\partial\phi_j}{\partial\nu}\, dS.\end{aligned}$$

Either Dirichlet or Neumann conditions will cause the second term to vanish, implying that $\lambda_j \geq 0$. If $\lambda_j = 0$ then the equation also shows that $\nabla\phi_j \equiv 0$, implying that ϕ_j is constant. In the Dirichlet case the only constant solution is trivial, $\phi_j \equiv 0$, but for Neumann conditions a nonzero constant is possible. □

Example 7.13 In Example 5.5 we found a set of eigenfunctions for a circular drumhead modeled by the unit disk, with Dirichlet boundary conditions. The eigenfunctions were given in polar coordinates by

$$\phi_{k,m}(r,\theta) := e^{ik\theta} J_k(j_{k,m}r), \quad k \in \mathbb{Z}, m \in \mathbb{N},$$

where $j_{k,m}$ is the mth positive zero of the Bessel function J_k. The eigenvalues of $-\Delta$ in this case are the values $j_{k,m}$. Since the only possible matches among the Bessel zeros are $j_{k,m} = j_{-k,m}$, these are the only potential non-orthogonal pairs.

Let us examine the orthogonality condition more explicitly. In polar coordinates, the L^2 inner product of two eigenfunctions is given by

$$\begin{aligned}\left\langle \phi_{k,m}, \phi_{k',m'}\right\rangle_{L^2} &= \int_0^1 \int_0^{2\pi} \phi_{k,m}(r,\theta)\overline{\phi_{k',m'}(r,\theta)}\, r\, d\theta\, dr \\ &= \int_0^1 \int_0^{2\pi} e^{i(k-k')\theta} J_k(j_{k,m}r) J_{k'}(j_{k',m'}r)\, r\, d\theta\, dr.\end{aligned}$$

Note that the eigenfunctions are clearly orthogonal when $k \neq k'$, because the θ integral vanishes in this case. If we set $k = k'$, then the θ integral is trivial and the inner product becomes

$$\left\langle \phi_{k,m}, \phi_{k,m'}\right\rangle_{L^2} = 2\pi \int_0^1 r J_k(j_{k,m}r) J_k(j_{k,m'}r)\, dr.$$

By Lemma 7.12 this integral vanishes for $m \neq m'$. The cancellations occur because of the oscillations, just as for sine functions, as Fig. 7.4 illustrates.

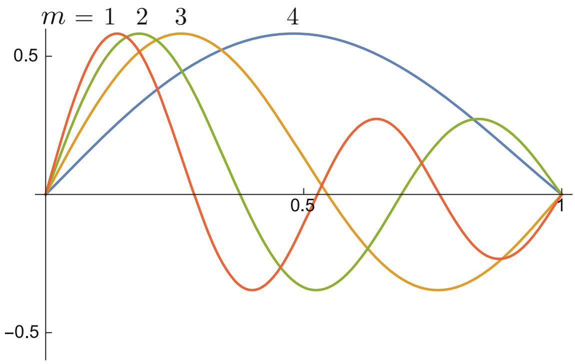


Fig. 7.4 Radial components $J_1(j_{1,m}r)$ of orthogonal eigenfunctions on the disk

7.7 Exercises

7.1 A norm $\|\cdot\|$ on a vector space V satisfies the *parallelogram law* if

$$\|v+w\|^2 + \|v-w\|^2 = 2\,\|v\|^2 + 2\,\|w\|^2,$$

for all $v, w \in V$.

(a) Show that a norm defined by an inner product as in (7.2) satisfies the parallelogram law.

(b) In $L^p(\mathbb{R})$, define the functions

$$f(x) = \chi_{[0,2]}, \qquad g(x) = \chi_{[0,1]} - \chi_{[1,2]}.$$

Use these to show that the parallelogram law fails for $\|\cdot\|_p$ if $p \neq 2$.

(c) Find an example to show that the parallelogram law fails for the sup norm (7.3).

7.2 Consider the sequence of functions on $\mathbb{R}$ defined by

$$f_n(x) = \begin{cases} ne^{-n^2x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Show that $f_n \to 0$ in $L^1(\mathbb{R})$ but not in $L^2(\mathbb{R})$.

7.3 Consider the sequence of functions on $\mathbb{R}$ defined by

$$g_n(x) = n^{-1}\chi_{[0,n]}.$$

Show that $g_n \to 0$ in $L^2(\mathbb{R})$ but not in $L^1(\mathbb{R})$.

7.4 Consider the sequence $f_n(x) = x^n$ for $x \in (0, 1)$. Show that $f_n \to 0$ in $L^p(0, 1)$ for each $p \in [1, \infty)$, but not for $p = \infty$.

7.5 Assume that $\Omega \subset \mathbb{R}^n$ is a bounded domain. Show that there is a constant $C > 0$ such that for $f \in L^2(\Omega)$,

$$\|f\|_1 \leq C\, \|f\|_2 .$$

This implies in particular that $L^2(\Omega) \subset L^1(\Omega)$. Find an example to show that this result does not hold for Ω unbounded.

7.6 As an application of the Cauchy-Schwarz inequality, we can use the quantity η defined in Exercise 6.4 to show that solutions of the heat equation with fixed boundary values are uniquely determined by the values at time $t = T > 0$. Under the hypotheses from that exercise, suppose that u solves the heat equation with

$$u|_{t=T} = 0, \quad u|_{x\in\partial\Omega} = 0.$$

The goal is to show that these assumptions imply $u = 0$ for all t.

(a) Use the Cauchy-Schwarz inequality to deduce that

$$\eta'(t)^2 \leq 4\eta(t) \int_\Omega \left|\frac{\partial u}{\partial t}\right|^2 d^n\boldsymbol{x}.$$

where η is defined as in (6.31).

(b) Show that

$$\eta''(t) = 4 \int_\Omega \left|\frac{\partial u}{\partial t}\right|^2 d^n\boldsymbol{x},$$

so that the inequality from (a) becomes

$$\eta'(t)^2 \leq \eta(t)\eta''(t). \tag{7.28}$$

(c) Suppose that $\eta(0) > 0$. Then by continuity $\log \eta(t)$ is defined at least in some neighborhood of $t = 0$. Using (7.28), show that

$$(\log \eta(t))'' \geq 0.$$

This implies that $\log \eta(t)$ is bounded below by its tangent lines. In particular

$$\log \eta(t) \geq \log \eta(0) + \frac{\eta'(0)}{\eta(0)} t,$$

which implies

$$\eta(t) \geq \eta(0)e^{-ct},$$

for $c = -\eta'(0)/\eta(0) > 0$. Thus if $\eta(0) > 0$ then η is strictly positive for all $t \geq 0$.

(d) Conclude from (c) that if $\eta(T) = 0$, then $\eta(t) = 0$ for all t, and deduce that $u = 0$.

7.7 Recall the radial decomposition formula (2.10). We can use this to get a basic picture of the degree of singularity or decay at infinity that is allowed in each L^p.

(a) For $\gamma \in \mathbb{R}$ consider the function

$$g(\boldsymbol{x}) := \begin{cases} r^\gamma, & r \leq 1, \\ 0, & r \geq 1. \end{cases}$$

For what values of γ and $p \in [1, \infty]$ is $g \in L^p(\mathbb{R}^n)$?

(b) For $\gamma \in \mathbb{R}$ consider the function

$$h(\boldsymbol{x}) := \begin{cases} 0, & r \leq 1, \\ r^\gamma, & r \geq 1. \end{cases}$$

For what values of γ and $p \in [1, \infty]$ is $h \in L^p(\mathbb{R}^n)$?

7.8 Consider the eigenfunctions given by (5.5) with $\ell = \pi$.

(a) Show that

$$\phi_n(x) := \sqrt{\frac{2}{\pi}} \sin(nx), \quad n \in \mathbb{N},$$

defines an orthonormal sequence in $L^2(0, \pi)$. (Hint: recall the trigonometric identity $\sin(\alpha)\sin(\beta) = \frac{1}{2}\big[\cos(\alpha - \beta) - \cos(\alpha + \beta)\big]$.)

(b) For the function $u \equiv 1$, compute the corresponding expansion coefficients,

$$c_k[1] := \langle 1, \phi_k \rangle .$$

Under Theorem 7.9, what explicit summation condition corresponds to the convergence $S_n[1] \to 1$ in $L^2(0, \pi)$?